Приложения дискретной математики

А. В. Пастор

Санкт-Петербургский политехнический университет Петра Великого

Осенний семестр 2022

Глава 1.
Теория чисел и криптография

- Дополнительные материалы:
  - И. М. Виноградов, *Основы теории чисел*. М.-Л., Гостехиздат, 1952.
  - Н. Коблиц, *Курс теории чисел и криптографии*. М.: ТВП, 2001.

- Слайды по данному курсу будут публиковаться по адресу
  - `https://logic.pdmi.ras.ru/~pastor/fmf/dm/`

Некоторые обозначения

Основные числовые множества

- $\mathbb{N}$ — множество натуральных чисел;
- $\mathbb{Z}$ — множество целых чисел;
- $\mathbb{Q}$ — множество рациональных чисел;
- $\mathbb{R}$ — множество вещественных чисел;
- $\mathbb{C}$ — множество комплексных чисел.

Некоторые подмножества $\mathbb{Z}$

- $\mathbb{N}_0 \stackrel{\text{def}}{=} \mathbb{N} \cup \{0\}$ — множество целых неотрицательных чисел;
- $[a..b] \stackrel{\text{def}}{=} \{x \in \mathbb{Z} \mid a \leq x \leq b\}$, где $a, b \in \mathbb{Z}$ и $a \leq b$ — множество целых чисел от a до b.

Делимость целых чисел (1/2)

Определение 1

- Пусть $a, b \in \mathbb{Z}$. Число a *делится* на b, если $b \neq 0$ и существует целое число c такое, что $a = bc$.
  - Обозначение: $a \vdots b$ или $b \mid a$.

Определение 2

- Пусть $a_1, a_2, \ldots, a_n \in \mathbb{Z}$, $d \in \mathbb{N}$. Число d называется
  - *общим делителем* чисел $a_1, a_2, \ldots, a_n$, если $\forall i \in [1..n]\ (a_i \vdots d)$;
  - *наибольшим общим делителем* чисел $a_1, a_2, \ldots, a_n$, если оно является общим делителем этих чисел, и больше любого другого их общего делителя.
    - Обозначение: $d = (a_1, a_2, \ldots, a_n)$.
- Числа $a, b \in \mathbb{Z}$ называются *взаимно простыми*, если $(a, b) = 1$.
- Числа $a_1, a_2, \ldots, a_n \in \mathbb{Z}$ называются
  - *взаимно простыми в совокупности*, если $(a_1, a_2, \ldots, a_n) = 1$;
  - *попарно взаимно простыми*, если $\forall i, j\ (i \neq j \to (a_i, a_j) = 1)$.

Делимость целых чисел (2/2)

Пример

- Числа 6, 10, 15 взаимно просты в совокупности, но не попарно.
  - Более того, никакие два из этих чисел не взаимно просты.

Определение 3

- Пусть $a_1, a_2, \ldots, a_n \in \mathbb{Z}$, $s \in \mathbb{N}$. Число s называется
  - *общим кратным* чисел $a_1, a_2, \ldots, a_n$, если $\forall i \in [1..n]$ $(s \vdots a_i)$.
  - *наименьшим общим кратным* чисел $a_1, a_2, \ldots, a_n$, если оно является общим кратным этих чисел, и меньше любого другого их общего кратного.
    - Обозначение: $s = [a_1, a_2, \ldots, a_n]$.

Деление с остатком

Теорема 1 (о делении с остатком)

Пусть $a, b \in \mathbb{Z}$, $b \neq 0$. *Тогда существует единственная пара целых чисел* q, r *такая, что* $a = bq + r$ *и* $0 \leq r < |b|$.

Определение 4

Число q из условия предыдущей теоремы называется *неполным частным*, а число r — *остатком* от деления a на b.

Лемма

$a, b, c, k \in \mathbb{Z}$, $a = bk + c$. *Тогда* $(a, b) = (b, c)$.

Алгоритм Евклида

- Вход: натуральные числа a и b.
  - На первом шаге алгоритма a делится с остатком на b. Обозначим получившиеся неполное частное и остаток через q_1 и r_1 соответственно.
  - На втором шаге b делится с остатком на r_1, получившиеся неполное частное и остаток обозначим через q_2 и r_2 соответственно.
  - На третьем шаге r_1 делится с остатком на r_2, получившиеся неполное частное и остаток обозначим через q_3 и r_3 соответственно. И т.д.
  - На k-м шаге r_{k-2} делится с остатком на r_{k-1}, получившиеся неполное частное и остаток обозначим через q_k и r_k соответственно.
  - Алгоритм заканчивает работу когда очередной остаток станет равным нулю.
- Выход: последний ненулевой остаток (обозначим его r_{n+1}).

Теорема 2

Для любых натуральных чисел a и b алгоритм Евклида заканчивает работу за конечное число шагов и дает на выходе наибольший общий делитель чисел a и b.

Схема работы алгоритма Евклида

- Вход: $a, b \in \mathbb{N}$.

$$
\begin{array}{ll}
a = bq_1 + r_1 & 0 \le r_1 < b \\
b = r_1 q_2 + r_2 & 0 \le r_2 < r_1 \\
r_1 = r_2 q_3 + r_3 & 0 \le r_3 < r_2 \\
\quad\vdots & \quad\vdots \\
r_{k-2} = r_{k-1} q_k + r_k & 0 \le r_k < r_{k-1} \\
\quad\vdots & \quad\vdots \\
r_{n-1} = r_n q_{n+1} + r_{n+1} & 0 \le r_{n+1} < r_n \\
r_n = r_{n+1} q_{n+2} + 0 &
\end{array}
$$

- Выход: $r_{n+1} = (a, b)$.

Время работы алгоритма Евклида (1/3)

- Напомним, что последовательность Фибоначчи задается рекуррентным соотношением $F_0 = F_1 = 1$ и $F_{n+1} = F_n + F_{n-1}$ при $n > 0$.

Теорема 3 (Ламе)

Пусть $a > b > 0$ и $b < F_{k+1}$. Тогда в алгоритме Евклида для чисел a и b производится не более k делений с остатком.

Доказательство. Индукция по k.
k=1: $b < F_2 = 2$, то есть $b = 1$. Тогда будет одно деление с остатком.
k=2: $b < F_3 = 3$, то есть $b \leq 2$. Тогда будет не более двух делений с остатком.
$k - 1, k \to k + 1$: Рассмотрим первое деление с остатком: $a = bq_1 + r_1$, $0 \leq r_1 < b$.

- Далее, возможны два случая: $r_1 < F_{k+1}$ или $r_1 \geq F_{k+1}$.
  1. $r_1 < F_{k+1}$. Тогда по индукционному предположению, в алгоритме Евклида для чисел b и r_1 будет не более k делений с остатком. То есть всего получится не более $k + 1$ делений с остатком.
  2. $r_1 \geq F_{k+1}$. Тогда $r_2 = b - r_1 q_2 \leq b - r_1 < F_{k+2} - F_{k+1} = F_k$. Следовательно, в алгоритме Евклида для чисел r_1 и r_2 будет не более $k - 1$ делений с остатком. То есть всего получится не более $k + 1$ делений с остатком. □

Время работы алгоритма Евклида (2/3)

Замечание

Если $a = F_{k+1}$ и $b = F_k$, то делений с остатком ровно k.
Таким образом, полученную оценку невозможно улучшить.

Утверждение

Для всех $n \in \mathbb{N}_0$ выполнено $F_{2n} \geq 2^n$.

Доказательство. Индукция по n. n=0: $F_0 = 1 = 2^0$; n=1: $F_2 = 2 = 2^1$.
$n \to n+1$: $F_{2n+2} = F_{2n+1} + F_{2n} > 2F_{2n} \geq 2 \cdot 2^n = 2^{n+1}$. □

Теорема 4

Если в двоичной записи каждого из чисел a и b не более β разрядов, то в ходе работы алгоритма Евклида для a и b будет выполнено $O(\beta^3)$ битовых операций.

Доказательство. Заметим, что $b < 2^\beta \leq F_{2\beta}$. Следовательно, при работе алгоритма Евклида будет выполнено не более $2\beta - 1$ делений с остатком.

- Поскольку каждое деление с остатком выполняется за $O(\beta^2)$ операций (деление в столбик), получаем $O(\beta^3)$ битовых операций. □

Время работы алгоритма Евклида (3/3)

Замечание

1. Более точную оценку для чисел Фибоначчи можно получить из формулы Бине (см. параграф 5.7.3 курса Дискретной математики):

$$F_n = \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2} \right)^{n+1} - \left(\frac{1-\sqrt{5}}{2} \right)^{n+1} \right).$$

   Из этой формулы следует, что $F_k \sim \frac{\varphi^{k+1}}{\sqrt{5}}$, где $\varphi = \frac{1+\sqrt{5}}{2}$ — *золотое сечение*.

2. Рассуждая несколько более аккуратно, можно получить более сильную оценку трудоемкости алгоритма Евклида: $O(\beta^2)$.

Линейное представление НОД (1/3)

Теорема 5

Пусть $a, b \in \mathbb{N}$. *Тогда существуют* $s, t \in \mathbb{Z}$, *такие, что* $as + bt = (a, b)$.

- Алгоритм нахождения линейного представления НОД основан на алгоритме Евклида.
  - Часто этот алгоритм называют расширенный алгоритм Евклида.

- $$\begin{aligned} a &= bq_1 + r_1 \\ b &= r_1 q_2 + r_2 \\ r_1 &= r_2 q_3 + r_3 \\ &\vdots \\ r_{n-2} &= r_{n-1} q_n + r_n \\ r_{n-1} &= r_n q_{n+1} + r_{n+1} \\ r_n &= r_{n+1} q_{n+2} + 0 \end{aligned}$$
- $$\begin{aligned} (a, b) = r_{n+1} &= r_{n-1} - r_n q_{n+1} = r_{n-1} - (r_{n-2} - r_{n-1} q_n) q_{n+1} = \\ &= r_{n-1}(1 + q_n q_{n+1}) - r_{n-2} q_{n+1} = \ldots = as + bt. \end{aligned}$$

Линейное представление НОД (2/3)

Замечание

- Коэффициенты s и t определены неоднозначно.
  - Например, $(3, 5) = 1 = 3 \cdot 2 - 5 \cdot 1 = -3 \cdot 3 + 5 \cdot 2$.

Следствие 1

Пусть $a, b \in \mathbb{N}$ и $d = (a, b)$. Тогда множество общих делителей чисел a и b совпадает с множеством делителей числа d.

Доказательство. Пусть $d = as + bt$ — линейное представление НОД.

- Тогда для любого общего делителя f чисел a и b получаем, что $a \vdots f$ и $b \vdots f$, откуда $d = as + bt \vdots f$.
- Обратно, если $d \vdots f$, то из $a \vdots d$ и $b \vdots d$ очевидно следует, что $a \vdots f$ и $b \vdots f$. $\square$

Линейное представление НОД (3/3)

Следствие 2

$a, b, c \in \mathbb{N}$, $(b, c) = 1$. *Тогда* $(ac, b) = (a, b)$.

Доказательство. Докажем, что $(ac, b) \vdots (a, b)$ и $(a, b) \vdots (ac, b)$.

- Заметим, что $ac \vdots (a, b)$ и $b \vdots (a, b)$.
  - Тогда по следствию 1 получаем, что $(ac, b) \vdots (a, b)$.
- Пусть $1 = bs + ct$ — линейное представление НОД.
  - Тогда $a = a(bs + ct) = abs + act \vdots (ac, b)$.
  - Очевидно также, что $b \vdots (ac, b)$.
  - Тогда по следствию 1 получаем, что $(a, b) \vdots (ac, b)$. $\square$

Простые числа

Определение 5

- Натуральное число называется *простым*,
если оно имеет ровно два натуральных делителя.
- Натуральное число называется *составным*,
если оно имеет больше двух натуральных делителей.
- Множество простых чисел обозначается $\mathbb{P}$.

Замечание

Число 1 не является ни простым, ни составным.

Теорема 6 (Евклид)

Простых чисел бесконечно много.

Теорема 7 (теорема Дирихле об арифметической прогрессии)

Пусть a и b — взаимно простые натуральные числа. Тогда среди чисел вида $ak + b$ (где $k \in \mathbb{N}$) бесконечно много простых. (б/д)

Основная теорема арифметики для целых чисел

Лемма

Пусть $a, b \in \mathbb{N}$, $p \in \mathbb{P}$, $ab \vdots p$. *Тогда либо* $a \vdots p$, *либо* $b \vdots p$.

Доказательство. Пусть $a \not\vdots p$. Тогда $(a, p) = 1$.

- Следовательно, $(b, p) = (ab, p) = p$, откуда $b \vdots p$. □

Следствие

Пусть $a_1, a_2, \ldots, a_n \in \mathbb{N}$, $p \in \mathbb{P}$, $a_1 a_2 \ldots a_n \vdots p$. *Тогда* $\exists i\, (a_i \vdots p)$.

Доказательство. Индукция по n. База для $n = 2$ доказана в лемме.
$n \to n + 1$: Пусть $(a_1 a_2 \ldots a_n) a_{n+1} \vdots p$.

- Тогда по лемме, либо $a_{n+1} \vdots p$ и все, что нужно, доказано,
либо $a_1 a_2 \ldots a_n \vdots p$, откуда по индукционному предположению
$\exists i \in [1..n]\, (a_i \vdots p)$. □

Теорема 8 (Основная теорема арифметики)

Любое натуральное число $n > 1$ *может быть представлено в виде произведения нескольких (возможно, одного) простых множителей, причем это представление единственно с точностью до порядка сомножителей.*

Основная теорема арифметики. Доказательство

Доказательство.

"$\exists$" Индукция по n. База для $n = 2$ очевидна.

$2, \ldots, n \to n+1$: Если $n+1 \in \mathbb{P}$, то разложение состоит из одного множителя.

- Пусть $n+1 = ab$, где $a, b \in [2..n]$.
- Тогда по индукционному предположению числа a и b можно разложить на простые множители: $a = p_1 \ldots p_k$ и $b = p_{k+1} \ldots p_{k+\ell}$.
- Следовательно, $n+1 = ab = p_1 \ldots p_k p_{k+1} \ldots p_{k+\ell}$.

"!" Пусть $n = p_1 \ldots p_s = q_1 \ldots q_t$ — наименьшее натуральное число, имеющее два различных разложения на простые множители.

- Тогда $p_1 \ldots p_s \vdots q_1$, следовательно, один из множителей p_i делится на q_1.
  - Поскольку p_i и q_1 — простые числа, имеем $p_i = q_1$.
- Пусть, не умаляя общности, $p_1 = q_1$. Тогда $p_2 \ldots p_s = q_2 \ldots q_t < n$.
- Получили меньшее, чем n, число, имеющее два разложения. Противоречие. $\square$

Каноническое разложение на простые множители

Замечание

• В разложении натурального числа на простые множители одно и то же простое число может встречаться несколько раз. Поэтому разложение числа на простые множители удобнее записывать в следующем виде:

$$n = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k},$$

где $p_1, p_2, \dots, p_k$ — различные простые числа и $a_1, a_2, \dots, a_k \in \mathbb{N}$.

• Такая запись называется *каноническим разложением* числа n на простые множители.

Пример

$360 = 2^3 \cdot 3^2 \cdot 5^1$.

Сравнения по модулю

Определение 6

- Пусть $a, b \in \mathbb{Z}$, $m \in \mathbb{N}$. Число a *сравнимо* с числом b по модулю m, если $a - b \vdots m$.
- Обозначение: $a \equiv b \pmod{m}$ или $a \equiv_m b$.

Теорема 9

Пусть $a, b \in \mathbb{Z}$, $m \in \mathbb{N}$. Тогда $a \equiv b \pmod{m}$, если и только если a и b дают одинаковые остатки при делении на m.

Теорема 10

Пусть $m \in \mathbb{N}$. Тогда отношение "a сравнимо с b по модулю m" является отношением эквивалентности на множестве целых чисел.

Из этого следует, что множество целых чисел можно разбить на классы эквивалентности по отношению сравнимости по модулю m.

Классы вычетов

Определение 7

- Пусть $m \in \mathbb{N}$. Классы эквивалентности по отношению $\equiv_m$ называются *классами вычетов* по модулю m.
- Класс вычетов, содержащий элемент $a \in \mathbb{Z}$ мы будем обозначать $\overline{a}$.

Замечание

1. В некоторых книгах вместо $\overline{a}$ пишут $[a]$.
   В случаях, когда нужно явно указывать модуль, пишут $\overline{a}_m$ или $[a]_m$.
2. Легко видеть, что существует ровно m классов вычетов по модулю m:
   это классы $\overline{0}$, $\overline{1}$, $\ldots$, $\overline{m-1}$. Каждый класс вычетов состоит из всех целых чисел, дающих некоторый фиксированный остаток при делении на m.

Арифметические свойства сравнений

Теорема 11

Пусть $a \equiv b \pmod{m}$ *и* $c \equiv d \pmod{m}$. *Тогда*

- $a \pm c \equiv b \pm d \pmod{m}$;
- $ac \equiv bd \pmod{m}$.

Следствие

Пусть $a \equiv b \pmod{m}$ *и* $n \in \mathbb{N}$. *Тогда* $a^n \equiv b^n \pmod{m}$.

Замечание

- Это означает, что мы можем корректно ввести операции сложения и умножения на множестве классов вычетов по модулю m.
- Тем самым, классы вычетов по модулю m образуют кольцо.

Полная и приведенная системы вычетов

Определение 8

Множество целых чисел, содержащее ровно по одному элементу из каждого класса вычетов по модулю m, называется *полной системой вычетов* по модулю m.

Утверждение

Если $a \equiv b \pmod{m}$, *то* $(a, m) = (b, m)$.

- Тем самым можно корректно говорить о том, что класс вычетов по модулю m *взаимно прост* с m.

Определение 9

Множество целых чисел, содержащее ровно по одному элементу из каждого класса вычетов по модулю m, взаимно простого с m, называется *приведенной системой вычетов* по модулю m.

Примеры

1. $\{0, 1, \ldots, m-1\}$ — полная система вычетов по модулю m.
2. $\{21, -13, 93, 3, -3, 12, -1\}$ — полная система вычетов по модулю 7.
3. $\{1, 2, \ldots, p-1\}$ — приведенная система вычетов по модулю p, где $p \in \mathbb{P}$.
4. $\{-1, 1\}$ — приведенная система вычетов по модулю 6.

Функция Эйлера

Определение 10

- Пусть $n \in \mathbb{N}$. Количество натуральных чисел, меньше либо равных n и взаимно простых с n обозначается через $\varphi(n)$.
- Функция $\varphi(n)$ называется *функцией Эйлера*.

Утверждение

Приведенная система вычетов по модулю m содержит ровно $\varphi(m)$ элементов.

Теорема 12

Пусть $n = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}$ — каноническое разложение n на простые множители. Тогда

$$\varphi(n) = n \left(1 - \frac{1}{p_1}\right) \left(1 - \frac{1}{p_2}\right) \dots \left(1 - \frac{1}{p_k}\right).$$

Свойства полной и приведенной систем вычетов

Теорема 13

Пусть $x_1, \ldots, x_m$ *— полная система вычетов по модулю* m*;* $a, b \in \mathbb{Z}$*,* $(a, m) = 1$*. Тогда* $ax_1 + b, ax_2 + b, \ldots, ax_m + b$ *— также полная система вычетов по модулю* m*.*

Теорема 14

Пусть $x_1, \ldots, x_{\varphi(m)}$ *— приведенная система вычетов по модулю* m*;* $a \in \mathbb{Z}$*,* $(a, m) = 1$*. Тогда* $ax_1, ax_2, \ldots, ax_{\varphi(m)}$ *— также приведенная система вычетов по модулю* m*.*

Малая теорема Ферма, теорема Эйлера, Китайская теорема об остатках

Теорема 15 (Малая теорема Ферма)

Пусть $a \in \mathbb{Z}$, $p \in \mathbb{P}$, $(a, p) = 1$. *Тогда* $a^{p-1} \equiv 1 \pmod{p}$.

Следствие (Другая формулировка малой теоремы Ферма)

Пусть $a \in \mathbb{Z}$, $p \in \mathbb{P}$. *Тогда* $a^p \equiv a \pmod{p}$.

Теорема 16 (Эйлер)

Пусть $a \in \mathbb{Z}$, $m \in \mathbb{N}$, $(a, m) = 1$. *Тогда* $a^{\varphi(m)} \equiv 1 \pmod{m}$.

Теорема 17 (Китайская теорема об остатках)

Пусть $m_1, m_2, \ldots, m_n \in \mathbb{N}$ — *попарно взаимно просты;* $M = m_1 m_2 \ldots m_n$.
Пусть также $b_1, b_2, \ldots, b_n \in \mathbb{Z}$. *Тогда все решения системы сравнений*

$$\begin{cases} x \equiv b_1 \pmod{m_1} \\ \ldots \\ x \equiv b_n \pmod{m_n} \end{cases}$$

образуют класс вычетов по модулю M.

Обратные вычеты (1/3)

Определение 11

- Пусть $a \in \mathbb{Z}$ и $m \in \mathbb{N}$. Число $b \in \mathbb{Z}$ называется *обратным вычетом* к a по модулю m, если $ab \equiv 1 \pmod{m}$.
- Обратный вычет к a по модулю m обозначается $a^{-1} \pmod{m}$.
  - То есть можно писать так: $b \equiv a^{-1} \pmod{m}$.

Теорема 18

1. *Обратный вычет к a по модулю m существует, если и только если $(a, m) = 1$.*
2. *Если обратный вычет к a по модулю m существует, то он единственен с точностью до сравнения по модулю m (т. е. если обратные вычеты существуют, то они образуют ровно один класс вычетов по модулю m).*

Доказательство. 1. "$\Rightarrow$": Пусть $b \equiv a^{-1} \pmod{m}$ и $d = (a, m)$.

- Тогда $ab \equiv 1 \pmod{m}$, следовательно, $ab - 1 = mc$, где $c \in \mathbb{Z}$.
  - Таким образом, $1 = ab - mc \vdots d$, откуда $d = 1$.

Обратные вычеты (2/3)

"$\Leftarrow$": Пусть $(a, m) = 1$.

- Рассмотрим линейное представление НОД a и m: $1 = as + mt$.
  - Тогда $as \equiv 1 \pmod{m}$, то есть $s \equiv a^{-1} \pmod{m}$.

2. Пусть b и c — два обратных вычета к a по модулю m.

- Тогда $ab \equiv 1 \equiv ac \pmod{m}$.
  - Следовательно, $a(b - c) = ab - ac \vdots m$.
  - Поскольку $(a, m) = 1$, получаем, что $b - c \vdots m$, то есть $b \equiv c \pmod{m}$. $\square$

Как найти обратный вычет?

Алгоритм поиска обратного вычета непосредственно вытекает из доказательства теоремы 18.

- Вход: $a \in \mathbb{Z}$; $m \in \mathbb{N}$.
- При помощи алгоритма Евклида ищем $d = (a, m)$;
  - если $d > 1$, то обратного вычета не существует;
  - если $d = 1$, то ищем линейное представление НОД a и m.
- Пусть $1 = as + mt$ — линейное представление НОД. Тогда $s \equiv a^{-1} \pmod{m}$.

Обратные вычеты (3/3)

Замечание

Часто в качестве ответа требуется дать число от 0 до $m-1$. В то же время, коэффициент s в линейном представлении может быть отрицательным.
В таком случае, на выход алгоритма следует передавать число $m+s$.

Пример

Найдем обратный вычет к числу 8 по модулю 19.

$$19 = 8 \cdot 2 + 3$$
$$8 = 3 \cdot 2 + 2$$
$$3 = 2 \cdot 1 + 1$$
$$2 = 1 \cdot 2 + 0$$

$$1 = 3 - 2 \cdot 1 = 3 - (8 - 3 \cdot 2) \cdot 1 = 3 \cdot 3 - 8 \cdot 1 = (19 - 8 \cdot 2) \cdot 3 - 8 \cdot 1 = \\ = 19 \cdot 3 - 8 \cdot 7.$$

Тогда $8^{-1} \equiv -7 \equiv 12 \pmod{19}$.

Возведение в степень по модулю (1/3)

- Даны числа $a, b, m \in \mathbb{N}$, где $a < m$. Нужно найти остаток от деления a^b на m.
- Это можно сделать методом повторного возведения в квадрат.
- На самом деле, есть два алгоритма, основанных на этом методе.
  - В обоих алгоритмах нам потребуется двоичное разложение числа b.
  - Пусть $b = \sum_{i=0}^{k} b_i 2^i$ (т. е. $b_0, b_1, \ldots, b_k$ — цифры двоичной записи числа b).

Алгоритм "старшими битами вперед"

- Пусть $s_j \stackrel{\text{def}}{=} \sum_{i=0}^{j} b_{k-j+i} 2^i$ — число, записанное первыми $j+1$ цифрами двоичной записи числа b.
  - Тогда $s_0 = 1$; $s_j = 2s_{j-1} + b_{k-j}$ при $j > 0$; $s_k = b$.
- Последовательно вычисляем $a^{s_0}, a^{s_1}, a^{s_2}, \ldots, a^{s_k} \pmod{m}$.
  - Пусть a_j — остаток от деления a^{s_j} на m.
  - Тогда $a_0 = a$; $\quad a_j \equiv_m \begin{cases} a_{j-1}^2, & \text{при } b_{k-j} = 0 \\ a_{j-1}^2 \cdot a, & \text{при } b_{k-j} = 1. \end{cases}$
  - При каждом умножении нужно делить его результат с остатком на m. Тогда все промежуточные числа будут меньше m^2.

Возведение в степень по модулю (2/3)

Пример

Возведем 13 в степень 23 по модулю 29. $23 = 10111_2$.

$$
\begin{aligned}
13^1 &= 13; \\
13^2 &\equiv 24 \pmod{29}; \\
13^5 &\equiv 24^2 \cdot 13 \equiv 25 \cdot 13 \equiv 6 \pmod{29}; \\
13^{11} &\equiv 6^2 \cdot 13 \equiv 7 \cdot 13 \equiv 4 \pmod{29}; \\
13^{23} &\equiv 4^2 \cdot 13 \equiv 16 \cdot 13 \equiv 5 \pmod{29}.
\end{aligned}
$$

Алгоритм “младшими битами вперед”

- Последовательно возводя в квадрат, вычисляем остатки от деления на m чисел a^{2^j}.
- Перемножаем те остатки, при которых $b_j = 1$.
  - Лучше это делать последовательно. То есть всякий раз, когда очередное $b_j = 1$, домножаем предыдущее произведение на a^{2^j}.

Возведение в степень по модулю (3/3)

Пример

$13^1 = 13;$ $\quad 13^1 = 13;$

$13^2 \equiv 24 \pmod{29};$ $\quad 13^3 \equiv 13^2 \cdot 13 \equiv 24 \cdot 13 \equiv 22 \pmod{29};$

$13^4 \equiv 24^2 \equiv 25 \pmod{29};$ $\quad 13^7 \equiv 13^4 \cdot 13^3 \equiv 25 \cdot 22 \equiv 28 \pmod{29};$

$13^8 \equiv 25^2 \equiv 16 \pmod{29};$

$13^{16} \equiv 16^2 \equiv 24 \pmod{29};$ $\quad 13^{23} \equiv 13^{16} \cdot 13^7 \equiv 24 \cdot 28 \equiv 5 \pmod{29}.$

Теорема 19

Каждый из приведенных выше алгоритмов имеет время работы $O(\log^2 m \cdot \log b)$.

Доказательство. Алгоритм производит не более $2k = O(\log b)$ операций умножения чисел, меньших m.

- Каждая операция умножения требует $O(\log^2 m)$ битовых операций. □

Замечание

- В случае, если b велико, $(a, m) = 1$ и нам известно число $\varphi(m)$, имеет смысл сначала поделить с остатком b на $\varphi(m)$, после чего возводить a в степень r (где r — остаток от деления b на $\varphi(m)$). Тогда время работы составит $O(\log^3 m + \log b \cdot \log m)$.
- Но мы далеко не всегда можем знать число $\varphi(m)$.

Криптосистема RSA (Rivest–Shamir–Adleman, 1977)

- Пусть p, q — большие простые числа;
- $n = pq$, $\varphi(n) = (p-1)(q-1)$;
- $e \in \mathbb{N}$ такое, что $e < \varphi(n)$ и $(e, \varphi(n)) = 1$;
- $d \in \mathbb{N}$ — обратный вычет к e по модулю $\varphi(n)$ ($d < \varphi(n)$ и $ed \equiv 1 \pmod{\varphi(n)}$).
  - Чаще всего числа e и d стараются выбирать так, чтобы число d было большим, а числа e — достаточно небольшим (но и не слишком маленьким).

- Пара (n, e) — *открытый ключ*. Он используется для шифрования сообщений и публикуется в открытом доступе.
- Пара (n, d) — *секретный ключ*. Он используется для дешифрования сообщений и должен храниться в секрете.

- Сообщение — число от 0 до $N-1$ (более длинные сообщения разбиваются на блоки, которые шифруются по отдельности).
- Шифрование — функция $P : [0..n-1] \to [0..n-1]$, где $P(m) \equiv m^e \pmod{n}$.
- Дешифрование — функция $S : [0..n-1] \to [0..n-1]$, где $S(m) \equiv m^d \pmod{n}$.

Криптосистема RSA. Доказательство корректности

Теорема 20

$S(P(m)) = P(S(m)) = m.$

Доказательство. Нужно доказать, что $m^{ed} \equiv m \pmod{n}$.

- Для этого достаточно доказать, что $m^{ed} \equiv m \pmod{p}$ и $m^{ed} \equiv m \pmod{q}$.
- Заметим, что $ed \equiv 1 \pmod{\varphi(n)}$. То есть, $ed = (p-1)(q-1)k + 1$, где $k \in \mathbb{N}$.
  - Пусть $m \not\vdots p$. Тогда $m^{p-1} \equiv 1 \pmod{p}$. Следовательно, $m^{ed} = m^{(p-1)(q-1)k+1} = (m^{p-1})^{(q-1)k} \cdot m \equiv 1^{(q-1)k} \cdot m \equiv m \pmod{n}$.
  - Пусть $m \vdots p$. Тогда $m^{ed} \equiv 0 \equiv m \pmod{n}$.
- Итак, во всех случаях получаем, что $m^{ed} \equiv m \pmod{p}$.
- То, что $m^{ed} \equiv m \pmod{q}$, доказывается аналогично. □

Замечание

- В 1977 году авторы алгоритма (R. L. Rivest, A. Shamir, L. M. Adleman) опубликовали тестовый пример, в котором число n состояло из 129 десятичных (425 двоичных) знаков.
- Тестовый пример был расшифрован в 1994 году при помощи распределенных вычислений: для этого потребовалось полгода работы сети из 1600 компьютеров.
- В настоящее время надежными считаются системы, в которых n содержит порядка 2000 двоичных знаков.

Криптосистема RSA. О выборе p и q

Выбирая простые числа p и q стоит придерживаться некоторых ограничений.

- Числа p и q не должны быть близки друг к другу. Обычно их выбирают так, чтобы длина их записи отличалась на несколько разрядов.
  - Действительно, $pq = \left(\frac{p+q}{2}\right)^2 - \left(\frac{p-q}{2}\right)^2$.
  - Если $|p-q|$ мал, то $\left(\frac{p+q}{2}\right)^2$ — точный квадрат, ненамного превосходящий n.
  - Тогда перебирая точные квадраты, большие n, мы быстро найдем такое a, что $a^2 - n$ — точный квадрат.
  - Далее, положив $a = \frac{p+q}{2}$ и $\sqrt{a^2 - n} = \frac{p-q}{2}$, мы легко найдем p и q.
- $(p-1, q-1)$ должен быть маленьким.
- Каждое из чисел $p-1, q-1$ должно иметь большой простой делитель.
  - Лучше всего рассматривать так называемые безопасные простые числа: число $p \in \mathbb{P}$ называют *безопасным*, если $\frac{p-1}{2} \in \mathbb{P}$.
  - Но такие числа значительно труднее искать.

Цифровая подпись на основе RSA

Рассмотрим следующую схему передачи данных: $A \xrightarrow{C} B$.

- A (Алиса): отправитель сообщения;
- B (Боб): получатель сообщения;
- C (Чарли): шпион, перехватывающий сообщения. В некоторых книгах шпиона называют E (Ева) — от слова eavesdropper (подслушивающий).
- Пусть P_A и S_A — открытый и секретный ключи Алисы; P_B и S_B — Боба.
- Пусть m — сообщение Алисы. Тогда
  - $S_A(m)$ — цифровая подпись для сообщения m;
  - $(m, S_A(m))$ — сообщение с цифровой подписью;
  - $(P_B(m), P_B(S_A(m)))$ — шифрованное сообщение с цифровой подписью.
- Для проверки подлинности сообщения m, B должен применить P_A к цифровой подписи и сравнить результат с m: если они совпадают, то сообщение подлинное.
- Здесь возможна проблема подделки открытого ключа: если C сможет передать B свой открытый ключ под видом открытого ключа A, то он сможет отправлять B подписанные этим ключом сообщения и B не сможет распознать подмену.
  - На самом деле, проверка цифровой подписи гарантирует лишь то, что отправитель имеет доступ к секретному ключу, соответствующему данному открытому ключу.

Как искать простые числа? (1/2)

Исторически, первым алгоритмом поиска простых чисел считается решето Эратосфена.

Решето Эратосфена

- Выписываем в ряд все натуральные числа от 2 до n;
- далее, пока есть непомеченные числа, повторяем следующие действия
  - выбираем наименьшее непомеченное число p, помечаем его как простое;
  - движемся с шагом p, помечаем числа $2p, 3p, \ldots$ как составные.

Замечание

- Легко видеть, что сложность этого алгоритма составляет $O(n^2)$ (внешний цикл выполняется не более n раз; rкаждая его итерация требует $O(n)$ операций).
- Посчитав аккуратнее, можно получить оценку $O(n \log \log n)$.
- Тем не менее, этот алгоритм экспоненциален относительно длины входа (здесь длина входа — это длина двоичной записи числа n, которая равна $\lceil \log n \rceil$). Для поиска простых чисел интересующего нас размера этот алгоритм неприменим.

Как искать простые числа? (2/2)

Чаще всего, на практике используют следующий алгоритм.

- Случайным образом выбираем число $a \in \mathbb{N}$ из нужного нам диапазона.
- Если $a \vdots 2$, то заменяем a на $a + 1$.
- Далее, последовательно перебираем числа $a, a + 2, a + 4, \ldots$ и проверяем каждое из них на простоту.

Определение 12

Количество простых чисел, лежащих в промежутке от 1 до n обозначается $\pi(n)$.

Теорема 21 (асимптотический закон распределения простых чисел)

$\pi(n) \sim \frac{n}{\ln n}$. *(б/д)*

Замечание

- Асимптотический закон сформулировал в 1848 году П. Л. Чебышёв. Доказан он был в 1896 году независимо Адамаром и де-ла-Валле-Пуссеном.
- Тем самым, плотность распределения простых чисел вблизи числа a составляет примерно $\frac{1}{\ln a}$. То есть при случайном выборе a можно рассчитывать на то, что для нахождения простого числа нужно будет перебрать $O(\log a)$ чисел.

Проверка простоты числа

- Тривиальный алгоритм: перебираем все числа от 2 до $\sqrt{n}$ и проверяем, делится ли n на каждое из них.
  - Этот алгоритм экспоненциален относительно длины входа ($\log n$) и для чисел интересующего нас размера неприменим.
- Полиномиальный алгоритм. На данный момент известен единственный алгоритм проверки простоты с доказанным полиномиаьным временем работы.
  - AKS тест (M. Agrawal, N. Kayal, N. Saxena, 2004).
  - Время работы последней модификации этого теста оценивается как $O(\log^6 n)$.
  - Это время, хотя и полиномиально, но весьма велико. По этой причине, а также из-за большого расхода памяти, AKS тест почти не применяют на практике.
- Квазиполиномиальный алгоритм. L. M. Adleman, C. Pomerance, R. S. Rumely, 1983.
  - Время работы оценивается как $O((\log n)^{c \log\log\log n})$.
  - В тех случаях, когда необходима точная проверка простоты числа, как правило применяют этот алгоритм или его модификации.
  - Тем не менее, чаще всего на практике используют вероятностные тесты, которые работают гораздо быстрее.

Вероятностные алгоритмы

- *Вероятностный алгоритм* — это алгоритм, ход и результаты работы которого, помимо входа, зависят от выбора некоторого случайного параметра a.
  - Как правило, a — это натуральное число, которое случайным образом выбирается из некоторого диапазона.
  - При определенных значениях a алгоритм может давать ошибочный результат, но вероятность этого должна быть не слишком велика (т. е. не превосходить некоторой заранее фиксированной константы).
    - Например, бывают вероятностные алгоритмы, для которых вероятность ошибки меньше $\frac{1}{2}$.
  - Для снижения вероятности ошибки можно многократно запустить вероятностный алгоритм. При каждом запуске алгоритма, параметр a выбирается заново, случайным и независимым от предыдущих запусков образом.
- Для проверки простоты числа нас будут интересовать *вероятностные алгоритмы с односторонней ошибкой*. Если такой алгоритм отвечает, что число составное, то это гарантировано так. А вот ответ простое может быть и ошибочным.

Тест Ферма

- Вход: нечётное натуральное число n.
- Выбираем случайным образом параметр $a \in [2..n-2]$;
- вычисляем a^{n-1} по модулю n;
  - если $a^{n-1} \equiv 1 \pmod{n}$, то ответ "простое",
  - если $a^{n-1} \not\equiv 1 \pmod{n}$, то ответ "составное".

Замечание

- Из малой теоремы Ферма следует, что ответ "составное" не может быть ошибочным.
- В то же время, ответ "простое" ошибочным быть может.
- Отметим, что сравнение $a^{n-1} \equiv 1 \pmod{n}$ может быть выполнено только в случае $(a, n) = 1$.

Определение 13

Нечетное составное число n называется *псевдопростым по основанию a*, если $a^{n-1} \equiv 1 \pmod{n}$.

Числа Кармайкла

- К сожалению, существуют такие нечетные составные числа n, для которых $a^{n-1} \equiv 1 \pmod{n}$ при всех a взаимно простых с n. Такие числа проходят тест Ферма почти при любом выборе a.

Пример

- Пусть $n = 561 = 3 \cdot 11 \cdot 17$ и $(a, n) = 1$. Тогда
  - $a \not\vdots 3 \;\Rightarrow\; a^2 \equiv 1 \pmod{3} \Rightarrow a^{560} \equiv (a^2)^{280} \equiv 1^{280} \equiv 1 \pmod{3}$;
  - $a \not\vdots 11 \Rightarrow a^{10} \equiv 1 \pmod{11} \Rightarrow a^{560} \equiv (a^{10})^{56} \equiv 1^{56} \equiv 1 \pmod{11}$;
  - $a \not\vdots 17 \Rightarrow a^{16} \equiv 1 \pmod{17} \Rightarrow a^{560} \equiv (a^{16})^{35} \equiv 1^{35} \equiv 1 \pmod{17}$.
- Следовательно, $a^{560} \equiv 1 \pmod{561}$.

Определение 14

Составное натуральное число n называется *числом Кармайкла*, если $a^{n-1} \equiv 1 \pmod{n}$ для любого a, такого, что $(a, n) = 1$.

Теорема 22 (W. R. Alford, A. Granville, C. Pomerance, 1994)

Чисел Кармайкла бесконечно много. (б/д)

Числа Кармайкла и тест Ферма (1/2)

- Хотя чисел Кармайкла бесконечно много, но они встречаются достаточно редко.
  - Например, среди чисел $n \leq 25 \cdot 10^9$ есть 2163 чисел Кармайкла.
  - При этом, простых чисел в данном промежутке около 10^9.
- В случае, когда n не число Кармайкла, тест Ферма работает достаточно хорошо.

Теорема 23

Пусть $n, b \in \mathbb{N}$ таковы, что $(n, b) = 1$ и $b^{n-1} \not\equiv 1 \pmod{n}$. Тогда в множестве $[1..n-1]$ есть не более $\frac{n-1}{2}$ чисел, удовлетворяющих условию

$$a^{n-1} \equiv 1 \pmod{n}. \tag{1}$$

Доказательство. Пусть $a_1, a_2, \ldots, a_k \in [1..n-1]$ — все числа, удовлетворяющие (1).

- Рассмотрим числа $c_1, c_2, \ldots, c_k \in [1..n-1]$, где $c_i \equiv a_i b \pmod{n}$.
  - Заметим, что $c_i^{n-1} \equiv (a_i b)^{n-1} = a_i^{n-1} b^{n-1} \equiv b^{n-1} \not\equiv 1 \pmod{n}$.
    То есть числа c_i не удовлетворяют условию (1).
  - Далее, все c_i различны. Действительно, если $c_i = c_j$, то $(a_i - a_j)b \vdots n$.
    Тогда, поскольку $(n, b) = 1$, получаем, что $a_i \equiv_n a_j$, откуда $i = j$.
- Таким образом, в промежутке $[1..n-1]$ есть k чисел, удовлетворяющих условию (1), и не менее k чисел, не удовлетворяющих (1). Следовательно, $k \leq \frac{n-1}{2}$. □

Числа Кармайкла и тест Ферма (2/2)

Следствие

Если составное число n не является числом Кармайкла, то при случайном выборе параметра a тест Ферма даст ответ "простое" с вероятностью меньше $\frac{1}{2}$.

Замечание

- Тем самым, многократно повторив тест Ферма для числа n мы с большой вероятностью узнаем, что оно составное, если, конечно, n не число Кармайкла.
- Поскольку числа Кармайкла редки, тест Ферма часто применяют на практике.
  - Более того, порой люди ограничиваются проверкой теста Ферма по основанию 2.
  - Известно, что среди $n \leq 25 \cdot 10^9$ есть 21853 составных чисел, которые проходят тест Ферма по основанию 2. А простых чисел в этом промежутке порядка 10^9, так что шанс ошибиться невелик.
- Тем не менее, наличие чисел Кармайкла заставляет искать более надежные тесты простоты.

Тест Миллера-Рабина (1/3)

- Вход: нечётное натуральное число n.
- Пусть $n - 1 = 2^t \cdot u$, где $t, u \in \mathbb{N}$ и $u \not\vdots 2$.
- Выбираем случайным образом параметр $a \in [2..n-2]$.
- Вычисляем $a^u, a^{2u}, \ldots, a^{2^t u}$ по модулю n (получившаяся последовательность называется *последовательность Миллера-Рабина*).
  - Ответ "простое" дается в следующих двух случаях:
    - если $a^u \equiv 1 \pmod{n}$,
    - либо если $a^{2^k u} \equiv -1 \pmod{n}$ при некотором $k \in [0..t-1]$.
  - Во всех остальных случаях, дается ответ "составное".

Замечание

Как и в случае теста Ферма, ответ "простое" при выполнении теста Миллера-Рабина может быть ошибочным.

Определение 15

Нечетное составное число n называется *сильно псевдопростым по основанию a*, если тест Миллера-Рабина для числа n с параметром a дает ответ "простое".

Тест Миллера-Рабина (2/3)

Утверждение

Если $n \in \mathbb{P}$, то тест Миллера-Рабина выдаст ответ "простое".

Доказательство. По малой теореме Ферма $a^{2^t \cdot u} = a^{n-1} \equiv 1 \pmod{n}$,
так что в последовательности Миллера-Рабина есть хотя бы одна единица.

- Рассмотрим такое наименьшее k, что $a^{2^k \cdot u} \equiv 1 \pmod{n}$.
  - Если $k = 0$, то $a^u \equiv 1 \pmod{n}$ и тогда ответ "простое".
  - Пусть $k > 0$. Тогда $a^{2^k \cdot u} \equiv 1 \pmod{n}$ и $a^{2^{k-1} \cdot u} \not\equiv 1 \pmod{n}$.
    - Следовательно, $(a^{2^{k-1} \cdot u} - 1)(a^{2^{k-1} \cdot u} + 1) = a^{2^k \cdot u} - 1 \vdots n$.
    - Поскольку $n \in \mathbb{P}$ и $a^{2^{k-1} \cdot u} \not\equiv 1 \pmod{n}$, получаем, что $a^{2^{k-1} \cdot u} + 1 \vdots n$.
      И тогда тоже ответ "простое". □

- Итак, тест Миллера-Рабина — вероятностный тест с односторонней ошибкой.

Теорема 24

Пусть n — нечетной составное число. Тогда в множестве $[1..n-1]$ есть не более $\frac{n-1}{4}$ чисел a, таких, что n является сильно псевдопростым по основанию a. (б/д)

Тест Миллера-Рабина (3/3)

Следствие

Пусть n — нечетной составное число. Тогда при случайном выборе параметра a тест Миллера-Рабина даст ответ "простое" с вероятностью меньше $\frac{1}{4}$.

Замечание

• В отличии от теста Ферма, для теста Миллера-Рабина нет "плохих" чисел, результаты проверки которых почти всегда ошибочные. При любом составном n вероятность ошибки будет меньше $\frac{1}{4}$.

• Время работы теста Миллера-Рабина мало отличается от времени работы теста Ферма. Действительно, если вычислять a^{n-1} по модулю n алгоритмом "старшие биты вперед", то остатки всех членов последовательности Миллера-Рабина будут получены в качестве промежуточных результатов.

• При этом как правило сильно псевдопростые числа по данному основанию встречаются значительно реже, чем псевдопростые. Например, среди $n \leq 25 \cdot 10^9$ есть 4842 числа, сильно псевдопростых по основанию 2, и 21853 числа, псевдопростых по основанию 2. Так что предпочтительно использовать тест Миллера-Рабина, а не тест Ферма.

• Из обобщенной гипотезы Римана следует, что нечетное составное число проваливает тест Миллера-Рабина хотя бы для одного из оснований $b < 2\log^2 n$. Если это доказать, то будет получен еще один полиномиальный алгоритм проверки простоты.

Глава 2.
Алгебраические структуры

- Дополнительные материалы:
  - Б. Л. ван дер Варден, *Алгебра*. М.: Мир, 1976.
  - С. Ленг, *Алгебра*. М.: Мир, 1968.

2.1. Кольца и поля

Определение 16

Полем называется упорядоченная тройка $(K, +, \cdot)$, состоящая из множества K и бинарных операций $+$ и $\cdot$ на K, называемых соответственно *сложением* и *умножением*, для которых выполняются следующие свойства:

1. $\forall a, b \in K \ (a + b = b + a)$ (*коммутативность сложения*);
2. $\forall a, b, c \in K \ (a + (b + c) = (a + b) + c)$ (*ассоциативность сложения*);
3. $\exists 0 \in K \ \forall a \in K \ (a + 0 = a)$ (*существование нулевого элемента*);
4. $\forall a \in K \ \exists (-a) \in K \ (a + (-a) = 0)$ (*существование обратного элемента по сложению*);
5. $\forall a, b \in K \ (a \cdot b = b \cdot a)$ (*коммутативность умножения*);
6. $\forall a, b, c \in K \ (a \cdot (b \cdot c) = (a \cdot b) \cdot c)$ (*ассоциативность умножения*);
7. $\exists 1 \in K \setminus \{0\} \ \forall a \in K \ (a \cdot 1 = 1 \cdot a = a)$ (*существование единичного элемента*);
8. $\forall a \in K \setminus \{0\} \ \exists a^{-1} \in K \ (a \cdot a^{-1} = a^{-1} \cdot a = 1)$ (*существование обратного элемента по умножению*);
9. $\forall a, b, c \in K \ (a \cdot (b + c) = a \cdot b + a \cdot c))$;
   $\forall a, b, c \in K \ ((b + c) \cdot a = b \cdot a + c \cdot a))$ (*дистрибутивность*).

Кольца и поля: типы колец

Определение 17

- Упорядоченная тройка $(K, +, \cdot)$ называется *кольцом*, если удовлетворяет свойствам 1-4, 6, 9 из предыдущего определения.
- Кольцо называется *коммутативным*, если оно удовлетворят также свойству 5.
- Кольцо называется *кольцом с единицей*, если удовлетворят также свойству 7.
- Кольцо называется *телом*, если выполняются свойства 1-4, 6-9.
  - То есть единственное из свойств поля, которое может не выполняться для тела — это коммутативность умножения.

Замечание

1. Иногда рассматривают также *неассоциативные кольца*, в которых выполняются только свойства 1-4, 9. Простейшим примером такого кольца является множество векторов в $\mathbb{R}^3$, со стандартной операцией сложения и векторным умножением в качестве умножения.
2. Символ "$\cdot$" часто опускают и вместо $a \cdot b$ пишут просто ab.

Кольца и поля: примеры (1/3)

1. Множество $\mathbb{Z}$ со стандартными операциями сложения и умножения является коммутативным кольцом с единицей.
2. Множества $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ являются полями.
3. Пусть $m \in \mathbb{N}$, $m > 1$. Мы доказывали, что на множестве классов вычетов по модулю m можно корректным образом ввести операции сложения и умножения. Тогда получится коммутативное кольцо с единицей.
   - Оно обозначается $\mathbb{Z}/m\mathbb{Z}$.

   Доказательство. Коммутативность, ассоциативность и дистрибутивность для сложения и умножения классов вычетов непосредственно следуют из аналогичных свойств в кольце $\mathbb{Z}$.
   - Нейтральным элементом по сложению будет класс $\overline{0}$;
   - обратным элементом по сложению для $\overline{a}$ будет класс $\overline{(-a)}$;
   - нейтральным элементом по умножению будет класс $\overline{1}$. □

Замечание

В ряде книг вместо обозначения $\mathbb{Z}/m\mathbb{Z}$ используется обозначение $\mathbb{Z}_m$. Но порой это приводит к конфликту обозначений: через $\mathbb{Z}_p$ обозначают кольцо целых p-адических чисел. Поэтому мы будем использовать обозначение $\mathbb{Z}/m\mathbb{Z}$.

Кольца и поля: примеры (2/3)

4. Пусть A — произвольное кольцо. Рассмотрим множество $A[x]$ многочленов с коэффициентами из кольца A от переменной x. Это множество является кольцом.

   • Легко видеть, что кольцо $A[x]$ коммутативно тогда и только тогда, когда коммутативно кольцо A, и что $A[x]$ — кольцо с единицей тогда и только тогда, когда A — кольцо с единицей.

5. Примером некоммутативного кольца является кольцо матриц $n \times n$, где $n > 1$.

   • Пусть A — коммутативное кольцо. Обозначим через $M_n(A)$ множество всех матриц $n \times n$ с коэффициентами из A. На этом множестве определены стандартные операции сложения и умножения матриц. Из курса линейной алгебры известно, что сложение и умножение матриц обладает всеми свойствами из определения кольца. Но при этом умножение матриц некоммутативно, поэтому кольцо $M_n(A)$ коммутативным не является.

6. Простейшим примером тела является *тело кватернионов*, то есть множество формальных сумм вида $a + bi + cj + dk$, где $a, b, c, d \in \mathbb{R}$, а *базисные элементы* i, j, k удовлетворяют соотношениям $i^2 = j^2 = k^2 = ijk = -1$.

   • Из этих соотношений следует, что например $ij = -ji$, так что тело кватернионов является некоммутативным, и, следовательно, не является полем.

Кольца и поля: примеры (3/3)

Теорема 25

Кольцо $\mathbb{Z}/p\mathbb{Z}$ является полем тогда и только тогда, когда $p \in \mathbb{P}$.

Доказательство. "$\Leftarrow$": Пусть $\overline{a} \in \mathbb{Z}/p\mathbb{Z}$ и $\overline{a} \neq \overline{0}$.

- Тогда $(a, p) = 1$, то есть существует вычет b, обратный к a по модулю p.
- Следовательно, $\overline{a} \cdot \overline{b} = \overline{1}$, то есть для $\overline{a}$ есть обратный по умножению элемент.

"$\Rightarrow$": Пусть $\mathbb{Z}/p\mathbb{Z}$ — поле.

- Тогда для любого его ненулевого элемента $\overline{a}$ есть обратный по умножению.
- Следовательно, к a есть обратный вычет по модулю p. Но тогда $(a, p) = 1$.
- Таким образом, любое некратное p число взаимно просто с p, откуда $p \in \mathbb{P}$. □

Замечание

- Поле $\mathbb{Z}/p\mathbb{Z}$ (другое обозначение этого поля — $\mathbb{F}_p$) является одним из примеров *конечных полей* (то есть полей, содержащих конечное число элементов).

Подробнее о конечных полях и их свойствах мы поговорим чуть позже.

- Конечные поля и многочлены над ними используются во многих алгоритмах.

В частности, на их основе можно строить помехоустойчивые коды.

Кольца и поля: простейшие свойства колец (1/2)

Теорема 26 (см. раздел 2.3.1 курса "Дискретная математика")

В кольце для любых элементов a и b выполняются следующие соотношения.

1. $0 \cdot a = a \cdot 0 = 0$;
2. $a \cdot (-b) = (-a) \cdot b = -(a \cdot b)$;
3. $(-a) \cdot (-b) = a \cdot b$;
4. $(-a) = a \cdot (-1)$;
5. $-(a + b) = (-a) + (-b)$;
6. *если* $a \neq 0$, *то* $(a^{-1})^{-1} = a$.

Теорема 27

В любом кольце K

1. *нулевой элемент единственен;*
2. *для любого $a \in K$ обратный по сложению единственен;*
3. *если K кольцо с единицей, то единичный элемент единственен;*
4. *если для $a \in K$ существует обратный по умножению, то он единственен.*

Кольца и поля: простейшие свойства колец (2/2)

Доказательство.

1. Пусть 0_1 и 0_2 — два нулевых элемента.
   - Тогда $0_1 = 0_1 + 0_2 = 0_2 + 0_1 = 0_2$.
2. Пусть b_1 и b_2 — обратные по сложению к a.
   - Тогда $b_1 = b_1 + 0 = b_1 + (a + b_2) = (b_1 + a) + b_2 = 0 + b_2 = b_2$.
3. Пусть 1_1 и 1_2 — два единичных элемента.
   - Тогда $1_1 = 1_1 \cdot 1_2 = 1_2$.
4. Пусть c_1 и c_2 — обратные по умножению к a.
   - Тогда $c_1 = c_1 \cdot 1 = c_1 \cdot (a \cdot c_2) = (c_1 \cdot a) \cdot c_2 = 1 \cdot c_2 = c_2$. □

Область целостности (1/2)

Определение 18

Пусть A — кольцо и элементы $x, y \in A \setminus \{0\}$ таковы, что $xy = 0$. Тогда элемент x называется *левым делителем нуля*, а элемент y — *правым делителем нуля*.

Пример

- В кольце $\mathbb{Z}/6\mathbb{Z}$ выполнено равенство $\overline{2} \cdot \overline{3} = \overline{0}$, следовательно, элементы $\overline{2}$ и $\overline{3}$ являются делителями нуля.
- Аналогично, для любого составного числа m в кольце $\mathbb{Z}/m\mathbb{Z}$ есть делители нуля: если $m = kl$, где $k, l < m$, то $\overline{k} \cdot \overline{l} = \overline{0}$ в $\mathbb{Z}/m\mathbb{Z}$.

Замечание

- Понятие делителя нуля важно с той точки зрения, что на ненулевые множители, не являющиеся делителями нуля, можно сокращать.
  - Пусть элементы $a, b, c \in A$ таковы, что $ab = ac$, причем $a \neq 0$ и a не делитель нуля. Тогда $a(b - c) = ab - ac = 0$, следовательно, $b - c = 0$, откуда $b = c$.
- На множитель, являющийся делителем нуля, сокращать нельзя.
  - Например, в $\mathbb{Z}/6\mathbb{Z}$ имеем $\overline{2} \cdot \overline{1} = \overline{2} \cdot \overline{4}$, но $\overline{1} \neq \overline{4}$.

Область целостности (2/2)

Определение 19

Коммутативное кольцо с единицей, не содержащее делителей нуля, называется *областью целостности* (или *целостным кольцом*).

Замечание

Условие о том, что область целостности должна быть кольцом с единицей, часто опускают.

Теорема 28

Поле является областью целостности.

Доказательство. Пусть K — поле и $a, b \in K$ таковы, что $ab = 0$.

- Предположим, что $b \neq 0$.
- Тогда $a = a \cdot 1 = a \cdot (b \cdot b^{-1}) = (a \cdot b) \cdot b^{-1} = 0 \cdot b^{-1} = 0$. □

Идеалы в коммутативном кольце (1/3)

Определение 20

Пусть R — коммутативное кольцо. Подмножество $I \subset R$ называется *идеалом*, если выполняются следующие условия:

1. $0 \in I$;
2. $\forall a, b \in I \ (a \pm b \in I)$;
3. $\forall a \in I \ \forall r \in R \ (ar \in I)$.

Замечание

1. Идеалы можно рассматривать также и в некоммутативных кольцах. Однако там ситуация несколько сложнее. Если кольцо R некоммутативно, то следует различать *правые идеалы*, определение которых в точности совпадает с приведенным выше, и *левые идеалы*, в которых третье условие заменяется на $\forall a \in I \ \forall r \in R \ (ra \in I)$. Идеал, одновременно являющийся и левым и правым, называется *двусторонним*. Тем не менее, изучение некоммутативных колец выходит за рамки нашего курса, поэтому далее мы будем рассматривать идеалы только в коммутативных кольцах.
2. Любой идеал в кольце R является его *подкольцом*, то есть сам является кольцом с теми же операциями, что и в кольце R. Однако, не всякое подкольцо является идеалом: например, $\mathbb{Z}$ — подкольцо в $\mathbb{Q}$, но не идеал в $\mathbb{Q}$.

Идеалы в коммутативном кольце (2/3)

Примеры

1. $\{0\}$ и R — идеалы в R.
2. Пусть $a \in R$. Тогда множество $(a) \stackrel{\text{def}}{=} \{ar \mid r \in R\}$ — идеал в R.
   Доказательство. $0 = a \cdot 0 \in (a)$;
   - если $ar_1, ar_2 \in (a)$, то $ar_1 \pm ar_2 = a(r_1 \pm r_2) \in (a)$;
   - если $ax \in (a)$ и $r \in R$, то $(ax)r = a(xr) \in (a)$. □

Определение 21

- Определенный выше идеал (a) называется *главным идеалом*, порожденным элементом a.
- Другое обозначение для главного идеала: aR.

Замечание

- Нулевой идеал является главным: $(0) = \{0\}$.
- Более того, если R — кольцо с единицей, то $R = (1)$ — также главный идеал.

Идеалы в коммутативном кольце (3/3)

Примеры

3. Пусть $a_1, \ldots, a_n \in R$. Тогда множество $(a_1, \ldots, a_n) \stackrel{\text{def}}{=} \{a_1 r_1 + \ldots + a_n r_n \mid r_1, \ldots, r_n \in R\}$ является идеалом в R.

   Доказательство.

   - $0 = a_1 \cdot 0 + \ldots + a_n \cdot 0 \in (a_1, \ldots, a_n)$;
   - если $b = a_1 r_1 + \ldots + a_n r_n \in (a_1, \ldots, a_n)$ и $c = a_1 s_1 + \ldots + a_n s_n \in (a_1, \ldots, a_n)$, то $b \pm c = a_1(r_1 \pm t_1) + \ldots + a_n(r_n \pm t_n) \in (a_1, \ldots, a_n)$;
   - если $b = a_1 x_1 + \ldots + a_n x_n \in (a_1, \ldots, a_n)$ и $r \in R$, то $br = a_1(x_1 r) + \ldots + a_n(x_n r) \in (a_1, \ldots, a_n)$. □

Определение 22

$(a_1, \ldots, a_n)$ — идеал, *порожденный* элементами $a_1, \ldots, a_n$.

Идеалы в поле и в кольце $\mathbb{Z}$

Теорема 29

Идеалы в поле — только (0) *и* (1).

Доказательство. Пусть K — поле; $I \subset K$ — идеал и $I \neq (0)$.

- Рассмотрим $a \in I \setminus \{0\}$. Тогда $1 = aa^{-1} \in I$.
- Следовательно, для любого $b \in K$ имеем $b = 1 \cdot b \in I$.
- Таким образом, $I = K = (1)$. □

Теорема 30

В кольце $\mathbb{Z}$ *все идеалы — главные.*

Доказательство. Пусть $I \subset \mathbb{Z}$ — идеал и $I \neq (0)$. Тогда $I \cap \mathbb{N} \neq \varnothing$.

- Рассмотрим $m = \min(I \cap \mathbb{N})$. Тогда очевидно, что $(m) \subset I$.
  - Пусть $a \in I$. Разделим a с остатком на m: $a = mq + r$, $0 \leq r < m$.
  - Тогда $r = a - mq \in I$. Поскольку $0 \leq r < m$, из этого следует, что $r = 0$.
  - Следовательно, $a = mq \in (m)$
- Таким образом, $I \subset (m)$, откуда $I = (m)$. □

Кольцо главных идеалов

Замечание

- Пусть a, b — ненулевые целые числа. Рассмотрим порожденный ими идеал (a, b).
- По теореме 30 этот идеал является главным, то есть $(a, b) = (c)$, где $c \in \mathbb{Z}$.
- Легко видеть, что тогда c — наибольший общий делитель a и b. Из этого факта, в частности, следует существование линейного представления наибольшего общего делителя двух целых чисел. Однако, в отличии от алгоритма Евклида, такой способ доказательства не дает алгоритма нахождения линейного представления.

Определение 23

Целостное кольцо с единицей, в котором все идеалы главные, называется *кольцом главных идеалов*.

Пример

- Кольцо $\mathbb{Z}[x]$ не является кольцом главных идеалов: идеал $(2, x)$ не главный.
- Действительно, если $(2, x) = (a)$, то $2 = ab$, где $a, b \in \mathbb{Z}[x]$.
- Тогда $a \in \{\pm 1, \pm 2\}$, но ни один из этих вариантов, очевидно, не подходит.

Факторкольцо (1/4)

Определение 24

- Пусть R — коммутативное кольцо с единицей, $x, y \in R$, I — идеал в R.
- Элемент x *сравним* с элементом y по модулю идеала I, если $x - y \in I$.
- Обозначение: $x \equiv y \pmod{I}$ или $x \equiv_I y$.

Замечание

- Понятие сравнения по модулю идеала является прямым обобщением понятия сравнения целых чисел по модулю целого числа: для целых чисел сравнимость по модулю m равносильна сравнимости по модулю идеала (m).
- Соответственно, сравнения по модулю идеала имеют практически те же свойства, что и сравнения по модулю целых чисел.

Факторкольцо (2/4)

Теорема 31

Пусть I — идеал в кольце R. Тогда отношение "a сравнимо с b по модулю I" является отношением эквивалентности.

Доказательство.

- Рефлексивность. $a - a = 0 \in I$, следовательно, $a \equiv a \pmod{I}$.
- Симметричность. Пусть $a \equiv b \pmod{I}$. Тогда $a - b \in I$, следовательно, $b - a = 0 - (a - b) \in I$, откуда $b \equiv a \pmod{I}$.
- Транзитивность. Пусть $a \equiv b \pmod{I}$ и $b \equiv c \pmod{I}$. Тогда $a - b \in I$ и $b - c \in I$. Следовательно, $a - c = (a - b) + (b - c) \in I$, откуда $a \equiv c \pmod{I}$. □

Замечание

- Из этого следует, что множество элементов кольца R можно разбить на классы эквивалентности по отношению сравнимости по модулю I. Эти классы эквивалентности называются *классами вычетов* или *смежными классами*.
- Легко видеть, что классом вычетов, содержащим элемент $a \in R$, является множество $a + I \stackrel{\text{def}}{=} \{a + r \mid r \in I\}$. Этот класс также часто обозначают $\overline{a}$.

Факторкольцо (3/4)

Теорема 32

Пусть $a \equiv b \pmod{I}$ *и* $c \equiv d \pmod{I}$. *Тогда*

1. $a \pm c \equiv b \pm d \pmod{I}$;
2. $ac \equiv bd \pmod{I}$.

Доказательство.

1. $(a \pm c) - (b \pm d) = (a - b) \pm (c - d) \in I$;
2. $ac - bd = (ac - bc) + (bc - bd) = (a - b)c + b(c - d) \in I$. □

Замечание

Это означает, что на множестве классов вычетов по модулю I можно корректно ввести операции сложения и умножения следующим образом:

- $(a + I) + (b + I) \stackrel{\text{def}}{=} (a + b) + I$;
- $(a + I)(b + I) \stackrel{\text{def}}{=} ab + I$.

Факторкольцо (4/4)

Теорема 33

Множество классов вычетов по модулю I с введенными выше операциями сложения и умножения образует коммутативное кольцо с единицей.

Доказательство. Свойства коммутативности, ассоциативности и дистрибутивности очевидно следуют из аналогичных свойств в кольце R.

- Нулевым элементом является класс $0 + I = I$.
- Обратным по сложению к $a + I$ будет класс $(-a) + I$.
- Единичным элементом является класс $1 + I$. □

Определение 25

Это кольцо называется *факторкольцом* и обозначается A/I.

Пример

Кольцо вычетов по модулю m является фактором кольца $\mathbb{Z}$ по модулю идеала $m\mathbb{Z} = (m)$. Этим объясняется обозначение $\mathbb{Z}/m\mathbb{Z}$.

Еще один пример факторкольца

- Рассмотрим кольцо $\mathbb{R}[x]$ многочленов с вещественными коэффициентами и его идеал $I = (x^2 + 1)$.
- Заметим, что $x^2 \equiv -1 \pmod{I}$, поэтому в каждом классе вычетов по модулю I есть ровно один многочлен вида $a + bx$, где $a, b \in \mathbb{R}$.
  - То есть $\mathbb{R}[x]/(x^2 + 1) = \{\overline{a + bx} \mid a, b \in \mathbb{R}\}$.
- Посмотрим на то, как выполняются арифметические действия в кольце $\mathbb{R}[x]/(x^2 + 1)$.
  - $\overline{a + bx} + \overline{c + dx} = \overline{(a + c) + (b + d)x}$;
  - $\overline{a + bx} \cdot \overline{c + dx} = \overline{ac + (ad + bc)x + bdx^2} = \overline{(ac - bd) + (ad + bc)x}$.
- Заметим, что эти формулы очень похожи на формулы сложения и умножения комплексных чисел: если всюду заменить x на i, то получатся в точности эти формулы.

Определение 26

- Кольца K и L называются *изоморфными*, если существует такая биекция $f : K \to L$, что $\forall a, b \in K\,(f(a + b) = f(a) + f(b))$ и $\forall a, b \in K\,(f(ab) = f(a)f(b))$.
- Обозначение: $K \cong L$. Биекция f называется *изоморфизмом* колец K и L.

Замечание

Фактически мы доказали, что $\mathbb{R}[x]/(x^2 + 1) \cong \mathbb{C}$.

Простые и максимальные идеалы

Определение 27

Собственный идеал кольца R — любой его идеал, отличный от (0) и R.

Определение 28

Пусть R — коммутативное кольцо с единицей; I — собственный идеал в R. Тогда

- идеал I — *простой*, если $\forall a, b \in R\,(ab \in I \to (a \in I \vee b \in I))$;
- идеал I — *максимальный*, если не существует такого идеала J, что $I \subsetneq J \subsetneq R$.

Замечание

- То есть максимальный идеал — это максимальный по включению собственный идеал кольца.
- В кольце $\mathbb{Z}$ понятия простого и максимального идеала совпадают. А именно, простыми и максимальными являются те и только те идеалы, которые порождены простыми числами.

Теорема о факторкольце по простому идеалу

Теорема 34

Собственный идеал I кольца R является простым,
если и только если R/I — область целостности.

Доказательство. "$\Rightarrow$": Пусть I — простой идеал.

- Рассмотрим такие $\overline{a}, \overline{b} \in R/I$, что $\overline{a}\overline{b} = \overline{0}$.
  - Тогда $\overline{ab} = \overline{a}\overline{b} = \overline{0}$, следовательно, $ab \in I$.
  - Поскольку I — простой идеал, получаем, что либо $a \in I$, либо $b \in I$.
  - Но тогда либо $\overline{a} = \overline{0}$, либо $\overline{b} = \overline{0}$.
- Таким образом, R/I — область целостности.

"$\Leftarrow$": Пусть R/I — область целостности.

- Рассмотрим такие $a, b \in R$, что $ab \in I$.
  - Тогда в R/I имеем $\overline{a}\overline{b} = \overline{ab} = \overline{0}$.
  - Поскольку R/I — область целостности, получаем, что либо $\overline{a} = \overline{0}$, либо $\overline{b} = \overline{0}$.
  - Но тогда либо $a \in I$, либо $b \in I$.
- Таким образом, I — простой идеал. $\square$

Теорема о факторкольце по максимальному идеалу (1/2)

Теорема 35

Собственный идеал I кольца R является максимальным,
если и только если R/I — поле.

Доказательство. "$\Rightarrow$": Пусть I — максимальный идеал.

- Заметим, что R/I — коммутативное кольцо с единицей.
- То есть нам нужно доказать, что все ненулевые элементы R/I обратимы.
- Пусть $\overline{a} \in R/I$ и $\overline{a} \neq \overline{0}$.
- Рассмотрим множество $J = (a) + I \stackrel{\text{def}}{=} \{ax + b \mid x \in R, b \in I\}$.
- Докажем, что J — идеал в R. Действительно,
  - $0 = a \cdot 0 + 0 \in J$;
  - если $x_1, x_2 \in R$ и $b_1, b_2 \in I$, то $(ax_1 + b_1) \pm (ax_2 + b_2) = a(x_1 \pm x_2) + (b_1 \pm b_2) \in J$;
  - если $x, y \in R$ и $b \in I$, то $(ax + b)y = a(xy) + (by) \in J$.
- Заметим, что $I \subsetneq J$, следовательно, $J = R$.
- Тогда $1 \in J$, следовательно, $1 = ax + b$, где $x \in R$ и $b \in I$.
- Но тогда $ax \equiv 1 \pmod{I}$, то есть $\overline{x} = (\overline{a})^{-1}$ в R/I.

Теорема о факторкольце по максимальному идеалу (2/2)

"$\Leftarrow$": Пусть R/I — поле.

- Рассмотрим такой идеал J кольца R, что $I \subsetneq J$. Нужно доказать, что $J = R$.
- Пусть $a \in J \setminus I$. Тогда $\overline{a} \neq \overline{0}$ в R/I.
- Рассмотрим такой элемент $x \in R$, что $\overline{x} = (\overline{a})^{-1}$ в R/I.
- Тогда $ax \equiv 1 \pmod{I}$, то есть $ax - 1 = b \in I$.
- Следовательно, $1 = ax - b \in J$, откуда $J = R$. □

Следствие

Любой максимальный идеал является простым.

Доказательство.

I — максимальный $\Rightarrow$ R/I — поле $\Rightarrow$ R/I — область целостности $\Rightarrow$ I — простой. □

Замечание

- Обратное неверно. Например, в кольце $\mathbb{Z}[x]$ идеал (x) — простой, но не максимальный.
  - Действительно, $(x) \subsetneq (x, 2) \subsetneq \mathbb{Z}[x]$.

2.2. Кольцо многочленов над полем

Пусть K — произвольное поле. Мы будем рассматриваться многочлены от одной переменной с коэффициентами из поля K (многочлены над K).

Определение 29

- *Многочленом* с коэффициентами из поля K называется бесконечная последовательность $(a_0, a_1, a_2, \dots)$ элементов поля K, все члены которой, начиная с некоторого места, равны нулю.
- Члены данной последовательности называются *коэффициентами* многочлена.
- Множество всех многочленов с коэффициентами из поля K обозначается $K[x]$.

Определение 30

Пусть $f = (a_0, a_1, \dots) \in K[x]$, $x \in K$. Тогда *значением* многочлена f в точке x называется величина $f(x) = a_0 + a_1 x + a_2 x^2 + \dots$.

Замечание

Таким образом, каждому многочлену $f \in K[x]$ можно поставить в соответствие функцию $f : K \to K$.

Формальное и функциональное равенство многочленов

Замечание

• Иногда многочленом называют описанную выше функцию. Однако здесь важно понимать существенное различие между этими двумя подходами к понятию многочлена.

- Если под многочленом понимать последовательность коэффициентов, то два многочлена $f = (a_0, a_1, \ldots)$ и $g = (b_0, b_1, \ldots)$ следует считать равными, если равны их соответствующие коэффициенты (то есть $\forall i \ (a_i = b_i)$ — так называемое *формальное равенство многочленов*).
- Если же под многочленом понимать функцию из K в K, то многочлены f и g следует считать равными, если при всех x их значения равны (то есть $\forall x \in K \ (f(x) = g(x))$ — *функциональное равенство многочленов*).

• Очевидно, что формальное равенство влечет функциональное, однако обратное верно не всегда. Ниже будет доказано, что это верно в случае, если поле K бесконечно. Однако для конечных полей это не верно.

Пример

• Пусть $p \in \mathbb{P}$. Рассмотрим многочлены x и x^p из $\mathbb{F}_p[x]$.

• Поскольку по малой теореме Ферма $\forall x \in \mathbb{Z} \ (x^p \equiv x \pmod p)$,

эти многочлены равны функционально, но, разумеется, не равны формально.

Кольцо многочленов над полем: корни и коэффициенты

Определение 31

Элемент $x_0 \in K$ называется *корнем* многочлена f, если $f(x_0) = 0$.

Определение 32

Пусть $f = (a_0, a_1, a_2, \ldots) \in K[x]$.

- Если у многочлена f есть ненулевые коэффициенты, то *степенью* многочлена f называется число

$$\deg f \stackrel{\text{def}}{=} \max\{n \mid a_n \neq 0\}.$$

  - В случае $\forall n\,(a_n = 0)$, будем считать, что $\deg f \stackrel{\text{def}}{=} -\infty$.
- *Свободным членом* многочлена f называется коэффициент a_0.
- *Старшим коэффициентом* многочлена f называется коэффициент a_n, где $n = \deg f$.
- Многочлен называется *унитарным*, если его старший коэффициент равен 1.

Кольцо многочленов над полем: арифметические действия (1/2)

Определение 33

Пусть $f = (a_0, a_1, \ldots), g = (b_0, b_1, \ldots) \in K[x]$.

- *Суммой* многочленов f и g называется многочлен $f + g \stackrel{\text{def}}{=} (a_0 + b_0, a_1 + b_1, \ldots)$.
- *Произведением* многочленов f и g называется многочлен $f \cdot g = (c_0, c_1, \ldots)$,

где $c_i \stackrel{\text{def}}{=} a_0 b_i + a_1 b_{i-1} + \ldots + a_i b_0$.

Замечание

Такие определения суммы и произведения многочленов полностью согласуются с их значениями. Т. е. для любого $x \in K$ имеем $(f + g)(x) = f(x) + g(x)$ и $(fg)(x) = f(x)g(x)$.

Теорема 36

Множество $K[x]$ с введенными выше операциями сложения и умножения является коммутативным кольцом с единицей.

Доказательство. Поскольку сложение многочленов осуществляется покоэффициентно, все свойства сложения (коммутативность, ассоциативность, существование нейтрального и обратного элементов) непосредственно следуют из аналогичных свойств в поле K.

Кольцо многочленов над полем: арифметические действия (2/2)

- Также легко видеть, что последовательность $(1, 0, 0, \ldots)$ является единичным элементом в $K[x]$.
- Осталось доказать коммутативность и ассоциативность умножения, а также дистрибутивность.
- Пусть $f = (a_0, a_1, \ldots), g = (b_0, b_1, \ldots), h = (c_0, c_1, \ldots) \in K[x]$;
  - также обозначим $fg = (d_0, d_1, \ldots)$, $fh = (e_0, e_1, \ldots)$.
- Коммутативность умножения. Пусть $gf = (d'_0, d'_1, \ldots)$.
  - Тогда $d_k = \sum_{i=0}^{k} a_i b_{k-i} = \sum_{j=0}^{k} b_j a_{k-j} = d'_k$.
- Дистрибутивность. Пусть $f \cdot (g + h) = (p_0, p_1, \ldots)$.
  - Тогда $p_k = \sum_{i=0}^{k} a_i (b_{k-i} + c_{k-i}) = \sum_{i=0}^{k} a_i b_{k-i} + \sum_{i=0}^{k} a_i c_{k-i} = d_k + e_k$.
- Ассоциативность умножения. Пусть $(fg)h = (q_0, q_1, \ldots)$ и $f(gh) = (r_0, r_1, \ldots)$.
  - Тогда $q_s = \sum_{i=0}^{s} d_i c_{s-i} = \sum_{i=0}^{s} \left(\sum_{j=0}^{i} a_j b_{i-j} \right) c_{s-i} = \sum_{i+j+k=s} a_i b_j c_k$.
  - Аналогично доказывается, что $r_s = \sum_{i+j+k=s} a_i b_j c_k$.
  - Следовательно, $(fg)h = f(gh)$. □

Константы и переменная

- Строго говоря, само поле K не является подмножеством кольца многочленов $K[x]$.
- Однако, в $K[x]$ есть многочлены вида $(a, 0, 0, \ldots)$, которые обладают теми же свойствами, что и элементы поля K.

Определение 34

Многочлен $f \in K[x]$ называется *константой*, если $\deg f \leq 0$.

Утверждение

Пусть $a, b \in K$. *Тогда* $(a, 0, 0, \ldots) + (b, 0, 0, \ldots) = (a + b, 0, 0, \ldots)$ *и*

$$(a, 0, 0, \ldots)(b, 0, 0, \ldots) = (ab, 0, 0, \ldots).$$

- Тем самым, множество K_0 всех констант является подполем кольца $K[x]$, причем $K \cong K_0$. Далее, мы будем отождествлять множества K и K_0 и говорить о константных многочленах как об элементах поля K.
- Введем еще одно обозначение: пусть $x = (0, 1, 0, 0, \ldots 0)$.
  - По индукции легко доказать, что $x^n = (\underbrace{0, \ldots, 0}_{n}, 1, 0, 0, \ldots)$.
  - Тогда $(a_0, a_1, a_2, \ldots) = a_0 + a_1 x + a_2 x^2 + \ldots$.
- Далее, мы будем использовать привычную форму записи: $f(x) = a_n x^n + \ldots + a_1 x + a_0$.

Деление многочленов с остатком (1/2)

Теорема 37 (о делении многочленов с остатком)

Пусть $f, g \in K[x]$, $g \neq 0$. *Тогда существует единственная пара многочленов* $q, r \in K[x]$ *такая, что* $f = gq + r$ *и* $\deg r < \deg g$.

Доказательство. "$\exists$" Индукция по $n = \deg f$.

База ($\deg f < \deg g$): $q = 0, r = f$ подходят.

Переход ($-\infty, 0, 1, \ldots, n-1 \to n$):

Пусть $f = a_n x^n + \ldots + a_0$, $g = b_m x^m + \ldots + b_0$, причём $a_n \neq 0$, $b_m \neq 0$, и $n \geq m$.

- Рассмотрим многочлен $f_1 \stackrel{\text{def}}{=} f - g \cdot \frac{a_n}{b_m} x^{n-m}$.
  - Заметим, что $\deg f_1 < n$. Тогда по индукционному предположению существуют такие q_1 и r_1, что $f_1 = gq_1 + r_1$ и $\deg r_1 < \deg g$.
  - Следовательно, $f = f_1 + g \cdot \frac{a_n}{b_m} x^{n-m} = g(q_1 + \frac{a_n}{b_m} x^{n-m}) + r_1$.
  - Таким образом, подходят многочлены $q \stackrel{\text{def}}{=} q_1 + \frac{a_n}{b_m} x^{n-m}$ и $r \stackrel{\text{def}}{=} r_1$.

Деление многочленов с остатком (2/2)

"!" Пусть $f = gq_1 + r_1 = gq_2 + r_2$.

- Тогда $g(q_1 - q_2) + (r_1 - r_2) = 0$.
  - Если $q_1 \neq q_2$, то $\deg(g(q_1 - q_2)) \geq \deg g > \deg(r_1 - r_2)$, что невозможно.
  - Следовательно, $q_1 = q_2$, а тогда и $r_1 = r_2$. □

Определение 35

Многочлен q из условия предыдущей теоремы называется *неполным частным*, а многочлен r — *остатком* от деления f на g.

Замечание

1. Доказательство существования неполного частного и остатка фактически дает алгоритм их построения (*деление в столбик*).
2. Поскольку в процессе деления с остатком нужно делить на старший коэффициент многочлена g, здесь важно то, что K — поле. В случае, если K — коммутативное кольцо, но не поле, теорема 37 неверна. В частности, при делении с остатком многочленов из $\mathbb{Z}[x]$, коэффициенты как неполного частного, так и остатка, могут оказаться рациональными.
   - Например, $x^2 + 1 = (2x - 1)(\frac{1}{2}x + \frac{1}{4}) + \frac{5}{4}$.
   - Однако, если многочлен g унитарный, то при делении на него неполное частное и остаток всегда будут многочленами с целыми коэффициентами.

Теорема Безу

Теорема 38 (Безу)

Пусть $f \in K[x]$ и $a \in K$. Тогда остаток от деления многочлена $f(x)$ на многочлен $x - a$ равен $f(a)$.

Доказательство. Поскольку $\deg(x - a) = 1$, остаток от деления $f(x)$ на $x - a$ будет константой. Обозначим её через r. Пусть $q(x)$ — неполное частное.

- Тогда $f(x) = (x - a)q(x) + r$.
- Подставив $x = a$, получим $f(a) = (a - a)q(a) + r = r$. □

Следствие

Если a — корень многочлена $f(x)$, то многочлен $f(x)$ можно представить в виде $f(x) = (x - a)q(x)$.

Замечание

Иногда теоремой Безу называют приведенное выше следствие.
Но это неправильно.

Делимость многочленов (1/2)

Определение 36

Пусть $f, g \in K[x]$. Многочлен f *делится* на многочлен g, если $g \neq 0$ и существует такой многочлен $h \in K[x]$, что $f = gh$. Обозначение: $f \vdots g$ или $g \mid f$.

Определение 37

- Делитель g многочлена f называется *тривиальным*, если $\deg g = 0$ или $\deg g = \deg f$.
- Многочлен f называется *неприводимым*, если у него нет нетривиальных делителей, и *приводимым* в противном случае.

Замечание

- Приводимость или неприводимость многочлена может зависеть от того, над каким полем он рассматривается.
- Например, многочлен $x^2 - 2$, очевидно, неприводим над $\mathbb{Q}$
  - то есть у него нет нетривиальных делителей с рациональными коэффициентами;
  - однако, над $\mathbb{R}$ он приводим: $x^2 - 2 = (x - \sqrt{2})(x + \sqrt{2})$.
- Еще более интересным примером является многочлен $x^4 + 1$. Он также неприводим над $\mathbb{Q}$;
  - раскладывается в произведение двух неприводимых множителей над $\mathbb{R}$: $x^4 + 1 = (x^2 - \sqrt{2}x + 1)(x^2 + \sqrt{2}x + 1)$;
  - и раскладывается в произведение четырех неприводимых множителей над $\mathbb{C}$: $x^4 + 1 = \left(x - \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i\right)\left(x - \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i\right)\left(x + \frac{\sqrt{2}}{2} - \frac{\sqrt{2}}{2}i\right)\left(x + \frac{\sqrt{2}}{2} + \frac{\sqrt{2}}{2}i\right)$.

Делимость многочленов (2/2)

Определение 38

• Пусть $f, g, h \in K[x]$. Многочлен h называется *общим делителем* многочленов f и g, если $f \vdots h$ и $g \vdots h$.

• Многочлен h называется *наибольшим общим делителем* многочленов f и g, если он является их общим делителем и имеет наибольшую степень среди всех общих делителей. Обозначение: $h = (f, g)$.

Замечание

• В отличии от наибольшего общего делителя целых чисел, наибольший общий делитель многочленов определен неоднозначно.

- Действительно, если h — наибольший общий делитель f и g, и $c \in K \setminus \{0\}$ то многочлен ch также, очевидно, является общим делителем f и g, и имеет ту же степень, что и h.
- Однако можно доказать, что все наибольшие общие делители f и g отличаются друг от друга домножением на константу.

• Этой неоднозначности можно избежать, если рассматривать в качестве наибольшего общего делителя только унитарный многочлен. Однако нам будет удобнее считать, что любой многочлен, удовлетворяющий условиям из определения, является наибольшим общим делителем f и g.

- При этом обозначение (f, g) фактически означает класс всех НОДов f и g, а запись $h = (f, g)$ — принадлежность многочлена h этому классу.

Алгоритм Евклида в кольце многочленов (1/3)

- Также как и для целых чисел, НОД двух многочленов f и g можно найти при помощи алгоритма Евклида. Работает этот алгоритм полностью аналогично алгоритму для целых чисел.
- Вход: многочлены $f, g \in K[x] \setminus \{0\}$.
  - На первом шаге нужно разделить с остатком f на g;
  - на втором — g на полученный на первом шаге остаток r_1;
  - на третьем — r_1 на r_2 (остаток, полученный на 2-м шаге). И т.д.
  - На k-м шаге r_{k-2} делится с остатком на r_{k-1} и получается остаток r_k.
  - Алгоритм заканчивает работу когда очередной остаток станет равным нулю.
- Выход: последний ненулевой остаток (обозначим его r_{n+1}).

Алгоритм Евклида в кольце многочленов (2/3)

Схема работы алгоритма Евклида в кольце многочленов

- Вход: $f, g \in K[x] \setminus \{0\}$.

$$
\begin{array}{ll}
f = gq_1 + r_1 & \deg r_1 < \deg g \\
g = r_1 q_2 + r_2 & \deg r_2 < \deg r_1 \\
r_1 = r_2 q_3 + r_3 & \deg r_3 < \deg r_2 \\
\quad\vdots & \quad\vdots \\
r_{k-2} = r_{k-1} q_k + r_k & \deg r_k < \deg r_{k-1} \\
\quad\vdots & \quad\vdots \\
r_{n-1} = r_n q_{n+1} + r_{n+1} & \deg r_{n+1} < \deg r_n \\
r_n = r_{n+1} q_{n+2} + 0 &
\end{array}
$$

- Выход: $r_{n+1} = (f, g)$.

Алгоритм Евклида в кольце многочленов (3/3)

Теорема 39

$f, g \in K[x] \setminus \{0\}$. *Тогда алгоритм Евклида для f и g заканчивает работу за конечное число шагов и дает на выходе наибольший общий делитель многочленов f и g.*

Доказательство. 1. Алгоритм Евклида заканчивает работу за конечное число шагов, поскольку степени остатков постоянно убывают:
$\deg g > \deg r_1 > \deg r_2 > \ldots$, а бесконечно убывать они не могут.

2. Докажем, что $(f, g) = (g, r_1) = (r_1, r_2) = \ldots = (r_n, r_{n+1})$.
   - Действительно, из равенства $r_{k-1} = r_k q_{k+1} + r_{k+1}$ следует, что множество общих делителей многочленов r_{k-1} и r_k совпадает с множеством общих делителей r_k и r_{k+1}.
   - Тогда и множества наибольших общих делителей у этих пар совпадают.
   - То есть равенство $(r_{k-1}, r_k) = (r_k, r_{k+1})$ выполнено при всех k (мы здесь считаем, что $f = r_{-1}$ и $g = r_0$).

- Осталось заметить, что $r_n \vdots r_{n+1}$, следовательно, $(r_n, r_{n+1}) = r_{n+1}$. □

Линейное представление НОД многочленов (1/3)

Следствие 1

Пусть $f, g \in K[x] \setminus \{0\}$. *Тогда существуют такие многочлены* $s, t \in K[x]$, *что* $fs + gt = (f, g)$.

Доказательство. Докажем индукцией по k, что при всех $k \in [0..n]$ существуют такие $s_k, t_k \in K[x]$, что $(f, g) = r_{n+1} = r_{k-1}s_k + r_k t_k$.

- База ($k = n$): $r_{n+1} = r_{n-1} - r_n q_{n+1}$. То есть $s_n = 1$ и $t_n = -q_{n+1}$ подходят.
- Переход ($k \to k-1$):
  - $r_{n+1} = r_{k-1}s_k + r_k t_k = r_{k-1}s_k + (r_{k-2} - r_{k-1}q_k)t_k = r_{k-2}t_k + r_{k-1}(s_k - q_k t_k)$.
  - То есть $s_{k-1} = t_k$ и $t_{k-1} = s_k - q_k t_k$ подходят.
- В итоге при $k = 0$ получим, что $r_{n+1} = r_{-1}s_0 + r_0 t_0 = fs_0 + gt_0$.
  - Таким образом, многочлены $s = s_0$ и $t = t_0$ удовлетворяют условию. □

Замечание

Как и в случае с натуральными числами, поиск линейного представления НОД многочленов удобно записывать как цепочку равенств.

- $(f, g) = r_{n+1} = r_{n-1} - r_n q_{n+1} = r_{n-1} - (r_{n-2} - r_{n-1}q_n)q_{n+1} =$
  $= r_{n-1}(1 + q_n q_{n+1}) - r_{n-2}q_{n+1} = \ldots = fs + gt.$

Линейное представление НОД многочленов (2/3)

Следствие 2

Любой общий делитель двух многочленов является делителем их НОДа.

Замечание

1. Из последнего следствия немедленно вытекает упоминавшееся выше утверждение о том, что все НОДы двух многочленов отличаются друг от друга домножением на константу.
2. Как и в случае натуральных чисел, алгоритм поиска линейного представления НОД многочленов часто называют *расширенным алгоритмом Евклида*.
3. В отличии от алгоритма Евклида для натуральных чисел, в случае многочленов нет нетривиальных оценок на число делений с остатком в алгоритме Евклида.

   • Из доказательства теоремы 39 очевидно следует, что число делений с остатком не превосходит $\deg g + 1$.

   • Однако улучшить эту оценку нельзя, поскольку легко построить пример, когда при каждом делении с остатком степень уменьшается ровно на 1.

Линейное представление НОД многочленов (3/3)

Пример

Пусть $f(x) = 2x^4 + 6x^3 + 4x^2 + 6x - 4$, $g(x) = 2x^3 + 2x - 2$.

$$
\begin{aligned}
2x^4 + 6x^3 + 4x^2 + 6x - 4 &= (2x^3 + 2x - 2)(x + 3) + (2x^2 + 2x + 2) \\
2x^3 + 2x - 2 &= (2x^2 + 2x + 2)(x - 1) + 2x \\
2x^2 + 2x + 2 &= 2x \cdot (x + 1) + 2 \\
2x &= 2 \cdot x + 0.
\end{aligned}
$$

$$
\begin{aligned}
(f, g) = 2 &= (2x^2 + 2x + 2) - 2x \cdot (x + 1) = \\
&= (2x^2 + 2x + 2) - ((2x^3 + 2x - 2) - (2x^2 + 2x + 2)(x - 1))(x + 1) = \\
&= (2x^2 + 2x + 2)x^2 - (2x^3 + 2x - 2)(x - 1) = \\
&= ((2x^4 + 6x^3 + 4x^2 + 6x - 4) - (2x^3 + 2x - 2)(x + 3))x^2 - (2x^3 + 2x - 2)(x - 1) = \\
&= (2x^4 + 6x^3 + 4x^2 + 6x - 4)x^2 - (2x^3 + 2x - 2)(x^3 + 3x^2 + x + 1) = \\
&= f(x)x^2 - g(x)(x^3 + 3x^2 + x + 1).
\end{aligned}
$$

Идеалы в кольце многочленов над полем

Теорема 40

$K[x]$ — *кольцо главных идеалов.*

Доказательство. Пусть $I \subset K[x]$ — такой идеал, что $I \neq (0)$.

- Тогда в I есть ненулевые многочлены.
- Рассмотрим многочлен $g \in I$, имеющий наименьшую степень среди всех ненулевых элементов I.
- Очевидно, что $(g) \subset I$. Докажем, что $I \subset (g)$.
  - Пусть $f \in I$.
  - Поделим f на g с остатком: $f = gq + r$, где $q, r \in K[x]$ и $\deg r < \deg g$.
  - Тогда $r = f - gq \in I$. Поскольку $\deg r < \deg g$ это означает, что $r = 0$.
  - Таким образом, $f = gq \in (g)$. □

Замечание

- Здесь важно, что K — поле. В случае, когда K — коммутативное кольцо, не являющееся полем, это утверждение неверно.
- Например, мы уже видели, что кольцо $\mathbb{Z}[x]$ — не КГИ.

2.3. Разложение на множители в кольце многочленов над полем

Обратимые элементы кольца

Определение 39

Пусть R — коммутативное целостное кольцо с единицей.

- Элемент $a \in R$ называется *обратимым*, если $\exists b \in R\ (ab = 1)$.
- Множество всех обратимых элементов кольца R обозначается через R^*.

Примеры

1. $\mathbb{Z}^* = \{-1, 1\}$;
2. если K — поле, то $K^* = K \setminus \{0\}$;
3. если K — поле, то $K[x]^* = K \setminus \{0\}$ (то есть обратимыми элементами в $K[x]$ являются ненулевые константы и только они).

Определение 40

- Элементы $a, b \in R$ называются *ассоциированными*, если $\exists \varepsilon \in R^*\ (a = b\varepsilon)$.
- Обозначение: $a \sim b$.

Факториальные кольца (1/2)

Определение 41

Элемент $p \in R$ называется *неприводимым*, если $p \notin R^*$ и $\forall a, b \in R\,(p = ab \to (a \in R^* \vee b \in R^*))$.

Замечание

Другими словами, элемент p неприводим, если $p \notin R^* \cup \{0\}$ и p нельзя разложить в произведение двух необратимых элементов.

Определение 42

Пусть R — целостное коммутативное кольцо с единицей.

• Кольцо R называется *факториальным*, если любой его ненулевой и необратимый элемент можно представить в виде произведения неприводимых, причем это представление единственно с точностью до порядка сомножителей и домножения их на обратимые элементы кольца.

Замечание

То есть, если $a = p_1 \dots p_n = q_1 \dots q_m$, где все p_i и q_j неприводимы, то $m = n$ и существуют такие $\varepsilon_1, \dots, \varepsilon_n \in R^*$ и биекция $\sigma : [1..n] \to [1..n]$, что $\forall i \in [1..n]\,(p_i = \varepsilon_i q_{\sigma(i)})$.

Факториальные кольца (2/2)

Примеры

1. Кольцо $\mathbb{Z}$ факториально (основная теорема арифметики).
   - Кольцо факториально, если в нем выполняется аналог основной теоремы арифметики.

2. Пример нефакториального кольца: $\mathbb{Z}[\sqrt{-5}] \stackrel{\text{def}}{=} \{a + b\sqrt{-5} \mid a, b \in \mathbb{Z}\}$.

• В этом кольце число 9 имеет два разложения: $9 = 3 \cdot 3 = (2 + \sqrt{-5})(2 - \sqrt{-5})$.

• Для того, чтобы доказать, что множители в этом разложении неприводимы и не ассоциированы, введем в этом кольце понятие нормы. Пусть $\alpha = a + b\sqrt{-5} \in \mathbb{Z}[\sqrt{-5}]$.

- Тогда *нормой* элемента α называется величина $N(a + b\sqrt{-5}) \stackrel{\text{def}}{=} \alpha \cdot \overline{\alpha} = a^2 + 5b^2$.
- Очевидно, что $N(\alpha) \in \mathbb{N}_0$, причем $N(\alpha) = 0 \Leftrightarrow \alpha = 0$.
- Также заметим, что $N(\alpha\beta) = \alpha\beta \cdot \overline{\alpha\beta} = (\alpha \cdot \overline{\alpha})(\beta \cdot \overline{\beta}) = N(\alpha)N(\beta)$.
- Далее, $\alpha \in \mathbb{Z}[\sqrt{-5}]^* \Leftrightarrow N(\alpha) = 1$.
  - Действительно, если $\alpha\beta = 1$, то $N(\alpha)N(\beta) = N(\alpha\beta) = 1$, следовательно, $N(\alpha) = 1$.
  - Обратно, если $N(\alpha) = 1$, то $\alpha \cdot \overline{\alpha} = 1$, то есть $\overline{\alpha} = \alpha^{-1}$.
- Это означает, что $\mathbb{Z}[\sqrt{-5}]^* = \{-1, 1\}$.

• Заметим, что $N(3) = N(2 + \sqrt{-5}) = N(2 - \sqrt{-5}) = 9$, а элементов с нормой 3 в данном кольце нет. Следовательно, все эти множители неприводимы и неассоциированы.

Факториальность кольца многочленов над полем (1/2)

Лемма

Пусть $f, g, h \in K[x]$, $fg \vdots h$ *и* h *— неприводим. Тогда либо* $f \vdots h$, *либо* $g \vdots h$.

Доказательство. Пусть $g \not\vdots h$. Тогда $(g, h) = 1$.

- Запишем линейное представление НОД: $1 = gs + ht$, где $s, t \in K[x]$.
- Тогда $f = (fg)s + h(ft) \vdots h$. □

Следствие

Пусть $f_1, f_2, \ldots, f_n, h \in K[x]$, h *— неприводим и* $f_1 f_2 \ldots f_n \vdots h$. *Тогда* $\exists i\, (f_i \vdots h)$.

Доказательство. Индукция по n.

- База для $n = 2$ доказана в лемме.
- Переход $n \to n + 1$: Пусть $(f_1 f_2 \ldots f_n) f_{n+1} \vdots h$. Тогда по лемме,
  - либо $f_{n+1} \vdots h$ и все, что нужно, доказано;
  - либо $f_1 f_2 \ldots f_n \vdots h$, откуда по индукционному предположению $\exists i \in [1..n]\, (f_i \vdots h)$.

□

Факториальность кольца многочленов над полем (2/2)

Теорема 41 (Основная теорема арифметики для многочленов)

Пусть K — поле. Тогда кольцо $K[x]$ факториально.

Доказательство. "$\exists$" Докажем, что у любого многочлена f, где $\deg f > 0$, есть разложение на неприводимые множители.

- Индукция по $n = \deg f$. База для $n = 1$ очевидна.
- $1, \ldots, n \to n+1$: Возможны два случая.
  - Если f неприводим, то разложение состоит из одного множителя.
  - Иначе $f = gh$, где $0 < \deg g < n$ и $0 < \deg h < n$. Тогда по индукционному предположению, g и h можно разложить на неприводимые множители. Перемножив, получим разложение f.

"!" Пусть $f = p_1 \ldots p_s = q_1 \ldots q_t$ — многочлен наименьшей степени, имеющее два различных разложения на неприводимые множители.

- Тогда $p_1 \ldots p_s \vdots q_1$, следовательно, один из множителей p_i делится на q_1.
  - Поскольку p_i и q_1 — неприводимы, имеем $p_i = \varepsilon q_1$, где $\varepsilon \in K[x]^*$.
- Пусть, не умаляя общности, $p_1 = \varepsilon q_1$. Тогда $(\varepsilon p_2)p_3 \ldots p_s = q_2 \ldots q_t$ — многочлен степени меньшее, чем n, имеющий два разложения. Противоречие. $\square$

Евклидовы кольца

• Можно заметить, что доказательства многих утверждений в кольцах $\mathbb{Z}$ и $K[x]$ очень похожи: они основаны на делении с остатком и алгоритме Евклида. На самом деле, эти кольца являются частными случаями более общего класса колец.

Определение 43

Целостное коммутативное кольцо R называется *евклидовым*, если существует такая функция $g : R \setminus \{0\} \to \mathbb{N}_0$, что

1. $\forall a, b \in R \setminus \{0\}\,(g(ab) \geq g(a))$;
2. Для любых $a, b \in R$, где $b \neq 0$, существуют такие $q, r \in R$, что $a = bq + r$ и либо $r = 0$, либо $g(r) < g(b)$.

Примеры

1. Кольцо $\mathbb{Z}$ евклидово. Здесь $g(a) = |a|$.
2. Кольцо $K[x]$, где K — поле, евклидово. Здесь $g(f) = \deg f$.

Замечание

• Неформально, евклидово кольцо — это кольцо, в котором можно делить с остатком.
• При этом не требуется единственность неполного частного и остатка — только существование!
• Можно доказать, что любое евклидово кольцо — КГИ и что любое КГИ — факториально. (б/д)

Кратные корни (1/2)

- Пусть $x_1, x_2, \ldots, x_s$ — корни многочлена $f(x) \in K[x]$.
- Тогда $x - x_1, x - x_2, \ldots, x - x_s$ — неприводимые делители многочлена f.
- Рассмотрим отдельно эти делители и показатели, с которыми они входят в разложение f на неприводимые множители. Многочлен f можно представить в виде

$$f(x) = (x - x_1)^{a_1}(x - x_2)^{a_2} \ldots (x - x_s)^{a_s} p(x),$$

где $p(x)$ — многочлен, не имеющий корней.

Определение 44

- Число a_i называется *кратностью* корня x_i многочлена f;
- сумма $a_1 + a_2 + \ldots + a_s$ называется *количеством корней многочлена f с учетом кратности*.
- Корень x_i называется *кратным*, если его кратность больше единицы.

Замечание

Во многих случаях бывает удобно считать, что корень кратности k — это k корней, значения которых совпадают.

Кратные корни (2/2)

Теорема 42

Количество корней ненулевого многочлена с учетом кратности не превосходит его степени.

Доказательство. Пусть $f(x) = (x - x_1)^{a_1}(x - x_2)^{a_2} \dots (x - x_s)^{a_s} p(x)$.

- Тогда $\deg f = a_1 + a_2 + \dots + a_s + \deg p \geq a_1 + a_2 + \dots + a_s$. □

Следствие

Пусть $|K| = \infty$, $f, g \in K[x]$ *и* $\forall a \in K\ (f(a) = g(a))$. *Тогда* $f = g$.

Доказательство. Пусть $h(x) = f(x) - g(x)$.

- Тогда все элементы поля K являются корнями многочлена h.
- То есть у h бесконечно много корней.
- По теореме 42 это возможно только при $h = 0$. □

Замечание

Тем самым, в случае $|K| = \infty$ формальное и функциональное равенство многочленов эквивалентны.

Формальная производная многочлена (1/3)

Определение 45

- Пусть $f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 \in K[x]$.
- *Формальной производной* многочлена $f(x)$ называется многочлен

$$f'(x) \stackrel{\text{def}}{=} n a_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \ldots + a_1.$$

Замечание

- Здесь символом n мы обозначаем элемент $\underbrace{1 + \ldots + 1}_{n} \in K$.
- Данное обозначение корректно, поскольку по свойствам ассоциативности и дистрибутивности имеем
  - $(\underbrace{1 + \ldots + 1}_{m}) + (\underbrace{1 + \ldots + 1}_{n}) = \underbrace{1 + \ldots + 1}_{m+n}$;
  - $(\underbrace{1 + \ldots + 1}_{m}) \cdot (\underbrace{1 + \ldots + 1}_{n}) = \underbrace{1 + \ldots + 1}_{mn}$.

Формальная производная многочлена (2/3)

- Наша цель состоит в том, чтобы проверить, что формальная производная обладает теми же свойствами, что и обычная производная функции.
- Для этого мы дадим эквивалентную переформулировку определения формальной производной, сделав его более похожим на классическое определение производной.
- Рассмотрим многочлен от двух переменных $f(x+t)$ и разложим его по степеням t.
  - Формально, *кольцо многочленов от двух переменных* над полем K — это кольцо $K[x,t] \stackrel{\text{def}}{=} (K[x])[t]$.
- Докажем, что коэффициент при t^1 равен как раз $f'(x)$.

Лемма

$f(x+t) = f(x) + f'(x)t + h(x,t)t^2$, *где* $h(x,t) \in K[x,t]$.

Доказательство. Пусть $f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$. Тогда

$$
\begin{aligned}
f(x+t) &= a_n(x+t)^n + a_{n-1}(x+t)^{n-1} + \ldots + a_1(x+t) + a_0 = \\
&= a_n\big(x^n + nx^{n-1}t + \textstyle\sum_{i=2}^{n} C_n^i x^{n-i}t^i\big) + \\
&\quad + a_{n-1}\big(x^{n-1} + (n-1)x^{n-2}t + \textstyle\sum_{i=2}^{n-1} C_{n-1}^i x^{n-1-i}t^i\big) + \ldots + a_1\big(x+t\big) + a_0 = \\
&= \big(a_n x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0\big) + \big(na_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + \ldots + a_1\big)t + \\
&\quad + \big(\textstyle\sum_{k=2}^{n}\sum_{i=2}^{k} a_k C_k^i x^{k-i}t^{i-2}\big)t^2.
\end{aligned}
$$

□

Формальная производная многочлена (3/3)

Следствие

$$f'(x) = \left.\frac{f(x+t)-f(x)}{t}\right|_{t=0}.$$

Теорема 43

1. $(f+g)'(x) = f'(x)+g'(x)$;
2. $(fg)'(x) = f'(x)g(x)+f(x)g'(x)$;
3. $(f(g(x)))' = f'(g(x))g'(x)$.

Доказательство. Пусть
$f(x+t) = f(x)+f'(x)t+h(x,t)t^2$ и $g(x+t) = g(x)+g'(x)t+s(x,t)t^2$. Тогда

1. $(f+g)(x+t) = (f+g)(x)+(f'(x)+g'(x))t+(h(x,t)+s(x,t))t^2$;
2. $$\begin{aligned}(fg)(x+t) &= (f(x)+f'(x)t+h(x,t)t^2)(g(x)+g'(x)t+s(x,t)t^2) = \\ &= (fg)(x)+(f'(x)g(x)+f(x)g'(x))t+w(x,t)t^2;\end{aligned}$$
3. Пусть $p(x,t) \stackrel{\text{def}}{=} g(x+t)-g(x) = t(g'(x)+s(x,t)t)$. Тогда
$$\begin{aligned}f(g(x+t)) &= f(g(x)+p(x,t)) = f(g(x))+f'(g(x))p(x,t)+h(g(x),p(x,t))p^2(x,t) = \\ &= f(g(x))+f'(g(x))g'(x)t+ \\ &\quad +(f'(g(x))s(x,t)+h(g(x),p(x,t))(g'(x)+s(x,t)t)^2)t^2.\end{aligned}$$

□

Формальная производная и кратные корни (1/2)

Теорема 44

Пусть $f \in K[x]$ и $a \in K$ — корень f. Тогда

1. *a — кратный корень f, если и только если a — корень f';*
2. *если a — корень f кратности $k > 1$, то a — корень f' кратности не менее $k-1$.*

Доказательство. Пусть $f(x) = (x-a)^k g(x)$, где $g(a) \neq 0$.

- Тогда $f'(x) = k(x-a)^{k-1}g(x) + (x-a)^k g'(x) = (x-a)^{k-1}(kg(x) + (x-a)g'(x))$.
  1. При $k = 1$ имеем $f'(a) = g(a) + (a-a)g'(a) = g(a) \neq 0$;
     при $k > 1$ имеем $f'(a) = (a-a)^{k-1}(kg(a) + (a-a)g'(a)) = 0$.
  2. Поскольку многочлен $f'(x)$ имеет множитель $(x-a)^{k-1}$,
     кратность корня a у этого многочлена не меньше $k-1$. □

Формальная производная и кратные корни (2/2)

Замечание

- Для таких полей, как $\mathbb{Q}$, $\mathbb{R}$ и $\mathbb{C}$ верно более сильное утверждение:
если a — корень f кратности $k > 1$, то a — корень f' кратности ровно $k - 1$.
  - Действительно, $f'(x) = (x-a)^{k-1}h(x)$, где $h(x) = kg(x) + (x-a)g'(x)$.
  - Но при этом $h(a) = kg(a) \neq 0$, откуда a — корень кратности $k-1$.

- Однако, для других полей это может быть неверным. Дело в том, что в некоторых полях $k = \underbrace{1 + \ldots + 1}_{k}$ может быть равно нулю.
  - Например, рассмотрим многочлен $f(x) = (x-1)^5(3x^2 - x) \in \mathbb{F}_5[x]$.
  - Заметим, что 1 — корень f кратности 5.
  - Но $f'(x) = 5(x-1)^4(3x^2 - x) + (x-1)^5(6x-1) = (x-1)^5(x-1) = (x-1)^6$.
  - То есть 1 — корень f' кратности 6!

Теорема Виета

Теорема 45 (Ф. Виет)

Пусть $f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0 \in K[x]$ *— многочлен, имеющий ровно* n *корней с учетом кратности. Пусть* $x_1, x_2, \ldots, x_n$ *— его корни, причем каждый корень выписан в этой последовательности столько раз, какова его кратность. Тогда*

$$\begin{cases} x_1 + x_2 + \ldots + x_n &= -\frac{a_{n-1}}{a_n} \\ x_1x_2 + x_1x_3 + \ldots + x_{n-1}x_n &= \frac{a_{n-2}}{a_n} \\ & \ldots \\ \sum\limits_{1 \le i_1 < i_2 < \ldots < i_k \le n} x_{i_1}x_{i_2}\ldots x_{i_k} &= (-1)^k \frac{a_{n-k}}{a_n} \\ & \ldots \\ x_1x_2\ldots x_n &= (-1)^n \frac{a_0}{a_n} \end{cases} .$$

Доказательство. $f(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0 = a_n(x - x_1)\ldots(x - x_n)$.

- Приравняв коэффициенты при x^{n-k}, получим $a_{n-k} = a_n(-1)^k \sum\limits_{1 \le i_1 < \ldots < i_k \le n} x_{i_1}\ldots x_{i_k}$.
- Разделив на $(-1)^k a_n$, получим требуемое. □

Основная теорема алгебры (1/2)

Теорема 46 (основная теорема алгебры)

Пусть $f(x) \in \mathbb{C}[x]$, $\deg f > 0$. *Тогда многочлен* f *имеет комплексный корень. (б/д)*

Следствие 1

Пусть $f(x) \in \mathbb{C}[x]$, $\deg f = n > 0$. *Тогда многочлен* f *имеет ровно* n *комплексных корней с учетом кратности.*

Доказательство. Индукция по n. База для $n = 1$ очевидна.

Переход ($n \to n+1$): пусть $f \in \mathbb{C}[x]$ и $\deg f = n+1$.

- По основной теореме алгебры, у f есть корень $x_1 \in \mathbb{C}$.
- Тогда $f(x) = (x - x_1)g(x)$, где $\deg g = n$.
- По индукционному предположению, g имеет ровно n комплексных корней с учетом кратности. Обозначим эти корни $x_2, \ldots, x_{n+1}$.
- Тогда $g(x) = a(x - x_2)\ldots(x - x_{n+1})$, где $a \in \mathbb{C} \setminus \{0\}$.
- Следовательно, $f(x) = a(x - x_1)(x - x_2)\ldots(x - x_{n+1})$. □

Основная теорема алгебры (2/2)

Следствие 2

1. *Неприводимыми многочленами в* $\mathbb{C}[x]$ *являются только линейные многочлены.*
2. *Неприводимыми многочленами в* $\mathbb{R}[x]$ *являются линейные многочлены и квадратные трехчлены с отрицательным дискриминантом.*

Доказательство.

1. Пусть $f \in \mathbb{C}[x]$, $\deg f > 1$ и $z \in \mathbb{C}$ — корень f.
   - Тогда $f(x) \vdots x - z$, следовательно, f — приводим.
2. Пусть $f(x) = a_n x^n + \ldots + a_1 x + a_0 \in \mathbb{R}[x]$ — неприводимый многочлен, $n = \deg f > 1$ и $z \in \mathbb{C}$ — корень f. Заметим, что $z \notin \mathbb{R}$, иначе многочлен $f(x)$ приводим над $\mathbb{R}$.
   - Тогда $f(\overline{z}) = a_n \overline{z}^n + \ldots + a_1 \overline{z} + a_0 = \overline{a_n z^n + \ldots + a_1 z + a_0} = \overline{f(z)} = 0$.
   - Следовательно, $f(x) \vdots x - z$ и $f(x) \vdots x - \overline{z}$.
   - Поскольку $(x - z, x - \overline{z}) = 1$, получаем, что $f(x) \vdots (x - z)(x - \overline{z})$.
   - Но $(x - z)(x - \overline{z}) = x^2 - (z + \overline{z})x + z\overline{z} \in \mathbb{R}[x]$.
   - Следовательно, $n = 2$ и $f(x) = a_2(x - z)(x - \overline{z})$ — квадратный трехчлен с корнями $z, \overline{z} \notin \mathbb{R}$. Тогда его дискриминант меньше нуля. □

2.4. Группы

Определение 46

Группой называется упорядоченная пара $(G, *)$,
где G — множество и $*$ — бинарная операция на G,
удовлетворяющая следующим свойствам:

1. $\forall a, b, c \in G \{a * (b * c) = (a * b) * c\}$;
2. $\exists e \in G \forall a \in G \{a * e = e * a = a\}$;
3. $\forall a \in G \exists a^{-1} \in G \{a * a^{-1} = a^{-1} * a = e\}$.

Определение 47

Группа G называется *коммутативной* (или *абелевой*), если $\forall a, b \in G \{a * b = b * a\}$.

Замечание

• Групповую операцию $*$ часто называют *умножением* и обозначают точкой, которую часто тоже опускают. То есть вместо $a * b$ обычно пишут $a \cdot b$ или просто ab. Это называется *мультипликативной формой записи*.

• Но бывает и *аддитивная форма записи*, когда групповую операцию обозначают значком $+$. Аддитивная форма записи используется в основном для абелевых групп.

Примеры групп (1/2)

1. Пусть $(A, +, \cdot)$ — произвольное кольцо.
   Тогда $(A, +)$ — группа (*аддитивная группа кольца*).
2. Пусть $(A, +, \cdot)$ — произвольное кольцо.
   Тогда $(A^*, \cdot)$ — группа (*мультипликативная группа кольца*).
3. Матричные группы. Пусть A — коммутативное кольцо с единицей, $n \in \mathbb{N}$. Тогда
   - $GL_n(A) \stackrel{\text{def}}{=} \{M \in M_n(A) \mid \text{матрица } M \text{ обратима}\}$ — *полная линейная группа*;
   - $SL_n(A) \stackrel{\text{def}}{=} \{M \in M_n(A) \mid \det M = 1\}$ — *специальная линейная группа*.

Замечание

- $GL_n(A) = M_n(A)^*$.
- Если K — поле, то матрица $M \in M_n(K)$ обратима,
если и только если $\det M \neq 0$.
  - То есть $GL_n(K) = \{M \in M_n(K) \mid \det M \neq 0\}$.
- Если A — коммутативное кольцо с единицей, то матрица $M \in M_n(K)$ обратима,
если и только если $\det M \in A^*$.
  - Например, $GL_n(\mathbb{Z}) = \{M \in M_n(\mathbb{Z}) \mid \det M = \pm 1\}$.

Примеры групп (2/2)

4. Группы перестановок. Пусть X — множество. *Перестановкой* (или *подстановкой*) на множестве X называется биекция $\sigma\colon X \to X$.

- Все перестановки на множестве X образуют группу относительно операции композиции. Эта группа обозначается $S(X)$.
- S_n — группа перестановок на множестве $[1..n]$ (*симметрическая группа*);
- A_n — группа всех *четных* (т. е. имеющих четное число инверсий) перестановок на множестве $[1..n]$ (*знакопеременная группа*).

5. Движения плоскости, т. е. биекции $f : \mathbb{R}^2 \to \mathbb{R}^2$, сохраняющие расстояния, образуют группу относительно операции композиции.

6. Группа автоморфизмов графа. Пусть $G = (V, E)$ — конечный неориентированный граф. Автоморфизмом графа G называется биекция $f : V \to V$, такая, что $\forall x, y \in V\,(xy \in E \leftrightarrow f(x)f(y) \in E)$.

- Автоморфизмы графа G образуют группу относительно операции композиции. Эта группа обозначается $\mathrm{Aut}(G)$.

Подгруппы (1/2)

Определение 48

• Пусть G — группа и $H \subset G$. Тогда H — *подгруппа* группы G, если H — группа с той же операцией, что и в G. Обозначение: $H < G$.

Лемма

Пусть $H < G$. Тогда

1. *единичные элементы в G и H совпадают;*
2. *для любого $x \in H$ элемент, обратный к x в группе H, совпадает с элементом, обратным к x в группе G.*

Доказательство.

1. Пусть e_G и e_H — единичные элементы в группах G и H соответственно.
   - Обозначим через e_H^{-1} элемент, обратный к e_H в группе G.
   - Тогда $e_H = e_H e_G = e_H(e_H e_H^{-1}) = (e_H e_H)e_H^{-1} = e_H e_H^{-1} = e_G$.
2. Пусть x_G^{-1} и x_H^{-1} — элементы, обратные к x в группах G и H соответственно.
   - Тогда $x_H^{-1} = x_H^{-1}e = x_H^{-1}(xx_G^{-1}) = (x_H^{-1}x)x_G^{-1} = ex_G^{-1} = x_G^{-1}$. □

Подгруппы (2/2)

Теорема 47

Подмножество H группы G является подгруппой, если и только если выполнены следующие три условия

1. $\forall x, y \in H\,(x \cdot y \in H)$;
2. $e \in H$;
3. $\forall x \in H\,(x^{-1} \in H)$.

Доказательство. "$\Rightarrow$": Условие 1. следует из определения подгруппы, а условия 2. и 3. — из предыдущей леммы.

"$\Leftarrow$": Поскольку выполнено условие 1., множество H *замкнуто* относительно операции $\cdot$. То есть $\cdot$ является бинарной операцией на множестве H.

- Эта операция ассоциативна, поскольку закон ассоциативности выполняется в G.
- Существование нейтрального и обратного элементов следует из условий 2. и 3.

□

Примеры подгрупп

1. $A_n < S_n$;
2. $SL_n(A) < GL_n(A)$.
3. Подгруппа, порожденная одним элементом. Пусть G — группа и $a \in G$.
   - Определим *степень* элемента a следующим образом.

     Пусть $n \in \mathbb{Z}$. Тогда $$a^n \stackrel{\text{def}}{=} \begin{cases} \underbrace{a \cdot \ldots \cdot a}_{n}, & n > 0 \\ e, & n = 0 \\ (a^{-1})^{-n}, & n < 0. \end{cases}$$
   - Тогда $\langle a \rangle \stackrel{\text{def}}{=} \{a^n \mid n \in \mathbb{Z}\}$ — подгруппа G.
     - Действительно, $a^m \cdot a^n = a^{m+n} \in \langle a \rangle$; $e = a^0 \in \langle a \rangle$; $(a^n)^{-1} \in \langle a \rangle$.

Определение 49

- $\langle a \rangle$ — подгруппа, *порожденная* элементом a.
- Количество элементов подгруппы $\langle a \rangle$ называется *порядком* элемента a.
  - А количество элементов группы называют *порядком* этой группы.

Циклические группы

Определение 50

Группа G называется *циклической*, если $\exists a \in G\,(G = \langle a \rangle)$.

Замечание

Циклическая группа всегда абелева.

• Для того, чтобы понять, как устроены циклические группы, введем понятие изоморфизма групп.

Определение 51

• Группы G и H называются *изоморфными*, если существует такая биекция $f : G \to H$, что $\quad \forall a, b \in G\,(f(ab) = f(a)f(b))$.

• Обозначение: $G \cong H$. Биекция f называется *изоморфизмом* групп G и H.

• Пусть $a \in G$. Возможны два варианта.

1° $\exists n \in \mathbb{N}\,(a^n = e)$. Тогда, пусть $m = \min\{n \in \mathbb{N} \mid a^n = e\}$.
В этом случае, $\langle a \rangle = \{e = a^0, a^1, a^2, \ldots, a^{m-1}\} \cong \mathbb{Z}/m\mathbb{Z}$.

2° $\forall n \in \mathbb{N}\,(a^n \neq e)$. Тогда $a^k \neq a^\ell$, при $k \neq \ell$. В этом случае, $\langle a \rangle \cong \mathbb{Z}$.

Смежные классы (1/2)

Определение 52

Пусть G — группа; $H < G$; $x \in G$. Тогда

- $xH \stackrel{\text{def}}{=} \{xh \mid h \in H\}$ — *левый смежный класс* подгруппы H, порожденный x.
- $Hx \stackrel{\text{def}}{=} \{hx \mid h \in H\}$ — *правый смежный класс* подгруппы H, порожденный x.

Теорема 48

Пусть $H < G$; $x, y \in G$. Тогда выполнено ровно одно из следующих двух утверждений: 1. $xH = yH$; 2. $xH \cap yH = \varnothing$.

Доказательство. Очевидно, что утверждения 1. и 2. одновременно выполняться не могут. Докажем, что выполнено хотя бы одно из них.

- Пусть утверждение 2. не выполнено. Тогда найдется элемент $a \in xH \cap yH$.
- То есть $a = xh_1 = yh_2$, где $h_1, h_2 \in H$. Следовательно, $x = y(h_2 h_1^{-1}) \in yH$.
- Но тогда для любого $h \in H$ имеем $xh = y(h_2 h_1^{-1} h) \in yH$. То есть $xH \subset yH$.
- Аналогично доказывается, что $yH \subset xH$. □

Смежные классы (2/2)

Следствие

Пусть $H < G$. Тогда G можно представить как объединение попарно непересекающихся левых смежных классов подгруппы H.

Замечание

- Другими словами, существует такое подмножество $X \subset G$, что

$$G = \bigcup_{x \in X} xH \quad \text{и} \quad \forall x, y \in X \, (x \neq y \to xH \cap yH = \varnothing).$$

- Утверждения, аналогичные теореме 48 и следствию из неё, верны также и для правых смежных классов.

Определение 53

Количество различных левых смежных классов подгруппы H группы G называется *индексом* подгруппы H и обозначается $(G : H)$.

Теорема Лагранжа

Теорема 49 (Лагранж)

Пусть G — конечная группа и $H < G$. Тогда $|G| = |H| \cdot (G : H)$.

Доказательство. Пусть $x_1H, x_2H, \ldots, x_nH$ — все различные левые смежные классы подгруппы H. Тогда $n = (G : H)$.

- Заметим, что $\forall i\,(|x_iH| = |H|)$.
  - Действительно, отображение $\varphi_i(y) = x_iy$ является биекцией из H в x_iH.
- Тогда $|G| = |x_1H| + \ldots + |x_nH| = n|H| = |H| \cdot (G : H)$. □

Следствие 1

Если $H < G$, то $|G| \vdots |H|$.

Следствие 2

Порядок группы G делится на порядок любого её элемента.

Доказательство. Пусть $a \in G$. Тогда в группе G есть циклическая подгруппа $\langle a \rangle < G$. Применяя к этой подгруппе следствие 1, получаем, что $|G| \vdots |\langle a \rangle|$. □

Теорема Лагранжа и тест Ферма

- В качестве примера, рассмотрим группу $G = (\mathbb{Z}/n\mathbb{Z})^*$, где n — составное число.
- Пусть $H = \{a \in G \mid a^{n-1} = 1\}$.
  - Очевидно, что $H < G$.
- По сути, элементы подгруппы H соответствуют тем основаниям, по которым тест Ферма для числа n дает ответ “простое”.
  - То есть то, что n — число Кармайкла, эквивалентно тому, что $H = G$.
  - В теореме 23 мы доказывали, что если n не число Кармайкла, то $|H| \leq \frac{n-1}{2}$.
  - Из теоремы Лагранжа следует более сильное утверждение: а именно, $|G| \vdots |H|$.

2.5. Конечные поля. Характеристика поля (1/3)

Определение 54

- *Характеристикой* поля K называется наименьшее натуральное число n, такое, что в поле K справедливо равенство

$$\underbrace{1+1+\ldots+1}_{n \text{ единиц}} = 0.$$

- Если такого n не существует, то характеристика поля K считается равной нулю.
- Характеристика поля K обозначается $\operatorname{char} K$.

Примеры

- $\operatorname{char} \mathbb{Q} = \operatorname{char} \mathbb{R} = \operatorname{char} \mathbb{C} = 0$;
- $\operatorname{char} \mathbb{F}_p = p$, где $p \in \mathbb{P}$.

Характеристика поля (2/3)

Лемма 1

Характеристика поля K может быть равна либо нулю, либо простому числу.

Доказательство. Пусть $\operatorname{char} K > 0$.

- Поскольку $0 \neq 1$ в поле K, $\operatorname{char} K > 1$.
- Пусть $\operatorname{char} K = ab$, где $a, b > 1$.
- Тогда $0 = \underbrace{1 + 1 + \ldots + 1}_{ab \text{ единиц}} = (\underbrace{1 + 1 + \ldots + 1}_{a \text{ единиц}})(\underbrace{1 + 1 + \ldots + 1}_{b \text{ единиц}})$.
- Поскольку в поле K нет делителей нуля, из этого следует, что либо $\underbrace{1 + 1 + \ldots + 1}_{a \text{ единиц}} = 0$, либо $\underbrace{1 + 1 + \ldots + 1}_{b \text{ единиц}} = 0$.
- Но это невозможно, поскольку $a < \operatorname{char} K$ и $b < \operatorname{char} K$. Противоречие. □

Характеристика поля (3/3)

Лемма 2

Если $|K| < \infty$, *то* $\operatorname{char} K \neq 0$.

Доказательство. Пусть $|K| = m$.

- Рассмотрим элементы 1, $1+1$, $\ldots$, $\underbrace{1+1+\ldots+1}_{m+1 \text{ единица}}$ поля K.
- Поскольку количество рассматриваемых элементов больше, чем $|K|$, среди них найдутся два равных.
- Пусть $\underbrace{1+1+\ldots+1}_{a \text{ единиц}} = \underbrace{1+1+\ldots+1}_{b \text{ единиц}}$, где $0 < a < b \leq m+1$.
- Тогда $\underbrace{1+1+\ldots+1}_{b-a \text{ единиц}} = 0$.

□

Изоморфизм полей

Определение 55

- Поля K и L называются *изоморфными*, если существует такая биекция $f : K \to L$, что
  - $\forall a, b \in K\,(f(a+b) = f(a) + f(b))$;
  - $\forall a, b \in K\,(f(ab) = f(a)f(b))$.
- Отображение f называется *изоморфизмом* полей K и L.
- То, что поля K и L изоморфны, обозначается так: $K \cong L$.

Утверждение

Пусть $f : K \to L$ *— изоморфизм. Тогда* 1) $f(0) = 0$; 2) $f(1) = 1$; 3) $\forall a \in K\,(f(-a) = -f(a))$; 4) $\forall a \in K \setminus \{0\}\,(f(a^{-1}) = (f(a))^{-1})$.

Доказательство. 1) $f(0) = f(0+0) = f(0) + f(0) \Longrightarrow 0 = f(0)$;
2) $f(1) = f(1 \cdot 1) = f(1) \cdot f(1)$ и $f(1) \neq f(0) = 0 \Longrightarrow 1 = f(1)$;
3) $f(a) + f(-a) = f(a + (-a)) = f(0) = 0 \Longrightarrow f(-a) = -f(a)$;
4) $f(a) \cdot f(a^{-1}) = f(a \cdot a^{-1}) = f(1) = 1 \Longrightarrow f(a^{-1}) = (f(a))^{-1}$. □

$\mathbb{Z}/p\mathbb{Z}$ как подполе поля характеристики p (1/2)

Определение 56

- Поле F называется *подполем* поля K, если $F \subset K$ и в полях F и K действуют одинаковые операции сложения и умножения.
- Поле K в этом случае называется *надполем* поля F.

Примеры

- $\mathbb{Q}$ — подполе $\mathbb{R}$;
- $\mathbb{R}$ — подполе $\mathbb{C}$.

Лемма 3

Пусть $\operatorname{char} K = p \in \mathbb{P}$. *Тогда в поле* K *есть подполе* F, *изоморфное* $\mathbb{Z}/p\mathbb{Z}$.

Доказательство. Пусть $F \stackrel{\text{def}}{=} \{\underbrace{1+1+\ldots+1}_{a \text{ единиц}} \mid 0 \le a < p\}$.

- Заметим, что равенство $\underbrace{1+1+\ldots+1}_{a \text{ единиц}} = \underbrace{1+1+\ldots+1}_{b \text{ единиц}}$ в K выполнено тогда и только тогда, когда $a \equiv b \pmod{p}$.

$\mathbb{Z}/p\mathbb{Z}$ как подполе поля характеристики p (2/2)

- Тогда сумма и произведение любых двух элементов F принадлежат F.
  - Более того, количества единиц при этом соответственно складываются или умножаются по модулю p.
- Таким образом, F — подполе K и отображение $f\colon \mathbb{Z}/p\mathbb{Z} \to F$, задаваемое формулой

$$f(\overline{a}) \stackrel{\text{def}}{=} \underbrace{1+1+\ldots+1}_{a \text{ единиц}},$$

(где $\overline{a}$ — класс вычетов по модулю p, содержащий a) является изоморфизмом полей $\mathbb{Z}/p\mathbb{Z}$ и F. □

Подполе и векторное пространство

Лемма 4

Пусть F — подполе K. Тогда K является векторным пространством над полем F.

Доказательство. В поле K определены операции сложения и умножения.

- В частности, любой элемент K можно умножить на любой элемент F.
- Все необходимые свойства следуют из свойств операций в поле. □

Примеры

- $\mathbb{C}$ — векторное пространство над $\mathbb{R}$ размерности 2;
- $\mathbb{R}$ — векторное пространство над $\mathbb{Q}$ бесконечной размерности.

Замечание

Можно доказать, что любой базис $\mathbb{R}$ над $\mathbb{Q}$ имеет мощность континуума.

Конечные поля

Теорема 50

Пусть K — конечное поле. Тогда $|K| = p^n$, где $p \in \mathbb{P}, n \in \mathbb{N}$.

Доказательство. Пусть $\operatorname{char} K = p$.

- Из лемм 1 и 2 следует, что $p \in \mathbb{P}$.
- По лемме 3 в K есть подполе F, где $F \cong \mathbb{Z}/p\mathbb{Z}$.
- Тогда, по лемме 4, K — векторное пространство над F.
- Пусть $n = \dim_F K$ — размерность K, как векторного пространства над F;
  - $x_1, \ldots, x_n$ — базис K, как векторного пространства над F.
- Тогда любой элемент поля K единственным образом представляется в виде $\lambda_1 x_1 + \ldots + \lambda_n x_n$, где $\lambda_1, \ldots, \lambda_n \in F$.
- Каждый из коэффициентов $\lambda_1, \ldots, \lambda_n$ можно выбрать p способами.
  - Следовательно, есть ровно p^n различных последовательностей $(\lambda_1, \ldots, \lambda_n)$.
- Таким образом, $|K| = p^n$. □

Существование конечных полей

Теорема 51

Пусть $p \in \mathbb{P}$, $n \in \mathbb{N}$, $q = p^n$. Тогда существует поле из q элементов.
Более того, такое поле единственно с точностью до изоморфизма. (б/д)

Замечание

- Поле из q элементов обозначается $\mathbb{F}_q$ или $GF(q)$.
- Конечные поля называют также *полями Галуа* — в честь Эвариста Галуа
  - Évariste Galois, 1811-1832.
- Здесь и всюду далее, когда речь будет идти о конечных полях, мы будем использовать обозначения $q = p^n$, где $p \in \mathbb{P}, n \in \mathbb{N}$.

Конечные поля и многочлены

Теорема 52

Любой элемент поля $\mathbb{F}_q$ является корнем многочлена $x^q - x$.

Доказательство. Очевидно, что $0^q - 0 = 0$.

- Поэтому далее мы будем рассматривать только ненулевые элементы поля $\mathbb{F}_q$.

- Рассмотрим $\mathbb{F}_q^*$ — *мультипликативную группу* поля $\mathbb{F}_q$.
  - Напомним, что $\mathbb{F}_q^* = \mathbb{F}_q \setminus \{0\}$ состоит из всех обратимых элементов поля $\mathbb{F}_q$;
  - $\mathbb{F}_q^*$ является группой по умножению.
- По следствию из теоремы Лагранжа, порядок любого элемента группы $\mathbb{F}_q^*$ является делителем порядка этой группы.
- Следовательно, для любого $a \in \mathbb{F}_q^*$ верно равенство $a^{q-1} = 1$.
- Но тогда $a^q - a = a(a^{q-1} - 1) = 0$. □

Конечное поле как поле разложения многочлена

Следствие

• *Многочлен $x^q - x$ раскладывается над $\mathbb{F}_q$ в произведение q различных линейных множителей.*

• *А именно, $x^q - x = \prod\limits_{a \in \mathbb{F}_q} (x - a)$.*

• *В частности, у многочлена $x^q - x$ над полем $\mathbb{F}_q$ нет кратных корней.*

Определение 57

Пусть $f \in K[x]$. *Полем разложения* многочлена f называется минимальное по включению надполе L поля K, над которым многочлен f раскладывается на линейные множители.

Замечание

• Заметим, что $(x^q - x)' = qx^{q-1} - 1 = -1$ в $\mathbb{F}_q[x]$. Из этого также следует, что у многочлена $x^q - x$ нет кратных корней.

• Фактически это означает, что $\mathbb{F}_q$ — поле разложения многочлена $x^q - x$.

Примитивные элементы конечного поля (1/3)

Определение 58

Элемент $\alpha \in \mathbb{F}_q^*$ называется *примитивным элементом* поля $\mathbb{F}_q$, если α имеет порядок $q-1$ в группе $\mathbb{F}_q^*$.

Теорема 53

Пусть $d \mid q-1$. Тогда в $\mathbb{F}_q^*$ ровно $\varphi(d)$ элементов порядка d.

Лемма

$\sum\limits_{d \mid n} \varphi(d) = n.$

Доказательство леммы. Рассмотрим дроби $\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}$.

- Приведем каждую из дробей к несократимому виду.
  - Получатся дроби, знаменатели которых — делители n.
- Заметим, что для каждого $d \mid n$ найдется ровно $\varphi(d)$ дробей со знаменателем d.
- Следовательно, всего дробей $\sum\limits_{d \mid n} \varphi(d)$.
- Но, с другой стороны, их n. □

Примитивные элементы конечного поля (2/3)

Доказательство теоремы. Индукция по d.

База: $d = 1$. Единица — единственный элемент $\mathbb{F}_q^*$ порядка 1.

Переход: от всех $s < d$ к d. Пусть $q - 1 = dm$. Рассмотрим многочлен

$$x^{q-1} - 1 = (x^d - 1)(x^{d(m-1)} + x^{d(m-2)} + \ldots + x^d + 1).$$

- По доказанному выше, многочлен $x^{q-1} - 1$ имеет ровно $q - 1$ различных корней.
  - При этом, у $x^d - 1$ не более d корней,
  - а у $x^{d(m-1)} + \ldots + x^d + 1$ — не более $d(m-1)$ корней.
- Следовательно, у $x^d - 1$ ровно d корней.
- Корни многочлена $x^d - 1$ — это элементы $\mathbb{F}_q^*$, порядок которых делит d.
- Посчитаем, сколько среди них имеют порядок меньше d.
- Количество интересующих нас элементов равно

$$\sum_{s|d,\, s<d} \varphi(s) = \sum_{s|d} \varphi(s) - \varphi(d) = d - \varphi(d).$$

- Следовательно, элементов порядка d ровно $\varphi(d)$. □

Примитивные элементы конечного поля (3/3)

Следствие

В $\mathbb{F}_q^*$ есть ровно $\varphi(q-1)$ примитивных элементов.

Замечание

- Это означает, что $\mathbb{F}_q^*$ — циклическая группа. Все элементы $\mathbb{F}_q^*$ являются степенями примитивного элемента. То есть $\mathbb{F}_q^* = \{1, \alpha, \alpha^2, \ldots, \alpha^{q-2}\}$, где $\alpha \in \mathbb{F}_q$ — примитивный элемент.
- Примитивный элемент поля $\mathbb{F}_p$, где $p \in \mathbb{P}$, называют также *первообразным корнем* по модулю p.

Автоморфизм Фробениуса (1/2)

Лемма

В поле $\mathbb{F}_q$ *выполнены равенства*
1) $(x+y)^p = x^p + y^p$; 2) $(xy)^p = x^p y^p$.

Доказательство. 1) $(x+y)^p = \sum\limits_{k=0}^{p} C_p^k x^{p-k} y^k = x^p + y^p$,
поскольку при $0 < k < p$ верно, что $C_p^k \vdots p$, то есть в $\mathbb{F}_q$
коэффициенты при таких одночленах обращаются в нуль.
2) Тривиально. □

Определение 59

Автоморфизмом поля K называются такая биекция $f : K \to K$, что

- $\forall a, b \in K\, (f(a+b) = f(a) + f(b))$;
- $\forall a, b \in K\, (f(ab) = f(a)f(b))$.

Замечание

То есть автоморфизм — это изоморфизм поля с собой.

Автоморфизм Фробениуса (2/2)

Теорема 54

Отображение $\varphi\colon \mathbb{F}_q \to \mathbb{F}_q$, *задаваемое формулой* $\varphi(x) \stackrel{\text{def}}{=} x^p$, *является автоморфизмом поля* $\mathbb{F}_q$.

Доказательство. То, что при всех $a, b \in \mathbb{F}_p$ выполнены равенства $\varphi(a+b) = \varphi(a) + \varphi(b)$ и $\varphi(ab) = \varphi(a)\varphi(b)$, следует из доказанной выше леммы.

- Докажем, что φ — биекция.
- Поскольку φ — отображение из конечного множества в себя, достаточно доказать, что φ — сюръекция.
- Пусть $y \in \mathbb{F}_q$. Тогда $\varphi(y^{p^{n-1}}) = (y^{p^{n-1}})^p = y^{p^n} = y^q = y$. □

Определение 60

Построенное выше отображение называется *автоморфизмом Фробениуса*.

- Ferdinand Georg Frobenius (1849–1917).

Построение поля из p^n элементов (1/3)

- Пусть $p \in \mathbb{P}$, $n \in \mathbb{N}$, $q = p^n$ (*эти обозначения будут использоваться на протяжении всей лекции*).
- Рассмотрим неприводимый многочлен $f \in \mathbb{F}_p[x]$ степени n.
- В кольце многочленов $\mathbb{F}_p[x]$ рассмотрим главный идеал (f).

Теорема 55

Факторкольцо $\mathbb{F}_p[x]/_{(f)}$ *является полем, состоящим из* q *элементов.*

Доказательство. Докажем, что (f) — максимальный идеал.

- Напомним, что все идеалы в $\mathbb{F}_p[x]$ главные.
- Пусть $(f) \subset (g)$. Тогда $f \vdots g$.
- Поскольку f неприводим, далее возможны два случая: $\deg g = \deg f$ и $\deg g = 0$.

1° Пусть $\deg g = \deg f$. Тогда $g = cf$, где $c \in \mathbb{F}_p \setminus \{0\}$. Следовательно, $(g) = (f)$.

2° Пусть $\deg g = 0$. Тогда $g \in \mathbb{F}_p \setminus \{0\}$. Следовательно, $(g) = (1) = \mathbb{F}_p[x]$.

Построение поля из p^n элементов (2/3)

- Итак, (f) — максимальный идеал в $\mathbb{F}_p[x]$. Следовательно, $\mathbb{F}_p[x]/_{(f)}$ — поле.
- Элементы полученного поля — классы вычетов по модулю многочлена f.
- В каждом таком классе есть ровно один многочлен, степень которого меньше n.
- Таким образом, мы получили биекцию между элементами построенного поля и многочленами вида $a_0 + a_1 x + \ldots + a_{n-1} x^{n-1}$, где $a_0, a_1, \ldots, a_{n-1} \in \mathbb{F}_p$.
- Каждый из коэффициентов a_i можно выбрать p способами.
  - Следовательно, всего имеем $q = p^n$ таких многочленов.

□

Замечание

Доказанная выше теорема не является доказательством существования поля из p^n элементов. Дело в том, что мы не доказывали существование неприводимого над $\mathbb{F}_p$ многочлена нужной нам степени.

- Тем не менее, это верно: можно доказать, что для любых $n \in \mathbb{N}$ и $p \in \mathbb{P}$ существует неприводимый многочлен степени n над полем $\mathbb{F}_p$.

Построение поля из p^n элементов (3/3)

- Итак, элементы поля $\mathbb{F}_q$ можно представлять как многочлены с целыми коэффициентами вида $a_0 + a_1 x + \ldots + a_{n-1} x^{n-1}$, где $a_0, a_1, \ldots, a_{n-1} \in [0..p-1]$.
- При этом, коэффициенты рассматриваются как вычеты по модулю p, а сами многочлены — по модулю f.
- То есть при сложении и умножении многочленов все операции над их коэффициентами производятся по модулю p.
- Если при перемножении многочленов получился многочлен, степень которого больше либо равна n, то его нужно поделить с остатком на f.
- Пусть g — многочлен, соответствующий ненулевому элементу поля $\mathbb{F}_q$. Как найти обратный ему элемент?
  - Заметим, что $(f, g) = 1$.
  - Находим линейное представление их НОД: $1 = f(x)a(x) + g(x)b(x)$.
  - Тогда $g(x)b(x) \equiv 1 \pmod{f(x)}$, следовательно, b — элемент $\mathbb{F}_q$, обратный к g.

Неприводимые многочлены степени n над $\mathbb{F}_p$

Лемма

Пусть $f \in \mathbb{F}_p[x]$ — неприводимый многочлен степени n. Тогда $x^q - x \vdots f(x)$.

Доказательство.

- По доказанному выше, $\mathbb{F}_p[x]/_{(f)}$ — поле из q элементов.
- Все элементы этого поля — корни многочлена $t^q - t$.
- Подставим в этот многочлен x (т. е. элемент поля $\mathbb{F}_p[x]/_{(f)}$, который соответствует многочлену x).
- Получаем, что в поле $\mathbb{F}_p[x]/_{(f)}$ выполнено равенство $x^q - x = 0$, а это и означает, что в кольце $\mathbb{F}_p[x]$ имеет место делимость $x^q - x \vdots f(x)$. □

Замечание

- Разложим многочлен $x^q - x$ в произведение унитарных неприводимых множителей над $\mathbb{F}_p$. В это разложение входят все унитарные неприводимые многочлены из $\mathbb{F}_p[x]$.
- Но в этом разложении есть и другие множители.

Минимальный многочлен (1/2)

Определение 61

Пусть $a \in \mathbb{F}_q$. *Минимальным многочленом* элемента a называется такой унитарный многочлен $f_a \in \mathbb{F}_p[x]$, что $f_a(a) = 0$ и многочлен f_a имеет наименьшую степень среди всех таких многочленов.

Теорема 56

Для каждого элемента $a \in \mathbb{F}_q$ существует единственный минимальный многочлен.

Доказательство. "$\exists$" Рассмотрим множество

$$I_a \stackrel{\text{def}}{=} \{f \in \mathbb{F}_p[x] \mid f(a) = 0\}.$$

- Заметим, что $x^q - x \in I_a$. Следовательно, $I_a \neq \{0\}$.
- Пусть $f \in I_a$ — ненулевой многочлен наименьшей степени.
- Разделим f на его старший коэффициент.
  - Получим унитарный многочлен $f_a \in I_a$, такой, что $\deg f_a = \deg f$.
- Тогда f_a — минимальный многочлен элемента a.

Минимальный многочлен (2/2)

"!" пусть f_1 и f_2 — два минимальных многочлена элемента a.

- Тогда $\deg f_1 = \deg f_2 > \deg(f_1 - f_2)$ и $(f_1 - f_2)(a) = 0$.
- Если $f_1 - f_2$ — ненулевой многочлен, сократим его на старший коэффициент и получим унитарный многочлен с корнем a, степень которого меньше, чем $\deg f_1$. Противоречие с минимальностью f_1.
- Таким образом, $f_1 - f_2 = 0$ и тогда $f_1 = f_2$. □

Замечание

- Легко видеть, что рассмотренное в доказательстве существования множество $I_a = \{f \in \mathbb{F}_p[x] \mid f(a) = 0\}$ — идеал в $\mathbb{F}_p[x]$.
- Тогда f_a — порождающий элемент этого идеала. Т. е. $I_a = (f_a)$.
  - Действительно, $\mathbb{F}_p[x]$ — КГИ, следовательно, $I_a = (g)$.
  - Тогда $f_a \vdots g$ и $\deg f_a \leq \deg g$, откуда $f_a = cg$, где $\deg c = 0$.
  - Таким образом, $(f_a) = (g) = I_a$.

Свойства минимального многочлена (1/2)

Теорема 57

Пусть f_a — минимальный многочлен элемента $a \in \mathbb{F}_q$. Тогда

1. *f_a — неприводим;*
2. *если $g(a) = 0$, где $g \in \mathbb{F}_p[x]$, то $g \vdots f_a$;*
3. $\deg f_a \leq n$;
4. *элементы a и a^p имеют один и тот же минимальный многочлен.*

Доказательство.

1. Пусть $f_a = gh$, где $\deg g < \deg f_a$ и $\deg h < \deg f_a$.
   - Тогда $0 = f_a(a) = g(a)h(a)$.
   - Следовательно, либо $g(a) = 0$, либо $h(a) = 0$.
   - В обоих случаях получаем противоречие с минимальностью f_a.
2. Если $g(a) = 0$, то $g \in I_a = (f_a)$. Следовательно, $g \vdots f_a$.

Свойства минимального многочлена (2/2)

3. Рассмотрим $\mathbb{F}_q$ как векторное пространство над $\mathbb{F}_p$.
   - Пусть $\deg f_a = m$.
   - Докажем, что элементы $1, a, a^2, \ldots, a^{m-1}$ — ЛНЗ.
     - Действительно, в противном случае существуют такие $c_0, \ldots, c_{m-1} \in \mathbb{F}_p$, что c_i не все нули и $c_0 + c_1 a + \ldots + c_{m-1} a^{m-1} = 0$.
     - Тогда $c(x) \stackrel{\text{def}}{=} c_0 + c_1 x + \ldots + c_{m-1} x^{m-1} \in \mathbb{F}_p[x]$ — ненулевой многочлен, такой, что $\deg c < m$ и $c(a) = 0$. Противоречие с минимальностью f_a.
   - Тогда $m \leq \dim_{\mathbb{F}_p} \mathbb{F}_q = n$.
4. Пусть $f_a(x) = x^m + b_{m-1} x^{m-1} + \ldots + b_0$. Тогда
$$\begin{aligned}(f_a(x))^p &= (x^m + b_{m-1} x^{m-1} + \ldots + b_0)^p = (x^m)^p + (b_{m-1} x^{m-1})^p + \ldots + (b_0)^p = \\ &= x^{mp} + (b_{m-1})^p x^{(m-1)p} + \ldots + (b_0)^p = \\ &= (x^p)^m + b_{m-1}(x^p)^{m-1} + \ldots + b_0 = f_a(x^p).\end{aligned}$$
   - Следовательно, $f_a(a^p) = (f_a(a))^p = 0$.
   - Поскольку f_a неприводим, это означает, что f_a — минимальный многочлен для a^p. □

Подполе конечного поля (1/2)

- Пусть $a \in \mathbb{F}_q$ и $m = \deg f_a$.
- Рассмотрим множество

$$\mathbb{F}_p(a) \stackrel{\text{def}}{=} \{c_0 + c_1 a + \ldots + c_{m-1} a^{m-1} \mid c_0, \ldots, c_{m-1} \in \mathbb{F}_p\}.$$

Теорема 58

$\mathbb{F}_p(a)$ — *подполе поля* $\mathbb{F}_q$, *изоморфное* $\mathbb{F}_p[x]/_{(f_a)}$; $|\mathbb{F}_p(a)| = p^m$.

Доказательство. Мы уже доказывали, что элементы $1, a, a^2, \ldots, a^{m-1}$ — ЛНЗ над $\mathbb{F}_p$.

- Следовательно, элементы вида $c_0 + c_1 a + \ldots + c_{m-1} a^{m-1}$ для разных наборов $c_0, \ldots, c_{m-1}$ различны.
- Элементы поля $\mathbb{F}_p[x]/_{(f_a)}$ представляются в виде многочленов $c_0 + c_1 x + \ldots + c_{m-1} x^{m-1} \in \mathbb{F}_p[x]$.
- Тогда отображение $\mathcal{F}\colon \mathbb{F}_p[x]/_{(f_a)} \to \mathbb{F}_p(a)$, задаваемое формулой $\mathcal{F}(c_0 + c_1 x + \ldots + c_{m-1} x^{m-1}) \stackrel{\text{def}}{=} c_0 + c_1 a + \ldots + c_{m-1} a^{m-1}$ — биекция.

Подполе конечного поля (2/2)

- Докажем, что отображение $\mathcal{F}$ переводит сумму в сумму и произведение в произведение.
  - Для суммы это очевидно, поскольку и в $\mathbb{F}_p[x]/_{(f_a)}$, и в $\mathbb{F}_p(a)$ сложение происходит покоэффициентно и коэффициенты принадлежат $\mathbb{F}_p$.
  - Для произведения тоже очевидно, поскольку элементы $\mathbb{F}_p[x]/_{(f_a)}$ представляются в виде многочленов, рассматриваемых по модулю f_a. А в $\mathbb{F}_q$ верно, что $g(a) = h(a) \Leftrightarrow g \equiv h \pmod{f_a}$.
- Следовательно, $\mathbb{F}_p(a)$ — подполе поля $\mathbb{F}_q$, изоморфное $\mathbb{F}_p[x]/_{(f_a)}$.
- Поскольку $\deg f_a = m$, имеем $|\mathbb{F}_p(a)| = |\mathbb{F}_p[x]/_{(f_a)}| = p^m$. □

Определение 62

Говорят, что подполе $\mathbb{F}(a)$ получено *присоединением* элемента a к полю $\mathbb{F}_p$.

Присоединение элемента к $\mathbb{F}_p$

Следствие 1

$\deg f_a \mid n$.

Доказательство. Поскольку $\mathbb{F}_p(a) \subset \mathbb{F}_q$, поле $\mathbb{F}_q$ является векторным пространством над $\mathbb{F}_p(a)$.

- Пусть $d = \dim_{\mathbb{F}_p(a)} \mathbb{F}_q$ и $m = \deg f_a$.
- Тогда $p^n = |\mathbb{F}_q| = |\mathbb{F}_p(a)|^d = p^{md}$, откуда $n = md$. □

Следствие 2

Если α — примитивный элемент $\mathbb{F}_q$, то $\deg f_\alpha = n$ *и* $\mathbb{F}_q \cong \mathbb{F}_p[x]/_{(f_\alpha)}$.

Доказательство. Докажем, что $\mathbb{F}_p(\alpha) = \mathbb{F}_q$.

- Действительно, все ненулевые элементы $\mathbb{F}_q$ представляются как степени α, а все степени α принадлежат $\mathbb{F}_p(\alpha)$. □

Следствие 3

Любые два поля из q элементов изоморфны. В частности, если $f, g \in \mathbb{F}_p[x]$ неприводимы и $\deg f = \deg g$, *то* $\mathbb{F}_p[x]/_{(f)} \cong \mathbb{F}_p[x]/_{(g)}$.

Примитивные многочлены

Определение 63

Многочлен $f \in \mathbb{F}_p[x]$ называется *примитивным*, если он является минимальным многочленом примитивного элемента.

Замечание

- Другими словами, это означает, что элемент x в поле $\mathbb{F}_p[x]/_{(f)}$ примитивен.
- Мы доказали, что минимальный многочлен примитивного элемента поля $\mathbb{F}_q$ имеет степень n. Однако обратное неверно: не всякий неприводимый многочлен степени n примитивен.

Пример

Рассмотрим многочлен $f(x) = x^4 + x^3 + x^2 + x + 1 \in \mathbb{F}_2[x]$. Можно доказать, что он неприводим (мы докажем это на практике). Однако, если $a \in \mathbb{F}_{16}$ — корень f, то $a^5 - 1 = (a-1)(a^4 + a^3 + a^2 + a + 1) = 0$. Следовательно, $a^5 = 1$ и элемент a — не примитивен.

Разложение на множители многочлена $x^q - x$

Теорема 59

Разложение многочлена $x^q - x$ на неприводимые унитарные множители над полем $\mathbb{F}_p$ состоит в точности из всех неприводимых унитарных многочленов, степени которых являются делителями n.

Доказательство. Пусть $p(x)$ — один из множителей разложения.

- Многочлен $x^q - x$ раскладывается на линейные множители над $\mathbb{F}_q$.
- Следовательно, и $p(x)$ тоже раскладывается на линейные множители.
  - Тогда $p(x)$ имеет корень $a \in \mathbb{F}_q$.
- Поскольку $p(x)$ неприводим, он является минимальным многочленом для a.
  - Но тогда $\deg p \mid n$.
- Обратно, пусть $m \mid n$ и $p(x)$ — неприводимый многочлен степени m.
- Тогда $p^n - 1 \vdots p^m - 1$, следовательно, $x^{p^n-1} - 1 \vdots x^{p^m-1} - 1$.
- Таким образом, $x^{p^n} - x \vdots x^{p^m} - x \vdots p(x)$. □

Циклотомические классы (1/2)

- Пусть $a \in \mathbb{F}_q$. Как найти минимальный многочлен f_a?
- Для этого нужно найти его корни.
  - Мы знаем, что корнями f_a являются элементы a, a^p, a^{p^2}, ...
  - Можно доказать, что все корни f_a имеют такой вид.
- Пусть α — примитивный элемент $\mathbb{F}_q$ и $a = \alpha^s$.
  - Тогда корни f_a имеют вид α^s, α^{ps}, $\alpha^{p^2 s}$, ...
- Поскольку α — элемент порядка $q-1$, показатели степеней s, ps, $p^2 s$, ... нужно рассматривать по модулю $q-1$.
- Пусть $m_s = \min\{m \in \mathbb{N} \mid p^m s \equiv s \pmod{q-1}\}$.
  - Такие m существуют, поскольку $p^n s = qs \equiv s \pmod{q-1}$.

Определение 64

Множество $C_s = \{s, ps, p^2 s, \ldots, p^{m_s - 1} s\}$ называется *циклотомическим классом* вычета s.

Циклотомические классы (2/2)

- Элементы циклотомического класса C_s можно рассматривать как вычеты по модулю $q-1$, то есть как элементы кольца $\mathbb{Z}/(q-1)\mathbb{Z}$.
- Легко видеть. что все числа $s, ps, p^2s, \ldots, p^{m_s-1}s$ различны по модулю $q-1$.
- Также нетрудно понять, что для любых $s, t \in \mathbb{Z}/(q-1)\mathbb{Z}$ либо $C_s = C_t$, либо $C_s \cap C_t = \varnothing$.
  - То есть элементы $\mathbb{Z}/(q-1)\mathbb{Z}$ (а также числа от 0 до $q-2$) можно разбить на непересекающиеся циклотомические классы.

Теорема 60

Пусть $a = \alpha^s$, где α — примитивный элемент поля $\mathbb{F}_q$. Тогда

$$f_a(x) = \prod_{j \in C_s} (x - \alpha^j).$$

(б/д)

Глава 3.
Коды, исправляющие ошибки

- Дополнительные материалы:
  - Ф. И. Соловьева, *Введение в теорию кодирования*. НГУ, 2011.
  - А. Е. Ромащенко, А. Ю. Румянцев, А. Шень, *Заметки по теории кодирования*. М.: МЦНМО, 2011.

3.1. Введение в теорию кодирования

- Пусть Σ_1 и Σ_2 — два конечных алфавита. *Сообщение* — произвольное слово $u \in \Sigma_1^*$.
  - Мы хотим *закодировать* сообщение u в алфавите Σ_2, то есть поставить ему в соответствие слово $F(u) \in \Sigma_2^*$, которое будет передаваться по каналам связи.
  - Для этого нам нужно задать отображение $F : \Sigma_1^* \to \Sigma_2^*$, которое называется *кодирующим отображением* или просто *кодированием*.
  - Требования к отображению F зависят от того, какую задачу мы решаем. Основные задачи теории кодирования таковы:
    - шифрование данных: требуется, чтобы вычисление обратного отображения F^{-1} было значительно более трудоемким, чем вычисление F;
    - помехоустойчивое кодирование: требуется, чтобы исходное сообщение u можно было восстановить даже в том случае, если при передаче $F(u)$ произошли ошибки (при условии, что ошибок было не слишком много);
    - сжимающие отображения: требуется, чтобы длина кодированного сообщения была как можно меньше.
  - В большинстве случаев, важным требованием является возможность однозначного декодирования (то есть F должно быть инъекцией). Но это требуется не всегда. Например, сжатие с потерей качества не предполагает однозначного декодирования.

Схемы кодирования

- В этой главе мы будем обсуждать вопросы, связанные с помехоустойчивым кодированием.
- Мы будем рассматривать *блочное* или *равномерное* кодирование, при котором сообщение $u \in \Sigma_1^*$ разбивается на блоки длины k, каждый из которых будет закодирован словом длины n в алфавите Σ_2.
  - Для этого нам нужно задать инъекцию $c : \Sigma_1^k \to \Sigma_2^n$, которая будет называться *схемой кодирования*.
  - В первую очередь нас будет интересовать множество кодовых слов $\mathcal{C} \stackrel{\text{def}}{=} \mathrm{Im}(c) = \{x \in \Sigma_2^n \mid \exists u \in \Sigma_2^n\,(c(u) = x)\}$, которое мы будем называть просто *кодом*.
  - Как правило, мы будем считать, что $\Sigma_1 = \Sigma_2 = \Sigma$ и $k < n$.
- Пусть $x = x_1 \dots x_n \in \Sigma^n$. Ошибки при передаче слова x могут быть трех типов:
  - замещение разряда: вместо символа x_i приняли другой символ x_i';
  - выпадение разряда: символ x_i не был распознан;
  - вставка разряда: между x_i и x_{i+1} прочитали "лишний" символ y.
- Мы будем рассматривать только ошибки типа замещения.

Кодовое расстояние (1/3)

Определение 65

Пусть Σ — конечный алфавит, $n \in \mathbb{N}$ и $x = x_1 \dots x_n, y = y_1 \dots y_n \in \Sigma^n$.
Тогда *расстоянием Хэмминга* между словами x и y называется величина $d(x, y) \stackrel{\text{def}}{=} |\{i \in [1..n] \mid x_i \neq y_i\}|$.

Замечание

- Легко видеть, что (Σ^n, d) — метрическое пространство.
- Важными объектами для изучения являются шары в этом метрическом пространстве.

Пусть $x \in \Sigma^n$ и $r \in \mathbb{N}_0$. Тогда *шаром* с центром x и радиусом r называется множество $B_r(x) \stackrel{\text{def}}{=} \{y \in \Sigma^n \mid d(x, y) \leq r\}$. Очевидно, что $|B_r(x)| = \sum_{i=0}^{r} C_n^i (q-1)^i$, где $q = |\Sigma|$.

Определение 66

- Пусть $\mathcal{C} \subset \Sigma^n$ — произвольный код. Тогда *кодовым расстоянием* кода $\mathcal{C}$ называется величина $d(\mathcal{C}) \stackrel{\text{def}}{=} \min\{d(x, y) \mid x, y \in \mathcal{C}, x \neq y\}$.
- Аналогично, *кодовым расстоянием* схемы кодирования $c : \Sigma^k \to \Sigma^n$ называется величина $d(c) \stackrel{\text{def}}{=} d(\mathrm{Im}(c))$.

Кодовое расстояние (2/3)

Теорема 61

Пусть при передаче сообщения длины n возникает не более r ошибок типа замещения, а для кодирования сообщений используется схема c. Тогда

1. *схема кодирования c обеспечивает гарантированное обнаружение ошибки, если и только если $d(c) > r$;*
2. *схема кодирования c обеспечивает гарантированное исправление всех ошибок, если и только если $d(c) > 2r$.*

Доказательство. Заметим, что при передаче слова x, результат может оказаться любым словом из $B_r(x)$.
1. Для гарантированного обнаружения ошибки необходимо и достаточно, чтобы никакое кодовое слово не лежало в шаре радиуса r с центром в другом кодовом слове. Но это и означает, что $d(c) > r$.
2. Для гарантированного исправления всех ошибок необходимо и достаточно, чтобы шары радиуса r с центрами в кодовых словах не пересекались. Докажем, что это эквивалентно тому, что $d(c) > 2r$.

Кодовое расстояние (3/3)

"$\Leftarrow$" Пусть $z \in B_r(x) \cap B_r(y)$.

- Тогда $d(x, y) \leq d(x, z) + d(z, y) \leq d + d = 2d$.

"$\Rightarrow$" Пусть $d(x, y) \leq 2r$.

- Рассмотрим те разряды, в которых слово x отличается от слова y.
- Пусть таких разрядов $d \leq 2r$.
- Заменим в слове x какие-нибудь $[d/2]$ из рассматриваемых разрядов на соответствующие разряды слова y.
- Получим слово z, такое, что $d(x, z) \leq r$ и $d(z, y) \leq r$.
- То есть $z \in B_r(x) \cap B_r(y)$. □

Замечание

Простейшим примером схемы кодирования с кодовым расстоянием d является схема, при которой каждый символ повторяется d раз.

- То есть слово $u = u_1 u_2 \dots u_k$ кодируется как $c(u) = \underbrace{u_1 \dots u_1}_{d} \underbrace{u_2 \dots u_2}_{d} \dots \underbrace{u_k \dots u_k}_{d}$.
- Но, разумеется, такая схема очень неэкономна.

3.2. Линейные коды

- Пусть q — степень простого числа p и $\Sigma = \mathbb{F}_q$.
  - Тогда множество $\mathbb{F}_q^n$ всех слов длины n в этом алфавите является векторным пространством размерности n над $\mathbb{F}_q$.

Определение 67

- Линейное подпространство $\mathcal{C}$ пространства $\mathbb{F}_q^n$ называется *линейным q-значным кодом длины n*.
- В случае $q = 2$ линейный код $\mathcal{C}$ называется *двоичным*.
- Линейный код $\mathcal{C}$ имеет следующие параметры:
  - n — *длина* кода (количество символов в каждом кодовом слове);
  - k — *размерность* кода (размерность $\mathcal{C}$ как векторного пространства над $\mathbb{F}_q$);
  - d — *кодовое расстояние*.
- Совокупности всех параметров линейного кода $\mathcal{C}$ мы будем обозначать $[n, k, d]$.
  - Код $\mathcal{C}$ в этом случае мы будем также называть *$[n, k, d]$-кодом*.
  - Иногда мы будем опускать параметр d и говорить об $[n, k]$-кодах.

Линейные коды и схемы кодирования

- Пусть дан линейный q-значный $[n, k, d]$-код $\mathcal{C}$.
  - Тогда кодовые слова представляются как векторы вида $x = (x_1, x_2, \ldots, x_n)$, где $x_i \in \mathbb{F}_q$.
- Поскольку $\dim_{\mathbb{F}_q} \mathcal{C} = k$, очевидно, что $|\mathcal{C}| = q^k$.
  - Это означает, что при помощи кода $\mathcal{C}$ можно кодировать сообщения длины k.
  - В таком случае, исходные сообщения также можно представлять как векторы вида $u = (u_1, u_2, \ldots, u_k)$, где $u_i \in \mathbb{F}_q$.
  - Схемой кодирования тогда будет линейное отображение $c : \mathbb{F}_q^k \to \mathbb{F}_q^n$.
  - Нам нужно, чтобы отображение c было инъекцией. В случае линейного отображения это равносильно тому, что $c(u) = 0$ только в случае $u = 0$.

Определение 68

- Линейные коды $\mathcal{C}_1$ и $\mathcal{C}_2$ *эквивалентны*, если они отличаются перестановкой координат.
- То есть $\exists \sigma \in S_n \{(x_1, x_2, \ldots, x_n) \in \mathcal{C}_1 \leftrightarrow (x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)}) \in \mathcal{C}_2\}$.

Замечание

У эквивалентных кодов все кодовые параметры одинаковы.

Кодовое расстояние линейного кода

Определение 69

- Пусть $x = (x_1, x_2, \ldots, x_n) \in \mathbb{F}_q^n$. *Весом Хэмминга* вектора x называется число его ненулевых координат. Вес Хэмминга обозначается $w(x)$.
- То есть $w(x) = |\{i \in [1..n] \mid x_i \neq 0\}|$.

Лемма

Пусть $x, y \in \mathbb{F}_q^n$. *Тогда* $d(x, y) = w(x - y)$.

Теорема 62

Пусть $\mathcal{C}$ *— линейный* q*-значный код с кодовым расстоянием* d.
Тогда $d = \min\{w(x) \mid x \in \mathcal{C} \setminus \{0\}\}$.

Доказательство. Пусть $\min\{w(x) \mid x \in \mathcal{C} \setminus \{0\} = d'$. Нужно доказать, что $d = d'$.

$d \geq d'$: Рассмотрим векторы $x, y \in \mathbb{F}_q^n$, такой, что $d(x, y) = d$.

- Тогда $d = d(x, y) = w(x - y) \geq d'$.

$d \leq d'$: Рассмотрим вектор $s \in \mathbb{F}_q^n$, такой, что $w(s) = d'$.

- Тогда $d \leq d(s, 0) = w(s - 0) = d'$. □

Порождающая матрица линейного кода

- Пусть $\mathcal{C}$ — линейный q-значный $[n, k]$-код.

Определение 70

- *Порождающей матрицей* кода $\mathcal{C}$ называется матрица G размером $k \times n$ (k строк и n столбцов), строки которой образуют базис подпространства $\mathcal{C}$.

Замечание

- Из определения очевидно, что у любого линейного кода есть порождающая матрица и её строки ЛНЗ (т. е. $\operatorname{rank} G = k$).
  - В то же время очевидно, что порождающая матрица не единственна.
- Порождающая матрица G задает схему кодирования.
  - Действительно, пусть $g_1, g_2, \ldots, g_k$ — строки G и $u \in \mathbb{F}_q^k$.
  - Тогда отображение c можно определить следующим образом: $c(u) \stackrel{\text{def}}{=} \sum_{i=1}^{k} g_k u_k$.
  - Это же отображение задается формулами $c(u) = uG$ или $c(u)^T = G^T u^T$.
  - Также легко видеть, что любая схема кодирования должна переводить стандартный базис пространства $\mathbb{F}_q^k$ в некоторый базис подпространства $\mathcal{C}$. Следовательно, любая схема кодирования представляется в описанном выше виде для некоторой порождающей матрицы кода $\mathcal{C}$.

Проверочная матрица линейного кода

Определение 71

• *Проверочной матрицей* кода $\mathcal{C}$ называется матрица H размером $(n-k)\times n$, удовлетворяющая следующему условию: $\forall x \in \mathbb{F}_q^n \, (x \in \mathcal{C} \leftrightarrow Hx^T = 0)$.

Замечание

В отличии от порождающей матрицы, существование проверочной матрицы не является очевидным. Для его доказательства нам нужно будет ввести несколько дополнительных определений.

Определение 72

Пусть $x = (x_1, x_2, \ldots, x_n), y = (y_1, y_2, \ldots, y_n) \in \mathbb{F}_q^n$. Тогда *скалярным произведением* векторов x и y будем называть величину $\langle x, y \rangle \stackrel{\text{def}}{=} \sum_{i=1}^{n} x_i y_i$.

Замечание

Скалярное произведение не всегда определяется по такой формуле.
Например, скалярное произведение векторов над $\mathbb{C}$ определяется иначе.
Но в данном случае нам удобно определить скалярное произведение именно так.

Ортогональное дополнение (1/2)

Определение 73

- Векторы $x, y \in \mathbb{F}_q^n$ *ортогональны*, если $\langle x, y \rangle = 0$.
- Пусть $\mathcal{C}$ — линейное подпространство $\mathbb{F}_q^n$. Тогда *ортогональным дополнением* к $\mathcal{C}$ называется множество $\mathcal{C}^\perp \stackrel{\text{def}}{=} \{y \in \mathbb{F}_q^n \mid \forall x \in \mathcal{C}\,(\langle x, y \rangle = 0)\}$.

Лемма 1

$\mathcal{C}^\perp$ — линейное подпространство $\mathbb{F}_q^n$. Его размерность равна $n - k$, где $k = \dim \mathcal{C}$.

Доказательство. Пусть $g_1, g_2, \ldots, g_k$ — базис $\mathcal{C}$.

- Рассмотрим матрицу G, строками которой являются векторы $g_1, g_2, \ldots, g_k$ (это порождающая матрица для $\mathcal{C}$). Её элементы будем обозначать g_{ij}.
- Заметим, что $y \in \mathcal{C}^\perp$, если и только если $\langle g_1, y \rangle = \langle g_2, y \rangle = \ldots = \langle g_k, y \rangle = 0$.
  - Действительно, если $\langle g_1, y \rangle = \langle g_2, y \rangle = \ldots = \langle g_k, y \rangle = 0$ и $\lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F}_q^n$, то $\left\langle \sum_{i=1}^k \lambda_i g_i, y \right\rangle = \sum_{i=1}^k \lambda_i \langle g_i, y \rangle = 0$.

Ортогональное дополнение (2/2)

- Это означает, что вектор y является решением однородной системы линейных уравнений

$$\begin{cases} g_{11}y_1 + g_{12}y_2 + \ldots + g_{1n}y_n &= 0 \\ g_{21}y_1 + g_{22}y_2 + \ldots + g_{2n}y_n &= 0 \\ \ldots & \\ g_{k1}y_1 + g_{k2}y_2 + \ldots + g_{kn}y_n &= 0. \end{cases}$$

- Множество решений этой системы является линейным подпространством $\mathbb{F}_q^n$.
- Его размерность равна $n - \operatorname{rank} G = n - k$. □

Лемма 2
$(\mathcal{C}^\perp)^\perp = \mathcal{C}$.

Доказательство. Из определения очевидно, что $\mathcal{C} \subset (\mathcal{C}^\perp)^\perp$.

- С другой стороны, $\dim(\mathcal{C}^\perp)^\perp = n - (n-k) = k = \dim \mathcal{C}$, следовательно, включение не может быть строгим. □

Существование проверочной матрицы

Теорема 63

У любого линейного q-значного кода $\mathcal{C}$ есть проверочная матрица.

Доказательство. Пусть H — матрица, строки которой образуют базис подпространства $\mathcal{C}^\perp$.

- Поскольку $\dim \mathcal{C}^\perp = n - k$, матрица H имеет размеры $(n-k) \times n$.

• Тогда очевидно, что векторы, удовлетворяющие условию $Hx^T = 0$ — это в точности векторы, принадлежащие подпространству $(\mathcal{C}^\perp)^\perp = \mathcal{C}$. □

Замечание

• Легко видеть, что любая проверочная матрица кода $\mathcal{C}$ будет иметь описанный выше вид: её строки должны образовывать базис подпространства $\mathcal{C}^\perp$. В частности, строки проверочной матрицы всегда ЛНЗ.

• Проверочная матрица линейного кода не единственна.

• Как порождающая, так и проверочная матрица однозначно задают код $\mathcal{C}$.

• Для любой порождающей матрицы G и любой проверочной матрицы H кода $\mathcal{C}$ выполнено равенство $HG^T = 0$ (здесь 0 — это нулевая матрица размером $(n-k) \times k$).

Канонический вид проверочной матрицы (1/3)

- Пусть нам требуется закодировать сообщение при помощи q-значного линейного кода с параметрами $[n, k]$.
  - На вход подается слово $u = (u_1, \ldots, u_k) \in \mathbb{F}_q^k$.
  - Нужно закодировать его в слово $x = (x_1, \ldots, x_n) \in \mathbb{F}_q^n$.
- Простейший способ кодирования:
  - Берем $x_1 = u_1, \ldots, x_k = u_k$ — *информационные символы*;
  - $x_{k+1}, \ldots, x_n$ — *проверочные символы*, они задаются как линейные комбинации информационных символов.
    - Такой способ кодирования называется *систематическим кодированием*.
- Пусть $a_{i1}x_1 + \ldots + a_{ik}x_k + x_{k+i} = 0$, где $i \in [1..n-k]$ и $a_{ij} \in \mathbb{F}_q$.
- Тогда проверочная матрица имеет вид

$$H = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1k} & 1 & 0 & \ldots & 0 \\ a_{21} & a_{22} & \ldots & a_{2k} & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mk} & 0 & 0 & \ldots & 1 \end{pmatrix},$$

где $m = n - k$.

Канонический вид проверочной матрицы (2/3)

Определение 74

Проверочная матрица H линейного кода $\mathcal{C}$ имеет *канонический вид*, если её можно записать в виде $H = (A_{n-k,k} \mid E_{n-k})$, где $A_{n-k,k}$ — произвольная матрица размера $(n-k) \times k$ над полем $\mathbb{F}_q$ и E_{n-k} — единичная матрица порядка $n-k$.

Теорема 64

Линейный код $\mathcal{C}$ задается проверочной матрицей $H = (A_{n-k,k} \mid E_{n-k})$,
если и только если $\mathcal{C}$ задается порождающей матрицей $G = (E_k \mid -A_{n-k,k}^T)$.

Доказательство. Очевидно, что строки каждой из матриц G и H линейно независимы.

- Тем самым, достаточно доказать, что $HG^T = 0$.
- Это эквивалентно тому, что скалярное произведение любой строки H на любую строку G равно нулю.

Канонический вид проверочной матрицы (3/3)

- Рассмотрим i-ю строку H и j-ю строку G:
  - $H_i = (a_{i1}, \ldots, a_{ij}, \ldots, a_{ik}, \underbrace{0, \ldots, 0, 1, 0, \ldots, 0}_{n-k \text{ чисел}})$, где 1 стоит на позиции $k + i$;
  - $G_j = (\underbrace{0, \ldots, 0, 1, 0, \ldots, 0}_{k \text{ чисел}}, -a_{1j}, \ldots, -a_{ij}, \ldots, -a_{n-k,j})$, где 1 стоит на позиции j.
- Тогда $\langle H_i, G_j \rangle = a_{ij} \cdot 1 + 1 \cdot (-a_{ij}) = 0$. □

Замечание

Не всякий линейный код имеет проверочную матрицу канонического вида. Но всегда существует линейный код, эквивалентный данному, который имеет проверочную матрицу канонического вида.

Доказательство. Приведем матрицу H к ступенчатому виду:

$$H = \begin{pmatrix} 0 & \ldots & 0 & 1 & * & \ldots & * & 0 & * & \ldots & * & 0 & \ldots & 0 & * & \ldots \\ 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 1 & * & \ldots & * & 0 & \ldots & 0 & * & \ldots \\ 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 1 & \ldots & 0 & * & \ldots \\ \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots & \vdots & \\ 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & \ldots & 1 & * & \ldots \end{pmatrix}.$$

Теорема о столбцах проверочной матрицы

Теорема 65 (О столбцах проверочной матрицы)

Пусть H — проверочная матрица линейного кода $\mathcal{C}$. Тогда код $\mathcal{C}$ имеет кодовое расстояние d, если и только если любые $d-1$ столбцов матрицы H линейно независимы и найдутся d линейно зависимых столбцов.

Доказательство. Пусть $h_1, h_2, \ldots, h_n$ — столбцы матрицы H.

- Рассмотрим вектор $a = (a_1, a_2, \ldots, a_n) \in \mathcal{C} \setminus \{0\}$.
  - Пусть $w(a) = s$;
  - $a_{i_1}, a_{i_2}, \ldots, a_{i_s}$ — все ненулевые координаты a.
- Тогда $\sum\limits_{j=1}^{s} a_{i_j} h_{i_j} = Ha^T = 0$.
  - Следовательно, столбцы $h_{i_1}, h_{i_2}, \ldots, h_{i_s}$ линейно зависимы.
- Обратно, если столбцы $h_{i_1}, h_{i_2}, \ldots, h_{i_s}$ линейно зависимы, то найдется такой вектор $a \in \mathbb{F}_q^n \setminus \{0\}$, что $Ha^T = 0$ и $w(a) \leq s$.
- Таким образом, d — это наименьшее число линейно зависимых столбцов в H.

□

Граница Синглтона

Теорема 66 (R. C. Singleton, 1964)

Для любого линейного кода $\mathcal{C}$ с параметрами $[n, k, d]$ выполнено соотношение $n - k \geq d - 1$.

Доказательство. Пусть H — проверочная матрица $\mathcal{C}$.

- В этой матрице $n - k$ строк, следовательно, $\operatorname{rank} H \leq n - k$.
- Тогда любые $n - k + 1$ столбец матрицы H линейно зависимы.
- Таким образом, по теореме о столбцах проверочной матрицы получаем, что $d \leq n - k + 1$. □

Замечание

1. Аналогичное утверждение верно и для нелинейных кодов.
   А именно, если $\mathcal{C}$ — произвольный q-значный код длины n с кодовым расстоянием d и $M = |\mathcal{C}|$ — *мощность* кода $\mathcal{C}$, то $\log_q M \leq n - d + 1$.
2. Существуют коды, для которых граница Синглтона достигается.
   Они называются MDS-кодами (maximum distance separable).

Граница Хэмминга

Теорема 67

Пусть $A_q(n, d)$ — наибольшая мощность q-значного кода длины n с кодовым расстоянием d и $r = [\frac{d-1}{2}]$. Тогда

$$A_q(n, d) \leq \frac{q^n}{\sum_{i=0}^{r} C_n^i (q-1)^i}.$$

Доказательство. Пусть Σ — алфавит кода $\mathcal{C}$; $\mathcal{C} \subset \Sigma^n$.

- Для любого $x \in \Sigma^n$ рассмотрим шар $B_r(x)$ в метрическом пространстве (Σ^n, d) с центром x и радиусом r.
  - Напомним, что $B_r(x) = \{y \in \Sigma^n \mid d(x, y) \leq r\}$.
  - Мы доказывали, что $|B_r(x)| = \sum_{i=0}^{r} C_n^i (q-1)^i$.
- Если $d(x, y) \geq d$, то $B_r(x) \cap B_r(y) = \varnothing$.
- Следовательно, шары с центрами во всех кодовых словах не должны пересекаться.

□

Границы Хэмминга и Синглтона: замечания

Замечание

1. При $q = 2$ граница Хэмминга дает лучшую оценку, чем граница Синглтона.
   - В частности, при $q = 2$ возможны только *тривиальные* MDS-коды:
     - код, состоящий только из одного кодового слова (все нули);
     - код, состоящий из двух кодовых слов (все нули и все единицы);
     - код, содержащий все возможные кодовые слова (все возможные последовательности нулей и единиц);
     - код, содержащий все последовательности с четной суммой.

   Однако при $q > 2$ это уже не так: при $q > 2$ граница Синглтона чаще всего дает лучшую оценку, чем граница Хэмминга. Особенно если d достаточно велико.
2. Коды, для которых достигается граница Хэмминга называются *совершенным* или *плотно упакованным*.
   - Известно довольно много примеров двузначных совершенных кодов. Однако при $q > 2$ таких примеров очень мало.

Граница Варшамова-Гилберта

Теорема 68 (P. P. Варшамов, Edgar N. Gilbert)

Если $\sum_{i=0}^{d-2} C_{n-1}^{i}(q-1)^i < q^r$, *то существует* q*-значный линейный код длины* n *с кодовым расстоянием не менее* d, *имеющий не более чем* r *проверочных символов (т. е.* $[n, k, d']$*–код, где* $k \geq n - r$ *и* $d' \geq d$*).*

Доказательство. Нужно построить матрицу H размера $r \times n$ так, чтобы любые $d-1$ её столбцов были линейно независимы.

- Выбираем последовательно столбцы высоты r.
- Очередной столбец не должен быть линейной комбинацией никаких $d-2$ из ранее выбранных столбцов.
- Пусть уже выбраны $t-1$ столбцов.
  - Тогда не подходит максимум $\sum_{i=0}^{d-2} C_{t-1}^{i}(q-1)^i$ столбцов.
  - А всего возможных столбцов — q^r.
- То есть если $\sum_{i=0}^{d-2} C_{t-1}^{i}(q-1)^i < q^r$, то можно выбрать еще один столбец.
- Таким образом, если $\sum_{i=0}^{d-2} C_{n-1}^{i}(q-1)^i < q^r$, то можно выбрать не менее n столбцов.

□

Двузначный код Хэмминга (1/3)

- Пусть $q = 2$ и $n = 2^m - 1$, где $m \in \mathbb{N}$.
- Рассмотрим линейный код, задаваемый проверочной матрицей H_m следующего вида:
  - матрица H_m имеет размер $m \times n$;
  - её столбцы — все ненулевые векторы длины m;
  - i-й столбец представляет из себя двоичную запись числа i
    - двоичная запись составляется из m разрядов, в случае необходимости, в её начало дописывается нужное число нулей;
    - разряды записываются "сверху вниз" — самый младших разряд должен оказаться в нижней строчке.

Пример

$$H_3 = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

- Поскольку все столбцы различны, $d = 3$.
- То есть получили линейный двузначный код с параметрами $[2^m - 1, 2^m - m - 1, 3]$.

Двузначный код Хэмминга (2/3)

Определение 75

Линейный код, заданный определенной выше проверочной матрицей H_m, называется *кодом Хэмминга*.

Теорема 69

Код Хэмминга является совершенным кодом.

Доказательство. $|B_1(u)| = n + 1 = 2^m$.

- Следовательно, $2^n / |B_1(u)| = 2^k$. □

Замечание

- Пусть $u = (u_1, u_2, \ldots, u_n)$ — кодовое слово кода Хэмминга. Его информационные символы — все u_i, где i — не степень двойки; $u_1, u_2, u_4, \ldots, u_{2^{m-1}}$ — проверочные символы.
- Значения проверочных символов можно выразить через информационные при помощи проверочной матрицы.
- Проверочная матрица здесь записана не в канонической форме. Её можно привести к каноническому виду, переставив столбцы.

Двузначный код Хэмминга (3/3)

- Пусть $x = (x_1, \ldots, x_n)$ — закодированное кодом Хэмминга сообщение.
- При передаче по каналу связи было получено сообщение y, которое может содержать ошибку.
  - Предполагаем, что могло произойти не более одной ошибки типа замещения.
- Пусть $\varepsilon = y - x$ — вектор ошибки.
  - Наше предположение означает, что вектор ε может быть либо нулевым, либо ровно одна его координата равна 1.
  - Обозначим через e_i вектор $(0, \ldots, 0, 1, 0, \ldots, 0)$, где единица находится на i-м месте.
- Тогда возможны следующие варианты.
  - Если $\varepsilon = 0$, то $H_m y = H_m x = 0$;
  - если $\varepsilon = e_i$, то $H_m y = H_m x + H_m e_i = B(i)$,
    где $B(i)$ — i-й столбец матрицы H_m, представляющий из себя двоичную запись числа i, то есть номера разряда, в котором произошла ошибка.

3.3. Циклические коды

Определение 76

Линейный код $\mathcal{C}$ длины n называется *циклическим*, если
$\forall x_1, x_2, \ldots, x_n \, ((x_1, x_2, \ldots, x_n) \in \mathcal{C} \to (x_2, \ldots, x_n, x_1) \in \mathcal{C})$.

Циклические коды удобно представлять при помощи многочленов

- Пусть $p \in \mathbb{P}$. Будем использовать в качестве алфавита поле $\mathbb{F}_p$.
- Пусть $a = (a_0, a_1, \ldots, a_{n-1}) \in \mathbb{F}_p^n$ — некоторое сообщение.
- Поставим ему в соответствие многочлен $a(x) = a_0 + a_1 x + \ldots + a_{n-1} x^{n-1} \in \mathbb{F}_p[x]$.
- Такие многочлены удобно рассматривать по модулю многочлена $x^n - 1$.
  - То есть мы будем смотреть на сообщение a как на класс вычетов $\overline{a(x)} \in \mathbb{F}_p[x]/(x^n - 1)$.
  - Для обозначения этого класса вычетов мы как правило будем использовать многочлен $a(x)$, степень которого меньше n (в каждом классе вычетов по модулю $x^n - 1$ есть ровно один такой многочлен).
- Итак, далее мы будем считать, что $\mathcal{C} \subset \mathbb{F}_p[x]/(x^n - 1)$.

Циклические коды и идеалы

Теорема 70

Подмножество $\mathcal{C} \subset \mathbb{F}_p[x]/(x^n - 1)$ является циклическим кодом тогда и только тогда, когда оно образует идеал.

Доказательство. Заметим, что в кольце $\mathbb{F}_p[x]/(x^n - 1)$ циклический сдвиг коэффициентов многочлена происходит при домножении на x.

• А именно, если $c(x) = c_0 + c_1 x + \ldots + c_{n-1}x^{n-1} \in \mathbb{F}_p[x]$, то $xc(x) = c_0 x + c_1 x^2 + \ldots + c_{n-1}x^n \equiv c_{n-1} + c_0 x + \ldots + c_{n-2}x^{n-1} \pmod{x^n - 1}$.

"$\Leftarrow$": пусть $\mathcal{C}$ — идеал в $\mathbb{F}_p[x]/(x^n - 1)$. Очевидно, что тогда $\mathcal{C}$ — линейное подпространство в $\mathbb{F}_p[x]/(x^n - 1)$.

• То, что код циклический, сразу следует из того, что если $c(x) \in \mathcal{C}$, то и $xc(x) \in \mathcal{C}$.

"$\Rightarrow$": пусть $\mathcal{C}$ — циклический код. Тогда очевидно, что $0 \in \mathcal{C}$ и если $f(x), g(x) \in \mathcal{C}$, то $f(x) \pm g(x) \in \mathcal{C}$ и $xf(x) \in \mathcal{C}$.

• Из этого следует, что $\mathcal{C}$ — идеал. □

Порождающий многочлен циклического кода (1/3)

Теорема 71

Пусть $\mathcal{C} \subset \mathbb{F}_p[x]/(x^n - 1)$ — циклический код; r — минимальная степень ненулевого многочлена из $\mathcal{C}$. Тогда

1. *в $\mathcal{C}$ есть ровно один унитарный многочлен $g(x)$ степени r;*
2. $x^n - 1 \vdots g(x)$;
3. $\mathcal{C} = (g) = \{ga \mid \deg a < n - r\}$.

Доказательство.

1. Пусть $g_1, g_2 \in \mathcal{C}$, $\deg g_1 = \deg g_2 = r$ и g_1, g_2 унитарны.
   - Тогда $g_1 - g_2 \in \mathcal{C}$ и $\deg(g_1 - g_2) < r$. Следовательно, $g_1 = g_2$.
2. Поделим с остатком.
   - Пусть $x^n - 1 = g(x)h(x) + s(x)$, где $\deg s < \deg g = r$.
   - Тогда $s(x) \in \mathcal{C}$, следовательно, $s(x) = 0$.
3. Пусть $c \in \mathcal{C}$, $c(x) = g(x)a(x) + s(x)$, где $\deg s < \deg g$.
   - Тогда $s(x) \in \mathcal{C}$, откуда $s(x) = 0$ и $c(x) = g(x)a(x)$.
   - Очевидно, что $\deg a < n - r$.

□

Порождающий многочлен циклического кода (2/3)

Определение 77

Определенный выше многочлен $g(x)$ называется *порождающим многочленом* циклического кода $\mathcal{C}$.

Следствие 1

Размерность кода $\mathcal{C}$ равна $n - r$.

Доказательство. Пусть $k = n - r$ и $a(x) = a_0 + a_1 x + \ldots + a_{k-1} x^{k-1}$.

- Тогда $g(x)a(x) = a_0 \cdot g(x) + a_1 \cdot xg(x) + \ldots + a_{k-1} \cdot x^{k-1} g(x)$ — линейная комбинация многочленов $g(x), xg(x), \ldots, x^{k-1} g(x)$.
- По пункту 3 теоремы 71, все многочлены из $\mathcal{C}$ представляются в виде таких линейных комбинаций.
  - То есть $g(x), xg(x), \ldots, x^{k-1} g(x)$ — порождающая система в $\mathcal{C}$.
- Далее, пусть $a \neq 0$. Тогда, поскольку $\deg a < n - r$, получаем, что $ga \not\vdots x^n - 1$.
  - То есть $ga \neq 0$ в $\mathbb{F}_p[x]/(x^n - 1)$.
  - Следовательно, многочлены $g(x), xg(x), \ldots, x^{k-1} g(x)$ — ЛНЗ.
- Таким образом, $g(x), xg(x), \ldots, x^{k-1} g(x)$ — базис в $\mathcal{C}$, откуда $\dim \mathcal{C} = k$. □

Порождающий многочлен циклического кода (3/3)

Следствие 2

Любой унитарный делитель $g(x)$ многочлена $x^n - 1$ является порождающим многочленом некоторого циклического кода длины n.

Доказательство. Рассмотрим идеал (g) в кольце $\mathbb{F}_p[x]/(x^n - 1)$. Нужно доказать, что g имеет наименьшую степень среди всех ненулевых элементов этого идеала.

- Пусть $a \in \mathbb{F}_p[x]$. Поделим с остатком ga на $x^n - 1$: $g(x)a(x) = (x^n - 1)q(x) + s(x)$.
- Тогда $s(x) \vdots g(x)$, следовательно, либо $s = 0$, либо $\deg s \geq \deg r$. □

Замечание

- Из теоремы 71 следует, что все идеалы в кольце $\mathbb{F}_p[x]/(x^n - 1)$ главные.
- Тем не менее, это кольцо нельзя называть кольцом главных идеалов, поскольку в нем есть делители нуля.
- В определении порождающего многочлена условие унитарности не обязательно. Но так удобнее.

Циклические коды: кодирование

- Пусть $a(x) = a_0 + a_1 x + \ldots + a_{k-1} x^{k-1}$ — исходное сообщение.
- Есть два способа закодировать его в сообщение $c(x) \in \mathcal{C}$.
  1. Несистематический кодер. $c(x) = a(x)g(x) \in \mathcal{C}$;
  2. Систематический кодер. $c(x) = x^r a(x) - s(x)$, где $s(x)$ — остаток от деления $x^r a(x)$ на $g(x)$.
     - То есть в этом случае мы заменяем вектор $(a_0, a_1, \ldots, a_k)$ на вектор $(\lambda_0, \ldots, \lambda_{r-1}, a_0, a_1, \ldots, a_k)$, где $-s(x) = \lambda_0 + \lambda_1 x + \ldots + \lambda_{r-1} x^{r-1}$.

Замечание

• Первый из указанных выше способов несистематический в том смысле, что коэффициенты многочлена $a(x)$ не обязаны присутствовать среди коэффициентов многочлена $c(x)$. Тем не менее, этот способ часто оказывается удобным из-за простоты кодирования.

• Второй способ систематический: поскольку $\deg s < r$, все коэффициенты многочлена $a(x)$ являются коэффициентами многочлена $c(x)$. А именно, $a_i = c_{i+r}$.

Порождающая матрица циклического кода

Теорема 72

Пусть $g(x) = g_0 + g_1 x + \ldots + g_r x^r$ *— порождающий многочлен циклического кода* $\mathcal{C}$*. Тогда матрица*

$$G = \begin{pmatrix} g_0 & g_1 & g_2 & \ldots & g_r & 0 & 0 & \ldots & 0 \\ 0 & g_0 & g_1 & g_2 & \ldots & g_r & 0 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & g_0 & g_1 & g_2 & \ldots & g_r & 0 \\ 0 & 0 & \ldots & 0 & g_0 & g_1 & g_2 & \ldots & g_r \end{pmatrix}$$

является порождающей матрицей кода $\mathcal{C}$*.* (*Матрица имеет размеры* $(n-r) \times n$*: в каждой её строке стоят* $r+1$ *коэффициент многочлена* g *и* $n-r-1$ *нулей.*)

Доказательство. Заметим, что все строки матрицы принадлежат $\mathcal{C}$: строка номер i соответствует многочлену $x^{i-1}g(x)$.

- Строки G — ЛНЗ. Действительно, $g_r = 1$, поэтому последние $n-r$ столбцов G образуют нижнетреугольную матрицу с единицами на главной диагонали.
- Поскольку $\dim \mathcal{C} = n-r$, получаем, что строки G образуют базис в $\mathcal{C}$. □

Проверочный многочлен циклического кода

Определение 78

Проверочным многочленом циклического кода $\mathcal{C}$ называется такой многочлен $h(x) \in \mathbb{F}_p[x]$, что $g(x)h(x) = x^n - 1$ (где g — порождающий многочлен кода $\mathcal{C}$).

Замечание

Легко видеть, что $\deg h = n - r = k$, где $r = \deg g$ и $k = \dim \mathcal{C}$.

Теорема 73

Пусть $c \in \mathbb{F}_p[x]$, $\deg c < n$. Тогда $c \in \mathcal{C}$, если и только если $h(x)c(x) \vdots x^n - 1$.

Доказательство. "$\Rightarrow$": пусть $c \in \mathcal{C}$. Тогда $c(x) = g(x)a(x)$, где $a \in \mathbb{F}_p[x]$.

- Следовательно, $h(x)c(x) = h(x)g(x)a(x) = (x^n - 1)a(x) \vdots x^n - 1$.

"$\Leftarrow$": пусть $h(x)c(x) = (x^n - 1)f(x)$, где $f \in \mathbb{F}_p[x]$.

- Тогда $h(x)c(x) = (x^n - 1)f(x) = h(x)g(x)f(x)$, откуда $c(x) = g(x)f(x) \in \mathcal{C}$. $\square$

Проверочная матрица циклического кода (1/2)

Теорема 74

Пусть $h(x) = h_0 + h_1 x + \ldots + h_k x^k$ *— проверочный многочлен циклического кода* $\mathcal{C}$*. Тогда матрица*

$$H = \begin{pmatrix} 0 & 0 & \ldots & 0 & h_k & \ldots & h_2 & h_1 & h_0 \\ 0 & \ldots & 0 & h_k & \ldots & h_2 & h_1 & h_0 & 0 \\ \vdots & \iddots & \iddots & & \iddots & \iddots & \iddots & \iddots & \vdots \\ 0 & h_k & \ldots & h_2 & h_1 & h_0 & 0 & \ldots & 0 \\ h_k & \ldots & h_2 & h_1 & h_0 & 0 & \ldots & 0 & 0 \end{pmatrix}$$

является проверочной матрицей кода $\mathcal{C}$*. (Матрица имеет размеры* $r \times n$ *: в каждой её строке стоят* $k+1$ *коэффициент многочлена* g *и* $r-1$ *нулей.)*

Доказательство. Заметим, что все строки матрицы ЛНЗ, поскольку $h_k = 1$.

- Пусть $c(x) = c_0 + c_1 x + \ldots + c_{n-1} x^{n-1} \in \mathcal{C}$.
- Мы знаем, что $c(x)h(x) \vdots x^n - 1$. При этом, $\deg(ch) < n + k$.
- Тогда коэффициенты при $x^k, x^{k+1}, \ldots, x^{n-1}$ многочлена ch равны нулю.

Проверочная матрица циклического кода (2/2)

- Заметим, что коэффициент при x^{k+t} многочлена ch равен $\sum_{i=0}^{k+t} c_i h_{k+t-i}$.
  - То есть $\sum_{i=0}^{k+t} c_i h_{k+t-i} = 0$ при $t \in [0..r-1]$.
  - Но написанная выше сумма — это скалярное произведение вектора c на $(r-t)$-ю строку матрицы H. □

Декодирование циклического кода

Пусть

- $a(x)$ — исходное сообщение;
- $c(x)$ — кодированное сообщение;
- $c'(x)$ — принятое сообщение (возможно, содержит ошибки);
- $\varepsilon(x) \stackrel{\text{def}}{=} c'(x) - c(x)$ — вектор ошибки.

- Тогда $\varepsilon(x) \equiv c'(x) \pmod{g(x)}$.
- Многочлен $\varepsilon(x)$ можно найти перебирая все векторы малого веса (начиная с нуля).

3.4. Коды БЧХ (Боуз-Чоудхури-Хоквингем)

- Пусть $p \in \mathbb{P}$. Мы будем рассматривать циклические коды над полем $\mathbb{F}_p$ длины $n = p^m - 1$, где $m \in \mathbb{N}$.
- Тогда $(x^n - 1)x = x^q - x$, где $q = p^m$. Следовательно, многочлен $x^n - 1$ не имеет кратных корней и его корнями являются все ненулевые элементы поля $\mathbb{F}_q$.

Теорема 75 (о нулях кода)

Пусть $\mathcal{C}$ — циклический код над $\mathbb{F}_p$ длины $n = p^m - 1$; $g(x)$ — порождающий многочлен кода $\mathcal{C}$; $\beta_1, \beta_2, \ldots, \beta_r \in \mathbb{F}_q$ — все корни многочлена g; $f(x) \in \mathbb{F}_p[x]$, $\deg f < n$. Тогда $f \in \mathcal{C}$, если и только если $f(\beta_1) = f(\beta_2) = \ldots = f(\beta_r) = 0$.

Доказательство. "$\Rightarrow$": по теореме 71 имеем $f = ga$, где $a \in \mathbb{F}_p[x]$.

- Следовательно, $f(\beta_i) = g(\beta_i)a(\beta_i) = 0$ при всех $i \in [1..r]$.

"$\Leftarrow$": разделим f на g с остатком: $f = ga + s$, где $\deg s < r$.

- Тогда $s(\beta_i) = f(\beta_i) - g(\beta_i)a(\beta_i) = 0$ при всех $i \in [1..r]$.
- То есть многочлен $s(x)$ имеет r различных корней и при этом $\deg s < r$.
- Следовательно, $s = 0$. Тогда $f(x) = g(x)a(x) \in \mathcal{C}$. $\square$

Теорема о границе БЧХ (1/3)

Определение 79

Нулями циклического кода $\mathcal{C}$ называются корни его порождающего многочлена.

• Далее, мы выберем в поле $\mathbb{F}_q$ примитивный элемент α и будем выражать ненулевые элементы $\mathbb{F}_q$ как степени α.

Теорема 76 (граница БЧХ)

Пусть $\mathcal{C}$ — p-значный циклический код длины n;
$g(x)$ — порождающий многочлен кода $\mathcal{C}$;
$b, \delta \in \mathbb{Z}$ таковы, что $b \geq 0$, $\delta > 1$ и $g(\alpha^b) = g(\alpha^{b+1}) = \ldots = g(\alpha^{b+\delta-2}) = 0$.
Тогда кодовое расстояние кода $\mathcal{C}$ не меньше δ.

Доказательство. Предположим противное: пусть в $\mathcal{C}$ есть ненулевой элемент, вес Хэмминга которого меньше δ.

• Этому элементу соответствует многочлен
$f(x) = c_1 x^{k_1} + c_2 x^{k_2} + \ldots + c_{\delta-1} x^{k_{\delta-1}} \in \mathcal{C}$,
где $c_1, c_2, \ldots, c_{\delta-1} \in \mathbb{F}_p$ — не все нули.

• По теореме 75 имеем $f(\alpha^b) = f(\alpha^{b+1}) = \ldots = f(\alpha^{b+\delta-2}) = 0$.

Теорема о границе БЧХ (2/3)

- Получаем следующие равенства:

$$\begin{cases} c_1\alpha^{k_1 b} & + & c_2\alpha^{k_2 b} & + \ldots + & c_{\delta-1}\alpha^{k_{\delta-1} b} & = & 0 \\ c_1\alpha^{k_1 b+k_1} & + & c_2\alpha^{k_2 b+k_2} & + \ldots + & c_{\delta-1}\alpha^{k_{\delta-1} b+k_{\delta-1}} & = & 0 \\ \ldots & & & & & & \\ c_1\alpha^{k_1 b+k_1(\delta-2)} & + & c_2\alpha^{k_2 b+k_2(\delta-2)} & + \ldots + & c_{\delta-1}\alpha^{k_{\delta-1} b+k_{\delta-1}(\delta-2)} & = & 0. \end{cases}$$

- На эти равенства можно смотреть как на однородную систему линейных уравнений, в которой $c_1, c_2, \ldots, c_{\delta-1}$ — неизвестные, и степени α — коэффициенты.
- Получилась однородная система линейных уравнений, имеющая нетривиальное решение. Тогда матрица этой системы должна быть вырожденной.
- Следовательно,

$$\begin{vmatrix} \alpha^{k_1 b} & \alpha^{k_2 b} & \ldots & \alpha^{k_{\delta-1} b} \\ \alpha^{k_1 b+k_1} & \alpha^{k_2 b+k_2} & \ldots & \alpha^{k_{\delta-1} b+k_{\delta-1}} \\ \vdots & \vdots & & \vdots \\ \alpha^{k_1 b+k_1(\delta-2)} & \alpha^{k_2 b+k_2(\delta-2)} & \ldots & \alpha^{k_{\delta-1} b+k_{\delta-1}(\delta-2)} \end{vmatrix} = 0.$$

Теорема о границе БЧХ (3/3)

- Преобразуем:

$$0 = \begin{vmatrix} \alpha^{k_1 b} & \alpha^{k_2 b} & \dots & \alpha^{k_{\delta-1} b} \\ \alpha^{k_1 b + k_1} & \alpha^{k_2 b + k_2} & \dots & \alpha^{k_{\delta-1} b + k_{\delta-1}} \\ \vdots & \vdots & & \vdots \\ \alpha^{k_1 b + k_1 (\delta-2)} & \alpha^{k_2 b + k_2 (\delta-2)} & \dots & \alpha^{k_{\delta-1} b + k_{\delta-1} (\delta-2)} \end{vmatrix} =$$

$$= \alpha^{(k_1 + k_2 + \ldots + k_{\delta-1}) b} \begin{vmatrix} 1 & 1 & \dots & 1 \\ \alpha^{k_1} & \alpha^{k_2} & \dots & \alpha^{k_{\delta-1}} \\ \vdots & \vdots & & \vdots \\ \alpha^{k_1 (\delta-2)} & \alpha^{k_2 (\delta-2)} & \dots & \alpha^{k_{\delta-1} (\delta-2)} \end{vmatrix} =$$

$$= \alpha^{(k_1 + k_2 + \ldots + k_{\delta-1}) b} \prod_{i<j} (\alpha^{k_i} - \alpha^{k_j}) \neq 0.$$

  - Последнее из написанных выше равенств — это определитель Вандермонда.
- Полученное противоречие завершает доказательство. □

Коды БЧХ (1/4)

Определение 80

Кодом БЧХ над полем $\mathbb{F}_p$ длины $n = p^m - 1$ с *конструктивным расстоянием* $\delta > 1$ называется циклический код с порождающим многочленом наименьшей степени, корнями которого являются элементы $\alpha^b, \alpha^{b+1}, \ldots, \alpha^{b+\delta-2}$, где α — примитивный элемент поля $\mathbb{F}_{p^m}$ и b — некоторое неотрицательное целое число.

Замечание

Это определение можно эквивалентно переформулировать следующим образом.

- Обозначим через $M^{(s)}(x)$ минимальный многочлен элемента α^s. По теореме 60 имеем

$$M^{(s)}(x) = (x - \alpha^s)(x - \alpha^{ps})(x - \alpha^{p^2 s}) \ldots = \prod_{i \in C_s} (x - \alpha^i).$$

- Тогда код БЧХ над полем $\mathbb{F}_p$ длины $n = p^m - 1$ с конструктивным расстоянием $\delta > 1$ — это циклический код с порождающим многочленом

$$g(x) \stackrel{\text{def}}{=} [M^{(b)}(x), M^{(b+1)}(x), \ldots, M^{(b+\delta-2)}(x)],$$

где b — некоторое неотрицательное целое число.

Коды БЧХ (2/4)

Теорема 77

Код БЧХ $\mathcal{C}$ над полем $\mathbb{F}_p$ длины $n = p^m - 1$ с конструктивным расстоянием $\delta > 1$ имеет параметры $d \geq \delta$ и $k \geq n - (\delta - 1)m$.

Доказательство.

- То, что $d \geq \delta$, непосредственно следует из теоремы 76.
- Рассмотрим порождающий многочлен

$$g(x) = [M^{(b)}(x), M^{(b+1)}(x), \ldots, M^{(b+\delta-2)}(x)]$$

кода $\mathcal{C}$.

  - Заметим, что

$$\deg g \leq \deg M^{(b)} + M^{(b+1)} + M^{(b+\delta-2)} \leq (\delta - 1)m.$$

  - Но тогда

$$k = n - \deg g \geq n - (\delta - 1)m.$$

□

Коды БЧХ (3/4)

Примеры

1. Пусть $p = 2$, $n = 2^m - 1$ и $g(x) = M^{(1)}(x)$.
   - Тогда $\deg g = m$, следовательно, $k = 2^m - m - 1$.
   - Далее, элементы α и α^2 являются корнями g. Следовательно, $d \geq 3$.

Замечание

Можно заметить, что построенный выше код эквивалентен коду Хэмминга.

- Действительно, поскольку $n = 2^m - 1$ и $k = 2^m - m - 1$, проверочная матрица должна состоять из $2^m - 1$ столбцов высоты m.
- Поскольку $d > 2$, все эти столбцы должны быть ненулевыми и различными.
- Но тогда столбцы проверочной матрицы — это в точности все ненулевые столбцы высоты m.
- Тем самым, проверочная матрица построенного выше кода, с точностью до перестановки столбцов, совпадает с проверочной матрицей кода Хэмминга.

Коды БЧХ (4/4)

2. Пусть $p = 2$, $n = 2^m - 1$ и $g(x) = M^{(1)}(x)M^{(3)}(x)$.
   - Тогда $\deg g \leq 2m$, следовательно, $k \geq 2^m - 2m - 1$.
     - На самом деле, при $m > 1$ верно, что $\deg M^{(3)} = m$, так что $k = 2^m - 2m - 1$.
   - Далее, элементы α, α^2 и α^4 являются корнями $M^{(1)}(x)$, а элемент α^3 является корнем $M^{(3)}(x)$.
     - То есть α, α^2, α^3 и α^4 — корни $g(x)$.
     - Следовательно, $d \geq 5$.

Коды Рида-Соломона

- Пусть $p \in \mathbb{P}$, $m \in \mathbb{N}$, $q = p^m > 2$, α — примитивный элемент поля $\mathbb{F}_q$.

Определение 81

Код Рида-Соломона — это код БЧХ длины $q-1$ над полем $\mathbb{F}_q$ с порождающим многочленом

$$g(x) = (x - \alpha^b)(x - \alpha^{b+1}) \dots (x - \alpha^{b+\delta-2}),$$

где $b, \delta \in \mathbb{Z}$, $b \geq 0$ и $\delta > 1$.

Теорема 78

Код Рида-Соломона имеет параметры $n = q-1$, $k = n - \delta + 1$ *и* $d = \delta = n - k + 1$.

Доказательство. $k = n - \deg g = n - \delta + 1$;

- $d \geq \delta$ по теореме о границе БЧХ;
- $d \leq \delta$ по теореме о границе Синглтона. □

Замечание

Код Рида-Соломона является MDS-кодом: он достигает границу Синглтона.

Глава 4.
Теория графов

- Дополнительные материалы:
  - Т. Кормен, Ч. Лейзерсон, Р. Ривест, *Алгоритмы: построение и анализ*. М.: МЦНМО, 2002.
  - Д. В. Карпов, *Теория графов*. М.: МЦНМО, 2022.
  - Ф. Харари, *Теория графов*. М.: Мир, 1973.
  - Ф. Харари, Э. Палмер, *Перечисление графов*. М.: Мир, 1977.

4.1. Сети и потоки

Мы дадим два эквивалентных определения сети.

Определение 82

- *Сетью* называется упорядоченная четверка (G, s, t, c), где
  - $G = (V, E)$ — конечный ориентированный граф;
  - $s, t \in V$ — такие вершины, что $d^-(s) = d^+(t) = 0$;
  - $c : E \to \mathbb{R}^+$.
- Вершина s называется *истоком*; вершина t — *стоком*; функция c — *пропускной способностью*.

Определение 83

- *Сетью* называется упорядоченная четверка (V, s, t, c), где V — конечное множество и $s, t \in V$; $c : V \times V \to \mathbb{R}$ — такая функция, что $\forall x, y \in V\,(c(x, y) \geq 0)$ и $\forall x \in V\,(c(x, s) = c(t, x) = 0)$.
- Как и в предыдущем определении, вершина s называется *истоком*; вершина t — *стоком*; функция c — *пропускной способностью*.

Определения сети: обсуждение

Замечание

- Легко видеть, что два приведенных выше определения сети эквивалентны.
  - Действительно, второе определение непосредственно получается из первого заменой орграфа G на множество его вершин V.
  - Пусть $N = (V, s, t, c)$ — сеть в смысле второго определения. Построим соответствующий этой сети орграф $G_N = (V_N, E_N)$ следующим образом: $V_N \stackrel{\text{def}}{=} V$ и $E_N \stackrel{\text{def}}{=} \{(x, y) \in V \times V \mid c(x, y) > 0\}$. Нетрудно проверить, что тогда четверка (G_N, s, t, c) удовлетворяет всем условиям первого определения.
- Условие о том, что нет дуг, входящих в исток, и нет дуг, исходящих из стока ($d^-(s) = d^+(t) = 0$ в первом определении или $\forall x \in V\,(c(x, s) = c(t, x) = 0)$ во втором) является естественным, но не обязательным. Далее мы будем рассматривать немного обобщенные сети, в которых это условие может не выполняться.
- Далее мы будем в основном использовать второе определение сети, но при этом будем также рассматривать соответствующий сети N орграф G_N.

Определение потока

Определение 84

- Пусть $N = (V, s, t, c)$ — сеть. *Потоком* в сети N называется функций $f : V \times V \to \mathbb{R}$, удовлетворяющая следующим условиям
  1. ограничение пропускной способности: $\forall x, y \in V \, (f(x, y) \leq c(x, y))$;
  2. кососимметричность: $\forall x, y \in V \, (f(x, y) = -f(y, x))$;
  3. сохранение потока: $\forall v \in V \setminus \{s, t\} \, \big(\sum\limits_{x \in V} f(v, x) = 0 \big)$.
- Число $|f| \stackrel{\text{def}}{=} \sum\limits_{x \in V} f(s, x)$ называется *величиной* потока f.
- Поток f называется *максимальным*, если он имеет наибольшую величину среди всех потоков в сети N.

Замечание

- Из свойства 2 получаем, что для любой вершины $x \in V$ выполнено $f(x, x) = -f(x, x)$. Следовательно, $f(x, x) = 0$.

Задача о максимальном потоке

Нас будет интересовать *задача о максимальном потоке*:

- Дана сеть $N = (V, s, t, c)$.
- Найти максимальный поток f^* в сети N.

Замечание

- Существование максимального потока, вообще говоря, не очевидно. Тем не менее, чуть позже мы докажем, что в любой сети есть максимальный поток.
- Максимальный поток может быть не единственен. Поэтом наша цель заключается в том, чтобы найти один из потоков максимальной величины.

- Основная идея метода нахождения максимального потока заключается в следующем: мы начинаем с нулевого потока и далее последовательно увеличиваем имеющийся поток до тех пор, пока он не станет максимальным.

Остаточная сеть

Определение 85

Пусть $N = (V, s, t, c)$ — сеть и f — поток в сети N.
Остаточной сетью потока f называется сеть $N_f = (V, s, t, c_f)$, где
$c_f(x, y) \stackrel{\text{def}}{=} c(x, y) - f(x, y)$ — *остаточная пропускная способность*.

Замечание

- Заметим, что остаточная пропускная способность $c_f(x, y)$ — это то, насколько мы можем потенциально увеличить поток из x в y.
- Отметим, что в орграфе, соответствующем остаточной сети N_f могут быть дуги, которых не было в исходном орграфе G_N.
  - Действительно, может быть так, что $c(x, y) = 0$ и $f(x, y) < 0$.
    Тогда $c_f(x, y) = c(x, y) - f(x, y) = -f(x, y) > 0$.
  - Однако, если $c(x, y) = c(y, x) = 0$, то тогда и $c_f(x, y) = c_f(y, x) = 0$.

Лемма о сложении потоков

Лемма 1

Пусть f — поток в сети N и f' — поток в остаточной сети N_f.
Тогда $f + f'$ — поток в сети N и $|f + f'| = |f| + |f'|$.

Доказательство.

- Проверим, что $f + f'$ удовлетворяет трем условиям из определения потока.
  1. $(f + f')(x, y) = f(x, y) + f'(x, y) \le f(x, y) + c_f(x, y) =$
     $= f(x, y) + (c(x, y) - f(x, y)) = c(x, y).$
  2. $(f + f')(x, y) = f(x, y) + f'(x, y) = -f(y, x) - f'(y, x) = -(f + f')(y, x).$
  3. Пусть $v \in V \setminus \{s, t\}$.
     Тогда $\sum\limits_{x \in V} (f + f')(v, x) = \sum\limits_{x \in V} f(v, x) + \sum\limits_{x \in V} f'(v, x) = 0.$
- $|f + f'| = \sum\limits_{x \in V} (f + f')(s, x) = \sum\limits_{x \in V} f(s, x) + \sum\limits_{x \in V} f'(s, x) = |f| + |f'|.$

□

Дополняющий путь (1/2)

Определение 86

- Пусть f — поток в сети $N = (V, s, t, c)$.
- *Дополняющим путем* называется простой путь p из s в t в остаточной сети N_f (точнее говоря, в соответствующем этой сети орграфе G_{N_f}).
- Число $c_f(p) \stackrel{\text{def}}{=} \min\{c_f(x, y) \mid (x, y) \in p\}$ называется *остаточной пропускной способностью* дополняющего пути p.

Лемма 2

Пусть f — поток в сети $N = (V, s, t, c)$; p — дополняющий путь в N_f.

Тогда функция $f_p(x, y) \stackrel{\text{def}}{=} \begin{cases} c_f(p), & (x, y) \in p \\ -c_f(p), & (y, x) \in p \\ 0, & \text{в остальных случаях} \end{cases}$ *задает поток в N_f.*

При этом, $|f_p| = c_f(p) > 0$.

Доказательство. Проверим условия из определения потока.

1. Пусть $x, y \in V$. Тогда, если $(x, y) \in p$, то $f_p(x, y) = c_f(p) \leq c_f(x, y)$; в противном случае, $f_p(x, y) \leq 0 \leq c_f(x, y)$.

Дополняющий путь (2/2)

2. Пусть $x, y \in V$.
   - Если $(x, y) \in p$ или $(y, x) \in p$, то равенство $f_p(x, y) = -f_p(y, x)$ следует непосредственно из определения функции f_p.
   - В противном случае, имеем $f_p(x, y) = 0 = -f_p(y, x)$.
3. Пусть $v \in V \setminus \{s, t\}$.
   - Если путь p не проходит через v, то $\sum\limits_{x \in V} f_p(v, x) = 0$, поскольку все слагаемые этой суммы равны нулю.
   - Пусть путь p проходит через v, а именно, $(u, v) \in p$ и $(v, w) \in p$. Тогда $\sum\limits_{x \in V} f_p(v, x) = f_p(v, u) + f_p(v, w) = -c_f(p) + c_f(p) = 0$.

- Пусть (s, a) — первая дуга пути p. Тогда $|f_p| = \sum\limits_{x \in V} f_p(s, x) = f_p(s, a) = c_f(p)$. □

Следствие 1

Пусть f — поток в сети N и p — дополняющий путь в N_f.
Тогда в сети N есть такой поток f', что $|f'| = |f| + c_f(p)$.

Разрезы

Определение 87

- *Разрезом* в сети $N = (V, s, t, c)$ называется упорядоченная пара множеств (S, T), такая, что $s \in S$, $t \in T$, $S \cap T = \varnothing$ и $S \cup T = V$.
- Величина $c(S, T) \stackrel{\text{def}}{=} \sum\limits_{x \in S, y \in T} c(x, y)$ называется *пропускной способностью* разреза (S, T).
  - Разрез, имеющий наименьшую пропускную способность, называется *минимальным*.
- Пусть f — поток в сети N. Тогда величина $f(S, T) \stackrel{\text{def}}{=} \sum\limits_{x \in S, y \in T} f(x, y)$ называется *потоком через разрез* (S, T).

Замечание

- Очевидно, что в любой сети существует минимальный разрез.

Но он может быть не единственным.

Лемма о потоке через разрез

Лемма 3

Пусть f — поток и (S, T) — разрез в сети $N = (V, s, t, c)$. Тогда $f(S, T) = |f|$.

Доказательство.

- Заметим, что $\sum\limits_{x,y \in S} f(x, y) = 0$, поскольку все ненулевые слагаемые этой суммы разбиваются на пары вида $f(x, y)$ и $f(y, x)$, которые сокращаются.
- Тогда

$$|f| = \sum_{y \in V} f(s, y) = \sum_{y \in V} f(s, y) + \sum_{x \in S \setminus \{s\}} \left(\sum_{y \in V} f(x, y) \right) = \sum_{x \in S, y \in V} f(x, y) =$$
$$= \sum_{x,y \in S} f(x, y) + \sum_{x \in S, y \in T} f(x, y) = \sum_{x \in S, y \in T} f(x, y) = f(S, T). \qquad \square$$

Замечание

В частности, $|f| = \sum\limits_{x \in V} f(x, t)$.

Следствие 2

Для любого потока f и разреза (S, T) в сети N выполнено $|f| \leq c(S, T)$.

Теорема Форда-Фалкерсона

Теорема 79 (L. R. Ford, D. R. Fulkerson, 1956)

Пусть f — поток в сети N. Тогда следующие три условия равносильны

1. *поток f максимален;*
2. *в остаточной сети N_f нет дополняющих путей;*
3. *существует такой разрез (S, T) в сети N, что $|f| = c(S, T)$.*

Доказательство.

"1. $\Rightarrow$ 2.": Предположим противное: пусть в N_f есть дополняющий путь p.

• Тогда по следствию 1 в сети N есть поток величины $|f| + c_f(p) > |f|$.

Противоречие.

"2. $\Rightarrow$ 3.": Пусть $S \stackrel{\text{def}}{=} \{x \in V \mid \text{в } N_f \text{ есть путь из } s \text{ в } x\}$ и $T = V \setminus S$.

• Очевидно, что пара (S, T) — разрез в N_f.

• Далее заметим, что $\forall x \in S\, \forall y \in T\, (c_f(x, y) = 0)$.

• Следовательно, $\forall x \in S\, \forall y \in T\, (f(x, y) = c(x, y))$, откуда $|f| = f(S, T) = c(S, T)$.

"3. $\Rightarrow$ 1.": Очевидно выводится из следствия 2. $\square$

Целочисленные сети (1/2)

Замечание

• Из теоремы Форда-Фалкерсона не следует существование максимального потока. Но из неё следует то, что если максимальный поток в сети существует, то его величина будет равна пропускной способности минимального разреза.

• Ниже мы рассмотрим практически значимый частный случай, при котором существование максимального потока прямо следует из теоремы Форда-Фалкерсона.

Определение 88

• Сеть $N = (V, s, t, c)$ называется *целочисленной*, если все значения функции c — целые неотрицательные числа (т. е. $c : V \times V \to \mathbb{N}_0$).

• Поток f в сети N называется *целочисленным*, если все значения функции f — целые числа (т. е. $f : V \times V \to \mathbb{Z}$).

Теорема 80

В любой целочисленной сети существует максимальный поток, причем среди максимальных потоков данной сети найдется целочисленный.

Целочисленные сети (2/2)

Доказательство.

Будем последовательно строить максимальный поток, начиная с нулевого потока.

- Пусть на очередном шаге мы имеем поток f.
- Найдем произвольный дополняющий путь p в остаточной сети N_f и заменим поток f на $f' = f + f_p$.
  - По доказанному выше, f' — поток в N и $|f'| = |f| + c_f(p)$.
  - Заметим, что если поток f был целочисленным, то и $c_f(p) \in \mathbb{N}$. Тогда поток f' также будет целочисленным.
- На каждом шаге величина потока увеличивается хотя бы на 1. Следовательно, рано или поздно процесс остановится и мы получим максимальный поток. □

Замечание

- Легко видеть, что доказательство теоремы 80 представляет из себя алгоритм поиска максимального потока в целочисленной сети. Время работы этого алгоритма можно оценить как $O(|E||f^*|)$, где f^* — максимальный поток в сети N. Действительно, дополняющий путь можно найти за $O(|E|)$ шагов, а искать его нужно не более $|f^*|$ раз.
- Такой алгоритм может оказаться неэффективным в том случае, когда $|f^*|$ велико. Ниже мы рассмотрим более совершенный алгоритм.

Алгоритм Эдмондса-Карпа (1/3)

- Вход: сеть $N = (V, s, t, c)$.
  - Последовательно строим максимальный поток, начиная с нулевого потока.
  - На каждом шаге алгоритма находим при помощи поиска в ширину кратчайший дополняющий путь и увеличиваем поток при помощи найденного пути.
  - Алгоритм завершает работу когда в остаточной сети не остается дополняющих путей.
- Выход: текущий поток на момент остановки алгоритма.

Лемма

Пусть p — кратчайший дополняющий путь в остаточной сети N_f; $f' = f + f_p$; p' — дополняющий путь в $N_{f'}$, которого нет в N_f. Тогда p' длиннее, чем p.

Доказательство.

Для каждой вершины $x \in V$ обозначим через $\delta_f(x)$ расстояние от s до x в сети N_f.

- Пусть $p = (x_0, x_1, \ldots, x_{n-1}, x_n)$, где $x_0 = s$ и $x_n = t$.
  - Очевидно, что тогда $\delta_f(x_i) = i$ при всех $i \in [0..n]$.
  - Также очевидно, что если $(x, y) \in E_f$, то $\delta_f(y) \leq \delta_f(x) + 1$.

Алгоритм Эдмондса-Карпа (2/3)

- Пусть $p' = (y_0, y_1, \ldots, y_{m-1}, y_m)$, где $y_0 = s$ и $y_m = t$.
  - По предположению, путь p' отсутствует в сети N_f. Следовательно, этот путь должен содержать дугу (y_k, y_{k+1}), такую, что $(y_k, y_{k+1}) \in E_{f'}$ и $(y_k, y_{k+1}) \notin E_f$.
  - Из построения потока f' очевидно, что такая дуга может быть только дугой вида $(x_{\ell+1}, x_\ell)$. То есть $\delta_f(y_{k+1}) - \delta_f(y_k) = -1$.
  - Для всех остальных дуг пути f' имеем $\delta_f(y_{i+1}) - \delta_f(y_i) \leq 1$.
- Тогда

$$n = \delta_f(t) - \delta_f(s) = \delta_f(y_m) - \delta_f(y_0) = \sum_{i=0}^{m-1}(-\delta_f(y_i) + \delta_f(y_{i+1})) =$$

$$= \sum_{i=0}^{k-1}(-\delta_f(y_i) + \delta_f(y_{i+1})) + (-\delta_f(y_k) + \delta_f(y_{k+1})) + \sum_{i=k+1}^{m-1}(-\delta_f(y_i) + \delta_f(y_{i+1})) \leq$$

$$\leq k + (-1) + (m - k - 1) = m - 2,$$

откуда $m \geq n + 2$. □

Алгоритм Эдмондса-Карпа (3/3)

Теорема 81

Время работы алгоритма Эдмондса-Карпа оценивается как $O(|V||E|^2)$.

Доказательство. Докажем, что в алгоритме Эдмондса-Карпа $O(|V||E|)$ шагов.

- Действительно, длина кратчайшего дополняющего пути находится в диапазоне $[1..|V|]$. По доказанной выше лемме, эта величина не убывает.
- На каждом шаге алгоритма мы удаляем из графа одну из дуг (а именно, дугу дополняющего пути, имеющую наименьшую пропускную способность). При этом, использовать дуги, которых ранее не было, мы сможем только после того, как увеличим длину кратчайшего дополняющего пути.
- Тем самым, длина кратчайшего дополняющего пути увеличится не более, чем через $|E|$ шагов. Таким образом, всего шагов будет $O(|V||E|)$.
- Каждый шаг можно выполнить за время $O(|E|)$. Итого получаем оценку времени работы алгоритма $O(|V||E|^2)$. □

Алгоритм Диница (1/4)

Определение 89

Пусть $N = (V, s, t, c)$ — сеть и f — поток в сети N.

- Дуга (x, y) называется *насыщенной* потоком f, если $c(x, y) = f(x, y)$.
- Поток f называется *блокирующим*, если любой путь из s в t в сети N содержит хотя бы одну насыщенную дугу.

Замечание

- Блокирующий поток не обязательно максимален. Однако в любом дополняющем пути для этого потока должна быть дуга, которой нет в исходной сети. Другими словами, для того, чтобы увеличить блокирующий поток, придется хотя бы один раз пройти по какой-либо дуге "в обратную сторону".

- Идея алгоритма Диница заключается в том, чтобы на каждом шаге строить в остаточной сети блокирующий поток, состоящий из нескольких дополняющих путей длины n, где n — длина кратчайшего дополняющего пути.

Алгоритм Диница (2/4)

- Вход: сеть $N = (V, s, t, c)$.
  - Последовательно строим максимальный поток, начиная с нулевого потока.
  - На каждом шаге алгоритма ищем в остаточной сети N_f блокирующий поток.
    - При помощи поиска в ширину вычисляем расстояния от s до всех вершин (как и раньше, обозначим расстояние от s до x через $\delta_f(x)$).
    - Строим вспомогательный граф $G' = (V, E')$, где $E' \stackrel{\text{def}}{=} \{(x, y) \in E_{N_f} \mid \delta_f(y) = \delta_f(x) + 1\}$.
    - При помощи поиска в глубину строим в графе G' путь p из s в t.
      - Если по ходу поиска в глубину мы пришли в *тупик* (вершину, из которой нет исходящих ребер), то возвращаясь из этой вершины назад мы удаляем из G' дугу, по которой мы в эту вершину пришли.
    - Рассматриваем максимальный поток вдоль пути p и удаляем из G' все дуги, насыщенные этим потоком.
    - Процесс продолжается до тех пор, пока в G' есть пути из s в t.
  - Прибавляем найденный блокирующий поток к найденному ранее потоку.
  - Алгоритм завершает работу когда в N_f нет дополняющих путей.
- Выход: текущий поток на момент остановки алгоритма.

Алгоритм Диница (3/4)

Теорема 82

Время работы алгоритма Диница оценивается как $O(|V|^2|E|)$.

Доказательство.

Докажем, что каждый шаг алгоритма Диница выполняется за время $O(|V||E|)$.

- Оценим время, необходимое на удаление из G' каждой отдельной дуги.
- Дуга (x, y) может быть удалена из графа G' в двух случаях.
  - В процессе поиска в глубину мы прошли по дуге (x, y) и оказались в тупике. В этом случае на удаление такой дуги потребуется время $O(1)$.
  - Был найден дополняющий путь p. Дуга (x, y) насыщена максимальным потоком вдоль пути p. Заметим, что нахождение пути p поиском в глубину требует время $O(|V|)$, не считая ходов, приводящих в тупик (такие ходы мы учли в предыдущем пункте).
- Итого, удаление каждой дуги потребует время $O(|V|)$. Поскольку в графе G' имеется $O(|E|)$ дуг, шаг алгоритма потребует время $O(|V||E|)$.
- Осталось заметить, что после каждого шага длина кратчайшего дополняющего пути увеличивается. Следовательно, будет $O(|V|)$ шагов. □

Алгоритм Диница (4/4)

Замечание

• Е. А. Диниц опубликовал исходную версию своего алгоритма в 1970 году. Позднее Ш. Эвен и А. Итаи внесли некоторые усовершенствования, в результате чего алгоритм приобрел нынешний вид.

• Можно доказать, что для 0-1 сетей (сетей, в которых все пропускные способности равны 0 или 1) алгоритм Диница будет работать за время $O(|E|^{3/2})$. Если, кроме этого, каждая вершина, за исключением s и t, имеет либо исходящую, либо входящую степень равную 1, то время работы алгоритма Диница можно оценить как $O(\sqrt{|V||E|})$.

• Есть и более быстрые алгоритмы решения задачи о максимальном потоке в общем случае. Наилучший результат на данный момент таков: $O\left(|V||E| \log\left(\frac{|V|^2}{|E|}\right)\right)$.

Теорема 83 (Е. А. Диниц, 1970)

В любой сети N есть максимальный поток.

Доказательство. Действуя как в алгоритме Эдмондса-Карпа (или как в алгоритме Диница), мы за конечное число шагов найдем максимальный поток. □

Максимальное паросочетание в двудольном графе (1/5)

Определение 90

Пусть $G = (V, E)$ — конечный неориентированный граф. Множество рёбер $M \subset E$ называется *паросочетанием*, если никакие два его ребра не инцидентны общей вершине.

- Нас будет интересовать задача о максимальном паросочетании: требуется в данном графе найти паросочетание, содержащее наибольшее возможное число ребер.
- В случае двудольного графа, эту задачу можно решить при помощи построения максимального потока в специальным образом выбранной сети.

- Пусть $G = (V, E)$ — двудольный граф; A и B — его доли (т. е. $V = A \cup B$, $A \cap B = \varnothing$ и каждое ребро соединяет вершину из A с вершиной из B).
- Построим орграф $G' = (V', E')$, где $V' = V \cup \{s, t\}$ (здесь $s, t \notin V$) и $E' = \{(s, x) \mid x \in A\} \cup \{(x, y) \mid x \in A, y \in B, (x, y) \in E\} \cup \{(y, t) \mid y \in B\}$.
- Далее, задав на всех дугах орграфа G' пропускные способности, равные 1, получим сеть $N' = (V', s, t, c)$.

Максимальное паросочетание в двудольном графе (2/5)

Алгоритм построения максимального паросочетания в двудольном графе

- Вход: двудольный граф $G = (V, E)$ с долями A и B.
- Строим орграф $G' = (V', E')$ и сеть $N' = (V', s, t, c)$, как это было показано на предыдущем слайде.
- Находим в сети N' максимальный целочисленный поток f.
- Выход: множество $M = \{(x, y) \in E \mid f(x, y) = 1\}$.

Лемма

- *Пусть $G = (V, E)$ — двудольный граф с долями A и B; $N' = (V', s, t, c)$ — соответствующая ему сеть; f — целочисленный поток в N'. Тогда множество $M = \{(x, y) \in E \mid f(x, y) = 1\}$ является паросочетанием в G и $|M| = |f|$.*
- *Обратно, если $M \subset E$ — паросочетание в G, то в N' есть такой целочисленный поток f, что $M = \{(x, y) \in E \mid f(x, y) = 1\}$ и $|M| = |f|$.*

Доказательство. Пусть f — целочисленный поток в N'. Поскольку пропускные способности всех дуг равны 1, функция f может принимать только три возможных значения: $-1, 0, 1$. При этом, $f(x, y) = -1$ возможно только при $(y, x) \in E'$.

Максимальное паросочетание в двудольном графе (3/5)

- Пусть $x \in A$. Тогда $\sum_{y \in V' \setminus \{s\}} f(x, y) = f(s, x) \leq 1$.
  - Заметим, что $f(x, y) \in \{0, 1\}$, при $y \in V' \setminus \{s\}$.
  - Следовательно, среди чисел $f(x, y)$, где $y \in V' \setminus \{s\}$, есть не более одного равного 1.
  - Таким образом, вершина x инцидентна не более, чем одному ребру из M.
- Аналогично доказывается, что любая вершина $y \in B$ также инцидентна не более, чем одному ребру из M. Следовательно, M — паросочетание в G.

- Обратно, пусть $M = \{(x_1, y_1), \ldots, (x_m, y_m)\}$ — паросочетание в G.
  - При этом, $x_1, \ldots, x_m \in A$ и $y_1, \ldots, y_m \in B$.
- Заметим, что при любом $i \in [1..m]$ путь $p_i = (s, x_i, y_i, t)$ является дополняющим в сети N' (с нулевым исходным потоком), причем эти пути не имеют общих рёбер.
- Сложив потоки вдоль этих путей, получим целочисленный поток f в сети G, для которого $M = \{(x, y) \in E \mid f(x, y) = 1\}$.

- Для доказательства того, что $|M| = |f|$, рассмотрим разрез (S, T), где $S = \{s\} \cup A$ и $T = B \cup \{t\}$. Тогда $|f| = f(S, T) = |M|$. □

Максимальное паросочетание в двудольном графе (4/5)

Теорема 84

Описанный выше алгоритм дает максимальное паросочетание в графе G.

Доказательство.

- По лемме, множество M, полученное на выходе алгоритма, является паросочетанием в графе G, причем $|M| = |f^*|$, где f^* — максимальный поток в сети N'.
- Докажем, что паросочетание M максимально.
- Предположим противное: пусть в графе G есть такое паросочетание M', что $|M'| > |M|$.
  - Тогда по лемме паросочетанию M' соответствует поток f' в сети N', для которого $|f'| = |M'| > |M| = |f^*|$.
  - Противоречие с максимальностью потока f^*.
- Таким образом, M — максимальное паросочетание в G. □

Максимальное паросочетание в двудольном графе (5/5)

Замечание

- Заметим, что $|f^*| = |M| \leq \frac{|V|}{2}$. Следовательно, даже использовав простейшую реализацию алгоритма Форда-Фалкерсона, мы получим время работы $O(|V||E|)$.
- Если искать максимальный поток при помощи алгоритма Диница, то время работы составит $O(\sqrt{|V|}|E|)$ (поскольку N' — это 0-1 сеть, в которой каждая вершина, кроме s и t, имеет либо исходящую, либо входящую степень равную 1).
  - На данный момент, это наилучший результат для задачи о максимальном паросочетании в двудольном графе.
  - Алгоритм поиска максимального паросочетания в двудольном графе с временем работы $O(\sqrt{|V|}|E|)$ впервые был предложен Дж. Хопкрофтом и Р. Карпом в 1973 году. Их алгоритм в целом аналогичен приведенному выше, но не использовал терминологию потоков в сетях.
- Задача о поиске максимального паросочетания в произвольном (не обязательно двудольном) графе также может быть решена за полиномиальное время (алгоритм Эдмондса).

4.2. Перечисление помеченных графов

- Напомним, что *помеченным графом* называется граф $G = (V, E)$, вершины которого занумерованы натуральными числами от 1 до p, где $p = |V|$.
  - Фактически, мы можем считать, что $V = [1..p]$.
- Сначала мы докажем несколько простейших фактов о числе помеченных графов.

1. Количество неориентированных графов
   - Обозначим через G_p количество помеченных неориентированных графов на p вершинах.
     - Тогда $G_p = 2^{\binom{p}{2}} = 2^{\frac{p(p-1)}{2}}$.
     - Действительно, есть $\binom{p}{2}$ пар вершин, каждая пара может быть либо соединена, либо не соединена ребром.
   - Пусть $m = \binom{p}{2}$ и $q \in [0..m]$. Тогда есть ровно $\binom{m}{q}$ помеченных графов на p вершинах с q рёбрами.
     - Действительно, есть m пар вершин, среди которых нужно выбрать q пар, которые будут соединены рёбрами.

Перечисление помеченных графов: орграфы и турниры (1/2)

2. Количество орграфов

• Обозначим через D_p количество помеченных орграфов на p вершинах. (Здесь мы считаем, что в орграфе не может быть кратных дуг, но могут быть *встречные* дуги. Другими словами, есть не более одной дуги из x в y, но могут быть одновременно дуги (x, y) и (y, x).)

- Тогда $D_p = 2^{p(p-1)}$.
- Действительно, есть $p(p-1)$ упорядоченных пар вершин, каждая пара может быть либо соединена, либо не соединена дугой.

• Пусть $q \in [0..p(p-1)]$. Тогда есть ровно $\binom{p(p-1)}{q}$ помеченных орграфов на p вершинах с q дугами.

- Действительно, есть $p(p-1)$ упорядоченных пар вершин, среди которых нужно выбрать q пар, которые будут соединены дугами.

• Напомним, что *турниром* называется орграф, в котором любые две вершины соединены ровно одной дугой (т. е. для любых x и y присутствует ровно одна из дуг (x, y) и (y, x)).

Перечисление помеченных графов: орграфы и турниры (2/2)

Утверждение

Количество помеченных турниров на p вершинах равно G_p.

Доказательство.

- Построим явную биекцию между неориентированными графами и турнирами с множеством вершин $[1..p]$.
- Пусть $G = (V, E)$ — неориентированный граф, где $V = [1..p]$.
- Построим соответствующий ему орграф $D = (V, A)$, где $A = \{(i, j) \mid ij \in E \text{ и } i < j\} \cup \{(i, j) \mid ij \notin E \text{ и } i > j\}$.
- По построению, для любых $i, j \in V$ в D есть ровно одна из дуг (i, j) и (j, i). Следовательно, D — турнир.
- Полученное соответствие — биекция, поскольку можно построить обратное отображение. А именно, турниру $D = (V, A)$ будет соответствовать граф $G = (V, E)$, где $E = \{ij \mid (i, j) \in A \text{ и } i < j\}$. □

Перечисление связных помеченных графов (1/2)

- Обозначим через C_p количество связных помеченных неориентированных графов на p вершинах.

Теорема 85

$$C_p = 2^{\binom{p}{2}} - \frac{1}{p}\sum_{k=1}^{p-1}\binom{p}{k}2^{\binom{p-k}{2}}kC_k = G_p - \frac{1}{p}\sum_{k=1}^{p-1}\binom{p}{k}G_{p-k}kC_k.$$

Доказательство. Будем рассматривать *корневые* помеченные графы, то есть помеченные графы вида $G = (V, E)$, в которых выделена вершина $x \in V$, называемая *корнем*.

- Очевидно, что есть pG_p корневых помеченных графов на p вершинах.
- Подсчитаем, в скольких из них корень находится в компоненте из k вершин.
  - $\binom{p}{k}$ способами выбираем k вершин для будущей компоненты;
  - k способами выбираем одну из k вершин, которая будет корнем;
  - C_k способами проводим ребра между выбранными k вершинами так, чтобы получился связный граф;
  - G_{p-k} способами проводим ребра между оставшимися вершинами.

Перечисление связных помеченных графов (2/2)

- Итого, получаем $\binom{p}{k} G_{p-k} k C_k$ корневых помеченных графов на p вершинах, в которых корень находится в компоненте связности из k вершин.
- Просуммировав по всем k от 1 до $p-1$, получим $\sum_{k=1}^{p-1} \binom{p}{k} G_{p-k} k C_k$ несвязных корневых помеченных графов на p вершинах.
- Вычитая это количество из pG_p и разделив на p, получим требуемую формулу.

□

Теорема 86

$C_p \sim G_p$ (*т. е.* $\frac{C_p}{G_p} \xrightarrow[p\to\infty]{} 1$).

Доказательство. Очевидно, что $\frac{C_p}{G_p} \le 1$. Оценим эту дробь снизу.

- $\frac{C_p}{G_p} = 1 - \frac{1}{p} \sum\limits_{k=1}^{p-1} k \binom{p}{k} \frac{G_{p-k} C_k}{G_p}$.
- $\frac{G_{p-k} C_k}{G_p} \le \frac{G_{p-k} G_k}{G_p} = \frac{2^{\frac{(p-k)(p-k-1)}{2}} \cdot 2^{\frac{k(k-1)}{2}}}{2^{\frac{p(p-1)}{2}}} = 2^{\frac{(p-k)(p-k-1)}{2} + \frac{k(k-1)}{2} - \frac{p(p-1)}{2}} =$

$$= 2^{\frac{(p-k)^2 - (p-k) + k^2 - k - p^2 + p}{2}} = 2^{\frac{(p-k)^2 + k^2 - p^2}{2}} = 2^{-k(p-k)}.$$

Перечисление связных помеченных графов (2/2)

- Итак, $\frac{C_p}{G_p} \geq 1 - \frac{1}{p}\sum\limits_{k=1}^{p-1} k\binom{p}{k}2^{-k(p-k)}$.

- Введем обозначение $F(p) = \frac{1}{p}\sum\limits_{k=1}^{p-1} k\binom{p}{k}2^{-k(p-k)}$. Нам нужно доказать, что $F(p) \xrightarrow[p\to\infty]{} 0$. Для этого оценим данную величину сверху.

- $F(p) = \frac{1}{p}\sum\limits_{k=1}^{p-1} k\binom{p}{k}2^{-k(p-k)} < \sum\limits_{k=1}^{p-1}\binom{p}{k}2^{-k(p-k)} \leq 2\sum\limits_{k=1}^{[p/2]}\binom{p}{k}2^{-k(p-k)} \leq$
$\leq 2\sum\limits_{k=1}^{[p/2]} p^k \cdot 2^{-k(p-k)} = 2\sum\limits_{k=1}^{[p/2]}\left(\frac{p}{2^{p-k}}\right)^k \leq 2\sum\limits_{k=1}^{[p/2]}\left(\frac{p}{2^{p/2}}\right)^k \leq 2\cdot\frac{\frac{p}{2^{p/2}}}{1-\frac{p}{2^{p/2}}} \xrightarrow[p\to\infty]{} 0.$

- Таким образом, $1 \geq \frac{C_p}{G_p} \geq 1 - F(p) \xrightarrow[p\to\infty]{} 1$, откуда по теореме о двух милиционерах $\frac{C_p}{G_p} \xrightarrow[p\to\infty]{} 1$. □

Замечание

Обозначим через B_p количество двусвязных помеченных неориентированных графов (блоков) на p вершинах. Можно доказать, что $B_p \sim C_p \sim G_p$. (б/д)

Перечисление деревьев (1/4)

Теорема 87 (A. Cayley, 1889)

Число помеченных деревьев на n вершинах равно n^{n-2}.

Доказательство (H. Prüfer, 1918).

- Пусть $V = [1..n]$. Мы построим взаимно однозначное соответствие между всеми деревьями на множестве V и всеми последовательностями длины $n-2$, в которых каждый член принимает натуральное значение от 1 до n.
  - Количество таких последовательностей равно в точности n^{n-2}.
- Пусть $T = (V, E)$ — дерево. Построим соответствующую ему последовательность $t_1, \ldots, t_{n-2}$.
- Пусть ℓ_1 — висячая вершина наименьшего номера в дереве T. Тогда
  - t_1 — единственная смежная с ℓ_1 вершина дерева T;
  - $T_1 = T - \ell_1$ (т. е. T_1 — дерево, полученное из T удалением вершины ℓ_1).
- Далее, найдём в T_1 висячую вершину наименьшего номера ℓ_2. Тогда
  - t_2 — единственная смежная с ℓ_2 вершина дерева T_1;
  - $T_2 = T_1 - \ell_2$.

Перечисление деревьев (2/4)

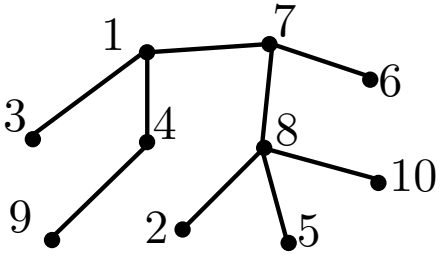
- И так далее, будем повторять процесс, пока не получим последовательность длины $n-2$ (при этом, останется дерево T_{n-2} на двух вершинах).

Определение 91

Построенная выше последовательность $t_1, \ldots, t_{n-2}$ называется *кодом Прюфера* дерева T.

Пример

Дерево на 10 вершинах и его код Прюфера.



8 1 8 7 4 1 7 8

Перечисление деревьев (3/4)

- Построим обратное соответствие.

Пусть дана последовательность $t_1, \ldots, t_{n-2}$ с элементами из $[1..n]$.

- Отметим, что по построению каждая вершина x встречается в последовательности дерева T ровно $d_T(x) - 1$ раз.
  - Следовательно, висячие вершины дерева T — это в точности вершины, которые не встречаются в последовательности $t_1, \ldots, t_{n-2}$.
- Выберем вершину ℓ_1 с наименьшим номером, не входящую в последовательность $t_1, \ldots, t_{n-2}$, и соединим её с t_1.
  - После этого, удалим ℓ_1 из списка номеров: $V_1 = V \setminus \{\ell_1\}$.
- Далее, выберем вершину $\ell_2 \in V_1$ с наименьшим номером, которая не встречается в последовательности $t_2, \ldots, t_{n-2}$ и соединим ℓ_2 с t_2.
  - Положим $V_2 = V_1 \setminus \{\ell_2\}$.
- И так далее. На k-м шаге мы имеем множество V_{k-1}, где $|V_{k-1}| = n - k + 1$, и последовательность $t_k, \ldots, t_{n-2}$, в которой $n - k - 1$ член. Следовательно, в V_{k-1} есть элементы, которых нет среди $t_k, \ldots, t_{n-2}$. Выберем наименьший такой элемент ℓ_k, соединим ℓ_k с t_k и положим $V_k = V_{k-1} \setminus \{\ell_k\}$.

Перечисление деревьев (4/4)

- Указанная выше операция повторяется $n-2$ раза. В результате будет использована вся последовательность и проведено $n-2$ ребра.
- На последнем шаге соединяем друг с другом две вершины множества V_{n-2}.
- Докажем, что граф, полученный описанным выше алгоритмом из любой последовательности $t_1, \ldots, t_{n-2}$, является деревом.
  - В этом графе n вершин и $n-1$ ребро. Нужно проверить, что в нем нет циклов.
  - Заметим, что на каждом шаге мы удаляем из множества V один из концов проведенного на этом шаге ребра. Следовательно, множество V_k содержит ровно по одной вершине из каждой компоненты связности имеющегося в данный момент графа. Но тогда на каждом шаге мы соединяем две вершины из разных компонент, следовательно, циклов не образуется.
- Осталось заметить, что если $t_1, \ldots, t_{n-2}$ — код Прюфера дерева T, то из $t_1, \ldots, t_{n-2}$ мы получим именно дерево T. На всех шагах, кроме последнего, мы проводим в точности то ребро, которое удаляли на соответствующем шаге построения кода Прюфера. А на последнем шаге мы соединяем две вершины из V_{n-2}, степени которых в имеющемся графе равны количеству вхождений этих вершин в $t_1, \ldots, t_{n-2}$, то есть на 1 меньше, чем их степени в дереве T. □

4.3. Перечисление графов с точностью до изоморфизма

- Напомним, что *изоморфизмом* графов $G_1 = (V_1, E_1)$ и $G_2 = (V_2, E_2)$ называется биекция $\varphi\colon V_1 \to V_2$, удовлетворяющая условию $\forall x, y \in V_1\,(xy \in E_1 \leftrightarrow \varphi(x)\varphi(y) \in E_2)$.
  - Сами графы G_1 и G_2 в этом случае называют *изоморфными*. Обозначение: $G_1 \cong G_2$.
- По сути, если мы стираем пометки на вершинах графа, то мы перестаем различать изоморфные друг другу графы. Тогда возникает вопрос о количестве графов *с точностью до изоморфизма*.
  - Легко видеть, что изоморфность двух помеченных графов — это отношение эквивалентности. А интересующее нас количество графов с точностью до изоморфизма — это число классов эквивалентности.

Пример

Есть $2^6 = 64$ помеченных графов на 4 вершинах, но всего 9 попарно неизоморфных графов на 4 вершинах.

Расстановка пометок и автоморфизмы графа

- Посмотрим на этот вопрос с другой стороны. Сколько есть способов расставить пометки на вершинах данного непомеченного графа?
- Другими словами, сколько помеченных графов входят в данный класс эквивалентности?
  - Количество классов эквивалентности было бы легко посчитать, если бы все классы содержали одинаковое число элементов. Однако, это, увы, не так.
  - Например, очевидно, что полный граф является единственным элементом своего класса эквивалентности. Но есть $n(n-1)/2$ помеченных графов на n вершинах ровно с одним ребром — и все они изоморфны.
- Всего есть $n!$ способов расставить пометки на данных n вершинах. Но некоторые из этих способов могут давать один и тот же помеченный граф.
  - То есть граф может оказаться изоморфен сам себе.
- Напомним, что *автоморфизмом* графа $G = (V, E)$ называется изоморфизм из G в G (т. е. такая биекция $f\colon V \to V$, что $\forall x, y \in V\,(xy \in E \leftrightarrow f(x)f(y) \in E)$).

Группа автоморфизмов графа

- Как уже отмечалось выше, множество всех автоморфизмов графа G образует группу относительно операции композиции. Эта группа называется *группой автоморфизмов* графа G и обозначается $\mathrm{Aut}(G)$.
  - Из определения видно, что автоморфизм графа — это перестановка на множестве его вершин, сохраняющая отношение смежности.
  - Пусть вершины графа $G = (V, E)$, где $V = [1..n]$.
    Тогда легко видеть, что $\mathrm{Aut}(G) < S_n$.

Утверждение

1. *Если* $G_1 \cong G_2$, *то* $\mathrm{Aut}(G_1) \cong \mathrm{Aut}(G_2)$;
2. *для любого графа* G *выполнено* $\mathrm{Aut}(G) \cong \mathrm{Aut}(\overline{G})$.

Замечание

- То есть группы автоморфизмов изоморфных графов всегда изоморфны.
- Но обратное наверное. Например, легко построить граф, не изоморфный своему дополнению. У этих графов группы автоморфизмов будут изоморфны, а сами графы — нет.

Действие группы на множестве (1/4)

Определение 92

Пусть A — группа и X — множество. *Действием* группы A на множестве X называется отображение $\cdot\colon A \times X \to X$, удовлетворяющее следующим свойствам:

1. $\forall x \in X\,(ex = x)$;
2. $\forall \alpha, \beta \in A \forall x \in X\,((\alpha\beta)x = \alpha(\beta x))$.

Замечание

- Отметим, что каждому элементу $\alpha \in A$ соответствует следующее отображение из X в X: $\begin{array}{ccc} \alpha\colon & X & \to & X \\ & \cup\!| & & \cup\!| \\ & x & \mapsto & \alpha x \end{array}$.
- Нейтральному элементу соответствует тождественное отображение $(ex = x)$;
- Произведению $\alpha\beta$ соответствует композиция отображений $((\alpha\beta)x = \alpha(\beta x))$;
- Элементу α^{-1} соответствует отображение, обратное к α.
  - В частности, это означает, что каждому элементу группы A соответствует биекция из X в X, то есть перестановка на множестве X.
- Если $X = [1..n]$, то фактически мы получаем гомоморфизм $F\colon A \to S_n$.

Действие группы на множестве (2/4)

Примеры

1. Группа S_n действует на множестве $[1..n]$;
2. Пусть $\mathcal{G}_n$ — множество всех помеченных графов на множестве $V = [1..n]$. Тогда мы можем задать следующее действие группы S_n на множестве $\mathcal{G}_n$:
   - паре (σ, H), где $\sigma \in S_n$ и $H = (V, E) \in \mathcal{G}_n$, будет соответствовать граф $\sigma H = (V, \sigma E)$, где $\sigma E \stackrel{\text{def}}{=} \{\sigma(x)\sigma(y) \mid xy \in E\}$.

Определение 93

Пусть группа A действует на множестве X и $x \in X$. Тогда

- $\langle x \rangle \stackrel{\text{def}}{=} \{\alpha x \mid \alpha \in A\}$ — *орбита* элемента x;
- $\mathrm{St}(x) \stackrel{\text{def}}{=} \{\alpha \in A \mid \alpha x = x\}$ — *стабилизатор* элемента x.

Действие группы на множестве (3/4)

Замечание

• Назовем элементы $x, y \in X$ *подобными*, если $\exists \alpha \in A\,(y = \alpha x)$. Легко видеть, что подобие — это отношение эквивалентности, а орбита $\langle x \rangle$ — это класс эквивалентности, содержащий x. Тем самым, множество X разбивается на непересекающиеся орбиты.

• В частности, в примере 2 подобными будут изоморфные графы.

Лемма

Пусть группа A действует на множестве X и $x \in X$. Тогда $\mathrm{St}(x) < A$.

Доказательство. Проверим, что выполняются три свойства из теоремы 47.

1. Пусть $\alpha, \beta \in \mathrm{St}(x)$. Тогда $(\alpha\beta)x = \alpha(\beta x) = \alpha x = x$, следовательно, $\alpha\beta \in \mathrm{St}(x)$.
2. По определению $ex = x$, следовательно, $e \in \mathrm{St}(x)$.
3. Пусть $\alpha \in \mathrm{St}(x)$. Тогда $\alpha^{-1}x = \alpha^{-1}(\alpha x) = (\alpha^{-1}\alpha)x = ex = x$, следовательно, $\alpha^{-1} \in \mathrm{St}(x)$.

□

Действие группы на множестве (4/4)

Теорема 88

Пусть группа A действует на множестве X и $x \in X$. Тогда $|\langle x\rangle| \cdot |\operatorname{St}(x)| = |A|$.

Доказательство. Пусть $y \in \langle x\rangle$ и $y = \alpha x$.

- Рассмотрим множество $A_y \stackrel{\text{def}}{=} \{\beta \in A \mid \beta x = y\}$.
- Докажем, что $A_y = \alpha \operatorname{St}(x) = \{\alpha\gamma \mid \gamma \in \operatorname{St}(x)\}$.
  - Действительно, если $\gamma \in \operatorname{St}(x)$, то $(\alpha\gamma)x = \alpha(\gamma x) = \alpha x = y$, следовательно, $\alpha\gamma \in \operatorname{St}(x)$.
  - Обратно, если $\beta \in A_y$, то $x = \alpha^{-1}y = \alpha^{-1}(\beta x) = (\alpha^{-1}\beta)x$, следовательно, $\alpha^{-1}\beta \in \operatorname{St}(x)$. Тогда обозначив $\alpha^{-1}\beta$ за γ получим, что $\beta = \alpha\gamma$, где $\gamma \in \operatorname{St}(x)$.
- Тем самым, все множества A_y — это в точности левые смежные классы подгруппы $\operatorname{St}(x)$ группы A.
- Тогда $|\langle x\rangle| = (A : \operatorname{St}(x))$, откуда по теореме Лагранжа получаем, что

$$|A| = |\operatorname{St}(x)| \cdot (A : \operatorname{St}(x)) = |\operatorname{St}(x)| \cdot |\langle x\rangle|.$$

□

Группа автоморфизмов графа и число способов расставить пометки

Лемма

Пусть $G = (V, E)$ — помеченный граф и $V = [1..n]$. Тогда существует ровно $\dfrac{n!}{|\operatorname{Aut}(G)|}$ помеченных графов на множестве V, изоморфных G.

Доказательство. Как и ранее, обозначим через $\mathcal{G}_n$ множество всех помеченных графов на множестве $[1..n]$.

- Рассмотрим действие на этом множестве, описанное в примере 2: паре (σ, H), где $\sigma \in S_n$ и $H = (V, E) \in \mathcal{G}_n$, будет соответствовать граф $\sigma H = (V, \sigma E)$, где $\sigma E \stackrel{\text{def}}{=} \{\sigma(x)\sigma(y) \mid xy \in E\}$.
- Тогда
  - $\langle G \rangle$ — множество всех графов на множестве $[1..n]$, изоморфных G;
  - $\operatorname{St}(G) = \operatorname{Aut}(G)$.
- Следовательно, по теореме 88 получаем, что $|\langle G \rangle| = \dfrac{|S_n|}{|\operatorname{St}(G)|} = \dfrac{n!}{|\operatorname{Aut}(G)|}$.

□

Лемма Бернсайда (1/2)

Определение 94

Пусть задано действие группы A на множестве X. Тогда для любого $\alpha \in A$

- $\mathrm{Fix}(\alpha) \stackrel{\mathrm{def}}{=} \{x \in X \mid \alpha x = x\}$ — *множество неподвижных точек* элемента α;
- элементы множества $\mathrm{Fix}(\alpha)$ — *неподвижные точки* элемента α.

Утверждение

$$\sum_{\alpha \in A} |\mathrm{Fix}(\alpha)| = \sum_{x \in X} |\mathrm{St}(x)|.$$

Доказательство. Обе части равны $|\{(\alpha, x) \in A \times X \mid \alpha x = x\}|$. □

Теорема 89 (Лемма Бернсайда)

Количество орбит действия группы A на множестве X равно $\frac{1}{|A|} \sum_{\alpha \in A} |\mathrm{Fix}(\alpha)|$.

Доказательство. Присвоим каждому элементу $x \in X$ вес $w(x) \stackrel{\mathrm{def}}{=} \frac{1}{|\langle x \rangle|}$.

- Тогда сумма весов элементов любой орбиты равна 1.

Лемма Бернсайда (2/2)

- Следовательно, сумма весов всех элементов множества X равна количеству орбит (обозначим его N).
- Тогда

$$N = \sum_{x \in X} \frac{1}{|\langle x \rangle|} = \sum_{x \in X} \frac{|\operatorname{St}(x)|}{|\langle x \rangle||\operatorname{St}(x)|} = \frac{1}{|A|} \sum_{x \in X} |\operatorname{St}(x)| = \frac{1}{|A|} \sum_{\alpha \in A} |\operatorname{Fix}(\alpha)|.$$

□

Замечание

Доказанную выше теорему обычно называют леммой Бернсайда. Но она была известна и ранее. Сам William Burnside в своей книге “Theory of Groups of Finite Order” 1897 года называл первооткрывателем этой леммы Фробениуса. Но судя по всему, это утверждение было известно еще раньше.

Асимптотика числа графов с точностью до изоморфизма (1/5)

- Обозначим через g_p количество графов на n вершинах с точностью до изоморфизма.
  - Напомним, что G_n — это количество помеченных графов на n вершинах.
  - Мы уже знаем, что $G_n = 2^{\frac{n(n-1)}{2}}$.
- Оказывается, что g_n примерно в $n!$ раз меньше, чем G_n.
  - Неформально это означает, что почти у всех графов группа автоморфизмов тривиальна (т. е. состоит из единственного элемента: тождественного преобразования).

Теорема 90

$g_n \sim \frac{G_n}{n!} = \frac{2^{\frac{n(n-1)}{2}}}{n!}$.

Доказательство. Пусть $\mathcal{G}_n$ — множество всех помеченных графов на множестве вершин $V = [1..n]$.

- Как и ранее, рассмотрим следующее действие группы S_n на множестве $\mathcal{G}_n$:
  - паре (σ, H), где $\sigma \in S_n$ и $H = (V, E) \in \mathcal{G}_n$, будет соответствовать граф $\sigma H = (V, \sigma E)$, где $\sigma E \stackrel{\text{def}}{=} \{\sigma(x)\sigma(y) \mid xy \in E\}$.

Асимптотика числа графов с точностью до изоморфизма (2/5)

- Нам нужно посчитать число неподвижных точек для перестановки $\sigma \in S_n$.
- Для этого рассмотрим множество $V^{(2)}$ двухэлементных подмножеств множества V.
  - Другими словами, $V^{(2)}$ — это множество ребер полного графа K_n на множестве вершин V.
- Заметим, что группа S_n действует также и на множестве $V^{(2)}$: $\sigma \cdot xy \stackrel{\text{def}}{=} \sigma(x)\sigma(y)$. Тем самым, каждая перестановка $\sigma \in S_n$ индуцирует перестановку $\sigma' \in S(V^{(2)})$, а группа S_n индуцирует подгруппу $S_n^{(2)} < S(V^{(2)})$, состоящую из всех перестановок множества $V^{(2)}$ вида σ'.
  - Группа $S_n^{(2)}$ называется *парной группой* группы S_n.
  - Фактически, мы построили гомоморфизм групп $S_n \to S(V^{(2)})$. Группа $S_n^{(2)}$ — это образ данного гомоморфизма.
  - Нетрудно проверить, что при $n > 2$ группы S_n и $S_n^{(2)}$ изоморфны.
- Для перестановки $\sigma \in S_n$ нас будут интересовать циклы соответствующей ей перестановки $\sigma' \in S_n^{(2)}$. Эти циклы мы будем называть *рёберными циклами* перестановки σ.

Асимптотика числа графов с точностью до изоморфизма (3/5)

- Заметим, что граф $G \in \mathcal{G}_n$ является неподвижной точкой для перестановки $\sigma \in S_n$, если и только если для любого рёберного цикла C перестановки σ либо $C \subset E(G)$, либо $C \cap E(G) = \varnothing$.
- Тем самым, $|\operatorname{Fix}(\sigma)| = 2^{q(\sigma)}$, где $q(\sigma)$ — число рёберных циклов перестановки σ.
- Тогда по лемме Бернсайда, $g_n = \frac{1}{n!} \sum\limits_{\sigma \in S_n} 2^{q(\sigma)}$.
- Обозначим через $S_{n.k}$ множество перестановок из S_n, имеющих ровно $n - k$ неподвижных точек.
- Пусть $g_n^{(k)} = \frac{1}{n!} \sum\limits_{\sigma \in S_{n,k}} 2^{q(\sigma)}$. Тогда $g_n = \sum\limits_{k=0}^{n} g_n^{(k)}$.
  - Очевидно, что $g_n^{(0)} = \frac{1}{n!} 2^{\frac{n(n-1)}{2}}$.
  - То есть нам нужно доказать, что $g_n \sim g_n^{(0)}$.

Асимптотика числа графов с точностью до изоморфизма (4/5)

Лемма

Если $\sigma \in S_{n,k}$, *то* $q(\sigma) \leq \binom{n}{2} + \frac{1}{2}(k - nk + \frac{k^2}{2})$.

Доказательство. Пусть перестановка σ имеет t рёберных циклов длины 1.

- Тогда оставшиеся $\frac{n(n-1)}{2} - t$ пар вершин разбиты на рёберные циклы длины хотя бы 2.
- Следовательно, рёберных циклов длины хотя бы 2 не более $\frac{n(n-1)}{4} - \frac{t}{2}$.
- Это означает, что $q(\sigma) \leq \frac{n(n-1)}{4} - \frac{t}{2} + t = \frac{n(n-1)}{4} + \frac{t}{2}$.
- Осталось заметить, что рёберными циклами длины 1 перестановки σ могут быть лишь
  - пары из двух неподвижных точек перестановки σ (их $\frac{(n-k)(n-k-1)}{2}$);
  - пары вершин, образующих цикл длины 2 перестановки σ (их не более, чем $\frac{k}{2}$).
- Итого, $t \leq \frac{(n-k)(n-k-1)}{2} + \frac{k}{2} = \frac{n^2-2nk+k^2-n+2k}{2} = \frac{n(n-1)}{2} + (k - nk + \frac{k^2}{2})$.
- Таким образом, $q(\sigma) \leq \frac{n(n-1)}{4} + \frac{t}{2} \leq \frac{n(n-1)}{2} + \frac{1}{2}(k - nk + \frac{k^2}{2})$. □

Асимптотика числа графов с точностью до изоморфизма (5/5)

- Заметим, что $|S_{n,k}| \leq \binom{n}{k} \cdot k! = \frac{n!}{(n-k)!} \leq n^k$.
- Тогда $g_n^{(k)} \leq \frac{1}{n!}|S_{n,k}|2^{\binom{n}{2}+\frac{1}{2}(k-nk+\frac{k^2}{2})} \leq g_n^{(0)} n^k 2^{\frac{k}{2}(1-n+\frac{k}{2})} \leq$

$$\leq g_n^{(0)} \left(\frac{n}{2^{\frac{1}{2}(n-1-\frac{k}{2})}}\right)^k \leq g_n^{(0)} \left(\frac{n}{2^{\frac{n-2}{4}}}\right)^k = g_n^{(0)} \left(\frac{n\sqrt{2}}{2^{\frac{n}{4}}}\right)^k.$$

- Следовательно, $1 \leq \frac{g_n}{g_n^{(0)}} \leq \sum\limits_{k=0}^{n} \left(\frac{n\sqrt{2}}{2^{\frac{n}{4}}}\right)^k \leq \frac{1}{1-\frac{n\sqrt{2}}{2^{\frac{n}{4}}}} \xrightarrow[n\to\infty]{} 1.$
- Тогда по теореме о двух милиционерах $\frac{g_n}{g_n^{(0)}} \xrightarrow[n\to\infty]{} 1$,

а это и означает, что

$$g_n \sim g_n^{(0)} = \frac{G_n}{n!} = \frac{2^{\frac{n(n-1)}{2}}}{n!}.$$

□