Министерство образования и науки
Российской Федерации
Федеральное государственное бюджетное образовательное
учреждение высшего образования
“Оренбургский государственный университет”

С. А. Пихтильков, О.А. Пихтилькова, Л.Б. Усова

ФУНДАМЕНТАЛЬНАЯ И КОМПЬЮТЕРНАЯ АЛГЕБРА

Рекомендовано Ученым советом
федерального государственного бюджетного образовательного
учреждения высшего образования
“Оренбургский государственный университет”
в качестве учебного пособия для студентов,
обучающихся по программе высшего
образования по направлению подготовки
02.03.01 Математика и компьютерные науки

Оренбург
2016

УДК 512.5(075.8)
ББК 22.14я73
П35

Рецензент – кандидат физико-математических наук, доцент кафедры компьютерной безопасности и математического обеспечения информационных систем Т.М. Отрыванкина

Пихтильков С.А.

П35 Фундаментальная и компьютерная алгебра:
учебное пособие для студентов физико-математических специальностей вузов/ С. А. Пихтильков, О.А. Пихтилькова, Усова Л.Б.; Оренбургский гос. ун-т.- Оренбург: ОГУ, 2016.- 116 с.

Цель данного пособия познакомить студентов с алгоритмами, лежащими в основе символьных вычислений, кодами, исправляющими ошибки и элементами криптографии. В пособии также приводятся основные сведения из фундаментальной алгебры.

УДК 512.5(075.8)
ББК 22.14я73

Содержание

Введение

Дисциплина "Фундаментальная и компьютерная алгебра" читается для студентов, обучающихся по направлению подготовки 02.03.01 Математика и компьютерные науки с 2011 года.

Появление данного курса связано с тем, что в последнее время компьютерная алгебра - активно развивающаяся область, лежащая на стыке математики и информатики. В отличие от численного анализа она сообщает результат в виде аналитического выражения. Само понятие "компьютерная алгебра" появилось в связи с разработкой и применением систем аналитических вычислений, в которых используются сложные алгоритмы. Она требует наличия знаний в различных областях математики, и, в первую очередь, в области абстрактной алгебры.

При разработке пособия преследовались цели: познакомить студентов с алгоритмами, лежащими в основе символьных вычислений, кодами, исправляющими ошибки, и элементами криптографии.

Реализация этих целей позволит студенту-бакалавру:

- сформировать необходимый терминологический запас, необходимый для самостоятельного изучения специальной математической литературы по программированию;

- овладеть методами решения практических задач и понять принцип построения систем компьютерной алгебры;

- уменьшить трудозатраты на элементарные вычисления;

- приобрести навыки конструирования алгоритмов;

- понять важность основных алгебраических структур при построении современного программного обеспечения.

Предполагается знание основных понятий линейной алгебры.

При составлении учебного пособия использовались источники [1], [4], [7] и другие.

Пособие предназначено для студентов высших учебных заведений, обучающихся по программе высшего образования по направлению подготовки 02.03.01 Математика и компьютерные науки.

1 Группы

1.1 Основные определения и понятия

Определение. *Группой называется множество G на котором задана алгебраическая операция "$*$" (т. е. для любых двум элементов $a, b \in G$ определен элемент $a * b \in G$), удовлетворяющая следующим аксиомам:*

(1) операция $*$ - ассоциативна, т. е. для всех $a, b, c \in G$ выполнено $a * (b * c) = (a * b) * c$;

(2) существует нейтральный элемент e, такой, что для всех $a \in G$ выполнено $a * e = e * a = a$;

(3) для каждого элемента $a \in G$ существует симметричный элемент $a' \in G$ такой, что $a * a' = a' * a = e$.

Группу G с операцией "$*$" обозначают $(G, *)$.

Справедлива следующая теорема.

Теорема 1.1.1 *Пусть G с операцией $*$ – группа. Тогда*

1. Нейтральный элемент в группе G только один.

2. Для любого элемента $a \in G$ существует единственный симметричный элемент $a' \in G$.

Доказательство. 1. Пусть e_1, e_2 – нейтральные элементы группы G. Тогда $e_1 * e_2 = e_1 = e_2$.

2. Пусть b_1, b_2 -симметричные для a элементы. Тогда

$$(b_1 * a) * b_2 = e * e_2 = e_2 = b_1 * (a * b_2) = e_1 * e = e_1.$$

Следует единственность симметричного элемента. □

Определение. *Если алгебраическая операция обозначается "$\cdot$", то группа называется мультипликативной, нейтральный по умножению элемент называется единицей, симметричный элемент называется обратным и обозначается a^{-1}.*

Определение. *Группа $(G, *)$ называется коммутативной или абелевой, если операция $*$ – коммутативна, т. е. $a * b = b * a$ для всех $a, b \in G$.*

Для абелевой группы обычно используют аддитивную запись: операция обозначают "$+$", нейтральный элемент называют нулем и обозначают 0, а обратный – противоположным и обозначается $-a$.

Пример. Абелевыми группами являются $(\mathbb{Z}, +)$, $(\mathbb{Q}, +)$, $(\mathbb{R}, +)$.

Здесь $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ обозначают соответственно множество целых, рациональных и действительных чисел.

Множество $(\mathbb{R}, \cdot)$ группой не является, так как для элемента 0 не существует обратного.

Определение. *Число элементов группы $(G, *)$ называется порядком группы и обозначается $\mid G \mid$. Если число элементов группы конечно и равно n, то пишут $\mid G \mid = n$ или $\mid G \mid < \infty$.*

Определение. *Непустое подмножество H группы $(G, *)$ называется подгруппой, если оно замкнуто относительно операции " $*$ " и $(H, *)$ само является группой.*

Справедлива следующая теорема.

Теорема 1.1.2 *Непустое подмножество H группы $(G, *)$ является подгруппой тогда и только тогда, когда выполнены следующие условия:*

1. если $a, b \in H$, то $a * b \in H$;

2. если $a \in H$, то $a^{-1} \in H$.

Ясно, что у каждой группы $(G, *)$ есть по крайней мере две подгруппы – G и $\{e\}$. Подгруппа H группы $(G, *)$ называется *собственной*, если она отлична от G и $\{e\}$.

Пример. Для каждого $m \in \mathbb{Z}$ множество $m\mathbb{Z} = \{mk | k \in \mathbb{Z}\}$ является подгруппой в $(\mathbb{Z}, +)$.

Конечную группу $(G, *)$ можно задать с помощью таблицы, называемой *таблицей Кэли*[1]

	g_1	g_2	$\dots$	g_n
g_1	$g_1 * g_1$	$g_1 * g_2$	$\dots$	$g_1 * g_n$
g_2	$g_2 * g_1$	$g_2 * g_2$	$\dots$	$g_2 * g_n$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
g_n	$g_n * g_1$	$g_n * g_2$	$\dots$	$g_n * g_n$

Напомним некоторые определения, относящиеся к функциям.

Определение. *1. Функция $f : A \to B$ отображающая множество A в множество B называется инъективной, если*

$$f(a_1) = f(a_2) \Rightarrow a_1 = a_2, \text{ где } a_1, a_2 \in A.$$

2. Функция $f : A \to B$ отображающая множество A в множество B называется сюръективной, если $\mathrm{Im}\ f = B$, *где*

$$\mathrm{Im}\ f = \{y \mid y \in B\ \&\ (\exists x \in A\ )y = f(x)\} -$$

множество значений функции (образ множества A).

3. Функция называется биективной, если она инъективна и сюръективна одновременно.

[1] Артур Кэли – английский математик (1821-1895).

Приведем пример некоммутативной группы, которая играет особую роль в теории конечных групп.

Определение. *Подстановкой непустого подмножества X называется любое биективное отображение множества X на себя.*

Множество всех подстановок множества X обозначим через $S(X)$.

Если $g_1,\ g_2 \in S(X)$, то можно определить их произведение, которое обозначается $g_1 \cdot g_2$ или $g_1 g_2$ и определяется следующим образом: $(g_1 g_2)(x) = g_2(g_1(x))$. Легко проверить, $(S(X), \cdot)$ является группой, называемой *симметрической группой* множества X.

Если множество $X = \{1, \ldots, n\}$ конечно, то группу $S(X)$ называют *симметрической группой подстановок степени n* и обозначают S_n. Подстановки из S_n записываются обычно в виде таблицы из двух строк:

$$g = \begin{pmatrix} 1 & 2 & \ldots & n \\ i_1 & i_2 & \ldots & i_n \end{pmatrix},$$

где i_k – образ элемента k при действии подстановки g.

Определение. *Пусть $(G_1, *)$ и $(G_2, \circ)$ – группы. Отображение $f : G_1 \to G_2$ называется гомоморфизмом групп, если $f(a * b) = f(a) \circ f(b)$ для любых $a, b \in G_1$.*

Легко проверить, что если f – гомоморфизм, то $f(e_{G_1}) = e_{G_2}$ и $f(a^{-1}) = (f(a))^{-1}$.

Определение. *Ядром гомоморфизма f называется множество*

$$\mathrm{Ker} f = \{a \in G_1 \mid f(a) = e_{G_2}\}.$$

Определение. *Гомоморфизм $f : G_1 \to G_2$ называется изоморфизмом, если функция f является биективной.*

Следующая теорема проясняет особое положение группы подстановок в теории групп.

Теорема 1.1.3 (Кэли) *Любая конечная группа G порядка n изоморфна некоторой подгруппе симметрической группы S_n.*

Упражнения

1. Показать, что непустое множество G с ассоциативной алгебраической операцией, удовлетворяющей условиям:

   a) $(\exists e \in G)(\forall a \in G)ae = e$;

   b) $(\forall a \in G)(\exists b \in G)ab = e$ – является группой.

2. Пусть H – мультипликативная подгруппа, все элементы a которой удовлетворяют условию $a^2 = e$. Доказать, что группа H является коммутативной.

3. Показать, что множество нечетных чисел $N = \{mk+1 | k \in \mathbb{Z}\}$ подгруппой в $(\mathbb{Z}, +)$ не является.

4. Показать, что любая подгруппа в $(\mathbb{Z}, +)$ имеет вид $m\mathbb{Z}, m \in \mathbb{Z}$.

5. Пусть K и H – две подгруппы группы $(\mathbb{Z}, +)$. Показать, что множество $K + H = \{k + h \mid k \in K, h \in H\}$ является группой.

6. Пусть a и b – произвольные целые числа, d – их наибольший общий делитель (точное определение см. в разделе 4.2). Используя формулировки упражнений 4 и 5, доказать утверждение о линейном представлении *НОД*, т. е. о существовании целых x и y таких, что $xa + yb = d$.

7. Доказать, что Imf и $Kerf$ являются подгруппами.

8. Доказать, что гомоморфизм групп $f : G_1 \to G_2$ является изоморфизмом, тогда и только тогда, когда $Imf = G_2$ и $Kerf = \{e_{G_1}\}$.

1.2 Разложение группы на смежные классы. Теорема Лагранжа

Определение. *Пусть H – подгруппа в G. Множества вида $gH = \{gh \mid h \in H\}$ (при фиксированном $g \in G$) называются левыми смежными классами группы G по подгруппе H. Аналогично, множества вида $Hg = \{hg \mid h \in H\}$ называются правыми смежными классами по подгруппе H.*

Заметим, что одним из смежных классов является сама подгруппа $H = eH = He$.

Предложение 1.2.1 *Левые смежные классы обладают следующими свойствами:*

1. Каждый смежный класс определяется любым своим элементом, т. е. если $g' \in gH$, то $g'H = gH$;

2. Два смежных класса либо не пересекаются, либо совпадают;

3. Объединение всех смежных классов равно G;

4. Отображение $H \to gH$, при котором $h \to gh$, биективно. Следовательно, все смежные классы по подгруппе – равномощны.

Доказательство. 1. Пусть $g' \in gH$ т. е. $g' = gh, \quad h \in H$. Тогда $g'H = ghH \subseteq gH$. С другой стороны, так как $g = g'h^{-1}$, то $g \in g'H$, и, следовательно, $gH \subseteq g'H$. Таким образом, $g'H = gH$.

2. Если смежные классы g_1H и g_2H пересекаются, то существует элемент $g \in g_1H \bigcap g_2H$. Из 1 получим, что $g_1H = gH = g_2H$.

3. Так как $g \in gH$, то $G = \bigcup_{g \in G} gH$.

4. Отображение $H \to gH$ сюръективно по определению левого смежного класса, но оно и инъективно, так как из того, что $gh_1 = gh_2$, следует, что $h_1 = h_2$. □

Аналогичными свойствами обладают и правые смежные классы.

Предложение 1.2.2 *Отображение $g \to g^{-1}$ группы G на себя задает биективное соответствие между множествами левых смежных классов и правых смежных классов группы G по подгруппе H.*

Доказательство следует из цепочки равенств

$$(gH)^{-1} = \{(gh)^{-1} \mid h \in H\} = \{h^{-1}g^{-1} \mid h \in H\} =$$

$$= \{hg^{-1} \mid h \in H\} = Hg^{-1}.$$

□

Определение. *Число (левых или правых) смежных классов группы G по подгруппе H называется индексом подгруппы H в G и обозначается через $(G : H)$.*

Из разложения группы на левые (или правые) смежные классы по подгруппе в силу предложения 1.2.1. получаем:

Теорема 1.2.1 (теорема Лагранжа) [1] *Порядок подгруппы H конечной группы G делит порядок группы и $\mid G \mid = (G : H) \mid H \mid$.*

Следствие 1.2.1 *Пусть G – конечная группа порядка n. Тогда $a^n = e$ для любого $a \in G$.*

[1] Жозеф Луи Лагранж – французский математик (1736-1813).

Упражнения

1. Доказать следствие 1.2.1.

2. Найти число смежных классов группы $(\mathbb{Z}, +)$ по подгруппе $m\mathbb{Z}, m \in \mathbb{Z}$ и указать их представители.

3. Найти правые смежные классы группы S_3 по подгруппе $\{e, (1,2)\}$, где через e обозначена тождественная подстановка (не меняет элементы), а через $(1,2)$ транспозиция элементов 1 и 2 (переставляет элементы 1 и 2, а элемент 3 не меняет).

1.3 Фактор-группы

Определение. *Подгруппа H группы G называется нормальной, если разбиение G на левые и правые смежные классы по подгруппе H совпадают, т. е. $gH = Hg$ для любого $g \in G$. Обозначается: $H \lhd G$.*

Предложение 1.3.1 *Для подгруппы H группы G следующие условия эквивалентны:*
1. $H \lhd G$;
2. $gHg^{-1} = H$ для любого $g \in G$;
3. $gHg^{-1} \subseteq H$ для любого $g \in G$.

Доказательство. $1 \Longrightarrow 2$ Так как $gH = Hg$, то

$$gHg^{-1} = Hgg^{-1} = H$$

для любого $g \in G$.

$2 \Longrightarrow 3$ очевидно.

$3 \Longrightarrow 1$ Так как $gHg^{-1} \subseteq H$, то $gHg^{-1}g \subseteq Hg$, следовательно, $gH \subseteq Hg$. Аналогично, применяя 3 для элемента g^{-1}, получим $g^{-1}Hg \subseteq H$, откуда $Hg \subseteq gH$. Таким образом, $gH = Hg$ для всех $g \in G$. □

Если H – нормальная подгруппа группы G, то можно естественным образом построить новую группу G/H, элементами которой являются смежные классы (левые или правые) группы G по подгруппе H. При этом операция на элементах группы G/H задается следующим образом:

$$g_1H \cdot g_2H = g_1g_2H.$$

Можно проверить, что это определение корректно и не зависит от выбора представителей смежных классов. Единицей группы G/H служит подгруппа H, обратным к элементу gH является элемент $g^{-1}H$.

Определение. *Группа G/H называется фактор-группой группы G по нормальной подгруппе H.*

Заметим, что отображение $\varphi : G \to G/H$, для которого $\varphi(g) = gH$, является сюръективным гомоморфизмом группы G на группу G/H, который называется *естественным*.

Сформулируем без доказательства теорему о гомоморфизме для групп.

Теорема 1.3.1 *Пусть $f : G \to H$ – сюръективный гомоморфизм групп, $\varphi : G \to G/\mathrm{Ker}\ f$ – естественный гомоморфизм, заданный формулой $\varphi(a) = a\mathrm{Ker}\ f$. Тогда существует единственный изоморфизм $\psi : G/\mathrm{Ker}\ f \to H$ такой, что $f = \psi \circ \varphi$.*

Упражнения

1. Доказать, что в абелевой группе всякая подгруппа является нормальной.
2. Доказать, что всякая подгруппа индекса 2 нормальна.
3. Доказать, что ядро гомоморфизма является нормальной подгруппой.
4. Найти таблицу умножения группы $\mathbb{Z}/2\mathbb{Z}$. С помощью теоремы о гомоморфизме показать, что группа $\mathbb{Z}/2\mathbb{Z}$ изоморфна группе $\{-1; 1\}$ по умножению.
5. Какой группе изоморфна мультипликативная группа $\mathbb{R}^*/\mathbb{R}_+$, где $\mathbb{R}^*$ множество ненулевых, $\mathbb{R}_+$ – множество положительных действительных чисел?
6. Какой группе изоморфна мультипликативная группа $\mathbb{R}^*/\{-1; 1\}$ (множество $\{-1; 1\}$ является подгруппой по умножению)?

1.4 Циклические группы. Функция Эйлера

Определение. *Пусть G мультипликативная группа, $a \in G$. Обозначим через a^n, где $n \in \mathbb{Z}$ – произвольное, следующий элемент группы*

$$a^n = \begin{cases} \underbrace{a \cdots a}_{n}, & \text{если } n > 0 \\ e, & \text{если } n = 0 \\ (a^{-n})^{-1}, & \text{если } n < 0 \end{cases} \tag{1}$$

Легко проверить, что операция возведения элемента группы в целочисленную степень для всех $a \in G, x, y \in \mathbb{Z}$ удовлетворяет следующим свойствам:

$$a^{x+y} = a^x a^y; (a^x)^y = a^{xy}. \tag{2}$$

Определение. *Мультипликативная группа G называется циклической с образующим элементом a, если $G = \{a^n \mid n \in \mathbb{Z}\}$.*

Напомним определение понятия делимости целых чисел.

Определение. *Пусть a и b – целые числа. Говорят, что a делится на b, если существует целое число c такое, что $a = bc$. Обозначается $a \vdots b$. В этом случае также говорят, что b делит a. Обозначается $b \mid a$.*

Известно, что все циклические группы равной мощности изоморфны. Для конечных циклических групп справедлива следующая теорема.

Теорема 1.4.1 *Пусть $G = \{e, a, a^2, ..., a^{n-1}\}, a^n = e$ – циклическая группа с образующим элементом a. Пусть $m, k \in \mathbb{Z}$ – произвольные. Тогда*

1. $a^m = e \Leftrightarrow m \vdots n$.

2. Элемент a^k является образующим элементом группы G тогда и только тогда, когда $(k, n) = 1$.

3. Для любого натурального $d \mid n$ существует единственная подгруппа H группы G порядка d. Подгруппа H является циклической.

Большую роль в теории чисел играет функция Эйлера[1].

Определение. *Функцией Эйлера $\varphi(n)$ натурального числа n называется количество натуральных чисел не превосходящих n и взаимно простых с ним.*

Пункт 2 теоремы 1.4.1 проясняет комбинаторный смысл функции Эйлера.

Следствие 1.4.1 *Число образующих циклической группы порядка n равно $\varphi(n)$.*

Одним из свойств функции Эйлера является ее мультипликативность.

Определение. *Теоретико-числовая функция φ называется мультипликативной, если для любых натуральных m и n таких, что $(m, n) = 1$ справедливо равенство $\varphi(mn) = \varphi(m)\varphi(n)$.*

Из мультипликативности функции Эйлера следует формула для ее вычисления:

$$\varphi(n) = n\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right)\cdots\left(1 - \frac{1}{p_k}\right) \text{ для } n = p_1^{\alpha_1}p_2^{\alpha_2}\cdot\cdot_k^{\alpha_k}, \qquad (3)$$

где $p_1, ..., p_k$ – различные простые числа, $\alpha_1, ..., \alpha_k$ – натуральные числа.

Напомним определение порядка элемента группы.

Определение. *Порядком $\mathrm{ord}(a)$ элемента a мультипликативной группы G называется наименьшее натуральное k такое, что $a^k = e$. Если такого k нет, то порядок – бесконечный.*

[1] Леонард Эйлер – щвейцарский математик (1707-1783).

Каждый элемент конечного порядка k порождает циклическую группу порядка k. Если группа G – конечная, то порядок любого ее элемента – конечен и является делителем порядка группы (последнее следует из теоремы Лагранжа).

Нам потребуется следующая формула для функции Эйлера.

Предложение 1.4.1 *Для любого натурального числа n имеет место соотношение*

$$\sum_{d|n} \varphi(d) = n, \tag{4}$$

где d – натуральные.

Доказательство. Рассмотрим циклическую группу G порядка n.

Подсчитаем число k, образующих элементов всех циклических подгрупп группы G, двумя способами.

Так как каждый элемент группы G является образующим элементом некоторой циклической подгруппы группы G имеет место равенство $k = n$.

Из 3 пункта теоремы 1.4.1 следует, что для каждого натурального делителя d числа n существует единственная подгруппа H порядка d группы G. Подгруппа H является циклической. Согласно следствию 1.4.1, число образующих подгруппы H равно $\varphi(d)$. Получаем равенство

$$k = \sum_{d|n} \varphi(d),$$

которое завершает доказательство предложения. □

Упражнения

1. Доказать теорему 1.4.1.
2. Вычислить $\varphi(120)$.
3. Доказать следующие формулы:
   a) $\varphi(p) = p - 1$, где p – простое;
   b) $\varphi(p^\alpha) = p^\alpha - p^{\alpha-1}$, где p – простое, α – натуральное;
   c) формулу (3).
4. Доказать, что каждый элемент a конечной группы G имеет конечный порядок и порождает циклическую подгруппу порядка $\mathrm{ord}(a)$.
5. Пусть $G = \{e, a, a^2, ..., a^{11}\}, a^{12} = e$ – циклическая группа порядка 12. Найти порядки элементов a^6, a^8, a^9 группы G и выписать элементы порожденных ими циклических подгрупп.

6. Доказать мультипликативность функции Эйлера.

   Указание. Пусть G и H группы. Прямым произведением $G \times H$ групп G и H называется декартово произведение множеств $G \times H$ по отношению к операции $(g_1, h_1)(g_2, h_2) = (g_1 g_2, h_1 h_2)$.

   a) Показать, что прямое произведение групп является группой.

   b) Пусть G и H – циклические группы порядков m и n соответственно, где $(m, n) = 1$. Тогда:

      i. группа $G \times H$ – циклическая порядка mn;

      ii. $O(G \times H) = O(G) \times O(H)$, где функция O обозначает множество образующих элементов циклической группы. Последнее равенство завершает доказательство мультипликативности функции Эйлера.

2 Кольца и поля

2.1 Основные определения и понятия

Определение. *Кольцом называется непустое множество R, на котором заданы две бинарные алгебраические операции: "$+$" – сложение и "$\cdot$" – умножение, такие, что:*

1. $(R, +)$ – абелева группа;

2. операция умножения ассоциативна, то есть для всех $a, b, c \in R$ выполнено $(a \cdot b) \cdot c = a \cdot (b \cdot c)$;

3. операция умножения дистрибутивна относительно сложения, то есть для всех $a, b, c \in R$ выполнено

$$(a + b) \cdot c = a \cdot c + b \cdot c, \qquad c \cdot (a + b) = c \cdot a + c \cdot b.$$

Говорят, что кольцо R коммутативно, если для всех $a, b \in R$ выполнено $a \cdot b = b \cdot a$ и с единицей, если в R существует нейтральный по умножению элемент $1 \neq 0$.

Для произвольного кольца справедливо следующее утверждение.

Предложение 2.1.1 *(простейшие свойства кольца) Пусть R – кольцо, элементы $a, b \in R$ – произвольные, 0_R – нуль кольца. Тогда справедливы равенства:*

1. $0_R a = a 0_R = 0_R$;

2. $a(-b) = (-a)b = -ab$.

Ненулевой элемент $a \in R$ называется делителем нуля, если существует $b \neq 0$ такой, что $a \cdot b = 0$.

Пример. В кольце квадратных матриц второго порядка элемент $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ является делителем нуля.

Коммутативное кольцо с единицей и без делителей нуля называется *областью целостности.*

Элемент a кольца R с единицей называется *обратимым*, если для него существует элемент $b \in R$ такой, что $a \cdot b = b \cdot a = 1$.

Множество всех обратимых элементов кольца R с единицей образует группу, которая обозначается $U(R)$.

Кольцо R с единицей называется *телом*, если всякий его ненулевой элемент обратим.

Коммутативное тело называется *полем.*

Пример. $(\mathbb{Q}, +, \cdot)$, $(\mathbb{R}, +, \cdot)$ являются полями.

Определение. *Пусть R – кольцо, $a \in A$, n – произвольное целое число. Обозначим через na элемент кольца равный*

$$na = \begin{cases} \underbrace{a + ... + a}_{n}, & \text{если } n > 0 \\ 0, & \text{если } n = 0 \\ -((-n)a), & \text{если } n < 0. \end{cases} \tag{5}$$

Операция (5) является аналогом дискретной экспоненты (1) для аддитивных групп.

Так же как и операция (1), операция (5) удовлетворяет свойствам аддитивной экспоненты:

$$(x + y)a = xa + ya; x(ya) = (xy)a, x(ab) = (xa)b = a(xb), \tag{6}$$

где $a, b \in R, x, y \in \mathbb{Z}$ – произвольные.

Определение. *Характеристикой* $\mathrm{char}\ F$ *поля* F *называется порядок элемента* 1 *в аддитивной группе поля* F*, то есть наименьшее натуральное* k *такое, что* $k \cdot 1 = 0$.

Если такого натурального k *не существует, то говорят, что характеристика поля равна бесконечности или* 0.

Несложно доказать следующее утверждение.

Лемма 2.1.1 *Если характеристика поля F конечная и равна p, то число p – простое.*

Пример. Рассмотрим множество, состоящее из двух элементов $F = \{0, 1\}$ и зададим операции сложения и умножения с помощью следующих таблиц Кэли:

$+$	0	1
0	0	1
1	1	0

$\cdot$	0	1
0	0	0
1	0	1

Легко проверить, что F является полем с нулем 0 и единицей 1, $char\ F = 2$. Оно называется полем Галуа и обозначается F_2.

Упражнения

1. Доказать предложение 2.1.1.
2. Доказать лемму 2.1.1.
3. Пусть $char\ F = p$,$a \in F$ – произвольный. Тогда $pa = 0$.

4. Доказать, что в поле нет делителей нуля.

5. Пусть $char\ F = p$,$a \in F, a \neq 0$ – произвольный. Тогда наименьшее натуральное k такое, что $ka = 0$ равно p.

2.2 Подкольца, идеалы и фактор-кольца

Аналогом понятия подгруппы в группе является понятие подкольца в кольце, а аналогом нормальной подгруппы – понятие идеала.

Определение. *Непустое подмножество S кольца R называется подкольцом, если оно само является кольцом относительно операций сложения и умножения, заданных на R.*

Также как и для групп, для колец можно доказать следующее утверждение.

Предложение 2.2.1 *Непустое подмножество S кольца R является подкольцом, тогда и только тогда, когда оно замкнуто относительно операций сложения, умножения и взятия противоположного элемента, определенных на R.*

Ясно, что $\mathbb{Z}$ и $\mathbb{Q}$ являются подкольцами в $\mathbb{R}$.

Определение. *Подмножество I кольца R называется идеалом (обозначение: $I \lhd R$), если выполнены следующие условия:*
1. если $a,\ b \in I$, то $a \pm b \in I$;
2. если $a \in I$, то $ar, ra \in I$ для любого $r \in R$.

Если I – идеал кольца R то на фактор-группе $(R/I, +)$ можно задать умножение следующим образом:

$$(a + I) \cdot (b + I) = ab + I.$$

Легко проверить, что $(R/I, +, \cdot)$ является кольцом, которое называется *факторкольцом* кольца R по идеалу I.

Заметим далее, что сумма и пересечение идеалов является идеалом.

Определение. *Идеалом, порожденным подмножеством S кольца R, называется пересечение всех идеалов кольца, содержащих S (обозначение: $(S)_R$).*

Определение. *Идеал $I \lhd R$ кольца с единицей называется главным, если существует элемент $a \in I$ такой, что любой элемент $b \in I$ может быть представлен в виде $b = ar$ или $d = ra$ для некоторого $r \in R$, т. е. $I = (a)_R$.*

Коммутативное кольцо с единицей называется *кольцом главных идеалов*, если все его идеалы главные.

Далее мы покажем, что кольцами главных идеалов являются кольцо $\mathbb{Z}$ и кольцо многочленов от одной переменной $\mathbb{R}[x]$. В то же время кольцо многочленов от двух переменных $\mathbb{R}[x, y]$ таковым не является.

Определение. *Обозначим через $\mathbb{Z}_m$ фактор-кольцо $\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z}$ кольца целых чисел $\mathbb{Z}$ по идеалу $m\mathbb{Z}$ и назовем его кольцом классов вычетов по модулю m, где $m > 1$ – натуральное. Обозначим через*

$$\hat{a} = a + m\mathbb{Z}, a \in \mathbb{Z}$$

элементы кольца $\mathbb{Z}_m$.

Определение. *Отображение $f : R \to S$ кольца R в кольцо S называется гомоморфизмом колец если выполнены условия:*

1. для всех $a, b \in R$ выполнено $f(a + b) = f(a) + f(b)$;

2. для всех $a, b \in R$ выполнено $f(ab) = f(a)f(b)$.

Гомоморфизм колец заданный биективным отображением называется изоморфизмом, а сами кольца – изоморфными.

Отметим, что гомоморфизм колец f является гомоморфизмом абелевых групп $(R, +)$ и $(S, +)$. В частности, при гомоморфизме колец нуль кольца R переходит в нуль кольца S.

Определение. *Пусть $f : R \to S$ – гомоморфизм колец. Назовем ядром гомоморфизма* Ker f *множество*

$$\text{Ker } f = \{x \mid x \in R, f(x) = 0\}.$$

Лемма 2.2.1 *Ядро гомоморфизма колец является идеалом.*

Сформулируем без доказательства теорему о гомоморфизме для колец.

Теорема 2.2.1 *Пусть $f : R \to S$ – сюръективный гомоморфизм колец, $\varphi : R \to R/\text{Ker } f$ – естественный гомоморфизм заданный формулой $\varphi(a) = a + \text{Ker } f$. Тогда существует единственный изоморфизм $\psi : R/\text{Ker } f \to S$ такой, что $f = \psi \circ \varphi$.*

Упражнения

1. Доказать лемму 2.2.1.

2. Показать, что кольцо $\mathbb{Z}_6$ содержит делители нуля.

3. Показать, что кольцо классов вычетов $\mathbb{Z}_m$ является полем тогда и только тогда, когда m – простое.

4. Найти обратные элементы для всех ненулевых элементов поля $\mathbb{Z}_7$.

5. Найти порядки элементов $\hat{3}, \hat{5}$ в мультипликативной группе поля $\mathbb{Z}_7$.

6. Доказать, что ядро гомоморфизма колец является идеалом.

3 Алгебраические структуры

3.1 Алгебры и алгебраические системы

В предыдущих разделах мы изучали алгебраические объекты с одной или двумя бинарными операциями, удовлетворяющими определенным условиям. Все они являются *алгебраическими структурами* или просто *алгебрами*.

Определение. *Алгеброй (универсальной алгеброй или алгебраической структурой) называется множество, наделенное системой операций. Чтобы подчеркнуть, какие операции заданы на данном множестве A пишут $\mathcal{A} = (A, \varphi_1, \ldots, \varphi_m)$. Множество операций $\Sigma = \{\varphi_1, \ldots, \varphi_m\}$ называется сигнатурой, а вектор арности $(n_1, \ldots, n_m)$ – типом алгебры.*

Пример. Кольцо является алгеброй $(R, +, \cdot)$ с двумя бинарными операциями, т. е. сигнатура $\Sigma = \{+, \cdot\}$, тип - $(2, 2)$.

Определение. *Пусть $\mathcal{A} = (A, \varphi_1, \ldots, \varphi_m)$ и $\mathcal{B} = (B, \psi_1, \ldots, \psi_m)$ – две алгебры одинакового типа. Отображение $f : A \to B$ называется гомоморфизмом алгебр $\mathcal{A}$ и $\mathcal{B}$, если*

$$f(\varphi_i(a_1, \ldots, a_{n_i}) = \psi_i(f(a_1), \ldots, f(a_{n_i}))$$

для всех $i = 1, \ldots, m$.

Гомоморфизмы, обладающие дополнительными свойствами, имеют специальные названия. В частности, биективный гомоморфизм называется изоморфизмом. Изоморфные алгебры обладают одинаковыми алгебраическими свойствами и поэтому алгебраические структуры принято рассматривать с точностью до изоморфизма.

Иногда кроме алгебраических операций на множестве заданы отношения. В компьютерной математики такие объекты называются *алгебраическими системами*.

Пример. $(\mathbb{N}, +, \cdot, \geq)$ – алгебраическая система.

Упражнения

1. Найти сигнатуру и тип мультипликативных и аддитивных групп. Как изменится сигнатура и тип, если добавить операции взятия симметричного и нейтрального элементов?

2. Рассмотрим операцию $s(\bar{x}, \bar{y}, \bar{z}) = (\bar{x} \times \bar{y}) \times \bar{z}$ векторном пространстве V_3. Какова арность этой операции? Найти тип алгебры $(V_3, +, \cdot, s)$.

3. Алгебры с одной унарной операцией называются унарами. Привести примеры унаров.

3.2 Булевы алгебры

Булевы алгебры возникли в трудах Дж. Буля[1] как аппарат символической логики. В последующем они нашли широкое применение в различных разделах математики – в теории вероятностей, топологии, функциональном анализе и др.

Определение. *Булевой алгеброй* $\mathcal{B} = (B, \vee, \wedge, C)$ *называется непустое множество* B *с двумя бинарными операциями* $\vee, \wedge$, *(дизъюнкция и конъюнкция) и одной унарной операцией* C*(не), которые удовлетворяют следующим аксиомам для любых* $a, b, c \in B$*:*

1) $a \vee b = b \vee a, \quad a \wedge b = b \wedge a;$

2) $a \vee (b \vee c) = (a \vee b) \vee c, \quad a \wedge (b \wedge c) = (a \wedge b) \wedge c;$

3) $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c), \quad a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c);$

4) $a \vee (a \wedge b) = a, \quad a \wedge (a \vee b) = a;$

5) $a \vee (b \wedge Cb) = a, \quad a \wedge (b \vee Cb) = a.$

Аксиомы 1 называются законами коммутативности (дизъюнкции и конъюнкции соответственно); 2 – ассоциативности; 3 – дистрибутивности; 4 – законами поглощения; 5 – законами комбинации с дополнениями. Заметим, что множество аксиом 1–5 не меняется при замене $\vee$ на $\wedge$, $\wedge$ на $\vee$. Откуда следует, то в теории булевых алгебр справедлив, так называемый, *принцип двойственности*:

любая теорема о булевых алгебрах, в формулировке которой участвуют только операции $\vee$, $\wedge$ *и* C, *остается справедливой, если в ее формулировке заменить* $\vee$ *на* $\wedge$, *и наоборот.*

Пример. Двухэлементная булева алгебра $\mathcal{B} = (\{0, 1\}, \vee, \wedge, C)$, в которой $0 \wedge 1 = 0,\ 0 \vee 1 = 1,\ C1 = 0,\ C0 = 1.$

Пример. Пусть M – некоторое множество, обозначим через 2^M – множество всех подмножеств множества M. Легко проверить, что $(2^M, \cup, \cap, C)$ является булевой алгеброй (здесь $\cup$, $\cap$ и C обозначают соответственно объединение, пересечение и взятие дополнения).

На булевых алгебрах вводится отношение порядка

$$x \leq y \Leftrightarrow x = x \wedge y.$$

В аксиомах булевой алгебры отражена аналогия между понятиями “множества”, “события”, “высказывания”. Отношение порядка в булевой алгебре может быть (в зависимости от выбора интерпретации) истолковано как теоретико-множественное включение, как причинное следование для событий, как логическое следование для высказываний и т.д.

Определим на булевой алгебре операцию

$$x \Delta y = (x \wedge Cy) \vee (y \wedge Cx),$$

[1] Джордж Буль – английский математик (1815-1864).

которая называется *симметрической разностью*. По отношению к операциям Δ и $\wedge$ булева алгебра является кольцом.

Любая булева алгебра сводится к алгебре подмножеств (не обязательно всех) некоторого множества.

Теорема 3.2.1 (Стоуна[1]) *Всякая булева алгебра изоморфна некоторой алгебре множеств.*

Упражнения

1. Доказать, что по отношению к операциям Δ и $\wedge$ булева алгебра является кольцом.

2. Рассмотреть булеву алгебру всех подмножеств множества из двух элементов. Написать таблицы сложения и умножения.

3. Сколько существует неизоморфных четырехэлементных булевых алгебр ?

4 Кольцо целых чисел

4.1 Делимость в кольце целых чисел

Определение. *Коммутативное кольцо $\mathbb{Z}$ с единицей, содеражащее подмножество $\mathbb{N} \subseteq \mathbb{Z}$ называется кольцом целых чисел, если выполнены условия:*
1. $(\mathbb{N}, +, \cdot, 1)$ является системой натуральных чисел;
2. для всех $z \in \mathbb{Z}$ существуют натуральные $a, b \in \mathbb{N}$ такие, что $z = a - b$.

Можно показать, что кольцо целых чисел $\mathbb{Z}$ не содержит делителей нуля и, следовательно, является областью целостности.

Определение. *Разделить с остатком целое число a на целое число b – это значит найти такие целые числа q и r, что*

$$a = bq + r, \quad 0 \leq r < \mid b \mid . \tag{7}$$

Числа q и r называются соответственно неполным частным и остатком от деления a на b.

Теорема 4.1.1 (о делении с остатком) *Если $a, b \in \mathbb{Z}$ и $b \neq 0$, то a можно разделить на b с остатком, причем неполное частное и остаток определяются однозначно.*

Доказательство. Можно показать, что

$$\mathbb{Z} = \{-n \mid n \in \mathbb{N}\} \cup \{0\} \cup \mathbb{N}\}.$$

Мы считаем, что для натуральных чисел можно проводить доказательство с помощью математической индукции.

Сначала докажем существование чисел q и r, удовлетворяющих условию (1).

Рассмотрим три случая:

1. $a \geq 0$, $b > 0$. Зафиксируем b и рассмотрим индукцию по a. При $a = 0$ имеем $0 = b \cdot 0 + 0$, т. е. утверждение справедливо. По индуктивному предположению, утверждение справедливо при $a = k$, т. е. существует пара чисел q_1 и r_1 такая, что

$$k = bq_1 + r_1, \quad 0 \leq r_1 < b \tag{8}$$

Покажем, что при $a = k+1$ также существует пара чисел, удовлетворяющая условию (7). Из (8) получим, $k + 1 = bq_1 + (r_1 + 1)$. Если $r_1 + 1 < b$, то искомая пара найдена, а именно: $q = q_1$, $r = r_1 + 1$. Если же $r_1 + 1 = b$, то $k + 1 = bq_1 + b = b(q_1 + 1)$ и в этом случае искомой будет пара чисел $q = q_1 + 1$, $r = 0$.

2. $b > 0$, $a < 0$. Тогда $-a > 0$ и поэтому существует пара целых чисел q' и r' такая, что $-a = bq' + r'$, где $0 \leq r' < b$. Следовательно $a = b(-q') - r' = b(-q' - 1) + (b - r')$, и искомая пара чисел найдена: $q = -q' - 1$, $r = b - r'$, если $r' > 0$ и $q = -q'$, $r = 0$, если $r' = 0$.

3. $b < 0$. Тогда $b = -\mid b \mid$. Найдем $q_1 \geq 0$ и $0 \leq r' < b$ такие, что $a = \mid b \mid q_1 + r_1 = b(-q_1) + r_1$. При $q = -q_1$ и $r = r_1$ получаем: $a = bq + r$.

Существование пары чисел, указанных в теореме, доказано.

Докажем единственность этой пары. Допустим, что $a = bq_1 + r_1$ и $a = bq_2 + r_2$. Тогда $bq_1 + r_1 = bq_2 + r_2$, откуда $b(q_1 - q_2) = r_2 - r_1$. Последнее соотношение показывает, что число $r_2 - r_1$ нацело делится на b, а это возможно лишь при $r_2 - r_1 = 0$. Следовательно, $r_1 = r_2$ и $b(q_1 - q_2) = 0$, поэтому $q_1 - q_2 = 0$ и $q_1 = q_2$. $\square$

Упражнения

1. Пусть a, b, c – произвольные целые числа. Доказать, следующие свойства делимости:.

   a) $a \vdots c, b \vdots c \Rightarrow a \pm b \vdots c$;

   b) $a \vdots c \Rightarrow ab \vdots c$;

   c) Если $c \neq 0$, то $ac \vdots bc \Leftrightarrow a \vdots b$.

2. Доказать, что $3^{50} + 4^{50} \vdots 25$.

3. Найти неполные частные и остатки от деления чисел: 123 на 37; -123 на 37; 123 на -37; -123 на -37.

4.2 Наибольший общий делитель. Алгоритм Евклида

Определение. *Целое число $d \neq 0$ называется наибольшим общим делителем чисел a_1, a_2,..., a_n, если выполнены два условия:*

1. d является общим делителем чисел $a_1,\ a_2, ..., a_n$, т. е.

$$d \mid a_1, d \mid a_2, ..., d \mid a_n;$$

2. d делится на любой другой общий делитель чисел $a_1,\ a_2, ..., a_n$.

Лемма 4.2.1 *Наибольший общий делитель не равных нулю одновременно чисел $a_1,\ a_2, ..., a_n$ определяется однозначно с точностью до знака.*

Доказательство. Пусть d и d_1 два наибольших общих делителя чисел $a_1,\ a_2, ..., a_n$.

Тогда $d_1 \mid d$ и $d \mid d_1$, откуда получим, что

$$d = d_1k_1, d_1 = dk \Rightarrow d = dkk_1 \Rightarrow d(1 - kk_1) = 0.$$

Отметим, что по условию одно из чисел a_i не равно нулю. Тогда $d \neq 0$.

Так как кольцо целых чисел является областью целостности, получим $kk_1 = 1$.

Следовательно, либо $d = d_1$, либо $d = -d_1$. □

Замечание. Обозначим через НОД $(a_1,\ a_2, ..., a_n)$ или $(a_1,\ a_2, ..., a_n)$ неотрицательный наибольший общий делитель, который определяется однозначно.

Легко проверить, что:

1. если $b \mid a$, то $(a, b) = b$;

2. если $a = bq + r$, то $(a, b) = (a, r) = (b, r)$.

Для нахождения (a, b), при условии, что $b \neq 0$, можно применить известный *алгоритм Евклида*[1] , сущность которого заключается в следующем:

Сначала делим с остатком a на b : $a = bq_1 + r_1,\ \ 0 \leq r_1 <\mid b \mid$. Если $r_1 = 0$, то алгоритм окончен. В этом случае $b \mid a$, и, очевидно, $(a, b) =\mid b \mid$. Если же $r_1 \neq 0$, то делим с остатком b на r_1, получим: $b = r_1q_2 + r_2,\ 0 \leq r_2 < r_1$. Если $r_2 = 0$, то алгоритм окончен, в противном случае делим с остатком r_1 на r_2 и т.д. до тех пор, пока не получим остаток, равный нулю. Такой момент обязательно наступит, поскольку получающиеся остатки являются неотрицательными числами и $r_1 > r_2 > \ldots$.

В случае, когда $r_1, r_2, \ldots, r_n$ отличны от нуля, а $r_{n+1} = 0$, получим следующую систему соотношений:

[1] Евклид – древнегреческий математик (365 - 300 лет до н.э.).

$$\begin{gathered} a = bq_1 + r_1, \quad 0 < r_1 <\mid b \mid, \\ b = r_1q_2 + r_2, \ 0 < r_2 < r_1, \\ r_1 = r_2q_3 + r_3, \ 0 < r_3 < r_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ r_{n-2} = r_{n-1}q_n + r_n, \ 0 < r_n < r_n - 1, \\ r_{n-1} = r_nq_{n+1} \end{gathered} \tag{9}$$

Двигаясь по системе равенств (9) снизу вверх, получим

$$r_n = (r_{n-1}, r_n) = (r_{n-2}, r_{n-1}) = ... = (b, r_1) = (a, b).$$

Таким образом, справедлива следующая теорема:

Теорема 4.2.1 *Для любых целых чисел a и b существует единственный неотрицательный наибольший общий делитель (a, b). При этом, если $b \mid a$ или $a \mid b$, то соответственно $(a, b) = a$ или $(a, b) = b$, в противном случае (a, b) совпадает с последним ненулевым остатком в алгоритме Евклида для чисел a и b.*

Поднимаясь по системе равенств (9) снизу вверх нетрудно представить наибольший общий делитель чисел a и b в виде целочисленной линейной комбинации этих чисел.

Теорема 4.2.2 (Линейное представление НОД) *Если $a,\ b \in \mathbb{Z}$ и $d{=}(a, b)$, то существуют такие целые числа $u,\ v$, что имеет место равенство:*

$$au + bv = d.$$

Замечание. Для нахождения НОД нескольких чисел используют следующее свойство $(a_1, a_2, \ldots, a_n) = (a_1, (a_2, \ldots, a_n))$, которое порождает рекурсивный алгоритм нахождения *НОД*, сущность которого заключается в следующем:

1. выбирают два числа a_1, a_2 в последовательности $(a_1, a_2, \ldots, a_n)$ и вычисляют (a_1, a_2);

2. заменяют эти два числа на их *НОД* и повторяют алгоритм.

Определение. *Целые числа a и b называются взаимно простыми, если $(a, b) = 1$.*

Упражнения

1. Показать, что целые a и b взаимно просты тогда и только тогда, когда существуют целые u и v такие, что $au + bv = 1$.

2. Доказать, что для целых a, b, c справедливо логическое следствие $ab \vdots c, (b, c) = 1 \Rightarrow a \vdots c$.

3. Используя упражнение 1, доказать теорему Евклида: если для произвольных целых a, b, c выполнено $(a, c) = 1, (b, c) = 1$, то $(ab, c) = 1$.

4. Найти наибольший общий делитель чисел 273 и 195 и его линейное представление.

4.3 Наименьшее общее кратное

Определение. *Наименьшим общим кратным НОК целых чисел* $a_1, a_2, \ldots, a_n$ *при* $n \geq 2$ *называется целое число* k*, удовлетворяющее условиям:*
1. k *является общим кратным чисел* $a_1,\ a_2, ..., a_n$*, т. е.*

$$a_1 \mid k, a_2 \mid k, ..., a_n \mid k;$$

2. k *делит любое общие кратное чисел* $a_1,\ a_2, ..., a_n$*.*

Как и в случае наибольшего общего делителя *НОК* ненулевых чисел $a_1,\ a_2, ..., a_n$ определяется однозначно с точностью до знака. Обозначим через НОК $(a_1,\ a_2, ..., a_n)$ или $[a_1,\ a_2, ..., a_n]$ положительное наименьшее общее кратное.

Предложение 4.3.1 *Для любых ненулевых чисел* a *и* b *справедливо следующее соотношение:*

$$[a, b] = \frac{\mid ab \mid}{(a, b)}.$$

Доказательство. Так как $[a, b] = [|a|, |b|]$, то достаточно доказать утверждение для положительных a и b. Пусть $d = (a, b)$, тогда существуют такие числа $k_1,\ k_2$, что $a = k_1 d,\ b = k_2 d$. Следовательно,

$$m = \frac{ab}{d} = k_1 k_2 d = k_2 a = k_1 b$$

является кратным чисел a и b. Несложно проверить, что k_1 и k_2 – взаимно просты.

Пусть теперь $x \in \mathbb{Z}$ и $a \mid x$ и $b \mid x$. Тогда, получим

$$x = ay = dk_1 y \Rightarrow dk_1 y \vdots b \Rightarrow dk_1 y \vdots dk_2,$$

где $y \in \mathbb{Z}$

НОД d ненулевых целых чисел отличен от 0. Из 3-го свойства делимости следует, что $k_1 y \vdots k_2$.

Так как k_1 и k_2 – взаимно просты, из 2-го упражнения предыдущего раздела получим $y \vdots k_2 \Rightarrow y = k_2 t, t \in \mathbb{Z}$.

Следовательно, $x = dk_1 k_2 t \Rightarrow x \vdots m$.

Мы доказали, что

$$m = \frac{|\, ab \,|}{(a, b)}$$

является наименьшим общим кратным целых чисел a и b. □

Для нахождения *НОК* нескольких чисел используют следующее свойство $[a_1, a_2, \ldots, a_n] = [a_1, [a_2, \ldots, a_n]]$, которое порождает рекурсивный алгоритм нахождения *НОК*.

Упражнения

1. Найти наименьшее общее кратное чисел 273 и 195.
2. Найти наименьшее общее кратное чисел 140, 168 и 210.
3. Найти наименьшее общее кратное чисел 105, 147 и 315.

4.4 Простые числа

Определение. *Натуральное числа $n > 1$ называется называется составным, если существуют натуральные $a, b < n$ такие, что $n = ab$. Натуральное число $p > 1$ называется простым, если оно не является составным.*

Можно показать, что простое число p не имеет других натуральных делителей кроме 1 и p.

Лемма 4.4.1 *Пусть a – целое число, p – простое. Если $a \not\vdots p$, то a и p – взаимно просты.*

Лемма 4.4.2 *Пусть $a_1, ..., a_r$ – целые числа, p – простое. Если $a_1 \cdots a_r \vdots p$, то одно из чисел $a_i (1 \leq i \leq r)$ делится на p.*

Лемма 4.4.1 следует непосредственно из определения простого числа. Лемма 4.4.2 доказывается по индукции с использованием упражнения 2 из раздела 4.2.

Теорема 4.4.1 (Основная теорема арифметики) *Любое натуральное число $n > 1$ либо простое, либо имеет единственное, с точностью до перестановки сомножителей, представление в виде $n = p_1 p_2 ... p_k$, где p_i – простое число.*

Доказательство проведем индукцией по числу n. Так как 2 - простое число, то при $n = 2$ утверждение теоремы верно.

Допустим, что оно верно и для всех $1 \leq k < n$ и докажем его истинность для $k = n$. Если n - простое число, то для него утверждение теоремы верно. Предположим, что n - составное. Тогда оно делится на некоторое число a, такое, что $1 < a < n$. Следовательно, $n = a \cdot b$, где $1 < b < n$. По предположению индукции каждое из чисел a и b либо простое, либо разлагается в произведение простых, т. е. $a = p_1 p_2 ... p_k$, $b = q_1 q_2 ... q_l$, а следовательно, $n = a \cdot b = p_1 p_2 ... p_k q_1 q_2 ... q_l$

Докажем единственность представления составного числа n в виде произведения простых. Пусть

$$n = p_1 p_2 ... p_k = q_1 q_2 ... q_l.$$

Так как $p_1 \mid n$, то $p_1 \mid q_i$ для некоторого $i = 1, ..., l$, при соответствующей нумерации можно считать, что $p_1 \mid q_1$, а так как q_1 - простое, то $p_1 = q_1$.

Получим, что

$$n/p_1 = p_2 ... p_k = q_2 ... q_l.$$

По предположению индукции эти разложения совпадают, а значит совпадают и исходные. □

Следствие 4.4.1 *Любое целое число z можно представить единственным образом в виде*

$$z = \varepsilon p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k},$$

где $\varepsilon = \pm 1$, p_i $(i = 1, ..., k)$ – простые числа, причем $p_1 < p_2 < ... < p_n$, $\alpha_i \in \mathbb{N}$.

Определение. *Такое представление числа z называется каноническим разложением.*

В связи с большой ролью, которую играют простые числа в арифметики, множество простых чисел привлекало к себе внимание многих математиков. Изучением его свойств занимались Евклид, П. Ферма[1], Л. Эйлер, А.М. Лежандр[2], П.Л. Чебышев[3], И.М. Виноградов[4] и др. Особенно много исследований было проведено относительно распределением простых чисел в натуральном ряду. Из имеющихся таблиц усматривалось, что простые числа распределены весьма неравномерно, так: в первой сотне их 25, во второй – 21, в сорок девятой – 8, в пятидесятой – 15. При удалении по натуральному ряду в сторону возрастания чисел появляются все более длинные промежутки, не содержащие простых чисел.

[1]Пьер Ферма – французский математик (1601-1665).

[2]А.М. Лежандр – французский математик (1752-1833).

[3]П.Л. Чебышев – русский математик (1821-1894).

[4]И.М. Виноградов – русский математик (1891 -1983).

Какого бы ни было натуральное n, можно найти n составных чисел непосредственно следующие друг за другом, например

$$(n+1)!+2, \quad (n+1)!+3, \ldots, (n+1)!+(n+1).$$

Теорема 4.4.2 (Евклид) *Множество простых чисел бесконечно.*

Доказательство. Предположим, что множество простых чисел конечно. Выпишем их все в порядке возрастания

$$2, 3, 5, \ldots p_s.$$

Рассмотрим число $N = 2 \cdot 3 \cdot 5 \cdot \ldots \cdot p_s + 1$. Ясно, что оно не делится ни на одно из чисел $2, 3, 5, \ldots p_s$, то есть не на одно простое число, что противоречит основной теореме арифметики. □

Множество простых чисел, не превосходящих x обозначается через $\pi(x)$.
Из теоремы Евклида следует, что

$$\lim_{x \to \infty} \pi(x) = \infty.$$

В 1896 году одновременно Ж. Адамаром[1] и В. Пуссеном [2] было доказано асимптотическое равенство $\pi(x) \sim x/\ln\ x$.

Упражнения

1. Доказать, леммы 4.4.1 и 4.4.1.
2. Пусть $p \geq 5$ – простое число. Показать, что тогда p имеет вид $p = 6k+1$ или $6k+5$, где k – целое.
3. Пусть $p \geq 5$ – простое число. Показать, что тогда $p^2 - 1 \vdots 24$.
4. Показать, что простых чисел вида $4k+3$ и $6k+5$, где k – натуральное, – бесконечно много.

[1] Ж. Адамар – французский математик (1865-1963).
[2] В. Пуссен – бельгийский математик (1866-1962).

5 Комплексные числа

5.1 Определение поля комплексных чисел

Определение. *Поле $\mathbb{C}$, содержащее подполе $\mathbb{R}$ действительных чисел и элемент i, удовлетворяющий условиям:*

1. $i^2 = -1$;

2. для любого $z \in \mathbb{C}$ существуют a и $b \in \mathbb{R}$ такие, что $z = a + bi$, называется полем комплексных чисел.

Элемент i называется "мнимой единицей".

Несложно доказать следующую лемму.

Лемма 5.1.1 *Пусть $\mathbb{C}$ – поле комплексных чисел, i – мнимая единица. Тогда для любого $z \in \mathbb{C}$ существуют единственные a и $b \in \mathbb{R}$ такие, что $z = a + bi$.*

Действительное число a называется действительной частью комплексного числа z и обозначается $Re\ z$, $b = Im\ z$ называется мнимой частью числа z.

Так же как и для всех числовых систем, для поля комплексных чисел можно доказать следующую теорему.

Теорема 5.1.1 *1. Поле комплексных чисел существует.*

2. Любые два поля комплексных чисел изоморфны.

Определение. *Комплексное число $a - bi$ называются сопряженным к числу $z = a + bi$ и обозначается через $\overline{z}$.*

Легко проверить, что для любых комплексных чисел z_1 и z_2 имеют место равенства:

1. $\overline{\overline{z_1}}$;
2. $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$;
3. $\overline{z_1 z_2} = \overline{z_1} \cdot \overline{z_2}$;
4. если $z_2 \neq 0$, то
$$\frac{(\overline{z_1})}{(\overline{z_2})} = \overline{\left(\frac{z_1}{z_2}\right)}.$$

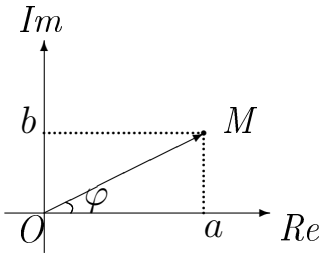
Упражнения

1. Доказать, лемму 5.1.1.
2. Доказать свойства операции комплексного сопряжения 1-4.
3. Вычислить $\frac{(1+i)^8-1}{(1-i)^8+1}$.
4. Вычислить $\frac{(1+2i)^3+(1-2i)^3}{(2-i)^2-(2+i)^2}$.
5. Вычислить $(-1+\sqrt{3}i)^{10}$.
6. Решить уравнение $z^2 = 3-4i$.

5.2 Тригонометрическая форма комплексного числа

Наряду с алгебраическим представлением комплексного числа $z = a + bi$ в математике и ее приложениях часто используются представление комплексного числа в тригонометрической и показательной форме. Для определение такого представления введем сначала геометрическую интерпретацию комплексных чисел.

Выберем на плоскости прямоугольную систему координат Oxy и изобразим комплексное число $z = a + bi$ точкой плоскости Oxy с координатами (a, b).



Таким образом, существует биективное соответствие между комплексными числами и точками координатной плоскости Oxy. Плоскость, точки которой отождествляются с комплексными числами, называется *комплексной плоскостью*. Ось абсцисс этой плоскостью называется *действительной осью*, а ось ординат – *мнимой осью*.

Расстояние от начала координат O до точки $M(a, b)$, изображающей комплексное число $z = a + bi$ называется *модулем числа* z и обозначается $|z|$. Угол образованный вектором $\overrightarrow{OM}$ с положительным направлением оси Ox называется *аргументом числа* $z \neq 0$ и обозначается $\arg(z)$. Для $z = 0$ аргумент не определяется. Аргумент комплексного числа удовлетворяет неравенству $0 \leq \arg(z) < 2\pi$.

Легко видеть, что

$$r = |z| = \sqrt{a^2 + b^2}, \tag{10}$$

а аргумент $\varphi = \arg(z)$ находится из соотношений:

$$\varphi = \begin{cases} arccos\left(\frac{a}{\sqrt{a^2+b^2}}\right), b \geq 0, a^2+b^2 \neq 0, \\ 2\pi - arccos\left(\frac{a}{\sqrt{a^2+b^2}}\right), b < 0. \end{cases} \tag{11}$$

и

$$z = r(\cos\varphi + i\sin\varphi) \tag{12}$$

Запись комплексного числа $z = a + bi$ в виде (12) называется *тригонометрической формой комплексного числа z*

В тригонометрической форме комплексного числа проще, чем в алгебраической, осуществлять умножение, деление, возведение в степень и извлечение корней.

Предложение 5.2.1 *Для любых комплексных чисел* $z_1 = r_1(\cos\varphi_1 + i\sin\varphi_1)$, $z_2 = r_2(\cos\varphi_2 + i\sin\varphi_2)$ *и целого* m *справедливы следующие равенства:*

$$z_1 z_2 = r_1 r_2(\cos(\varphi_1+\varphi_2) + i\sin(\varphi_1+\varphi_2)); \tag{13}$$

$$\text{если } z_1 \neq 0, \text{ то } z_1^m = r_1^m(\cos m\varphi_1 + i\sin m\varphi_1); \tag{14}$$

$$\text{если } z_2 \neq 0, \quad \text{то } \frac{z_1}{z_2} = \frac{r_1}{r_2}(\cos(\varphi_1-\varphi_2) + i\sin(\varphi_1-\varphi_2)). \tag{15}$$

Доказательство. Равенства (13) и (15) проверяются непосредственно, а (14) является их следствием. □

Равенство (14) называется *формулой Муавра* в честь А. де Муавра[1]. Им же была выведена и формула извлечения корня n-ой степени из комплексного числа $z = r(\cos\varphi + i\sin\varphi)$.

Теорема 5.2.1 *Для любого* $n \in \mathbb{N}$ *уравнение*

$$z^n = c, c = r(\cos\varphi + i\sin\varphi),$$

где $c \neq 0$, *имеет ровно* n *различных корней, расположенных в вершинах правильного* n*-угольника с центром в начале координат, которые находятся по формуле:*

$$z_k = \sqrt[n]{r}(\cos\frac{\varphi+2\pi k}{n} + i\sin\frac{\varphi+2\pi k}{n}), \quad k = 0, 1, ..., n-1 \tag{16}$$

[1] А. де Муавр – английский математик (1667-1754).

Доказательство. Пусть $z = \varrho(\cos\psi + i\sin\psi)$ является корнем n-ой степени из комплексного числа c, т. е. $z^n = c$. Воспользовавшись формулой Муавра (10), получим:

$$\varrho^n(\cos n\psi + i\sin n\psi) = r(\cos\varphi + i\sin\varphi).$$

Следовательно,

$$\varrho^n = r, \quad n\psi = \varphi + 2\pi k,$$

для некоторого целого числа k или

$$\varrho = \sqrt[n]{r}, \quad \psi = \frac{\varphi + 2\pi k}{n}.$$

Ясно, что для любого $k \in \mathbb{Z}$

$$z_k = \sqrt[n]{r}\left(\cos\frac{\varphi + 2\pi k}{n} + i\sin\frac{\varphi + 2\pi k}{n}\right) \tag{17}$$

является корнем n-ой степени из комплексного числа c. Выясним, сколько среди них различных чисел.

По теореме о делении целых чисел с остатком любое число $k \in \mathbb{Z}$ можно представить в виде $k = nq + r$, где $r \in \{0, 1, ..., n-1\}$. Так как $z_{nq+r} = z_r$, то среди чисел (17) будет n различных при $k = 0, 1, ..., n-1$. □

Следствие 5.2.1 *Корни n-ой степени из 1 выражаются формулой*

$$\varepsilon_k = \cos\frac{2\pi k}{n} + i\sin\frac{2\pi k}{n}, \quad k = 0, 1, ..., n-1 \tag{18}$$

Они расположены в вершинах правильного n-угольника, вписанного в окружность с центром в начале координат и радиуса 1.

Замечание. В силу знаменитой формулы Эйлера

$$e^{i\varphi} = \cos\varphi + i\sin\varphi \tag{19}$$

любое комплексное число, представленное в тригонометрической форме, можно также представить также в показательной форме

$$z = re^{i\varphi}. \tag{20}$$

Упражнения

1. Представить в тригонометрической форме комплексные числа: а) $1-\sqrt{3}i$; б) $-\sqrt{3}+i$;в) $1-i$;а) $-1-\sqrt{3}i$.

2. Найдите аргументы следующих комплексных чисел: а) $cos\frac{\pi}{6}-i\sin\frac{\pi}{6}$; б) $-cos\frac{\pi}{3}+i\sin\frac{\pi}{3}$; в) $sin\frac{\pi}{3}+i\cos\frac{\pi}{3}$.

3. Возведите в степень: а) $\left(\frac{4}{\sqrt{3}+i}\right)^{13}$; б) $\left(\frac{\sqrt{3}}{2}-\frac{1}{2}i\right)^{100}$; в) $\left(-\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{2}}i\right)^{99}$.

4. Решите уравнение $z^4=\frac{-1+i}{1-i\sqrt{3}}$.

5. Решите уравнение $z^6=\frac{-\sqrt{3}+i}{-2-2i}$.

6. Решите уравнение $z^8=\frac{-1+i\sqrt{3}}{1+i\sqrt{3}}$.

6 Кольца и поля вычетов

6.1 Сравнение целых чисел по модулю

Зафиксируем натуральное число $m \in \mathbb{N}$

Определение. *Два целых числа a, b называются сравнимыми по модулю m, если они при делении на m дают одинаковые остатки.*

Кратко записывают $a \equiv b \ (mod\ m)$, и называют это соотношение сравнением.

Теорема 6.1.1 (Критерий сравнимости) *Для любых целых чисел a, b*

$$a \equiv b \ (mod\ m) \Longleftrightarrow m \mid (a - b)$$

Доказательство. Необходимость. Пусть $a \equiv b \ (mod\ m)$, тогда $a = m\ a_1 + r$ и $b = m\ b_1 + r$. Найдем их разность: $a - b = m\ (a_1 - b_1)$. Следовательно, $m \mid (a - b)$.

Достаточность. Пусть $m \mid (a - b)$ и $a = m\ a_1 + r_1, \quad b = m\ b_1 + r_2$ и $0 \leq r_2 \leq r_1 < m$, тогда $a - b = m\ (a_1 - b_1) + (r_1 - r_2)$. Но по условию $m \mid (a - b)$, следовательно, $r_1 = r_2$. □

Напомним следующее определение

Определение. *Отношение $\sim$, определенное на множестве M, называется отношением эквивалентности, если оно обладает следующими свойствами:*

1. рефлексивно, то есть для всех $a \in M$ выполнено $a \sim a$;

2. симметрично, то есть из того, что $a \sim b$ следует, что $b \sim a$;

3. транзитивно, то есть если $a \sim b$ и $b \sim c$, то тогда и $b \sim c$.

Отношение эквивалентности $\sim$ позволяет разбить все элементы множества M на не пересекающиеся классы так, чтобы элементы, принадлежащие одному классу находились в данном отношении $\sim$, а элементы из двух разных классов нет.

Теорема 6.1.2 *1. Отношение сравнимости целых чисел по модулю m является отношением эквивалентности на $\mathbb{Z}$.*

2. Для любых целых чисел $a, b, c, d \in \mathbb{Z}$

$$a \equiv b \ (mod\ m), \quad c \equiv d \ (mod\ m) \Longrightarrow a * c \equiv b * d \ (mod\ m),$$

где $*$ – любая из операций $+, -, \cdot$ (т. е. сравнения можно почленно складывать, вычитать и перемножать)

3. Если d – общий делитель чисел $a,\ b,\ m$, то

$$a \equiv b\ (mod\ \ m) \Longleftrightarrow a/d \equiv b/d (mod\ \ m/d),$$

4. Если d – общий делитель чисел a, b и $(d, m) = 1$, то

$$a \equiv b\ (mod\ \ m) \Longleftrightarrow a/d \equiv b/d\ (mod\ \ m),$$

Доказательство. 1. Следует непосредственно из определения.

2–4. Легко доказать, применяя критерий сравнимости.

В качестве примера докажем 3. Так как d является общим делителем чисел $a,\ b,\ m$, то существуют такие числа $a_1,\ b_1,\ m_1$, что: $a = a_1 d,\ b = b_1 d, m = m_1 d$. Отсюда получим, что

$$m \mid (a - b) \Longleftrightarrow m_1 \mid (a_1 - b_1)d \Longleftrightarrow m_1 \mid (a_1 - b_1).$$

Следствие 6.1.1 *Для любых целых чисел $a, b, c \in \mathbb{Z}$ и натурального k*

$$a \equiv b\ (mod\ \ m) \Longrightarrow a * c \equiv b * c\ (mod\ \ m),\quad a^k \equiv b^k\ (mod\ \ m)$$

где $*$ – любая из операций $+, -, \cdot$.

Упражнения

1. Найдите остаток при делении 15^{231} на 14.
2. Найдите остаток при делении 15^{231} на 16.
3. Найдите остаток при делении $12^{1231} + 14^{4234}$ на 13.
4. Найдите остаток при делении 208^{208} на 23.
5. Найдите две последние цифры числа 2^{341}.
6. Найдите две последние цифры числа 289^{289}.

6.2 Кольцо классов вычетов

Так как отношение сравнимости по модулю m является отношением эквивалентности на множестве целых чисел $\mathbb{Z}$, то $\mathbb{Z}$ разбивается на не пересекающиеся классы чисел, сравнимых по модулю m.

Определение. *Класс всех целых чисел, сравнимых с числом a по модулю m, называется классом вычетов по модулю m и обозначается $[a]_m$ или $\hat{a}$. Множество всех классов вычетов по модулю m будем обозначать $\mathbb{Z}_m$.*

Из определения следует, что $[a]_m = [b]_m$ в том и только том случае, если $a \equiv b \ (mod \ m)$.

Так как количество различных остатков от деления целых чисел на m равно m: $0, 1, 2, ..., m-1$, то и и число классов вычетов по модулю m равно m, и

$$\mathbb{Z}_m = \{[0]_m, [1]_m, ..., [m-1]_m\}.$$

Определим на множестве $\mathbb{Z}_m$ операции сложения и умножения следующим образом:

$$[a]_m + [b]_m = [a+b]_m;$$
$$[a]_m \cdot [b]_m = [a \cdot b]_m.$$

Из теоремы 6.1.2 следует, что это определение не зависит от выбора представителей в каждом из классов.

Теорема 6.2.1 *Множество $\mathbb{Z}_m$ всех классов вычетов по модулю m с определенными выше операциями является коммутативным кольцом с единицей.*

Кольцо $\mathbb{Z}_m$ называется *кольцом вычетов по модулю m.*

Заметим, что $\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z}$, где $m\mathbb{Z} = \{mz \mid z \in \mathbb{Z}\}$ – идеал в $\mathbb{Z}$, порожденный элементом m.

Теорема 6.2.2 *В кольце $\mathbb{Z}_m$ каждый элемент $[a]_m \neq [0]_m$ или обратим, или делитель нуля, причем:*

1. $[a]_m$ - обратим тогда и только тогда $(a, m) = 1$;

2. $[a]_m$ - делитель нуля тогда и только тогда $(a, m) \neq 1$.

Доказательство. 1. Пусть $(a, m) = 1$, тогда существуют такие числа u и v, что $1 = au + mv$ (теорема 4.2.2). Отсюда $au \equiv 1 (mod \ m)$, то есть $[a]_m \cdot [u]_m = [1]_m$. Следовательно, $[a]_m$ - обратим и $[a]_m^{-1} = [u]_m$.

2. Пусть $(a, m) = d > 1$. Тогда $a = da_1$, где $a_1 \in \mathbb{Z}$ и

$$[a]_m \cdot \left[\frac{m}{d}\right]_m = \left[a\frac{m}{d}\right]_m = [a_1 m]_m = [0]_m.$$

Так как $[a]_m \neq [0]_m$ по условию, $\left[\frac{m}{d}\right]_m \neq [0]_m$ в силу неравенства $d > 1$, то $[a]_m$ - делитель нуля.

Доказательство пунктов 1 и 2 в обратную сторону предоставляется читателю. □

Следствие 6.2.1 *Если p – простое число, то кольцо $\mathbb{Z}_p$ вычетов по модулю p является полем.*

Поле $\mathbb{Z}_p$ называется *полем вычетов по модулю p.*

Следствие 6.2.2 *Порядок мультипликативной группы кольца* $\mathbb{Z}_m$ *равен* $\varphi(m)$.

Из следствия 1.2.1 к теореме Лагранжа получим одно из замечательных свойств функции Эйлера.

Теорема 6.2.3 (Эйлера) *Если натуральные числа* a *и* m *взаимно просты, то* $a^{\varphi(m)} \equiv 1(mod\ m)$.

Следствие 6.2.3 (малая теорема Ферма) *Если* p *– простое число, и* $a \in \mathbb{Z}$ *имеют место следующие сравнения:*

a) $a^{p-1} \equiv 1(mod\ p)$ *при* $a \not\vdots p$,

b) $a^p \equiv a(mod\ p)$ *при любом* a.

Упражнения

1. Доказать теорему 6.2.3.
2. Перечислить обратимые элементы кольца $\mathbb{Z}_{30}$.
3. Пусть p – простое. Найти характеристику поля $\mathbb{Z}_p$.

7 Элементы теории многочленов

7.1 Кольцо многочленов над кольцом с единицей

Пусть R - кольцо с единицей.

Определение. *Выражение*

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0 = \sum_{i=0}^{n} a_i x^i$$

где $a_i \in R$, называется многочленом от x над кольцом R. Элементы a_i называются его коэффициентами, a_0 – свободным членом.

Определение. *Степенью* $\deg f(x)$ *многочлена $f(x)$ называется наибольшее натуральное n, для которого $a_n \neq 0$; при этом a_n называется старшим коэффициентом, а $f_C = a_n x$ – старшим членом многочлена $f(x)$. Если $f(x) = 0$, то будем считать, что степень многочлена $f(x)$ – не определена.*

Определение. *Многочлен $f(x)$ называется унитарным (или нормированным), если его старший коэффициент равен 1, то есть* $f(x) = x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0$.

Если $R = F$ является полем, то любой многочлен $f(x) \in F[x]$ может быть записан в виде

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0 =$$

$$a_n(x^n + \frac{a_{n-1}}{a_n}x^{n-1} + ... + \frac{a_1}{a_n}x + \frac{a_0}{a_n}) = a_n\ f_1(x),$$

где $f_1(x)$ – унитарный многочлен.

Замечание. Иногда для удобства многочлен $f(x)$ записывается в виде:

$$f(x) = \sum_{i=0}^{\infty} a_i x^i,$$

при этом подразумевается, что в этой сумме лишь конечное число ненулевых коэффициентов a_i.

Множество всех многочленов от переменной x над кольцом R обозначается через $R[x]$.

Два многочлена

$$f(x) = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n, a_n \neq 0,$$

$$g(x) = b_0 + b_1 x + b_2 x^2 + ... + b_m x^m, b_m \neq 0$$

равны, если $m = n$ и $a_i = b_i$ для всех $i = 1, ...n$.

Можно определить сумму и произведение многочленов:

$$f(x) + g(x) = \sum_{i=0}^{n} (a_i + b_i)x^i;$$

$$f(x) \cdot g(x) = a_0b_0 + (a_0b_1 + a_1b_0)x + (a_jb_2 + a_1b_1 + a_2b_0)x^2 + ...+$$
$$+(a_{n-1}b_m + a_nb_{m-1})x^{m+n-1} + a_nb_mx^{m+n}.$$

Отсюда легко следует следующее предложение

Предложение 7.1.1 *Для любых многочленов* $f(x),\ g(x) \in R[x]$ *выполнены свойства:*

а) $\deg\ (f(x) + g(x)) \leq max(\deg\ f(x),\ \deg\ g(x))$, *если все три многочлена* $f(x), g(x), f(x)+g(x)$ *отличны от нуля. При этом неравенство будет строгим тогда и только тогда, когда* $f_C = -g_C$;

б) если в кольце R *нет делителей нуля (в частности, если* R *является полем) и многочлены* $f(x)$ *и* $g(x)$ *отличны от нуля, то* $\deg\ (f(x) \cdot g(x)) = \deg\ f(x) + \deg\ g(x)$.

Теорема 7.1.1 *Множество* $(R[x], +, \cdot)$ *многочленов над кольцом* R *с единицей является кольцом с единицей. Кольцо* $R[x]$ *коммутативно тогда и только тогда, когда коммутативно* R *и содержит делители нуля тогда и только тогда, когда* R *содержит делители нуля.*

Доказательство. Так как $(R, +)$ – абелева группа, то $(R[x], +)$ также является абелевой группой с нулем 0, в которой противоположным элементом для многочлена $f(x) = \sum_{i=0}^{n} a_ix^i$ является многочлен $-f(x) = \sum_{i=0}^{n}(-a_i)x^i$

Дистрибутивность $f(x) \cdot (g(x) + h(x)) = f(x) \cdot g(x) + f(x) \cdot h(x)$ и $(g(x) + h(x)) \cdot (f(x) = g(x) \cdot f(x) + h(x) \cdot f(x)$ проверяется непосредственно.

Легко проверить, что

$$(f(x) \cdot g(x)) \cdot h(x)) = (\sum_{i\geq 0} a_ix^i \sum_{j\geq 0} b_jx^j) \sum_{k\geq 0} c_kx^k =$$

$$= \sum_{i\geq 0}\sum_{j\geq 0}\sum_{k\geq 0}(a_ib_j)c_kx^{i+j+k} =$$

$$= \sum_{i\geq 0}\sum_{j\geq 0}\sum_{k\geq 0} a_i(b_jc_k)x^{i+j+k} = f(x) \cdot (g(x)) \cdot h(x)).$$

Таким, образом, $(R[x], +, \cdot)$ является кольцом.

Единицей кольца $(R[x], +, \cdot)$, очевидно, является многочлен $f(x) = x^0 = 1$.

Если R коммутативно, то коммутативность $R[x]$ доказывают равенства:

$$f(x) \cdot g(x) = \sum a_i b_j x^{i+j} = \sum b_j a_i x^{j+i} = g(x) \cdot f(x).$$

Если же $a \cdot b \neq b \cdot a$ для некоторых $a,\ b \in R$, то $ax^0 \cdot bx^0 \neq bx^0 \cdot ax^0$ в кольце $R[x]$.

Если в R нет делителей нуля, то из предложения 7.1.1 б) следует, что для любых ненулевых $f(x),\ g(x) \in R[x]$ справедливы соотношения $\deg(f(x) \cdot g(x)) = \deg f(x) + \deg g(x)$ и поэтому $f(x) \cdot g(x) \neq 0$. □

Упражнения

1. Найти степень каждого многочлена и их суммы: $2x^3 + x^2 - 3x + 1$; $-2x^3 + 3x^2 - 3$.

2. Найти степень каждого многочлена и их произведения: $2x - 3$; $x^3 - x^2 + 2x - 1$.

3. Чему равна степень нулевого многочлена ?

7.2 Многочлены над полем. Теорема о делении с остатком

Пусть теперь $R = F$ – поле и $F[x]$ – кольцо многочленов от одной переменной над полем F. Из теоремы 7.1.1 следует, что $F[x]$ является коммутативным кольцом с единицей без делителей нуля.

Определение. *Говорят, что многочлен $f(x)$ делится с остатком на многочлен $g(x)$, если $f(x) = g(x)\ q(x) + r(x)$ для некоторых многочленов $q(x), r(x) \in F[x]$, причем $\deg\ r(x) < \deg\ g(x)$ или $r(x) = 0$.*

Многочлены $q(x)$ и $r(x)$ называются соответственно неполным частным и остатком.

Если $r(x) = 0$, то говорят, что многочлен $f(x)$ делится на $g(x)$. Обозначается $f(x) \vdots g(x)$ или $g(x) \mid f(x)$ ($g(x)$ делит $f(x)$).

Справедлива следующая теорема

Теорема 7.2.1 *Для любых многочленов $f(x), g(x) \in F[x]$ существуют единственные многочлены $q(x)$ и $r(x)$, удовлетворяющие условию $\deg\ r(x) < \deg\ g(x)$ или $r(x) = 0$, такие, что*

$$f(x) = g(x)\ q(x) + r(x) \tag{21}$$

Доказательство. Аналогично доказательству теоремы о делении с остатком в кольце целых чисел $\mathbb{Z}$ и проводится индукцией по степени многочлена $f(x)$. $\square$

Заметим, что на практике, чтобы разделить многочлен $f(x) = 3x^5 + 2x^4 - x^2 + x + 2$ на многочлен $g(x) = x^3 - x + 1$ с остатком применяют хорошо известный метод деления "уголком":

$$\begin{array}{rrrr|l}
3x^5 + 2x^4 & - x^2 & +x + 2 & & x^3 - x + 1 \\
\cline{5-5}
3x^5 & -3x^3 + 3x^2 & & & 3x^2 + 2x + 3 \\
\cline{1-3}
 & 2x^4 + 3x^3 - 4x^2 & + x & & \\
 & 2x^4 & - 2x^2 + 2x & & \\
\cline{2-4}
 & & 3x^3 - 2x^2 & - x + 2 & \\
 & & 3x^3 & - 3x + 3 & \\
\cline{3-4}
 & & & -2x^2 + 2x - 1 &
\end{array}$$

Здесь неполное частное и остаток соответственно равны $q(x) = 3x^2 + 2x + 3$ и $r(x) = -2x^2 + 2x - 1$.

Определение. *Значением многочлена*

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + ... + a_1 x + a_0$$

с коэффициентами из поля F в точке $\alpha \in F$, называется элемент

$$f(\alpha) = a_n \alpha^n + a_{n-1} \alpha^{n-1} + ... + a_1 \alpha + a_0 \in F.$$

Элемент $\alpha \in F$ называется корнем многочлена $f(x)$, если $f(\alpha) = 0$.

Теорема 7.2.2 (Безу[1]) *Пусть $f(x) \in F[x]$ – многочлен, $\alpha \in F$ – произвольные. Тогда остаток от деления $f(x)$ на $x - \alpha$ равен $f(\alpha)$.*

Из теоремы Безу можно вывести следующее полезное следствие.

Следствие 7.2.1 *Элемент α поля F является корнем многочлена $f(x) \in F[x]$ тогда и только тогда, когда $f(x) \vdots x - \alpha$.*

Упражнения

1. Доказать теорему 7.2.2 и следствие 7.2.1.
2. Разделить с остатком многочлен $x^4 + 2x^2 + 20x + 7$ на $x^2 - 2x + 3$.
3. Найти все корни многочлена $x^3 - 3x + 2$.

7.3 Наибольший общий делитель многочленов. Алгоритм Евклида

Пусть $F[x]$ – кольцо многочленов над полем F.

Аналогично кольцу целых чисел $\mathbb{Z}$ можно определить наибольший общий делитель двух многочленов $f(x)$ и $g(x)$.

Ясно, что если в качестве наибольших общих делителей рассматривать только нормированные многочлены, то *НОД* определяется однозначно.

Два многочлена $f(x), g(x) \in F[x]$ называются взаимно простыми, если их наибольший общий делитель является многочленом нулевой степени.

Так же, как и для кольца целых чисел $\mathbb{Z}$ существует алгоритм Евклида для нахождения наибольшего общего делителя двух многочленов и его линейного представления.

Теорема 7.3.1 (Линейное представление НОД) *Пусть $f(x)$ и $g(x)$ не равные нулю одновременно многочлены с коэффициентами из поля F. Тогда существует единственный нормированный наибольший общий делитель $d(x)$ многочленов $f(x)$ и $g(x)$ и многочлены $u(x),\ v(x) \in F[x]$, такие, что*

$$d(x) = u(x)f(x) + v(x)g(x).$$

Нормированный наибольший общий делитель многочленов $f(x)$ и $g(x)$ обозначается НОД $(f(x), g(x))$ или $(f(x), g(x))$.

Упражнения

1. Доказать теорему 7.3.1.

2. Найти наибольший общий делитель многочленов $f(x) = 2x^5 + x^4 - 12x^3 - 6x^2 - 34x + 33$, $g(x) = 2x^3 - x^2 - 17x + 12$ и его линейное представление.

3. Найти наибольший общий делитель многочленов $f(x) = x^5 - x^4 - 4x^3 - 2x^2 - 2x + 8$, $g(x) = x^3 - 2x^2 - 4x + 5$ и его линейное представление.

7.4 Неприводимые многочлены

Определение. *Многочлен $f(x) \in F[x]$ называется неприводимым (или простым) над полем F, если* $\deg f(x) > 0$ *и его нельзя представить в виде произведения двух многочленов меньшей степени с коэффициентами из F.*

Многочлен называется приводимым или составным, если его можно представить в виде произведения двух многочленов меньшей степени.

Пример. Если $\deg f(x) = 1$, то многочлен $f(x)$ является неприводимым над любым полем, так как меньшая степень – нулевая, а произведение двух констант является константой.

Многочлен $f(x) = x^2 + 1$ является неприводимым над полем $\mathbb{R}$ и приводимым над полем $\mathbb{C}$, так как

$$f(x) = x^2 + 1 = (x + i)(x - i).$$

Теорема 7.4.1 *Любой нормированный многочлен $f(x) \in F[x]$ положительной степени является либо неприводимым, либо единственным способом (с точностью до перестановки сомножителей) представим в виде произведения нормированных неприводимых многочленов.*

Доказательство. Рассмотрим многочлен $f(x) \in F[x]$ положительной степени. Доказательство будем проводить индукцией по степени многочлена $n = \deg f(x)$.

При $n = 1$ многочлен $f(x)$ имеет первую степень – неприводим.

Предположим, что утверждение теоремы справедливо для любого многочлена положительной степени, меньшей n и $\deg f(x) = n$.

Предположим, что многочлен $f(x)$ приводим. Тогда $f(x) = f_1(x)f_2(x)$, где $f_1(x)$ и $f_2(x)$ – многочлены положительной степени, меньшей n.

Так как степени многочленов $f_1(x)$ и $f_2(x)$ меньше n, для них справедливо предположение индукции и, следовательно,

$$f_1(x) = p_1(x)p_2(x)\cdots p_k(x), \quad f_2(x) = q_1(x)q_2(x)\cdots q_m(x),$$

где многочлены $p_i(x), q_j(x)$ – неприводимые нормированные.

Отсюда получим, что

$$f(x) = f_1(x)\ f_2(x) = p_1(x)\ p_2(x)\cdots p_k(x)\ q_1(x)\ q_2(x)\cdots q_m(x).$$

Единственность разложения следует из того, что если $p(x),\ q(x)$ – неприводимые нормированные многочлены, то либо $p(x) = q(x)$, либо $(p(x), q(x)) = 1$. □

Упражнения

1. Разложить многочлен $x^4 + 1$ на неприводимые сомножители над полями $\mathbb{Q}, \mathbb{R}, \mathbb{C}$.

2. Разложить многочлен $x^4 + x^2 + 1$ на неприводимые сомножители над полями $\mathbb{Q}, \mathbb{R}, \mathbb{C}$.

3. Разложить многочлен $(x-1)x(x+1)(x+2) - 24$ на неприводимые сомножители над полями $\mathbb{Q}, \mathbb{R}, \mathbb{C}$.

7.5 Схема Горнера. Основная теорема алгебры

Определение. *Поле F называется алгебраически замкнутым, если любой многочлен $f(x) \in F[x]$ положительной степени имеет хотя бы один корень $\alpha \in F$.*

Замечание. Поле действительных чисел $\mathbb{R}$ не является алгебраически замкнутым, поскольку многочлен $f(x) = x^2 + 1 \in \mathbb{R}[x]$ не имеет действительных корней.

Рассмотрим схему Горнера[1] деления многочлена

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0$$

на двучлен $x - \alpha$.

Пусть $f(x) = (x - \alpha)q(x) + r$, где $q(x) = b_{n-1}x^{n-1} + ... + b_1 x + b_0$. Тогда

$$a_n x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0 = (x - \alpha)(b_{n-1}x^{n-1} + ... + b_1 x + b_0) + r.$$

Отсюда получим:

$$\begin{array}{ccc} a_n & = & b_{n-1}, \\ a_{n-1} & = & b_{n-2} - \alpha b_{n-1}, \\ \dots & \dots & \dots, \\ a_1 & = & b_0 - \alpha b_1, \\ a_0 & = & r - \alpha b_0. \end{array}$$

Найдем отсюда b_i и запишем их в таблицу, которая называется схемой **Горнера**:

	a_n	a_{n-1}	...	a_1	a_0
α	$b_{n-1} = a_n$	$b_{n-2} = \alpha b_{n-1} + a_{n-1}$	...	$b_0 = \alpha b_1 + a_1$	$r = \alpha b_0 + a_0$

Определение. *Кратностью корня $\alpha \in F$ многочлена $f(x) \in F[x]$ называется натуральное число k такое, что многочлен $(x - \alpha)^k$ является делителем многочлена $f(x)$, а многочлен $(x - \alpha)^{k+1}$ – нет.*

Если кратность корня α равна 1, то α называется простым корнем.

Определение. *Формальной производной многочлена*

$$f(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0 \in F[x]$$

называется многочлен

$$f'(x) = n\ a_n x^{n-1} + (n-1)a_{n-1}x^{n-2} + ... + a_1 \in F[x].$$

[1] Уильям Джордж Горнер – английский математик (1786-1837).

Заметим, что при этом все свойства производной при таком определении остаются верными, то есть:
(1) $(f(x)+g(x))' = f'(x)+g'(x)$;
(2) $(f(x)\ g(x))' = f'(x)\ g(x) + f(x)\ g'(x)$.

Теорема 7.5.1 *Корень α многочлена $f(x)$ с коэффициентами из поля F является простым в том и только том случае, если α не является корнем производной $f'(x)$.*

Доказательство. Пусть α является корнем кратности k для многочлена $f(x)$, тогда $f(x) = (x-\alpha)^k\ g(x)$ для некоторого многочлена $g(x)$, такого, что $g(\alpha) \neq 0$.

Производная представима в виде

$$f'(x) = k\ (x-\alpha)^{k-1}\ g(x) + (x-\alpha)^k\ g'(x).$$

При $k = 1$, получим $f'(x) = g(x) + (x-\alpha)\ g'(x)$, и, следовательно, $f'(\alpha) = g(\alpha) \neq 0$.

Если же $k > 1$, то $f'(\alpha) = k\ (\alpha-\alpha)^{k-1}\ g(\alpha) + (\alpha-\alpha)^k\ g'(\alpha) = 0$. □

Теорема 7.5.2 (Гаусс[1]) *Над полем комплексных чисел $\mathbb{C}$ любой многочлен положительной степени имеет хотя бы один корень (другими словами, поле $\mathbb{C}$ является алгебраически замкнутым).*

Эта теорема, называемая основной теоремой алгебры, не имеет чисто алгебраического доказательства. Она может быть доказана с использованием понятия непрерывности функции одного или двух действительных переменных.

Следствие 7.5.1 *Любой многочлен n-ой степени имеет в $\mathbb{C}$ ровно n корней с учетом их кратности.*

Доказательство. Пусть $f(x) \in \mathbb{C}[x]$ и $\deg\ f(x) = n$. Из основной теоремы алгебры следует, что он имеет по крайней мере один корень, пусть $\alpha \in \mathbb{C}$ – его корень. Тогда по теореме Безу $f(x) = (x-\alpha)g(x)$ и $\deg\ g(x) = n-1$, если $n-1 > 0$, то многочлен $g(x)$ тоже имеет по крайней мере один корень и т. д. Продолжая этот процесс, получим, что многочлен $f(x)$ имеет ровно n корней, некоторые из которых могут совпадать. □

Упражнения

1. Разложить многочлен $x^4 + 16$ на сомножители первой степени над полем $\mathbb{C}$.

2. Разложить многочлен $x^4 - 10x^2 + 1$ на сомножители первой степени над полем $\mathbb{C}$.

3. Разложить многочлен $(x^2-x+1)(x^2-x+2)-12$ на сомножители первой степени над полем $\mathbb{C}$.

7.6 Многочлены над полем действительных чисел

Теорема 7.6.1 *В кольце $\mathbb{R}[x]$ неприводимыми являются все многочлены первой степени, многочлены второй степени не имеющие действительных корней и только они.*

Доказательство. Неприводимость указанных многочленов очевидна. Покажем, что других неприводимых многочленов в кольце $\mathbb{R}[x]$ нет.

Пусть $f(x) = a_n x^n + a_{n-1}x^{n-1} + ... + a_1 x + a_0 \in \mathbb{R}[x]$ – неприводимый многочлен степени $n > 1$. Тогда он не имеет корней в $\mathbb{R}$, но так как $\mathbb{R} \subseteq \mathbb{C}$, то из основной теоремы алгебры он имеет по крайней мере один комплексный корень $\alpha \in \mathbb{C}$.

Так как $\alpha \notin \mathbb{R}$, то $\alpha \neq \overline{\alpha}$. Покажем, что $\overline{\alpha}$ также является корнем многочлена $f(x)$. Так как $\overline{a_i} = a_i$ для всех $i = 0, 1, ..., n$ получим

$$f(\overline{\alpha}) = a_n\overline{\alpha}^n + a_{n-1}\overline{\alpha}^{n-1} + ... + a_1\overline{\alpha} + a_0 =$$
$$= \overline{a_n}\ \overline{\alpha}^n + \overline{a_{n-1}}\ \overline{\alpha}^{n-1} + ... + \overline{a_1}\ \overline{\alpha} + \overline{a_0} = \overline{f(\alpha)} = \overline{0} = 0.$$

По теореме Безу $f(x) \vdots x-\alpha$ и $f(x) \vdots x-\overline{\alpha}$. Многочлены $x-\alpha$ и $x-\overline{\alpha}$ являются различными нормированными неприводимыми сомножителями многочлена $f(x)$.

Следовательно, $f(x) \vdots (x-\alpha)(x-\overline{\alpha})$ над полем $\mathbb{C}$.

Отметим, что $(x-\alpha)(x-\overline{\alpha}) = x^2 + px + q$, где $p = -(\alpha + \overline{\alpha})$, $q = \alpha\overline{\alpha} \in \mathbb{R}$. То есть многочлен $(x-\alpha)(x-\overline{\alpha})$ имеет действительные коэффициенты.

Следовательно, $f(x) \vdots (x-\alpha)(x-\overline{\alpha})$ над полем $\mathbb{R}$ (при делении уголком не могут возникнуть коэффициенты не являющиеся действительными).

Из неприводимости $f(x)$ получаем равенство

$$f(x) = a(x^2 + px + q), \quad a \in \mathbb{R}.$$

Следовательно, $f(x)$ – многочлен второй степени, не имеющий действительных корней. □

Следствие 7.6.1 *Любой многочлен нечетной степени с действительными коэффициентами имеет действительный корень.*

Упражнения

1. Разложить многочлен $x^4 + 3x^2 + 1$ на неприводимые сомножители над полем $\mathbb{R}$.

2. Какие из многочленов неприводимы над $\mathbb{R}$: а) $x^6 - 3x^2 + 3$; б) $x^3 + x + 1$; в) $4x^2 + 11x + 8$.

7.7 Кольцо многочленов от нескольких переменных

Если R – коммутативное кольцо с единицей, то кольцо многочленов $R_1 = R[x]$ от одной переменной над кольцом R является коммутативным кольцом с единицей (теорема 7.1.1).

Кольцо $R_1[y] = R[x, y]$ называется *кольцом многочленов от двух переменных x и y над кольцом R.*

Ясно, что любой многочлен из кольца $R[x, y]$ может быть записан в виде

$$f(x, y) = \sum_{j=0}^{n} \sum_{i=0}^{m} a_{ij} x^i y^j,$$

где $a_{ij} \in R$ или в виде бесконечной суммы

$$f(x, y) = \sum_{j\geq 0} \sum_{i\geq 0} a_{ij} x^i y^j = \sum_{(i,j)} a_{ij} x^i y^j,$$

имеющей лишь конечное число ненулевых слагаемых.

Элементы a_{ij} называются *коэффициентами многочлена $f(x, y)$.*

При этом, так же как и в случае многочленов от одной переменной, многочлены $f(x, y) = \sum_{(i,j)} a_{ij} x^i y^j$ и $g(x, y) = \sum_{(i,j)} b_{ij} x^i y^j$ считаются равными, если $a_{ij} = b_{ij}$ для всех $i \geq 0, j \geq 0$.

Аналогично индуктивным методом строится кольцо многочленов от произвольного конечного числа переменных.

Если $R[x_1, ..., x_{n-1}]$ – кольцо многочленов от $n - 1$ переменных $x_1, ..., x_{n-1}$ над кольцом R, то кольцо многочленов

$$R[x_1, ..., x_n] = R[x_1, ..., x_{n-1}][x_n]$$

называется *кольцом многочленов от n переменных $x_1, ..., x_n$ над кольцом R.*

Каждый элемент $f(x_1, ..., x_n) \in R[x_1, ..., x_n]$ может быть представлен в виде

$$f(x_1, ..., x_n) = \sum_{i_1=0}^{m_1} ... \sum_{i_n=0}^{m_n} a_{i_1...i_n} x_1^{i_1} ... x_n^{i_n}, \quad a_{i_1...i_n} \in R.$$

Представление многочлена в таком виде называется *канонической записью*, а слагаемые $a_{i_1...i_n} x_1^{i_1} ... x_n^{i_n}$ – *одночленами* или *мономами.*

Определение. *Степенью одночлена $a x_1^{i_1} ... x_n^{i_n}$ называется число $i_1 + ... + i_n$, а степенью одночлена $a x_1^{i_1} ... x_n^{i_n}$ по переменной x_s называется число i_s.*

Степенью многочлена $f(x_1, ..., x_n) = \sum_{i_1=0}^{m_1} ... \sum_{i_n=0}^{m_n} a_{i_1...i_n} x_1^{i_1} ... x_n^{i_n}$ называется наибольшая степень входящих в него одночленов.

Степень нулевого многочлена не определена.

Кольцо многочленов $R[x_1, ..., x_n]$, как и кольцо многочленов от одного переменного, сохраняет некоторые свойства исходного кольца R.

Теорема 7.7.1 *Кольцо $R[x_1, ..., x_n]$ коммутативно тогда и только тогда, когда коммутативно кольцо R, и содержит делители нуля тогда и только тогда, когда R содержит делители нуля.*

Доказательство. При $n = 1$ это было доказано. Доказательство в общем случае легко проводится индукцией по числу переменных n. □

Упражнения

1. Найти степень многочлена $x_1^2x_2x_3^3 - 7x_1^3x_2^3 + 6x_2^5 - 3$. Какой одночлен имеет наибольшую степень? Какой одночлен имеет наибольшую степень по переменной: а) x_1; б) x_2?

2. Пусть многочлены $f,\ g \in R[x_1, ..., x_n]$ – произвольные ненулевые. Доказать следующие свойства степени:

   а) $\deg\,(f + g) \leq max(\deg\,f,\ \deg\,g)$, если $f + g \neq 0$;

   б) если в кольце R нет делителей нуля (в частности, если R является полем), то $\deg\,(f \cdot g) = \deg\,f + \deg\,g$.

3. Показать, что многочлен $x^2 + y^2 + 1$ нельзя разложить в произведение многочленов первой степени даже с комплексными коэффициентами.

8 Элементы теории полей

8.1 Расширения полей

Определение. *Пусть F, G – два поля и $F \subset G$. Поле G называется расширением поля F, а F – подполем поля G.*

Примеры.
1. Поле $\mathbb{C}$ является расширением поля $\mathbb{R}$, а $\mathbb{R}$, в свою очередь, – расширением поля $\mathbb{Q}$.
2. В любом поле F есть хотя бы одно подполе – само поле F.

Определение. *Поле называется простым, если в нем нет подполей, кроме него самого.*

Всякое поле F является расширением своего простого подполя, порожденного единицей. Простое подполе характеристики p изоморфно $\mathbb{Z}_p$, а характеристики нуль – $\mathbb{Q}$.

Определение. *Пусть G – расширение поля F и T – подмножество поля G. Пересечение всех подполей поля G, содержащих F и T, называется расширением поля F, порожденным подмножеством T и обозначается $F(T)$.*

Если множество T состоит из одного элемента, расширение $F \subseteq F(T)$ называется простым.

Пример. В поле $\mathbb{C}$ расширение поля $\mathbb{R}$, порожденное элементом $i \in \mathbb{C}$, совпадает с полем $\mathbb{C}$. Действительно, $\mathbb{R}(i)$ – подполе поля $\mathbb{C}$, содержащее $\mathbb{R}$ и i. Поэтому $\mathbb{R}(i)$ содержит все элементы вида $a + bi,\ a, b \in \mathbb{R}$. Следовательно, $\mathbb{R}(i) = \mathbb{C}$.

Упражнения

1. Пусть $F \subseteq G$ – расширение полей. Показать, что множество G является векторным пространством над F по отношению к операциям сложения и умножения, заданным в поле G.

2. Пусть $z \in \mathbb{C} \setminus \mathbb{R}$. Что можно сказать про расширение $\mathbb{R}(z)$ поля $\mathbb{R}$?

3. Найти базис поля $\mathbb{Q}(\sqrt{2}, \sqrt{3})$ как векторного простанства над $\mathbb{Q}$.

8.2 Алгебраические и конечные расширения

Определение. *Расширение полей $F \subseteq G$ называется конечным, если G является конечномерным векторным пространством над F. Размерность $\dim_F G$ называется степенью расширения полей и обозначается $\dim_F G = [G : F]$.*

Если G – бесконечномерное векторное пространство над F, то говорят о расширении бесконечной степени: $[G : F] = \infty$.

Пример. В силу предыдущего примера $\mathbb{C} = \mathbb{R}(i)$. А поскольку $\{1,\ i\}$ образуют базис векторного пространства $\mathbb{C}$ над $\mathbb{R}$, то $[\mathbb{C} : \mathbb{R}] = 2$.

Определение. *Пусть F, G – поля и $F \subseteq G$. Элемент $\alpha \in G$ называется алгебраическим над полем F, если α является корнем некоторого многочлена $p(x)$ положительной степени из $F[x]$. Если такого многочлена не существует, то α называется трансцендентным над полем F.*

Многочлен наименьшей степени из $F[x]$, корнем которого является α, называется минимальным многочленом алгебраического элемента α, а его степень степенью α.

Комплексное число, которое является алгебраическим элементом расширения $\mathbb{Q} \subseteq \mathbb{C}$ называется алгебраическим.

Определение. *Расширение G поля F называется алгебраическим, если все элементы поля G являются алгебраическими над полем F, и трансцендентным, если в G существует элемент, не являющийся алгебраическим над полем F.*

Пример. Элемент x поля рациональных функций $F(x)$ трансцендентен над полем F, следовательно $F(x)$ – трансцендентное расширение поля F. Поле $\mathbb{C}$ является алгебраическим расширением поля $\mathbb{R}$, так как произвольный элемент $a + bi \in \mathbb{C}$ является корнем ненулевого многочлена $(x - a)^2 + b^2 \in \mathbb{R}[x]$.

Приведем без доказательства теорему о строении простого алгебраического расширения поля.

Теорема 8.2.1 *Пусть $F \subseteq K$ – поля, элемент $\alpha \in K$ – алгебраический степени n над F. Тогда простое алгебраическое расширение $F \subseteq F(\alpha)$ имеет размерность n, а его базисом являются элементы $1, \alpha, \alpha^2, ..., \alpha^{n-1}$.*

Упражнения

1. Показать, что конечное расширение полей является алгебраическим.

2. Пусть $F \subseteq G, G \subseteq K$ – конечные расширения полей. Доказать, что расширение $F \subseteq K$ также является конечным и его степень равна $[K : F] = [K : G][G : F]$.

3. Показать, что алгебраические числа образуют подполе поле комплексных чисел. Какова его мощность?

4. Пусть $F \subseteq G$ – расширение полей, элемент $\alpha \in G$ – алгебраический над полем F, $p(x)$ – минимальный многочлен элемента α, $f(x) \in F[x]$ – произвольный. Доказать следующие свойства минимальных многочленов:

   а) $f(\alpha) = 0 \Rightarrow f(x) \vdots p(x)$;

   б) многочлен $p(x)$ – неприводимый над полем F;

   в) если $f(x)$ – неприводимый над полем F и $f(\alpha) = 0$, то многочлен $f(x)$ является минимальным многочленом элемента α.

5. Найти минимальные многочлены и степени следующих алгебраических чисел: а) $\sqrt{2}$; б) $\sqrt[3]{2}$. в) $\sqrt{2} + \sqrt{5}$.

8.3 Конечные поля

Большую роль в компьютерной алгебре играют конечные поля – поля содержащие конечное множество элементов.

Несложно доказать следующее утверждение.

Лемма 8.3.1 *Характеристика конечного поля – конечна.*

Теорема 8.3.1 *Число элементов конечного поля F равно p^n, где p – простое число, а n – натуральное.*

Доказательство. Рассмотрим гомоморфизм $f : \mathbb{Z} \to F$ заданный формулой $f(z) = z \cdot 1_F$, где 1_F – единица поля F.

Покажем, что отображение f является гомоморфизмом колец.

Действительно, согласно свойствам аддитивной экспоненты (6),

$$f(z_1 + z_2) = (z_1 + z_2) \cdot 1_F = z_1 \cdot 1_F + z_2 \cdot 1_F = f(z_1) + f(z_2),$$

$$f(z_1 z_2) = (z_1 z_2) \cdot 1_F = z_1 \cdot (z_2 \cdot 1_F) = z_1 \cdot (1_F \cdot (z_2 \cdot 1_F)) =$$
$$= (z_1 \cdot 1_F) \cdot (z_2 \cdot 1_F) = f(z_1) f(z_2),$$

где $z_1, z_2 \in \mathbb{Z}$ – произвольные.

Пусть $p = \mathrm{char}\ F$ – характеристика поля F (ее конечность следует из леммы 8.3.1). Из леммы 2.1.1 следует, что число p – простое.

Обозначим через $K = \mathrm{Im}\ f$ – образ кольца целых чисел при отображении f. Известно, что образ кольца при гомоморфизме является кольцом.

Покажем, что $\mathrm{Ker}\ f = p\mathbb{Z}$. Это следует из части 1 теоремы 1.4.1, где $a = 1_F$ – образующий элемент аддитивной циклической группы порядка p.

Из теоремы о гомоморфизме для колец (теорема 2.2.1) следует изоморфизм $\mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z} \simeq K$. Это означает, что подкольцо K поля F является простым полем из p элементов.

Рассмотрим поле F как векторное пространство над полем K. Из конечности поля F следует, что $\dim_K F < \infty$. Пусть $\dim_K F = n$. Тогда существует базис $e_1, ..., e_n$ векторного пространства F над K.

Для каждого $r \in F$ существуют единственные элементы

$$\alpha_1, ..., \alpha_n \in K$$

такие, что $r = \alpha_1 e_1 + ... + \alpha_n e_n$. Поле F содержит столько же элементов сколько существует последовательностей

$$(\alpha_1, ..., \alpha_n), \alpha_1, ..., \alpha_n \in K.$$

Согласно комбинаторному принципу умножения, таких последовательностей p^n. Теорема доказана. □

Приведем без доказательства следующую теорему [4, стр. 382].

Теорема 8.3.2 *Для любого простого p и натурального n существует поле из p^n элементов и все такие поля изоморфны.*

Идея доказательства теоремы 8.3.2 основана на следующем утверждении.

Предложение 8.3.1 *Пусть F – конечное поле и $\mid F \mid = p^n$. Тогда:*
1. для любого элемента $a \in F$ справедливо равенство: $a^{p^n} = a$;
2. Множество элементов поля F совпадает с множеством корней многочлена $f(x) = x^{p^n} - x$;
3. $x^{p^n} - x = \prod_{a_i \in G}(x - a_i)$.

Доказательство. 1. Поскольку любой ненулевой элемент $a \in F$ лежит в мультипликативной группе $F^\star$ поля F, порядок которой равен $p^n - 1$, то для него $a^{p^n - 1} = 1$, и, следовательно, $a^{p^n} = a$. Ясно, что нуль также удовлетворяет данному соотношению.

2. Сразу следует из 1 и того факта, что многочлен степени p^n не может иметь больше, чем p^n корней в поле.

3. Из 1 следует, что любой элемент поля F является корнем уравнения $x^{p^n} - x = 0$. Тогда для любого $a \in F$ получим

$$\begin{gathered}(x^{p^n} - x) = x^{p^n} - x - (a^{p^n} - a) = (x^{p^n} - a^{p^n}) - (x - a) = \\ = (x - a)(x^{p^n - 1} + ax^{p^n - 2} + \cdots + a^{p^n - 2}x + a^{p^n - 1}) - (x - a) = \\ = (x - a)(x^{p^n - 1} + ax^{p^n - 2} + \cdots + a^{p^n - 2}x + a^{p^n - 1} - 1).\end{gathered}$$

Отсюда получим, что многочлен $x^{p^n} - x$ делится на произведение $\prod_{a_i \in G}(x - a_i)$. А поскольку степени этих многочленов одинаковы и старшие коэффициенты равны 1, то они равны. □

Обозначим поле, содержащее q элементов через F_q.

Для построения конечных полей может быть использована следующая теорема, которую мы также приведем без доказательства.

Теорема 8.3.3 *Пусть p – простое и $m(x)$ – неприводимый многочлен степени r над полем F_p. Тогда $F_p[x]/m(x)F_p[x]$ – поле из p^r элементов, содержащее поле F_p и корень полинома $m(x)$.*

Введем обозначение $F_p[x]_{m(x)} = F_p[x]/m(x)F_p[x]$.

Пример. Многочлен $m(x) = x^3+x+1$ является неприводимым многочленом над полем F_2 (третьей степени и не имеет корней в поле F_2). Пусть $\alpha = x + m(x)F_2[x]$ – корень многочлена $m(x)$ в поле $F_2[x]_{m(x)}$. Найдем минимальный многочлен элемента $\beta = \alpha + 1$.

Элемент α удовлетворяет соотношению $\alpha^3 = \alpha + 1$.

Вычислим элементы β^2, β^3. Получим $\beta^2 = (\alpha + 1)^2 = \alpha^2 + 1$, $\beta^3 = (\alpha + 1)(\alpha^2 + 1) = \alpha^3 + \alpha^2 + \alpha + 1 = \alpha + 1 + \alpha^2 + \alpha + 1 = \alpha^2$.

Известно, что если алгебраический элемент α является корнем неприводимого многочлена $m(x)$, то $m(x)$ является минимальным многочленом элемента α [4]. По теореме 8.2.1 элементы $1, \alpha, \alpha^2$ образуют базис поля $F_2[x]_{m(x)}$.

Приравнивая нулю линейную комбинацию элементов $1, \beta, \beta^2, \beta^3$ и решая соответсвующую систему трех однородных линейных уравнений с 4 неизвестными, находим минимальный многочлен $g(x) = x^3 + x^2 + 1$ элемента β.

Приведем без доказательства следующее очевидное и полезное предложение.

Предложение 8.3.2 *Пусть $f(x)$ многочлен второй или третьей степени с коэффициентами из поля F. Многочлен $f(x)$ неприводим тогда и только тогда, когда $f(x)$ не имеет корней в поле F.*

Пример. Построим конечное поле из четырех элементов. Рассмотрим многочлен $m(x) = x^2+x+1$ с коэффициентами из поля F_2. Согласно предложению 8.3.2, многочлен $m(x)$ неприводим над полем F_2.

Построим поле $F_2[x]_{x^2+x+1}$. Обозначим через α образ элемента x в факторкольце $F_2[x]_{x^2+x+1} : \alpha = \hat{x} = x + (x^2 + x + 1)F_2[x]$.

Элемент α является корнем многочлена $m(x) : \alpha^2 + \alpha + 1 = 0$, учитывая равенство $-1 = 1$ в поле F_2, получим

$$\alpha^2 = \alpha + 1. \tag{22}$$

Согласно теореме 8.2.1, о строении простого алгебраического расширения, элементы 1 и α образуют базис поля $F_2[x]_{x^2+x+1}$ как векторного пространства над F_2. Получим $F_2[x]_{x^2+x+1} = \{0; 1; \alpha; \alpha + 1\}$.

Найдем произведения α^2,$\alpha(\alpha+1)$,$(\alpha+1)(\alpha+1)$. Первое получается из формулы (22). Для второго

$$\alpha(\alpha+1) = \alpha^2 + \alpha = \alpha + 1 + \alpha = 1.$$

Окончательно,

$$(\alpha+1)(\alpha+1) = \alpha^2 + 1 = \alpha + 1 + 1 = \alpha.$$

С учетом того, что нуль нейтральный по сложению, а 1 – по умножению, получим таблицу Кэли умножения для кольца $F_2[x]_{x^2+x+1}$:

$\times$	0	1	α	$\alpha+1$
0	0	0	0	0
1	0	1	α	$\alpha+1$
α	0	α	$\alpha+1$	1
$\alpha+1$	0	$\alpha+1$	1	α

Из таблицы умножения видно, что элементы α и $\alpha+1$ являются образующими элементами мультипликативной группы поля

$$F_2[x]_{x^2+x+1} \simeq F_4.$$

Определение. *Назовем группу ненулевых элементов поля F по умножению – мультипликативной группой поля. Обозначим ее F^*.*

Лемма 8.3.2 *Пусть F – поле, $char\ F = p$, элементы $a_1, a_2 \in F$ – произвольные. Тогда $(a_1 + a_2)^p = a_1^p + a_2^p$.*

Лемма 8.3.2 доказывается с помощью бинома Ньютона. Следующая лемма обобщает лемму 8.3.2 и может быть доказана по индукции.

Лемма 8.3.3 *Пусть F – поле, $char\ F = p$, элементы $\alpha_1, ..., \alpha_k \in F$ – произвольные, r – натуральное число. Тогда*

$$(\alpha_1 + ... + \alpha_k)^{p^r} = \alpha_1^{p^r} + ... + \alpha_k^{p^r}.$$

Теорема 8.3.4 *Пусть F – поле, $char\ F = p$, элемент $\alpha \in F$ является корнем многочлена $g(x) \in F_p[x]$. Тогда элемент α^p также является корнем многочлена $g(x)$.*

Доказательство. Пусть $g(x) = a_n x^n + ... + a_1 x + a_0$, где

$$a_0, a_1, ..., a_n \in F_p.$$

Тогда $a_n\alpha^n + ... + a_1\alpha + a_0 = 0$. Возведем полученное равенство в степень p.

Используя лемму 8.3.3, получим равенство

$$a_n^p \alpha^{np} + ... + a_1^p \alpha^p + a_0^p = 0.$$

Как известно, $F_p \simeq \mathbb{Z}/p\mathbb{Z} = \mathbb{Z}_p$ – кольцо вычетов по простому модулю p. Согласно малой теореме Ферма (теорема 6.2.3), для кольца вычетов по простому модулю, для всех $\alpha \in \mathbb{Z}_p$ справедливо равенство $\alpha^p = \alpha$.

Следовательно, $\alpha_i^p = \alpha_i, i = 0, 1, ..., n$.

Получили равенство

$$a_n(\alpha^p)^n + ... + a_1\alpha^p + a_0 = 0.$$

Доказано, что элемент α^p является корнем многочлена $g(x)$. □

Теорема 8.3.5 *Мультипликативная группа конечного поля – циклическая.*

Доказательство. Обозначим через n число элементов поля F. Тогда мультипликативная группа поля F^* содержит $n-1$ элемент. Согласно следствию из теоремы Лагранжа, для всех $x \in F^*$ выполнено равенство $x^{n-1} = 1$.

Пусть d – натуральное число, $d \mid n-1$. Отметим, что $n > 1$.

Обозначим через $\psi(d)$ число элементов порядка d в мультипликативной группе поля.

Каждый элемент x мультипликативной группы поля имеет порядок $\mathrm{ord}(x)$, который по теореме Лагранжа делит $n-1$.

Пересчитывая элементы F^* по натуральным делителям числа $n-1$ получим равенство

$$\sum_{d|n} \psi(d) = n. \tag{23}$$

Пусть $\psi(d) > 0$. Это означает, что существует элемент $a \in F^*$ порядка d.

Тогда $1, a, ..., a^{d-1}$ – различные решения уравнения

$$x^d = 1 \tag{24}$$

в поле F. Из теории многочленов мы знаем, что уравнение степени d имеет не более d решений в поле.

Следовательно, единственными решениями уравнения (24) в поле F являются элементы $1, a, ..., a^{d-1}$, среди которых, согласно пункту 2 теоремы 1.4.1, $\varphi(d)$ элементов порядка d.

Все элементы порядка d поля F являются решениями уравнения (24).

Получили равенство $\psi(d) = \varphi(d)$.

Следовательно, для всех натуральных делителей d числа $n-1$ справедливо неравенство $\psi(d) \leq \varphi(d)$.

Вычитая равенства (8) и (4), получим равенство

$$\sum_{d|n}(\varphi(d) - \psi(d)) = 0, \tag{25}$$

где d – натуральные.

Из неотрицательности слагаемых, стоящих под знаком суммы в формуле (10) следуют равенства $\psi(d) = \varphi(d)$ для всех натуральных делителей d числа $n-1$.

В частности, при $d = n-1$ получаем, что $\psi(n-1) = \varphi(n-1) > 0$.

Следовательно, в группе F^* существуют элементы порядка $n-1$ и она является циклической. □

Упражнения

1. Доказать лемму 8.3.1.

2. Доказать, используя теоремы 8.3.2, 8.3.3, что над полем F_p, где p – простое существуют неприводимые многочлены любой натуральной степени.

3. Выпишите таблицы сложения и умножения элементов поля $F_8 \simeq F_2[x]_{m(x)}$, определенного неприводимым полиномом $m(x) = x^3 + x + 1$ над F_2.

4. Постройте поле F_9.

5. Для каждой степени $d \leq 5$ найдите число неприводимых над F_2 полиномов степени d и составьте их список.

6. Для каждой степени $d \leq 6$ найдите число нормированных неприводимых над F_3 полиномов степени d и для $d \leq 3$ составьте их список.

9 Коды исправляющие ошибки

Ошибки при передаче данных по каналу могут возникнуть по самым разным причинам. Для защиты от ошибок мы можем кодировать посылаемую информацию и декодировать ее таким образом, чтобы максимизировать вероятность исправления или по крайней мере обнаружения таких ошибок.

История кодирования, контролирующего ошибки, началось в 1948 году публикацией статьи Клода Шеннона[1]. Согласно фундаментальной теореме Шеннона за счет увеличения длины элементарных сообщений можно устранить влияние помех.

Цель применения кодов исправляющих ошибки – передавать избыточную информацию так, чтобы при возникновении в полученном сообщении небольшого фиксированного числа ошибок типа замещения разряда (одной, двух, трех,...) информация могла быть восстановлена. Первый такой код был разработан Хеммингом[2] в 1950 году.

9.1 Коды Хемминга

Определение. *Назовем (n, k)-кодом C любое k-мерное подпространство n-мерного векторного пространства V_n над полем F_2.*

Будем представлять себе векторное пространство V_n как множество последовательностей из нулей и единиц длины n.

Любой вектор пространства V_n, который может быть передан называется кодовым словом.

Так как кодовые слова передаются по зашумленному каналу, будем считать, что мы можем получить любой вектор пространства V_n. После получения некоторого вектора пользователь должен решить, в какое из кодовых слов его декодировать.

Определение. *Расстоянием по Хеммингу $d(u, v)$ между двумя двоичными последовательностями u и v длины n называется число координат k, в которых они различаются.*

Теорема 9.1.1 *Расстояние Хемминга $d(u, v)$ является метрикой на пространстве V_n, т. е. для любых векторов $u_1, u_2, u_3 \in V_n$ выполнено:*

a) $d(u_1, u_2) \geq 0$, $d(u_1, u_2) = 0$ тогда и только тогда, когда $u_1 = u_2$;

b) $d(u_1, u_2) = d(u_2, u_1)$;

c) $d(u_1, u_3) \leq d(u_1, u_2) + d(u_2, u_3)$ (неравенство треугольника).

[1] Клод Элвуд Шеннон – американский математик (1916-2001).

[2] Ричард Весли Хемминг – американский математик (1915-1998).

Доказательство. a) Расстояние Хемминга определяется как количество координат k, в которых векторы отличаются. Оно не может быть отрицательным.

Расстояние Хемминга $d(u_1, u_2)$ равно нулю тогда и только тогда, когда количество координат, в которых векторы u_1 и u_2 отличаются равно нулю.

b) Количество координат, в которых векторы u_1 и u_2 отличаются не зависит от того, какой вектор рассматривать первым, а какой вторым.

c) Пусть $d_k(u_1, u_2), k = 1, ..., n$ равно нулю, если k-ые координаты векторов u_1 и u_2 совпадают и равно единице, если k-ые координаты векторов u_1 и u_2 отличаются. Тогда

$$d(u_1, u_2) = \sum_{k=1}^{n} d_k(u_1, u_2).$$

Для доказательства пункта c) достаточно доказать неравенство треугольника для величин $d_k(u_1, u_2), k = 1, ..., n$.

Пусть k-ая координата вектора $u_s, s = 1, 2, 3$ равна x_s.

Если $x_1 = x_3$, то $d_k(u_1, u_3) = 0$ и свойство c) – выполнено.

Если $x_1 \neq x_3$, то $d_k(u_1, u_3) = 1$, $d_k(u_1, u_2) + d_k(u_2, u_3) = 1$ и свойство c) – выполнено. □

Определение. *Вес Хемминга $w(u), u \in V_n$ – это количество отличных от нуля координат вектора u. Очевидно, что*

$$w(u) = d(u, 0).$$

Предложение 9.1.1 *Для любых векторов $u, v \in V_n$ выполнены соотношения:*

a) $w(u) \geq 0$, $w(u) = 0$ тогда и только тогда, когда $u = 0$;

b) $w(v) = d(u, u + v)$;

c) $w(u + v) \leq w(u) + w(v)$.

Доказательство. Свойство a) следует непосредственно из свойства a) теоремы 9.1.1.

b) Количество координат, в которых отличаются векторы u $u + v$ совпадает с количеством ненулевых координат вектора v.

c) Рассмотрим векторы $0, u, u + v$. Из свойства c) теоремы 9.1.1 следует неравенство

$$w(u + v) = d(0, u + v) \leq d(0, u) + d(u, u + v) = w(u) + w(v).$$

□

Определение. *Минимальное расстояние*

$$d_{min} = \min_{\substack{u_i, u_j \in C, \\ u_i \neq u_j}} d(u_i, u_j)$$

между различными кодовыми словами двоичного кода C называется кодовым расстоянием.

Предложение 9.1.2 *Кодовое расстояние двоичного кода равно наименьшему весу ненулевых кодовых слов.*

Доказательство. Следует из пункта b) предложения 9.1.1. □

Теорема 9.1.2 *1. Двоичный код C с кодовым расстоянием d_{min} может исправлять все комбинации не более t ошибок типа замещения координаты тогда и только тогда, когда $d_{min} \geq 2t+1$.*

2. Если $d_{min} \geq 2t+2$, то код исправляет комбинации не более t ошибок типа замещения координаты и обнаруживает $t+1$ ошибку.

Доказательство. 1. Если u – переданное кодовое слово, v – полученный вектор с $t' \leq t$ ошибками, то $d(u,v) = t'$. Пусть u' – кодовое слово отличное от u. Тогда

$$d(u',v) \geq d(u',u) - d(v,u) \geq 2t+1-t = t+1 > t \geq t' = d(u,v).$$

Первое неравенство следует из неравенства треугольника.

Полученное соотношение показывает, что вектор v будет декодирован в кодовое слово u, что и означает возможность исправления t и менее ошибок.

2. Из пункта один следует возможность исправлять все комбинации не более t ошибок типа замещения координаты. Предположим, что произошла $t+1$ ошибка. Тогда получим неравенство $d(u',v) \geq d(u,v)$, которое означает, что либо $t+1$ ошибка исправляется, либо обнаруживается при $d(u',v) = d(u,v)$ для некоторого u'. □

Следующая теорема устанавливает верхнюю границу для числа кодовых слов для (n,k) кода с заданным кодовым расстоянием.

Теорема 9.1.3 *Если двоичный код C, имеющий s кодовых слов длины n, может исправлять все комбинации t или менее ошибок, то*

$$s \leq \frac{2^n}{\sum_{i=0}^{n} C_n^i}.$$

Доказательство. Пусть $u_1, u_2, ..., u_s$ – кодовые слова длины n в C. Для каждого кодового слова рассмотрим шар

$$S_t(u_i) = \{x \mid x \in V_n, d(x,u_i) \leq t\}$$

радиуса t с центром в u_i.

Так как код исправляет все комбинации t или менее ошибок, шары не пересекаются. Шар с центром u_i содержит все векторы, которые отличаются от u_i в 0,1,2,...,t координатах.

Тогда

$$\mid S_t(u_i) \mid = 1 + n + C_n^2 + ... + C_n^t = \sum_{i=0}^{t} C_n^t.$$

Получаем неравенство для числа слов всех шаров $s \sum_{i=0}^{t} C_n^t \leq 2^n$, которое доказывает искомое утверждение. □

Назовем *скоростью передачи информации* кода отношение $\frac{k}{n}$.

Теорема 9.1.2 показывает, что если код исправляет много ошибок, то кодовое расстояние должно быть как можно больше. Из теоремы 9.1.3 и требования иметь большое кодовое расстояние следует, что кодовых слов должно быть мало. Для большой скорости передачи информации хорошо иметь много кодовых слов. Большое количество кодовых слов ведет к снижению скорости декодирования (надо искать расстояние от полученного вектора до каждого кодового слова).

Из сказанного выше следует, что требования исправления большого числа ошибок и высокой скорости декодирования вступают в противоречие с требованием большой скорости передачи информации. Приходится идти на компромисс: исправление одной (как правило) ошибки при некоторой скорости передачи информации.

Упражнения

1. Двоичный код C длины n состоит из кодовых слов

   $$u_0 = (0, ..., 0), u_1 = (1, ..., 1).$$

   Найти кодовое расстояние. При каком n код исправляет все комбинации не более двух ошибок при наибольшей скорости передачи информации? Опишите процедуру декодирования.

2. Дан двоичный код, состоящий из векторов $u_0 = (0, 0, 0, 0, 0)$, $u_1 = (1, 0, 0, 1, 1), u_2 = (1, 1, 1, 0, 0), u_3 = (0, 1, 1, 1, 1)$.

   Какова размерность подпространства кодовых слов? Сколько ошибок может исправлять данный код? Какова скорость передачи информации?

3. Постройте код, состоящий из восьми слов длины 7, такой, что его кодовое расстояние равно 4. Какова размерность подпространства кодовых слов? Сколько ошибок может исправлять данный код? Какова скорость передачи информации?

4. Какое максимальное количество слов содержит код длины 8 с кодовым расстоянием 5? Постройте такой код. Какова размерность подпространства кодовых слов? Сколько ошибок может исправлять данный код? Какова скорость передачи информации?

5. Какую длину должны иметь кодовые слова, чтобы код с кодовым расстоянием 5 содержал восемь кодовых слов? Постройте такой код. Какова размерность подпространства кодовых слов? Сколько ошибок может исправлять данный код? Какова скорость передачи информации?

9.2 Проверочная и порождающая матрицы линейного кода

Как мы уже говорили в предыдущем разделе скорость декодирования зависит от числа кодовых слов: чем больше кодовых слов тем ниже скорость декодирования. Кроме того при размерности кода k (напомним, что код является подпространством V_n) в памяти компьютера должно быть сохранено 2^k кодовых слов.

Для увеличения скорости кодирования, декодирования и уменьшения используемой оперативной памяти компьютера применяются методы линейной алгебры.

Определение. *Линейными называются коды, в которых проверочные биты представляют собой линейные комбинации информационных битов.*

Пусть требуется передать вектор $(m_1, ..., m_k)$. Составим кодовое слово u из информационных и проверочных битов

$$u = (m_1, ..., m_k, c_1, ..., c_{n-k}).$$

Обозначим количество проверочных битов через $r = n - k$. Зададим проверочные биты соотношениями

$$\begin{array}{l} c_1 = a_{11}m_1 + ... + a_{1k}m_k, \\ c_2 = a_{21}m_1 + ... + a_{2k}m_k, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ c_r = a_{r1}m_1 + ... + a_{rk}m_k. \end{array} \tag{26}$$

Так как в поле F_2 справедливо равенство $-1 = 1$ получим соотношения

$$a_{i1}m_1 + ... + a_{ik}m_k + c_i = 0, i = 1, 2, ..., r. \tag{27}$$

Обозначим через E_r единичную матрицу порядка $r \times r$. Пусть

$$A = \begin{pmatrix} a_{11}...a_{1k} \\ a_{21}...a_{2k} \\ \cdots\cdots \\ a_{r1}...a_{rk} \end{pmatrix} -$$

матрица коэффициентов соотношений (26) порядка $r \times k$,

$$H = (A \mid E_r) -$$

матрица порядка $r \times n$, составленная из матриц A и E_r. Назовем матрицу H *проверочной матрицей* линейного кода.

Запишем в матричном виде соотношения (27). Пусть

$$v = (m_1, ..., m_k, c_1, ..., c_r).$$

Тогда уравнение

$$Hv^T = 0 \text{ или } vH^T = 0 \tag{28}$$

называется *проверочным*.

Через H^T мы будем обозначать матрицу, транспонированную к матрице H.

Напомним, что для произвольных матриц A и B, для которых определено произведение AB, справедливо равенство $(AB)^T = B^T A^T$.

Из равносильности проверочного уравнения (28) соотношениям (27) следует, что вектор $v \in V_n$ является кодовым словом тогда и только тогда, когда он удовлетворяет проверочному уравнению (28).

Для кодирования информации выберем базис $g_1, ..., g_k$ подпространства C кодовых слов. Назовем *порождающей матрицей* G кода C матрицу, строки которой являются координатами векторов $g_1, ..., g_k$ в стандартном базисе V_n (одна из координат векторов стандартного базиса равна 1, остальные 0).

Векторы $g_1, ..., g_k$ являются кодовыми словами. Следовательно, они удовлетворяют проверочному уравнению (28):

$$g_i H^T = 0, i = 1, ..., k \Rightarrow GH^T = 0.$$

Пример. Пусть $n = 7, k = 4$ и проверочные биты заданы соотношениями:

$$\begin{aligned} c_1 &= m_1 + m_2 + m_3, \\ c_2 &= m_1 + m_2 + \quad m_4, \\ c_3 &= m_1 + \quad m_3 + m_4. \end{aligned}$$

Получаем проверочную матрицу

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

Для получения порождающей матрицы можно найти фундаментальный набор решений однородной системы линейных уравнений (28).

Одна из порождающих матриц (порождающая матрица находится неоднозначно) имеет вид

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

Определение. *Системы векторов U и V называются эквивалентными, если их линейные оболочки совпадают (они порождают одинаковые подпространства).*

Известно, что при выполнении элементарных операций над векторами системы (умножения любого вектора на ненулевой элемент поля, прибавления к любому вектору любого другого, умноженного на элемент поля) она переходит в эквивалентную.

Следовательно, порождающая матрица переходит в порождающую при выполнении элементарного преобразования строк.

Из всех возможных порождающих матриц нас больше всего интересует порождающая матрица вида

$$G = \begin{pmatrix} 1 & 0 & 0 & ... & 0 & p_{11} & ... & p_{1,n-k} \\ 0 & 1 & 0 & ... & 0 & p_{21} & ... & p_{2,n-k} \\ . & . & . & ... & . & ... & ... & ..... \\ 0 & 0 & 0 & ... & 1 & p_{k1} & ... & p_{k,n-k} \end{pmatrix}. \tag{29}$$

Такую порождающую матрицу можно найти не для всех подпространств V_n над F_2.

Пример. Пусть порождающая матрица $G = (0, 1, 1, ..., 1)$ состоит из одной строки. У рассмотренного кода не существует порождающей матрицы вида (29).

Отметим, что такой код никому не придет в голову использовать – у всех кодовых слов первый бит нулевой, что снижает скорость передачи информации.

У каждого естественного кода (каждый бит может быть ненулевым для некоторого кодового слова) существует порождающая матрица вида (29).

Определение. *Скажем, что двоичный код является упорядоченным, если в каждом кодовом слове первые k битов информационные, а остальные проверочные.*

Пусть $m = (m_1, ..., m_k)$ и порождающая матрица кода имеет вид (29). Тогда передаваемое кодовое слово u равно

$$u = m \begin{pmatrix} 1 & 0 & 0 & ... & 0 & p_{11} & ... & p_{1,n-k} \\ 0 & 1 & 0 & ... & 0 & p_{21} & ... & p_{2,n-k} \\ . & . & . & ... & . & ... & ... & ..... \\ 0 & 0 & 0 & ... & 1 & p_{k1} & ... & p_{k,n-k} \end{pmatrix} = (m_1, ..., m_k, c_1, ..., c_r),$$

где $c_1, ..., c_r$ – проверочные биты.

Мы заметили, что для кодирования упорядоченного кода, можно применить формулу $u = mG$.

Следующая теорема показывает как связаны проверочная и порождающая матрицы упорядоченного кода.

Теорема 9.2.1 *Пусть упорядоченный двоичный (n,k)-код C имеет порождающую матрицу $G = (E_k \mid P)$, где P – произвольная матрица порядка $k \times (n-k)$. Тогда проверочная матрица H кода C имеет вид $H = (P^T \mid E_{n-k})$ и наоборот.*

Доказательство. 1. Пусть порождающая матрица G равна

$$G = \begin{pmatrix} 1 & 0 & 0 & ... & 0 & p_{11} & ... & p_{1,n-k} \\ 0 & 1 & 0 & ... & 0 & p_{21} & ... & p_{2,n-k} \\ . & . & . & ... & . & ... & ... & ..... \\ 0 & 0 & 0 & ... & 1 & p_{k1} & ... & p_{k,n-k} \end{pmatrix},$$

$m = (m_1, ..., m_k)$ – передаваемый вектор, $c = (c_1, ..., c_r)$ – вектор проверочных битов.

Тогда из формулы кодирования $u = mG$ получаем следующие соотношения для проверочных битов

$$c_i = p_{1i}m_1 + p_{2i}m_2 + ... + p_{k,i}m_k, i = 1, 2, ..., r(r = n-k). \tag{30}$$

2. Проверочная матрица H имеет вид

$$H = \begin{pmatrix} p_{11} & p_{21} & ... & p_{k1} & 1 & 0 & 0 & ... & 0 \\ p_{12} & p_{22} & ... & p_{k2} & 0 & 1 & 0 & ... & 0 \\ ... & ... & ... & ... & . & . & . & ... & . \\ p_{1,n-k} & p_{2,n-k} & ... & p_{k,n-k} & 0 & 0 & 0 & ... & 1 \end{pmatrix}.$$

Из проверочного уравнения (28) получаем

$$H \begin{pmatrix} m^T \\ c^T \end{pmatrix} = 0.$$

Получаем следующие соотношения для проверочных битов

$$p_{1i}m_1 + p_{2i}m_2 + ... + p_{k,i}m_k + c_i = 0, i = 1, 2, ..., r. \tag{31}$$

Осталось отметить, что над полем F_2 соотношения (30) и (31) эквивалентны. □

Для нахождения кодового расстояния используется следующая теорема.

Теорема 9.2.2 *Код C с проверочной матрицей H имеет кодовое расстояние w тогда и только тогда, когда любое множество из $w-1$ столбцов проверочной матрицы H линейно независимо и найдутся w линейно зависимых столбцов H.*

Доказательство. Любое кодовое слово u кода C, удовлетворяет проверочному уравнению (28): $Hu^T = 0$.

Если вектор u имеет ровно t единичных координат, то в матрице H есть t линейно зависимых столбцов.

Подсистема линейно независимой системы векторов линейно независима. Следовательно, в матрице H линейно независимо любое множество из $w-1$ и меньшего числа столбцов. Поэтому наименьший вес кодовых слов кода C больше $w-1$.

Из существования w линейно зависимых столбцов проверочной матрицы H следует равенство w наименьшего веса ненулевых кодовых слов.

Остается применить предложение 9.1.2. □

Пример. Приведем пример кода C, исправляющего одну ошибку. Пусть длина кодовых слов

$$n = 2^m - 1, k = 2^m - m - 1, r = n - k = 2^m - 1 - (2^m - m - 1) = m,$$

где $m \geq 2$ – натуральное.

Тогда проверочная матрица H имеет порядок $m \times (2^m - 1)$. Составим проверочную матрицу из всех ненулевых векторов-столбцов m-мерного арифметического векторного пространства над полем F_2.

Все столбцы матрицы H ненулевые, попарно различны, есть 3 линейно зависимых столбца. Согласно теореме 9.2.2, кодовое расстояние кода C равно трем и он исправляет одну ошибку.

Определение. *Пусть H – проверочная матрица двоичного (n,k)-кода. Если v – полученный вектор, то вектор-столбец $s = Hv^T$, содержащий $n-k$ элементов называется синдромом вектора v.*

Из проверочного уравнения (28) следует, что вектор является кодовым словом тогда и только тогда, когда его синдром равен нулю.

Опишем процесс декодирования. Пусть u – переданное кодовое слово, v – полученный вектор, а $e = v - u = v + u$ – вектор ошибок. Тогда

$$s = Hv^T = H(u+e)^T = H(u^T + e^T) = He^T.$$

Если произошла одна ошибка, то вектор e содержит только одну единичную k-ую координату, если k-ые координаты u и v отличаются.

Следовательно, синдром s совпадает с k-ым столбцом матрицы H. В случае, когда код исправляет одну ошибку его кодовое расстояние больше двух и любые два столбца матрицы H отличаются, согласно теореме 9.2.2. В этом случае номер координаты, в которой произошла ошибка находится по синдрому однозначно.

Разберем решение типовой задачи.

Задача. Пусть

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} -$$

порождающая матрица двоичного кода.

1. Найти проверочную матрицу.

2. Найти длину кодовых слов и размерность их подпространства. Какова скорость передачи информации?

3. Каково его кодовое расстояние?

4. Сколько ошибок исправляет данный код?

5. Закодировать сообщение $m = (1, 0, 1, 0)$.

6. Раскодировать сообщения в предположении, что при передаче произошло не более одной ошибки:

а) $v = (0, 0, 1, 0, 0, 1, 1)$; б) $v = (1, 1, 0, 0, 0, 1, 0)$.

Решение. 1. Найдем проверочную матрицу, используя теорему 9.2.1

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

2. Порождающая матрица имеет порядок 4×7. Следовательно, длина кодовых слов равна 7, подпространство кодовых слов имеет размерность 4 и мы изучаем $(7, 4)$-код. Скорость передачи информации равна $4/7$.

3. Так как все столбцы проверочной матрицы ненулевые, попарно различны и есть при линейно зависимых столбца (например, первый, пятый и шестой), то согласно теореме 9.2.2 кодовое расстояние равно 3.

4. Согласно теореме 9.1.2, так как $3 \geq 2 \cdot 1 + 1$, код исправляет одну ошибку.

5. По формуле $u = mG = (1, 0, 1, 0, 1, 0, 1)$.

6. а) Синдром равен $s = Hv^T = 0$. Следовательно, сообщение u является кодовым словом и передано без ошибок.

б) Синдром равен $s = Hv^T = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Он совпадает с последним столбцом проверочной матрицы. Следовательно, ошибка произошла в 7-ом бите и было передано сообщение $u = (1, 1, 0, 0, 0, 1, 1)$.

Упражнения

1. Пусть
$$G = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix} -$$
порождающая матрица двоичного кода. Найти: длину кодовых слов и размерность их подпространства; скорость передачи информации; проверочную матрицу; кодовое расстояние; количество исправляемых ошибок. Закодировать сообщение $u = (1, 0)$. Раскодировать сообщения в предположении, что при передаче произошло не более одной ошибки: а) $v = (0, 1, 0, 1, 1)$;
б) $v = (1, 0, 1, 1, 0)$.

2. Пусть
$$H = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} -$$
проверочная матрица двоичного кода. Найти: длину кодовых слов и размерность их подпространства; скорость передачи информации; порождающую матрицу; кодовое расстояние; количество исправляемых ошибок. Закодировать сообщение $u = (0, 1, 1, 0)$. Раскодировать сообщения в предположении, что при передаче произошло не более одной ошибки:
а) $v = (0, 1, 0, 0, 1, 0, 1)$; б) $v = (0, 1, 0, 0, 0, 1, 0)$.

3. Пусть
$$H = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} -$$
проверочная матрица двоичного кода. Найти: длину кодовых слов и размерность их подпространства; скорость передачи информации; порождающую матрицу; кодовое расстояние; количество исправляемых ошибок. Закодировать сообщение $u = (0, 1, 0, 1)$. Раскодировать сообщения в предположении, что при передаче произошло не более двух ошибок:
а) $v = (0, 0, 1, 0, 1, 1, 1, 0)$; б) $v = (1, 1, 0, 1, 0, 1, 0, 0)$;
в) $v = (0, 0, 0, 0, 1, 0, 1, 1)$.

9.3 Циклические коды

Циклические коды представляют собой подмножество кодов Хемминга.

Для увеличения скорости передачи информации и количества исправляемых ошибок используется реализация векторного пространства V_n над полем F_2 как множество многочленов ограниченной степени.

А. Хоквингем (1959 г.)[1], Р. К. Боуз[2] и Д. К. Рой-Чоудхури (1960 г.)[3] нашли большой класс кодов, обеспечивающий произвольное минимальное кодовое расстояние $d_{min} \geq 5$. Они получили название БЧХ кодов (Боуза-Чоудхури-Хоквингема) и являются наиболее известным подклассом циклических кодов.

Определение. *Подпространство C векторного пространства V_n над полем F_2 называется циклическим подпространством или циклическим кодом, если для каждого вектора $v = (v_{n-1}, v_{n-2}, ..., v_0)$ кода C вектор $v' = (v_{n-2}, ..., v_0, v_{n-1})$, полученный из v циклическим сдвигом, также принадлежит C.*

Будем реализовывать векторное пространство V_n над полем F_2 как факторалгебру $F_2[x]/f(x)F_2[x] = F_2[x]_{f(x)}$, где $f(x)$ – многочлен степени n. В качестве представителей смежных классов алгебры $F_2[x]_{f(x)}$ рассмотрим множество T_n – всех многочленов кольца $F_2[x]$, степени не выше $n-1$, включающее нулевой многочлен.

Множество многочленов T_n является векторным пространством над F_2 размерности n. Можно считать, что на T_n задано умножение следующим образом: произведением двух многочленов $t(x), s(x) \in T_n$ является остаток от деления $t(x)s(x)$ на $f(x)$. Множество T_n с естественной структурой векторного пространства на нем и заданной операцией умножения является алгеброй, изоморфной алгебре $F_2[x]_{f(x)}$.

Теория циклических кодов основана на следующей важной теореме.

Теорема 9.3.1 *Пусть F – поле. Подпространство I фактор-алгебры $F[x]_{x^n-1}$ является циклическим тогда и только тогда, когда I – идеал.*

Доказательство. Рассмотрим равенство многочленов

$$x(a_{n-1}x^{n-1} + ... + a_0) = a_{n-1}(x^n - 1) + a_{n-2}x^{n-1} + ... + a_0x + a_{n-1},$$

где $a_{n-1}, ..., a_0 \in F_2$ – произвольные.

[1] А. Хоквингем – французский математик.

[2] Р. К. Боуз – американский математик.

[3] Д. К. Рой-Чоудхури – американский математик.

В алгебре T_n остатков от деления на $x^n - 1$ получим равенство

$$x(a_{n-1}x^{n-1} + ... + a_0) = a_{n-2}x^{n-1} + ... + a_0x + a_{n-1}. \quad (32)$$

1. Пусть C – циклическое подпространство T_n, $v \in C$. Из формулы (32) следует, что $xv, x^2v, ..., x^{n-1}v \in C$.

Тогда $(c_{n-1}x^{n-1} + ... + c_0)v \in C$, где $c_{n-1}, ..., c_0 \in F_2$ – произвольные. Следовательно, множество C является идеалом алгебры T_n. Из изоморфизма алгебр T_n и $F_2[x]_{x^n-1}$ вытекает, что циклическое подпространство алгебры $F_2[x]_{x^n-1}$ является идеалом.

2. Пусть множество C является идеалом алгебры T_n. Из того, что $v \in C$ следует, что $xv \in C$. Из формулы (32) вытекает, что C циклическое подпространство T_n. Дальше так же, как в 1, применяется изоморфизм алгебр T_n и $F_2[x]_{x^n-1}$. □

Напомним следующую важную теорему.

Теорема 9.3.2 ([4]) *Любой идеал I кольца многочленов над полем F представляет собой множество многочленов кратных некоторому многочлену $I = f(x)F[x], f(x) \in F[x]$.*

Теорема 9.3.2 утверждает, что кольцо многочленов от одной переменной над полем является кольцом главных идеалов.

Следствие 9.3.1 *Пусть $f(x) \in F[x]$ – ненулевой, F – поле. Алгебра $F[x]_{f(x)}$ является кольцом главных идеалов (элементы любого идеала кратны некоторому $[g(x)], g(x) \in F[x]$).*

Доказательство. Пусть $\varphi : K \to H$ – сюръективный гомоморфизм колец, K – кольцо главных идеалов. Покажем, что H тоже является кольцом главных идеалов.

Пусть $\bar{I}$ – идеал кольца H. Обозначим через I полный прообраз $\bar{I}$. Тогда множество I является идеалом кольца K. Следовательно, в K существует элемент a такой, что $I = \{ra \mid r \in K\}$.

Рассмотрим произвольный элемент $s \in \bar{I}$. Тогда $s = \varphi(t), t \in K$.

Существует элемент $r \in K$ такой, что $t = ra$. Тогда

$$s = \varphi(ra) = \varphi(r)\varphi(a).$$

Следовательно, идеал $\bar{I}$ является главным $\bar{I} = \varphi(a)K$. □

Определение. *Пусть $I \lhd K$ – идеал кольца K. Скажем, что элемент $r \in K$ ортогонален идеалу I, если для всех $s \in I$ выполнено $rs = 0$.*

Теорема 9.3.3 *Рассмотрим полиномы $f(x), g(x), h(x) \in F[x]$, где F – поле, $f(x) = g(x)h(x), f(x) \neq 0$. Тогда элемент $[a(x)]$ алгебры $F[x]_{f(x)}$ ортогонален идеалу I, порожденному $[h(x)]$ тогда и только тогда, когда $[a(x)]$ принадлежит идеалу J, порожденному $[g(x)]$.*

Доказательство. 1. Пусть $[a(x)]$ принадлежит идеалу J. Тогда $[a(x)] = [s(x)][g(x)]$ для некоторого многочлена $s(x)$.

Пусть $[t(x)][h(x)]$ – произвольный элемент I. Получим

$$[a(x)][t(x)][h(x)] = [s(x)][g(x)][t(x)][h(x)] = [s(x)t(x)g(x)h(x)] =$$

$$= [s(x)t(x)f(x)] = f(x)F[x] = [0].$$

2. Рассмотрим элемент $[a(x)]$ ортогональный идеалу I. Тогда $[a(x)][h(x)] = [0]$. Следовательно,

$$a(x)h(x) = t(x)g(x)h(x) \Rightarrow (a(x) - t(x)g(x))h(x) = 0,$$

для некоторого многочлена $t(x)$. Так как кольцо многочленов над полем является областью целостности (не содержит делителей нуля) [4] и $h(x) \neq 0$, получим

$$a(x) = t(x)g(x) \Rightarrow [a(x)] = [t(x)][g(x)] \Rightarrow [a(x)] \in J.$$

Последнее включение завершает доказательство теоремы. □

Рассмотрим полиномы $x^n - 1 = g(x)h(x)$ над полем F_2, где $deg\ h(x) = k, deg\ g(x) = r = n - k$. Тогда $dim_{F_2}(F_2[x]_{x^n-1}) = n$. Как мы уже говорили выше множество T_n – всех многочленов кольца $F_2[x]$, степени не выше $n-1$, включающее нулевой многочлен, изоморфно фактор-алгебре $F_2[x]_{x^n-1}$.

Будем считать многочлен $g(x)$ порождающим многочленом двоичного кода C. Код C состоит из всех многочленов векторного пространства $V_n = T_n$ кратных $g(x)$. Его базис образуют многочлены

$$x^{k-1}g(x), x^{k-2}g(x), ..., xg(x), g(x), dim_{F_2}C = k.$$

Кодирование

Сообщение $(a_{k-1}, a_{k-2}, ..., a_0)$ может быть закодировано с помощью кода C путем вычисления произведения

$$(a_{k-1}x^{k-1} + a_{k-2}x^{k-2} + ... + a_0)g(x).$$

Пример. Пусть $n = 7, k = 4$. Рассмотрим порождающий многочлен $g(x) = x^3 + x + 1$. Отметим, что $x^7 - 1 \vdots x^3 + x + 1$.

Закодируем сообщение $m = (0, 1, 0, 1)$. Получим многочлен

$$u(x) = (x^2 + 1)(x^3 + x + 1) = x^5 + x^3 + x^2 + x^3 + x + 1 = x^5 + x^2 + x + 1,$$

которому соответствует кодовое слово $u = (0, 1, 0, 0, 1, 1, 1)$ длины 7.

Многочлен $h(x)$ является проверочным. Из теоремы 9.3.3 следует, что многочлен $t(x) \in T_n$ является кодовым словом тогда и только когда, когда $t(x)h(x) \vdots x^n - 1$. Более простой проверкой является выполнение делимости $t(x) \vdots g(x)$.

Декодирование

Пусть получено сообщение v длины n. Представим его в виде многочлена $v(x) \in T_n$.

Разделим многочлен $v(x)$ с остатком на порождающий многочлен $g(x)$. Получим $v(x) = d(x)g(x) + s(x)$, где $deg\ s(x) \leq deg\ g(x)$ или $s(x) = 0$.

Остаток $s(x)$ называется синдромом сообщения $v(x)$. Если $s(x) = 0$, то $v(x)$ – кодовое слово.

Предположим, что было передано сообщение $u(x)$ и многочлен ошибок имеет вид $e(x) = x^{i_1} + x^{i_2} + ... + x^{i_t}$. Это означает, что ошибки произошли в коэффициентах многочлена $u(x)$ при степенях $i_1, i_2, ..., i_t$.

Получим равенство $v(x) = u(x)+e(x)$. Подставляя выражение для синдрома имеем $d(x)g(x)+s(x) = u(x)+e(x)$. Так как многочлен $u(x)$ кратен $g(x)$ получим

$$u(x) = d_1(x)g(x) \Rightarrow e(x) = (d(x) - d_1(x))g(x) + s(x).$$

Учитывая, что $deg\ s(x) \leq deg\ g(x)$ или $s(x) = 0$, синдром $s(x)$ полученного сообщения $v(x)$ является остатком от деления $e(x)$ на $g(x)$.

Если многочлен $g(x)$ – неприводим, $deg\ g(x) = m$ и $n = 2^m - 1$ и α – элемент порядка n в мультипликативной группе поля F_{2^m} является корнем $g(x)$, то код может исправлять одну ошибку и его кодовое расстояние не меньше 3.

Покажем как в этом случае производится декодирование.

Имеет место следующий изоморфизм

$$F_2[x]_{x^n-1}/[g(x)]F_2[x]_{x^n-1} \simeq F_2[x]/g(x)F_2[x] \simeq F_2[x]_{g(x)} -$$

поле из 2^m элементов (напомним, что многочлен $g(x)$ является делителем $x^n - 1$).

Пусть $\alpha = [x] = x + g(x)F_2[x]$ элемент поля $F_2[x]_{g(x)}$. В поле $F_2[x]_{g(x)}$ имеет место равенство $e(\alpha) = s(\alpha)$. В случае наступления одной ошибки $e(x) = x^i$.

Мультипликативная группа поля $F_2[x]_{g(x)}$ имеет простой порядок n, его неединичный элемент α является образующим циклической группы $(F_2[x]^*_{g(x)}, \cdot)$.

Вычисляя $e(\alpha) = \alpha^i$ находим одночлен, в коэффициенте которого произошла ошибка.

Пример. Пусть $n = 7, k = 4$, $g(x) = x^3 + x + 1$.

Предположим, что было получено сообщение $v = (0, 1, 0, 0, 0, 1, 0)$, которому соответствует многочлен $v(x) = x^5 + x$. Поделив $v(x)$ с остатком на $g(x) =$

$x^3 + x + 1$ получим синдром $s(x) = x^2 + 1$. Так как синдром отличен от 0, произошла ошибка. Декодируем ее в предположении, что произошло не более одной ошибки типа замещения разряда.

Тогда $e(\alpha) = s(\alpha) = \alpha^2 + 1$, где $\alpha = x + (x^3 + x + 1)F_2[x]$. Учитывая, что $\alpha^3 = \alpha + 1$, получим равенства

$$\alpha^4 = \alpha \cdot \alpha^3 = \alpha(\alpha + 1) = \alpha^2 + \alpha,$$

$$\alpha^5 = \alpha \cdot \alpha^4 = \alpha(\alpha^2 + \alpha) = \alpha^3 + \alpha^2 = \alpha^2 + \alpha + 1,$$

$$\alpha^6 = \alpha \cdot \alpha^5 = \alpha(\alpha^2 + \alpha + 1) = \alpha^3 + \alpha^2 + \alpha = \alpha + 1 + \alpha^2 + \alpha = \alpha^2 + 1$$

или

$$\alpha^6 = (\alpha^3)^2 = (\alpha + 1)^2 = \alpha^2 + 1.$$

Мы установили, что $s(\alpha) = \alpha^6$. Следовательно, ошибка произошла в коэффициенте при x^6, был передан многочлен

$$u(x) = x^6 + x^5 + x = (x^3 + x^2 + x)(x^3 + x + 1),$$

информационная часть сообщения имела вид $m = (1, 1, 1, 0)$.

Код, исправляющий 2 ошибки типа замещения разряда

Пусть $g(x)$ – многочлен из $F_2[x]$, для которого элементы α и α^3 поля F_{2^m} являются корнями, где α – образующий элемент мультипликативной группы поля.

Покажем, что циклический код длины $n = 2^m - 1$ с порождающим полиномом $g(x)$ исправляет все одиночные и двойные ошибки типа замещения разряда и, следовательно, имеет кодовое расстояние не меньше 5.

Предположим, что был получен многочлен $v(x) = d(x)g(x) + s(x)$, где $s(x)$ – синдром. Пусть $e(x)$ – многочлен ошибок. Как мы знаем, $e(\alpha) = s(\alpha)$.

Мы считаем, что многочлен $e(x)$ имеет не более двух ненулевых коэффициентов, то есть $e(x) = 0$ или $e(x) = x^i$ или $e(x) = x^i + x^{i'}$.

Элементы $X_1 = \alpha^i$ и $X_2 = \alpha^{i'}$ поля F_{2^m} называются локаторами.

Так как α – образующий элемент мультипликативной группы поля F_{2^m}, степени i и i' находятся однозначно по элементам X_1 и X_2.

Пусть $S_1 = v(\alpha)$ и $S_2 = v(\alpha^3)$. Элементы S_1 и S_2 вычисляются по полученному многочлену $v(x)$. Так как элементы α и α^3 являются корнями многочлена $g(x)$, справедливы равенства $S_1 = e(\alpha)$ и $S_2 = e(\alpha^3)$.

В предположении двух ошибок получаем

$$S_1 = \alpha^i + \alpha^{i'}, S_2 = \alpha^{3i} + \alpha^{3i'}.$$

Получаем следующую систему для нахождения X_1 и X_2

$$\begin{cases} S_1 = X_1 + X_2, \\ S_2 = X_1^3 + X_2^3. \end{cases}$$

Рассмотрим многочлен

$$(x - X_1)(x - X_2) = x^2 - (X_1 + X_2)x + X_1X_2.$$

Если найти его коэффициенты, то локаторы определятся однозначно. В поле характеристики два получаем

$$S_1^3 + S_2 = X_1^2X_2 + X_1X_2^2 = S_1X_1X_2,$$

$$(x - X_1)(x - X_2) = x^2 - S_1x + \frac{S_1^3 + S_2}{S_1}.$$

При наличии ошибок $S_1 \neq 0$.

Следовательно, рассмотренные циклические коды исправляют все комбинации не более двух ошибок типа замещения разряда.

Отметим, что рассмотренные ранее $(2^m, 2^m - m - 1)$ коды Хемминга эквивалентны циклическим кодам.

Приведем без доказательства теорему, позволяющую строить циклический код с заданным кодовым расстоянием.

Теорема 9.3.4 *Пусть $g(x)$ – порождающий полином двоичного циклического кода длины $n = 2^m - 1$, α – элемент порядка n в мультипликативной группе поля F_{2^m} и $\alpha^{\varepsilon_1}, \alpha^{\varepsilon_2}, ..., \alpha^{\varepsilon_s}$ являются корнями полинома $g(x)$. Тогда кодовое расстояние этого кода больше, чем максимальное число последовательных целых чисел в множестве $\{\alpha^{\varepsilon_1}, \alpha^{\varepsilon_2}, ..., \alpha^{\varepsilon_s}\}$.*

Рассмотрим примеры применения теоремы 9.3.4.

Примеры. Пусть $g(x)$ – порождающий полином двоичного циклического кода длины $n = 2^m - 1$, α – элемент порядка n в мультипликативной группе поля F_{2^m}.

1. Если α является корнем многочлена $g(x)$, то код может исправлять одну ошибку и его кодовое расстояние не меньше 3.

Согласно теореме 8.3.4, элемент α^2 является корнем $g(x)$. Из теоремы 9.3.4 следует, что $d_{min} \geq 3$.

2. Пусть $g(x)$ – многочлен, для которого элементы α и α^3 поля F_{2^m} являются корнями многочлена $g(x)$.

Тогда циклический код длины n с порождающим полиномом $g(x)$ исправляет все одиночные и двойные ошибки типа замещения разряда и его кодовое расстояние не меньше 5.

Согласно теореме 8.3.4, элементы α^2, α^4 является корнями $g(x)$. Элементы α, α^3 является корнями $g(x)$ по условию. Из теоремы 9.3.4 следует, что $d_{min} \geq 5$.

3. Пусть $m = 4$, элементы $\alpha, \alpha^2, \alpha^3, \alpha^4, \alpha^5, \alpha^6$ корни многочлена $g(x)$.

Согласно теореме 9.3.4, кодовое расстояние не меньше 7.

Мультипликативная группа поля F_{16} содержит 15 элементов. Следовательно, элемент α удовлетворяет соотношению $\alpha^{15} - 1 = 0$.

Вычислим коэффициенты минимального многочлена $g(x)$ с заданными корнями.

Многочлен имеющий корень α будет иметь корнями элементы $\alpha^2, \alpha^4, \alpha^8$.

Многочлен с корнем α^3 будет иметь корни

$$\alpha^6, \alpha^{12}, \alpha^{24} = \alpha^{15}\alpha^9 = \alpha^9.$$

Многочлен с корнем α^5 будет иметь корень α^{10}.

Получаем

$$g(x) = (x+\alpha)(x+\alpha^2)(x+\alpha^4)(x+\alpha^8)(x+\alpha^3)(x+\alpha^6)(x+\alpha^{12})(x+\alpha^9)\cdot$$

$$\cdot(x+\alpha^5)(x+\alpha^{10}) = (x^4+x+1)(x^4+x^3+x^2+x+1)(x^2+x+1).$$

Если рассмотреть неприводимый над F_2 многочлен $x^4 + x + 1$ и взять в качестве $\alpha = x + (x^4 + x + 1)F_2[x]$, то элемент α удовлетворяет соотношению $\alpha^4 = \alpha + 1$.

Получаем следующие соотношения для различных степеней α:

$$\alpha^5 = \alpha^2 + \alpha; \alpha^6 = \alpha^3 + \alpha^2; \alpha^7 = \alpha^3 + \alpha + 1; \alpha^8 = \alpha^2 + 1;$$

$$\alpha^9 = \alpha^3 + \alpha; \alpha^{10} = \alpha^2 + \alpha + 1; \alpha^{11} = \alpha^3 + \alpha^2 + \alpha;$$

$$\alpha^{12} = \alpha^3 + \alpha^2 + \alpha + 1; \alpha^{13} = \alpha^3 + \alpha^2 + 1; \alpha^{14} = \alpha^3 + 1.$$

Вычислим коэффициенты многочлена

$$(x+\alpha^5)(x+\alpha^{10}) = x^2 + (\alpha^5 + \alpha^{10})x + \alpha^5\alpha^{10} = x^2 + x + 1.$$

Аналогично проверяются равенства

$$(x+\alpha)(x+\alpha^2)(x+\alpha^4)(x+\alpha^8) = (x^4+x+1) \text{ и}$$

$$(x+\alpha^3)(x+\alpha^6)(x+\alpha^{12})(x+\alpha^9) = (x^4+x^3+x^2+x+1).$$

Многочлен $g(x)$ имеет степень 10, длина кода n равна 15, Код содержит $k = 15 - 10$ информационных битов.

Для решения упражнений полезно знать следующее разложение многочлена $x^{15} - 1$ в произведение неприводимых над F_2 многочленов

$$x^{15} - 1 = (x-1)(x^2+x+1)(x^4+x+1)(x^4+x^3+1)(x^4+x^3+x^2+x+1).$$

Упражнения

1. Определите все кодовые слова кода с порождающим полиномом $g(x) = x^3+x+1$ над F_2, если подпространство кодовых слов имеет размерность 4. Закодировать сообщение $(1,0,1,0)$. Имеет ли сообщение $(0,1,1,1,0,1,1)$ обнаруживаемые ошибки ?

2. Многочлен $g(x) = x^6 + x^5 + x^4 + x^3 + 1$ является порождающим многочленом циклического кода длины 15. Найдите проверочный многочлен. Сколько ошибок может исправлять этот код? Закодировать и декодировать сообщения, составленные самостоятельно.

3. Постройте циклический код длины 15 с 7 информационными и 8 проверочными битами, исправляющий две ошибки. Закодировать и декодировать сообщения, составленные самостоятельно.

10 Элементы криптографии

В переводе с греческого языка слово криптография означает тайнопись. Дадим следуя [1], определение.

Определение. *Криптология – это искусство проектирования и взлома секретных систем. Часть, связанная с проектированием, называется криптографией, а "взламывающая" часть – криптографическим анализом, или криптоанализом.*

Кодирование данных с целью защиты от несанкционированного доступа осуществляется с помощью криптографической системы (криптосистемы или шифра).

Основное требование к шифру состоит в том, чтобы декодирование было возможно только при наличии санкции, т. е. некоторой дополнительной информации, которая называется ключом шифра.

Вскрытие (взламывание) шифра – процесс декодирования без наличия ключа.

Способность шифра противостоять попыткам его вскрытия называется стойкостью шифра и обычно измеряется сложностью алгоритма декодирования.[1]

Долгое время занятие криптографией было уделом чудаков-одиночек. Известны случаи, когда криптография считалась даже черной магией. Свой след в истории криптографии оставили многие исторические личности. Первые сведения об использовании шифров в военном деле связаны с именем спартанского полководца Лицандра (*шифр Сцитала*, для его его реализации использовался сцитала - железный жезл, имеющий форму цилиндра). Отметим, что в этом шифре преобразование открытого текста в шифрованный заключается в определенной перестановке букв открытого текста. Такие шифры называются шифрами перестановки.

Шифр Цезаря реализует следующее преобразование открытого текста: каждая буква открытого текста заменяется третьей после нее буквой в алфавите, который считается написанным по кругу. Шифры такого типа называются шифрами замены.

В 1566 году известный математик Д. Кардано[1] опубликовал работу с описанием изобретенной им системы шифрования (решетка Кардано).

Петром I, была придумана цифирная азбука для кодирования сообщений.

Известный теоретико-числовик, член Петербургской академии наук Х. Гольдбах[2] служил криптографом при Екатерине II.

---

[1]В практической криптографии стойкость шифра оценивается из экономических соображений. Если раскрытие шифра стоит (в денежном выражении, включая необходимые компьютерные ресурсы, специальные устройства и т.п.) больше, чем сама зашифрованная информация, то шифр считается достаточно надежным.

[1]Джероламо Кардано – итальянский математик (1501-1576).

[2]Христан Гольдбах – российский математик, немец по национальности (1690-1764).

Хорошее подробное описание *шифров замены* и методов его вскрытия содержится в двух известных рассказах: Э. По “Золотой жук“ и А. Конан-Дойля “Пляшущие человечки“.

Без использования криптографии сегодня немыслимо решение задач по обеспечению безопасности информации, связанных с конфиденциальностью и целостностью, аутентификацией и невозможностью отказа сторон от соавторства.

Если до 1990 года криптография в России обеспечивала закрытие исключительно государственных линий связи, то в наши дни использование криптографических методов получило широкое распространение благодаря развитию компьютерных сетей и электронного обмена данными в различных областях: финансах; банковском деле; торговле и других [2].

10.1 Симметричные криптосистемы

Схему передачи кодированных сообщений можно изобразить следующим образом (рис. 1).

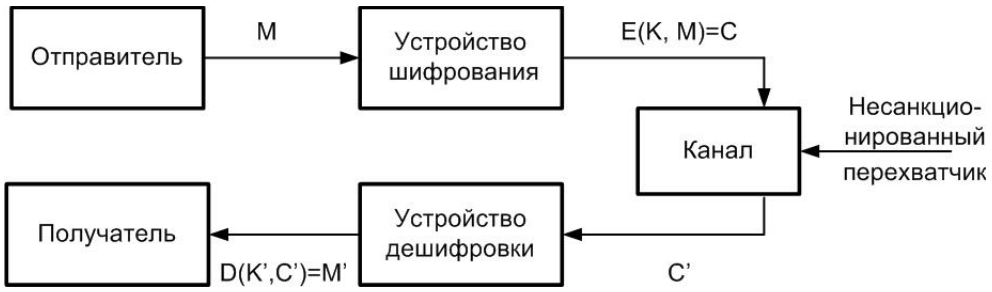


Рис. 1: Передача кодированного сообщения M.

Сообщение M, которое называется открытым текстом, кодируется с помощью ключа K в шифрованный текст $C = E(K, M)$. Шифрованный текст C передается по зашумленному незащищенному каналу связи. Во время передачи несанкционированный перехватчик может перехватить текст C и заменить его на C'.

Декодер основан на алгоритме расшифровки D, который использует в качестве аргументов полученный шифрованный текст C' и ключ для дешифровки K'. Ключ и шифрованный текст должны определять открытый текст M однозначно. В результате декодирования получится сообщение $M' = D(K', C')$, которое может отличатся от исходного сообщения M, если вмешался несанкционированный перехватчик или канал сильно зашумлен.

Назовем системы, удовлетворяющие условию, $K = K'$ системами единого ключа или симметричными криптосистемами. Системы в которых $K \neq K'$ называются асимметричными криптосистемами или системами открытого ключа.

В классической криптографии имеется два основных преобразования открытого текста сообщений вместе с их комбинациями:

1) Шифры перестановки, переупорядочивают группу символов в соответствии с некоторым правилом, не меняя их;

2) Шифры замены изменяют символы отк1рытого текста соответствующими символами из алфавита шифрованного текста (ключ задает отображение);

3) Комбинируя 1) и 2), мы получаем шифры перестановки-замены.

Еще Клода Шеннон в своей основополагающей для теории информации работе 1948 года "Математическая теория связи" писал, что даже в сложных шифрах в качестве типичных компонентов можно выделить такие простые шифры, как шифры замены, шифры перестановки или их сочетания.

Легко дать математическое описание шифра замены.

Пусть X и Y – два алфавита (открытого и шифрованного текстов соответственно), состоящие из одинакового количества букв, и пусть $g : X \to Y$ – биективное отображение X в Y. Тогда шифр замены действует следующим образом: открытый текст $x_1x_2\cdots x_n$ преобразуется в шифрованный текст $g(x_1)g(x_2)\cdots g(x_n)$.

Шифр перестановки действует обычно следующим образом: открытый текст разбивается на отрезки равной длины, каждый из которых шифруется независимо. Пусть, например, длина отрезков равна n и σ – подстановка на множестве $\{1, 2, \ldots, n\}$, тогда отрезок открытого текста $x_1x_2\cdots x_n$ преобразуется в отрезок шифрованного текста $x_{\sigma(1)}x_{\sigma(2)}\cdots x_{\sigma(n)}$.

Важнейшим для развития криптографии был результат Шеннона о существовании и единственности абсолютно стойкого шифра (невозможно раскодировать, не зная ключа). Единственным таким шифром является какая-нибудь форма так называемой *ленты однократного использования*, в котором открытый текст "объединяется" с полностью случайным ключом такой же длины.

Типичным и наиболее простым примером реализации абсолютно стойкого шифра является *шифр Вермана*, который осуществляет побитовое сложение n-битового открытого текста и n-битового ключа:

$$y_i = x_i + k_i, \qquad i = 1, 2, \ldots, n.$$

Здесь $x_1x_2\cdots x_n$ – открытый текст, $k_1, k_2, \cdots, k_n$ – ключ, $y_1y_2\cdots y_n$ – шифрованный текст.

Заметим, что для абсолютной стойкости такого шифра необходимо выполнение следующих условий:

1) полная случайность ключа;

2) равенство длины ключа и длины открытого текста;

3) однократность использования ключа.

Но именно эти условия делают абсолютно стойкий шифр очень дорогим и непрактичным. Прежде чем им пользоваться, необходимо обеспечить всех абонентов достаточным запасом случайных ключей и исключить возможность их повторного применения, что необычайно трудно и дорого.

Пример 1 (перестановка). Предположим, что мы хотим закодировать сообщение "компьютерная алгебра". Запишем текст в обратном порядке без пробе-

лов, группируя его традиционным способом по пять символов для затруднения раскодирования. Получим следующий шифрованный текст

арбег лаяан ретюь пмок.

Приведенный пример подчеркивает лаконичность английского языка по сравнению с русским. Аналогичный пример на английском языке содержит 3 группы по 5 символов [1].

Пример 2 (перестановка). Рассмотрим шифр изгороди. Распишем текст побуквенно в две строки, а затем перепишем его построчно, группируя его по пять символов.

Получим

к м ь т р а а г б а
о п ю е н я л е р ,

шифрованный текст имеет вид

кмьтр аагба омюен ялер.

Пример 3 (замена). Мы уже обсуждали шифр Цезаря. Для математической записи этого шифра пронумеруем буквы русского алфавита (Цезарь пользовался латинским алфавитом) следуя западной традиции, начиная с нуля. Получим таблицу

а	б	в	г	д	е	ё	ж	з	и	й
0	1	2	3	4	5	6	7	8	9	10

к	л	м	н	о	п	р	с	т	у	ф
11	12	13	14	15	16	17	18	19	20	21

х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
22	23	23	25	26	27	28	29	30	31	32

Пусть A – множество букв русского алфавита. Определим следующие функции:

1) $n : A \to \mathbb{Z}$, где $n(x), x \in A$ – номер символа x согласно таблице;

2) $chr : \{0, 1, 2, ..., 32\} \to A$, где $chr(n)$ – символ с номером n согласно таблице;

3) $r_m(n)$ – остаток от деления целого числа n на ненулевое целое m.

Определим для шифра Цезаря функцию замены $g : A \to A$ по следующей формуле $g(x) = chr(r_{33}(n(x)+3))$ (к номеру символа прибавляем число 3, берем остаток от деления на 33, получаем символ с вычисленным номером).

Легко проверить, что функция $g(x)$ – биективна.

Закодируем сообщение "компьютерная алгебра". Получим следующий шифрованный текст

нсптя ахзур гвгоё здуг.

Проверим кодирование некоторых символов

$$chr(r_{33}(n(\text{к})+3)) = chr(r_{33}(11+3)) = chr(r_{33}(14)) = chr(14) = \text{н};$$

$$chr(r_{33}(n(\text{ю})+3)) = chr(r_{33}(31+3)) = chr(r_{33}(34)) = chr(1) = \text{б}.$$

Запишем обратную к $g(x)$ функцию $g^{-1} : A \to A$ по следующей формуле $g(x) = chr(r_{33}(n(x) - 3))$.

Напомним определение частного и остатка при делении целого числа a на целое b отличное от нуля. Согласно теореме 4.1.1 о делении с остатком, существуют единственные целые q и r такие, что $a = qb + r, 0 \leq r <\mid b \mid$. Число q называется неполным частным, а r остатком от деления a на b.

Приведем пример раскодирования символа:

$$chr(r_{33}(n(\text{в})-3)) = chr(r_{33}(2-3)) = chr(r_{33}(-1)) =$$

$$= chr(r_{33}(-1 = -1 \cdot 33 + 32)) = chr(32) = \text{я}.$$

При использовании одного алфавита для шифрованных сообщений, криптосистема называется одноалфавитной. Назовем криптосистемы, в которых буква шифрованного текста может представлять более одной буквы открытого текста – многоалфавитными.

Следует также различать потоковые и блочные шифры. Потоковые шифры рассматривают открытый текст как последовательность символов, подлежащих шифровке, в то время как блочные шифры делят сообщения на блоки равной длины и производят шифрование, действуя на блоках символов открытого текста. В потоковом шифре основная допустимая операция над сообщением – подстановка одного символа вместо другого, в то время как в блочном шифре помимо подстановки мы можем выполнять также перестановку. Хотя потоковые шифры сохраняют свое значение для многих приложений, блочные шифры последнее время получили широкое распространение. В качестве примеров в этом разделе рассмотрены только потоковые шифры.

Рассмотренный нами в примере 3 шифр Цезаря является одноалфавитным – все одинаковые символы сообщения кодируются одним символом в шифрованном тексте. В приведенном примере 3 в сообщении три раза встречается символ а, который в шифрованном тексте заменяется на один символ г.

При достаточно длинном сообщении одноалфавитные замены можно раскодировать, наблюдая частоту распределения символов в шифрованном тексте.

Имеются таблицы различных частот букв, например, частоты первых букв в слове, частоты последних букв в слове, частоты диграфов (т. е. частоты сочетания: буква а, за которой следует б) и т. д.

Приведем частоты появления некоторых букв в русском языке.

	%		%
Буква	Частота	Буква	Частота
о	$9,28\%$	т	$6,30\%$
а	$8,66\%$	р	$5,53\%$
е	$8,10\%$	с	$5,45\%$
и	$7,45\%$	л	$4,32\%$
н	$6,35\%$	в	$4,15\%$

Если в шифрованном тексте чаще встречается буква р, а следующая по частоте буква е, то можно предположить, что р раскодируется как о, а е как а. После угадывания нескольких букв можно угадать отдельные слова и т. д.

Будем использовать многоалфавитный шифр подстановки, скрывающий частоты букв за счет кратных подстановок. Здесь при шифровании сообщения используется более одного алфавита, и ключ включает указание, какая подстановка должна использоваться для каждого символа. Эти шифры известны также как шифры Виженера[1]

Пример 4 (многоалфавитная замена). В качестве примера рассмотрим шифр Виженера, в котором ключом является слово.

Предположим, что мы должны закодировать сообщение "компьютерная алгебра" с ключом "студент".

Пусть сообщение состоит из n символов $w_0w_1...w_{n-1}$, а ключ $k_0k_1...k_{p-1}$ – это слово длины p. Тогда i-ый символ сообщения можно закодировать при помощи функции $chr(r_{33}(n(w_i) + n(k_{r_p(i)})))$, где i – номер символа сообщения.

Рассмотрим кодирование двух первых символов сообщения:

$$chr(r_{33}(n(\text{к}) + n(\text{с}))) = chr(r_{33}(11 + 18)) = chr(r_{33}(29)) = chr(29) = \text{ь};$$

$$chr(r_{33}(n(\text{о}) + n(\text{т}))) = chr(r_{33}(15 + 19)) = chr(r_{33}(34)) = chr(1) = \text{б}.$$

Получим следующий шифрованный текст

ьбауб лоцгб дднюф чффе.

При кодировании восьмой буквы сообщения р используется первая буква ключа с (7 – длина ключа), далее вторая и так далее по циклу. При раскодировании шифров Виженера номера букв ключа отнимаются по циклу.

До XIX века шифры Виженера считались невзламываемыми. При угадывании длины ключа шифрованный текст распадается на несколько одноалфавитных текстов. Об определении длины ключа для шифров Виженера с ключевым словом рассказано в [1].

[1]Блез де Виженер – французский криптограф шестнадцатого столетия.

5(абсолютно стойкий шифр). Предположим, что ключом является достаточно длинная последовательность случайных чисел от 0 до 32, причем эта последовательность используется один раз и имеется у отправителя и получателя.

Требуется закодировать сообщение "компьютерная алгебра" с помощью последовательности случайных чисел

$$12, 31, 29, 2, 4, 15, 9, 27, 5, 8, 23, 17, 13, 5, 11, 19, 24, 16, 8, 22.$$

Если сообщение состоит из n символов $w_0w_1...w_{n-1}$, а ключ содержит первые случайные числа $l_0, l_1, ..., l_{n-1}$, то кодирование можно провести с помощью функции $chr(r_{33}(n(w_i) + l_i)))$, где i – номер символа сообщения.

Получим следующий шифрованный текст

цмиса мыяхх цпмрн чшаз.

Алгоритм раскодирования понятен без объяснения.

Упражнения

1. Функция $g(x) = chr(r_{33}(a \cdot n(x) + k)), x \in A$, где a и k – целые числа, $(a, 33) = 1$, задает модулярный шифр на русском алфавите A. Выберите a взаимно простое с числом 33 и при k равном номеру студента в списке закодируйте сообщение "криптографический анализ" с помощью модулярного шифра.

2. Закодируйте сообщение "криптографический анализ" с помощью шифра Виженера, используя в качестве ключа свою фамилию.

3. Напишите на любом алгоритмическом языке программу кодирующую и раскодирующую тексты с помощью модулярного шифра и шифра Виженера.

10.2 Коды с открытым ключом

В 1976 году была опубликована работа У. Диффи[1] и М.Э.Хеллмана[2] "Новые направления в криптографии", которая существенно изменила криптографию.

Центральным понятием данной работы являлась, так называемая, функция с секретом.

Функцией с секретом K называется функция $F_K : X \to Y$, зависящая от параметра K и обладающая следующими свойствами:

[1] У. Диффи – американский математик.

[2] М.Э.Хеллмана – американский математик.

1) существует полиномиальный алгоритм вычисления (алгоритм с временем вычисления равным Cx^t, где C – константа, x – аргумент функции, t – натуральной число) значений $F_K(x)$ для любых K и x;

2) не существует полиномиального алгоритма инвертирования функции F_K (т. е. решения уравнения $F_K(x) = y$ относительно x) при неизвестном K;

3) существует полиномиальный алгоритм инвертирования F_K при известном K.

Применение функции с секретом позволяет организовать обмен шифрованными сообщениями с использованием только открытых каналов связи.

Пользователь, который хочет получать шифрованные сообщения, должен выбрать какую-нибудь функцию F_K с секретом K. Он сообщает всем заинтересованным описание функции F_K в качестве своего алгоритма шифрования, но при этом K он никому не сообщает. А так как никто другой секрета K не знает, то никто другой и не сможет расшифровать сообщение.

Такие системы называются криптосистемами с открытым ключом, поскольку алгоритм шифрования F_K является общедоступным и открытым.

Асимметричные системы шифрования обеспечивают значительно меньшие скорости шифрования, нежели симметричные, в силу чего они обычно используются не столько для шифрования сообщений, сколько для шифрования пересылаемых между корреспондентами ключей, которые затем используются в симметричных системах.

В качестве примера применения криптосистем с открытым ключом рассмотрим крупный банк, который имеет филиалы по всему миру. Предположим, что филиал банка передает информацию о переводе некоторой суммы денег с одного номерного счета на другой. При этом информация о счетах имеет не очень высокий уровень секретности (неизвестно кому принадлежит номерной счет), но перечисляемая сумма должна быть передана точно и гарантирована от изменений несанкционированным перехватчиком. Ключ кодировки основного сообщения или электронная подпись кодируются с помощью системы с открытым ключом. В головном банке будут уверены, что необходимую сумму требуется перечислить на указанный счет.

Шифросистема *RSA*

В 1978 году Р. Ривест[1], А. Шамир[2] и Л. Адлеман[3] предложили пример функции с секретом f, на основе которой была построена реально используемая система шифрования, получившая название по первым буквам имен авторов – система RSA.

[1] Р. Ривест – американский математик.

[2] А. Шамир – американский математик.

[3] Л. Адлеман – американский математик.

Пусть n и e – натуральные числа. Функция f устроена следующим образом:

$$f(x) \equiv x^e (mod\ n), 0 \leq x, f(x) < n. \tag{33}$$

Для декодирования сообщения $a = f(x)$ достаточно решить сравнение

$$x^e \equiv a (mod\ n). \tag{34}$$

При некоторых условиях на n и e это сравнение имеет единственное решение.

Если $(e, \varphi(n)) = 1$ (здесь $\varphi(n)$ – функция Эйлера), то сравнение (34) имеет единственное решение. Для того, чтобы его найти, определим целое число d, удовлетворяющее условиям:

$$de \equiv 1 (mod\ \varphi(n)), \quad 1 \leq d < \varphi(n). \tag{35}$$

Число d находится эффективно с помощью алгоритма Евклида. Из сравнения (35) следует существование натурального r такого, что $ed = r\varphi(n) + 1$.

Теорема Эйлера утверждает, что для каждого x, взаимно простого с n выполнено сравнение $x^{\varphi(n)} \equiv 1 (mod\ n)$ и, следовательно,

$$a^d \equiv x^{de} \equiv x^{r\varphi(n)+1} \equiv \left(x^{\varphi(n)}\right)^r x \equiv x (mod\ n). \tag{36}$$

Таким образом, в предположении $(a, n) = 1$, единственное решение сравнения (34) может быть найдено в виде

$$x \equiv a^d (mod\ n). \tag{37}$$

Если дополнительно предположить, что n состоит из различных простых сомножителей, то сравнение (37) будет выполняться и без предположения $(a, n) = 1$.

Пусть все простые делители p – различны, p простой делитель числа $n = pk$. Тогда $\varphi(n) = (p-1)\varphi(k) = (p-1)t$.

Согласно малой теореме Ферма, $x^p \equiv x (mod\ p)$ для всех натуральных x. Предположим, что $x \not\vdots p$. Получим сравнения

$$a^d \equiv x^{ed} = x^{r\varphi(n)+1} \equiv x^{r(p-1)t+1} \equiv x^{rpt} x^{-rt+1} \equiv$$

$$\equiv (x^p)^{rt} x^{-rt+1} \equiv x^{rt} x^{-rt+1} \equiv x (mod\ p).$$

Возведение сравнения в отрицательную степень возможно, так как число x обратимо по модулю p.

При $x \vdots p$ сравнение (37) также справедливо по модулю p, так как его правая и левая части делятся на p.

Итак, мы установили, что сравнение (37) справедливо по модулю p при всех натуральных x и всех простых делителях p числа n. Так как n является произведением различных простых сомножителей, сравнение (37) справедливо по модулю n.

Из условия $0 \leq x < n$ следует однозначность определения числа x.

Заметим, что для вычисления функции (33) достаточно знать лишь числа n и e. Именно они и составляют открытый ключ для шифрования. А для вычисления обратной функции требуется кроме того еще знание числа d, которое и является секретом.

В шифросистеме RSA n представлено в виде произведения двух различных больших (сотни и, даже, тысячи значащих цифр) простых множителей p и q. Так как

$$\varphi(n) = \varphi(pq) = (p-1)(q-1),$$

то e нужно выбрать так, чтобы

$$(e, p-1) = (e, q-1) = 1.$$

Занумеруем буквы парами цифр, например, следующим способом.

а	б	в	г	д	е	ё	ж	з	и	й
01	02	03	04	05	06	07	08	09	10	11

к	л	м	н	о	п	р	с	т	у	ф
12	13	14	15	16	17	18	19	20	21	22

х	ц	ч	ш	щ	ъ	ы	ь	э	ю	я
23	24	25	26	27	28	29	30	31	32	33

Можно также закодировать знаки препинания и другие символы.

Для более информативного кодирования лучше использовать последовательности из 5 двоичных цифр ($33 > 2^4$).

Запишем сообщение x последовательностями цифр. Если $x > n$ разобъем сообщение на блоки $x_i, i = 1, ..., r$, где каждый блок удовлетворяет неравенству $0 < x_i < n, i = 1, ..., r$. Каждый блок преобразуем в остаток $a_i, i = 1, ..., r$ от деления x_i^e на n. Числа $a_i, i = 1, ..., r$ передаются получателю и, затем, декодируются с помощью возведения в степень d и нахождения остатка от деления на n.

Отправитель знает: n, e, сообщения $x_i, i = 1, ..., r$.

Получатель знает: n, p, q, e, кодированные сообщения $a_i, i = 1, ..., r$.

С помощью p и q вычисляется $\varphi(n) = (p-1)(q-1)$ и, затем, с помощью алгоритма Евклида d. Далее шифротекст $a_i, i = 1, ..., r$ раскодируется.

Пример. Пусть $p = 17, q = 31$, тогда $n = pq = 527$ и, следовательно, $\varphi(n) = 16 \cdot 30 = 2^5 \cdot 3 \cdot 5 = 480$.

Положим $e = 7$.

Реализуем алгоритм последовательного деления с остатком (Евклида) для чисел $\varphi(n)$ и e. Получим

$$\begin{aligned} 480 &= 68 \cdot 7 + 4, \\ 7 &= 1 \cdot 4 + 3, \\ 4 &= 1 \cdot 3 + 1, \\ 3 &= 3 \cdot 1. \end{aligned}$$

Выражая остатки, из равенств получим

$$\begin{aligned} 4 &= \varphi(n) - 68e, \\ e &= \varphi(n) - 68e + 3 \Rightarrow 3 = 69e - \varphi(n), \\ \varphi(n) - 68e &= 69e - \varphi(n) + 1 \Rightarrow -137e + 2\varphi(n) = 1. \end{aligned}$$

Преобразуя последнее равенство, получим

$$e(\varphi(n) - 137) + (2 - e)\varphi(n) = 1 \Rightarrow e \cdot 343 \equiv 1 (mod\ 527).$$

Рассмотрим сообщение "буква". Ему соответствует число 0221120301. Разобъем сообщение на блоки

$$x_1 = 221, x_2 = 120, x_3 = 301.$$

Все они меньше $n = 527$.

Отметим, что только первый блок может начинаться с нуля (иначе невозможно однозначно раскодировать первую цифру блока).

Закодируем первый блок. Число 7 имеет следующее представление в двоичной системе $7_{10} = 111_2$. Ему соответствует следующее разложение $7 = 4+2+1$.

Произведем вычисления

$$221^2 \equiv 357 (mod\ 527), 221^4 \equiv 357^2 = 442 (mod\ 527),$$

$$221^7 \equiv 442 \cdot 357 \cdot 221 \equiv 357 (mod\ 527).$$

Следовательно, $a_1 = 357$. Аналогично вычислим

$$a_2 = 511,\ a_3 = 517.$$

Для раскодирования переведем число 343 в двоичную систему счисления, используя промежуточную шестнадцатиричную

$$343_{10} = 21 \cdot 16 + 7, 21 = 1 \cdot 16 + 5,$$

$$343 = (16 + 5)16 + 7 = 16^2 + 5 \cdot 16 + 7 = 157_{16} = 1\ 0101\ 0111_2.$$

Получаем представление

$$343 = 2^8 + 2^6 + 2^4 + 2^2 + 2 + 1 = 256 + 64 + 16 + 4 + 2 + 1.$$

Раскодируем первый блок

$$357^2 \equiv 442(mod\ 527),\ 357^4 \equiv 442^2 = 374(mod\ 527),$$

$$357^8 \equiv 374^2 \equiv 221(mod\ 527),\ 357^{16} \equiv 221^2 \equiv 357(mod\ 527),$$
$$357^{32} \equiv 357^2 \equiv 442(mod\ 527),\ 357^{64} \equiv 442^2 \equiv 374(mod\ 527),$$
$$357^{128} \equiv 374^2 \equiv 221(mod\ 527),\ 357^{256} \equiv 221^2 \equiv 357(mod\ 527).$$

Окончательно получаем

$$357^{343} \equiv 357^{256+64+16+4+2+1} \equiv 357 \cdot 374 \cdot 357 \cdot 374 \cdot 442 \cdot 357 \equiv 357(mod\ 527).$$

Аналогично раскодируются остальные блоки.

Сложность нахождения секретного ключа системы RSA определяется сложностью разложения числа n на простые множители. Рекомендуется выбирать числа p и q таким образом, чтобы задача разложения числа n была достаточно сложна в вычислительном плане. Для этого рекомендуются следующие требования:

1) числа p и q должны быть достаточно большими, не слишком сильно отличаться друг от друга и в то же время быть не слишком близкими друг другу;

2) числа p и q должны быть такими, чтобы наибольший общий делитель чисел $p-1$ и $q-1$ был небольшим; желательно, чтобы $(p-1, q-1) = 2$;

3) p и q должны быть сильно простыми числами; (сильно простым называется такое простое число r, что $r+1$ имеет большой простой делитель, $r-1$ имеет большой простой делитель s такой, что число $s-1$ также обладает достаточно большим простым делителем).

В случае когда не выполнено хотя бы одно из указанных условий, имеются эффективные алгоритмы разложения числа n на простые множители.

В настоящее время самые большие простые числа, вида $n = pq$, которые удается разложить на множители известными методами, содержат в своей записи 140 десятичных знаков. Поэтому, рекомендуется выбирать числа p и q для системы RSA, содержащие не менее 100 десятичных знаков.

Упражнения

1. Пусть в шифросистеме RSA $n = 35$. Вычислите $\varphi(35)$. Пусть $e = 5$. Найдите d. Сколько символов содержит один блок? Закодировать любое трехбуквенное слово и поручить соседу раскодировать его.

2. Пусть в шифросистеме RSA $n = 55$. Вычислите $\varphi(55)$. Пусть $e = 3$. Найдите d. Сколько символов содержит один блок? Закодировать любое трехбуквенное слово и поручить соседу раскодировать его.

Шифросистема Эль-Гамаля

Шифросистема Эль-Гамаля была предложена в 1985 г. и является фактически одним из вариантов метода выработки открытых ключей Диффи-Хеллмана. Криптографическая стойкость данной системы основана на сложности проблемы логарифмирования в мультипликативной группе конечного простого поля.

Шифросистема Эль-Гамаля определяется следующим образом.

Пусть p – достаточно большое простое число, α – порождающий элемент группы $\mathbb{Z}_p^*$ – мультипликативной группы конечного поля, a – целое число из интервала $1 \leq a \leq p-2$.

Пусть $\beta \equiv \alpha^a (mod\ p)$.

Отправитель знает: p, α, β, сообщения $x_i, i = 1, ..., k$.

Получатель знает: p, a, α, β, кодированные сообщения $(b_i, c_i), i = 1, ..., k$.

Предполагается, что все сообщения x_i и шифротекст (b_i, c_i) удовлетворяют неравенствам $1 \leq x_i, b_i, c_i \leq p-1, i = 1, ..., k$.

Правило кодирования сообщений определяется формулами $b \equiv \alpha^r (mod\ p), c \equiv x\beta^r (mod\ p)$, где r – случайное число из интервала $1 \leq r \leq p-2$, которое называется рандомизатором.

Правило раскодирования определяется формулой

$$d \equiv c(b^a)^{-1} (mod\ p),$$

где число d удовлетворяет неравенству $1 \leq d \leq p-1$.

Несложно проверить, что выполнено равенство $d = x$, то есть система кодирования-раскодирования работает корректно.

В самом деле справедлива формула

$$d \equiv c(b^a)^{-1} \equiv x(\alpha^a)^r((\alpha^r)^a)^{-1} \equiv x\alpha^{ar}\alpha^{-ra} \equiv x (mod\ p).$$

Следовательно, $d - x \vdots p$. Из ограничений на x и d получаем неравенство $0 \leq d - x < p$. Вытекает равенство $d = x$.

Следует отметить, что в приведенной системе необходимо использовать различные значения рандомизатора r для кодирования различных открытых текстов x и x'. В противном случае соответствующие шифротексты (b, c) и (b', c') оказываются связанными соотношением $c(c')^{-1} = x(x')^{-1}$ и сообщение x' может быть вычислено, если известно x.

Как уже отмечалось, стойкость системы Эль-Гамаля определяется сложностью решения задачи дискретного логарифмирования в $\mathbb{Z}_p^*$. В настоящее время эта задача практически нереализуема для значений p, содержащих не менее 150 десятичных знаков. Рекомендуется также, чтобы число $p-1$ содержало бы большой простой делитель.

Система Эль-Гамаля может быть обобщена для применения в любой конечной циклической группе G. Криптографическая стойкость такой обобщенной схемы определяется сложностью задачи логарифмирования в группе G. При

этом групповая операция в G должна быть легко реализуемой. В качестве G чаще всего выбираются следующие три группы:

1) мультипликативная группа $\mathbb{Z}_p^*$ кольца классов вычетов по модулю простого числа p;

2) мультипликативная группа F_{2p}^* конечного кольца F_{2p}, где p – большое простое число;

3) группа точек эллиптической кривой над конечным полем.

Вероятностный характер шифрования можно отнести к достоинствам системы Эль-Гамаля, так как схемы вероятностного шифрования обладают, как правило, большей стойкостью по сравнению со схемами с детерминированным процессом шифрования.

Недостатком системы является удвоение длины открытого текста при шифровании.

Упражнения

1. Пусть в шифросистеме Эль-Гамаля $p = 37$, $\alpha = 5$, $r = 3$. Вычислите β. Сколько символов содержит один блок? Закодировать любое трехбуквенное слово и поручить соседу раскодировать его.

2. Пусть в шифросистеме Эль-Гамаля $p = 41$, $\alpha = 7$, $r = 2$. Вычислите β. Сколько символов содержит один блок? Закодировать любое трехбуквенное слово и поручить соседу раскодировать его.

11 Представление данных в компьютере

11.1 Представление целых чисел

Обычно в системах компьютерной алгебры рассматриваются точные аналитические преобразования и никакие округления или другие искажения целых чисел недопустимы. А потому при проведении аналитических преобразований промежуточные результаты могут потребовать значительной памяти, хотя исходные данные были невелики.

В качестве иллюстрации рассмотрим известный пример Д. Кнута и В. Брауна[1] [8] вычисления наибольшего общего делителя многочленов $(f(x), g(x))$, где

$$f(x) = x^8 + x^6 - 3x^4 - 3x^3 + 8x^2 + 2x - 5;$$

$$g(x) = 3x^6 + 5x^4 - 4x^2 - 9x + 21.$$

Проводя вычисления по алгоритму Евклида, получим в результате целое число, содержащее 35 десятичных цифр (т. е. 117 бит), и, следовательно, полиномы взаимно просты. Однако исходные данные малы (наибольшее из чисел равно 21), результат "взаимно простые" является ответом типа да/нет и требует для записи всего 1 бит. Такая проблема *разбухания промежуточных данных* возникает и в других ситуациях (например, в задаче интегрирования), и является одной из основных проблем компьютерной алгебры.

Также бывает необходимо рассматривать целые числа произвольной длины. Почти все системы компьютерной алгебры допускают такую возможность и для представления чисел выбирают в качестве основания некоторое число N, и представляют числа, по аналогии с обычной десятичной системой, относительно этого основания (с помощью "цифр" от 0 до $N-1$) с добавлением знакового бита. Например, на 32-битовых компьютерах в качестве N можно выбрать 10^9, или 2^{30}, или 2^{31}. Как правило, 2^{32} не используется, чтобы не возникали трудности при сложении двух чисел.

После выбора такого представления сложение, вычитание и умножение целых чисел становится, в принципе, простыми – достаточно воспользоваться хорошо известными из школы алгоритмами. Правда, умножение требует специально написанную на машинном языке для этого действия функцию, так как для записи в память произведения необходимы уже два слова.

Деление представляет собой значительные трудности, так как метод, изучаемый в школе, требует угадывания цифры частного. Д. Кнут в 1981 году детально описал эту проблему и предложил достаточно надежный алгоритм получения необходимой цифры.

[1] Д. Кнут и В. Браун – американские математики.

11.2 Представление дробей

Обыкновенные дроби представляются в виде пары целых чисел: числителя и знаменателя $(p/q, q \neq 0)$. Вообще говоря, не следует заменять их приближенными значениями (например, числом с плавающей точкой), так как это может привести к неверным результатам. Например, если

$$(x^3 - 8;\ \ \frac{1}{3}x^2 - \frac{4}{3}) = \frac{1}{3}x - \frac{2}{3},$$

то $(x^3 - 8;\ \ 0,333333\ x^2 - 1,33333) = 0,000001$ из-за округления.

Все операции с рациональными числами требуют вычисления наибольшего общего делителя целых чисел, что приводит к большим затратам времени. Если удается каким-то образом избежать таких действий, то это даст серьезную экономию времени. В нашем примере можно ограничиться вычислением $(x^3 - 8;\ \ x^2 - 4)$, которое не требует применения обыкновенных дробей.

Алгоритмы сложения и умножения дробей достаточно просты, но и тут можно оптимизировать процесс вычисления. Рассмотрим, например, умножение дробей, представленных в несократимом виде:

$$\frac{a}{b} \cdot \frac{c}{d} = \frac{p}{q}.$$

Очевидный способ: вычислить $p = ac,\ \ q = bd$, и затем сократить на наибольший общий делитель этих чисел.

Однако, используя равенство

$$(p, q) = (a, d) \cdot (b, c)$$

можно более эффективно вычислить два *НОД* в правой части равенства, чем один *НОД* в левой, так как аргументы у них меньше (предполагается, что $(a, b) = 1$ и.$(c, d) = 1$).

Аналогично для сложения

$$\frac{a}{b} + \frac{c}{d} = \frac{p}{q}.$$

Вместо вычисления $p = ad + bc$ и $q = bd$ и последующего сокращения на (p, q) более эффективно вычислить сначала

$$q = \frac{bd}{(b, d)} \ \text{ и } \ p = a\frac{d}{(b, d)} + c\frac{b}{(b, d)}.$$

Легко видеть, что и в этом случае мы работаем с меньшими значениями p и q, чем при прямых вычислениях.

11.3 Представление многочленов

Именно работа с полиномами (многочленами) отличает системы компьютерной алгебры от других систем.

Отметим, что термин "полиномиальные" относится к программным вычислениям, а не просто к типам данных, над которыми производятся вычисления. Например, вычисление

$$(x-y)(x+y) = x^2 - y^2$$

является полиномиальным, но им является также и

$$(\cos a - \sin b)(\cos a + \sin b) = (\cos^2 a - \sin^2 b),$$

в котором фактически произведено то же вычисление, но с заменой x на $\cos a$, а y – на $\sin b$.

Все системы компьютерной алгебры могут работать с полиномами от произвольного числа переменных. Их можно складывать, вычитать, умножать, делить (по крайней мере, если деление выполняется без остатка), но в действительности самой интересной представляется операция *упрощения.*

Отметим в связи с этим, что запись $(x-y)(x+y)$ не слишком интересна; запись x^2-y^2 гораздо более полезна. Однако запись $(x-1)$ более "простая" чем $\frac{x^2-1}{x+1}$, но является ли выражение $x^{999} - x^{998} + x^{997} - ... - 1$ более "простым", чем $\frac{x^{1000}-1}{x+1}$ далеко не очевидно. Чтобы разобраться с понятием "упрощение", сначала уточним два связанных связанных с ним термина.

Представление математических объектов (в рассматриваемом случае это многочлены) называется *каноническим,* если две различные записи соответствуют всегда двум различным объектам. Если совокупность объектов обладает структурой моноида (а почти любая совокупность в компьютерной алгебре является по меньшей мере моноидом), то можно ввести следующее определение. Представление называется *нормальным,* если нуль представляется единственным образом.

Любое каноническое представление является нормальным. Таким образом, можно сказать, что упрощение должно давать, по крайней мере, нормальное представление, а если возможно, то и каноническое. В этом случае для определения совпадения двух объектов моноида составляют их разность и смотрят ее представление: если это представление нуля, то элементы считают совпадающими, в противном случае — различными.

Кроме того, желательно, чтобы представление было "регулярным", "естественным" и "компактным" при соответствующем определении этих понятий. Мы не будем развивать эти рассуждения. Ограничимся замечанием, что двум последним требованиям, по-видимому, не удовлетворяет представление целого числа единицами, например, 7 = "1111111".

Существует несколько представлений, отвечающих указанным требованиям, и всякая система компьютерной алгебры располагает, по крайней мере, одним из них.

Для полиномов от одной переменной такие представления с математической точки зрения достаточно очевидны: каждая степень переменной x присутствует не более одного раза, и равенство полиномов сводится к равенству коэффициентов. Но в представлении ситуация не настолько проста.

Представление называется разреженным, если нулевые члены явно в нем не представлены и – плотным, если в нем явно представлены все члены. Обычное математическое представление является разреженным: мы пишем $8x^2+7$ вместо $8x^2 + 0x + 7$.

Наиболее очевидным компьютерным представлением полинома

$$a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0$$

является его представление таблицей коэффициентов

$$[a_n, a_{n-1}, \ldots, a_1, a_0].$$

Так как в нем записаны все коэффициенты, то такое представление является плотным. В этом случае для сложения двух многочленов степени n необходимо $n + 1 = O(n)$ сложений, а для умножения – $O(n^2)$ умножений коэффициентов.

Для разреженного представления каждый ненулевой член $a_i x^i$ запоминается в виде пары: (i, a^i). Например, полином $8x^2 + 7$ представляется в виде $((2, 8), (0, 7))$. Порядок задается обычно убыванием показателей степеней.

Большинство встречающихся на практике полиномов не являются плотными, и разреженное представление дает существенную экономию. Например, плотный метод потребует миллион умножений для проверки равенства

$$(x^{1000} + 1)(x^{1000} - 1) = x^{2000} - 1,$$

а разреженный – всего четыре.

Если же речь идет о полиномах от нескольких переменных, то всякий реальный полином должен быть разряжен, так как плотное представление, например, полинома от пяти переменных $x^5 y^5 z^5 u^5 v^5$ содержит 7776 членов, а разреженное — всего один.

Операции	Степень результата	Количество членов в результате
$f+g$	$max\{n_f,\ n_g\}$	m_f+m_g
$f-g$	$max\{n_f,\ n_g\}$	m_f+m_g
$f\cdot g$	n_f+n_g	$m_f\cdot m_g$
f/g	n_f-n_g	n_f-n_g+1
(f,g)	$min\{n_f,\ n_g\}$	$\leq min\{n_f,\ n_g\}+1$
Подстановка полинома f вместо переменной в полином g	$n_f\cdot n_g$	$n_f\cdot n_g+1$

Отметим, что время счета, по крайней мере для сложения и умножения, является скорее функцией количества членов, чем степени. Если f и g – два полинома степеней n_f и n_g, содержащие m_f и m_g членов, то примитивные операции над ними удовлетворяют ограничениям, представленным в приведенной выше таблице.

11.4 Представление рациональных функций

Большинство вычислений используют не только полиномы, но и их отношения, т. е. рациональные функции. К действиям с рациональными функциями естественно отнести и преобразования вида

$$\frac{1}{\sin x}+\frac{1}{\cos x}=\frac{\cos x+\sin x}{\sin x\cos x}\ ,$$

так как это то же самое, что и

$$\frac{1}{a}+\frac{1}{b}=\frac{a+b}{ab}\ .$$

Если представлять рациональную функцию как полином (числитель), деленный на другой полином (знаменатель), то получается нормальное представление, так как функция есть нуль тогда и только тогда, когда ее числитель есть нуль.

Если мы хотим получить каноническое представление, то возникают следующие сложности: например, формулы

$$\frac{x-1}{x+1}\quad\text{и}\quad\frac{x^2-2x+1}{x^2-1}$$

представляют один и тот же элемент, но это разные формулы, так как у них разные области определения.

Естественно потребовать, чтобы в каноническом представлении не существовало какого-либо общего делителя числителя и знаменателя. В общем случае приходим к представлению с минимально возможной степенью числителя (то же верно и для знаменателя). Но этого условия мало, как следует из следующего примера:

$$\frac{-2x+1}{2x+1} = \frac{2x-1}{-2x-1} = \frac{4x-2}{-4x-2} = \frac{-x+1/2}{x+1/2} = \frac{x-1/2}{-x-1/2}\ .$$

Для устранения этой неоднозначности большинство существующих систем принимают но внимание следующие правила для рациональных функций с коэффициентами из $\mathbb{Q}$:

(1) в выражении не должно быть рациональных коэффициентов (что отбрасывает последние два выражения);

(2) никакое целое число не может делить как числитель, так и знаменатель (что отбрасывает третье выражение);

(3) старший коэффициент знаменателя должен быть положительным (что отбрасывает второе выражение).

Имеются и другие возможности, но эти правила наиболее распространены и достаточны для получения канонической формы.

Упражнения

1. Написать программы для сложения, вычитания, умножения и деления с остатком целых чисел.

2. Написать программы для выполнения арифметических операций над дробями.

3. Написать программы для сложения, вычитания, умножения и деления с остатком многочленов представленных в плотном или разреженном виде.

12 Решение алгоритмических задач в кольце многочленов

12.1 Системы уравнений и идеалы в кольцах многочленов

Как было отмечено ранее, кольцо целых чисел $\mathbb{Z}$ и кольцо многочленов от одной переменной $F[x]$ над полем F являются кольцами главных идеалов. Следующий пример показывает, что $F[x,y]$ таковым не является.

Пример. В кольце многочленов $F[x,y]$ множество многочленов, свободный член которых равен нулю, образуют идеал I_0. Покажем, что он не является главным. Предположим, что $I_0 = (f)$ для некоторого $f \in F[x,y]$. Так как $x, y \in I_0$, существуют многочлены $t, s \in F[x,y]$ такие, что $x = ft, y = fs$.

Следовательно, $x - y \vdots f$. Единственный общий делитель различных неприводимых многочленов x и y – это ненулевая константа.

Если $f \in F, f \neq 0$, то $(f) = F[x,y]$. Получили противоречие.

Следовательно, идеал I_0 не является главным.

В коммутативных кольцах можно уточнить понятие множества, порождающего идеал.

Определение. *Элементы $a_1, a_2, ..., a_m \in I$ образуют базис идеала I коммутативного кольца R, если любой элемент $a \in I$ можно представить в виде:*

$$a = a_1\ r_1 + a_2\ r_2 + \cdots + a_m\ r_n \text{ для некоторых } r_i \in R.$$

Обозначается $I = (a_1, a_2, ..., a_m)$.

Определение. *Говорят, что идеал $I \lhd R$ допускает конечный базис, если найдутся такие элементы $a_1, a_2, ..., a_m \in I$, что $I = (a_1,\ a_2, ..., a_m)$.*

Заметим, что в определение базиса идеала (в отличие от векторного пространства) нет требования минимальности на число элементов базиса. Добавив к базису произвольный элемент идеала, мы вновь получим базис идеала.

На рубеже XIX и XX веков Давидом Гильбертом[1] была доказана следующая фундаментальная теорема:

Теорема 12.1.1 (Теорема Гильберта о базисе) *Каждый идеал кольца многочленов $F[x_1, ..., x_n]$ от n переменных над полем F допускает конечный базис.*

Пусть задана некоторая система алгебраических уравнений:

$$\begin{cases} f_1(x_1, \ldots, x_n) = 0 \\ f_2(x_1, \ldots, x_n) = 0 \\ \quad \ldots\ldots\ldots\ldots\ldots\ldots \end{cases} \tag{38}$$

Мы можем сопоставить ей некоторый идеал $I \lhd F[x_1, ..., x_n]$, порожденный многочленами, отвечающими уравнениям системы:

$$I = (f_1(x_1, \ldots, x_n),\ f_2(x_1, \ldots, x_n),\ \ldots).$$

Если систему (38) обозначить для краткости буквой S, то соответствующий идеал будем обозначать $I(S)$.

[1] Давидом Гильберт – немецкий математик (1862-1943).

Лемма 12.1.1 *Если $f \in I(S)$, то $f(\alpha_1, \ldots, \alpha_n) = 0$ для всякого решения $(\alpha_1, \ldots, \alpha_n)$ системы S.*

Доказательство. Из того что $f \in I(S)$ следует, что $f = f_1 r_1 + \ldots + f_m r_m$. Таким образом,

$$f(\alpha_1, \ldots, \alpha_n) = f_1(\alpha_1, \ldots, \alpha_n) r_1(\alpha_1, \ldots, \alpha_n) + \ldots$$

$$\ldots + f_m(\alpha_1, \ldots, \alpha_n) r_m(\alpha_1, \ldots, \alpha_n) =$$

$$= 0 \cdot r_1(\alpha_1, \ldots, \alpha_n) + \ldots + 0 \cdot r_m(\alpha_1, \ldots, \alpha_n) = 0.$$

□

Предложение 12.1.1 *Пусть $\{f_1, \ldots, f_k\}$ и $\{g_1, \ldots, g_m\}$ – два базиса одного и того же идеала $I \lhd F[x_1, ..., x_n]$. Тогда следующие системы*

$$\begin{cases} f_1(x_1, \ldots, x_n) = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ f_k(x_1, \ldots, x_n) = 0 \end{cases} \tag{39}$$

и

$$\begin{cases} g_1(x_1, \ldots, x_n) = 0 \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ g_m(x_1, \ldots, x_n) = 0 \end{cases} \tag{40}$$

эквивалентны.

Доказательство. Пусть $(\alpha_1, \ldots, \alpha_n)$ является решением системы (39), тогда так как $g_i \in I = (f_1, \ldots, f_k)$, то в силу леммы 12.1.1 следует, что $g_i(\alpha_1, \ldots, \alpha_n) = 0$. Аналогично, каждое решение системы (40) является решением системы (39). □

Таким образом, множество решений системы однозначно определяется идеалом системы. Различные базисы одного идеала отвечают эквивалентным системам. А из теоремы Гильберта о базисе получим

Следствие 12.1.1 *Каждая система алгебраических уравнений от n неизвестных эквивалентна конечной системе.*

Всякая система алгебраических уравнений от одного неизвестного эквивалентна системе из одного уравнения.

12.2 Базис Грёбнера полиномиального идеала

При решение различных алгоритмических задач, связанных с идеалами, порожденными конечными множествами многочленов от n переменных, большую роль играет понятие базиса Грёбнера полиномиального идеала. Это понятие возникло в математике сравнительно недавно: оно было введено Б. Бухбергером[1] в 1965 году и названо в честь В.Грёбнера[2] – научного руководителя Бухбергера. Бухбергер также разработал основные алгоритмы построения базисов.

При рассмотрении алгоритма деления в кольце $F[x]$ и алгоритма приведения системы линейных уравнений к ступенчатому виду методом Гаусса ключевым моментом является понятие *упорядочивания членов многочлена* (хотя это, как правило, не подчеркивается).

Для определения базиса Грёбнера идеала $I \lhd F[x_1, ..., x_n]$ нам потребуется ввести понятие упорядочивания членов многочлена из $F[x_1, ..., x_n]$ и понятие старшего члена многочлена $f(x_1, \ldots, x_n)$.

Заметим, что каждому одночлену $ax_1^{i_1}...x_n^{i_n} \in F[x_1, ..., x_n]$ можно сопоставить набор $(i_1, ..., i_n)$ целых неотрицательных чисел, который называется *набором степеней*. Если задать упорядочивание набора степеней, то получим упорядочивание и одночленов.

Определение. *Будем говорить, что набор* $(i_1, ..., i_n)$ *больше набора* $(j_1, ..., j_n)$, *если существует такое* k, $k \leq n$, *что* $i_1 = j_1$, $i_2 = j_2$, ..., $i_{k-1} = j_{k-1}$, $i_k > j_k$.

Такой способ упорядочивания наборов называется *чисто лексикографическим.*

Определение. *Будем говорить, что одночлен* $ax_1^{i_1}...x_n^{i_n}$ *больше* $bx_1^{j_1}...x_n^{j_n}$, *если* $(i_1, ..., i_n)$ *больше* $(j_1, ..., j_n)$.

Замечание. Необходимо отметить, что существует много лексикографических упорядочиваний - каждому упорядочиванию переменных отвечает свое.

Рассматривается также *однородный лексикографический* порядок: у двух наборов $(i_1, ..., i_n)$ и $(j_1, ..., j_n)$ вначале сравниваются степени $\sum_{t=1}^{n} i_t$ и $\sum_{t=1}^{n} j_t$ и только в случае их совпадения наборы сравниваются лексикографически.

Определение. *Старшим членом* f_C *многочлена*

$$f(x_1, ..., x_n) = \sum a_{i_1...i_n} x_1^{i_1}...x_n^{i_n}$$

называется наибольший ненулевой одночлен, встречающийся в $f(x_1, ..., x_n)$.

[1] Б. Бухбергер – немецкий математик.

[2] В.Грёбнер – немецкий математик.

Пусть $\mathcal{F} = \{f_1, \ldots, f_m\}$ – конечное множество многочленов из $F[x_1, ..., x_n]$. Говорят, что многочлен g *редуцирован относительно* $\mathcal{F}$, если ни один из старших членов $f_{1C}, \ldots, f_{mC}$ не делит старшего члена g_C.

Если же g не является редуцированным относительно $\mathcal{F}$, т. е. $g_C = f_{iC}q$ для некоторого f_{iC} (q – одночлен), то положив $g_1 = g - f_i q$, получим $g_{1C} < g_C$. Этот процесс называется *редукцией многочлена* g *с помощью многочлена* f_{iC}. Будем обозначать $g \xrightarrow{f_i} g_1$

Отметим, что многочлены g и g_1 лежат в одном смежном классе относительно идеала I, порожденного множеством $\mathcal{F}$.

Следовательно, $g \in I = (f_1, ..., f_m)$ в том и только том случае, если $g_1 \in I$.

Определение. *Базис* $\mathcal{F} = \{f_1, \ldots, f_m\}$ *идеала* I *называется базисом Грёбнера этого идеала, если всякий многочлен* $g \in I$ *редуцируется относительно* $\mathcal{F}$ *к нулю.*

Заметим, что не всякий базис идеала является его базисом Грёбнера. Для построения базиса Грёбнера нам потребуются следующие определения.

Пусть $\mathcal{F} = \{f_1, \ldots, f_m\}$ – базис идеала I. Говорят, что многочлены f_i и f_j имеют *зацепление*, если $(f_{iC}, f_{jC}) \notin F$. В этом случае рассмотрим многочлен

$$S(f_i, f_j) = \alpha \frac{h}{f_{iC}}\ f_i - \beta \frac{h}{f_{jC}}\ f_j,$$

где $h = [f_{iC}, f_{jC}]$, α и β выбираются так, чтобы уравнять коэффициенты при старших членах уменьшаемого и вычитаемого, называемый S-*полиномом* пары f_i и f_j.

Будем говорить, что зацепление *разрешимо*, если S-полиномом $S(f_i, f_j)$ редуцируется относительно $\mathcal{F}$ к нулю.

Справедлива следующая теорема.

Теорема 12.2.1 *Базис* $\mathcal{F} = \{f_1, \ldots, f_m\}$ *идеала* I *является базисом Грёбнера тогда и только тогда, когда в нем нет зацеплений или каждое зацепление редуцируется к нулю.*

Алгоритм Бухбергера построения базиса Грёбнера полиномиального идеала:

Пусть $\mathcal{F} = \{f_1, \ldots, f_m\}$ – базис идеала I.

1) Проверим, есть ли в наборе многочленов зацепления. Если нет, то $\mathcal{F}$ является базисом Грёбнера идеала I.

2) Для найденного зацепления многочленов f_i и f_j строим S-полином $S(f_i, f_j)$ и редуцируем его относительно $\mathcal{F}$. Если он редуцируется к нулю, то все хорошо. Если же он редуцируется к ненулевому многочлену f, то добавим его к набору $\mathcal{F}$ в качестве f_{m+1}.

3) В построенном множестве многочленов $\mathcal{F}' = \{f_1, \ldots, f_m, f_{m+1}\}$ рассматриваем зацепление, которое не было рассмотрено ранее и т.д.

За конечное число шагов мы получим набор многочленов $f_1, \ldots, f_m, f_{m+1}, \ldots, f_k$, в котором каждое зацепление разрешимо. Это и будет базисом Грёбнера идеала $I = (f_1, \ldots, f_m)$.

Следующая лемма позволяет ускорить алгоритм Бухбергера нахождения базиса Грёбнера.

Лемма 12.2.1 *Пусть требуется найти базис Грёбнера идеала, порожденного множеством $\mathcal{F} = \{f_1, \ldots, f_m\}$. Пусть*

$$[f_{iC}, f_{jC}] \vdots f_{kC}, \text{ где } f_i, f_j, f_k \in \mathcal{F}.$$

Тогда если рассмотрены S-полиномы $S(f_i, f_k)$ и $S(f_j, f_k)$, то не требуется рассматривать S-полином $S(f_i, f_j)$.

Доказательство леммы можно прочитать в [7].

Пример. Найти базис Грёбнера идеала $I = (f_1, f_2, f_3)$ в кольце $\mathbb{R}[x, y, z]$, если $f_1 = x^2 + y^2 + z^2$, $f_2 = x + y - z$, $f_3 = y + z^2$ (для чисто лексикографического упорядочивания при $x > y > z$).

Так как $f_{1C} = x^2, f_{2C} = x$, то многочлены f_1 и f_2 имеют зацепление.

$$S(f_1, f_2) = f_1 - xf_2 = y^2 + z^2 - xy + xz \xrightarrow{f_2}$$

$$\xrightarrow{f_2} 2y^2 + 2z^2 - 2yz \xrightarrow{f_3} 2z^4 + 2z^3 + 2z^2 \neq 0.$$

Добавив его в качестве многочлена $f_4 = 2z^4 + 2z^3 + 2z^2$ (на константу можно сократить и считать $f_4 = z^4 + z^3 + z^2$), получим набор $\mathcal{F} = \{f_1, f_2, f_3, f_4\}$. Так как больше зацеплений нет, то $\mathcal{F}$ является базисом Грёбнера идеала I.

Пусть $\mathcal{F} = \{f_1, \ldots, f_m\}$ – базис Грёбнера идеала I. Можно ли его упростить?

Во-первых, если f_1 и f_2 – элементы базиса Грёбнера идеала I и f_{1C} делится на f_{2C}, то f_1 можно удалить. Оставшиеся элементы $f_2, \ldots, f_m$ по-прежнему образуют базис Грёбнера идеала I.

Определение. *Базис Грёбнера $\mathcal{F} = \{f_1, \ldots, f_m\}$ идеала I называется минимальным, если f_{iC} не делится на f_{jC} при всех $i \neq j$.*

Второе упрощение касается нестарших членов многочленов $f_1, \ldots, f_m$. Предположим, что некоторый член q многочлена f_i делится на старший член многочлена f_j. Редуцируем q с помощью f_j и полученный результат запишем в f_i вместо q. При данной операции число элементов в базисе Грёбнера не измениться, но понижаются степени членов многочленов $f_1, \ldots, f_m$.

Определение. *Базис Грёбнера $\mathcal{F} = \{f_1, \ldots, f_m\}$ идеала I называется редуцированным, если все старшие коэффициенты равны 1 и ни один член многочлена f_i не делится на f_{jC} при всех $i, j = 1, \ldots, m,\ i \neq j$.*

Каждый базис Грёбнера можно свести к редуцированному, более того имеет место следующая теорема.

Теорема 12.2.2 *Редуцированный базис Грёбнера идеала $I \lhd F[x_1, \ldots, x_n]$ определен однозначно, т. е. не зависит от выбора исходного базиса идеала I и от последовательности проводимых операций (но зависит от упорядочивания мономов).*

Пример. Рассмотрим базис Грёбнера

$$f_1 = x^2 + y^2 + z^2,\ f_2 = x + y - z,\ f_3 = y + z^2,\ f_4 = z^4 + z^3 + z^2$$

из примера 12.2.

1) Так как $f_{1C} = x^2$ делится на $f_{2C} = x$, то многочлен f_1 можно отбросить, и $\{f_1, f_2, f_3\}$ является минимальным базисом Грёбнера.

2) Применив редукцию $y \xrightarrow{f_3} -z^2$ и заменив y на $-z^2$ в f_2, получим $f_2 \dashrightarrow x - z^2 - z$.

Таким образом, $\{x - z^2 - z,\ y + z^2,\ z^4 + z^3 + z^2\}$ – редуцированный базис Грёбнера.

С помощью базисов Грёбнера можно решать следующие задачи:

1) задача о принадлежности идеалу;

2) решение систем полиномиальных уравнений;

3) задача нахождения неявного представления.

В приложении дано описание пакета для работы с базисами Грёбнера в система *Maple*.

Упражнения

1. Найти старший член многочлена относительно лексикографического порядка $x_1^2x_2x_3^3 + 3x_1x_2^4x_3^2 - 2x_1^2x_3$. Как изменится старший член если рассмотреть: а) упорядочивание по степени; б) лексикографический порядок со следующим порядком на образующих: $x_2 > x_1 > x_3$?

2. Найти базис Грёбнера идеала, порожденного многочленами $f_1 = x^2 - 1, f_2 = (x-1)y, f_3 = (x+1)z$.

3. Найти базис Грёбнера идеала, порожденного многочленами $f_1 = xz - 2y + 1, f_2 = yz - 1 + z, f_3 = yz + xyz + z$.

4. Найти базис Грёбнера идеала, порожденного многочленами $f_1 = x^3yz - xz^2, f_2 = xy^2z - xyz, f_3 = x^2y^2 - z$.

13 Формальное дифференцирование и интегрирование

13.1 Постановка задачи

Компьютеры хорошо справляются с задачами численного интегрирования, т. е. находят определенные интегралы. Решение таких задач было одним из первых применений вычислительных машин. Чуть позднее реализовано формальное дифференцирование и значительно позднее – формальное интегрирование, т. е. нахождение неопределенного интеграла. В этом и состоит одно из главных достижений компьютерной алгебры.

Заметим, что формальное дифференцирование принципиально отличается от формального интегрирования. Дифференцирование является алгоритмической процедурой. Знание производных элементарных функций и следующих четырех правил:

1) $(f \pm g)' = f' \pm g'$;

2) $(f \cdot g)' = f' \cdot g + f \cdot g'$;

3) $(\frac{f}{g})' = \frac{f' \cdot g - f \cdot g'}{g^2}$;

4) $(f(g(x)))' = f'(g(x)) \cdot g'(x)$.

дает возможность продифференцировать произвольную функцию. Основной проблемой при дифференцировании является упрощение результата, так как если его не упрощать, то, производная от $3x^2+1$ выдается в виде $0 \cdot x^2 + 6 \cdot x + 0$.

Интегрирование же не является алгоритмической процедурой, а выглядит случайным набором правил и частных случаев. Существует только одно более или менее общее правило

$$\int (f \pm g)\, dx = \int f\, dx \pm \int g\, dx,$$

которое, кстати, не всегда работает. Для других комбинации функций (отличных от сложения и вычитания) общих правил не существует.

Например, известно, как интегрировать функции $f(x) = e^x$ и $g(x) = x^2$. Но их суперпозиция $f(g(x)) = e^{x^2}$ уже не поддается интегрированию в том смысле, что её невозможно выразить в виде суперпозиции четырех арифметических действий и элементарных функции, к которым причисляют

$$x^a, \quad a^x, \quad \sin x, \quad \cos x, \quad \ln x, \quad |\, x \,|,$$

а также обратные тригонометрические функции. Такие интегралы называют “неберущимися” в виде конечной комбинации элементарных функций.

Отметим, что часто встречающиеся функции, выражаемые такими "неберущимися" интегралами (например, $\int e^{-x^2}\,dx$, $\int \frac{\sin x}{x}\,dx$) изучают отдельно и называются *специальными функциями*. Отметим в связи с этим, что на практике постоянно встречаются новые неберущиеся интегралы.

Алгоритмическая трудность формального интегрирования состоит в том, что априори неизвестно, какую комбинацию известных методов нужно применить, чтобы найти данный интеграл. Более того, мы не знаем, можно ли это сделать в принципе, располагая описанным набором элементарных функций. В настоящее время разработана развитая система формального интегрирования.

Так как дифференцирование проще, чем интегрирование, имеет смысл переформулировать проблему интегрирования как "обратную задачу" для дифференцирования, т. е. для данной функции f вместо отыскания ее интеграла будем искать такую функцию g, что $g' = f$.

Для двух данных классов функций A и B *задача интегрирования из A в B* состоит в том, чтобы найти алгоритм, который для любого элемента $f \in A$ либо выдает элемент $g \in B$, такой, что $g' = f$, либо доказывает, что в B не существует такого элемента g, что $f = g'$.

Пример. Пусть $A = B = \mathbb{Q}(x)$ – множество рациональных функций, где в качестве числителя и знаменателя могут быть произвольные полиномы. Тогда, если $f(x) = 1/x^2$, то $g(x) = -1/x$. Если же $f(x) = 1/x$, то элемента g не существует. Если же $A = B = \mathbb{Q}(x, \ln x)$ т. е. B – множество рациональных дробей, где в качестве числителя и знаменателя могут быть произвольные полиномы и логарифмы, то для $f(x) = 1/x$ ответом будет $g(x) = \ln x$.

Заметим, что не для любых классов A и B проблема интегрирования разрешима. Ричардсон доказал, что она неразрешима в классах

$$A = B = \mathbb{Q}(i, \pi, \exp, \log, \mathrm{abs}),$$

где abs обозначает абсолютную величину.

13.2 Интегрирование рациональных функций

Любая рациональная функция f может быть записана в виде $f = p + q/r$, где $p, q,\ r$ – полиномы, и q и r – взаимно просты и $\deg q < \deg r$. Мы уже отмечали, что $\int (f + g)\,dx = \int f\,dx + \int g\,dx$, но необходимо соблюдать осторожность, так как первый интеграл может быть выражен в явном виде, а оба интеграла в правой части будут "неберущимися". Хорошо известен следующий пример: $\int (x^x + (\ln x)x^x)\,dx = x^x + c$, но оба интеграла $\int x^x dx$ и $\int \ln x\; x^x\,dx$ неинтегрируемы в описанном выше смысле. Таким образом, приведенным соотношением можно пользоваться только в том случае, если известно, что существуют два из трех интегралов.

Для любого полинома p существует $\int p\,dx$, следовательно, проблема интегрирования сводится к задачам интегрирования полинома p, которая совсем проста, и правильной рациональной функции q/r.

Если полином r разлагается на линейные множители $r(x) = \prod_{i=1}^{n_i}(x - a_i)^{n_i}$, то, как известно, правильная рациональная функция q/r разлагается на простейшие дроби, причем единственным образом:

$$q/r = \sum_{i=1}^{n}\sum_{j=1}^{n_i} \frac{b_{ij}}{(x - a_i)^j} \quad ,$$

где b_{ij} – числа.

В этом случае рациональная дробь имеет интеграл, который выражается в виде суммы рациональных дробей и логарифмов от рациональных функций с постоянными коэффициентами.

Отметим недостатки этого подхода с алгоритмической точки зрения:

1) провести разложение полинома r на линейные множители не всегда удается в исходном поле; требуется сделать его алгебраическое расширение, что весьма нежелательно;

2) разложение на множители полинома r достаточно высокой степени – очень дорогостоящая операция;

3) достаточно сложно найти представление q/r в виде суммы простейших дробей.

В то же время некоторые интегралы от рациональных дробей легко находятся и без разложения на сумму простейших дробей. Например,

$$\int \frac{5x^4 + 1}{(x^5 + x + 1)^2}\, dx = -\frac{1}{x^5 + x + 1} + C.$$

Данный пример показывает, что даже в тех случаях, когда знаменатель не разлагается на множители над $\mathbb{Q}$, можно провести интегрирование.

Отсюда вытекает следующая проблема: найти алгоритм интегрирования рациональных дробей, который работает только с теми алгебраическими величинами, которые необходимы для выражения интеграла.

Перед тем как изложить методы решения этой проблемы, рассмотрим *разложение полинома на свободные от квадратов множители.*

Предположим, что полином r разлагается в произведение линейных множителей $r(x) = \prod_{i=1}^{n_i}(x - a_i)^{n_i}$. Ясно, что r можно представить в виде $\prod r_i^i$, где r_i – произведение сомножителей кратности i. В этом разложении полинома r у каждого полинома r_i нет кратных сомножителей и различные полиномы r_i взаимно просты. Это разложение называется *разложение полинома r на свободные от квадратов множители.*

Метод Эрмита позволяет определить рациональную часть интеграла рациональной функции q/r без использования дополнительных алгебраических величин.

Предположим, что r имеет разложение на свободные от квадратов множители $\prod_{i=1}^{n} r_i^i$. Так как полиномы r_i взаимно просты, то подынтегральную дробь можно разложить на простейшие дроби:

$$\frac{q}{r} = \frac{q}{\prod_{i=1}^{n} r_i^i} = \sum_{i=1}^{n} \frac{q_i}{r_i^i} \ .$$

Так как каждый элемент в правой части обладает интегралом, достаточно проинтегрировать поочередно каждый элемент. В силу взаимной простоты полиномов r_i и r_i' существуют такие полиномы a и b, что $ar_i + br_i' = 1$. Тогда (при $i \neq 1$) получим

$$\int \frac{q_i}{r_i^i}\, dx = \int \frac{q_i(ar_i + br_i')}{r_i^i}\, dx = \int \frac{q_i a}{r_i^{i-1}}\, dx + \int \frac{q_i b r_i'}{r_i^i}\, dx =$$

$$= \int \frac{q_i a}{r_i^{i-1}}\, dx + \int q_i b\, d(\frac{1}{(i-1)r_i^{i-1}}) =$$

$$= \int \frac{q_i a}{r_i^{i-1}}\, dx - \frac{q_i b}{(i-1)r_i^{i-1}} + \int \frac{(q_i b)'}{(i-1)r_i^{i-1}}\, dx =$$

$$= -\frac{q_i b}{(i-1)r_i^{i-1}} + \int \frac{q_i a + (q_i b/(i-1))'}{r_i^{i-1}}\, dx \ ,$$

и нам удалось понизить степень полинома r_i.

Продолжим подобным образом до тех пор, пока этот показатель не станет равным 1, а оставшийся интеграл – суммой логарифмов.

Приложение: вычисление базисов Гребнера в системе компьютерной алгебры *Maple*

Во многих системах компьютерной алгебры реализован алгоритм Бухбергера для вычисления базисов Грёбнера, обычно эти системы находят базис, элементы которого отличаются от элементов редуцированного базиса постоянным множителем. К таким системам можно отнести *Maple* и *Mathematica*.

Мы не пытаемся дать общее введение для пользователя данной системы – это задача специального курса, а ставим целью лишь познакомить с пакетом для работы с базисами Грёбнера.

Данный пакет вызывается командой:

$$> \ with(grobner);$$

($>$ – это приглашение Maple, каждый оператор в Maple заканчивается точкой с запятой). После загрузки пакета можно делить полиномы, вычислять базис Грёбнера и производить другие вычисления, описанные далее.

Мономиальное упорядочение в Maple называется *termorder*. Из мономиальных упорядочений, рассмотренных нами, Maple работает с чисто (pure) лексикографическим *plex* и однородно лексикографическим *tdeg* (от total degree – полная степень).

Так как мономиальное упорядочение зависит от порядка переменных, то Maple должен знать *termorder* (*plex* или *tdeg*) и список переменных. Например, если вы хотите использовать чисто (pure) лексикографическое упорядочивание с $x > y > z$, то вам необходимо ввести *plex* и список $[x, y, z]$ (списки в *Maple* заключаются в квадратные скобки). Если *termorder* не введен, то по умолчанию используется *tdeg*. Список переменных должен быть обязательно введен, это нельзя сделать по умолчанию.

Наиболее часто используемые команды из пакета Грёбнер – *normalf* (алгоритм деления) и *gbasis* (вычисление базиса Грёбнера). Имя *normalf* (нормальная форма) означает редукцию многочлена . Эта команда имеет следующий синтаксис:

$$> \ normalf(f, polylist, varlist, termorder);$$

Результатом является редукция полинома f относительно полиномов из списка *polylist* при использовании мономиального упорядочения, заданного *termorder* и *varlist*. Например, редукция полинома $x^3 + 3y^2$ относительно полиномов $x^2 + y$ и $x + 2xy$ для однородного лексикографического упорядочения с $x > y$ осуществляется следующей командой:

$$> \ normalf(x\hat{}3 + 3 * y\hat{}2, [x\hat{}2 + y, x + 2 * x * y], [x, y]);$$

Мы опустили указание *termorder*, так как *tdeg* используется по умолчанию. Основным полем здесь является $\mathbb{Q}$. Обратите внимание на то, что *normalf* не выдает частных в алгоритме деления.

Команда *gbasis* означает, разумеется, "базис Грёбнера". Она имеет следующий синтаксис:

$$> \ gbasis(polylist, varlist, termorder);$$

и вычисляет базис Грёбнера идеала, порожденного полиномами из списка *polylist* по отношению к мономиальному упорядочению, заданному *termorder* и *varlist*. Результатом является редуцированный базис Грёбнера с целыми коэффициентами. Например, в результате выполнения команды

$$> \ gb \ := \ gbasis([x\hat{}2 + y, 2 * x * y + y\hat{}2], [x, y], plex);$$

выдается список (с именем *gb*) полиномов, которые образуют базис Грёбнера идеала $(x^2 + y, 2xy + y^2) \ \subset \ \mathbb{Q}[x, y]$ для чисто лексикографического упорядочения с $x > y$.

Если вы работаете с полиномами с целыми или рациональными коэффициентами и командами *normalf* и *gbasis*, то *Maple* будет считать основным полем поле $\mathbb{Q}$.

Пакет для вычисления базиса Грёбнера содержит и другие полезные команды:

- *leadmon* для вычисления старшего коэффициента $LC(f)$ и старшего монома $LM(f)$ для полинома f ($f_C = LC(f) \cdot LM(f)$);
- *spoly* для вычисления $S-$полинома $S(f, g)$;
- *solvable* для определения существования решений данной полиномиальной системы над алгебраически замкнутым полем;
- *finit e* для определения конечности или бесконечности числа решений данной полиномиальной системы над алгебраически замкнутым полем;
- *solve* для решения (нахождения полного списка решений) систем уравнений.

Библиотека содержит также пакет *"GB"*, который вычисляет остатки от деления и базисы Грёбнера по модулю простых чисел. Для его загрузки необходимо ввести команды:

$$> \ with(share);$$

$$> \ readshare(`mod/GB`);$$

Maple содержит отличную диалоговую систему help, позволяющую легко освоить команды. Например, команда

$$> \ help(GB);$$

объяснит, как пользоваться командой *"GB"*. Интересной опцией этой команды является переменная *infolevel[GB]*. Задав ее значение, можно получать информацию о ходе вычисления базиса Грёбнера.

Список использованных источников

1. Акритас, А. Основы компьютерной алгебры с приложениями / А. Акритас.- Москва: Мир, 1994.- 544 с.

2. Алферов, А.П. Основы криптографии / А.П. Алферов, А.Ю. Зубов, А.С. Кузьмин, А.В. Черемушкин.- Москва: Гелиос АРВ, 2001.- 480 с.

3. Аржанцев, И.В. Базисы Грёбнера и системы алгебраических уравнений / И.В. Аржанцев.- Москва: МЦНМО, 2003.- 68 с.

4. Винберг, Э.Б. Курс алгебры / Э.Б. Винберг.- М.: Факториал Пресс, 2001.- 557 с.

5. Глухов, М.М. Алгебра / М.М. Глухов, В.П. Елизаров, А.А. Нечаев.- Т. 1.: Гелиос АРВ, 2003.- 336 с.

6. Глухов, М.М. Алгебра / М.М. Глухов, В.П. Елизаров, А.А. Нечаев.- Т. 2.: Гелиос АРВ, 2003.- 416 с.

7. Дэвенпорт, Дж. Компьютерная алгебра / Дж. Дэвенпорт, И. Сирэ, Э. Турнье.- Системы и алгоритмы алгебраических вычислений.- Москва: Мир, 1991.- 350 с.

8. Кнут, Д. Искусство прогаммирования на ЭВМ / Д. Кнут.- Т. 2: Получисленные алгоритмы.- М.: Мир, 1977.- 724 с.

9. Матрос, Д.Ш. Элементы абстрактной и компьютерной алгебры / Д.Ш. Матрос, Г.Б. Поднебесова.- М.: Издательский центр "Академия", 2004.-240 с.

10. Новиков, Ф.А. Дискретная математика для программистов / Ф.А. Новиков.- СПб.: Изд-во "Питер 2000.- 304 с.

11. Устян А.Е. Алгебра и теория чисел. В 2-х частях. Тула: Изд-во Тул.гос.пед.ун-та, 2002.

Список обозначений

$S(X)$ – симметрическая группа множества X

S_n – симметрическая группа подстановок множества $\{1, ..., n\}$

$H \lhd G$ – нормальная подгруппа группы G

G/H – фактор-группа группы G по нормальной подгруппе H

$\varphi(n)$ – функция Эйлера натурального числа n

$U(R)$ – группа обратимых элементов кольца

$a \equiv b(mod\ m)$ – целые числа a и b сравнимы по модулю $m \Leftrightarrow a - b \vdots m$

$\mathbb{Z}_m$ – фактор-кольцо $\mathbb{Z}_m = \mathbb{Z}/m\mathbb{Z}$ кольца целых чисел $\mathbb{Z}$ по идеалу $m\mathbb{Z}$ (кольцо классов вычетов по модулю m)

$\hat{a} = a + m\mathbb{Z}, a \in \mathbb{Z}$ – элемент кольца $\mathbb{Z}_m$

$F_p[x]_{m(x)} = F_p[x]/m(x)F_p[x]$ – фактор-кольцо кольца многочленов $F_p[x]$ по идеалу $m(x)F_p[x]$

$x\Delta y = (x \wedge Cy) \vee (y \wedge Cx)$ – симметрическая разность для булевых алгебр

H^T – матрица, транспонированная к матрице H

(a, b) – наибольший общий делитель целых чисел a и b

$[a, b]$ – наибольшее общее кратное целых чисел a и b

$(f(x), g(x))$ – наибольший общий делитель многочленов $f(x)$ и $g(x)$

$[f(x), g(x)]$ – наибольшее общее кратное многочленов $f(x)$ и $g(x)$

$n(x)$ – номер символа x в алфавите A начиная с нуля

$chr(n)$ – символ алфавита A с номером n

$r_m(n)$ – остаток от деления целого числа n на ненулевое целое m

Предметный указатель