ЭЛЕМЕНТЫ АЛГЕБРАИЧЕСКОЙ ТЕОРИИ КОДОВ

Коды и их основные параметры.

Пусть Ω — конечное множество (алфавит), $q = |\Omega| > 1$, Ω^n — множество *слов* длины n.

Кодом длины n над алфавитом Ω называется любое непустое подмножество C множества строк Ω^n. Элементы множества C называются *кодовыми словами* кода C.

Размерностью кода C называется (действительное) число $\log_q |C|$.

Расстоянием Хэмминга между словами $a = (a_1, \ldots, a_n)$ и $b = (b_1, \ldots, b_n)$ называется число $d(a, b) = |\{i \in \overline{1, n} : a_i \neq b_i\}|$. *Расстоянием* кода C называется число $d(C) = \min\{d(a, b) : a, b \in C \text{ и } a \neq b\}$.

Код длины n над алфавитом из q элементов, имеющий размерность k и расстояние d, называется $[n, k, d]_q$*-кодом*.

Свойства расстояния Хэмминга.

Для любых слов $a, b, c \in \Omega^n$ справедливы следующие утверждения:

1) $d(a, b) = d(b, a)$;

2) $d(a, b) = 0$ тогда и только тогда, когда $a = b$.

3) *(Неравенство треугольника)* $d(a, c) \leq d(a, b) + d(b, c)$.

Исправление ошибок.

Теорема 1. Пусть d — расстояние Хэмминга кода C и $2r < d$. Тогда для любого слова $a \in \Omega^n$ существует не более одного слова $c \in C$, для которого выполнено неравенство $d(a, c) \leq r$.

□ Действительно, пусть $c_1, c_2 \in C$, $d(a, c_1) \leq r$ и $d(a, c_2) \leq r$. Тогда $d(c_1, c_2) \leq d(c_1, a) + d(a, c_2) \leq 2r < d$, откуда $c_1 = c_2$. □

Принцип максимального правдоподобия. Если известно, что переданное слово a принадлежит коду C, а полученное слово b ему не принадлежит, то слово b заменяется на такое слово $c \in C$, что $d(b, c)$ минимально. Если при передаче искажено менее чем $d/2$ знаков, то $c = a$.

Граница Синглтона.

Теорема 2. Если C — код длины n размерности k с расстоянием d, то $d + k \leq n + 1$.

□ Пусть $m = |C| = q^k$. Выпишем все слова кода C в виде матрицы и разделим ее столбцы

$$\begin{array}{c|c} (a_{1,1} \dots a_{1,d-1} & a_{1,d} \dots a_{1,n}) \\ \dots & \dots \\ (a_{i,1} \dots a_{i,d-1} & a_{i,d} \dots a_{i,n}) \\ \dots & \dots \\ (a_{j,1} \dots a_{j,d-1} & a_{j,d} \dots a_{j,n}) \\ \dots & \dots \\ (\underbrace{a_{m,1} \dots a_{m,d-1}}_{d-1} & \underbrace{a_{m,d} \dots a_{m,n}}_{n-(d-1)}) \end{array} \qquad \begin{array}{c} (a_{i,d} \dots a_{i,n}) = (a_{j,d} \dots a_{j,n}) \\ \Downarrow \\ i = j \end{array}$$

Следовательно, $m = q^k \leq q^{n-(d-1)} \Rightarrow k \leq n - d + 1$. □

МДР-коды.

Определение 1. Код C длины n размерности k с расстоянием d называется *МДР-кодом* (или, полностью, *кодом с максимальным достижимым расстоянием*), если

$$d + k = n + 1.$$

Простейшие примеры: 1) *Код констант* $C = \{(c, c, \dots, c) : c \in \Omega\}$ — $[n, 1, n]_q$-код. 2) Код Ω^n — $[n, n, 1]_q$-код. 3) *Код проверки на чётность*, состоящий из слов $(a_1, \dots, a_n)$ над алфавитом $\{0, 1, \dots, q-1\}$, для которых

$$a_1 + \dots + a_n \equiv 0 \pmod{q} \quad —$$

$[n, n-1, 2]_q$-код.

Параметры МДР-кодов.

Нерешенная задача.
Для заданных значений q и k найти максимальную возможную длину $n(k, q)$ МДР-кода размерности k над алфавитом из q элементов.

Простые оценки $n(k, q)$**.**
1) $n(1, q) = \infty$ (код констант имеет любую заданную длину). 2) $n(k, q) \geq k + 1$ (код проверки на чётность). 3) Если $q \leq k$, то $n(k, q) = k + 1$. 4) Если $k < q$, то $n(k, q) \leq q + k - 1$. 5) Если $q = q_1 q_2$, то

$$n(k, q) \geq \min\{n(k, q_1), n(k, q_2)\}.$$

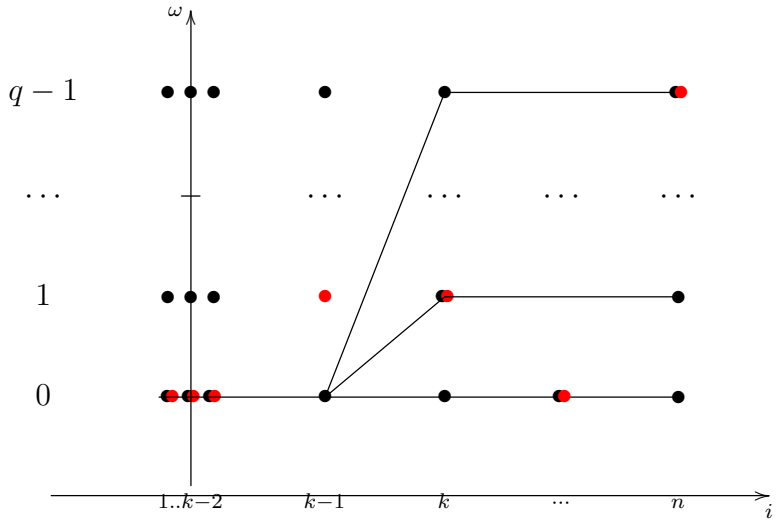
Пример доказательства оценки.

Докажем утверждение 4. Пусть C — МДР-код длины n и размерности k над Ω.

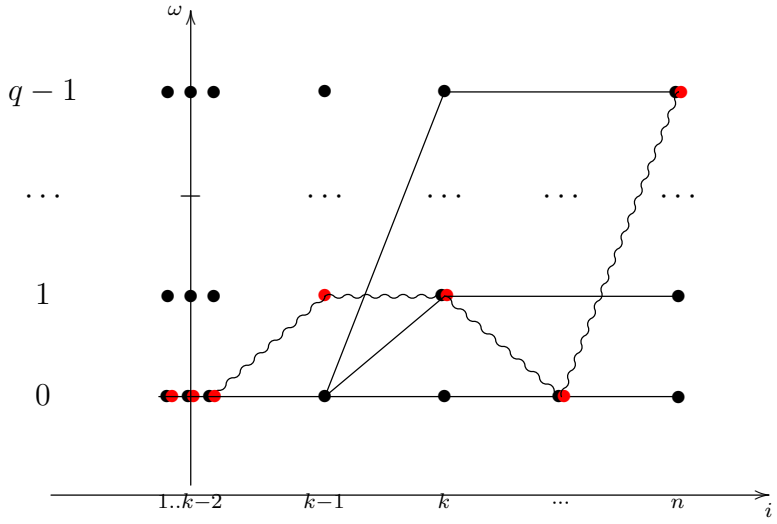
Будем считать, что $\Omega = \{0, 1, ..., q-1\}$ и каждое кодовое слово $(a_1, \ldots, a_n)$ представим его графиком, т.е. линией, соединяющей точки (i, a_i).

Любые две линии, представляющие различные кодовые слова, имеют не более чем $(k-1)$ общих точек.

Рассмотрим все слова, имеющие общее начало $(a_1, \ldots, a_{k-1})$.



Рассмотрим кодовое слово, которое имеет начало $(a_1, \ldots, a_{k-2}, a'_{k-1})$, где $a'_{k-1} \neq a_{k-1}$. Его график имеет не более одной точки пересечения с продолжениями рассмотренных слов.



Из рисунка видно, что $n-(k-1) \leq q$, что и требовалось.

Линейные коды.

Пусть теперь $\Omega = F$ — *конечное поле*, $q = |F| = p^r$, p — простое число.

Код C над F называется *линейным*, если C — подпространство линейного пространства строк над полем F.

Примеры кодов, приведенные ранее, — линейные коды.

Замечание 1. Если C — линейный код, то $d(a,b) = d(a-b,0)$ для любых $a, b \in C$, поэтому $d(C) = \min\{d(c,0) : 0 \neq c \in C\}$. Число $d(c,0)$ называется *весом* слова c.

Определение 2. Пусть $k \leq q-1$. Выберем $n \geq k$ различных элементов $\alpha_1, \ldots, \alpha_n$ поля F и n обратимых элементов $u_1, \ldots, u_n$ поля F. Множество слов $\{(u_1 f(\alpha_1), \ldots, u_n f(\alpha_n)) : f(x) \in F[x] \text{ и } \deg f(x) < k\}$ называется *обобщённым кодом Рида–Соломона* $GRS(k,n)$.

Обобщённый код Рида–Соломона — МДР-код.

Теорема 3. Код $GRS(k,n)$ имеет размерность k и расстояние $n-k+1$, т.е. является МДР-кодом.

□ Пусть $0 \neq f(x) \in F[x]$ и $\deg f(x) < k$. Тогда число корней многочлена $f(x)$ в F не превосходит $\deg f(x)$, т.е. меньше k. Значит, вес слова $\{(u_1 f(\alpha_1), \ldots, u_n f(\alpha_n))$ больше $n-k$, значит $d(GRS(k,n)) \geq n-k+1$.

Приведённое рассуждение также показывает, что число слов в коде $GRS(k,n)$ равно числу многочленов $f(x) \in F[x]$, таких, что $\deg f(x) < k$, т.е. q^k, т.е. k — размерность кода $GRS(k,n)$. Граница Синглтона дает $d(C) \leq n+1-k$. Следовательно, $d(C) = n+1-k$. □

Проверочная матрица линейного кода.

Из курса линейной алгебры известно, что любое подпространство линейного пространства над полем задается системой однородных линейных уравнений. Записывая ее в матричном виде, получаем следующее утверждение:

Теорема 4. Если C — линейный код длины n и размерности k над полем F, то существует матрица H размера $m \times n$, такая, что

$$C = \{a \in F^n : Ha^T = 0\}.$$

При этом $\operatorname{rank}(H) = n-k$.

Матрица H называется *проверочной матрицей* кода C.

Декодирование по таблице синдромов

Пусть C — линейный код длины n и размерности k над полем F, имеющий расстояние d. H — его проверочная матрица размера $(n-k)\times k$.

Синдромом вектора $a \in F^n$ называется вектор $aH^T = (Ha^T)^T$. Синдром — это вектор длины $n-k$.

Один раз можно составить таблицу: каждому возможному синдрому b сопоставляется вектор $e(b)$, являющийся решением системы $Hx^T = b^T$ и имеющий наименьший вес среди всех таких решений.

Пусть послано слово $a \in C$, получено слово a'. Если синдром $a'H^T = 0$, то, предполагая, что число ошибок при передаче меньше, чем d, получаем $a' = a$ *(обнаружение $< d$ ошибок)*.

Если же $b = c'H^T \neq 0$, то находим по таблице $e(b)$. Тогда $(c' - e(b))H^T = b - b = 0 \Rightarrow c = c' - e(b) \in C$. Предполагая, что число ошибок при передаче меньше, чем $d/2$, получаем $c = a$.

Список литературы

[1] Ф. Дж. МакВильямс, Н. Дж. А. Слоэн, Теория кодов, исправляющих ошибки. Москва, “Связь”, 1979.

[2] Э. Р. Берлекамп, Алгебраическая теория кодирования, Москва, “Мир”, 1971.