Конспект лекций по теории кодирования

Александр Дайняк

Актуальная версия файла доступна на www.dainiak.com

Моим родителям, сумевшим передать мне то,
что систематическому кодированию не поддаётся

Оглавление

Введение 5
1. Алфавитное кодирование 5
1.1. Математическая постановка, однозначность кодов 5
1.1.1. Критерий однозначности алфавитного кодирования 6
1.1.2. Оценка длины неоднозначно декодируемых слов 7
1.2. Коды с минимальной избыточностью 7
1.2.1. Свойства оптимальных кодов 9
2. Коды, исправляющие ошибки 10
2.1. Модель канала с ошибками 10
2.2. Основные определения и обозначения 10
2.3. Границы мощностей кодов 12
2.3.1. Граница Хемминга (граница сферической упаковки) 12
2.3.2. Граница Синглтона 12
2.3.3. Граница Плоткина 13
2.3.4. Граница Элайеса—Бассалыго 13
3. Линейные коды 15
3.1. Основные понятия 15
3.2. Теорема Варшамова—Гилберта 17
3.3. Двоичный код Хемминга 18
3.4. Границы мощностей для линейных кодов 19
3.4.1. Граница Синглтона для линейных кодов 19
3.4.2. Граница Грайсмера—Соломона—Штиффлера 19
3.5. Графы-расширители и коды на их основе 20
3.5.1. Графы-расширители 20
3.5.2. Коды на основе расширителей 21
3.5.3. Алгоритм Сипсера—Шпильмана 22
4. Коды Рида—Соломона и Рида—Маллера 23
4.1. Коды Рида—Соломона (I.S. Reed, G. Solomon) 23
4.1.1. Определение 23
4.1.2. Декодирование RS-кодов 24
4.2. Коды Рида—Маллера (I.S. Reed, D.E. Muller) 25
4.2.1. Определение 25
4.2.2. Кодовое расстояние 25
4.2.3. Декодирование RM-кодов 26
4.3. Понятие об алгеброгеометрических кодах (кодах В.Д. Гоппы) 28

5. Циклические коды 28
5.1. Определение 28
5.2. Порождающий многочлен 29
5.3. Порождающая и проверочная матрицы 30
5.4. Граница БЧХ и коды БЧХ 32
5.4.1. Граница БЧХ 32
5.4.2. Коды БЧХ 33
5.5. Циклическое представление кодов Хемминга 34
5.6. Восстановление синхронизации для смежных классов циклических кодов 34
5.6.1. Задача восстановления синхронизации 34
6. Совершенные коды 35
6.1. Совершенство кодов Голея и Хемминга 36
6.2. Некоторые теоремы о совершенных кодах 36
6.2.1. Доказательство теоремы Васильева при $q = 2$ 37
7. Теоремы Шеннона о скорости кодирования 37
7.1. Шенноновская ёмкость графов 37
7.2. Теоремы Шеннона для каналов с фиксированной вероятностью ошибок 40
7.2.1. Теоремы Шеннона 40
7.2.2. Доказательство теоремы о существовании хороших кодов 41
8. Некоторые специальные семейства кодов 42
8.1. Коды Варшамова—Тененгольца 42
8.1.1. Исправление одной ошибки выпадения в кодах Варшамова—Тененгольца 43
8.1.2. Исправление одной ошибки вставки в кодах Варшамова—Тененгольца 43
8.2. Матрицы Адамара и коды Адамара 44
8.2.1. Матрицы Адамара 44
8.2.2. Конструкция Пэли на основе квадратичных вычетов 45
8.2.3. Коды Адамара 46
8.3. Каскадные коды 47
8.3.1. Определение 47
8.3.2. Асимптотически хорошие коды с полиномиальным декодированием 47
9. О приложениях теории кодирования в информатике 48
9.1. Коммуникационная сложность 48
9.2. Криптография с открытым ключом: криптосистема МакЭлиса (R. McEliece ’1978) 49
9.3. l-однородные множества и дерандомизация 49
9.4. Задача о разделении секрета 51
Литература 52

Введение

Теория кодирования изучает модели хранения и передачи «дискретной» информации и предлагает способы оптимального её кодирования. Основные требования, которые предъявляются к способам кодирования, перечислены ниже.

- **Однозначность.** По закодированному сообщению нужно уметь однозначно восстанавливать исходное. Это требование обязательно.
- **Минимальная избыточность.** Закодированное сообщение должно при прочих равных условиях иметь как можно меньший объём для скорейшей передачи по каналам связи.
- **Устойчивость к ошибкам.** Возможность расшифровать закодированное сообщение даже при возникновении ошибок при его передаче.

Теории кодов, исправляющих ошибки, будет посвящено основное внимание, однако начнём мы с рассмотрения первых двух пунктов.

1. Алфавитное кодирование

Чтобы хранить в компьютере тексты на естественном языке, нужно их кодировать. Простейший подход состоит в следующем: каждой букве языка и знаку препинания сопоставим по двоичному слову, и тогда текст закодируем, записав друг за другом коды отдельных букв.

1.1. Математическая постановка, однозначность кодов

Даны алфавиты $\mathbb{A} = \{a_1, \dots, a_n\}$ и $\mathbb{B} = \{b_1, \dots, b_q\}$.

Алфавит $\mathbb{A}$ — *кодируемый*, «естественный»; алфавит $\mathbb{B}$ — *кодовый* (например, $\mathbb{B} = \{0,1\}$). *Алфавитное кодирование* — это отображение ϕ: $\mathbb{A}^* \to \mathbb{B}^*$, такое, что для любых $a_{i_1}, \dots, a_{i_r}$ выполнено $\phi(a_{i_1} \dots a_{i_r}) = \phi(a_{i_1}) \dots \phi(a_{i_r})$.

Видно, что достаточно определить ϕ на отдельных символах алфавита $\mathbb{A}$:

$$\begin{gathered}\phi(a_1) = B_1 \\ \vdots \\ \phi(a_n) = B_n\end{gathered}$$

Слова $B_1, \dots, B_n$ называются *кодовыми*, совокупность $\{B_1, \dots, B_n\}$ называется *кодом*.

Кодирование ϕ называется *однозначным*, если $\phi(w') \neq \phi(w'')$ при $w' \neq w''$. В противном случае кодирование почти бесполезно. Для однозначности необходимо, чтобы все B_i были различны (и везде далее мы будем это предполагать), но этого недостаточно. Однозначность никак не зависит от алфавита $\mathbb{A}$, а целиком определяется набором $\{B_1, \dots, B_n\}$.Кодирование однозначное т. и т.т., когда никакое слово $b_{i_1} b_{i_2} \dots b_{i_r}$ нельзя двумя разными способами разбить на кодовые слова:

Каждое из перечисленных условий является достаточным для однозначности кода:

- Равномерность: $|B_1| = |B_2| = \dots = |B_n|$.
- Свойство префикса: $\nexists i, j \left(i \neq j \text{ и } B_i = B_j w, \text{ где } w \in \mathbb{B}^*\right)$.
- Свойство суффикса: $\nexists i, j \left(i \neq j \text{ и } B_i = w B_j, \text{ где } w \in \mathbb{B}^*\right)$.

Префиксные коды называют ещё *мгновенными*, так как закодированные с их помощью сообщения можно декодировать по мере приёма, без задержек.

Пример того, что равномерность, префиксность и суффиксность не являются необходимыми условиями для однозначности:

$$\mathbb{A} = \{a_1, a_2\},$$
$$\mathbb{B} = \{0,1\},$$
$$\phi(a_1) = 0,$$
$$\phi(a_2) = 010.$$

1.1.1. Критерий однозначности алфавитного кодирования

В этом разделе мы выведем критерий, позволяющий по любому заданному кодированию (или, что то же, набору кодовых слов) определять, однозначно ли оно. Для этого подробнее рассмотрим свойства неоднозначных кодов.

Код неоднозначен, если найдётся слово $B \in \mathbb{B}^*$, которое не менее чем двумя разными способами можно разбить на кодовые слова. Рассмотрим самое короткое такое «неоднозначное» B и два его различных разбиения на кодовые слова:

Заметим, что точки «верхнего» и «нижнего» разбиений, кроме крайних, все различны, иначе слово B можно было бы укоротить:

Также, среди отрезков B, концы которых принадлежат разным разбиениям, нет кодовых слов, иначе B также можно было бы укоротить:

Минимальные отрезки слова B, концы которых принадлежат разным разбиениям, назовём *промежуточными*.

Первый промежуточный отрезок (п.о.) получается, если из начала некоторого кодового слова «отнять» некоторую последовательность кодовых слов. Любой из остальных отрезков получается, если из некоторого кодового слова отнять предыдущий отрезок и последовательность (возможно, пустую) кодовых слов. Последний п.о. таков, что если его отнять из начала некоторого кодового слова, получится последовательность кодовых слов.

Обозначим через $w_1, \dots, w_k$ все промежуточные отрезки. Через β будем обозначать последовательность (возможно, пустую) кодовых слов. Имеем:

$$\exists i, \beta(B_i = \beta w_1)$$
$$\exists i, \beta(B_i = w_1 \beta w_2)$$
$$\vdots$$
$$\exists i, \beta(B_i = w_{k-1} \beta w_k)$$
$$\exists i, \beta(B_i = w_k \beta)$$

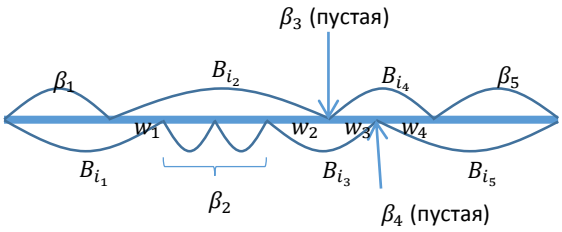
Наоборот, пусть нашлись непустые слова $w_1, \dots, w_k \in \mathbb{B}^*$, кодовые слова $B_{i_1}, \dots, B_{i_{k+1}}$ и последовательности кодовых слов $\beta_1, \dots, \beta_{k+1}$, такие, что выполнены соотношения

$$B_{i_1} = \beta_1 w_1$$
$$B_{i_2} = w_1 \beta_2 w_2$$
$$\vdots$$
$$B_{i_{k+1}} = w_k \beta_{k+1}$$

Тогда слово $\beta_1 w_1 \beta_2 w_2 \dots w_{k-1}\beta_k w_k \beta_{k+1}$ можно разбить на кодовые слова двумя способами.

$$B_{i_1} = \varepsilon\beta_1 w_1 B_{i_2} = w_1\beta_2 w_2$$
$$\vdots$$
$$B_{i_{k+1}} = w_k \beta_{k+1}\varepsilon$$

Пример:



Сказанное выше позволяет нам теперь сформулировать следующее утверждение.

Критерий однозначности алфавитного кодирования (Ал. А. Марков ’1963)
Код $C = \{B_1, \dots, B_n\}$ *не* однозначный т. и т.т., когда найдутся непустые слова $w_1, \dots, w_k \in \mathbb{B}^* \setminus C$, кодовые слова $B_{i_1}, \dots, B_{i_{k+1}}$ и последовательности кодовых слов $\beta_1, \dots, \beta_{k+1}$, такие, что $k \geq 1$ и выполнены соотношения

$$B_{i_1} = \beta_1 w_1$$
$$B_{i_2} = w_1\beta_2 w_2$$
$$\vdots$$
$$B_{i_{k+1}} = w_k \beta_{k+1}$$

(или $k = 0$ и $B_{i_1} = \beta_1$, где β_1 составлено не менее чем из двух кодовых слов).

1.1.2. Оценка длины неоднозначно декодируемых слов

Через ε будем обозначать пустое слово. Пусть $C = \{B_1, \dots, B_n\}$ — код, который нужно проверить на однозначность. Построим орграф $G_C = (V, E)$, где
$V \coloneqq \{\varepsilon$, а также все слова из $\mathbb{B}^* \setminus C$, являющиеся началами и концами кодовых слов$\}$,
$E \coloneqq \left\{(\alpha', \alpha'') \mid \exists \beta \in C^* \left((\alpha'\beta\alpha'' \in C) \wedge \left((\beta \neq \varepsilon) \vee \left((\alpha' \neq \varepsilon) \wedge (\alpha'' \neq \varepsilon)\right)\right)\right)\right\}$.

Доказанный нами критерий однозначности можно переформулировать так: код C однозначный т. и т.т., когда в орграфе G_C нет (ориентированного) цикла, проходящего через вершину ε. Имеем

$$|V| \leq 1 + \sum_{B \in C}(|B| - 1) \leq |C| \cdot \max_{B \in C}|B|.$$

Получим отсюда оценку длины минимального неоднозначно декодируемого слова. Если в G_C есть цикл через ε, то есть и цикл, число вершин в котором не больше, чем $|C| \cdot \max_{B \in C}|B|$. Рассмотрим соответствующее этому циклу неоднозначно декодируемое слово $W_{\text{неодн.}} = \beta_1 w_1 \beta_2 w_2 \dots w_{k-1}\beta_k w_k \beta_{k+1}$. Каждая пара $\beta_i w_i$ умещается в некотором кодовом слове, поэтому

$$\left|W_{\text{неодн.}}\right| \leq (k+1) \cdot \max_{B \in C}|B| \leq |C| \cdot \left(\max_{B \in C}|B|\right)^2.$$

Из предыдущих рассуждений вытекает следующая оценка на длину неоднозначно декодируемого слова.

Теорема. (А.А. Марков)
Если C — неоднозначный код, длина слов которого не превосходит l, то найдётся слово длины не более $|C| \cdot l^2$, декодируемое неоднозначно.

1.2. Коды с минимальной избыточностью

Обычно, кодируемые символы $a_1, \dots, a_n$ встречаются в кодируемых сообщениях не одинаково часто, а с разными частотами. Например, в английском языке буква *e* встречается примерно в 180 раз чаще,

чем z. Естественно при построении кодирования ϕ кодировать более частые буквы более короткими словами.

Поставим задачу математически. Пусть в кодируемых сообщениях символы $a_1, \dots, a_n$ встречаются с частотами $p_1, \dots, p_n$ соответственно. Считаем, что $\sum p_i = 1$ и $\forall i\ p_i > 0$.

Пусть символ a_i кодируется словом B_i. Рассмотрим сообщение $A \in \mathbb{A}^*$. Каждый из символов a_i встретится в $|A|$ примерно $|A| \cdot p_i$ раз. Отсюда

$$|\phi(A)| \approx \sum_i |A| \cdot p_i \cdot |B_i| = |A| \cdot \sum_i p_i \cdot |B_i|.$$

То есть «среднестатистическое» сообщение A при кодировании «разбухает» примерно в $\sum_i p_i|B_i|$ раз. Величина $\sum_i p_i|B_i|$ называется *коэффициентом избыточности (к.и.) кода*.

Задача построения кода с минимальной избыточностью: по заданным $p_1, \dots, p_n$ построить однозначно декодируемый код $B_1, \dots, B_n \in \mathbb{B}^*$, для которого коэффициент избыточности минимален.

Такой код называется *кодом с минимальной избыточностью (к.м.и.) для набора частот* $p_1, \dots, p_n$.

Все слова кода не получится взять слишком короткими, иначе код не будет однозначным. Количественно это выражает следующая теорема.

Теорема. (L.G. Kraft, B. McMillan)
Пусть $l_1, \dots, l_n$ — длины слов однозначного кода в алфавите $\mathbb{B}$, где $|\mathbb{B}| = q$. Тогда выполнено неравенство

$$\sum_{i=1}^{n} q^{-l_i} \le 1.$$

Доказательство. Пусть $B_1, \dots, B_n$ — однозначный код в q-значном алфавите, и пусть $|B_i| = l_i$. Пусть $t \in \mathbb{N}$. Положим $L \coloneqq t \cdot \max\limits_i l_i$. Рассмотрим выражение

$$\left(\sum_{i=1}^{n} q^{-l_i}\right)^t = \sum_{1 \le i_1, \dots, i_t \le n} q^{-(l_{i_1} + \dots + l_{i_t})} = \sum_{l=1}^{L} s_l q^{-l},$$

где s_l — количество наборов $(i_1, \dots, i_t)$, таких, что $l_{i_1} + \dots + l_{i_t} = l$.

Каждому набору $(i_1, \dots, i_t) \in S_l$ поставим в соответствие слово $B_{i_1} \dots B_{i_t} \in \mathbb{B}^*$. Тогда разным наборам из S_l соответствуют разные слова (т.к. код однозначный). Отсюда $s_l \le q^l$ и следовательно

$$\left(\sum_{i=1}^{n} q^{-l_i}\right)^t = \sum_{l=1}^{L} s_l q^{-l} \le \sum_{l=1}^{L} 1 = L.$$

Получили, что для любого $t \in \mathbb{N}$ выполнено $\sum_{i=1}^{n} q^{-l_i} \le \left(t \cdot \max\limits_i l_i\right)^{1/t}$. Устремляя t к бесконечности, получаем $\sum_{i=1}^{n} q^{-l_i} \le 1$.

Теорема доказана.

Докажем обратное утверждение:
Теорема.
Пусть натуральные числа $l_1, \dots, l_n$ и q таковы, что $\sum_{i=1}^{n} q^{-l_i} \le 1$. Тогда существует *префиксный* код $B_1, \dots, B_n$ в q-значном алфавите, такой, что $|B_i| = l_i$.

Доказательство. Будем считать, что среди $l_1, \dots, l_n$ всего m различных, и при этом $l_1 < \dots < l_m$. Для каждого $j \in [1, m]$ положим $n_j \coloneqq \left|\{i \in [1, n] \mid l_i = l_j\}\right|$. Тогда из условия теоремы следует неравенство $\sum_{j=1}^{m} n_j q^{-l_j} \le 1$. Отсюда

$$\sum_{j=1}^{k} n_j q^{-l_j} \le 1$$

для любого $k \in [1, m]$. Домножив обе части последнего неравенства на q^{l_k}, получим

$$q^{l_k} \ge \sum_{j=1}^{k} n_j q^{l_k - l_j} = n_k + \sum_{j=1}^{k-1} n_j q^{l_k - l_j}.$$

Следовательно, для любого $k \in [1, m]$ имеем $n_k \le q^{l_k} - \sum_{j=1}^{k-1} n_j q^{l_k - l_j}$. Будем строить префиксный код, сначала выбирая n_1 слов длины l_1, затем n_2 слов длины l_2, и т.д. Пусть уже набраны все кодовые слова с длинами $l_1, \dots, l_{k-1}$. Слов длины l_k, для которых выбранные кодовые слова являются префиксами, не более $n_1 q^{l_k - l_1} + \cdots + n_{k-1} q^{l_k - l_{k-1}}$, то есть «пригодных для выбора» слов длины l_k не меньше, чем $q^{l_k} - (n_1 q^{l_k - l_1} + \cdots + n_{k-1} q^{l_k - l_{k-1}})$. Из неравенства $n_k \le q^{l_k} - \sum_{j=1}^{k-1} n_j q^{l_k - l_j}$ вытекает, что мы сможем выбрать n_k слов длины l_k, так, чтобы никакие из ранее выбранных слов не были их префиксами. По индукции получаем утверждение теоремы.

Теорема доказана.

Следствие из двух доказанных теорем.
Для любого однозначного кода существует префиксный код *в том же алфавите и с теми же длинами кодовых слов*. Значит, *к.м.и. можно искать только среди префиксных кодов*.

1.2.1. Свойства оптимальных кодов

Лемма о монотонности длин слов к.м.и.
Если $B_1, \dots, B_n$ — к.м.и. для набора частот $p_1, \dots, p_n$, то

$$\forall i, j \; (p_i > p_j \quad \Rightarrow \quad |B_i| \le |B_j|).$$

Доказательство. В противном случае, поменяв B_i и B_j местами, получили бы код с коэффициентом избыточности

$$\sum_{i=1}^{n} p_i |B_i| - (p_i - p_j)(|B_i| - |B_j|) < \sum_{i=1}^{n} p_i |B_i|.$$

Лемма доказана.

Теорема «о редукции». (D.A. Huffman)
Пусть $p_1 \ge \cdots \ge p_{n-1} \ge p_n$ и $p := p_{n-1} + p_n$. Если $B_1, \dots, B_{n-2}, B \in \{0,1\}^*$ — префиксный к.м.и. для частот $p_1, \dots, p_{n-2}, p$, то $B_1, \dots, B_{n-2}, B0, B1$ — префиксный к.м.и. для частот $p_1, \dots, p_n$.

Доказательство. Пусть к.и. кода $B_1, \dots, B_{n-2}, B$ для частот $p_1, \dots, p_{n-2}, p$ равен k. К.и. кода $B_1, \dots, B_{n-2}, B0, B1$ для частот $p_1, \dots, p_n$ равен

$$\sum_{i=1}^{n-2} p_i |B_i| + (p_{n-1} + p_n)(|B| + 1) = \sum_{i=1}^{n-2} p_i |B_i| + p|B| + p = k + p.$$

Допустим, что нашёлся код $B'_1, \dots, B'_n$, к.и. которого для набора частот $p_1, \dots, p_n$ равен $k' < k + p$, и придём к противоречию. Б.о.о. будем считать код $\{B'_i\}_{i=1}^n$ префиксным к.м.и. для набора $p_1, \dots, p_n$. Т.к. $p_1 \ge \cdots \ge p_n$, то $|B'_1| \le \cdots \le |B'_n|$.

Пусть $B'_n = B'0$, где B' — некоторое слово. Заметим, что $B' \notin \{B'_i\}_{i=1}^n$ и B' является префиксом одного из слов $B'_1, \dots, B'_{n-1}$. Б.о.о. будем считать, что $B'_{n-1} = B'1$. Тогда код $B'_1, \dots, B'_{n-2}, B'$ префиксный. К.и. кода $B'_1, \dots, B'_{n-2}, B'$ для набора частот $p_1, \dots, p_{n-2}, p$ равен

$$(p_{n-1} + p_n)|B'| + \sum_{i=1}^{n-2} p_i |B'_i| = \sum_{i=1}^{n} p_i |B'_i| - (p_{n-1} + p_n) = k' - p < k,$$

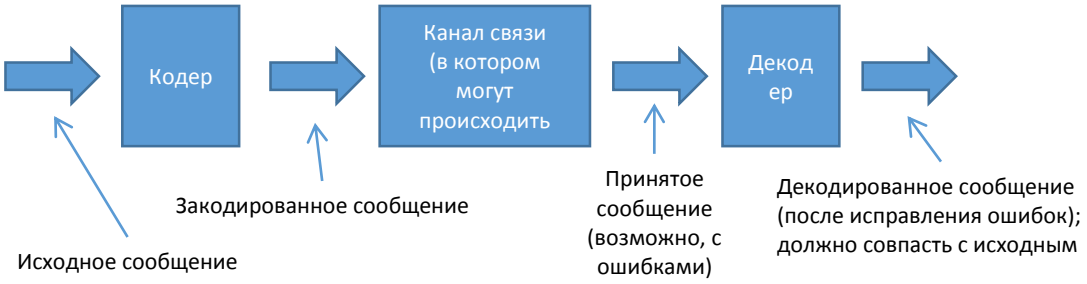
— противоречие с тем, что код $B_1, \dots, B_{n-2}, B$ является к.м.и. для частот $p_1, \dots, p_{n-2}, p$.

Теорема доказана.

2. Коды, исправляющие ошибки

2.1. Модель канала с ошибками

Основная модель канала связи:



Чаще всего рассматривают следующие **типы ошибок:**

- Ошибки замещения: *муха* → *мука*
  - Симметричные (если символ x может замениться на y, то возможно и обратное)
  - Несимметричные
- Ошибки стирания: *муха* → *му?а*
- Ошибки выпадения: *муха* → *уха*
- Ошибки вставки: *мука* → *мурка*
- Комбинации перечисленных типов

Всегда задаются *ограничения на «ненадёжность» канала*. Ограничения можно разделить на два типа:

- вероятностные (например, вероятность возникновения ошибки в одном символе сообщения),
- детерминированные (например, верхняя оценка числа ошибок на одно сообщение).

Естественный язык весьма устойчив к ошибкам: «веть ву мжте прчтттть эт ткст п поняц го!». Причины этой устойчивости: избыточность (например, наличие «малоинформативных» гласных) и разреженность («вблизи» слов обычно нет других слов). Если эти свойства где-то нарушаются, то ошибки исправлять тяжело: сравните *чемодан зарыт* и *чемодан закрыт.*

2.2. Основные определения и обозначения

Пусть $\mathbb{A}_q$ — кодовый алфавит, *алфавит канала*, и пусть $|\mathbb{A}_q| = q$. Будем называть *q-ичным кодом* любое подмножество $C \subseteq \mathbb{A}_q^n$, где n — *длина кода* (длина кодовых слов), $|C|$ — *мощность кода* (число кодовых слов).

Чаще всего рассматривают двоичные коды, т.е. когда $q = 2$ и $\mathbb{A}_q = \{0,1\}$. Таким же важным является случай, когда $\mathbb{A}_q$ является полем. В этих случаях для произвольного слова $\boldsymbol{a}$ будем через $\|\boldsymbol{a}\|$ обозначать *вес* этого слова, то есть величину $\#\{i \mid a_i \neq 0\}$.

Пусть $\boldsymbol{a}$ и $\boldsymbol{b}$ — слова в алфавите канала. Обозначим через $\tilde{d}(\boldsymbol{a}, \boldsymbol{b})$ минимальное число ошибок, в результате которых $\boldsymbol{a}$ может перейти в $\boldsymbol{b}$.

Способ кодирования позволяет *обнаруживать k ошибок*, если для любых различных кодовых сообщений $\boldsymbol{a}'$ и $\boldsymbol{a}''$ при передаче в канал $\boldsymbol{a}'$ на выходе из канала не может получиться $\boldsymbol{a}''$ (если в канале произошло не более k ошибок). Иначе говоря, $\tilde{d}(\boldsymbol{a}', \boldsymbol{a}'') > k$.

Способ кодирования позволяет *исправлять k ошибок*, если при передаче в канал различных кодовых сообщений $\boldsymbol{a}'$ и $\boldsymbol{a}''$ на выходе из канала будут получаться различные сообщения (при условии, что с каждым отдельным сообщением в канале происходит не более k ошибок). Формально:

$$\nexists \boldsymbol{a}', \boldsymbol{a}'' \in C, \boldsymbol{a}: \boldsymbol{a}' \neq \boldsymbol{a}'' \quad \wedge \quad \tilde{d}(\boldsymbol{a}', \boldsymbol{a}) \leq k \quad \wedge \quad \tilde{d}(\boldsymbol{a}'', \boldsymbol{a}) \leq k.$$

Особенно удобно, когда $\tilde{d}$ является *метрикой*:

- $\forall \boldsymbol{a}, \boldsymbol{b}\ \tilde{d}(\boldsymbol{a}, \boldsymbol{b}) = \tilde{d}(\boldsymbol{b}, \boldsymbol{a})$ — симметричность
- $\forall \boldsymbol{a} \neq \boldsymbol{b}\ \tilde{d}(\boldsymbol{a}, \boldsymbol{b}) > 0$
- $\forall \boldsymbol{a}\ \tilde{d}(\boldsymbol{a}, \boldsymbol{a}) = 0$
- $\forall \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}\ \tilde{d}(\boldsymbol{a}, \boldsymbol{b}) \leq \tilde{d}(\boldsymbol{a}, \boldsymbol{c}) + \tilde{d}(\boldsymbol{c}, \boldsymbol{b})$ — неравенство треугольника

Так бывает не всегда. Например, если в канале могут происходить лишь ошибки вставки, то при $\boldsymbol{a} \neq \boldsymbol{b}$ по крайней мере одна из двух величин $\tilde{d}(\boldsymbol{a}, \boldsymbol{b}), \tilde{d}(\boldsymbol{b}, \boldsymbol{a})$ вовсе не определена.

Пусть $\tilde{d}(\dots, \dots)$ — метрика и C — код. Положим

$$\tilde{d}(C) \coloneqq \min_{\substack{\boldsymbol{a} \neq \boldsymbol{b} \\ \boldsymbol{a}, \boldsymbol{b} \in C}} \tilde{d}(\boldsymbol{a}, \boldsymbol{b}).$$

Величина $\tilde{d}(C)$ называется *кодовым расстоянием* кода C.

Утверждение (о связи кодового расстояния с устойчивостью к ошибкам).

- Код C обнаруживает t ошибок $\Leftrightarrow\ \tilde{d}(C) > t$.
- Код C исправляет t ошибок $\Leftarrow\ \tilde{d}(C) > 2t$. Для метрик Хемминга и Левенштейна можно «$\Leftarrow$» заменить на «$\Leftrightarrow$».

Докажем только первую часть.
$\Leftarrow$ Пусть в канал передавалось сообщение $\boldsymbol{a}$. Если в канале происходит не более t ошибок, то даже если при передаче возникли ошибки и получилось слово $\boldsymbol{a}' \neq \boldsymbol{a}$, то нам гарантирована оценка $\tilde{d}(\boldsymbol{a}, \boldsymbol{a}') \leq t$. Но тогда $\boldsymbol{a}' \notin C$, а значит, мы увидим, что при передаче произошли ошибки.
$\Rightarrow$ С другой стороны, если бы в коде была пара слов $\boldsymbol{a}, \boldsymbol{b}$, таких, что $\tilde{d}(\boldsymbol{a}, \boldsymbol{b}) \leq t$, то при неудачной ситуации, когда в канал передавалось слово $\boldsymbol{a}$, ошибки могли распределиться таким образом, что получилось бы в точности слово $\boldsymbol{b}$. Но тогда, приняв слово $\boldsymbol{b}$, мы наивно полагали бы, что «само слово $\boldsymbol{b}$ и передавалось изначально, причём при передаче не было ошибок».
Доказательство второй части утверждения — упражнение (придётся использовать и симметричность метрики, и неравенство треугольника).

Утверждение доказано.

Если рассматриваются слова одной и той же длины, а в канале возможны только ошибки типа замещения (любые), то $\tilde{d}(\boldsymbol{a}, \boldsymbol{b}) = d(\boldsymbol{a}, \boldsymbol{b})$, где

$$d(\boldsymbol{a}, \boldsymbol{b}) \coloneqq \#\{i \mid a_i \neq b_i\}.$$

Функционал d называется *метрикой Хемминга*, а величина $d(\boldsymbol{a}, \boldsymbol{b})$ — *расстоянием Хемминга между* $\boldsymbol{a}$ *и* $\boldsymbol{b}$.

Везде далее по умолчанию будем считать, что в канале возможны произвольные ошибки типа замещения, т.е. в качестве метрики на словах везде далее используется метрика Хемминга.
Обозначим

$$d(C) \coloneqq \min_{\substack{\boldsymbol{a} \neq \boldsymbol{b} \\ \boldsymbol{a}, \boldsymbol{b} \in C}} d(\boldsymbol{a}, \boldsymbol{b}).$$

Шар радиуса r с центром в $\boldsymbol{a}$ — это множество $S_r(\boldsymbol{a}) \coloneqq \{\boldsymbol{b} \mid d(\boldsymbol{a}, \boldsymbol{b}) \leq r\}$.

Если в канал передавалось $\boldsymbol{a}$, то на выходе из канала может быть любое слово $\boldsymbol{b} \in S_t(\boldsymbol{a})$. Значит, код обнаруживает t ошибок т. и т.т., когда никакое кодовое слово не попадает в шар радиуса t с центром в другом кодовом слове. Аналогично, код исправляет t ошибок т. и т.т., когда при передаче в канал различных кодовых слов на выходе получаются различные слова, то есть когда шары радиуса t с центрами в кодовых словах не пересекаются.

Основная задача теории кодов, исправляющих ошибки: строить коды, для которых число кодовых слов как можно больше, кодовое расстояние как можно больше, а длина кодовых слов как можно

меньше. Также можно рассматривать **задачи, связанные с ресурсами:** процессы кодирования и декодирования (исправление ошибок) должны быть возможно менее трудоёмкими по количеству операций и по памяти.

Геометрически основная задача — это **задача об упаковке** возможно большего числа шаров, возможно большего радиуса, в пространстве возможно меньшей размерности.

Обозначение кода с заданными параметрами
Если C — q-ичный код с длиной слов n, числом слов M и кодовым расстоянием d, то пишут: «C является $(n, M, d)_q$-кодом». Если код двоичный, то символ q не указывают.

2.3. Границы мощностей кодов

Полезно знать границы своих возможностей при построении кодов. С несколькими такими границами мы познакомимся в этом разделе.

2.3.1. Граница Хемминга (граница сферической упаковки)

Теорема. (R.W. Hamming)
Для любого $(n, M, d)_q$-кода имеем

$$M \leq \frac{q^n}{\left|S_{\lfloor (d-1)/2 \rfloor}(\mathbf{0})\right|}$$

В двоичном случае

$$M \leq \frac{2^n}{\sum_{k=0}^{\lfloor (d-1)/2 \rfloor} \binom{n}{k}}$$

Доказательство. Пусть $C = \{\boldsymbol{a}_1, \boldsymbol{a}_2, \dots, \boldsymbol{a}_M\}$ — $(n, M, d)_q$-код. Так как $d(C) = d$, то шары радиуса $\lfloor (d-1)/2 \rfloor$ с центрами в кодовых словах не пересекаются. Отсюда $q^n \geq \sum_{j=1}^{M} \left|S_{\lfloor (d-1)/2 \rfloor}(\boldsymbol{a}_j)\right| = M \cdot \left|S_{\lfloor (d-1)/2 \rfloor}(\mathbf{0})\right|$.

Теорема. (В некотором смысле, обратная границе Хемминга)
Пусть числа $q,\ n, M, d \in \mathbb{N}$ таковы, что $M \leq \frac{q^n}{|S_d(\mathbf{0})|}$. Тогда существует $(n, M, d)_q$-код.

Доказательство. Пусть $C = \{\boldsymbol{a}_1, \dots, \boldsymbol{a}_{|C|}\}$ — код максимальной мощности с кодовым расстоянием d и длиной слов n. Тогда шары радиуса d с центрами в кодовых словах покрывают целиком множество $\mathbb{A}_q^n$ (иначе код C можно было пополнить любым из слов, не лежащих ни в одном из этих шаров). Отсюда $\sum_{j=1}^{|C|} \left|S_d(\boldsymbol{a}_j)\right| \geq q^n$, следовательно, $|C| \geq M$.

Коды, достигающие границу Хемминга, называются *совершенными* или *плотно упакованными*.

2.3.2. Граница Синглтона

Теорема. (R.C. Singleton)
Для любого $(n, M, d)_q$-кода имеем $M \leq q^{n-d+1}$.

Доказательство. Пусть $C = \{\boldsymbol{a}_1, \dots, \boldsymbol{a}_M\}$ — $(n, M, d)_q$-код. Рассмотрим слова $\{\boldsymbol{a}_i'\}_{i=1}^{M}$, где $\boldsymbol{a}_i'$ получено из $\boldsymbol{a}_i$ отбрасыванием $(d-1)$ последних координат. Так как $d(\boldsymbol{a}_i, \boldsymbol{a}_j) \geq d$ для любых i, j, то все слова $\boldsymbol{a}_i'$ различны. Их количество не превосходит числа всех q-ичных слов длины $(n-d+1)$. Поэтому $M \leq q^{n-d+1}$.

Теорема доказана.

Коды, на которых достигается граница Синглтона, называются *MDS-кодами* (maximum distance separable codes).

2.3.3. Граница Плоткина

Теорема. (M. Plotkin)

Пусть $nr < d$, где $r \coloneqq 1 - \frac{1}{q}$. Тогда для любого $(n, M, d)_q$-кода имеем $M \leq \left\lfloor \frac{d}{d-nr} \right\rfloor$.

Доказательство. Рассмотрим матрицу, в которой по строкам выписаны все кодовые слова:

$$\begin{pmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_M \end{pmatrix}.$$

Элементы этой матрицы будем обозначать a_{ij}. Оценим снизу и сверху следующую сумму:

$$T \coloneqq \sum_{\substack{1 \leq k \leq n \\ 1 \leq i < j \leq M}} \mathbb{1}_{a_{ik} \neq a_{jk}}$$

(через $\mathbb{1}_{\dots}$ обозначен индикатор того, что условие … выполняется).

Имеем

$$T = \sum_{1 \leq i < j \leq M} \sum_{1 \leq k \leq n} \mathbb{1}_{a_{ik} \neq a_{jk}} = \sum_{1 \leq i < j \leq M} d(\boldsymbol{a}_i, \boldsymbol{a}_j) \geq \frac{M \cdot (M-1)}{2} \cdot d.$$

С другой стороны,

$$T = \sum_{1 \leq k \leq n} \sum_{1 \leq i < j \leq M} \mathbb{1}_{a_{ik} \neq a_{jk}}.$$

Зафиксируем произвольное k. Пусть среди кодовых слов ровно x_s слов имеют k-ю координату, равную s. Тогда

$$\sum_{1 \leq i < j \leq M} \mathbb{1}_{a_{ik} \neq a_{jk}} = \sum_{s' \neq s''} x_{s'} \cdot x_{s''}.$$

Имеем

$$\sum_{s' \neq s''} x_{s'} \cdot x_{s''} = \frac{1}{2} \cdot \left(\left(\sum_s x_s \right)^2 - \sum_s x_s^2 \right) = \frac{1}{2} \cdot \left(M^2 - \sum_s x_s^2 \right) \leq \frac{1}{2} \cdot \left(M^2 - q \cdot \frac{M^2}{q^2} \right) = \frac{M^2}{2} \left(1 - \frac{1}{q} \right).$$

(Выше мы учли, что минимум выражения $\sum_s x_s^2$ достигается, когда все x_s равны M/q, — это неравенство Коши—Буняковского.)

Из выведенных неравенств мы получаем оценку

$$T = \sum_{1 \leq k \leq n} \sum_{1 \leq i < j \leq M} \mathbb{1}_{a_{ik} \neq a_{jk}} \leq \frac{nM^2}{2} \left(1 - \frac{1}{q} \right).$$

Сопоставим верхнюю и нижнюю оценки для T:

$$\frac{M \cdot (M-1)}{2} \cdot d \leq T \leq \frac{nM^2}{2} \left(1 - \frac{1}{q} \right).$$

Отсюда $(M-1) \cdot d \leq nrM$, и, стало быть, $M(d - nr) \leq d$. Так как $d - nr > 0$ по условию и $M \in \mathbb{Z}$, то $M \leq \left\lfloor \frac{d}{d-rn} \right\rfloor$.

Теорема доказана.

2.3.4. Граница Элайеса—Бассалыго

2.3.4.1. Вложения метрических пространств

Метрическое пространство — это множество с заданной на нём метрикой.

Примеры:

- $\left(\{0,1\}^n, d(\boldsymbol{a}, \boldsymbol{b}) \right)$ — *метрическое пространство Хемминга* (здесь d — метрика Хемминга).

- $\left(\mathbb{R}^n,\ \tilde{d}(\boldsymbol{a},\boldsymbol{b})\right)$ — евклидово метрическое пространство (здесь $\tilde{d}(\boldsymbol{a},\boldsymbol{b}) \coloneqq \sqrt{\sum_i (a_i - b_i)^2}$ — обычная евклидова метрика).

Вложение метрического пространства U в метрическое пространство V — это отображение $\phi\colon U \to V$, сохраняющее метрику:

$$\operatorname{dist}_U(x,y) = \operatorname{dist}_V\big(\phi(x),\phi(y)\big).$$

Вложение n-мерного хеммингова пространства в евклидово n-мерное пространство при $n > 1$ сделать не получится, но можно выполнить отображение, сохраняющее определённую информацию о метрике. Сопоставим каждому вектору $\boldsymbol{a} \in \{0,1\}^n$ вектор $\boldsymbol{x^a} \in \mathbb{R}^n$ по правилу:

$$x_i^{\boldsymbol{a}} = \begin{cases} 1, & \text{если } a_i = 1 \\ -1, & \text{если } a_i = 0 \end{cases}.$$

При этом

- $\tilde{d}\big(\boldsymbol{x^a},\boldsymbol{x^b}\big) = 2\cdot\sqrt{d(\boldsymbol{a},\boldsymbol{b})}$,
- $\big\langle \boldsymbol{x^a},\boldsymbol{x^b}\big\rangle = n - 2\cdot d(\boldsymbol{a},\boldsymbol{b})$,
- $\|\boldsymbol{x^a}\| = \sqrt{n}$ (здесь $\|\cdot\|$ — евклидова норма).

2.3.4.2. Системы векторов в евклидовом пространстве

Лемма (о тупоугольной системе векторов).
Пусть $\boldsymbol{y},\boldsymbol{x}_1,\dots,\boldsymbol{x}_m \in \mathbb{R}^n$ таковы, что выполнено

- $\langle \boldsymbol{x}_i,\boldsymbol{y}\rangle > 0$ для $i = 1,\dots,m$,
- $\big\langle \boldsymbol{x}_i,\boldsymbol{x}_j\big\rangle \le 0$ при $i \ne j$.

Тогда $\boldsymbol{x}_1,\dots,\boldsymbol{x}_m$ линейно независимы и, в частности, $m \le n$.

Доказательство. Рассмотрим произвольную нулевую линейную комбинацию: $c_1\boldsymbol{x}_1 + \dots + c_m\boldsymbol{x}_m = \boldsymbol{0}$. Положим

$$\mathrm{Pos} \coloneqq \{\, i \mid c_i > 0 \,\},\quad \mathrm{Neg} \coloneqq \{\, i \mid c_i < 0 \,\}.$$

Нам нужно доказать, что $\mathrm{Pos} = \mathrm{Neg} = \emptyset$. Допустим, что это не так, и придём к противоречию. Пусть, например, $\mathrm{Pos} \ne \emptyset$ (быть может, при этом $\mathrm{Neg} = \emptyset$). Положим

$$\boldsymbol{z} \coloneqq \sum_{i\in\mathrm{Pos}} c_i\boldsymbol{x}_i = \sum_{j\in\mathrm{Neg}} \big(-c_j\big)\boldsymbol{x}_j.$$

Имеем

$$\langle \boldsymbol{z},\boldsymbol{y}\rangle = \sum_{i\in\mathrm{Pos}} c_i\langle \boldsymbol{x}_i,\boldsymbol{y}\rangle > 0.$$

Отсюда следует, что $\boldsymbol{z} \ne \boldsymbol{0}$. Имеем

$$\boldsymbol{z} \coloneqq \sum_{i\in\mathrm{Pos}} c_i\boldsymbol{x}_i = \sum_{j\in\mathrm{Neg}} \big(-c_j\big)\boldsymbol{x}_j \ne \boldsymbol{0}.$$

Рассмотрим теперь соотношения

$$0 < \langle \boldsymbol{z},\boldsymbol{z}\rangle = \left\langle \sum_{i\in\mathrm{Pos}} c_i\boldsymbol{x}_i\,,\ \sum_{j\in\mathrm{Neg}} \big(-c_j\big)\boldsymbol{x}_j \right\rangle = \sum_{\substack{i\in\mathrm{Pos}\\ j\in\mathrm{Neg}}} c_i\big(-c_j\big)\cdot\big\langle \boldsymbol{x}_i,\boldsymbol{x}_j\big\rangle \le 0$$

— противоречие!

Лемма доказана.

2.3.4.3. Доказательство теоремы Элайеса—Бассалыго

Теорема. (P. Elias, Л.А. Бассалыго)
Для любого $(n, M, d)_2$-кода, где $d \le n/2$, выполнено неравенство

$$M \le \frac{n2^n}{|S_{\lfloor \tau n - 1 \rfloor}|},$$

где $\tau = \frac{1-\sqrt{1-2\delta}}{2}$, $\delta = \frac{d}{n}$. Через $S_{\lfloor \tau n-1 \rfloor}$ мы сокращённо обозначаем шар $S_{\lfloor \tau n-1 \rfloor}(\mathbf{0})$.

Доказательство. Положим $\delta := \frac{d}{n}$, $\tau := \frac{1-\sqrt{1-2\delta}}{2}$ и $t := \lfloor \tau n - 1 \rfloor$.

Пусть C — (n, M, d)-код. Положим $\deg_t C := \max\limits_{\boldsymbol{b}\in\{0,1\}^n} |C \cap S_t(\boldsymbol{b})|$. Имеем

$$|C| \cdot |S_t| = \sum_{\boldsymbol{a}\in C} \sum_{\boldsymbol{b}\in\{0,1\}^n} \mathbb{1}_{d(\boldsymbol{a},\boldsymbol{b})\le t} = \sum_{\boldsymbol{b}\in\{0,1\}^n} \sum_{\boldsymbol{a}\in C} \mathbb{1}_{d(\boldsymbol{a},\boldsymbol{b})\le t} \le 2^n \cdot \deg_t C.$$

Отсюда $M \le \frac{2^n \cdot \deg_t C}{|S_t|}$. Осталось доказать, что при выбранном t выполнено неравенство $\deg_t C \le n$.
Пусть $\boldsymbol{b} \in \{0,1\}^n$, и $\boldsymbol{a}_1, \dots, \boldsymbol{a}_m \in C \cap S_t(\boldsymbol{b})$ ($\boldsymbol{a}_i \ne \boldsymbol{a}_j$ при $i \ne j$). Нам нужно доказать, что $m \le n$.

Сопоставим словам $\boldsymbol{b}, \boldsymbol{a}_1, \dots, \boldsymbol{a}_m$ векторы $\boldsymbol{y}, \boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$ так (на примере $\boldsymbol{b}$):

$$y_i = \begin{cases} 1/\sqrt{n}, & \text{если } b_i = 1, \\ -1/\sqrt{n}, & \text{если } b_i = 0. \end{cases}$$

При этом

$$\langle \boldsymbol{x}_i, \boldsymbol{y} \rangle = \frac{1}{n}\big(n - 2d(\boldsymbol{a}_i, \boldsymbol{b})\big) \ge \frac{1}{n}(n - 2t) > 1 - 2\tau$$

и

$$\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle = \frac{1}{n}\Big(n - 2d\big(\boldsymbol{a}_i, \boldsymbol{a}_j\big)\Big) \le \frac{1}{n}(n - 2d) = 1 - 2\delta.$$

Похоже, можно применить лемму о векторах в $\mathbb{R}^n$, но для этого придётся «подправить» векторы $\boldsymbol{y}$ и $\boldsymbol{x}_1, \dots \boldsymbol{x}_m$. Для этого перейдём к векторам

$$2\tau\boldsymbol{y}, (\boldsymbol{x}_1 - (1-2\tau)\boldsymbol{y}), \dots, (\boldsymbol{x}_m - (1-2\tau)\boldsymbol{y}).$$

Для этих векторов получаем

$$\langle (\boldsymbol{x}_i - (1-2\tau)\boldsymbol{y}), 2\tau\boldsymbol{y} \rangle = 2\tau\langle \boldsymbol{x}_i, \boldsymbol{y} \rangle - 2\tau(1-2\tau)\langle \boldsymbol{y}, \boldsymbol{y} \rangle = 2\tau\langle \boldsymbol{x}_i, \boldsymbol{y} \rangle - 2\tau(1-2\tau) > 0$$

и

$$\big\langle (\boldsymbol{x}_i - (1-2\tau)\boldsymbol{y}), \big(\boldsymbol{x}_j - (1-2\tau)\boldsymbol{y}\big) \big\rangle = \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle + (1-2\tau)^2\langle \boldsymbol{y}, \boldsymbol{y} \rangle - (1-2\tau)\big(\langle \boldsymbol{x}_i, \boldsymbol{y} \rangle + \langle \boldsymbol{x}_j, \boldsymbol{y} \rangle\big)$$
$$\le 1 - 2\delta + (1-2\tau)^2 - 2(1-2\tau)^2 = -2(2\tau^2 - 2\tau + \delta) = 0.$$

Отсюда, по лемме о тупоугольной системе векторов, следует, что $m \le n$.

Теорема доказана.

3. Линейные коды

3.1. Основные понятия

Пусть q — степень простого числа и символы кодовых слов — элементы конечного поля $\mathbb{F}_q$.
$(n, M, d)_q$-код C называется *линейным*, если он является линейным подпространством пространства $\mathbb{F}_q^n$, то есть линейная комбинация кодовых слов также является кодовым словом.
Если $\dim C = k$, то говорят, что задан *линейный* $[n, k, d]_q$*-код*.

Пример линейного двоичного кода — код с проверкой чётности:

$$C := \{(a_1, \dots, a_n) \in \mathbb{F}_2^n \mid \textstyle\sum a_i = 0\}$$

Один из базисов этого кода:

$$(1,0,0,\dots,0,1)$$
$$(0,1,0,\dots,0,1)$$
$$(0,0,1,\dots,0,1)$$

$$\vdots$$
$$(0,0,0,\dots,1,1)$$

Если выписать базис линейного $[n,k,d]_q$-кода построчно в виде матрицы размера $k \times n$, получим *порождающую матрицу* кода.

Итак, для задания $[n,k,d]_q$-кода достаточно указать его порождающую матрицу $G \in \mathbb{F}_q^{k\times n}$. Число линейных комбинаций k базисных векторов с коэффициентами из $\mathbb{F}_q$ равно q^k, поэтому каждый $[n,k,d]_q$-код является $(n,q^k,d)_q$-кодом. Если линейный $[n,k,d]_q$-кода задан порождающей матрицей G, а исходное сообщение представлено как вектор $\boldsymbol{x} \in \mathbb{F}_q^k$, то закодировать его можно быстро и просто:

$$\boldsymbol{x} \xrightarrow{\text{кодирование}} \boldsymbol{x}^T G \in \mathbb{F}_q^n.$$

Обратно, если закодированное сообщение $\boldsymbol{a} \in \mathbb{F}_q^n$ было принято *без ошибок*, декодируем его, решая, например, методом Гаусса систему $\boldsymbol{x}^T G = \boldsymbol{a}$.

Порождающую матрицу $G \in \mathbb{F}_q^{k\times n}$ линейными преобразованиями и перестановками строк и столбцов можно привести к *каноническому виду*:

$$\left(\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & \\ 0 & 1 & \cdots & 0 & \\ \vdots & \vdots & \ddots & \vdots & \tilde{G} \\ 0 & 0 & \cdots & 1 & \end{array}\right),$$

где $\tilde{G} \in \mathbb{F}_q^{k\times(n-k)}$.
Тогда кодирование будет *систематическим*:

$$(x_1,\dots,x_k) \xrightarrow{\text{кодирование}} \left(x_1,\dots,x_k \mid \boldsymbol{x}\tilde{G}\right).$$

Если порождающая матрица кода задана в каноническом виде, то кодирование будет систематическим и слово $(x_1,\dots,x_k)$ переходит в

$$\left(x_1,\dots,x_k \mid \boldsymbol{x}\tilde{G}\right) = (x_1,\dots,x_k,\tilde{x}_1,\dots,\tilde{x}_{n-k}).$$

Разряды $x_1,\dots,x_k$ называются *информационными*, а $\tilde{x}_1,\dots,\tilde{x}_{n-k}$ — *проверочными*.

Если матрица G исходно задана не в каноническом виде, а мы приводим её к каноническому виду, то получается в общем случае не тот же код, а *эквивалентный* ему. Формально, коды C_1 и C_2 *эквивалентны*, если существует перестановка π и константы $r_1,\dots,r_n \in \mathbb{F}_q \setminus \{0\}$, такие, что

$$(a_1,\dots,a_n) \in C_1 \Leftrightarrow \left(r_1 a_{\pi(1)},\dots,r_n a_{\pi(n)}\right) \in C_2.$$

Для эквивалентных кодов $d(C_1) = d(C_2)$.

Утверждение (о кодовом расстоянии линейных кодов).
Для любого линейного кода C имеем

$$d(C) = \min_{\substack{\boldsymbol{a}\in C\\ \boldsymbol{a}\neq\boldsymbol{0}}} \|\boldsymbol{a}\|.$$

Доказательство. Поскольку $\boldsymbol{0} \in C$, то

$$d(C) \stackrel{\text{def}}{=} \min_{\substack{\boldsymbol{a},\boldsymbol{b}\in C\\ \boldsymbol{a}\neq\boldsymbol{b}}} d(\boldsymbol{a},\boldsymbol{b}) \le \min_{\substack{\boldsymbol{a}\in C\\ \boldsymbol{a}\neq\boldsymbol{0}}} d(\boldsymbol{a},\boldsymbol{0}) = \min_{\substack{\boldsymbol{a}\in C\\ \boldsymbol{a}\neq\boldsymbol{0}}} \|\boldsymbol{a}\|.$$

В обратную сторону. Пусть кодовое расстояние достигается на паре слов $\boldsymbol{a}^*, \boldsymbol{b}^*$. Тогда так как $(\boldsymbol{a}^* - \boldsymbol{b}^*) \in C$, получаем

$$d(C) = d(\boldsymbol{a}^*,\boldsymbol{b}^*) = \|\boldsymbol{a}^* - \boldsymbol{b}^*\| \ge \min_{\substack{\boldsymbol{a}\in C\\ \boldsymbol{a}\neq\boldsymbol{0}}} \|\boldsymbol{a}\|.$$

Утверждение доказано.

Проверочная матрица H линейного кода — это матрица однородной системы линейных уравнений, которым удовлетворяет код. Например, для кода с проверкой чётности $H = (11\dots1)$. Для любого

кодового слова $\boldsymbol{a}$, по определению матрицы H, выполнено равенство $H\boldsymbol{a}^T = \mathbf{0}$. Для $[n,k,d]_q$-кода $H \in \mathbb{F}_q^{(n-k)\times n}$.

Утверждение.
Если G и H — порождающая и проверочная матрицы линейного $[n,k,d]_q$-кода, то $HG^T = \mathbf{0}^{(n-k)\times k}$.

Доказательство: достаточно заметить, что каждая строка матрицы G — это кодовое слово, а значит, она удовлетворяет системе, задаваемой H.

Если коды C_1 и C_2 таковы, что проверочная матрица C_1 является порождающей матрицей для C_2, то эти коды называют *двойственными*. Коды C_1 и C_2 двойственны т. и т.т., когда $\langle \boldsymbol{a}_1, \boldsymbol{a}_2 \rangle = 0$ для любых слов $\boldsymbol{a}_1 \in C_1, \boldsymbol{a}_2 \in C_2$. Поэтому двойственные коды называются также *ортогональными*.

Пусть H — проверочная матрица кода C, и пусть $\boldsymbol{a} \in C$. Если при передаче $\boldsymbol{a}$ по каналу произошло t ошибок, на выходе из канала имеем слово $\boldsymbol{b}$. Вектор $\boldsymbol{e} \coloneqq \boldsymbol{b} - \boldsymbol{a}$ называется *вектором ошибок*. Очевидно, $\|\boldsymbol{e}\| = t$. Исправить ошибки в $\boldsymbol{b}$ — то же самое, что найти вектор $\boldsymbol{e}$. Используем равенство

$$H\boldsymbol{b}^T = H\boldsymbol{a}^T + H\boldsymbol{e}^T = H\boldsymbol{e}^T.$$

Получаем задачу: найти вектор $\boldsymbol{e}$, такой, что

$$\begin{cases} \|\boldsymbol{e}\| < \frac{d(C)}{2}, \\ H\boldsymbol{e}^T = H\boldsymbol{b}^T. \end{cases}$$

Утверждение.
Если решение этой системы существует, то оно единственное.

Доказательство. Пусть нашлись разные решения $\boldsymbol{e}_1 \neq \boldsymbol{e}_2$. Тогда $\|\boldsymbol{e}_1 - \boldsymbol{e}_2\| \leq \|\boldsymbol{e}_1\| + \|\boldsymbol{e}_2\| < d(C)$ и $H(\boldsymbol{e}_1 - \boldsymbol{e}_2)^T = 0$ — противоречие.

Можно составить таблицу решений системы

$$\begin{cases} \|\boldsymbol{e}\| < \frac{d(C)}{2}, \\ H\boldsymbol{e}^T = \boldsymbol{s}. \end{cases}$$

для всевозможных $\boldsymbol{s}$. При получении из канала слова $\boldsymbol{b}$ мы вычисляем $\boldsymbol{s} \coloneqq H\boldsymbol{b}^T$ (*синдром* слова $\boldsymbol{b}$), и для этого $\boldsymbol{s}$ смотрим в таблице соответствующий вектор $\boldsymbol{e}$. Декодированное сообщение — решение системы

$$\boldsymbol{x}^T G = (\boldsymbol{b} - \boldsymbol{e}).$$

3.2. Теорема Варшамова—Гилберта

Утверждение. (О связи кодового расстояния и проверочной матрицы)
Линейный код, определяемый проверочной матрицей H, имеет расстояние d т. и т.т., когда любые $(d-1)$ столбцов H линейно независимы, и найдутся d линейно зависимых столбцов в H.

Доказательство. Пусть $\boldsymbol{a} = (a_1, \dots, a_n) \neq \mathbf{0}$ — произвольное кодовое слово кода с проверочной матрицей H. Пусть $H_1, \dots, H_n$ — столбцы H. Имеем

$$H\boldsymbol{a}^T = a_1 \cdot H_1 + \cdots + a_n \cdot H_n.$$

Если $i_1, \dots, i_t$ — все ненулевые координаты $\boldsymbol{a}$, то

$$a_1 H_1 + \cdots + a_n H_n = a_{i_1} H_{i_1} + \cdots + a_{i_t} H_{i_t}.$$

Так как $\boldsymbol{a}$ — кодовое слово, то $H\boldsymbol{a}^T = \mathbf{0}$, то есть

$$a_{i_1} H_{i_1} + \cdots + a_{i_t} H_{i_t} = \mathbf{0}.$$

Мы получили, что в коде есть слово веса не более t т. и т.т., когда некоторые t столбцов матрицы H линейно зависимы. Осталось воспользоваться формулой, справедливой для любых линейных кодов:

$$d(C) = \min_{\substack{\boldsymbol{a} \in C \\ \boldsymbol{a} \neq \mathbf{0}}} \|\boldsymbol{a}\|.$$

Утверждение доказано.

Одним из важнейший достаточных условий существования линейных кодов является следующая теорема.

Теорема. (P.P. Варшамов, E.N. Gilbert)

Пусть натуральные числа n, k, d' таковы, что

$$\sum_{j=0}^{d'-1} \binom{n-1}{j} < 2^{n-k}.$$

Тогда существует $[n, k, d]$-код, где $d > d'$.

Доказательство. Покажем, что в условиях теоремы можно построить матрицу $H \in \mathbb{F}_2^{(n-k)\times n}$, у которой любые d' столбцов линейно независимы. Будем строить матрицу по столбцам. Пусть уже выбраны m столбцов (где $m < n$) и требуется выбрать $(m+1)$-й столбец. Этот новый столбец не должен образовывать нулевую линейную комбинацию с $(d'-1)$ или менее из уже выбранных столбцов. Так как мы работаем в $\mathbb{F}_2$, то это равносильно тому, что выбираемый столбец не равен сумме $(d'-1)$ или менее уже выбранных столбцов. Количество таких сумм равно

$$\sum_{j=0}^{d'-1} \binom{m}{j} \le \sum_{j=0}^{d'-1} \binom{n-1}{j} < 2^{n-k}.$$

Итак, запрещённых для выбора столбцов у нас оказывается строго меньше 2^{n-k}, а всего векторов длины $(n-k)$ ровно 2^{n-k}. Значит, найдётся вектор из $\mathbb{F}_2^{n-k}$, который можно добавить в качестве очередного столбца.

Теорема доказана.

3.3. Двоичный код Хемминга

В этом разделе мы введём коды Хемминга, пожалуй, самые простые нетривиальные линейные коды.

Рассмотренный нами ранее простейший двоичный код с проверкой чётности

$$\{(a_1, \dots, a_n) \in \mathbb{F}_2^n \mid \textstyle\sum a_i = 0\}$$

может обнаруживать одну ошибку, т.к. если один разряд a_i заменить на противоположный, соотношение $\sum a_i = 0$ нарушится. Но исправить ошибку не удастся. Хочется построить двоичный код, *исправляющий* хотя бы одну ошибку. Для этого вместо «глобального» контроля чётности применим несколько «дихотомических» проверок на чётность.

Пример для $n = 7$. Рассмотрим код, удовлетворяющий соотношениям

$$\begin{aligned} a_4 + a_5 + a_6 + a_7 &= 0, \\ a_2 + a_3 + a_6 + a_7 &= 0, \\ a_1 + a_3 + a_5 + a_7 &= 0. \end{aligned}$$

Проверочная матрица этого кода:

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

Столбцы матрицы — всевозможные ненулевые векторы высоты 3: j-й столбец суть двоичная запись числа j. Любая пара столбцов л.н.з., значит $d(C) \ge 3$, значит, этот код исправляет одну ошибку и обнаруживает две. Если ошибка случается в a_j, то можно вычислить левые части проверочных соотношений, и они дадут двоичную запись j, например:

$$\begin{aligned} a_4 + \overline{a_5} + a_6 + a_7 &= 1 \\ a_2 + a_3 + a_6 + a_7 &= 0 \\ a_1 + a_3 + \overline{a_5} + a_7 &= 1 \end{aligned}$$

Общий случай: $n \coloneqq 2^m - 1$ для некоторого m. *Двоичный код Хемминга* длины n с параметрами $\left[2^m - 1, 2^m - m - 1,\ 3\right]$ определяется проверочной матрицей $H \in \mathbb{F}_2^{m \times n}$, столбцы которой — всевозможные двоичные векторы высоты m:

$$\begin{pmatrix} 1 & 0 & 1 & \cdots & 1 \\ 0 & 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

Теорема Хемминга утверждает, что для любого (n, M, d)-кода выполнено неравенство $M \leq \frac{2^n}{\sum_{k=0}^{\lfloor (d-1)/2 \rfloor} \binom{n}{k}}$. Напомним, что коды, на которых достигается граница Хемминга, называются *совершенными* или *плотно упакованными*. Нетрудно показать, что код Хемминга совершенен. Действительно, для кода Хемминга имеем $n = 2^m - 1$, $M = 2^{2^m - m - 1}$ и $d = 3$. Отсюда $\sum_{k=0}^{\lfloor (d-1)/2 \rfloor} \binom{n}{k} = n + 1 = 2^m = 2^n / M$.

3.4. Границы мощностей для линейных кодов

3.4.1. Граница Синглтона для линейных кодов

Утверждение. (Граница Синглтона для линейных кодов)
Для любого $[n, k, d]_q$-кода выполнено неравенство $k \leq n - d + 1$.

Доказательство. По теореме Синглтона, для любого $(n, M, d)_q$-кода выполнено $M \leq q^{n-d+1}$. С другой стороны, для линейного кода $M = q^k$.

3.4.2. Граница Грайсмера—Соломона—Штиффлера

Теорема об остаточном коде. (G. Solomon, J.J. Stiffler)
Если существует $[n, k, d]_q$-код, то существует и $[n - d, k - 1, d']_q$-код, где $d' \geq d/q$.

Доказательство. Пусть G — порождающая матрица некоторого $[n, k, d]_q$-кода C. Б.о.о. будем считать, что первая строка G содержит ровно d ненулевых элементов и имеет вид $(r_1 \ldots r_d 0 \ldots 0)$. Порождающая матрица кода C:

$$G = \begin{pmatrix} r_1 \ldots r_d & 0 \ldots 0 \\ \ldots & G' \end{pmatrix}.$$

Имеем $G' \in \mathbb{F}_q^{(k-1)\times(n-d)}$. Покажем, что $\operatorname{rk} G' = k - 1$. Допустим противное: некоторая нетривиальная линейная комбинация строк G' равняется $\mathbf{0}$. Тогда линейная комбинация соответствующих строк G равна $(t_1 \ldots t_d 0 \ldots 0)$, где $\forall i\ (t_i \neq 0)$.

Линейная комбинация U некоторых строк G равна $(t_1 \ldots t_d 0 \ldots 0)$, где $\forall i\ (t_i \neq 0)$. Т.к. $\mathbb{F}_q$ — поле, то $\exists s \in \mathbb{F}_q$, такой, что $st_d = -r_d$. Тогда $s \cdot U + (r_1 \ldots r_d 0 \ldots 0)$ — линейная комбинация строк G, равная

$$\left((st_1 + r_1) \ldots (st_{d-1} + r_{d-1})\ 0 0 \ldots 0\right).$$

Это противоречит условию $d(C) = d$.

Итак, $G' \in \mathbb{F}_q^{(k-1)\times(n-d)}$ и $\operatorname{rk} G' = k - 1$. Значит, G' является порождающей матрицей некоторого $[n - d, k - 1, d']_q$-кода C' (этот код называется *остаточным* для исходного кода C).

Рассмотрим любой ненулевой вектор кода C':

$$\boldsymbol{a}' \coloneqq (a'_1, \ldots, a'_{n-d}) \neq \mathbf{0},$$

такой, что $\|\boldsymbol{a}'\| = d'$.

В коде C есть вектор вида

$$(a_1, \ldots, a_d, a'_1, \ldots, a'_{n-d}).$$

Пусть $f_1, \ldots, f_q$ — все элементы поля $\mathbb{F}_q$.

Коду C принадлежат все векторы вида

$$\big((a_1 - f_i r_1), \dots, (a_d - f_i r_d), a'_1, \dots, a'_{n-d}\big).$$

Запишем эти q векторов построчно в виде матрицы и оценим количество ненулевых элементов в ней:

$$\begin{pmatrix} (a_1 - f_1 r_1) & \dots & (a_d - f_1 r_d) & a'_1 & \dots & a'_{n-d} \\ (a_1 - f_2 r_1) & \dots & (a_d - f_2 r_d) & a'_1 & \dots & a'_{n-d} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (a_1 - f_q r_1) & \dots & (a_d - f_q r_d) & a'_1 & \dots & a'_{n-d} \end{pmatrix}.$$

В каждой подстроке вида $a'_1, \dots, a'_{n-d}$ ровно d' элементов отличны от нуля. В каждом из первых d столбцов ровно один ноль. Поэтому в рассмотренной матрице $d(q-1) + qd'$ ненулевых элементов. С другой стороны, каждая строка матрицы — ненулевой вектор кода C, значит, в матрице не менее чем qd ненулевых элементов. Отсюда $d(q-1) + qd' \geq qd$, и, следовательно $d' \geq d/q$.

Теорема доказана.

Теорема. (J.H. Griesmer, G. Solomon, J.J. Stiffler)
Для любого $[n, k, d]_q$-кода имеем

$$n \geq \sum_{i=0}^{k-1} \left\lceil \frac{d}{q^i} \right\rceil.$$

Доказательство. Утверждение очевидно при $k = 1$. Предположим, что оно выполнено для кодов с размерностью $\leq k-1$ и докажем его для $[n, k, d]_q$-кодов. Обозначим через $N(k, d)$ минимальную длину слов у кода с размерностью k и расстоянием d. Пусть C — какой-нибудь $[N(k, d), k, d]_q$-код. Остаточный для C код имеет параметры $[N(k, d) - d, k-1, d']_q$, и, по предположению индукции, для него справедливо неравенство

$$N(k, d) - d \geq \sum_{i=0}^{k-2} \left\lceil \frac{d'}{q^i} \right\rceil.$$

Отсюда, с учётом предположения индукции и соотношения $d' \geq d/q$, получаем

$$N(k, d) \geq d + \sum_{i=0}^{k-2} \left\lceil \frac{d'}{q^i} \right\rceil \geq d + \sum_{i=0}^{k-2} \left\lceil \frac{d}{q^{i+1}} \right\rceil = \sum_{i=0}^{k-1} \left\lceil \frac{d}{q^i} \right\rceil.$$

Теорема доказана.

3.5. Графы-расширители и коды на их основе

3.5.1. Графы-расширители

Двудольный граф с долями L и R называется $(n, m, \Delta, \alpha, c)$-*расширителем*, если

- $|L| = n$, $|R| = m$,
- $\deg u = \Delta$ для любого $u \in L$,
- для любого $S \subseteq L$ при $|S| \leq \alpha n$ выполнено неравенство $|N(S)| \geq c \cdot |S|$, где $N(S)$ — множество вершин в R, смежных с вершинами из S.

Графы-расширители также называются *расширяющими*, *экспандерными графами* или *экспандерами* (от англ. *expander*). Начало применению расширительных свойств графов в кодировании положили работы советских математиков М.С. Пинскера, Л.А. Бассалыго, Г.А. Маргулиса в 1970-х годах.

Теорема о существовании расширителей.
Пусть $\Delta \geq 3$, $c \leq \Delta - 2$, $\alpha < 1$ и $m \geq 4n\Delta^2\sqrt{\alpha}$.
Тогда при всех натуральных n существуют $(n, m, \Delta, \alpha, c)$-расширители.

Доказательство. Построим случайный двудольный граф G и докажем, что он с большой вероятностью будет искомым. Зафиксируем множества L и R ($|L| = n, |R| = m$) и проведём по Δ

рёбер из каждой вершины в L в выбираемые равновероятно и независимо вершины в R (в итоге некоторые из этих Δ рёбер могут попасть в одни и те же вершины R).

Рёбра G имеют естественную нумерацию, в том порядке, в котором мы определяли их концы в R (сначала Δ рёбер из 1-й вершины L, затем Δ рёбер из 2-й вершины L и т.д.).

Если G не является расширителем, то найдётся такое $S \subset L$, для которого $|N(S)| < (\Delta - 2) \cdot |S|$. Оценим вероятность того, что фиксированное множество S оказалось таким «плохим» при случайном выборе концов рёбер из S в R. А затем оценим вероятность того, что G не расширитель, по формуле

$$\Pr[G \text{ плохой}] = \Pr[\exists \text{плохое } S] \le \sum_S \Pr[S \text{ — плохое}].$$

Каждое ребро вида (u, v), где $u \in S, v \in N(S)$, отнесём к одному из двух типов. Если никакое ребро из S в $N(S)$ с меньшим номером не ведёт в v, то ребро (u, v) назовём «*первопроходцем*». В противном случае назовём (u, v) «*дублем*». Очевидно, всего будет $N(S)$ «первопроходцев» и $(\Delta \cdot |S| - |N(S)|)$ «дублей».

Мы предполагаем, что $|N(S)| < (\Delta - 2) \cdot |S|$, а значит, «дублей» будет не менее $2 \cdot |S|$.

Вероятность того, что среди рёбер из S в $N(S)$ есть $2s$ дублей, не превосходит

$$\binom{\Delta \cdot |S|}{2|S|} \left(\frac{\Delta \cdot |S|}{m}\right)^{2|S|}.$$

Первый из сомножителей оценивает число способов выбрать рёбра-дубли, второй — вероятность попадания конца «дубля» в одну из вершин в $|N(S)|$.

Значит,

$$\Pr[G \text{ не расширитель}] \le \sum_{\substack{S \subset L \\ 1 \le |S| \le \alpha n}} \Pr[S \text{ «плохое»}] \le \sum_{1 \le s \le \alpha n} \binom{n}{s} \binom{\Delta \cdot s}{2s} \left(\frac{\Delta \cdot s}{m}\right)^{2s}.$$

С учётом оценки $\binom{a}{b} < \left(\frac{ea}{b}\right)^b$ и того, что при $m \ge 4n\Delta^2\sqrt{\alpha}$ выполняется неравенство $\frac{e^3\Delta^4\alpha n^2}{4m^2} < \frac{1}{3}$, получаем

$$\Pr[G \text{ не расширитель}] \le \sum_{1 \le s \le \alpha n} \left(\frac{en}{s}\right)^s \left(\frac{e\Delta s}{2s}\right)^{2s} \left(\frac{\Delta s}{m}\right)^{2s} = \sum_{1 \le s \le \alpha n} \left(\frac{e^3\Delta^4 s n}{4m^2}\right)^s \le \sum_{1 \le s \le \alpha n} \left(\frac{e^3\Delta^4 \alpha n^2}{4m^2}\right)^s$$
$$\le \sum_{1 \le s \le \alpha n} \left(\frac{1}{3}\right)^s < \frac{1}{2}.$$

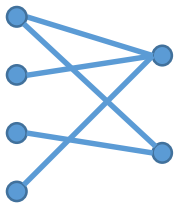
Итак, случайный двудольный мультиграф будет расширителем с вероятностью не менее $\frac{1}{2}$. Чтобы от мультиграфа перейти к обычному графу, достаточно перенаправить концы рёбер-дублей в произвольные вершины R. Свойства расширительности от этого могут только улучшиться.

Теорема доказана.

3.5.2. Коды на основе расширителей

Код на основе двудольного графа — это линейный двоичный код, строящийся следующим образом. Вершинам из L соответствуют переменные $x_1, \dots, x_n$. Вершинам из R соответствуют уравнения: если в вершину $v \in R$ входят рёбра из вершин $u_{i_1}, \dots, u_{i_l}$, то уравнение будет $x_{i_1} + \dots + x_{i_l} = 0$. Искомый код состоит из всех слов $(x_1 \dots x_n)$, удовлетворяющих системе этих уравнений.

Например, для графа



соответствующий код будет выглядеть так: $\{(x_1x_2x_3x_4) \mid x_1 + x_2 + x_4 = 0, \quad x_1 + x_3 = 0\}$.

Коды на основе расширителей введены М. Сипсером и Д. Шпильманом, являются обобщением низкоплотностных кодов (LDPC-codes) Р. Галлагера (R.G. Gallager).

Утверждение.
Код, построенный по двудольному графу, в котором $|L| = n$ и $|R| = m$, является двоичным линейным $[n, k, d]$-кодом, где $k \geq n - m$.

Доказательство: код является множеством решений системы из m уравнений с n неизвестными, а значит, он образует линейное пространство размерности не менее чем $n - m$.

Теорема о кодовом расстоянии. (M. Sipser, D.A. Spielman)
Если $c > \frac{\Delta}{2}$ и C — код, построенный на основе $(n, m, \Delta, \alpha, c)$-расширителя, то $d(C) > \alpha n$.

Доказательство от противного. Допустим, что $d(C) \leq \alpha n$. Тогда найдётся слово $\boldsymbol{a} \in C$, такое, что $\boldsymbol{a} \neq \mathbf{0}$ и $\|\boldsymbol{a}\| \leq \alpha n$.
Пусть $I \coloneqq \{u_1, \ldots, u_{\|\boldsymbol{a}\|}\}$ — вершины из L, соответствующие единичным координатам $\boldsymbol{a}$. Так как наш граф — расширитель и $\|\boldsymbol{a}\| \leq \alpha n$, то

$$|N(I)| \geq c \cdot |I| > \frac{\Delta}{2} \cdot |I|.$$

Всего из I в $N(I)$ ведёт ровно $\Delta \cdot |I|$ рёбер. Поэтому *среднее* число рёбер, входящее в вершины $N(I)$ из I, равно

$$\frac{\Delta \cdot |I|}{|N(I)|} < \frac{\Delta \cdot |I|}{\frac{\Delta}{2} \cdot |I|} = 2.$$

Значит, в $N(I)$ найдётся вершина, в которую входит ровно одно ребро из I. Получается, что среди задающих код уравнений есть такое уравнение $x_{i_1} + \cdots + x_{i_l} = 0$, в котором ровно одна из переменных на слове $\boldsymbol{a}$ обращена в единицу. Но этого не может быть в предположении, что $\boldsymbol{a}$ является решением этого уравнения. Противоречие.

Теорема доказана.

3.5.3. Алгоритм Сипсера—Шпильмана

3.5.3.1. Формулировка алгоритма декодирования

Пусть $c > \frac{3\Delta}{4}$, и C — код, построенный на основе $(n, m, \Delta, \alpha, c)$-расширителя. Пусть слово $\boldsymbol{a}'$ получено из некоторого кодового слова $\boldsymbol{a}$ искажением не более чем $\frac{\alpha n}{4}$ битов. Тогда восстановить $\boldsymbol{a}$, зная $\boldsymbol{a}'$, можно с помощью следующего алгоритма **(M. Sipser, D.A. Spielman)**:

1. Если $\boldsymbol{a}' \in C$, то выводим $\boldsymbol{a}'$ и завершаем работу.
2. Если $\boldsymbol{a}' \notin C$, то для $\boldsymbol{a}'$ некоторые из уравнений (отвечающих вершинам в R) нарушены. Считаем поочерёдно для каждого бита $\boldsymbol{a}'$ число нарушенных уравнений, в которых он участвует. Если их $> \Delta/2$, инвертируем этот бит и идём на шаг 1.

3.5.3.2. Лемма о результативном бите

Чтобы доказать корректность алгоритма Сипсера—Шпильмана, нам потребуется следующая лемма.

Лемма о «результативном бите».
Пусть $c > \frac{3\Delta}{4}$, и C — код, построенный на основе $(n, m, \Delta, \alpha, c)$-расширителя. Пусть $\boldsymbol{a}' \notin C$, но при этом $d(\boldsymbol{a}, \boldsymbol{a}') \leq \alpha n$ для некоторого $\boldsymbol{a} \in C$. Тогда в $\boldsymbol{a}'$ найдётся бит, обращение которого на

противоположный строго уменьшает число невыполненных для $\boldsymbol{a}'$ уравнений. (Имеются в виду уравнения, построенные по графу-расширителю.)

Доказательство. Пусть $\boldsymbol{a}'$ — не кодовое слово, находящееся от ближайшего кодового на расстоянии $\leq \alpha n$. Пусть $I \subset L$ — множество вершин, соответствующих координатам, в которых $\boldsymbol{a}'$ отличается от ближайшего кодового слова.

Обозначим через $N_{\text{pass}}(I)$ вершины из $N(I)$, соответствующие уравнениям, выполненным на слове $\boldsymbol{a}'$. Аналогично, пусть $N_{\text{fail}}(I)$ — вершины из $N(I)$, отвечающие нарушенным уравнениям.
Так как наш граф расширитель, и $|I| \leq \alpha n$, то

$$\left|N_{\text{pass}}(I)\right| + |N_{\text{fail}}(I)| = |N(I)| \geq c|I| > \frac{3\Delta}{4} \cdot |I|.$$

Из каждой вершины $N(I)$ в I ведёт хотя бы одно ребро. При этом из каждой вершины $N_{\text{pass}}(I)$ в I ведёт хотя бы два ребра: чтобы «обмануть» уравнение, нужно инвертировать в нём чётное количество переменных. С другой стороны, число рёбер между I и $N(I)$ в точности равно $\Delta \cdot |I|$. Из всего сказанного вытекает цепочка соотношений

$$\Delta \cdot |I| \geq |N_{\text{fail}}(I)| + 2 \cdot \left|N_{\text{pass}}(I)\right| > |N_{\text{fail}}(I)| + 2 \cdot \left(\frac{3\Delta}{4} \cdot |I| - |N_{\text{fail}}(I)|\right).$$

Отсюда $|N_{\text{fail}}(I)| > \frac{\Delta}{2} \cdot |I|$. Значит, в I найдётся вершина, для которой нарушены *больше* половины тех уравнений, в которых она участвует. То есть даже не зная I, можно утверждать следующее: среди координат $\boldsymbol{a}'$ есть хотя бы одна такая, обратив значение которой мы уменьшим число нарушенных уравнений.

Лемма доказана.

3.5.3.3. Завершение доказательства корректности алгоритма

Лемма о результативном бите говорит, что если к очередному шагу алгоритма мы пришли с некодовым словом $\boldsymbol{a}'$, находящимся от ближайшего кодового на расстоянии $\leq \alpha n$, то очередной бит для изменения мы найдём. Осталось доказать, что, начав со слова $\boldsymbol{a}'$ на расстоянии $\leq \frac{\alpha n}{4}$ от ближайшего кодового слова $\boldsymbol{a}$, мы не «притянемся» случайно к какому-то другому кодовому слову $\boldsymbol{b} \neq \boldsymbol{a}$.

До начала работы алгоритма $|I| \leq \frac{\alpha n}{4}$, и, значит, число нарушенных уравнений не превосходит $\Delta \cdot |I| \leq \frac{\alpha n \Delta}{4}$. В ходе работы алгоритма число нарушенных уравнений уменьшается.

Пусть на очередном шаге получено слово $\boldsymbol{a}''$, и пусть I'' — биты, в которых $\boldsymbol{a}''$ отличается от ближайшего кодового слова. Имеем

$$\frac{\alpha n \Delta}{4} \geq \#\text{наруш. ур.} = N_{\text{fail}}(I'') > \frac{\Delta}{2} \cdot |I''|,$$

отсюда $|I''| < \frac{\alpha n}{2}$.

Итак, на каждом шаге алгоритма получаем слово, отличающееся от ближайшего кодового менее чем в $\frac{\alpha n}{2}$ битах. Т.к. на каждом шаге в слове меняется только один бит и $d(C) > \alpha n$, то кодовое слово, к которому мы стремимся, всё время одно и то же: если $d(\boldsymbol{a}, \boldsymbol{b}) > t$, то, находясь в шаре $S_{t/2}(\boldsymbol{a})$ и смещаясь на один бит, мы не «вывалимся» в шар $S_{t/2}(\boldsymbol{b})$.

4. Коды Рида—Соломона и Рида—Маллера

4.1. Коды Рида—Соломона (I.S. Reed, G. Solomon)

4.1.1. Определение

Пусть $k \leq n \leq q$. Пусть $t_1, \ldots, t_n \in \mathbb{F}_q$ — фиксированные, попарно различные элементы. Рассмотрим такое множество слов:

$$C \coloneqq \{(P(t_1), \dots, P(t_n)) \mid P \in \mathbb{F}_q[x] \wedge \deg P < k\}.$$

Непосредственно проверяется, что C — линейное пространство:

$$\alpha \cdot (P_1(t_1), \dots, P_1(t_n)) + \beta \cdot (P_2(t_1), \dots, P_2(t_n)) = ((\alpha P_1 + \beta P_2)(t_1), \dots, (\alpha P_1 + \beta P_2)(t_n))$$

У многочлена степени $< k$ может быть не более $(k-1)$ корней, поэтому если $P \not\equiv 0$, то в векторе $(P(t_1), \dots, P(t_n))$ не более $(k-1)$ нулевых координат. Отсюда

$$d(C) = \min_{\substack{\boldsymbol{a} \in C \\ \boldsymbol{a} \neq \boldsymbol{0}}} \|\boldsymbol{a}\| = \min_{P \not\equiv 0} \#\{i \mid P(t_i) \neq 0\} = n - (k-1).$$

Векторы $(P(t_1), \dots, P(t_n))$ при разных P различны: если выполнено $(P_1(t_1), \dots, P_1(t_n)) = (P_2(t_1), \dots, P_2(t_n))$,
то у многочлена $(P_1 - P_2)$ не менее n корней, а т.к. $\deg(P_1 - P_2) < k \leq n$, то $(P_1 - P_2) \equiv 0$. Значит, $\dim C = k$.

Итак, для любых $k \leq n \leq q$ множество C является $[n, k, d]_q$-кодом, где $d = n - k + 1$. Этот код называется *кодом Рида—Соломона* или *RS-кодом*.

Вспомним границу Синглтона: *для любого $[n, k, d]_q$-кода выполнено $k \leq n - d + 1$*. То есть построенный код достигает границы Синглтона и, стало быть, является MDS-кодом. Недостаток кодов Рида—Соломона состоит в том, что кодовый алфавит нужно брать очень большим, т.к. $q \geq n$.

4.1.2. Декодирование RS-кодов

Т.к. $d(C) = n - k + 1$, то код может исправлять $\left\lfloor \frac{n-k}{2} \right\rfloor$ ошибок. В этом разделе мы покажем, как можно эффективно исправлять ошибки в словах RS-кодов. Формально поставить задачу декодирования можно следующим образом:

- Дано искажённое кодовое слово RS-кода $(\hat{p}_1, \dots, \hat{p}_n) \in \mathbb{F}_q^n$.
- Найти $P \in \mathbb{F}_q[x]$, такой, что $\deg P \leq k - 1$ и $\#\{i \mid P(t_i) \neq \hat{p}_i\} \leq \left\lfloor \frac{n-k}{2} \right\rfloor$ (нам гарантируется, что такой P существует).

Приводимая ниже техника называется *алгоритмом Берлекэмпа—Велча* (E.R. Berlekamp, L.R. Welch).

Рассмотрим *многочлен ошибок*

$$E(x) \coloneqq \prod_{i:\, P(t_i) \neq \hat{p}_i} (x - t_i)$$

и вспомогательный многочлен $U(x) \coloneqq E(x) \cdot P(x)$.

Обозначим $s \coloneqq \#\{i \mid P(t_i) \neq \hat{p}_i\}$. Заметим, что $\deg E = s,\ \operatorname{coef}_{x^s} E = 1,\ \deg U \leq \deg E + \deg P \leq s + k - 1$, и для любого $i \in \{1, \dots, n\}$ выполнено равенство $U(t_i) = E(t_i) \cdot \hat{p}_i$.

Идея: мы не знаем P, так что попытаемся найти *какие-то* многочлены $\tilde{E}$ и $\tilde{U}$, для которых

- $\deg \tilde{E} = \tilde{s}$ и $\operatorname{coef}_{x^{\tilde{s}}} \tilde{E} = 1$, где $\tilde{s} \leq (n-k)/2$,
- $\deg \tilde{U} \leq \tilde{s} + k - 1$,
- для любого $i \in \{1, \dots, n\}$ выполнено равенство $\tilde{U}(t_i) = \tilde{E}(t_i) \cdot \hat{p}_i$.

Указанную тройку условий назовём условиями Берлекэмпа—Велча, или БВ-условиями. Мы знаем, что многочлены $\tilde{E}$ и $\tilde{U}$, удовлетворяющие БВ-условиям, точно найдутся (например, такова пара многочленов E, U, удовлетворяющая этой системе при $\tilde{s} = s$). Остаётся вопрос: как эффективно найти какие-нибудь $\tilde{E}$ и $\tilde{U}$ и что делать, если найденные $\tilde{E}$ и $\tilde{U}$ не совпадут с нужными нам E и U?

Зафиксируем $\tilde{s}$ и положим $\tilde{E} = x^{\tilde{s}} + \sum_{j \leq \tilde{s}-1} e_j x^j$ и $\tilde{U} = \sum_{j \leq \tilde{s}+k-1} u_j x^j$, где $e_0, \dots, e_{\tilde{s}-1}, u_0, \dots, u_{\tilde{s}+k-1}$ — неопределённые коэффициенты. Получим систему

$$\begin{cases} \hat{p}_1 t_1^{\tilde{s}} + \sum\limits_{0\le j\le \tilde{s}-1} \hat{p}_1 e_j t_1^j = \sum\limits_{0\le j\le k+\tilde{s}-1} u_j t_1^j, \\ \vdots \\ \hat{p}_n t_n^{\tilde{s}} + \sum\limits_{0\le j\le \tilde{s}-1} \hat{p}_n e_j t_n^j = \sum\limits_{0\le j\le k+\tilde{s}-1} u_j t_n^j. \end{cases}$$

При любом фиксированном $\tilde{s} \le (n-k)/2$ эта система линейная относительно $e_0, \dots, e_{\tilde{s}-1}$, $u_0, \dots, u_{s+k-1}$. Перебирая $\tilde{s} = 0,1, \dots$, найдём то $\tilde{s}$, при котором решение системы есть (такое $\tilde{s}$ найдётся, хотя бы, при $\tilde{s} = s$). Тем самым найдём пару $\tilde{E}$ и $\tilde{U}$, удовлетворяющую БВ-условиям. Мы нашли *какие-то* $\tilde{E}$ и $\tilde{U}$. Если бы это были *те самые* E и U, то можно было бы выразить $P(x) = \frac{U(x)}{E(x)}$. Оказывается, и в ином случае P будет выражаться так же, как утверждает следующая лемма.

Лемма.
Если пары (E_1, U_1) и (E_2, U_2) удовлетворяют БВ-условиям, то $\frac{U_1}{E_1} \equiv \frac{U_2}{E_2}$.

Доказательство. Пусть (E_1, U_1) и (E_2, U_2) удовлетворяют БВ-условиям. Имеем $\deg U_1E_2 \le \deg U_1 + \deg E_2 \le \left(\frac{n-k}{2} + k - 1\right) + \frac{n-k}{2} \le n-1$. Аналогично $\deg E_1U_2 \le n-1$. Отсюда следует, что

$$\deg(U_1E_2 - E_1U_2) \le n-1.$$

Далее, для любого i имеем $U_1(t_i)E_2(t_i) = \big(\hat{p}_i E_1(t_i)\big)E_2(t_i) = E_1(t_i)\big(\hat{p}_i E_2(t_i)\big) = E_1(t_i)U_2(t_i)$. То есть для $i = 1, \dots, n$ выполнено

$$U_1(t_i)E_2(t_i) - E_1(t_i)U_2(t_i) = 0.$$

Отсюда следует, что многочлен $\big(U_1(x)E_2(x) - E_1(x)U_2(x)\big)$ тождественный ноль, а это эквивалентно тождеству $\frac{U_1}{E_1} \equiv \frac{U_2}{E_2}$.

Лемма доказана.

4.2. Коды Рида—Маллера (I.S. Reed, D.E. Muller)

Можно представить два пути обобщения конструкции Рида и Соломона:

- рассматривать многочлены не от одной, а от многих переменных,
- рассматривать не все возможные многочлены, а специально выбранное их подмножество.

Идя по первому пути, мы получаем коды Рида—Маллера, а идя по второму — коды Гоппы.

4.2.1. Определение

Зафиксируем параметры (r, m), где $r \le m$. Положим $q := 2$ и рассмотрим многочлены от m переменных степени $\le r$. Базис в пространстве $\{P \in \mathbb{F}_2[x_1, \dots, x_m], \deg P \le r\}$:

$$\{1\} \cup \{x_1, x_2, \dots, x_m\} \cup \{x_1x_2, x_1x_3, \dots, x_{m-1}x_m\} \cup \big\{x_{i_1}x_{i_2}\cdots x_{i_r} \mid 1 \le i_1, \dots, i_r \le m\big\}.$$

Размерность этого пространства равна $k = \sum_{t\le r}\binom{m}{t}$.

Рассмотрим множество векторов значений многочленов во всех точках $\mathbb{F}_2^m$:

$$C := \{\big(P(0\dots00), P(0\dots01), \dots, P(1\dots11)\big), \text{где } P \in \mathbb{F}_2[x_1, \dots, x_m] \text{ и } \deg P \le r\}.$$

Множество C образует $[n, k, d]$-код, где $n = 2^m$ и $k = \sum_{t\le r}\binom{m}{t}$. Этот код называется *кодом Рида—Маллера* или *RM-кодом* с параметрами (r, m).

4.2.2. Кодовое расстояние

Чтобы оценить кодовое расстояние d, нам понадобится доказать лемму.

Лемма.
Если $P \in \mathbb{F}_2[x_1, \dots, x_m]$, $P \not\equiv 0$ и $\deg P \le r$, то

$$\#\{(s_1, \dots, s_m) \in \mathbb{F}_2^m \mid P(s_1, \dots, s_m) = 1\} \ge 2^{m-r}.$$

Доказательство индукцией по m.
База: $m = 1$. Тогда $P \in \{1, x_1, x_1 + 1\}$ — очевидно.
Переход: $m - 1 \to m$. Б.о.о. будем считать, что P существенно зависит от x_m. Распишем

$$P(x_1, \dots, x_m) = P_1(x_1, \dots x_{m-1}) + x_m P_2(x_1, \dots x_{m-1}).$$

Так как $P_2 \not\equiv 0$ и $\deg P_2 \le r - 1$, то

$$\#\{(s_1, \dots, s_{m-1}) \mid P_2(s_1, \dots, s_{m-1}) = 1\} \ge 2^{(m-1)-(r-1)} = 2^{m-r},$$

$$P(x_1, \dots, x_m) = P_1(x_1, \dots x_{m-1}) + x_m P_2(x_1, \dots x_{m-1}),$$

Каждый набор $(s_1, \dots, s_{m-1})$ на котором $P_2 = 1$, можно дополнить до набора, на котором $P = 1$:

- если $P_1(s_1, \dots, s_{m-1}) = 0$, то возьмём набор $(s_1, \dots, s_{m-1}, 1)$,
- если $P_1(s_1, \dots, s_{m-1}) = 1$, то возьмём набор $(s_1, \dots, s_{m-1}, 0)$.

Значит, $P = 1$ не менее чем на 2^{m-r} наборах.

Лемма доказана.

Из леммы следует, что РМ-код с параметрами (r, m) является $\left[2^m, \sum_{t \le r} \binom{m}{t}, 2^{m-r}\right]$-кодом.

4.2.3. Декодирование RM-кодов

Так как кодовое расстояние RM-кода равно 2^{m-r}, то он способен исправлять вплоть до $(2^{m-r-1} - 1)$ ошибок. Оказывается, это можно делать очень быстро *многоэтапным голосованием* (этот способ декодирования также называют *мажоритарным*).

Постановка задачи: в векторе из RM-кода (т.е. векторе значений многочлена степени $\le r$) изменяются менее 2^{m-r-1} координат (т.е. значение многочлена искажается менее чем в стольких точках), нужно восстановить по искажённому вектору значений исходный вектор значений (т.е. исходный многочлен).

По определению кодовое слово RM-кода — это значения многочлена, выразимого линейной комбинацией в базисе

$$\{1\} \cup \{x_1, x_2, \dots, x_m\} \cup \dots \cup \left\{x_{i_1} x_{i_2} \cdots x_{i_r} \mid 1 \le i_1, \dots, i_r \le m\right\}.$$

Восстановить кодовое слово — это то же, что найти коэффициенты этой линейной комбинации.

Лемма.
Для любого $P \in \mathbb{F}_2[x_1, \dots, x_r]$ справедлива формула

$$\operatorname{coef}_{x_1 \cdot \dots \cdot x_r} P = \sum_{\alpha_1, \dots, \alpha_r \in \boldsymbol{F}_2} P(\alpha_1, \dots, \alpha_r).$$

Доказательство. Многочлен P можно представить в виде

$$P = c \cdot x_1 \dots x_r + P_1 + \dots + P_r,$$

где в P_i *не* входит x_i.

Рассмотрим сумму

$$\sum_{\alpha_1, \dots, \alpha_r \in \mathbb{F}_2} P_i(\alpha_1, \dots, \alpha_r) = \sum_{\alpha_1, \dots \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_r \in \mathbb{F}_2} \left(P_i(\alpha_1, \dots \alpha_{i-1}, 0, \alpha_{i+1}, \dots, \alpha_r) + P_i(\alpha_1, \dots \alpha_{i-1}, 1, \alpha_{i+1}, \dots, \alpha_r)\right).$$

Так как в слагаемые многочлена P_i переменная x_i не входит, то $P_i(\alpha_1, \dots \alpha_{i-1}, 0, \alpha_{i+1}, \dots, \alpha_r) = P_i(\alpha_1, \dots \alpha_{i-1}, 1, \alpha_{i+1}, \dots, \alpha_r)$, а значит, по модулю 2 каждое слагаемое в последней сумме равно нулю. Отсюда

$$\sum_{\alpha_1, \dots, \alpha_r \in \mathbb{F}_2} P(\alpha_1, \dots, \alpha_r) = \underbrace{\sum_{\alpha_1, \dots, \alpha_r \in \mathbb{F}_2} c \cdot \alpha_1 \dots \alpha_r}_{=c} + \sum_{1 \le i \le r} \underbrace{\sum_{\alpha_1, \dots, \alpha_r \in \mathbb{F}_2} P_i(\alpha_1, \dots, \alpha_r)}_{=0} = c.$$

Лемма доказана.

Пусть P — произвольный многочлен из RM-кода. Зафиксируем произвольные $\beta_1, \dots, \beta_{m-r} \in \mathbb{F}_2$ и положим

$$P_{\beta_1,\dots,\beta_{m-r}} \coloneqq P(x_1, \dots, x_r, \beta_1, \dots, \beta_{m-r}).$$

Имеем $P_{\beta_1,\dots,\beta_{m-r}} \in \mathbb{F}_2[x_1, \dots, x_r]$, и по только что доказанной лемме мы получаем

$$\sum_{\alpha_1,\dots,\alpha_r \in \mathbb{F}_2} P_{\beta_1,\dots,\beta_{m-r}}(\alpha_1, \dots, \alpha_r) = \operatorname{coef}_{x_1 \cdot \dots \cdot x_r} P_{\beta_1,\dots,\beta_{m-r}} = \operatorname{coef}_{x_1 \cdot \dots \cdot x_r} P.$$

Если нам дано кодовое слово с не более чем $(2^{m-r-1} - 1)$ ошибками, это означает, что нам дан некий набор величин $\hat{P}(t_1, \dots, t_m)$, где $\hat{P}(t_1, \dots, t_m) = P(t_1, \dots, t_m)$ для всех $(t_1, \dots, t_m) \in \mathbb{F}_2^m \setminus T_{\text{bad}}$, где $|T_{\text{bad}}| \leq 2^{m-r-1} - 1$. Подставим $\hat{P}(t_1, \dots, t_m)$ вместо $P(t_1, \dots, t_m)$ в нашу формулу. Для каждого набора $(\beta_1, \dots, \beta_{m-r}) \in \mathbb{F}_2^{m-r}$ рассмотрим сумму

$$\hat{S}_{\beta_1,\dots,\beta_{m-r}} \coloneqq \sum_{\alpha_1,\dots,\alpha_r \in \mathbb{F}_2} \hat{P}(\alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_{m-r}).$$

У сумм $\hat{S}_{\beta_1,\dots,\beta_{m-r}}$ при разных $(\beta_1, \dots, \beta_{m-r})$ нет общих слагаемых. Поэтому $\hat{S}_{\beta_1,\dots,\beta_{m-r}} = \operatorname{coef}_{x_1 \cdot \dots \cdot x_r} P$ для всех $(\beta_1, \dots, \beta_{m-r}) \in \mathbb{F}_2^{m-r}$, кроме, быть может, $|T_{\text{bad}}|$ штук. Всего сумм 2^{m-r}, и $|T_{\text{bad}}| < 2^{m-r-1}$, а значит, *большинство* этих сумм равны $\operatorname{coef}_{x_1 \cdot \dots \cdot x_r} P$.

В итоге способ нахождения $\operatorname{coef}_{x_1 \cdot \dots \cdot x_r} P$ таков:

- для каждого $(\beta_1, \dots, \beta_{m-r}) \in \mathbb{F}_2^{m-r}$ вычисляем соответствующую сумму $\hat{S}_{\beta_1,\dots,\beta_{m-r}}$,
- находим $\operatorname{coef}_{x_1 \cdot \dots \cdot x_r} P$ *голосованием*, то есть как то значение, которое встречается чаще всего среди $\{\hat{S}_{\beta_1,\dots,\beta_{m-r}}\}$.

Ясно, что так можно определить *любой* из коэффициентов $\operatorname{coef}_{x_{i_1} \cdot \dots \cdot x_{i_r}} P$.

Пусть уже найдены все $\operatorname{coef}_{x_{i_1} \cdot \dots \cdot x_{i_r}} P$. Рассмотрим многочлен

$$P_{[r-1]} \coloneqq P - \sum_{i_1,\dots,i_r} \left(\operatorname{coef}_{x_{i_1} \cdot \dots \cdot x_{i_r}} P\right) \cdot x_{i_1} \dots x_{i_r}.$$

В $P_{[r-1]}$ уже все слагаемые степени $\leq r - 1$. Рассмотрим величины $\hat{P}_{[r-1]}(t_1, \dots, t_m)$, равные

$$\hat{P}(t_1, \dots, t_m) - \sum_{i_1,\dots,i_r} \left(\operatorname{coef}_{x_{i_1} \cdot \dots \cdot x_{i_r}} P\right) \cdot t_{i_1} \dots t_{i_r}.$$

Так как $\hat{P}_{[r-1]}(t_1, \dots, t_m) = P_{[r-1]}(t_1, \dots, t_m)$ на множестве $\mathbb{F}_2^m \setminus T_{\text{bad}}$, то, как и ранее, голосованием можно определить $\operatorname{coef}_{x_{i_1} \cdot \dots \cdot x_{i_{r-1}}} P_{[r-1]}$. Но $\operatorname{coef}_{x_{i_1} \cdot \dots \cdot x_{i_{r-1}}} P_{[r-1]} = \operatorname{coef}_{x_{i_1} \cdot \dots \cdot x_{i_{r-1}}} P$, то есть теперь мы знаем коэффициенты при слагаемых P степени $\leq r - 1$. Действуя дальше в том же духе, можно найти и остальные коэффициенты P.

Общая схема декодирования RM-кодов:

- определяем все $\operatorname{coef}_{x_{i_1} \cdot \dots \cdot x_{i_r}} P$,
- рассматриваем $P_{[r-1]} \coloneqq P - \sum_{i_1,\dots,i_r} \left(\operatorname{coef}_{x_{i_1} \cdot \dots \cdot x_{i_r}} P\right) \cdot x_{i_1} \dots x_{i_r}$,
- определяем все $\operatorname{coef}_{x_{i_1} \cdot \dots \cdot x_{i_{r-1}}} P$,
- рассматриваем $P_{[r-2]} \coloneqq P_{[r-1]} - \sum_{i_1,\dots,i_{r-1}} \left(\operatorname{coef}_{x_{i_1} \cdot \dots \cdot x_{i_{r-1}}} P\right) \cdot x_{i_1} \dots x_{i_{r-1}}$,
- определяем все $\operatorname{coef}_{x_{i_1} \cdot \dots \cdot x_{i_{r-2}}} P$,
- и так далее…

4.3. Понятие об алгеброгеометрических кодах (кодах В.Д. Гоппы)

Код Рида—Соломона определяется так: $\{(P(t_1), \dots, P(t_n)) \mid P \in \mathbb{F}_q[x] \wedge \deg P < k\}$. Идея оценки кодового расстояния RS-кодов состояла в том, что многочлен маленькой степени имеет мало нулей. Эти соображения можно уточнить: во-первых, можно тщательно выбрать множество точек $\{t_1, \dots, t_n\}$, в которых вычисляется значение P, и во-вторых, можно брать не всевозможные многочлены ограниченной степени, а специально выбранное их подмножество.

Идея: в качестве точек $t_1, \dots, t_n$ брать нули некоторого многочлена P_{base} небольшой степени, а в качестве многочленов, по которым строится C, брать многочлены, имеющие мало общих нулей с P_{base}. Реализуя полученные идеи, мы приходим к *кодам Гоппы*. Ниже мы лишь вскользь коснёмся этой красивой конструкции, рассмотрев конкретный пример.

Будем работать в конечном поле $\mathbb{F}_{13}$, взяв $P_{base} \coloneqq y^2 - 2x^3 + 2x$. Множество нулей этого многочлена:

$S_{base} = \{(0,0), (\pm 1;\ 0), (2;\pm 5), (3;\pm 3), (4;\pm 4), (6;\pm 2), (7;\pm 3), (9;\pm 6), (10;\pm 2), (11;\pm 1)\}$,
$q \coloneqq 13$,
$n \coloneqq |S_{base}| = 19$.

Рассмотрим множество многочленов

$$\tilde{P} \coloneqq \{\alpha_1 + \alpha_2 x + \alpha_3 x^2 + \alpha_4 x^3 + \alpha_5 y + \alpha_6 xy\}.$$

Читатель может самостоятельно доказать, что если $P \in \tilde{P}$ и $P \not\equiv 0$, то у многочленов P и P_{base} не больше шести общих нулей. Отсюда вытекает, что множество

$$C \coloneqq \left\{ \left(P(x_0, y_0)\right)_{(x_0,y_0)\in S_{\text{base}}} \mid P \in \tilde{P} \right\}$$

является $[19,6,13]_{13}$-кодом.

Сравнение с конструкцией Рида—Соломона

Чтобы с помощью конструкции Рида—Соломона получить $k \geq 6$ и $d \geq 13$, пришлось бы взять $q \geq n \geq k + d - 1 \geq 18$, и это дало бы $[18,6,13]_{19}$-код или $[19,6,14]_{19}$-код. То есть мы выгадали бы единицу в длине слов или расстоянии, но проиграли бы в мощности алфавита в полтора раза.

5. Циклические коды

5.1. Определение

Циклический код — это линейный код, такой, что для любого кодового слова $(a_0, a_1 \dots, a_{n-1})$ слово $(a_{n-1}, a_0 \dots, a_{n-2})$ также является кодовым. Из определения следует, что тогда *любой* циклический сдвиг кодового слова также является кодовым словом.

Например, $[7,4,3]$-код Хемминга эквивалентен циклическому коду с проверочной матрицей

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

Сопоставим слову $(a_0, \dots, a_{n-1}) \in \mathbb{F}_q^n$ многочлен

$$f \coloneqq a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1} \in \mathbb{F}_q[x].$$

Тогда слову $(a_{n-1}, a_0 \dots, a_{n-2})$ отвечает многочлен

$$a_{n-1} + a_0 x + a_1 x^2 + \cdots + a_{n-2} x^{n-1} = x \cdot f - a_{n-1}(x^n - 1).$$

Перейдём в кольцо $\mathbb{F}_q[x]/(x^n - 1)$. Слову $(a_0, \dots, a_{n-1})$ отвечает элемент кольца

$$f = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1},$$

а слову $(a_{n-1}, a_0 \dots, a_{n-2})$ отвечает элемент

$$x \cdot f - a_{n-1}(x^n - 1) \equiv \{\text{в кольце}\} \equiv x \cdot f.$$

Вывод: циклический сдвиг слова эквивалентен умножению соответствующего многочлена на x в кольце $\mathbb{F}_q[x]/(x^n - 1)$.

Таким образом, можно сформулировать *алгебраическое определение циклических кодов*: *циклический код* — это подмножество C кольца $\mathbb{F}_q[x]/(x^n-1)$, такое, что

- $f_1, f_2 \in C \quad \Rightarrow \quad \forall \alpha, \beta \in \mathbb{F}_q \quad \alpha f_1 + \beta f_2 \in C$,
- $f \in C \quad \Rightarrow \quad x \cdot f \in C$.

Утверждение.
Для любого ц.к. $C \subseteq \mathbb{F}[x]/(x^n-1)$ выполнено

$$f \in C \quad \Rightarrow \quad \forall g \in \mathbb{F}[x]/(x^n-1) \quad f \cdot g \in C.$$

Доказательство: утверждение непосредственно следует из алгебраического определения ц.к.

5.2. Порождающий многочлен

Утверждение.
Любой циклический код $C \subseteq \mathbb{F}[x]/(x^n-1)$ может быть представлен в виде

$$\{f \cdot g \quad | \quad f \in \mathbb{F}[x]/(x^n-1)\}$$

для некоторого фиксированного многочлена g.

Доказательство. Пусть C — ц.к. Рассмотрим $g_0 \in C$, такой, что

$$\deg g_0 = \min_{\substack{g \in C \\ g \neq 0}} \deg g$$

Тогда любой многочлен $f \in C$ кратен g_0. Действительно, поделим f на g_0 с остатком:

$$f(x) = g_0(x) \cdot \tilde{f}(x) + r(x),$$

где $\deg r < \deg g_0$.

Но $r = f - \tilde{f} \cdot g_0 \in C$, а значит $r \equiv 0$.

Утверждение доказано.

Нормированный многочлен — это многочлен с коэффициентом 1 при мономе старшей степени.

Утверждение.
В любом ц.к. ненулевой нормированный многочлен минимальной степени единственен. Этот многочлен называется *порождающим многочленом* циклического кода.

Доказательство. Допустим, что в коде C нашлись два разных нормногочлена минимальной степени:

$$g_1(x) = x^l + \cdots \; g_2(x) = x^l + \cdots$$

Но тогда $(g_1 - g_2) \in C$ и $\deg(g_1 - g_2) < l$ — это противоречит минимальности l.

Теорема.
Нормногочлен $g \in \mathbb{F}[x]/(x^n-1)$ может быть порождающим многочленом циклического кода т. и т.т., когда он является делителем многочлена (x^n-1) в кольце $\mathbb{F}[x]$.

Доказательство $g|(x^n-1) \quad \Rightarrow \quad \exists$ ц. к.:

Пусть $g(x) \mid (x^n-1)$. Положим $C \coloneqq \{fg, \quad \text{где } f \in \mathbb{F}[x]/(x^n-1)\}$. Очевидно, C — циклический код. Осталось доказать, что g — порождающий многочлен кода C, то есть что в C любой ненулевой многочлен имеет степень $> \deg g$, либо равен $\text{const} \cdot g$.

Рассмотрим произвольный многочлен $\tilde{g} \in C$. Имеем $\tilde{g} = fg$ для некоторого $\mathbb{F}[x]/(x^n-1)$. Тогда в кольце $\mathbb{F}[x]$ для тех же самых f и $\tilde{g}$ и некоторого s выполнено равенство $\tilde{g} = f \cdot g + s \cdot (x^n-1)$. По условию, $(x^n-1) = r \cdot g$ для некоторого $r \in \mathbb{F}[x]$, следовательно

$$\tilde{g} = f \cdot g + sr \cdot g = (f + sr) \cdot g.$$

Итак, в кольце $\mathbb{F}[x]$ для некоторых f, r, s имеем $\tilde{g} = (f + sr) \cdot g$. Возможны случаи:

- $(f + sr) \equiv 0$ — тогда $\tilde{g} \equiv 0$,
- $(f + sr) \equiv \text{const} \neq 0$ — тогда $\tilde{g} = \text{const} \cdot g$,

- $\deg(f + sr) \geq 1$ — тогда $\deg \tilde{g} > \deg g$.

Доказательство ∃ц. к. $\Rightarrow$ $g|(x^n - 1)$*:*
Пусть C — циклический код в $\mathbb{F}[x]/(x^n - 1)$ с порождающим многочленом g. Поделим в кольце $\mathbb{F}[x]$ с остатком $(x^n - 1)$ на g: $x^n - 1 = f \cdot g + r$, где $\deg r < \deg g$. Тогда в кольце $\mathbb{F}[x]/(x^n - 1)$ имеем $r = (-f) \cdot g \in C$. Отсюда $r \equiv 0$, то есть $g|(x^n - 1)$.

5.3. Порождающая и проверочная матрицы

Утверждение.
Пусть порождающий многочлен циклического кода $C \subseteq \mathbb{F}_q[x]/(x^n - 1)$ имеет вид

$$c_0 + c_1 x + \cdots + c_{\alpha-1} x^{\alpha-1} + x^{\alpha}.$$

Тогда если рассматривать C как подпространство $\mathbb{F}_q^n$, то $\dim C = n - \alpha$ и порождающая матрица кода имеет вид

$$G \coloneqq \begin{pmatrix} c_0 & c_1 & \dots & c_{\alpha-1} & 1 & 0 & \dots & \dots & 0 \\ 0 & c_0 & c_1 & \dots & c_{\alpha-1} & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & c_0 & c_1 & \dots & c_{\alpha-1} & 1 \end{pmatrix}.$$

Доказательство. Очевидно, что строки матрицы G линейно независимы и её ранг равен $(n - \alpha)$. Остаётся доказать равенство $\dim C = n - \alpha$.

Для произвольного $f \in \mathbb{F}_q[x]/(x^n - 1)$ положим $C_f \coloneqq \{f + h \mid h \in C\}$. Докажем, что если $f_1 \neq f_2$ и $\deg f_i < \alpha$, то $C_{f_1} \cap C_{f_2} = \emptyset$.

Допустим, что $C_{f_1} \cap C_{f_2} \neq \emptyset$. Это означает, что $f_1 + h_1 = f_2 + h_2$ для некоторых $h_1, h_2 \in C$. Тогда $f_1 - f_2 = h_2 - h_1 \in C$. Но тогда из условия $\deg(f_1 - f_2) < \alpha$ вытекает, что $f_1 - f_2 \equiv 0$ — противоречие.

Пусть

$$f_1, \dots, f_{q^\alpha} \in \mathbb{F}_q[x]/(x^n - 1)$$

— всевозможные многочлены степени $< \alpha$.

Так как $C_{f_i} \cap C_{f_j} = \emptyset$ при $i \neq j$, то

$$\left|C_{f_1}\right| + \cdots + \left|C_{f_{q^\alpha}}\right| \leq \left|\mathbb{F}_q[x]/(x^n - 1)\right| = q^n.$$

Очевидно, $\left|C_{f_i}\right| = |C|$ для каждого i, а значит, $|C| \leq \frac{q^n}{q^\alpha} = q^{n-\alpha}$. Следовательно, $\dim C \leq n - \alpha$.

Утверждение доказано.

Следствие.
Код C можно представить в виде

$$\{f \cdot g \quad | \quad f \in \mathbb{F}[x]/(x^n - 1), \quad \deg f < n - \alpha\}.$$

Утверждение.
У любого циклического кода существует порождающая матрица канонического вида, то есть любой циклический код допускает систематическое кодирование. Важно, что *сам* код допускает систематическое кодирование; не нужно переходить к эквивалентному коду.

Доказательство. Пусть g — порождающий многочлен, $\deg g = \alpha$. Поделим многочлены $x^\alpha, x^{\alpha+1}, \dots, x^{n-1}$ с остатком на g:

$$\begin{aligned} x^{\alpha} &= h_0 \cdot g + r_0 \\ &\vdots \\ x^{n-1} &= h_{n-\alpha-1} \cdot g + r_{n-\alpha-1} \end{aligned}$$

Перепишем эти равенства:

$$\begin{aligned} h_0 \cdot g &= x^{\alpha} - r_0 \\ &\vdots \\ h_{n-\alpha-1} \cdot g &= x^{n-1} - r_{n-\alpha-1} \end{aligned}$$

Каждый из многочленов $h_i \cdot g$ принадлежит C и имеет вид

$$x^{\alpha+i} + c_{i,\alpha-1}x^{\alpha-1} + c_{i,\alpha-2}x^{\alpha-2} + \cdots + c_{i,0},$$

где $c_{i,j}$ — некоторые коэффициенты.

Многочлены $h_i \cdot g$ принадлежат C и имеют вид $x^{\alpha+i} + c_{i,\alpha-1}x^{\alpha-1} + c_{i,\alpha-2}x^{\alpha-2} + \cdots + c_{i,0}$.

Составим из их коэффициентов порождающую матрицу кода C, имеющую вид

$$\begin{pmatrix} c_{0,0} & \dots & c_{0,\alpha-1} & 1 & 0 & \dots & 0 \\ c_{1,0} & \dots & c_{1,\alpha-1} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & & \ddots & \\ c_{n-\alpha-1,0} & \dots & c_{n-\alpha-1,\alpha-1} & 0 & 0 & \dots & 1 \end{pmatrix}.$$

Т.к. код C циклический, то можно циклически переставить столбцы в этой матрице, и получится искомая матрица вида $(I|\tilde{G})$, где I — единичная матрица порядка $(n-\alpha)$.

Утверждение доказано.

Пусть g — порождающий многочлен кода C. Так как $g|(x^n-1)$, то в кольце $\mathbb{F}[x]$ имеем $x^n - 1 = g \cdot h$ для некоторого $h \in \mathbb{F}[x]$. Многочлен $h(x)$ называется *проверочным многочленом* кода C. Для любого $f \in C$ в кольце $\mathbb{F}[x]/(x^n-1)$ выполнено равенство $f \cdot h = 0$.

Утверждение.
Пусть проверочный многочлен циклического кода $C \subseteq \mathbb{F}[x]/(x^n-1)$ имеет вид

$$h_0 + h_1x + \cdots + h_{n-\alpha}x^{n-\alpha}.$$

Тогда если рассматривать C как обычный линейный код, то его проверочная матрица будет иметь вид

$$\begin{pmatrix} h_{n-\alpha} & \dots & h_1 & h_0 & 0 & \dots & \dots & 0 \\ 0 & h_{n-\alpha} & \dots & h_1 & h_0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & h_{n-\alpha} & \dots & h_1 & h_0 \end{pmatrix}.$$

Доказательство. Пусть проверочный многочлен циклического кода $C \subseteq \mathbb{F}[x]/(x^n-1)$ имеет вид

$$h_0 + h_1x + \cdots + h_{n-\alpha}x^{n-\alpha}.$$

Для удобства формально введём $h_{n-\alpha+1} = h_{n-\alpha+2} = \cdots = 0$. В кольце $\mathbb{F}[x]/(x^n-1)$ выполнены равенства

$$0 = (c_0 + c_1x + \cdots + c_{n-1}x^{n-1}) \cdot (h_0 + h_1x + \cdots + h_{n-1}x^{n-1}) = \sum_{m=0}^{2n-2} x^m \sum_{i=0}^{m} c_i h_{m-i}.$$

В $\mathbb{F}[x]/(x^n-1)$ выполнено $x^{n+t} = x^t$, отсюда

$$0 = \sum_{m=0}^{2n-2} x^m \sum_{i=0}^{m} c_i h_{m-i} = \sum_{m=0}^{n-1} x^m \sum_{i=0}^{m} c_i h_{m-i} + \sum_{m=0}^{n-1} x^m \sum_{i=m+1}^{n-1} c_i h_{m+n-i}.$$

Следовательно при каждом $m \in \{0, \dots, n-1\}$ должно быть выполнено

$$\sum_{i=0}^{m} c_i h_{m-i} + \sum_{i=m+1}^{n-1} c_i h_{m+n-i} = \sum_{i=0}^{n-1} c_i h_{(m-i) \bmod n} = 0.$$

При $m \in \{n-\alpha, \dots, n-1\}$ уравнения

$$\sum_{i=0}^{n-1} c_i h_{(m-i) \bmod n} = 0$$

как раз и задаются матрицей

$$\begin{pmatrix} h_{n-\alpha} & \dots & h_1 & h_0 & 0 & \dots & \dots & 0 \\ 0 & h_{n-\alpha} & \dots & h_1 & h_0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & h_{n-\alpha} & \dots & h_1 & h_0 \end{pmatrix}.$$

То, что эта матрица проверочная (т.е. никакие «лишние» слова не удовлетворяют системе), следует из того, что её ранг равен α, а размерность кода равна $(n-\alpha)$.

Утверждение доказано.

5.4. Граница БЧХ и коды БЧХ

5.4.1. Граница БЧХ

Формула Вандермонда.

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_r \\ \vdots & \vdots & \vdots & \vdots \\ \lambda_1^{r-1} & \lambda_2^{r-1} & \cdots & \lambda_r^{r-1} \end{vmatrix} = \prod_{1\le i<j\le r} (\lambda_j - \lambda_i).$$

Доказательство индукцией по r. База $r=1$ очевидна: $\begin{vmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{vmatrix} = \lambda_2 - \lambda_1$.

Индуктивный переход:

$$\begin{vmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_r \\ \vdots & \vdots & \cdots & \vdots \\ \lambda_1^{r-1} & \lambda_2^{r-1} & \cdots & \lambda_r^{r-1} \end{vmatrix} = \begin{vmatrix} 1 & 0 & \cdots & 0 \\ \lambda_1 & \lambda_2-\lambda_1 & \cdots & \lambda_r-\lambda_1 \\ \vdots & \vdots & \cdots & \vdots \\ \lambda_1^{r-1} & \lambda_2^{r-1}-\lambda_1^{r-1} & \cdots & \lambda_r^{r-1}-\lambda_1^{r-1} \end{vmatrix} = \begin{vmatrix} \lambda_2-\lambda_1 & \cdots & \lambda_r-\lambda_1 \\ \lambda_2^2-\lambda_1^2 & \cdots & \lambda_r^2-\lambda_1^2 \\ \vdots & \cdots & \vdots \\ \lambda_2^{r-2}-\lambda_1^{r-2} & \cdots & \lambda_r^{r-2}-\lambda_1^{r-2} \\ \lambda_2^{r-1}-\lambda_1^{r-1} & \cdots & \lambda_r^{r-1}-\lambda_1^{r-1} \end{vmatrix}$$

$$= \begin{vmatrix} \lambda_2-\lambda_1 & \cdots & \lambda_r-\lambda_1 \\ \lambda_2^2-\lambda_1\lambda_2 & \cdots & \lambda_r^2-\lambda_1\lambda_r \\ \vdots & \cdots & \vdots \\ \lambda_2^{r-2}-\lambda_1\lambda_2^{r-3} & \cdots & \lambda_r^{r-2}-\lambda_1\lambda_r^{r-3} \\ \lambda_2^{r-1}-\lambda_1\lambda_2^{r-2} & \cdots & \lambda_r^{r-1}-\lambda_1\lambda_r^{r-2} \end{vmatrix} = \left(\prod_{i=2}^{r}(\lambda_i-\lambda_1)\right)\cdot \det \begin{vmatrix} 1 & 1 & \cdots & 1 \\ \lambda_2 & \lambda_3 & \cdots & \lambda_r \\ \vdots & \vdots & \cdots & \vdots \\ \lambda_2^{r-2} & \lambda_3^{r-2} & \cdots & \lambda_r^{r-2} \end{vmatrix}$$

$$= \prod_{1\le i<j\le r} (\lambda_j - \lambda_i).$$

В первом переходе вычитаем первый столбец из всех остальных. Во втором — раскладываем определитель по первой строке. В третьем — вычитаем из каждой строки предыдущую, домноженную на λ_1. Далее из каждого i-го столбца выносим множитель $(\lambda_{i+1}-\lambda_1)$.

Формула доказана.

Рассмотрим поле $\mathbb{F}_q$, где $q=p^m$, p простое. Известно, что множество $\mathbb{F}_q \setminus \{0\}$ образует циклическую группу по умножению. Каждый образующий элемент этой группы (порядок которого равен $(q-1)$) называется *примитивным элементом поля*. Иными словами, примитивный элемент — это такой $\lambda \in \mathbb{F}_q$, что $\{1, \lambda, \lambda^2, \dots, \lambda^{q-2}\} = \mathbb{F}_q \setminus \{0\}$.

Теорема. (A.Hocquenghem'1959, R.C. Bose & D.K. Ray-Chaudhuri'1960)
Пусть λ — примитивный элемент $\mathbb{F}_q$, и $\delta \le q$. Пусть порождающий многочлен g кода $C \subseteq \mathbb{F}_q^n$ таков, что в $\mathbb{F}_q$ среди его корней есть числа $\lambda^b, \lambda^{b+1}, \dots, \lambda^{b+\delta-2}$. Тогда $d(C) \ge \delta$.

Доказательство. Рассмотрим произвольный $f(x) \in C$. Найдётся многочлен $s(x) \in \mathbb{F}_q[x]/(x^n-1)$, такой, что $\deg s < n - \deg g$ и в кольце $\mathbb{F}_q[x]/(x^n-1)$ выполнено равенство $f(x) = s(x)\cdot g(x)$. Так как $\deg s + \deg g < n$, то это равенство выполнено и в кольце $\mathbb{F}_q[x]$. Пусть $\lambda^b, \dots, \lambda^{b+\delta-2}$ — различные корни $g(x)$. Они же будут корнями f. Пусть $f = c_0 + c_1 x + \dots + c_{n-1}x^{n-1}$. Вектор $(c_0, \dots, c_{n-1})$ удовлетворяет системе линейных уравнений с матрицей

$$\widetilde{H} = \begin{pmatrix} 1 & \lambda^{b} & \dots & \lambda^{b(n-1)} \\ 1 & \lambda^{b+1} & \dots & \lambda^{(b+1)(n-1)} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \lambda^{b+\delta-2} & \dots & \lambda^{(b+\delta-2)(n-1)} \end{pmatrix}.$$

Матрица $\widetilde{H}$ не обязательно проверочная матрица кода, но её можно дополнить до проверочной. Достаточно теперь доказать, что любые $(\delta-1)$ столбцов матрицы $\widetilde{H}$ линейно независимы. Выберем в $\widetilde{H}$ произвольные столбцы $i_1, \dots, i_{\delta-1}$. Получим матрицу $\widetilde{H}_{i_1,\dots,i_{\delta-1}}$, определитель которой равен

$$\det \widetilde{H}_{i_1,\dots,i_{\delta-1}} = \begin{vmatrix} \lambda^{b\cdot i_1} & \lambda^{b\cdot i_2} & \dots & \lambda^{b\cdot i_{\delta-1}} \\ \lambda^{(b+1)\cdot i_1} & \lambda^{(b+1)\cdot i_2} & \dots & \lambda^{(b+1)\cdot i_{\delta-1}} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda^{(b+\delta-2)\cdot i_1} & \lambda^{(b+\delta-2)\cdot i_2} & \dots & \lambda^{(b+\delta-2)\cdot i_{\delta-1}} \end{vmatrix}$$
$$= \lambda^{b\cdot(i_1+\dots+i_{\delta-1})} \begin{vmatrix} 1 & 1 & \cdots & 1 \\ \lambda^{i_1} & \lambda^{i_2} & \dots & \lambda^{i_{\delta-1}} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda^{(\delta-2)\cdot i_1} & \lambda^{(b-2)\cdot i_2} & \dots & \lambda^{(\delta-2)\cdot i_{\delta-1}} \end{vmatrix}.$$

Из формулы Вандермонда следует, что полученный определитель отличен от нуля.

Теорема доказана.

5.4.2. Коды БЧХ

Проблема: Если применять теорему БЧХ «в лоб», то невозможно доказать, что кодовое расстояние больше мощности кодового алфавита.
Решение: Код рассмотрим как подмножество в $\mathbb{F}_p^n$, но при применении границы БЧХ погрузим поле $\mathbb{F}_p$ в $\mathbb{F}_{p^m}$. При любом простом p и любом m поле $\mathbb{F}_p$ можно вложить как подполе в $\mathbb{F}_{p^m}$.

Обычно поле $\mathbb{F}_p$ — это поле вычетов $\mathbb{Z}_p$, а $\mathbb{F}_{p^m}$ строится как множество многочленов с коэффициентами из $\mathbb{Z}_p$, которые складываются и умножаются по модулю некоторого многочлена степени m, неприводимого над $\mathbb{Z}_p$. Тогда вложение $\mathbb{F}_p$ в $\mathbb{F}_{p^m}$ очевидно: элементам $\mathbb{F}_p$ соответствуют многочлены степени ≤ 0. Элементам $\mathbb{F}_{p^m}$ можно сопоставить векторы из $\mathbb{F}_p^m$, так, что сумме элементов $\mathbb{F}_{p^m}$ соответствует сумма векторов в $\mathbb{F}_p^m$.

Раз $\mathbb{F}_{p^m}$ — многочлены с коэффициентами из $\mathbb{F}_p$ степени $\leq m$, то каждому элементу $\mathbb{F}_{p^m}$ сопоставим вектор коэффициентов многочлена.

Пусть p простое. Рассмотрим циклический код $C \subseteq \mathbb{F}_p[x]/(x^n-1)$ с порождающим многочленом g. Коэффициенты g берутся из $\mathbb{F}_p$, но их можно считать одновременно элементами $\mathbb{F}_{p^m}$. Рассмотрим код $\tilde{C} \subseteq \mathbb{F}_{p^m}[x]/(x^n-1)$, порождённый многочленом g (если считать, что $g \in \mathbb{F}_{p^m}[x]/(x^n-1)$). Можно в том же духе считать, что $C \subseteq \tilde{C}$. Коэффициенты g берутся из $\mathbb{F}_p$, но их можно считать одновременно элементами $\mathbb{F}_{p^m}$. Код $\tilde{C} \subseteq \mathbb{F}_{p^m}[x]/(x^n-1)$ порождён g. Пусть λ — примитивный элемент $\mathbb{F}_{p^m}$, и $g(\lambda^b) = \dots = g\left(\lambda^{b+\delta-2}\right) = 0$. Тогда граница БЧХ гласит: $d\left(\tilde{C}\right) \geq \delta$. Так как $C \subseteq \tilde{C}$, то и $d(C) \geq \delta$. Итак, предлагается действовать по такой схеме:

- Подбираем $g \in \mathbb{F}_{p^m}[x]/(x^n-1)$ с коэффициентами из $\mathbb{F}_p$, так, чтобы в $\mathbb{F}_{p^m}$ было выполнено $g(\lambda^b) = \dots = g\left(\lambda^{b+\delta-2}\right) = 0$ и $g|(x^n-1)$.
- На основе g строим циклический код в $\mathbb{F}_p[x]/(x^n-1)$, для которого, по теореме БЧХ, выполнено $d(C) \geq \delta$.

Возникают вопросы: существует ли вообще такой g, и если существует, то как оценить $\dim C$? К счастью, теория конечных полей позволяет выбрать «хороший» g.

Утверждение (без доказательства).
Для любого $\alpha \in \mathbb{F}_{p^m} \setminus \{0\}$ существует многочлен $f \in \mathbb{F}_{p^m}[x] \setminus \{\mathbf{0}\}$ с коэффициентами из $\mathbb{F}_p$,

для которого $f(\alpha) = 0$. Если взять такой f минимальной степени, то f неприводим над $\mathbb{F}_p$, $\deg f \leq m$, и $f \mid \left(x^{p^m-1} - 1\right)$. Такой f называется *минимальным многочленом* элемента α.

Следствие.
Если $n = p^m - 1$, то существует g, такой, что$g(\lambda^b) = \cdots = g\left(\lambda^{b+\delta-2}\right) = 0$ и $g|(x^n - 1)$, причём $\deg g \leq (\delta - 1)m$. Такой g можно взять как $\mathrm{LCM}\left(f_{\lambda^b}, \dots, f_{\lambda^{b+\delta-2}}\right)$, где $f_{\lambda^b}, \dots, f_{\lambda^{b+\delta-2}}$ — минимальные многочлены соответствующих элементов. Через LCM обозначаем наименьшее общее кратное многочленов.

Окончательный способ построения БЧХ-кода:

- Берём $n \coloneqq p^m - 1$ и $g \coloneqq \mathrm{LCM}\left(f_{\lambda^b}, \dots, f_{\lambda^{b+\delta-2}}\right)$, где $f_{\lambda^b}, \dots, f_{\lambda^{b+\delta-2}}$ — минимальные многочлены соответствующих элементов в $\mathbb{F}_{p^m}$.
- Строим циклический код в $\mathbb{F}_p$, используя g в качестве порождающего многочлена наименьшей возможной степени.

Получаем код с параметрами $[p^m - 1, k, d]_p$, где $k = n - \deg g \geq n - (\delta - 1)m$ и $d \geq \delta$.

5.5. Циклическое представление кодов Хемминга

Пусть λ — примитивный элемент поля $\mathbb{F}_{2^m}$. Пусть g — минимальный многочлен для λ, и $C \subset \mathbb{F}_2^{2^m-1}$ — код, порождённый g. Любой кодовый многочлен $f \in C$ имеет корень, равный λ. Поэтому кодовые векторы $(c_0, \dots, c_{2^m-2})$ удовлетворяют соотношению

$$c_0 + c_1\lambda + c_2\lambda^2 + \cdots + c_{2^m-2}\lambda^{2^m-2} = 0.$$

Так как $\left\{\lambda^0, \lambda^1, \dots, \lambda^{2^m-2}\right\} = \mathbb{F}_{2^m} \setminus \{0\}$, и так как каждому элементу $\mathbb{F}_{2^m}$ отвечает вектор из $\mathbb{F}_2^m$, то проверочная матрица кода равна $\left(v_{\lambda^0}, v_{\lambda^1}, \dots, v_{\lambda^{2^m-2}}\right) \in \mathbb{F}_2^{m\times(2^m-1)}$, где v_{λ^i} — вектор, отвечающий λ^i. Столбцы матрицы — все ненулевые векторы $\mathbb{F}_2^m$.

Утверждение.
Двоичный код Хемминга с параметрами $[2^m - 1, 2^m - 1 - m, 3]$ эквивалентен циклическому коду, порождённому минимальным многочленом примитивного элемента $\mathbb{F}_{2^m}$.

5.6. Восстановление синхронизации для смежных классов циклических кодов

5.6.1. Задача восстановления синхронизации

До сих пор мы рассматривали коды, устойчивые к ошибкам замещения; в этом разделе опишем модель ошибок другого типа. Пусть по каналу передаются слова … $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ …:

$$a_1a_2 \dots a_nb_1b_2 \dots b_nc_1c_2 \dots c_n \dots$$

Если в канале выпадают символы, может произойти *ошибка синхронизации*:

$$a_{i+1}a_{i+2} \dots a_nb_1b_2 \dots b_nc_1c_2 \dots c_n \dots$$

и есть шанс неправильно разбить принятую последовательность на слова:

$$a_{i+1} \dots a_nb_1 \dots b_i \mid b_{i+1} \dots b_nc_1 \dots c_i \mid c_{i+1} \dots$$

Если при этом такие слова окажутся кодовыми, то мы далеко не сразу обнаружим ошибку!

Циклические коды очень плохие с точки зрения восстановления синхронизации: если $\boldsymbol{a} \in C$ и в канал передавалась последовательность $\boldsymbol{aaa}$ …, то при потере синхронизации мы обнаружим ошибку только в самом конце приёма.

Заметим, что если при приёме слова $\boldsymbol{b}$ «запоздать» на i тактов, то мы примем слово $b_{i+1} \dots b_nc_1 \dots c_i$, а если «забежать вперёд» на i тактов, примем $a_{n-i+1} \dots a_nb_1 \dots b_{n-i}$.

Код обладает *свободой от запятой степени* r, если для любых кодовых слов $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c}$ и любого $i \leq r$ коду не принадлежат слова $b_{i+1} \dots b_nc_1 \dots c_i$ и $a_{n-i+1} \dots a_nb_1 \dots b_{n-i}$. Если $r \geq \frac{n}{2}$, то считаем, что $r = \infty$, а код называется *кодом без запятой*. Если $r < \infty$, то при приёме можно (перебором) исправить синхросдвиг на $\leq \frac{r}{2}$ символов. Если $r = \infty$, то может быть исправлен любой синхросдвиг.

Циклические коды совершенно не способны исправлять ошибки потери синхронизации как раз в силу своего определения: циклический сдвиг любого кодового слова сам является кодовым словом. Зато оказывается, что смежные классы циклических кодов отлично справляются с задачей. Пусть $C \subseteq \mathbb{F}^n$ — линейный код. Его *смежный класс* — это множество вида $C + \boldsymbol{a} \coloneqq \{\boldsymbol{c} + \boldsymbol{a} \mid \boldsymbol{c} \in C\}$. Отметим, что при $\boldsymbol{a} \notin C$ этот код не является линейным.

Если определять циклические коды как идеалы в кольце многочленов, то для циклического кода с порождающим многочленом g смежный класс имеет вид

$$\{f \cdot g + s \mid f \in \mathbb{F}[x]/(x^n - 1)\}$$

для некоторого $s \in \mathbb{F}[x]/(x^n - 1)$.

Теорема.
При $s \equiv 1$ степень свободы от запятой смежного класса циклического кода равна $(\deg g - 1)$ при $n \geq 2 \deg g$ и равна ∞ при $n < 2 \deg g$.

Доказательство (подробно обоснуем только нижнюю оценку). Пусть передавались слова $\boldsymbol{a}, \boldsymbol{b} \in C$. Если при приёме слова $\boldsymbol{a}$ произошло запаздывание на i тактов, то принято будет слово-многочлен $x^{-i} \cdot (f_{\boldsymbol{a}} - t_1) + x^{n-i} \cdot t_2$, где t_1 и t_2 — многочлены, образованные i первыми координатами слов $\boldsymbol{a}$ и $\boldsymbol{b}$ соответственно, а $f_{\boldsymbol{a}}$ — многочлен, отвечающий слову $\boldsymbol{a}$.

В кольце $\mathbb{F}[x]/(x^n - 1)$ имеем

$$x^{-i} \cdot (f_{\boldsymbol{a}} - t_1) + x^{n-i} \cdot t_2 = x^{n-i} \cdot (f_{\boldsymbol{a}} - t_1 + t_2).$$

Аналогично, если при приёме слова $\boldsymbol{b}$ произошло «забегание вперёд», то будет принято слово $x^i \cdot f_{\boldsymbol{b}} + (t_3 - t_4)$, где t_3 и t_4 — многочлены, образованные i старшими разрядами слова $\boldsymbol{a}$ и i младшими разрядами $\boldsymbol{b}$ соответственно. При этом $\deg t_1, \deg t_2, \deg t_3, \deg t_4 < i$.

Итак, в случае рассинхронизации принятое слово будет иметь вид $x^{n-i} \cdot (\tilde{f} + \Delta)$ или $x^i \cdot \tilde{f} + \Delta$, где $\tilde{f} \in C$ и $\deg \Delta < i$. Нужно, чтобы $x^{n-i} \cdot (\tilde{f} + \Delta)$, $x^i \cdot \tilde{f} + \Delta \notin C$ при $0 < i \leq r$.

Пусть C — смежный класс ц.к. вида $\{f \cdot g + 1 \mid f \in \mathbb{F}[x]/(x^n - 1)\}$.

Нужно, чтобы для любых f_1, f_2, Δ, i, таких, что $\deg \Delta < i$ и $0 < i \leq r$ в кольце $\mathbb{F}[x]/(x^n - 1)$ было выполнено $f_1 \cdot g + 1 \neq x^{n-i} \cdot (f_2 \cdot g + 1 + \Delta)$ и $f_1 \cdot g + 1 \neq x^i \cdot (f_2 \cdot g + 1) + \Delta$. Это равносильно тому, что для любых f, Δ, i выполнено $f \cdot g \neq x^i + \Delta - 1$.

Пусть C — смежный класс ц.к. вида $\{f \cdot g + 1 \mid f \in \mathbb{F}[x]/(x^n - 1)\}$. Нужно, чтобы для любых f, Δ, i, таких, что $\deg \Delta < i$ и $0 < i \leq r$ в кольце $\mathbb{F}[x]/(x^n - 1)$ было выполнено $f \cdot g \neq x^i + \Delta - 1$.

Имеем $x^i + \Delta - 1 \neq 0$ при любом $i > 0$. Т.к. $\deg(x^i + \Delta - 1) = i \leq r$, то достаточно, чтобы $\deg g > r$. Это и требовалось доказать.

6. Совершенные коды

Напомним, что совершенными кодами называются коды, на которых достигается граница Хемминга. Эти коды интересны тем, что шары с центрами в кодовых словах не пересекаются и заполняют собой всё пространство слов. В этом разделе многие утверждения будут приведены без доказательства, но пропускать их не стоит, поскольку они имеют «мировоззренческое» значение.

Под *тривиальными* мы будем понимать следующие следующие совершенные коды:

- $(n, 1, n)_q$-код (состоит из одного слова),
- $(n, 2, n)_2$-код при нечётных n (пара слов-антиподов),
- $(n, q^n, 1)_q$-код (состоит из всех слов).

Для дальнейшего нам будет полезно отдельно сформулировать границу Хемминга для линейных кодов: для любого линейного $[n, k, d]_q$-кода $\left|S_{\lfloor (d-1)/2 \rfloor}(\mathbf{0})\right| \leq q^{n-k}$.
Таким образом, линейный $[n, k, d]_q$-код *совершенный*, если и только если $\left|S_{\lfloor (d-1)/2 \rfloor}(\mathbf{0})\right| = q^{n-k}$.

6.1. Совершенство кодов Голея и Хемминга

М. Голей (M. J. E. Golay) предложил два кода, и позже было замечено, что они циклические:

- $[23,12,7]$-код с порождающим многочленом $1 + x + x^5 + x^6 + x^7 + x^9 + x^{11}$,
- $[11,6,5]_3$-код с порождающим многочленом $2 + x^2 + 2x^3 + x^4 + x^5$.

Доказывать, что эти коды имеют именно такие параметры, мы не будем. По кодам Голея стандартным образом можно получить расширенные коды Голея:

- $[23,12,7]$-код $\to$ $[24,12,8]$-код,
- $[11,6,5]_3$-код $\to$ $[12,6,6]_3$-код.

Утверждение.
Коды Голея и Хемминга являются совершенными.

Доказательство. Утверждение доказывается непосредственной проверкой. Для $[23,12,7]$-кода Голея: $\sum_{i=0}^{3}\binom{23}{i} = 2^{23-12}$. Для $[11,6,5]_3$-кода Голея: $\sum_{i=0}^{2}\binom{n}{i}\cdot 2^i = 3^{11-6}$.

Совершенство двоичных кодов Хемминга мы уже доказывали ранее. Докажем его теперь в общем случае. Проверочная матрица q-ичного кода Хемминга содержит все линейно независимые столбцы высоты m. Получается $\left[n = \frac{q^m-1}{q-1}, k = n-m, 3\right]_q$-код. Проверяем соотношение $\left|S_{\lfloor (d-1)/2 \rfloor}(\mathbf{0})\right| = q^{n-k}$:

$$|S_1(\mathbf{0})| = 1 + n(q-1) = q^m = q^{n-k}.$$

Утверждение доказано.

6.2. Некоторые теоремы о совершенных кодах

Следующая теорема утверждает, что, по большому счёту, всё разнообразие кодов над конечными полями исчерпывается (с точки зрения параметров) тривиальными кодами и кодами Хемминга и Голея.

Теорема. (В. А. Зиновьев, В. К. Леонтьев ’1972, A. Tietäväinen ’1973, J. H. van Lint ’1971, M. R. Best ’1983, Y. Hong ’1983, V. Pless ’1968) — *Без доказательства.*

- Нетривиальных совершенных кодов с расстоянием > 7 не существует.
- Единственным с точностью до эквивалентности совершенным кодом с расстоянием 7 является $[23,12,7]$-код Голея.
- Любой нетривиальный код над алфавитом мощности $q = p^m$ с расстоянием ≤ 5 либо эквивалентен $[11,6,5]_3$-коду Голея, либо имеет те же длину слов, число слов и кодовое расстояние, что и $\left[\frac{q^t-1}{q-1}, \frac{q^t-1}{q-1} - t, 3\right]_q$-код Хемминга.

До сих пор неизвестно, существуют ли совершенные коды над алфавитами мощности $\neq p^m$ с кодовым расстоянием 3 или 5.

Следующая теорема Васильева утверждает, что, в отличие от кодов Голея, у кодов Хемминга есть множество «близнецов».

Теорема. (Ю. Л. Васильев ’1962, обобщения: J. Schonheim ’1968, B. Lindstrom ’1969)
Для любого $q = p^m$ существуют совершенные q-ичные коды с расстоянием 3, не эквивалентные линейным кодам.

Теорему Васильева в частном случае $q = 2$ мы докажем в следующем разделе.

Из теоремы Зиновьева—Леонтьева—Тиетавайнена следует, что для любого совершенного кода над $\mathbb{F}_{p^m}$ существует *линейный* код с той же длиной слов, числом слов и кодовым расстоянием. Однако не стоит думать, что линейные коды оказываются лучшими всегда, как подтверждают следующие результаты.

Теорема. (F. P. Preparata ’1968, J.-M. Goethals & S. L. Snover ’1972)

- Для любого $m \geq 2$ существует $\left(4^m, 2^{4^m-4m}, 6\right)$-код (коды с такими параметрами называют *кодами Препа́раты*).
- Любой линейный код длины 4^m с расстоянием 6 имеет меньшую мощность.

6.2.1. Доказательство теоремы Васильева при $q = 2$

Лемма. (Конструкция Васильева)
Пусть $C', C'' \subseteq \mathbb{F}_2^n$ — коды с расстояниями d' и d'', причём d' нечётно. Положим $\pi(\boldsymbol{a}) \coloneqq \sum_i a_i$. Для произвольного $\gamma\colon C'' \to \mathbb{F}_2$ рассмотрим код

$$C \coloneqq \left\{\left(\boldsymbol{c}' \mid \boldsymbol{c}' + \boldsymbol{c}'' \mid \pi(\boldsymbol{c}') + \gamma(\boldsymbol{c}'')\right), \text{ где } \boldsymbol{c}' \in C', \boldsymbol{c}'' \in C''\right\}.$$

Тогда C является $(2n+1, |C'| \cdot |C''|, d)$-кодом, где $d \geq \min\{2d'+1, d''\}$.

Доказательство. Нетривиальна только оценка $d(C)$ — её доказательством и займёмся. Возьмём пару различных слов кода C:

$$\boldsymbol{a} = \left(\boldsymbol{c}' \mid \boldsymbol{c}' + \boldsymbol{c}'' \mid \pi(\boldsymbol{c}') + \gamma(\boldsymbol{c}'')\right),$$
$$\widehat{\boldsymbol{a}} = \left(\hat{\boldsymbol{c}}' \mid \hat{\boldsymbol{c}}' + \hat{\boldsymbol{c}}'' \mid \pi(\hat{\boldsymbol{c}}') + \gamma(\hat{\boldsymbol{c}}'')\right).$$

Возможны случаи:

- $\boldsymbol{c}' \neq \hat{\boldsymbol{c}}'$ и $\boldsymbol{c}'' = \hat{\boldsymbol{c}}''$. Если $d(\boldsymbol{c}', \hat{\boldsymbol{c}}') = d'$, то $\pi(\boldsymbol{c}') \neq \pi(\hat{\boldsymbol{c}}')$, и, следовательно, $d(\boldsymbol{a}, \widehat{\boldsymbol{a}}) = 2d' + 1$. Если $d(\boldsymbol{c}', \hat{\boldsymbol{c}}') > d'$, то $d(\boldsymbol{a}, \widehat{\boldsymbol{a}}) > 2d'$.
- $\boldsymbol{c}' = \hat{\boldsymbol{c}}'$ и $\boldsymbol{c}'' \neq \hat{\boldsymbol{c}}''$. Тогда, очевидно, $d(\boldsymbol{a}, \widehat{\boldsymbol{a}}) \geq d(\boldsymbol{c}'', \hat{\boldsymbol{c}}'') \geq d''$.
- Остался случай $\boldsymbol{c}' \neq \hat{\boldsymbol{c}}'$ и $\boldsymbol{c}'' \neq \hat{\boldsymbol{c}}''$. Пусть $\boldsymbol{c}'$ и $\hat{\boldsymbol{c}}'$ отличаются на множестве позиций D_1, а $\boldsymbol{c}''$ и $\hat{\boldsymbol{c}}''$ отличаются на множестве позиций D_2. Тогда $\boldsymbol{c}' + \boldsymbol{c}''$ и $\hat{\boldsymbol{c}}' + \hat{\boldsymbol{c}}''$ отличаются по крайней мере на множестве $D_2 \setminus D_1$. Следовательно, $d(\boldsymbol{a}, \widehat{\boldsymbol{a}}) \geq |D_1| + |D_2 \setminus D_1| \geq |D_2| \geq d''$.

Лемма доказана.

Следствие из леммы Васильева.
Если существует $(n, M, 3)$-код, то существует $(2n+1, 2^n \cdot M, 3)$-код, не эквивалентный никакому линейному.

Доказательство: достаточно применить конструкцию Васильева с $C' \coloneqq \mathbb{F}_2^n$ и заметить, что если отображения γ и $(\gamma + 1)$ нелинейны, то получаемый с помощью конструкции код не эквивалентен никакому линейному.

Теорема Васильева для $q = 2$.
Для любого $m \geq 2$ существует совершенный $\left(2^m - 1, 2^{2^m-m-1}, 3\right)$-код, не эквивалентный никакому линейному.

Доказательство: заметим, что при каждом $m \geq 2$ есть $\left(2^{m-1} - 1, 2^{2^{m-1}-m}, 3\right)$-код Хемминга, и применим Следствие с $n \coloneqq 2^{m-1} - 1$ и $M \coloneqq 2^{2^{m-1}-m}$.

Теорема доказана.

7. Теоремы Шеннона о скорости кодирования

7.1. Шенноновская ёмкость графов

Напомним типы ошибок замещения:

- Односторонние ошибки. Например, если в двоичном канале возможны только замещения вида $0 \to 1$ или только $1 \to 0$. Канал связи в этом случае называется *несимметричным*.
- Двусторонние (симметричные) ошибки. Если возможно замещение символов $b_1 \to b_2$, то возможно и замещение $b_2 \to b_1$. Канал связи, в котором ошибки только симметричные, называется *симметричным*.

Можно задать модель канала графом, вершины которого — символы алфавита канала. Дуга идёт из x в y, если в канале возможна ошибка замещения x на y. Если в канале для каждой пары символов (x, y) возможны либо оба замещения $x \to y$, $y \to x$, либо ни одно из них, то граф канала можно считать неориентированным.

Пусть G — граф канала. Если считать каждый символ канала отдельным сообщением, то во избежание ошибок придётся использовать множество символов C_1, являющееся *независимым множеством* в G. При этом в лучшем случае пропускная способность канала получается такой же, как у канала без ошибок, алфавит которого имеет мощность $\alpha(G)$.

Если по каналу мы посылаем пары символов, то следует выбирать для использования такое множество пар C_2, чтобы

$$\nexists (x', y'), (x'', y'') \in C_2 \colon \big((x' = x'' \vee x'x'' \in E) \wedge (y' = y'' \vee y'y'' \in E)\big).$$

Если мы рассмотрим такое C_2, имеющее максимально возможную мощность, то данный код позволяет достичь пропускной способности $\sqrt{|C_2|}$. Смысл извлечения квадратного корня из $|C_2|$ в следующем: это скорость передачи данных в канале в расчёте на один передаваемый символ. Если бы у нас был *безошибочный* канал, в алфавите которого ровно $\sqrt{|C_2|}$ символов, то мы могли бы передавать по нему как раз $|C_2|$ сообщений длины 2.

Аналогично, если G — граф канала, и мы посылаем тройки символов, то следует выбирать такое множество троек C_3, чтобы в нём не было двух троек, соответствующие компоненты которых совпадают или образуют ребро в G. Если мы рассмотрим такое C_3, имеющее максимально возможную мощность, то данный код позволит достичь пропускной способности $\sqrt[3]{|C_3|}$. И так далее… Чтобы рассмотреть описанные конструкции в пределе, нужно ввести понятие произведения графов.

Шенноновское произведение графов G и H — это граф $G \times H$ со множеством вершин

$$V(G \times H) = \{(x, y) \mid x \in V(G), y \in V(H)\}.$$

Рёбра в $G \times H$ между парами (x', y') и (x'', y''), для которых

$$\big(x' = x'' \vee x'x'' \in E(G)\big) \wedge \big(y' = y'' \vee y'y'' \in E(H)\big).$$

Очевидно, $(G_1 \times G_2) \times G_3 \simeq G_1 \times (G_2 \times G_3)$ и $G_1 \times G_2 \simeq G_2 \times G_1$ для любых G_1, G_2, G_3.

Шенноновская ёмкость графа G определяется так:

$$\operatorname{cap}(G) = \sup_{n \in \boldsymbol{N}} \sqrt[n]{\alpha(G \times G \times \ldots \times G)},$$

где произведение берётся от n копий графа G. Это наибольшая возможная (в пределе) «скорость», с которой можно передавать без ошибок данные по каналу, моделью которого является граф G.

Пусть G — произвольный граф. Будем рассматривать наборы весов $\boldsymbol{w}$ на вершинах графа G, такие, что

- $\forall x\ w(x) \geq 0$,
- $\sum_{x \in V(G)} w(x) = 1$.

Для любого множества $A \subseteq G$ положим $w(A) \coloneqq \sum_{x \in A} w(x)$ и для произвольного графа G рассмотрим величину

$$\nu(G) \coloneqq \min_{\boldsymbol{w}} \max_{K \text{ — клика в } G} w(K).$$

Отметим, что максимум в формуле выше достаточно брать только по *максимальным по включению* кликам.

Теорема Шеннона о верхней оценке ёмкости:

Для любого G справедливо неравенство $\operatorname{cap}(G) \leq \big(\nu(G)\big)^{-1}$.

Доказательство. Докажем, что $\sqrt[n]{\alpha(G^n)} \le \big(\nu(G)\big)^{-1}$ для каждого n. Пусть $A \subseteq V(G)$ — независимое множество размера $\alpha(G)$. Рассмотрим набор весов $\boldsymbol{w}_0$, в котором

- $w_0(x) \coloneqq 0$, если $x \notin A$,
- $w_0(x) \coloneqq \big(\alpha(G)\big)^{-1}$, если $x \in A$.

Для таких весов получим

$$\nu(G) \le \max_{K \text{ — клика в } G} w_0(K) = \big(\alpha(G)\big)^{-1},$$

откуда $\alpha(G) \le \big(\nu(G)\big)^{-1}$.

Лемма. Для любых графов G, H выполнено $\nu(G \times H) = \nu(G) \cdot \nu(H)$.

Заметим, что если мы докажем эту лемму, то докажем и теорему, поскольку

$$\nu(G^n) = \big(\nu(G)\big)^n \Rightarrow \sqrt[n]{\alpha(G^n)} \le \sqrt[n]{\big(\nu(G^n)\big)^{-1}} = \big(\nu(G)\big)^{-1}.$$

Доказательство леммы. Пусть $\boldsymbol{w}_G^*$ и $\boldsymbol{w}_H^*$ — наборы весов на вершинах графов G и H, на которых достигаются минимумы в $\nu(G)$ и $\nu(H)$. Рассмотрим набор весов $\widehat{\boldsymbol{w}}_{G\times H}$ на вершинах графа $G \times H$, такой, что

$$\widehat{w}_{G\times H}(x,y) \coloneqq w_G^*(x) \cdot w_H^*(y).$$

Любая максимальная по включению клика K в графе $G \times H$ является множеством пар вида

$$\{(x,y) \mid x \in K_G, y \in K_H\},$$

где K_G и K_H — клики в G и H соответственно. Для таких клик K имеем

$$\widehat{w}_{G\times H}(K) = \sum_{(x,y)\in K} w_G^*(x) \cdot w_H^*(y) = \left(\sum_{x\in K_G} w_G^*(x)\right)\left(\sum_{y\in K_H} w_H^*(y)\right) = w_G^*(K_G) \cdot w_H^*(K_H).$$

Имеем

$$\begin{aligned}\nu(G\times H) = \min_{\boldsymbol{w}} \max_{\substack{K-\text{клика}\\ \text{в } G\times H}} w(K) &\le \max_{\substack{K-\text{клика}\\ \text{в } G\times H}} \widehat{w}_{G\times H}(K) = \max_{\substack{K-\text{клика}\\ \text{в } G\times H}} w_G^*(K_G) \cdot w_H^*(K_H)\\ &\le \max_{\substack{K-\text{клика}\\ \text{в } G}} w_G^*(K) \cdot \max_{\substack{K-\text{клика}\\ \text{в } H}} w_H^*(K) = \nu(G) \cdot \nu(H).\end{aligned}$$

В итоге, $\nu(G \times H) \le \nu(G) \cdot \nu(H)$. Осталось доказать обратное неравенство. Обозначим через $\widetilde{\boldsymbol{w}}$ набор весов на кликах графа (неотрицательны, сумма равна единице). По теореме о свойствах двойственных задач линейного программирования, для любого графа имеет место равенство

$$\min_{\boldsymbol{w}} \max_{K} \sum_{x\in K} w(x) = \max_{\widetilde{\boldsymbol{w}}} \min_{x} \sum_{K \ni x} \widetilde{w}(K)$$

Пусть $\widetilde{\boldsymbol{w}}_G^*$ и $\widetilde{\boldsymbol{w}}_H^*$ — наборы весов, на которых достигается $\max$ для графов G и H. Построим набор весов $\widetilde{w}_{G\times H}$ на кликах графа $G \times H$:

- Пусть клика K может быть представлена в виде $K_G \times K_H$, где K_G и K_H — клики в G и H. Тогда полагаем $\widetilde{w}_{G\times H}(K) \coloneqq \widetilde{w}_G^*(K_G) \cdot \widetilde{w}_H^*(K_H)$.
- Для остальных клик K полагаем $\widetilde{w}_{G\times H}(K) \coloneqq 0$.

Для такого набора весов и для любой фиксированной вершины $(x,y) \in V(G \times H)$ имеем

$$\begin{aligned}\sum_{K\ni(x,y)} \widetilde{w}_{G\times H}(K) &\ge \sum_{\substack{K\ni(x,y)\\ \exists K_G,K_H:\, K=K_G\times K_H}} \widetilde{w}_{G\times H}(K) = \sum_{\substack{K_G\ni x,\\ K_H\ni y}} \widetilde{w}_G^*(K_G) \cdot \widetilde{w}_H^*(K_H)\\ &= \left(\sum_{K_G\ni x} \widetilde{w}_G^*(K_G)\right)\left(\sum_{K_H\ni y} \widetilde{w}_H^*(K_H)\right) \ge \nu(G) \cdot \nu(H).\end{aligned}$$

Отсюда следует неравенство $\nu(G \times H) \geq \nu(G) \cdot \nu(H)$, а значит верно равенство $\nu(G \times H) = \nu(G) \cdot \nu(H)$. *Доказательство леммы, а вместе с ним, и теоремы, завершено.*

7.2. Теоремы Шеннона для каналов с фиксированной вероятностью ошибок

В этом разделе мы будем работать с вероятностной моделью канала связи, и предполагать следующее:

- Канал двоичный и симметричный (алфавит канала $\mathbb{F}_2$, вероятность замены $0 \to 1$ и $1 \to 0$ одинакова).
- В каждом символе, переданном в канал, может произойти ошибка независимо от других символов с вероятностью p, лежащей в интервале $\left(0, \frac{1}{2}\right)$.
- Кодовые слова имеют одинаковую длину n.

Число p называется *вероятностью ошибки на символ*.

Пусть заданы способы кодирования и декодирования, т. е. пара функций:

$$E\colon \mathbb{F}_2^k \to \mathbb{F}_2^n,$$
$$D\colon \mathbb{F}_2^n \to \mathbb{F}_2^k.$$

Скоростью кодирования называется величина $\mathrm{rate}(E, D) \coloneqq \frac{k}{n}$ — это доля информационных символов в сообщении.

Поданное в канал кодовое слово $\boldsymbol{b}$ в результате ошибок в канале преобразуется в слово $\boldsymbol{b}'$. *Вероятностью ошибки на слово* называется величина

$$\frac{1}{\#\text{кодовых слов}} \cdot \sum_{\boldsymbol{b} \text{ — к.сл.}} P[D(\boldsymbol{b}') \neq \boldsymbol{b}].$$

Через $\boldsymbol{e}$ будем обозначать вектор ошибок, произошедших в канале с сообщением.

7.2.1. Теоремы Шеннона

Положим $H(p) \coloneqq -p\log_2 p - (1-p)\log_2(1-p)$. Функция H называется *функцией двоичной энтропии*.

Теорема (о существовании кода). (C.E. Shannon)
Пусть $\varepsilon > 0$ — сколь угодно малое число. Пусть вероятность ошибки на символ $p \in \left(0, \frac{1}{2}\right)$. Тогда существуют такие (большие!) n, k и такие функции E, D, что

$$\mathrm{rate}(E, D) \geq 1 - H(p) - \varepsilon$$

и при случайном выборе слова $\boldsymbol{a} \in \mathbb{F}_2^k$

$$\Pr[D(E(\boldsymbol{a}) + \boldsymbol{e}) \neq \boldsymbol{a}] < \varepsilon.$$

Теорема (о пределе скорости). (C.E. Shannon)
Пусть $\varepsilon, \delta > 0$ — сколь угодно малые числа. Пусть вероятность ошибки на символ $p \in \left(0, \frac{1}{2}\right)$. Тогда для любых достаточно больших n и k и любых E, D, таких, что

$$\mathrm{rate}(E, D) \geq 1 - H(p) + \varepsilon,$$

и при случайном выборе слова $\boldsymbol{a} \in \mathbb{F}_2^k$

$$\Pr[D(E(\boldsymbol{a}) + \boldsymbol{e}) = \boldsymbol{a}] < \delta,$$

где $\delta > 0$ зависит только от ε.

Вывод из теорем Шеннона:
Для двоичного канала связи с постоянной вероятностью ошибки в отдельном символе p максимально достижимая теоретическая скорость передачи сообщений равна $\left(1 - H(p)\right)$ (при условии стремления к нулю вероятности ошибки декодирования).

Далее мы докажем только первую из двух упомянутых теорем; доказательство второй аналогично.

7.2.2. Доказательство теоремы о существовании хороших кодов

Утверждение.

При любом $\lambda \in \left(0, \frac{1}{2}\right)$ справедлива оценка

$$\sum_{i=0}^{\lambda n} \binom{n}{i} \leq 2^{n \cdot H(\lambda)}.$$

Доказательство. Положив $t \coloneqq \lambda n$ и $x \coloneqq \frac{\lambda}{1-\lambda} \in (0,1)$, получаем

$$2^{n \cdot H(\lambda)} = x^{-t} \cdot (1+x)^n \geq \sum_{i=0}^{\lambda n} \binom{n}{i} x^{i-t} \geq \sum_{i=0}^{t} \binom{n}{i}.$$

Лемма о числе ошибок.

Пусть p и γ — произвольные числа, такие, что $p \in (0, \frac{1}{2})$ и $\gamma > 0$. Пусть $\boldsymbol{e} \in \mathbb{F}_2^n$ — случайный вектор, каждая компонента которого равна единице с вероятностью p и нулю с вероятностью $(1-p)$. Тогда $\boldsymbol{P}[\|\boldsymbol{e}\| > n(p+\gamma)] \leq c^n$, где $c < 1$ и c зависит только от p и γ.

Неформально: если в двоичный канал, в котором отдельные символы искажаются с вероятностью p, передать длинное слово, то с большой вероятностью доля ошибок в переданном слове будет не сильно превосходить своё матожидание.

Доказательство. Пусть t — произвольное положительное число. Пусть $e_1, \ldots, e_n$ — компоненты вектора $\boldsymbol{e}$. Пользуясь неравенством Маркова, имеем

$$\begin{aligned}\Pr[\|\boldsymbol{e}\| > n(p+\gamma)] &= \Pr\left[\exp(t \cdot \|\boldsymbol{e}\|) > \exp\left(t \cdot n(p+\gamma)\right)\right] = \Pr\left[\exp(t \cdot \textstyle\sum e_i) > \exp\left(t \cdot n(p+\gamma)\right)\right] \\ &\leq \frac{\boldsymbol{E}[\exp(t \cdot \sum e_i)]}{\exp\left(tn(p+\gamma)\right)} = \exp\left(-tn(p+\gamma)\right) \cdot \boldsymbol{E}[\textstyle\prod \exp(te_i)] = \exp\left(-tn(p+\gamma)\right) \cdot \prod \boldsymbol{E}[\exp(te_i)] \\ &= \left(\exp\left(-t(p+\gamma)\right) \cdot \boldsymbol{E}[\exp(te_1)]\right)^n = \left(\exp\left(-t(p+\gamma)\right) \cdot (1-p+p\exp(t))\right)^n.\end{aligned}$$

Осталось показать, что при достаточно малых t то, что возводится в степень, меньше единицы. Это и будет то самое c. Мы хотим доказать, что при малых t выполнено неравенство

$$\exp\left(-t(p+\gamma)\right) \cdot (1-p+p\exp(t)) < 1$$

или, что то же, неравенство

$$1-p+p\exp(t) < \exp\left(t(p+\gamma)\right).$$

Левую часть неравенства можно оценить так:

$$1-p+p\exp(t) = 1-p+p\left(1+t+O(t^2)\right) = 1+pt+O(t^2),$$

а правую часть следующим образом:

$$\exp\left(t(p+\gamma)\right) = 1+(p+\gamma)t+O(t^2).$$

Отсюда вытекает доказываемое свойство.

Лемма доказана.

Доказательство теоремы Шеннона о существовании кодов. Зафиксируем сколь угодно малое $\varepsilon > 0$. Возьмём γ, удовлетворяющее условиям

$$\begin{aligned}&p+\gamma < \tfrac{1}{2}, \\ &H(p+\gamma) < H(p) + \tfrac{\varepsilon}{2}.\end{aligned}$$

Пусть n и k достаточно велики и при этом

$$\frac{k}{n} = 1 - H(p) - \varepsilon.$$

Зададим функцию $E\colon \mathbb{F}_2^k \to \mathbb{F}_2^n$ *случайным образом*, т.е. для каждого $\boldsymbol{a} \in \mathbb{F}_2^k$ значение $E(\boldsymbol{a})$ выбирается равновероятно среди $\mathbb{F}_2^n$. Для заданной E функцию D определим так:

- $D(\boldsymbol{b}) = \boldsymbol{a}$, если $\boldsymbol{a}$ — единственное слово, для которого $d(E(\boldsymbol{a}), \boldsymbol{b}) \le n(p+\gamma)$.
- $D(\boldsymbol{b}) = \boldsymbol{0}$, если указанного $\boldsymbol{a}$ не существует или оно не единственное.

Временно зафиксируем $\boldsymbol{a} \in \mathbb{F}_2^k$. Через $\Pr_{\boldsymbol{a}}[\cdot]$ обозначим вероятность события при фиксированном $\boldsymbol{a}$. Оценим $\Pr_{\boldsymbol{a}}[D(E(\boldsymbol{a}) + \boldsymbol{e}) \ne \boldsymbol{a}]$. Декодирование может пройти неверно только в одной из следующих ситуаций:

- Ошибок слишком много: $d\big(E(\boldsymbol{a}) + \boldsymbol{e}, E(\boldsymbol{a})\big) > n(p+\gamma)$. Тогда, по Лемме о числе ошибок, вероятность такого события не больше c^n, где $c < 1$.
- Ошибок немного, но схема кодирования выбрана неудачно: $d\big(E(\boldsymbol{a}) + \boldsymbol{e}, E(\boldsymbol{a})\big) \le n(p+\gamma)$, и при этом $\exists \boldsymbol{a}' \ne \boldsymbol{a}$ т. ч. $d\big(E(\boldsymbol{a}) + \boldsymbol{e}, E(\boldsymbol{a}')\big) \le n(p+\gamma)$. Вероятность этого события мы и будем оценивать ниже.

Для каждого фиксированного $\boldsymbol{a}' \in \mathbb{F}_2^n$ имеем

$$\begin{aligned}
&\Pr_{\boldsymbol{a}}\big[d\big(E(\boldsymbol{a}) + \boldsymbol{e}, E(\boldsymbol{a}')\big) \le n(p+\gamma)\big] \\
&\quad = \sum_{\boldsymbol{x},\boldsymbol{y}} 2^{-2n} \cdot \Pr_{\boldsymbol{a}}\big[d\big(E(\boldsymbol{a}) + \boldsymbol{e}, E(\boldsymbol{a}')\big) \le n(p+\gamma) \mid E(\boldsymbol{a}) = \boldsymbol{x}, \boldsymbol{e} = \boldsymbol{y}\big] \\
&\quad = \Pr_{\boldsymbol{a}}\big[d\big(E(\boldsymbol{a}) + \boldsymbol{e}, E(\boldsymbol{a}')\big) \le n(p+\gamma) \mid E(\boldsymbol{a}) = \boldsymbol{e} = \boldsymbol{0}\big] = 2^{-n} \cdot \sum_{k=0}^{(p+\gamma)n} \binom{n}{k} \\
&\quad \le 2^{n\cdot(H(p+\gamma)-1)} < 2^{n\cdot\left(H(p)+\frac{\varepsilon}{2}-1\right)}.
\end{aligned}$$

Отсюда

$$\begin{aligned}
\Pr_{\boldsymbol{a}}\big[\exists \boldsymbol{a}' \text{ т. ч. } d\big(E(\boldsymbol{a}), E(\boldsymbol{a}')\big) \le n(p+\gamma)\big] &< 2^k \cdot 2^{n\cdot\left(H(p)+\frac{\varepsilon}{2}-1\right)} = 2^{n\cdot\left(\frac{k}{n}+H(p)+\frac{\varepsilon}{2}-1\right)} \\
&= 2^{n\cdot\left(1-H(p)-\varepsilon+H(p)+\frac{\varepsilon}{2}-1\right)} = 2^{-\frac{\varepsilon}{2}\cdot n}.
\end{aligned}$$

Итак, при *фиксированном* $\boldsymbol{a}$ и *случайно выбираемых* E и $\boldsymbol{e}$ имеем

$$\Pr_{\boldsymbol{a}}[D(E(\boldsymbol{a}) + \boldsymbol{e}) \ne \boldsymbol{a}] < c^n + 2^{-\frac{\varepsilon}{2}\cdot n} < \hat{c}^n$$

для некоторой константы $\hat{c} < 1$.

При случайном выборе $\boldsymbol{a}, \boldsymbol{e}$ и E имеем

$$\Pr[D(E(\boldsymbol{a}) + \boldsymbol{e}) \ne \boldsymbol{a}] = \sum_{\boldsymbol{x} \in \mathbb{F}_2^k} 2^{-k} \cdot \Pr_{\boldsymbol{a}}[D(E(\boldsymbol{a}) + \boldsymbol{e}) \ne \boldsymbol{a} \mid \boldsymbol{a} = \boldsymbol{x}] < \sum_{\boldsymbol{x} \in \mathbb{F}_2^k} 2^{-k} \cdot \hat{c}^n = \hat{c}^n$$

для некоторой константы $\hat{c} < 1$. Значит, *существует* такое кодирование E, что при случайном выборе $\boldsymbol{a}$ и $\boldsymbol{e}$ выполнено $\Pr[D(E(\boldsymbol{a}) + \boldsymbol{e}) \ne \boldsymbol{a}] < \hat{c}^n$. При всех достаточно больших n выполнено неравенство $\hat{c}^n < \varepsilon$, откуда и следует утверждение теоремы.

Теорема доказана.

8. Некоторые специальные семейства кодов

8.1. Коды Варшамова—Тененгольца

В этом разделе мы рассмотрим *коды Варшамова—Тененгольца*, которые могут исправлять ошибки вставки и выпадения символов.

Код Варшамова—Тененгольца длины n определяется как множество двоичных слов

$$C \coloneqq \left\{ a_1 a_2 \dots a_n \mid \sum_{i=1}^{n} i a_i \equiv 0 \pmod{(n+1)} \right\}.$$

Для мощности кода справедлива формула (приводимая здесь без доказательства)

$$|C| = \frac{1}{2(n+1)} \sum_{\substack{d \mid (n+1) \\ d \text{ нечётно}}} \phi(d) 2^{(n+1)/d},$$

где ϕ — функция Эйлера. Известно, что асимптотически это максимально возможная мощность кода, исправляющего одну ошибку выпадения/вставки символа.

8.1.1. Исправление одной ошибки выпадения в кодах Варшамова—Тененгольца

Пусть C — код В.—Т. длины n, и пусть $\boldsymbol{a} \in C$. Пусть в канал передали $\boldsymbol{a} = a_1 \dots a_n$, и на выходе получили слово

$$\boldsymbol{a}' \coloneqq a'_1 \dots a'_{n-1} = a_1 \dots a_{k-1} a_{k+1} \dots a_n.$$

Наша задача: по $\boldsymbol{a}'$ восстановить $\boldsymbol{a}$. Отметим, что восстановить $\boldsymbol{a}$ — не то же самое, что восстановить пару (k, a_k). Например, если $\boldsymbol{a}' = 1001$, то $\boldsymbol{a} = 10001$, но мы не узнаем, какой именно из нулей выпал.

Положим

$$n_0 \coloneqq \#\{ i > k \mid a_i = 0 \},$$
$$n_1 \coloneqq \#\{ i > k \mid a_i = 1 \}.$$

Заметим, что если $a_k = 0$, то $\boldsymbol{a}$ можно восстановить по $\boldsymbol{a}'$, если известно n_1. Аналогично, если $a_k = 1$, то $\boldsymbol{a}$ можно восстановить по $\boldsymbol{a}'$, если известно n_0. Рассмотрим суммы

$$S \coloneqq \sum_{i=1}^{n} i a_i \text{ и } S' \coloneqq \sum_{i=1}^{n-1} i a'_i.$$

Заметим, что

$$S - S' = \sum_{i=1}^{n} i a_i - \left(\sum_{i=1}^{k-1} i a_i + \sum_{i=k}^{n-1} i a_{i+1} \right) = \sum_{i=k}^{n} i a_i - \sum_{i=k+1}^{n} (i-1) a_i = k a_k + \sum_{i=k+1}^{n} a_i.$$

Получаем

$$S' = S - \left(k a_k + \sum_{i=k+1}^{n} a_i \right) = S - k a_k - n_1.$$

Так как $S \equiv 0 \pmod{(n+1)}$, то $S' \equiv -n_1 - k a_k \pmod{(n+1)}$. Если $a_k = 0$, то $-S' \equiv n_1$. Если $a_k = 1$, то $-S' \equiv n_1 + k = (n - k - n_0) + k = n - n_0$. Осталось научиться определять, чему именно равно a_k. Заметим, что $\|\boldsymbol{a}'\| \geq n_1$ и $\|\boldsymbol{a}'\| \leq (n-1) - n_0$. Отсюда $n_1 \leq \|\boldsymbol{a}'\| < n - n_0$. То есть если $(-S') \bmod (n+1) \leq \|\boldsymbol{a}'\|$, то это n_1, а в противном случае это $n - n_0$.

Итоговый алгоритм восстановления $\boldsymbol{a}$ по $\boldsymbol{a}'$:

- вычисляем величину $T \coloneqq (-\sum_{i=1}^{n-1} i a'_i) \bmod (n+1)$,
- если $T \leq \|\boldsymbol{a}'\|$, то в слово $\boldsymbol{a}'$ вставляем перед T-й с конца единицей символ 0,
- если $T > \|\boldsymbol{a}'\|$, то в слово $\boldsymbol{a}'$ вставляем перед $(n-T)$-м с конца нулём символ 1.

8.1.2. Исправление одной ошибки вставки в кодах Варшамова—Тененгольца

Теперь рассмотрим задачу, когда $\boldsymbol{a}'$ получено из $\boldsymbol{a}$ вставкой символа:

$$\boldsymbol{a}' = \dots a_k x a_{k+1} \dots$$

(Если $k = 0$, то $\boldsymbol{a}' = x\boldsymbol{a}$; если $k = n$, то $\boldsymbol{a}' = \boldsymbol{a}x$.)

Тогда $S' = S + (k+1)x + \sum_{i>k} a_i$, и значит, $S' \equiv (k+1)x + \sum_{i>k} a_i = (k+1)x + n_1$.

Положим $T \coloneqq S' \bmod (n+1)$. Есть два случая, когда $T = 0$:

- $(k+1)x + n_1 = 0$. Тогда $x = 0$ и $a_{k+1} = \cdots = a_n = 0$.
- $(k+1)x + n_1 = n+1$. Тогда $x = 1$ и $a_{k+1} = \cdots = a_n = 1$.

В обоих случаях $\boldsymbol{a}$ получается из $\boldsymbol{a}'$ удалением последнего символа.

Теперь рассмотрим случай, когда $T = \|\boldsymbol{a}'\| > 0$. Это возможно только в одном из двух подслучаев:

- $a_1 = \cdots = a_k = x = 0$
- $a_1 = \cdots = a_k = x = 1$

В любом случае, если $T = \|\boldsymbol{a}'\|$, то $\boldsymbol{a}$ получается из $\boldsymbol{a}'$ удалением первого символа.

Остался случай $0 < T \neq \|\boldsymbol{a}'\|$.

- Если $x = 0$, то $T = n_1 < \|\boldsymbol{a}'\|$.
- Если $x = 1$, то $T = k + 1 + n_1 > \|\boldsymbol{a}'\|$.

При этом оказывается, что $T = k + 1 + (n - k - n_0) = n + 1 - n_0$. В обоих подслучаях нужная для восстановления $\boldsymbol{a}$ информация у нас есть.

8.2. Матрицы Адамара и коды Адамара

8.2.1. Матрицы Адамара

Матрица Адамара — это квадратная матрица из $\{-1,\ 1\}^{n\times n}$, в которой любые две строки ортогональны.

Примеры:

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{pmatrix}.$$

Матрицы Адамара получили своё название благодаря следующей теореме Адамара.

Теорема Адамара.
Пусть $A = (a_{ij}) \in \mathbb{R}^{n\times n}$, и пусть $|a_{ij}| \le 1$. Тогда $|\det A| \le n^{n/2}$, причём неравенство превращается в равенство в точности на матрицах Адамара.

Доказательство. Заметим, что определитель матрицы равен по абсолютному значению объёму параллелепипеда, построенного на векторах-строках матрицы. Длины сторон такого параллелепипеда при условиях теоремы не превосходят $\sqrt{n}$ (и достигают этого значения только если $a_{ij} = \pm 1$). Объём такого параллелепипеда не может превышать величину $\left(\sqrt{n}\right)^n$, и равен ей только если стороны (т.е. векторы-строки матрицы A) попарно ортогональны.

Если H — матрица Адамара, то матрица, полученная из H перестановками строк/столбцов и/или умножением строк/столбцов на -1, тоже является матрицей Адамара. Матрицы Адамара, получаемые друг из друга такими преобразованиями, *эквивалентны*.

Любую матрицу Адамара умножением строк/столбцов на -1 можно привести к *нормализованному* виду

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & & & \\ \vdots & & & \\ 1 & & & \end{pmatrix}.$$

Утверждение.
Если $H \in \{-1,\ 1\}^{n\times n}$ — матрица Адамара, и $n > 2$, то $4|n$.

Доказательство. От матрицы H перейдём к эквивалентной матрице, в которой первые три строки такие:

$$\begin{matrix} 1 & 1 & \cdots & 1 & 1 & 1 & \cdots & 1 & 1 & 1 & \cdots & 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 & 1 & 1 & \cdots & 1 & -1 & -1 & \cdots & -1 & -1 & -1 & \cdots & -1 \\ \underbrace{1 \quad 1 \quad \cdots \quad 1}_{i} & & & & \underbrace{-1 \quad -1 \quad \cdots \quad -1}_{j} & & & & \underbrace{1 \quad 1 \quad \cdots \quad 1}_{k} & & & & \underbrace{-1 \quad -1 \quad \cdots \quad -1}_{l} & & & \end{matrix}$$

Отсюда

$$\begin{cases} i+j+k+l=n \\ i+j-k-l=0 \\ i-j-k+l=0 \\ i-j+k-l=0 \end{cases}$$

Решение этой системы: $i=j=k=l=n/4$.

Утверждение доказано.

Гипотеза Адамара (не доказана).
Матрицы Адамара порядка n существуют(?) для *всех* натуральных n, кратных четырём.

Наименьший порядок, для которого пока не доказано существование матрицы Адамара, равен 668.

Утверждение.
Матрица Адамара порядка n существует для любого $n = 2^k$.

Доказательство (J. J. Sylvester). Заметим, что если H — матрица Адамара, то матрицей Адамара будет и такая: $\begin{pmatrix} H & H \\ H & -H \end{pmatrix}$. Утверждение теперь следует по индукции из того факта, что $\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ — матрица Адамара.

8.2.2. Конструкция Пэли на основе квадратичных вычетов

8.2.2.1. Вспомогательные утверждения

Элемент $a \in \mathbb{F}_q \setminus \{0\}$ называется *квадратичным вычетом*, если $a = x^2$ для некоторого $x \in \mathbb{F}_q$. Остальные элементы из $\mathbb{F}_q \setminus \{0\}$ называются *квадратичными невычетами*.

Например, в $\mathbb{Z}_7$ элементы $1,2,4$ — к.в., а $3,5,6$ — к.н.

Утверждение.

- Если λ — примитивный элемент $\mathbb{F}_q$, то элементы вида λ^{2t} являются к.в., а вида λ^{2t+1} — к.н.
- Если $q = p^m$ и $p > 2$, то ровно половина элементов из $\mathbb{F}_q \setminus \{0\}$ являются к.в., а половина — к.н.

Везде далее будем предполагать, что $p > 2$.

Символом Лежандра называется функция на элементах поля, определяемая так:

$$\chi(a) = \begin{cases} 0, \text{ если } a = 0 \\ 1, \text{ если } a \text{ к. в.} \\ -1, \text{ если } a \text{ к. н.} \end{cases}$$

Нетрудно показать, что символ Лежандра является мультипликативной функцией: для любых $a, b \in \mathbb{F}_q$ имеет место равенство $\chi(a) \cdot \chi(b) = \chi(ab)$.

Утверждение.
Для любого $c \in \mathbb{F}_q \setminus \{0\}$ имеет место равенство $\sum_{b\in\mathbb{F}_q} \chi(b) \cdot \chi(b+c) = -1$.

Доказательство. Т.к. *ровно* половина элементов $\mathbb{F}_q \setminus \{0\}$ квадратичными вычетами, то $\sum_{a\in\mathbb{F}_q}\chi(a)=0$. Также заметим, что

$$\sum_{b\in\mathbb{F}_q}\chi(b)\cdot\chi(b+c)=\sum_{b\in\mathbb{F}_q\setminus\{0\}}\chi(b)\cdot\chi(b+c).$$

С учётом замеченного, получаем

$$\begin{aligned}\sum_{b\in\mathbb{F}_q\setminus\{0\}}\chi(b)\cdot\chi(b+c)&=\sum_{b\in\mathbb{F}_q\setminus\{0\}}\chi(b)\cdot\chi\big(b\cdot b^{-1}(b+c)\big)=\sum_{b\in\mathbb{F}_q\setminus\{0\}}\big(\chi(b)\big)^2\cdot\chi\big(b^{-1}(b+c)\big)\\&=\sum_{b\in\mathbb{F}_q\setminus\{0\}}\chi\big(b^{-1}(b+c)\big)=\sum_{b\in\mathbb{F}_q\setminus\{0\}}\chi(1+b^{-1}c)=\sum_{a\in\mathbb{F}_q\setminus\{1\}}\chi(a)=\sum_{a\in\mathbb{F}_q}\chi(a)-\chi(1)\\&=-1.\end{aligned}$$

Утверждение доказано.

8.2.2.2. Построение матрицы Адамара

Рассмотрим матрицу $\left(t_{a,b}\right)_{a,b\in\mathbb{F}_q}\in\{-1,\ 0,\ 1\}^{q\times q}$, в которой $t_{a,b}\coloneqq\chi(a-b)$. Скалярное произведение любых двух различных строк $\left(t_{a',b}\right)_{b\in\mathbb{F}_q}$ и $\left(t_{a'',b}\right)_{b\in\mathbb{F}_q}$ с учётом доказанного утверждения равно

$$\sum_{b\in\mathbb{F}_q}\chi(a'-b)\cdot\chi(a''-b)=\sum_{b\in\mathbb{F}_q}\chi(b)\cdot\chi\big(b+(a''-a')\big)=-1.$$

Рассмотрим матрицу $T'\coloneqq\left(t'_{a,b}\right)_{a,b\in\mathbb{F}_q}\in\{-1,\ 1\}^{q\times q}$, в которой $t'_{a,b}=\chi(a-b)$, если $a\neq b$ и $t'_{a,b}=-1$ иначе. Скалярное произведение различных строк $\left(t'_{a',b}\right)_{b\in\mathbb{F}_q}$ и $\left(t'_{a'',b}\right)_{b\in\mathbb{F}_q}$ равно

$$\left(\sum_{b\in\mathbb{F}_q}\chi(a'-b)\cdot\chi(a''-b)\right)-\chi(a'-a'')-\chi(a''-a')=-1-\chi(a'-a'')-\chi(a''-a').$$

Если (-1) является квадратичным *не*вычетом в $\mathbb{F}_q$, то $\chi(a''-a')=\chi(-1)\cdot\chi(a'-a'')=-\chi(a'-a'')$, и скалярное произведение получается равным -1. Тогда матрица

$$\begin{pmatrix}1 & 1 & \cdots & 1\\ 1 & & & \\ \vdots & & T' & \\ 1 & & & \end{pmatrix}$$

является нормализованной матрицей Адамара. Известно, что при $4|(q+1)$ элемент (-1) является квадратичным невычетом в $\mathbb{F}_q$, а значит, из приведённых нами конструкций вытекает справедливость следующей теоремы.

Теорема. (R. E. A. C. Paley '1933)
Если p простое и $4|(p^m+1)$, то существует матрица Адамара порядка (p^m+1).

8.2.3. Коды Адамара

Несколько способов построения кодов на основе матриц Адамара были предложены R. C. Bose и S. S. Shrikhande в 1959 г. В любой матрице Адамара любые две строки $\boldsymbol{a},\boldsymbol{b}$ ортогональны. Так как $a,b\in\{-1,\ 1\}^n$, это значит, что ровно половина координат у них совпадает, а половина противоположны. Пусть $A\in\{0,1\}^{n\times n}$ — матрица, полученная из нормализованной матрицы Адамара заменой элементов -1 на 0. Тогда

- множество строк матрицы A с отброшенной первой координатой образует двоичный $\left(n-1,n,\frac{n}{2}\right)$-код,
- множество строк матрицы A и их дополнений образует $\left(n,2n,\frac{n}{2}\right)$-код.

Напомним границу Плоткина (в двоичном случае): если $N < 2d$, то для любого (N, M, d)-кода $M \leq \frac{2d}{2d-N}$. Нетрудно видеть, что коды Адамара с параметрами $\left(n-1, n, \frac{n}{2}\right)$ достигают границы Плоткина, имея максимально число слов при заданных длине и кодовом расстоянии.

8.3. Каскадные коды

8.3.1. Определение

Каскадные коды впервые предложены G. D. Forney ’1966.

Пусть даны два кода: $C_{\text{internal}} — (n, m, d)_q$-код (*внутренний код*) и $C_{\text{external}} — (N, M, D)_m$-код (*внешний код*). Символам алфавита кода C_{ext} сопоставим слова кода C_{int}. Тогда кодовым словам кода C_{ext} соответствуют слова длины Nn в алфавите кода C_{int}. Получаем *каскадный* $(Nn, M, d')_q$*-код*, где $d' \geq Dd$.

Аналогично, пусть $C_{\text{int}} — [n, k, d]$-код и $C_{\text{ext}} — [N, K, D]_{2^k}$-код. Элементам $\mathbb{F}_{2^k}$ сопоставим слова кода C_{int} так, чтобы линейная комбинация элементов $\mathbb{F}_{2^k}$ соответствовала линейной комбинации слов кода C_{int}. Тогда кодовым словам кода C_{ext} соответствуют слова длины Nn в алфавите кода C_{int}. Получаем *каскадный линейный* $[Nn, kK, dD]$*-код*.

В качестве внешнего кода удобно брать оптимальный (например, MDS) код над алфавитом большой мощности. В качестве внутреннего кода можно брать близкий к оптимальному код с не очень большим числом кодовых слов.

8.3.2. Асимптотически хорошие коды с полиномиальным декодированием

Пусть задано семейство двоичных кодов $\tilde{C} \coloneqq \{C_n\}_{n=1}^{\infty}$.

Асимптотической скоростью семейства $\tilde{C}$ называется величина

$$\text{rate}(\tilde{C}) \coloneqq \varliminf_{n\to\infty} \frac{\log_2|C_n|}{n}.$$

Асимптотическим относительным кодовым расстоянием семейства $\tilde{C}$ называется величина

$$\delta(\tilde{C}) \coloneqq \varliminf_{n\to\infty} \frac{d(C_n)}{n}.$$

Семейство кодов называется *асимптотически хорошим*, если для него $\text{rate}(\tilde{C}) > 0$ и $\delta(\tilde{C}) > 0$. Из доказательства теоремы Шеннона следует, что для любого ε существуют семейства кодов, такие, что $\text{rate}(\tilde{C}) \geq 1 - H\left(\delta(\tilde{C})\right) - \varepsilon$. Отсюда сразу следует, что асимптотически хорошие семейства кодов существуют. Однако до появления кодов Форни было неизвестно, существуют ли такие семейства с *полиномиальными алгоритмами декодирования*.

Утверждение.
Если $\delta < 0.5$ и ρ таковы, что $H(\delta) \leq 1 - \rho$, то *существует* семейство линейных кодов $\tilde{C}$, для которого $\text{rate}(\tilde{C}) \geq \rho$ и $\delta(\tilde{C}) \geq \delta$.

Доказательство. Если n, k, d' таковы, что $H(d'/n) \leq 1 - \frac{k}{n}$, то

$$\sum_{j=0}^{d'-1} \binom{n-1}{j} < 2^{n \cdot H(d'/n)} \leq 2^{n-k},$$

и, значит, для чисел n, k, d' выполнены условия теоремы Варшамова—Гилберта, а стало быть, существует линейный $[n, k, d]$-код, где $d > d'$. Если $H(\delta) \leq 1 - \rho$, то берём для каждого n параметры $k \coloneqq \lfloor \rho n \rfloor$, $d' \coloneqq \lfloor \delta n \rfloor$ и получаем требуемое.

Утверждение доказано.

Теорема Варшамова—Гилберта по сути экзистенциальная, и не даёт полиномиальных (по длине кодовых слов) алгоритмов построения кода и декодирования. А хотелось бы, во-первых, для каждого n за полиномиальное от n время строить порождающую/проверочную матрицу некоторого

линейного кода с длиной слов, размерностью и кодовым расстоянием $\Omega(n)$ и, во-вторых, исправлять $\Omega(n)$ ошибок в кодовых словах за полиномиальное от n время. Покажем, как достигнуть этой цели.

Пусть $t \in \mathbb{N}$, и пусть $\delta < 0.5$ — произвольное фиксированное число. По теореме Варшамова—Гилберта, существует линейный код с параметрами $\left[\frac{t}{1-H(\delta)}, t, \frac{\delta t}{1-H(\delta)}\right]$. Возьмём этот код в качестве внутреннего в каскадной конструкции. В качестве внешнего возьмём RS-код с параметрами $[2^t, 2^{t-1}, 2^{t-1}+1]_{2^t}$. Получаем каскадный $[n, k, d]$-код, для которого $n = \frac{t \cdot 2^t}{1-H(\delta)}$, $k = \frac{t \cdot 2^t}{2}$ и $d > \frac{\delta t \cdot 2^t}{2(1-H(\delta))}$. Этот код является асимптотически хорошим, т.к. $\frac{d}{n} \geq \frac{\delta}{2(1-H(\delta))} > 0$ и $\frac{k}{n} \geq \frac{1}{2(1-H(\delta))} > 0$.
Вычислить порождающую матрицу кода можно при каждом t за полиномиальное время, т.к. коды Рида—Соломона строятся за полиномиальное время от своих параметров, а проверочная матрица кода в теореме Варшамова—Гилберта строится хотя и перебором, но параметры этого кода логарифмичны по n.

Покажем, что существует «почти тривиальный» полиномиальный алгоритм, декодирующий кодовые слова, принятые с не более чем $\frac{\delta t \cdot 2^t}{8(1-H(\delta))}$ ошибками. Каждое слово кода Форни имеет вид $\boldsymbol{c} = \boldsymbol{a}_1\boldsymbol{a}_2 \ldots \boldsymbol{a}_{2^t}$, где $\boldsymbol{a}_i$ — слова кода Варшамова—Гилберта длины $\frac{t}{1-H(\delta)}$. Если в слове $\boldsymbol{c}$ произошло $\leq \frac{\delta t \cdot 2^t}{8(1-H(\delta))}$ ошибок, и в результате принято слово $\tilde{\boldsymbol{c}} = \tilde{\boldsymbol{a}}_1\tilde{\boldsymbol{a}}_2 \ldots \tilde{\boldsymbol{a}}_{2^t}$, то слово $\boldsymbol{c}$ восстановим в два шага:

1. Для каждого $\tilde{\boldsymbol{a}}_i$ перебором ищем ближайшее к нему слово кода Варшамова—Гилберта, при этом неверно восстановленных слов может быть не более $\frac{\delta t \cdot 2^t}{8(1-H(\delta))} \Big/ \frac{\delta t}{2(1-H(\delta))} = 2^{t-2}$.
2. Кодовое расстояние внешнего кода равно $(2^{t-1}+1)$, поэтому даже 2^{t-2} ошибок он успешно исправит.

Пусть слово $\boldsymbol{c} = \boldsymbol{a}_1\boldsymbol{a}_2 \ldots \boldsymbol{a}_{2^t}$ исказилось в s разрядах и перешло в слово $\tilde{\boldsymbol{c}} = \tilde{\boldsymbol{a}}_1\tilde{\boldsymbol{a}}_2 \ldots \tilde{\boldsymbol{a}}_{2^t}$, где $\boldsymbol{a}_i$ — слова кода Варшамова—Гилберта с расстоянием d_{int}. Пусть $I = \left\{i \mid d(\tilde{\boldsymbol{a}}_i, \boldsymbol{a}_i) \geq \frac{d_{\text{int}}}{2}\right\}$, где $|I| \leq \frac{2s}{d_{\text{int}}}$. Так как на первом шаге проблемы с исправлением могут быть только у слов $\tilde{\boldsymbol{a}}_i$, у которых $i \in I$, то на втором шаге, рассматривая каждое $\tilde{\boldsymbol{a}}_i$ как один элемент поля $\mathbb{F}_{2^t}$, мы получаем задачу восстановления слова кода Рида—Соломона с ошибками не более чем в $\frac{2s}{d_{\text{int}}}$ разрядах.

9. О приложениях теории кодирования в информатике

9.1. Коммуникационная сложность

- У Аси есть слово X, а у Бори слово Y, где $X, Y \in \{0,1\}^k$. (Ася не знает Y, а Боря не знает X.)
- Задана функция f, определённая на $\{0,1\}^k \times \{0,1\}^k$.
- Ася и Боря хотят вычислить значение $f(X,Y)$, переслав для этого друг другу минимум данных.
- $L_{\text{comm}}(f) :=$ число битов, которые в сумме Ася и Боря перешлют друг другу в худшем случае при использовании фиксированного алгоритма вычисления f.
- Пример: $L_{\text{comm}}(\mathbb{1}_{X=Y}) = k+1$. (Доказательство: принцип Дирихле.)

Построим *рандомизированный* алгоритм вычисления $\mathbb{1}_{X=Y}$, при котором Ася и Боря пересылаю друг другу всего $O(\log k)$ битов.

Идея: используем код, исправляющий ошибки, в качестве «усилителя различия» слов.

Ася и Боря кодируют X и Y в одном и том же $[n, k, d]_q$-коде (кодируя биты 0 и 1 различными элементами из $\mathbb{F}_q$), получая слова $X', Y' \in \mathbb{F}_q^n$.

Ася выбирает *случайную* позицию в X' и пересылает её номер (в двоичной записи объёмом $\lceil \log_2 n \rceil$) и её значение (объёмом $\lceil \log_2 q \rceil$).

Боря проверяет, совпадает ли принятое от Аси значение с соответствующим значением в Y', и затем результат сравнения он одним битом пересылает Асе. Вероятность ошибки не превосходит $(n-d)/n$.

Пусть $\varepsilon \in (0,1)$. Если Ася и Боря используют $\left[\frac{1}{\varepsilon}\cdot k, k, \frac{1-\varepsilon}{\varepsilon}\cdot k + 1\right]_q$-код Рида—Соломона, где $q \in \left[\frac{k}{\varepsilon}, \frac{2k}{\varepsilon}\right]$, то вероятность ошибки будет не больше ε, а количество бит, которые Ася и Боря перешлют друг другу, равно $O\left(\log \frac{k}{\varepsilon}\right)$.

9.2. Криптография с открытым ключом: криптосистема МакЭлиса (R. McEliece '1978)

- Ася вывешивает в интернете алгоритм с открытыми исходниками, который преобразует сообщения X в $\phi(X)$.
- У Аси есть алгоритм, *который знает только она*, позволяющий по коду вида $\phi(X)$ эффективно восстановить сам X.
- Никто, кроме Аси (т.е. у кого нет секретного алгоритма декодирования) не должен уметь эффективно восстанавливать X по $\phi(X)$. Так что собеседник Аси Боря может выкладывать в открытый доступ сообщения, к которым сможет обращаться кто угодно, но расшифровать (за приемлемое время) Борины послания сможет только Ася.

МакЭлис предложил следующую схему. Ася выбирает (не раскрывая никому)

- произвольный $[n,k,d]$-код C, где $d \geq 2t+1$; этот код должен обладать эффективными алгоритмами построения порождающей матрицы $G \in \mathbb{F}_2^{k\times n}$ и декодирования с исправлением не более t ошибок,
- *случайную* невырожденную матрицу S из $\mathbb{F}_2^{k\times k}$,
- *случайную* перестановочную матрицу P из $\mathbb{F}_2^{n\times n}$.

Затем Ася вычисляет матрицу $\hat{G} \coloneqq SGP \in \mathbb{F}_2^{k\times n}$ и выкладывает в открытый доступ алгоритм, который

- по сообщению $X \in \mathbb{F}_2^k$ вычисляет вектор $X\hat{G} \in \mathbb{F}_2^n$ и искажает его в t случайных битах.

Ася может восстановить X, декодировав с исправлением ошибок вектор $\tilde{X}P^{-1}$ (ведь это искажённое слово кода C), и домножив результат на S^{-1}.

Почему именно так:

- Предполагается, что задача NCP даже при известной порождающей матрице кода трудна для «почти всех» кодов.
  Значит, даже зная хороший алгоритм декодирования кода с матрицей G, трудно декодировать код с матрицей GP.
- Домножение X на S перед кодированием призвано разрушить внутреннюю структуру X, чтобы трудно было «угадать» X.

9.3. l-однородные множества и дерандомизация

Множество наборов $U \subseteq \{0,1\}^n$ называется l-однородным, если для любых $i_1, \dots, i_l \in \{1, \dots, n\}$ и любых $t_1, \dots, t_l \in \{0,1\}$ выполнено

$$\frac{\left|\{(a_1, \dots, a_n) \in U \mid a_{i_1} = t_1, a_{i_2} = t_2, \dots, a_{i_l} = t_l\}\right|}{|U|} = 2^{-l}.$$

То есть при случайном равномерном выборе набора $\boldsymbol{a} \in U$ любые l бит в $\boldsymbol{a}$ будут равны фиксированным значениям с той же вероятностью, что и при случайном выборе из «полного» множества $\{0,1\}^n$.

Лемма.
Пусть $C \subseteq \mathbb{F}_2^n$ — линейный $[n, k, \dots]$-код. Множество C является l-однородным т. и т.т., когда любые l столбцов порождающей матрицы кода линейно независимы.

Доказательство. Пусть $G_1, \dots, G_n \in \mathbb{F}_2^k$ — столбцы порождающей матрицы G кода C.
Пусть, например, $G_1 + G_2 + \dots + G_s = \mathbf{0}$, где $s \le l$. Рассмотрим тогда любые $t_1, \dots, t_l$, такие, что $t_1 + \dots + t_s = 1$. Имеем

$$\frac{\left|\{(a_1, \dots, a_n) \in C \mid a_1 = t_1,\ \dots, a_l = t_l\}\right|}{|C|} = 0 \neq 2^{-l}.$$

Пусть теперь $G_1, \dots, G_l$ линейно независимы. Тогда ранг матрицы $(G_1|G_2|\dots|G_l)$ равен l, и значит в G найдутся строки, — пусть это строки $\boldsymbol{g}_1, \dots, \boldsymbol{g}_l \in \mathbb{F}_2^n$, — такие, что их начальные куски длины l линейно независимы.

Обозначим $C_{t_1,\dots,t_l} \coloneqq \{\boldsymbol{c} \in C \mid c_1 = t_1, \dots, c_l = t_l\}$. Для любых $t_1', \dots, t_l', t_1'', \dots, t_l''$ найдётся кодовое слово $\boldsymbol{a}$ (линейная комбинация строк $\boldsymbol{g}_1, \dots, \boldsymbol{g}_l$), для которого $a_1 = t_1' + t_1'', \dots, a_l = t_l' + t_l''$. Тогда $C_{t_1',\dots,t_l'} = C_{t_1'',\dots,t_l''} + \boldsymbol{a}$, отсюда $\left|C_{t_1',\dots,t_l'}\right| = \left|C_{t_1'',\dots,t_l''}\right|$. В итоге получаем $\left|C_{t_1,\dots,t_l}\right| = \frac{|C|}{2^l}$ для любых $t_1, \dots, t_l$.

Лемма доказана.

Теорема.
Двоичный линейный код C является l-однородным множеством т. и т.т., когда $d(C^\perp) > l$.

Доказательство. Применяем лемму, заметив, что порождающая матрица C является проверочной для $C^\perp$, и используем утверждение о связи кодового расстояния с проверочной матрицей линейного кода.

Замечание.
Аналогично можно вести понятие q-ичного l-однородного множества и доказать похожую теорему: линейный код $C \subseteq \mathbb{F}_q^n$ образует l-однородное множество т. и т.т., когда $d(C^\perp) > l$.

Интересны l-однородные множества малой (полиномиальной по n) мощности.

Для того, чтобы получить такое множество, нужно взять линейный код C у которого $d(C) > l$, а $\dim C$ велико, и затем рассмотреть $C^\perp$.

Возьмём в качестве C код Рида—Маллера с параметрами $m \coloneqq \lceil \log_2 n \rceil$, $r \coloneqq m - 2$. Для такого C имеем $d(C) = 2^{m-r} = 4$. Код $C^\perp$ тоже является РМ-кодом, с параметром $r' = m - r - 1 = 1$. При этом $|C^\perp| = 2^{1+m} \le 4n$.

Задача $3_{\ge\gamma}$-SAT: для заданной 3-КНФ найти набор, на котором не менее чем γ-я доля всех скобок обращается в единицу.

Обычная задача 3-SAT — это $3_{\ge 1}$-SAT.

Теорема.
Задача $3_{\ge 7/8}$-SAT полиномиально разрешима.

Доказательство. Пусть 3-КНФ содержит n переменных и m скобок. При случайном выборе набора из $\{0,1\}^n$ имеем

$$\Pr[\text{фиксированная скобка равна нулю}] = \frac{1}{8}.$$

Отсюда

$$\mathbb{E}\,\#\text{скобок, равных нулю} = \frac{m}{8}.$$

Заметим, что $\Pr[\text{фикс. ск. } = 0] = \frac{1}{8}$ и в том случае, когда берётся случайный набор из произвольного 3-однородного множества. Получается, что в любом 3-однородном множестве найдётся набор, на котором $\leq \frac{m}{8}$ скобок равны нулю. Мы приходим к очень простому алгоритму:

- Перебираем всевозможные наборы РМ-кода с параметрами $m := \lceil \log_2 n \rceil, r := m - 2$, подставляем эти наборы в 3-КНФ. Хотя бы один из наборов должен сгодиться.

9.4. Задача о разделении секрета

Задача. Есть несколько человек и *секрет*. Нужно сообщить людям некоторую информацию, так, чтобы все вместе они могли бы восстановить секрет, но никакая компания из меньшего числа человек не могла бы восстановить секрет.

Пусть нужно разделить секрет между m людьми. Сопоставим секрету элемент $s \in \mathbb{F}_q$. Рассмотрим q-ичный код C, такой, что $d(C^\perp) = m + 1$. В порождающей матрице G кода C найдутся $(m+1)$ линейно зависимых столбцов, пусть это первые $(m+1)$ столбцов. Тогда найдутся такие $\alpha_1, \dots, \alpha_{m+1} \in \mathbb{F}_q \setminus \{0\}$, что в любом кодовом слове $\boldsymbol{c}$ первые $(m+1)$ разрядов удовлетворяют соотношению $\alpha_1 c_1 + \dots + \alpha_{m+1} c_{m+1} = 0$. Т.е. c_{m+1} всегда можно однозначно определить по $c_1, \dots, c_m$.

Выберем теперь *случайные* элементы $t_1, \dots, t_{m-1}$. Элемент t_m однозначно выбирем так, чтобы в коде C нашлось слово вида

$$(t_1, \dots t_{m-1}, t_m, s, \dots)$$

Поскольку C является q-ичным m-однородным множеством, зная любые $(m-1)$ из чисел $t_1, \dots, t_m$, об s ничего нельзя сказать.

Литература

Заинтересованным в более глубоком изучении теории кодирования можно рекомендовать ознакомиться со следующими книгами.

1. Ф. Мак-Вильямс, Н. Слоэн. Теория кодов, исправляющих ошибки. М.: Связь, 1979.
2. У. Питерсон, Э. Уэлдон. Коды, исправляющие ошибки. М.: Мир, 1976.
3. А. Ромащенко, А. Румянцев, А. Шень. Заметки по теории кодирования. М.: МЦНМО, 2011.
4. Ю. Л. Сагалович. Введение в алгебраические коды. М.: изд-во МФТИ: 2007.
5. Ф. И. Соловьева. Введение в теорию кодирования. Новосибирск: изд-во НГУ, 2006.

Книги [1] и [2] являются классическими монографиями по теории кодирования, при этом в [1] основное внимание уделено кодам, исправляющим ошибки замещения, а в [2] круг тем более широкий. В частности, в [2] рассмотрены коды, исправляющие ошибки синхронизации и ошибки, возникающие в арифметических устройствах.

Книга [3], пожалуй, наиболее близка к настоящему пособию по набору тем. В [3] вошли также такие актуальные вопросы, как декодирование списком.

Учебники [4] и [5] хорошо подходят для ознакомления с основными разделами теории кодирования. В [4] хорошо изложены основные алгебраические понятия, большое внимание уделено БЧХ-кодам. В [5] рассмотрены коды Препараты и уделено внимание совершенным кодам и каскадным конструкциям.

При подготовке настоящего текста помимо перечисленных книг были использованы следующие источники.

- М. Н. Аршинов, Л. Е. Садовский. Коды и математика. М.: Наука, 1983.
- С. Б. Гашков. Графы-расширители и их применение в теории кодирования // Математическое просвещение, сер. 3, вып. 13, 2009. С. 104–126.
- Ye. Lindell. Introduction to Coding Theory. Lecture notes. Bar-Ilan University.
- J. H. van Lint. Introduction to Coding Theory. Third edition. Springer-Verlag, 1999.
- M. Sudan. Algorithmic Introduction to Coding Theory. Lecture notes. MIT, 2001.