А.Г.Дьячков

Теория информации

§3. Теоремы кодирования для комбинаторной модели двоичного канала с ошибками

Содержание

3.1 Постановка задачи

Пусть $1 \leq t \leq D \leq N$, $M \geq 2$ — натуральные числа. Используя обозначения из **§2**, дадим следующие определения.

Определение 1. Двоичный канал связи

$$\mathbf{y} = \mathbf{x}(m) \oplus \mathbf{z} \ ,$$

где $\mathbf{x}(m) = (x_1(m), x_2(m), \ldots, x_N(m))$ — кодовое слово (сигнал на входе) при передаче сообщения $\xi = m$, $m = \overline{1, M}$, слово $\mathbf{y} = (y_1, \ldots, y_N)$ — сигнал на выходе, $\mathbf{z} = (z_1, \ldots, z_N)$ — шум, назовем *каналом с t ошибками*, если для любого $m = \overline{1, M}$ вес шума $|\mathbf{z}| = \sum_{i=1}^{N} z_i \leq t$.

Определение 2. Код $X = (\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(M))$ объема M, длины N называется *кодом с расстоянием* D, если для любых $m \neq m'$ расстояние Хэмминга

$$\rho\left(\mathbf{x}(m)\,,\,\mathbf{x}(m')\right) \geq D\ .$$

Такой код будем называть также (N, D)-*кодом* объема M.

Важность этих определений объясняет очевидная

Лемма 1. *Если при передаче сообщения $\xi = m$, $m = \overline{1, M}$ по каналу с t ошибками используется (N, D)-код объема M и MP-декодирование этого кода, то при условии*

$$2t + 1 \leq D \iff t \leq \left\lfloor \frac{D-1}{2} \right\rfloor$$

на выходе канала произойдет правильное решение о любом переданном значении $m = \overline{1, M}$ сообщения ξ.

Обозначим через $M(N, D)$ максимально возможный объем (N, D)-кода. Другими словами, объем *максимального (оптимального)* кода. Наша цель — изучение этой величины при больших значениях параметров $1 \leq D \leq N$. Очевидно, что в двух крайних случаях

$$M(N, N) = 2\ ;\ M(N, 1) = 2^N\ .$$

Для исследования асимптотики $(N \to \infty)$ границ $M(N, D)$ введем понятие *скорости* кода

$$R(d) := \overline{\lim_{N \to \infty}} \quad \frac{\log_2 M(N, \lfloor Nd \rfloor)}{N},$$

где d , $0 < d < 1$ — фиксированный параметр. Здесь и далее в **§3** используются двоичные логарифмы.

Отметим,что скорость кода, т.е. функция $R(d)$, дает главный член логарифмической асимптотики объема максимального кода с линейным расстоянием $D \sim dN$, $N \to \infty$.

3.2 Нижние границы

Теорема 1. (Граница исчерпывания)

$$M(N, D) \geq \left\lceil \frac{2^N}{\sum_{i=0}^{D-1} C_N^i} \right\rceil .$$

Доказательство. Пусть $\mathbf{x}(1)$ - произвольное слово из $E_1^{(N)} = (0,1)^N$, а

$$S_r(\mathbf{x}) = \{\mathbf{e} : \rho(\mathbf{x},\mathbf{e}) \le r\}$$

— шар радиуса r с центром в $\mathbf{x} \in (0,1)^N$. Введем множество $E_2^{(N)} = E_1^{(N)} \backslash S_{D-1}(\mathbf{x}(1))$, в котором произвольным образом выберем слово $\mathbf{x}(2)$. Далее, пусть

$$E_m^{(N)} = E_{m-1}^{(N)} \backslash S_{D-1}(\mathbf{x}(m-1)) \ , \ m = 2, 3, \dots ,$$

а $\mathbf{x}(m)$ — произвольное слово из $E_m^{(N)}$, если $E_m^{(N)} \ne \emptyset$. Поскольку $|S_r(x)| = \sum_{i=0}^r C_N^i$, то такой процесс выбора слов (*исчерпывание*) можно продолжать, пока

$$m \le M_D(N) := \left\lceil \frac{2^N}{\sum_{i=0}^{D-1} C_N^i} \right\rceil \ .$$

Теорема 1 доказана.

Несколько более худшую границу дает доказываемая более общим методом *случайного кодирования*

Теорема 2. (**Граница случайного кодирования**).

$$M(N,D) \ge \left\lfloor \frac{2^N}{4\sum_{i=0}^{D-1} C_N^i} \right\rfloor \ .$$

Доказательство. Зафиксируем произвольные целые числа $1 \le D \le N$, $M \ge 2$. Рассмотрим *ансамбль* из 2^{2MN} кодов объема $2M$, длины N. Пусть $X = (\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(2M))$ — произвольный код этого ансамбля. Слово $\mathbf{x}(m)$, $m = \overline{1,2M}$ кода X назовем "плохим", если среди остальных $2M-1$ слов существует хотя бы одно слово $\mathbf{x}(m')$, $m' \ne m$ такое, что $\rho(\mathbf{x}(m)\,;\,\mathbf{x}(m')) \le D-1$. Для $m = \overline{1,2M}$ определим на ансамбле кодов X характеристическую функцию

$$\chi_m(X) = \begin{cases} 1 \quad , & \text{если существует } m' \ne m, \\ & \text{такое что } \rho(\mathbf{x}(m), \mathbf{x}(m')) \le D-1, \\ 0 \quad , & \text{в противном случае.} \end{cases}$$

Тогда

$$M_2 = M_2(X) = \sum_{m=1}^{2M} \chi_m(X) \text{ —}$$

число "плохих"слов кода X, а $M_1 = M_1(X) = 2M - M_2(X)$ — число "хороших"слов кода X. Заметим, что множество "хороших"слов образует (N, D)-код объема M_1.

Как и в §2, рассмотрим на ансамбле кодов X *равномерное* распределение вероятностей $\mathbf{Q}(X)$, т.е. для любого X определим

$$\mathbf{Q}(X) := 2^{-2MN}, \qquad X = (\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(2M)),$$

а символом $\mathbf{Q}\{.\}$ будем обозначать вероятности событий в ансамбле. Тогда среднее значение (математическое ожидание) числа "плохих"слов

$$\overline{M_2(X)} = \sum_{m=1}^{2M} \mathbf{Q}\{\chi_m(X) = 1\} \ .$$

Поскольку для любой пары $m \neq m'$ в данном ансамбле кодов

$$\mathbf{Q}\{\rho(\mathbf{x}(m)\,,\, \mathbf{x}(m')) \leq D - 1\} = \sum_{i=0}^{D-1} C_N^i (1/2)^N := Q_N(D) \ ,$$

то применяя аддитивную границу для вероятности объединенения событий, имеем для любого $m = \overline{1, 2M}$ вероятность

$$\mathbf{Q}\{\chi_m(X) = 1\} \leq (2M - 1) Q_N(D) \ .$$

Поэтому среднее значение случайной величины $M_2(X)$ удовлетворяет неравенству

$$\overline{M_2(X)} \leq 2M(2M-1)Q_N(D) < 4M^2 Q_N(D) \ . \qquad (1)$$

Положим теперь

$$M := M_D(N) := \Big\lfloor 1/\Big(4Q_N(D)\Big)\Big\rfloor = \left\lfloor \frac{2^N}{4\sum_{i=0}^{D-1} C_N^i} \right\rfloor \ .$$

Неравенство (1) тогда означает, что для средних по ансамблю значений справедливы границы

$$\overline{M_2(X)} \leq M_D(N) \iff \overline{M_1(X)} \geq M_D(N) \ .$$

Из нижней оценки для $\overline{M_1(X)}$ вытекает существование кода X (элементарного исхода ансамбля) объема $2M_D(N)$, у которого по крайней мере $M_D(N)$ слов "хорошие", т.е. образуют (N, D)-код.

Теорема 2 доказана.

Задача 3.1. Доказать следующую границу случайного кодирования:

$$M(N,D) \geq \left\lfloor \frac{C_N^{\lfloor N/2 \rfloor}}{4\sum_{i=0}^{D-1}(C^i_{\lfloor N/2 \rfloor})^2} \right\rfloor \ .$$

Указание. Рассмотрите ансамбль кодов $X = (\mathbf{x}(1), \mathbf{x}(2), \ldots, \mathbf{x}(2M))$ объема $2M$ и длины N, который называется *ансамблем равновесных кодов* и распределение вероятностей $\mathbf{Q}(X)$ на элементах ансамбля задается соотношением

$$\mathbf{Q}(X) = \prod_{m=1}^{2M} Q_N(\mathbf{x}(m)) \ ,$$

$$Q_N(\mathbf{x}) = \begin{cases} \left(C_N^{\lfloor N/2 \rfloor}\right)^{-1} & , \quad \text{если } |\mathbf{x}| = \lfloor N/2 \rfloor \ , \\ 0 & , \quad \text{если } |\mathbf{x}| \neq \lfloor N/2 \rfloor \ . \end{cases}$$

Это означает, что слова $\mathbf{x}(m)$, $m = \overline{1, 2M}$ взаимно независимы, одинаково распределены и имеют равномерное распределение на множестве из $C_N^{\lfloor N/2 \rfloor}$ двоичных слов веса $\lfloor N/2 \rfloor$. Проверьте, что для любой пары $m \neq m'$ в данном ансамбле вероятность

$$\mathbf{Q}\{\rho(\mathbf{x}(m), \mathbf{x}(m')) \leq D-1\} = \frac{\sum_{i=0}^{D-1}(C^i_{\lfloor N/2 \rfloor})^2}{C_N^{\lfloor N/2 \rfloor}} \ .$$

Из теоремы 1(2) и задачи 2.3 из **§2** вытекает

Следствие 1. *Для любого* d , $0 < d \leq 1/2$ *скорость кода*

$$R(d) \geq 1 - h(d) \ . \tag{2}$$

Граница (2) называется границей **Варшамова-Гильберта**.

Из леммы 1 и границы Варшамова - Гильберта следует, что для исследуемого в **§2** канала с равновесным шумом (КРШ) при любой скорости R , $0 < R < 1-h(2p) \leq C = 1-h(p)$ существует (N,R)-код X, у которого при достаточно большом значении N вероятность ошибки $\mathcal{P}^{(t)}(X) = 0$. Отсюда вытекает

Следствие 2. *Для любой скорости* R , $0 < R < 1-h(2p)$ *существует число* $N_0 = N_0(p,R)$ *такое, что минимально возможная вероятность ошибки в КРШ* $\mathcal{E}_{N_0}(p,R) = 0$.

3.3 Верхние границы

Теорема 3. (**Граница Хэмминга**).

$$M(N,D) \le 2^N / \sum_{i=0}^{\lfloor \frac{D-1}{2} \rfloor} C_N^i \ . \tag{3}$$

Доказательство. Пусть существует (N,D)-код $X = (\mathbf{x}(1),\ldots,\mathbf{x}(M))$ объема M. Надо проверить, что число M должно быть ограничено сверху правой частью (3). Положим, для краткости, $t := \lfloor (D-1)/2 \rfloor$. Очевидно, что сферы $S_t(\mathbf{x}(m))$, $m = \overline{1,M}$ радиуса t с центрами в кодовых словах $\mathbf{x}(m)$ попарно не пересекаются. Следовательно, сумма их объемов не превосходит 2^N, т.е.

$$2^N \ge \sum_{m=1}^{M} |S_t(\mathbf{x}(m))| = M \sum_{i=0}^{t} C_N^i \ , \ t = \left\lfloor \frac{D-1}{2} \right\rfloor \ .$$

Теорема 3 доказана.

Из теоремы 3, по аналогии со следствием 1, вытекает асимптотическая граница Хэмминга

Следствие 3. *Для любого* d , $0 < d < 1$ *скорость кода*

$$R(d) \le 1 - h(d/2) \ .$$

Пусть $d_0, 0 < d_0 < 1/2$,—единственный корень уравнения $2d = h(d/2)$. Тогда при $d > d_0$ асимптотическую границу Хэмминга уточняет

Теорема 4. (**Граница Плоткина**). *Для любого* $d \ge 1/2$ *скорость кода* $R(d) = 0$, *а при* $0 < d < 1/2$ *скорость*

$$R(d) \le 1 - 2d \ .$$

Для вывода границы Плоткина нам потребуется

Лемма 2. *Параметры произвольного* (N,D)*-кода объема* M *удовлетворяют неравенству*

$$N \ge 2D\left(1 - \frac{1}{M}\right) \ .$$

Доказательство леммы 2. Пусть $\mathbf{x}(m) = (x_1(m), x_2(m), \ldots, x_N(m))$, $m = \overline{1, M}$, кодовое слово (N, D)-кода $X = (\mathbf{x}(1), \mathbf{x}(2), \ldots \mathbf{x}(M))$. Для каждого $n = \overline{1, N}$ и пары $m \neq m'$ определим

$$\xi_n(m, m') = \begin{cases} 1 & , \text{ если } x_n(m) \neq x_n(m'), \\ 0 & , \text{ если } x_n(m) = x_n(m'). \end{cases}$$

Тогда расстояние Хэмминга

$$\rho\left(\mathbf{x}(m)\,,\,\mathbf{x}(m')\right) = \sum_{n=1}^{N} \xi_n(m\,;\,m')\ .$$

Рассмотрим сумму всех C_M^2 попарных расстояний кода X

$$\sum_{m<m'} \rho\left(\mathbf{x}(m)\,,\,\mathbf{x}(m')\right) = \sum_{n=1}^{N} \left[\sum_{m<m'} \xi_n(m; m')\right] = \sum_{n=1}^{N} r_n(M - r_n)\ , \qquad (4)$$

где $r_n \triangleq \sum_{m=1}^{M} x_n(m)$ — число единиц в n-ой строке кода X. Мы воспользовались тем, что сумма, стоящая в квадратных скобках, представляет собой число пар слов кода X, которые не совпадают в n-ой строке. Число таких пар, очевидно, равно $r_n(M - r_n)$. Заметим,что $C_{r_n}^2$ — число пар $m < m'$, у которых $x_n(m) = x_n(m') = 1$, число $C_{M-r_n}^2$ — количество пар $m < m'$, у которых $x_n(m) = x_n(m') = 0$, а общее число пар $C_M^2 = C_{r_n}^2 + C_{M-r_n}^2 + r_n(M - r_n)$.

С другой стороны, из определения (N, D)-кода следует, что сумма попарных расстояний

$$\sum_{m<m'} \rho\left(\,\mathbf{x}(m)\,;\,\mathbf{x}(m')\,\right) \geq D \cdot C_M^2\ . \qquad (5)$$

Из (4) и (5) имеем

$$D\,\frac{M(M-1)}{2} \leq \sum_{n=1}^{N} r_n(M - r_n) \leq \frac{NM^2}{4}\ ,$$

т.е. параметры любого (N, D)-кода объема M удовлетворяют неравенству $N \geq 2D\left(1 - \frac{1}{M}\right)$.

Лемма 2 доказана.

Доказательство теоремы 4. 1) Рассмотрим сначала случай, когда $d > 1/2$. Положив в неравенстве леммы 2 параметр $D = \lfloor Nd \rfloor$, а параметр $M = M(N, \lfloor Nd \rfloor)$, имеем

$$N \geq 2\lfloor Nd \rfloor \left(1 - \frac{1}{M(N, \lfloor Nd \rfloor)}\right). \qquad (6)$$

Предположим *противное*, т.е. при $d > 1/2$ скорость $R(d) > 0$. Тогда из определения скорости следует, что

$$\lim_{N\to\infty} M = \lim_{N\to\infty} M(N, \lfloor Nd \rfloor) = \infty. \tag{8}$$

Поэтому, переходя в (6) к пределу при $N \to \infty$, получаем $d \le 1/2$, что завершает доказательство теоремы 4 в случае $d > 1/2$.

2) Рассмотрим теперь случай, когда $0 < d < 1/2$. Зафиксируем произвольное натуральное n, $1 \le n \le \log_2 M$. Заметим, что среди всех 2^n двоичных слов длины n существует хотя бы одно слово, которое является "началом"по крайней мере $M' \ge M/2^n$ кодовых слов (N, D)-кода. Следовательно, "хвосты"этих M' слов должны образовывать $(N-n, D)$-код объема M'. Поэтому, в силу леммы 2, для любого n, $1 \le n \le \log_2 M$, имеем

$$N - n \ge 2D\left(1 - \frac{1}{M'}\right) \ge 2D\left(1 - \frac{2^n}{M}\right) \ ,$$

т.е.

$$n \le N - 2D\left(1 - \frac{2^n}{M}\right) \ .$$

Полагая $n := \lceil \log_2(M/\log_2 M) \rceil < \log_2 M$, получаем, что параметры любого (N, D)-кода объема N удовлетворяют неравенству

$$\log_2 M \le N - 2D\left(1 - \frac{1}{\log_2 M}\right) + \log_2 \log_2 M + 1 \ .$$

Поскольку слова (N, D)-кода объема M отличны друг от друга, то $N \ge \log_2 M$. Поэтому из последнего неравенства имеем

$$\frac{\log_2 M}{N} \le 1 - 2\,\frac{D}{N}\left(1 - \frac{1}{\log_2 M}\right) + \frac{\log_2 \log_2 M + 1}{\log_2 M} \ . \tag{9}$$

Если при $0 < d < 1/2$ скорость $R(d) = 0$, то теорема 4 очевидна. Пусть при $0 < d < 1/2$ скорость $R(d) > 0$. Положим в (9) параметр $D = \lfloor Nd \rfloor$, а параметр $M = M(N, \lfloor Nd \rfloor)$. Переходя в (9) к пределу при $N \to \infty$ и учитывая (8), устанавливаем справедливость теоремы 4 при $0 < d < 1/2$.

3) В случае $d = 1/2$ равенство $R(1/2) = 0$ очевидным образом (об'ясните почему?) доказывается от противного с помощью (9) и рассуждений, аналогичных п.1).

Теорема 4 доказана.

3.4 Границы для линейных кодов

3.4.1 Определения и обозначения

Пусть $1 \le k < N$, $M = 2^k$. Согласно разделу **2.2** из **§2**, линейный код $X = \Big(\mathbf{x}(1), \ldots, \mathbf{x}(M)\Big)$ длины N есть линейное подпространство размерности k, базисом которого являются столбцы

$$\mathbf{g}(i) = \Big(g_1(i), g_2(i), \ldots, g_N(i)\Big) \ , \ i = \overline{1, k}$$

порождающей матрицы $G = \|g_n(i)\|$, $n = \overline{1, N}$, $i = \overline{1, k}$ размера $N \times k$. Если $m-1 = \sum_{i=1}^{k} u_i 2^{i-1}$, то информационному слову $\mathbf{u} = (u_1, u_2, \ldots, u_k)$ соответствует кодовое слово

$$\mathbf{x}(m) = \sum_{i=1}^{k} u_i \, \mathbf{g}(i)$$

В линейном коде слово $\mathbf{x}(1)$ — нулевое, а остальные $2^k - 1$ слов $\mathbf{x}(2), \mathbf{x}(3), \ldots, \mathbf{x}(2^k)$ — ненулевые.

Из определения линейного кода следует, что если $\mathbf{x}, \mathbf{y} \in X$, то их сумма $\mathbf{x} \oplus \mathbf{y} \in X$. Поскольку расстояние $\rho(\mathbf{x}, \mathbf{y}) = |\mathbf{x} \oplus \mathbf{y}|$, то справедлива

Лемма 3. *Линейный код объема $M = 2^k$ является (N, D)-кодом тогда и только тогда, когда для любого $m = \overline{2, 2^k}$ вес $|\mathbf{x}(m)| \ge D$.*

Обозначим через $\overline{X}$ нулевое пространство для подпространства X. Пусть $r := N - k$ — размерность $\overline{X}$, т.е. число элементов нулевого пространства $|\overline{X}| = 2^r = 2^{N-k}$, и столбцы

$$\mathbf{h}(i) = \Big(h_1(i), h_2(i), \ldots, h_N(i)\Big) \ , \ i = \overline{1, r},$$

образуют базис $\overline{X}$. Рассмотрим $(N \times r)$-матрицу $H = \|h_j(i)\|$, $j = \overline{1, N}$, $i = \overline{1, r}$, составленную из столбцов $\mathbf{h}(1), \mathbf{h}(2), \ldots, \mathbf{h}(r)$:

$$H = \underbrace{\begin{pmatrix} h_1(1) & h_1(2) & \ldots & h_1(r) \\ \ldots & \ldots & \ldots & \ldots \\ h_n(1) & h_n(2) & \ldots & h_n(r) \\ \ldots & \ldots & \ldots & \ldots \\ h_N(1) & h_N(2) & \ldots & h_N(r) \end{pmatrix}}_{r = N - k} \left.\vphantom{\begin{pmatrix} a \\ a \\ a \\ a \\ a \end{pmatrix}}\right\} N$$

Эта матрица H называется *проверочной матрицей* линейного кода X. Согласно определению H для любого кодового слова $\mathbf{x}(m) =$

$\Big(x_1(m), x_2(m), \ldots, x_N(m)\Big)$

$$\sum_{j=1}^{N} x_j(m)\,\mathbf{h}_j = \mathbf{0},$$

где $\mathbf{h}_j := \Big(h_j(1), h_j(2), \ldots, h_j(r)\Big)$, $j = \overline{1,N}$ обозначает j-ю строку матрицы H.

Из данного свойства проверочной матрицы H и леммы 3 очевидным образом вытекает

Лемма 4. *Линейный код X является (N,D)-кодом объема 2^k тогда и только тогда, когда в его проверочной матрице H любые $n \le D-1$ строк линейно независимы.*

Задача 3.2. Пусть E_k — единичная матрица размера $(k \times k)$, A — произвольная матрица размера $(N-k) \times k$ и порождающая $(N \times k)$-матрица G линейного кода имеет *систематический* вид:

$$G = \underbrace{\begin{pmatrix} E_k \\ A \end{pmatrix}}_{k} \begin{matrix} \} & k \\ \} & N-k=r \end{matrix} \quad , \qquad A = \|a_l(j)\| \ , \ l = \overline{1,r} \ , \ j = \overline{1,k}.$$

Докажите, что проверочная матрица H этого кода может быть записана в виде

$$H = \begin{pmatrix} A^T \\ E_{N-k} \end{pmatrix} \begin{matrix} \} & k \\ \} & N-k \end{matrix} \quad .$$

Для этого проверьте, что скалярное произведение $\Big(\mathbf{g}(j) \cdot \mathbf{h}(l)\Big) = a_l(j) \oplus a_l(j) = 0$.

3.4.2 Код Хэмминга

Этот код имеет следующие параметры

$$N = 2^r - 1 \ , \quad k = N - r \ , \quad M = 2^{N-r} \ , \quad r = 2, 3, \ldots \ ,$$

и задается своей проверочной матрицей H, строки которой $\mathbf{h}_j = \Big(h_j(1), \ldots, h_j(r)\Big)$, $j = \overline{1, 2^r - 1}$ представляют собой двоичные разложения чисел из множетва $[2^r - 1]$: строка $\mathbf{h}_j$ есть разложение числа $j \in [2^r - 1]$. Согласно лемме 4, расстояние этого кода $D = 3$.

Если при $D = 3$, $N = 2^r - 1$, $r = 2, 3, \ldots$ выписать верхнюю границу Хэмминга из теоремы 3, то получим

$$2^N / \sum_{i=0}^{\lfloor \frac{D-1}{2} \rfloor} C_N^i = 2^N/(1+N) = 2^N/2^r = 2^{N-r} = M \ ,$$

т.е. граница Хэмминга совпадает с объемом построенного кода. Следовательно, объем максимального кода $M(2^r - 1\,;\, 3) = \exp_2\{2^r - r - 1\}$ и достигается на кодах Хэмминга.

Декодирование. По выходу канала $\mathbf{y} = \mathbf{x}(m) + \mathbf{z}$ вычисляется *синдром*

$$\mathbf{s} = \mathbf{y}H = \underbrace{\mathbf{x}(m)H}_{\mathbf{0}} \oplus \mathbf{z}H = \mathbf{z}H.$$

Заметим, что если произошла одиночная ошибка, т.е. $|\mathbf{z}| = 1$, то синдром $\mathbf{s} = \mathbf{z}H$ представляет собой строку матрицы H, которая есть двоичное разложение номера позиции, где произошла эта одиночная ошибка. Далее по $\mathbf{y}$ и $\mathbf{z}$, $|\mathbf{z}| = 1$ правильно вычисляется переданное кодовое слово $\mathbf{x}(m) = \mathbf{z} \oplus \mathbf{y}$.

Кодирование. В проверочной матрице H кода Хэмминга строки с номерами $n = 2^i$, $i = \overline{0, r-1}$, содержат ровно по одной единице. Поэтому при кодировании

$$\mathbf{u} = (u_1, u_2, \ldots, u_k) \implies \mathbf{x} = (x_1, x_2, \ldots, x_N)$$

удобно $k = N - r = 2^r - r - 1$ позиций слова $\mathbf{x}$ с номерами $n \neq 2^i$, $i = \overline{0, r-1}$, заполнить поочередно символами $u_1, u_2, \ldots, u_k$, а символ x_n при $n = 2^i$, $i = \overline{0, r-1}$, вычислить как соответствующую матрице H линейную комбинацию (см. пример) символов $u_1, u_2, \ldots, u_k$. Это дает возможность *удобного* восстановления информационного слова $\underline{u} = (u_1, u_2, \ldots, u_k)$ по вычисленному на выходе канала кодовому слову $\mathbf{x}$.

Пример.

а) $r = 2$, $N = 3$, $k = 1$, $M = 2$, т.е. код $X = \Big[(0,0,0)\,;\, (1,1,1) \Big]$ и $D = 3$.

б) $r = 3$, $N = 7$, $k = N - r = 4$, $M = 2^4 = 16$;

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

Поскольку 1-ая, 2-ая и 4-ая строки H содержат ровно по *одной* 1, то слово $\mathbf{x} = (x_1, x_2, \dots, x_7)$ является словом данного кода Хэмминга, если

$$\begin{aligned} x_1 &= x_3 \oplus x_5 \oplus x_7 \ , \\ x_2 &= x_3 \oplus x_6 \oplus x_7 \ , \\ x_4 &= x_5 \oplus x_6 \oplus x_7 \ . \end{aligned}$$

Данные соотношения означают, что отображение информационного слова $\mathbf{u} = (u_1, u_2, u_3, u_4)$ в соответствующее кодовое слово $\mathbf{x}(\mathbf{u}) = (x_1, x_2, \dots, x_7)$ задается следующим образом:

$$\begin{aligned} x_3 = u_1 &\implies (1000) = \mathbf{g}_3; \\ x_5 = u_2 &\implies (0100) = \mathbf{g}_5; \\ x_6 = u_3 &\implies (0010) = \mathbf{g}_6; \\ x_7 = u_4 &\implies (0001) = \mathbf{g}_7; \end{aligned} \qquad \begin{aligned} x_1 = u_1 \oplus u_2 \oplus u_4 &\implies (1101) = \mathbf{g}_1; \\ x_2 = u_1 \oplus u_3 \oplus u_4 &\implies (1011) = \mathbf{g}_2; \\ x_4 = u_2 \oplus u_3 \oplus u_4 &\implies (0111) = \mathbf{g}_4. \end{aligned}$$

Поэтому порождающая матрица кода Хэмминга при $N = 7$, $k = 4$, $r = 3$, $D = 3$, $M = 16$ имеет вид

$$G = \begin{pmatrix} \mathbf{g}_1 \\ \mathbf{g}_2 \\ \mathbf{g}_3 \\ \mathbf{g}_4 \\ \mathbf{g}_5 \\ \mathbf{g}_6 \\ \mathbf{g}_7 \end{pmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

Задача 3.3. Докажите, что линейный код объема $M = 2^k$ длины $N = k + 1$, $k = 2, 3, \dots$, задаваемый проверочной матрицей $H = (1, 1, \dots, 1)$, имеет кодовое расстояние $D = 2$. Найдите порождающую матрицу G этого кода, который называется кодом с *проверкой на четность.* Докажите, что в классе линейных кодов с расстоянием $D = 2$ и длины $N \geq 2$ код с проверкой на четность имеет максимально возможный объем.

3.4.3 Граница Варшамова – Гильберта для линейных кодов

Теорема 5. *Существует линейный (N, D)-код объема*

$$M = 2^k \geq \frac{2^N}{\sum_{i=0}^{D-2} C_N^i},$$

т.е. об‘ем максимального линейного (N, D)-кода

$$M_{linear}(N; D) \geq \frac{2^N}{\sum_{i=0}^{D-2} C_N^i}.$$

Замечание. Теорема 5 дает более сильную нижнюю границу для объема максимального кода, чем граница исчерпывания теоремы 1. Отметим также, что нижняя граница для скорости $R(d)$, которая вытекает из теоремы 5, совпадает с нижней границей для скорости, вытекающей из теоремы 1 и сформулированной в следствии 1.

Доказательство. В силу леммы 2, достаточно доказать существование последовательности из N двоичных строк

$$\begin{aligned}
&\mathbf{h}_1 = \Big(h_1(1), h_1(2), \ldots, h_1(r) \Big),\\
&\mathbf{h}_2 = \Big(h_2(1), h_2(2), \ldots, h_2(r) \Big),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&\mathbf{h}_N = \Big(h_N(1), h_N(2), \ldots, h_N(r) \Big),
\end{aligned}$$

которая удовлетворяет следующим двум условиям:

1. каждая строка имеет длину r такую, что
$$2^r \leq \sum_{i=0}^{D-2} C_N^i \ ;$$
2. строку $\mathbf{h}_n$, $n = \overline{1, N}$ нельзя записать в виде линейной комбинации из $D-2$ предыдущих строк, т.е. $\mathbf{h}_n \neq \sum_{t=1}^{i} \mathbf{h}_{s_t}$ для любого значения i , $1 \leq i \leq D-2$ и любых чисел $1 \leq s_1 < s_2 < \cdots < s_t \leq n-1$.

Пусть уже построены строки $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_{n-1}$, удовлетворяющие условиям 1) и 2). Из построенных $n-1$ строк можно составить $\sum_{i=1}^{D-2} C_{n-1}^i$ различных линейных комбинаций, содержащих $\leq D-2$ элементов. Поэтому при выполнении условия

$$\sum_{i=1}^{D-2} C_{n-1}^i < 2^r - 1 \ , \ n = 1, 2, \ldots, N$$

существует строка $\mathbf{h}_n$ длины r такая, что строки $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_n$ удовлетворяют условию 2). Процесс добавления строк прекратится, т.е. нельзя добавить строку $\mathbf{h}_{N+1}$, если N — наименьшее из чисел, для которых

$$\sum_{i=1}^{D-2} C_N^i \geq 2^r - 1 \iff \sum_{i=0}^{D-2} C_N^i \geq 2^r \ .$$

Таким образом, мы показали, что при заданном $r = 1, 2, \ldots$ существует число N и последовательность из N строк $\mathbf{h}_1, \mathbf{h}_2, \ldots, \mathbf{h}_N$, удовлетворяющая условиям 1) и 2).

Теорема 5 доказана.