§ 4. Линейные коды

Определения.

Чтобы иметь эффективные алгоритмы кодирования и декодирования, необходимо ввести определенную структуру в кодовое пространство. Пусть F — конечное поле (в качестве примеров чаще всего будем рассматривать поле $\mathbb{F}_2 = \{0, 1\}$). *Линейным кодом* называется векторное пространство над полем F. Линейный код C длины n и размерности k вкладывается как подпространство в векторное пространство F^n. Таким образом, слова кодового пространства F^n суть векторы, поэтому кодовые слова (т.е. элементы C) будем называть кодовыми векторами.

Линейный код с параметрами n, k, d (d — наименьшее расстояние кода) называется $[n, k, d]$-кодом. Квадратные скобки, в отличие от круглых, применяются для обозначения линейности кода.

Примеры. 1. Код повторения длины n есть $[n, 1, n]$-код.

2. Код с проверкой на четность длины n есть $[n, n-1, 2]$-код.

3. Коды Хэмминга $[7, 4]$, $[8, 4]$ и $[4, 2]$, поскольку были определены посредством проверочных уравнений, также являются линейными.

4. Код Рида–Соломона, рассмотренный в §1, п.6, есть $[27, 7, 21]$-код над полем $\mathbb{R}$.

В случае линейных кодов вместо понятия расстояния можно использовать понятие веса. *Вес Хэмминга* $w_H(v)$ вектора v есть число его ненулевых координат. Очевидно, что $w_H(v) = d_H(x, 0)$. *Минимальный вес* кода C есть минимальное ненулевое значение веса всех кодовых слов:

$$w_{\min}(C) = w(C) = \min_{0 \neq x \in C} w_H(x).$$

Для минимального веса также справедливо совпадение $w_{\min}(v) = d_{\min}(x, 0)$, что следует из равенства $d_H(x, y) = d_H(x - y, 0)$.

Примеры. 1. Минимальный вес кода повторения длины n равен n.

2. Минимальный вес кода с проверкой на четность равен 2, независимо от длины кода.

3. Коды Хэмминга $[7, 4]_2$, $[8, 4]_2$ и $[4, 2]_3$ обладают минимальными весами 3, 4 и 3 соответственно.

Порождающая матрица кода.

Векторное подпространство $C \subset F^n$ размерности k можно задать, указав его базис $c_1, \ldots, c_k \in C$. *Порождающей матрицей* кода C называется матрица G размера $k \times n$, строки которой суть базисные векторы $c_1, \ldots, c_k$:

$$C = \{\alpha G \mid \alpha \in F^k\} = \Big\{\sum_{i=1}^{k} \alpha_i c_i \mid \alpha_i \in F\Big\}.$$

Порождающая матрица, так же как и базис, неединственна.

Примеры. 1. Порождающая матрица кода повторения есть $G = (1, 1, \ldots, 1)$.

2. Код с проверкой на четность обладает порождающей матрицей вида

$$\begin{pmatrix} 1 & 0 & 0 & \ldots & 0 & 0 & 1 \\ 0 & 1 & 0 & \ldots & 0 & 0 & 1 \\ 0 & 0 & 1 & \ldots & 0 & 0 & 1 \\ & \vdots & & \ddots & & \vdots & \\ 0 & 0 & 0 & \ldots & 1 & 0 & 1 \\ 0 & 0 & 0 & \ldots & 0 & 1 & 1 \end{pmatrix}$$

3. Для кода Хэмминга $[7, 4]_2$ положим символы сообщения (x_3, x_5, x_6, x_7) равными $(1, 0, 0, 0)$, затем $(0, 1, 0, 0)$, $(0, 0, 1, 0)$ и $(0, 0, 0, 1)$. Соответствующие кодовые слова образуют порождающую матрицу

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

4. Порождающая матрица расширенного $[8, 4]_2$-кода Хэмминга может быть получена добавлением столбца в предыдущую матрицу с целью дополнить каждый базисный вектор (строку матрицы) до четного веса:

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

5. Чтобы построить порождающую матрицу троичного $[4, 2]_3$-кода Хэмминга, в качестве (a, b) выберем $(1, 0)$ и затем $(0, 1)$:

$$\begin{pmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$

Дуальный код к коду $C \subset F^n$ (не обязательно линейному) — это код

$$C^\perp = \{x \in F^n \mid \langle x, c\rangle = 0 \ \forall c \in C\},$$

где $\langle x, c\rangle = \sum_{i=1}^n x_i c_i$ — скалярное произведение.

Дуальный код $C^\perp$ всегда линейный, даже если C не линеен. Рассматривая код, дуальный к дуальному, легко проверить, что $(C^\perp)^\perp = C^{\perp\perp} \supseteq C$. Причем, если код C линейный, то $C^{\perp\perp} = C$. Например, дуальный к двоичному коду повторения длины n — код проверки на четность длины n, дуальный к коду проверки на четность длины n — код повторения длины n.

Лемма. Если C — линейный $[n, k]$-код над полем F, то дуальный код $C^\perp$ также линейный код над полем F, и $C^{\perp\perp} = C$.

Доказательство. Пусть G — порождающая матрица для C. Тогда $x \in C^\perp \iff Gx^t = 0$. Следовательно, векторы из $C^\perp$ — это в точности транспонированные к векторам из нулевого пространства (ядра) матрицы G, что доказывает первое утверждение. Так как $\dim C + \dim C^\perp = n$, размерность $C^\perp$ равна $n-k$. Повторяя вычисление еще раз, получим, что $C^{\perp\perp}$ имеет размерность k. Поскольку это подпространство содержит C и имеет ту же размерность, оно совпадает с C.

Линейный код C называется *самоортогональным*, если $C^\perp \supseteq C$, и *самодуальным*, если $C^\perp = C$. Например, двоичный код повторения четной длины самоортогонален, так же как и двоичный дуальный $[7, 3]$-код Хэмминга. Размерность кода и размерность дуального к нему в сумме дают длину кода, поэтому самодуальный код должен быть линейным $[2k, k]$-кодом при некотором k. Легко установить, что расширенный $[8, 4]_2$-код Хэмминга и $[4, 2]_3$-код Хэмминга оба самодуальны.

Задачи.

1. Для двоичного кода длины 16 кодовые слова удобно записывать в виде квадратных (0,1)-матриц размера 4×4. Пусть код E состоит из таких матриц M, в которых

a) каждая строка содержит четное число единиц;

b) каждый столбец в пределах одной матрицы содержит одинаковое по четности число единиц (либо четное, либо нечетное).

Докажите, что E — линейный код, и найдите его размерность и минимальное расстояние. Отыщите все кодовые матрицы в E, наи-

менее удаленные от матрицы

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

2. Пусть C — линейный $[n,k]_2$-код. Доказать, что либо веса всех кодовых слов в C четны, либо кодовые слова четного веса образуют в C линейный $[n,k-1]_2$-подкод.

3. Пусть C — самоортогональный двоичный линейный код. Доказать, что все его кодовые слова имеют четный вес.

4. Доказать, что троичный линейный код самоортогонален тогда, и только тогда, когда веса всех его кодовых слов кратны 3.

Граница Плоткина.

Теорема. (Граница Плоткина). Для линейного q-ричного $[n,k,d]_q$-кода C, параметры которого удовлетворяют неравенству $d/n > (q-1)/q$, имеет место

$$|C| \leqslant \frac{d}{d - \frac{q-1}{q}n}.$$

Доказательство. Шаг 1. В линейном коде над полем $\mathbb{F}_q$ либо все кодовые векторы начинаются с 0, либо ровно $1/q$-я их часть начинается с 0.

Шаг 2. Сумма весов всех кодовых векторов не превосходит $n(q-1)q^{k-1}$.

Шаг 3. Поскольку минимальный вес не превосходит среднего значения из всех ненулевых весов,

$$d \leqslant \frac{n(q-1)q^{k-1}}{q^k - 1}.$$