ПРИКЛАДНАЯ ТЕОРИЯ КОДИРОВАНИЯ

УДК 519.72

СРАВНЕНИЕ КОДА ГОЛЕЯ С АЛГЕБРОГЕОМЕТРИЧЕСКИМ КОДОМ

П. М. Ширяев

Московский государственный университет им. М. В. Ломоносова, г. Москва, Россия

Рассматриваются два двоичных кода, код Голея $\mathcal{G} = [23, 12, 7]_2$ и предложенный автором алгеброгеометрический код C, для кодирования информации в двоичном симметричном канале с шириной $W = 50\,\text{КБ/с}$, тактовой частотой кодера/декодера $1\,\text{ГГц}$, вероятностью битовой ошибки $p = 0{,}005$ и требуемой вероятностью успешного декодирования передаваемого кодового слова не менее 0,9999. Показывается, что оба кода подходят под эти условия и что скорость передачи по этому каналу информации, закодированной по коду C, примерно в 1,12 раз выше, чем для информации, закодированной по коду $\mathcal{G}$. Показано также, как за счёт выбора дивизора D и базиса $L(D)$ при построении кода C можно ускорить стандартный алгоритм декодирования.

Ключевые слова: *AG-код, код Голея, L-конструкция, эллиптическая кривая.*

DOI 10.17223/20710410/30/7

COMPARISON OF THE BINARY GOLAY CODE WITH THE ALGEBRO-GEOMETRIC CODE

P. M. Shiriaev

Lomonosov Moscow State University, Moscow, Russia

E-mail: shiryaev.petr@gmail.com

The binary Golay code $\mathcal{G} = [23, 12, 7]_2$ and a binary algebro-geometric code C, proposed by the author, are considered for coding information in a binary symmetric channel with bandwidth $W = 50\,\text{KB/s}$, coder/decoder clock rate $1\,\text{GHz}$, bit error ratio $p = 0.005$, and required decoding probability 0.9999. It is shown that both codes fit this channel and the code C rate is $12\,\%$ greater than the code $\mathcal{G}$ rate. It is also shown how you can increase the decoding speed of the standard decoding algorithm by a proper choice of a divisor D and the basis of $L(D)$ for constructing C. The decoding complexity of C is estimated and the message transmission durations for C and $\mathcal{G}$ are compared.

Keywords: *AG-code, Golay code, L-construction, elliptic curve.*

Введение

Введём следующие обозначения:

— $\mathcal{G} = [23, 12, 7]_2$ — двоичный код Голея;

— $\mathbb{F} = \mathbb{Z}_2[t]/(t^4 + t^3 + 1)$ — поле из 16 элементов;
— C — алгеброгеометрический $[24, 14, 10]_{16}$-код;
— $L(D) = \{0\} \cup \{f \in k(\mathbb{X})^* : (f) + D \geqslant 0\}$ — пространство Римана — Роха дивизора D кривой $\mathbb{X}$ над полем k;
— $l(D)$ — размерность $L(D)$.

Для передачи информации в некоторых случаях требуются коды с вероятностью успешного декодирования не меньше 0,9999 при условии вероятности битовой ошибки 0,005 — например, код Голея $\mathcal{G} = [23, 12, 7]_2$. Опишем способ построения алгеброгеометрического кода на эллиптической кривой с параметрами $[24, 14, 10]_{16}$, который удовлетворяет этому условию и быстрее $\mathcal{G}$ примерно в 1,12 раз. Для декодирования полученного кода воспользуемся стандартным алгоритмом [1, с. 279]. Алгоритм декодирования использует базис пространства Римана — Роха $L(D)$ дивизора D, где D задаётся при построении кода. Выбор в качестве базиса множества, построенного в [2, теорема 3.3], позволяет уменьшить количество арифметических операций в алгоритме декодирования по сравнению с произвольным базисом. Вычислим сложность алгоритма декодирования и проверим, что указанный алгоритм гарантирует требуемую вероятность успешного декодирования. Все рассматриваемые дивизоры являются приведёнными [2, с. 48] и имеют положительную степень, поэтому для вычисления размерности пространства Римана — Роха будем пользоваться равенством $l(D) = \deg D + 1$ [1, 2.2.23] (род эллиптической кривой равен 1).

В ходе построений используются определённые на кривой $\mathbb{X}$ рациональные функции, в частности базис $L(D)$. Стоит отметить, что при этом используются не сами функции, а наборы их значений на некотором множестве точек $\mathcal{P}$, иными словами, векторы длины $|\mathcal{P}|$.

1. Построение АГ-кода

Многочлен t^4+t^3+1 неприводим над $\mathbb{Z}_2$, следовательно, $\mathbb{F} = \mathbb{Z}_2[t]/(t^4+t^3+1)$ является полем $\mathrm{GF}(2^4)$. Элементы этого поля будем интерпретировать как числа в двоичной записи. В частности, $t^3 + t + 1 \to 1011_2 = 11$, $t^2 + t + 1 \to 0111_2 = 7$. Рассмотрим над $\mathbb{F}$ эллиптическую кривую

$$\mathbb{X} = \{\infty\} \cup \{(x, y) : y^2 + y = x^3 + 11x + 7\},\ x, y \in \mathbb{F}.$$

Введём обозначение $\mathcal{P} = \mathbb{X}\backslash\infty$. Отметим, что $\mathcal{P}$ разбивается на пары сопряжённых точек [3, def. 5]. В дальнейшем нам понадобится зависимость координат сопряжённой точки от координат исходной. По определению, для $\mathbb{X}$ эта зависимость устроена следующим образом.

Замечание 1. Если $P = (x_P, y_P)$, то $\overline{P} = (x_P, y_P + 1)$.

Зафиксируем дивизор [3, def. 31] $D = 10 \cdot \infty$. Составим матрицу

$$H_D = (f_i(P))_{\substack{i=1\ldots l(D),\\ P\in\mathcal{P}}}$$

где $\{f_i\}$ — базис $L(D)$, и рассмотрим задаваемый ею код

$$C = \{c \in \mathbb{F}^{24} : H_D \cdot c = 0\}.$$

Согласно [1, теоремы 4.1.1 и 4.1.25], C является кодом с параметрами

$$[24, 24 - \deg(D), \deg(D)]_{16} = [24, 14, 10]_{16}.$$

Поскольку минимальное расстояние C равно 10, рассматриваемый код может исправить до четырёх произвольных ошибок. Ниже приведён алгоритм, исправляющий такое количество ошибок.

2. Скорость декодирования

По следствию 4.1.42 [1], стандартный алгоритм исправления ошибок [1, с. 279] позволяет декодировать до $\min(l(D'), \deg(D) - \deg(D')) - 1 = 4$ ошибок.

Опишем стандартный алгоритм. В качестве вспомогательного дивизора зафиксируем $D' = 5 \cdot \infty$. Для каждого шага алгоритма вычислим требуемое количество операций над элементами $\mathbb{F}$. Для краткости будем записывать операции как тройку чисел: количество (сложений, умножений, делений). Входящий вектор обозначим v, искомый вектор ошибок — e.

1) Обозначим базисы $L(D')$ и $L(D - D')$ как $\{g_j\}$ и $\{h_k\}$ соответственно. По теореме 3.3 из [2] для рассматриваемых дивизоров можно выбрать следующие базисы:

$$\{f_i\} = \{1, x, x^2, x^3, x^4, x^5, y, yx, yx^2, yx^3\}; \tag{1}$$

$$\{g_j\} = \{h_k\} = \{1, x, x^2, y, yx\}. \tag{2}$$

Замечание 2. Матрица $H_{D'} = (g_j(P))_{\substack{j=1\ldots l(D') \\ P \in \mathcal{P}}}$, используемая в алгоритме, получается из H_D вычёркиванием строк, поскольку базис (2) является подмножеством базиса (1).

2) Вычислим синдромы (вектор и матрицу)

$$s = H_D \cdot v = \left(\sum_{P \in \mathcal{P}} f_i(P) v_P \right)_{i=1\ldots l(D)}, \quad S = \left(\sum_{P \in \mathcal{P}} g_j(P) h_k(P) v_P \right)_{\substack{j=1\ldots l(D') \\ k=1\ldots l(D')}}.$$

Как упоминалось выше, $\mathcal{P}$ является объединением пар сопряжённых точек. Используем это для ускорения вычисления синдромов. Для упрощения записи введём обозначение $\alpha_n(P) = \begin{pmatrix} 1 \\ x_P \\ \cdots \\ x_P^n \end{pmatrix}$ и рассмотрим два столбца H_D, соответствующие паре сопряжённых точек P и $\overline{P}$. Их вклад в s выглядит следующим образом:

$$\begin{pmatrix} \alpha_5(P) & \alpha_5(P) \\ y_P \alpha_3(P) & (y_P + 1)\alpha_3(P) \end{pmatrix} \begin{pmatrix} v_P \\ v_{\overline{P}} \end{pmatrix} = \begin{pmatrix} \alpha_5(P) \\ y_P \alpha_3(P) \end{pmatrix} (v_P + v_{\overline{P}}) + \begin{pmatrix} 0 \\ \alpha_3(P) \end{pmatrix} v_{\overline{P}} =$$

$$= \begin{pmatrix} (v_P + v_{\overline{P}})\alpha_5(P) \\ (v_P + v_{\overline{P}}) y_P \alpha_3(P) \end{pmatrix} + \begin{pmatrix} 0 \\ v_{\overline{P}} \alpha_3(P) \end{pmatrix}.$$

Векторы $\alpha_5(P)$ и $y_P\alpha_3(P)$ составляют столбец матрицы H_D и потому известны; $\alpha_3(P)$ является подвектором $\alpha_5(P)$ и потому тоже известен. Отметим, что в $\alpha_3(P)$ и $\alpha_5(P)$ количество отличных от 1 координат не превосходит 3 и 5 соответственно.

За одно сложение вычислим $(v_P + v_{\overline{P}})$. Теперь координаты $\begin{pmatrix} (v_P + v_{\overline{P}})\alpha_5(P) \\ (v_P + v_{\overline{P}}) y_P \alpha_3(P) \end{pmatrix}$ можно найти за $5 + 4$ умножений, а $\begin{pmatrix} 0 \\ v_{\overline{P}} \alpha_3(P) \end{pmatrix}$ — за 3 умножения. На вычисление суммы векторов нужно ещё 4 сложения. Таким образом, вклад точек P и $\overline{P}$ может быть найден за $(5, 12, 0)$ операций. Такие вычисления нужно проделать 12 раз и сложить все

результаты, что даёт итоговую сложность в $12 \cdot (5, 12, 0) + 11 \cdot (10, 0, 0) = (170, 144, 0)$ операций для вычисления s.

Из очевидного равенства $D = D' + (D - D')$ следует, что $g_j h_k \in L(D)$. Значит, можно разложить $g_j h_k$ по базису $L(D)$, что позволяет выразить S через s:

$$\sum_{P \in \mathcal{P}} g_j(P) h_k(P) = \sum_i a_i f_i(P) \Rightarrow$$
$$\Rightarrow S_{jk} = \sum_{P \in \mathcal{P}} g_j(P) h_k(P) v_P = \sum_{P \in \mathcal{P}} \sum_i a_i f_i(P) v_P = \sum_i a_i \sum_{P \in \mathcal{P}} f_i(P) v_P = \sum_i a_i s_i.$$

Посмотрим, какие именно функции требуется так выразить. Запишем

$$(g_j h_k) = \begin{pmatrix} 1 & x & x^2 & y & yx \\ x & x^2 & x^3 & yx & yx^2 \\ x^2 & x^3 & x^4 & yx^2 & yx^3 \\ y & yx & yx^2 & y^2 & y^2 x \\ yx & yx^2 & yx^3 & y^2 x & y^2 x^2 \end{pmatrix}.$$

Все функции, за исключением y^2, $y^2 x$, $y^2 x^2$, уже принадлежат (1), поэтому результаты их скалярного произведения с v являются компонентами s. Из уравнения кривой имеем

$$y^2 = y + x^3 + 11x + 7,$$
$$y^2 x = yx + x^4 + 11x^2 + 7x,$$
$$y^2 x^2 = yx^2 + x^5 + 11x^3 + 7x^2,$$

где правые части равенств являются линейными комбинациями функций из (1). Таким образом, для вычисления каждой линейной комбинации понадобится не более $(3, 2, 0)$ операций, что в сумме даёт $(9, 6, 0)$.

3) Найдём z — произвольное нетривиальное решение однородной линейной системы уравнений

$$z^T S = 0\,.$$

Решение будем искать методом Гаусса. Для системы размера 5×5 возьмём оценку в $(50, 50, 15)$ операций [4, с. 17].

4) Найдём множество $I = \{P : \sum z_j g_j(P) = 0, P \in \mathcal{P}\}$ нулевых компонент вектора $z^T H_{D'}$. Для этого воспользуемся разбиением $\mathcal{P}$ на пары сопряжённых точек.

Вычислить $\sum\limits_{j=1}^{5} z_j g_j(P)$ можно за $(4, 5, 0)$ операций. Воспользовавшись замечанием 1, запишем условие для $\overline{P}$ следующим образом:

$$\sum_{j=1}^{5} z_j g_j(\overline{P}) = 0 \Leftrightarrow \sum_{j=1}^{5} z_j g_j(P) = z_4 \cdot 1 + z_5 \cdot x_P.$$

Левая часть последнего равенства уже найдена при проверке точки P, а правую можно вычислить за $(1, 1, 0)$ операций. Таким образом, $z^T H_{D'}$ можно найти за $12 \cdot (5, 6, 0) = (60, 72, 0)$ операций.

5) Найдём e как решение системы

$$\sum_{i=1}^{10} f_i(P) e_i = s_i, P \in I\,.$$

Для оценки размера системы докажем следующую лемму.

Лемма 1. $|I| \leqslant 5$.

Доказательство. По теореме 4.1.25 в [1] $H_{D'}$ можно рассматривать как порождающую матрицу для кода $C_{D'}$ с параметрами

$$[24, \deg(D'), 24 - \deg(D')]_{16} = [24, 5, 19]_{16}.$$

Тогда для всякого $z \in \mathbb{F}^5$ вектор $z^T H_{D'}$ является словом $C_{D'}$ и имеет не меньше 19 ненулевых компонент. Следовательно, нулевых компонент не больше $24 - 19$. Поскольку I есть обозначение множества нулевых компонент $z^T H_{D'}$, получаем $|I| \leqslant 24 - 19 = 5$. ■

Эту систему также будем решать методом Гаусса. По лемме 1 матрица системы состоит из не более чем 5 столбцов. Для системы размера 10×5 воспользуемся оценкой $(375, 375, 55) - (50, 50, 0) = (325, 325, 55)$ операций; $(375, 375, 55)$ — оценка [4, с. 17] для системы 10×10, при этом 5 столбцов можно считать нулевыми, поэтому на обнуление каждого элемента не потребуется тратить как минимум одно сложение и одно умножение, что и даёт $(50, 50, 0)$.

Суммарное количество операций равно

$$(170, 144, 0) + (9, 6, 0) + (50, 50, 15) + (60, 72, 0) + (325, 325, 55) = (614, 597, 70).$$

3. Сравнение C и $\mathcal{G}$

Убедимся, что код Голея удовлетворяет требованиям данной модели. Вероятность правильного декодирования сообщения не меньше вероятности того, что при передаче произойдёт не более 3 ошибок, то есть

$$\sum_{t=0}^{3} C_{23}^t p^t (1-p)^{23-t} \approx 0{,}999997.$$

Передадим сообщение размером $T = 50\,\text{КВ}$. Если передавать без кодирования, то на передачу уйдёт 1 с. Если же передавать кодом $\mathcal{G}$, то потребуется $23/12 \approx 1{,}92$ с.

Для кодирования одного блока требуется 84 операции `xor`, по одной на единичный элемент порождающей матрицы [5, рис. 2.13], поэтому на кодирование T при скорости вычислений 1 ГГц уйдёт $\dfrac{84 \cdot 50 \cdot 8 \cdot 2^{10}}{23} \cdot 10^{-9}$ с.

Декодирование можно осуществлять, например, методом вылавливания ошибок [1, с. 267–268]. Метод требует выполнить до $2 \cdot 23$ операций `xor`. Значит, на декодирование потребуется $\dfrac{46 \cdot 50 \cdot 8 \cdot 2^{10}}{23} \cdot 10^{-9}$ с. Общее время кодирования и декодирования составляет $\dfrac{130 \cdot 50 \cdot 8 \cdot 2^{10}}{23} \cdot 10^{-9} \approx 0{,}002\,\text{с}$. Общее время на передачу сообщения, таким образом, составит $\approx 1{,}922\,\text{с}$.

Поскольку для передачи информации достаточно хранить порождающую матрицу размера $23 \cdot 12 = 276$ бит [5, рис. 2.13], затратами памяти можно пренебречь.

Теперь проделаем то же самое для кода C. Сначала убедимся, что вероятность успешного декодирования достаточно высока. Поскольку каждый элемент $\mathbb{F}$ записывается четырьмя битами, число $\hat{p} = 1 - (1-p)^4$ естественно интерпретировать как вероятность возникновения ошибки в элементе $\mathbb{F}$. Тогда вероятность правильного декодирования сообщения произвольным кодом с $n = 24$, $d \geqslant 9$ над GF(16) равна

$$\sum_{t=0}^{4} \mathrm{C}_{24}^t \, \hat{p}^t (1-\hat{p})^{24-t} \approx 0{,}999904.$$

Будем считать, что известны таблицы умножения и деления элементов $\mathbb{F}$ (операция сложения элементов поля эквивалентна `xor` и потому выполняется за одну операцию без таблиц). Далее, поскольку конкретный вид порождающей матрицы C значения не имеет, будем считать, что в ней выделена единичная матрица, то есть $G_C = (I_{14}|A)$ для некоторой матрицы A.

Передадим сообщение размера T с помощью C. На передачу уйдёт $24/14 \approx 1{,}71$ с. Сложность декодирования одного блока длины $24{\cdot}4$ бита вычислена выше и составляет $(614, 597, 70)$ операций сложения, умножения и деления элементов поля соответственно. Кодирование информационного сообщения k заключается в вычислении $k^T A$, которое можно выполнить за $(140, 140, 0)$ операций. Таким образом, время на кодирование и декодирование сообщения составит $\dfrac{50 \cdot 8 \cdot 2^{10}}{96}(614{+}597{+}70{+}140{+}140){\cdot}10^{-9} \approx 0{,}007$ с. Общее время на передачу сообщения составит $\approx 1{,}717$ с.

Для кодирования требуется хранить матрицу A размером $14 \cdot 10 = 140$ байт, для декодирования, согласно замечанию 2, — матрицу H_D размером $24 \cdot 10 = 240$ байт. Дополнительно на таблицы умножения и деления достаточно выделить $2 \cdot 16 \cdot 16$ байт, поэтому затратами памяти для C также можно пренебречь.

Таким образом, в рамках данной модели код C оказывается быстрее в $\dfrac{1{,}922}{1{,}717} \approx$ $\approx 1{,}12$ раз при пренебрежимо малых для обоих кодов затратах памяти.

ЛИТЕРАТУРА

1. *Влэдуц С. Г., Ногин Д. Ю., Цфасман М. А.* Алгеброгеометрические коды. Основные понятия. М.: МЦНМО, 2002. 504 с.
2. *Семеновых Д. Н.* О теоретико-числовых вопросах в теории кодирования: дис. ... канд. физ.-мат. наук. М.: МГУ, 2005. 60 с.
3. *Menezes A., Wu Y.-H., and Zuccherato R.* An Elementary Introduction to Hyperelliptic Curves. Research report. Waterloo: Faculty of Mathematics, University of Waterloo, 1996. No. 19. 35 p.
4. *Богачев К. Ю.* Практикум на ЭВМ. Методы решения линейных систем и нахождения собственных значений. М.: МГУ, 1998. 79 с.
5. *Мак-Вильямс Ф. Дж., Слоэн Н. Дж. А.* Теория кодов, исправляющих ошибки. М.: Связь, 1979. 774 с.

REFERENCES

1. *Vleduts S. G., Nogin D. Yu., Tsfasman M. A.* Algebrogeometricheskie kody. Osnovnye ponyatiya [Algebro-Geometric Codes. Basic Concepts]. Moscow, MCCME Publ., 2002. 504 p. (in Russian)
2. *Semenovykh D. N.* O teoretiko-chislovykh voprosakh v teorii kodirovaniya [On Number-Theoretic Problems in Coding Theory]. PhD Thesis, Moscow, MSU Publ., 2005. 60 p. (in Russian)
3. *Menezes A., Wu Y.-H., and Zuccherato R.* An Elementary Introduction to Hyperelliptic Curves. Research report. Waterloo, Faculty of Mathematics, University of Waterloo, 1996, no. 19. 35 p.
4. *Bogachev K. Yu.* Praktikum na EVM. Metody resheniya lineynykh sistem i nakhozhdeniya sobstvennykh znacheniy [Workshop on the Computer. Methods for Linear Systems Solving and Eigenvalues Finding]. Moscow, MSU Publ., 1998. 79 p. (in Russian)
5. *Mak-Vil'yams F. Dzh., Sloen N. Dzh. A.* Teoriya kodov, ispravlyayushchikh oshibki [The Theory of Error-Correcting Codes]. Moscow, Svyaz' Publ., 1979. 774 p. (in Russian)