数理解析研究所講究録
1060 巻 1998 年 137-144

Introduction to the theory of error-correcting codes

Tsuyoshi Uehara (上原　健)

Saga University

0. Introduction

In this note we introduce the basic theory of error-correcting codes, showing especially the constructions and decoding processes of Hamming codes and Reed-Solomon codes. We are also concerned with simply constructed algebraic-geometric codes, and give a result on their minimum distances.

1. Codes, Encoding and Decoding

In this section we denote by F a finite set and call it an *alphabet.* Fix positive integers k, n with $k \leq n$. Let $S \subset F^k$ and $C \subset F^n$ and call them a *source* and a *code*, respectively. An *encoding rule* ε is a bijective mapping from S to C, and a *decoding rule* δ is a mapping from a set C' with $C \subset C' \subset F^n$ to C. The following diagram provides a rough idea of an information transmission system.

$$S \xrightarrow{\varepsilon} C \longrightarrow C' \xrightarrow{\delta} C \xrightarrow{\varepsilon^{-1}} S$$

A sender sends a source symbol $x \in S$ through a channel to a receiver. If an error is caused by noise on the channel, the receiver will receive a wrong symbol. Therefore using an encoder one transletes the symbol x into a codeword $w = \varepsilon(x) \in C$. From the received word $w' \in C'$ a decoder induces a decoded word $w'' = \delta(w') \in C$, and finally the receiver get the symbol $x'' = \varepsilon^{-1}(w'')$. To see a simple example we set

$$F = S = \{0, 1\}, \quad C = \{(0,0,0), (1,1,1)\}, \quad C' = F^3.$$

For instance suppose that 0 is sent and is encoded as $(0,0,0)$, and that the received word is $(0,0,1)$. The decoder decides by majority that $(0,0,0)$ is sent, and gives 0 to the receiver.

Those codes constructed for the above purpose are called *error-correcting codes.*

2. Linear Codes

In what follows $F = GF(q)$ denotes the finite field with q elements. For a fixed positive integer n we consider the linear space

$$F^n = \{(x_1, \cdots, x_n) : x_i \in F\}.$$

If $C \subset F^n$ is a subspace of F^n then we call C a q-ary *linear code* of length n. When referring the length and the dimension $k = \dim C$ of a q-ary linear code C we call C a q-ary (n, k) code. Any element of a code is called a *codeword*. The information rate $R(C)$ of C is defined as k/n. If C is an (n, k) code, we may take a linear mapping ε from F^k to C as an encoding rule. Hence there is no problem about encoding for linear codes.

For any vectors $x = (x_1, \cdots, x_n)$ and $y = (y_1, \cdots, y_n)$ in F^n we denote by $d(x, y)$ the number of indices i, $1 \leq i \leq n$ such that $x_i \neq y_i$ and call it the *Hamming distance* of x and y. The *minimum distance* $d(C)$ of a linear code $C \subset F^n$ is defined as

$$d(C) = \min\{d(x, y) : x, y \in C, \ x \neq y\} = \min\{d(x, 0) : 0 \neq x \in C\}.$$

The minimum distance of a linear code has a close relation to its ability of error-correcting. In fact, if a codeword $x \in C$ is sent and is received as $x' \in F^n$ and if

$$d(x, x') \leq t(C) = [(d(C) - 1)/2],$$

then any codeword $y \in C$ with $y \neq x$ satisfies $d(y, x') > d(x, x')$, i.e. x is the nearest codeword to x'. This implies that C can correct at most $t(C)$ errors. We say also that C is $t(C)$-error-correcting.

Roughly speaking, an (n, k) code C is "good" if both $R(C)$ and $d(C)/n$ are large. Various bounds for $d(C)$ are known. In particular, Singleton's bound [6] (cf. [5]) says that

$$d(C) \leq n - k + 1.$$

If the equality holds in the above, C is called a *maximum distance separable* (*MDS*) *code.*

A main purpose of study of error-correcting codes is to construct explicitly a class of linear codes C with k/n, $d(C)/n$ large and with good decoding algorithms. For the general theory of error-correcting codes, for intance see [5], [7].

3. Hamming Codes

To explain a method of decoding of linear codes, we give a simple example in this section. We denote by $L[v_1, \cdots, v_m] \subset F^n$ the subspace of F^n spanned by vectors $v_1, \cdots, v_m \in F^n$.

Let $F = GF(2)$ and define a binary linear code C by

$$C = L[(1,0,1,0,1,0,1),\ (0,1,1,0,0,1,1),\ (0,0,0,1,1,1,1)]^{\perp}.$$

Namely, a vector $(x_1, \cdots, x_7) \in F^7$ is a codeword of C if and only if

$$x_1 + x_3 + x_5 + x_7 = x_2 + x_3 + x_6 + x_7 = x_4 + x_5 + x_6 + x_7 = 0.$$

We easily see that C is a binary $(7,4)$ code and has the following 16 codewords.

$$\begin{array}{llll}
(0,0,0,0,0,0,0), & (1,0,0,0,0,1,1), & (0,1,0,0,1,0,1), & (1,1,0,0,1,1,0), \\
(0,0,1,0,1,1,0), & (1,0,1,0,1,0,1), & (0,1,1,0,0,1,1), & (1,1,1,0,0,0,0), \\
(0,0,0,1,1,1,1), & (1,0,0,1,1,0,0), & (0,1,0,1,0,1,0), & (1,1,0,1,0,0,1), \\
(0,0,1,1,0,0,1), & (1,0,1,1,0,1,0), & (0,1,1,1,1,0,0), & (1,1,1,1,1,1,1).
\end{array}$$

Then the minimum distance $d(C)$ of C is equal to 3 and hence C is single-error-correcting. This code is called a *Hamming code.* Let $x' = (x'_1, \cdots, x'_7)$ be a received word of a codeword $x \in C$. We first compute the three values

$$s_1 = x'_1 + x'_3 + x'_5 + x'_7, \quad s_2 = x'_2 + x'_3 + x'_6 + x'_7, \quad s_3 = x'_4 + x'_5 + x'_6 + x'_7$$

in F which are called *syndromes* of x'. Next we compute $r = s_1 + 2s_2 + 4s_3$ in the ordinary way. The decoding is performed as follows: If $r = 0$ we conclude that x' has no error. If $r \neq 0$ we correct the r-th component x'_r. The trick is simple. One can see that x'_r is a summand only of those sums s_i equal to 1. For example suppose that $x' = (1,0,1,1,1,0,0)$ is a received word. We compute the syndromes $s_1 = 1$, $s_2 = 1$, $s_3 = 0$. Hence $r = 3$. We decode x' to $(1,0,0,1,1,0,0)$.

4. Reed-Solomon Codes

We introduce an important linear code which is now practically used.

Denote by $a_1, a_2, \cdots, a_q \in F$ the elements of $F = GF(q)$. For every polynomial $f \in F[X]$ over F we put

$$\psi(f) = (f(a_1), f(a_2), \cdots, f(a_q)) \in F^q.$$

It is clear that $L[\psi(1), \psi(X), \cdots, \psi(X^{q-1})] = F^q$. We define for $1 \leq k \leq q-1$ a linear code

$$C = C(q,k) = L[\psi(1), \psi(X), \cdots, \psi(X^{q-k-1})]^{\perp}$$

and call it a *Reed-Solomon* (RS) *code.* Then C is a q-ary (q,k) code. It is easy to see that

$$C = L[\psi(1), \psi(X), \cdots, \psi(X^{k-1})].$$

By Singleton's bound we have $d(C) \leq q-k+1$. On the other hand, if $x = \psi(f) \in C$ with $f \in F[X]$ then $\deg f \leq k-1$ and hence $d(x,0) \geq q-k+1$. This means that $d(C) \geq q-k+1$. Thus $d(C) = q-k+1$ and so C is an MDS code. For example the minimum distance of $C(8,4)$ is equal to 5, i.e. it is double-error-correcting.

We write $GF(8) = \{0, 1, \alpha, \cdots, \alpha^6\}$ with $\alpha^3 = \alpha + 1$, and put $a_1 = 0, a_2 = 1, a_3 = \alpha, \cdots, a_8 = \alpha^6$. To explain a method of decoding of RS codes, let $C = C(8,4)$ and suppose that

$$w' = (1, 1, \alpha^3, 1, 1, 1, 0, 1)$$

is a received word of a codeword $w \in C$. We write $w' = w + e$ with $e = (e_1, \cdots, e_8)$. For any vector $z \in GF(8)^8$ we define

$$s_i(z) = (z, \psi(X^i)),$$

where $(x,y) = x_1y_1 + \cdots + x_ny_n$ denotes the inner product of $x = (x_1, \cdots, x_n)$ and $y = (y_1, \cdots, y_n)$. By definition

$$s_i(w') = s_i(w) + s_i(e) = s_i(e)$$

is valid if $0 \leq i \leq 3$. As the first step of decoding we compute

$$s_0(e) = s_0(w') = \alpha^3, \quad s_1(e) = s_1(w') = \alpha^3, \quad s_2(e) = s_2(w') = 0, \quad s_3(e) = s_3(w') = \alpha^2.$$

These values are also called syndromes of w'. Assuming that the number of errors is at most two, i.e. $d(e,0) \leq 2$, we put

$$g(X) = \prod_{e_i \neq 0} (X - a_i) = c_0 + c_1 X + c_2 X^2$$

with c_0, c_1, c_2 unknown. Since $e_i g(a_i) = 0$ for any i we have

$$(e, \psi(g(X)X^j)) = \sum_{i=1}^{8} e_i g(a_i) a_i^j = c_0 s_j(e) + c_1 s_{j+1}(e) + c_2 s_{j+2}(e) = 0.$$

We secondly solve the simultaneous equations

$$c_0 s_0(e) + c_1(e) s_1(e) + c_2 s_2(e) = c_0 s_1(e) + c_1 s_2(e) + c_2 s_3(e) = 0$$

to get $c_0 = c_1 = \alpha^6$, $c_2 = 1$. Hence we see that $g(X) = X^2 + \alpha^6 X + \alpha^6 = (X - \alpha)(X - \alpha^5)$. This means that the errors occured at third and seventh components, i.e. $e_3, e_7 \neq 0$. We finally find the values e_3, e_7. Solving

$$s_0(e) = e_3 + e_7 = \alpha^3, \qquad s_1(e) = e_3 \alpha + e_7 \alpha^5 = \alpha^3$$

we obtain $e_3 = \alpha, e_7 = 1$. Thus we decode w' as

$$w = w' - e = (1, 1, \alpha^3, 1, 1, 1, 0, 1) - (0, 0, \alpha^3, 0, 0, 0, 1, 0) = (1, 1, 1, 1, 1, 1, 1).$$

There are various methods of fast decoding of RS codes. We remark that the length of the RS code $C(q, k)$ does not exceed q. This is a defect of RS codes.

5. Algebraic-Geometric Codes

In 1981 Goppa [2], [3] presented a new class of linear codes applying the theory of algebraic curves, which are called *algebraic-geometric* (AG) *codes.* Let V be a non-singular projective curve over $F = GF(q)$ of genus g. We take a positive integer n less than the number of F-rational points of V. Then Goppa's theorem guarantees the existence of q-ary linear codes C such that

$$k(C) + d(C) \geq n + 1 - g,$$

where $k(C), d(C)$ denote the dimension, the minimum distance of C, respectively. It is possible to find many q-ary AG codes each with length larger than q. For the general theory of AG codes for instance see [4], [7], [8].

Recently, Feng and Rao [1] proposed a method of simple construction of AG codes. We are concerned here with such AG codes. Let $X_1, \cdots, X_m$ be independent indeterminates and put $X = (X_1, \cdots, X_m)$. Denote by $F[X] = F[X_1, \cdots, X_m]$ the polynomial ring over $F = GF(q)$. We define

$$\Gamma = \{(z_1, \cdots, z_m) \,:\, z_i \in \boldsymbol{Z},\ z_i \geq 0\}.$$

For every $a = (a_1, \cdots, a_m) \in \Gamma$ we put $X^a = X_1^{a_1} \cdots X_m^{a_m}$. We further consider a *monomial ordering* $\prec$ on X which is a total ordering of the set $\{X^a \,:\, a \in \Gamma\}$ such that

$$(1)\ \ X^a \neq 1 \Longrightarrow 1 \prec X^a, \qquad (2)\ \ X^a \prec X^{a'} \Longrightarrow X^{a+b} \prec X^{a'+b}\ \ (b \in \Gamma).$$

Take n distinct points $P_1, \cdots, P_n \in F^m$ and put

$$D = \{P_1, P_2, \cdots, P_n\}.$$

For any polynomial $f \in F[X]$ we define

$$\psi(f) = (f(P_1), f(P_2), \cdots, f(P_n)) \in F^n.$$

Then ψ is a surjective linear mapping from $F[X]$ to F^n.

Proposition 1. *There exist points $b(1), \cdots, b(k) \in \Gamma$ such that*

$$\{\psi(X^a) \,:\, a \in \Gamma_{b(1)} \cup \cdots \cup \Gamma_{b(k)}\}$$

is a basis of F^n, where $\Gamma_b = \{a \in \Gamma \,:\, a_i \leq b_i\ (1 \leq i \leq m)\}$ for any $b = (b_1, \cdots, b_m) \in \Gamma$.

We now state the definition of AG codes by Feng and Rao. We fix a monimial ordering $\prec$ on X and write

$$B = \{X^a \,:\, a \in \Gamma_{b(1)} \cup \cdots \cup \Gamma_{b(k)}\} = \{f_1, f_2, \cdots, f_n\},$$

where $f_1 \prec f_2 \prec \cdots \prec f_n$. Let $B_l = \{f_1, \cdots, f_l\}$ with $1 \leq l \leq n-1$. For any k, $1 \leq k \leq n$ we define a q-ary linear code $C(D, B_{n-k})$ by

$$C(D, B_{n-k}) = L[\psi(f_1), \psi(f_2), \cdots, \psi(f_{n-k})]^\perp$$

and call it an AG code in the sense of Feng-Rao. This is a q-ary (n, k) code.

In the rest of this note we will give a result on the minimum distances of AG codes. From now on let $m = 2$. For any $x \in F$ we denote by $m(x)$ the number of points $(a, b) \in D$

such that $a = x$. For simplicity we suppose that either $m(x) = 0$ or $m(x) = s$ holds for every $x \in F$, where s is a fixed positive integer. We also denote by t the number of $x \in F$ such that $m(x) = s$. In this case we have $n = st$ and we may take B as

$$B = \{X_1^i X_2^j : 0 \le i \le t-1, 0 \le j \le s-1\} = \{f_1, f_2, \cdots, f_n\}.$$

By the definition of B there are elements $c_i \in F$ such that

$$\psi(X_2^s) = \sum_{i=1}^{n} c_i \psi(f_i).$$

For any polynomials $g, h \in F[Y]$ in one indeterminate Y let $M(g,h)$ be the number of those points $(a,b) \in D$ such that $g(a) \neq 0$, $h(b) \neq 0$, and put

$$M_{i,j} = \min\{M(g,h) : g, h \in F[Y], \deg g = i, \deg h = j\}$$

for $i, j \ge 0$. If $f_l = X_1^i Y_2^j \in B$ then we define $M_l = M_{i,j}$.

Proposition 2. *Let $C = C(D, B_{n-k})$ with $1 \le k \le n$. If $f_{i_0} \prec X_2^s$ is valid where $i_0 = \max\{i : c_i \neq 0\}$, then*

$$d(C) \le d'(C) = \min\{M_1, \cdots, M_k\}.$$

We now consider a polynomial $H(X_1, X_2) = g(X_1) + h(X_2)$, where $g(Y), h(Y) \in F[Y]$. Suppose that $H(X_1, X_2)$ satisfies

(1) $r = \deg g \ge 1$, $s = \deg h \ge 1$ and $(r,s) = 1$,

(2) the number of solutions of $H(a, X_2) = 0$ is equal to 0 or s for any $a \in F$,

(3) the number of solutions of $H(X_1, b) = 0$ is equal to r for some $b \in F$.

Furthermore we put $D = \{(a,b) \in F^2 : H(a,b) = 0\}$. Then $m(x) = 0$ or $m(x) = s$. Hence in this case we can write $D = \{P_1, \cdots, P_n\}$ with $n = st$, where t is the number of $x \in F$ such that $m(x) = s$, and take $B = \{X_1^i X_2^j : 0 \le i \le t-1, 0 \le i \le s-1\}$.

Theorem. *Let the assumption be as above and put $C = C(D, B_{n-k})$ with $1 \le k \le n-1$. Then $d(C) = d'(C)$ is valid for a certain monomial ordering on (X_1, X_2).*

We remark that if (1), (2), (3) are all valid then for $f_l = X_1^i X_2^j$ we can compute

$$M_l = M_{i,j} = \begin{cases} (t-i)(s-j) & \text{when } i \ge t-r, \\ s(t-r-i) + r(s-j) & \text{when } i < t-r. \end{cases}$$

For example, when $q = p^2$, $H(X_1, X_2) = X_1^{p+1} + X_2^p + X_2$, the conditions (1), (2), (3) are all valid with $r = p+1$, $s = p$, $t = p^2 = q$. Hence $n = p^3$. In this case the associated AG codes $C(D, B_{n-k})$ are called *Hermitian codes.* Such a Hermitian code is originally one of Goppa's AG codes induced from the projective curve V defined by $X_1^{p+1} + X_0X_2^p + X_0^pX_2 = 0$. The minimum distances of Hermitian codes are known. Our theorem covers this result. In particular let $q = 9$. Then $C = C(D, B_7)$ is a $(27, 20)$ code. Since the genus of V is $3 = (4-1)(3-1)/2$, by Goopa's theorem we have $d(C) \geq 5 = 27 + 1 - 3 - 20$. On the other hand by Theorem we can deduce $d(C) = 6$.

References

[1] G.-L. Feng and T. R. N. Rao, A simple approach for construction of algebraic-geometric codes from affine plane curve, IEEE Trans. Info. Theory, 40 (1994) 1003 - 1012

[2] V. G. Goppa, Codes on algebraic curves, Soviet Math. Dokl, 24 (1981) 170 - 172.

[3] V. G. Goppa, Codes and information, Russ. Math. Surveys, 39:1 (1984) 87 - 141.

[4] V. G. Goppa, Geometry and codes, Kluwer Acad. Publ., 1988.

[5] F. J. MacWilliams and N. J. A. Sloane, The theory of error-correcting codes, North-Holland, 1977.

[6] R. C. Singleton, Maximum distance q-nary codes, IEEE Trans. Info. Theory, 10 (1964) 116 - 118.

[7] M. A. Tsfasman and S. G. Vlăduţ, Algebraic-geometric codes, Kluwer Acad. Publ., 1991.

[8] H. van Lint and G. van der Geer, Introduction to coding theory and algeraic geometry, Birkäuser Verlag, 1988.

Department of information science
Faculty of Science and Engineering
Saga University
Saga 840-8502, Japan
email : uehara@is.saga-u.ac.jp