

代数系と符号理論 第9回板書プリント

植松友彦 2007.6.1

1 2元対称通信路と復号誤り確率の限界

図1のように送信された2元シンボル0,1が確率 p で誤って受信される通信路のモデルを2元対称通信路 (Binary Symmetric Channel (BSC)) と呼ぶ。BSC で生じる誤りは送信された符号語あ

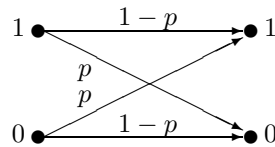


図 1: 2 元対称通信路

るいは送信シンボルには依存しない。従って、符号長 n の線形ブロック符号 C を用いた場合、符号語中に j 個の誤りが生じる確率は、

$$P[n, j] = \binom{n}{j} p^j (1-p)^{n-j}$$

となる。また、一方、符号 C が最小距離 d を持てば、受信語中の $t = \lfloor (d-1)/2 \rfloor$ 個までの誤りは訂正できる。従って、復号に失敗する確率は $t+1$ 個以上の誤りが生じたときなので、次の定理を得る。

定理 3.2.1 BSC に符号長 n 、最小距離 d のブロック符号を用い、 $t = \lfloor (d-1)/2 \rfloor$ 個までの誤りを訂正したときの復号失敗確率は、

$$P_{fail} = \sum_{j=t+1}^n \underbrace{\binom{n}{j} p^j (1-p)^{n-j}}_{t+1 \text{ 個以上の誤りの生じる確率}} \approx \binom{n}{t+1} p^{t+1} (1-p)^{n-t-1}$$

で与えられる。

2 リード・ソロモン符号とその応用

2.1 Singleton 限界

定理 5.1.1 (Singleton 限界) C を最小距離 d の (n, k) 符号とする。そのとき、

$$d \leq n - k + 1$$

である。また、等号が成立する符号を「最大距離分離符号」という。

2.2 q 元線形符号 (教科書 p.2)

定義 1.1.1 (q 元線形ブロック符号) 有限体 F_q 上の符号長 n , 情報記号数 k の (n, k) 線形ブロック符号とは、ベクトル空間 F_q^n の k 次元部分空間である。すなわち、 $C \subset F_q^n$ が線形符号ならば、

$$c_i \in C \text{ かつ } c_j \in C \implies c_i + c_j \in C$$

および

$$c_i \in C \text{ かつ } f \in F_q \implies fc_i \in C$$

が成り立つ。ここで、零ベクトルは常に符号語であり、符号語の総数は、

$$|C| = q^k$$

に等しい。

2.3 リード・ソロモン (RS) 符号

定義 5.1.1 (Reed-Solomon (RS) 符号) $x_1, \dots, x_n$ を有限体 F_q の相異なる元とする。但し、 $n \leq q$ である。 $k \leq n$ であるような k に対し、 $F_q[x]$ における次数が k 次未満の多項式の集合

$$P_k = \{f(x) \in F_q[x] : \deg(f(x)) < k\}$$

を考える。このとき、RS 符号は、以下の符号語の全体からなる。

$$(f(x_1), f(x_2), \dots, f(x_n)) \in F_q^n \quad f(x) \in P_k$$

RS 符号の符号長が n であることは明らかである。また、

$$c_1 = (f_1(x_1), \dots, f_1(x_n)), \quad c_2 = (f_2(x_1), \dots, f_2(x_n))$$

に対し、 $a, b \in F_q$ ならば、 $g(x) = af_1(x) + bf_2(x)$ に対し、

$$\begin{aligned} (g(x_1), \dots, g(x_n)) &= (af_1(x_1) + bf_2(x_1), \dots, af_1(x_n) + bf_2(x_n)) \\ &= a(f_1(x_1), \dots, f_1(x_n)) + b(f_2(x_1), \dots, f_2(x_n)) \\ &= ac_1 + bc_2 \end{aligned}$$

であるので、RS 符号は線形符号である。

注意 $n \leq q$ から 2 元 RS 符号は符号長 2 以下となり、意味を持たない。従って、実用的な RS 符号は、 $q = 2^m$ の場合であり、符号長を $n = q - 1$ と選び、 α を F_q の原始元として、

$$x_1 = 1, x_2 = \alpha, x_3 = \alpha^2, \dots, x_n = \alpha^{n-1}$$

とすることが多い。実際には、 $m = 8$ と選び、 F_{2^8} 上の符号長 $n = 255$ の RS 符号がよく用いられている。

P_k に含まれる多項式の係数は k 個あるから、 F_q 上の k 次元のベクトル空間をなす。ここで、相異なる 2 つ多項式 $f(x), g(x) \in P_k$ は同一の符号語を生成しない。なぜなら、もし生成すると

$$(f(x_1), \dots, f(x_n)) = (g(x_1), \dots, g(x_n)) \iff (f(x_1) - g(x_1), \dots, f(x_n) - g(x_n)) = 0$$

であり、これは k 次未満の多項式 $f(x) - g(x)$ が $n \geq k$ 個の相異なる根を持つことを意味し、定理 2.2.2 に矛盾する。従って、符号の次元は k となる。また、定理 2.2.2 から $f(x) \in P_k$ の根は高々 $k - 1$ 個以下なので、零ベクトルではない符号語の重みは $n - k + 1$ 以上である。これと定理 5.1.1 から次の定理を得る。

定理 5.1.2 (n, k) RS 符号の最小距離は、 $n - k + 1$ である。

RS 符号の自然な符号化法の一つとして、 k 個の情報記号 $i_0, \dots, i_{k-1} \in F_q$ に対して、多項式

$$i(x) = i_{k-1}x^{k-1} + i_{k-2}x^{k-2} + \dots + i_0 \in P_k$$

を作り、この多項式に符号語

$$(i(x_1), \dots, i(x_n))$$

を対応させる方法がある。但し、これは組織符号ではない。

[例] F_5 上の RS 符号を考えよう。符号長を $n = 4$ とし、簡単の為

$$x_1 = 1, \quad x_2 = 2, \quad x_3 = 3, \quad x_4 = 4$$

とする。 $k = 2$ とし、情報多項式を $i(x) = x + 1$ とする。対応する符号語は、

$$(i(1), i(2), i(3), i(4)) = (1 + 1, 2 + 1, 3 + 1, 4 + 1) = (2, 3, 4, 0)$$

である。 1 と x は、それぞれ

$$1 \rightarrow (1, 1, 1, 1), \quad x \rightarrow (1, 2, 3, 4)$$

に符号化されるので、情報多項式 $i(x) = i_1x + i_0$ は

$$i_1x + i_0 \rightarrow i_0(1, 1, 1, 1) + i_1(1, 2, 3, 4) = (i_0 + i_1, i_0 + 2i_1, i_0 + 3i_1, i_0 + 4i_1)$$

に符号化される。

2 元符号と同様に q 元符号でも生成行列 G を定義することができる。この例では、符号の基底となるベクトルは、 $(1, 1, 1, 1)$ と $(1, 2, 3, 4)$ の 2 つであるから、定義 1.1.2 に従って、

$$G = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix}$$

とすれば良い。このとき、情報（メッセージ）を表すベクトル $i = (i_1, i_2) \in F_5^2$ に対応する符号語 $c \in F_5^4$ は、

$$c = iG$$

で与えられる。

5.2 リード・ソロモン符号の復号

ここでは、Sudan によって 97 年に提案された、リードソロモン符号の画期的な復号法について述べる。

符号語 c を送信し、 $r = (r_1, r_2, \dots, r_n) = c + e$ を受信語とする。このとき、誤りベクトル $e = (e_1, e_2, \dots, e_n) \in F_q^n$ のハミング重みは、

$$w(e) \leq t = \left\lfloor \frac{n-k}{2} \right\rfloor$$

を満足するとする。このとき、2 変数多項式 (受信語に対する補間多項式という)

$$Q(x, y) = Q_0(x) + yQ_1(x) \in F_q[x, y] \setminus \{0\}$$

であって、条件

1. $Q(x_i, r_i) = 0, \quad i = 1, 2, \dots, n$
2. $\deg(Q_0(x)) \leq n-1-t$
3. $\deg(Q_1(x)) \leq n-1-t-(k-1)$

を満足するものを一つ選ぶ。このとき次の定理が成り立つ。

定理 5.2.1 条件 1-3 を満足する非ゼロの多項式 $Q(x, y)$ が少なくとも 1 個存在する。

(証明) $n-k \geq 2t$ に注意すれば、未知数の個数について

$$(n-1-t+1) + (n-1-t-k+1+1) = 2n-2t-k+1 \geq 2n-(n-k)-k+1 = n+1$$

が得られ、式の数 n よりも大きいので、非ゼロの解を持つ。

Q.E.D.

定理 5.2.2 送信された符号語 c が多項式 $g(x)$ から $c = (g(x_1), g(x_2), \dots, g(x_n))$ によって生成され、(通信路で生じた) 誤りの個数が t 以下 (すなわち、 $w(e) \leq t$) ならば、

$$g(x) = -\frac{Q_0(x)}{Q_1(x)}$$

が成り立つ。

(証明)

$$c = (g(x_1), \dots, g(x_n)), \quad r = c + e$$

であり、 $w(e) \leq t$ とする。 $Q(x, y) = Q_0(x) + yQ_1(x)$ は、

$$Q(x_i, r_i) = Q(x_i, g(x_i) + e_i) = 0, \quad i = 1, \dots, n$$

を満足することに注意する。 $n-t$ 個以上のインデックス i について $e_i = 0$ であるから、そのようなインデックスに対しては、 $g(x_i) = r_i$ が成り立ち、 $Q(x_i, g(x_i)) = Q(x_i, r_i) = 0$ なので、1 変数多項式 $Q(x, g(x))$ は $n-t$ 個以上の根を有する。

一方、 $Q(x, g(x))$ の次数は高々 $n-1-t$ であるから、恒等的に $Q(x, g(x)) = 0$ でなければならず、

$$Q(x, g(x)) = Q_0(x) + g(x)Q_1(x) = 0$$

から、

$$g(x) = -\frac{Q_0(x)}{Q_1(x)}$$

を得る。

Q.E.D.

以上のことから、RS 符号の復号には、受信語に対応する補間多項式 $Q(x, y)$ を求めれば良いことが分かる。これをアルゴリズムにしたのが、次の復号法である。

まず、

$$\begin{aligned} Q_0 \text{の次数の最大値} &: l_0 = n - 1 - t \\ Q_1 \text{の次数の最大値} &: l_1 = n - 1 - t - (k - 1) \end{aligned}$$

とおく。また、 $Q_1(x)$ のことを「誤り位置多項式 (error locator polynomial)」とも呼ぶ。この理由は、

$$\begin{aligned} Q(x, y) &= Q_0(x) + yQ_1(x) \\ &= Q_1(x) \left(y + \frac{Q_0(x)}{Q_1(x)} \right) \\ &= Q_1(x)(y - g(x)) \end{aligned}$$

であり、条件 $Q(x_i, r_i) = 0$ から、 $r_i \neq g(x_i)$ なら $Q_1(x_i) = 0$ でなければならない。このことは、符号語の i 番目の成分 $g(x_i)$ に誤りが生じた場合、 $r_i \neq g(x_i)$ であるから $Q_1(x_i) = 0$ を意味し、 $Q_1(x)$ は全ての誤り位置 i に対応する x_i を根として持つ。

アルゴリズム 5.2.1

入力: 受信語 $r = (r_1, r_2, \dots, r_n)$

1. 次の一次方程式の非零の解を求める。

$$\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^{l_0} & r_1 & r_1 x_1 & \cdots & r_1 x_1^{l_1} \\ 1 & x_2 & x_2^2 & \cdots & x_2^{l_0} & r_2 & r_2 x_2 & \cdots & r_2 x_2^{l_1} \\ \vdots & \vdots & & & \vdots & \vdots & & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^{l_0} & r_n & r_n x_n & \cdots & r_n x_n^{l_1} \end{bmatrix} \begin{bmatrix} Q_{0,0} \\ Q_{0,1} \\ \vdots \\ Q_{0,l_0} \\ Q_{1,0} \\ Q_{1,1} \\ \vdots \\ Q_{1,l_1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

2. 多項式 $Q_0(x)$ と $Q_1(x)$ を

$$Q_0(x) = \sum_{j=0}^{l_0} Q_{0,j} x^j, \quad Q_1(x) = \sum_{j=0}^{l_1} Q_{1,j} x^j$$

によって定める。

3. $Q_0(x)$ が $Q_1(x)$ で割り切れるならば、

$$g(x) = -\frac{Q_0(x)}{Q_1(x)}$$

とし、 $(g(x_1), g(x_2), \dots, g(x_n))$ を送信符号語の推定とする。もし割り切れなければ、復号に失敗したとする。

[例 2] 先ほどの例の符号語 $c = (2, 3, 4, 0)$ に誤り $e = (3, 0, 0, 0)$ が生じて受信語 $r = (0, 3, 4, 0)$ が受信されたとする。この場合、連立方程式の係数行列は、

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \cdot 1 \\ 1 & 2 & 2^2 & 3 & 3 \cdot 2 \\ 1 & 3 & 3^2 & 4 & 4 \cdot 3 \\ 1 & 4 & 4^2 & 0 & 0 \cdot 4 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 4 & 3 & 1 \\ 1 & 3 & 4 & 4 & 2 \\ 1 & 4 & 1 & 0 & 0 \end{bmatrix}$$

となる。これを行基本変形すると

$$\begin{aligned} & \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 4 & 3 & 1 \\ 1 & 3 & 4 & 4 & 2 \\ 1 & 4 & 1 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 3 & 3 & 1 \\ 0 & 2 & 3 & 4 & 2 \\ 0 & 3 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 3 & 2 & 4 \\ 0 & 1 & 3 & 3 & 1 \\ 0 & 0 & -3 & -2 & 0 \\ 0 & 0 & -4 & -4 & -3 \end{bmatrix} \\ & \rightarrow \begin{bmatrix} 1 & 0 & 3 & 2 & 4 \\ 0 & 1 & 3 & 3 & 1 \\ 0 & 0 & -3 & -2 & 0 \\ 0 & 0 & 1 & 1 & 2 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 3 & 2 & 4 \\ 0 & 1 & 3 & 3 & 1 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & -3 & -2 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & -1 & -2 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 2 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \\ & \rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 & 4 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{aligned}$$

を得る。従って、与えられた連立方程式は、次の連立方程式と等価であり、

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 4 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} Q_{0,0} \\ Q_{0,1} \\ Q_{0,2} \\ Q_{1,0} \\ Q_{1,1} \end{bmatrix} = 0$$

の非自明な解は、例えば、 $Q_{1,1} = 1$ と置くことで、

$$\begin{bmatrix} Q_{0,0} \\ Q_{0,1} \\ Q_{0,2} \\ Q_{1,0} \end{bmatrix} = - \begin{bmatrix} 4 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

となる。従って、

$$Q_0(x) = -4 - x^2 = 1 + 4x^2, \quad Q_1(x) = -1 + x = 4 + x$$

$$g(x) = -\frac{1+4x^2}{4+x} = 1+x$$

が得られ、 $(g(1), g(2), g(3), g(4)) = (2, 3, 4, 0)$ が送信された符号語の推定となる。