■1 群（信号・システム）- 2 編（符号理論）

3 章 符号の性能

（執筆者：和田山正）［2012 年 3 月 受領］

■概要■

線形符号の訂正能力と符号化率の間のトレードオフ関係を明確にすることは符号理論における重要な課題の一つである．線形符号の符号長を n，符号語数を M，最小距離を $d_{\min}$ とするとき，これらのパラメータ間に，種々の制約関係が存在することが知られている．最も基本的なハミング限界式は，与えられた $n, d_{\min}$ に対して，符号語数 M を上から押さえる限界式となっており，2 元空間における球充填の考え方に基づいて導出される．更に同様の限界式としてプロトキン限界式，シングルトン限界式が知られている．これらの限界式は，あるパラメータにおける符号の非存在性を与えているものと考えることができる．一方，$n, M, d_{\min}$ がある関係式を満たすならばそのパラメータをもつ符号の存在を保証できる場合がある．この保証を与えるのがバルシャモフ-ギルバート限界である．

線形符号 C の符号語のうち，重み w の符号の個数を A_w とする．このとき，$A_0, A_1, \ldots, A_n$ を符号 C の重み分布と呼ぶ．重み分布は符号の性質を決める重要な量の一つであり，符号の復号誤り率の評価や誤り検出における見逃し誤り率の評価などに利用される．線形符号 C の重み分布とその双対符号 $C^{\perp}$ の重み分布の間にはマクウィリアムスの恒等式と呼ばれる線形の制約関係式が成立する．このマクウィリアムスの恒等式は，2 元原始 BCH 符号の一部，低次のリード-マラー符号の重み分布公式の導出など重み分布に関する理論的研究の基盤となっている．

【本章の構成】

本章では，3-1 節において，まず重要な符号の限界式（ハミング・プロトキン・シングルトン・バルシャモフ-ギルバート限界式）について述べたのち，それらの漸近的性質について議論する．3-2 節においては，重み分布に関する事項を紹介する．マクウィリアムスの恒等式を述べたのちに，ハミング符号やいくつかの原始 BCH 符号の重み分布を示す．3-3 節では，重み分布に基づく復号誤り率の評価について解説する．

■1 群 - 2 編 - 3 章

3-1 符号の限界式

（執筆者：森井昌克）[2012 年 3 月 受領]

符号の設計において，符号長 n，符号語数 M，最小距離 $d_{\min}$ のうち，n と M が与えられたとき d をできる限り大きくする，あるいは n と $d_{\min}$ が与えられたとき M をできる限り大きくすることが望まれる．これらのパラメータは任意の値を取れるわけではなく，パラメータ間には取り得る値の限界が存在する．このような限界式を用いることによって設計した符号を評価でき，更にはより性能の良い符号の設計方法について示唆を与えることができる．以下，符号語数 M に対応する情報シンボル数を $k = \log_q M$ とし，誤り訂正能力 $t = \lfloor (d_{\min} - 1)/2 \rfloor$ を有する q 元 $(n, k, d_{\min})$ 線形符号 C について述べる．

符号のパラメータに関する最も基本的な限界式としてハミング限界がある．符号 C において符号語数 M は

$$M \leq \frac{q^n}{\sum_{i=0}^{t} \binom{n}{i} (q-1)^i} \tag{3·1}$$

を満たす．このハミング限界は符号長 n と誤り訂正能力 t が与えられたとき符号語数 M の上界を与える．この限界式において等号が成立する符号は完全符号と呼ばれ，2 元完全符号としてはハミング符号，ゴーレイ符号がよく知られている．

ハミング限界と同様によく知られた限界式としてプロトキン限界がある．符号 C において最小距離 $d_{\min}$ は

$$d_{\min} \leq \frac{nM(q-1)}{(M-1)q} \tag{3·2}$$

を満たす．これは n，M に対して最小距離 $d_{\min}$ の上界を与えており，最小距離が任意の異なる符号語間の距離の平均値より大きくならないことから導ける．異なる符号語間の距離がすべて等しい符号は等距離符号と呼ばれ，等距離符号はこの限界式において等号を満たす．プロトキン限界の等号を満たす符号として重要なものにシンプレックス符号があり，これは等距離符号である．

また，q 元 $(n, k, d_{\min})$ 符号 C においては

$$d_{\min} \leq n - k + 1 \tag{3·3}$$

が成り立つ．この限界式をシングルトン限界と呼び，この限界式の等号を満たす符号を最大距離分離符号（maximum distance separable code），あるいは略して MDS 符号と呼ぶ（1.4 節参照）．

一方，n，$d_{\min}$ に対して符号語数 M の最大値の下界を与える限界式として次のものがある．

$$q^{n-k} \leq \sum_{i=0}^{d_{\min}-2} \binom{n}{i} (q-1)^i \tag{3·4}$$

を満たすならば最小距離 $d_{\min}$ の q 元 (n,k) 符号が存在する．これをバルシャモフ-ギルバート限界（Varshamov-Gilbert bound），また略して VG 限界と呼ぶ．ハミング限界，プロトキン限界，シングルトン限界は符号が存在するための必要条件を与えているのに対し，VG 限界は十分条件として符号の存在性を与えている．VG 限界は符号長が短かい場合は厳密な限界式にはならない．実際に符号長 1000 以下の BCH 符号は多くの場合この限界を超えている．一方，符号長が長い範囲ではこの限界式に達する符号を構成することは容易ではないが，代数幾何符号においていくつかの構成法が与えられている．

次に，符号長を無限に長くしたときの符号化率 $R = k/n$ と最小距離比 $\delta = d_{\min}/n$ に関する限界式の漸近的性質について，特に 2 元線形符号の漸近的限界式について述べる．

ハミング限界において符号長 n を十分大きくしたとき

$$H\left(\frac{\delta}{2}\right) \le 1 - R \tag{3·5}$$

という関係が得られる．ただし，$H(x)$ はエントロピー関数 $H(x) = -x\log_2 x - (1-x)\log_2(1-x)$ である．2 元 $(n,k,d_{\min})$ 符号の符号化率 R と最小距離比 δ はこの漸近的なハミング限界を満たす．また，プロトキン限界において符号長 n を十分大きくしたとき

$$\delta \le \frac{1}{2}(1 - R) \tag{3·6}$$

が得られ，2 元 $(n,k,d_{\min})$ 符号はこの漸近的なプロトキン限界を満たす．一方，VG 限界において符号長 n を十分大きくしたときには

$$H(\delta) \ge 1 - R \tag{3·7}$$

が得られ，この漸近的な VG 限界を満たす 2 元 $(n,k,d_{\min})$ 符号が存在する．これらの漸近的限界式を図 3·1 に示す．

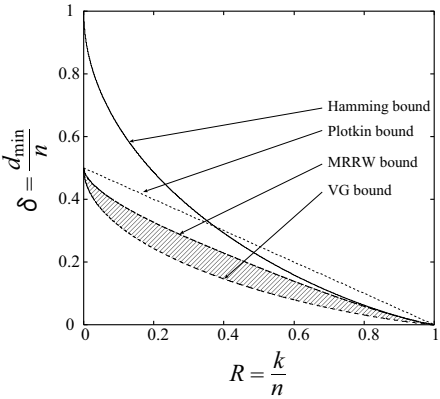

図 3·1　2 元 (n,k) 線形符号の最小距離 $d_{\min}$ の漸近的限界式

この図から分かるように，δ の上界と下界には差が生じている．このため，δ の上界や下界を改善する研究が盛んに行われている．下界に関しては現在のところ VG 限界より良い限界式は得られていないが，上界に関してはより厳密な限界式としてマクエリース-ロディミッチ-ルンセィ-ウェルチ限界（McEliece-Rodemich-Rumsey-Welch bound，略して MRRW 限界）がある．

■1 群 - 2 編 - 3 章

3-2 重み分布

（執筆者：森井昌克）[2012 年 3 月 受領]

q 元 (n,k) 線形符号 C の符号語のうちハミング重みが w である符号語の数を A_w で表す．このとき $A_0, A_1, \ldots, A_n$ を符号 C の重み分布という．また，

$$A(X) = \sum_{w=0}^{n} A_w X^w \tag{3·8}$$

を符号 C の重み母関数という．$A_w\,(0 < w \le n)$ が非零となる最小の w は符号 C の最小重み（最小距離）である．重み分布は符号の復号特性を解析する際に重要な役割を果たす．例えば，重み分布から 2 元対称通信路上での復号誤り確率や見逃し誤り確率などを正確に計算できる．

符号 C とその双対符号 $C^\perp$ の重み分布には次の重要な関係が成り立つ．符号 C の重み母関数 $A(X)$ が得られたとき，その双対符号である q 元 $(n, n-k)$ 線形符号 $C^\perp$ の重み母関数は

$$A^\perp(X) = q^{-k}(1+(q-1)X)^n A\left(\frac{1-X}{1+(q-1)X}\right) \tag{3·9}$$

で与えられる．これをマクウィリアムスの恒等式と呼ぶ．この恒等式を用いることで，C あるいは $C^\perp$ のいずれか一方の重み分布が分かれば他方の重み分布は計算によって求めることができる．符号 C の符号語をすべて生成し重みを調査するような全数探索的手法によって重み分布を求める計算量は $O(nq^k)$ である．一方，符号化率 $k/n > 1/2$ の符号 C に対しては，その双対符号 $C^\perp$ の重み分布母関数 $A^\perp(X)$ を求め，その $A^\perp(X)$ からマクウィリアムスの恒等式を用いて $A(X)$ を計算する方が効率が良くなり，その計算量は $O(nq^{n-k})$ である．よって，情報点数 k また検査シンボル数 $n-k$ が小さい符号では全数探索的手法によって重み分布を求めることができるが，k 及び $n-k$ のいずれも大きい場合には計算量が膨大となり，重み分布を求めることは困難になる．

線形符号のうち，一部の符号に対して重み分布の公式が明らかにされている．よく知られているものに次の符号がある．

- 2 元ハミング符号とその双対符号及び拡大符号
  $(2^m-1, 2^m-m-1)$ ハミング符号の双対符号である $(2^m-1, m)$ 符号の重み母関数は

$$A^\perp(X) = 1 + (2^m-1)X^{2^{m-1}} \tag{3·10}$$

  である．これよりマクウィリアムスの恒等式を用いることによって，符号長 $n = 2^m-1$，情報ビット数 $k = 2^m-m-1$ のハミング符号の重み母関数

$$\begin{aligned} A(X) &= 2^{-m}(1+X)^{2^m-1} A^\perp\left(\frac{1-X}{1+X}\right) \\ &= \frac{1}{n+1}\left[(1+X)^n + n(1-X)(1-X^2)^{\frac{n-1}{2}}\right] \end{aligned} \tag{3·11}$$

が得られる．また，この符号を拡大した (n,k) 拡大ハミング符号の重み母関数は

$$A(X) = \frac{1}{2n}[(1+X)^n + (1-X)^n + 2(n-1)(1-X^2)^{\frac{n}{2}}] \qquad (3\cdot12)$$

となる．

- 誤り訂正能力の低い 2 元原始 BCH 符号の双対符号
  2 重誤り及び 3 重誤り訂正原始 BCH 符号の双対符号の重み分布は表 3･1，3･2 のように公式で与えられている．よって，2 重誤り及び 3 重誤り訂正原始 BCH 符号の重み分布はこれらの公式からマクウィリアムスの恒等式を用いて計算できる．

表 3･1　符号長 2^m-1 の 2 重誤り訂正原始 BCH 符号の双対符号の重み分布

m が 3 以上の奇数の場合

重み w	符号語数 $A_w^{\perp}$
0	1
$2^{m-1}-2^{\frac{m+1}{2}}-1$	$(2^{m-2}+2^{\frac{m-1}{2}}-1)(2^m-1)$
2^{m-1}	$(2^m-2^{m-1}+1)(2^m-1)$
$2^{m-1}+2^{\frac{m+1}{2}}-1$	$(2^{m-2}-2^{\frac{m-1}{2}}-1)(2^m-1)$

m が 4 以上の偶数の場合

重み w	符号語数 $A_w^{\perp}$
0	1
$2^{m-1}-2^{\frac{m+2}{2}}-1$	$\frac{1}{3}\cdot2^{\frac{m-2}{2}}-1(2^{\frac{m-2}{2}}+1)(2^m-1)$
$2^{m-1}-2^{\frac{m}{2}}-1$	$\frac{1}{3}\cdot2^{\frac{m+2}{2}}-1(2^{\frac{m}{2}}+1)(2^m-1)$
2^{m-1}	$(2^{m-2}+1)(2^m-1)$
$2^{m-1}+2^{\frac{m}{2}}-1$	$\frac{1}{3}\cdot2^{\frac{m+2}{2}}-1(2^{\frac{m}{2}}-1)(2^m-1)$
$2^{m-1}+2^{\frac{m+2}{2}}-1$	$\frac{1}{3}\cdot2^{\frac{m-2}{2}}-1(2^{\frac{m-2}{2}}-1)(2^m-1)$

表 3･2　符号長 2^m-1 の 3 重誤り訂正原始 BCH 符号の双対符号の重み分布

m が 5 以上の奇数の場合

重み w	符号語数 $A_w^{\perp}$
0	1
$2^{m-1}-2^{\frac{m+1}{2}}$	$\frac{1}{3}\cdot2^{\frac{m-5}{2}}(2^{\frac{m-3}{2}}+1)(2^{m-1}-1)(2^m-1)$
$2^{m-1}-2^{\frac{m-1}{2}}$	$\frac{1}{3}\cdot2^{\frac{m-3}{2}}(2^{\frac{m-1}{2}}+1)(5\cdot2^{m-1}+4)(2^m-1)$
2^{m-1}	$(9\cdot2^{2m-4}+3\cdot2^{m-3}+1)(2^m-1)$
$2^{m-1}+2^{\frac{m-1}{2}}$	$\frac{1}{3}\cdot2^{\frac{m-3}{2}}(2^{\frac{m-1}{2}}-1)(5\cdot2^{m-1}+4)(2^m-1)$
$2^{m-1}+2^{\frac{m+1}{2}}$	$\frac{1}{3}\cdot2^{\frac{m-5}{2}}(2^{\frac{m-3}{2}}-1)(2^{m-1}-1)(2^m-1)$

m が 6 以上の偶数の場合

重み w	符号語数 $A_w^{\perp}$
0	1
$2^{m-1}-2^{\frac{m+4}{2}-1}$	$\frac{1}{960}(2^{m-1}+2^{\frac{m+4}{2}-1})(2^m-4)(2^m-1)$
$2^{m-1}-2^{\frac{m+2}{2}-1}$	$\frac{7}{48}(2^{m-1}+2^{\frac{m+2}{2}-1})2^m(2^m-1)$
$2^{m-1}-2^{\frac{m}{2}-1}$	$\frac{2}{15}(2^{m-1}+2^{\frac{m}{2}-1})(3\cdot2^m+8)(2^m-1)$
2^{m-1}	$\frac{1}{64}(29\cdot2^{2m}-4\cdot2^m+64)(2^m-1)$
$2^{m-1}+2^{\frac{m}{2}-1}$	$\frac{2}{15}(2^{m-1}-2^{\frac{m}{2}-1})(3\cdot2^m+8)(2^m-1)$
$2^{m-1}+2^{\frac{m+2}{2}-1}$	$\frac{7}{48}(2^{m-1}-2^{\frac{m+2}{2}-1})2^m(2^m-1)$
$2^{m-1}+2^{\frac{m+4}{2}-1}$	$\frac{1}{960}(2^{m-1}-2^{\frac{m+4}{2}-1})(2^m-4)(2^m-1)$

- 低次の 2 元リード-マラー（RM）符号
  符号長 2^m の 1 次 RM 符号と 2 次 RM 符号の重み分布は公式で与えられている．また，r 次 RM 符号の双対符号が $m-r-1$ 次 RM 符号であることから，1 次及び 2 次 RM 符号の重み分布からマクウィリアムスの恒等式を用いて $m-2$ 次及び $m-3$ 次 RM 符号の重み分布が計算できる．更に，r 次 RM 符号の最小重みとその符号語数も明らかにされている．

以上のように，一部の限られた符号に対して重み分布の公式が与えられているが，一般の符号に対して重み分布を与えることは容易ではない．また，符号長 n 及び検査シンボル数 $n-k$ が大きい符号の重み分布を求める計算量は大きく，導出は困難になる．よって，重み分布を効率的に計算する方法が希求されている．

■1 群 - 2 編 - 3 章

3-3 誤り率の評価

（執筆者：森井昌克）[2012 年 3 月 受領]

符号の性能は各種通信路に適用したときの誤り率によって評価することができる．通信路をビット誤り率 $p\,(0 \le p \le 1/2)$ の 2 元対称通信路とし，2 元 (n,k) 線形符号 C の見逃し誤り確率，復号誤り確率などの誤り率について考える．ここで，符号 C の重み分布は $A_0, A_1, \ldots, A_n$ とし，C の最小距離を $d_{\min}$ とする．

符号 C の符号語 $\boldsymbol{c}$ をビット誤り率 p の 2 元対称通信路に送信し，受信語 $\boldsymbol{r}\,(=\boldsymbol{c}+\boldsymbol{e})$ を得たとする．このとき，誤りベクトル $\boldsymbol{e}$ が符号 C の非零符号語と一致すると受信語 $\boldsymbol{r}$ は送信語 $\boldsymbol{c}$ とは異なる符号語になり，誤りを検出できない．よって，符号 C の見逃し誤り確率は

$$P_u = \sum_{i=1}^{n} A_i p^i (1-p)^{n-i} \tag{3·13}$$

となる．また，誤り検出確率は P_u を用いて

$$\begin{aligned} P_d &= \sum_{i=1}^{n} \binom{n}{i} p^i (1-p)^{n-i} - P_u \\ &= 1 - (1-p)^n - P_u \end{aligned} \tag{3·14}$$

と与えられる．設計距離 d に対して $t = \lfloor (d-1)/2 \rfloor$ なる t 個までの誤りを訂正する限界距離復号を行った場合について考える．設計距離 d が最小距離 $d_{\min}$ と等しくなり $t = \lfloor (d_{\min}-1)/2 \rfloor$ であるとき，この復号は最小距離復号と呼ばれる．限界距離復号を行う場合，受信語 $\boldsymbol{r}$ に t 個以上の誤りが生起すると受信語を送信語に訂正できず，復号に失敗する．その復号失敗確率は

$$P_F = \sum_{i=t+1}^{n} \binom{n}{i} p^i (1-p)^{n-i} \tag{3·15}$$

である．また，符号 C に限界距離復号を適用し送信語とは異なる符号語に誤訂正される確率，すなわち，復号誤り確率は

$$P_E = \sum_{w=1}^{n} \sum_{i=w-t}^{w+t} \sum_{j=0}^{t} A_w T(i,j,w) p^i (1-p)^{n-i} \tag{3·16}$$

となる．ただし，$T(i,j,w)$ は復号語の重み w のとき送信語から距離 i であり復号語から距離 j であるベクトルの個数を表し，$i+j-w$ が偶数かつ $w-j \le i \le w+j$ のとき $T(i,j,w) = \binom{w}{i-(i+j-w)/2} \binom{n-w}{(i+j-w)/2}$，そのほかのとき $T(i,j,w) = 0$ である．更に，2 元対称通信路において符号 C に最ゆう復号を適用した場合の復号誤り確率は

$$\begin{aligned} P_E \le & \sum_{w=1}^{n} A_w \sum_{i=\lfloor \frac{w}{2} \rfloor + 1}^{I} \sum_{j=0}^{i-1} T(i,j,w) p^i (1-p)^{n-i} \\ & + \sum_{i=I+1}^{n} \binom{n}{i} p^i (1-p)^{n-i} + \frac{1}{2} \sum_{j=1}^{\lfloor \frac{n}{2} \rfloor} \sum_{i=j}^{i=I} A_{2j} T(i,i,2j) p^i (1-p)^{n-i} \end{aligned} \tag{3·17}$$

となる．ただし，I は右辺を最小とする値である．この復号誤り確率の上界はポルティレヴ限界と呼ばれている．最ゆう復号の場合，誤り検出ではとどめず，受信語は距離が近い符号語に訂正するため，復号失敗確率は復号誤り確率に一致する．

一方，通信路が加法的白色ガウス雑音（AWGN）通信路である場合，符号化率 $R = k/n$ の 2 元線形符号 C に最ゆう復号を用いたときの復号誤り確率は

$$P_E \leq \sum_{w=d_{\min}}^{n} A_w Q\left(2wR\frac{E_b}{N_0}\right) \tag{3･18}$$

と与えられる．ただし，$Q(x)$ はガウスの誤差関数 $Q(x) = \int_x^\infty \frac{1}{2\pi} e^{-\frac{z^2}{2}} dz$ である．これをユニオン限界と呼ぶ．

次に，ビット誤り率 p の 2 元対称通信路における誤り率の漸近的限界式について述べる．2 元対称通信路において限界距離復号を用いる場合，符号 C の復号失敗確率は式 (3･15) のようになる．ここで，$\tau = t/n$ とおき，符号長 n を大きくすると

$$P_F \approx \binom{n}{t+1} p^{t+1}(1-p)^{n-t-1} \approx 2^{-nH(\tau,p)} \tag{3･19}$$

と近似できる．ただし，$H(\tau, p) = -\tau \log_2 p - (1-\tau)\log_2(1-p) - H(\tau)\,(0 \leq p < \tau \leq \frac{1}{2})$ である．この式と VG 限界及び MRRW 限界を組み合わせることで限界距離復号における P_F の漸近的な上界，下界を求めることができる．一方，最ゆう復号の場合は復号失敗確率 P_F と復号誤り確率 P_E が一致し，復号誤り確率 P_E が最も小さくなる符号では

$$P_E \approx 2^{-nE(R)} \tag{3･20}$$

と考えられている．この $E(R)$ は最ゆう復号における誤り指数と呼ばれている．誤り指数 $E(R)$ の下界として

$$E_r(R) = \max_{0<\rho\leq 1}[E_0(\rho) - \rho R] \tag{3･21}$$

がある．ただし，$E_0(\rho)$ は 2 元対称通信路におけるギャラガーの信頼性関数 $E_0(\rho) = \rho - (1+\rho)\log_2[p^{\frac{1}{1+\rho}} + (1-p)^{\frac{1}{1+\rho}}]$ である．$P_E \leq 2^{-nE_r(R)}$ はギャラガーの上界と呼ばれ，符号化率が高い範囲で厳密な上界を与える．また，下界や低符号化率においてより厳密になる上界についても種々の方法で求められている．