6.3 最小距離

最近傍復号法を用いるとき、送信された符号語 $\mathbf{u}$ は受信語 $\mathbf{v} \in \mathcal{V}$ から最も近い符号語 $\Delta(\mathbf{v})$ になるため、他の符号語から互いに離れた符号語 $\mathbf{u}$ を用いることで誤り確率 $\Pr_E$ を低く保つことができる。

よって、C の最小距離を任意の異なる 2 つの符号語のハミング距離の最小値

$d = d(C) = \min\{d(\mathbf{u}, \mathbf{u}') \mid \mathbf{u}, \mathbf{u}' \in C,\ \mathbf{u} \neq \mathbf{u}'\}$ (6.3)

として定める。

ある符号が符号長 n、符号語数 M で最小距離 d のとき、これを (n, M, d) 符号と呼ぶことがある。もし符号が線形で次元が k ならば、これは $[n, k, d]$ 符号と呼ばれる。

我々の目的は $\Pr_E$ を小さくするために、符号語 C を d が大きくなるように選ぶことである。

もし、C が M 個の符号語を持つ場合、(6.3) を用いて d を求めることは、$\binom{M}{2} = M(M-1)/2$ 通りの距離の計算と比較を要求する。これはとてもうんざりする。

しかしながら、この作業は C が線形の場合次に示すようにとても単純になる。

まず、任意のベクトル $\mathbf{v} = v_1 v_2 \dots v_n \in \mathcal{V}$ に対する重みを、$\mathbf{0} = 00\dots0$ として以下のように定義する。

$\mathrm{wt}(\mathbf{v}) = d(\mathbf{v}, \mathbf{0})$ (6.4)

言い換えると、$\mathrm{wt}(\mathbf{v})$ は単純に $v_i \neq 0$ となるような添字 i の数である。任意の $\mathbf{u}, \mathbf{u}' \in \mathcal{V}$ に対して、

$d(\mathbf{u}, \mathbf{u}') = \mathrm{wt}(\mathbf{u} - \mathbf{u}')$

となることは容易にわかる。

系 6.8

C が線形符号のとき、その最小距離 d は

$d = \min\{\mathrm{wt}(\mathbf{v}) \mid \mathbf{v} \in C,\ \mathbf{v} \neq \mathbf{0}\}$

で与えられる。

証明　$\mathbf{v} = \mathbf{u} - \mathbf{u}'$ の下、$d(\mathbf{u}, \mathbf{u}') = \mathrm{wt}(\mathbf{v})$ とする。今、C は $\mathcal{V}$ の線形部分空間なので、$\mathbf{u}$ と $\mathbf{u}'$ が全ての C 内の異なる組み合わせに及ぶと、それらの差 $\mathbf{v} = \mathbf{u} - \mathbf{u}'$ は C の全ての非零な要素に及ぶ。

これは $d(C)$ なるこのような $\mathbf{u}, \mathbf{u}'$ の組の間の最小距離は、これらの重み $\mathrm{wt}(\mathbf{v})$ に等しくなる。

この結果の利点は、非線型な符号で $M(M-1)/2$ 回の計算と比較をする場合に比べて、$M-1$ 回の計算と比較で済むことである。§7.3 では、より良い線形符号の最小距離計算の方法があることを確認する。

練習 6.3

二元ハミング符号 $\mathcal{H}_7$ (例 6.5) の符号語を全て列挙せよ。そして系 6.8 を用いて最小距離が 3 であることを確認せよ。

練習 6.4

C が最小距離 d の二元線形符号のとき、拡大符号 $\bar{C}$ は d が偶数または奇数に応じて、それぞれ d または $d+1$ なる最小距離を持つことを示せ。拡大二元ハミング符号 $\bar{\mathcal{H}_7}$ の符号語を列挙し、その最小距離を求めよ。

今どのように符号の最小距離がその符号の誤り訂正能力に影響するかを考える。符号 C は t 個の誤りを訂正する、または t 重誤り訂正とは、符号語 $\mathbf{u} \in C$ が伝送され、多くとも t 個のシンボルが誤って受信された場合はいつでも結果として受信語 $\mathbf{v}$ は正しく $\mathbf{u}$ として復号されることを言う。これは $\mathbf{u} \in C$ かつ $\mathbf{v} \in \mathcal{V}$ で、$d(\mathbf{u}, \mathbf{v}) \leq t$ を満たす場合いつでも、決定則 Δ が $\Delta(\mathbf{v}) = \mathbf{u}$ を与えることと同値である。

例 6.9

反復符号 $\mathcal{R}_3$ は (どんなアルファベットの上でも) 一つの誤りを訂正するが、2 つの誤りは訂正しない (§5.2 の $q = 2$ の場合を確認せよ)。例えば、$\mathbf{u} = 111$ が転送され $\mathbf{v} = 101$ が受信された (つまり 1 つの誤りがある) 場合、最近傍復号は $\Delta(\mathbf{v}) = 111 = \mathbf{u}$ を与える。しかし、$\mathbf{v} = 001$ が受信される (つまり 2 つの誤りがある) 場合、$\Delta(\mathbf{v}) = 000 \neq \mathbf{u}$ が得られる。

$\mathbf{u}$ が送信され、$\mathbf{v}$ が受信される場合、$\mathbf{e} = \mathbf{v} - \mathbf{u}$ なるベクトルを誤りパターンと呼ぶ、なぜならこれの非零の要素は伝送中にどの位置にどういった誤りが起きたかを示すからである。等式 $\mathbf{v} = \mathbf{u} + \mathbf{e}$ は、$\mathbf{v}$ が送信された語 $\mathbf{u}$ と $\mathbf{e}$ で示される誤りの和からなることを示している。誤ったシンボルの数は $d(\mathbf{u}, \mathbf{v}) = \mathrm{wt}(\mathbf{e})$ で、符号が t 個の誤りを訂正するというのは、符号が $\mathrm{wt}(\mathbf{e}) \leq t$ なる全ての誤りパターン $\mathbf{e}$ を訂正することの必要十分条件である。

定理 6.10

最小距離 d の符号 C が t 個の誤りを訂正するのは、$d \geq 2t + 1$ であることの必要十分条件である。(同様に C が $\lfloor \frac{d-1}{2} \rfloor$ 個以下の誤りを訂正する。)

証明 ($\Leftarrow$)C を最小距離 $d \geq 2t + 1$ の符号と置く。$\mathbf{u} \in C$ が送信され、誤りパターン $\mathbf{e} = \mathbf{v} - \mathbf{u}$ が重み $\mathrm{wt}(\mathbf{e}) \leq t$ を持つ、つまり $d(\mathbf{u}, \mathbf{v}) \leq t$ として、$\mathbf{v} = \mathbf{u} + \mathbf{e}$ が受信されたとする。全ての C の元 $\mathbf{u}' \neq \mathbf{u}$ に対し、

$$d(\mathbf{u}, \mathbf{u}') \geq d \geq 2t + 1$$

である。今、三角不等式 (系 5.8 (c)) は

$$d(\mathbf{u}, \mathbf{u}') \leq d(\mathbf{u}, \mathbf{v}) + d(\mathbf{v}, \mathbf{u}')$$

を与える。つまり

$$d(\mathbf{v}, \mathbf{u}') \geq d(\mathbf{u}, \mathbf{u}') - d(\mathbf{u}, \mathbf{v}) \geq (2t + 1) - t = t + 1 > d(\mathbf{u}, \mathbf{v}).$$

よって、$\Delta(\mathbf{v}) = \mathbf{u}$ なので、復号は正しく、C は t 個の誤りを訂正する。

($\Rightarrow$)C が最小距離 $d < 2t + 1$ を持つ、つまり $d \leq 2t$ であるとする。$d(\mathbf{u}, \mathbf{u}') = d$ となるように $\mathbf{u}, \mathbf{u}' \in C$ を選ぶことができる。この時、以下のようなベクトル $\mathbf{v} \in \mathcal{V}$ が存在する。

$$d(\mathbf{u}, \mathbf{v}) \leq t \text{ かつ } d(\mathbf{u}', \mathbf{v}) \leq t$$

(例えば、$\mathbf{u}$ と $\mathbf{u}'$ はちょうど d 個のシンボルが異なっていて、それらの $\lfloor d/2 \rfloor$ 個の $\mathbf{u}$ のシンボル u_i を対応する $\mathbf{u}'$ のシンボル u_i' に変えることによってそのような $\mathbf{v}$ なるベクトルを得ることができる。) 今 $\Delta(\mathbf{v})$ は $\mathbf{u}$ と $\mathbf{u}'$ の両方になることはできないので、少なくともそれら 2 つの符号語のうち 1 つが、伝送され $\mathbf{v}$ として受信されたとき、誤って復号される。よって C は t 個の誤りを訂正できない。

例 6.11

符号長 n の反復符号 $\mathcal{R}_n$ は最小距離 $d = n$ をを持つ。これは全ての $\mathcal{R}_n$ 内の $\mathbf{u} \neq \mathbf{u}'$ に対して、$d(\mathbf{u}, \mathbf{u}') = n$ からわかる。よってこの符号は $t = \lfloor \frac{n-1}{2} \rfloor$ 個の誤りを訂正する。

例 6.12

練習 6.3 ではハミング符号 $\mathcal{H}_7$ が最小距離 $d = 3$ を持つことを示した、よってこれは $t = 1$ (§6.2 で示したように) である。似たように、$\bar{\mathcal{H}_7}$ は $d = 4$ (練習 6.4 より) で、よってこの符号もまた $t = 1$ である。

例 6.13

符号長 n のパリティ検査符号 $\mathcal{P}_n$ は最小距離 $d = 2$ を持つ。例えば、符号語 $\mathbf{u} = 1(-1)0 \ldots 0$ と $\mathbf{u}' = \mathbf{0} = 00 \ldots 0$ は 2 だけ離れているが、1 だけ離れている組は無い。その結果として、$\mathcal{P}$ にとって訂正される誤りの数は $t = \lfloor \frac{d-1}{2} \rfloor = 0$ 個である。例えば、$\mathbf{v} = 10 \ldots 0$ は $\mathbf{u}$ と $\mathbf{u}'$ の両方 (さらに他の符号もある) に復号され、それらの符号は 1 つの誤りで $\mathbf{v}$ になることが出来る。

$\mathcal{P}_n$ は誤りを訂正する目的では使えないが、少なくとも 1 つの誤りを検出する。より一般的にいうと、符号 C の最小距離が d で、符号語 $\mathbf{u} \in C$ が送信され、$\mathbf{v} = \mathbf{u} + \mathbf{e}$ が受信され、$1 \leq \mathrm{wt}(\mathbf{e}) \leq d - 1$ とする。このとき、$0 < d(\mathbf{u}, \mathbf{v}) < d$ であるため、$\mathbf{v}$ は符号語ではない。よって受信者は少なくとも 1 つの誤りが $\mathbf{v}$ に存在することを知る。もし、$\mathrm{wt}(\mathbf{e}) = d$ である場合、$\mathbf{v}$ は符号語の可能性があり、このような場合は受信者は $\mathbf{v}$ が正しく伝送されたのか、あるいは $\mathbf{u}$ (あるいはその他の符号語) が誤って転送されたのかがわからない。このような場合 C は $d - 1$ 個の誤りを検出するという。

例 6.14

反復符号 $\mathcal{R}_n$ とパリティ検査符号 $\mathcal{P}_n$ はそれぞれ、$d = n$ 及び 2 なので、$\mathcal{R}_n$ は $n - 1$ 個の誤りを検出し、一方 $\mathcal{P}_n$ は 1 つの誤りを検出する。$\mathcal{H}_7$ は $d = 3$ なので、2 つの誤りを検出する。