概要

我々の目的は、C という符号を、シャノンの基本定理に基づいて転送効率 (R) 良くかつ誤り確率 (Pr_E) を低く構成することである。

本章では前項の情報理論と対応して、符号理論 (あるいは誤り訂正符号) について触れていく。

このような符号の生成は極めて異なる作業で、より発展的な符号を生成するためのいくつかの手法についてを表現する、いくつかの単純な例に集中することになるだろう。

導入

今後、入力 A と出力 B の符号アルファベットが同一の、通信路 Γ、BSC(2 元対称通信路) という符号について想定していく。

これを考える際一般性を失うことはない、もし同一でなければ A と B を共通の符号アルファベット $A \cup B$ で置き換えることができるからである。

この共通の有限の符号アルファベットを F で示す、これは代数のテクニックを使うためにこれをよく体として取るためである。

体となるためには、F は $ab = ba, a(b+c) = ab+ac, etc\ldots$ などの公理により、加算・減算・積算・除算 (0 除算を除く) の演算が閉じていなければならない。

標準的な体を含む例は、有理数 $\mathbf{Q}$・実数 $\mathbf{R}$・複素数 $\mathbf{C}$ などがある。

これらは無限な体であるが、我々の目的のためには有限体を使うことが必要である、例えば素数 p に対し、 p を法とする整数の体 $\mathbf{Z}_p$ のようなものを使うことである。

必要となる有限体についての事項は以下である

定理 6.1

- (a) ある素数 p に対し、$^{\exists}e \geq 1, q = p^e$ となる場合のみ、位数が q となる有限体が存在する。
- (b) 位数が同じな任意の 2 つの有限体は同型である。

多くの代数学の教科書では (KR83 のような) がこの結果を証明しているので、証明無しで使っていく。

本質的な唯一の位数 q の体はガロア体 F_q ないし $GF(q)$ として知られている。もし $e = 1$ で $q = p$ が素数であれば $F_q = F_p = \mathbf{Z}_p$ で、 $\bmod p$ からなる整数の体である。

しかしながら、もし $e > 1$ ならば、 q は合成数で $\mathbf{Z}_q$ は体ではない。

例えば、 たとえ $p \neq 0$ だとしても、$\mathbf{Z}_q$ の中で $pe = 0$ であり、p は零因子となってしまう。 これは $e > 1$ に対して $F_q \neq \mathbf{Z}_q$ を意味する。代わりに F_q は、$\mathbf{Z}_p$ に既約多項式 $f(x)$ の解 α を添加したものとして得られる、ちょうど複素数体 $\mathbf{C}$ が $\mathbf{R}$ に $f(x) = x^2 + 1$ の解 $i = \sqrt{-1}$ を添加したものから得られるようなものである。

F_q の要素は、このとき $a_0, a_1, \cdots, a_{e-1} \in \mathbf{Z}_p$ として、$a_0 + a_1\alpha + \cdots + a_{e-1}\alpha^{e-1}$ の形であり、明白に加算と減算の演算をもつ。

このような要素 2 つの積は、α の次数を下げるため、$f(\alpha) = 0$ の方程式を用いてこの形式に入れることができる。$f(x)$ は F_q の零因子を避けるために既約多項式である必要がある。

例 6.2　二次方程式 $f(x) = x^2 + x + 1$ は $\mathbf{Z}_2$ 上の解を持たず $(f(0) = f(1) = 1$ なので)、よって一次の解を持たず、そして故に $\mathbf{Z}_2$ 上で既約である。

$\mathbf{Z}_2$ に解 α を添加することで、位数 $q = 4$ の体

$$F_4 = \{a + b\alpha \mid a, b \in \mathbf{Z}_2\} = \{0, 1, \alpha, 1 + \alpha\}$$

を得る。$\alpha^2 + \alpha + 1 = 0$ となるため、$\alpha^2 = -1 - \alpha = 1 + \alpha$ である。

例えば、 $\alpha(1 + \alpha) = \alpha + \alpha^2 = 1 + 2\alpha = 1$ なので、α と $1 + \alpha$ は F_4 上で互いに積の逆元となる。

似たような有限体の構成については、演習問題 6.16 と 6.17 を見よ。

我々の目的では、正確に F_q を構成することは通常重要ではなく、それぞれの素数の累乗を位数とするガロア体が存在することを簡単に知ることが重要である。

しかしながら、本書で扱う内容を超えたより発展的で、より深い有限体の知識を必要とする符号が存在する。

F_q の計算は $q = p^e$ で $p = 0$ な F_q を除いて、他の体と同じである。更に、$\mathbf{R}, \mathbf{Q}$ に存在し、$\mathbf{C}$ には存在しないように、自然な順序関係 $<$ は存在しない。

多くの場合、我々は $F_q = \mathbf{Z}_2 = \{0, 1\}$ で $1 + 1 = 0$ となるような 2 元符号に専念していく。

これからは、シャノンの基本定理に従って、すべての符号語が同じ長さとなるようなブロック符号を使っていく。

これは、本書でこれまでに効率のために可変長の符号を扱っていたことに矛盾しない。

最初にそのような符号を使い、それを固定長 n の符号語とみなして、連続した同じ長さ k の符号に分けることができる。

結果として符号が良い誤り訂正能力を持つために、(ハミング距離の意味で) これらの符号語を可能な限り離れるように選んでいく。

もし符号語長 n とすると、長さ n の符号 C は、$\mathcal{V} = F^n$ なるすべての F の要素の $n-$タプルの集合の部分集合になる。

もし F が体であれば、$\mathcal{V}$ は n 次元の F 上の線形空間となり、加算及びスカラー倍の演算を持つ。

ここで、$\mathbf{u} = u_1 \ldots u_n,\ \mathbf{v} = v_1, \ldots v_n\ \in \mathcal{V}$ で、 更に $a, b \in F$ ならば、$a\mathbf{u} + b\mathbf{v}$ は符号語ないし、ベクトルであり、$i = 1 \ldots n$ に対して i 番目の要素が $au_i + bv_i$ となる。

もし C が $\mathcal{V}$ の線形部分空間であれば、このような C を線形符号 (ないし組織符号) と呼び、このことは C が空集合でなく、更に $\mathbf{u}, \mathbf{v} \in C$ ならば任意の $a, b \in F$ に対して、$a\mathbf{u} + b\mathbf{v} \in C$ となることを意味している。

特に、すべての線形符号は零ベクトル $\mathbf{0} = 000 \ldots 0$ を含んでいる。なぜなら任意の $\mathbf{u}, \mathbf{v} \in C$ に対して $\mathbf{0} = 0\mathbf{u} + 0\mathbf{v}$ だからである。

多くの符号は非線形で、$\mathcal{V}$ の線形な部分集合は比較的少ない。

しかしながら、現在研究され、利用されている多くの符号は線形である。なぜならそれらは理解しやすく使いやすいためである。

以下のようにして線形符号に対するシャノンの基本定理の類似系を証明することが出来る。

定理 5.9 の様にランダムな符号 $C \subseteq \mathcal{V}$ をとる代わりに、$C \subseteq \mathcal{V}$ の基底となるように $\mathcal{V}$ の部分空間をとる。

そして、$n \to \infty$ のとき C が要求される性質を持つことを示す。

今後 $\mid C \mid$ を M で示す。C が線形である場合、$k = \dim(C)$ を線形空間 C の次元とすると、$M = q^k$ となる。

すべての C の要素は、$a_1, \cdots, a_k \in F$ かつ C の基底 $\mathbf{u}_1, \cdots \mathbf{u}_k \in C$ の下、それぞれ一意な $a_1\mathbf{u}_1 + \cdots + a_k\mathbf{u}_k$ なる式を持ち、 更に、それぞれの a_i に対して独立した $\mid F \mid = q$ 通りの選択があるからである。

このような符号を線形 $[n, k]$ 符号と呼ぶ。

符号 C の伝送レートは

$\frac{\log_q M}{n}$ (6.1)

で、線形 $[n, k]$ 符号の場合は

$\frac{k}{n}$ (6.2)

となる。

このことは、符号語の n 桁のうち k 桁を情報桁及び伝達したい情報を運ぶ部分とみなし、そして残りの $n-k$ 桁を検査桁 (チェックディジット) とみなし、情報の確認あるいは保護をする部分として解釈できる。

この先では、C 中のすべての符号語が等しい確率で、($\mathcal{V}$ 上のハミング距離に関して) 最近傍復号を使うものと仮定する。