誤り訂正符号の理論への代数幾何学の応用

丸山 直昌　　データ科学研究系　准教授

1　誤り訂正符号とは

雑音の多い電話回線を通じて、例えば「たちかわ」という言葉を相手に伝えたい時、「たけのこ」の「た」、「ちくわ」の「ち」、「からし」の「か」、「わらび」の「わ」、というふうに伝える方法があります。聞き手が「たけのこ」を「らけのこ」に聞き違えたとしても、「らけのこ」という食物はありませんから、元の単語「たけのこ」を容易に想像でき、最初の文字「た」を認識できます。このやり方では、送信側と受信側が共通の辞書を持っていることが重要な鍵となります。誤り訂正符号とは、このようなやり方を数学的な手法を使って実現する仕組みです。

2　線形符号

辞書としては、通常有限体上のベクトル空間の部分空間が用いられます。qを素数巾、$\mathbb{F}_q$をq個の元から成る有限体、Cを$\mathbb{F}_q{}^n$のk次元部分空間とするとき、$(x_1, x_2, \ldots, x_k)$に$n - k$個の成分を付加して$(x_1, x_2, \ldots, x_n)$がCの元となるようにします。射影

$$p : (x_1, x_2, \ldots, x_n) \mapsto (x_1, x_2, \ldots, x_k)$$

によりCが$\mathbb{F}_q^k$に全射で写っていればそれは可能で、$(x_1, x_2, \ldots, x_n)$は一意的です。この場合、送信したい情報$(x_1, x_2, \ldots, x_k)$に対して$(x_1, x_2, \ldots, x_n)$を送信し、受信した情報$(y_1, y_2, \ldots, y_n)$に一番「近い」Cの元$(z_1, z_2, \ldots, z_n)$を探して$(z_1, z_2, \ldots, z_k)$を$(x_1, x_2, \ldots, x_k)$の推定値とします。$(y_1, y_2, \ldots, y_n)$から$(z_1, z_2, \ldots, z_n)$を見つける手続きを復号と言います。

3　代数幾何的符号、一点符号

更に、代数幾何学の枠組を使い、次のようなベクトル空間とその部分空間を用いて符号理論を考えることができます。

Xを有限体$\mathbb{F}_q$上の特異点が無い完備代数曲線、GをX上の因子(X上の点の形式的な和)とし、$\{p_1, p_2, \ldots, p_n\}$は$X$上の相異なる$n$個の点で、$G$を構成するどの点とも異なるものとします。$X$上の有理関数$f$で、その因子$(f)$と$G$を加えたものが正の因子となるもの全体と$0$がなす線型空間を$L$とします(この$f$の条件は、「$f$の極が$G$を構成する点だけにあり、その位数が対応する$G$の点の係数を越えない」、と言い換えることができます)。写像$\psi : L \to \mathbb{F}_q{}^n$を、$f \in L$に対して$\psi(f) = (f(p_1), f(p_2), \ldots, f(p_n))$と決めると、$\psi(L) \subset \mathbb{F}_q{}^n$が符号の役割を果たします。このことは次のGoppaによる命題によって理解できます。

Proposition(Goppa). $max(0, 2g-1) \leq degG \leq min(n-1, n+2g-2)$の時$\psi$は単射で、

$$dim\ \psi(L) = degG + 1 - g.$$

さらに、$\psi(L)$の最小距離は$n - degG$以上である。ただしここにgは$\overline{X}$の種数(genus)である。

このGoppaの命題は良く知られたリーマン・ロッホの定理から簡単に導き出せます。このようにして構成される符号は一般に代数幾何的符号と呼ばれます。

さらに、Gが一点p_0の定数倍の場合を一点符号と呼びます。すなわち$G = (k-1)p_0$の場合で、言い換えると$\overline{X}$上の有理関数でp_0で$k-1$位以下の極を持ち、他に極を持たないもの全体が成す線形空間Lを考えることになり、上と同様に写像$\psi : L \to \mathbb{F}_q{}^n$を、$f \in L$に対して$\psi(f) = (f(p_1), f(p_2), \ldots, f(p_n))$と決めると、$\psi(L) \subset \mathbb{F}_q{}^n$が符号の役割を果たします。この場合Goppaによる命題は次のようになります。

Proposition(Goppa). $max(0, 2g-1) \leq k-1 \leq min(n-1, n+2g-2)$の時$\psi$は単射で、

$$dim\ \psi(L) = k - g.$$

さらに、$\psi(L)$の最小距離は$n - k + 1$以上である。ただしここにgは$\overline{X}$の種数(genus)である。

4　リードソロモン符号

リードソロモン符号は元々巡回符号の一種として定義されますが、一点符号として表現できることも知られています。Xが射影直線でp_0が無限遠点とした場合の一点符号は、リードソロモン符号と等価です。具体的に書きますと、有限体$\mathbb{F}_q$のq個の元を$x_0, x_1, \ldots, x_n$とし、

$$RS(n,k) = \{(f(x_1), f(x_2), \ldots, f(x_n)) | f(x) \text{は}$$

$$\text{次数}k-1\text{以下の}\mathbb{F}_q\text{係数多項式}\}$$

とすると、$RS(n,k)$は$n(=q-1)$次元ベクトル空間の中のk次元部分空間で、これがリードソロモン符号の代数幾何学的な表示を与えます。

5　リードソロモン符号の復号

符号の復号とは、数学的には$\mathbb{F}_q{}^n$の元を一つ与えたとき、それに一番近い$\psi(L)$の元を探すこと、と言うことができます。これは写像ψで移すもとの元を探すこととと等価です。リードソロモン符号$RS(n,k)$について述べると、与えられた$(r_1, r_2, \ldots, r_n)$に対して、$(f(x_1), f(x_2), \ldots, f(x_n))$が一番近くなるなるような$f$を求めることです。実際の通信の状況では、$\mathbb{F}_q$の$k$個分の情報$(m_1, m_2, \ldots, m_k)$を送信したい場合、

$$f(x) = \sum_{i=1}^{i=k} m_i x^{i-1}$$

として、$(s_1, s_2, \ldots, s_n) = (f(x_1), f(x_2), \ldots, f(x_n))$を送信します。受信情報$(r_1, r_2, \ldots, r_n)$から送信情報$(s_1, s_2, \ldots, s_n)$を推定することは、$f(x)$或は$(m_1, m_2, \ldots, m_k)$を求めることと等価で、これによって送信したかった情報$(m_1, m_2, \ldots, m_k)$が求まります。

6　Gruswami-Sudanのアルゴリズム

Gruswami, V. and Sudan, M.(1999)はリードソロモン符号$RS(n,k)$の復号を多項式の因数分解を使って行う次のようなアルゴリズムを提案しました。

Step1. n個の点(x_i, r_i) $(i = 1, 2, \ldots, n)$で大きな次数の零点を持つ2変数多項式$Q(x,y)$を探す。

Step2. $Q(x,y)$を因数分解して$y - f(x)$の形の因子を見つける。

もし伝送誤差が全くなければ、$s_i = r_i$であり、Step1で作る$Q(x,y)$が、目的とする$y - f(x)$の倍元となっていることは想像し易いでしょう。

7　BMS復号法

S. Sakata(2009)はBMS(Berlekamp-Massey-Salkata)復号法という有力な手法を使い、Gruswami-Sudanのアルゴリズムに証明を与えています。その手法は一点符号の場合にも一般化できるものです。一連の議論は符号理論における代数幾何学的手法の有用性を示すもので、さらにこれらの復号法が最近流行のグレーブナー基底の理論とも結び付いて、興味が尽きない研究対象となっています。