誤り訂正符号と数学

河東　泰之

東京大学大学院数理科学研究科

2011 年 11 月 30 日

前回の質問について

1. RSA 方式と DH 方式では，コストや PC 性能を勘案すると，将来的にはどちらが継続して使われる可能性が高いのか．→ 常時さまざまな改良法が提案されており，どちらがいいとは言いかねる．
2. 現実的な時間内では解読できないとあったが，スパコンなど演算能力がはるかに上昇した場合，解読してしまうということはあり得るのか．→ ありうる．もっと難しい暗号にする．
3. RSA 暗号で素数の積をとるとき，数によっては簡単に素数がわかると言うことだったが，どのようにすればよいのか．→ 知られている方法を試す．

4. たまたま偶然，暗号が破られる可能性はどのくらいあるのか．
   → いくらでも自分で小さくできるが，たとえば自分が突然死ぬ可能性の方がずっと高い．
5. RSA 暗号によるデジタル署名の，M の値は何でもいいのか．
   → そうである．
6. RSA 暗号の p と q がもし合成数だったらどのような問題が起こるのか．→ Fermat の小定理などが使えないので，元のメッセージが復元できない．
7. 失敗した場合に，失敗したことを検知することはできるのか．
   → 失敗すればでたらめなメッセージが出てくるので，検知できる．

8. 量子コンピュータで RSA 暗号が破られると聞いたことがあるが，そのとき社会全体としての対策はあるのか．$\longrightarrow$ 現在差し迫った脅威ではないが，もしそうなったらもっと解読が難しい方法を探す．

9. 暗号の仕組みや作成を学ぶ，関わること自体を制限した方がよいのではないか．$\longrightarrow$ 軍事研究と同様だが，費用も掛からず少人数で研究できるので制限は極めて困難と思われる．自由に研究させた方が得ではないか．

10. n が特別な形のときに $n = pq$ が分かる可能性がある，というが，それはどのようなときか．$\longrightarrow$ 場合による．一口には言えない．

誤り訂正符号

データ転送の設定と，想定している誤りの種類について

2 進数のデータを転送する．0 と 1 のうちのいくつかが反対になってしまうことを考える．

これがデータ転送の最小単位に生じる誤りなので，それを発見，訂正したいというのが今の設定である．

0 と 1 の大きなブロックが丸ごとおかしくなるとか，失われるとかいうことは，今想定していない．

データ転送以外にも，デジタルデータを CD-ROM などに記録する際も状況としては同じであり，同様の手法が有効である．

パリティビット

まず古典的かつ単純な方法から始める．

2 進数 7 桁で $2^7 = 128$ 通りのデータが表せるので，英数字には十分である．2 進数 8 桁をしばしば 1 単位にするので，8 桁目には，他の 7 桁に 1 が偶数個あるときは 0，奇数個の時は 1 をつける．これによって，8 桁分の中に常に 1 が偶数個あることになる．この余計な 1 桁をパリティ・ビットという．

1 桁分の誤りが発生すれば，それを含む 8 桁分の中で，1 の数が奇数個になってしまうので，誤りが起きたことがわかる．

しかし，どの桁で間違えたのかはわからないので，これでは誤り訂正は不可能である．

水平垂直パリティ

今度は 2 進数 16 桁分をひとまとめにして，4×4 の形に配置し，縦横 4 桁分についてそれぞれ上の考え方を適用する．

0100	1	$\leftarrow$ 1 の数が偶数になるようにする
1010	0	$\leftarrow$ 1 の数が偶数になるようにする
1110	1	$\leftarrow$ 1 の数が偶数になるようにする
0101	0	$\leftarrow$ 1 の数が偶数になるようにする
0101		$\leftarrow$ 縦方向にも 1 の数が偶数になるようにする

もし，左上隅の 0 が間違って 1 になってしまったとする．このとき，パリティ・ビットから 1 行目と 1 列目に間違いがあることがわかり，左上の誤りを訂正できる．これによって誤りが 1 か所であれば訂正ができる．

Hamming 距離

同じ桁数の 2 進数が二つあったとする．桁ごとに比べ，対応する数字が違っている場合の個数を二つの 2 進数の間の Hamming 距離と言う．たとえば，10100011 と 10110010 の間の Hamming 距離は 2 である．(右から数えて，1 桁目と 5 桁目の 2 か所が食い違っている．)

誤り訂正の考え方

本来のデータである 2 進数に訂正用の桁をいくつか付けて送る．誤りが出れば，正しい数字列のうち，受信したものからもっとも Hamming 距離が小さいものを，本来のデータ (プラス訂正用の桁) と推定する．

線形符号

今回もまた線形代数が有効である．訂正用のデータもつけて，n 桁の 2 進数を 1 回に送る単位とする．「2 で割った余り」の世界で考えれば，n 桁の 2 進数は n 次元ベクトルと思える．「**正しい n 次元ベクトル**」の全体が線形空間をなすとき，そのような符号システムを**線形符号**と言う．

線形空間というのは，定数倍と加法で閉じているということだが，今は 0 倍と 1 倍しかないので，これは考えなくてよい．そこで，$\mathbb{Z}/2\mathbb{Z}$ 上の n 次元線形空間の部分線形空間 V が，空集合でなく，次の条件を満たしているとき V は線形符号である．

$$\vec{x}, \vec{y} \in V \Rightarrow \vec{x} + \vec{y} \in V.$$

例を考えよう．

前に考えた 16 個の $0, 1$ にパリティ・ビット 8 個をつけたものを考える．できるデータは 24 桁の 2 進数なので，これを 24 次元ベクトルと思う．8 個のパリティ・ビットが誤りを示していないベクトルの全体が V である．これが線形符号の条件を満たしていることがわかる．これは 24 次元空間の中の 16 次元部分線形空間である．

前に考えたように，左上の 1 か所を間違えたとする．この時できる 24 次元ベクトル $\vec{x}$ は V の中に入っていない．しかし V の中に $\vec{x}$ からの Hamming 距離が 1 のもの $\vec{y}$ が一つだけあり，これが一番近い．そこで，このいちばん近い $\vec{y}$ が本来のデータであったと推定し，そのうち 16 桁分を採用するのである．

Golay 符号

上の例では，16 桁の 2 進数に 8 桁も余計につけて，1 か所の誤りを訂正できるだけであったが，もっと賢い方法がある．それが Golay 符号である．正確には拡張 Golay 符号と言うものを考える．これは 2 進数 12 桁のデータにさらに 12 桁の訂正用データをつける線形符号で，3 か所までの誤りに対し，「Hamming 距離で一番近いものを選ぶ」という方法で誤りが訂正できる．これは 24 次元空間の中の 12 次元線形部分空間になっている．

12 桁も余計につけて無駄のようだが 3 か所までの誤りを自動的に訂正できるのは強力である．

具体的な式は「Golay 符号」で検索すればすぐに見つかる．

例外構造と 24 次元

上に出てきた「24 次元」というのは特別な意味のある数字である．直接的には Golay 符号は 23 桁で考えてパリティ・ビットをつけたものであり，3 か所までの誤りがぴったり訂正できるのは次の式のおかげである．(C は 2 項係数を表す．)

$$2^{12}({}_{23}\mathrm{C}_0 + {}_{23}\mathrm{C}_1 + {}_{23}\mathrm{C}_2 + {}_{23}\mathrm{C}_3) = 2^{23}.$$

下でも述べるように 24 次元でだけ成り立つ不思議な現象がいろいろあり，24 は特別な数字である．

『博士の愛した数式』で靴のサイズが 24 であるとき，$24 = 4!$ だからいい数字だという話がある．

球の最密充填問題

平面に 10 円玉を重ならないようぎっしり並べることを考える．誰がやってもやる方法が実際に一番ぎっしり詰め込む並べ方である．(数学的な証明を要するが，古くから証明されている．)

3 次元ではどうだろうか．今度は同じ半径の球をできるだけぎっしり詰めて並べる問題である．誰がやっても，これが一番ぎっしり詰める方法だろうというものがある．これが本当に一番ぎっしり詰め込む方法であろう，というのが Kepler 問題であり，2006 年になってやっと， Hales によって本当にそうであることが示された．

ただ，球を「規則的に」並べる場合に限定すれば問題はずっと易しい．この場合を正規充填という．

高次元球最密充填問題

一般に n 次元空間で，同じ半径の球をできるだけぎっしり詰め込むという問題を考える．でたらめな詰め方を許すととても難しいので，「規則的」な詰め方だけを考える．

8 次元まではもっともぎっしり詰め込む方法が知られているが，もっと上の次元では難しい．

しかし，24 次元空間では特別な詰め方があり，それが最良であることが知られている．このとき，球の中心は規則的に並ぶので，その配置を「格子」と呼ぶ．24 次元空間には Leech 格子と呼ばれる特別の並び方があり，中心がこの Leech 格子に並ぶとき，球はもっともぎっしり詰まるのである．

24 次元空間で，球の中心が Leech 格子に並ぶときに，もっともぎっしり球を詰められるであろう，ということはかなり前から予想されていたが，本当に証明されたのは 2009 年のことで，Cohn と Kumar による．

他の次元には Leech 格子の類似物は知られておらず，24 次元でだけ，特別に高い密度で球を詰め込む方法があるのである．(24 次元空間でも 3 次元空間の場合もまねして詰めることはできるが，Leech 格子の方がより高い密度で詰め込める．)

24 次元空間にはほかにも例外的な格子がいくつもあり，現代数学の高度な話題と関連して，現在も盛んに研究されている．

Leech 格子と Golay 符号

Leech 格子は Golay 符号から作ることができる．

拡張 Golay 符号は成分が $0, 1$ の 24 次元ベクトルの集合であったことを思い出す．そこで，通常の 24 次元空間の中で成分が整数のベクトル $\vec{x}$ だけを考え，各成分を 2 で割った余りを考えた時に，それが拡張 Golay 符号として正しいものであるような $\vec{x}$ の集合を考える．

こうしてできる $\vec{x}$ の集合は格子であるが，まだ Leech 格子ではない．この格子を「半分」にして「ねじってふくらませる」方法が知られており，これによって Leech 格子が作れるのである．

有限群

有限集合 $\{1, 2, \ldots, n\}$ をとり，この集合の上の「置換」を考える．たとえば，$1 \longrightarrow 2$, $2 \longrightarrow 3$, $\ldots$, $n-1 \longrightarrow n$, $n \longrightarrow 1$ といったものである．置換は続けて実行することができ，これを置換の積と言う．

たとえば $n = 3$ とし，$1, 2, 3$ を順に $3, 2, 1$ と移す置換 a を $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$ と書く．別の置換として，$1, 2, 3$ を順に $1, 3, 2$ と移す置換 b を $\begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ とする．積 ba は $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ である．

上の話で n を一つ決める．このとき上のような置換がいくつか集まった集合 G を考え，次の性質を満たすものを考える．

1. $1, 2, \ldots, n$ を全部動かさないものも置換と考え，これが G に入っている．
2. 置換 a が G に入っていれば，入れ替え方を逆にした a の逆置換も G に入る．
3. 置換 a, b が G に入っていれば，上の意味での積 ba も G に入る．

このような G を**有限群** という．通常の有限群の定義はこれとは違うが実質的に同じことである．

群と対称性

たとえば立方体を一つ考える．立方体を中心の周りに回転することを考える．たいていの回転では，回転した後の立方体は元の立方体に重ならないが，たまたまぴったり重なることがある．このような特別な回転だけを集めてくれば,「二つの回転を続けて行う」という操作を「積」と思うことにより，有限群とみなせる．(前ページの置換に基づく説明とは表面上違うものだが，実際上同じものである．)

「立方体」のような数学的対象を一つ決めた時，その「対称性」を表す群がしばしば決まる．(有限ではないことも多い．) このような対称性の研究は現代数学において大変重要な手法である．

有限単純群

$n = 6$ としたとき，12，34, 56 を常に一かたまりにして動かす置換だけを考えると，それは $n = 3$ とした置換を考えていることと同じである．このような有限群同士は同じものと思う．掛け算の規則だけが重要である．

また，$n = 7$ として，1234 と 567 のブロックをそれぞれその中だけで入れ替える置換を考えることができるが，このような有限群は基本的なものではなくより基本的なものに分解できる．

もっと複雑な分解も考えたうえでなお，分解不可能な有限群を**有限単純群**と言う．これが最も基本的な有限群であると考えられる．

有限単純群の分類定理

有限単純群をすべて，分類，列挙することが重要な問題として 100 年以上研究されてきた．

有限体の要素を成分とする行列を考えることにより，有限単純群の族が 16 系列作れることがわかっている．そのほかにもっと簡単な例が 2 系列ある．

それ以外にどれだけ有限単純群があるか，というのが根本問題であったが，26 個あってそれで全部であるということが 20 世紀末に証明された．

証明は極めて大部であり，1 万ページ以上に及び，何度も不完全な点が見つかって訂正されている．

Monster

26 個の例外型有限単純群の中でサイズが最大のものは Monster と呼ばれ，その存在は 1980 年に Griess によって示された．そのサイズは約 8×10^{53} で，素因数分解した形で書くと

$$2^{46} \cdot 3^{20} \cdot 5^{9} \cdot 7^{6} \cdot 11^{2} \cdot 13^{3} \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71$$

である．

これを置換を使って作るのはとても困難だが，196883 次元の行列を使っても Monster は作ることができる．

Monster は現代数学におけるもっとも例外的な数学的対象物の一つである．

j-関数

一方，19 世紀から研究されている重要な j-関数というものがある．これは虚部が正である複素数 τ の関数で，$q = e^{2\pi i\tau}$ とおいて，

$$j(\tau) = q^{-1} + 744 + 196884q + 21493760q^2 + \cdots$$

と表される．(これは無限級数で，係数はある規則に従って決まっている．)

McKay は 1970 年代，ここに出てくる係数 196884 と，前ページに出てきた Monster を作る際の行列のサイズ 196883 が「とても近い」ことに気づいた．

当時，これは無意味な偶然と思った人が多かったが，そうではないことが徐々に明らかになっていった．

Moonshine 予想

Conway と Norton の二人は，McKay の指摘を真剣に検討し，19 世紀から伝わるさまざまな無限級数を調べた結果，Moonshine 予想と呼ばれる次のアイディアに到達した．

1. 未知の無限次元の数学的対象があり，その対称性を表す群が Monster である．
2. 群 Monster の各要素に対して，(1) の数学的対象を通じてある無限級数が定まり，それは 19 世紀以来研究されているよい性質を満たす．

もちろん，これを研究するには「未知の数学的対象」が何かが大きなポイントである．

なお，Moonshine と言う名前は「月の輝き」のことではなく，「たわごと」といった意味の俗語である．有限単純群と j-関数に関係があるというのが当時の数学の常識からはずれたたわごとだったからである．「密造酒」の意味もあると言われている．

「未知の数学的対象」は，Frenkel, Lepowsky, Meurman の 3 人によって 1980 年代に発見され，頂点作用素代数と名付けられた．これは大きな前進であったが，その後も Moonshine 予想の残りの部分の解明は極めて難しいままであった．

Borcherds は 1992 年に，予想の残りの無限級数に関する部分を証明し，1998 年にフィールズ賞を受賞した．

Leech 格子と頂点作用素代数

頂点作用素代数とは，理論物理で考えられていた数学的構造を元に，抽象的枠組みとして考え出されたものである．一般に n 次元空間に格子があるとき，そこから頂点作用素代数を作る方法がある．(あとで述べるように理論物理の超弦理論と関係している．)

しかしこれだけでは，目標によく似たものまでしか作れず，群 Monster とは直接対応させられない．

Golay 符号から Leech 格子を作ったとき，標準的な構成法のあと,「半分にしてねじってふくらませる」という方法を取ったのであった．ここでも (同じではないが) 似た方法があり，それによって Monster と関係した「未知の数学的対象」が作れるのである．

量子場の理論

この考え方の背景にある理論物理について簡単に説明する．

「場」とは，電磁場などのように，時空の各点で数値 (あるいはベクトル) が定まっているものである．つまり時空の上の関数を考えている．

量子力学では，物理量を表すものは数ではなく，作用素 (演算子 — 無限次元空間にはたらく行列) である．そこで，時空の上の作用素に値を取る関数を考える．

これを物理的に考えるものが「場の量子論」であり，大きな成功を収めているが，完全な理論的理解にはほど遠い．特に数学的な理解は，根本のところから全く不十分である．

超弦理論と共形場の理論

時空，物質の最小単位として点ではなく，1 次元のひも (弦) を考えるのが弦理論である．これが本当に我々の実際の時空やその中の物質を理解するための理論であるかどうかについてはいろいろな議論があるが，数学的に大変興味深い理論であることは間違いない．(物理学者の Witten はこの関連の業績によりフィールズ賞を 1990 年に受賞した．)

これに関連して，空間 1 次元，時間 1 次元の時空での相対論的場の量子論が現れる．共形対称性と呼ばれる対称性を持つ理論なので，共形場理論と言う．群 Monster に関連して現れた頂点作用素代数は，共形場理論の数学的構造を抽象化したものと考えられる．

私が研究しているのは，共形場理論の数学的構造であり，頂点作用素代数とは別のアプローチである．それは作用素環論と呼ばれる手法を用いるもので，作用素環論は 20 世紀前半に von Neumann によって創始された．

von Neumann は 20 世紀最大の科学者の一人であり，純粋数学，理論物理学，コンピュータ科学，理論経済学などきわめて幅広い分野での偉大な業績がある．純粋数学における彼の最大の業績が作用素環論を Murray と共に創始したことである．

作用素環論は，Connes, Jones の二人のフィールズ賞受賞者をだし，位相幾何学，微分幾何学，表現論，整数論，確率論，理論物理学など多くの分野と関連して発展している．

無限次元線形代数

最後に作用素環論の枠組みを簡単に解説して終わる．

これまでも線形代数の有効性は多く見た．ここでの話では行列のサイズも有限で，中に入る数字も有限の範囲を動くものが重要であった．

今度はサイズが無限，中に入る数字は複素数と言うものを考えたい．これについてはどうやって掛け算をするかも明らかではないが，歴史的には微分方程式や量子力学の研究がきっかけとなって 20 世紀前半に研究が始まった．

数学や理論物理学を学べば，学部上級生，大学院初年級で学ぶ．現在も盛んに研究されており，日本人研究者の貢献も大きい．