

無限列の圧縮可能性

東工大 理学部 小林孝次郎

1. はじめに α を, i 番目のビット $\alpha(i)$ が次の規則で定義されるような 2 進無限列 (以後単に 無限列 と呼ぶ) とする.

$$\alpha(i) = \begin{cases} 1 & i \text{ 番目の Turing 機械 (あるいは FORTRAN プログラム) がある決まった入力 (例えば 0) を与えると停止するとき,} \\ 0 & \text{そうでないとき.} \end{cases}$$

よく知られているように, α は計算可能ではない. その意味で, α は我々人間にとって, 神の啓示によってのみ与えられる一種の 聖なるメッセージ であるといえる. しかし α に含まれるすべてのビットが, 神の啓示を必要とする不可知の情報であるわけではない.

例えば, $n_0, n_1, n_2, \dots$ を次のような FORTRAN プログラムの番号としよう.

```

I = 0
STOP
END

```

```

I = 0
I = 0
STOP
END

```

```

I = 0
I = 0
I = 0
STOP
END

```

...

これらのプログラムすべてが停止することは容易にわかる。
従って $\alpha(n_0)$, $\alpha(n_1)$, $\alpha(n_2)$, ... の値はすべて1である。
つまり α のビットのうちのいくつかは、我々にも知る得るの
である。

もう1つの例をあげよう。今任意の FORTRAN プログラ
ムが与えられたものとし、その番号を n とする。そのプログ
ラムは STOP 文を含んでいると仮定する。その STOP 文を
2つの STOP 文、3つの STOP 文、... 等で置かえて得ら
れる次のようなプログラムを考え、その番号を m_0 , m_1 , ...
とする。

```

:
STOP
:

```

```

:
STOP
STOP
:

```

```

:
STOP
STOP
STOP
:

```

...

(もとの
プログラム)

我々は $\alpha(n)$ の値は知らないかも知れないが、 $\alpha(n)$, $\alpha(m_0)$, $\alpha(m_1)$, ... の値すべてが等しいことは容易にわかる。従って、無限個のビット $\alpha(n)$, $\alpha(m_0)$, $\alpha(m_1)$, ... のうちの1ビットの値がわかるだけで十分である。このことから特に、 α を印刷した聖典から有限ビットが虫食いによって失われるとしても、われわれはそれを修復することができることがわかる。

以上のような意味で、無限列 α は冗長性を持っている。ところで有限列が冗長性を持つ場合にはそれを利用してより短い有限列で表現することができた。我々は、無限列の場合にもそのようなことが可能であるか、という問題を考える。

2. 基本概念 f を自然数から自然数への関数とする。無限列 α は、次の条件を満足するようなアルゴリズム $\mathcal{A}$ と無限列 β が存在するとき f 圧縮可能であるという：任意の n に対し、 $\mathcal{A}$ は β の最初の $f(n)$ ビット $\beta(1)$, $\beta(2)$, ..., $\beta(f(n))$ より $\alpha(n)$ の値を計算する。もう少し正確な定義は次の通りである。

$M_0, M_1, M_2, \dots$ を、無限列を oracle として使う Turing 機械の、ある自然数 n への上げとする。任意の i と無限列 β に対し、 N から N への部分関数 $\psi_{i,f}^\beta$ を次のように定義する。 M_i に oracle として β , 入力として n を与えたとき、 M_i が

β の最初の $f(n)$ ビットの部分以外の部分を見分けてある値 m を出力して停止する場合には $\psi_{i,f}^{\beta}(n) = m$ である。それ以外の場合には $\psi_{i,f}^{\beta}(n)$ の値は定義されない。 α は、 $\alpha = \psi_{i,f}^{\beta}$ なる β が存在するとき、 f 圧縮可能 (f compressible) であるという。

圧縮可能性と Kolmogorov, Chaitin, Martin-Löf, Daley, Schnorr, Loveland らによって研究された randomness の間には、強い関係 (類似性といくつかの本質的な差) がある。このことについては [1] を参照されたい。

3. いくつかの基本的な結果 本節では圧縮可能性に関するいくつかの基本的な結果を要約する。

まず α が $f(n)$ 圧縮可能なら、任意の定数 k に対し α は $f(n) - k$ 圧縮可能である。(α の最初の k ビットを、アルゴリズムの中に繰りこんで記憶しておけばよい。) これは 'constant compressibility theorem' とでもいべきもので、有用な道具である。

このことから直ちに、 $n - f(n)$ が上に有界なら (つまり任意の n に対し $n - f(n) \leq k$ なる定数 k が存在すれば)、任意の α は f 圧縮可能であることがでてくる。従って、我々に興味があるのは、 $n - f(n)$ が上に有界でない場合のみで

ある。

$n - f(n)$ が上に有界でない場合には、「ほとんどすべて」の α が f 圧縮可能でないことを示すことができる。

【定理 1】 $n - f(n)$ が上に有界でない場合には、集合 $\{\alpha \mid \alpha \text{ は } f \text{ 圧縮可能}\}$ は測度 0 のある集合の部分集合になっている。

ここで、すべての無限列の集合 Σ^∞ は 2 体集合 $\Sigma = \{0, 1\}$ の無限積 $\Sigma \times \Sigma \times \dots$ であると考え、 Σ^∞ には Σ の測度 μ' ($\mu'(\{0\}) = \mu'(\{1\}) = 1/2$) の積測度 μ が入れられているものと考えている。上の定理の証明は容易であるので省略する。

次に我々は、無限列のいろいろなクラス C といろいろな関数 f に対し

$$\alpha \in C \implies \alpha \text{ は } f \text{ 圧縮可能} \quad (*)$$

というかたちの定理がなりたつかどうかを考えてみる。特に C が、集合 $\{i \mid \alpha(i) = 1\}$ の算術的階層 (arithmetical hierarchy) によって特徴づけられる場合に興味がある。

集合 $\{i \mid \alpha(i) = 1\}$ が帰納的集合 (recursive set) であるとき、 α は 帰納的 であるという。その集合が帰納的に枚挙可能な集合 (recursively enumerable ^{set}) であるとき、 α は 帰納的に枚挙可能 である (r. e. と略する) という。

同じ集合が算術的階層の Δ_2 に入っているとき、つまり

$$i \in \{ i \mid \alpha(i) = 1 \}$$

$$\leftrightarrow \exists x \forall y R(i, x, y) \leftrightarrow \forall x \exists y R'(i, x, y)$$

がなりたつような帰納的述語 (recursive predicate) R ,

R' が存在するとき、 α は Δ_2 無限列 であるという。明らかに

$$\alpha \text{ が帰納的} \Rightarrow \alpha \text{ が r.e.} \Rightarrow \alpha \text{ が } \Delta_2 \text{ 無限列}$$

がなりたつ。つまり、帰納的な α の集合、r.e. な α の集合、

Δ_2 無限列の集合をそれぞれ C_{rec} , $C_{\text{r.e.}}$, C_{Δ_2} で表わすと、

$$C_{\text{rec}} \subseteq C_{\text{r.e.}} \subseteq C_{\Delta_2}$$

である。

α が帰納的な α は 0-圧縮可能である。(α は, oracle

を全く見ないで計算できる。) 従って, $C = C_{\text{rec}}$ なし, どんな

な f に対しても上記の式 (*) はなりたつ。一方 C_{Δ_2} に関し

ては, 次の定理がなりたつ。

【定理 2】 f が帰納的関数で $n - f(n)$ が上に有界である場合には, f 圧縮可能である無限列が C_{Δ_2} 内に存在する。

この定理は, 単純な対角線論法で証明することができる。

この定理は, f が帰納的関数の場合には, 全く自明な場合 (

$n - f(n)$ が上に有界である場合) 以外には上記の式 (*)

はなりたたないことを示している。

そこで残されている問題は, $C = C_{\text{r.e.}}$ の場合に, どんな

f に対し上記の式 (*) が定理にならうるか、ということである。次節でこの問題を考える。

4. r.e. な無限列の圧縮可能性 我々は本節で、 f に関するある付帯条件のもとで、 $\sum_{i=0}^{\infty} 2^{-f(i)}$ が収束するとき、そのときに限り

α が r.e. $\Rightarrow \alpha$ は f 圧縮可能

がなりたつことを示す。このことから、例えば $f(n) = n/2$, $\sqrt{n}$, $2 \log n$, $\log n + 2 \log \log n$ などについては上のことがなりたち、 $f(n) = \log n$, $\log n + \log \log n$ などについては上のことがなりたたないことがわかる。(本稿では、特に指定しない限り $\log$ の底は 2 である。)

【定理 3】 f が $\sum_{i=0}^{\infty} 2^{-f(i)} < \infty$ なる帰納的関数なら、任意の r.e. な無限列 α は f 圧縮可能である。

(証明) α を r.e. な無限列とする。 $\sum_{i=0}^{\infty} 2^{-f(i)}$ は収束するから $\sum_{\substack{1 \leq i \\ \alpha(i)=1}} 2^{-f(i)}$ なる値 (つまり、 $\alpha(i) = 1$ なる i ($1 \leq i$) に対する $2^{-f(i)}$ の和) は有限のある値である。それを Ω とする。 Ω の 2 進表現から小数点を除いて得られる無限列を β とする。ただし $\Omega < 1$ なら β は小数点の右の位置からはじまるものとし、 $\Omega \geq 1$ なら β は 1 ではじまるように整えるものとする。 β のビットのうち、小数点より左のものの数を n_0

とする。

つまり，もし Ω の 2 進表現が

0 . 0 0 1 0 1 1 0 1 1 1 0 1 1 0 ...

なら β は

0 0 1 0 1 1 0 1 1 1 0 1 1 0 ...

で $n_0 = 0$ であり，もし Ω の 2 進表現が

1 1 0 1 . 0 0 1 1 1 0 1 0 0 1 1 0 ...

なら β は

1 1 0 1 0 0 1 1 1 0 1 0 0 1 1 0 ...

で $n_0 = 4$ である．あるところから後 0 が続く表現とあるところから後 1 が続く表現のどちらもが可能な場合（つまり Ω が整数を 2 のべき乗で割ったかたちの場合）は，どちらを採用してもよい．

我々は以下で， β の最初の $n_0 + f(n) + 2$ ビットから $\alpha(n)$ を計算できることを示す．このことから α が $n_0 + f(n) + 2$ 圧縮可能，従って constant compressibility theorem により $f(n)$ 圧縮可能であることがでてくる．

β の最初の $n_0 + f(n) + 2$ ビットの後に all 0 を追加した無限列が 2 進表現であるような値を Ω_1 ，all 1 を追加した無限列が 2 進表現であるような値を Ω_2 とする．（小数点はどの位置を採用する．） Ω_1, Ω_2 は有理数で，その値は β の最

初の $n_0 + f(n) + 2$ ビットから計算できる。さらに

$$\Omega_1 \leq \Omega \leq \Omega_2,$$

$$\Omega_2 - \Omega_1 = 2^{-f(n)-2} = 2^{-f(n)}/4$$

がなりたつ。

次に、 Ω_3, Ω_4 を次の式で定義する。

$$\Omega_3 = \Omega_1 - 2^{-f(n)}/8,$$

$$\Omega_4 = \Omega_2 + 2^{-f(n)}/8.$$

そうすると、この Ω_3, Ω_4 に対し

$$\Omega_3 < \Omega < \Omega_4,$$

$$\Omega_4 - \Omega_3 = 2^{-f(n)}/2$$

がなりたつ。

$\alpha(n)$ の値を決めるには、次のようにすればよい。我々は集合 $\{i \mid \alpha(i) = 1\}$ の要素を次々と enumerate してゆく。 α は r.e. であるから、これは可能である。同時に、どの時刻 t においても、その時刻までに enumerate した $\{i \mid \alpha(i) = 1\}$ の要素に対する $2^{-f(i)}$ の和 $\Omega^{(t)}$ を計算しておく。この $\Omega^{(t)}$ に対しては

$$\Omega^{(t)} \leq \Omega,$$

$$\lim_{t \rightarrow \infty} \Omega^{(t)} = \Omega$$

がなりたつ。

もし $\alpha(n) = 1$ なら、 n はいつか enumeration の中にで

てくるから、必ず我々はそのことを知る。

$\alpha(n) = 0$ なら n は決して enumeration にでてこない。

しかも $\Omega_3 < \Omega = \lim_{t \rightarrow \infty} \Omega^{(t)}$ であるから、 n が enumeration にでてこないまま $\Omega_3 < \Omega^{(t)}$ となる時刻 t がある。このとき我々は $\alpha(n) = 0$ であることを知る。なぜなら、 $\alpha(n) = 1$ なら次のような矛盾が得られるからである。

$$\Omega^{(t)} + 2^{-f(n)} \leq \Omega < \Omega_4 = \Omega_3 + 2^{-f(n)}/2$$

$$\therefore \Omega^{(t)} < \Omega_3 - 2^{-f(n)}/2 < \Omega_3.$$

(証明終り)

Ω は G. Chaitin によって導入された ([2])。 Ω の esoteric を解説については文献 [3] を見られたい。文献 [4] は上記文献 [3] の紹介を含んでいる。

次の定理に進む前に、若干の記号を導入し補助定理を証明する。

f を、上に有界でないう単調非減少関数とする。このとき $f^{-1}(m)$ を

$$f^{-1}(m) = \max \{ n \mid f(n) \leq m \}$$

によ、て定義する。 $f^{-1}(m)$ はある値以上のすべての m に対し定義され、 f^{-1} 自身上に有界でないう非減少関数である。さらに f が帰納的関数なら f^{-1} も帰納的部分関数である。

【補助定理 1】 f を上に有界でない単調非減少関数とし、
 m_0 を $f^{-1}(m)$ がすべての $m \geq m_0$ に対し定義されているような値
 m の最小
とす。このとき $\sum_{i=0}^{\infty} 2^{-f(i)}$ が発散すれば、 $\sum_{m=m_0+1}^{\infty} 2^{-m} (f^{-1}(m) - f^{-1}(m-1))$ も発散する。

(証明) 集合 A の要素の数を $\#A$ で表わすことにすれば、

$$\begin{aligned} & \sum_{i=0}^{\infty} 2^{-f(i)} \\ &= \sum_{m=m_0}^{\infty} 2^{-m} \cdot \#\{i \mid f(i) = m\} \\ &= \sum_{m=m_0+1}^{\infty} 2^{-m} \cdot \#\{i \mid f(i) = m\} \\ & \quad + 2^{-m_0} \cdot \#\{i \mid f(i) = m_0\} \\ &= \sum_{m=m_0+1}^{\infty} 2^{-m} (f^{-1}(m) - f^{-1}(m-1)) \\ & \quad + 2^{-m_0} \cdot \#\{i \mid f(i) = m_0\} \end{aligned}$$

がなりたつ。このことより明らか。

(証明終り)

【定理 4】 f を、上に有界でない単調非減少な帰納的関数とする。このとき、 $\sum_{i=0}^{\infty} 2^{-f(i)}$ が発散すれば、 f 圧縮可能でない r.e. な無限列 α が存在する。

(証明) 目的は

$$\forall i \forall \beta [\alpha \neq \psi_{i,f}^{\beta}]$$

がなりたつような、r.e. な α を作ることである。そのために、 i, β に関する対角線論法を使う。今、 α の値をあるところ

まで定義して

$$\forall \beta [\alpha \neq \psi_{0,f} \beta],$$

$$\forall \beta [\alpha \neq \psi_{1,f} \beta],$$

⋮

$$\forall \beta [\alpha \neq \psi_{i-1,f} \beta]$$

がなりたつようにできたとする。我々は、

$$\forall \beta [\alpha \neq \psi_{i,f} \beta]$$

がなりたつようにするには、 α をどう定義すればよいかを考える。

各 m に対し、集合 I_m を $I_m = \{n \mid f^{-1}(m-1) < n \leq f^{-1}(m)\}$ によって定義する。 $\alpha(n)$ の値は、 $I_j \cup I_{j+1} \cup I_{j+2} \cup \dots$ に含まれる n に対しては定義されていないものとする。

n を I_j に含まれるある値、 u を長さ j の 0, 1 のある有限列とし、 $\alpha(n)$ の値を次のように定義してみる。

$$\alpha(n) = \begin{cases} 1 & \psi_{i,f}^{u000\dots}(n) = 0 \text{ のとき,} \\ 0 & \text{そうでないとき.} \end{cases}$$

そうすると、 u の拡張になっていくような β (つまり、 $\beta = u\beta'$ と表わせる β) に対しては、

$$\begin{aligned} \alpha(n) &\neq \psi_{i,f}^{u000\dots}(n) \\ &= \psi_{i,f}^{\beta}(n) \end{aligned}$$

であるから、 $\alpha \neq \psi_{i,f}^{\beta}$ がなりたつ。(ここで、 $n \in I_j$ で

あるから $f(n) = j =$ ‘ u の長さ’ であることを使った。) ところで, u の拡張となっていて β の集合の測度は 2^{-j} である。従って, I_j に含まれる 1 個の n に対してうまく $\alpha(n)$ の値を定義することにより, 測度が 2^{-j} のある集合に含まれる すべての β に対し, $\alpha \neq \psi_{i,f} \beta$ がなりたつようにできる。(しかもその集合は, 長さ j のある 0, 1 の有限列のすべての の集合 の拡張, という極めて簡単なかたちをしている。)

このことから, I_j に含まれるすべての n (それは $f^{-1}(j) - f^{-1}(j-1)$ 個ある) に対してうまく $\alpha(n)$ の値を定義することにより, 測度が $2^{-j} \cdot (f^{-1}(j) - f^{-1}(j-1))$ のある集合に含まれるすべての β に対し, $\alpha \neq \psi_{i,f} \beta$ がなりたつようにできる。このことを I_{j+1} , I_{j+2} , ... についてもくり返してゆけば, $I_j \cup I_{j+1} \cup \dots \cup I_{j'}$ に含まれるすべての n に対してうまく $\alpha(n)$ の値を定義することにより, 測度が

$$\sum_{m=j}^{j'} 2^{-m} \cdot (f^{-1}(m) - f^{-1}(m-1))$$

のある集合に含まれるすべての β に対し, $\alpha \neq \psi_{i,f} \beta$ がなりたつようにできる。

ところで補助定理 1 により, 上の値は j' を大きくするといくらでも大きくなる。このことは, ある j' の段階で, $\alpha \neq \psi_{i,f} \beta$ であるような β の集合 ^X の測度が 1 になることを意味する。しかも X は, ‘ u のすべての拡張の集合’ というかたち

の集合 (α は 0, 1 のある有限列) の有限個の和 (しかも、互に素なものの和) になっている。従って、 α はすべての無限列の集合である。

これで、 α の値をうまく定義して

$$\forall \beta [\alpha \neq \psi_{i,f} \beta]$$

がなりたつようにできることがわかった。 α が r.e. であることは、 α の作り方より明らかである。

(証明終り)

【系 1】 f を、上に有界でないう単調非減少な帰納的関数とする。このとき、次の 2 つの条件は同値である。

$$(1) \sum_{i=0}^{\infty} 2^{-f(i)} \text{ が収束する。}$$

(2) すべての r.e. な α が f 圧縮可能である。

この系を、具体的な f に対して適用すると、次の結果を得る。

【系 2】 すべての整数 $k \geq 0$ と実数 $\varepsilon > 0$ に対し、任意の r.e. な α は $\log n + \log^2 n + \dots + \log^k n + (1+\varepsilon) \log^{k+1} n$ 圧縮可能である。すべての整数 $k \geq 0$ に対し、 $\log n + \log^2 n + \dots + \log^k n + \log^{k+1} n$ 圧縮可能でない、r.e. な α が存在する。

(証明) $k \geq 0$, $\varepsilon \geq 0$ に対し $f_{k,\varepsilon}$ を次のように定義する。

$$f_{k,\varepsilon}(n) = \log n + \log^2 n + \dots + \log^k n + (1+\varepsilon) \log^{k+1} n.$$

この系は，次の式と系1よりすぐでてくる．

$$\begin{aligned} & \sum_{i=i_0}^{\infty} 2^{-f_{k,\varepsilon}(i)} \\ &= \sum_{i=i_0}^{\infty} \frac{1}{n(\log n)(\log^2 n) \dots (\log^{k-1} n)(\log^k n)^{1+\varepsilon}} \\ &\doteq \int_{i_0}^{\infty} \frac{dx}{x(\log x)(\log^2 x) \dots (\log^{k-1} x)(\log^k x)^{1+\varepsilon}} \\ &= \frac{1}{(\log e)^k} \int_{\log^k i_0}^{\infty} \frac{dy}{y^{1+\varepsilon}} \end{aligned}$$

$$\begin{cases} < \infty & \varepsilon > 0 \text{ のとき,} \\ = \infty & \varepsilon = 0 \text{ のとき.} \end{cases}$$

ただしここで i_0 は， $\log^{k+1} n$ が定義されるような n の最小値である．(証明終)

文献

1. K. Kobayashi, On compressibility of infinite sequences, Tokyo Institute of Technology, Dept. of Information Sciences, Research Reports on Information Sciences, No. C-34, March 1981.
2. G. Chaitin, A theory of program size formally identical to information theory, JACM 22

(1975), 329-340.

3. C. Bennett, On random and hard-to-describe numbers, IBM Research Report, RC 7483, May 1979.

4. M. ガードナー, 乱数オメガの値を知れば宇宙の神秘も解明される, サイエンス, 1980年1月. (文献[3]の紹介を含んでいる.)