ショートノート

変動式多重分割による自然数表現法†

中森 眞理雄†† 萩原 洋一††† 高田 正之††

上界が未知の自然数を表現する方式を考察し，変動式多重分割による新しい方式を提案する．一般に，自然数の表現には，指数部可変長の浮動小数点表現（URR など）の指数部を利用することが考えられるが，それらの方式では分割の多重度が固定されている．本論文の方式では，数値が大きくなるほど多重度が大きくなり，比較的少ないビット数で大きな自然数を表現することができる．本論文の方式は次数が 1，2 の場合は既存の自然数表現方式と同じものであり，本論文は既存の方式の優秀さを示すことにもなっている．

1. まえがき

大きさの上限が未知の自然数を表現する方式を決める問題は自然数の表現の問題[1]とよばれる．横尾は自然数の表現法に求められる特に重要な要請として次の 4 つを挙げている[2),3)]．

（C1）語頭条件：いかなる自然数 i の表現 $R(i)$ も他の自然数 j の表現 $R(j)$ の語頭とならないこと；

（C2）辞書式順序：$i<j$ ならば $R(i)<R(j)$（ただし，$R(i)$ と $R(j)$ の間の順序は辞書式順序）；

（C3）完全性：自然数の表現全体の集合 R^* は一意復号可能な表現の極大な集合である；

（C4）長さの単調性：$i<j$ ならば $|R(i)|<|R(j)|$．

指数部可変長の浮動小数点表現は，指数部分だけを見れば，自然数表現と解釈される．このうち，浜田の二重指数分割による URR の指数部分は，（C1-4）をすべて満たすが，指数の増大に伴いビット数が急速に増加する[4)]．中森，土井らは三重指数分割により URR における指数部の長さを改善する方式を提案している[5)]．横尾は，一般的な多重指数分割を検討し，その中では三重指数分割（の指数部分）が実用的な大きさの自然数に対して良好な性質をもつことを示している．

しかし，これらの方法はいずれも，分割の段数（多重度）を固定している．中森は，多重度を可変にし，指数が大きくなるほど多重度を大きくする“変動式多重指数分割”方式を提案している[6)]．ただし，それは Knuth により提案されている方式[7)]と（定義のしかたを除けば）大きく異なるものではない．

本報告では，中森[6)]と Knuth[7)] を特殊な場合として含む一般的な自然数表現方式を提案する．本報告で提案する方式は（C1-4）をすべて満たすものであり，“次数”を適当に設定することにより，出現頻度の高い範囲の自然数を少ないビット数で表現することができる．なお，文献 8）の方式は表現される数の上限が既知であることを前提としており，本論文の対象ではない．

以下では，2 章で自然数の新しい表現方式を提案し，3，4 章で既存の自然数表現と比較する．

2. 変動式多重分割による自然数表現

変動式多重分割は，1 次の場合は中森[6)]により提案されている方式（の指数部）と本質的に同じものであるが，説明をわかりやすくするために 2.1 節で一応述べる．一般化した k 次の方式は 2.2 節で提案する．

2.1 1次の変動式多重分割

本表現方式では語を上位ビットから L, E のフィールドに分け，E をいくつかのサブフィールドに分ける（図 1）．各フィールドの意味は次のとおりである．

L：$1^m\ 0$ の形（$m\geqq 1$）で，値 m を表す．

E：L 部が 1^m0 のとき m 個のフィールド $E_1, E_2, \cdots, E_m$ から成る．E_1 は E_2 の長さを，E_2 は E_3 の長さを，E_{m-1} は E_m の長さを，E_m は表現する自然数を，表す．ただし，$E_1, E_2, \cdots, E_m$ はいずれもけち表現（語頭の 1 を省略）を用い，$E_1, \cdots, E_{m-2}$ が表すのは $E_2, \cdots, E_{m-1}$ のけち表現で省略されるビットを**含まない**長さであり，E_{m-1} が表すのはけち表現で省略されるビットも

† A New Representation of Natural Numbers Based on Variably Multiple Decomposition by MARIO NAKAMORI (Department of Electronics and Computer Science, Faculty of Technology, Tokyo University of Agriculture and Technology), YOICHI HAGIWARA (Computer Center, Tokyo University of Agriculture and Technology) and MASAYUKI TAKATA (Department of Electronics and Computer Science, Faculty of Technology, Tokyo University of Agriculture and Technology).

†† 東京農工大学工学部電子情報工学科
††† 東京農工大学情報処理センター

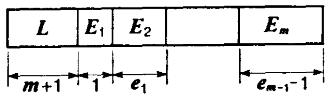

図 1 1次の変動式自然数表現
Fig. 1 Representation of natural numbers of first order based on variably multiple decomposition.

含めた E_m の長さである（E_m の長さについてだけ，けち表現で省略されるビットの扱いが異なることに注意）．$m=1$ のときは E_1 は1ビット（けち表現のため，実際には表示されない），$m\geqq 2$ のときは E_1 は2ビット（けち表現のため，実際には1ビット）．

例 1 表現 11100011101 は各フィールドに分離すると 1110 (1)0 (1)01 (1)1101 となる（括弧内はけち表現のため省略）．すなわち，

L部：1110（すなわち，$m=3$）

E部：$m=3$ であるから，E_1, E_2, E_3 から成る．

定義により，E_1 の長さ$=2$

$E_1=(10)_2=2$ ∴ E_2 の長さ$=E_1+1=3$

$E_2=(101)_2=5$ ∴ E_3 の 長さ$=E_2=5$

$E_3=(11101)_2=29$

したがって，この表現の表す値は 29 である．

2.2 高次の変動式多重分割

k 次 ($k\geqq 2$) の場合も L, E のフィールドに分ける．1次の場合と同様に L部は 1^m0 の形である．$m<k$ ならば URR の指数部分と同じであり（Eを複数のサブフィールドに分けない），$m\geqq k$ ならばEを複数のサブフィールドに分ける（図 2）．すなわち，

E：L 部が $1^m0(m<k)$ のときは長さが m ビット（けち表現により実際には $m-1$ ビット）で，表現する自然数を表す．L部が $1^m0(m\geqq k)$ のときは $m-k+1$ 個のフィールド $E_1, E_2, \cdots, E_{m-k+1}$ から成り，E_1 は E_2 の長さを，E_2 は E_3 の長さを，E_{m-k} は E_{m-k+1} の長さを，E_{m-k+1} は表現する自然数を，表す．ただし，E_1 $E_2, \cdots, E_{m-k+1}$ はいずれもけち表現を用い，$E_1, \cdots,$ E_{m-k-1} が表すのはそれぞれ $E_2, \cdots, E_{m-k}$ のけち表現で省略されるビットを**含まない**長さであり，E_{m-k} が表すのは E_{m-k+1} のけち表現で省略されるビットを**含む**長さから $k-1$ を減じたものである．$m\geqq k$ かつ $m\geqq 2$ のときは E_1 の長さは2ビット（けち表現のため，実際には1ビット）．

例 2 3次（すなわち $k=3$）について一例を示す．表現 111110001110101 は各フィールドに分離すると 111110 (1)0 (1)01 (1)110101 となる（括弧内はけち表現のため省略）．すなわち，

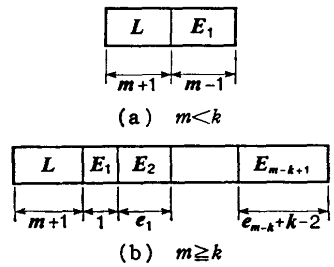

図 2 k次の変動式自然数表現
Fig. 2 Representation of natural numbers of k-th order based on variably multiple decomposition.

L部：111110（すなわち，$m=5$）

E部：$m=5$, $k=3$ であるから，3つのフィールド E_1, E_2, E_3 から成る（∵ $m-k+1=3$）．

定義により，E_1 の長さは2ビットである．

$E_1=(10)_2=2$ ∴ E_2 の長さ$=E_1+1=3$

$E_2=(101)_2=5$ ∴ E_3 の長さ$=E_2+k-1=7$

$E_3=(1110101)_2=117$

したがって，この表現の表す値は 117 である．

3. Knuth の方式について

自然数の表現には Knuth による方式もある[7]．

Knuth の方式では，自然数 i の2進展開を $1\alpha(i)$ とするとき，i を $K(i)$ と表現する．ここで，

$K(0)=0,$

$K(i)=1K(|\alpha(i)|)\alpha(i) \quad (i\geqq 1)$

ただし，$|\alpha(i)|$ はビット列 $\alpha(i)$ の長さである．

Knuth の方式は本論文の2次の変動式多重分割による自然数表現に等しい．

4. 各次の変動式多重分割の比較

1次の変動式多重分割による自然数表現（2.1 節），Knuth の方式すなわち2次の変動式多重分割による自然数表現（2.2 節），URR[4] の指数部の例を**表 1** に示す．また，それらに要する長さを**図 3** に示す．一般の k 次 ($k\geqq 3$) の変動式多重分割による自然数表現に要する長さは図3の範囲では1次，2次とほとんど重複しており描くことが不可能であるため，図3には示していない．なお，図3の範囲で k 次 ($k\geqq 3$) の長さが1次，2次の長さより小さくなることはない．

k 次の変動式多重分割表現に要する長さ $V_k(i)$ は

$$V_k(i)\simeq k+1+\log i+\log\ \log(i/2^{k-2}) + \cdots+\log^n(i/2^{k-2})+\cdots \quad (1)$$

ただし，上式で "≃" は両辺の差が i によらない定数

表 1 変動式多重分割による自然数表現と URR (指数部) の比較
Table 1 Representation of natural number based on variably multiple decomposition and URR (exponent part).

i	URR (exponent)	$k=1$	$k=2$ (Knuth)
1	10	10	10
$1*$	$110\ *$	$110\ 0\ *$	$110\ *$
$1**$	$1110\ **$	$110\ 1\ **$	$1110\ 0\ **$
$1*^{3}$	$11110\ *^{3}$	$1110\ 0\ 00\ *^{3}$	$1110\ 1\ ***$
$1*^{4}$	$111110\ *^{4}$	$1110\ 0\ 01\ *^{4}$	$11110\ 0\ 00\ *^{4}$
$1*^{5}$	$1111110\ *^{5}$	$1110\ 0\ 10\ *^{5}$	$11110\ 0\ 01\ *^{5}$
$1*^{6}$	$11111110\ *^{6}$	$1110\ 0\ 11\ *^{6}$	$11110\ 0\ 10\ *^{6}$
$1*^{7}$	$111111110\ *^{7}$	$1110\ 1\ 000\ *^{7}$	$11110\ 0\ 11\ *^{7}$
$1*^{8}$	$1111111110\ *^{8}$	$1110\ 1\ 001\ *^{8}$	$11110\ 1\ 000\ *^{8}$
…	…………………	…………………	…………………
$1*^{14}$	$111111111111110\ *^{14}$	$1110\ 1\ 111\ *^{14}$	$11110\ 1\ 110\ *^{14}$
$1*^{15}$	$1111111111111110\ *^{15}$	$11110\ 0\ 00\ 0000\ *^{15}$	$11110\ 1\ 111\ *^{15}$
$1*^{16}$	$11111111111111110\ *^{16}$	$11110\ 0\ 00\ 0001\ *^{16}$	$111110\ 0\ 00\ 0000\ *^{16}$
…	…………………	…………………	…………………

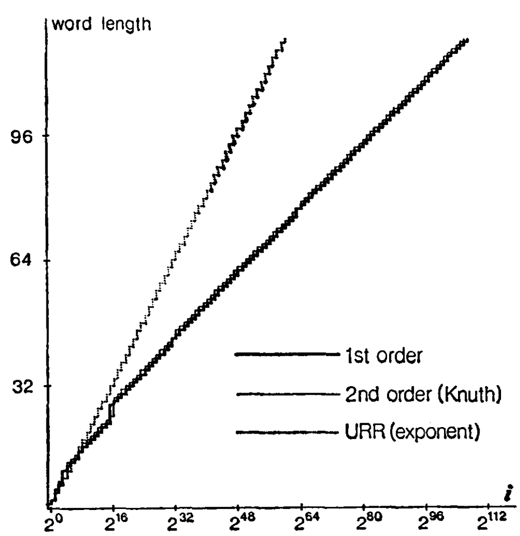

図 3 自然数を表現するのに要するビット数
Fig. 3 The number of bits required to represent natural numbers.

で抑えられることを意味する．また，$\log x$ は2を底とする対数（小数点以下を切り捨て）で，$x<1$ のとき $\log x=1$ と解釈する（以下の式についても同様）．URR の指数部の長さ $U_2(i)$ は

$$U_2(i)=2\log i+2 \quad (2)$$

三重指数分割の指数部の長さ $U_3(i)$ は

$$U_3(i)=\log i+2\log\ \log i+2 \quad (3)$$

一般の多重指数分割の指数部の長さ $U_m(i)$ は

$$U_m(i)=\log i+\log\ \log i+\cdots+2\log^{m-1}i+2 \quad (4)$$

である（m は多重度）[2]．

5. あとがき

自数数の表現の問題を考察し，Knuth や中森の方式を一般化した k 次の変動式多重分割方式を提案し，その中では，Knuth や中森の方式が良いことを示した．高次の変動式多重分割方式は必ずしも効率が良いわけではないが，そのような一般化を通じて，Knuthや中森の方式の良さを裏付けることができた．中森の方式は，実用的な範囲の自然数に対しては，Knuth の方式より1ビット程度性能が良いが，後者は定義が美しい．本方式を暗号へ応用することも検討する価値があろう．

謝辞 本研究について討論いただいた群馬大学横尾英俊講師，東京農工大学並木美太郎助手，中川正樹助教授，高橋延匡教授，植村俊亮教授に感謝します．本研究の一部は，文部省科学研究費補助金（一般研究（C）01550280）の援助を受けた．

参考文献

1) Elias, P.: Universal Codeword Sets and Representations of the Integers, *IEEE Trans. Information Theory*, Vol. IT-21, No. 2, pp. 194-203 (1975).
2) 横尾英俊：自然数の表現の立場から見た多重指数分割浮動小数点表示方式，情報処理学会論文誌，Vol. 30, No. 6, pp. 792-794 (1989).
3) 横尾英俊：自然数の表現に基づく多重指数分割浮動小数点表示方式のクラス，信学論(A)，Vol. 72-A, No. 12, pp. 1998-2004 (1989).
4) 浜田穂積：二重指数分割に基づくデータ長独立実数値表現法 II，情報処理学会論文誌，Vol. 24, No. 2, pp. 149-156 (1983).
5) 中森眞理雄，土井 孝：三重指数分割による数値表現方式について，信学論(A)，Vol. J71-A, No. 7, pp. 1468-1469 (1988).
6) 中森眞理雄：変動型多重指数分割による数値表現方式について，信学論(A)，Vol. J72-A, No. 6, pp. 1009-1011 (1989).
7) Knuth, D. E.: Supernatural Numbers, in Klarner, D. A. (ed.), *The Mathematical Gardner*, pp. 310-325, Prindle Weber and Schmidt, Boston (1980).
8) 松井正一，伊理正夫：あふれのない浮動小数点表示方式，情報処理学会論文誌，Vol. 21, No. 4, pp. 303-313 (1980).

(平成2年2月5日受付)
(平成2年4月17日採録)

中森眞理雄（正会員）

昭和 46 年東京大学工学部計数卒業．昭和 52 年同大学院修了．工学博士．同年東京農工大学工学部数理情報講師．現在助教授．昭和 60 年度文部省在外研究員として西ドイツ，ボン大学で研究．組合せ的数理計画問題，グラフ・ネットワークフロー，アルゴリズム・データ構造，地理情報処理，集積回路の配置配線設計，CAI，ソフトウェア開発手法などの研究と教育に従事．

荻原　洋一（正会員）

昭和 54 年東京電機大学工学部電気通信工学卒業．同年東京農工大学工学部数理情報工学科技官．現在助手．情報処理センターのシステム運用に従事．情報システム，データベース，ネットワークに興味を持っている．著書「パソコン端末利用マニュアル」（オーム社，共著）．

高田　正之（正会員）

昭和 54 年早稲田大学理工学部数学修士修了．同年東京農工大学工学部数理情報工学科助手．以来，プログラム言語，知識処理，音楽情報処理などの研究に従事．現在江戸川大学社会学部応用社会学科講師．情報規格調査会 FORTRAN WG 委員，音楽情報科学研究グループ連絡委員．