

## 対称ヒープの実現とその応用

二村夏彦 寛捷彦 二村良彦  
早稲田大学 理工学部 情報学科

数列に関連しかつ数の大小関係を含む広範囲のプログラム作成問題を分割統治法により解決し易くするために対称ヒープと呼ばれるデータ構造を利用する方法を提案する。まず対称ヒープとは何かを詳しく説明し、かつ与えられた長さ $n$ の数列から $O(n)$ 時間(最良 $n-1$ かつ最悪 $2n-1$ )で対称ヒープを作成するアルゴリズムを示す(ヒープに要するメモリー所要量は $3n$ である)。次に対称ヒープの操作に必要な基本関数およびその計算量を示す。最後に数列に関する6つのプログラム作成問題(そのうち3つは本稿オリジナルと考えられる)を対称ヒープを用いて解決する。

## IMPLEMENTATION OF SYMMETRIC HEAP AND ITS APPLICATIONS

Natuhiko Futamura Katuhiko Kakehi Yoshihiko Futamura  
Department of Information and Computer Science  
School of Science and Engineering, Waseda University  
Ookubo 3-4-1, Sinjuku-ku, Tokyo 169, Japan

Solving programming problems using data structures called symmetric heaps is proposed. Symmetric heaps make the classic divide and conquer method easier to apply to a class of problems concerning number sequences. The definition of the heap and an  $O(n)$  worst case time (between  $n-1$  and  $2n-1$ ) algorithm to generate a symmetric heap from a given sequence of length  $n$  is described. Extra memories to implement a heap is  $3n$  pointers. Then, primitive functions on symmetric heaps and their time complexities are presented. Finally, solutions of six programming problems concerning number sequences are shown as example applications of the proposed method. Three of the problems are considered to be originated from this manuscript.

## 1. はじめに

数列に関連しかつ数の大小関係を含むプログラム作成問題は、整列法に代表されるように理論的にも実用的にも重要である。従ってその種の問題はプログラム作成法またはプログラム変換法の例題としてよく現れる<sup>3, 4, 9, 14)</sup>。一方高性能プログラムの作成法の典型としては分割統治法がある。整列法においてはQUICKソート<sup>10)</sup>、リストマージソート<sup>12)</sup>等がそれを直接的に利用している。しかしその他の多くの例題の解法においては、分割統治法が明白な形で使われていない。分割統治法が問題の解決に直接適用できる場合には、その解法は理解が容易であり、しかもかなり性能のよいプログラムを作り出すことが期待できる。

本稿では数列に関連しかつ数の大小関係を含む広範囲のプログラム作成問題を分割統治法により解決し易くするために、対称ヒープと呼ばれるデータ構造を利用する方法を提案する。以下ではまず対称ヒープとは何かを詳しく説明し、かつ与えられた長さ $n$ の数列から $O(n)$ 時間(最良 $n-1$ かつ最悪 $2n-1$ )で対称ヒープを作成するアルゴリズムを示す(ヒープに要するメモリ所要量は $3n$ である)。次に対称ヒープの操作に必要な基本関数およびその計算量を示す。最後に下記の6つのプログラム作成問題を対称ヒープを用いて解決する。但し以下においては数列の連続した部分列をセグメント、そしてセグメントの長さとセグメントが含む最小値との積をセグメントの面積と呼ぶ。

- (1) セグメントの最大面積<sup>10)</sup>
- (2) 長さ=最小値なるセグメントの最大面積<sup>10)</sup>
- (3) 最大値がその長さ以下のセグメントの最大長<sup>10)</sup>
- (4) 最大値と最小値の差がその長さ以下のセグメントの最大長
- (5) 最大値と最小値の差がその長さを越えるセグメントの最大長
- (6) 純粹上昇列の最大長

上記問題(1)~(5)の素朴な、即ち非能率的な解法はBirdの最大セグメント和<sup>3)</sup>の素朴な解法とよく似ている。即ち $O(n^3)$ アルゴリズムである。しかし最大部分列の和の場合と異なり<sup>1)</sup>、素朴な解法から一般部分計

算<sup>9)</sup>等のプログラム変換によって $O(n)$ アルゴリズムを導出することに筆者等はまだ成功していない。一方、本稿で述べる解法は単純明快に $O(n)$ アルゴリズムを求めることに成功している。それらの本稿における解法は全て直接的な分割統治法に基づいており、(1)~(3)に関しては元の解法とは異なる新しいものである。また問題(4)~(6)は本稿オリジナルであると考えられる(4は寛により、5と6は二村による)。

対称ヒープ自身は増加2進木(increasing binary tree)という名前でも知られている数学的概念<sup>15)</sup>を計算機のデータ構造として実現したものに過ぎない。しかし筆者等の調査の範囲においては、本稿で述べるような高速実現法と応用は他に見出すことが出来なかった。なお以下においてアルゴリズムの記述にはメタLISPと呼ばれるPAD/LISP<sup>5)</sup>表現(付録参照)を用いる。

## 2. 対称ヒープ

数列に対応する減少対称ヒープとは次の3つの性質を有する木構造である：

- (1) 一つの節から出る枝の数は2本以下である。
- (2) 親の値は子供の値以上である。
- (3) 対称(即ち中順)ツリーウォーク<sup>11)</sup>によって元の数列を生成することができる。

上記(1)および(2)は、所謂ヒープ(または優先木)の特質である<sup>1, 2, 12, 17)</sup>。性質(3)を有する故に本稿で用いるヒープを対称ヒープと呼ぶ。これは元の数列の全ての要素の位置関係がヒープの中で保存されることを意味する。但し対称ツリーウォーク(symmetric or in-order tree traversal)とは、木構造をその根から出発し縦優先(depth-first)に辿り、葉に到達したときまたは節に2度目に到達したときにその値を取り出す方法である。数列10, 1, 5, 3, 6, 2, 7, 3, 15, 4, 6に対応する減少対称ヒープを図1に示した。

上記(2)を逆に、即ち(2)'親の値は子供の値以下である、としたものを増加対称ヒープと呼ぶ。これは増加2進木(increasing binary tree)と同じである<sup>15)</sup>が、本稿においてはこれが対称性を持つことを強調する目的で対称ヒープと呼ぶことにする。以下においては増加かあるいは減少かを特に区別する必要がないと

きには、単に対称ヒープと呼ぶことにする。また減少対称ヒープおよび増加対称ヒープを各々減少ヒープおよび増加ヒープと省略して呼ぶ。

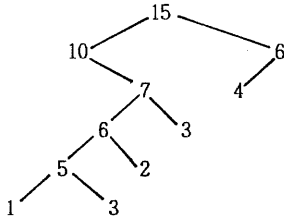


図1: 数列[10 15 3 6 2 7 3 15 4 6] に対する(減少)対称ヒープの図表現

与えられた数列と対応する対称ヒープとの間には次の関係が成立する(証明は付録2参照)。

定理1: 与えられた長さnの数列を $O(n)$ 時間で対応する対称ヒープに変換することが出来る。実際に必要な数値の比較回数は最良 $n-1$ かつ最悪 $2n-1$ である。

ここで数列および対称ヒープに関する基本的関数名とその値を定義する。但し以下において任意の数列を $\sigma$ そして任意の対称ヒープを $h$ で表す。また $\sigma$ の要素数および $h$ の節数を $n$ とする。

(1) `makeheap( $\sigma$ )`:  $\sigma$ に対応する対称ヒープ。

この関数の計算量は定理1より $O(n)$ である。特に減少ヒープを作る場合には`makeheap( $\sigma$ )`、そして増加ヒープを作る場合には`makeinheap( $\sigma$ )`と呼ぶ。

(2) `root(h)`:  $h$ の根の値。

この関数の計算量は $O(1)$ である。

(3) `lchild(h)`:  $h$ の左の子供。

この関数の計算量は $O(1)$ である。

(4) `rchild(h)`:  $h$ の右の子供。

この関数の計算量は $O(1)$ である。

(5) `length(h)`:  $h$ に含まれる節数即ち $n$ 。

必要ならば`makeheap( $\sigma$ )`においてヒープの根に値と共にそれに含まれる節数を入れておくことが、計算量を増やさずに出来る(付録参照)。従ってこの関数の計算量は $O(1)$ である。

(6) `meld(h1, h2)`: 2つの対称ヒープ $h1$ と $h2$ を融合させ

た対称ヒープ。即ち`meld(lchild(h), rchild(h)) = h`。この関数の計算量は平均 $O(\log(n))$ かつ最悪 $O(n)$ である(証明は省略)。

(7) `delete(i, h)`:  $h$ から値 $i$ を含む節を除去することによって得られる対称ヒープ。この関数の計算量は平均 $O(\log(n))$ かつ最悪 $O(n)$ である(証明は省略)。

(8) `split(i, h)`: 値 $i$ を含む節により $h$ を左右に2つの対称ヒープ $h1$ と $h2$ に分割し、その対 $[h1, h2]$ を値とする。但し $i$ は除去される。この関数の計算量は平均 $O(\log(n))$ かつ最悪 $O(n)$ である(証明は省略)。

(9) `null(h)`:  $h$ が空なら`true`, そうでなければ`false`。この関数の計算量は $O(1)$ である。

### 3. 対称ヒープの応用

対称ヒープの重要な応用としては整列法(sorting)があるが、それについては文献<sup>9)</sup>で報告した。ここではセグメントに関して5つ、そして上昇列に関して1つのプログラム作成問題の解を前節で述べた基本関数とPAD/LISP(付録1)を用いて記述する。

#### (1) セグメントの最大面積<sup>10)</sup>

与えられた数列 $\sigma$ に含まれる任意のセグメントを $s$ とする。この時、

$$s \text{の面積} = (s \text{の長さ}) * (s \text{に含まれる最小の値})$$

と定義する。例えば $s = [1 \ 2 \ 3]$ ならばその面積は $3 * 1 = 3$ 。この時間問題は「 $\sigma$ に含まれるセグメントのうちで最大の面積を持つものの面積を求める関数`msa( $\sigma$ )`を作成せよ」である。ここで`msa`はmaximum segment areaの略である。

例: `msa([1]) = 1 * 1 = 1`

$$\text{msa}([1 \ 2 \ 3]) = 2 * 2 = 4$$

$$\text{msa}([1 \ 2 \ 3 \ 1 \ 1]) = 5 * 1 = 5$$

一般に長さ $n$ の数列に含まれるセグメントの個数は $(n * (n + 1)) / 2$ である。従って全てのセグメントの面積を求めてからその最大値を求める手順の計算量は $O(n^3)$ である。しかし次のように考えれば $O(n)$ の手順を作ることができる。

$\sigma$ における最小値を $m_0$ とする。この時 $m_0$ を含む最長のセグメントは $\sigma$ 自身である。従って $m_0$ を含むセグメントの最大面積は $\text{length}(\sigma) * m_0$ である。また $\sigma$ における $m_0$ より大きい値を $m$ とすれば、 $m$ を最小値とするセグメントは $m_0$ を含むことはできない。従って他の値を含む最大面積セグメントは $m_0$ の左側か右側かのどちらかに含まれる。そこで $\sigma$ を $m_0$ によって左右に2つの数列 $\sigma_1$ と $\sigma_2$ に分割し、各々の最大面積を求め、その2者と $\text{length}(\sigma) * m_0$ の中で最大なものが求める面積である。従って $\sigma$ が既に増加ヒープになっているならば $\text{msa}(\sigma)$ は図2のようなになる(増加ヒープは $O(n)$ 時間で作成出来、かつその根には最小値があることに注意)。図2に示された $\text{msa}(\sigma)$ が最悪 $n$ 回の再帰呼び出しにより終了することは $n$ に関する数学的帰納法により容易に証明できる。

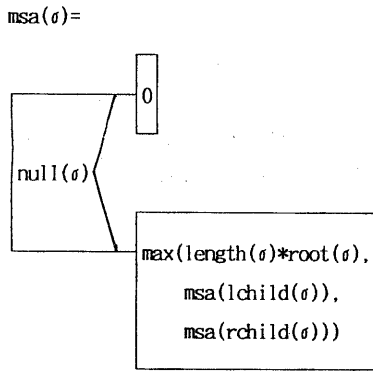


図2: セグメントの最大面積の $O(n)$ 時間の解法

(2) 正方セグメントの最大面積<sup>19)</sup>

その含む最小値と自分自身の長さが等しいセグメントを正方セグメント (square segment) と呼ぶ。この時間問題は「 $\sigma$ に含まれる正方セグメントのうちで最大の面積を持つものの面積を求める関数 $\text{llsl}(\sigma)$ を作成せよ」である。ここで $\text{msa}$ はmaximum square segment areaの略である。

例:  $\text{msa}([1]) = 1 * 1 = 1$

$\text{msa}([1 \ 2 \ 3]) = 2 * 2 = 4$

この問題の解法は、面積を求める際にセグメントの長さで最小値が等しくなければ面積を0にすることを

除いては上記(1)の解法と同じである(図3)。

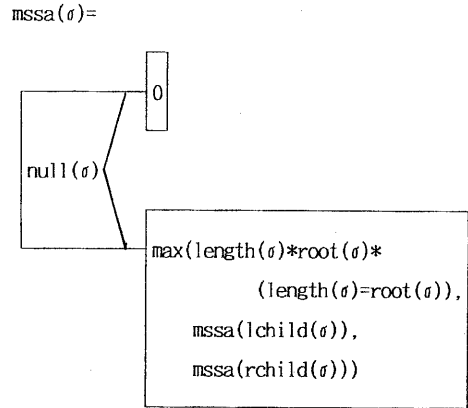


図3: 正方セグメントの最大面積の $O(n)$ 時間の解法  
( $a=b$ )の値は $a=b$ ならば1, そうでなければ0

(3) 低セグメントの最大長<sup>19)</sup>

含まれる最大の値が長さ以下のセグメントを低セグメントと呼ぶ。この時間問題は「 $\sigma$ に含まれる低セグメントのうちで最長のものの長さを求める関数 $\text{llsl}(\sigma)$ を作成せよ」である。ここで $\text{llsl}$ はlongest low segment lengthの略である。

$\text{llsl}(\sigma) =$

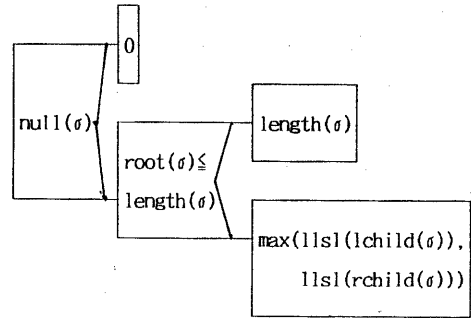


図4: 低セグメントの最大長の $O(n)$ 時間の解法  
( $a \leq b$ )の値は $a \leq b$ ならば1, そうでなければ0

例:  $\text{llsl}([1]) = 1$

$\text{llsl}([1 \ 2 \ 3]) = 3$

$\text{llsl}([1 \ 2 \ 3 \ 1 \ 1]) = 5$

$\sigma$ における最大値を $m_0$ とする。この時 $m_0$ を含む最長のセグメントは $\sigma$ 自身である。従って $m_0 \leq \text{length}(\sigma)$ ならば $\text{length}(\sigma)$ が求める最大セグメント長である。また $\sigma$ における $m_0$ より小さい値を $m$ とすれば、 $m$ を最大値とするセグメントは $m_0$ を含むことはできない。従って $m$ を最大値として含むセグメントは $m_0$ の左側か右側かのどちらかに含まれる。そこで $\sigma$ を $m_0$ によって左右に2つの数列 $\sigma_1$ と $\sigma_2$ に分割し、各々の最長低セグメント長を求め、その中で最大なものが求める長さである。これは上記(1)と丁度双対の議論である。従ってこの問題の解法には減少ヒープを利用すれば良い(図4)。

#### (4) 平セグメントの最大長

含まれる最大の値と最小の値の差がその長さ以下のセグメントを平セグメントと呼ぶ。逆に含まれる最大の値と最小の値の差がその長さを越えるセグメントを粗セグメントと呼ぶ。この時間問題は「 $\sigma$ に含まれる平セグメントのうちで最長のものの長さを求める関数 $\text{lfsl}(\sigma)$ を作成せよ」である。ここで $\text{lfsl}$ はlongest flat segment lengthの略である。

例:  $\text{lfsl}([1])=1$

$\text{lfsl}([1\ 2\ 3])=3$

$\text{lfsl}([1\ 2\ 3\ 1\ 1])=5$

$\text{lfsl}([10\ 1\ 5\ 3\ 6\ 2\ 7\ 3\ 15\ 4\ 6])=7$

以下ではセグメントにおける最大値と最小値との差をセグメントの落差と呼ぶ。落差を $O(1)$ 時間で求めるためにセグメントに対する減少ヒープ $\text{maxheap}$ と増加ヒープ $\text{minheap}$ の両方を常に用意しておく必要がある。またヒープの節には要素の値のみならず、その要素が元の数列において占める位置(先頭から何番目か)も入れておく必要がある。ここで初めに $\sigma$ から $O(n)$ 時間で作られた対応する減少ヒープおよび増加ヒープは各々 $\text{maxheap}$ および $\text{minheap}$ に入っているものとする。

$\sigma$ における最大値を $m_0$ そして最小値を $m_0'$ とする。この時 $m_0$ と $m_0'$ を含む最長のセグメントは $\sigma$ 自身である。従って $m_0 - m_0' \leq \text{length}(\sigma)$ ならば $\text{length}(\sigma)$ が求める最大セグメント長である。 $m_0 - m_0' > \text{length}(\sigma)$ ならば、関数 $\text{lfsl}(\sigma) = \text{lfsl}(\text{root}(\text{maxheap}), \text{minheap})$ により $m_0$ を最大値とする平セグメントの最大長を求める(なければ0)。

$\text{lfsl}(\sigma)$ は $\text{minheap}$ 中の $m_0'$ から $m_0$ に至る経路上の節に対応するセグメントで平なるものがあれば最初のものの長さを返し、経路上に平なるものがなければ0を返す。この関数は平均長 $O(\log(n))$ の経路をたどるだけであるので、その計算量が平均 $O(\log(n))$ であることは明らかである。

また $\sigma$ における $m_0$ より小さい値を $m$ とすれば、 $m$ を最大値とするセグメントは $m_0$ を含むことはできない。従って $m$ を最大値として含むセグメントは $m_0$ の左側か右側かのどちらかに含まれる。そこで $\sigma$ を $m_0$ によって左右に2つの数列 $\sigma_1$ と $\sigma_2$ に分割し、各々の最大平セグメント長を求め、それ等と $\text{lfsl}(\text{root}(\text{maxheap}), \text{minheap})$ との中で最大なものが求める長さである。但し $\sigma_1$ と $\sigma_2$ に対応する増加ヒープは $\text{split}(\text{root}(\text{maxheap}), \text{minheap})$ により平均 $O(\log(n))$ 時間で作る(図5)。図5が平均 $O(n)$ 時間の手順であることの証明は付録3で述べる。

$\text{lfsl}(\sigma) = \text{lfsl}(\text{maxheap}, \text{minheap})$

$\text{lfsl}(a, i) =$

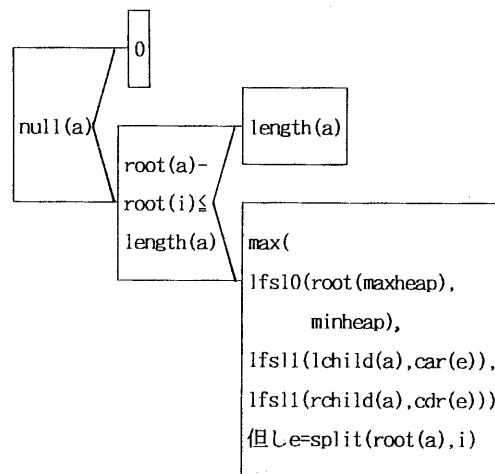


図5: 平セグメントの最大長の平均 $O(n)$ 時間の解法

#### (5) 粗セグメントの最大長

平セグメントと粗セグメントの濃度が同じ(付録3参照)である故に落差の判定に関する大小関係を入れ替えるだけで、平セグメントの最大長を求める方法から平均 $O(n)$ 時間の手順が得られる。

### (6) 純粋上昇列の最大長

$1 \leq i_1 < i_2 < \dots < i_j \leq n$  とするとき、数列  $[x_{i_1} \ x_{i_2} \ \dots \ x_{i_j}]$  の部分上昇列  $[x_{i_1} \ x_{i_2} \ \dots \ x_{i_j}]$  のうちで下記の性質を持つものを純粋上昇列と呼ぶ。

$$1 \leq i_k \leq i_{k+1} \leq i_{k+2} \leq \dots \leq i_{k+j} \leq i_{k+j+1} \quad (1 \leq k < j)$$

例：数列  $[1 \ 2 \ 5 \ 3]$  において  $[1 \ 2 \ 3]$  は上昇列であるが純粋上昇列ではない。 $[1 \ 5]$  は純粋上昇列である。

即ち純粋上昇列は Dijkstra の上昇列<sup>4)</sup> の部分集合である。この時間問題は「 $\sigma$  に含まれる純粋上昇列のうちで最も長いものの長さを求める関数  $llsu(\sigma)$  を作成せよ」である。ここで  $llsu$  は length of longest straight up-sequence の略である。

例： $llsu([1 \ 2 \ 5 \ 3])=3$

$$llsu([3 \ 1 \ 1 \ 2 \ 5 \ 3 \ 4 \ 3])=4$$

$\sigma$  における最大値を  $m_0$  とする。この時最長の純粋上昇列は  $m_0$  で終わるかあるいは  $m_0$  の右にある。そこで  $\sigma$  を  $m_0$  によって左右に2つの数列  $\sigma_1$  と  $\sigma_2$  に分割し、各々の純粋上昇列の最大長を求め、左側のものに1を加えたものと、右側のものとの大きい方を選べばよい(図6)。図6が  $O(n)$  時間の手順であることの証明は上記(1)と同様である。

$$llsu(\sigma) =$$

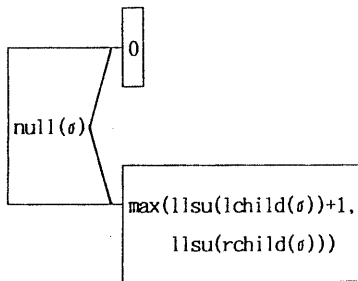


図6：純粋上昇列の最大長の  $O(n)$  時間の解法

### 4. おわりに

数列に関連しかつ数の大小関係を含むプログラム作成問題を統一的に解決するために対称ヒープを使用することを提案した。従来は様々な方法で解決されてい

たセグメントに関する5つの問題が、対称ヒープの利用と分割統治法により単純明快に解決できることを示した。また数列の上昇列に関する新たな問題を作成しその問題も前5者と全く同様に解決可能なことを示した。さらに与えられた数列に対する対称ヒープを  $O(n)$  時間で作成するアルゴリズムも示した。

対称ヒープの実用的応用としては整列法<sup>9)</sup>がある。そのほかの応用については調査中である。また対称ヒープの数学的性質の一部は文献<sup>8)</sup>に述べたが、さらに調査を進めている。

### 5. 参考文献

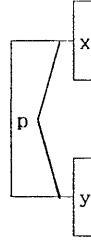
- 1) Aho A.V., Hopcroft J.E. and Ullman J.D.: The Design and Analysis of Computer Algorithms, Addison-Wesley, Reading, Massachusetts, U.S.A. (1974).
- 2) Aho A.V., Hopcroft J.E. and Ullman J.D.: Data Structures and Algorithms, Addison-Wesley Reading, Massachusetts, U.S.A., (1983).
- 3) Bird, R.S.: Algebraic identities for program calculation, The Computer Journal 32(2), 1989.
- 4) Dijkstra, E.W. and Feijen, W.H.J.: A method of programming, Addison-Wesley, Reading, Massachusetts, U.S.A., (1988).
- 5) 二村良彦：プログラム技法-PADによる構造化プログラミング，オーム社，(1984)。
- 6) Futamura, Y., Nogi, K. and Takano, A.: Essence of generalized partial computation, Theoretical Computer Science 90 (1991) 61-79.
- 7) Futamura, Y.: Linear maximum-sum program: an example of generalized partial computation, RIMS workshop on partial evaluation, November, 1991.
- 8) 二村良彦，二村夏彦，寛捷彦：対称ソート，情報処理学会アルゴリズム研究会資料，1992年7月。
- 9) Gries, D.: The Science of Programming, Springer-Verlag, New York, U.S.A., (1981).
- 10) Hoare, C.A.R.: Quicksort, Computer J., Vol.1, No.1, 10-15, (1962).

- 11) Knuth, D.E.: The Art of Computer programming, Vol.1 :Fundamental Algorithms, Addison-Wesley, Reading, Massachusetts, U.S.A.,(1968).
- 12) Knuth, D.E.: The Art of Computer programming, Vol.3 :Sorting and Searching, Addison-Wesley, Reading, Massachusetts, U.S.A.,(1973).
- 13) muLISP-86 Reference Manual, LISP Language Programming Environment Soft Warehouse inc., (1986).
- 14) 永塚弘毅, 山本秀宣: プログラミング演習, プログラムの演繹的導出法の調査研究中間報告書第二分冊, 情報処理振興事業協会技術センター, 1992.
- 15) Stanley, R.T.: Enumerative Combinatorics Vol.1, Wadsworth & Books/Cole Advanced Books, Monterey, California, U.S.A., (1986).
- 16) Swierstra, D. and de Moor, O.: Virtual data structures, RUU-CS-92-16, Department of Computer Science, Utrecht University, 1992.
- 17) Tarjan, R.E.: Data Structures and Algorithms, Society for Industrial and Applied Mathematics, Philadelphia, U.S.A., (1983).
- 18) van der Woud, J.C.S.P.: Rabbitcount:=rabbitcount-1, in J.L.A. van de Snepscheut, Editor, Mathematics of Program Construction, LNCS375, Springer-Verlag, 409-420, 1989.
- 19) 米田信夫: 私信

#### 付録1 メタLISP

LISPプログラムの文法的構文は括弧の重なり具合に依存しており, 非常に読みにくい。その問題点を改善する目的で下記の2点においてLISPを変更した言語がメタLISPである。

- (1) 関数名と引数の関係を示す括弧に丸括弧を, そしてS式を表すためには角括弧を用いる(即ちMcCarthyのM式とは逆)。例えば[]はNIL, [A.B]はAとBとのドット対そして[A B]はAとBのリストを表す。そしてcar([A B])=A, cons(A,B)=[A.B]。
- (2) 条件式(COND (p x) (T y))はPAD<sup>5)</sup>の条件式を用いる。即ち



例: appendのメタLISPによる記述を図7に示した。

append(x,y)=

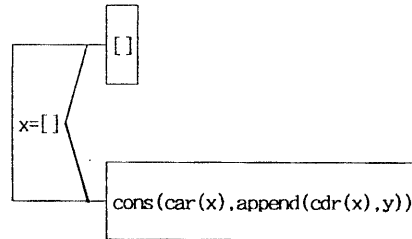


図7: メタLISPによるappendの記述

#### 付録2 対称ヒープの実現例

ここでは対称ヒープの一実現例を示し, それに基づいて諸関数の計算量を評価する。

(1) 対称ヒープのリスト表現: 対称ヒープhを次のようなリストで表現する。

h=[] または

h=[[root(h).length(h)] lchild(h) rchild(h)]

例えば図1に対応するリスト表現は[[15.11][[10.8][[7.7][[6.5][[5.3][[1.1][[]][[3.1][[]]][[2.1][[]]][[3.1][[]]][[6.2][[4.1][[]]][[]]]]]である。

必要ならばroot(h)として要素の値ばかりではなくその要素が元の数列内で占める位置(例えば先頭から何番目であるか)も返すことができる。この時は

root(h)=[value(h).location(h)].

(2) 対称ヒープの作成法

与えられた数列 $[x_n x_{n-1} \dots x_1 x_0]$ を左から右に進みながら対応する(減少)対称ヒープを作る手順makeheapの概略は次の通りである(但し $x_0 = \infty$ )。現在先頭(左端)からj番目の要素 $x_j$ を見ているとする。そして $x_j$ の右に最初に現れる $x_j$ 以上の要素を $x_{m(j)}$ とする。この時 $x_j$ が $x_{m(j)}$ のlchildとなりかつ $[x_{j-1} \dots$

$x_{m(j)+1}$ ]をヒープにしたもの、即ち

$$\text{makeheap}([x_{j-1} \dots x_{m(j)+1}])$$

が $x_j$ のrchildとなる。 $x_j$ のlchildは $x_j$ に至る迄に $x_{m(j)}$ と同様にして決まっていることに注意されたい。従って関数 $m(j)$ が $1 \leq j \leq n$ に関して定義されていれば $x_0$ に到達するまでに対応するヒープは完成する、即ちヒープの作成時間は $O(n)$ であることは明らかである。

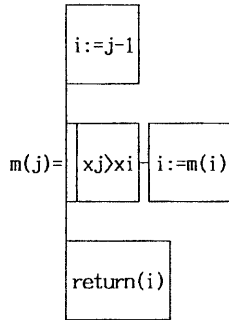


図8 :  $m(j)$ の計算手順

次に関数 $m(j)$ を表として作成する手順 $nln(x)$ を定義しその計算量を評価する。ここで $nln$ はnext larger numberの省略である( $m(j)$ の計算手順は図8)。

但し $m(j)$ の計算中に使われる $m(i)$ の値は前もって計算済みとする。ここで $m(j)$ の値を $j=1 \dots n$ の順で計算する際に要する計算量(即ち反復の回数)を $cm(j)$ と表す。また $cands(j)$ を次のように定義する。これは $x_j$ のnext larger numberの候補者数( $m(i)$ の連鎖を辿って到達出来る要素の個数)である。

$$m^{cands(j)-1}(j-1)=0$$

この時、

$$cm(j) \leq cands(j) \quad \dots (1)$$

$$\begin{aligned} cands(j+1) &= 1 + cands(j) - (cm(j)-1) \\ &= 2 + cands(j) - cm(j) \quad \dots (2) \end{aligned}$$

上記(2)は、 $m(j)$ が求められたことによりその計算の途中の連鎖に現れた $m(i)$ が $x_{j+1}$ のnext larger numberの候補から除外されたことを意味する。

上記(2)より、

$$cands(n+1) + \sum_{j=1}^{n-1} cands(j+1) = 2n + \sum_{j=1}^n cands(n) - \sum_{j=1}^n cm(j)$$

従って、

$$cands(n+1) = 2n+1 - \sum_{j=1}^n cm(j)$$

上記(1)より、 $cm(n+1) \leq cands(n+1)$

従って、

$$cm(n+1) \leq 2n+1 - \sum_{j=1}^n cm(j)$$

故に、 $n+1$

$$\sum_{j=1}^n cm(j) \leq 2n+1 = 2(n+1)-1$$

故に、 $n$

$$\sum_{j=1}^n cm(j) \leq 2n-1$$

$nln(x)$ の計算量は $\sum_{j=1}^n cm(j)$ であるから、その最悪計算

時間は $2n-1$ 、すなわち $O(n)$ である。ちなみに $cm(i)$ のamortized time boundは2である。

対称ヒープを作る上述の議論において数値の比較が行われるのは $nln$ においてのみであり、 $m(j)$ が確定した後は数値の比較は行われない。従って数列からヒープを作る際の数値の比較回数は $O(n)$ である(定理1)。

上述の手続きは実際にmuLISP<sup>13)</sup>により作成され、要素を5000以上持つ数列の整列実験に利用された<sup>6)</sup>。

### 付録3 最長セグメント関数lfs1の計算量

要素数 $n$ の対称ヒープに対する $lfs1(a, i)$ の計算量を $T(n)$ で表す。いま平均的に $T(n) \leq n+k \cdot \log(n+1)$ であることを数学的帰納法により証明する。但し $k$ と $q$ は適当な常数とする。ここで任意のセグメントが平セグメントである確率は $1/2$ であるという数学的事実を使う(粗セグメント\*(長さ/落差<sup>2</sup>)が平セグメント空間への単射となり、逆に平セグメントの最大値を(長さ+1)増やして得られる粗セグメントが平セグメントからの単射となるから)。またlchildの要素数は $0$ と $n-1$ の間を等確率で取るものとする。

(1) Base:  $n=0$ のとき $T(0)=1 \leq k$ 。

(2) Induction step:



$$\begin{aligned}
T(n) &= (\frac{1}{2}n) * \sum_{i=0}^{n-1} (T(i) + T(n-1-i)) + \frac{1}{2} * q * \log(n) \\
&\leq ((\frac{1}{2}n) * \sum_{i=0}^{n-1} (n-1+2k * \log(i+1))) + \frac{1}{2} * q * \log(n) \\
&= \frac{1}{2}(n-1) + (k/n) * \sum_{i=1}^n \log(i) + \frac{1}{2} * q * \log(n) \\
&= \frac{1}{2}(n-1) + (k/n) * \log(n!) + \frac{1}{2} * q * \log(n) \\
&\leq \frac{1}{2}(n-1) + (k/n) * (n * \log(n) - n + \frac{1}{2} \log(2\pi n)) \\
&\quad + \frac{1}{2} * q * \log(n) \\
&= \frac{1}{2}(n-1) + k * \log(n) - k + (k/2n) * \log(2\pi n) \\
&\quad + \frac{1}{2} * q * \log(n) \\
&\leq n + k * \log(n+1) \quad (\text{QED}).
\end{aligned}$$