プログラミング-言語・基礎・実践-　21－2
（1995.　3.　9）

# P区間表とそのプログラミング教育における効果

二村　良彦[†]　白井　千恵子[†]　劉　咏梅[††]　二村　夏彦[‡]　筧　捷彦[†]
[†] 早稲田大学理工学部
[††] 武漢大学
[‡] School of Computer and Information Science, Syracuse University

**e-mail:** futamara@cfi.waseda.ac.jp, cshirai@futamura.info.waseda.ac.jp

あらまし　数列の区間および数平面上の矩形のうちで，与えられた性質を持つ最長あるいは最大のものを求めるプログラミング問題は多い．本稿ではそれらの問題のサブクラスを統一的に解決する目的でP区間表と呼ばれるデータ構造を提案する．P区間表とは数列の各点に対しそれを左（または右）端とし，かつ性質Pを有する最短（もしくは最長）の区間をあらわす表である．本稿では，まずP区間表とその作成法について述べ，次に区間や矩形に関する問題に対する適用例を報告する．また，P区間表を用いた新しいアルゴリズムを解説し，最後に，プログラミング教育にP区間表を応用した効果について報告する．

キーワード　**データ構造，アルゴリズム，プログラミング，教育，区間，矩形**

# P-Segment Table and its Effectiveness to Programming Education

Yoshihiko FUTAMURA[†] Chieko SHIRAI[†] Yongmei LIU[††] Natsuhiko FUTAMURA[‡] KatsuhikoKAKEHI[†]
[†] School of Science & Engineering, Waseda University
[††] Wuhan University
[‡] School of Computer and Information Science, Syracuse University

**e-mail:** futamara@cfi.waseda.ac.jp, cshirai@futamura.info.waseda.ac.jp,

Abstract　We will propose a new data structure called P-segment tables to solve a class of problems in computational geometry concerning lines and rectangles. P-segment table is such a one dimensional array as t, each i-th element of which points to the end of segment (i, t(i)) having some property P. This paper describes P-segment tables and their applications to several well known problems and to new problems developed by ourselves. The data structure has been taught to students in our programming course to evaluate its effectiveness to students' problem solving ability. Results of the evaluation are also described.

key words　**data structure, algorithm, programming, education, segment, rectangle**

## 1　はじめに

数列の区間および数平面上の矩形の中で，与えられた関係Pを満たす最長あるいは最大なものを求めるプログラミング問題は多い．これらの問題は実用性があり，かつ，問題自体が明白であるので，プログラミングの教科書等 [3, 7] にしばしば引用される．問題が明白である一方，その効率の良い解法は複雑になることが多い．また，これらの問題を出題している文献は解法が特殊で適用範囲が狭かったり [10]，あるいはガイドラインのみで具体性が欠如している [3]．したがって，多くの学生はそれらの問題に対する最適な解法を得ることや理解することに困難を感じている．

本稿では数列の区間および数平面上の矩形に関する多くの問題に対し最適な解法を系統的に得る目的でP区間表と呼ばれるデータ構造を提案する．P区間表とは数列の各点に対しそれを左（または右）端とし，かつ性質Pを有する最短（もしくは最長）の区間をあらわす表である．また，P区間表を用いて新しいアルゴリズムを開発すると同時に，プログラミング教育にも応用した結果について報告する．以下ではまず最短P区間表とその作成法について述べ，区間や矩形に関する問題に対する適用例を報告する．その中には，最長素均衡区間問題及び最大矩形体積問題の二つの新例題を含む．次に最長P間表とその作成法について述べ，その適用例を報告する．その中には，数列が単調な場合の最長平区間問題の $O(n)$ の解法も含まれる．最後にP区間表のプログラミング教育における効果について報告する．なお本稿では区間とは数列の連続する部分列をさす．また，アルゴリズムはPAD（ISO DIS8631）によって記す．

## 2　最短P区間表

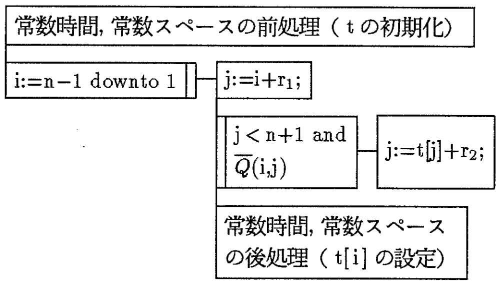
与えられた数列xの最短P区間表tとは，数列上の各点i（$1 \leq i \leq n$）に対して与えられた関係Pを満足する最小の右端j（$i \leq j \leq n$）(もしくは左端j（$1 \leq j \leq i$）) を対応させた表である．右端を対応させた表を最短右P区間表，左端を対応させた表を最短左P区間表と呼ぶ．例えば最短右P区間表は表1のようになる．

表1：最短右P区間表

| 座標 | 1 | 2 | … | i | … | n |
|---|---|---|---|---|---|---|
| 値 | x[1] | x[2] | … | x[i] | … | x[n] |
| 右端 | $t_1$ | $t_2$ | … | $t_i$ | … | $t_n$ |

ただし，$t_i = \min\{j \mid P(i,j) \wedge i \leq j \leq n\}$

与えられた関係Pに対し，常数時間およびスペースで計算できるある関係Qと整数 $r_{i(i=1,2)} (0 \leq r_i \leq 1)$ を設定することができれば，最短右P区間表は下記のアルゴリズムにより $O(n)$ 時間，およびスペースで作成することができる（証明 付録1参照）．ただし，前処理，後処理では制御変数の変更や区間表に対し必要以上の操作を行わないものとする．




ちなみに，最短左P区間表は上記のアルゴリズムで左端から調べるようにし（i:=2 to n），$j:=i-r_1$，$0<j$，$j:=t[j]-r_2$ と書き換えるだけで作成できる．

次に最短P区間表の適用例として，数列の区間および数平面上の矩形に関する問題において最長あるいは最大なものを求めるプログラム作成問題の解法を示す．問題の難易度を分かりやすくするために，[3] の問題に関しては [3] の問題番号をEの後に記す．ダイクストラは問題番号が大きいほど，難しい問題としている．

**最長素均衡区間** 均衡区間とは，0と1のみからなる数列（0–1数列）において，0と1の個数が等しい区間である．素均衡区間とは，均衡区間であって，左右に2分割することによって2つの均衡区間を得ることができないものをいう．問題は，与えられた0–1数列において長さが最大の素均衡区間を $O(n)$ 時間かつスペースで求めることである．例えば，数列<u>110100</u>10 の均衡区間最大長 [3] は，8であるが，素均衡区間最大長は6（下線が引かれた区間の長さ）である．

この問題は均衡区間最大長 [3]E55 と似ているが，こちらの方が難しい問題である．

この問題は，最短P区間表として素均衡区間表を作成することによって解決できる．これは，前述の最短右P区間表作成のプログラムにおいて，関係P(i,j) として「[i,j] は均衡区間」，Q(i,j) として「$x[i] \neq x[j]$」そして $r_1 = r_2 = 1$ と組み込むだけでよい（証明 付録2参照）．したがって最長均衡区間はそのような区間の中で最長のものを選べばよい．例えば 11010010 に対する素均衡区間表は表2のようになる．

表2：素均衡区間表

| 座標 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| 値 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| 右端 | 6 | 3 | 4 | 5 | nil | 7 | 8 | nil |

**最短右区間表作成問題** 最短右区間表とは数列 x の各点 i に対し，i よりも右にある x[i] よりも小さい要素の中で最も i に近い要素 j を対応させた表である．問題はこの表を $O(n)$ 時間及びスペースで作成することである．例えば，数列 6,3,5,8,6,3 に対する最短右区間表は下記のようになる．

表 3：最短右区間表

| 座標 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 値 | 6 | 3 | 5 | 8 | 6 | 3 |
| 右端 | 2 | nil | 6 | 5 | 6 | nil |

この問題は，前述の最短右 P 区間表のプログラムにおいて関係 P(i,j) として「i から始まり右端（x[j]）が左端（x[i]）より小さい」，Q(i,j) として「x[i] > x[j]」そして $r_1 = 1$, $r_2 = 0$, nil=n+1 と組み込むことによって解決できる．

**最大区間面積 [10]** 区間の面積とは，区間の長さと区間の最小値との積である．問題は与えられた数列に含まれる全ての区間の中で最大面積を持つものを $O(n)$ で求めることである．

この問題は数列の各要素 i に対して，それが最小である最長の区間 [el, er] を求めることによって容易に解決できる．最長の区間 [el, er] は上述の最短右区間表を利用すれば er=t[i]−1, el に対しても同様に最短左区間表 $t'$ を作成すれば el=$t'$[i]+1 となる．よって最大区間面積はその様な区間の中で面積最大なものを選べばよい [4, 5].

**最大空部分行列** 0 と 1 のみからなる行列が与えられたときその空部分行列とは，要素がすべて 0 である部分行列である．問題は与えられた行列に含まれる空部分行列の中で，面積が最大なものを $O(N)$ 時間かつスペースで求めることである（ただし N は行列に含まれる要素数）．例えば，表 4 における行列 M の最大空部分行列の面積は 5 である．

この問題は [3] の問題 E63 を一般化したものであり，シミのついた布からシミを含まない最大な正置矩形を見つけるなどの応用も考えられる問題である．同様な問題として計算幾何学の分野では，(1 要素の個数)$*\log^2$(1 要素の個数) の方法 [1] が知られていて，N が大きい場合に有効な方法と考えられる．

この問題は下記に定義する累積表を利用し，問題を最大区間面積問題に還元することによって容易に解決できる．

**定義 1** M を 0 と 1 のみからなる行列とする．この時各 0 要素 M(i,j) から上方に向かう j 列上の 0 の連の長さを A(i,j) とする．また 1 要素に対しては A(i,j)=0 とする．このようにして得られる行列 A を M の累積表と呼ぶ（表 4 参照）．

表 4：行列 M とその累積表 A

$$M = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

| | | | | | |
|---|---|---|---|---|---|
| 1 | 0 | 1 | 1 | 0 | 1 |
| 2 | 1 | 2 | 0 | 1 | 2 |
| 3 | 0 | 3 | 0 | 0 | 0 |
| 4 | 0 | 4 | 1 | 0 | 0 |
| 5 | 0 | 5 | 2 | 1 | 1 |

行列の要素数を N とするとその累積表は $O(N)$ 時間で出来ることは明らかである．最大空部分行列問題は与えられた行列の累積表を作成し，累積表の各行に対して最大区間面積を求め，その中から面積最大のものを選べばよい．したがって，最大空部分行列問題も $O(N)$ 時間およびスペースで解決できる．

**特定の点を含む最大空部分行列** これは指定された点（例えば A）を含む最大空部分列を求める問題である．

この問題は A を含むという制約により，A を含む行上の 0 の連（以下基準線と呼ぶ）の上側と下側の多角形から，上辺および底辺を選び，それらを組み合わせて矩形を作り，そのような矩形の中で面積が最大なものを選べば良い．上辺を選ぶには上方向の累積表の基準線の行に対し，A の左側の最短左区間表および A の右側の最短右区間表を作成し，A から左右の最短区間表を辿り，累積表の値が大きい方へ辺を順次伸ばしていけば良い．底辺に対しても下方向の累積表を同様に利用すれば良い．上辺と底辺の本数は基準線の長さを u とすれば，次の定理により平均 $O(\log u)$ 個（最悪 u 個）である．したがって，この問題は既に上下方向への累積表が作成されているならば，平均の場合は上辺と底辺をそれぞれ $O(\log u)$ 個組み合わせるので，$O(\log^2 u)$ 時間，区間表を作成するのに $O(u)$ 時間かかるので平均 $O(u)$ 時間である．最悪の場合は上辺と底辺をそれぞれ u 個組み合わせなければならないので，$O(u^2)$ 時間かかる．スペースは $O(u)$ である．

**定理 1** 長さ n の一様乱数列が与えられた時，その最大値から始まり左端（または右端）で終わる単調純減少列の最大長は平均 $O(\log n)$ である（証明 付録 3 参照）．

上の問題を計算幾何学の用語で言えば次の定理が得られる．

**定理 2** x 軸に関して単調な多角形 P および多角形内の x 軸上の点 A が与えられた時，A を含みかつ P に含まれる最大正置矩形は平均 $O(u)$，最悪 $O(u^2)$ 時間および O(u) スペースで求めることが出来る（ただし u は P に含まれる x 軸の長さ）．

**最大矩形体積** 矩形とは与えられた行列の部分行列で，矩形体積は矩形の最小値と矩形の面積の積である．問題は，矩形の中で体積が最大なものを求めることである．

この問題は最大区間面積問題の3次元化した問題で，この問題およびその最適解は筆者等の調査の範囲では見出すことが出来なかった．ここでは(1)最悪$O(\mathrm{N}^{1.5})$時間および(2)平均$O(\mathrm{N}\log\mathrm{N})$時間，最悪$O(\mathrm{N}^2)$の2解法を示す（共に$O(\mathrm{N})$スペース）．

**解法1（[5]の簡略版）** 与えられた行列をM(n,m)とする（$\mathrm{N}=\mathrm{n}*\mathrm{m}$かつ$\mathrm{n}\leq\mathrm{m}$とする）．$1\leq x,y\leq \mathrm{n}$なる任意のx,yの組合せ（$O(\mathrm{n}^2)$個）に対して，長さmの数列$\min(\{\mathrm{M}(i,j)\mid j=1, x\leq i\leq y\}),\ldots,\min(\{\mathrm{M}(i,j)\mid j=\mathrm{m}, x\leq i\leq y\})$に対する最大区間面積問題を解き，$O(\mathrm{n}^2)$個の中から最大値を求めればよい．従ってこの手順の計算量は$O(\mathrm{mn}^2)$である．$\min(\{\mathrm{M}(i,j)\mid x\leq i\leq y\})$の値は，x=1,y=1から始めx=1,y=2,...,x=1,y=n,...,x=2,y=2,...,x=n,y=nの順序で計算すれば各々が常数時間で計算できることに注意されたい．この方法の特徴は非常に単純なことである．

**解法2** Mの各点M(i,j)を最小値とする全ての矩形は，その点から上下左右各方向へ探して最初に出会うその点より小さい4点に囲まれる矩形に含まれる．その矩形の大きさは定理3より平均$O(\log\mathrm{n}\times\log\mathrm{m})$である．この矩形を各点から常数時間で求める為に，前もって各行に対して左右の最小区間表，そして各列については上下の最小区間表を作っておく．更にMの要素を降順にソートしておき，値の大きな点からMと同サイズの別の行列M′に入れてゆく（M′の初期値は0）．すると各点M(i,j)が処理される時にはM′上ではその点が0点を除いて最小である．ここで0点を1そしてその他の点を0と考えなおせば問題はM(i,j)を含む空部分行列を，平均サイズ$O(\log\mathrm{n}\times\log\mathrm{m})$の範囲で探せば良いことになる．M(i,j)が新たにM′に入った時点で累積表におけるM(i,j)とその上下の部分だけ変更すれば，累積表は正しく維持できる．従って定理3よりこの累積表の書き換え作業は$O(\log\mathrm{n})$である．定理2より累積表が出来ていれば指定された点を含む最大空部分行列問題は$O(\mathrm{u})$（この場合$\mathrm{u}=\log\mathrm{m}$）で解ける．従って$O(\mathrm{N}\log\mathrm{m})$で各点に対する最大矩形を求めることが出来る．ただしソートに$O(\mathrm{N}\log\mathrm{N})$掛かる為，その時間が上界となる．

**定理3** 一様乱数列が与えられた時，上述の最大区間面積問題に現れた区間[el,er]の平均長は$O(\log\mathrm{n})$である（証明 付録4参照）．

**最長平区間[7]** 平区間とは，最大値と最小値の差が長さ以下の区間である．問題は与えられた数列に含まれるすべての平区間の中で長さが最大になるものを求めることである．例えば，数列6,<u>2,3,1</u>の平区間最大長は，3（下線が引かれた区間の長さ）である．

筆者等の調査の範囲では本問題に対する最適解は不明である．ここでは平均$O(\mathrm{n}\log\log\mathrm{n})$時間，最悪$O(\mathrm{n}^2)$時間の解法を示す．なお，与えられた数列が単調の場合については，$O(\mathrm{n})$の解答を得ている．その解法は最長P区間表の箇所ににおいて記す．

**一般の数列に対する解法** まず，与えられた数列に対する最短右区間表rsおよび最短左区間表lsを作成する．同様に大きい要素を対応させた最短右区間表rbと最短左区間表lbを作成する．次に数列xの各点iに対し，x[i]が最大値となる区間[el,er]において，最長の平区間を求める．これはiから最短左区間表lsを辿るポインタをleft，最短右区間表rsを辿るポインタをrightとし，x[left]とx[right]の大きい方のポインタを一つずつ動かしながら，区間[left,right]が区間[el,er]と等しくなるか，あるいは越えるまで，x[p]=max(x[left],x[right])なるpを選び，x[p]が最小値，x[i]が最大値となる最長の区間[fl,fr]が平区間であるか調べることによって求められる．ただし，区間[el,er]および[fl,fr]はel=lb[i]+1, er=rb[i]−1, fl=max(el,ls[p]+1), fr=min(er,rs[p]−1)とすれば良いことに注意されたい．この解法は数列の各点iに対し，x[i]が最大値となる最長平区間を求めるのに，定理1と定理3により平均$O(\log\log\mathrm{n})$時間要するので，数列全体に対する最長平区間を求めるには，平均$O(\mathrm{n}\log\log\mathrm{n})$時間要する．また，与えられた数列が単調な場合に最悪となり$O(\mathrm{n}^2)$時間かかる．スペースは$O(\mathrm{n})$である．

## 3 最長P区間表

与えられた数列xの最長P区間表tとは，数列上の各点i（1≤i≤n）に対して与えられた関係Pを満足する最大の右端j（i≤j≤n）（もしくは左端j（1≤j≤i））を対応させた表である．右端を対応させた表を最長右P区間表，左端を対応させた表を最長左P区間表と呼ぶ．

与えられた関係Pに対し，最長右P区間表は下記のアルゴリズムにより作成することができる．ただし，前処理，後処理では制御変数の変更や区間表に対し必要以上の操作を行わないものとする．

常数時間，常数スペースの前処理（tの初期化）

i:=n−1 downto 1 ── iよりも大きなjに対しt[j]を常数回参照しt[i]を設定

常数時間，常数スペースの後処理

この時反復の中が常数時間とスペースで計算できれば最長右P区間表は$O(\mathrm{n})$時間およびスペースで作成できることは明らかである．ちなみに，最長左P区間表は上記のアルゴリズムを数列の左から調べるようにし（i:=2 to n），i よりも小さい j に対し t[j] を常数回参照し t[i] を設定と書き換えれば作成できる．

次に最長 P 区間表の適用例として次の二つの問題を報告する．

**均衡区間最大長 [3]E55** 均衡区間とは，0 と 1 のみからなる数列が与えられたとき，0 と 1 の個数が等しい区間である．問題は均衡区間の中で長さが最大なものを求めることである．例えば，数列<u>11010010</u>の均衡区間最大長は，8（下線が引かれた区間の長さ）である．

この問題は下記に述べる性質 1,2 を利用して，数列の各点 i から始まる最長の均衡区間表を対応させた表，即ち最長均衡区間表を作成することで解決できる．ちなみに，数列の各点 i に対し高々2 回区間表を参照することによって，i から始まる最長均衡区間右端を決定できるので，最長均衡区間表は $O(\mathrm{n})$ 時間及びスペースで作成できる．

**性質 1** [a,b] を均衡区間とする．このとき [b+1,c] が b+1 から始まる最長均衡区間ならば，a から始まる最長均衡区間は [a,c] である．

**証明** $\mathrm{d} > \mathrm{c}$ が a から始まる均衡区間 $[\mathrm{a}, \mathrm{d}]$ をなすと仮定する．区間 $[\mathrm{a}, \mathrm{d}]$ は，

$$[\mathrm{a}, \mathrm{d}] = [\mathrm{a}, \mathrm{b}] + [\mathrm{b}+1, \mathrm{d}]$$

に分けられる．ここで，$[\mathrm{a}, \mathrm{b}]$ は均衡区間である．$[\mathrm{a}, \mathrm{d}]$ が均衡区間であるためには，$[\mathrm{b}+1, \mathrm{d}]$ も均衡区間でなければならない．しかし，これは区間 $[\mathrm{b}+1, \mathrm{c}]$ の最長性に矛盾する．よって，$[\mathrm{a}, \mathrm{d}]$ は均衡区間ではない． Q.E.D.

**性質 2** [a,a+1] を均衡区間ではないとし，a+1 から始まる最長均衡区間を [a+1,b] とする．このとき，a と b+1 が均衡をなさないならば，a から始まる均衡区間は存在しない．

**証明** $\mathrm{c} > \mathrm{b}+1$ が a から始まる均衡区間 $[\mathrm{a}, \mathrm{c}]$ をなすと仮定する．区間 $[\mathrm{a}, \mathrm{c}]$ は，

$$[\mathrm{a}, \mathrm{c}] = [\mathrm{a}, \mathrm{a}] + [\mathrm{a}+1, \mathrm{b}] + [\mathrm{b}+1, \mathrm{c}]$$

に分けられる．ここで，$[\mathrm{a}+1, \mathrm{b}]$ は均衡区間である．今，$\mathrm{x}[\mathrm{a}] = 1$ とすると a と b+1 が均衡をなさないことより，$\mathrm{x}[\mathrm{b}+1] = 1$ と書ける．ところで，$[\mathrm{a}, \mathrm{c}]$ が均衡をなすためには区間 $[\mathrm{b}+1, \mathrm{c}]$ は，0 の個数が一つ多くなければならない．しかし，区間 $[\mathrm{a}+1, \mathrm{b}]$ の最長性により b+1 から始まる全ての区間は 1 の個数の方が多い．よって，区間 $[\mathrm{a}, \mathrm{c}]$ は均衡区間ではない． Q.E.D.

**最長平区間 [7]** 平区間とは，最大値と最小値の差が長さ以下の区間である．問題は与えられた数列に含まれるすべての平区間の中で長さが最大になるものを求めることである．

ここでは与えられた数列が単調の場合について $O(\mathrm{n})$ の解答を示す．それは筆者等の共同研究者の一人により発表されたもの [9] とはまったく異なる新解法である．最初にこの問題の解法に必要な関数，定義，定理を導入する．ただし，扱う数列は単調増加な数列のみとする．

区間 $[\mathrm{a}, \mathrm{b}]$ の平坦度を表す関数 $\mathrm{f}(\mathrm{a}, \mathrm{b})$ を $\mathrm{f}(\mathrm{a}, \mathrm{b}) = (\mathrm{b} - \mathrm{a} + 1) - (\mathrm{x}[\mathrm{b}] - \mathrm{x}[\mathrm{a}])$ とする．

**定義 2** 次の 3 つの条件を満たすとき，[a,b] を最平区間（flattest segment）という．

1. $\mathrm{f}(\mathrm{a}, \mathrm{b}) \geq 1$
2. $\mathrm{c} \leq \mathrm{b}$ に対して，$\mathrm{f}(\mathrm{a}, \mathrm{c}) \leq \mathrm{f}(\mathrm{a}, \mathrm{b})$
3. $\mathrm{b} < \mathrm{d}$ に対して，$\mathrm{f}(\mathrm{a}, \mathrm{d}) < \mathrm{f}(\mathrm{a}, \mathrm{b})$

定義 2 より，a から始まる最平区間を求める関数 g(a) は，下記のようになる（証明 付録 5 参照）．

$$\mathrm{g}(\mathrm{a}) = \begin{cases} \mathrm{g}(\mathrm{a}+1) & : \mathrm{f}(\mathrm{a}, \mathrm{g}(\mathrm{a}+1)) \geq 1 \\ \mathrm{a} & : \mathrm{otherwise} \end{cases}$$

**定理 4** [a,b] を平区間，[b+1,c] を最平区間とする．このとき，[a,c] が平区間でないならば [a,b] は最長平区間である．

**証明** $\mathrm{d} > \mathrm{b}$ が a から始まる最長平区間 $[\mathrm{a}, \mathrm{d}]$ をなすとする．

$$\begin{aligned} \mathrm{f}(\mathrm{a,d}) &= \mathrm{f}(\mathrm{a}, \mathrm{b}) - (\mathrm{x}[\mathrm{b}+1] - \mathrm{x}[\mathrm{b}]) + \mathrm{f}(\mathrm{b}+1, \mathrm{d}) \\ &\leq \mathrm{f}(\mathrm{a}, \mathrm{b}) - (\mathrm{x}[\mathrm{b}+1] - \mathrm{x}[\mathrm{b}]) + \mathrm{f}(\mathrm{b}+1, \mathrm{c}) \\ &= \mathrm{f}(\mathrm{a}, \mathrm{c}) < 0 \end{aligned}$$

よって，区間 [a,d] は平区間ではない．したがって，a から始まる最長平区間は [a,b] である． Q.E.D.

**定理 5** [a,b] を平区間，[b+1,c] を最平区間とする．このとき，f(a,c)=0 ならば，[a,c] は最長平区間である．

**証明** $\mathrm{d} > \mathrm{c}$ が a から始まる最長平区間 $[\mathrm{a}, \mathrm{c}]$ をなすとする．

$$\begin{aligned} \mathrm{f}(\mathrm{a,d}) &= \mathrm{f}(\mathrm{a}, \mathrm{b}) - (\mathrm{x}[\mathrm{b}+1] - \mathrm{x}[\mathrm{b}]) + \mathrm{f}(\mathrm{b}+1, \mathrm{d}) \\ &< \mathrm{f}(\mathrm{a}, \mathrm{b}) - (\mathrm{x}[\mathrm{b}+1] - \mathrm{x}[\mathrm{b}]) + \mathrm{f}(\mathrm{b}+1, \mathrm{c}) \\ &= \mathrm{f}(\mathrm{a}, \mathrm{c}) = 0 \end{aligned}$$

よって，区間 [a,d] は平区間ではない．したがって，a から始まる最長平区間は [a,c] である． Q.E.D.

**単調数列に対する解法 1** まず，与えられた数列に対する最平区間表を作成する．これは最平区間を求める関数 g により $O(\mathrm{n})$ で作成できる．次に数列の各点 i から始まる最長平区間を求める．これは，ある点 A（初期値は数列の左端）との平坦度を平坦度が 0 になるか，即ち定理 5 が成り立つか，あるいは平坦度が負になるか，即ち定理 4 が成り立つまで調べていく．ここで定理 4 または定理 5 が成り立つ点を A′ と呼ぶ．そして区間 [A′,A] において平坦度が小さい順にソートし（ビンソートをする），小

さい方から順に平区間に最平区間を繋いでいく．繋ぎ終えたら，A′を新たにＡとして，同様の作業を数列の右端まで行う．数列が単調な場合は平坦度の増え方は高々1である．この性質より，最平区間を繋ぐ毎に定理4または5が成立し，各点から始まる最長平区間を$O(1)$で求めることが出来る．従って$O(\mathrm{n})$で全ての点から始まる最長平区間を求めることが出来る．

## 4　教育上の有効性と問題点

Ｐ区間表のプログラミング教育上の有効性を調査するために，学部2年生のプログラミング演習において数列の区間や数平面上の矩形に関するプログラミング作成問題を次の順序で出題した．最初に，最短右区間表を作成させた．すると半数の学生が独力で最短右区間表を$O(\mathrm{n})$時間かつスペースで作成した．次に最大区間面積を出題した．すると最短右区間表を$O(\mathrm{n})$で作成した学生の多くがこの問題を$O(\mathrm{n})$時間かつスペースで解答した．なお，最大区間面積は以前に問題のみを出題すると$O(\mathrm{n})$で解答する学生は皆無で，$O(\mathrm{n}^2)$で解答する学生も数人であったため出題を中止していた問題である．更に，Ｐ区間表の作成法を指導した後，難問として有名な最大空部分行列問題を出題した．すると102人中88人が$O(\mathrm{N})$時間かつスペースで解答できた．更により難しい最大矩形体積問題を出題した．すると102人中21人が$O(\mathrm{N}^{1.5})$時間および$O(\mathrm{N})$スペース，1人が平均$O(\mathrm{N}\log\mathrm{N})$時間，最悪$O(\mathrm{N}^2)$時間で解答できた．なお，この数字は学生のレポートを採点した結果である．以上より，筆者等はＰ区間表が区間や矩形に関する問題を解く上で重要なデータ構造であると実感した．

しかし，Ｐ区間表が簡単なデータ構造で，かつＰ区間表を用いると容易に効率の良い解法を得ることができるため，Ｐ区間表に固執し，より効率の良い解法を見落とす問題が生じた．与えられた数列が単調な場合の最長平区間問題がその例である（付録6参照）．他に類似の例があり，Ｐ区間表を教える際にはそのような点に留意するよう付言する必要性を感じた．

## 5　おわりに

(1) Ｐ区間表を提案し，数列の区間および数平面上の矩形に関するプログラミング問題を効率よく解決できることを示した．(2) 2つの新例題，即ち最長素均衡区間および最大矩形体積とそのＰ区間表の技法を利用した解法を提案した．(3) 単調な数列に対する最長平区間問題を最長Ｐ区間表を利用して$O(\mathrm{n})$で解決した．(4) 従来は大半の学生が従来解答を示されても理解できなかった問題を，本稿で述べた教育方法により自力で解答できるようになった．したがって筆者等は区間や矩形に関するプログラミング教育にはデータ構造としてＰ区間表が必須であるという結論を得た．

今後の課題は引き続きＰ区間表を適用できる問題を調べること，およびプログラミング教育に有効な技法を考案することである．

## 参考文献


[1] Aggarwal, A. and Suri, S. : Fast Algorithms for Computing the Largest Empty Rectangle, *Proc. of the Third Annual Symposium on Computational Geometry*,(1987) pp278–290

[2] Aho, A.V., Hopcroft, J.E. and Ullman, J.D. : *Data Structures and Algorithms*, Addison–Wesley Reading, Massachusetts, U.S.A.,1983.

[3] Dijkstra, E.W. and Feijen, W.H.J. : *A method of programming*, Addison–Wesley Reading, Massachusetts, U.S.A.,1988.

[4] 二村夏彦，二村良彦，筧捷彦：対称ヒープの実現とその応用，情報処理学会アルゴリズム研究会 28–7, 1992年7月.

[5] 二村夏彦，二村良彦，筧捷彦：最長区間表とその応用，日本ソフトウェア科学会第10回大会，B9–3,1993.

[6] 二村良彦，白井千恵子，劉咏梅，二村夏彦，筧捷彦：ダイクストラの演習問題の系統的解決法，日本ソフトウェア科学会第11回大会，C1–3, 1994

[7] 筧捷彦：プログラミング演習A，早稲田大学理工学部情報学科講義,1991.

[8] Knuth, D.E. : *The Art of Computer programming*, Vol.3 : Sorting and Searching, Addison–Wesley, Reading, Massachusetts, U.S.A.,1973.

[9] Liu Yongmei : Linear Algorithm to the Monotonic Case of the Maximum Flat Segment Length Problem,*Softwere Engineering Institiute Research Report, Wuhan University*, 1994年5月

[10] van der Woude : Rabbitcount := Rabbitcount − 1, in J.L.A. van de Snepscheut, Editor, Mathematics of Program Construction, LNCS375, Springer–Verlag, 1989, pp409–420

[11] Vuillemin, J. : A Unifying Look at Data Structures, *Comm. of ACM*, Vol.23, No.4, April (1980) pp229–239

**付録 1** 最短右 P 区間表の計算量 [4]

**証明** 次の関数 t[i] により各点に対する最短右 P 区間表を作成する．ただし，t[i] の計算中に使われる t[j] の値は前もって計算済みとする（t[n+1]=nil）．

$$t[i]=\boxed{j:=i+r_1}$$
$$\boxed{j \le n+1 \text{ and } \overline{Q}(i,j)} - \boxed{j:=t[j]+r_2}$$
$$\boxed{\text{return}(j);}$$

ここで，t[i] の値を i=n,...,1 の順で計算する際に実行される比較回数を ct(i) と表す．また cands(i) を次のように定義する．これは関数 t[i] 実行前の値で，x[i] に対する右端の右隣になり得る候補者の数（即ち t[j+1] の連鎖を辿って到達できる要素の個数）である．

$$\text{cands}(n)=1 \quad \text{かつ} \quad t^{\text{cands}(i)-1}[i+1]=0$$

このとき

$$\text{ct}(i) \le \text{cands}(i) \tag{1}$$

$$\begin{aligned}\text{cands}(i-1) &= 1+\text{cands}(i)-(\text{ct}(i)-1)\\ &= 2+\text{cands}(i)-\text{ct}(i)\end{aligned} \tag{2}$$

上式 (2) は，t[i] 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 t[j] が x[i−1] の右端の右隣の候補から除外されたことを意味する．上式 (2) より，

$$\sum_{i=2}^{n}\text{cands}(i-1)=2(n-1)+\sum_{i=2}^{n}\text{cands}(i)-\sum_{i=2}^{n}\text{ct}(i)$$

$$\text{cands}(1)+\sum_{i=2}^{n-1}\text{cands}(i)$$
$$=2(n-1)+\sum_{i=2}^{n}\text{cands}(i)-\sum_{i=2}^{n}\text{ct}(i)$$

したがって，

$$\text{cands}(1)=2n-1-\sum_{i=2}^{n}\text{ct}(i)$$

上式 (1) より，

$$\text{ct}(1) \le \text{cands}(1)$$
$$\text{ct}(1) \le 2n-1-\sum_{i=2}^{n}\text{ct}(i)$$
$$\text{故に} \quad \sum_{i=1}^{n}\text{ct}(i) \le 2n-1$$

したがって，全ての右端の最悪計算時間は 2n−1，すなわち $O(n)$ である．ちなみに ct(i) の amortized time bound は 2 である．よって最短 P 区間表の計算量は $O(n)$ 時間かつスペースである．

**付録 2** 素均衡区間表作成の証明

**証明** b を a から区間表を辿って初めて，x[a] ≠ x[b] となる点とする．

(1) b = a + 1 のとき，明らかに区間 [a, b] は，素均衡区間である．

(2) b ≠ a+1 のとき，区間 [a, b] が二つの均衡区間 [a, c] と [c + 1, b] に分けられると仮定する．

$$[a,b]=[a,a]+[a+1,c]+[c+1,b]$$

このとき，x[a] = 1 とすると点 b の条件により x[a+1] = 1 となる．よって，区間 [a, c] が均衡区間であるためには，区間 [a + 1, c] は 0 の個数が一つ多くなければならない．しかし，b の条件により a から始まる全ての区間（c まで）においては 0 の個数は 1 の個数以下である．よって，区間 [a, c] は均衡区間でない．以上より，区間 [a, b] が素均衡区間であることが証明された．　Q.E.D.

**付録 3** 定理 1 の証明 [11]

**証明** 長さ n の一様乱数に対して，その最大値から始まり左端（または右端）で終わる単調純減少列の最大長の平均長を ed(n) とする．

まず，x[i] が最大値であるとし，i の左部分木（1〜i−1）と右部分木（i+1〜n）に分けて考える．左部分木，右部分木の単調純減少列の平均長はそれぞれ，ed(i−1),ed(n−i) と書ける．よって，x[i] が最大値であるときの単調純減少列の最大長の平均は，

$$\frac{ED(i)+(i-1)ed(i-1)+(n-i)ed(n-i)}{n} \tag{3}$$

となる．ただし，

$$ED(i)=\begin{cases} ed(n-i)+1 & : 1 \le i \le \frac{\lceil n \rceil}{2} \\ ed(i-1)+1 & : \frac{\lceil n \rceil}{2}+1 \le i \le n \end{cases}$$

とする．1 から n までの i に関して (3) 式の和を平均すると ed(n) は，

$$\begin{aligned} ed(n) &= \frac{1}{n}\sum_{i=1}^{n}\frac{ED(i)+(i-1)ed(i-1)+(n-i)ed(n-i)}{n} \\ &= \frac{1}{n}+\frac{1}{n^2}\sum_{i=1}^{\frac{\lceil n \rceil}{2}} ed(n-i) \\ &\quad +\frac{1}{n^2}\sum_{i=\frac{\lceil n \rceil}{2}+1}^{n} ed(i-1)+\frac{2}{n^2}\sum_{i=1}^{n-1} i\,ed(i) \end{aligned}$$

次に数学的帰納法により，$ed(n) \le 1+k\log n$ となることを証明する．

base) $n = 1$ のとき，

$ed(1)=1 \le 1+k\log n$

$k > 0$ で成り立つ

induction step) $i < n$ に対して，

$ed(i) \le 1+k\log i$

が成り立つと仮定する．

$$
\begin{aligned}
\mathrm{ed}(n) =& \frac{1}{n} + \frac{1}{n^2}\sum_{i=1}^{\lceil\frac{n}{2}\rceil}\mathrm{ed}(n-i) \\
&+\frac{1}{n^2}\sum_{i=\lceil\frac{n}{2}\rceil+1}^{n}\mathrm{ed}(i-1) + \frac{2}{n^2}\sum_{i=1}^{n-1} i\,\mathrm{ed}(i) \\
\le& \frac{1}{n} + \frac{1}{n^2}\sum_{i=1}^{\lceil\frac{n}{2}\rceil}(1 + k\log(n-i)) \\
&+\frac{1}{n^2}\sum_{i=\lceil\frac{n}{2}\rceil+1}^{n}(1 + k\log(i-1)) \\
&+\frac{2}{n^2}\sum_{i=1}^{n-1} i(1 + k\log i) \\
\le& \frac{2}{n} + \frac{k}{n^2}\sum_{i=1}^{\lceil\frac{n}{2}\rceil}\log n + \frac{k}{n^2}\sum_{i=\lceil\frac{n}{2}\rceil+1}^{n}\log n \\
&+\frac{2}{n^2}\times\left\{\frac{n(n-1)}{2} + k\sum_{i=1}^{n-1} i\log i\right\} \\
\le& \frac{2}{n} + \frac{k}{n}\log n + \frac{n-1}{n} \\
&+\frac{2k}{n^2}\left(\sum_{i=1}^{\lceil\frac{n}{2}\rceil-1} i\log\frac{n}{2} + \sum_{i=\lceil\frac{n}{2}\rceil}^{n-1} i\log n\right) \\
\le& 1 + \frac{1}{n} + \frac{k}{n}\log n + \frac{2k}{n^2}\left(\sum_{i=1}^{n-1} i\log n - \sum_{i=1}^{\lceil\frac{n}{2}\rceil-1} i\right) \\
\le& 1 + \log n - \frac{k(n-2)-4}{4n}
\end{aligned}
$$

$k = 4$ とすれば，$n \ge 2$ に対して

$\mathrm{ed}(n) \le 1 + 4\log n$ が成り立つ．

よって，$O(\log n)$ であることが証明された．Q.E.D.

**付録 4** 定理 3 の証明 [2, 8]

**証明** 長さ n の一様乱数に対して，最大区間面積問題に現れた区間 [el,er] の平均長を $s(n)$ とする．

まず，$x[i]$ が最小値であるとする．$x[i]$ を最小値とする区間は，[1,n] となり，$i \le j \le i-1$ なる $x[j]$ を最小値とする区間，および $i+1 \le j \le n$ なる $x[j]$ を最小値とする区間は，i を含まない．

そこで，i の左部分木（1〜i−1）と右部分木（i+1〜n）に分けて考える．左部分木，右部分木の区間 [el,er] の平均長はそれぞれ，$s(i-1)$, $s(n-i)$ と書ける．よって，$x[i]$ が最小値であるときの区間 [el,er] の平均長は，

$$\frac{n + (i-1)s(i-1) + (n-i)s(n-i)}{n}$$

となる．$x[i]$ から $x[n]$ のいずれかが最小値となる確率は等確率であるから $s(n)$ は，

$$
\begin{aligned}
s(n) &= \frac{1}{n}\sum_{i=1}^{n}\frac{n + (i-1)s(i-1) + (n-i)s(n-i)}{n} \\
&= 1 + \frac{2}{n^2}\sum_{i=1}^{n-1} i s(i)
\end{aligned}
$$

次に数学的帰納法により，$s(n) \le 1 + k\log n$ となることを証明する．

base) $n = 1$ のとき，

$s(1) = 1 \le 1 + k\log n$

$k > 0$ で成り立つ

induction step) $i < n$ に対して，

$s(i) \le 1 + k\log i$

が成り立つと仮定する．

$$
\begin{aligned}
s(n) =& 1 + \frac{2}{n^2}\sum_{i=1}^{n-1} i s(i) \\
\le& 1 + \frac{2}{n^2}\sum_{i=1}^{n-1} i(1 + k\log i) \\
\le& 1 + \frac{2}{n^2}\times\left\{\frac{n(n-1)}{2} + k\sum_{i=1}^{n-1} i\log i\right\} \\
\le& 2 - \frac{1}{n} + \frac{2k}{n^2}\left(\sum_{i=1}^{\lceil\frac{n}{2}\rceil-1} i\log\frac{n}{2} + \sum_{i=\lceil\frac{n}{2}\rceil}^{n-1} i\log n\right) \\
\le& 2 - \frac{1}{n} + \frac{2k}{n^2}\left(\sum_{i=1}^{n-1} i\log n - \sum_{i=1}^{\lceil\frac{n}{2}\rceil-1} i\right) \\
\le& 2 - \frac{k}{4} + k\log n - \frac{1}{n}\left(1 + k\log n - \frac{k}{2}\right)
\end{aligned}
$$

$k = 4$ とすれば，$s(n) \le 1 + 4\log n$

よって，$O(\log n)$ であることが証明された．Q.E.D.

**付録 5** 関数 g の証明

**証明** (1) $f(a, g(a+1)) \ge 1$ のとき，

$$
\begin{aligned}
f(a, g(a+1)) =& f(a, a+1) + f(a+1, g(a+1)) - 1 \\
>& f(a, a+1) + f(a+1, d) - 1 \\
>& f(a, d)
\end{aligned}
$$

(2) $f(a, g(a+1)) < 1$ のとき，

$$
\begin{aligned}
f(a, d) =& f(a, a+1) + f(a+1, d) - 1 \\
<& f(a, a+1) + f(a+1, g(a+1)) - 1 \\
<& f(a, g(a+1)) < 1
\end{aligned}
$$

$f(a, d) < f(a, a)$ Q.E.D.

**付録 6 単調数列に対する解法 2** 定理 4 と同様に，区間に最平区間を繋いで平区間にならなければそれより後まで伸びる平区間は存在しない．したがって最長平区間の長さ max は，下記のアルゴリズムにより求めることができる（解法 1 と比べソーティングが不要なことに注意）．