数理解析研究所講究録 1325 巻 2003 年 221-226

MPR 束における「結び既約元」の数理的性質について

– On mathematical properties of join-irreducibles in MPR-Lattices –

宮川 幹平　成嶋 弘
東海大学福岡短期大学

Abstract

With biological sciences such as taxonomy, cladistics and phylogeny as a background, the principle of maximum parsimony also called Wagner Parsimony has been mathematically formulated and then a mathematical and algorithmic theory has been developing. Recently, a clear method for the character-state minimization problem called the First MPR (Most-Parsimonious Reconstruction) Problem under linearly ordered character-states has been presented by Hanazawa et al. (1995), Narushima and Hanazawa (1997). From a phylogenetic point of view, Minaka (1993) has introduced a partial ordering on the set of MPRs to investigate the relationships among the MPRs. Miyakawa and Narushima (1999) have shown that an MPR-poset induced by this ordering is a complete distributive lattice. In lattice theory, a comcept of "join-irreducible" is very important, because the fundamental theorem for finite distributive lattices states that the set of all order ideals of the subposet which consist of the join-irreducibles in finite distributive lattice L, ordered by inclusion, is lattice-isomorphic to L. In this paper, we show some mathematical properties of join-irreducibles in a finite MPR-lattice.

1. 導入

近年，系統分類学，分子系統学等の各分野において，対象となる種や分類単位の形質状態（形態的もしくは遺伝子的な情報）から進化系統樹（進化の道筋）を推定する問題が盛んに議論されている．この問題を大まかに言うと，その形質状態を実際に観察できる種の分類単位（操作的分類単位: Operational taxonomic unit:OTU）の集合が与えられたときに，進化生物学的に適切と考えられる系統樹形と，樹形上の内部分岐点に対応する仮想的分類単位（Hypothetical taxonomic unit:HTU）に適切な形質状態を復元するという問題と考えられる．またその手法としては，何を「進化生物学的に適切」と考える基準とするのかという点も含めて様々な手法が提案されており，本論文ではこの中で最節約法に焦点を当てる．なお，系統樹推定に関する進化生物学的議論について，より詳しくは文献[12, 3, 13, 11]等を参照されたい．

ここで簡単に最節約法について説明する．復元された系統樹において，枝の長さとはその両端点に対応する分類単位の形質状態間距離[1]のことであり，系統樹全体の長さとは各枝の長さの総和のことである．ここで最節約法とは，「対立仮説の中で，最も単純に形質分布を説明できる仮説を選択する」という最節約原理に従い，系統樹全体の長さが最小となる

[1]この距離の定義には，背景となる進化モデルや形質状態集合の定義が反映される．例えば核酸塩基配列データのようにどのような変化も１ステップで行える形質状態を扱う場合にはハミング距離を，形態学的形質によく見られるように，離れた形質状態間の変化にはそれだけ多くのステップがかかるような形質状態を扱う場合にはマンハッタン距離が代表的なものとして挙げられている．より複雑なモデルとしては，各形質状態間の変換に確率的重みを付与したものなどがある．

ように樹形や仮想的分類単位の形質状態を復元する系統樹推定手法である．なお，そのような条件を満たす復元を最節約復元（MPR:Most-Parsimonious Reconstruction）と呼ぶ．この問題は，与えられた樹形の下での仮想的分類単位に対して，最節約原理の下で最適な形質状態を求める問題と，最適な樹形とその仮想的分類単位に対する形質状態を同時に求める問題，という 2 個の問題に分けて論じられている [1]．なお，我々は前者を第 1 最節約復元問題，後者を第 2 最節約復元問題と呼んでいる．本論文では第 1 最節約復元問題について論じる．

第 1 最節約復元問題に関する研究は Farris[1] に始まり，Swofford and Maddison[10] を経て，Hanazawa et. al.[2] によって数理的に厳密な問題の定式化がなされ，またそれと共に，与えられた樹形のもとでの最節約復元を全て再帰的に列挙する簡明なアルゴリズム及びその時間計算量に関する解析が与えられている．また，[10] では，対象としている生物群の共通祖先（系統樹の根となる分類単位）を仮定するとき，系統樹解析において意味のあると思われる復元として，以下の二つが提案された．

ACCTRAN 復元 形質状態の変化ができるだけ根の近くで生じるように形質状態を設定する（形質変換を促進する:Accelerated Transformation）．形質進化の逆転[2]回数は最大化され，逆に形質の収斂[3]は最小化される．

DELTRAN 復元 形質状態の変化ができるだけ末端近くで生じるように形質状態を設定する（形質変換を抑制する:Delayed Transformation）．形質の収斂回数は最大化され，逆に形質進化の逆転は最小化される．

これらの復元に関して，数理的な立場からもその特徴が解析されている [7, 8, 5]．

なお，与えられた系統樹の MPR は上記で挙げた ACCTRAN，DELTRAN だけではなく，一般的には樹形の頂点数について指数関数的に存在する．これら複数の MPR は，それぞれ形質進化に関して異なる背景の仮定[4]を持っていると考えられており，各 MPR の数学的特徴づけと進化生物学的特徴づけの対応が試みられている．

また三中 [12] は，各 MPR の比較や MPR 全体の特徴を調べるため，MPR 全体集合にある半順序を導入した．この順序による MPR 半順序集合 (MPR-poset) は，数学的に見ると完備な分配束 (MPR 束) になっていることが証明されており [4]，そのため，MPR 全体集合に対して，束論的なアプローチが可能になった．本論文では，MPR 束において，束論的に重要な要素である結び既約元を数理的に特徴づけ，結び既約元を構築するためのアルゴリズムを示す．

2. 準備

外点に付値されている樹形が与えられたとき，その木の長さを最小化するような内点への付値を定める問題を第 1 最節約復元問題と呼ぶ．まず，この問題を以下のように定式化する．

[2]逆転 (reversal) とは，一旦変化した形質状態がその子孫で再び元の状態に戻るといった，一般に言う先祖がえり現象のことを指す．

[3]収斂 (convergence) とは，遺伝的に異なる複数の生物が，類似する形質を個別に進化させることを言う．

[4]例えば，ある部分木においては進化が促進又は逆に抑制されるなど．

T を単純な無向木とする．$V(T)$ を T の頂点集合，$V_O(T)$ を外点（次数 1 の頂点）集合，$V_H(T)$ を内点集合，$E(T)$ を T の辺集合とする．混乱する恐れの無い限り，これらを単にそれぞれ V, V_O, V_H, E と書く．また Ω を実数集合もしくは整数集合とする．ここで外点に対する付値関数を $\sigma : V_O \to \Omega$ としたとき，この木構造を我々は **el-木**(*el-tree*) と呼び，$T = (V, E, \sigma)$ と書く．なお，以下の例では Ω を整数集合とする．本論文においての結果は Ω が可算的でなければ成り立たないことに注意されたい．el-木 T が与えられたとき，$\lambda | V_O$（V_O に定義域を制限した λ）が σ と等しいような T の各頂点への付値関数 $\lambda : V \to \Omega$ を T の**復元**（*reconstruction*）と呼ぶ．ここで 1 節で述べた進化生物学的議論との関係を考えると，V_O が操作的分類単位（OTU）の集合，V_H が仮想的分類単位（HTU）の集合，σ が（操作的分類単位に対して）観測された形質状態，λ が各仮想的分類単位に対する形質状態の仮説，にそれぞれ対応している．次に，el-木 T に復元 λ が与えられたとき，各辺 $e \in E$ に対して**長さ** $\ell(e|\lambda)$ を $\ell(e|\lambda) = |\lambda(u) - \lambda(v)|$, $e = \{u, v\}$ と定義する．また復元 λ が与えられたときの T の長さを各辺の長さの総和と定義する．即ち，$L(T|\lambda) = \sum_{e \in E} \ell(e)$．さらに，$T$ の長さの最小値 $L^*(T)$ を以下のように定義する:

$$L^*(T) = \min\{L(T|\lambda) \mid \lambda \text{は} T \text{上の復元}\}.$$

ここで，$L(T|\lambda) = L^*(T)$ となるような復元 λ を T 上の**最節約復元**（以下では *MPR* と書く）と呼び，T 上の MPR 全体の集合を $\mathbf{Rmp}(T)$ と書く．また，各 MPR が頂点 u で取り得る値の集合 $\{\lambda(u) | \lambda \in \mathbf{Rmp}(T)\}$ を u の **MPR-集合**（*MPR-set*）と呼び，S_u と書く．

与えられた el-木 T について，ある頂点 r を根 (root) と定めることで，根付き el-木 $T^{(r)}$ も自然に定義できる．根付き el-木上において，頂点 u が v の親であるとき，$u = p(v)$ と書く．また，根でない外点を特に葉 (leaf) と呼ぶ．ここで注意として，その定義から明らかなように MPR は根に依存しない．根を定めるのは前節で述べたように，対象とする操作的分類単位に対して共通的な祖先となる分類単位を仮定するという進化生物学的観点からか，もしくは（より純粋な数理的問題として）MPR を決定するためのアルゴリズム的必要性からかのどちらかである．

$\mathcal{I} = \{I_i | 1 \leq i \leq n\}$ を Ω 上の（多重）閉区間族としたとき，$\mathcal{I}$ に含まれる全ての閉区間 I_i の両端点を以下のように昇順に並べたとする．

$$x_1 \leq \ldots \leq x_n \leq x_{n+1} \leq \ldots \leq x_{2n}.$$

このとき，x_n と x_{n+1} を閉区間族 $\mathcal{I}$ の**中間 2 点**(*median two points*) と呼び，その列 $\langle x_n, x_{n+1} \rangle$ を $\mathrm{med2}\langle I_1, I_2, \ldots, I_n \rangle$ もしくは $\mathrm{med2}\langle I_i | 1 \leq i \leq n \rangle$ と書く．また，この中間 2 点を端点とする閉区間 $[x_n, x_{n+1}]$ を，閉区間族 $\mathcal{I}$ の**中間区間**(*median interval*) と呼び，$\mathrm{med}\langle I_1, I_2, \ldots, I_n \rangle$ もしくは $\mathrm{med}\langle I_i | 1 \leq i \leq n \rangle$ と書く．

次に，根付き el-木の本体に属する各頂点 u について，Ω 上の閉区間 $I(u)$ を以下のように再帰的に与える:

$$I(u) = \begin{cases} [\sigma(u), \sigma(u)] & (u \text{ が葉のとき}), \\ \mathrm{med}\langle I(v) | u = p(v) \rangle & (\text{それ以外}). \end{cases}$$

但し，$\langle I(v) | u = p(v) \rangle$ は，u の子を $v_1, v_2, \ldots, v_m$ としたときの $\langle I(v_1), I(v_2), \ldots, I(v_m) \rangle$ なる区間列を表すものとする．以下でも同様の表現を用いる．この閉区間 $I(u)$ を u の**特性区

間，また I を T 上の**特性区間写像**と呼ぶ．これらは第1最節約復元問題に関する一連の論文のキーコンセプトである．

次の結果は el-木 T に対して，T の最小長と MPR 全体の集合を特徴づける，第1最節約復元問題の基本定理と呼ぶべき定理である．なお再度の注意として，T の根 r をどのように選んでも，T の最小長そのものの値や MPR 全体の集合には影響しない．

定理 1 (花沢-成嶋-三中 [2]) *T を根付き el-木 $T^{(r)}$ とし，λ を T 上の復元とする．λ が T 上の MPR であるための必要十分条件は各内点 u において，*

$$\lambda(u) \in \begin{cases} I(u) & (u \text{ が根である}), \\ med\langle\ [\lambda(p(u)), \lambda(p(u))], I(v) | u = p(v)\ \rangle & (\text{それ以外の内点}). \end{cases}$$

を満たすことである．

なお，以下では任意の $x \in S_{p(u)}$ に対して $\mathrm{med}\langle [x,x], I(v) | u = p(v)\rangle$ を $S_u|x$ と書く．ここで，本論文での定理の証明に有用な命題を次に引用する．

命題 2 (宮川-成嶋 [4]) *T を根付き el-木，I を T 上の特性区間写像としたとき，T の頂点 u とその子 v に対して，次の 3 通りのいずれか 1 つのみが成立する．*

1. $\min S_u < \min I(v)$,
2. $S_u \subseteq I(v)$,
3. $\max I(v) < \max S_u$.

ここで，V 上の弧集合 $A(E)$ を以下のように定義する：$u, v \in V$ に対して，$(u,v) \in A(E)$ となるのは，$\{u,v\} \in E$ であり，かつ，T を $u = p(v)$ となるような任意の頂点で根付けしたとき，u, v が命題2の 1. を満たすときである．このとき，MPR を以下のように特徴づけすることもできる：

命題 3 *el-木 $T = (V, E, \sigma)$ の復元 λ が T の MPR であるための必要十分条件は，各辺 $e = \{u, v\}$ において，*

$$\begin{cases} \lambda(u) \le \lambda(v) & ((u,v) \in A(E) \text{ のとき}), \\ \lambda(u) \ge \lambda(v) & ((v,u) \in A(E) \text{ のとき}), \\ \lambda(u) = \lambda(v) & ((u,v), (v,u) \notin A(E) \text{ のとき}). \end{cases}$$

を満たすことである．

3. MPR 束と結び既約元

MPR 全体集合 $\mathbf{Rmp}(T)$ 上の二項関係 $\le$ を三中 [3] に従って以下のように定義する：任意の MPR λ, μ に対して，$\lambda \le \mu$ とは，任意の頂点 u において $\lambda(u) \le \mu(u)$ が成り立つことである．このとき，明らかに $\le$ は MPR 全体集合上の半順序である．この順序 $\le$ によって誘導された半順序集合 (**MPR-poset**) に関して，以下のような束論的特徴づけがなされ

定理 4 (**宮川-成嶋** [4]) *el-木 T において，MPR-poset* $(\boldsymbol{Rmp}(T), \leq)$ は完備分配束である．

ここで，MPR 全体集合上の束演算 (結び ($\vee$) と交わり ($\wedge$)) は以下のように定義される：任意の MPR λ, μ とする．u を任意の頂点とするとき，

$$\begin{aligned}(\lambda \vee \mu)(u) &= \min\{\lambda(u), \mu(u)\}\\(\lambda \wedge \mu)(u) &= \max\{\lambda(u), \mu(u)\}\end{aligned}$$

なお本論文であつかう範囲では Ω が可算であるため，MPR-poset は有限である．有限分配束において，重要な概念のひとつが既約 (irreducible) の概念である．ある最小元でない元 x が結び既約 (join-irreducible) であるとは，他の任意の二つの元の結び演算 ($\vee$) によって x を表現できないことを言う．交わり既約も同様である．有限分配束の基本定理として，有限分配束 L において，ある有限半順序集合 P が順序同型を除いて唯一存在し，L と $J(P)$ が束として同型であることが良く知られている．この P とは L の結び既約 (もしくは交わり既約) 元の集合からなる部分半順序集合 (と順序同型) である．つまり，結び既約 (もしくは交わり既約) 元とは有限分配束の生成元ということになる．

次節において，MPR 束における結び既約元を構築する為のアルゴリズムを示す．

4. 主定理

el-木 $T = (V, E, \sigma)$ において，V 上の関係 $\sim$ を，V 上の頂点 x と y を結ぶ T 上のパス P の任意の辺 $e = \{u, v\}$ について，$(u, v), (v, u) \notin A(E)$ であるとき $x \sim y$，と定義する．このとき，明らかに $\sim$ は V 上の同値関係である．ここで，$\sim$ による V の同値類を $B_1, B_2, \cdots, B_k$ とする．$A(E)$ の定義より，任意の MPR λ, $i \in [1, k]$ について，$u, v \in B_i$ であるなら $\lambda(u) = \lambda(v)$ である．よって，$u, v \in B_i$ ならば $S_u = S_v$ が成り立つ．そこで，$S_{B_i} = S_u$ (但し $u \in B_i$) と定義する．ここで， 任意の $i \in [1, k]$, $x \in (\min S_{B_i}, \max S_{B_i}]$ に対して，T の復元 $\lambda_{i,x}$ を以下のように定義する：

1. B_i 上の任意の点 r を根とする，根付き el-木を構成する．
2. $u := r$, $\lambda_{i,x}(u) := x$ と定める．
3. u の子 v について，v が葉であれば停止する．そうでないとき，

$$\lambda_{i,x}(v) := \begin{cases} \mathrm{median}\langle\ \lambda(u), \min S_v, \max S_v\ \rangle & ((u,v) \in A(E) \text{ のとき}),\\ \min S_v & ((v,u) \in A(E) \text{ のとき}),\\ \lambda(u) & ((u,v),(v,u) \notin A(E) \text{ のとき}).\end{cases}$$

   と定める．

4. $v := u$ として，3. を再帰的に繰り返す．

このとき，次の命題が成り立つ．

命題 5 el-木 $T=(V,E,\sigma)$ において，V 上の関係 $\sim$ による同値類を $B_1,B_2,\cdots,B_k$ とする．$i,j\in[1,k]$，$x\in(\min S_{B_i},\max S_{B_i}]$，$y\in(\min S_{B_j},\max S_{B_j}]$ としたとき，$i=j$ かつ $x=y$ は，$\lambda_{i,x}=\lambda_{j,y}$ と必要十分である．

最後に，本論文における主定理を述べる．

定理 6 el-木 $T=(V,E,\sigma)$ において，T の復元 λ が MPR 束 $(\boldsymbol{Rmp}(T),\vee,\wedge)$ の結び既約元である為の必要十分条件は，V 上の関係 $\sim$ による同値類を $B_1,B_2,\cdots,B_k$ としたとき，ある $i\in[1,k]$，$x\in(\min S_{B_i},\max S_{B_i}]$ が存在して，$\lambda=\lambda_{i,x}$ となることである．

References

[1] J. M. Farris, Methods for computing Wagner trees, Systematic Zoology vol.19, pp.83–92, 1970.

[2] M. Hanazawa, H. Narushima and N. Minaka, Generating most parsimonious reconstructions on a tree: a generalization of the Farris-Swofford-Maddison method, Discrete Applied Mathematics vol.56, pp.245–265, 1995.

[3] N. Minaka, Algebraic properties of the most parsimonious reconstructions of the hypothetical ancestors on a given tree, Forma vol.8 no.2, pp.277–296, 1993.

[4] K. Miyakawa and H. Narushima, Lattice-theoretic properties of MPR-posets in phylogeny, 投稿中.

[5] K. Miyakawa and H. Narushima, On mathematical properties of ancestral character-state reconstructions under the delayed transformation optimization, 投稿中.

[6] H. Narushima and M. Hanazawa, A more efficient algorithm for MPR problems in phylogeny, Discrete Applied Mathematics vol.80, pp.231–238, 1997.

[7] H. Narushima and N. Misheva, On characteristics of ancestral character-state reconstructions under the accelerated transformation optimization, Discrete Applied Mathematics vol.122, pp.195–209, 2002.

[8] H. Narushima, On extremal properties of ACCTRAN reconstructions in phylogeny, preprint.

[9] D. F. Robinson, Extending a function on a graph, Discrete Mathematics 6 (1973) 89–99.

[10] D. L. Swofford and W. P. Maddison, Reconstructing ancestral character states under Wagner parsimony, Mathematical Biosciences vol.87, pp.199–229, 1987.

[11] 馬渡峻輔 編著，動物の自然史—現代分類学の多様な展開，pp.183–267，北海道大学図書刊行会，1995.

[12] 三中信宏，組合せ論的視点から見た系統推定—最節約法と離散数学の接点—，千葉中央博物館自然誌研究報告，第 2 巻 第 2 号 pp.83–98，1993.

[13] 三中信宏，生物系統学，東京大学出版会，1997.