Parametric maximum parsimonious reconstruction on trees

Gilles Didier
Institut de Mathématiques de Luminy CNRS-UMR 6206
Campus de Luminy, Case 907
13288 MARSEILLE Cedex 9
didier@iml.univ-mrs.fr

December 7, 2009

Abstract

We give a formal study of the relationships between the transition cost parameters and the generalized maximum parsimonious reconstructions of unknown (ancestral) binary character states $\{0, 1\}$ over a phylogenetic tree. As a main result, we show there are two thresholds λ_n^1 and λ_n^0, generally confounded, associated to each node n of the phylogenetic tree and such that there exists a maximum parsimonious reconstruction associating state 1 to n (*resp.* state 0 to n) if the ratio "10-cost"/"01-cost" is smaller than λ_n^1 (*resp.* greater than λ_n^0). We propose a dynamic programming algorithm computing these thresholds in a time quadratic with the size of tree.

We briefly illustrated some possible applications of this work over a biological dataset. In particular, the thresholds provide a natural way to quantify the degree of support for states reconstructed as well as to determine what kind of evolutionary assumptions in terms of costs are necessary to a given reconstruction.

Keywords : maximum parsimony, ancestor state reconstruction, character mapping, parametric methods

1 Introduction

Testing hypothesis about evolutionary mechanisms like environment influence, homoplasy *etc.* calls for information not only about the contemporary organisms, which is - at least potentially - available, but also about the ancestral ones, which is by nature inaccessible, with few exceptions when related fossils can be found. The only possibility to achieve this kind of analysis is to infer the information about the ancestral organisms [14, 8, 3]. This inference is sometimes called character mapping or ancestral-state optimization or reconstruction. In short, the general purpose is to puzzle out the evolution history of a character from its contemporary states. Besides its relevance to questions about theoretical aspects of evolution, a practical interest of character mapping is it provides a natural way to transfer what is known about some organisms to another ones (ancestral or contemporary) using phylogenetic information.

Basically the problem of character-state reconstruction can be stated as follows. The evolutionary history of a set of organisms is assumed known and represented as a

rooted phylogenetic tree where contemporary organisms are leaves and ancestors ones are internal nodes. We consider a given character for which the states are known only over some of the nodes (organisms) of the phylogenetic tree. The present work deals only with binary characters: typically the presence/absence of a given feature. The question is to infer the character states of nodes for which this information is missing. Note that we consider the problem in a general way and make no assumption over which nodes the character states are known or unknown: they can be indifferently either leaves (contemporary organisms) or internal nodes (ancestors).

The generalized maximum parsimonious reconstruction and is variants like Dollo parsimony are certainly the most common methods used to solve the question of character-state inference [14, 8, 3]. Parsimony methods are easy and fast to compute by dynamics programming [12, 13] and several implementations are available [7]. A predictable issue, which was pointed out in [2, 3, 11], is that the results obtained by these methods depend heavily on the cost parameters chosen by the user (the usual choice is to consider equal costs for gain and loss, with no real biological motivation – just because it sounds like a neutral choice) or on method's assumptions, somehow hiding the parameters, like in Dollo parsimony (see Section 5.2). This dependency was systematically studied in an empirical way in [11]. On the other hand, parametric approaches have been developed to overcome the difficulty in choosing parameters used in optimization methods. They was successfully applied to alignment of sequences [15, 6] or statistical models [9, 10]. The main idea is to not just return the result (*i.e.* the alignment with the greatest score in [15, 6] or the maximum parsimonious reconstruction here) optimal with respect to an user-defined instance of parameters but to work on the whole set of the optimization process results one can obtain from any instance of parameters.

We apply here the parametric approach to character-state reconstruction by addressing essentially the same questions as [11]. Although we provide formal results which establish several properties of the relationships between cost parameters and the resulting reconstructions. In particular we show the existence of two thresholds λ_n^1 and λ_n^0, confounded in most cases, associated to all nodes n of the tree which are such that the state of node n can be reconstructed by state "present" (*resp.* "absent") only if the ratio "loss cost"/"gain cost" is smaller than λ_n^1 (*resp.* greater than λ_n^0). These properties are used to develop an efficient algorithm which computes these thresholds in a time quadratic with the size of the phylogenetic tree, by using a linear memory space.

The rest of the paper is organized as follows. We introduce basic concepts and notations in the next section. The formal results about relations between the parameters and the maximum parsimonious reconstructions are exposed in Section 3. An algorithm computing the "parametric maximum parsimonious reconstruction" is presented in Section 4. Finally, in Section 5, we show how the notions studied in this work are related to [11] and illustrate some possible uses of the parametric reconstruction over a dataset from [2].

A software implementing the algorithm presented here is freely available at `http://iml.univ-mrs.fr/~didier/recons`

2 Definitions and notations

Let T be a finite rooted tree and r its root node. As it will not lead to confusion, we still note T the set of nodes of T. For a node $n \in T$, $\mathcal{C}_n$ denotes the set of child nodes of n and T_n the subtree of T rooted at n. Let $E \subset T$ be a subset of nodes and s a function from E to $\{0, 1\}$. The nodes of E are said *known* relatively to s which is called the *initial function*. We consider the set of functions f from T to $\{0, 1\}$ extending s: the functions f such that $f(n) = s(n)$ for all nodes $n \in E$. Such functions are called

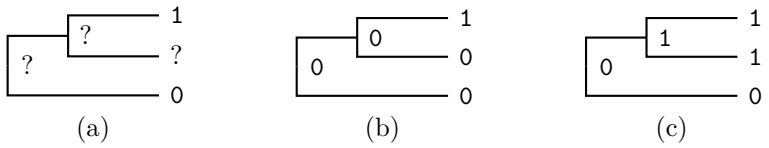

Figure 1: (a) a tree and an initial function ; (b) and (c) two alternative assignments with configuration $(0, 1)$ which has minimal γ-costs for all $\gamma \in [1, +\infty[$.

assignments of T relatively to s. In the following, the tree T and the function s are assumed fixed (all the variables defined depend on them but they do not appear in the notations). An assignment of a subtree T_n can be understood as an assignment of T_n relatively to the restriction of s to T_n or as the restriction of assignment of T to T_n.

With these notations the problem of missing data reconstruction over a tree can be stated: "from the given of a tree T and a function s, find the most relevant assignment of T relatively to s".

In the generalized parsimony framework, the relevance of an assignment is expressed in terms of cost. In practice for the binary case (character states are `0` or `1`), one defines a step matrix of the form:

$$c = \begin{pmatrix} 0 & \mathrm{c}_{01} \\ \mathrm{c}_{10} & 0 \end{pmatrix}$$

where the entries c_{01} and c_{10} are two positive real numbers representing the costs of transitions respectively from `0` to `1` and from `1` to `0`. Steadiness is assumed free of cost and we have $\mathrm{c}_{00} = \mathrm{c}_{11} = 0$. The cost of an assignment f of T is then the sum of all the costs of the ancestor/child transitions observed:

$$\sum_{n \in T} \sum_{m \in \mathcal{C}_n} \mathrm{c}_{\mathrm{f}(n)\mathrm{f}(m)}$$

The most parsimonious reconstruction on T relatively to s remains to find, from a given transition step matrix, an assignment f of T relatively to s with a minimal cost. Since the tree T is finite, so is the set of possible assignments of T whatever s and at least such an assignment exists.

Multiplying all the entries of the step matrix by a positive constant factor does not change the relative order of the costs of the assignments. Without loss of generality we will assume in the following $\mathrm{c}_{01} = 1$ and $\mathrm{c}_{10} = \gamma$: since we have $\mathrm{c}_{01} > 0$, we can always divide the transition costs by c_{01} in order to be in this case. In the following, the maximum parsimonious assignments depend on only one parameter: $\gamma = \frac{\mathrm{c}_{10}}{\mathrm{c}_{01}}$, the relative cost of a loss and a gain.

To a given assignment f of a subtree T_n we associate the pair $a = ({}^{10}a, {}^{01}a)$ where ${}^{10}a$ is the number of transitions "ancestor/child" from `1` to `0` (called the `10`-coordinate of a) and ${}^{01}a$ the number of transitions from `0` to `1` (called the `01`-coordinate of a) observed in T_n. This pair is designed as the *configuration* of f and is sufficient to compute the cost of f over T_n. For a parameter γ and a configuration a we define $\Delta_a(\gamma) = \gamma\,{}^{10}a + {}^{01}a$, which is naturally the cost of an assignment f with the configuration a using the parameter γ. Below we will say γ-cost of f or γ-cost of a. An assignment f and its configuration a are said with a minimal γ-cost if there is no assignment g such that the γ-cost of g is strictly smaller than $\Delta_a(\gamma)$. They are said with a minimal cost if they are with a minimal γ-cost for at least a positive real number γ. Remark that a configuration, even with a minimal cost, can correspond to more than one assignment (Figure 1).

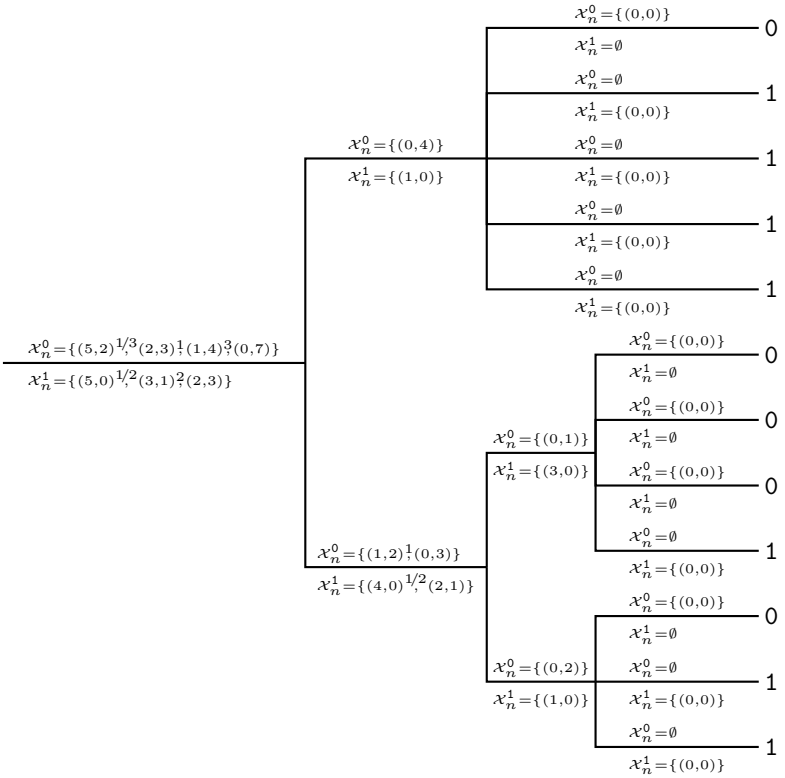

Figure 2: A tree T with a function s which is defined only on the leaves of T. The sets of configurations $\mathcal{X}_n^0$ and $\mathcal{X}_n^1$ are displayed around the branches bearing nodes n. Above the commas separating two configurations in $\mathcal{X}_n^0$ or $\mathcal{X}_n^1$ are displayed the values α_i^n and β_j^n.

For a node n of T and a real number $\gamma > 0$, $\mathcal{A}_{\gamma,n}^0$ (*resp.* $\mathcal{A}_{\gamma,n}^1$) denotes the set of assignments f of T_n with minimal γ-costs such that $\mathrm{f}(n) = \mathtt{0}$ (*resp.* $\mathrm{f}(n) = \mathtt{1}$). Notice that $\mathcal{A}_{\gamma,n}^0$ (*resp.* $\mathcal{A}_{\gamma,n}^1$) is empty when n is known with $\mathrm{s}(n) = \mathtt{1}$ (*resp.* $\mathrm{s}(n) = \mathtt{0}$). We note $\mathcal{X}_n^0$ (*resp.* $\mathcal{X}_n^1$) the set of the configurations of all the assignments in $\bigcup_{\gamma>0} \mathcal{A}_{\gamma,n}^0$ (*resp.* in $\bigcup_{\gamma>0} \mathcal{A}_{\gamma,n}^1$). The sets $\mathcal{X}_n^0$ and $\mathcal{X}_n^1$ are finite and an example is displayed in Figure 2. The set of configurations of the assignments of T_n with minimal costs (*i.e.* with no constraint on the image of n) is denoted $\mathcal{X}_n$.

If $\mathrm{s}(n) \neq \mathtt{1}$ (*resp.* $\mathrm{s}(n) \neq \mathtt{0}$) the minimal γ-cost which can be reached by an assignment f of T_n such that $\mathrm{f}(n) = \mathtt{0}$ (*resp.* $\mathrm{f}(n) = \mathtt{1}$) is unique and noted $\Gamma_n^0(\gamma)$ (*resp.* $\Gamma_n^1(\gamma)$), otherwise there is no assignment f of T_n with $\mathrm{f}(n) = \mathtt{0}$ (*resp.* $\mathrm{f}(n) = \mathtt{1}$) and $\Gamma_n^0(\gamma)$ (*resp.* $\Gamma_n^1(\gamma)$) is said not defined and set by convention to $+\infty$. A minimal γ-cost can correspond to more than one assignment (and possibly to more than one configuration). We note $\Gamma_n(\gamma)$ the minimal γ-cost of an assignment of T_n with no constraint on the image of n: $\Gamma_n(\gamma) = \min\{\Gamma_n^0(\gamma), \Gamma_n^1(\gamma)\}$. The functions Γ_r^0 and Γ_r^1 for the example depicted in Figure 2 are plotted in Figure 3.

3 Preliminary results

Our main interest in this section is to study how the configurations of the assignments with minimal γ-costs evolve with γ. The following remark comes with the positivity of the cost parameters.

Remark 1 *Let n be a node of T, f, g two assignments of T_n, a and b their respective configurations. If $\mathrm{f} \in \mathcal{A}^0_{\gamma,n}$ and $\mathrm{g} \in \mathcal{A}^0_{\gamma',n}$ for two positive real numbers γ and γ' then*

$$^{01}a \geq {}^{01}b \iff {}^{10}a \leq {}^{10}b$$

This remark still holds if $\mathrm{f} \in \mathcal{A}^1_{\gamma,n}$ and $\mathrm{g} \in \mathcal{A}^1_{\gamma',n}$. It is still true when f and g are two configurations with respectively a minimal γ-costs and a minimal γ'-cost.

Two different configurations of T_n with minimal costs (*resp.* belonging to $\mathcal{X}^0_n$, *resp.* belonging to $\mathcal{X}^1_n$) cannot have the same `01`- or `10`-coordinates. It follows the number of elements in $\mathcal{X}_n$ is basically bounded by the maximal number of gains (or losses) we can observe over T_n, which is itself strictly smaller of the number nodes of T_n.

Remark 2 *Let n be a node of T. The number of elements in $\mathcal{X}_n$ (resp. in $\mathcal{X}^0_n$, resp. in $\mathcal{X}^1_n$) is strictly smaller than the number of nodes in T_n.*

We assume in the following that the configurations of $\mathcal{X}^0_n = \{a_1, a_2, \ldots, a_k\}$ are indiced with respect to their (strictly) decreasing `10`-coordinates, that is ${}^{10}a_i > {}^{10}a_j$ for all $1 \leq i < j \leq n$. From Remark 1 we then have ${}^{01}a_i < {}^{01}a_j$ for all $1 \leq i < j \leq k$. We define the sequence of positive real numbers $(\alpha^n_i)_{1 \leq i < k}$ by:

$$\alpha^n_i = \frac{{}^{01}a_{i+1} - {}^{01}a_i}{{}^{10}a_i - {}^{10}a_{i+1}}$$

The value α^n_i is displayed over the comma separating the i^{th} and the $(i+1)^{\text{th}}$ configurations of $\mathcal{X}^0_n$ in Figure 2.

We make the same assumption about the way the configurations of $\mathcal{X}^1_n = \{b_1, b_2, \ldots, b_\ell\}$ are indiced and define $(\beta^n_i)_{1 \leq i < \ell}$ in the same manner as $(\alpha^n_i)_{1 \leq i < k}$. The values β^n_i are also displayed over the commas of the sets $\mathcal{X}^1_n$ in Figure 2.

Lemma 1 *Let n be a node of T, $\mathcal{X}^0_n = \{a_1, a_2, \ldots, a_k\}$ and $\mathcal{X}^1_n = \{b_1, b_2, \ldots, b_\ell\}$ the corresponding sets of optimal configurations indiced following the decreasing order of their `10`-coordinates. The sequences $(\alpha^n_i)_{1 \leq i < k}$ and $(\beta^n_i)_{1 \leq i < \ell}$ are increasing.*

Proof: We prove the increasingness only for $(\alpha^n_i)_{1 \leq i < k}$. Assume that there exists an integer $1 \leq i < k$ such that $\alpha^n_{i+1} < \alpha^n_i$. From the definition of the sequence $(\alpha^n_i)_{1 \leq i < k}$, we have $\gamma \geq \alpha^n_{i+1}$ if and only if $\Delta_{a_{i+1}}(\gamma) \leq \Delta_{a_{i+2}}(\gamma)$ and $\gamma \geq \alpha^n_i$ if and only if $\Delta_{a_i}(\gamma) \leq \Delta_{a_{i+1}}(\gamma)$. It follows that:

1. if $0 < \gamma \leq \alpha^n_{i+1}$ then $\Delta_{a_i}(\gamma) < \Delta_{a_{i+1}}(\gamma) \leq \Delta_{a_{i+2}}(\gamma)$,
2. if $\alpha^n_{i+1} < \gamma \leq \alpha^n_i$ then $\Delta_{a_i}(\gamma) \leq \Delta_{a_{i+1}}(\gamma)$ and $\Delta_{a_{i+2}}(\gamma) < \Delta_{a_{i+1}}(\gamma)$,
3. if $\gamma > \alpha^n_i$ then $\Delta_{a_i}(\gamma) > \Delta_{a_{i+1}}(\gamma) > \Delta_{a_{i+2}}(\gamma)$.

In all the cases, which altogether cover the whole set of positive real values, we observe that $\Delta_{a_{i+1}}(\gamma)$ is strictly greater than $\Delta_{a_i}(\gamma)$ or $\Delta_{a_{i+2}}(\gamma)$. In other words, the configuration a_{i+1} cannot have a minimal γ-cost for any parameter $\gamma > 0$. This contradicts the fact that a_{i+1} belongs to $\mathcal{X}^0_n$. □

Theorem 1 *For all nodes n of T, the functions Γ^0_n, Γ^1_n and Γ_n are either continuous piecewise affine or not defined.*

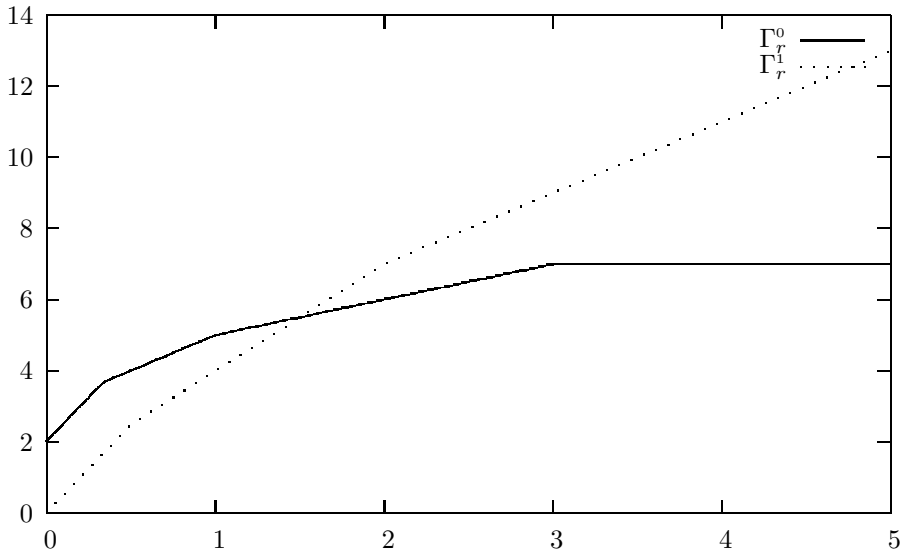

Figure 3: Representation of functions Γ_r^0 and Γ_r^1 for the example depicted in Figure 2 (r is the root node).

Proof: Let us assume $\mathrm{s}(n) \neq \mathtt{1}$ (*i.e.* $\mathrm{s}(n) = \mathtt{0}$ or n is unknown) thus Γ_n^0 is defined. It comes from the definition of $\mathcal{X}_n^0$ that for all $\gamma > 0$ there is a configuration $a \in \mathcal{X}_n^0$ such that $\Gamma_n^0(\gamma) = \Delta_a(\gamma)$ and reciprocally. The cost $\Delta_a(\gamma)$ of a configuration a is an affine function of γ. Let $\mathcal{X}_n^0 = \{a_1, a_2, \ldots, a_k\}$. The sequence $(\alpha_i^n)_{1 \leq i < k}$ is defined in such a way that for all $1 \leq i < k$, $\gamma \geq \alpha_i^n$ if and only if $\Delta_{a_i}(\gamma) \leq \Delta_{a_{i+1}}(\gamma)$. Since, moreover, Lemma 1 ensures that $(\alpha_i^n)_{1 \leq i < k}$ is increasing, Γ_n^0 can be alternatively defined by:

- for $0 < \gamma \leq \alpha_1^n$, $\Gamma_n^0(\gamma) = \Delta_{a_1}(\gamma)$,
- for $1 \leq i < k-1$ and $\alpha_i^n \leq \gamma \leq \alpha_{i+1}^n$, $\Gamma_n^0(\gamma) = \Delta_{a_i}(\gamma)$,
- for $\gamma \geq \alpha_{k-1}^n$, $\Gamma_n^0(\gamma) = \Delta_{a_k}(\gamma)$.

Remark that for all $1 \leq i < k-1$, we have $\Delta_{a_i}(\alpha_i^n) = \Delta_{a_{i+1}}(\alpha_i^n)$. This alternative definition expresses Γ_n^0 as a continuous piecewise affine function. All the preceding considerations hold for Γ_n^1 and lead to the same conclusion. At least Γ_n^0 or Γ_n^1 is defined and otherwise set to $+\infty$. Finally, $\Gamma_n = \min\{\Gamma_n^0, \Gamma_n^1\}$ is also continuous piecewise affine. The functions Γ_r^0 and Γ_r^1 for the example of Figure 2 are plotted in Figure 3. □

The following remark shows why considering the assignments of $\mathcal{A}_{\gamma,n}^0$ and $\mathcal{A}_{\gamma,n}^1$ can be useful.

Remark 3 *Let n be a node of T, m a node of T_n, γ a positive real number and* f *an assignment of T_n with a minimal γ-cost. If* $\mathrm{f}(m) = \mathtt{0}$ *(resp.* $\mathrm{f}(m) = \mathtt{1}$*) then the restriction of* f *to T_m is an assignment of $\mathcal{A}_{\gamma,m}^0$ (resp. of $\mathcal{A}_{\gamma,m}^1$).*

Again, this remark still holds if we replace "f an assignment of T_n with a minimal γ-cost" by "$\mathrm{f} \in \mathcal{A}_{\gamma,n}^0$" or by "$\mathrm{f} \in \mathcal{A}_{\gamma,n}^1$" in its statement.

Lemma 2 *Let n be a node of T, δ and ζ two positive real numbers such that $\zeta > \delta$. If there exist two assignments of T_n,* f *and* g *such that* $\mathrm{f} \in \mathcal{A}_{\gamma,n}^0$ *and* $\mathrm{g} \in \mathcal{A}_{\gamma,n}^1$ *for*

all $\gamma \in]\delta, \zeta[$*, then their respective configurations* a *and* b *are such that* ${}^{10}a \leq {}^{10}b$ *and* ${}^{01}a \geq {}^{01}b$*. Moreover we have* $a = b$ *if and only if* $a = b = (0, 0)$*.*

Proof: This lemma concerns only nodes n which are unknown: otherwise there cannot exist simultaneously two configurations $\mathrm{f} \in \mathcal{A}^0_{\gamma,n}$ and $\mathrm{g} \in \mathcal{A}^1_{\gamma,n}$. We proceed by induction over the number of nodes of the subtrees. The property is basically true when T_n contains only one node: it is enough to remark that in this case, n is an unknown leaf and $a = b = (0, 0)$, whatever δ and ζ. Let us assume the lemma is true for all subtrees with up to k nodes and T_n contains $k + 1$ nodes.

For all $m \in \mathcal{C}_n$ we note f_m (*resp.* g_m) the restriction of f (*resp.* of g) to T_m and a_m (*resp.* b_m) the corresponding configuration. If $\mathrm{f}(m) = \mathtt{0}$ (*resp.* $\mathrm{f}(m) = \mathtt{1}$) then $\mathrm{f}_m \in \mathcal{A}^0_{\gamma,m}$ (*resp.* $\mathrm{f}_m \in \mathcal{A}^1_{\gamma,m}$) for all $\gamma \in]\delta, \zeta[$ (Remark 3). In the same way, we have $\mathrm{g}_m \in \mathcal{A}^0_{\gamma,m}$ or $\mathrm{g}_m \in \mathcal{A}^1_{\gamma,m}$ depending on whether $\mathrm{g}(m) = \mathtt{0}$ or $\mathrm{g}(m) = \mathtt{1}$.

If $\mathrm{f}(m) = \mathrm{g}(m) = \mathtt{0}$, then f_m and g_m are both in $\mathcal{A}^0_{\gamma,m}$ for all $\gamma \in]\delta, \zeta[$. It implies that $\Delta_{a_m}(\gamma) = \Delta_{b_m}(\gamma) = \Gamma^0_m(\gamma)$ again for all $\gamma \in]\delta, \zeta[$. Since $\zeta > \delta$, it comes that $a_m = b_m$. The same arguments prove that $\mathrm{f}(m) = \mathrm{g}(m) = \mathtt{1}$ implies $a_m = b_m$.

For $\gamma \in]\delta, \zeta[$, the γ-costs of a and b can be written:

$$\begin{aligned}
\Delta_a(\gamma) &= \sum_{\substack{m \in \mathcal{C}_n \\ \mathrm{f}(m)=0}} \Delta_{a_m}(\gamma) + \sum_{\substack{m \in \mathcal{C}_n \\ \mathrm{f}(m)=1}} (\Delta_{a_m}(\gamma) + 1), \\
\Delta_b(\gamma) &= \sum_{\substack{m \in \mathcal{C}_n \\ \mathrm{g}(m)=0}} (\Delta_{b_m}(\gamma) + \gamma) + \sum_{\substack{m \in \mathcal{C}_n \\ \mathrm{g}(m)=1}} \Delta_{b_m}(\gamma).
\end{aligned}$$

The expression of $\Delta_a(\gamma)$ shows that if $\mathrm{f}(m) = 1$ then $\Delta_{a_m}(\gamma)$ is strictly smaller than the γ-cost of any configuration h of T_m such that $\mathrm{h}(m) = 0$ (otherwise there is an assignment in $\mathcal{A}^0_{\gamma,n}$ with a strictly smaller γ-cost than f). It follows that for all $m \in \mathcal{C}_n$, if $\mathrm{f}(m) = \mathtt{1}$ then $\mathrm{g}(m) = \mathtt{1}$.

We distinguish two kinds of child nodes:

- children m such that $\mathrm{f}(m) = \mathrm{g}(m)$ (*Kind 1*). We have shown that it implies $a_m = b_m$,
- children m such that $\mathrm{f}(m) = \mathtt{0}$, $\mathrm{g}(m) = \mathtt{1}$ (*Kind 2*). All the subtrees T_m have less than k nodes and we can use the induction assumption to get ${}^{01}a_m \geq {}^{01}b_m$ and ${}^{10}a_m \leq {}^{10}b_m$.

The $\mathtt{01}$-coordinates of a and b can be decomposed as follows:

$$\begin{aligned}
{}^{01}a &= \sum_{\substack{m \in \mathcal{C}_n \\ \mathrm{f}(m)=0}} {}^{01}a_m + \sum_{\substack{m \in \mathcal{C}_n \\ \mathrm{f}(m)=1}} ({}^{01}a_m + 1), \\
{}^{01}b &= \sum_{\substack{m \in \mathcal{C}_n \\ \mathrm{g}(m)=0}} {}^{01}b_m + \sum_{\substack{m \in \mathcal{C}_n \\ \mathrm{g}(m)=1}} {}^{01}b_m.
\end{aligned}$$

Using these decompositions, we write the difference ${}^{01}a - {}^{01}b$ as a sum of three terms:

$${}^{01}a - {}^{01}b = \sum_{\substack{m \in \mathcal{C}_n \\ \mathrm{f}(m)=0 \\ \mathrm{g}(m)=0}} ({}^{01}a_m - {}^{01}b_m) + \sum_{\substack{m \in \mathcal{C}_n \\ \mathrm{f}(m)=1 \\ \mathrm{g}(m)=1}} ({}^{01}a_m + 1 - {}^{01}b_m) + \sum_{\substack{m \in \mathcal{C}_n \\ \mathrm{f}(m)=0 \\ \mathrm{g}(m)=1}} ({}^{01}a_m - {}^{01}b_m). \quad (1)$$

The preceding considerations ensures that these three terms are non-negative (the two first terms involve child nodes of Kind 1 and the third, child nodes of Kind 2) and allow us to conclude ${}^{01}a \geq {}^{01}b$. The inequality ${}^{10}a \leq {}^{10}b$ is proved in the same way.

It remains to prove that $a = b$ implies $a = b = (0, 0)$. Consider the three terms of the right side of Equation (1):

1. $\sum_{\substack{m\in\mathcal{C}_n\\ \mathrm{f}(m)=0\\ \mathrm{g}(m)=0}} ({}^{01}a_m - {}^{01}b_m)$ is always 0,
2. $\sum_{\substack{m\in\mathcal{C}_n\\ \mathrm{f}(m)=1\\ \mathrm{g}(m)=1}} ({}^{01}a_m + 1 - {}^{01}b_m)$ is positive unless there is no child m with $\mathrm{f}(m) = \mathtt{1}$ and $\mathrm{g}(m) = \mathtt{1}$,
3. $\sum_{\substack{m\in\mathcal{C}_n\\ \mathrm{f}(m)=0\\ \mathrm{g}(m)=1}} ({}^{10}a_m - {}^{10}b_m)$ is 0 only if ${}^{01}a_m = {}^{01}b_m$ for all children m with $\mathrm{f}(m) = \mathtt{0}$ and $\mathrm{g}(m) = \mathtt{1}$ (since such nodes m are children of Kind 2, we have ${}^{01}a_m \geq {}^{01}b_m$).

In summary, we have ${}^{01}a = {}^{01}b$ only if the two following assertions hold:

- there is no child node m such that $\mathrm{f}(m) = 1$ and $\mathrm{g}(m) = 1$,
- for all child nodes m such that $\mathrm{f}(m) = 0$ and $\mathrm{g}(m) = 1$ then ${}^{01}a_m = {}^{01}b_m$.

Let us write the difference ${}^{10}a - {}^{10}b$ in the same way as in Equation (1):

$$ {}^{10}a - {}^{10}b = \sum_{\substack{m\in\mathcal{C}_n\\ \mathrm{f}(m)=0\\ \mathrm{g}(m)=0}} ({}^{10}a_m - {}^{10}b_m - 1) + \sum_{\substack{m\in\mathcal{C}_n\\ \mathrm{f}(m)=1\\ \mathrm{g}(m)=1}} ({}^{10}a_m - {}^{10}b_m) + \sum_{\substack{m\in\mathcal{C}_n\\ \mathrm{f}(m)=0\\ \mathrm{g}(m)=1}} ({}^{10}a_m - {}^{10}b_m) \quad (2) $$

Symmetrically, from Equation (2), we have ${}^{10}a = {}^{10}b$ only if the two following assertions hold:

- there is no child node m such that $\mathrm{f}(m) = 0$ and $\mathrm{g}(m) = 0$,
- for all child nodes m such that $\mathrm{f}(m) = 0$ and $\mathrm{g}(m) = 1$ then ${}^{10}a_m = {}^{10}b_m$.

Altogether we have $a = b$ only if all child nodes m are such that $\mathrm{f}(m) = 0$, $\mathrm{g}(m) = 1$ and $a_m = b_m$. The induction assumption gives $a_m = b_m = (0, 0)$ for all child nodes m which implies $a = b = (0, 0)$. □

The results of Lemma 2 do not hold if the assignments and configurations considered have a minimal γ-cost for a single value of parameter (*i.e.* not over an interval of positive length as assumed in the lemma). A (counter-)example is displayed in Figure 4. The configurations $(2, 3)$, $(1, 4)$ and $(0, 5)$ stand for assignments of $\mathcal{A}^0_{1,r}$ while $(3, 2)$, $(2, 3)$ and $(1, 4)$ stand for assignments of $\mathcal{A}^1_{1,r}$. In particular we have a configuration $a = (2, 3)$ of an assignment in $\mathcal{A}^0_{1,r}$ and a configuration $b = (1, 4)$ of an assignment in $\mathcal{A}^0_{1,r}$ such that ${}^{10}a > {}^{10}b$ and ${}^{01}a < {}^{01}b$. Moreover, the configuration $(1, 4)$ (as well as $(2, 3)$) corresponds both to an assignment of $\mathcal{A}^0_{1,r}$ and an assignment of $\mathcal{A}^1_{1,r}$, without being equal to $(0, 0)$.

Lemma 3 *Let n be an unknown node of T. The function $(\Gamma^1_n - \Gamma^0_n)$ is increasing.*

Proof: Since n is unknown, both Γ^0_n and Γ^1_n are defined. Theorem 1 ensures Γ^0_n and Γ^1_n are continuous piecewise affine. The same holds for $(\Gamma^1_n - \Gamma^0_n)$. For all intervals I over which both Γ^1_n and Γ^0_n are simply affine, there are a configuration a of an assignment of $\mathcal{A}^0_{\gamma,n}$ and a configuration b of an assignment of $\mathcal{A}^1_{\gamma,n}$ such that $(\Gamma^1_n - \Gamma^0_n)(\gamma) = \Delta_b(\gamma) - \Delta_a(\gamma)$ for all $\gamma \in I$. From Lemma 2 we have ${}^{10}a \leq {}^{10}b$ and $(\Gamma^1_n - \Gamma^0_n)$ is increasing over I. Moreover $(\Gamma^1_n - \Gamma^0_n)$ is continuous over the set of positive real numbers which ends the proof. □

Theorem 2 *Let n be a node of T. One of the following assertions holds:*

1. *For all positive real numbers γ, $\Gamma^0_n(\gamma) < \Gamma^1_n(\gamma)$.*
2. *For all positive real numbers γ, $\Gamma^0_n(\gamma) > \Gamma^1_n(\gamma)$.*
3. *For all positive real numbers γ, $\Gamma^0_n(\gamma) = \Gamma^1_n(\gamma) = 0$.*
4. *There exists a unique positive real number λ such that $\Gamma^0_n(\gamma) \leq \Gamma^1_n(\gamma)$ if and only if $\gamma \geq \lambda$.*

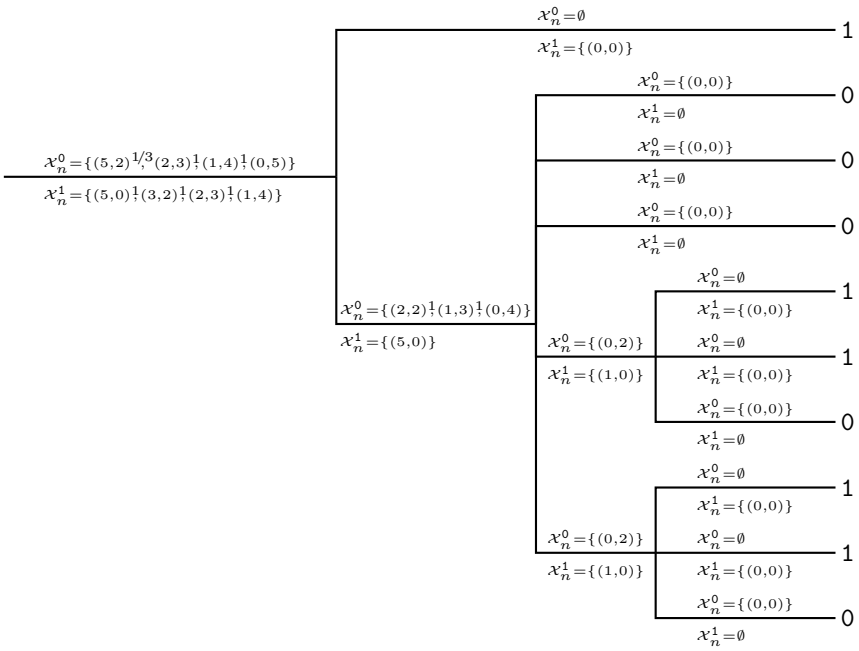

Figure 4: A tree and an initial function defined on leaves for which many configurations and assignments have a minimal 1-cost.

Proof: All these assertions mutually exclude themselves. The first two ones are verified in particular when Γ_n^0 or Γ_n^1 is not defined. Assume that neither Assertion 1 nor Assertion 2 holds. It means Γ_n^0 and Γ_n^1 are defined and there are two positive real numbers θ and θ' with $\theta \neq \theta'$ such that $\Gamma_n^0(\theta) \leq \Gamma_n^1(\theta)$ and $\Gamma_n^0(\theta') \geq \Gamma_n^1(\theta')$. We distinguish here two possibilities depending upon there are more than one positive real number γ such that $\Gamma_n^0(\gamma) = \Gamma_n^1(\gamma)$.

Assume first there are two positive real numbers θ and θ' with $\theta < \theta'$, $\Gamma_n^0(\theta) = \Gamma_n^1(\theta)$ and $\Gamma_n^0(\theta') = \Gamma_n^1(\theta')$. From the increasingness of $(\Gamma_n^1 - \Gamma_n^0)$ (Lemma 3), we have for all $\gamma \in [\theta, \theta']$, $\Gamma_n^0(\gamma) = \Gamma_n^1(\gamma)$. Since $\theta < \theta'$, there are two numbers δ and ζ in $[\theta, \theta']$ with $\delta < \zeta$ and such that the interval $]\delta, \zeta[$ is small enough that the conditions of Lemma 2 are granted: *i.e.* there are two assignments with minimal γ-costs for all $\gamma \in]\delta, \zeta[$, differing in their images of n. Let a and b their respective configurations. We have $\Gamma_n^0(\gamma) = \Delta_a(\gamma)$, $\Gamma_n^1(\gamma) = \Delta_b(\gamma)$ and $\Delta_a(\gamma) = \Delta_b(\gamma)$ for all $\gamma \in]\delta, \zeta[$. It implies $a = b$ and Lemma 2 ensures the configurations of these assignments are both $(0, 0)$. Whatever the value of the cost parameter, there cannot be any assignment with a smaller cost. For all positive real numbers γ we have $\Gamma_n^0(\gamma) = \Gamma_n^1(\gamma) = \Delta_{(0,0)}(\gamma) = 0$ and Assertion 3 holds.

It remains the case where there is a unique real number γ such that $\Gamma_n^0(\gamma) = \Gamma_n^1(\gamma)$ (Assertions 1 and 2 do not hold). Two respective direct consequences are $(\Gamma_n^1 - \Gamma_n^0)$ increases strictly in a neighborhood of γ and there are two different positive real numbers θ and θ' with $0 < \theta < \gamma < \theta'$ such that $\Gamma_n^0(\theta) < \Gamma_n^1(\theta)$ and $\Gamma_n^0(\theta') > \Gamma_n^1(\theta')$. The continuity and the increasingness of $(\Gamma_n^1 - \Gamma_n^0)$ finish to prove Assertion 4. In particular the functions plotted in Figure 3 satisfy Assertion 4. □

Theorem 3 *Let m be a node of T different from the root node and n its direct ancestor. One of the following assertions holds:*

1. *there is no positive real number* γ *such that there exists an assignment* $\mathrm{f} \in \mathcal{A}^{0}_{\gamma,n}$ *with* $\mathrm{f}(m) = \mathtt{1}$ *;*
2. *for all positive real numbers* γ*, there exists an assignment* $\mathrm{f} \in \mathcal{A}^{0}_{\gamma,n}$ *with* $\mathrm{f}(m) = \mathtt{1}$ *;*
3. *there is a positive real number* μ^{0}_{m} *such that there exists an assignment* $\mathrm{f} \in \mathcal{A}^{0}_{\gamma,n}$ *with* $\mathrm{f}(m) = \mathtt{1}$ *if* $\gamma \leq \mu^{0}_{m}$ *and no such assignment otherwise.*

Symmetrically, one of the following assertions holds:

1. *there is no positive real number* γ *such that there exists an assignment* $\mathrm{f} \in \mathcal{A}^{0}_{\gamma,n}$ *with* $\mathrm{f}(m) = \mathtt{1}$ *;*
2. *for all positive real numbers* γ*, there exists an assignment* $\mathrm{f} \in \mathcal{A}^{1}_{\gamma,n}$ *with* $\mathrm{f}(m) = \mathtt{1}$ *;*
3. *there is a positive real number* μ^{1}_{m} *such that there exists an assignment* $\mathrm{f} \in \mathcal{A}^{1}_{\gamma,n}$ *with* $\mathrm{f}(m) = \mathtt{1}$ *if* $\gamma \leq \mu^{1}_{m}$ *and no such assignment otherwise.*

Proof: We need two additional notations here and first design by $\Phi_{\mathrm{f},n}(\gamma)$ the γ-cost of an assignment f of T_n. The γ-cost of the restriction of f to T_m where m is a node of T_n is still noted $\Phi_{\mathrm{f},m}(\gamma)$. This cost $\Phi_{\mathrm{f},n}(\gamma)$ can be split into two terms in such a way that the first one, noted $\Psi_{\mathrm{f},m}(\gamma)$, depends on the value of $\mathrm{f}(m)$ and the second one $\Phi_{\mathrm{f},n}(\gamma) - \Psi_{\mathrm{f},m}(\gamma)$ does not. Formally $\Psi_{\mathrm{f},m}(\gamma)$ is defined by:

$$\Psi_{\mathrm{f},m}(\gamma) = \begin{cases} \Phi_{\mathrm{f},m}(\gamma) & \text{if } \mathrm{f}(p), \mathrm{f}(m) = \mathtt{0},\mathtt{0} \\ \Phi_{\mathrm{f},m}(\gamma) + \gamma & \text{if } \mathrm{f}(p), \mathrm{f}(m) = \mathtt{1},\mathtt{0} \\ \Phi_{\mathrm{f},m}(\gamma) + 1 & \text{if } \mathrm{f}(p), \mathrm{f}(m) = \mathtt{0},\mathtt{1} \\ \Phi_{\mathrm{f},m}(\gamma) & \text{if } \mathrm{f}(p), \mathrm{f}(m) = \mathtt{1},\mathtt{1} \end{cases}$$

With Remark 3, if f is an assignment with a minimal γ-cost, we have $\Phi_{\mathrm{f},n}(\gamma) = \Gamma_n(\gamma)$ and:

$$\Psi_{\mathrm{f},m}(\gamma) = \begin{cases} \Gamma^{0}_{m}(\gamma) & \text{if } \mathrm{f}(p), \mathrm{f}(m) = \mathtt{0},\mathtt{0} \\ \Gamma^{0}_{m}(\gamma) + \gamma & \text{if } \mathrm{f}(p), \mathrm{f}(m) = \mathtt{1},\mathtt{0} \\ \Gamma^{1}_{m}(\gamma) + 1 & \text{if } \mathrm{f}(p), \mathrm{f}(m) = \mathtt{0},\mathtt{1} \\ \Gamma^{1}_{m}(\gamma) & \text{if } \mathrm{f}(p), \mathrm{f}(m) = \mathtt{1},\mathtt{1} \end{cases}$$

Let consider assignments of $\mathcal{A}^{0}_{\gamma,n}$ for a parameter γ. The point to prove is that if there exists an assignment $\mathrm{g} \in \mathcal{A}^{0}_{\theta,n}$ with $\mathrm{g}(m) = \mathtt{1}$ for a positive parameter θ then there exists an assignment $\mathrm{g}' \in \mathcal{A}^{0}_{\theta',n}$ with $\mathrm{g}'(m) = \mathtt{1}$ for all $\theta' \leq \theta$.

Let $\mathrm{g} \in \mathcal{A}^{0}_{\theta,n}$ with $\mathrm{g}(m) = \mathtt{1}$. We have $\Psi_{\mathrm{g},m}(\theta) = \Gamma^{1}_{m}(\theta) + 1$ and, since $\mathrm{g} \in \mathcal{A}^{0}_{\theta,n}$, $\Gamma^{1}_{m}(\theta) + 1 \leq \Gamma^{0}_{m}(\theta)$. Let us consider a parameter $\theta' \leq \theta$ and an assignment $\mathrm{h} \in \mathcal{A}^{0}_{\theta',n}$. We have $\Psi_{\mathrm{h},m}(\theta') = \Gamma^{1}_{m}(\theta') + 1$ or $\Psi_{\mathrm{h},m}(\theta') = \Gamma^{0}_{m}(\theta')$ depending on whether $\mathrm{h}(m) = \mathtt{1}$ or $\mathrm{h}(m) = \mathtt{0}$. The increasingness of $(\Gamma^{1}_{m} - \Gamma^{0}_{m})$ implies $\Gamma^{1}_{m}(\theta') + 1 \leq \Gamma^{0}_{m}(\theta')$. Let us define the assignment g' in the following way:

- $\mathrm{g}'(p) = \mathrm{h}(p)$ for all nodes $p \in T_{n \setminus T_m}$,
- the restriction of g' to T_m belongs to $\mathcal{A}^{\mathtt{1}}_{\theta',m}$: $\mathrm{g}'(m) = \mathtt{1}$ and $\Psi_{\mathrm{g}',m}(\theta') = \Gamma^{1}_{m}(\theta') + 1$.

Under this definition, we have $\Phi_{\mathrm{h},n}(\gamma) - \Psi_{\mathrm{h},m}(\gamma') = \Phi_{\mathrm{g}',n}(\gamma) - \Psi_{\mathrm{g}',m}(\gamma')$ and $\Psi_{\mathrm{g}',m}(\gamma') \leq \Psi_{\mathrm{h},m}(\theta')$. The conjunction ensures $\mathrm{g}' \in \mathcal{A}^{0}_{\theta',n}$. The case of assignments of $\mathcal{A}^{1}_{\gamma,n}$ is proved in the same way. □

In the following, the cases of Assertions 1 and 2 will be treated as degenerated cases of Assertion 3 where μ^{0}_{m} (*resp.* μ^{1}_{m}) is 0 or $+\infty$. The values μ^{0}_{m} and μ^{1}_{m} will be referred to as the *conditional thresholds* of m (Figure 5).

Basic considerations show we have $\mu^{0}_{m} \leq \mu^{1}_{m}$ for all nodes m different from the root node of T.

Theorem 4 *Let n be a node of T. One of the following assertions holds:*

1. *there is no positive real number γ such that there exists an assignment* f *of T with a minimal γ-cost and* $\mathrm{f}(n) = \mathtt{1}$ *;*
2. *for all positive real numbers γ, there exists an assignment* f *of T with a minimal γ-cost and* $\mathrm{f}(n) = \mathtt{1}$ *;*
3. *there is a positive real number $\lambda_n^{\mathtt{1}}$ such that there exists an assignment* f *of T with a minimal γ-cost and* $\mathrm{f}(n) = \mathtt{1}$ *if* $\gamma \leq \lambda_n^{\mathtt{1}}$ *and no such assignment if* $\gamma > \lambda_n^{\mathtt{1}}$.

The proof of this theorem will be given in Section 4.2 since it is also useful to explain a part of the algorithm. The following corollary comes with symmetry. The 1-thresholds and the conditional thresholds for the tree and the initial function of Figure 2 are displayed in Figure 5.

Corollary 1 *Let n be a node of T. One of the following assertions holds:*

1. *there is no positive real number γ such that there exists an assignment* f *of T with a minimal γ-cost and* $\mathrm{f}(n) = \mathtt{0}$ *;*
2. *for all positive real numbers γ, there exists an assignment* f *of T with a minimal γ-cost and* $\mathrm{f}(n) = \mathtt{0}$ *;*
3. *there is a positive real number $\lambda_n^{\mathtt{0}}$ such that there exists an assignment* f *of T with a minimal γ-cost and* $\mathrm{f}(n) = \mathtt{0}$ *if* $\gamma \geq \lambda_n^{\mathtt{0}}$ *and no such assignment if* $\gamma < \lambda_n^{\mathtt{0}}$.

Again, both in Theorem 4 and Corollary 1, the cases of Assertions 1 and 2 will be considered as degenerated cases of Assertion 3 where $\lambda_n^{\mathtt{1}}$ is 0 or $+\infty$ and $\lambda_n^{\mathtt{0}}$ is $+\infty$ or 0, respectively. The values $\lambda_n^{\mathtt{1}}$ and $\lambda_n^{\mathtt{0}}$ are called 1-*threshold* and 0-*threshold* of n. Though we have $\lambda_n^{\mathtt{1}} = \lambda_n^{\mathtt{0}}$ in most cases, it is not true in general. For instance, if we consider n, the internal (not root) node of Figure 1, we have $\lambda_n^{\mathtt{1}} = 0$ and $\lambda_n^{\mathtt{0}} = 1$. However since $\lambda_n^{\mathtt{1}} = \lambda_n^{\mathtt{0}}$ is verified in all the following examples (Figures 6 and 7), we display a single value per node.

4 Algorithm

In this section we present a way to compute for each node n, the 1-threshold $\lambda_n^{\mathtt{1}}$: the value such that there exists an assignment f of T with a minimal γ-cost and $\mathrm{f}(n) = \mathtt{1}$ if $0 < \gamma \leq \lambda_n^{\mathtt{1}}$ and no such assignment otherwise ($\lambda_n^{\mathtt{1}}$ taking values among the non-negative real numbers and $+\infty$). We don't write how to compute the 0-thresholds since it is done in a symmetric way.

The algorithm follows the same general outline as the Sankoff dynamic programming algorithm for the classic maximum parsimonious ancestral reconstruction [12, 13]. It involves two stages. The first one consists in computing recursively the sets $\mathcal{X}_n^{\mathtt{0}}$ and $\mathcal{X}_n^{\mathtt{1}}$ and the thresholds $\mu_m^{\mathtt{0}}$ and $\mu_m^{\mathtt{1}}$ for all nodes n with children m of the trees from leaves to root node (Algorithm 1). After an intermediate step where we compute the 1-threshold of the root node, the second stage uses the conditional thresholds $\mu_m^{\mathtt{0}}$ and $\mu_m^{\mathtt{1}}$ to somehow broadcast 1-thresholds from root node to leaves (Algorithm 2).

4.1 Algorithm 1 - Pretreatment

The sets of configurations computed here do not match exactly the definitions given in Section 2. The difference is they do not necessarily contain all the "constrained" configurations with minimal costs : they contain only the configurations which have minimal γ-costs for more than a single value of parameter γ. Let say configurations having minimal costs with positive supports. By abuse of notation, they are still denoted $\mathcal{X}_n^{\mathtt{0}}$ and $\mathcal{X}_n^{\mathtt{1}}$. For instance the sets computed by the algorithm for the root node

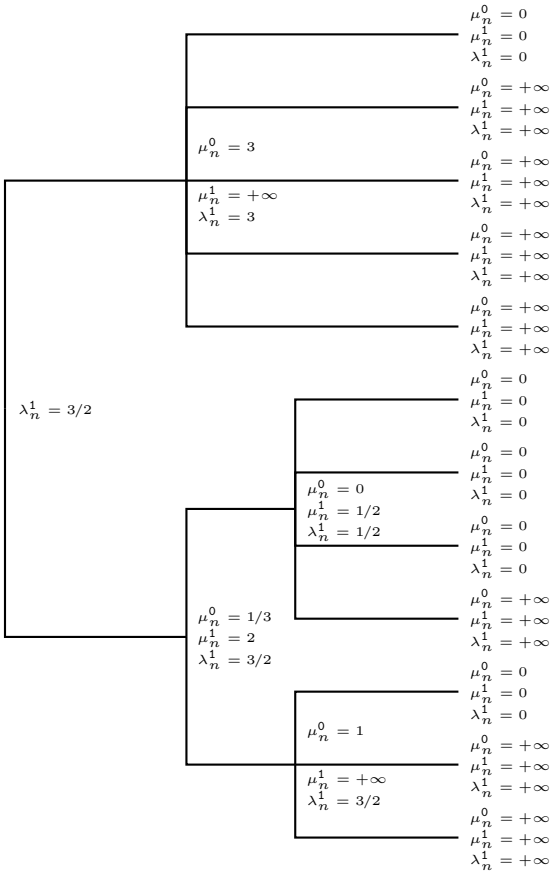

Figure 5: The 1-thresholds and the conditional thresholds for the tree and the initial function of Figure 2.

of the example of Figure 4 are $\mathcal{X}_r^0 = \{(5,2) \overset{\frac{1}{3}}{;} (2,3) \overset{1}{;} (0,5)\}$ and $\mathcal{X}_r^1 = \{(5,0) \overset{1}{;} (1,4)\}$ instead of $\mathcal{X}_r^0 = \{(5,2) \overset{\frac{1}{3}}{;} (2,3) \overset{1}{;} (1,4) \overset{1}{;} (0,5)\}$ and $\mathcal{X}_r^1 = \{(5,0) \overset{1}{;} (3,2) \overset{1}{;} (2,3) \overset{1}{;} (1,4)\}$.

We focus on computing $\mathcal{X}_n^0$ and μ_m^0 (the case of $\mathcal{X}_n^1$ and μ_m^1 being symmetrical). Except for the recursive calls computing the configurations with minimal costs of the children of n, the task of Algorithm 1 is to compute the set $\mathcal{X}_n^0$ from the sets $\mathcal{X}_m^0$ and $\mathcal{X}_m^1$ as well as the conditional thresholds μ_m^0 for all children m of the entry node n. The sets $\mathcal{X}_m^0$ and $\mathcal{X}_m^1$ are written respectively $\{a_1^m, a_2^m, \ldots, a_{k_m}^m\}$ and $\{b_1^m, b_2^m, \ldots, b_{\ell_m}^m\}$, these configurations being indiced with respect to their decreasing 10-coordinates as usual. Algorithm 1 benefits from Remark 3 which says that if an assignment of T_n belongs to $\mathcal{A}_{\gamma,n}^0$ then its restrictions to T_m for all children m of n belong to $\mathcal{A}_{\gamma,m}^0$ or $\mathcal{A}_{\gamma,m}^1$ depending on whether the images of m are 0 or 1. It follows that the configuration of an assignment in $\mathcal{A}_{\gamma,n}^0$ can be computed by finding the combination of images of its children which is such that the configurations of the corresponding assignments in $\mathcal{A}_{\gamma,m}^0$ or $\mathcal{A}_{\gamma,m}^1$ lead to minimize its γ-cost. To get all the configurations of $\mathcal{X}_n^0$, we somehow make run the parameter γ from 0 (excluded) to $+\infty$ while maintaining which configuration of $\mathcal{X}_m^0$ and which configuration of $\mathcal{X}_m^1$ are corresponding to assignments in $\mathcal{A}_{\gamma,m}^0$ and $\mathcal{A}_{\gamma,m}^1$ respectively. Since the minimal γ-costs are piecewise affine functions of γ, we have in fact to consider successive intervals of parameters over which both Γ_m^0 and Γ_m^1 are affine for all children nodes m, *i.e.* where the configurations of assignments in $\mathcal{A}_{\gamma,m}^0$ and $\mathcal{A}_{\gamma,m}^1$ doesn't change. The bounds of these intervals are given by the sequences $(\alpha_i^m)_{1 \leq i < k_m}$ and $(\beta_j^m)_{1 \leq j < \ell_m}$ defined in the beginning of Section 3. Another useful property is given by Theorem 3 which ensures that there are a threshold μ_m^0 for each child m, such that there is an assignment f in $\mathcal{A}_{\gamma,n}^0$ with $\mathrm{f}(m) = 1$ if $\gamma \leq \mu_m^0$ and no such assignment otherwise (this threshold is basic to compute if it belongs to an interval where configurations of assignments in $\mathcal{A}_{\gamma,m}^0$ and $\mathcal{A}_{\gamma,m}^1$ doesn't change). In addition to the fact these thresholds μ_m^0 has to be determined for further use, there cannot be an assignment in $\mathcal{A}_{\gamma,n}^0$ associating 1 to m over this threshold and we no longer have to take into account the elements of $\mathcal{X}_n^1$ in the next iterations. In practice, the variables i_m and j_m contain respectively the indices of the configurations of $\mathcal{X}_m^0$ and $\mathcal{X}_m^1$ corresponding to assignments in $\mathcal{A}_{\gamma,m}^0$ and $\mathcal{A}_{\gamma,m}^1$ for a parameter γ in the current interval, $\mathrm{h}(m)$ is equal to 1 if there is an assignment in $\mathcal{A}_{\gamma,n}^0$ associating 1 to n and to 0 otherwise.

The first current interval to consider has left bound 0 (excluded). We leave its right bound unknown and start by initializing all the variables to their values relatively to an infinitesimal positive value of the parameter (lines 3-13). Next, we are ready to start the iterations. The first stage (lines 15-18) of the main loop is for computing the right bound of the current interval, or equivalently the left bound of the next interval to consider, that is the greatest positive real number such that all the variables satisfy their definitions. It remains to find ω, the minimum, among all children m, of $\alpha_{i_m}^m$, $\beta_{j_m}^m$ (just noted α_m and β_m in Algorithm 1), beyond which a change over i_m or j_m is needed, and δ_m which becomes μ_m^0 if it is smaller than the two preceding bounds, meaning we have to change $\mathrm{h}(m)$. Remark that δ_m cannot be strictly smaller that the left bound of the current interval because we have ${}^{10}b_{j_m}^m \geq {}^{10}a_{i_m}^m$. The second stage of the main loop (lines 19-30) updates all the variables in order to make it satisfy their definitions with regard to the new left bound and compute the corresponding configuration of $\mathcal{X}_n^0$. Since a change over variable i_m has no effect over this configuration, we test it to avoid storing several times the same configuration (line 30). The iterations stop when no change is observed over the variables.

The computation of $\mathcal{X}_n^1$ and μ_m^1 for the children of n follows the same outline that above and is not displayed in Algorithm 1.

Before exiting the algorithm, we free the memory space used to store the sets of configurations of all the children (lines 34-35). These sets are only needed to compute $\mathcal{X}_n^0$, $\mathcal{X}_n^1$, μ_n^0 and μ_n^1. As we will see further, the rest of the computation needs, besides

the conditional thresholds of the nodes, only the sets of configurations of the root node which are not freed during the execution of Algorithm 1 since it has no parent.

4.2 Algorithm 2 - Computation of 1-thresholds

We explain this stage of the algorithm by giving a proof of Theorem 4, which states the existence of 1-thresholds for all nodes. To this end, we proceed by induction over the depths of the nodes and first show the existence of the 1-threshold λ_r^1 of the root node of T. When applied to the root node, Theorem 2 distinguishes 4 possibilities:

1. For all positive real numbers γ, $\Gamma_r^0(\gamma) < \Gamma_r^1(\gamma)$, which corresponds to $\lambda_r^1 = 0$.
2. For all positive real numbers γ, $\Gamma_r^0(\gamma) > \Gamma_r^1(\gamma)$, which corresponds to $\lambda_r^1 = +\infty$.
3. For all positive real numbers γ, $\Gamma_r^0(\gamma) = \Gamma_r^1(\gamma) = 0$, which corresponds to $\lambda_r^1 = +\infty$.
4. There exists a unique positive real number λ such that $\Gamma_n^0(\gamma) \leq \Gamma_n^1(\gamma)$ if and only if $\gamma \geq \lambda$, which corresponds to $\lambda_r^1 = \lambda$.

Let us assume the existence of 1-threshold λ_n^1 for a node n of T. In other words, for all $0 < \gamma \leq \lambda_n^1$ there is an assignment f of T with a minimal γ-cost and $\mathrm{f}(n) = 1$ and no such assignment otherwise. Consider now the conditional thresholds μ_m^0 and μ_m^1, defined in Theorem 3, of the children m of n. We have $\mu_m^0 \leq \mu_m^1$. This leaves the following possibilities:

- $\lambda_n^1 < \mu_m^0$: since an assignment f with a minimal γ-cost for $\gamma > \lambda_n^1$ is such that $\mathrm{f}(n) = 0$, Theorem 3 ensures that there is a assignment with a minimal γ-cost and $\mathrm{f}(m) = 0$ if and only if $\gamma \leq \mu_m^0$, *i. e.* $\lambda_m^1 = \mu_m^0$.
- $\mu_m^0 \leq \lambda_n^1 \leq \mu_m^1$: If $\gamma \leq \lambda_n^1$ then there is an assignment f with a minimal γ-cost and $\mathrm{f}(n) = 1$. Theorem 3 ensures that the threshold μ_m^1 applies and there is an assignment with a minimal γ-cost associating 1 to m. If $\gamma > \lambda_n^1$ then an assignment with a minimal γ-cost is such that $\mathrm{f}(n) = 0$ and we have to use the threshold μ_m^0 which ensures there is no assignment with minimal γ-cost and an image of m equal to 1. In short, we have $\lambda_m^1 = \lambda_n^1$.
- $\lambda_n^1 > \mu_m^1$: since there is an assignment f with a minimal γ-cost and $\mathrm{f}(n) = 1$ for γ up to λ_m^1, there is an assignment f with a minimal γ-cost and $\mathrm{f}(m) = 1$ up to μ_m^1 and no such assignment for $\mu_m^1 < \gamma \leq \lambda_m^1$.If $\gamma > \lambda_n^1$ then there is no assignment with a minimal γ-cost associating 1 to n. The threshold μ_m^0, which is here smaller than λ_n^1, applies for such γ and there is no assignment f with a minimal γ-cost and $\mathrm{f}(m) = 1$. Finally we have $\lambda_m^1 = \mu_m^1$.

The existence of a 1-threshold for the child m arises in all these possibilities, which ends our induction and the proof of Theorem 4.

Moreover, the preceding considerations give an explicit way of computing the 1-threshold λ_m^1 of a node m (different from the root node) from the values λ_n^1, the 1-threshold of its direct ancestor, μ_m^0 and μ_m^1, its conditional thresholds. Algorithm 2 follows this way to recursively compute the 1-thresholds of the whole tree from the 1-threshold of the root node.

4.3 Algorithm 3 - Main algorithm

The main algorithm just calls sequentially the two preceding ones with an intermediate step: the computation of λ_r^1, the 1-threshold of the root node. Since this computation remains essentially in finding the point (if it exists) where the two piecewise affine functions Γ_r^0 and Γ_r^1, defined from $\mathcal{X}_r^0$ and $\mathcal{X}_r^1$, intersect one another, we don't write the formal algorithm (very similar to the part of the main loop of Algorithm 1 which determines δ_m).

Algorithm 1: **compute_bottom_up**(n) : compute recursively $\mathcal{X}_n^0$ and $\mathcal{X}_n^1$ as well as μ_m^0 and μ_m^1 for all children m of n

for all $m \in \mathcal{C}_n$ **do** /* Recursive calls over the children of n */
1 **compute_bottom_up**(m);
2 **if** s$(n) = 1$ **then** $\mathcal{X}_n^0 \leftarrow \emptyset$ **else** /* Compute $\mathcal{X}_n^0$ and μ_m^0 for all $m \in \mathcal{C}_n$ */
3 $k_n \leftarrow 1$; $a_1^n \leftarrow (0,0)$ /* Initializations */
4 **for all** $m \in \mathcal{C}_n$ **do**
5 $i_m \leftarrow 1$; $j_m \leftarrow 1$
6 **if** $\ell_m \geq 1$ **and** ($k_m = 0$ **or** ${}^{01}a_1^m \geq {}^{01}b_1^m + 1$) **then**
7 h$(m) \leftarrow 1$; $\mu_m^0 \leftarrow +\infty$
8 **else**
9 h$(m) \leftarrow 0$; $\mu_m^0 \leftarrow 0$; $j_m \leftarrow \ell_m + 1$
10 **if** h$(m) = 1$ **then**
11 $a_1^n \leftarrow ({}^{10}a_1^n + {}^{10}b_1^m, {}^{01}a_1^n + {}^{01}b_1^m + 1)$
12 **else**
13 $a_1^n \leftarrow ({}^{10}a_1^n + {}^{10}a_1^m, {}^{01}a_1^n + {}^{01}a_1^m)$
14 **repeat** /* Main loop */
15 **for all** $m \in \mathcal{C}_n$ **do** /* Compute upper bounds and thresholds */
16 **if** $i_m < k_m$ **then** $\alpha_m \leftarrow \dfrac{{}^{01}a_{i_m+1}^m - {}^{01}a_{i_m}^m}{{}^{10}a_{i_m}^m - {}^{10}a_{i_m+1}^m}$ **else** $\alpha_m \leftarrow +\infty$
17 **if** $j_m < \ell_m$ **then** $\beta_m \leftarrow \dfrac{{}^{01}b_{j_m+1}^m - {}^{01}b_{j_m}^m}{{}^{10}b_{j_m}^m - {}^{10}b_{j_m+1}^m}$ **else** $\beta_m \leftarrow +\infty$
18 **if** $i_m \leq k_m$ **and** $j_m \leq \ell_m$ **then** $\delta_m \leftarrow \dfrac{{}^{01}a_{i_m}^m + 1 - {}^{01}b_{j_m}^m}{{}^{10}b_{j_m}^m - {}^{10}a_{i_m}^m}$ **else**
$\delta_m \leftarrow +\infty$
19 **if** $(\omega \leftarrow \min_{m \in \mathcal{C}_n}\{\min\{\alpha_m, \beta_m, \delta_m\}\}) < +\infty$ **then** /* Update */
20 $k_n \leftarrow k_n + 1$; $a_{k_n}^n \leftarrow (0,0)$
21 **for all** $m \in \mathcal{C}_n$ **do**
22 **if** $\alpha_m \leq \omega$ **then** $i_m \leftarrow i_m + 1$
23 **if** $\beta_m \leq \omega$ **then** $j_m \leftarrow j_m + 1$
24 **if** $\delta_m \leq \omega$ **then**
25 h$(m) \leftarrow 0$; $\mu_m^1 \leftarrow \delta_m$; $j_m \leftarrow \ell_m + 1$
26 **if** h$(m) = 1$ **then**
27 $a_{k_n}^n \leftarrow ({}^{10}a_{k_n}^n + {}^{10}b_{j_m}^m, {}^{01}a_{k_n}^n + {}^{01}b_{j_m}^m + 1)$
28 **else**
29 $a_{k_n}^n \leftarrow ({}^{10}a_{k_n}^n + {}^{10}a_{i_m}^m, {}^{01}a_{k_n}^n + {}^{01}a_{i_m}^m)$
30 **if** $a_{k_n}^n = a_{k_n-1}^n$ **then** $k_n \leftarrow k_n - 1$
31 **until** no change over i_m, j_m or h(m) for all $m \in \mathcal{C}_n$ inside the loop;
32 **if** s$(n) = 0$ **then** $\mathcal{X}_n^1 \leftarrow \emptyset$ **else** /* Compute $\mathcal{X}_n^1$ and μ_m^1 for all $m \in \mathcal{C}_n$ */
33 Symmetrical...
34 **for all** $m \in \mathcal{C}_n$ **do** /* Exit */
35 Free memory space used by $\mathcal{X}_m^0$ and $\mathcal{X}_m^1$

Algorithm 2: **broadcast_top_down**(n): broadcast the threshold λ_n^1 of node n to its children

```
for all m ∈ C_n do
    if μ_m^0 < λ_n^1 then
        λ_m^1 = μ_m^0
    else
        if λ_n^1 > μ_m^1 then
            λ_m^1 = μ_m^1
        else
            λ_m^1 = λ_n^1
    broadcast_top_down(m)
```

Algorithm 3: **main**(): main algorithm - r is the root node of T

```
compute_bottom_up(r)
get_root_1-threshold()        /* Compute λ_r^1 from X_r^0 and X_r^1 */
broadcast_top_down(r)
```

4.4 Complexity analysis

For a node n of T, $|T_n|$ and $|\mathcal{C}_n|$ design respectively the total number of nodes in T_n and the number of children nodes of n (naturally $|T| = |T_r|$).

Algorithm 1 is recursive. During the execution of the non-recursive part algorithm with node n as parameter, the total amount of memory space used (apart the space used for local variables which is linear with the number of children nodes of n) is for storing $\mathcal{X}_m^0$ and $\mathcal{X}_m^1$, the sets of configurations of the children of n, as well as the resulting configurations $\mathcal{X}_n^0$ and $\mathcal{X}_n^1$. Since the memory space used for children is freed at the exit of the algorithm, the sets $\mathcal{X}_n^0$, $\mathcal{X}_n^1$, $\mathcal{X}_m^0$ and $\mathcal{X}_m^1$ for all children m of n are the only ones stored in memory during an execution of Algorithm 1. From Remark 2, the total memory space used for storing these configurations is linear with the number of nodes in T_n. Let us now consider the time spent in the non-recursive part of Algorithm 1. Initializations are done in a time linear with the number of children of n. At each iteration of the main loop (except the last one), at least one of the following events occurs:

- an increment of i_m,
- an increment of j_m,
- a change of $\mathrm{h}(m)$.

The last event occurs at most once for each child node. The total number of the two first ones is given by the total number of configuration in the sets $\mathcal{X}_m^0$ and $\mathcal{X}_m^1$ which is linear with the number of unknown nodes in T_n, thus with $|T_n|$. Moreover the time spent in each iteration of the main loop is linear with the number of children of n. It comes that the non-recursive time complexity of Algorithm 1 is, up to a constant factor, bounded by $|\mathcal{C}_n||T_n|$.

The total time complexity of a call of Algorithm 1 over the root node of T is obtained by summing the non-recursive time complexity over all the nodes of T. This sum is smaller, again up to a constant factor, than $\sum_{n\in T} |\mathcal{C}_n||T_n| \leq |T| \sum_{n\in T} |\mathcal{C}_n| = |T|(|T|-1)$. Finally the time complexity of Algorithm 1 is quadratic with the size of T.

Algorithm 2 does not allow any proper memory space and is basically a depth-first tree traversal. Its time complexity is $O(|T|)$.

The complexity of finding the 1-threshold of the root node from $\mathcal{X}_r^0$ and $\mathcal{X}_r^1$ is linear with the total number of elements in $\mathcal{X}_r^0$ and $\mathcal{X}_r^1$, each of these sets having a cardinal smaller than the number of nodes of T (Remark 2).

When considering the complexities of all the stages of the main algorithm, it comes it has $O(|T|)$ memory space complexity and $O(|T|^2)$ time complexity. These complexities have to be compared with the ones of the Sankoff algorithm which performs the maximum parsimonious reconstruction with fixed parameters in time $O(|T|)$ using $O(|T|)$ memory space. A quadratic time complexity is actually not critical with regard to the typical size of evolutionary trees available. However this complexity has to be better understood. In particular it could be worthy to express it with respect to the number of unknown nodes.

It is possible to improve the way in which Algorithm 1 computes the sets of configurations by storing the values α_m, β_m, δ_m for all the children m of n, in a min heap of size $(3|\mathcal{C}_n|)$ and by keeping track of the variables to update at each iteration of the main loop. The time complexity of the non recursive part of Algorithm 1 would become $O(|T_n| \log |\mathcal{C}_n|)$ but it does not lead to a better bound for the overall complexity which remains *a priori* quadratic.

5 Applications

To be consistent with the figures of the papers cited, the states 0 and 1 are respectively displayed '$\circ$' and '$\bullet$' in the figures of this section.

5.1 Dependency of gains and losses on parameters

The main purpose of the present work is to explore the relationship between the transition costs and the corresponding ancestral reconstructions. This is also the aim of [11] where Ree and Donoghue summarize an ancestral reconstruction by the numbers of gains and losses (its configuration) and plot the evolution of the configurations with minimal costs *versus* the ratio $\frac{c_{01}}{c_{10}}$ (unfortunately we consider in this paper, the inverse ratio $\gamma = \frac{c_{10}}{c_{01}}$). In particular they determine this so-called "cost-change graph" for the ancestral reconstruction of the tree depicted in Figure 6. In our framework, the configurations with minimal costs can be determined from the sets $\mathcal{X}_r^0$ and $\mathcal{X}_r^1$, which are, for the tree and the initial function of Figure 6:

$$\begin{aligned}
\mathcal{X}_r^0 &= \{(13,2)\ ^{\frac{1}{4}}_{;}\ (9,3)\ ^{\frac{2}{7}}_{;}\ (2,5)\ ^{3}_{;}\ (1,8)\ ^{6}_{;}\ (0,14)\} \\
\mathcal{X}_r^1 &= \{(19,0)\ ^{\frac{1}{5}}_{;}\ (14,1)\ ^{\frac{1}{3}}_{;}\ (5,4)\ ^{1}_{;}\ (4,5)\ ^{3}_{;}\ (3,8)\ ^{6}_{;}\ (2,14)\}
\end{aligned}$$

Above the commas are displayed the thresholds bounding the domains of parameters over which each configuration corresponds to an assignment with a minimal cost.

The set of unconstrained configurations with a minimal cost $\mathcal{X}_r$ is then easy to compute:

$$\mathcal{X}_r = \{(19,0)\ ^{\frac{1}{5}}_{;}\ (14,1)\ ^{\frac{1}{3}}_{;}\ (2,5)\ ^{3}_{;}\ (1,8)\ ^{6}_{;}\ (0,14)\}$$

The set $\mathcal{X}_r$ and the thresholds are the same as the ones plotted in Figure 2(a) of [11]. Ree and Donoghue obtained these configurations by iteratively computing assignments with minimal costs over a user defined range of values of parameters. Besides computational efficiency issues, a drawback of this kind of approaches is that, depending upon the bounds and the step used to explore this range of values, some of

the configurations with minimal costs can be missed and the thresholds such obtained are approximative (see below).

Ree and Donoghue define two thresholds $C_{G\max}$ and $C_{L\max}$ over the nodes which are related to our thresholds λ_n^0 and λ_n^1 in the following way. They take the reconstruction obtained with $c_{01} = c_{10}$ (*i.e.* $\gamma = 1$) as reference and associate to each node n assigned to `0` by this particular reconstruction, the threshold $C_{G\max}$ which is the maximum value of $\frac{c_{01}}{c_{10}}$ such that the reconstructed state of n remains `0`: $C_{G\max}$ is the inverse of λ_n^0. For a node n reconstructed as `1`, they define $C_{L\max}$ as the maximum value of $\frac{c_{10}}{c_{01}}$ such that the reconstructed state of n remains `1`, which is actually the threshold λ_n^1. Certainly for the approximative way these thresholds are computed, the values of $C_{L\max}$ reported in Figure 1 of [11] do not match the corresponding exact values of λ_n^1, computed using our algorithm, we display in Figure 6 (they are quite close: 2.8 instead of 3 and 5.7 instead of 6). Although values of $C_{G\max}$ are consistent with our results.

5.2 Dollo parsimony, support of reconstructed states and cost assumptions

The Dollo approach of character-state reconstruction is based on the (strong) assumption of irreversible evolution, called Dollo's law [4, 5]:

> *An organism never returns exactly to a former state, even if it finds itself placed in conditions of existence identical to those in which it has previously lived.*
>
> Louis Dollo (1893)

In Dollo parsimony reconstruction, this law is interpreted in a narrowed but more operative way, that is: "a character that has been lost during the evolution of a particular lineage cannot be regained". The Dollo reconstruction remains essentially to find an assignment with a minimal number of `10`-transitions among the assignments containing at most one `01`-transition.

The Dollo reconstruction is quite simple and fast to compute, again by dynamic programming, and is still widely used to infer ancestral states in phylogenetic studies, even if the irreversibility hypothesis is still controversial [5, 1].

In order to show how the Dollo reconstruction takes place in the general parsimony framework, we remark that if there exists an assignment with less than one `01`-transition (this is in particular always the case when the initial function is defined only on the leaves), then the assignment given by the Dollo reconstruction has a minimal γ-cost for a parameter γ in a certain range. Remark that, since a single gain is allowed, the generalized maximum parsimonious reconstruction obtained with an infinite gain cost, or equivalently with an infinitesimal parameter γ, does not always coincide with the Dollo reconstruction.

In [2], Cunningham studies the ancestral reconstruction of the character "larval feeding" in starfish. The tree and the initial function are displayed in Figure 7. He compares the ancestral reconstruction given by Dollo parsimony with the ones computed by other methods, in particular the general parsimony with equal transition costs (*i.e.* $\gamma = 1$). These two ancestors reconstructions differ and Cunningham deplores the lack of degree of support for the character states reconstructed with one or another method.

The `1`-thresholds provide a natural way to quantify the degree of support for reconstructed states. This can be done by considering a given *a priori* probability distribution over the parameter γ, or equivalently over the transition costs c_{01} and c_{10}, and by defining the probability for a node n to be reconstructed to `1` as the cumulative probability of λ_n^1 in the *a priori* distribution.

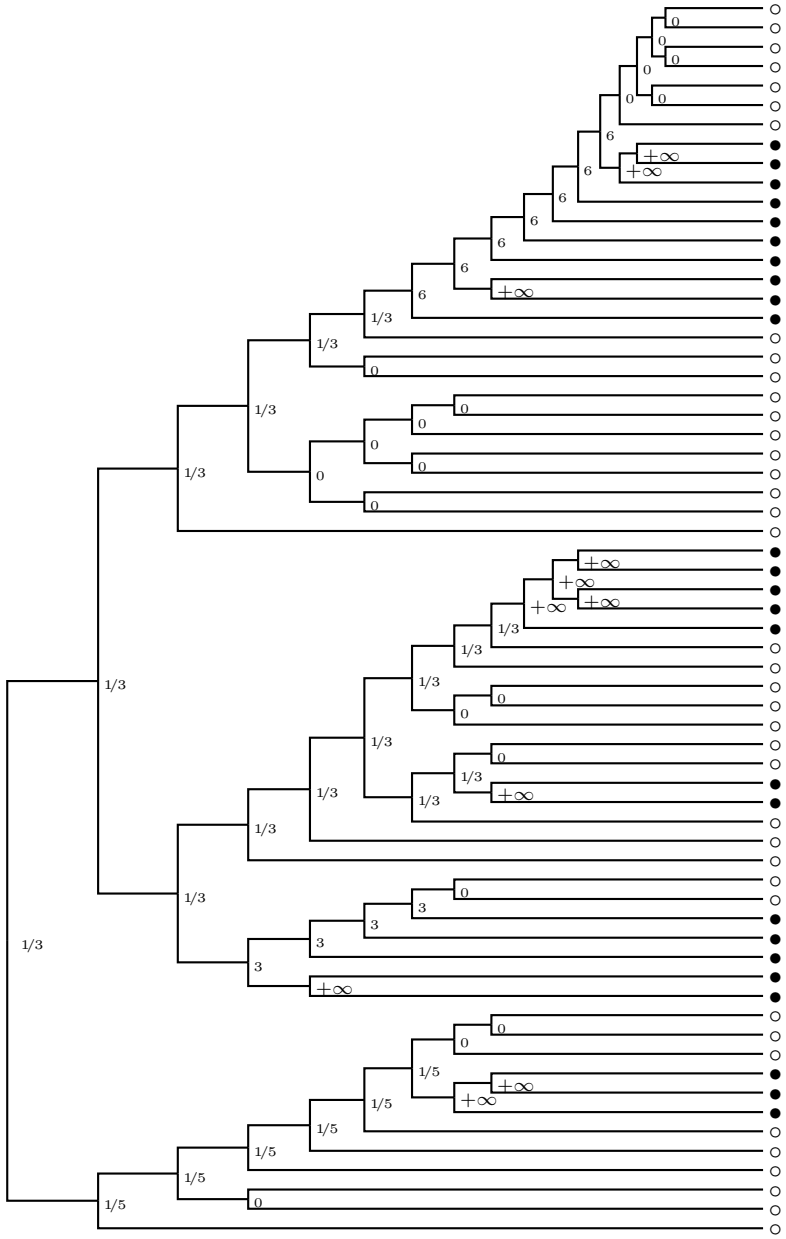

Figure 6: 0- and 1-thresholds (confounded) of internal nodes for a tree and an initial function defined only on the leaves [11].

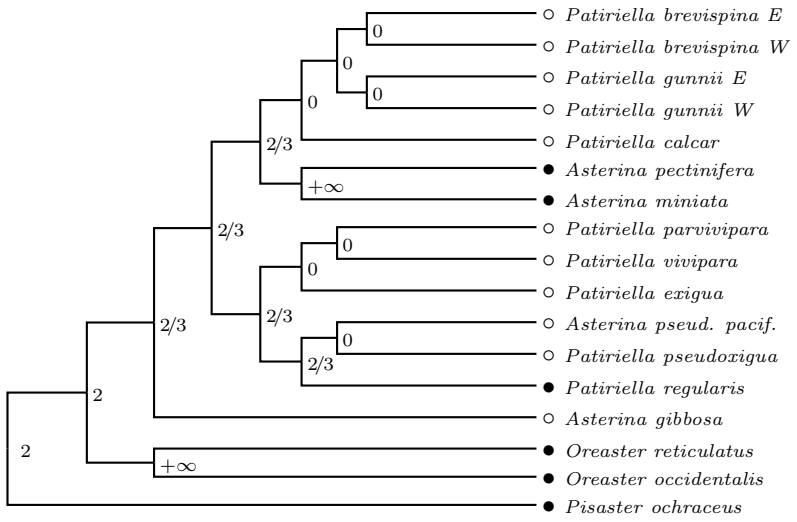

Figure 7: 0- and 1-thresholds (confounded) of internal nodes for a tree and an initial function defined only on the leaves [2].

In Figure 7 we display the 1-thresholds of all nodes. We observe that the Dollo reconstruction has a minimal γ-cost for γ in $]0, \frac{2}{3}]$. In terms of transition costs, it correspond to pairs of positive values such that $c_{10} \leq \frac{2}{3}c_{01}$. Roughly speaking, it means a third of the transition costs space supports the Dollo reconstruction. In other words, if we consider *a priori* distributions uniform of growing supports over the space of parameters (c_{01}, c_{10}), the limit probability of the Dollo reconstruction is $\frac{1}{3}$, which can be considered as not small enough to reject the irreversibility hypothesis.

Another, and immediate, benefit of the 0- and 1-thresholds it that they allow us to see at a glance witch are the constraints the transition costs have to satisfy if we want the corresponding reconstruction to be consistent with a given evolutionary hypothesis like irreversibility.

Moreover a maximum parsimonious reconstruction is necessarily consistent with the 0- and 1-thresholds. In particular there is no maximum parsimonious reconstruction assigning 1 to the right child of the root node (with 1-threshold $\frac{1}{5}$) without assigning the same state to the left child (with 1-threshold $\frac{1}{3}$) in the tree of Figure 6, unless we relax the assumption that the cost parameters do not change along the tree.

A last remark is that if the 0- and 1-thresholds are confounded for all nodes, there are exactly $(j+1)$ alternative maximum parsimonious reconstructions of positive supports, where j is the number of "0-1-thresholds" observed on the tree and "of positive supports" means there is an interval I of positive length such that the reconstruction considered has minimal γ-costs for all $\gamma \in I$. These alternative reconstructions are then straightforward to enumerate from the 0-1-thresholds. For instance, there are only 3 different such maximum parsimonious reconstructions on the tree of Figure 7.

Acknowledgements

I thank Elisabeth Remy and Pierre Pontarotti for helpful discussions and comments on earlier drafts of this work.

References

[1] Rachel Collin and Maria Pia Miglietta. Reversing opinions on dollos law. *Trends in Ecology & Evolution*, 23(11):602–609, 2008.

[2] Clifford W. Cunningham. Some limitations of ancestral character-state reconstruction when testing evolutionary hypotheses. *Systematic Biology*, 48(3):665–674, 1999.

[3] Clifford W. Cunningham, Kevin E. Omland, and Todd H. Oakley. Reconstructing ancestral character states : a critical reappraisal. *Trends in ecology & evolution*, 13:361–366, 1998.

[4] L. Dollo. Les lois de l'évolution. *Bulletin de la Société Belge de Géologie de Paléontologie et d'Hydrologie*, 7:164–166, 1893.

[5] E. E. Goldberg and B. Igic. On phylogenetic tests of irreversible evolution. *Evolution*, 62:2727–2741, 2008.

[6] D Gusfield, K Balasubramanian, and D Naor. Parametric optimization of sequence alignment. *Algorithmica*, 12:312–326, 1994.

[7] W. P. Maddison and D.R. Maddison. Mesquite: a modular system for evolutionary analysis. version 2.71, 2009.

[8] Kevin E. Omland. The assumptions and challenges of ancestral state reconstruction. *Systematic Biology*, 48(3):665–674, 1999.

[9] Lior Pachter and B Sturmfels. Parametric inference for biological sequence analysis. *Proc Natl Acad Sci U S A*, 101:16138–43, 2004.

[10] Lior Pachter and B Sturmfels, editors. *Algebraic Statistics for Computational Biology*. Cambridge University Press, Cambridge, 2005.

[11] Richard H. Ree and Michael J. Donoghue. Step matrices and the interpretation of homoplasy. *Systematic Biology*, 47(4):582–588, 1998.

[12] D Sankoff. Minimal mutation trees of sequences. *SIAM J. Appl. Math.*, 28:35–42, 1975.

[13] D. Sankoff and R. J. Cedergren. Simultaneous comparison of three or more sequences related by a tree. In D. Sankoff and J. B. Kruskal, editors, *Time warps, string edits, and macromolecules: The theory and practice of sequence comparison*, pages 253–263. Addison-Wesley, Reading, Massachusetts, 1983.

[14] D.L. Swofford and W. P. Maddison. Parsimony, character-state reconstructions, and evolutionary inferences. In R. L. Mayden, editor, *Systematics, Historical Ecology, and North American Freshwater Fishes*, pages 187–223. Stanford University Press, Stanford, California, 1992.

[15] M Waterman, M Eggert, and E Lander. Parametric sequence comparisons. *Proc Natl Acad Sci U S A*, 89:6090–6093, 1992.