Parsimonious identification of evolutionary shifts for quantitative characters

Olivier Chabrol and Gilles Didier

Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France

{olivier.chabrol,gilles.didier}@univ-amu.fr

September 5, 2017

Abstract

Detecting shifts in the rate or the trend of a character evolution is an important question which has been widely addressed. To our knowledge, all the approaches developed so far for detecting such shifts from a quantitative character strongly involved stochastic models of evolution.

We propose here a novel method based on an asymmetric version of the linear parsimony (aka Wagner parsimony) for identifying the most parsimonious split of a tree into two parts between which the evolution of the character is allowed to differ. To this end, we evaluate the cost of splitting a phylogenetic tree at a given node as the integral, over all pairs of asymmetry parameters, of the most parsimonious cost which can be achieved by using the first parameter on the subtree pending from this node and the second parameter elsewhere. By testing all the nodes, we then get the most parsimonious split of a tree with regard to the character values at its tips.

A study of the partial costs of the tree enabled us to develop a polynomial algorithm for determining the most parsimonious splits and their evolutionary costs. Applying this algorithm on two toy examples shows that using more than one asymmetry parameter does not always lower parsimonious costs. By using the approach on biological datasets, we obtained splits consistent with those identified by previous stochastic approaches.

1 Introduction

Changes in the rate and/or the trend of a character evolution are often associated with major biological phenomena such as adaptation or evolutionary convergence [10]. Detecting shifts in the evolution of a trait is thus an important question, which has been widely addressed, notably in the case of quantitative characters. All the approaches developed so far for studying this question are based on stochastic models of evolution for continuous characters, mainly Brownian motion or Ornstein-Uhlenbeck processes [1, 9, 3, 6, 11, 10, 13, 14, 15, 16, 17]. The method presented in [2] used parsimony approaches but only for reconstructing ancestral states which were next compared with simulations under Brownian models. Other approaches addressed the related question of testing the significance of splits given *a priori* (e.g. from environmental or dietary considerations) with regard to stochastic models [21, 22, 18].

We propose here a novel approach based on an asymmetric version of the linear parsimony introduced in [4] and further studied in [5], in both cases in an ancestral reconstruction context. We do not discuss the pros and cons of parsimonious approaches with regard to those based on stochastic models [19, 12, 8]. From a technical point of view, ideas behind parsimonious algorithms can generally be transposed to stochastic frameworks. We plan to adapt principles used in our computations for determining the most likely split of a phylogenetic tree with regard to Brownian and Ornstein-Uhlenbeck models.

The problem that we are studying here can be formally stated as follows. Being given a phylogenetic tree and a continuous character which is known only for extant species (i.e. here the tips of the tree), we aim to identify the node of the tree from which the evolution of the character differs (or differs the most) from the rest of tree, if such a node exists.

A parsimonious framework enables us to compute the minimal evolutionary cost of the extant character values. This cost plays the role of the likelihood in probabilistic models. In the asymmetric linear parsimony case, the evolutionary cost depends on the asymmetry parameter α . In order to evaluate the

relevance of putting an evolutionary shift at a node n , we consider evolutionary costs obtained by using an asymmetry parameter on the subtree $\mathcal{T}_n$ and another parameter on the rest of the tree. Testing all the nodes of the tree enables us to identify the subtree on which the evolution of the character differs the most. We first provide a polynomial algorithm for computing such costs and for integrating their over all pairs of asymmetry parameters. This way we get a parsimonious method for finding the subtree on which the evolution differs the most. We show on toy examples that considering more than one asymmetric does not always lead to lower the evolutionary cost. Finally, our approach is applied on two previous biological datasets from [18] and [20]. Shifts detected with our approach are consistent with those obtained from methods based on stochastic models and well separate clades with specific evolution like whales from other cetaceans.

We developed *ParSplit* a software performing the parsimonious split of a phylogenetic tree with regard to extant values of characters written in C language. Sources of the software and toy examples are available at <https://github.com/gilles-didier/ParSplit>.

The rest of the paper is organized as follows. Section 2 presents definitions, notations and asymmetric linear parsimony then outlines the approach. The *partial cost functions*, which are the bases of our computations, are defined and studied in Section 3. Next, Section 4 shows how to integrate the (split) costs over all the possible values of the asymmetry parameter(s). Finally, results obtained with our approach on two toy examples and on two biological datasets are presented in Section 5 which ends with a short discussion.

2 Asymmetric parsimony

2.1 Definitions and Notations

We use the same notations as [5]. The cardinal of a finite set S is noted $|S|$. In what follows, $\mathcal{T}$ designates a rooted tree which may or may not be binary. As it should lead to no confusion, we still write $\mathcal{T}$ for its set of nodes. For all nodes $n \in \mathcal{T}$, we put

- $\mathcal{C}_n$ for the set of child nodes of n ,
- τ_n for the length of the branch ending at n ,
- $\mathcal{T}_n$ for the subtree of $\mathcal{T}$ rooted at n .

Let us consider a subset $\mathcal{K}$ of nodes of $\mathcal{T}$ and a map ϑ from $\mathcal{K}$ to the set of real numbers $\mathbb{R}$. The map ϑ will be referred to as the *initial function* and the nodes of $\mathcal{K}$ are said *known*. For all nodes n of $\mathcal{T}$, we put $\mathcal{K}_n$ for the subset of known nodes of the subtree $\mathcal{T}_n$, i.e. $\mathcal{K}_n = \mathcal{K} \cap \mathcal{T}_n$. Though the general case can be treated as in [5], we assume that all and only the tips of the tree have known states in order to lighten the statements. In plain English, $\mathcal{K}$ (resp. $\mathcal{K}_n$) is the set of tips of $\mathcal{T}$ (resp. of $\mathcal{T}_n$).

The values of $\{\vartheta(k) \mid k \in \mathcal{K}\}$ are the *known states* of $\mathcal{T}$. For all nodes n , we put $\vartheta(\mathcal{K}_n)$ for the set $\{\vartheta(k) \mid k \in \mathcal{K}_n\}$.

A ϑ -assignment of $\mathcal{T}$ is a map ξ from $\mathcal{T}$ to the set of real numbers which extends ϑ (i.e. such that $\xi(n) = \vartheta(n)$ for all nodes $n \in \mathcal{K}$). The set of all ϑ -assignments of $\mathcal{T}$ is noted Ξ_{ϑ} .

2.2 Asymmetric parsimonious cost

A parsimony framework is based on a way of computing the evolutionary cost of an assignment. In the time-dependent-asymmetric-linear parsimony (*TDALP*, see [5]) case, an ancestor/child transition from value x at node n to value y at its child m is associated with the cost:

$$\Delta_{\gamma,\lambda}(x, y, \tau_m) = \begin{cases} \gamma\phi(\tau_m)(y - x) & \text{if } x < y, \\ \lambda\phi(\tau_m)(x - y) & \text{if } x \geq y, \end{cases}$$

where λ and γ are two nonnegative real numbers and ϕ is a function from $\mathbb{R}_{>0}$ to $\mathbb{R}_{>0}$ (in what follows, we make the assumption that $\tau_n > 0$ for all nodes $n \in \mathcal{T}$). The case where ϕ is constant with $\phi(\tau) = 1$ for all τ corresponds to the standard parsimony scheme [4, 7]. In an evolutionary context, it makes sense to choose a decreasing function ϕ , but any positive function can be used. In the current implementation of the approach, users can choose between $\phi(\tau) = \frac{1}{\tau}$ and $\phi(\tau) = 1$.

The cost $\Delta_{\gamma,\lambda}(\xi)$ of an assignment ξ of $\mathcal{T}$ is then the sum of all the costs of its ancestor/child transitions:

$$\Delta_{\gamma,\lambda}(\xi) = \sum_{n \in \mathcal{T}} \sum_{m \in \mathcal{C}_n} \Delta_{\gamma,\lambda}(\xi(n), \xi(m), \tau_m)$$

Let us first remark that multiplying both λ and α by a positive constant factor just lead to multiply the assignments' cost by the same factor. Since computing an evolutionary cost with $\alpha = \lambda = 0$ makes no sense, we can assume $\gamma + \lambda > 0$ and divide both parameters by $\gamma + \lambda$ in order to normalize the parsimonious costs. From now on, we assume that $\gamma + \lambda = 1$. The cost of an assignment then only depends on a single parameter $\alpha \in [0, 1]$, and is called α -cost:

$$\Delta_{\alpha}(x, y, \tau_m) = \begin{cases} \alpha \phi(\tau_m)(y - x) & \text{if } x < y, \\ (1 - \alpha) \phi(\tau_m)(x - y) & \text{if } x \geq y. \end{cases} \quad (1)$$

Below, α will be referred to as the *asymmetry parameter*. Reconstructing with $\alpha < \frac{1}{2}$ (resp. with $\alpha = \frac{1}{2}$, with $\alpha > \frac{1}{2}$) makes the cost of an increase smaller than (resp. equal as, greater than) that of a decrease of the same amount. Intuitively, it corresponds to the assumption that the character evolves with a positive trend (resp. without trend, with a negative trend). Note that a slightly different parametrization is considered in [5].

The (*parsimonious*) α -cost of the pair $(\mathcal{T}, \vartheta)$ is defined as the minimal cost which can be achieved by a ϑ -assignment:

$$\check{\Delta}_{\alpha}(\mathcal{T}, \vartheta) = \min_{\xi \in \Xi_{\vartheta}} \Delta_{\alpha}(\xi). \quad (2)$$

2.3 Outline of the approach

Our aim is to find the most parsimonious way of splitting a phylogenetic tree $\mathcal{T}$ into two parts according to an initial function ϑ (i.e., the characters values of tips). In what follows, we will consider the two ways of splitting a phylogenetic tree at a node n displayed in Figure 1 and below referred to as *A-split* and *B-split*. Devising a parsimonious approach for identifying evolutionary shifts first requires to be able to associate an evolutionary cost at a given split.

Figure 1: Two different ways of splitting a phylogenetic tree at a node n .

In a parsimonious context, a natural idea should be to define the cost of a split of a tree into two parts as the minimum cost, over all pairs $(\alpha, \alpha') \in [0, 1]^2$, which can be achieved by using parameter α on a part of the tree and parameter α' on the other part. Unfortunately, even without splitting the tree, minimizing the parsimonious cost over all parameters α leads to trivial solutions: by setting $\alpha = 0$ (resp. $\alpha = 1$) the assignment which associates the greatest (resp. the smallest) of the known values to all unknown nodes has a null α -cost!

In order to overcome this issue, we first define the *no-split cost* of $(\mathcal{T}, \vartheta)$ as the sum over all parameters $\alpha \in [0, 1]$ of the α -parsimonious costs $\check{\Delta}_{\alpha}(\mathcal{T}, \vartheta)$. Note that this cost is parameter-free and depends only on $\mathcal{T}$ and ϑ . It somehow reflects the disparity of the character with regard to the tree. We then define the cost of a split of $\mathcal{T}$ into two parts as the sum over all pairs of parameters $(\alpha, \alpha') \in [0, 1]^2$ of the smallest cost which can be achieved by using parameter α on the first part and parameter α' on the second one. The most parsimonious split of a phylogenetic tree with regard to a given initial function is either the whole tree if the no-split cost is lower than all the costs of the A- and B-splits, or the A- or B-split with the smallest cost otherwise.

The next two sections show how to compute the no-, A- and B-splits costs.

3 Partial cost functions

For all nodes n of $\mathcal{T}$, we shall define and study the partial cost functions $f_n(\alpha, x)$, $\hat{f}_n(\alpha, x)$, $h_n(\alpha, x)$ and $\hat{h}_n(\alpha, x)$ that give the minimal costs which be obtained by associating x to n or its direct ancestor and by using parameter α on the corresponding parts of the tree as displayed in Figure 1.

Namely, following [5, 4], for all nodes n of $\mathcal{T}$, we put $f_n(\alpha, x)$ for the minimal cost which can be achieved by an assignment ξ of the subtree $\mathcal{T}_n$ (red part of Figure 1-left) such that $\xi(n) = x$. For all nodes n of $\mathcal{T}$ but its root, we put $\hat{f}_n(\alpha, x)$ for the minimal cost which can be achieved by an assignment ξ of the tree consisting of m , the direct ancestor of n and of the subtree $\mathcal{T}_n$, which will be referred to as the *stem subtree of n* and noted $\hat{\mathcal{T}}_n$ (red part of Figure 1-right), verifying $\xi(m) = x$.

Cost functions $h_n(\alpha, x)$ and $\hat{h}_n(\alpha, x)$ give the complementary costs of $\hat{f}_n(\alpha, x)$ and $f_n(\alpha, x)$. Namely, for all nodes n of $\mathcal{T}$, we put $h_n(\alpha, x)$ for the minimum reconstruction cost of $\mathcal{T}$ without n , the branch bearing n and the subtree rooted at n (blue part of Figure 1-right), which associates x to the direct ancestor of n . Last, we put $\hat{h}_n(\alpha, x)$ for the minimal cost which can be achieved by an assignment ξ of the whole tree with n and the branch bearing n but without the subtree rooted at n (blue part of Figure 1-left) which is such that $\xi(n) = x$.

We first recall and slightly adapt a Theorem from [5].

Theorem 1 ([5]). *Let n be an unknown node of $\mathcal{T}$. The maps f_n and $\hat{f}_n$ are piecewise-linear and continuous and, for all $\alpha > 0$, the maps $x \rightarrow f_n(\alpha, x)$ and $x \rightarrow \hat{f}_n(\alpha, x)$ are both convex.*

More precisely, if all the nodes of $\mathcal{T}_n$ are unknown then $f_n(\alpha, x) = 0$ for all $\alpha > 0$ and all x . Otherwise, there exist:

- an integer u_n and a strictly increasing positive real sequence $(\Gamma_i^n)_{1 \leq i \leq u_n}$,
- an integer sequence $(v_i^n)_{0 \leq i \leq u_n}$ and for, all $0 \leq i \leq u_n$, a sequence $(\kappa_{i,j}^n)_{1 \leq j \leq v_i^n}$ of known states of $\mathcal{T}_n$ (i.e. in $\mathcal{V}(\mathcal{K}_n)$) verifying $\kappa_{i,j}^n < \kappa_{i,j'}^n \Leftrightarrow j < j'$; by convention we set $\kappa_{i,0}^n = -\infty$ and $\kappa_{i,v_i^n+1}^n = +\infty$,
- two nonnegative real bi-sequences $(A_{i,j}^n)_{0 \leq i \leq u_n, 0 \leq j \leq v_i^n}$ and $(B_{i,j}^n)_{0 \leq i \leq u_n, 0 \leq j \leq v_i^n}$,
- two real bi-sequences $(C_{i,j}^n)_{0 \leq i \leq u_n, 0 \leq j \leq v_i^n}$ and $(D_{i,j}^n)_{0 \leq i \leq u_n, 0 \leq j \leq v_i^n}$

such that, by setting $\Gamma_0^n = 0$ and $\Gamma_{u_n+1}^n = 1$ and for all $0 \leq i \leq u_n$, all $0 \leq j \leq v_i^n$, all $\alpha \in (\Gamma_i^n, \Gamma_{i+1}^n]$ and all $x \in (\kappa_{i,j}^n, \kappa_{i,j+1}^n]$, we have

$$f_n(\alpha, x) = -A_{i,j}^n \alpha x + B_{i,j}^n x + C_{i,j}^n \alpha + D_{i,j}^n,$$

all the coefficients being such that f_n is continuous. Moreover, the sequence $(-A_{i,j}^n \alpha + B_{i,j}^n)_{0 \leq j \leq v_i^n}$ (i.e. the x -coefficients of f_n) is increasing with $A_{i,0}^n = A_{i,v_i^n}^n = B_{i,v_i^n}^n \leq \sum_{m \in \mathcal{C}_n} \phi(\tau_m)$ and $B_{i,0}^n = 0$.

In the same way, if all the nodes of $\mathcal{T}_n$ are unknown then $\hat{f}_n(\alpha, x) = 0$ for all $\alpha > 0$ and all x . Otherwise there exist $\hat{u}_n$, $(\hat{\Gamma}_i^n)_{1 \leq i \leq \hat{u}_n}$, $(\hat{v}_i^n)_{0 \leq i \leq \hat{u}_n}$, $(\hat{\kappa}_{i,j}^n)_{0 \leq i \leq \hat{u}_n, 1 \leq j \leq \hat{v}_i^n}$, $(\hat{A}_{i,j}^n)_{0 \leq i \leq \hat{u}_n, 0 \leq j \leq \hat{v}_i^n}$, $(\hat{B}_{i,j}^n)_{0 \leq i \leq \hat{u}_n, 0 \leq j \leq \hat{v}_i^n}$, $(\hat{C}_{i,j}^n)_{0 \leq i \leq \hat{u}_n, 0 \leq j \leq \hat{v}_i^n}$ and $(\hat{D}_{i,j}^n)_{0 \leq i \leq \hat{u}_n, 0 \leq j \leq \hat{v}_i^n}$ verifying the same properties as their f_n -counterparts except that we have $\hat{A}_{i,0}^n = \hat{A}_{i,\hat{v}_i^n}^n = \hat{B}_{i,\hat{v}_i^n}^n \leq \phi(\tau_n)$ and $\hat{B}_{i,0}^n = 0$, and such that, for all $0 \leq i \leq \hat{u}_n$, all $0 \leq j \leq \hat{v}_i^n$, all $\alpha \in (\hat{\Gamma}_i^n, \hat{\Gamma}_{i+1}^n]$ and all $x \in (\hat{\kappa}_{i,j}^n, \hat{\kappa}_{i,j+1}^n]$, we have

$$\hat{f}_n(\alpha, x) = -\hat{A}_{i,j}^n \alpha x + \hat{B}_{i,j}^n x + \hat{C}_{i,j}^n \alpha + \hat{D}_{i,j}^n.$$

Proof. In [5], this theorem was established by considering another parametrization of the parsimony cost, namely

$$\Delta_\gamma(x, y, \tau_m) = \begin{cases} \gamma \phi(\tau_m)(y - x) & \text{if } x < y, \\ \phi(\tau_m)(x - y) & \text{if } x \geq y. \end{cases} \quad (3)$$

The version above was obtained by dividing the cost just above and the corresponding partial cost functions by $\gamma + 1$ and by applying the substitution $\alpha = \frac{\gamma}{\gamma+1}$ to transform the coefficients and the bounds of these cost functions. $\square$

A similar result can be stated for h_n and $\hat{h}_n$.

Theorem 2. Let n be a non-root unknown node of $\mathcal{T}$. The maps h_n and $\hat{h}_n$ are piecewise-linear and continuous and, for all $\alpha > 0$, the maps $x \rightarrow h_n(\alpha, x)$ and $x \rightarrow \hat{h}_n(\alpha, x)$ are both convex.

More precisely, if all the nodes of $\mathcal{T}_n$ are unknown then $h_n(\alpha, x) = 0$ for all $\alpha > 0$ and all x . Otherwise, there exist:

- an integer w_n and a strictly increasing positive real sequence $(\Upsilon_i^n)_{1 \leq i \leq w_n}$,
- an integer sequence $(z_i^n)_{0 \leq i \leq w_n}$ and for, all $0 \leq i \leq w_n$, a sequence $(\rho_{i,j}^n)_{1 \leq j \leq z_i^n}$ of known states of $\mathcal{T} \setminus \mathcal{T}_n$ (i.e. in $\vartheta(\mathcal{K} \setminus \mathcal{K}_n)$) verifying $\rho_{i,j}^n < \rho_{i,j'}^n \Leftrightarrow j < j'$; by convention we set $\rho_{i,0}^n = -\infty$ and $\rho_{i,z_i^n+1}^n = +\infty$,
- four real bi-sequences $(O_{i,j}^n)_{0 \leq i \leq w_n, 0 \leq j \leq z_i^n}$, $(P_{i,j}^n)_{0 \leq i \leq w_n, 0 \leq j \leq z_i^n}$, $(Q_{i,j}^n)_{0 \leq i \leq w_n, 0 \leq j \leq z_i^n}$ and $(R_{i,j}^n)_{0 \leq i \leq w_n, 0 \leq j \leq z_i^n}$

such that, by setting $\Upsilon_0^n = 0$ and $\Upsilon_{w_n+1}^n = 1$ and for all $0 \leq i \leq w_n$, all $0 \leq j \leq z_i^n$, all $\alpha \in (\Upsilon_i^n, \Upsilon_{i+1}^n]$ and all $x \in (\rho_{i,j}^n, \rho_{i,j+1}^n]$, we have

$$h_n(\alpha, x) = -O_{i,j}^n \alpha x + P_{i,j}^n x + Q_{i,j}^n \alpha + R_{i,j}^n,$$

all the coefficients being such that h_n is continuous. Moreover, the sequence $(-O_{i,j}^n \alpha + P_{i,j}^n)_{0 \leq j \leq z_i^n}$ (i.e. the x -coefficients of h_n) is increasing with $-O_{i,z_i^n}^n \alpha + P_{i,z_i^n}^n \geq 0$ for all $\alpha \in (\Upsilon_i^n, \Upsilon_{i+1}^n]$.

In the same way, if all the nodes of $\mathcal{T}_n$ are unknown then $\hat{h}_n(\alpha, x) = 0$ for all $\alpha > 0$ and all x . Otherwise there exist $\hat{w}_n$, $(\hat{\Upsilon}_i^n)_{1 \leq i \leq \hat{w}_n}$, $(\hat{z}_i^n)_{0 \leq i \leq \hat{w}_n}$, $(\hat{\rho}_{i,j}^n)_{0 \leq i \leq \hat{w}_n, 1 \leq j \leq \hat{z}_i^n}$, $(\hat{O}_{i,j}^n)_{0 \leq i \leq \hat{w}_n, 0 \leq j \leq \hat{z}_i^n}$, $(\hat{P}_{i,j}^n)_{0 \leq i \leq \hat{w}_n, 0 \leq j \leq \hat{z}_i^n}$, $(\hat{Q}_{i,j}^n)_{0 \leq i \leq \hat{w}_n, 0 \leq j \leq \hat{z}_i^n}$ and $(\hat{R}_{i,j}^n)_{0 \leq i \leq \hat{w}_n, 0 \leq j \leq \hat{z}_i^n}$ verifying the same properties as their h_n -counterparts and such that, for all $0 \leq i \leq \hat{w}_n$, all $0 \leq j \leq \hat{z}_i^n$, all $\alpha \in (\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n]$ and all $x \in (\hat{\rho}_{i,j}^n, \hat{\rho}_{i,j+1}^n]$, we have

$$\hat{h}_n(\alpha, x) = -\hat{O}_{i,j}^n \alpha x + \hat{P}_{i,j}^n x + \hat{Q}_{i,j}^n \alpha + \hat{R}_{i,j}^n.$$

Proof. The proof follows the same ideas as that of Theorem 1 from [5], a noticeable difference being that here the induction goes from the root to the tips. Namely, we shall prove the theorem by proving the three following assertions:

1. if n is a child of the root then h_n satisfies the properties of the theorem;
2. for all nodes n , if h_n satisfies the properties of the theorem then so it is for $\hat{h}_n$;
3. for all nodes n , if $\hat{h}_n$ satisfies the properties of the theorem then so it is for all h_k with k child of n .

By construction, if n is a direct descendant of the root r , we have that

$$h_n(\alpha, x) = \sum_{k \in \mathcal{C}_r \setminus \{n\}} \hat{f}_k(\alpha, x). \quad (4)$$

Assertion 1 then follows from Theorem 1.

In order to prove Assertion 2, let us remark that for all nodes n , by setting

$$g_x(y) = \Delta_\alpha(y, x, \tau_n) + h_n(\alpha, y),$$

we have

$$\hat{h}_n(\alpha, x) = \inf_{y \in \mathbb{R}} g_x(y).$$

Let us assume that h_n satisfies the properties of the theorem. In particular h_n is piecewise linear with, for all $0 \leq i < w_n$, all $\alpha \in [\Upsilon_i^n, \Upsilon_{i+1}^n]$, all $0 \leq j \leq z_i^n$ and all $y \in (\rho_{i,j}^n, \rho_{i,j+1}^n]$:

$$h_n(\alpha, y) = -O_{i,j}^n \alpha y + P_{i,j}^n y + Q_{i,j}^n \alpha + R_{i,j}^n.$$

Let us first consider the case where $y \geq x$ for which we have $\Delta_\alpha(x, y, \tau_n) = (1 - \alpha)\phi(\tau_n)(y - x)$.

For all $0 \leq i < w_n$, all $\alpha \in [\Upsilon_i^n, \Upsilon_{i+1}^n]$, all $0 \leq j \leq z_i^n$ and all $y \in (\rho_{i,j}^n, \rho_{i,j+1}^n]$, we get that

$$\begin{aligned} g_x(y) &= (1 - \alpha)\phi(\tau_n)(y - x) - O_{i,j}^n \alpha y + P_{i,j}^n y + Q_{i,j}^n \alpha + R_{i,j}^n \\ &= [-(O_{i,j}^n + \phi(\tau_n))\alpha + P_{i,j}^n + \phi(\tau_n)] y - (1 - \alpha)\phi(\tau_n)x + Q_{i,j}^n \alpha + R_{i,j}^n \end{aligned}$$

Since h_n satisfies the properties of the theorem, it is convex with respect to x . The sequence $(-(O_{i,j}^n + \phi(\tau_n))\alpha + P_{i,j}^n + \phi(\tau_n))_{0 \leq j \leq z_i^n}$ thus increases with j . Let j_α^+ be defined as

- $j_\alpha^+ = 0$ if $-(O_{i,j}^n + \phi(\tau_n))\alpha + P_{i,j}^n + \phi(\tau_n) \geq 0$ for all $0 \leq j \leq \hat{z}_i^n$;
- the greatest integer smaller or equal to z_i^n and such that $-(O_{i,j_\alpha^+ - 1}^n + \phi(\tau_n))\alpha + P_{i,j_\alpha^+ - 1}^n + \phi(\tau_n) < 0$ otherwise. In other words, $j_\alpha^+ - 1$ is the index of the last interval on which g_x strictly decreases with $y \geq x$ and is the greatest index such that $\alpha \leq \frac{P_{i,j_\alpha^+ - 1}^n + \phi(\tau_n)}{O_{i,j_\alpha^+ - 1}^n + \phi(\tau_n)}$.

Since g_x decreases before $\rho_{i,j_\alpha^+}^n$ and increases after this value, we get that

$$\inf_{y \geq x} g_x(y) = \begin{cases} g_x(\rho_{i,j_\alpha^+}^n) & \text{if } x < \rho_{i,j_\alpha^+}^n, \\ g_x(x) & \text{otherwise,} \end{cases} \quad (5)$$

and that

$$g_x(\rho_{i,j_\alpha^+}^n) \leq g_x(x) \quad \text{for all } x < \rho_{i,j_\alpha^+}^n. \quad (6)$$

Let us now consider the case where $y < x$ for which we have $\Delta_\alpha(x, y, \tau_n) = \alpha\phi(\tau_n)(x - y)$. For all $0 \leq i < w_n$, all $\alpha \in [\Upsilon_i^n, \Upsilon_{i+1}^n]$, all $0 \leq j \leq z_i^n$ and all $y \in (\rho_{i,j}^n, \rho_{i,j+1}^n]$, we get that

$$\begin{aligned} g_x(y) &= \alpha\phi(\tau_n)(x - y) - O_{i,j}^n\alpha y + P_{i,j}^n y + Q_{i,j}^n\alpha + R_{i,j}^n \\ &= [-(O_{i,j}^n + \phi(\tau_n))\alpha + P_{i,j}^n] y + \alpha\phi(\tau_n)x + Q_{i,j}^n\alpha + R_{i,j}^n \end{aligned}$$

Let us define j_α^- as

- $j_\alpha^- = z_i^n + 1$ if $-(O_{i,j}^n + \phi(\tau_n))\alpha + P_{i,j}^n < 0$ for all $0 \leq j \leq \hat{z}_i^n$;
- the smallest index such that $-(O_{i,j}^n + \phi(\tau_n))\alpha + P_{i,j}^n \geq 0$ otherwise. In other words, j_α^- corresponds to the first interval on which g_x does not decrease with $y < x$ and is the smallest index such that $\alpha \geq \frac{P_{i,j_\alpha^-}^n}{O_{i,j_\alpha^-}^n + \phi(\tau_n)}$.

We have that

$$\inf_{y < x} g_x(y) = \begin{cases} g_x(x) & \text{if } x < \rho_{i,j_\alpha^-}^n, \\ g_x(\rho_{i,j_\alpha^-}^n) & \text{otherwise.} \end{cases} \quad (7)$$

and that

$$g_x(\rho_{i,j_\alpha^-}^n) \leq g_x(x) \quad \text{for all } x \geq \rho_{i,j_\alpha^-}^n. \quad (8)$$

Since $\phi(\tau_n)$ is positive, $-(O_{i,j}^n + \phi(\tau_n))\alpha + P_{i,j}^n \geq 0$ implies that $-(O_{i,j}^n + \phi(\tau_n))\alpha + P_{i,j}^n + \phi(\tau_n) \geq 0$ for all i and j . It follows that that

$$j_\alpha^+ \leq j_\alpha^- \quad \text{and} \quad \rho_{i,j_\alpha^+}^n \leq \rho_{i,j_\alpha^-}^n.$$

The fact that

$$\begin{aligned} \hat{h}_n(\alpha, x) &= \inf_{y \in \mathbb{R}} g_x(y) \\ &= \min\{\inf_{y \geq x} g_x(y), \inf_{y < x} g_x(y)\}, \end{aligned}$$

Equations 5 and 7 and Inequalities 6 and 8 imply that

$$\hat{h}_n(\alpha, x) = \begin{cases} (1 - \alpha)\phi(\tau_n)(\rho_{i,j_\alpha^+}^n - x) + h_n(\alpha, \rho_{i,j_\alpha^+}^n) & \text{if } x < \rho_{i,j_\alpha^+}^n, \\ h_n(\alpha, x) & \text{if } \rho_{i,j_\alpha^+}^n \leq x < \rho_{i,j_\alpha^-}^n, \\ \alpha\phi(\tau_n)(x - \rho_{i,j_\alpha^-}^n) + h_n(\alpha, \rho_{i,j_\alpha^-}^n) & \text{if } x \geq \rho_{i,j_\alpha^-}^n. \end{cases} \quad (9)$$

We assume here that h_n satisfies the properties of the theorem. This first implies that for all α , the map $x \rightarrow \hat{h}_n(\alpha, x)$ is piecewise linear and continuous. Next, the map $x \rightarrow h_n(\alpha, x)$ is convex, in particular between $\rho_{i,j_\alpha^+}^n$ and $\rho_{i,j_\alpha^-}^n$. Moreover, definitions of j_α^+ and j_α^- ensure that if $j_\alpha^+ > 0$ then $-O_{i,j_\alpha^+}^n\alpha + P_{i,j_\alpha^+}^n \geq (1 - \alpha)\phi(\tau_n)$ and if $j_\alpha^- \leq z_i^n$ then $-O_{i,j_\alpha^- - 1}^n\alpha + P_{i,j_\alpha^- - 1}^n \leq \alpha\phi(\tau_n)$. It follows that the map $x \rightarrow \hat{f}_n(\alpha, x)$ is still convex. Moreover, if $j_\alpha^- \leq z_i^n$ then the x -coefficient of $\hat{h}_n$ for $x \geq \rho_{i,j_\alpha^-}^n$ is $\alpha\phi(\tau_n)$ which

is nonnegative. Otherwise, we have in particular that $-\hat{O}_{i,z_i^n}^n \alpha + \hat{P}_{i,z_i^n}^n = -O_{i,z_i^n}^n \alpha + P_{i,z_i^n}^n > \alpha \phi(\tau_n) \geq 0$, which implies that the x -coefficient of $\hat{h}_n$ for $x \in (\rho_{i,z_i^n}^n, \infty]$ is positive.

Let us put $\Phi_1, \dots, \Phi_p$ for the elements of

$$\left(\left\{ \frac{P_{i,j}^n + \phi(\tau_n)}{O_{i,j}^n + \phi(\tau_n)} \middle| 0 \leq j \leq z_i^n \right\} \cup \left\{ \frac{P_{i,j}^n}{O_{i,j}^n + \phi(\tau_n)} \middle| 0 \leq j \leq z_i^n \right\} \right) \cap (\Upsilon_i^n, \Upsilon_{i+1}^n)$$

indexed in increasing order. By construction, the indexes j_α^+ and j_α^- are both constant over all the sub-intervals $(\Upsilon_i^n, \Phi_1]$, $(\Phi_1, \Phi_2]$, $\dots$, $(\Phi_{p-1}, \Phi_p]$, $(\Phi_p, \Upsilon_{i+1}^n]$. The fact that h_n satisfies the properties of the theorem and Equation 9 imply that for all values x , the map $\alpha \rightarrow \hat{h}_n(\alpha, x)$ is linear over all the sub-intervals above. We get that the map $\hat{h}_n$ is piecewise linear and its continuity with regard to α is straightforward to verify at all bounds Φ_k for $1 \leq k \leq p$.

Last, Assertion 3 follows straightforwardly from Theorem 1 and from the fact that, for all children k of n , we have

$$h_k(\alpha, x) = \hat{h}_n(\alpha, x) + \sum_{c \in \mathcal{C}_n \setminus \{k\}} \hat{f}_c(\alpha, x), \quad (10)$$

which ends the proof. $\square$

Theorem 2 suggests the procedure sketched in Algorithm 1 for computing the functions h_n and $\hat{h}_n$ for all non-root nodes n .

Like the algorithm computing $(f_n)_{n \in \mathcal{T}}$ and $(\hat{f}_n)_{n \in \mathcal{T}}$ provided in [5], the complexity of Algorithm 1 depends on the total number of bounds (over asymmetry parameters and over character values) required for defining $(h_n)_{n \in \mathcal{T}}$ and $(\hat{h}_n)_{n \in \mathcal{T}}$. The number of bounds over character values, below referred to as x -bounds, is, by construction, smaller than $|\mathcal{K}|$ (we don't count the two extrema $-\infty$ and $+\infty$). We shall bound the number of intervals over asymmetry parameters, below referred to as α -bounds, following the same ideas as in [5].

From now on, we make the implicit assumption that all bounds of the partial cost functions are actually required to define it. Namely, under the notations of Theorem 2 and for all nodes n of $\mathcal{T}$, the cost function h_n is such that:

- for all $0 \leq i \leq w_n$ and all $0 \leq j < z_i^n$, we have that

$$(O_{i,j}^n, P_{i,j}^n, Q_{i,j}^n, R_{i,j}^n) \neq (O_{i,j+1}^n, P_{i,j+1}^n, Q_{i,j+1}^n, R_{i,j+1}^n);$$

- for all $0 \leq i < w_n$, there exists $0 \leq j \leq z_i^n$ and $0 \leq j' \leq z_{i+1}^n$ such that

$$(\rho_{i,j}^n, \rho_{i,j+1}^n) \cap (\rho_{i+1,j'}^n, \rho_{i+1,j'+1}^n) \neq \emptyset \text{ and } (O_{i,j}^n, P_{i,j}^n, Q_{i,j}^n, R_{i,j}^n) \neq (O_{i+1,j'}^n, P_{i+1,j'}^n, Q_{i+1,j'}^n, R_{i+1,j'}^n);$$

and that the same holds for the partial cost functions $\hat{h}_n$, f_n and $\hat{f}_n$.

Remark that since h_n is continuous, the first assertions implies that for all $0 \leq i \leq w_n$ and all $0 \leq j < z_i^n$, we have that

$$(O_{i,j}^n, P_{i,j}^n) \neq (O_{i,j+1}^n, P_{i,j+1}^n);$$

Lemma 1. *Let n be an unknown node of $\mathcal{T}$. For all $x \in \mathbb{R}$, the x -coefficient of h_n (resp. of $\hat{h}_n$) decreases with γ .*

Proof. We shall follow the same three steps as for the proof of Theorem 2 and the same ideas as for that of Lemma 1 from [5]. Let us first recall that Lemma 1 from [5] ensures that the x -coefficient of $\hat{h}_k$ decreases with α for all children k of the root different from n . Actually, a slightly different parametrization is used in [5] but the same change of variable as for Theorem 1 shows that this lemma remains true with the parametrization used here.

First, for all children n of the root of $\mathcal{T}$, Equation 4 and Lemma 1 from [5] ensure that the lemma holds for h_n .

Second, let us prove that if the lemma is true for h_n then it is true for $\hat{h}_n$. Let $\alpha \in [0, 1]$ be an asymmetry parameter and i be the index such that $\Upsilon_{i-1}^n < \alpha \leq \Upsilon_i^n$. From the proof of Theorem 2,

```

Function compute_h_and_h
  input : a tree  $\mathcal{T}$  of root  $r$ 
  /* Compute functions  $h_k$  and  $\hat{h}_k$  for all children  $k$  of the root */
  forall  $k \in \mathcal{C}_r$  do
     $h_k \leftarrow \sum_{c \in \mathcal{C}_r \setminus \{k\}} \hat{f}_c$ ;
    recursion_h_and_h( $k$ );

Function recursion_h_and_h
  input : a node  $n$  of  $\mathcal{T}$ 
  /* Call function to_stem on  $h_n$  for computing  $\hat{h}_n$ , use it for computing  $h_k$  for all children  $k$  of
   $n$  then recursively treat all children  $k$  */
   $\hat{h}_n \leftarrow \text{to\_stem}(h_n)$ ;
  forall  $k \in \mathcal{C}_n$  do
     $h_k \leftarrow \hat{h}_n + \sum_{c \in \mathcal{C}_n \setminus \{k\}} \hat{f}_c$ ;
    recursion_h_and_h( $k$ );

Function to_stem
  input : a piecewise linear function  $h_n$  given under notations of Theorem 2
  output: the piecewise linear function  $\hat{h}_n$  given under notations of Theorem 2
  /* Compute  $\hat{h}_n$  from  $h_n$  */
   $\hat{w}_n \leftarrow -1$ ;  $\hat{\Upsilon}_0^n \leftarrow \Upsilon_0^n$ ;
  for  $i \leftarrow 0$  to  $w_n$  do
     $j^+ \leftarrow \min \left\{ j \mid \frac{P_{i,j-1}^n + \phi(\tau_n)}{O_{i,j-1}^n + \phi(\tau_n)} > \Upsilon_i^n \right\}$ ;  $M_+ \leftarrow \max \left\{ j \mid \frac{P_{i,j-1}^n + \phi(\tau_n)}{O_{i,j-1}^n + \phi(\tau_n)} < \Upsilon_{i+1}^n \right\}$ ;
     $j^- \leftarrow \min \left\{ j \mid \frac{P_{i,j}^n}{O_{i,j}^n + \phi(\tau_n)} > \Upsilon_i^n \right\}$ ;  $M_- \leftarrow \max \left\{ j \mid \frac{P_{i,j}^n}{O_{i,j}^n + \phi(\tau_n)} < \Upsilon_{i+1}^n \right\}$ ;
    while  $j^- \leq M_-$  or  $j^+ \leq M_+$  do
       $\hat{w}_n \leftarrow \hat{w}_n + 1$ ;
      for  $\hat{\Upsilon}_{\hat{w}_n}^n < \gamma < \hat{\Upsilon}_{\hat{w}_n+1}^n$ , set
         $\hat{h}_n(\alpha, x) = \begin{cases} (1 - \alpha)\phi(\tau_n)(\rho_{i,j^+}^n - x) + h_n(\alpha, \rho_{i,j^+}^n) & \text{if } x < \rho_{i,j^+}^n, \\ h_n(\alpha, x) & \text{if } \rho_{i,j^+}^n \leq x < \rho_{i,j^-}^n, \\ \alpha\phi(\tau_n)(x - \rho_{i,j^-}^n) + h_n(\alpha, \rho_{i,j^-}^n) & \text{if } x \geq \rho_{i,j^-}^n. \end{cases}$ ;
      if  $j^- \leq M_-$  or  $j^+ \leq M_+$  then
         $\hat{\Upsilon}_{\hat{w}_n+1}^n \leftarrow \min \left\{ \frac{P_{i,j^+-1}^n + \phi(\tau_n)}{O_{i,j^+-1}^n + \phi(\tau_n)}, \frac{P_{i,j^-}^n}{O_{i,j^-}^n + \phi(\tau_n)} \right\}$ ;
        if  $\frac{P_{i,j^+-1}^n + \phi(\tau_n)}{O_{i,j^+-1}^n + \phi(\tau_n)} \leq \hat{\Upsilon}_{\hat{w}_n}^n$  then  $j^+ \leftarrow j^+ + 1$ ;
        if  $\frac{P_{i,j^-}^n}{O_{i,j^-}^n + \phi(\tau_n)} \leq \hat{\Upsilon}_{\hat{w}_n}^n$  then  $j^- \leftarrow j^- + 1$ ;
      else  $\hat{\Upsilon}_{\hat{w}_n+1}^n \leftarrow \Upsilon_{i+1}^n$ ;
    return  $\hat{h}_n$ 

```

Algorithm 1: Computation of h_n and $\hat{h}_n$ for all non-root nodes n of $\mathcal{T}$.

there exists an index $0 \leq j \leq z_i^n$ such that, by setting $\pi_\alpha^+ = \rho_{i,j}^n$, the x -coefficient of $\hat{h}_n(\alpha, x)$ is equal to $-(1 - \alpha)\phi(\tau_n)$ for $x < \pi_\alpha^+$ and greater than $-(1 - \alpha)\phi(\tau_n)$ otherwise (π_α^+ is possibly equal to $-\infty$).

Symmetrically, there exists an index $0 \leq j \leq z_i^n$ such that, by setting $\pi_\alpha^- = \rho_{i,j}^n$, the x -coefficient of $\hat{h}_n(\alpha, x)$ is equal to $\alpha\phi(\tau_n)$ for $x > \pi_\alpha^-$ and smaller than $\alpha\phi(\tau_n)$ otherwise.

By construction and from the induction hypothesis, we get that both π_α^+ and π_α^- increase with α .

Let us now consider two asymmetry parameters $0 \leq \alpha' \leq \alpha'' \leq 1$. We then have $\pi_{\alpha'}^+ \leq \pi_{\alpha''}^+$, $\pi_{\alpha'}^- \leq \pi_{\alpha''}^-$, $\pi_{\alpha'}^+ \leq \pi_{\alpha''}^+$ and $\pi_{\alpha'}^- \leq \pi_{\alpha''}^-$. The first two inequalities come from construction (see the proof of Theorem 2) and the two last ones from the remark just above. This leaves only two cases to investigate:

1. $\pi_{\alpha'}^+ \leq \pi_{\alpha'}^- \leq \pi_{\alpha''}^+ \leq \pi_{\alpha''}^-$ and
2. $\pi_{\alpha'}^+ \leq \pi_{\alpha''}^+ \leq \pi_{\alpha'}^- \leq \pi_{\alpha''}^-$.

By checking all the possible positions of x with regard to the bounds $\pi_{\alpha'}^+$, $\pi_{\alpha'}^-$, $\pi_{\alpha''}^+$ and $\pi_{\alpha''}^-$ like in the proof of Lemma 1 of [5], we verify that the x -coefficient of $\hat{h}_n(\alpha', x)$ is always greater than that of $\hat{h}_n(\alpha'', x)$ either directly or from the induction hypothesis both in Cases 1 and 2. The x -coefficient of $\hat{h}_n(\alpha, x)$ does decrease with α .

Third, we remark that if the x -coefficient of $\hat{h}_n(\alpha, x)$ decreases with α then, since the same holds for all children k of n from Lemma 1 of [5], Equation 10 implies that the x -coefficient of $\hat{h}_c(\alpha, x)$ decreases with α for all children c of n . $\square$

Proposition 1. *Under the notations of Theorem 2 and for all nodes n of $\mathcal{T}$, we have that*

$$\hat{w}_n - w_n \leq 2|\mathcal{K} \setminus \mathcal{K}_n|.$$

Proof. Let n be a node of $\mathcal{T}$ with $\hat{w}_n > w_n$. There exists an index ℓ in $\{1, \dots, \hat{w}_n\}$ such that $\hat{\Upsilon}_\ell^n \notin \{\Upsilon_k^n \mid 1 \leq k \leq w_n\}$. Let i be the index such that $\Upsilon_i^n < \hat{\Upsilon}_\ell^n < \Upsilon_{i+1}^n$. From the proof of Theorem 2, there exists an index $0 \leq j \leq z_i^n$ such that at least one of the following assertions holds:

1. $\hat{\Upsilon}_\ell^n = \frac{P_{i,j}^n + \phi(\tau_n)}{O_{i,j}^n + \phi(\tau_n)},$
2. $\hat{\Upsilon}_\ell^n = \frac{P_{i,j}^n}{O_{i,j}^n + \phi(\tau_n)}.$

In other words, there are two types of α -bounds of $\hat{h}_n$ which are not α -bounds of h_n .

If $\hat{\Upsilon}_\ell^n = \frac{P_{i,j}^n + \phi(\tau_n)}{O_{i,j}^n + \phi(\tau_n)}$ (Type 1) and still from the proof of Theorem 2, we have that $\rho_{\ell-1,1}^n = \rho_{i,j}^n < \rho_{i,j+1}^n = \rho_{\ell,1}^n$. Lemma 1 then ensures that for all $i' \geq i$ and all j' such that $(\rho_{i',j'}^n, \rho_{i',j'+1}^n] \cap (\rho_{i,j}^n, \rho_{i,j+1}^n] \neq \emptyset$, we have that $-O_{i',j'}^n\alpha' + P_{i',j'}^n \leq -O_{i,j}^n\alpha + P_{i,j}^n$ for all $\alpha' \in (\Upsilon_{i'}^n, \Upsilon_{i'+1}^n]$ and all $\alpha \in (\Upsilon_i^n, \Upsilon_{i+1}^n]$ with $\alpha' \geq \alpha$. It follows that, if $\rho_{i',j'}^n \leq \rho_{i,j}^n$ then $-(O_{i',j'}^n\alpha' - \phi(\tau_n)) + P_{i',j'}^n \leq 0$ for all $\alpha' \geq \hat{\Upsilon}_\ell^n$. This implies that $\rho_{\ell',1}^n \leq \rho_{\ell,1}^n$ for all $\ell' \geq \ell$. In plain English, each time that a α -bound $\hat{\Upsilon}_\ell^n$ of Type 1 appears, there is a x -bound which is no longer required for defining $\hat{\Upsilon}_\ell^n(\alpha, x)$ for all $\alpha > \hat{\Upsilon}_\ell^n$.

The situation is symmetrical if $\hat{\Upsilon}_\ell^n = \frac{P_{i,j}^n}{O_{i,j}^n + \phi(\tau_n)}$. Each time that a α -bound $\hat{\Upsilon}_\ell^n$ of Type 2 appears, there is a x -bound of h_n which is no longer required for defining $\hat{\Upsilon}_\ell^n(\alpha, x)$ for all $\alpha < \hat{\Upsilon}_\ell^n$.

Since all x -bounds used in the definition of h_n belong to $\mathcal{V}(\mathcal{K} \setminus \mathcal{K}_n)$ and can be involved at most once in the appearance of α -bounds of Type 1 and at most once in the appearance of α -bounds of Type 2, we get that the number of “new” α -bounds of $\hat{h}_n$ is smaller than $2|\mathcal{K} \setminus \mathcal{K}_n|$. $\square$

Corollary 1. *Under the notations of Theorem 2 and for all nodes n of $\mathcal{T}$, both w_n and $\hat{w}_n$ are smaller than $4|\mathcal{T}| \cdot |\mathcal{K}|$*

Proof. From Equation 4 and Theorem 3 from [5], if n is a child of the root then $w_n \leq 2|\mathcal{T}_n| \cdot |\mathcal{K} \setminus \mathcal{K}_n| \leq 2|\mathcal{T}| \cdot |\mathcal{K}|$. Let m is a non-root node of $\mathcal{T}$, there exists a path $k_1, k_2, \dots, k_\ell$ of $\mathcal{T}$ with $\ell < |\mathcal{T}|$ going from a child of the root k_0 to $k_\ell = m$. The number w_m of α -bounds of h_m is obtained by adding all the increases along the path $k_1, k_2, \dots, k_\ell$ to w_{k_1} :

$$w_m = w_{k_1} + \sum_{i=1}^{\ell-1} (\hat{w}_{k_i} - w_{k_i}) + \sum_{i=1}^{\ell-1} (w_{k_{i+1}} - \hat{w}_{k_i}).$$

From Equation 4 and Theorem 3 from [5] and since k_1 is a child of the root, we have

$$w_{k_1} \leq 2|\mathcal{T}_{k_1}| \cdot |\mathcal{K} \setminus \mathcal{K}_{k_1}|.$$

Proposition 1 ensures that

$$\hat{w}_{k_i} - w_{k_i} \leq 2|\mathcal{K} \setminus \mathcal{K}_{k_i}|, \quad \text{for all } 1 \leq i \leq \ell.$$

Equation 10 and Theorem 3 of [5] imply that

$$w_{k_{i+1}} - \hat{w}_{k_i} \leq \sum_{c \in \mathcal{C}_{k_i} \setminus \{k_{i+1}\}} 2|\mathcal{T}_c| \cdot |\mathcal{K}_c|, \quad \text{for all } 1 \leq i < \ell.$$

We get that

$$\begin{aligned} w_m &\leq 2|\mathcal{T}_{k_1}| \cdot |\mathcal{K} \setminus \mathcal{K}_{k_1}| + \sum_{i=1}^{\ell-1} 2|\mathcal{K} \setminus \mathcal{K}_{k_i}| + \sum_{i=1}^{\ell-1} \left(\sum_{c \in \mathcal{C}_{k_i} \setminus \{k_{i+1}\}} 2|\mathcal{T}_c| \cdot |\mathcal{K}_c| \right) \\ &\leq 2|\mathcal{T}| \left(|\mathcal{K} \setminus \mathcal{K}_{k_1}| + \sum_{i=1}^{\ell-1} \left(\sum_{c \in \mathcal{C}_{k_i} \setminus \{k_{i+1}\}} 2|\mathcal{T}_c| \cdot |\mathcal{K}_c| \right) \right) + \sum_{i=1}^{\ell-1} 2|\mathcal{K} \setminus \mathcal{K}_{k_i}| \\ &\leq 2|\mathcal{T}| \cdot |\mathcal{K} \setminus \mathcal{K}_m| + 2(\ell-1)|\mathcal{K}| \\ &\leq 4|\mathcal{T}| \cdot |\mathcal{K}|. \end{aligned}$$

Since $\hat{w}_m - w_m \leq 2|\mathcal{K} \setminus \mathcal{K}_m|$ (Proposition 1), we have that

$$\begin{aligned} \hat{w}_m &\leq 2|\mathcal{T}| \cdot |\mathcal{K} \setminus \mathcal{K}_m| + 2\ell|\mathcal{K}| \\ &\leq 4|\mathcal{T}| \cdot |\mathcal{K}|. \end{aligned}$$

□

Proposition 2. *Under the assumption that the number of children of any node is bounded independently of the size of the tree, the algorithmic complexity of the computation of the cost functions h_n and $\hat{h}_n$ for all nodes n of a tree $\mathcal{T}$ with a set of known nodes $\mathcal{K}$ is $O(|\mathcal{T}|^2 \cdot |\mathcal{K}|^2)$ both in time and memory space.*

Proof. From Theorem 1, less than $4|\mathcal{T}| \cdot |\mathcal{K}|$ α -bounds are required for defining $\hat{h}_n$. Over intervals between two successive α -bounds, $\hat{h}_n$ requires at most $|\mathcal{K}_n|$ x -bounds. It follows that the total number of bounds required for defining $\hat{h}_n$ is $O(|\mathcal{T}| \cdot |\mathcal{K}|^2)$. Since the part of Algorithm 1 computing $\hat{h}_n$ from h_n is linear with the total number of bounds required to define $\hat{h}_n$, its complexity is $O(|\mathcal{T}| \cdot |\mathcal{K}|^2)$ both in time and memory space.

The cost function h_n is obtained by summing the partial cost functions $\hat{h}_m$ of its direct ancestor m and the partial cost functions $\hat{f}_c$ of its siblings c , which is performed by using a procedure similar to that of merging sorted lists. Under the assumption that the number of children m is bounded independently of the size of the tree, this operation is linear with the sum of all the bounds of these cost functions cost functions which is $O(|\mathcal{T}| \cdot |\mathcal{K}|^2)$.

In conclusion, the computation of all the partial cost functions of $\mathcal{T}$ is $O(|\mathcal{T}|^2 \cdot |\mathcal{K}|^2)$ both in time and memory space. □

4 Integrating parsimonious costs over asymmetry parameter(s)

4.1 Single parameter case: no-split cost

We show here how to sum the parsimonious α -cost $\check{\Delta}_\alpha(\mathcal{T}, \vartheta)$ over all asymmetry parameters α . Let us start by recalling Remark 3 from [5].

Remark 1 ([5]). Let r be the root of $\mathcal{T}$ and $(\dot{\Gamma}_k^r)_{0 \leq k \leq w_r+1}$ be the elements of

$$\left(\bigcup_{1 \leq i \leq u_r} \left(\left\{ \frac{B_{i,j}^r}{A_{i,j}^r} \mid 0 \leq j \leq v_i^r \right\} \cap (\Gamma_i^r, \Gamma_{i+1}^r) \right) \right) \cup \{\Gamma_i^r \mid 0 \leq i \leq u_r + 1\},$$

indexed in increasing order. There exists a sequence $(\rho_k^r)_{0 \leq k \leq w_r}$ of increasing values in $\vartheta(\mathcal{K})$ such that for all $0 \leq k \leq w_r$ and all $\dot{\Gamma}_k^r \leq \alpha < \dot{\Gamma}_{k+1}^r$, a α -parsimonious reconstruction associates ρ_k^r to r . If $\dot{\Gamma}_k^r < \alpha < \dot{\Gamma}_{k+1}^r$ then all α -parsimonious reconstructions associate ρ_k^r to r .

An algorithm computing $(\dot{\Gamma}_k^r)_{0 \leq k \leq w_r+1}$ and $(\rho_k^r)_{0 \leq k \leq w_r}$ from f_r is provided in [5]. Its complexity is linear with the total number of bounds required for defining f_r .

From Theorem 1, we have that the most parsimonious α -cost is, for all $\dot{\Gamma}_k^r < \alpha < \dot{\Gamma}_{k+1}^r$,

$$\begin{aligned} \check{\Delta}_\alpha(\mathcal{T}, \vartheta) &= \min_x f_n(\alpha, x) \\ &= \left(-A_{\mathbf{i}(k), \mathbf{j}(k)}^r \alpha + B_{\mathbf{i}(k), \mathbf{j}(k)}^r \right) \rho_k^r + C_{\mathbf{i}(k), \mathbf{j}(k)}^r \alpha + D_{\mathbf{i}(k), \mathbf{j}(k)}^r \end{aligned}$$

where $\mathbf{i}(k)$ is the index i such that $(\dot{\Gamma}_k^r, \dot{\Gamma}_{k+1}^r] \subseteq (\Gamma_i^r, \Gamma_{i+1}^r]$ and $\mathbf{j}(k)$ is the index j such that $\rho_k^r \in (\kappa_{\mathbf{i}(k), \mathbf{j}}^n, \kappa_{\mathbf{i}(k), \mathbf{j}+1}^n]$.

Integrating the parsimonious cost of $(\mathcal{T}, \vartheta)$ over all possible values of the asymmetry parameter can then be done explicitly:

$$\begin{aligned} \int_0^1 \check{\Delta}_\alpha(\mathcal{T}, \vartheta) d\alpha &= \sum_{0 \leq k \leq w_r} \int_{\dot{\Gamma}_k^r}^{\dot{\Gamma}_{k+1}^r} \check{\Delta}_\alpha(\mathcal{T}, \vartheta) d\alpha \\ &= \sum_{0 \leq k \leq w_r} \left[\left(-A_{\mathbf{i}(k), \mathbf{j}(k)}^r \rho_k^r + C_{\mathbf{i}(k), \mathbf{j}(k)}^r \right) \left(\frac{(\dot{\Gamma}_{k+1}^r)^2 - (\dot{\Gamma}_k^r)^2}{2} \right) \right. \\ &\quad \left. + \left(B_{\mathbf{i}(k), \mathbf{j}(k)}^r \rho_k^r + D_{\mathbf{i}(k), \mathbf{j}(k)}^r \right) \left(\dot{\Gamma}_{k+1}^r - \dot{\Gamma}_k^r \right) \right] \end{aligned}$$

The quantity such obtained will be referred to as the *no-split cost* of $(\mathcal{T}, \vartheta)$.

4.2 Two parameters case: A- and B-split costs

In order to test a trend change at a node n of $\mathcal{T}$, we shall consider the two ways of splitting $\mathcal{T}$ at n displayed in Figure 1, namely, the most parsimonious costs obtained with

- (A) an asymmetry parameter α' on the subtree $\mathcal{T}_n$ and an asymmetry parameter α on the rest of $\mathcal{T}$ (i.e., with α' on the red part and with α on the blue part of Figure 1-left).
- (B) an asymmetry parameter α' on the subtree consisting of the direct ancestor of n the branch bearing n and $\mathcal{T}_n$ (i.e., the stem subtree of n) and an asymmetry parameter α on the rest of $\mathcal{T}$ (i.e., with α' on the red part and with α on the blue part of Figure 1-right);

In Case A, given two asymmetry parameters α and α' , we are able to compute the smallest cost which can be achieved by associating x to n with asymmetry parameters α' on $\mathcal{T}_n$ and α on the rest of $\mathcal{T}$. Namely, this cost is

$$\Omega_n^A(\alpha, \alpha', x) = \hat{h}_n(\alpha, x) + f_n(\alpha', x).$$

In the same way for Case B, the smallest cost which can be achieved by associating x to the direct ancestor of n and by using the asymmetry parameter α' on the stem subtree of n and α on the rest of $\mathcal{T}$ is given by

$$\Omega_n^B(\alpha, \alpha', x) = h_n(\alpha, x) + \hat{f}_n(\alpha', x).$$

We shall detail only Case A, Case B being very similar.

From Theorems 1 and 2, the map $\Omega_n^A : (\alpha, \alpha', x) \rightarrow \hat{h}_n(\alpha, x) + f_n(\alpha', x)$ is piecewise linear and continuous. Namely, under the notations of these theorems, for all $0 \leq i \leq \hat{w}_n$, all $0 \leq j \leq u_n$ and by putting $(\pi_{i,j,k}^n)_{0 \leq k \leq c_{i,j}^n}$ for the elements of $\{\hat{\rho}_{i,k}^n \mid 0 \leq k \leq \hat{z}_i^n\} \cup \{\kappa_{i,k}^n \mid 0 \leq k \leq v_j^n\}$ sorted in increasing order, we have that, for all $0 \leq i \leq \hat{w}_n$, all $\alpha \in (\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n]$, all $0 \leq j \leq u_n$, $\alpha' \in (\Gamma_j^n, \Gamma_{j+1}^n]$, all $0 \leq k \leq c_{i,j}^n$ and all $x \in (\pi_{i,j,k}^n, \pi_{i,j,k+1}^n]$,

$$\begin{aligned}\Omega_n^A(\alpha, \alpha', x) &= (-\hat{O}_{i,k_{\hat{h}}}^n \alpha + \hat{P}_{i,k_{\hat{h}}}^n)x + \hat{Q}_{i,k_{\hat{h}}}^n \alpha + \hat{R}_{i,k_{\hat{h}}}^n + (-A_{i,k_f}^n \alpha + B_{i,k_f}^n)x + C_{i,k_f}^n \alpha + D_{i,k_f}^n \\ &= (-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n)x + W_{i,j,k}^n \alpha + X_{i,j,k}^n \alpha' + Y_{i,j,k}^n,\end{aligned}$$

where $k_{\hat{h}}$ is the index such that $\pi_{i,j,k}^n = \hat{\rho}_{i,k_{\hat{h}}}^n$, k_f is the index such that $\pi_{i,j,k}^n = \kappa_{i,k_f}^n$, $T_{i,j,k}^n = \hat{O}_{i,k_{\hat{h}}}^n$, $U_{i,j,k}^n = A_{i,k_f}^n$, $V_{i,j,k}^n = \hat{P}_{i,k_{\hat{h}}}^n + B_{i,k_f}^n$, $W_{i,j,k}^n = \hat{Q}_{i,k_{\hat{h}}}^n$, $X_{i,j,k}^n = C_{i,k_f}^n$, $Y_{i,j,k}^n = \hat{R}_{i,k_{\hat{h}}}^n + D_{i,k_f}^n$.

Let us remark that, from Theorems 1 and 2, for all asymmetry parameters α and α' , the map $x \rightarrow \Omega_n^A(\alpha, \alpha', x)$ is convex.

The smallest cost obtained with asymmetry parameter α' on $\mathcal{T}_n$ and with asymmetry parameter α elsewhere is

$$\check{\Omega}_n^A(\alpha, \alpha') = \min_x \Omega_n^A(\alpha, \alpha', x).$$

Since Ω_n^A is convex with respect to x , for all $0 \leq i \leq \hat{w}_n$, all $\alpha \in (\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n]$, all $0 \leq j \leq u_n$ and all $\alpha' \in (\Gamma_j^n, \Gamma_{j+1}^n]$, we have that

$$\check{\Omega}_n^A(\alpha, \alpha') = (-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n) \pi_{i,j,\ell}^n + W_{i,j,\ell}^n \alpha + X_{i,j,\ell}^n \alpha' + Y_{i,j,\ell}^n,$$

where ℓ is the smallest index such that $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n \geq 0$. Such an index ℓ always exists since Theorems 1 and 2 imply that $-T_{i,j,c_{i,j}^n}^n \alpha - U_{i,j,c_{i,j}^n}^n \alpha' + V_{i,j,c_{i,j}^n}^n \geq 0$.

For all indexes $0 \leq i \leq \hat{w}_n$, $0 \leq j \leq u_n$, the lines of equation $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ with $0 \leq k \leq c_{i,j}^n$ partition $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$ into subsets on which the smallest index ℓ such that $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n \geq 0$ remains the same, i.e. in which $\check{\Omega}_n^A$ is linear with respect to α and α' (see Figure 2 without taking into account the dotted line).

Proposition 3. For all indexes $0 \leq i \leq \hat{w}_n$, $0 \leq j \leq u_n$, if there exists $0 \leq \ell \leq c_{i,j}^n$ such that $U_{i,j,\ell}^n \neq 0$ and such that the line of equation $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n$ intersects $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$ then the points (α, α') for which $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n \geq 0$ are those which are below the line of equation $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n = 0$.

Proof. Since from Theorem 1, $U_{i,j,k}^n$ is nonnegative for all $0 \leq k \leq c_{i,j}^n$, we have that $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n \geq 0$ if and only if $\alpha' \leq \frac{-T_{i,j,k}^n \alpha + V_{i,j,k}^n}{U_{i,j,k}^n}$, i.e., the point (α, α') is below the line of equation $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ (Figure 2). $\square$

Proposition 4. For all indexes $0 \leq i \leq \hat{w}_n$, $0 \leq j \leq u_n$ and $0 \leq k < k' \leq c_{i,j}^n$, the lines of equations $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ and $-T_{i,j,k'}^n \alpha - U_{i,j,k'}^n \alpha' + V_{i,j,k'}^n = 0$ do not intersect each other at any point in $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$.

Proof. Since $x \rightarrow \Omega_n^A(\alpha, \alpha', x)$ is convex, we have that $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n \leq -T_{i,j,k'}^n \alpha - U_{i,j,k'}^n \alpha' + V_{i,j,k'}^n$. It follows that $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n > 0$ implies $-T_{i,j,k'}^n \alpha - U_{i,j,k'}^n \alpha' + V_{i,j,k'}^n > 0$ for all $(\alpha, \alpha') \in (\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$.

From the implicit assumption that all the bounds of partial functions are required, if $k \neq k'$ then $(T_{i,j,k}^n, U_{i,j,k}^n, V_{i,j,k}^n) \neq (T_{i,j,k'}^n, U_{i,j,k'}^n, V_{i,j,k'}^n)$, i.e., two lines intersect each other at a single point. If the lines of equations $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ and $-T_{i,j,k'}^n \alpha - U_{i,j,k'}^n \alpha' + V_{i,j,k'}^n = 0$ intersect each other at a point of $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$, they split $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$ into four non-empty parts among which at least one contains points (α, α') such that $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n > 0$ and $-T_{i,j,k'}^n \alpha - U_{i,j,k'}^n \alpha' + V_{i,j,k'}^n < 0$ which is in contradiction with the convexity of Ω_n^A with respect to x since we assume $k < k'$. $\square$

Proposition 5. Let i and j be two indexes with $0 \leq i \leq \hat{w}_n$ and $0 \leq j \leq u_n$.

- a. If there exists $0 \leq \ell \leq c_{i,j}^n$ such that both $T_{i,j,\ell}^n = U_{i,j,\ell}^n = 0$ and $V_{i,j,\ell}^n \geq 0$ then there is no line of equation $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ with $k > \ell$ intersecting $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$.
- b. If there exists $0 \leq \ell \leq c_{i,j}^n$ such that $T_{i,j,\ell}^n \neq 0$ and $U_{i,j,\ell}^n = 0$ then
- if $T_{i,j,\ell}^n > 0$ and $\frac{V_{i,j,\ell}^n}{T_{i,j,\ell}^n} > \hat{\Upsilon}_i^n$ then there is no line of equation $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ with $k > \ell$ intersecting $(\hat{\Upsilon}_i^n, \min\{\hat{\Upsilon}_{i+1}^n, \frac{V_{i,j,\ell}^n}{T_{i,j,\ell}^n}\}] \times (\Gamma_j^n, \Gamma_{j+1}^n]$.
 - if $T_{i,j,\ell}^n < 0$ and $\frac{V_{i,j,\ell}^n}{T_{i,j,\ell}^n} < \hat{\Upsilon}_{i+1}^n$ then there is no line of equation $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ with $k > \ell$ intersecting $(\max\{\hat{\Upsilon}_i^n, \frac{V_{i,j,\ell}^n}{T_{i,j,\ell}^n}\}, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$.
- c. For all $0 \leq k < k' \leq c_{i,j}^n$, the part of the line corresponding to k' intersecting $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$ is above that of k on $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$.

Proof. These three properties are again consequences of the convexity of the map $x \rightarrow \Omega_n^A(\alpha, \alpha', x)$ for all $(\alpha, \alpha') \in (\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$, i.e., of the fact that the sequence $(-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n)_{0 \leq k \leq c_{i,j}^n}$ increases with k . In particular, this implies that if there exist a part $I \subseteq (\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$ and an index ℓ is such that $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n \geq 0$ for all $(\alpha, \alpha') \in I$ then there is no line of equation $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ with $k > \ell$ splitting I into two non-empty parts, since this would imply the existence of points $(\alpha, \alpha') \in I$ with $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n < 0$.

The three assertions are proved as follows.

- a. If $T_{i,j,\ell}^n = U_{i,j,\ell}^n = 0$ and $V_{i,j,\ell}^n \geq 0$ then we have that $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n \geq 0$ for all $(\alpha, \alpha') \in (\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$.
- b. Let us assume that $T_{i,j,\ell}^n \neq 0$ and $U_{i,j,\ell}^n = 0$.
- If $T_{i,j,\ell}^n > 0$ and $\frac{V_{i,j,\ell}^n}{T_{i,j,\ell}^n} > \hat{\Upsilon}_i^n$ then we have that $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n \geq 0$ for all $(\alpha, \alpha') \in (\hat{\Upsilon}_i^n, \min\{\hat{\Upsilon}_{i+1}^n, \frac{V_{i,j,\ell}^n}{T_{i,j,\ell}^n}\}] \times (\Gamma_j^n, \Gamma_{j+1}^n]$.
 - If $T_{i,j,\ell}^n < 0$ and $\frac{V_{i,j,\ell}^n}{T_{i,j,\ell}^n} < \hat{\Upsilon}_{i+1}^n$ then we have that $-T_{i,j,\ell}^n \alpha - U_{i,j,\ell}^n \alpha' + V_{i,j,\ell}^n \geq 0$ for all $(\alpha, \alpha') \in (\max\{\hat{\Upsilon}_i^n, \frac{V_{i,j,\ell}^n}{T_{i,j,\ell}^n}\}, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$.
- c. The property directly follows from Proposition 3 and from the opening remark of this proof. $\square$

The A -split cost of $(\mathcal{T}, \vartheta)$ at n is the integral of $\check{\Omega}_n^A$ over all pairs $(\alpha, \alpha') \in [0, 1]^2$:

$$\int_0^1 \int_0^1 \check{\Omega}_n^A(\alpha, \alpha') d\alpha' d\alpha = \sum_{i=0}^{\hat{w}_n} \sum_{j=0}^{u_n} \int_{\hat{\Upsilon}_i^n}^{\hat{\Upsilon}_{i+1}^n} \int_{\Gamma_j^n}^{\Gamma_{j+1}^n} \check{\Omega}_n^A(\alpha, \alpha') d\alpha' d\alpha.$$

For all $0 \leq i \leq \hat{w}_n$ and all $0 \leq j \leq u_n$, computing

$$\int_{\hat{\Upsilon}_i^n}^{\hat{\Upsilon}_{i+1}^n} \int_{\Gamma_j^n}^{\Gamma_{j+1}^n} \check{\Omega}_n^A(\alpha, \alpha') d\alpha' d\alpha$$

can be performed by splitting $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$ into trapezoids of the form $\{(\alpha, \alpha') \mid \beta < \alpha \leq \omega \text{ and } a\alpha + b < \alpha' \leq c\alpha + d\}$ (i.e., with two vertical sides) in which the coefficients of $\check{\Omega}_n^A$ are constant and of the form:

$$\check{\Omega}_n^A(\alpha, \alpha') = (-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n) \kappa_{i,j,k}^n + W_{i,j,k}^n \alpha + X_{i,j,k}^n \alpha' + Y_{i,j,k}^n,$$

for a certain index k . The different lines of Figure 2, here including the dotted one, partition $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$ into such trapezoids (the triangle below the line of equation $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ is seen as a degenerated trapezoid with a left side of length 0).

Figure 2: Partition of $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$ resulting from the lines of equations $-T_{i,j,k+1}^n \alpha - U_{i,j,k+1}^n \alpha' + V_{i,j,k+1}^n = 0$ and $-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n = 0$ (the top and the bottom solid lines respectively).

The algorithm partitioning $(\hat{\Upsilon}_i^n, \hat{\Upsilon}_{i+1}^n] \times (\Gamma_j^n, \Gamma_{j+1}^n]$ into trapezoids is not presented here since it is quite basic thanks to Propositions 3, 4 and 5 which limit the number of situations that we have to deal with.

The integral of $\check{\Omega}_n^A$ over a trapezoid $\{(\alpha, \alpha') \mid \beta < \alpha \leq \omega \text{ and } a\alpha + b < \alpha' \leq c\alpha + d\}$ can be explicitly computed:

$$\begin{aligned} \int_{\beta}^{\omega} \int_{a\alpha+b}^{c\alpha+d} \check{\Omega}_n^A(\alpha, \alpha') d\alpha' d\alpha &= \int_{\beta}^{\omega} \int_{a\alpha+b}^{c\alpha+d} ((-T_{i,j,k}^n \alpha - U_{i,j,k}^n \alpha' + V_{i,j,k}^n) \kappa_{i,j,k}^n + W_{i,j,k}^n \alpha + X_{i,j,k}^n \alpha' \\ &\quad + Y_{i,j,k}^n) d\alpha' d\alpha \\ &= \frac{1}{2} \left[\frac{\omega^3 - \beta^3}{3} (2(-T_{i,j,k}^n \kappa_{i,j,k}^n + W_{i,j,k}^n)(c-a) + (-U_{i,j,k}^n \kappa_{i,j,k}^n + X_{i,j,k}^n)(c^2 - a^2)) \right. \\ &\quad + (\omega^2 - \beta^2) ((-T_{i,j,k}^n \kappa_{i,j,k}^n + W_{i,j,k}^n)(d-b) + (V_{i,j,k}^n \kappa_{i,j,k}^n + Y_{i,j,k}^n)(c-a) \\ &\quad + (-U_{i,j,k}^n \kappa_{i,j,k}^n + X_{i,j,k}^n)(cd - ab)) \\ &\quad \left. + (\omega - \beta) (2(V_{i,j,k}^n \kappa_{i,j,k}^n + Y_{i,j,k}^n)(d-b) + (-U_{i,j,k}^n \kappa_{i,j,k}^n + X_{i,j,k}^n)(d^2 - b^2)) \right] \end{aligned}$$

Symmetrically, by setting

$$\check{\Omega}_n^B(\alpha, \alpha') = \min_x \Omega_n^B(\alpha, \alpha', x),$$

we define the *B-split cost of $(\mathcal{T}, \vartheta)$ at n* as

$$\int_0^1 \int_0^1 \check{\Omega}_n^B(\alpha, \alpha') d\alpha' d\alpha = \sum_{i=0}^{w_n} \sum_{j=0}^{u_n} \int_{\Gamma_i^n}^{\Gamma_{i+1}^n} \int_{\hat{\Gamma}_j^n}^{\hat{\Gamma}_{j+1}^n} \check{\Omega}_n^B(\alpha, \alpha') d\alpha' d\alpha.$$

All the remarks stated above about A-split costs still hold for B-split costs which are computed in the very same way.

5 Applications

For all examples and datasets below, the parsimonious splits identification was performed by setting $\phi(\tau) = \frac{1}{\tau}$ for computing the evolutionary costs.

5.1 Two small examples

Figure 3: A tree with two sets of character values (in brackets) for tips. Left: A A-split at node ‘n’ improves the parsimonious cost of the whole tree. Right: no split improve the parsimonious cost of the whole tree (see Table 1).

node	A-split cost	B-split cost
r	0.32	-
m	0.32	0.40
n	0.31	0.35
a	-	0.34
b	-	0.34
c	-	0.35
p	0.34	0.40
d	-	0.35
e	-	0.35

node	A-split cost	B-split cost
r	0.50	-
m	0.52	0.59
n	0.54	0.59
a	-	0.62
b	-	0.62
c	-	0.57
p	0.54	0.59
d	-	0.53
e	-	0.53

Table 1: Left: (resp. Right:) A- and B-split costs for all nodes of the tree with character values of Figure 3-left (resp. of Figure 3-right). The first line of the tables displays the A-split cost of the root ‘r’, i.e., the no-split cost computed as presented in Section 4.1.

Figure 3 and Table 1 illustrate the fact that considering more than one asymmetry parameter on a tree may or may not lead to lower the parsimonious cost. The two examples displayed in Figure 3 are based on the same tree but with two different initial functions (in brackets just after idents of the tips). Table 1 displays the A- and B-split costs for all nodes of the tree with regard to the character values of Figure 3.

In the example at the left of the figure, the A-split cost at node ‘n’ is smaller than the no-split cost (i.e., the A-split at the root ‘r’ – first line of Table 1-left). Conversely, in the example at the right, there is no A- or B-split cost lower than the no-split cost.

From a general point of view, it may be worth paying interest not only on the most parsimonious split but also on the other splits with costs smaller than the no-split cost.

5.2 Two biological datasets

Figures 4 and 5 display all the splits that lead to a cost lower than the no-split one. Splits are represented with ‘x’ at the corresponding nodes. Figures also display the ranks of the splits with regard to their costs (ordered from lowest to highest). The color of the subtree pending from a split depends on the corresponding split cost, more precisely on the cost of the “most recent” split including it. The color

scale starts from blue which corresponds to the no-split cost or higher to red for the lowest cost observed. These two figures were output by our software *ParSplit*.

For the two datasets studied here, all the most parsimonious splits are A-splits. Moreover, no B-split costs were smaller than the no-split cost. This is not a general behavior since we did encounter cases in which some of the B-split costs were lower than the no-split cost and even smaller than the A-split costs at the same nodes.

We observe in Figure 4 that the split with the lowest cost occurs at the most recent common ancestor of the mysticetes (i.e., *Balaenidae*, *Neobalaenidae* and *Balaenopteridae*). First, this split is consistent with a probable shift in diet type, since *mysticetes* are filter feeders. Second, authors of [18] also distinguished this clade and found that “The largest phylogenetic mean body size was recovered for *Mysticeti*...” by using approaches based on heterogeneous stochastic models from [21, 22].

Figure 5 displays the parsimonious splits of the tree of *Anolis* lizards according to the body size of males. This dataset comes from [20] in which authors used two methods for identifying rate shifts both based on stochastic models. With the first one, they detected two shifts at the same nodes as those of our two splits with the smallest costs, ranked in the same order as with our approach. With their second approach, they identified a third shift at the same node as that of the split ranked 5 with our method for which the costs of split of ranks 3, 4 and 5 are very close.

5.3 Discussion

It is worth noting that, as observed on the two biological datasets above, a parsimony-based approach may identify the evolutionary shifts detected from methods based on stochastic models of evolution. In all the situations where parsimony is preferred, our approach provides a parsimonious solution for identifying evolutionary shifts.

We emphasize the fact that the different splits displayed in Figures 4 and 5 have to be considered as different alternatives for splitting the tree into two parts and absolutely not as a split of the tree into more than two parts. Finding the most parsimonious split of the tree into $k > 2$ parts in the same sense as above could be performed with ideas similar as those developed in this work. The algorithmic complexity increases exponentially with k but computations would be still feasible for small k (e.g. smaller than 5 for usual phylogenetic trees). In particular, one could seek for the most parsimonious cost obtained with a same asymmetry parameter on k subtrees and another parameter for the rest of the tree, following the ideas of [10] for studying convergence with Ornstein-Uhlenbeck processes.

References

- [1] P. Bastide, M. Mariadassou, and S. Robin. Detection of adaptive shifts on phylogenies by using shifted stochastic processes on a tree. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1067–1093, 2017.
- [2] M. A. Butler and J. B. Losos. Testing for unequal amounts of evolution in a continuous character on different branches of a phylogenetic tree using linear and squared-change parsimony: An example using lesser antillean *ANOLIS* lizards. *Evolution*, 51(5):1623–1635, 1997.
- [3] C. E. Cressler, M. A. Butler, and A. A. King. Detecting adaptive evolution in phylogenetic comparative analysis using the Ornstein-Uhlenbeck model. *Systematic Biology*, 64(6):953–968, 2015.
- [4] M. Csűrös. Ancestral Reconstruction by Asymmetric Wagner Parsimony over Continuous Characters and Squared Parsimony over Distributions. In C. Nelson and S. Vialette, editors, *Comparative Genomics*, volume 5267 of *Lecture Notes in Computer Science*, pages 72–86. Springer Berlin / Heidelberg, 2008.
- [5] G. Didier. Time-dependent-asymmetric-linear-parsimonious ancestral state reconstruction. *Bulletin of Mathematical Biology*, Aug 2017.
- [6] J. M. Eastman, M. E. Alfaro, P. Joyce, A. L. Hipp, and L. J. Harmon. A novel comparative method for identifying shifts in the rate of character evolution on trees. *Evolution*, 65(12):3578–3589, 2011.
- [7] J. S. Farris. Methods for Computing Wagner Trees. *Systematic Biology*, 19(1):83–92, 1970.

- [8] J. S. Farris. Parsimony and explanatory power. *Cladistics*, 24(5):825–847, 2008.
- [9] J. A. Fuentes-G., E. A. Housworth, A. Weber, and E. P. Martins. Phylogenetic ancova: Estimating changes in evolutionary rates as well as relationships between traits. *The American Naturalist*, 188(6):615–627, 2016.
- [10] T. Ingram and D. Mahler. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein-Uhlenbeck models with stepwise Akaike Information Criterion. *Methods in Ecology and Evolution*, 4(5):416–425, 2013.
- [11] M. Khabbazian, R. Kriebel, K. Rohe, and C. An. Fast and accurate detection of evolutionary shifts in Ornstein-Uhlenbeck models. *Methods in Ecology and Evolution*, 7(7):811–824, 2016.
- [12] B. Kolaczkowski and J. W. Thornton. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. *Nature*, 431(7011):980–984, Oct 2004.
- [13] M. J. Landis, J. G. Schraiber, and M. Liang. Phylogenetic analysis using Lévy processes: Finding jumps in the evolution of continuous traits. *Systematic Biology*, 62(2):193–204, 2013.
- [14] B. C. O’Meara, C. Ané, M. J. Sanderson, and P. C. Wainwright. Testing for different rates of continuous trait evolution using likelihood. *Evolution*, 60(5):922–933, 2006.
- [15] L. Revell. On the analysis of evolutionary change along single branches in a phylogeny. *The American Naturalist*, 172(1):140–147, 2008. PMID: 18554139.
- [16] L. J. Revell and D. C. Collar. Phylogenetic analysis of the evolutionary correlation using likelihood. *Evolution*, 63(4):1090–1100, 2009.
- [17] G. J. Slater. Phylogenetic evidence for a shift in the mode of mammalian body size evolution at the cretaceous-palaeogene boundary. *Methods in Ecology and Evolution*, 4(8):734–744, 2013.
- [18] G. J. Slater, S. A. Price, F. Santini, and M. E. Alfaro. Diversity versus disparity and the radiation of modern cetaceans. *Proceedings of the Royal Society of London B: Biological Sciences*, 2010.
- [19] M. Steel and D. Penny. Parsimony, likelihood, and the role of models in molecular phylogenetics. *Molecular Biology and Evolution*, 17(6):839–850, 2000.
- [20] G. H. Thomas and R. P. Freckleton. MOTMOT: models of trait macroevolution on trees. *Methods in Ecology and Evolution*, 3(1):145–151, 2012.
- [21] G. H. Thomas, R. P. Freckleton, and T. Székely. Comparative analyses of the influence of developmental mode on phenotypic diversification rates in shorebirds. *Proceedings of the Royal Society of London B: Biological Sciences*, 273(1594):1619–1624, 2006.
- [22] G. H. Thomas, S. Meiri, and A. B. Phillimore. Body size diversification in anolis: Novel environment and island effects. *Evolution*, 63(8):2017–2030, 2009.

Figure 4: Parsimonious splits of the tree of cetaceans according to their body sizes with their ranks. All the parsimonious splits are A-splits. Dataset from [18]

Figure 5: Parsimonious splits of the tree of *Anolis* lizards according to their body sizes with their ranks. All the parsimonious splits are A-splits. Dataset from [20]