

Rémy's diffusion on Brownian trees

Nicolas CURIEN[♦] & Cyril MARZOUK[♥]

Abstract

Rémy's algorithm is a famous recursive construction of uniform random binary trees of growing size by a local grafting operation. In this work we construct a continuous version, a new local diffusion on the space of real trees of growing Brownian Continuum Random Trees (CRT's). It appears as the scaling limit of a variant of Rémy's algorithm due to Bacher, Bodini, and Jacquot. Once the trees are rescaled to have constant mass, this diffusion uncovers an ergodic dynamics on trees with the Brownian CRT as unique invariant law.

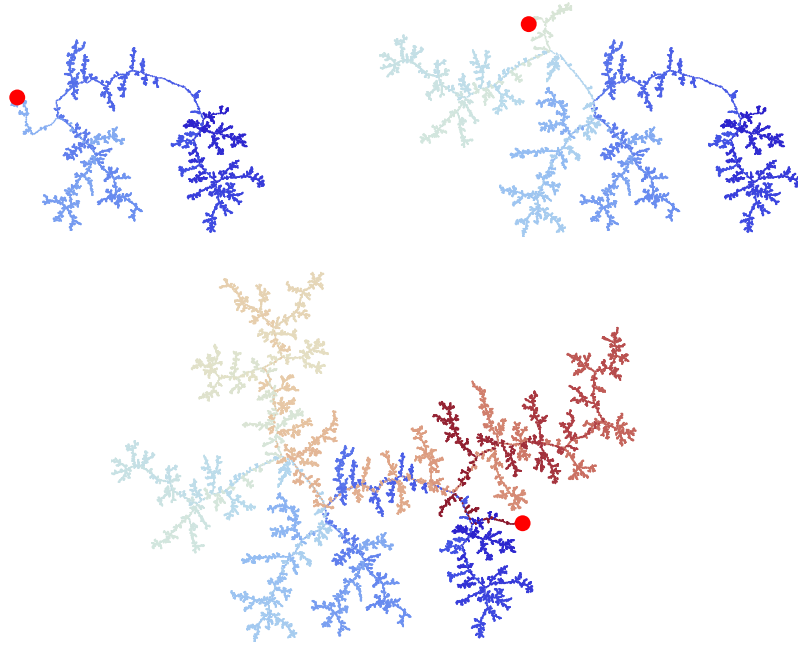


Figure 1: Snapshot at times 10k, 20k, and 40k of the Rémy diffusion. Colours are ranging from dark blue to red as new vertices are created. The red point is the current marked point where the growth continues. *Click here to see a video.*

[♦]Institut de Mathématique d'Orsay, Université Paris-Saclay, 91400 Orsay, France.
 nicolas.curien@universite-paris-saclay.fr

[♥]CMAP, CNRS, École polytechnique, Institut Polytechnique de Paris, Palaiseau, France.
 cyril.marzouk@polytechnique.edu

Contents

1	Introduction	2
1.1	One-step Rémy algorithm & Bacher–Bodini–Jacquot projection	2
1.2	Continuum tree growth	4
2	Preliminaries on the discrete dynamics	7
2.1	On the tree and resampling steps	8
2.2	The dynamics along the spine between two resampling steps	9
2.3	Height of the marked vertex	11
2.4	Stretching and gluing between two resampling steps	13
3	Scaling limit of the tree dynamics	15
3.1	Preliminaries on continuum trees	15
3.2	Gluing and stretching	17
3.3	Scaling limit of the discrete dynamics	20
4	Rémy’s diffusion is ergodic	23
4.1	A large Brownian subtree	24
4.2	Proof of ergodicity	29
	References	29

1 Introduction

This work studies a variant of the celebrated Rémy algorithm to construct a Markov chain of marked binary trees with growing size. Its scaling limit gives rise to a novel diffusion on marked continuum trees, which we call RÉMY’S GROWING DIFFUSION, and whose unique stationary distribution is the Brownian CRT of Aldous. In this introduction, we first describe precisely the discrete dynamics, then, more informally, the continuum one, before stating our main results.

1.1 One-step Rémy algorithm & Bacher–Bodini–Jacquot projection

Throughout this work, all discrete trees that we consider will be rooted plane binary trees (each vertex has 0 or 2 children) and we shall simply call them **binary trees**. The **size** n of the tree will be its number of internal vertices (it then has $n + 1$ leaves). We shall often distinguish a vertex, in which case the tree, and this vertex, are said to be **marked**. Rémy’s algorithm [Ré85] is a recursive construction of random binary trees based on the following one-step growth: Given a binary plane tree T_n with size n and a marked vertex X_n that can be either an internal vertex or a leaf, one constructs a random tree T_{n+1} by replacing in T_n the vertex X_n by a new internal vertex with two offspring: one is X_n and the other one is a new leaf, which is placed either to the left or to the right of X_n with probability $1/2$ each, see Figure 2 left.

It is straightforward to check that if T_n is uniformly distributed on binary trees with size n and if, conditionally on T_n , the vertex X_n is uniformly distributed over the $2n + 1$ possible vertices, then T_{n+1} has the uniform distribution on binary trees with size $n + 1$. A famous theorem originally due to Aldous [Ald93] (see also [LGo5] for this version) states that such a random binary tree T_n

with size n satisfies the scaling limit:

$$\frac{1}{\sqrt{2n}} \cdot T_n \xrightarrow[n \rightarrow \infty]{(d)} \mathcal{T}_{2e}, \quad (1.1)$$

in the Gromov–Hausdorff–Prokhorov sense where the limit is Aldous’ original Brownian CRT, coded by **twice** the standard Brownian excursion e ; this excursion is informally Brownian motion on $[0, 1]$ conditioned to stay nonnegative and end at 0. The goal of this work is to define the analogue of the discrete Rémy algorithm on Brownian trees obtained heuristically by “grafting and stretching” locally in $\mathcal{T}_{2e}$ as in the discrete Figure 2 left. We shall do so by taking appropriate scaling limit of the discrete Rémy algorithm.

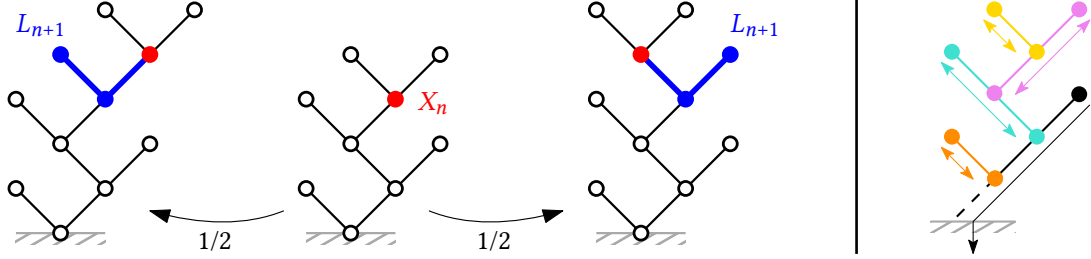


Figure 2: The two steps of the BBJ dynamics. Left: we replace the marked red vertex X_n in T_n by a blue cherry and choose the side of the new leaf L_{n+1} with probability $1/2$. Right: the bijective projection of a leaf on the internal vertices, except the rightmost leaf whose projection “exits the tree”, which is used to determine the next marked vertex.

The Bacher–Bodini–Jacquot coupled version

In the original **Rémy algorithm**, the one-step growth as in Figure 2 left is iterated at each step n by choosing independently of the rest the marked vertex X_n uniformly at random in T_n and performing the growth procedure at this vertex to get T_{n+1} . Unfortunately, the Markov chain $(T_n)_{n \geq 0}$ that is produces converges *almost surely* after suitable normalisation towards a Brownian CRT, see [Pit06, Exercises 7.4.10 and 7.4.11] or [CH13, Theorem 5], so this chain *cannot* be used to define a continuous version.

However, one can also couple the growing points X_n and X_{n+1} while preserving the crucial property that X_n is a uniform random vertex of T_n at each time n . In [BBJ17], **Bacher, Bodini, and Jacquot** introduced such a coupling in such a way that in some sense X_{n+1} is as close to X_n as possible. Their work was motivated by the fact that this coupling is nearly optimal in terms of randomness, for it requires $\sim 2n$ bits to generate a random tree of size n , which best possible up to lower order corrections, see [BBJ17]. Our motivation is that this coupling makes the growth steps accumulate locally and this will provide a nontrivial scaling limit. Let us now describe this coupling, which is represented in Figure 2 right.

Since a binary tree of size n has $n + 1$ leaves, it is possible to associate in a bijective way every leaf but one with an internal vertex: Every vertex of the tree is either a right vertex or a left vertex, the root being a right vertex by convention. Then define the **projection** of a leaf on the internal vertices as its last ancestor that lies on the opposite side. This indeed defines a bijection between the leaves, except the rightmost one, and the internal vertices, that one can easily invert: send each left internal vertex to the rightmost leaf among its descendants and vice versa. From the convention that the root is a right vertex, it is the projection of the leftmost leaf; however the rightmost leaf that has no projection: we say that the projection “exits the tree”.

We will study in this work the following two-step algorithm, which will be henceforth referred to as the **BBJ dynamics**. First initialise the process with $T_0^\bullet := (T_0, X_0)$ being a single marked vertex (a leaf) and then for any $n \geq 0$, build $T_{n+1}^\bullet := (T_{n+1}, X_{n+1})$ from $T_n^\bullet := (T_n, X_n)$, as follows:

- First build T_{n+1} by applying the one-step Rémy growth as in Figure 2 left: replace the marked vertex X_n in $T_n^\bullet$ by a new branchpoint with two offspring: X_n with its entire progeny and a new leaf L_{n+1} , placed either on the left or on the right independently with probability $1/2$;
- Then mark this tree at X_{n+1} chosen as follows: again independently, with probability $1/2$ each, either set $X_{n+1} = L_{n+1}$, or let X_{n+1} be the projection of L_{n+1} on the internal vertices of T_{n+1} as in Figure 2 right; since the rightmost leaf of T_{n+1} has no projection, then in this case we sample X_{n+1} uniformly at random and independently among all vertices of T_{n+1} .

The last case where the new leaf is the rightmost one and we need to sample the marked vertex uniformly at random will be henceforth called **resampling steps**. The chain $(T_n, X_n)_{n \geq 0}$ produced this way has the property that at any given time $n \geq 0$ the marked tree $T_n^\bullet = (T_n, X_n)$ has the uniform distribution on its size (see Lemma 2.3 below). Figure 3 displays a possible example of the first steps of such a sequence of random marked trees.

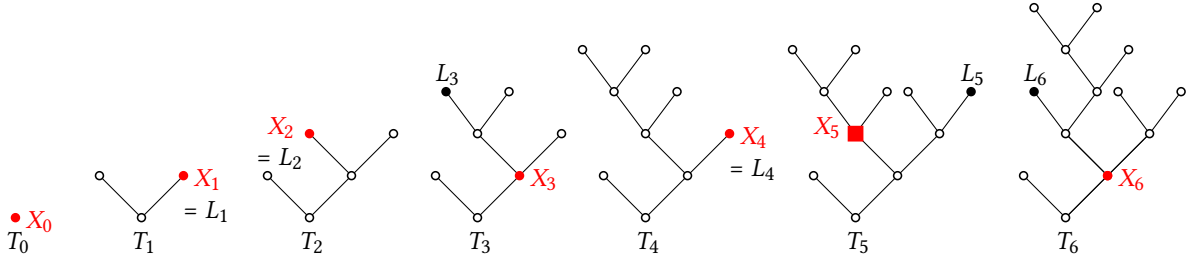


Figure 3: A few steps of the BBJ dynamics. The current marked vertex X_n is shown in red and the newly added leaf L_n in black, unless it coincides with X_n . The fifth step is a resampling step.

An important tool in our study will be the process $(S_n)_{n \geq 0}$ recording the height (i.e. the generation) of the marked vertex in the successive trees. Its scaling limit will enable us to build the continuum analogue of the discrete tree dynamics that we shall now describe.

1.2 Continuum tree growth

The **Rémy continuous diffusion** will be obtained by taking the scaling limit of the discrete algorithm above. Let us describe informally this continuum random dynamics $(\mathcal{T}_t^\bullet)_{t \geq 0}$ and refer to Section 3 for details. This dynamics takes value in the space $\mathbb{T}^\bullet$ of compact real trees¹ $(\mathcal{T}, d)$ equipped with a finite measure μ , a root ρ , and an extra marked point x . This space is endowed with the marked Gromov–Hausdorff–Prokhorov topology, see Section 3.1.

We introduce a gluing–stretching procedure Ψ that takes as an input such a marked real tree $\mathcal{T}^\bullet = (\mathcal{T}, d, \mu, \rho, x) \in \mathbb{T}^\bullet$ and a continuous function $g : [0, \infty) \rightarrow \mathbb{R}$ such that $g(0) = 0$ and returns a *random* càdlàg function $\Psi(\mathcal{T}^\bullet, g) = (\mathcal{T}_t^g)_{t \geq 0}$ with values in $\mathbb{T}^\bullet$, with $\mathcal{T}_0^g = \mathcal{T}^\bullet$. It can be heuristically described as follows, see also Figure 4 for an illustration. We start from the point x in $\mathcal{T}$ and follow the dynamics of a point whose height (distance to the root) variation is given by g :

¹To help the reader we shall always use bold fonts to denote measured rooted trees, and use the symbol $\bullet$ when they have an additional marked point.

- When g decreases, we go down the tree towards the root and, doing so, we *stretch* the distances along the ancestral line by a factor of 2;
- When g increases, we *glue* new pieces of the tree given by the contour of g .

In particular, when g completes an excursion above its running infimum, this corresponds to the creation of a subtree coded by **twice** that excursion in the usual sense, which is glued on the tree $\mathcal{T}^\bullet$ on the ancestral line of its marked point x , at a distance from x given by the negative of the minimum of g at this time. Notice that locally the effect is similar to the one-step Rémy algorithm of Figure 2 left: we stretch the ancestral line and graft new branches.

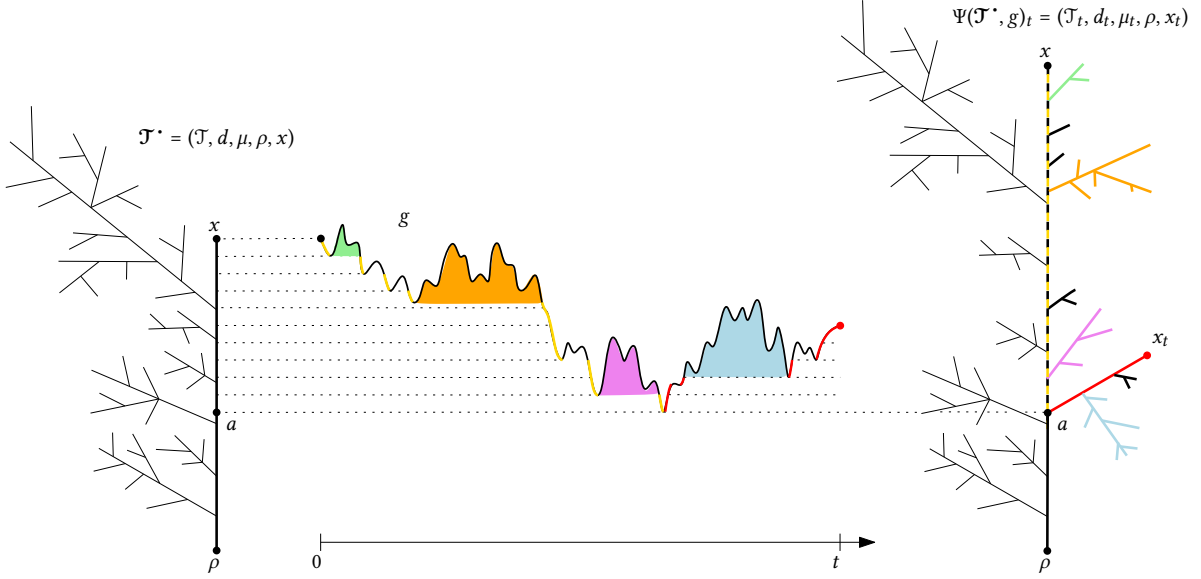


Figure 4: Illustration of the application $\Psi(\mathcal{T}^\bullet, g)$ that produces the tree on the right at time t . All the sub-excursions of g above its running minimum create new subtrees (here in different colours) coded by twice the corresponding excursion, and the instants of a new minimum double the distances along the ancestral line of the point x (the yellow parts). The new tree $\Psi(\mathcal{T}^\bullet, g)_t$ is marked at x_t , corresponding to the tip of g at time t , while the root ρ is left unchanged.

Let $\text{Ht}(x) = d(\rho, x)$ denote the height of x in the initial tree $(\mathcal{T}, d)$. We shall prove in Section 3 that, up to time

$$\xi = \inf\{t > 0 : g(t) < -\text{Ht}(x)\},$$

this construction is well-defined and provides a function $(\Psi(\mathcal{T}^\bullet, g)_t)_{t \in [0, \xi]}$ which is *deterministic and continuous* in t and which depends in a Lipschitz-continuous way in both the initial tree $\mathcal{T}^\bullet$ and the function g . In addition, it enjoys the semi-group property: for every $s, t \geq 0$ such that $t + s < \xi$, letting $g|_{[u, u+v]} : w \in [0, v] \mapsto g(u + w) - g(u)$, we have:

$$\Psi(\mathcal{T}^\bullet, g)_{t+s} = \Psi(\Psi(\mathcal{T}^\bullet, g)_s, g|_{[s, s+t]})_t = \Psi(\Psi(\mathcal{T}^\bullet, g)_t, g|_{[t, s+t]})_s.$$

This time ξ is the analogue of the resampling events in the discrete dynamics, when we exit below the root; accordingly, we must now resample the marked point randomly in the current tree. This is where both discontinuity and randomness enter the definition of Ψ . To do this, recall that the initial tree $\mathcal{T}^\bullet$ is endowed with a finite measure μ ; the latter evolves then with g to provide a finite measure on $\mathcal{T}_{\xi-}^g$ that we use to resample the marked point at time ξ . After resampling at this

time $\xi = \xi_1$, we then continue the construction as above with the remaining part of the function g up to the next time ξ_2 this situation occurs again, etc. We shall see that these resampling events do not accumulate (Lemma 3.2), so the process is well-defined for every time $t > 0$.

Consider as initial tree a single point $\mathbf{o}$ and as driving function g a standard Brownian motion $B = (B_t)_{t \geq 0}$. Then $\xi = 0$ and the previous construction $\Psi(\mathbf{o}, B)$ is ill-defined near $t = 0$, but the existence of such a process will be proved in the following theorem by approximations with the BBJ dynamics. For every $n \geq 0$ we let $T_n^\bullet$ denote the measured rooted and marked tree obtained by equipping the $2n + 1$ vertices of the discrete marked binary tree $T_n^\bullet$ with the graph distance and the counting measure. In the rest of the paper, if $\mathcal{T}^\bullet = (\mathcal{T}, d, \mu, \rho, x)$ is any measured rooted marked tree and $\alpha > 0$ any positive number, we denote by $\alpha \cdot \mathcal{T}^\bullet$ the tree

$$\alpha \cdot \mathcal{T}^\bullet = (\mathcal{T}, \alpha d, \alpha^2 \mu, \rho, x).$$

Theorem 1.1 (R  my's diffusion as the scaling limit of the Bacher–Bodini–Jacquot chain). *The convergence in distribution*

$$\left(\frac{1}{\sqrt{2n}} \cdot T_{[nt]}^\bullet \right)_{t \geq 0} \xrightarrow[n \rightarrow \infty]{(d)} (\mathcal{T}_t^\bullet)_{t \geq 0},$$

holds for the Skorokhod topology on $\mathbb{D}(\mathbb{R}_+, \mathbb{T}^\bullet)$. The limit process is such that for every $t > s > 0$, the tree $\mathcal{T}_t^\bullet$ has the same law as $\Psi(\mathcal{T}_s^\bullet, B|_{[s,t]})$ where B is an independent standard Brownian motion.

The process $(\mathcal{T}_t)_{t \geq 0}$ obtained from $(\mathcal{T}_t^\bullet)_{t \geq 0}$ by forgetting the marked point is continuous: the only source of discontinuities are the discrete resampling instants of the marked point when it collides with the root. Since T_n is uniformly distributed for every n , then Aldous' invariance principle (1.1) shows that for each $t > 0$, the tree $\mathcal{T}_t$ has the law of the Brownian CRT with mass t .

A consequence of Theorem 1.1 and the convergence of trees (1.1) is that if $\mathcal{T}_{2\varepsilon}^\bullet = (\mathcal{T}_{2\varepsilon}, X)$ is a Brownian CRT of mass 1 with an independent random marked point $X \in \mathcal{T}_{2\varepsilon}$ sampled from its mass measure, and if B is an independent Brownian motion then for any $\varepsilon > 0$,

$$\Psi(\mathcal{T}_{2\varepsilon}^\bullet, B)_\varepsilon \text{ has the law of a marked Brownian CRT of mass } 1 + \varepsilon. \quad (1.2)$$

We consider this fact as the infinitesimal version of the continuous R  my diffusion on Brownian CRT's. Although it may have a direct proof in the setting of continuum trees (but we do not know how to do it), we deduce it here from the discrete setting by passing to the limit.

In particular, by Brownian scaling, the rescaled dynamics $\mathcal{T}^\bullet \mapsto \frac{1}{\sqrt{1+t}} \cdot \Psi(\mathcal{T}^\bullet, B)_t$ leaves invariant the law of the marked Brownian CRT $\mathcal{T}_{2\varepsilon}^\bullet = (\mathcal{T}_{2\varepsilon}, X)$. Even though the law of processes started from two distinct initial conditions may remain mutually singular at every finite time, the rescaled dynamics exhibits an **asymptotic coupling** in the sense of [HMS11]: the two processes can be coupled so that their distance converges to zero. This yields in particular the following unique ergodicity property.

Theorem 1.2 (Unique ergodicity). *Let $\mathcal{T}^\bullet$ be any rooted real tree equipped with a mass measure with total mass 1 and a marked point different from the root. Let B be an independent Brownian motion. Then $\frac{1}{\sqrt{1+t}} \cdot \Psi(\mathcal{T}^\bullet, B)_t$ converges in distribution towards $\mathcal{T}_{2\varepsilon}^\bullet = (\mathcal{T}_{2\varepsilon}, X)$ as $t \rightarrow \infty$ in the marked Gromov–Hausdorff–Prokhorov topology.*

Diffusions on Brownian CRT's

Several Markovian dynamics on real trees with the Brownian CRT as invariant measure can be found in the literature, which, interestingly, all arise as scaling limits of discrete tree-valued Markov chains, but are constructed and analysed through different lenses.

The **prune-and-regraft** dynamics was introduced by Evans, Pitman, & Winter [EPWo6] as a scaling limit of the Aldous–Broder algorithm, see also [EWo6]. The so-called **Aldous’ diffusion** has been constructed rigorously by Forman, Pal, Rizzolo, & Winkel in the culmination of a series of work [FPRW23]; see also the work of Lohr, Mytnik, & Winter [LMW20] in a weaker topology. This diffusion is the scaling of the discrete chains on cladograms which consists heuristically to perform Rémy’s one-step move back and forth at i.i.d. locations in the tree. Another diffusion can be obtained by projecting Zambotti’s dynamics [Zam01] on paths through the Duquesne–Le Gall coding of trees. This dynamics is the scaling of a **corner flip dynamics** as proved in [EL15].

Arguably, our new dynamic is simpler to construct and has the advantage of being completely local. A companion paper [CCFA26] introduces yet another such local dynamics, with the strong property of being monotone, but whose growth has the disadvantage of evolving in a càdlàg fashion (on trees, even with no marked point). Our work gives a great latitude to define a growth procedure on continuum trees and we plan to study in the future a continuous version of the present one.

More broadly, understanding Markovian dynamics whose invariant distribution is a canonical probability measure $\mathcal{L}$ is of fundamental importance, as such dynamics can provide valuable information about $\mathcal{L}$ itself, including monotonicity properties, fixed-point characterisations, and tools for proving convergence to equilibrium. In the Gaussian setting, classical examples include Langevin dynamics and the Markovian interpretation of Stein’s method. In the context of Yang–Mills theory, the stochastic quantisation program initiated by Parisi & Wu seeks to construct and study a difficult probability measure $\mathcal{L}$ through a comparatively tractable Markovian dynamics having $\mathcal{L}$ as its invariant measure; see [CCHS22] for a recent breakthrough in this direction.

In the more modest setting of real trees, one of our aims is to *develop an infinitesimal, or ‘generator-like’, approach to Markovian dynamics on the space of real trees*. More precisely, we seek a way to characterise nice Markov processes through their infinitesimal action in the vein of (1.2). Such an approach might provide a common framework for comparing the various dynamics described above.

Acknowledgments: We thank Alessandra Caraceni, Serte Donderwinkel, William Fleurat, Robin Stephenson, and Adrianus Twigt for stimulating discussions. The first author is supported by “SuperGrandMa”, the ERC Consolidator Grant No 101087572.

2 Preliminaries on the discrete dynamics

In this section we study the discrete dynamics $(T_n, X_n)_{n \geq 0}$ defined in the Introduction using the Bacher–Bodini–Jacquot coupling of Rémy’s algorithm. Recall in particular the definition of resampling events when exiting the tree from the rightmost leaf.

First, the fact that for each $n \geq 0$ the random marked tree $T_n^* = (T_n, X_n)$ is uniformly distributed is a direct consequence of the following two lemmas whose proofs are clear from Figure 2.

Lemma 2.1. *The one-step Rémy growth operation of Figure 2 left is a bijection between:*

- *binary trees with size n marked at a vertex together with a coin flip, i.e. triples (t_n, x_n, ϵ_n) where x_n is any vertex of t_n and $\epsilon_n \in \{\text{left}, \text{right}\}$,*
- *binary trees with size $n + 1$ marked at a leaf, i.e. pairs (t_{n+1}, ℓ_{n+1}) where ℓ_{n+1} is a leaf of t_{n+1} .*

Lemma 2.2. *Let t_n be a binary tree with size n and L_n a uniform random leaf of t_n . Independently and with probability $1/2$ each, let either $X_n = L_n$ or let X_n be the projection of L_n on t_n as in Figure 2 right; if L_n is the rightmost leaf, then let X_n be an independent uniform random vertex of t_n instead in the second case. Then X_n is uniformly distributed among the $2n + 1$ vertices of t_n .*

In the BBJ chain $(T_n^\bullet)_{n \geq 0}$, the set of vertices of T_{n+1} is naturally identified to those of T_n together with an additional internal vertex and the leaf L_{n+1} , hence if $x \in T_n$, we will abuse notation and still write $x \in T_{n+1}$ for the corresponding vertex in T_{n+1} . Beware, if $x, y \in T_n$, their distance in T_{n+1} might differ from that in T_n (it may only increase, though).

2.1 On the tree and resampling steps

The combination of the two previous lemmas shows that the marked tree produced by the algorithm at time n has the uniform distribution. Quite surprisingly, this marked tree is independent of the instants of the resampling steps that have occurred so far, which are also independent of each other. As a consequence, not only is (T_n, X_n) uniformly distributed for any fixed n , but this is also the case at the resampling instants.

For every $k \geq 1$, let $I_k = 1$ if the k 'th step is a resampling step and $I_k = 0$ otherwise.

Lemma 2.3 (Independence of the final tree and resampling steps). *For every $n \geq 1$, the variables $(T_n^\bullet, I_1, \dots, I_n)$ are independent, each I_k has the Bernoulli distribution with parameter $1/(2k + 2)$ respectively, and $T_n^\bullet = (T_n, X_n)$ has the uniform distribution on marked binary trees with size n .*

Proof. The key observation is that the final marked tree (t_n, x_n) and the indices of resampling steps determine uniquely the entire sequence $(t_1, x_1), \dots, (t_n, x_n)$. Indeed, let us consider the dynamics run backward. For any $k = n, \dots, 1$, if the k 'th step is not a resampling step, then given (t_k, x_k) one can recover the intermediate marked leaf of t_k : it is the right-most leaf among the descendants of x_k if the latter is a left child and vice versa. If the k 'th step is a resampling step, then this means that the marked leaf was the rightmost leaf of t_k . In each case one thus recovers the marked leaf and we can then invert the growth procedure to recover (t_{k-1}, x_{k-1}) .

At each step, one makes a uniform random choice among 4 possibilities and independently of the rest, and in case of resampling at step j , one extra choice of a vertex with $2j + 1$ total possibilities is needed. Then the probability to produce (t_n, x_n) and to resample exactly at times in a given subset $J \subset \{1, \dots, n\}$, or equivalently to produce the unique corresponding sequence $(t_1, x_1), \dots, (t_n, x_n)$ with resampling precisely at the steps with index in J , equals

$$4^{-n} \prod_{j \in J} \frac{1}{2j + 1}.$$

Let $\text{Cat}(n) = \frac{1}{n+1} \binom{2n}{n}$ denote the n 'th Catalan numbers, which counts the number of rooted binary plane trees with size n . After some simple manipulations, we obtain:

$$4^{-n} \prod_{j \in J} \frac{1}{2j + 1} = \frac{1}{(2n + 1)\text{Cat}(n)} \cdot \prod_{j \in J} \frac{1}{2j + 2} \cdot \prod_{j \in J^c} \left(1 - \frac{1}{2j + 2}\right),$$

which proves our claim. $\square$

Let us mention that the algorithm originally designed in [BBJ17] actually consisted, when exiting the tree from the rightmost leaf, to start all over from a single vertex, again and again,

until for the first time one reaches a binary tree with size n without any exit. As argued there, this is (nearly) optimal in terms of random number generation, informally because resampling steps mostly occur when the tree is small. Compared to the present dynamics, this is equivalent to conditioning not to require resampling before time n and the previous lemma shows that this indeed produces a uniform random binary tree with size n . Note that the same holds if one conditions instead the dynamics to require resampling for the first time precisely at time n .

In these pages, we shall only use the algorithm with resampling a uniform random vertex when exiting the tree. The reason comes from the observation that resampling at time n , although independent of the tree at time n , is however not independent of the tree at time $n - 1$. In particular, if one conditions the dynamics to not require resampling before time n , then Lemma 2.3 does show that T_n has the uniform distribution, however T_k for $k < n$ does *not* have the uniform distribution and in the scaling limit, this does *not* yield Brownian CRT's before time 1.

2.2 The dynamics along the spine between two resampling steps

Let us now describe precisely how the marked tree is transformed between two consecutive resampling steps. Equivalently, instead of starting the dynamics from a single vertex, let it start from an arbitrary marked binary tree (t, x) , and let us run it until the first resampling step, at time $\Xi = \inf\{k \geq 1 : I_k = 1\}$.

The statements and proofs of this section should become intuitive once one has a clear view of how the dynamics work. A moment of thought leads to the following description, which is represented in Figure 5. At the first step, the vertex x is replaced by a new internal vertex; then x and its progeny is pushed up by one generation, and a new leaf is created next to it. Suppose that this leaf is placed on opposite side of its parent and that we keep this leaf as the new marked vertex (each has a probability $1/2$ of occurring). Then after that, we create an entire new subtree rooted at this leaf before exiting it; at this moment, the mark lands precisely where x was at time 0. If at the first step we do not keep the mark on the leaf but instead we project on its parent, then the only difference is that the new subtree rooted at the leaf is reduced to just that leaf.

After that, the dynamics then continues in the same way: We replace the current marked vertex by a new internal one, move everything above it up by one generation, and create a new leaf. If this leaf is placed on opposite side of its parent again, then again we next create an entire new subtree before reaching back again the original position of x .

Hence, after a random geometric number, possibly zero, of subtrees created on opposite side of x , we replace the current vertex sitting at the original position of x by a new internal vertex and place the leaf on the *same side* as its parent. Now again we create after that an entire subtree rooted at that leaf (possibly reduced to just that leaf) before exiting it. The difference with the previous case is that, when exiting it this subtree, we do not land on the original position of x , but on its last ancestor which lies on opposite side, which we call a *turning vertex*.

After that, the same story takes place at this turning vertex: we first create a random geometric number, possibly zero, of subtrees on the opposite side (in red in Figure 5), then exactly one subtree on the same side (in green in Figure 5), then move down to the next turning vertex, etc. until we leave the entire tree, at time Ξ . Notice that a no time before Ξ does the dynamics enter the subtrees that are attached on the ancestral line of x in the original tree (t, x) , and thus leaves them completely intact.

Let us record for every $n \geq 0$:

S_n the height of X_n in T_n .

We are going to see that this sequence gives a precise description of the process $(T_n, X_n)_{n \geq 0}$, similar to the role played by the function g in the continuous dynamics informally described in the Introduction. For the moment, from the previous description, we see that for every time $0 \leq k < \Xi$,

$$X_k \text{ is an ancestor of } x \text{ or } X_k = x \iff S_k = \min_{[0,k]} S.$$

In addition, this ancestor X_k at a weak negative record of S is either x or a turning vertex along its ancestral line. Precisely, it is x whenever $S_k = \min_{[0,k]} S = S_0$, and it is the i 'th turning vertex (from top to bottom) when S has made a total of i strict negative records by time k . See again Figure 5 for a pictorial description.

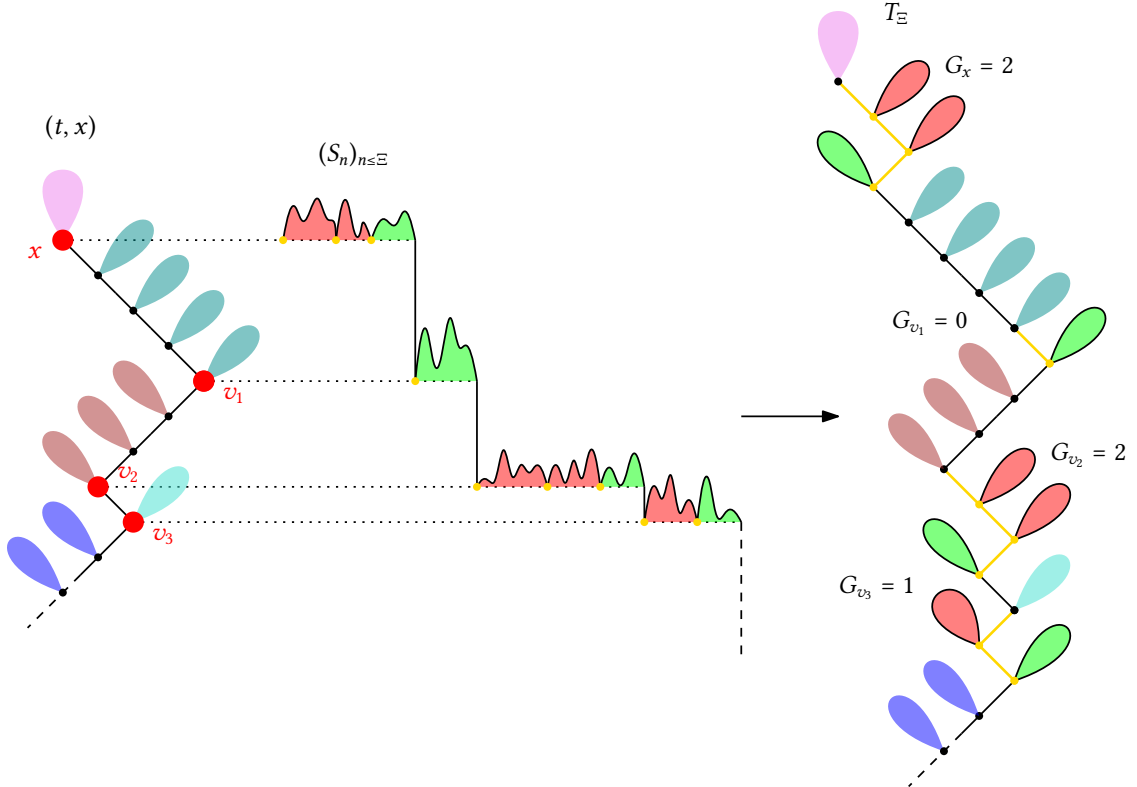


Figure 5: Left: the initial marked tree (t, x) with the turning vertices on its ancestral lines in large red. Middle: the height $(S_n)_n$ of the marked vertex in the dynamics. In gold are shown the instants where it sits at its running minimum. Right: the tree T_Ξ created by the dynamics at the first resampling step Ξ . It has only visited the vertex x and the turning vertices, where it has created subtrees coded by the excursions of S above its minimum. All coloured trees on the left are unchanged in the right picture.

In addition to being a natural quantity, let us motivate the study of $(S_n)_{n \geq 0}$ with the following observation: at each step $j \geq 1$, the dynamics creates a new internal vertex and a leaf, and the internal vertex replaces X_{j-1} so the height of this internal vertex, when created, is S_{j-1} , while the new leaf is placed at height $S_{j-1} + 1$. We shall see in a second step how these heights evolve in the subsequent trees $(T_{n+j})_{n \geq 1}$.

Before that, let us label these new vertices: $2j$ refers to the internal one and $2j + 1$ the leaf, both created at step j . In this way, if the initial tree t is composed of a single vertex labelled 1, then this provides a labelling of all the vertices of T_n , from 1 to $2n + 1$, that is consistent as n varies, and where internal vertices receive an even label and leaves receive an odd label. This labelling

corresponds to the colour scheme in the simulations from Figure 1. The previous description of the ancestral lines yields the following result, for all initial trees (t, x) .

Lemma 2.4. *Let $1 \leq i < j \leq n$. On the event that $\Xi \geq n$, the most recent common ancestor of the internal vertices $2i$ and $2j$ in the tree T_n is the internal vertex $2(i \wedge_n j)$, where*

$$i \wedge_n j = \sup\{k \in \{i, \dots, j\} : S_{k-1} = \min_{[i-1, k-1]} S\}.$$

Proof. Recall that at step i , the vertex $2i$ replaces the previous marked vertex X_{i-1} , at height S_{i-1} . At time $(i \wedge_n j) - 1$, since S is at its minimum since time $i - 1$, then the marked vertex $X_{(i \wedge_n j)-1}$ is an ancestor of $2i$. Then at the next step we replace this marked vertex by $2(i \wedge_n j)$, and up to time j , the path S remains strictly above this value, meaning that it creates a subtree rooted at $2(i \wedge_n j)$. This shows our claim when $n = j$. Between time j and Ξ , the dynamics never modifies the subtrees attached along the ancestral line of j , so the vertex $2(i \wedge_n j)$ will remain the most recent common ancestor of $2i$ and $2j$. $\square$

2.3 Height of the marked vertex

Let us now describe the law of the sequence $(S_n)_{n \geq 0}$ recording the height of the successive marked vertices, when again we let the dynamics start from an arbitrary marked binary tree (t, x) . The construction of the algorithm suggests that it does not evolve as a random walk since downward steps, in the case of projection, depend on the current tree structure. However besides this, it does exhibit independent increments.

Definition 2.5 (coding the ancestral lineage). Let (t, x) be a binary tree with a marked vertex at height $h \geq 1$. For every $i = 0, \dots, h - 1$, let $a_i = 0$ if the ancestor of x at height $h - i$ lies on the same side as its parent, and $a_i = 1$ if it lies on the opposite side. Let $A = \sum_{i=0}^{h-1} a_i$. For every $1 \leq k \leq A$, let $p_k = 1 + \inf\{j \geq 0 : \sum_{i=0}^j a_i \geq k\}$ and then $\ell_k = p_k - p_{k-1}$, where $p_0 = 0$. Let finally $\ell_{A+1} = h + 1 - p_A$.

Since the root lies on the right by convention, the sequence $(a_i)_{0 \leq i \leq h-1}$ allows to recover the left/right position of all the ancestors of x . Those with $a_i = 1$ are the children of the turning vertices of Figure 5 from top to bottom. The quantities $(\ell_k)_{1 \leq k \leq A+1}$ count the length of the intervals between these turning vertices, if we count x and the root as turning vertices.

Let us introduce some notation for the next result. Let G be a positive geometric random variable, whose law is $\mathbb{P}(G = k) = 2^{-k}$ for every $k \geq 1$. Given the tree (t, x) , let W denote a random walk with step distribution $2 - G \in \mathbb{Z}_{\leq 1}$ started at $W_0 = h$, the height of x in t . Then modify this path W to create a path W^x as follows: for every $1 \leq k \leq A + 1$, change the value at the k 'th time n such that $W_n < \inf_{[0, n-1]} W$ and instead set the value of W^x at this time to be equal to $h - p_k$. All the other increments are unchanged. The path W^x then stops after its $A + 1$ 'st strict negative record, by reaching the value $h - p_1 - \ell_{A+1} = -1$ for the first time.

Let us run the dynamics starting from (t, x) and until the first resampling step Ξ , but without performing the resampling of the vertex at the end and simply declare $S_\Xi = -1$.

Proposition 2.6 (The law of height process). *The path $(S_n)_{n \leq \Xi}$ has the same law as W^x .*

Proof. Let us imagine that the a_i 's are unknown. When performing the first one-step Rémy growth starting from the tree (t, x) , the new leaf is placed one generation above x ; let us not reveal right

away if it is placed on the same side as x or not but instead append this information to the previous unknown vector. Note that this entry is 0 or 1 with probability $1/2$ and is independent of the rest.

With probability $1/2$ we keep this leaf as the next marked vertex, in which case S increases by 1, whatever the side of this leaf and the unknown vector. With a probability $1/2$ we project the mark from this leaf and now we start revealing our unknown vector, from the leaf we just added to the root, until we find an entry equal to 1. In this case S makes a nonpositive increment, whose value equals 1 minus the number of entries we revealed before the first 1. Notice that when we stop and find the ancestor that will be the next marked vertex, we have not revealed the a_i 's corresponding to this vertex and its ancestors in our vector. We may then iterate this argument.

The description of the law of S follows since, during the construction, the 0's and 1's that we add are i.i.d. Bernoulli distributed with probability $1/2$ and independent of the rest, but the original a_i 's are deterministic, and this corresponds to the modification W^x described above. $\square$

Let us deduce an invariance principle from this description. Informally, we say that a vertex is *good* if it admits the same law of large numbers along its spine as in a uniform random spine.

Definition 2.7 (Good vertices). Let us fix once and for all a value $\mathfrak{s} \in (0, 1/6)$. Let (t, x) be a binary tree with a marked vertex at height $h \geq 1$. We say that (t, x) is *n-good* if

$$h \leq n^{\mathfrak{s}+1/2} \quad \text{and} \quad \sup_{t \in [0,1]} \left| \frac{h}{2} t - \sum_{i=1}^{\lfloor ht \rfloor} a_i \right| \leq n^{\mathfrak{s}+1/4},$$

where the a_i 's are as in Definition 2.5. A non-marked binary tree t is *n-good* when (t, x) is *n-good* for all $x \in t$.

This property will be in force for all points considered in our dynamics.

Lemma 2.8 (All good). *The following events occur with a probability tending to 1 as $n \rightarrow \infty$:*

- (i) $T_0^\bullet, \dots, T_n^\bullet$ are all *n-good* when $T_0^\bullet$ is a single vertex;
- (ii) $T_0^\bullet, \dots, T_{\Xi-1}^\bullet$ are all $\max(n, \Xi)$ -good when $T_0^\bullet = (t, x)$ is *n-good*.
- (iii) T_Ξ is $\max(n, \Xi)$ -good when the non-marked tree T_0 is *n-good*.

Proof. (i) Fix $k \leq n$; we know that $T_k^\bullet$ is uniformly distributed, so the a_i 's corresponding to X_k are i.i.d. Bernoulli random variables with parameter $1/2$. We can then apply a large deviation result for the binomial random walk $(\text{Bin}_n)_{n \geq 1}$, namely for every $N \geq 1$ and $t \geq 0$,

$$\mathbb{P}(\max_{1 \leq i \leq N} |\text{Bin}_i - i/2| \geq Nt) \leq 2e^{-2Nt^2},$$

see e.g. [McD98] at the very end of Section 2 there. In addition, the height H_k of X_k divided by $\sqrt{k}$ has sub-Gaussian tails (for binary trees, this can be traced back to [FGOR93, Theorem 1.3]) so the probability that X_k is not *n-good* decays as the exponential of some positive power of n . We conclude from a union bound.

- (ii) Fix $k < \Xi$. The marked vertex X_k in $T_k^\bullet$ belongs to a subtree made only of vertices created since T_0 (in either green or red in Figure 5) and that is grafted on the ancestral line of X_0 in T_0 , at a turning vertex V_k (or X_k is this turning vertex). The ancestral line of V_k has not yet changed since time 0 and is therefore *n-good*. As for the new subtree grafted above V_k and

marked at X_k , the previous argument shows that it is n_k -good where $n_k \leq \Xi$ is its number of internal vertices. More precisely, conditionally on n_k , this subtree has the law of $T_{n_k}^\bullet$ started from a single vertex and conditioned of the first resampling instant being greater than n_k . According to Lemma 2.3, this tree is uniformly distributed so the argument above applies.

- (iii) Recall that $T_\Xi^\bullet$ is marked at an independent uniform random vertex. Due to the second item, if the initial marked vertex is n -good and the random marked vertex at time Ξ is either of its ancestors or belongs to one of the new subtrees that we created, then $T_\Xi^\bullet$ is $\max(n, \Xi)$ -good. If all vertices of T_0 are n -good, then $T_\Xi^\bullet$ is indeed $\max(n, \Xi)$ -good. $\square$

In the next proposition, for every $n \geq 1$, we let the dynamics start from a marked binary tree $(t^{(n)}, x^{(n)})$, not necessarily with n internal vertices, and where $x^{(n)}$ has height h_n . Let $\Xi^{(n)}$ denote the first resampling time and $(S_k^{(n)})_{k < \Xi^{(n)}}$ record the height of the marked vertex before the first resampling step, started from $S_0^{(n)} = h_n$. For every $h > 0$, let B^h denote a Brownian motion started from $B_0^h = h$ and let $\xi^h = \inf\{t > 0 : B_t^h = 0\}$ denote the first hitting time of 0.

Proposition 2.9 (Quenched Brownian limit). *Suppose that the trees $(t^{(n)}, x^{(n)})$ are all n -good and that*

$$\frac{h_n}{\sqrt{2n}} \xrightarrow{n \rightarrow \infty} h > 0.$$

Then the following convergence in distribution holds:

$$\left(\frac{\Xi^{(n)}}{n}, \left(\frac{1}{\sqrt{2n}} S_{\lfloor nt \rfloor \wedge (\Xi^{(n)} - 1)}^{(n)} \right)_{t \geq 0} \right) \xrightarrow[n \rightarrow \infty]{(d)} (\xi^h, (B_{t \wedge \xi^h}^h)_{t \geq 0}). \quad (2.1)$$

Proof. If the $a_i^{(n)}$'s were random i.i.d. Bernoulli with parameter $1/2$ independent of the rest, then according to Proposition 2.6, the path $S^{(n)}$ would just be a random walk started from $W_0 = h_n$ and with step distribution $2 - G$, stopped when hitting $\mathbb{Z}_{<0}$ and the convergence (2.1) would then follow since $2 - G$ has mean 0 and variance 2. In addition, in this case our assumption on the $a_i^{(n)}$'s holds by the Law of Large Numbers.

Now to derive a quenched version when the $a_i^{(n)}$'s are given, Proposition 2.6 shows that the comparison with this random walk holds for the excursions of S above its past minimum. The $a_i^{(n)}$'s only encode the negative overshoots $S_k^{(n)} - \min_{[0, k-1]} S^{(n)}$ at the instants of negative strict records. Precisely, the j 'th such non-zero quantity is the length $\ell_j^{(n)}$ of the interval between two 1's in the sequence $(a_i^{(n)})_{i=0, \dots, h_n-1}$. The n -good assumption implies

$$\sup_{t \in [0, 1]} \frac{1}{h_n} \left| S_{\lfloor nt \rfloor \wedge \Xi^{(n)}}^{(n)} - \tilde{S}_{\lfloor nt \rfloor \wedge \tilde{\xi}^{(n)}}^{(n)} \right| \xrightarrow[n \rightarrow \infty]{\mathbb{P}} 0,$$

where $\tilde{S}^{(n)}$ is obtained from $S^{(n)}$ by replacing the deterministic $a_i^{(n)}$'s by random i.i.d. Bernoulli with parameter $1/2$ independent of the rest, but keeping the same increments otherwise, and $\tilde{\xi}^{(n)}$ is its hitting time of $\mathbb{Z}_{<0}$. We conclude from that case. $\square$

2.4 Stretching and gluing between two resampling steps

In this section, we run again the BBJ dynamics $(T_k^\bullet)_{k \geq 0}$ starting from an arbitrary marked binary tree $T_0^\bullet = (t, x)$. Recall that the vertices of a tree can naturally be identified with some vertices in the subsequent trees. We then consider here the sequence of the height of the initial marked vertex x in the successive trees, which we denote by $(\text{Ht}_{T_n}(x))_{n \geq 0}$. Recall that $(S_n)_{n \geq 0}$ follows the height of the marked vertex and Ξ denotes the first resampling step.

Lemma 2.10 (Stretching). *Fix $n \geq 1$ and suppose that the initial tree (t, x) is n -good. Then the height of x in the tree at time $n \wedge \Xi$ satisfies:*

$$\mathbb{P}(|\text{Ht}_{T_{n \wedge \Xi}}(x) - \text{Ht}_t(x) - (S_0 - \min_{[0, n \wedge \Xi - 1]} S)| \geq n^{1/2-5}) \leq 2 \exp\left(-\frac{1}{32} n^{1/2-35}\right).$$

Proof. The vertex x only moves, and up by 1, at a step $i \geq 1$ when the marked vertex X_{i-1} is x or an ancestor of x (necessarily a turning vertex). This occurs precisely when the path S lies at its running infimum. Hence

$$\text{Ht}_{T_{n \wedge \Xi}}(x) - \text{Ht}_t(x) = \#\{k \in \{0, \dots, n \wedge \Xi - 1\} : S_k = \min_{[0, k]} S\}.$$

Our claim thus reduces to studying the local time at the minimum of the path S and controlling the error between the previous display and $S_0 - \min_{[0, n \wedge \Xi - 1]} S$. Let us simplify the notation and set:

- $\Delta_n S = S_0 - \min_{[0, n \wedge \Xi - 1]} S$ the lower deviation of S ,
- $N_n = \#\{k \in \{0, \dots, n \wedge \Xi - 1\} : S_k = \min_{[0, k]} S\}$ the number of weak records,
- $N_n^* = \#\{k \in \{0, \dots, n \wedge \Xi - 1\} : S_k < \min_{[0, k-1]} S\}$ the number of strict records.

We thus aim to upper bound

$$|\Delta_n S - N_n| \leq |\Delta_n S - 2N_n^*| + |N_n - 2N_n^*|.$$

To this end, first notice that the values of the strict records are encoded in the sequence $(a_i)_{0 \leq i \leq \text{Ht}_t(x)-1}$ describing the ancestral line of x . Precisely, the absolute size of the strict descending record equal the length of the intervals in this sequence between two consecutive 1's. Definition 2.7 of n -good implies that

$$|N_n^* - \frac{1}{2} \Delta_n S| \leq n^{5+1/4}.$$

Next recall from the proof of Proposition 2.6 that outside the strict negative records, the path S behaves as the random walk whose increments are i.i.d. with the same law as $2 - G$ where G has the geometric distribution with parameter $1/2$. By lack of memory of this law, a new negative record is strict or weak with probability $1/2$ independently of the rest. Therefore, conditionally on the number N_n^* of strict records, the error

$$|N_n - 2N_n^*|,$$

has the law of the absolute value of the sum of N_n^* independent i.i.d. random variables with law $G - 2$, and this has sub-Gaussian tails. This is easily obtained by exploiting the explicit Laplace transform

$$\log \mathbb{E}[e^{\lambda(G-2)}] = -2\lambda - \log(2e^{-\lambda} - 1) \leq 2\lambda^2,$$

for any $|\lambda| < 1/10$. □

Removing the leaves

The walk S describes the height of the marked vertices, which is also the height of the new internal vertices at the time they are created. Lemma 2.10 then shows how this height changes as the dynamics continues. When dealing with the sole large scale geometry of a tree, leaves do not play any role since they lie at distance 1 from their parent. However leaves play an important role in the resampling and cannot be neglected anymore since an internal vertex can have 0, 1, or 2 offspring that are leaves. The next lemma controls the graph distance between the leaf $2i + 1$ and the internal vertex $2i$ that are created at the same step i , at the time Ξ of the first resampling step.

Lemma 2.11. *For every initial tree (t, x) , for every $i \geq 1$, the graph distance at the first resampling step after i between the internal vertex $2i$ and the leaf $2i + 1$ has the geometric distribution with parameter $1/2$.*

Proof. This follows from the description of the dynamics at the beginning of Section 2.2. At step i , we replace the current marked vertex X_{i-1} with the new internal vertex $2i$ with two offspring: X_{i-1} and the leaf $2i + 1$ and then either we mark $X_i = 2i + 1$ or we project the mark on an ancestor, with probability $1/2$ each. In the latter case, even if this ancestor is $2i$, the dynamics until the next resampling step will not affect the subtree formed by $2i$ and its descendants, so in particular $2i$ and $2i + 1$ will stay at distance 1 from each other. If, on the contrary, we mark $X_i = 2i + 1$, then the dynamics will create first an entire subtree above this point, before exiting it and reaching $2i$ or one of its ancestors. In this case, the distance between $2i$ and $2i + 1$ has then the same law as 1 plus the height of $2i + 1$ in this subtree. We claim that this height in the second case is geometrically distributed, which yields our claim.

Indeed, if $X_i = 2i + 1$ is the new leaf, then after that, the height of this vertex grows by 1 exactly when the current marked vertex X_j is at the position where X_i was, and all that matters is, at these moments, if the new leaf is placed on opposite side of its parent, in which case we will eventually reach back this position, or on the same side, in which case we will eventually exit the subtree without modifying further the height of the vertex $2i + 1$. Since the probability to place the leaf on the left or on the right each time we are at the root is $1/2$, independently of the rest, the final height of $2i + 1$ in this subtree is indeed geometrically distributed. $\square$

3 Scaling limit of the tree dynamics

In this section, we shall make formal the construction of the continuum dynamics described in the Introduction. This will be the analogue of the coding by the height $(S_n)_n$ of the marked vertex, using a continuous function g . We shall then prove, as stated in Theorem 1.1, that this dynamics is the scaling limit of the discrete one, when the function g is the Brownian motion, which appears as the scaling limit of $(S_n)_n$ in (2.1).

3.1 Preliminaries on continuum trees

Let us recall in this subsection the classical setting of continuum trees and refer to e.g. [LGo5] for a gentle introduction. A real tree, or R-tree, or continuum tree, is a geodesic metric space $(\mathcal{T}, d)$ without any cycle. We shall denote by $\llbracket a, b \rrbracket$ the geodesic path between a and b . Continuum trees considered in this work will be compact, equipped with a Borel finite measure μ called the *mass measure*, and a distinguished element ρ called the *root*. We write $\mathcal{T}$ for such a space

$(\mathcal{T}, d, \mu, \rho)$, which we call a rooted measured tree, or sometimes simply a tree. We shall also often distinguish a second element x of the tree (which may or not coincide with ρ), and we shall then write $\mathcal{T}^\bullet = (\mathcal{T}, x) = (\mathcal{T}, d, \mu, \rho, x)$, which we call a marked and rooted measured tree, or simply a marked tree. This point x will play the role of the marked point where the infinitesimal growth takes place. We shall denote by $\text{Ht}(z) = d(\rho, z)$ the height of any element $z \in \mathcal{T}$.

The space of such marked or non-marked trees can be equipped with a pseudo-distance, which is a variant of the Gromov–Hausdorff–Prokhorov pseudo-distance on measured compact metric spaces, taking into account the, one or two, distinguished points, as considered in e.g. [ABBGM17, Section 2]. A correspondence between two sets A and B is a subset $C \subset A \times B$ of pairs such that for every $a \in A$, there exists $b \in B$ such that $(a, b) \in C$ and conversely for every $b \in B$, there exists $a \in A$ such that $(a, b) \in C$. Then the distance $d_{\text{GHP}}^\bullet(\mathcal{T}_1^\bullet, \mathcal{T}_2^\bullet)$ between two marked trees $\mathcal{T}_1^\bullet = (\mathcal{T}_1, d_1, \mu_1, \rho_1, x_1)$ and $\mathcal{T}_2^\bullet = (\mathcal{T}_2, d_2, \mu_2, \rho_2, x_2)$ is the infimum of those $\lambda > 0$ such that:

- There exists a correspondence $\mathcal{C}$ between $\mathcal{T}_1$ and $\mathcal{T}_2$ that contains both pairs (ρ_1, ρ_2) and (x_1, x_2) and that satisfies $|d_1(a_1, b_1) - d_2(a_2, b_2)| \leq 2\lambda$ for all pairs $(a_1, a_2), (b_1, b_2) \in \mathcal{C}$;
- There exists a measure ν on $\mathcal{T}_1 \times \mathcal{T}_2$ such that if ν_i is the projection of ν on $\mathcal{T}_i$, then $\|\nu_1 - \mu_1\| + \|\nu_2 - \mu_2\| + \nu(\mathcal{C}^c) \leq \lambda$, where $\|\cdot\|$ denotes the total variation norm.

It can be checked that this function $d_{\text{GHP}}^\bullet(\cdot, \cdot)$ defines a pseudo-distance on marked trees. In addition two trees lie at distance zero from each other exactly when there exists an isometric map from one to the other that fixes the root and the marked point, as well as the measure. We let finally $\mathbb{T}^\bullet$ denote the corresponding quotient space, which is separable and complete when equipped with the distance $d_{\text{GHP}}^\bullet$. We refer again to [ABBGM17] and references therein for details.

When dealing with rooted but not marked trees, this definition is adapted in the obvious way by only requiring that a correspondence contains the roots (ρ_1, ρ_2) , and we let $(\mathbb{T}, d_{\text{GHP}})$ denote the corresponding (again, separable and complete) quotient space.

A usual simple way to construct such continuum trees consists in taking a continuous function h from a compact interval $[0, \zeta]$ to $\mathbb{R}$, and considering the pseudo-distance:

$$d_h(s, t) = d_h(t, s) = h(s) + h(t) - 2 \min_{[s, t]} h, \quad (3.1)$$

for every $0 \leq s \leq t \leq \zeta$. If one identifies points of $[0, \zeta]$ that lie at distance 0, then one obtains a compact metric space, which in this case is a continuum tree [DLGo5, Theorem 2.1]. We shall call this tree $(\mathcal{T}_h, d_h)$ the one *coded by h* . The mass measure μ_h in this case is the image of the Lebesgue measure on $[0, \zeta]$ by the canonical projection, and the tree is naturally rooted at the image of $\text{argmin } h$ by this projection. The mapping $h \mapsto \mathcal{T}_h$ is Lipschitz-continuous for the uniform topology on continuous functions and the distance d_{GHP} , see [DLGo5, Lemma 2.3] for the GH distance and [ADH13, Proposition 2.10] for the GHP distance (with no distinguished root, but the generalisation is straightforward).

Finally, the Brownian Continuum Random Tree, or here simply Brownian tree or CRT, is the random tree $\mathcal{T}_{2e}$ coded in the previous sense by **twice** the standard Brownian excursion e . Note that its mass measure is a probability measure, with total mass 1. We shall also consider CRT's with mass $m > 0$, which are coded in the same way by a Brownian excursion with duration m . The scaling property of Brownian motion shows that if $\mathcal{T} = (\mathcal{T}, d, \mu, \rho)$ is such a CRT with mass m , then

$$\frac{1}{\sqrt{m}} \cdot \mathcal{T} = (\mathcal{T}, m^{-1/2} d, m^{-1} \mu, \rho)$$

has the same law as $\mathcal{T}_{2e}$. We shall also consider marked versions of such trees, in which case the extra marked point x will be an independent point sampled from the mass measure, that is, simply the image in the tree of an independent uniform random time.

3.2 Gluing and stretching

Let us now make formal the continuous tree growth presented in the Introduction, which informally starts from a marked tree and follows the evolution of a point whose height variation is given by the function g :

- When g decreases, we go down the tree towards the root and, doing so, we *stretch* the distances along the ancestral line of the current point by a factor of 2;
- When g increases, we *graft* new pieces of the tree given by the contour of g .

We start with the case where there is no ‘resampling’, i.e. when the height of the point stays nonnegative. Let us refer to Figure 6 for a graphical representation of the construction.

Without resampling

Let $\mathcal{T}^* = (\mathcal{T}, d, \mu, \rho, x)$ be a marked measured rooted tree and let $g : [0, \zeta] \rightarrow \mathbb{R}$ be a continuous function with compact support such that

$$\text{Ht}(x) - g(0) + \min_{[0, \zeta]} g \geq 0. \quad (3.2)$$

Define the following modification:

$$h(t) = 2g(t) - \min_{[t, \zeta]} g, \quad (3.3)$$

for every $t \in [0, \zeta]$. Following (3.1), let us then define the modified tree distance:

$$D^g = d_h \quad \text{where } h \text{ is given by (3.3) and } d_h \text{ is given by (3.1).} \quad (3.4)$$

As for (3.1), it is easy to see that D^g is a pseudo-distance on $[0, \zeta]$ and its quotient is a measured real tree denoted by $\tilde{\mathcal{T}}_g^\bullet$, which should not be confused with the tree coded by g that uses d_g . Note that the mass measure on $\tilde{\mathcal{T}}_g^\bullet$ has total mass ζ . This tree comes with three natural distinguished points: the projection $\tilde{\rho}$ of the argmin of g where the tree is rooted, the projection $\tilde{x}$ of 0 (which will be identified with x), and that $\tilde{\zeta}$ of ζ (which will be the marked point of the new tree).

Let $\delta = g(0) - \min g$ and for every $r \in [0, \delta]$, let then $\tilde{x}_r \in \tilde{\mathcal{T}}_g^\bullet$ be the projection of the time $\inf\{t \in [0, \zeta] : g(0) - g(t) = r\}$. Because of the modification (3.4) the path $[\tilde{x}_0, \tilde{x}_\delta] = [\tilde{x}, \tilde{\rho}]$ describes the ancestral line of $\tilde{x}$, with $\tilde{x}_r$ lying at distance $2r$ from $\tilde{x}$ for every $r \in [0, \delta]$.

Similarly in the tree $\mathcal{T}$, for every $r \in [0, \delta]$, let x_r denote the point on the path from x to the root ρ that lies at distance r from x . Its existence follows because Assumption (3.2) states that $d(x, \rho) \geq \delta$. Then let us double in $\mathcal{T}$ the distances along the path from x to x_δ without any change anywhere else and then, for every $r \in [0, \delta]$, let us identify the point x_r to the point $\tilde{x}_r \in \tilde{\mathcal{T}}_g^\bullet$. We let finally

$$\Phi(\mathcal{T}^*, g) \quad (3.5)$$

be the metric space obtained after these identifications. An easy exercise shows that this is still a tree. We equip $\Phi(\mathcal{T}^*, g)$ with the natural measure, in which the measures on $\mathcal{T}^*$ and $\tilde{\mathcal{T}}_g^\bullet$ are simply

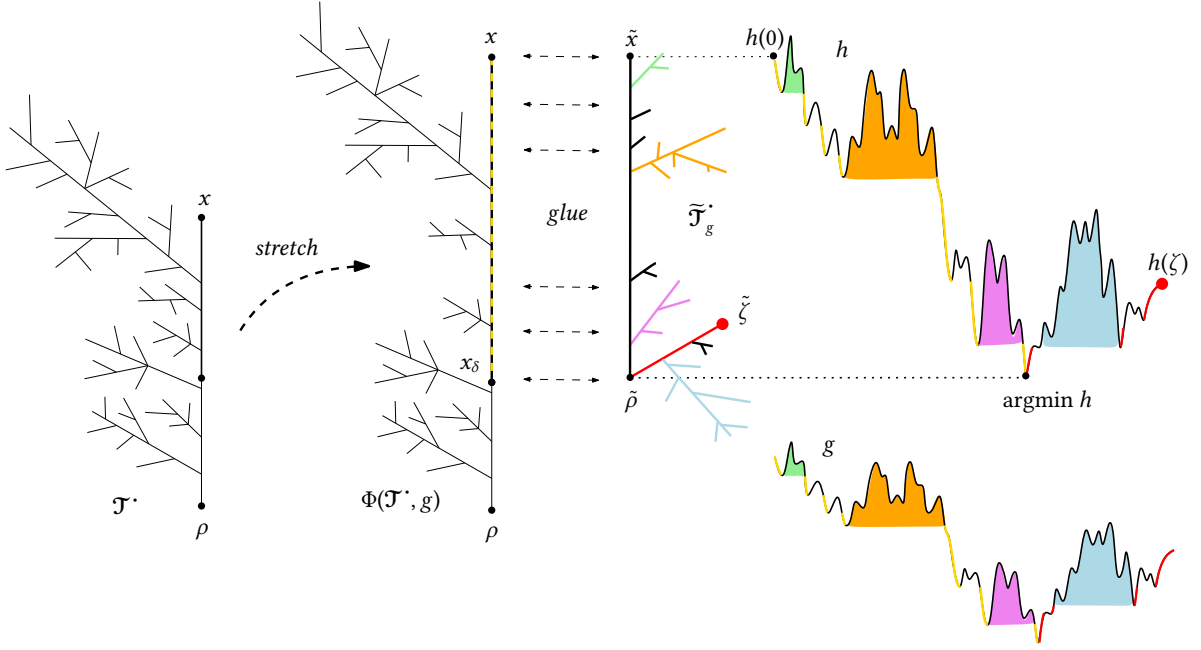


Figure 6: The construction of the continuum stretching and gluing mapping Φ from (3.5).

(pushed-forward by the stretching and) added together. We shall keep the root ρ of $\mathcal{T}$ as the root of $\Phi(\mathcal{T}^*, g)$ and forget the point $x = \tilde{x}$ and instead mark the point $\tilde{\zeta}$.

If f_1, f_2 are two continuous functions defined on possibly different intervals $[0, \zeta_1]$ and $[0, \zeta_2]$ respectively and with values in the same metric space (E, d) , then we extend the uniform distance as:

$$d_\infty(f_1, f_2) = |\zeta_1 - \zeta_2| + \sup_{s \in [0, 1]} d(f_1(s\zeta_1), f_2(s\zeta_2)).$$

We shall use this notion in both cases $E = \mathbb{R}$ and $E = \mathbb{T}^*$.

Proposition 3.1 (Continuity of Φ). *Let $\mathcal{T}_1^* = (\mathcal{T}_1, x_1)$ and $\mathcal{T}_2^* = (\mathcal{T}_2, x_2)$ be two marked measured compact real trees, and let $g_1 : [0, \zeta_1] \rightarrow \mathbb{R}$ and $g_2 : [0, \zeta_2] \rightarrow \mathbb{R}$ be two continuous functions with compact support satisfying $g_1(0) = g_2(0) = 0$ and*

$$\text{Ht}(x_1) + \min_{[0, \zeta_1]} g_1 \geq 0 \quad \text{and} \quad \text{Ht}(x_2) + \min_{[0, \zeta_2]} g_2 \geq 0.$$

Then there exists a universal constant $C > 0$ such that

$$d_{\text{GHP}}^*(\Phi(\mathcal{T}_1^*, g_1), \Phi(\mathcal{T}_2^*, g_2)) \leq C(d_{\text{GHP}}^*(\mathcal{T}_1^*, \mathcal{T}_2^*) + d_\infty(g_1, g_2)).$$

Proof. In the previous subsection, we mentioned that the mapping $h \mapsto \mathcal{T}_h$ was Lipschitz-continuous [ADH13, Proposition 2.10]. So is the stretching by a factor 2 along a branch applied to $\mathcal{T}^*$, and combining these two arguments, we obtain that the mapping $g \mapsto \tilde{\mathcal{T}}_g$ is Lipschitz-continuous as well. Finally the gluing operation is Lipschitz-continuous. $\square$

In addition to continuity, it is easy to check from the construction that Φ enjoys the following semi-group property: let $g \oplus h$ be the concatenation of the functions g and h , and assume that $g \oplus h$ is continuous and satisfies (3.2), then

$$\Phi(\Phi(\mathcal{T}^*, g), h) = \Phi(\mathcal{T}^*, g \oplus h).$$

This allows to define a process version of the previous construction. Let us denote by $C([0, \zeta], \mathbb{T}^*)$ the space of continuous functions on $[0, \zeta]$ with values in the space of rooted and marked measured real trees, equipped as usual with the uniform distance d_∞ . Given $\mathcal{T}^*$ and $g : [0, \zeta] \rightarrow \mathbb{R}$ satisfying (3.2), let us define a function $\Psi(\mathcal{T}^*, g) \in C([0, \zeta], \mathbb{T}^*)$ by setting for every $t \in [0, \zeta]$:

$$\Psi(\mathcal{T}^*, g)_t = \Phi(\mathcal{T}^*, g|_{[0, t]}).$$

Continuity of Φ shows that this function is indeed continuous in t : simply apply Proposition 3.1 to the same trees $\mathcal{T}_1^* = \mathcal{T}_2^* = \mathcal{T}^*$ and to the functions $g|_{[0, t]}$ and $g|_{[0, t+s]}$ with $s > 0$ small.

With resampling

Let us now incorporate random resamplings in the previous construction.

We shall extend the definition of Ψ without requiring Condition (3.2), but instead that the marked point $x \in \mathcal{T}$ be different from the root ρ . Also, we now let g be a continuous real-valued function, defined on the infinite half-line $[0, \infty)$ instead of a compact interval. The construction below is adapted in the obvious way if g is only defined on a compact interval $[0, \zeta]$: simply stop at time $t = \zeta$. Without loss of generality, let us also assume that $g(0) = 0$.

Let us define by induction a *random càdlàg* function $\Psi(\mathcal{T}^*, g) = (\mathcal{T}_t^g)_{t \geq 0} = (\mathcal{T}_t, d_t, \mu_t, \rho_t, x_t)_{t \geq 0}$. First let $\xi_0 = 0$ and $\mathcal{T}_0^g = \mathcal{T}^*$. Fix $i \geq 1$, suppose constructed $(\mathcal{T}_t^g)_{t \leq \xi_{i-1}}$ and let us construct $(\mathcal{T}_t^g)_{\xi_{i-1} < t \leq \xi_i}$. Let us denote by δ_i the distance between the root $\rho_{\xi_{i-1}}$ and the marked point $x_{\xi_{i-1}}$ in $\mathcal{T}_{\xi_{i-1}}^g$, which we assume is nonzero. Let then

$$\xi_i = \inf\{t \geq 0 : g(t) < -(\delta_1 + \dots + \delta_i)\}.$$

Let g_i be the function defined by $g_i(s) = g(\xi_{i-1} + s) - g(\xi_{i-1})$ for every $s \in [0, \xi_i - \xi_{i-1}]$, so g_i satisfies (3.2) with the tree $\mathcal{T}_{\xi_{i-1}}^g$. We then define for $s \in [0, \xi_i - \xi_{i-1}]$:

$$\mathcal{T}_{\xi_{i-1}+s}^g = \Phi(\mathcal{T}_{\xi_{i-1}}^g, g_i|_{[0, s]}).$$

By taking left-limit we can then define the tree $\mathcal{T}_{\xi_i-}^g \in \mathbb{T}^*$ but the marked point x_{ξ_i-} coincides with the root $\rho_{\xi_i-} = \rho$, so the function Φ cannot apply anymore. We then perform a **resampling step** (this uses extra-randomness and destroys the continuity for the marked GHP topology): we define $\mathcal{T}_{\xi_i}^g$ as the same rooted tree as $\mathcal{T}_{\xi_i-}^g$, simply replacing the marked point $x_{\xi_i-} = \rho_{\xi_i-} = \rho$ by a random point $x_{\xi_i} \in \mathcal{T}_{\xi_i-}^g$ **different from the root**, sampled independently of the past proportionally to the measure $\mu_{\xi_i-}(\cdot \cap (\mathcal{T}_{\xi_i-}^g \setminus \{\rho\}))$. Observe that the latter is not the null measure, even if the original measure μ on $\mathcal{T}$ was concentrated on the root ρ . These times ξ_i are called the *resampling instants*.

It could be that $\xi_i = \infty$ for some $i \geq 1$, in which case the dynamics continues forever using the function Φ and without resampling the marked point anymore. The next lemma rules out however the possibility that resampling instants accumulate, so the random dynamics is well-defined for every time $t \geq 0$.

Lemma 3.2 (Ψ is well-defined). *With the above notation, almost surely either there exists $n \geq 1$ such that $\xi_n = \infty$ or $\xi_n \rightarrow \infty$ as $n \rightarrow \infty$. In particular Ψ is a well-defined random continuous function of marked trees.*

Proof. Recall that we assume that in the initial tree $\mathcal{T}$, the marked point x is different from the root ρ . Then even if the initial measure μ is concentrated on ρ , for some small time $t > 0$, the dynamics has not required yet any resampling and the measure μ_t on the current tree is not concentrated

on the root, so for some $\varepsilon, \delta > 0$, the mass of points at distance greater than δ from the root is at least ε .

Then at each resampling instant ξ_i before time K , the probability to resample the new point x_{ξ_i} in the image of the previous set is bounded below by $\varepsilon/(\mu(\mathcal{T}) + K) > 0$. When this occurs, since the dynamics only increases the distances, then x_{ξ_i} lies at distance greater than δ from the root in the current tree; since the function g is uniformly continuous on $[0, K]$, then in this case, the next resampling instant cannot be arbitrarily close to the present one. Consequently we can only resample finitely many times before time K a new marked point at height greater than δ , and since this has a probability bounded away from 0 to occur at each resampling step, then there can only be finitely many resampling steps at all before time K . $\square$

This construction gives rise to a process of growing random trees $(\mathcal{T}_t^g)_{t \geq 0}$ started from any tree with a marked point distinct from the root. Beware though, the law of this process is *not* continuous with respect to the function g . This is due mainly to the behaviour of g near the resampling instants, e.g. if it makes a local minimum or a plateau where it stays constant. Such pathological behaviours however do not occur when considering a random function such as Brownian motion B as we will do below.

3.3 Scaling limit of the discrete dynamics

Let us finally prove that this continuum construction with an independent Brownian motion as driving function is the scaling limit of the discrete dynamics as claimed in Theorem 1.1.

Let us first consider the dynamics between two consecutive resampling steps. The framework is the same as in Proposition 2.9: for every $n \geq 1$, we let the discrete dynamics $(T_k^{(n)}, X_k^{(n)})_{k \geq 0}$ start from a marked binary tree $(T_0^{(n)}, X_0^{(n)}) = (t^{(n)}, x^{(n)})$, not necessarily with n internal vertices, and where $x^{(n)}$ has height h_n . Assume that $t^{(n)}$ is equipped with a finite measure $\mu^{(n)}$ on its vertices. Let $\Xi^{(n)}$ denote the first resampling time. We shall equip the tree $T_k^{(n)}$ with the measure $\mu_k^{(n)}$ given by the image of $\mu^{(n)}$ on the vertices of $t^{(n)}$ and a mass 1 on each new vertex.

Theorem 3.3. *Assume that the initial non-marked tree $t^{(n)}$ is n -good in the sense of Definition 2.7 and that, as a marked tree, it satisfies the convergence in the marked GHP topology on $\mathbb{T}^*$:*

$$\left(\frac{1}{\sqrt{2n}} t^{(n)}, \frac{1}{2n} \mu^{(n)}, x^{(n)} \right) \xrightarrow{n \rightarrow \infty} \mathcal{T}^* = (\mathcal{T}, d, \mu, \rho, x),$$

for some limit tree in which $\text{Ht}(x) > 0$. Let $B = (B_t)_{t \geq 0}$ be an independent standard Brownian motion and $\xi = \inf\{t > 0 : B_t < -\text{Ht}(x)\}$. Then we have

$$\left(\frac{\Xi^{(n)}}{n}, \left(\frac{1}{\sqrt{2n}} T_{[tn] \wedge (\Xi^{(n)} - 1)}^{(n)}, \frac{1}{2n} \mu_{[tn] \wedge (\Xi^{(n)} - 1)}^{(n)}, X_{[tn] \wedge (\Xi^{(n)} - 1)}^{(n)} \right)_{0 \leq t < [\Xi^{(n)}/n]} \right) \xrightarrow[n \rightarrow \infty]{(d)} (\xi, (\Psi(\mathcal{T}^*, B)_t)_{0 \leq t < \xi}), \quad (3.6)$$

for the Skorokhod topology.

Proof. The strategy of the proof consists in replacing in $T_k^{(n)}$ the graph distance $d_k^{(n)}$ and the measure $\mu_k^{(n)}$ by the discrete analogue of the construction of Ψ , using the height process $S^{(n)}$ of the marked vertex, that we shall denote by $\tilde{d}_k^{(n)}$ and $\tilde{\mu}_k^{(n)}$ respectively, and proving that the GHP distance between $(T_k^{(n)}, \frac{1}{\sqrt{2n}} d_k^{(n)}, \frac{1}{2n} \mu_k^{(n)}, X_k^{(n)})$ and $(T_k^{(n)}, \frac{1}{\sqrt{2n}} \tilde{d}_k^{(n)}, \frac{1}{2n} \tilde{\mu}_k^{(n)}, X_k^{(n)})$ tends to 0 in probability as $n \rightarrow \infty$ uniformly over $k \leq \Xi^{(n)}$. We then prove our claim for this modified tree-process, using the convergence of $S^{(n)}$ and $\Xi^{(n)}$ from Proposition 2.9.

Fix $\varepsilon > 0$. With probability at least $1 - \varepsilon$, the first resampling time $\Xi^{(n)}$ is bounded by some large constant C times n (thanks to (2.1)) and all marked trees $(T_k^{(n)}, X_k^{(n)})$ for $k < \Xi^{(n)}$ are Cn -good, thanks to Lemma 2.8. We henceforth assume that these events do occur.

STEP 1: MODIFYING THE IMAGE OF THE INITIAL TREE. Let us first focus on the image of the vertices of the initial tree $t^{(n)}$ in the tree $T_k^{(n)}$. Recall the description of $t^{(n)}$ as the ancestral lineage of $x^{(n)}$ along which are grafted subtrees that will not change at all until time $\Xi^{(n)}$ (included). Only the ancestral lineage of $x^{(n)}$ will be stretched by grafting new subtrees on it. Precisely, recall that at each time of a new strict negative record by $S^{(n)}$, the discrete tree is marked at the next turning vertex along the ancestral line of $x^{(n)}$; since these vertices are n -good by our assumption that the initial tree is, then Lemma 2.10 applies to show that, up to a $o(\sqrt{n})$ -error, their ancestral lineage will be stretched at time k by the negative of the infimum of $S^{(n)}$ since their visit. A union bound shows that the amount of stretching between $x^{(n)}$ and the first turning vertex, then between two turning vertices, is approximately, up to a $o(\sqrt{n})$ factor, given by the differences in the successive strict negative records of $S^{(n)}$, uniformly over time up to $\Xi^{(n)}$. Let $\tilde{d}_k^{(n)}$ denote the distance between the vertices of $t^{(n)}$ after this stretch using $(S_i^{(n)})_{i \leq k}$; then,

$$\frac{1}{\sqrt{n}} \sup_{k \leq \Xi^{(n)}} \sup_{a, b \in t^{(n)}} |d_k^{(n)}(a, b) - \tilde{d}_k^{(n)}(a, b)| \xrightarrow[n \rightarrow \infty]{P} 0.$$

As far as measures are concerned, let us define $\tilde{\mu}_k^{(n)}(a) = \mu_k^{(n)}(a) = \mu^{(n)}(a)$ for every $a \in t^{(n)}$, without any modification.

STEP 2: REMOVING THE NEW LEAVES. Let us turn to the subtrees that are created by the dynamics. At time $k \leq \Xi^{(n)}$, they form a forest $\tilde{T}_k^{(n)}$ that we equip with the graph distance $d_k^{(n)}$ inherited from $T_k^{(n)}$ (equivalently, we keep $x^{(n)}$ and its ancestral line) and the measure $\mu_k^{(n)}$ that gives a mass 1 to every vertex. As in the previous section, let us order these vertices from 2 to $2k + 1$ in order of appearance: the vertices $2i$ and $2i + 1$ are respectively the internal one and the leaf created at step i . For definiteness, let 0 and 1 also denote the vertex 2.

We shall focus on internal vertices only and define $\bar{2i} = \overline{2i + 1} = 2i$ for every $1 \leq i \leq k$, i.e. we project each new leaf on the internal vertex with preceding label. Define then $\bar{\mu}_k^{(n)}$ as the image of $\mu_k^{(n)}$ by this projection and $\bar{d}_k^{(n)}(i, j) = d_k^{(n)}(\bar{i}, \bar{j})$ for every $i, j \leq 2k + 1$. According to Lemma 2.11 the distance between the leaf $2i + 1$ and the internal vertex $2i$ can only increase over time and, at time $\Xi^{(n)}$, it has the geometric distribution with parameter $1/2$. We infer from a union bound (recall the Hoeffding inequality in the proof of Lemma 2.10) that, uniformly over all time up to $\Xi^{(n)}$, which we recall is at most Cn with high probability, and uniformly over all leaves up to this time, their distance to the internal vertex with the preceding label is $o(\sqrt{n})$. In other words,

$$\frac{1}{\sqrt{n}} \sup_{k \leq \Xi^{(n)}} \sup_{a, b \in \tilde{T}_k^{(n)}} |d_k^{(n)}(a, b) - \bar{d}_k^{(n)}(a, b)| \xrightarrow[n \rightarrow \infty]{P} 0.$$

STEP 3: MODIFYING THE NEW INTERNAL VERTICES. Let us modify now the distances $\bar{d}_k^{(n)}$ on $\tilde{T}_k^{(n)}$ as in the first step.

In the forest $\tilde{T}_k^{(n)}$, the distance between the corresponding internal vertices is given by the sum of their height in the tree $T_k^{(n)}$ minus the height of their last common ancestor. Recall that the height of an internal vertex $2i$ at the time it is created is given by the path $S_{i-1}^{(n)}$; also Lemma 2.4 allows to find the last common ancestor using this path, and finally Lemma 2.10 shows how the height evolves after it is created. For the latter fact, recall that with high probability, the marked trees up to time $\Xi^{(n)} \leq Cn$ are all Cn -good. Just as in Step 1, define then the distance $\tilde{d}_k^{(n)}$ between

two internal vertices $2i$ and $2j$ of $\tilde{T}_k^{(n)}$ as

$$\tilde{d}_k^{(n)}(2i, 2j) = (2S_{i-1}^{(n)} - \min_{[i-1, k]} S^{(n)}) + (2S_{j-1}^{(n)} - \min_{[j-1, k]} S^{(n)}) - 2 \min_{i-1 \leq \ell \leq j-1} (2S_\ell^{(n)} - \min_{[\ell, k]} S^{(n)}). \quad (3.7)$$

As in Step 1, according to Lemma 2.10, with high probability this distance is close to $\bar{d}_k^{(n)}$ up to $o(\sqrt{n})$ error, uniformly in $k \leq \Xi^{(n)}$. As for the measures, as previously let $\tilde{\mu}_k^{(n)} = \bar{\mu}_k^{(n)}$, without further modification.

STEP 4: COMPARISON WITH THE CONTINUOUS DYNAMICS. By definition of the GHP, using as correspondence the identity on the image of the vertices of $t^{(n)}$ and the projection $i \mapsto \bar{i}$ for the new vertices, the first two steps have shown that the GHP distance between the original tree-process

$$\left(T_{[tn] \wedge \Xi^{(n)}}^{(n)}, \frac{1}{\sqrt{2n}} d_{[tn] \wedge \Xi^{(n)}}^{(n)}, \frac{1}{2n} \mu_{[tn] \wedge \Xi^{(n)}}^{(n)}, X_{[tn] \wedge \Xi^{(n)}}^{(n)} \right)_{t \geq 0}$$

and the modified one

$$\left(T_{[tn] \wedge \Xi^{(n)}}^{(n)}, \frac{1}{\sqrt{2n}} \tilde{d}_{[tn] \wedge \Xi^{(n)}}^{(n)}, \frac{1}{2n} \tilde{\mu}_{[tn] \wedge \Xi^{(n)}}^{(n)}, X_{[tn] \wedge \Xi^{(n)}}^{(n)} \right)_{t \geq 0}$$

tends to 0 in probability. Let us finally argue that the latter converges to our desired limit $(\Psi(\mathcal{T}^\bullet, B)_{t \wedge \xi})_{t \geq 0}$.

Let us fix some time $t > 0$ and assume $t < \xi$; recall that $\Psi(\mathcal{T}^\bullet, B)_t$ is constructed from $\mathcal{T}^\bullet$ by stretching the distances by a factor 2 along the ancestral line of the marked point and gluing this path to a new tree $\tilde{\mathcal{T}}_t^\bullet$ coded by the pseudo-distance D^g from (3.4) where $g = (B_s)_{s \leq t}$. Precisely $\tilde{\mathcal{T}}_t^\bullet$ is the quotient of the interval $[0, t]$ by the equivalence relation $D^g = 0$. This provides a natural correspondence between $\tilde{\mathcal{T}}_t^\bullet$ and $\tilde{T}_{[nt]}^{(n)}$ by letting correspond the image of a time $s \in [0, t]$ in $\tilde{\mathcal{T}}_t^\bullet$ and the vertex labelled $[ns]$ in $\tilde{T}_{[nt]}^{(n)}$. Comparing the formula (3.7) for the discrete distances and (3.4) for the continuum one, and recalling that $S^{(n)}$ and $\Xi^{(n)}$ converge after scaling to B and ξ by Proposition 2.9, we infer that, jointly, the processes $(\frac{1}{\sqrt{2n}} \tilde{T}_{[nt] \wedge \Xi^{(n)}}^{(n)})_{t \geq 0}$ converge in distribution to $(\tilde{\mathcal{T}}_{t \wedge \xi}^\bullet)_{t \geq 0}$ in the GHP topology. Similarly, since $t^{(n)}$ converges after scaling in the GHP topology towards $\mathcal{T}$, using the image of the correspondence between these trees at time t , we conclude by comparing the discrete distance $\tilde{d}_{[tn] \wedge \Xi^{(n)}}^{(n)}$ and the continuum one that the trees at time t converge. $\square$

Theorem 3.3 does not readily extend to the first resampling time. More precisely, (3.6) shows that, just before resampling, the marked tree $(\frac{1}{\sqrt{2n}} T_{\Xi^{(n)}-1}^{(n)}, \frac{1}{2n} \mu_{\Xi^{(n)}-1}^{(n)}, X_{\Xi^{(n)}-1}^{(n)})$ converges towards $\Psi(\mathcal{T}^\bullet, B)_{\xi-}$ in the marked GHP sense. In general this *does not imply* that

$$\left(\frac{1}{\sqrt{2n}} T_{\Xi^{(n)}}^{(n)}, \frac{1}{2n} \mu_{\Xi^{(n)}}^{(n)}, X_{\Xi^{(n)}}^{(n)} \right) \xrightarrow[n \rightarrow \infty]{(d)} \Psi(\mathcal{T}^\bullet, B)_\xi. \quad (3.8)$$

The issue is that if the measure of $\mathcal{T}^\bullet$ possesses an atom at the root, then, in the discrete setting we may resample $X_{\Xi^{(n)}}^{(n)}$ close to the root, while the marked point in $\Psi(\mathcal{T}^\bullet, B)_\xi$ is sampled according to the measure of $\Psi(\mathcal{T}^\bullet, B)_{\xi-}$ *except* its possible atom at the root. However, as soon as $\mathcal{T}^\bullet$ has no atom at the root, then (3.8) is granted. By the Markovian nature of the discrete and continuous process, one may then extend the convergence up to finitely many resampling times: the convergence (3.8) serves as a new initial condition, and by Lemma 2.8, the non-marked tree $T_{\Xi^{(n)}}^{(n)}$ is $\max(n, \Xi^{(n)})$ -good and $\Xi^{(n)}$ is of order n by (3.6), leading to the convergence (3.6) up to the second resampling time, etc. Finally, recall from Lemma 3.2 that resampling instants do not accumulate, then the

convergence (3.6) extends to convergence over any compact intervals, i.e. in the Skorokhod topology on $\mathbb{D}(\mathbb{R}_+, \mathbb{T}^\bullet)$.

Theorem 1.1 in the Introduction claims that such a convergence also holds when the discrete dynamics starts from a single vertex, which does not fit the assumption of Theorem 3.3 but it can easily be deduced from this.

Proof of Theorem 1.1. Fix $\delta > 0$. The discrete dynamics started from a single vertex produces at time $\lfloor \delta n \rfloor$ a uniform random binary tree with size $\lfloor \delta n \rfloor$ and equipped with an independent uniform random vertex. By Aldous' invariance principle (1.1) (extended to take care of the measure), we then have

$$\frac{1}{\sqrt{2n}} \cdot T_{\lfloor \delta n \rfloor}^\bullet \xrightarrow[n \rightarrow \infty]{(d)} \mathcal{T}_\delta^\bullet,$$

where the limit is a Brownian CRT with mass δ marked at an independent random point sampled from its mass measure. We may then apply Theorem 3.3 and the above discussion to derive the convergence of processes:

$$\left(\frac{1}{\sqrt{2n}} \cdot T_{\lfloor tn \rfloor}^\bullet \right)_{t \geq \delta} \xrightarrow[n \rightarrow \infty]{(d)} (\mathcal{T}_t^\bullet)_{t \geq \delta} = \Psi(\mathcal{T}_\delta^\bullet, B|_{[\delta, \infty)}).$$

To finish, let us remark that, as $\delta \rightarrow 0$, the sequence $(\frac{1}{\sqrt{2n}} \cdot T_{\lfloor tn \rfloor}^\bullet)_{0 \leq t \leq \delta}$ converges in probability towards a singleton uniformly in n . These facts imply the convergence stated in Theorem 1.1. $\square$

4 Rémy's diffusion is ergodic

In this section we prove Theorem 1.2 on unique ergodicity of Rémy's continuous dynamics. Henceforth, we assume that the initial tree $\mathcal{T}_0^\bullet$ has mass 1 and is marked at a point different from its root. We let B be an independent Brownian motion. In order to lighten the notation, we shall set for every $t \geq 0$:

$$\mathcal{T}_t^\bullet = \Psi(\mathcal{T}_0^\bullet, B)_t \quad \text{and} \quad \mathcal{S}_t^\bullet = \frac{1}{\sqrt{1+t}} \cdot \mathcal{T}_t^\bullet.$$

In this way $\mathcal{S}_t^\bullet$ has mass 1 and its marked point differs from its root for every $t \geq 0$. Let $\mathcal{T}_{2e}^\bullet$ denote the CRT coded by twice the standard Brownian excursion in the usual sense, together with a marked point sampled independently from its mass measure. Throughout this section, by a *marked Brownian CRT*, we mean a possibly scaled version $\mathcal{T}_{2e}^\bullet$, in which the marked point is always sampled independently from the mass measure. Then Theorem 1.2 claims that for any (possibly random) such initial tree $\mathcal{T}_0^\bullet$, we have the convergence in distribution

$$\mathcal{S}_t^\bullet \xrightarrow[t \rightarrow \infty]{(d)} \mathcal{T}_{2e}^\bullet,$$

in the marked GHP topology.

To prove this result, we first observe that, from the discrete approximation, if $\mathcal{T}_0^\bullet$ has the law of $\mathcal{T}_{2e}^\bullet$, then so does $\mathcal{S}_t^\bullet$ for any $t > 0$. The convergence for a generic initial distribution then follows if we can prove that $\mathcal{S}_t^\bullet$ stays eventually arbitrarily close to the analogue started from $\mathcal{T}_{2e}^\bullet$ when using as much as possible the same randomness from Brownian motion and resampling for both processes. We stress however that the dynamics never entirely forgets its past and so we only provide an asymptotic coupling in the spirit of [HMS11]. Although we could try to fit the very general framework of this reference, we have opted for a (rather short and) completely self-contained argument.

Our strategy consists in showing that, given any initial tree and $\varepsilon > 0$, with high probability, there exists a (random) finite time t at which the tree $\mathcal{T}_t^\bullet$ contains a subtree whose mass represents a fraction greater than $1 - \varepsilon$ of the total tree and that has the law of a scaled version of $\mathcal{T}_{2\varepsilon}^\bullet$ (and whose mark is the mark of $\mathcal{T}_t^\bullet$), and such that in addition the subsequent evolution of this subtree and the entire tree stay close to each other. Precisely, we shall prove that the GHP distance between the rescaled entire tree and the rescaled Brownian subtree tends to 0 on this event with large probability, which allows to conclude.

4.1 A large Brownian subtree

Let us first make formal the previous idea that the entire tree will contain at some finite time a large Brownian subtree that will occupy most of it. Let us first introduce some denomination that we will use throughout this section.

Definition 4.1. We say that a random marked tree $\mathcal{T}^\bullet$ with mass 1 is ε -good where $\varepsilon \in (0, 1)$, if it is composed of three parts, as illustrated in Figure 7:

- (i) a *spine*, in red, which is a segment started from the root and with length smaller than ε ;
- (ii) a *good tree*, in green, grafted at the extremity of the spine, with mass greater than $1 - \varepsilon$, which has the law of a scaled marked Brownian CRT, whose marked point is that of the entire tree;
- (iii) a *bad part* made of an arbitrary forest of blue trees grafted all along the spine.

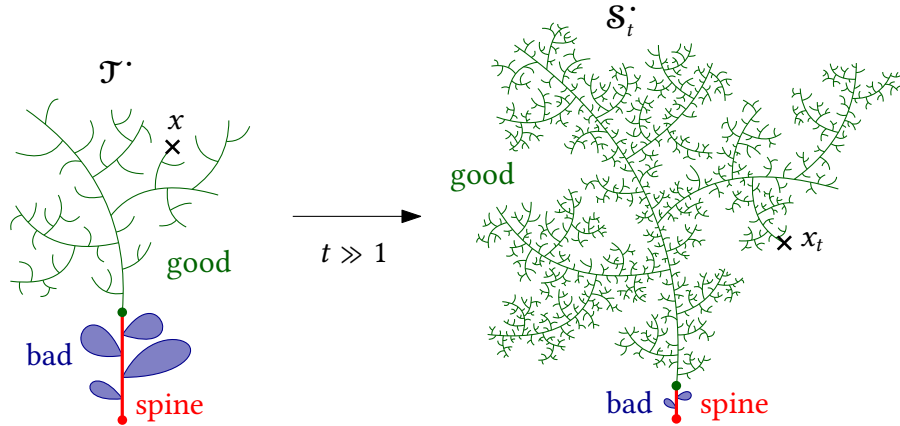


Figure 7: Strategy of the proof of Theorem 1.2: If we start from the situation on the left with a big green Brownian subtree sitting at the end of a short red spine, then with high probability this Brownian subtree will take over the entire tree when applying Rémy's growing diffusion. A key input is that with high probability we always resample the marked point in the green Brownian part.

If at some time $t \geq 0$ in Rémy's diffusion, the rescaled tree $\mathcal{S}_t^\bullet$ is ε -good, then we shall still call *good* and *bad* the parts of the subsequent trees that will come from stretching or grafting on these initial good and bad parts respectively, and call *spine* the parts coming from stretching the initial red spine, but new subtrees grafted on the spine will belong to the bad part.

Lemma 4.2 (Birth of a Brownian subtree). *Fix any marked measured tree $\mathcal{T}_0^\bullet$ with mass 1 and a marked point different from the root and fix some $\varepsilon \in (0, 1)$. Then there exists an almost surely finite*

stopping time $N \geq 1$ (for the discrete-time filtration generated by the dynamics at the resampling instants) such that $\mathcal{S}_{\xi_N}^\bullet$ is ε -good.

Proof. Suppose first that the marked point in the initial tree lies at height $h \geq 1$ and let us prove the conclusion occurs with positive probability for $N = 1$. Fix $K > 0$ arbitrarily large and observe that the following event (depicted in Figure 8 left) about a Brownian motion B started from $B_0 = 1$ occurs with positive probability: first $\tau = \inf\{t > 0 : B_t = 0\} \in (K + 1, K + 2)$ and second, there exists a subexcursion above the running minimum, on the time interval say $[g, d]$, such that $g \leq 1, d \geq K + 1$ and $B_g = B_d \leq 1$. Let us scale this picture by a factor h in space and h^2 in time, then $h^2\tau = \xi_1$ is the first resampling time. Conditionally on the previous event, the sub-excursion on the interval (g, d) codes a Brownian subtree that occupies a fraction of the mass of the tree $\Psi(\mathcal{T}_0^\bullet, B)_{\xi_1}$ equal to, since $h^2 \geq 1$,

$$\frac{h^2d - h^2g}{1 + h^2\tau} \geq \frac{d - g}{1 + \tau} \geq \frac{K}{K + 3}.$$

Since $1 + h^2\tau \geq h^2\tau \geq h^2(K + 1)$, then the root of this Brownian subtree lies at height at most

$$h \leq \frac{1}{\sqrt{K + 1}} \sqrt{1 + h^2\tau} = \frac{1}{\sqrt{K + 1}} \sqrt{1 + \xi_1}.$$

We conclude that given any $K > 0$, there exists $c_K > 0$, depending only on K , such that, whatever the initial tree $\mathcal{T}_0^\bullet$, provided that the marked point lies at height no smaller than 1, at the first resampling instant ξ_1 , with probability at least c_K , the rescaled tree $\mathcal{S}_{\xi_1}^\bullet$ contains a Brownian subtree that occupies a fraction of the mass at least $1 - 3/(K + 3)$ and whose root lies at height at most $1/\sqrt{K + 1}$; finally the new marked point is sampled in the Brownian subtree from its mass measure.

Consider now the case of an initial marked point lying at height smaller than 1 and recall that the initial tree $\mathcal{T}_0^\bullet$ has mass 1. Now for every $k \geq 1$, a Brownian motion started from $B_0 = 0$ has the same positive probability, independent of k , to satisfy $B_{2^k} > 2 \times \sqrt{2^k - 2^{k-1}} + \inf_{[2^{k-1}, 2^k]} B$; see Figure 8 right for a representation of this event for $k = 1$. Hence almost surely this occurs for some finite index $k_0 \geq 1$ and then, whatever the initial tree $\mathcal{T}_0^\bullet$, the marked point in the tree $\mathcal{S}_{2^{k_0}}^\bullet$ lies at height greater than $2 \times \sqrt{2^{k_0} - 2^{k_0-1}} / \sqrt{1 + 2^{k_0}} \geq 2/\sqrt{3} > 1$. According to the first part of this proof, there is a probability at least $c_K > 0$ that at the first resampling time after 2^{k_0} , the rescaled tree contains a Brownian subtree that occupies a fraction of the mass at least $1 - 3/(K + 3)$ and whose root lies at height at most $1/\sqrt{K + 1}$. If this fails, then either the marked point at this resampling time lies at height greater than 1 in the rescaled tree and we have the same probability c_K that this succeeds at the next resampling time. Else, if the marked point at the resampling time lies at height smaller than 1, then we wait for a finite time until we are sure to reach a height greater than 1 before applying the first part of the argument again. In any case, after an almost surely finite time, the same conclusion as the first paragraph holds. $\square$

Rate of growth of Brownian trees

The next key idea is to understand the rate of growth of the discrete-time chain made of the random trees in-between resampling times when the initial tree is a Brownian CRT. This will eventually show that the Brownian subtree obtained from the previous lemma will grow more rapidly than the rest. Precisely, let $(\xi_n)_{n \geq 1}$ denote the resampling instants in the process $\Psi(\mathcal{T}_0^\bullet, B)$ and also $\xi_0 = 0$.

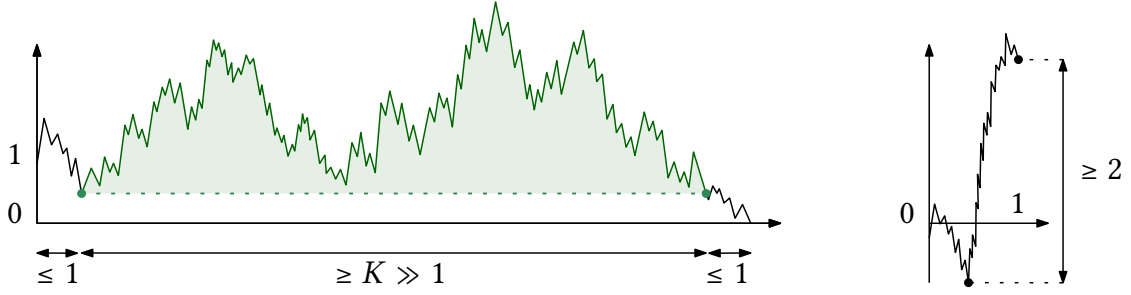


Figure 8: How to get a good tree from an arbitrary tree: on the left a Brownian motion started from 1 makes an enormous excursion above its minimum at a low height before reaching 0; on the right a Brownian motion started from 0 ends at time 1 at least 2 unit above its minimum.

It follows from Theorem 3.3 and Lemma 2.3 that if $\mathcal{T}_0^\bullet = \mathcal{T}_{2e}^\bullet$ is a standard marked Brownian CRT, then conditionally on $(\xi_n)_{n \geq 1}$, the renormalised trees $(\mathcal{S}_{\xi_n}^\bullet)_{n \geq 0}$ form a stationary sequence. In particular, the ratios of their masses $(1 + \xi_{n+1})/(1 + \xi_n)$ for $n \geq 0$ also form a stationary sequence. We first compute this stationary law.

For every $a > 0$, let $\tau_a = \inf\{t > 0 : B_t = -a\}$ denote the first hitting time of $-a$ by the standard Brownian motion, then the first resampling time ξ_1 is distributed as $\tau_{\text{Ht}(X_0)}$ where $\text{Ht}(X_0)$ is the height of the random point X_0 in $\mathcal{T}_0^\bullet$, which is independent of B .

Lemma 4.3. *When $\mathcal{T}_0^\bullet = \mathcal{T}_{2e}^\bullet$, the random variable $\log(1 + \tau_{\text{Ht}(X_0)})$ is exponentially distributed with expectation 2.*

Proof. For standard Brownian motion, τ_a has density $\frac{a}{\sqrt{2\pi t^3}} e^{-a^2/(2t)}$ on $[0, \infty)$. Then for twice the standard Brownian excursion, it is well-known that the height of a uniform random point follows the Rayleigh distribution with density $x e^{-x^2/2}$ on $[0, \infty)$. This can be obtained e.g. from the Bismut decomposition or by taking the scaling limit of the height of a uniform random vertex in a uniform random binary tree.

A straightforward integration, using the change of variables $(u, y) = (\log(1 + t), x \sqrt{(1 + t)/t})$, whose Jacobian determinant equals $(t(1 + t))^{-1/2}$, shows that for any measurable and nonnegative function f , we have

$$\begin{aligned} \mathbb{E}[f(\log(1 + \tau_{\text{Ht}(X)}))] &= \int_0^\infty \int_0^\infty f(\log(1 + t)) \frac{x}{\sqrt{2\pi t^3}} e^{-x^2/(2t)} x e^{-x^2/2} dx dt \\ &= \int_0^\infty \int_0^\infty f(u) e^{-u/2} \frac{y^2}{\sqrt{2\pi}} e^{-y^2/2} dy du \\ &= \int_0^\infty \frac{f(u)}{2} e^{-u/2} du. \end{aligned}$$

This proves the claim. $\square$

In the next proposition, we let $\mathcal{G}_t$ denote the good part of $\mathcal{T}_t$ and we denote by γ_t its mass and by $\beta_t = (t + 1) - \gamma_t$ the combined mass of the bad part and the spine. If the marked point x_t of $\mathcal{T}_t^\bullet$ belongs to $\mathcal{G}_t$, then the latter is marked at x_t ; otherwise we mark $\mathcal{G}_t$ at its root.

Proposition 4.4 (The Brownian subtree may take over). *For any $\varepsilon > 0$, there exists $c_\varepsilon > 0$ such that if the initial tree $\mathcal{T}_0^\bullet$ is ε -good, then with (annealed) probability at least c_ε , the following two events hold:*

- (i) For every $n \geq 1$, the point that is resampled at time ξ_n in the tree $\mathcal{T}_{\xi_n}^\bullet$ always belongs to the good part $\mathcal{G}_{\xi_n}$,
- (ii) and the proportion of the good mass $y_t/(1+t)$ tends to 1 and the marked GHP distance between the rescaled entire tree $\frac{1}{\sqrt{1+t}} \cdot \mathcal{T}_t^\bullet$ and the good part $\frac{1}{\sqrt{y_t}} \cdot \mathcal{G}_t^\bullet$ tends to 0.

Proof. Recall that if we start the dynamics with a marked Brownian CRT of mass 1, then the ratios $((1 + \xi_n)/(1 + \xi_{n-1}))_{n \geq 1}$ of the masses between two resampling times form a stationary sequence. Without assuming ergodicity of the rescaled trees $\mathcal{S}_{\xi_n}^\bullet$ (which is what we are eventually aiming to prove!), Birkhoff's ergodic theorem implies the convergence

$$\frac{1}{n} \sum_{k=1}^n \log \frac{1 + \xi_k}{1 + \xi_{k-1}} \xrightarrow[n \rightarrow \infty]{a.s.} \Theta, \quad (4.1)$$

where, by uniform integrability, the limit Θ is a random variable that has expectation $\mathbb{E}[\log(1 + \xi_1)] = 2$ according to Lemma 4.3. Notice that this expectation is relative to both the initial tree and the independent Brownian motion B driving Rémy's diffusion (and the randomness needed for the resamplings).

Now suppose that the initial tree contains a good part, which is a marked Brownian CRT with mass m , whose marked point is that of the entire tree, sitting on top of a spine with length h . If the dynamics Ψ only resamples points in the good tree, then in-between resampling events, it first stretches and glues in this good subtree, before reaching the spine, stretching it and gluing new bad subtrees up to the next resampling event. By the strong Markov property, we may split the Brownian motion B in two: the part B^1 affecting the good tree and the part B^2 affecting the spine and bad trees and both are independent Brownian motions. Since the good part is initially a Brownian CRT, then after scaling by a factor m , its mass evolution satisfies (4.1) with the ξ_k^1 's corresponding to B^1 and with a limit Θ^1 that satisfies $\mathbb{E}[\Theta^1] = 2$, so in particular $\mathbb{P}(\Theta^1 \geq 2) > 0$. As for the rest, at each step the length of the spine is simply doubled, and we add a forest coded by the Brownian motion B^2 started from the current length of the spine and killed upon reaching 0.

Then after n resampling steps, if all in the good part, this good part has mass

$$y_{\xi_n} = m(1 + \xi_n^1) = m \prod_{k=1}^n \frac{1 + \xi_k^1}{1 + \xi_{k-1}^1} = m e^{n(\Theta^1 + o(1))},$$

where $o(1)$ is a term that tends almost surely to 0, while the spine and bad part have a combined mass

$$\beta_{\xi_n} = (1 - m) + \tau_{h(1+2+4+\dots+2^{n-1})}^2 \stackrel{(d)}{=} (1 - m) + h^2 \tau_{2^{n-1}}^2,$$

where τ_a^2 is the hitting time of $-a$ by the Brownian motion B^2 .

Suppose that the initial tree is ε -good, namely that $h < \varepsilon$ and $m > 1 - \varepsilon$. Let us fix $\eta > 0$ small enough so $e^2 > 4 + 4\eta$. Easy estimates show that, with a positive probability, depending on ε , we have $\beta_{\xi_n} \leq (4 + \eta)^n$ for all $n \geq 0$ large enough. Recall that $\mathbb{P}(\Theta \geq 2) > 0$; since $e^2 > 4 + 4\eta$, then again, with a positive probability, depending on ε , we have $y_{\xi_n} \geq (4 + 2\eta)^n$ for all $n \geq 0$ large enough. We infer that, a posteriori, the conditional probability to always resample the points in the good part equals

$$\prod_{n \geq 1} \frac{y_{\xi_n}}{\beta_{\xi_n} + y_{\xi_n}},$$

which is positive with positive probability. This proves the first claim.

For the second claim, note that on the previous event, after n steps and without rescaling, the good tree is a Brownian CRT with mass $\gamma_{\xi_n} \geq (4 + 2\eta)^n$, whereas the bad part is made of a spine with length $h(2^n - 1)$ on which are grafted the initial bad trees that have not changed as well as a forest of Brownian CRT's with total mass at most $\beta_{\xi_n} \leq (4 + \eta)^n$. In addition, with a positive probability, depending again only on ε , this bad forest of Brownian CRT's only contains trees with diameter smaller than $(4 + 3\eta/2)^{n/2}$. If we rescale the entire tree by $1/(\gamma_{\xi_n} + \beta_{\xi_n}) = 1/(\xi_n + 1)$, then the spine plus bad part only thus represent a tiny fraction in terms of both the mass and distances so the good part is eventually close to the entire tree for large enough resampling instants.

In order to conclude about any time t large enough, we compare the situation at such a time and at the two surrounding resampling times using monotonicity. Indeed, if the point that is resampled at time ξ_n belongs to the good part, then the dynamics will first affect this part only, so the rescaled good part will be even closer to the rescaled entire tree, until we leave the good part and reach the spine: then the dynamics will grow the bad part so the good part and the entire tree start to differ more and more, but only up to time ξ_{n+1} where they are close again. $\square$

Let us prove now that the conclusion of Proposition 4.4 in fact holds with high probability when ε is small. To this end, let us first improve the ergodic theorem (4.1) by showing that the limit Θ is maximal. This relies on Proposition 4.4 (that used (4.1) in the first place!).

Corollary 4.5 (Ergodic theorem for the growth of Brownian trees). *When $\mathcal{T}_0^\bullet = \mathcal{T}_{2e}^\bullet$, we have the almost sure convergence:*

$$\frac{1}{n} \sum_{k=1}^n \log \frac{1 + \xi_k}{1 + \xi_{k-1}} \xrightarrow[n \rightarrow \infty]{a.s.} \mathbb{E}[\log(1 + \tau_{\text{Ht}(X_0)})] = 2.$$

Proof. According to (4.1), the almost sure convergence holds with a limit Θ that has expectation $\mathbb{E}[\log(1 + \tau_{\text{Ht}(X_0)})] = 2$. It suffices then to prove that $\Theta \geq 2$ almost surely.

Let us fix $\eta > 0$ such that $e^{2-\eta} > 4$ and such that $\mathbb{P}(\Theta = 2 - \eta) = 0$. Let $A = \{\Theta \leq 2 - \eta\}$, and let us suppose for contradiction that $\mathbb{P}(A) > 0$. Let us denote by $\Theta_n = \frac{1}{n} \sum_{k=1}^n \log \frac{1 + \xi_k}{1 + \xi_{k-1}}$ the left side of (4.1) and let $A_n = \{\Theta_n \leq 2 - \eta/2\}$. By the convergence $\Theta_n \rightarrow \Theta$, the probability of the symmetric difference $A \Delta A_n$ tends to 0. Fix $\delta > 0$ smaller than $\mathbb{P}(A)$ and let n be large enough so $\mathbb{P}(A \Delta A_n) < \delta$. In particular $\mathbb{P}(A_n) > \mathbb{P}(A) - \delta > 0$. Note that the event A_n belongs to $\mathcal{F}_n = \sigma(\mathcal{T}_0^\bullet, (B_t)_{t \leq \xi_n}, (x_{\xi_k})_{k \leq n})$ the filtration generated by Rémy's diffusion up to the n 'th resampling time.

Now let us fix $\varepsilon > 0$. Applying Lemma 4.2 and Proposition 4.4 to the tree $\mathcal{T}_{\xi_n}^\bullet$ shows that with probability $c_{\varepsilon, \eta} > 0$, the following holds conditionally on $\mathcal{F}_n$: there exists a time ξ_N for some finite stopping time $N \geq n$ at which the tree is ε -good and the $\tilde{\Theta}$ corresponding to the good Brownian subtree satisfies $\{\tilde{\Theta} > 2 - \eta\}$, and after that all resampled points belong to the good part. On this event, the bad part grows at a lower rate, of order 2^{2n} , so the average growth rate Θ_n of the entire tree converges almost surely to $\tilde{\Theta} > 2 - \eta$. This shows that $\mathbb{P}(A_n \cap A^c) \geq c_{\varepsilon, \eta} \mathbb{P}(A_n) \geq c_{\varepsilon, \eta} (\mathbb{P}(A) - \delta)$. Since this probability is also upper bounded by $\mathbb{P}(A \Delta A_n) < \delta$, then we must have $c_{\varepsilon, \eta} (\mathbb{P}(A) - \delta) < \delta$. Since δ can be chosen arbitrarily small without affecting $c_{\varepsilon, \eta}$ and $\mathbb{P}(A)$, this cannot always be true, unless $\mathbb{P}(A) = 0$.

We have thus shown that $\mathbb{P}(\Theta \leq 2 - \eta) = 0$ for each $\eta > 0$ such that $e^{2-\eta} > 4$ and such that $\mathbb{P}(\Theta = 2 - \eta) = 0$. We may choose an decreasing sequence of such η 's that converge to 0 to conclude that $\mathbb{P}(\Theta < 2) = 0$. $\square$

Bouncing back one more time, we may now improve Proposition 4.4 using this improved version of (4.1) by showing that the event in that proposition occurs with high probability when ε is small enough.

Corollary 4.6 (The Brownian subtree does take over). *The lower bound $c_\varepsilon > 0$ from Proposition 4.4 tends to 1 as $\varepsilon \rightarrow 0$.*

Proof. Examining the proof of Proposition 4.4 shows that, now that we know that $\Theta = 2$, the key point is to only resample points in the good part, which occurs with conditional probability

$$\prod_{n \geq 1} \frac{\gamma_{\xi_n}}{\beta_{\xi_n} + \gamma_{\xi_n}} = \prod_{n \geq 1} \frac{1}{1 + \beta_{\xi_n}/\gamma_{\xi_n}},$$

given their respective masses. Our claim then reduces to showing that this product is close to 1 with high probability when ε is arbitrarily small.

Recall that the bad mass β_{ξ_n} is distributed as

$$(1 - m) + h^2 \tau_{2^{n-1}}^2 \leq \varepsilon(1 + \tau_{2^{n-1}}^2),$$

while the good mass is

$$\gamma_{\xi_n} \geq (1 - \varepsilon) \prod_{k=1}^n \frac{1 + \xi_k^1}{1 + \xi_{k-1}^1} = (1 - \varepsilon) e^{n\Theta_n^1},$$

where the last equality serves as the definition of Θ_n^1 .

It is important to note here that $\tau_{2^{n-1}}^2$ and Θ_n^1 do not depend on ε . Fix a small $\delta > 0$. On the one hand $1 + \tau_{2^{n-1}}^2$ grows slower than $2^{n(2+\delta)}$ and the other hand, by Corollary 4.5, we know that $\Theta_n^1 \rightarrow 2$ almost surely, so $e^{n\Theta_n^1}$ grows faster than $e^{n(2-\delta)}$. It follows that if δ is small enough, then the ratio $\beta_{\xi_n}/\gamma_{\xi_n}$ is upper bounded by $\varepsilon c^n Z$ with $c = 2^{(2+\delta)} e^{-(2-\delta)} < 1$ and where Z is an almost surely finite random variable independent of ε . This implies that indeed the product $\prod_{n \geq 1} \frac{1}{1 + \beta_{\xi_n}/\gamma_{\xi_n}}$ is positive and tends to 1 in probability as $\varepsilon \rightarrow 0$. $\square$

4.2 Proof of ergodicity

We now have all the ingredients to prove the claimed unique ergodicity, see [HMS11] for general arguments of this kind.

Proof of Theorem 1.2. Let $\mathcal{T}_0^\bullet$ be an arbitrary marked tree with mass 1 and a marked point different from the root and let $\varepsilon > 0$ be arbitrarily small. According to Lemma 4.2, almost surely there exists a finite stopping time t_0 , of the form ξ_N , such that the rescaled tree $\mathcal{S}_{t_0}^\bullet$ is ε -good. Applying Proposition 4.4 to this tree $\mathcal{S}_{t_0}^\bullet$ shows that with probability bounded below by $c_\varepsilon > 0$, all subsequent resampled points will belong to the good Brownian subtree which will then (by Corollary 4.5) grow more rapidly than the rest. Hence the GHP distance between the rescaled entire tree $(1 + t_0)^{-1/2} \cdot \mathcal{T}_{t_0}^\bullet$ and the good part $\gamma_{t_0}^{-1/2} \cdot \mathcal{G}_{t_0}^\bullet$ tends to 0.

According to Theorem 3.3 and Lemma 2.3, since the good part at time t_0 is a marked Brownian CRT conditionally on its size, then this remains true at each subsequent time. This proves the convergence of the process $\mathcal{S}_t^\bullet = (1 + t)^{-1/2} \cdot \mathcal{T}_t^\bullet$ to the standard Brownian CRT on this event with probability c_ε , and then in general since c_ε can be made arbitrarily close to 1 by taking ε small enough by Corollary 4.6. $\square$

References

- [ABBGM17] Louigi Addario-Berry, Nicolas Broutin, Christina Goldschmidt, and Grégory Miermont. The scaling limit of the minimum spanning tree of the complete graph. *Ann. Probab.*, 45(5):3075–3144, 2017.
- [ADH13] Romain Abraham, Jean-François Delmas, and Patrick Hoscheit. A note on the Gromov-Hausdorff-Prokhorov distance between (locally) compact metric measure spaces. *Electron. J. Probab.*, 18:no. 14, 21, 2013.
- [Ald93] David Aldous. The continuum random tree. III. *Ann. Probab.*, 21(1):248–289, 1993.
- [BBJ17] Axel Bacher, Olivier Bodini, and Alice Jacquot. Efficient random sampling of binary and unary-binary trees via holonomic equations. *Theor. Comput. Sci.*, 695:42–53, 2017.
- [CCFA26] Alessandra Caraceni, Nicolas Curien, William Fleurat, and Twigt Adrianus. Sweet Trims are made of Threes - A càdlàg erasure of the Brownian tree. *Preprint available at arXiv:2604.14138*, 2026.
- [CCHS22] Ajay Chandra, Ilya Chevyrev, Martin Hairer, and Hao Shen. Langevin dynamic for the 2D Yang-Mills measure. *Publ. Math., Inst. Hautes Étud. Sci.*, 136:1–147, 2022.
- [CH13] Nicolas Curien and Bénédicte Haas. The stable trees are nested. *Probab. Theory Related Fields*, 157(3-4):847–883, 2013.
- [DLGo5] Thomas Duquesne and Jean-François Le Gall. Probabilistic and fractal aspects of Lévy trees. *Probab. Theory Related Fields*, 131(4):553–603, 2005.
- [EL15] Alison Etheridge and Cyril Labbé. Scaling limits of weakly asymmetric interfaces. *Commun. Math. Phys.*, 336(1):287–336, 2015.
- [EPWo6] Steven Evans, Jim Pitman, and Anita Winter. Rayleigh processes, real trees, and root growth with re-grafting. *Probab. Theory Related Fields*, 134(1):81–126, 2006.
- [EWo6] Steven Evans and Anita Winter. Subtree prune and regraft: A reversible real tree-valued Markov process. *The Annals of Probability*, 34(3):918–961, 2006.
- [FGOR93] Philippe Flajolet, Zhicheng Gao, Andrew Odlyzko, and Bruce Richmond. The distribution of heights of binary trees and other simple trees. *Comb. Probab. Comput.*, 2(2):145–156, 1993.
- [FPRW23] Noah Forman, Soumik Pal, Douglas Rizzolo, and Matthias Winkel. The Aldous diffusion: a stationary evolution of the Brownian CRT. *Preprint available at arXiv:2305.17269*, 2023.
- [HMS11] Martin Hairer, Jonathan Mattingly, and Michael Scheutzow. Asymptotic coupling and a general form of Harris’ theorem with applications to stochastic delay equations. *Probab. Theory Relat. Fields*, 149(1-2):223–259, 2011.
- [LGo5] Jean-François Le Gall. Random trees and applications. *Probab. Surv.*, 2:245–311, 2005.
- [LMW20] Wolfgang Löhr, Leonid Mytnik, and Anita Winter. The Aldous chain on cladograms in the diffusion limit. *Ann. Probab.*, 48(5):2565–2590, 2020.
- [McD98] Colin McDiarmid. Concentration. In *Probabilistic methods for algorithmic discrete mathematics*, volume 16 of *Algorithms Combin.*, pages 195–248. Springer, Berlin, 1998.
- [Pit06] Jim Pitman. Combinatorial stochastic processes. In *École d’été de probabilités de Saint-Flour, XXXII – 2002*, volume 1875 of *Lecture Notes in Math.*, pages ix + 256. Springer, Berlin, 2006.
- [Rém85] Jean-Luc Rémy. Un procédé itératif de dénombrement d’arbres binaires et son application à leur génération aléatoire. *RAIRO, Inf. Théor.*, 19:179–195, 1985.
- [Zam01] Lorenzo Zambotti. A reflected stochastic heat equation as symmetric dynamics with respect to the 3-d Bessel bridge. *J. Funct. Anal.*, 180(1):195–209, 2001.