Topological spaces of trees as state spaces for stochastic processes

Habilitationsschrift

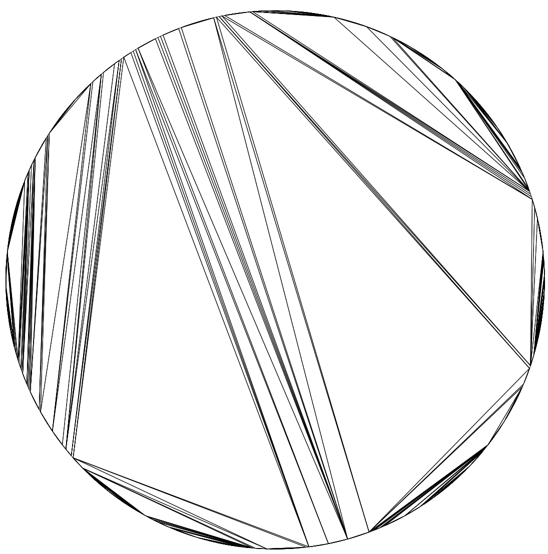
Der Fakultät für Mathematik
der Universität Duisburg-Essen
eingereicht von

Dr. Wolfgang Löhr

zur Erlangung der Lehrbefähigung
im Lehrgebiet

Mathematik

am 17. August 2020 angenommen



Abstract

This thesis presents some aspects of the theory of continuum limits of tree-valued stochastic processes, as well as limits of stochastic processes on (possibly random) trees as the size of the underlying trees tends to infinity. In both situations, in order to get the full picture, one has to work in a suitable, sufficiently nice, abstract topological state space of (continuum) trees. The focus of the articles collected in this thesis is on the construction and analysis of such state spaces, as well as their usage to obtain general limit theorems.

The state spaces fall into two categories. Firstly, the more classical framework using metric measure spaces and various generalizations. Secondly, a new, more algebraic/combinatorial approach of coding the tree-structure by a branch point map without reference to a canonical or pre-specified metric. We call these objects algebraic measure trees. An important classical tool is to code metric measure trees by excursions. Similarly, we can code binary algebraic measure trees by triangulations of the circle.

Some stochastic processes are considered as examples to highlight the applicability and usefulness of the developed state spaces and tools. These are tree-valued multi-type Fleming-Viot processes, tree-valued pruning processes, random walks on trees, and the Aldous chain on cladograms. In all these cases we obtain limit theorems in suitable state spaces of measure trees.

Acknowledgements

First of all, I want to thank Anita Winter for all the advice she gave me, for introducing me to the topic of tree-valued stochastic processes, for the discussions, the plenty of fascinating ideas and her incredible optimism, which motivated me to carry on. I also thank my co-authors of the articles forming the content of this thesis, Siva Athreya, Sandra Kliem, Leonid Mytnik, Thomas Rippl, Guillaume Voisin and Anita Winter, for the not only fruitful but also most enjoyable collaborations. The same applies to my collaborators Hui He, Anton Klimovsky, Sara Mazzonetto and Jan Nagel, who are not co-authors yet, but the discussions with whom certainly contributed to this thesis in one way or another.

Further thanks go to my colleagues Fabian Gerle, Roland Meizis and Josué Nussbaumer for scientific discussions, and to them as well as to a lot of other present and former colleagues (in particular including those mentioned above) for the great, friendly atmosphere they created – and are still creating – in our working group. I am also indebted to Nihat Ay for still supporting me even though I changed the scientific field a bit. Last but not least many thanks to Dagmar Goetz, the secretary of our group, who generally does a great and indispensable job, and on the same account to Nicole Obszanski.

Essen, November 27, 2019

Contents

Part I

Introduction

Overview

About the thesis

This cumulative habilitation thesis is organized as follows. Part II, the main part, is a collection of eight articles. They are published in different journals, or on the arXiv preprint server and submitted to a journal. The reprints included here differ from the published versions in formatting, typesetting and minor corrections, but not in essential content. Preceding the paper reprints, I give in Section 1 a short overview over the topics and common theme of the articles, as well as a summary of some of their main results within a broader context in Sections 2 – 6.

1 The big picture

Graph-theoretic trees are abundant in mathematics and its applications, from computer science to theoretical biology. Often, the considered trees are finite but huge. Think, for instance, of search trees in big data analysis or the "tree of live" containing more than one million known species (mostly insects, see [ROO$^+$19]). Therefore, a natural question is how to define limit objects and appropriate notions of convergence, as the size of the trees tends to infinity. On the one hand, there are *local* approaches yielding countably infinite graphs, or generalized so-called graphings with a Benjamini-Schramm-type approach (going back to [BS01], see also [Lov12, Part 4]). On the other hand, if one takes a more *global* point of view, as we are doing here, the predominant (but, as we will see, not the only possible) approach is to consider graph-theoretic trees as metric spaces equipped with the (rescaled) graph distance. Then the limit objects are certain "tree-like" metric spaces, most prominently so-called $\mathbb{R}$-*trees* introduced by Tits in [Tit77], and we are talking about convergence of metric spaces.

1.1 Trees as metric (measure) spaces

In the field of metric geometry, working with spaces of (sufficiently well-behaved, e.g. compact) metric spaces, and turning them into metric spaces themselves, dates back to the invention of the Gromov-Hausdorff metric in [Gro81, GP81]. It was also realized that, for many applications, more structure is required on the metric space, namely a measure (on its Borel σ-algebra). Hence, theory for spaces of *metric measure spaces* was developed, most notably by Gromov in [Gro99, Chapter $3\frac{1}{2}$], and with a different perspective by Sturm in [Stu06]. A more recent and detailed treatment is given by Shioya in [Shi16]. Originally, the measure was motivated by the volume measure on Riemannian manifolds. It can, e.g., be used to generalize Ricci curvature from manifolds to more general spaces, study transportation problems ([Vil09, LV09]), and define generalizations of Laplace operators as well as the associated diffusion processes ([Stu98, KS05]).

More recently, metric measure spaces have been becoming important in probability theory and its applications. While the research in the area of metric geometry primarily focuses on geodesic spaces with lower curvature bounds (e.g. *Alexandrov spaces* or Ricci curvature bounds) that behave manifold-like, here tree-like spaces, in particular $\mathbb{R}$-trees, are more important. These are loop-free geodesic spaces or, equivalently, 0-hyperbolic connected spaces. The curvature of $\mathbb{R}$-trees in branch points is $-\infty$, thus they do not fall into the usual classes of spaces with curvature bounds. They are, however, essentially one-dimensional (the Hausdorff dimension can be infinite, but the topological dimension is one), which simplifies the analysis. For instance, it is not too difficult to define a gradient (w.r.t. an arbitrary fixed point, the root) and, via a partial integration formula, a Laplace operator. This way, using Dirichlet form theory, Athreya, Eckhoff and Winter constructed Brownian motions on $\mathbb{R}$-trees (depending on a speed measure on the tree) in [AEW13]. See also **[6]** for a generalization to so-called metric trees. This naturally leads to the question if there is a "Donsker theorem for trees", i.e. does a sequence of random walks on discrete trees with an increasing number of vertices (after suitable rescaling) converge in Skorokhod path space to Brownian motion on a limiting $\mathbb{R}$-tree? And if so in what sense? What does "convergence in path space" actually mean if the stochastic processes are defined on different spaces? These questions are the topic of **[6]**, and obviously answering them requires a suitable topological space of trees, which we constructed and analyzed in **[5]**.

Our philosophy, here and elsewhere, is to prove – instead of a collection of specific cases – a general *continuity theorem* of the type

$$\text{Whenever the initial condition converges,} \\ \text{the corresponding processes converge, weakly in Skorokhod path space.} \tag{$*$}$$

The Skorokhod path space of càdlàg paths is of course equipped with the standard Skorokhod J_1-topology. Here, the "initial condition" includes, apart from the starting point of the process, the underlying tree. And indeed, in **[6]** we show a very general invariance principle of this type: For every (sufficiently nice) metric measure tree (T, r, ν) there is a unique associated stochastic process, called *speed-ν motion* on the tree, in natural scale with the given speed measure ν. This process can be defined via its occupation time formula (together with the strong Markov property), or by its Dirichlet form. It coincides with the random walk and Brownian motion in the cases of discrete trees and $\mathbb{R}$-trees, respectively, and turns out to depend continuously (in path space) on the underlying metric measure tree, provided the space of metric measure trees is equipped with the Gromov-Hausdorff-vague topology introduced in **[5]**. Of course, we have to make precise what we mean by this, see Section 6.

Long before (topological) spaces of metric measure spaces were introduced into probability theory by Greven, Pfaffelhuber and Winter in [GPW09], limits of random trees were extensively studied by Aldous in the seminal sequence of papers [Ald91a, Ald91b, Ald93] by embedding them into ℓ_1. In particular, he showed that finite-variance Galton-Watson trees conditioned to have N vertices, with suitably rescaled edge-lengths, tend in law to a particular, compact, binary random tree called the *Brownian continuum random tree (CRT)*. His work, in particular [Ald93], contains many of the ideas which were put into the more elegant framework of metric measure spaces later. Another way of thinking about continuum trees is to see them as coded by a continuous excursion, which is the "contour process" of the tree. In this way, the Brownian CRT is coded by the normalized Brownian excursion. This idea allows one to use the more familiar theory of random continuous functions (with uniform convergence on compacta on the space of functions), a viewpoint taken around the same time by Le Gall in [LG91, LG93]. He also obtained the theorem of Aldous mentioned above with quite different techniques. Continuing this line of research, Le Gall and Duquesne also showed that Galton-Watson trees with infinite variance converge to so called Lévy trees, which can be defined using Lévy processes. See [DLG02, Duq03]. Despite being quite successful, the approach with continuous functions has the drawbacks that the function coding a continuum tree is not unique, and that using the uniform topology on excursions leads to a rather strong notion of convergence for the trees, as it turned out too strong for some applications in the theory of tree-valued processes.

1.2 Tree-valued stochastic processes

These reasons lead to the introduction of the space of metric measure spaces as state space for tree-valued processes into probability theory by Greven, Pfaffelhuber and Winter in [GPW09]. They introduced the so-called *Gromov-weak* topology on it, which can be defined by a complete metric similar to the Gromov-Hausdorff metric,[1] just replacing the Hausdorff metric for subsets by the Prokhorov metric for the measures. But, maybe even more importantly, it can alternatively be defined as the weak topology induced by a natural class of test functions, called *polynomials*. It follows from a general theorem due to Le Cam that the set of polynomials is even convergence determining for measures, as observed in [1]. This simplifies many weak convergence arguments considerably. In [4], we extended Le Cam's theorem to infinite, boundedly-finite measures and weak$^\#$ convergence, which finds applications, e.g., in the decomposition analysis of branching trees ([GGR19]). Because several aspects (especially compactness criteria) of the space of metric measure spaces with Gromov-weak topology were quite parallel to the space of metric measure spaces with Gromov's metric $\underline{\square}_1$ introduced in [Gro99], the two topologies were conjectured to be actually identical. This is shown to be true in [1]. A major technical difficulty for working with spaces of metric measure spaces is that they are generally far from being locally compact, even though many of them are Polish spaces.

Discrete-tree-valued stochastic processes play an important rôle for algorithms that sample random trees, e.g. for Markov chain Monte Carlo simulations to fit phylogenetic trees to data. They also appear in population or speciation models in mathematical biology, where one considers an evolving genealogical tree in a model. In this case, the rôle of the measure in the continuum limit is especially clear: it is a way of sampling a finite sub-population (after all, statistics of randomly drawn, comparatively small sub-populations is all one can compare to actual measurements) and encodes a population density. The construction and investigation of scaling limits of such tree-valued Markov chains within a metric space setup started with the continuum analogs of the Aldous-Broder-algorithm for sampling a uniform spanning tree from the complete graph, which Evans, Pitman and Winter constructed in [EPW06], and of the tree-valued subtree-prune and regraft Markov chain used for the reconstruction of phylogenetic trees, which Evans and Winter constructed in [EW06]. Note that proving convergence of (rescaled) Markov chains in an abstract state space of metric measure spaces is notoriously difficult, and also [EW06] does not prove such a convergence; it constructs (and analyzes) the potential limit process.

Proving convergence of tree-valued, Moran-type population models turned out to be somewhat less hopeless: Greven, Pfaffelhuber and Winter constructed in [GPW13] an infinite population, tree-valued Fleming-Viot process and showed convergence of the finite population, tree-valued Moran models to it in the space of metric measure spaces with Gromov-weak convergence. This line of research was continued in [Pio11, Glö13]. In order to include mutation and selection, additional information about the *types* of the individuals had to be encoded in the state space, which lead to the development of the space of *marked metric measure spaces* with *marked Gromov-weak* topology in [DGP11]. In this space, the tree-valued Fleming-Viot process with mutation and selection was constructed and obtained as limit of finite population models in [DGP12, DGP13].

[1] The Gromov-Hausdorff distance between two metric spaces is the infimal Hausdorff distance between the images of isometric embeddings into a common metric space, where the infimum is taken over all embeddings.

Kliem and myself showed in [2] that this convergence actually occurs in a subspace, where we can exclude the pathology of genetically indistinguishable individuals having different types in the limit process. We did so by first proving general criteria for checking this, which are applicable also in different situations ([KW19]), and then verifying the conditions. A lookdown representation is obtained in [Guf18], and interacting processes in different geographic locations are considered by Greven, Sun and Winter in [GSW16]. Competition between the individuals, which is a major difficulty because it destroys the branching property, is considered in [KW19].

Another approach to incorporate competition was pursued by Berestycki, Fittipaldi and Fontbona using an excursion setup in [BFF18]. They use a pruning process of the genealogical tree without competition in order to incorporate additional deaths due to competition between the individuals (the pruning rate has to be adaptive and solve a fixed point equation because of the lack of the branching property induced by competition). Continuum analogues of tree-valued pruning processes had been considered before by Aldous and Pitman in [AP98] and by Abraham, Delmas, Serlet, and Voisin in [AS02, ADV10, AD12]. They considered different random trees (Brownian CRT, Lévy trees with and without Brownian part) and different pruning intensities (node-pruning, skeleton-pruning). They did not, however, show convergence of the discrete to the continuous pruning processes, which is the objective of [3]. Consistently with our philosophy mentioned above in ($*$), we did neither consider a specific class of random trees (a priori), nor treat discrete and continuous pruning processes with various pruning intensities differently. Instead, we obtained them as instances of one and the same pruning process with different initial conditions. This means that the pruning measure had to become part of the state. A major difficulty here is that natural pruning measures, such as the length measure on the Brownian CRT, are not locally finite. Therefore, we introduced in [3] a new state space of *metric bi-measure trees* with two measures: the usual finite sampling measure and a second, potentially locally infinite pruning measure. In this state space, with a convenient topology which we called *leaf-sampling weak-vague topology*, we were able to define THE pruning process and show that its paths (with Skorokhod topology) depend continuously on the initial condition. Further convergences of pruning processes were obtained afterwards in [HW14, HW19].

1.3 Continuum trees without a metric

Even though considering trees as metric spaces has been a very successful approach for proving limit theorems of tree-valued stochastic processes, and all of the examples mentioned above fall into this category, there are some draw-backs in some situations. First, checking compactness or tightness criteria for (random) metric (measure) spaces is not always easy, and some natural sequences of trees do not converge as metric (measure) spaces with a uniform rescaling of edge-lengths, while intuitively there should be some "limit". This was for instance observed by Curien in [Cur14], where he also suggested reducing the considered structure and weakening the topology on the space of trees in order to obtain better compactness properties. Second, the metric is often less canonical than the tree structure in situations where it is not clear that every edge should have the same length. For instance, in a phylogenetic tree, edges might correspond to very different evolutionary time spans. Thus, the length of an edge should rather be proportional to this time span than be the same for all edges, and using the graph distance might lead to "wrong", misleading scaling limits (or their non-existence). The time spans are, however, much more difficult to estimate from data than the qualitative information about the tree structure. Third, one might want to preserve certain functionals of the tree structure in the limit. For instance, the limit of binary trees is not always binary in the metric space setup, because several branch points (even infinitely many) may collapse to one branch point of arbitrarily high degree if their distance tends to zero. Also, functions such as the centroid function[2] used in [Ald00] are not continuous on spaces of metric measure trees with any of the common topologies.

In [7], we address these issues and provide a new state space of *algebraic measure trees*, which do not possess a pre-specified or canonical metric. In this state space, the subset of binary measure trees is compact (in particular closed), the centroid function is continuous, and the state only contains information about the "tree structure", so that we do not need to know any information about distances to work in this state space. We code this tree structure by a branch point map, which associates to a triplet of vertices the corresponding branch point, so that an algebraic tree is a set T with a function $c\colon T^3 \to T$ satisfying certain axioms. We have shown that, under a separability constraint, every algebraic tree can be represented by a (non-unique) metric tree. Using this result, one sees that algebraic measure trees are intimately related to the random exchangeable *didendritic systems* introduced by Evans, Grübel, and Wakolbinger in [EGW17], and to the *mass-structural* equivalence classes of $\mathbb{R}$-trees introduced independently of us in [For18]. An important tool is the coding of binary algebraic measure trees by *triangulations of the circle*, in the same spirit as metric trees can be coded by excursions. In principle, such a coding was introduced by Aldous in [Ald94a, Ald94b], but it had never been identified which structure precisely is coded by a triangulation. We show in [7] that it is the structure of a binary algebraic measure tree and obtain a surjective coding map.

In this new state space of algebraic measure trees, we have shown in [8] the path space convergence of the *Aldous chain on cladograms* considered in [Ald00] to an ergodic Feller process, which we call *Aldous diffusion*. Thereby, we solved an old open problem of Aldous.

[2]Given a tree T with measure μ, and a vertex $v \in T$, let $m_T(v)$ be the mass of the largest of the components obtained by splitting T at v. Then $f(T,\mu) := \inf_{v\in T} m_T(v)$ is the centroid function.

Main contributions

In this chapter, I summarize some of what I think are the main contributions of the articles collected in this thesis. The theorems in this introductory part are not in one-to-one correspondence to theorems in the papers, but results distributed over several lemmata, propositions and theorems of one article may be collected into a single theorem. Furthermore, some of them are substantially simplified to reduce their technicality (I hope the reader will forgive me that maybe even some sloppiness resulted from the simplification) and do not give the complete results.

2 Spaces of metric measure spaces

A *metric measure space* (*mm-space*), sometimes also known as *Gromov metric triple*, as used in [Gro99] is a complete, separable metric space (X, r) together with a finite measure μ on its Borel σ-algebra. The separability assumption is of course crucial, whereas the completeness assumption is – as far as spaces of mm-spaces are concerned – more for convenience, because the necessary formation of equivalence classes of mm-spaces naturally identifies an incomplete space with its completion. In the probabilistic literature, most notably in [GPW09], the measure μ is often restricted to be a probability measure, but for most of the theory with the topologies we are considering, this restriction makes no essential difference. For convenience and the purpose of this introduction, I will assume μ to be a probability unless specified otherwise. The finiteness assumption for μ, however, is more important and mm-spaces with certain (usually locally-finite) infinite measures have also been considered, for example in [Stu06], [Vil09], [ADH13], [5], [6].

In order to define a topology on a space of mm-spaces, one needs to have a suitable equivalence relation between mm-spaces and consider the set of *equivalence classes* instead of the mm-spaces themselves. Of course, in praxis, one usually proceeds the other way round: given a notion of distance between or convergence of mm-spaces, one identifies mm-spaces that cannot be distinguished. For most of the topologies considered here, two mm-spaces $\mathcal{X} = (X, r, \mu)$, $\mathcal{X}' = (X', r', \mu')$ are *equivalent* (or *isomorphic*) if there is a measure preserving isometry $f\colon \operatorname{supp}(\mu) \to X'$, i.e.

$$\mu' = \mu \circ f^{-1} \qquad \text{and} \qquad r(x, y) = r'\big(f(x), f(y)\big) \;\; \forall x, y \in \operatorname{supp}(\mu).$$

We denote the space of equivalence classes of mm-spaces by $\mathbb{M}$, and usually do not distinguish between an equivalence class and a representative. This approach neglects parts of the space that are not "seen" by the measure. An alternative, which we call *strongly equivalent*, is to require existence of a measure preserving, bijective isometry between the whole metric spaces.

2.1 Topologies on $\mathbb{M}$

The idea of Gromov's metric on $\mathbb{M}$ is the following. By the Skorokhod representation theorem (e.g. [Bog07, Thm. 8.5.4]), every measure μ on a Polish space X can be parameterized by a measure preserving function $\varphi\colon [0, 1] \to X$, where the unit interval is equipped with Lebesgue measure. To define a distance between mm-spaces, Gromov evaluates for every choice of parametrizations of the measures the distance between the pullbacks of the metrics by the respective parametrizations, using a suitable metric for functions on $[0,1]^2$. Then the parametrizations are chosen optimally. More formally, denoting by $\mathcal{F}(X,r,\mu)$ the set of measure preserving parametrizations of μ:

Definition 1 (Gromov's $\underline{\square}_1$ metric; [Gro99]). Let $\mathcal{X}_i = (X_i, r_i, \mu_i) \in \mathbb{M}$, $i = 1,2$. For $\varphi \in \mathcal{F}(\mathcal{X}_i)$, let $d_i^\varphi(s,t) := r_i(\varphi(s), \varphi(t))$. Then,

$$\underline{\square}_1(\mathcal{X}, \mathcal{X}') := \inf_{\varphi_i \in \mathcal{F}(\mathcal{X}_i)} \square_1(d_1^{\varphi_1}, d_2^{\varphi_2}),$$

where for $f, g\colon [0,1]^2 \to \mathbb{R}_+$ and the outer Lebesgue measure λ^* on $[0,1]$

$$\square_1(f,g) := \inf\Big\{ \varepsilon > 0 \,\Big|\, \exists A \subseteq [0,1] : \|f\restriction_{A\times A} - g\restriction_{A\times A})\|_\infty \le \varepsilon,\ \lambda^*(X \setminus A) \le \varepsilon \Big\}.$$

One of the ways to define the Gromov-weak topology is by the Gromov-Prokhorov metric. The logic is precisely the other way round: instead of parameterizing the measure optimally and using a distance between the pullbacks of the metrics, the metric spaces are embedded optimally into a common space and the distance between the pushforwards of the measures is used.

Definition 2 (Gromov-Prokhorov metric; [GPW09]). Let $\mathcal{X}_i = (X_i, r_i, \mu_i) \in \mathbb{M}$, $i = 1,2$. The *Gromov-Prokhorov metric* is

$$d_{\mathrm{GP}}(\mathcal{X}_1, \mathcal{X}_2) := \inf_{f_1, f_2} d_{\mathrm{Pr}}\big(\mu_1 \circ f_1^{-1}, \mu_2 \circ f_2^{-1}\big),$$

where the infimum is taken over all isometries $f_i\colon X_i \to X$ into a common separable metric space (X,r). Here, d_{Pr} denotes the Prokhorov metric.

This definition does not lead to set-theoretic problems, because separable metric spaces have the cardinality of at most the continuum, and hence we may assume w.l.o.g. the set X to be $[0,1]$. One of the main results in [GPW09] is that the topology induced by the Gromov-Prokhorov metric coincides with the Gromov-weak topology, which is defined as the weak topology induced by a certain set of test functions. The possibility to represent the topology by a tractable set of test functions is one of the main reasons why the Gromov-weak topology and its derivatives and extensions have become so successful in probability theory.

Definition 3 (Polynomials & Gromov-weak topology; [GPW09]). A *polynomial* (on $\mathbb{M}$) is a function $\Phi\colon \mathbb{M} \to \mathbb{R}$ of the form

$$\Phi(\mathcal{X}) = \int_{X^n} \phi\big(\big(r(x_i, x_j)\big)_{i,j \le n}\big)\ \mu^{\otimes n}(\mathrm{d}x),$$

where $n \in \mathbb{N}$ and $\phi \in \mathcal{C}_b\big(\mathbb{R}^{n\times n}\big)$. Let Π be the set of such functions. The *Gromov-weak topology* is the topology induced by Π on $\mathbb{M}$.

The set Π of polynomials is an algebra, by definition inducing Gromov-weak topology, but it is straight-forward to see that it is not dense in the set $\mathcal{C}_b(\mathbb{M})$ of bounded continuous functions on $\mathbb{M}$, as observed in [1]. Nevertheless, it follows from a general, apparently (at least a few years ago) not so well-known theorem due to Le Cam that Π is *convergence determining* for probability measures on $\mathbb{M}$, see [1]. This means that mm-space valued random variables X_n converge in law with respect to Gromov-weak topology to an mm-space valued random variable X if and only if

$$\mathbb{E}\big(\Phi(X_n)\big) \underset{n\to\infty}{\longrightarrow} \mathbb{E}\big(\Phi(X)\big) \qquad \forall \Phi \in \Pi,$$

which is obviously a very nice property to have when working with convergence of random mm-spaces, and would have substantially simplified some arguments in [GPW13] had it been known. We note that it had already been shown with some effort and a much less flexible and extensible proof by Depperschmidt, Greven and Pfaffelhuber in [DGP11]. One main result of [1] is:

Theorem 4 ([1]). *$\square_1$ and d_{GP} are bi-Lipschitz equivalent, and the set Π of polynomials is convergence determining for probability measures on $\mathbb{M}$ with the induced Gromov-weak topology.*

2.2 Marked metric measure spaces & tree-valued Fleming-Viot

In order to model types or locations of individuals in population models, additional marks were included in the state space by Depperschmidt, Greven and Pfaffelhuber in [DGP11]. The mark space is a fixed, complete, separable metric space (I, d), and an *I-marked metric measure space* (*mmm-space*) is a complete, separable metric space (X, r) with a measure μ on $X \times I$. Then equivalence and marked Gromov-weak convergence are defined analogously to the case of ordinary mm-spaces, with the difference that the mark space has to be fixed by the isometry/embedding. As is the case for $\mathbb{M}$ with Gromov-weak topology, the space $\mathbb{M}_I$ of equivalence classes of mmm-spaces with marked Gromov-weak topology is a Polish space.

Because the marks are modelled by the measure instead of a separate function,[3] it is not true in general that a point $x \in X$ possesses a unique mark, which could be given by a *mark function* $\kappa\colon X \to I$. Instead, there is only a kernel K from X to I obtained by the disintegration $\mu = \nu \otimes K$ for the marginal ν of μ on X. It is a natural question to ask in which cases this kernel is actually deterministic, i.e. induced by a mark function.

Definition 5 (Fmm-spaces; [2]). We call an mmm-space (X, r, μ) *functionally marked* (*fmm-space*) if there is a measurable mark function $\kappa\colon X \to I$ such that $\mu(\mathrm{d}x, \mathrm{d}u) = \nu(\mathrm{d}x)\delta_{\kappa(x)}(\mathrm{d}u)$. Let $\mathbb{M}_I^{\mathrm{fct}} \subseteq \mathbb{M}_I$ be the set of (equivalence classes of) fmm-spaces.

The problem is that $\mathbb{M}_I^{\mathrm{fct}}$ is not closed (nor open) in $\mathbb{M}_I$. Thus, it is not obvious if $\mathbb{M}_I^{\mathrm{fct}}$ (with marked Gromov weak topology) is a Polish space. And even though Markov chains of population models for finite populations naturally take values in $\mathbb{M}_I^{\mathrm{fct}}$, it is not clear if the same is true for a marked Gromov-weak limit process. It was claimed with a false proof in [DGP13] that the tree-valued Fleming-Viot with mutation and selection constructed in [DGP12] stays in $\mathbb{M}_I^{\mathrm{fct}}$. We were able to prove this claim. In order to do so, we derived general criteria which can be checked to verify that a marked Gromov-weak limit of $\mathbb{M}_I^{\mathrm{fct}}$-valued processes is $\mathbb{M}_I^{\mathrm{fct}}$-valued as well. They seem a bit technical (see Theorems 3.9 and 3.11 in [2]), but are general and tractable enough to be applicable also in other situations ([GSW16, KW19]).

Theorem 6 ([2]). *$\mathbb{M}_I^{\mathrm{fct}}$ with marked Gromov-weak topology is a Polish space, and the tree-valued Fleming-Viot process with mutation and selection $(X_t)_{t\geq 0}$ defined in [DGP12] admits a mark function at all times, i.e.*

$$\mathbb{P}\big(\big\{\, X_t \in \mathbb{M}_I^{\mathrm{fct}} \ \ \forall t > 0 \,\big\}\big) = 1.$$

2.3 Gromov-Hausdorff-weak and related topologies

We can also combine Gromov-weak (Gw) and Gromov-Hausdorff convergence. If we want to stay in $\mathbb{M}$, as opposed to considering strong equivalence classes, this means that we require only Hausdorff convergence of the *supports* of the measures (together with weak convergence of the measures) in

[3] This makes the general theory easier and more elegant. For a quite restrictive approach using mark functions see [Pio11].

an optimal embedding. The resulting *Gromov-Hausdorff-weak (GHw) topology* is non-separable on $\mathbb{M}$, therefore it makes sense to consider it on the subset $\mathbb{M}_c$ of compact mm-spaces (more precisely where the equivalence class contains a compact representative).

Closely related to the GHw topology on $\mathbb{M}_c$ is the *measured Hausdorff topology* introduced by Fukaya in [Fuk87]. It was rediscovered and shown to be Polish under the name *weighted Gromov-Hausdorff topology* by Evans and Winter in [EW06], and is most well-known under the name *Gromov-Hausdorff-Prokhorov (GHP) topology*, going back to [Mie09]. The difference is that it requires Hausdorff convergence of the *whole spaces* instead of only supports, which leads to different equivalence classes, defined by strong equivalence. Therefore, $\mathbb{M}_c$ with GHw topology is not homeomorphic to the space of strong equivalence classes of compact mm-spaces with GHP topology, but to the subspace of those spaces, where the measure has full support. This subspace is not GHP-closed, thus the Gromov-Hausdorff-Prokhorov metric is not complete on $\mathbb{M}_c$.

Gromov-weak convergence has the advantage of the nice, convergence determining family Π of polynomials which often makes proving convergence easier, but sometimes one is interested in the stronger Gromov-Hausdorff-weak convergence. For instance, convergence of supports is necessary for the invariance principle on trees in **[6]**. Therefore, it is natural to ask what is missing to conclude from Gw the GHw convergence. Uniform volume doubling, a standard assumption in metric geometry, is strong enough, but it is rather too strong and quite restrictive if we think about Lévy trees. In **[5]**, we obtained an if and only if condition in terms of a uniform lower mass bound which is much weaker than uniform volume doubling. More concretely, a set $A \subseteq \mathbb{M}$ satisfies a *uniform lower mass-bound* if

$$\inf_{\mathcal{X}\in A} \mathfrak{m}_\delta(\mathcal{X}) > 0 \qquad \text{for} \qquad \mathfrak{m}_\delta(\mathcal{X}) := \inf_{x\in\operatorname{supp}(\mu)} \mu\big(\overline{B}_r(x,\delta)\big), \tag{1}$$

where $\overline{B}_r(x,\delta)$ is the closed r-ball with radius δ around x. The methodology of first showing Gw convergence and then verifying (1) was central for Bahmidi, van der Hostad and Sen in [BvdHS18] to analyze the structure of the maximal components of the multiplicative coalescent. The last claim of the following Theorem is not stated explicitly in **[5]**, but follows directly from Corollary 4.3 together with Theorem 6.1 there.

Theorem 7 ([**5**]). *$\mathcal{X} = (X, r, \mu) \in \mathbb{M}$ is compact if and only if $\mathfrak{m}_\delta(\mathcal{X}) > 0$. $\mathbb{M}_c$ with Gromov-Hausdorff-weak topology is a Polish space, and a sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ in this space converges if and only if it converges Gromov-weakly and satisfies a uniform lower mass-bound. A set $A \subseteq \mathbb{M}_c$ is Gromov-Hausdorff-weakly relatively compact if and only if it satisfies a uniform lower mass-bound.*

2.4 Infinite measures

Considering mm-spaces with infinite, usually locally finite, measures becomes important for population models if one considers a population which is spread over an infinite geographic space, such that only the local population density is finite, whereas globally, there is an infinite population size. It is also important for our invariance principle for random walks on trees, as there the measure is the speed measure of the process, and we certainly want to include the classical Donsker case with the Brownian motion on $\mathbb{R}$ as limit process.

In **[4]**, we provided a tool for working with *boundedly-finite measures* (i.e. they are bounded on bounded sets, which is the same as locally finite if the underlying space is Heine-Borel) on separable metric spaces (note that we do not require Polishness). Namely, we extended Le Cam's theorem that a multiplicatively closed subset of $\mathcal{C}_b(X)$ is convergence determining for finite measures if it induces the topology of X. It also holds in the case of boundedly finite measures and *weak$^\#$ convergence*, with an additional non-vanishing hypothesis, which we need due to the fact that the one function is not integrable w.r.t. infinite measures. Our main tool is the classical embedding of the space into the Hilbert cube $[0,1]^{\mathbb{N}}$, which allows us to avoid assuming the measures to be Radon (inner compact-regular) or the space to be Polish. Our result has proven to be effective by Gloede Greven and Rippl in [GGR19], and by Basrak and Planinić in [BP19]. We also obtained sets of functions that are *separating* for boundedly finite measures under weaker assumptions.

In [5], we considered the space of pointed mm-spaces with boundedly-finite measures and extended both the Gromov-weak and the Gromov-Hausdorff-weak to a *Gromov-vague* and a *Gromov-Hausdorff-vague* topology, respectively. In less generality, a similar extension of the Gromov-Hausdorff-Prokhorov topology had been defined before by Abraham, Delmas and Hoscheit in [ADH13]. We obtained results similar to Theorem 7 for the vague instead of weak topologies, including the characterization of the difference between Gomov-vague and Gromov-Hausdorff-vague convergence, where we only had to replace the lower mass bound by a *local* lower mass bound.

3 Trees & coding by excursions

The standard continuum analog of graph-theoretic trees (with graph distance) are so-called $\mathbb{R}$-*trees*, which are "tree-like" metric spaces introduced in [Tit77], investigated under the heading *T-theory* by Dress and coauthors (a summary is given in [DMT96]) and generalized by Chiswell in [Chi01]. For an introduction with a view on probability theory, see the lecture notes [Eva08] by Evans. There are many equivalent definitions, the maybe most intuitive one is that the metric space has to be uniquely path-wise connected and geodesic. More convenient for us is the following one: it is a connected metric space (T,r) satisfying the *four point condition*

$$r(v_1,v_2)+r(v_3,v_4) \;\le\; \max\bigl\{\, r(v_1,v_3)+r(v_2,v_4),\; r(v_1,v_4)+r(v_2,v_3) \,\bigr\} \qquad \forall v_1,\dots,v_4\in T.$$

Not necessarily connected spaces satisfying the four point condition are called 0-*hyperbolic*. In [6], we introduced *metric trees* in order to treat $\mathbb{R}$-trees, discrete (graph-theoretic) trees and "mixtures" thereof simultaneously in the metric space setup. They are 0-hyperbolic, and for any $x,y,z\in T$, a corresponding "branch point" has to exist in T. One possibility to construct 0-hyperbolic spaces is to code them by excursions, see [Ald93, LG93, DLG02]. Often, excursions are assumed to be continuous or strictly positive in the interior, but we do not make these restrictions, because it restricts the possible (rooted) measure trees we can code. Instead, we consider the set

$$\mathcal{E} \;:=\; \bigl\{\, h\colon [0,1]\to\mathbb{R}_+ \;\big|\; h(0)=0,\ h \text{ lower semi-continuous} \,\bigr\}.$$

Given $h\in\mathcal{E}$, consider the pseudo metric on $[0,1]$ given by

$$d_h(s,t) \;:=\; h(s)+h(t)-2\inf_{u\in[s\wedge t,\, s\vee t]} h(u).$$

The quotient metric space $T_h := [0,1]/d_h$ is 0-hyperbolic and, due to lower semi-continuity, it is separable with measurable canonical projection $\pi_h\colon [0,1]\to T_h$ (see [1] for a short argument).

Definition 8 (The glue function)**.** The *glue function* or *coding function* is

$$\mathfrak{C}\colon \mathcal{E}\to\mathbb{M}, \qquad h \;\mapsto\; (T_h, d_h, \mu_h),$$

where (T_h,d_h) is defined above and $\mu_h = \lambda_{[0,1]}\circ\pi_h^{-1}$ is the pushforward of Lebesgue measure.

If $h\in\mathcal{E}$ is continuous, $\mathfrak{C}(h)$ is compact. Abraham, Delmas and Hoscheit have shown in [ADH14] that on the subspace of continuous excursions with uniform norm, the coding function is Lipschitz continuous w.r.t. the Gromov-Hausdorff-Prokhorov metric. The uniform topology, however, is a rather strong one, especially if we want to consider all of $\mathcal{E}$ (to code non-compact trees), where it is not separable. Therefore, I introduced in [1] a new, strictly weaker, metrizable topology on $\mathcal{E}$, which I call *excursion topology*. It is the weakest topology which is stronger than both, convergence in measure, and epigraph convergence. For details on the latter see [Bee94], it is Kuratowski convergence of the epigraphs.

Theorem 9 (Excursion topology and glue function; [1]). *$\mathcal{E}$ with excursion topology is a non-Polish, metrizable Lusin space, and the glue function $\mathfrak{C}\colon \mathcal{E} \to \mathbb{M}$ is continuous w.r.t. excursion topology on $\mathcal{E}$ and Gromov-weak topology on $\mathbb{M}$.*

In [5], we also extended the continuity result of [ADH14] to transient excursions on $\mathbb{R}_+$ coding measure trees with locally finite measure and Gromov-Hausdorff-vague topology.

4 Bi-measure trees, LWV-topology & the pruning process

In order to consider pruning processes in a unified way, as mentioned in Section 1, we need a sufficiently nice topology on a space of *bi-measure trees*, where one of the measures has to be general enough to include the length measure on the tree, which is not even locally finite in general (e.g. for the Brownian CRT). Of course, we need some finiteness hypothesis on the pruning measure, and we assume it to be finite on arcs. The topology should at least be separable, metrizable so that probabilistic standard theorems apply, and we want to be able to work with a tractable class of convergence determining test functions, similar to the set of polynomials.

Definition 10 (Bi-measure trees; [3]). (T, r, μ, ν) is a *(rooted) bi-measure $\mathbb{R}$-tree* if (T, r) is a complete, separable $\mathbb{R}$-tree, μ is a finite measure on T called *sampling measure*, and ν is a measure on T called *pruning measure* which satisfies $\nu([x, y]) < \infty$ for every arc $[x, y]$ in T, and $\nu(\mathrm{lf}(T) \setminus \mathrm{at}(\mu)) = 0$. Denote the set of (equivalence classes of) bi-measure $\mathbb{R}$-trees by $\mathbb{H}^{f,\sigma}$.

Of course, we need an appropriate notion of equivalence of bi-measure $\mathbb{R}$-trees, see Definition 2.12 in [3], and I suppress here the detail that we are actually considering rooted spaces. We can re-interpret the definition of Gromov-weak topology via polynomials as special case of the following general idea (dense graph limits, [LS06], also fall into this category). For every cardinality $n \in \mathbb{N}$, use the sampling measure to generate a random subspace by n independently chosen points, and require convergence in distribution of this random subspace. In the case of Gromov-weak topology, the finite subspace is just the set of (up to) n points with the metric inherited from the mm-space, which can be seen as $n \times n$ distance matrix. The idea behind the *leaf-sampling weak-vague (LWV) topology* is to go “one level higher” and use the sampling measure to sample a subtree which, equipped with the restricted pruning measure, is an mm-space. For these random, sampled subtrees we require Gromov-weak convergence in distribution (more precisely a multi-pointed version of it). This effectively leads to a two-step sampling procedure.

Definition 11 (LWV-topology; [3]). For $(T, r, \mu, \nu) \in \mathbb{H}^{f,\sigma}$, $n \in \mathbb{N}$, $\underline{u} \in T^n$, let $[\![\underline{u}]\!]$ be the connected subtree spanned by $\underline{u}$. Let $\tau^n_{\mathcal{X}}(\underline{u}) := ([\![\underline{u}]\!], r\restriction_{[\![\underline{u}]\!]}, \underline{u}, \nu\restriction_{[\![\underline{u}]\!]})$, which is a multi pointed mm-space with finite measure. A sequence $(\mathcal{X}_k)_{k\in\mathbb{N}}$ in $\mathbb{H}^{f,\sigma}$ converges w.r.t. *LWV-topology* to $\mathcal{X} \in \mathbb{H}^{f,\sigma}$ if $\mu_k^{\otimes n} \circ (\tau^n_{\mathcal{X}_k})^{-1}$ converges weakly w.r.t. multi pointed Gromov-weak topology to $\mu^{\otimes n} \circ (\tau^n_{\mathcal{X}})^{-1}$.

Similarly to the polynomials, consider the set $\mathcal{F}$ of functions $\Psi\colon \mathbb{H}^{f,\sigma} \to \mathbb{R}$ of the form

$$\Psi(\mathcal{X}) := \gamma(\|\mu\|) \cdot \int_{T^n} \Phi\big(\tau^n_{\mathcal{X}}(\underline{u})\big)\, \mu^{\otimes n}(\mathrm{d}\underline{u}),$$

where $\gamma \in \mathcal{C}_b(\mathbb{R}_+)$ satisfies $\lim_{x\to\infty} x^k\gamma(x) = 0$ for all $k \in \mathbb{N}$, and Φ is a polynomial (I am cheating a bit here, Φ should be from a slightly more general set, see (2.32) of [3]). For an $\mathbb{R}$-tree (T, r), let λ_T be the *length measure* on T, i.e. the unique measure with $\lambda_T([x, y]) = r(x, y)$ and $\lambda_T(\mathrm{lf}(T)) = 0$.

Theorem 12 (The LWV-topology; [3])**.** *$\mathbb{H}^{f,\sigma}$ with LWV-topology is a separable, metrizable space, and $\mathcal{F}$ is convergence determining for measures on $\mathbb{H}^{f,\sigma}$. Furthermore, for every $\mathcal{X} = (T, r, \mu, \nu) \in \mathbb{H}^{f,\sigma}$, the set $\{(T, r, \mu', \nu') \in \mathbb{H}^{f,\sigma} \mid \mu' \le \mu,\ \nu' \le \nu\}$ is compact, and the map $\mathbb{H} \to \mathbb{H}^{f,\sigma}$, $(T, r, \mu) \mapsto (T, r, \mu, \lambda_T)$ is a homeomorphism onto its image. Here, $\mathbb{H}$ is the set of metric measure trees with Gromov-weak topology.*

In this new state space $\mathbb{H}^{f,\sigma}$, we can define one pruning process which, depending on the initial bi-measure $\mathbb{R}$-tree, contains all the discrete and continuous pruning process considered in [AP98, AS02, ADV10, AD12] as special cases.

Theorem 13 (The pruning process; [3])**.** *The $\mathbb{H}^{f,\sigma}$-valued pruning process is a strong Markov process with càdlàg paths, and the law of the path depends continuously (w.r.t. Skorokhod topology) on the initial $\mathcal{X} \in \mathbb{H}^{f,\sigma}$.*

Furthermore, we obtain an expression for the generator of the process acting on test functions from $\mathcal{F}$. As a corollary to the continuity of the pruning process together with the homeomorphic embedding from Theorem 12, we obtain for every Gromov-weakly convergent sequence of metric measure trees that the corresponding pruning processes with the length measure converge Gromov-weakly in path space. Note that He and Winkel showed in [HW14] the convergence of node pruning processes of discrete trees to pruning processes of Lévy trees which we did not obtain because we were not able to show LWV-convergence of the initial trees.

5 Tree structure without a metric

As already noted in Section 1.3, thinking of trees as metric spaces is not always the most natural way, and a priori it is not clear why the non-discrete generalization of a combinatorial object should be a metric space. And working with them as metric measure spaces is not always the easiest thing to do because of the lack of compactness inherent in metric measure spaces. Curien suggest in [Cur14], without working out the details, that one might look at homeomorphism equivalence classes, and base a notion of convergence on the coding of trees by triangulations of the circle introduced by Aldous in [Ald94a, Ald94b]. It turns out, however, that the topological information is not really coded naturally, and we suggest a different idea of "tree structure". Characterizing the topological structures induced by $\mathbb{R}$-trees has received considerable attention ([MO90, MNO92, Fab15]). But, unlike for metric spaces, we do not know any useful notion of convergence of topological spaces or of topological measure spaces. Therefore, in [7], we change the focus not only away from the metric, but also from the topological structures. Instead, we formalize the "tree structure" with a branch point map. We take an axiomatic approach, which in particular allows us to characterize which branch point maps may arise from separable $\mathbb{R}$-trees.

5.1 Algebraic measure trees

Definition 14 (Algebraic trees; [7])**.** An *algebraic tree* is a set T with a symmetric *branch point map* $c\colon T^3 \to T$ satisfying $c(x, y, y) = y$, $c(x, y, c(x, y, z)) = c(x, y, z)$ and the four point condition

$$c(x_1, x_2, x_3) \in \big\{c(x_1, x_2, x_4),\ c(x_1, x_3, x_4),\ c(x_2, x_3, x_4)\big\}.$$

As example, one should have in mind an $\mathbb{R}$-tree (T, r) with branch point map given by $[x, y] \cap [y, z] \cap [z, x] = \{c(x, y, z)\}$, where $[x, y]$ is the arc in (T, r). We call (T, r) *metric representation* of (T, c) if this relation holds. Given an algebraic tree (T, c), we can go the other way round and define analogues of arcs by $[x, y] := \{z \in T \mid c(x, y, z) = z\}$. From there, we can naturally define other concepts, such as for $x, y \in T$ the component of $T \setminus \{x\}$ containing y, namely $\mathcal{S}_x(y) := \{z \in T \setminus \{x\} \mid x \notin [y, z]\}$, the degree of a vertex, a notion of separability, ...

Theorem 15 ([7])**.** *Every separable algebraic tree has a (non-unique) metric representation.*

The set of components also defines a natural topology on T. On the corresponding Borel σ-algebra, we can define probability measures, and using an appropriate notion of equivalence, we obtain a set $\mathbb{T}$ of (equivalence classes of) separable *algebraic measure trees*, and the subset $\mathbb{T}_2 \subseteq \mathbb{T}$ of *binary algebraic measure trees* with no atoms except (possibly) at leaves. Using the branch point distribution $\nu = \mu^{\otimes 3} \circ c^{-1}$ of an algebraic measure tree (T, c, μ), we can define the metric $r_\nu(x, y) := \nu([x, y]) - \frac{1}{2}(\nu\{x\} + \nu\{y\})$. The *n-sample shape distribution* is (essentially) the law of the combinatorial tree obtained by sampling n points with μ and adding all branch points.

Theorem 16 (The space $\mathbb{T}$; [7])**.** *The selection map $\iota\colon \mathbb{T} \to \mathbb{M}$, $\iota(T, c, \mu) = (T, r_\nu, \mu)$ is injective, $\iota(\mathcal{X})$ is a metric representation of $\mathcal{X}$, and the topology induced by ι on $\mathbb{T}$ is separable and metrizable. Furthermore, the subspace $\mathbb{T}_2$ is compact, and a sequence in $\mathbb{T}_2$ converges if and only if for every n, the n-sample shape distribution converges weakly.*

Furthermore, we obtained a continuous, surjective coding function from the space of sub-triangulations of the circle (with Hausdorff metric topology) onto $\mathbb{T}_2$.

5.2 The Aldous diffusion

The Aldous chain on n-cladograms considered in [Ald00] is a simple Markov chain on the set of n-cladograms (semi-labelled, binary trees with n leaves) with the uniform distribution as reversible measure. In every time step, a randomly chosen leaf is removed, and re-attached to a randomly chosen edge, so that we obtain again a binary tree with n leaves. It has been an open problem of Aldous's since 1999 to construct a limit process on "some space of trees". We solved this problem in [8] in the space $\mathbb{T}_2$ of binary algebraic measure trees. A different approach has been pursued independently by Forman, Pal, Rizzolo and Winkel in the sequence of papers [FPRW16, FPRW18c, FPRW18b, FPRW18a]. In the last one, they obtained an analogue of the Aldous chain in a metric measure space setup. They use quite different techniques, include the information about the metric which we neglect, and do not obtain convergence of the discrete chains.

Theorem 17 (Aldous diffusion; [8])**.** *The Aldous chain on n-cladograms, seen as $\mathbb{T}_2$-valued process and run at speed n^2, converges weakly in path space to a Feller process with continuous paths, which we call* Aldous diffusion*. The Aldous diffusion is ergodic with the algebraic measure Brownian CRT as unique stationary distribution and solves a well-posed martingale problem.*

6 An invariance principle on trees

As mentioned in Section 1, state spaces of continuum trees are not only needed for limits of Markov chains with values in the set of trees, but also for limits of chains with values in a given tree. The easiest such Markov chain is of course the simple, nearest neighbour random walk. There is an extensive theory about random walks on Galton-Watson (GW) trees, see Chapter 17 of [LP16]. They are also used to study random walks on percolation clusters (e.g. [vdHHN19]). In the view of Aldous's result that conditioned, finite variance GW trees converge to the Brownian CRT, it is natural to ask if the random walks converge to some kind of Brownian motion on the CRT as well. This particular case, and a generalization including infinite variance GW trees converging to Lévy trees, was answered affirmatively by Croydon in [Cro08, Cro10], using a construction of Brownian motion on the Brownian CRT given by Krebs in [Kre95].

One technicality here is that, unlike in the classical Donsker theorem, the rescaled discrete trees are not isometrically embeddable into the limiting $\mathbb{R}$-tree. Another one is about the very definition of Brownian motion. The generalization of Lebesgue measure on an $\mathbb{R}$-tree is the length measure, which is the one-dimensional Hausdorff measure restricted to the skeleton of the tree. This, however, is not locally finite for the Brownian CRT, and thus the Brownian motion with "natural speed" would get stuck immediately. Therefore, a locally finite *speed measure* is required. In [Kre95], implicitly, the mass measure of the Brownian CRT was used. A more general ν-Brownian motion on $\mathbb{R}$-trees with given speed measure ν was constructed by Athreya, Eckhoff and Winter in [AEW13]. A speed measure ν on a discrete tree can also be used to govern the speed of a random walk, so that we obtain ν-random walks.

6.1 Speed-ν motions

What do ν-random walks and ν-Brownian motions have in common? They are on natural scale, do not jump over points of $\mathrm{supp}(\nu)$, and the local speed is (in some sense) inversely proportional to ν. Stone analyzed in [Sto63] the corresponding class of processes, which we call *speed-ν motions*, on $\mathbb{R}$. He defined them by a time transformation of the same standard Brownian motion for every Radon measure ν on $\mathbb{R}$, and showed that they converge almost surely in path space whenever the speed measures converge in what we would call Hausdorff-vague topology. Note that the class of speed-ν motions also contains sticky Brownian motions (ν has an atom at the sticky point), Brownian motion on the Cantor set (ν is the uniform distribution on the Cantor set), and the snapping out Brownian motion defined by Lejay in [Lej16] (ν is Lebesgue measure restricted to $(-\infty, a] \cup \mathbb{R}_+$, $a < 0$, and we have to identify a with 0).

In [6], we define the speed-ν motion on a complete, locally compact metric tree (T, r), using a different approach. We define it as the unique strong Markov process satisfying the *occupation time formula*

$$\mathbb{E}^x\Big[\int_0^{\tau_y} f(X_s)\,\mathrm{d}s\Big] = 2\int_T r\big(y, c(x,y,z)\big) f(z)\,\nu(\mathrm{d}z),$$

for all $x, y \in T$ and sufficiently nice f, where τ_y is the hitting time of y and c is the branch point map. We construct it for all Radon measures ν with $\mathrm{supp}(\nu) = T$ using a Dirichlet form. If (T, r) is discrete, it is a random walk, if (T, r) is an $\mathbb{R}$-tree it is the ν-Brownian motion, and if $T = \mathbb{R}$ it is the process considered in [Sto63].

6.2 Convergence of speed-ν motions

Recall that both Gromov-weak and Gromov-Hausdorff-weak topology can be defined using embeddings into a common metric space, and the same is true for the vague versions ([5]). To define convergence of processes living on different spaces, we take a similar approach and say that it *converges weakly in path space* (or f.d.d.) if the underlying metric spaces can be embedded isometrically into a common metric space, such that the embedded processes converge weakly in Skorokhod path space (or f.d.d.).

Theorem 18 (Invariance principle on trees; [6])**.** *Let $\mathcal{X}_n = (T_n, r_n, \nu_n)$ be Heine-Borel spaces with a Radon measure of full support, $n \in \mathbb{N} \cup \{\infty\}$, $\mathcal{X} := \mathcal{X}_\infty$.*

1. *If $\mathcal{X}_n \to \mathcal{X}$ Gromov-Hausdorff-vaguely, then the speed-ν_n motions converge weakly in path space to the speed-ν motion (provided there is no local edge-length explosion) until the explosion time of the speed-ν motion.*

2. *If $\mathcal{X}_n \to \mathcal{X}$ Gromov-vaguely and the diameters are uniformly bounded, then the speed-ν_n motions converge f.d.d. to the speed-ν motion.*

This invariance principle contains the one of [Sto63] for speed-ν motions on $\mathbb{R}$ (Stone allows killing, which we do not consider), and the convergences of random walks to Brownian motions on Lévy trees obtained in [Cro08, Cro10]. The basic ideas of **[6]** turned out to be flexible enough to be generalized further from trees to so-called *resistance metric* spaces by Croydon in [Cro18]. Besides trees, some fractals are examples of resistance metric spaces (see [Kig01]).

Speed-ν motions are on natural scale. Nevertheless, the invariance principle can also be used to derive convergence of some processes with drift: we can use the scale function to distort the metric of the tree in such a way that the process becomes on natural scale. Then we can use the invariance principle if the distorted spaces converge, and finally undo the distortion of the metric. Results about random walks in random potentials were recently obtained using this approach by Andriopoulos in [And18].

Bibliography

[1] – [8] Articles constituting the main part of this habilitation. See Contents or Part II.

[AD12] Romain Abraham and Jean-François Delmas, *A continuum-tree-valued Markov process*, Ann. Probab. **40** (2012), no. 3, 1167–1211.

[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** (2013), no. 14, 21 pp.

[ADH14] ______, *Exit times for an increasing Lévy tree-valued process*, Probab. Theory Related Fields **159** (2014), no. 1-2, 357–403. MR 3201925

[ADV10] Romain Abraham, Jean-François Delmas, and Guillaume Voisin, *Pruning a Lévy continuum random tree*, Electron. J. Probab. **15** (2010), no. 46, 1429–1473.

[AEW13] Siva Athreya, Michael Eckhoff, and Anita Winter, *Brownian motion on $\mathbb{R}$-trees*, Trans. Amer. Math. Soc. **365** (2013), 3115–3150.

[Ald91a] David Aldous, *The continuum random tree I*, Ann. Probab. **19** (1991), no. 1, 1–28.

[Ald91b] ______, *The continuum random tree II: An overview*, Stochastic Analysis, London Math. Soc. Lecture Note Ser., vol. 167, Cambridge Univ. Press, Cambridge, 1991, pp. 23–70.

[Ald93] ______, *The continuum random tree III*, Ann. Probab. **21** (1993), no. 1, 248–289.

[Ald94a] ______, *Recursive self-similarity for random trees, random triangulations and Brownian excursion*, Ann. Probab. **22** (1994), no. 2, 527–545. MR 1288122

[Ald94b] ______, *Triangulating the circle, at random*, Amer. Math. Monthly **101** (1994), no. 3, 223–233. MR 1264002

[Ald00] ______, *Mixing time for a Markov chain on cladograms*, Combin. Probab. Comput. **9** (2000), no. 3, 191–204. MR MR1774749 (2001f:05129)

[And18] George Andriopoulos, *Invariance principles for random walks in random environment on trees*, 2018, arXiv:1812.10197v1.

[AP98] David Aldous and Jim Pitman, *The standard additive coalescent*, Ann. Probab. **26** (1998), no. 4, 1703–1726. MR 1675063

[AS02] Romain Abraham and Laurent Serlet, *Poisson snake and fragmentation*, Electron. J. Probab. **7** (2002), 1–15.

[Bee94] Gerald Beer, *A note on epi-convergence*, Canad. Math. Bull. **37** (1994), no. 3, 294–239.

[BFF18] J. Berestycki, M. C. Fittipaldi, and J. Fontbona, *Ray-Knight representation of flows of branching processes with competition by pruning of Lévy trees*, Probab. Theory Related Fields **172** (2018), no. 3-4, 725–788. MR 3877546

[Bog07] V. I. Bogachev, *Measure theory, volume II*, Springer, 2007.

[BP19] Bojan Basrak and Hrvoje Planinić, *A note on vague convergence of measures*, Statist. Probab. Lett. **153** (2019), 180–186. MR 3979308

[BS01] Itai Benjamini and Oded Schramm, *Recurrence of distributional limits of finite planar graphs*, Electron. J. Probab. **6** (2001), no. 23, 13. MR 1873300

[BvdHS18] Shankar Bhamidi, Remco van der Hofstad, and Sanchayan Sen, *The multiplicative coalescent, inhomogeneous continuum random trees, and new universality classes for critical random graphs*, Probab. Theory Related Fields **170** (2018), no. 1-2, 387–474. MR 3748328

[Chi01] Ian Chiswell, *Introduction to Λ-trees*, World Scientific Publishing Co. Inc., River Edge, NJ, 2001.

[Cro08] David Croydon, *Convergence of simple random walks on random discrete trees to Brownian motion on the continuum random tree*, Ann. Inst. H. Poincaré Probab. Statist. **44** (2008), no. 6, 987–1019.

[Cro10] ———, *Scaling limits for simple random walks on random ordered graph trees*, Adv. Appl. Probab. **42** (2010), no. 2, 528–558.

[Cro18] ———, *Scaling limits of stochastic processes associated with resistance forms*, Ann. Inst. H. Poincaré Probab. Statist. **54** (2018), no. 4, 1939–1968. MR 3865663

[Cur14] Nicolas Curien, *Dissecting the circle, at random*, Journées MAS 2012, ESAIM Proc., vol. 44, EDP Sci., Les Ulis, 2014, pp. 129–139. MR 3178612

[DGP11] Andrej Depperschmidt, Andreas Greven, and Peter Pfaffelhuber, *Marked metric measure spaces*, Electron. Commun. Probab. **16** (2011), 174–188.

[DGP12] Andrej Depperschmidt, Andreas Greven, and Peter Pfaffelhuber, *Tree-valued Fleming-Viot dynamics with selection*, Ann. Appl. Probab. **22** (2012), no. 6, 2560–2615.

[DGP13] Andrej Depperschmidt, Andreas Greven, and Peter Pfaffelhuber, *Path-properties of the tree-valued Fleming-Viot processes*, Electron. J. Probab. **18** (2013), no. 84, 47 pp.

[DLG02] Thomas Duquesne and Jean-François Le Gall, *Random trees, Lévy processes and spatial branching processes*, Astérisque **281** (2002), vi+147.

[DMT96] Andreas Dress, Vincent Moulton, and Werner Terhalle, *T-theory: an overview*, European J. Combin. **17** (1996), no. 2-3, 161–175, Discrete metric spaces (Bielefeld, 1994). MR 97e:05069

[Duq03] Thomas Duquesne, *A limit theorem for the contour process of conditioned Galton-Watson trees*, Annals of Prob. **31** (2003), no. 2, 996–1027.

[EGW17] Steven N. Evans, Rudolf Grübel, and Anton Wakolbinger, *Doob-Martin boundary of Rémy's tree growth chain*, Ann. Probab. **45** (2017), no. 1, 225–277. MR 3601650

[EPW06] Steven N. Evans, Jim Pitman, and Anita Winter, *Rayleigh processes, real trees, and root growth with re-grafting*, Probab. Theory Related Fields **134** (2006), no. 1, 81–126.

[Eva08] Steven N. Evans, *Probability and real trees*, Lecture Notes in Mathematics, vol. 1920, Springer, Berlin, 2008, Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005. MR 2351587

[EW06] Steven N. Evans and Anita Winter, *Subtree prune and re-graft: A reversible real-tree valued Markov chain*, Ann. Probab. **34** (2006), no. 3, 918–961.

[Fab15] Paul Fabel, *A topological characterization of the underlying spaces of complete R-trees*, Michigan Math. J. **64** (2015), no. 4, 881–887. MR 3426619

[For18] Noah Forman, *Mass-structure of weighted real trees*, arXiv:1801.02700v1, 2018.

[FPRW16] Noah Forman, Soumik Pal, Douglas Rizzolo, and Matthias Winkel, *Diffusions on a space of interval partitions with Poisson-Dirichlet stationary distributions*, arXiv:1609.06707, 2016.

[FPRW18a] ———, *Aldous diffusion I: a projective system of continuum k-tree evolutions*, arXiv:1809.07756, 2018.

[FPRW18b] ———, *Interval partition evolutions with emigration related to the Aldous diffusion*, arXiv:1804.01205, 2018.

[FPRW18c] ———, *Projections of the Aldous chain on binary trees: Intertwining and consistency*, arXiv:1802.00862, 2018.

[Fuk87] Kenji Fukaya, *Collapsing of Riemannian manifolds and eigenvalues of Laplace operator*, Invent. Math. **87** (1987), no. 3, 517–547.

[GGR19] Patric Glöde, Andreas Greven, and Thomas Rippl, *Branching trees I: Concatenation and infinite divisibility*, Electron. J. Probab. **24** (2019), no. 52, 55 pp. MR 3968714

[Glö13] Patric Glöde, *Dynamics of genealogical trees for autocatalytic branching processes*, Ph.D. thesis, FAU Erlangen-Nürnberg, 2013, http://d-nb.info/1033029912/34.

[GP81] M. Gromov and P. Pansu, *Structures métriques pour les variétés Riemanniennes*, Textes Mathématiques, Cedic/Fernand Nathan, 1981.

[GPW09] Andreas Greven, Peter Pfaffelhuber, and Anita Winter, *Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees)*, Probab. Theory Related Fields **145** (2009), 285–322. MR 2520129

[GPW13] ———, *Tree-valued resampling dynamics: Martingale problems and applications*, Probab. Theory Related Fields **155** (2013), no. 3–4, 789–838.

[Gro81] Misha Gromov, *Groups of polynomial growth and expanding maps*, Inst. Hautes Études Sci. Publ. Math. **53** (1981), 53–73.

[Gro99] Misha Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, vol. 152, Birkhäuser Boston Inc., Boston, MA, 1999.

[GSW16] Andreas Greven, Rongfeng Sun, and Anita Winter, *Continuum space limit of the genealogies of interacting Fleming-Viot processes on $\mathbb{Z}$*, Electron. J. Probab. **21** (2016), no. 58, 64 pp. MR 3563886

[Guf18] Stephan Gufler, *Pathwise construction of tree-valued Fleming-Viot processes*, Electron. J. Probab. **23** (2018), no. 42, 58 pp. MR 3806410

[HW14] Hui He and Matthias Winkel, *Invariance principles for pruning processes of Galton-Watson trees*, 2014, arxiv:1409.1014v1.

[HW19] ———, *Gromov-Hausdorff-Prokhorov convergence of vertex cut-trees of n-leaf Galton-Watson trees*, Bernoulli **25** (2019), no. 3, 2301–2329.

[Kig01] Jun Kigami, *Analysis on fractals*, Cambridge Tracts in Mathematics, vol. 143, Cambridge University Press, Cambridge, 2001. MR 1840042

[Kre95] W. B. Krebs, *Brownian motion on the continuum tree*, Probab. Theory Related Fields **101** (1995), 421–433.

[KS05] Takashi Kumagai and Karl-Theodor Sturm, *Construction of diffusion processes on fractals, d-sets, and general metric measure spaces*, J. Math. Kyoto Univ. **45** (2005), no. 2, 307–327. MR 2161694

[KW19] Sandra Kliem and Anita Winter, *Evolving phylogenies of trait-dependent branching with mutation and competition, Part I: Existence*, Stochastic Process. Appl. **129** (2019), no. 12, 4837–4877. MR 4025694

[Lej16] Antoine Lejay, *The snapping out Brownian motion*, Ann. Appl. Probab. **26** (2016), no. 3, 1727–1742. MR 3513604

[LG91] Jean-François Le Gall, *Brownian excursions, trees and measure-valued branching processes*, Ann. Probab. **19** (1991), no. 4, 1399–1439. MR 1127710

[LG93] Jean-François Le Gall, *The uniform random tree in a Brownian excursion*, Probab. Theory Related Fields **96** (1993), no. 3, 369–383.

[Lov12] László Lovász, *Large networks and graph limits*, American Mathematical Society Colloquium Publications, vol. 60, AMS, Providence, RI, 2012. MR 3012035

[LP16] Russell Lyons and Yuval Peres, *Probability on trees and networks*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 42, Cambridge University Press, New York, 2016. MR 3616205

[LS06] László Lovász and Balázs Szegedy, *Limits of dense graph sequences*, J. Combin. Theory Ser. B **96** (2006), no. 6, 933–957. MR 2274085

[LV09] John Lott and Cédric Villani, *Ricci curvature for metric-measure spaces via optimal transport*, Ann. Math. **169** (2009), no. 3, 903–991.

[Mie09] Grégory Miermont, *Tessellations of random maps of arbitrary genus*, Ann. Sci. Éc. Norm. Supér. (4) **42** (2009), no. 5, 725–781. MR 2571957

[MNO92] John C. Mayer, Jacek Nikiel, and Lex G. Oversteegen, *Universal spaces for* $\mathbf{R}$*-trees*, Trans. Amer. Math. Soc. **334** (1992), no. 1, 411–432. MR 1081940

[MO90] John C. Mayer and Lex G. Oversteegen, *A topological characterisation of* $\mathbf{R}$*-trees*, Trans. Amer. Math. Soc. **320** (1990), no. 1, 395–415.

[Pio11] Sven Piotrowiak, *Dynamics of genealogical trees for type- and state-dependent resampling models*, Ph.D. thesis, FAU Erlangen-Nürnberg, 2011, http://d-nb.info/1009840509/34.

[ROO+19] Y. Roskov, G. Ower, T. Orrell, D. Nicolson, N. Bailly, P.M. Kirk, T. Bourgoin, R.E. DeWalt, W. Decock, E. van Nieukerken, J. Zarucchi, and L. Penev, *ITIS catalogue of life, 2019 annual checklist*, www.catalogueoflife.org/annual-checklist/2019, 2019.

[Shi16] Takashi Shioya, *Metric measure geometry*, IRMA Lectures in Mathematics and Theoretical Physics, vol. 25, EMS Publishing House, Zürich, 2016. MR 3445278

[Sto63] Charles Stone, *Limit theorems for random walks, birth and death processes, and diffusion processes*, Illinois J. Math. **7** (1963), no. 4, 638–660.

[Stu98] Karl-Theodor Sturm, *Diffusion processes and heat kernels on metric spaces*, Ann. Probab. **26** (1998), no. 1, 1–55.

[Stu06] ———, *On the geometry of metric measure spaces I*, Acta Math. **196** (2006), no. 1, 65–131.

[Tit77] J. Tits, *A "theorem of Lie-Kolchin" for trees*, Contributions to algebra, Academic Press, New York, 1977, pp. 377–388. MR 0578488

[vdHHN19] Remco van der Hofstad, Tim Hulshof, and Jan Nagel, *Random walk on barely supercritical branching random walk*, Probability Theory and Related Fields **online first** (2019).

[Vil09] Cédric Villani, *Optimal transport*, Grundlehren der mathematischen Wissenschaften, vol. 338, Springer, Berlin-Heidelberg, 2009.

Part II

Articles

EQUIVALENCE OF GROMOV-PROHOROV- AND GROMOV'S $\underline{\Box}_\lambda$-METRIC ON THE SPACE OF METRIC MEASURE SPACES

WOLFGANG LÖHR

ABSTRACT. The space of metric measure spaces (complete separable metric spaces with a probability measure) is becoming more and more important as state space for stochastic processes. Of particular interest is the subspace of (continuum) metric measure trees. Greven, Pfaffelhuber and Winter introduced the Gromov-Prohorov metric d_{GP} on the space of metric measure spaces and showed that it induces the Gromov-weak topology. They also conjectured that this topology coincides with the topology induced by Gromov's $\underline{\Box}_1$ metric. Here, we show that this is indeed true, and the metrics are even bi-Lipschitz equivalent. More precisely, $d_{\mathrm{GP}} = \frac{1}{2}\underline{\Box}_{\frac{1}{2}}$, and hence $d_{\mathrm{GP}} \leq \underline{\Box}_1 \leq 2d_{\mathrm{GP}}$. The fact that different approaches lead to equivalent metrics underlines their importance and also that of the induced Gromov-weak topology.

As an application, we give a shorter proof of the known fact that the map associating to a lower semi-continuous excursion the coded $\mathbb{R}$-tree is Lipschitz continuous when the excursions are endowed with the (non-separable) uniform metric. We also introduce a new, weaker, metric topology on excursions, which has the advantage of being separable and making the space of bounded excursions a Lusin space. We obtain continuity also for this new topology.

1. Introduction

Tree-valued stochastic processes frequently appear in probability theory and its application areas, such as theoretical biology. For instance, in an evolutionary model, the development of the genealogical tree is of interest. In the continuum limit of infinite population size, the finite tree becomes a continuum tree ($\mathbb{R}$-tree) and the normalised counting measure of individuals becomes a probability measure on it. This measure is needed to describe the population density on the tree and to sample individuals from it. See Aldous' seminal paper [Ald93] for the convergence of finite variance Galton-Watson trees to a (Brownian) continuum measure tree, and results of Duquesne and Le Gall ([DLG02, Duq03]) for the convergence of infinite variance Galton-Watson trees to Lévy trees.

More generally than $\mathbb{R}$-trees, we can consider random metric (probability) measure spaces, an approach introduced by Greven, Pfaffelhuber and Winter in [GPW09] and applied by the authors and Depperschmidt to obtain tree-valued Fleming-Viot dynamics in [GPW13, DGP12]. Here, $\mathcal{X} = (X, d, \mu)$ is a **metric measure space** (**mm-space**) if (X, d) is a complete, separable metric space and μ a probability measure on the Borel σ-algebra of X. To work with mm-space valued processes, it is crucial to have an appropriate topology on the set of mm-spaces, or rather the set $\mathfrak{X}$ of isometry classes of mm-spaces. A fruitful topology is given by the Gromov-weak topology introduced in [GPW09]. In the same paper, the authors conjectured that it coincides with the topology induced by Gromov's metric $\underline{\Box}_1$, which is defined in [Gro99, Chapter $3\frac{1}{2}$]. They also introduced a complete metric, the Gromov-Prohorov metric d_{GP}, that metrises the Gromov-weak topology.

Here, we show that $\underline{\Box}_1$ and d_{GP} are bi-Lipschitz equivalent, which in particular implies that the conjecture is true and $\underline{\Box}_1$ indeed metrises Gromov-weak topology. Furthermore, we use this result to prove that the measure $\mathbb{R}$-tree coded by an excursion depends continuously on the excursion. To this end, we consider two topologies on the space of lower semi-continuous excursions. For the uniform topology, Lipschitz continuity is already shown by Abraham, Delmas and Hoscheit in [ADH14, Prop. 2.9] (with their metric on trees, which implies the result for ours), but we obtain a much shorter proof using the equivalence of d_{GP} and $\underline{\square}_1$. The uniform topology has the disadvantage of being non-separable, therefore we introduce a new, weaker, separable, metrisable topology, which is Lusin on the subset of bounded excursions. We also show continuous dependence of the tree on the excursion in this weaker topology.

In the next section, we recall the definition of the metrics d_{GP} and $\underline{\square}_1$, as well as of Gromov-weak topology, and emphasize that the algebra of polynomials used to define Gromov-weak topology is convergence determining albeit not dense in the bounded continuous functions. We also give a short comparison to related, but slightly different topologies used on spaces of mm-spaces. The third section contains the proof of the equivalence of d_{GP} and $\underline{\square}_1$. In the last section, we apply the equivalence to measure trees coded by excursions and define the new topology on the space of excursions.

2. Metrics and topologies on the space of mm-spaces

We do not distinguish between isomorphic mm-spaces. Here, two mm-spaces $\mathcal{X} = (X, d, \mu)$ and $\mathcal{X}' = (X', d', \mu')$ are called **isomorphic** if there is a measure preserving map $f\colon X \to X'$ such that the restriction to the support of μ is an isometry, i.e.

$$\mu' = \mu \circ f^{-1} \qquad \text{and} \qquad d(x,y) = d'\big(f(x), f(y)\big) \;\; \forall x, y \in \operatorname{supp}(\mu).$$

We denote the space of (isometry classes of) mm-spaces by $\mathfrak{X}$.

Remark 2.1. Because (X, d) is complete, an isomorphism f from $\mathcal{X}$ to $\mathcal{X}'$ is an isometric bijection between $\operatorname{supp}(\mu)$ and $\operatorname{supp}(\mu')$. In particular, there is also an inverse isomorphism g from $\mathcal{X}'$ to $\mathcal{X}$ with $g \circ f = \mathrm{id}$ on $\operatorname{supp}(\mu)$.

Gromov-Prohorov metric. The Gromov-Prohorov metric is obtained by embedding the metric spaces underlying the mm-spaces optimally into a common metric space and taking the Prohorov distance between the push forward measures.

Definition 2.2 (Prohorov metric)**.** Let μ, ν be probability measures on a metric space (X, d). Then the **Prohorov distance** is

$$d_{\mathrm{Pr}}(\mu,\nu) := \inf\big\{\, \varepsilon > 0 \;\big|\; \mu(A) \le \nu(A^\varepsilon) + \varepsilon \;\; \forall A \in \mathfrak{B}(X) \,\big\},$$

where $A^\varepsilon := \big\{\, x \in X \;\big|\; d(A, x) < \varepsilon \,\big\}$.

Remark 2.3. Below, we use the following equivalent expression for the Prohorov metric. A **coupling** between μ and ν is a measure ξ on $X^2 = X \times X$ with marginals μ and ν on X. Then

$$d_{\mathrm{Pr}}(\mu,\nu) = \inf\Big\{\, \varepsilon > 0 \;\Big|\; \exists \text{ coupling } \xi \text{ of } \mu,\, \nu : \xi\big(\{\, (x,y) \in X^2 \mid d(x,y) \ge \varepsilon \,\}\big) \le \varepsilon \,\Big\}.$$

Definition 2.4 (Gromov-Prohorov metric)**.** Let $\mathcal{X}_i = (X_i, d_i, \mu_i) \in \mathfrak{X}$, $i = 1, 2$, be mm-spaces. The **Gromov-Prohorov metric** is defined by

$$d_{\mathrm{GP}}(\mathcal{X}_1, \mathcal{X}_2) := \inf_{f,g} d_{\mathrm{Pr}}\big(\mu_1 \circ f^{-1}, \mu_2 \circ g^{-1}\big),$$

where the infimum is taken over all isometries $f\colon X_1 \to X$ and $g\colon X_2 \to X$ into a common separable metric space (X, d).

Gromov-weak topology. The idea of Gromov-weak topology is to use convergence in distribution of finite metric subspaces, which are sampled from X with the measure μ. A very nice property of the Gromov-Prohorov metric is that it induces precisely the Gromov-weak topology, as shown in [GPW09]. This alternative characterisation of convergence provides us with a sub-algebra of $\mathcal{C}_b(\mathfrak{X})$, called algebra of polynomials. The usefulness of this algebra stems from the fact that it is rich enough to determine convergence of measures on $\mathfrak{X}$. To emphasize that polynomials are an essential tool for working with convergence in distribution of $\mathfrak{X}$-valued random variables, we remark that one cannot use the space $\mathcal{C}_c(\mathfrak{X})$ of continuous functions with compact support, because no point in $\mathfrak{X}$ has a compact neighbourhood, and hence $\mathcal{C}_c(\mathfrak{X}) = \{0\}$ is trivial.

Definition 2.5. A **polynomial** (on $\mathfrak{X}$) is a function $\Phi\colon \mathfrak{X} \to \mathbb{R}$ of the form

$$\Phi(\mathcal{X}) \;=\; \Phi^\phi(\mathcal{X}) \;:=\; \int_{X^n} \phi\big(\big(d(x_i,x_j)\big)_{i,j\le n}\big)\; \mu^{\otimes n}(\mathrm{d}x),$$

where $n \in \mathbb{N}$ and $\phi \in \mathcal{C}_b\big(\mathbb{R}^{n\times n}\big)$. Let Π be the set of such functions. **Gromov-weak topology** is the topology induced by Π on $\mathfrak{X}$.

Remark 2.6 (Polynomials are not dense)**.** Π is obviously an algebra, but it is not dense in $\mathcal{C}_b(\mathfrak{X})$. To see this, assume it is dense and consider the subspace $\mathfrak{X}_r$ of mm-spaces with essential diameter bounded by a fixed $r > 0$. Because $\mathfrak{X}_r$ is closed, the set $\Pi_r := \{\, \Phi\restriction_{\mathfrak{X}_r} \mid \Phi \in \Pi \,\}$ of restrictions of polynomials to $\mathfrak{X}_r$ is dense in $\mathcal{C}_b(\mathfrak{X}_r)$. Because Π_r is clearly separable, this means that $\mathcal{C}_b(\mathfrak{X}_r)$ is separable, and hence $\mathfrak{X}_r$ is compact. This is a contradiction (e.g. the set of finite spaces with discrete metric and uniform distribution has no limit point).

We say that a set $\mathcal{F} \subseteq \mathcal{C}_b(\mathfrak{X})$ is **convergence determining** if for probability measures ξ_n, ξ on $\mathfrak{X}$, the weak convergence $\xi_n \xrightarrow{w} \xi$ is equivalent to

$$\int f \,\mathrm{d}\xi_n \underset{n\to\infty}{\longrightarrow} \int f \,\mathrm{d}\xi \qquad \forall f \in \mathcal{F}.$$

Since $\mathcal{C}_b(\mathfrak{X})$ is difficult to describe, it is important to have such a set with a more tractable description. That Π is indeed convergence determining is shown with some effort by Depperschmidt, Greven and Pfaffelhuber in [DGP11]. We can also deduce it from an apparently not so well-known general theorem due to Le Cam.

Theorem 2.7 (Le Cam, [LC57]; see also [HJ77, Lem. 4.1])**.** *Let X be a completely regular Hausdorff space, and $\mathcal{F} \subseteq \mathcal{C}_b(X)$ multiplicatively closed. Then $\mathcal{F}$ is convergence determining for Radon probability measures if and only if $\mathcal{F}$ generates the topology of X.*

Corollary 2.8. *The set Π of polynomials is convergence determining.*

Proof. $\mathfrak{X}$ is a Polish space, hence completely regular and all probability measures on it are Radon. Π is an algebra, thus multiplicatively closed, and we can apply the Le Cam theorem. ☐

Gromov's metric $\underline{\Box}_\lambda$. To obtain the Gromov-Prohorov metric, we embed the metric spaces and measure the distance of the resulting push forward measures with the Prohorov metric. For Gromov's $\underline{\Box}_\lambda$ metric, it works the opposite way. Namely, the measure spaces are parametrised by a measure preserving map from $[0,1]$ (with Lebesgue measure), and then the distance of the resulting pull backs of the metrics is evaluated with the following metric.

Definition 2.9 ($\Box_\lambda$ metric)**.** Let $(X, \mathfrak{B}, \mu)$ be a probability space. For functions $r, s\colon X \times X \to \mathbb{R}$, we define

$$\Box_\lambda(r,s) \;:=\; \inf\Big\{\, \varepsilon > 0 \;\Big|\; \exists X_\varepsilon \in \mathfrak{B} : \|r\restriction_{X_\varepsilon\times X_\varepsilon} - s\restriction_{X_\varepsilon\times X_\varepsilon})\|_\infty \le \varepsilon,\;\; \mu(X \setminus X_\varepsilon) \le \lambda\varepsilon \,\Big\}.$$

Obviously, we have

$$\Box_\lambda \;\le\; \Box_{\lambda'} \;\le\; \tfrac{\lambda}{\lambda'}\Box_\lambda \qquad \forall \lambda > \lambda'.$$

Definition 2.10 (Gromov's $\underline{\Box}_\lambda$ metric)**.** Let $\mathcal{X}, \mathcal{X}'$ be mm-spaces, and $I := [0,1]$, equipped with Lebesgue measure. Let $\mathcal{F}(\mathcal{X}) := \big\{\, \varphi\colon I \to X \;\big|\; \varphi \text{ is measure preserving} \,\big\}$ be the set of **parametrisations** of (X, μ), and for $\varphi \in \mathcal{F}(\mathcal{X})$ let $d_\varphi(s,t) := d\big(\varphi(s), \varphi(t)\big)$ be the pull back of d with φ. Then we define

$$\underline{\Box}_\lambda(\mathcal{X}, \mathcal{X}') \;:=\; \inf_{\substack{\varphi\in\mathcal{F}(\mathcal{X})\\ \varphi'\in\mathcal{F}(\mathcal{X}')}} \Box_\lambda(d_\varphi, d'_{\varphi'}).$$

Remark 2.11. Because (X, d) is a Polish space, the set $\mathcal{F}(\mathcal{X})$ of (measure preserving) parametrisations is non-empty. This follows for example from the version of the Skorohod representation on I given in [Bog07, Thm. 8.5.4].

Related topologies.

1. In [Fuk87], Fukaya introduced the **measured Hausdorff topology** (often cited as measured Gromov-Hausdorff topology) for compact mm-spaces. The same topology is called **weighted Gromov-Hausdorff topology**, and a complete metric inducing it is constructed by Evans and Winter in [EW06]. The idea is that spaces are close if there is an ε-isometry mapping one measure Prohorov-close to the other. Convergence in measured Hausdorff topology implies Gromov-weak convergence, but not vice versa, because the former implies Gromov-Hausdorff convergence of the underlying metric spaces, which is not the case for Gromov-weak topology. Note that the underlying equivalence classes are also different: For two mm-spaces to be equivalent in the measured Hausdorff topology, the whole spaces have to be isometric, while in a Gromov-weak sense, this is required only for the supports of the measures.
2. Recently, Abraham, Delmas and Hoscheit ([ADH13]) extended the measured Hausdorff topology to complete, locally compact, rooted length spaces with locally finite measures. Note that these measures are finite on all balls, because closed balls are compact in such spaces. The authors introduced the **Gromov-Hausdorff-Prohorov metric**, first on compact spaces using an embedding and measuring the sum of Hausdorff and Prohorov distance. That this metrises measured Hausdorff topology is easy to see from the definitions, using the same connection between ε-isometries and Hausdorff-close embeddings that is frequently applied in the context of Gromov-Hausdorff convergence. In the locally compact setting, they integrate the weighted distances of the measures restricted to balls. Note that this extended topology is vague in the sense that the total mass is not preserved. Thus, on spaces with finite (not necessarily probability) measures, it is not stronger than the natural extension of Gromov-weak topology, where the measures in Definition 2.5 are no longer required to be probabilities.
3. In [Stu06], Sturm defines the **L_2-transportation distance** analogously to d_{GP}, but with the (2-)Wasserstein metric instead of the Prohorov metric. It induces a topology on $\mathfrak{X}$ that is strictly stronger than Gromov-weak topology, but coincides with it on subspaces of $\mathfrak{X}$ consisting of spaces with uniformly bounded (essential) diameter. Its restriction to the space of compact mm-spaces is strictly weaker than measured Hausdorff topology.

3. Equivalence of d_{GP} and $\underline{\Box}_1$

Theorem 3.1. $d_{\mathrm{GP}} = \frac{1}{2}\underline{\Box}_{\frac{1}{2}}$.

Proof. Let $\mathcal{X}_i = (X_i, d_i, \mu_i)$, $i = 1, 2$, be mm-spaces.

"$\geq$": Assume $d_{\mathrm{GP}}(\mathcal{X}_1, \mathcal{X}_2) < \varepsilon$ for some $\varepsilon > 0$. Then we can embed (X_i, d_i), $i = 1, 2$, into a (common) complete, separable metric space (X, d), such that the push forward measures ν_i satisfy $d_{\mathrm{Pr}}(\nu_1, \nu_2) < \varepsilon$. Thus, there is a coupling ν of ν_1 and ν_2 on X^2 with

$$\nu(Y_\varepsilon) \leq \varepsilon \qquad \text{for} \qquad Y_\varepsilon := \big\{ (x, y) \in X^2 \bigm| d(x, y) \geq \varepsilon \big\}.$$

Now choose a parametrisation φ of (X^2, ν), i.e. $\varphi\colon [0, 1] \to X^2$ is measurable and $\nu = \lambda \circ \varphi^{-1}$ for Lebesgue measure λ. Let π_i, $i = 1, 2$, be the canonical projections from X^2 to X. Then $\varphi_i := \pi_i \circ \varphi$ is a parametrisation of $\mathcal{X}_i$ (or its isomorphic image in X). Let r_i be the pull back of d under φ_i. We show $\Box_{\frac{1}{2}}(r_1, r_2) \leq 2\varepsilon$. Indeed, $\lambda\big(\varphi^{-1}(Y_\varepsilon)\big) = \nu(Y_\varepsilon) \leq \varepsilon = \frac{1}{2}2\varepsilon$, and for $s, t \in [0, 1] \setminus \varphi^{-1}(Y_\varepsilon)$ we have by definition of Y_ε that $d\big(\varphi_1(s), \varphi_2(s)\big) \leq \varepsilon$. Thus,

$$r_1(s, t) = d\big(\varphi_1(s), \varphi_1(t)\big) \leq d\big(\varphi_2(s), \varphi_2(t)\big) + 2\varepsilon = r_2(s, t) + 2\varepsilon,$$

and by symmetry, $\big|r_1(s, t) - r_2(s, t)\big| \leq 2\varepsilon$. In total, $\underline{\Box}_{\frac{1}{2}}(\mathcal{X}_1, \mathcal{X}_2) \leq \Box_{\frac{1}{2}}(r_1, r_2) \leq 2\varepsilon$.

"$\leq$": Let $\underline{\square}_{\frac{1}{2}}(\mathcal{X}_1, \mathcal{X}_2) < 2\varepsilon$ and $\varphi_i\colon [0,1] \to X_i$ parametrisations of $\mathcal{X}_i$, $i = 1,2$, with $\square_{\frac{1}{2}}(r_1, r_2) < 2\varepsilon$, where r_i is the pull back of d_i with φ_i. There is a set $S \subseteq [0,1]$ with $\lambda(S) \geq 1 - \varepsilon$ and $|r_1 - r_2| \leq 2\varepsilon$ on S^2. On the disjoint union $X := X_1 \uplus X_2$, we define a metric d by

$$(1) \qquad d\restriction_{X_i^2} := d_i \qquad \text{and} \qquad d(x,y) := \inf_{s \in S} d_1\big(x, \varphi_1(s)\big) + d_2\big(\varphi_2(s), y\big) + \varepsilon \ \ \forall x \in X_1,\, y \in X_2.$$

We check that d satisfies the $\triangle$-inequality in Lemma 3.3 below. Extend the μ_i to measures on X with support in X_i. To estimate their Prohorov distance in (X, d), let $F \subseteq X$ be measurable. Note that by definition, $d\big(\varphi_1(s), \varphi_2(s)\big) = \varepsilon$ for every $s \in S$. Consequently, for every $\varepsilon_0 > \varepsilon$,

$$\varphi_2\big(\varphi_1^{-1}(F) \cap S\big) \subseteq F^{\varepsilon_0} \qquad \text{where} \qquad F^{\varepsilon_0} = \big\{\, x \in X \bigm| d(x,F) < \varepsilon_0 \,\big\}.$$

Therefore,

$$\mu_1(F) = \lambda\big(\varphi_1^{-1}(F)\big) \leq \lambda\big(\varphi_1^{-1}(F) \cap S\big) + \varepsilon \leq \mu_2\Big(\varphi_2\big(\varphi_1^{-1}(F) \cap S\big)\Big) + \varepsilon \leq \mu_2(F^{\varepsilon_0}) + \varepsilon.$$

Since $\varepsilon_0 > \varepsilon$ is arbitrary, $d_{\mathrm{Pr}}(\mu_1, \mu_2) \leq \varepsilon$ and thus $d_{\mathrm{GP}}(\mathcal{X}_1, \mathcal{X}_2) \leq \varepsilon$. $\square$

Corollary 3.2. *For every $\lambda > 0$, we have*

$$\min\{2, \tfrac{1}{\lambda}\} \cdot d_{\mathrm{GP}} \leq \underline{\square}_\lambda \leq \max\{2, \tfrac{1}{\lambda}\} \cdot d_{\mathrm{GP}}.$$

In particular, $\underline{\square}_1$ induces the Gromov-weak topology.

Proof. For $\lambda \geq \frac{1}{2}$, the equation $\underline{\square}_{\frac{1}{2}} \leq 2\lambda \underline{\square}_\lambda \leq 2\lambda \underline{\square}_{\frac{1}{2}}$ is obvious from the definition of $\underline{\square}_\lambda$. For $\lambda \leq \frac{1}{2}$, we get the same inequality with "$\geq$" instead of "$\leq$". Now the theorem implies the claim. $\square$

We still have to check that (1) in the proof of Theorem 3.1 defines a metric.

Lemma 3.3. *The d defined in* (1) *satisfies the $\triangle$-inequality. Thus, it is a metric.*

Proof. For $x, m \in X_1$, $y \in X_2$, we have

$$d(x,y) \leq \inf_{s \in S} d_1(x,m) + d_1\big(m, \varphi_1(s)\big) + d_2\big(\varphi_2(s), y\big) + \varepsilon = d(x,m) + d(m,y).$$

For $x, y \in X_1$, $m \in X_2$, we have

$$\begin{aligned} d(x,y) &\leq \inf_{s,t \in S} d_1\big(x, \varphi_1(s)\big) + d_1\big(\varphi_1(s), \varphi_1(t)\big) + d_1\big(\varphi_1(t), y\big) \\ &\leq \inf_{s,t \in S} d_1\big(x, \varphi_1(s)\big) + d_2\big(\varphi_2(s), \varphi_2(t)\big) + d_1\big(\varphi_1(t), y\big) + 2\varepsilon \\ &\leq \inf_s d_1\big(x, \varphi_1(s)\big) + d_2\big(\varphi_2(s), m\big) + \varepsilon + \inf_t d_2\big(m, \varphi_2(t)\big) + d_1\big(\varphi_1(t), y\big) + \varepsilon \\ &= d(x,m) + d(m,y). \end{aligned}$$

All other cases follow by symmetry or by the $\triangle$-inequalities in X_1 and X_2. $\square$

4. Continuity of the coding of $\mathbb{R}$-trees by excursions

An $\mathbb{R}$-tree (see [DMT96]) is a complete, connected 0-hyperbolic metric space (T, d). One of the possible definitions of 0-hyperbolicity is that it satisfies the four point condition, i.e.

$$d(v_1, v_2) + d(v_3, v_4) \leq \max\big\{\, d(v_1, v_3) + d(v_2, v_4),\ d(v_1, v_4) + d(v_2, v_3) \,\big\} \qquad \forall v_1, \ldots, v_4 \in T.$$

Note that every 0-hyperbolic space can be embedded isometrically into a unique smallest $\mathbb{R}$-tree (see [Eva07, Thm. 3.38]), which is separable whenever the original space was separable. Because d_{GP} (unlike the measured Hausdorff topology) identifies a metric measure space with every subspace containing the support of the measure, the equivalence class of every 0-hyperbolic space contains an $\mathbb{R}$-tree.

One possibility to construct 0-hyperbolic spaces is to code them by excursions, see [Ald93, LG93, DLG02]. To this end, let $h\colon [0,1] \to \mathbb{R}_+$ be a positive function with $h(0) = 0$, and consider the semi-metric

$$d_h(s,t) := h(s) + h(t) - 2I_h(s,t), \qquad I_h(s,t) := \inf_{u \in [s\wedge t,\, s\vee t]} h(u),$$

on $[0,1]$. Then the quotient space $T_h := [0,1]/d_h$ is a 0-hyperbolic metric space. We additionally assume that h is lower semi-continuous. Then T_h is separable and the natural projection

$$\pi_h\colon [0,1] \to T_h$$

is measurable. To see this, note that the canonical projection from the graph $\mathrm{gr}(h) = \{ (t, h(t)) \mid t \in [0,1] \} \subseteq \mathbb{R}^2$ of h onto the tree T_h is continuous due to lower semi-continuity of h. T_h needs to be neither complete nor connected, but we identify it with its completion and, once we have put a measure on it, the equivalence class contains a connected representative.

Remark 4.1. 1. If the graph of h is connected, then T_h is complete and connected to begin with. We do not, however, make this restriction.
2. If h is continuous, π_h is continuous and T_h is compact. Conversely, every compact $\mathbb{R}$-tree can be coded by a (non-unique) continuous excursion ([EW06, Rem. 3.2]). To code compact *measured* trees, continuous excursions are not sufficient. See [Duq06] for a detailed account on coding compact, rooted, ordered, measured $\mathbb{R}$-trees in a unique way by upper semi-continuous càglàd excursions.

Definition 4.2. We define the set of (generalised) **excursions** on $[0,1]$ as

$$\mathcal{E} := \{ h\colon [0,1] \to \mathbb{R}_+ \mid h(0) = 0,\ h \text{ lower semi-continuous} \}.$$

Let $\mathcal{E}_b$ be the subset of bounded functions in $\mathcal{E}$. For $h \in \mathcal{E}$, let the mass measure μ_h on T_h be the image of Lebesgue measure λ under π_h and define the **coding function**

$$\mathfrak{C}\colon \mathcal{E} \to \mathfrak{X}, \qquad h \mapsto \mathcal{T}_h := (T_h, d_h, \mu_h).$$

It is shown in [ADH14, Prop. 2.9] that the coding function $\mathfrak{C}$ is Lipschitz continuous when the space of excursions is equipped with the uniform metric and the space of trees with the Gromov-Hausdorff-Prohorov metric. For the Gromov-Prohorov metric, this is a slightly weaker statement. The proof, however, becomes trivial in this case if we use Theorem 3.1, because the trees are already given in a parameterised form.

Proposition 4.3. *Let $h, g \in \mathcal{E}$. Then*

$$d_{\mathrm{GP}}(\mathcal{T}_h, \mathcal{T}_g) \le 2\|h - g\|_\infty = 2 \sup_{t\in[0,1]} \bigl|h(t) - g(t)\bigr|.$$

Proof. $d_{\mathrm{GP}}(\mathcal{T}_h, \mathcal{T}_g) = \frac{1}{2}\underline{\square}_{\frac{1}{2}}(\mathcal{T}_h, \mathcal{T}_g) \le \frac{1}{2}\square_{\frac{1}{2}}(d_h, d_g) \le 2\|h - g\|_\infty$. $\square$

The uniform metric on $\mathcal{E}$ is a rather strong one, in particular $\mathcal{E}$ and $\mathcal{E}_b$ are not separable in this metric. The coding function turns out to be still continuous if we equip $\mathcal{E}$ with a weaker, separable, metrisable topology, namely the weakest topology which is stronger than convergence in measure and epigraph convergence. For $h, h' \in \mathcal{E}$, let

$$d_\lambda(h, h') := \inf\Bigl\{ \varepsilon > 0 \Bigm| \lambda\bigl(\{ t \mid |h(t) - h'(t)| > \varepsilon \}\bigr) < \varepsilon \Bigr\},$$

which metrises convergence in Lebesgue measure, d_H the Hausdorff metric in $\mathbb{R}^2$, and

$$d_\Gamma(h, h') := d_H\bigl(\mathrm{epi}(h), \mathrm{epi}(h')\bigr), \qquad \mathrm{epi}(h) := \{ (t, y) \in [0,1] \times \mathbb{R}_+ \mid y \ge h(t) \}.$$

Note that the epigraph of a function is closed if and only if the function is lower semi-continuous. Epigraph convergence is usually defined as convergence in Fell topology (or equivalently Kuratowski convergence) of the epigraphs, see e.g. [Bee93]. It is a compact, metrisable topology on the set $\overline{\mathcal{E}}$ of $(\mathbb{R}_+ \cup \{\infty\})$-valued, lower semi-continuous functions on $[0,1]$. On $\overline{\mathcal{E}}$, the topology induced by d_Γ is strictly stronger. Restricted to $\mathcal{E}$, however, the topologies coincide, which follows from [Bee94, Thm. 1] using compactness of $[0,1]$ and $\mathbb{R}$-valuedness of excursions. Epigraph convergence also coincides with Γ-convergence (see e.g. [Mas93]), whence the name d_Γ.

Definition 4.4. We endow $\mathcal{E}$ with the **excursion metric** $d_{\mathcal{E}} := d_\Gamma + d_\lambda$.

Recall that a metrisable topological space X is called *Lusin space* if it is the continuous, injective image of a Polish space, i.e. if there exists a Polish space Y and a continuous bijection $f\colon Y \to X$. X is Lusin if and only if it is homeomorphic to a Borel subset of a Polish space (see [Coh80, Sec. 8.6] for details).

Proposition 4.5. *$\mathcal{E}$ is a separable metric space, and the set of continuous excursions is dense. Furthermore, $\mathcal{E}_b$ is a Lusin space.*

Proof. $d_{\mathcal{E}}$ is obviously a metric, and the continuous excursions are both d_Γ-dense (increasing pointwise convergence implies d_Γ-convergence) and d_λ-dense in $\mathcal{E}$. Hence, $\mathcal{E}$ is separable, and it remains to show that $\mathcal{E}_b$ is a Borel subset of a Polish space. First note that this is the case for $(\mathcal{E}_b, d_\Gamma)$, because the set of excursions bounded by a fixed $M \in \mathbb{N}$ is closed in the compact metric space $\overline{\mathcal{E}}$ with epigraph topology. Now we can identify $(\mathcal{E}_b, d_{\mathcal{E}})$ with the graph of the function $\pi\colon (\mathcal{E}_b, d_\Gamma) \to L^0 := \big(L^0(\lambda), d_\lambda\big)$, which maps an excursion to its λ-a.e. equivalence class. It is enough to show that π is measurable, because then $(\mathcal{E}_b, d_{\mathcal{E}}) \cong \mathrm{gr}(\pi)$ is an injective measurable image of a Lusin space, hence Lusin itself by [Coh80, Thm. 8.3.7].

To show measurability, choose a fixed dense sequence $(f_n)_{n\in\mathbb{N}}$ of continuous excursions, and define $\pi_n\colon \mathcal{E}_b \to L^0$, $h \mapsto \sup_{f_k \le h,\, k\le n} f_k$. Then π_n is a simple function and measurable, because $\{\, h \in \mathcal{E}_b \mid h \ge f_k \,\}$ is closed in $(\mathcal{E}_b, d_\Gamma)$. Because $h = \sup_{f_n \le h} f_n$, π is the pointwise limit of the π_n, thus also measurable. $\square$

Example 4.6 ($d_{\mathcal{E}}$ is not complete and $\mathfrak{C}$ is not uniformly continuous)**.** Let $h_n(t) = 1 - \mathbb{1}_{\mathbb{N}_0}(nt)$, $t \in [0,1]$. Then h_n codes the discrete space of n points with uniform distribution or, equivalently, the star-shaped tree with n leaves and uniform distribution on the leaves. h_n converges in epigraph topology to the zero function, while $d_\lambda(h_n, \mathbb{1}) = 0$ for each n. Thus, $(h_n)_{n\in\mathbb{N}}$ is Cauchy w.r.t. $d_{\mathcal{E}}$, but does not converge. $\big(\mathfrak{C}(h_n)\big)_{n\in\mathbb{N}}$ is not a Cauchy sequence in $\mathfrak{X}$, hence $\mathfrak{C}$ is not uniformly continuous.

Remark 4.7. We do not know if $\mathcal{E}$ is Lusin or even Polish. $\mathcal{E}_b$ is not Polish, because it is a dense $\mathcal{F}_\sigma$-set (countable union of closed sets) with dense complement (in $\mathcal{E}$).

That such a set cannot be Polish can be seen as follows. Let A_n be closed with dense complement in $\mathcal{E}$. Then its closure $\overline{A}_n$ in $\overline{\mathcal{E}}$ is closed with empty interior in the Polish space $\overline{\mathcal{E}}$. Assume that $A := \bigcup_{n\in\mathbb{N}} A_n$ is Polish. By the Mazurkiewicz theorem ([Coh80, Thm. 8.1.4]), A is a $\mathcal{G}_\delta$-set in $\overline{\mathcal{E}}$, i.e. $A = \bigcap_{n\in\mathbb{N}} U_n$ for some open sets $U_n \subseteq \overline{\mathcal{E}}$. Let $A'_n := \overline{\mathcal{E}} \setminus U_n$. Then $\overline{\mathcal{E}} = \bigcup_{n\in\mathbb{N}} \big(\overline{A}_n \cup A'_n\big)$ and by the Baire category theorem ([Coh80, Thm. D.37]), at least one A'_n has to have non-empty interior. This means that A is not dense.

Theorem 4.8. *The coding function $\mathfrak{C}\colon \mathcal{E} \to \mathfrak{X}$ is continuous (w.r.t. $d_{\mathcal{E}}$ and d_{GP}).*

Proof. Fix $h \in \mathcal{E}$, $\varepsilon > 0$. We construct a $\delta > 0$ such that $\square_1(d_h, d_g) \le 6\varepsilon$ for every $g \in \mathcal{E}$ with $d_{\mathcal{E}}(h,g) \le \delta$. Then Corollary 3.2 implies the result.

1. Let $A_\eta := \big\{\, t \in [0,1] \,\big|\, I_h(t-\eta, t+\eta) < h(t) - \varepsilon \,\big\}$. Because h is lower semi-continuous, $A_\eta \searrow \emptyset$ for $\eta \to 0$. Thus, there is a $0 < \delta < \varepsilon$ with $\lambda(A_\delta) < \varepsilon$. Fix $g \in \mathcal{E}$ with $d_{\mathcal{E}}(h,g) \le \delta$ and let $X_\varepsilon := [0,1] \setminus \big(A_\delta \cup \{\, |h-g| > \delta \,\}\big)$. Then $\lambda\big([0,1]\setminus X_\varepsilon\big) \le 2\varepsilon$ and it is enough to show $\big|d_h(s,t) - d_g(s,t)\big| \le 6\varepsilon$ for $s,t \in X_\varepsilon$. Because h and g are ε-close at s and t, this is satisfied once we have shown $\big|I_h(s,t) - I_g(s,t)\big| \le 2\varepsilon$.

2. "$I_g \le I_h + 2\varepsilon$": Because h is lower semi-continuous, the infimum $I_h(s,t)$ is attained and there is a $u \in [s,t]$ with $h(u) = I_h(s,t)$. From $d_\Gamma(h,g) \le \delta$, we obtain the existence of $u' \in [u-\delta, u+\delta]$ with $g(u') \le h(u) + \delta$. If $u' \in [s,t]$, then $I_g(s,t) \le g(u') \le h(u) + \delta \le I_h(s,t) + \varepsilon$. For the case $u' \notin [s,t]$, assume w.l.o.g. $u' < s$, and therefore $u \in [s, s+\delta]$. Then, because s is not in A_δ, we have $I_h(s,t) = h(u) \ge h(s) - \varepsilon \ge g(s) - 2\varepsilon \ge I_g(s,t) - 2\varepsilon$.
3. "$I_h \le I_g + 2\varepsilon$": Choose $u \in [s,t]$ with $g(u) = I_g(s,t)$ and $u' \in [u-\delta, u+\delta]$ with $h(u') \le g(u) + \delta$. As above, we can assume $u \in [s, s+\delta]$, $u' \in [s-\delta, s]$ and obtain $I_h(s,t) \le h(s) \le h(u') + \varepsilon \le g(u) + 2\varepsilon = I_g(s,t) + 2\varepsilon$. □

Acknowledgements. I am thankful to Anita Winter for discussions, encouragement, and helpful comments on the previous version of the manuscript. I also thank Guillaume Voisin for many discussions about trees, Patrick Hoscheit for a discussion about topologies on the space of excursions, and the referees for helpful comments.

References

[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** (2013), no. 14, 21 pp.

[ADH14] ———, *Exit times for an increasing Lévy tree-valued process*, Probab. Theory Related Fields **159** (2014), no. 1-2, 357–403. MR 3201925

[Ald93] David Aldous, *The continuum random tree III*, Ann. Probab. **21** (1993), no. 1, 248–289.

[Bee93] Gerald Beer, *Topologies on closed and closed convex sets*, Kluwer Acad. Publ., 1993.

[Bee94] ———, *A note on epi-convergence*, Canad. Math. Bull. **37** (1994), no. 3, 294–239.

[Bog07] V. I. Bogachev, *Measure theory, volume II*, Springer, 2007.

[Coh80] Donald L. Cohn, *Measure theory*, Birkhäuser, 1980.

[DGP11] Andrej Depperschmidt, Andreas Greven, and Peter Pfaffelhuber, *Marked metric measure spaces*, Electron. Commun. Probab. **16** (2011), 174–188.

[DGP12] Andrej Depperschmidt, Andreas Greven, and Peter Pfaffelhuber, *Tree-valued Fleming-Viot dynamics with selection*, Ann. Appl. Probab. **22** (2012), no. 6, 2560–2615.

[DLG02] Thomas Duquesne and Jean-François Le Gall, *Random trees, Lévy processes and spatial branching processes*, Astérisque **281** (2002), vi+147.

[DMT96] Andreas W.M. Dress, V. Moulton, and W.F. Terhalle, *T-theory: An overview*, Europ. J. Combinatorics **17** (1996), no. 2-3, 161–175.

[Duq03] Thomas Duquesne, *A limit theorem for the contour process of conditioned Galton-Watson trees*, Annals of Prob. **31** (2003), no. 2, 996–1027.

[Duq06] Thomas Duquesne, *The coding of compact real trees by real valued functions*, 2006, arXiv:0604106.

[Eva07] Steven N. Evans, *Probability and real trees*, École d'Été de Probabilités de Saint Flour XXXV-2005, Lecture Notes in Mathematics, vol. 1920, Springer, 2007.

[EW06] Steven N. Evans and Anita Winter, *Subtree prune and re-graft: A reversible real-tree valued Markov chain*, Ann. Probab. **34** (2006), no. 3, 918–961.

[Fuk87] Kenji Fukaya, *Collapsing of Riemannian manifolds and eigenvalues of Laplace operator*, Invent. Math. **87** (1987), no. 3, 517–547.

[GPW09] Andreas Greven, Peter Pfaffelhuber, and Anita Winter, *Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees)*, Probab. Theory Related Fields **145** (2009), 285–322. MR 2520129

[GPW13] ———, *Tree-valued resampling dynamics: Martingale problems and applications*, Probab. Theory Related Fields **155** (2013), no. 3–4, 789–838.

[Gro99] Misha Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, vol. 152, Birkhäuser Boston Inc., Boston, MA, 1999.

[HJ77] J. Hoffmann-Jørgensen, *Probability in Banach spaces*, École d'Été de Probabilités de Saint Flour VI-1976, Lecture Notes in Mathematics, vol. 598, Springer, 1977.

[LC57] L. Le Cam, *Convergence in distribution of stochastic processes*, University of California Publications in Statistics **2** (1957), 207–236.

[LG93] Jean-François Le Gall, *The uniform random tree in a Brownian excursion*, Probab. Theory Related Fields **96** (1993), no. 3, 369–383.

[Mas93] Gianni Dal Maso, *An introduction to Γ-convergence*, Birkhäuser, 1993.

[Stu06] Karl-Theodor Sturm, *On the geometry of metric measure spaces I*, Acta Math. **196** (2006), no. 1, 65–131.

EXISTENCE OF MARK FUNCTIONS IN MARKED METRIC MEASURE SPACES

SANDRA KLIEM AND WOLFGANG LÖHR

ABSTRACT. We give criteria on the existence of a so-called mark function in the context of marked metric measure spaces (mmm-spaces). If an mmm-space admits a mark function, we call it functionally-marked metric measure space (fmm-space). This is not a closed property in the usual marked Gromov-weak topology, and thus we put particular emphasis on the question under which conditions it carries over to a limit. We obtain criteria for deterministic mmm-spaces as well as random mmm-spaces and mmm-space-valued processes. As an example, our criteria are applied to prove that the tree-valued Fleming-Viot dynamics with mutation and selection from [DGP12] admits a mark function at all times, almost surely. Thereby, we fill a gap in a former proof of this fact, which used a wrong criterion.

Furthermore, the subspace of fmm-spaces, which is dense and not closed, is investigated in detail. We show that there exists a metric that induces the marked Gromov-weak topology on this subspace and is complete. Therefore, the space of fmm-spaces is a Polish space. We also construct a decomposition into closed sets which are related to the case of uniformly equicontinuous mark functions.

CONTENTS

1. INTRODUCTION

A metric (finite) measure spaces (*mm-space*) is a complete, separable metric space (X, r) together with a finite measure ν on it. Considering the space of (equivalence classes of) mm-spaces itself as a metric space dates back to Gromov's invention of the $\underline{\square}_\lambda$-metric in [Gro99, Chapter $3\frac{1}{2}$]. Motivated by Aldous' work on the Brownian continuum random tree ([Ald93]), it was realised in [GPW09] that the space of mm-spaces is a useful state space for tree-valued stochastic processes, and Polish when equipped with the Gromov-weak topology. That the Gromov-weak topology actually coincides with the one induced by the $\underline{\square}_\lambda$-metric was shown in [Löh13].

Important examples for the use of mm-spaces within probability theory are individual-based populations X with given mutual genealogical distances r between individuals. Here, r can for instance measure the time to the most recent common ancestor (MRCA) (cf. [DGP12, (2.7), Remark 3.3]), where the resulting metric space is ultrametric. Another possibility is the number of mutations back to the MRCA (cf. [KW15]), where the resulting space is not ultrametric. Finally, there is a sampling measure ν on the space X which models population density. This means that the state of the process is an mm-space (X, r, ν). Such individual-based models are often formulated for infinite population size (with diffuse measures ν) but obtained as the high-density limit of approximating models with finite populations (where ν is typically the uniform distribution on all individuals).

For encoding more information about the individuals, such as an (allelic) type or location (which may change over time), marked metric measure spaces (mmm-spaces) and the corresponding marked Gromov-weak topology (mGw-topology) have been introduced in [DGP11]. For a fixed complete, separable metric space (I, d) of marks, the sampling measure ν is replaced by a measure μ on $X \times I$, which models population density in combination with mark distribution.

A natural question in this context is whether or not every point of the limiting population X has a single mark almost surely, that is, does genetic distance zero imply the same type/location? Put differently, we ask ourselves if μ factorizes into a "population density" measure ν on X and a mark function $\kappa\colon X \to I$ assigning each individual its mark. If this is the case, we call the mmm-space functionally-marked (fmm-space). This property is often desirable, and one might want to consider the space of fmm-spaces, rather than mmm-spaces, as the state space. Unfortunately, the subspace of fmm-spaces is not closed in the mGw-topology, which means that limits of finite-population models that are constructed as fmm-spaces might not admit mark functions themselves. It is therefore of interest to know, if the space of fmm-spaces with marked Gromov-weak topology is a Polish space (that is a "good" state space). Here, we show in Theorem 2.2 that this is indeed the case. We also produce criteria to enable one to check if an mmm-space admits a mark function. For limiting populations, they are given in terms of the approximating mmm-spaces. We derive such criteria for deterministic spaces (Theorem 3.1), random spaces (Theorem 3.7) and mmm-space-valued processes (Theorem 3.9 and Theorem 3.11).

An important example of such a high-density limit of approximating models with finite populations is the tree-valued Fleming-Viot dynamics. In the neutral case, it is constructed in [GPW13] using the formalism of mm-spaces. In [DGP12], (allelic) types – encoded as marks of mmm-spaces – are included, in order to model mutation and selection. For this process, the question of existence of a mark function has already been posed. In [DGP12, Remark 3.11] and [DGP13, Theorem 6] it is stated that the tree-valued Fleming-Viot process admits a mark function at all times, almost surely. The given proof, however, contains a gap, because it relies on the criterion claimed in [DGP13, Lemma 7.1], which is wrong in general, as we show in Example 4.1. We fill this gap by applying our criteria and showing in Theorem 4.3 that the claim is indeed true and the tree-valued Fleming-Viot process with mutation and selection (TFVMS) admits a mark function at all times, almost surely. We also show in Theorem 4.4 that the same arguments apply to the Λ-version of the TFVMS in the neutral case, that is where selection is not present.

Intuitively, the existence of a mark function in the case of the TFVMS holds because mutations are large but rare in the approximating sequence of tree-valued Moran models. Hence, as genealogical distance becomes small, the probability that any mutation happened at all in the close past becomes small as well (recall that distance equals time to the MRCA). In contrast, in [KW15], where evolving phylogenies of trait-dependent branching with mutation and competition are under investigation, mutations happen at a high rate but are small which justifies the hope for the existence of a mark function also for the limiting model. Our criteria are also well-suited for this kind of situation.

Outline. The paper is organized as follows. In the subsections of the introduction we first introduce notations and basic results for the Prohorov metric for finite measures. Then, we give a short introduction to the space $\mathbb{M}_I$ of marked metric measure spaces (mmm-spaces) with the marked Gromov-weak topology, as well as the marked Gromov-Prohorov metric d_{mGP} on it. We continue with defining the so-called functionally-marked metric measure spaces (fmm-spaces) $\mathbb{M}_I^{\mathrm{fct}} \subseteq \mathbb{M}_I$, and finally investigate the case of equicontinuous mark functions as an illustrative example. We emphasize that the restriction of the marked Gromov-Prohorov metric d_{mGP} to $\mathbb{M}_I^{\mathrm{fct}}$ is not complete.

In Section 2, we therefore show that there exists another metric on $\mathbb{M}_I^{\mathrm{fct}}$ that induces the marked Gromov-weak topology and is complete. As one sees in Subsection 1.4, the situation becomes easy if we restrict to a subspace of $\mathbb{M}_I$ containing spaces with uniformly equicontinuous mark functions. We introduce in Subsection 2.2 several related subspaces capturing some aspect of equicontinuity, and obtain a decomposition of $\mathbb{M}_I^{\mathrm{fct}}$ into closed sets. This decomposition is used to prove Polishness of $\mathbb{M}_I^{\mathrm{fct}}$, and in Section 3 to formulate criteria for the existence of mark functions.

Section 3 gives criteria for the existence of mark functions. Based on the construction of the complete metric and the decomposition of $\mathbb{M}_I^{\mathrm{fct}}$, we derive in Subsection 3.1 criteria to check if an mmm-space admits a mark function, especially in the case where it is given as a marked Gromov-weak limit. We then transfer the results in Subsection 3.2 to random mmm-spaces and in Subsection 3.3 to $\mathbb{M}_I$-valued stochastic processes.

To conclude, Section 4 gives examples. We first show that the criterion in [DGP13] is wrong in general by means of counterexamples. Our criteria are then applied in Subsection 4.1 to prove the existence of a mark function for the tree-valued Fleming-Viot dynamics with mutation and selection. To this goal, we verify the necessary assumptions for a sequence of approximating tree-valued Moran models. In Subsection 4.2 we show that a similar strategy applies if we replace the tree-valued Moran models by so-called tree-valued Λ-Cannings models. Finally, in Subsection 4.3, a future application to evolving phylogenies of trait-dependent branching with mutation and competition is indicated.

1.1. **Notations and prerequisites.** In this paper, let all topological spaces be equipped with their Borel σ-algebras. We use the following notation throughout the article.

Notation 1.1. *For a Polish space E, let $\mathcal{M}_1(E)$ respectively $\mathcal{M}_{\mathrm{f}}(E)$ denote the space of probability respectively finite measures on the Borel σ-algebra $\mathfrak{B}(E)$ on E. The space $\mathcal{M}_{\mathrm{f}}(E)$ is always equipped with the topology of weak convergence, which is denoted by $\xrightarrow{w}$. We also use the distance in variational norm of $\mu, \nu \in \mathcal{M}_{\mathrm{f}}(E)$, which is*

$$\|\mu - \nu\| := \sup_{B \in \mathfrak{B}(E)} \big|\mu(B) - \nu(B)\big|. \tag{1.1}$$

In particular, $\|\mu\| = \mu(E)$, and $\|\mu - \nu\| = \nu(E) - \mu(E)$ if $\mu \leq \nu$, that is $\mu(A) \leq \nu(A)$ for all $A \in \mathfrak{B}(E)$.

For $Y \in \mathfrak{B}(E)$ and $\mu \in \mathcal{M}_{\mathrm{f}}(E)$, denote by $\mu|_Y \in \mathcal{M}_{\mathrm{f}}(E)$ the restriction of μ to Y, that is $\mu|_Y(B) := \mu(B \cap Y)$ for all $B \in \mathfrak{B}(E)$. Because $\mu|_Y \leq \mu$, we have $\|\mu|_Y - \mu\| = \mu(E \setminus Y)$.

For $\varphi\colon E \to F$ measurable, with F some other Polish space, denote the image measure of μ under φ by $\varphi_*\mu := \mu \circ \varphi^{-1}$. Finally, for the product space $X := E \times F$, the canonical projection operators from X onto E and F are denoted by π_E and π_F, respectively.

Definition 1.2 (Prohorov metric)**.** *For finite measures μ_0, μ_1 on a metric space (E, r), the* Prohorov metric *is defined as*

$$d_{\mathrm{Pr}}(\mu_0, \mu_1) := \inf\big\{\varepsilon > 0 : \mu_i(A) \leq \mu_{1-i}(A^\varepsilon) + \varepsilon \ \ \forall A \in \mathfrak{B}(E),\, i \in \{0, 1\}\big\}, \tag{1.2}$$

where $A^\varepsilon := \{\, x \in E : r(A, x) < \varepsilon \,\}$ is the ε-neighbourhood of A.

It is well-known that the Prohorov metric metrizes the weak convergence of measures if and only if the underlying metric space is separable. The following equivalent expression for the Prohorov metric turns out to be useful.

Remark 1.3 (coupling representation of the Prohorov metric)**.** Let (E,r) be a separable metric space and $\mu_1,\mu_2 \in \mathcal{M}_1(E)$. For a finite measure ξ on E^2, we denote the marginals as $\xi_1 := \xi(\cdot \times E)$ and $\xi_2 := \xi(E \times \cdot)$. It is well-known (see, e.g., [EK05, Theorem III.1.2]) that

$$d_{\mathrm{Pr}}(\mu_1,\mu_2) = \inf\bigl\{\varepsilon > 0 : \exists \xi \in \mathcal{M}_1(E^2) \text{ with } \xi(N_\varepsilon) \le \varepsilon,\ \xi_i = \mu_i,\ i=1,2\bigr\}, \tag{1.3}$$

where $N_\varepsilon := \{\, (x,y) \in E^2 : r(x,y) \ge \varepsilon \,\}$. We obtain from this equation

$$d_{\mathrm{Pr}}(\mu_1,\mu_2) = \inf\bigl\{\varepsilon > 0 : \exists \xi' \in \mathcal{M}_{\mathrm{f}}(E^2) \text{ with } \xi'(N_\varepsilon) = 0,\ \xi_i' \le \mu_i,\ \|\mu_i - \xi_i'\| \le \varepsilon,\ i=1,2\bigr\}. \tag{1.4}$$

Indeed, consider $\xi' := \xi|_{E^2 \setminus N_\varepsilon}$ respectively $\xi := \xi' + (1-\|\xi'\|)^{-1}\bigl((\mu_1 - \xi_1') \otimes (\mu_2 - \xi_2')\bigr)$ to obtain equality in the above. Following the ideas of the proof of the representation (1.3) in [EK05], the representation (1.4) for the Prohorov metric $d_{\mathrm{Pr}}(\mu_0,\mu_1)$ is easily seen to hold true for measures $\mu_1,\mu_2 \in \mathcal{M}_{\mathrm{f}}(E)$ as well, which are not necessarily probability measures.

From (1.4), we can easily deduce the following lemma, which we use below.

Lemma 1.4 (rectangular lemma)**.** *Let (E,r) be a separable, metric space, $\varepsilon,\delta > 0$, and $\mu_1,\mu_2 \in \mathcal{M}_{\mathrm{f}}(E)$. Assume that $d_{\mathrm{Pr}}(\mu_1,\mu_2) < \delta$ and there is $\mu_1' \le \mu_1$ with $\|\mu_1 - \mu_1'\| \le \varepsilon$. Then*

$$\exists \mu_2' \le \mu_2 : d_{\mathrm{Pr}}(\mu_1',\mu_2') < \delta,\ \|\mu_2 - \mu_2'\| \le \varepsilon. \tag{1.5}$$

Proof. According to (1.4), we find $\xi \in \mathcal{M}_{\mathrm{f}}(E^2)$ with marginals $\xi_i \le \mu_i$, $i=1,2$, $\|\mu_i - \xi_i\| < \delta$, and $\xi(\{r \ge \delta\}) = 0$. Let L be a probability kernel from E to E (for existence see [Kle14, Theorems 8.36–8.38]) with $\xi = \mu_1 \otimes L$ and define $\xi' := (\mu_1' \wedge \xi_1) \otimes L$. Obviously, $\xi_1' \le \mu_1'$ and $\|\mu_1' - \xi_1'\| \le \|\mu_1 - \xi_1\| < \delta$. Now set

$$\mu_2' := \xi_2' + \mu_2 - \xi_2. \tag{1.6}$$

Then $\xi_2' \le \mu_2'$, $\|\mu_2' - \xi_2'\| = \|\mu_2 - \xi_2\| < \delta$ and thus $d_{\mathrm{Pr}}(\mu_2,\mu_2') < \delta$ by (1.4). Furthermore, $\mu_2' \le \mu_2$ and $\|\mu_2 - \mu_2'\| = \|\xi_2 - \xi_2'\| \le \|\mu_1 - \mu_1'\| \le \varepsilon$. □

1.2. **The space of marked metric measure spaces (mmm-spaces).** In this subsection, we recall the space $\mathbb{M}_I$ of marked metric measure spaces, and the marked Gromov-Prohorov metric d_{mGP}, which induces the marked Gromov-weak topology on it. This space, $(\mathbb{M}_I, d_{\mathrm{mGP}})$, will be the basic space used in the rest of the paper. These concepts have been introduced in [DGP11], and are based on the corresponding non-marked versions introduced in [GPW09]. In contrast to [DGP11], we allow the measures of the marked metric measure spaces to be finite, that is do not restrict ourselves to probability measures only. Because a sequence of finite measures converges weakly if and only if their total masses and the normalized measures converge, or the masses converge to zero, this straight-forward generalization requires only minor modifications (compare [LVW14, Section 2.1], where this generalization is done for metric measure spaces without marks).

In what follows, fix a complete, separable metric space (I,d), called the *mark space*. It is the same for all marked metric measure spaces in $\mathbb{M}_I$.

Definition 1.5 (mmm-spaces, $\mathbb{M}_I$)**.**
(i) *An (I-)*marked metric measure space (mmm-space) *is a triple (X,r,μ) such that (X,r) is a complete, separable metric space, and $\mu \in \mathcal{M}_{\mathrm{f}}(X \times I)$, where $X \times I$ is equipped with the product topology.*
(ii) *Let $\mathcal{X}_i = (X_i,r_i,\mu_i)$, $i=1,2$, be two mmm-spaces, and $\nu_i := \mu_i(\cdot \times I)$ the marginal of μ_i on X_i. For a map $\varphi\colon X_1 \to X_2$ we use the notation*

$$\tilde\varphi\colon X_1 \times I \to X_2 \times I, \quad (x,u) \mapsto \tilde\varphi(x,u) := (\varphi(x),u). \tag{1.7}$$

We call $\mathcal{X}_1$ and $\mathcal{X}_2$ equivalent *if they are measure- and mark-preserving isometric, that is there is an isometry $\varphi\colon \operatorname{supp}(\nu_1) \to \operatorname{supp}(\nu_2)$, such that*

$$\tilde{\varphi}_* \mu_1 = \mu_2. \tag{1.8}$$

(iii) *Finally, define*

$$\mathbb{M}_I := \big\{\textit{equivalence classes of mmm-spaces}\big\}. \tag{1.9}$$

With a slight abuse of notation, we identify an mmm-space with its equivalence class and write $\mathcal{X} = (X, r, \mu) \in \mathbb{M}_I$ for both mmm-spaces and equivalence classes thereof.

Next, we recall the marked Gromov-weak topology from [DGP11, Section 2.2] that turns $\mathbb{M}_I$ into a Polish space (cf. [DGP11, Theorem 2]). To this goal, we first recall

Definition 1.6 (marked distance matrix distribution)**.** *Let $\mathcal{X} := (X, r, \mu) \in \mathbb{M}_I$ and*

$$R^{(X,r)} := \begin{cases} (X \times I)^{\mathbb{N}} & \to & \mathbb{R}_+^{\binom{\mathbb{N}}{2}} \times I^{\mathbb{N}}, \\ \big((x_k, u_k)_{k\geq 1}\big) & \mapsto & \big(\big(r(x_k, x_l)\big)_{1\leq k<l}, (u_k)_{k\geq 1}\big). \end{cases} \tag{1.10}$$

The marked distance matrix distribution *of $\mathcal{X}$ is defined as*

$$\boldsymbol{\nu}^{\mathcal{X}} := \|\mu\| \cdot \big(R^{(X,r)}\big)_* \big(\tfrac{\mu}{\|\mu\|}\big)^{\mathbb{N}} \in \mathcal{M}_{\mathrm{f}}\big(\mathbb{R}_+^{\binom{\mathbb{N}}{2}} \times I^{\mathbb{N}}\big). \tag{1.11}$$

The marked Gromov-weak topology is the one induced by the map $\mathcal{X} \mapsto \boldsymbol{\nu}^{\mathcal{X}}$.

Definition 1.7 (marked Gromov-weak topology)**.** *Let $\mathcal{X}, \mathcal{X}_1, \mathcal{X}_2, \ldots \in \mathbb{M}_I$. We say that $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}$ in the* marked Gromov-weak topology*, $\mathcal{X}_n \xrightarrow[n\to\infty]{\text{mGw}} \mathcal{X}$, if and only if*

$$\boldsymbol{\nu}^{\mathcal{X}_n} \xrightarrow[n\to\infty]{w} \boldsymbol{\nu}^{\mathcal{X}} \tag{1.12}$$

in the weak topology on $\mathcal{M}_{\mathrm{f}}\big(\mathbb{R}_+^{\binom{\mathbb{N}}{2}} \times I^{\mathbb{N}}\big)$.

Finally, let us recall the Gromov-Prohorov metric from [DGP11, Section 3.2]. It is complete and metrizes the marked Gromov-weak topology, as shown in [DGP11, Proposition 3.7].

Definition 1.8 (marked Gromov-Prohorov metric, d_{mGP})**.** *For $\mathcal{X}_i = (X_i, r_i, \mu_i) \in \mathbb{M}_I, i = 1, 2$, set*

$$d_{\mathrm{mGP}}(\mathcal{X}_1, \mathcal{X}_2) := \inf_{(E,\varphi_1,\varphi_2)} d_{\mathrm{Pr}}\big((\tilde{\varphi}_1)_*\mu_1, (\tilde{\varphi}_2)_*\mu_2\big), \tag{1.13}$$

where the infimum is taken over all complete, separable metric spaces (E, r) and isometric embeddings $\varphi_i\colon X_i \to E$, and $\tilde{\varphi}_i$ is as in (1.7), $i = 1, 2$. The Prohorov metric d_{Pr} is the one on $\mathcal{M}_{\mathrm{f}}(E \times I)$, based on the metric $\tilde{r} = r + d$ on $E \times I$, metrizing the product topology. The metric d_{mGP} is called the marked Gromov-Prohorov metric*.*

A direct consequence of the fact that d_{mGP} induces the marked Gromov-weak topology is the following characterization of marked Gromov-weak convergence obtained in [DGP11, Lemma 3.4].

Lemma 1.9 (embedding of marked Gromov-weakly converging sequences)**.** *Let $\mathcal{X}_n \in \mathbb{M}_I$, $\mathcal{X}_n = (X_n, r_n, \mu_n)$ for $n \in \mathbb{N} \cup \{\infty\}$. Then $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}_\infty$ Gromov-weakly if and only if there is a complete, separable metric space (E, r), and isometric embeddings $\varphi_n\colon X_n \to E$, such that for $\tilde{\varphi}_n$ as in (1.7),*

$$(\tilde{\varphi}_n)_*\mu_n \xrightarrow{w} \mu. \tag{1.14}$$

1.3. **Functionally-marked metric measure spaces (fmm-spaces).** Consider an I-marked metric measure space $\mathcal{X} = (X, r, \mu) \in \mathbb{M}_I$. Since μ is a finite measure on the Polish space $X \times I$, regular conditional measures exist (cf. [Kle14, Theorems 8.36–8.38]), and we write

$$\mu(\mathrm{d}x, \mathrm{d}u) = \nu(\mathrm{d}x) \cdot K_x(\mathrm{d}u), \tag{1.15}$$

in short $\mu = \nu \otimes K$, for the marginal $\nu := \mu(\cdot \times I) \in \mathcal{M}_\mathrm{f}(X)$, and a ($\nu$-a.s. unique) probability kernel K from X to I.

In the present article we investigate criteria for the *existence of a mark function* for $\mathcal{X}$, that is (cf. [DGP13, Section 3.3]) a measurable function $\kappa\colon X \to I$ such that

$$\mu(\mathrm{d}x, \mathrm{d}u) = \nu(\mathrm{d}x) \cdot \delta_{\kappa(x)}(\mathrm{d}u), \tag{1.16}$$

or equivalently, $K_x = \delta_{\kappa(x)}$ for ν-almost every x. Obviously, $\mathcal{X}$ admits a mark function if and only if K_x is a Dirac measure for ν-almost every x. Recall that the complete, separable mark space (I, d) is fixed once and for all.

Definition 1.10 (fmm-spaces, $\mathbb{M}_I^{\mathrm{fct}}$)**.** *We call $\mathcal{X} = (X, r, \nu, \kappa)$ an* (I-)functionally-marked metric measure space *(*fmm-space*) if (X, r) is a complete, separable metric space, $\nu \in \mathcal{M}_\mathrm{f}(X)$, and $\kappa\colon X \to I$ is measurable. We identify $\mathcal{X}$ with the marked metric measure space $(X, r, \mu) \in \mathbb{M}_I$, where μ satisfies* (1.16)*. With a slight abuse of notation, we write $(X, r, \nu, \kappa) = (X, r, \mu)$ if* (1.16) *is satisfied. Denote by $\mathbb{M}_I^{\mathrm{fct}} \subseteq \mathbb{M}_I$ the space of (equivalence classes of) fmm-spaces.*

A first, simple observation is that $\mathbb{M}_I^{\mathrm{fct}}$ is a dense subspace of $\mathbb{M}_I$.

Lemma 1.11. *The subspace $\mathbb{M}_I^{\mathrm{fct}}$ is dense in $\mathbb{M}_I$ with marked Gromov-weak topology.*

Proof. For $\mathcal{X} = (X, r, \mu) \in \mathbb{M}_I$, define $\mathcal{X}_n = (X \times I, r_n, \nu_n, \kappa_n) \in \mathbb{M}_I^{\mathrm{fct}}$ with $\nu_n = \mu$, $\kappa_n(x, u) = u$, and $r_n\big((x, u), (y, v)\big) := r(x, y) + e^{-n} \wedge d(u, v)$, for $x, y \in X,\ u, v \in I$. It is easy to see that $\mathcal{X}_n \to \mathcal{X}$ in the marked Gromov-weak topology. $\square$

1.4. **The equicontinuous case.** It directly follows from Lemma 1.11 that the subspace $\mathbb{M}_I^{\mathrm{fct}}$ is not closed in $\mathbb{M}_I$, meaning that if $\mathcal{X}_n \xrightarrow{\mathrm{mGw}} \mathcal{X}$ is a marked Gromov-weakly converging sequence in $\mathbb{M}_I$, and all $\mathcal{X}_n$ admit a mark function, this need not be the case for $\mathcal{X}$. In applications, however, the limit $\mathcal{X}$ is often not known explicitly, and it would be important to have (sufficient) criteria for the existence of a mark function in terms of the $\mathcal{X}_n$ alone. An easy possibility is Lipschitz equicontinuity: if all $\mathcal{X}_n$ admit a mark function that is Lipschitz continuous with a common Lipschitz constant $L > 0$, the same is true for $\mathcal{X}$ (see [Pio11]). More generally, this holds for uniformly equicontinuous mark functions as introduced below. We briefly discuss the equicontinuous case in this subsection, because it is straightforward and illustrates the main ideas.

Recall that a *modulus of continuity* is a function $h\colon \mathbb{R}_+ \to \mathbb{R}_+ \cup \{\infty\}$ that is continuous in 0 and satisfies $h(0) = 0$. A function $f\colon X \to I$, where (X, r) is a metric space, is *h-uniformly continuous* if $d\big(f(x), f(y)\big) \le h\big(r(x, y)\big)$ for all $x, y \in X$. Note that for every modulus of continuity h, there exists another modulus of continuity $h' \ge h$ which is increasing and continuous with respect to the topology of the one-point compactification of $\mathbb{R}_+$. Therefore, we can restrict ourselves without loss of generality to moduli of continuity from

$$\mathcal{H} := \big\{\, h\colon \mathbb{R}_+ \to \mathbb{R}_+ \cup \{\infty\} \;\big|\; h(0) = 0,\ h \text{ is continuous and increasing} \,\big\}. \tag{1.17}$$

For $h \in \mathcal{H}$ and a metric space (X, r), we define

$$A_h^X := A_h^{(X,r)} := \big\{(x_i, u_i)_{i=1,2} \in (X \times I)^2 : d(u_1, u_2) \le h(r(x_1, x_2))\big\} \subseteq (X \times I)^2. \tag{1.18}$$

Note that $f\colon X \to I$ is h-uniformly continuous if and only if $\big((x, f(x)), (y, f(y))\big) \in A_h^X$ for all $x, y \in X$, and that A_h^X is a closed set in $(X \times I)^2$ with product topology.

Definition 1.12 ($\mathfrak{M}_I^h$)**.** *For $h \in \mathcal{H}$, let $\mathfrak{M}_I^h \subseteq \mathbb{M}_I^{\mathrm{fct}}$ be the space of marked metric measure spaces admitting an h-uniformly continuous mark function.*

The next lemma states that a marked metric measure space (X, r, μ) admits an h-uniformly continuous mark function if and only if a pair of independent samples from μ is almost surely in A_h^X. Furthermore, if a sequence with h-uniformly continuous mark functions converges marked Gromov-weakly, the limit space also admits an h-uniformly continuous mark function.

Lemma 1.13 (uniform equicontinuity)**.** *Fix a modulus of continuity $h \in \mathcal{H}$.*

(i) $\mathfrak{M}_I^h = \{(X, r, \mu) \in \mathbb{M}_I : \mu^{\otimes 2}(A_h^X) = \|\mu^{\otimes 2}\|\}$.

(ii) $\mathfrak{M}_I^h$ *is closed in the marked Gromov-weak topology.*

Proof. The mmm-space $\mathcal{X} = (X, r, \mu)$ is in $\mathfrak{M}_I^h$ if and only if $\mathrm{supp}(\mu)$ is the graph of an h-uniformly continuous function. This is clearly equivalent to $\mu^{\otimes 2}\big((X \times I)^2 \setminus A_h^X\big) = 0$. Item (ii) is obvious from (i), because A_h^X is a closed set. $\square$

This preliminary result is quite restrictive because of the condition to have the same modulus of continuity for all occurring spaces. In fact, the mark function of the tree-valued Fleming Viot dynamic considered in Subsection 4.1 is not even continuous.

At the heart of the following generalisation to measurable mark functions lies the fact that measurable functions are "almost continuous" by Lusin's celebrated theorem (see for instance [Bog07, Theorem 7.1.13]). Here, we give a version tailored to our setup:

Lusin's theorem. *Let X, Y be Polish spaces, μ a finite measure on X, and $f\colon X \to Y$ a measurable function. Then, for every $\varepsilon > 0$, there exists a compact set $K_\varepsilon \subseteq X$ such that $\mu(X \setminus K_\varepsilon) < \varepsilon$ and $f|_{K_\varepsilon}$ is continuous.*

2. The space of fmm-spaces is Polish

The subspace $\mathbb{M}_I^{\mathrm{fct}}$ is not closed in $\mathbb{M}_I$ in the marked Gromov-weak topology, and hence the restriction of the marked Gromov-Prohorov metric d_{mGP} to $\mathbb{M}_I^{\mathrm{fct}}$ is not complete. In this section, we show that there exists another metric on $\mathbb{M}_I^{\mathrm{fct}}$ that induces the marked Gromov-weak topology and is complete. This shows that $\mathbb{M}_I^{\mathrm{fct}}$ is a Polish space in its own right.

2.1. A complete metric on the space of fmm-spaces.

For a measure ξ on I, we define

$$\beta_\xi := \int_I \int_I (1 \wedge d(u, v))\ \xi(\mathrm{d}u)\, \xi(\mathrm{d}v). \tag{2.1}$$

Note that $\beta_\xi = 0$ if and only if ξ is a Dirac measure. For $\mathcal{X} = (X, r, \mu) \in \mathbb{M}_I$, with $\mu = \nu \otimes K$ as in (1.15), we define

$$\beta(\mathcal{X}) := \int_X \beta_{K_x}\, \nu(\mathrm{d}x) = \int_{X \times I} \int_I (1 \wedge d(u, v))\ K_x(\mathrm{d}v)\, \mu\big(\mathrm{d}(x, u)\big). \tag{2.2}$$

Proposition 2.1 (characterization of $\mathbb{M}_I^{\mathrm{fct}}$ as continuity points)**.** *Let $\mathrm{cont}(\beta) \subseteq \mathbb{M}_I$ be the set of continuity points of $\beta\colon \mathbb{M}_I \to \mathbb{R}_+$, where $\mathbb{M}_I$ carries the marked Gromov-weak topology. Then*

$$\mathrm{cont}(\beta) = \beta^{-1}(0) = \mathbb{M}_I^{\mathrm{fct}}. \tag{2.3}$$

Proof (first part). As seen before, $\mathcal{X} = (X, r, \nu \otimes K) \in \mathbb{M}_I$ admits a mark function if and only if K_x is a Dirac measure for ν-almost every $x \in X$, which is the case if and only if $\beta(\mathcal{X}) = 0$. Hence $\beta^{-1}(0) = \mathbb{M}_I^{\mathrm{fct}}$. Because $\mathbb{M}_I^{\mathrm{fct}}$ is dense in $\mathbb{M}_I$ by Lemma 1.11, no $\mathcal{X} \in \mathbb{M}_I \setminus \beta^{-1}(0)$ can be a continuity point of β. Thus $\mathrm{cont}(\beta) \subseteq \beta^{-1}(0)$.

We defer the proof of the inclusion $\beta^{-1}(0) \subseteq \mathrm{cont}(\beta)$ to Subsection 2.2, because it requires a technical estimate on β derived in Proposition 2.7. $\square$

In view of (2.3), we can use standard arguments to construct a complete metric on $\mathbb{M}_I^{\mathrm{fct}}$ that metrizes marked Gromov-weak topology. Namely consider the sets

$$F_m := \overline{\beta^{-1}\big([\tfrac{1}{m}, \infty)\big)} \subseteq \mathbb{M}_I, \quad m \in \mathbb{N}, \tag{2.4}$$

where the closure is in the marked Gromov-weak topology. Then, due to Proposition 2.1, F_m is disjoint from $\mathbb{M}_I^{\mathrm{fct}}$, and $\mathbb{M}_I^{\mathrm{fct}} = \mathbb{M}_I \setminus \bigcup_{m\in\mathbb{N}} F_m$. Because F_m is also closed by definition, we obtain

$$\mathbb{M}_I^{\mathrm{fct}} = \bigcap_{m\in\mathbb{N}} \bigl\{\mathcal{X} \in \mathbb{M}_I : d_{\mathrm{mGP}}(\mathcal{X}, F_m) > 0\bigr\}. \tag{2.5}$$

We consider the metric d_{fGP} on $\mathbb{M}_I^{\mathrm{fct}}$ defined for $\mathcal{X}, \mathcal{Y} \in \mathbb{M}_I^{\mathrm{fct}}$ by

$$d_{\mathrm{fGP}}(\mathcal{X}, \mathcal{Y}) := d_{\mathrm{mGP}}(\mathcal{X}, \mathcal{Y}) + \sup_{m\in\mathbb{N}} 2^{-m} \wedge \Bigl|\frac{1}{d_{\mathrm{mGP}}(\mathcal{X}, F_m)} - \frac{1}{d_{\mathrm{mGP}}(\mathcal{Y}, F_m)}\Bigr|. \tag{2.6}$$

Theorem 2.2 ($\mathbb{M}_I^{\mathrm{fct}}$ is Polish)**.** *The space $\mathbb{M}_I^{\mathrm{fct}}$ of I-functionally-marked metric measure spaces with marked Gromov-weak topology is a Polish space. Namely, d_{fGP} is a complete metric on $\mathbb{M}_I^{\mathrm{fct}}$ inducing the marked Gromov-weak topology.*

Proof. First, we show that d_{fGP} induces the marked Gromov-weak topology on $\mathbb{M}_I^{\mathrm{fct}}$. For $m \in \mathbb{N}$, $\mathcal{X} \in \mathbb{M}_I$, define

$$\rho_m(\mathcal{X}) := d_{\mathrm{mGP}}(\mathcal{X}, F_m), \tag{2.7}$$

with F_m defined in (2.4). Note that ρ_m is a continuous function on $\mathbb{M}_I$. Let $\mathcal{X}_n, \mathcal{X} \in \mathbb{M}_I^{\mathrm{fct}}$. Then $\rho_m(\mathcal{X}) > 0$ for all $m \in \mathbb{N}$ because of (2.5). Therefore, by definition, $d_{\mathrm{fGP}}(\mathcal{X}_n, \mathcal{X}) \underset{n\to\infty}{\longrightarrow} 0$ if and only if the two conditions $d_{\mathrm{mGP}}(\mathcal{X}_n, \mathcal{X}) \underset{n\to\infty}{\longrightarrow} 0$ and

$$\rho_m(\mathcal{X}_n) \underset{n\to\infty}{\longrightarrow} \rho_m(\mathcal{X}) \qquad \forall m \in \mathbb{N} \tag{2.8}$$

hold. We have to show that the marked Gromov-weak convergence already implies (2.8). This, however, follows from the continuity of the ρ_m.

It remains to show that d_{fGP} is a complete metric on $\mathbb{M}_I^{\mathrm{fct}}$. Consider a d_{fGP}-Cauchy sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ in $\mathbb{M}_I^{\mathrm{fct}}$. By completeness of d_{mGP} on $\mathbb{M}_I$, it converges marked Gromov-weakly to some $\mathcal{X} = (X, r, \mu) \in \mathbb{M}_I$. Furthermore, for every fixed $m \in \mathbb{N}$, (2.6) implies that $1/\rho_m(\mathcal{X}_n)$ converges as $n \to \infty$, and hence $d_{\mathrm{mGP}}(\mathcal{X}_n, F_m)$ is bounded away from zero. Thus $\mathcal{X} \notin F_m$. Because $\mathbb{M}_I^{\mathrm{fct}} = \mathbb{M}_I \setminus \bigcup_{m\in\mathbb{N}} F_m$, this means that $\mathcal{X} \in \mathbb{M}_I^{\mathrm{fct}}$, and by the first part of the proof $d_{\mathrm{fGP}}(\mathcal{X}_n, \mathcal{X}) \underset{n\to\infty}{\longrightarrow} 0$. ☐

With $B_\delta^{\mathbb{M}_I}(\mathcal{X}) := \bigl\{\mathcal{Y} \in \mathbb{M}_I : d_{\mathrm{mGP}}(\mathcal{X}, \mathcal{Y}) < \delta\bigr\}$ we denote the open δ-ball in $\mathbb{M}_I$ with respect to d_{mGP}. The following corollary gives formal criteria for a limiting space to admit a mark function, which are useful only together with estimates on β.

Corollary 2.3. *Let $(\mathcal{X}_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{M}_I$ which converges marked Gromov-weakly to $\mathcal{X}$. Then the following four conditions are equivalent:*

(i) $\mathcal{X} \in \mathbb{M}_I^{\mathrm{fct}}$.

(ii) $\limsup_{n\to\infty} \rho_m(\mathcal{X}_n) > 0$ *for all* $m \in \mathbb{N}$, *with* ρ_m *defined in* (2.7).

(iii) *For every* $\delta > 0$,

$$\limsup_{n\to\infty} \inf_{\mathcal{Y}\in\beta^{-1}([\delta,\infty[)} d_{\mathrm{mGP}}(\mathcal{X}_n, \mathcal{Y}) > 0. \tag{2.9}$$

(iv)

$$\lim_{\delta\downarrow 0} \liminf_{n\to\infty} \sup_{\mathcal{Y}\in B_\delta^{\mathbb{M}_I}(\mathcal{X}_n)} \beta(\mathcal{Y}) = 0. \tag{2.10}$$

Proof. *"(i) ⇔ (ii)":* We have $\rho_m(\mathcal{X}) = \lim_{n\to\infty} \rho_m(\mathcal{X}_n)$, and $\rho_m(\mathcal{X}) > 0$ for all $m \in \mathbb{N}$ if and only if $\mathcal{X} \in \mathbb{M}_I^{\mathrm{fct}}$.

"(ii) ⇔ (iii)": follows directly from the definition of ρ_m.

"(iii) $\Leftrightarrow$ (iv)": Using monotonicity in δ we obtain

(2.11)

$$\begin{aligned}\text{(iii)} &\iff \forall\delta>0\,\exists\varepsilon>0\,\forall(\mathcal{Y}_n)_{n\in\mathbb{N}}\subseteq\mathbb{M}_I \text{ with } \beta(\mathcal{Y}_n)\geq\delta : \limsup_{n\to\infty} d_{\mathrm{mGP}}(\mathcal{X}_n,\mathcal{Y}_n)\geq\varepsilon\\ &\iff \forall\delta>0\,\exists\varepsilon>0\,\forall(\mathcal{Y}_n)_{n\in\mathbb{N}}\subseteq\mathbb{M}_I : \liminf_{n\to\infty}\beta(\mathcal{Y}_n)<\delta \text{ or } \limsup_{n\to\infty} d_{\mathrm{mGP}}(\mathcal{X}_n,\mathcal{Y}_n)\geq\varepsilon\\ &\iff \forall\varepsilon>0\,\exists\delta>0\,\forall(\mathcal{Y}_n)_{n\in\mathbb{N}}\subseteq\mathbb{M}_I \text{ with } \mathcal{Y}_n\in B_\delta^{\mathbb{M}_I}(\mathcal{X}_n) : \liminf_{n\to\infty}\beta(\mathcal{Y}_n)<\varepsilon \iff \text{(iv)},\end{aligned}$$

where, in the third equivalence, we renamed δ to ε and ε to δ. □

2.2. **A decomposition of $\mathbb{M}_I^{\mathrm{fct}}$ into closed sets and estimates on β.** In this subsection, we derive some estimates on β and use them to complete the proof of Proposition 2.1. Furthermore, we construct a decomposition of $\mathbb{M}_I^{\mathrm{fct}}$ into closed sets which are related to the sets $\mathfrak{M}_I^h$.

As we have seen in Subsection 1.4, the situation becomes easy if we restrict to the uniformly equicontinuous case, that is to the subspace $\mathfrak{M}_I^h$ for some $h\in\mathcal{H}$ as in Definition 1.12. We introduce in what follows several related subspaces capturing some aspect of equicontinuity. In analogy to the definition of A_h^X in (1.18), we use for a metric space (X,r), and $\delta,\varepsilon>0$, the notation

$$A_{\delta,\varepsilon}^X := A_{\delta,\varepsilon}^{(X,r)} := \big\{(x_i,u_i)_{i=1,2}\in(X\times I)^2 : r(x_1,x_2)\geq\delta \text{ or } d(u_1,u_2)\leq\varepsilon\big\}\subseteq(X\times I)^2. \tag{2.12}$$

Note that $A_{\delta,\varepsilon}^X$ is a closed set. For every $h\in\mathcal{H}$, using monotonicity and continuity of h, we observe that

$$A_h^X = \bigcap_{\delta>0} A_{\delta,h(\delta)}^X. \tag{2.13}$$

Definition 2.4 ($\mathfrak{M}_I^{\delta,\varepsilon}$, $\mathbb{M}_I^{\delta,\varepsilon}$, $\mathbb{M}_I^h$)**.** *Let $\delta,\varepsilon>0$ and $h\in\mathcal{H}$. We define*

$$\mathfrak{M}_I^{\delta,\varepsilon} := \big\{(X,r,\mu)\in\mathbb{M}_I : \mu^{\otimes2}(A_{\delta,\varepsilon}^X)=\|\mu^{\otimes2}\|\big\}, \tag{2.14}$$

$$\mathbb{M}_I^{\delta,\varepsilon} := \big\{(X,r,\mu)\in\mathbb{M}_I : \exists\mu'\in\mathcal{M}_{\mathrm{f}}(X\times I) : \mu'\leq\mu,\ \|\mu-\mu'\|\leq\varepsilon,\ (X,r,\mu')\in\mathfrak{M}_I^{\delta,\varepsilon}\big\}, \tag{2.15}$$

and $\mathbb{M}_I^h := \bigcap_{\delta>0}\mathbb{M}_I^{\delta,h(\delta)}$.

The intuition is that for spaces in $\mathfrak{M}_I^{\delta,h(\delta)}$, the measure behaves as if it admitted an h-uniformly continuous mark function when distances of order δ are observed. The same holds for the spaces in $\mathbb{M}_I^{\delta,h(\delta)}$ if we are additionally allowed to neglect a portion $h(\delta)$ of mass.

Remark 2.5.
(i) Clearly $\mathfrak{M}_I^h\subseteq\mathbb{M}_I^h$. We will see in Lemma 2.8 that $\mathbb{M}_I^h\subseteq\mathbb{M}_I^{\mathrm{fct}}$.
(ii) The space $\mathbb{M}_I^h$ is *much* larger than $\mathfrak{M}_I^h$: while $\bigcup_{h\in\mathcal{H}}\mathfrak{M}_I^h$ contains only mmm-spaces admitting a uniformly *continuous* mark function, we will see in Lemma 2.8 that every element of $\mathbb{M}_I^{\mathrm{fct}}$ is in some $\mathbb{M}_I^h$.
(iii) The spaces $\mathfrak{M}_I^{\delta,\varepsilon}$ and $\mathbb{M}_I^{\delta,\varepsilon}$ are not contained in $\mathbb{M}_I^{\mathrm{fct}}$. For instance, consider $I=\mathbb{R}$ and $\mathcal{X}=\big(\{0\},0,\delta_{(0,0)}+\delta_{(0,\varepsilon)}\big)\in\mathfrak{M}_I^{\delta,\varepsilon}\subseteq\mathbb{M}_I^{\delta,\varepsilon}$.

We have the following stability of $\mathbb{M}_I^{\delta,\varepsilon}$ with respect to small perturbations in the marked Gromov-Prohorov metric.

Lemma 2.6 (perturbation of $\mathbb{M}_I^{\delta,\varepsilon}$)**.** *Let $\delta,\varepsilon>0$, $\mathcal{X}\in\mathbb{M}_I^{\delta,\varepsilon}$ and $\hat{\mathcal{X}}\in\mathbb{M}_I$. Then*

$$\delta' := d_{\mathrm{mGP}}(\mathcal{X},\hat{\mathcal{X}}) < \tfrac12\delta \implies \hat{\mathcal{X}}\in\mathbb{M}_I^{\delta-2\delta',\,\varepsilon+2\delta'}. \tag{2.16}$$

Proof. Let $\mathcal{X}=(X,r,\mu)$, $\hat{\mathcal{X}}=(\hat X,\hat r,\hat\mu)$. We may assume that $X,\hat X$ are subspaces of some separable, metric space (E,r_E) such that $d_{\mathrm{Pr}}(\mu,\hat\mu)<\delta'$. By definition of $\mathbb{M}_I^{\delta,\varepsilon}$, there is $\mu'\leq\mu$ with $\|\mu-\mu'\|\leq\varepsilon$ and $\mathcal{X}':=(X,r,\mu')\in\mathfrak{M}_I^{\delta,\varepsilon}$. Due to Lemma 1.4, we find $\hat\mu'\leq\hat\mu$ with $\|\hat\mu-\hat\mu'\|\leq\varepsilon$ and $d_{\mathrm{Pr}}(\mu',\hat\mu')<\delta'$, where $\hat{\mathcal{X}}'=(\hat X,\hat r,\hat\mu')$. By the coupling representation of the Prohorov metric, (1.4), we obtain a measure ξ on $(E\times I)^2$ with marginals $\xi_1 \le \mu'$ and $\xi_2 \le \hat\mu'$ such that $\|\hat\mu' - \xi_2\| \le \delta'$ and

$$\xi\big(\big\{\big((x,u),(\hat x,\hat u)\big) \in (X\times I)\times(\hat X\times I) : r_E(x,\hat x) + d(u,\hat u) \ge \delta'\big\}\big) = 0. \tag{2.17}$$

By definition, $(\mu')^{\otimes 2}$ is supported by $A^X_{\delta,\varepsilon}$. Therefore, the same is true for $\xi_1^{\otimes 2}$ and we obtain

$$\begin{aligned} \|\xi_2^{\otimes 2}\| = \|\xi^{\otimes 2}\| &= \xi^{\otimes 2}\big(\big\{(x_i,u_i,\hat x_i,\hat u_i)_{i=1,2} \in ((X\times I)\times(\hat X\times I))^2 : (x_i,u_i)_{i=1,2} \in A^X_{\delta,\varepsilon}\big\}\big) \\ &\le \xi_2^{\otimes 2}\big(\big\{(\hat x_i,\hat u_i)_{i=1,2} \in (\hat X\times I)^2 : r_E(\hat x_1,\hat x_2) \ge \delta - 2\delta' \text{ or } d(\hat u_1,\hat u_2) \le \varepsilon + 2\delta'\big\}\big) \\ &= \xi_2^{\otimes 2}(A^{\hat X}_{\delta-2\delta',\varepsilon+2\delta'}), \end{aligned} \tag{2.18}$$

where the inequality follows from (2.17) together with the triangle-inequality. Therefore, we have $(\hat X,\hat r,\xi_2) \in \mathfrak{M}_I^{\delta-2\delta',\,\varepsilon+2\delta'}$. Now the claim follows from $\|\hat\mu - \xi_2\| \le \|\hat\mu - \hat\mu'\| + \|\hat\mu' - \xi_2\| \le \varepsilon + \delta'$. $\square$

Proposition 2.7 (estimates on β)**.** *Let $\delta,\varepsilon > 0$ and consider $\mathcal{X} = (X,r,\mu) \in \mathbb{M}_I$. Then the following hold:*

(i) *If $\mu' \in \mathcal{M}_\mathrm{f}(X\times I)$, then $\beta(\mathcal{X}) \le \beta\big((X,r,\mu')\big) + 2\|\mu - \mu'\|$.*
(ii) *If $\mathcal{X} \in \mathfrak{M}_I^{2\delta,\varepsilon}$, then $\beta(\mathcal{X}) \le \varepsilon\|\mu\|$.*
(iii) *If $\mathcal{X} \in \mathbb{M}_I^{2\delta,\varepsilon}$ and $\hat{\mathcal{X}} \in \mathbb{M}_I$ with $d_\mathrm{mGP}(\mathcal{X},\hat{\mathcal{X}}) < \delta$, then $\beta(\mathcal{X}) \le \varepsilon\big(\|\mu\| + 2\big)$ and*

$$\beta(\hat{\mathcal{X}}) \le (\varepsilon + 2\delta)(2 + \|\mu\| + \delta). \tag{2.19}$$

Proof. (i) follows directly from the definition.
(ii) If $x \in X$ and $u,v \in I$ satisfy $\big((x,u),(x,v)\big) \in A^X_{2\delta,\varepsilon}$, then $d(u,v) \le \varepsilon$ by definition of $A^X_{2\delta,\varepsilon}$. Thus $\beta(\mathcal{X}) = \int_{X\times I}\int_I (1\wedge d(u,v))\, K_x(\mathrm{d}v)\,\mu\big(\mathrm{d}(x,u)\big) \le \varepsilon\|\mu\|$.
(iii) Combining (i) and (ii) yields $\beta(\mathcal{X}) \le 2\varepsilon + \varepsilon\|\mu\|$. Let $\delta' = d_\mathrm{mGP}(\mathcal{X},\hat{\mathcal{X}})$. By Lemma 2.6, we have $\hat{\mathcal{X}} \in \mathbb{M}_I^{2\delta-2\delta',\,\varepsilon+2\delta'}$ and thus $\beta(\hat{\mathcal{X}}) \le (2 + \|\hat\mu\|)(\varepsilon + 2\delta') \le (2 + \|\mu\| + \delta)(\varepsilon + 2\delta)$. $\square$

In order to complete the proof of Proposition 2.1 with the help of Proposition 2.7, we first observe that, as a consequence of Lusin's theorem, every functionally marked metric measure space is an element of $\mathbb{M}_I^h$ for some $h \in \mathcal{H}$. Together with Lemma 2.9 below, this means that we have a nice (though uncountable) decomposition of $\mathbb{M}_I^\mathrm{fct}$ into closed sets.

Lemma 2.8 (decomposition of $\mathbb{M}_I^\mathrm{fct}$)**.** *The following equality holds: $\mathbb{M}_I^\mathrm{fct} = \bigcup_{h\in\mathcal{H}} \mathbb{M}_I^h$.*

Proof. We have $\mathbb{M}_I^h \subseteq \beta^{-1}(0) = \mathbb{M}_I^\mathrm{fct}$ for every $h \in \mathcal{H}$. Indeed, the equality was shown in the first part of the proof of Proposition 2.1. To obtain the inclusion, that is $\beta(\mathcal{X}) = 0$ for all $\mathcal{X} \in \mathbb{M}_I^h$, recall $\mathbb{M}_I^h$ from Definition 2.4 and choose $\varepsilon = h(2\delta)$ in Proposition 2.7(iii).

Conversely, let $\mathcal{X} = (X,r,\nu,\kappa) \in \mathbb{M}_I^\mathrm{fct}$. According to Lusin's theorem, we find for every $\varepsilon > 0$ a compact set $K_\varepsilon \subseteq X$, and a modulus of continuity $h_\varepsilon \in \mathcal{H}$, such that $\nu(X\setminus K_\varepsilon) \le \varepsilon$ and $\kappa|_{K_\varepsilon}$ is h_ε-uniformly continuous. In particular,

$$\mathcal{X} \in \mathbb{M}_I^{\delta,\,h_\varepsilon(\delta)\vee\varepsilon} \quad \forall \varepsilon,\delta > 0. \tag{2.20}$$

We may assume without loss of generality that $\varepsilon \mapsto h_\varepsilon(\delta)$ is decreasing and right-continuous for every $\delta > 0$. We define

$$h(\delta) := \inf\big\{\varepsilon > 0 : h_\varepsilon(\delta) < \varepsilon\big\} \in \mathbb{R}_+ \cup \{\infty\}. \tag{2.21}$$

Clearly, $h(\delta)$ converges to 0 as $\delta \downarrow 0$ because $h_\varepsilon \in \mathcal{H}$. Furthermore, $h_{h(\delta)}(\delta) \le h(\delta)$, and hence (2.20) with $\varepsilon = h(\delta)$ implies $\mathcal{X} \in \mathbb{M}_I^h$. $\square$

Proof of Proposition 2.1 *(completion).* We still have to show continuity of β in $\mathcal{X} \in \beta^{-1}(0)$. Due to Lemma 2.8, there is $h \in \mathcal{H}$ with $\mathcal{X} \in \mathbb{M}_I^h$. Now Proposition 2.7 yields for $\delta > 0$ the estimate $\sup_{\hat{\mathcal{X}} \in B_\delta^{\mathbb{M}_I}(\mathcal{X})} \beta(\hat{\mathcal{X}}) \le (h(2\delta) + 2\delta)(2 + \|\mu\| + \delta)$, which converges to 0 as $\delta \downarrow 0$. $\square$

It directly follows from Proposition 2.7(iii) that the marked Gromov-weak closure of $\mathbb{M}_I^h$ is contained in $\mathbb{M}_I^{\mathrm{fct}}$. In fact, $\mathbb{M}_I^h$ is even Gromov-weakly closed, which will be used in the proof of Theorem 3.11 below.

Lemma 2.9 (closedness of $\mathbb{M}_I^h$)**.** *For every $\delta, \varepsilon > 0$, $\mathbb{M}_I^{\delta,\varepsilon}$ is marked Gromov-weakly closed in $\mathbb{M}_I$. In particular, $\mathbb{M}_I^h$ is closed for every $h \in \mathcal{H}$.*

Proof. Fix $\varepsilon, \delta > 0$ and let $(\mathcal{X}_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{M}_I^{\delta,\varepsilon}$ converging marked Gromov-weakly to some $\mathcal{X} = (X, r, \mu) \in \mathbb{M}_I$. Using Lemma 1.9, we may assume that X_n, $n \in \mathbb{N}$, and X are subspaces of a common separable, metric space (E, r_E), such that $\mu_n \xrightarrow{w} \mu$ on $E \times I$. By definition of $\mathbb{M}_I^{\delta,\varepsilon}$, we find $\mu_n' \le \mu_n$, $\|\mu_n' - \mu_n\| \le \varepsilon$, such that $(\mu_n')^{\otimes 2}$ is supported by $A_{\delta,\varepsilon}^E$ for all $n \in \mathbb{N}$. Since $(\mu_n')_{n\in\mathbb{N}}$ is tight, we may assume, by passing to a subsequence, that $\mu_n' \xrightarrow{w} \mu'$ for some $\mu' \in \mathcal{M}_{\mathrm{f}}(E)$. Obviously, $\mu' \le \mu$ and $\|\mu - \mu'\| = \lim_{n\to\infty} \|\mu_n\| - \|\mu_n'\| \le \varepsilon$. Because $A_{\delta,\varepsilon}^E$ is closed, $(\mu')^{\otimes 2}$ is supported by $A_{\delta,\varepsilon}^E$ and hence $\mathcal{X} \in \mathbb{M}_I^{\delta,\varepsilon}$. $\square$

3. Criteria for the existence of mark functions

Based on the construction of the complete metric and the decomposition $\mathbb{M}_I^{\mathrm{fct}} = \bigcup_{h\in\mathcal{H}} \mathbb{M}_I^h$ into closed sets obtained in Section 2, we now derive criteria to check if a marked metric measure space admits a mark function, especially in the case where it is given as a marked Gromov-weak limit. We then transfer the results to random mmm-spaces and $\mathbb{M}_I$-valued stochastic processes.

3.1. **Deterministic criteria.** Our main criterion for deterministic spaces is a direct consequence of the results in Section 2. Recall that $\mathcal{H}$ is the set of moduli of continuity defined in (1.17).

Theorem 3.1 (characterization of existence of a mark function in the limit)**.** *Let $(\mathcal{X}_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{M}_I$ with $\mathcal{X}_n \xrightarrow{\mathrm{mGw}} \mathcal{X} \in \mathbb{M}_I$. Then $\mathcal{X} \in \mathbb{M}_I^{\mathrm{fct}}$ if and only if there exists $h \in \mathcal{H}$ such that for every $\delta > 0$*

$$\mathcal{X}_n \in \mathbb{M}_I^{\delta,h(\delta)} \quad \textit{for infinitely many } n \in \mathbb{N}. \tag{3.1}$$

In this case, $\mathcal{X} \in \mathbb{M}_I^h$.

Proof. First assume there is $h \in \mathcal{H}$ such that (3.1) is satisfied. Since $\mathbb{M}_I^{\delta,h(\delta)}$ is closed by Lemma 2.9, (3.1) implies that $\mathcal{X} \in \mathbb{M}_I^{\delta,h(\delta)}$ for every δ, that is $\mathcal{X} \in \mathbb{M}_I^h$. By Lemma 2.8, $\mathbb{M}_I^h \subseteq \mathbb{M}_I^{\mathrm{fct}}$.

Conversely, assume $\mathcal{X} \in \mathbb{M}_I^{\mathrm{fct}}$. Then, by Lemma 2.8, we find $h \in \mathcal{H}$ with $\mathcal{X} \in \mathbb{M}_I^h$. We claim that (3.1) holds with h replaced by $\hat{h}(\delta) := h(3\delta) + 2\delta$. Indeed, fix $\delta > 0$ and observe that $\mathcal{X} \in \mathbb{M}_I^h \subseteq \mathbb{M}_I^{3\delta,h(3\delta)}$. Lemma 2.6 yields $\mathcal{X}_n \in \mathbb{M}_I^{\delta,\hat{h}(\delta)}$ for all n with $d_{\mathrm{mGP}}(\mathcal{X}, \mathcal{X}_n) < \delta$. $\square$

We will use Theorem 3.1 in the following form.

Corollary 3.2. *Let $\mathcal{X}_n = (X_n, r_n, \nu_n, \kappa_n) \in \mathbb{M}_I^{\mathrm{fct}}$, $\mathcal{X}_n \xrightarrow{\mathrm{mGw}} \mathcal{X} \in \mathbb{M}_I$. Let $Y_{n,\delta} \subseteq X_n$ measurable for $n \in \mathbb{N}$, $\delta > 0$, and $h \in \mathcal{H}$. Then $\mathcal{X} \in \mathbb{M}_I^{\mathrm{fct}}$ if the following two conditions hold for every $\delta > 0$:*

$$\liminf_{n\to\infty} \nu_n(X_n \setminus Y_{n,\delta}) \le h(\delta), \tag{3.2}$$

$$\forall n \in \mathbb{N},\ x, y \in Y_{n,\delta} : r_n(x, y) < \delta \implies d\big(\kappa_n(x), \kappa_n(y)\big) \le h(\delta). \tag{3.3}$$

Proof. Let $\mu_n' := \mu_n|_{Y_{n,\delta}\times I}$, where $\mu_n = \nu_n \otimes \delta_{\kappa_n}$. Then (3.3) implies $(X_n, r_n, \mu_n') \in \mathfrak{M}_I^{\delta,h(\delta)}$ and (3.2) yields $\|\mu_n' - \mu_n\| \le h(\delta)$ for infinitely many n. Hence we can apply Theorem 3.1. $\square$

Remark 3.3. To obtain $\mathcal{X} \in \mathbb{M}_I^{\mathrm{fct}}$, it is clearly enough to show in Theorem 3.1 and Corollary 3.2, (3.1) respectively (3.2) and (3.3) only for $\delta = \delta_m$ for a sequence $(\delta_m)_{m\in\mathbb{N}}$ with $\delta_m \downarrow 0$ as $m \to \infty$.

We illustrate the rôle of the exceptional set $X_n \setminus Y_{n,\delta}$, and the importance of its dependence on δ, with a simple example.

Example 3.4. Consider $X = [0,1]$ with Euclidean metric r, $\nu = \lambda + \delta_0$, where λ is Lebesgue-measure, and $\kappa_n(x) = (nx) \wedge 1$. Obviously, $\mathcal{x}_n = (X, r, \nu, \kappa_n)$ converges marked Gromov-weakly and the limit admits the mark function $\mathbb{1}_{(0,1]}$. To see this from Corollary 3.2, we choose $h(\delta) = \delta$ and $Y_{n,\delta} = \{0\} \cup [\delta \vee \frac{1}{n}, 1]$. Note that we cannot choose $Y_{n,\delta}$ independent of δ.

Remark 3.5 (equicontinuous case)**.** If, in Corollary 3.2, $Y_{n,\delta} = Y_n$ does not depend on δ, then (3.3) means that κ_n is h-uniformly continuous on Y_n. Consequently, the mark function of $\mathcal{x}$ is in this case h-uniformly continuous. If we restrict to $Y_n = X_n$ for all n, we recover part (ii) of Lemma 1.13.

Corollary 3.6. *Let $\mathcal{x}_n = (X_n, r_n, \nu_n, \kappa_n) \in \mathbb{M}_I^{\mathrm{fct}}$ and assume that $\mathcal{x}_n$ converges to $\mathcal{x} = (X, r, \mu) \in \mathbb{M}_I$ marked Gromov-weakly. Further assume that for $n \in \mathbb{N}$, $\delta > 0$, there are measurable sets $Z_{n,\delta} \subseteq X_n$, such that*

$$\lim_{\delta\downarrow 0} \liminf_{n\to\infty} \left(\nu_n(X_n \setminus Z_{n,\delta}) + \int_{Z_{n,\delta}} \Big(1 \wedge \mathrm{diam}\big(\kappa_n\big(B_\delta^{X_n}(x) \cap Z_{n,\delta}\big)\big)\Big)\ \nu_n(\mathrm{d}x) \right) = 0, \tag{3.4}$$

where diam *is the diameter of a set. Then $\mathcal{x}$ admits a mark function, that is $\mathcal{x} \in \mathbb{M}_I^{\mathrm{fct}}$.*

Proof. For $\delta > 0$ let

$$g(\delta) := \sup_{0<\delta'\le\delta} \liminf_{n\to\infty} \left(\nu_n(X_n \setminus Z_{n,\delta'}) + \int_{Z_{n,\delta'}} \Big(1 \wedge \mathrm{diam}\big(\kappa_n\big(B_{\delta'}^{X_n}(x) \cap Z_{n,\delta'}\big)\big)\Big)\ \nu_n(\mathrm{d}x) \right). \tag{3.5}$$

By (3.4), $\lim_{\delta\downarrow 0} g(\delta) = 0$ and g is increasing with $\|g\|_\infty \le \|\mu\|$. Let $h \in \mathcal{H}$ be such that $g(\delta) \le \frac{h(\delta)}{2}(1 \wedge h(\delta))$ for all $\delta > 0$. Then

$$\nu_n\big(\big\{x \in Z_{n,\delta} : \mathrm{diam}\big(\kappa_n(B_\delta^{X_n}(x) \cap Z_{n,\delta})\big) > h(\delta)\big\}\big) \le \frac{g(\delta)}{1 \wedge h(\delta)} \le h(\delta)/2. \tag{3.6}$$

Now apply Corollary 3.2 with

$$Y_{n,\delta} := \big\{x \in Z_{n,\delta} : \mathrm{diam}\big(\kappa_n\big(B_\delta^{X_n}(x) \cap Z_{n,\delta}\big)\big) \le h(\delta)\big\}. \tag{3.7}$$

Then (3.3) follows from the definition of $Y_{n,\delta}$ in (3.7), and $\nu_n(X_n \setminus Y_{n,\delta}) \le \nu_n(X_n \setminus Z_{n,\delta}) + h(\delta)/2 \le g(\delta) + h(\delta)/2 \le h(\delta)$ holds by (3.6) and (3.5). ☐

3.2. **Random fmm-spaces.** The following theorem is a randomized version of Theorem 3.1. It is our main criterion for $\mathbb{M}_I$-valued random variables.

Theorem 3.7 (random fmm-spaces as limits in distribution)**.** *Let $(\mathcal{X}_n)_{n\in\mathbb{N}}$ be a sequence of $\mathbb{M}_I$-valued random variables which converges in distribution (w.r.t. marked Gromov-weak topology) to an $\mathbb{M}_I$-valued random variable $\mathcal{X}$. Further assume that for every $\varepsilon > 0$, there exists a modulus of continuity $h_\varepsilon \in \mathcal{H}$ such that*

$$\limsup_{\delta\downarrow 0} \limsup_{n\to\infty} \mathbb{P}\big(\big\{\ \mathcal{X}_n \in \mathbb{M}_I^{\delta, h_\varepsilon(\delta)}\ \big\}\big) \ge 1 - \varepsilon. \tag{3.8}$$

Then $\mathcal{X}$ admits almost surely a mark function, that is $\mathcal{X} \in \mathbb{M}_I^{\mathrm{fct}}$ almost surely.

If additionally $\mathcal{X}_n = (X_n, r_n, \nu_n, \kappa_n) \in \mathbb{M}_I^{\mathrm{fct}}$ almost surely for all $n \in \mathbb{N}$, we can replace (3.8) *by existence of random measurable sets $Y_{n,\delta}^\varepsilon \subseteq X_n$, $n \in \mathbb{N}$, $\delta > 0$, in addition to the $h_\varepsilon \in \mathcal{H}$, such that the following two conditions hold for every $\varepsilon > 0$:*

$$\limsup_{\delta\downarrow 0} \limsup_{n\to\infty} \mathbb{P}\big(\big\{\ \nu_n(X_n \setminus Y_{n,\delta}^\varepsilon) \le h_\varepsilon(\delta)\ \big\}\big) \ge 1 - \varepsilon. \tag{3.9}$$

$$\forall n \in \mathbb{N},\ \delta > 0,\ x, y \in Y_{n,\delta}^\varepsilon : r_n(x,y) < \delta \implies d\big(\kappa_n(x), \kappa_n(y)\big) \le h_\varepsilon(\delta). \tag{3.10}$$

Remark 3.8. In (3.9), we need not worry about measurability of the "event" $B_{n,\delta} := \{\nu_n(X_n \setminus Y^{\varepsilon}_{n,\delta}) \le h_\varepsilon(\delta)\}$ due to the choice of $Y^{\varepsilon}_{n,\delta}$. The inequality (3.9) is to be understood in the sense of inner measure, that is we require that there are measurable sets $C_{n,\delta} \subseteq B_{n,\delta}$ with $\limsup_{\delta\downarrow 0}\limsup_{n\to\infty}\mathbb{P}(C_{n,\delta}) \ge 1-\varepsilon$.

Proof. The second statement follows in the same way as Corollary 3.2. We divide the proof of the main part in two steps. First, we show $\mathcal{X} \in \mathbb{M}_I^{\mathrm{fct}}$ if, instead of (3.8), even

$$\mathbb{P}\Big(\bigcap_{m\in\mathbb{N}}\big\{\mathcal{X}_n \in \mathbb{M}_I^{\delta_m, h_\varepsilon(\delta_m)} \text{ for infinitely many } n\big\}\Big) \ge 1-\varepsilon \tag{3.11}$$

holds for a sequence $\delta_m = \delta_m(\varepsilon) \downarrow 0$ as $m \to \infty$. In the second step, we show that, given (3.8), we can modify h_ε to $\hat{h}_\varepsilon \in \mathcal{H}$ such that (3.11) holds with h_ε replaced by $\hat{h}_\varepsilon$.

Step 1. By Skorohod's representation theorem, we may assume that the $\mathcal{X}_n$ are coupled such that they converge almost surely to $\mathcal{X}$ in the marked Gromov-weak topology. The inequality (3.11) implies that with probability at least $1-\varepsilon$, for all $m \in \mathbb{N}$, $\mathcal{X}_n \in \mathbb{M}_I^{\delta_m, h_\varepsilon(\delta_m)}$ for infinitely many n. By Theorem 3.1 and Remark 3.3, this means that the probability that $\mathcal{X}$ admits a mark function is at least $1-\varepsilon$. Because ε is arbitrary, this implies $\mathcal{X} \in \mathbb{M}_I^{\mathrm{fct}}$ almost surely.

Step 2. Let $T(\varepsilon,\delta) := \limsup_{n\to\infty}\mathbb{P}\big(\{\mathcal{X}_n \in \mathbb{M}_I^{\delta, h_\varepsilon(\delta)}\}\big)$ in (3.8). Set

$$\delta_1 := \sup\big\{\delta \in [0,1] : T(\varepsilon/4,\delta) \ge 1-\varepsilon/2 \text{ and } h_{\varepsilon/4}(\delta) < 1\big\}. \tag{3.12}$$

By (3.8) and as $h_{\varepsilon/4} \in \mathcal{H}$, the set inside the supremum is non-empty. Next define recursively

$$\delta_m := \sup\big\{\delta \in [0,\delta_{m-1}/2] : T(\varepsilon 2^{-(m+1)},\delta) \ge 1-\varepsilon 2^{-m} \text{ and } h_{\varepsilon 2^{-(m+1)}}(\delta) < 1/m\big\} \tag{3.13}$$

for $m \in \mathbb{N}, m \ge 2$. Again, the set inside the supremum is non-empty by (3.8) and as $h_{\varepsilon 2^{-(m+1)}} \in \mathcal{H}$. Moreover, $\delta_m = \delta_m(\varepsilon) > 0$, $\delta_m \downarrow 0$ for $m \to \infty$ and $h_{\varepsilon 2^{-(m+1)}}(\delta_m) \le 1/m$ follows. We can therefore set

$$\hat{h}_\varepsilon(\delta_m) := h_{\varepsilon 2^{-(m+1)}}(\delta_m) \tag{3.14}$$

and extend this to $\hat{h}_\varepsilon \in \mathcal{H}$. Using Fatou's lemma, we obtain

$$\begin{aligned}
\mathbb{P}\Big(\bigcup_{m\in\mathbb{N}}\big\{\mathcal{X}_n \notin \mathbb{M}_I^{\delta_m,\hat{h}_\varepsilon(\delta_m)} \text{ eventually}\big\}\Big) &\le \sum_{m\in\mathbb{N}} \mathbb{E}\big(\liminf_{n\to\infty} \mathbb{1}_{\mathbb{M}_I\setminus\mathbb{M}_I^{\delta_m,\hat{h}_\varepsilon(\delta_m)}}(\mathcal{X}_n)\big) \\
&\le \sum_{m\in\mathbb{N}} \liminf_{n\to\infty}\mathbb{P}\big(\{\mathcal{X}_n \notin \mathbb{M}_I^{\delta_m,\hat{h}_\varepsilon(\delta_m)}\}\big) \\
&= \sum_{m\in\mathbb{N}} \big(1 - T(\varepsilon 2^{-(m+1)},\delta_m)\big) \\
&\le \sum_{m\in\mathbb{N}} \varepsilon 2^{-m} = \varepsilon.
\end{aligned} \tag{3.15}$$

Thus (3.11) holds with h_ε replaced by $\hat{h}_\varepsilon$. □

3.3. **Fmm-space-valued processes.** Let $J \subseteq \mathbb{R}_+$ be a (closed, open or half-open) interval and consider a stochastic process $\mathcal{X} = (\mathcal{X}_t)_{t\in J}$ with values in $\mathbb{M}_I$ and càdlàg paths, where $\mathbb{M}_I$ is equipped with the marked Gromov-weak topology. We say that $\mathcal{X}$ is an $\mathbb{M}_I^{\mathrm{fct}}$*-valued càdlàg process* if

$$\mathbb{P}\big(\{\mathcal{X}_t, \mathcal{X}_{t-} \in \mathbb{M}_I^{\mathrm{fct}} \text{ for all } t \in J\}\big) = 1, \tag{3.16}$$

where $\mathcal{X}_{t-}$ is the left limit of $\mathcal{X}$ at t ($\mathcal{X}_{\ell-} := \mathcal{X}_\ell$ if ℓ is the left endpoint of J). In the following, we give sufficient criteria for $\mathcal{X}$ to be an $\mathbb{M}_I^{\mathrm{fct}}$-valued càdlàg process. We are particularly interested in the situation where $\mathcal{X}$ is the limit of $\mathbb{M}_I^{\mathrm{fct}}$-valued processes $\mathcal{X}^n$.

Unsurprisingly, if the set of $\mathbb{P}$-measure smaller or equal to ε in Theorem 3.7 is independent of t, the result is true for all t simultaneously, almost surely. The modulus of continuity may also depend on t in a continuous way; or be arbitrary if the limiting process has continuous paths:

Theorem 3.9. *Let $J \subseteq \mathbb{R}_+$ be an interval, and $\mathcal{X}^n = (\mathcal{X}^n_t)_{t\in J}$, $n \in \mathbb{N}$, a sequence of $\mathbb{M}_I$-valued càdlàg processes converging in distribution to an $\mathbb{M}_I$-valued càdlàg process $\mathcal{X} = (\mathcal{X}_t)_{t\in J}$. Assume that for every $t \in J$, $\varepsilon > 0$, there exists $h_{t,\varepsilon} \in \mathcal{H}$ such that*

$$\limsup_{\delta\downarrow 0} \limsup_{n\to\infty} \mathbb{P}\big(\big\{ \mathcal{X}^n_t \in \mathbb{M}_I^{\delta,h_{t,\varepsilon}(\delta)} \ \forall t \in J \big\}\big) \geq 1 - \varepsilon. \tag{3.17}$$

Then $\mathcal{X}$ is an $\mathbb{M}_I^{\mathrm{fct}}$-valued càdlàg process, that is (3.16) *is satisfied, if at least one of the following two conditions holds:*

(i) *$\mathcal{X}$ has continuous paths a.s.*
(ii) *$t \mapsto h_{t,\varepsilon}(\delta)$ is continuous for every $\varepsilon, \delta > 0$.*

If additionally $\mathcal{X}^n$ is $\mathbb{M}_I^{\mathrm{fct}}$-valued almost surely for all $n \in \mathbb{N}$, (3.17) *can be replaced by existence of random measurable sets $Y^n_{t,\varepsilon,\delta} \subseteq X^n_t$, in addition to the $h_{t,\varepsilon} \in \mathcal{H}$, satisfying the following two conditions for every $\varepsilon > 0$:*

$$\limsup_{\delta\downarrow 0} \limsup_{n\to\infty} \mathbb{P}\big(\big\{ \nu^n_t(X^n_t \setminus Y^n_{t,\varepsilon,\delta}) \leq h_{t,\varepsilon}(\delta) \ \forall t \in J \big\}\big) \geq 1 - \varepsilon, \tag{3.18}$$

$$\forall n \in \mathbb{N},\ t \in J,\ \delta > 0,\ x, y \in Y^n_{t,\varepsilon,\delta} : r_n(x,y) < \delta \implies d\big(\kappa_n(x), \kappa_n(y)\big) \leq h_{t,\varepsilon}(\delta). \tag{3.19}$$

Proof. Due to the Skorohod representation theorem, we may assume that $\mathcal{X}^n \to \mathcal{X}$ almost surely in the Skorohod topology. For condition (i) respectively (ii) we obtain

(i) If $\mathcal{X}$ has continuous paths a.s., the convergence in Skorohod topology implies uniform convergence of $\mathcal{X}^n_t(\omega)$ on J a.s. with respect to d_{mGP}. Hence we have $\mathcal{X}^n_t \xrightarrow[n\to\infty]{\mathrm{mGw}} \mathcal{X}_t$ for all $t \in J$, almost surely, and we can proceed as in the proof of Theorem 3.7.
(ii) There are (random) continuous $w^n \colon J \to J$, converging to the identity uniformly on compacta, such that $\mathcal{X}^n_{w^n(t)} \to \mathcal{X}_t$ for all $t \in J$, almost surely. We can use the moduli of continuity $\hat{h}_{t,\varepsilon}(\delta) := h_{t,\varepsilon}(\delta) + \delta$ and proceed as in the proof of Theorem 3.7. Note here that, due to continuity of $h_{t,\varepsilon}(\delta)$ in t, there is for every compact subinterval $\mathcal{J}$ of J an $N_{\mathcal{J},\varepsilon,\delta} \in \mathbb{N}$ such that $\hat{h}_{t,\varepsilon}(\delta) \geq h_{w^n(t),\varepsilon}(\delta)$ for all $n \geq N_{\mathcal{J},\varepsilon,\delta}$ and $t \in \mathcal{J}$.
The same arguments apply for left limits with w^n_- such that $\mathcal{X}^n_{w^n_-(t)} \to \mathcal{X}_{t-}$. □

To use Theorem 3.9, we have to check in (3.17) or (3.18) a condition for uncountably many t simultaneously, which is often much more difficult than for every t individually. One situation, where it is easy to pass from individual t to all t simultaneously is the case where the moduli of continuity $h_{t,\varepsilon}$ actually do not depend on t and ε (see Corollary 3.13). The independence of ε, however, is a strong requirement. Therefore, we relax it to not blowing up too fast as $\varepsilon \downarrow 0$, where the "too fast" is determined by the following modulus of càdlàgness of the limiting process.

Definition 3.10 (modulus of càdlàgness)**.** *Let J be an interval, (E, r) a metric space, and $e = (e_t)_{t\in J} \in \mathcal{D}_E(J)$ a càdlàg path on J with values in E. Following* [Bil68, (14.44)], *set*

$$w''(e,\delta) := \sup_{t,t_1,t_2 \in J : t_1 \leq t \leq t_2, t_2 - t_1 \leq \delta} \min\big\{ r(e(t), e(t_1)),\, r(e(t_2), e(t))\big\}. \tag{3.20}$$

We say that e admits $w \in \mathcal{H}$ as modulus of càdlàgness *if $w''(e,\delta) \leq w(\delta)$ for all $\delta > 0$.*

Theorem 3.11. *Fix an interval $J \subseteq \mathbb{R}_+$. Let $\mathcal{X} = (\mathcal{X}_t)_{t\in J}$ and $\mathcal{X}^n = (\mathcal{X}^n_t)_{t\in J}$, $n \in \mathbb{N}$, be $\mathbb{M}_I$-valued càdlàg processes such that $\mathcal{X}^n$ converges in distribution to $\mathcal{X}$. Furthermore, assume that there is a dense set $Q \subseteq J$ and $w_\varepsilon, h_\varepsilon \in \mathcal{H}$, such that for all $\varepsilon > 0$*

$$\limsup_{n\to\infty} \mathbb{P}(\{\mathcal{X}^n_t \in \mathbb{M}_I^{\delta,h_\varepsilon(\delta)}\}) \geq 1 - \varepsilon \qquad \forall \delta > 0,\ t \in Q, \tag{3.21}$$

$$\mathbb{P}(\{t \mapsto \mathcal{X}_t \text{ admits } w_\varepsilon \text{ as modulus of càdlàgness w.r.t. } d_{\mathrm{mGP}}\}) \geq 1 - \varepsilon, \quad \text{and} \tag{3.22}$$

$$\liminf_{\delta \downarrow 0} h_{\varepsilon \cdot \delta}\big(2w_\varepsilon(\delta)\big) = 0. \tag{3.23}$$

Then $\mathcal{X}$ is an $\mathbb{M}_I^{\mathrm{fct}}$-valued càdlàg process, that is (3.16) *holds.*

Recall the decomposition $\mathbb{M}_I \setminus \mathbb{M}_I^{\mathrm{fct}} = \bigcup_{m \in \mathbb{N}} F_m$ with F_m defined in (2.4). The basic idea of the proof is to use the following lemma about càdlàg paths to show that, almost surely, the path of $\mathcal{X}$ avoids F_m. The assertion of the lemma follows easily using the triangle-inequality.

Lemma 3.12. *Let J be an interval, (E, r) a metric space, and $e = (e_t)_{t \in J} \in \mathcal{D}_E(J)$ a càdlàg path admitting modulus of càdlàgness $w \in \mathcal{H}$. Let $F \subseteq E$ be any set, $\delta > 0$, and $Q \subseteq J$ such that for all $t \in J$ there is $t_1, t_2 \in Q$ with $t_1 \leq t \leq t_2 \leq t_1 + \delta$. Then*

$$r(e_t, F) > w(\delta) \;\; \forall t \in Q \quad \Longrightarrow \quad e_t \notin F \text{ and } e_{t-} \notin F \;\; \forall t \in J. \tag{3.24}$$

Proof of Theorem 3.11. Because $\mathbb{M}_I^{h_\varepsilon} = \bigcap_{\delta > 0} \mathbb{M}_I^{\delta, h_\varepsilon(\delta)}$ is closed by Lemma 2.9, the Portmanteau theorem and (3.21) imply

$$\mathbb{P}(\{\mathcal{X}_t \notin \mathbb{M}_I^{h_\varepsilon}\}) < \varepsilon \qquad \forall t \in Q,\, \varepsilon > 0. \tag{3.25}$$

Due to the Skorohod representation theorem, we may assume that $\mathcal{X}^n \to \mathcal{X}$ almost surely in Skorohod topology. In order to simplify notation, we assume $J = [0,1]$ and $Q = \bigcup_{k \in \mathbb{N}} Q_k$ with $Q_k = \{\, i2^{-k} : i = 0, \ldots, 2^k \,\}$. It is enough to show for every $\varepsilon > 0$, $m \in \mathbb{N}$ and F_m as defined in (2.4) that

$$p_m := \mathbb{P}\big(\{\, \exists t \in [0,1] : \mathcal{X}_t \text{ or } \mathcal{X}_{t-} \in F_m \,\}\big) \leq 3\varepsilon. \tag{3.26}$$

To show (3.26), fix $\varepsilon > 0$ and $m \in \mathbb{N}$, and let $\mathcal{X}_t = (X_t, r_t, \mu_t)$. Because $\mathcal{X}$ has càdlàg paths, we find $K = K(\varepsilon) < \infty$ such that

$$\mathbb{P}\big(\{\, \sup_{t \in [0,1]} \|\mu_t\| \geq K - 3 \,\}\big) < \varepsilon. \tag{3.27}$$

According to (3.23) and (3.25), we can choose $k \in \mathbb{N}$ big enough such that for $h := h_{\varepsilon 2^{-k}}$ we have

$$h\big(2w_\varepsilon(2^{-k})\big) < (Km)^{-1} - 2w_\varepsilon(2^{-k}) \qquad \text{and} \qquad \mathbb{P}(\{\mathcal{X}_t \notin \mathbb{M}_I^h\}) < \varepsilon 2^{-k}. \tag{3.28}$$

Assume without loss of generality that $w_\varepsilon(2^{-k}) \leq 1$. Now Proposition 2.7(iii) implies that, whenever $\mathcal{X}_t \in \mathbb{M}_I^h$ and $\|\mu_t\| < K - 3$, we have

$$d_{\mathrm{mGP}}(\mathcal{X}_t, F_m) > w_\varepsilon(2^{-k}). \tag{3.29}$$

Combining (3.22) and Lemma 3.12, we obtain

$$p_m \leq \varepsilon + \mathbb{P}\big(\{\, \exists t \in Q_k : d_{\mathrm{mGP}}(\mathcal{X}_t, F_m) \leq w_\varepsilon(2^{-k}) \,\}\big). \tag{3.30}$$

Using (3.27), (3.29), and (in the last step) (3.28), we conclude

$$p_m \leq 2\varepsilon + 2^k \sup_{t \in Q_k} \mathbb{P}\big(\{\, \|\mu_t\| < K - 3,\, \mathcal{X}_t \notin \mathbb{M}_I^h \,\}\big) \leq 3\varepsilon. \tag{3.31}$$

Thus (3.26) holds for all $\varepsilon > 0$, and $\mathbb{P}(\{\exists t \in [0,1] : X_t \notin \mathbb{M}_I^{\mathrm{fct}}\}) = \sup_{m \in \mathbb{N}} p_m = 0$ follows. □

If, in Theorem 3.11, we can choose the modulus of continuity $h_\varepsilon = h \in \mathcal{H}$, independent of ε, such that (3.21) holds, we do not need to check (3.22) and (3.23).

Corollary 3.13 (ε-independent modulus of continuity)**.** *Assume that $\mathcal{X}^n = (\mathcal{X}_t^n)_{t \in J}$ converges in distribution to an $\mathbb{M}_I$-valued càdlàg process $\mathcal{X}$, and $Q \subseteq J$ is dense. Then $\mathcal{X}$ is an $\mathbb{M}_I^{\mathrm{fct}}$-valued càdlàg process if, for some $h \in \mathcal{H}$,*

$$\limsup_{n \to \infty} \mathbb{P}(\{\mathcal{X}_t^n \in \mathbb{M}_I^h\}) = 1 \qquad \forall t \in Q. \tag{3.32}$$

Proof. Let $h \in \mathcal{H}$ be such that (3.32) is satisfied and set $h_\varepsilon := h$. Then (3.23) is satisfied for every choice of $w_\varepsilon \in \mathcal{H}$, $\varepsilon > 0$. For every càdlàg process, in particular for $\mathcal{X}$, there exist moduli of càdlàgness w_ε such that (3.22) holds (cf. [Bil68, (14.6),(14.8) and (14.46)]). Thus, Theorem 3.11 yields the claim. □

4. EXAMPLES

The (neutral) tree-valued Fleming-Viot dynamics is constructed in [GPW13] using the formalism of metric measure spaces. In [DGP12], (allelic) types – encoded as marks of marked metric measure spaces – are included, in order to be able to model mutation and selection.

In [DGP12, Remark 3.11] and [DGP13, Theorem 6] it is stated that the resulting tree-valued Fleming-Viot dynamics with mutation and selection (TFVMS) admits a mark function at all times, almost surely. The given proof, however, contains a gap, because it relies on the criterion claimed in [DGP13, Lemma 7.1], which is wrong in general (see Example 4.1). The reason why the criterion may fail is a lack of homogeneity of the measure ν, in the sense that there are parts with high and parts with low mass density. Consequently, if we condition two samples to have distance less than ε, the probability that they are from the high-density part tends to one as $\varepsilon \downarrow 0$, and we do not "see" the low-density part. This phenomenon occurs if ν has an atom but is not purely atomic. We also give two non-atomic examples, one a subset of Euclidean space, and the other one ultrametric.

Example 4.1 (counterexamples)**.** In both examples, it is straight-forward to see that (X, r, μ), with $\mu = \nu \otimes K$, satisfies the assumptions of [DGP13, Lemma 7.1], but does not admit a mark function. The mark space is $I = \{0, 1\}$.

(i) Let λ_A be Lebesgue measure of appropriate dimension on a set A. Define $X := [0,1]^2 \cup [2,3]$, where $[2,3]$ is identified with $[2,3] \times \{0\} \subseteq \mathbb{R}^2$,

$$\nu := \tfrac{1}{2}(\lambda_{[0,1]^2} + \lambda_{[2,3]}) \quad \text{and} \quad K_x := \begin{cases} \frac{1}{2}(\delta_0 + \delta_1), & x \in [0,1]^2, \\ \delta_0, & x \in [2,3]. \end{cases} \tag{4.1}$$

(ii) In this example think of a tree consisting of a left part with tertiary branching points and a right part with binary branching points. The leaves correspond to $X := A \cup B$ with $A = \{0,1,2\}^{\mathbb{N}}$ and $B = \{3,4\}^{\mathbb{N}}$, and we choose as a metric

$$r\big((x_n)_{n\in\mathbb{N}}, (y_n)_{n\in\mathbb{N}}\big) := \max_{n\in\mathbb{N}} e^{-n} \cdot \mathbb{1}_{x_n \neq y_n}. \tag{4.2}$$

Note that (X, r) is a compact, ultrametric space. The measure ν is constructed as follows: choose the left respectively right part of the tree with probability $\frac{1}{2}$ each. Going deeper in the tree, at each branching point a branch is chosen uniformly. That is, let ν_A and ν_B be the Bernoulli measures on A and B with uniform marginals on $\{0,1,2\}$ and $\{3,4\}$, respectively. Define

$$\nu := \tfrac{1}{2}(\nu_A + \nu_B) \quad \text{and} \quad K_x := \begin{cases} \frac{1}{2}(\delta_0 + \delta_1), & x \in A, \\ \delta_0, & x \in B. \end{cases} \tag{4.3}$$

4.1. **The tree-valued Fleming-Viot dynamics with mutation and selection.** In the following, we prove the existence of a mark function for the TFVMS by verifying the assumptions of Theorem 3.9 for a sequence of approximating tree-valued Moran models. Due to the Girsanov transform given in [DGP12, Theorem 2], it is enough to consider the neutral case, that is without selection.

We briefly recall the construction of the tree-valued Moran model with mutation (TMMM) with finite population $U_N = \{1, \dots, N\}$, $N \in \mathbb{N}$, and types from the mark space I. For details and more formal definitions, see [DGP12, Subsections 2.1–2.3]. In the underlying Moran model with mutation (MMM), every pair of individuals "resamples" independently at rate $\gamma > 0$. Here,

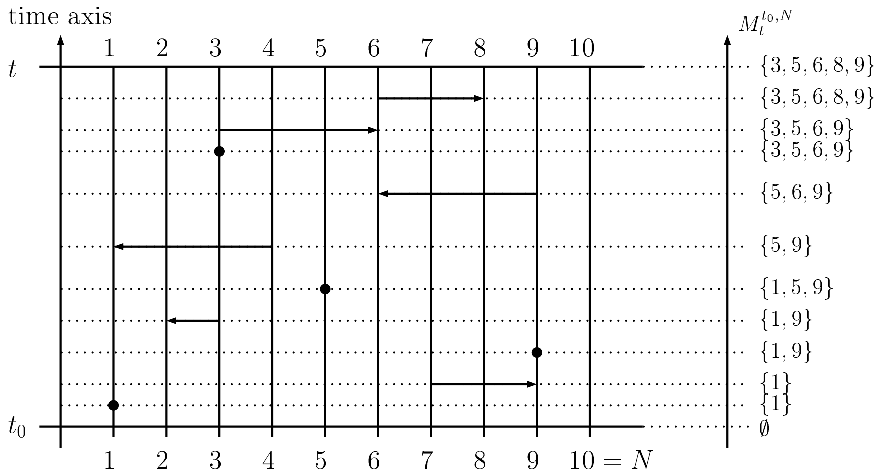


FIGURE 1. Graphical construction of the MMM for $N = 10$ for the time-period $[t_0, t]$, and the resulting process $(M_s^{t_0,N})_{s\in[t_0,t]}$. Resampling arrows are drawn at points of $\eta_{\mathrm{res}}^{k,\ell}$, and mutation dots at points of η_{mut}^{k}.

resampling means that one of the individuals (chosen uniformly at random among the two) is replaced by an offspring of the other one, and the offspring gets the same type as the parent. Furthermore, every individual mutates independently at rate $\vartheta \geq 0$, which means that it changes its type according to a fixed stochastic kernel $\beta(\cdot,\cdot)$ on I. Denote the resulting type of individual $x \in U_N$ at time $t \geq 0$ by $\kappa_t^N(x)$. To obtain the tree-valued dynamics, define the distance $r_t^N(x,y)$ between two individuals $x, y \in U_N$ at time $t \geq 0$ as twice the time to the most recent common ancestor (MRCA) (cf. [DGP12, (2.7)]), provided that a common ancestor exists, and as $2t+r_0^N(x,y)$ otherwise. The TMMM is the resulting process $\mathcal{X}_t^N = (U_N, r_t^N, \nu_N, \kappa_t^N)$, with sampling measure $\nu_N = \frac{1}{N}\sum_{k=1}^N \delta_k$. It is easy to check that, by definition, (U_N, r_t^N) is an ultrametric space, provided that the initial metric space (U_N, r_0^N) is ultrametric. This explains the name *tree-valued* (cf. [DGP12, Remark 2.7]).

Next recall the graphical construction of the MMM from [DGP12, Definition 2.2]. A resampling event is modeled by means of a family of independent Poisson point processes $\{\eta_{\mathrm{res}}^{k,\ell} : k, \ell \in U_N\}$ on $\mathbb{R}_+$, where each $\eta_{\mathrm{res}}^{k,\ell}$ has rate $\gamma/2$. If $t \in \eta_{\mathrm{res}}^{k,\ell}$, draw an arrow from (k,t) to (ℓ,t) to represent a resampling event at time t, where ℓ is an offspring of k. Similarly, model mutation times by a family of independent Poisson point processes $\{\eta_{\mathrm{mut}}^{k} : k \in U_N\}$, where each η_{mut}^{k} has rate ϑ. If $t \in \eta_{\mathrm{res}}^{k,\ell}$, draw a dot at (k,t) to represent a mutation event changing the type of individual k (see Figure 1).

Let $(M_t^{t_0,N})_{t\geq t_0}$, $M_t^{t_0,N} \subseteq U_N$ with $M_{t_0}^{t_0,N} = \emptyset$ be the process that records the individuals of the population at time t with an ancestor at a time $t_0 < s \leq t$ involved in a mutation event. By a coupling argument, this process can be constructed by means of the Poisson point processes $(\eta_{\mathrm{res}}^{k,\ell}, \eta_{\mathrm{mut}}^{k}, k, \ell \in U_N)$ as follows (compare Figures 1–2):
(4.4)
$$M_t^{t_0,N} = \begin{cases} M_{t-}^{t_0,N} \cup \{\ell\} & \text{if there is a resampling arrow from } k \in M_{t-}^{t_0,N} \text{ to } \ell \in U_N \text{ at time } t, \\ M_{t-}^{t_0,N} \cup \{k\} & \text{if there is a mutation event at } k \in U_N \text{ at time } t, \\ M_{t-}^{t_0,N} \backslash \{\ell\} & \text{if there is a resampling arrow from } k \notin M_{t-}^{t_0,N} \text{ to } \ell \in U_N \text{ at time } t. \end{cases}$$

Let $\xi_t^N := \frac{1}{N}\# M_{t_0+t}^{t_0,N}$ be the proportion of individuals at time $t_0 + t, t \geq 0$ whose ancestors have mutated after (the for the moment fixed) time t_0.

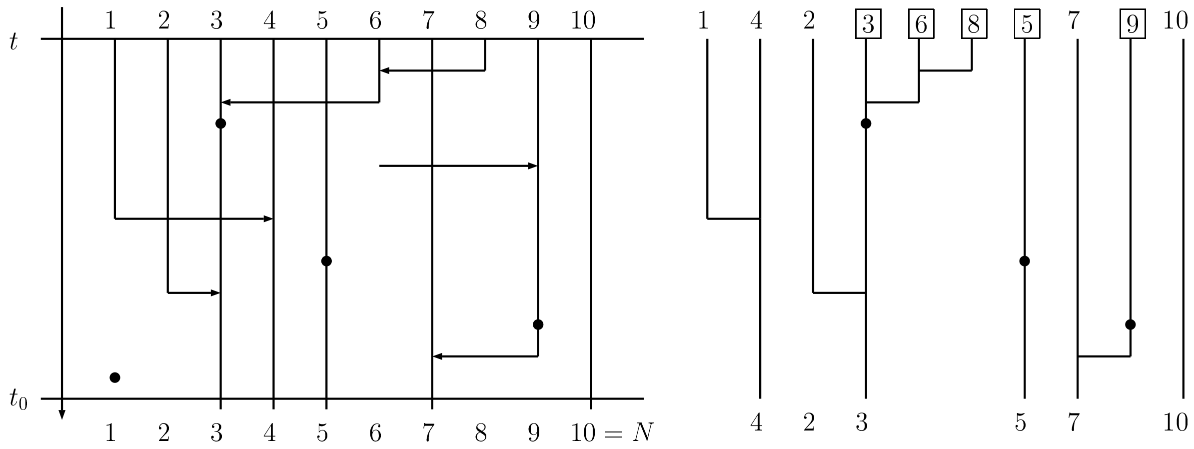


FIGURE 2. Tracing the ancestor backwards in time in Figure 1: This dual construction is also known as the coalescent backwards in time. Reverse the arrows to see for instance that 3 at time t_0 is an ancestor of 8 at time t. The elements of $M_t^{t_0,N} \subseteq U_N$ are highlighted by boxes in the right part of the picture.

Lemma 4.2. *Let* $C := \frac{1}{2}\vartheta(2\vartheta + \gamma)$*. Then for all* $a, \delta > 0$

$$\limsup_{N\to\infty} \mathbb{P}\big(\sup_{t\in[0,\delta]} \xi_t^N \geq a\big) \leq Ca^{-2}\delta^2. \tag{4.5}$$

Proof. By definition, $\big(\xi_t^N\big)_{t\geq 0}$ is a (continuous time) Markov jump process on $[0,1]$ with $\xi_0^N = 0$ and transitions

$$\begin{cases} x \mapsto x - 1/N & \text{at rate } \frac{\gamma}{2}N^2 x(1-x), \\ x \mapsto x + 1/N & \text{at rate } \frac{\gamma}{2}N^2 x(1-x) + \vartheta N(1-x). \end{cases} \tag{4.6}$$

This process converges weakly with respect to the Skorohod topology to the solution $(Z_t)_{t\geq 0}$ of the stochastic differential equation (SDE)

$$\mathrm{d}Z_t = \vartheta(1 - Z_t)\mathrm{d}t + \sqrt{\gamma Z_t(1 - Z_t)}\,\mathrm{d}B_t, \quad Z_0 = 0. \tag{4.7}$$

Indeed, to establish tightness use [EK05, Theorem III.9.4]. Note that, as $[0,1]$ is compact, it suffices to show the convergence of the generators applied to a set of appropriate test-functions. For existence and uniqueness of solutions to (4.7) reason as for the Bessel SDE in [RW00, (48.1) and below]. Moreover, $Z_t \in [0,1]$ is a bounded non-negative right-continuous submartingale. Hence, with Doob's submartingale inequality (see for instance [EK05, Proposition II.2.16(a)]), we obtain

$$\mathbb{P}\big(\sup_{t\in[0,\delta]} Z_t \geq a\big) = \mathbb{P}\big(\sup_{t\in[0,\delta]} Z_t^2 \geq a^2\big) \leq a^{-2}\mathbb{E}[Z_\delta^2]. \tag{4.8}$$

As $Z_t \in [0,1]$, we further deduce using Itô's formula that for all $t \geq 0$,

$$\mathbb{E}[Z_t] \leq \vartheta t \quad \text{and} \tag{4.9}$$

$$\mathbb{E}[Z_t^2] = \mathbb{E}\Big[\int_0^t 2Z_s\vartheta(1 - Z_s) + \gamma Z_s(1 - Z_s)\,\mathrm{d}s\Big] \leq Ct^2. \tag{4.10}$$

Then

$$\limsup_{N\to\infty} \mathbb{P}\big(\sup_{t\in[0,\delta]} \xi_t^N \geq a\big) \leq \mathbb{P}\big(\sup_{t\in[0,\delta]} Z_t \geq a\big) \leq Ca^{-2}\delta^2 \tag{4.11}$$

follows. $\square$

As the construction of the TFVMS in [DGP12] is only given for a compact type-space I, we make the same assumption. Note, however, that our proof itself does not use compactness and is therefore valid for non-compact I, provided that the TFVMS is the limit of the corresponding Moran models, and there exists a Girsanov transform allowing us to reduce to the neutral case.

Theorem 4.3 (the TFVMS admits a mark-function)**.** *Let I be compact and $\mathcal{X} = (\mathcal{X}_t)_{t\geq 0}$ be the tree-valued Fleming-Viot dynamics with mutation and selection as defined in* [DGP12]*. Then*

$$\mathbb{P}(\mathcal{X}_t \in \mathbb{M}_I^{\mathrm{fct}} \text{ for all } t > 0) = 1. \tag{4.12}$$

In particular, $(\mathcal{X}_t)_{t>0}$ is an $\mathbb{M}_I^{\mathrm{fct}}$-valued càdlàg process.

Proof. By [DGP12, Theorem 2], there exists a Girsanov transform that enables us to assume without loss of generality that selection is not present. In this case, according to [DGP12, Theorem 3], $\mathcal{X}$ is the limit in distribution of TMMMs $\mathcal{X}^N = (\mathcal{X}_t^N)_{t\geq 0}$, as discussed above. Let $\mathcal{X}_t^N = (U_N, r_t^N, \nu_N, \kappa_t^N)$ with $U_N = \{1, \ldots, N\}$ and ν_N the uniform distribution on U_N. Let $\delta > 0$ be fixed for the moment, and recall that the distance $r_t^N(x, y)$ between two individuals $x, y \in U_N$ at time $t \geq \delta/2$ is twice the time to the MRCA. Hence, if $r_t^N(x, y) < \delta$, then x and y at time t have a common ancestor at time $t - \delta/2$. Further recall that $(M_t^{t_0,N})_{t\geq t_0}$, with $M_t^{t_0,N} \subseteq U_N$ and $M_{t_0}^{t_0,N} = \emptyset$, records the individuals of the population at time t with an ancestor at a time $s \in (t_0, t]$ involved in a mutation event (cf. (4.4)).

Fix an arbitrary time horizon $T > 0$ and $i \in \mathbb{N}$, $i \leq 2T/\delta$. Using the notation of Theorem 3.9, for $t \in [i\delta/2, (i + 1)\delta/2)$, let $Y_{t,\varepsilon,\delta}^N := U_N \backslash M_t^{(i-1)\delta/2,N}$, independent of $\varepsilon > 0$. Set $Y_{t,\varepsilon,\delta}^N := \emptyset$ for $t < \delta/2$. We claim that (3.19) is satisfied for any choice of $h_{t,\varepsilon} \in \mathcal{H}$. Indeed, if $x, y \in Y_{t,\varepsilon,\delta}^N$ satisfy $r_t^N(x, y) < \delta$, then they have a common ancestor at time $t_0 := (i - 1)\delta/2 \leq t - \delta/2$, and after this point in time no mutation occurred along their ancestral lineages. In particular, $d(\kappa_t^N(x), \kappa_t^N(y)) = 0$, and (3.19) is obvious. Moreover, $\mathcal{X}^N$ is $\mathbb{M}_I^{\mathrm{fct}}$-valued by construction, and $\mathcal{X}$ has continuous paths by [DGP12, Theorem 1]. According to Theorem 3.9, it is therefore enough to find moduli of continuity $h_{t,\varepsilon} \in \mathcal{H}$ such that (3.18) holds for every $\varepsilon > 0$.

By Lemma 4.2, we obtain a constant $C > 0$ such that for every $a > 0$,

$$\limsup_{N\to\infty} \mathbb{P}\big(\sup_{t\in[i\delta/2,(i+1)\delta/2)} \nu_N\big(U_N \setminus Y_{t,\varepsilon,\delta}^N\big) \geq a\big) \leq Ca^{-2}\delta^2. \tag{4.13}$$

After summation over $i \in \{1, \ldots, \lfloor 2T/\delta \rfloor\}$, we obtain

$$\limsup_{N\to\infty} \mathbb{P}\big(\sup_{t\in[\delta/2,T]} \nu_N\big(U_N \backslash Y_{t,\varepsilon,\delta}^N\big) \geq a\big) \leq 2TC\delta a^{-2}. \tag{4.14}$$

For $\varepsilon > 0$ arbitrary, we use this inequality with $a := \sqrt{\varepsilon^{-1}2TC\delta}$, together with $\|\nu_N\| \leq 1$ for $t < \delta/2$, to see that (3.18) is satisfied for $h_{t,\varepsilon} \in \mathcal{H}$ with

$$h_{t,\varepsilon}(\delta) \geq \sqrt{\varepsilon^{-1}2TC\delta} + \mathbb{1}_{[2t,\infty[}(\delta). \tag{4.15}$$

□

4.2. **The tree-valued Λ-Fleming-Viot process.** Let Λ be a finite measure on $[0, 1]$, and recall the Λ-coalescent, introduced in [Pit99]. It is a coalescent process, where each k-tuple out of N blocks merges independently at rate

$$\lambda_{N,k} := \int_0^1 y^{k-2}(1 - y)^{N-k}\, \Lambda(\mathrm{d}y). \tag{4.16}$$

For fixed N, it is elementary to construct a finite, random (ultra-)metric measure space encoding the random genealogy of the Λ-coalescent, where the distance is defined as the time to the MRCA (recall the construction of Figures 1–2 and see Figure 3). In [GPW09, Theorem 4], existence and uniqueness of a Gromov-weak limit in distribution, as $N \to \infty$, is proven to be equivalent to the so-called "dust-free"-property, namely $\int_0^1 y^{-1}\, \Lambda(\mathrm{d}y) = \infty$. The resulting limit is called Λ-coalescent measure tree.

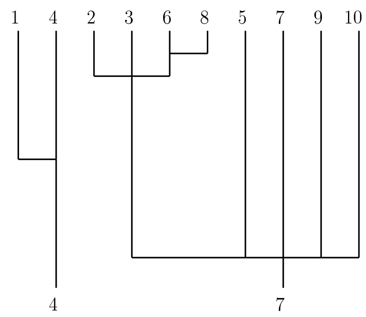


Figure 3. Tracing the ancestor backwards in time: The Λ-coalescent allows for one parent to have more than one child.

Now, replace the tree-valued Moran models considered in Subsection 4.1 and [DGP12] by so-called tree-valued Λ-Cannings models with Λ satisfying the dust-free-property. That is, leave the mutation- and selection-part as it is and change the resampling-part of the Moran models as follows: For $k = 2, \dots, N$, at rate $\binom{N}{k}\lambda_{N,k}$ a block of k individuals is chosen uniformly at random among the N individuals of the population. Upon such a resampling event, all individuals in this block are replaced by an offspring of a single individual which is chosen uniformly from this block. Note that the genealogy (disregarding types) of the resulting Λ-Cannings model with N individuals is dual to the Λ-coalescent starting with N blocks. We call any limit point (in path space) of the tree-valued Λ-Cannings processes, as N tends to infinity and Λ is fixed, *tree-valued Λ-Fleming-Viot process* (TLFV). In the neutral case, existence and uniqueness of such a limit point follows as a special case of the forthcoming work [GKW15]. Here, we show that, whenever limit points exist, all of them admit mark functions.

Theorem 4.4 (the TLFV admits a mark-function)**.** *Suppose there is no selection, that is $\alpha = 0$, and $\mathcal{X} = (\mathcal{X}_t)_{t\geq 0}$ is a tree-valued Λ-Fleming-Viot process with mutation. Then*

$$\mathbb{P}(\mathcal{X}_t \in \mathbb{M}_I^{\text{fct}} \text{ for all } t > 0) = 1. \tag{4.17}$$

Proof. By passing to a subsequence if necessary, we may assume that the Λ-Cannings models converge in distribution to $\mathcal{X}$. We proceed as in Subsection 4.1. Again, let $(M_t^{t_0,N})_{t\geq t_0}$, $M_t^{t_0,N} \subseteq U_N$ with $M_{t_0}^{t_0,N} = \emptyset$ be the process that records the individuals of the population at time t with an ancestor at a time $t_0 < s \leq t$ involved in a mutation event and $\xi_t^N := \frac{1}{N}\# M_{t_0+t}^{t_0,N}$ be the proportion of individuals at time $t_0 + t, t \geq 0$ whose ancestors have mutated after (the for the moment fixed) time t_0. By definition, $\big(\xi_t^N\big)_{t\geq 0}$ is a (continuous time) Markov jump process on $[0,1]$ with $\xi_0^N = 0$ and generator

$$\begin{aligned}(\Omega^N f)(x) = \vartheta N(1-x)\big(f(x+1/N) - f(x)\big) + \sum_{k=2}^{N} \lambda_{N,k} \sum_{m=0}^{(Nx)\wedge k} \binom{Nx}{m}\binom{N(1-x)}{k-m} \\ \times \Big(\frac{m}{k}\big(f(x+(k-m)/N) - f(x)\big) + \frac{k-m}{k}\big(f(x-m/N) - f(x)\big)\Big),\end{aligned} \tag{4.18}$$

where $x \in [0,1], N\cdot x \in \mathbb{N}\cup\{0\}, f \in \mathcal{C}_b^2([0,1])$. Due to Taylor's formula, there is $x_{m,k,N}^+ \in [x, x+(k-m)/N]$, $x_{m,k,N}^- \in [x - m/N, x]$ with

$$\begin{aligned}(\Omega^N f)(x) = \vartheta N(1-x)\big(f(x+1/N) - f(x)\big) + \sum_{k=2}^{N} \lambda_{N,k} \sum_{m=0}^{(Nx)\wedge k} \binom{Nx}{m}\binom{N(1-x)}{k-m} \\ \times \Big(\frac{f''(x_{m,k,N}^+)}{2}\frac{m(k-m)^2}{kN^2} + \frac{f''(x_{m,k,N}^-)}{2}\frac{(k-m)m^2}{kN^2}\Big)\end{aligned} \tag{4.19}$$

$$= \vartheta(1-x)f'(x) + O(N^{-1}) + x(1-x)\sum_{k=2}^{N} \lambda_{N,k}\Delta_{N,k}(x),$$

where, using $\binom{n}{i} = \frac{n}{i}\binom{n-1}{i-1}$ for $i \geq 1$,

$$\Delta_{N,k}(x) = \sum_{m=1}^{(Nx)\wedge(k-1)} \binom{Nx-1}{m-1}\binom{N(1-x)-1}{k-m-1}\Big(f''(x^{+}_{m,k,N})\frac{k-m}{2k} + f''(x^{-}_{m,k,N})\frac{m}{2k}\Big). \tag{4.20}$$

Recall that $\sum_{m=0}^{k}\binom{\ell}{m}\binom{N-\ell}{k-m} = \binom{N}{k}$ and $\lambda_{N,k} = \int_0^1 y^{k-2}(1-y)^{N-k}\,\Lambda(\mathrm{d}y)$ with a finite measure Λ on $[0,1]$ to see that

$$\begin{aligned} \sum_{k=2}^{N} |\Delta_{N,k}(x)| &\leq \|f''\|_\infty \sum_{k=2}^{N} \lambda_{N,k} \sum_{m=0}^{k-2} \binom{Nx-1}{m}\binom{N(1-x)-1}{k-2-m} \\ &= \|f''\|_\infty \int_0^1 \sum_{k=2}^{N} \binom{N-2}{k-2} y^{k-2}(1-y)^{N-k}\,\Lambda(\mathrm{d}y) \\ &= \|f''\|_\infty\,\Lambda([0,1]). \end{aligned} \tag{4.21}$$

Therefore,

$$\big(\Omega^N f\big)(x) = \vartheta(1-x)f'(x) + O(N^{-1}) + x(1-x)O(1). \tag{4.22}$$

Use $f(x) = x, x \in [0,1]$ in (4.19) to see that $(\xi_t^N)_{t\geq 0}$ is a non-negative right-continuous submartingale with $\xi_0^N = 0$ and $\mathbb{E}[\xi_t^N] \leq \vartheta t$. Use $f(x) = x^2$ to deduce from (4.22) that

$$\mathbb{E}\big[(\xi_t^N)^2\big] \leq Ct^2 + O(N^{-1})t. \tag{4.23}$$

Now reason as for the TFVMS in the proofs of Lemma 4.2 and Theorem 4.3 to complete the claim. □

4.3. **Future application: Evolving phylogenies of trait-dependent branching.** In [KW15] the results of the present paper will be applied in a context of evolving genealogies to establish the existence of a mark function with the help of Theorem 3.9. These genealogies are random marked metric measure spaces, constructed as the limit of approximating particle systems. The individual birth- respectively death-rates in the N^{th}-approximating population depend on the present trait of the individuals alive and are of order $O(N)$. At each birth-event, mutation happens with a fixed probability. Each individual is assigned mass $1/N$. The metric under consideration is genetic distance: in the N^{th}-approximating population genetic distance is increased by $1/N$ at each birth with mutation. Hence, genetic distance of two individuals is counted in terms of births with mutation backwards in time to the MRCA rather than in terms of the time to the MRCA.

Because of the use of exponential times in the modeling of birth- and death-events in this therefore non-ultrametric setup the analysis of the modulus of continuity of the trait-history of a particle in combination with the evolution of its genetic age plays a major role in establishing tightness of the approximating systems and existence of a mark function. In [Kli14, Lemma 3.9], control on the modulus of continuity is obtained by transferring the model to the context of historical particle systems. In a first step, time is related to genetic distance by means of the modulus of continuity. The extend of the change of trait of an individual in a small amount of time (recall (3.9) and (3.3)) can then be controlled by means of the modulus of continuity of its trait-path in combination with a control on the height of the largest jump during this period of time. This can in turn be ensured by appropriate assumptions on the mutation transition kernels of the approximating systems.

Acknowledgements. We are thankful to Anita Winter for discussions in the initial phase of the project, and to the referee for helpful comments. The research of Sandra Kliem was supported by the DFG through the SPP Priority Programme 1590.

References

[Ald93] David Aldous, *The continuum random tree III*, Ann. Probab. **21** (1993), no. 1, 248–289.

[Bil68] Patrick Billingsley, *Convergence of probability measures*, John Wiley & Sons, Inc., New York-London-Sydney, 1968. MR 0233396 (38 #1718)

[Bog07] V. I. Bogachev, *Measure theory, volume II*, Springer, 2007.

[DGP11] Andrej Depperschmidt, Andreas Greven, and Peter Pfaffelhuber, *Marked metric measure spaces*, Electron. Commun. Probab. **16** (2011), no. 17, 174–188.

[DGP12] ———, *Tree-valued Fleming-Viot dynamics with mutation and selection*, Ann. Appl. Probab. **22** (2012), no. 6, 2560–2615.

[DGP13] ———, *Path-properties of the tree-valued Fleming-Viot processes*, Electron. J. Probab. **18** (2013), no. 84, 1–47.

[EK05] S. N. Ethier and T. G. Kurtz, *Markov processes, characterization and convergence*, Wiley, 2005.

[GKW15] Andreas Greven, Anton Klimovsky, and Anita Winter, *Evolving genealogies for spatial Λ-Fleming-Viot processes with mutation*, In preparation, 2015.

[GPW09] Andreas Greven, Peter Pfaffelhuber, and Anita Winter, *Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees)*, Probab. Theory Related Fields **145** (2009), no. 1-2, 285–322.

[GPW13] ———, *Tree-valued resampling dynamics. Martingale problems and applications*, Probab. Theory Related Fields **155** (2013), 789–838.

[Gro99] Misha Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, vol. 152, Birkhäuser Boston, Inc., Boston, MA, 1999. MR 1699320 (2000d:53065)

[Kle14] Achim Klenke, *Probability theory*, second ed., Universitext, Springer, London, 2014. MR 3112259

[Kli14] Sandra Kliem, *A compact containment result for nonlinear historical superprocess approximations for population models with trait-dependence*, Electron. J. Probab. **19** (2014), no. 97, 1–13.

[KW15] Sandra Kliem and Anita Winter, *Evolving phylogenies of trait-dependent branching with mutation and competition*, In preparation, 2015.

[Löh13] Wolfgang Löhr, *Equivalence of Gromov-Prohorov- and Gromov's $\underline{\square}_\lambda$-metric on the space of metric measure spaces*, Electron. Commun. Probab. **18** (2013), no. 17, 1–10. MR 3037215

[LVW14] Wolfgang Löhr, Guillaume Voisin, and Anita Winter, *Convergence of bi-measure $\mathbb{R}$-trees and the pruning process*, Ann. Inst. H. Poincaré Probab. Statist. **in press** (2014), 36 pages, arXiv:1304.6035.

[Pio11] Sven Piotrowiak, *Dynamics of genealogical trees for type- and state-dependent resampling models*, Ph.D. thesis, University of Erlangen-Nuremberg, 2011.

[Pit99] Jim Pitman, *Coalescents with multiple collisions*, Ann. Probab. **27** (1999), no. 4, 1870–1902. MR 1742892 (2001h:60016)

[RW00] L. C. G. Rogers and D. Williams, *Diffusions, markov processes and martingales, volume II*, second ed., Cambridge University Press, 2000.

CONVERGENCE OF BI-MEASURE $\mathbb{R}$-TREES AND THE PRUNING PROCESS

WOLFGANG LÖHR, GUILLAUME VOISIN, AND ANITA WINTER

Abstract. In [AP98b] a tree-valued Markov chain is derived by pruning off more and more subtrees along the edges of a Galton-Watson tree. More recently, in [AD12], a continuous analogue of the tree-valued pruning dynamics is constructed along Lévy trees. In the present paper, we provide a new topology which allows to link the discrete and the continuous dynamics by considering them as instances of the same strong Markov process with different initial conditions. We construct this pruning process on the space of so-called bi-measure trees, which are metric measure spaces with an additional pruning measure. The pruning measure is assumed to be finite on finite trees, but not necessarily locally finite. We also characterize the pruning process analytically via its Markovian generator and show that it is continuous in the initial bi-measure tree. A series of examples is given, which include the finite variance offspring case where the pruning measure is the length measure on the underlying tree.

Contents

1. Introduction and motivation

Let $\mathcal{G}_1$ be a rooted Galton-Watson tree with an offspring generating function g. For $0 \le u \le 1$, let $\mathcal{G}_u$ be the subtree of $\mathcal{G}_1$ obtained by retaining each edge with probability u. Lyons ([Lyo92]) showed that $\mathcal{G}_u$ is again a Galton-Watson tree which corresponds to the offspring generating function $g_u = g(1 - u + u\cdot)$. As one can couple the pruning procedures for several $u \in [0,1]$ in such a way that $\mathcal{G}_{u'}$ is a rooted subtree of $\mathcal{G}_u$ whenever $u' \le u$, they give rise to a non-decreasing tree-valued Markov process $(\mathcal{G}_u)_{u\in[0,1]}$ which was further studied in Aldous and Pitman ([AP98b]). Recently, Abraham, Delmas and He consider in [ADH12] another pruning procedure on Galton Watson trees where cut points fall on the branch points to the effect that the subtree above is pruned. Here each node of the initial Galton-Watson tree is cut independently with probability $1 - u^{n-1}$ where n is the number of children of the node.

In the same spirit some authors consider continuum tree analogues of pruning dynamics. Compare, for example, [AP98a, AS02] for a pruning proportional to the length on the skeleton of a Brownian CRT, [Mie05] for a pruning on the infinite branch points of a stable Lévy tree, [AD08] for a pruning on the infinite branch points of a Lévy tree without Brownian part, [ADV10, AD12] for a combined pruning proportional to the length and on the infinite branch points of a general Lévy tree.

In [AD12] it is conjectured that the pruning procedure presented in the same paper is the continuous analogue of a mixture of the pruning procedures suggested in [AP98b] and [AD12], that is of pruning procedures on Galton-Watson trees where cut points fall on edges as well as on nodes. However, no precise link between the discrete and the continuum tree-valued dynamics has been given so far. The main goal of the present paper is to present ONE Markov process, which in the following is referred to as THE *pruning process*. We shall give an analytic characterization via a Markovian generator and provide with the so-called *leaf-sampling weak vague topology* a notion of convergence which allows to state convergence of the discrete tree-valued dynamics to the associated continuous tree-valued dynamics.

It had been a long tradition to encode trees via continuous excursions, and to use uniform topology as a notion of convergence. A more recent and conceptional approach is to think of trees as "tree-like" metric spaces, the so-called $\mathbb{R}$-trees, and to use the Gromov-Hausdorff topology as a notion of convergence (compare, for example, [DT96] for an introduction into $\mathbb{R}$-trees and [Gro99, EPW06] for details on the Gromov-Hausdorff distance). For a long time convergence of suitably rescaled Galton-Watson processes were established for very particular offspring distributions only. To be in a position to prove an invariance principle, Aldous developed in [Ald91, Ald93] a notion of convergence by encoding trees as closed subsets of ℓ^1_+, the space of summable sequences of positive numbers which were additionally equipped with a sampling measure. Convergence was then proposed as the convergence of all subtrees spanned by finite samples from the tree. Once more, this very neat and powerful idea had been generalized to the more conceptional encoding of trees as metric probability measure spaces where the tree space was equipped with the so-called *Gromov-weak topology* (compare [Gro99, GPW09]). Further developments which combine the Gromov-Hausdorff and Gromov-weak topology and allow for sampling measures that are finite on bounded sets can be found, for example, in [EW06, Mie09, ADH13].

In the present paper, we provide a unified framework by regarding these pruning processes as the *same* Feller-continuous Markov process on a (non locally compact) space of $\mathbb{R}$-trees with different initial conditions, and to establish convergence in Skorohod space whenever the initial distribution converges. For that purpose, we introduce (rooted) *bi-measure $\mathbb{R}$-trees*, i.e., metric measure spaces (T, r, μ), which are additionally equipped with a so-called *pruning* measure, ν. Here, the so-called *sampling measure* μ is a finite measure (allowing for a varying total mass), while the pruning measure is only assumed to be finite on finite subtrees. As the pruning measure is already part of the state, we are in a position to construct one (universal) pruning process. This process is a pure jump process which, given a bi-measure $\mathbb{R}$-tree, lets rain down successively more and more cut points according to a Poisson process whose intensity measure is equal to the pruning measure. At each cut point, the subtree above (seen from the fixed root) is cut off and removed, and the sampling and pruning measures are simultaneously updated by simply restricting them to the remaining, pruned part of the tree.

A major difficulty is that important examples for the pruning measures, such as the length measure on the Brownian CRT, are not locally finite. Therefore, we introduce with the *leaf-sampling weak vague topology* a new topology on the spaces of bi-measure $\mathbb{R}$-trees. We give equivalent characterizations of convergence and provide convergence determining classes of functions. We also obtain a simple sufficient (but far from necessary) compactness condition.

Outline. The paper is organized as follows. In Section 2 we introduce the leaf-sampling weak-vague topology and give a characterization of convergence. In Section 3 we construct the pruning process, calculate its Markovian generator and verify that the law of the process on Skorohod space depends continuously on the initial condition. Finally, in Section 4 we apply our main result to obtain convergence of various pruning processes that appeared in the literature.

2. Bi-measure $\mathbb{R}$-trees and the LWV-topology

In this section we introduce the space of $\mathbb{R}$-trees equipped with a finite *sampling measure* and a *pruning measure* which is assumed to be finite on finite subtrees. Moreover, we define the *leaf-sampling weak vague topology* (LWV-topology) on this space of bi-measure $\mathbb{R}$-trees. The idea behind our topology is to first sample a finite number of points from the tree according to the sampling measure. These points span a finite subtree. In many relevant examples they are actually the leaves of this subtree. Then we equip this finite subtree with the restriction of the pruning measure and obtain a random metric measure tree. For convergence of bi-measure trees, we require that these random metric measure trees converge together with the sampled points as *n-pointed metric measure $\mathbb{R}$-trees* in the Gromov-weak topology.

We therefore recall in Subsection 2.1 the notion of Gromov-weak topology on metric measure spaces and extend it to the *n-pointed Gromov-weak topology.* In Subsection 2.2 we then define a stronger topology on n-pointed metric measure $\mathbb{R}$-trees, the subtree Gromov-weak topology. Finally, in Subsection 2.3 we define the LWV-convergence. It turns out that it can be characterized by both the pointed as well as the subtree Gromov-weak convergence of samples from the bi-measure $\mathbb{R}$-tree and defines a separable, metrizable topology. In Subsection 2.4, we introduce classes of test functions that induce the LWV-topology. One of them turns out to be convergence determining. Using these test functions, we derive several convergence results.

2.1. **The n-pointed Gromov-weak topology.** Greven, Pfaffelhuber and Winter [GPW09] define the space of metric probability measure spaces equipped with the Gromov-weak topology. In this subsection, we define a slightly more general space using finite measures instead of probability measures and considering n-pointed metric measure spaces. We do not give proofs, because the extension is straightforward.

We start recalling basic notation. As usual, given a topological space X, we denote by $\mathcal{C}(X)$ ($\mathcal{C}_b(X)$) the space of (bounded) continuous, $\mathbb{R}$-valued functions on X, and by $\mathcal{M}_1(X)$ ($\mathcal{M}_f(X)$) the space of probability (finite) measures, defined on the Borel σ-algebra of X. For $x \in X$, $\delta_x \in \mathcal{M}_1(X)$ is the Dirac measure in the point x. "$\Rightarrow$" means weak convergence in $\mathcal{M}_1(X)$ or in $\mathcal{M}_f(X)$. Recall that the support of μ, $\mathrm{supp}(\mu)$, is the smallest closed set $X_0 \subseteq X$ such that $\mu(X_0) = \mu(X) =: \|\mu\|$. For $\mu \in \mathcal{M}_f(X)$, we denote the normalization by

$$\mu^\circ := \tfrac{\mu}{\|\mu\|} \in \mathcal{M}_1(X). \tag{2.1}$$

The *push forward* of μ under a measurable map ϕ from X into another topological space Z is the finite measure $\phi_*\mu \in \mathcal{M}_f(Z)$ defined by

$$\phi_*\mu(A) := \mu\big(\phi^{-1}(A)\big), \tag{2.2}$$

for all measurable subsets $A \subseteq Z$. For the integral of an integrable function φ with respect to μ, we sometimes use the notation

$$\langle \mu, \varphi \rangle := \int \varphi \, \mathrm{d}\mu. \tag{2.3}$$

A *metric measure space* is a triple (X, r, μ), where (X, r) is a metric space such that $\big(\mathrm{supp}(\mu), r\big)$ is complete and separable and $\mu \in \mathcal{M}_f(X)$ is a finite measure on $(X, \mathcal{B}(X))$. If $\mathrm{supp}(\mu)$ is separable but not complete, we simply identify it with its completion.

Branching trees such as Galton-Watson trees and the CRT are often rooted. We therefore define a *rooted metric measure space* (X, r, ρ, μ) as a metric measure space (X, r, μ) together with a distinguished point $\rho \in X$ which is referred to as the *root.* To avoid heavy notations, in the following we suppress the metric and the root, i.e. we abbreviate, for example,

$$X = (X, r, \rho), \qquad (X, \mu) = (X, r, \rho, \mu). \tag{2.4}$$

The definition of metric measure spaces given in [GPW09] can easily be extended to rooted metric measure spaces. In the context of metric spaces, rooted spaces are often referred to as *pointed spaces* (compare, for example, Section 8 in [BBI01]).

We want to extend these rooted metric measure spaces (X, r, μ) by fixing n additional points $u_1, \ldots, u_n \in X$, and call $(X, r, \rho, (u_1, \ldots, u_n), \mu)$ a (rooted) *n-pointed metric measure space.* The support of an n-pointed metric measure space $(X, r, \rho, (u_1, \ldots, u_n), \mu)$ is defined by

$$\operatorname{supp}\big((X, r, \rho, (u_1, \ldots, u_n), \mu)\big) := \operatorname{supp}(\mu) \cup \{\rho, u_1, \ldots, u_n\}. \tag{2.5}$$

In the following we identify two n-pointed metric measure spaces if there is a measure preserving isometry between their supports that also preserves the root and the fixed points.

Definition 2.1 (Equivalence and the space $\mathbb{M}_n$)**.** *We call two n-pointed metric measure spaces, $\mathcal{X} = (X, r, \rho, (u_1, \ldots, u_n), \mu)$ and $\mathcal{X}' = (X', r', \rho', (u'_1, \ldots, u'_n), \mu')$,* equivalent *if there exists an isometry ϕ between* $\operatorname{supp}(\mathcal{X})$ *and* $\operatorname{supp}(\mathcal{X}')$ *such that $\phi_*\mu = \mu'$, $\phi(\rho) = \rho'$ and $\phi(u_k) = u'_k$ for all $1 \le k \le n$. It is clear that this defines an equivalence relation.*

We denote by $\mathbb{M}_n$ the set of equivalence classes of n-pointed metric measure spaces.

Remark 2.2. Notice that for a notion of equivalence of metric measure spaces (X, r, μ) and (X', r', μ') there are two canonical choices. Either we insist that the metric spaces (X, r) and (X', r') are isometric or we are satisfied with their supports to be isometric thereby neglecting sets of measure zero (compare, for example, [Vil09, Section 27]). Here we take the second approach which allows for a characterization of convergence in $\mathbb{M}_n$ through convergence determining classes of functions. The gap between such a notion of (weak) convergence and a stronger topology which also requires the convergence of supports of the measures is closed in [ALW14]. $\Diamond$

To simplify notations, we do not distinguish between an n-pointed metric measure space and its equivalence class. That is, we write

$$\mathcal{X} = (X, (x_1, \ldots, x_n), \mu) \in \mathbb{M}_n. \tag{2.6}$$

Remark 2.3 (The space $\mathbb{M}_0$)**.** $\mathbb{M}_0$ is the usual space of rooted metric measure spaces (with finite measures). $\Diamond$

For a rooted metric space X, we define a map R^X that associates to a sequence of points the matrix of their distances to the root and to each other, i.e.,

$$R^X \colon \begin{cases} X^{\mathbb{N}} & \to \quad \mathbb{R}_+^{\binom{\mathbb{N}_0}{2}} \\ (x_i)_{i \ge 1} & \mapsto \quad \big(r(x_i, x_j)\big)_{0 \le i < j} \text{ with } x_0 := \rho. \end{cases} \tag{2.7}$$

The *distance matrix distribution* of an n-pointed metric measure space $\mathcal{X} = (X, (u_1, \ldots, u_n), \mu)$ is then given by

$$\boldsymbol{\upsilon}^{\mathcal{X}} := \|\mu\| \cdot \big(R^X\big)_* \Big(\bigotimes_{k=1}^{n} \delta_{u_k} \otimes (\mu^\circ)^{\otimes \mathbb{N}}\Big) \in \mathcal{M}_f(\mathbb{R}_+^{\binom{\mathbb{N}_0}{2}}), \tag{2.8}$$

which obviously depends only on the equivalence class. Vershik's proof of Gromov's reconstruction theorem (see [Gro99, $3\frac{1}{2}$.7]) directly carries over to n-pointed metric measure spaces. Therefore, $\mathcal{X} \in \mathbb{M}_n$ is uniquely determined by its distance matrix distribution $\boldsymbol{\upsilon}^{\mathcal{X}}$.

Definition 2.4 (pGw-topology)**.** *A sequence of n-pointed metric measure spaces $\mathcal{X}_N \in \mathbb{M}_n$ converges n-pointed Gromov-weakly (*pGw*) to $\mathcal{X} \in \mathbb{M}_n$ if*

$$\boldsymbol{\upsilon}^{\mathcal{X}_N} \underset{N \to \infty}{\Longrightarrow} \boldsymbol{\upsilon}^{\mathcal{X}} \tag{2.9}$$

in the weak topology on $\mathcal{M}_f\big(\mathbb{R}_+^{\binom{\mathbb{N}_0}{2}}\big)$.

We see directly from the definition that functions of the form $\Phi\colon \mathbb{M}_n \to \mathbb{R}$, $\mathcal{X} \mapsto \langle \boldsymbol{v}^{\mathcal{X}}, f\rangle$ with $f \in \mathcal{C}_b\big(\mathbb{R}_+^{\binom{\mathbb{N}_0}{2}}\big)$ are continuous. If f depends only on finitely many coordinates, Φ is called a *polynomial*, and there exists $m \in \mathbb{N}$, $\varphi \in \mathcal{C}_b\big(\mathbb{R}_+^{\binom{n+m+1}{2}}\big)$ such that for $\mathcal{X} = (X, \underline{u}, \mu) \in \mathbb{M}_n$,

$$\Phi(\mathcal{X}) = \Phi^{m,\varphi}(\mathcal{X}) := \int_{X^m} \mu^{\otimes m}(\mathrm{d}\underline{v})\, \varphi\big(R^X(\underline{u}, \underline{v})\big), \tag{2.10}$$

where $\underline{u} = (u_1, ..., u_n)$, $\underline{v} = (v_1, ..., v_m)$ and $(\underline{u}, \underline{v}) := (u_1, \dots, u_n, v_1, \dots, v_m)$. Note that

$$\Phi^{m,1}(\mathcal{X}) = \|\mu\|^m, \tag{2.11}$$

and, in particular, $\Phi^{0,1} \equiv 1$. Moreover, as polynomials are not bounded (compare (2.11)), we define a class $\Pi_n \subseteq \mathcal{C}_b(\mathbb{M}_n)$ of bounded test functions by

$$\begin{aligned} \Pi_n := \big\{ &\Phi^{\gamma,m,\varphi}(\mathcal{X}) := \gamma(\|\mu\|) \cdot \Phi^{m,\varphi}(\mathcal{X}) : \\ &\Phi^{m,\varphi} \text{ is a polynomial, } \gamma \in \mathcal{C}_b(\mathbb{R}_+),\ \lim_{x\to\infty} x^k \gamma(x) = 0,\ \forall k \in \mathbb{N}\big\}. \end{aligned} \tag{2.12}$$

Recall the *Prohorov distance* d_{Pr} between two finite measures μ, ν on a metric space (Z, d),

$$\begin{aligned} &d_{\mathrm{Pr}}^{(Z,d)}(\mu, \nu) \\ &:= \inf\big\{\varepsilon > 0 :\ \mu(A^\varepsilon) + \varepsilon \geq \nu(A),\ \nu(A^\varepsilon) + \varepsilon \geq \mu(A)\, \forall A \text{ closed}\big\}, \end{aligned} \tag{2.13}$$

where $A^\varepsilon := \{x \in Z \mid d(x, A) < \varepsilon\}$.

Definition 2.5 (n-pointed Gromov-Prohorov distance)**.** *Define the n-pointed Gromov-Prohorov distance between $\mathcal{X} = (X, \underline{u}, \mu)$ and $\mathcal{Y} = (Y, \underline{v}, \nu)$ in $\mathbb{M}_n$ by*

$$d_{\mathrm{pGP}}\big(\mathcal{X}, \mathcal{Y}\big) := \inf_d \big\{ d_{\mathrm{Pr}}^{(X \uplus Y, d)}(\mu, \nu) + d(\rho_X, \rho_Y) + \sum_{k=1}^n d(u_k, v_k)\big\}, \tag{2.14}$$

where the infimum is taken over all metrics d on the disjoint union $X \uplus Y$ that extends r_X and r_Y. If there is no confusion, we simply write d_{Pr} for $d_{\mathrm{Pr}}^{(X\uplus Y,d)}$.

Recall that a set $\mathcal{F} \subseteq \mathcal{C}_b(X)$ is *convergence determining* (on the topological space X) if, for probability measures μ_N, μ on X, the weak convergence $\mu_N \underset{N\to\infty}{\Longrightarrow} \mu$ is equivalent to $\int f \,\mathrm{d}\mu_N \underset{N\to\infty}{\longrightarrow} \int f \,\mathrm{d}\mu$ for all $f \in \mathcal{F}$.

Proposition 2.6 (Π_n is convergence determining)**.** *Let $\mathcal{X}, \mathcal{X}_1, \mathcal{X}_2, \ldots \in \mathbb{M}_n$. The following conditions are equivalent:*

(i) $\mathcal{X}_N \xrightarrow{\mathrm{pGw}} \mathcal{X}$, *as* $N \to \infty$.
(ii) $\Phi(\mathcal{X}_N) \underset{N\to\infty}{\longrightarrow} \Phi(\mathcal{X})$, *for all polynomials* Φ.
(iii) $d_{\mathrm{pGP}}(\mathcal{X}_N, \mathcal{X}) \underset{N\to\infty}{\longrightarrow} 0$.

Furthermore, $\mathbb{M}_n$ is separable, d_{pGP} is a complete metric on $\mathbb{M}_n$, and the class $\Pi_n \subseteq \mathcal{C}_b(\mathbb{M}_n)$ is convergence determining on $\mathbb{M}_n$.

Proof. The proof of the equivalences is an obvious modification of that of Theorem 5 in [GPW09]. Notice that μ° in the definition of the pGw-topology can be replaced by μ in the definition of polynomials because $\|\mu\| = \Phi^{1,1}(\mathcal{X})$. Separability and completeness follow in the same way as Proposition 5.6 in [GPW09].

To see that Π_n induces the pGw-topology, note that for $\gamma(x) = e^{-x}$, we have $\Phi^{\gamma,0,1} \in \Pi_n$, and convergence of $\Phi^{\gamma,0,1}(\mathcal{X}_N) = \gamma\big(\|\mu_N\|\big)$ implies convergence of $\|\mu_N\|$. Hence, the topology induced by Π_n coincides with the topology induced by the polynomials. Using the fact that Π_n is multiplicatively closed, we see that it is convergence determining with the same proof as for the set of polynomials on the space of metric probability measure spaces (see [DGP11, Löh13]), or directly from Le Cam's theorem (see [LC57], [HJ77, Lem. 4.1]). □

2.2. **Measure $\mathbb{R}$-trees and subtree Gromov-weak topology.** In this subsection we define the *subtree Gromov-weak topology.* As "tree-like" metric spaces are 0-hyperbolic, throughout the paper we work with the subspaces

$$\mathbb{H}_n := \big\{\mathcal{X} \in \mathbb{M}_n : \mathcal{X} \text{ is 0-hyperbolic}\big\} \subseteq \mathbb{M}_n, \tag{2.15}$$

and

$$\mathbb{H} := \mathbb{H}_0 \subseteq \mathbb{M}_0, \tag{2.16}$$

where a metric measure space $\mathcal{X} \in \mathbb{M}_n$ is called 0-*hyperbolic* iff

$$r(x_1,x_2)+r(x_3,x_4) \le \max\{r(x_1,x_3)+r(x_2,x_4), r(x_1,x_4)+r(x_2,x_3)\}, \tag{2.17}$$

for all $x_1,x_2,x_3,x_4 \in \mathrm{supp}(\mathcal{X})$. It follows immediately from Theorem 2.5 in [EPW06] that for each $n \in \mathbb{N}$, $(\mathbb{H}_n, d_{\mathrm{pGP}})$ is complete.

Recall that a 0-hyperbolic space is called an $\mathbb{R}$-*tree* if it is connected (see [DMT96] for equivalent definitions and background on $\mathbb{R}$-trees). Given a (rooted) $\mathbb{R}$-tree (T,r,ρ), we denote the unique path between two points $x,y \in T$ by $[x,y]$, and $[x,y[:= [x,y] \setminus \{y\}$. The set of *leaves* of the tree is

$$\mathrm{Lf}(T) := T \setminus \bigcup_{x\in T} [\rho, x[. \tag{2.18}$$

We also use the notation $[\![\underline{v}]\!]$ for the tree spanned by the root ρ and the vector $\underline{v} \in T^n$, i.e.,

$$[\![\underline{v}]\!] := \bigcup_{i=1}^{n} [\rho, v_i]. \tag{2.19}$$

Here and in the following we refer to any $\mathbb{R}$-tree of the form (2.19) as a *finite tree.*

Remark 2.7 (0-hyperbolic spaces are equivalent to $\mathbb{R}$-trees)**.** According to Theorem 3.38 of [Eva07], every 0-hyperbolic space can be isometrically embedded into an $\mathbb{R}$-tree. Since our notion of equivalence of two n-pointed metric measure spaces $\mathcal{X}$ and $\mathcal{X}'$ requires only a (measure and point preserving) isometry between $\mathrm{supp}(\mathcal{X})$ and $\mathrm{supp}(\mathcal{X}')$, this means that every n-pointed 0-hyperbolic metric measure space is equivalent to an n-pointed, measured $\mathbb{R}$-tree. In the following we assume, without loss of generality, that $\mathcal{X} \in \mathbb{H}_n$ is an $\mathbb{R}$-tree, by choosing a connected representative of the equivalence class.

Also note that, given two $\mathbb{R}$-trees (T,r), (T',r') with subsets $A \subseteq T$, $A' \subseteq T'$, and an isometry $\phi\colon A \to A'$, there is a unique extension of ϕ to an isometry between the generated $\mathbb{R}$-trees, $\overline{\phi}\colon [\![A]\!] \to [\![A']\!]$. Indeed, for $v \in [\![A]\!]$ there exist (non-unique) $x,y \in A$ with $v \in [x,y]$, and a unique $w_v \in [\phi(x),\phi(y)]$ with $r(x,v) = r'\big(\phi(x),w_v\big)$. It is straightforward to check that w_v does not depend on the choice of x,y and $\overline{\phi}(v) := w_v$ is an isometry. In particular, for $\mathcal{X} \in \mathbb{H}_n$, the $\mathbb{R}$-tree $[\![\mathrm{supp}(\mathcal{X})]\!]$ is unique up to isometry. $\Diamond$

We now define a topology on $\mathbb{H}_n$ which requires that every subtree generated by a subset of the n distinguished points converges. For that purpose, we define a projection map which sends a list $\underline{u}$ to the sublist indexed by $I = \{i_1,...,i_k\}$ for given $1 \le i_1 < \cdots < i_k \le n$. That is,

$$\pi_I^n\colon \begin{cases} T^n & \to \quad T^k \\ \underline{u} & \mapsto \quad (u_{i_1},\dots,u_{i_k}) \end{cases}. \tag{2.20}$$

The sublist $(u_1,\dots,u_k)$ of $\underline{u} \in T^n$ is simply denoted $\underline{u}^k$. With a slight abuse of notation, we also write

$$\pi_I^n(T,\underline{u},\mu) := \big([\![\pi_I^n(\underline{u})]\!], \pi_I^n(\underline{u}), \mu\big) := \big([\![\pi_I^n(\underline{u})]\!], \pi_I^n(\underline{u}), \mu\restriction_{[\![\pi_I^n(\underline{u})]\!]}\big), \tag{2.21}$$

where the measure μ in the middle expression is tacitly understood to be restricted to the appropriate space, $[\![\pi_I^n(\underline{u})]\!]$.

Definition 2.8 (sGw-topology)**.** *Consider n-pointed measure $\mathbb{R}$-trees $\mathcal{X}$, $\mathcal{X}_1, \mathcal{X}_2, \ldots \in \mathbb{H}_n$. We say that $(\mathcal{X}_N)_{N\in\mathbb{N}}$ converges subtree Gromov-weakly* (sGw) *to $\mathcal{X}$ iff $\mathcal{X}_N \xrightarrow[N\to\infty]{\mathrm{pGw}} \mathcal{X}$ and*

$$\pi_I^n(\mathcal{X}_N) \xrightarrow[N\to\infty]{\mathrm{pGw}} \pi_I^n(\mathcal{X}), \quad \forall I \subseteq \{1, ..., n\}. \tag{2.22}$$

Put

$$\widetilde{\mathbb{H}}_n := \big\{(T, \underline{u}, \mu) \in \mathbb{H}_n : \operatorname{supp}(\mu) \subseteq [\![\underline{u}]\!]\big\} \subseteq \mathbb{H}_n, \tag{2.23}$$

and note that $\widetilde{\mathbb{H}}_n$ consists only of finite trees with at most n leaves.

Remark 2.9 (Related topologies)**.** The sGw-topology is strictly stronger than the pGw-topology. On $\widetilde{\mathbb{H}}_n$, sGw-convergence implies measured Gromov-Hausdorff convergence ([Fuk87]), also known as weighted Gromov-Hausdorff convergence ([EW06, Mie09]). $\Diamond$

Lemma 2.10 (Sufficient condition for sGw-convergence)**.** *Consider random n-pointed measure $\mathbb{R}$-trees $\mathcal{X} = (T, \underline{u}, \mu)$, $\mathcal{X}_N = (T_N, \underline{u}_N, \mu_N) \in \widetilde{\mathbb{H}}_n$, $N \in \mathbb{N}$ (in particular $T_N = [\![\underline{u}_N]\!]$). Assume that $(\mathcal{X}_N)_{N\in\mathbb{N}}$ converges almost surely (a.s.) to $\mathcal{X}$ in the n-pointed Gromov-weak topology, as $N \to \infty$. Furthermore, assume that there is a strictly increasing function $\psi\colon \mathbb{R}_+ \to \mathbb{R}_+$ such that $\psi\big(\|\mu\|\big)$ is integrable and*

$$\mathbb{E}\Big[\psi\Big(\mu_N\big([\![\pi_I^n(\underline{u}_N)]\!]\big)\Big)\Big] \underset{N\to\infty}{\longrightarrow} \mathbb{E}\Big[\psi\Big(\mu\big([\![\pi_I^n(\underline{u})]\!]\big)\Big)\Big], \;\; \forall I \subseteq \{1, ..., n\}. \tag{2.24}$$

Then $(\mathcal{X}_N)_{N\in\mathbb{N}}$ converges also subtree Gromov-weakly to $\mathcal{X}$, a.s., as $N \to \infty$.

To prepare the proof, we state the following:

Remark 2.11 (pGw-convergence yields a tree homomorphism)**.** Consider a sequence of n-pointed measure $\mathbb{R}$-trees $\mathcal{X} = (T, \underline{u}, \mu)$, $\mathcal{X}_1 = (T_1, \underline{u}_1, \mu_1)$, $\mathcal{X}_2 = (T_2, \underline{u}_2, \mu_2), \ldots \in \widetilde{\mathbb{H}}_n$. Assume furthermore that $(\mathcal{X}_N)_{N\in\mathbb{N}}$ converges n-pointed Gromov-weakly to $\mathcal{X}$, a.s., as $N \to \infty$.

For sufficiently large $N \in \mathbb{N}$, we can define a function $f_N\colon T_N \to T$ by sending the root to the root, letting $f_N(u_{N,k}) = u_k$ and $f_N(u_{N,k} \wedge u_{N,l}) = u_k \wedge u_l$, $k, l = 1, ..., n$, and then stretching linearly. Here, as usual, $u \wedge v$ denotes the unique branch point such that $[\rho, u \wedge v] = [\rho, u] \cap [\rho, v]$.

By construction, $\operatorname{dis}(f_N) \underset{N\to\infty}{\longrightarrow} 0$ where

$$\operatorname{dis}(f) := \sup_{x,y\in T} \big|r(x, y) - r'(f(x), f(y))\big| \tag{2.25}$$

denotes the *distortion* of a map $f : (T, r) \to (T', r')$. $\Diamond$

Proof of Lemma 2.10. Assume that N is large enough, such that the function $f_N\colon T_N \to T$ from Remark 2.11 is a tree homomorphism with $f_N(u_{N,k}) = u_k$, $k \in \{1, 2, \ldots, n\}$ and such that $\operatorname{dis}(f_N) \underset{N\to\infty}{\longrightarrow} 0$. We can therefore choose a metric d on $T_N \uplus T$ extending r_N and r such that $d(x, f_N(x)) \to 0$ for all $x \in T_N$ (compare, for example, [BBI01, Corollary 7.3.28]).

Thus $d_{\mathrm{Pr}}((f_N)_*\mu_N, \mu_N) \le \sup_x d(x, f_N(x)) \to 0$, as $N \to \infty$, and we obtain that

$$d_{\mathrm{Pr}}\big((f_N)_*\mu_N, \mu\big) \le d_{\mathrm{Pr}}\big((f_N)_*\mu_N, \mu_N\big) + d_{\mathrm{Pr}}\big(\mu_N, \mu\big) \underset{N\to\infty}{\longrightarrow} 0. \tag{2.26}$$

Fix now $I \subseteq \{1, \ldots, n\}$ and define the subtree

$$S := [\![\pi_I^n(\underline{u})]\!] \subseteq T. \tag{2.27}$$

Because S is closed in T, we have $\limsup_{N\to\infty}(f_N)_*\mu_N(S) \le \mu(S)$ by the Portmanteau theorem (see Theorem 2.1 in [Bil99]). Because ψ is increasing, this implies

$$\limsup_{N\to\infty} \psi\big((f_N)_*\mu_N(S)\big) \le \psi\big(\mu(S)\big). \tag{2.28}$$

By assumption (2.24),

$$\mathbb{E}\big[\psi\big((f_N)_*\mu_N(S)\big)\big] = \mathbb{E}\big[\psi\big(\mu_N([\![\pi_I^n(\underline{u}_N)]\!])\big)\big] \underset{N\to\infty}{\longrightarrow} \mathbb{E}\big[\psi\big(\mu(S)\big)\big]. \tag{2.29}$$

(2.29) and (2.28) together yield $\psi\big((f_N)_*\mu_N(S)\big) \underset{N\to\infty}{\longrightarrow} \psi\big(\mu(S)\big)$, almost surely. Because ψ is strictly increasing, also $(f_N)_*\mu_N(S) \to \mu(S)$. Using once more the Portmanteau theorem and closedness of S, we obtain that

$$(f_N)_*\mu_N{\restriction_S} \underset{N\to\infty}{\Longrightarrow} \mu{\restriction_S}. \tag{2.30}$$

The inequality

$$\begin{aligned} &d_{\mathrm{pGw}}\big(\pi_I^n(\llbracket \underline{u}_N\rrbracket, \underline{u}_N, \mu_N),\, \pi_I^n(\llbracket \underline{u}\rrbracket, \underline{u}, \mu)\big) \\ &\le d_{\mathrm{Pr}}\big(\mu_N{\restriction_{\llbracket \pi_I^n(\underline{u}_N)\rrbracket}},\, (f_N)_*\mu_N{\restriction_S}\big) + d_{\mathrm{Pr}}\big((f_N)_*\mu_N{\restriction_S},\, \mu{\restriction_S}\big), \end{aligned} \tag{2.31}$$

then gives the sGw-convergence. □

As for the pGw-topology, we define an associated set of test functions $\tilde{\Phi} : \mathbb{H}_n \to \mathbb{R}$ by

$$\tilde{\Phi}(T, \underline{u}, \mu) := \prod_{I\subseteq\{1,\ldots,n\}} \Phi_I\big(\pi_I^n(T, \underline{u}, \mu)\big), \tag{2.32}$$

where the Φ_I are polynomials on $\mathbb{H}_{\#I}$. Obviously, this class of test functions induces the sGw-topology on $\widetilde{\mathbb{H}}_n$, and together with the polynomials on $\mathbb{H}_n$, the sGw-topology on $\mathbb{H}_n$. We also define

$$\widetilde{\Pi}_n := \Big\{ \prod_{I\subseteq\{1,\ldots,n\}} \Phi_I^{\gamma,m,\varphi} \circ \pi_I^n \,:\, \Phi_I^{\gamma,m,\varphi} \in \Pi_{\#I} \Big\}. \tag{2.33}$$

2.3. **The LWV-topology.** In this subsection we give the definition of *bi-measure* $\mathbb{R}$*-trees* and equip the space of equivalence classes of bi-measure $\mathbb{R}$-trees with the *leaf-sampling weak vague topology*, in the following referred to as the *LWV-topology*.

Given a rooted measure $\mathbb{R}$-tree (T, μ), denote by

$$\mathrm{Sk}_\mu(T) := \bigcup_{v\in\mathrm{supp}(\mu)} [\rho, v[\,\cup \big\{v \in T :\ \mu\big(\{v\}\big) > 0\big\} \tag{2.34}$$

the μ*-skeleton* of (T, μ), and by

$$\mathrm{Lf}_\mu(T) := \llbracket \mathrm{supp}(\mu)\rrbracket \setminus \mathrm{Sk}_\mu(T) \tag{2.35}$$

the set of μ*-leaves* of (T, μ).

We call (T, μ, ν) a (rooted) *bi-measure* $\mathbb{R}$*-tree* if (T, μ) is a (rooted) measure $\mathbb{R}$-tree and ν is a (σ-finite) measure on T which satisfies the following two conditions:

(i) $\nu\big([\rho, u]\big)$ is μ-a.s. finite for $u \in T$,
(ii) ν vanishes on the set of μ-leaves, i.e., $\nu(\mathrm{Lf}_\mu(T)) = 0$.

Note that (i) implies that ν is finite on subtrees of T with a finite number of leaves sampled with μ, a.s, and that $\nu{\restriction_{\mathrm{Sk}_\mu(T)}}$ is σ-finite (because our definition of measure $\mathbb{R}$-trees includes separability of $\mathrm{supp}(\mu)$). In many interesting cases, however, ν is not locally finite.

Definition 2.12 (The spaces $\mathbb{H}^{f,\sigma}$ and $\mathbb{H}^{K,\sigma}$)**.** *Two bi-measure* $\mathbb{R}$*-trees* (T, μ, ν) *and* (T', μ', ν') *are called* equivalent *if there exists an isometry* $\phi\colon \llbracket \mathrm{supp}(\mu)\rrbracket \to T'$ *preserving the root and* μ *and preserving* ν *on the* μ*-skeleton, i.e.,* $\phi_*(\mu) = \mu'$ *and* $\phi_*(\nu{\restriction_{\mathrm{Sk}_\mu(T)}}) = \nu'{\restriction_{\mathrm{Sk}_{\mu'}(T')}}$*. In particular,* (T, μ, ν) *is equivalent to* $(T, \mu, \nu{\restriction_{\mathrm{Sk}_\mu(T)}})$*.*

We denote by $\mathbb{H}^{f,\sigma}$ *the space of equivalence classes of (rooted) bi-measure* $\mathbb{R}$*-trees, and by* $\mathbb{H}^{K,\sigma} := \big\{(T, \mu, \nu) \in \mathbb{H}^{f,\sigma} \;\big|\; \|\mu\| \le K\big\}$*,* $K > 0$*, the subspace where the total mass of the sampling measure is bounded by* K*.*

Similar to the distance matrix distribution $\boldsymbol{\upsilon}^{(T,\mu)}$ introduced in (2.8), which characterizes n-pointed measure $\mathbb{R}$-trees and is used to define the pGw-topology, we want to characterize bi-measure $\mathbb{R}$-trees by the so-called *subtree-vector-distribution.* To introduce this, consider for a given bi-measure $\mathbb{R}$-tree (T,μ,ν) the function

$$\tau_{(T,\mu,\nu)}\colon \begin{cases} \bigcup_{n\in\mathbb{N}} T^n & \to \bigcup_{n\in\mathbb{N}} \mathbb{H}_n, \\ (u_1,u_2,\dots,u_n) & \mapsto \big([\![u_1,\dots,u_n]\!],(u_1,\dots,u_n),\nu\big), \end{cases} \tag{2.36}$$

which sends a vector of n points in T to the n-pointed $\mathbb{R}$-tree spanned by these points and equipped with ν, which we tacitly understand to be restricted to the appropriate space, i.e. $[\![u_1,...,u_n]\!]$. We also define the function

$$\varsigma_{(T,\mu,\nu)}\colon \begin{cases} T^{\mathbb{N}} & \to \prod_{n\in\mathbb{N}} \mathbb{H}_n, \\ \underline{u} & \mapsto \big(\tau_{(T,\mu,\nu)}(u_1),\, \tau_{(T,\mu,\nu)}(u_1,u_2),\,\dots\big), \end{cases} \tag{2.37}$$

which sends a sequence of points to the sequence of pointed measure $\mathbb{R}$-trees spanned and pointed by the first 1, 2, etc. points and each of these is equipped with the appropriate restriction of ν. Note that $\tau_{(T,\mu,\nu)}$ does not depend on the measure μ and is in general not continuous.

Lemma 2.13 (Measurability)**.** *Equip $\mathbb{H}_n$ with the n-pointed Gromov-weak topology, and $\prod_{n\in\mathbb{N}}\mathbb{H}_n$ with the product topology. Then the function $\varsigma_{\mathcal{X}}$ is measurable for all $\mathcal{X}\in\mathbb{H}^{f,\sigma}$.*

Proof. It is enough to show that $\tau_{\mathcal{X}}$ is measurable on T^n for each $n\in\mathbb{N}$. Fix therefore $n\in\mathbb{N}$.

Since $\mathbb{H}_n$ is separable (Proposition 2.6), and the space of all polynomials induces the n-pointed Gromov-weak topology on $\mathbb{H}_n$, it is enough to show that $\Phi\circ\tau_{\mathcal{X}}$ is measurable for every polynomial Φ (compare (2.10)). As for each $m\in\mathbb{N}$, $\varphi\in\mathcal{C}_b(\mathbb{R}_+^{\binom{m+n+1}{2}})$ and $\mathcal{X}=(T,\mu,\nu)$,

$$\Phi^{m,\varphi}\circ\tau_{\mathcal{X}}(\underline{u}) = \int \nu^{\otimes m}(\mathrm{d}\underline{v})\,\mathbb{1}_{\{v_1,\dots,v_m\in[\![\underline{u}]\!]\}}\,\varphi\big(R^T(\underline{u},\underline{v})\big), \tag{2.38}$$

this follows from joint measurability of $(\underline{u},\underline{v})\mapsto \mathbb{1}_{\{v_1,\dots,v_m\in[\![\underline{u}]\!]\}}(\varphi\circ R^T)(\underline{u},\underline{v})$. □

We are now in a position to define the *subtree vector distribution*, $\varpi^{\mathcal{X}}$, of a bi-measure $\mathbb{R}$-tree $\mathcal{X}=(T,\mu,\nu)$ as

$$\varpi^{\mathcal{X}} := \|\mu\|\cdot(\varsigma_{\mathcal{X}})_*\big((\mu^\circ)^{\otimes\mathbb{N}}\big)\in\mathcal{M}_f\big(\prod_{n\in\mathbb{N}}\mathbb{H}_n\big). \tag{2.39}$$

Definition 2.14 (LWV-topology)**.** *We say that a sequence $(\mathcal{X}_N)_{N\in\mathbb{N}}$ converges to $\mathcal{X}$ in $\mathbb{H}^{f,\sigma}$ in the leaf-sampling weak vague topology (*LWV*-topology) if the corresponding subtree vector distributions converge, i.e.,*

$$\varpi^{\mathcal{X}_N} \underset{N\to\infty}{\Longrightarrow} \varpi^{\mathcal{X}}, \tag{2.40}$$

where convergence is weak convergence of finite measures on $\prod_n(\mathbb{H}_n,\mathrm{pGw})$.

Remark 2.15. Obviously, $\mathbb{H}^{K,\sigma}$ is closed in $\mathbb{H}^{f,\sigma}$ with LWV-topology, $\mathbb{H}^{f,\sigma}=\bigcup_{K\in\mathbb{N}}\mathbb{H}^{K,\sigma}$, and for every compact set $\mathbb{K}\subseteq\mathbb{H}^{f,\sigma}$ there exists $K\in\mathbb{N}$ with $\mathbb{K}\subseteq\mathbb{H}^{K,\sigma}$. ◊

Remark 2.16 (Relation with Gromov-weak topology)**.**

(i) LWV-convergence of $(T_N,\mu_N,\nu_N)_{N\in\mathbb{N}}$ implies Gromov-weak convergence of $(T_N,\mu_N)_{N\in\mathbb{N}}$.

(ii) Gromov-weak convergence of $(T_N,\mu_N)_{N\in\mathbb{N}}$ does not imply LWV-convergence of the bi-measure trees $(T_N,\mu_N,\mu_N)_{N\in\mathbb{N}}$ (compare Example 2.21). ◊

Recall from Definition 2.4 and Definition 2.8 the n-pointed Gromov-weak topology (pGw) and the subtree Gromov-weak topology (sGw), respectively. Let $(U_{N,k})_{k\in\mathbb{N}}$ be an i.i.d. sequence of μ_N°-distributed random variables, and $\underline{U}_N^n := (U_{N,1},\dots,U_{N,n})$. The definition of LWV-convergence requires, in addition to convergence of $\|\mu_N\|$, the *joint* convergence in law with respect to the pGw-topology of $\tau_{\mathcal{X}_N}(\underline{U}_N^n)$, $n \in \mathbb{N}$. The next proposition shows that we can, on one hand, weaken this requirement to individual convergence of all $\tau_{\mathcal{X}_N}(\underline{U}_N^n)$, and, on the other hand, strengthen it to require convergence in law with respect to the sGw-topology.

Proposition 2.17 (Characterization of LWV-convergence)**.** *Consider a sequence of bi-measure $\mathbb{R}$-trees $\mathcal{X}_N = (T_N, \mu_N, \nu_N) \in \mathbb{H}^{f,\sigma}$ and another bi-measure $\mathbb{R}$-tree $\mathcal{X} \in \mathbb{H}^{f,\sigma}$ such that $\|\mu_N\| \to \|\mu\|$, as $N \to \infty$. The three following statements are equivalent:*

(i) $\mathcal{X}_N \xrightarrow{\text{LWV}} \mathcal{X}$*, as* $N \to \infty$.

(ii) *For all* $n \in \mathbb{N}$,

$$(\tau_{\mathcal{X}_N})_* \big(\mu_N^\circ\big)^{\otimes n} \underset{N\to\infty}{\overset{\text{pGw}}{\Longrightarrow}} (\tau_{\mathcal{X}})_* \big(\mu^\circ\big)^{\otimes n}. \tag{2.41}$$

(iii) *Equipping* $\prod_{n\in\mathbb{N}} \mathbb{H}_n$ *with the product topology* $\prod(\text{sGw})$,

$$\big(\varsigma_{\mathcal{X}_N}\big)_* \big(\mu_N^\circ\big)^{\otimes \mathbb{N}} \underset{N\to\infty}{\overset{\prod(\text{sGw})}{\Longrightarrow}} \big(\varsigma_{\mathcal{X}}\big)_* \big(\mu^\circ\big)^{\otimes \mathbb{N}}. \tag{2.42}$$

Proof. First remark that (iii) $\Rightarrow$ (i) $\Rightarrow$ (ii) is straightforward.

We prove that (ii) implies (iii). Fix therefore $n \in \mathbb{N}$. By Skorohod's representation theorem (Theorem 6.7 in [Bil99]), there exists a list $\underline{U}^n = (U_1, \ldots, U_n)$ of n i.i.d. random variables with common distribution μ° and $\underline{U}_N^n = (U_{N,1}, \ldots, U_{N,n})$ i.i.d. random variables with distribution μ_N° such that

$$\tau_{\mathcal{X}_N}(\underline{U}_N^n) \underset{N\to\infty}{\overset{\text{pGw}}{\longrightarrow}} \tau_{\mathcal{X}}(\underline{U}^n), \quad \text{almost surely.} \tag{2.43}$$

In order to obtain sGw-convergence, by Lemma 2.10, it is sufficient to prove for all $I \subseteq \{1, \ldots, n\}$ that $\nu_N\big([\![\pi_I^n(\underline{U}_N^n)]\!]\big)$ converges weakly (as $\mathbb{R}_+$-valued random variable) to $\nu\big([\![\pi_I^n(\underline{U}^n)]\!]\big)$. Because $\pi_I^n(\underline{U}_N^n)$ has the same distribution as $\underline{U}_N^{\#I}$, and similarly for $\underline{U}^n$ instead of $\underline{U}_N^n$, this follows from (2.41) for $n = \#I$, where we use that the total mass of an n-pointed measure $\mathbb{R}$-tree is continuous in the pGw-topology. Finally, we conclude from Lemma 2.10 that

$$\tau_{\mathcal{X}_N}(\underline{U}_N^n) \xrightarrow{\text{sGw}} \tau_{\mathcal{X}}(\underline{U}^n), \text{ as } N \to \infty, \text{ almost surely.} \tag{2.44}$$

In particular, the one-dimensional marginals of $(\varsigma_{\mathcal{X}_N})_*(\mu_N^\circ)^{\otimes \mathbb{N}_0}$ converge as measures on $(\mathbb{H}_n, \text{sGw})$. In order to obtain convergence of laws on the product space, we have to show convergence of finite-dimensional marginals. This comes directly from the definition of sGw-convergence. $\square$

We are now in a position to show that the subtree vector distribution characterizes bi-measure $\mathbb{R}$-trees uniquely.

Proposition 2.18 (Reconstruction theorem for $\mathbb{H}^{f,\sigma}$)**.** *If $\mathcal{X}, \mathcal{X}' \in \mathbb{H}^{f,\sigma}$ are such that $\varpi^{\mathcal{X}} = \varpi^{\mathcal{X}'}$, then $\mathcal{X} = \mathcal{X}'$.*

Proof. Let $\mathcal{X} = (T, \mu, \nu), \mathcal{X}' = (T', \mu', \nu') \in \mathbb{H}^{f,\sigma}$ with $\varpi^{\mathcal{X}} = \varpi^{\mathcal{X}'}$. It follows immediately that $\|\mu\| = \|\mu'\|$. Assume w.l.o.g. that $\|\mu\| = \|\mu'\| = 1$.

We will first adapt Vershik's proof of Gromov's reconstruction theorem for metric measure spaces to show that $(T, \mu) = (T', \mu')$ (compare [Gro99, $3\frac{1}{2}$.7]). Recall that a sequence $\underline{u} = (u_n)_{n\in\mathbb{N}}$ in T is called *μ-uniformly distributed* if

$$\tfrac{1}{n} \sum_{i=1}^n \delta_{u_i} \underset{n\to\infty}{\Longrightarrow} \mu, \tag{2.45}$$

and note that, due to separability of T, $\mu^{\otimes \mathbb{N}}$-almost every sequence is μ-uniformly distributed (see, for example, [Dud02, Theorem 11.4.1]).

Of course, the corresponding statement is also true for μ' instead of μ, and as $\varpi^{\mathcal{X}} = \varpi^{\mathcal{X}'}$, we can find a μ-uniformly distributed sequence $\underline{u} = (u_n)_{n\in\mathbb{N}}$ in T, and a μ'-uniformly distributed sequence $\underline{u}' = (u'_n)_{n\in\mathbb{N}}$ in T' with $\varsigma_{\mathcal{X}}(\underline{u}) = \varsigma_{\mathcal{X}'}(\underline{u}')$.

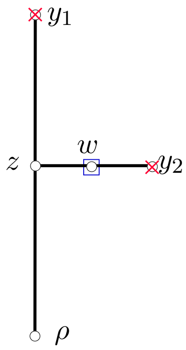


FIGURE 1

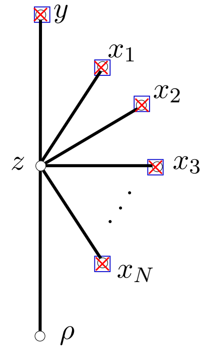


FIGURE 2

The crosses $\times$ are μ-masses and the squares $\square$ are ν-masses.

Put $u_0 := \rho$, $u'_0 := \rho'$. Then $f(u_k) := u'_k$, for all $k \in \mathbb{N}_0$, defines a root-preserving isometry from $\{u_0, u_1, \ldots\}$ onto $\{u'_0, u'_1, \ldots\}$, which can be extended to an isometry (still denoted by f) from $[\![\mathrm{supp}(\mu)]\!]$ onto $[\![\mathrm{supp}(\mu')]\!]$ (see Remark 2.7). Because the sequences are uniformly distributed and f_* is continuous,

$$
\begin{aligned}
f_*(\mu) &= f_*\Big(\lim_{n\to\infty} \tfrac{1}{n} \sum_{i=1}^{n} \delta_{u_i}\Big) = \lim_{n\to\infty} f_*\Big(\tfrac{1}{n}\sum_{i=1}^{n} \delta_{u_i}\Big) \\
&= \lim_{n\to\infty} \tfrac{1}{n} \sum_{i=1}^{n} \delta_{u'_i} = \mu'.
\end{aligned} \tag{2.46}
$$

We still need to show that $f_*(\nu') = \nu$ (on the μ-skeleton), or equivalently, $f_*(\nu')(S) = \nu(S)$ for all finite trees $S \subseteq \mathrm{Sk}_\mu(T)$. By definition of $\mathrm{Sk}_\mu(T)$ and the fact that $(u_n)_{n\in\mathbb{N}}$ is uniformly distributed, we have $S \subseteq [\![\underline{u}^n]\!]$ for sufficiently large n. Because $([\![\underline{u}^n]\!], \underline{u}^n, \nu)$ and $([\![\underline{u}'^n]\!], \underline{u}'^n, \nu')$ are equivalent as n-pointed metric measure spaces, $f_*(\nu')\restriction_{[\![\underline{u}^n]\!]} = \nu\restriction_{[\![\underline{u}^n]\!]}$. $\square$

We can now immediately conclude that $\mathbb{H}^{f,\sigma}$ is separable and metrizable. We are not able to come up, however, with a complete metric. "Polishness" of the state space will not be used throughout the paper.

Corollary 2.19 (Separability & metrizability)**.** *The space $\mathbb{H}^{f,\sigma}$ equipped with the* LWV*-topology is separable and metrizable.*

Proof. As the map which sends a bi-measure $\mathbb{R}$-tree to its subtree vector distribution is injective, we can identify $\mathbb{H}^{f,\sigma}$ with a subspace of $\mathcal{M}_f(\prod_{n\in\mathbb{N}} \mathbb{H}_n)$. $\mathbb{H}_n$ is separable, metrizable according to Proposition 2.6, hence the same holds for the countable product and the space of finite measures on it (with weak topology). $\square$

It is important to note that μ and ν play different rôles in the LWV-topology, even if ν happens to be finite and μ is supported on the skeleton. While the convergence is *weak* with respect to μ, it is *vague* with respect to ν in the sense that the total ν-mass is not preserved under convergence, but mass may get lost in the limit. We give two examples of this phenomenon.

Example 2.20. Consider the (finite) $\mathbb{R}$-tree shown in Figure 1 and define the probability measures $\mu_N := (1 - \frac{1}{N})\delta_{y_1} + \frac{1}{N}\delta_{y_2}$. Then (T, μ_N) converges Gromov-weakly to $\big(\{\rho, y_1\}, \delta_{y_1}\big)$. We endow (T, μ_N) with a constant measure $\nu := \delta_w$, then (T, μ_N, δ_w) converges in the LWV-topology to $(\{\rho, y_1\}, \delta_{y_1}, 0)$. $\Diamond$

Example 2.21 (Figure 2)**.** We define a sequence of $\mathbb{R}$-trees

$$T_N := \big\{\rho, z, y, x_1, x_2, \ldots, x_N\big\}$$

shown in Figure 2 where $r_N(\rho, z) = r_N(z, y) = 1$ and $r(z, x_i) = \frac{1}{N}$, for all $i = 1, \ldots, N$. We define a probability measure μ_N on the leaves of T_N by $\mu_N = \lambda\delta_y + (1-\lambda)\sum_i \frac{1}{N}\delta_{x_i}$, then (T_N, μ_N) converges Gromov-weakly to $(\{\rho, z, y\}, \lambda\delta_y + (1-\lambda)\delta_z)$. If we endow this measure $\mathbb{R}$-treewith the measure $\nu_N = \mu_N$, then (T_N, μ_N, ν_N) converges in the LWV-topology to $(\{\rho, z, y\}, \lambda\delta_y + (1-\lambda)\delta_z, \lambda\delta_y)$.
$\Diamond$

2.4. **Convergence determining classes for the LWV-topology.** In this subsection, we introduce important classes of test functions and use them to obtain several convergence results. Namely, we consider functions $\Psi = \Psi^{\gamma,n,\Phi} \colon \mathbb{H}^{f,\sigma} \to \mathbb{R}$ of the form

$$\Psi(\mathcal{X}) := \Psi^{\gamma,n,\Phi}(\mathcal{X}) := \gamma(\|\mu\|) \cdot \int_{T^n} \mu^{\otimes n}(\mathrm{d}\underline{u})\, \Phi\big(\tau_{\mathcal{X}}(\underline{u})\big), \tag{2.47}$$

where $\gamma \in \mathcal{C}_b(\mathbb{R}_+)$ and $\Phi \in \mathcal{C}_b(\mathbb{H}_n)$.

Recall Π_n and $\widetilde{\Pi}_n$ from (2.12) and (2.33). As we will see later, the following subspaces of test functions are helpful in characterizing LWV-convergence. Put

$$\mathcal{F} := \big\{\Psi^{1,n,\Phi} \mid \Phi \in \Pi_n\big\}, \tag{2.48}$$

and

$$\tilde{\mathcal{F}}^1 := \big\{\Psi^{1,n,\tilde{\Phi}} \mid \tilde{\Phi} \in \widetilde{\Pi}_n\big\}, \tag{2.49}$$

and

$$\tilde{\mathcal{F}} := \big\{\Psi^{\gamma,n,\tilde{\Phi}} \mid \tilde{\Phi} \in \widetilde{\Pi}_n,\ \lim_{x\to\infty} x^k\gamma(x) = 0\ \forall k \in \mathbb{N}\big\}. \tag{2.50}$$

Lemma 2.22 (LWV-convergence via test functions)**.** *Both $\mathcal{F}$ and $\tilde{\mathcal{F}}$ induce the* LWV*-topology, i.e., for a sequence of bi-measure $\mathbb{R}$-trees $\mathcal{X}_N \in \mathbb{H}^{f,\sigma}$ and another bi-measure $\mathbb{R}$-tree $\mathcal{X} \in \mathbb{H}^{f,\sigma}$, the following statements are equivalent.*

(i) $\mathcal{X}_N \xrightarrow{\text{LWV}} \mathcal{X}$*, as $N \to \infty$.*
(ii) $\Psi(\mathcal{X}_N) \to \Psi(\mathcal{X})$*, as $N \to \infty$, for all $\Psi \in \mathcal{F}$.*
(iii) $\tilde{\Psi}(\mathcal{X}_N) \to \tilde{\Psi}(\mathcal{X})$*, as $N \to \infty$, for all $\tilde{\Psi} \in \tilde{\mathcal{F}}$.*

Proof. The equivalence of (i) and (ii) is clear, as by Proposition 2.17, LWV-convergence is equivalent to the convergence of $\|\mu_N\| \to \|\mu\|$ together with $\big\langle (\tau_{\mathcal{X}_N})_* (\mu_N^\circ)^{\otimes n}, f\big\rangle \to \big\langle (\tau_{\mathcal{X}})_* (\mu^\circ)^{\otimes n}, f\big\rangle$, as $N \to \infty$, for all $n \in \mathbb{N}$ and for a class of functions f which determine the n-pointed Gromov-weak convergence. Moreover, by Proposition 2.6, Π_n is such a convergence determining class. As $\Psi(\mathcal{X}_N) = \big\langle (\tau_{\mathcal{X}_N})_* (\mu_N)^{\otimes n}, \Phi\big\rangle$, the claim follows.

By Proposition 2.17, $\tilde{\mathcal{F}}$ contains only functions which are continuous with respect to the LWV-topology, and thus (i) clearly implies (iii). To see that (iii) implies (ii), note that for $\gamma(x) := e^{-x}$, convergence of $\Psi^{\gamma,0,1}(\mathcal{X}_N) = \gamma\big(\|\mu_N\|\big)$ implies convergence of $\|\mu_N\|$. Hence convergence of $\tilde{\Psi}(\mathcal{X}_N)$, for all $\tilde{\Psi} \in \tilde{\mathcal{F}}$, implies convergence of $\Psi(\mathcal{X}_N)$, for all $\Psi \in \mathcal{F}$. $\square$

Proposition 2.23 (Convergence determining classes)**.** *The following hold:*

(i) *The class of test functions $\tilde{\mathcal{F}}$ is convergence determining on $\mathbb{H}^{f,\sigma}$.*
(ii) *The class of test functions $\tilde{\mathcal{F}}^1$ is convergence determining on $\mathbb{H}^{K,\sigma}$ for all $K > 0$.*

Proof. We apply Theorem 6 from [BK10], a slight extension of Le Cam's theorem (see [LC57]) in the separable, metrizable case: if a set of bounded real-valued functions is multiplicatively closed and induces a separable, metrizable topology, then it is a convergence determining class with respect to this topology. By Lemma 2.22, $\tilde{\mathcal{F}}$ induces the LWV-topology, which is separable, metrizable by Corollary 2.19. We therefore need to verify that if $\tilde{\Psi}_1$, $\tilde{\Psi}_2 \in \tilde{\mathcal{F}}$, then $\tilde{\Psi}_1 \cdot \tilde{\Psi}_2 \in \tilde{\mathcal{F}}$.

Let $\tilde{\Psi}_i = \Psi^{\gamma_i,n_i,\tilde{\Phi}_i}$ for some $n_i \in \mathbb{N}_0$, $\gamma_i \in \mathcal{C}_b(\mathbb{R}_+)$ with $\lim_{x\to\infty} x^k\gamma_i(x) = 0$, for all $k \in \mathbb{N}$, and $\tilde{\Phi}_i \in \widetilde{\Pi}_n$, $i = 1, 2$. Then

$$\begin{aligned}
&\Psi^{\gamma_1,n_1,\tilde{\Phi}_1} \cdot \Psi^{\gamma_2,n_2,\tilde{\Phi}_2}(\mathcal{X}) \\
&= \big(\gamma_1\gamma_2\big)(\|\mu\|) \cdot \int_{T^{n_1+n_2}} \mu^{\otimes(n_1+n_2)}(\mathrm{d}\underline{u}_1, \mathrm{d}\underline{u}_2)\, \tilde{\Phi}_1\big([\![\underline{u}_1]\!], \underline{u}_1, \nu\big)\tilde{\Phi}_2\big([\![\underline{u}_2]\!], \underline{u}_2, \nu\big).
\end{aligned} \tag{2.51}$$

For $\underline{u} = (\underline{u}_1, \underline{u}_2)$, let $\tilde{\Phi}\left([\![\underline{u}]\!], \underline{u}, \nu\right) := \tilde{\Phi}_1\left([\![\underline{u}_1]\!], \underline{u}_1, \nu\right) \cdot \tilde{\Phi}_2\left([\![\underline{u}_2]\!], \underline{u}_2, \nu\right)$. As $\underline{u}_1$ and $\underline{u}_2$ are sublists of $\underline{u}$, $\tilde{\Phi} \in \widetilde{\Pi}_n$ and therefore $\tilde{\Psi}_1 \cdot \tilde{\Psi}_2 \in \tilde{\mathcal{F}}$.

To get the second statement in the same way, note that functions $\Psi^{1,n,\tilde{\Phi}} \in \tilde{\mathcal{F}}^1$ are bounded on $\mathbb{H}^{K,\sigma}$. □

An important fact about the LWV-topology is that Gromov-weak convergence of measure $\mathbb{R}$-trees implies LWV-convergence if the trees are additionally equipped with their respective length measures (see Example 2.24 for a definition of length measure and Proposition 2.25 for the statement). We obtain the same also for a slightly more general class of measures. Given a family $(T_i, \mu_i)_{i\in I}$ of measure $\mathbb{R}$-trees, we say that a family $(\nu_i)_{i\in I}$ of measures on respective T_i *depends continuously on the distances* if, for all $n \in \mathbb{N}$, there exists a continuous mapping $F_n : \mathbb{R}^{\binom{n+1}{2}} \to \mathbb{H}_n$, where $\mathbb{H}_n$ is endowed with the pGw-topology, such that

$$\big([\![\underline{u}]\!], \underline{u}, \nu_i\big) = F_n\big(R^{T_i}(\underline{u})\big), \quad \forall \underline{u} \in T_i^n,\ \forall i \in I. \tag{2.52}$$

Example 2.24 (Length measure)**.** The length measure, λ_T, on a separable 0-hyperbolic and connected metric space T generalizes the Lebesgue measure on $\mathbb{R}$ in an obvious way (compare [EPW06]). Recall the set of leaves of T from (2.18). The length measure can be defined by the following two requirements:

$$\forall x, y \in T : \lambda_T([x, y]) = r(x, y) \ \text{ and } \ \lambda_T\big(\mathrm{Lf}(T)\big) = 0. \tag{2.53}$$

Obviously, the family of length measures $(\lambda_T)_{T\in\{\mathbb{R}\text{-trees}\}}$ depends continuously on the distances. The same is true if we replace λ_T by $\nu_T = f_T \cdot \lambda_T$, where f_T is a density that depends only on the height, i.e., $f_T(v) := h\big(r(\rho, v)\big)$ for a bounded measurable function h (which does not depend on T). ◊

We can relax the continuity of the F_n, $n \in \mathbb{N}$, a little. Let $(T, \mu) \in \mathbb{H}$. We say that a family $(\nu_i)_{i\in I}$ as above *depends $\boldsymbol{v}^{(T,\mu)}$-almost continuously on the distances* if it satisfies (2.52) with functions F_n that are not necessarily continuous, but where the set of discontinuity points is a null set with respect to the distance matrix distribution induced by (T, μ), i.e. $(R^T)_*\mu^{\otimes n}(\mathrm{Discont}(F_n)) = 0$.

Proposition 2.25 (LWV-convergence from Gromov-weak convergence)**.** *Consider a sequence $(\mathcal{X}_N)_{N\in\mathbb{N}} := (T_N, \mu_N, \nu_N)_{N\in\mathbb{N}}$, $\mathcal{X}_\infty := (T_\infty, \mu_\infty, \nu_\infty)$ in $\mathbb{H}^{f,\sigma}$ such that the measures $\nu_\infty, \nu_1, \nu_2, \ldots$ depend $\boldsymbol{v}^{(T_\infty,\mu_\infty)}$-almost continuously on the distances. If $(T_N, \mu_N) \underset{N\to\infty}{\overset{\mathrm{Gw}}{\longrightarrow}} (T_\infty, \mu_\infty)$, then*

$$(T_N, \mu_N, \nu_N) \underset{N\to\infty}{\overset{\mathrm{LWV}}{\longrightarrow}} (T_\infty, \mu_\infty, \nu_\infty). \tag{2.54}$$

In particular, the embedding defined by

$$\begin{array}{rcl} \mathbb{H} & \to & \mathbb{H}^{f,\sigma}, \\ (T, \mu) & \mapsto & (T, \mu, \lambda_T), \end{array} \tag{2.55}$$

where λ_T is the length measure, is a homeomorphism onto its image.

Proof. Given $n \in \mathbb{N}$, fix a function $F_n : \mathbb{R}_+^{\binom{n+1}{2}} \to \mathbb{H}_n$ as in (2.52), such that the set of discontinuity points of F_n is a zero set with respect to $(R^{T_\infty})_*(\mu_\infty^{\otimes n})$. For $N \in \mathbb{N} \cup \{\infty\}$, let $\underline{U}_N$ be a random

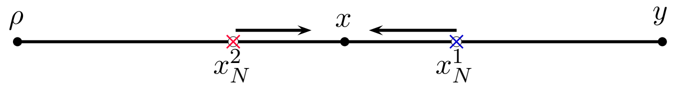


FIGURE 3. The tree T_N with the two sequences x_N^1 and x_N^2 that converge to x.

vector in T_N^n with distribution $(\mu_N^\circ)^{\otimes n}$. Then the assumed Gromov-weak convergence means that $\|\mu_N\| \to \|\mu_\infty\|$ and

$$R^{T_N}(\underline{U}_N) \underset{N\to\infty}{\overset{\mathcal{L}}{\Longrightarrow}} R^{T_\infty}(\underline{U}_\infty), \tag{2.56}$$

where $\overset{\mathcal{L}}{\Rightarrow}$ denotes convergence in law. By the continuous mapping theorem (see Theorem 5.1 in [Bil99]), we obtain

$$\big([\![\underline{U}_N]\!], \underline{U}_N, \nu_N\big) = F_n\big(R^{T_N}(\underline{U}_N)\big) \underset{N\to\infty}{\overset{\mathcal{L}}{\Longrightarrow}} F_n\big(R^{T_\infty}(\underline{U}_\infty)\big) = \big([\![\underline{U}_\infty]\!], \underline{U}_\infty, \nu_\infty\big). \tag{2.57}$$

Using that $\big([\![\underline{U}_N]\!], \underline{U}_N, \nu_N\big)$ has law $(\tau_{\mathcal{X}_N})_*\big((\mu_N^\circ)^{\otimes n}\big)$ for $N \in \mathbb{N}\cup\{\infty\}$, the claimed LWV-convergence $\mathcal{X}_N \xrightarrow{\text{LWV}} \mathcal{X}_\infty$ now follows from Proposition 2.17. That (2.55) defines a homeomorphism onto its image is now obvious, because the length measure depends continuously on the distances (see Example 2.24). $\square$

Corollary 2.26 (Sampling measure perturbation)**.** *Consider two sequences of bi-measure $\mathbb{R}$-trees $\mathcal{X}_N^i := (T_N, \mu_N^i, \nu_N)$, $i = 1,2$ that differ by their sampling measures μ_N^1 and μ_N^2. Assume that $\mathcal{X}_N^1 \underset{N\to\infty}{\xrightarrow{\text{LWV}}} \mathcal{X}$, and that the pruning measures $(\nu_N)_{N\in\mathbb{N}}$ depend $\boldsymbol{v}^{(T,\mu)}$-almost continuously on the distances. If $d_{\text{Pr}}(\mu_N^1, \mu_N^2) \underset{N\to\infty}{\longrightarrow} 0$, then also $\mathcal{X}_N^2 \underset{N\to\infty}{\xrightarrow{\text{LWV}}} \mathcal{X}$.*

Proof. Because $\mathcal{X}_N^1 \underset{N\to\infty}{\xrightarrow{\text{LWV}}} \mathcal{X}$ implies that $(T_N, \mu_N^1) \underset{N\to\infty}{\longrightarrow} (T,\mu)$ in the Gw-topology, we obtain $d_{\text{pGP}}((T_N, \mu_N^1), (T,\mu)) \underset{N\to\infty}{\longrightarrow} 0$ by Proposition 2.6. Since μ_N^1 and μ_N^2 are defined on the same space T_N, the latter implies that also

$$\lim_{N\to\infty} d_{\text{pGP}}\big((T_N, \mu_N^2), (T,\mu)\big) = 0, \tag{2.58}$$

(compare (2.14)). Proposition 2.25 allows us to endow these metric measure spaces with the associated measures ν_N and some ν_∞ on T, defined by (2.52). Because of uniqueness of LWV-limits, we have $(T,\mu,\nu_\infty) = (T,\mu,\nu)$. $\square$

Example 2.27 (Counterexample)**.** We cannot extend the result of Corollary 2.26 to pruning measures which do not depend only on the distances.

As illustrated in Figure 3, we consider a constant rooted metric space T and two fixed points $x, y \in T$ such that $x \in [\rho, y]$. We construct two sequences of points $(x_N^1)_{N\in\mathbb{N}}$ and $(x_N^2)_{N\in\mathbb{N}}$ that converge to x, the first from above, the second from below; i.e. $x_N^1 \in [x,y]$ and $x_N^2 \in [\rho, x]$ for all $N \in \mathbb{N}$, and $r(x_N^i, x) \underset{N\to\infty}{\longrightarrow} 0$ for $i = 1,2$. We then define the two sequences of measures $\mu_N^i := \frac{1}{2}\delta_{x_N^i} + \frac{1}{2}\delta_y$ for $i = 1,2$ and a constant measure $\nu_N = \nu = \delta_x$. Clearly, $\mathcal{X}_N^1 \underset{N\to\infty}{\xrightarrow{\text{LWV}}} \mathcal{X}$ and $d_{\text{Pr}}(\mu_N^1, \mu_N^2) \underset{N\to\infty}{\longrightarrow} 0$, but the sequence (T, μ_N^2, ν) does not converge, since the subtree $[\rho, x_N^2]$ never contains the point x, except at the limit. Thus $([\rho, x_N^2], \{x_N^2\}, \nu)$ does not converge pointed Gromov-weakly. $\Diamond$

Lemma 2.28 (Sum of pruning measures)**.** *Let $\mathcal{X}_N^i = (T_N, \mu_N, \nu_N^i) \in \mathbb{H}^{f,\sigma}$ with $(T_N, \mu_N, \nu_N^i) \xrightarrow{\text{LWV}} \mathcal{X}^i = (T, \mu, \nu^i) \in \mathbb{H}^{f,\sigma}$, as $N \to \infty$, for $i = 1,2$. If $(\nu_N^1)_{N\in\mathbb{N}}$ depends $\boldsymbol{v}^{(T,\mu)}$-almost continuously on the distances, we obtain*

$$(T_N, \mu_N, \nu_N^1 + \nu_N^2) \underset{N\to\infty}{\xrightarrow{\text{LWV}}} (T, \mu, \nu^1 + \nu^2). \tag{2.59}$$

Proof. Fix $n \in \mathbb{N}$. Because $(\nu_N^1)_{N\in\mathbb{N}}$ depends $\boldsymbol{v}^{(T,\mu)}$-almost continuously on the distances, we can choose F_n as in (2.52). Let $\underline{U}_N$, $\underline{U}$ be random variables with distribution $\mu_N^{\otimes n}$, $\mu^{\otimes n}$, respectively. By the LWV-convergence and the Skorohod representation theorem, we can couple them such that $\tau_{\mathcal{X}_N^2}(\underline{U}_N) \xrightarrow[N\to\infty]{\text{pGw}} \tau_{\mathcal{X}^2}(\underline{U})$, a.s., which implies $R^{T_N}(\underline{U}_N) \xrightarrow[N\to\infty]{} R^T(\underline{U})$. Because $R^T(\underline{U})$ is a.s. a continuity point of F_n, we also have

$$\tau_{\mathcal{X}_N^1}(\underline{U}_N) = F_n \circ R^{T_N}(\underline{U}_N) \xrightarrow[N\to\infty]{\text{pGw}} F_n \circ R^T(\underline{U}) = \tau_{\mathcal{X}_N^1}(\underline{U}), \text{ a.s.} \tag{2.60}$$

As explained in Remark 2.11, we can define functions $f_N\colon [\![\underline{U}_N]\!] \to [\![\underline{U}]\!]$ such that a.s. $f_N(\underline{U}_N) = \underline{U}$ for large enough N, $\operatorname{dis}(f_N) \xrightarrow[N\to\infty]{} 0$, and $(f_N)_*(\nu_N^i) \underset{N\to\infty}{\Longrightarrow} \nu^i$. Then also $f_{N*}(\nu_N^1 + \nu_N^2) \underset{N\to\infty}{\Longrightarrow} \nu^1 + \nu^2$, which implies $\big([\![\underline{U}_N]\!], \underline{U}_N, \nu_N^1 + \nu_N^2\big) \xrightarrow[N\to\infty]{\text{pGw}} \big([\![\underline{U}]\!], \underline{U}, \nu^1 + \nu^2\big)$, a.s. By Proposition 2.17, this implies the claimed LWV-convergence. □

Remark 2.29 (Assumption on $\boldsymbol{v}^{(T,\mu)}$-almost continuity is important)**.** In Lemma 2.28, we cannot drop the assumption that one of the measures depends $\boldsymbol{v}^{(T,\mu)}$-almost continuously on the distances, because then we cannot use the same coupling of $\underline{U}_N$ to get almost sure convergence of $\tau_{\mathcal{X}_N^i}(\underline{U}_N)$ for $i = 1$ and for $i = 2$. ◊

If we get LWV-convergence of a sequence of bi-measure $\mathbb{R}$-trees, the following lemma asserts that the limit is stable under a small perturbation of ν_N in a certain sense.

Lemma 2.30 (Pruning measure perturbation)**.** *Consider two sequences of bi-measure $\mathbb{R}$-trees $\mathcal{X}_N^i := (T_N, \mu_N, \nu_N^i)$, $i = 1, 2$ that differ by their pruning measures ν_N^1 and ν_N^2. If the two pruning measures are Prohorov merging on subtrees sampled by $\mu_N^{\otimes n}$, i.e.,*

$$\lim_{N\to\infty} d_{\mathrm{Pr}}\big(\nu_N^1\restriction_{[\![U_N^n]\!]}, \nu_N^2\restriction_{[\![U_N^n]\!]}\big) = 0, \ \ \mu_N^{\otimes n}\text{-a.s. }, \ \forall n \in \mathbb{N}, \tag{2.61}$$

then $\mathcal{X}_N^1 \xrightarrow[N\to\infty]{\text{LWV}} \mathcal{X}$, for some $\mathcal{X} = (T, \mu, \nu)$, implies $\mathcal{X}_N^2 \xrightarrow[N\to\infty]{\text{LWV}} \mathcal{X}$.

Proof. Let $\underline{U}_N$ and $\underline{U}$ be sequences of independent μ_N- and μ-distributed random variables in T_N and T, respectively. Because ν^1 and ν^2 are defined on the same measure $\mathbb{R}$-tree, the Prohorov distance in (2.61) is an upper bound for the pGP-distance, and we obtain

$$\begin{aligned} &d_{\mathrm{pGP}}\left(\tau_{\mathcal{X}_N^2}(\underline{U}_N^n), \tau_{\mathcal{X}}(\underline{U}^n)\right) \\ &\le d_{\mathrm{Pr}}\left(\nu_N^1\restriction_{[\![\underline{U}_N^n]\!]}, \nu_N^2\restriction_{[\![\underline{U}_N^n]\!]}\right) + d_{\mathrm{pGP}}\left(\tau_{\mathcal{X}_N^1}(\underline{U}_N^n), \tau_{\mathcal{X}}(\underline{U}^n)\right) \xrightarrow[N\to\infty]{} 0, \end{aligned} \tag{2.62}$$

almost surely, for all $n \in \mathbb{N}$. This implies $(\tau_{\mathcal{X}_N^2})_*(\mu_N^\circ)^{\otimes n} \overset{\text{pGw}}{\Longrightarrow} (\tau_{\mathcal{X}})_*(\mu^\circ)^{\otimes n}$ for all $n \in \mathbb{N}$, and Proposition 2.17 gives the LWV-convergence. □

We conclude this section by giving a simple, sufficient (but far from necessary) condition for relative compactness of a set $\mathbb{K} \subseteq \mathbb{H}^{f,\sigma}$. Assume that for all $\mathcal{X}' = (T', \mu', \nu') \in \mathbb{K}$, there is an isometric embedding of T' into some common $\mathbb{R}$-tree T, and there are measures μ and ν on T dominating all the (push forwards of) μ' and ν', respectively. Further assume that $\mathcal{X} := (T, \mu, \nu) \in \mathbb{H}^{f,\sigma}$. In other words,

$$\mathbb{K} \subseteq \mathbb{S}_{\mathcal{X}} := \big\{(T, \mu', \nu') \in \mathbb{H}^{f,\sigma} \bigm| \mu' \le \mu, \ \nu' \le \nu\big\}. \tag{2.63}$$

Then $\mathbb{K}$ is relatively compact, as the following lemma shows.

Lemma 2.31 (Compactness of $\mathbb{S}_{\mathcal{X}}$)**.** *Let $\mathcal{X} = (T, \mu, \nu) \in \mathbb{H}^{f,\sigma}$. Then $\mathbb{S}_{\mathcal{X}}$, defined in* (2.63)*, is compact in the* LWV*-topology.*

Proof. Consider measures $\mu_N \le \mu$, $\nu_N \le \nu$, $N \in \mathbb{N}$. We have to find a subsequence of $\mathcal{X}_N := (T, \mu_N, \nu_N)$ that converges in $\mathbb{S}_{\mathcal{X}}$. Fix finite subtrees $T_n \subseteq T$, $n \in \mathbb{N}$, with $T_n \subseteq T_{n+1}$ and $\bigcup_{n\in\mathbb{N}} T_n \supseteq \mathrm{Sk}_\mu(T)$.

Because the family $(\mu_N)_{N\in\mathbb{N}}$ is uniformly σ-additive and norm bounded, there exists a setwise convergent subsequence ([Bog07, Thm. 4.7.25]). Assume w.l.o.g. that there is $\mu_\infty \in \mathcal{M}_f(T)$ with $\mu_N(A) \underset{N\to\infty}{\longrightarrow} \mu_\infty(A)$ for all measurable $A \subseteq T$. Similarly, using Cantor's diagonalization argument, we may assume that $\nu_N\restriction_{T_n}$ converges setwise to some $\hat{\nu}_n \in \mathcal{M}_f(T_n)$, for every $n \in \mathbb{N}$. Define

$$\nu_\infty(A) := \sup_{n\in\mathbb{N}} \hat{\nu}_n\bigl(T_n \cap A \cap \mathrm{Sk}_{\mu_\infty}(T)\bigr). \tag{2.64}$$

Because $\hat{\nu}_n\restriction_{T_{n-1}} = \hat{\nu}_{n-1}$, we can easily check that ν_∞ is a measure on T and $\mathcal{X}_\infty := (T, \mu_\infty, \nu_\infty) \in \mathbb{S}_{\mathcal{X}}$. Furthermore, for measurable $A \subseteq \mathrm{Sk}_{\mu_\infty}(T) \subseteq \mathrm{Sk}_\mu(T) \subseteq \bigcup_{n\in\mathbb{N}} T_n$, we obtain

$$\nu_\infty(A) = \sup_{n\in\mathbb{N}} \lim_{N\to\infty} \nu_N(A \cap T_n) \begin{cases} \le \liminf\limits_{N\to\infty} \nu_N(A) \\ \ge \limsup\limits_{N\to\infty} \nu_N(A) - \sup\limits_{n\in\mathbb{N}} \nu(A \setminus T_n) \end{cases} \tag{2.65}$$

Using $A \subseteq \bigcup_{n\in\mathbb{N}} T_n$, this implies

$$\nu_\infty(A) = \lim_{N\to\infty} \nu_N(A). \tag{2.66}$$

We shall show that $\mathcal{X}_N \xrightarrow[N\to\infty]{\mathrm{LWV}} \mathcal{X}_\infty$. By Lemma 2.22, it is enough to show that $\Psi(\mathcal{X}_N) \to \Psi(\mathcal{X}_\infty)$ for all $\Psi \in \mathcal{F}$. Let

$$G := \bigl\{\underline{u} \in T^n : \nu\bigl([\![\underline{u}]\!] \setminus \mathrm{Sk}_{\mu_\infty}(T)\bigr) = 0\bigr\}. \tag{2.67}$$

Fix $\Psi = \Psi^{n,\Phi} \in \mathcal{F}$. Then (2.66) implies

$$\Phi \circ \tau_{\mathcal{X}_N}(\underline{u}) \underset{N\to\infty}{\longrightarrow} \Phi \circ \tau_{\mathcal{X}_\infty}(\underline{u}) \quad \forall \underline{u} \in G, \tag{2.68}$$

and with $B := T^n \setminus G$ we estimate

$$\begin{aligned} \bigl|\Psi(\mathcal{X}_N) - \Psi(\mathcal{X}_\infty)\bigr| &\le \mu_N^{\otimes n}(B) 2\|\Phi\|_\infty + \int_G |\Phi \circ \tau_{\mathcal{X}_N} - \Phi \circ \tau_{\mathcal{X}_\infty}| \, \mathrm{d}\mu_N^{\otimes n} \\ &\quad + \int |\Phi \circ \tau_{\mathcal{X}_\infty}| \, \mathrm{d}(\mu_N^{\otimes n} - \mu_\infty^{\otimes n}). \end{aligned} \tag{2.69}$$

The last term converges to zero because of the setwise convergence of μ_N to μ_∞, and the second term is bounded by $\int_G |\Phi \circ \tau_{\mathcal{X}_N} - \Phi \circ \tau_{\mathcal{X}_\infty}| \, \mathrm{d}\mu^{\otimes n}$, which converges to zero according to (2.68), using the dominated convergence theorem.

For every $(u_1, \ldots, u_n) \in \mathrm{supp}(\mu_\infty)^n \setminus G$, there is an index $k \in \{1, \ldots, n\}$ with $u_k \in \mathrm{At}(\nu) \setminus \mathrm{At}(\mu_\infty)$, where At denotes the set of atoms of a measure. Because $\mathrm{At}(\nu)$ is countable, this implies that B is a μ_∞-null set. Again using setwise convergence of μ_N, we obtain

$$\lim_{N\to\infty} \mu_N^{\otimes n}(B) = \mu_\infty^{\otimes n}(B) = 0. \qquad \square$$

3. The Pruning Process

In this section, we present the construction of the bi-measure valued pruning process, $(X_t)_{t\ge0}$. In Subsection 3.1, we carry out an explicit construction given a realization of the Poisson point process which gives rise to a càdlàg path. We continue the construction in Subsection 3.2 by adding randomness and establishing that the stochastic process obtained this way has the strong Markov property. In Subsection 3.3, we establish the Feller property from which we can conclude that the law of the pruning process on Skorohod space is weakly continuous in the initial distribution on bi-measure $\mathbb{R}$-trees. Finally, in Subsection 3.4 we give an analytic characterization via the infinitesimal generator.

3.1. **Getting the construction started: pruning moves.** It is convenient to introduce randomness later and work initially in a setting where the *cut times* and *cut points* are fixed. Given a bi-measure $\mathbb{R}$-tree, $(T,\mu,\nu) \in \mathbb{H}^{f,\sigma}$, consider a subset $\pi \subseteq \mathbb{R}_+ \times T$. Although π is associated with a particular class representative, it corresponds, of course, to a similar set for any representative of the same equivalence class by mapping across using the appropriate root invariant isometry. Then the set of cut points up to time t is the projection of $\pi \cap \big([0,t] \times T\big)$ onto the tree, i.e.

$$\pi_t := \big\{ v \in T \;\big|\; \exists s \le t : (s,v) \in \pi \big\}. \tag{3.1}$$

For every $v \in T$, the *tree pruned at* v is defined by

$$T^v := \big\{ w \in T \;\big|\; v \notin [\rho, w] \big\}. \tag{3.2}$$

The pruned tree at the set $\pi_t \subseteq T$, T^{π_t}, is the intersection of the trees T^v pruned at $v \in \pi_t$, i.e.,

$$T^{\pi_t} := \bigcap_{v \in \pi_t} T^v. \tag{3.3}$$

We equip the pruned tree T^{π_t} with the restrictions of the measures μ and ν. As always, we write (T^{π_t},μ,ν) instead of $(T^{\pi_t}, \mu\restriction_{T^{\pi_t}}, \nu\restriction_{T^{\pi_t}})$ and easily verify $(T^{\pi_t},\mu,\nu) \in \mathbb{H}^{f,\sigma}$.

Lemma 3.1 (Càdlàg paths)**.** *Fix $\mathcal{X} = (T,\mu,\nu) \in \mathbb{H}^{f,\sigma}$ and a set $\pi \subseteq \mathbb{R}_+ \times T$. The map $t \mapsto \mathcal{X}_t := (T^{\pi_t},\mu,\nu)$ is càdlàg with respect to the* LWV*-topology.*

Proof. Let $0 < s < t$. As $T^{\pi_t} \subseteq T^{\pi_s}$, we obtain for all $\Psi = \Psi^{1,n,\Phi} \in \mathcal{F}$,

$$\begin{aligned} \big|\Psi(\mathcal{X}_s) - \Psi(\mathcal{X}_t)\big| &= \Big| \int_{(T^{\pi_s})^n \setminus (T^{\pi_t})^n} \Phi \circ \tau_{\mathcal{X}} \, \mathrm{d}\mu^{\otimes n} \Big| \\ &\le \|\Phi\|_\infty \cdot \mu^{\otimes n}\big((T^{\pi_s})^n \setminus (T^{\pi_t})^n\big) \\ &\le \|\Phi\|_\infty \cdot n \cdot \|\mu\|^{n-1} \cdot \mu(T^{\pi_s} \setminus T^{\pi_t}). \end{aligned} \tag{3.4}$$

For fixed s, $\bigcap_{t>s} T^{\pi_s} \setminus T^{\pi_t} = \emptyset$, which implies that $\mu(T^{\pi_s} \setminus T^{\pi_t}) \to 0$, as $t \to s$ from the right. Because $\mathcal{F}$ induces the LWV-topology, this implies *right continuity.*

To construct the *left limit*, define $T_{t-} := \cap_{0 \le s < t} T^{\pi_s} \supseteq T^{\pi_t}$ for each $t > 0$, and define $\mathcal{Y}_t := (T_{t-},\mu,\nu)$, which is obviously an element of $\mathbb{H}^{f,\sigma}$. Similarly as before, for all $0 < s < t$ and $\Psi \in \mathcal{F}$, there exists a constant $C = C^\Psi$ such that

$$\big|\Psi(\mathcal{X}_s) - \Psi(\mathcal{Y}_t)\big| \le C \cdot \mu\big(T^{\pi_s} \setminus T_{t-}\big). \tag{3.5}$$

As, for fixed t, $\bigcap_{s<t} T^{\pi_s} \setminus T_{t-} = \emptyset$, $\mathcal{Y}_t$ is indeed the left limit. ☐

3.2. **Continuing the construction: adding randomness.** In this subsection we define, given a bi-measure $\mathbb{R}$-tree $\mathcal{X} = (T,\mu,\nu)$, the pruning process of $\mathcal{X}$, where π is now the (random) Poisson point measure with intensity $\lambda \otimes \nu$ on $\mathbb{R}_+ \times T$. Here, we identify an atomic measure $\mathfrak{m}$ on T with the set $\mathrm{At}(\mathfrak{m})$ of its atoms and define

$$T^{\mathfrak{m}} := T^{\mathrm{At}(\mathfrak{m})} = \bigcap_{v \in \mathrm{At}(\mathfrak{m})} T^v. \tag{3.6}$$

Definition 3.2 (The pruning process)**.** *Fix a bi-measure $\mathbb{R}$-tree $\mathcal{X} := (T,\mu,\nu) \in \mathbb{H}^{f,\sigma}$. Let $\pi^{\mathcal{X}}$ be the Poisson point measure on $\mathbb{R}_+ \times T$ with intensity measure $\lambda \otimes \nu$, where λ is the Lebesgue measure on $\mathbb{R}_+$. We define* the pruning process, *$X := (X_t)_{t \ge 0}$, as the bi-measure $\mathbb{R}$-tree-valued process obtained by pruning $X_0 := \mathcal{X}$ at the points of the Poisson point process $\pi_t(\cdot) := \pi_t^{\mathcal{X}}(\cdot) := \pi^{\mathcal{X}}\big([0,t] \times \cdot\big)$, i.e.,*

$$X_t := \big(T^{\pi_t},\mu,\nu\big) := \big(T^{\pi_t}, \mu\restriction_{T^{\pi_t}}, \nu\restriction_{T^{\pi_t}}\big). \tag{3.7}$$

$\mathbb{E}^{\mathcal{X}}$, or $\mathbb{E}$ if there is no confusion, denotes the distribution of the process X starting from $X_0 = \mathcal{X}$.

Lemma 3.3 (Strong Markov property)**.** *The pruning process X is a strong Markov process.*

Proof. Denote by $(\mathcal{A}_t)_{t\geq 0}$ the filtration generated by the Poisson point process $(\pi_t)_{t\geq 0}$. Note that X is adapted to this filtration. Using the strong Markov property of the Poisson process, we get for every $t \geq 0$, stopping time σ, and $\underline{u} \in T^n$, $n \in \mathbb{N}$,

$$\mathbb{P}\big(\pi_{\sigma+t}([\![\underline{u}]\!]) = 0 \bigm| \mathcal{A}_\sigma\big) = \mathbb{1}_{\{\pi_\sigma([\![\underline{u}]\!])=0\}}\mathbb{P}\big(\pi_t([\![\underline{u}]\!]) = 0\big). \tag{3.8}$$

For every $\tilde{\Psi} = \Psi^{1,n,\tilde{\Phi}} \in \tilde{\mathcal{F}}^1$, this implies

$$\begin{aligned} \mathbb{E}\Big[\tilde{\Psi}(X_{\sigma+t}) \Bigm| \mathcal{A}_\sigma\Big] &= \int_{T^n} \mu^{\otimes n}(\mathrm{d}\underline{u})\, \mathbb{P}\big(\pi_{\sigma+t}([\![\underline{u}]\!]) = 0 \bigm| \mathcal{A}_\sigma\big) \cdot \tilde{\Phi}(\tau_{\mathcal{X}}(\underline{u})) \\ &= \int_{(T^{\pi_\sigma})^n} \mu^{\otimes n}(\mathrm{d}\underline{u})\, e^{-t\nu([\![\underline{u}]\!])} \cdot \tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big). \end{aligned} \tag{3.9}$$

On the other hand, we also have

$$\mathbb{E}^{X_\sigma}\Big[\tilde{\Psi}(X_t)\Big] = \int_{(T^{\pi_\sigma})^n} \mu^{\otimes n}(\mathrm{d}\underline{u})\, e^{-t\nu([\![\underline{u}]\!])} \cdot \tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big). \tag{3.10}$$

Because $X_t \in \mathbb{H}^{\|\mu\|,\sigma}$, for all $t \geq 0$, and $\tilde{\mathcal{F}}^1$ is a separating class on this space, we obtain the strong Markov property. □

3.3. **Continuity of the pruning process.** In this subsection we show that the law of X_t under $\mathbb{P}^{\mathcal{X}}$ is weakly continuous in the initial value $\mathcal{X}$ for each $t \geq 0$. This property is sometimes referred to as the *Feller property* of the corresponding semigroup $(S_t)_{t\geq 0}$, although this terminology is often restricted to the case of a locally compact state space and transition operators that map the space of continuous functions that vanish at infinity into itself. In the latter, more restrictive case, the Feller property implies that the law of the whole process (as random variable on Skorohod space) depends continuously on the initial value. If S_t maps only $\mathcal{C}_b$ into itself, this is no longer the case in general, and one needs an extra argument. The pruning process $(X_t)_{t\geq 0}$, however, does depend continuously on the initial condition (Theorem 3.6).

Let $(S_t)_{t\geq 0}$ be the semi-group associated to the pruning process $(X_t)_{t\geq 0}$, i.e. for $t \geq 0$ and a bounded measurable function $G : \mathbb{H}^{f,\sigma} \to \mathbb{R}$,

$$S_t G(\mathcal{X}) := \mathbb{E}^{\mathcal{X}}\big[G(X_t)\big]. \tag{3.11}$$

Proposition 3.4 (Feller continuity)**.** *The process $X := (X_t)_{t\geq 0}$ is Feller continuous in the sense that $S_t\big(\mathcal{C}_b(\mathbb{H}^{f,\sigma})\big) \subseteq \mathcal{C}_b(\mathbb{H}^{f,\sigma})$.*

Proof. Consider the convergence of bi-measure $\mathbb{R}$-trees $\mathcal{X}_N \xrightarrow{\mathrm{LWV}} \mathcal{X}$. Write $K := \sup\{\|\mu_N\|, N \in \mathbb{N}\}$, then the sequence converges in $\mathbb{H}^{K,\sigma}$. Because $\tilde{\mathcal{F}}^1$ is convergence determining on $\mathbb{H}^{K,\sigma}$ (see Proposition 2.23), it is enough to prove for all $\tilde{\Psi} \in \tilde{\mathcal{F}}^1$, $t > 0$ that

$$\mathbb{E}^{\mathcal{X}_N}\big[\tilde{\Psi}(X_t)\big] \underset{N\to\infty}{\longrightarrow} \mathbb{E}^{\mathcal{X}}\big[\tilde{\Psi}(X_t)\big]. \tag{3.12}$$

Fix therefore $\tilde{\Psi} = \Psi^{1,n,\tilde{\Phi}} \in \tilde{\mathcal{F}}^1$. Then

$$\begin{aligned} \mathbb{E}^{\mathcal{X}_N}\big[\tilde{\Psi}(X_t)\big] &= \mathbb{E}^{\mathcal{X}_N}\Big[\int_{(T_N^{\pi_t})^n} \tilde{\Phi} \circ \tau_{\mathcal{X}_N}\, \mathrm{d}\mu_N^{\otimes n}\Big] \\ &= \int_{T_N^n} \mu_N^{\otimes n}(\mathrm{d}\underline{u})\, \mathbb{P}\big(\pi_t^{\mathcal{X}_N}([\![\underline{u}]\!]) = 0\big) \cdot \tilde{\Phi}\big([\![\underline{u}]\!], \underline{u}, \nu_N\big). \end{aligned} \tag{3.13}$$

Using $\mathbb{P}\left(\pi_t^{\mathcal{X}_N}([\![\underline{u}]\!]) = 0\right) = \exp(-t\nu_N([\![\underline{u}]\!]))$, we see that $\mathbb{E}^{\mathcal{X}_N}\Big[\tilde{\Psi}(X_t)\Big] = \tilde{\Psi}'(\mathcal{X}_N)$ for some $\tilde{\Psi}' \in \tilde{\mathcal{F}}^1$. The convergence follows therefore from the LWV-convergence of $(\mathcal{X}_N)_{N\in\mathbb{N}}$. □

Consider a separable, metrizable space E and a contraction semigroup $S = (S_t)_{t\geq 0}$ on $\mathcal{C}_b(E)$. We define

$$\mathcal{D}(S) := \big\{f \in \mathcal{C}_b(E) : \lim_{t\to 0}\|S_t f - f\|_\infty = 0\big\}. \tag{3.14}$$

Note that $\mathcal{D}(S)$ is uniformly closed, S_t maps $\mathcal{D}(S)$ into itself, and the restriction of $(S_t)_{t\geq 0}$ to $\mathcal{D}(S)$ is a strongly continuous contraction semigroup. In particular, the restricted semigroup has a generator $\Omega_S\colon \mathcal{D}(\Omega_S) \to \mathcal{D}(S)$ with dense domain $\mathcal{D}(\Omega_S) \subseteq \mathcal{D}(S)$.

Lemma 3.5. *Let E be a separable, metrizable space, and $Y^x = (Y^x_t)_{t\geq 0}$, $x \in E$, an E-valued, Feller-continuous (time-homogeneous) Markov process with càdlàg paths and semigroup $S = (S_t)_{t\geq 0}$ on $\mathcal{C}_b(E)$. Assume that there is a set $\mathcal{G} \subseteq \mathcal{D}(S)$ that is multiplicatively closed and induces the topology of E. Then the map*

$$\begin{array}{ccc} \mathcal{M}_1(E) & \to & \mathcal{M}_1\big(D_E(\mathbb{R}_+)\big), \\ \eta & \mapsto & \mathcal{L}(Y^\eta) \end{array} \tag{3.15}$$

is continuous, where $D_E(\mathbb{R}_+)$ is the space of càdlàg paths with Skorohod topology, $\mathcal{L}$ is the law of a process, and Y^η is the process with initial condition $\mathcal{L}(Y^\eta_0) = \eta$, i.e., $\mathcal{L}(Y^\eta) = \int \mathcal{L}(Y^x)\, \eta(\mathrm{d}x)$.

Proof. It is sufficient to prove that $\mathcal{L}(Y^{x_N}) \underset{N\to\infty}{\Longrightarrow} \mathcal{L}(Y^x)$ for every convergent sequence $x_N \underset{N\to\infty}{\longrightarrow} x$ in E. Because $\mathcal{G}$ induces the topology of E, it strongly separates points (see Lemma 1 in [BK10]). According to Theorem 10 of [BK10], it is therefore enough to prove that for all $f_1, \dots, f_k \in \mathcal{G}$,

$$\big(f_1(Y^{x_N}_t), \dots, f_k(Y^{x_N}_t)\big)_{t\geq 0} \underset{N\to\infty}{\overset{\mathcal{L}}{\Longrightarrow}} \big(f_1(Y^{x}_t), \dots, f_k(Y^{x}_t)\big)_{t\geq 0} \tag{3.16}$$

in Skorohod space as $\mathbb{R}^k$-valued processes. The assumed Feller continuity implies f.d.d. convergence, hence it is enough to prove tightness.

To this end, we apply Theorem 3.9.4 of [EK86]. The linear span $C_a := \operatorname{span}(\mathcal{G})$ of $\mathcal{G}$ is an algebra contained in $\mathcal{D}(S)$, and the domain $\mathcal{D}(\Omega_S)$ of the generator Ω_S of S is dense in $\mathcal{D}(S)$. For every $f \in \mathcal{D}(\Omega_S)$, we define $Z^N_t := \Omega_S f(Y^{x_N}_t)$. Then the following hold:

(i) The processes $\Big(f(Y^{x_N}_t) - \displaystyle\int_0^t Z^N_s \,\mathrm{d}s\Big)_{t\geq 0}$ are martingales.

(ii) For all $T \geq 0$, $\sup_{N\in\mathbb{N}} \mathbb{E}\big[\operatorname*{ess\,sup}_{0\leq t\leq T} |Z^N_t|\big] \leq \|\Omega_S f\|_\infty < \infty$.

Now tightness of the processes $\big(f_1(Y^{x_N}_t), \dots, f_k(Y^{x_N}_t)\big)_{t\geq 0}$, $N \in \mathbb{N}$, for every fixed $f_1, \dots, f_k \in C_a \supseteq \mathcal{G}$ follows from [EK86, Thm. 3.9.4]. □

Theorem 3.6 (Continuity in the initial distribution)**.** *The law of X on the Skorohod space depends continuously on the initial condition.*

Proof. It is sufficient to prove continuity for deterministic initial conditions. Every convergent sequence $\mathcal{X}_N \underset{N\to\infty}{\overset{\mathrm{LWV}}{\longrightarrow}} \mathcal{X}$ in $\mathbb{H}^{f,\sigma}$ is contained in $\mathbb{H}^{K,\sigma}$ for some $K > 0$, and the pruning process stays a.s. in that subspace. We verify the conditions of Lemma 3.5 for the $\mathbb{H}^{K,\sigma}$-valued pruning process. It has càdlàg paths (Lemma 3.1), is Feller-continuous (Proposition 3.4), and $\tilde{\mathcal{F}}^1 \subseteq \mathcal{C}_b(\mathbb{H}^{K,\sigma})$ is multiplicatively closed and induces the LWV-topology. It remains to show that $\tilde{\mathcal{F}}^1 \subseteq \mathcal{D}(S)$, where S is the $\mathcal{C}_b(\mathbb{H}^{K,\sigma})$-semigroup.

For $\tilde{\Phi} \in \tilde{\Pi}_n$, $x \in \mathbb{R}_+$, we define

$$\gamma_{\tilde{\Phi}}(x) := \sup_{(T,\underline{u},\nu)\in\mathbb{H}_n,\, \|\nu\|=x} \big|\tilde{\Phi}(T,\underline{u},\nu)\big| \tag{3.17}$$

and note that $\lim_{x\to\infty} \gamma_{\tilde{\Phi}}(x) = 0$. Using Fubini's theorem, we obtain for $\tilde{\Psi} = \Psi^{1,n,\tilde{\Phi}} \in \tilde{\mathcal{F}}^1$ and $\mathcal{X} = (T,\mu,\nu) \in \mathbb{H}^{K,\sigma}$

$$\begin{aligned} S_t\tilde{\Psi}(\mathcal{X}) &= \int \mathbb{P}^{\mathcal{X}}\big(\pi^{\mathcal{X}}_t([\![\underline{u}]\!]) = 0\big) \cdot \tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big)\, \mu^{\otimes n}(\mathrm{d}\underline{u}) \\ &= \int e^{-t\nu([\![\underline{u}]\!])} \cdot \tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big)\, \mu^{\otimes n}(\mathrm{d}\underline{u}). \end{aligned} \tag{3.18}$$

Therefore,

$$\sup_{\mathcal{X}\in\mathbb{H}^{K,\sigma}} \big|S_t\tilde{\Psi}(\mathcal{X}) - \tilde{\Psi}(\mathcal{X})\big| \le K^n \sup_{x\in\mathbb{R}_+} \gamma_{\tilde{\Phi}}(x)\big(1-e^{-tx}\big) \underset{t\to 0}{\longrightarrow} 0. \qquad \square$$

3.4. **The infinitesimal generator.** In this subsection we calculate the action of the generator on the test functions $\tilde{\Psi} \in \tilde{\mathcal{F}}^1$. For these functions to be bounded, we have to work on the space $\mathbb{H}^{K,\sigma}$. Note that $\mathbb{H}^{K,\sigma}$ is a good state space for the pruning process, as once started in $\mathbb{H}^{K,\sigma}$, it will never leave the space. In the following we write

$$\big(\Omega, \mathcal{D}(\Omega)\big) \quad \text{and} \quad \big(\Omega_K, \mathcal{D}(\Omega_K)\big) \tag{3.19}$$

for the infinitesimal generators of the pruning process with state spaces $\mathbb{H}^{f,\sigma}$ and $\mathbb{H}^{K,\sigma}$ respectively.

Proposition 3.7 (Infinitesimal Generator)**.** *For every $K > 0$, we have $\tilde{\mathcal{F}}^1 \subseteq \mathcal{D}(\Omega_K)$. Furthermore, for $\tilde{\Psi} = \Psi^{1,n,\tilde{\Phi}} \in \tilde{\mathcal{F}}^1$ and $\mathcal{X} = (T,\mu,\nu) \in \mathbb{H}^{K,\sigma}$,*

$$\Omega\tilde{\Psi}\big(\mathcal{X}\big) = \int \nu(\mathrm{d}v)\big[\tilde{\Psi}\big((T^v,\mu,\nu)\big) - \tilde{\Psi}\big(\mathcal{X}\big)\big] \tag{3.20}$$

$$= -\int \mu^{\otimes n}(\mathrm{d}\underline{u})\; \nu\big([\![\underline{u}]\!]\big)\tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big). \tag{3.21}$$

Proof. Using Formula (3.18), we obtain for $\tilde{\Psi} = \Psi^{1,n,\tilde{\Phi}} \in \tilde{\mathcal{F}}^1$, $\mathcal{X} \in \mathbb{H}^{K,\sigma}$,

$$\tfrac{1}{t}\big(S_t\tilde{\Psi}(\mathcal{X}) - \tilde{\Psi}(\mathcal{X})\big) = -\tfrac{1}{t}\int \mu^{\otimes n}(\mathrm{d}\underline{u})\,\big(1-e^{-t\nu([\![\underline{u}]\!])}\big)\tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big). \tag{3.22}$$

Note that $|1-e^{-x}-x| \le x^2$, for all $x \ge 0$ and recall the definition of $\gamma_{\tilde{\Phi}}$ from (3.17). Comparing (3.22) to (3.21), we see that

$$\begin{aligned} &\sup_{\mathcal{X}\in\mathbb{H}^{K,\sigma}} \Big|\tfrac{1}{t}\big(S_t\tilde{\Psi}(\mathcal{X}) - \tilde{\Psi}(\mathcal{X})\big) + \int \mu^{\otimes n}(\mathrm{d}\underline{u})\,\nu\big([\![\underline{u}]\!]\big)\tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big)\Big| \\ &\le \sup_{\mathcal{X}\in\mathbb{H}^{K,\sigma}} t\cdot\int \mu^{\otimes n}(\mathrm{d}\underline{u})\,\nu\big([\![\underline{u}]\!]\big)^2\,\big|\tilde{\Phi}(\tau_{\mathcal{X}}(\underline{u}))\big| \le tK^n \sup_{x\in\mathbb{R}_+} x^2\gamma_{\tilde{\Phi}}(x). \end{aligned} \tag{3.23}$$

Due to our assumptions on $\tilde{\Phi} \in \widetilde{\Pi}_n$, $x^2\gamma_{\tilde{\Phi}}(x)$ is bounded, and we obtain uniform convergence of $\frac{1}{t}(S_t\tilde{\Psi} - \tilde{\Psi})$ on $\mathbb{H}^{K,\sigma}$ for $t \to 0$. Hence $\tilde{\mathcal{F}}^1 \subseteq \mathcal{D}(\Omega_K)$ and Formula (3.21) are proven.

We next prove Formula (3.20). Notice that for all $\underline{u} \in T^n$,

$$\nu([\![\underline{u}]\!]) = \int_T \mathbb{1}_{\{v\in[\![\underline{u}]\!]\}}\,\nu(\mathrm{d}v) = \int_T 1 - \mathbb{1}_{\{\underline{u}\in(T^v)^n\}}\,\nu(\mathrm{d}v). \tag{3.24}$$

Inserting the latter into (3.21) and using Fubini's theorem yields

$$\Omega\tilde{\Psi}(\mathcal{X}) = \int_T \nu(\mathrm{d}v)\left(\int_{(T^v)^n} \mu^{\otimes n}(\mathrm{d}\underline{u})\,\tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big) - \int_{T^n} \mu^{\otimes n}(\mathrm{d}\underline{u})\,\tilde{\Phi}\big(\tau_{\mathcal{X}}(\underline{u})\big)\right), \tag{3.25}$$

which gives (3.20). $\square$

4. EXAMPLES

In this section we want to apply Theorem 3.6 to obtain convergence of various pruning processes that appear in the literature. We first recall the excursion representation of a measure $\mathbb{R}$-tree. We denote by

$$\mathcal{E} := \big\{e\colon [0,1] \to \mathbb{R}_+ \;\big|\; e \text{ is l.s.c., } e(0) = e(1) = 0\big\} \tag{4.1}$$

the set of lower semi-continuous excursions on $[0,1]$. From each excursion $e \in \mathcal{E}$, we can define a measure $\mathbb{R}$-tree in the following way:

- $r_e(x,y) := e(x) + e(y) - 2\inf_{[x,y]} e$ is a pseudo-distance on $[0,1]$,
- $x,y \in [0,1]$ are said to be equivalent, $x \sim_e y$, if $r_e(x,y) = 0$,

- the image of the projection $\pi_e : [0,1] \to [0,1]/\sim_e$ endowed with the push forward of r_e (again denoted r_e), i.e. $T_e := (T_e, r_e, \rho_e) := \big(\pi_e([0,1]), r_e, \pi_e(0)\big)$, is a 0-hyperbolic space (for example, [EW06, Lemma 3.1]).
- We endow this space with the probability measure $\mu_e := \pi_{e*}\lambda_{[0,1]}$ which is the push forward of the Lebesgue measure on $[0,1]$.

We denote by $g : \mathcal{E} \to \mathbb{H}_\rho$ the resulting "glue function",

$$g(e) := \big(T_e, \mu_e\big), \tag{4.2}$$

which sends an excursion to a rooted probability measure $\mathbb{R}$-tree. The map g is continuous if $\mathbb{H}_\rho$ is endowed with the Gromov-weak topology, and $\mathcal{E}$ with the uniform topology (see [ADH14, Prop. 2.9] for the case of continuous excursions) or, more generally, with the weaker *excursion topology* introduced in [Löh13] (see Theorem 4.8 there).

Example 4.1 (An approach via excursions)**.** Consider a sequence of random excursions $e_N = (e_N(s), s \in [0,1]) \in \mathcal{E}$, $N \in \mathbb{N}$, that converges in distribution (with respect to the uniform, respectively the excursion topology) to $e \in \mathcal{E}$. For each $N \in \mathbb{N}$, we denote by $(X_t^N)_{t\geq 0}$ the pruning process started in the bi-measure tree $\mathcal{X}_{e_N} := (T_{e_N}, \mu_{e_N}, \lambda_{T_{e_N}}) \in \mathbb{H}^{f,\sigma}$, where $\lambda_{T_{e_N}}$ is the length measure on T_{e_N}, and similarly for $(X_t)_{t\geq 0}$ and $\mathcal{X}_e$.

Due to continuity of g, we have that $g(e_N)$ converges Gromov-weakly in distribution to $g(e)$. By Proposition 2.25, we obtain the LWV-convergence in distribution of $\mathcal{X}_{e_N}$ to $\mathcal{X}_e$, and by Theorem 3.6, we get the Skorohod convergence

$$(X_t^N)_{t\geq 0} \overset{\text{Sk}}{\Longrightarrow} (X_t)_{t\geq 0}$$

as $\mathbb{H}^{f,\sigma}$-valued processes with LWV-topology. Note that this, in particular, implies Skorohod convergence of the pruning processes $(T_{e_N}^{\pi_t}, \mu_{e_N})_{t\geq 0}$ as measure $\mathbb{R}$-tree-valued processes in the usual Gromov-weak topology, where we do not keep track of the pruning measure. ◊

We shall apply this example to Galton-Watson trees. Consider a critical or sub-critical Galton-Watson tree $\mathcal{G}$ with offspring distribution η on $\mathbb{N}_0$, i.e., every node in the discrete tree has a random number of children given independently by the distribution η, where $\mathbb{E}[\eta] \leq 1$. Encode $\mathcal{G}$ as a rooted $\mathbb{R}$-tree with unit length edges. For each $N \in \mathbb{N}$, let $\mathcal{G}_N$ be the tree $\mathcal{G}$ conditioned to have N nodes (in addition to the root). We consider two different sampling measures μ on $\mathcal{G}_N$: one is the *normalized length measure*

$$\mu_N^{\text{ske}} := \tfrac{1}{N}\lambda_{\mathcal{G}_N}, \tag{4.3}$$

and the second is the *uniform measure on the nodes*,

$$\mu_N^{\text{nod}} := \tfrac{1}{N}\sum_{i=1}^{N} \delta_{x_i}, \tag{4.4}$$

where $\{x_1, ..., x_N\}$ are the nodes of $\mathcal{G}_N$. Notice that

$$\mu_N^{\text{nod}}(A) = \sum_{x\in\text{nod}(A)} \mu_N^{\text{ske}}\big([x_-, x]\big) \leq \mu_N^{\text{ske}}\big(\{v \in \mathcal{G}_N \mid r_{\mathcal{G}_N}(v, A) < 1\}\big) \tag{4.5}$$

where $\text{nod}(A)$ is the set of nodes in A and x_- is the parent of x.

In order to obtain convergence, we rescale the tree $\mathcal{G}_N$ to have edge lengths $a_N > 0$, i.e., we leave the set unchanged and multiply the metric by a_N. We denote the rescaled tree by $a_N\mathcal{G}_N$. As

$$d_{\text{Pr}}^{a_N\mathcal{G}_N}\big(\mu_N^{\text{ske}}, \mu_N^{\text{nod}}\big) \leq a_N \tag{4.6}$$

on the rescaled tree by (4.5), μ_N^{nod} and μ_N^{ske} become arbitrary close whenever a_N converges to zero, as $N \to \infty$.

We also consider two different pruning measures ν: one is the *length measure on the rescaled tree*,

$$\nu_N^{\mathrm{ske}} := \lambda_{a_N \mathcal{G}_N} = a_N \cdot N \cdot \mu_N^{\mathrm{ske}}, \tag{4.7}$$

and the second is a suitably rescaled *uniform measure on the nodes*,

$$\nu_N^{\mathrm{nod}} := a_N \cdot N \cdot \mu_N^{\mathrm{nod}}. \tag{4.8}$$

In order to be in a position to apply Example 4.1, we associate the conditioned and rescaled bi-measure Galton-Watson tree with an excursion. That is, by the depth-first search algorithm we obtain a graph-theoretic path $\rho = y_0, y_1, ..., y_{2N-1}, y_{2N} = \rho$ in the discrete tree, which traverses each edge exactly twice. The *contour process* $(C_N(t), 0 \le t \le 1)$ of $\mathcal{G}_N$ is the linear interpolation of $C_N(\frac{k}{2N}) := h(y_k) := r_{\mathcal{G}_N}(\rho, y_k)$, $k = 0, ..., 2N$. Note that in our definition of C_N, the domain is normalized to $[0,1]$, and we obtain that

$$g(C_N) = \big(\mathcal{G}_N, \mu_N^{\mathrm{ske}}\big). \tag{4.9}$$

Example 4.2 (Brownian CRT)**.** Let the variance σ^2 of η be finite and choose

$$a_N := \tfrac{\sigma}{\sqrt{N}}. \tag{4.10}$$

We know from Theorem 23 in [Ald93] that $(a_N C_N(t), 0 \le t \le 1)$ converges uniformly in distribution to $(2B(t), 0 \le t \le 1)$, where B is the standard Brownian excursion. We now apply Example 4.1 and get the LWV-convergence in distribution of the bi-measure $\mathbb{R}$-trees

$$\big(\tfrac{\sigma}{\sqrt{N}}\mathcal{G}_N, \mu_N^{\mathrm{ske}}, \nu_N^{\mathrm{ske}}\big) \underset{N\to\infty}{\overset{\mathrm{LWV}}{\longrightarrow}} \big(CRT, \mu, \lambda_{CRT}\big), \tag{4.11}$$

where $(CRT, \mu) = g(2B)$ is the $\mathbb{R}$-tree called Brownian continuum random tree, and λ_{CRT} is the length measure on the Brownian CRT.

By Corollary 2.26 and Lemma 2.30, we also have the convergence

$$\big(\tfrac{\sigma}{\sqrt{N}}\mathcal{G}_N, \mu_N, \nu_N\big) \underset{N\to\infty}{\overset{\mathrm{LWV}}{\longrightarrow}} \big(CRT, \mu, \lambda_{CRT}\big) \tag{4.12}$$

for all choices of $\mu_N \in \{\mu_N^{\mathrm{ske}}, \mu_N^{\mathrm{nod}}\}$ and $\nu_N \in \{\nu_N^{\mathrm{ske}}, \nu_N^{\mathrm{nod}}\}$. Finally we have the convergence of the pruning processes in Skorohod space:

$$\Big(\tfrac{\sigma}{\sqrt{N}}\mathcal{G}_N^{\pi_t}, \mu_N, \nu_N\Big)_{t\ge 0} \underset{\mathrm{LWV}}{\overset{\mathrm{Sk}}{\Longrightarrow}} \left(CRT^{\pi_t}, \mu, \lambda_{CRT}\right)_{t\ge 0}. \tag{4.13}$$

In particular,

$$\Big(\tfrac{\sigma}{\sqrt{N}}\mathcal{G}_N^{\pi_t}, \mu_N^{\mathrm{nod}}\Big)_{t\ge 0} \underset{\mathrm{Gw}}{\overset{\mathrm{Sk}}{\Longrightarrow}} \left(CRT^{\pi_t}, \mu\right)_{t\ge 0}. \tag{4.14}$$

Notice that for $\nu_N = \nu_N^{\mathrm{ske}}$, the pruning process $(\mathcal{G}_N^{\pi_t})_{t\ge0}$ is, up to the time transformation $u = e^{-t/\sqrt{N}}$, the same as the pruning process $(\mathcal{G}_u^{\mathrm{AP}})_{u\in[0,1]}$ uniformly on the edges of Aldous and Pitman in [AP98b]. The process on the right hand side is the one considered by Aldous and Pitman [AP98a] and by Abraham and Serlet [AS02] for example. ◊

Example 4.3 (α-stable Lévy tree)**.** We know from Theorem 3.1 of [Duq03] that if η is in the domain of attraction of an α-stable distribution with $\alpha \in (1,2]$, then there exists a sequence a_N such that $(a_N C_N(t), 0 \le t \le 1)$ converges uniformly in distribution to $(H(t), 0 \le t \le 1)$, where H is a continuous excursion that codes an α-stable Lévy tree, $(LT_\alpha, \mu) := g(H)$. More precisely, for $\eta(k) \sim_{k\to\infty} Ck^{1-\alpha}$, we have $a_N = N^{-\bar\alpha}\left(\frac{\alpha(\alpha-1)}{C\Gamma(2-\alpha)}\right)^{-1/\alpha}$ with $\bar\alpha = 1 - 1/\alpha$ (see Section 1.2 in [CH12]). As in Example 4.2, we obtain

$$\big(a_N\mathcal{G}_N^{\pi_t}, \mu_N, \nu_N\big)_{t\ge 0} \underset{\mathrm{LWV}}{\overset{\mathrm{Sk}}{\Longrightarrow}} \left(LT_\alpha^{\pi_t}, \mu, \lambda_{LT_\alpha}\right)_{t\ge 0} \tag{4.15}$$

or more precisely

$$\left(\frac{1}{N^{\bar{\alpha}}}\mathcal{G}_N^{\pi_t}, \mu_N, \nu_N\right)_{t\geq 0} \xRightarrow[\text{LWV}]{\text{Sk}} \left(\left(\frac{\alpha(\alpha-1)}{C\Gamma(2-\alpha)}\right)^{1/\alpha} LT_\alpha^{\pi_t}, \mu, \lambda_{LT_\alpha}\right)_{t\geq 0}. \tag{4.16}$$

where $\mu_N = \mu_N^{\text{ske}}$ or μ_N^{nod} and $\nu_N = \nu_N^{\text{ske}}$ or ν_N^{nod}. ◊

Example 4.4 (Pruning at a height)**.** As before, we assume that the Gromov-weak convergence $(a_N\mathcal{G}_N, \mu_N) \xrightarrow{\text{Gw}} (LT_\alpha, \mu)$ holds. For $a \geq 0$, we define the pruning measure

$$\nu_N^a := \sum_{x\in\mathcal{G}_N^a} \delta_x, \tag{4.17}$$

where $\mathcal{G}_N^a = \{x \in \mathcal{G}_N \mid r_N(\rho, x) = a\}$, and the corresponding measure

$$\nu_\infty^a := \sum_{x\in LT_\alpha^a \cap \text{Sk}_\mu(LT_\alpha)} \delta_x \tag{4.18}$$

on LT_α. Here, we restrict the pruning measure to the points of LT_α which are not leaves in order to ensure the condition $\nu_\infty^a\big(\text{Lf}_\mu(LT_\alpha^a)\big) = 0$. Because the probability that $\mu(LT_\alpha^a) \neq 0$ is zero for fixed a, the sequence $(\nu_N^a)_{N\in\mathbb{N}\cup\{\infty\}}$ almost surely depends $\boldsymbol{v}^{(LT_\alpha,\mu)}$-almost continuously on the distances, i.e.

$$R^{LT_\alpha}{}_*\mu^{\otimes n}\left(\text{Discont}(F_n)\right) = 0 \;\; a.s.,$$

for F_n as in (2.52). We use Proposition 2.25 and the previous construction to get

$$\left(a_N\mathcal{G}_N^{\pi_t}, \mu_N, \nu_N^a\right)_{t\geq 0} \xRightarrow{\text{Sk}} \left(LT_\alpha^{\pi_t}, \mu, \nu_\infty^a\right)_{t\geq 0}. \tag{4.19}$$

It is easy to check that $(LT_\alpha^{\pi_t}, \mu, \nu_\infty^a)$ converges almost surely, as $t \to \infty$, in the LWV-topology to $\big(LT_\alpha^{\leq a}, \mu, 0\big)$ where $LT_\alpha^{\leq a} = \{x \in LT_\alpha \mid r(\rho, x) \leq a\}$. This is the pruning construction at the height a of Miermont [Mie03]. ◊

Remark 4.5 (Pruning based on other scaling results)**.** Some authors give other convergence of Galton-Watson trees to continuous trees. For example a sequence of Galton-Watson trees $(\mathcal{G}_N)_{N\in\mathbb{N}}$ conditioned to have maximum height at least $\gamma_N T$ converges to a general Lévy tree conditioned to have maximum height at least T, see Proposition 2.5.2 in [DLG02]. Or a sequence of Galton-Watson trees that converges to a forest of Lévy trees, see Theorem 2.4.1 in [DLG02]. In the first case, the previous results clearly applies. In the second case, in general we do not have an excursion with finite length anymore, i.e., the measure μ^{ske} might become infinite. However, if we restrict the domain of the contour processes to a finite interval, we can still apply the previous results. ◊

Example 4.6 (More general pruning)**.** A non-uniform pruning process on the branch points of a general Galton-Watson tree has been defined by Abraham, Delmas and He [ADH12]: they cut a branch point v and its subtree above independently with probability $1 - u^{c(v)-1}$, where $c(v)$ is the number of children of v. This corresponds to taking the pruning measure ν_N^{ADH} on $\mathcal{G}_N$ that is supported on the branch points and satisfies

$$\nu_N^{ADH}\big(\{v\}\big) := c(v) - 1. \tag{4.20}$$

A pruning process on the infinite branch points of a Lévy tree has been defined by Abraham and Delmas [AD12]: they cut each infinite branch point and its subtree above independently with probability $1 - e^{-t\Delta_x}$ where Δ_x is the weight of the node x that can be defined using the jumps of the Lévy process. This corresponds to taking a measure ν^{AD} on the infinite branch points of the Lévy tree.

Because we know that a properly renormalized sequence of conditioned Galton-Watson trees converges to a Lévy tree, we conjecture that there exists a sequence b_N such that

$$(4.21)\qquad \big(a_N\mathcal{G}_N, \mu_N^{\text{nod}}, \nu_N^{\text{ske}} + b_N\nu_N^{\text{ADH}}\big) \underset{N\to\infty}{\xrightarrow{\text{LWV}}} \big(LT, \mu^{\text{nod}}, \nu^{\text{ske}} + \nu^{\text{AD}}\big)$$

where LT is a Lévy tree or at least an α-stable Lévy tree with b_N of the order $N^{-1/\alpha}$ up to a slowly varying function. The Poisson point process with intensity $\nu^{\text{ske}} + \nu^{\text{AD}}$ used in the pruning of the Lévy tree is the Poisson point process given in Subsection 4.2 of [Voi11]. $\Diamond$

Example 4.7 (Cutting down trees)**.** Random deconstruction of trees is an old topic which has recently gained a lot of attention (compare, [MM70, Pan06, Jan06, DIMR09, Hol10, Ber12, BM13]). The main result of [Jan06] is the following. Given a finite-variance Galton-Watson tree conditioned to have N nodes, select an edge at random and delete the subtree above. Repeat the procedure until the root is isolated. Then the suitably rescaled number of cuts needed converges jointly with the rescaled tree to some random couple (Z_T, T). It is known that the limiting tree T is the Brownian CRT, while (unconditioned) Z_T is Rayleigh distributed. In a very recent paper, Abraham and Delmas [AD13] used a pruning with the length measure on the Brownian CRT (compare Example 4.2) and showed that given T, Z_T equals in distribution the averaged time it takes to separate a point from the root. The latter quantity was used in the proof given by Janson [Jan06]. In this example, we show that whenever bi-measure $\mathbb{R}$-trees converge – provided some extra tightness conditions hold – Janson's quantities converge as well.

Let $(\mathcal{G}_N, \mu_N, \nu_N)_{N\in\mathbb{N}\cup\{\infty\}}$ be a sequence of random bi-measure $\mathbb{R}$-trees such that

$$(4.22)\qquad \big(\mathcal{G}_N, \mu_N, \nu_N\big) \underset{N\to\infty}{\overset{\text{LWV}}{\Longrightarrow}} \big(\mathcal{G}_\infty, \mu_\infty, \nu_\infty\big).$$

For each $N \in \mathbb{N}\cup\{\infty\}$, let the pruning process $(X_t^N)_{t\geq 0}$ start in $X_0^N = (\mathcal{G}_N, \mu_N, \nu_N)$. Denote by Θ_N the *averaged time until a point gets separated from the root* ρ_N, where the average is taken with respect to the sampling measure μ_N. Given a realization $\mathcal{X} \in \mathbb{H}^{f,\sigma}$ of X_0^N, consider for each $u \in \operatorname{supp}(\mu_N)$ the (random) time $\mathcal{E}_{\mathcal{X}}^u$ until u gets separated from ρ_N, i.e., until a cut point falls on $[\rho_N, u[$. We abbreviate $\mathcal{E}_N^u := \mathcal{E}_{X_0^N}^u$ and obtain

$$(4.23)\qquad \Theta_N = \int_{\mathcal{G}_N} \mu_N(\mathrm{d}u)\, \mathcal{E}_N^u.$$

For all finite subsets $\{u_1, ..., u_n\} \subseteq \mathcal{G}_N$ and $t_1, ..., t_n \geq 0$, the distribution of $\mathcal{E}_N^{u_1}, \dots, \mathcal{E}_N^{u_n}$ is given by

$$(4.24)\qquad \mathbb{P}\big(\mathcal{E}_N^{u_1} \geq t_1, ..., \mathcal{E}_N^{u_n} \geq t_n \mid (\mathcal{G}_N, \mu_N, \nu_N)\big) = \prod_{l=1}^n e^{-t_{p(l)}\cdot \nu_N(S_l\setminus S_{l+1})},$$

where $p\colon \{1,2,...,n\} \to \{1,2,...,n\}$ is any permutation such that $t_{p(1)} \leq \dots \leq t_{p(n)}$, and $S_l := [\![u_{p(l)}, ..., u_{p(n)}]\!]$. Then for all $n \in \mathbb{N}$,

$$(4.25)\qquad \begin{aligned} \mathbb{E}\big[\Theta_N^n\big] &= \mathbb{E}\Big[\int_{\mathcal{G}_N^n} \mu_N^{\otimes n}(\mathrm{d}(u_1, ..., u_n))\, \mathbb{E}\big[\prod_{l=1}^n \mathcal{E}_N^{u_l} \mid (\mathcal{G}_N, \mu_N, \nu_N)\big]\Big] \\ &= n! \cdot \mathbb{E}\Big[\int_{\mathcal{G}_N^n} \mu_N^{\otimes n}(\mathrm{d}(u_1, ..., u_n)) \prod_{j=1}^n \frac{1}{\nu_N\big([\![u_1, ..., u_j]\!]\big)}\Big], \end{aligned}$$

where the last equality is obtained by using (4.24) and easy computations with the formula $\mathbb{E}\big[\prod_{i=1}^n X_i\big] = \int_{\mathbb{R}_+^n} \mathbb{P}(X_i > t_i, \forall i)\, \mathrm{d}(t_1 \dots t_n)$. Now assume the following:

(i) For all $n \in \mathbb{N}$, $\varepsilon > 0$ there is an $M > 0$ such that

$$(4.26)\qquad \sup_{N\in\mathbb{N}} \mathbb{E}\Big[\int_{\mathcal{G}_N^n} \mu_N^{\otimes n}(\mathrm{d}(u_1, ..., u_n)) \Big(\prod_{j=1}^n \frac{1}{\nu_N\big([\![u_1, ..., u_j]\!]\big)} - M\Big)^+\Big] \leq \varepsilon.$$

(ii) There is only one probability measure $\mathbb{Q}$ on $\mathbb{R}_+$ with moments

$$\int_{\mathbb{R}_+} \mathbb{Q}(\mathrm{d}\theta)\,\theta^n = n! \cdot \mathbb{E}\Big[\int \mu_\infty^{\otimes n}(\mathrm{d}(u_1,...,u_n)) \prod_{j=1}^{n} \frac{1}{\nu_\infty([\![u_1,...,u_j]\!])}\Big] \quad \forall n \in \mathbb{N}. \tag{4.27}$$

Note that these assumptions are in particular satisfied in the case of conditioned finite variance Galton-Watson trees converging to the Brownian CRT if ν_N is the length measure and μ_N the uniform distribution on the nodes (see, e.g., [Jan06, proof of Lem. 4.5, Thm. 1.9]).

For each $n, M \in \mathbb{N}$, define $\gamma_M^n \colon \mathbb{H}_n \to \mathbb{R}_+$ by

$$\gamma_M^n\big(T,(u_1,\dots,u_n),\nu\big) := M \wedge \prod_{j=1}^{n} \nu\big([\![u_1,\dots,u_j]\!]\big)^{-1}. \tag{4.28}$$

Then $\gamma_M^n \in \mathcal{C}_b(\mathbb{H}_n)$ if $\mathbb{H}_n$ is equipped with the sGw-topology, and the LWV-convergence (4.22) together with Proposition 2.17 implies that

$$\mathbb{E}\Big[\int \gamma_M^n \circ \tau_{(\mathcal{G}_N,\mu_N,\nu_N)}\,\mathrm{d}\mu_N^{\otimes n}\Big] \underset{N\to\infty}{\longrightarrow} \mathbb{E}\Big[\int \gamma_M^n \circ \tau_{(\mathcal{G}_\infty,\mu_\infty,\nu_\infty)}\,\mathrm{d}\mu_\infty^{\otimes n}\Big]. \tag{4.29}$$

Thus, we also have $\mathbb{E}\big[\Theta_N^n\big] \underset{N\to\infty}{\longrightarrow} \mathbb{E}\big[\Theta_\infty^n\big]$ for each $n \in \mathbb{N}$, provided that (4.26) holds. By assumption (ii), the moments of Θ_∞ determine its distribution, and the method of moments yields

$$\Theta_N \underset{N\to\infty}{\Longrightarrow} \Theta_\infty. \qquad \lozenge$$

References

[AD08] Romain Abraham and Jean-François Delmas, *Fragmentation associated with Lévy processes using snake*, Probability Theory and Related Fields **141** (2008), 113–154 (Anglais).

[AD12] Romain Abraham and Jean-François Delmas, *A continuum-tree-valued Markov process*, Ann. of Probab. **40** (2012), no. 3, 1167–1211.

[AD13] Romain Abraham and Jean-François Delmas, *Record process on the continuum random tree*, ALEA Lat. Am. J. Probab. Math. Stat. **10** (2013), no. 1, 225–251. MR 3083925

[ADH12] Romain Abraham, Jean-François Delmas, and Hui He, *Pruning Galton-Watson trees and tree-valued Markov process*, Ann. Inst. H. Poincaré Probab. Statist. **48** (2012), no. 3.

[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** (2013), no. 14, 1–21.

[ADH14] Romain Abraham, Jean-François Delmas, and Patrick Hoscheit, *Exit times for an increasing Lévy tree-valued process*, Probab. Theory Related Fields **159** (2014), no. 1-2, 357–403. MR 3201925

[ADV10] Romain Abraham, Jean-François Delmas, and Guillaume Voisin, *Pruning a Lévy continuum random tree*, Electron. J. Probab. **15** (2010), no. 46, 1429–1473.

[Ald91] David Aldous, *The continuum random tree I*, Ann. Probab. **19** (1991), 1–28.

[Ald93] ______, *The continuum random tree III*, Ann. Probab. **21** (1993), 248–289.

[ALW14] Siva Athreya, Wolfgang Löhr, and Anita Winter, *The gap between Gromov-vague and Gromov-Hausdorff-vague topology*, arXiv:1407.6309, 2014.

[AP98a] David Aldous and Jim Pitman, *The standard additive coalescent*, Ann. Probab **26(4)** (1998), 1703–1726.

[AP98b] ______, *Tree-valued Markov chains derived from Galton-Watson processes*, Ann. Inst. H. Poincaré Probab. Statist. **34(5)** (1998), 637–686.

[AS02] Romain Abraham and Laurent Serlet, *Poisson snake and fragmentation*, Elect. Journal of Probab. **7** (2002), 1–15.

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov, *A course in metric geometry, graduate studies in mathematics*, AMS, Boston, MA **33** (2001).

[Ber12] Jean Bertoin, *Fires on trees*, Ann. Inst. Henri Poincaré Probab. Stat. **48** (2012), no. 4, 909–921. MR 3052398

[Bil99] P. Billingsley, *Convergence of Probability Measures*, Wiley, New York, 1999.

[BK10] Douglas Blount and Michael A. Kouritzin, *On convergence determining and separating classes of functions*, Stochastic Processes and their Applications **120** (2010), no. 10, 1898–1907.

[BM13] Jean Bertoin and Grégory Miermont, *The cut-tree of large Galton-Watson trees and the Brownian CRT*, Ann. Appl. Probab. **23** (2013), no. 4, 1469–1493. MR 3098439

[Bog07] V. I. Bogachev, *Measure theory, volume I*, Springer, 2007.

[CH12] Nicolas Curien and Bndicte Haas, *The stable trees are nested*, Probability Theory and Related Fields (2012), 1–37 (English).

[DGP11] Andrej Depperschmidt, Andreas Greven, and Peter Pfaffelhuber, *Marked metric measure spaces*, Electron. Commun. Prob. **16** (2011), 174–188.

[DIMR09] M. Dromota, A. Iksanov, Martin Mohle, and Uwe Rosler, *A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree*, Random Structures and Algorithms **34** (2009), 319–336.

[DLG02] Thomas Duquesne and Jean-François Le Gall, *Random trees, Lévy processes and spatial branching processes*, Astérisque **281** (2002), vi+147. MR 1 954 248

[DMT96] Andreas W.M. Dress, V. Moulton, and W.F. Terhalle, *T-theory: An overview*, Europ. J. Combinatorics **17** (1996), no. 2-3.

[DT96] Andreas W.M. Dress and W.F. Terhalle, *The real tree*, Adv. Math. **120** (1996), 283–301.

[Dud02] M. Dudley, Richard, *Real analysis and probability*, Cambridge University Press, Cambridge, 2002.

[Duq03] Thomas Duquesne, *A limit theorem for the contour process of conditioned Galton-Watson trees*, The Annals of Probab. **31** (2003), no. 2, 996–1027.

[EK86] S.N. Ethier and T. Kurtz, *Markov Processes. Characterization and Convergence*, John Wiley, New York, 1986.

[EPW06] Steven N. Evans, Jim Pitman, and Anita Winter, *Rayleigh processes, real trees, and root growth with re-grafting*, Prob. Theo. Rel. Fields **134** (2006), no. 1, 81–126.

[Eva07] Steven N. Evans, *Probability and real trees*, École d'Été de Probabilités de Saint Flour XXXV-2005, Lecture Notes in Mathematics **1920** (2007), 1–193.

[EW06] Steven N. Evans and Anita Winter, *Subtree prune and re-graft: A reversible real-tree valued Markov chain*, Ann. Prob. **34** (2006), no. 3, 918–961.

[Fuk87] Kenji Fukaya, *Collapsing of Riemannian manifolds and eigenvalues of Laplace operator*, Inventiones Math. **87** (1987), no. 3, 517–547.

[GPW09] Andreas Greven, Peter Pfaffelhuber, and Anita Winter, *Convergence in distribution of random metric measure spaces (the Λ-coalescent measure tree)*, Probab. Theory Related Fields **145** (2009), no. 1-2, 285–322.

[Gro99] Misha Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, vol. 152, Birkhäuser Boston Inc., Boston, MA, 1999.

[HJ77] J. Hoffmann-Jørgensen, *Probability in Banach spaces*, École d'Été de Probabilités de Saint Flour VI-1976, Lecture Notes in Mathematics, vol. 598, Springer, 1977.

[Hol10] C. Holmgren, *Random records and cuttings in binary search trees*, Combinatorics, Probability and Computing **19** (2010), 391–424.

[Jan06] S. Janson, *Random cuttings and records in deterministic and random trees*, Random Structures Algorithms **29** (2006), 139–179.

[LC57] Lucien Le Cam, *Convergence in distribution of stochastic processes*, Univ. California Publ. Statist. **2** (1957), 207–236.

[Löh13] Wolfgang Löhr, *Equivalence of Gromov-Prohorov- and Gromov's $\square_\lambda$-metric on the space of metric measure spaces*, Electron. Commun. Probab. **18** (2013), no. 17, 10. MR 3037215

[Lyo92] Russel Lyons, *Random walks, capacities, and percolation on trees*, Annals of Probability **20** (1992), 2043–2088.

[Mie03] Grégory Miermont, *Self-similar fragmentations derived from the stable tree I: splitting at heights*, Probability Theory and Related Fields **127** (2003), 423–454 (Anglais), 30 pages.

[Mie05] Grégory Miermont, *Self-similar fragmentations derived from the stable tree II: splitting at nodes*, Probability Theory and Related Fields **131** (2005), 341–375 (English).

[Mie09] Grégory Miermont, *Tessellations of random maps of arbitrary genus*, Ann. Sci. Ec. Norm. Sup **42** (2009), 725–781 (Anglais).

[MM70] A. Meir and J.W. Moon, *Cutting down random trees*, J. Australian Math. Soc. **11** (1970).

[Pan06] A. Panholzer, *Cutting down very simple trees*, Quaestiones Mathematicae **29** (2006), 211–227.

[Vil09] Cédric Villani, *Optimal transport*, Grundlehren der mathematischen Wissenschaften, vol. 338, Springer, Berlin-Heidelberg, 2009.

[Voi11] Guillaume Voisin, *Dislocation measure of the fragmentation of a general Lévy tree*, ESAIM Probab. Stat. **15** (2011), 372–389. MR 2870521

BOUNDEDLY FINITE MEASURES: SEPARATION AND CONVERGENCE BY AN ALGEBRA OF FUNCTIONS

WOLFGANG LÖHR, THOMAS RIPPL

ABSTRACT. We prove general results about separation and weak$^{\#}$-convergence of boundedly finite measures on separable metric spaces and Souslin spaces. More precisely, we consider an algebra of bounded real-valued, or more generally a $*$-algebra $\mathcal{F}$ of bounded complex-valued functions and give conditions for it to be separating or weak$^{\#}$-convergence determining for those boundedly finite measures that integrate all functions in $\mathcal{F}$. For separation, it is sufficient if $\mathcal{F}$ separates points, vanishes nowhere, and either consists of only countably many measurable functions, or of arbitrarily many continuous functions. For convergence determining, it is sufficient if $\mathcal{F}$ induces the topology of the underlying space, and every bounded set A admits a function in $\mathcal{F}$ with values bounded away from zero on A.

CONTENTS

1. INTRODUCTION

Boundedly finite measures play an increasingly important role in probability theory. Classical examples are Itô excursion measures, or Lévy measures, which we will come back to in a moment. But they also appear in very recent research as speed measures of Brownian motions on $\mathbb{R}$-trees [AEW13, ALW15] or sampling measures for spatial Fleming-Viot processes [GSW15]. Convergence of metric measure spaces with boundedly finite measures has been analysed with a view towards probabilistic applications in [ALW16].

For the purpose of illustration, we quickly recall the situation for Lévy measures of infinitely divisible random variables. A real-valued random variable X is called *infinitely divisible* if, for any $n \in \mathbb{N}$, we can write $X \stackrel{\mathrm{d}}{=} X_{1,n} + \cdots + X_{n,n}$ for an i.i.d. sequence of random variables $(X_{i,n})_{1\le i\le n}$. A famous theorem by Lévy and Khintchine states that in that case the Fourier transform of the random variable has a very explicit form:

$$(1) \qquad -\log \mathbb{E}\big[\exp(-iuX)\big] = iub + \frac{c}{2}u^2 - \int_{\mathbb{R}\setminus\{0\}} e^{-iuy} - 1 + iuy\mathbb{1}_{|y|\le 1}\,\mu(\mathrm{d}y), \quad u \in \mathbb{R},$$

where $b \in \mathbb{R}$, $c \ge 0$ and μ is a measure on $\mathbb{R}\setminus\{0\}$ satisfying $\int_{\mathbb{R}\setminus\{0\}} 1 \wedge |y|^2\,\mu(\mathrm{d}y) < \infty$, see Theorem 8.1 of [Sat99]. Let us look at the measure μ, called *Lévy measure*. The obvious questions are: (a) is the Lévy measure unique?[1] (b) If we have a sequence of infinitely divisible random variables, do their Lévy measures converge and in which sense? Lévy measures are not necessarily

[1]this question has an affirmative answer with a neat proof in [BCR76] Theorem 3.7

finite measures, but are required to be finite on any set which is not close to the origin 0. This motivates to consider measures which are finite on a certain class of sets. Daley and Vere-Jones used Appendix A2.6 in [DVJ03] to present a framework for such questions which they call *boundedly finite measures*, because the measures are assumed to be finite on bounded sets. The space of these measures is equipped with weak$^{\#}$-convergence, defined as convergence of integrals over bounded continuous functions with bounded support (see Section 2 for definitions). Some extensions are given in [HL06] and [LRR14]. Lévy measures fit into this framework if we change the Euclidean metric on $\mathbb{R} \setminus \{0\}$ such that 0 is sent infinitely far away, an idea also used in [BP06].

How does one prove weak convergence $\mu_n \xrightarrow{w} \mu$ of probability measures on a topological space X in situations where it is not feasible to show convergence of $\int f \, d\mu_n$ for *all* bounded continuous $f \in \mathcal{C}_b(X)$ directly? One possibility is to find a class $\mathcal{F} \subseteq \mathcal{C}_b(X)$ of sufficiently "nice" functions, which is still rich enough to be *convergence determining*, i.e.

$$\int f \, d\mu_n \to \int f \, d\mu \; \forall f \in \mathcal{F} \quad \Longrightarrow \quad \mu_n \xrightarrow{w} \mu.$$

This approach has proven to be particularly fruitful if the topology on X itself is defined in terms of a class of functions. A classical example for such a topology is the weak (weak-$*$) topology on (the dual of) a Banach space. A more modern one is the Gromov-weak topology on the space of metric measure spaces, which is induced both by a complete metric and by a class of functions called *polynomials* (see [GPW09]). That polynomials do not only induce the topology but are even convergence determining was shown with some effort in [DGP11]. But it also follows directly from a general result due to Le Cam, as pointed out in [Löh13]. Le Cam's theorem goes back to [LC57] and states that on a completely regular Hausdorff space X, a set of functions $\mathcal{F} \subseteq \mathcal{C}_b(X)$ is convergence determining for Radon probability measures if it is multiplicatively closed and induces the topology of X. The proof can be found in [HJ77, Proposition 4.1]. A version of Le Cam's theorem for separable metric spaces dropping the "Radon" assumption on the probability measures is given in [BK10]. This version was used extensively for the construction of a tree-valued pruning process in [LVW15].

Our main goal is to extend Le Cam's result to the case of boundedly finite measures and weak$^{\#}$-convergence and, because convergence determining is sometimes too much to ask for, to obtain (weaker) sufficient conditions for $\mathcal{F}$ to at least *separate* boundedly finite measures. A separating class of functions can also be used to prove weak (or weak$^{\#}$) convergence if tightness is known by other methods. In particular, our results allow to give an answer on the question of uniqueness and convergence of μ in (1) within a general framework. More importantly, they will find applications in future work about spaces of metric measure spaces and $\mathbb{R}$-trees such as in the upcoming paper [GGR] which was a driving motivation for the present article. The results will hopefully also facilitate the analysis of spatial population models on unbounded spaces with infinite total population size such as the one in [GSW15], and of other models appearing in modern areas of probability theory.

The rest of the paper is organized as follows. In Section 2, we give our main results about convergence determining (Theorem 2.3) and separating (Theorem 2.7 and Corollary 2.8) classes of functions for boundedly finite measures. In Section 3, we illustrate in four examples how our results can be applied. There, we consider Lévy measures, excursion theory and mass fragmentations.

2. Separation and convergence of boundedly finite measures

Let (X, d) be a separable metric space, endowed with the Borel σ-field induced by d. By $\mathcal{C}_b(X)$, we denote the set of bounded continuous functions on the metric space (X, d) with values in $\mathbb{C}$. For real-valued functions we write $\mathcal{C}_b(X; \mathbb{R})$. Note that $\mathcal{C}_b(X; \mathbb{R}) \subseteq \mathcal{C}_b(X)$.

Definition 2.1 (Boundedly finite measures and weak$^{\#}$-convergence)**.** The set of *boundedly finite measures* $\mathcal{M}^{\#}(X)$ on X w.r.t. d is given as

$$\mathcal{M}^{\#}(X) = \{\mu \in \mathcal{M}(X) \mid \mu(A) < \infty \text{ for all } d\text{-bounded, measurable } A \subseteq X\}\,.$$

A sequence $(\mu_n)_{n\in\mathbb{N}}$ in $\mathcal{M}^{\#}(X)$ is said to be *weak$^{\#}$-convergent* to $\mu \in \mathcal{M}^{\#}(X)$, denoted by $\mu_n \xrightarrow{\mathrm{w}^{\#}} \mu$, if $\int f \,\mathrm{d}\mu_n \to \int f, \mathrm{d}\mu$ holds for all $f \in \mathcal{C}_b(X;\mathbb{R})$ with d-bounded support. ♢

Remark 2.2 (Weak$^{\#}$-convergence versus vague convergence)**.** If (X,d) is a *Heine-Borel space*, i.e. every closed, bounded set is compact, then $\mathcal{M}^{\#}(X)$ coincides with the set of Radon measures on X, and weak$^{\#}$-convergence with vague convergence. For a general separable metric space, however, $\mathcal{M}^{\#}(X)$ is a subset of the Radon measures and weak$^{\#}$-convergence is a potentially much stronger convergence than vague convergence. ♣

Consider a set $\mathcal{F}$ of measurable, $\mathbb{C}$-valued functions on X and define

$$\mathcal{M}_{\mathcal{F}}(X) = \Big\{\mu \in \mathcal{M}(X) \;\Big|\; \int |f(x)|\,\mu(\mathrm{d}x) < \infty \;\; \forall f \in \mathcal{F}\Big\}, \quad \mathcal{M}^{\#}_{\mathcal{F}}(X) = \mathcal{M}^{\#}(X) \cap \mathcal{M}_{\mathcal{F}}(X). \tag{2}$$

Theorem 2.3 (Convergence determining for boundedly finite measures)**.** *Let (X,d) be a separable metric space and $\mathcal{F} \subset \mathcal{C}_b(X)$. Assume that*

(**T.1**) *$\mathcal{F}$ is multiplicatively closed and closed under complex conjugation.*
(**T.2**) *$\mathcal{F}$ induces the topology of X.*
(**T.3**) *For every bounded set $A \subset X$ there exists $f \in \mathcal{F}$ and $\delta > 0$ with $\inf_{x\in A} |f(x)| > \delta$.*

Then $\mathcal{F}$ is weak$^{\#}$-convergence determining for measures in $\mathcal{M}^{\#}_{\mathcal{F}}(X)$, i.e.

$$\mu, \mu_n \in \mathcal{M}^{\#}_{\mathcal{F}}(X),\; \int f \,\mathrm{d}\mu_n \to \int f \,\mathrm{d}\mu \;\forall f \in \mathcal{F} \quad\Longrightarrow\quad \mu_n \xrightarrow{\mathrm{w}^{\#}} \mu.$$

Remark 2.4. (1) (T.1) and (T.2) are classical assumptions for these kind of theorems, see Proposition 4.1 in [HJ77]. Something like (T.3) is necessary to replace the fixed total mass in the weak convergence of probabilities. At least, we have to ensure that $\mathcal{F}$ "vanishes nowhere", because if there was $x \in X$ with $f(x) = 0$ for all $f \in \mathcal{F}$, then $\mathcal{F}$ could not even separate $a \cdot \delta_x$ for different $a \geq 0$. For the purpose of separation of measures, we can do with this weaker requirement S.3 in Theorem 2.7 below. We do not know, however, if it would be enough for Theorem 2.3.
(2) For real-valued functions, the part "closed under complex conjugation" is always satisfied. ♣

For the proof, we embed everything in the Hilbert cube H, a technique going back to Urysohn's work on metrisation and also used in [BK10]. Recall that

$$H = [0,1]^{\mathbb{N}}, \quad \text{with product topology.}$$

Denote the uniform norm on $\mathcal{C}_b(H;\mathbb{R})$ by $\|\cdot\| := \|\cdot\|_\infty$. For $0 < \delta < 1$, we consider the subspace

$$H_\delta := \big\{x = (x_n)_{n\in\mathbb{N}} \in H \;\big|\; x_1 \geq \delta\big\}$$

and use the following variant of the Stone-Weierstrass theorem.

Definition 2.5 ($\mathcal{P}$, $\mathcal{P}_0$)**.** Let $\mathcal{P} \subseteq \mathcal{C}_b(H;\mathbb{R})$ be the set of polynomials on H (i.e. functions depending on finitely many coordinates and an algebraic multivariate polynomial in these coordinates). Let $\mathcal{P}_0 := \big\{p \in \mathcal{P} \;\big|\; p(x) = 0 \;\forall x = (x_n)_{n\in\mathbb{N}} \in H \text{ with } x_1 = 0\big\}$. ♢

Lemma 2.6 (Stone-Weierstrass variant)**.** *Let $g\colon H \to [0,1]$ be continuous with $\operatorname{supp} g \subset H_\delta$ for some $\delta > 0$. Then, for every $\varepsilon > 0$ there exists a polynomial $p_\varepsilon \in \mathcal{P}_0$ such that*

$$|g(x) - p_\varepsilon(x)| \leq \varepsilon x_1 \quad \forall x = (x_n)_{n\in\mathbb{N}} \in H.$$

Proof. For $x = (x_n)_{n\in\mathbb{N}} \in H$ define

$$\tilde{g}(x) := \begin{cases} x_1^{-1} g(x) & \text{if } x_1 > 0, \\ 0 & \text{if } x_1 = 0. \end{cases}$$

Then $\tilde{g} \in \mathcal{C}_b(H;\mathbb{R})$ and we can use the Stone-Weierstrass theorem. So, for $\varepsilon > 0$ we find $\tilde{p}_\varepsilon \in \mathcal{P}$ such that

$$|\tilde{g}(x) - \tilde{p}_\varepsilon(x)| < \varepsilon \ \forall x \in H.$$

Define $p_\varepsilon(x) := x_1 \tilde{p}_\varepsilon(x)$, $x \in H$. Then $p_\varepsilon \in \mathcal{P}_0$ and we get for any $a \in (0,1]$ the estimate

$$\sup_{x\in H,\, x_1=a} a^{-1}|g(x) - p_\varepsilon(x)| = \sup_{x\in H,\, x_1=a} |\tilde{g}(x) - \tilde{p}_\varepsilon(x)| < \varepsilon.$$

That is what we needed to show, as the case $x_1 = 0$ is trivial. □

Proof of Theorem 2.3. Step 1. It is enough to consider functions with values in $[0,1]$: First, $\mathcal{F}$ may be replaced by $\mathcal{F}' := \{\Re(f), \Im(f) \mid f \in \mathcal{F}\}$, where $\Re(f)$ and $\Im(f)$ are the real and imaginary parts of f, respectively. Because $\mathcal{F}$ is closed under complex conjugation due to (T.1), the conditions (T.1), (T.2) and (T.3) are also satisfied for $\mathcal{F}'$ instead of $\mathcal{F}$. Thus we may assume $\mathcal{F}$ to consist of real-valued functions. Second, $\mathcal{F}$ may be replaced by $\mathcal{F}' := \{f^2 \mid f \in \mathcal{F}\} \cup \{f^2(\|f\|_\infty - f) \mid f \in \mathcal{F}\}$, which is contained in the vector space generated by $\mathcal{F}$ and easily seen to satisfy the prerequisites of the theorem provided that $\mathcal{F}$ does. Because $\mathcal{F}'$ maps to $\mathbb{R}_+$, we can assume, by normalisation, that the elements of $\mathcal{F}$ map into $[0,1]$.

Step 2. By Assumption (T.2), $\mathcal{F}$ induces the topology of X. Because X is a separable, metric space, it has a countable basis, and thus we can choose a countable subfamily of $\mathcal{F}$ that still induces the topology of X. Indeed, the family of sets of the form $\bigcap_{i=1}^n f_i^{-1}(U_i)$ for $n \in \mathbb{N}$, $f_i \in \mathcal{F}$, $U_i \subseteq [0,1]$ open is a base for the topology, and because X has a countable base, we can select a countable subfamily that is still a base[2]. Therefore there exist $f_1, f_2, \ldots \in \mathcal{F}$ with $0 \le f_m \le 1$, such that $(f_m)_{m\in\mathbb{N}}$ induces the topology of X. Then $\iota\colon X \to H$, $x \mapsto (f_m(x))_{m\in\mathbb{N}}$ is a topological embedding (i.e. a homeomorphism onto its image) of X into H. Identifying X with $\iota(X)$, we assume w.l.o.g. $X \subseteq H$ and f_n to be the (restriction of the) n^{th} coordinate projection. In particular, being an algebra, the linear span of $\mathcal{F}$ contains $\mathcal{P}_0$ (defined in Definition 2.5).

Step 3. Let $\mu_n, \mu \in \mathcal{M}^{\#}_{\mathcal{F}}(X)$ with

$$\int f \,\mathrm{d}\mu_n \to \int f \,\mathrm{d}\mu \quad \forall f \in \mathcal{F}. \tag{3}$$

For the claimed weak$^{\#}$-convergence, it is enough to show $\int g \,\mathrm{d}\mu_n \to \int g \,\mathrm{d}\mu$ for all $g\colon X \to [0,1]$ which have d-bounded support and are uniformly continuous (by the Portmanteau theorem, Theorem 2.1 in [LRR14]). Because weak convergence depends on the metric only through the induced topology, we may assume g to be uniformly continuous w.r.t. any other metric on X inducing the same topology as d. To this end, we take any metric on H inducing its topology, and assume that g is uniformly continuous w.r.t. its restriction to X (recall that X is a subspace of H by Step 2). By Assumption (T.3), there is $f \in \mathcal{F}$ and $\delta > 0$ such that for all $x \in \operatorname{supp}(g)$ we have $|f(x)| > \delta$. We may assume w.l.o.g. that $f = f_1$ (if not, we define $f_1' = f$, $f_{m+1}' = f_m$ and observe that ι' defined with these f_m' is still an embedding and g uniformly continuous w.r.t. the restriction of the metric on H). Then $\operatorname{supp}(g) \subseteq H_\delta$. Furthermore, since g is uniformly continuous, it can be extended continuously to the closure of X in H, and by the Tietze extension theorem (e.g.

[2] One can show this with standard arguments: If $\mathcal{B}, \mathcal{B}'$ are bases, $\mathcal{B}$ countable, $B \in \mathcal{B}$, then $B = \bigcup I$ for some $I \subseteq \mathcal{B}'$, and for every $U \in I$ there is $J_U \subseteq \mathcal{B}$ with $U = \bigcup J_U$. For every $V \in J := \bigcup_{U\in I} J_U$, we select one $U_V \in I$ with $V \in J_{U_V}$. J is a subset of $\mathcal{B}$, hence countable. Because $B = \bigcup J = \bigcup_{V\in J} U_V$, we obtain a countable basis by taking all U_V for all joices of $B \in \mathcal{B}$.

[Mun00, Theorem 35.1]) to a continuous function from H to $[0,1]$ with support in H_δ. We denote the extension again by g. We also identify μ_n and μ with their natural extensions to H.

Step 4. g satisfies the assumptions of Lemma 2.6. For $\varepsilon > 0$, choose $p_\varepsilon \in \mathcal{P}_0 \subseteq \operatorname{span}(\mathcal{F})$ as in the lemma. Because $\mu_n(f_1) \to \mu(f_1)$, we have $M := \sup_{n\in\mathbb{N}} \mu_n(f_1) < \infty$ and obtain for all $n \in \mathbb{N}$

$$\left|\int p_\varepsilon - g \,\mathrm{d}\mu_n\right| \le \int_H \varepsilon x_1 \,\mu_n(\mathrm{d}x) = \varepsilon \int f_1 \,\mathrm{d}\mu_n \le \varepsilon M.$$

Because $\mu_n(p_\varepsilon) \to \mu(p_\varepsilon)$ by (3), we conclude for every $\varepsilon > 0$

$$\limsup_{n\to\infty} |\mu_n(g) - \mu(g)| \le \limsup_{n\to\infty} |\mu_n(p_\varepsilon) - \mu(p_\varepsilon)| + |\mu_n(p_\varepsilon - g)| + |\mu(p_\varepsilon - g)| \le 2\varepsilon M.$$

Because ε is arbitrary, the claimed convergence follows. □

It is desirable to have a result which separates two boundedly finite measures but requires less than the previous theorem. While "boundedly finite" is essential for the definition of weak$^\#$-convergence, we can drop this assumption for the purpose of separation and work with $\mathcal{M}_\mathcal{F}$, the space of measures integrating $\mathcal{F}$ as defined in (2), instead of $\mathcal{M}^\#_\mathcal{F}$. We can also relax the metrisability assumption on the space X, but do need some topological assumption. Recall that a topological space is Hausdorff if any two distinct points can be separated by open sets, and a Hausdorff topological space X is, by definition, a *Souslin space* if there exists a Polish space Y and a continuous surjective map from Y onto X. Note that a Souslin space is separable but need not be metrisable. An example is a separable Banach space in its weak topology, which is clearly Souslin but not metrisable. Conversely, not every separable metrisable space is Souslin. In the case of a Souslin space X, and a countable family of functions $\mathcal{F}$, we can drop the topological assumptions on $\mathcal{F}$ from the prerequisites of Theorem 2.3 and still obtain the weaker conclusion that $\mathcal{F}$ is separating for measures in $\mathcal{M}_\mathcal{F}$. More precisely, we have

Theorem 2.7 (Separation of boundedly finite measures with measurable functions)**.** *Let X be a Souslin space (for example a Polish space), and $\mathcal{F}$ a* countable *set of bounded, measurable $\mathbb{C}$-valued functions. Assume that*

(**S.1**) *$\mathcal{F}$ is multiplicatively closed and closed under complex conjugation.*
(**S.2**) *$\mathcal{F}$ separates points of X.*
(**S.3**) *$\mathcal{F}$ vanishes nowhere, i.e. for every $x \in X$ there exists an $f_x \in \mathcal{F}$ with $f_x(x) \neq 0$.*

Then $\mathcal{F}$ is separating for measures in $\mathcal{M}_\mathcal{F}(X)$, i.e.

$$\mu_1, \mu_2 \in \mathcal{M}_\mathcal{F}(X),\ \int f \,\mathrm{d}\mu_1 = \int f \,\mathrm{d}\mu_2 \ \forall f \in \mathcal{F} \quad \Longrightarrow \quad \mu_1 = \mu_2. \tag{4}$$

Proof. Assume that $\mu_1, \mu_2 \in \mathcal{M}_\mathcal{F}(X)$ are such that $\int f \,\mathrm{d}\mu_1 = \int f \,\mathrm{d}\mu_2$ holds for all $f \in \mathcal{F}$. We have to show $\mu_1 = \mu_2$. Enumerate $\mathcal{F} = \{f_n \mid n \in \mathbb{N}\}$. Using Step 1 from the proof of Theorem 2.3, we may (and do) assume that f_n takes values in $[0,1]$ for all $n \in \mathbb{N}$. We proceed in two steps: first we show that (4) holds if we assume that instead of (S.3) the following stronger condition holds,

(**S.3†**) $f_1(x) \neq 0$ for all $x \in X$.

In the second step, we reduce the general case to the one where (S.3†) holds.

Step 1. Assume that (S.3†) holds and define $\iota\colon X \to H$, $x \mapsto \big(f_n(x)\big)_{n\in\mathbb{N}}$. Then ι is measurable and injective by assumption (S.2). Because X is a Souslin space, it is an analytic measurable space ([Coh80, Proposition 8.6.13]) and so is $Y := \iota(X)$ ([Coh80, Corollary 8.6.9]). By [Coh80, Proposition 8.6.2], ι is a Borel isomorphism onto Y, i.e. $\iota^{-1}\colon Y \to X$ is measurable. Therefore,

$$\mu_1 = \mu_2 \iff \mu_1 \circ \iota^{-1} = \mu_2 \circ \iota^{-1}. \tag{5}$$

Because of (S.3†), every $x = (x_n)_{n\in\mathbb{N}} \in Y$ satisfies $x_1 \neq 0$. We define the metric

$$r(x,y) := |x_1^{-1} - y_1^{-1}| + \sum_{n\in\mathbb{N}} 2^{-n}|x_n - y_n| \wedge 1, \qquad x, y \in Y,$$

which induces on Y the topology inhereted as a subspace of H. Let $\mathcal{G} := \{f \circ \iota^{-1} \mid f \in \mathcal{F}\}$. We show that (Y, r) and $\mathcal{G}$ satisfy the prerequisites of Theorem 2.3.

$\mathcal{G}$ satisfies (T.1) because $\mathcal{F}$ does by assumption. By construction of ι, $\mathcal{G}$ coincides with the set of restrictions of coordinate projections to Y. Therefore, $\mathcal{G}$ induces the topology of Y, i.e. (T.2) is satisfied. An r-bounded set A in Y satisfies $\delta := \inf_{x \in A} |x_1| > 0$, and (T.3) is satisfied with $f = f_1 \circ \iota^{-1}$. Thus we can apply Theorem 2.3 and obtain that $\mathcal{G}$ is weak$^{\#}$-convergence determining and a fortiori separating for measures in $\mathcal{M}^{\#}_{\mathcal{G}}(Y)$.

If $\mu_1, \mu_2 \in \mathcal{M}_{\mathcal{F}}(X)$, then they integrate f_1, and $\hat{\mu}_i := \mu_i \circ \iota^{-1}$, $i = 1, 2$, are boundedly finite measures on (Y, r). Thus obviously $\hat{\mu}_i \in \mathcal{M}^{\#}_{\mathcal{G}}(Y)$ and the claim of the theorem follows with (5).

Step 2. Now consider the general case, where (S.3†) does not necessarily hold. Define

$$f_0 := \sum_{n \in \mathbb{N}} 2^{-n} \frac{f_n}{1 \vee \int f_n \, \mathrm{d}\mu_1},$$

and let $\mathcal{F}'$ be the set of finite products of elements of $\mathcal{F} \cup \{f_0\}$. Then $\mathcal{F}'$ is a countable set of measurable functions satisfying (S.1), (S.2) and, because $\mathcal{F}$ vanishes nowhere, also (S.3†) (with f_1 replaced by $f_0 \in \mathcal{F}'$). Hence, by Step 1, $\mathcal{F}'$ is separating for measures in $\mathcal{M}_{\mathcal{F}'}(X)$.

According to the monotone convergence theorem, we have

$$\int f_0 \, \mathrm{d}\mu_1 = \int f_0 \, \mathrm{d}\mu_2 \leq 1.$$

Because every element of $\mathcal{F}'$ is dominated by an element of $\mathcal{F} \cup \{f_0\}$ (recall that elements of $\mathcal{F}'$ map to $[0, 1]$), this implies $\mu_1, \mu_2 \in \mathcal{M}_{\mathcal{F}'}(X)$. Moreover, dominated convergence yields

$$\int g \, \mathrm{d}\mu_1 = \int g \, \mathrm{d}\mu_2 \quad \forall g \in \mathcal{F}',$$

which implies $\mu_1 = \mu_2$ by Step 1. □

In the case of *continuous* functions, we can drop the countability of $\mathcal{F}$.

Corollary 2.8 (Separation of boundedly finite measures with continuous functions)**.** *Let X be a Souslin space (e.g. a Polish space), and $\mathcal{F} \subseteq \mathcal{C}_b(X)$. Assume* (S.1), (S.2), *and* (S.3) *from Theorem 2.7. Then $\mathcal{F}$ is separating for measures in $\mathcal{M}_{\mathcal{F}}(X)$, i.e.* (4) *holds.*

Proof. For $x \in X$, let f_x be as in (S.3). There is an open neighbourhood U_x of x with $f_x(y) \neq 0$ for all $y \in U_x$. Recall that a topological space is called Lindelöf if every open cover has a countable subcover, and every Polish space has this property (because it has a countable base). Because the property is obviously preserved by continuous maps, every Souslin space is Lindelöf as well. Hence $(U_x)_{x \in X}$ has a countable subcover, and there exists a countable subfamily $\mathcal{F}_1$ of $\mathcal{F}$ satisfying (S.3).

Similarly, for $x, y \in X$ let $f_{xy} \in \mathcal{F}$ be such that $f_{xy}(x) \neq f_{xy}(y)$. Then there is an open neighbourhood U_{xy} of (x, y) in X^2 with $f_{xy}(u) \neq f_{xy}(v)$ for all $(u, v) \in U_{xy}$. Because X^2 is also Souslin and hence Lindelöf, we find a countable subfamily $\mathcal{F}_2$ of $\mathcal{F}$ satisfying (S.2). Let $\mathcal{F}'$ be the closure of $\mathcal{F}_1 \cup \mathcal{F}_2$ under multiplication and complex conjugation. Then $\mathcal{F}'$ satisfies the prerequisites of Theorem 2.7 and the conclusion follows. □

3. Examples

3.1. **Example 1: Lévy-Khintchine formula on $\mathbb{R}^D$.** Let Z be an infinitely divisible random variable with values in $\mathbb{R}^D$ for $D \in \mathbb{N}$. That means for any $n \in \mathbb{N}$ there are i.i.d. random variables $Z_{1,n}, \ldots, Z_{n,n}$ such that $Z \overset{\mathrm{d}}{=} Z_{1,n} + \cdots + Z_{n,n}$. Consider

$$X := \mathbb{R}^D \setminus \{0\} \quad \text{with metric} \quad d(x, y) := \|x - y\|_\infty + \Big| \|x\|_\infty^{-1} - \|y\|_\infty^{-1} \Big|, \; x, y \in X.$$

It is well-known that there exist $b \in \mathbb{R}^D$, $C \in S_+(\mathbb{R}^D)$ a symmetric, positive semidefinite matrix, and $\mu \in \mathcal{M}^\#(X)$ with $\int (1 \wedge \|x\|_\infty^2)\, \mu(\mathrm{d}x) < \infty$ such that

$$\text{(6)} \quad \Psi(u) := \log \mathbb{E}[\exp(iu^t Z)] = iu^t b - \frac{1}{2} u^t C u + \int_{\mathbb{R}^D \setminus \{0\}} \exp(iu^t x) - 1 - iu^t x \mathbb{1}_{|x| \le 1}\, \mu(\mathrm{d}x),\ u \in \mathbb{R}^D.$$

This formula is called the Lévy-Khintchine formula, see [Sat99, Theorem 8.1]. The function $\mathbb{R}^D \to \mathbb{C}$, $u \mapsto \Psi(u)$ characterizes the distribution of the random vector Z. On a first glance, however, it is not clear why the Lévy triple (b, C, μ) should be unique. Theorem 2.3 allows to give simple verification of that known fact in a general setup.

Proposition 3.1 (Uniqueness and convergence of Lévy measures)**.** (6) *determines the Lévy triple* $(b, C, \mu) \in \mathbb{R}^D \times S_+(\mathbb{R}^D) \times \mathcal{M}^\#(X)$ *uniquely. Furthermore, if* Z_n *are infinitely divisible random variables converging in distribution to* Z*, and* (b_n, C_n, μ_n) *is the Lévy triple of* Z_n*, then* $\mu_n \xrightarrow{\mathrm{w}^\#} \mu$*.*

Proof. We start with the unique identification of the law.

Step 1. First note that

$$C_{k,j} = -\lim_{m \to \infty} m^{-2} \left[\Psi(m(\mathrm{e}_k + \mathrm{e}_j)) - \Psi(m\mathrm{e}_k) - \Psi(m\mathrm{e}_j) \right],\ 1 \le k, j \le D,$$

where e_k, $k = 1, \dots, D$ are unit vectors in $\mathbb{R}^D$. Moreover, for $k = 1, \dots, D$:

$$b_k = \lim_{m \to \infty} -\frac{i}{m} \left(\Psi(m\mathrm{e}_k) - \frac{1}{2} m^2 C_{k,k} \right).$$

Hence C and b are unique.

Step 2. Now suppose $\mu_1, \mu_2 \in \mathcal{M}^\#(X)$ both satisfy $\int (1 \wedge \|x\|_\infty^2)\, \mu_i(\mathrm{d}x) < \infty$ and (6) with μ replaced by μ_i, $i = 1, 2$.

For $u \in \mathbb{R}^D$, define $F_u, \psi_u, G_u \colon \mathbb{R}^D \to \mathbb{C}$ by

$$\text{(7)} \qquad F_u(x) := \exp(iu^t x) - 1, \qquad \psi_u(x) := iu^t x \mathbb{1}_{|x| \le 1}, \qquad G_u(x) := F_u(x) - \psi_u(x),\ x \in \mathbb{R}^D$$

and consider the following two classes of functions, where span denotes the linear span:

$$\mathcal{G} := \mathrm{span}\{G_u \mid u \in \mathbb{R}^D\}, \qquad \mathcal{F} := \mathrm{span}\{F_u \cdot F_v \mid u, v \in \mathbb{R}^D\}.$$

Then with (6) and the uniqueness of b and C from Step 1, we have $\int G\, \mathrm{d}\mu_1 = \int G\, \mathrm{d}\mu_2$ for all $G \in \mathcal{G}$. Now observe that, using linearity of $u \mapsto \psi_u(x)$ for every $x \in \mathbb{R}^D$,

$$F_u \cdot F_v = F_{u+v} - F_u - F_v = G_{u+v} - G_u - G_v \in \mathcal{G} \qquad \forall u, v \in \mathbb{R}^D.$$

Hence, $\mathcal{F}$ is multiplicatively closed (by the first equality) and $\mathcal{F} \subseteq \mathcal{G}$. In particular, (S.1) holds,

$$\int f\, \mathrm{d}\mu_1 = \int f\, \mathrm{d}\mu_2 \qquad \forall f \in \mathcal{F},$$

and $\mu_1, \mu_2 \in \mathcal{M}^\#_{\mathcal{F}}$, because functions from $\mathcal{G}$ are integrable. Furthermore, $\mathcal{F}$ is contained in $\mathcal{C}_b(X)$ and (S.2) and (S.3) are easily verified. Thus $\mu_1 = \mu_2$ follows from Corollary 2.8.

Now we show the convergence result.

Step 3. First, $\Psi_n(u) = \log \mathbb{E}[\exp(iu^t Z_n)] \to \Psi(u)$ pointwise since $x \mapsto \exp(iu^t x)$ is a bounded continuous function.

Step 4. Recall that for the Lévy-triple (b, C, μ) the linear part b depends on the choice of the compensation function ψ_u, but C and μ do not. So in order to show $\mu_n \xrightarrow{\mathrm{w}^\#} \mu$ we may choose any (admissible) ψ_u we like. Replace ψ_u in (7) by $\hat{\psi}_u(x) = iu^t x h(x)$ for a $\mathcal{C}^1$-function $h \colon \mathbb{R}^D \to \mathbb{R}$ with $h(0) = 1$ and compact support. Then the argument from above still works: $\mathcal{F}$ is multiplicatively closed and $\mathcal{F} \subset \mathcal{C}_b(X)$, so (T.1) holds.

Moreover, Assumption (T.2) holds by the fact that $F_u(x_n) \to F_u(x) \Leftrightarrow \exp(iu^t x_n) \to \exp(iu^t x)$ for all $u \in \mathbb{R}^D$. The latter is nothing else than the convergence of the characteristic function of the measures $\delta(x_n)$ to $\delta(x)$. But this implies that $x_n \to x$ in $\mathbb{R}^D$, so (T.2) holds.

Finally, let $A \subset \mathbb{R}^D \setminus \{0\}$ be bounded w.r.t. d. Then there is $\varepsilon > 0$ s.t. $\varepsilon < \inf\{\|x\|_\infty \mid x \in A\} \leq \sup\{\|x\|_\infty \mid x \in A\} < \varepsilon^{-1}$. Consider $u = (u^*, \ldots, u^*) \in \mathbb{R}^D$ with $u^* = (\varepsilon\pi)/(2D)$. Then $u^t x \in (\pi\varepsilon^2/(2D), \pi/2)$ for $x \in A$ and moreover:

$$\begin{aligned}\inf_{x\in A}|F_u(x)|^2 &= \inf_{x\in A}\left|e^{iu^t x} - 1\right|^2 \geq \inf_{x\in A}\left|\cos(u^t x) - 1\right|^2 \\ &= \inf_{z\in(\pi\varepsilon^2/(2D),\pi/2)}\left|\cos(z) - 1\right|^2 = \big(1 - \cos(\pi\varepsilon^2/(2D))\big)^2 =: \delta^2.\end{aligned}$$

Thus, (T.3) holds and Theorem 2.3 applies to show $\mu_n \xrightarrow{\mathrm{w}^\#} \mu$. □

Remark 3.2. Of course, the previous result is well-known, see [Sat99, Theorem 8.7]. ♣

3.2. **Example 2: Lévy-Khintchine formula on $\mathcal{M}_f(E)$.** Let E be a Polish space and $\mathcal{M}_f(E)$ denote the finite measures on E. Suppose that Z takes values in finite measures on E ($E = \{1, \ldots, D\}$ is a special case of Example 1 if we restrict to nonnegative random variables there) and that Z is infinitely divisible. Then, under the assumption that $\mathbb{E}[Z(E)] < \infty$, Theorem 6.1 of [Kal83] states that there exists $b \in \mathcal{M}_f(E)$ and $\mu \in \mathcal{M}^\#(\mathcal{M}_f(E) \setminus \{0\})$ such that

$$L(\phi) := -\log \mathbb{E}[\exp(-\langle \phi, Z\rangle)] = \langle \phi, b\rangle + \int 1 - \exp(-\langle \phi, \nu\rangle)\, \mu(\mathrm{d}\nu), \quad \phi \in \mathcal{C}_b\big(E, [0,\infty)\big), \tag{8}$$

where the set $\mathcal{C}_b\big(E, [0,\infty)\big)$ denotes the nonnegative, bounded, continuous functions on E. We use the metric space $(X, d) = (\mathcal{M}_f(E) \setminus \{0\}, d)$ with

$$d(\nu, \nu') = d_{\text{Prohorov}}(\nu, \nu') + |\nu(E)^{-1} - \nu'(E)^{-1}|$$

for the definition of $\mathcal{M}^\#(X)$. The uniqueness of the pair (b, μ) in the Lévy-Khintchine formula in (8) can be shown with our methodology.

Proposition 3.3. *The pair $(b, \mu) \in \mathcal{M}_f(E) \times \mathcal{M}^\#(\mathcal{M}_f(E) \setminus \{0\})$ in (8) is unique.*

Proof. Step 1. First, b can be identified via $\langle b, \phi\rangle = \lim_{m\to\infty} \frac{1}{m} L(m\phi)$.

Step 2. To identify μ we want to use Corollary 2.8. Since (X, d) is a Polish space it is also a Souslin space. Define for $\phi \in \mathcal{C}_b(E, [0,\infty))$ the function $F_\phi : X \to [0, 1]$ via

$$F_\phi(\nu) = 1 - \exp(-\langle \phi, \nu\rangle),\ \nu \in X.$$

The linear span of functions of this kind is defined

$$\mathcal{F} := \operatorname{span}\{F_\phi \mid \phi \in \mathcal{C}_b(E, [0,\infty))\}.$$

Then it is easy to see that $\mathcal{F} \subset \mathcal{C}_b(X)$.

Step 3. Now we want to verify the remaining conditions of Corollary 2.8. (S.1) holds since $\mathcal{F}$ are real-valued functions and $F_\phi \cdot F_\psi = F_{\phi+\psi} - F_\phi - F_\psi$ for $\phi, \psi \in \mathcal{C}_b(E, [0,\infty))$. (S.2) and (S.3) trivially hold. So we can apply the corollary and deduce the uniqueness of μ. □

Remark 3.4. The previous proof also works with Theorem 2.3 and thus allows to deduce a result on the convergence of the characteristics for sequences of infinitely divisible random measures. ♣

3.3. **Example 3: Excursion measure of Brownian motion.** Let $P_x \in \mathcal{M}_1\big(\mathcal{C}(\mathbb{R}_+, \mathbb{R})\big)$ be the law of a 1-dimensional Brownian motion started in $x \in \mathbb{R}$ and denote the canonical process by $(B_t)_{t\geq 0}$. Let $T_0 := \inf\{t > 0 :\ B_t = 0\}$ be the first hitting time of the origin 0. It is a folklore fact that the measure $\mu_n := nP_{1/n}((B_{t\wedge T_0})_{t\geq 0} \in \cdot)$ converges, as $n \to \infty$, to the Itô excursion measure μ_{exc} of the reflected Brownian motion $(|B_t|)_{t\geq 0}$. There are several ways to define μ_{exc}. We use the characterisation given in Theorem XII.4.2 of [RY99] (where $\mu_{\text{exc}} = 2n_+$ for n_+ used in [RY99]) as definition.

Definition 3.5 (Brownian excursion measure μ_{exc})**.** Let $X' := \mathcal{C}(\mathbb{R}_+;\mathbb{R}_+)$ be equipped with the topology of uniform convergence on compacta. For $r > 0$, let $\nu_r \in \mathcal{M}_1(X')$ be the law of a 3-dimensional Bessel bridge of length r. Define

$$\mu_{\mathrm{exc}} := \int_{\mathbb{R}_+} \nu_r \, \kappa(r) \, \mathrm{d}r \qquad \text{for} \qquad \kappa(r) := (2\pi r^3)^{-1/2}.$$

Then the σ-finite measure μ_{exc} on X' is called Brownian excursion measure. $\diamondsuit$

Since μ_{exc} is obviously not a finite measure, we have to be more precise about what we mean by convergence of μ_n to μ_{exc}. In Theorem 1 of [Hut09] it is shown (for a more general class of diffusions) that $\int F \, \mathrm{d}\mu_n \to \int F \, \mathrm{d}\mu_{\mathrm{exc}}$ holds for every $F \in \mathcal{C}_b(X')$ with the property that there is an $\varepsilon > 0$ with $F(e) = 0$ whenever $\|e\|_\infty < \varepsilon$. This looks very much like weak$^\#$-convergence on $X' \setminus \{0\}$, where 0 denotes the zero function and is sent infinitely far away.[3] This is, however, not precisely the case, because the map $e \mapsto \|e\|_\infty$ is not continuous w.r.t. uniform convergence on compacta.

In this subsection we give a setup, where we can apply Theorem 2.3 to obtain a weak$^\#$-convergence $\mu_n \xrightarrow{\mathrm{w}^\#} \mu_{\mathrm{exc}}$. To this end, we have to modify the topology on (a subspace of) X' in two ways. First, we weaken uniform convergence on compacta to convergence in Lebesgue measure, because the latter is induced by "nice" functions and therefore much easier to handle in our framework. This, of course, substantially weakens our result, so that it does not imply the one in [Hut09]. Second, we strengthen the topology (and therefore our result) a bit by additionally requiring convergence of excursion lengths for the convergence of excursions. This allows us to send the zero function infinitely far away by a continuous function, and the result in [Hut09] does not directly include ours.

Definition 3.6 (Our excursion space)**.** Define the excursion length $\zeta\colon X' \to [0,\infty]$ by

$$\zeta(e) = \sup\{t > 0 \mid e(t) \neq 0\} \cup \{0\},$$

the space of excursions $X := \zeta^{-1}\big((0,\infty)\big)$, and the metric

$$d(e,\hat{e}) = \Big(\int_0^\infty |e(t) - \hat{e}(t)| \wedge 1 \, \mathrm{d}t\Big) \wedge 1 + \big|\zeta(e)^{-1} - \zeta(\hat{e})^{-1}\big|$$

on X. $\diamondsuit$

The topology induced by d on X is the Meyer-Zheng topology (or pseudo-path topology) introduced in [MZ84] plus convergence of excursion lengths as we show in Lemma 3.8.

Definition 3.7 (Meyer-Zheng topology)**.** Let λ be the probability measure on $\mathbb{R}_+$ with Lebesgue-density $t \mapsto e^{-t}$, and $e_n, e\colon \mathbb{R}_+ \to \mathbb{R}_+$ measurable. Then e_n is said to converge to e in Meyer-Zheng topology if the image measures of λ under e_n converge weakly to the one under e. $\diamondsuit$

Lemma 3.8 (The topology induced by d)**.** *Let $e_n, e \in X$. Then the following are equivalent:*

(1) $e_n \to e$ with respect to d.

(2) e_n converges to e in Lebesgue-measure, and $\zeta(e_n) \to \zeta(e)$.

(3) e_n converges to e in Meyer-Zheng topology, and $\zeta(e_n) \to \zeta(e)$.

In particular, (X,d) is a separable metric space.

Proof. For the "in particular" note that separability of convergence in Lebesgue-measure is well-known and carries over to d-convergence by the first equivalence.

[3]Given a metric space (X,d) and $x \in X$, "sending x infinitely far away" is a figure of speech for considering $X \setminus \{x\}$ with a metric d' topologically equivalent to d, but making every sequence that d-converges to x leave every d'-ball. A possible choice is $d'(y,z) = d(y,z) + |d(x,y)^{-1} - d(x,z)^{-1}|$. Formally, $d'(x,y) = \infty$ for all $y \in X \setminus \{x\}$.

1⇔2: Since $\zeta(e) \in (0,\infty)$ for all $e \in X$, we have that $\left|\zeta(e)^{-1} - \zeta(\hat{e})^{-1}\right| \to 0$ is equivalent to $\zeta(e_n) \to \zeta(e)$. It is well-known that $d_{\mathrm{Leb}}(e,\hat{e}) = \int |e(t) - \hat{e}(t)| \wedge 1 \, \mathrm{d}t$ induces convergence in Lebesgue-measure (e.g. [Bog07, Exercise 4.7.61]), so the same is true for $d_{\mathrm{Leb}} \wedge 1$.
3⇔2: In [MZ84, Lemma 1] it is shown that Meyer-Zheng topology coincides with convergence in λ-measure. Now $\zeta(e_n) \to \zeta(e)$ implies $M := \sup_{n\in\mathbb{N}} \zeta(e_n) < \infty$, and λ is equivalent to Lebesgue-measure on $[0,M]$. □

Theorem 3.9 (Brownian excursion measure)**.** *Let (X,d) be the excursion space introduced in Definition 3.6, μ_{exc} the Brownian excursion measure, and $\widehat{B} = (\widehat{B}_t)_{t\geq 0}$ Brownian motion killed in 0. In $\mathcal{M}^{\#}(X)$,*

$$\mu_n := nP_{1/n}(\widehat{B} \in \cdot) \xrightarrow{\mathrm{w}^{\#}} \mu_{exc} \qquad \text{as } n \to \infty.$$

In order to use Theorem 2.3 to prove Theorem 3.9, we need a set $\mathcal{F}$ of continuous functions on X satisfying (T.1) – (T.3). To this end, define for $f \in L^1(\mathbb{R}_+)$ and $g \in \mathcal{C}_b(\mathbb{R}_+)$

$$F_{f,g}(e) = \int_0^\infty f(t) g(e(t)) \, \mathrm{d}t.$$

Denote by $\mathcal{C}_c$ the continuous functions with compact support and define a set of bounded functions on X by

$$\mathcal{F}' = \left\{ F_{f,g} \mid f \in \mathcal{C}_c(\mathbb{R}_+),\, x \mapsto (1 \wedge x)^{-1} g(x) \in \mathcal{C}_b(\mathbb{R}_+) \right\} \cup \{ h \circ \zeta \mid h \in \mathcal{C}_b(\mathbb{R}_+) \}.$$

Definition 3.10 ($\mathcal{F}$)**.** Let $\mathcal{F}$ be the multiplicative closure of $\mathcal{F}'$. ♢

Lemma 3.11. *$\mathcal{F}$ is weak$^{\#}$-convergence determining for measures in $\mathcal{M}^{\#}_{\mathcal{F}}(X)$.*

Proof. $\mathcal{F}$ obviously satisfies (T.1) and (T.3). Indeed, $\mathcal{F}$ is multiplicatively closed by definition, and $A \subseteq X$ is d-bounded if and only if $\inf_{e\in A} \zeta(e) > 0$. Once we have shown (T.2), i.e. that $\mathcal{F}$ induces the same topology as d, the claim follows from Theorem 2.3. To this end, note that $\zeta(e_n) \to \zeta(e)$ is necessary for both d- and $\mathcal{F}$-convergence, and recall that under this condition, by Lemma 3.8, d-convergence is equivalent to

$$\int_0^\infty \varphi(t, e_n(t)) \, \lambda(\mathrm{d}t) \to \int_0^\infty \varphi(t, e(t)) \, \lambda(\mathrm{d}t) \qquad \forall \varphi \in \mathcal{C}_b(\mathbb{R}_+^2). \tag{9}$$

The set of $\varphi_{f,g}$ of the form $\varphi_{f,g}(t,x) = f(t)g(x)e^t$ with $f \in \mathcal{C}_c(\mathbb{R}_+)$, $g \in \mathcal{C}_b(\mathbb{R}_+)$ and $x \mapsto x^{-1}g(x) \in \mathcal{C}_b(\mathbb{R}_+)$ is a multiplicatively closed subset of $\mathcal{C}_b(\mathbb{R}_+^2)$, and induces the Euclidean topology on $\mathbb{R}_+^2$. Thus, by the classical Le Cam theorem, it is convergence determining, and (9) is equivalent to

$$F_{f,g}(e_n) = \int_0^\infty \varphi_{f,g}(t, e_n(t)) \, \lambda(\mathrm{d}t) \to F_{f,g}(e) \qquad \forall F_{f,g} \in \mathcal{F}',$$

which implies the claim. □

Proof of Theorem 3.9. In view of Lemma 3.11 it is sufficient to show

$$\lim_{\varepsilon\to 0} \tfrac{1}{\varepsilon} \mathbb{E}_\varepsilon\big[F(\widehat{B})\big] = \int F \, \mathrm{d}\mu_{\mathrm{exc}} \qquad \forall F \in \mathcal{F}. \tag{10}$$

Fix $F \in \mathcal{F}$. There is $h \in \mathcal{C}_b(\mathbb{R}_+)$, $n \in \mathbb{N}_0$, $f_i \in L^1 \cap \mathcal{C}_c(\mathbb{R}_+)$, $x \mapsto (1 \wedge x)^{-1} g_i(x) \in \mathcal{C}_b(\mathbb{R}_+)$ for $1 \leq i \leq n$, such that

$$F(e) = h \circ \zeta(e) \cdot \prod_{i=1}^n F_{f_i,g_i}(e) = \int_{\mathbb{R}_+^n} f(\underline{t}) h(\zeta(e)) g\big(e_1(t_1), \ldots, e_n(t_n)\big) \, \mathrm{d}\underline{t},$$

where we set $\underline{t} = (t_1, \ldots, t_n)$, $f(\underline{t}) = \prod_{i=1}^n f_i(t_i)$ and $g(\underline{t}) = \prod_{i=1}^n g_i(t_i)$. Let $\bar{t} = \max\{t_1, \ldots, t_n\}$. Using that $\zeta(e) = r$ ν_r-a.s., and $g_i(e(t_i)) = 0$ whenever $t_i > \zeta(e)$, we obtain

$$\int F \, \mathrm{d}\mu_{\mathrm{exc}} = \int_0^\infty \kappa(r) h(r) \int \prod_{i=1}^n F_{f_i,g_i} \, \mathrm{d}\nu_r \, \mathrm{d}r \tag{11}$$

$$= \int_{\mathbb{R}_+^n} f(\underline{t}) \int_0^\infty h(\bar{t}+r)\kappa(\bar{t}+r) \int g\big(e_1(t_1),\dots,e_n(t_n)\big)\, \nu_{\bar{t}+r}(\mathrm{d}e)\,\mathrm{d}r\,\mathrm{d}\underline{t}$$
$$= \int_{\mathbb{R}_+^n} f(\underline{t}) \int_0^\infty h(\bar{t}+r)\mathbb{E}_0\Big[\frac{g(\rho_{t_1},\dots,\rho_{t_n})}{\rho_{\bar{t}}}\ell^{\rho_{\bar{t}}}(r)\Big]\,\mathrm{d}r\,\mathrm{d}\underline{t},$$

where $\rho = (\rho_t)_{t\geq 0}$ denotes a 3-dimensional Bessel process, ℓ^α the density of a Lévy distribution with scale parameter α, and we have used the relation of densities of the Bessel bridge and process (as obtained, e.g., in [RY99, XI.§3]).

On the other hand, recall that $T_0 = \zeta(\widehat{B})$ is the hitting time of 0 of $\widehat{B}$ and observe

$$\tfrac{1}{\varepsilon}\mathbb{E}_\varepsilon\big[F(\widehat{B})\big] = \int f(\underline{t})\mathbb{E}_\varepsilon\big[h(T_0)\tfrac{1}{\varepsilon}\prod_{i=1}^n g_i(\widehat{B}_{t_i})\big]\,\mathrm{d}\underline{t}. \tag{12}$$

For fixed $\underline{t}$, reordering if necessary, we assume for notational convenience that $t_1 \leq \dots \leq t_n = \bar{t}$ and set $t_0 = 0$. Then, using the Markov property at t_n in the first step,

$$\mathbb{E}_\varepsilon\Big[h(T_0)\tfrac{1}{\varepsilon}\prod_{i=1}^n g_i(\widehat{B}_{t_i})\Big] = \mathbb{E}_\varepsilon\Big[\mathbb{E}_{\widehat{B}_{t_n}}\big[h(t_n+T_0)\big]\frac{1}{\widehat{B}_{t_n}}\prod_{i=1}^n \frac{\widehat{B}_{t_i}}{\widehat{B}_{t_{i-1}}} g_i(\widehat{B}_{t_i})\Big]$$
$$= \mathbb{E}_\varepsilon\Big[\mathbb{E}_{\rho_{t_n}}\big[h(t_n+T_0)\big]\frac{1}{\rho_{t_n}}\prod_{i=1}^n g_i(\rho_{t_i})\Big], \tag{13}$$

where we have used that the sub-Markovian semigroup $(Q_t)_{t\geq 0}$ of the killed Brownian motion $\widehat{B}$ and the Markovian semigroup $(H_t)_{t\geq 0}$ of the Bessel process ρ are related by

$$H_t(x,\mathrm{d}y) = \begin{cases} x^{-1}Q_t(x,\mathrm{d}y)y & \text{for } x>0,\\ 2\kappa(t)y^2\exp(-y^2/(2t))\,\mathrm{d}y & \text{for } x=0.\end{cases}$$

Because $\widehat{B}$ is a Feller-process and $h \in \mathcal{C}_b$, the function $x \mapsto \mathbb{E}_x\big[h(t_n+T_0)\big]$ is also a bounded continuous function. Because also $x \mapsto x^{-1}g_n(x)$ is bounded and continuous by assumption, we see that the term inside the outer expectation in (13) is a bounded continuous function in $\rho_{t_1},\dots,\rho_{t_n}$. Using that ρ is a Feller process, we obtain from (12) and (13)

$$\lim_{\varepsilon\to 0}\tfrac{1}{\varepsilon}\mathbb{E}_\varepsilon\big[F(\widehat{B})\big] = \int f(\underline{t})\mathbb{E}_0\Big[\mathbb{E}_{\rho_{\bar{t}}}\big[h(\bar{t}+T_0)\big]\frac{g(\rho_{t_1},\dots,\rho_{t_n})}{\rho_{\bar{t}}}\Big]\,\mathrm{d}\underline{t}. \tag{14}$$

The hitting time T_0 of 0 under P_x is known to be Lévy-distributed with scale parameter x, hence

$$\mathbb{E}_{\rho_{\bar{t}}}\big[h(\bar{t}+T_0)\big] = \int_0^\infty h(\bar{t}+r)\ell^{\rho_{\bar{t}}}(r)\,\mathrm{d}r. \tag{15}$$

Inserting (15) into (14), and applying (11), we obtain the claimed convergence (10). □

3.4. **Example 4: Mass fragmentations.** Consider the set

$$S^\downarrow = \Big\{\underline{s} = (s_1,s_2,\dots) \in [0,1]^{\mathbb{N}} \;\Big|\; \sum_{i\geq 1} s_i \leq 1,\ s_1 \geq s_2 \geq s_3 \geq \dots \geq 0\Big\}$$

of decreasing sequences $(s_1, s_2, \dots)$ with sum less than 1. This set is used to model (mass) fragmentation processes. A perfect introduction to the topic is given in Bertoin's book [Ber06]. In his Section 2.1, the set $S^\downarrow$ is introduced and endowed with the topology of pointwise convergence. Our goal is to present two sets of real-valued convergence-determining functions on $S^\downarrow$, namely

$$\mathcal{G}_1 = \Big\{S^\downarrow \ni \underline{s} = (s_1,s_2,\dots) \mapsto G_p(\underline{s}) = \sum_{i\geq 1} s_i^p \;\Big|\; p \in \mathbb{N}\Big\},$$
$$\mathcal{G}_2 = \Big\{S^\downarrow \ni \underline{s} = (s_1,s_2,\dots) \mapsto H_\alpha(\underline{s}) = \sum_{i\geq 1}(1-e^{-\alpha s_i}) \;\Big|\; \alpha > 0\Big\}.$$

Theorem 3.12. *The sets $\mathcal{G}_1$ and $\mathcal{G}_2$ are convergence-determining on $S^\downarrow$ in the sense that for $\underline{s}(n), \underline{s} \in S^\downarrow$, $n \in \mathbb{N}$ the following holds for $j = 1, 2$:*

$$G(\underline{s}(n)) \xrightarrow{n\to\infty} G(\underline{s}) \; \forall G \in \mathcal{G}_j \qquad \Longrightarrow \qquad \underline{s}(n) \xrightarrow{n\to\infty} \underline{s}.$$

Before we give the proof, we relate the set $S^\downarrow$ to boundedly finite measures. Therefore, let

$$X = (0,1] \text{ with metric } d(x,y) = \left|x^{-1} - y^{-1}\right|.$$

Consider the mapping

$$\Phi : \begin{cases} S^\downarrow & \to \mathcal{M}^\#(X), \\ (s_1, s_2 \dots) & \mapsto \sum_{i\geq 1} \delta_{s_i}. \end{cases}$$

Lemma 3.13. *The mapping Φ is a homeomorphism from $S^\downarrow$ to $\Phi(S^\downarrow)$.*

Proof. Step 1. The mapping Φ is injective: For $\underline{s} \in S^\downarrow$, all $s_i, i \geq 1$ can be easily reconstructed.

Step 2. The mapping Φ is continuous: Let $\underline{s}(n), \underline{s} \in S^\downarrow$ with $\underline{s}(n) \to \underline{s}$. If A is a bounded set in X w.r.t. d with $\Phi(\underline{s})(\partial A) = 0$, then it is clear that $\Phi(\underline{s}(n))(A) \to \Phi(\underline{s})(A)$ and we can use Proposition A.2.6.II (d) in [DVJ03].

Step 3. The mapping $\Phi^{-1}|_{\Phi(S^\downarrow)}$ is continuous: Suppose $\Phi(\underline{s}(n)) \xrightarrow{\mathrm{w}^\#} \Phi(\underline{s})$. Fix $z \in (0,1)$ with $z \notin \{\underline{s}_i \mid i \in \mathbb{N}\}$. Then we have $\Phi(\underline{s}(n))|_{[z,1]} \xrightarrow{w} \Phi(\underline{s})|_{[z,1]}$. But this implies convergence of $\underline{s}(n)$ to $\underline{s}$ on those coordinates that lie in $[z, 1]$. Since we may choose z arbitrarily small, this suffices. □

Proof of Theorem 3.12. We only provide the proof for $\mathcal{G}_1$, since the other proof is similar.

Step 1. Note that we can write

$$G_p(\underline{s}) = \int x^p \, \Phi(\underline{s})(\mathrm{d}x), \; p \in \mathbb{N}.$$

Thus, $G(\underline{s}(n)) \to G(\underline{s}) \, \forall G \in \mathcal{G}_1$ can be written as

$$\int x^p \, \Phi(\underline{s}(n))(\mathrm{d}x) \to \int x^p \, \Phi(\underline{s})(\mathrm{d}x), \; p \in \mathbb{N}.$$

We use Theorem 2.3 to show that $\Phi(\underline{s}(n)) \xrightarrow{\mathrm{w}^\#} \Phi(\underline{s})$ and this suffices for $\underline{s}(n) \to \underline{s}$ by Lemma 3.13.

Step 2. Of course (X, d) as before Lemma 3.13 is a separable metric space. Moreover, the class $\mathcal{F} = \mathrm{span}\{x \mapsto x^p \mid p \in \mathbb{N}\} \subset \mathcal{C}_b(X)$ of polynomials without constant term on X satisfies the prerequisites of Theorem 2.3, which is easy to check. Finally, $\Phi(\underline{s}) \in \mathcal{M}^\#_{\mathcal{F}}(X)$ for all $\underline{s} \in S^\downarrow$ since for all $p \in \mathbb{N}$:

$$\int x^p \, \Phi(\underline{s})(\mathrm{d}x) \leq \int x \, \Phi(\underline{s})(\mathrm{d}x) = \sum_{i\geq 1} s_i \leq 1 < \infty.$$

Thus, Theorem 2.3 applies and yields the claim. □

Remark 3.14. Note that $\mathcal{G}_1 \not\subset \mathcal{C}(S^\downarrow)$: for $\underline{s}_i(n) = n^{-1}\mathbb{1}_{1\leq i\leq n}$, $i \in \mathbb{N}$, we have $\underline{s}(n) \to \underline{0}$, but $G_1(\underline{s}(n)) = 1 \not\to 0 = G_1(\underline{0})$. ♣

The stronger statement than Theorem 3.12 including the continuity holds on the subset of decreasing sequences summing up to 1.

Corollary 3.15. *On the subset $S_1^\downarrow = \{\underline{s} \in S^\downarrow \mid \sum_{i\geq 1} s_i = 1\}$, the set of functions $\mathcal{G}_1$ generates the topology of pointwise convergence.*

Proof. First, the subset is a Polish space with the relative topology since it is a G_δ-subset of $S^\downarrow$.

Let $\underline{s}(n), \underline{s} \in S_1^\downarrow$, $n \in \mathbb{N}$.

Step 1. First suppose that $\underline{s}(n) \to \underline{s}$ as $n \to \infty$. We can use an approximation argument to get $G(\underline{s}(n)) \to G(\underline{s})$ for all $G \in \mathcal{G}_1$: Let $\varepsilon > 0$. Fix $z > 0$ s.t. $\sum_{i\geq 1} s_i \mathbb{1}_{s_i \leq z} < \varepsilon$ and $s_i \neq z$ for all

$i \geq 1$. By Proposition A.2.6.II (d) from [DVJ03] and our Proposition 3.13, we may choose n so large that $\Phi(\underline{s}(n))|_{[z,1]}$ and $\Phi(\underline{s})|_{[z,1]}$ are very close in the following sense:

$$\sup_{q\in\{1,p\}} \Big| \sum_{i\geq 1} s_i(n)^q \, \mathbb{1}_{s_i(n)>z} - \sum_{i\geq 1} s_i^q \, \mathbb{1}_{s_i(n)>z} \Big| = A(\varepsilon) < \varepsilon.$$

Use that in the following equation

$$\begin{aligned} G_p(\underline{s}(n)) &= \sum_{i\geq 1} s_i(n)^p \, \mathbb{1}_{s_i(n)>z} + \sum_{i\geq 1} s_i(n)^p \, \mathbb{1}_{s_i(n)\leq z} \\ &= A(\varepsilon) + \sum_{i\geq 1} s_i^p \, \mathbb{1}_{s_i>z} + \sum_{i\geq 1} s_i(n)^p \, \mathbb{1}_{s_i(n)\leq z}. \end{aligned}$$

Once we establish that the last term is small, we see that $G_p(\underline{s}(n)) \to G_p(\underline{s})$. But the last term is bounded by:

$$\sum_{i\geq 1} s_i(n)^p \, \mathbb{1}_{s_i(n)\leq z} \leq \sum_{i\geq 1} s_i(n) \mathbb{1}_{s_i(n)\leq z} = 1 - \sum_{i\geq 1} s_i(n) \mathbb{1}_{s_i(n)>z} \leq 1 - (1-\varepsilon) = \varepsilon.$$

Note that it was crucial to know that $\sum_{i\geq 1} s_i = 1$ in the last proof.

Step 2. The converse direction was established in Theorem 3.12. □

Remark 3.16. An application in a similar spirit is given in [GGR], where X is the set of ultrametric measure spaces with diameter in $[0, 2h)$ for certain $h > 0$. Any ultrametric measure space with diameter in $[0, 2h]$ can be written as a boundedly finite measure on X (in a unique way similar to a prime factorization). This relation can be used for a result analogous to Corollary 3.15. ♣

References

[AEW13] Siva Athreya, Michael Eckhoff, and Anita Winter, *Brownian motion on $\mathbb{R}$-trees*, Trans. Amer. Math. Soc. **365** (2013), 3115–3150. MR 3034461

[ALW15] Siva Athreya, Wolfgang Löhr, and Anita Winter, *Invariance principle for variable speed random walks on trees*, Ann. Probab. **in press** (2015), 49 pages, arXiv:1404.6290.

[ALW16] ———, *The gap between Gromov-vague and Gromov-Hausdorff-vague topology*, Stochastic Process. Appl. **126** (2016), no. 9, 2527–2553. MR 3522292

[BCR76] Christian Berg, Jens Peter Reus Christensen, and Paul Ressel, *Positive definite functions on abelian semigroups*, Mathematische Annalen **223** (1976), no. 3, 253–274.

[Ber06] Jean Bertoin, *Random fragmentation and coagulation processes*, Cambridge Studies in Advanced Mathematics, vol. 102, Cambridge University Press, Cambridge, 2006. MR 2253162 (2007k:60004)

[BK10] Douglas Blount and Michael A. Kouritzin, *On convergence determining and separating classes of functions*, Stochastic Process. Appl. **120** (2010), no. 10, 1898–1907.

[Bog07] V. I. Bogachev, *Measure theory, volume I*, Springer, 2007.

[BP06] Mátyás Barczy and Gyula Pap, *Portmanteau theorem for unbounded measures*, Statist. Probab. Lett. **76** (2006), no. 17, 1831 – 1835.

[Coh80] Donald L. Cohn, *Measure theory*, Birkhäuser, 1980.

[DGP11] A. Depperschmidt, A. Greven, and P. Pfaffelhuber, *Marked metric measure spaces*, Electron. Commun. Probab. **16** (2011), 174–188.

[DVJ03] D. J. Daley and D. Vere-Jones, *An introduction to the theory of point processes. Vol. I*, second ed., Springer-Verlag, New York, 2003.

[GGR] A. Greven, P. Glöde, and T. Rippl, *Branching trees I: Concatenation and infinite divisibility*, in preparation.

[GPW09] Andreas Greven, Peter Pfaffelhuber, and Anita Winter, *Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees)*, Probab. Theory Related Fields **145** (2009), no. 1-2, 285–322. MR 2520129 (2011c:60008)

[GSW15] Andreas Greven, Rongfeng Sun, and Anita Winter, *Continuum space limit of the genealogies of interacting Fleming-Viot processes on $\mathbb{Z}$*, arXiv:1508.07169, 2015.

[HJ77] J. Hoffmann-Jørgensen, *Probability in banach space*, Springer, 1977.

[HL06] Henrik Hult and Filip Lindskog, *Regular variation for measures on metric spaces*, Publ. Inst. Math. (Beograd)(NS) **80** (2006), no. 94, 121–140.

[Hut09] M. Hutzenthaler, *Interacting diffusions and trees of excursions: convergence and comparison*, Electron. J. Probab. **17** (2009), 1–49.

[Kal83] Olav Kallenberg, *Random measures*, Akademie-Verlag/Academic Press, Berlin/London, 1983.

[LC57] L. Le Cam, *Convergence in distribution of stochastic processes*, University of California Publications in Statistics **2** (1957), 207–236.

[Löh13] Wolfgang Löhr, *Equivalence of Gromov-Prohorov- and Gromov's $\underline{\square}_\lambda$-metric on the space of metric measure spaces*, Electron. Commun. Probab. **18** (2013), no. 17, 1–10. MR 3037215

[LRR14] Filip Lindskog, Sidney I. Resnick, and Joyjit Roy, *Regularly varying measures on metric spaces: Hidden regular variation and hidden jumps*, Probab. Surveys **11** (2014), 270–314.

[LVW15] Wolfgang Löhr, Guillaume Voisin, and Anita Winter, *Convergence of bi-measure $\mathbb{R}$-trees and the pruning process*, Ann. Inst. H. Poincaré Probab. Statist. **51** (2015), no. 4, 1342–1368. MR 3414450

[Mun00] James Munkres, *Topology*, 2nd ed., Pearson, 2000.

[MZ84] P.A. Meyer and W. Zheng, *Tightness criteria for laws of semimartingales*, Ann. Inst. H. Poincaré Probab. Statist. **20** (1984), no. 4, 353–372.

[RY99] Daniel Revuz and Marc Yor, *Continuous martingales and Brownian motion*, third ed., Grundlehren der Mathematischen Wissenschaften, vol. 293, Springer-Verlag, Berlin, 1999. MR 1725357

[Sat99] Ken-iti Sato, *Lévy processes and infinitely divisible distributions*, Cambridge Studies in Advanced Mathematics, vol. 68, Cambridge University Press, Cambridge, 1999. MR 1739520

THE GAP BETWEEN GROMOV-VAGUE AND GROMOV-HAUSDORFF-VAGUE TOPOLOGY

SIVA ATHREYA, WOLFGANG LÖHR, AND ANITA WINTER

Abstract. In [ALW17] an invariance principle is stated for a class of strong Markov processes on tree-like metric measure spaces. It is shown that if the underlying spaces converge Gromov vaguely, then the processes converge in the sense of finite dimensional distributions. Further, if the underlying spaces converge Gromov-Hausdorff vaguely, then the processes converge weakly in path space. In this paper we systematically introduce and study the Gromov-vague and the Gromov-Hausdorff-vague topology on the space of equivalence classes of metric boundedly finite measure spaces. The latter topology is closely related to the Gromov-Hausdorff-Prohorov metric which is defined on different equivalence classes of metric measure spaces.

We explain the necessity of these two topologies via several examples, and close the gap between them. That is, we show that convergence in Gromov-vague topology implies convergence in Gromov-Hausdorff-vague topology if and only if the so-called lower mass-bound property is satisfied. Furthermore, we prove and disprove Polishness of several spaces of metric measure spaces in the topologies mentioned above (summarized in Figure 1)

As an application, we consider the Galton-Watson tree with critical offspring distribution of finite variance conditioned to not get extinct, and construct the so-called Kallenberg-Kesten tree as the weak limit in Gromov-Hausdorff-vague topology when the edge length are scaled down to go to zero.

Contents

1. Introduction

The paper introduces the *Gromov-vague* and the *Gromov-Hausdorff-vague* topology. These are two notions of convergence of (equivalence classes of) metric boundedly finite measure spaces. These are "localized" versions of the Gromov-weak topology and a topology closely related to the Gromov-Hausdorff-Prohorov topology on (equivalence classes of) metric finite measure spaces.

Gromov-weak convergence and sampling. The Gromov-weak topology originates from the weak topology in the space of probability measures on a fixed metric space. It is an example of a topology which comes with a canonical family of measures and convergence determining test functions. That is, given a complete, separable metric space, (X, r), we denote by $\mathcal{M}_1(X)$ the space of all Borel probability measures on X and by $\bar{\mathcal{C}}(X) := \bar{\mathcal{C}}_{\mathbb{R}}(X)$ the space of bounded, continuous $\mathbb{R}$-valued functions. A sequence of probability measures (μ_n) converges *weakly* to μ in $\mathcal{M}_1(X)$ (abbreviated $\mu_n \Longrightarrow \mu$), as $n \to \infty$, if and only if $\int \mathrm{d}\mu_n\, f \to \int \mathrm{d}\mu\, f$ in $\mathbb{R}$, as $n \to \infty$, for all $f \in \bar{\mathcal{C}}(X)$.

We wish to consider sequence of measures that live on different spaces. In such a case an immediate analogue of bounded continuous functions is not available. To still be in a position to imitate the notion of weak convergence, we rely on the following useful fact: for a sequence (μ_n) in $\mathcal{M}_1(X)$ and $\mu \in \mathcal{M}_1(X)$,

$$\mu_n \underset{n\to\infty}{\Longrightarrow} \mu \qquad \text{if and only if} \qquad \mu_n^{\otimes\mathbb{N}} \underset{n\to\infty}{\Longrightarrow} \mu^{\otimes\mathbb{N}}. \tag{1.1}$$

Indeed, the *"if"* direction follows by the fact that projections to a single coordinate are continuous. The *"only if"* direction follows as the set of bounded continuous functions $\varphi\colon X^{\mathbb{N}} \to \mathbb{R}$ of the form $\varphi((x_n)_{n\in\mathbb{N}}) = \prod_{i=1}^{N} \varphi_i(x_i)$ for some $N \in \mathbb{N}$, $\varphi_i\colon X \to \mathbb{R}$, $i = 1,\dots,N$, separates points in $X^{\mathbb{N}}$ and is multiplicatively closed (see, for example, [Löh13, Theorem 2.7] for an argument how to use [LC57] to conclude from here that integration over such test functions is even convergence determining for measures on $\mathcal{M}_1(X)$).

Consider now the set of bounded continuous functions $\varphi\colon X^{\mathbb{N}} \to \mathbb{R}$ of the following form

$$\varphi = \tilde{\varphi} \circ R^{(X,r)}, \tag{1.2}$$

where $R^{(X,r)}$ denotes the map that sends a vector $(x_n)_{n\in\mathbb{N}} \in X^{\mathbb{N}}$ to the matrix $(r(x_i,x_j))_{1\le i<j} \in \mathbb{R}_+^{\binom{\mathbb{N}}{2}}$ of mutual distances, and $\tilde{\varphi} \in \bar{\mathcal{C}}(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})$ depends on finitely many coordinates only. A *(complete, separable) metric measure space* (X,r,μ) consists of a complete, separable metric space (X,r) and a Borel measure μ on X. Denote by $\mathbb{X}_1$ the space of measure preserving isometry classes of metric spaces equipped with a Borel probability measure. Then for each representative (X,r,μ) of an isometry class $\mathcal{X} \in \mathbb{X}_1$ the image measure $R_*^{(X,r)}\mu^{\otimes\mathbb{N}} = \mu^{\otimes\mathbb{N}} \circ (R^{(X,r)})^{-1} \in \mathcal{M}_1(\mathbb{R}_+^{\binom{\mathbb{N}}{2}})$ is the same and is referred to as the *distance matrix distribution* $\nu^{\mathcal{X}}$ of $\mathcal{X}$. It turns out that if the distance matrix distributions of two metric measure spaces coincide, then the metric measure spaces fall into the same isometry class. This is known as Gromov's reconstruction theorem (compare [Gro99, Chapter $3\frac{1}{2}$]), and suggests to consider the *Gromov-weak topology*, which is the topology induced by the set of functions of the form

$$\Phi(X,r,\mu) = \int_{X^{\mathbb{N}}} \mathrm{d}\mu^{\otimes\mathbb{N}}\, \varphi = \int \mathrm{d}\nu^{\mathcal{X}}\, \tilde{\varphi}, \tag{1.3}$$

where φ is of the form (1.2). As this set is multiplicatively closed we can conclude once more that it is also convergence determining for metric measure spaces on $\mathbb{X}_1$.

The Gromov-weak topology on spaces of metric measure spaces, prescribed by test functions as in (1.3), originates from the work of Gromov in the context of metric geometry, where it is induced by so-called box metrics. In [GPW09a] the Gromov-weak topology on complete, separable metric measure spaces was reintroduced via convergence of the functions of the form (1.3), and metrized by the so-called *Gromov-Prohorov metric*. Recently, in [Löh13], it was shown that Gromov's box metric and the Gromov-Prohorov metric are bi-Lipschitz equivalent.

Independently of Gromov's work, however, the idea of proving convergence of random 0-hyperbolic metric measure spaces (that means trees) via "finite dimensional distributions", i.e., with the help of test functions of the form (1.3), has been used before in probability theory. As the land mark we consider [Ald93, Theorem 23], which states Gromov-weak convergence of suitably rescaled Galton-Watson trees towards the so-called Brownian continuum random tree (CRT), where the Galton-Watson trees are associated with an offspring distribution of finite variance, conditioned on a growing number of nodes and equipped with the uniform distribution on its nodes. Other results using test functions which imitate sampling include [GPW09a, Theorem 4], where the so-called Λ-coalescent tree is constructed as a Gromov-weak limit of finite trees. Furthermore, distributions of finite samples from metric measure spaces are used in hypothesis testing and for providing confidence intervals in the field of topological data analysis (e.g., [BGMP14, Car14]).

From Gromov-weak to Gromov-Hausdorff weak convergence. The following embedding result is known from [GPW09a, Lemma 5.8]. A sequence $(\mathcal{X}_n)$ converges Gromov-weakly to $\mathcal{X}$ in

$\mathbb{X}_1$ if and only if there is a complete, separable metric space (E, d) such that (representatives of) all $\mathcal{X}_n$ and $\mathcal{X}$ can be embedded measure-preserving isometrically into (E, d) in a way that the image measures under the isometries converge weakly to the image limit measure. Using this embedding procedure, we can also define a stronger topology: We say that a sequence $(\mathcal{X}_n)$ converges *Gromov-Hausdorff-weakly* to $\mathcal{X}$ in $\mathbb{X}_1$ if and only if there is a metric space (E, d) such that we can do the above embedding in a way that, additionally, the supports of the measures converge in Hausdorff distance.

This topology is closely related to the one introduced under the name *measured Hausdorff topology* in [Fuk87] in the context of studying the asymptotics of eigenvalues of the Laplacian on collapsing Riemannian manifolds, and extended from compact to Heine-Borel measure spaces in [KS03]. The difference to the Gromov-Hausdorff weak topology is that, instead of the supports, the whole spaces are required to converge in Hausdorff metric topology. This leads to different equivalence classes, and the connection is discussed extensively in Section 5. In probability theory, the measured (Gromov-)Hausdorff topology was reintroduced and further discussed in [EW06, Mie09], and recently extended in [ADH13] to complete, locally compact length spaces equipped with locally finite measures.

Verification of convergence. As for the Gromov-Hausdorff-weak topology no canonical family of convergence determining functions is available, a key question is how to actually verify convergence in Gromov-Hausdorff-weak topology? According to the definition, first an embedding of the whole sequence in to the same metric space must be provided. For random forests there has been the tradition to encode them (if possible) as excursions on compact intervals, and showing then convergence of the associated excursions in the uniform topology. As the map that sends an excursion to a tree-like metric measure space is continuous with respect to the Gromov-(Hausdorff)-weak topology ([ADH14, Proposition 2.9], [Löh13, Theorem 4.8]), convergence statements obtained by re-scaling the associated excursions always imply convergence Gromov-Hausdorff-weakly. This approach has been successfully applied to branching forests with a particular offspring distribution (see, for example, [DLG02, Duq03, GPW09b]). However, except for a few prototype models there is no obvious way to assign to a random graph model an excursion coming from a Markov process. In such a situation, Gromov-Hausdorff convergence and Gromov-weak convergence are shown separately (for example, [HM12, CH13, ABBGM13]), or the scaling results are stated either without the measure, using Gromov-Hausdorff convergence (for example, [LG07, MM11, HW14]), or only in the Gromov-weak topology (for example, [GPW13]).

Closing the gap. It is known that, if all considered metric measure spaces satisfy a (common) uniform volume doubling property, then Gromov-weak and Gromov-Hausdorff-weak topology are the same ([Vil08, Corollary 27.27]). "Volume doubling" is a standard property for Riemannian manifolds and regular, self-similar fractals. It is quite restrictive for random spaces, such as random recursive fractals or, important for us, random $\mathbb{R}$-trees. In particular, Aldous's Brownian CRT almost surely does not have the doubling property, as can be seen from the estimates in [Cro08, Theorem 1.3] (see also [DW14] for stable Lévy trees).

If the uniform volume doubling property fails, Gromov-Hausdorff-weak convergence is in general not implied by Gromov-weak convergence. The gap between Gromov-weak and Gromov-Hausdorff-weak topology, however, sometimes matters a lot.

Important example. We have recently considered in [ALW17] a class of strong Markov processes on natural scale with values in 0-hyperbolic compact metric spaces, which are uniquely determined by their speed measures. We obtained in [ALW17, Theorem 1] an invariance principle which states convergence of such processes in path space provided the underlying metric (speed-)measure spaces converge Gromov-Hausdorff-weakly. If we only assume Gromov-weak convergence, the processes still converge in their finite dimensional distributions, but without the additional convergence of the supports, convergence in paths space fails.

The main goal of the present paper is to close this gap between Gromov-Hausdorff-weak and Gromov-weak topology. We show that provided metric measure spaces converge Gromov-weakly, they also converge Gromov-Hausdorff-weakly if and only if the so-called (global) lower mass-bound property (Definition 3.1) is satisfied. This allows to verify Gromov-Hausdorff weak convergence via the following two steps (Theorem 6.1):

1. Verify convergence of the test functions from (1.3) together with
2. an extra "tightness condition" given by this lower mass-bound property.

The same lower mass function also turns out to be useful for characterizing the metric measure spaces which are compact and Heine-Borel, respectively, and for proving that the subspaces consisting of these metric measure spaces are Lusin spaces but not Polish if equipped with the Gromov-weak topology. The lower mass-bound property also appears in a compactness condition for the Gromov-Hausdorff-weak topology (Corollary 5.7). Furthermore, we also extend the space of complete, separable metric probability measure spaces to complete, separable, metric boundedly finite measure spaces and equip the latter with the so-called Gromov-(Hausdorff)-vague topologies.

Outline. The paper is organized as follows: In Section 2 we recall the Gromov-weak topology on the space of metric finite measure spaces and then use it to define the Gromov-vague topology on the space of metric boundedly finite measure spaces. In Section 3 the global and local lower mass-bound properties are defined and used to characterize compact metric (finite) measure spaces and Heine-Borel metric boundedly finite measure spaces. In Section 4 we characterize Gromov-vague convergence via isometric embeddings and deduce criteria for Gromov-vague compactness and Gromov-vague tightness, as well as Polishness of the space of metric boundedly finite measure spaces in Gromov-vague topology. Furthermore, we show that the subspaces of all compact and all Heine-Borel spaces, respectively, are Lusin but not Polish. In Section 5 we introduce the stronger Gromov-Hausdorff-vague topology, and clarify its relation to the measured Gromov-Hausdorff topology and the Gromov-Hausdorff-Prohorov metric. We also show that it is a Polish topology on the space of Heine-Borel boundedly finite measure spaces. For the measured Gromov-Hausdorff topology and the Gromov-Hausdorff-Prohorov metric, this means that restricting to spaces with measures of full support yields again a Polish space. In Section 6 we prove our main convergence criterion for Gromov-Hausdorff-weak and -vague convergence. Namely, given convergence in Gromov-weak or Gromov-vague topology, Gromov-Hausdorff-weak or Gromov-Hausdorff-vague convergence is equivalent to the global or local lower mass-bound property, respectively. In Section 7 we consider the construction of trees coded by continuous, transient excursions, and show that the map which sends an excursion to the corresponding metric boundedly finite measure space is continuous with respect to the Gromov-Hausdorff-vague topology. Finally, as an example, we present the Gromov-Hausdorff-vague convergence in distribution of suitably re-scaled finite-variance, critical Galton-Watson trees, which are conditioned on survival, to the so-called *continuum Kallenberg-Kesten tree.*

2. The Gromov-vague topology

In this section we define the (pointed) Gromov-vague topology. We first introduce pointed metric boundedly finite measure spaces, and the subspaces of interest. We recall the pointed Gromov-weak topology on pointed metric finite measure spaces (Definition 2.5). The pointed Gromov-vague topology is then defined based on the Gromov-weak topology via a "localization procedure" (Definition 2.7). We discuss the connection between both topologies (Remark 2.8), and present a perturbation result (Lemma 2.9).

A *(pointed, complete, separable) metric measure space* (X, r, ρ, μ) consists of a complete, separable metric space (X, r), a distinguished point $\rho \in X$ called the *root*, and a Borel measure μ on X. Since all our spaces are pointed, complete and separable, we usually drop these adjectives in the following when referring to metric measure spaces.

Definition 2.1 (equivalence of metric measure spaces)**.** *Two metric measure spaces* (X, r, ρ, μ) *and* (X', r', ρ', μ') *are said to be* equivalent *if and only if there is an isometry* $\phi\colon \operatorname{supp}(\mu) \cup \{\rho\} \to \operatorname{supp}(\mu') \cup \{\rho'\}$ *such that* $\phi(\rho) = \rho'$ *and* $\phi_*\mu = \mu'$*, where as usual we denote by*

$$\phi_*\mu := \mu \circ \phi^{-1} \tag{2.1}$$

the push forward *of the measure* μ *under the measurable map* ϕ*. We denote the equivalence of metric measure spaces by* $\cong$*. Most of the time, however, we do not distinguish between a metric measure space and its equivalence class.*

Recall that a *Heine-Borel space* is a metric space in which every bounded, closed set is compact. A Heine-Borel space is obviously complete, separable and locally compact. We consider the following subclasses of metric measure spaces.

Definition 2.2 ($\mathbb{X}$, $\mathbb{X}_{\mathrm{HB}}$, $\mathbb{X}_{\mathrm{c}}$)**.**

1. *A metric measure space* (X, r, ρ, μ) *is called* boundedly finite *if the measure* μ *is finite on all bounded subsets of* X*. Let* $\mathbb{X}$ *be the set of (equivalence classes of) metric boundedly finite measure spaces.*
2. $\mathcal{X} \in \mathbb{X}$ *is called* Heine-Borel locally finite measure space *if the equivalence class contains a representative* $\mathcal{X} = (X, r, \rho, \mu)$ *such that* (X, r) *is a Heine-Borel space. Denote the subspace of Heine-Borel spaces in* $\mathbb{X}$ *by* $\mathbb{X}_{\mathrm{HB}}$*.*
3. *An equivalence class* $\mathcal{X} \in \mathbb{X}_{\mathrm{HB}}$ *is called* compact metric finite measure space *if it contains a representative* $\mathcal{X} = (X, r, \rho, \mu)$ *such that* (X, r) *is a compact space. Denote the subspace of compact spaces in* $\mathbb{X}_{\mathrm{HB}}$ *by* $\mathbb{X}_{\mathrm{c}}$*.*

We illustrate this definition with an example which is useful for considering continuum limits of trees.

Example 2.3 (locally compact geodesic spaces and $\mathbb{R}$-trees)**.** *Recall that a geodesic space is a metric space in which every two points are connected by an isometric path, i.e. a path with length equal to the distance between these points. A geodesic space is called* $\mathbb{R}$*-tree if there is, up to reparametrization, only one simple path between every pair of points. It is a classical fact that every complete, locally compact geodesic space is a Heine-Borel space. In particular,* $\mathbb{X}_{\mathrm{HB}}$ *contains the subclass of complete, locally compact* $\mathbb{R}$*-trees with Radon measures.* □

As every Heine-Borel space is locally compact, the local compactness assumption on the geodesic space is obviously essential. The following remark discusses why the completeness assumption is important as well.

Remark 2.4 (non-complete spaces)**.** *We can allow also non-complete spaces as elements of* $\mathbb{X}$ *by identifying them with their respective completions. Note, however, that Radon measures on non-complete metric spaces are not boundedly finite in general. Consider for example the binary tree* $T := \{\rho\} \cup \bigcup_{n\in\mathbb{N}}\{0,1\}^n$ *with edges connecting* $w \in \{0,1\}^n$ *with* $(w,0) \in \{0,1\}^{n+1}$ *and* $(w,1) \in \{0,1\}^{n+1}$*,* $n \in \mathbb{N}_0$*, equipped with a metric determined by* $r(w,(w,0)) = r(w,(w,1)) := c^{-n}$ *if* $w \in \{0,1\}^n$*, for some* $c \in [\frac{1}{2}, 1)$*, and equipped with the length measure (see, Example 5.15 for a detailed definition). The length measure is indeed a Radon measure as all compact subtrees are contained in a subtree spanned by finitely many vertices. On the other hand* (T, r) *is bounded, but the length measure is not finite. Thus non-complete, locally compact* $\mathbb{R}$*-trees with a Radon measure are not elements of* $\mathbb{X}$ *in general.*

Moreover, non-complete, locally compact $\mathbb{R}$*-trees with a boundedly finite measure are not elements of* $\mathbb{X}_{\mathrm{HB}}$ *in general, as their completions do not need to be locally compact. Take for example* $T := (0,1] \times \{0\} \cup \bigcup_{n\in\mathbb{N}}\{\frac{1}{n}\} \times [0,1] \subseteq \mathbb{R}^2$*, and let* r *be the intrinsic length metric on* T *(i.e.,* $r(x,y)$ *is the Euclidean length of the shortest path within* T *connecting* x *and* y*). Then* (T, r) *is a non-complete* $\mathbb{R}$*-tree, and it is easy to see that it is locally compact. Its completion* $\overline{T} = T \cup \{(0,0)\}$*, however, is not locally compact, because* $(0,0)$ *does not possess any compact neighborhood.* □

We next recall the definition of the (pointed) Gromov-weak topology on metric finite measure spaces (see [GPW09a] and [LVW15, Section 2.1] for more details). As with the metric measure spaces, we drop the adjective "pointed" in the following when referring to topologies on spaces of (pointed) metric measure spaces.

Definition 2.5 ((pointed) Gromov-weak topology)**.** *For $m \in \mathbb{N}$, the* m-point distance matrix distribution *of a metric finite measure space $\mathcal{X} = (X, r, \rho, \mu)$ is the finite measure on $\mathbb{R}_+^{\binom{m+1}{2}}$ defined by*

$$\nu_m(\mathcal{X}) := \int_{X^m} \mu^{\otimes m}(\mathrm{d}(x_1, \ldots, x_m))\, \delta_{(r(x_i,x_j))_{0 \le i < j \le m}}, \tag{2.2}$$

where $x_0 := \rho$ and δ is the Dirac measure. A sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ of metric finite measure spaces converges *to a metric finite measure space $\mathcal{X}$* Gromov-weakly *if all m-point distance matrix distributions converge, i.e., if*

$$\nu_m(\mathcal{X}_n) \underset{n\to\infty}{\Longrightarrow} \nu_m(\mathcal{X}), \tag{2.3}$$

for all $m \in \mathbb{N}$, where we write $\Rightarrow$ for weak convergence of finite measures.

Next we define the Gromov-vague topology on the space $\mathbb{X}$ of metric boundedly finite measure spaces. The construction is a straight-forward "localization" procedure, similar to the one used by Gromov for Gromov-Hausdorff convergence of pointed locally compact spaces (compare [Gro99, Section 3B]).

Given a metric space (X, r), we use the notations $B_r(x, R)$ and $\overline{B}_r(x, R)$ for the open respectively closed ball around $x \in X$ of radius $R \geq 0$. If there is no risk of confusion, we sometimes drop the subscript r. The restriction of a metric measure space $\mathcal{X} = (X, r, \rho, \mu) \in \mathbb{X}$ to the closed ball $\overline{B}(\rho, R)$ of radius $R \geq 0$ around the root is denoted by

$$\mathcal{X}{\restriction}_R := (X, r, \rho, \mu{\restriction}_{\overline{B}(\rho,R)}) \cong \big(\overline{B}(\rho, R), r{\restriction}_{\overline{B}(\rho,R)^2}, \rho, \mu{\restriction}_{\overline{B}(\rho,R)}\big). \tag{2.4}$$

Generally (and informally), localization works as follows: given a topology on some class of spaces, the localized form of convergence is defined for those spaces $\mathcal{X}$, where for all $R > 0$, the restriction $\mathcal{X}{\restriction}_R$ falls into the original class. Such spaces converge in the localized topology if, for almost all $R > 0$, the restrictions converge. If d is a metric inducing the original topology, the localized convergence can therefore, for example, be induced by the metric

$$d^{\#}(\mathcal{X}, \mathcal{Y}) := \int_{\mathbb{R}_+} \mathrm{d}R\, e^{-R}\big(1 \wedge d(\mathcal{X}{\restriction}_R, \mathcal{Y}{\restriction}_R)\big). \tag{2.5}$$

We need the following lemma for our definition of Gromov-vague topology. Denote the Gromov-Prohorov metric, which we define in Section 4 below, by d_{GP}. For the moment, it is enough to know that it induces the Gromov-weak topology by [GPW09a, Theorem 5].

Lemma 2.6. *Let $(\mathcal{X}_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{X}$ and $\mathcal{X} = (X, r, \rho, \mu) \in \mathbb{X}$. The following are equivalent:*

1. $(\mathcal{X}_n){\restriction}_R \underset{n\to\infty}{\longrightarrow} \mathcal{X}{\restriction}_R$ *Gromov-weakly for all $R > 0$ with $\mu\big(S_r(\rho, R)\big) = 0$, where $S_r(\rho, R) = \overline{B}_r(\rho, R) \setminus B_r(\rho, R)$ is the sphere of radius R around ρ.*
2. $(\mathcal{X}_n){\restriction}_R \underset{n\to\infty}{\longrightarrow} \mathcal{X}{\restriction}_R$ *Gromov-weakly for all but countably many $R > 0$.*
3. $(\mathcal{X}_n){\restriction}_R \underset{n\to\infty}{\longrightarrow} \mathcal{X}{\restriction}_R$ *Gromov-weakly for Lebesgue-almost all $R > 0$.*
4. *There exists a sequence $R_k \to \infty$ such that $(\mathcal{X}_n){\restriction}_{R_k} \underset{n\to\infty}{\longrightarrow} \mathcal{X}{\restriction}_{R_k}$ Gromov-weakly for all $k \in \mathbb{N}$.*
5. $d_{\mathrm{GP}}^{\#}(\mathcal{X}_n, \mathcal{X}) \underset{n\to\infty}{\longrightarrow} 0$.

Proof. The implications *"1.⇒2.⇒3.⇒4."* are trivial.

"4.⇒1." is a consequence of the Portmanteau theorem. Indeed, assume that $(\mathcal{X}_n)\lceil_{R_k} \underset{n\to\infty}{\longrightarrow} \mathcal{X}\lceil_{R_k}$ Gromov-weakly along a sequence $R_k \to \infty$, and fix $R > 0$. Choose $k \in \mathbb{N}$ large enough such that $R_k \geq R$. Then, for every $m \in \mathbb{N}$, $\nu_m((\mathcal{X}_n)\lceil_{R_k}) \underset{n\to\infty}{\Longrightarrow} \nu_m(\mathcal{X}\lceil_{R_k})$. The first row of the m-point distance matrix ν_m contains, by definition, the distances to the root. Hence $\nu_m((\mathcal{X}_n)\lceil_R)$ is equal to the restriction of $\nu_m((\mathcal{X}_n)\lceil_{R_k})$ to the set of matrices with no entry in the first row exceeding R. The set of these matrices is closed, hence, by the Portmanteau theorem, the condition $\mu\big(S_r(\rho,R)\big) = 0$ implies the claimed convergence.

"3.⇔5." follows directly from the fact that d_{GP} induces the Gromov-weak topology, the definition of $d^{\#}_{\mathrm{GP}}$ in (2.5), and the dominated convergence theorem. □

We are now in a position to define the Gromov-vague topology.

Definition 2.7 ((pointed) Gromov-vague topology)**.** *We say that a sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ in $\mathbb{X}$ converges to $\mathcal{X} \in \mathbb{X}$* Gromov-vaguely *if the equivalent conditions of Lemma 2.6 hold.*

Note that usually localized convergence is not strictly a generalization of the original one, because parts can "vanish at infinity" in the limit. For example, consider Gromov-Hausdorff convergence of (pointed) compact metric spaces, and a sequence of two-point spaces, where the distance between the two points tends to infinity. Such a sequence does not converge. In the localized Gromov-Hausdorff topology, however, it converges to the compact space consisting of only one point. A similar phenomenon arises for the Gromov-vague topology.

Remark 2.8 (Gromov-vague versus Gromov-weak)**.** *Consider the subspaces $\mathbb{X}_{\mathrm{fin}}$ and $\mathbb{X}_1$ of $\mathbb{X}$, consisting of spaces $\mathcal{X} = (X, r, \rho, \mu)$ where μ is a finite measure, respectively a probability. Then on $\mathbb{X}_1$, the induced Gromov-vague topology coincides with the Gromov-weak topology. On $\mathbb{X}_{\mathrm{fin}}$, and even on $\mathbb{X}_{\mathrm{c}}$, however, this is not the case, because the total mass is not preserved in the Gromov-vague convergence. In fact, for $\mathcal{X} = (X, r, \rho, \mu)$, $\mathcal{X}_n = (X_n, r_n, \rho_n, \mu_n) \in \mathbb{X}_{\mathrm{fin}}$, The Gromov-weak convergence $\mathcal{X}_n \to \mathcal{X}$ is equivalent to $\mathcal{X}_n \to \mathcal{X}$ Gromov-vaguely and $\mu_n(X_n) \to \mu(X)$.* □

For a given metric space (X, r), denote by $d^{(X,r)}_{\mathrm{Pr}}$ the Prohorov-metric on the space of all finite measures on $(X, \mathcal{B}(X))$, i.e.,

$$\begin{aligned} &d^{(X,r)}_{\mathrm{Pr}}(\mu,\mu') \\ &:= \inf\big\{\varepsilon > 0 :\ \mu(A) \leq \mu'(A^\varepsilon) + \varepsilon,\ \mu'(A) \leq \mu(A^\varepsilon) + \varepsilon\ \ \forall A \text{ closed}\big\}, \end{aligned} \tag{2.6}$$

where $A^\varepsilon = \{\, x : d(x, A) \leq \varepsilon \,\}$ is the closed ε-neighborhood of A. Recall that the Prohorov metric induces weak convergence.

We conclude this section with a simple stability property of Gromov-vague convergence under perturbations of the measures in a localized Prohorov sense. We will illustrate this later in Section 7 with Example 5.15.

Lemma 2.9 (perturbation of measures)**.** *Consider $\mathcal{X} = (X, r, \rho, \mu)$, $\mathcal{X}_n = (X_n, r_n, \rho_n, \mu_n) \in \mathbb{X}$, and another sequence of boundedly finite measure μ'_n on X_n, $n \in \mathbb{N}$. Assume that $\mathcal{X}_n \underset{n\to\infty}{\longrightarrow} \mathcal{X}$ Gromov-vaguely, and that there exists a sequence $R_k \to \infty$ such that for all $k \in \mathbb{N}$,*

$$\lim_{n\to\infty} d^{(X_n,r_n)}_{\mathrm{Pr}}\big(\mu_n\lceil_{R_k}, \mu'_n\lceil_{R_k}\big) = 0. \tag{2.7}$$

Then $\mathcal{X}'_n := (X_n, r_n, \rho_n, \mu'_n)$ converges Gromov-vaguely to $\mathcal{X}$.

Proof. Notice that for every fixed $k, n \in \mathbb{N}$,

$$\lim_{R\downarrow R_k} d^{(X_n,r_n)}_{\mathrm{Pr}}\big(\mu_n\lceil_R, \mu_n\lceil_{R_k}\big) = 0. \tag{2.8}$$

We may therefore assume w.l.o.g. that (2.7) and $(\mathcal{X}_n)\lceil_{R_k} \underset{n\to\infty}{\longrightarrow} \mathcal{X}\lceil_{R_k}$, Gromov-weakly, hold along the same sequence $(R_k)_{k\in\mathbb{N}}$. Thus for any fixed $k \in \mathbb{N}$,

$$(\mathcal{X}'_n)\lceil_{R_k} \underset{n\to\infty}{\longrightarrow} \mathcal{X}\lceil_{R_k}, \tag{2.9}$$

Gromov-weakly, by Theorem 5 of [GPW09a]. This, however, implies the claimed Gromov-vague convergence. □

3. The lower mass-bound property

In this section we introduce the local and global lower mass-bound properties, and use them to characterize compact spaces and Heine-Borel spaces, respectively. These properties are formulated in terms of the following lower mass functions on the space of metric boundedly finite measure spaces. For $\delta, R > 0$, we define $\mathfrak{m}_\delta^R \colon \mathbb{X} \to \mathbb{R}_+ \cup \{\infty\}$ as

$$\mathfrak{m}_\delta^R\big((X, r, \rho, \mu)\big) := \inf\big\{\, \mu\big(\overline{B}_r(x, \delta)\big) : x \in B_r(\rho, R) \cap \operatorname{supp}(\mu) \,\big\}, \tag{3.1}$$

with the convention that the infimum of the empty set is ∞ (which may happen if $\rho \notin \operatorname{supp}(\mu)$). Furthermore, set

$$\mathfrak{m}_\delta := \lim_{R\to\infty} \mathfrak{m}_\delta^R = \inf_{R>0} \mathfrak{m}_\delta^R. \tag{3.2}$$

The following property plays an important rôle at several places in later arguments.

Definition 3.1 (lower mass-bound property)**.** *A set $\mathbb{K} \subseteq \mathbb{X}$ of metric boundedly finite measure spaces satisfies the* local lower mass-bound property *if and only if*

$$\inf_{\mathcal{X}\in\mathbb{K}} \mathfrak{m}_\delta^R(\mathcal{X}) > 0, \tag{3.3}$$

for all $R > \delta > 0$. It satisfies the global lower mass-bound property *if and only if*

$$\inf_{\mathcal{X}\in\mathbb{K}} \mathfrak{m}_\delta(\mathcal{X}) > 0, \tag{3.4}$$

for all $\delta > 0$. We say that a single metric measure space $\mathcal{X} \in \mathbb{X}$ satisfies the local/global mass-bound property if and only if $\mathbb{K} := \{\mathcal{X}\}$ does.

Notice that in the definition of $\mathfrak{m}_\delta^R$, we could have replaced the closed ball by an open ball and/or the open ball by a closed ball without changing the conditions (3.3) and (3.4). We made our choice such that $\mathfrak{m}_\delta^R$ is upper semi-continuous, which will be convenient in some proofs.

Lemma 3.2 (upper semi-continuity)**.** *For every $R, \delta > 0$, the lower mass functions $\mathfrak{m}_\delta^R$ and $\mathfrak{m}_\delta$ are upper semi-continuous with respect to the Gromov-vague topology.*

Proof. Fix $R, \delta > 0$, and let $\mathcal{X}_n = (X_n, r_n, \rho_n, \mu_n) \to \mathcal{X} = (X, r, \rho, \mu)$ be a Gromov-vaguely converging sequence in $\mathbb{X}$. Then we can choose $R' > R + \delta$ such that $\mathcal{X}'_n := \mathcal{X}_n\lceil_{R'}$ converges Gromov-weakly to $\mathcal{X}' := \mathcal{X}\lceil_{R'}$. By Lemma 5.8 of [GPW09a], we can assume w.l.o.g. that X, X_1, X_2, ... are subspaces of some metric space (E, d), and $\mu'_n := \mu_n\lceil_{\overline{B}(\rho,R')}$ converges weakly to $\mu' := \mu\lceil_{\overline{B}(\rho,R')}$ on (E, d). We can then find for every $x \in \operatorname{supp}(\mu) \cap B_r(\rho, R)$ a sequence $x_n \to x$ with $x_n \in \operatorname{supp}(\mu_n)$ for all $n \in \mathbb{N}$. Thus

$$\begin{aligned}
\mu\big(\overline{B}(x, \delta)\big) &= \inf_{\varepsilon>0} \mu'\big(\overline{B}(x, \delta + \varepsilon)\big) \\
&\geq \inf_{\varepsilon>0} \liminf_{n\to\infty} \mu'_n\big(\overline{B}(x, \delta + \varepsilon)\big) \\
&\geq \liminf_{n\to\infty} \mu_n\big(\overline{B}(x_n, \delta)\big) \\
&\geq \mathfrak{m}_\delta^R(\mathcal{X}_n),
\end{aligned} \tag{3.5}$$

where we have applied the Portmanteau theorem in the second step, and used in the last step that $x_n \in B(\rho_n, R)$ for large enough n. Hence $\mathfrak{m}_\delta^R$ is upper semi-continuous. Therefore, $\mathfrak{m}_\delta = \inf_{R>0} \mathfrak{m}_\delta^R$ is also upper semi-continuous. □

Corollary 3.3 (lower mass-bound property is preserved under closure)**.** *If $\mathbb{K} \subseteq \mathbb{X}$ satisfies the global or local lower mass-bound property, the same is true for its Gromov-vague closure $\overline{\mathbb{K}}$.*

Lemma 3.4 (characterization of compact mm-spaces)**.** *Let $\mathcal{X} \in \mathbb{X}$. Then $\mathcal{X}$ is a* compact *metric finite measure space if and only if it has finite total mass, and satisfies the global lower mass-bound property.*

Proof. "⇒" Assume that $\mathcal{X} = (X, r, \rho, \mu)$ is compact. Then X is bounded, and hence μ is a finite measure. For every $\delta > 0$, the function $x \mapsto \mu\big(B(x, \delta)\big)$ is lower semi-continuous. Therefore, it attains its minimum on the compact set $\mathrm{supp}(\mu)$, and thus the global lower mass-bound property holds.

"⇐" Assume that μ is finite, and that the global lower mass-bound property holds. Then for all $\delta > 0$, we can cover $\mathrm{supp}(\mu)$ with finitely many balls of radius 2δ. To see this, notice that we can choose an at most countable covering $\{B(x, 2\delta);\, x \in S \subseteq X\}$ of $\mathrm{supp}(\mu)$ with the property that the points in S have mutual distances at least 2δ. As $\{B(x, \delta);\, x \in S \subseteq X\}$ then consists of pairwise disjoint sets, each carrying μ-mass at least $\mathfrak{m}_\delta(\mathcal{X})$, the total mass of μ is at least $\mathfrak{m}_\delta(\mathcal{X}) \cdot \#S$. As μ is a finite measure, $\{B(x, 2\delta);\, x \in S \subseteq X\}$ must be a finite set. Since $\mathrm{supp}(\mu)$ is complete, this means that $\mathrm{supp}(\mu)$ is actually compact. □

Lemma 3.5 (characterization of Heine-Borel mm-spaces)**.** *Let $\mathcal{X} \in \mathbb{X}$. Then $\mathcal{X}$ is a* Heine-Borel *locally finite measure space if and only if it satisfies the local lower mass-bound property.*

Proof. Given $R > 0$, $\mathcal{X}$ satisfies $\mathfrak{m}_\delta^R(\mathcal{X}) > 0$ for every $\delta > 0$ if and only if $\mathcal{X}\!\restriction_R$ satisfies the global lower mass-bound property. Hence by Lemma 3.4, $\mathcal{X}\!\restriction_R$ satisfies the global lower mass-bound property if and only if $\mathcal{X}\!\restriction_R$ is compact. Obviously, $\mathcal{X}\!\restriction_R$ is compact for all $R > 0$ if and only if $\mathcal{X}$ is Heine-Borel. □

Corollary 3.6 ($\mathbb{X}_{\mathrm{HB}}$ and $\mathbb{X}_{\mathrm{c}}$ are measurable)**.** *Both $\mathbb{X}_{\mathrm{HB}}$ and $\mathbb{X}_{\mathrm{c}}$ are measurable subsets of $\mathbb{X}$ with respect to Borel σ-field generated by the Gromov-vague topology.*

Proof. Notice that

$$\mathbb{X}_{\mathrm{HB}} = \bigcap\nolimits_{R\in\mathbb{N}} \bigcap\nolimits_{\delta>0} \bigcup\nolimits_{a>0} \big\{\mathcal{X} \in \mathbb{X} : \mathfrak{m}_\delta^R(\mathcal{X}) \geq a\big\}, \tag{3.6}$$

by Lemma 3.5. Since the lower mass functions are upper semi-continuous by Lemma 3.2, $A_{R,\delta,a} := \{\, \mathcal{X} \in \mathbb{X} : \mathfrak{m}_\delta^R(\mathcal{X}) \geq a \,\}$ is closed for all $\delta, R > 0$. Hence $\mathbb{X}_{\mathrm{HB}}$ is measurable. The measurability of $\mathbb{X}_{\mathrm{c}}$ follows analogously by noticing that

$$\mathbb{X}_{\mathrm{c}} = \bigcap\nolimits_{\delta>0} \bigcup\nolimits_{a>0} \{\, \mathcal{X} \in \mathbb{X} : \mathfrak{m}_\delta(\mathcal{X}) \geq a,\ \mu(X) \leq a^{-1} \,\}, \tag{3.7}$$

by Lemma 3.4. □

4. Embeddings, compactness and Polishness

Recall that weak convergence of finite measures on a complete, separable metric space is induced by the complete Prohorov metric (see, (2.6)). In the same spirit, the Gromov-weak topology is induced by the complete *Gromov-Prohorov metric*, which is defined for two metric finite measure spaces $\mathcal{X} = (X, r, \mu)$ and $\mathcal{X}' = (X', r', \mu')$ by

$$d_{\mathrm{GP}}(\mathcal{X}, \mathcal{X}') := \inf_d d_{\mathrm{Pr}}^{(X \sqcup X', d)}\big(\mu, \mu'\big), \tag{4.1}$$

where the infimum is taken over all metrics d on $X \sqcup X'$ that extend both r and r', and $\sqcup$ denotes the disjoint union (see [GPW09a, Theorem 5]).

The fact that d_{GP} induces the Gromov-weak topology immediately implies the following embedding result: for every Gromov-weakly convergent sequence, $((X_n, r_n, \mu_n))_{n\in\mathbb{N}}$, there exists a common complete, separable metric space (E, d) in which all (X_n, r_n) can be isometrically embedded such that (the push-forwards of) the measures μ_n converge weakly to a measure μ on (E, d) (compare [GPW09a, Lemma 5.8]).

In this section we show that an analogous statement (Proposition 4.1) is true for the Gromov-vague topology and, if the sequence satisfies the local lower mass-bound, (E, d) can be chosen as Heine-Borel space. We will apply this to characterize compact sets in $\mathbb{X}$ (Corollary 4.3), and to show that $\mathbb{X}$ is Polish (Proposition 4.8), while $\mathbb{X}_{\mathrm{c}}$ and $\mathbb{X}_{\mathrm{HB}}$ are Lusin spaces but not Polish (Corollary 4.9). We also prove a tightness criterion for probability measures on $\mathbb{X}$ (Corollary 4.6). We start with the embedding result.

Proposition 4.1 (characterization via isometric embeddings). *For each $n \in \mathbb{N}\cup\{\infty\}$, let $\mathcal{X}_n = (X_n, r_n, \rho_n, \mu_n) \in \mathbb{X}$. Then $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}_\infty$ Gromov-vaguely if and only if there exists a pointed complete, separable metric space (E, d, ρ) and isometries $\varphi_n\colon \mathrm{supp}(\mu_n) \to E$ such that $\varphi_n(\rho_n) = \rho$ for $n \in \mathbb{N}\cup\{\infty\}$, and*

$$\big((\varphi_n)_*\mu_n\big)\restriction_{\overline{B}_d(R,\rho)} \underset{n\to\infty}{\Longrightarrow} \big((\varphi_\infty)_*\mu_\infty\big)\restriction_{\overline{B}_d(R,\rho)}, \tag{4.2}$$

for all but countably many $R \ge 0$. Furthermore, if $\{\, \mathcal{X}_n : n \in \mathbb{N}\,\}$ satisfies the local lower mass-bound property, then $\mathcal{X}_\infty \in \mathbb{X}_{\mathrm{HB}}$ and E can be chosen as Heine-Borel space. In this case, (4.2) *is equivalent to*

$$(\varphi_n)_*\mu_n \xrightarrow{vag} (\varphi_\infty)_*\mu_\infty, \tag{4.3}$$

where $\xrightarrow{vag}$ denotes vague convergence of Radon measures on E.

Before we give the proof, we illustrate with an example that the local lower mass-bound property cannot be dropped without replacement in the second part of the proposition, even if the limit is assumed to be compact.

Example 4.2 (E is not Heine-Borel without lower mass-bound). *Consider the (rooted) metric measure spaces $\mathcal{X}_n = \big([0,1]^n, r_n, 0, \frac{n-1}{n}\delta_{0^n} + \frac{1}{n}\lambda_n\big)$, where r_n is the Euclidean metric, δ_{0^n} is the Dirac measure in $0^n = (0,\dots,0) \in [0,1]^n$, and λ_n is the n-dimensional Lebesgue measure. Then $\mathcal{X}_n$ is compact and obviously converges Gromov-vaguely (and Gromov-weakly) to the compact probability space consisting of only one point, but the embedding space (E, d) cannot be chosen as Heine-Borel space.* □

Proof of Proposition 4.1. It is easy to see that (4.2) implies the Gromov-vague convergence, and that if E is a Heine-Borel space, (4.2) is equivalent to (4.3).

Conversely, assume that $\mathcal{X}_n \underset{n\to\infty}{\longrightarrow} \mathcal{X}$ Gromov-vaguely, and abbreviate $X := X_\infty$, $r := r_\infty$ and $\rho := \rho_\infty$. Let $(R_k)_{k\in\mathbb{N}}$ be an increasing sequence of radii with $\lim_{k\to\infty} R_k = \infty$ and $\mu\big(\overline{B}_r(\rho, R_k) \setminus B_r(\rho, R_k)\big) = 0$. Using that the Gromov-Prohorov metric metrizes the Gromov-weak topology by [GPW09a, Theorem 5], we can construct for $n, k \in \mathbb{N}$ a metric $d_{n,k}$ on $X_n \sqcup X$ extending both r_n and r such that for all $l \in \{1,\dots,k\}$,

$$\lim_{n\to\infty} d_{\mathrm{Pr}}^{(X_n\sqcup X, d_{n,k})}\big(\mu_n\restriction_{R_l}, \mu\restriction_{R_l}\big) = 0, \tag{4.4}$$

where we use the abbreviation

$$\mu\restriction_R := \mu\restriction_{\overline{B}_d(\rho,R)}. \tag{4.5}$$

It is easy to check that we can do it such that ρ_n and ρ are identified. Using Cantor's diagonal argument, we find a subsequence (k_n) such that $d_n := d_{n,k_n}$ satisfies $\lim_{n\to\infty} d_{\mathrm{Pr}}^{(X_n\sqcup X, d_n)}(\mu_n\restriction_{R_k}, \mu\restriction_{R_k}) = 0$, for every $k \in \mathbb{N}$. Let $E := \bigsqcup_{n\in\mathbb{N}\cup\{\infty\}} X_n$, d the largest metric on E which extends all d_n, and

$\varphi_n\colon X_n \to E$ the canonical injection. Then it is easy to check that (E,d) is a complete, separable metric space and (4.2) is satisfied.

Now assume that $\{\mathcal{X}_n : n \in \mathbb{N}\}$ satisfies the local lower mass-bound property. Then it is also satisfied for $\{\mathcal{X}_n : n \in \mathbb{N} \cup \{\infty\}\}$ by Corollary 3.3. Due to Lemma 3.5, we may assume that X_n and X are Heine-Borel spaces. We have to show that E is a Heine-Borel space as well. To this end, we show that every bounded sequence $(x_i)_{i\in\mathbb{N}}$ in E has an accumulation point. If infinitely many x_i are in a single X_n, $n \in \mathbb{N}\cup\{\infty\}$, this follows from the Heine-Borel property of X_n. Therefore, we can assume w.l.o.g. that $x_n \in X_n \cap B_d(\rho, R_k)$ for all n and some k. By (3.3) together with $(\mu_n)\restriction_{R_k} \Rightarrow \mu\restriction_{R_k}$ on E, we obtain $d(x_n, X) \to 0$. Hence there is $y_n \in X$ with $d(x_n, y_n) \to 0$ and, by the Heine-Borel property of X, $(y_n)_{n\in\mathbb{N}}$ has an accumulation point, which is also an accumulation point of $(x_n)_{n\in\mathbb{N}}$. □

From here we can easily characterize the relatively compact sets.

Corollary 4.3 (Gromov-vague compactness)**.** *For a set $\mathbb{K} \subseteq \mathbb{X}$ the following are equivalent:*

1. *$\mathbb{K}$ is relatively compact in $\mathbb{X}$ equipped with the Gromov-vague topology.*
2. *For all $R > 0$, the set of restrictions $\mathbb{K}\restriction_R := \{\mathcal{X}\restriction_R : \mathcal{X} \in \mathbb{K}\}$ is relatively compact in the Gromov-weak topology.*
3. *$\mathbb{K}\restriction_{R_k}$ is relatively compact in the Gromov-weak topology for a sequence $R_k \to \infty$.*

Furthermore, a set $\mathbb{K} \subseteq \mathbb{X}_{\mathrm{HB}}$ which satisfies the local lower mass-bound property is relatively compact in $\mathbb{X}_{\mathrm{HB}}$ equipped with Gromov-vague topology if and only if the total masses of large balls are uniformly bounded, i.e., for all $R > 0$,

$$\sup_{(X,r,\rho,\mu)\in\mathbb{K}} \mu\big(B_r(\rho,R)\big) < \infty. \tag{4.6}$$

Remark 4.4 (Gromov-weak compactness)**.** *Criteria for relative compactness in the Gromov-weak topology are given in Theorem 2 and Proposition 7.1 of* [GPW09a]. □

Remark 4.5 (convergence without the lower-mass bound property)**.** *As we have seen in Example 4.2, a Gromov-vaguely convergent sequence in $\mathbb{X}_{\mathrm{HB}}$ does not have to satisfy the local lower mass-bound property. Hence the local lower mass-bound property is not necessary for relative compactness in $\mathbb{X}_{\mathrm{HB}}$.* □

Proof of Corollary 4.3. "1.⇒2." Assume that $\mathbb{K}$ is relatively compact. Let $R > 0$, and consider a sequence $\big(\mathcal{X}_n = (X_n, r_n, \rho_n, \mu_n)\big)_{n\in\mathbb{N}}$ in $\mathbb{K}$. Then it possesses a Gromov-vague limit point $\mathcal{X} \in \mathbb{X}$, and, by passing to a subsequence, we may assume w.l.o.g. that $\mathcal{X}_n \underset{n\to\infty}{\longrightarrow} \mathcal{X}$ Gromov-vaguely. By Proposition 4.1, we can assume that $X_n \subseteq E$ for some separable metric space E, and $\mu'_n := \mu_n\restriction_{B(\rho,R')} \underset{n\to\infty}{\Longrightarrow} \mu\restriction_{B(\rho,R')} =: \mu'$ for some $R' > R$. By the Prohorov theorem, the sequence $(\mu'_n)_{n\in\mathbb{N}}$ is tight. Thus $\big(\mu_n\restriction_{\overline{B}(\rho,R)}\big)_{n\in\mathbb{N}}$ is also tight. We can conclude once more with the Prohorov theorem that $\big(\mu_n\restriction_{\overline{B}(\rho,R)}\big)_{n\in\mathbb{N}}$ is relatively weakly compact. Consequently, $(\mathcal{X}_n\restriction_R)_{n\in\mathbb{N}}$ has a Gromov-weak limit point.

"2.⇒3." is obvious.

"3.⇒1." Let $(\mathcal{X}_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{K}$. By passing to a subsequence, we may assume that $\mathcal{X}_n\restriction_{R_k}$ converges Gromov-weakly to some metric finite measure space $\mathcal{X}^{(k)}$ for all k. Now it is easy to check that $\mathcal{X}^{(i)} = \mathcal{X}^{(k)}\restriction_{R_i}$ whenever $R_i \le R_k$ and that we can therefore construct $\mathcal{X} \in \mathbb{X}$ with $\mathcal{X}\restriction_{R_k} = \mathcal{X}^{(k)}$ for every $k \in \mathbb{N}$. By definition, $\mathcal{X}_n \to \mathcal{X}$ in Gromov-vague topology.

Now assume that $\mathbb{K} \subseteq \mathbb{X}_{\mathrm{HB}}$ satisfies the local lower mass-bound property and (4.6). Fix $R > 0$. Then for every $\varepsilon > 0$ we can find $N = N(\varepsilon, \mathbb{K}) \in \mathbb{N}$ such that for every $\mathcal{X} = (X, r, \rho, \mu) \in \mathbb{K}$, we can cover $\overline{B}_r(\rho, R)$ by N balls of radius ε. Hence $\mathbb{K}\restriction_R$ is relatively compact in Gromov-weak topology by Proposition 7.1 of [GPW09a]. Therefore, $\mathbb{K}$ is relatively compact in $\mathbb{X}$ with Gromov-vague topology. As also $\overline{\mathbb{K}}$ satisfies the local lower mass-bound property by Corollary 3.3, $\overline{\mathbb{K}} \subseteq \mathbb{X}_{\mathrm{HB}}$ by Lemma 3.5. $\square$

Having a characterization of compactness at hand, we can also characterize tightness of probability measures on $\mathbb{X}$. Denote by $\mathbb{X}_{\mathrm{fin}}$ the subspace of metric finite measure spaces.

Corollary 4.6 (tightness of measures on $\mathbb{X}$)**.** *Let Γ be a family of probability measures on $\mathbb{X}$, and consider for each $R > 0$ the restriction map $\psi_R \colon \mathbb{X} \to \mathbb{X}_{\mathrm{fin}}$ given by $\mathcal{X} \mapsto \mathcal{X}\restriction_R$. Then the following are equivalent:*

1. *Γ is Gromov-vaguely tight.*
2. *The set $(\psi_R)_*(\Gamma)$ is Gromov-weakly tight for all $R > 0$.*
3. *For all $R, \varepsilon > 0$, there is a $\delta > 0$ such that*

$$\sup_{\mathbb{P}\in\Gamma} \mathbb{P}\big\{(X,r,\rho,\mu) \in \mathbb{X} : \mu\big(B_r(\rho,R)\big) > \tfrac{1}{\delta}\big\} \le \varepsilon, \tag{4.7}$$

$$\sup_{\mathbb{P}\in\Gamma} \mathbb{P}\big\{(X,r,\rho,\mu) \in \mathbb{X} : \mu\big\{x \in B_r(\rho,R) : \mu(B_r(x,\varepsilon)) \le \delta\big\} \ge \varepsilon\big\} \le \varepsilon.$$

Remark 4.7 (Gromov-weak tightness)**.** *A characterization of Gromov-weak tightness of probability measures of metric finite measures spaces is given in Theorem 3 in* [GPW09a] *(compare also* [GPW09a, Remark 7.2(ii)]*).* $\square$

Proof of Corollary 4.6. "1.⇒2." Assume that the family Γ is Gromov-vaguely tight. Then we can find for all $\varepsilon > 0$ a compact set $\mathbb{K}_\varepsilon$ with $\mathbb{P}(\mathbb{K}_\varepsilon) \ge 1 - \varepsilon$ for all $\mathbb{P} \in \Gamma$. In particular, by Corollary 4.3, the sets $\mathbb{K}_\varepsilon\restriction_R$ are Gromov-weakly relatively compact for all $R > 0$. Because $(\psi_R)_*(\mathbb{P})(\mathbb{K}_\varepsilon\restriction_R) \ge \mathbb{P}(\mathbb{K}_\varepsilon) \ge 1 - \varepsilon$ for all $\mathbb{P} \in \Gamma$, the set $(\psi_R)_*(\Gamma)$ is Gromov-weakly tight.

"2.⇒1." Conversely assume that, for $\varepsilon, R > 0$, $\mathbb{K}_{\varepsilon,R}$ is a Gromov-weakly compact set satisfying $(\psi_R)_*(\mathbb{P})(K_{\varepsilon,R}) \ge 1 - \varepsilon$. Then, for all $\varepsilon > 0$, $\mathbb{K}_\varepsilon := \big\{ \mathcal{X} \in \mathbb{X} : \mathcal{X}\restriction_n \in \mathbb{K}_{2^{-n}\varepsilon,n}\ \forall n \in \mathbb{N} \big\}$ is a Gromov-vaguely relatively compact set which satisfies $\mathbb{P}(K_\varepsilon) \ge 1 - \varepsilon$ for all $\mathbb{P} \in \Gamma$.

"2.⇔3." follows from Theorem 3 in [GPW09a]. $\square$

Constructing a complete metric on $\mathbb{X}$ that metrizes the Gromov-vague topology is now standard.

Proposition 4.8 ($\mathbb{X}$ is Polish)**.** *The space $\mathbb{X}$ of metric boundedly finite measure spaces equipped with the Gromov-vague topology is a Polish space.*

Proof. One possible choice of a complete metric is

$$d_{\mathrm{GP}}^{\#}(\mathcal{X},\mathcal{Y}) := \int_{\mathbb{R}_+} \mathrm{d}R\, e^{-R}\big(1 \wedge d_{\mathrm{GP}}(\mathcal{X}\restriction_R, \mathcal{Y}\restriction_R)\big). \tag{4.8}$$

Indeed, that $d_{\mathrm{GP}}^{\#}$ induces the Gromov-vague topology is shown in Lemma 2.6, and separability is obvious. To see completeness, consider a Cauchy sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ in $\mathbb{X}$. Then $\mathcal{X}_n\restriction_R$ is a Cauchy sequence with respect to d_{GP} for Lebesgue-almost all $R > 0$. By completeness of d_{GP}, $\{ \mathcal{X}_n\restriction_R : n \in \mathbb{N} \}$ is relatively compact in the Gromov-weak topology for these $R > 0$. By Corollary 4.3, this implies relative compactness of $\{ \mathcal{X}_n : n \in \mathbb{N} \}$ in the Gromov-vague topology. Hence the sequence converges Gromov-vaguely. $\square$

Unfortunately, the subspaces $\mathbb{X}_{\mathrm{HB}}$ and $\mathbb{X}_{\mathrm{c}}$ are not Polish, and hence it is impossible to find a complete metric inducing Gromov-vague topology on them. They are, however, Lusin spaces. Recall that a metrizable topological space is, by definition, a Lusin space if it is the image of a Polish space under a continuous, bijective map.

Corollary 4.9 ($\mathbb{X}_{\mathrm{HB}}$ and $\mathbb{X}_{\mathrm{c}}$ are Lusin)**.** *The space $\mathbb{X}_{\mathrm{HB}}$ of Heine-Borel locally finite measure spaces, equipped with the Gromov-vague topology, is a Lusin space but not Polish. The same is true for the space $\mathbb{X}_{\mathrm{c}}$ of compact metric finite measure spaces.*

Proof. $\mathbb{X}_{\mathrm{HB}}$ and $\mathbb{X}_{\mathrm{c}}$ are measurable subsets (Corollary 3.6) of the Polish space $\mathbb{X}$. Hence they are Lusin by Theorem 8.2.10 of [Coh80].

To see that $\mathbb{X}_{\mathrm{HB}}$ is not Polish, note that it is a dense subspace of $\mathbb{X}$, and using Lemma 3.5 we see that its complement, $\mathbb{X} \setminus \mathbb{X}_{\mathrm{HB}}$, contains a countable intersection of open dense sets, namely $G := \bigcap_{a>0,\, a\in\mathbb{Q}} \{\, \mathcal{X} \in \mathbb{X} : \mathfrak{m}_1^1(\mathcal{X}) < a \,\}$. Such a subspace cannot be Polish by standard arguments (see also [Löh13, Remark 4.7]), which we recall for the reader's convenience. Assume for a contradiction that $\mathbb{X}_{\mathrm{HB}}$ is Polish. By the Mazurkiewicz theorem ([Coh80, Theorem 8.1.4]), it is a countable intersection of open sets, $\mathbb{X}_{\mathrm{HB}} = \bigcap_{n\in\mathbb{N}} U_n$, say. Obviously, the U_n have to be dense, because $\mathbb{X}_{\mathrm{HB}}$ is. Now $\mathbb{X}_{\mathrm{HB}} \cap G$ is also a countable intersection of open dense sets, hence it is dense by the Baire category theorem ([Coh80, Theorem D.37]). This is a contradiction, because $G \subseteq \mathbb{X} \setminus \mathbb{X}_{\mathrm{HB}}$.

The same reasoning also applies to $\mathbb{X}_{\mathrm{c}}$, hence $\mathbb{X}_{\mathrm{c}}$ is also not Polish. □

5. The Gromov-Hausdorff-vague topology

In this section we introduce with the *Gromov-Hausdorff-vague topology* a topology which is stronger than the Gromov-vague topology. The need for such a topology can be motivated by situations as in Example 4.2, and by the fact that there are sequences of finite measures $(\mu_n)_{n\in\mathbb{N}}$ on a common compact space (E,d), such that $\mu_n \Rightarrow \mu$, as $n\to\infty$, but their supports do not converge. The convergence of supports, however, plays a crucial rôle for the convergence of associated random walks to a Brownian motion on the limit space (see [ALW17]). We define the stronger topology based on isometric embeddings, discuss its connection to the related measured (Gromov-)Hausdorff topology and to the Gromov-Hausdorff-Prohorov metric known from the literature, state a stability result, and characterize compact sets. A main result of this section is Polishness of the Gromov-Hausdorff-weak and -vague topologies (Propositions 5.5 and 5.12). Interpreted in terms of the Gromov-Hausdorff-Prohorov metric as used in [Mie09], this means that the subspace of metric measure spaces with full support of the measure is Polish although it is not closed (Corollary 5.6).

We cannot, of course, build such a strong notion of convergence on the notion of sampling alone, and therefore rather use an isometric embedding approach (compare Proposition 4.1). Recall that the Hausdorff distance between two closed subsets A, B of a metric space (E,d) of bounded diameter is defined by

$$d_{\mathrm{H}}\big(A,B\big) := \inf\big\{\varepsilon>0 :\ A^\varepsilon \supseteq B,\ \text{and}\ B^\varepsilon \supseteq A\big\}, \tag{5.1}$$

where once more $A^\varepsilon := \{x \in A :\ d(x,A) \le \varepsilon\}$ denotes the closed ε-neighborhood of A. Recall that $\mathbb{X}_{\mathrm{fin}}$ denotes the space of metric finite measure spaces.

Definition 5.1 ((pointed) Gromov-Hausdorff-weak topology)**.** *Let for each* $n \in \mathbb{N}\cup\{\infty\}$, $\mathcal{X}_n := (X_n, r_n, \rho_n, \mu_n) \in \mathbb{X}_{\mathrm{fin}}$. *We say that* $(\mathcal{X}_n)_{n\in\mathbb{N}}$ *converges to* $\mathcal{X}_\infty$ *in* Gromov-Hausdorff weak topology *if and only if there exists a pointed metric space* (E, d_E, ρ_E) *and, for each* $n\in\mathbb{N}\cup\{\infty\}$, *an isometry* $\varphi_n\colon \operatorname{supp}(\mu_n) \to E$ *with* $\varphi_n(\rho_n) = \rho_E$, *and such that in addition to*

$$(\varphi_n)_*\mu_n \underset{n\to\infty}{\Longrightarrow} (\varphi_\infty)_*\mu_\infty, \tag{5.2}$$

also

$$d_{\mathrm{H}}\Big(\varphi_n\big(\operatorname{supp}(\mu_n)\big),\, \varphi_\infty\big(\operatorname{supp}(\mu_\infty)\big)\Big) \underset{n\to\infty}{\longrightarrow} 0. \tag{5.3}$$

A very similar topology for compact metric measure spaces was first introduced in [Fuk87] under the name *measured Hausdorff topology* (often referred to as measured Gromov-Hausdorff topology) and further discussed in [EW06, Mie09]. The definition of this topology is exactly the same as that of the Gromov-Hausdorff-weak topology, except that $\operatorname{supp}(\mu_n)$ is replaced by X_n, $n\in\mathbb{N}\cup\{\infty\}$. The consequence is that, when comparing compact metric measure spaces, the geometric structure outside the support is taken into account, while it is ignored by our definition.

It is important to note that this leads to different equivalence classes, i.e., the measured Hausdorff topology is *not* defined on $\mathbb{X}_c$, but rather on a space of equivalence classes with respect to the following equivalence relation. We say that two metric measure spaces (X, r, ρ, μ) and (X', r', ρ', μ') are *strongly equivalent* if and only if there is a surjective isometry $\phi\colon X \to X'$ such that $\phi(\rho) = \rho'$ and $\phi_*(\mu) = \mu'$. Define

$$\mathfrak{X}_c := \{\, \text{strong equivalence classes of compact metric measure spaces} \,\}. \tag{5.4}$$

it is well-known that the measured Hausdorff topology is induced by the so-called *Gromov-Hausdorff-Prohorov metric* defined on $\mathfrak{X}_c$ as follows. For $\mathcal{X} = (X, r, \rho, \mu)$, $\mathcal{X}' = (X', r', \rho', \mu') \in \mathfrak{X}_c$,

$$d_{\mathrm{GHP}}(\mathcal{X}, \mathcal{X}') := \inf_d d_{\mathrm{Pr}}^{(X \sqcup X', d)}(\mu, \mu') + d_{\mathrm{H}}^{(X \sqcup X', d)}(X, X') + d(\rho, \rho'), \tag{5.5}$$

where the infimum is taken over all metrics d on $X \sqcup X'$ that extend both r and r'. Note that $(\mathfrak{X}_c, d_{\mathrm{GHP}})$ is a complete, separable metric space (see [Mie09, Proposition 8]).

Now we can easily identify $\mathbb{X}_c$ with the subspace of $\mathfrak{X}_c$ that consists of all (strong equivalence classes of) compact metric spaces with a measure of full support, i.e. with

$$\mathfrak{X}_c^{\mathrm{supp}} := \big\{\, (X, r, \rho, \mu) \in \mathfrak{X}_c : X = \mathrm{supp}(\mu) \,\big\}, \tag{5.6}$$

by choosing representatives with full support from the larger equivalence classes of $\mathbb{X}_c$, i.e. via the injective map

$$\iota\colon \mathbb{X}_c \to \mathfrak{X}_c^{\mathrm{supp}}, \qquad (X, r, \rho, \mu) \mapsto \big(\mathrm{supp}(\mu), r, \rho, \mu\big). \tag{5.7}$$

It is obvious that ι is a homeomorphism if we equip $\mathbb{X}_c$ with the Gromov-Hausdorff-weak and $\mathfrak{X}_c^{\mathrm{supp}}$ with the measured Hausdorff topology. Its inverse ι^{-1} can naturally be extended to all of $\mathfrak{X}_c$, but this extension looses continuity, as we show in the following remark.

Remark 5.2 (support projection)**.** *Equip $\mathfrak{X}_c$ with the measured Hausdorff topology and $\mathbb{X}_c$ with the Gromov-Hausdorff-weak topology. The* support projection

$$\pi^{\mathrm{supp}}\colon \mathfrak{X}_c \to \mathfrak{X}_c^{\mathrm{supp}}, \qquad (X, r, \rho, \mu) \mapsto \big(\mathrm{supp}(\mu), r, \rho, \mu\big). \tag{5.8}$$

is an open map, but neither continuous nor closed. In particular, associating to a strong equivalence class of metric measure spaces in $\mathfrak{X}_c$ the corresponding equivalence class in $\mathbb{X}_c$ is not a continuous operation, although it induces a homeomorphism from $\mathfrak{X}_c^{\mathrm{supp}}$ onto $\mathbb{X}_c$.

Remark 5.3 (full support assumption)**.** *The requirement that the measure on a metric space has full support is not unnatural. It plays, for instance, a crucial rôle for defining Markov processes via Dirichlet forms (a particular example is* [ALW17]*), and is even included in the definition of "Radon measure" in* [FOT11]. $\square$

Note that $\mathfrak{X}_c^{\mathrm{supp}}$ is not closed in $\mathfrak{X}_c$, hence transporting the Gromov-Hausdorff-Prohorov metric d_{GHP} with ι back to $\mathbb{X}_c$ does not yield a complete metric. The following proposition shows, however, that we can find a different, complete metric for the Gromov-Hausdorff-weak topology on $\mathbb{X}_c$. This also implies that, although $(\mathfrak{X}_c^{\mathrm{supp}}, d_{\mathrm{GHP}})$ is not complete, it can still be used as a Polish state-space, because the induced topological space is Polish. To define the complete metric on $\mathbb{X}_c$, we use the global lower mass function $\mathfrak{m}_\delta$ from (3.2), the Gromov-Hausdorff-Prohorov metric d_{GHP} from (5.5), and the homeomorphism ι from (5.7). Recall that $\mathfrak{m}_\delta > 0$ on $\mathbb{X}_c$ for every $\delta > 0$ by Lemma 3.4.

Definition 5.4. *For $\mathcal{X}, \mathcal{X}' \in \mathbb{X}_c$, let*

$$d_{\mathrm{sGHP}}(\mathcal{X}, \mathcal{X}') := d_{\mathrm{GHP}}\big(\iota(\mathcal{X}), \iota(\mathcal{X}')\big) + \int_0^1 \mathrm{d}\delta\; 1 \wedge \left| \frac{1}{\mathfrak{m}_\delta(\mathcal{X})} - \frac{1}{\mathfrak{m}_\delta(\mathcal{X}')} \right|. \tag{5.9}$$

We call d_{sGHP} the support Gromov-Hausdorff-Prohorov metric.

Proposition 5.5 (($\mathbb{X}_c, d_{sGHP}$) is a complete metric space)**.** *The metric d_{sGHP} induces the Gromov-Hausdorff-weak topology on $\mathbb{X}_c$. Furthermore, $(\mathbb{X}_c, d_{sGHP})$ is a complete, separable metric space.*

Proof. Let $(\mathcal{X}_n)_{n\in\mathbb{N}}$ and $\mathcal{X}$ be in $\mathbb{X}_c$. Then $\mathfrak{m}_\delta(\mathcal{X}) > 0$, for all $\delta > 0$. Thus by definition, $d_{sGHP}(\mathcal{X}_n, \mathcal{X}) \underset{n\to\infty}{\longrightarrow} 0$ if and only if

$$d_{GHP}\big(\iota(\mathcal{X}_n), \iota(\mathcal{X})\big) \underset{n\to\infty}{\longrightarrow} 0, \tag{5.10}$$

and for almost all $\delta > 0$,

$$\mathfrak{m}_\delta(\mathcal{X}_n) \underset{n\to\infty}{\longrightarrow} \mathfrak{m}_\delta(\mathcal{X}). \tag{5.11}$$

Because ι is a homeomorphism, (5.10) is equivalent to the Gromov-Hausdorff-weak convergence $\mathcal{X}_n \underset{n\to\infty}{\longrightarrow} \mathcal{X}$. We have to show that this already implies (5.11), i.e. that $\mathcal{X}$ is continuity point of $\mathfrak{m}_\delta$ w.r.t. Gromov-Hausdorff-weak topology for almost all $\delta > 0$. To see this, recall that $\mathfrak{m}_\delta$ is upper semi-continuous w.r.t. Gromov-vague topology (Lemma 3.2), and a fortiori also w.r.t. Gromov-Hausdorff-weak topology. Assume that all $\mathcal{X}_n = (X_n, r_n, \rho_n, \mu_n)$ and $\mathcal{X} = (X, r, \rho, \mu)$ are embedded in some common space (E, d, ρ) such that μ_n converges weakly to μ and $\mathrm{supp}(\mu_n)$ in Hausdorff metric to $\mathrm{supp}(\mu)$. Then, for every $\hat{\delta} < \delta$ and n sufficiently large, every δ-ball around some $y \in \mathrm{supp}(\mu_n)$ contains a $\hat{\delta}$-ball around some $x \in \mathrm{supp}(\mu)$. Therefore, $\liminf_{n\to\infty} \mathfrak{m}_\delta(\mathcal{X}_n) \geq \mathfrak{m}_{\hat{\delta}}(\mathcal{X})$. This means that $\mathfrak{m}_\delta$ is Gromov-Hausdorff-weakly lower semi-continuous in $\mathcal{X}$ for every $\delta > 0$ with $\mathfrak{m}_\delta(\mathcal{X}) = \sup_{\hat{\delta}<\delta} \mathfrak{m}_{\hat{\delta}}(\mathcal{X})$. Because $\delta \mapsto \mathfrak{m}_\delta(\mathcal{X})$ is an increasing function, this is the case for almost all $\delta > 0$. This means that (5.11) is implied by Gromov-Hausdorff-weak convergence, and hence d_{sGHP} induces Gromov-Hausdorff-weak topology as claimed.

That $(\mathbb{X}_c, d_{sGHP})$ is a *separable* metric space is obvious, and it remains to show its *completeness*. Consider a d_{sGHP}-Cauchy sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ in $\mathbb{X}_c$. Then, by completeness of d_{GHP} on $\mathfrak{X}_c$, the sequence $\big(\iota(\mathcal{X}_n)\big)_{n\in\mathbb{N}}$ converges in measured Hausdorff topology to some $\mathcal{Y} = (X, r, \rho, \mu) \in \mathfrak{X}_c$. We have to show $\mathcal{Y} \in \iota(\mathbb{X}_c) = \mathfrak{X}_c^{\mathrm{supp}}$. Assume for a contradiction that this is not the case, i.e. there exists $x \in X \setminus \mathrm{supp}(\mu)$. Then there is a $\delta > 0$ with $B(x, 2\delta) \cap \mathrm{supp}(\mu) = \emptyset$. By the measured Hausdorff convergence and the fact that $\iota(\mathcal{X}_n) \in \mathfrak{X}_c^{\mathrm{supp}}$ for all n, this clearly implies $\mathfrak{m}_\delta(\mathcal{X}_n) \underset{n\to\infty}{\longrightarrow} 0$. This, however, cannot be the case because $(\mathcal{X}_n)_{n\in\mathbb{N}}$ is a Cauchy sequence w.r.t. d_{sGHP}. □

Corollary 5.6. *The set $\mathfrak{X}_c^{\mathrm{supp}}$ of (strong equivalence classes of) compact metric full-support measure spaces with the topology induced by the Gromov-Hausdorff-Prohorov metric d_{GHP} is a Polish space (although d_{GHP} restricted to $\mathfrak{X}_c^{\mathrm{supp}}$ is not complete).*

Corollary 5.7 (Gromov-Hausdorff-weak compactness)**.** *A set $\mathbb{K} \subseteq \mathbb{X}_c$ is relatively compact in the Gromov-Hausdorff-weak topology if and only if the following hold*

1. *The set of the total masses is uniformly bounded, i.e.,*

$$\sup_{(X,r,\rho,\mu)\in\mathbb{X}_c} \mu(X) < \infty. \tag{5.12}$$

2. *For all $\varepsilon > 0$ there exists an $N_{R,\varepsilon} \in \mathbb{N}$ such that for all $(X, r, \rho, \mu) \in \mathbb{K}$, $\mathrm{supp}(\mu)$ can be covered by $N_{R,\varepsilon}$ many balls of radius ε.*
3. *$\mathbb{K}$ satisfies the global lower mass-bound property.*

Proof. From Proposition 5.5, the definition of d_{sGHP}, and the fact that d_{GHP} induces the measured Hausdorff topology, we see that $\mathbb{K}$ is Gromov-Hausdorff-weakly relatively compact if and only if $\iota(\mathbb{K})$ is relatively compact in measured Hausdorff topology, and $1/\mathfrak{m}_\delta$ is bounded on $\mathbb{K}$. The latter is obviously equivalent to the global lower mass-bound 3. If $\mathbb{K} \subseteq \mathbb{X}_1$, the measured Hausdorff relative compactness of $\iota(\mathbb{K})$ is equivalent to 2. by [EW06, Proposition 2.4] together with [BBI01, Theorem 7.4.15]. It is therefore easy to see that it is in general equivalent to 2. together with 1. (compare [GPW09a, Remark 7.2(ii)]). □

In the same way as we used the Gromov-weak topology to define the Gromov-vague topology, we also define the Gromov-Hausdorff-vague topology on $\mathbb{X}$ based on the Gromov-Hausdorff-weak topology on $\mathbb{X}_{\mathrm{fin}}$.

Definition 5.8 ((pointed) Gromov-Hausdorff-vague topology). *Let for each $n \in \mathbb{N} \cup \{\infty\}$, $\mathcal{X}_n := (X_n, r_n, \rho_n, \mu_n)$ be in $\mathbb{X}$. We say that $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}_\infty$ in* Gromov-Hausdorff vague topology *if and only if $(\mathcal{X}_n)\restriction_R \underset{n\to\infty}{\longrightarrow} (\mathcal{X}_\infty)\restriction_R$ Gromov-Hausdorff-weakly for all but countably many $R > 0$.*

The following embedding result and its corollary about Gromov-Hausdorff-vaguely compact sets are proved in the same way as Proposition 4.1 and Corollary 4.3.

Proposition 5.9 (isometric embeddings; Gromov-Hausdorff-Prohorov metric). *Let for each $n \in \mathbb{N} \cup \{\infty\}$, $\mathcal{X}_n := (X_n, r_n, \rho_n, \mu_n)$ be in $\mathbb{X}_{\mathrm{HB}}$. The following are equivalent:*

1. *$\mathcal{X}_n \underset{n\to\infty}{\longrightarrow} \mathcal{X}_\infty$, Gromov-Hausdorff vaguely.*
2. *There exists a rooted Heine-Borel space (E, d_E, ρ_E) and for each $n \in \mathbb{N} \cup \{\infty\}$ isometries $\varphi_n \colon \operatorname{supp}(\mu_n) \to E$ with $\varphi_n(\rho_n) = \rho_E$, and such that in addition to (4.2), also*

$$d_{\mathrm{H}}\Big(\varphi_n(\operatorname{supp}\mu_n) \cap \overline{B}_{d_E}(\rho_E, R),\, \varphi_\infty(\operatorname{supp}\mu_\infty) \cap \overline{B}_{d_E}(\rho_E, R)\Big) \underset{n\to\infty}{\longrightarrow} 0, \tag{5.13}$$

for all but countably many $R > 0$.

3. *$d^{\#}_{\mathrm{sGHP}}(\mathcal{X}_n, \mathcal{X}_\infty) \underset{n\to\infty}{\longrightarrow} 0$, where for $\mathcal{X}, \mathcal{X}' \in \mathbb{X}_{\mathrm{HB}}$,*

$$d^{\#}_{\mathrm{sGHP}}(\mathcal{X}, \mathcal{X}') := \int \mathrm{d}R\, e^{-R}\big(1 \wedge d_{\mathrm{sGHP}}(\mathcal{X}\restriction_R, \mathcal{X}'\restriction_R)\big). \tag{5.14}$$

Corollary 5.10 (Gromov-Hausdorff-vague compactness). *For a set $\mathbb{K} \subseteq \mathbb{X}$ the following are equivalent:*

1. *$\mathbb{K}$ is relatively compact in $\mathbb{X}$ equipped with the Gromov-Hausdorff-vague topology.*
2. *For all $R > 0$, the set of restrictions $\mathbb{K}\restriction_R := \big\{ \mathcal{X}\restriction_R : \mathcal{X} \in \mathbb{K} \big\}$ is relatively compact in the Gromov-Hausdorff-weak topology.*
3. *$\mathbb{K}\restriction_{R_k}$ is relatively compact in the Gromov-Hausdorff-weak topology for a sequence $R_k \to \infty$.*

Remark 5.11 (Gromov-Hausdorff-Prohorov and length spaces). *The measured Hausdorff topology was recently extended – under the name* Gromov-Hausdorff-Prohorov topology *– in* [ADH13] *to the space of complete, locally compact length spaces equipped with locally finite measures. The extension was done with the same localization procedure that we use. Note the following:*

1. *Complete locally compact length spaces are Heine-Borel spaces and well suited for applications concerning $\mathbb{R}$-trees. The assumption of being a length space and thereby path-connected, however, is too restrictive in general. For example, in Theorem 1 of* [ALW17] *we establish convergence in path space of continuous time random walks on discrete trees to time-changed Brownian motion on $\mathbb{R}$-trees (appearing as the Gromov-Hausdorff-vague limit of the discrete trees), where the underlying trees are encoded as metric spaces and jump rates and/or time-changes are encoded by the so-called speed measure. Since we need the speed measure to have full support, the situation is incompatible with a connectedness requirement.*
2. *In a generalized setting, the name Gromov-Hausdorff-Prohorov topology might be a bit misleading, as "Prohorov" suggests weak convergence, while the localized convergence is vague in the sense that mass can get lost. Also note that, if we drop the assumption of being length spaces, the localized convergence is not really an extension of measured Hausdorff convergence any more (compare Remark 2.8).* □

Proposition 5.12 ($\mathbb{X}_{\mathrm{HB}}$ with Gromov-Hausdorff-vague topology is Polish). *The space $\mathbb{X}_{\mathrm{HB}}$ of Heine-Borel boundedly finite measure spaces equipped with the Gromov-Hausdorff-vague topology is a Polish space.*

	$\mathbb{X}$	$\mathbb{X}_{\mathrm{HB}}$	$\mathbb{X}_{\mathrm{c}}$	$\mathbb{X}_{\mathrm{fin}}$	$\mathbb{X}_1$	$\mathbb{X}_{\mathrm{c}} \cap \mathbb{X}_1$
Gv	Polish	Lusin	Lusin	Lusin	Polish	Lusin
Gw	–	–	Lusin	Polish	Polish	Lusin
GHv	non-sep.	Polish	Lusin	non-sep.	non-sep.	Lusin
GHw	–	–	Polish	non-sep.	non-sep.	Polish

FIGURE 1. The table shows topological properties of different spaces of metric measure spaces in different topologies. Entries: "–" means not defined; "non-sep." means non-separable; "Lusin" means Lusin but *not* Polish. Spaces: the spaces are defined in Definition 2.2 and Remark 2.8. Topologies: Gv=Gromov-vague, Gw=Gromov-weak, GHv=Gromov-Hausdorff-vague, GHw=Gromov-Hausdorff-weak.

Proof. We follow the proof of Proposition 4.8 and define

$$d^{\#}_{\mathrm{sGHP}}(\mathcal{X}, \mathcal{Y}) := \int_{\mathbb{R}_+} \mathrm{d}R\, e^{-R}\big(1 \wedge d_{\mathrm{sGHP}}(\mathcal{X}\restriction_R, \mathcal{Y}\restriction_R)\big). \tag{5.15}$$

We know from Proposition 5.9 that $d^{\#}_{\mathrm{sGHP}}$ induces the Gromov-Hausdorff-vague topology. Separability and completeness follow from the corresponding properties of d_{sGHP} (Proposition 5.5) and the compactness criterion given in Corollary 5.10, in the same way as in the proof of Proposition 4.8. □

Even though the Gromov-Hausdorff-vague topology is nice (i.e. Polish) on $\mathbb{X}_{\mathrm{HB}}$ and defined on all of $\mathbb{X}$, it appears to be too strong to be useful on the larger space.

Remark 5.13 (Gromov-Hausdorff-vague topology is non-separable on $\mathbb{X}$)**.** *The spaces $\mathbb{X}$ and $\mathbb{X}_1$, equipped with the Gromov-Hausdorff-vague topology, are not separable. In particular they are not Lusin spaces. Indeed, we can topologically embed the non-separable space l^∞_+ into $\mathbb{X}_1$ as follows: for $n \in \mathbb{N}$ and $a \in \mathbb{R}_+$, let $A^n_a := \{n\} \times [0, a]^n$, and μ^n_a some measure on A^n_a with full support and total mass 2^{-n}. Define $\psi\colon l^\infty_+ \to \mathbb{X}_1$ by $\psi(a) := \big(\bigcup_{n\in\mathbb{N}} A^n_{a_n}, r, \rho, \sum_{n\in\mathbb{N}} \mu^n_{a_n}\big)$, where $\rho = (1, 0)$, and r is the supremum of the discrete metric on the first component and the Euclidean metric on the second component. It is straightforward to check that ψ is a homeomorphism onto its image.* □

We know from Propositions 4.8, 5.5, and 5.12 that $\mathbb{X}$ with Gromov-vague topology, $\mathbb{X}_{\mathrm{c}}$ with Gromov-Hausdorff-weak topology, and $\mathbb{X}_{\mathrm{HB}}$ with Gromov-Hausdorff-vague topology are Polish spaces. Furthermore, it is known from [GPW09a, Theorem 1] that $\mathbb{X}_1$ and $\mathbb{X}_{\mathrm{fin}}$ with Gromov-weak topology are Polish. In the case of $\mathbb{X}_1$, this is also true for the Gromov-vague topology, although $\mathbb{X}_1$ is not Gromov-vaguely closed in $\mathbb{X}$ (see Remark 2.8). On the other hand, Corollary 4.9 proves that $\mathbb{X}_{\mathrm{c}}$ and $\mathbb{X}_{\mathrm{HB}}$ with Gromov-vague topology are Lusin but *not* Polish. Similar arguments also show that $\mathbb{X}_{\mathrm{fin}}$ with Gromov-vague topology and $\mathbb{X}_{\mathrm{c}}$ with Gromov-weak as well as with Gromov-Hausdorff-vague topology are Lusin and not Polish. Gromov-Hausdorff-vague topology is not even separable on $\mathbb{X}$ and $\mathbb{X}_1$ by Remark 5.13. We summarize the situation in Figure 1.

We conclude this section with a stability property, which is the analogue of Lemma 2.9. It is an immediate consequence of the definition of the Gromov-Hausdorff-vague topology by means of isometric embeddings. The proof follows the same lines as the proof of Lemma 2.9 and is therefore omitted.

Lemma 5.14 (perturbation of measures)**.** *Consider $\mathcal{X} = (X, r, \rho, \mu)$, $\mathcal{X}_n = (X_n, r_n, \rho_n, \mu_n) \in \mathbb{X}$, and another boundedly finite measure μ'_n on X_n, $n \in \mathbb{N}$. Assume that $\mathcal{X}_n \underset{n\to\infty}{\longrightarrow} \mathcal{X}$ Gromov-Hausdorff-vaguely, and that there exists a sequence $R_k \to \infty$ such that for all $k \in \mathbb{N}$,*

$$d^{(X_n,r_n)}_{\mathrm{Pr}}\big(\mu_n\restriction_{R_k}, \mu'_n\restriction_{R_k}\big) \underset{n\to\infty}{\longrightarrow} 0, \quad \textit{and} \quad d^{(X_n,r_n)}_{\mathrm{H}}\big(\mathrm{supp}(\mu_n\restriction_{R_k}), \mathrm{supp}(\mu'_n\restriction_{R_k})\big) \underset{n\to\infty}{\longrightarrow} 0. \tag{5.16}$$

Then $(X_n, r_n, \rho_n, \mu'_n)$ converges Gromov-Hausdorff-vaguely to $\mathcal{X}$.

Example 5.15 (Normalized length measure versus degree measure)**.** *Consider a graph theoretic tree T' which is locally finite, i.e.* $\deg(v) < \infty$ *for all* $v \in T'$*, where* deg *is the degree of a node. Equip T' with the graph distance r', i.e. the length of the shortest path, and fix a root $\rho' \in T'$. Recall the notion of $\mathbb{R}$-tree from Example 2.3. It is well known that (T', r') can be embedded isometrically into a complete, locally compact $\mathbb{R}$-tree (T, r) in an essentially unique way. Denote the image of ρ' by ρ and the image of T' by* $\mathrm{nod}(T)$*. On T', we consider two natural measures. The* node measure $\mu_{T'}^{\mathrm{nod}}$*, which is just the counting measure on the nodes (except the root), and the* degree measure $\mu_{T'}^{\mathrm{deg}}$*, which is proportional to the degree of the node. The push-forwards on T are given by*

$$\mu_T^{\mathrm{nod}} := \sum_{x \in \mathrm{nod}(T) \setminus \{\rho\}} \delta_x \quad \text{and} \quad \mu_T^{\mathrm{deg}} := \tfrac{1}{2} \sum_{x \in \mathrm{nod}(T)} \deg(x) \cdot \delta_x. \tag{5.17}$$

Note that $(T', r', \rho', \mu_{T'}^{\mathrm{deg}}) \cong (T, r, \rho, \mu_T^{\mathrm{deg}})$*, and similarly for the node measure. On T, there is also a third natural measure, namely the* length measure $\lambda = \lambda_{(T,r)}$*, which is the 1-dimensional Hausdorff measure on* $T \setminus \mathrm{lf}(T)$*, where* $\mathrm{lf}(T) = \{\, x \in T : T \setminus \{x\} \text{ is connected} \}$ *is the set of leaves of T. Note that* $\lambda_{(T,r)}(T) = \mu_T^{\mathrm{nod}}(T) = \mu_T^{\mathrm{deg}}(T)$.

Now consider a sequence $(T_n')_{n\in\mathbb{N}}$ of locally finite, graph theoretic trees, and the rooted $\mathbb{R}$-trees (T_n, r_n, ρ_n) constructed as above. We assume that there are two sequences $(\alpha_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ of positive numbers, both of which converge to 0*, such that*

$$\big(T_n, \alpha_n r_n, \rho_n, \beta_n \lambda_{(T_n,r_n)}\big) \underset{n\to\infty}{\longrightarrow} \mathcal{X}, \tag{5.18}$$

Gromov-Hausdorff-vaguely for some $\mathcal{X} = (T, r, \rho, \mu) \in \mathbb{X}_{\mathrm{HB}}$, which is necessarily an $\mathbb{R}$-tree. Such a convergence can often be deduced via convergence of excursions, see Proposition 7.5 and Example 7.6 below. We claim that in this case, the length measure can be replaced by the degree measure or the node measure, i.e. that (5.18) *implies the Gromov-Hausdorff-vague convergences*

$$\big(T_n, \alpha_n r_n, \rho_n, \beta_n \mu_{T_n}^{\mathrm{deg}}\big) \underset{n\to\infty}{\longrightarrow} \mathcal{X} \quad \text{and} \quad \big(T_n, \alpha_n r_n, \rho_n, \beta_n \mu_{T_n}^{\mathrm{nod}}\big) \underset{n\to\infty}{\longrightarrow} \mathcal{X}. \tag{5.19}$$

Indeed, we have $\mathrm{supp}(\mu_{T_n}^{\mathrm{nod}}) = \mathrm{supp}(\mu_{T_n}^{\mathrm{deg}}) = \mathrm{nod}(T_n)$, $\mathrm{supp}(\lambda_{(T_n,r_n)}) = T_n$ *and, for every* $R > 0$,

$$d_{\mathrm{H}}\big(\mathrm{nod}(T_n) \cap B(\rho_n, R),\, B(\rho_n, R)\big) \le \alpha_n \underset{n\to\infty}{\longrightarrow} 0. \tag{5.20}$$

For the Prohorov distance, assume first that the diameter of T_n is smaller than R. Then

$$d_{\mathrm{Pr}}^{(T_n,r_n)}(\mu_{T_n}^{\mathrm{deg}}, \lambda_{(T_n,r_n)}) \le \tfrac{1}{2}\alpha_n \quad \text{and} \quad d_{\mathrm{Pr}}^{(T_n,r_n)}(\mu_{T_n}^{\mathrm{nod}}, \lambda_{(T_n,r_n)}) \le \alpha_n. \tag{5.21}$$

In the general case, we have to take boundary effects into account. Using the annulus $S^\varepsilon(\rho_n, R) := \overline{B}(\rho_n, R + \frac{1}{2}\varepsilon) \setminus B(\rho_n, R - \frac{1}{2}\varepsilon)$*, we obtain*

$$d_{\mathrm{Pr}}^{(T_n,r_n)}(\mu_{T_n}^{\mathrm{deg}}\restriction_R,\, \lambda_{(T_n,r_n)}\restriction_R) \;\le\; \tfrac{1}{2}\alpha_n \vee \beta_n \cdot \lambda_{(T_n,r_n)}\big(S^{\alpha_n}(\rho_n, R)\big), \tag{5.22}$$

and a similar estimate for $\mu_{T_n}^{\mathrm{nod}}$ instead of $\mu_{T_n}^{\mathrm{deg}}$. Using (5.18) *we see that* $\beta_n \lambda_{(T_n,r_n)}\big(S^{\alpha_n}(\rho_n, R)\big)$ *tends to zero for all R with* $\mu\big(S(\rho, R)\big) = 0$*. Therefore the claimed Gromov-Hausdorff-vague convergences* (5.19) *follow from* (5.22), (5.20) *and Lemma 5.14.* □

6. Closing the gap

In this section we prove the main criterion for convergence in Gromov-Hausdorff-vague topology. We shall use notation used in (1.2), (1.3) and the definitions of the lower mass functions $\mathfrak{m}_\delta^R$ and $\mathfrak{m}_\delta$ from (3.1) and (3.2), respectively.

In order for a sequence $\big(\mathcal{X}_n\big)_{n\in\mathbb{N}} := \big(X_n, r_n, \rho_n, \mu_n\big)_{n\in\mathbb{N}}$ of compact metric finite measure spaces to converge in Gromov-Hausdorff-weak topology to a space $\mathcal{X} = (X, r, \rho, \mu) \in \mathbb{X}_{\mathrm{c}}$, it certainly has to converge in the weaker Gromov-weak topology. This is a kind of "finite-dimensional convergence", which is expressible in terms of sampling finite sub-spaces:

1. For all $k \in \mathbb{N}$, and $\varphi \in \bar{\mathcal{C}}(\mathbb{R}_+^{\binom{k+1}{2}})$

$$\begin{aligned}&\int \mu_n^{\otimes k}(\mathrm{d}(x_1^n, \dots, x_k^n))\, \varphi\big((r_n(x_i^n, x_j^n))_{0 \le i < j \le k}\big)\\ &\underset{n\to\infty}{\longrightarrow} \int \mu^{\otimes k}(\mathrm{d}(x_1, \dots, x_k))\, \varphi\big((r(x_i, x_j))_{0 \le i < j \le k}\big)\end{aligned} \tag{6.1}$$

where we put $x_0^n := \rho$ and $x_0 := \rho$.

We show in Theorem 6.1 below that, given 1., Gromov-Hausdorff-weak convergence follows from a simple "tightness condition", which is given in terms of the lower mass function:

2. For all $\delta > 0$, $\liminf_{n\to\infty} \mathfrak{m}_\delta(\mathcal{X}_n) > 0$.

Note that for checking 1. and 2., we do not have to find any embedding into a common metric space. We actually show that 1. and 2. together are even equivalent to Gromov-Hausdorff-weak convergence, and this characterization even holds if the $\mathcal{X}_n$ are not compact (but $\mathcal{X}$ is).

Theorem 6.1 (Gromov-weak versus Gromov-Hausdorff-weak convergence)**.** *Let $\mathcal{X} = (X, r, \rho, \mu)$ and $\mathcal{X}_n = (X_n, r_n, \rho_n, \mu_n)$, $n \in \mathbb{N}$, be metric finite measure spaces. Then the following are equivalent.*

1. *$(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges in Gromov-weak topology to $\mathcal{X}$, and for all $\delta > 0$,*

$$\liminf_{n\to\infty} \mathfrak{m}_\delta(\mathcal{X}_n) > 0. \tag{6.2}$$

2. *$\mathcal{X}$ is compact, and $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges in Gromov-Hausdorff-weak topology to $\mathcal{X}$.*

If $\mathcal{X}_n$ is compact for all $n \in \mathbb{N}$, the following is also equivalent:

3. *$(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges in Gromov-weak topology to $\mathcal{X}$, and $\{\mathcal{X}_n : n \in \mathbb{N}\}$ satisfies the global lower mass-bound property (Definition 3.1).*

Proof. *"2.⇒1."* Assume $(\mathrm{supp}(\mu_n), r_n, \rho_n, \mu_n) \underset{n\to\infty}{\longrightarrow} (\mathrm{supp}(\mu), r, \rho, \mu)$ Gromov-Hausdorff-weakly. We may assume w.l.o.g. that $X = \mathrm{supp}(\mu)$, $X_n = \mathrm{supp}(\mu_n)$, and that X_n and X are embedded into a complete, separable metric space (E, d) such that

$$d_{\mathrm{Pr}}^{(E,d)}\big(\mu_n, \mu\big) \underset{n\to\infty}{\longrightarrow} 0, \quad \text{and} \quad d_{\mathrm{H}}^{(E,d)}\big(X_n, X\big) \underset{n\to\infty}{\longrightarrow} 0. \tag{6.3}$$

Furthermore let (X, r) be compact. We need to show (6.2). Assume to the contrary that there exists $\delta > 0$ and $x_n \in X_n$ such that $\liminf_{n\to\infty} \mu_n\big(B(x_n, 2\delta)\big) = 0$. Due to (6.3) we can find $y_n \in X$ with $d(x_n, y_n) \underset{n\to\infty}{\longrightarrow} 0$. Moreover by (6.3),

$$\liminf_{n\to\infty} \mu\big(B(y_n, \delta)\big) \le \liminf_{n\to\infty} \mu_n\big(B(x_n, 2\delta)\big) = 0. \tag{6.4}$$

As X is compact, we may assume w.l.o.g. that y_n converges to some $y \in X$. Then $\mu\big(B(y, \delta)\big) \le \liminf \mu\big(B(y_n, \delta)\big) = 0$, which contradicts $X = \mathrm{supp}(\mu)$.

"1.⇒2." Assume that $\mathcal{X}_n \underset{n\to\infty}{\longrightarrow} \mathcal{X}$ Gromov-weakly, and w.l.o.g. that $X = \mathrm{supp}(\mu)$, $X_n = \mathrm{supp}(\mu_n)$, and that X_n and X are embedded into a complete, separable metric space (E, d) such that $d_{\mathrm{Pr}}^{(E,d)}\big(\mu_n, \mu\big) \underset{n\to\infty}{\longrightarrow} 0$, and that (6.2) holds. Then, for all $\varepsilon > 0$, we can find $n_0 = n_0(\varepsilon) \in \mathbb{N}$ such that for all $n \ge n_0$,

$$d_{\mathrm{Pr}}^{(E,d)}\big(\mu_n, \mu\big) < \varepsilon \wedge \inf_{y\in X_n} \mu_n\big(\overline{B}(y, \varepsilon)\big) \wedge \inf_{x\in X} \mu\big(\overline{B}(x, \varepsilon)\big), \tag{6.5}$$

where we used that

$$\inf_{x\in X} \mu\big(\overline{B}(x, \varepsilon)\big) \ge \liminf_{n\to\infty} \inf_{y\in X_n} \mu_n\big(\overline{B}(y, \tfrac{1}{2}\varepsilon)\big) > 0. \tag{6.6}$$

Then, for all $y \in X_n$, $\overline{B}(y, \varepsilon) \cap X^\varepsilon \ne \emptyset$, and thus also $\overline{B}(y, 2\varepsilon) \cap X \ne \emptyset$. Similarly, $\overline{B}(x, 2\varepsilon) \cap X_n \ne \emptyset$ for all $x \in X$, and hence $d_{\mathrm{H}}^{(E,d)}(X_n, X) \le 2\varepsilon$.

Compactness of $\mathcal{X}$ follows directly from (6.6) and Lemma 3.4.

"3.⇔1." Obviously, the global lower mass-bound is equivalent to (6.2) together with $\mathfrak{m}_\delta(\mathcal{X}_n) > 0$ for all $n \in \mathbb{N}$ and $\delta > 0$. The last condition is satisfied for compact spaces by Lemma 3.4. □

The following corollaries are now obvious.

Corollary 6.2 (Gromov-vague versus Gromov-Hausdorff-vague convergence)**.** *Let $\mathcal{X} = (X, r, \rho, \mu)$ and $\mathcal{X}_n = (X_n, r_n, \rho_n, \mu_n)$, $n \in \mathbb{N}$, be metric boundedly finite measure spaces. Then the following are equivalent.*

1. $(\mathcal{X}_n)_{n\in\mathbb{N}}$ *converges in Gromov-vague topology to $\mathcal{X}$, and for all $\delta > 0$ and $R > 0$,*

$$\liminf_{n\to\infty} \mathfrak{m}_\delta^R(\mathcal{X}_n) > 0. \tag{6.7}$$

2. $\mathrm{supp}(\mu)$ *is Heine-Borel, and* $\big(\mathrm{supp}(\mu_n), r_n, \mu_n\big)_{n\in\mathbb{N}}$ *converges in Gromov-Hausdorff-vague topology to* $\big(\mathrm{supp}(\mu), r, \mu\big)$*.*

If $\mathcal{X}_n$ is Heine-Borel for all $n \in \mathbb{N}$, the following is also equivalent:

3. $(\mathcal{X}_n)_{n\in\mathbb{N}}$ *converges in Gromov-vague topology to $\mathcal{X}$, and $\{\, \mathcal{X}_n : n \in \mathbb{N} \}$ satisfies the local lower mass-bound property (3.3).*

Corollary 6.3 (Polish subspaces)**.** *Let $\mathbb{K} \subseteq \mathbb{X}_{\mathrm{HB}}$ be a space of Heine-Borel locally finite measure spaces satisfying the local lower mass-bound property (3.3). Then its closure $\overline{\mathbb{K}}$ in $\mathbb{X}$ (w.r.t. the Gromov-vague topology) is a Polish subspace of $\mathbb{X}_{\mathrm{HB}}$. Furthermore, the Gromov-vague topology and the Gromov-Hausdorff-vague topology coincide on $\overline{\mathbb{K}}$.*

Corollary 6.4 (Gromov-vague implies Gromov-Hausdorff-vague on a large subset)**.** *Let $\mathcal{X} = (X, r, \rho, \mu)$ and $\mathcal{X}_n = (X_n, r_n, \rho_n, \mu_n)$, $n \in \mathbb{N}$, be in $\mathbb{X}$. Then the following are equivalent:*

1. $\mathcal{X}_n \underset{n\to\infty}{\longrightarrow} \mathcal{X}$, *Gromov-vaguely.*
2. *For each $n \in \mathbb{N}$ there is $A_n \subseteq X_n$ such that $\mu_n(A_n \cap B_{r_n}(\rho_n, R)) \underset{n\to\infty}{\longrightarrow} 0$ for all $R > 0$, and*

$$\big(X_n \setminus A_n, r_n, \rho_n, \mu_n{\restriction}_{X_n\setminus A_n}\big) \underset{n\to\infty}{\longrightarrow} \mathcal{X}, \tag{6.8}$$

   Gromov-Hausdorff-vaguely.
3. *For each $n \in \mathbb{N}$ there is $A_n \subseteq X_n$ such that $\mu_n(A_n \cap B_{r_n}(\rho_n, R)) \underset{n\to\infty}{\longrightarrow} 0$ for all $R > 0$, and (6.8) holds Gromov-vaguely.*

7. Application to trees coded by excursions

In this section we consider encodings of trees by means of excursions. To be in a position to consider locally compact rather than just compact $\mathbb{R}$-trees we consider possibly transient excursions, and conclude from uniform convergence on compacta of a sequence of excursions that the corresponding rooted boundedly finite $\mathbb{R}$-trees converge Gromov-Hausdorff-vaguely (Proposition 7.5). As an example we present a representation of the scaling limit of a size-biased Galton-Watson tree (Example 7.6).

Definition 7.1 ((transient) excursions)**.** *A continuous function $e \colon \mathbb{R}_+ \to \mathbb{R}_+$ is called* (continuous) excursion *if $e(0) = 0$ and e is not identically 0. We refer to $\zeta(e) := \sup\{\, s > 0 : e(s) > 0 \,\}$ as the* excursion length*, and to $I_e := [0, \zeta(e))$ as the* excursion interval*. If the excursion length is finite, we call the excursion* compactly supported*. If $\lim_{s\to\infty} e(s) = \infty$, the excursion is called* transient*.*

Let

$$\mathcal{E} := \big\{ e \colon \mathbb{R}_+ \to \mathbb{R}_+ \;\big|\; e \text{ is a continuous excursion} \big\}. \tag{7.1}$$

Given $e \in \mathcal{E}$, we define the pseudo-metric r'_e by letting for all $0 \le s \le t < \zeta(e)$,

$$r'_e(s,t) := e(s) + e(t) - 2\inf_{u\in[s,t]} e(u). \tag{7.2}$$

We write $s \sim_e t$ if $r'_e(s,t) = 0$. Obviously, $s \sim_e t$ is an equivalence relation.

Definition 7.2 (glue map)**.** *The* glue map $g\colon \mathcal{E} \to \mathbb{X}$ *sends an excursion to the complete, separable, rooted measure* $\mathbb{R}$*-tree*

$$g(e) := (T_e, r_e, \rho_e, \mu_e), \tag{7.3}$$

where $T_e := I_e/{\sim_e}$*, and* r_e*,* μ_e*,* ρ_e *are the push-forwards of* r'_e*, the Lebesgue measure* λ_{I_e}*, and* 0*, respectively, under the canonical projection* $\pi_e\colon I_e \to T_e$*.*

Lemma 7.3 (excursions and associated $\mathbb{R}$-trees)**.** *Let* $e \in \mathcal{E}$*.*

1. *If* e *is compactly supported, then* $g(e)$ *is a pointed compact finite measure* $\mathbb{R}$*-tree, i.e.* $g(e) \in \mathbb{X}_{\mathrm{c}}$*.*
2. *If* e *is transient, then* $g(e)$ *is a pointed Heine-Borel boundedly finite measure* $\mathbb{R}$*-tree, i.e.* $g(e) \in \mathbb{X}_{\mathrm{HB}}$*.*
3. *If* e *is neither compactly supported nor transient, then* $g(e) \notin \mathbb{X}_{\mathrm{HB}}$*.*

Proof. 1. Follows from Lemma 3.1 in [EW06].

2. Assume that e is transient. Then for all $R > 0$, $\xi_e(R) := \sup\{s \ge 0 : e(s) < R\} < \infty$, and $A_R := \{ s \in [0,\infty) : e(s) \le R \}$ is a closed subset of $[0, \xi_e(R)]$, and hence compact. Note that continuity of e implies continuity of the projection π_e. Therefore, $\overline{B}(\rho, R) = \pi_e(A_R)$ is also compact. Moreover, $\mu_e\big(\overline{B}(\rho,R)\big) \le \xi(R) < \infty$. As any closed and bounded subset of (T_e, r_e) is a closed subset of a closed ball $\overline{B}(\rho, R)$ for some $R > 0$, it is compact as well. Thus $g(e) \in \mathbb{X}_{\mathrm{HB}}$.
3. Assume that e is such that $\zeta(e) = \infty$ but $a := \liminf_{t\to\infty} e(t) < \infty$, and define $b := \limsup_{t\to\infty} e(t)$. In the case $b > a$, (T_e, r_e) is not Heine-Borel (and therefore not locally compact). Indeed, there is an $\varepsilon > 0$ with $a + 3\varepsilon < b$, and an increasing sequence (t_n) in $\mathbb{R}_+$ with $e(t_n) \in [a + 2\varepsilon, a + 3\varepsilon]$ and $\inf_{u\in[t_n,t_{n+1}]} e(u) \le a + \varepsilon$ for all $n \in \mathbb{N}$. This means that $x_n := \pi_e(t_n)$ defines a sequence of points in $\overline{B}(\rho, a + 3\varepsilon)$ with mutual distances at least ε. In the case $b = a$,

$$\mu_e\big(\overline{B}(\rho, b+1)\big) = \lambda\big\{ s \in \mathbb{R}_+ : e(s) \le b+1 \big\} = \infty, \tag{7.4}$$

which means that μ_e is not boundedly finite. In both cases $g(e) \notin \mathbb{X}_{\mathrm{HB}}$. □

Denote the space of continuous, transient excursions on $\mathbb{R}_+$ by

$$\mathcal{E}_{\mathrm{trans}} := \big\{e\colon \mathbb{R}_+ \to \mathbb{R}_+ \;\big|\; e \text{ is continuous, } e(0) = 0,\ \lim_{x\to\infty} e(x) = \infty\big\}, \tag{7.5}$$

and let for $e \in \mathcal{E}_{\mathrm{trans}}$ and $R > 0$,

$$\xi_e(R) := \sup\{s \ge 0 : e(s) < R\} < \infty \tag{7.6}$$

denote the last visit to height $R > 0$.

Remark 7.4 ($\mathbb{R}$-trees under transient excursions)**.** *Let* $e \in \mathcal{E}_{\mathrm{trans}}$*. Then* $g(e)$ *is a Heine-Borel boundedly finite measure* $\mathbb{R}$*-tree with precisely one end at infinity, i.e., there is a unique isometry* $\varphi\colon [0,\infty) \to T_e$ *with* $\varphi(0) = \rho_e$*. Indeed, the map* $\varphi_e := \pi_e \circ \xi_e$ *is such an isometry. Assume that* ψ *is a further such isometry and fix* $R > 0$*. We show that* $\psi(R) = \varphi_e(R)$*. Choose* $t \in \mathbb{R}_+$ *with* $\pi_e(t) = \psi(R)$*. Because* ψ *is an isometry, we have* $e(t) = R$*, and consequently* $t \le \xi_e(R)$*. Choose* $S > \sup_{u\in[0,\xi_e(R)]} e(u)$ *and* $s \in \pi_e^{-1}\big(\psi(S)\big)$*. Note that* $s > \xi_e(R)$ *and* $e(s) = S$*. Therefore* $S - R = r_e\big(\psi(S), \psi(R)\big) = S + R - 2\inf_{u\in[t,s]} e(u)$*, and hence* $\inf_{u\in[t,\xi_e(R)]} e(u) \ge \inf_{u\in[t,s]} e(u) = R$*. This implies* $r_e\big(\psi(R), \varphi_e(R)\big) = 2R - \inf_{u\in[t,\xi_e(R)]} e(u) = 0$*.* □

Proposition 7.5 (continuity of glue map). *The glue map $g\colon \mathcal{E}_{\mathrm{trans}} \to \mathbb{X}_{\mathrm{HB}}$ is continuous if $\mathcal{E}_{\mathrm{trans}}$ is equipped with the topology of uniform convergence on compacta, and $\mathbb{X}_{\mathrm{HB}}$ with the Gromov-Hausdorff-vague topology.*

Proof. Let $(e_n)_{n\in\mathbb{N}}$ and e in $\mathcal{E}_{\mathrm{trans}}$ be such that $e_n \underset{n\to\infty}{\longrightarrow} e$ uniformly on compacta. Put $\mathbb{W}_+ := \{R \geq 0 : \lambda\{s \geq 0 : e(s) = R\} > 0\}$. Standard arguments show that $\mathbb{W}_+$ is at most countable. Recall $\xi_e(R)$ from (7.6) and note that for all $R > 0$, the map $e \mapsto \xi_e(R)$ is continuous with respect to the uniform topology on compacta. Thus for all $R \in [0,\infty) \setminus \mathbb{W}_+$,

$$e_n\restriction_{[0,\xi_{e_n}(R)]} \underset{n\to\infty}{\longrightarrow} e\restriction_{[0,\xi_e(R)]}, \tag{7.7}$$

which in turn implies that $g(e_n)\restriction_R \to g(e)\restriction_R$ Gromov-Hausdorff-weakly (see, for example, [ADH14, Proposition 2.9]). Therefore $g(e_n) \underset{n\to\infty}{\longrightarrow} g(e)$ Gromov-Hausdorff-vaguely by Definition 5.8. □

We illustrate the usefulness of Proposition 7.5 with an example about the scaling limit of a size-biased branching tree (compare [Kal77, GW91] for a probabilitistic representation of this tree).

Example 7.6 (Kallenberg-Kesten tree). *Consider a (discrete time) Galton-Watson tree with a finite variance, mean 1 offspring distribution $p = (p_n)_{n\in\mathbb{N}_0}$. Let T' be the so-called* Kallenberg-Kesten tree, *which is a random graph theoretic tree that is distributed like this Galton-Watson tree conditioned on survival. The simple, nearest neighbor random walk on T', and scaling limits thereof, are of interest because of the "subdiffusive" behavior (see* [Kes86, BK06]*). The random walk is associated to the degree measure, defined in Example 5.15, as "speed measure" (see* [ALW17, Section 7.4]*). As in Example 5.15, we construct the (equivalent) rooted, measured $\mathbb{R}$-tree $(T, r, \rho, \mu_T^{\mathrm{deg}})$, corresponding to T'. In the particular case of a geometric offspring distribution, i.e., $p_n := 2^{-(1+n)}$ for all $n \in \mathbb{N}_0$, we can code the tree with the length measure instead of the degree measure as follows:*

$$(T, r, \rho, \lambda_{(T,r)}) \overset{\mathcal{L}}{=} g(\tilde{W}), \tag{7.8}$$

where $\overset{\mathcal{L}}{=}$ denotes equivalence in law and, for all $t \geq 0$, $\tilde{W}_t := W_t - 2\inf_{s\in[0,t]} W_s$, with a simple random walk path $(W_n)_{n\in\mathbb{N}}$ linearly interpolated. We refer to $\mathcal{T}_{\mathrm{K}}^{\mathrm{geom}(1/2)} := (T, r, \rho, \mu_T^{\mathrm{deg}})$ as the discrete Kallenberg tree *with geometric offspring distribution.*

As W converges, after Brownian rescaling, weakly in path space towards standard Brownian motion B, we have

$$\big(n^{-1}\tilde{W}_{n^2 t}\big)_{t\geq 0} \underset{n\to\infty}{\Longrightarrow} \big(\tilde{B}_t\big)_{t\geq 0}, \tag{7.9}$$

where $\tilde{B}_t := B_t - 2\inf_{s\in[0,t]} B_s$. It is shown in [Pit75] *that $(\tilde{B}_t)_{t\geq 0}$ equals in law the unique strong solution of the stochastic differential equation*

$$X_t := \tfrac{1}{X_t}\mathrm{d}t + \mathrm{d}B_t,\ t > 0, \qquad X_0 = 0. \tag{7.10}$$

Note that this solution is a three dimensional Bessel process (i.e. the radial path of a three dimensional Brownian motion). We refer to $g(X)$ as the continuum Kallenberg-Kesten tree, *$\mathcal{T}_{\mathrm{K}}$.*

Because, almost surely, a realization $e := n^{-1}\tilde{W}_{n^2\cdot}$ has slope $\pm n$ almost everywhere, we have $\mu_e = n^{-1}\lambda_{(T_e,r_e)}$. Hence, by Proposition 7.5, (7.9) implies

$$(T, n^{-1}r, \rho, n^{-2}\lambda_{(T,r)}) \overset{\mathcal{L}}{=} g\big(n^{-1}\tilde{W}_{n^2\cdot}\big) \underset{n\to\infty}{\Longrightarrow} g(X) = \mathcal{T}_{\mathrm{K}}, \tag{7.11}$$

Gromov-Hausdorff-vaguely. By Example 5.15, this also implies

$$(T, n^{-1}r, \rho, n^{-2}\mu_T^{\mathrm{deg}}) \underset{n\to\infty}{\Longrightarrow} \mathcal{T}_{\mathrm{K}}, \tag{7.12}$$

Gromov-Hausdorff-vaguely. In words, if we consider the discrete Kallenberg-Kesten tree and rescale the edge length to become n^{-1}, and then equip it with the measure which assigns mass

$\frac{1}{2}n^{-2}\deg(x)$ *to each branch point* x*, then this discrete measure tree converges weakly with respect to the Gromov-Hausdorff-vague topology to the continuum Kallenberg-Kesten tree. This implies that the simple, nearest neighbor random walk on the rescaled discrete Kallenberg-Kesten tree converges, if sped up by a factor of* n^3*, to the Brownian motion on* $\mathcal{T}_{\mathrm{K}}$*, according to Theorem 1 of* [ALW17]*. See also Section 7.4 there.* □

References

[ABBGM13] Lougi Addario-Berry, Nicolas Broutin, Christina Goldschmidt, and Grégory Miermont, *The scaling limit of the minimum spanning tree of the complete graph*, arXiv:1301.1664, 2013.

[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** (2013), no. 14, 1–21.

[ADH14] ———, *Exit times for an increasing Lévy tree-valued process*, Probab. Theory Related Fields **159** (2014), no. 1-2, 357–403. MR 3201925

[Ald93] David Aldous, *The continuum random tree III*, Ann. Probab. **21** (1993), 248–289.

[ALW17] Siva Athreya, Wolfgang Löhr, and Anita Winter, *Invariance principle for variable speed random walks on trees*, Ann. Probab. **45** (2017), no. 2, 625–667.

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov, *A course in metric geometry*, Graduate studies in mathematics, vol. 33, AMS, Boston, MA, 2001.

[BGMP14] Andrew J. Blumberg, Itamar Gal, Michael A. Mandell, and Matthew Pancia, *Robust statistics, hypothesis testing, and confidence intervals for persistent homology on metric measure spaces*, Found. Comput. Math. **14** (2014), no. 4, 745–789. MR 3230014

[BK06] Martin T. Barlow and Takashi Kumagai, *Random walk on the incipient infinite cluster on trees*, Illinois J. Math. **50** (2006), no. 1, 33–65.

[Car14] Gunnar Carlson, *Topological pattern recognition for point cloud data*, Acta Numerics **23** (2014), 289–368.

[CH13] Nicolas Curien and Bénédicte Haas, *The stable trees are nested*, Probab. Theory Related Fields **157** (2013), 847–883.

[Coh80] Donald L. Cohn, *Measure theory*, Birkhäuser, 1980.

[Cro08] David A. Croydon, *Volume growth and heat kernel estimates for the continuum random tree*, Probab. Theory Related Fields **140** (2008), 207–238.

[DLG02] Thomas Duquesne and Jean-François Le Gall, *Random trees, Lévy processes and spatial branching processes*, Astérisque **281** (2002), vi+147. MR MR1954248 (2003m:60239)

[Duq03] Thomas Duquesne, *A limit theorem for the contour process of conditioned Galton-Watson trees*, Ann. Probab. **31** (2003), 996–1027.

[DW14] Thomas Duquesne and Guanying Wang, *Exceptionally small balls in stable trees*, Bull. Soc. Math. France **142** (2014), no. 2, 223–254. MR 3269345

[EW06] Steven N. Evans and Anita Winter, *Subtree prune and re-graft: A reversible real-tree valued Markov chain*, Ann. Probab. **34** (2006), no. 3, 918–961.

[FOT11] Masatoshi Fukushima, Yoichi Oshima, and Masayoshi Takeda, *Dirichlet forms and symmetric Markov processes*, second ed., de Gruyter Studies in Mathematics, vol. 19, Walter de Gruyter & Co., Berlin, 2011. MR 2778606 (2011k:60249)

[Fuk87] Kenji Fukaya, *Collapsing of Riemannian manifolds and eigenvalues of Laplace operators*, Invent. Math. **87** (1987), 517–547.

[GPW09a] Andreas Greven, Peter Pfaffelhuber, and Anita Winter, *Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees)*, Probab. Theory Related Fields **145** (2009), no. 1-2, 285–322.

[GPW09b] Andreas Greven, Lea Popovic, and Anita Winter, *Genealogy of catalytic branching models*, Annals of Applied Probability **19** (2009), no. 3, 1232–1272.

[GPW13] Andreas Greven, Peter Pfaffelhuber, and Anita Winter, *Tree-valued resampling dynamics: Martingale Problems and applications*, Probab. Theory Related Fields **155** (2013), no. 3–4, 789–838.

[Gro99] Misha Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, vol. 152, Birkhäuser Boston Inc., Boston, MA, 1999, Based on the 1981 French original.

[GW91] Luis G. Gorostiza and Anton Wakolbinger, *Persistence criteria for a class of critical branching particle systems in continuous time*, Ann. Probab. **19** (1991), no. 1, 266–288.

[HM12] Bénédicte Haas and Grégory Miermont, *Scaling limits of Markov branching trees with applications to Galton-Watson and random unordered trees*, Ann. Probab. **40** (2012), no. 6, 2589–2666. MR 3050512

[HW14] Hui He and Matthias Winkel, *Invariance principles for pruning processes of Galton-Watson trees*, arXiv:1409.1014, 2014.

[Kal77] Olav Kallenberg, *Stability of critical cluster fields*, Math. Nachr. **77** (1977), 7–43.

[Kes86] Harry Kesten, *Subdiffusive behavior of random walk on a random cluster*, Ann. Inst. H. Poincaré Probab. Statist. **22** (1986), no. 4, 425–487.

[KS03] Kazhiro Kuwae and Takashi Shioya, *Convergence of spectral structure: a functional analytic theory and its application to spectral geometry*, Analysis and Geometry **11** (2003), no. 4, 599–673.

[LC57] Lucien Le Cam, *Convergence in distribution of stochastic processes*, Univ. California Publ. Statist. **2** (1957), 207–236.

[LG07] Jean-François Le Gall, *The topological structure of scaling limits of large planar maps*, Invent. Math. **169** (2007), no. 3, 621–670. MR 2336042 (2008i:60022)

[Löh13] Wolfgang Löhr, *Equivalence of Gromov-Prohorov- and Gromov's $\underline{\square}_\lambda$-metric on the space of metric measure spaces*, Electron. Commun. Probab. **18** (2013), no. 17, 1–10. MR 3037215

[LVW15] Wolfgang Löhr, Guillaume Voisin, and Anita Winter, *Convergence of bi-measure $\mathbb{R}$-trees and the pruning process*, Ann. Inst. H. Poincaré Probab. Statist. **51** (2015), no. 4, 1342–1368. MR MR3414450

[Mie09] Grégory Miermont, *Tessellations of random maps of arbitrary genus*, Ann. Sci. Éc. Norm. Supér. (4) **42** (2009), no. 5, 725–781.

[MM11] Jean-François Marckert and Grégory Miermont, *The CRT is the scaling limit of unordered binary trees*, Random Structures Algorithms **38** (2011), no. 4, 467–501.

[Pit75] James W. Pitman, *One-dimensional Brownian motion and the three-dimensional Bessel process*, Adv. in Appl. Probab. **7** (1975), 511–526. MR MR0375485

[Vil08] Cédric Villani, *Optimal transport, old and new*, École d'Été de Probabilités de Saint Flour XXXV-2005, Lecture Notes in Mathematics, vol. 1920, Springer-Verlag, 2008.

INVARIANCE PRINCIPLE FOR VARIABLE SPEED RANDOM WALKS ON TREES

SIVA ATHREYA, WOLFGANG LÖHR, AND ANITA WINTER

Abstract. We consider stochastic processes on complete, locally compact tree-like metric spaces (T,r) on their "natural scale" with boundedly finite speed measure ν. Given a triple (T,r,ν) such a speed-ν motion on (T,r) can be characterized as the unique strong Markov process which if restricted to compact subtrees satisfies for all $x,y\in T$ and all positive, bounded measurable f,

$$\mathbb{E}^x\Big[\int_0^{\tau_y}\mathrm{d}s\, f(X_s)\Big] = 2\int_T \nu(\mathrm{d}z)\, r\big(y, c(x,y,z)\big) f(z) < \infty, \tag{0.1}$$

where $c(x,y,z)$ denotes the branch point generated by x,y,z. If (T,r) is a discrete tree, X is a continuous time nearest neighbor random walk which jumps from v to $v'\sim v$ at rate $\frac{1}{2}\cdot\big(\nu(\{v\})\cdot r(v,v')\big)^{-1}$. If (T,r) is path-connected, X has continuous paths and equals the ν-Brownian motion which was recently constructed in [AEW13]. In this paper we show that speed-ν_n motions on (T_n,r_n) converge weakly in path space to the speed-ν motion on (T,r) provided that the underlying triples of metric measure spaces converge in the Gromov-Hausdorff-vague topology introduced in [ALW16].

Contents

1. Introduction and main result (Theorem 1)

Fifty years ago in [Sto63] Markov processes were considered which have in common that their state spaces are closed subsets of the real line and that their random trajectories *"do not jump over points"*. When put in their *"natural scale"* these processes are determined by their *"speed measure"*.

Stone argues that in some sense the processes depend continuously on the speed measures. The most classical example is the symmetric simple random walk on $\mathbb{Z}$ which, after a suitable rescaling, converges to standard Brownian motion. If you rescale edge lengths by a factor $\frac{1}{\sqrt{n}}$ and speed up time by a factor n, then you might think of the rescaled random walk as such a process with speed measure $\frac{1}{\sqrt{n}}q(\sqrt{n}\,\boldsymbol{\cdot})$, where q denotes the counting measure on $\mathbb{Z}$, and of the standard Brownian motion as such a process whose speed measure equals the Lebesgue measure on $\mathbb{R}$.

In the present paper we want to extend this result from $\mathbb{R}$-valued Markov processes to Markov processes which take values in tree-like metric spaces. Before we state our main result precisely, we do the preliminary work and define the space of rooted metric boundedly finite measure trees equipped with pointed Gromov-vague topology and give our notion of convergence in path space.

Definition 1.1 (Rooted metric boundedly finite measure trees)**.**

(i) *A* pointed Heine-Borel space (X, r, ρ) *consists of a Heine-Borel space*[1] (X, r) *and a distinguished point* $\rho \in X$.

(ii) *A* rooted metric tree *is a pointed Heine-Borel space* (T, r, ρ)*, which is both* 0-hyperbolic*, or equivalently, satisfies the* four point condition*, i.e.,*

$$\begin{aligned} &r(x_1, x_2) + r(x_3, x_4) \\ &\le \max\bigl\{r(x_1, x_3) + r(x_2, x_4),\, r(x_1, x_4) + r(x_2, x_3)\bigr\}, \end{aligned} \tag{1.1}$$

holds for all $x_1, x_2, x_3, x_4 \in T$*, and* fine*, i.e., for all* $x_1, x_2, x_3 \in T$ *there is a (necessarily unique) point* $c(x_1, x_2, x_3) \in T$*, such that for* $i, j \in \{1, 2, 3\}$*,* $i \ne j$*,*

$$r\bigl(x_i, c(x_1, x_2, x_3)\bigr) + r\bigl(x_j, c(x_1, x_2, x_3)\bigr) = r(x_i, x_j). \tag{1.2}$$

The point $c(x_1, x_2, x_3)$ *is referred to as* branch point*, and the distinguished point* $\rho \in T$ *as the* root.

(iii) *In a rooted metric tree* (T, r, ρ) *we define for* $a, b \in T$ *the intervals*

$$[a, b] := \bigl\{x \in T \,:\, r(a, x) + r(x, b) = r(a, b)\bigr\}, \tag{1.3}$$

$(a, b) := [a, b] \setminus \{a, b\}$, $[a, b) := [a, b] \setminus \{b\}$ *and* $(a, b] := [a, b] \setminus \{a\}$*. We say that* $x, y \in T$ *are connected by an* edge*, in symbols* $x \sim_T y$ *or simply* $x \sim y$*, iff*

$$x \ne y \qquad \textit{and} \qquad [x, y] = \{x, y\}. \tag{1.4}$$

If $x \sim y$ *and* $x \in [\rho, y]$*, we call the pair* (x, y) *an* oriented edge *of length* $r(x, y)$.

(iv) *A* rooted metric boundedly finite measure tree (T, r, ρ, ν) *consists of a rooted metric tree* (T, r, ρ) *and a measure* ν *on* $(T, \mathcal{B}(T))$ *which is finite on bounded sets and has full support,* $\mathrm{supp}(\nu) = T$.

Remark 1.2 ($\mathbb{R}$-trees versus trees with edges)**.** *A metric tree is connected (i.e. is an $\mathbb{R}$-tree) if and only if it has no edges. Due to separability, there can be only countably many edges.* □

We will establish a one-to-one correspondence between rooted metric boundedly finite measure trees (T, r, ρ, ν) and strong Markov processes $X = (X_t)_{t \ge 0}$ with values in (T, r) starting at ρ. When (T, r) is compact such a process can be characterized by the occupation time formula given in (0.1) (see Proposition 5.1). For general rooted metric boundedly finite measure trees the corresponding Markov process is associated with a regular Dirichlet form (see Definition 2.7). We will refer to this Markov process as *speed-ν motion on* (T, r) or *variable speed motion associated to* ν *on* (T, r). If (T, r) is path-connected, then X has continuous paths and equals the so-called ν-Brownian motion on (T, r), which was recently constructed in [AEW13]. On the other hand, if (T, r) is discrete, X is a continuous time nearest neighbor Markov chain which jumps from v to $v' \sim v$ at rate

$$\gamma_{vv'} := \tfrac{1}{2} \cdot \bigl(\nu(\{v\}) \cdot r(v, v')\bigr)^{-1} \tag{1.5}$$

[1] Recall that a *Heine-Borel space* is a metric space in which every bounded closed subset is compact. Note that every Heine-Borel space is complete, separable and locally compact.

(see Lemma 2.11).

The invariance principle which we are going to state says that a sequence of variable speed motions converges in path space to a limiting variable speed motion whenever the underlying metric measure trees converge in the pointed Gromov-Hausdorff-vague topology which was recently introduced in [ALW16]. In particular, it was shown that convergence in pointed Gromov-Hausdorff-vague topology is equivalent to convergence in pointed Gromov-vague topology together with the *uniform local lower mass-bound property*, i.e., for each $\delta, R > 0$,

$$\liminf_{n\to\infty} \inf_{x\in B_n(\rho_n,R)} \nu_n\big(B_n(x,\delta)\big) > 0 \tag{1.6}$$

(see Proposition 3.8). Here, $B_n(x,R) = \big\{y \in T_n : r_n(x,y) < R\big\}$ is the ball around x with radius R in the metric space (T_n, r_n). In the introduction we recall only the definition of Gromov-vague topology. For a more elaborate discussion of the topology, we refer the reader to Section 3.

We call two rooted metric measure trees (T, r, ρ, ν) and (T', r', ρ', ν') equivalent iff there is an isometry φ between (T, r) and (T', r') such that $\varphi(\rho) = \rho'$ and $\nu \circ \varphi^{-1} = \nu'$. Denote

$$\mathbb{T} := \big\{\text{equivalence classes of rooted metric boundedly finite measure trees}\big\}. \tag{1.7}$$

Let $\mathcal{X} := (T, r, \rho, \nu)$, $\mathcal{X}_1 := (T_1, r_1, \rho_1, \nu)$, $\mathcal{X}_2 := (T_2, r_2, \rho_2, \nu), \ldots$ be in $\mathbb{T}$. We say that $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}$ in *pointed Gromov-vague topology* iff there are a pointed metric space (E, d_E, ρ_E) and isometries $\varphi_n \colon T_n \to E$ with $\varphi_n(\rho_n) = \rho_E$, for all $n \in \mathbb{N}$, as well as an isometry $\varphi \colon T \to E$ with $\varphi(\rho) = \rho_E$ such that the sequence of image measures $(\varphi_{n*}\nu_n)\restriction_{B(\rho_E,R)}$ restricted to the ball of radius R around the root converges weakly for all but countably many $R > 0$.

Before we are in a position to state our main scaling result, notice that the approximating Markov processes may live on different spaces. We therefore agree on the following:

Definition 1.3 (A notion of convergence in path space)**.** *For every $n \in \mathbb{N} \cup \{\infty\}$, let X^n be a càdlàg process with values in a metric space (T_n, r_n).*

(i) *We say that $(X^n)_{n\in\mathbb{N}}$ converges to X^∞ weakly in path space (resp. f.d.d.) if there exists a metric space (E, d_E) and isometric embeddings $\phi_n \colon T_n \to E$, $n \in \mathbb{N} \cup \{\infty\}$, such that $(\phi_n \circ X^n)_{n\in\mathbb{N}}$ converges to $\phi_\infty \circ X^\infty$ weakly in Skorohod path space (resp. f.d.d.).*

(ii) *We say that $(X^n)_{n\in\mathbb{N}}$ converges to X^∞ in the one-point compactification weakly in path space (resp. f.d.d.) if there exists a locally compact space (E, d_E) and embeddings as in (i) such that we have weak path space (resp. f.d.d.) convergence in the one-point compactification $E \cup \{\infty\}$, where the processes are defined to take the value ∞ after their lifetimes.*

To be in a position to state our invariance principle, we recall the notion of the one-point compactification $\widehat{E} := E \cup \{\infty\}$ of a separable, locally compact (but non-compact) metric space E, and the life time ζ of a E-valued strong Markov process, i.e.,

$$\zeta := \inf\big\{t \ge 0 : \ X_t = \infty\big\}. \tag{1.8}$$

Our main result is the following:

Theorem 1 (Invariance principle)**.** *Let $\mathcal{X} := (T, r, \rho, \nu)$, $\mathcal{X}_n := (T_n, r_n, \rho_n, \nu_n) \in \mathbb{T}$, $n \in \mathbb{N}$. Let X be the speed-ν motion on (T, r) starting in ρ, and for all $n \in \mathbb{N}$, let X^n be the speed-ν_n motion on (T_n, r_n) started in ρ_n. Assume that the following conditions hold:*

(A0) *For all $R > 0$,*

$$\limsup_{n\to\infty} \sup\big\{r_n(x,z) \,:\, x \in B_n(\rho_n, R),\, z \in T_n,\, x \sim z\big\} < \infty. \tag{1.9}$$

(A1) *The sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}$ pointed Gromov-vaguely.*

(A2) *The uniform local lower mass-bound property (1.6) holds.*

Then the following hold:

(i) *X^n converges in the one-point compactification weakly in path space to a process Y, such that Y stopped at infinity has the same distribution as the speed-ν motion X. In particular, if X is conservative (i.e. does not hit infinity), then X^n converges weakly in path space to X.*

(ii) *If $\sup_{n\in\mathbb{N}} \mathrm{diam}(T_n, r_n) < \infty$, where* diam *is the diameter, and we assume (A1) but not (A2), then X^n converges f.d.d. to X.*

Remark 1.4 (Entrance law)**.** *Let $\mathcal{X} := (T, r, \rho, \nu)$, $\mathcal{X}_1 := (T_1, r_1, \rho_1, \nu_1)$, $\mathcal{X}_2 := (T_2, r_2, \rho_2, \nu_2), \ldots$ in $\mathbb{T}$ be such that $\mathcal{X}_n \xrightarrow[n\to\infty]{} \mathcal{X}$ Gromov-Hausdorff-vaguely. The statement of Theorem 1(i) reflects the fact that it is possible that the approximating speed-ν_n motions on (T_n, r_n), as well as their limit processes on the one-point compactification, are recurrent but the speed-ν motion on (T, r) is not. Note that in such a situation we obtain an entrance law and that the limit processes cannot be a strong Markov processes. We explain this in detail in Example 5.5.* □

We want to briefly illustrate this invariance principle with a first non-trivial example which was established in [Cro08]. Further examples and the relation of Theorem 1 to the existing literature are discussed in Section 7.

Example 1.5 (RWs on GW-trees converge to BM on the CRT)**.** *Consider a Galton-Watson process in discrete time whose offspring distribution is critical and has finite (positive) variance σ^2. For each $n \in \mathbb{N}$, let $\mathcal{T}_n$ be the corresponding GW-tree conditioned on having n vertices. Given $\mathcal{T}_n$, whenever $v' \sim_{\mathcal{T}_n} v$, put $r_n(v, v') := \frac{\sigma}{\sqrt{n}}$, and let $\nu_n(\{v\}) := \frac{\deg(v)}{2n}$ for all $v \in \mathcal{T}_n$, where* deg *denotes the degree of node. Notice that given $\mathcal{T}_n$, the speed-ν_n random walk on $(\mathcal{T}_n, r_n)$ is the symmetric nearest neighbor random walk on $\mathcal{T}_n$ with edge lengths rescaled by a factor $\frac{\sigma}{\sqrt{n}}$ and with exponential jump rates*

$$\gamma_n(v) = \tfrac{1}{2\nu_n(\{v\})} \sum_{v' \sim v} r_n^{-1}(v, v') = \tfrac{1}{2} \cdot \tfrac{2n}{\deg(v)} \cdot \deg(v) \tfrac{\sqrt{n}}{\sigma} = \sigma^{-1} \cdot n^{\frac{3}{2}}. \tag{1.10}$$

Denote by μ_n^{ske} the normalized length-measure (see Section 2.1) on the path-connected tree $\overline{\mathcal{T}}_n$ spanned by $\mathcal{T}_n$. Then it is known that $(\overline{\mathcal{T}}_n, r_n, \mu_n^{\mathrm{ske}})$ converges Gromov-vaguely in distribution to some random, compact, path-connected metric measure tree $(\mathcal{T}, r, \mu)$, where $(\mathcal{T}, r)$ is the so-called Brownian continuum random tree (or shortly, the CRT), and μ the "leaf-measure" (see, for example, [Ald93, Theorem 23]*). As the Prohorov distance between ν_n and μ_n^{ske} is not greater than $\frac{\sigma}{2\sqrt{n}}$, $(\mathcal{T}_n, r_n, \nu_n)$ also converges Gromov-vaguely to $(\mathcal{T}, r, \mu)$ by* [ALW16, Lemma 2.10]*. Furthermore it is known that the family $\{\nu_n;\, n \in \mathbb{N}\}$ satisfies the uniform local lower mass-bound property (compare* [Ald93, Corollary 19] *together with Proposition 3.8).*

We can therefore conclude from Theorem 1 that given a realization of a sequence $(\mathcal{T}_n)_{n\in\mathbb{N}}$ converging Gromov-weakly to some $\mathcal{T}$, the symmetric random walk with jumps rescaled by $\frac{1}{\sqrt{n}}$ and time speeded up by a factor of $n^{\frac{3}{2}}$ converges to μ-Brownian motion on the CRT. This was first conjectured in [Ald91, Section 5.1] *and proved in* [Cro08]*. A more general result on homogeneous scaling limits of random walks on graph trees towards diffusions on continuum trees was established in* [Cro10]*. We will discuss in Section 7.3 how this result is covered by our invariance principle.* □

For the proof of the invariance principle we use the following approach. We first use techniques from Dirichlet forms to construct the speed-ν motion on (T, r). We continue showing tightness based on a version of Aldous' stopping time criterion (Proposition 4.2), and then identify the limit. As we are working with Dirichlet forms, one might be tempted to show f.d.d.-convergence of the motions by verifying the Mosco-convergence introduced in [Mos69] (compare also [Mos94] for its application to Dirichlet forms). It turns out, however, that this is tedious, and we rather identify the limit via the occupation time formula (0.1). For that, we first restrict ourselves to limit metric (finite) measure trees which are compact, and show that any limit point must be a strong Markov process satisfying (0.1). We then reduce the general case to the case of compact limit trees by showing that there are suitably many hitting times which converge.

The rest of the paper is organized as follows: In Section 2 we construct the speed-ν motion on (T,r) and present occupation time formula (0.1). In Section 3 we introduce all the topological concepts needed to deal with convergence of the underlying metric measure spaces. In Section 4 we prove the tightness of a sequence of speed-ν_n motions on (T_n,r_n) provided that the underlying spaces $(T_n,r_n,\nu_n)_{n\in\mathbb{N}}$ converges. In Section 5 we show that any limit point satisfies the strong Markov property and that its occupation time formula agrees with that of the limit variable speed motion. In Section 6 we collect all the ingredients to present the proof of Theorem 1. Finally, in Section 7 we present examples and relate our result to the existing literature.

2. The speed-ν motion on (T,r) and its Dirichlet form

In this section we will use Dirichlet form techniques to construct the variable speed motions. We will follow the lines of [AEW13] where the variable speed motion was constructed on path-connected rooted metric measure trees, or rooted measure $\mathbb{R}$-*trees* for short. The main idea behind the generalization to arbitrary rooted metric measure trees is the presentation of a universal notion of the length measure and the gradient. This will be given in Subsection 2.1. In Subsection 2.2 we associate the variable speed motion with a Dirichlet form and establish in Subsection 2.3 the occupation time formula. We will revise (where necessary) the proofs given in [AEW13] to the larger class of underlying rooted metric measure trees.

2.1. **The set-up.** In this subsection we discuss preliminaries that are required to construct the variable speed motions.

Recall rooted metric trees and rooted $\mathbb{R}$-trees from Definition 1.1, and notice that a rooted metric tree (T,r,ρ) can be embedded isometrically into an $\mathbb{R}$-*tree*, i.e. a path-connected rooted metric tree (see, for example, Theorem 3.38 in [Eva08]). Furthermore, there is a unique (up to isometry) smallest rooted $\mathbb{R}$-tree, $(\bar{T},\bar{r},\rho)$, which contains (T,r,ρ) (compare, e.g., [LVW15, Remark 2.7]). $(\bar{T},\bar{r})$ is the smallest $\mathbb{R}$-tree in the following sense: if $(\hat{T},\hat{r})$ is another $\mathbb{R}$-tree with $T\subseteq\hat{T}$, and $\hat{r}$ extends r, then there is a unique isometric embedding $\phi\colon\bar{T}\to\hat{T}$ such that $\phi{\restriction}_T$ is the identity on T. Heuristically, $(\bar{T},\bar{r})$ is obtained from (T,r) by replacing edges with line segments of the appropriate length.

Given a rooted metric tree (T,r,ρ), we can define a partial order (with respect to ρ), $\le_\rho$, on T by saying that $x\le_\rho y$ for all $x,y\in T$ with $x\in[\rho,y]$.

To be in a position to capture that our variable speed motions are processes on "natural scale" we need the notion of a length measure. For $\mathbb{R}$-trees it was first introduced in [EPW06]. It turns out that this measure can be constructed on any separable 0-hyperbolic metric space provided that we have fixed a reference point, say the root ρ. Let therefore (T,r,ρ) be a rooted metric tree, and $\mathcal{B}(T)$ the Borel-σ-algebra of (T,r). We denote the set of *isolated points* (other than the root) by $\mathrm{Iso}(T,r,\rho)$, and define the *skeleton* of (T,r,ρ) as

$$T^o:=\mathrm{Iso}(T,r,\rho)\cup\bigcup\nolimits_{a\in T}(\rho,a). \tag{2.1}$$

Recall that rooted metric trees are Heine-Borel spaces and thus separable, and observe that if $T'\subset T$ is a dense countable set, then (2.1) holds with T replaced by T'. In particular, $T^o\in\mathcal{B}(T)$ and $\mathcal{B}(T){\restriction}_{T^o}=\sigma(\{(a,b);\,a,b\in T'\})$, where $\mathcal{B}(T){\restriction}_{T^o}:=\{A\cap T^o;\,A\in\mathcal{B}(T)\}$. Hence, there exist a unique σ-finite measure $\lambda^{(T,r,\rho)}$ on T, such that $\lambda^{(T,r,\rho)}(T\setminus T^o)=0$ and for all $a\in T$,

$$\lambda^{(T,r,\rho)}\big((\rho,a]\big)=r(\rho,a). \tag{2.2}$$

Definition 2.1 (Length measure)**.** *Let (T, r, ρ) be a rooted metric tree. The unique σ-finite measure $\lambda^{(T,r,\rho)}$ satisfying (2.2) and $\lambda^{(T,r,\rho)}(T \setminus T^o) = 0$ is called the* length measure *of (T, r, ρ).*

Remark 2.2 (Length measure; particular instances)**.**

(i) *If (T, r) is an $\mathbb{R}$-tree, then $\lambda^{(T,r,\rho)}$ does not depend on the root ρ, and is the trace onto T^o of the 1-dimensional Hausdorff-measure on T.*

(ii) *If (T, r) is discrete as a topological space, i.e. all points in T are isolated, the length measure shifts all the "length" sitting on an edge to the end point which is further away from the root. In this case it does explicitly depend on the root.*

(iii) *In general, let $(\bar{T}, \bar{r})$ be the $\mathbb{R}$-tree spanned by (T, r) and $\pi\colon \bar{T} \to T$ defined by*

$$\pi(x) := \inf \big\{y \in T : x \le_\rho y\big\}, \tag{2.3}$$

for all $x \in \bar{T}$. Note that π is well defined because T is closed and satisfies (1.2)*. It is therefore easy to check that*

$$\lambda^{(T,r,\rho)} = \pi_* \lambda^{(\bar{T},\bar{r})}. \tag{2.4}$$

□

In order to characterize the variable speed motion analytically (via Dirichlet forms), we use a concept of weak differentiability. Denote the space of continuous functions $f\colon T \to \mathbb{R}$ by $\mathcal{C}(T)$. We call a function $f \in \mathcal{C}(T)$ *locally absolutely continuous* if and only if for all $\varepsilon > 0$ and all subsets $S \subseteq T$ with $\lambda^{(T,r,\rho)}(S) < \infty$ there exists a $\delta = \delta(\varepsilon, S)$ such that if $[x_1, y_1], \ldots, [x_n, y_n] \subseteq S$ are disjoint arcs with $\sum_{i=1}^n r(x_i, y_i) < \delta$ then $\sum_{i=1}^n \big|f(x_i) - f(y_i)\big| < \varepsilon$. Put

$$\mathcal{A} = \mathcal{A}^{(T,r)} := \big\{f \in \mathcal{C}(T) : f \text{ is locally absolutely continuous}\big\}. \tag{2.5}$$

Of course, if (T, r) is discrete, then $\mathcal{A}$ equals the space $\mathcal{C}(T)$ of continuous functions.

The definition of the gradient is then based on the following observation which was proved for $\mathbb{R}$-trees in [AEW13, Proposition 1.1].

Proposition 2.3 (Gradient)**.** *Let $f \in \mathcal{A}$. There exists a unique (up to $\lambda = \lambda^{(T,r,\rho)}$-zero sets) function $g \in L^1_{\text{loc}}(\lambda^{(T,r,\rho)})$ such that*

$$f(y) - f(x) = \int_{[\rho,y]} \lambda(\mathrm{d}z)\, g(z) - \int_{[\rho,x]} \lambda(\mathrm{d}z)\, g(z), \tag{2.6}$$

for all $x, y \in T$. Moreover, g is already uniquely determined (up to $\lambda^{(T,r,\rho)}$-zero sets) if we only require (2.6) to hold for all $x \le_\rho y$.

Proof. For $f \in \mathcal{A}$, we define the linear extension $\bar{f}\colon \bar{T} \to \mathbb{R}$ by $\bar{f}\restriction_T := f$ and

$$\bar{f}(v) := \tfrac{r(v,y)}{r(x,y)} f(x) + \tfrac{r(v,x)}{r(x,y)} f(y), \tag{2.7}$$

whenever (x, y) is an edge of T and $v \in [x, y] \subseteq \bar{T}$. By [AEW13, Proposition 1.1], there is $\bar{g}\colon \bar{T} \to \mathbb{R}$ such that (2.6) holds for $x, y \in \bar{T}$ and $\bar{\lambda} := \lambda^{(\bar{T},\bar{r})}$ instead of λ. It is easy to see from the definition of $\bar{f}$ that $\bar{g}$ is constant on edges of T and hence $g\colon T \to \mathbb{R}$ is well defined by $g \circ \pi := \bar{g}$, with π defined in Remark 2.2(iii). By (2.4),

$$f(y) - f(x) = \int_{[\rho,y]} \mathrm{d}\bar{\lambda}\, \bar{g} - \int_{[\rho,x]} \mathrm{d}\bar{\lambda}\, \bar{g} = \int_{[\rho,y]} \mathrm{d}\lambda\, g - \int_{[\rho,x]} \mathrm{d}\lambda\, g. \tag{2.8}$$

Uniqueness and integrability of g follow from the corresponding properties of $\bar{g}$. □

The statement of Proposition 2.3 yields a general notion of a gradient.

Definition 2.4 (Gradient)**.** *The gradient, $\nabla f = \nabla^{(T,r,\rho)} f$, of $f \in \mathcal{A}$ is the unique up to $\lambda^{(T,r,\rho)}$-zero sets function g which satisfies (2.6) for all $x, y \in T$.*

2.2. **The regular Dirichlet form.** In this subsection we recall the construction of the so-called ν-Brownian motion on an $\mathbb{R}$-tree given in [AEW13], and extend it to arbitrary rooted metric measure trees.

As usual, we denote by $\mathcal{C}(T)$ the space of continuous functions $f\colon T \to \mathbb{R}$, and the subspace of functions vanishing at infinity by

$$\mathcal{C}_\infty(T) := \big\{f \in \mathcal{C}(T) : \forall \varepsilon > 0\, \exists K \text{ compact } \forall x \in T \setminus K : |f(x)| \le \varepsilon\big\}. \tag{2.9}$$

Consider the bilinear form $(\mathcal{E}, \mathcal{D}(\mathcal{E}))$ where

$$\mathcal{E}(f,g) := \tfrac{1}{2}\int \mathrm{d}\lambda\, \nabla f \nabla g, \tag{2.10}$$

and

$$\mathcal{D}(\mathcal{E}) := \big\{f \in L^2(\nu) \cap \mathcal{A} \cap \mathcal{C}_\infty(T) : \nabla f \in L^2(\lambda)\big\}. \tag{2.11}$$

For technical purposes we also introduce for all closed subsets $A \subseteq T$ the domain

$$\mathcal{D}_A(\mathcal{E}) := \big\{f \in \mathcal{D}(\mathcal{E}) :\ f|_A \equiv 0\big\}. \tag{2.12}$$

We first note that the bilinear form $(\mathcal{E}, \mathcal{D}_A(\mathcal{E}))$ is closable for all closed sets $A \subseteq T$. Indeed, let $(f_n)_{n\in\mathbb{N}}$ be an $\mathcal{E}$-Cauchy sequence in $\mathcal{D}_A(\mathcal{E}) \subseteq L^2(\nu)$ with $\|f_n\|_{L^2(\nu)} \to 0$. Then, by passing to a subsequence if necessary, we may assume $\nabla f_n \to 0$, $\lambda^{(T,r,\rho)}$-almost surely and $\mathcal{E}(f_n, f_n)$ is uniformly bounded in $n \in \mathbb{N}$ (see for example, [AEW13, (2.15),(2.16)]).

Let $\big(\mathcal{E}, \bar{\mathcal{D}}_A(\mathcal{E})\big)$ be the closure of $\big(\mathcal{E}, \mathcal{D}_A(\mathcal{E})\big)$, i.e., $\bar{\mathcal{D}}_A(\mathcal{E})$ is the closure of $\mathcal{D}_A(\mathcal{E})$ with respect to $\mathcal{E}_1 = \mathcal{E} + \langle \cdot, \cdot \rangle_\nu$.

Remark 2.5 (Closing the form might not be necessary)**.** *The procedure of closing the form is unnecessary if the global lower mass-bound property holds on $T \setminus A$, i.e., for all $\delta > 0$,*

$$\inf_{x \in T\setminus A} \nu\big(B(x,\delta)\big) > 0. \tag{2.13}$$

In this case, $\bar{\mathcal{D}}_A(\mathcal{E}) = \mathcal{D}_A(\mathcal{E})$. □

The following lemma is an immediate consequence of Proposition 2.4, Lemma 2.8, Lemma 3.4, and Proposition 4.1 in [AEW13].

Lemma 2.6 (Regular Dirichlet form)**.** *Let (T, r, ν) be a metric boundedly finite measure tree, and $A \subseteq T$ a closed subset. Then the following hold:*

(i) *The bilinear form $(\mathcal{E}, \bar{\mathcal{D}}_A(\mathcal{E}))$ is a regular Dirichlet form.*

(ii) *Dirac measures are of finite energy integral, there exists a constant $C_x > 0$ such that for all $f \in \mathcal{D}(\mathcal{E}) \cap \mathcal{C}_0(T)$,*

$$f(x)^2 \le C_x\, \mathcal{E}_1(f,f) \tag{2.14}$$

(See (2.2.1) in [FOT11]*).*

(iii) *If A is non-empty the Dirichlet form is transient.*

It follows immediately from [FOT11, Theorem 7.2.1] that there is a unique (up to ν-equivalence) ν-symmetric strong Markov process

$$X = ((X_t)_{t\ge 0}, (\mathbb{P}^x)_{x\in T}) \tag{2.15}$$

on (T, r) associated with the regular Dirichlet form $(\mathcal{E}, \bar{\mathcal{D}}(\mathcal{E}))$.

Definition 2.7 (Speed-ν motion on (T, r))**.** *Let (T, r, ν) be a metric boundedly finite measure tree. In the following we refer to the unique ν-symmetric strong Markov process associated with $(\mathcal{E}, \bar{\mathcal{D}}(\mathcal{E}))$ as the speed-ν motion on (T, r).*

- *If (T, r) is discrete, then the speed-ν motion on (T, r) is referred to as speed-ν random walk on (T, r).*

- *If (T,r) is an $\mathbb{R}$-tree, then the speed-ν motion on (T,r) agrees with the ν-Brownian motion on (T,r) constructed in* [AEW13].

Remark 2.8 (Variable speed motion does not depend on root)**.** *Notice that although the definition of the length measure and the gradient depend on the root, the Dirichlet form does not. Therefore the variable speed motion is independent on the choice of the root.* □

Remark 2.9 (Connectedness and continuous paths)**.** *Notice that the Dirichlet form satisfies the local property if and only if the underlying space is connected. Thus the variable speed motion on (T,r) has continuous paths if and only if (T,r) is an $\mathbb{R}$-tree.* □

Recall the explosion time ζ from (1.8). Notice that the Dirichlet form need not be conservative, which means that the speed-ν motion might exist only for a (random) finite *life time*. This happens when it explodes in finite time, i.e., $\zeta < \infty$.

Remark 2.10 (Finite versus infinite life time)**.** *Let (T,r,ν) be a rooted boundedly finite measure tree, and X the speed-ν motion on (T,r). Whether or not $\zeta = \infty$, almost surely, depends on the tree topology and the measure ν.*

(i) *The speed-ν motion on (T,r) cannot explode if it is recurrent. Recurrence depends on (T,r,ν) only through (T,r). See* [AEW13, Theorem 4] *for recurrence criteria.*
(ii) *An example of a transient variable speed motion with finite life time will be discussed in Example 5.5.* □

Lemma 2.11 (Variable speed motion on discrete trees is a Markov chain)**.** *Let (T,r,ν) be a metric boundedly finite measure tree such that (T,r) is discrete. Then the speed-ν random walk on (T,r) is a continuous time nearest neighbor Markov chain with jumps from v to $v' \sim v$ at rate $\gamma_{vv'} := \frac{1}{2\cdot\nu(\{v\})\cdot r(v,v')}$.*

Proof. Recall from Definition 1.1 that (T,r) is a Heine-Borel space. Thus each ball around ρ contains only a finite number of branch points, and in consequence the nearest neighbor random walk with the jump rates $(\gamma_{vv'})_{v\sim v'}$ is a well-defined strong Markov process. Its generator Ω acts on the space $\mathcal{C}_c(T)$ of continuous functions which depend only on finitely many $v \in T$ as follows:

$$\Omega f(v) := \tfrac{1}{2\cdot\nu(\{v\})} \sum_{v'\sim v} \tfrac{1}{r(v,v')}\big(f(v') - f(v)\big). \tag{2.16}$$

Notice that for all $f, g \in \mathcal{C}_c(T)$,

$$\begin{aligned}
\mathcal{E}(f,g) &= \tfrac{1}{2}\int \mathrm{d}\lambda \nabla f \nabla g \\
&= \tfrac{1}{2}\sum_{v\in T} \tfrac{1}{2}\sum_{v'\sim v} \tfrac{1}{r(v,v')}\big(f(v') - f(v)\big)\big(g(v') - g(v)\big) \\
&= -\sum_{v\in T} \nu(\{v\}) \tfrac{1}{2\nu(\{v\})} \sum_{v'\sim v} \tfrac{1}{r(v,v')}\big(f(v') - f(v)\big) g(v) \\
&= -\big(\Omega f, g\big)_\nu.
\end{aligned} \tag{2.17}$$

The statement therefore follows from Example 1.2.5 together with Exercise 4.4.1 in [FOT11]. □

2.3. **The occupation time formula.** We conclude this section by recalling here the occupation time formula as known from speed-ν motions on $\mathbb{R}$ or the ν-Brownian motion on compact metric trees (see, for example, [AEW13, Proposition 1.9]).

As usual, we denote for each $x \in T$ by

$$\tau_x = \tau_x(X) := \inf\{\, t \ge 0 : X_t = x \,\} \tag{2.18}$$

the *first hitting time* of x. A standard calculation shows the following:

Proposition 2.12 (Occupation time formula)**.** *Let X be a speed-ν motion on (T,r). If X is recurrent, then for all $x,z\in T$,*

$$\mathbb{E}^x\Big[\int_0^{\tau_z} f(X_t)\,\mathrm{d}t\Big] = 2\int_T f(y)\cdot r\big(z,c(x,z,y)\big)\,\nu(\mathrm{d}y), \tag{2.19}$$

for all bounded, measurable $f\colon T\to\mathbb{R}$. Moreover, the process $X_{\cdot\wedge\tau_z}$ is transient for all $z\in T$.

Proof. Let (T,r,ν) be a metric boundedly finite measure tree, and $z\in T$ fixed. By Lemma 2.6(iii), the Dirichlet form $(\mathcal{E},\mathcal{D}_{\{z\}}(\mathcal{E}))$ is transient. Thus, by Theorem 4.4.1(ii) in [FOT11], $R_{\{z\}}f(x) := \mathbb{E}^x[\int_0^{\tau_z}\mathrm{d}s\, f(X_s)]$ is the resolvent of the speed-ν motion killed on hitting z, i.e.,

$$\mathcal{E}(R_{\{z\}}f,h) = \int \mathrm{d}\nu\, h\cdot f, \tag{2.20}$$

for all $h\in\bar{\mathcal{D}}_{\{z\}}(\mathcal{E})$ and $f\in\mathcal{D}(\mathcal{E})$ with $(R_{\{z\}}f,f)_\nu<\infty$. The resolvent of a Markov process has the form

$$R_{\{z\}}f(x) = \int_T \nu(\mathrm{d}y)\,\tfrac{h^*_{\{z\},y}(x)}{\mathrm{cap}_{\{z\}}(y)} f(y), \tag{2.21}$$

where $\mathrm{cap}_{\{z\}}(y) := \inf\{\mathcal{E}(f,f):\ f\in\bar{\mathcal{D}}(\mathcal{E}),\ f(z)=0,\ f(y)=1\}$ and $h^*_{\{z\},y}$ is the unique minimizer for $\mathrm{cap}_{\{z\}}(y)$. This can be shown by essentially rewriting the argument laid out in [AEW13, Section 3]. Moreover, for our particular Dirichlet form we find that $h^*_{\{z\},y}(x) := \frac{r(c(x,y,z),z)}{r(y,z)}$ and $\mathrm{cap}_{\{z\}}(y) = \frac{1}{2r(y,z)}$, and thus that

$$\mathbb{E}^x\Big[\int_0^{\tau_z}\mathrm{d}s\, f(X_s)\Big] = 2\int \nu(\mathrm{d}y)\, r(z,c(x,y,z))f(y). \tag{2.22}$$

$\square$

3. Preliminaries on the Gromov-vague topology

Recall the notion of a rooted metric (boundedly finite) measure space (T,r,ρ,ν) from Definition 1.1. Once more, we call two rooted metric measure trees (T,r,ρ,ν) and (T',r',ρ',ν') equivalent iff there is an isometry φ between $\mathrm{supp}(\nu)\cup\{\rho\}$ and $\mathrm{supp}(\nu')\cup\{\rho'\}$ such that $\varphi(\rho)=\rho'$ and $\nu\circ\varphi^{-1}=\nu'$, and denote by

$$\mathbb{T} := \text{ the space of equivalence classes of rooted metric measure trees.} \tag{3.1}$$

In this section we want to equip $\mathbb{T}$ with the so-called *Gromov-Hausdorff-vague topology* on which the convergence of the underlying spaces in our invariance principle is based. We refer the reader to [ALW16] for many detailed discussions. We recall the definition of the pointed Gromov-weak topology on finite metric measure spaces in Subsection 3.1 and then extend it to a Gromov-vague topology on $\mathcal{T}$ in Subsection 3.2. Finally we compare the notions of Gromov-weak and Gromov-vague convergence in Subsection 3.3.

3.1. Gromov-weak and Gromov-Hausdorff-weak topology.

In this subsection we restrict to compact metric spaces and recall the Gromov-weak topology. This topology originates from the work of Gromov [Gro99] who considers topologies allowing to compare metric spaces who might not be subspaces of a common metric space. The *Gromov-weak topology* on complete and separable metric measure spaces was introduced in [GPW09]. In the same paper the Gromov-weak topology was metrized by the so-called *Gromov-Prohorov-metric* which is equivalent to Gromov's box metric introduced in [Gro99], as was shown in [Löh13]. The topology is closely related to the so-called *measured Gromov-Hausdorff topology* which was first introduced by [Fuk87], and further discussed in [KS03, EW06].

Remark 3.1 (Full-support assumption)**.** *Note that, by our definition, the measure ν of a metric boundedly finite measure tree (T,r,ρ,ν) is required to have full support. This is usually not assumed* *for metric measure spaces, but it is only a minor restriction, because, whenever $\rho \in \operatorname{supp}(\nu)$ we can choose representatives with full support.* □

Consider also the subspace

$$\mathbb{T}_c := \big\{(T, r, \rho, \nu) \in \mathbb{T} : \, (T, r) \text{ is compact}\big\}. \tag{3.2}$$

We shortly recall the basic definitions of the Gromov-weak and Gromov-Hausdorff-weak topologies on $\mathbb{T}_c$.

Definition 3.2 (Gromov-weak and Gromov-Hausdorff-weak topology)**.** *Let for each $n \in \mathbb{N} \cup \{\infty\}$, $\mathcal{X}_n := (T_n, r_n, \rho_n, \nu_n)$ be in $\mathbb{T}_c$. We say that $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}_\infty$ in*

(i) pointed Gromov-weak topology *if and only if there exists a complete, separable rooted metric space (E, d_E, ρ_E) and for each $n \in \mathbb{N} \cup \{\infty\}$ isometries $\varphi_n : T_n \to E$ with $\varphi_n(\rho_n) = \rho_E$, and such that*

$$(\varphi_n)_* \nu_n \underset{n\to\infty}{\Longrightarrow} (\varphi_\infty)_* \nu_\infty. \tag{3.3}$$

(ii) pointed Gromov-Hausdorff-weak topology *if and only if there exists a compact metric space (E, d_E, ρ_E) and for each $n \in \mathbb{N} \cup \{\infty\}$ isometries $\varphi_n \colon T_n \to E$ with $\varphi_n(\rho_n) = \rho_E$, such that* (3.3) *holds and*

$$\operatorname{supp}\big((\varphi_n)_* \nu_n\big) \xrightarrow[n\to\infty]{\text{Hausdorff}} \operatorname{supp}\big((\varphi_\infty)_* \nu_\infty\big). \tag{3.4}$$

Remark 3.3 (Supports do not converge under Gromov-weak convergence)**.** *Consider, for example, $T_n :\equiv \{\rho, \rho'\}$ and $r_n(\rho, \rho') \equiv 1$, and put $\nu_n := \frac{n-1}{n}\delta_\rho + \frac{1}{n}\delta_{\rho'}$ for all $n \in \mathbb{N}$. Clearly, $\big((T_n, r_n, \rho, \nu_n)\big)_{n\in\mathbb{N}}$ converges pointed Gromov-weakly to the unit mass pointed singleton $(\{\rho\}, \rho, \delta_\rho)$. The supports, however, do not converge. This shows that Gromov-weak is in general weaker than Gromov-Hausdorff-weak convergence.* □

In order to close the gap between Gromov-weak and Gromov-Hausdorff-weak convergence, we define for each $\delta > 0$ the *lower mass-bound function* $m_\delta \colon \mathbb{T} \to \mathbb{R}_+$ as

$$m_\delta\big((T, r, \rho, \nu)\big) := \inf\big\{\nu\big(\overline{B}_r(x, \delta)\big) \, : \, x \in T\big\}. \tag{3.5}$$

It follows from our full-support assumption, $\operatorname{supp}(\nu) = T$, that $m_\delta(\mathcal{X}) > 0$ for all $\delta > 0$ if $\mathcal{X} \in \mathbb{T}_c$.

Definition 3.4 (Global lower mass-bound property)**.** *We say that a family $\Gamma \subseteq \mathbb{T}_c$ satisfies the global lower mass-bound property if and only if the lower mass-bound functions are all bounded away from zero uniformly in Γ, i.e., for each $\delta > 0$,*

$$m_\delta\big(\Gamma\big) := \inf_{\mathcal{X}\in\Gamma} m_\delta(\mathcal{X}) > 0. \tag{3.6}$$

The following is Theorem 6.1 in [ALW16].

Proposition 3.5 (Gromov-weak versus Gromov-Hausdorff-weak topology)**.** *Let for each $n \in \mathbb{N} \cup \{\infty\}$, $\mathcal{X}_n := (T_n, r_n, \rho_n, \nu_n)$ be in $\mathbb{T}_c$ such that $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}_\infty$ pointed Gromov-weakly. Then the following are equivalent:*

(i) *The sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}_\infty$ pointed Gromov-Hausdorff-weakly.*
(ii) *The sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ satisfies the global lower mass-bound property.*

3.2. **Gromov-vague and Gromov-Hausdorff-vague topology.**
Recently, in [ADH13], the Gromov-Hausdorff-weak topology on rooted compact length spaces was extended to complete locally compact length spaces equipped with locally finite measures. In this subsection we want, in similar spirit, extend the Gromov(-Hausdorff)-weak topology on $\mathbb{T}_c$ to the Gromov(-Hausdorff)-vague topology on $\mathbb{T}$.

The restriction of $\mathcal{X} = (X, r, \rho, \nu) \in \mathbb{T}$ to the closed ball $\overline{B}(\rho, R)$ of radius $R > 0$ around the root is denoted by

$$\mathcal{X}\restriction_R := \big(\overline{B}(\rho, R), r, \rho, \nu\restriction_{\overline{B}_r(\rho,R)}\big). \tag{3.7}$$

Definition 3.6 (Gromov-vague topology)**.** *Let for each* $n \in \mathbb{N} \cup \{\infty\}$*,* $\mathcal{X}_n := (T_n, r_n, \rho_n, \nu_n)$ *be in* $\mathbb{T}$*. We say that* $(\mathcal{X}_n)_{n\in\mathbb{N}}$ *converges to* $\mathcal{X}_\infty$ *in*

(i) pointed Gromov-vague topology *if and only if there exists a complete, separable rooted metric space* (E, d_E, ρ_E) *and for each* $n \in \mathbb{N}\cup\{\infty\}$ *isometries* $\varphi_n : T_n \to E$ *with* $\varphi_n(\rho_n) = \rho_E$*, and such that*

$$\big((\varphi_n)_*\nu_n\big)\restriction_R \underset{n\to\infty}{\Longrightarrow} \big((\varphi_\infty)_*\nu_\infty\big)\restriction_R \tag{3.8}$$

for all but countably many $R > 0$*.*

(ii) pointed Gromov-Hausdorff-vague topology *if and only if there exists a rooted Heine-Borel space* (E, d_E, ρ_E) *and for each* $n \in \mathbb{N} \cup \{\infty\}$ *isometries* $\varphi_n : T_n \to E$ *with* $\varphi_n(\rho_n) = \rho_E$*, and such that* (3.8) *and*

$$\varphi_n(T_n) \cap \overline{B}_{d_E}(\rho_E, R) \xrightarrow[n\to\infty]{Hausdorff} \varphi(T) \cap \overline{B}_{d_E}(\rho_E, R) \tag{3.9}$$

hold for all but countably many $R > 0$*.*

Once more we want to close the gap between Gromov-vague and Gromov-Hausdorff-vague convergence. Define therefore for all $\delta > 0$ and $R > 0$, the *local lower mass-bound function* $m_\delta^R \colon \mathbb{T} \to \mathbb{R}_+ \cup \{\infty\}$ as

$$m_\delta^R\big((T, r, \rho, \nu)\big) := \inf\big\{\nu\big(\overline{B}_r(x, \delta)\big) \,:\, x \in B(\rho, R)\big\}. \tag{3.10}$$

Notice that $m_\delta^R(\mathcal{X}) > 0$ for all $\mathcal{X} \in \mathbb{T}$, and $\delta, R > 0$.

Definition 3.7 (Local lower mass-bound property)**.** *We say that a family* $\Gamma \subseteq \mathbb{T}$ *satisfies the local lower mass-bound property if and only if the lower mass-bound functions are all bounded away from zero uniformly in* Γ*, i.e., for each* $\delta > 0$ *and* $R > 0$*,*

$$m_\delta^R(\Gamma) := \inf_{\mathcal{X}\in\Gamma} m_\delta^R(\mathcal{X}) > 0. \tag{3.11}$$

The following is Corollary 5.2 in [ALW16].

Proposition 3.8 (Gromov-vague versus Gromov-Hausdorff-vague)**.** *Let for each* $n \in \mathbb{N} \cup \{\infty\}$*,* $\mathcal{X}_n := (T_n, r_n, \rho_n, \nu_n)$ *be in* $\mathbb{T}$ *such that* $(\mathcal{X}_n)_{n\in\mathbb{N}}$ *converges to* $\mathcal{X}_\infty$ *pointed Gromov-vaguely. Then the following are equivalent:*

(i) *The sequence* $(\mathcal{X}_n)_{n\in\mathbb{N}}$ *converges to* $\mathcal{X}_\infty$ *pointed Gromov-Hausdorff-vaguely.*
(ii) *The sequence* $(\mathcal{X}_n)_{n\in\mathbb{N}}$ *satisfies the local lower mass-bound property.*

3.3. **Gromov-weak versus Gromov-vague convergence.** Note that the concept of Gromov-vague convergence on $\mathbb{T}$ is not strictly an extension of the concept of Gromov-weak convergence on $\mathbb{T}_c$ because in the limit parts might "vanish at infinity", and hence a non-converging sequence of compact spaces with respect to the Gromov-weak or the Gromov-Hausdorff-weak topology may converge in the "locally compact version" of the corresponding topology.

Remark 3.9 (Gromov-vague versus Gromov-weak)**.** *Consider the subspaces $\mathbb{T}_{finite}$ and $\mathbb{T}_{probability}$ of $\mathbb{T}$ consisting of spaces $\mathcal{X} = (T, r, \rho, \nu) \in \mathbb{T}$ where ν is a finite or a probability measure, respectively. Then on $\mathbb{T}_{probability}$ the induced Gromov-vague topology coincides with the Gromov-weak topology. However, on $\mathbb{T}_{finite}$ and even on $\mathbb{T}_c$ this is not the case as the total mass might not be preserved under Gromov-vague convergence. In fact, for $\mathcal{X}, \mathcal{X}_n = (T_n, r_n, \rho_n, \nu_n) \in \mathbb{T}_{finite}$ the following are equivalent:*

(i) *$\mathcal{X}_n \to \mathcal{X}$ Gromov-weakly.*
(ii) *$\mathcal{X}_n \to \mathcal{X}$ Gromov-vaguely and $\nu_n(T_n) \to \nu(T)$.*

Moreover, $\mathcal{X}_n \to \mathcal{X} \in \mathbb{T}_c$ Gromov-Hausdorff-weakly if and only if $\mathcal{X}_n \to \mathcal{X}$ Gromov-Hausdorff-vaguely and the diameters of (T_n, r_n) are bounded uniformly in n (except for finitely many n). □

4. TIGHTNESS

Recall the speed-ν motion on (T, r), $X^{(T,r,\nu)}$, from Definition 2.7. In this section we prove that the sequence $\{X^{(T_n,r_n,\nu_n)};\, n \in \mathbb{N}\}$ is tight provided that Assumptions (A0), (A1) and (A2) from Theorem 1 are satisfied. The main result is the following:

Proposition 4.1 (Tightness)**.** *Let $\mathcal{X} := (T, r, \rho, \nu)$ and $\mathcal{X}_n := (T_n, r_n, \rho_n, \nu_n)$, $n \in \mathbb{N}$, be rooted metric boundedly finite measure trees. Assume that for all $n \in \mathbb{N}$, $\mathcal{X}_n$ is discrete, and that the following conditions hold:*

(A0) *For all $R > 0$,*

$$\limsup_{n\to\infty} \sup\big\{ r_n(x,z) \,:\, x \in B_n(\rho_n, R),\, z \in T_n,\, x \sim z \big\} < \infty. \tag{4.1}$$

(A1) *The sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}$ in the pointed Gromov-vague topology as $n \to \infty$.*
(A2) *The local lower mass-bound property holds uniformly in $n \in \mathbb{N}$.*

Then there is a Heine Borel space (E, d), such that T and all T_n, $n \in \mathbb{N}$, are embedded in (E, d) and the sequence X^n, $n \in \mathbb{N}$, of speed-ν_n random walks on (T_n, r_n) is tight in the one-point compactification of E.

For the proof we rely on the following version of the Aldous tightness criterion (see, [Kal02, Theorem 16.11+16.10]).

Proposition 4.2 (Aldous tightness criterion)**.** *Let $X^n = (X^n_t)_{t\geq 0}$, $n \in \mathbb{N}$, be a sequence of càdlàg processes on a complete, separable metric space (E, d). Assume that the one-dimensional marginal distributions are tight, and for any bounded sequence of X^n-stopping times τ_n and any $\delta_n > 0$ with $\delta_n \to 0$ we have*

$$d\big(X^n_{\tau_n}, X^n_{\tau_n+\delta_n}\big) \xrightarrow{n\to\infty} 0 \quad \textit{in probability.} \tag{4.2}$$

Then the sequence $(X^n)_{n\in\mathbb{N}}$ is tight.

To verify Proposition 4.2, we have to show that it is unlikely that the walk has moved more than a certain distance in a sufficiently small amount of time, uniformly in n and the starting point.

Corollary 4.3. *Let (E, d) be a locally compact, separable metric space. For each $n \in \mathbb{N}$, let $T_n \subseteq E$ and $(X^n, (\mathbb{P}^x)_{x\in T_n})$ a strong Markov process on T_n. Assume that for every $\varepsilon > 0$*

$$\lim_{t\to 0} \lim_{n\to\infty} \sup_{x\in T_n} \mathbb{P}^x\big\{ d(x, X^n_t) > \varepsilon \big\} = 0. \tag{4.3}$$

Then for every sequence of initial distributions $\mu_n \in \mathcal{M}_1(T_n)$ the sequence $(X^n)_{n\in\mathbb{N}}$ is tight as processes on the one-point compactification of E.

Proof. Let $\widehat{E} = E \cup \{\infty\}$ be the one-point compactification of E. $\widehat{E}$ is metrizable, and we can choose a metric $\hat{d}$ with $\hat{d} \le d$ on $E \times E$. A possible choice is

$$\hat{d}(x,y) := \inf_{n\in\mathbb{N}} \inf_{z_1,\ldots,z_n\in E} \sum_{k=0}^{n} e^{-\inf\{j:z_k\in U_j \text{ or } z_{k+1}\in U_j\}} \big(1 \wedge d(z_k, z_{k+1})\big), \tag{4.4}$$

where $z_0 := x$, $z_{n+1} := y$, and $U_1 \subseteq U_2 \subseteq \cdots$ are (fixed) open, relatively compact subsets of E with $E = \bigcup_{n\in\mathbb{N}} U_n$. By the strong Markov property, (4.3) implies (4.2) for d and hence also for $\hat{d}$. By Proposition 4.2, $(X^n)_{n\in\mathbb{N}}$ is tight on $\widehat{E}$. $\square$

From here we proceed in several steps. We first give an estimate for the probability to reach a particular point in a small amount of time. We are then seeking an estimate for the probability that the walk has moved more than a given distance away from the starting point. For that we will need a bound on the number of possible directions the random walk might have taken until reaching that distance.

Recall from (2.18) the first hitting time τ_x of a point $x \in T$.

Lemma 4.4 (Hitting time bound)**.** *Let (T, r, ρ, ν) be a discrete rooted metric boundedly finite measure tree, $x \in T$, X the speed-ν random walk on (T, r) started at x. Fix $v \in T$ and $\delta \in (0, r(x, v))$. Denote by $S := B(x, \delta)$ the subtree δ-close to x and let $R := r(S, v)$. Then, for all $t \ge 0$,*

$$\mathbb{P}^x\big\{\tau_v \le t\big\} \le 2\Big(1 - \tfrac{R}{R+2\delta} e^{-\frac{t}{R\nu(S)}}\Big). \tag{4.5}$$

Proof. Assume w.l.o.g. that X is recurrent and let w be the unique point in S with $r(v, w) = R$. Obviously, if X starts in x then it must pass w before hitting v. Neglect the time until w and assume X starts in w instead of x. For $u \in S$, let t_u be the (random) amount of time spent in u before hitting v, $r_u := r(w, u)$, and $m_u := \nu\big(\{u\}\big)$. Using that a geometric sum of independent, exponentially distributed random variables is again exponentially distributed, it is easy to see that the law of t_u is

$$\mathcal{L}^w(t_u) = \tfrac{r_u}{R+r_u}\delta_0 + \tfrac{R}{R+r_u} \operatorname{Exp}\big(\tfrac{1}{2(R+r_u)m_u}\big), \tag{4.6}$$

where $\operatorname{Exp}(\lambda)$ denotes an exponential distribution with expectation $\frac{1}{\lambda}$, and δ_0 the Dirac measure in 0.

As $\tau_v \ge \sum_{u\in S} t_u$, we find that for every $a > 0$,

$$\tau_v \ge \sum_{u\in S} am_u \mathbf{1}_{\{t_u\ge am_u\}} = a\,\nu\big(\{u \in S :\, t_u \ge am_u\}\big). \tag{4.7}$$

Now we pick $a := \frac{2t}{\nu(S)}$ and obtain

$$\begin{aligned}
\mathbb{P}^x\big\{\tau_v \le t\big\} &\le \mathbb{P}^w\big\{\nu\{u \in S :\, t_u \ge am_u\} \le \tfrac{1}{2}\nu(S)\big\} \\
&= \mathbb{P}^w\big\{\nu\{u \in S :\, t_u < am_u\} \ge \tfrac{1}{2}\nu(S)\big\} \\
&\le \tfrac{2}{\nu(S)}\mathbb{E}^w\big[\nu\{u \in S :\, t_u < am_u\}\big] \\
&= \tfrac{2}{\nu(S)} \sum_{u\in S} m_u \mathbb{P}^w\big\{t_u < 2t\tfrac{m_u}{\nu(S)}\big\},
\end{aligned} \tag{4.8}$$

which together with (4.6) and the fact that $r_u \le 2\delta$ gives the claim. $\square$

To get bounds on the probability to move sufficiently far from bounds on the probability to hit a pre-specified point, we need a bound on the number of directions the random walk can take in order to get far away. With ε-degree of a node x we mean the number of edges that intersect the ε-sphere around x and are connected to points at least 2ε away from x.

Definition 4.5 (ε-degree)**.** *Let (T,r) be a discrete metric tree. For $\varepsilon > 0$, $x \in T$, let $B := B(x,\varepsilon)$ be the ε-ball around x. The ε-degree of x is*

$$\begin{aligned} \deg_\varepsilon(x) &:= \deg_\varepsilon^T(x) \\ &:= \#\big\{v \in T \setminus B : \exists u \in B, w \in T \setminus B(x,2\varepsilon) : u \sim v,\, v \in [u,w]\big\}. \end{aligned} \tag{4.9}$$

We also define the maximal degree as

$$\deg_\varepsilon(T) := \sup_{x \in T} \deg_\varepsilon^T(x). \tag{4.10}$$

Lemma 4.6 (Topological bound)**.** *Let $\mathcal{X}_n := (T_n, r_n)$, $n \in \mathbb{N}$, be discrete metric trees, and $\mathcal{X} := (T,r)$ a compact metric tree. If $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}$ in Gromov-Hausdorff topology, then for every $\varepsilon > 0$,*

$$\limsup_{n\to\infty} \deg_\varepsilon(T_n) < \infty. \tag{4.11}$$

Proof. Fix $\varepsilon > 0$. As $\mathcal{X}_n \to \mathcal{X}$ in Gromov-Hausdorff topology, there exists a finite ε-net S in T, and ε-nets S_n in T_n, such that for all sufficiently large $n \in \mathbb{N}$, S_n has the same cardinality as S (see, for example, [BBI01, Proposition 7.4.12]). Obviously, this common cardinality is an upper bound for $\{\deg_\varepsilon(T_n);\, n \in \mathbb{N}\}$. $\square$

With the notion of an ε-degree of a tree, we can immediately conclude the following.

Lemma 4.7 (Speed bound)**.** *Let (T,r,ρ,ν) be a discrete metric boundedly finite measure tree, $x \in T$, and X the speed-ν random walk on (T,r). Then for every $\varepsilon > 0$, $\delta \in (0,\varepsilon)$ and $t < (\varepsilon-\delta)m$, where $m := \nu\big(B(x,\delta)\big)$,*

$$\mathbb{P}^x\Big\{\sup_{s\in[0,t]} r(X_s,x) > 2\varepsilon\Big\} \le 2\deg_\varepsilon(x)\Big(1 - \tfrac{\varepsilon-\delta}{\varepsilon+\delta}\exp\big(-\tfrac{t}{\varepsilon m}\big)\Big). \tag{4.12}$$

Proof. Let $v_1,\dots,v_N$ be the points outside $B(x,\varepsilon)$ that are neighbours of a point inside $B(x,\varepsilon)$ and on the way from x to a point outside $B(x,2\varepsilon)$. Then $N \le \deg_\varepsilon(x)$. Under $\mathbb{P}^x$, if $r(X_s,x) > 2\varepsilon$ for some $s \le t$, X must have hit at least one point in $\{v_1,\dots,v_N\}$ before time s. Hence the claim follows from Lemma 4.4. $\square$

Proof of Proposition 4.1. According to Proposition 3.8, $\mathcal{X}_n \to \mathcal{X}$ in Gromov-Hausdorff-vague topology. Hence, we may assume that there is a rooted Heine-Borel space (E,d,ρ_E), such that $T_n, T \subseteq E$, $\rho_E = \rho = \rho_n$ for all $n \in \mathbb{N}$, and, for all but countably many $R > 0$, we have both

$$T_n \cap \overline{B}_d(\rho,R) \to T \cap \overline{B}_d(\rho,R) \tag{4.13}$$

as subsets of E in Hausdorff topology, and

$$\nu_n\restriction_R \Rightarrow \nu\restriction_R. \tag{4.14}$$

Let $\widehat{E} = E \cup \{\infty\}$ be the one-point compactification of E, metrized by a metric $\hat{d}$ with $\hat{d} \le d$ on E^2 (see, for example, (4.4)). For each $x \in \widehat{E}$ and $N \in \mathbb{N}$, write $B_{\hat{d}}(x,\frac{1}{N}) := \{y \in \widehat{E} : \hat{d}(x,y) < \frac{1}{N}\}$ and put

$$K_N := \widehat{E} \setminus B_{\hat{d}}(\infty,\tfrac{1}{N}) \subseteq E. \tag{4.15}$$

Notice that K_N is compact by definition.

To show tightness, we show that condition (4.3) of Corollary 4.3 is satisfied for the metric $\hat{d}$, i.e., for given $\varepsilon, \hat{\varepsilon} > 0$, we can construct $t_0 > 0$ such that

$$\sup_{x\in T_n} \mathbb{P}^x\big\{\hat{d}(x,X_t^n) > \varepsilon\big\} \le \hat{\varepsilon}, \tag{4.16}$$

for all $t \in [0,t_0]$ and all $n \in \mathbb{N}$.

Fix $\varepsilon > 0$, and choose $N > \frac{4}{\varepsilon}$. Then the diameter of $\widehat{E} \setminus K_N$ with respect to $\hat{d}$ is at most $\frac{1}{2}\varepsilon$. Let

$$e_N := \sup_{n\in\mathbb{N}} \sup_{x\in T_n\cap K_N} \sup_{y\sim x} d(x,y) \tag{4.17}$$

be the supremum of edge-lengths emanating from points in $T_n \cap K_N$, and note that $e_N < \infty$ by assumption. Now choose $M > N$ such that K_M contains the e_N-neighbourhood of K_N, i.e., $\{x' \in E : d(K_N, x') < e_N\} \subseteq K_M$. Then all points of K_M which are connected to a point in $E \setminus K_M$ (within some T_n) are actually in $K_M \setminus K_N$.

Consider the hitting time of K_M, $\tau_{K_M} := \inf\{s \geq 0 : X^n_s \in K_M\}$, and recall that the $\hat{d}$-diameter of $\widehat{E} \setminus K_N$ is at most $\frac{\varepsilon}{2}$. Therefore, if X^n starts in $x \in T_n$, then $\hat{d}(x, X^n_t) > \varepsilon$ implies $\tau_{K_M} < t$ and $\hat{d}(x, X^n_{\tau_{K_M}}) \leq \frac{\varepsilon}{2}$. Using the strong Markov property at τ_{K_M}, we obtain for all $n \in \mathbb{N}$, $x \in T_n$,

$$\mathbb{P}^x\big\{\hat{d}(x, X^n_t) > \varepsilon\big\} \leq \sup_{y\in T_n\cap K_M} \sup_{s\in[0,t]} \mathbb{P}^y\big\{\hat{d}(y, X^n_s) > \tfrac{1}{2}\varepsilon\big\}. \tag{4.18}$$

Applying Lemma 4.7, we conclude for all $\delta \in (0,\varepsilon)$ and $t < \frac{1}{4}(\varepsilon-\delta)m_\delta$, where $m_\delta := \inf_{n\in\mathbb{N}} \inf_{y\in T_n\cap K_M} \nu_n\big(B(y, \frac{\delta}{4})\big)$,

$$\mathbb{P}^x\big\{\hat{d}(x, X^n_t) > \varepsilon\big\} \leq 2\deg_{\frac{\varepsilon}{4}}(T_n \cap K_M)\big(1 - \tfrac{\varepsilon-\delta}{\varepsilon+\delta}\exp\big(-\tfrac{4t}{\varepsilon m_\delta}\big)\big). \tag{4.19}$$

As $D := \sup_{n\in\mathbb{N}} \deg_{\frac{\varepsilon}{4}}(T_n \cap K_M) < \infty$ by Lemma 4.6, and $m_\delta > 0$ by the local lower mass-bound property (A2), we can choose $\delta > 0$ small enough such that $\frac{\varepsilon-\delta}{\varepsilon+\delta} > 1 - \frac{\hat{\varepsilon}}{4D}$, and subsequently $t_0 < \frac{1}{4}(\varepsilon-\delta)m_\delta$ such that $\exp(-\frac{4t_0}{\varepsilon m_\delta}) > 1 - \frac{\hat{\varepsilon}}{4D}$. Inserting this into (4.19), we obtain (4.16) and tightness follows from Corollary 4.3. $\square$

5. Identifying the limit

In this section we identify the limit process. For this purpose, we use a characterization from [Ald91, Section 5], where the existence of a diffusion process on a particular non-trivial continuum tree, the so-called Brownian CRT (T, r, ν) from Example 1.5, was shown. Aldous defines this diffusion as a strong Markov process on T with continuous path such that ν is the reversible equilibrium and it satisfies the following two properties:

(i) For all $a, b, x \in T$ with $x \in [a,b]$, $\mathbb{P}^x\{\tau_a < \tau_b\} = \frac{r(x,b)}{r(a,b)}$.
(ii) The occupation time formula (0.1) holds.

While (i) reflects the fact that this diffusion is on "natural scale", (ii) recovers ν as the "speed" measure. At several places in the literature constructions of diffusions on the CRT and more general continuum random trees rely on Aldous' characterisation (see, for example, [Kre95, Cro08, Cro10]). Albeit the diffusions can be indeed characterised by (i) and (ii) uniquely, a formal proof for this fact has to the best of our knowledge never been given anywhere. We want to close this gap, and even show that the requirement (i) is redundant.

The following result will be proven in Subsection 6.1.

Proposition 5.1 (Characterization via occupation time formula)**.** *Assume that (T, r) is a compact metric tree, and that we are given two T-valued strong Markov processes X and Y such that for all $x, y \in T$, and bounded measurable $f : T \to \mathbb{R}_+$,*

$$\mathbb{E}^x\Big[\int_0^{\tau_y} \mathrm{d}t\, f(X_t)\Big] = \mathbb{E}^x\Big[\int_0^{\tau_y} \mathrm{d}t\, f(Y_t)\Big]. \tag{5.1}$$

Assume further that $X_{\cdot\wedge\tau_y}$ is transient for all $y \in T$. Then the laws of X and Y agree.

We will rely on Proposition 5.1 and show for compact limiting trees that any limit point satisfies the strong Markov property in Subsection 5.1 and the occupation time formula (0.1) in Subsection 5.2. Note that, if $\mathcal{X}_n = (T_n, r_n, \rho_n, \nu_n)$ converges to $\mathcal{X} = (T, r, \rho, \nu)$ pointed Gromov-Hausdorff-vaguely (i.e. we assume (A1) and (A2) of Theorem 1), then compactness of $\mathcal{X}$ together with assumption (A0) of Theorem 1 is equivalent to the uniform diameter bound $\sup_{n\in\mathbb{N}} \mathrm{diam}(T_n, r_n) < \infty$.

5.1. **The strong Markov property of the limit.** In this subsection we show that any limit point has the strong Markov property. To be more precise, the main result is the following:

Proposition 5.2 (Strong Markov property)**.** *Let $\mathcal{X} := (T, r, \nu)$ and $\mathcal{X}_n := (T_n, r_n, \nu_n)$, $n \in \mathbb{N}$, be metric boundedly finite measure trees. Assume that all $\mathcal{X}_n$, $n \in \mathbb{N}$, are discrete with $\sup_{n\in\mathbb{N}} \mathrm{diam}(T_n, r_n) < \infty$, and that the sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}$ Gromov-Hausdorff-vaguely as $n \to \infty$. If X^n is the speed-ν_n random walk on (T_n, r_n) and $X^n \underset{n\to\infty}{\Longrightarrow} \tilde{X}$ in path space, then $\tilde{X}$ is a (strong Markov) Feller process.*

In order to prove Proposition 5.2, we will first show that under its assumptions the family of functions $\{P_n : n \in \mathbb{N}\}$, where for each $n \in \mathbb{N}$

$$P_n \colon \begin{cases} T_n \times \mathbb{R}_+ & \to \quad \mathcal{M}_1(E), \\ \quad (x, t) & \mapsto \quad \mathcal{L}^x(X^n_t) =: P^x_{n,t} \end{cases}, \tag{5.2}$$

is uniformly equicontinuous. Here, $\mathcal{L}^x(X^n_t)$ denotes the law of X^n_t, where X^n is started in $x \in T_n$, E is a metric space containing all T_n, and $\mathcal{M}_1(E)$ is equipped with the Prohorov metric.

Lemma 5.3 (Equicontinuity)**.** *Let $\mathcal{X} := (T, r, \nu)$ and $\mathcal{X}_n := (T_n, r_n, \nu_n)$, $n \in \mathbb{N}$, be metric boundedly finite measure trees. Assume that all $\mathcal{X}_n$, $n \in \mathbb{N}$, are discrete with $\sup_{n\in\mathbb{N}} \mathrm{diam}(T_n, r_n) < \infty$, and that $\mathcal{X}_n \to \mathcal{X}$ Gromov-Hausdorff-vaguely. If for each $n \in \mathbb{N}$, X^n is the speed-ν_n random walk on (T_n, r_n), and $P_n \colon \mathbb{R}_+ \times T_n \to \mathcal{M}_1(E)$ is defined as in (5.2), then the family $\{P_n : n \in \mathbb{N}\}$ is uniformly equicontinuous.*

Proof. Fix $\varepsilon > 0$. We construct a $\delta > 0$, independent of n, such that $P^x_{n,s}$ and $P^y_{n,t}$ are ε-close whenever $x, y \in T_n$, $s, t \in \mathbb{R}_+$ are such that $r_n(x, y) < \delta$ and $s \le t \le s + \delta$.

Fix $n \in \mathbb{N}$, and denote for any two $x, y \in T_n$ by X^x and X^y speed-ν_n random walks on (T_n, r_n) starting in x and y, respectively, which are coupled as follows: let the random walks X^x, X^y run independently until X^x hits y for the first time, i.e., until $\tau := \inf\{t \ge 0 : X^x_t = y\}$, and put $X^x_{\tau+\cdot} = X^y$. In particular, whenever $s \ge \tau$, we obtain $X^x_s = X^y_{t-u}$ for $u = \tau + t - s$.

Using the strong Markov property of X^y, we can estimate for any $c \in [t - s, t]$

$$\mathbb{P}\big\{r_n(X^x_s, X^y_t) > \varepsilon\big\} \le \mathbb{P}\{\tau > c - t + s\} + \sup_{z\in T_n} \mathbb{P}\big\{\sup_{u\in[0,c]} r_n(z, X^z_u) > \varepsilon\big\}. \tag{5.3}$$

For small t, we need another estimate, namely for $r_n(x, y) \le \frac{1}{3}\varepsilon$ we have

$$\mathbb{P}\big\{r_n(X^x_s, X^y_t) > \varepsilon\big\} \le 2 \sup_{z\in T_n} \mathbb{P}\big\{\sup_{u\in[0,t]} r_n(z, X^z_u) > \tfrac{\varepsilon}{3}\big\} =: 2q_t. \tag{5.4}$$

Combining (5.3) and (5.4), we obtain, under the condition $r_n(x, y) \le \frac{1}{3}\varepsilon$, for any $c \ge t - s$

$$\mathbb{P}\big\{r_n(X^x_s, X^y_t) > \varepsilon\big\} \le q_c + q_c \vee \mathbb{P}\big\{\tau > c - (t - s)\big\}. \tag{5.5}$$

Note that this estimate depends on x, y, s, t only through $r_n(x, y)$ and $t - s$.

The Gromov-Hausdorff-vague convergence together with the uniform diameter bound on (T_n, r_n) implies that (T, r) is compact and (T_n, r_n) converges to (T, r) in Gromov-Hausdorff topology. Hence, by Lemma 4.6, $\sup_{n\in\mathbb{N}} \deg_{\frac{\varepsilon}{6}}(T_n) < \infty$. Furthermore, the global lower mass-bound property is satisfied, i.e. for every $\varepsilon' > 0$, $m_{\varepsilon'} := \inf_{n\in\mathbb{N}, x\in T_n} \nu_n\big(B_n(x, \varepsilon')\big) > 0$. We can thus apply Lemma 4.7 to obtain a sufficiently small $c = c(\varepsilon) > 0$, independent of n, such that $q_c \le \frac{\varepsilon}{2}$. To estimate (for this c) $\mathbb{P}\{\tau > c-(t-s)\}$, we note that $M := \sup_{n\in\mathbb{N}} \nu_n(T_n) < \infty$ because of the diameter bound, and obtain for $t-s \le \frac{1}{2}c$

$$\mathbb{P}\big\{\tau > c-(t-s)\big\} \le \tfrac{2}{c}\mathbb{E}[\tau] \le \tfrac{4}{c}M \cdot r_n(x,y). \tag{5.6}$$

Choose therefore $\delta := \frac{\varepsilon}{8M}c \wedge \frac{\varepsilon}{3} \wedge \frac{1}{2}c$. Then for all $x, y \in T_n$ with $r_n(x,y) < \delta$, and $0 \le s \le t < s+\delta$, (5.5) implies $\mathbb{P}\big\{r_n(X_s^x, X_t^y) > \varepsilon\big\} \le \varepsilon$, and hence $d_{\mathrm{Pr}}(P^x_{n,s}, P^y_{n,t}) \le \varepsilon$, which is the claimed equicontinuity. □

The proof of Proposition 5.2 relies on the following modification of the Arzelà-Ascoli theorem, which is proven in the same way as the classical theorem.

Lemma 5.4 (Arzelà-Ascoli)**.** *Let (E,d) be a compact metric space, (F,d_F) a metric space, $T, T_n \subseteq E$ closed and $f_n\colon T_n \to F$ for $n \in \mathbb{N}$. Further assume that the family $\{f_n;\, n\in\mathbb{N}\}$ is uniformly equicontinuous with modulus of continuity h, and that for all $x \in T$ there exists $x_n \in T_n$ such that $x_n \to x$ and $\{f_n(x_n) : n \in \mathbb{N}\}$ is relatively compact in F. Then there is a function $f\colon T \to F$, a subsequence of $(f_n)_{n\in\mathbb{N}}$, again denoted by (f_n), and $\varepsilon_n > 0$ with $\varepsilon_n \to 0$ such that for all $n \in \mathbb{N}$, for all $x \in T$ and $y \in T_n$,*

$$d_F\big(f(x), f_n(y)\big) \le h\big(d(x,y)\big) + \varepsilon_n. \tag{5.7}$$

Note that (5.7) in particular implies that f is continuous with the same modulus of continuity h, and that $f_n(x_n) \to f(x)$ whenever $x_n \to x$.

Proof of Proposition 5.2. By assumption, there is a compact metric space (E,d) such that $T, T_n \subseteq E$, $d{\restriction}_{T_n} = r_n$ for all $n \in \mathbb{N}$, $d{\restriction}_T = r$, and $(T_n, r_n, \nu_n)_{n\in\mathbb{N}}$ converges Hausdorff-weakly to (T, r, ν).

According to Proposition 4.1 and Lemma 5.3, the assumptions of Arzelà-Ascoli are satisfied for the family of functions P_n, $n \in \mathbb{N}$, defined in (5.2). Thus we obtain a continuous subsequential limit $P\colon T \times \mathbb{R}_+ \to \mathcal{M}_1(E)$, $(x,t) \mapsto P_t^x$. Let $S = (S_t)_{t\ge 0}$ and $S^n = (S_t^n)_{t\ge 0}$ be the corresponding operators on $\mathcal{C}(T)$ and $\mathcal{C}(T_n)$, respectively. That is $S_t f(x) := \int_T f\,\mathrm{d}P_t^x$ and $S_t^n f(x) := \int_{T_n} f\,\mathrm{d}P^x_{n,t}$, $n \in \mathbb{N}$. We show that S is indeed a strongly continuous semigroup.

To this end, it is enough to show $\lim_{t\to 0} \|S_t f - f\|_\infty = 0$ and $S_{t+s}f = S_t(S_s f)$, $s,t > 0$, for Lipschitz continuous $f \in \mathcal{C}(T)$ with Lipschitz constant (at most) 1 and $\|f\|_\infty \le 1$. We can extend every such f to a function on E with the same properties. Let $\mathrm{Lip}_1 = \mathrm{Lip}_1(E)$ be the space of such (extended) f and recall that the Kantorovich-Rubinshtein metric between two measures $\mu, \hat\mu \in \mathcal{M}_1(E)$,

$$d_{\mathrm{KR}}\big(\mu, \hat\mu\big) := \sup_{f\in\mathrm{Lip}_1} \int f\,\mathrm{d}(\mu - \hat\mu), \tag{5.8}$$

is uniformly equivalent to the Prohorov metric (see [Bog07, Thm. 8.10.43]). For the rest of the proof, $\mathcal{M}_1(E)$ is equipped with d_{KR}. Let h be a common modulus of continuity for all P_n, $n \in \mathbb{N}$, which exists according to Lemma 5.3. Due to Lemma 5.4, P has the same modulus of continuity and hence, for all $f \in \mathrm{Lip}_1$,

$$\|S_t f - f\|_\infty \le \sup_{x\in T} d_{\mathrm{KR}}(P_t^x, P_0^x) \le h(t) \xrightarrow[t\to 0]{} 0, \tag{5.9}$$

i.e. S is strongly continuous.

Because T_n converges to T in the Hausdorff metric, we find $g_n\colon T_n \to T$ such that

$$\alpha_n := \sup_{y\in T_n} d\big(y, g_n(y)\big) \xrightarrow[n\to\infty]{} 0. \tag{5.10}$$

W.l.o.g. we may also assume that $T_1, T_2, \ldots$, are disjoint. As the spaces (T_n, r_n), $n \in \mathbb{N}$, are discrete, the map

$$g\colon T \cup \bigcup_{n\in\mathbb{N}} T_n \to T, \quad x \mapsto \begin{cases} x, & x \in T \\ g_n(x), & x \in T_n \end{cases} \tag{5.11}$$

is continuous. Now we apply (5.7) to P_n and P and obtain for all $n \in \mathbb{N}$, $f \in \mathrm{Lip}_1$ and $s > 0$

$$\begin{aligned} \sup_{y\in T_n} \big| S_s^n f(y) - (S_s f)\big(g(y)\big)\big| &\le \sup_{y\in T_n} d_{\mathrm{KR}}(P^y_{n,s}, P_s^{g(y)}) \\ &\le h(\alpha_n) + \varepsilon_n \xrightarrow[n\to\infty]{} 0, \end{aligned} \tag{5.12}$$

where ε_n is obtained in Lemma 5.4. For $x \in T$, there exists $x_n \in T_n$ with $x_n \to x$ and thus, using (5.12) and the semigroup property of S^n,

$$\begin{aligned} S_{t+s} f(x) &= \lim_{n\to\infty} S^n_{t+s} f(x_n) = \lim_{n\to\infty} S_t^n (S_s^n f)(x_n) \\ &= \lim_{n\to\infty} S_t^n (S_s f \circ g)(x_n) = S_t(S_s f \circ g)(x) \\ &= S_t(S_s f)(x). \end{aligned} \tag{5.13}$$

Now it is standard to see that S comes from a Feller process, and this process has to be $\tilde{X}$. $\square$

We can conclude immediately from Proposition 5.2 that in the general locally compact case any limit process has the strong Markov property, at least up to the first time it hits the boundary at infinity.

The following example shows that in general we loose the strong Markov property once we hit infinity.

Example 5.5 (Entrance law)**.** *Let (T, r, ρ) be the discrete binary tree with unit edge-lengths, i.e.,*

$$T := \bigcup_{n\in\mathbb{N}} \{0,1\}^n \cup \{\rho\}, \tag{5.14}$$

$r(\rho, x) := n$ for all $x \in \{0,1\}^n$, and there is an edge $x \sim y$ if and only if $y = (x, i)$ or $x = (y, i)$ for $i \in \{0, 1\}$.

Put $h(x) := r(\rho, x)$, and consider the speed measure $\nu(\{x\}) := e^{-h(x)}$, $x \in T$. Obviously, the speed-ν random walk on X is transient, as $h(X)$ is a reflected random walk on $\mathbb{N}$ with constant drift to the right.

Now consider (T_n, r, ρ, ν) with $T_n := \{x \in T : h(x) \le n\}$, where the metric and the measure are understood to be restricted to T_n. Because T_n is finite and the speed-ν_n random walk X^n has no absorbing points, it is positive recurrent. We may therefore conclude from Proposition 2.12 that for all $x \in T$, $n \in \mathbb{N}$ suitably large,

$$\mathbb{E}^x\big[\tau_\rho^n\big] = 2 \sum_{y\in T_n} h\big(c(\rho, x, y)\big) e^{-h(y)} \le \sum_{k=1}^n k 2^k e^{-k} < \infty. \tag{5.15}$$

Therefore, in contrast to the transience of the speed-ν random walk on (T, r), any "limiting" process Y of the speed-ν_n random walks on (T_n, r_n) is also positive recurrent. This shows that in Theorem 1 we indeed have to stop limiting processes at infinity in order for them to coincide with the speed-ν motion on (T, r). Consequently, this also means that the speed-ν motion has an entrance law on (T, r) from infinity, which we obtain by considering excursions of Y away from infinity. Finally, the limit Y obviously looses its strong Markov property at hitting infinity, because, in the one-point compactification, we are identifying all ends at infinity. $\square$

5.2. **The occupation time formula of the limit.** In this section we assume that the limiting tree is compact and show that all limit points satisfy the occupation time formula (0.1). The main result is the following:

Proposition 5.6 (Occupation time formula)**.** *Let $\mathcal{X} := (T, r, \nu)$ and $\mathcal{X}_n := (T_n, r_n, \nu_n)$, $n \in \mathbb{N}$, be metric boundedly finite measure trees. Assume that all $\mathcal{X}_n$, $n \in \mathbb{N}$, are discrete with $\sup_{n\in\mathbb{N}} \operatorname{diam}(T_n, r_n) < \infty$, and that $\mathcal{X}_n \to \mathcal{X}$ Gromov-Hausdorff-vaguely as $n \to \infty$. If X^n is the speed-ν_n random walk on (T_n, r_n) and $X^n \underset{n\to\infty}{\Longrightarrow} \tilde{X}$ in path space, then $\tilde{X}$ satisfies (0.1).*

To prove this formula, we need a lemma about semi-continuity of hitting times in Skorohod space. This semi-continuity does not hold in general, but we rather have to use that the limiting path satisfies a certain regularity property.

If $\operatorname{supp}(\nu)$ is not connected, the paths of the limit process are obviously not continuous. They satisfy, however, the following weaker *closedness condition.*

Definition 5.7 (Closed-interval property)**.** *Let E be a topological space. We say that a function $w\colon \mathbb{R}_+ \to E$ has the* closed-interval property *if $w\big([s,t]\big) \subseteq E$ is closed for all $0 \le s < t$.*

Lemma 5.8 (Speed-ν motions have the closed-interval property)**.** *The path of the limit process $\tilde{X}$ has the closed-interval property, almost surely.*

Proof. Let $A \subseteq T$ be the set of endpoints of edges of T. Recall from Remark 1.2 that A is at most countable. Jumps of the limit process $\tilde{X}$ can only occur over edges of T, hence $\tilde{X}_{t-} := \lim_{s\nearrow t} \tilde{X}_s \neq \tilde{X}_t$ implies $\tilde{X}_{t-} \in A$.

Fix $a \in A$. We first show that if $\tau_a^- := \inf\{t > 0 : \tilde{X}_{t-} = a\}$ denotes the first time when the left limit of $\tilde{X}$ reaches a, we have $\tilde{X}_{\tau_a^-} = a$ almost surely, i.e., $\tilde{X}$ does not jump at time τ_a^- almost surely. Indeed, for every $\varepsilon > 0$ we can use the right-continuity of the paths of $\tilde{X}$ together with Feller-continuity to find $s_0 > 0$ and $\delta > 0$ such that for all $x \in B(a, \delta)$,

$$\mathbb{P}^x\Big\{ \sup_{s\in[0,s_0]} r(a, \tilde{X}_s) > \varepsilon \Big\} < \tfrac{1}{2}\varepsilon. \tag{5.16}$$

Define the stopping times $\tau_n := \inf\big\{t \ge 0 : r(\tilde{X}_t, a) \le \frac{1}{n}\big\}$, and note that $\tau_n \uparrow \tau_a^-$. If $n > \frac{1}{\delta}$ is such that $\mathbb{P}^x\{\tau_a^- - \tau_n > s_0\} < \frac{1}{2}\varepsilon$, then by Proposition 5.2,

$$\mathbb{P}^x\big\{ r(\tilde{X}_{\tau_a^-}, a) > \varepsilon \big\} \le \tfrac{1}{2}\varepsilon + \mathbb{E}^x\Big[\mathbb{P}^{\tilde{X}_{\tau_n}}\Big\{ \sup_{s\in[0,s_0]} r(a, \tilde{X}_s) > \varepsilon \Big\}\Big] \le \varepsilon. \tag{5.17}$$

Since ε is arbitrary, this proves $\tilde{X}_{\tau_a^-} = a$ almost surely.

Because A is countable, this implies that $\big\{\tilde{X}_u : u \in [0,t]\big\}$ is closed for all $t \ge 0$, almost surely. Again using the Markov property, we also obtain almost surely closedness of $\big\{\tilde{X}_u : u \in [s,t]\big\}$ for all $t \ge 0$, $s \in \mathbb{Q}_+$, which implies closedness for all $s \ge 0$ by right-continuity. □

We omit the proof of the following lemma, because it is straight-forward.

Lemma 5.9 (Semi-continuity of the hitting time functional)**.** *Let E be a Polish space and $\mathcal{D}_E = \mathcal{D}_E(\mathbb{R}_+)$ the corresponding Skorohod space. For a set $A \subseteq E$, define*

$$\sigma_A \colon \mathcal{D}_E \to \mathbb{R}_+ \cup \{\infty\}, \quad w \mapsto \inf\big\{ t \in \mathbb{R}_+ : w(t) \in A \big\}. \tag{5.18}$$

Then if A is open, σ_A is upper semi-continuous, and if A is closed, the set of lower semi-continuity points of σ_A contains the set of paths with the closed-interval property.

Remark 5.10. *For $A \subseteq E$ closed, σ_A is in general not lower semi-continuous.* □

Proof of Proposition 5.6. Fix $x, y \in T$ and let τ_y be the first time when $\tilde{X}$ hits y. It is enough to show (0.1) for non-negative $f \in \mathcal{C}_b(T)$. Because T is closed in E, we can extend f to a bounded continuous function on E, again denoted by f. For $A \subseteq E$, recall the definition of σ_A from (5.18) and consider the function

$$F_A\colon \mathcal{D}_E \to \mathbb{R}_+ \cup \{\infty\}, \quad w \mapsto \int_0^{\sigma_A(w)} f\big(w(t)\big)\,\mathrm{d}t. \tag{5.19}$$

Note that the left-hand side of (0.1) coincides with $\mathbb{E}^x[F_y(\tilde{X})]$, where we abbreviate $F_y := F_{\{y\}}$. The strategy is to approximate F_y by F_A for small neighbourhoods A of y and then use semi-continuity properties of F_A and the occupation time formula of the approximating X^n.

Denote for each $\varepsilon > 0$ the closed ε-ball in E around y by A_ε. We claim that almost surely

$$\tau := \sup_{\varepsilon>0} \sigma_{A_\varepsilon}(\tilde{X}) = \sigma_{\{y\}}(\tilde{X}) = \tau_y. \tag{5.20}$$

Indeed, $\tau \le \tau_y$ is obvious. For the converse inequality, recall that the path of $\tilde{X}$ almost surely has the closed-interval property by Lemma 5.8, which means that $\big\{\tilde{X}_t : t \in [0,\tau]\big\}$ is almost surely a closed set containing points in every A_ε, $\varepsilon > 0$, hence also y. Therefore $\tau_y \le \tau$ almost surely.

Because f is non-negative, (5.20) implies that

$$\sup_{\varepsilon>0} F_{A_\varepsilon}(\tilde{X}) = F_y(\tilde{X}), \tag{5.21}$$

almost surely. Furthermore, it follows from the definition of the Skorohod topology that whenever w is a lower- or upper semi-continuity point of σ_A, the same is true for F_A. Hence Lemma 5.9 together with Lemma 5.8 implies that the path of $\tilde{X}$ is almost surely a lower semi-continuity point of F_A for closed sets A, and an upper semi-continuity point for open sets A.

Choose $x_n, y_n \in T_n$ with $y_n \to y$ and $x_n \to x$, and note that $y_n \in A_\varepsilon$ for all sufficiently large n. Since $X^n \underset{n\to\infty}{\Longrightarrow} \tilde{X}$, and $\tilde{X}$ is almost surely a lower semi-continuity point of F_A,

$$\begin{aligned} \mathbb{E}^x\big[F_y(\tilde{X})\big] &= \sup_{\varepsilon>0} \mathbb{E}^x\big[F_{A_\varepsilon}(\tilde{X})\big] \\ &\le \sup_{\varepsilon>0} \liminf_{n\to\infty} \mathbb{E}^{x_n}\big[F_{A_\varepsilon}(X^n)\big] \\ &\le \liminf_{n\to\infty} \mathbb{E}^{x_n}\big[F_{y_n}(X^n)\big]. \end{aligned} \tag{5.22}$$

Note that the functions $(x_n, y_n, z_n) \mapsto 2r_n\big(y_n, c_n(x_n, y_n, z_n)\big)$ on T_n^3 and the corresponding function $(x,y,z) \mapsto 2r\big(y, c(x,y,z)\big)$ on T^3 have a common Lipschitz continuous extension to E given by

$$\xi(x,y,z) := d(y,x) + d(y,z) - d(z,x). \tag{5.23}$$

Therefore, we obtain from (5.22) and the occupation time formula for X^n (Proposition 2.12) that

$$\begin{aligned} \mathbb{E}^x\big[F_y(\tilde{X})\big] &\le \liminf_{n\to\infty} \int \nu_n(\mathrm{d}z)\, \xi(x_n, y_n, z) f(z) \\ &= 2\int \nu(\mathrm{d}z)\, r\big(y, c(x,y,z)\big) f(z). \end{aligned} \tag{5.24}$$

On the other hand, for every sufficiently small $\varepsilon > 0$ and large $n \in \mathbb{N}$, there is a unique point $y_n' \in B(y_n, 2\varepsilon) \cap T_n$ closest to x_n, and using that $\tilde{X}$ is almost surely an upper semi-continuity point of $F_{B(y,\varepsilon)}$, we obtain

$$\begin{aligned} \mathbb{E}^x\big[F_y(\tilde{X})\big] &\ge \limsup_{n\to\infty} \mathbb{E}^{x_n}\big[F_{B(y,\varepsilon)}(X^n)\big] \\ &\ge \limsup_{n\to\infty} \mathbb{E}^{x_n}\big[F_{B(y_n,2\varepsilon)}(X^n)\big] \\ &= \limsup_{n\to\infty} \mathbb{E}^{x_n}\big[F_{y_n'}(X^n)\big] \\ &\ge 2\int \nu(\mathrm{d}z)\, \big(r(y, c(x,y,z)) - 2\varepsilon\big) f(z). \end{aligned} \tag{5.25}$$

The claim follows with $\varepsilon \to 0$. $\square$

6. Proof of Theorem 1

In this section, we collect all the pieces we have proven so far and present the proof of our invariance principle.

As we have stated all the results which characterize the limiting process for approximating rooted metric measure trees $(T_n, r_n, \rho_n, \nu_n)$ where (T_n, r_n) was assumed to be discrete, we start with a lemma which states that each rooted metric boundedly finite measure tree can be approximated by discrete trees.

Lemma 6.1 (Approximation by discrete trees)**.** *Let (T, r, ρ, ν) be a rooted metric boundedly finite measure tree $\mathcal{X}$. Then we can find a sequence $\mathcal{X}_n := (T_n, r_n, \rho, \nu_n)$ of rooted discrete metric boundedly finite measure trees such that $\mathcal{X}_n \to \mathcal{X}$ pointed Gromov-Hausdorff-vaguely.*

Proof. Let (T, r, ρ, ν) be a rooted metric boundedly finite measure tree, and for each $n \in \mathbb{N}$, S_n a finite $\frac{1}{n}$-net of $B(\rho, n)$ containing $\{\rho\}$. Let $T_n \subseteq T$ be the smallest metric tree containing S_n, i.e. the union of S_n and all branching points $x \in T$ with

$$r(x, s_1) = \tfrac{1}{2}\big(r(s_1, s_2) + r(s_1, s_3) - r(s_2, s_3)\big), \tag{6.1}$$

for some $s_1, s_2, s_3 \in S_n$. As usual, let r_n be the restriction of r to T_n, and note that T_n is a finite set, hence (T_n, r_n) is a discrete metric tree.

Consider for each $n \in \mathbb{N}$ the map $\psi_n : T \to T_n$ which sends a point in T to the nearest point on the way from x to ρ which belongs to T_n, i.e.,

$$\psi_n(x) := \sup\big\{y \in T_n : y \in [\rho, x]\big\}. \tag{6.2}$$

Finally, put

$$\nu_n := \big(\psi_n\big)_* \nu{\restriction}_{B(\rho,n)}. \tag{6.3}$$

Then, obviously, the Prohorov distance between $\nu{\restriction}_{B(\rho,n)}$ and ν_n is not larger than $\frac{1}{n}$. Thus (T_n, r_n, ρ, ν_n) converges pointed Gromov-vaguely and also pointed Gromov-Hausdorff-vaguely to (T, r, ρ, ν). $\square$

6.1. Compact limit trees.

In this subsection we restrict to the case where the limiting tree is compact. We start with the proof of Proposition 5.1, on which we shall rely the characterization of the limit process.

Proof of Proposition 5.1. Consider X and Y satisfying the assumption on Proposition 5.1. In particular, assume that $X_{\cdot \wedge \tau_y}$ is transient for all $y \in T$. Consider for each $y \in T$ the family of resolvent operators $\{G_\alpha^{X,y};\, \alpha > 0\}$ and $\{G_\alpha^{Y,y}\ \alpha > 0\}$ associated with $\{X_{\cdot \wedge \tau_y};\, y \in T\}$ and $Y_{\cdot \wedge \tau_y}$, and put $G_X^y := \lim_{N\to\infty} G_{1/N}^{X,y}$ and $G_Y^y := \lim_{N\to\infty} G_{1/N}^{Y,y}$, respectively. By transience, $G_X^y < \infty$ for all $y \in T$. Moreover, for all $x \in T$, and bounded, measurable $f\colon T \to \mathbb{R}_+$,

$$G_X^y f(x) = \mathbb{E}^x\Big[\int_0^{\tau_y} \mathrm{d}s\, f(X_s)\Big]. \tag{6.4}$$

By (5.1), $G_Y^y f(x) < \infty$ as well.

As X is a strong Markov processes, the resolvent identity holds, i.e.,

$$G_\alpha^{X,y} = G_\beta^{X,y} + (\alpha - \beta) G_\alpha^{X,y} G_\beta^{X,y}. \tag{6.5}$$

Iterating the latter with $\alpha > \beta > 0$ and $|\alpha - \beta| \leq \frac{1}{2\|G_\beta^{X,y}\|}$, we have

$$G_\alpha^{X,y} = G_\beta^{X,y} + (\alpha - \beta)\big(G_\beta^{X,y}\big)^2 + (\alpha - \beta)^2\big(G_\beta^{X,y}\big)^3 + \cdots \tag{6.6}$$

We note that $\|G_\beta^{X,y}\| \le \|G^{X,y}\|$ for all $\beta \ge 0$. So it is bounded above independent of β. Hence (6.6) holds for $\beta = 0$ by taking limits. Further, by the same arguments, (6.6) also holds for Y instead of X, and by (5.1) $G_0^{Y,y} := G_Y^y = G_X^y$. Therefore, for all small enough $\alpha > 0$, $G_\alpha^{X,y} = G_\alpha^{Y,y}$. Thus for all small enough $\alpha > 0$,

$$\mathbb{E}^x\Big[\int_0^{\tau_y} \mathrm{d}t\, e^{-\alpha t}\cdot f(X_t)\Big] = \mathbb{E}^x\Big[\int_0^{\tau_y} \mathrm{d}t\, e^{-\alpha t}\cdot f(Y_t)\Big]. \tag{6.7}$$

Therefore by uniqueness of the Laplace transform,

$$\mathbb{E}^x\big[f(X_t);\,\{t<\tau_y\}\big] = \mathbb{E}^x\big[f(Y_t);\,\{t<\tau_y\}\big] \tag{6.8}$$

for all $y \in T$ and for all $t > 0$. Therefore the one dimensional distributions of $X_{\cdot\wedge\tau_y}$ and $Y_{\cdot\wedge\tau_y}$ are the same for all $y \in T$. By the strong Markov property, this implies that the laws of X and Y agree. □

To show f.d.d. convergence, we need to control the probability that X_t is in an "exceptional" set of small ν-measure. To this end, we use the following simple heat-kernel bound. We will see in Corollary 6.4 below that the technical assumption $\nu(\{x\}) > 0$ can be dropped.

Lemma 6.2. *Let $\mathcal{X} := (T, r, \nu)$ be a compact metric finite measure tree, $x \in T$ with $\nu(\{x\}) > 0$, and X the speed-ν motion on (T, r) started in x. Then the law of X_t has for every $t > 0$ a density $q_t(x,\cdot) \in L^2(\nu)$ w.r.t. ν, and*

$$\big\|q_t(x,\cdot)\big\|_2^2 \le \nu(T)^{-1} + \mathrm{diam}(T)\cdot t^{-1} \qquad \forall t>0, \tag{6.9}$$

where $\|\cdot\|_2$ is the norm in $L^2(\nu)$. In particular, for any $A \subseteq T$, we have

$$\mathbb{P}^x\{X_t \in A\} \le \gamma_t\sqrt{\nu(A)} \qquad \forall t>0, \tag{6.10}$$

where the constant $\gamma_t := 1 + \nu(T)^{-1} + \mathrm{diam}(T)\cdot t^{-1}$ is independent of x and depends on (T, r, ν) only through $\nu(T)$ and $\mathrm{diam}(T)$.

Proof. 1. Let $f := \nu(\{x\})^{-1}\mathbf{1}_{\{x\}}$ be the density of δ_x w.r.t. ν, and

$$f_t := P_t f, \qquad g(t) := \|f_t\|_2^2, \tag{6.11}$$

where $(P_t)_{t\ge0}$ is the semi-group of the speed-ν motion. Due to reversibility of ν it is easy to see that $f_t = q_t(x,\cdot)$ is the density of X_t w.r.t. ν. Furthermore,

$$g'(t) = 2\langle Gf_t, f_t\rangle_\nu = -2\mathcal{E}(f_t, f_t), \tag{6.12}$$

where G is the generator of $(P_t)_{t\ge0}$. Let $a := \mathrm{diam}(T)^{-1}$. Because $\|f_t\|_1 = 1$, we find a point $y \in T$ with $f_t(y) \le b := \nu(T)^{-1}$. For every $z \in T$ with $f_t(z) \ge b$, we have

$$\mathcal{E}(f_t, f_t) \ge \big(f_t(z) - f_t(y)\big)^2\cdot\big(2r(z,y)\big)^{-1} \ge \tfrac12 a\big(f_t(z)-b\big)^2. \tag{6.13}$$

Combining (6.13) and (6.12), and using $g(t) = \|f_t\|_2^2 \le \|f_t\|_\infty\|f_t\|_1 = \|f_t\|_\infty$, we obtain the differential inequality

$$g'(t) \le -a(\|f_t\|_\infty - b)^2 \le -a\big(g(t)-b\big)^2. \tag{6.14}$$

In the above, we have used that $g(t) \ge b$. Solving $h_u'(t) = -a(h_u(t)-b)^2$, $h_u(0) = u$, and using monotonicity of the solution in u, we conclude

$$g(t) \le \lim_{u\to\infty} h_u(t) = \lim_{u\to\infty}\frac{u(1+abt) - ab^2t}{uat - bat + 1} = b + (at)^{-1}, \tag{6.15}$$

which is the desired bound (6.9).

2. For $u := \nu(A)^{-1/2}$ we obtain

$$\mathbb{P}^x\{X_t \in A\} \le u\,\nu(A) + \int_{\{f_t>u\}}\frac{f_t^2}{u}\,\mathrm{d}\nu \le \sqrt{\nu(A)}\big(1 + \|f_t\|_2^2\big). \tag{6.16}$$

Together with (6.9) this implies the desired bound (6.10). □

Proposition 6.3 (Theorem 1 for compact limit trees)**.** *Let $\mathcal{X} = (T, r, \rho, \nu)$, $\mathcal{X}_1 = (T_1, r_1, \rho_1, \nu_1)$, $\mathcal{X}_2 = (T_2, r_2, \rho_2, \nu_2), \dots$ be rooted metric boundedly finite measure trees with $\sup_{n\in\mathbb{N}} \mathrm{diam}(T_n, r_n) < \infty$. Let X be the speed-ν motion on (T, r) starting in ρ, and for all $n \in \mathbb{N}$, X^n the speed-ν_n motion on (T_n, r_n) started in ρ_n. Assume that the following conditions hold:*

- (A1) *The sequence $(\mathcal{X}_n)_{n\in\mathbb{N}}$ converges to $\mathcal{X}$ pointed Gromov-vaguely.*
- (A2) *The uniform local lower mass-bound property (1.6) holds.*

Then the following hold:

- (i) *X^n converges weakly in path space to X.*
- (ii) *If we assume only (A1) but not (A2), then X^n converges in finite dimensional distributions to X.*

Proof. Assume w.l.o.g. that $(\mathcal{X}_n)_{n\in\mathbb{N}}$ are discrete trees (the general result is then obtained by Lemma 6.1 and a diagonal argument). Let X^n be a sequence of ν_n-random walks on (T_n, r_n) starting in ρ_n.

(i) By Proposition 4.1 we know that the sequence is tight. Let $\tilde{X}$ be a weak subsequential limit on (T, r). Then in particular, $\tilde{X}_0 = \rho$ almost surely. From Proposition 5.2 together with Proposition 5.6, we know that $\tilde{X}$ is a strong Markov process and we have $\mathbb{E}^x[\int_0^{\tau_z} \mathrm{d}s\, f(\tilde{X}_s)] = 2\int \nu(\mathrm{d}y)\, r(z, c(x, y, z)) f(y)$.

Let X be the speed-ν motion on (T, r) starting in ρ. Then X is the strong Markov process associated with the Dirichlet form $(\mathcal{E}, \mathcal{D}(\mathcal{E}))$. X is recurrent as clearly $\mathbf{1} \in \mathcal{D}(\mathcal{E})$ and $\mathcal{E}(\mathbf{1}, \mathbf{1}) = 0$. Thus X satisfies (0.1) by Proposition 2.12. Moreover, it follows from Lemma 2.6 that $X_{\cdot\wedge\tau_y}$ is transient for all $y \in T$. Therefore the laws of $\tilde{X}$ and X agree by Proposition 5.1.

(ii) Using that $\mathcal{X}_n$ converges Gromov-weakly to $\mathcal{X}$, and $\mathcal{X}$ is compact, we can construct subsets $A_n \subseteq T_n$ with $\nu_n(A_n) \to 0$, $\rho_n \notin A_n$ and the following property. The measure trees $\tilde{\mathcal{X}}_n := (\tilde{T}_n, r_n, \rho_n, \nu_n)$, where $\tilde{T}_n := T_n \setminus A_n$, satisfy the lower mass-bound (1.6) and still converge Gromov-weakly to $\mathcal{X}$. Let $\tilde{X}^n$ be the ν_n-random walk on $(\tilde{T}_n, r_n)$. Then $\tilde{X}^n$ converges in distribution to X by part (i). We show that every finite-dimensional marginal of $\tilde{X}^n$ is weakly merging with the corresponding marginal of X^n. For this it is enough to show for all $t \geq 0$ the uniform merging of one-dimensional marginals, i.e.

$$\lim_{n\to\infty} \sup_{x\in\tilde{T}_n} d_{\mathrm{Pr}}^{(T_n,r_n)}\big(\mathcal{L}^x(X_t^n),\, \mathcal{L}^x(\tilde{X}_t^n)\big) = 0, \tag{6.17}$$

where $d_{\mathrm{Pr}}^{(T_n,r_n)}$ is the Prohorov metric associated to r_n. The finite-dimensional statement then follows from the Markov property of the speed-ν motions together with the Feller continuity of the limiting process (proven in Proposition 5.2).

Recall that (T_n, r_n) is discrete and thus $\nu_n(\{x\}) > 0$ for all $x \in T_n$. Using Lemma 6.2, and the fact that $\mathrm{diam}(T_n)$ and $\nu_n(T_n)^{-1}$ are bounded uniformly in n, we obtain $\gamma_t > 0$, independent of n, such that

$$\sup_{x\in T_n} \mathbb{P}^x\{X_t^n \in A_n\} \leq \gamma_t \sqrt{\nu_n(A_n)}. \tag{6.18}$$

We can couple X^n and $\tilde{X}^n$ by a time transformation such that $\tilde{X}_t^n = X_{L_n^{-1}(t)}^n$, where $L_n^{-1}(t) = \inf\{s \geq 0 : \int_0^s \mathbf{1}_{\tilde{T}_n}(X_u^n)\,\mathrm{d}u > t\}$. For (6.17) it is enough to show for every fixed $t, \varepsilon > 0$ that

$$\sup_{x\in\tilde{T}_n} \mathbb{P}^x\big\{r_n(X_t^n, \tilde{X}_t^n) > \varepsilon\big\} \leq 4\varepsilon, \tag{6.19}$$

for all sufficiently large $n \in \mathbb{N}$. The idea is that X_t^n and $\tilde{X}_t^n$ do not differ too much, because $\tilde{X}_t^n$ cannot move far in a short amount of time and will be ahead of X_t^n only a small amount of time, controlled via the occupation time formula by the (small) ν_n-measure of $A_n = T_n \setminus \tilde{T}_n$.

Because $\tilde{\mathcal{X}}_n$ converges Gromov-Hausdorff weakly, we can use the speed bound, Lemma 4.7, to find $c > 0$ such that the probability that $\tilde{X}^n$ moves ε within time c is bounded by ε, i.e.,

$$\sup_{x\in\tilde{T}_n} \mathbb{P}^x\big\{ \sup_{s\in[0,c]} r_n(\tilde{X}^n_s, x) > \varepsilon \big\} \le \varepsilon. \tag{6.20}$$

In order to use the occupation time formula, we fix two points $y_n, z_n \in \tilde{T}_n$ with $r_n(y_n, z_n) > \varepsilon$ and define recursively the times where X^n hits y_n and z_n in alternation, i.e. $\tau_n^0 := 0$, $\tau_n^k := \inf\{t > \tau_n^{k-1} : X_t^n = y_n\}$ for k odd and $\tau_n^k := \inf\{t > \tau_n^{k-1} : X^n_t = z_n\}$ for k even. Let $\tilde{\tau}_n^k$, $k \in \mathbb{N}$, be the analogous stopping times for $\tilde{X}^n$ instead of X^n. Because the lower bound for the distance of y_n and z_n is independent of n, we can use Lemma 4.7 again to find $k \in \mathbb{N}$, independent of n, such that $\mathbb{P}\{\tilde{\tau}^k_n < t\} < \varepsilon$. Because $\tau_n^k \ge \tilde{\tau}_n^k$, we also obtain

$$\sup_{x\in\tilde{T}_n} \mathbb{P}^x\{\tau_n^k < t\} < \varepsilon. \tag{6.21}$$

Now consider the accumulated time difference between X^n and $\tilde{X}^n$ until τ_n^k, i.e.,

$$\delta_n := \int_0^{\tau_n^k} \mathbf{1}_{A_n}(X^n_t)\,\mathrm{d}t. \tag{6.22}$$

Then, by the occupation time formula,

$$\sup_{x\in\tilde{T}_n} \mathbb{E}^x[\delta_n] \le k \cdot 2\,\mathrm{diam}(T_n)\nu_n(A_n). \tag{6.23}$$

The right-hand side tends to zero as n tends to infinity, because $\mathrm{diam}(T_n)$ is uniformly bounded by assumption and k is independent of n. Therefore, for sufficiently large n depending on c chosen in (6.20),

$$\sup_{x\in\tilde{T}_n} \mathbb{P}^x\{\delta_n > c\} < \varepsilon. \tag{6.24}$$

On the event $\{X_t^n \notin A_n\}$, we have $X^n_t = \tilde{X}^n_{L_n(t)}$, and on the event $\{\tau_n^k \ge t\}$, we have $t - L_n(t) < \delta_n$. Hence, using (6.18) and (6.21), we obtain for all $x \in \tilde{T}_n$,

$$\begin{aligned}
&\mathbb{P}^x\big\{r_n(X^n_t, \tilde{X}^n_t) > \varepsilon\big\} \\
&\le \mathbb{P}^x\{X^n_t \in A_n\} + \mathbb{P}^x\{\tau_n^k < t\} + \mathbb{P}^x\big\{t - L_n(t) < \delta_n,\ r_n(\tilde{X}^n_{L_n(t)}, \tilde{X}^n_t) > \varepsilon\big\} \\
&\le \gamma_t\sqrt{\nu_n(A_n)} + \varepsilon + \mathbb{P}^x\{\delta_n > c\} + \mathbb{P}^x\big\{\sup_{s\in[t-c,t]} r_n(\tilde{X}^n_s, \tilde{X}^n_t) > \varepsilon\big\},
\end{aligned} \tag{6.25}$$

which is bounded by 4ε for large n due to $\nu_n(A_n) \to 0$, (6.24) and (6.20) together with the Markov property of $\tilde{X}^n$. This proves (6.19) and hence the claimed f.d.d. convergence. □

Corollary 6.4 (pointwise L^2-heat-kernel bound)**.** *Lemma 6.2 remains correct if we drop the assumption $\nu(\{x\}) > 0$. In particular, for every compact metric finite measure tree $\mathcal{X} := (T, r, \nu)$, the following bound on the $L^2(\nu)$-norm of the heat-kernel q_t (defined in Lemma 6.2) holds:*

$$\big\|q_t(x,\cdot)\big\|_2^2 \le \nu(T)^{-1} + \mathrm{diam}(T)\cdot t^{-1} \qquad \forall x \in T,\ t > 0. \tag{6.26}$$

Proof. Fix $x \in T$, $t > 0$, and let $\nu_n := \nu + \frac{1}{n}\delta_x$. Let X^n and X be the speed-ν_n and speed-ν motion on (T, r), respectively, all started in x. According to Proposition 6.3 for $\mathcal{X}_n := (T_n, r_n, \rho_n, \nu_n) := (T, r, x, \nu + \frac{1}{n}\delta_x)$, the law $\mu_{n,t}$ of X^n_t converges weakly to the law μ_t of X_t. According to Lemma 6.2, there is $f_{n,t} \in L^2(\nu)$ with $\mu_{n,t} = f_{n,t}\cdot\nu$, and $\|f_{n,t}\|_2$ is bounded uniformly in n. Therefore, the weak limit μ_t also admits a density with the same bound on its $L^2(\nu)$-norm. □

We conclude this subsection with examples showing how the violation of the tightness condition (A2) destroys convergence in path space, while f.d.d. convergence still holds.

Example 6.5 (f.d.d. convergence but not path-wise)**.** *Let $r, r_1, r_2, \dots$ be the Euclidean metric on $[0,1]$.*

(i) *Let $T_n = \{0,1\}$, and $\nu_n = \delta_0 + \frac{1}{n}\delta_1$ for $n \in \mathbb{N}$. Then $\mathcal{X}_n := (T_n, r, 0, \nu_n)$ converges pointed Gromov-vaguely to $\mathcal{X} := (\{0\}, r, 0, \delta_0)$. The speed-$\nu_n$ motion X^n is a two-state Markov chain that jumps from 0 to 1 at rate $\frac{1}{2}$ and from 1 to 0 at rate $\frac{n}{2}$. It obviously converges f.d.d. to the constant process, but not in path space.*

(ii) *Let $T_n = [0,1]$, and $\nu_n = \delta_0 + \delta_1 + \frac{1}{n}\lambda_{[0,1]}$, where $\lambda_{[0,1]}$ is Lebesgue measure on $[0,1]$. Then $(T_n, r, 0, \nu_n)$ converges pointed Gromov-vaguely to $\big(\{0,1\}, r, 0, \nu\big)$ with $\nu = \delta_0 + \delta_1$. The speed-$\nu$ motion X is the symmetric Markov chain on $\{0,1\}$ with jump-rate $\frac{1}{2}$, and the speed-ν_n motions X^n are sticky Brownian motions on $[0,1]$ with diverging speed on $(0,1)$, as n tends to ∞. As X^n has continuous paths for each $n \in \mathbb{N}$ but X has discontinuous paths, the convergence cannot be in path space. The finite dimensional distributions of X^n, however, converge to those of X, as the processes X^n spend less and less times in discontinuity points.* □

6.2. **From compact to locally compact limit trees.** In this subsection we extend the proof of Theorem 1 to locally compact trees equipped with boundedly finite speed measures. In order to reduce this to the compact case, we stop the processes upon reaching a height R. For that purpose we need the following lemma whose proof is straight-forward and will therefore be omitted.

Recall the closed interval property from Definition 5.7.

Lemma 6.6 (Continuity points)**.** *Let (E,d) be a Polish space, $\rho \in E$, and $R > 0$. Define the function*

$$\psi_R \colon \mathcal{D}_E \to \mathcal{D}_E, \quad \psi_R(w)(t) := w\big(t \wedge \inf\big\{s \,:\, d(\rho, w(s)) \geq R\big\}\big). \tag{6.27}$$

Assume that $w \in \mathcal{D}_E$ has the closed-interval property, and that the map $t \mapsto d\big(\rho, w(t)\big)$ does not have a local maximum at height R. Then w is a continuity point of ψ_R.

Proof of Theorem 1. (ii) has already been shown in Proposition 6.3.

(i) We call a point $v \in T$ *extremal leaf* of T if the height function $h \colon T \to \mathbb{R}_+$, $x \mapsto r(\rho, x)$ has a local maximum at v. Note that, although there can be uncountably many extremal leaves, the set of heights of extremal leaves is at most countable due to separability of T. Now choose $R_k > 0$, $k \in \mathbb{N}$, with $R_k \to \infty$ such that there is no extremal leaf of T at height R_k and $\nu\big\{x' \in T \,:\, r(\rho, x') = R_k\big\} = 0$.

Let X be the speed-ν motion on (T,r) started in ρ, and recall that $X = X_{\cdot\wedge\zeta}$, where $\zeta := \inf\{t \geq 0 \,:\, r(\rho, X_t) = \infty\}$. We show that the law of X coincides with the law of $\tilde{X}_{\cdot\wedge\zeta} := \psi_\infty(\tilde{X})$, where $\tilde{X}$ is any limit process. Using that there is no extremal leaf of T at height R_k and that $\tilde{X}$ and X have the closed-interval property, we obtain from Lemma 6.6 that (the paths of) $\tilde{X}$ and X are almost surely continuity points of ψ_{R_k}.

Let X_k^n be the speed-ν_n motion on the compact metric measure tree $T_n\!\restriction_{B(\rho_n, R_k)}$ and X_k the speed-ν motion on the compact metric measure tree $T\!\restriction_{B(\rho, R_k)}$. Then, for every $k \in \mathbb{N}$, $X_k^n \underset{n\to\infty}{\Longrightarrow} X_k$, as $n \to \infty$, by Proposition 6.3. Furthermore, for every k there is an $\ell = \ell_k$, such that the laws of $\psi_{R_k}(X^n)$ and $\psi_{R_k}(X_\ell^n)$ coincide; and the same is true for $\psi_{R_k}(X)$ and $\psi_{R_k}(X_\ell)$.

By continuity of ψ_{R_k} in $\tilde{X}$ and X, we obtain

$$\psi_{R_k}\big(X_\ell^n\big) \overset{\mathcal{L}}{=} \psi_{R_k}\big(X^n\big) \underset{n\to\infty}{\overset{\mathcal{L}}{\Longrightarrow}} \psi_{R_k}\big(\tilde{X}\big), \tag{6.28}$$

and on the other hand

$$\psi_{R_k}\big(X_\ell^n\big) \underset{n\to\infty}{\overset{\mathcal{L}}{\Longrightarrow}} \psi_{R_k}\big(X_\ell\big) \overset{\mathcal{L}}{=} \psi_{R_k}\big(X\big). \tag{6.29}$$

Hence $\psi_{R_k}(\tilde{X}) \overset{\mathcal{L}}{=} \psi_{R_k}(X)$ for all $k \in \mathbb{N}$, and therefore $\psi_\infty(\tilde{X}) \overset{\mathcal{L}}{=} \psi_\infty(X) = X$ as claimed. □

7. Examples and related work

We conclude the paper with a discussion on how our invariance principle relates to results from the existing literature. These results have often been proven via quite different techniques but they all follow in a unified way from Theorem 1.

In Subsection 7.1 we revisit [Sto63] which (including a killing part) proves the invariance principle in the particular situation when the underlying metric trees are closed subsets of $\mathbb{R}$, or equivalently, linear trees. In Subsection 7.2 we connect our invariance principle with the construction of diffusions on so-called dendrites, or equivalently, $\mathbb{R}$-trees, which is given in [Kig95]. We continue in Subsection 7.3 with [Cro10], where the classical convergence of rescaled simple random walks on $\mathbb{Z}$ to Brownian motion on $\mathbb{R}$ is generalized in a different direction than in [Sto63]. Namely, simple random walks on discrete trees with uniform edge-lengths are proven to converge to Brownian motion on a limiting rooted compact $\mathbb{R}$-tree which additionally has to satisfy some conditions. Finally, in Subsection 7.4 we consider the nearest neighbor random walk on a size-biased branching tree for which the suitably rescaled height process averaged over all realizations is tight according to [Kes86], while for almost every fixed realization it is not tight by [BK06].

7.1. **Invariance principle on $\mathbb{R}$.** In this subsection, we consider the special case of linear trees, i.e., closed subsets of $\mathbb{R}$.

Let ν, ν_n, $n \in \mathbb{N}$, be locally finite measures on $\mathbb{R}$, $T := \mathrm{supp}(\nu)$ and $T_n := \mathrm{supp}(\nu_n)$. Denote the Euclidean metric on $\mathbb{R}$ by r. Then $(T, r, 0, \nu)$ and $(T_n, r, 0, \nu_n)$ are obviously rooted metric boundedly finite measure trees in the sense of Definition 1.1. Also note that the speed-ν motion is conservative (i.e. does not hit infinity), because the tree (T, r) is recurrent (see, e.g., [AEW13, Theorem 4]). Now if ν_n converges vaguely to ν, and the uniform local lower mass-bound (1.6) holds, Theorem 1 implies that the speed-ν_n motions converge in path space to the speed-ν motion. This (essentially) is Theorem 1 (i) obtained in [Sto63] in the special case, where the killing measures are not present.

The methods used in [Sto63] are quite different from ours. In that paper all processes are represented as time-changes of standard Brownian motion and a jointly continuous version of local times is used.

Example 7.1 (Standard motion on disconnected sets)**.** *A particular instance of Stone's invariance principle was studied in detail in* [BEPR08]*. Put for each $q > 1$, $T_q := \{\pm q^k;\, k \in \mathbb{Z}\} \cup \{0\}$ and $\rho_q = 0$. Then $(T_q)_{q>1}$ converges, as $q \downarrow 1$, to $\mathbb{R}$ with respect to the localized Hausdorff distance. Recall the length measure from (2.2). Obviously, as the length measure is always boundedly finite on linear trees, the embedding which sends a rooted tree (T, ρ) with $T \subseteq \mathbb{R}$ to the measure tree $(T, \rho, \lambda^{(T,\rho)})$ is a homeomorphism onto its image. Thus $(T_q, 0, \lambda^{(T_q,0)})$ converges Hausdorff-vaguely to $(\mathbb{R}, 0, \lambda)$, as $q \downarrow 1$, where λ is the Lebesgue measure. It therefore follows that the speed-$\lambda^{(T_q,0)}$ motion on T_q converges in path space to the standard Brownian motion on $\mathbb{R}$ by Theorem 1. The latter is Proposition 5.1 in* [BEPR08]. $\square$

7.2. **Diffusions on dendrites.** In [Kig95] diffusions on dendrites (which are $\mathbb{R}$-trees) are constructed via approximating Dirichlet forms rather than processes. In this subsection we relate our invariance principle to this construction.

Let (T, r, ρ, ν) be a complete, locally compact, rooted boundedly finite measure $\mathbb{R}$-tree. Let furthermore $(T_m)_{m\in\mathbb{N}}$ be an increasing family of finite subsets of T. Put for all $f, g : T_m \to \mathbb{R}$

$$\mathcal{E}_m\big(f, g\big) := \tfrac{1}{2} \int_{T_m} \lambda^{(T_m, r_m, \rho)}(\mathrm{d}y)\, \nabla f(y) \nabla g(y). \tag{7.1}$$

Assume for each $m \in \mathbb{N}$ that T_m contains all the branch points of the subtree spanned by T_m (see our condition (1.2)). Then for all $m \le m'$, and for all $f : T_m \to \mathbb{R}$,

$$\mathcal{E}_m\big(f, f\big) = \min\big\{\mathcal{E}_{m'}\big(g, g\big) : \, g : T_{m'} \to \mathbb{R},\, g{\restriction}_{T_m} = f\big\}. \tag{7.2}$$

That is, the sequence $(T_m, \mathcal{E}_m)_{m\in\mathbb{N}}$ is compatible in the sense of Definition 0.2 (and the following paragraph) in [Kig95]. Assume further that $T^* := \cup_{m\in\mathbb{N}} T_m$ is dense in T, and consider the bilinear form

$$\mathcal{E}^{\mathrm{Kigami}}(f,g) := \lim_{m\to\infty} \mathcal{E}_m\big(f{\restriction}_{T_m}, g{\restriction}_{T_m}\big) \tag{7.3}$$

with domain

$$\mathcal{F}^{\mathrm{Kigami}} := \big\{f : T^* \to \mathbb{R} : \text{ limit on r.h.s. of (7.3) exists}\big\}. \tag{7.4}$$

Let $\mathcal{D}(\mathcal{E}^{\mathrm{Kigami}})$ be the completion of $\mathcal{F}^{\mathrm{Kigami}} \cap \mathcal{C}_c(T)$ with respect to the $\mathcal{E}^{\mathrm{Kigami}} + (\cdot,\cdot)_\nu$-norm. By Theorem 5.4 in [Kig95], $(\mathcal{E}^{\mathrm{Kigami}}, \bar{\mathcal{D}}(\mathcal{E}^{\mathrm{Kigami}}))$ is a regular Dirichlet form.

It was shown in Remark 3.1 in [AEW13] that the unique ν-symmetric strong Markov process associated with $(\mathcal{E}^{\mathrm{Kigami}}, \bar{\mathcal{D}}(\mathcal{E}^{\mathrm{Kigami}}))$ is the speed-ν motion on (T, r).

The bilinear form $\mathcal{E}^{\mathrm{Kigami}}$ describes the discrete time embedded Markov chains evaluated at T_n, $n \in \mathbb{N}$. The fact that it is a resistance form means that the projective limit diffusion is on "natural scale", which we additionally equip with speed measure ν. We can, of course, also approximate the speed-ν motion on (T, r) by continuous time Markov chains evaluated at T_n, $n \in \mathbb{N}$. Similar as in the proof of Lemma 6.1, consider for each $n \in \mathbb{N}$ the map $\psi_n : T \to T_n$ which sends a point in T to the nearest point on the way from x to ρ which belongs to T_n, i.e.,

$$\psi_n(x) := \sup\big\{y \in T_n : y \in [\rho, x]\big\}, \tag{7.5}$$

and equip T_n with

$$\nu_n := \big(\psi_n\big)_* \nu. \tag{7.6}$$

As T^* is dense, $(\nu_n)_{n\in\mathbb{N}}$ converges vaguely to ν, and thus $(T_n, r, \nu_n)_{n\in\mathbb{N}}$ converges Gromov-Hausdorff-vaguely to (T, r, ν). It therefore follows from our invariance principle that the continuous time Markov chains which jump from $v \in T_n$ to a neighboring $v \sim v'$ at rate $(2\nu_n(\{v\}) r(v, v'))^{-1}$ converges weakly in path space to the speed-ν motion on (T, r).

7.3. **Invariance principle with homogeneous rescaling.** In this subsection we relate our invariance principle to the one obtain earlier in [Cro10]. We first recall the excursion representation of a rooted compact measure $\mathbb{R}$-tree. We denote by

$$\mathcal{E} := \big\{e \colon [0,1] \to \mathbb{R}_+ \;\big|\; e \text{ is continuous, } e(0) = e(1) = 0\big\} \tag{7.7}$$

the set of continuous excursions on $[0,1]$. From each excursion $e \in \mathcal{E}$, we can define a measure $\mathbb{R}$-tree in the following way:

- $r_e(x,y) := e(x) + e(y) - 2\inf_{[x,y]} e$ is a pseudo-distance on $[0,1]$,
- $x, y \in [0,1]$ are said to be equivalent, $x \sim_e y$, if $r_e(x,y) = 0$,
- the image of the projection $\pi_e \colon [0,1] \to [0,1]/{\sim_e}$ endowed with the pushforward of r_e (again denoted r_e), i.e. $T_e := (T_e, r_e, \rho_e) := \big(\pi_e([0,1]), r_e, \pi_e(0)\big)$, is a rooted compact $\mathbb{R}$-tree.
- We endow this space with the probability measure $\mu_e := \pi_{e*}\lambda_{[0,1]}$ which is the pushforward of the Lebesgue measure on $[0,1]$.

We denote by $g : \mathcal{E} \to \mathbb{T}_c$ the resulting "glue function",

$$g(e) := \big(T_e, r_e, \rho_e, \mu_e\big), \tag{7.8}$$

which sends an excursion to a rooted probability measure $\mathbb{R}$-tree.

Recall $\mathbb{T}_c$ from (3.2). Given $\mathcal{X} := (T, r, \rho, \nu) \in \mathbb{T}_c$, we say that $\mathcal{X}$ satisfies a polynomial lower bound for the volume of balls, or short a *polynomial lower bound* if there is a $\kappa > 0$ such that

$$\liminf_{\delta\downarrow 0} \inf_{x\in T} \delta^{-\kappa} \nu\big(B_r(x,\delta)\big) > 0. \tag{7.9}$$

In [Cro10] the following subspace of $\mathbb{T}_c$ is considered:

$$\begin{aligned} \mathbb{T}^* := \big\{ \mathcal{X} = (T, r, \rho, \nu) \in \mathbb{T}_c : \\ &\text{(a) } \nu \text{ is non-atomic, (b) } \nu \text{ is supported on the leaves, and} \\ &\text{(c) } \nu \text{ satisfies a polynomial lower bound.} \big\} \end{aligned} \tag{7.10}$$

Let $((T_n, \rho_n))_{n\in\mathbb{N}}$, be a sequence of rooted graph trees with $\#T_n = n$, whose search-depth functions e_n in $\mathcal{E}$ with uniform topology satisfy

$$\tfrac{1}{a_n} e_n \xrightarrow[n\to\infty]{} e \tag{7.11}$$

for a sequence $(a_n)_{n\in\mathbb{N}}$ and some $e \in \mathcal{E}$ with $(T_e, r_e, 0, \mu_e) \in \mathbb{T}^*$. In Theorem 1.1 of [Cro10], it is shown that the discrete-time simple random walks on T_n starting in ρ_n with jump sizes rescaled by $1/a_n$ and speeded up by a factor of $n \cdot a_n$ converge to the μ_e-Brownian motion on T_e starting in 0.

To connect the above construction with Theorem 1 notice that the map g from (7.8) is continuous if $\mathbb{T}_c$ is endowed with the rooted Gromov-Hausdorff-weak topology, and $\mathcal{E}$ with the uniform topology (see [ADH14, Proposition 2.9]; compare also [Löh13, Theorem 4.8] for a generalization to lower semi-continuous excursions). Thus it follows from (7.11) that if we put $\nu_n := \mu_{a_n^{-1} e_n}$, then (T_n, ν_n) converges to (T_e, μ_e) rooted Gromov-Hausdorff-weakly. Analogously to Example 1.5 we obtain that $d_{\mathrm{Pr}}^{(T_n, r_n)}(\nu_n, \tilde{\nu}_n) \le a_n^{-1}$, where

$$\tilde{\nu}_n(\{v\}) := \tfrac{\deg(v)}{2n}, \tag{7.12}$$

and that thus also $(T_n, \tilde{\nu}_n)$ converges to (T_e, μ_e) rooted Gromov-Hausdorff-weakly by [ALW16, Lemma 2.10]. Theorem 1 then implies that unit rate simple random walks with edge lengths rescaled by a_n^{-1} and speeded up by $n \cdot a_n$ converge to the speed-μ_e motion on (T_e, r_e). As μ_e always has full support, the requirement that μ_e is supported on the leaves already implies that (T_e, r_e) is an $\mathbb{R}$-tree and thus the speed-μ_e motion on (T_e, r_e) has continuous paths.

Note that in contrast to [Cro10] our Theorem 1 does not require any additional assumptions on the limiting tree, which also does not have to be an $\mathbb{R}$-tree. The polynomial lower bound or that ν is non-atomic and supported on the leaves are not required. Also note that Theorem 1.1 of [Cro10] does only allow for homogeneous (non-state-dependent) rescaling. This means, for example, that in the particular case where the trees (T_n, r_n) are subsets of $\mathbb{R}$, only the case $T_n = a_n^{-1}\mathbb{Z} \cap [0, na_n^{-1}]$ and $\nu_n\big(\{x\}\big) = n^{-1}$, $x \in T_n$, is covered.

7.4. **Random walk on the size-biased branching tree.** Theorem 1 applies to trees that are complete and locally compact. The extension from compact to complete, locally compact trees is relatively straight forward. However this extension helps us to cover the random walk on the size-biased Galton-Watson tree studied in [Kes86] in the annealed regime and in [BK06] in the quenched regime. In this subsection we want to illuminate these results and put them in the context of our invariance principle.

Consider a random graph theoretical tree $\mathcal{T}_{\mathrm{Kesten}}$ which is distributed like the rooted Galton-Watson process with finite variance mean 1 offspring distribution conditioned to never die out. Let X be the (discrete-time) nearest neighbor random walk on $\mathcal{T}_{\mathrm{Kesten}}$ and d the graph distance on $\mathcal{T}_{\mathrm{Kesten}}$. Consider the rescaled height process

$$Z_t^{(n)} := n^{-\frac{1}{3}} \cdot d\big(\rho, X_{\lfloor nt \rfloor}\big), \qquad t \ge 0. \tag{7.13}$$

In [Kes86] it is shown that if $\tau_{B^c(\rho,N)} := \inf\big\{n \ge 0 : d(\rho, X_n) = N\big\}$, then for all $\varepsilon > 0$ there exists λ_1, λ_2 such that under the annealed law $\mathbb{P}^*$,

$$\mathbb{P}^*\big\{\lambda_1 \le N^{-3}\tau_{B^c(\rho,N)} \le \lambda_2\big\} \ge 1 - \varepsilon,$$

for all $N \geq 1$. Moreover, under $\mathbb{P}^*$, the process $Z^{(n)}$ converges weakly in path space to a non-trivial process Z with continuous paths.

In contrast to this annealed regime, in [BK06] (in the continuous time setting) it is shown that for almost all realizations of $\mathcal{T}_{\text{Kesten}}$, the family $\{Z^{(n)};\, n \in \mathbb{N}\}$ is not tight.

These two statements relate to our invariance principle as follows. Recall from (7.7) the space of continuous excursions on $[0,1]$ and from (7.8) the glue map g which sends an excursion $e \in \mathcal{E}$ to a rooted metric tree $([0,1]/{\sim_e}, r_e, 0)$ as well the map π_e which, given $e \in \mathcal{E}$, sends a point from the excursion interval $[0,1]$ to T_e. We can easily extend the maps g and π_e to the space

$$\mathcal{E}_\infty := \bigl\{e\colon \mathbb{R} \to \mathbb{R}_+ \bigm| e \text{ is continuous, } e(0) = 0,\ \lim_{x\to\pm\infty} e(x) = \infty\bigr\} \tag{7.14}$$

of continuous, two-sided, transient excursions on $\mathbb{R}$. To this end, we use the semimetric defined by

$$r_e(x,y) := \begin{cases} e(x) + e(y) - 2\inf_{z\in[x,y]} e(z), & xy \geq 0, \\ e(x) + e(y) - 2\inf_{z\in\mathbb{R}\setminus[x,y]} e(z), & xy < 0 \end{cases} \tag{7.15}$$

for $x \leq y$ (see [Duq09]). Then $g(e)$ is a rooted locally compact metric measure tree with a boundedly finite measure, for all $e \in \mathcal{E}_\infty$. It is not hard to show that the map g from (7.8) is continuous if $\mathbb{T}$ is endowed with the rooted Gromov-Hausdorff-vague topology, and $\mathcal{E}_\infty$ with the uniform topology on compact sets (see [ALW16, Proposition 7.5]).

In the particular case of a geometric offspring distribution, $\mathcal{T}_{\text{Kesten}}$ can be associated with the (two-sided) random excursion $\tilde{W}$, where for all $t \in \mathbb{R}$,

$$\tilde{W}_t := \begin{cases} W_t - 2\inf_{s\in[0,t]} W_s, & t \geq 0 \\ W_t - 2\inf_{s\in[t,0]} W_s, & t < 0, \end{cases} \tag{7.16}$$

with a simple two-sided random walk path $(W_n)_{n\in\mathbb{Z}}$, $W_0 = 0$, linearly interpolated. As W converges, after Brownian rescaling, weakly in path space towards (two-sided) standard Brownian motion $(B_t)_{t\in\mathbb{R}}$, we have

$$\bigl(n^{-1/3}\tilde{W}_{n^{2/3}t}\bigr)_{t\in\mathbb{R}} \underset{n\to\infty}{\Longrightarrow} \bigl(\tilde{B}_t\bigr)_{t\in\mathbb{R}}, \tag{7.17}$$

where $\tilde{B}_t := B_t - 2\inf_{s\in[0\wedge t,\, t\vee 0]} B_s$.

Given a realization e of $\tilde{W}$, define $e_n := n^{-1/3}e(n^{2/3}\boldsymbol{\cdot}) \in \mathcal{E}_\infty$ and denote by ν_n the rescaled degree measure on T_{e_n}, i.e., for all $A \subseteq T_{e_n}$,

$$\nu_n(A) := n^{-2/3}\sum_{v\in A} \tfrac{1}{2}\deg(v). \tag{7.18}$$

By Proposition 2.8 in [BK06], for almost all realizations e of $\tilde{W}$,

$$\liminf_{n\to\infty} \nu_n\bigl(B(\rho,R)\bigr) = 0, \quad \text{and} \quad \limsup_{n\to\infty} \nu_n\bigl(B(\rho,R)\bigr) = \infty, \tag{7.19}$$

and thus the sequence $\{\nu_n;\, n \in \mathbb{N}\}$ does not converge. Consider once more the map which sends all points of a half edge to its end point, and notice that the image measure of $\mu_{e_n} = (\pi_{e_n})_*\lambda_{\mathbb{R}_+}$ under this map equals ν_n. Thus the Prohorov distance between μ_{e_n} and ν_n is at most $n^{-1/3}$, and thus for almost all realizations e of $\tilde{W}$, also the sequence $\{\mu_{e_n};\, n \in \mathbb{N}\}$ does not converge. Hence the assumptions on our invariance principle fail for almost all realizations of $\mathcal{T}_{\text{Kesten}}$.

Notice that we can choose for each $n \in \mathbb{N}$ a realization e_n of $n^{-1/3}\tilde{W}_{n^{2/3}\cdot}$, and a realization e of $\tilde{B}$, such that $e_n \xrightarrow[n\to\infty]{} e$, almost surely. To understand why the quenched rescaling failed, notice that $e_n \xrightarrow[n\to\infty]{} e$ cannot be realized via a coupling such that all the e_n come from the same realization of $\tilde{W}$. As now $g(e_n)$ clearly converges to $g(e)$ by continuity of g, Theorem 1 implies that the speed-μ_{e_n} random walk X^n on (T_{e_n}, r_{e_n}) starting in ρ_{e_n} converges weakly in path space to the μ_e-Brownian motion $X = (X_t)_{t\geq 0}$ on (T_e, r_e) started in ρ_e for almost all realizations.

We can interpret this as *annealed convergence* in law of X^n to X, which we define – in analogy to Definition 1.3 and in view of Skorohod's representation theorem – as follows. There exists a coupling of the underlying random spaces $\mathcal{X} = (T_e, r_e, \mu_e)$, $\mathcal{X}_n = (T_{e_n}, r_{e_n}, \mu_{e_n})$, $n \in \mathbb{N}$, such that almost surely, conditioned on these spaces, X^n converges weakly in path space to X in the sense of Definition 1.3. In particular, the rescaled height processes $Z^{(n)}$, defined in (7.13), converge under the annealed law to the height process $Z = (Z_t)_{t \ge 0}$ defined by $Z_t := r_e(\rho_e, X_t)$. As X is recurrent by Theorem 4 in [AEW13], its life time is infinite, and Z is non-trivial.

7.5. **Motions on Λ-coalescent measure trees.** We conclude the example section with the example of speed-ν motions on the Λ-coalescent measure trees for appropriate measures ν. These have not been considered in the literature so far.

Let Λ be a finite measure on $([0,1], \mathcal{B}([0,1]))$ which satisfies

$$\sum_{n=2}^{\infty} \Big(\int_0^1 \sum_{k=2}^{n} \binom{n}{k} (k-1) x^{k-2} (1-x)^{n-k} \Lambda(\mathrm{d}x) \Big)^{-1} < \infty. \tag{7.20}$$

Denote by $\mathbb{S}$ the set of all partitions of $\mathbb{N}$, and for each $n \in \mathbb{N}$ by $\mathbb{S}_n$ the set of all partitions of $\{1, ..., n\}$. Write ρ_n for the restriction map from $\mathbb{S}$ to $\mathbb{S}_n$.

The Λ-coalescent is the unique $\mathbb{S}$-valued strong Markov process ζ, such that for each $n \in \mathbb{N}$ the restricted process $\rho_n(\zeta)$ is the following $\mathbb{S}_n$-valued continuous time Markov chain. Given the current partition $\mathcal{P} \in \mathbb{S}_n$, every k-tuple of its partition elements merges independently at rate

$$\lambda_{k,\#\mathcal{P}} := \int \Lambda(\mathrm{d}x)\, x^{k-2} (1-x)^{\#\mathcal{P}-k} \tag{7.21}$$

into one partition element, thereby forming a new partition. It is known that condition (7.20) is equivalent to the Λ-coalescent coming down from infinity, i.e., under (7.20), $\#\zeta_t < \infty$ for each $t > 0$, almost surely ([Sch00]). Furthermore, (7.20) implies the so-called *dust-free property*, i.e., $\int_0^1 \Lambda(\mathrm{d}x)\, x^{-1} = \infty$.

Equip for each realization of the Λ-coalescent started in $\mathcal{P}_0 := \{\{i\} : i \in \mathbb{N}\}$ the set $\mathbb{N}$ with the genealogical distances, i.e., $r(i,j)$ is for all $i, j \in \mathbb{N}$ the first time when i and j belong to the same partition element. Denote the completion of $(\mathbb{N}, r)$ by $(\mathcal{T}_\Lambda, r)$. Obviously, coming down from infinity implies (and is in fact equivalent to) the compactness of $\mathcal{T}_\Lambda$. Further, equip for each $n \in \mathbb{N}$, $\mathcal{T}_\Lambda$ with the sampling measure $\mu^n := \frac{1}{n} \sum_{i=1}^n \delta_i$. By Theorem 4 in [GPW09] the sequence $((\mathcal{T}_\Lambda, r, \mu^n))_{n \in \mathbb{N}}$ converges weakly in Gromov-weak topology towards the so-called Λ-*coalescent measure tree*, $(\mathcal{T}_\Lambda, r, \mu)$.

Consider next the $\mathbb{R}$-tree $(\bar{\mathcal{T}}_\Lambda, \bar{r})$ spanned by $(\mathcal{T}_\Lambda, r)$, and notice that $\mathcal{T}_\Lambda$ is ultra-metric. We therefore find a unique point $\rho \in \bar{\mathcal{T}}_\Lambda$ whose distance to $\mathcal{T}_\Lambda$ equals $\mathrm{diam}(\bar{\mathcal{T}}_\Lambda)/2$, which we choose as the root. For each point $x \in \bar{\mathcal{T}}_\Lambda$ denote by

$$S^x := \big\{ z \in \mathcal{T}_\Lambda : x \in [\rho, z] \big\} \tag{7.22}$$

the (leaves of the) subtree above x, and recall from (2.2) the notion of the length measure $\lambda^{(T,r,\rho)}$ of a rooted compact metric tree (T, r, ρ).

Define the speed measures ν^n, $n \in \mathbb{N}$, and ν on $\bar{\mathcal{T}}_\Lambda$ as being absolutely continuous with respect to the length measure with densities

$$\tfrac{\mathrm{d}\nu^n}{\mathrm{d}\lambda^{\bar{\mathcal{T}}_\Lambda}}(x) := \mu^n\big(S^x\big), \quad \text{and} \quad \tfrac{\mathrm{d}\nu}{\mathrm{d}\lambda^{\bar{\mathcal{T}}_\Lambda}}(x) := \mu\big(S^x\big). \tag{7.23}$$

for all $x \in \bar{\mathcal{T}}_\Lambda$. Obviously, ν^n, $n \in \mathbb{N}$, and ν are finite measures with total masses at most (and in fact due to the dust-free property equal to) $\mathrm{diam}(\bar{\mathcal{T}}_\Lambda)/2$. Note that for every ultrametric space (T, r), the map $\xi^{(T,r)}$ which sends a pair $(t, x) \in [0, \infty) \times T$ to the unique "ancestor" of x a time t back, i.e., the unique $y \in \bar{T}$ ($\bar{T}$ denoting the span of T) with $\bar{r}(y, x) = t \wedge \frac{1}{2}\mathrm{diam}(\bar{T})$ is continuous. Hence using the convergence alluded to earlier (Theorem 4 in [GPW09]) the sequence $((\bar{\mathcal{T}}_\Lambda, \nu^n))_{n \in \mathbb{N}}$ converges weakly in Gromov-weak topology towards $(\bar{\mathcal{T}}_\Lambda, \nu)$. Our invariance principle therefore implies that the ν_n-Brownian motion on $(\mathrm{supp}(\nu^n), \bar{r})$ converges weakly to the ν-Brownian motion on $(\bar{\mathcal{T}}_\Lambda, \bar{r})$ in the sense of finite dimensional marginals (provided all Brownian motions start at the same point). Applying once more the dust-free property implies that the global lower mass-bound holds, and thus the convergence holds even in path space.

We can modify the example such that we obtain path-wise convergence of a continuous time Markov chain to a motion on a totally disconnected (limiting) tree. For that purpose, denote by $\mathrm{Br}(\bar{\mathcal{T}}_\Lambda)$ the set of branch points of $\bar{\mathcal{T}}_\Lambda$, i.e., the set of those $x \in \bar{\mathcal{T}}_\Lambda$ such that either $x = \rho$ or $\bar{\mathcal{T}}_\Lambda \setminus \{x\}$ consists of at least 3 connected components. Consider now the (atomic) length measure on $\mathrm{Br}(\bar{\mathcal{T}}_\Lambda)$ and the Dirac measure δ_ρ, and define

$$\hat{\lambda} := \lambda^{(\mathrm{Br}(\bar{\mathcal{T}}_\Lambda), \bar{r}, \rho)} + \delta_\rho. \tag{7.24}$$

We use the speed measures $\tilde{\nu}^n$, $n \in \mathbb{N}$, and $\tilde{\nu}$ on $\bar{\mathcal{T}}_\Lambda$ which are absolutely continuous with respect to $\hat{\lambda}$ with densities

$$\tfrac{\mathrm{d}\tilde{\nu}^n}{\mathrm{d}\hat{\lambda}}(x) := \mu^n\bigl(S^x\bigr), \quad \text{and} \quad \tfrac{\mathrm{d}\tilde{\nu}}{\mathrm{d}\hat{\lambda}}(x) := \mu\bigl(S^x\bigr) \tag{7.25}$$

for all $x \in \mathrm{Br}(\bar{\mathcal{T}}_\Lambda)$. For each $\varepsilon \in (0, \frac{1}{2}\mathrm{diam}(\bar{\mathcal{T}}_\Lambda)))$ and for all suitably large $n \in \mathbb{N}$, we have $\mathrm{supp}(\tilde{\nu}^n) \cap \{x \in \mathrm{Br}(\bar{\mathcal{T}}_\Lambda) : \bar{r}(x, \mathcal{T}_\Lambda) \geq \varepsilon\} = \{x \in \mathrm{Br}(\bar{\mathcal{T}}_\Lambda) : \bar{r}(x, \mathcal{T}_\Lambda) \geq \varepsilon\}$. Therefore, the sequence $((\bar{\mathcal{T}}_\Lambda, \tilde{\nu}^n))_{n\in\mathbb{N}}$ also converges weakly in Gromov-weak topology towards $(\bar{\mathcal{T}}_\Lambda, \tilde{\nu})$. Thus our invariance principle applies to the speed-$\tilde{\nu}^n$ random walk on $\mathrm{supp}(\tilde{\nu}^n)$ and the speed-$\tilde{\nu}$ motion on $\mathrm{supp}(\tilde{\nu}) = \mathrm{Br}(\bar{\mathcal{T}}_\Lambda) \cup \mathcal{T}_\Lambda$.

Acknowledgements. We would like to thank Steve Evans and Fabian Gerle for discussions, and the anonymous referees for their detailed reports, which enabled us to improve the paper.

References

[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** (2013), no. 14, 1–21.

[ADH14] ———, *Exit times for an increasing Lévy tree-valued process*, Probab. Theory Related Fields **159** (2014), no. 1-2, 357–403. MR 3201925

[AEW13] Siva Athreya, Michael Eckhoff, and Anita Winter, *Brownian motion on $\mathbb{R}$-trees*, Trans. Amer. Math. Soc. **365** (2013), 3115–3150.

[Ald91] David Aldous, *The continuum random tree. II: An overview*, Stochastic analysis, Proc. Symp., Durham/UK 1990, Lond. Math. Soc. Lect. Note Ser. **167** (1991), 23–70.

[Ald93] David Aldous, *The continuum random tree III*, Ann. Probab. **21** (1993), 248–289.

[ALW16] Siva Athreya, Wolfgang Löhr, and Anita Winter, *The gap between Gromov-vague and Gromov-Hausdorff-vague topology*, Stochastic Process. Appl. **126** (2016), no. 9, 2527–2553. MR 3522292

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov, *A course in metric geometry*, Graduate studies in mathematics, vol. 33, AMS, Boston, MA, 2001.

[BEPR08] Shankar Bhamidi, Steven N. Evans, Ron Peled, and Peter Ralph, *Brownian motion on disconnected sets, basic hypergeometric functions, and some continued fractions of Ramanujan*, Probability and statistics: Essays in honor of David A. Freedman, Inst. Math. Stat. Collect., vol. 2, Inst. Math. Statist., Beachwood, OH, 2008, pp. 42–75.

[BK06] Martin T. Barlow and Takashi Kumagai, *Random walk on the incipient infinite cluster on trees*, Illinois J. Math. **50** (2006), no. 1, 33–65.

[Bog07] V. I. Bogachev, *Measure theory, volume II*, Springer, 2007.

[Cro08] David Croydon, *Convergence of simple random walks on random discrete trees to Brownian motion on the continuum random tree*, Ann. Inst. H. Poincaré Probab. Statist. **44** (2008), no. 6, 987–1019.

[Cro10] ———, *Scaling limits for simple random walks on random ordered graph trees*, Adv. in Appl. Probab. **42** (2010), no. 2, 528–558.

[Duq09] Thomas Duquesne, *Continuum random trees and branching processes with immigration*, Stochastic Process. Appl. **119** (2009), no. 1, 99–129. MR 2485021 (2010k:60119)

[EPW06] Steven N. Evans, Jim Pitman, and Anita Winter, *Rayleigh processes, real trees, and root growth with re-grafting*, Probab. Theory Related Fields **134** (2006), no. 1, 81–126.

[Eva08] Steven N. Evans, *Probability and real trees*, École d'Été de Probabilités de Saint Flour XXXV-2005, Lecture Notes in Mathematics, vol. 1920, Springer-Verlag, 2008.

[EW06] Steven N. Evans and Anita Winter, *Subtree prune and re-graft: A reversible real-tree valued Markov chain*, Ann. Probab. **34** (2006), no. 3, 918–961.

[FOT11] Masatoshi Fukushima, Yoichi Oshima, and Masayoshi Takeda, *Dirichlet forms and symmetric Markov processes*, second ed., de Gruyter Studies in Mathematics, vol. 19, Walter de Gruyter & Co., Berlin, 2011. MR 2778606 (2011k:60249)

[Fuk87] Kenji Fukaya, *Collapsing of Riemannian manifolds and eigenvalues of Laplace operators*, Invent. Math. **87** (1987), 517–547.

[GPW09] Andreas Greven, Peter Pfaffelhuber, and Anita Winter, *Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees)*, Probab. Theory Related Fields **145** (2009), no. 1-2, 285–322.

[Gro99] Misha Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, vol. 152, Birkhäuser Boston Inc., Boston, MA, 1999, Based on the 1981 French original.

[Kal02] Olav Kallenberg, *Foundations of modern probability*, second ed., Probability and its Applications (New York), Springer-Verlag, New York, 2002.

[Kes86] Harry Kesten, *Subdiffusive behavior of random walk on a random cluster*, Ann. Inst. H. Poincaré Probab. Statist. **22** (1986), no. 4, 425–487.

[Kig95] Jun Kigami, *Harmonic calculus on limits of networks and its applications to dentrides*, J. Funct. Anal. **125** (1995), 48–86.

[Kre95] W.B. Krebs, *Brownian motion on the continuum tree*, Probab. Theory Related Fields **101** (1995), 421–433.

[KS03] Kazhiro Kuwae and Takashi Shioya, *Convergence of spectral structure: a functional analytic theory and its application to spectral geometry*, Analysis and Geometry **11** (2003), no. 4, 599–673.

[Löh13] Wolfgang Löhr, *Equivalence of Gromov-Prohorov- and Gromov's $\underline{\square}_\lambda$-metric on the space of metric measure spaces*, Electron. Commun. Probab. **18** (2013), no. 17, 1–10. MR 3037215

[LVW15] Wolfgang Löhr, Guillaume Voisin, and Anita Winter, *Convergence of bi-measure $\mathbb{R}$-trees and the pruning process*, Ann. Inst. H. Poincaré Probab. Statist. **51** (2015), no. 4, 1342–1368. MR MR3414450

[Mos69] Umberto Mosco, *convergence of convex sets and of solutions of variational inequalities*, Adv. Math. **3** (1969), no. 4, 510–585.

[Mos94] ———, *Composite media and asymptotic dirichlet forms*, J. Funct. Anal. **123** (1994), no. 2, 368–421.

[Sch00] Jason Schweinsberg, *A necessary and sufficient condition for the Λ-coalescent to come down from infinity*, Electron. Commun. Probab. **5** (2000), 1–11.

[Sto63] Charles Stone, *Limit theorems for random walks, birth and death processes, and diffusion processes*, Illinois J. Math. **7** (1963), no. 4, 638–660.

SPACES OF ALGEBRAIC MEASURE TREES AND TRIANGULATIONS OF THE CIRCLE

WOLFGANG LÖHR AND ANITA WINTER

Abstract. In this paper we present with *algebraic trees* a novel notion of (continuum) trees which generalizes countable graph-theoretic trees to (potentially) uncountable structures. For that purpose we focus on the tree structure given by the branch point map which assigns to each triple of points their branch point. We give an axiomatic definition of algebraic trees, define a natural topology, and equip them with a probability measure on the Borel-σ-field.

Under an order-separability condition, algebraic (measure) trees can be considered as tree structure equivalence classes of metric (measure) trees (i.e. subtrees of $\mathbb{R}$-trees). Using Gromov-weak convergence (i.e. sample distance convergence) of the particular representatives given by the metric arising from the distribution of branch points, we define a metrizable topology on the space of equivalence classes of algebraic measure trees.

In many applications binary trees are of particular interest. We introduce on that subspace with the sample shape and the sample subtree mass convergence two additional, natural topologies. Relying on the connection to triangulations of the circle, we show that all three topologies are actually the same, and the space of binary algebraic measure trees is compact. To this end, we provide a formal definition of triangulations of the circle, and show that the coding map which sends a triangulation to an algebraic measure tree is a continuous surjection onto the subspace of binary algebraic non-atomic measure trees.

Contents

1. Introduction

Graph-theoretic trees are abundant in mathematics and its applications, from computer science to theoretical biology. A natural question is how to define limits and limit objects as the size of the trees tends to infinity. On the one hand, there are *local* approaches yielding countably infinite graphs, or generalized so-called graphings with a Benjamini-Schramm-type approach (going back to [BS01], see [Lov12, Part 4]). On the other hand, if one takes a more *global* point of view, as we are doing here, the predominant approach is to consider graph-theoretic trees as metric spaces equipped with the (rescaled) graph distance. Then the limit objects are certain "tree-like" metric spaces, most prominently so-called $\mathbb{R}$-trees introduced in [Tit77]. They are also of independent interest, e.g. for studying isometry groups of hyperbolic space ([MS84]), or as generalized universal covering spaces in the study of the fundamental groups of one-dimensional spaces ([FZ13]). Characterizing the topological structures induced by $\mathbb{R}$-trees has received considerable attention ([MO90, MNO92, Fab15]). Here, instead of the topological structures, we are more interested in the "tree structures" induced by $\mathbb{R}$-trees. We formalize the tree structure with a branch point map and call the resulting axiomatically defined objects *algebraic trees*. While, unlike for metric spaces, we do not know any useful notion of convergence for topological spaces or topological measure spaces, it is essential for us that we can define a very useful convergence of algebraic measure trees.

Our main motivation lies in suitable state spaces for tree-valued stochastic processes. The construction and investigation of scaling limits of tree-valued Markov chains within a metric space setup started with the continuum analogs of the Aldous-Broder-algorithm for sampling a uniform spanning tree from the complete graph ([EPW06]), and of the tree-valued subtree-prune and regraft Markov chain used in the reconstruction of phylogenetic trees ([EW06]). It continued with the construction of evolving genealogies of infinite size populations in population genetics ([GPW13, DGP12, Pio10, GSW16]) and in population dynamics ([Glö12, KW]). Moreover, continuum analogues of pruning procedures were constructed ([ADV10, AD12, LVW15, HW, HW19]). All these constructions have in common that they encode trees as metric (measure) spaces or bi-measure $\mathbb{R}$-trees, and equip the respective space of trees with the Gromov-Hausdorff ([Gro99]), Gromov-weak ([Fuk87, GPW09, Löh13]), Gromov-Hausdorff-weak ([Vil09, ADH13, ALW17]), or leaf-sampling weak-vague topology ([LVW15]).

In the present paper, we shift the focus from the metric to the tree structure for several reasons. First, checking compactness or tightness criteria for (random) metric (measure) spaces is not always easy, and some natural sequences of trees do not converge as metric (measure) spaces with a uniform rescaling of edge-lengths. At least for the subspace of binary algebraic measure trees we introduce, the situation is much more favorable, because it turns out to be compact. Second, the metric is often less canonical than the tree structure in situations where it is not clear that every edge should have the same length, e.g. in a phylogenetic tree, where edges might correspond to very different evolutionary time spans. Third, one might want to preserve certain functionals of the tree structure in the limit. For instance, the limit of binary trees is not always binary in the metric space setup, while this will be the case for our algebraic measure trees. Also, the centroid function used in [Ald00] is not continuous on spaces of metric measure trees, but it is continuous on our space.

The starting point of our construction is the notion of an $\mathbb{R}$-tree (see [Tit77, DMT96, Chi01, Eva08]). There are many equivalent definitions, but the following one is the most convenient for us:

Definition 1.1 ($\mathbb{R}$-trees)**.** A metric space (T, r) is an $\mathbb{R}$-*tree* iff it satisfies the following:

(RT1) (T, r) satisfies the so-called 4-*point condition*, i.e., for all $x_1, x_2, x_3, x_4 \in T$,

$$r(x_1, x_2) + r(x_3, x_4) \le \max\big\{r(x_1, x_3) + r(x_2, x_4),\, r(x_1, x_4) + r(x_2, x_3)\big\}. \tag{1.1}$$

(RT2) (T, r) is a connected metric space.

Notice that any metric space (T, r) satisfying (RT1) and (RT2) admits a *branch point map* $c\colon T^3 \to T$, i.e., for all $x_1, x_2, x_3 \in T$ there exists a unique point $c(x_1, x_2, x_3) \in T$ such that

$$\big\{c(x_1, x_2, x_3)\big\} = [x_1, x_2] \cap [x_1, x_3] \cap [x_2, x_3], \tag{1.2}$$

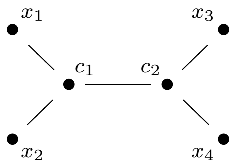


Figure 1. The only possible tree shape spanned by four points separates them into two pairs. Here, $r(x_1, x_2) + r(x_3, x_4) < \max\{r(x_1, x_3) + r(x_2, x_4),\ r(x_1, x_4) + r(x_2, x_3)\}$, while any other permutation yields equality. Furthermore, $c_1 = c(x_1, x_2, x_3) = c(x_1, x_2, x_4)$ and $c_2 = c(x_1, x_3, x_4) = c(x_2, x_3, x_4)$.

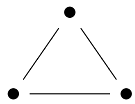


Figure 2. The graph shown here is not a tree, but the vertices satisfy the 4-point condition with respect to the graph-distance. Condition (MT2) fails.

where for $x, y \in T$ the *interval* $[x, y]$ is defined as

$$[x, y] := \big\{z \in T :\ r(x, z) + r(z, y) = r(x, y)\big\}. \tag{1.3}$$

Given the branch point map c, we can recover the intervals via the identity

$$[x, y] = \big\{z \in T :\ c(x, y, z) = z\big\}. \tag{1.4}$$

While condition (RT1) is crucial for trees as it reflects the fact that there is only one possible shape for the subtree spanned by four points (as shown in Figure 1), the assumption of connectedness can be relaxed. In [ALW17], the notion of a *metric tree* was introduced to allow for a unified set-up in discrete and continuous situations. A metric tree (T, r) is defined as a metric space which can be embedded isometrically into an $\mathbb{R}$-tree such that it contains all branch points $c(x_1, x_2, x_3)$, $x_1, x_2, x_3 \in T$, as defined by (1.2). To exclude non-tree graphs satisfying the 4-point condition (see Figure 2), we have to require the property of containing the branch points explicitly.

Definition 1.2 (metric trees)**.** A metric space (T, r) is a *metric tree* if the following holds:

(MT1) (T, r) satisfies the 4-point condition (RT1).

(MT2) (T, r) admits all branch points, i.e., for all $x_1, x_2, x_3 \in T$ there exists a (necessarily unique) $c(x_1, x_2, x_3) \in T$ such that

$$r\big(x_i,\ c(x_1, x_2, x_3)\big) + r\big(c(x_1, x_2, x_3),\ x_j\big) = r(x_i, x_j) \quad \forall\, i, j \in \{1, 2, 3\},\ i \neq j. \tag{1.5}$$

Our main goal is to forget the metric while keeping the tree structure encoded by the branch point map. To axiomatize the latter, notice that for metric trees the branch point map satisfies the following obvious properties:

(BPM1) The map $c\colon T^3 \to T$ is symmetric.

(BPM2) The map $c\colon T^3 \to T$ satisfies the 2-*point condition* that for all $x, y \in T$

$$c(x, y, y) = y. \tag{1.6}$$

(BPM3) The map $c\colon T^3 \to T$ satisfies the 3-*point condition* that for all $x, y, z \in T$

$$c\big(x, y,\ c(x, y, z)\big) = c(x, y, z). \tag{1.7}$$

(BPM4) The map $c\colon T^3 \to T$ satisfies the 4-*point condition* that for all $x_1, x_2, x_3, x_4 \in T$,

$$c(x_1, x_2, x_3) \in \big\{c(x_1, x_2, x_4),\ c(x_1, x_3, x_4),\ c(x_2, x_3, x_4)\big\}. \tag{1.8}$$

Definition 1.3 (algebraic tree)**.** An *algebraic tree* (T, c) consists of a set $T \neq \emptyset$ and a branch point map $c\colon T^3 \to T$ satisfying (BPM1)–(BPM4).

We define a natural topology on an algebraic tree (T, c) as follows. For each $x \in T$, we define an equivalence relation $\sim_x$ on $T \setminus \{x\}$ such that for all $y, z \in T \setminus \{x\}$, $y \sim_x z$ iff $c(x, y, z) \neq x$. For $y \in T \setminus \{x\}$, we denote by

$$\mathcal{S}_x(y) := \{z \in T : z \sim_x y\} \tag{1.9}$$

the equivalence class w.r.t. $\sim_x$ which contains y. $\mathcal{S}_x(y)$ should be thought of as a subtree rooted at (but not containing) x. We consider the topology generated by sets of the form (1.9) with $x \neq y$ and denote by $\mathcal{B}(T, c)$ the corresponding Borel σ-algebra.

Our first main result (Theorem 1) relates metric trees with algebraic trees. On the one hand, if (T, r) is a metric tree, then it is clear that T together with the map c from (MT2) yields an algebraic tree. On the other hand, we show that every order separable algebraic tree (Definition 2.20) is induced by a metric tree in this way. More concretely, if ν is a measure on $\mathcal{B}(T, c)$ which is finite and non-zero on non-degenerate intervals, i.e., on sets of the form

$$[x, y] := \big\{z \in T : c(x, y, z) = z\big\} \tag{1.10}$$

for $x, y \in T$, $x \neq y$, then a metric representation of (T, c) is given by

$$r_\nu(x, y) := \nu\big([x, y]\big) - \tfrac{1}{2}\nu\big(\{x\}\big) - \tfrac{1}{2}\nu\big(\{y\}\big). \tag{1.11}$$

Next, we equip an algebraic tree (T, c) with a sampling probability measure μ on $\mathcal{B}(T, c)$, and call the resulting triple (T, c, μ) *algebraic measure tree*. Two algebraic measure trees (T, c, μ) and (T', c', μ') are equivalent (compare with Definition 3.2) if there are $A \subseteq T$, $A' \subseteq T'$ and a bijection $\phi\colon A \to A'$ such that the following holds.

- $\mu(A) = \mu'(A') = 1$, $c(A^3) \subseteq A$ and $c'((A')^3) \subseteq A'$.
- ϕ is measure preserving, and $c'(\phi(x), \phi(y), \phi(z)) = \phi(c(x, y, z))$ for all $x, y, z \in T$.

Denote by $\mathbb{T}$ the space of all equivalence classes of order separable algebraic measure trees. We equip $\mathbb{T}$ with a topology based on the Gromov-weak topology (introduced in [GPW09] and shown in [Löh13] to be equivalent to Gromov's $\underline{\Box}_1$-topology from [Gro99]). For that purpose, we introduce a particular metric representation of an algebraic measure tree. As metric representations are far from being unique, we will consider the intrinsic metric r_ν which comes from the branch point distribution, i.e., the image measure $\nu := c_*\mu^{\otimes 3}$ of $\mu^{\otimes 3}$ under the branch point map c. We declare that

$$(T_n, c_n, \mu_n) \underset{n\to\infty}{\longrightarrow} (T, c, \mu) \quad \text{iff} \quad (T, r_{(c_n)_*\mu_n^{\otimes 3}}, \mu_n) \to (T, r_{c_*\mu^{\otimes 3}}, \mu) \text{ Gromov-weakly,} \tag{1.12}$$

or equivalently, $\Phi((T_n, c_n, \mu_n)) \underset{n\to\infty}{\longrightarrow} \Phi((T, c, \mu))$ for all test functions of the form

$$\Phi(T, c, \mu) = \Phi^{n,\phi}(T, c, \mu) := \int_{T^n} \phi\big((r_{c_*\mu^{\otimes 3}}(x_i, x_j))_{1\le i,j\le n}\big)\, \mu^{\otimes n}(\mathrm{d}\underline{x}), \tag{1.13}$$

where $n \in \mathbb{N}$ and $\phi \in \mathcal{C}_b(\mathbb{R}^{n\times n})$. We refer to this convergence as branch point distribution distance Gromov-weak convergence, or shortly, *bpdd-Gromov-weak convergence*.

A particular subclass of interest is the space of binary algebraic measure trees. Similar to encoding compact $\mathbb{R}$-trees by a continuous excursion on the unit interval, binary algebraic trees can be encoded by *sub-triangulations of the circle* (see Figure 3), where a sub-triangulation of the circle $\mathbb{S}$ is a closed, non-empty subset C of $\mathbb{D}$ satisfying the following two conditions:

(Tri1) The complement of the convex hull of C consists of open interiors of triangles.
(Tri2) C is the union of non-crossing (non-intersecting except at endpoints), possibly degenerate closed straight line segments with endpoints in $\mathbb{S}$.

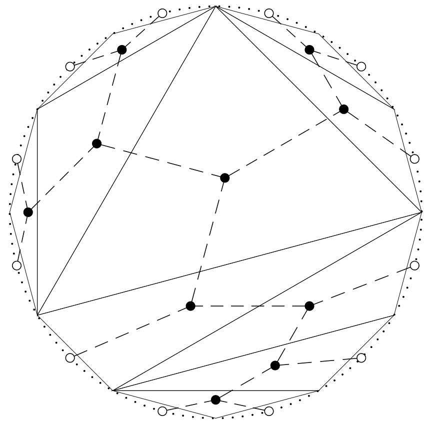


FIGURE 3. A triangulation of the 12-gon and the tree coded by it.

Such an encoding was introduced by David Aldous in [Ald94a, Ald94b], and there has since then been an increasing amount of research in the random tree community using this approach (e.g. [CLG11, BS15, CK15]). Also more general -angulations and dissections have been considered which allow for encoding not necessarily binary trees ([Cur14, CHK15]). Note, however, that the relation between triangulations and trees has never been made explicit, except for the finite case, where the tree is the dual graph.

Aldous originally defines a triangulation of the circle as a closed subset of the disc the complement of which is a disjoint union of open triangles with vertices on the circle ([Ald94b, Definition 1]). We modify his definition in two respects. First, we add Condition (Tri2) which enforces existence of branch points and under which triangulations of the circle are precisely the Hausdorff-metric limits of triangulations of n-gons as $n \to \infty$. Second, we extend the definitions to sub-triangulation of the circle (triangulations of a subset of the circle) which allow for encoding algebraic measure trees with point masses on leaves. In fact, triangulations of the whole circle encode binary trees with non-atomic measures, which is relevant in the case of Aldous's CRT. We formally construct the coding map that associates to a sub-triangulation of the circle the corresponding binary algebraic measure tree with point-masses restricted to the leaves. Furthermore, we show that – similar to the case of coding compact $\mathbb{R}$-trees by continuous excursions – the coding map is *surjective* and *continuous* when the set of sub-triangulations is equipped with the Hausdorff metric topology and the set of binary algebraic measure trees with our bpdd-Gromov-weak topology (Theorem 2).

We also analyze the subspace of binary algebraic measure trees with point-masses restricted to the leaves in more detail. Our third main result (Theorem 3) states that this space in the bpdd-Gromov-weak topology is topologically as nice as it gets, namely a compact, metrizable space. We also give two more notions of convergence which turn out to be equivalent to bpdd-Gromov-weak convergence on this subspace. One is of combinatorial nature and based on the weak convergence of test functions of the form

$$\Phi(T,c,\mu) = \Phi^{n,\mathfrak{t}}(T,c,\mu) := \mu\big(\big\{(u_1,...,u_n) \in T^n : \mathfrak{s}_{(T,c)}(u_1,...,u_n) = \mathfrak{t}\big\}\big), \tag{1.14}$$

where $\mathfrak{t}$ is an n-cladogram (a binary graph-theoretic tree with n leaves) and $\mathfrak{s}_{(T,c)}$ denotes the shape spanned by a finite sample in (T,c) (Definitions 5.1 and 5.2). The other one is more in the spirit of stochastic analysis and based on weak convergence of the *tensor of subtree-masses* read off the algebraic measure subtree spanned by a finite sample (see Definition 5.12). This equivalence allows to switch between different perspectives and turns out to be very useful for the following reasons:

- Using convergence of sample bpd-distance matrices allows to exploit well-known results about Gromov-weak convergence.
- Showing convergence of graph theoretic tree-valued Markov chains as the number of vertices tends to infinity is, due to the combinatorial nature of the Markov chains, often easiest by showing convergence of the sample shape distributions. This has recently been successfully applied in the construction of the conjectured continuum limit of the Aldous chain ([Ald00]) in [LMW20], and of the continuum limit of the $\alpha = 1$-Ford chain ([For05]) in [Nus].
- The convergence of sample subtree-mass tensor distributions allows to analyze the limit process with stochastic analysis methods and gives more insight into the global structure of the evolving random trees.

Related work. As an alternative with better compactness properties to Gromov-Hausdorff convergence of discrete trees, Curien suggested in [Cur14] to look at convergence of coding triangulations (in Hausdorff metric topology). He also proposed to read off a measured, ordered, *topological tree* from the limit triangulation, and sketched the construction as quotient w.r.t. some equivalence relation in the special case of the Brownian triangulation. Note, however, that the topological information cannot be completely encoded by the triangulation, because the latter only encodes the algebraic measure tree by Theorem 2, and the algebraic structure does not determine the topological structure uniquely (see Example 2.35). Therefore, Curien did not obtain a map from the space of triangulations to a space of trees.

The random exchangeable *didendritic systems* introduced recently by Evans, Grübel, and Wakolbinger in [EGW17] can be considered as rooted, ordered versions of binary algebraic measure trees with diffuse measure on the set of leaves. A didendritic system is an equivalence relation on $\mathbb{N} \times \mathbb{N}$ together with two partial orders on the set of equivalence classes. An exchangeable didendritic system is similar to our sequence of sample-shape distributions. The authors also introduce a particular metric representation as an $\mathbb{R}$-tree. Even though it is somewhat implicit in their work that they think of the set of exchangeable didendritic systems as equipped with a kind of sample shape convergence, they do not define it explicitly and there are no statements about properties of the resulting space. Also note that there is no notion of "continuum tree" without a particular measure in this framework.

Close relatives of algebraic measure trees have recently been studied independently by Forman in [For]. He uses ideas from [FHP18] to represent rooted trees by so-called *hierarchies* (certain sets of subsets) on $\mathbb{N}$, which are similar to the didendritic systems in [EGW17], but unordered. Thus, exchangeable random hierarchies can be thought of as rooted versions of algebraic measure trees. Forman shows that the resulting equivalence classes of rooted measure $\mathbb{R}$-trees coincide with the so-called *mass-structural* equivalence classes, which he defines by bijections preserving intervals as well as masses of points, intervals and certain sub-trees. He also singles out a particular representative, which he calls *interval partition tree*, with the essentially same metric as in [EGW17] (not restricted to the binary case). This metric follows a similar idea to but is different from our r_ν. Note that [For] does not talk about convergence of trees or introduce a notion of "continuum tree" without a measure.

Outline. The rest of the paper is organized as follows. In Section 2, we introduce our concept of *algebraic trees* by formalising the branch point map as a tertiary operation on the tree. We also introduce an intrinsic Hausdorff topology and characterize compactness (Proposition 2.18) and second countability (Proposition 2.19). We show that under a separability constraint, algebraic trees can be seen as metric trees (subtrees of $\mathbb{R}$-trees), where the metric structure has been "forgotten" (Theorem 1), and give an example that the separability condition cannot be dropped without replacement.

In Section 3, we introduce the space of (equivalence classes of) order separable *algebraic measure trees*, and equip it with the Gromov-weak topology with respect to the metric associated with the branch point distribution. We show that the resulting space is separable and metrizable (Corollary 3.9). Furthermore, we prove a Carathéodory-type extension theorem, which is helpful for constructing algebraic measure trees (Propositions 3.12 and 3.13).

In Section 4, we give a definition of *triangulations of the circle*, and show that they are precisely the limits of triangulations of n-gons (Proposition 4.3). We also formalize the notion of the algebraic measure tree associated with a given triangulation of the circle. This correspondence has been allured to in the literature, but it has never been made precise (except for finite trees), and it has never been shown a tree in what sense is coded by a triangulation of the circle. We show that the resulting *coding map* (mapping triangulations to trees) is well-defined and surjective onto the space of binary algebraic measure trees with non-atomic measure. Furthermore, the coding map is *continuous* if the space of triangulations is equipped with the Hausdorff metric topology, and the space of trees with the bpdd-Gromov-weak topology (Theorem 2).

In Section 5, we consider the subspace of *binary* algebraic measure trees, and introduce two other, natural notions of convergence. We use the construction of the coding map from Section 4 to show that on this subspace all three notions of convergence are actually equivalent and define the same topology (Theorem 3). This topology turns the subspace of binary algebraic measure trees into a *compact, metrizable* space, which in particular implies that it is a closed subset of the space of algebraic measure trees. In this section, we also finish the proof of Theorem 2 by showing continuity of the coding map.

In Section 6, we consider the example of the continuum limits of sampling consistent families of random trees and illustrate it with the example of so-called β-splitting trees introduced in [Ald96]. This family includes the uniform binary tree (converging to the Brownian CRT) and the Yule tree (aka Kingman tree or random binary search tree).

2. Algebraic trees

In this section we introduce algebraic trees. In Subsection 2.1 we formalize the "tree structure" common to both graph-theoretic trees and metric trees by a function that maps every triplet of points in the tree to the corresponding branch point. We show that the set of defining properties is rich enough to obtain known concepts such as leaves, branch points, degree, edges, intervals, subtrees spanned by a set, discrete and continuum trees, etc. In Subsection 2.2 we introduce the notion of structure preserving morphisms. In Subsection 2.3 we equip algebraic trees with a canonical Hausdorff topology. We also characterize compactness and a concept we call order separability, which is closely related to second countability of the topology. Finally, in Subsection 2.4, we show that any order separable algebraic tree is induced by a metric tree (which is not true without order separability), and establish the condition under which this metric tree can be chosen to be a compact $\mathbb{R}$-tree.

2.1. The branch point map.

In this subsection we introduce algebraic trees. Recall from Definition 1.2 the definition of a metric tree, and the properties (BPM1)–(BPM4) of the map which sends a triplet of 3 points in a metric tree to its branch point.

Definition 2.1 (algebraic trees)**.** An *algebraic tree* (T, c) consists of a set $T \neq \emptyset$ and a branch point map $c\colon T^3 \to T$ satisfying (BPM1)–(BPM4).

The following useful property reflects the fact that any four points in an algebraic tree can be associated with a shape as illustrated in Figure 1 above.

Lemma 2.2 (a consequence of (BPM4))**.** *Let (T, c) be an algebraic tree. Then for all $x_1, \ldots, x_4 \in T$ the following hold:*

(i) If $c(x_1, x_2, x_3) = c(x_1, x_2, x_4)$, then $c(x_1, x_3, x_4) = c(x_2, x_3, x_4)$.

(ii) If $c(x_1,x_2,x_3) = c(x_1,x_2,x_4)$*, then* $c(x_1,x_2,x_3) = c(x_1,x_2,c(x_1,x_3,x_4))$*.*

Proof. Let $x_1,x_2,x_3,x_4 \in T$ with $c_1 := c(x_1,x_2,x_3) = c(x_1,x_2,x_4)$, and $c_2 := c(x_1,x_3,x_4)$.

(i) Condition (BPM4) implies that

$$c_2 \in \big\{c_1 = c(x_1,x_3,x_2),\ c(x_2,x_3,x_4),\ c_1 = c(x_1,x_2,x_4)\big\}. \tag{2.1}$$

Thus $c_1 = c_2$, or $c_2 = c(x_2,x_3,x_4)$. The second case is the claim. In the first case, we apply Condition (BPM4) once more to find that

$$c(x_2,x_3,x_4) \in \big\{c_1 = c(x_1,x_2,x_3),\ c_2 = c(x_1,x_3,x_4),\ c_1 = c(x_1,x_2,x_4)\big\} = \{c_1,c_2\} = \{c_2\}, \tag{2.2}$$

so that the claim also holds in this case.

(ii) Condition (BPM3) implies that

$$c(x_1,x_3,c_2) = c\big(x_1,\ x_3,\ c(x_1,x_3,x_4)\big) = c(x_1,x_3,x_4) = c_2, \tag{2.3}$$

and similarly also $c(x_2,x_3,c_2) = c(x_2,x_3,x_4) = c_2$. Now part *(i)* with x_4 replaced by c_2 yields $c(x_1,x_2,x_3) = c(x_1,x_2,c_2)$ as claimed. □

We have seen that the four axiomatizing properties of the branch point map are necessary. In many respects they are also sufficient to capture the tree structure. For example, in analogy to (1.3) we can define for each $x,y \in T$ the *interval* $[x,y]$ by

$$[x,y] := \big\{w \in T :\ c(x,y,w) = w\big\}. \tag{2.4}$$

We also use the notation $(x,y) := [x,y] \setminus \{x,y\}$, and similarly $(x,y]$, $[x,y)$. The following properties of intervals are known to hold in $\mathbb{R}$-trees (compare, e.g., to [Chi01, Chapter 2] or [Eva08, Chapter 3]):

Lemma 2.3 (properties of intervals)**.** *Let* (T,c) *be an algebraic tree. Then the following hold:*

(i) If $x,v,w,z \in T$ *are such that* $w \in [x,z]$ *and* $v \in [x,w]$*, then* $v \in [x,z]$*.*

(ii) If $x,y,z \in T$*, then*

$$[x,y] \cap [y,z] = \big[c(x,y,z),\ y\big]. \tag{2.5}$$

In particular,

$$\big[x,\ c(x,y,z)\big] \cap \big[c(x,y,z),\ z\big] = \big\{c(x,y,z)\big\}. \tag{2.6}$$

(iii) If $x,y,z \in T$*, then*

$$[x,y] \cup [y,z] = [x,z] \uplus \big(c(x,y,z),\ y\big]. \tag{2.7}$$

In particular,

$$[x,y] \cup [y,z] = [x,z] \quad \textit{iff} \quad y \in [x,z]. \tag{2.8}$$

(iv) For all $x,y,z \in T$*,*

$$[x,y] \cap [y,z] \cap [z,x] = \big\{c(x,y,z)\big\}. \tag{2.9}$$

Proof. (i) Let $x,v,w,z \in T$ with $w = c(x,w,z)$ and $v = c(x,v,w)$. Then by Condition (BPM4),

$$c(x,v,z) \in \big\{c(x,v,w),\ w = c(x,w,z),\ c(v,w,z)\big\}. \tag{2.10}$$

We discuss the three cases separately. If $c(x,v,z) = c(x,v,w)$, then $c(v,w,z) = c(x,w,z) = w$ by Lemma 2.2(i). It then follows that $c(x,v,z) = c(x,v,c(x,w,z)) = c(x,v,w) = v$ by Lemma 2.2(ii), which gives the claim in this case.

If $c(x,v,z) = w$ then $v = c(v,w,x) = c(v,w,z)$ by Lemma 2.2(i). It then follows that $c(x,v,z) = c(x,z,c(z,w,v)) = c(x,v,v) = v$ by Lemma 2.2(ii), which gives the claim in this case.

If $c(x,v,z) = c(v,w,z)$ then $v = c(x,w,v) = c(x,w,z) = w$ by Lemma 2.2(i). Thus $v = w \in [x,z]$, and the claim holds also in this case.

(ii) Let $x, y, z \in T$, and $v \in [x, y] \cap [y, z]$. That is, $v = c(x, v, y) = c(y, v, z)$. It follows from Lemma 2.2(i) that $c(x, z, v) = c(x, z, y)$, and then from Lemma 2.2(ii) together with Condition (BPM2) that

$$v = c(x, v, y) = c\big(v, y, c(y, x, z)\big). \tag{2.11}$$

Equivalently, $v \in [c(x, y, z), y]$. This proves the inclusion $[x, y] \cap [y, z] \subseteq [c(x, y, z), y]$. The other inclusion follows from (i).

Notice that (2.6) follows from (2.5) with the special choice $y = c(x, y, z)$.

(iii) Notice first that it follows immediately from (i) that the union on the right hand side is disjoint. We claim that

$$[x, z] \subseteq [x, y] \cup [y, z]. \tag{2.12}$$

Indeed, let $v \in [x, z]$, i.e. $c(x, z, v) = v$. Then by (BPM4) applied to $\{v, x, y, z\}$,

$$v = c(x, z, v) \in \big\{c(x, y, v),\ c(x, y, z),\ c(y, z, v)\big\}, \tag{2.13}$$

which implies that $v \in [x, y]$ (if $v = c(x, y, v)$) or $v \in [x, z] \cap [x, y]$ (if $v = c(x, y, z)$) or $v \in [y, z]$ (if $v = c(y, z, v)$). Second, we claim that for all $x, y, z \in T$,

$$[x, z] \cup \big[c(x, y, z),\ y\big] \subseteq [x, y] \cup [y, z]. \tag{2.14}$$

To see this, recall from (ii) that $[c(x, y, z), y] = [x, y] \cap [z, y] \subseteq [x, y] \cap [z, y]$. As $[x, c(x, y, z)] \subseteq [x, y]$ by (i), we have $[x, y] \subseteq [x, c(x, y, z)] \cup [c(x, y, z), y] \subseteq [x, y] \uplus (c(x, y, z), y]$. The corresponding inclusion for $[y, z]$ is shown in the same way, and we have proven Equation (2.7).

(iv) This follows immediately from (ii). □

We say that $\{x, y\} \subseteq T$ with $x \neq y$ is an *edge* of (T, c) if and only if there is "nothing in between", i.e. $[x, y] = \{x, y\}$, and denote by

$$\mathrm{edge}(T, c) := \big\{\{x, y\} \subseteq T :\ x \neq y,\ [x, y] = \{x, y\}\big\} \tag{2.15}$$

the *set of edges*. The following example explains that there is no need to distinguish between finite algebraic trees and graph-theoretical trees, and the definitions of edges are consistent.

Example 2.4 (finite algebraic trees correspond to graph-theoretic trees)**.** Finite algebraic trees are in one to one correspondence with finite (undirected) graph-theoretic trees. Let (T, E) be a graph-theoretic tree with vertex set T and edge set E. Then (T, E) corresponds to the algebraic tree (T, c_E) with $c_E(u, v, w)$ defined as the unique vertex that is on the (graph-theoretic) path between any two of u, v, w. Conversely, if (T, c) is an algebraic tree with T finite, then (T, c) corresponds to the graph-theoretic tree (T, E_c) with $E_c := \mathrm{edge}(T, c)$. Obviously, $c_{E_c} = c$. ♢

For a graph-theoretic tree (T, E), we can allow the vertex set T to be countably infinite, and still obtain a corresponding algebraic tree as in the previous example. Note, however, that countable algebraic trees do not necessarily correspond to graph-theoretic trees. Indeed, it is possible that T is countably infinite and $\mathrm{edge}(T, c) = \emptyset$. This can be seen by taking $T = \mathbb{Q}$ in the following example, which shows that every totally ordered space naturally corresponds to an algebraic tree.

Example 2.5 (totally ordered spaces as algebraic trees)**.** For a totally ordered space $(T, \leq)$, define $c_{\leq}(x, y, z) := y$ whenever $x \leq y \leq z$, $(x, y, z \in T)$. Then it is trivial to check that $(T, c_{\leq})$ is an algebraic tree and the interval $[x, y]$ coincides with the order interval $\{z \in T :\ x \leq z \leq y\}$. ♢

Conversely, given an algebraic tree (T, c) and any fixed point ρ (often referred to as *root*), we can define a *partial order* $\leq_\rho$ by letting for $x, y \in T$,

$$x \leq_\rho y \quad \text{iff } x \in [\rho, y]. \tag{2.16}$$

Lemma 2.6 (algebraic trees as semi-lattices). *Let (T,c) be an algebraic tree, and $\rho,x,y \in T$. Then $(T,\le_\rho)$ is a partially ordered set, and a meet semi-lattice with infimum*

$$x \wedge y = c(\rho,x,y). \tag{2.17}$$

Furthermore, $\le_\rho$ is a total order on $[\rho,x]$ for all $x \in T$.

Proof. Let $x,y \in T$ with $x \le_\rho y$ and $y \le_\rho x$. That is, $x = c(\rho,x,y)$ and $y = c(\rho,y,x)$ which implies that $x = y$, and proves that $\le_\rho$ is *antisymmetric*. As $x = c(\rho,x,x)$, $x \le_\rho x$ which proves that $\le_\rho$ is *reflexive*. Finally, to show *transitivity*, let $x,y,z \in T$ with $x \le_\rho y$ and $y \le_\rho z$. That is $x \in [\rho,y]$ and $y \in [\rho,z]$, which implies that $x \in [\rho,z]$ by Lemma 2.3(i). Equivalently, $x \le_\rho z$ which proves the transience, and thus that $\le_\rho$ is a partial order.

For the *infimum*, notice that $v \le_\rho x$ and $v \le_\rho y$ if and only if $v \in [\rho,x] \cap [\rho,y]$, or equivalently by Lemma 2.3(ii), $v \in [\rho, c(\rho,x,y)]$. As for all $v \in [\rho, c(\rho,x,y)]$ we have $v \le c(\rho,x,y)$, the claim (2.17) follows.

Fix $x \in T$. For *totality* on $[\rho,x]$, let $v,w \in [\rho,x]$, i.e., $v = c(\rho,v,x)$ and $w = c(\rho,w,x)$. Applying Condition (BPM4) to $\{\rho,v,w,x\}$ we find that one of the following three cases must occur: $c(\rho,v,w) = c(\rho,v,x)$ (which implies that $v = c(\rho,v,w)$, or equivalently, $v \le_\rho w$), $c(\rho,w,v) = c(\rho,w,x)$ (which implies that $w = c(\rho,w,v)$, or equivalently, $w \le_\rho v$), or $c(\rho,x,v) = c(\rho,x,w)$ (which implies that $w = v$). □

Corollary 2.7. *Let (T,c) be an algebraic tree, and $\rho,x,y \in T$. If $v \in [x,y]$, then $v \ge_\rho c(x,y,\rho)$.*

Proof. Let $\rho,x,y \in T$ and $v \in [x,y]$, i.e., $v = c(x,v,y)$. We need to show that $c(\rho,v,c(\rho,x,y)) = c(\rho,x,y)$.

By Condition (BPM4) applied to $\{x,y,\rho,v\}$ we have one of the following three cases: $c(x,y,\rho) = c(x,y,v)$ (in which case $c(\rho,x,y) = v$) or $c(\rho,y,x) = c(\rho,y,v)$ (in which case $c(x,v,\rho) = c(x,v,y) = v$ by Lemma 2.2(i) and thus $v \in [\rho,x]$; the claim then follows since this implies that $v \in [\rho,x] \cap [x,y] = [c(\rho,x,y),y]$ by Lemma 2.3(ii)), or $c(\rho,x,y) = c(\rho,x,v)$ (in which we conclude similar to the second case that $v \in [c(\rho,x,y),x]$). □

The partial orders $\le_\rho$ allow us to define a notion of completeness of algebraic trees.

Definition 2.8 (directed order completeness). Let (T,c) be an algebraic tree. We call (T,r) *(directed) order complete* if for all $\rho \in T$ the supremum of every totally ordered, non-empty subset exists in the partially ordered set $(T,\le_\rho)$.

Obviously, in an order complete algebraic tree, infima of totally ordered sets exists, because they are either ρ if the set is empty or a non-empty supremum w.r.t. a different root. This notion of completeness allows us to define the analogs of complete $\mathbb{R}$-trees.

Definition 2.9 (algebraic continuum tree). We call an algebraic tree (T,c) *algebraic continuum tree* if the following two conditions hold:

(ACT1) (T,c) is order complete.
(ACT2) $\mathrm{edge}(T,c) = \emptyset$.

2.2. **Morphisms of algebraic trees.** Like any decent algebraic structure (or in fact mathematical structure), algebraic trees come with a notion of structure-preserving morphisms.

Definition 2.10 (morphisms). Let (T,c) and $(\widehat{T},\hat{c})$ be algebraic trees. A map $f\colon T \to \widehat{T}$ is called a *tree homomorphism* (from T into $\widehat{T}$) if for all $x,y,z \in T$,

$$f\big(c(x,y,z)\big) = \hat{c}\big(f(x),f(y),f(z)\big). \tag{2.18}$$

We refer to a bijective tree homomorphism as *tree isomorphism.*

As we have seen that the tree structure can be expressed also in terms of intervals or partial orders rather than the branch point map, and we obtain the following equivalences.

Lemma 2.11 (equivalent definitions)**.** *Let (T,c) and $(\widehat{T},\hat{c})$ be algebraic trees, and $f\colon T\to\widehat{T}$. Then the following are equivalent:*

1. *f is a tree homomorphism.*
2. *For all $\rho\in T$, f is an order preserving map from $(T,\le_\rho)$ to $(\widehat{T},\le_{f(\rho)})$.*
3. *For all $x,y\in T$, $f([x,y])\subseteq[f(x),f(y)]$.*

Proof. *"1⇒2".* Let $x,y,\rho\in T$ with $x\le_\rho y$, i.e., $x=c(\rho,x,y)$. Then $f(x)=\hat{c}(f(\rho),f(x),f(y))$, and therefore $f(x)\le_{f(\rho)} f(y)$.

"2⇒3". Let $x,y,z\in T$ with $z\in[x,y]$. Then $z\le_x y$ and thus $f(z)\le_{f(x)} f(y)$, i.e. $f(z)\in[f(x),f(y)]$.

"3⇒1". Let $x,y,z\in T$. Then $\{c(x,y,z)\}=[x,y]\cap[x,z]\cap[y,z]$. Hence

$$\big\{f(c(x,y,z))\big\}\subseteq\big[f(x),f(y)\big]\cap\big[f(y),f(z)\big]\cap\big[f(x),f(z)\big]=\big\{\hat{c}\big(f(x),f(y),f(z)\big)\big\}. \tag{2.19}$$

Therefore, $f(c(x,y,z))=\hat{c}(f(x),f(y),f(z))$. □

The image of an algebraic tree under a homomorphism is a subtree in the following sense.

Definition 2.12 (subtree)**.** Let (T,c) be an algebraic tree, and $\emptyset\neq A\subseteq T$. A is called a *subtree (of (T,c))* if

$$c(A^3)\subseteq A. \tag{2.20}$$

We refer to $c(A^3)$ as the *algebraic subtree generated by A*.

Obviously, a subtree A of (T,c), implicitly equipped with the restriction of c to A^3, is an algebraic tree in its own right. Furthermore, the following lemma is easy to check.

Lemma 2.13 (tree homomorphisms)**.** *Let (T,c) and $(\widehat{T},\hat{c})$ be two algebraic trees, and $f\colon T\to\widehat{T}$ a homomorphism. Then the image $f(T)$ is a subtree of $\widehat{T}$. If f is injective, f^{-1} is a tree homomorphism from $f(A)$ into T.*

In particular, if $(\tilde{T},\tilde{c})$ is another algebraic tree, and g is a homomorphism form $(\widehat{T},\hat{c})$ to $(\tilde{T},\tilde{c})$, then $g\circ f$ is a homomorphism from (T,c) to $(\tilde{T},c_{\tilde{T}})$.

2.3. **Algebraic trees as topological spaces.** In contrast to metric trees, there is a priori no topology defined on a given algebraic tree. In this section, we therefore equip algebraic trees with a canonical topology.

For each $x\in T$, we introduce a (component) relation $\sim_x$ by letting $y\sim_x z$ if and only if $x\notin[y,z]$, where $y,z\in T$. Let for each $y\in T\setminus\{x\}$

$$\mathcal{S}_x(y)=\mathcal{S}_x^{(T,c)}(y):=\big\{z\in T\setminus\{x\}:\,z\sim_x y\big\} \tag{2.21}$$

be the equivalence class of $T\setminus\{x\}$ containing y, and note that $\mathcal{S}_x(y)$ is a subtree for all $x,y\in T$, and $\mathcal{S}_x(y)=\mathcal{S}_x(z)$ whenever $z\in\mathcal{S}_x(y)$. We refer to $\mathcal{S}_x(y)$ as the *component* of $T\setminus x$ containing y. Now and in the following, we equip (T,c) with the topology

$$\tau:=\tau\big(\big\{\mathcal{S}_x(y):\,x,y\in T,\ x\neq y\big\}\big) \tag{2.22}$$

generated by the set of components, i.e. with the coarsest topology such that all components are open sets. We call τ the *component topology* of (T,c).

Example 2.14 (on totally ordered trees, τ is the order topology)**.** If $(T,\le)$ is a totally ordered space, and $(T,c_\le)$ the corresponding algebraic tree as in Example 2.5, then τ coincides with the *order topology* (i.e. the one generated by sets of the form $\{y\in T:\,y>x\}$ and $\{y\in T:\,y<x\}$ for $x\in T$). ◊

Example 2.15 (intervals are closed sets)**.** Let (T,c) be an algebraic tree, and $x,y \in T$. Then

$$T \setminus [x,y] = \bigcup \big\{ \mathcal{S}_u(v) : \, u \in [x,y], \, v \in T, \, \mathcal{S}_u(v) \cap [x,y] = \emptyset \big\} \in \tau. \tag{2.23}$$

This means that $[x,y]$ is closed in the component topology τ. $\Diamond$

Lemma 2.16. *Let (T,c) be an algebraic tree. Then c is continuous w.r.t. the component topology τ.*

Proof. By definition of τ, it is sufficient to show that for any $x,y \in T$, $x \neq y$, the set $c^{-1}(\mathcal{S}_x(y))$ is open in T^3. By definition of $\mathcal{S}_x(y)$ and the property $c(u,v,w) \in [u,v] \cap [v,w] \cap [w,u]$ shown in Lemma 2.3, $c(u,v,w) \in \mathcal{S}_x(y)$ if and only if (at least) two of u,v,w are in $\mathcal{S}_x(y)$. Because $\mathcal{S}_x(y)$ is open, the same is true for $\{(u,v,w) \in T^3 : \, u,v \in \mathcal{S}_x(y)\}$ in the product topology. Hence $c^{-1}(\mathcal{S}_x(y))$ is a union of open set and thus open. $\square$

Next, we show that τ is a Hausdorff topology and characterize compactness of algebraic trees in this topology.

Lemma 2.17 (τ is Hausdorff)**.** *Let (T,c) be an algebraic tree. Then the component topology τ defined in* (2.22) *is a Hausdorff topology on T.*

Proof. To show that (T,τ) is Hausdorff, let $x,y \in T$ be distinct. If $\mathcal{S}_y(x) \cap \mathcal{S}_x(y) = \emptyset$, then $\mathcal{S}_y(x)$ and $\mathcal{S}_x(y)$ are clearly disjoint neighbourhoods of x and y, respectively. Assume that this is not the case, and choose $z \in \mathcal{S}_x(y) \cap \mathcal{S}_y(x)$. Then $\rho := c(x,y,z) \notin \{x,y\}$. Furthermore, $c(x,\rho,y) = c(x,y,z) = \rho$, and hence $x \not\sim_\rho y$. Thus $\mathcal{S}_\rho(x)$ and $\mathcal{S}_\rho(y)$ are disjoint neighbourhoods of x and y, respectively. Hence τ is Hausdorff. $\square$

Proposition 2.18 (characterizing compactness)**.** *Let (T,c) be an algebraic tree with component topology τ. Then (T,τ) is compact if and only if (T,c) is directed order complete.*

Proof. "only if". Assume first that (T,c) is not order complete. Then we can choose $\rho \in T$ and $\emptyset \neq A \subseteq T$ such that A is totally ordered w.r.t. $\leq_\rho$ but does not have a supremum in $(T,\leq_\rho)$. For $x,y \in T$, let $U_x := \{z \in T : \, z \not\geq_\rho x\}$ and $V_y := \{z \in T : \, z >_\rho y\}$. Then U_x and V_y are open sets. We claim that $\mathcal{U} := \{U_x : \, x \in A\} \cup \{V_y : \, y \geq A\}$ is an open cover of T. Indeed, if $z \geq_\rho A$, then, because A has no supremum, there is $y \in T$ with $A \leq_\rho y \leq_\rho z$, hence $z \in V_y \in \mathcal{U}$. Otherwise, if $z \not\geq_\rho A$, there is $x \in A$ with $z \in U_x \in \mathcal{U}$. Thus $\mathcal{U}$ is a cover of T.

$\mathcal{U}$ has no finite sub-cover, because if $\mathcal{U}' = \{U_{x_1},...,U_{x_n},V_{y_1},...,V_{y_m}\}$ were such a finite sub-cover, then $\{U_{x_1},...,U_{x_n}\}$ would cover A. This, however, would imply that $\max\{x_1,...,x_n\}$ would be a supremum of A, contradicting our assumption. Hence (T,τ) is not compact.

"if". Assume that (T,c) is order complete. Consider a cover $\mathcal{U}$ of T with components, i.e. $\mathcal{U} \subseteq \{\mathcal{S}_y(x) : \, x,y \in T, \, x \neq y\}$. By the Alexander subbase theorem, for compactness of τ, it is sufficient to show that $\mathcal{U}$ has a finite sub-cover.

To this end, fix an element $\rho \in T$ and consider the set $\mathcal{U}_\rho := \{U \in \mathcal{U} : \, \rho \in U\} \neq \emptyset$. By Hausdorff's maximal chain theorem (or Zorn's lemma), there is a maximal chain I in the partially ordered set $(\mathcal{U}_\rho, \subseteq)$. For every $U \in I$, we have $\rho \in U$, and thus there is $x_U \in T$ such that $U = \mathcal{S}_{x_U}(\rho)$. We claim that $U \subseteq V$ implies $x_U \leq_\rho x_V$. Indeed, $x_V \notin V$ and hence $x_V \notin U$ which is equivalent to $x_V \geq_\rho x_U$. Therefore, $z := \sup\{x_U : \, U \in I\}$ exists in $(T,\leq_\rho)$ by directed order completeness of T. Because $\mathcal{U}$ is a cover, there is $V \in \mathcal{U}$ with $z \in V$, hence $V = \mathcal{S}_y(z)$ for some $y \in T$. Because $V \notin I$ and I is a maximal chain, $y \not\geq_\rho z$. Hence there is $U \in I$ with $y \not\geq_\rho x_U =: x$. We claim that $T = \mathcal{S}_y(z) \cup \mathcal{S}_x(\rho)$. Indeed, let $w \in T \setminus \mathcal{S}_x(\rho)$. Then $w \geq_\rho x$. Using $z \geq_\rho x$ and $c(w,z,y) \in [w,z]$, we obtain $c(w,z,y) \geq_\rho x$, and hence $c(w,z,y) \neq y$. Thus $w \in \mathcal{S}_y(z)$ as claimed, and $\{\mathcal{S}_y(z), \, \mathcal{S}_x(\rho)\}$ is the desired sub-cover. $\square$

It turns out that the following separability condition, which we call order separability, is crucial for us.

Proposition 2.19 (order separability). *Let (T, c) be an algebraic tree with component topology τ. Then the following are equivalent:*

1. *There exists a countable set D such that for all $x, y \in T$ with $x \neq y$,*

$$D \cap [x, y) \neq \emptyset. \tag{2.24}$$

2. *The topological space (T, τ) is second countable (i.e. τ has a countable base), and* $\mathrm{edge}(T, c)$ *is countable.*
3. *The topological space (T, τ) is separable, and* $\mathrm{edge}(T, c)$ *is countable.*

Proof. "3 ⇒ 1". Assume that $\mathrm{edge}(T, c)$ is countable, and that (T, τ) is separable. Then there exists a countable, dense subset $\tilde{D} \subseteq T$. We claim that

$$D := c\big(\tilde{D}^3\big) \cup \big\{z \in T : \exists\, x \in T \text{ such that } \{x, z\} \in \mathrm{edge}(T, c)\big\} \tag{2.25}$$

satisfies (2.24). Indeed, D is countable by assumption. Moreover, let $x, y \in T$. Then two cases are possible: either $\mathcal{S}_x(y) \cap \mathcal{S}_y(x) = \emptyset$. In this case, $\{x, y\} \in \mathrm{edge}(T, c)$, which implies that $[x, y) \cap D \neq \emptyset$. Or $\mathcal{S}_x(y) \cap \mathcal{S}_y(x) \neq \emptyset$. In this case, as $\mathcal{S}_x(y) \cap \mathcal{S}_y(x)$ is open by definition of τ, there is $z \in \tilde{D} \cap \mathcal{S}_x(y) \cap \mathcal{S}_y(x)$. Let $v := c(x, y, z)$. Then $v \in (x, y)$, and either $v = z \in D$, or the three components $\mathcal{S}_v(x)$, $\mathcal{S}_v(y)$, $\mathcal{S}_v(z)$ are distinct. In the second case, we can choose $x' \in \tilde{D} \cap \mathcal{S}_v(x)$ and $y' \in \tilde{D} \cap \mathcal{S}_v(y)$ to see that $v = c(x', y', z) \in D$. In any case, $v \in [x, y) \cap D$.

"1 ⇒ 2". Let D be a countable set satisfying (2.24). Then for all $\{x, y\} \in \mathrm{edge}(T, c)$, $D \cap [x, y) = \{x\}$. This implies that $\mathrm{edge}(T, c)$ is countable. We consider the countable set $\mathcal{U} = \big\{\mathcal{S}_v(u) : u, v \in D\big\} \subseteq \tau$ and claim that it is a subbase for τ (i.e. generates τ). To this end, let $x, y \in T$. We show that $U := \mathcal{S}_x(y)$ is a union of sets from $\mathcal{U}$, i.e. for every $z \in U$ we construct $V \in \mathcal{U}$ with $z \in V \subseteq U$. By assumption on D, there is $v \in D \cap [x, z)$ and $u \in D \cap (v, z]$. Let $V := \mathcal{S}_v(u) \in \mathcal{U}$. Because $c(u, v, z) = u \neq v$, we have $z \in V$. Let $w \in T \setminus U$. Because $u \in U$, we have $U = \mathcal{S}_x(u)$ and therefore $x \in [u, w]$. Similarly, $x \in [v, w]$. In particular, by Lemma 2.2, $c(u, v, w) = c(u, v, c(u, x, w)) = c(u, v, x) = v$, and thus $w \notin V$. Because $w \in T \setminus U$ is arbitrary, we obtain $V \subseteq U$.

"2 ⇒ 3". Trivial, because every second countable topological space is separable. □

Definition 2.20 (order separability). We call an algebraic tree (T, c) *order separable* if the equivalent conditions of Proposition 2.19 are satisfied. We call a set $D \subseteq T$ *order dense* if it satisfies (2.24).

Example 2.21 (uncountable star tree). This example shows that in (2.24) we can not replace $[x, y)$ by $[x, y]$. Let $T := \{0\} \cup [1, 2]$ with $c(x, y, z) := 0$ whenever $x, y, z \in T$ are distinct. Then if $D \subseteq T$ is such that $D \cap [x, 0) \neq \emptyset$ for all $x \in [1, 2]$ then $[1, 2] \subseteq D$, and thus D is uncountable and (T, c) not order separable. On the other hand, $D := \{0\}$ has the property that $D \cap [x, y] \neq \emptyset$ for all $x, y \in T$ with $x \neq y$. ◊

An order complete, order separable algebraic tree is, in its component topology τ, a compact, second countable Hausdorff space by Propositions 2.18 and 2.19. In particular, it is metrizable. In fact, order separability already implies metrizability, as we will see in Subsection 2.4. The following example shows that (topological) separability of (T, τ) alone, without requiring the number of edges to be countable, is neither sufficient for order separability nor for metrizability of (T, τ).

Example 2.22 (a continuum ladder). Let $T = [0, 1] \times \{0, 1\}$ with the lexicographic order $\leq$ on T, and define the canonical branch point map $c_{\leq}$ as in Example 2.5. Then $\mathrm{edge}(T, c_{\leq}) = \big\{\{(x, 0), (x, 1)\} : x \in [0, 1]\big\}$ is uncountable, and hence $(T, c_{\leq})$ is not order separable. Because $(\mathbb{Q} \cap [0, 1]) \times \{0, 1\}$ is a countable dense set, (T, τ) is (topologically) separable. The topological subspace $[0, 1] \times \{1\}$ is the *Sorgenfrey line*, which is known to be non-metrizable (see [SS78, Counterexample 51]). Thus also (T, τ) cannot be metrizable. ◊

Definition 2.23 (Borel σ-algebra $\mathcal{B}(T,c)$)**.** Let (T,c) be an algebraic tree. We denote the Borel σ-algebra of the component topology τ by $\mathcal{B}(T,c)$ and call it *Borel σ-algebra of* (T,c).

In general, $\mathcal{B}(T,c)$ is not generated by the set of components. Order separability, however, is sufficient to ensure this property because it implies second countability of the component topology.

Corollary 2.24 (Borel σ-algebra generated by components)**.** *Let (T,c) be an order separable algebraic tree, and $D \subseteq T$ an order dense set. Then its Borel σ-algebra is generated by the set of components indexed by D, i.e.*

$$\mathcal{B}(T,c) = \sigma\big(\{\mathcal{S}_x(y) : x,y \in D,\ x \neq y\}\big). \tag{2.26}$$

Proof. Define $\mathcal{U} := \{\mathcal{S}_x(y) : x,y \in D,\ x \neq y\}$. By Proposition 2.19, (T,τ) is second countable. Hence $\mathcal{B}(T,c)$ is generated by any subbase of τ. If D is order dense, $\mathcal{U}$ is such a subbase as shown in the proof of Proposition 2.19. □

2.4. **Metric tree representations of algebraic trees.** In this subsection, we discuss the connection of metric trees with algebraic trees. Let (T,r) be a metric tree (recall from Definition 1.2). Then by (MT2), there exists to any three points $x_1,x_2,x_3 \in T$ a unique branch point $c_{(T,r)}(x_1,x_2,x_3)$ satisfying (1.5). We refer to $(T,c_{(T,r)})$ as the algebraic tree *induced by* (T,r), and to (T,r) as a *metric representation* of $(T,c_{(T,r)})$.

Lemma 2.25 (the algebraic tree induced by a metric tree)**.** *Let (T,r) be a metric tree, and $c_{(T,r)}$ the map which sends any three distinct points to their branch point. Then the following hold:*

- *(i)* $(T,c_{(T,r)})$ *is an algebraic tree.*
- *(ii)* $(T,c_{(T,r)})$ *is order separable if and only if* (T,r) *is separable.*
- *(iii)* $(T,c_{(T,r)})$ *is directed order complete if and only if* (T,r) *is bounded and complete. In particular, it is an algebraic continuum tree if and only if* (T,r) *is a bounded, complete* $\mathbb{R}$*-tree.*

Proof. (i) It can be easily checked that $(T,c_{(T,r)})$ is an algebraic tree.

(ii) Let (T,r) be separable. Then $\mathrm{edge}(T,c_{(T,r)})$ is countable. The topology induced by r is obviously stronger than the topology τ introduced in (2.22), hence τ is separable and therefore the algebraic tree $(T,c_{(T,r)})$ is order separable. Conversely, if $(T,c_{(T,r)})$ is order separable, then any countable set D satisfying (2.24) is also dense in (T,r).

(iii) Clearly, $(T,c_{(T,r)})$ admits suprema along any linearly ordered set with respect to some root if and only if (T,r) is bounded and complete. The "in particular" follows because a complete metric tree (T,r) is an $\mathbb{R}$-tree if and only if $\mathrm{edge}(T,r) := \mathrm{edge}(T,c_{(T,r)}) = \emptyset$ ([ALW17, Remark 1.2]). □

Our first main result states that under the assumption of order separability any algebraic tree can be embedded by an injective homomorphism into a compact $\mathbb{R}$-tree and hence is isomorphic to (the algebraic tree induced by) a totally bounded metric tree.

Theorem 1 (characterisation of order separable algebraic trees)**.** *Let T be a set, $c\colon T^3 \to T$.*

- *(i)* (T,c) *is an order separable algebraic continuum tree if and only if there exists a metric r on T such that (T,r) is a compact $\mathbb{R}$-tree with*

$$c = c_{(T,r)}. \tag{2.27}$$

- *(ii)* (T,c) *is an order separable algebraic tree if and only if there is an order separable algebraic continuum tree $(\overline{T},\bar{c})$ such that (T,c) is a subtree of $(\overline{T},\bar{c})$. In particular, every order separable algebraic tree is induced by a totally bounded metric tree.*

The separability hypothesis in Theorem 1 is crucial and cannot be dropped. In Example 2.22, we have already seen an algebraic tree where the component topology τ is not metrizable. Moreover, in this example, τ coincides with the order topology which is also the case for the metric topology of any metric tree without branch points. Thus the algebraic tree cannot be induced by a metric tree. The following example shows that also algebraic continuum trees need not be induced by metric trees.

Example 2.26 (algebraic continuum tree that is not induced by a metric tree)**.** Let $T = [0,1] \times [0,1]$ with lexicographic order, and (T,c) the corresponding algebraic tree as in Example 2.5. It is easy to check that (T,c) is an algebraic continuum tree. It cannot be induced by a metric tree because in its order topology τ, it is connected but not path-wise connected. These two properties are equivalent for metric trees (see [Eva08, Theorem 2.20]). $\Diamond$

In order to prove Theorem 1, given an algebraic tree (T,c), we need to provide a metric r such that (2.27) holds. For that purpose, we consider for any measure ν on $(T,\mathcal{B}(T,c))$ such that ν is finite on every interval, the following pseudometric,

$$r_\nu(x,y) := \nu([x,y]) - \tfrac12\nu(\{x\}) - \tfrac12\nu(\{y\}), \quad x,y \in T. \tag{2.28}$$

Lemma 2.27 (r_ν is a pseudometric)**.** *Let (T,c) be an algebraic tree, and ν a measure on (T,c) with $\nu([x,y]) < \infty$ for all $x,y \in T$. Then r_ν is a pseudometric on T.*

Proof. By Lemma 2.3 for all $x,y,z \in T$,

$$\begin{aligned} \nu\big([x,y]\big) + \nu\big([y,z]\big) &= \nu\big([x,y] \cup [y,z]\big) + \nu\big([x,y] \cap [y,z]\big) \\ &= \nu\big([x,z]\big) + \nu\big(\big(c(x,y,z),\, y\big]\big) + \nu\big(\big[c(x,y,z),\, y\big]\big). \end{aligned} \tag{2.29}$$

Hence

(2.30)

$$\begin{aligned} &r_\nu(x,y) + \tfrac12\nu\{x\} + \tfrac12\nu\{y\} + r_\nu(y,z) + \tfrac12\nu\{y\} + \tfrac12\nu\{z\} \\ &\quad = r_\nu(x,z) + \tfrac12\nu\{x\} + \tfrac12\nu\{z\} + 2r_\nu\big(c(x,y,z),\, y\big) + \nu\{c(x,y,z)\} + \nu\{y\} - \nu\{c(x,y,z)\}, \end{aligned}$$

or equivalently,

$$r_\nu(x,y) + r_\nu(y,z) = r_\nu(x,z) + 2r_\nu\big(c(x,y,z),\, y\big). \tag{2.31}$$

This implies that r_ν satisfies the triangle inequality. $\square$

We denote the quotient metric space by (T_ν, r_ν), i.e. T_ν is the set of equivalence classes of points in T with r_ν-distance zero, and the quotient metric on T_ν is again denoted by r_ν. Furthermore, let $\pi_\nu \colon T \to T_\nu$ be the canonical projection.

Lemma 2.28 ((T_ν, r_ν) is a metric tree)**.** *Let (T,c) be an algebraic tree, and ν a measure on (T,c) with $\nu([x,y]) < \infty$ for all $x,y \in T$. Then the quotient space (T_ν, r_ν) is a metric tree, and the canonical projection π_ν is a tree homomorphism.*

Proof. Let $x_1,\ldots,x_4 \in T$. By Condition (BPM4), we can assume w.l.o.g. that $c(x_1,x_2,x_3) = c(x_1,x_2,x_4)$. Then by Lemma 2.2(ii), $c(x_1,x_2,x_3) \in [x_1,x_2] \cap [x_1,x_3] \cap [x_2,x_3] \cap [x_1,x_4] \cap [x_2,x_4]$, and (2.31) yields that for $\{i,j\} \in \{\{1,2\},\ \{1,3\},\ \{1,4\},\ \{2,3\},\ \{2,4\}\}$,

$$r_\nu(x_i,x_j) = r_\nu\big(x_i, c(x_1,x_2,x_3)\big) + r_\nu\big(c(x_1,x_2,x_3), x_j\big). \tag{2.32}$$

Therefore,

(2.33)

$$\begin{aligned} &r_\nu(x_1,x_3) + r_\nu(x_2,x_4) \\ &= r_\nu\big(x_1,\, c(x_1,x_2,x_3)\big) + r_\nu\big(c(x_1,x_2,x_3),\, x_3\big) + r_\nu\big(x_2,\, c(x_1,x_2,x_3)\big) + r_\nu\big(c(x_1,x_2,x_3),\, x_4\big) \\ &= r_\nu(x_1,x_2) + r_\nu\big(c(x_1,x_2,x_3),\, x_3\big) + r_\nu\big(c(x_1,x_2,x_3),\, x_4\big) \\ &\ge r_\nu(x_1,x_2) + r_\nu(x_3,x_4), \end{aligned}$$

and analogously,

$$\begin{aligned} r_\nu(x_1,x_4)+r_\nu(x_2,x_3) &= r_\nu\big(x_3,\, c(x_1,x_2,x_3)\big)+r_\nu\big(c(x_1,x_2,x_3),\, x_4\big)+r_\nu(x_1,x_2) \\ &\ge r_\nu(x_1,x_2)+r_\nu(x_3,x_4). \end{aligned} \tag{2.34}$$

This means that the four point Condition (MT1) is satisfied. Moreover, (2.32) implies Condition (MT2) with branch point $\pi_\nu(c(x_1,x_2,x_3))$. In particular, π_ν is a tree homomorphism. □

Remark 2.29. Lemma 2.28 also explains why we had defined r_ν as in (2.28) and not just as $r'_\nu := \nu([x,y])$ for $x \neq y$. Namely, in the latter case we would still have (MT1), but (MT2) might fail. Take, for example, $T := \{1,2,3\}$, $c(1,2,3) = 2$, and $\nu = \delta_2$. In this case, r'_ν is the discrete metric on T, thus 2 does not lie on the interval $[1,3]$ anymore. ◊

Let (T,c) be an algebraic tree. For all $v \in T$, define the *degree* of v in (T,c) by

$$\deg(v) := \deg_{(T,c)}(v) := \#\big\{\mathcal{S}_v(y) : y \in T\big\}. \tag{2.35}$$

We say that $v \in T$ is a *leaf* if $\deg_{(T,c)}(v) = 1$, and a *branch point* if $\deg_{(T,c)}(v) \ge 3$. Notice that

$$\mathrm{lf}(T,c) := \big\{u \in T : c(u,v,w) \neq u \;\forall v,w \in T \setminus \{u\}\big\} \tag{2.36}$$

equals the set of leaves of T, and

$$\mathrm{br}(T,c) := \big\{u \in T : c(x,v,w) = u \;\text{ for some } x,v,w \in T \setminus \{u\}\big\} \tag{2.37}$$

the set of branch points. Moreover, note that any ν-mass on $\mathrm{lf}(T,c)$ that is not atomic does not contribute to r_ν.

Proposition 2.30 (metric representations of algebraic trees)**.** *Let (T,c) be an algebraic tree, ν a measure on $(T,\mathcal{B}(T,c))$ with $\nu([x,y]) < \infty$ for all $x,y \in T$, and r_ν defined by (1.11). Then the following hold:*

- *(i) If (T,c) is order separable and ν has at most countably many atoms, then (T_ν,r_ν) is separable.*
- *(ii) If $\#T > 1$, (T,c) is order complete, and $[x,y]$ is order separable for every $x,y \in T$, then (T_ν,r_ν) is connected if and only if ν is non-atomic. In this case, (T_ν,r_ν) is a complete $\mathbb{R}$-tree.*

Proof. Throughout the proof denote by $\pi_\nu : T \to T_\nu$ the canonical projection.

(i) It is easy to see that if a set $A \subseteq T$ satisfies (2.24) and contains all atoms of ν, then $\pi_\nu(A)$ is dense in (T_ν,r_ν). Therefore, by Proposition 2.19 order separability of (T,c) implies separability of (T_ν,r_ν).

(ii) For all $x,y \in T$ with $x \neq y$, $r_\nu(x,y) \ge \frac{1}{2}\nu\{x\}$. Hence (T_ν,r_ν) cannot be connected if ν has atoms. Conversely, assume that ν is non-atomic. For $x,z \in T$, consider $([x,z],\le_x)$, which is a totally ordered space according to Lemma 2.6, and define $y := \sup\{v \in [x,z] : 2\nu([x,v]) \le \nu([x,z])\}$. The supremum exists due to order completeness of (T,c). Because of the order separability of $[x,z]$ and the non-atomicity of ν, we obtain $2\nu([x,y]) = \nu([x,z]) = 2\nu([y,z])$ and therefore $2r_\nu(x,y) = r_\nu(x,z) = 2r_\nu(y,z)$. From this equality, connectedness follows once we have shown completeness, and every connected metric tree is an $\mathbb{R}$-tree.

Recall from Lemma 2.28 that (T_ν,r_ν) is a metric tree. The same holds for its metric completion $\overline{T}_\nu$. Assume for a contradiction that there is a sequence $(x_n)_{n\in\mathbb{N}}$ in T_ν converging to some $x \in \overline{T}_\nu \setminus T_\nu$. Then x cannot be a branch point and one of the at most two components of $\overline{T}_\nu \setminus \{x\}$ contains infinitely many x_n. Thus we may assume w.l.o.g. that $x \in \mathrm{lf}(\overline{T}_\nu)$. Define $y_n := c_{\overline{T}_\nu}(x_1,x_n,x)$. Then $y_n \to x$ and, for large enough m, we have $y_n = c_{\overline{T}_\nu}(x_1,x_n,x_m)$. Hence $y_n \in T_\nu$ for all $n \in \mathbb{N}$ and we may choose representatives $x'_n \in \pi_\nu^{-1}(y_n)$ such that $x'_n = c(\rho,x'_n,x'_m)$ for $\rho := x'_1$ and all sufficiently large m. By Lemma 2.6, $\{x'_n : n \in \mathbb{N}\}$ is totally ordered w.r.t. $\le_\rho$, and hence $x' := \sup\{x'_n : n \in \mathbb{N}\} \in T$ exists by order completeness. Obviously, $\pi_\nu(x') = x$ and $x \in T_\nu$. □

In order to prove Theorem 1(i) using Proposition 2.30, we need a non-atomic probability measure ν (to ensure connectedness of (T_ν, r_ν)) charging all intervals (so that π_ν is injective). Such a measure always exists in the case of *order separable* algebraic continuum trees.

Lemma 2.31. *Let (T,c) be an order separable algebraic continuum tree with $\#T > 1$. Then there exists a non-atomic probability measure ν on $(T, \mathcal{B}(T,c))$ with $\nu(\mathrm{lf}(T,c)) = 0$ and*

$$\nu\big([x,y]\big) > 0 \quad \forall\, x,y \in T,\ x \neq y. \tag{2.38}$$

Proof. Fix $\rho \in T$. Then, for every $x \in T \setminus \{\rho\}$, the interval $([\rho,x], \leq_\rho)$ is a separable *linear continuum* in the sense of order theory, i.e. a totally ordered space (proven in Lemma 2.6) without *jumps* (what we call here edges) or *gaps* (which follows from directed order completeness). Due to Cantor's order characterisation of $\mathbb{R}$ (e.g. [Das14, Theorem 560]), this means that $[\rho,x]$ is order isomorphic to the unit interval. Obviously, every order isomorphism is measurable and bijective, and the image of Lebesgue measure on the unit interval is a non-atomic probability measure ν_x on $[\rho,x]$. Then $\sum_{n\in\mathbb{N}} 2^{-n}\nu_{x_n}$, where $\{x_n : n \in \mathbb{N}\}$ satisfies (2.24), is a non-atomic probability satisfying (2.38) and $\nu(\mathrm{lf}(T,c)) = 0$. □

Any separable $\mathbb{R}$-tree (T,r) comes with an intrinsic measure, called length measure, that generalizes the Lebesgue-measure on $\mathbb{R}$. More generally, if (T,r) is a complete, separable metric tree and $\rho \in T$ a fixed root, the *length measure* $\lambda = \lambda^{(T,r,\rho)}$ is uniquely defined by the two properties $\lambda([\rho,x]) = r(\rho,x)$ for all $x \in T$, and $\lambda(\mathrm{lf}_0(T,r)) = 0$, where lf_0 is the set of non-isolated leaves (see [ALW17, Section 2.1]). Note that the total mass $\lambda(T)$ (the "total length" of the metric tree) does not depend on the choice of ρ.

Proposition 2.32 (total length of (T_ν, r_ν))**.** *Let (T,c) be an order separable, order complete algebraic tree, ν a measure on $(T,\mathcal{B}(T,c))$ with $\nu([x,y]) < \infty$ for all $x,y \in T$ and such that $\nu\restriction_{\mathrm{lf}(T,c)}$ is purely atomic, and r_ν be defined by (1.11). Then the following hold:*

(i) The total length of the metric tree (T_ν, r_ν) is given by

$$\lambda(T_\nu) = \tfrac{1}{2}\int_T \deg_{(T,c)}\, \mathrm{d}\nu. \tag{2.39}$$

(ii) $\int_T \deg_{(T,c)}\, \mathrm{d}\nu = \int_{T_\nu} \deg_{(T_\nu,r_\nu)} \circ \pi_\nu\, \mathrm{d}\nu$.

Proof. (i) Let $D := \{v_n : n \in \mathbb{N}\}$ be a subset of (T,c) which contains the atoms of ν and satisfies (2.24), and $\pi_\nu \colon T \to T_\nu$ be the canonical projection. We use $\rho := \pi_\nu(v_1)$ as the root of (T_ν, r_ν). Then

$$T \setminus \mathrm{lf}(T,c) \subseteq [\![D]\!] = \bigcup_{n\in\mathbb{N}} [\![v_1,...,v_n]\!], \tag{2.40}$$

where $[\![A]\!] := \bigcup_{x,y\in A}[x,y]$. Hence $\nu(T \setminus [\![D]\!]) = 0$, and

$$\lambda^{(T_\nu,r_\nu,\rho)}(T_\nu) = \lim_{n\to\infty} \lambda^{(T_\nu,r_\nu,\rho)}\big(\pi_\nu([\![v_1,...,v_n]\!])\big). \tag{2.41}$$

Abbreviate $T_n := [\![v_1,...,v_n]\!]$ and $\ell_n := \lambda^{(T_\nu,r_\nu,\rho)}\big(\pi_\nu([\![v_1,...,v_n]\!])\big)$. If $v_{n+1} \in T_n$, then $T_{n+1} = T_n$ and $\lambda^{(T_\nu,r_\nu,\rho)}\big(\pi_\nu(T_{n+1})\big) = \lambda^{(T_\nu,r_\nu,\rho)}\big(\pi_\nu(T_n)\big)$. Otherwise, there exists a unique $u_n \in T$ with $T_{n+1} = T_n \uplus (u_n, v_{n+1}]$, and thus

$$\ell_{n+1} = \ell_n + r_\nu(u_n, v_{n+1}) = \ell_n + \nu\big((u_n, v_{n+1}]\big) - \tfrac{1}{2}\nu\{v_{n+1}\} + \tfrac{1}{2}\nu\{u_n\}. \tag{2.42}$$

For $v \in T_n$, let $\deg_n(v)$ be the degree of v in the tree $(T_n, c\restriction_{T_n})$. In the case $v_{n+1} \notin T_n$, we have $\deg_{n+1}(v) = \deg_n(v)$ for $v \in T_n \setminus \{u_n\}$, and $\deg_{n+1}(u_n) = \deg_n(u_n) + 1$. By induction over n, we obtain

$$\ell_n = \tfrac{1}{2}\int_{T_n} \deg_n\, \mathrm{d}\nu \tag{2.43}$$

Note that $\deg_n(v)$ is monotonically increasing in n, and $\deg(v) = \lim_{n\to\infty} \deg_n(v)$ holds for all $v \in [\![D]\!]$. Thus using the monotone convergence theorem, combining (2.41) and (2.43) yields (2.39).

(ii) If $\deg_{(T,c)}(v) \neq \deg_{(T_\nu,r_\nu)}(\pi_\nu(v))$, then either $\pi(\mathcal{S}_v(y)) = \{\pi(v)\}$ for some $y \in T$ (and thus $\deg_{(T,c)}(v) > \deg_{(T_\nu,r_\nu)}(\pi_\nu(v))$), or $\pi(v) = \pi(v')$ for some $v' \in \mathrm{Br}(T,c)$ (and thus $\deg_{(T,c)}(v) < \deg_{(T_\nu,r_\nu)}(\pi_\nu(v))$). In both cases we have that $\nu\{v\} = \nu\{\pi_\nu(v)\} = 0$, and thus the claim follows. $\square$

Corollary 2.33 (compactness for bounded degree trees)**.** *Let (T,c) be an order separable algebraic tree, and ν a finite measure on $(T,\mathcal{B}(T,c))$ with $\nu\{v \in T : \deg(v) > d\} = 0$ for some $d \in \mathbb{N}$. Then the completion of (T_ν, r_ν) is compact.*

Proof. W.l.o.g. assume that $\nu\restriction_{\mathrm{lf}(T,c)}$ is non-atomic (if $\nu\restriction_{\mathrm{lf}(T,c)}$ has a non-atomic part, we can remove it without changing r_ν). Then by Proposition 2.32(i), (T_ν, r_ν) has finite total length. As complete metric trees with finite total length are necessarily compact, the statement follows. $\square$

We are now in a position to prove Theorem 1.

Proof of Theorem 1. (i) "$\Longleftarrow$" Since every compact metric space is bounded, complete and separable, this step follows from Lemma 2.25.

"$\Longrightarrow$" Let (T,c) be an order separable algebraic continuum tree. To avoid trivialities, assume that T contains more than two points. By Lemma 2.31 we can choose a non-atomic probability measure ν on $(T,\mathcal{B}(T,c))$ satisfying (2.38). Define r_ν by (1.11). Then the equivalence classes in T_ν are singletons by (2.38), and we may identify T_ν with T.

By Proposition 2.30, (T, r_ν) is a complete $\mathbb{R}$-tree and the identity is a tree homomorphism by Lemma 2.28. Thus c is induced by r_ν. Moreover, $\nu(\mathrm{br}(T,c)) = 0$ because $\mathrm{br}(T,c)$ is countable and ν is non-atomic. We can therefore conclude with Corollary 2.33 that (T, r_ν) is also compact.

(ii) "$\Longleftarrow$" This is obvious because every order separable algebraic continuum tree is induced by a separable $\mathbb{R}$-tree according to part(i), and subspaces of separable metric spaces are separable.

"$\Longrightarrow$" Let (T,c) be an order separable algebraic tree and $D \subseteq T$ a countable set satisfying (2.24). Let ν be any probability measure on D with $\nu\{x\} > 0$ for all $x \in D$, and r_ν defined by (1.11). The equivalence classes in T_ν are singletons, and we may again identify T_ν with T. By Proposition 2.32, (T, r_ν) is a metric tree with (2.27). As (T,c) is order separable, (T, r_ν) is separable by Proposition 2.30(i). Moreover, the diameter of (T, r_ν) is bounded by 1. Hence, by [Eva08, Theorem 3.38] there is a bounded, separable $\mathbb{R}$-tree $(\overline{T}, \bar{r})$ such that $T \subseteq \overline{T}$ and r_ν is the restriction of $\bar{r}$ to T. By Lemma 2.25, this $\mathbb{R}$-tree induces an algebraic continuum tree $(\overline{T}, \bar{c})$, and T is a subtree of $\overline{T}$.

"in particular". According to part *(i)*, there is a metric $\tilde{r}$ on $\overline{T}$ such that $(\overline{T}, \tilde{r})$ is a compact $\mathbb{R}$-tree inducing $(\overline{T}, \bar{c})$. Let r be the restriction of $\tilde{r}$ to T. Then (T, r) is a totally bounded metric tree inducing (T,c). $\square$

2.5. **Tree homomorphisms versus homeomorphisms.** Since order separable algebraic continuum trees are $\mathbb{R}$-trees where we have "forgotten" the metric, the question arises how homeomorphisms of $\mathbb{R}$-trees relate to tree homomorphisms of the corresponding algebraic trees. A first observation is that homeomorphisms are necessarily tree homomorphisms. This statement relies on connectedness of the $\mathbb{R}$-trees and we cannot replace "$\mathbb{R}$-tree" by "metric tree": every bijection between finite metric trees is obviously a homeomorphism because the topologies are discrete, but not necessarily a tree homomorphism.

Lemma 2.34 (homeomorphisms are tree isomorphisms)**.** *Let (T,r), $(\widehat{T}, \hat{r})$ be $\mathbb{R}$-trees, and $f\colon T \to \widehat{T}$ a homeomorphism. Then f is a tree homomorphism.*

Proof. The branch point map can be expressed in terms of intervals by (1.2). In an $\mathbb{R}$-tree (T, r), the interval $[x, y]$, $x, y \in T$, is the unique simple path from x to y, which is a purely topological notion, and hence preserved by homeomorphisms. $\square$

Example 2.35 (tree isomorphisms need not be homeomorphisms)**.** In Lemma 2.34, the converse is not true: bijective tree homomorphisms need not be homeomorphisms, even if the trees are order separable. To see this, let $r, \hat{r}$ the metrics on $\mathbb{N}$ defined by $r(n, m) = \frac{1}{n} + \frac{1}{m}$, $\hat{r}(n, m) = 2$ for distinct $n, m \in \mathbb{N}$. Let T and $\widehat{T}$ be the $\mathbb{R}$-trees generated by $(\mathbb{N}, r)$ and $(\mathbb{N}, \hat{r})$, respectively. Then both $\widehat{T}$ and T are the countable star with set $\mathbb{N}$ of leaves. In T, the distance from the branch point to leaf n is $\frac{1}{n}$, while it is 1 in $\widehat{T}$. Hence T is compact while $\widehat{T}$ is not. The identity on $\mathbb{N}$ can be extended to a bijective tree homomorphism $f : T \to \widehat{T}$ which cannot be continuous. $\Diamond$

Example 2.35 shows that it is possible for non-homeomorphic (topologically non-equivalent) $\mathbb{R}$-trees to induce isomorphic (equivalent) algebraic continuum trees. This can only happen if at least one of the trees is non-compact.

Proposition 2.36 (tree isomorphisms of compact $\mathbb{R}$-trees are homeomorphisms)**.** *Let $T, \widehat{T}$ be $\mathbb{R}$-trees, and $f : T \to \widehat{T}$.*

(i) If $\widehat{T}$ is compact, $f(T)$ is connected, and f a tree homomorphism, then f is continuous.
(ii) If both T and $\widehat{T}$ are compact and f is bijective, then f is a homeomorphism if and only if it is a tree homomorphism.

Proof. (ii) is obvious from (i) and Lemma 2.34.

Assume f is a tree homomorphism, $f(T)$ is connected, and $\widehat{T}$ is compact. Choose a root $\rho \in T$. Let $v_n \to v$ be a convergent sequence in T, and $w \in \widehat{T}$ an accumulation point of $f(v_n)$. Then there is a subsequence $(n_k)_{k \in \mathbb{N}}$ with $f(v_{n_k}) \to w$. We have

$$v = \sup_{k \in \mathbb{N}} \inf_{i > k} v_{n_i} \qquad \text{and} \qquad w = \sup_{k \in \mathbb{N}} \inf_{i > k} f(v_{n_i}), \tag{2.44}$$

where sup and inf are w.r.t. the partial orders $\leq_\rho$ and $\leq_{f(\rho)}$ in the first and second equality, respectively. In the following, we show $w = f(v)$. Because f is order preserving for these partial orders due to Lemma 2.11, we obtain $w \leq_{f(\rho)} f(v)$. Assume for a contradiction $w \neq f(v)$. Because $f(T)$ is connected, there is $y \in \widehat{T}$ with $w <_{f(\rho)} y <_{f(\rho)} f(v)$ and $x \in T$ with $y = f(x)$. For $u := c(\rho, x, v)$, we have $f(u) = \hat{c}(f(\rho), y, f(v)) = y$, $u \leq_\rho v$, and $u \neq v$. Therefore, $u \leq_\rho v_{n_i}$ for all sufficiently large i, and thus $y = f(u) \leq_{f(\rho)} f(v_{n_i})$ for those i. Now (2.44) implies $y \leq_{f(\rho)} w$ in contradiction to the choice of y, finishing the proof of $w = f(v)$. Compactness of $\widehat{T}$ and uniqueness of accumulation points implies $f(v_n) \to f(v)$, and f is continuous. $\square$

In view of Theorem 1, Proposition 2.36 implies that order separable algebraic continuum trees are in one-to-one correspondence with homeomorphism classes of compact $\mathbb{R}$-trees. Furthermore, the unique metric topology induced by the compact $\mathbb{R}$-tree coincides with the component topology τ introduced in Subsection 2.3. But be aware that there may be other, non-homeomorphic, non-compact $\mathbb{R}$-trees inducing the same order separable algebraic continuum tree, as shown in Example 2.35.

Corollary 2.37 (uniqueness of inducing $\mathbb{R}$-tree)**.** *Every order separable algebraic continuum tree is induced by a compact $\mathbb{R}$-tree that is unique up to homeomorphism, and the unique induced topology coincides with the component topology τ defined in* (2.22)*.*

Proof. That an order separable algebraic continuum tree is induced by a compact $\mathbb{R}$-tree is Theorem 1(i). Any two such compact $\mathbb{R}$-trees are isomorphic as algebraic trees, hence homeomorphic by Proposition 2.36. The component topology is a Hausdorff topology and clearly weaker than the topology induced by the $\mathbb{R}$-tree, because components are open sets of $\mathbb{R}$-trees. Hence, by compactness of the $\mathbb{R}$-tree, the two topologies coincide. $\square$

3. The space of algebraic measure trees

In this section, we define algebraic measure trees, and equip the space of (equivalence classes of) algebraic measure trees with a topology. In what follows, the order separability of the underlying algebraic tree is crucial. Therefore, we include it already in the following definition of algebraic measure trees.

Definition 3.1 (algebraic measure trees)**.** An *algebraic measure tree* (T, c, μ) is an order separable algebraic tree (T, c) together with a probability measure μ on $\mathcal{B}(T, c)$.

Definition 3.2 (equivalence of algebraic measure trees)**.** (i) We call two algebraic measure trees (T_i, c_i, μ_i), $i = 1, 2$, *equivalent* if there exist subtrees A_i of T_i with $\mu_i(A_i) = 1$, and a measure preserving tree isomorphism f from A_1 onto A_2. In this case, we call f *isomorphism* of the algebraic measure trees.

(ii) A metric measure tree (T, r, μ) is called a *metric representation* of the algebraic measure tree (T', c', μ') if its induced algebraic measure tree $(T, c_{(T,r)}, \mu)$ is equivalent to (T', c', μ').

In the following, we denote for an algebraic measure tree $\chi := (T, c, \mu)$ by $\mathrm{supp}(\chi)$ the algebraic subtree generated by the support of μ, i.e.

$$\mathrm{supp}(\chi) := c\big(\mathrm{supp}(\mu)^3\big), \tag{3.1}$$

and by

$$\mathrm{br}(\chi) := \mathrm{br}(T, c) \cap \mathrm{supp}(\chi) \tag{3.2}$$

the set of *branch points* of χ. It is easy to check that an isomorphism f from $\chi = (T, c, \mu)$ to $\chi' = (T', c', \mu')$ induces a bijection between $\mathrm{br}(\chi)$ and $\mathrm{br}(\chi')$ (although it need neither be defined nor injective on all of $\mathrm{supp}(\chi)$). Also note that χ is equivalent to $\mathrm{supp}(\chi)$ equipped with the appropriate restrictions of c and μ.

Remark 3.3 (a note on our definition of equivalence)**.** Every algebraic measure tree is equivalent to an algebraic continuum measure tree, and has a metric representation with a compact $\mathbb{R}$-tree by Theorem 1. For the definition of equivalence of algebraic measure trees it is important that we do not require the whole trees to be isomorphic (see Example 3.11 below). On the other hand, it is also important that the isomorphism is injective on a subtree (as opposed to only a subset) of full measure, because otherwise it would not be an equivalence relation and every tree with n leaves and uniform distribution on them would be equivalent to the n-star. $\Diamond$

Example 3.4 (the linear non-atomic measure tree)**.** There is only one equivalence class of linearly ordered algebraic measure trees with non-atomic measure. Indeed, let (T, c, μ) be an algebraic measure tree with $\mathrm{br}(T, c) = \emptyset = \mathrm{at}(\mu)$. Then, by Theorem 1, there is a tree isomorphism from T into $[0, 1]$ and we may assume $T \subseteq [0, 1]$ to begin with. Let $F_\mu \colon [0, 1] \to [0, 1]$ be the distribution function of μ. Then F_μ is continuous and maps μ to Lebesgue-measure $\lambda_{[0,1]}$. Let $A := \{x \in \mathrm{supp}(\mu) : \text{there is no } y_n \in [0, 1] \setminus \mathrm{supp}(\mu) : y_n < x,\ y_n \to x\}$ be the support of μ with left boundary points removed. Then F_μ restricted to A is bijective and hence a measure preserving tree isomorphism onto $[0, 1]$ (with Lebesgue measure and canonical branch point map). Thus (T, c, μ) is equivalent to $[0, 1]$. $\Diamond$

Let

$$\mathbb{T} := \{\text{equivalence classes of algebraic measure trees}\}. \tag{3.3}$$

Next, we equip $\mathbb{T}$ with a topology. We shall base this notion of convergence on the fact that algebraic measure trees allow for metric representations (see Theorem 1), and require convergence in Gromov-weak topology of particular representations. To this end, let

$$\mathbb{H} := \{\text{equivalence classes of (separable) metric measure trees}\}, \tag{3.4}$$

where we consider two metric measure trees (T, r, μ) and (T', r', μ') as *equivalent* if there exists a measure preserving isometry between the metric completions of $\mathrm{supp}(\mu)$ and $\mathrm{supp}(\mu')$.

In order to get a useful topology on $\mathbb{T}$, we cannot take arbitrary (optimal) metric representations. Instead, given an algebraic measure tree (T, c, μ), we use the metric r_ν defined in (2.28) for the *branch point distribution* ν, namely the distribution of the random branch point obtained by sampling three points with the sampling measure μ.

Definition 3.5 (branch point distribution)**.** The *branch point distribution* of an algebraic measure tree (T, c, μ) is the push-forward of $\mu^{\otimes 3}$ under the branch point map,

$$\nu := c_*\mu^{\otimes 3}. \tag{3.5}$$

Note that the branch point distribution is not necessarily supported by $\mathrm{br}(T, c)$. For instance, every atom of μ is also an atom of ν. If (T, c, μ) and (T', c', μ') are equivalent algebraic measure trees with branch point distributions ν and ν', respectively, then the isomorphism is also an isometry w.r.t. r_ν and $r_{\nu'}$. Therefore, the following selection map, which associates a particular metric representation to every algebraic measure tree, is well-defined.

Definition 3.6 (selection map ι)**.** Define the map $\iota\colon \mathbb{T} \to \mathbb{H}$ by

$$\iota(T, c, \mu) := (T_\nu, r_\nu, \mu_\nu), \tag{3.6}$$

where $\nu = c_*\mu^{\otimes 3}$ is the branch point distribution of (T, c, μ), (T_ν, r_ν) is the quotient metric space, and μ_ν is the image of μ under the canonical projection π_ν.

The topology we use on $\mathbb{T}$ is the Gromov-weak topology w.r.t. the branch point distribution distance. That is, it is the topology induced by the selection map ι, i.e., the weakest (coarsest) topology on $\mathbb{T}$ such that ι is continuous.

Definition 3.7 (bpdd-Gromov-weak topology)**.** Let $\mathbb{H}$ be equipped with the Gromov-weak topology. We call the topology induced on $\mathbb{T}$ by the selection map ι *branch point distribution distance Gromov-weak topology* (*bpdd-Gromov-weak topology*).

The following reconstruction theorem is crucial for the usefulness of bpdd-Gromov-weak convergence. It shows that the selection map ι is an embedding and indeed selects metric representations.

Proposition 3.8 (ι is injective)**.** *The selection map $\iota\colon \mathbb{T} \to \mathbb{H}$ is injective, and $\iota(\chi)$ is a metric representation of $\chi \in \mathbb{T}$.*

Proof. If we show that $\iota(\chi)$ is a metric representation of $\chi = (T, c, \mu) \in \mathbb{T}_2$, it is obvious that ι is injective, because equivalence of metric measure spaces implies equivalence of the corresponding algebraic measure trees by Lemma 2.34.

Choosing an appropriate representative, we can assume that $\nu\{v\} > 0$ for all $v \in \mathrm{br}(T, c)$. The canonical projection $\pi_\nu\colon T \to T_\nu$ is a tree homomorphism by Lemma 2.28. To show equivalence of (T, c, μ) and $(T_\nu, c_{(T_\nu, r_\nu)}, \mu_\nu)$, we have to show that π_ν is injective on a subtree $A \subseteq T$ with $\mu(A) = 1$. Let $N := \{v \in T : \pi_\nu(v) \neq \{v\}\}$. Then $\mu(\pi_\nu(v)) = 0$ for all $v \in N$, and $w \in \pi_\nu(v)$ implies $[v, w] \subseteq \pi_\nu(v)$ because π_ν is a tree homomorphism. Because there are at most countably many non-degenerate, disjoint closed intervals in T due to order separability, this implies that $\pi_\nu(N)$ is countable, and thus $\mu(N) = 0$. Define $A = T \setminus N$. Then $\mu(A) = 1$, and π_ν is injective on $T \setminus N$. To see that A is a subtree, pick $x, y, z \in A$. If $v := c(x, y, z) \in \{x, y, z\}$, then $v \in A$. Otherwise, $v \in \mathrm{br}(T, c)$, and hence $\nu\{v\} > 0$. This implies $\pi_\nu(v) = \{v\}$, i.e. $v \in A$. □

Corollary 3.9 (metrizability)**.** $\mathbb{T}$ *equipped with bpdd-Gromov-weak topology is a separable, metrizable space.*

Proof. The Gromov-weak topology on $\mathbb{H}$ is separable, and metrizable, e.g. by the Gromov-Prohorov metric d_{GP} (see [GPW09]). Because ι is injective by Proposition 3.8, $d_{\mathrm{BGP}}(\chi, y) := d_{\mathrm{GP}}(\iota(\chi), \iota(y))$, $\chi, y \in \mathbb{T}$, is a metric on $\mathbb{T}$ inducing bpdd-Gromov-weak topology. □

Remark 3.10 (distance polynomials)**.** By definition, a sequence $(\chi_n)_{n\in\mathbb{N}}$ in $\mathbb{T}$ converges to $\chi\in\mathbb{T}$ bpdd-Gromov-weakly if and only if $\iota(\chi_n)\underset{n\to\infty}{\longrightarrow}\iota(\chi)$ Gromov-weakly. It has been shown that the Gromov-weak convergence is equivalent to the convergence of the distribution of the distance matrix ([GPW09, Theorem 5]). Therefore, the bpdd-Gromov-weak convergence is equivalent to

$$\Phi(\chi_n)\underset{n\to\infty}{\longrightarrow}\Phi(\chi) \tag{3.7}$$

for all so-called *polynomials* $\Phi\colon\mathbb{T}\to\mathbb{R}$, which are test functions of the form (1.13). Note that the set Π_ι of all polynomials is an algebra, and therefore also convergence determining for $\mathbb{T}$-valued random variables (see [Löh13, BK10]). $\Diamond$

As pointed out in Remark 3.3, the equivalence class of every algebraic measure tree contains an algebraic *continuum* measure tree. The following example shows that ι would not be injective if we had defined it on the set of algebraic continuum measure trees with the stricter notion of equivalence where the whole algebraic continuum trees have to be measure preserving isomorphic.

Example 3.11. For $x\geq 0$, let T_x be the $\mathbb{R}$-tree generated by the interval $I_x=[-x,1]$ together with additional leaves $\{v_n\}$, $n\in\mathbb{N}$, where $c(0,1,v_n)=\frac{1}{n}$ and $r(\frac{1}{n},v_n)=\frac{1}{n}$, i.e. at each point $\frac{1}{n}\in I_x$ there is a branch of length $\frac{1}{n}$ attached. Then T_x is a compact $\mathbb{R}$-tree for every $x\geq 0$, hence induces an algebraic continuum tree by Theorem 1. Let $\mu_x\{-x\}=\frac{1}{2}$, and $\mu_x\{v_n\}=2^{-n-1}$ for $n\in\mathbb{N}$. Then $\chi_x:=(T_x,\mu_x)\in\mathbb{T}_2$. Now $\iota(\chi_x)=\iota(\chi_y)$ for every $x,y\geq 0$, but T_x and T_0 are not homeomorphic, hence not isomorphic by Proposition 2.36.

Note that $A_x:=\{x\}\cup\{v_n:\,n\in\mathbb{N}\}\cup\{\frac{1}{n}:\,n\in\mathbb{N}\}$ is a subtree of T_x with $\mu_x(A_x)=1$, and A_x is isomorphic (although not homeomorphic) to A_0. $\Diamond$

In order to construct algebraic measure trees, it is of course not necessary to specify the mass of every Borel subset. To the contrary, we can use the following Carathéodory-type extension result. To this end, recall for $x,y\in T$ with $x\neq y$ from (2.21) the component $\mathcal{S}_x(y)=\mathcal{S}_x^{(T,c)}(y)$ of $T\setminus\{x\}$ which contains y. In this section, it is convenient to define

$$\mathcal{S}_x(x):=\{x\}. \tag{3.8}$$

Then T is the disjoint union of the $\deg(x)+1$ sets in

$$\mathcal{C}_x:=\big\{\mathcal{S}_x(y):\,y\in T\big\}. \tag{3.9}$$

Note that $\mathcal{C}_x=\{\mathcal{S}_x(y):\,y\in V\}$ for order dense $V\subseteq T$ with $x\in V$. In particular, $\mathcal{C}_x$ is countable if (T,c) is order separable. For $y\in T$, $V\subseteq T$, we call a function $f\colon V\to\mathbb{R}$ *order left-continuous* on V w.r.t. $\leq_y$ if the following holds. For all $x,x_n\in V$ with $x_1\leq_y x_2\leq_y\cdots$ and $x=\sup_{n\in\mathbb{N}}x_n$ w.r.t. $\leq_y$ (in short $x_n\uparrow x$), we have $\lim_{n\to\infty}f(x_n)=f(x)$. Recall the notion of algebraic continuum tree from Definition 2.9.

Proposition 3.12 (extension to a measure)**.** *Let (T,c) be an order separable algebraic continuum tree, and $V\subseteq T$ order dense. Then a set-function $\mu_0\colon\mathcal{C}_V:=\bigcup_{x\in V}\mathcal{C}_x\to[0,1]$ has a unique extension to a probability measure on $\mathcal{B}(T,c)$ if it satisfies*

1. For all $x\in V$, $\sum_{A\in\mathcal{C}_x}\mu_0(A)=1$

2. For all $x,y\in V$ with $x\neq y$,

$$\mu_0\big(\mathcal{S}_x(y)\big)+\mu_0\big(\mathcal{S}_y(x)\big)\geq 1 \tag{3.10}$$

3. For every $y\in V$, the function $\psi_y\colon x\mapsto\mu_0(\mathcal{S}_x(y))$ is order left-continuous on V w.r.t. $\leq_y$.

Proof. Note that $\psi_y(x)=\psi_z(x)$ for $z\in\mathcal{S}_x(y)$. We therefore may write $\psi_A(x):=\psi_y(x)$ for any $A\subseteq\mathcal{S}_x(y)$. Define the $\cap$-stable set system

$$\mathcal{A}:=\Big\{\bigcap_{k=1}^{n}A_k:n\in\mathbb{N},\,A_k\in\mathcal{C}_V\Big\}. \tag{3.11}$$

By Corollary 2.24, $\mathcal{A}$ generates the Borel σ-algebra $\mathcal{B}(T,c)$. Let $\emptyset \neq A \in \mathcal{A}$ and $y \in A$. Because (T,c) has no edges and is order complete, we have $A = \bigcap_{x\in\partial A} \mathcal{S}_x(y)$, where ∂ denotes the boundary w.r.t. the component topology τ, which is a finite set in the case of A. Using (3.10), we obtain for $v \in V$, $x_0,\dots,x_n \in V \setminus \{v\}$ such that $\mathcal{S}_v(x_0),\dots,\mathcal{S}_v(x_n)$ are distinct, that

$$\psi_{x_0}(v) \leq 1 - \sum_{k=1}^{n} \psi_{x_k}(v) \leq 1 - \sum_{k=1}^{n}\big(1-\psi_v(x_k)\big). \tag{3.12}$$

This implies for $\emptyset \neq A \in \mathcal{A}$, by induction over $\#\partial A$, that

$$\mu(A) := 1 - \sum_{x\in\partial A}\big(1-\psi_A(x)\big) \geq 0, \tag{3.13}$$

hence μ is a non-negative extension of μ_0 to $\mathcal{A}$. We claim that μ is super-additive, additive and inner regular for compact sets. From this it follows by standard arguments that it has a unique extension to a measure on the generated σ-algebra $\sigma(\mathcal{A}) = \mathcal{B}(T,c)$.

Additivity. Let $n \in \mathbb{N}\setminus\{1\}$, and $A_1,\dots,A_n \in \mathcal{A}\setminus\{\emptyset\}$ disjoint with $A := \biguplus_{k=1}^{n} A_k \in \mathcal{A}$. Define $D := \bigcup_{k=1}^{n} \partial A_k$. Then $\partial A \subseteq D$ and there is $x \in D\setminus\partial A \subseteq A$. Let $I_x := \{k \in \{1,\dots,n\} : x \in \partial A_k\}$ and choose $y_k \in A_k$. Then, because the A_k are disjoint, the $\mathcal{S}_x(y_k)$, $k \in I$, are distinct, and because the A_k cover A, we have $\{\mathcal{S}_x(y_k) : k \in I_x\} = \mathcal{C}_x$. In particular, $\sum_{k\in I_x} \psi_{y_k}(x) = 1$, and $B_x := \bigcup_{k\in I_x} A_k \in \mathcal{A}$ with $\partial B_x = \biguplus_{k\in I_x} \partial A_k \setminus \{x\}$. We obtain

$$\begin{aligned}\sum_{k\in I_x}\mu(A_k) &= \sum_{k\in I_x}\Big(1-\big(1-\psi_{y_k}(x)\big) - \sum_{z\in\partial A_k\setminus\{x\}}\big(1-\psi_{y_k}(z)\big)\Big)\\ &= \sum_{k\in I_x}\psi_{y_k}(x) - \sum_{z\in\partial B_x}\big(1-\psi_x(z)\big) = \mu(B_x).\end{aligned} \tag{3.14}$$

By induction over n, this implies additivity of μ.

Super-additivity. Let $A_1,\dots,A_n \in \mathcal{A}\setminus\{\emptyset\}$ be disjoint and $\biguplus_{k=1}^{n} A_k \subseteq A \in \mathcal{A}$. The case $n=1$ is trivial, and we proceed by induction over n. Choose $y \in A_1$ and let $D := \partial A_1 \setminus \partial A$. For $x \in D$, $C \in \mathcal{C}'_x := \mathcal{C}_x \setminus \mathcal{S}_x(y)$ and $k \in \{2,\dots,n\}$, either $A_k \subseteq C$, or $A_k \cap C = \emptyset$. Therefore, we have the decomposition $\{2,\dots,n\} = \biguplus_{x\in D}\biguplus_{C\in\mathcal{C}'_x} I_C$ with $I_C := \{k : A_k \subseteq C\}$. Because $C\cap A \in \mathcal{A}$, and $A_k \subseteq C \cap A$ for $k \in I_C$, we can use the induction hypothesis to obtain

$$\sum_{k\in I_C}\mu(A_k) \leq \mu(C\cap A) = \psi_C(x) - \sum_{z\in\partial A\cap C}\big(1-\psi_A(x)\big). \tag{3.15}$$

Therefore,

$$\begin{aligned}\mu(A_1) &= 1 - \sum_{x\in\partial A_1\cap\partial A}\big(1-\psi_y(x)\big) - \sum_{x\in D}\big(1-\psi_y(x)\big)\\ &= \mu(A) + \sum_{x\in\partial A\setminus\partial A_1}\big(1-\psi_y(x)\big) - \sum_{x\in D}\sum_{C\in\mathcal{C}'_x}\psi_C(x)\\ &\leq \mu(A) - \sum_{x\in D}\sum_{C\in\mathcal{C}'_x}\sum_{k\in I_C}\mu(A_k)\\ &= \mu(A) - \sum_{k=2}^{n}\mu(A_k).\end{aligned} \tag{3.16}$$

Compact regularity. According to Proposition 2.18, all closed subsets of T are compact. Let $y \in A \in \mathcal{A}$. Because (T,c) is an order separable algebraic continuum tree, and V is order dense, we find for $z \in \partial A$ a sequence $(x_n(z))_{n\in\mathbb{N}}$ in $A\cap V$ with $x_n(z) \uparrow z$ w.r.t. $\leq_y$ as $n\to\infty$. Define $A_n := \bigcap_{z\in\partial A}\mathcal{S}_{x_n(z)}(y) \in \mathcal{A}$ and $K_n := A_n \cup \partial A_n$. Then K_n is compact, $A_n \subseteq K_n \subseteq A$, and because ∂A is finite, we have $\partial A_n = \{x_n(z) : z \in \partial A\}$ for sufficiently large n. Thus, by order left-continuity of ψ_y,

$$\lim_{n\to\infty} \mu(A_n) = 1 - \lim_{n\to\infty} \sum_{z\in\partial A} \big(1 - \psi_y(x_n(z))\big) = 1 - \sum_{z\in\partial A} \big(1 - \psi_y(z)\big) = \mu(A), \tag{3.17}$$

and μ is inner compact regular as claimed. □

We conclude this section with an extension result, which will be very useful for reading off algebraic measure trees from (sub-)triangulations of the circle in Section 4 below. In Proposition 3.12, we assumed the whole tree to be known, and considered the question of constructing a probability measure on it. Now, we assume that not the whole tree is given a priori, but only the (countably many) branch points. The question is, whether there is an extension of the tree which is rich enough to carry a measure with the specified masses of components.

Proposition 3.13 (construction of algebraic measure trees)**.** *Let (V, c_V) be a countable algebraic tree, and for each $x \in V$, let $A \mapsto \psi_A(x)$ be a probability measure on $\mathcal{C}_x$. Define $\psi_y(x) := \psi_{\mathcal{S}_x(y)}(x)$. Assume that for $x, y \in V$ with $x \neq y$,*

$$\psi_x(y) + \psi_y(x) \geq 1. \tag{3.18}$$

Then there is a unique (up to equivalence) algebraic measure tree $\chi = (T, c, \mu)$ such that

(i) $V \subseteq T$, $\mathrm{br}(T, c) = \mathrm{br}(V, c_V)$,
(ii) $\mu\big(\mathcal{S}_x^{(T,c)}(y)\big) = \psi_y(x)$ *for all* $x, y \in V$,
(iii) $\mathrm{at}(\mu) \subseteq V$, *where* $\mathrm{at}(\mu)$ *denotes the set of atoms of* μ.

Note that in general we cannot obtain $\mathrm{lf}(T, c) \subseteq \mathrm{lf}(V, c_V)$. To the contrary, $\mathrm{lf}(T, c)$ can be uncountable (for every representative of χ).

Proof. Existence. First note that for $y \in V$, ψ_y is monotonic w.r.t. $\leq_y$. Indeed, $z \leq_y x$ implies $\psi_y(z) \leq 1 - \psi_x(z) \leq \psi_z(x) = \psi_y(x)$.

We need to enlarge the tree to make ψ_y order left-continuous. Because V is countable, we may consider one y and one point x at a time. If $x, y \in V$ are such that there exists $x_n \in V$ with $x_n \uparrow x$, then by monotonicity $\phi_y(x) := \lim_{n\to\infty} \psi_y(x_n) \leq \psi_y(x)$ exists and is independent of the choice of x_n. If $\phi_y(x) \neq \psi_y(x)$, we extend the tree by adding one extra point $z \notin V$, i.e. we consider $\tilde{V} := V \uplus \{z\}$ with the unique extension $\tilde{c}$ of c_V such that $(\tilde{V}, \tilde{c})$ is an algebraic tree with $x_n \leq_y z \leq_y x$ for all n. Furthermore, we extend ψ to $\tilde{\psi}$ on $\tilde{V}$ by defining $\tilde{\psi}_y(z) := \phi_y(x)$, $\tilde{\psi}_z(z) = 0$ and $\tilde{\psi}_x(z) = 1 - \phi_y(x)$. It is easy to check that $(\tilde{V}, \tilde{c})$ together with $\tilde{\psi}$ satisfies the prerequisites of the Proposition, $\mathrm{br}(\tilde{V}, \tilde{c}) = \mathrm{br}(V, c)$, and $\{x \in \tilde{V} : \tilde{\psi}_x(x) > 0\} = \{x \in V : \psi_x(x) > 0\} \subseteq V$.

Now assume w.l.o.g. that ψ_y is already order left-continuous for all $y \in V$. Because V is countable, it is in particular order separable and according to Theorem 1, there is an order separable algebraic continuum tree (T, c) such that (V, c_V) is a subtree. We can choose (T, c) such that $\mathrm{br}(T, c) = \mathrm{br}(V, c_V)$. Consider the closure $\overline{V}$ of V w.r.t. the component topology τ. For $x \in \overline{V} \setminus V$, we define

$$\psi_y(x) := \sup\{\psi_y(z) : z \in V \cap \mathcal{S}_x(y)\}. \tag{3.19}$$

Then (3.18) holds for $x, y \in \overline{V}$, $x \neq y$, and ψ_y is order left-continuous. For every $\{x, y\} \in \mathrm{edge}(\overline{V}, \bar{c})$, where $\bar{c}$ is the restriction of c to $\overline{V}^3$, we fix an order isomorphism $\varphi_{x,y} : [x, y] \to [0, 1]$, which exists by Cantor's order characterization of $\mathbb{R}$ because $[x, y]$ is a linearly ordered, separable algebraic continuum tree. For every $z \in T \setminus \overline{V}$, there exists $\{x, y\} \in \mathrm{edge}(\overline{V}, \bar{c})$, with $z \in [x, y]$. We define

$$\psi_y(z) := (1 - \varphi_{x,y}(z))\psi_y(x) + \varphi_{x,y}(z)\big(1 - \psi_x(y)\big), \tag{3.20}$$

$\psi_x(z) := 1 - \psi_y(z)$ and $\psi_z(z) := 0$. Now we can use Proposition 3.12 to see that

$$\mu_0\big(\mathcal{S}_x(y)\big) := \psi_y(x) \tag{3.21}$$

has a unique extension to a probability measure μ on $\mathcal{B}(T,c)$.

The last step in the construction is to remove point-masses outside V by expanding them to intervals. To this end, let $P := \mathrm{at}(\mu) \setminus V$, and $\overline{T} := (T \setminus P) \uplus (P \times [0,1])$. Because $P \subseteq T \setminus V$ contains no branch points, we can extend the restriction of c to $T \setminus P$ to a branch point map $\tilde{c}$ on $\overline{T}$ in a canonical way such that $[(x,0),(x,1)] = \{x\} \times [0,1]$ for $x \in P$. Define the Markov kernel κ from T to $\overline{T}$ by

$$\kappa(x) := \begin{cases} \delta_x, & x \in T \setminus P, \\ \delta_x \otimes \lambda_{[0,1]}, & x \in P, \end{cases} \tag{3.22}$$

where δ_x is the Dirac measure in x and $\lambda_{[0,1]}$ is Lebesgue measure. Let $\bar{\mu} := \kappa_*(\mu)$ be the push-forward of μ under κ. Then $(\overline{T}, \tilde{c}, \bar{\mu})$ is a separable algebraic measure tree, and by construction $\mathrm{br}(\overline{T}, \tilde{c}) = \mathrm{br}(V, c_V)$ as well as $\mathrm{at}(\bar{\mu}) = \mathrm{at}(\mu) \cap V \subseteq V$. Furthermore, for $x, y \in V$, we have $\bar{\mu}(\mathcal{S}_x^{(\overline{T},\tilde{c})}(y)) = \mu(\mathcal{S}_x^{(T,c)}(y)) = \psi_y(x)$ as claimed.

Uniqueness. Follows similarly, where we note that it does not matter how we distribute the mass on an edge of $(\overline{V}, \bar{c})$ in a non-atomic way, because all algebraic measure trees without branch points and non-atomic measure are equivalent by Example 3.4. □

4. Triangulations of the circle

In this section, we encode binary algebraic measure trees by triangulations of subsets of the circle. This is comparable with the encoding of compact (ordered, rooted) metric (probability) measure trees by excursions over the unit interval, where the height profile encodes the branch point map as well as the metric distances. Moreover, also the measure can be encoded by the excursion by identifying the lengths of sub-excursions with the mass of the corresponding subtrees. Similarly, it turns out that we can encode binary algebraic measure trees by what we call sub-triangulations of the circle. As in the case of coding metric measure trees with excursions, the resulting *coding map* associating to a sub-triangulation the algebraic measure tree is continuous.

In Subsection 4.1, we introduce the space of sub-triangulations of the circle. In Subsection 4.2, we construct the coding map.

4.1. **The space of sub-triangulations of the circle.** Let $\mathbb{D}$ be a (fixed) closed disc of circumference 1, and $\mathbb{S} := \partial\mathbb{D}$ the circle. As usual, for a subset $A \subseteq \mathbb{D}$, we denote by $\bar{A}$, $\mathring{A}$, ∂A and $\mathrm{conv}(A)$ the closure, the interior, the boundary and the convex hull of A, respectively. Furthermore, let

$$\Delta(A) := \big\{\text{connected components of } \mathrm{conv}(A) \setminus A\big\}, \tag{4.1}$$

and

$$\nabla(A) := \big\{\text{connected components of } \mathbb{D} \setminus \mathrm{conv}(A)\big\}. \tag{4.2}$$

Then we have the disjoint decomposition $\mathbb{D} = A \uplus \bigcup \Delta(A) \uplus \bigcup \nabla(A)$.

Definition 4.1 ((sub-)triangulations of the circle)**.** A *sub-triangulation of the circle* is a closed, non-empty subset C of $\mathbb{D}$ satisfying the following two conditions:

- **(Tri1)** $\Delta(C)$ consists of open interiors of triangles.
- **(Tri2)** C is the union of non-crossing (non-intersecting except at endpoints), possibly degenerate closed straight line segments with endpoints in $\mathbb{S}$.

We denote the set of sub-triangulations of the circle by $\mathcal{T}$, i.e.

$$\mathcal{T} := \big\{\text{sub-triangulations of the circle}\big\}, \tag{4.3}$$

and call $C \in \mathcal{T}$ *triangulation of the circle* if and only if $\mathbb{S} \subseteq C$.

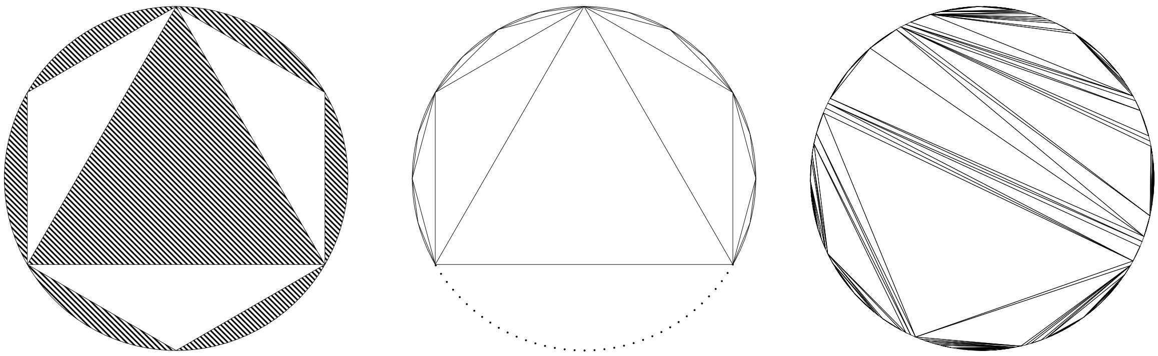


FIGURE 4. *left:* An Aldous-triangulation of the circle that is not a triangulation of the circle (Condition (Tri2) does not hold as the black triangle in the middle is not the union of non-crossing straight lines with endpoints on the circle). *middle:* A sub-triangulation of the circle (compare with Example 4.11). *right:* A triangulation of the circle. It is a realisation of the Brownian triangulation (compare with Example 4.5).

In particular, (Tri1) implies that $\partial \operatorname{conv}(C) \subseteq C$, and we may call C triangulation of $\partial \operatorname{conv}(C)$. Given (Tri1), (Tri2) implies that $\nabla(A)$ consists of circular segments with the bounding straight line excluded and the rest of the bounding arc included. We want to point out that our definition of triangulation of the circle differs from the one given by Aldous in [Ald94b, Definition 1]. Namely, Aldous required only Condition (Tri1). For the characterization of triangulations of the circle as limits of triangulations of n-gons given in Proposition 4.3 below, however, Condition (Tri2) is necessary. See Figure 4 for an example of a triangulation in the sense of Aldous that is excluded by Condition (Tri2), a sub-triangulation of the circle that is no triangulation of the circle, and a triangulation of the circle.

For a metric space (X, d), let

$$\mathcal{F}(X) := \{F \subseteq X : F \neq \emptyset,\, F \text{ closed}\}, \tag{4.4}$$

and equip $\mathcal{F}(X)$ with the *Hausdorff metric topology.* That is, we say that a sequence $(F_n)_{n\in\mathbb{N}}$ converges to F in $\mathcal{F}(X)$ if and only if for all $\varepsilon > 0$ and all large enough $n \in \mathbb{N}$,

$$F_n^{\epsilon} \supseteq F \qquad \text{and} \qquad F^{\epsilon} \supseteq F_n, \tag{4.5}$$

where for all $A \in \mathcal{F}(X)$, as usual, $A^{\epsilon} := \{x \in X : d(x, A) < \epsilon\}$. It is well-known that if (X, d) is compact, then $\mathcal{F}(X)$ is a compact metrizable space as well. As sub-triangulations of the circle are elements of $\mathcal{F}(\mathbb{D})$, we naturally equip $\mathcal{T}$ with the Hausdorff metric topology. A first observation is that $\mathcal{T}$ is actually a closed, and therefore compact subspace of $\mathcal{F}(\mathbb{D})$.

Lemma 4.2 (compactness of $\mathcal{T}$)**.** *Both the space of triangulations of the circle, and the space $\mathcal{T}$ of sub-triangulations of the circle, are compact metrizable spaces in the Hausdorff metric topology.*

Proof. Because $\mathbb{D}$ is compact, $\mathcal{F}(\mathbb{D})$ is compact as well, and it is sufficient to show that $\mathcal{T}$ and the set of triangulations of the circle are closed subsets of $\mathcal{F}(\mathbb{D})$.

Let $C_n \in \mathcal{T}$ with $C_n \underset{n\to\infty}{\longrightarrow} C \in \mathcal{F}(\mathbb{D})$ in the Hausdorff metric topology. (Tri1) is easily seen to be a closed property, thus C satisfies (Tri1). Let L_n be a set of non-crossing line segments with endpoints in $\mathbb{S}$ such that $C_n = \bigcup L_n$. The closure of L_n in $\mathcal{F}(\mathbb{D})$ has the same property (it possibly differs from L_n by a set of degenerated one-point segments contained in non-degenerate segments of L_n), so we may assume L_n is closed to begin with, so that $L_n \in \mathcal{F}(\mathcal{F}(\mathbb{D}))$. Because $\mathcal{F}(\mathcal{F}(\mathbb{D}))$ is compact, we may assume, taking a subsequence if necessary, that $L_n \to L$ for some $L \in \mathcal{F}(\mathcal{F}(\mathbb{D}))$. Obviously, L_n consists of non-crossing line segments with endpoints in $\mathbb{S}$. Because the union operator $\bigcup\colon \mathcal{F}(\mathcal{F}(\mathbb{D})) \to \mathcal{F}(\mathbb{D})$ is continuous, we have $\bigcup L = C$. In particular, (Tri2) holds for C, and $C \in \mathcal{T}$. Obviously, also the property that $\mathbb{S} \subseteq C$ is preserved by Hausdorff metric limits, thus the set of triangulations of the circle is closed as well. □

We now show two characterizations of sub-triangulations of the circle. Namely, condition (Tri2) can be replaced by existence of “triangles in the middle” which is the major technical ingredient for the construction of the branch point map in the next subsection. Furthermore, they are precisely the limits of finite sub-triangulations, where we consider a sub-triangulation C as *finite* if $C \cap \mathbb{S}$ is a finite set, or equivalently, C consists of finitely many line segments.

Proposition 4.3 (characterization of (sub-)triangulations)**.** *Let $\emptyset \neq C \subseteq \mathbb{D}$ be closed. Then the following are equivalent.*

1. *C is a sub-triangulation of the circle.*
2. *Condition (Tri1) holds, all extreme points of* $\operatorname{conv}(C)$ *are contained in $\mathbb{S}$, and*
   (Tri2)’ *For $x, y, z \in \Delta(C) \cup \nabla(C)$ pairwise distinct, there exists a unique $c_{xyz} \in \Delta(C)$ such that x, y, z are subsets of pairwise different connected components of $\mathbb{D} \setminus \partial c_{xyz}$.*
3. *There exists a sequence $(C_n)_{n\in\mathbb{N}}$ of finite sub-triangulations of the circle with $C_n \underset{n\to\infty}{\longrightarrow} C$ in the Hausdorff metric topology.*

Furthermore, C is a triangulation of the circle if and only if C_n in 3. can be chosen as a triangulation of a regular n-gon inscribed in $\mathbb{S}$.

Remark 4.4 (condition (Tri2)’)**.** That x, y, z are subsets of different connected components of $\mathbb{D} \setminus \partial c_{xyz}$ means that either $c_{xyz} \in \{x, y, z\}$ and the two elements of $\{x, y, z\} \setminus \{c_{xyz}\}$ are subsets of different connected components of $\mathbb{D} \setminus \overline{c_{xyz}}$, or $c_{xyz} \notin \{x, y, z\}$ and x, y, z are subsets of pairwise different connected components of $\mathbb{D} \setminus \overline{c_{xyz}}$. ♢

Proof of Proposition 4.3. “1 ⇒ 2”. Because C is the union of line segments with endpoints on $\mathbb{S}$, it is obvious that the extreme points of $\operatorname{conv}(C)$ are contained in $\mathbb{S}$. We have to show (Tri2)’, so let $x, y, z \in \Delta(C) \cup \nabla(C)$ be pairwise distinct and note that uniqueness is obvious. If one of the elements of $\{x, y, z\}$, say x, is such that the other two are subsets of two different connected components of $\mathbb{D} \setminus \bar{x}$, then necessarily $x \in \Delta(C)$, and $c_{xyz} := x$ has the desired properties. So assume this is not the case.

Fix a set L of non-crossing, closed lines with endpoints in $\mathbb{S}$ such that $C = \bigcup L$. Define

$$L_x := \{\ell \in L : \ell \text{ separates } x \text{ from } y \cup z \text{ in } \mathbb{D}\}, \tag{4.6}$$

note that $L_x \neq \emptyset$ because y and z are in the same connected component of $\mathbb{D} \setminus \bar{x}$ by assumption, and order L_x by distance from x. Similarly, define L_y as set of lines separating y from $x \cup z$ ordered by distance from y, and L_z as set of lines separating z from $x \cup y$, ordered by distance from z. Define $\ell_x := \sup L_x$, $\ell_y := \sup L_y$, and $\ell_z := \sup L_z$, which exist because C is closed. In particular, they are non-crossing, and because $\operatorname{conv}(C) \setminus C$ may only consist of triangles, they have to be the sides of some $c_{xyz} \in \Delta(C)$ which has the desired properties.

“2 ⇒ 3”. Because the extreme points of $\operatorname{conv}(C)$ are on the circle, for every $x \in \nabla(C)$, the boundary $\partial_{\mathbb{D}} x$ in $\mathbb{D}$ is a single straight line with endpoints in $\mathbb{S}$. Let $(V_n)_{n\in\mathbb{N}}$ be an increasing sequence of finite subsets of $\Delta(C) \cup \nabla(C)$ such that $c_{xyz} \in V_n$ for pairwise distinct $x, y, z \in V_n$, and $V_n \uparrow \Delta(C) \cup \nabla(C)$. Let $A_n := \mathbb{D} \setminus \bigcup V_n$. Then $A_n \to C$ in the Hausdorff metric topology. Because $c_{xyz} \in V_n$ for distinct $x, y, z \in V_n$, the boundary of each of the finitely many connected components of $A_n \setminus \mathbb{S}$ consists of one or two line segments and one or two connected sub-arcs of $\mathbb{S}$. Therefore, there is a finite sub-triangulation $C_n \subseteq A_n$ of the circle with Hausdorff distance from A_n less than e^{-n}. Thus $C_n \to C$.

“3 ⇒ 1”. Obvious, because $\mathcal{T}$ is a closed subset of $\mathcal{F}(\mathbb{D})$ by Lemma 4.2.

“Furthermore”. If C_n is a triangulation of the n-gon, it contains the n-gon, and hence any Hausdorff metric limit as $n \to \infty$ contains the circle, and hence is a triangulation of the circle.

That triangulations of the circle can be approximated by triangulations of regular n-gons is a slight modification of the arguments above. Details are left to the reader. $\square$

The most prominent random tree is Aldous's Brownian CRT, which is the limit of uniform random trees. Similarly, one can define the Brownian triangulation of the circle.

Example 4.5 (Brownian triangulation)**.** The uniform random triangulation of the n-gon converges in law with respect to the Hausdorff metric topology to the so-called *Brownian triangulation* C_{CRT}, see [Ald94a, Ald94b, CK14]. A realisation is shown in the right of Figure 4. It has a.s. Hausdorff dimension $\frac{3}{2}$ (see [Ald94a]). $\Diamond$

4.2. **Coding binary measure trees with (sub-)triangulations of the circle.** Given an algebraic tree (T, c), recall the set of leaves $\mathrm{lf}(T, c)$, and the degree $\deg_{(T,c)}(v)$ of $v \in T$ from (2.36) and (2.35), respectively. In this section, we are interested in the following subspace of the space of all binary algebraic measure trees.

Definition 4.6 (our space $\mathbb{T}_2$)**.** Let $\mathbb{T}_2 \subseteq \mathbb{T}$ be the set of (equivalence classes of) algebraic measure trees (T, c, μ) with (T, c) binary (i.e. $\deg_{(T,c)}(v) \le 3$ for all $v \in T$) and $\mathrm{at}(\mu) \subseteq \mathrm{lf}(T, c)$.

The space $\mathbb{T}_2$ is of particular interest to us, as it is invariant under the dynamics of the Aldous diffusion on cladograms, the construction of which was one of the motivations for studying algebraic measure trees, and because, as we will see, it is precisely the space of algebraic measure trees that can be coded by sub-triangulations of the circle.

To illustrate the construction of the tree coded by a sub-triangulation, we first consider a triangulation C of the regular n-gon into necessarily $n-2$ triangles (see Figure 3). Here, the coded tree is the dual graph. That is, every triangle corresponds to a branch point of the tree, and two branch points are connected by an edge if and only if the triangles share a common edge. We then add a leaf for every edge of the n-gon and obtain a graph-theoretic binary tree with n leaves and $n-2$ internal vertices. Recall from Example 2.4 that the finite graph-theoretic tree corresponds to a unique algebraic tree. We finally assign to each leaf mass n^{-1} (which corresponds to the length of the arcs of the circle connecting two endpoints of edges of the n-gon if we inscribe it in a circle of unit length), and obtain an algebraic measure tree.

The main result of this section is that there is a natural, surjective coding map from $\mathcal{T}$ onto $\mathbb{T}_2$, which is also continuous. To state that formally, we need further notation. Given a sub-triangulation $C \subseteq \mathbb{D}$, recall $\Delta(C)$ and $\nabla(C)$ from (4.2) and (4.1), respectively. For $x \in \Delta(C) \cup \nabla(C)$, and $y \subseteq \mathbb{D}$ connected and disjoint from $\partial_{\mathbb{D}} x$, where $\partial_{\mathbb{D}}$ denotes the boundary in the space $\mathbb{D}$, let

$$\mathrm{comp}_x(y) := \text{the connected component of } \mathbb{D} \setminus \partial_{\mathbb{D}} x \text{ which contains } y. \tag{4.7}$$

For $x \in \Delta(C)$, let $p_i(x)$, $i = 1, 2, 3$, be the mid-points of the three arcs of $\mathbb{S} \setminus \partial x$, and define

$$\square(C) := \big\{\{p_i(x)\} : x \in \Delta(C),\, i \in \{1,2,3\},\, \mathrm{comp}_x(\{p_i(x)\}) \subseteq C\big\}, \tag{4.8}$$

as well as $\mathrm{comp}_p(p) := \{p\}$ for $p \in \square(C)$ (see Figure 5). Recall the definition of components $\mathcal{S}_v(w)$ in an algebraic tree from (2.21).

Lemma 4.7 (induced branch point map)**.** *For $C \in \mathcal{T}$, let $V_C := \Delta(C) \cup \nabla(C) \cup \square(C)$. If $V_C \neq \emptyset$, then there is a unique branch point map $c_V : V_C^3 \to V_C$ such that (V_C, c_V) is an algebraic tree with $\mathcal{S}_x^{(V_c,c_V)}(y) = \{v \in V_C : \mathrm{comp}_x(y) = \mathrm{comp}_x(v)\}$ for $x, y \in V_C$. Furthermore,* $\deg(x) = 3$ *for all $x \in \Delta(C)$, and* $\deg(x) = 1$ *for $x \in \nabla(C) \cup \square(C)$.*

Proof. Recall from Proposition 4.3 that for a sub-triangulation C of the circle and pairwise distinct $x, y, z \in \Delta(C) \cup \nabla(C)$, there is a triangle $c_{xyz} \in \Delta(C)$ "in the middle". It is straight-forward to see that this defines a branch point map and can naturally be extended to V_C^3. $\square$

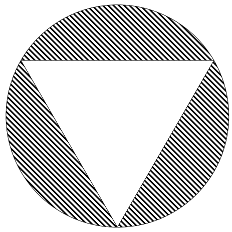


FIGURE 5. Triangulation C with $\#\Delta(C) = 1$, $\nabla(C) = \emptyset$, and $\#\square(C) = 3$. The coded tree consists of three line segments with non-atomic measure of $\frac{1}{3}$ each, glued together at one branch point.

The following theorem states that all sub-triangulations C of the circle can be associated with an element in $\mathbb{T}_2$ for which $\Delta(C)$ corresponds to the set of branch points, $\nabla(C)$ corresponds to the set

$$\mathrm{lf}_{\mathrm{atom}}(\chi) := \big\{x \in \mathrm{lf}(T, c) : \mu(\{x\}) > 0\big\} \tag{4.9}$$

of leaves which carry an atom, and $\mathrm{comp}_v(w)$ corresponds to the component $\mathcal{S}_v(w)$.

Theorem 2 (algebraic measure tree associated to a sub-triangulation)**.**

(i) For every sub-triangulation $C \subseteq \mathbb{D}$ of the circle, there is a unique (up to equivalence) algebraic measure tree $\chi_C = (T_C, c_C, \mu_C) \in \mathbb{T}_2$ with the following properties:

(CM1) $V_C \subseteq T_C$, $\mathrm{br}(\chi_C) = \Delta(C)$, *and c_C is an extension of c_V, where (V_C, c_V) is defined in Lemma 4.7.*

(CM2) $\mu_C\big(\mathcal{S}_x^{(T_C,c_C)}(y)\big) = \lambda_{\mathbb{S}}\big(\mathbb{S} \cap \mathrm{comp}_x(y)\big)$ *for all $x, y \in V_C$, where $\lambda_{\mathbb{S}}$ denotes the Lebesgue measure on $\mathbb{S}$.*

(CM3) $\mathrm{at}(\mu_C) = \nabla(C)$.

(ii) The coding map $\tau \colon \mathcal{T} \to \mathbb{T}_2$, $C \mapsto \chi_C$ *is surjective and continuous, where $\mathcal{T}$ is equipped with the Hausdorff metric topology and $\mathbb{T}_2$ with the bpdd-Gromov-weak topology.*

Proof. (i) Let C be a sub-triangulation of the circle. If $C = \mathbb{D}$, then $\Delta(C) = \nabla(C) = \emptyset$, which requires by (CM1) that $\mathrm{br}(\chi_C) = \emptyset$, and by (CM3) that $\mathrm{at}(\mu) = \emptyset$. There is a unique algebraic measure tree without branch points and atoms, namely the line segment with no atoms (see Example 4.10). We may therefore assume w.l.o.g. that $C \neq \mathbb{D}$, and consequently that $T_C \neq \emptyset$.

We claim that (V_C, c_V) together with $\psi_y(x) := \lambda_{\mathbb{S}}\big(\mathbb{S} \cap \mathrm{comp}_x(y)\big)$ satisfies the assumptions of Proposition 3.13. Indeed, V_C is obviously countable and an algebraic tree by Lemma 4.7, $\psi_y(x)$ depends on y only through its equivalence class w.r.t. $\sim_x$, and the lengths of all the arcs add up to the total length of $\lambda_{\mathbb{S}}(\mathbb{S}) = 1$. Furthermore, $\psi_x(y) + \psi_y(x) \geq \lambda_{\mathbb{S}}(\mathbb{S}) = 1$, and Proposition 3.13 yields existence and uniqueness of the desired algebraic measure tree.

(ii) Let $\chi = (T, c, \mu) \in \mathbb{T}_2$. We construct a sub-triangulation C such that $\tau(C) = \chi$. Fix $\rho \in \mathrm{lf}(T, c)$, and recall that ρ induces a partial order relation $\leq_\rho$. We can extend this partial order to a total (planar) order $\leq$ by picking for every $v \in \mathrm{br}(T, c)$ an order of the two components of $T \setminus \{v\}$ that do not contain ρ. That is, we define $S_0(v) := \mathcal{S}_v(\rho)$, denote the two remaining components of $T \setminus \{v\}$ by $S_1(v)$, $S_2(v)$, and define

$$v \leq w \quad :\Leftrightarrow \quad v \leq_\rho w \text{ or } v \in S_1\big(c(x, y, \rho)\big),\ w \in S_2\big(c(x, y, \rho)\big). \tag{4.10}$$

Identify $\mathbb{S}$ with $[0, 1]$, where the endpoints are glued. For $a \in [0, 1]$ and $b, c > 0$ with $a + b + c \leq 1$, let $\Delta(a, b, c) \subseteq \mathbb{D}$ be the open triangle with vertices $a, a + b, a + b + c \in \mathbb{S}$, $\ell(a, b) \subseteq \mathbb{D}$ the straight line from a to $a + b$, and $L(a, b)$ the connected component of $\mathbb{D} \setminus \ell(a, b)$ containing $a + \frac{b}{2} \in \mathbb{S}$. The first vertex of the triangle or circular segment corresponding to $v \in \mathrm{br}(T, c) \cup \mathrm{lf}_{\mathrm{atom}}(\chi)$ is given by the total mass before (w.r.t. $\leq$ defined in (4.10)), i.e. by

$$\alpha(v) := \mu\big(\{u \in T : u < v\}\big). \tag{4.11}$$

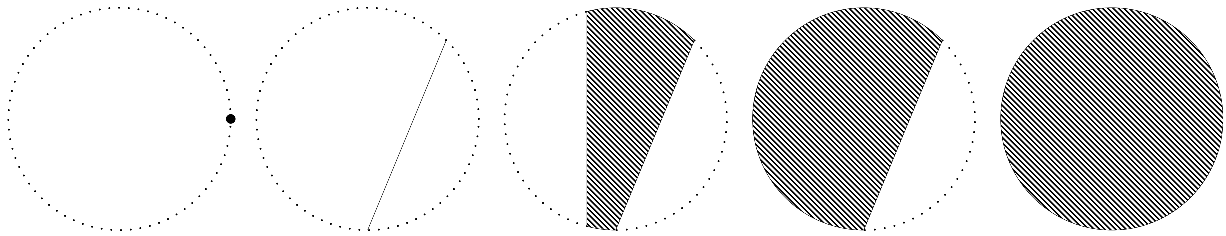


Figure 6. Sub-triangulations of the circle which correspond to the five cases of algebraic measure trees without branch points as explained in Example 4.10.

Define

$$\mathbb{D} \setminus C := \biguplus_{v \in \mathrm{br}(T,c)} \Delta\big(\alpha(v), \mu(S_1(v)), \mu(S_2(v))\big) \uplus \biguplus_{v \in \mathrm{lf}_{\mathrm{atom}}(x)} L\big(\alpha(v), \mu\{v\}\big) \tag{4.12}$$

By definition of C, $\mathrm{conv}(C) \setminus C$ consists of open triangles, i.e. condition (Tri1) is satisfied. Furthermore, the extreme points of $\mathrm{conv}(C)$ are contained in $\mathbb{S}$, and for $x, y, z \in \Delta(C) \cup \nabla(C)$ distinct, there are corresponding $u, v, w \in T$, and a triangle $c_{xyz} \in \Delta(C)$ corresponding to $c(u, v, w)$, which satisfies the requirements of (Tri2)'. Thus, by Proposition 4.3, C is a sub-triangulation of the circle. It is straight-forward to check that $\tau(C) = x$.

We defer the proof of continuity of τ to the next section, where we prove it in Lemma 5.20. □

The following is obvious now.

Lemma 4.8 (non-atomicity). *A sub-triangulation C of the circle is a triangulation of the circle if and only if, for $(T_C, c_C, \mu_C) := \tau(C)$, the measure μ_C is non-atomic.*

Corollary 4.9 (finite tree approximation). *Let $x = (T, c, \mu) \in \mathbb{T}_2$. Then there is a sequence $(x_n)_{n \in \mathbb{N}}$ of finite algebraic measure trees in $\mathbb{T}_2$ with $x_n \to x$ bpdd-Gromov-weakly. Furthermore, if μ is non-atomic, then x_n can be chosen as a tree with n leaves and uniform distribution on the leaves.*

Proof. By Theorem 2, there is a sub-triangulation $C \in \mathcal{T}$ with $\tau(C) = x$, and by Proposition 4.3, there are finite sub-triangulations C_n with $C_n \to C$. Obviously, $x_n := \tau(C_n)$ is a finite algebraic measure tree and by continuity of τ we have $x_n \to x$. If μ is non-atomic, then, by Lemma 4.8, C is a triangulation of the circle, and hence, by Proposition 4.3, C_n can be chosen as triangulation of the n-gon, which means that x_n has n leaves and uniform distribution on them. □

We conclude this section with a few illustrative examples.

Example 4.10 (coding algebraic measure trees without branch points). Let x be an algebraic measure tree without branch points. If $x = x_C$ for some sub-triangulation C, then $\Delta(C) = \mathrm{br}(x_C) = \emptyset$ and the following five cases can occur (see Figure 6): a) x_C consists of one single point of mass 1. Then $C = \{x\}$ for some $x \in \mathbb{S}$. b) x_C consists of an interval with two leaves, where each carries positive mass adding up to 1, in which case C is a single line segment dividing the circle into two arcs with length corresponding to the masses of the two leaves. c) x_C consists of an interval with two leaves, where each has positive mass adding up to $a < 1$. In this case, C is the area of the disc bounded by two distinct line segments and two arcs (possibly one of them degenerated) of $\mathbb{S}$, and the lengths of the remaining two arcs are given by the masses of the leaves. d) x_C consists of an interval with two leaves, where one has positive mass $a < 1$ and the other one has zero mass. Then C is a circular segment with arc length $1 - a$. e) x_C consists of an interval with no atoms on the leaves, which implies $C = \mathbb{D}$. ◊

Example 4.11 (a complete binary tree). Let C be the sub-triangulation of the circle drawn in the middle of Figure 4. Then $\#\nabla(C) = \#\,\mathrm{lf}_{\mathrm{atom}}(\tau(C)) = 1$. We refer to this only leaf with

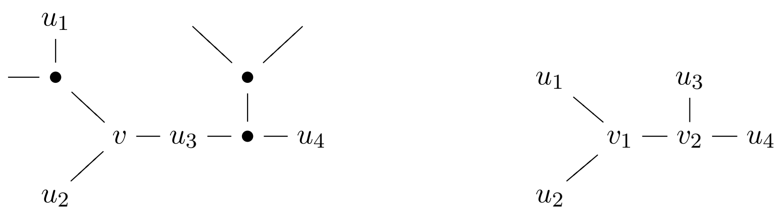


FIGURE 7. A tree T and the shape $\mathfrak{s}_T(u_1, u_2, u_3, u_4)$. Here, we are considering the homomorphism $f: C \to c^3(\{u_1, ..., u_4\}^3)$ given by $f(u_i) := u_i$, $i = 1, ..., 4$, and then necessarily $f(v_1) = v$, $f(v_2) = u_3$. f is clearly no isomorphism, and the cladogram is not isomorphic to the subtree $c(\{u_1, u_2, u_3, u_4\}^3)$ because $c(u_1, u_4, u_3) = u_3$.

positive mass as the root ρ, and obtain $\mu(\{\rho\}) = \frac{1}{3}$, corresponding to the length of the dotted arc. Moreover, $\tau(C)$ consists of a complete rooted binary tree in the sense of graph theory (with the convention that the root has degree one), together with an uncountable set of leaves given by the ends at infinity and carrying the remaining $\frac{2}{3}$ of the mass. $\diamondsuit$

Example 4.12 (coding the Brownian CRT)**.** Recall the Brownian triangulation C_{CRT} from Example 4.5, which is defined as the limit in distribution of uniform random triangulations C_n of the n-gon. A realization is shown in the right of Figure 4. It is easy to see that $\tau(C_n)$ is the uniform binary tree with n leaves and uniform distribution on the leaves. Thus, by Theorem 2, the uniform binary tree converges bpdd-Gromov-weakly to $\tau(C_{\mathrm{CRT}})$. At this point it is not entirely clear that $\tau(C_{\mathrm{CRT}})$ is the algebraic measure tree induced by the metric measure Brownian CRT. We will see in Section 6 that this is indeed the case. $\diamondsuit$

5. The subspace of binary algebraic measure trees

In this section we introduce in Subsections 5.1 and 5.2 with the *sample shape convergence* and the *sample subtree-mass convergence* two more notions of convergence of algebraic measure trees which seem more natural when thinking of algebraic trees as combinatorial objects. We then show in Subsection 5.3 that on $\mathbb{T}_2$, both of these notions are equivalent to the bpdd-Gromov-weak convergence. The main tools are a uniform Glivenko Cantelli argument, and that the coding map sending a sub-triangulation of the circle to an element in $\mathbb{T}_2$ is continuous.

5.1. Convergence in distribution of sampled tree shapes.

The basic idea behind Gromov-weak convergence for metric measure spaces is to sample finite metric sub-spaces with the sampling measure μ and then require these to converge in distribution. In this section, we propose a corresponding construction for binary algebraic measure trees, where we sample finite tree shapes with μ.

First, we have to make precise what we mean by "tree shape", which we understand to be a cladogram with the peculiarity that leaves may carry more than one label. The multi-label case is necessary to allow for sampling the same point several times due to a possible atom at that point.

Definition 5.1 (m-labelled cladogram)**.** For $m \in \mathbb{N}$, an *m-labelled cladogram* is a binary, finite algebraic tree $C = (C, c)$ together with a surjective labelling map $\ell: \{1, ..., m\} \to \mathrm{lf}(C)$. Two m-labelled cladograms (C_1, ℓ_1) and (C_2, ℓ_2) are equivalent if they are label preserving isomorphic i.e., there exists a tree isomorphism $f: C_1 \to C_2$ with $f(\ell_1(i)) = \ell_2(i)$ for all $i = 1, ..., m$.

Define

$$\mathfrak{C}_m := \{\text{isomorphism classes of } m\text{-labelled cladograms}\}. \tag{5.1}$$

In the following we will use cladograms to encode the shape of a subtree spanned by a finite sample of leaves.

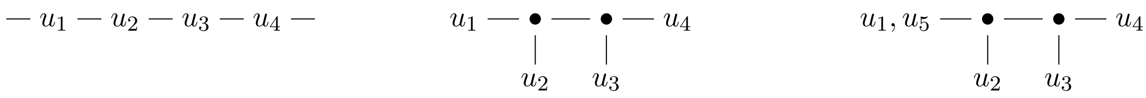


FIGURE 8. The left shows a totally ordered binary algebraic tree and four different points $u_1, ..., u_4$. The middle shows the shape $\mathfrak{s}_T(u_1, ..., u_4)$ of the cladogram which forms a comb tree. The right illustrates what happens if a fifth point is equal to u_1. Now one of the leaves of $\mathfrak{s}_T(u_1, ..., u_5)$ has two labels.

Definition 5.2 (tree shape)**.** For a binary algebraic tree (T, c), $m \in \mathbb{N}$, and $u_1, ..., u_m \in T \setminus \mathrm{br}(T, c)$, the *tree shape* $\mathfrak{s}_T(u_1, ..., u_m)$ of the m-labelled cladogram spanned by $(u_1, ..., u_m)$ in (T, c) is the unique (up to isomorphism) m-labelled cladogram $\mathfrak{s}_T(u_1, ..., u_m) = (C, c_C, \ell)$ with $\mathrm{lf}(C) = \{u_1, ..., u_m\}$ and $\ell(i) = u_i$ for all $i = 1, ..., m$, and such that the identity on $\mathrm{lf}(C)$ extends to a tree homomorphism from C onto $c\big(\{u_1, ..., u_m\}^3\big)$.

Remark 5.3 (spanned subtree and cladogram are not necessarily isomorphic)**.** The tree homomorphism from $\mathfrak{s}_T(u_1, ..., u_m)$ onto $c(\{u_1, ..., u_m\}^3)$ does not need to be injective. This is the case if (and only if) $u_i \in (u_j, u_k)$ for some $i, j, k \in \{1, ..., m\}$. See Figure 7. ◊

Example 5.4 (shape of a totally ordered algebraic tree)**.** Let (T, c) be a totally ordered algebraic tree, and $u_1, ..., u_m \in T$. Then $\mathfrak{s}_T(u_1, ..., u_m)$ is a so-called *comb tree* which has a totally ordered spine of binary branch points with attached leaves (see Figure 8). ◊

In the following, we build a topology on the convergence of tree shapes of m randomly sampled points. We therefore need the measurability of the shape map.

Lemma 5.5 (measurability of the shape map)**.** *For every binary algebraic tree (T, c) and $m \in \mathbb{N}$, the tree shape map $\mathfrak{s}_T : (T \setminus \mathrm{br}(T, c))^m \to \mathfrak{C}_m$ is a measurable function.*

Proof. Restricted to the open subset $\big\{v \in (T \setminus \mathrm{br}(T, c))^m : v_1, ..., v_m \text{ distinct}\big\}$, $\mathfrak{s}_T$ is locally constant, hence continuous. The same is true restricted to the set $\big\{v \in (T \setminus \mathrm{br}(T, c))^m : v_1 = v_2,\ v_2, ..., v_m \text{ distinct}\big\}$, which is an intersection of a closed and an open set, hence measurable. We can continue this way to see that $\mathfrak{s}_T$ is measurable on $(T \setminus \mathrm{br}(T, c))^m$. □

Definition 5.6 (tree shape distribution)**.** For $\chi = (T, c, \mu) \in \mathbb{T}_2$ and $m \in \mathbb{N}$, the *m-tree shape distribution* of χ is defined by

$$\mathfrak{S}_m(\chi) := \mu^{\otimes m} \circ \mathfrak{s}_T^{-1} \in \mathcal{M}_1(\mathfrak{C}_m). \tag{5.2}$$

Example 5.7 (shape of the linear non-atomic measure tree)**.** Let $\chi = (T, c, \mu)$ be the linear non-atomic algebraic measure tree (Example 3.4). Then any sample $(u_1, ..., u_m)$ with μ consists of pairwise different points, and $\mathfrak{S}_m(\chi)$ is the mixture of Dirac measures on labelled comb trees where the mixture is over all (up to isometry) permutations of the labels. ◊

We refer to the weakest topology on $\mathbb{T}_2$ such that for every $m \in \mathbb{N}$ the m-tree shape distribution is continuous as sample shape topology.

Definition 5.8 (sample shape topology)**.** The topology induced on $\mathbb{T}_2$ by the set $\{\mathfrak{S}_m : m \in \mathbb{N}\}$ of tree shape distributions is called *sample shape topology.*

We say that a sequence $(\chi_n)_{n \in \mathbb{N}}$ is *sample shape convergent* to χ in $\mathbb{T}_2$ if it converges w.r.t. the sample shape topology, i.e. if $\mathfrak{S}_m(\chi_n)$ converges to $\mathfrak{S}_m(\chi)$ as $n \to \infty$ for every $m \in \mathbb{N}$.

In analogy to the set Π_ι of polynomials introduced in Remark 3.10, we also introduce a set of test functions which evaluate the tree shape distributions. We refer to $\Phi = \Phi^{m,\varphi} : \mathbb{T}_2 \to \mathbb{R}$,

$$\Phi(\chi) = \int_{\mathfrak{C}_m} \varphi \, \mathrm{d}\mathfrak{S}_m(\chi) = \int_{T^m} \varphi \circ \mathfrak{s}_T \, \mathrm{d}\mu^{\otimes m}, \tag{5.3}$$

where $m \in \mathbb{N}$ and $\varphi\colon \mathfrak{C}_m \to \mathbb{R}$, as *shape polynomial.* We also define

$$\Pi_{\mathfrak{s}} := \{\,\text{shape polynomials on } \mathbb{T}_2\,\}. \tag{5.4}$$

Obviously, the sample shape topology is induced by the set $\Pi_{\mathfrak{s}}$ of shape polynomials.

Proposition 5.9 (sample shape implies bpdd-Gromov-weak convergence)**.** *On $\mathbb{T}_2$, the sample shape topology is stronger than the bpdd-Gromov-weak topology (i.e. any open set in the bpdd-Gromov-weak topology is open in the sample shape topology).*

Proof. The bpdd-Gromov-weak topology is induced by the set Π_ι of polynomials (see Remark 3.10). Because the set of $\phi \in \mathcal{C}_b(\mathbb{R}^{m\times m})$ which are Lipschitz continuous is convergence determining for probability measures on $\mathbb{R}^{m\times m}$, the subset of those $\Psi \in \Pi_\iota$ with

$$\Psi(T,c,\mu) = \int_{T^m} \phi\big((\nu[u_i,u_j] - \tfrac12\nu\{u_i\} - \tfrac12\nu\{u_j\})_{i,j=1,\dots,m}\big)\,\mu^{\otimes m}(\mathrm{d}\underline{u}) \tag{5.5}$$

for some $m \in \mathbb{N}$ and Lipschitz continuous $\phi \in \mathcal{C}_b(\mathbb{R}^{m\times m})$ also induces the bpdd-Gromov-weak topology. Therefore, it is enough to show that such a Ψ is continuous on $\mathbb{T}_2$ w.r.t. the sample shape topology. We do so by showing that the restriction to $\mathbb{T}_2$ of Ψ is in the uniform closure of $\Pi_{\mathfrak{s}}$. Let L be the Lipschitz constant of ϕ w.r.t. the ℓ_∞-norm on $\mathbb{R}^{m\times m}$.

For $n \in \mathbb{N}$ with $3n \geq m$, we define

$$\Phi_n(T,c,\mu) := \int_{T^{3n}} \phi\big((\nu_{n,\underline{u}}[u_i,u_j] - \tfrac12\nu_{n,\underline{u}}\{u_i\} - \tfrac12\nu_{n,\underline{u}}\{u_j\})_{i,j=1,\dots,m}\big)\,\mu^{\otimes 3n}(\mathrm{d}\underline{u}), \tag{5.6}$$

with the empirical branch point distribution

$$\nu_{n,\underline{u}} := \tfrac1n \sum_{k=0}^{n-1} \delta_{c(u_{3k+1},u_{3k+2},u_{3k+3})}. \tag{5.7}$$

Note that the restriction of Φ_n to $\mathbb{T}_2$ belongs to $\Pi_{\mathfrak{s}}$ because whether or not $c(u_{k+1},u_{k+2},u_{k+3})$ lies on $[u_i,u_j]$, $k \in \{0,\dots,n-1\}$, $i,j \in \{1,\dots,m\}$ only depends on the shape $\mathfrak{s}_{3n}(\underline{u})$.

Finally, we observe

$$\|\Psi - \Phi_n\|_\infty \leq \sup_{(T,c,\mu)\in\mathbb{T}_2} \int_{T^{3n}} L\cdot 3 \sup_{I\in\mathcal{I}_T} |\nu(I) - \nu_{n,\underline{u}}(I)|\,\mu^{\otimes 3n}(\mathrm{d}\underline{u}) \leq 3L\cdot\epsilon_n \underset{n\to\infty}{\longrightarrow} 0, \tag{5.8}$$

with $\mathcal{I}_T := \{[x,y];\, x,y \in T\}$ and $(\epsilon_n)_{n\in\mathbb{N}} \underset{n\to\infty}{\longrightarrow} 0$, where we have used a uniform Glivenko-Cantelli estimate which upper bounds the distance of the empirical branch point distribution to the branch point distribution. Such an estimate should be known, but as we could not come up with a reference, we show it in Lemma A.4 in the appendix. We note that $\dim_{\mathrm{VC}}(\mathcal{I}_T) = 2$ (compare Example A.2). □

Corollary 5.10 (metrizability)**.** *The sample shape topology is metrizable.*

Proof. Because the sample shape topology is induced by a countable family of functions $(\mathfrak{S}_m)_{m\in\mathbb{N}}$ with values in metrizable spaces, it is pseudo-metrizable. By Proposition 5.9, it is stronger than the bpdd-Gromov-weak topology, hence a Hausdorff topology. Therefore, it is metrizable. □

5.2. **Convergence in distribution of sampled subtree masses.** In this subsection, we introduce yet another notion of convergence of algebraic measure trees which, in contrast to sampling tree shapes, is based on sampling branch points and evaluating the masses of the subtrees that are joined at these branch points. This approach might be more similar to the case of metric measure spaces and distance matrix distributions, because we sample a tensor of real numbers (masses of subtrees) as opposed to a combinatorial object (tree shape). Thus, the typical tools of analysis are more readily applicable for the corresponding class of test functions.

Let $(T, c, \mu) \in \mathbb{T}_2$, and recall from (2.21) for $u, v, w \in T$ the subtree component $\mathcal{S}_{c(u,v,w)}(x)$ of $T \setminus \{c(u, v, w)\}$ which contains $x \neq c(u, v, w)$. Here, we always take the component containing $x = u$, and consider its mass

$$\eta(u, v, w) := \mathbb{1}_{u \neq c(u,v,w)} \cdot \mu\big(\mathcal{S}_{c(u,v,w)}(u)\big). \tag{5.9}$$

Lemma 5.11 (measurability of the subtree masses)**.** *For every binary algebraic measure tree $\chi = (T, c, \mu) \in \mathbb{T}_2$ and $m \in \mathbb{N}$, the function $\eta\colon T^3 \to [0, 1]$ is measurable.*

Proof. First, we claim that the map $\psi\colon T^2 \to [0, 1]$,

$$\psi(u, v) := \mathbb{1}_{u \neq v} \cdot \mu\big(\mathcal{S}_v(u)\big) \tag{5.10}$$

is lower semi-continuous. Indeed, let (u_n, v_n) be a sequence converging to (u, v). We may assume w.l.o.g. that $v \neq u$, $u_n \in \mathcal{S}_v(u)$, and either $v_n \notin \mathcal{S}_v(u)$ for all $n \in \mathbb{N}$, or $v_n \in \mathcal{S}_v(u)$ for all $n \in \mathbb{N}$. In the first case, $\mathcal{S}_v(u) \subseteq \mathcal{S}_{v_n}(u_n)$, and hence $\psi(u, v) \leq \psi(u_n, v_n)$. In the second case, for every $x \in \mathcal{S}_v(u)$ and $n \geq n_x$ sufficiently large, we have $u \in \mathcal{S}_{v_n}(u_n)$ and $v_n \notin [x, u]$. This means $x \in \mathcal{S}_{v_n}(u) = \mathcal{S}_{v_n}(u_n)$ and hence

$$\psi(u, v) - \liminf_{n\to\infty} \psi(u_n, v_n) \leq \lim_{n\to\infty} \mu\big(\mathcal{S}_v(u) \setminus \mathcal{S}_{v_n}(u_n)\big) = 0. \tag{5.11}$$

Therefore, ψ is lower semi-continuous. Because the branch point map c is continuous due to Lemma 2.16, the same applies to $\eta(u, v, w) = \psi((u, c(u, v, w))$, and η is measurable. □

Given a vector $\underline{u} = (u_1, \dots, u_m) \in T^m$, $m \in \mathbb{N}$, we consider the masses of all the subtrees we obtain as branch points of entries of $\underline{u}$. To this end, let

$$\underline{\eta}(u, v, w) := \big(\eta(u, v, w),\ \eta(v, u, w),\ \eta(w, u, v)\big) \tag{5.12}$$

and define the function $\mathfrak{m}_\chi\colon T^m \to [0, 1]^{3\cdot\binom{m}{3}}$, given by

$$\mathfrak{m}_\chi(\underline{u}) := \big(\underline{\eta}(u_i, u_j, u_k)\big)_{1\leq i<j<k\leq m} \tag{5.13}$$

Definition 5.12 (subtree-mass tensor distribution)**.** For $\chi = (T, c, \mu) \in \mathbb{T}_2$ and $m \in \mathbb{N}$, the *m-subtree-mass tensor distribution* of χ is defined by

$$\vartheta_m(\chi) := \mu^{\otimes m} \circ \mathfrak{m}_\chi^{-1} \in \mathcal{M}_1\big([0, 1]^{3\cdot\binom{m}{3}}\big), \tag{5.14}$$

Example 5.13 (symmetric binary tree)**.** Let for each $n \in \mathbb{N}$, $\chi_n = (T_n, c_n, \mu_n)$ the symmetric binary tree with $N = 2^n$ leaves and the uniform distribution on the set of leaves. Then the 3-*subtree-mass tensor distribution* of χ_n is equal to

$$\begin{aligned}
\vartheta_3(\chi_n) &= \mu_n^{\otimes 3} \circ \mathfrak{m}_\chi^{-1} \\
&= \sum_{k=1}^{n-1} \frac{1-2^{-k}}{2^{k+1}} \Big(\delta_{(\frac{1}{2^{k+1}},\frac{1}{2^{k+1}},1-\frac{1}{2^k})} + \delta_{(\frac{1}{2^{k+1}},1-\frac{1}{2^k},\frac{1}{2^{k+1}})} + \delta_{(1-\frac{1}{2^k},\frac{1}{2^{k+1}},\frac{1}{2^{k+1}})}\Big) \\
&\quad + \tfrac{1}{N}(1-\tfrac{1}{N})\big(\delta_{(\frac{1}{N},\frac{1}{N},1)} + \delta_{(\frac{1}{N},1,\frac{1}{N})} + \delta_{(1,\frac{1}{N},\frac{1}{N})}\big) + \tfrac{1}{N^2}\delta_{(\frac{1}{N},\frac{1}{N},\frac{1}{N})} \\
&\underset{n\to\infty}{\longrightarrow} \sum_{k=1}^{\infty} \frac{1-2^{-k}}{2^{k+1}} \Big(\delta_{(\frac{1}{2^{k+1}},\frac{1}{2^{k+1}},1-\frac{1}{2^k})} + \delta_{(\frac{1}{2^{k+1}},1-\frac{1}{2^k},\frac{1}{2^{k+1}})} + \delta_{(1-\frac{1}{2^k},\frac{1}{2^{k+1}},\frac{1}{2^{k+1}})}\Big)
\end{aligned} \tag{5.15}$$

◊

Remark 5.14 (3-subtree-mass tensor distribution is not enough)**.** It is not enough to consider only the 3-*subtree-mass tensor distribution.* Indeed, ϑ_3 cannot distinguish all non-isomorphic binary algebraic measure trees, i.e. it does not separate the points of $\mathbb{T}_2$. To see this, take the tree from Figure 9 with uniform distribution on its 12 leaves, and the same tree with the subtrees marked by $\times$ and $\circ$, respectively, exchanged. These two trees are clearly non-isomorphic, and because the two marked subtrees have the same number of leaves, every vertex in one tree corresponds to a vertex in the other with the same value for $\mathfrak{m}_\chi$. ◊

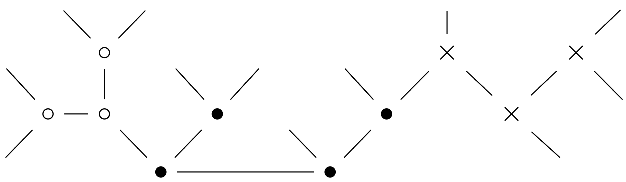


FIGURE 9. μ is the uniform distribution on the leaves. Swap the $\circ$-part with the $\times$-part to obtain a non-isomorphic tree giving the same value for ϑ_3.

We consider the weakest topology on $\mathbb{T}_2$ such that for every $m \in \mathbb{N}$ the m-subtree-mass tensor distribution is continuous. Here, as usual, we equip $\mathcal{M}_1([0,1]^{3\cdot\binom{m}{3}})$ with the weak topology.

Definition 5.15 (sample subtree-mass topology)**.** The topology induced on $\mathbb{T}_2$ by the set $\{\vartheta_m : m \in \mathbb{N}\}$ of subtree-mass tensor distributions is called *sample subtree-mass topology.*

We say that a sequence $(\chi_n)_{n\in\mathbb{N}}$ is *sample subtree-mass convergent* to χ in $\mathbb{T}_2$ if it converges w.r.t. the sample subtree-mass topology, i.e. if $\vartheta_m(\chi_n)$ converges to $\vartheta_m(\chi)$ as $n \to \infty$ for every $m \in \mathbb{N}$. To see that the sample subtree-mass topology is a Hausdorff topology on $\mathbb{T}_2$, we need the following reconstruction theorem.

Proposition 5.16 (reconstruction theorem)**.** *The set of subtree-mass tensor distributions $\{\vartheta_m : m \in \mathbb{N}\}$ separates points of $\mathbb{T}_2$, i.e., if $\chi_1, \chi_2 \in \mathbb{T}_2$ are such that $\vartheta_m(\chi_1) = \vartheta_m(\chi_2)$ for all $m \in \mathbb{N}$, then $\chi_1 = \chi_2$.*

Proof. We always assume that the representative (T, c, μ) of an algebraic measure tree is chosen such that $\mu(\mathcal{S}_v(u)) > 0$ whenever $u, v \in T$, $u \neq v$.

Because the set $\{\mathfrak{S}_m : m \in \mathbb{N}\}$ of tree shape distributions separates points by Corollary 5.10, it is enough to show that $\mathfrak{S}_m$ is determined by the m-subtree-mass tensor distribution ϑ_m for every $m \in \mathbb{N}$. We do so by showing that there exists a (non-continuous) function $h\colon [0,1]^{3\cdot\binom{m}{3}} \to \mathfrak{C}_m$ such that for every $\chi = (T, c, \mu) \in \mathbb{T}_2$ we have $\mathfrak{s}_T = h \circ \mathfrak{m}_\chi$ on $\big(T \setminus \mathrm{br}(T,c)\big)^m$. This is enough, because $\mu(\mathrm{br}(T,c)) = 0$ by countability of $\mathrm{br}(T,c)$ and the assumption that $\mathrm{at}(\mu) \subseteq \mathrm{lf}(T,c)$.

Fix $\underline{u} = (u_1, ..., u_m) \in (T \setminus \mathrm{br}(T,c))^m$ and set $C = (C, c_C, \ell) := \mathfrak{s}_T(\underline{u})$. For $i \neq j$, we have $u_i = u_j$ if and only if $\eta(u_i, u_j, u_k) = \eta(u_j, u_i, u_k) = 0$ for any and hence all $k \in \{1, ..., m\} \setminus \{i, j\}$. Thus, we can determine multiple labels of C by $\mathfrak{m}_\chi(\underline{u})$ and may assume in the following that $u_1, ..., u_m$ are distinct. Then, the m-labelled cladogram C is uniquely determined by the set of pairs $(\underline{x}_1, \underline{x}_2)$ of triples $\underline{x}_i = (x_{i,1}, x_{i,2}, x_{i,3}) \in \{u_1, ..., u_m\}^3$, $x_{i,j} \neq x_{i,k}$ for $j \neq k$, $i = 1, 2$, such that

$$c_C\big(x_{1,1},\, x_{1,2},\, x_{1,3}\big) = c_C\big(x_{2,1},\, x_{2,2},\, x_{2,3}\big). \tag{5.16}$$

We claim that (5.16) holds if and only if we can reorder the three entries of $\underline{x}_2$ such that we can replace every entry of $\underline{x}_1$ by the corresponding entry of $\underline{x}_2$ and obtain the same masses of subtrees. More precisely,

$$\underline{\eta}(x_{1,1}, x_{1,2}, x_{1,3}) = \underline{\eta}(x_{i,1}, x_{j,2}, x_{k,3}) \quad \forall i, j, k \in \{1, 2\}. \tag{5.17}$$

Indeed, if $c_C(\underline{x}_1) = c_C(\underline{x}_2)$, then $c(\underline{x}_1) = c(\underline{x}_2)$ by definition of $\mathfrak{s}_T$. Because none of the u_i is a branch point, every component of $T \setminus \{c(\underline{x}_1)\}$ contains precisely one of the $x_{1,i}$, as well as one of the $x_{2,i}$. We can reorder the entries of x_2 such that $x_{1,i}$ is in the same component as $x_{2,i}$, $i = 1, \ldots, 3$. Then it is easy to check that (5.17) holds.

Conversely, assume that $c_C(\underline{x}_1) \neq c_C(\underline{x}_2)$. Because the restriction of the tree homomorphism $C \to c(\{u_1, \ldots, u_m\}^3)$ to the branch points of C is injective, this implies $v_1 := c(\underline{x}_1) \neq c(\underline{x}_2) =: v_2$. There must be an i with $x_{1,i} \in \mathcal{S}_{v_1}(v_2)$, say $i = 3$. Also, $x_{2,j} \in \mathcal{S}_{v_1}(v_2)$ for at least two different j, so at least one which is different from i, say $j = 2$ (see Figure 10). Then $v_3 := c(x_{1,1}, x_{2,2}, x_{1,3}) \in$

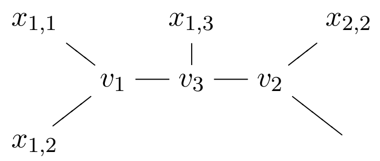


FIGURE 10. The situation in the proof of Proposition 5.16.

$\mathcal{S}_{v_1}(v_2)$, and in particular, $x_{1,1}, x_{1,2} \in \mathcal{S}_{v_3}(x_{1,1})$. Thus $\eta(\underline{x}_1) < \eta(x_{1,1}, x_{2,2}, x_{1,3})$, and (5.17) does not hold. □

Corollary 5.17 (metrizability)**.** *The sample subtree-mass topology is metrizable.*

Proof. Because the sample subtree-mass topology is induced by a countable family of functions $(\vartheta_m)_{m\in\mathbb{N}}$ with values in metrizable spaces, it is pseudo-metrizable. By Proposition 5.16, it is a Hausdorff topology, hence it is metrizable. □

In analogy to the sets Π_ι and $\Pi_\mathfrak{s}$ of polynomials and shape polynomials, respectively, the sample subtree-mass topology also comes with a canonical set of test functions. We call $\Psi\colon \mathbb{T}_2 \to \mathbb{R}$ *subtree-mass polynomial* if there is $m \in \mathbb{N}$ and $\psi \in \mathcal{C}_b([0,1]^{3\cdot\binom{m}{3}})$ with

$$\Psi(\chi) = \int_{[0,1]^{3\cdot\binom{m}{3}}} \psi \,\mathrm{d}\vartheta_m(\chi) = \int_{T^m} \psi \circ \mathfrak{m}_\chi \,\mathrm{d}\mu^{\otimes m} \tag{5.18}$$

We also define

$$\Pi_\mathfrak{m} := \{\, \text{subtree-mass polynomials on } \mathbb{T}_2 \,\}. \tag{5.19}$$

Obviously, the sample subtree-mass topology is induced by the set $\Pi_\mathfrak{m}$ of subtree-mass polynomials.

Proposition 5.18 (sample shape convergence implies sample subtree-mass convergence)**.** *The sample shape topology is stronger than the sample subtree-mass topology.*

Proof. The proof is similar to that of Proposition 5.9. We will show that each subtree-mass polynomial in $\Psi \in \Pi_\mathfrak{m}$,

$$\Psi(T,c,\mu) = \int_{T^m} \psi\big(\big(\underline{\eta}(u_i,u_j,u_k)\big)_{1\le i<j<k\le m}\big)\,\mu^{\otimes m}(\mathrm{d}\underline{u}), \tag{5.20}$$

with $m \in \mathbb{N}$ and $\psi \in \mathcal{C}([0,1]^{3\cdot\binom{m}{3}})$ Lipschitz continuous w.r.t. the ℓ_∞-Norm on $[0,1]^{3\cdot\binom{m}{3}}$ is in the uniform closure of $\Pi_\mathfrak{s}$. Let L be the Lipschitz constant of Ψ. For $n \in \mathbb{N}$ with $n \ge m$, we define

$$\Phi_n(T,c,\mu) := \int_{T^n} \psi\big(\big(\underline{\eta}^{\mu_{n,\underline{u}}}(u_i,u_j,u_k)\big)_{1\le i<j<k\le m}\big)\,\mu^{\otimes n}(\mathrm{d}\underline{u}), \tag{5.21}$$

where $\underline{\eta}^{\mu_{n,\underline{u}}}$ is defined in the same way as $\underline{\eta}$, but with μ replaced by the empirical sample distribution

$$\mu_{n,\underline{u}} := \tfrac{1}{n}\sum_{\ell=1}^{n} \delta_{u_\ell}. \tag{5.22}$$

Note that $\Phi_n \in \Pi_\mathfrak{s}$ because whether or not $u_\ell \in \mathcal{S}_{c(u_i,u_j,u_k)}(u_i)$ for some $\ell \in \{1,\dots,n\}$, $i,j,k \in \{1,\dots,m\}$ depends only on the shape $\mathfrak{s}_T(\underline{u})$.

Finally, applying the uniform Glivenko-Cantelli estimate Lemma A.4, we have

$$\|\Psi - \Phi_n\|_\infty \le \sup_{(T,c,\mu)\in\mathbb{T}_2} \int_{T^n} L \cdot \sup_{S\in\mathcal{S}_T} \big|\mu(S) - \mu_{n,\underline{u}}(S)\big|\,\mu^{\otimes n}(\mathrm{d}\underline{u}) \le L\epsilon_n \underset{n\to\infty}{\longrightarrow} 0, \tag{5.23}$$

where $\mathcal{S}_T := \big\{\mathcal{S}_v(u) : u,v \in T\big\}$ and $(\epsilon_n)_{n\in\mathbb{N}} \underset{n\to\infty}{\longrightarrow} 0$. We note that $\dim_{\mathrm{VC}}(\mathcal{S}_T) \le 3$ (compare Example A.3). □

5.3. **Equivalence and compactness of topologies.** In this section, we show that sample shape convergence (Definition 5.8), sample subtree-mass convergence (Definition 5.15) and branch point distribution distance Gromov-weak convergence (Definition 3.7) on $\mathbb{T}_2$ are equivalent. While spaces of metric measure spaces are usually far from being locally compact, $\mathbb{T}_2$ is in this topology even a compact metrizable space.

Theorem 3 (equivalence of topologies and compactness)**.** *The sample shape topology, the sample subtree-mass topology, and the bpdd-Gromov-weak topology coincide on $\mathbb{T}_2$. Furthermore, $\mathbb{T}_2$ is compact and metrizable in this topology.*

Because compact subsets of a Hausdorff space are closed, a direct corollary is that unlike the situation in the space of metric measure trees (with Gromov-weak or Gromov-Hausdorff-weak topology), the set of binary trees is closed w.r.t. the bpdd-Gromov-weak topology.

Corollary 5.19. *The subspace $\mathbb{T}_2$ of binary algebraic measure trees with atoms restricted to leaves is closed in $\mathbb{T}$ (with bpdd-Gromov-weak topology).*

As a preparation of the proof for the theorem, we show that binary algebraic measure trees depend continuously on their encoding as sub-triangulations of the circle. Together with Proposition 5.9, this also finishes the proof of Theorem 2. Recall the space $\mathcal{T}$ of sub-triangulations of the circle equipped with the Hausdorff metric topology from (4.3), and the coding map $\tau\colon \mathcal{T} \to \mathbb{T}_2$ from Theorem 2.

Lemma 5.20 (continuity of the coding map)**.** *Let $\mathbb{T}_2$ be equipped with the sample shape topology, and $\mathcal{T}$ with the Hausdorff metric topology. Then the coding map $\tau\colon \mathcal{T} \to \mathbb{T}_2$ is continuous.*

Proof. Fix $C \in \mathcal{T}$ and $m \in \mathbb{N}$. By definition of the sample shape topology, it is enough to show that $\mathfrak{S}_m \circ \tau\colon \mathcal{T} \to \mathcal{M}_1(\mathfrak{C}_m)$ is continuous at C. Let $U_1, ..., U_m$ be i.i.d. points on the circle $\mathbb{S}$ chosen with the Lebesgue measure.

Recall from (4.2) the set $\nabla(C)$ of connected components of $\mathbb{D}\setminus\mathrm{conv}(C)$, from (4.7) the connected component $\mathrm{comp}_x(y)$ of $\mathbb{D} \setminus \partial_{\mathbb{D}} x$ which contains y, where $x \in \Delta(C) \cup \nabla(C)$, and $y \subseteq \mathbb{D}$ connected and disjoint from $\partial_{\mathbb{D}} x$. Furthermore, recall the set $\square(C)$ from (4.8), and the subtree components $\mathcal{S}_x(y)$ from (1.9).

For $\epsilon > 0$, there exists $N = N_{C,m,\epsilon} \in \mathbb{N}$ and $v_1, ..., v_N \in \Delta(C) \cup \nabla(C)$ distinct such that with probability at least $1 - \epsilon$ the following holds:

- if $\{U_1, ..., U_m\} \cap v \neq \emptyset$ for $v \in \nabla(C)$, then $v \in \{v_1, ..., v_N\}$, and
- if $\{U_1, ..., U_m\} \cap \mathrm{comp}_v(w) \neq \emptyset$ for some $v \in \Delta(C)$ and all $w \in \Delta(C) \cup \nabla(C) \cup \square(C)$ with $w \neq v$, then $v \in \{v_1, ..., v_N\}$.

Put $\epsilon' := \epsilon \cdot (12mN)^{-1}$. Then

$$\mathbb{P}\big(\big\{d(U_i, \partial v_j) \geq \epsilon',\, \forall\, i = 1, ..., m; j = 1, ..., N\big\}\big) \geq 1 - \epsilon. \tag{5.24}$$

There is a $\delta = \delta(\epsilon) > 0$ sufficiently small such that for any $C' \in \mathcal{T}$ with Hausdorff metric $d_{\mathrm{H}}(C, C') < \delta$ there are distinct $v'_1, ..., v'_N \in \Delta(C') \cup \nabla(C')$ such that $d_{\mathrm{H}}(v_i, v'_i) \leq \epsilon'$ for $i = 1, ..., N$. Let $\chi = (T, c, \mu) := \tau(C)$, and $V_1, ..., V_m$ be i.i.d. μ-distributed, coupled to $U_1, ..., U_m$ such that $V_k \in \mathcal{S}_v(w)$ if and only if $U_k \in \mathrm{comp}_v(w)$, which is possible due to the properties of τ established in Theorem 2. Define χ' and $V'_1, ..., V'_m$ similarly with C' instead of C. Then

$$\mathbb{P}\big(\big\{\mathfrak{s}_T(V_1, ..., V_m) = \mathfrak{s}_{T'}(V'_1, ..., V'_m)\big\}\big) \geq 1 - 2\epsilon, \tag{5.25}$$

which implies that $d_{\mathrm{Pr}}\big(\mathfrak{S}_m(\tau(C)),\, \mathfrak{S}_m(\tau(C'))\big) \leq 2\epsilon$ (with d_{Pr} denoting the Prokhorov distance). This shows that $\mathfrak{S}_m \circ \tau$ is continuous at C and, since m and C are arbitrary, that τ is continuous. $\square$

Now we are in a position to combine our results to a proof of the main theorem of Section 5.

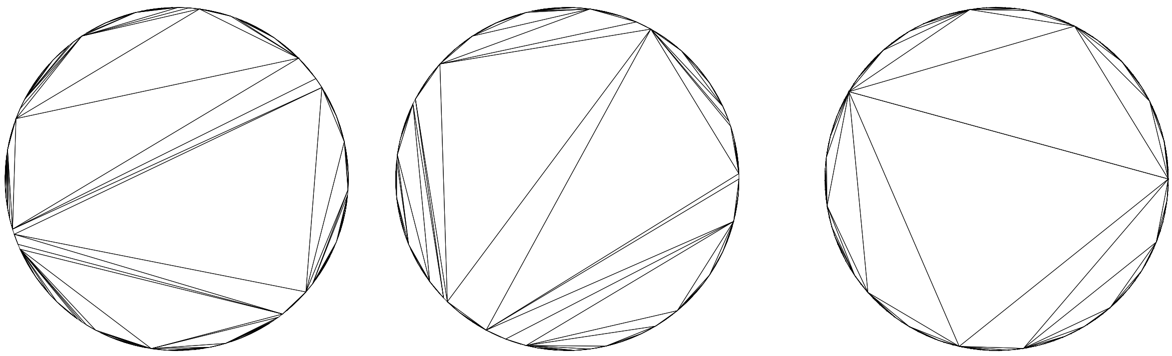


FIGURE 11. Realisations of β-splitting trees for (from left to right) $\beta = -1$, $\beta = 0$ (Yule tree), $\beta = 10$.

Proof of Theorem 3. The space $\mathcal{T}$ of sub-triangulations of the circle with Hausdorff metric topology is compact according to Lemma 4.2. The coding map $\tau\colon \mathcal{T} \to \mathbb{T}_2$ is surjective by Theorem 2, and continuous when $\mathbb{T}_2$ is equipped with the sample shape topology by Lemma 5.20. Therefore, the sample shape topology is a compact topology on $\mathbb{T}_2$. Moreover, the sample shape topology is Hausdorff by Corollary 5.10. As the sample subtree-mass topology is a weaker Hausdorff topology by Proposition 5.18 and Corollary 5.17, it coincides with the sample shape topology. The same is true for the bpdd-Gromov-weak topology by Proposition 5.9. □

Recall from Remark 3.10 that the set of distance polynomials is convergence determining for measures on $\mathbb{T}_2$. It directly follows from the construction that the same is true for the sets of shape polynomials and subtree-mass polynomials. This property is very useful for proving convergence in law of random variables.

Corollary 5.21 (convergence determining classes of functions)**.** *The sets* $\Pi_{\mathfrak{s}} \subseteq \mathcal{C}_b(\mathbb{T}_2)$ *(defined in* (5.3)*) and* $\Pi_{\mathfrak{m}}$ *(defined in* (5.18)*) are convergence determining for measures on* $\mathbb{T}_2$ *with bpdd-Gromov-weak topology.*

Proof. $\mathbb{T}_2$ is a compact metrizable space, and both $\Pi_{\mathfrak{s}}$ and $\Pi_{\mathfrak{m}}$ induce the bpdd-Gromov-weak topology on $\mathbb{T}_2$ by Theorem 3. Furthermore, each of $\Pi_{\mathfrak{s}}$ and $\Pi_{\mathfrak{m}}$ is closed under multiplication. Thus the claim follows by the Stone-Weierstrass theorem. □

6. EXAMPLE: SAMPLING CONSISTENT FAMILIES

Consider a family $(T_n, c_n)_{n\in\mathbb{N}}$ of random, finite binary (algebraic) trees, where (T_n, c_n) has n leaves. Let K_n be the Markov kernel that takes such a tree and removes a leaf uniformly chosen at random, together with the branch point it is attached to, thus obtaining a binary tree with $n-1$ leaves. We say that the family is *sampling consistent* if $K_n(T_n, \cdot\,) = \mathcal{L}(T_{n-1})$, where $\mathcal{L}$ denotes the law of a random variable.

Example 6.1 (β-splitting trees)**.** For every $\beta \in [-2, \infty]$, let T_n^β be the β-splitting tree on n leaves from [Ald96] (with forgotten labels). For $-2 < \beta < \infty$, the β-splitting tree T_n^β can be constructed recursively as follows. T_2^β consists of two leaves connected by a distinguished root edge. If $n > 2$, choose $i \in \{1, \dots, n-1\}$ with probability

$$q_n^\beta(i) = \frac{1}{a_n(\beta)} \binom{n}{i} \int_0^1 x^{i+\beta}(1-x)^{n-i+\beta}\,\mathrm{d}x, \tag{6.1}$$

where $a_n(\beta)$ is a normalisation constant. Then construct two independent β-splitting trees T_i^β and T_{n-i}^β, introduce a new branch point in the middle of each of the two root edges, and connect these new branch points with the new root edge to obtain T_n^β.

It is easy to see (and observed in [Ald96]) that $(T_n^\beta)_{n\in\mathbb{N}}$ is sampling consistent. Note the special cases $\beta = -2$ which is the *comb tree*, $\beta = -\frac{3}{2}$ which is the *uniform cladogram*, $\beta = 0$ which is the *Yule tree* and $\beta = \infty$ which is the *symmetric binary tree*. See Figure 11 for triangulations of a realization of β-splitting trees for different values of β and large n. The Aldous Brownian CRT, which is the limit for $\beta = -\frac{3}{2}$, is shown in Figure 4. $\Diamond$

Lemma 6.2 (convergence of sampling consistent families)**.** *Let $((T_n, c_n))_{n\in\mathbb{N}}$ be a sampling consistent family of random binary trees, and μ_n the uniform distribution on $\mathrm{lf}(T_n, c_n)$. Then we have the convergence in law*

$$(T_n, c_n, \mu_n) \xrightarrow[n\to\infty]{\mathcal{L}} (T, c, \mu) \quad \text{on } \mathbb{T}_2 \text{ with bpdd-Gromov-weak topology} \tag{6.2}$$

for some random algebraic measure tree $(T, c, \mu) \in \mathbb{T}_2$ with non-atomic measure μ.

Proof. Recall the m-tree shape distribution $\mathfrak{S}_m$ from Definition 5.8. Let $n, m \in \mathbb{N}$ with $m < n$ and define

$$\epsilon_{n,m} := \mu_n^{\otimes m}\big\{x \in T^m : x_1, \ldots, x_m \text{ not distinct}\big\} \le \tfrac{m^2}{n}. \tag{6.3}$$

Because (T_n) is sampling consistent, we obtain for the annealed shape distribution

$$\mathbb{E}\big(\mathfrak{S}_m(T_n, c_n, \mu_n)\big) = (1 - \epsilon_{n,m})\mathcal{L}(T_m^*) + \epsilon_{n,m}\mu_{n,m}, \tag{6.4}$$

where T_m^* is obtained from T_m by randomly labelling the leaves, and $\mu_{n,m} \in \mathcal{M}_1(\mathfrak{C}_m)$ is some law of m-labelled cladograms supported by cladograms where at least one leaf has more than one label. This shows that, for every fixed m, the expected m-tree shape distribution converges as $n \to \infty$. Because the m-tree shape distribution is convergence determining for the bpdd-Gromov-weak topology by Corollary 5.21, all limit points of $\mathcal{L}(T_n, c_n, \mu_n)$ in $\mathcal{M}_1(\mathbb{T}_2)$ coincide. According to Theorem 3, $\mathbb{T}_2$, and hence $\mathcal{M}_1(\mathbb{T}_2)$, is compact and thus a unique limit exists. That the limiting measure is non-atomic is obvious, because the probability that a sampled shape is single-labelled tends to one by (6.4). $\square$

Example 6.3 (β-splitting trees continued)**.** By Lemma 6.2, for every $\beta \in [-2, \infty]$, the sequence $(T_n^\beta, c_n^\beta, \mu_n^\beta)_{n\in\mathbb{N}}$ of increasing β-splitting trees converges in distribution to some limiting random algebraic measure tree $(T^\beta, c^\beta, \mu^\beta)$. In the case of the uniform cladogram ($\beta = -\frac{3}{2}$), the limit is the Brownian algebraic continuum random tree which can be obtained as tree $\tau(C_{\mathrm{CRT}})$ coded by the Brownian triangulation (see Example 4.5), or as the algebraic measure tree induced by the metric measure Brownian CRT which is known to have uniform shape distribution ([Ald93]). In the case of the comb tree ($\beta = -2$), the limit is the unit interval with Lebesgue measure (a coding triangulation is shown in the very right of Figure 6). $\Diamond$

Appendix A. A uniform Glivenko-Cantelli theorem

In Subsections 5.1 and 5.2 we made use of uniform estimates of the speed of convergence in the approximation of the branch point distribution and the measure of a algebraic measure tree by empirical distribution. Such uniform Glivenko-Cantelli estimates under a bound on the Vapnik-Chervonenkis dimension (VC-dimension) of the type presented below should be well-known. As we did not find it explicitly in sufficient generality in the literature, we will present it here.

We recall the definition of VC-dimension, going back to the seminal work of Vapnik and Chervonenkis, [VC71]. Let E be a non-empty set and $\mathcal{I}$ a non-empty collection of subsets of E. For $n \in \mathbb{N}$ and $x \in E^n$, put

$$\mathcal{I}(x) := \big\{(\mathbb{1}_I(x_1), \ldots, \mathbb{1}_I(x_n)) : I \in \mathcal{I}\big\} \subseteq \{0,1\}^n. \tag{A.1}$$

Then obviously, $1 \le \#\mathcal{I}(x) \le 2^n$.

Definition A.1 (Vapnik-Chervonenkis dimension)**.** The *Vapnik-Chervonenkis dimension* of $\mathcal{I}$ is defined as

$$\dim_{\mathrm{VC}}(\mathcal{I}) := \sup\bigl\{n \in \mathbb{N} : \max_{x \in E^n} \#\mathcal{I}(x) = 2^n\bigr\}. \tag{A.2}$$

Example A.2 (collection of intervals of an algebraic tree)**.** Let (T, c) be a separable algebraic tree with $\#T > 2$, and

$$\mathcal{I} := \mathcal{I}_T := \bigl\{[u, v] : u, v \in T\bigr\}. \tag{A.3}$$

For $x_1, x_2, u \in T$ distinct, we have $\#\mathcal{I}(x) \geq \#\{[u, u], [x_1, x_1], [x_2, x_2], [x_1, x_2]\} = 2^2$, hence $\dim_{\mathrm{VC}}(\mathcal{I}_T) \geq 2$. Conversely, for $x \in T^3$, either there is $u, v \in T$ with $x_1, x_2, x_3 \in [u, v]$. Then w.l.o.g. $x_2 \in [x_1, x_3]$ and $(1, 0, 1) \notin \mathcal{I}_T(x)$. Or there is no such $u, v \in T$, which means $(1, 1, 1) \notin \mathcal{I}_T(x)$. Therefore,

$$\dim_{\mathrm{VC}}(\mathcal{I}_T) = 2. \tag{A.4}$$

$\diamondsuit$

Recall the notion $\mathcal{S}_x(y)$ of the equivalence class of $T \setminus \{x\}$ containing y.

Example A.3 (collection of subtrees branching of a branch point)**.** Let (T, c) be a separable algebraic tree, and

$$\mathcal{I} := \mathcal{S}_T := \bigl\{\mathcal{S}_v(u) : u, v \in T\bigr\}. \tag{A.5}$$

We claim that

$$\dim_{\mathrm{VC}}(\mathcal{S}_T) \leq 3. \tag{A.6}$$

For this upper bound, let $x = (x_1, x_2, x_3, x_4) \in T^4$. By the 4-point condition of the branch point map, we can assume w.l.o.g. that

$$c(x_1, x_2, x_3) = c(x_1, x_2, x_4). \tag{A.7}$$

In this case, it is not possible to cover $\{x_1, x_3\}$ but neither x_2 nor x_4 with a single subtree in $\mathcal{S}_T$, which proves the claim. $\diamondsuit$

The constant in front in the following Glivenko-Cantelli lemma is clearly not optimal. For us it is only important that it is universal and not depending on the measure space (E, μ).

Lemma A.4 (rate of convergence in Glivenko-Cantelli)**.** *Let E be a Polish space, μ a probability measure on E, $(X_n)_{n\in\mathbb{N}}$ i.i.d. μ-distributed, and $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$ the empirical measure. Then, for every $\mathcal{I} \subseteq \mathcal{B}(E)$ with $\dim_{\mathrm{VC}}(\mathcal{I}) < \infty$ and $n > 1$,*

$$\mathbb{E}\Bigl(\sup_{I \in \mathcal{I}} \bigl|\mu(I) - \mu_n(I)\bigr|\Bigr) \leq 96 \sqrt{\frac{\dim_{\mathrm{VC}}(\mathcal{I})}{n}}. \tag{A.8}$$

Proof. By the celebrated Kuratowski isomorphism theorem, all uncountable Polish spaces are Borel-isomorphic. Therefore, we may assume w.l.o.g. that $E = \mathbb{R}$. Theorem 3.2 in [DL01] yields

$$\Delta := \mathbb{E}\bigl(\sup_{I \in \mathcal{I}} \bigl|\mu(I) - \mu_n(I)\bigr|\bigr) \leq \frac{24}{\sqrt{n}} \sup_{x \in \mathbb{R}^n} \int_0^1 \sqrt{\log\bigl(2N(r, \mathcal{I}(x))\bigr)}\, \mathrm{d}r, \tag{A.9}$$

where $N(r, \mathcal{I}(x))$ is the covering number of $\mathcal{I}(x)$ w.r.t. the metric $\frac{1}{\sqrt{n}} \cdot d_{\ell^2}$, where d_{ℓ^2} is the Euclidean metric on $\{0, 1\}^n$. This covering number can be upper-bounded in terms of the separation number $M(r, \mathcal{I})$ w.r.t. the metric $\frac{1}{n} \cdot d_{\ell^1}$ used by Haussler in [Hau95], and Theorem 1 there yields

$$N\bigl(r, \mathcal{I}(x)\bigr) \leq M\bigl(r^2, \mathcal{I}(x)\bigr) \leq e(\dim_{\mathrm{VC}}(\mathcal{I}) + 1)\Bigl(\frac{2e}{r^2}\Bigr)^{\dim_{\mathrm{VC}}(\mathcal{I})}, \tag{A.10}$$

provided that $nr^2 \in \mathbb{N}$. For $r^2 \le \frac{1}{n}$, we use the trivial estimate $M(r^2, \mathcal{I}(x)) \le 2^n$. For general $r^2 \ge \frac{1}{n}$, we estimate $M(r^2, \mathcal{I}(x)) \le M(\frac{1}{n}\lfloor nr^2 \rfloor, \mathcal{I}(x))$, and inserting (A.10) into (A.9) yields

$$\begin{aligned}\Delta &\le \frac{24}{\sqrt{n}}\Big(\sqrt{\tfrac{n+1}{n}} + \int_{\frac{1}{\sqrt{n}}}^{1} \sqrt{\log(2e(\dim_{\mathrm{VC}}(\mathcal{I})+1)) + \dim_{\mathrm{VC}}(\mathcal{I})\log\big(2e(r^2-\tfrac{1}{n})^{-1}\big)}\,\mathrm{d}r\Big) \\ &\le \frac{24}{\sqrt{n}}\sqrt{\dim_{\mathrm{VC}}(\mathcal{I})}\Big(\sqrt{\tfrac{n+1}{n}} + \int_0^1 \sqrt{3+\log(2e)-2\log(r)}\,\mathrm{d}r\Big),\end{aligned} \tag{A.11}$$

where we used that $\log(2e(d+1)) \le 3d$ for $d \ge 1$, and $r^2 - \frac{1}{n} \ge (r - \frac{1}{\sqrt{n}})^2$. The last bracket is less than 4 for $n > 1$, and the claim follows. □

References

[AD12] Romain Abraham and Jean-François Delmas, *A continuum-tree-valued Markov process*, Ann. Probab. **40** (2012), no. 3, 1167–1211.

[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** (2013), no. 14, 1–21.

[ADV10] Romain Abraham, Jean-François Delmas, and Guillaume Voisin, *Pruning a Lévy continuum random tree*, Electron. J. Probab. **15** (2010), no. 46, 1429–1473.

[Ald93] David Aldous, *The continuum random tree III*, Ann. Probab. **21** (1993), 248–289.

[Ald94a] ———, *Recursive self-similarity for random trees, random triangulations and Brownian excursion*, Ann. Probab. **22** (1994), no. 2, 527–545. MR 1288122

[Ald94b] ———, *Triangulating the circle, at random*, Amer. Math. Monthly **101** (1994), no. 3, 223–233.

[Ald96] David Aldous, *Probability distributions on cladograms*, Random discrete structures (Minneapolis, MN, 1993), IMA Vol. Math. Appl., vol. 76, Springer, New York, 1996, pp. 1–18. MR 1395604

[Ald00] David Aldous, *Mixing time for a Markov chain on cladograms*, Combin. Probab. Comput. **9** (2000), no. 3, 191–204.

[ALW17] Siva Athreya, Wolfgang Löhr, and Anita Winter, *Invariance principle for variable speed random walks on trees*, Ann. Probab. **45** (2017), no. 2, 625–667. MR 3630284

[BK10] Douglas Blount and Michael A. Kouritzin, *On convergence determining and separating classes of functions*, Stochastic Process. Appl. **120** (2010), no. 10, 1898–1907.

[BS01] Itai Benjamini and Oded Schramm, *Recurrence of distributional limits of finite planar graphs*, Electron. J. Probab. **6** (2001), no. 23, 13. MR 1873300

[BS15] Nicolas Broutin and Henning Sulzbach, *The dual tree of a recursive triangulation of the disk*, Ann. Probab. **43** (2015), no. 2, 738–781. MR 3306004

[Chi01] Ian Chiswell, *Introduction to Λ-trees*, World Scientific Publishing Co. Inc., River Edge, NJ, 2001.

[CHK15] Nicolas Curien, Bénédicte Haas, and Igor Kortchemski, *The CRT is the scaling limit of random dissections*, Random Structures Algorithms **47** (2015), no. 2, 304–327. MR 3382675

[CK14] Nicolas Curien and Igor Kortchemski, *Random non-crossing plane configurations: a conditioned Galton-Watson tree approach*, Random Structures Algorithms **45** (2014), no. 2, 236–260. MR 3245291

[CK15] ———, *Percolation on random triangulations and stable looptrees*, Probab. Theory Relalted Fields **163** (2015), no. 1-2, 303–337. MR 3405619

[CLG11] Nicolas Curien and Jean-François Le Gall, *Random recursive triangulations of the disk via fragmentation theory*, Ann. Probab. **39** (2011), no. 6, 2224–2270. MR 2932668

[Cur14] Nicolas Curien, *Dissecting the circle, at random*, Journées MAS 2012, ESAIM Proc., vol. 44, EDP Sci., Les Ulis, 2014, pp. 129–139. MR 3178612

[Das14] Abhijit Dasgupta, *Set theory*, Birkhäuser/Springer, New York, 2014, With an introduction to real point sets. MR 3157351

[DGP12] Andrej Depperschmidt, Andreas Greven, and Peter Pfaffelhuber, *Tree-valued Fleming-Viot dynamics with selection*, Ann. Appl. Probab. **22** (2012), no. 6, 2560–2615.

[DL01] Luc Devroye and Gábor Lugosi, *Combinatorial methods in density estimation*, Springer Series in Statistics, Springer-Verlag, New York, 2001. MR 1843146

[DMT96] Andreas Dress, Vincent Moulton, and Werner Terhalle, *T-theory: an overview*, European J. Combin. **17** (1996), no. 2-3, 161–175, Discrete metric spaces (Bielefeld, 1994). MR 97e:05069

[EGW17] Steven N. Evans, Rudolf Grübel, and Anton Wakolbinger, *Doob-Martin boundary of Rémy's tree growth chain*, Ann. Probab. **45** (2017), no. 1, 225–277. MR 3601650

[EPW06] Steven N. Evans, Jim Pitman, and Anita Winter, *Rayleigh processes, real trees, and root growth with re-grafting*, Probab. Theory Relalted Fields **134** (2006), no. 1, 81–126.

[Eva08] Steven N. Evans, *Probability and real trees*, Lecture Notes in Mathematics, vol. 1920, Springer, Berlin, 2008, Lectures from the 35^{th} Summer School on Probability Theory in Saint-Flour, 2005. MR 2351587

[EW06] Steven N. Evans and Anita Winter, *Subtree prune and re-graft: A reversible real-tree valued Markov chain*, Ann. Probab. **34** (2006), no. 3, 918–961.

[Fab15] Paul Fabel, *A topological characterization of the underlying spaces of complete R-trees*, Michigan Math. J. **64** (2015), no. 4, 881–887. MR 3426619

[FHP18] Noah Forman, Chris Haulk, and Jim Pitman, *A representation of exchangeable hierarchies by sampling from random real trees*, Probab. Theory Related Fields **172** (2018), no. 1-2, 1–29. MR 3851828

[For] Noah Forman, *Mass-structure of weighted real trees*, arXiv:1801.02700.

[For05] Daniel J. Ford, *Probabilities on cladograms: introduction to the alpha model*, 2005, arxiv:0511246.

[Fuk87] Kenji Fukaya, *Collapsing of Riemannian manifolds and eigenvalues of Laplace operators*, Invent. Math. **87** (1987), 517–547.

[FZ13] Hanspeter Fischer and Andreas Zastrow, *Combinatorial $\mathbb{R}$-trees as generalized Cayley graphs for fundamental groups of one-dimensional spaces*, Geom. Dedicata **163** (2013), 19–43. MR 3032684

[Glö12] Patric Karl Glöde, *Dynamics of Genealogical Trees for Autocatalytic Branching Processes*, Ph.D. thesis, FAU Erlangen-Nürnberg, 2012, `http://d-nb.info/1033029912/34`.

[GPW09] Andreas Greven, Peter Pfaffelhuber, and Anita Winter, *Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees)*, Probab. Theory Relalted Fields **145** (2009), 285–322.

[GPW13] ———, *Tree-valued resampling dynamics: Martingale Problems and applications*, Probab. Theory Relalted Fields **155** (2013), no. 3–4, 789–838.

[Gro99] Misha Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, vol. 152, Birkhäuser Boston Inc., Boston, MA, 1999.

[GSW16] Andreas Greven, Rongfeng Sun, and Anita Winter, *Continuum space limit of the genealogies of interacting Fleming-Viot processes on $\mathbb{Z}$*, Electron. J. Probab. **21** (2016), no. 58, 1–64. MR 3563886

[Hau95] David Haussler, *Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension*, J. Combin. Theory Ser. A **69** (1995), no. 2, 217–232. MR 1313896

[HW] Hui He and Matthias Winkel, *Invariance principles for pruning processes of GW-trees*, arxiv:1409.1014.

[HW19] ———, *Gromov-Hausdorff-Prokhorov convergence of vertex cut-trees of n-leaf Galton-Watson trees*, Bernoulli **25** (2019), no. 3, 23012329.

[KW] Sandra Kliem and Anita Winter, *Evolving phylogenies of trait-dependent branching with mutation and competition; Part I: Existence*, arxiv:1705.03277.

[LMW20] Wolfgang Löhr, Leonid Mytnik, and Anita Winter, *The Aldous chain on cladograms*, Ann. Probab. **in press** (2020^+).

[Löh13] Wolfgang Löhr, *Equivalence of Gromov-Prohorov- and Gromov's $\underline{\square}_\lambda$-metric on the space of metric measure spaces*, Electron. Commun. Probab. **18** (2013), no. 17, 1–10. MR 3037215

[Lov12] László Lovász, *Large networks and graph limits*, American Mathematical Society Colloquium Publications, vol. 60, American Mathematical Society, Providence, RI, 2012. MR 3012035

[LVW15] Wolfgang Löhr, Guillaume Voisin, and Anita Winter, *Convergence of bi-measure $\mathbb{R}$-trees and the pruning process*, Ann. Inst. H. Poincaré Probab. Statist. **51** (2015), no. 4, 1342–1368. MR 3414450

[MNO92] John C. Mayer, Jacek Nikiel, and Lex G. Oversteegen, *Universal spaces for* **R***-trees*, Trans. Amer. Math. Soc. **334** (1992), no. 1, 411–432. MR 1081940

[MO90] John C. Mayer and Lex G. Oversteegen, *A topological characterisation of R-trees*, Trans. Amer. Math. Soc. **320** (1990), no. 1, 395–415.

[MS84] John W. Morgan and Peter B. Shalen, *Valuations, trees, and degenerations of hyperbolic structures. I*, Ann. of Math. (2) **120** (1984), no. 3, 401–476. MR 769158

[Nus] Josué Nussbaumer, *The Kingman coalescent: old and new aspects*, `http://www-stud.uni-due.de/~snjonuss/master_thesis_kingman_coalescent.pdf`, Master thesis.

[Pio10] Sven Piotrowiak, *Dynamics of Genealogical Trees for Type- and State-dependent Resampling Models*, Ph.D. thesis, FAU Erlangen-Nürnberg, 2010, `http://d-nb.info/1009840509/34`.

[SS78] Lynn Arthur Steen and J. Arthur Seebach, Jr., *Counterexamples in topology*, second ed., Springer-Verlag, New York-Heidelberg, 1978. MR 507446

[Tit77] J. Tits, *A "theorem of Lie-Kolchin" for trees*, Contributions to algebra (collection of papers dedicated to Ellis Kolchin), Academic Press, New York, 1977, pp. 377–388. MR 0578488

[VC71] V. N. Vapnik and A. Ya. Chervonenkis, *On the uniform convergence of relative frequencies of events to their probabilities*, Theory Probab. Appl. **16** (1971), no. 2, 264–280, Translation of Russian original.

[Vil09] Cédric Villani, *Optimal transport*, Grundlehren der mathematischen Wissenschaften, vol. 338, Springer, Berlin-Heidelberg, 2009.

THE ALDOUS CHAIN ON CLADOGRAMS IN THE DIFFUSION LIMIT

WOLFGANG LÖHR, LEONID MYTNIK, AND ANITA WINTER

Abstract. In [Ald00], Aldous investigates a symmetric Markov chain on cladograms and gives bounds on its mixing and relaxation times. The latter bound was sharpened in [Sch02]. In the present paper we encode cladograms as binary, algebraic measure trees and show that this Markov chain on cladograms with a fixed number of leaves converges in distribution as the number of leaves tends to infinity. We give a rigorous construction of the limit as the solution of a well-posed martingale problem. The existence of a continuum limit diffusion was conjectured by Aldous, and we therefore refer to it as Aldous diffusion. We show that the Aldous diffusion is a Feller process with continuous paths, and the algebraic measure Brownian CRT is its unique invariant distribution.

Furthermore, we consider the vector of the masses of the three subtrees connected to a sampled branch point. In the Brownian CRT, its annealed law is known to be the Dirichlet distribution. Here, we give an explicit expression for the infinitesimal evolution of its quenched law under the Aldous diffusion.

Contents

1. Introduction

An N-*cladogram* is a semi-labelled, unrooted, binary tree with $N \geq 2$ *leaves* labelled $\{1, 2, ..., N\}$ and with $N-2$ unlabelled internal vertices. Cladograms are particular phylogenetic trees for which no information on the edge lengths is available, and which therefore only capture the tree structure. Reconstructing cladograms from DNA data is of major interest in population genetics. An important ingredient for several algorithms are Markov chains that move through a space of finite trees (see, for example, [Fel03] for a survey on Markov chain Monte Carlo algorithms in maximum likelihood tree reconstruction). Usually, such chains are based on a set of simple rearrangements that transform a tree into a "neighboring" tree (see, for example, [Fel03, BRST02, BHV01, AS01]).

The present paper considers (a continuous-time version of) the *Aldous chain* on cladograms, which is a Markov chain on the space $\mathfrak{C}_N$ of all N-cladograms. It has the following transition rates: for each pair (u, e) consisting of a leaf and an edge, at rate 1, the Markov chain jumps from its current state $\mathfrak{t}$ to $\mathfrak{t}^{(u,e)}$, where that latter is obtained as follows (see Figures 1 and 2). If u is not incident to e, then

- Erase the unique edge (including the incident vertices) incident to u,
- split the remaining subtree at the edge e into two pieces, and
- reintroduce the above edge (including u and the branch point) at the split point.

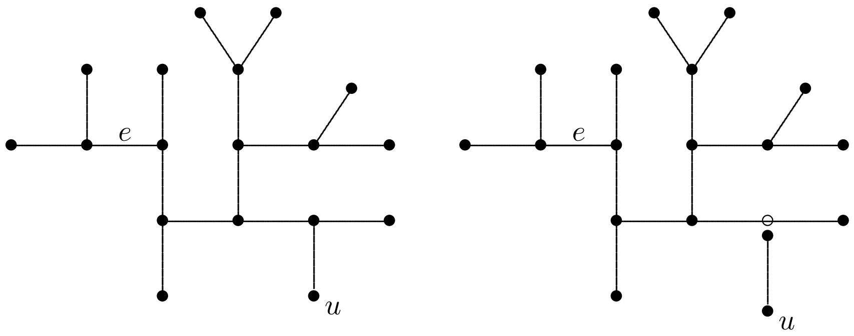


Figure 1. At rate $N(2N-3)$, a) a leaf u and an edge e are picked at random, and if e and u are not adjacent, b) the edge incident to u is taken away, leaving behind a branch point of degree 2. (continued in Figure 2)

Otherwise, if u is incident to e, we set $\mathfrak{t}^{(u,e)} = \mathfrak{t}$. In total, these so-called *Aldous moves* from $\mathfrak{t}$ to $\mathfrak{t}^{(u,e)}$ happen at rate $N(2N-3)$, and the rate of actual jumps of the Markov chain (where $\mathfrak{t}^{(u,e)} \neq \mathfrak{t}$) is $N(2N-6)$.

This Markov chain has the generator Ω_N, acting on all functions $\phi\colon \mathfrak{C}_N \to \mathbb{R}$ as follows:

$$\Omega_N \phi(\mathfrak{t}) = \sum_{(u,e)} \Big(\phi\big(\mathfrak{t}^{(u,e)}\big) - \phi(\mathfrak{t})\Big), \tag{1.1}$$

where the sum runs over all pairs (u,e) consisting of a leaf and an edge, and $\mathfrak{t} \in \mathfrak{C}_N$. Obviously, the Aldous chain is reversible, and the uniform distribution on $\mathfrak{C}_N$ is the stationary distribution. It was shown in [Ald00] that both mixing and relaxation time of the discrete-time chain are of order at least $\mathcal{O}(N^2)$, but at most of order $\mathcal{O}(N^3)$. [Sch02] verified that the relaxation time is of order $\mathcal{O}(N^2)$. Therefore, our continuous-time version has relaxation time of order 1.

As [Ald93] shows that a random N-cladogram with uniform edge lengths $\frac{1}{\sqrt{N}}$ converges weakly to the Brownian Continuum Random Tree (CRT), Aldous conjectured the existence of a CRT-symmetric diffusion limit of the Aldous chain on N-cladograms observed at time scale of order $\mathcal{O}(N^2)$ as $N \to \infty$. This conjecture was presented in a talk in March 1999 given at the Fields Institute, and is supported by the following calculation: suppose we start the Markov chain in some initial N-cladogram, fix a branch point, and consider the relative sizes (η_1, η_2, η_3) of the three subtrees attached to this branch point. Then, as the Markov chain runs, these proportions change as a certain Markov chain, until the branch point disappears. On the proposed time-rescaling of N^2, the $N \to \infty$ limit is the diffusion with generator

$$\Omega f(\underline{\eta}) = \sum_{1 \le i,j \le 3} \eta_i(\delta_{i,j} - \eta_i)\partial^2_{i,j} f(\underline{\eta}) - \tfrac{1}{2}\sum_{i=1}^{3}\big(1 - 3\eta_i\big)\partial_i f(\underline{\eta}) \tag{1.2}$$

which records certain aspects of a diffusion on the continuum tree. Aldous raised the question of how this diffusion should be constructed rigourously and what more can we calculate from there? On Aldous open problem website the construction was rated as straightforward provided the right set-up is chosen.

The present paper is demonstrating that indeed a straightforward construction can be given once we choose the right state space. A classical starting point would be to think of continuum trees as real trees which are particular metric spaces. A metric space is called a real tree if it is path connected and satisfies the so-called 4-point condition. For convergence results one would like to be in a position to treat the approximating discrete trees and their path-connected scaling limits in a unified way. One therefore also considers *metric trees* (introduced in [ALW17]), which are metric spaces differing from a real tree by not necessarily being path connected. A metric

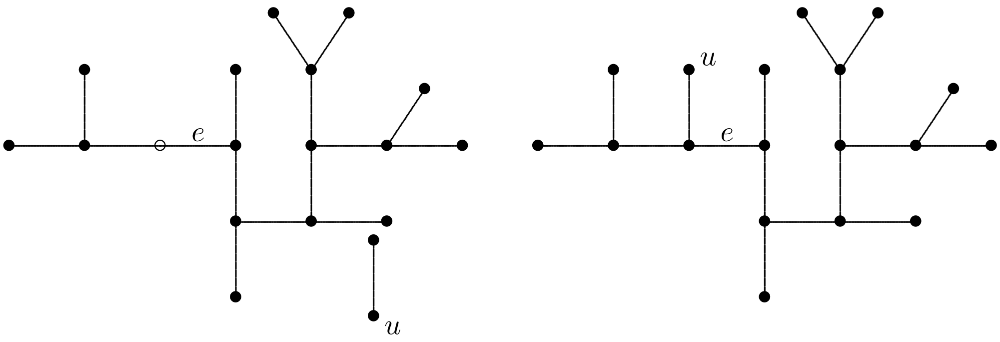


Figure 2. c) The two edges containing the branch point of degree 2 are identified, the edge e is opened, and d) the free edge is reattached there.

space is a metric tree if it can be embedded isometrically into a real tree in such a way that for every choice of three points in the metric tree, the corresponding branch point (defined in the real tree) belongs to the metric tree.

In many applications it is useful to have metric trees equipped with a probability measure as, for example, the definition of the discrete Aldous chain dynamics requires to sample leaves according to some probability measure. One therefore considers the space $\mathbb{M}$ of isometry classes of *metric measure spaces* and equips it with the Gromov-weak topology. In fact, Aldous's CRT arises as the Gromov-weak scaling limit of uniformly chosen N-cladograms with the uniform distribution on the leaves and edge lengths scaled down by the factor $\frac{1}{\sqrt{N}}$.

One of the equivalent definitions of the *Gromov-weak topology* is by convergence of the distance matrix distributions, i.e., a sequence $(\chi_N)_{N\in\mathbb{N}}$ of metric measure spaces converges to a metric measure space $\chi \in \mathbb{M}$ if and only if $\Phi(\chi_N) \underset{N\to\infty}{\longrightarrow} \Phi(\chi)$ for all test functions $\Phi\colon \mathbb{M} \to \mathbb{R}$ of the form

$$\Phi(\chi) := \int \phi\big((r(u_i,u_j))_{1\le i<j\le m}\big)\, \mu^{\otimes m}(\mathrm{d}\underline{u}), \tag{1.3}$$

where $\chi = (X, r, \mu)$, $m \in \mathbb{N}$, and $\phi \in \mathcal{C}_b(\mathbb{R}_+^{\binom{m}{2}})$ (see [GPW09, Löh13]).

In this set-up, many tree-valued Markov processes have been constructed and in some cases also the convergence of approximating discrete tree-valued dynamics has been established (see, for example, [EW06, GPW13, DGP12, KL15, LVW15]). One could think that metric (measure) trees are the natural framework for rescaling the Aldous chain as well. However, the Aldous chain resists this approach. An easy calculation shows that the quadratic variation of the averaged distance process rescales at time scale $N^{\frac{3}{2}}$. But how does it relate to the conjecture that the Aldous chain rescales on the time scale N^2? One reason might be that distances behave too wildly for tightness on that time scale to hold. Which in turn might be a hint that the naively used graph distance is not the notion of distance intrinsic to the Aldous chain dynamics. And indeed, one can argue that two points are close if the mass branching off the line segment connecting them is small rather than if the length of that line segment is small. The idea for our new state space is to overcome the metric issue by focusing on the tree structure only.

In what follows, we refer to (T, c) as an *algebraic tree* if $T \neq \emptyset$ is a set equipped with a *branch point map* $c\colon T^3 \to T$ satisfying consistency conditions (see Definition 2.1). Even though algebraic trees can be seen as metric trees where one has "forgotten" the metric (i.e., equivalence classes of metric trees), the branch point map is defined such that the notions of leaves, branch points, degree, subtrees, line segments, etc. can be formalized without reference to a metric (and agree with the corresponding notions in the metric tree). The Aldous diffusion takes values in the new state space $\mathbb{T}$ of (equivalence classes of) *algebraic measure trees* introduced in [LW20] (see Section 2 for algebraic trees as topological spaces and for equivalence classes of algebraic measure trees). An algebraic measure tree (T, c, μ) consists of an algebraic tree (T, c) satisfying a separability condition, together with a probability measure μ on it (see Definition 2.2). For a notion of convergence in $\mathbb{T}$, we first introduce the *branch point distribution* on T,

$$\nu_{(T,c,\mu)} := \mu^{\otimes 3} \circ c^{-1}, \tag{1.4}$$

and then associate an algebraic measure tree $\chi = (T, c, \mu) \in \mathbb{T}$ with the metric measure tree $(T, r_\mu, \mu) \in \mathbb{M}$. To this end, define the pseudometric

$$r_\mu(x, y) := \nu_\chi\big([x, y]\big) - \tfrac{1}{2}\nu_\chi\big(\{x\}\big) - \tfrac{1}{2}\nu_\chi\big(\{y\}\big), \tag{1.5}$$

where $x, y \in T$, and $[x, y]$ is the interval ("line segment") from x to y. We define convergence of the algebraic measure trees in $\mathbb{T}$ as Gromov-weak convergence of these associated metric measure trees, i.e., we say

$$(T_N, c_N, \mu_N)_{N\in\mathbb{N}} \text{ converges in } \mathbb{T} \quad \text{iff} \quad (T_N, r_{\mu_N}, \mu_N)_{N\in\mathbb{N}} \text{ converges in } \mathbb{M}. \tag{1.6}$$

The space $\mathbb{T}$ equipped with the so-called *branch point distribution distance* Gromov-weak topology (or, for short, bpdd-Gromov-weak topology) is introduced and further studied in [LW20]. Because cladograms are by definition binary, it is for the purpose of the present paper enough to consider the subspace of $\mathbb{T}$ consisting of binary trees. More precisely, we consider the subspaces

$$\mathbb{T}_2 := \big\{(T, c, \mu) \in \mathbb{T} : \text{degrees at most 3, atoms of } \mu \text{ only at leaves}\big\} \tag{1.7}$$

of binary trees with no atoms on the skeleton, and

$$\mathbb{T}_2^{\mathrm{cont}} := \big\{(T, c, \mu) \in \mathbb{T}_2 : \mu \text{ non-atomic}\big\} \tag{1.8}$$

of continuum binary trees. It is shown in [LW20, Theorem 3] that both $\mathbb{T}_2$ and $\mathbb{T}_2^{\mathrm{cont}}$ are compact, which is very convenient for showing tightness of the approximating processes. Furthermore, on $\mathbb{T}_2$, we have equivalent formulations of bpdd-Gromov-weak convergence which we can use to prove our limit statements (see Section 2 for more details).

Let $\mathfrak{C}_m$ denotes the set of m-cladograms (see (2.9)). For an algebraic tree (T, c) and $\underline{u} = (u_1, \ldots, u_m) \in T^m$, let $\mathfrak{s}_{(T,c)}(\underline{u})$ denote the m-cladogram generated by the points $u_1, ..., u_m$ in (T, c) (see Definition 2.5 for a precise definition). For $m \in \mathbb{N}$ and $\mathfrak{t} \in \mathfrak{C}_m$, let $\Phi^{m,\mathfrak{t}}$ be the function which sends an algebraic measure tree to the probability that m points sampled independently with μ generate the cladogram $\mathfrak{t}$, i.e.,

$$\Phi^{m,\mathfrak{t}}(T, c, \mu) := \mu^{\otimes m}\big(\mathfrak{s}^{-1}_{(T,c)}(\mathfrak{t})\big), \tag{1.9}$$

where $(T, c, \mu) \in \mathbb{T}_2$. We refer to $\mu^{\otimes m} \circ \mathfrak{s}^{-1}_{(T,c)}$ as m-sample shape distribution, and to functions in the linear span of functions of the form (1.9) as *shape polynomials*. One of the main results of [LW20] is that $\Phi^{m,\mathfrak{t}} \in \mathcal{C}_b(\mathbb{T}_2)$, and moreover, the set of shape polynomials is convergence determining for measures on $\mathbb{T}_2^{\mathrm{cont}}$. Therefore, it is a convenient set of test functions.

We characterize the Aldous diffusion analytically as the unique solution of a martingale problem. We use the following terminology (see Sections 4.3 and 4.4 of [EK86]). Let E be a polish space, $B(E)$ be the set of bounded, measurable, real-valued functions on E.

Definition 1.1 (Martingale problem). Let $A \colon \mathcal{D}(A) \to B(E)$ with $\mathcal{D}(A) \subseteq B(E)$ be a linear operator, and P a probability measure on E.

(i) A *solution of* the $(A, \mathcal{D}(A), P)$-martingale problem is an E-valued, measurable stochastic process $X = (X_t)_{t\geq 0}$ such that P is the law of X_0 and, for all $\Phi \in \mathcal{D}(A)$, the process $M := (M_t)_{t\geq 0}$ given by

$$M_t := \Phi(X_t) - \int_0^t A\Phi(X_s)\,\mathrm{d}s \tag{1.10}$$

is a martingale (w.r.t. the natural, augmented filtration of X).

(ii) The $(A, \mathcal{D}(A), P)$-martingale problem is *well-posed* if there exists a unique (in finite dimensional distribution) solution of it.

If the $(A, \mathcal{D}(A), P)$-martingale problem is well-posed for every probability measure P on E, the solution X is necessarily a Markov process by [EK86, Theorem 4.4.2]. We sometimes call the operator A *pre-generator* of X, because it is the restriction of the full generator to $\mathcal{D}(A)$. As pre-generator of the Aldous diffusion, we introduce the operator Ω_{Ald} acting on functions of the form (1.9) as follows:

$$\Omega_{\text{Ald}}\Phi^{m,\mathfrak{t}}(T,c,\mu) := \int \Omega_m \mathbb{1}_{\mathfrak{t}}\big(\mathfrak{s}_{(T,c)}(\underline{u})\big)\, \mu^{\otimes m}(\mathrm{d}\underline{u}). \tag{1.11}$$

Obviously, Ω_{Ald} can be extended linearly to the set of shape polynomials, i.e. to

$$\mathcal{D}(\Omega_{\text{Ald}}) := \operatorname{span}\big\{ \text{functions } \Phi^{m,\mathfrak{t}} \text{ of the form (1.9) with } m \in \mathbb{N},\ \mathfrak{t} \in \mathfrak{C}_m \big\}, \tag{1.12}$$

where span denotes the linear span of a set of functions. Our first main result is the following:

Theorem 1 (The well-posed martingale problem)**.** *For all probability measures P_0 on $\mathbb{T}_2^{\text{cont}}$, the $(\Omega_{\text{Ald}}, \mathcal{D}(\Omega_{\text{Ald}}), P_0)$-martingale problem is well-posed. Its unique solution is a Feller process with continuous paths in the compact state space $\mathbb{T}_2^{\text{cont}}$. In particular, it is a strong Markov process. Moreover, this solution is ergodic with the algebraic measure Brownian CRT as unique invariant distribution.*

We refer to the process from Theorem 1 as *Aldous diffusion*:

Definition 1.2 (Aldous diffusion on binary algebraic measure trees)**.** The unique solution of the $(\Omega_{\text{Ald}}, \mathcal{D}(\Omega_{\text{Ald}}), P_0)$-martingale problem is called *Aldous diffusion on binary algebraic non-atomic measure trees*, or simply *Aldous diffusion*, started in P_0.

It is important to mention that the Aldous diffusion is dual to the Aldous chain, as for all $m \in \mathbb{N}$ and m-cladograms $\mathfrak{t}$, the Aldous diffusion $X_t = (T_t, c_t, \mu_t)$ started in $X_0 = (T, c, \mu) \in \mathbb{T}_2^{\text{cont}}$ satisfies

$$\mathbb{E}_{(T,c,\mu)}\Big[\mu_t^{\otimes m}\big\{\underline{u} \in T_t^m : \mathfrak{s}_{(T_t,c_t)}(\underline{u}) = \mathfrak{t}\big\}\Big] = \mathbb{E}_{\mathfrak{t}}\Big[\mu^{\otimes m}\big\{\underline{u} \in T^m : \mathfrak{s}_{(T,c)}(\underline{u}) = \mathcal{T}_t\big\}\Big], \tag{1.13}$$

where $(\mathcal{T}_t)_{t\geq 0}$ denotes the Aldous chain on m-cladograms started in $\mathfrak{t}$.

The name *Aldous diffusion* is justified by the following convergence result. Here, we identify the $\mathfrak{C}_N$-valued Aldous chain on N-cladograms with the $\mathbb{T}_2$-valued Markov chain obtained by forgetting the labels of the cladograms and equipping it with the uniform distribution on the leaves.

Theorem 2 (Diffusion approximation)**.** *For each $N \in \mathbb{N}$, let χ_N be an N-cladogram with the uniform distribution on the leaves. Assume that $\chi_N \to \chi \in \mathbb{T}_2^{\text{cont}}$. Then the Aldous chain X^N starting in $X_0^N = \chi_N$ converges weakly in Skorokhod path space w.r.t. the bpdd-Gromov-weak topology to the Aldous diffusion starting in χ.*

Our last result makes a connection to Aldous's original calculation (1.2) of the evolution of the relative sizes of the three subtrees attached to a fixed branch point until that branch point disappears. Instead of fixing a branch point in the beginning, we take the average over branch points w.r.t. the branch point distribution (1.4). Our topology on $\mathbb{T}_2$ turns out to be strong enough for us to use the diffusion approximation from Theorem 2 to extend the martingale probelm for the Aldous diffusion to the corresponding test functions. Thus we can do explicit calculations which show the missing term compensating for the disappearance of branch points.

To state the result, we need some notation. For a branch point $v \in \operatorname{br}(T)$, consider the three subtrees (components) attached to v, and denote by $\mathcal{S}_v(u)$ the one containing $u \in T$ with $u \neq v$ (see (2.2) below for a precise definition). For $\underline{u} = (u_1, u_2, u_3) \in T^3$, let

$$\underline{\eta}(\underline{u}) := \big(\eta_i(\underline{u})\big)_{i=1,2,3} := \big(\mu\big(\mathcal{S}_{c(\underline{u})}(u_i)\big)\big)_{i=1,2,3} \tag{1.14}$$

be the vector of the three masses of the components connected to the branch point $c(\underline{u})$ of $\underline{u}$. We consider test functions of the following form, called *mass polynomials of degree* 3: For $f\colon [0,1]^3 \to \mathbb{R}$ continuous define

$$\Phi^f(T,c,\mu) := \int f\big(\underline{\eta}(\underline{u})\big)\,\mu^{\otimes 3}(\mathrm{d}\underline{u}), \tag{1.15}$$

where $(T,c,\mu) \in \mathbb{T}_2$. One of the main results of [LW20] is that $\Phi^f \in \mathcal{C}(\mathbb{T}_2)$.

We can extend the domain of the pre-generator Ω_{Ald} to the set of those mass polynomials Φ^f of degree 3 with $f\colon [0,1]^3 \to \mathbb{R}$ twice continuously differentiable. To this end, consider the migration operators $\Theta_{i,j}\colon \mathcal{C}^2([0,1]^3) \to \mathcal{C}^1([0,1]^3)$, $i,j \in \{1,2,3\}$, $i \neq j$

$$\Theta_{1,2} f(\underline{x}) := \frac{\mathbb{1}_{x_1>0}}{x_1}\big(f(0, x_2+x_1, x_3) - f(\underline{x})\big) + \mathbb{1}_{x_1=0}\big(\partial_2 f(\underline{x}) - \partial_1 f(\underline{x})\big), \tag{1.16}$$

and $\Theta_{i,j} f$ defined analogously with the indices 1 and 2 replaced by i and j, respectively. Let $e_i = (\delta_{ij})_{j=1,2,3}$ be the i^{th} unit vector and define

$$\begin{aligned} \Omega_{\mathrm{Ald}}\Phi^f(T,c,\mu) := \int_{T^3} \mathrm{d}\mu^{\otimes 3}\,\Big(& 2\sum_{i,j=1}^{3} \eta_i(\delta_{ij} - \eta_j)\partial^2_{ij} f(\underline{\eta}) + 3\sum_{i=1}^{3}(1-3\eta_i)\partial_i f(\underline{\eta}) \\ & + \tfrac{1}{2}\sum_{i,j=1,\, i\neq j}^{3} \Theta_{i,j} f(\underline{\eta}) + \sum_{i=1}^{3}\big(f(e_i) - f(\underline{\eta})\big)\Big). \end{aligned} \tag{1.17}$$

Theorem 3 (Extended martingale problem for subtree masses)**.** *Let $X = (X_t)_{t\geq 0}$ be the Aldous diffusion on $\mathbb{T}_2^{\mathrm{cont}}$. Then for all test functions Φ^f of the form* (1.15) *with $f \in \mathcal{C}^2\big([0,1]^3\big)$, the process $M^f := (M^f_t)_{t\geq 0}$ given by*

$$M^f_t := \Phi^f(X_t) - \Phi^f(X_0) - \int_0^t \Omega_{\mathrm{Ald}}\Phi^f(X_s)\,\mathrm{d}s \tag{1.18}$$

is a martingale.

Related work. We note that a construction related to the Aldous diffusion has been recently established independently in a sequence of papers [FPRWb, FPRWd, FPRWc, FPRWa]. A discussion of the differences is therefore in order. Their construction was first sketched in [Pal]. Pal suggests to first take a finite number of branch points, consider the cladogram spanned by them, and decompose the lines connecting any two neighboring branch points of this cladogram into subtrees. Then study the suitably rescaled subtree masses as partitions of an interval of random length while relaxing the constrain that the total number of vertices must be preserved by letting removing and inserting of edges happen independently. When applying the time change which reverses the described Poissonization, on the proposed time scale the masses converge to an evolving interval partition described by a family of diffusions indexed by $\mathbb{N}$. However, if this chain runs, then the mass branching off one of the external edges of the cladogram gets exhausted. When this happens, the dynamics breaks down, and one needs to find a slightly different set of branch points to proceed. To resolve the problem of disappearing vertices [FPRWd] suggests a smart way of swapping labels of the cladograms in such a way that the resulting dynamics preserves stationarity when one starts from the uniform distribution.

Our construction is related in spirit but differs in some important aspects. First, rather than sampling cladograms and describing their dynamics under the Aldous chain, we describe the behavior of the *average* of the quantities of interest over uniformly sampled cladograms. This allows us to give a nice characterization of the Aldous diffusion as a unique solution of some martingale problem. As a consequence, we do not require the initial distribution for the Aldous diffusion to be uniform but can rather let it start in any deterministic continuum tree. We can show that the Aldous chain converges weakly in path space to the Aldous diffusion, and that the latter is a Feller process. Note that [FPRWa] never states explicitly that the $\mathbb{R}$-tree-valued diffusion constructed is a strong Markov process. Furthermore, we are also able to state a duality relation, which allows us to conclude convergence to the uniform distribution for all starting points as time tends to infinity. In [LW20], we put some effort in establishing with the space of algebraic measure trees a new state space and invested in a detailed study of topological aspects. As a result, we obtained equivalent formulations of our notion of convergence on the subspace of binary trees which made martingale convergence statements very much straightforward. Finally, the framework provided is not restricted to the construction of the continuum limit of the Aldous Markov chain. It can also be applied to other (not necessarily symmetric) sampling consistent tree dynamics. For example, in [Nus] a tree-valued dynamics is constructed which has the algebraic measure Kingman tree as its stationary distribution.

Other approaches of encoding relatives of binary algebraic trees can be found in [For] and [EGW17]. The *Rémy chain* considered in [EGW17] is a Markov chain of growing (ordered) trees that is somewhat related to the Aldous chain: it is the process obtained by successively inserting new leaves at randomly chosen edges without removing a leaf before.

Outline. The rest of the paper is organized as follows. In Section 2, we introduce our state space of algebraic measure trees and recall its most important properties from [LW20].

In Section 3, we show tightness of the Aldous chains and existence of solutions of the martingale problem from Theorem 1. We do so by using and proving uniform convergence of (pre-)generators. In Section 4, we obtain the duality for the Aldous diffusion (Proposition 4.1), and use it to show uniqueness of solutions of the martingale problem. In Section 5, we show that the Aldous diffusion has a unique invariant measure, namely the algebraic measure Brownian CRT, and that the Aldous diffusion converges to it in law as time goes to infinity (Proposition 5.3). We also finish the proof of Theorems 1 and 2. In Section 6, we prove Theorem 3 and apply it to calculate the annealed average distance of two points in the Brownian CRT with respect to our intrinsic metric.

2. The state space of binary, algebraic measure trees

In this section we introduce the state space. The goal is to overcome the metric issue raised in the introduction by focusing on the algebraic tree structure only. We encode the cladograms as binary, algebraic trees, and use the space of these trees together with the bpdd-Gromov-weak topology studied in [LW20]. All proofs can be found there.

Definition 2.1 (Algebraic tree)**.** An *algebraic tree* is a non-empty set T together with a symmetric map $c\colon T^3 \to T$ satisfying the following:

- **(2pc)** For all $x_1, x_2 \in T$, $c(x_1, x_2, x_2) = x_2$.
- **(3pc)** For all $x_1, x_2, x_3 \in T$, $c\big(x_1, x_2, c(x_1, x_2, x_3)\big) = c(x_1, x_2, x_3)$.
- **(4pc)** For all $x_1, x_2, x_3, x_4 \in T$,

$$c(x_1, x_2, x_3) \in \big\{c(x_1, x_2, x_4),\ c(x_1, x_3, x_4),\ c(x_2, x_3, x_4)\big\}. \tag{2.1}$$

We refer to the map c as *branch point map.* A *tree isomorphism* between two algebraic trees (T_i, c_i), $i = 1, 2$, is a bijective map $\phi\colon T_1 \to T_2$ with $\phi\big(c_1(x_1, x_2, x_3)\big) = c_2\big(\phi(x_1), \phi(x_2), \phi(x_3)\big)$ for all $x_1, x_2, x_3 \in T_1$.

For each point $x \in T$, we define an equivalence relation $\sim_x$ on $T \setminus \{x\}$ such that for all $y, z \in T \setminus \{x\}$, $y \sim_x z$ iff $c(x, y, z) \neq x$. For $y \in T \setminus \{x\}$, we denote by

$$\mathcal{S}_x(y) := \big\{z \in T \setminus \{x\} : z \sim_x y\big\} \tag{2.2}$$

the equivalence class w.r.t. $x \in T$ which contains y. We also call $\mathcal{S}_x(y)$ the *component* of $T \setminus \{x\}$ containing y. An algebraic tree (T, c) allows for all kinds of notions which capture the tree structure, e.g.,

- we say that $S \subseteq T$ is a *subtree* of T iff $c(S^3) = S$,

- we call the number of components of $T \setminus \{x\}$ the *degree* of $x \in T$ and write $\deg(x) = \#\{\mathcal{S}_x(y) : y \in T \setminus \{x\}\}$,
- we say that $u \in T$ is a *leaf* iff $\deg(u) = 1$, and write $\mathrm{lf}(T)$ for the set of leaves,
- we say that $v \in T$ is a *branch point* iff $\deg(v) \geq 3$, or equivalently, $v = c(x_1, x_2, x_3)$ for some $x_1, x_2, x_3 \in T \setminus \{v\}$, and write $\mathrm{br}(T)$ for the set of branch points,
- we write $[x, y]$, $x, y \in T$, for the *interval*

$$[x, y] := \{z \in T : c(x, y, z) = z\}, \tag{2.3}$$

- and we say that $\{x, y\}$ is an edge iff $x \neq y$ and $[x, y] = \{x, y\}$.

There is a natural topology on a given algebraic tree, namely the *component topology* generated by the set of all components $\mathcal{S}_x(y)$ as defined in (2.2) with $x \neq y$, $x, y \in T$. In what follows we refer to an algebraic tree (T, c) as *order separable* if it is separable w.r.t. this topology and has at most countably many edges. We further equip order separable algebraic trees with a probability measure on the Borel σ-algebra $\mathcal{B}(T)$ of the component topology. This so-called *sampling measure* allows to sample vertices from the tree.

Definition 2.2 (Algebraic measure trees)**.** A (separable) *algebraic measure tree* (T, c, μ) is an order separable algebraic tree (T, c) together with a probability measure μ on $\mathcal{B}(T)$.

In what follows we call two algebraic measure trees (T_i, c_i, μ_i), $i = 1, 2$, *equivalent* if there exist subtrees $S_i \subseteq T_i$ with $\mu_i(S_i) = 1$, $i = 1, 2$, and a measure preserving tree isomorphism ϕ from S_1 onto S_2, i.e., $c_2(\phi(x), \phi(y), \phi(z)) = \phi(c_1(x, y, z))$ for all $x, y, z \in S_1$, and $\mu_1 \circ \phi^{-1} = \mu_2$. We define

$$\mathbb{T} := \text{set of equivalence classes of algebraic measure trees.} \tag{2.4}$$

With a slight abuse of notation, we will write $\chi = (T, c, \mu)$ for the algebraic tree as well as the equivalence class. Note that $\deg(x)$, $\mathrm{lf}(T)$, $\mathrm{edge}(T), \ldots$ are properties of the particular representative and not preserved under equivalence, because we do not require the whole trees to be isomorphic. For instance, every equivalence class contains a representative without edges (informally, we can replace edges by line segments carrying no measure).

Recall the branch point distribution $\nu = \nu_\chi = \mu^{\otimes 3} \circ c^{-1}$ and the pseudometric r_μ from Equations (1.4) and (1.5), respectively. For every equivalence class of algebraic measure trees, a representative (T, c, μ) can be chosen such that r_μ is a metric (by identifying points of distance zero in any representative). One can check that in this case, r_μ induces the component topology, and (T, r_μ) is a separable metric tree in the sense of [ALW17] (i.e., isometric to a subset of an $\mathbb{R}$-tree containing all branch points) satisfying for all $x, y, z \in T$,

$$[x, y]_{r_\mu} \cap [x, z]_{r_\mu} \cap [y, z]_{r_\mu} = \{c(x, y, z)\}, \tag{2.5}$$

where $[x, y]_{r_\mu} = \{v \in T : r_\mu(x, y) = r_\mu(x, v) + r_\mu(v, y)\}$ denotes the interval in (T, r_μ). In particular, $[x, y]_{r_\mu} = [x, y]$. Note that any point which carries positive mass is an isolated point in the metric space (T, r_μ).

As in any metric tree, we can define for a fixed reference point (root) $\rho \in T$ a unique measure $\ell^{(T,c,\mu,\rho)}$ on $(T, \mathcal{B}(T))$ which is characterized by the two properties $\ell^{(T,c,\mu,\rho)}((\rho, y]) := r_\mu(\rho, y)$ and $\ell^{(T,c,\mu,\rho)}\big(\mathrm{lf}(T) \setminus \mathrm{at}(\mu)\big) = 0$, where $\mathrm{at}(\mu)$ denotes the set of atoms of μ. The measure $\ell^{(T,c,\mu,\rho)}$ is referred to as *length measure* w.r.t. ρ. Note that it depends on the choice of the distinguished point ρ. However, the total mass of the length measure does not depend on the choice of ρ and equals

$$\|\ell^{(T,c,\mu,\rho)}\| := \ell^{(T,c,\mu,\rho)}(T) = \tfrac{1}{2} \int_T \deg(v)\, \nu(\mathrm{d}v). \tag{2.6}$$

We define convergence in $\mathbb{T}$ as follows.

Definition 2.3 (Bpdd-Gromov-weak topology)**.** We say that a sequence $(\chi_n)_{n\in\mathbb{N}}$ of (equivalence classes of) algebraic measure trees $\chi_n = (T_n, c_n, \mu_n) \in \mathbb{T}$ converges *branch point distribution distance Gromov-weakly* (*bpdd-Gromov-weakly*) to the algebraic measure tree $(T, c, \mu) \in \mathbb{T}$ iff the sequence $(\tilde{\chi}_n)_{n\in\mathbb{N}}$ of (equivalence classes of) metric measure trees $\tilde{\chi}_n := (T_n, r_{\mu_n}, \mu_n) \in \mathbb{M}$ converges to the metric measure tree $(T, r_\mu, \mu) \in \mathbb{M}$ Gromov-weakly, i.e., if for $U_1^n, U_2^n, ...$ independent and μ_n-distributed, and $U_1, U_2, ...$ independent and μ-distributed, for all $m \in \mathbb{N}$,

$$\big(r_{\mu_n}(U_i^n, U_j^n)\big)_{1\le i<j\le m} \underset{n\to\infty}{\Longrightarrow} \big(r_\mu(U_i, U_j)\big)_{1\le i<j\le m}. \tag{2.7}$$

In this paper we are only considering binary algebraic measure trees with the property that the measure has atoms only (if at all) on the leaves of the tree, i.e. the subspace of $\mathbb{T}$ given by

$$\mathbb{T}_2 = \big\{(T, c, \mu) \in \mathbb{T} : \deg(v) \le 3\,\forall v \in T,\ \mathrm{at}(\mu) \subseteq \mathrm{lf}(T)\big\}, \tag{2.8}$$

(compare (1.7)). Even though the equivalence class $\chi \in \mathbb{T}_2$ contains algebraic measure trees which are not binary, we will implicitly assume that the chosen representative (T, c, μ) satisfies $\deg(v) \le 3$. In this subspace, it turns out that bpdd-Gromov-weak convergence is equivalent to another very useful notion of convergence, namely the so-called sample shape convergence, which we introduce next.

Definition 2.4 (m-labelled cladogram)**.** For $m \in \mathbb{N}$, an *m-labelled cladogram* is a binary, finite tree (C, c) consisting only of leaves and branch points, together with a surjective labelling map $\zeta\colon \{1, ..., m\} \to \mathrm{lf}(C)$.

Note that an m-labelled cladogram has at most m leaves (and $m-2$ branch points), but can have less if a leaf has multiple labels. An m-labelled cladogram (C, c, ζ) is an m-cladogram if and only if ζ is injective. We call two m-labelled cladograms (C_1, c_1, ζ_1) and (C_2, c_2, ζ_2) *isomorphic* if there exists a tree isomorphism ϕ from (C_1, c_1) onto (C_2, c_2) such that $\zeta_2 = \phi \circ \zeta_1$. Furthermore, we denote the sets of isomorphism classes of m-labelled cladogram and m-cladograms by $\overline{\mathfrak{C}}_m$ and $\mathfrak{C}_m$, respectively, i.e.,

$$\overline{\mathfrak{C}}_m := \big\{\text{isomorphism classes of } m\text{-labelled cladograms}\big\} \tag{2.9}$$

and

$$\mathfrak{C}_m := \big\{(C, c, \zeta) \in \overline{\mathfrak{C}}_m : \zeta \text{ injective}\big\}. \tag{2.10}$$

Definition 2.5 (The shape function)**.** For a binary algebraic tree (T, c), $m \in \mathbb{N}$, and $u_1, ..., u_m \in T\backslash\mathrm{br}(T)$ (not necessarily distinct), there exists a unique (up to isomorphism) m-labelled cladogram

$$\mathfrak{s}_{(T,c)}(u_1, ..., u_m) = (C, c_C, \zeta) \tag{2.11}$$

with $\mathrm{lf}(C) = \{u_1, ..., u_m\}$ and $\zeta(i) = u_i$, such that the identity on $\mathrm{lf}(C)$ extends to a tree homomorphism π from C onto $c\big(\{u_1, ..., u_m\}^3\big)$, i.e., for all $i, j, k = 1, ..., m$,

$$\pi\big(c_C(u_i, u_j, u_k)\big) = c(u_i, u_j, u_k). \tag{2.12}$$

We refer to $\mathfrak{s}_{(T,c)}(u_1, ..., u_m) \in \overline{\mathfrak{C}}_m$ as the *shape* of $u_1, ..., u_m$ in (T, c).

Definition 2.6 (Sample shape convergence)**.** We say that a sequence $(\chi_n)_{n\in\mathbb{N}}$ of (equivalence classes of) binary algebraic measure trees (T_n, c_n, μ_n) *converges in sample shape* to the (equivalence class of the) algebraic measure tree (T, c, μ) iff for $U_1^n, U_2^n, ...$ independent and μ_n-distributed, and $U_1, U_2, ...$ independent and μ-distributed, for all $m \in \mathbb{N}$,

$$\mathfrak{s}_{(T,c)}\big(U_1^n, ..., U_m^n\big) \underset{n\to\infty}{\Longrightarrow} \mathfrak{s}_{(T,c)}\big(U_1, ..., U_m\big). \tag{2.13}$$

To be later in a position to recover the calculations of Aldous and others concerning the dynamics of subtree masses, we introduce yet another notion of convergence.

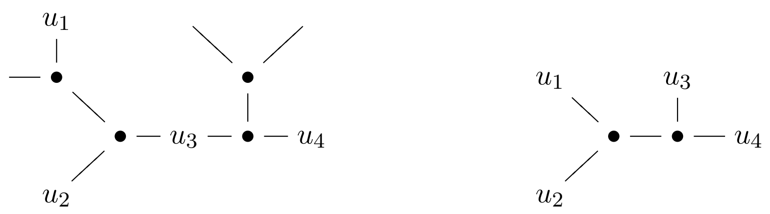


Figure 3. A tree T and the shape $\mathfrak{s}_{(T,c)}(u_1,u_2,u_3,u_4)$. The cladogram is not isomorphic to the subtree $c(\{u_1,u_2,u_3,u_4\}^3)$ because $u_3 \in\,]u_1,u_4[$.

Definition 2.7 (Sample subtree mass convergence)**.** We say that a sequence $(\chi_n)_{n\in\mathbb{N}}$ of (equivalence classes of) algebraic measure trees (T_n,c_n,μ_n) *converges in sample subtree mass* to the (equivalence class of the) algebraic measure tree (T,c,μ) iff for $U_1^n,U_2^n,...$ independent and μ_n-distributed, and $U_1,U_2,...$ independent and μ-distributed, for all $m\in\mathbb{N}$,

$$\big(\mu_n(\mathcal{S}_{c_n(U_i^n,U_j^n,U_k^n)}(U_i^n))\big)_{i,j,k=1,...,m} \underset{n\to\infty}{\Longrightarrow} \big(\mu(\mathcal{S}_{c(U_i,U_j,U_k)}(U_i))\big)_{i,j,k=1,...,m}. \tag{2.14}$$

The following results are crucial for the construction of Aldous diffusion and stated in [LW20, Proposition 2.32, Theorem 3, Corollary 5.21]. On $\mathbb{T}_2$, all of the above notions of convergence are equivalent. By (2.6), the total length of binary algebraic measure trees is uniformly bounded by 3, and one can show that the space $\mathbb{T}_2$ is compact.

Proposition 2.8. *Let $(\chi_N = (T_N,c_N,\mu_N))_{N\in\mathbb{N}}$ and $\chi=(T,c,\mu)$ be in $\mathbb{T}_2$. The following are equivalent:*

(1) $\chi_N \underset{n\to\infty}{\longrightarrow} \chi$ w.r.t. sample shape convergence.

(2) For all $m\in\mathbb{N}$ and $\mathfrak{t}\in\overline{\mathfrak{C}}_m$,

$$\mu_N^{\otimes m}\big\{(u_1,...,u_m):\mathfrak{s}_{(T_N,c_N)}\big(\underline{u}\big)=\mathfrak{t}\big\} \underset{n\to\infty}{\longrightarrow} \mu^{\otimes m}\big\{(u_1,...,u_m):\mathfrak{s}_{(T,c)}\big(\underline{u}\big)=\mathfrak{t}\big\}. \tag{2.15}$$

(3) $\chi_N \underset{n\to\infty}{\longrightarrow} \chi$ Gromov-weakly w.r.t. the branch point distribution distance.

(4) For all $m\in\mathbb{N}$ and $\phi\in\mathcal{C}_b(\mathbb{R}_+^{m\times m})$,

$$\int \mu_N^{\otimes m}(\mathrm{d}\underline{u})\,\phi\big((r_{\mu_N}(u_i,u_j))_{1\le i,j\le m}\big) \underset{n\to\infty}{\longrightarrow} \int \mu^{\otimes m}(\mathrm{d}\underline{u})\,\phi\big((r_{\mu}(u_i,u_j))_{1\le i,j\le m}\big). \tag{2.16}$$

(5) $\chi_N \underset{n\to\infty}{\longrightarrow} \chi$ w.r.t. sample subtree mass convergence.

(6) For all $m\in\mathbb{N}$ with $m\ge 3$ and $f\in\mathcal{C}_b([0,1]^{m^3})$,

$$\begin{aligned}&\int \mu_N^{\otimes m}(\mathrm{d}\underline{u})\, f\Big(\big(\mu_N(\mathcal{S}_{c_N(u_i,u_j,u_k)}(u_i))\big)_{i,j,k=1,...,m}\Big)\\ &\qquad \underset{n\to\infty}{\longrightarrow} \int \mu^{\otimes m}(\mathrm{d}\underline{u})\, f\Big(\big(\mu(\mathcal{S}_{c(u_i,u_j,u_k)}(u_i))\big)_{i,j,k=1,...,m}\Big)\end{aligned} \tag{2.17}$$

In what follows, we will need the following two subspaces of $\mathbb{T}_2$. Let for each $N\in\mathbb{N}$,

$$\mathbb{T}_2^N := \Big\{(T,c,\mu)\in\mathbb{T}_2 : \#\,\mathrm{lf}(T)=N \text{ and } \mu=\tfrac{1}{N}\sum_{u\in\mathrm{lf}(T)}\delta_u\Big\}, \tag{2.18}$$

and

$$\mathbb{T}_2^{\mathrm{cont}} := \big\{(T,c,\mu)\in\mathbb{T}_2 : \mathrm{at}(\mu)=\emptyset\big\}. \tag{2.19}$$

The Aldous chain on N-cladograms is naturally defined on $\mathbb{T}_2^N$: for $\chi\in\mathbb{T}_2^N$, there is a unique (up to measure preserving tree isomorphism) minimal representative (T,c,μ) of χ (i.e. no subset with the restrictions of c and μ is an algebraic measure tree) with $2N-2$ vertices and $2N-3$ edges. We identify $\chi\in\mathbb{T}_2^N$ with this minimal representative and interpret it as "N-caldogram without labels" with uniform distribution on the leaves. We define the Aldous chain on $\mathbb{T}_2^N$ in the same way as the one on $\mathfrak{C}_N$, via its generator Ω_N in (1.1). With a slight abuse of notation, we use the same notation for the generators of the $\mathbb{T}_2^N$-valued and of the $\mathfrak{C}_N$-valued chain. We will define the Aldous diffusion on the space $\mathbb{T}_2^{\mathrm{cont}}$ in view of the following approximation result:

Proposition 2.9 (Approximations with $\mathbb{T}_2^N$)**.** *Let $\chi \in \mathbb{T}_2$. Then $\chi \in \mathbb{T}_2^{\mathrm{cont}}$ if and only if there exists for each $N \in \mathbb{N}$ an $\chi_N \in \mathbb{T}_2^N$ such that $\chi_N \to \chi$ in one (and thus all) of the equivalent notions of convergence on $\mathbb{T}_2$ given above.*

Proposition 2.10 (Compactness and metrizability)**.** *$\mathbb{T}_2$ is a compact, metrizable space. Both $\mathbb{T}_2^N$ and $\mathbb{T}_2^{\mathrm{cont}}$ are closed subspaces of $\mathbb{T}_2$, and thus compact as well.*

To deal with the Aldous chain and diffusions, it is convenient to introduce the following set of test functions on $\mathbb{T}_2$.

Definition 2.11 (Shape polynomials)**.** A *shape polynomial* is a linear combination of functions $\Phi^{m,\mathfrak{t}}\colon \mathbb{T}_2 \to \mathbb{R}$ of the form

$$\Phi^{m,\mathfrak{t}}(\chi) := \mu^{\otimes m}\big(\mathfrak{s}_{(T,c)}^{-1}(\mathfrak{t})\big), \tag{2.20}$$

where $\chi = (T, c, \mu)$, $m \in \mathbb{N}$ and $\mathfrak{t} \in \overline{\mathfrak{C}}_m$. Let $\Pi_{\mathfrak{s}}$ be the set of all shape polynomials.

Apart from its combinatorial nature, the usfulness of shape polynomials stems from the fact that every real-valued continuous function on $\mathbb{T}_2$ can be approximated by them.

Lemma 2.12. *The set $\Pi_{\mathfrak{s}}$ of shape polynomials is a uniformly dense sub-algebra of $\mathcal{C}(\mathbb{T}_2)$.*

Proof. It is immediate from Propositions 2.8 and 2.10 that $\Pi_{\mathfrak{s}}$ is contained in the space $\mathcal{C}(\mathbb{T}_2)$ of continuous functions and separates the points of $\mathbb{T}_2$. To see that $\Pi_{\mathfrak{s}}$ is multiplicatively closed, consider for $k, m \in \mathbb{N}$, $k < m$ the projection $\pi_{m,k}\colon \overline{\mathfrak{C}}_m \to \overline{\mathfrak{C}}_k$, mapping $\mathfrak{t} = (C, c, \zeta) \in \overline{\mathfrak{C}}_m$ to the subcladogram spanned by $\zeta(1), \dots, \zeta(k)$, as well as the projection $\tilde{\pi}_{m,k}\colon \overline{\mathfrak{C}}_m \to \overline{\mathfrak{C}}_{m-k}$, mapping to the subcladogram spanned by $\zeta(k+1), \dots, \zeta(m)$ (relabelled to have labels in $\{1, \dots, m-k\}$). Because μ is a probability and the product measure is used in (2.20), we have for $m, n \in \mathbb{N}$, $\mathfrak{t} \in \overline{\mathfrak{C}}_m$, $\tilde{\mathfrak{t}} \in \overline{\mathfrak{C}}_n$

$$\Phi^{m,\mathfrak{t}} \cdot \Phi^{n,\tilde{\mathfrak{t}}} = \sum_{\mathfrak{t}' \in \pi_{n+m,m}^{-1}(\mathfrak{t}) \cap \tilde{\pi}_{n+m,n}^{-1}(\tilde{\mathfrak{t}})} \Phi^{m+n,\mathfrak{t}'} \in \Pi_{\mathfrak{s}}. \tag{2.21}$$

Because $\mathbb{T}_2$ is compact by Proposition 2.10, $\Pi_{\mathfrak{s}}$ is dense in $\mathcal{C}(\mathbb{T}_2)$ by the Stone-Weierstrass theorem. $\square$

Consider $m \in \mathbb{N}$ and $\mathfrak{t} = (C, c, \zeta) \in \overline{\mathfrak{C}}_m \setminus \mathfrak{C}_m$, i.e. there is at least one leaf in the m-labelled cladogram $\mathfrak{t}$ with multiple labels. Then $\mathfrak{s}_{(T,c)}(u_1, \dots, u_m) = \mathfrak{t}$ implies that the $u_1, \dots, u_m$ are not distinct. Hence $\Phi^{m,\mathfrak{t}}(\chi) = 0$ for all $\chi \in \mathbb{T}_2^{\mathrm{cont}}$. This is in fact the reason why we restricted the domain of the pre-generator of the Aldous diffusion $\mathcal{D}(\Omega_{\mathrm{Ald}})$ to shape polynomials using m-cladograms instead of m-labelled cladograms (see (1.12)). Note that the set of restrictions to $\mathbb{T}_2^{\mathrm{cont}}$ of functions in $\mathcal{D}(\Omega_{\mathrm{Ald}})$ is dense in $\mathcal{C}(\mathbb{T}_2^{\mathrm{cont}})$.

3. Convergence of generators, tightness and existence

In this section we prepare the proofs of our main results by showing the uniform convergence of the generators of the discrete chains to the pre-generator $(\mathcal{D}(\Omega_{\mathrm{Ald}}), \Omega_{\mathrm{Ald}})$, and deduce tightness of the Aldous chains (provided tightness of initial conditions) as well as existence of solutions of the limiting martingale problem by general theory. We also obtain continuous paths of all limit processes.

A first simple observation about the pre-generator is that it maps $\mathcal{D}(\Omega_{\mathrm{Ald}})$ into itself.

Lemma 3.1. *For every $\Phi \in \mathcal{D}(\Omega_{\mathrm{Ald}})$, we have $\Omega_{\mathrm{Ald}}\Phi \in \mathcal{D}(\Omega_{\mathrm{Ald}})$. In particular, $(\Phi, \Omega_{\mathrm{Ald}}\Phi) \in \mathcal{C}(\mathbb{T}_2) \times \mathcal{C}(\mathbb{T}_2)$.*

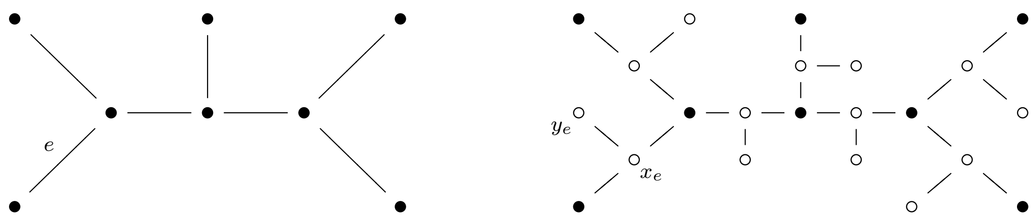


FIGURE 4. A finite algebraic tree (T, c) and the extended tree $(\overline{T}, \bar{c})$.

Proof. Both Φ and $\Omega_{\mathrm{Ald}}\Phi$ are shape polynomials, hence continuous by definition of sample shape convergence. □

For $N \in \mathbb{N}$, recall from (1.1) the generator Ω_N of the Aldous chain on N-cladograms and from (2.18) the space $\mathbb{T}_2^N$ of algebraic measure trees with N leaves and uniform distribution on the leaves.

Proposition 3.2 (Convergence of generators)**.** *For all $\Phi \in \mathcal{D}(\Omega_{\mathrm{Ald}})$, we have*

$$\lim_{N\to\infty} \sup_{\chi \in \mathbb{T}_2^N} \left|\Omega_N \Phi(\chi) - \Omega_{\mathrm{Ald}}\Phi(\chi)\right| = 0. \tag{3.1}$$

Proof. Consider $\Phi \in \mathcal{D}(\Omega_{\mathrm{Ald}})$. By linearity, we may assume w.l.o.g. that $\Phi = \Phi^{m,\mathfrak{t}}$ for some $m \in \mathbb{N}$ and $\mathfrak{t} \in \mathfrak{C}_m$. In particular, $\mathfrak{t}$ is such that no leaf has multiple labels, and consequently for $\underline{u} \in T^m$, $\mathfrak{s}_{(T,c)}(\underline{u}) = \mathfrak{t}$ implies that $u_1, ..., u_m$ are distinct.

Fix $N \in \mathbb{N}$ and $\chi \in \mathbb{T}_2^N$, and let (T, c, μ) be the unique (up to measure preserving tree isomorphism) minimal representative (i.e. $\#T = 2N - 2$). Then $\#\,\mathrm{lf}(T) = N$ and $\#\,\mathrm{edge}(T) = 2N - 3$). In the following we abbreviate the inverse numbers of leaves and edges respectively by

$$\epsilon = \epsilon_N := \tfrac{1}{N}, \qquad \text{and} \qquad \delta = \delta_N := \tfrac{1}{2N-3}. \tag{3.2}$$

We extend the algebraic tree to allow for potential new branch points (due to inserting an edge) and new leaves. To this end, for every edge $e \in \mathrm{edge}(T)$, we introduce two additional points x_e, y_e. Informally, x_e is inserted in the middle of e, and y_e is attached to x_e as a leaf. More precisely, we consider

$$\overline{T} = T \uplus \biguplus_{e \in \mathrm{edge}(T)} \{x_e, y_e\}, \tag{3.3}$$

and extend c to $\bar{c}\colon \overline{T}^3 \to \overline{T}$ which is uniquely defined by the following. $(\overline{T}, \bar{c})$ is an algebraic tree such that for $e = \{u, v\} \in \mathrm{edge}(T)$, we have $x_e \in [u, v]$ in $(\overline{T}, \bar{c})$, and

$$\bar{c}(y_e, x_e, z) = x_e \quad \forall z \in \overline{T} \setminus \{y_e\}. \tag{3.4}$$

The construction is illustrated in Figure 4. Note that $(\overline{T}, \bar{c}, \mu)$ is a binary algebraic measure tree equivalent to (T, c, μ).

For $k \in \{1, ..., m\}$ and $x \in \overline{T}$, let $\theta_{k,x}\colon T^m \to \overline{T}^m$ be the *replacement operator* which replaces the k^{th}-coordinate by x, i.e.,

$$\theta_{k,x}(u_1, ..., u_m) = (u_1, ..., u_{k-1}, x, u_{k+1}, ..., u_m). \tag{3.5}$$

For $z = (x, e) \in \mathrm{lf}(T) \times \mathrm{edge}(T)$, let

$$\chi_z := (\overline{T}, \bar{c}, \mu + \epsilon\delta_{y_e} - \epsilon\delta_x) \tag{3.6}$$

be the binary algebraic measure tree after the Aldous move with z. The difference between sampling with the new and old measure is given by

$$
\begin{aligned}
&\big(\mu+\epsilon\delta_{y_e}-\epsilon\delta_x\big)^{\otimes m}-\mu^{\otimes m}\\
&\quad=\epsilon\sum_{k=1}^m \mu^{\otimes(k-1)}\otimes\big(\delta_{y_e}-\delta_x\big)\otimes\mu^{\otimes(m-k)}\\
&\qquad+\epsilon^2\sum_{1\le k<j\le m}\mu^{\otimes(k-1)}\otimes\big(\delta_{y_e}-\delta_x\big)\otimes\mu^{\otimes(j-k-1)}\otimes\big(\delta_{y_e}-\delta_x\big)\otimes\mu^{\otimes(m-j)}+\tilde\mu\\
&\quad=\epsilon\sum_{k=1}^m\big(\mu^{\otimes m}\circ\theta_{k,y_e}^{-1}-\mu^{\otimes m}\circ\theta_{k,x}^{-1}\big)-\epsilon^2\sum_{j,k=1,\,j\neq k}^m\mu^{\otimes m}\circ\theta_{k,y_e}^{-1}\circ\theta_{j,x}^{-1}+\tilde\mu',
\end{aligned}
\tag{3.7}
$$

where $\tilde\mu,\tilde\mu'$ are signed measures on $\overline{T}^m$ vanishing on $\{(u_1,\dots,u_m):u_1,\dots,u_m \text{ distinct}\}$. Thus

$$
\Omega_N\Phi^{m,\mathfrak{t}}(\chi)=\sum_{z\in\mathrm{lf}(T)\times\mathrm{edge}(T)}\big(\Phi^{m,\mathfrak{t}}(\chi_z)-\Phi^{m,\mathfrak{t}}(\chi)\big)=\sum_{k=1}^m A_k-\sum_{j,k=1,\,j\neq k}^m B_{k,j},
\tag{3.8}
$$

with

$$
A_k:=\epsilon\sum_{(x,e)\in\mathrm{lf}(T)\times\mathrm{edge}(T)}\int_{T^m}\mu^{\otimes m}(\mathrm{d}\underline{u})\,\Big(\mathbb{1}_{\mathfrak{t}}\big(\mathfrak{s}_{(\overline{T},\bar c)}(\theta_{k,y_e}\underline{u})\big)-\mathbb{1}_{\mathfrak{t}}\big(\mathfrak{s}_{(\overline{T},\bar c)}(\theta_{k,x}\underline{u})\big)\Big),
\tag{3.9}
$$

and

$$
B_{k,j}:=\epsilon^2\sum_{(x,e)\in\mathrm{lf}(T)\times\mathrm{edge}(T)}\int_{T^m}\mu^{\otimes m}(\mathrm{d}\underline{u})\,\mathbb{1}_{\mathfrak{t}}\big(\mathfrak{s}_{(\overline{T},\bar c)}(\theta_{k,y_e}\circ\theta_{j,x}\underline{u})\big).
\tag{3.10}
$$

We use the notation $\mathfrak{t}_{\wedge k}\in\mathfrak{C}_{m-1}$ for the $(m-1)$-cladogram obtained from $\mathfrak{t}$ by deleting the leaf with label k (and relabelling the labels $j>k$ to $j-1$), i.e., if $\mathfrak{t}=\mathfrak{s}_{(\overline{T},\bar c)}(\underline{u})$, then $\mathfrak{t}_{\wedge k}=\mathfrak{s}_{(\overline{T},\bar c)}(\underline{u}_{\wedge k})$ with $\underline{u}_{\wedge k}:=(u_1,\dots,u_{k-1},u_{k+1},\dots,u_m)$. Furthermore, for $\underline{u}\in T^m$, we define

$$
E_{\mathfrak{t},k}(\underline{u}):=\big\{v\in\overline{T}:\mathfrak{s}_{(\overline{T},\bar c)}(\theta_{k,v}\underline{u})=\mathfrak{t}\big\}.
\tag{3.11}
$$

Note that $E_{\mathfrak{t},k}(\underline{u})$ does not depend on u_k, contains no u_j for $j\neq k$, and that $E_{\mathfrak{t},k}(\underline{u})\neq\emptyset$ only if $\mathfrak{s}_{(T,c)}(\underline{u}_{\wedge k})=\mathfrak{t}_{\wedge k}$. In this case, $E_{\mathfrak{t},k}(\underline{u})$ "corresponds to" an edge of $\mathfrak{t}_{\wedge k}$. Let $\ell:=\delta\sum_{e\in\mathrm{edge}(T)}\delta_{y_e}$ be the uniform distribution on $\{y_e:e\in\mathrm{edge}(T)\}$. By Fubini's theorem and using that $\epsilon\delta\sum_{(x,e)\in\mathrm{lf}\times\mathrm{edge}}\delta_x\otimes\delta_{y_e}=\mu\otimes\ell$, we obtain

$$
\begin{aligned}
A_k&=\delta^{-1}\int_{\overline{T}^m}\mu^{\otimes m}(\mathrm{d}\underline{u})\big(\ell(E_{\mathfrak{t},k}(\underline{u}))-\mu(E_{\mathfrak{t},k}(\underline{u}))\big)\\
&=\int_{T^m}\mu^{\otimes m}(\mathrm{d}\underline{u})\mathbb{1}_{\mathfrak{t}_{\wedge k}}\big((\mathfrak{s}_{(\overline{T},\bar c)}\underline{u}_{\wedge k})\big)\cdot\big(3\mu(E_{\mathfrak{t},k}(\underline{u}))+1\big)\\
&=3\Phi^{m,\mathfrak{t}}(\chi)+\Phi^{m-1,\mathfrak{t}_{\wedge k}}(\chi),
\end{aligned}
\tag{3.12}
$$

where we have used in the second step that, because (T,c) is binary,

$$
\#\big\{e\in\mathrm{edge}(T):c(y_e,z,z')\in(z,z')\big\}=2\#\big\{x\in\mathrm{lf}(T):c(x,z,z')\in(z,z')\big\}+1
\tag{3.13}
$$

for every $z,z'\in T$, and hence $\delta^{-1}\ell(E_{\mathfrak{t},k}(\underline{u}))=2N\mu(E_{\mathfrak{t},k}(\underline{u}))+1$ if $\mathfrak{s}_{(\overline{T},\bar c)}(\underline{u}_{\wedge k})=\mathfrak{t}_{\wedge k}$. Similarly,

$$
\begin{aligned}
B_{k,j}&=\tfrac{\epsilon}{\delta}\int_{\overline{T}}\ell(\mathrm{d}y)\int_T\mu(\mathrm{d}x)\int_{T^m}\mu^{\otimes m}(\mathrm{d}\underline{u})\,\mathbb{1}_{\mathfrak{t}}\big(\mathfrak{s}_{(\overline{T},\bar c)}\circ\theta_{k,y}\circ\theta_{j,x}(\underline{u})\big)\\
&=\tfrac{\epsilon}{\delta}\Phi^{m,\mathfrak{t}}(\chi)+\epsilon A_k\\
&=2\Phi^{m,\mathfrak{t}}(\chi)+\epsilon A_k+3\epsilon\Phi^{m,\mathfrak{t}}(\chi).
\end{aligned}
\tag{3.14}
$$

Combining (3.8), (3.12) and (3.14), we obtain that

(3.15)

$$\begin{aligned}
&\Omega_N \Phi^{m,\mathfrak{t}}(\chi) \\
&= \sum_{k=1}^{m} \Phi^{m-1,\mathfrak{t}_{\wedge k}}(\chi) + \big(3m - 2m(m-1)\big)\Phi^{m,\mathfrak{t}}(\chi) - \epsilon(m-1)\sum_{k=1}^{m} A_k - 3\epsilon m(m-1)\Phi^{m,\mathfrak{t}}(\chi) \\
&= \sum_{k=1}^{m} \Phi^{m-1,\mathfrak{t}_{\wedge k}}(\chi) - m(2m-5)\Phi^{m,\mathfrak{t}}(\chi) - \epsilon(m-1)\sum_{k=1}^{m} \Phi^{m-1,\mathfrak{t}_{\wedge k}}(\chi) - 6\epsilon m(m-1)\Phi^{m,\mathfrak{t}}(\chi).
\end{aligned}$$

For an edge e of $\mathfrak{t}_{\wedge k}$, denote by $\mathfrak{t}^{(k,e)}$ the cladogram obtained by inserting a leaf labelled k in $\mathfrak{t}_{\wedge k}$ at the edge e (and relabelling the labels $j \geq k$ to $j+1$). In particular, $\mathfrak{t}^{(k,e)}$ is the cladogram obtained from $\mathfrak{t}$ by the Aldous move (k,e). For $\underline{u} \in T^m$, we have $\mathfrak{s}_{(T,c)}(\underline{u}_{\wedge k}) = \mathfrak{t}_{\wedge k}$ if and only if there is an edge e of $\mathfrak{t}_{\wedge k}$ such that $\mathfrak{s}_{(T,c)}(\underline{u}) = \mathfrak{t}^{(k,e)}$, and this e is unique. Hence,

$$\begin{aligned}
\sum_{k=1}^{m} \Phi^{m-1,\mathfrak{t}_{\wedge k}}(\chi) &= \int_{T^m} \mu^{\otimes m}(\mathrm{d}\underline{u}) \sum_{k=1}^{m} \mathbb{1}_{\mathfrak{t}_{\wedge k}}\big(\mathfrak{s}_{(T,c)}(\underline{u}_{\wedge k})\big) \\
&= \int_{T^m} \mu^{\otimes m}(\mathrm{d}\underline{u}) \sum_{k=1}^{m} \sum_{e\in \mathrm{edge}(\mathfrak{t}_{\wedge k})} \mathbb{1}_{\mathfrak{t}^{(k,e)}}\big(\mathfrak{s}_{(T,c)}(\underline{u})\big) \qquad (3.16)\\
&= \int_{T^m} \mu^{\otimes m}(\mathrm{d}\underline{u})\, \Omega_m \mathbb{1}_{\mathfrak{t}}\big(\mathfrak{s}_{(T,c)}(\underline{u})\big) + m\,\#\,\mathrm{edge}(\mathfrak{t}_{\wedge k})\, \Phi^{m,\mathfrak{t}}(\chi).
\end{aligned}$$

Inserting this into (3.15) and using that $\#\,\mathrm{edge}(\mathfrak{t}_{\wedge k}) = 2m-5$, we see that

$$\Big|\Omega_N \Phi^{m,\mathfrak{t}}(\chi) - \int_{T^m} (\Omega_m \mathbb{1}_{\mathfrak{t}}) \circ \mathfrak{s}_{(T,c)}\, \mathrm{d}\mu^{\otimes m}\Big| \leq 7m(m-1)\epsilon, \tag{3.17}$$

which gives the claim. □

As $\mathbb{T}_2$ is compact by Proposition 2.10, we can immediately conclude the following from the convergence of the generators.

Corollary 3.3 (The limiting martingale problem)**.** *Let $(\chi_N)_{N\in\mathbb{N}}$ be a sequence of random binary algebraic measure trees with $\chi_N \in \mathbb{T}_2^N$, such that*

$$\chi_N \Rightarrow \chi, \quad \text{as } N \to \infty, \tag{3.18}$$

where χ is distributed according to P_0 on $\mathbb{T}_2^{\mathrm{cont}}$. Let $X^N = (X_t^N)_{t\geq 0}$ be the Aldous chain started in χ_N. Then the sequence $(X^N)_{N\in\mathbb{N}}$ is $\mathcal{C}$-tight in Skorokhod path space, and any limit process $(X_t)_{t\geq 0}$ has paths in $\mathbb{T}_2^{\mathrm{cont}}$ and satisfies the $(\Omega_{\mathrm{Ald}}, \mathcal{D}(\Omega_{\mathrm{Ald}}), P_0)$-martingale problem.

Proof. Tightness. Tightness follows, in view of the approximation result Proposition 2.9 and Lemma 3.1, with the exactly same proof as Theorems 3.9.1 and 3.9.4 in [EK86] (see also [EK86, Remark 4.5.2]).

Continuous paths. For $\Phi \in \mathcal{D}(\Omega_{\mathrm{Ald}})$, let $\Phi(X^N) = \big(\Phi(X_t^N)\big)_{t\geq 0}$. By definition, $\mathcal{D}(\Omega_{\mathrm{Ald}})$ induces the topology of sample-shape convergence on $\mathbb{T}_2^{\mathrm{cont}}$. Hence, continuity of the paths of the limit process $X = (X_t)_{t\geq 0}$ in $\mathbb{T}_2^{\mathrm{cont}}$ is equivalent to path-continuity of $\Phi(X)$ for all $\Phi \in \mathcal{D}(\Omega_{\mathrm{Ald}})$. Because $\Phi(X)$ is the limit of $\Phi(X^N)$, this follows from the obvious estimate $\big|\Phi(X_t^N) - \Phi(X_{t-}^N)\big| \leq \frac{m}{N}$ for $\Phi = \Phi^{m,\mathfrak{t}}$.

Paths in $\mathbb{T}_2^{\mathrm{cont}}$. That any limit point has paths in $\mathbb{T}_2^{\mathrm{cont}}$ follows directly from the fact that X^N has paths in $\mathbb{T}_2^N$ and the approximation result Proposition 2.9.

Martingale problem. That all limit points satisfy the martingale problem follows with the same proof as Lemma 4.5.1 in [EK86]. □

The following corollary is immediate from the previous corollary and the approximation result Proposition 2.9.

Corollary 3.4 (Existence)**.** *For any probability measure P_0 on $\mathbb{T}_2^{\mathrm{cont}}$ there exists a solution in $\mathcal{C}_{\mathbb{T}_2^{\mathrm{cont}}}(\mathbb{R}_+)$ of the $(\Omega_{\mathrm{Ald}}, \mathcal{D}(\Omega_{\mathrm{Ald}}), P_0)$-martingale problem.*

4. Duality, uniqueness and convergence

In this section we first obtain a duality result that in turn allows to conclude the uniqueness of the martingale problem. We also use duality to show that the Aldous diffusion is a Feller process on $\mathbb{T}_2^{\mathrm{cont}}$. For $m \in \mathbb{N}$ let $Y^m := (Y_t^m)_{t\geq 0}$ be the $\mathfrak{C}_m$-valued Aldous chain with generator Ω_m from (1.1). If $Y_0^m = \mathfrak{t} \in \mathfrak{C}_m$, then $\mathbb{E}_{\mathfrak{t}}^Y$ denotes the corresponding expectation.

Proposition 4.1 (Duality)**.** *Let P_0 be an arbitrary probability measure on $\mathbb{T}_2^{\mathrm{cont}}$ and let $X := ((T_t, c_t, \mu_t))_{t\geq 0}$ be a solution of the $(\Omega_{\mathrm{Ald}}, \mathcal{D}(\Omega_{\mathrm{Ald}}), P_0)$-martingale problem in $\mathcal{D}_{\mathbb{T}_2^{\mathrm{cont}}}(\mathbb{R}_+)$. For arbitrary $m \in \mathbb{N}$ and $\mathfrak{t} \in \mathfrak{C}_m$, let $Y^m := (Y_t^m)_{t\geq 0}$ be the $\mathfrak{C}_m$-valued Aldous chain with $Y_0^m = \mathfrak{t}$. Assume that Y^m is independent of X. Then*

$$
\begin{aligned}
&\mathbb{E}_{P_0}^X\big[\mu_t^{\otimes m}\big\{\underline{u} \in T_t^m : \mathfrak{s}_{(T_t,c_t)}(\underline{u}) = \mathfrak{t}\big\}\big] \\
&\qquad = \int_{\mathbb{T}_2^{\mathrm{cont}}} \mathbb{E}_{\mathfrak{t}}^Y\big[\mu^{\otimes m}\big\{\underline{u} \in T^m : \mathfrak{s}_{(T,c)}(\underline{u}) = Y_t^m\big\}\big]\, P_0(\mathrm{d}(T,c,\mu)).
\end{aligned} \tag{4.1}
$$

Proof. Let $m \in \mathbb{N}$. For $\chi = (T,c,\mu) \in \mathbb{T}_2^{\mathrm{cont}}$ and $\mathfrak{t} \in \mathfrak{C}_m$, we define $H(\chi,\mathfrak{t}) := \mu^{\otimes m}\big\{\underline{u} \in T^m : \mathfrak{s}_{(T,c)}(\underline{u}) = \mathfrak{t}\big\}$. Then

$$
\Omega_{\mathrm{Ald}} H(\cdot,\mathfrak{t})(\chi) = \int_{T^m} \mu^{\otimes m}(\mathrm{d}\underline{u})\, \Omega_m \mathbb{1}_{\mathfrak{t}}\big(\mathfrak{s}_{(T,c)}(\underline{u})\big) = \Omega_m H(\chi,\cdot)(\mathfrak{t}). \tag{4.2}
$$

By our assumptions on the test functions H and definitions of Ω_{Ald} and Ω_m, the result follows by [EK86, Lemma 4.4.11, Corollary 4.4.13]. □

Corollary 4.2 (Uniqueness of the martingale problem)**.** *Let P_0 be an arbitrary probability measure on $\mathbb{T}_2^{\mathrm{cont}}$. Then uniqueness holds for the $(\Omega_{\mathrm{Ald}}, \mathcal{D}(\Omega_{\mathrm{Ald}}), P_0)$-martingale problem in $\mathcal{D}_{\mathbb{T}_2^{\mathrm{cont}}}(\mathbb{R}_+)$.*

Proof. As the set of all shape polynomials is separating (for probability measures), the result is immediate by the previous proposition and Proposition 4.4.7 from [EK86]. □

Corollary 4.3 (Feller process)**.** *For $F \in \mathcal{C}(\mathbb{T}_2^{\mathrm{cont}})$, $t \geq 0$, and $\chi \in \mathbb{T}_2^{\mathrm{cont}}$, let*

$$
S_t F(\chi) := \mathbb{E}_\chi\big(F(X_t)\big), \tag{4.3}
$$

where, under $\mathbb{E}_\chi$, $X = (X_t)_{t\geq 0}$ is the Aldous diffusion on $\mathbb{T}_2^{\mathrm{cont}}$ started in χ. Then $(S_t)_{t\geq 0}$ is a Feller semi-group. In particular, the Aldous diffusion is a strong Markov process.

Proof. $(S_t)_{t\geq 0}$ is well-defined by existence and uniqueness shown in Corollaries 3.4 and 4.2. It is a semi-group on the set of bounded measurable functions on $\mathbb{T}_2^{\mathrm{cont}}$ by the Markov property of X, which in turn follows from uniqueness and Theorem 4.4.2(a) in [EK86]. Recall from Proposition 2.10 and Lemma 2.12 that the state space $\mathbb{T}_2^{\mathrm{cont}}$ is compact, and the set $\mathcal{D}(\Omega_{\mathrm{Ald}})$ of shape polynomials is uniformly dense in $\mathcal{C}(\mathbb{T}_2^{\mathrm{cont}})$. Hence, in order to show that S_t maps $\mathcal{C}(\mathbb{T}_2^{\mathrm{cont}})$ into itself, it is enough to show $S_t F \in \mathcal{C}(\mathbb{T}_2^{\mathrm{cont}})$ for $F = \Phi^{m,\mathfrak{t}} \in \mathcal{D}(\Omega_{\mathrm{Ald}})$. Using duality, we have $S_t \Phi^{m,\mathfrak{t}}(\chi) = \mathbb{E}\big[\Phi^{m,Y_t}(\chi)\big]$ for the $\mathfrak{C}_m$-valued Aldous chain started in $\mathfrak{t}$. Thus $S_t\Phi^{m,\mathfrak{t}} \in \mathcal{C}(\mathbb{T}_2^{\mathrm{cont}})$, because it is a finite linear combination of the continuous functions $\Phi^{m,\mathfrak{t}'}$ for different $\mathfrak{t}' \in \mathfrak{C}_m$.

We have shown that $(S_t)_{t\geq 0}$ is a contraction semigroup on $\mathcal{C}(\mathbb{T}_2^{\mathrm{cont}})$. Its weak continuity follows directly from right-continuity of the sample paths of X. Weak continuity implies strong continuity, e.g., by Theorem 19.6 of [Kal02]. Therefore, X is a Feller process. This also implies the strong Markov property (e.g. [EK86, Theorem 4.2.7]). □

5. LONG TERM BEHAVIOR AND THE BROWNIAN CRT

In this section, we define the algebraic measure Brownian CRT, and provide the joint density of the cladogram shape spanned by a sample of finite size together with the vector of subtree masses branching off the edges of the cladogram. Moreover, we show that the algebraic measure Brownian CRT is invariant under the Aldous diffusion and that for any initial $\chi \in \mathbb{T}_2^N$, the Aldous diffusion converges in law to the algebraic measure Brownian CRT as time goes to infinity.

Recall the definitions of the set $\mathfrak{C}_m$ of m-cladograms (after Definition 2.4) and of the shape $\mathfrak{s}_{(T,c)}(u_1,...,u_m)$ spanned by the vector of m points $u_1,...,u_m \in T$ (Definition 2.5). We define

Definition 5.1 (Algebraic measure Brownian CRT)**.** The algebraic measure Brownian CRT is the unique (in distribution) random binary algebraic measure tree $\mathcal{X}_{\mathrm{CRT}} = (T,c,\mu)$ with uniform annealed sample shape distribution, i.e., for all $m \in \mathbb{N}$, for all $\mathfrak{t} \in \mathfrak{C}_m$,

$$\mathbb{E}_{\mathrm{CRT}}\Big[\mu^{\otimes m}\big\{(u_1,...,u_m) : \mathfrak{s}_{(T,c)}(u_1,.,,,.u_m) = \mathfrak{t}\big\}\Big] = \tfrac{1}{\#\mathfrak{C}_m}. \tag{5.1}$$

Note that there is a unique law on $\mathbb{T}_2$ satisfying (5.1) because the sample shape distribution separates probability measures on $\mathbb{T}_2$, and it is realized through the well-known Brownian CRT once we ignore the distances (compare, [Ald93, Theorem 23]).

Now we provide the analog of [Ald93, Theorem 23] by considering, together with the sample shape, the vector of masses of the subtrees branching off the edges of the shape cladogram. As expected, under the annealed law of the Brownian CRT, we obtain that this vector is Dirichlet distributed and independent of the shape. To state the result more precisely, for $\underline{u} = (u_1,\dots,u_m) \in T$ let $T(\underline{u}) = c(\{u_1,\dots,u_m\}^3)$ be the generated subtree, and for $e = \{e_x,e_y\} \in \mathrm{edge}(T(\underline{u}))$ let

$$\eta_{(T,c,\mu)}(\underline{u},e) := \mu\Big(\Big\{v \in T : c(v,e_x,e_y) \in (e_x,e_y) \cup \big(\{e_x,e_y\} \cap \mathrm{lf}(T(\underline{u}))\big)\Big\}\Big). \tag{5.2}$$

Let $\eta_{(T,c,\mu)}(\underline{u}) = \big(\eta_{(T,c,\mu)}(\underline{u},e)\big)_{e\in \mathrm{edge}(T(\underline{u}))}$ be the vector of these $2m-3$ masses (assuming that $u_1,\dots,u_m$ are distinct). We obtain the following, which is proven in the special case $m = 3$ in [Ald94, Theorem 2].

Proposition 5.2 (Brownian CRT and $\mathrm{Dir}(\frac{1}{2},...,\frac{1}{2})$)**.** *Let $\mathcal{X}_{\mathrm{CRT}}$ be the Brownian CRT, $m \in \mathbb{N}$, $\mathfrak{t} \in \mathfrak{C}_m$ and $f\colon \Delta_{2m-3} \to \mathbb{R}$ bounded measurable, where Δ_k is the k-simplex for $k \in \mathbb{N}$, i.e.,*

$$\Delta_k := \big\{x \in [0,1]^k : x_1 + ... + x_k = 1\big\}. \tag{5.3}$$

Then the following holds:

$$\begin{aligned}
&\mathbb{E}_{\mathrm{CRT}}\Big[\int \mu^{\otimes m}(\mathrm{d}\underline{u})\mathbb{1}_{\mathfrak{t}}\big(\mathfrak{s}_{(T,c)}(\underline{u})\big)f\big(\eta_{(T,c,\mu)}(\underline{u})\big)\Big] \\
&= \tfrac{1}{\#\mathfrak{C}_m}\int_{\Delta_{2m-3}} f(\underline{x})\ \mathrm{Dir}(\tfrac12,\dots,\tfrac12)(\mathrm{d}\underline{x}) \\
&= \tfrac{\Gamma(m-\frac32)}{\#\mathfrak{C}_m\Gamma(\frac12)^{2m-3}}\int_{\Delta_{2m-3}} f(\underline{x})\big(x_1\cdot\ldots\cdot x_{2m-3}\big)^{-\frac12}\,\mathrm{d}\underline{x},
\end{aligned} \tag{5.4}$$

where $\mathrm{Dir}(\frac12,\dots,\frac12)$ is the Dirichlet distribution and the last integral in (5.4) *is w.r.t. Lebesgue measure on Δ_{2m-3}.*

Proof. We follow Aldous's proof of [Ald94, Theorem 2] and show a local limit theorem for the subtree mass distribution of a uniform N-cladogram (with uniform distribution on the leaves) as $N \to \infty$. Because the uniform N-cladogram converges to the Brownian CRT, this implies the claim.

Fix $m \in \mathbb{N}$ and $\mathfrak{t} \in \mathfrak{C}_m$. Let $N \geq m$. Denote by $\pi_{N,m}\colon \mathfrak{C}_N \to \mathfrak{C}_m$ the projection map which sends an N-cladogram (T,c,ζ) to the m-cladogram spanned by the first m leaves, i.e., for $T_m := c\big(\zeta(\{1,...,m\})^3\big)$,

$$\pi_{N,m}(T,c,\zeta) := \big(T_m,\ c\restriction_{T_m},\ \zeta\restriction_{\{1,...,m\}}\big). \tag{5.5}$$

For $n = (n_e)_{e\in\text{edge}(\mathfrak{t})}$ with $\sum_e n_e = N$, let $q_N(n)$ be the probability that the first m leaves of a uniform N-cladogram span the m-cladogram $\mathfrak{t}$, and the numbers of leaves of the N-cladogram in the subtrees corresponding to the edges of $\mathfrak{t}$ are given by the vector n, i.e.,

$$\begin{aligned} q_N(n) &:= \frac{1}{\#\mathfrak{C}_N}\#\Big\{(T,c,\zeta)\in\pi^{-1}_{N,m}(\mathfrak{t}) : N\cdot\eta_{(T,c,\zeta)}\big(\zeta(1),\dots,\zeta(m)\big)=n\Big\} \\ &= \frac{1}{\#\mathfrak{C}_N}\cdot\frac{(N-m)!\prod_{e\in\text{ex-edge}(\mathfrak{t})}\#\mathfrak{C}_{n_e+1}\prod_{e\in\text{in-edge}(\mathfrak{t})}\#\mathfrak{C}_{n_e+2}}{\prod_{e\in\text{ex-edge}(\mathfrak{t})}(n_e-1)!\prod_{e\in\text{in-edge}(\mathfrak{t})}n_e!}. \end{aligned} \tag{5.6}$$

The first factor in the numerator together with the denominator counts the number of possibilities to distribute the $N-m$ remaining leaves to the edges of the m-cladogram $\mathfrak{t}$ (with quantities specified by n), and the products in the numerator count the possibilities to give cladogram structure to the leaves associated to every edge of $\mathfrak{t}$. For an external edge $e\in$ ex-edge($\mathfrak{t}$), this is the number of (n_e+1)-cladograms (identify the additional leaf with the branch point of $\mathfrak{t}$ it is attached to). For an internal edge, we need two additional leaves. Recall that

$$\#\mathfrak{C}_N = (2N-5)!! = \tfrac{(2N-4)!}{2^{N-2}(N-2)!} \approx (N-2)!\cdot 2^{(N-2)}(\pi(N-2))^{-\frac12}, \tag{5.7}$$

where $\approx$ means that the multiplicative error tends to zero as $N\to\infty$, and we have applied the Stirling formula.

Fix $\eta=(\eta_1,...,\eta_{2m-3})\in\Delta_{2m-3}$ with $\eta_i>0$ for $i=1,\dots,2m-3$. For $n_i=n_i(N)$ with $\sum_{i=1}^{2m-3}n_i=N$ and $N^{-1}n_i\to\eta_i$, as $N\to\infty$, we obtain (using the convention that the first m edges are external)

$$\begin{aligned} q_N(n) &= \frac{1}{\#\mathfrak{C}_N}\cdot(N-m)!\cdot\prod_{i=1}^{m}\frac{\#\mathfrak{C}_{n_i+1}}{(n_i-1)!}\prod_{i=m+1}^{2m-3}\frac{\#\mathfrak{C}_{n_i+2}}{n_i!} \\ &\approx \frac{\sqrt{\pi(N-2)}}{(N-2)!\cdot 2^{N-2}}\cdot(N-m)!\prod_{i=1}^{m}2^{n_i-1}(\pi(n_i-1))^{-\frac12}\prod_{i=m+1}^{2m-3}2^{n_i}(\pi n_i)^{-\frac12} \\ &= \sqrt{(N-2)}\cdot\tfrac{(N-m)!}{(N-2)!}\cdot 2^{2-m}\pi^{\frac12-\frac{2m-3}{2}}\prod_{i=1}^{m}(n_i-1)^{-\frac12}\prod_{i=m+1}^{2m-3}n_i^{-\frac12} \\ &\approx N^{-(m-\frac52)}\cdot(2\pi)^{2-m}\cdot N^{-(m-\frac32)}\cdot\big(\eta_1\eta_2\cdot\ldots\cdot\eta_{2m-3}\big)^{-\frac12} \\ &= N^{-(2m-4)}\cdot\frac{1}{\#\mathfrak{C}_m}\cdot\frac{\Gamma(\frac{2m-3}{2})}{\Gamma(\frac12)^{2m-3}}\big(\eta_1\eta_2\cdot\ldots\cdot\eta_{2m-3}\big)^{-\frac12}. \end{aligned} \tag{5.8}$$

This gives the claimed Dirichlet density on the $(2m-3)$-simplex in the limit. □

Proposition 5.3 (Convergence to the CRT)**.** *Let $(X_t)_{t\ge0}$ be the Aldous diffusion started in $\chi\in\mathbb{T}_2^{\text{cont}}$, and $\mathcal{X}_{\text{CRT}}$ the algebraic measure Brownian CRT. Then*

$$X_t \underset{t\to\infty}{\Longrightarrow} \mathcal{X}_{\text{CRT}}. \tag{5.9}$$

In particular, the algebraic measure Brownian CRT is the unique invariant distribution of the Aldous diffusion.

Proof. Fix $m\in\mathbb{N}$ and $\mathfrak{t}\in\mathfrak{C}_m$. Let $(Y_t)_{t\ge0}$ be the Aldous chain on m-caldograms started in $Y_0=\mathfrak{t}$. Then, for $\Phi^{m,\mathfrak{t}}\in\mathcal{D}(\Omega_{\text{Ald}})$ as in (2.20), we have by duality (Proposition 4.1)

$$\mathbb{E}\big(\Phi^{m,\mathfrak{t}}(X_t)\big) = \sum_{\mathfrak{t}'\in\mathfrak{C}_m}\mathbb{P}\{Y_t=\mathfrak{t}'\}\Phi^{m,\mathfrak{t}'}(\chi). \tag{5.10}$$

Because the uniform distribution on $\mathfrak{C}_m$ is the unique reversible distribution of the Aldous chain (see [Ald00]), $\lim_{t\to\infty}\mathbb{P}\{Y_t=\mathfrak{t}'\}=\frac{1}{\#\mathfrak{C}_m}$ for every $\mathfrak{t}'\in\mathfrak{C}_m$. Because $\sum_{\mathfrak{t}'\in\mathfrak{C}_m}\Phi^{m,\mathfrak{t}'}=1$ on $\mathbb{T}_2^{\text{cont}}$, this means

$$\lim_{t\to\infty} \mathbb{E}\big(\Phi^{m,\mathrm{t}}(X_t)\big) = \frac{1}{\#\mathfrak{C}_m} = \mathbb{E}\big(\Phi^{m,\mathrm{t}}(\mathcal{X}_{\mathrm{CRT}})\big). \tag{5.11}$$

Because $\mathcal{D}(\Omega_{\mathrm{Ald}})$ is convergence determining for probability measures on $\mathbb{T}_2^{\mathrm{cont}}$, this proves (5.9). Invariance of the law of $\mathcal{X}_{\mathrm{CRT}}$ follows from the convergence (5.9) together with the Feller property (Corollary 4.3). □

In summary, we have now proven the first two theorems.

Proof of Theorems 1 and 2. Well-posedness of the martingale problem is shown in Corollaries 3.4 and 4.2. Continuous paths and tightness of the sequence of Aldous chains are shown in Corollary 3.3. Furthermore, every limit process satisfies the martingale problem (Corollary 3.3) and this implies convergence because of the uniqueness shown in Corollary 4.2. The Feller property is shown in Corollary 4.3, and unique ergodicity with the algebraic measure Brownian CRT as invariant distribution is shown in Proposition 5.3. □

6. On the dynamics of the sample subtree mass vector

In this section, we further study the Aldous diffusion on binary, algebraic non-atomic measure trees and prove Theorem 3. Recall from Proposition 5.2 that under the annealed law of the Brownian CRT, the sample tree shape is uniform and independent of the vector of subtree masses branching of the cladogram spanned by the sample. Furthermore, the vector of subtree masses is Dirichlet distributed. Next, we study the infinitesimal evolution of the quenched law of this vector under the dynamics of the Aldous diffusion in the case of sample size $m = 3$.

Recall the definition of the components $\mathcal{S}_v(u)$, $u, v \in T$, from (2.2), and from (1.14) that $\underline{\eta}(\underline{u})$ with $\underline{u} = (u_1, u_2, u_3) \in T^3$ denotes the vector of the three masses of the components connected to $c(\underline{u})$, i.e.

$$\underline{\eta}(\underline{u}) = \big(\eta_i(\underline{u})\big)_{i=1,2,3} = \big(\mu\big(\mathcal{S}_{c(\underline{u})}(u_i)\big)\big)_{i=1,2,3}. \tag{6.1}$$

With a slight abuse of notation, we also denote for $v \in \mathrm{br}(T)$ by $\underline{\eta}(v)$ the μ-masses of the three components of $T \setminus \{v\}$ (ordered decreasingly for definiteness), so that $\underline{\eta}(\underline{u}) = \underline{\eta}(c(\underline{u}))$ up to a permutation of the entries of the vector. Also recall that for mass-polynomials

$$\Phi^f(\mathcal{X}) = \int f\big(\underline{\eta}(c(\underline{u}))\big)\, \mu^{\otimes 3}(\mathrm{d}\underline{u}) \tag{6.2}$$

with $f \in \mathcal{C}^2\big([0,1]^3\big)$ and $\mathcal{X} = (T, c, \mu) \in \mathbb{T}_2$, we define

$$\begin{aligned}\Omega_{\mathrm{Ald}}\Phi^f(T,c,\mu) = \int_{T^3} \mathrm{d}\mu^{\otimes 3} \Big(2\sum_{i,j=1}^{3} \eta_i(\delta_{ij} - \eta_j)\partial^2_{ij} f(\underline{\eta}) + 3\sum_{i=1}^{3}(1-3\eta_i)\partial_i f(\underline{\eta}) \\ + \tfrac{1}{2}\sum_{i,j=1,\, i\neq j}^{3} \Theta_{i,j} f(\underline{\eta}) + \sum_{i=1}^{3}\big(f(e_i) - f(\underline{\eta})\big)\Big).\end{aligned} \tag{6.3}$$

see (1.15) and (1.17), and the definition of migration operators $\Theta_{i,j}$ in (1.16). Note that $\Phi^f \in \mathcal{C}(\mathbb{T}_2)$ by Proposition 2.8, and because the functions $\Theta_{ij} f$ are continuous due to continuous differentiability of f, $\Omega_{\mathrm{Ald}}\Phi^f$ is also a mass-polynomial. In particular, $\Omega_{\mathrm{Ald}}\Phi^f \in \mathcal{C}(\mathbb{T}_2^{\mathrm{cont}})$.

Remark 6.1. 1. If $f \in \mathcal{C}^2([0,1]^3)$ is symmetric, we can use the symmetry of the sampling procedure and rewrite (6.3) as

$$\begin{aligned}\Omega_{\mathrm{Ald}}\Phi^f(\mathcal{X}) = 3\int_{T^3} \mathrm{d}\mu^{\otimes 3} \Big(2\eta_1(1-\eta_1)\partial^2_{11} f(\underline{\eta}) - 4\eta_1\eta_2\partial^2_{12} f(\underline{\eta}) + 3(1-3\eta_1)\partial_1 f(\underline{\eta}) \\ + \Theta_{1,2} f(\underline{\eta}) + f(1,0,0) - f(\underline{\eta})\Big).\end{aligned} \tag{6.4}$$

This helps to reduce the number of terms in explicit calculations.

2. We can often assume f to be symmetric. If f is not necessarily symmetric, we use that $\Phi^f = \Phi^{\tilde{f}}$ for the symmetrization $\tilde{f}$ of f defined as follows. Given a permutation π of $\{1,2,3\}$, define $f_\pi(x_1,x_2,x_3) := f(x_{\pi(1)}, x_{\pi(2)}, x_{\pi(3)})$. Then $\tilde{f} = \frac{1}{6}\sum_{\pi\in S_3} f_\pi$. $\Diamond$

For the proof of Theorem 3, we do not use the martingale problem of Theorem 1, because approximating the mass polynomial of degree three by shape polynomials, though possible in theory, seems difficult in praxis. Instead, we show that uniform convergence of generators holds also for mass polynomials of degree three, and use the diffusion approximation of Theorem 2. For $N \in \mathbb{N}$, recall the state space $\mathbb{T}_2^N$ from (2.18) and the Aldous chain with generator Ω_N from (1.1).

Proposition 6.2 (Subtree mass under the Aldous diffusion)**.** *For all test functions Φ^f of the form (1.15) with $f\colon [0,1]^3 \to \mathbb{R}$ twice continuously differentiable,*

$$\lim_{N\to\infty} \sup_{\chi\in\mathbb{T}_2^N} \left|\Omega_N \Phi^f(\chi) - \Omega_{\mathrm{Ald}}\Phi^f(\chi)\right| = 0. \tag{6.5}$$

From here we can prove Theorem 3 by standard arguments.

Proof of Theorem 3. Let $X = (X_t)_{t\geq 0}$ be the Aldous diffusion on $\mathbb{T}_2^{\mathrm{cont}}$. Due to Proposition 2.9, there exist $\mathbb{T}_2^N$-valued random variables X_0^N such that X_0^N converges in law to X_0. By Theorem 2, the Aldous chains X^N started in X_0^N converge in law to X. Furthermore, Ω_{Ald} maps into $\mathcal{C}(\mathbb{T}_2)$. Hence the same proof as Lemma 4.5.1 in [EK86] shows that (1.18) follows from Proposition 6.2. $\square$

Recall the intrinsic metric r_μ on an algebraic measure tree (T,c,μ), as defined in (1.5). Before proving Proposition 6.2, we show how to use the extended martingale problem from Theorem 3 to calculate the annealed average r_μ-distance in the algebraic measure Brownian CRT.

Corollary 6.3 (Mean average distance of the Brownian CRT)**.** *Let $\chi_{\mathrm{CRT}} = (T,c,\mu)$ be the algebraic measure Brownian CRT. Then*

$$\mathbb{E}_{\mathrm{CRT}}\Big[\int_{T^2} r_\mu(x,y)\,\mu^{\otimes 2}(\mathrm{d}(x,y))\Big] = \tfrac{2}{5}. \tag{6.6}$$

Proof. First, we express the average distance $\Phi(\chi) := \int r_\mu(x,y)\,\mu^{\otimes 2}(\mathrm{d}(x,y))$ for $\chi = (T,c,\mu) \in \mathbb{T}_2^{\mathrm{cont}}$ as mass polynomial of degree three. Recall the branch point distribution $\nu = \mu^{\otimes 3}\circ c^{-1}$ used in the definition of r_μ.

$$\begin{aligned}
\Phi(\chi) &= \int \sum_{v\in \mathrm{br}(T)\cap(x,y)} \nu(\{v\})\,\mu^{\otimes 2}(\mathrm{d}(x,y)) \\
&= \sum_{v\in\mathrm{br}(T)} \nu(\{v\})\mu^{\otimes 2}\big(\{(x,y) : v\in(x,y)\}\big) \\
&= 2\sum_{v\in\mathrm{br}(T)} \nu(\{v\})\big(\eta_1(v)\eta_2(v) + \eta_2(v)\eta_3(v) + \eta_1(v)\eta_3(v)\big) \\
&= 2\int \big(\eta_1(\underline{u})\eta_2(\underline{u}) + \eta_2(\underline{u})\eta_3(\underline{u}) + \eta_1(\underline{u})\eta_3(\underline{u})\big)\,\mu^{\otimes 3}(\mathrm{d}\underline{u}),
\end{aligned} \tag{6.7}$$

i.e. $\Phi = 2\Phi^f$ (on $\mathbb{T}_2^{\mathrm{cont}}$) with $f(x,y,z) := xy + yz + xz$. From here, we could obtain (6.6) by a direct computation using Proposition 5.2. But it also follows easily from the invariance of the CRT under the Aldous diffusion: Because f is symmetric, we obtain from Remark 6.1

$$\begin{aligned}
\Omega_{\mathrm{Ald}}\Phi^f(\chi) &= 3\int \Big(-4\eta_1\eta_2 + 3(1-3\eta_1)(\eta_2+\eta_3) + \frac{(\eta_1+\eta_2)\eta_3 - f(\underline{\eta})}{\eta_1} - 3\eta_1\eta_2\Big)\,\mathrm{d}\mu^{\otimes 3} \\
&= 3\int (5\eta_1 - 25\eta_1\eta_2)\,\mathrm{d}\mu^{\otimes 3} = 5 - 25\Phi^f(\chi).
\end{aligned} \tag{6.8}$$

Thus $\mathbb{E}_{\mathrm{CRT}}[\Omega_{\mathrm{Ald}}\Phi^f] = 0$ implies $\mathbb{E}_{\mathrm{CRT}}[\Phi^f(\chi)] = \frac{1}{5}$. $\square$

To prove Proposition 6.2, fix $N \in \mathbb{N}$ and $\chi = (T, c, \mu) \in \mathbb{T}_2^N$. We use some notation from the proof of Proposition 3.2, in particular recall from (3.2) that ϵ and δ denote the inverse numbers of leaves and edges, respectively, and the extended tree $(\overline{T}, \bar{c})$ which allows to represent one Aldous move on the same tree (see Figure 4). We consider μ, ν, and $\underline{\eta}$ to be defined on $(\overline{T}, \bar{c})$ and, for $z \in \mathrm{lf}(T) \times \mathrm{edge}(T)$, we denote by μ_z, ν_z, and $\underline{\eta}^z$ the corresponding objects after the Aldous move z. Because our trees are binary, the relation between the fraction of leaves and the fraction of edges in a subtree can be easily related as follows.

Lemma 6.4 (Proportion of leaves versus edges)**.** *Let $\chi = (T, c, \mu) \in \mathbb{T}_2^N$, and $S = \mathcal{S}_v(u)$ for some $v \in \mathrm{br}(T)$, $u \in T \setminus \{v\}$ a component. Let $\ell(S)$ be the fraction of the edges contained in S (including the edge to v). Then*

$$\ell(S) = \mu(S) \cdot (1 + 3\delta) - \delta. \tag{6.9}$$

Recall the branch point distribution $\nu = \mu^{\otimes 3} \circ c^{-1}$. The next Lemma shows the effect we would see if the branch point distribution remained unchanged. This corresponds exactly to Aldous's original calculation (compare with (1.2) but notice that our chain runs at total rate $N(2N-3)$ rather than N^2 as in Aldous's case). In what follows, we write $O(\epsilon)$ for terms which divided by $\epsilon = \epsilon_N$ are bounded uniformly in the tree (and N), while the bound may depend on f, and similarly for $o(\epsilon)$ and so on.

Lemma 6.5 (Wright-Fisher term with negative drift)**.** *Let f be as in Proposition 6.2. Then*

$$\begin{aligned}&\sum_{z \in \mathrm{lf}(T) \times \mathrm{edge}(T)} \sum_{v \in \mathrm{br}(T)} \nu\{v\}\big(f(\underline{\eta}^z(v)) - f(\underline{\eta}(v))\big) \\ &\quad = \int \Big(2 \sum_{i,j=1}^{3} \eta_i(\delta_{ij} - \eta_j)\partial^2_{ij} f(\underline{\eta}) - \sum_{i=1}^{3}(1 - 3\eta_i)\partial_i f(\underline{\eta})\Big)\, \mathrm{d}\mu^{\otimes 3} + o(1),\end{aligned} \tag{6.10}$$

and the $o(1)$-term tends to zero as $N \to \infty$ uniformly in the binary tree with N leaves.

Proof. According to Remark 6.1, we may and do assume w.l.o.g. that f is symmetric so that $f(\underline{\eta}(\underline{u}))$ depends on $\underline{u} \in T^3$ only through $v = c(\underline{\eta}(\underline{u}))$. Fix $v \in \mathrm{br}(T)$. To make the calculation more readable, we abbreviate $\eta_i = \eta_i(v)$ (ordered decreasingly) in the following equations as long as v is fixed. Denote the three components of $T \setminus \{v\}$ by $S_i(v)$, $i = 1, 2, 3$, ordered such that $\eta_i = \mu(S_i)$. For all $z = (u, e) \in \mathrm{lf}(T) \times \mathrm{edge}(T)$ with $u \in S_i(v)$ and $e \in S_j(v)$, a Taylor expansion yields

$$f\big(\underline{\eta}^z(v)\big) = \big(1 - \epsilon(\partial_i - \partial_j) + \tfrac{1}{2}\epsilon^2(\partial_i - \partial_j)(\partial_i - \partial_j)\big) f(\underline{\eta}) + o(\epsilon^2). \tag{6.11}$$

Using first this expansion and then Lemma 6.4, we obtain

$$\begin{aligned}A_v &:= \sum_{z \in \mathrm{lf}(T) \times \mathrm{edge}(T)} \big(f(\underline{\eta}^z) - f(\underline{\eta})\big) \\ &= \sum_{i,j=1,\, i \neq j}^{3} \tfrac{\eta_i}{\epsilon} \tfrac{\ell_j}{\delta} \epsilon\Big(\big(\partial_j - \partial_i + \tfrac{\epsilon}{2}(\partial_{ii} + \partial_{jj} - 2\partial_{ij})\big) f(\underline{\eta}) + o(\epsilon)\Big) \\ &= \sum_{i,j=1,\, i \neq j}^{3} \eta_i \frac{\eta_j(1 + 3\delta) - \delta}{\delta}\Big(\big(\partial_j - \partial_i + \tfrac{\epsilon}{2}(\partial_{ii} + \partial_{jj} - 2\partial_{ij})\big) f(\underline{\eta}) + o(\epsilon)\Big).\end{aligned} \tag{6.12}$$

As the highest order term is anti-symmetric in $i \neq j$, i.e. $\sum_{i\neq j=1}^{3} \eta_i\eta_j(\partial_j - \partial_i)f(\underline{\eta}) = 0$, and $\frac{\epsilon}{\delta} = 2 + O(\epsilon)$, we obtain

$$\begin{aligned} A_v &= -\sum_{i,j=1,\,i\neq j}^{3} \eta_i(\partial_j - \partial_i)f(\underline{\eta}) + \tfrac{\epsilon}{\delta}\sum_{i,j=1,\,i\neq j}^{3} \eta_i\eta_j\big(\partial_{ii} - \partial_{ij}\big)f(\underline{\eta}) + o(1) \\ &= -\sum_{i=1}^{3}\big(1-3\eta_i\big)\partial_i f(\underline{\eta}) + 2\sum_{i=1}^{3}\eta_i\big(1-\eta_i\big)\partial_{ii}f(\underline{\eta}) - 2\sum_{i,j=1,\,i\neq j}^{3}\eta_i\eta_j\partial_{ij}f(\underline{\eta}) + o(1) \\ &= -\sum_{i=1}^{3}(1-3\eta_i)\partial_i f(\underline{\eta}) + 2\sum_{i,j=1}^{3}\eta_i(\delta_{ij} - \eta_j)\partial_{ij}f(\underline{\eta}) + o(1), \end{aligned} \tag{6.13}$$

and the claim follows by (weighted) summation over v and Fubini's Theorem. □

Recall that, for $e \in \mathrm{edge}(T)$, we introduced $x_e, y_e \in \overline{T} \setminus T$, where x_e is "in the middle" of e and y_e is a leaf attached to x_e. Because μ is supported by T and $\eta(x_e)$ is ordered decreasingly, we always have $\eta_3(x_e) = 0$ and $\eta_1(x_e) + \eta_2(x_2) = 1$. The following lemma is easily obtained by associating a branch point to its three adjacent edges.

Lemma 6.6 (Matching lemma)**.** *Let $g\colon [0,1]^2 \to \mathbb{R}$ be symmetric. Then*

$$\sum_{e\in\mathrm{edge}(T)} g(\eta_1(x_e), \eta_2(x_e)) = \tfrac{1}{2}\sum_{v\in\mathrm{br}(T)}\sum_{i=1}^{3} g\big(\eta_i(v), 1-\eta_i(v)\big) + \tfrac{1}{2}Ng(1-\epsilon,\epsilon) \tag{6.14}$$

Proof. If $e \in \mathrm{edge}(T)$ is adjacent to $v \in \mathrm{br}(T)$, there is $i \in \{1,2,3\}$, $j \in \{1,2\}$ with $\eta_j(x_e) = \eta_i(v)$ and $\eta_{3-j}(x_e) = 1 - \eta_i(v)$. The edge e is either adjacent to precisely two branch points, or it is an external edge, i.e. adjacent to a leaf of T. In the latter case, we have $\big(\eta_1(x_e), \eta_2(x_e)\big) = (1-\epsilon, \epsilon)$, and there are N external edges. Therefore,

$$2\sum_{e\in\mathrm{edge}(T)} g(\eta_1(x_e), \eta_2(x_e)) - Ng(1-\epsilon,\epsilon) = \sum_{v\in\mathrm{br}(T)}\sum_{i=1}^{3} g\big(\eta_i(v), 1-\eta_i(v)\big) \tag{6.15}$$

as claimed. □

Proof of Proposition 6.2. We assume w.l.o.g. that f is symmetric (see Remark 6.1).

Step 1. Recall that for $z = (u,e) \in \mathrm{lf}(T) \times \mathrm{edge}(T)$, ν_z and $\underline{\eta}^z$ are the branch point distribution and mass vector after the Aldous move z, respectively. In the first step, we calculate the effect of the branch point "created" by the Aldous move z due to the fact that $\nu_z(\{x_e\})$ might be non-zero, whereas $\nu(\{x_e\}) = 0$ for all $e \in \mathrm{edge}(T)$. To this end, set

$$C_{x_e} := \sum_{z\in\mathrm{lf}(T)\times\mathrm{edge}(T)} \nu_z\{x_e\} f\big(\underline{\eta}^z(x_e)\big). \tag{6.16}$$

Recall that we order the entries of $\underline{\eta}$ decreasingly, so that $\eta_1(x_e) + \eta_2(x_e) = 1$ and $\eta_3(x_e) = 0$. For $(x,y,z) \in \Delta_3$, let

$$\begin{aligned} h_\epsilon(x,y,z) &:= \big(1 - \epsilon(2 + x\partial_1 + y\partial_2 - \partial_3)\big)f(x,y,z), \\ g_\epsilon(x,y) &:= 6xy \cdot h_\epsilon(x,y,0). \end{aligned} \tag{6.17}$$

Then g_ϵ is a symmetric function. Let $e \in \mathrm{edge}(T)$. For $z = (u, e') \in \mathrm{lf}(T) \times \mathrm{edge}(T)$, we have $\nu_z\{x_e\} \neq 0$ if and only if $e = e'$, and hence, using symmetry of f,

$$\begin{aligned} C_{x_e} &= \sum_{u \in \mathrm{lf}(T)} \nu_{(u,e)}\{x_e\} f\big(\underline{\eta}^{(u,e)}(x_e)\big) \\ &= \sum_{i=1}^{2} N\eta_i(x_e) \cdot 6\epsilon(\eta_i(x_e) - \epsilon)\eta_{3-i}(x_e) \cdot f\big(\eta_i(x_e) - \epsilon,\, \eta_{3-i}(x_e),\, \epsilon\big) \\ &= g_\epsilon\big(\eta_1(x_e),\, \eta_2(x_e)\big) + O(\epsilon^2), \end{aligned} \tag{6.18}$$

where we used, in the last equality, a first order Taylor expansion of f and the identity $\eta_1(x_e) + \eta_2(x_e) = 1$. For $v \in \mathrm{br}(\overline{T}) \setminus T$, there is a unique edge $e \in \mathrm{edge}(T)$ with $v = x_e$. Thus, using Lemma 6.6,

$$\begin{aligned} \sum_{v \in \mathrm{br}(\overline{T}) \setminus T} C_v &= \sum_{e \in \mathrm{edge}(T)} g_\epsilon\big(\eta_1(x_e),\, \eta_2(x_e)\big) + O(\epsilon) \\ &= \tfrac{1}{2} \sum_{v \in \mathrm{br}(T)} \sum_{i=1}^{3} g_\epsilon\big(\eta_i(v),\, 1 - \eta_i(v)\big) + \tfrac{1}{2} N g_\epsilon(1-\epsilon, \epsilon) + O(\epsilon) \end{aligned} \tag{6.19}$$

Now we use that $\nu\{v\} = 6\eta_1(v)\eta_2(v)\eta_3(v)$ for $v \in \mathrm{br}(T)$, and hence for any permutation $(i,j,k) \in S_3$, we have $g_\epsilon\big(\eta_i(v),\, 1 - \eta_i(v)\big) = \nu\{v\}\big(\frac{1}{\eta_j(v)} + \frac{1}{\eta_k(v)}\big) h_\epsilon(\eta_i, 1-\eta_i, 0)$, and obtain

$$\sum_{v \in \mathrm{br}(\overline{T}) \setminus T} C_v = \tfrac{1}{2} \sum_{v \in \mathrm{br}(T)} \nu\{v\} \sum_{i,j=1,\, i \neq j}^{3} \frac{h_\epsilon\big(\eta_i(v), 1 - \eta_i(v), 0\big)}{\eta_j(v)} + 3f(1,0,0) + O(\epsilon). \tag{6.20}$$

Step 2. In the second step, we calculate the effect of the change in $\nu\{v\}$ for the "old" branch points. Fix $v \in \mathrm{br}(T)$. To make the calculation more readable, we abbreviate $\eta_i = \eta_i(v)$ as long as v is fixed. We use Lemma 6.4 in the first transformation, and a first order Taylor expansion of f in the second.

$$\begin{aligned} B_v &:= \sum_{z \in \mathrm{lf}(T) \times \mathrm{edge}(T)} \big(\nu_z\{v\} - \nu\{v\}\big) f\big(\underline{\eta}^z(v)\big) \\ &= \sum_{(i,j,k) \in S_3} \frac{\eta_i}{\epsilon} \frac{\eta_j(1+3\delta) - \delta}{\delta} \cdot \nu\{v\}\Big(\frac{(\eta_i - \epsilon)(\eta_j + \epsilon)\eta_k}{\eta_i\eta_j\eta_k} - 1\Big) f(\eta_i - \epsilon, \eta_j + \epsilon, \eta_k) \\ &= \nu\{v\} \sum_{(i,j,k) \in S_3} \big(\tfrac{1}{\delta}\eta_j(1+3\delta) - 1\big) \frac{\eta_i - \eta_j - \epsilon}{\eta_j} \big(f(\underline{\eta}) + \epsilon(\partial_j - \partial_i) f(\underline{\eta}) + O(\epsilon^2)\big) \end{aligned} \tag{6.21}$$

Cancelling all terms which are anti-symmetric in (i,j), we obtain

$$\begin{aligned} B_v = \nu\{v\} \sum_{(i,j,k) \in S_3} \Big(& \tfrac{\epsilon}{\delta}(\eta_i - \eta_j)(\partial_j - \partial_i) f(\underline{\eta}) - \big(\frac{\eta_i - \epsilon}{\eta_j} - 1 + \tfrac{\epsilon}{\delta}\big) f(\underline{\eta}) \\ & - \epsilon \frac{\eta_i}{\eta_j}(\partial_j - \partial_i) f(\underline{\eta}) + O(\epsilon)\Big). \end{aligned} \tag{6.22}$$

Using $\epsilon/\delta = 2 + O(\epsilon)$, that $\sum_{(i,j,k) \in S_3} \frac{\eta_i}{\eta_j} = \sum_{(i,j,k) \in S_3} \big(\frac{1}{2\eta_j} - \frac{1}{2}\big)$, and

$$\sum_{(i,j,k) \in S_3} \tfrac{\eta_i}{\eta_j}(\partial_j - \partial_i) f = \sum_{j=1}^{3} \big(\tfrac{1-\eta_j}{\eta_j}\partial_j - \tfrac{\eta_i}{\eta_j}\partial_i - \tfrac{\eta_k}{\eta_j}\partial_k\big) f = \sum_{(i,j,k) \in S_3} \tfrac{1}{2\eta_j}(\partial_j - \eta_i\partial_i - \eta_k\partial_k) f + O(1), \tag{6.23}$$

we continue

(6.24)
$$B_v = \nu\{v\} \sum_{(i,j,k)\in S_3} \Big(4(\eta_j - \eta_i)\partial_i f(\underline{\eta}) - \big(\tfrac{1}{2\eta_j} + \tfrac{1}{2} - \tfrac{\epsilon}{\eta_j} + \tfrac{\epsilon}{2\eta_j}(\partial_j - \eta_i\partial_i - \eta_k\partial_k)\big) f(\underline{\eta}) + O(\epsilon) \Big)$$
$$= 4\nu\{v\}\sum_{i=1}^{3}(1-3\eta_i)\partial_i f(\underline{\eta}) - \nu\{v\}\sum_{(i,j,k)\in S_3} \frac{h_\epsilon(\eta_i,\eta_j,\eta_k)}{2\eta_k} - \nu\{v\}\big(3f(\underline{\eta}) + O(\epsilon)\big).$$

Step 3. Because f is twice continuously differentiable, $\frac{1}{z}(2 + x\partial_1 + y\partial_2 - \partial_3)\big(f(x,y+z,0) - f(x,y,z)\big)$ is bounded and hence

$$\begin{aligned} \tfrac{1}{z}\big(h_\epsilon(x,y+z,0) - h_\epsilon(x,y,z)\big) &= \tfrac{1}{z}\big(f(x,y+z,0) - f(x,y,z)\big) + O(\epsilon) \\ &= \Theta_{3,2} f(x,y,z) + O(\epsilon). \end{aligned} \tag{6.25}$$

Therefore, using Fubini's Theorem and combining (6.19) with (6.24) yields

(6.26)
$$\sum_{z\in \mathrm{lf}(T)\times \mathrm{edge}(T)} \sum_{v\in \mathrm{br}(\overline{T})} \big(\nu_z\{v\} - \nu\{v\}\big) f\big(\underline{\eta}^z(v)\big)$$
$$= \sum_{v\in \mathrm{br}(T)} B_v + \sum_{v\in \mathrm{br}(\overline{T})\setminus T} C_v$$
$$= \sum_{v\in \mathrm{br}(T)} \nu\{v\}\Big(4\sum_{i=1}^{3}(1-3\eta_i)\partial_i f\big(\underline{\eta}(v)\big) + \tfrac{1}{2}\sum_{i,j=1,\, i\neq j}^{3} \Theta_{i,j} f\big(\underline{\eta}(v)\big) + 3f(1,0,0) - 3f\big(\underline{\eta}(v)\big)\Big)$$
$$+ O(\epsilon).$$

Together with Lemma 6.5 (and using symmetry of f), we have obtained for $\chi \in \mathbb{T}_2^N$

$$\begin{aligned} \Omega_N \Phi^f(\chi) &= \sum_{z\in \mathrm{lf}(T)\times \mathrm{edge}(T)} \sum_{v\in \mathrm{br}(\overline{T})} \Big(\nu_z\{v\} f\big(\underline{\eta}^z(v)\big) - \nu\{v\} f\big(\underline{\eta}(v)\big)\Big) \\ &= \Omega_{\mathrm{Ald}} \Phi^f(\chi) + o(1), \end{aligned} \tag{6.27}$$

which shows the claim of Proposition 6.2. □

References

[Ald93] David Aldous, *The continuum random tree III*, Ann. Probab. **21** (1993), 248–289.

[Ald94] ———, *Recursive self-similarity for random trees, random triangulations and Brownian excursion*, Ann. Probab. **22** (1994), no. 2, 527–545. MR 1288122

[Ald00] David Aldous, *Mixing time for a Markov chain on cladograms*, Combin. Probab. Comput. **9** (2000), no. 3, 191–204. MR MR1774749 (2001f:05129)

[ALW17] Siva Athreya, Wolfgang Löhr, and Anita Winter, *Invariance principle for variable speed random walks on trees*, Ann. Probab. **45** (2017), no. 2, 625–667.

[AS01] Benjamin L. Allen and Mike Steel, *Subtree transfer operations and their induced metrics on evolutionary trees*, Ann. Comb. **5** (2001), no. 1, 1–15. MR MR1841949 (2003c:68175)

[BHV01] Louis J. Billera, Susan P. Holmes, and Karen Vogtmann, *Geometry of the space of phylogenetic trees*, Adv. in Appl. Math. **27** (2001), no. 4, 733–767. MR MR1867931 (2002k:05229)

[BRST02] Oliver Bastert, Dan Rockmore, Peter F. Stadler, and Gottfried Tinhofer, *Landscapes on spaces of trees*, Appl. Math. Comput. **131** (2002), no. 2-3, 439–459. MR MR1920237 (2003g:92001)

[DGP12] Andrej Depperschmidt, Andreas Greven, and Peter Pfaffelhuber, *Tree-valued Fleming-Viot dynamics with selection*, Ann. Appl. Probab. **22** (2012), no. 6, 2560–2615.

[EGW17] Steven N. Evans, Rudolf Grübel, and Anton Wakolbinger, *Doob-Martin boundary of Rémy's tree growth chain*, Ann. Probab. **45** (2017), no. 1, 225–277.

[EK86] Stewart N. Ethier and Thomas G. Kurtz, *Markov Processes. Characterization and Convergence*, John Wiley, New York, 1986.

[EW06] Steven N. Evans and Anita Winter, *Subtree prune and re-graft: A reversible real-tree valued Markov chain*, Ann. Probab. **34** (2006), no. 3, 918–961.

[Fel03] Joseph Felsenstein, *Inferring phylogenies*, Sinauer, 2003.

[For] Noah Forman, *Mass structure of weighted real trees*, arXiv:1801.02700v1.

[FPRWa] Noah Forman, Soumik Pal, Douglas Rizzolo, and Matthias Winkel, *Aldous diffusion I: a projective system of continuum k-tree evolutions*, arXiv:1809.07756.

[FPRWb] ———, *Diffusions on a space of interval partitions with Poisson-Dirichlet stationary distributions*, arXiv:1609.06707v2.

[FPRWc] ———, *Interval partition evolutions with emigration related to the Aldous diffusion*, arXiv:1804.01205v1.

[FPRWd] ———, *Projections of the Aldous chain on binary trees: Intertwining and consistency*, arXiv:1802.00862v1.

[GPW09] Andreas Greven, Peter Pfaffelhuber, and Anita Winter, *Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees)*, Probab. Theory Related Fields **145** (2009), 285–322. MR 2520129

[GPW13] ———, *Tree-valued resampling dynamics: Martingale problems and applications*, Probab. Theory Related Fields **155** (2013), no. 3–4, 789–838.

[Kal02] Olav Kallenberg, *Foundations of modern probability*, Springer, 2002.

[KL15] Sandra Kliem and Wolfgang Löhr, *Existence of mark functions in marked metric measure spaces*, Electron. J. Probab. **20** (2015), no. 73, 1–24. MR 3371432

[Löh13] Wolfgang Löhr, *Equivalence of Gromov-Prohorov- and Gromov's $\underline{\square}_\lambda$-metric on the space of metric measure spaces*, Electron. Commun. Probab. **18** (2013), no. 17, 1–10.

[LVW15] Wolfgang Löhr, Guillaume Voisin, and Anita Winter, *Convergence of bi-measure $\mathbb{R}$-trees and the pruning process*, Ann. Inst. H. Poincaré Probab. Statist. **51** (2015), no. 4, 1342–1368.

[LW20] Wolfgang Löhr and Anita Winter, *Spaces of algebraic measure trees and triangulations of the circle*, Bull. Soc. Math. France **in press** (2020^+).

[Nus] Josué Nussbaumer, *The Kingman coalescent: old and new aspects*, `http://www-stud.uni-due.de/~snjonuss/master_thesis_kingman_coalescent.pdf`, Master thesis.

[Pal] Soumik Pal, *On the Aldous diffusion on continuum tree, I*, arXiv:1104.4186v1.

[Sch02] Jason Schweinsberg, *An $O(n^2)$ bound for the relaxation time of a Markov chain on cladograms*, Random Structures Algorithms **20** (2002), no. 1, 59–70. MR MR1871950 (2002j:60126)

DOI: 10.17185/duepublico/72669
URN: urn:nbn:de:hbz:464-20200821-083323-0