

Laboratoire Paul Painlevé
École Doctorale MADIS

Fakultät für Mathematik

A thesis submitted for the degree of
DOCTOR OF THE UNIVERSITY OF LILLE
and for the degree of
DR. RER. NAT. OF THE UNIVERSITY OF DUISBURG-ESSEN

Discipline
Mathematics

by
JOSUÉ NUSSBAUMER

Algebraic and nested tree-valued processes

Publicly defended on May 13, 2022

PR. DAVID DEREUDRE	UNIVERSITÉ DE LILLE	EXAMINER
PR. BÉNÉDICTE HAAS	UNIVERSITÉ SORBONNE PARIS NORD	REFeree
PR. PETER PFAFFELHUBER	UNIVERSITÄT FREIBURG	REFeree
PR. VIET CHI TRAN	UNIVERSITÉ GUSTAVE EIFFEL	SUPERVISOR
DR AMANDINE VÉBER	UNIVERSITÉ PARIS CITÉ	EXAMINER
PR. ANTON WAKOLBINGER	GOETHE-UNIVERSITÄT FRANKFURT	EXAMINER
PR. ANITA WINTER	UNIVERSITÄT DUISBURG-ESSEN	SUPERVISOR

Laboratoire Paul Painlevé
École Doctorale MADIS

Fakultät für Mathematik

Thèse préparée pour l'obtention du grade de
DOCTEUR DE L'UNIVERSITÉ DE LILLE
et le grade de
DR. RER. NAT. DE L'UNIVERSITÉ DE DUISBURG-ESSEN

Discipline
Mathématiques et leurs interactions

par
JOSUÉ NUSSBAUMER

Processus aléatoires à valeurs dans les arbres algébriques et hiérarchiques

Soutenue publiquement le 13 mai 2022

PR. DAVID DEREUDRE
PR. BÉNÉDICTE HAAS
PR. PETER PFAFFELHUBER
PR. VIET CHI TRAN
DR AMANDINE VÉBER
PR. ANTON WAKOLBINGER
PR. ANITA WINTER

UNIVERSITÉ DE LILLE
UNIVERSITÉ SORBONNE PARIS NORD
UNIVERSITÄT FREIBURG
UNIVERSITÉ GUSTAVE EIFFEL
UNIVERSITÉ PARIS CITÉ
GOETHE-UNIVERSITÄT FRANKFURT
UNIVERSITÄT DUISBURG-ESSEN

EXAMINATEUR
RAPPORTEURE
RAPPORTEUR
DIRECTEUR
EXAMINATRICE
EXAMINATEUR
DIRECTRICE

Acknowledgements

*The earth is the Lord's, and everything in it,
the world, and all who live in it.*

The Bible, Psalm 24:1

Anita and Chi, I feel greatly privileged to have had you as my supervisors. I have benefited a lot from working with both of you, from your understanding of probability theory and your love of research. You have been very available for me and I really enjoyed our weekly meetings, as a source of fresh ideas and renewed motivation. Thank you for trusting me, for encouraging me and for guiding me by your answers and suggestions. These past years would have been much harder without your faithful support, and I appreciate that both our relationships go beyond research matters.

Many thanks to Bénédicte Haas and Peter Pfaffelhuber for the time you devoted to reading my thesis and for your insightful comments. I warmly thank David Dereudre, Amandine Véber and Anton Wakolbinger as well, for accepting to be part of the jury. I am grateful for the interest you showed in my work.

The conferences, seminars and schools that I have attended were great opportunities to meet inspiring colleagues, discover stimulating research topics and share my own humble contribution. Therefore, I thank the Research Training Group 2131, the SPP 1590 (Probabilistic Structures in Evolution) and the Laboratoire d'analyse et de mathématiques appliquées (LAMA) for the many financial supports I received for travels. Je remercie aussi la Chaire Modélisation Mathématique et Biodiversité pour les nombreuses rencontres organisées, et en particulier pour l'école de printemps à Aussois. Merci également au groupe de travail Grands graphes.

Cette thèse a fait l'objet d'un soutien financier de l'Université franco-allemande (UFA), que je remercie chaleureusement.

I am very thankful for the time I spent in Essen, as it was a pleasure being part of a research team with Luis, Tommaso, Geronimo, Fabian, Wolfgang, Anton, Roland. I enjoyed the probability seminars, the conference trips, the discussions over lunch and the board game nights with you. Thank you Anita for supervising my master thesis as it allowed me to discover the beautiful theory of algebraic trees. Many thanks to Dagmar for your constant help, especially regarding the cotutelle agreement!

Même si mon arrivée à Marne-la-Vallée a coïncidé avec celle de la pandémie, je garderai de très bons souvenirs des moments passés au Laboratoire d'analyse et de mathématiques appliquées (LAMA). Je suis reconnaissant envers l'ensemble des collègues de l'équipe et en particulier ceux qui m'ont aidé dans mes missions d'enseignements, et de qui j'ai beaucoup appris. Chi, je veux te remercier tout spécialement pour ton accueil, qui a grandement contribué à mon intégration au sein de l'équipe.

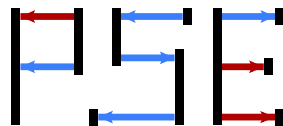
Je souhaiterais aussi exprimer toute ma gratitude aux doctorants et ATER du LAMA avec qui j'ai eu le plaisir d'échanger lors de repas, pendant le groupe de lecture ou dans les couloirs du quatrième étage. Merci à Arafat pour notre amitié, ainsi qu'à Ahmed, Benjamin, Quentin, Toan, Elias, Jean, Kacem, Gayane, Maxime, Eddy, Eddy, Han-Mai, Martin, Aurélien.

Je suis rempli de reconnaissance envers ma famille, et en particulier envers Papa, Maman, David et Clément. Merci pour votre amour et votre soutien constant durant ces dernières années. Vous m'apprenez tellement et vous m'aidez à me rappeler ce qui compte vraiment.

Finally, to my wife Rūta, who has been my greatest help and best friend during the whole PhD. Thank you for being by my side, on beautiful days, and on the other ones. With you I am at home. Et merci Noah pour toute la vie et la joie que tu apportes dans mon quotidien.

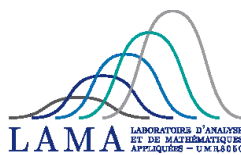


Université
franco-allemande
Deutsch-Französische
Hochschule



PROBABILISTIC STRUCTURES
IN EVOLUTION

DFG SPP 1590



Contents

Abstract	iii
Résumé	v
Zusammenfassung	vii
General introduction	ix
0.1 Chapter 1: The α -Ford algebraic measure trees	xii
0.2 Chapter 2: Resampling dynamics on metric two-level measure trees	xvi
0.3 Chapter 3: Algebraic two-level measure trees	xix
0.4 Perspectives	xxiii
1 The α-Ford algebraic measure trees	1
1.1 Introduction and motivation	1
1.2 The state space: algebraic measure trees	6
1.3 Continuum limits of the α -Ford trees	10
1.4 The α -Ford chain on fixed size cladograms	16
1.5 The α -Ford chain in the diffusion limit	19
1.6 Application on sample subtree masses	25
2 Resampling dynamics on metric two-level measure trees	33
2.1 Introduction	33
2.2 State space: metric two-level measure spaces	35
2.3 Particle approximation	39
2.4 The diffusion limit: the two-level tree-valued Fleming-Viot process	45
2.5 Expected length distribution	51
2.6 Proofs of the main results	56
3 Algebraic two-level measure trees	65
3.1 Introduction	65
3.2 Preliminaries	68
3.3 The space of algebraic two-level measure trees	73
3.4 Triangulations of the circle	76
3.5 Topologies on the subspace of binary algebraic two-level measure trees	87
3.6 Example: the Kingman algebraic two-level measure tree	92
A Results on the subtree mass distribution of the rooted α-Ford trees	95
Bibliography	97

Abstract

This thesis aims to study several tree-valued stochastic processes modeling the genealogical relationships within a population.

The first chapter is devoted to the infinite limit of the alpha model introduced by D. Ford [For]. It is a one-parameter family of random binary trees with a fixed number of leaves which interpolates between the coalescent tree (also known as the Yule tree) and the branching tree (also known as the uniform tree). To construct the α -Ford models with an infinite number of leaves, we see them as elements of the space \mathbb{T}_2 of *binary algebraic measure trees* and equip \mathbb{T}_2 with the *sample shape convergence* introduced by W. Löhrr and A. Winter [LW21]. We show that the sequence of the α -Ford trees with an increasing number of leaves converges weakly in \mathbb{T}_2 . We then determine the annealed law of the statistics of subtree masses in the particular case of the Kingman algebraic measure tree. We also introduce through a well-posed martingale problem the α -Ford diffusion which generalizes the version of the Aldous diffusion constructed by W. Löhrr, L. Mytnik and A. Winter [LMW20]. Finally, using that the α -Ford tree with infinitely many leaves is an invariant distribution of the α -Ford diffusion, we give a complete description of the annealed law of the statistics of subtree masses for any α -Ford tree through recursive relations on its moments.

In the second chapter, we are interested in the two-level version of two tree-valued resampling dynamics introduced by A. Greven, P. Pfaffelhuber and A. Winter [GPW13]. We first build the two-level tree-valued Moran dynamics as a stochastic process with values in the space of *(ultra-)metric two-level measure spaces* equipped with the *two-level Gromov-weak topology*, defined by R. Meizis [Mei19]. Under this model, a finite population of parasites divided in finitely many hosts undergoes resampling, both on the parasite and the host levels. Then, we show that the operator of this dynamics uniformly converge as the numbers of hosts and parasites both tend to infinity and that the martingale problem associated with the limit operator is well posed. The uniqueness of the solution results from a duality to the nested Kingman coalescent. We call the solution of the martingale problem the *two-level tree-valued Fleming-Viot dynamics*. Finally, we give formulas describing the evolution of the lengths of sampled subtrees under this dynamics.

The last chapter focuses on the space $\mathbb{T}^{(2)}$ of *algebraic two-level measure trees*, which are the two-level analogues of the algebraic measure trees introduced by W. Löhrr and A. Winter [LW21]. Associating each algebraic (two-level measure) tree to the metric (two-level measure) space given by the distance arising from the distribution of branch points, we use the two-level Gromov-weak topology to define a metrizable topology on $\mathbb{T}^{(2)}$. On the subspace of binary trees, we also introduce with the *two-level sample shape convergence* a more natural topology. We encode binary algebraic two-level measure trees with a triangulation of the circle together with a two-level measure on the circle line. Through this encoding, we prove that the two notions of topologies we defined on the subspace of binary algebraic two-level measure trees are equivalent and compact. We finish the chapter with a construction of the random algebraic two-level measure tree corresponding to the nested Kingman coalescent.

Résumé

Cette thèse a pour objet l'étude de processus stochastiques à valeurs arbres qui modélisent les relations généalogiques au sein d'une population.

Le premier chapitre est consacré à la limite infinie du modèle alpha introduit par D. Ford [For]. Il s'agit d'une famille à un paramètre d'arbres binaires aléatoires avec un nombre fini de feuilles, qui interpole l'arbre coalescent (aussi connu sous le nom d'arbre de Yule) et l'arbre de branchement (également connu sous le nom d'arbre uniforme). Pour construire les modèles alpha de Ford avec un nombre infini de feuilles, ils sont vus comme des variables aléatoires prenant leurs valeurs dans l'espace \mathbb{T}_2 des *arbres algébriques binaires mesurés* introduits par W. Löhr et A. Winter [LW21]. Nous montrons que les modèles alpha de Ford convergent en distribution dans \mathbb{T}_2 muni de la *convergence des formes des sous-arbres échantillonnés*. Nous déterminons ensuite la loi de distribution des masses des sous-arbres autour des points d'embranchement dans le cas particulier de l'arbre algébrique mesuré de Kingman. Nous introduisons également, via un problème de martingale bien posé, la diffusion alpha de Ford qui généralise la diffusion d'Aldous construite par W. Löhr, L. Mytnik et A. Winter [LMW20]. Enfin, en utilisant le fait que l'arbre alpha de Ford avec un nombre infini de feuilles est une distribution invariante de la diffusion alpha de Ford, nous donnons une description complète de la loi de distribution des masses des sous-arbres autour des points d'embranchement pour tout arbre alpha Ford.

Dans le deuxième chapitre, nous nous intéressons à la version à deux niveaux de deux dynamiques de rééchantillonnage sur des espaces d'arbres introduites par A. Greven, P. Pfaffelhuber et A. Winter [GPW13]. Nous construisons d'abord le modèle de Moran à valeurs arbres à deux niveaux comme un processus stochastique à valeurs dans l'espace $\mathbb{U}^{(2)}$ des *espaces (ultra-)métriques mesurés à deux niveaux*, muni de la *topologie faible-Gromov à deux niveaux*, définie par R. Meizis [Mei19]. Dans ce modèle, une population finie de parasites divisée en un nombre fini d'hôtes évolue lors d'événements de naissance-mort, à la fois au niveau des parasites et des hôtes. Nous montrons que les opérateurs de ces processus convergent uniformément lorsque les nombres d'hôtes et de parasites tendent tous les deux vers l'infini, et que le problème de martingale associé à l'opérateur limite est bien posé. L'unicité de la solution se montre par un résultat de dualité au coalescent de Kingman à deux niveaux. Nous appelons la solution du problème de martingale le *processus de Fleming-Viot à valeurs arbres à deux niveaux*. Enfin, nous donnons des formules décrivant l'évolution des longueurs des sous-arbres échantillonnés sous cette dynamique.

Le dernier chapitre est consacré à l'espace $\mathbb{T}^{(2)}$ des *arbres algébriques mesurés à deux niveaux*, qui sont les analogues à deux niveaux des arbres algébriques mesurés introduits par W. Löhr et A. Winter [LW21]. En associant chaque arbre algébrique (mesuré à deux niveaux) à l'espace métrique (mesuré à deux niveaux) donné par la distance provenant de la distribution des points de branchement, nous utilisons la topologie faible-Gromov à deux niveaux pour définir une topologie métrisable sur $\mathbb{T}^{(2)}$. Sur le sous-espace des arbres binaires, nous introduisons également une topologie plus naturelle appelée *convergence des formes des sous-arbres échantillonnés à deux niveaux*. Nous encodons les arbres algébriques binaires mesurés à deux niveaux par un couple formé d'une triangulation du cercle et d'une mesure aléatoire sur le cercle, ce qui nous permet de montrer que les deux notions de topologies sur l'espace des arbres algébriques binaires mesurés à deux niveaux sont équivalentes et compactes. Nous terminons le chapitre par la construction de l'arbre algébrique mesuré aléatoire à deux niveaux correspondant au coalescent de Kingman imbriqué.

Zusammenfassung

In dieser Dissertation werden mehrere baumwertige stochastische Prozesse untersucht, die die genealogischen Beziehungen innerhalb einer Population modellieren.

Das erste Kapitel ist dem Kontinuumslimit des von D. Ford [For] eingeführten Alpha-Modells gewidmet. Dies ist eine einparametrische Familie von zufälligen binären Bäumen mit einer festen Anzahl von Blättern, die zwischen dem Koaleszenzbaum (auch unter dem Namen Yule-Baum bekannt) und dem Verzweigungsbaum (auch bekannt als der zufällige Baum) interpoliert. Um das α -Ford-Modelle mit unendlich vielen Blättern zu konstruieren, kodieren wir binäre Bäume als *binäre algebraische Maßbäume* und versehen den Raum \mathbb{T}_2 dieser mit der sogenannten *sample-shape-Konvergenz*, welche von W. Löhr und A. Winter in [LW21] eingeführt wurde. Wir zeigen, dass eine Folge von α -Ford-Bäumen mit zunehmender Blattzahl in \mathbb{T}_2 schwach konvergiert. Wir studieren dann die Statistik der Teilbaummassen im speziellen Falle des Kingman-Koaleszentenbaumes und konstruieren dann mit Hilfe eines gut gestellten Martingalproblems die α -Ford-Diffusion, die die sogenannte Aldous-Diffusion verallgemeinert, welche zuvor von W. Löhr, L. Mytnik und A. Winter in [LMW20] konstruiert wurde. Unter Verwendung der Tatsache, dass der α -Ford-Baum mit unendlich vielen Blättern eine invariante Verteilung der α -Ford-Diffusion ist, können wir dann die Statistik der Teilbaummassen für jeden α -Ford-Baum durch rekursive Beziehungen ihrer Momente beschreiben.

Im zweiten Kapitel interessieren wir uns für die zwei-Level Version baumwertiger Resampling-Dynamiken, deren ein-Level Version von A. Greven, P. Pfaffelhuber und A. Winter in [GPW13] eingeführt wurden. Wir konstruieren zunächst die zwei-Level baumwertige Moran-Dynamik als stochastischen Prozess mit Werten im Raum *(ultra-)metrischer zwei-Level Maßräume* und statten diese mit der *zwei-Level Gromov-schwachen Topologie* aus, die von R. Meizis in [Mei19] eingeführt wurde. Bei diesem Modell wird eine endliche Population von Parasiten, die in endlich viele Wirte aufgeteilt ist, einem Resampling unterzogen, welches sowohl auf Parasiten- als auch auf Wirtsebene wirkt. Dann zeigen wir, dass die Operatoren dieser Dynamik gleichmäßig konvergieren, wenn die Anzahl der Wirte und Parasiten gegen unendlich strebt, und dass das mit dem Grenzwertoperator verbundene Martingalproblem gut gestellt ist. Die Eindeutigkeit der Lösung ergibt sich aus einer Dualität zum sogenannten *nested Kingman-Koaleszenten*. Wir nennen die Lösung des Martingalproblems die *zwei-Level baumwertige Fleming-Viot-Dynamik*. Schließlich geben wir Formeln an, die die Entwicklung der Längen von gesampelten Teilbäumen unter dieser Dynamik beschreiben.

Das letzte Kapitel konzentriert sich auf den Raum $\mathbb{T}^{(2)}$ von *algebraischen zwei-Level Maßbäumen*, die die zwei-Level Analoga der von W. Löhr und A. Winter in [LW21] eingeführten algebraischen Maßbäume sind. Wir assoziieren jeden algebraischen (zwei-Level) Maßbaum mit dem metrischen (zwei-Level) Maßraum, der durch den Abstand gegeben ist, der sich aus der Verteilung der Verzweigungspunkte ergibt, und verwenden die zwei-Level Gromov-schwache Topologie, um eine metrisierbare Topologie auf $\mathbb{T}^{(2)}$ zu definieren. Auf dem Unterraum von Binärbäumen führen wir auch mit der *zwei-Level sample-shape-Konvergenz* eine natürlichere Topologie ein. Wir codieren binär algebraische zwei-Level Maßbäume mit einem Paar bestehend aus einer Subtriangulation des Kreises zusammen mit einem zwei-Level Maß auf der Kreislinie. Mit Hilfe dieser Kodierung beweisen wir, dass die zwei Begriffe von Topologien, die wir auf dem Teilraum von binären algebraischen zwei-Level Maßbäumen definiert haben, äquivalent und kompakt sind. Wir beenden das Kapitel mit einer Konstruktion des zufälligen algebraischen zwei-Level Maßbaums des nested Kingman-Koaleszentenbaumes.

General introduction

This thesis is devoted to several mathematical tree-valued models and their limits as the sizes of the objects tend to infinity. We focus on so-called algebraic trees, which are trees without edge length generalizing countable graph-theoretic trees, and two-level models, where individuals of the first level are grouped together in clusters to form the second level and both levels undergo resampling mechanisms. The different stochastic processes we study are of interest in many branches of biology (e.g. population genetics, epidemiology, evolutionary biology), as they might describe genealogical (or phylogenetic) trees and their evolution over time.

One of the first stochastic models describing the genealogy of a large population is the Kingman coalescent. It is a partition-valued process introduced in 1982 [Kin82] to model an evolutionary tree by looking at the genealogy backwards in time. More precisely, for $N \in \mathbb{N}$, the Kingman N -coalescent is a continuous-time Markov chain $(\Pi_t^N)_{t \geq 0}$ on the space of partitions of $\{1, \dots, N\}$ defined as follows:

- Start with a partition Π_0^N (for example the trivial partition in singletons).
- For each pair of blocks (π, π') in the current partition Π_t^N at rate 1, the Markov chain jumps to the partition where the blocks π and π' have merged, i.e. $(\Pi_t^N \setminus \{\pi, \pi'\}) \cup \{\pi \cup \pi'\}$.

Thinking of the partition blocks as particles, every pair of particles coalesce independently at a constant rate to form a new particle. The following property of *sampling consistency* plays an important role in constructing the infinite limit of this model: the restriction of the N -coalescent to any subset of $\{1, \dots, m\}$ of size m is an m -coalescent. This property allows, through Kolmogorov's extension theorem, to consider the projective limit and define the Kingman coalescent $(\Pi_t)_{t \geq 0}$ of a countable population with the property that for each m , its restriction to $\{1, \dots, m\}$ is an m -coalescent.

In [Eva00], Evans encoded the evolutionary tree of the Kingman coalescent as a random metric space by considering the natural metric on \mathbb{N} induced by the coalescent: the distance $r(i, j)$ between two individuals i and j is the first time until they belong to the same block of the partition:

$$r(i, j) := \inf\{t \geq 0 : \exists \pi \in \Pi_t, i, j \in \pi\}. \quad (1)$$

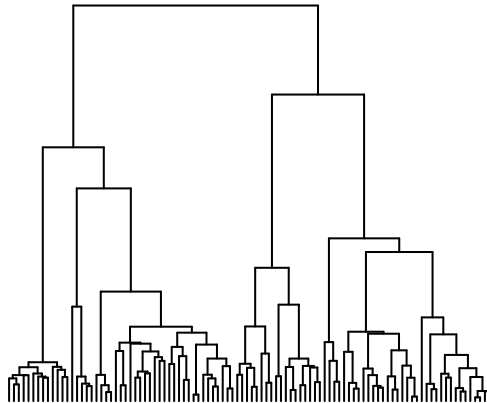


Figure 1: A tree representation of a simulation of the Kingman N -coalescent with $N = 100$. Notice that many partition blocks coalesce very quickly (see (2)).

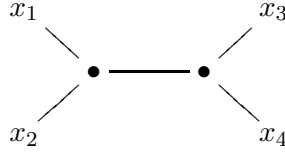


Figure 2: The only possible tree shape spanned by four points separates them into two pairs. The points x_1, x_2, x_3, x_4 are such that $r_X(x_1, x_2) + r_X(x_3, x_4) \leq \max\{r_X(x_1, x_3) + r_X(x_2, x_4), r_X(x_1, x_4) + r_X(x_2, x_3)\}$.

Looking at the genealogy forward in time, $r(i, j)$ is how long before the present the respective lines of descent of i and j diverged. Evans showed that a.s. the completion of (\mathbb{N}, r) is a compact metric space with Hausdorff dimension 1, using in particular that the Kingman coalescent *comes down from infinity*, that is,

$$\mathbb{P}\{\forall t > 0, \Pi_t \text{ has finitely many blocks}\} = 1. \quad (2)$$

Moreover, r is an ultrametric and thus satisfies the so-called four-point condition, which gives that the shape spanned by four points is “tree-like” (see Figure 2). More precisely, we say that a metric space (X, r_X) satisfies the *four-point condition* if for all $x_1, x_2, x_3, x_4 \in X$,

$$r_X(x_1, x_2) + r_X(x_3, x_4) \leq \max\{r_X(x_1, x_3) + r_X(x_2, x_4), r_X(x_1, x_4) + r_X(x_2, x_3)\}, \quad (3)$$

and if (X, r_X) is also connected, we call it an \mathbb{R} -tree. From this point of view, the Kingman coalescent defines a random \mathbb{R} -tree, whose leaves are the particles (or individuals) alive at time 0.

In the context of population genetics, lengths of sampled subtrees contain important information about genealogies. For example, Watterson proposed an estimator for the mutation rate relying on the subtree length spanned by finite samples [Wat75]. In [GPW09] was developed a theory that allows to sample leaves in the tree and evaluate the length of the subtrees spanned by finite samples. This was done by equipping ultrametric spaces (X, r) encoding trees with a probability measure μ . In order to focus on the structure of the tree rather than the labels, we say that two metric measure spaces (X, r, μ) and (X', r', μ') are *equivalent* if there exists an isometry ϕ between the supports of μ on (X, r) and of μ' on (X', r') such that $\mu' = \phi_*\mu$. We denote by \mathbb{M} the space of equivalence classes of Polish metric spaces equipped with a probability measure. Note that by considering equivalence classes, we are only interested in the restriction of r and μ to the support of μ . It is thus the measure μ that represents the population, and it allows to pick individuals (i.e. leaves in the tree) at random. If the number of individuals N is finite, a uniform sample can simply be obtained by considering the uniform distribution on the individuals

$$\mu_N := \frac{1}{N} \sum_{i=1}^N \delta_i. \quad (4)$$

If the metric space is infinite and compact as in the case of the Kingman random metric tree, one can extend the notion of a uniform distribution by approximation, i.e. by considering the weak limit of the probability measures μ_N . This requires to define a notion of (weak)-convergence on the space of metric measure spaces. Several such notions have been introduced, either only quantitatively (e.g. via Gromov’s \square_λ -metric [Gro99] or the *measured Hausdorff distance* [Fuk87, KS03]) or qualitatively as in the *Gromov-weak topology* by requiring all finite samples to converge in distribution [GPW09]. The latter gives rise to a family of convergence determining classes of functions which are very useful to us as we want to study tree-valued stochastic processes.

In the Gromov-weak topology, a sequence of trees converges to a limit tree if and only if all randomly sampled finite subtrees converge to the corresponding limit subtrees. A similar notion of convergence had already been defined by David Aldous in [Ald93]. More precisely, his approach is analogous to the classical way of considering weak convergence of processes, where finite-dimensional distributions must satisfy a consistency property (as in Kolmogorov’s extension theorem above) and the initial processes a tightness condition. In the case of trees, he replaces the property on the finite-dimensional distributions by a condition on subtrees spanned by finitely many randomly chosen leaves. By embedding trees into

the space l_1^+ of positive real-valued sequences whose series is absolutely convergent, the notion of continuous trees is formalized and the tightness criterion is expressed on l_1^+ . In particular, Aldous then applied this notion of convergence to construct the continuum random tree obtained from the Kingman coalescent (see [Ald93, Section 4.1]).

Rather than relying on the embedding of trees into l_1^+ , the Gromov-weak topology on \mathbb{M} is introduced in [GPW09] as the topology induced by *distance polynomials*, which are functions $\Phi: \mathbb{M} \rightarrow \mathbb{R}$ of the form

$$\Phi((X, r, \mu)) = \int \mu^{\otimes m}(d(x_1, \dots, x_n)) \phi((r(x_i, x_j))_{1 \leq i, j \leq m}), \quad (5)$$

where $m \in \mathbb{N}$ and $\phi: [0, \infty)^{m \times m} \rightarrow \mathbb{R}$ is a bounded continuous function. With polynomials, we consider averages over all subspaces spanned by finitely many points sampled according to μ . The Gromov-weak topology on \mathbb{M} is separable and metrizable, e.g. by the *Gromov-Prohorov metric*, which makes it a suitable state space for stochastic processes, and the set Π of polynomials separates metric measure spaces and is even convergence determining (compare [DGP11, Theorem 5] and [Löh13, Corollary 2.8]), which can be used to define Markov processes on \mathbb{M} through well-posed martingale problems. In [GPW09], a criterion for tightness was given and applied to show that Λ -coalescents satisfying a “dust-free” property (which holds for the Kingman coalescent) define an infinite (random) metric measure space.

The theory of [GPW09] has since been extended to several other state spaces: metric spaces with finite measures ([Glö12]), *marked* metric measure spaces ([DGP11, GSW16, KW19]), metric spaces equipped with locally finite measures ([ALW16]). Yet another extension of metric measure spaces was introduced in [Mei19], namely the metric *two-level* measure trees, which are Polish metric spaces (X, r) equipped with a probability measure ν on the set of probability measures on X .

This change to a two-level measure is motivated by the study of two-level systems in biology, where individuals are divided in colonies, and both colonies and individuals within colonies are subject to branching dynamics. For example, in a host-parasite system, parasites form the first level and are grouped into hosts to form the second level. Multi-level models allow to take into account the different rates at which events might happen in each level, or even the differences of mechanisms operating at each level. By equipping metric spaces with a two-level measure ν , we allow for a two-level sampling: first we sample with ν a colony $\mu \in \mathcal{M}_1(X)$, and then with μ we sample an individual within the sampled colony. The space of two-level measure spaces is then equipped with the two-level analogue of the Gromov-weak topology, which is the one induced by the set of two-level distance polynomials. In Chapter 2 of this thesis, we use it as state space to construct the two-level tree-valued Moran dynamics for finite populations and its diffusion limit as the population becomes infinite.

All the notions mentioned above have in common that they encode the structure of the genealogical tree through a metric. However, data on branch lengths in the phylogenetical trees might be uncertain and rather than focusing on pairwise metric distances, one might want to bring the attention to the tree structure. In [LW21], the metric component is ignored to focus on the so-called *branch point map* which assigns to each triplets of points their branch point. A set T together with such a branch point map c is called an *algebraic tree* and is then equipped with a natural topology. As for the metric measure spaces, a probability measure is added on algebraic trees and a notion of structure-preserving morphisms is introduced to consider equivalence classes of algebraic measure trees. Similarly to the Gromov-weak topology, a sequence of trees converges to a limit tree if all random tree shapes spanned by finite samples converge weakly to the corresponding limit shapes. Equipped with this topology, the space of equivalence classes of *binary* algebraic measure trees is compact, making it a very convenient state space to construct trees with an infinite number of leaves. This is the object of Chapter 1, as we study the diffusion limit of the α -Ford model ([For]), which is a one-parameter family of random finite trees interpolating between two well-known algebraic trees, the Kingman coalescent tree and the uniform tree.

Much like the construction of metric two-level measure spaces, we extend in the Chapter 3 the theory of algebraic measure trees to a two-level setup. We show that the space of algebraic two-level measure trees can again be equipped with a compact topology, and use it to define the Kingman algebraic two-level measure tree.

Outline. We divide the work in three chapters. Chapter 1 is devoted to the infinite limit of the α -Ford model in the space of algebraic measure trees. In Chapter 2, we introduce the two-level tree-valued Fleming-Viot dynamics as a stochastic process in the space of metric two-level measure spaces. Finally, we define in Chapter 3 the space of algebraic two-level measure trees and equip it with a compact topology. The three chapters are strongly related but can be read independently.

0.1 Chapter 1: The α -Ford algebraic measure trees

This chapter is the subject of a paper in preprint [NW].

In [Ald96], the β -splitting model was introduced in an attempt to find a one-parameter family of distributions on rooted cladograms (i.e. finite binary trees with labelled leaves) that models the “neutral evolution of species”. For $\beta \in (-2, \infty)$, the β -splitting model with N leaves is constructed by splitting the number of leaves between the two subtrees of each branch point as follows. At the root, the N leaves are split in i and $N - i$ according to the distribution

$$q_N(i) = \frac{1}{a_N(\beta)} \frac{\Gamma(\beta + i + 1)\Gamma(\beta + N - i + 1)}{\Gamma(i + 1)\Gamma(N - i + 1)}, \quad 1 \leq i \leq N - 1, \quad (6)$$

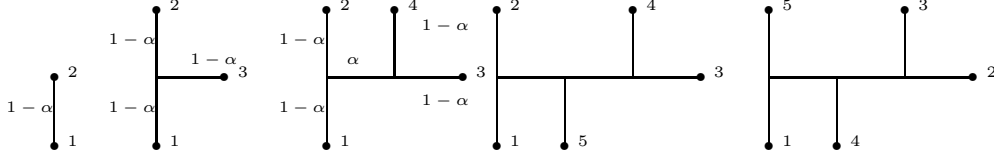
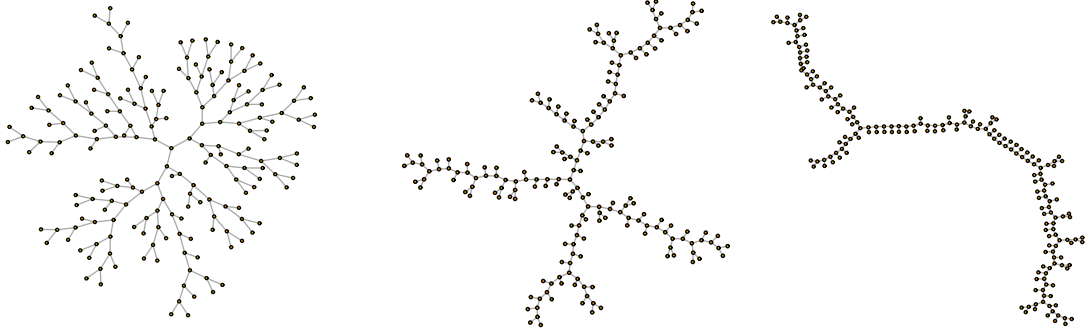
where $a_N(\beta)$ is a normalizing constant. Repeat recursively at each branch point with N replaced by the correct number of leaves. The β -splitting model interpolates between three popular models ranging from the coalescent tree (also known as Yule tree) in the case $\beta = 0$ via the branching tree (also known as uniform tree) in the case $\beta = -\frac{3}{2}$ to the totally unbalanced tree (also known as comb tree) when β converges to -2 . Furthermore, the model satisfies the two following properties which are natural properties to hope for neutral evolutionary trees: the random cladogram is *exchangeable* in the labels of the leaves and *Markovian self-similar*, i.e. the subtree above any edge has distribution the β -splitting model with the correct size, and is independant of the rest of the tree. By considering the split at the root of 30 large cladograms, Aldous observed that the model $\beta = -1$ fits the data much better than the usual models ($\beta = 0$ and $\beta = -\frac{3}{2}$).

In [For], Ford defined another one-parameter family of probability models on cladograms satisfying the properties mentioned above. For $\alpha \in [0, 1]$ and $N \in \mathbb{N}$, the α -Ford tree of size N is a cladogram with N leaves constructed recursively as follows (see Figure 3):

1. Start with one edge, and label its leaves by $\{1, 2\}$ (yielding the only 2-cladogram).
2. Given the α -Ford tree with $k \geq 2$ leaves, assign weight $1 - \alpha$ to each external edge and weight α to each internal edge.
3. Choose an edge at random according to these weights and to the middle of this edge, insert a new leaf together with an edge. Label the new leaf $k + 1$.
4. Stop when the current binary combinatorial tree has N leaves.
5. Randomly permute the leaf labels.

As for the β -splitting model, the α -Ford model interpolates between the Yule tree for $\alpha = 0$ via the uniform tree in the case $\alpha = \frac{1}{2}$ to the comb tree in the case $\alpha = 1$. We notice qualitatively different behaviours for different values of α (see Figure 4). In order to specify this difference, we are interested in characterizing the mass distribution of the α -Ford model. Since the space of all N -cladograms grows super-exponentially with the number of leaves N , we describe the statistics of mass distribution for the α -Ford trees with infinite number of leaves. But for this, one first needs a state space that allows to construct the limit of the model as the number of leaves goes to infinity.

One possible approach to define this limit is to see the cladograms as metric spaces (as \mathbb{R} -trees) by equipping them with the graph distance and use convergence in distribution for the Gromov-Hausdorff topology of the suitably rescaled trees (see [HMPW08, CFW09, PW09]). We use here another approach, that focuses on the “tree structure” rather than the metric of the considered \mathbb{R} -trees. It relies on the notion of algebraic measure trees, introduced in [LW21], which is based on the following property


 Figure 3: The 5 steps in constructing the α -Ford tree with 5 leaves.

 Figure 4: Simulations of the α -Ford models with $N = 100$ leaves for different values of the parameter α ($\alpha = 0$ on the left; $\alpha = 0.5$ in the middle; $\alpha = 0.9$ on the right).

of \mathbb{R} -trees. If (T, r) is an \mathbb{R} -tree, then for all $x_1, x_2, x_3 \in T$, there exists a unique branch point $c(x_1, x_2, x_3) \in T$ such that

$$\{c(x_1, x_2, x_3)\} = [x_1, x_2] \cap [x_2, x_3] \cap [x_3, x_1], \quad (7)$$

where for $x, y \in T$, the *interval* $[x, y]$ is defined as

$$[x, y] := \{z \in T : r(x, z) + r(z, y) = r(x, y)\}. \quad (8)$$

The function $c: T^3 \rightarrow c$, $(x_1, x_2, x_3) \mapsto c(x_1, x_2, x_3)$, called *branch point map* is symmetric and satisfies the following conditions:

- (2pc) For all $x_1, x_2 \in T$, $c(x_1, x_2, x_2) = x_2$.
- (3pc) For all $x_1, x_2, x_3 \in T$, $c(x_1, x_2, c(x_1, x_2, x_3)) = c(x_1, x_2, x_3)$.
- (4pc) For all $x_1, x_2, x_3, x_4 \in T$,

$$c(x_1, x_2, x_3) \in \{c(x_1, x_2, x_4), c(x_1, x_3, x_4), c(x_2, x_3, x_4)\}. \quad (9)$$

Algebraic trees are defined by the tree structure encoded in such a branch point map. That is, an *algebraic tree* (T, c) is a non empty set T together with a symmetric map $c: T^3 \rightarrow T$ satisfying conditions (2pc)-(4pc). Even though algebraic trees can be seen as metric trees where one has “forgotten” the metric, the branch point map is defined such that the notion of leaves, branch points, degree, subtrees, line segments, etc. can be formalized without reference to a metric and agree with the corresponding notion in the metric tree (see Section 1.2).

In order to sample leaves from an algebraic tree, we equip it with a measure. For this, we consider on an algebraic tree (T, c) the topology generated by so-called *subtree components*, which are sets of the form

$$\mathcal{S}_x(y) := \{z \in T \setminus \{x\} : c(x, z, y) \neq x\}, \quad (10)$$

with $x, y \in T$. An *algebraic measure tree* (T, c, μ) then consists of a separable algebraic tree (T, c) together with a probability measure μ on the Borel σ -algebra $\mathcal{B}(T, c)$. Associating each algebraic measure

tree to the metric measure space given by the distance arising from the distribution of branch points, we can use the Gromov-weak topology to define a metrizable topology on the set \mathbb{T} of (equivalence classes of) algebraic measure trees (see Definition 1.9 and below). Since cladograms are binary by definition, we are more specifically interested in the subspace

$$\mathbb{T}_2 := \{(T, c, \mu) \in \mathbb{T} : \text{degrees at most 3, atoms of } \mu \text{ only at leaves}\} \quad (11)$$

of (equivalence classes of) binary algebraic measure trees with no atoms on the skeleton. On \mathbb{T}_2 , another topology was introduced, which relies on the notion of tree *shape*

$$\mathfrak{s}_{(T,c)}(u_1, \dots, u_m) \quad (12)$$

spanned by m points (u_1, \dots, u_m) in an algebraic tree (T, c) . We say that a sequence of trees $(\chi_N)_N$ converges in sample shape to χ in \mathbb{T}_2 if the random sub-cladograms spanned by finite samples in χ_N converge weakly with respect to the discrete topology to the corresponding limit sub-cladograms in χ (see Definition 1.9 for a rigorous definition). One of the main results in [LW21] states that both topologies are equivalent on \mathbb{T}_2 and compact.

Back to the α -Ford model, a cladogram with N leaves defines an algebraic measure tree when we forget the leaf labels and equip it with the uniform measure on the N leaves. Therefore, the α -Ford tree with N leaves can be seen as a random algebraic measure tree $\chi_N^\alpha \in \mathbb{T}_2$. Using the sampling consistency of the model, we showed that the α -Ford trees converge weakly to a tree in the set $\mathbb{T}_2^{\text{cont}}$ of binary algebraic measure trees without atoms.

Proposition 0.1 (Continuum α -Ford trees). *Fix $\alpha \in [0, 1]$. Then there exists a random continuum binary measure tree $\chi^\alpha \in \mathbb{T}_2^{\text{cont}}$ such that*

$$\chi_N^\alpha \xRightarrow[n \rightarrow \infty]{} \chi^\alpha, \quad (13)$$

where \Rightarrow stands for weak convergence on \mathbb{T}_2 equipped with the sample shape convergence. We call χ^α the α -Ford algebraic measure tree (with infinite number of leaves).

For $\alpha = 0$, we call χ^0 the *Kingman algebraic measure tree* as it equals in law the algebraic measure tree read off from the Kingman coalescent.

In phylogenetic trees with edge lengths, all sufficient information about genealogies is contained in the lengths of subtrees spanned by a finite sample. We want to introduce a similar statistics which is more suited for algebraic measure trees, for which a priori edge lengths are not defined. For that, consider for a branch point $v \in \text{br}(T)$ the three subtree components attached to v and denote for each $u \neq v$ by $\mathcal{S}_v(u)$ the subtree component that contains $u \in T$. For $\underline{u} = (u_1, u_2, u_3) \in T^3$, let

$$\underline{\eta}(\underline{u}) := (\eta_i(\underline{u}))_{i=1,2,3} := (\mu(\mathcal{S}_{c(\underline{u})}(u_i)))_{i=1,2,3} \quad (14)$$

be the vector in the two-simplex Δ_2 of the three masses of the components connected to $c(\underline{u})$. We refer to its annealed law as *sample subtree mass distribution*. It allows to distinguish between α -Ford models for different $\alpha \in [0, 1]$. For $\alpha = \frac{1}{2}$ a combinatorial argument shows that it is equal to the Dirichlet distribution with parameters $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ (compare [Ald94a, Theorem 2] or [LMW20, Proposition 5.2]). We showed for the case $\alpha = 0$ that the sample subtree mass distribution of the Kingman algebraic measure tree χ^0 is given by the products of independent beta distributions.

Proposition 0.2 (Subtree mass distribution in the Kingman algebraic measure tree). *Let $B_{1,2}$ and $B_{2,2}$ be two independent beta random variables, such that $B_{1,2}$ has law $\text{Beta}(1, 2)$ and $B_{2,2}$ has law $\text{Beta}(2, 2)$. Then for all $f: \Delta_2 \rightarrow \mathbb{R}$ continuous bounded,*

$$\mathbb{E} \left[\int_{(T_{\chi^0})^3} (\mu_{\chi^0})^{\otimes 3}(\underline{d}\underline{u}) f(\underline{\eta}_{\chi^0}(\underline{u})) \right] = \frac{1}{6} \sum_{\pi \in \mathcal{S}_3} \mathbb{E}[f \circ \pi^*(1 - B_{1,2}, B_{1,2}B_{2,2}, B_{1,2}(1 - B_{2,2}))], \quad (15)$$

where \mathcal{S}_3 is the set of permutations of $\{1, 2, 3\}$, and for $\pi \in \mathcal{S}_3$, $\pi^*: \Delta_2 \rightarrow \Delta_2$ is the induced map $\pi^*(\underline{x}) = (x_{\pi(1)}, x_{\pi(2)}, x_{\pi(3)})$.

For a general $\alpha \in [0, 1]$, we obtained a characterization of the sample subtree mass distribution through recursive relations giving all the mixed moments of the vector of masses (see (24) below).

Another goal was to generalize with the α -Ford diffusion the results from [LMW20] on the *Aldous diffusion*. Rather than adding new leaves, we now keep the number of leaves constant by first removing a leaf picked uniformly at random and then inserting it into an edge chosen at random according to the α -Ford weights. This dynamics defines a one-parameter family of Markov chains on the space of cladograms \mathfrak{C}_N with a fixed number of leaves N , called α -Ford chain. For $\alpha = \frac{1}{2}$, the chain is reversible and is called *Aldous chain*, as it was studied in detail in [Ald00] (see also [Sch01]). Aldous conjectured in 1999 the existence of a diffusion limit of the Aldous chain as the number of leaves N goes to infinity. Such a diffusion was since constructed in two independent and different approaches ([FPRW21, FPRW20, FPRWb, FPRWa] versus [LMW20]). This conjecture was one of the motivations for introducing algebraic measure trees and we will therefore follow the approach of [LMW20] to construct the α -Ford diffusion.

To do so, we consider, for $\alpha \in [0, 1]$, the operator Ω_α acting on the following test functions called *sample shape polynomials*

$$\Phi^{m, \mathfrak{t}}(\chi) := \int_{T^m} \mu^{\otimes m}(\mathrm{d}\underline{u}) \mathbf{1}_{\mathfrak{t}}(\mathfrak{s}_{(T, c)}(\underline{u})), \quad (16)$$

with $m \in \mathbb{N}$, $\chi = (T, c, \mu) \in \mathbb{T}_2$ and $\mathfrak{t} \in \mathfrak{C}_m$, as follows:

$$\Omega_\alpha \Phi^{m, \mathfrak{t}}(\chi) := \int_{T^m} \mu^{\otimes m}(\mathrm{d}\underline{u}) \tilde{\Omega}_\alpha^m \mathbf{1}_{\mathfrak{t}}(\mathfrak{s}_{(T, c)}(\underline{u})), \quad (17)$$

where $\mathbf{1}_{\mathfrak{t}}$ plays the role of the test function for $\tilde{\Omega}_\alpha^m$, which denotes the generator of the α -Ford Markov chain on the space of m -cladograms \mathfrak{C}_m (see (1.69) for a precise definition).

Before giving the main result of the chapter, we recall the definition of a well-posed martingale problem:

Definition 0.3 (Well-posed martingale problem). Let (E, r) be a Polish space, $P_0 \in \mathcal{M}_1(E)$, \mathcal{F} a subspace of the space $\mathcal{B}(E)$ of bounded measurable functions on E and Ω a linear operator on $\mathcal{B}(E)$ with domain \mathcal{F} . An E -valued stochastic process $X = (X_t)_{t \geq 0}$ is called a solution of the $(P_0, \Omega, \mathcal{F})$ -martingale problem if X_0 has distribution P_0 , X has paths in the space $\mathcal{D}_E([0, \infty))$ of E -valued càdlàg functions, almost surely (where $\mathcal{D}_E([0, \infty))$ is equipped with the Skorohod topology) and for all $F \in \mathcal{F}$,

$$\left(F(X_t) - F(X_0) - \int_0^t \Omega F(X_s) \mathrm{d}s \right)_{t \geq 0} \quad (18)$$

is a martingale. Moreover, the $(P_0, \Omega, \mathcal{F})$ -martingale problem is said to be *well-posed* if there is a unique solution.

Our main result states that the α -Ford chains converge weakly to the unique solution of the $(\Omega_\alpha, \mathcal{D}(\Omega_\alpha))$ -martingale problem.

Theorem 0.4 (The well-posed martingale problem). Let $\alpha \in [0, 1]$ and P_0 be a probability measure on $\mathbb{T}_2^{\text{cont}}$. For each $N \in \mathbb{N}$, let $X_0^N \in \mathbb{T}_2^N$ and assume that $X_0^N \rightarrow \chi$, where χ is distributed according to P_0 . Then the α -Ford chain $X^{N, \alpha}$ starting in X_0^N converges weakly in Skorokhod path space w.r.t. the sample shape convergence to a $\mathbb{T}_2^{\text{cont}}$ -valued Feller process X^α with continuous paths.

Furthermore, X^α is the unique $\mathbb{T}_2^{\text{cont}}$ -valued Markov process $(X_t)_{t \geq 0}$ such that P_0 is the distribution of X_0 , and for all $\Phi \in \mathcal{D}(\Omega_\alpha)$, the process $M := (M_t)_{t \geq 0}$ given by

$$M_t := \Phi(X_t) - \Phi(X_0) - \int_0^t \Omega_\alpha \Phi(X_s) \mathrm{d}s \quad (19)$$

is a martingale.

We call the limit process the α -Ford diffusion. The uniqueness of the solution is shown through a Feynman-Kac duality, that is, the α -Ford diffusion is dual to the time-reversed α -Ford chain on the space of cladograms (see (1.71) and (1.73) for a definition of the dual chain). Furthermore, the α -Ford algebraic measure tree is an invariant distribution of the α -Ford diffusion.

Then we also generalized the result of [LMW20] on the evolution of the sample subtree mass distribution as follows. We extend the domain of the operator Ω_α to the set of *mass polynomials* of the form

$$\Phi^f(T, c, \mu) := \int_{T^3} f(\underline{\eta}(c(\underline{u}))) \mu^{\otimes 3}(\underline{d}\underline{u}), \quad (20)$$

where $f: [0, 1]^3 \rightarrow \mathbb{R}$ is twice continuously differentiable and $(T, c, \mu) \in \mathbb{T}_2$. We then put

$$\begin{aligned} \Omega_\alpha \Phi^f(\chi) = & \int \mu^{\otimes 3}(\underline{d}\underline{u}) \left(\sum_{i,j=1}^3 \eta_i(\delta_{ij} - \eta_j) \partial_{ij}^2 f(\underline{\eta}(\underline{u})) + (2 - \alpha) \sum_{i=1}^3 (1 - 3\eta_i) \partial_i f(\underline{\eta}(\underline{u})) \right. \\ & + (2 - 3\alpha) \sum_{i=1}^3 (f(e_i) - f(\underline{\eta}(\underline{u}))) + \frac{\alpha}{2} \sum_{i \neq j=1}^3 \frac{\mathbf{1}_{\eta_i \neq 0}}{\eta_i} (f \circ \theta_{i,j}(\underline{\eta}(\underline{u})) - f(\underline{\eta}(\underline{u}))) \\ & \left. + \frac{\alpha}{2} \sum_{i \neq j=1}^3 (\mathbf{1}_{\eta_j=0} - \mathbf{1}_{\eta_i=0}) \partial_i f(\underline{\eta}(\underline{u})) \right) \end{aligned} \quad (21)$$

where $\theta_{i,j}: \Delta_2 \rightarrow \Delta_2$ denotes the migration operator on the two-simplex which sends the vector $\underline{\eta}$ to the vector where we subtract η_i from the i th entry (resulting in the entry zero) and add it to the j th entry (resulting in $\eta_i + \eta_j$), and $e_i = (\delta_{ij})_{i=1,2,3}$ is the i th unit vector.

Theorem 0.5 (Extended martingale problem for subtree masses). *Let $\alpha \in [0, 1]$ and $X = (X_t)_{t \geq 0}$ be the α -Ford diffusion on $\mathbb{T}_2^{\text{cont}}$. Then for all mass polynomials Φ^f with $f \in \mathcal{C}^3([0, 1])$, the process $M^f := (M_t^f)_{t \geq 0}$ given by*

$$M_t^f := \Phi^f(X_t) - \Phi^f(X_0) - \int_0^t \Omega_\alpha \Phi^f(X_s) ds \quad (22)$$

is a martingale.

Since the α -Ford algebraic measure tree is an invariant distribution of the α -Ford diffusion, this result also allowed us to provide representations of the sample subtree mass distribution for general $\alpha \in [0, 1]$. For $\underline{k} = (k_1, k_2, k_3) \in \mathbb{N}_0^3$, define $f^{\underline{k}}: \Delta_2 \rightarrow [0, 1]$ by

$$f^{\underline{k}}(\underline{\eta}) = \eta_1^{k_1} \eta_2^{k_2} \eta_3^{k_3}. \quad (23)$$

Then, for all $\alpha \in [0, 1]$ and $\underline{k} \in \mathbb{N}_0^3$,

$$\begin{aligned} \mathbb{E} \left[\Phi^{f^{\underline{k}}}(\chi^\alpha) \right] = & \frac{1}{(|\underline{k}| + 3)(|\underline{k}| + 2 - 3\alpha)} \left(\sum_{i=1}^3 \mathbf{1}_{\{k_i \neq 0\}} (k_i + 1)(k_i - \alpha) \mathbb{E} \left[\Phi^{f^{\underline{k} - e_i}}(\chi^\alpha) \right] \right. \\ & + (2 - 3\alpha) (\mathbf{1}_{\{k_1=k_2=0\}} + \mathbf{1}_{\{k_2=k_3=0\}} + \mathbf{1}_{\{k_3=k_1=0\}}) \\ & \left. + \frac{\alpha}{2} \sum_{i=1}^3 \mathbf{1}_{k_i=0} \sum_{j \neq i=1}^3 \sum_{l_j=1}^{k_j} \binom{k_j}{l_j} \mathbb{E} \left[\Phi^{f^{\underline{k} + (l_j-1)e_i - l_j e_j}}(\chi^\alpha) \right] \right), \end{aligned} \quad (24)$$

where $|\underline{k}| = k_1 + k_2 + k_3$.

0.2 Chapter 2: Resampling dynamics on metric two-level measure trees

This chapter is the subject of a paper in prep [Nus].

Though the Moran model is a very basic resampling dynamics, it aroused a significant interest since its introduction in [Mor58]. It models the evolution of a finite population of fixed size, from which each pair of individuals resamples at constant rate. At a resampling event, one individual is

chosen at random from the pair, and replaces the other individual by a copy (or clone) of herself. If we trace backward in time the genealogy of the population at a given time, we recover the Kingman finite coalescent. Considering again the large population limit, the Moran model leads to the measure-valued Fleming-Viot process, which arises as the infinite limit of a larger class of finite resampling models (see [Daw93, EK93, FV79]).

In [GPW13], the space of metric measure trees served as state space to define the *tree-valued Fleming-Viot dynamics* and its particle approximation the *tree-valued Moran dynamics*. Considering tree-valued stochastic processes allows to study the evolution of the entire genealogical structure. Both dynamics were defined using well-posed martingale problems. They showed that the Kingman metric measure tree is dual to the tree-valued Fleming-Viot dynamics, as well as its unique equilibrium distribution. Finally, they gave formulas to describe the evolution of the length distribution of random finite subtrees under the dynamics. We sought to extend these results to nested populations undergoing a two-level resampling dynamics.

Multi-level models in population dynamics can be found for example in ecology to account for the levels of individuals and species, or in epidemiology for the modelling of hosts and parasites. The phylogenies of such systems bring an important light on the paths and dynamical phenomenas that are interplaying. For example in anthropology these models have served as basis for statistical studies and understand collected data (e.g. [BJ10, JBA19, LBP⁺21, VAE⁺09]). For cell models, we can refer for example to works by Kimmel [Kim97].

In the mathematical literature, we can find several two-level dynamics modelling various systems: for instance particles grouped in so-called superparticles and subject to a birth-and-death process [DHW90, Wu91, GHW95, DHV96, GH00, DGW04], parasites inside a population of cells which divide [BT11], individuals carrying two types of cells [MR13], a population undergoing mutation, selection, resampling and migration mechanisms, and divided in colonies that are also subject to selection and resampling mechanisms [Daw18]. All these models have in common that the population is represented by a measure or a two-level measure, that is, a measure on the set of measures. The dynamics we are interested in is fitted for a host-parasite system, that is, the resampling dynamics at the lower level, i.e. of parasites is constrained by the upper level: only pairs of parasites belonging to the same hosts can resample. At the upper level, resampling of hosts occurs as for the usual Moran model. It is a version of the dynamics found in [Daw18] where we focus strictly on the resampling mechanisms. However, as we want to describe the evolution of the ancestral relationships under the dynamics, we define stochastic processes on the space of metric two-level measure spaces, where the metric encodes the genealogical distances between individuals.

A *metric two-level measure space* (X, r, ν) (*m2m space* for short) is defined in [Mei19] as a Polish metric space (X, r) equipped with a probability measure $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$ on the set of probability Borel measures on X . For example by taking $X := \mathbb{N}^2$, we can represent a population of parasites in M hosts by the two-level measure

$$\nu := \frac{1}{M} \sum_{i=1}^M \delta_{\tilde{\mu}_i} = \frac{1}{M} \sum_{i=1}^M \delta_{\frac{1}{N} \sum_{j=1}^N \delta_{(i,j)}}, \quad (25)$$

where $(i, j) \in X$ represents the parasite j of the host i . The measure ν allows for a two-level sampling: first we sample with ν a host $\mu \in \mathcal{M}_1(X)$, and then with μ we sample a parasite within the sampled host. The two-level Moran model we introduce fits to this example, where each parasite belongs to a unique host. Notice however that the space $\mathcal{M}_1(\mathcal{M}_1(X))$ allows for much more general two-level measures where an individual might belong to several clusters (see Figure 5, left).

In the theory of two-level measures, the *intensity measure* $M_\nu \in \mathcal{M}_1(X)$ plays a key role to adapt results from the one-level setup. It is defined by

$$M_\nu(\cdot) := \int \nu(d\mu) \mu(\cdot). \quad (26)$$

If ν is given by (25), then

$$M_\nu = \frac{1}{M} \sum_{i=1}^M \frac{1}{N} \sum_{j=1}^N \delta_{(i,j)}. \quad (27)$$

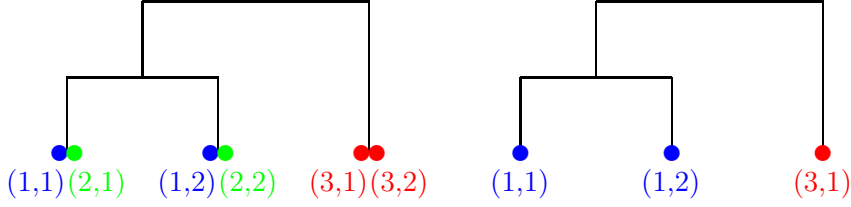


Figure 5: The two (pseudo)-m2m spaces (X_1, r_1, ν_1) (on the left) and (X_2, r_2, ν_2) (on the right) are equivalent, with $\nu_1 = \frac{1}{3}(\delta_{\frac{1}{2}(\delta_{(1,1)} + \delta_{(1,2)})} + \delta_{\frac{1}{2}(\delta_{(2,1)} + \delta_{(2,2)})} + \delta_{\frac{1}{2}(\delta_{(3,1)} + \delta_{(3,2)})})$ and $\nu_2 = \frac{2}{3}\delta_{\frac{1}{2}(\delta_{(1,1)} + \delta_{(1,2)})} + \frac{1}{3}\delta_{\delta_{(3,1)}}$.

In this case, the intensity measure allows to sample a parasite uniformly at random from the whole population, regardless of their hosts. Since the population is represented by the two-level measure ν , we are not interested in what happens on the complementary of the support of the intensity measure. Therefore, and to focus on the tree-structure rather than the labels, we say that two m2m spaces (X, r, ν) and (X', r', ν') are *equivalent* if there exists an isometry ϕ between the supports of M_ν on (X, r) and of $M_{\nu'}$ on (X', r') such that ν' is the two-level push-forward of ν under ϕ (see Figure 5 for an example of two equivalent m2m spaces). On the set $\mathbb{M}^{(2)}$ of all equivalence classes of m2m spaces, the *two-level Gromov-weak topology* is introduced as the topology induced by test functions $\Phi : \mathbb{M}^{(2)} \rightarrow \mathbb{R}$ of the form

$$\Phi((X, r, \nu)) = \int_{(\mathcal{M}_1(X))^m} \nu^{\otimes m}(\underline{d}\mu) \int_{T^{|\underline{n}|}} \bigotimes_{i=1}^m \mu_i^{\otimes n_i}(\underline{d}x_i) \phi \circ R(\underline{x}), \quad (28)$$

where $m \in \mathbb{N}$, $\underline{n} = (n_1, \dots, n_m) \in \mathbb{N}^m$ and ϕ is a real bounded continuous function on $\mathbb{R}_+^{|\underline{n}| \times |\underline{n}|}$. As in the one-level case, the two-level Gromov-weak topology on \mathbb{M} is separable and metrizable, e.g. by the so-called two-level Gromov-Prohorov metric.

We are now able to define the two-level tree-valued Fleming-Viot and Moran dynamics as stochastic processes with values in $\mathbb{M}^{(2)}$. On each level, resampling events might occur (see Figure 6):

- **Resampling of hosts:** at rate γ_H , a resampling event occurs between any two hosts. One host dies (with all the parasites it contains) and is replaced by a copy of the other host (with all the parasites it contains).
- **Resampling of parasites:** at rate γ_P , any two parasites belonging to the same host resample. One parasite dies and is replaced by a copy of the other.

We first give a construction of the two-level tree-valued Moran process following this dynamics for a population with a fixed number of hosts each containing the same number of parasites. We then define a corresponding operator acting on test functions of the form (28) with three components, one for the growth of distances between distinct parasites (belonging to the same host or not), and one for the resampling at each level. We also introduce a similar operator Ω^\uparrow for the diffusion dynamics, that we obtain as uniform limit of the operators for the finite dynamics.

Our main result states that when the numbers of hosts and of parasites simultaneously go to infinity, the *two-level tree-valued Moran process* converges weakly to the unique solution of the $(\Omega^\uparrow, \mathcal{D}(\Omega^\uparrow))$ -martingale problem.

Theorem 0.6 (The well-posed martingale problem). *Let P_0 be a probability measure on $\mathbb{U}^{(2)}$. For each $M, N \in \mathbb{N}$, let $\mathcal{U}^{M,N} := (\mathcal{U}_t^{M,N})_{t \geq 0}$ be the two-level tree-valued Moran dynamics of size (M, N) . Assume that $(\mathcal{U}_0^{M,N})_{M,N}$ converges weakly with respect to the two-level Gromov-weak topology, as $(M, N) \rightarrow \infty$, to \mathcal{U}_0 distributed according to P_0 . Then $(\mathcal{U}^{M,N})_{M,N}$ converges weakly in Skorohod path space w.r.t. the two-level Gromov-weak topology to a $\mathbb{U}^{(2)}$ -valued Markov process \mathcal{U} with càdlàg paths.*

Furthermore, \mathcal{U} is the unique $\mathbb{U}^{(2)}$ -valued Markov process $(\mathcal{X}_t)_{t \geq 0}$ such that P_0 is the distribution of \mathcal{X}_0 , and for all $\Phi \in \mathcal{D}(\Omega^\uparrow)$, the process

$$\left(\Phi(\mathcal{X}_t) - \Phi(\mathcal{X}_0) - \int_0^t \Omega^\uparrow \Phi(\mathcal{X}_s) ds \right)_{t \geq 0} \quad (29)$$

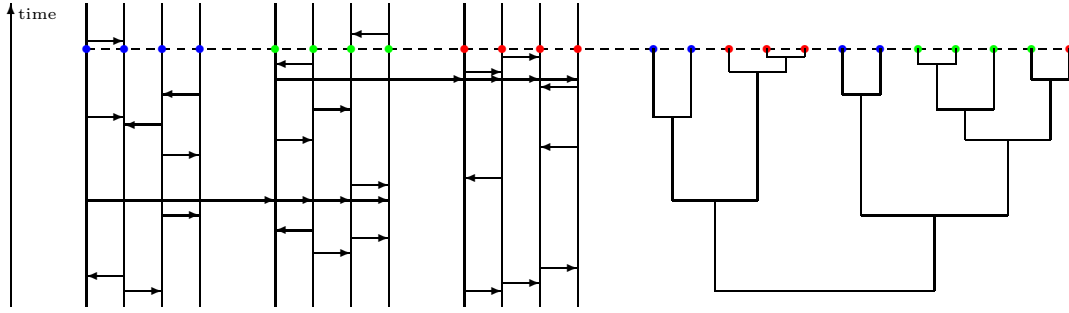


Figure 6: The graphical representation of the two-level Moran model with $M = 3$ hosts and $N = 4$ parasites in each host. An arrow from (i, k) to (i, l) indicates a parasite resampling event within host i , where (i, l) dies and (i, k) reproduces. For the resampling of hosts k and l , we draw N arrows, from (k, j) to (l, j) for each $1 \leq j \leq N$. The different colors represent different hosts.

is a martingale.

To prove the weak convergence, we showed that the sequence of two-level tree-valued Moran processes satisfies a compact containment condition and is hence tight. To show the uniqueness of the solution, we reformulated, in terms of two-level measure trees, the well-known duality between the Fleming-Viot process and the Kingman coalescent, which is the nested Kingman coalescent measure tree [Mei19, Section 10] (see also [BDLS18, BRSSJ19, LS20]).

With the tools provided in [Mei19] for the extension to two-level models, the proofs are very similar to the ones for the (one-level) tree-valued resampling dynamics in [GPW13]. However, a major obstacle arises when expressing the diffusion limit operator in a convenient way for studying the evolution of length statistics. We were still able to express formulas for the evolution of the length distribution of the subtrees spanned by small samples of parasites and we compared them to the ones obtained in [GPW13]. But when the number of hosts involved in the sampled gets larger, it becomes quickly difficult to write more general formulas. The formulas on statistics we obtained also provided information about the length of random subtrees in the nested Kingman coalescent, since it is the unique equilibrium distribution of the two-level Fleming-Viot dynamics. In particular, it allows us to estimate the rates of coalescence at each level on a real world genealogical tree, which we assume a priori to be distributed according to a nested coalescent tree.

0.3 Chapter 3: Algebraic two-level measure trees

This chapter is the subject of a paper in prep [NTW].

In the first chapter of the thesis, we study the Kingman algebraic measure tree with infinite number of leaves, and its generalization to the α -Ford algebraic measure trees. For this, we heavily relied on the compactness of the space \mathbb{T}_2 of binary algebraic measure trees. Motivated by the study of two-level models as in the second chapter, we would like to consider another extension of the Kingman algebraic measure tree, that is, where the sampling measure is replaced by a two-level measure. But for this, one first needs to define a proper state space and equip it with a notion of convergence. The goal of the third chapter is to provide these tools by extending the notion of algebraic measure trees in [LW21], in a way that is similar to the work on metric two-level measure spaces in [Mei19]. In particular, our main emphasis here is to establish a compact state space for the two-level situation.

An *algebraic two-level measure tree* (T, c, ν) (a2m tree for short) is a separable algebraic tree (T, c) equipped with a probability measure $\nu \in \mathcal{M}_1(\mathcal{M}_1(T))$ on the set of probability measures on $\mathcal{B}(T, c)$. We then say that two a2m trees (T, c, ν) and (T', c', ν') are *equivalent* if there exists a tree isomorphism ϕ between M_ν of (T, c) and the support of $M_{\nu'}$ on (T', c') such that ν' is the two-level push-forward of ν under ϕ . Denote by $\mathbb{T}^{(2)}$ the set of equivalence classes of a2m trees.

We equip $\mathbb{T}^{(2)}$ with a topology based on the two-level Gromov-weak topology introduced in [Mei19]

on the space of metric two-level measure spaces. To do so, we define the *branch point distribution* of a given a2m tree (T, c, ν) as the pushforward $\lambda_\nu := c_*(M_\nu)^{\otimes 3}$ and we associate (T, c, ν) to the metric two-level measure space (T, r_ν, ν) where r_ν is defined by

$$r_\nu(x, y) := \lambda_\nu[x, y] - \frac{1}{2}\lambda_\nu\{x\} - \frac{1}{2}\lambda_\nu\{y\}, \quad x, y \in T. \quad (30)$$

We then say that a sequence $(T_n, c_n, \nu_n)_n$ converges to (T, c, ν) if $(T_n, r_{\nu_n}, \nu_n)_n$ converges to (T, r_ν, ν) with respect to the two-level Gromov-weak topology on $\mathbb{M}^{(2)}$. We refered to this convergence as *two-level branch point distribution distance (bpdd) Gromov-weak convergence* and so we equip $\mathbb{T}^{(2)}$ with a separable and metrizable topology.

It is quite straightforward to adapt most of the results in [LW21] to a2m trees. As pointed out in [Mei19], it is often sufficient to replace the measure μ by the intensity measure M_ν . However, difficulties arise when trying to extend the coding of binary algebraic measure trees by sub-triangulations of the circle. This coding was a crucial ingredient to show that the subspace \mathbb{T}_2 of binary algebraic measure trees with no atoms on the skeleton is a compact space. Similar to encoding compact \mathbb{R} -trees by continuous excursions on the unit interval, this idea was first used by David Aldous in [Ald94a, Ald94b]. It has since then aroused significant interest (see [CLG11, BS15, CK13]) and similar approaches have been introduced for not necessarily binary trees ([Cur14, CHK14]).

We give now a brief recall of the definitions and results concerning the coding of binary algebraic measure trees by sub-triangulations of the circle from [LW21] and point out the differences in the two-level setup. For a subset A of the disc \mathbb{D} , we define

$$\Delta(A) := \{\text{connected components of } \text{conv}(A) \setminus A\}, \quad (31)$$

and

$$\nabla(A) := \{\text{connected components of } \mathbb{D} \setminus \text{conv}(A)\}. \quad (32)$$

We then say that a closed, non-empty subset C of the disc is a *sub-triangulation of the circle* \mathbb{S} if:

1. $\Delta(C)$ consists of open interiors of triangles.
2. C is the union of non-crossing (non-intersecting except at endpoints), possibly degenerate closed straight line segments with endpoints in the circle line \mathbb{S} .

We denote by \mathcal{T} the set of sub-triangulations of the circle.

From this definition, $\nabla(C)$ consists of circular segments with the bounding straight line excluded and the rest of the bounding arc included. We also define for a sub-triangulation the analogue of the components in an algebraic tree. For $x \in \Delta(C) \cup \nabla(C)$, and $y \subseteq \mathbb{D}$ connected and disjoint from $\partial_{\mathbb{D}}x$, let

$$\text{comp}_x(y) := \text{the connected component of } \mathbb{D} \setminus \partial_{\mathbb{D}}x \text{ which contains } y. \quad (33)$$

A sub-triangulation of the circle C induces a binary algebraic tree (V_C, c_V) , where $V_C := \Delta(C) \cup \nabla(C) \cup \square(C)$, triangles in $\Delta(C)$ are the branch points, circular segments in $\Delta(C) \cup \square(C)$ are leaves and the branch point of $x, y, z \in V_C$ is the triangle $c_V(x, y, z) \in \Delta(C)$ “in the middle” (see Figure 7). Furthermore, $\text{comp}_x(y)$ corresponds to the tree component $\mathcal{S}_x(y)$ in (V_C, c_V) .

Using (V_C, c_V) as a “skeleton”, the following result gives the algebraic measure tree associated to a sub-triangulation.

Proposition 0.7 (One-level coding map). *(i) For every $C \in \mathcal{T}$, there is a unique (up to equivalence) binary algebraic measure tree $\chi_C = (T_C, c_C, \mu_C)$ such that:*

- (CM1) $V_C \subseteq T_C$, $\text{br}(T_C, c_C) = \Delta(C)$, and c_C is an extension of c_V .
- (CM2) For all $x, y \in V_C$ and $\sigma \geq 0$, $\mu_C(\mathcal{S}_x(y)) = \lambda_{\mathbb{S}}(\text{comp}_x(y) \cap \mathbb{S})$.
- (CM3) $\text{at}(\mu_C) = \nabla(C)$.

(ii) The coding map $\tau : \mathcal{T} \rightarrow \mathbb{T}_2$, $C \mapsto \chi_C$ is surjective and continuous, where \mathcal{T} is equipped with the Hausdorff metric topology and \mathbb{T}_2 with the bpdd-Gromov-weak topology.

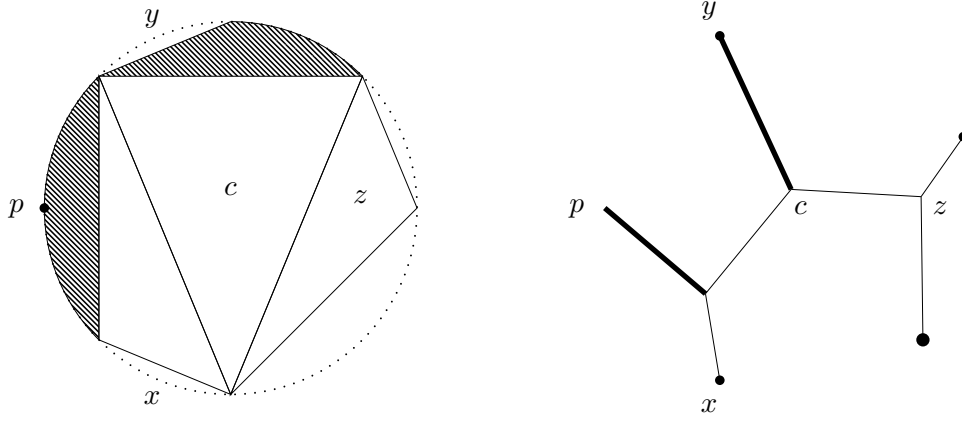


Figure 7: *On the left:* A sub-triangulation of the circle such that $\#\Delta(C) = 3$ (empty triangles), $\#\nabla(C) = 4$ (empty circular segments), $\#\blacksquare(C) = 2$ (filled areas) and $\#\square(C) = 1$ (more precisely, $\square(C) = \{p\}$). The triangle $c = c_V(x, y, z)$ is “in the middle” of x , y and z . *On the right:* The tree coded by the sub-triangulation. The leaf p does not carry an atom, but the four other leaves carry a weight given by the lengths of the corresponding arcs. The two thick segments carry a non-atomic mass.

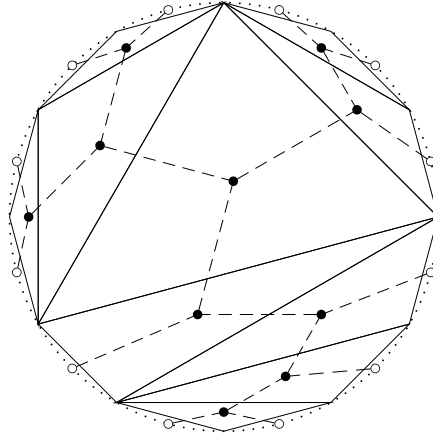


Figure 8: A triangulation of the 12-gon. Here, the coded tree is the dual graph, with uniform distribution on the leaves.

The condition (CM2) expresses that the arc lengths encode the way the mass is distributed in the algebraic measure tree. For example, if C is a triangulation of an n -gon, the Lebesgue measure of an arc of the circle line is the mass on the corresponding leaf in the tree (see Figure 8).

We introduced a similar coding of binary a2m trees in the subspace

$$\mathbb{T}_2^{(2)} := \{(T, c, \nu) \in \mathbb{T}^{(2)} : \begin{array}{l} \text{degrees at most 3,} \\ \text{atoms of } \mu \text{ only at leaves for } \nu\text{-almost every } \mu \end{array}\}. \quad (34)$$

In order to encode the information on the distribution of the random ν -mass in the tree, one cannot rely on the Lebesgue measure of arc lengths anymore. Thus, we replaced the Lebesgue measure by a two-level measure K on the circle. For technical reasons and to avoid degenerate cases, we require the intensity measure M_K to be the Lebesgue measure on the circle. Therefore, we formally construct the *coding map* that associates an a2m tree in $\mathbb{T}_2^{(2)}$ to a pair $(C, K) \in \mathfrak{D}$ where C is a subtriangulation of the circle, and $K \in \mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$ with $M_K = \lambda_{\mathbb{S}}$. In the theorem below, the conditions are very similar to the ones of Proposition 0.7, where we used the Laplace transforms of the two-level measures ν and K to define the analogue of (CM2). Notice in particular that if (C, K) encodes the a2m tree (T, c, ν) , then the sub-triangulation C encodes the algebraic (one-level) measure tree (T, c, M_ν) in the sense of Proposition 0.7.

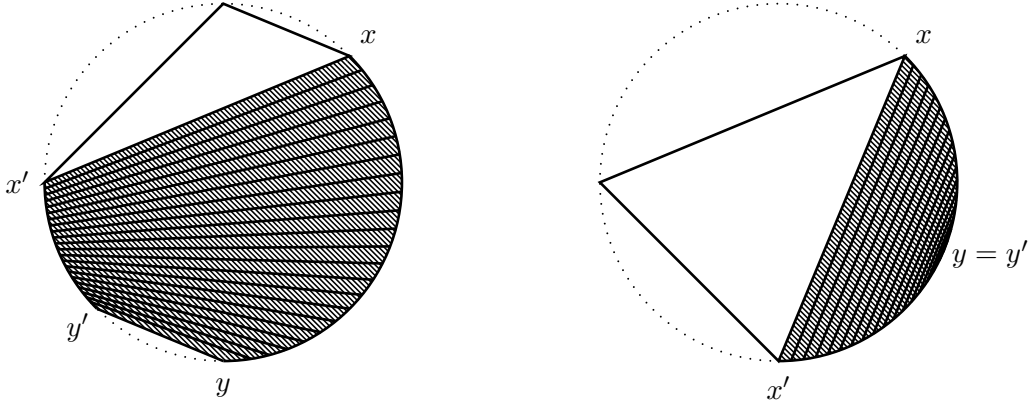


Figure 9: Two examples of “filled” areas $b \in \blacksquare(C)$ delimited by the circular segments $[x, y]$ and $[x', y']$. They are partitioned into straight line segments with endpoints in $[x, y]$ and $[x', y']$.

A major difference lies however in the definition of the skeleton V_C of the encoded tree. Contrary to the one-level case, where there is only one way (up to equivalence) to assign a non-atomic measure to a line segment in the tree, it is important in the two-level case to know how the random mass is distributed along such a line segment. Therefore, one needs to add in (CM2) the information about the mass distribution of K along the boundaries of filled areas (see Figure 7). To do this, we partition each filled area into infinitely many straight segments with endpoints in \mathbb{S} (see Figure 9) and define $V_C := \Delta(C) \cup \nabla(C) \cup \square(C) \cup \parallel(C)$, where $\parallel(C)$ is the set of all the straight segments partitioning filled areas. With this adjustment, the a2m tree encoded by (C, K) is indeed unique in our main result.

Theorem 0.8 (Two-level coding map). *(i) For all $\Gamma = (C, K) \in \mathfrak{D}$, there is a unique (up to equivalence) a2m tree $\chi_\Gamma = (T_\Gamma, c_\Gamma, \nu_\Gamma) \in \mathbb{T}_2^{(2)}$ such that:*

(CM1) $V_C \subseteq T_\Gamma$, $\text{br}(T_\Gamma, c_\Gamma) = \Delta(C)$, and c_Γ is an extension of c_V .

(CM2) For all $x, y \in V_C$ and $\sigma \geq 0$,

$$\int \nu(d\mu) e^{-\sigma \mu(\mathcal{S}_x(y))} = \int K(d\kappa) e^{-\sigma \kappa(\text{comp}_x(y) \cap \mathbb{S})}. \quad (35)$$

(CM3) $\text{at}(M_{\nu_\Gamma}) = \nabla(C)$.

(ii) The coding map $\tau : \mathfrak{D} \rightarrow \mathbb{T}_2^{(2)}$, $\Gamma \mapsto \chi_\Gamma$ is surjective.

(iii) Let \mathcal{T} be equipped with the Hausdorff metric topology, $\mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$ with the weak topology, $\mathcal{T} \times \mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$ with the product topology and $\mathbb{T}_2^{(2)}$ with the two-level bpdd-Gromov-weak topology. Then the coding map τ is continuous.

On the subspace $\mathbb{T}_2^{(2)}$, we introduced another topology that is based on labelled sub-cladograms spanned by finite samples of points in the tree. It is the two-level analogue of the sample shape topology on \mathbb{T}_2 . More precisely, we now consider subtrees coming from a two-level sampling, i.e. we first sample a finite number of measures according to ν and then we sample a finite number of leaves according to each sampled measure. Therefore, the notion of cladograms we define keeps track of this two-level sampling by using double indices (see Figure 10). Using the coding from Theorem 0.8 of binary trees in $\mathbb{T}_2^{(2)}$, we showed the following result.

Theorem 0.9 (Equivalence of topologies and compactness). *The two-level sample shape topology and the two-level bpdd-Gromov-weak topology coincide on $\mathbb{T}_2^{(2)}$. Furthermore, $\mathbb{T}_2^{(2)}$ is compact and metrizable in this topology.*

As an example, we finish by constructing the *Kingman algebraic two-level measure tree*, which is the nested Kingman coalescent measure tree without branch length (see [Mei19]). It is defined as the weak limit of a sequence of finite random a2m trees that is sampling consistent, due to the properties of finite Kingman coalescents.

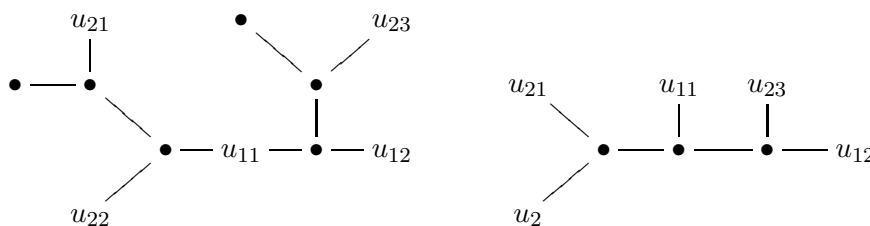


Figure 10: A tree T and the sub-cladogram spanned by $u_{11}, u_{12}, u_{21}, u_{22}, u_{23}$.

0.4 Perspectives

We give here a few open questions and ideas for future work related to the content of the thesis, many of which concern applications to real world data. For this reason, we particularly engaged in conversations with Patrick Hoscheit about possible uses of our theory. We warmly thank him for his suggestions and for providing workable data on phylogenetic trees.

0.4.1 On metric two-level measure trees

Several results from [GPW13] on tree-valued resampling dynamics still remain to be extended to the two-level setup. In particular, contrary to the one-level case, the duality we provide between the two-level tree-valued Fleming-Viot process and the nested Kingman measure tree cannot be easily adapted to the particle approximation. Therefore such a duality relation still has to be investigated in the finite case.

We also hope to develop richer formulas on the subtree length distribution of the nested Kingman coalescent and apply them to estimate the rates γ_H and γ_P of coalescence on phylogenies.

We mentioned above Watterson's mutation rate estimator that relies on the length of the subtrees spanned by samples, but also on the number of segregating sites or SNPs in these samples [Wat75]. More detailed information concerning mutations is provided by the *site frequency spectrum*, which gives how many sites are carried by each given number of individuals in the sample. It has been studied in several tree models such as the Kingman coalescent [STH⁺11, SSV15], and more generally Aldous' β -splitting model [SV18], and is often used for population genetics inference (see [GJB13] and references therein). Therefore, one might be interested in describing the site frequency spectrum of the nested Kingman coalescent to allow for inferences of two-level genealogies.

0.4.2 On algebraic (two-level) measure trees

In Chapter 1, we prove that the α -Ford algebraic measure tree is an invariant distribution of the α -Ford diffusion. In the case $\alpha = \frac{1}{2}$, it has been shown in [LMW20, Proposition 5.3] that the Aldous diffusion converges toward the algebraic measure Brownian CRT, which is thus its unique invariant distribution. It was proved by relying on a duality relation and the symmetry of the Aldous diffusion. However, the diffusion is no longer symmetric for other values of α and the argument breaks due to the exponential term in the Feynman-Kac duality equation. Therefore convergence toward the α -Ford continuum tree remains an open question for $\alpha \in [0, 1] \setminus \{\frac{1}{2}\}$.

David Aldous observed that the β -splitting model fits the data the best when $\beta = -1$ (see [Ald96, Section 4.3]). In a similar way, is there a parameter α which gives a better fit for the α -Ford model?

We give in Appendix A a few results concerning the subtree mass distribution of the rooted α -Ford algebraic measure trees. We hope to obtain full descriptions of this distribution for static trees as well as its evolution under the α -Ford diffusion. In particular, this might yield more explicit expressions of the subtree mass distribution of the unrooted α -Ford algebraic measure tree for general $\alpha \in [0, 1]$.

The notion of sampled subtree masses is an important part of the theory of algebraic measure trees developed in [LW21]. It still remains to investigate the analogue in the two-level setup. After defining this object, which needs to take into account a two-level sampling, it might be used to describe the

subtree mass distribution in Kingman algebraic two-level measure tree depending on the coalescence rates γ_H and γ_P . A natural question to ask is whether one can estimate these rates on phylogenies without branch lengths. Furthermore, the two-level resampling dynamics of Chapter 2 can easily be adapted to algebraic trees, and the evolution of the subtree mass distribution under such a dynamics is of interest.

Finally, the space of algebraic (two-level) measure trees could serve for the study of more complex static and dynamic models. For instance in [HS21], growing models of random trees were investigated which are built by gluing finite trees (rather than single edges in the α -Ford model) on the current structure.

The α -Ford algebraic measure trees

1.1 Introduction and motivation

An N -cladogram is a semi-labeled, un-rooted and binary tree with $N \geq 2$ leaves labeled $\{1, 2, \dots, N\}$ and with $N - 2$ unlabeled internal nodes. Cladograms are particular phylogenetic trees for which no information on the edge lengths is available, and which therefore only capture the tree structure.

As prototype models are needed for testing real world phylogenies, parametric families of random cladograms have been studied (compare [Ald96, For]). One such family introduced in [For] is today referred to as the α -Ford model (see also [HMPW08, CFW09, PW09, Ste09, CMR18]). Fix $\alpha \in [0, 1)$ and $N \in \mathbb{N}$. The α -Ford tree with N leaves is an N -cladogram constructed recursively as follows (compare Figure 1.1):

1. Start with one edge, and label its leaves by $\{1, 2\}$ (yielding the only 2-cladogram).
2. Given the α -Ford tree with $k \geq 2$ leaves, assign weight $1 - \alpha$ to each external edge and weight α to each internal edge.
3. Choose an edge at random according to these weights and to the middle of this edge, insert a new leaf together with an edge. Label the new leaf $k + 1$.
4. Stop when the current binary combinatorial tree has N leaves.
5. Randomly permute the leaf labels.

Note that permuting the labels in the last step ensures *consistency*. That is, for all $1 \leq m \leq N$, restricting to the sub-cladogram spanned by a uniform sample of size m from the leaf set $\{1, \dots, N\}$ yields an m -cladogram which equals in law the α -Ford tree with m leaves.

The case $\alpha = 1$ is excluded as for $k = 2, 3$ all edges have weight 0, and therefore the above construction is not well-defined. However, we can extend the construction with some care. To overcome the issue, let us simply choose the edge, at which we are inserting the next edge, uniformly among the external edges. As soon as $k = 4$, there is only one possible tree shape with exactly one inner edge and the problem disappears.

The α -Ford model interpolates between three popular models ranging from the coalescent tree (also known as Yule tree) in the case $\alpha = 0$ via the branching tree (also known as uniform tree) in the case $\alpha = \frac{1}{2}$ to the totally unbalanced tree (also known as comb tree) in the case $\alpha = 1$. In this paper we are interested in limit cladograms as the number of leaves goes to infinity. For that we will rely on the notion of continuum algebraic measure trees recently introduced in [LW21].

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* $c: T^3 \rightarrow T$ satisfying consistency conditions (see Definition 1.4). Even though algebraic trees can be seen as metric trees where one has “forgotten” the metric, the branch point map is defined such that the notion of leaves, branch points, degree, subtrees, line segments, open sets, etc. can be formalized without reference to a metric and agree with the corresponding notion in the metric tree. An algebraic measure tree (T, c, μ) consists of a separable algebraic tree (T, c) together with a probability measure

μ on the Borel σ -algebra $\mathcal{B}(T)$. The α -Ford infinite limit takes values in the state space

$$\mathbb{T}_2 := \{(T, c, \mu) \in \mathbb{T} : \text{degrees at most 3, atoms of } \mu \text{ only at leaves}\} \quad (1.1)$$

of (equivalence classes of) binary algebraic measure trees with no atoms on the skeleton, and more specifically in its subspace

$$\mathbb{T}_2^{\text{cont}} := \{(T, c, \mu) \in \mathbb{T}_2 : \mu \text{ non-atomic}\} \quad (1.2)$$

of so called *continuum* binary algebraic measure trees. We equip \mathbb{T}_2 with the so-called *sample shape convergence* (Definition 1.9), which says that a sequence $(t_N)_{N \in \mathbb{N}}$ converges to t in \mathbb{T}_2 if the random shapes $\mathfrak{s}_{(T,c)}(x_1, \dots, x_m)$ of sub-cladograms spanned by finite samples (x_1, \dots, x_m) of size m converge weakly with respect to the discrete topology (compare Definition 1.8 and Figure 1.4). It is shown in [LW21] that both \mathbb{T}_2 and $\mathbb{T}_2^{\text{cont}}$ are compact, which is very convenient for showing tightness.

To get started we first introduce the α -Ford models with an infinite number of leaves in \mathbb{T}_2 . To do this, we consider the α -Ford tree with N leaves as a random element in the subspace

$$\mathbb{T}_2^N := \{(T, c, \mu) \in \mathbb{T}_2 : \#\text{lf}(T, c) = N \text{ and } \mu = \frac{1}{N} \sum_{u \in \text{lf}(T, c)} \delta_u\}, \quad (1.3)$$

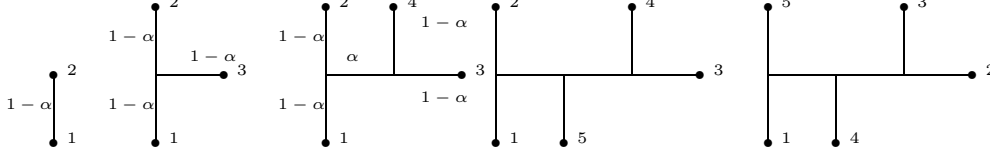
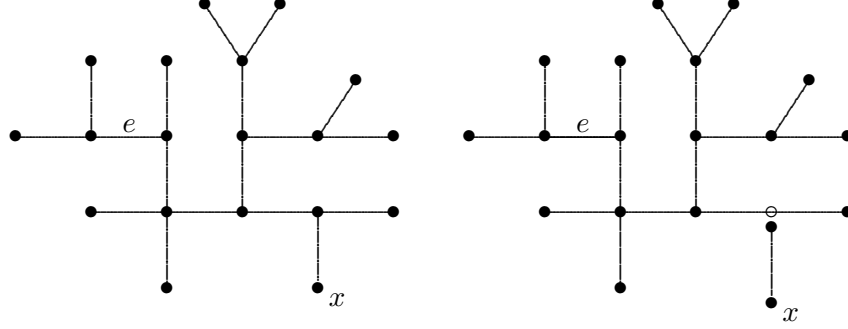
where $\text{lf}(T, c)$ denotes the set of leaves. Then, using the consistency property of Ford models, we can show that, for each $\alpha \in [0, 1]$, the sequence of such constructed random binary algebraic measure trees converges to an element of $\mathbb{T}_2^{\text{cont}}$, that we call *α -Ford algebraic measure tree* (with infinite number of leaves). For $\alpha = \frac{1}{2}$, we get the algebraic measure Brownian CRT, which is the unique continuum random algebraic measure tree whose i.d.d. samples span uniform binary trees. For $\alpha = 0$, we call this tree the *Kingman algebraic measure tree* as it equals in law the algebraic measure tree read off from the Kingman coalescent.

In statistical applications of phylogenies with edge lengths, it has been exploited that all sufficient information about genealogies is contained in the lengths of subtrees spanned by a finite sample. One such example is the Watterson estimator for the mutation rate of a neutral population, which counts the number of segregating sites that is often represented by the edge lengths ([Wat75, BvH99]). In this paper we want to introduce with the *sample subtree mass distribution* a similar statistics which is more suited for algebraic measure trees, for which a priori edge lengths are not defined. For that, consider for a branch point $v \in \text{br}(T)$ the three subtree components attached to v and denote for each $u \neq v$ by $\mathcal{S}_v(u)$ the subtree component that contains $u \in T$ (see (1.19) below for a precise definition). For $\underline{u} = (u_1, u_2, u_3) \in T^3$, let

$$\underline{\eta}(\underline{u}) := (\eta_i(\underline{u}))_{i=1,2,3} := (\mu(\mathcal{S}_{c(\underline{u})}(u_i)))_{i=1,2,3} \quad (1.4)$$

be the vector of the three masses of the components connected to $c(\underline{u})$. We refer to its annealed law as *sample subtree mass distribution*. It allows to distinguish between α -Ford models for different $\alpha \in [0, 1]$. For $\alpha = 1$ it can be easily read off from the associated comb tree (see Proposition 1.20). For $\alpha = \frac{1}{2}$ a more elaborate combinatorial argument shows that it is equal to the Dirichlet distribution with parameters $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ (compare [Ald94a, Theorem 2] or [LMW20, Proposition 5.2]). The case $\alpha = 0$ is treated in Proposition 1.22 where we show that the sample subtree mass distribution of the Kingman algebraic measure tree equals in distribution the symmetrization of $(B_{1,2}B_{2,2}, B_{1,2}(1 - B_{2,2}), 1 - B_{1,2})$, where $B_{1,2}$ and $B_{2,2}$ are independent beta distributed random variables with parameters $(1, 2)$ and $(2, 2)$, respectively.

An important ingredient for several algorithms that reconstruct cladograms from DNA data are Markov chains that move through a space of finite trees (see, for example, [Fel05] for a survey on Markov chain Monte Carlo algorithms in maximum likelihood tree reconstruction). The present paper has a focus on the one-parameter family of Markov chains on the space \mathfrak{C}_m of all m -cladograms which are related to the α -Ford model in the following way. Fix $\alpha \in [0, 1]$. Rather than adding new leaves, we keep the number of leaves constant by first removing a leaf picked uniformly at random and then inserting it into an edge chosen at random according to the α -Ford weights. More detailed, for each pair (x, e) consisting of a leaf and an edge (other than the edge adjacent to x) at rate 1, the Markov chain jumps from its current state t to $t^{(x,e)}$, where the latter is obtained as follows (see Figures 1.2 and 1.3):


 Figure 1.1: The 5 steps in constructing the α -Ford tree with 5 leaves.

 Figure 1.2: At rate $(1 - \alpha)m(m - 1)$ and rate $\alpha(m - 4)$ a) a leaf x and an external respectively internal edge e are picked at random, and b) the edge adjacent to x is taken away (leaving behind a branch point of degree 2).

- erase the unique edge (including the incident vertices) which connects x to the sub-tree spanned by all leaves but x ,
- split the remaining subtree at the edge e into two pieces, and
- reintroduce the above edge (including x and the branch point) at the split point.

We call this Markov chain the α -Ford chain on m -cladograms. One can easily check that the α -Ford model is the stationary distribution, and that the α -Ford chain is symmetric if and only if $\alpha = \frac{1}{2}$. In the latter case the mixing and relaxation time has been studied in detail in [Ald00, Sch01]. This case is therefore often referred to as the *Aldous move* or the *Aldous chain on cladograms*. To see why the α -Ford chain is not symmetric for general $\alpha \in [0, 1]$, notice that inserting a leaf at an edge creates a *cherry leaf* if and only if the edge was external. Therefore the time reversed α -Ford chain picks at rate $(1 - \alpha)$ a pair consisting of a cherry leaf and an edge, and at rate α a pair consisting of a non-cherry leaf and an edge at random, and inserts the picked leaf at the chosen edge. The discrepancy $\beta_\alpha^m(\mathbf{t})$ between the total backward and forward rate at the current state \mathbf{t} is a potential which links the forward α -Ford chain $X^{m,\alpha}$ and the backward α -Ford chain $Y^{m,\alpha}$ via a *Feynman-Kac duality*: for all $\mathbf{s}, \mathbf{t} \in \mathfrak{C}_m$,

$$\mathbb{P}_{\mathbf{s}}(\{X_t^{m,\alpha} = \mathbf{t}\}) = \mathbb{E}_{\mathbf{t}}\left[\mathbf{1}_{\mathbf{s}}(Y_t^{m,\alpha}) \exp\left(\int_0^t \beta_\alpha^m(Y_s^{m,\alpha}) ds\right)\right] \quad (1.5)$$

(compare Proposition 1.25).

One of the main goals of this paper is to construct the diffusion limit of the α -Ford chain as the number of leaves goes to infinity, and to provide analytic characterizations. In the case $\alpha = \frac{1}{2}$ the existence of such a diffusion limit was conjectured by David Aldous in a seminar held at the Field Institute in 1999 and had been listed on his open problem list since. Only recently such a Aldous diffusion was constructed in two independent and different approaches ([FPRW21, FPRW20, FPRWb, FPRWa] versus [LMW20]). We will here follow the approach of [LMW20] which relies on the notion of algebraic measure trees and the sample shape convergence, and generalize their construction to all $\alpha \in [0, 1]$.

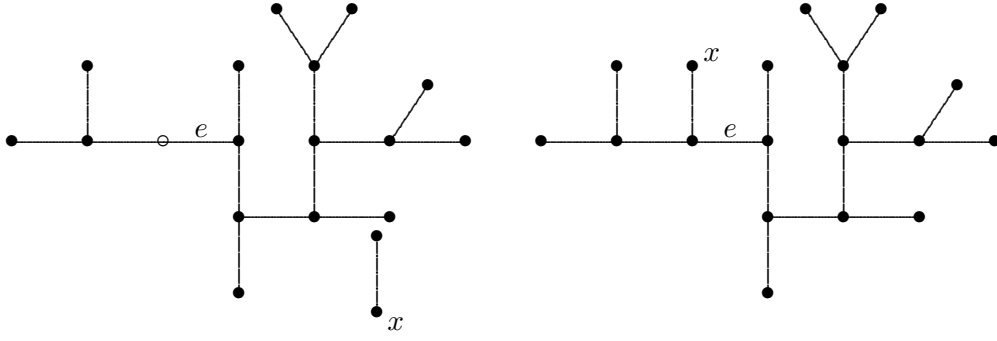


Figure 1.3: c) the two edges containing the branch point of degree 2 are identified while the edge e gets opened, and d) the free edge gets shuffled there and reattached.

Consider the operator Ω_α acting on test functions of so-called *sample shape polynomials*

$$\Phi^{m,t}(\chi) := \int_{T^m} \mu^{\otimes m}(\underline{du}) \mathbf{1}_t(\mathfrak{s}_{(T,c)}(\underline{u})), \quad (1.6)$$

with $m \in \mathbb{N}$, $\chi = (T, c, \mu) \in \mathbb{T}_2$ and $\mathfrak{t} \in \mathfrak{C}_m$, as follows:

$$\Omega_\alpha \Phi^{m,t}(\chi) := \int_{T^m} \mu^{\otimes m}(\underline{du}) \tilde{\Omega}_\alpha^m \mathbf{1}_t(\mathfrak{s}_{(T,c)}(\underline{u})), \quad (1.7)$$

where $\mathbf{1}_t$ plays the role of the test function for $\tilde{\Omega}_\alpha^m$, which denotes the generator of the α -Ford Markov chain on the space of m -cladograms (see (1.69) for a precise definition).

We state here our first main result. To do so, we identify as before an N -cladogram with an element of \mathbb{T}_2^N by forgetting the leaf labels and adding the uniform distribution on the leaves. That is, in what follows the α -Ford chain is a \mathbb{T}_2^N -valued Markov chain.

Theorem 1.1 (The well-posed martingale problem). *Let $\alpha \in [0, 1]$ and P_0 be a probability measure on $\mathbb{T}_2^{\text{cont}}$. For each $N \in \mathbb{N}$, let $X_0^N \in \mathbb{T}_2^N$ and assume that $X_0^N \rightarrow \chi$, where χ is distributed according to P_0 . Then the α -Ford chain $X^{N,\alpha}$ starting in X_0^N converges weakly in Skorokhod path space w.r.t. the sample shape convergence to a $\mathbb{T}_2^{\text{cont}}$ -valued Feller process X^α with continuous paths.*

Furthermore, X^α is the unique $\mathbb{T}_2^{\text{cont}}$ -valued Markov process $(X_t)_{t \geq 0}$ such that P_0 is the distribution of X_0 , and for all $\Phi \in \mathcal{D}(\Omega_\alpha)$, the process $M := (M_t)_{t \geq 0}$ given by

$$M_t := \Phi(X_t) - \Phi(X_0) - \int_0^t \Omega_\alpha \Phi(X_s) ds \quad (1.8)$$

is a martingale.

We refer to the process from Theorem 1.1 as α -Ford diffusion, which is justified by the first part of the theorem. We point out that the α -Ford diffusion is dual to the backward α -Ford chain through the following Feynman-Kac-duality relation: for all $m \in \mathbb{N}$ and $\mathfrak{t} \in \mathfrak{C}_m$, the α -Ford diffusion $X := ((T_t, c_t, \mu_t))_{t \geq 0}$ with initial law $P_0 = \delta_\chi$, $\chi \in \mathbb{T}_2^{\text{cont}}$, satisfies

$$\mathbb{E}_\chi^X [\Phi^{m,t}(X_t)] = \mathbb{E}_\mathfrak{t}^{Y^m} \left[\Phi^{m,Y_t^m}(\chi) \exp \left(\int_0^t \beta_\alpha^m(Y_s^m) ds \right) \right], \quad (1.9)$$

where $Y^m := (Y_t^m)_{t \geq 0}$ is the α -Ford backward chain on m -cladograms started in $Y_0^m = \mathfrak{t}$ (Proposition 1.30).

In order to provide representations of the sample subtree mass distribution for general $\alpha \in [0, 1]$, we extend this martingale problem as follows. We consider test functions of the following form, called *mass polynomials of degree 3*: for $f: [0, 1]^3 \rightarrow \mathbb{R}$ continuous,

$$\Phi^f(T, c, \mu) := \int_{T^3} f(\underline{\eta}(c(\underline{u}))) \mu^{\otimes 3}(\underline{du}), \quad (1.10)$$

where $(T, c, \mu) \in \mathbb{T}_2$. One of the main results of [LW21] is that $\Phi^f \in \mathcal{C}(\mathbb{T}_2)$.

For all $\alpha \in [0, 1]$, we extend the domain of the operator Ω_α to the set of mass polynomials Φ^f with f twice continuously differentiable on $[0, 1]$. We then put

$$\begin{aligned} \Omega_\alpha \Phi^f(\chi) = & \int_{T^3} \mu^{\otimes 3}(d\underline{u}) \left(\sum_{i,j=1}^3 \eta_i(\delta_{ij} - \eta_j) \partial_{ij}^2 f(\underline{\eta}(\underline{u})) + (2 - \alpha) \sum_{i=1}^3 (1 - 3\eta_i) \partial_i f(\underline{\eta}(\underline{u})) \right. \\ & + (2 - 3\alpha) \sum_{i=1}^3 (f(e_i) - f(\underline{\eta}(\underline{u}))) + \frac{\alpha}{2} \sum_{i \neq j=1}^3 \frac{\mathbf{1}_{\eta_i \neq 0}}{\eta_i} (f \circ \theta_{i,j}(\underline{\eta}(\underline{u})) - f(\underline{\eta}(\underline{u}))) \\ & \left. + \frac{\alpha}{2} \sum_{i \neq j=1}^3 (\mathbf{1}_{\eta_j=0} - \mathbf{1}_{\eta_i=0}) \partial_i f(\underline{\eta}(\underline{u})) \right) \end{aligned} \quad (1.11)$$

where $\theta_{i,j}: \Delta_2 \rightarrow \Delta_2$ denotes the migration operator on the two-simplex

$$\Delta_2 := \{\underline{x} \in [0, 1]^3 : x_1 + x_2 + x_3 = 1\}, \quad (1.12)$$

which sends the vector $\underline{\eta}$ to the vector where we subtract η_i from the i th entry (resulting in the entry zero) and add it to the j th entry (resulting in $\eta_i + \eta_j$), and $e_i = (\delta_{ij})_{i=1,2,3}$ is the i th unit vector. See Remark 1.35 for some heuristics of the expression (1.11).

Our second main result is the following:

Theorem 1.2 (Extended martingale problem for subtree masses). *Let $\alpha \in [0, 1]$ and $X = (X_t)_{t \geq 0}$ be the α -Ford diffusion on $\mathbb{T}_2^{\text{cont}}$. Then for all mass polynomials Φ^f with $f \in \mathcal{C}^3([0, 1])$, the process $M^f := (M_t^f)_{t \geq 0}$ given by*

$$M_t^f := \Phi^f(X_t) - \Phi^f(X_0) - \int_0^t \Omega_\alpha \Phi^f(X_s) ds \quad (1.13)$$

is a martingale.

Since the α -Ford algebraic measure tree is an invariant distribution of the α -Ford diffusion, this result also yields representations of the sample subtree mass distribution for general $\alpha \in [0, 1]$. For $\underline{k} = (k_1, k_2, k_3) \in \mathbb{N}_0^3$, define $f^{\underline{k}}: \Delta_2 \rightarrow [0, 1]$ by

$$f^{\underline{k}}(\underline{\eta}) = \eta_1^{k_1} \eta_2^{k_2} \eta_3^{k_3}. \quad (1.14)$$

Obviously, $\mathbb{E}[\Phi^{f^{(0,0,0)}}(\chi^\alpha)] = 1$ and

$$\mathbb{E}[\Phi^{f^{(1,0,0)}}(\chi^\alpha)] = \mathbb{E}[\Phi^{f^{(0,1,0)}}(\chi^\alpha)] = \mathbb{E}[\Phi^{f^{(0,0,1)}}(\chi^\alpha)] = \frac{1}{3}, \quad (1.15)$$

for all $\alpha \in [0, 1]$. Moreover, the following recursive relations hold:

Corollary 1.3 (Moments of the subtree mass distribution of the α -Ford tree). *Then, for all $\alpha \in [0, 1]$ and $\underline{k} \in \mathbb{N}_0^3$,*

$$\begin{aligned} \mathbb{E}[\Phi^{f^{\underline{k}}}(\chi^\alpha)] = & \frac{1}{(|\underline{k}| + 3)(|\underline{k}| + 2 - 3\alpha)} \left(\sum_{i=1}^3 \mathbf{1}_{\{k_i \neq 0\}} (k_i + 1)(k_i - \alpha) \mathbb{E}[\Phi^{f^{\underline{k} - e_i}}(\chi^\alpha)] \right. \\ & + (2 - 3\alpha) (\mathbf{1}_{\{k_1=k_2=0\}} + \mathbf{1}_{\{k_2=k_3=0\}} + \mathbf{1}_{\{k_3=k_1=0\}}) \\ & \left. + \frac{\alpha}{2} \sum_{i=1}^3 \mathbf{1}_{\{k_i=0\}} \sum_{j \neq i=1}^3 \sum_{l_j=1}^{k_j} \binom{k_j}{l_j} \mathbb{E}[\Phi^{f^{\underline{k} + (l_j-1)e_i - l_j e_j}}(\chi^\alpha)] \right), \end{aligned} \quad (1.16)$$

where $|\underline{k}| = k_1 + k_2 + k_3$.

Outline. The rest of the paper is organized as follows. In Section 1.2 we introduce our state space of algebraic measure trees and recall its most important properties from [LW21]. In Section 1.3 we consider the static α -Ford model and give a description of the sample subtree mass distribution in the particular case of the Kingman algebraic tree. In Section 1.4 we then consider the α -Ford chain on cladograms with a fixed number of leaves and state the Feynman-Kac duality relation to the time reversed chain. In Section 1.5 we construct the diffusion limit of the α -Ford chain as the number of leaves goes to infinity as a solution of a well-posed martingale problem. In Section 1.6 we extend this martingale problem to test functions which evaluate the sample subtree mass distribution and derive our recursive relations for the moments of the sample subtree mass distribution.

1.2 The state space: algebraic measure trees

In this section we introduce the state space. For that we rely on the framework of algebraic measure trees, which was introduced in [LW21]. All proofs can be found there.

1.2.1 Algebraic trees

In order to focus on the algebraic tree structure rather than the metric, the definition of a tree is based on axioms on the map which sends any three points to their branch point.

Definition 1.4 (Algebraic tree). An *algebraic tree* is a non-empty set T together with a symmetric map $c: T^3 \rightarrow T$ satisfying the following:

$$(2pc) \text{ For all } x_1, x_2 \in T, c(x_1, x_2, x_2) = x_2.$$

$$(3pc) \text{ For all } x_1, x_2, x_3 \in T, c(x_1, x_2, c(x_1, x_2, x_3)) = c(x_1, x_2, x_3).$$

$$(4pc) \text{ For all } x_1, x_2, x_3, x_4 \in T,$$

$$c(x_1, x_2, x_3) \in \{c(x_1, x_2, x_4), c(x_1, x_3, x_4), c(x_2, x_3, x_4)\}. \quad (1.17)$$

We call c the *branch point map*. A *tree isomorphism* between two algebraic trees (T_i, c_i) , $i = 1, 2$, is a bijective map $\phi: T_1 \rightarrow T_2$ such that for all $x_1, x_2, x_3 \in T_1$,

$$\phi(c_1(x_1, x_2, x_3)) = c_2(\phi(x_1), \phi(x_2), \phi(x_3)). \quad (1.18)$$

For each point $x \in T$, we define an equivalence relation \sim_x on $T \setminus \{x\}$ such that for all $y, z \in T \setminus \{x\}$, $y \sim_x z$ if and only if $c(x, y, z) \neq x$. For $y \in T \setminus \{x\}$, we denote by

$$\mathcal{S}_x(y) := \{z \in T \setminus \{x\} : z \sim_x y\} \quad (1.19)$$

the equivalence class of y for this equivalence relation \sim_x . We also call $\mathcal{S}_x(y)$ the *component* of $T \setminus \{x\}$ containing y . We introduce the following definitions to describe the tree structure of an algebraic tree (T, c) :

- a *subtree* of T is a set $S \subseteq T$ such that $c(S^3) = S$,
- the *degree* of $x \in T$ is the number of components of $T \setminus \{x\}$, and we write $\deg(x) := \#\{\mathcal{S}_x(y) : y \in T \setminus \{x\}\}$,
- a *leaf* is a point $u \in T$ such that $\deg(u) = 1$, and we write $\text{lf}(T)$ for the set of leaves,
- a *branch point* is a point $v \in T$ such that $\deg(v) \geq 3$, or equivalently such that $v = c(x_1, x_2, x_3)$ for some $x_1, x_2, x_3 \in T \setminus \{v\}$, and we denote by $\text{br}(T)$ the set of branch points,
- for $x, y \in T$, we define the interval $[x, y]$ as

$$[x, y] := \{z \in T : c(x, y, z) = z\}, \quad (1.20)$$

- and we say that $\{x, y\}$ is an *edge* if $x \neq y$ and $[x, y] = \{x, y\}$.

There is a natural Hausdorff topology on a given algebraic tree, namely the topology generated by the set of all components $\mathcal{S}_x(y)$ with $x \neq y$, $x, y \in T$. We say that an algebraic tree (T, c) is *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 $\mathcal{B}(T, c)$, which allows to sample leaves from the tree.

Definition 1.5 (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 $\mathcal{B}(T, c)$.

We say that two algebraic measure trees (T_i, c_i, μ_i) , $i = 1, 2$ are *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

$$\mathbb{T} := \text{set of equivalence classes of algebraic measure trees.} \quad (1.21)$$

With an abuse of notation, we will write $\chi = (T, c, \mu)$ for the algebraic tree as well as the equivalence class.

A first way to equip \mathbb{T} with a topology is by associating an algebraic measure tree with a metric measure tree and define the convergence of algebraic measure trees in \mathbb{T} as the Gromov-weak convergence (compare, for example, [GPW09]) of these associated metric measure trees. We first need to define the metric measure tree associated to an algebraic measure tree.

Definition 1.6 (branch point distribution). For an algebraic measure tree $\chi = (T, c, \mu)$, the *branch point distribution* on T is defined as

$$\nu_{(T, c, \mu)} := \mu^{\otimes 3} \circ c^{-1}, \quad (1.22)$$

and we associate χ with the metric measure tree (T, r_μ, μ) , where we put for $x, y \in T$,

$$r_\mu(x, y) := \nu_\chi([x, y]) - \frac{1}{2}\nu_\chi(\{x\}) - \frac{1}{2}\nu_\chi(\{y\}). \quad (1.23)$$

The choice of the metric r_μ can be understood as follows: two points are close if the mass branching off the line segment connecting them is small rather than if the length of this line segment is small. We then say that a sequence of algebraic measure trees converges in the *branch point distribution distance Gromov-weak topology* if the associated (through r_μ) sequence of metric measure trees converges Gromov-weakly.

Because cladograms are by definition binary, it is enough for the purpose of the present paper to consider the subspace of \mathbb{T} consisting of binary trees. More precisely, we consider the subspace of binary algebraic measure trees with the property that the measure has atoms only (if at all) on the leaves on the tree:

$$\mathbb{T}_2 = \{(T, c, \mu) \in \mathbb{T} : \deg(v) \leq 3 \ \forall v \in T, \text{at}(\mu) \subseteq \text{lf}(T)\}, \quad (1.24)$$

where we write $\text{at}(\mu)$ for the set of atoms of μ . Under this extra condition, the notion of Gromov-weak convergence with respect to r_ν is equivalent to a more combinatorial notion of convergence. In contrast to the Gromov-weak convergence which relies on sample distance matrices, this combinatorial notion make use of sample shapes. To introduce the latter, we first extend our previous definition of cladograms as follows.

Definition 1.7 (m -cladogram). For $m \in \mathbb{N}$, an m -labelled *cladogram* is a binary, finite tree $C = (C, c)$ consisting only of leaves and branch points together with a surjective labelling map $\zeta : \{1, \dots, m\} \rightarrow \text{lf}(C)$. An m -cladogram (C, c, ζ) is an m -labelled cladogram such that ζ is also 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 $\zeta_2 = \phi \circ \zeta_1$. We then write

$$\overline{\mathfrak{C}}_m := \{\text{isomorphism classes of } m\text{-labelled cladograms}\} \quad (1.25)$$

and

$$\mathfrak{C}_m := \{(C, c, \zeta) \in \overline{\mathfrak{C}}_m : \zeta \text{ injective}\}. \quad (1.26)$$

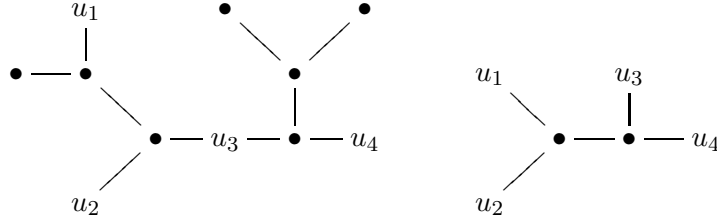


Figure 1.4: A tree T and the shape $\mathfrak{s}_{(T,c)}(u_1, u_2, u_3, u_4)$. The cladogram is not isomorphic to the subtree $c(\{u_1, u_2, u_3, u_4\}^3)$ because $u_3 \in [u_1, u_4]$.

Note that an m -cladogram has exactly m leaves (and $m-2$ branch points). An m -labelled cladogram can have less than m leaves (and $m-2$ branch points) if a leaf has multiple labels.

We next define the shape function, which allows to associate m ordered distinct leaves with a unique m -cladogram.

Definition 1.8 (Shape function). For a binary algebraic tree (T, c) , $m \in \mathbb{N}$, and $u_1, \dots, u_m \in T \setminus \text{br}(T)$, there exists a unique (up to isomorphism) m -labelled cladogram

$$\mathfrak{s}_{(T,c)}(u_1, \dots, u_m) = (C, c_C, \zeta) \quad (1.27)$$

with $\text{lf}(C) = \{u_1, \dots, u_m\}$ and $\zeta(i) = u_i$, such that the identity on $\text{lf}(C)$ extends to a tree homomorphism π from C onto $c(\{u_1, \dots, u_m\}^3)$, i.e. for all $i, j, k = 1, \dots, m$,

$$\pi(c_C(u_i, u_j, u_k)) = c(u_i, u_j, u_k). \quad (1.28)$$

We will refer to $\mathfrak{s}_{(T,c)}(u_1, \dots, u_m) \in \overline{\mathfrak{C}}_m$ as the *shape* of u_1, \dots, u_m in (T, c) (compare with Figure 1.4).

We are now in a position to define the sample shape convergence.

Definition 1.9 (Sample shape convergence). A sequence $(\chi_N)_{N \in \mathbb{N}}$ of binary algebraic measure trees (T_N, c_N, μ_N) converges in sample shape to the algebraic measure tree (T, c, μ) if and only if for U_1^N, U_2^N, \dots i.i.d. of law μ_N , and U_1, U_2, \dots i.i.d. of law μ , for all $m \in \mathbb{N}$,

$$\mathfrak{s}_{(T_N, c_N)}(U_1^N, \dots, U_m^N) \xrightarrow[N \rightarrow \infty]{} \mathfrak{s}_{(T, c)}(U_1, \dots, U_m). \quad (1.29)$$

Since for any $m \in \mathbb{N}$ the space of m -cladograms is finite, we have the following equivalence ([LW21, Corollary 5.21]).

Proposition 1.10 (Convergence determining class of functions). Consider $\chi = (T, c, \mu)$ and $(\chi_N = (T_N, c_N, \mu_N))_{N \in \mathbb{N}}$ in \mathbb{T}_2 . Then $(\chi_N)_{N \in \mathbb{N}}$ converges to χ w.r.t. the sample shape convergence if and only if for all $m \in \mathbb{N}$ and $\mathfrak{t} \in \overline{\mathfrak{C}}_m$,

$$\mu_N^{\otimes m}(\{(u_1, \dots, u_m) : \mathfrak{s}_{(T_N, c_N)}(\underline{u}) = \mathfrak{t}\}) \xrightarrow[N \rightarrow \infty]{} \mu^{\otimes m}(\{(u_1, \dots, u_m) : \mathfrak{s}_{(T, c)}(\underline{u}) = \mathfrak{t}\}). \quad (1.30)$$

In what follows we will consider α -Ford trees with N leaves as random algebraic measure tree which belong to the following subspace:

$$\mathbb{T}_2^N := \{(T, c, \mu) \in \mathbb{T}_2 : \#\text{lf}(T) = N \text{ and } \mu = \frac{1}{N} \sum_{u \in \text{lf}(T)} \delta_u\}, \quad (1.31)$$

and let then N tend to infinity. The next proposition claims that the limit points are elements in the following closed subspace ([LW21, Corollary 4.9]):

$$\mathbb{T}_2^{\text{cont}} := \{(T, c, \mu) \in \mathbb{T}_2 : \text{at}(\mu) = \emptyset\}. \quad (1.32)$$

Proposition 1.11 (Approximations with $\mathbb{T}_2^{\text{cont}}$). *Let $\chi \in \mathbb{T}_2$. Then $\chi \in \mathbb{T}_2^{\text{cont}}$ if and only if there exists for each $N \in \mathbb{N}$ a $\chi_N \in \mathbb{T}_2^N$ such that $\chi_N \rightarrow \chi$ in one and thus all of the equivalent notions of convergence on \mathbb{T}_2 given above.*

Finally, the following result will be very helpful in constructing the diffusions in Section 1.5 ([LW21, Theorem 3, Corollary 5.19]):

Proposition 1.12 (Compactness and metrizability). *\mathbb{T}_2 and $\mathbb{T}_2^{\text{cont}}$ are compact, metrizable spaces.*

1.2.2 Rooted algebraic trees

In many applications rooted trees are of interest. Even though we mainly work with unrooted trees in this paper, we want to shortly explain how algebraic trees are extended to rooted algebraic trees.

Definition 1.13 (Rooted and partially ordered algebraic tree). A *partially ordered algebraic tree* is a non-empty set T together with a symmetric map $c_\wedge : T \times T \rightarrow T$ such that:

- (M1) For all $x \in T$, $c_\wedge(x, x) = x$.
- (M2) For all $x_1, x_2, x_3 \in T$, $c_\wedge(x_1, c_\wedge(x_2, x_3)) = c_\wedge(c_\wedge(x_1, x_2), x_3)$.
- (M3) For all $x_1, x_2, x_3 \in T$, $\#\{c_\wedge(x_1, x_2), c_\wedge(x_1, x_3), c_\wedge(x_2, x_3)\} \leq 2$ and if $c_\wedge(x_1, x_2) = c_\wedge(x_1, x_3)$, then

$$c_\wedge(x_1, x_2) = c_\wedge(c_\wedge(x_1, x_2), c_\wedge(x_2, x_3)). \quad (1.33)$$

We refer to c_\wedge as the *minimum map*.

A *rooted algebraic tree* (T, c_\wedge) is a partially ordered algebraic tree for which there exists a point $\rho \in T$ with $c_\wedge(\rho, x) = \rho$ for all $x \in T$. We will refer to (this unique) ρ as the *root* of (T, c_\wedge) .

Remark 1.14 ((M1) and (M2) define a partial order). Let (T, c_\wedge) be a partially ordered algebraic tree. In what follows, we write for $x, y \in T$, $x \leq y$ if and only if $x = c_\wedge(x, y)$ respectively, $x < y$ if and only if $x = c_\wedge(x, y) \neq y$. Notice that the first two conditions (M1) and (M2) ensure that \leq defines a partial order relation. Indeed, reflexivity follows from (M1), antisymmetry follows from the fact that c_\wedge is a symmetric map and transitivity follows from (M2), i.e., if $x, y, z \in T$ are such that $x \leq y$ and $y \leq z$, or equivalently, if $x = c_\wedge(x, y)$ and $y = c_\wedge(y, z)$, then $x = c_\wedge(x, y) = c_\wedge(x, c_\wedge(y, z)) = c_\wedge(c_\wedge(x, y), z) = c_\wedge(x, z)$.

The following lemma gives key arguments for Proposition 1.16.

Lemma 1.15. *Let (T, c_\wedge) be a partially ordered, algebraic tree.*

- (i) For $x, y \in T$,

$$c_\wedge(x, y) \leq x. \quad (1.34)$$

- (ii) Then for all $x, y, z \in T$,

$$c_\wedge(x, y) \leq c_\wedge(x, z) \quad \text{or} \quad c_\wedge(x, y) > c_\wedge(x, z). \quad (1.35)$$

In particular, $\max\{c_\wedge(x, y), c_\wedge(x, z), c_\wedge(y, z)\}$ is well-defined.

Proof. “(i)” By (M2) and (M1),

$$c_\wedge(x, c_\wedge(x, y)) = c_\wedge(c_\wedge(x, x), y) = c_\wedge(x, y). \quad (1.36)$$

“(ii)” If $x, y, z \in T$ are such that $c_\wedge(x, y) = c_\wedge(x, z)$, the claim is trivial. By (M3) we can therefore assume w.l.o.g. that $c_\wedge(x, y) = c_\wedge(y, z)$. Then using (M2) together with part (i), $c_\wedge(c_\wedge(x, y), c_\wedge(x, z)) = c_\wedge(y, c_\wedge(x, c_\wedge(x, z))) = c_\wedge(y, c_\wedge(x, z)) = c_\wedge(x, y)$. \square

When we add the third condition (M3) to the partially ordered set defined by (M1) and (M2), we ensure that there are no loops, so that we can relate algebraic trees to rooted, algebraic trees.

Proposition 1.16 (Rooted versus unrooted algebraic trees). *Let $T \neq \emptyset$.*

(i) *If (T, c) is an algebraic tree, $\rho \in T$, and $c_\wedge : T^2 \rightarrow T$ the symmetric map defined as*

$$c_\wedge(x, y) := c(x, y, \rho), \quad x, y \in T, \quad (1.37)$$

then (T, c_\wedge, ρ) is a rooted algebraic tree.

(ii) *If (T, c_\wedge) is a partially ordered, algebraic tree and $c : T^3 \rightarrow T$ the symmetric map defined as*

$$c(x, y, z) := \max\{c_\wedge(x, y), c_\wedge(x, z), c_\wedge(y, z)\}, \quad x, y, z \in T, \quad (1.38)$$

then (T, c) is an algebraic tree.

Proof. “(i)” Let (T, c) be an algebraic tree, $\rho \in T$, and define c_\wedge as in (1.37). (M1) follows from (2pc) as $c_\wedge(x, x) = c(\rho, x, x) = x$ for all $x \in T$. To verify (M2), let $x, y, z \in T$. We need to show that $c(\rho, x, c(\rho, y, z)) = c(\rho, c(\rho, x, y), z)$. By (4pc), three cases are possible: a) if $c(\rho, y, z) = c(\rho, x, y)$, then by (3pc), $c(\rho, x, c(\rho, y, z)) = c(\rho, x, y) = c(\rho, c(\rho, x, y), z)$, b) if $c(\rho, y, z) = c(\rho, x, z)$, then by (3pc) together with [LW21, Lemma 2.2], $c(\rho, x, c(\rho, y, z)) = c(\rho, x, z) = c(\rho, c(\rho, x, y), z)$ or c) if $c(\rho, x, y) = c(\rho, x, z)$, we use the same line of arguments as in case b). Finally, to see (M3), let $x, y, z \in T$. By [LW21, Lemma 2.2], $\{c(x, y, z), c(\rho, x, y), c(\rho, x, z), c(\rho, y, z)\}$ consists of two pairs of equal points or one point only. Moreover, if $c_\wedge(x, y) = c_\wedge(x, z)$, that is, $c(\rho, x, y) = c(\rho, x, z)$, then by [LW21, Lemma 2.2], $c(\rho, x, y) = c(\rho, x, c(\rho, y, z))$ and with (3pc), $c(\rho, c(\rho, x, y), c(\rho, y, z)) = c(\rho, c(\rho, x, c(\rho, y, z)), c(\rho, y, z)) = c(\rho, x, c(\rho, y, z)) = c(\rho, x, y)$.

“(ii)” Let (T, c_\wedge) be a partially ordered, algebraic tree, and define c as in (1.37). To see (2pc), let $x_1, x_2, x_3 \in T$ such that $x_2 = x_3$. Then $c_\wedge(x_1, x_2) = c_\wedge(x_1, x_3)$, and therefore $c(x_1, x_2, x_3) = c_\wedge(x_2, x_3) = x_2$, by (M3) and (M1).

As for (3pc), let $x_1, x_2, x_3 \in T$. We distinguish three cases: a) If $c_\wedge(x_1, x_3) = c_\wedge(x_2, x_3)$, we have with (1.34), $c_\wedge(x_1, c_\wedge(x_1, x_2)) = c_\wedge(x_1, x_2) = c_\wedge(c_\wedge(x_1, x_2), x_2)$, so that, with (1.37),

$$c(x_1, x_2, c(x_1, x_2, x_3)) = c(x_1, x_2, c_\wedge(x_1, x_2)) = c_\wedge(x_1, x_2) = c(x_1, x_2, x_3). \quad (1.39)$$

b) If $c_\wedge(x_1, x_2) = c_\wedge(x_1, x_3)$, then using (M2) and (1.34), we have that

$$c_\wedge(x_1, c_\wedge(x_2, x_3)) = c_\wedge(c_\wedge(x_1, x_2), x_3) = c_\wedge(c_\wedge(x_1, x_3), x_3) = c_\wedge(x_1, x_3) = c_\wedge(x_1, x_2) \quad (1.40)$$

and therefore, with (1.37) and (1.34),

$$c(x_1, x_2, c(x_1, x_2, x_3)) = c(x_1, x_2, c_\wedge(x_2, x_3)) = c_\wedge(x_2, c_\wedge(x_2, x_3)) = c_\wedge(x_2, x_3) = c(x_1, x_2, x_3). \quad (1.41)$$

c) If $c_\wedge(x_1, x_2) = c_\wedge(x_2, x_3)$, we use the same line of arguments as in case b).

Finally for (4pc), let $x_1, \dots, x_4 \in T$ and assume w.l.o.g. that $c_\wedge(x_1, x_2) = c_\wedge(x_1, x_3)$ and thus that

$$c(x_1, x_2, x_3) = c_\wedge(x_2, x_3) \geq c_\wedge(x_1, x_2) = c_\wedge(x_1, x_3). \quad (1.42)$$

We show that $c_\wedge(x_2, x_3) \in \{c(x_1, x_2, x_4), c(x_1, x_3, x_4), c(x_2, x_3, x_4)\}$. If $c(x_2, x_3, x_4) = c_\wedge(x_2, x_3)$, this obviously holds. Assume therefore that $c(x_2, x_3, x_4) \neq c_\wedge(x_2, x_3)$. Then we can assume w.l.o.g. that $c_\wedge(x_2, x_3) = c_\wedge(x_2, x_4) < c_\wedge(x_3, x_4)$. By (1.42), $c_\wedge(x_3, x_4) > c_\wedge(x_2, x_3) \geq c_\wedge(x_1, x_3)$, so that with (M3) applied to $\{x_1, x_3, x_4\}$, $c_\wedge(x_1, x_4) = c_\wedge(x_1, x_3) = c_\wedge(x_1, x_2)$ and with (1.37), $c(x_1, x_2, x_4) = c_\wedge(x_2, x_4) = c_\wedge(x_2, x_3)$ and we are done. \square

1.3 Continuum limits of the α -Ford trees

Recall from (1.31) and from (1.32) the spaces \mathbb{T}_2^N and $\mathbb{T}_2^{\text{cont}}$ of all binary algebraic measure trees with N leaves and equipped with the uniform distribution on the set of leaves respectively of all continuum binary algebraic measure trees. In this section we consider the α -Ford trees with N leaves as elements in \mathbb{T}_2^N , and establish their continuum limits as N tends to infinity.

Proposition 1.17 (Continuum α -Ford trees). *Fix $\alpha \in [0, 1]$, and let for $N \in \mathbb{N}$, χ_N^α denote the α -Ford tree with N leaves. Then there exists a random continuum binary measure tree $\chi^\alpha \in \mathbb{T}_2^{\text{cont}}$ such that*

$$\chi_N^\alpha \xrightarrow[N \rightarrow \infty]{} \chi^\alpha, \quad (1.43)$$

where \Rightarrow stands for weak convergence on \mathbb{T}_2 equipped with the sample shape convergence.

Proof. By Proposition 1.12, the space \mathbb{T}_2 of binary algebraic measure trees with atoms only on the set of leaves is compact. Therefore the sequence $\{\chi_N^\alpha; N \in \mathbb{N}\}$ is clearly tight. Moreover, all limit points are elements of $\mathbb{T}_2^{\text{cont}}$ by Proposition 1.11.

For uniqueness of the limit we show that the family $\{\chi_N^\alpha = (T_N^\alpha, c_N^\alpha); n \in \mathbb{N}\}$ is *sampling consistent*, i.e., for all $N \in \mathbb{N}$ and $m \leq N$, the shape $\mathfrak{s}_{(T_N^\alpha, c_N^\alpha)}(U_1, \dots, U_m)$ of the subtree spanned by m leaves (U_1, \dots, U_m) sampled at random from χ_N^α has the distribution of the α -Ford tree with m -leaves. Indeed, it has been shown in [For, Proposition 42] through a combinatorial argument that the α -Ford models are *deletion stable*, which means that the cladogram obtained by removing the leaf with label m from the α -Ford tree with m leaves has the distribution of the α -Ford tree with $m - 1$ leaves. Furthermore, the last step of the construction of the α -Ford cladogram assures that we have exchangeability, i.e. the resulting distribution on cladograms is symmetric under permutation of leaf labels. This implies the claimed sampling consistency. \square

Definition 1.18 (Continuum α -Ford tree). The α -Ford algebraic measure tree χ^α is the unique limit in $\mathbb{T}_2^{\text{cont}}$ of the sequence $(\chi_N^\alpha)_N$, where χ_N^α is the random algebraic measure tree in \mathbb{T}_2^N obtained from the random N -cladogram distributed according to the α -Ford model.

Remark 1.19 (Rooted continuum α -Ford tree). In [For], the alpha model is introduced as a random rooted m -cladogram, in which the edge adjacent to the root is an internal edge and thus has weight α . Once more, we can consider rooted α -Ford measure trees as a random rooted algebraic tree with a fixed finite number of leaves equipped with the uniform distribution on the set of leaves (not including the root leaf). One can show by similar arguments as in [LW21] that the space of rooted binary algebraic measure trees equipped with the rooted sample shape topology is compact. As the family of α -Ford rooted trees indexed by the number of leaves is sampling consistent, Proposition 1.17 still holds in the case of rooted algebraic measure trees. For each $\alpha \in [0, 1]$, we refer to the continuum limit $\chi^{\alpha, \rho}$ as the *rooted α -Ford algebraic measure tree*.

Note that by Definition 1.18, the distribution of the shape spanned by a sample of size m from the α -Ford algebraic measure tree equals the α -Ford model on m -cladograms. In this paper we want to promote with the sample subtree mass distribution a further statistics for testing hypotheses. More precisely, recall the definition of the components $\mathcal{S}_v(u)$, $u, v \in T$, from (1.19) and for $\underline{u} = (u_1, u_2, u_3) \in T^3$, denote by $\underline{\eta}(\underline{u})$ the vector of the μ -masses of the components of $T \setminus \{c(\underline{u})\}$, that is

$$\underline{\eta}(\underline{u}) = (\eta_i(\underline{u}))_{i=1,2,3} = (\mu(\mathcal{S}_{c(\underline{u})}(u_i)))_{i=1,2,3} \in \Delta_2, \quad (1.44)$$

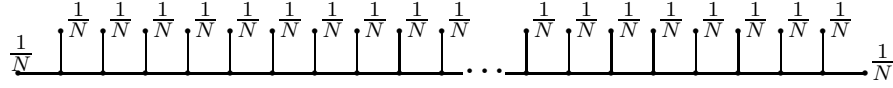
where Δ_k stands for the k -simplex for $k \in \mathbb{N}$, i.e.,

$$\Delta_k := \{\underline{x} \in [0, 1]^{k+1} : x_1 + \dots + x_{k+1} = 1\}. \quad (1.45)$$

We here want to point out that in the following three cases we know the shape distribution as well as the subtree mass distribution explicitly:

- $\alpha = 0$: The Kingman algebraic measure tree.
- $\alpha = \frac{1}{2}$: The algebraic measure Brownian CRT.
- $\alpha = 1$: The comb algebraic measure tree.

Before we continue in the next sections developing analytical tools for a description of subtree mass distribution for a general $\alpha \in [0, 1]$, we close this section by summarizing what can be obtained for the particular cases by means of known probabilistic representations.


 Figure 1.5: The comb tree with N leaves.

For $k \in \mathbb{N}$ and $\gamma_1, \gamma_2, \dots, \gamma_k > 0$ denote by $\text{Dir}(\gamma_1, \dots, \gamma_k)$ the *Dirichlet distribution*, i.e., the probability distribution on Δ_{k-1} with density

$$f_{(\gamma_1, \dots, \gamma_k)}(x_1, \dots, x_k) := \frac{\Gamma(\gamma_1 + \dots + \gamma_k)}{\Gamma(\gamma_1) \cdot \dots \cdot \Gamma(\gamma_k)} x_1^{\gamma_1-1} x_2^{\gamma_2-1} \cdot \dots \cdot x_k^{\gamma_k-1}. \quad (1.46)$$

Recall that for $k = 2$ and $\gamma_1, \gamma_2 > 0$, the $\text{Dir}(\gamma_1, \gamma_2)$ is referred to as the beta distribution with parameters γ_1 and γ_2 .

1.3.1 The comb algebraic measure tree

We will start with the simplest case $\alpha = 1$. Let for each $N \in \mathbb{N}$, χ_N^1 be (the equivalence class of) the algebraic comb tree with N leaves and uniform distribution on the leaves (see Figure 1.5 for an illustration). Let further χ^1 be (the equivalence class of) the unit interval with the uniform distribution. Clearly, if a sample of size m from χ_N^1 consists of m distinct points, then it spans a shape that equals the shape spanned by a sample of size m from χ^1 . As the probability that a sample of size m from χ_N^1 has double points converges to 0 as $N \rightarrow \infty$, $\chi_N^1 \rightarrow \chi^1$ in sample shape convergence.

Notice that χ^1 is deterministic. We can therefore easily describe the subtree mass distribution.

Proposition 1.20 (Subtree mass distribution in the comb tree). *Let $\text{Beta}(2, 2)$ be the beta distribution on $[0, 1]$. Then (almost surely), for all $f: \Delta_2 \rightarrow \mathbb{R}$ continuous bounded,*

$$\int_{(T_{\chi^1})^3} (\mu_{\chi^1})^{\otimes 3}(\underline{du}) f(\underline{\eta}_{\chi^1}(\underline{u})) = \frac{1}{6} \sum_{\pi \in \mathcal{S}_3} \int_{[0,1]} f \circ \pi^*(x, 1-x, 0) \text{Beta}(2, 2)(dx), \quad (1.47)$$

where \mathcal{S}_3 is the set of permutations of $\{1, 2, 3\}$, and for $\pi \in \mathcal{S}_3$, $\pi^*: \Delta_2 \rightarrow \Delta_2$ is the induced map $\pi^*(\underline{x}) = (x_{\pi(1)}, x_{\pi(2)}, x_{\pi(3)})$.

Proof. Let U_1, U_2, U_3 be independent random variables uniformly distributed on $[0, 1]$. Denote by $U_{(1)}, U_{(2)}, U_{(3)}$ the increasingly ordered random variables. It is well-known that $U_{(2)}$ has law $\text{Beta}(2, 2)$. Now, if we sample three points from χ^1 , the points correspond to U_1, U_2, U_3 and their branch point to $U_{(2)}$. The subtree masses around this branch point are $U_{(2)}$, $1 - U_{(2)}$ and 0, and the claim follows by exchangeability. \square

1.3.2 The algebraic measure Brownian CRT

The case $\alpha = \frac{1}{2}$ is well-studied. Consider for each $N \in \mathbb{N}$ the equivalence class $\chi_N^{\frac{1}{2}}$ of the $\frac{1}{2}$ -Ford tree with N leaves and uniform distribution on the leaves. Clearly, if a sample of size m from $\chi_N^{\frac{1}{2}}$ consists of m distinct points, then it spans a shape which is uniformly distributed among all possible shapes. Once more, as the probability that a sample of size m from $\chi_N^{\frac{1}{2}}$ has double points converges to 0 as $N \rightarrow \infty$, the m -sample shape distribution of $\chi^{\frac{1}{2}}$ is also uniform. It therefore follows from [Ald93, Corollary 22] that $\chi^{\frac{1}{2}}$ is (up to equivalence) equal to the algebraic tree read off from the Brownian CRT.

Moreover, it is known that the sample subtree masses of the algebraic measure Brownian CRT is Dirichlet distributed (see [Ald94a, Proposition 1], or [LMW20, Proposition 5.2] which gives a more

general result for a sample of size m , proved with combinatorial arguments). More precisely, we have, for all $f: \Delta_2 \rightarrow \mathbb{R}$ continuous bounded,

$$\mathbb{E} \left[\int_{(T_{\chi^{\frac{1}{2}}})^3} (\mu_{\chi^{\frac{1}{2}}})^{\otimes 3} (d\underline{u}) f \left(\eta_{\chi^{\frac{1}{2}}}(\underline{u}) \right) \right] = \int_{\Delta_2} f(\underline{x}) \text{Dir} \left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2} \right) (d\underline{x}). \quad (1.48)$$

1.3.3 The Kingman algebraic measure tree

The case $\alpha = 0$ can be associated with the Kingman coalescent tree. More precisely, let $K = (K_t)_{t \geq 0}$ be the unit rate Kingman coalescent which takes values in the space of partitions of the set \mathbb{N} and starts in the finest partition $\{\{i\}; i \in \mathbb{N}\}$. Given K , we define the *Kingman rooted algebraic tree* as the random rooted algebraic tree (T, c_\wedge, ρ) with the vertex set

$$T := \{\varpi_\rho\} \uplus \bigcup_{t \geq 0} \bigcup_{\varpi \in K_t} \varpi, \quad (1.49)$$

with $\varpi_\rho = \{1, \dots, m\} \uplus \{\rho\}$ for a point $\rho \notin \mathbb{N}$ and the minimal map c_ρ which sends two elements $\varpi, \varpi' \in T$ to the smallest $\tilde{\varpi} \in T$ which contains both ϖ and ϖ' , i.e.,

$$c_\wedge(\varpi, \varpi') := \bigcap_{\tilde{\varpi} \in T: \varpi, \varpi' \subseteq \tilde{\varpi}} \tilde{\varpi}. \quad (1.50)$$

Further, we define the *Kingman algebraic tree* as the random algebraic space (T, c) obtained from the rooted Kingman algebraic tree $(T, c_\wedge, \varpi_\rho)$ as in (1.38). We also define the *rooted Kingman algebraic measure tree* and the *Kingman algebraic measure tree* as the weak limit as $m \rightarrow \infty$ (with respect to the sample shape convergence) of the rooted algebraic measure Kingman tree respectively the algebraic measure Kingman tree which are in addition equipped with

$$\mu^m := \frac{1}{m} \sum_{i=1}^m \delta_{\{i\}}. \quad (1.51)$$

Proposition 1.21. *The rooted Kingman algebraic measure tree and the Kingman algebraic measure tree equal in distribution with the rooted $\alpha = 0$ -Ford algebraic measure tree respectively the $\alpha = 0$ -Ford algebraic measure tree.*

Even if this fact is known, a proof does not seem easy to spot in the literature, and we therefore give its proof.

Proof. We first consider the rooted finite models. Since there is only one equivalence class in \mathbb{T}_2^N for $N = 1, 2, 3$, the (rooted) Kingman and $(\alpha = 0)$ -Ford algebraic measure trees with $N = 1, 2, 3$ leaves have the same distribution.

Fix now $N \in \mathbb{N}$. Given the Kingman N -coalescent $(K^N)_{t \geq 0}$ and a time $s \geq 0$, we can define as above the random rooted algebraic tree (T_s^N, c_\wedge, ρ) with the vertex set

$$T_s^N := \{\varpi_\rho\} \uplus \bigcup_{t \geq s} \bigcup_{\varpi \in K_t^N} \varpi, \quad (1.52)$$

with $\varpi_\rho = \{1, \dots, N\} \uplus \{\rho\}$ for a point $\rho \notin \mathbb{N}$ and the minimal map c_ρ which sends two elements $\varpi, \varpi' \in T$ to the smallest $\tilde{\varpi} \in T$ which contains both ϖ and ϖ' , i.e.,

$$c_{\wedge, s}^N(\varpi, \varpi') := \bigcap_{\tilde{\varpi} \in T_s^N: \varpi, \varpi' \subseteq \tilde{\varpi}} \tilde{\varpi}. \quad (1.53)$$

We further equip T_s^N with the uniform measure $\mu_{K_s^N}$ on the blocks of the partition K_s^N .

Let τ be the random time of the first coalescence event in the Kingman N -coalescent. Thus, $(T_\tau^N, c_{\wedge, \tau}^N)$ is the algebraic tree obtained from the N -coalescent where we removed the two leaves

corresponding to the two blocks that merged first (together with the adjacent edges). Then, by a property of the coalescent (see [Kin82]), $(T_\tau^N, c_{\wedge, \tau}^N, \rho, \mu_{K_\tau^N})$ and $(T_0^{N-1}, c_{\wedge, 0}^{N-1}, \rho, \mu_{K_0^{N-1}})$ have the same distribution (up to equivalence), and $(T_\tau^N, c_{\wedge, \tau}^N, \rho, \mu_{K_\tau^N})$ is independent (up to isomorphism) of the first coalescence. This means that in (T_τ^N, c_\wedge) , the leaf corresponding to the only block of K_τ^N of size two is uniformly distributed among all the leaves.

On the other hand, in the construction of the $(\alpha = 0)$ -Ford model, the N -th leaf is inserted into the middle of an edge chosen uniformly at random among the $N - 1$ external edges. It follows by induction that the rooted Kingman and $(\alpha = 0)$ -Ford algebraic measure trees with N leaves have same distribution. Taking the limit when N goes to infinity yields the result. \square

However, the following result on the subtree mass distribution is new.

Proposition 1.22 (Subtree mass distribution in the Kingman algebraic measure tree). *Let $B_{1,2}$ and $B_{2,2}$ be two independent beta random variables, such that $B_{1,2}$ has law $\text{Beta}(1, 2)$ and $B_{2,2}$ has law $\text{Beta}(2, 2)$. Then for all $f: \Delta_2 \rightarrow \mathbb{R}$ continuous bounded,*

$$\mathbb{E} \left[\int_{(T_{\chi^0})^3} (\mu_{\chi^0})^{\otimes 3} (d\underline{u}) f(\underline{\eta}_{\chi^0}(\underline{u})) \right] = \frac{1}{6} \sum_{\pi \in \mathcal{S}_3} \mathbb{E} [f \circ \pi^* (1 - B_{1,2}, B_{1,2}B_{2,2}, B_{1,2}(1 - B_{2,2}))], \quad (1.54)$$

where \mathcal{S}_3 is the set of permutations of $\{1, 2, 3\}$, and for $\pi \in \mathcal{S}_3$, $\pi^*: \Delta_2 \rightarrow \Delta_2$ is the induced map $\pi^*(\underline{x}) = (x_{\pi(1)}, x_{\pi(2)}, x_{\pi(3)})$.

This result is obtained by symmetrizing the subtree mass distribution in the rooted Kingman algebraic measure tree, which is the rooted α -Ford algebraic measure tree with $\alpha = 0$. In the case of rooted algebraic trees, one of the three components contains the root ρ and we need to distinguish it in the vector of subtree masses. For a branch point $v \in \text{br}(T)$, we write $(\eta_\rho(v), \eta_1(v), \eta_2(v))$ for the masses of the components of $T \setminus \{v\}$, with $\eta_\rho(v)$ being for the component which contains the root.

Proposition 1.23 (Subtree mass distribution in the rooted Kingman algebraic measure tree). *Let $B_{1,2}$ and $B_{2,2}$ be two independent beta random variables, such that $B_{1,2}$ has law $\text{Beta}(1, 2)$ and $B_{2,2}$ has law $\text{Beta}(2, 2)$. For all $f: \Delta_2 \rightarrow \mathbb{R}$ continuous and bounded,*

$$\mathbb{E} \left[\int_{(T_{\chi^0, \rho})^3} (\mu_{\chi^0, \rho})^{\otimes 3} (d\underline{u}) f(\underline{\eta}_{\chi^0, \rho}(\underline{u})) \right] = \mathbb{E} [f(1 - B_{1,2}, B_{1,2}B_{2,2}, B_{1,2}(1 - B_{2,2}))]. \quad (1.55)$$

We will need the following lemma for the proof.

Lemma 1.24. *Let $\mathbb{P}^{0, \rho}$ note the law of the rooted Kingman algebraic measure tree. $\mathbb{P}^{0, \rho}$ -almost surely, for all $v \in \text{br}(T)$, we have*

$$\underline{\eta}(v) = (\eta_\rho(v), (1 - \eta_\rho(v))U_v, (1 - \eta_\rho(v))(1 - U_v)), \quad (1.56)$$

where U_v is a random variable uniformly distributed on $(0, 1)$, independent of $\eta_\rho(v)$.

Proof. Given a rooted algebraic tree (T, c_\wedge, ρ) denote for each $v \in T$ by $S^v := \{z \in T : c_\wedge(z, v) = v\}$ the subtree “above” v . Notice that the finite α -Ford tree is *Markovian self-similar* (see [For, Definition 24, Lemma 27]), and that this implies that there exists a function $q(\cdot, \cdot)$ called *conditional split distribution* such that for all $1 \leq m \leq n - 1$, $q_\alpha(m, n - m)$ is the probability that if there are n leaves in the subtree above a fixed vertex v , m leaves are put on the left side of v and $n - m$ leaves are put on the right side of v . For the alpha model with $\alpha = 0$, this probability is simply

$$q_0(m, n - m) = \frac{1}{n - 1} \quad (1.57)$$

for all $1 \leq m \leq n - 1$ (see [For, Lemma 27]).

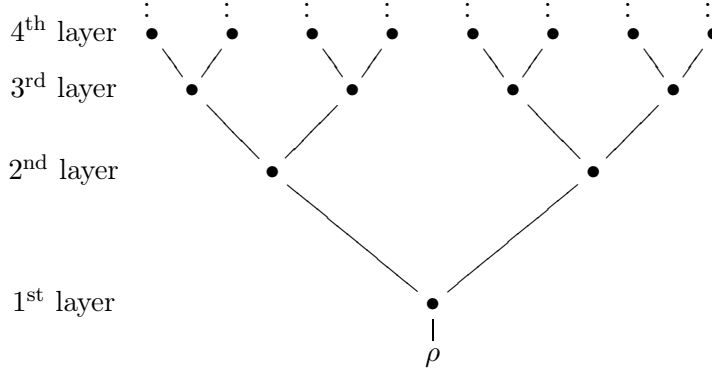


Figure 1.6: The n^{th} layer consists in the 2^{n-1} branch points that are at graphical distance n from the root.

We are interested in the limit of this split distribution as the number of leaves n goes to infinity. Let $v \in \text{br}(T)$ and consider U_v such that given $\eta_\rho(v)$, the vector $(\eta_1(v), \eta_2(v))$ has the same distribution as

$$((1 - \eta_\rho(v))U_v, (1 - \eta_\rho(v))(1 - U_v)). \quad (1.58)$$

Since the $(\alpha = 0)$ -Ford model is Markovian self-similar, U_v is independent of $\eta_\rho(v)$ and U_v has the same distribution for all $v \in \text{br}(T)$. Moreover, for all $x \in [0, 1]$,

$$\mathbb{P}(U_v \leq x) = \lim_{n \rightarrow \infty} \sum_{m=1}^{\lfloor xn \rfloor} q_0(m, n-m) = x, \quad (1.59)$$

and the result follows. \square

Proof of Proposition 1.23. Due to exchangeability in the α -Ford model, for each $v \in T$ the distribution of $\underline{\eta}(v)$ and thus also

$$c_* \mu^{\otimes 3}(v) = 3! \eta_\rho(v) \eta_1(v) \eta_2(v) \quad (1.60)$$

depend on v only through $\#[\rho, v]$. Denote therefore for each $n \in \mathbb{N}$ by

$$T^{(n)} := \{z \in T : \#[\rho, z] = n\} \quad (1.61)$$

the n^{th} layer and note that $\#T^{(n)} = 2^{n-1}$ (Figure 1.6).

Let $(U_i)_{i \geq 0}$ be a sequence of i.i.d. uniform random variables with values in $(0, 1)$. Using the property 1.24 recursively, we find for each $v \in T^{(n)}$, $n \geq 1$, that $(\eta_\rho(v), \eta_1(v), \eta_2(v))$ equals in distribution

$$(1 - \prod_{i=0}^{n-2} U_i, U_{n-1} \prod_{i=0}^{n-2} U_i, (1 - U_{n-1}) \prod_{i=0}^{n-2} U_i) = (1 - A_{n-1}, A_{n-1} U_{n-1}, A_{n-1} (1 - U_{n-1})) \quad (1.62)$$

with $A_n := \prod_{i=0}^{n-1} U_i$ for $n \in \mathbb{N}$. Notice that A_n has density

$$g_{A_n}(x) = \frac{(-1)^{n-1} \log^{n-1}(x)}{(n-1)!}, \quad (1.63)$$

and thus

$$\sum_{n=2}^{\infty} 2^{n-1} g_{A_{n-1}}(x) = \sum_{n=1}^{\infty} 2^{n-1} \frac{(-1)^{n-2} \log^{n-2}(x)}{(n-2)!} = 2e^{-2 \log x} = \frac{2}{x^2}. \quad (1.64)$$

Summing over all the layers we therefore find that for all continuous $f: \Delta_2 \rightarrow \mathbb{R}$,

$$\begin{aligned}
& \mathbb{E} \left[\int_{T^3} \mu_{\chi^{0,\rho}}^{\otimes 3}(\underline{d}\underline{u}) f(\underline{\eta}(c_{\chi^{0,\rho}}(\underline{u}))) \right] \\
&= 3! \sum_{n=2}^{\infty} 2^{n-1} \mathbb{E} \left[(A_{n-1})^2 (1 - A_{n-1}) U_{n-1} (1 - U_{n-1}) \right. \\
&\quad \left. \times f(1 - A_{n-1}, A_{n-1} U_{n-1}, A_{n-1} (1 - U_{n-1})) \right] \\
&= 3! \int_0^1 x^2 (1 - x) \left(\sum_{n=2}^{\infty} 2^{n-1} g_{A_{n-1}}(x) \right) \int_0^1 y (1 - y) f(1 - x, xy, x(1 - y)) dy dx \\
&= 2 \cdot 3! \int_0^1 (1 - x) \int_0^1 y (1 - y) f(1 - x, xy, x(1 - y)) dy dx \\
&= \int_0^1 \frac{\Gamma(3)}{\Gamma(1)\Gamma(2)} (1 - x) \int_0^1 \frac{\Gamma(4)}{\Gamma(2)\Gamma(2)} y (1 - y) f(1 - x, xy, x(1 - y)) dy dx,
\end{aligned} \tag{1.65}$$

which gives the claim. \square

Further outline. Probabilistic representations as those used in the previous subsections for $\alpha = 0$, $\alpha = \frac{1}{2}$ and $\alpha = 1$ are to the best of our knowledge not known to other $\alpha \in [0, 1]$. However, we will provide in Section 1.6 a complete description of the subtree mass distribution in the general case through a recursive relation giving all the moments of the distribution (see Proposition 1.3). To obtain the latter we will consider in Section 1.4 a Markov chain which has the finite α -Ford model as its invariant distribution. In Section 1.5 we then find an analytic expression for its diffusion limit in terms of a well-posed martingale problem (see Proposition 1.31) with respect to an operator which acts on sample subtree mass polynomials (see (1.11)).

1.4 The α -Ford chain on fixed size cladograms

In this section, we introduce for each $m \in \mathbb{N}$ and $\alpha \in [0, 1]$ with the α -Ford chain a Markov chain on \mathfrak{C}_m , which has the α -Ford model as its stationary distribution. This result will be a key argument to prove in Section 1.5 that the α -Ford algebraic measure tree is an invariant distribution of the α -Ford diffusion (see Proposition 1.32).

Recall from Section 1.1 that the α -Ford chain has the following transition rate: for a pair (x, e) consisting of a leaf and an external (resp. internal) edge not adjacent to x at rate $1 - \alpha$ (resp. α), the Markov chain jumps from its current state \mathfrak{t} to $\mathfrak{t}^{(x,e)}$, which is obtained as follows (see Figures 1.2 and 1.3):

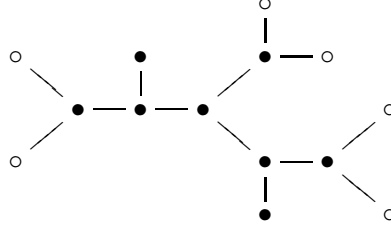
- erase the edge (including the incident vertices) which connects x to the subtree spanned by all leaves but x ,
- split the remaining subtree at the edge e into two pieces,
- reintroduce the above edge (including x and the branch point) at the split point.

The transition rates of the α -Ford chain are, for $\mathfrak{t}, \mathfrak{t}' \in \mathfrak{C}_m$,

$$\begin{aligned}
& q_\alpha^m(\mathfrak{t}, \mathfrak{t}') \\
&= \sum_{x \in \text{Lf}(\mathfrak{t})} \sum_{e \in \text{edge}(\mathfrak{t}_{\wedge x})} \left((1 - \alpha) \mathbf{1}_{\text{ext-edge}(\mathfrak{t}_{\wedge x})}(e) + \alpha \mathbf{1}_{\text{int-edge}(\mathfrak{t}_{\wedge x})}(e) \right) \mathbf{1}_{\mathfrak{t}'}(\mathfrak{t}^{(x,e)}),
\end{aligned} \tag{1.66}$$

where $\text{ext-edge}(\mathfrak{t})$ (resp. $\text{int-edge}(\mathfrak{t})$) denotes the set of external edges of \mathfrak{t} (resp. internal edges), and where we write

$$\mathfrak{t}_{\wedge x} \in \mathfrak{C}_{m-1} \tag{1.67}$$

Figure 1.7: An 8-cladogram with 6 cherry leaves, denoted by \circ .

for the $(m-1)$ -cladogram obtained from \mathbf{t} by deleting the leaf x (and relabelling the labels $j > k$ to $j-1$, with k the label of the leaf x). As for all $\mathbf{t} \in \mathfrak{C}_m$, $\#\text{ext-edge}(\mathbf{t}) = m$ and $\#\text{int-edge}(\mathbf{t}) = m-3$ (for $m \geq 4$), we find for the total jump rate

$$\begin{aligned} \bar{q}_\alpha^m(\mathbf{t}, \mathbf{t}') &:= \sum_{\mathbf{t}' \in \mathfrak{C}_m} q_\alpha^m(\mathbf{t}, \mathbf{t}') \\ &= m((1-\alpha)(m-1) + \alpha(m-4)) = m(m-1-3\alpha). \end{aligned} \quad (1.68)$$

The generator $\tilde{\Omega}_\alpha^m$ of the α -Ford chain acts on all functions $\phi : \mathfrak{C}_m \rightarrow \mathbb{R}$ as follows:

$$\begin{aligned} \tilde{\Omega}_\alpha^m \phi(\mathbf{t}) &= (1-\alpha) \sum_{x \in \text{Lf}(\mathbf{t})} \sum_{e \in \text{ext-edge}(\mathbf{t}_{\wedge x})} \left(\phi(\mathbf{t}^{(x,e)}) - \phi(\mathbf{t}) \right) \\ &\quad + \alpha \sum_{x \in \text{Lf}(\mathbf{t})} \sum_{e \in \text{int-edge}(\mathbf{t}_{\wedge x})} \left(\phi(\mathbf{t}^{(x,e)}) - \phi(\mathbf{t}) \right). \end{aligned} \quad (1.69)$$

Note that we have, for all $\alpha \in [0, 1]$,

$$\tilde{\Omega}_\alpha^m = (1-2\alpha)\tilde{\Omega}_{\text{Kin}}^m + 2\alpha\tilde{\Omega}_{\text{Ald}}^m, \quad (1.70)$$

where $\tilde{\Omega}_{\text{Ald}}^m := \tilde{\Omega}_{\frac{1}{2}}^m$ is the generator of the Aldous chain and $\tilde{\Omega}_{\text{Kin}}^m := \tilde{\Omega}_0^m$ the one of the Kingman chain, where the edges are introduced into external edges only. Although $(1-2\alpha)\tilde{\Omega}_{\text{Kin}}^m$ is not a generator for $\alpha > \frac{1}{2}$, we will make use of the linearity relation (1.70) to extend to any $\alpha \in [0, 1]$ some results that are shown for the generators in the cases $\alpha = 0, \frac{1}{2}$. In particular, since the Aldous chain was studied in detail in discrete time in [Ald00, Sch01], and in continuous time (up to a time change by a factor of 2) in [LMW20], it will sometimes be enough to show the result for the Kingman case $\alpha = 0$.

In order to state a duality result for the α -Ford diffusion (see Proposition 1.30), we also introduce the *backward Markov chain*, i.e. the Markov chain with reversed transition rates

$$q_{\alpha\downarrow}^m(\mathbf{t}', \mathbf{t}) := q_\alpha^m(\mathbf{t}, \mathbf{t}'). \quad (1.71)$$

To describe this chain through its generator, notice that after modifying the cladogram \mathbf{t} according to the forward chain, we can go back by a similar move: picking a leaf and inserting it to a given edge. But in this case, the rates for a pair (x, e) will differ depending on the position of the leaf (not of the edge) in the tree. Consider for example the Kingman chain with $\alpha = 0$. Since we choose any leaf and put it to an external edge in the forward chain, the edge we pick for the reverse move can be any edge (not only external edge), but the leaf we pick has to be a *cherry leaf*. We call x a *cherry leaf* of \mathbf{t} if there exists y another leaf of \mathbf{t} such that x and y are both adjacent to the same internal branch point (Figure 1.7). We write

$$\text{ch-Lf}(\mathbf{t}) \subseteq \text{Lf}(\mathbf{t}) \quad (1.72)$$

for the set of cherry leaves of \mathbf{t} .

Thus, the generator $\tilde{\Omega}_{\alpha\downarrow}^m$ of the α -Ford backward chain acts on functions $\phi : \mathfrak{C}_m \rightarrow \mathbb{R}$ as follows:

$$\begin{aligned} \tilde{\Omega}_{\alpha\downarrow}^m \phi(\mathbf{t}) &:= (1-\alpha) \sum_{x \in \text{ch-Lf}(\mathbf{t})} \sum_{e \in \text{edge}(\mathbf{t}_{\wedge x})} \left(\phi(\mathbf{t}^{(x,e)}) - \phi(\mathbf{t}) \right) \\ &\quad + \alpha \sum_{x \notin \text{ch-Lf}(\mathbf{t})} \sum_{e \in \text{edge}(\mathbf{t}_{\wedge x})} \left(\phi(\mathbf{t}^{(x,e)}) - \phi(\mathbf{t}) \right). \end{aligned} \quad (1.73)$$

Note that for $m \geq 5$, the α -Ford chain is symmetric if and only if $\alpha = \frac{1}{2}$ (Aldous chain).

We have the following relation between the α -Ford forward and backward chains:

Proposition 1.25 (Feynman-Kac duality). *Let $X^{m,\alpha} = (X_t^{m,\alpha})_{t \geq 0}$ be the α -Ford forward chain and $Y^{m,\alpha} = (Y_t^{m,\alpha})_{t \geq 0}$ the backward chain. Then for all $\mathbf{t}, \mathbf{s} \in \mathfrak{C}_m$,*

$$\mathbb{E}_{\mathbf{s}} [\mathbf{1}_{\mathbf{t}}(X_t^{m,\alpha})] = \mathbb{E}_{\mathbf{t}} \left[\mathbf{1}_{\mathbf{s}}(Y_t^{m,\alpha}) \exp \left(\int_0^t \beta_{\alpha}^m(Y_s^{m,\alpha}) ds \right) \right], \quad (1.74)$$

where

$$\beta_{\alpha}^m(\mathbf{t}) := (1 - 2\alpha)(\#(\text{ch-lf}(\mathbf{t}))(2m - 5) - m(m - 1)). \quad (1.75)$$

Proof. For $m \in \mathbb{N}$ and $\mathbf{t}, \mathbf{t}' \in \mathfrak{C}_m$, we write $H(\mathbf{t}', \mathbf{t}) := \mathbf{1}_{\{\mathbf{t}' = \mathbf{t}\}}$. We claim that

$$\tilde{\Omega}_{\alpha}^m H(\cdot, \mathbf{t})(\mathbf{t}') = \tilde{\Omega}_{\alpha \downarrow}^m H(\mathbf{t}', \cdot)(\mathbf{t}) + \beta_{\alpha}^m(\mathbf{t}) H(\mathbf{t}', \cdot)(\mathbf{t}). \quad (1.76)$$

Using (1.70) and an analogous relation for backward chains, it is enough to show the claim for $\alpha = \frac{1}{2}$ and $\alpha = 0$. The case $\alpha = \frac{1}{2}$ is obvious, since the Aldous chain is symmetric and $\beta_{\frac{1}{2}}^m(\mathbf{t}) = 0$ for all \mathbf{t} . In the case $\alpha = 0$, we have

$$\begin{aligned} & \tilde{\Omega}_{\text{Kin}}^m H(\cdot, \mathbf{t})(\mathbf{t}') - \tilde{\Omega}_{\text{Kin} \downarrow}^m H(\mathbf{t}', \cdot)(\mathbf{t}) \\ &= \sum_{x \in \text{lf}(\mathbf{t}')} \sum_{e \in \text{ext-edge}(\mathbf{t}'_{\wedge x})} \left(\mathbf{1}_{\mathbf{t}}(\mathbf{t}'^{(x,e)}) - \mathbf{1}_{\mathbf{t}}(\mathbf{t}') \right) - \sum_{y \in \text{ch-lf}(\mathbf{t})} \sum_{f \in \text{edge}(\mathbf{t}_{\wedge y})} \left(\mathbf{1}_{\mathbf{t}'}(\mathbf{t}^{(y,f)}) - \mathbf{1}_{\mathbf{t}'}(\mathbf{t}) \right) \\ &= - \sum_{x \in \text{lf}(\mathbf{t}')} \sum_{e \in \text{ext-edge}(\mathbf{t}'_{\wedge x})} \mathbf{1}_{\mathbf{t}}(\mathbf{t}') + \sum_{y \in \text{ch-lf}(\mathbf{t})} \sum_{f \in \text{edge}(\mathbf{t}_{\wedge y})} \mathbf{1}_{\mathbf{t}'}(\mathbf{t}) \\ &= \beta_0^m(\mathbf{t}) \mathbf{1}_{\mathbf{t}'}(\mathbf{t}), \end{aligned} \quad (1.77)$$

where we used for the second equality that if there exists one forward move to go from \mathbf{t}' to \mathbf{t} , then there exists one backward move to go from \mathbf{t} to \mathbf{t}' , and reciprocally. And if this is the case, then both moves are unique.

This proves (1.76) and the result follows by [EK86, Lemma 4.4.11, Corollary 4.4.13]. \square

We have defined a family of Markov chains on a finite state space. For all $\alpha \in [0, 1]$, the chain is irreducible recurrent and thus has a unique invariant distribution. But the following result stills holds for $\alpha = 1$.

Proposition 1.26 (Invariance). *For all $\alpha \in [0, 1]$ and $m \in \mathbb{N}$, the α -Ford model, $\tilde{\chi}_m^{\alpha}$, on \mathfrak{C}_m is the unique invariant distribution of the α -Ford (forward) chain. In particular, for all $\phi: \mathfrak{C}_m \rightarrow \mathbb{R}$ and $\mathbf{t}' \in \mathfrak{C}_m$,*

$$\sum_{\mathbf{t} \in \mathfrak{C}_m} \mathbb{P}(\tilde{\chi}_m^{\alpha} = \mathbf{t}) q_{\alpha}^m(\mathbf{t}, \mathbf{t}') = \bar{q}_{\alpha}^m(\mathbf{t}) \mathbb{P}(\tilde{\chi}_m^{\alpha} = \mathbf{t}'). \quad (1.78)$$

Proof. The proof relies heavily on the consistency property of the α -Ford model. Let $\alpha \in [0, 1]$. By (1.68), we want to prove for all $\mathbf{t}' \in \mathfrak{C}_m$ that

$$\sum_{\mathbf{t} \in \mathfrak{C}_m} \mathbb{P}(\{\tilde{\chi}_m^{\alpha} = \mathbf{t}\}) q_{\alpha}^m(\mathbf{t}, \mathbf{t}') = m(m - 1 - 3\alpha) \mathbb{P}(\{\tilde{\chi}_m^{\alpha} = \mathbf{t}'\}). \quad (1.79)$$

For $\mathbf{t} = (T, c, \zeta) \in \mathfrak{C}_m$, we denote the set of labels of cherry leaves of \mathbf{t} by

$$\text{ch-lb}(\mathbf{t}) := \{k \in \{1, \dots, m\} : \zeta(k) \in \text{ch-lf}(\mathbf{t})\}, \quad (1.80)$$

and for $k \in \{1, \dots, m\}$, we write

$$\mathbf{t}^{(k,e)} := \mathbf{t}^{(\zeta(k), e)} \quad \text{and} \quad \mathbf{t}_{\wedge k} := \mathbf{t}_{\wedge(\zeta(k))}. \quad (1.81)$$

For the left hand side of the equation, we exploit that for $\mathbf{t}, \mathbf{t}' \in \mathfrak{C}_m$ and $k \in \{1, \dots, m\}$, $\mathbf{t}^{(k,e)} = \mathbf{t}'$ for some $e \in \text{edge}(\mathbf{t}_{\wedge k})$ if and only if $\mathbf{t}_{\wedge k} = \mathbf{t}'_{\wedge k}$ and more precisely, $\mathbf{t}^{(k,e)} = \mathbf{t}'$ for some $e \in \text{ext-edge}(\mathbf{t}_{\wedge k})$ if and only if $\mathbf{t}_{\wedge k} = \mathbf{t}'_{\wedge k}$ and $k \in \text{ch-lb}(\mathbf{t}')$, and in this case, the edge e is unique. This also implies that $\mathbf{t}^{(k,e)} = \mathbf{t}'$ for some $e \in \text{int-edge}(\mathbf{t}_{\wedge k})$ if and only if $\mathbf{t}_{\wedge k} = \mathbf{t}'_{\wedge k}$ and $k \in \mathfrak{C}\text{ch-lb}(\mathbf{t}')$, and in this case again, the edge e is unique. Therefore,

$$\begin{aligned}
& \sum_{\mathbf{t} \in \mathfrak{C}_m} \mathbb{P}(\tilde{\chi}_m^\alpha = \mathbf{t}) q_\alpha^m(\mathbf{t}, \mathbf{t}') \\
&= \sum_{\mathbf{t} \in \mathfrak{C}_m} \mathbb{P}(\tilde{\chi}_m^\alpha = \mathbf{t}) \sum_{k=1}^m \left((1-\alpha) \sum_{e \in \text{ext-edge}(\mathbf{t}_{\wedge k})} \mathbf{1}_{\mathbf{t}'}(\mathbf{t}^{(k,e)}) + \alpha \sum_{e \in \text{int-edge}(\mathbf{t}_{\wedge k})} \mathbf{1}_{\mathbf{t}'}(\mathbf{t}^{(k,e)}) \right) \\
&= \sum_{k=1}^m \left((1-\alpha) \mathbf{1}_{\text{ch-lb}(\mathbf{t}')(k)} + \alpha \mathbf{1}_{\mathfrak{C}\text{ch-lb}(\mathbf{t}')(k)} \right) \sum_{\substack{\mathbf{t} \in \mathfrak{C}_m \\ \mathbf{t}_{\wedge k} = \mathbf{t}'_{\wedge k}}} \mathbb{P}(\tilde{\chi}_m^\alpha = \mathbf{t}) \\
&= \sum_{k=1}^m \left((1-\alpha) \mathbf{1}_{\text{ch-lb}(\mathbf{t}')(k)} + \alpha \mathbf{1}_{\mathfrak{C}\text{ch-lb}(\mathbf{t}')(k)} \right) \mathbb{P}(\tilde{\chi}_{m-1}^\alpha = \mathbf{t}'_{\wedge k}),
\end{aligned} \tag{1.82}$$

where we used, for the last equality, the consistency property of the α -Ford model. Moreover

For the right hand side of equation (1.79), we make use of the fact that we can construct $\tilde{\chi}_m^\alpha$ as follows (see [For, Proposition 22]):

- take $\tilde{\chi}_{m-1}^\alpha$ the α -Ford $(m-1)$ -cladogram,
- pick an edge e of $\tilde{\chi}_{m-1}^\alpha$ randomly according to the weights of the α -Ford model,
- insert a leaf labelled m together with an edge at e , and denote this new m -cladogram by $(\tilde{\chi}_{m-1}^\alpha)^{(e,m)}$,
- apply a uniform permutation σ to the leaf labels of $(\tilde{\chi}_{m-1}^\alpha)^{(e,m)}$. We write $\sigma((\tilde{\chi}_{m-1}^\alpha)^{(e,m)})$ for the new m -cladogram.

Therefore, writing \mathcal{S}_m for the set of permutations of $\{1, \dots, m\}$, and using similar arguments as above,

$$\begin{aligned}
& \mathbb{P}(\chi_m^\alpha = \mathbf{t}') \\
&= \sum_{\mathfrak{z} \in \mathfrak{C}_{m-1}} \mathbb{P}(\tilde{\chi}_{m-1}^\alpha = \mathfrak{z}) \frac{1}{m!} \sum_{\sigma \in \mathcal{S}_m} \left(\frac{(1-\alpha)}{m-1-3\alpha} \sum_{e \in \text{ext-edge}(\mathfrak{z})} \mathbf{1}_{\mathbf{t}'}(\sigma(\mathfrak{z}^{(e,m)})) \right. \\
&\quad \left. + \frac{\alpha}{m-1-3\alpha} \sum_{e \in \text{int-edge}(\mathfrak{z})} \mathbf{1}_{\mathbf{t}'}(\sigma(\mathfrak{z}^{(e,m)})) \right) \\
&= \sum_{\mathfrak{z} \in \mathfrak{C}_{m-1}} \mathbb{P}(\tilde{\chi}_{m-1}^\alpha = \mathfrak{z}) \frac{1}{m} \sum_{k=1}^m \frac{1}{(m-1)!} \sum_{\substack{\sigma \in \mathcal{S}_m \\ \sigma(m)=k}} \frac{\mathbf{1}_{\mathbf{t}'_{\wedge k}}(\sigma|_{\{1, \dots, m-1\}}(\mathfrak{z}))}{m-1-3\alpha} \\
&\quad \times \left((1-\alpha) \sum_{e \in \text{ext-edge}(\mathfrak{z})} \mathbf{1}_{\mathbf{t}'}(\sigma(\mathfrak{z}^{(e,m)})) + \alpha \sum_{e \in \text{int-edge}(\mathfrak{z})} \mathbf{1}_{\mathbf{t}'}(\sigma(\mathfrak{z}^{(e,m)})) \right) \\
&= \frac{1}{m(m-1-3\alpha)} \sum_{k=1}^m \mathbb{P}(\tilde{\chi}_{m-1}^\alpha = \mathbf{t}'_{\wedge k}) \left((1-\alpha) \mathbf{1}_{\text{ch-lb}(\mathbf{t}')(k)} + \alpha \mathbf{1}_{\mathfrak{C}\text{ch-lb}(\mathbf{t}')(k)} \right),
\end{aligned} \tag{1.83}$$

which together with (1.82) proves (1.78). \square

1.5 The α -Ford chain in the diffusion limit

In this section, we are interested in the diffusion limit of the Ford chains as the number of leaves goes to infinity. We prove our first main result (Theorem 1.1) and thereby give an analytic construction of what we shall call the α -Ford diffusion on cladograms.

Recall from (1.1) the space \mathbb{T}_2 of binary algebraic measure tree with no atoms on the skeleton as well as from (1.31) and from (1.32) the subspace \mathbb{T}_2^N of all binary algebraic measure trees with N leaves and equipped with the uniform distribution on the set of leaves respectively the subspace $\mathbb{T}_2^{\text{cont}}$ of all continuum binary algebraic measure trees. As we have done in Section 1.3, we here use that by forgetting the leaf labelling and adding the uniform distribution on the leaves, cladograms with N leaves can be seen as elements in $\mathbb{T}_2^N \subseteq \mathbb{T}_2$, where the space \mathbb{T}_2 is equipped with the sample shape convergence and thus compact (see Definition 1.9 and Proposition 1.12). This allows to consider the α -Ford chains with values in \mathbb{T}_2^N . Recall from Proposition 1.10 that a convergence determining class of function is given by the class of sample shape polynomials $\Phi^{m,\mathfrak{t}}$ (defined in (1.6)) for some $m \in \mathbb{N}$ and $\mathfrak{t} \in \overline{\mathfrak{C}}_m$. In what follows we extend this class to the space $\Pi_{\mathfrak{s}}$ of linear combination of sample shape polynomials. Note that $\Pi_{\mathfrak{s}}$ forms an algebra which contains the function which is constant $\mathbf{1}$.

Notice that the generator Ω_{α}^N of the \mathbb{T}_2^N -valued α -Ford chain acts on sample shape polynomials as

$$\begin{aligned} \Omega_{\alpha}^N \Phi^{m,\mathfrak{t}}(\chi) &= (1 - \alpha) \sum_{x \in \text{lf}(T)} \sum_{e \in \text{ext-edge}(T,c)} \left(\Phi^{m,\mathfrak{t}}(\chi^{(x,e)}) - \Phi^{m,\mathfrak{t}}(\chi) \right) \\ &+ \alpha \sum_{x \in \text{lf}(T)} \sum_{e \in \text{int-edge}(T,c)} \left(\Phi^{m,\mathfrak{t}}(\chi^{(x,e)}) - \Phi^{m,\mathfrak{t}}(\chi) \right). \end{aligned} \quad (1.84)$$

Recall from (1.69) that $\tilde{\Omega}_{\alpha}^m$ is the generator of the α -Ford forward chain on $\overline{\mathfrak{C}}_m$. Recall from (1.7) the operator Ω_{α} acting on the class of sample shape polynomials, and extend his action linearly to $\Pi_{\mathfrak{s}}$. Since the infinite trees we consider are limits as $N \rightarrow \infty$ of trees in \mathbb{T}_2^N , Proposition 1.11 provides that we can limit our work to binary algebraic measure trees in $\mathbb{T}_2^{\text{cont}}$. Now consider $m \in \mathbb{N}$ and $\mathfrak{t} \in \overline{\mathfrak{C}}_m \setminus \mathfrak{C}_m$. Then if $\mathfrak{s}_{(T,c)}(u_1, \dots, u_m) = \mathfrak{t}$ for some $\chi = (T, c, \mu) \in \mathbb{T}_2^{\text{cont}}$, we have that u_1, \dots, u_m are not distinct, so that $\Phi^{m,\mathfrak{t}}(\chi) = 0$ because $\text{at}(\mu) = \emptyset$. For this reason, we will limit the domain of the operator of the α -Ford forward chain $\mathcal{D}(\Omega_{\alpha})$ to shape polynomials using m -cladograms instead of m -labelled cladograms:

$$\mathcal{D}(\Omega_{\alpha}) := \text{span}\{\Phi^{m,\mathfrak{t}} : m \in \mathbb{N}, \mathfrak{t} \in \mathfrak{C}_m\}. \quad (1.85)$$

Our first observation is the following:

Lemma 1.27. *For all $\Phi \in \mathcal{D}(\Omega_{\alpha})$, we have $\Omega_{\alpha}\Phi \in \mathcal{D}(\Omega_{\alpha})$. In particular,*

$$(\Phi, \Omega_{\alpha}\Phi) \in \mathcal{C}_b(\mathbb{T}_2) \times \mathcal{C}_b(\mathbb{T}_2). \quad (1.86)$$

Proof. Let $\Phi \in \mathcal{D}(\Omega_{\alpha})$. The functions Φ and $\Omega_{\alpha}\Phi$ are linear combinations sample shape polynomials, hence continuous by definition of sample shape convergence. Furthermore, for all $m \in \mathbb{N}$, $\mathfrak{t} \in \mathfrak{C}_m$ and $\chi \in \mathbb{T}_2$, $|\Phi^{m,\mathfrak{t}}(\chi)| \leq 1$ and by definition of $\tilde{\Omega}_{\alpha}^m$ (see (1.69)), $|\Omega_{\alpha}\Phi^{m,\mathfrak{t}}(\chi)| \leq m(2m - 5)$. Thus, Φ and $\Omega_{\alpha}\Phi$ are bounded. \square

To prove the existence, we will follow, as in [LMW20], an argument similar to [EK86, Lemma 4.5.1]. For this, we relate the generators of α -Ford Markov chain and the prospective diffusion limit.

Proposition 1.28 (Uniform convergence of generators). *Let $\alpha \in [0, 1]$. For all $\Phi \in \mathcal{D}(\Omega_{\alpha})$, we have*

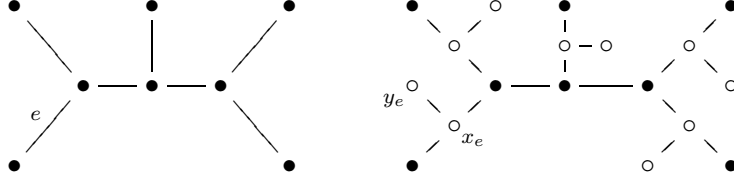
$$\lim_{N \rightarrow \infty} \sup_{\chi \in \mathbb{T}_2^N} |\Omega_{\alpha}^N \Phi(\chi) - \Omega_{\alpha}\Phi(\chi)| = 0. \quad (1.87)$$

Proof. Similar as in (1.70), note here that for all $\alpha \in [0, 1]$,

$$\Omega_{\alpha} = (1 - 2\alpha)\Omega_{\text{Kin}} + 2\alpha\Omega_{\text{Ald}}. \quad (1.88)$$

Since the result was shown by [LMW20] for the Aldous case $\alpha = \frac{1}{2}$, and using (1.88), we need only to show it for the Kingman case $\alpha = 0$.

Consider $\Phi \in \mathcal{D}(\Omega_{\alpha})$. By linearity, we can assume w.l.o.g. that $\Phi = \Phi^{m,\mathfrak{t}}$ for some $m \in \mathbb{N}$ and $\mathfrak{t} \in \mathfrak{C}_m$. If $m = 1, 2, 3$, $\#\mathfrak{C}_m = 1$, so that $\Phi^{m,\mathfrak{t}}$ is constant on \mathbb{T}_2^N , for each $N \in \mathbb{N}$. Therefore,


 Figure 1.8: A finite algebraic tree (T, c) and the extended tree (\bar{T}, \bar{c}) .

$\Omega_{\text{Kin}}^N \Phi^{m, \mathfrak{t}}(\chi) = 0$ for all $\chi \in \mathbb{T}_2^N$, and the convergence holds since $\Omega_{\text{Kin}} \Phi^{m, \mathfrak{t}}$ also equals to zero for $m = 1, 2, 3$. Thus, we suppose $m \geq 4$. Fix $N \in \mathbb{N}$ and $\chi = (T, c, \mu) \in \mathbb{T}_2^N$. We write

$$\epsilon := \frac{1}{N}. \quad (1.89)$$

We extend the algebraic tree to allow for potential new branch points and new leaves due to the chain moves on binary trees. To this end, for each external edge $e \in \text{ext-edge}(T, c)$, we introduce two additional points x_e, y_e , i.e., we consider

$$\bar{T} = T \cup \bigcup_{e \in \text{ext-edge}(T, c)} \{x_e, y_e\}, \quad (1.90)$$

and extend c to $\bar{c} : \bar{T}^3 \rightarrow \bar{T}$ which is uniquely defined as follows (see Figure 1.8). (\bar{T}, \bar{c}) is an algebraic tree such that for $e = \{a, b\} \in \text{ext-edge}(T, c)$, we have $x_e \in (a, b)$ in (\bar{T}, \bar{c}) , and

$$\bar{c}(y_e, x_e, z) = x_e, \quad \forall z \in \bar{T} \setminus \{y_e\}. \quad (1.91)$$

For $k \in \{1, \dots, m\}$ and $x \in \bar{T}$, let

$$\theta_{k, x} : T^m \rightarrow \bar{T}^m \quad (1.92)$$

be the replacement operator which replaces the k^{th} -coordinate by x . For $\chi = (T, \mu) = (\bar{T}, \mu)$ and $(x, e) \in \text{lf}(T, c) \times \text{ext-edge}(T, c)$, we write $\chi^{(x, e)}$ the binary algebraic measure tree obtained by the chain move with (x, e) , i.e.,

$$\chi^{(x, e)} := (\bar{T}, \bar{c}, \mu + \epsilon \delta_{y_e} - \epsilon \delta_x). \quad (1.93)$$

The difference between sampling with the new and old measure is given by

$$\begin{aligned} & (\mu + \epsilon \delta_{y_e} - \epsilon \delta_x)^{\otimes m} - \mu^{\otimes m} \\ &= \epsilon \sum_{k=1}^m \mu^{\otimes(k-1)} \otimes (\delta_{y_e} - \delta_x) \otimes \mu^{\otimes(m-k)} \\ & \quad + \epsilon^2 \sum_{1 \leq k < j \leq m} \mu^{\otimes(k-1)} \otimes (\delta_{y_e} - \delta_x) \otimes \mu^{\otimes(j-1)} \otimes (\delta_{y_e} - \delta_x) \otimes \mu^{\otimes(m-j-k)} + \tilde{\mu} \\ &= \epsilon \sum_{k=1}^m (\mu^{\otimes m} \circ \theta_{k, y_e}^{-1} - \mu^{\otimes m} \circ \theta_{k, x}^{-1}) - \epsilon^2 \sum_{j \neq k=1}^m \mu^{\otimes m} \circ \theta_{k, y_e}^{-1} \circ \theta_{j, x}^{-1} + \tilde{\mu}, \end{aligned} \quad (1.94)$$

where $\tilde{\mu}$ is a signed measure on \bar{T}^m with

$$\tilde{\mu}(\{u_1, \dots, u_m\} : u_1, \dots, u_m \text{ distinct}\} = 0. \quad (1.95)$$

As $\mathfrak{t} \in \mathfrak{C}_m$, we therefore find that

$$\begin{aligned} \Omega_{\text{Kin}}^N \Phi^{m, \mathfrak{t}}(\chi) &= \sum_{x \in \text{lf}(T, c)} \sum_{e \in \text{ext-edge}(T, c)} (\Phi^{m, \mathfrak{t}}(\chi^{(x, e)}) - \Phi^{m, \mathfrak{t}}(\chi)) \\ &= \sum_{k=1}^m A_k - \sum_{k \neq j=1}^m B_{k, j}, \end{aligned} \quad (1.96)$$

with

$$\begin{aligned} A_k &:= \epsilon \sum_{x \in \text{lf}(T, c)} \sum_{e \in \text{ext-edge}(T, c)} \int_{T^m} \mu^{\otimes m}(\text{d}\underline{u}) \left(\mathbf{1}_{\mathfrak{t}(\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, y_e} \underline{u}))} - \mathbf{1}_{\mathfrak{t}(\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, x} \underline{u}))} \right) \\ &= (N-1) \sum_{k=1}^m \int_{T^m} \mu^{\otimes k-1} \otimes (\nu - \mu) \otimes \mu^{\otimes m-k}(\text{d}\underline{u}) \mathbf{1}_{\mathfrak{t}(\mathfrak{s}_{(\overline{T}, \overline{c})}(\underline{u}))}, \end{aligned} \quad (1.97)$$

and

$$\begin{aligned} B_{k,j} &:= \epsilon^2 \sum_{x \in \text{lf}(T, c)} \sum_{e \in \text{ext-edge}(T, c)} \int_{T^m} \mu^{\otimes m}(\text{d}\underline{u}) \mathbf{1}_{\mathfrak{t}(\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, y_e} \circ \theta_{j, x} \underline{u}))} \\ &= \epsilon \int_{T^m} \mu^{\otimes m}(\text{d}\underline{u}) \sum_{e \in \text{ext-edge}(T, c)} \mathbf{1}_{\mathfrak{t}(\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, y_e} \underline{u}))} \\ &= \sum_{k=1}^m \int_{T^m} \mu^{\otimes k-1} \otimes \nu \otimes \mu^{\otimes m-k}(\text{d}\underline{u}) \mathbf{1}_{\mathfrak{t}(\mathfrak{s}_{(\overline{T}, \overline{c})}(\underline{u}))} \end{aligned} \quad (1.98)$$

and with

$$\nu := \epsilon \sum_{e \in \text{ext-edge}(T, c)} \delta_{y_e} \quad (1.99)$$

which is different from μ . However, notice that if we consider u_1, \dots, u_m sampled from (T, c) according to μ and if e is an external edge, then with $v_e \in \text{lf}(T)$ being the unique leaf with $v_e \in e$, $\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, y_e} \underline{u}) = \mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, v_e} \underline{u})$ if and only if $e \cap \{u_i : i = 1, \dots, u_m, i \neq k\} = \emptyset$. Furthermore, if $e \cap \{u_i : i = 1, \dots, u_m, i \neq k\} \neq \emptyset$, then $\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, v_e} \underline{u}) \neq \mathfrak{t}$ because $\mathfrak{t} \in \mathfrak{C}_m$. Thus,

$$\begin{aligned} &\sum_{k=1}^m \int_{T^m} \mu^{\otimes k-1} \otimes (\nu - \mu) \otimes \mu^{\otimes m-k}(\text{d}\underline{u}) \mathbf{1}_{\mathfrak{t}(\mathfrak{s}_{(\overline{T}, \overline{c})}(\underline{u}))} \\ &= \epsilon \int_{T^m} \mu^{\otimes m}(\text{d}\underline{u}) \sum_{\substack{i=1 \\ i \neq k}}^m \mathbf{1}_{\mathfrak{t}(\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, y_{e_{u_i}}} \underline{u}))}. \end{aligned} \quad (1.100)$$

It follows, in particular, that for all $1 \leq k \neq j \leq m$,

$$B_{k,j} = \Phi^{m, \mathfrak{t}}(\chi) + \mathcal{O}(\epsilon), \quad (1.101)$$

where the term $\mathcal{O}(\epsilon)$ goes to 0 as $N \rightarrow \infty$ uniformly over all $\chi \in \mathbb{T}_2^N$.

To further simplify the right hand side of (1.100), fix $\underline{u} \in T^m$. Recall 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 $j-1$). If $\mathfrak{t} = \mathfrak{s}_{(\overline{T}, \overline{c})}(\underline{u})$, then $\mathfrak{t}_{\wedge k} = \mathfrak{s}_{(\overline{T}, \overline{c})}(\underline{u}_{\wedge k})$ with

$$\underline{u}_{\wedge k} = (u_1, \dots, u_{k-1}, u_{k+1}, \dots, u_m). \quad (1.102)$$

If $\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, y_{e_{u_i}}} \underline{u}) = \mathfrak{t}$, then $\mathfrak{s}_{(\overline{T}, \overline{c})}(\underline{u}_{\wedge k}) = \mathfrak{t}_{\wedge k}$. Reciprocally, if $\mathfrak{s}_{(\overline{T}, \overline{c})}(\underline{u}_{\wedge k}) = \mathfrak{t}_{\wedge k}$, then $\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, y_{e_{u_i}}} \underline{u}) = \mathfrak{t}$ if and only if k is the label of a cherry leaf of \mathfrak{t} and the leaves with labels i and k are both adjacent to the same internal branch point of \mathfrak{t} . Therefore,

$$\int_{T^m} \mu^{\otimes m}(\text{d}\underline{u}) \mathbf{1}_{\mathfrak{t}_{\wedge k}}(\mathfrak{s}_{(\overline{T}, \overline{c})}(\underline{u}_{\wedge k})) = \mathbf{1}_{\text{ch-lb}(\mathfrak{t})}(k) \Phi^{m-1, \mathfrak{t}_{\wedge k}}(\chi). \quad (1.103)$$

All in all,

$$\Omega_{\text{Kin}}^N \Phi^{m, \mathfrak{t}}(\chi) = \sum_{k \in \text{ch-lb}(\mathfrak{t})} \Phi^{m-1, \mathfrak{t}_{\wedge k}}(\chi) - m(m-1) \Phi^{m, \mathfrak{t}}(\chi) + \mathcal{O}(\epsilon). \quad (1.104)$$

We have

$$\sum_{k \in \text{ch-lb}(\mathfrak{t})} \Phi^{m-1, \mathfrak{t}_{\wedge k}}(\chi) = \int_{T^m} \mu^{\otimes m}(\text{d}\underline{u}) \sum_{k \in \text{ch-lb}(\mathfrak{t})} \mathbf{1}_{\mathfrak{t}_{\wedge k}}(\mathfrak{s}_{(T, c)}(\underline{u}_{\wedge k})). \quad (1.105)$$

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 $\mathfrak{s}_{(T,c)}(\underline{u}) = \mathfrak{t}^{(k,e)}$, where $\mathfrak{t}^{(k,e)}$ is the m -cladogram obtained by inserting a leaf with label k at the edge e in $\mathfrak{t}_{\wedge k}$ (and relabelling the labels $j \geq k$ to $j + 1$). If such an edge e exists, it is unique, and we have

$$\mathbf{1}_{\mathfrak{t}_{\wedge k}}(\mathfrak{s}_{(T,c)}(\underline{u}_{\wedge k})) = \sum_{e \in \text{edge}(\mathfrak{t}_{\wedge k})} \mathbf{1}_{\mathfrak{t}^{(k,e)}}(\mathfrak{s}_{(T,c)}(\underline{u})). \quad (1.106)$$

Recall from (1.73) with $\alpha = 0$ the generator $\tilde{\Omega}_{\text{Kin}\downarrow}^m$ of the backward Kingman chain. By linearity, we have

$$\begin{aligned} & \sum_{k \in \text{ch-lb}(\mathfrak{t})} \Phi^{m-1, \mathfrak{t}_{\wedge k}}(\chi) \\ &= \int_{T^m} \mu^{\otimes m}(\underline{u}) \sum_{k \in \text{ch-lb}(\mathfrak{t})} \sum_{e \in \text{edge}(\mathfrak{t}_{\wedge k})} \mathbf{1}_{\mathfrak{s}_{(T,c)}(\underline{u})}(\mathfrak{t}^{(k,e)}) \\ &= \int_{T^m} \mu^{\otimes m}(\underline{u}) \left(\tilde{\Omega}_{\text{Kin}\downarrow}^m \mathbf{1}_{\mathfrak{s}_{(T,c)}(\underline{u})}(\mathfrak{t}) + \#(\text{ch-lf}(\mathfrak{t})) \# \text{edge}(\mathfrak{t}_{\wedge k}) \mathbf{1}_{\mathfrak{s}_{(T,c)}(\underline{u})}(\mathfrak{t}) \right). \end{aligned} \quad (1.107)$$

With the notation $\beta_0^m(\mathfrak{t}) := \#(\text{ch-lf}(\mathfrak{t}))(2m - 5) - m(m - 1)$,

$$\Omega_{\text{Kin}}^N \Phi^{m, \mathfrak{t}}(\chi) = \int_{T^m} \mu^{\otimes m}(\underline{u}) \left(\tilde{\Omega}_{\text{Kin}\downarrow}^m \mathbf{1}_{\mathfrak{s}_{(T,c)}(\underline{u})}(\mathfrak{t}) + \beta_0^m(\mathfrak{t}) \mathbf{1}_{\mathfrak{s}_{(T,c)}(\underline{u})}(\mathfrak{t}) \right) + \mathcal{O}(\epsilon). \quad (1.108)$$

Finally, using the relation between the Kingman forward and backward chains stated in (1.76) with $\alpha = 0$,

$$\begin{aligned} \Omega_{\text{Kin}}^N \Phi^{m, \mathfrak{t}}(\chi) &= \int_{T^m} \mu^{\otimes m}(\underline{u}) \tilde{\Omega}_{\text{Kin}}^m \mathbf{1}_{\mathfrak{s}_{(T,c)}(\underline{u})}(\mathfrak{t}) + \mathcal{O}(\epsilon) \\ &= \Omega_{\text{Kin}} \Phi^{m, \mathfrak{t}}(\chi) + \mathcal{O}(\epsilon), \end{aligned} \quad (1.109)$$

and we have the result for all $m \geq 4$. \square

We can now deduce from this convergence the existence of a solution to the martingale problem.

Proposition 1.29 (Existence). *Let $\alpha \in [0, 1]$. 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 $\chi_N \Rightarrow \chi$, as $N \rightarrow \infty$, where χ is a random tree in $\mathbb{T}_2^{\text{cont}}$ with distribution P_0 . Let $X^N := (X_t^N)_{t \geq 0}$ be the α -Ford forward chain started in χ_N . Then the sequence $(X^N)_{N \in \mathbb{N}}$ is tight in $\mathcal{D}_{\mathbb{T}_2}$, and any limit point $(X_t)_{t \geq 0}$ has continuous paths in $\mathbb{T}_2^{\text{cont}}$ and satisfies the $(\Omega_\alpha, \mathcal{D}(\Omega_\alpha), P_0)$ -martingale problem.*

In particular, for any probability measure P_0 on $\mathbb{T}_2^{\text{cont}}$ there exists a solution in $\mathcal{C}_{\mathbb{T}_2^{\text{cont}}}(\mathbb{R}_+)$ to the $(\Omega_\alpha, \mathcal{D}(\Omega_\alpha), P_0)$ -martingale problem.

Proof. The proof follows the same line of arguments of [LMW20, Corollary 3.3] with replacing [LMW20, Proposition 3.2] by Proposition 1.28. In particular, to show that all limit points satisfy the martingale problem, we adapt the proof of [EK86, Lemma 4.5.1] in the following way.

Let $(X_t)_{t \geq 0}$ be a limit point of $(X^N)_{N \in \mathbb{N}}$. Let $0 \leq t_i \leq t < s$ and $h_i \in \mathcal{C}_b(\mathbb{T}_2)$, $i = 1, \dots, k$. For all $N \in \mathbb{N}$ and $\Phi \in \mathcal{D}(\Omega_\alpha)$,

$$\begin{aligned} & \left| \int_t^s \Omega_\alpha \Phi(X_u) du - \int_t^s \Omega_\alpha^N \Phi(X_u^N) du \right| \\ & \leq \int_t^s |\Omega_\alpha \Phi(X_u) - \Omega_\alpha^N \Phi(X_u^N)| du + \int_t^s |\Omega_\alpha \Phi(X_u^N) - \Omega_\alpha^N \Phi(X_u^N)| du, \end{aligned} \quad (1.110)$$

and we know from Lemma 1.27 that $\Omega_\alpha \Phi$ is continuous and bounded. Therefore, using that $(X_t)_{t \geq 0}$ has continuous paths (see the proof of [LMW20, Corollary 3.3]), we have by Proposition 1.28 that for all $\Phi \in \mathcal{D}(\Omega_\alpha)$,

$$\begin{aligned} & \mathbb{E} \left[\left(\Phi(X_t) - \Phi(X_s) - \int_t^s \Omega_\alpha \Phi(X_u) du \right) \prod_{i=1}^k h_i(X_{t_i}) \right] \\ &= \lim_{N \rightarrow \infty} \mathbb{E} \left[\left(\Phi(X_t^N) - \Phi(X_s^N) - \int_t^s \Omega_\alpha^N \Phi(X_u^N) du \right) \prod_{i=1}^k h_i(X_{t_i}^N) \right] = 0. \end{aligned} \quad (1.111)$$

Thus X is a solution of the martingale problem for $(\Omega_\alpha, \mathcal{D}(\Omega_\alpha))$. \square

To prove the uniqueness of the solution for the martingale problem, we will use a result of duality which appeared in the proof of Proposition 1.28. We claim that we have a duality between a diffusion and a Markov chain on a finite state space: the dual of the α -Ford diffusion is the dual α -Ford Markov chain, that is, the backward chain. For this, consider the duality functions given by the sample shape polynomials, i.e. for $m \in \mathbb{N}$, $\chi = (T, c, \mu) \in \mathbb{T}_2^{\text{cont}}$ and $\mathbf{t} \in \mathfrak{C}_m$, let

$$H^m(\chi, \mathbf{t}) := \Phi^{m, \mathbf{t}}(\chi) = \int_{T^m} \mu^{\otimes m}(\mathrm{d}\underline{u}) \mathbf{1}_{\mathbf{t}}(\mathfrak{s}_{(T, c)}(\underline{u})). \quad (1.112)$$

Proposition 1.30 (Feynman-Kac duality). *Let $\alpha \in [0, 1]$. Let P_0 be a probability measure on $\mathbb{T}_2^{\text{cont}}$, let $X := ((T_t, c_t, \mu_t))_{t \geq 0}$ be a solution to the $(\Omega_\alpha, \mathcal{D}(\Omega_\alpha), P_0)$ -martingale problem in $\mathcal{D}_{\mathbb{T}_2^{\text{cont}}}(\mathbb{R}_+)$. For $m \in \mathbb{N}$ and $\mathbf{t} \in \mathfrak{C}_m$, we denote by $Y^m := (Y_t^m)_{t \geq 0}$ the α -Ford backward chain on \mathfrak{C}_m -cladograms started in $Y_0^m = \mathbf{t}$. Then if Y^m is independent of X ,*

$$\mathbb{E}_{P_0}^X [H^m(X_t, \mathbf{t})] = \int_{\mathbb{T}_2^{\text{cont}}} \mathbb{E}_{\mathbf{t}}^{Y^m} \left[H^m(\chi, Y_t^m) \exp \left(\int_0^t \beta_\alpha^m(Y_s^m) \mathrm{d}s \right) \right] P_0(\mathrm{d}\chi), \quad (1.113)$$

where $\beta_\alpha^m(\mathbf{t}) = (1 - 2\alpha) (\#(\text{ch-lf}(\mathbf{t}))(2m - 5) - m(m - 1))$.

Proof. Fix $m \in \mathbb{N}$, $\chi = (T, c, \mu) \in \mathbb{T}_2^{\text{cont}}$ and $\mathbf{t} \in \mathfrak{C}_m$. Using (1.76),

$$\begin{aligned} \Omega_\alpha H^m(\cdot, \mathbf{t})(\chi) &:= \int_{T^m} \mu^{\otimes m}(\mathrm{d}\underline{u}) \tilde{\Omega}_\alpha^m \mathbf{1}_{\mathbf{t}}(\mathfrak{s}_{(T, c)}(\underline{u})) \\ &= \int_{T^m} \mu^{\otimes m}(\mathrm{d}\underline{u}) \left(\tilde{\Omega}_{\alpha \downarrow}^m \mathbf{1}_{\mathfrak{s}_{(T, c)}(\underline{u})}(\mathbf{t}) + \beta_\alpha^m(\mathbf{t}) \mathbf{1}_{\mathfrak{s}_{(T, c)}(\underline{u})}(\mathbf{t}) \right) \\ &= \tilde{\Omega}_{\alpha \downarrow}^m \left(\int_{T^m} \mu^{\otimes m}(\mathrm{d}\underline{u}) \mathbf{1}_{\mathfrak{s}_{(T, c)}(\underline{u})}(\mathbf{t}) \right) + \beta_\alpha^m(\mathbf{t}) \Phi^{m, \mathbf{t}}(\chi) \\ &= \tilde{\Omega}_{\alpha \downarrow}^m H^m(\chi, \cdot)(\mathbf{t}) + \beta_\alpha^m(\mathbf{t}) H^m(\chi, \mathbf{t}). \end{aligned} \quad (1.114)$$

The result then follows by [EK86, Theorem 4.4.11, Corollary 4.4.13]. \square

Proposition 1.31 (Uniqueness of the martingale problem). *Let $\alpha \in [0, 1]$. For all probability measures P_0 on $\mathbb{T}_2^{\text{cont}}$, uniqueness holds for the $(\Omega_\alpha, \mathcal{D}(\Omega_\alpha), P_0)$ -martingale problem in $\mathcal{D}_{\mathbb{T}_2^{\text{cont}}}(\mathbb{R}_+)$.*

Proof. By [EK86, Proposition 4.4.2], the uniqueness of the solution follows from the uniqueness of the one-dimensional distributions of solutions of the martingale problem. But if $(X_t)_{t \geq 0}$ is a solution of the $(\Omega_\alpha, \mathcal{D}(\Omega_\alpha), P_0)$ -martingale problem, then by Proposition 1.30 it satisfies (1.113) for all $m \in \mathbb{N}$ and $\mathbf{t} \in \mathfrak{C}_m$. By Proposition 1.10, the set of functions

$$\{H^m(\cdot, \mathbf{t}) : m \in \mathbb{N}, \mathbf{t} \in \mathfrak{C}_m\} = \{\Phi^{m, \mathbf{t}} : m \in \mathbb{N}, \mathbf{t} \in \mathfrak{C}_m\} \quad (1.115)$$

is separating for probability measures on $\mathbb{T}_2^{\text{cont}}$, and the uniqueness of the one-dimensional distributions follows. \square

Summing up the results of Proposition 1.29 and Proposition 1.31, we have shown Theorem 1.1.

Recall the continuum α -Ford algebraic measure tree χ^α (compare Definition 1.18). The following result states that the distribution of χ^α is invariant under the α -Ford diffusion.

Proposition 1.32 (continuum α -Ford tree is invariant). *Let $\alpha \in [0, 1]$. The α -Ford algebraic measure tree is an invariant distribution of the α -Ford diffusion.*

Proof. Fix $m \in \mathbb{N}$ and $\mathbf{t} \in \mathfrak{C}_m$. We have that

$$\mathbb{E}[\Phi^{m, \mathbf{t}}(\chi^\alpha)] = \mathbb{P}(\tilde{\chi}_m^\alpha = \mathbf{t}), \quad (1.116)$$

where $\tilde{\chi}_m^\alpha$ is the α -Ford model on m -cladograms. Therefore for all $m \in \mathbb{N}$ and $\mathbf{t} \in \mathfrak{C}_m$,

$$\begin{aligned} \mathbb{E}[\Omega_\alpha \Phi^{m,\mathbf{t}}(\chi^\alpha)] &= \mathbb{E} \left[\int_{T_{\chi^\alpha}^m} \mu_{\chi^\alpha}^{\otimes m}(\mathrm{d}\underline{u}) \tilde{\Omega}_\alpha^m \mathbf{1}_{\mathbf{t}}(\mathfrak{s}_{(T,c)}(\underline{u})) \right] \\ &= \tilde{\Omega}_\alpha^m \mathbb{P}(\tilde{\chi}_m^\alpha = \cdot)(\mathbf{t}) = 0, \end{aligned} \quad (1.117)$$

since the α -Ford model $\tilde{\chi}_m^\alpha$ is the invariant distribution of the α -Ford chain (see Proposition 1.26). This proves the claim. \square

Open question 1.33 (ergodicity). A natural question to ask is then whether the α -Ford diffusion converges toward the α -Ford invariant distribution, for any initial distribution. This result would give in the same time the uniqueness of the invariant distribution. It was proven in [LMW20] for the case $\frac{1}{2}$ relying on the duality relation and the symmetry of the Aldous diffusion. However, the α -Ford diffusion is symmetric if and only if $\alpha = \frac{1}{2}$ and in the non-symmetric cases the argument is destroyed due to the exponential term in the Feynman-Kac duality equation (1.113). Ergodicity therefore remains an open question for $\alpha \in [0, 1] \setminus \{\frac{1}{2}\}$.

1.6 Application on sample subtree masses

In this section we are interested in the infinitesimal evolution of the law of the vector of subtree masses under the α -Ford diffusion. Recall from (1.44) the definition of the subtree masses $\underline{\eta}(\underline{u})$ for $\underline{u} = (u_1, u_2, u_3) \in T^3$, that is,

$$\underline{\eta}(\underline{u}) = (\eta_i(\underline{u}))_{i=1,2,3} = (\mu(\mathcal{S}_{c(\underline{u})}(u_i)))_{i=1,2,3}. \quad (1.118)$$

Recall also from (1.10) that for mass polynomials

$$\Phi^f(\chi) = \int_{T^3} \mu^{\otimes 3}(\mathrm{d}\underline{u}) f(\underline{\eta}(\underline{u})) \quad (1.119)$$

where $f \in \mathcal{C}^2([0, 1]^3)$ and $\chi = (T, c, \mu) \in \mathbb{T}_2$, we extend the generator of the α -Ford diffusion by defining

$$\begin{aligned} \Omega_\alpha \Phi^f(\chi) &= \int \mu^{\otimes 3}(\mathrm{d}\underline{u}) \left(\sum_{i,j=1}^3 \eta_i(\delta_{ij} - \eta_j) \partial_{ij}^2 f(\underline{\eta}(\underline{u})) + (2 - \alpha) \sum_{i=1}^3 (1 - 3\eta_i) \partial_i f(\underline{\eta}(\underline{u})) \right. \\ &\quad + (2 - 3\alpha) \sum_{i=1}^3 (f(e_i) - f(\underline{\eta}(\underline{u}))) + \frac{\alpha}{2} \sum_{i \neq j=1}^3 \frac{\mathbf{1}_{\eta_i \neq 0}}{\eta_i} (f \circ \theta_{i,j}(\underline{\eta}(\underline{u})) - f(\underline{\eta}(\underline{u}))) \\ &\quad \left. + \frac{\alpha}{2} \sum_{i \neq j=1}^3 (\mathbf{1}_{\eta_j=0} - \mathbf{1}_{\eta_i=0}) \partial_i f(\underline{\eta}(\underline{u})) \right). \end{aligned} \quad (1.120)$$

Proposition 1.34. *Let $\alpha \in [0, 1]$. For all test functions Φ^f of the form (1.119) with $f: [0, 1]^3 \rightarrow \mathbb{R}$ twice continuously differentiable,*

$$\lim_{N \rightarrow \infty} \sup_{\chi \in \mathbb{T}_2^N} \left| \Omega_\alpha^N \Phi^f(\chi) - \Omega_\alpha \Phi^f(\chi) \right| = 0. \quad (1.121)$$

Remark 1.35 (Heuristics on the dynamics of the sample subtree mass vector). We want to stress that the sampled vector of the three subtree masses does not perform an autonomous dynamics, and therefore is not Markovian. It is true that for a while the subtree masses change as a result of one α -Ford move by taking a leaf away from one subtree and simultaneously adding it to another subtree. This however breaks down when the last leaf of one of the subtrees is taken because in that moment also the branch point of the three originally sampled leaves jumps, and the new position depends one

the whole tree rather than just the subtree spanned by the sample. One can therefore not expect the expression under the integral of the right hand side of (1.120) to be the generator of a Markov process.

We suggest to read the expression as follows: given the sample, as long as we are shuffling around leaves without hereby destroying a subtree, each subtree edge is attracting another subtree leaf which leads in the continuum limit to a Wright-Fisher diffusion with a drift towards the center of the simplex taking care of the discrepancy between the number of leaves and the number of edges in a subtree. The moment we need to find a new position at that we attach the single edge (which was the last edge of the original subtree), we make use of the fact that the generator acts on test functions which take averages over all possible samples of three leaves. Comparing these averages before and after such a jump brings us in a position to use a coupling which more or less takes one of the branch point or leaves which are adjacent to the current branch point of the sample. We therefore see in the continuum limit terms for which we jump from the inner of the simplex to its edges (in case the neighboring vertex was a branch point) or even to one of its corners (in case the neighboring vertex was a leaf) of the simplex.

Proof. The result has already been proved by [LMW20] for the Aldous case $\alpha = \frac{1}{2}$ so it is enough to show it for the Kingman case. Consider $f \in \mathcal{C}^2([0, 1]^3)$, and recall the notations introduced in the proof of Proposition 1.28. For each permutation π of $\{1, 2, 3\}$, define $\pi_*: \Delta_2 \rightarrow \Delta^3$ by $\pi_*(\underline{\eta}) = (\eta_{\pi(1)}, \eta_{\pi(2)}, \eta_{\pi(3)})$. Since $\Phi^f = \Phi^{f \circ \pi_*}$ and $\Omega_{\text{Kin}} \Phi^f = \Omega_{\text{Kin}} \Phi^{f \circ \pi_*}$ for every permutation of $\{1, 2, 3\}$, we may and do assume w.l.o.g. that f is symmetric.

Let \mathbf{t} be the only 3-cladogram. Since $\text{at}(\mu) = \emptyset$, we can introduce a term $\mathbf{1}_{\mathbf{t}}(\mathfrak{s}_{(T,c)}(\underline{u}))$, which we do for later purpose. We thus have

$$\begin{aligned}
 & \Omega_{\text{Kin}}^N \Phi^f(\chi) \\
 &= \Omega_{\text{Kin}}^N \int_{T^3} \mu^{\otimes 3}(\mathrm{d}\underline{u}) \mathbf{1}_{\mathbf{t}}(\mathfrak{s}_{(T,c)}(\underline{u})) f(\underline{\eta}_{\chi}(\underline{u})) \\
 &:= \sum_{x \in \text{lf}(T,c)} \sum_{e \in \text{ext-edge}(T,c)} \left(\int_{T^3} (\mu^{(x,e)})^{\otimes 3}(\mathrm{d}\underline{u}) \mathbf{1}_{\mathbf{t}}(\mathfrak{s}_{(\overline{T},\overline{c})}(\underline{u})) f(\underline{\eta}_{\chi^{(x,e)}}(\underline{u})) \right. \\
 &\quad \left. - \int_{T^3} \mu^{\otimes 3}(\mathrm{d}\underline{u}) \mathbf{1}_{\mathbf{t}}(\mathfrak{s}_{(T,c)}(\underline{u})) f(\underline{\eta}_{\chi}(\underline{u})) \right) \\
 &= \sum_{x \in \text{lf}(T,c)} \sum_{e \in \text{ext-edge}(T,c)} \int_{T^3} ((\mu^{(x,e)})^{\otimes 3} - \mu^{\otimes 3})(\mathrm{d}\underline{u}) \mathbf{1}_{\mathbf{t}}(\mathfrak{s}_{(\overline{T},\overline{c})}(\underline{u})) f(\underline{\eta}_{\chi^{(x,e)}}(\underline{u})) \\
 &\quad + \sum_{x \in \text{lf}(T,c)} \sum_{e \in \text{ext-edge}(T,c)} \int_{T^3} \mu^{\otimes 3}(\mathrm{d}\underline{u}) \left(f(\underline{\eta}_{\chi^{(x,e)}}(\underline{u})) - f(\underline{\eta}_{\chi}(\underline{u})) \right) \\
 &=: \Delta_{\mu} + \Delta_f.
 \end{aligned} \tag{1.122}$$

The term Δ_f appears when the measure μ is left unchanged, but there is change in the three masses. It gives the Wright-Fisher term. To see this, fix $\underline{u} \in T^3$. We abbreviate $\eta_i = \eta_i(\underline{u})$ as long as \underline{u} is fixed. We denote the components of $T \setminus \{c(\underline{u})\}$ by S_i , $i = 1, 2, 3$, ordered such that $\eta_i = \mu(S_i)$. For all $(x, e) \in \text{lf}(T, c) \times \text{ext-edge}(T, c)$ with $x \in S_i$ and $e \in S_j$, we have by a Taylor expansion that

$$f(\underline{\eta}_{\chi^{(x,e)}}(\underline{u})) - f(\underline{\eta}_{\chi}(\underline{u})) = \left(\epsilon(\partial_j - \partial_i) + \frac{\epsilon^2}{2}(\partial_{ii}^2 + \partial_{jj}^2 - 2\partial_{ij}^2) \right) f(\underline{\eta}) + o(\epsilon^2), \tag{1.123}$$

and the $o(\epsilon^2)$ -term is uniform in the binary trees with N leaves as $N \rightarrow \infty$. Now summing over all

$(x, e) \in \text{lf}(T, c) \times \text{ext-edge}(T, c)$, we have

$$\begin{aligned}
& \sum_{x \in \text{lf}(T, c)} \sum_{e \in \text{ext-edge}(T, c)} \left(f(\eta_{\chi(x, e)}(\underline{u})) - f(\eta_{\chi}(\underline{u})) \right) \\
&= \sum_{i \neq j=1}^3 \frac{\eta_i \eta_j}{\epsilon} \epsilon \left((\partial_j - \partial_i + \frac{\epsilon}{2}(\partial_{ii}^2 + \partial_{jj}^2 - 2\partial_{ij}^2)) f(\underline{\eta}) + o(\epsilon) \right) \\
&= \sum_{i \neq j=1}^3 \eta_i \eta_j (\partial_{ii}^2 - \partial_{ij}^2) f(\underline{\eta}) + o(1) \\
&= \sum_{i, j=1}^3 \eta_i (\delta_{ij} - \eta_j) \partial_{ij}^2 f(\underline{\eta}) + o(1).
\end{aligned} \tag{1.124}$$

where we used for the second equality that the highest order term is anti-symmetric in $i \neq j$.

Finally, Fubini's Theorem gives

$$\Delta_f = \int_{T^m} \mu^{\otimes m}(\underline{du}) \mathbf{1}_{\mathfrak{t}}(\mathfrak{s}_{(T, c)}(\underline{u})) \sum_{i, j=1}^3 \eta_i (\delta_{ij} - \eta_j) \partial_{ij}^2 f(\underline{\eta}) + o(1). \tag{1.125}$$

The term Δ_μ gives the effect of the change in μ , when the subtree masses are considered after the chain move. We use for this the same decomposition as in the proof of Proposition 1.28. For $k = 1, 2, 3$ and $x \in \overline{T}$, recall that $\theta_{k, x} : T^3 \rightarrow \overline{T}^3$ is the replacement operator which replaces the k^{th} -coordinate by x . The difference between sampling with the new and old measure is given by

$$(\mu + \epsilon \delta_{y_e} - \epsilon \delta_x)^{\otimes 3} - \mu^{\otimes 3} = \epsilon \sum_{k=1}^3 (\mu^{\otimes 3} \circ \theta_{k, y_e}^{-1} - \mu^{\otimes 3} \circ \theta_{k, x}^{-1}) - \epsilon^2 \sum_{j \neq k=1}^3 \mu^{\otimes 3} \circ \theta_{k, y_e}^{-1} \circ \theta_{j, x}^{-1} + \tilde{\mu}, \tag{1.126}$$

where $\tilde{\mu}$ is a signed measure on \overline{T}^3 with $\tilde{\mu}\{(u_1, u_2, u_3) : u_1, u_2, u_3 \text{ distinct}\} = 0$. But since $\mathfrak{t} \in \mathfrak{C}_3$, the leaf labels are distinct. The purpose of introducing $\mathbf{1}_{\mathfrak{t}}(\mathfrak{s}_{(T, c)}(\underline{u}))$ is to be able to ignore the term due to $\tilde{\mu}$. Thus,

$$\Delta_\mu =: \sum_{k=1}^3 A_k - \sum_{k \neq j=1}^3 B_{k, j}, \tag{1.127}$$

with

$$\begin{aligned}
A_k = & \epsilon \sum_{x \in \text{lf}(T, c)} \sum_{e \in \text{ext-edge}(T, c)} \left(\int_{T^3} \mu^{\otimes 3}(\underline{du}) \mathbf{1}_{\mathfrak{t}}(\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, y_e} \underline{u})) f(\eta_{\chi(x, e)}(\theta_{k, y_e} \underline{u})) \right. \\
& \left. - \int_{T^3} \mu^{\otimes 3}(\underline{du}) \mathbf{1}_{\mathfrak{t}}(\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, x} \underline{u})) f(\eta_{\chi(x, e)}(\theta_{k, x} \underline{u})) \right),
\end{aligned} \tag{1.128}$$

and

$$B_{k, j} = \epsilon^2 \sum_{x \in \text{lf}(T, c)} \sum_{e \in \text{ext-edge}(T, c)} \int_{T^3} \mu^{\otimes 3}(\underline{du}) \mathbf{1}_{\mathfrak{t}}(\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, y_e} \circ \theta_{j, x} \underline{u})) f(\eta_{\chi(x, e)}(\theta_{k, y_e} \circ \theta_{j, x} \underline{u})). \tag{1.129}$$

Let $k \neq j$. We calculate first $B_{k, j}$. Note that the element u_j sampled according to μ does not appear in the integrand, but is replaced by the leaf x which is sampled according to

$$\epsilon \sum_{x \in \text{lf}(T)} \delta_x = \mu. \tag{1.130}$$

Thus, using first Fubini's theorem, we can write

$$B_{k, j} = \epsilon \int_{T^3} \mu^{\otimes 3}(\underline{du}) \sum_{e \in \text{ext-edge}(T, c)} \mathbf{1}_{\mathfrak{t}}(\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, y_e} \underline{u})) f(\eta_{\chi(u_j, e)}(\theta_{k, y_e} \underline{u})). \tag{1.131}$$

In the same way, the element u_k does not appear in the integrand. However, it is this time replaced by the leaf y_e which is sampled according to

$$\nu := \epsilon \sum_{e \in \text{ext-edge}(T)} \delta_{y_e} \neq \mu. \quad (1.132)$$

But we can still find a similar relation. Indeed, consider (u_1, u_2, u_3) sampled from (T, c) according to μ . One can notice that if $e \notin \{e_{u_i}, i = 1, 2, 3\}$, where e_u denotes the external edge connected to the leaf u , then

$$\mathbf{1}_{\mathfrak{t}(\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, y_e} \underline{u}))} f(\eta_{\chi^{(u_j, e)}}(\theta_{k, y_e} \underline{u})) = \mathbf{1}_{\mathfrak{t}(\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, u} \underline{u}))} f(\eta_{\chi^{(u_j, e_u)}}(\theta_{k, u} \underline{u})), \quad (1.133)$$

where $u \in \text{lf}(T)$ is such that $e_u = e$. In other words, sampling according to ν instead of μ leaves the above quantity unchanged when $e \notin \{e_{u_i}, i = 1, 2, 3\}$. This is also true when $e = e_{u_k}$. But if $e = e_{u_i}$ for some $i \in \{1, 2, 3\} \setminus \{k\}$, then

$$\mathbf{1}_{\mathfrak{t}(\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, u_i} \underline{u}))} = 0, \quad (1.134)$$

because $\mathfrak{t} \in \mathfrak{C}_3$ has distinct leaf labels, but not $\mathfrak{s}_{(\overline{T}, \overline{c})}(\theta_{k, u_i} \underline{u})$. Therefore using that $\text{at}(\mu) = \emptyset$, we can write

$$\begin{aligned} B_{k,j} &= \int_{T^3} \mu^{\otimes 3}(\text{d}\underline{u}) f(\eta_{\chi^{(u_j, e_{u_k})}}(\underline{u})) + \epsilon \int_{T^3} \mu^{\otimes 3}(\text{d}\underline{u}) \sum_{\substack{i=1 \\ i \neq k}}^3 f(\eta_{\chi^{(u_j, e_{u_i})}}(\theta_{k, y_{e_{u_i}}} \underline{u})) \\ &= \int_{T^3} \mu^{\otimes 3}(\text{d}\underline{u}) f(\eta_j - \epsilon, \eta_k + \epsilon, \eta_i) + O(\epsilon) \\ &= \int_{T^3} \mu^{\otimes 3}(\text{d}\underline{u}) f(\underline{\eta}_{\chi}(\underline{u})) + O(\epsilon), \end{aligned} \quad (1.135)$$

where we used for the second equality that f is bounded on $[0, 1]^3$, and for the last equality a Taylor expansion and that f' is also bounded on $[0, 1]^3$. We proceed in the same way to calculate A_k : since $\text{at}(\mu) = \emptyset$,

$$\begin{aligned} A_k &= \int_{T^3} \mu^{\otimes 3}(\text{d}\underline{u}) \sum_{x \in \text{lf}(T, c)} f(\eta_{\chi^{(x, e_{u_k})}}(\underline{u})) - \int_{T^3} \mu^{\otimes 3}(\text{d}\underline{u}) \sum_{e \in \text{ext-edge}(T, c)} f(\eta_{\chi^{(u_k, e)}}(\underline{u})) \\ &\quad + \epsilon \int_{T^3} \mu^{\otimes 3}(\text{d}\underline{u}) \sum_{x \in \text{lf}(T, c)} \sum_{\substack{i=1 \\ i \neq k}}^3 f(\eta_{\chi^{(x, e_{u_i})}}(\theta_{k, y_{e_{u_i}}} \underline{u})) \\ &= \int_{T^3} \mu^{\otimes 3}(\text{d}\underline{u}) \sum_{i=1}^3 \frac{\eta_i}{\epsilon} (f(\eta_i - \epsilon, \eta_k + \epsilon, \eta_j) - f(\eta_i + \epsilon, \eta_k - \epsilon, \eta_j)) + a_k(\epsilon) \\ &= \int_{T^3} \mu^{\otimes 3}(\text{d}\underline{u}) 2 \sum_{i=1}^3 \eta_i (\partial_k - \partial_i) f(\underline{\eta}_{\chi}(\underline{u})) + a_k(\epsilon) + O(\epsilon), \end{aligned} \quad (1.136)$$

where $a_k(\epsilon)$ denotes the correction term due to a difference between the sampling of leaves and the sampling of external edges:

$$\begin{aligned} a_k(\epsilon) &:= \epsilon \int_{T^3} \mu^{\otimes 3}(\text{d}\underline{u}) \sum_{x \in \text{lf}(T, c)} \sum_{\substack{i=1 \\ i \neq k}}^3 f(\eta_{\chi^{(x, e_{u_i})}}(\theta_{k, y_{e_{u_i}}} \underline{u})) \\ &= \int_{T^3} \mu^{\otimes 3}(\text{d}\underline{u}) \sum_{\substack{i=1 \\ i \neq k}}^3 \epsilon \left(\sum_{x \in \text{lf}(T, c) \setminus \{u_i\}} f(\epsilon, \epsilon, 1 - 2\epsilon) + f(0, \epsilon, 1 - \epsilon) \right) \\ &= \int_{T^3} \mu^{\otimes 3}(\text{d}\underline{u}) 2f(0, 0, 1) + O(\epsilon). \end{aligned} \quad (1.137)$$

In order to calculate $\sum_{k=1}^3 A_k$, we use that

$$\begin{aligned} \sum_{k=1}^3 \sum_{i=1}^3 \eta_i (\partial_k - \partial_i) f(\underline{\eta}_\chi(\underline{u})) &= \sum_{k=1}^3 \partial_k f(\underline{\eta}_\chi(\underline{u})) - 3 \sum_{i=1}^3 \eta_i \partial_i f(\underline{\eta}_\chi(\underline{u})) \\ &= \sum_{i=1}^3 (1 - 3\eta_i) \partial_i f(\underline{\eta}_\chi(\underline{u})). \end{aligned} \quad (1.138)$$

Adding up the results for Δ_f (1.125) and Δ_μ , the proposition is proved for $\alpha = 0$, and so for all $\alpha \in [0, 1]$. \square

We can now give the proof of Theorem 1.2.

Proof of Theorem 1.2. Let $\alpha \in [0, 1]$ and $X^\alpha = (X_t^\alpha)_{t \geq 0}$ be the α -Ford diffusion on $\mathbb{T}_2^{\text{cont}}$. By Proposition 1.11, there exists a sequence $(X_0^{\alpha, N})_{N \in \mathbb{N}}$ of random values converging in law to X_0^α such that for all $N \in \mathbb{N}$, $X_0^{\alpha, N} \in \mathbb{T}_2^N$. As a consequence of Theorem 1.1, the sequence $(X^{\alpha, N})_{N \in \mathbb{N}}$ of α -Ford chains started in $(X_0^{\alpha, N})_{N \in \mathbb{N}}$ converges weakly in Skorokhod path space to X^α . Furthermore, by [LW21, Proposition 5.18], $\Phi^f \in \mathcal{C}_b(\mathbb{T}_2)$ for all $f \in \mathcal{C}^2([0, 1])$ and we can easily see from its definition (1.120) that $\Omega_\alpha \Phi^f \in \mathcal{C}_b(\mathbb{T}_2)$ as well. Therefore, using Proposition 1.34, we can easily adapt the proof of Proposition 1.29 to show that X^α is a solution of the extended martingale problem. \square

We finish by giving the proof of Corollary 1.3 which gives representations of the sample subtree mass distribution of the α -Ford algebraic measure tree for general $\alpha \in [0, 1]$. They are given by recursive relations on the moments of the subtree mass distribution. For $\underline{k} = (k_1, k_2, k_3) \in \mathbb{N}^3$, we define $f^{\underline{k}}: \Delta_2 \rightarrow \mathbb{R}$ by

$$f^{\underline{k}}(\underline{\eta}) = \eta_1^{k_1} \eta_2^{k_2} \eta_3^{k_3}. \quad (1.139)$$

Proof of Corollary 1.3. If we have $\underline{k} = (0, 0, 0)$, then (1.16) is just $\mathbb{E}[\Phi^{f^{(0,0,0)}}(\chi^\alpha)] = 1$ and whenever $\underline{k} \in \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$, we have by symmetry

$$\mathbb{E}[\Phi^{f^{(1,0,0)}}(\chi^\alpha)] = \mathbb{E}[\Phi^{f^{(0,1,0)}}(\chi^\alpha)] = \mathbb{E}[\Phi^{f^{(0,0,1)}}(\chi^\alpha)] = \frac{1}{3}, \quad (1.140)$$

for all $\alpha \in [0, 1]$, so (1.16) holds as well. Choose $\underline{k} \in \mathbb{N}_0^3 \setminus \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1)\}$, and let $\alpha \in [0, 1]$. As for all $\underline{\eta} \in (0, 1)^3$ and $i = 1, 2, 3$,

$$\partial_i f^{\underline{k}}(\underline{\eta}) = \mathbf{1}_{\{k_i \neq 0\}} k_i f^{\underline{k} - e_i}(\underline{\eta}), \quad (1.141)$$

and all $i, j \in \{1, 2, 3\}$,

$$\partial_{i,j} f^{\underline{k}}(\underline{\eta}) = \mathbf{1}_{\{k_i, k_j \neq 0\}} (k_i - \delta_{i,j}) k_j f^{\underline{k} - e_i - e_j}(\underline{\eta}), \quad (1.142)$$

it follows that for all $\underline{\eta} \in (0, 1)^3$,

$$\begin{aligned} \Omega_\alpha \Phi^{f^{\underline{k}}}(\chi) &= \int \mu^{\otimes 3}(\underline{u}) \left(\sum_{i,j=1}^3 \mathbf{1}_{\{k_i, k_j \neq 0\}} (k_i - \delta_{i,j}) k_j (\delta_{ij} f^{\underline{k} - e_i}(\underline{\eta}) - f^{\underline{k}}(\underline{\eta})) \right. \\ &\quad + (2 - \alpha) \sum_{i=1}^3 \mathbf{1}_{\{k_i \neq 0\}} k_i (f^{\underline{k} - e_i}(\underline{\eta}) - 3f^{\underline{k}}(\underline{\eta})) \\ &\quad + (2 - 3\alpha) \sum_{1 \leq i < j \leq 3} \mathbf{1}_{\{k_i = k_j = 0\}} - 3(2 - 3\alpha) f^{\underline{k}}(\underline{\eta}) \\ &\quad + \frac{\alpha}{2} \sum_{i=1}^3 \mathbf{1}_{\{k_i = 0\}} \sum_{j=1; j \neq i}^3 \binom{k_j}{p_j} f^{\underline{k} + (p_j - 1)e_i - p_j e_j}(\underline{\eta}) \\ &\quad \left. - \alpha \sum_{i=1}^3 \mathbf{1}_{\{k_i \neq 0\}} \sum_{j=1; j \neq i}^3 \binom{k_j}{p_j} f^{\underline{k} - e_i}(\underline{\eta}) \right). \end{aligned} \quad (1.143)$$

By Proposition 1.32, $\Omega_\alpha \Phi^{f^{\underline{k}}}(\chi^\alpha) = 0$ for all \underline{k} , so that

$$\begin{aligned}
 \mathbb{E} \left[\Phi^{f^{\underline{k}}}(\chi^\alpha) \right] & \left(\sum_{i,j=1}^3 \mathbf{1}_{\{k_i, k_j \neq 0\}} (k_i - \delta_{i,j}) k_j + 3(2 - \alpha) \sum_{i=1}^3 \mathbf{1}_{\{k_i \neq 0\}} k_i + 3(2 - 3\alpha) \right) \\
 &= \left(\sum_{i,j=1}^3 \mathbf{1}_{\{k_i, k_j \neq 0\}} (k_i - \delta_{i,j}) k_j \delta_{ij} + (2 - \alpha) \sum_{i=1}^3 \mathbf{1}_{\{k_i \neq 0\}} k_i \right) \mathbb{E} \left[\Phi^{f^{\underline{k} - e_i}}(\chi^\alpha) \right] \\
 &+ (2 - 3\alpha) \sum_{1 \leq i < j \leq 3} \mathbf{1}_{\{k_i = k_j = 0\}} \\
 &+ \frac{\alpha}{2} \sum_{i=1}^3 \mathbf{1}_{\{k_i = 0\}} \sum_{j=1; j \neq i}^3 \binom{k_j}{p_j} \mathbb{E} \left[\Phi^{f^{\underline{k} + (p_j - 1)e_i - p_j e_j}}(\chi^\alpha) \right].
 \end{aligned} \tag{1.144}$$

Note that

$$\begin{aligned}
 & \sum_{i,j=1}^3 \mathbf{1}_{\{k_i, k_j \neq 0\}} (k_i - \delta_{i,j}) k_j + 3(2 - \alpha) \sum_{i=1}^3 \mathbf{1}_{\{k_i \neq 0\}} k_i + 3(2 - 3\alpha) \\
 &= \left(\sum_{i=1}^3 k_i \right)^2 - \sum_{i=1}^3 k_i + 3(2 - \alpha) \sum_{i=1}^3 k_i + 3(2 - 3\alpha) \\
 &= \left(\sum_{i=1}^3 k_i + 3 \right) \left(\sum_{i=1}^3 k_i + (2 - 3\alpha) \right)
 \end{aligned} \tag{1.145}$$

and

$$\begin{aligned}
 & \sum_{i,j=1}^3 \mathbf{1}_{\{k_i, k_j \neq 0\}} (k_i - \delta_{i,j}) k_j \delta_{ij} + (2 - \alpha) \sum_{i=1}^3 \mathbf{1}_{\{k_i \neq 0\}} k_i - \alpha \sum_{i=1}^3 \mathbf{1}_{\{k_i \neq 0\}} \\
 &= \sum_{i=1}^3 \mathbf{1}_{\{k_i \neq 0\}} (k_i + 1) (k_i - \alpha),
 \end{aligned} \tag{1.146}$$

which finishes the proof. \square

For $k_2 = k_3 = 0$, the recurrence relation (1.16) becomes

$$\begin{aligned}
 \mathbb{E} \left[\Phi^{f^{(k_1, 0, 0)}}(\chi^\alpha) \right] &= \frac{1}{(k_1 + 3)(k_1 + 2 - 3\alpha)} \left((k_1 + 1)(k_1 - \alpha) \mathbb{E} \left[\Phi^{f^{(k_1 - 1, 0, 0)}}(\chi^\alpha) \right] \right. \\
 &\quad \left. + (2 - 3\alpha) + \alpha \sum_{p=1}^{k_1} \binom{k_1}{p} \mathbb{E} \left[\Phi^{f^{(k_1 - p, p - 1, 0)}}(\chi^\alpha) \right] \right),
 \end{aligned} \tag{1.147}$$

and we get in particular

$$\begin{aligned}
 \mathbb{E} \left[\Phi^{f^{(1, 0, 0)}}(\chi^\alpha) \right] &= \frac{1}{3}, \\
 \mathbb{E} \left[\Phi^{f^{(2, 0, 0)}}(\chi^\alpha) \right] &= \frac{1}{5}, \\
 \mathbb{E} \left[\Phi^{f^{(3, 0, 0)}}(\chi^\alpha) \right] &= \frac{11 - 7\alpha}{15(5 - 3\alpha)}, \\
 \mathbb{E} \left[\Phi^{f^{(4, 0, 0)}}(\chi^\alpha) \right] &= \frac{37 - 25\alpha}{63(5 - 3\alpha)}, \\
 \mathbb{E} \left[\Phi^{f^{(5, 0, 0)}}(\chi^\alpha) \right] &= \frac{145 - 165\alpha + 44\alpha^2}{42(5 - 3\alpha)(7 - 3\alpha)}.
 \end{aligned} \tag{1.148}$$

Remark 1.36 (Löhr's conjecture). While the first moment being $\frac{1}{3}$ simply follows from the symmetry of the sampling in the mass polynomials, we do not yet have an argument to explain why the second moment does not depend on the parameter α . More generally, Wolfgang Löhr conjectured that this

holds for all continuum algebraic measure trees, even deterministic ones. That is, for all $\chi = (T, c, \mu) \in \mathbb{T}_2^{\text{cont}}$,

$$\Phi^{f^{(2,0,0)}}(\chi) = \int_{T^3} \mu^{\otimes 3}(\mathrm{d}\underline{u}) (\eta(\underline{u}))^2 = \frac{1}{5}. \quad (1.149)$$

Resampling dynamics on metric two-level measure trees

2.1 Introduction

Multi-level models in population dynamics can be found for example in ecology to account for the levels of individuals and species, or in epidemiology for the modelling of hosts and parasites. The phylogenies of such systems bring an important light on the paths and dynamical phenomenas that are interplaying. For example in anthropology these models have served as basis for statistical studies and understand collected data (e.g. [BJ10, JBA19, LBP⁺21, VAE⁺09]). For cell models, we can refer for example to works by Kimmel [Kim97]. The dynamics on each level do not need to be symmetric. In system of species for example, speciation can be seen as a result of underlying dynamics at the individual level, such as mutations and selection for instance.

In the mathematical literature, we can find several two-level dynamics modelling various systems: for instance particles grouped in so-called superparticles and subject to a birth-and-death process [DHW90, Wu91, GHW95, DHV96, GH00, DGW04], parasites inside a population of cells which divide [BT11], individuals carrying two types of cells [MR13], a population undergoing mutation, selection, resampling and migration mechanisms, and divided in colonies that are also subject to selection and resampling mechanisms [Daw18]. All these models have in common that the population is represented by a measure or a two-level measure, that is, a measure on the set of measures. But different approaches are used to define the population dynamics, such as the space of marked measure-valued processes (see e.g. [DH91, Eth93, MR13, GdHKK14, BGK⁺21]), branching diffusions or random walks [Kim97, Ban08, BT11, OW20]. We will rather use the idea of [DHV96, Wu94, DDV95, Daw18], where the two-level population is described by a measure on the set of measures.

The dynamics we study in this paper is fitted for a host-parasite system, that is, the resampling dynamics at the lower level, i.e. of parasites is constrained by the upper level. The host-parasite population will be represented by a probability measure $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$ on the set of probability Borel measures of a set X . For example, we can represent a population of parasites in M hosts by the two-level measure

$$\nu := \frac{1}{M} \sum_{i=1}^M \delta_{\tilde{\mu}_i} = \frac{1}{M} \sum_{i=1}^M \delta_{\frac{1}{N} \sum_{j=1}^N \delta_{(i,j)}}, \quad (2.1)$$

where $(i, j) \in X$ represents the parasite j of the host i . The measure ν allows for a two-level sampling: first we sample with ν a host $\mu \in \mathcal{M}_1(X)$, and then with μ we sample a parasite within the sampled host. This is related to a two-stage experiment in applications: if one wants to gain information about a parasite population, one would first sample hosts and then sample parasites within these hosts.

Additionally, our purpose is to describe the evolution of genealogies. In order to keep track of the genealogical structure in the population, several approaches have been developed. In [DK96, DK99a, DK99b], the so-called *look-down processes* contain (in an implicit way) all information about the genealogies. For this, one relies on labels as coordinates, which can be constraining as it requires the

population model to be exchangeable. Otherwise, the *historical processes* of [DP91, GLW05] do not use labels to encode the genealogical relationships but this approach only applies to spatially structured populations.

In this paper, we rather adopt the point of view of [GPW09], where the metric defined by the genealogical distances between individuals (see [Eva00]) was added to the measure-valued approach to encode trees as *metric measure spaces* (X, r, μ) . As we are interested in a two-level dynamics, we will rely on the extension of this theory in [Mei19] to a hierarchical setup. A *metric two-level measure space* (X, r, ν) (*m2m space* for short) is defined as a Polish metric space (X, r) equipped with a probability measure $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$ on the set of probability measures on X . The host-parasite population is represented by ν and the metric r encodes the ancestral tree of the parasites. In the theory of two-level measures, the *intensity measure* $M_\nu \in \mathcal{M}_1(X)$ plays a key role to adapt results from the one-level setup. It is defined by

$$M_\nu(\cdot) := \int \nu(d\mu) \mu(\cdot). \quad (2.2)$$

If ν is given by (2.1), then

$$M_\nu = \frac{1}{M} \sum_{i=1}^M \frac{1}{N} \sum_{j=1}^N \delta_{(i,j)}. \quad (2.3)$$

In this case, the intensity measure allows to sample a parasite uniformly at random from the whole population, regardless of their hosts. On the set $\mathbb{M}^{(2)}$ of (equivalence classes of) m2m spaces, the *two-level Gromov-weak topology* is introduced as the topology induced by test functions $\Phi : \mathbb{M}^{(2)} \rightarrow \mathbb{R}$ of the form

$$\Phi((X, r, \nu)) = \int \nu^{\otimes m}(d\underline{\mu}) \int \bigotimes_{i=1}^m \mu_i^{\otimes n_i}(d\underline{x}_i) \phi \circ R(\underline{x}), \quad (2.4)$$

where $m \in \mathbb{N}$, $\underline{n} = (n_1, \dots, n_m) \in \mathbb{N}^m$ and ϕ is a real bounded continuous function on $\mathbb{R}_+^{|\underline{n}| \times |\underline{n}|}$. In the Gromov-weak topology, a sequence of trees converges to a limit tree if and only if all randomly sampled finite subtrees converge to the corresponding limit subtrees. This topology on $\mathbb{M}^{(2)}$ is separable and metrizable, e.g. by the so-called two-level Gromov-Prohorov metric. Furthermore, the set of functions of the form (2.4) separates m2m spaces and is even convergence determining, which can be used to define Markov processes on $\mathbb{M}^{(2)}$ through well-posed martingale problems.

In this paper, the space $\mathbb{M}^{(2)}$ of metric two-level measure spaces serves as state space to study a host-parasite population undergoing resampling dynamics at each level. We rely for that on the Moran model [Mor58], a very simple resampling dynamics, that aroused a significant interest since its introduction. It models the evolution of a finite population of fixed size, from which each pair of individuals resamples at constant rate. At a resampling event, one individual is chosen at random from the pair, and replaces the other individual by a copy (or clone) of herself. In the two-level model we consider, resampling of hosts (at the upper level) occurs as for the usual Moran model and at the lower level, only pairs of parasites belonging to the same hosts can resample. It is a tree-valued version of the dynamics found in [Daw18] where we focus strictly on the resampling mechanisms. Our results extend to a two-level setup the tree-valued resampling dynamics described in [GPW13], namely *tree-valued Fleming-Viot dynamics* and its particle approximation the *tree-valued Moran dynamics*.

We define the two-level tree-valued Fleming-Viot and Moran dynamics as stochastic processes with values in $\mathbb{M}^{(2)}$. On each level, resampling events might occur (see Figure 2.1):

- **Resampling of hosts:** at rate γ_H , a resampling event occurs between any two hosts. One host dies (with all the parasites it contains) and is replaced by a copy of the other host (with all the parasites it contains).
- **Resampling of parasites:** at rate γ_P , any two parasites belonging to the same host resample. One parasite dies and is replaced by a copy of the other.

We first give a pathwise construction of the two-level tree-valued Moran process following this dynamics for a population with a fixed number of hosts each containing the same number of parasites. We then define a corresponding operator acting on test functions of the form (2.4) with three components, one

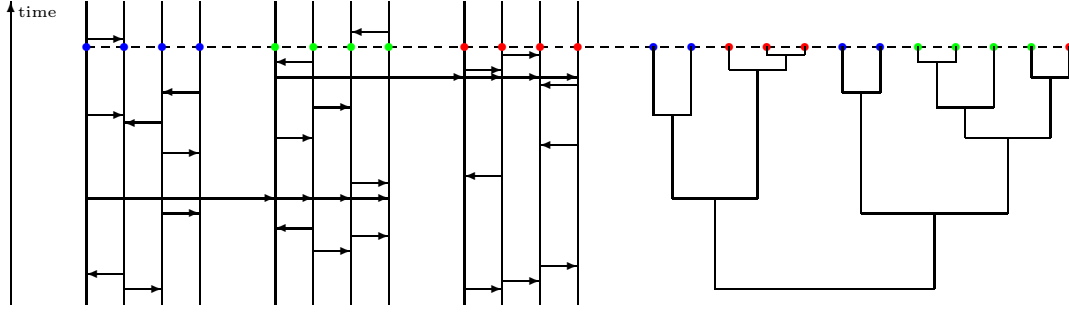


Figure 2.1: The graphical representation of the two-level Moran model with $M = 3$ hosts and $N = 4$ parasites in each host. An arrow from (i, k) to (i, l) indicates a parasite resampling event within host i , where (i, l) dies and (i, k) reproduces. For the resampling of hosts k and l , we draw N arrows, from (k, j) to (l, j) for each $1 \leq j \leq N$. The different colors represent different hosts.

for the growth of distances between distinct parasites (belonging to the same host or not), and one for the resampling at each level. We also introduce a similar operator for the diffusion dynamics, that we obtain as uniform limit of the operators for the finite dynamics.

Our main result states that when the numbers of hosts and of parasites simultaneously go to infinity, the *two-level tree-valued Moran process* converges weakly to the unique solution of a well-posed martingale problem. We call *two-level tree-valued Fleming-Viot process* this unique solution. For this, we showed that the sequence of two-level tree-valued Moran processes satisfies a compact containment condition and is hence tight. To show the uniqueness of the solution, we reformulated, in terms of two-level measure trees, the well-known duality between the Fleming-Viot process and the Kingman coalescent, which is the nested Kingman coalescent measure tree [Mei19, Section 10]. More general nested models of trees within trees have been investigated in [BDLS18, BRSSJ19, Duc20, LS20] to account for the evolutionary events at both the level of genes and species.

With the tools provided in [Mei19] for the extension to two-level models, the proofs are very similar to the ones for the (one-level) tree-valued resampling dynamics in [GPW13]. However, a major obstacle arises when expressing the diffusion limit operator in a convenient way for studying the evolution of length statistics. We were still able to express formulas for the evolution of the length distribution of the subtrees spanned by small samples of parasites and we compared them to the ones obtained in [GPW13]. But when the number of hosts involved in the sampled gets larger, it becomes quickly difficult to write more general formulas. The formulas on statistics we obtained also provided information about the length of random subtrees in the nested Kingman coalescent, since it is the unique equilibrium distribution of the two-level Fleming-Viot dynamics. In particular, it allows us to estimate the rates of coalescence at each level on a real world genealogical tree, which we assume a priori to be distributed according to a nested coalescent tree.

Outline. We start in Section 2.2 by describing the state space $\mathbb{M}^{(2)}$ and recalling some results from [Mei19]. We use it to define the particle approximation in Section 2.3 and show in Theorem 2.11 that the sequence of two-level tree-valued Moran processes is tight. In Section 2.4, we construct the diffusion limit of the Moran model as the unique solution of a well-posed martingale problem in Theorem 2.14. In particular, we give a duality relation of the two-level tree-valued Fleming-Viot dynamics to the nested Kingman coalescent. We study the subtree length distribution under these dynamics and at equilibrium in Section 2.5. Finally, we gather in Section 2.6 ingredients for the proofs of the two main results.

2.2 State space: metric two-level measure spaces

We consider trees as metric spaces together with a *two-level probability measure*, which can be seen as random probability measures as defined below. This measure allows a two-level sampling, that is, the sampling of a host and then the sampling of a parasite within the sampled host. The set of metric

two-level measure spaces was investigated in [Mei19]. We give here some definitions and basic results.

Given a metric space (X, r) we denote by $\mathcal{M}_1(X)$ the space of all probability measures defined on the Borel- σ -algebra of X . As our purpose here is to deal with two-level Moran models, we can restrict to probability measures in the whole paper. Recall that the support $\text{supp}(\mu)$ of $\mu \in \mathcal{M}_1(X)$ is the smallest closed set $X_0 \subseteq X$ such that $\mu(X_0) = 1$. We will deal with *two-level probability measures* of the form $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$. To fix ideas, let us describe the case of a population with a finite number of hosts each containing the same number of parasites, as in the two-level Moran model that will interest us.

Example 2.1. Let $M \in \mathbb{N}$ be the number of hosts and for $i \in \{1, \dots, M\}$, let $N \in \mathbb{N}$ be the number of parasites in each host. Then, the parasite population in the host i is represented by the measure

$$\tilde{\mu}_i = \frac{1}{N} \sum_{j=1}^N \delta_{x_{ij}}, \quad (2.5)$$

where $x_{ij} \in X$ represents the parasite j of the host i , and the entire two-level population is:

$$\nu = \frac{1}{M} \sum_{i=1}^M \delta_{\tilde{\mu}_i} = \frac{1}{M} \sum_{i=1}^M \delta_{\frac{1}{N} \sum_{j=1}^N \delta_{x_{ij}}}. \quad (2.6)$$

We denote by $\mathcal{M}_1^N(X)$ the set of measures of the form (2.5), i.e., for $N \in \mathbb{N}$,

$$\mathcal{M}_1^N(X) := \left\{ \frac{1}{N} \sum_{j=1}^N \delta_{x_j} : x_1, \dots, x_N \in X \right\}, \quad (2.7)$$

and in particular, if $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$ is of the form (2.6), then $\nu \in \mathcal{M}_1^M(\mathcal{M}_1^N(X))$.

In the sequel, we will introduce the two-level Moran dynamics with the $x_{ij} = (i, j)$ being the labels given to the parasites. In particular, the $x_{ij} = (i, j)$ will be distinct and the support of the measures $(\tilde{\mu}_i)_i$ representing the hosts will also be disjoint. However, notice that the space $\mathcal{M}_1(\mathcal{M}_1(X))$ contains much more general two-level measures than ν as in (2.6). For instance, we could consider the space $X = \{x_1, x_2, x_3\}$ and the two-level measure

$$\nu = \frac{1}{2} (\delta_{\frac{1}{2}\delta_{x_1} + \frac{1}{2}\delta_{x_2}} + \delta_{\frac{1}{2}\delta_{x_1} + \frac{1}{2}\delta_{x_3}}). \quad (2.8)$$

□

For a two-level measure $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$, the *intensity measure*, also called *first moment measure*, is the Borel measure on X defined by

$$M_\nu(\cdot) = \int_{\mathcal{M}_1(X)} \mu(\cdot) \nu(d\mu). \quad (2.9)$$

Example 2.2. For a measure ν of the form (2.6), the corresponding intensity measure is a probability measure on X given by:

$$M_\nu = \frac{1}{M} \sum_{i=1}^M \frac{1}{N} \sum_{j=1}^N \delta_{x_{ij}}.$$

This measure can be seen as the parasite population irrespectively of their hosts. In particular, all the x_{ij} have the same weight, so that M_ν allows to sample a parasite uniformly at random from the whole population, regardless of their hosts. □

The push forward of a measure $\mu \in \mathcal{M}_1(X)$ under a measurable map φ from X into another metric space Y is the probability measure $\varphi_*\mu \in \mathcal{M}_1(Y)$ defined by $\varphi_*\mu = \mu \circ \varphi^{-1}$. In the same way, the *two-level push forward* of a measure $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$ under a measurable map φ from X into Y is the probability two-level measure $\varphi_{**}\nu \in \mathcal{M}_1(\mathcal{M}_1(Y))$ defined by

$$\varphi_{**}\nu = \nu \circ (\varphi_*)^{-1}, \quad (2.10)$$

where φ_* is the (one-level) push-forward operator of φ . We denote by $\mathcal{B}(X)$ and $\mathcal{C}_b(X)$ the bounded real-valued functions on X which are measurable and continuous, respectively.

A triple (X, r, ν) is called a *metric two-level measure space (m2m space)* if $X \subset \mathbb{R}^{\mathbb{N}}$ is non-empty, (X, r) is a Polish metric space and $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$. Two m2m spaces (X, r, ν) and (Y, d, λ) are called *equivalent* if there exists a measurable function $\varphi: X \rightarrow Y$ such that $\lambda = \overline{\varphi_{**}\nu}$ and φ is isometric on the set $\text{supp}(M_\nu)$ (but not necessarily on the whole space X). We write $\overline{(X, r, \nu)}$ for the equivalence class of a m2m space (X, r, μ) , but we will often use (X, r, ν) for the equivalence class. Now define the set of (equivalence classes of) metric two-level measure spaces

$$\mathbb{M}^{(2)} := \{\chi = \overline{(X, r, \mu)} : (X, r, \mu) \text{ metric two-level measure space}\}. \quad (2.11)$$

Notice that when we will consider equivalence classes in the sequel, we will no longer keep track of the labels of the parasites (in the sense of Example 2.1), but only of the structure of the (genealogical) trees.

If (X, r, ν) is such that r is only a pseudo-metric on X (i.e. $r(x, y) = 0$ is possible for $x \neq y$), we can still define its equivalence class. Since the class of such a pseudo-metric two-level measure space contains also m2m spaces, there is a bijection between the equivalence classes of pseudo-metric two-level measure spaces and the ones of m2m spaces. Therefore, we will use both notions interchangeably.

Example 2.3 (A pseudo-m2m space equivalent to an m2m space). Consider the two following spaces: $X_1 := \{(1, 1), (1, 2), (2, 1), (2, 2), (3, 1), (3, 2)\}$ and $X_2 := \{(1, 1), (1, 2), (3, 1)\}$. On X_1 , define the pseudo-metric r_1 by

$$(r_1((i, j), (i', j'))_{(i, j), (i', j') \in X_1} := \begin{pmatrix} 0 & 1 & 0 & 1 & 2 & 2 \\ 1 & 0 & 1 & 0 & 2 & 2 \\ 0 & 1 & 0 & 1 & 2 & 2 \\ 1 & 0 & 1 & 0 & 2 & 2 \\ 2 & 2 & 2 & 2 & 0 & 0 \\ 2 & 2 & 2 & 2 & 0 & 0 \end{pmatrix} \quad (2.12)$$

and the two-level measure

$$\nu_1 := \frac{1}{3} \left(\delta_{\frac{1}{2}(\delta_{(1,1)} + \delta_{(1,2)})} + \delta_{\frac{1}{2}(\delta_{(2,1)} + \delta_{(2,2)})} + \delta_{\frac{1}{2}(\delta_{(3,1)} + \delta_{(3,2)})} \right). \quad (2.13)$$

On X_2 , define the metric r_2 by

$$(r_2((i, j), (i', j'))_{(i, j), (i', j') \in X_2} := \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 2 \\ 2 & 2 & 0 \end{pmatrix} \quad (2.14)$$

and the two-level measure

$$\nu_2 := \frac{2}{3} \delta_{\frac{1}{2}(\delta_{(1,1)} + \delta_{(1,2)})} + \frac{1}{3} \delta_{\delta_{(3,1)}}. \quad (2.15)$$

Defined this way, the (pseudo)-m2m space (X_1, r_1, ν_1) and the m2m space (X_2, r_2, ν_2) are equivalent (see Figure 2.2). \square

Distance matrix distribution. We now introduce the distance matrix distribution of m2m spaces, which is heuristically the distribution of pairwise distances for a sequence of points sampled independently, in two stages, according to the two-level measure ν . We will see that this distribution characterizes the m2m space.

For a metric space (X, r) we define the following *distance operator*:

$$R^{(X, r)} : \begin{array}{ccc} X^{\mathbb{N} \times \mathbb{N}} & \longrightarrow & \mathbb{R}^{\mathbb{N}^4} \\ (x_{ij})_{(i, j) \in \mathbb{N}^2} & \longmapsto & (r(x_{ij}, x_{kl}))_{(i, j), (k, l) \in \mathbb{N}^4} \end{array} \quad (2.16)$$

We use double indices (i, j) to take into account the two-level sampling we will consider. More precisely, the index i will correspond to the sampled host, and j to the sampled parasite within the sampled host i . We then define the *distance matrix distribution* of m2m space (X, r, ν) by

$$R^{(X, r)}_* (M_\nu^{\infty, \infty}), \quad (2.17)$$

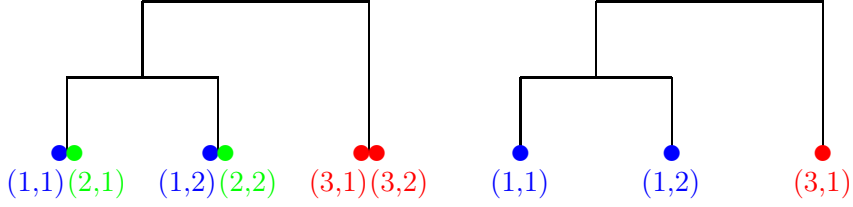


Figure 2.2: The two (pseudo)-m2m spaces (X_1, r_1, ν_1) (on the left) and (X_2, r_2, ν_2) (on the right) of Example 2.3 are equivalent, with $\nu_1 = \frac{1}{3}(\delta_{\frac{1}{2}(\delta_{(1,1)}+\delta_{(1,2)})} + \delta_{\frac{1}{2}(\delta_{(2,1)}+\delta_{(2,2)})} + \delta_{\frac{1}{2}(\delta_{(3,1)}+\delta_{(3,2)})})$ and $\nu_2 = \frac{2}{3}\delta_{\frac{1}{2}(\delta_{(1,1)}+\delta_{(1,2)})} + \frac{1}{3}\delta_{(3,1)}$.

where $M_\nu^{\infty, \infty} \in \mathcal{M}_1(X^{\mathbb{N} \times \mathbb{N}})$ is the *infinite mixed moment measure* of ν defined by

$$M_\nu^{\infty, \infty}(\cdot) = \int \nu^{\otimes \mathbb{N}}(d\mu) \bigotimes_{i=1}^{\infty} \mu_i^{\otimes \mathbb{N}}(\cdot). \quad (2.18)$$

Since the distance matrix distribution $R^{(X,r)}_*(M_\nu^{\infty, \infty})$ of (X, r, ν) is constant for all elements of its equivalence class, we can define the distance matrix distribution of elements of $\mathbb{M}^{(2)}$.

From [Mei19, Section 3], m2m spaces are uniquely determined by their distance matrix distribution in the sense that given a random infinite matrix with distribution $M_\nu^{\infty, \infty}$ and their mutual distances $(r(x_{ij}, x_{kl}))_{(i,j),(k,l) \in \mathbb{N}^4}$, it is possible to reconstruct the two-level measure ν and the metric space (X, r) . The reconstruction of ν is based on a Glivenko-Cantelli theorem for random probability measures (see [Mei19, Proposition 3.5], see also [Gro99, Section 3 $\frac{1}{2}$.7]). For this reason, we define the two-level Gromov-weak topology on $\mathbb{M}^{(2)}$ as follows.

We say that a sequence $(\chi_n)_{n \in \mathbb{N}} = (\overline{X_n, r_n, \nu_n})$ converges to $\chi = \overline{X, r, \nu}$ in two-level Gromov-weak topology if

$$R^{(X_n, r_n)}_*(M_{\nu_n}^{\infty, \infty}) \xrightarrow[n \rightarrow \infty]{} R^{(X, r)}_*(M_\nu^{\infty, \infty}) \quad (2.19)$$

in the weak topology on $\mathcal{M}_1(\mathbb{R}^{\mathbb{N}^4})$ (with $\mathbb{R}^{\mathbb{N}^4}$ equipped with the product topology). This topology is equivalent to the topology induced by the two-level Gromov-Prokhorov metric (see [Mei19, Definition 4.1, Theorem 8.1]) and one can show that the space $\mathbb{M}^{(2)}$ is Polish (see [Mei19, Proposition 4.6]).

Finally, we call (the equivalence class of) an m2m space (X, r, ν) *compact* whenever the metric space $(\text{supp}(M_\nu), r)$ is compact, and define

$$\mathbb{M}_c^{(2)} := \{\chi \in \mathbb{M}^{(2)} : \chi \text{ is compact}\}. \quad (2.20)$$

If $\chi = \overline{X, r, \nu}$ is a *finite* m2m space, i.e., $\#\text{supp}(M_\nu) < \infty$, then χ is compact. Moreover, every element of $\mathbb{M}^{(2)}$ can be approximated by a sequence of finite m2m spaces (see the proof of Proposition 4.6 in [Mei19]). Thus the sub-space $\mathbb{M}_c^{(2)}$ is not closed.

Relative compactness characterization in $\mathbb{M}^{(2)}$. For μ a probability Borel measure on a Polish metric space (X, r) , we define the *distance distribution* $w_\mu \in \mathcal{M}_1(\mathbb{R}_+)$ of μ by

$$w_\mu := r_* \mu^{\otimes 2}. \quad (2.21)$$

By comparing the relative compactness criteria of [Mei19, Theorem 7.2] for $\mathbb{M}^{(2)}$ and the ones of [GPW09, Proposition 7.1] for the space \mathbb{M} of metric (probability) measure spaces (see also [Kle08, EW06]), it is easy to see that the following holds (in the case of metric two-level *probability* measure spaces).

Proposition 2.4. *A set $\Gamma \subset \mathbb{M}^{(2)}$ is relatively compact in the two-level Gromov-weak topology if and only if the set*

$$\left\{ \overline{(X, r, M_\nu)} \in \mathbb{M} : \overline{(X, r, \nu)} \in \Gamma \right\} \quad (2.22)$$

is relatively compact in the Gromov-weak topology in \mathbb{M} .

It follows that the results stated in [GPW13, Propositions 6.1,6.2] can be adapted to our two-level setting by replacing the (one-level) measure μ by the intensity measure M_ν .

Corollary 2.5. *A set $\Gamma \subset \mathbb{M}^{(2)}$ is relatively compact in the two-level Gromov-weak topology if and only if the following two conditions hold:*

- (i) $\{w_{M_\nu} : (X, r, \nu) \in \Gamma\}$ is relatively compact in $\mathcal{M}_1(\mathbb{R}_+)$.
- (ii) For all $\epsilon > 0$, there exists $C_\epsilon > 0$ such that $\sup_{\mathcal{X} \in \Gamma} \tilde{S}_\epsilon(\mathcal{X}) \leq C_\epsilon$, where for $\mathcal{X} = (X, r, \nu) \in \mathbb{M}^{(2)}$,

$$\tilde{S}_\epsilon(\mathcal{X}) := \min \left\{ K : \exists x_1, \dots, x_K \in X : M_\nu \left(\bigcup_{k=1}^K B_\epsilon(x_k) \right) > 1 - \epsilon \right\}. \quad (2.23)$$

Proposition 2.6. *A set $\Gamma \subset \mathbb{M}_c^{(2)}$ is relatively compact in the two-level Gromov-weak topology if the following two conditions hold:*

- (i) $\{w_{M_\nu} : (X, r, \nu) \in \Gamma\}$ is relatively compact in $\mathcal{M}_1(\mathbb{R}_+)$.
- (ii) For all $\epsilon > 0$, there exists $N_\epsilon \in \mathbb{N}$ such that $\sup_{\mathcal{X} \in \Gamma} S_\epsilon(\mathcal{X}) \leq N_\epsilon$, where, for $\mathcal{X} = (X, r, \nu) \in \mathbb{M}^{(2)}$, $S_\epsilon(\mathcal{X})$ denotes the minimal number of open ϵ -balls needed to cover $\text{supp}(M_\nu)$.

Ultra-metric and compact m2m spaces. Among m2m spaces, a particular attention will be given in the sequel to ultra-metric spaces. If the m2m space (X, r, ν) is such that for all $x, y, z \in \text{supp}(M_\nu)$,

$$r(x, z) \leq r(x, y) \vee r(y, z), \quad (2.24)$$

then we call its equivalence class $\overline{(X, r, \nu)}$ *ultra-metric*. We define

$$\mathbb{U}^{(2)} := \{\chi \in \mathbb{M}^{(2)} : \chi \text{ is ultra-metric}\}. \quad (2.25)$$

We introduce this notion because ultra-metric spaces are closely related to \mathbb{R} -trees, i.e., complete path-connected metric spaces (X, r_X) which satisfy the *four-point condition*

$$r_X(x_1, x_2) + r_X(x_3, x_4) \leq \max\{r_X(x_1, x_3) + r_X(x_2, x_4), r_X(x_1, x_4) + r_X(x_2, x_3)\}, \quad (2.26)$$

for all $x_1, x_2, x_3, x_4 \in X$ (see, for example, [Dre84, Ter97]). Indeed, every metric space that satisfy the four-point condition can be isometrically embedded in an \mathbb{R} -tree (see Lemma 3.12 and Theorem 3.38 in [Eva08]) and it is easy to see that ultra-metric spaces satisfy the four-point conditions. Reciprocally, if (X, r_X) is an \mathbb{R} -tree and ρ is a distinguished point in X , then for all $t \geq 0$, the level set $X^t := \{x \in X : r(\rho, x) = t\}$ is an ultra-metric space.

Since the set $\mathbb{U}^{(2)}$ is closed in $\mathbb{M}^{(2)}$ (same as Lemma 2.3 in [GPW13]), the set $\mathbb{U}^{(2)}$ equipped with the two-level Gromov-weak topology is again Polish.

Finally, we denote by

$$\mathbb{U}_c^{(2)} := \mathbb{U}^{(2)} \cap \mathbb{M}_c^{(2)} \quad (2.27)$$

the set of (equivalence classes of) ultra-metric spaces (X, r, ν) for which the metric space $(\text{supp}(M_\nu), r)$ is compact. As for $\mathbb{M}_c^{(2)}$, $\mathbb{U}_c^{(2)}$ is not closed.

2.3 Particle approximation

2.3.1 The tree-valued two-level Moran process

We now describe the Moran model as a stochastic process on the set of m2m spaces, which is the particle approximation of the two-level tree-valued Fleming-Viot process in the limit of large population size. This is already known in the case of measure-valued processes (see [FV79] and [Daw93] for a survey) and in the case of metric measure spaces (see [GPW13]).

In the classical Moran model, each pair of individuals, taken from a finite population of fixed size, resamples at constant rate. At a resampling event, one individual is chosen at random from the pair, and replaces the other individual by a copy (or clone) of herself. Here the main difference is the two-level resampling, one on hosts and one on parasites, inside the hosts. The resampling rates for hosts and parasites might be different.

Pathwise construction. The criterion for the compact containment condition we provide in Proposition 2.12 applies to sequences of m2m-spaces describing finite population dynamics in a general setting (stated in Definition 2.7). For example, the population size might not be constant and might even go extinct, which is not the case in the Moran model. Therefore, we first introduce this general class of finite population dynamics and then use it to give a pathwise construction of the tree-valued two-level Moran process. Let us introduce the notation for labelling the parasites and the hosts:

$$\mathbb{N}_{M,\underline{N}} := \{(i, j) : i \in \{1, \dots, M\}, j \in \{1, \dots, N_i\}\}, \quad (2.28)$$

with $M \in \mathbb{N}$ and $\underline{N} := (N_1, \dots, N_M) \in \mathbb{N}^M$, i denoting the host index and j the parasite index within a host. If $N_1 = \dots = N_M$, we will write $\mathbb{N}_{M,N} := \mathbb{N}_{M,\underline{N}}$. In some cases, we will use the letters α or β to denote an element of $\mathbb{N}_{M,\underline{N}}$ that is, a specific parasite within a host. In the following definition, each parasite is given a unique label $(i, j) \in \mathbb{N}_{M,\underline{N}}$.

Definition 2.7 (Finite population dynamics). Let $(\Omega, (\mathcal{A}_t)_{t \geq 0}, \mathbb{P})$ be a filtered probability space. Let $\mathcal{I} = (\mathcal{I}_t)_{t \geq 0}$ be an adapted process with values in $\{\mathbb{N}_{M,\underline{N}} : M \in \mathbb{N}, \underline{N} \in \mathbb{N}^m\}$. For each $t \in \mathbb{R}$, we refer to \mathcal{I}_t as the population at time t . Furthermore, let $\preceq = (\preceq_t)_{t \geq 0}$ be a family of partial order on

$$\{(\alpha, s) : s \in (-\infty, t], \alpha \in \mathcal{I}_s\} \quad (2.29)$$

which defines the genealogical relationships at all times before t . Suppose (\mathcal{I}, \preceq) satisfy the following:

- (i) For all $r, s, t \in \mathbb{R}$, with $r \leq s \leq t$, $\alpha_r \in \mathcal{I}_r$ and $\alpha_s \in \mathcal{I}_s$ such that $(\alpha_r, r) \preceq_s (\alpha_s, s)$, we also have $(\alpha_r, r) \preceq_t (\alpha_s, s)$. In other words, order relations from earlier times are preserved.
- (ii) For all $\alpha \in \mathcal{I}_t$ and $s \leq t$ there is a unique $A_s(\alpha, t) \in \mathcal{I}_s$ such that $(A_s(\alpha, t), s) \preceq_t (\alpha, t)$. We say that $A_s(\alpha, t)$ is the ancestor of α at time s .
- (iii) For all $\alpha, \beta \in \mathcal{I}_0$, there is an almost surely finite time $-\infty < T_{\alpha\beta}^0 \leq 0$ such that $A_{T_{\alpha\beta}^0}(\alpha, 0) = A_{T_{\alpha\beta}^0}(\beta, 0)$, i.e., all individuals at time $t = 0$ are related.

From such a process (\mathcal{I}, \preceq) , we construct a random process with values in the set of m2m spaces. Let $\tau := \inf\{s \geq 0 : \mathcal{I}_s = \emptyset\}$ be the lifetime of the population. Put then for all $t \leq \tau$ and $\alpha, \beta \in \mathcal{I}_t$,

$$r_t(\alpha, \beta) := 2(t - \sup\{s \leq t : A_s(\alpha, t) = A_s(\beta, t)\}). \quad (2.30)$$

The two-level tree-valued population dynamics $(\mathcal{U}_t)_{t \in [0, \tau]}$ reads off from (\mathcal{I}, \preceq) and is defined as follows: for all $t \in [0, \tau)$,

$$\mathcal{U}_t := (\overline{\mathcal{I}_t, r_t, \nu_{\mathcal{I}_t}}) \in \mathbb{U}^{(2)}, \quad (2.31)$$

where, for $\mathcal{I}_t = \mathbb{N}_{M,\underline{N}}$,

$$\nu_{\mathcal{I}_t} := \frac{1}{M} \sum_{i=1}^M \delta_{\frac{1}{N_i} \sum_{j=1}^{N_i} \delta_{(i,j)}}. \quad (2.32)$$

The tree-valued two-level Moran process is obtained for the following choice of (\mathcal{I}, \preceq) . In this model, the number of hosts M and the numbers of parasites per host N are constant and we consider the case where all hosts have the same number of parasites. The process evolves through resamplings of hosts and parasites.

Definition 2.8 (Two-level tree-valued Moran dynamics of population size (M, N)). Fix $M, N \in \mathbb{N}$. The two-level tree-valued population Moran dynamics with population size (M, N) , $(\mathcal{U}_t^{M,N})_{t \geq 0}$ is the two-level tree-valued population dynamics read off from (\mathcal{I}, \preceq) as in (2.31), where (\mathcal{I}, \preceq) is as follows. Let \preceq_0 be a partial order on $\mathbb{N}_{M,\underline{N}} \times (-\infty, 0]$ which satisfies Condition (iii) in Definition 2.7, almost surely. Consider also the two following families of Poisson processes η_H and η_P on (Ω, \mathbb{P}) :

- (a) Let $\eta_H := \{\eta_H^{k,l} : 1 \leq k, l \leq M\}$ be an independent family of rate $\frac{\gamma_H}{2}$ -Poisson processes. At time $\eta_H^{k,l}$, the host l is replaced by a copy of the host k , that is, for all $1 \leq j \leq N$, the parasite (l, j) is replaced by a copy of the parasite (k, j) .
- (b) Let $\eta_P := \{\eta_P^{(i,k),(i,l)} : 1 \leq i \leq M, 1 \leq k, l \leq N\}$ be an independent family of rate $\frac{\gamma_P}{2}$ -Poisson processes, independent the family $\eta_H := \{\eta_H^{k,l} : 1 \leq k, l \leq M\}$. At time $\eta_P^{(i,k),(i,l)}$, the parasite (i, l) is replaced by a copy of the parasite (i, k) .

For any $s, t \in \mathbb{R}$ with $0 \leq s \leq t$ and $\alpha_s, \alpha_t \in \mathcal{I}^{M,N}$, we say that $(\alpha_s, s) \preceq_t (\alpha_t, t)$ if and only if there is path of descent from (α_s, s) to (α_t, t) , that is, there exists $n \in \mathbb{N}$, $s = u_0 \leq u_1 < u_2 < \dots < u_n = t$ and $\beta_0 = \alpha_s$, $\beta_n = \alpha_t$, $\beta_1, \dots, \beta_{n-1} \in \mathcal{I}^{M,N}$ such that for all $k \in \{1, \dots, n\}$, at time u_k the parasite α_{k-1} is replaced by a copy of the parasite α_k and during the time interval (u_{k-1}, u_k) , the parasite α_{k-1} is not replaced.

Through this definition, it is then possible to construct a graphical representation of the underlying phylogenies (see Figure 2.1) thanks to a popular graphical representation: the lookdown graph developed by Donnelly and Kurtz [DK99a]. The bridge between the lookdown approach and the representations of the underlying phylogenetical trees with ultrametric measure spaces has been developed with [Eva00, GPW09, Guf18] among others.

Remark 2.9. Note that resampling events in Definition 2.8 can happen in both directions, that is for $k \leq l$ and $k \geq l$. In particular, we write the rates $\gamma_H/2$ and $\gamma_P/2$ for the Poisson processes, so that γ_H and γ_P would be the rates of the nested Kingman coalescent which is dual to the diffusion limit (see Section 2.4.3).

A stochastic differential equation for the tree-valued two-level Moran dynamics. We introduced the two-level tree-valued Moran dynamics as a process whose values are the equivalence classes of triplets of the form

$$\left(\{1, \dots, M\} \times \{1, \dots, N\}, \tilde{r}, \frac{1}{M} \sum_{i=1}^M \delta_{\frac{1}{N} \sum_{j=1}^N \delta_{(i,j)}} \right), \quad (2.33)$$

where \tilde{r} is some pseudo-metric on $\mathbb{N}_{M,N}$. For this reason, the states of the two-level Moran dynamics with population size (M, N) belong to the set

$$\mathbb{U}_{M,N}^{(2)} := \left\{ \chi = \overline{(X, r, \nu)} \in \mathbb{U}^{(2)} : \nu \in \mathcal{M}_1^M(\mathcal{M}_1^N(X)) \right\}, \quad (2.34)$$

with $\mathcal{M}_1^N(X)$ defined in (2.5). Note that $\mathbb{U}_{M,N}^{(2)}$ is a subset of $\mathbb{U}_c^{(2)}$. Conversely, each equivalence class $\chi \in \mathbb{U}_{M,N}^{(2)}$ contains a representative of the form (2.33). For such a representative, the two-level tree-valued Moran dynamics can also be expressed as changes on \tilde{r} in the following way:

- **Distance growth** The distance between any two distinct parasites (belonging to the same host or not) grows at speed 2, where the parasites are all the $(i, j) \in \mathbb{N}_{M,N}$.
- **Resampling of parasites** For each host i and for each pair of parasites $(i, k), (i, l)$, at rate $\frac{\gamma_P}{2}$, the distances to (i, l) are replaced by the distances to (i, k) . In particular, the distances between (i, l) and (i, k) are set to be zero. At the time t of the resampling, the pseudo-metric jumps from \tilde{r}_{t-} to $\tilde{r}_t = \Theta_{(i,k),(i,l)}(\tilde{r}_{t-})$ where $\Theta_{(i,k),(i,l)}$ is defined by:

$$(\Theta_{(i,k),(i,l)}(\tilde{r}))((u, j), (u', j')) = \begin{cases} \tilde{r}((i, k), (u', j')) & \text{if } (u, j) = (i, l) \text{ and } (u', j') \neq (i, l) \\ \tilde{r}((u, j), (i, k)) & \text{if } (u', j') = (i, l) \text{ and } (u, j) \neq (i, l) \\ \tilde{r}((u, j), (u', j')) & \text{otherwise.} \end{cases} \quad (2.35)$$

- **Resampling of hosts.** For each pair of hosts k, l , at rate $\frac{\gamma_H}{2}$, the distances to (l, j) are replaced by the distances to (k, j) , for all $j \in \{1, \dots, N\}$. In particular, the distances between (l, j) and

(k, j) are set to be zero. At the time t of the resampling, the pseudo-metric jumps from \tilde{r}_{t-} to $\tilde{r}_t = \Theta^{k,l}(\tilde{r}_{t-})$ where $\Theta^{k,l}$ is defined by:

$$(\Theta^{k,l}(\tilde{r}))((i, j), (i', j')) := \begin{cases} \tilde{r}((k, j), (i', j')) & \text{if } i = l \text{ and } i' \neq l \\ \tilde{r}((i, j), (k, j')) & \text{if } i' = l \text{ and } i \neq l \\ \tilde{r}((i, j), (i', j')) & \text{otherwise.} \end{cases} \quad (2.36)$$

Using the Poisson point processes η_H and η_P , we can write a stochastic differential equation describing the evolution of the $\mathbb{U}_{M,N}^{(2)}$ -valued stochastic process $(\mathcal{U}_t^{M,N})_{t \geq 0}$. Considering again the pseudo-metric $\tilde{r}_t^{M,N}$ associated with $\mathcal{U}_t^{M,N}$ through (2.33), the equation is

$$\begin{aligned} \tilde{r}_t^{M,N} = & \tilde{r}_0^{M,N} + 2t + \sum_{1 \leq i \leq M} \sum_{1 \leq k, l \leq N} \int_0^t (\Theta_{(i,k),(i,l)}(\tilde{r}_{s-}^{M,N}) - \tilde{r}_{s-}^{M,N}) \eta_P^{(i,k),(i,l)}(ds) \\ & + \sum_{1 \leq k, l \leq M} \int_0^t (\Theta^{k,l}(\tilde{r}_{s-}^{M,N}) - \tilde{r}_{s-}^{M,N}) \eta_H^{k,l}(ds). \end{aligned} \quad (2.37)$$

For given M and N , and for given Poisson point processes η_H and η_P , the process $\mathcal{U}^{M,N}$ associated with $(\tilde{r}_t^{M,N})_{t \geq 0}$ is a jump process with a deterministic drift and admits a unique strong solution. The latter can be constructed algorithmically (see e.g. [FT09, DKL20]).

In order to write in the sequel the generators corresponding to this dynamics, we use that we can equivalently express the jumps at resampling events by keeping the distance constant and changing the two-level measure ν . For the replacement of parasite (i, l) by parasite (i, k) , the two-level measure changes from ν to

$$\vartheta_{(i,k),(i,l)}(\nu) := \nu - \frac{1}{M} \delta_{\frac{1}{N} \sum_{j=1}^N \delta_{(i,j)}} + \frac{1}{M} \delta_{\frac{1}{N} \sum_{j=1}^N \delta_{(i,j)} - \frac{1}{N} \delta_{(i,l)} + \frac{1}{N} \delta_{(i,k)}} \quad (2.38)$$

and for the replacement of parasite l by parasite k , the two-level measure changes from ν to

$$\vartheta^{k,l}(\nu) := \nu - \frac{1}{M} \delta_{\frac{1}{N} \sum_{j=1}^N \delta_{(l,j)}} + \frac{1}{M} \delta_{\frac{1}{N} \sum_{j=1}^N \delta_{(k,j)}}. \quad (2.39)$$

Notice that in this case, the modified representative is no longer of the form (2.33) (see also Example 2.3).

The operator of the tree-valued two-level Moran dynamics. We introduce now the operators corresponding to the two-level Moran dynamics described above. For finite populations, it is more convenient to proceed by sampling without repetition. We compute our generator for such sampling. To provide the convergence of the generators to the diffusion limit operator, we will use in Section 2.4.2 that sampling without or with repetition is asymptotically equivalent in large populations.

Fix $M, N \in \mathbb{N}$. For $\mu = \frac{1}{M} \sum_{i=1}^M \delta_{y_i}$ a probability measure on some space X and $m \leq M$, we denote by

$$\mu^{\otimes \downarrow m}(\underline{dx}) := \mu(dx_1) \otimes \frac{\mu - \frac{1}{M} \delta_{x_1}}{1 - \frac{1}{M}}(dx_2) \otimes \cdots \otimes \frac{\mu - \frac{1}{M} \sum_{i=1}^{m-1} \delta_{x_i}}{1 - \frac{m-1}{M}}(dx_m) \quad (2.40)$$

the *sampling (without replacement) measure*. Now if $m \leq M$ and if for all $1 \leq i \leq m$, $n_i \leq N$ we can define the polynomial (without replacement) $\Phi_{M,N}^{m,\underline{n},\phi}$, for all $\chi = (\overline{T}, r, \nu) \in \mathbb{U}_{M,N}^{(2)}$,

$$\Phi_{M,N}^{m,\underline{n},\phi}(\chi) = \langle \langle \nu, \phi \rangle \rangle_{\downarrow} := \int_{(\mathcal{M}_1(T))^m} \nu^{\otimes \downarrow m}(d\underline{\mu}) \int_{T^{\lfloor \underline{n} \rfloor}} \bigotimes_{i=1}^m \mu_i^{\otimes \downarrow n_i}(d\underline{u}_i) \phi \circ R(\underline{u}), \quad (2.41)$$

where $\phi \in \mathcal{C}_b(\mathbb{R}^{\lfloor \underline{n} \rfloor \times \lfloor \underline{n} \rfloor})$ and $R = R^{(X,r)}$ is the distance operator defined in (2.16). Define

$$\Pi_{M,N} := \left\{ \Phi_{M,N}^{m,\underline{n},\phi} : 1 \leq m \leq M, \forall i, 1 \leq n_i \leq N, \phi \in \mathcal{C}_b(\mathbb{R}^{\lfloor \underline{n} \rfloor \times \lfloor \underline{n} \rfloor}) \right\} \quad (2.42)$$

and

$$\Pi_{M,N}^1 := \left\{ \Phi_{M,N}^{m,\underline{n},\phi} : 1 \leq m \leq M, \forall i, 1 \leq n_i \leq N, \phi \in \mathcal{C}_b^1(\mathbb{R}^{|\underline{n}| \times |\underline{n}|}) \right\}, \quad (2.43)$$

where $\mathcal{C}_b^1(\mathbb{R}^{|\underline{n}| \times |\underline{n}|})$ is the space of bounded and continuously differentiable real-valued functions on $\mathbb{R}^{|\underline{n}| \times |\underline{n}|}$ with bounded derivatives. Notice that the set $\Pi_{M,N}^1$ separates point in $\mathbb{U}_{M,N}^{(2)}$.

We can now introduce the operator $\Omega_{M,N}^\uparrow$ that acts on polynomials in $\Pi_{M,N}^1$ as follows

$$\Omega_{M,N}^\uparrow = \Omega_{M,N}^{\uparrow,H} + \Omega_{M,N}^{\uparrow,P} + \Omega_{M,N}^{\uparrow,\text{grow}}, \quad (2.44)$$

where the distance growth operator $\Omega_{M,N}^{\uparrow,H}$ is defined as

$$\Omega_{M,N}^{\uparrow,\text{grow}} \Phi_{M,N}^{\phi,m,\underline{n}}(\chi) := \langle \langle \nu, \Lambda \phi \rangle \rangle_\downarrow, \quad (2.45)$$

with

$$\Lambda \phi := 2 \sum_{\substack{(i,j),(i',j') \in \mathbb{N}_{M,N} \\ (i,j) \neq (i',j')}} \frac{\partial \phi}{\partial r_{(i,j)(i',j')}}. \quad (2.46)$$

and the operators for the resampling of parasites and hosts are

$$\Omega_{M,N}^{\uparrow,P} \Phi_{M,N}^{\phi,m,\underline{n}}(\chi) = \frac{\gamma_P}{2} \sum_{i=1}^M \sum_{1 \leq k,l \leq N} \left(\langle \langle \vartheta_{(i,k),(i,l)}(\nu), \phi \rangle \rangle_\downarrow - \langle \langle \nu, \phi \rangle \rangle_\downarrow \right) \quad (2.47)$$

and

$$\Omega_{M,N}^{\uparrow,H} \Phi_{M,N}^{\phi,m,\underline{n}}(\chi) = \frac{\gamma_H}{2} \sum_{1 \leq k,l \leq M} \left(\langle \langle \vartheta^{k,l}(\nu), \phi \rangle \rangle_\downarrow - \langle \langle \nu, \phi \rangle \rangle_\downarrow \right), \quad (2.48)$$

where $\vartheta_{(i,k),(i,l)}$ and $\vartheta^{k,l}$ where defined in (2.38) and (2.39).

The following rewriting of the operators will be useful when showing the uniform convergence to the generator of the tree-valued two-level Fleming-Viot process. It states in particular that for the resampling operators, we can sum only on the indices of the sample appearing in the integrand. This is true because all the other terms are compensated when we substract averages over all the possible samples before and after the resampling. For $1 \leq k, l \leq m$, denote by $\theta^{k,l} : (\mathcal{M}_1(T))^m \rightarrow (\mathcal{M}_1(T))^m$ the replacement map defined by

$$\theta^{k,l}(\mu_1, \dots, \mu_m) = (\mu_1, \dots, \mu_{l-1}, \mu_k, \mu_{l+1}, \dots, \mu_m), \quad (2.49)$$

and, for $1 \leq i \leq m$ and $1 \leq k, l \leq n_i$, denote by $\theta_{(i,k),(i,l)} : \mathbb{R}^{|\underline{n}| \times |\underline{n}|} \rightarrow \mathbb{R}^{|\underline{n}| \times |\underline{n}|}$ the one defined by

$$(\theta_{(i,k),(i,l)}(r))_{(s,t),(s',t')} = \begin{cases} r_{(s,t),(i,k)} & \text{if } (s',t') = (i,l) \\ r_{(i,l),(s',t')} & \text{if } (s,t) = (i,l) \\ r_{(s,t),(s',t')} & \text{otherwise.} \end{cases} \quad (2.50)$$

Lemma 2.10. *For all polynomials in $\Pi_{M,N}^1$,*

$$\Omega_{M,N}^{\uparrow,\text{grow}} \Phi_{M,N}^{\phi,m,\underline{n}}(\chi) := \left\langle \left\langle \nu, 2 \sum_{(i,j) \neq (i',j')} \frac{\partial \phi}{\partial r_{(i,j)(i',j')}} \right\rangle \right\rangle_\downarrow, \quad (2.51)$$

$$\Omega_{M,N}^{\uparrow,P} \Phi_{M,N}^{\phi,m,\underline{n}}(\chi) = \frac{\gamma_P}{2} \sum_{i=1}^m \sum_{1 \leq k,l \leq n_i} \left(\langle \langle \nu, \phi \circ \theta_{(i,k),(i,l)} \rangle \rangle_\downarrow - \langle \langle \nu, \phi \rangle \rangle_\downarrow \right), \quad (2.52)$$

and

$$\Omega_{M,N}^{\uparrow,H} \Phi_{M,N}^{\phi,m,\underline{n}}(\chi) = \frac{\gamma_H}{2} \sum_{1 \leq k,l \leq m} \left(\langle \langle \nu^{\otimes \downarrow m}, \psi_\downarrow \circ \theta^{k,l} \rangle \rangle - \langle \langle \nu^{\otimes \downarrow m}, \psi_\downarrow \rangle \rangle \right), \quad (2.53)$$

where for all $\underline{\mu} \in (\mathcal{M}_1(T))^m$,

$$\psi_\downarrow(\underline{\mu}) := \int_{T^{|\underline{n}|}} \bigotimes_{i=1}^m \mu_i^{\otimes \downarrow n_i} (d\underline{u}_i) \phi \circ R(\underline{u}). \quad (2.54)$$

2.3.2 Tightness

In this subsection we state that the family of two-level tree-valued Moran dynamics is tight. This will be a key ingredient in showing the existence of a solution to the diffusion limit martingale problem introduced in Section 2.4.

Since the particle approximation above is defined for $M, N \in \mathbb{N}$, we first need to discuss the ways of convergence we are interested in. We can consider the following:

- (C1) **Finite number of hosts with infinite parasites populations:** the number of hosts is of fixed size M and in each host the number of parasites N goes to infinity.
- (C2) **Infinite number of hosts with finite parasites populations:** the number of parasites within a host is of fixed size N and we let the number of hosts M go to infinity.
- (C3) **Infinite number of hosts with infinite parasites populations:** we let the number both the number of hosts M and parasites N go to infinity simultaneously, for example in a coupled way, that is, M is a function of N .

We made the choice of considering (C3) in the sequel, because the definitions and results of Section 2.4 can easily be adapted from (C3) to (C1) or (C2) by keeping M or N fixed.

Theorem 2.11 (Tightness). *Let, for each $M, N \in \mathbb{N}$, $(\mathcal{U}_t^{M,N})_{t \geq 0}$ be the two-level tree-valued Moran dynamics with population sizes (M, N) such that the family $\{\mathcal{U}_0^{M,N} : M, N \in \mathbb{N}\}$ is tight in $\mathbb{U}^{(2)}$. Then, the family $\{\mathcal{U}^{M,N} : M, N \in \mathbb{N}\}$ is tight in $\mathcal{D}_{\mathbb{U}^{(2)}}[0, \infty)$.*

An important part of the proof of the tightness stated in Theorem 2.11 consists in showing that the family of two-level tree-valued Moran dynamics satisfies a criterion for a compact containment condition to hold. This criterion applies to the more general class of finite population dynamics described in Definition 2.7, where the population size might not be constant and might even go extinct (in this case $\tau < \infty$).

The criterion uses the following notions. Let $\mathcal{U} = (\mathcal{U}_t)_{t \in [0, \tau]}$ be the tree-valued population dynamics read off from (\mathcal{I}, \preceq) , where (\mathcal{I}, \preceq) and τ are as in Definition 2.7. For $t \in [0, \tau]$ and $\epsilon > 0$, denote by

$$S_{2\epsilon}(\mathcal{U}_t) := \# \{A_{t-\epsilon}(\alpha, t) : \alpha \in \mathcal{I}_t\} \quad (2.55)$$

the number of ancestors of \mathcal{I}_t at time $t - \epsilon$, and by

$$\tilde{S}_{2\epsilon}(\mathcal{U}_t) := \inf_{\mathcal{J} \subseteq \mathcal{I}_t : M_{\nu_t}(\mathcal{J}) \leq 2\epsilon} \# \{A_{t-\epsilon}(\alpha, t) : \alpha \in \mathcal{I}_t \setminus \mathcal{J}\} \quad (2.56)$$

the minimal number of ancestors at time $t - \epsilon$ whose descendants cover a fraction of at least $1 - 2\epsilon$ of the population at time t . Note that $S_{2\epsilon}(\mathcal{U}_t)$ and $\tilde{S}_{2\epsilon}(\mathcal{U}_t)$ are the minimal numbers of 2ϵ -balls needed to cover respectively $\text{supp}(M_{\nu_t})$ and $\text{supp}(M_{\nu_t}) \setminus \mathcal{J}$ where the subset $\mathcal{J} \subseteq \mathcal{I}_t$ is such that $M_{\nu_t}(\mathcal{J}) \leq 2\epsilon$. For $t \geq \tau$ and $\epsilon > 0$, we set $S_{2\epsilon}(\mathcal{U}_t) = \tilde{S}_{2\epsilon}(\mathcal{U}_t) = 0$. Moreover for $\mathcal{J} \subseteq \mathcal{I}_s$ and $s \leq t$, denote by

$$D_t(\mathcal{J}, s) := \# \{\alpha \in \mathcal{I}_t : A_s(\alpha, t) \in \mathcal{J}\} \quad (2.57)$$

the number of descendants of the set \mathcal{J} at time t .

We can now enunciate the criterion for compact containment.

Proposition 2.12 (Compact containment for population dynamics). *For each $M, N \in \mathbb{N}$, let $\tau^{M,N}$ and $(\Omega^{M,N}, (\mathcal{A}_t^{M,N})_{t \in \mathbb{R}}, \mathbb{P}^{M,N})$, $(\mathcal{I}^{M,N}, \preceq^{M,N})$ be as in Definition 2.7.*

Let $\mathcal{U}^{M,N} := (\mathcal{U}_t^{M,N})_{t \in [0, \tau^{M,N}]}$ be the tree-valued population dynamics read off from $(\mathcal{I}^{M,N}, \preceq^{M,N})$ and assume that the family $\{\mathcal{U}_0^{M,N} : M, N \in \mathbb{N}\}$ is tight in $\mathbb{U}^{(2)}$. Furthermore, fix $T > 0$, and consider the following assumptions:

- (i) *For all $0 < \epsilon < T$ there exists a $\delta = \delta(\epsilon) > 0$ such that for all $s \in [0, T)$, $M, N \in \mathbb{N}$ and $\mathcal{A}_s^{M,N}$ -measurable random subsets $\mathcal{J}^{M,N} \subseteq \mathcal{I}_s^{M,N}$ with $M_{\nu_s^{M,N}}(\mathcal{J}^{M,N}) \leq \delta$,*

$$\sup_{M, N \in \mathbb{N}} \mathbb{P}^{M,N} \left\{ \sup_{t \in [s, T \wedge \tau^{M,N})} M_{\nu_t^{M,N}}(D_t(\mathcal{J}^{M,N}, s)) > \epsilon \right\} \leq \epsilon. \quad (2.58)$$

(ii.i) For all $0 < \epsilon \leq t < T$, the family $\left\{S_{2\epsilon}(\mathcal{U}_t^{M,N}) : M, N \in \mathbb{N}\right\}$ is tight.

(ii.ii) For all $0 < \epsilon \leq t < T$, the family $\left\{\tilde{S}_{2\epsilon}(\mathcal{U}_t^{M,N}) : M, N \in \mathbb{N}\right\}$ is tight.

Then the following compact containment conditions hold:

(a) Under (i) and (ii.i), for all $\epsilon > 0$ there exists a set $\Gamma_{\epsilon,T} \subseteq \mathbb{U}_c^{(2)}$ which is compact in $\mathbb{U}_c^{(2)}$ such that

$$\inf_{M,N \in \mathbb{N}} \mathbb{P}^{M,N} \left\{ \mathcal{U}_t^{M,N} \in \Gamma_{\epsilon,T} \text{ for all } t \in [\epsilon, T \wedge \tau^{M,N}) \right\} > 1 - \epsilon. \quad (2.59)$$

(b) Under (i) and (ii.ii), for all $\epsilon > 0$ there exists a set $\tilde{\Gamma}_{\epsilon,T} \subseteq \mathbb{U}^{(2)}$ which is compact in $\mathbb{U}^{(2)}$ such that

$$\inf_{M,N \in \mathbb{N}} \mathbb{P}^{M,N} \left\{ \mathcal{U}_t^{M,N} \in \tilde{\Gamma}_{\epsilon,T} \text{ for all } t \in [0, T \wedge \tau^{M,N}) \right\} > 1 - \epsilon. \quad (2.60)$$

2.4 The diffusion limit: the two-level tree-valued Fleming-Viot process

In this section, we introduce the two-level tree-valued Fleming-Viot process as the unique solution of a well-posed martingale problem. This process is approximated by the two-level tree-valued Moran dynamics in the sense of Proposition 2.16 which gives the convergence of the respective generators. Furthermore, the duality of the diffusion process to the nested Kingman coalescent stated in Section 2.4.3 is a key ingredient in showing the uniqueness of the solution to the martingale problem.

2.4.1 The martingale problem

In order to introduce operators for the two-level tree-valued Fleming-Viot process, we need a separating set of test functions on $\mathbb{M}^{(2)}$. As opposed to the definition of polynomials without replacement (2.41), we allow to sample several times the same host measure or the same parasite. A *polynomial* is a function $\Phi : \mathbb{M}^{(2)} \rightarrow \mathbb{R}$ that is of the following form: for all $\chi = (X, r, \nu) \in \mathbb{M}^{(2)}$,

$$\Phi(\chi) = \Phi^{m,\underline{n},\phi}(\chi) := \int_{(\mathcal{M}_1(X))^m} \nu^{\otimes m}(\underline{d}\mu) \int_{X^{m|\underline{n}|}} \bigotimes_{i=1}^m \mu_i^{\otimes n_i}(\underline{d}\underline{x}_i) \phi \circ R(\underline{x}), \quad (2.61)$$

where $m \in \mathbb{N}$, $\underline{n} \in \mathbb{N}^m$, $|\underline{n}| = n_1 + \dots + n_m$, $\phi \in \mathcal{C}_b(\mathbb{R}^{|\underline{n}| \times |\underline{n}|})$ and $R = R^{(X,r)}$ is the distance operator defined in (2.16). We will sometimes denote $\Phi^{m,\underline{n},\phi}(X, r, \nu)$ by $\langle \langle \nu, \phi \rangle \rangle$. It has been shown in [Mei19] that these test functions are well-defined, that is, $\Phi(\chi)$ does not depend on the representative of the equivalence class χ .

Define now

$$\Pi := \left\{ \Phi^{m,\underline{n},\phi} : m \in \mathbb{N}, \underline{n} \in \mathbb{N}^m, \phi \in \mathcal{C}_b(\mathbb{R}^{|\underline{n}| \times |\underline{n}|}) \right\} \quad (2.62)$$

and

$$\Pi^1 := \left\{ \Phi^{m,\underline{n},\phi} : m \in \mathbb{N}, \underline{n} \in \mathbb{N}^m, \phi \in \mathcal{C}_b^1(\mathbb{R}^{|\underline{n}| \times |\underline{n}|}) \right\}, \quad (2.63)$$

where $\mathcal{C}_b^1(\mathbb{R}^{|\underline{n}| \times |\underline{n}|})$ is the space of bounded and continuously differentiable real-valued functions on $\mathbb{R}^{|\underline{n}| \times |\underline{n}|}$ with bounded derivatives.

Remark 2.13 (Π^1 is an algebra that separates points). It has been shown in [Mei19, Theorem 3.8] that the set Π is separating for $\mathcal{M}_1(\mathbb{M}^{(2)})$. Since the set $\mathcal{C}_b^1(\mathbb{R}^{|\underline{n}| \times |\underline{n}|})$ is dense in $\mathcal{C}_b(\mathbb{R}^{|\underline{n}| \times |\underline{n}|})$ in the topology of point-wise convergence, Π^1 also separates points in $\mathbb{M}^{(2)}$. Furthermore, it is easy to see that Π and Π^1 are algebras.

We define the following operator

$$\Omega^\uparrow = \Omega^{\uparrow,H} + \Omega^{\uparrow,P} + \Omega^{\uparrow,\text{grow}} \quad (2.64)$$

on the set of test functions Π^1 . Let us thus consider $m \in \mathbb{N}$, $\underline{n} \in \mathbb{N}^m$ and $\phi \in \mathcal{C}_b^1(\mathbb{R}^{mn})$.

The operator for the resampling of hosts is

$$\Omega^{\uparrow, H} \Phi^{m, \underline{n}, \phi}(\chi) = \frac{\gamma_H}{2} \sum_{1 \leq k, l \leq m} \left(\langle \nu^{\otimes m}, \psi \circ \theta^{k, l} \rangle - \langle \nu^{\otimes m}, \psi \rangle \right), \quad (2.65)$$

where $\theta^{k, l}$ defined in (2.49) is the map which replaces the measure μ_l by the measure μ_k and for all $\underline{\mu} \in (\mathcal{M}_1(T))^m$,

$$\psi(\underline{\mu}) := \int_{T^{|\underline{n}|}} \mu_1^{\otimes n_1}(\mathrm{d}\underline{u}_1) \otimes \cdots \otimes \mu_m^{\otimes n_m}(\mathrm{d}\underline{u}_m) \phi \circ R(\underline{u}). \quad (2.66)$$

The operator for the resampling of parasites is

$$\Omega^{\uparrow, P} \Phi^{m, \underline{n}, \phi}(\chi) = \frac{\gamma_P}{2} \sum_{i=1}^m \sum_{1 \leq k, l \leq n_i} \left(\langle \langle \nu, \phi \circ \theta_{(i, k), (i, l)} \rangle \rangle - \langle \langle \nu, \phi \rangle \rangle \right), \quad (2.67)$$

where $\theta_{(i, k), (i, l)}$ defined in (2.50) is the map which replaces the distances to (i, l) by the distances to (i, k) .

The distance growth operator $\Omega^{\uparrow, \text{grow}}$ is defined as

$$\Omega^{\uparrow, \text{grow}} \Phi^{m, \underline{n}, \phi}(\chi) := \Phi^{m, \underline{n}, \Lambda \phi}(\chi) = \langle \langle \nu, \Lambda \phi \rangle \rangle, \quad (2.68)$$

where

$$\Lambda \phi := 2 \sum_{\substack{(i, j), (i', j') \in \mathbb{N}_{m, \underline{n}} \\ (i, j) \neq (i', j')}} \frac{\partial \phi}{\partial r_{(i, j)(i', j')}}. \quad (2.69)$$

The main result of this section reads as follows.

Theorem 2.14 (The well-posed martingale problem). *Let P_0 be a probability measure on $\mathbb{U}^{(2)}$. For each $M, N \in \mathbb{N}$, let $\mathcal{U}^{M, N} := (\mathcal{U}_t^{M, N})_{t \geq 0}$ be the two-level tree-valued Moran dynamics of size (M, N) . Assume that $(\mathcal{U}_0^{M, N})_{M, N}$ converges weakly with respect to the two-level Gromov-weak topology, as $(M, N) \rightarrow \infty$, to \mathcal{U}_0 distributed according to P_0 . Then $(\mathcal{U}^{M, N})_{M, N}$ converges weakly in Skorohod path space w.r.t. the two-level Gromov-weak topology to a $\mathbb{U}^{(2)}$ -valued Markov process \mathcal{U} with càdlàg paths.*

Furthermore, \mathcal{U} is the unique $\mathbb{U}^{(2)}$ -valued Markov process $(\mathcal{X}_t)_{t \geq 0}$ such that P_0 is the distribution of \mathcal{X}_0 , and for all $\Phi \in \Pi^1$, the process

$$\left(\Phi(\mathcal{X}_t) - \Phi(\mathcal{X}_0) - \int_0^t \Omega^{\uparrow} \Phi(\mathcal{X}_s) \mathrm{d}s \right)_{t \geq 0} \quad (2.70)$$

is a martingale.

We refer to the process from Theorem 2.14 as the *two-level tree-valued Fleming-Viot process* and finish this section with sample path properties of the process.

Proposition 2.15. *The two-level tree-valued Fleming-Viot process has the following properties:*

- (i) almost surely, \mathcal{U} has sample paths in $\mathcal{C}_{\mathbb{U}}[0, \infty)$.
- (ii) almost surely, for all $t > 0$, $\mathcal{U}_t \in \mathbb{U}_c^{(2)}$.

2.4.2 Convergence of generators

We show here that the generators $(\Omega_{M, N}^{\uparrow})_{M, N}$ for the two-level tree-valued Moran dynamics converge uniformly to the generator Ω^{\uparrow} for the two-level tree-valued Fleming-Viot dynamics. This convergence is an important part to prove the existence of a solution to the two-level tree-valued Fleming-Viot martingale problem. Recall that we consider the type of convergence (C3) where both the numbers of hosts and parasites go to infinity simultaneously.

Proposition 2.16 (Uniform convergence of the generators). *Let $\Phi \in \Pi^1$. For all $M, N \in \mathbb{N}$, there exists $\Phi_{M,N} \in \Pi_{M,N}^1$ such that*

$$\lim_{M,N \rightarrow \infty} \sup_{\chi \in \mathbb{U}_{M,N}^{(2)}} |\Phi_{M,N}(\chi) - \Phi(\chi)| = 0, \quad (2.71)$$

and

$$\lim_{M,N \rightarrow \infty} \sup_{\chi \in \mathbb{U}_{M,N}^{(2)}} |\Omega_{M,N}^\uparrow \Phi_{M,N}(\chi) - \Omega^\uparrow \Phi(\chi)| = 0. \quad (2.72)$$

Proof. The proof consists in showing that sampling with or without replacement is asymptotically equivalent in large populations. We follow the line of arguments of [Fre77]. Let $\mu = \frac{1}{M} \sum_{i=1}^M \delta_{y_i}$ a probability measure on some space $Y = \{y_1, \dots, y_M\}$ and $m \leq M$. Define

$$E := \{(z_1, \dots, z_m) \in Y^m : \forall i, j, z_i \neq z_j\}. \quad (2.73)$$

By definition, $d_{VT}(\mu^{\otimes \downarrow m}, \mu^{\otimes m}) := \sup_A |\mu^{\otimes \downarrow m}(A) - \mu^{\otimes m}(A)|$. For all $\underline{z} \in E$, $\mu^{\otimes \downarrow m}(\underline{z}) \geq \mu^{\otimes m}(\underline{z})$ and for all $\underline{z} \in \mathbb{C}E$, $\mu^{\otimes \downarrow m}(\underline{z}) \leq \mu^{\otimes m}(\underline{z})$. Therefore

$$d_{VT}(\mu^{\otimes \downarrow m}, \mu^{\otimes m}) = \mu^{\otimes \downarrow m}(E) - \mu^{\otimes m}(E) = 1 - \prod_{i=1}^{m-1} \left(1 - \frac{i}{M}\right). \quad (2.74)$$

Using that $1 - \sum_i x_i \leq \prod_i (1 - x_i)$ for all $x_1, \dots, x_n \in \mathbb{R}$ with $0 < x_i < 1$, we get the following bound on the difference between sampling with or without replacement

$$d_{VT}(\mu^{\otimes \downarrow m}, \mu^{\otimes m}) \leq \frac{m(m-1)}{2M}. \quad (2.75)$$

Then, for all $f: Y^m \rightarrow \mathbb{R}$ bounded,

$$\left| \int (\mu^{\otimes \downarrow m} - \mu^{\otimes m})(d\underline{z}) f(\underline{z}) \right| \leq 2d_{VT}(\mu^{\otimes \downarrow m}, \mu^{\otimes m}) \|f\| \leq \frac{m(m-1)}{M} \|f\|. \quad (2.76)$$

Fix $m \in \mathbb{N}$, $\underline{n} \in \mathbb{N}^m$ and $\phi \in \mathcal{C}_b^1(\mathbb{R}^{|\underline{n}| \times |\underline{n}|})$. Let $\Phi := \Phi^{m, \underline{n}, \phi}$. Now for $M \geq m$ and $N \geq \max_i n_i$, let $\Phi_{M,N} := \Phi_{M,N}^{m, \underline{n}, \phi}$. For $\chi = (X, r, \nu) \in \mathbb{U}_{M,N}^{(2)}$,

$$\begin{aligned} & |\Phi_{M,N}(\chi) - \Phi(\chi)| \\ &= \left| \int \nu^{\otimes \downarrow m}(d\underline{\mu}) \int \bigotimes_{i=1}^m \mu_i^{\otimes \downarrow n_i}(d\underline{u}_i) \phi \circ R(\underline{u}) - \int \nu^{\otimes m}(d\underline{\mu}) \int \bigotimes_{i=1}^m \mu_i^{\otimes n_i}(d\underline{u}_i) \phi \circ R(\underline{u}) \right| \\ &\leq \sum_{j=1}^m \int \nu^{\otimes \downarrow m}(d\underline{\mu}) \left| \int (\mu_j^{\otimes \downarrow n_j} - \mu_j^{\otimes n_j})(d\underline{u}_j) \int \bigotimes_{i=1}^{j-1} \mu_i^{\otimes \downarrow n_i}(d\underline{u}_i) \int \bigotimes_{i=j+1}^m \mu_i^{\otimes n_i}(d\underline{u}_i) \phi \circ R(\underline{u}) \right| \\ &\quad + \left| \int (\nu^{\otimes \downarrow m} - \nu^{\otimes m})(d\underline{\mu}) \int \bigotimes_{i=1}^m \mu_i^{\otimes n_i}(d\underline{u}_i) \phi \circ R(\underline{u}) \right|. \end{aligned} \quad (2.77)$$

Therefore, with (2.76),

$$|\Phi_{M,N}(\chi) - \Phi(\chi)| \leq \sum_{j=1}^m \frac{n_j(n_j-1)}{N} \|\phi\| + \frac{m(m-1)}{M} \|\phi\|, \quad (2.78)$$

and (2.71) follows.

For (2.72), observe first that $\Omega_{M,N}^{\uparrow, \text{grow}} \Phi_{M,N}^{m, \underline{n}, \phi} = \Phi_{M,N}^{m, \underline{n}, 2\Lambda\phi}$ and $\Omega^{\uparrow, \text{grow}} \Phi^{m, \underline{n}, \phi} = \Phi^{m, \underline{n}, 2\Lambda\phi}$ with $2\Lambda\phi$ bounded since the derivatives of $\phi \in \mathcal{C}_b^1(\mathbb{R}^{|\underline{n}| \times |\underline{n}|})$ are bounded. Hence, we can use (2.77) to bound

the difference of the two generators. This also applies to the parasite resampling operator with $2\Lambda\phi$ replaced by $\frac{\gamma_P}{2} \sum_{i=1}^m \sum_{1 \leq k, l \leq m} (\phi \circ \theta_{(i,k),(i,l)} - \phi)$. For the resampling of hosts,

$$\begin{aligned} \left| \Omega_{M,N}^{\uparrow,H} \Phi_{M,N}^{m,\underline{n},\phi}(\chi) - \Omega^{\uparrow,H} \Phi^{m,\underline{n},\phi}(\chi) \right| &\leq \sum_{1 \leq k, l \leq m} \left| \left\langle \nu^{\otimes \downarrow m}, \psi_{\downarrow} \circ \theta^{k,l} \right\rangle - \left\langle \nu^{\otimes m}, \psi \circ \theta^{k,l} \right\rangle \right| \\ &\quad + m^2 \left| \left\langle \nu^{\otimes \downarrow m}, \psi_{\downarrow} \right\rangle - \left\langle \nu^{\otimes m}, \psi \right\rangle \right|, \end{aligned} \quad (2.79)$$

where ψ_{\downarrow} and ψ are defined in (2.54) and (2.66). Then (2.72) follows by we applying (2.77) to each term $\left| \left\langle \nu^{\otimes \downarrow m}, \psi_{\downarrow} \circ \theta^{k,l} \right\rangle - \left\langle \nu^{\otimes m}, \psi \circ \theta^{k,l} \right\rangle \right|$. \square

2.4.3 Duality

It is well known that the Kingman coalescent is dual to the neutral measure-valued Fleming-Viot process (see [DK96]). This is also true of the tree-valued Fleming-Viot dynamics (see [GPW13]) and the two-level measure-valued Fleming-Viot process (see [Daw18]). We show here that this duality relation can be formulated in the two-level tree-valued case. That is, the nested Kingman coalescent is dual to the two-level tree-valued Fleming-Viot dynamics.

The nested Kingman coalescent we consider takes value in the space \mathcal{P}_{nest} of all pairs $(\mathbf{p}_H, \mathbf{p}_P)$ where \mathbf{p}_H and \mathbf{p}_P are two partitions of \mathbb{N}^2 such that for all $(i_1, j_1), (i_2, j_2) \in \mathbb{N}^2$,

- if $i_1 = i_2$, then (i_1, j_1) and (i_2, j_2) belong to the same block of \mathbf{p}_H ,
 - if (i_1, j_1) and (i_2, j_2) belong to the same block of \mathbf{p}_P , then they belong to the same block of \mathbf{p}_H .
- In other words, each block of \mathbf{p}_P is contained in a single block of \mathbf{p}_H .

Adding a component which measures genealogical distances, the tree-valued nested Kingman coalescent $\mathcal{K} = (\mathcal{K}_t)_{t \geq 0}$ with

$$\mathcal{K}(t) = (\kappa_H(t), \kappa_P(t), \underline{r}'(t)) \in \mathbb{G} := \mathcal{P}_{nest} \times \mathbb{R}^{\mathbb{N}^4} \quad (2.80)$$

is defined as follows:

- **Coalescence:** $(\kappa_H(t), \kappa_P(t))_{t \geq 0}$ is the nested Kingman coalescent with rates (γ_H, γ_P) taking values in \mathcal{P}_{nest} starting in $(\kappa_H(0), \kappa_P(0)) \in \mathcal{P}_{nest}$.
- **Distance growth:** At time t , for all (i, j) and (i', j') with $(i, j) \approx_{\kappa_P(t)} (i', j')$, the genealogical distance $r'_{((i,j),(i',j'))}(\cdot)$ grows with constant speed 2.

We now define a martingale problem associated with this dynamics. For this, we consider

$$\mathcal{G} := \left\{ G \in \mathcal{B}(\mathbb{G}) : \forall \underline{r}' \in \mathbb{R}^{\mathbb{N}^4}, G(\cdot, \underline{r}') \in \mathcal{C}(\mathcal{P}_{nest}) \text{ and depends on } \mathbf{p}_H \text{ and } \mathbf{p}_P \text{ only through their restrictions to } \mathbb{N}_{m,\underline{n}} \text{ for some } m \in \mathbb{N}, \underline{n} \in \mathbb{N}^m \right\} \quad (2.81)$$

and its subset

$$\mathcal{G}^1 := \{ G \in \mathcal{G} : \Lambda_{\mathbf{p}_P}^{r'} G \text{ exists for each partition } \mathbf{p}_P \text{ of } \mathbb{N}^2 \}, \quad (2.82)$$

with

$$\Lambda_{\mathbf{p}_P}^{r'} G := 2 \sum_{\substack{(i,j) \neq (i',j') \\ \mathbf{p}_P(i,j) \neq \mathbf{p}_P(i',j')}} \frac{\partial G}{\partial r'_{(i,j)(i',j')}}, \quad (2.83)$$

where $\mathbf{p}_P(i, j)$ denotes the block of the partition \mathbf{p}_P that contains (i, j) .

We consider the martingale problem associated with the operator Ω^{\downarrow} defined on \mathcal{G} and with domain \mathcal{G}^1 , where $\Omega^{\downarrow} := \Omega^{\downarrow, \text{grow}} + \Omega^{\downarrow, s} + \Omega^{\downarrow, g}$ with $\Omega^{\downarrow, \text{grow}}$ the growth operator

$$\Omega^{\downarrow, \text{grow}} G(\mathbf{p}_H, \mathbf{p}_P, \underline{r}') := \Lambda_{\mathbf{p}_P}^{r'} G(\mathbf{p}_H, \mathbf{p}_P, \underline{r}').$$

The two terms for coalescence events are

$$\begin{aligned}\Omega^{\downarrow, H} G(\mathbf{p}_H, \mathbf{p}_P, \underline{r}') &:= \gamma_H \sum_{\substack{\{\pi, \pi'\} \subset \mathbf{p}_H \\ \pi \neq \pi'}} (G(\pi_{\mathbf{p}_H}(\pi, \pi'), \mathbf{p}_P, \underline{r}') - G(\mathbf{p}_H, \mathbf{p}_P, \underline{r}')), \\ \Omega^{\downarrow, P} G(\mathbf{p}_H, \mathbf{p}_P, \underline{r}') &:= \gamma_P \sum_{\substack{\{\pi, \pi'\} \subset \mathbf{p}_P \\ \pi \sqcup_{\mathbf{p}_H} \pi', \pi \neq \pi'}} (G(\mathbf{p}_H, \beta_{\mathbf{p}_P}(\pi, \pi'), \underline{r}') - G(\mathbf{p}_H, \mathbf{p}_P, \underline{r}')), \end{aligned} \quad (2.84)$$

where

$$\beta_{\mathbf{p}}(\pi, \pi') := (\mathbf{p} \setminus \{\pi, \pi'\}) \cup \{\pi \cup \pi'\}, \quad (2.85)$$

and $\pi \sqcup_{\mathbf{p}_H} \pi'$ denotes that $\pi \cup \pi'$ is contained in a single block of \mathbf{p}_H . For P_0 a probability measure on \mathbb{G} , the tree-valued nested Kingman coalescent we defined solves the $(P_0, \Omega^{\downarrow}, \mathcal{G}^1)$ -martingale problem.

The duality relation. In order to give the duality relation between the two-level tree-valued Fleming-Viot process and the tree-valued nested Kingman coalescent, we need to define suitable duality functions.

We will need the following notation. For $m \in \mathbb{N}$, $\underline{n} \in \mathbb{N}^m$, \mathbf{p} a partition of $\mathbb{N}_{m, \underline{n}}$ and r a distance matrix indexed by the blocks of \mathbf{p} , we write

$$\left((r_{\varpi, \varpi'})_{\varpi, \varpi' \in \mathbf{p}} \right)^{\mathbf{p}} := (r_{\mathbf{p}(i, j), \mathbf{p}(i', j')})_{(i, j), (i', j') \in \mathbb{N}_{m, \underline{n}}}. \quad (2.86)$$

Example 2.17. Let $m = 3$, $n_1 = n_2 = 1$ and $n_3 = 2$. Consider \mathbf{p}_P a partition of \mathbb{N}^2 whose restriction to $\mathbb{N}_{3, (2, 1, 1)}$ is

$$\mathbf{p}_P|_{\mathbb{N}_{3, (2, 1, 1)}} \left\{ \{(1, 1), (3, 2)\}, \{(2, 1)\}, \{(3, 1)\} \right\}. \quad (2.87)$$

Then the matrix $\left((r_{\varpi, \varpi'})_{\varpi, \varpi' \in \mathbf{p}_P|_{\mathbb{N}_{3, (2, 1, 1)}}} \right)^{\mathbf{p}_P|_{\mathbb{N}_{3, (2, 1, 1)}}}$ is

$$\begin{pmatrix} r_{\{(1, 1), (3, 2)\}, \{(1, 1), (3, 2)\}} & r_{\{(1, 1), (3, 2)\}, \{(2, 1)\}} & r_{\{(1, 1), (3, 2)\}, \{(3, 1), (3, 2)\}} & r_{\{(1, 1), (3, 2)\}, \{(1, 1), (3, 2)\}} \\ r_{\{(2, 1)\}, \{(1, 1), (3, 2)\}} & r_{\{(2, 1)\}, \{(2, 1)\}} & r_{\{(2, 1)\}, \{(3, 1), (3, 2)\}} & r_{\{(2, 1)\}, \{(1, 1), (3, 2)\}} \\ r_{\{(3, 1)\}, \{(1, 1), (3, 2)\}} & r_{\{(3, 1)\}, \{(2, 1)\}} & r_{\{(3, 1)\}, \{(3, 1), (3, 2)\}} & r_{\{(3, 1)\}, \{(1, 1), (3, 2)\}} \\ r_{\{(1, 1), (3, 2)\}, \{(1, 1), (3, 2)\}} & r_{\{(1, 1), (3, 2)\}, \{(2, 1)\}} & r_{\{(1, 1), (3, 2)\}, \{(3, 1), (3, 2)\}} & r_{\{(1, 1), (3, 2)\}, \{(1, 1), (3, 2)\}} \end{pmatrix}. \quad (2.88)$$

For each $m \in \mathbb{N}$, $\underline{n} \in \mathbb{N}^m$, $\phi \in \mathcal{C}_b^1(\mathbb{R}^{|\underline{n}| \times |\underline{n}|})$, define the function $H^{m, \underline{n}, \phi}: \mathbb{U}^{(2)} \times \mathbb{G} \rightarrow \mathbb{R}$ as

$$\begin{aligned} & H^{m, \underline{n}, \phi}(\chi, (\mathbf{p}_H, \mathbf{p}_P, \underline{r}')) \\ &:= \int \bigotimes_{\pi \in \mathbf{p}_H|_{\mathbb{N}_{m, \underline{n}}}} \nu(d\mu_\pi) \int \bigotimes_{\pi \in \mathbf{p}_H|_{\mathbb{N}_{m, \underline{n}}}} \bigotimes_{\substack{\varpi \in \mathbf{p}_P|_{\mathbb{N}_{m, \underline{n}}} \\ \varpi \subset \pi}} \mu_\pi(du_\varpi) \phi \left(\left(R \left((u_{\varpi'})_{\varpi' \in \mathbf{p}_P|_{\mathbb{N}_{m, \underline{n}}}} \right) \right)^{\mathbf{p}_P|_{\mathbb{N}_{m, \underline{n}}}} + \underline{r}' \right). \end{aligned} \quad (2.89)$$

For each block π in the partition $\mathbf{p}_H|_{\mathbb{N}_{m, \underline{n}}}$, we sample a measure (a host) μ_π and for each block ϖ in the partition $\mathbf{p}_P|_{\mathbb{N}_{m, \underline{n}}}$ that is a subset of π , we sample a parasite u_ϖ according to the measure μ_π . We denote by \mathcal{H} the set of duality functions.

Note that for $m \in \mathbb{N}$, $\underline{n} \in \mathbb{N}^m$, $\phi \in \mathcal{C}_b^1(\mathbb{R}^{|\underline{n}| \times |\underline{n}|})$ and $(\mathbf{p}_H, \mathbf{p}_P, \underline{r}') \in \mathbb{G}$, we can express the function

$$H^{m, \underline{n}, \phi}(\cdot, (\mathbf{p}_H, \mathbf{p}_P, \underline{r}')) \quad (2.90)$$

as $\Phi^{\widehat{m}, \widehat{\underline{n}}, \widehat{\phi}} \in \Pi^1$ with $\widehat{m} = \#(\mathbf{p}_H|_{\mathbb{N}_{m, \underline{n}}})$, $\widehat{\underline{n}} = (\#(\mathbf{p}_P|_\pi))_{\pi \in \mathbf{p}_H|_{\mathbb{N}_{m, \underline{n}}}}$ and $\widehat{\phi} = (\phi((\cdot)^{\mathbf{p}_P} + \underline{r}'))$. Conversely, we can write each function in Π^1 as a function of the form (2.90), so that the subset

$$\left\{ H^{m, \underline{n}, \phi}(\cdot, (\mathbf{p}_H, \mathbf{p}_P, \underline{r}')) \in \mathcal{H} : m \in \mathbb{N}, \underline{n} \in \mathbb{N}^m, (\mathbf{p}_H, \mathbf{p}_P, \underline{r}') \in \mathbb{G}, \phi \in \mathcal{C}_b^1(\mathbb{R}^{|\underline{n}| \times |\underline{n}|}) \right\} \quad (2.91)$$

is equal to Π^1 and thus separates points in $\mathbb{U}^{(2)}$. Moreover, we need to make explicit how we will apply Ω^\uparrow to a function of the form (2.90). Indeed the operator Ω^\uparrow acts on polynomials $\Phi^{m, \underline{n}, \phi}$ where ϕ acts on distance matrix indexed by $\mathbb{N}_{m, \underline{n}}$ (see (2.64) and below). But since the order in the indices play no part in the definition of the generator, we simply replace the indices in $\mathbb{N}_{m, \underline{n}}$ by partition-block indices.

Proposition 2.18 (Duality relation). *For $P_0 \in \mathcal{M}_1(\mathbb{U}^{(2)})$ and $\zeta \in \mathbb{G}$, let $\mathcal{U} = (\mathcal{U}_t)_{t \geq 0}$ and $\mathcal{K} = (\mathcal{K}_t)_{t \geq 0}$ be solutions of the $(P_0, \Omega^\uparrow, \Pi^1)$ and $(\delta_\zeta, \Omega^\downarrow, \mathcal{G}^{1,0})$ -martingale problems, respectively. Then, if \mathcal{U} and \mathcal{K} are independent,*

$$\mathbb{E}[H(\mathcal{U}_t, \zeta)] = \mathbb{E}[H(\mathcal{U}_0, \mathcal{K}_t)], \quad (2.92)$$

for all $t \geq 0$ and $H \in \mathcal{H}$.

Proof. We apply [EK86, Theorem 4.4.11, Corollary 4.4.13] (with $\alpha = \beta = 0$) to the functions in \mathcal{H} which are bounded. For this, we show that for all $m \in \mathbb{N}$, $\underline{n} \in \mathbb{N}^m$ and $\phi \in \mathcal{C}_b^1(\mathbb{R}^{|\underline{n}| \times |\underline{n}|})$,

$$\Omega^\uparrow H^{m, \underline{n}, \phi}(\cdot, (\mathfrak{p}_H, \mathfrak{p}_P, \underline{r}'))(\chi) = \Omega^\downarrow H^{m, \underline{n}, \phi}(\chi, \cdot)(\mathfrak{p}_H, \mathfrak{p}_P, \underline{r}'). \quad (2.93)$$

We verify this for each of the three components separately. To make the calculation more readable and since it leads to no confusion, we omit the restriction notation $\mathfrak{p}_{|\mathbb{N}_{m, \underline{n}}}$ and simply use \mathfrak{p} in the following equations. First, we have that

$$\begin{aligned} & \Omega^{\uparrow, \text{grow}} H^{m, \underline{n}, \phi}(\cdot, (\mathfrak{p}_H, \mathfrak{p}_P, \underline{r}'))(\chi) \\ &= \int \bigotimes_{\pi \in \mathfrak{p}_H} \nu(d\mu_\pi) \int \bigotimes_{\pi \in \mathfrak{p}_H} \bigotimes_{\substack{\varpi \in \mathfrak{p}_P \\ \varpi \subset \pi}} \mu_\pi(du_\varpi) \sum_{\varpi' \neq \varpi''} \frac{\partial(\phi((\cdot)^{\mathfrak{p}_P} + \underline{r}'))}{\partial r_{\varpi', \varpi''}} \left(R((u_\varpi)_{\varpi \in \mathfrak{p}_P}) \right) \\ &= \int \bigotimes_{\pi \in \mathfrak{p}_H} \nu(d\mu_\pi) \int \bigotimes_{\pi \in \mathfrak{p}_H} \bigotimes_{\substack{\varpi \in \mathfrak{p}_P \\ \varpi \subset \pi}} \mu_\pi(du_\varpi) \sum_{\varpi' \neq \varpi''} \sum_{\substack{(i, j) \neq (i', j') \\ \mathfrak{p}_P(i, j) = \varpi' \\ \mathfrak{p}_P(i', j') = \varpi''}} \frac{\partial(\phi((R((u_\varpi)_{\varpi \in \mathfrak{p}_P}))^{\mathfrak{p}_P} + \cdot))}{\partial r'_{(i, j)(i', j')}}(\underline{r}') \\ &= \int \bigotimes_{\pi \in \mathfrak{p}_H} \nu(d\mu_\pi) \int \bigotimes_{\pi \in \mathfrak{p}_H} \bigotimes_{\substack{\varpi \in \mathfrak{p}_P \\ \varpi \subset \pi}} \mu_\pi(du_\varpi) \sum_{\substack{(i, j) \neq (i', j') \\ \mathfrak{p}_P(i, j) \neq \mathfrak{p}_P(i', j')}} \frac{\partial(\phi((R((u_\varpi)_{\varpi \in \mathfrak{p}_P}))^{\mathfrak{p}_P} + \cdot))}{\partial r'_{(i, j)(i', j')}}(\underline{r}') \\ &= \sum_{\substack{(i, j) \neq (i', j') \\ \mathfrak{p}_P(i, j) \neq \mathfrak{p}_P(i', j')}} \frac{\partial H^{m, \underline{n}, \phi}(\chi, \cdot)}{\partial r'_{(i, j)(i', j')}}(\mathfrak{p}_H, \mathfrak{p}_P, \underline{r}') \\ &= \Omega^{\downarrow, \text{grow}} H^{m, \underline{n}, \phi}(\chi, \cdot)(\mathfrak{p}_H, \mathfrak{p}_P, \underline{r}'). \end{aligned} \quad (2.94)$$

For the generators of resampling of parasites and coalescence of parasites, we have

$$\begin{aligned} & \Omega^{\uparrow, P} H^{m, \underline{n}, \phi}(\cdot, (\mathfrak{p}_H, \mathfrak{p}_P, \underline{r}'))(\chi) \\ &= \frac{\gamma_P}{2} \sum_{\pi' \in \mathfrak{p}_H} \sum_{\substack{\varpi', \varpi'' \in \mathfrak{p}_P \\ \varpi', \varpi'' \subset \pi'}} \int \bigotimes_{\pi \in \mathfrak{p}_H} \nu(d\mu_\pi) \int \bigotimes_{\pi \in \mathfrak{p}_H} \bigotimes_{\substack{\varpi \in \mathfrak{p}_P \\ \varpi \subset \pi}} \mu_\pi(du_\varpi) \\ & \quad \left(\phi((\theta_{\varpi' \varpi''} \circ R(\underline{u}))^{\mathfrak{p}_P} + \underline{r}') - \phi((R(\underline{u}))^{\mathfrak{p}_P} + \underline{r}') \right) \\ &= \frac{\gamma_P}{2} \sum_{\pi' \in \mathfrak{p}_H} \sum_{\substack{\varpi', \varpi'' \in \mathfrak{p}_P \\ \varpi', \varpi'' \subset \pi'}} \left(\int \bigotimes_{\pi \in \mathfrak{p}_H} \nu(d\mu_\pi) \int \bigotimes_{\pi \in \mathfrak{p}_H} \bigotimes_{\substack{\varpi \in \mathfrak{p}_P \\ \varpi \subset \pi}} \mu_\pi(du_\varpi) \phi((R(\underline{u}))^{\beta_{\mathfrak{p}_P}(\varpi', \varpi'')} + \underline{r}') \right. \\ & \quad \left. - \int \bigotimes_{\pi \in \mathfrak{p}_H} \nu(d\mu_\pi) \int \bigotimes_{\pi \in \mathfrak{p}_H} \bigotimes_{\substack{\varpi \in \mathfrak{p}_P \\ \varpi \subset \pi}} \mu_\pi(du_\varpi) \phi((R(\underline{u}))^{\mathfrak{p}_P} + \underline{r}') \right), \end{aligned} \quad (2.95)$$

which is $\Omega^{\downarrow, P} H^{m, \underline{n}, \phi}(\chi, \cdot)(\mathfrak{p}_H, \mathfrak{p}_P, \underline{r}')$.

Finally for the generators of resampling of hosts and coalescence of hosts, we have

$$\begin{aligned}
& \Omega^{\uparrow, H} H^{m, \underline{n}, \phi}(\cdot, (\mathfrak{p}_H, \mathfrak{p}_P, \underline{r}'))(\chi) \\
&= \frac{\gamma_H}{2} \sum_{\pi', \pi'' \in \mathfrak{p}_H} \left(\int \bigotimes_{\pi \in \mathfrak{p}_H} \nu(d\mu_\pi) \int \bigotimes_{\substack{\pi \in \mathfrak{p}_H \\ \pi \neq \pi''}} \bigotimes_{\substack{\varpi \in \mathfrak{p}_P \\ \varpi \subset \pi}} \mu_\pi(du_\varpi) \otimes \bigotimes_{\substack{\varpi \in \mathfrak{p}_P \\ \varpi \subset \pi''}} \mu_{\pi'}(du_\varpi) \phi((R(\underline{u}))^{\mathfrak{p}_P} + \underline{r}') \right. \\
&\quad \left. - \int \bigotimes_{\pi \in \mathfrak{p}_H} \nu(d\mu_\pi) \int \bigotimes_{\pi \in \mathfrak{p}_H} \bigotimes_{\substack{\varpi \in \mathfrak{p}_P \\ \varpi \subset \pi}} \mu_\pi(du_\varpi) \phi((R(\underline{u}))^{\mathfrak{p}_P} + \underline{r}') \right) \\
&= \frac{\gamma_H}{2} \sum_{\pi', \pi'' \in \mathfrak{p}_H} \left(\int \bigotimes_{\pi \in \beta_{\mathfrak{p}_H}(\pi', \pi'')} \nu(d\mu_\pi) \int \bigotimes_{\pi \in \beta_{\mathfrak{p}_H}(\pi', \pi'')} \bigotimes_{\substack{\varpi \in \mathfrak{p}_P \\ \varpi \subset \pi}} \mu_\pi(du_\varpi) \phi((R(\underline{u}))^{\mathfrak{p}_P} + \underline{r}') \right. \\
&\quad \left. - \int \bigotimes_{\pi \in \mathfrak{p}_H} \nu(d\mu_\pi) \int \bigotimes_{\pi \in \mathfrak{p}_H} \bigotimes_{\substack{\varpi \in \mathfrak{p}_P \\ \varpi \subset \pi}} \mu_\pi(du_\varpi) \phi((R(\underline{u}))^{\mathfrak{p}_P} + \underline{r}') \right)
\end{aligned} \tag{2.96}$$

which is $\Omega^{\downarrow, H} H^{m, \underline{n}, \phi}(\chi, \cdot)(\mathfrak{p}_H, \mathfrak{p}_P, \underline{r}')$, and the result follows by combining the three terms. \square

2.5 Expected length distribution

In population genetics, a simple popular model is the infinitely allele model ([Kim69, Wat75] or [Ber09] for a review). It describes the evolution of a sample of genes at a single locus. Mutations are assumed to occur according to a Poisson point process with intensity $\theta > 0$ along the tree, and each time, a new allele appears. Thus, the number of mutations follows a Poisson distribution with parameter θL where L is the total length of the tree. Several works have studied the length of a genealogical tree, with applications to the estimation of θ or to testing hypotheses, for various one-level tree models (Kingman process [Fel71, Section I.6], Bolthausen-Sznitman coalescent processes [DIMR07], Lambda-coalescent processes [Mö6], Beta-coalescent processes [BBS08, DDSJ08, DFSJY13]...).

We seek here to derive some analog results on the subtree length distribution in the two-level tree-valued Fleming-Viot dynamics. We first consider its action on specific polynomials evaluating the length distribution of the subtrees spanned by finite samples. Then, we study the unique stationary distribution of the dynamics, which is the nested Kingman measure tree, and we compare its subtree length distribution to the one of the coalescent without hierarchy.

2.5.1 Length of a subtree of the evolving two-level Fleming-Viot process

In this section, we adapt a result from [GPW13] concerning the subtree length distribution. For a metric space (X, r) satisfying the four point condition (2.26), the length of the subtree of (X, r) spanned by $x_1, \dots, x_n \in X$ is given by

$$\frac{1}{2} \inf \left\{ \sum_{i=1}^n r_{x_i, x_{\sigma(i)}} : \sigma \in \Sigma_n^1 \right\}, \tag{2.97}$$

where Σ_n^1 is the set of all permutations of $\{1, \dots, n\}$ with one cycle. Therefore, to describe the length distribution of subtrees spanned by finite two-level samples, we define for $m \in \mathbb{N}$, $\underline{n} \in \mathbb{N}^m$ and $\underline{r} \in \mathbb{R}_+^{\mathbb{N}_{m, \underline{n}} \times \mathbb{N}_{m, \underline{n}}}$,

$$l_{m, \underline{n}}(\underline{r}) := \frac{1}{2} \inf \left\{ \sum_{i=1}^m \sum_{j=1}^{n_i} r_{(i, j), \sigma(i, j)} : \sigma \in \mathcal{S}_{m, \underline{n}} \right\}, \tag{2.98}$$

where $\mathcal{S}_{m, \underline{n}}$ denotes the set of all permutations of $\mathbb{N}_{m, \underline{n}} := \{(i, j) : 1 \leq i \leq m, 1 \leq j \leq n_i\}$ with one cycle. We also define for $\chi \in \mathbb{U}^{(2)}$, $\sigma > 0$ and $m \in \mathbb{N}$, $\underline{n} \in \mathbb{N}^m$,

$$\Psi^{m, \underline{n}, l}(\chi, \sigma) := \int_{(\mathcal{M}_1(X))^m} \nu^{\otimes m}(d\mu) \int_{X^{\underline{n}}} \bigotimes_{i=1}^m \mu_i^{\otimes n_i}(d\underline{x}_i) e^{-\sigma l_{m, \underline{n}} \circ R(\underline{x})}. \tag{2.99}$$

The following result describes the evolution of the length distribution of a sampled subtree under the two-level tree-valued Fleming-Viot dynamics.

Proposition 2.19. *For $m \geq 1$ and $\underline{n} \in \mathbb{N}^m$, we have*

$$\begin{aligned} \Omega^\uparrow \Psi^{m, \underline{n}, l}(\chi, \sigma) &= -\sigma |\underline{n}| \Psi^{m, \underline{n}, l}(\chi, \sigma) + \frac{\gamma_P}{2} \sum_{i=1}^m n_i (n_i - 1) (\Psi^{m, \underline{n} - e_i, l}(\chi, \sigma) - \Psi^{m, \underline{n}, l}(\chi, \sigma)) \\ &\quad + \frac{\gamma_H}{2} \sum_{\substack{1 \leq k, l \leq m \\ k \neq l}} (\Psi^{m-1, (n_1, \dots, n_{k-1}, n_k + n_l, n_{k+1}, \dots, n_{l-1}, n_{l+1}, \dots, n_m), l}(\chi, \sigma) - \Psi^{m, \underline{n}, l}(\chi, \sigma)). \end{aligned} \quad (2.100)$$

Proof. The result follows from the definition of Ω^\uparrow (2.64) as the sum of three operators and an easy computation of the action of each operator on $\Psi^{m, \underline{n}, l}(\chi)$. \square

2.5.2 Subtree length distribution of the nested Kingman measure tree

First, we state that in long time, the two-level Fleming-Viot process converges in distribution w.r.t. the Gromov-weak topology to the nested Kingman measure tree and provides statistical representations of its subtree length distribution through a recursive relation.

The nested Kingman measure tree $\mathcal{U}_\infty^{(2)}$ is a metric two-level measure space introduced in [Mei19, Section 10] as the weak limit of a sequence of finite nested Kingman coalescent processes with rates (γ_H, γ_P) when the number of leaves tends to infinity.

Proposition 2.20 (Convergence to the nested Kingman measure tree). *When $t \rightarrow +\infty$, the two-level tree-valued Fleming-Viot process $(\mathcal{U}_t^{(2)})_{t \geq 0}$ converges in distribution w.r.t. the two-level Gromov-weak topology to the two-level Kingman process $\mathcal{U}_\infty^{(2)}$. In particular, the distribution of $\mathcal{U}_\infty^{(2)}$ is the unique equilibrium distribution of the two-level tree-valued Fleming-Viot dynamics.*

Proof. By [Mei19, Theorem 9.2], the set Π^1 is convergence determining so it is enough to prove that for all $\Phi \in \Pi_1$,

$$\lim_{t \rightarrow +\infty} \mathbb{E}[\Phi(\mathcal{U}_t^{(2)})] = \mathbb{E}[\Phi(\mathcal{U}_\infty^{(2)})]. \quad (2.101)$$

To show that, we use the duality relation from Proposition 2.18. Fix $m \in \mathbb{N}$, $\underline{n} \in \mathbb{N}^m$, $|\underline{n}| = n_1 + \dots + n_m$ and $\phi \in \mathcal{C}_b^1(\mathbb{R}^{n \times |n|})$. We consider the nested Kingman coalescent $(\mathcal{K}(t))_{t \geq 0}$ which starts in $\mathcal{K}(0) := (\kappa_H(0), \kappa_P(0), \underline{r}'(0)) \in \mathbb{G}$, where $\kappa_H(0) := \{((i, j), (i, j)) | i, j \in \mathbb{N}\}$, $\kappa_P(0) := \{((i, j), (i', j')) | i, j, i', j' \in \mathbb{N}\}$ and $\underline{r}'(0) \equiv 0$. By construction of \mathcal{K} , $\lim_{t \rightarrow +\infty} (\kappa_H(t), \kappa_P(t)) = (\{\mathbb{N}^2\}, \{\mathbb{N}^2\})$ and

$$\lim_{t \rightarrow +\infty} \mathbb{E}[\phi(\underline{r}'(t))] = \mathbb{E}[\Phi(\mathcal{U}_\infty^{(2)})]. \quad (2.102)$$

Therefore by the duality relation (2.92),

$$\begin{aligned} \lim_{t \rightarrow +\infty} \mathbb{E}[\Phi(\mathcal{U}_t^{(2)})] &= \lim_{t \rightarrow +\infty} \mathbb{E}\left[H^{m, \underline{n}, \phi}\left(\mathcal{U}_t^{(2)}, (\kappa_H(t), \kappa_P(t), \underline{r}'(t))\right)\right] \\ &= \lim_{t \rightarrow +\infty} \mathbb{E}\left[H^{m, \underline{n}, \phi}\left(\mathcal{U}_t^{(2)}, (\{\mathbb{N}^2\}, \{\mathbb{N}^2\}, \underline{r}'(t))\right)\right] \\ &= \lim_{t \rightarrow +\infty} \mathbb{E}[\phi(\underline{r}'(t))] \\ &= \mathbb{E}[\Phi(\mathcal{U}_\infty^{(2)})]. \end{aligned} \quad (2.103)$$

This concludes the proof. \square

For $m \geq 1$ and $\underline{n} \in \mathbb{N}^m$, let us denote by $L_{m, \underline{n}}$ the length of the finite nested Kingman coalescent started with m hosts and n_i parasites in host i . From the definition of the nested Kingman coalescent (see [Mei19, Definition 10.2]), we have in particular that for $\sigma > 0$,

$$\mathbb{E}[e^{-\sigma L_{m, \underline{n}}}] = \Psi^{m, \underline{n}, l}(\mathcal{U}_\infty, \sigma). \quad (2.104)$$

Proposition 2.21 (Subtree length distribution in the nested Kingman tree). *For all $m \in \mathbb{N}$ and $\underline{n} \in \mathbb{N}^m$, we have*

$$\mathbb{E} [e^{-\sigma L_{m,\underline{n}}}] = \frac{1}{\sigma \sum_{i=1}^m n_i + \sum_{i=1}^m \frac{n_i(n_i-1)}{2} \gamma_P + \frac{m(m-1)}{2} \gamma_H} \left(\sum_{i=1}^m \frac{n_i(n_i-1)}{2} \gamma_P \mathbb{E} [e^{-\sigma L_{m,\underline{n}-e_i}}] \right. \\ \left. + \sum_{1 \leq k < l \leq m} \gamma_H \mathbb{E} [e^{-\sigma L_{m-1,(n_1, \dots, n_{k-1}, n_k+n_l, n_{k+1}, \dots, n_{l-1}, n_{l+1}, \dots, n_m)}]}] \right). \quad (2.105)$$

Proof. We condition on the first event of resampling that occurs in the nested coalescent. This event occurs after a random time \mathcal{E}_1 , which is exponentially distributed with parameter $\sum_{i=1}^m \frac{n_i(n_i-1)}{2} \gamma_P + \frac{m(m-1)}{2} \gamma_H$. After this event, the process has law the nested Kingman coalescent started with one host less, or one parasite less, so that we get the following induction relations:

$$\mathbb{E} [e^{-\sigma L_{m,\underline{n}}}] = \mathbb{E} [e^{-\sigma(\sum_{i=1}^m n_i) \mathcal{E}_1}] \frac{1}{\sum_{i=1}^m \frac{n_i(n_i-1)}{2} \gamma_P + \frac{m(m-1)}{2} \gamma_H} \left(\sum_{i=1}^m \frac{n_i(n_i-1)}{2} \gamma_P \mathbb{E} [e^{-\sigma L_{m,\underline{n}-e_i}}] \right. \\ \left. + \sum_{1 \leq k < l \leq m} \gamma_H \mathbb{E} [e^{-\sigma L_{m-1,(n_1, \dots, n_{k-1}, n_k+n_l, n_{k+1}, \dots, n_{l-1}, n_{l+1}, \dots, n_m)}]}] \right) \\ = \frac{1}{\sigma \sum_{i=1}^m n_i + \sum_{i=1}^m \frac{n_i(n_i-1)}{2} \gamma_P + \frac{m(m-1)}{2} \gamma_H} \left(\sum_{i=1}^m \frac{n_i(n_i-1)}{2} \gamma_P \mathbb{E} [e^{-\sigma L_{m,\underline{n}-e_i}}] \right. \\ \left. + \sum_{1 \leq k < l \leq m} \gamma_H \mathbb{E} [e^{-\sigma L_{m-1,(n_1, \dots, n_{k-1}, n_k+n_l, n_{k+1}, \dots, n_{l-1}, n_{l+1}, \dots, n_m)}]}] \right). \quad (2.106)$$

This is the announced equation. \square

Notice that this result is consistent with the action of the operator Ω^\uparrow on $\Psi^{m,\underline{n},l}$ from (2.100). Indeed, since \mathcal{U}_∞ is the stationary distribution of the two-level Fleming-Viot dynamics, $\Omega^\uparrow \Psi^{m,\underline{n},l}(\mathcal{U}_\infty, \sigma) = 0$ for all $\sigma > 0$.

Let us specify the subtree length distribution in the case where $m = 1$ or $m = 2$.

Proposition 2.22. *For $m = 1$, the nested Kingman measure tree satisfies:*

$$\mathbb{E} [e^{-\sigma L_{1,n}}] = \prod_{i=1}^{n-1} \frac{\frac{\gamma_P}{2} i}{\sigma + \frac{\gamma_P}{2} i} = \mathbb{E} [e^{-\sigma \sum_{i=2}^n \mathcal{E}_P^i}], \quad (2.107)$$

where $(\mathcal{E}_P^i)_{2 \leq i \leq n}$ are independent and \mathcal{E}_P^i is exponentially distributed with parameter $\frac{\gamma_P(i-1)}{2}$. Moreover, for $m = 2$,

$$\mathbb{E} [e^{-\sigma L_{2,(n,1)}}] = \frac{\frac{\gamma_H}{2}}{\sigma + \frac{\gamma_H}{2}} \prod_{i=1}^n \frac{\frac{\gamma_P}{2} i}{\sigma + \frac{\gamma_P}{2} i} = \mathbb{E} [e^{-\sigma(\mathcal{E}_H^2 + \sum_{i=2}^{n+1} \mathcal{E}_P^i)}], \quad (2.108)$$

where \mathcal{E}_H^2 is exponentially distributed with parameter $\frac{\gamma_H}{2}$ and independent from $(\mathcal{E}^i)_{2 \leq i \leq n}$.

Proof. Let $n \in \mathbb{N}$. For $m = 1$, (2.105) rewrites:

$$\mathbb{E} [e^{-\sigma L_{1,n}}] = \frac{1}{\sigma n + \frac{n(n-1)}{2} \gamma_P} \frac{n(n-1)}{2} \gamma_P \mathbb{E} [e^{-\sigma L_{1,n-1}}]. \quad (2.109)$$

and a direct recursion gives the announced result for $m = 1$. For $m = 2$ and $\underline{n} = (n, 1)$, (2.105) becomes:

$$\mathbb{E} [e^{-\sigma L_{2,(n,1)}}] = \frac{1}{\sigma(n+1) + \frac{n(n-1)}{2} \gamma_P + \gamma_H} \left(\frac{n(n-1)}{2} \gamma_P \mathbb{E} [e^{-\sigma L_{2,(n-1,1)}}] + \gamma_H \mathbb{E} [e^{-\sigma L_{1,n+1}}] \right). \quad (2.110)$$

A straightforward computation shows that 2.108 solves this equation. \square

2.5.3 Results on the expected tree length: comparison with the one-level case

Here, we no longer consider the mean sample Laplace transform of the subtree length distribution, but only its expected value. It allows in particular for estimation of the coalescence rates γ_H and γ_P genealogical trees, which we assume a priori to be distributed according to a nested coalescent tree.

As a direct consequence of Proposition 2.21, for all $m \in \mathbb{N}$ and $\underline{n} \in \mathbb{N}^m$,

$$\begin{aligned} \mathbb{E}[L_{m,\underline{n}}] = & \frac{1}{\sum_{i=1}^m \frac{n_i(n_i-1)}{2} \gamma_P + \frac{m(m-1)}{2} \gamma_H} \left(\sum_{i=1}^m n_i + \sum_{i=1}^m \frac{n_i(n_i-1)}{2} \gamma_P \mathbb{E}[L_{m,\underline{n}-e_i}] \right. \\ & \left. + \sum_{1 \leq k < l \leq m} \gamma_H \mathbb{E}[L_{m,(n_1, \dots, n_{k-1}, n_k+n_l, n_{k+1}, \dots, n_{l-1}, n_{l+1}, \dots, n_m)}] \right). \end{aligned} \quad (2.111)$$

Furthermore, from Proposition 2.22, we have for all $n \in \mathbb{N}$,

$$\mathbb{E}[L_{1,n}] = \sum_{k=1}^{n-1} \frac{2}{k\gamma_P}, \quad (2.112)$$

$$\mathbb{E}[L_{2,(n,1)}] = \frac{2}{\gamma_H} + \sum_{k=1}^n \frac{2}{k\gamma_P}. \quad (2.113)$$

In order to compare these results with the subtree length distribution of the simple Kingman coalescent, we denote by L_m the length of the Kingman m -coalescent with rate γ , and recall that $L_{m,\underline{1}}$ is the length of the nested Kingman coalescent started with m parasites in distinct hosts. The following give a comparison of the mean values of L_m and $L_{m,\underline{1}}$ for $m \in \{2, 3, 4, 5\}$:

$$\mathbb{E}[L_2] = \frac{2}{\gamma}, \quad \mathbb{E}[L_{2,\underline{1}}] = \frac{2}{\gamma_H} + \frac{2}{\gamma_P}, \quad (2.114)$$

$$\mathbb{E}[L_3] = \frac{3}{\gamma}, \quad \mathbb{E}[L_{3,\underline{1}}] = \frac{3}{\gamma_H} + \frac{3}{\gamma_P}, \quad (2.115)$$

$$\mathbb{E}[L_4] = \frac{11}{3\gamma}, \quad \mathbb{E}[L_{4,\underline{1}}] = \frac{11}{3\gamma_H} + \frac{11}{3\gamma_P} + \frac{1}{3} \frac{1}{\gamma_P + 3\gamma_H} \frac{2\gamma_P + 3\gamma_H}{2\gamma_P + \gamma_H}, \quad (2.116)$$

$$\mathbb{E}[L_5] = \frac{25}{6\gamma}, \quad \mathbb{E}[L_{5,\underline{1}}] = \frac{25}{6\gamma_H} + \frac{25}{6\gamma_P} + \frac{5}{6} \frac{1}{\gamma_P + 3\gamma_H}. \quad (2.117)$$

Notice that when $n = 2$ or $n = 3$, the sampled subtree lengths have the same mean value for $1/\gamma = 1/\gamma_H + 1/\gamma_P$. Therefore, in order to compare the lengths between both models, γ is equal to the harmonic mean of γ_H and γ_P . The relations for $n = 4$ and $n = 5$ brings out that the length of the nested model deviates from the one-level neutral model. Indeed, the subtrees are longer in the two-level setup, which is not surprising, as the coalescence of parasites (in the lower level) is constrained by the beforehand coalescence of their respective hosts (in the upper level).

Let us solve the recursion equation (2.111) for $m = 2$ and $\underline{n} = (n_1, n_2)$. Recall that $|\underline{n}| = n_1 + n_2$. For $\ell \leq |\underline{n}|$, define by

$$\Gamma_{\ell,\underline{n}} = \{\underline{\pi} \in (\mathbb{N}^2)^{\ell+1} \mid \pi_0 = \underline{n}, \text{ and } \forall 0 \leq t \leq \ell, \pi_{t+1} - \pi_t \in \{(-1, 0), (0, -1)\}\}$$

the set of west-south paths of length ℓ on \mathbb{N}^2 started at $\pi_0 = (n_1, n_2)$. We can associate to each of these paths a weight depending on π_ℓ and the directions taken by the paths between time 0 and ℓ :

$$w(\pi) = \prod_{t=0}^{\ell-1} \left(\beta^{\leftarrow}(\pi_t) \mathbf{1}_{\{\pi_{t+1}-\pi_t=(-1,0)\}} + \beta^{\downarrow}(\pi_t) \mathbf{1}_{\{\pi_{t+1}-\pi_t=(0,-1)\}} \right) \times \alpha(\pi_\ell), \quad (2.118)$$

where for $(x, y) \in \mathbb{N}^2$, the weight of edges β and of the final vertex α are given by:

$$\begin{aligned}\alpha(x, y) &= \frac{1}{\left(\frac{x(x-1)}{2} + \frac{y(y-1)}{2}\right)\gamma_P + \gamma_H}, \\ \beta^{\leftarrow}(x, y) &= \frac{x(x-1)}{2} \times \gamma_P \times \alpha(x, y), \\ \beta^{\downarrow}(x, y) &= \frac{y(y-1)}{2} \times \gamma_P \times \alpha(x, y).\end{aligned}\tag{2.119}$$

By convention, if π is of length 0, we define $w(\pi) = \alpha(\pi_0)$.

Proposition 2.23. *With the expressions of (2.112), we have:*

$$\mathbb{E}[L_{2,(n_1,n_2)}] = \sum_{\ell=0}^{n_1+n_2-2} \left(n_1 + n_2 - \ell + \gamma_H \mathbb{E}[L_{1,n_1+n_2-\ell}] \right) \times W_\ell(n_1, n_2),\tag{2.120}$$

where

$$W_\ell(n_1, n_2) = \sum_{\pi \in \Gamma_{\ell, \mathbf{z}}} w(\pi).\tag{2.121}$$

Proof. From (2.111), we have:

$$\begin{aligned}\mathbb{E}[L_{2,(n_1,n_2)}] &= \frac{1}{\left(\frac{n_1(n_1-1)}{2} + \frac{n_2(n_2-1)}{2}\right)\gamma_P + \gamma_H} \left(n_1 + n_2 + \gamma_H \mathbb{E}[L_{1,n_1+n_2}] \right) \\ &+ \frac{1}{\left(\frac{n_1(n_1-1)}{2} + \frac{n_2(n_2-1)}{2}\right)\gamma_P + \gamma_H} \left(\frac{n_1(n_1-1)}{2} \gamma_P \mathbb{E}[L_{2,(n_1-1,n_2)}] + \frac{n_2(n_2-1)}{2} \gamma_P \mathbb{E}[L_{2,(n_1,n_2-1)}] \right).\end{aligned}\tag{2.122}$$

In the above equation, the first line in the right hand side is known by (2.112) and we can recognize $W_0(n_1, n_2)$ in the first fraction. Iterating the formula (2.122) for $\mathbb{E}[L_{2,(n_1-1,n_2)}]$ and $\mathbb{E}[L_{2,(n_1,n_2-1)}]$, we obtain for the second line:

- the term $(n_1 + n_2 - 1 + \gamma_H \mathbb{E}[L_{1,n_1+n_2-1}])$ (which depends only on the sum of the coordinates $(n_1 - 1, n_2)$ or $(n_1, n_2 - 1)$, and is hence constant on the diagonal $x + y = n_1 + n_2 - 1$), in factor of

$$\begin{aligned}&\frac{\frac{n_1(n_1-1)}{2} \gamma_P}{\left(\frac{n_1(n_1-1)}{2} + \frac{n_2(n_2-1)}{2}\right)\gamma_P + \gamma_H} \times \frac{1}{\left(\frac{(n_1-1)(n_1-2)}{2} + \frac{n_2(n_2-1)}{2}\right)\gamma_P + \gamma_H} \\ &+ \frac{\frac{n_2(n_2-1)}{2} \gamma_P}{\left(\frac{n_1(n_1-1)}{2} + \frac{n_2(n_2-1)}{2}\right)\gamma_P + \gamma_H} \times \frac{1}{\left(\frac{n_1(n_1-1)}{2} + \frac{(n_2-1)(n_2-2)}{2}\right)\gamma_P + \gamma_H},\end{aligned}$$

which is the weight $W_1(n_1, n_2)$.

- terms involving the four paths of length 2 $((\leftarrow, \leftarrow), (\downarrow, \downarrow), (\leftarrow, \downarrow), (\downarrow, \leftarrow))$:

$$\begin{aligned}&\frac{\frac{n_1(n_1-1)}{2} \gamma_P}{\left(\frac{n_1(n_1-1)}{2} + \frac{n_2(n_2-1)}{2}\right)\gamma_P + \gamma_H} \times \frac{\frac{(n_1-1)(n_1-2)}{2} \gamma_P}{\left(\frac{(n_1-1)(n_1-2)}{2} + \frac{n_2(n_2-1)}{2}\right)\gamma_P + \gamma_H} \times \mathbb{E}[L_{2,(n_1-2,n_2)}] \\ &+ \frac{\frac{n_2(n_2-1)}{2} \gamma_P}{\left(\frac{n_1(n_1-1)}{2} + \frac{n_2(n_2-1)}{2}\right)\gamma_P + \gamma_H} \times \frac{\frac{(n_2-1)(n_2-2)}{2} \gamma_P}{\left(\frac{n_1(n_1-1)}{2} + \frac{(n_2-1)(n_2-2)}{2}\right)\gamma_P + \gamma_H} \times \mathbb{E}[L_{2,(n_1,n_2-2)}] \\ &+ \left\{ \frac{\frac{n_1(n_1-1)}{2} \gamma_P}{\left(\frac{n_1(n_1-1)}{2} + \frac{n_2(n_2-1)}{2}\right)\gamma_P + \gamma_H} \times \frac{\frac{n_2(n_2-1)}{2} \gamma_P}{\left(\frac{(n_1-1)(n_1-2)}{2} + \frac{n_2(n_2-1)}{2}\right)\gamma_P + \gamma_H} \right. \\ &\left. + \frac{\frac{n_2(n_2-1)}{2} \gamma_P}{\left(\frac{n_1(n_1-1)}{2} + \frac{n_2(n_2-1)}{2}\right)\gamma_P + \gamma_H} \times \frac{\frac{n_1(n_1-1)}{2} \gamma_P}{\left(\frac{n_1(n_1-1)}{2} + \frac{(n_2-1)(n_2-2)}{2}\right)\gamma_P + \gamma_H} \right\} \times \mathbb{E}[L_{2,(n_1-1,n_2-1)}].\end{aligned}$$

Using formula (2.122) again, the three expectations above can be replaced by a term involving $\mathbb{E}[L_{1,n_1+n_2-2}]$ and terms corresponding to paths of length 3.

More generally, reiterating the formula (2.122), we obtain the terms $(n_1 + n_2 - \ell + \gamma_H \mathbb{E}[L_{1,n_1+n_2-\ell}])$, for $0 \leq \ell \leq n_1 + n_2 - 2$. The multiplicative factor for the term associated with ℓ corresponds to the weights of all south-west paths linking (n_1, n_2) to the diagonal $x + y = n_1 + n_2 - \ell$. The reiteration stops when reaching $\mathbb{E}[L_{2,1}]$ that is computed in (2.113). \square

To conclude this section, let us mention that the formulas computed above could serve for the statistical estimation of the parameters γ_H and γ_P . If we observe the lengths $(L_{m,\underline{n}}^{(k)}, 1 \leq k \leq K)$ of K sampled subtrees of the two-level Kingman measure tree, then, it is possible to approximate $\mathbb{E}[L_{m,\underline{n}}]$ by the empirical mean of the tree lengths:

$$\frac{1}{K} \sum_{k=1}^K L_{m,\underline{n}}^{(k)}.$$

The moment estimators of γ_H and γ_P are then defined as the values that minimize the distance between the vectors of empirical means and the vector of expectations, for various values of m and \underline{n} .

2.6 Proofs of the main results

2.6.1 Proof of Theorem 2.11

We prove here that the sequence of two-level tree-valued Moran dynamics is tight in $\mathcal{D}_{\mathbb{U}(2)}[0, \infty)$ and that the potential limit points take values in the space of compact ultra-metric m2m spaces. For this, we prove the criterion for compact containment of Proposition 2.12 and that this criterion is satisfied by the family of two-level tree-valued Moran dynamics.

Recall the distance distribution w_μ of μ from (2.21). Recall also from Proposition 2.4 that a characterization of relative compactness in $\mathbb{M}^{(2)}$ can be expressed with respect to the corresponding metric measure spaces arising from the intensity measure. For this reason, the proof follows literally (apart from a few additional comments) the proof of Proposition 2.22 in [GPW13] with the (one-level) measure μ replaced by the intensity measure M_ν .

Compact containment. We now prove Proposition 2.12. The proof relies on two lemmata that give bounds on the distance distribution and the number of ancestors, that are used in showing the sufficient conditions for relative compactness in Proposition 2.5. The first bounds are related to Assumption (i) of the criterion for the compact containment condition to hold.

Lemma 2.24 (Bounds on the distance distribution under Assumption (i)). *Fix $T > 0$, and assume that $\{\mathcal{U}_0^{M,N} : M, N \in \mathbb{N}\}$ is tight in $\mathbb{U}^{(2)}$. If Condition (i) of Proposition 2.12 holds, then for all $\epsilon > 0$, there is a $C_\epsilon > 0$ such that*

$$\sup_{M,N \in \mathbb{N}} \mathbb{P}^{M,N} \left\{ \sup_{t \in [0, T \wedge \tau^{M,N})} w_{M_{\nu_t}^{M,N}}([C_\epsilon, +\infty)) > \epsilon \right\} \leq \epsilon. \quad (2.123)$$

The proof uses the following lemma, applied to $\{\mathcal{U}_0^{M,N} : M, N \in \mathbb{N}\}$.

Lemma 2.25. *If a family $\{\mathcal{X}_i\}_{i \in I} = \{(X_i, r_i, \nu_i)\}_{i \in I}$ of random variables in $\mathbb{M}^{(2)}$ is tight, then for all $\epsilon > 0$, there exists $C_\epsilon > 0$ such that for all $i \in I$, there is a random subset $X_{\epsilon,i} \subseteq X_i$ with $M_{\nu_i}(X_{\epsilon,i}) \geq 1 - \epsilon$ and*

$$\mathbb{P} \{ \text{diam}(X_{\epsilon,i}) \leq C_\epsilon \} \geq 1 - \epsilon, \quad (2.124)$$

where $\text{diam}(X_{\epsilon,i})$ denotes the diameter of the set $X_{\epsilon,i}$.

Proof. Suppose that $\{\mathcal{X}_i\}_{i \in I}$ is tight and fix $\epsilon > 0$. By definition, we can find a compact set $\Gamma_\epsilon \in \mathbb{M}^{(2)}$ such that $\inf_{i \in I} \mathbb{P}(\mathcal{X}_i \in \Gamma_\epsilon) > 1 - \epsilon$. Since Γ_ϵ is compact, there exists, by [Mei19, Theorem 7.2], $C_\epsilon > 0$ such that for all $\mathcal{X} = (X, r, \nu) \in \Gamma_\epsilon$, there exists $X_{\epsilon, \mathcal{X}} \subseteq X$ with $M_\nu(X_{\epsilon, \mathcal{X}}) \geq 1 - \epsilon$ and $X_{\epsilon, \mathcal{X}}$ has diameter at most C_ϵ .

Therefore, for all $i \in I$,

$$\mathbb{P}\{\exists X_{\epsilon, i} \subseteq X_i, \ M_{\nu_i}(X_{\epsilon, i}) \geq 1 - \epsilon, \ \text{diam}(X_{\epsilon, i}) \leq C_\epsilon\} \geq 1 - \epsilon. \quad (2.125)$$

On the event that $\mathcal{X}_i \notin \Gamma_\epsilon$, it is always possible to take $X_{\epsilon, i} = X_i$ that verifies $M_{\nu_i}(X_{\epsilon, i}) \geq 1 - \epsilon$ (but not necessarily $\text{diam}(X_{\epsilon, i}) \leq C_\epsilon$). And the assertion follows. \square

Proof of Lemma 2.24. Since $\{\mathcal{U}_0^{M, N} : M, N \in \mathbb{N}\}$ is tight, we control with high probability the diameter of a subset that carry most of the mass (for the intensity measure), due to the previous lemma. We then control the mass of the descendants of the complement of this subset with Assumption (i).

Let $\epsilon > 0$. Choose $\delta = \delta(\frac{\epsilon}{4}) > 0$ such that (2.58) holds with $\epsilon/4$ (instead of ϵ) for all $M, N \in \mathbb{N}$ and $\mathcal{A}_s^{M, N}$ -measurable random subsets $\mathcal{J}^{M, N} \subseteq \mathcal{I}_0^{M, N}$ with $M_{\nu_t^{M, N}}(\mathcal{J}^{M, N}) \leq \delta$, that is,

$$\sup_{M, N \in \mathbb{N}} \mathbb{P}^{M, N} \left\{ \sup_{t \in [s, T \wedge \tau^{M, N})} M_{\nu_t^{M, N}}(D_t(\mathcal{J}^{M, N}, s)) > \frac{\epsilon}{4} \right\} \leq \frac{\epsilon}{4}. \quad (2.126)$$

Since $\{\mathcal{U}_0^{M, N} : M, N \in \mathbb{N}\}$ is tight in $\mathbb{U}^{(2)}$, we can find, with Lemma 2.25, a constant $\tilde{C}_\epsilon > 0$ and a sequence of random sets $\{\mathcal{J}^{M, N}\}_{M, N}$ such that for all $M, N \in \mathbb{N}$, $\mathcal{J}^{M, N}$ is a measurable random subset $\mathcal{J}^{M, N} \subseteq \mathcal{I}_0^{M, N}$ with $M_{\nu_t^{M, N}}(\mathcal{J}^{M, N}) \leq \delta$, and

$$\sup_{M, N \in \mathbb{N}} \mathbb{P}^{M, N} \left\{ \text{diam}(\mathcal{I}_0^{M, N} \setminus \mathcal{J}^{M, N}) \leq \tilde{C}_\epsilon \right\} > 1 - \frac{\epsilon}{2}. \quad (2.127)$$

On the event that $\mathcal{I}_0^{M, N} \setminus \mathcal{J}^{M, N}$ has diameter at most \tilde{C}_ϵ , the set $D_t(\mathcal{I}_0^{M, N} \setminus \mathcal{J}^{M, N}, 0)$ of descendants of $\mathcal{I}_0^{M, N} \setminus \mathcal{J}^{M, N}$ at time t has diameter at most $\tilde{C}_\epsilon + 2t$. Therefore, on this event,

$$\begin{aligned} w_{M_{\nu_t^{M, N}}}([\tilde{C}_\epsilon + 2T, \infty)) &= M_{\nu_t^{M, N}}^{\otimes 2} \left\{ (x, x') \in \mathcal{I}_t^{M, N} \times \mathcal{I}_t^{M, N} : r_t^{M, N}(x, x') \geq \tilde{C}_\epsilon + 2T \right\} \\ &\leq M_{\nu_t^{M, N}}^{\otimes 2} \left\{ (x, x') \in \mathcal{I}_t^{M, N} \times \mathcal{I}_t^{M, N} : x \in D_t(\mathcal{J}^{M, N}, 0) \text{ or } x' \in D_t(\mathcal{J}^{M, N}, 0) \right\} \\ &\leq 2M_{\nu_t^{M, N}}(D_t(\mathcal{J}^{M, N}, 0)). \end{aligned} \quad (2.128)$$

Now, partitioning on the event that $\mathcal{I}_0^{M, N} \setminus \mathcal{J}^{M, N}$ has diameter at most \tilde{C}_ϵ and its complement,

$$\begin{aligned} &\sup_{M, N \in \mathbb{N}} \mathbb{P}^{M, N} \left\{ \sup_{t \in [0, T \wedge \tau^{M, N})} w_{M_{\nu_t^{M, N}}}([\tilde{C}_\epsilon + 2T, \infty)) > \epsilon \right\} \\ &\leq \sup_{M, N \in \mathbb{N}} \mathbb{P}^{M, N} \left\{ \sup_{t \in [0, T \wedge \tau^{M, N})} w_{M_{\nu_t^{M, N}}}([\tilde{C}_\epsilon + 2T, \infty)) > \epsilon, \ \text{diam}(\mathcal{I}_0^{M, N} \setminus \mathcal{J}^{M, N}) \leq \tilde{C}_\epsilon \right\} \\ &\quad + \sup_{M, N \in \mathbb{N}} \mathbb{P}^{M, N} \left\{ \text{diam}(\mathcal{I}_0^{M, N} \setminus \mathcal{J}^{M, N}) \geq \tilde{C}_\epsilon \right\} \\ &\leq \sup_{M, N \in \mathbb{N}} \mathbb{P}^{M, N} \left\{ \sup_{t \in [0, T \wedge \tau^{M, N})} M_{\nu_t^{M, N}}(D_t(\mathcal{J}^{M, N}, 0)) > \frac{\epsilon}{4}, \ \text{diam}(\mathcal{I}_0^{M, N} \setminus \mathcal{J}^{M, N}) \leq \tilde{C}_\epsilon \right\} + \frac{\epsilon}{2} \\ &\leq \epsilon. \end{aligned} \quad (2.129)$$

\square

Lemma 2.26 (Uniform bounds on $S_{2\epsilon}$ and $\tilde{S}_{2\epsilon}$). *Fix $T > 0$.*

(a) Assume Condition (ii.i) from Proposition 2.12. Then for all $\epsilon > 0$ we can find $C_\epsilon > 0$ such that

$$\sup_{M,N \in \mathbb{N}} \mathbb{P}^{M,N} \left\{ \sup_{t \in [\epsilon, T \wedge \tau^{M,N})} S_{2\epsilon}(\mathcal{U}_t^{M,N}) > C_\epsilon \right\} \leq 2\epsilon. \quad (2.130)$$

(b) Assume that the family $\{\mathcal{U}_0^{M,N} : M, N \in \mathbb{N}\}$ is tight in $\mathbb{U}^{(2)}$ and Conditions (i) and (ii.ii) from Proposition 2.12. Then for all $\epsilon > 0$ we can find $C_\epsilon > 0$ such that

$$\sup_{M,N \in \mathbb{N}} \mathbb{P}^{M,N} \left\{ \sup_{t \in [0, T \wedge \tau^{M,N})} \tilde{S}_{2\epsilon}(\mathcal{U}_t^{M,N}) > C_\epsilon \right\} \leq 2\epsilon. \quad (2.131)$$

Proof. (a) Fix $\epsilon > 0$. Without loss of generality, we assume that $T = k\epsilon$ for some $k \in \mathbb{N}$. Since we set $S_{2\epsilon}(\mathcal{U}_t) = 0$ for $t \geq \tau$, we can also assume that $\tau^{M,N} \geq T$ for all $M, N \in \mathbb{N}$. For each $t \in [\epsilon, T)$, we can control the number of ancestors at time $t - \epsilon$ with high probability. In order to get a uniform control over all $t \in [\epsilon, T)$, we partition $[\epsilon, T)$ in k intervals and control $S_{2\epsilon}(\mathcal{U}_t^{M,N})$ on each of these intervals by using the fact that for all t, t', ϵ and ϵ' such that $[t - \epsilon, t] \subseteq [t' - \epsilon', t']$,

$$S_{2\epsilon}(\mathcal{U}_t^{M,N}) \geq S_{2\epsilon'}(\mathcal{U}_{t'}^{M,N}). \quad (2.132)$$

Since for all $t \in [\epsilon, T)$ the family $\{S_{2\epsilon}(\mathcal{U}_t^{M,N}) : M, N \in \mathbb{N}\}$ is tight by assumption, there exists a $C_\epsilon > 0$ such that for all $M, N \in \mathbb{N}$,

$$\sum_{i=2}^{2k-1} \mathbb{P}^{M,N} \left\{ S_\epsilon(\mathcal{U}_{i\frac{\epsilon}{2}}^{M,N}) > C_\epsilon \right\} \leq 2\epsilon. \quad (2.133)$$

Therefore, applying (2.132) yields that for all $M, N \in \mathbb{N}$,

$$\begin{aligned} \mathbb{P}^{M,N} \left\{ \sup_{t \in [\epsilon, T)} S_{2\epsilon}(\mathcal{U}_t^{M,N}) > C_\epsilon \right\} &\leq \sum_{i=2}^{2k-1} \mathbb{P}^{M,N} \left\{ \sup_{t \in [i\frac{\epsilon}{2}, (i+1)\frac{\epsilon}{2})} S_{2\epsilon}(\mathcal{U}_t^{M,N}) > C_\epsilon \right\} \\ &\leq \sum_{i=2}^{2k-1} \mathbb{P}^{M,N} \left\{ S_\epsilon(\mathcal{U}_{i\frac{\epsilon}{2}}^{M,N}) > C_\epsilon \right\} \leq 2\epsilon, \end{aligned} \quad (2.134)$$

and the assertion follows.

(b) We use a line of arguments similar to case (a). The difference here is that we need to add a control of the part of the population that is not covered by the descendants, which is possible with Assumption (i). We define for $\epsilon > 0$ and $0 < \zeta < 1$,

$$\tilde{S}_{2\epsilon, \zeta}(\mathcal{U}_t^{M,N}) := \inf_{\mathcal{J} \subseteq \mathcal{I}_t : M_{\nu_t^{M,N}}(\mathcal{J}) \leq \zeta} \# \{A_{t-\epsilon}(i, t) : i \in \mathcal{I}_t \setminus \mathcal{J}\}. \quad (2.135)$$

We have in particular, $\tilde{S}_{2\epsilon}(\mathcal{U}_t^{M,N}) = \tilde{S}_{2\epsilon, 2\epsilon}(\mathcal{U}_t^{M,N})$, and thus for all $0 < \zeta < 1$ and $t \in [\epsilon, T)$, the family $\{\tilde{S}_{2\epsilon, \zeta}(\mathcal{U}_t^{M,N}) : M, N \in \mathbb{N}\}$ is tight by Assumption (ii.ii).

Let t, t', δ and δ' be such that $[t - \delta, t] \subseteq [t' - \delta', t']$. By definition of $\tilde{S}_{2\delta, \zeta}(\mathcal{U}_t^{M,N})$, for all $0 < \zeta < 1$, $t < \tau^{M,N}$ and $M, N \in \mathbb{N}$ there is a $\mathcal{A}_t^{M,N}$ -measurable subset $\mathcal{J}^{M,N, \zeta, t} \subseteq \mathcal{I}_t^{M,N}$ such that $M_{\nu_t^{M,N}}(\mathcal{J}^{M,N, \zeta, t}) \leq \zeta$ and $\mathcal{I}_t^{M,N} \setminus \mathcal{J}^{M,N, \zeta, t}$ can be covered by $\tilde{S}_{2\delta, \zeta}(\mathcal{U}_t^{M,N})$ balls of radius 2δ . Moreover, for all $\zeta, \zeta' \in (0, 1)$,

$$\left\{ \tilde{S}_{2\delta, \zeta}(\mathcal{U}_t^{M,N}) < \tilde{S}_{2\delta', \zeta'}(\mathcal{U}_{t'}^{M,N}) \right\} \subseteq \left\{ M_{\nu_{t'}^{M,N}}(D_{t'}(\mathcal{J}^{M,N, \zeta, t}, t)) > \zeta' \right\}, \quad (2.136)$$

and hence

$$\begin{aligned} \mathbb{P}^{M,N} \left\{ \tilde{S}_{2\delta, \zeta}(\mathcal{U}_t^{M,N}) < \sup_{t' \in [t, (t-\delta)+\delta')} \tilde{S}_{2\delta', \zeta'}(\mathcal{U}_{t'}^{M,N}) \right\} \\ \leq \mathbb{P}^{M,N} \left\{ \sup_{t' \in [t, (t-\delta)+\delta')} M_{\nu_{t'}^{M,N}}(D_{t'}(\mathcal{J}^{M,N, \zeta, t}, t)) > \zeta' \right\}. \end{aligned} \quad (2.137)$$

Fix $T > 0$ and $\epsilon > 0$. Without loss of generality, we assume that $T = k\epsilon$ for some $k \in \mathbb{N}$. Since we set $\tilde{S}_{2\epsilon}(\mathcal{U}_t) = 0$ for $t \geq \tau$, we can also assume that $\tau^{M,N} \geq T$ for all $M, N \in \mathbb{N}$. By Condition (i) of Proposition 2.12 applied ($2k$ times) with $s = i\frac{\epsilon}{2}$, $i = 0, \dots, 2k-1$, we can choose a $\zeta = \zeta(\epsilon, T)$ suitably small such that for each $i = 0, \dots, 2k-1$ and for all $\mathcal{A}_{i\frac{\epsilon}{2}}^{M,N}$ -measurable subsets $\mathcal{J}^{M,N,\zeta,i\frac{\epsilon}{2}} \subseteq \mathcal{I}_{i\frac{\epsilon}{2}}^{M,N}$ with $M_{\nu_s^{M,N}}(\mathcal{J}^{M,N,\zeta,i\frac{\epsilon}{2}}) \leq \zeta$,

$$\sup_{M,N} \mathbb{P}^{M,N} \left\{ \sup_{t \in [i\frac{\epsilon}{2}, (i+1)\frac{\epsilon}{2})} M_{\nu_s^{M,N}} \left(D_t(\mathcal{J}^{M,N,\zeta,i\frac{\epsilon}{2}}, i\frac{\epsilon}{2}) \right) > \epsilon \right\} \leq \frac{\epsilon}{2k}. \quad (2.138)$$

Thus, inserting (2.138) into (2.137) applied with $t = i\frac{\epsilon}{2}$, $\delta = \frac{\epsilon}{2}$, $\delta' = \zeta' = \epsilon$, and ζ from (2.138),

$$\sup_{M,N} \mathbb{P}^{M,N} \left\{ \tilde{S}_{\epsilon,\zeta}(\mathcal{U}_{i\frac{\epsilon}{2}}^{M,N}) < \sup_{t \in [i\frac{\epsilon}{2}, (i+1)\frac{\epsilon}{2})} \tilde{S}_{2\epsilon,2\epsilon}(\mathcal{U}_t^{M,N}) \right\} \leq \frac{\epsilon}{2k}. \quad (2.139)$$

Since for all $\zeta \in (0, 1)$, $t \in [\epsilon, T)$ the family $\{\tilde{S}_{2\epsilon,\zeta}(\mathcal{U}_t^{M,N}) : M, N \in \mathbb{N}\}$ is tight by Assumption (ii.ii), and $\{\mathcal{U}_0^{M,N} : M, N \in \mathbb{N}\}$ is assumed to be tight as well, there exists a $C_\epsilon > 0$ such that for all $N \in \mathbb{N}$,

$$\sum_{i=0}^{2k-1} \mathbb{P}^{M,N} \left\{ \tilde{S}_{\epsilon,\zeta}(\mathcal{U}_{i\frac{\epsilon}{2}}^{M,N}) > C_\epsilon \right\} \leq \epsilon. \quad (2.140)$$

Therefore,

$$\begin{aligned} & \sup_{M,N \in \mathbb{N}} \mathbb{P}^{M,N} \left\{ \sup_{t \in [0,T)} \tilde{S}_{2\epsilon}(\mathcal{U}_t^{M,N}) > C_\epsilon \right\} \\ & \leq \sup_{M,N \in \mathbb{N}} \sum_{i=0}^{2k-1} \mathbb{P}^{M,N} \left\{ \sup_{t \in [i\frac{\epsilon}{2}, (i+1)\frac{\epsilon}{2})} \tilde{S}_{2\epsilon}(\mathcal{U}_t^{M,N}) > C_\epsilon \right\} \\ & \leq \sup_{M,N \in \mathbb{N}} \sum_{i=0}^{2k-1} \mathbb{P}^{M,N} \left\{ \tilde{S}_{\epsilon,\zeta}(\mathcal{U}_{i\frac{\epsilon}{2}}^{M,N}) < \sup_{t \in [i\frac{\epsilon}{2}, (i+1)\frac{\epsilon}{2})} \tilde{S}_{2\epsilon}(\mathcal{U}_t^{M,N}) \right\} \\ & \quad + \sum_{i=0}^{2k-1} \mathbb{P}^{M,N} \left\{ \tilde{S}_{\epsilon,\zeta}(\mathcal{U}_{i\frac{\epsilon}{2}}^{M,N}) > C_\epsilon \right\} \\ & \leq 2\epsilon, \end{aligned} \quad (2.141)$$

and the assertion follows. \square

We can now conclude the:

Proof of Proposition 2.12. Fix $T > 0$ and $\delta > 0$.

(a) Since Conditions (i) and (ii.i) from Proposition 2.12 hold, we find for all $n \in \mathbb{N}$ a $C_{\delta 2^{-n}} > 0$ such that (2.123) and (2.130) hold with $\epsilon = \delta 2^{-n}$. Put

$$\Gamma_{1,\delta} := \left\{ (X, r, \nu) \in \mathbb{U}^{(2)} : w_{M_\nu}([C_{\delta 2^{-n}}, \infty)) \leq \delta 2^{-n}, \text{ for all } n \in \mathbb{N} \right\} \quad (2.142)$$

and

$$\Gamma_{2,\delta} := \left\{ (X, r, \nu) \in \mathbb{U}_c^{(2)} : S_{2\delta 2^{-n}}((X, r, \nu)) \leq C_{\delta 2^{-n}}, \text{ for all } n \in \mathbb{N} \right\}, \quad (2.143)$$

where $S_{2\delta 2^{-n}}((X, r, \nu))$ denotes the number of balls of radius $\delta 2^{-n}$ needed to cover $\text{supp}(M_\nu)$. By Proposition 2.6, we know that $\Gamma_{1,\delta} \cap \Gamma_{2,\delta}$ is relatively compact in $\mathbb{U}_c^{(2)}$. Since for all $n \in \mathbb{N}$ (2.123)

holds for $C_{\delta 2^{-n}}$ and $\epsilon = \delta 2^{-n}$, we have

$$\begin{aligned} & \inf_{M,N \in \mathbb{N}} \mathbb{P}^{M,N} \left\{ \mathcal{U}_t^{M,N} \in \Gamma_{1,\delta} \text{ for all } t \in [0, T \wedge \tau^{M,N}) \right\} \\ & \geq 1 - \sum_{n=1}^{\infty} \sup_{M,N \in \mathbb{N}} \mathbb{P}^{M,N} \left\{ \sup_{t \in [0, T \wedge \tau^{M,N})} w_{M_{\nu_T^{M,N}}}([C_{\delta 2^{-n}}, \infty)) > \delta 2^{-n} \right\} \\ & \geq 1 - \sum_{n=1}^{\infty} \delta 2^{-n} = 1 - \delta. \end{aligned} \quad (2.144)$$

With (2.130) we show in a similar way that

$$\inf_{M,N \in \mathbb{N}} \mathbb{P}^{M,N} \left\{ \mathcal{U}_t^{M,N} \in \Gamma_{1,\delta} \text{ for all } t \in [\delta, T \wedge \tau^{M,N}) \right\} \geq 1 - 2\delta, \quad (2.145)$$

and thus,

$$\inf_{M,N \in \mathbb{N}} \mathbb{P}^{M,N} \left\{ \mathcal{U}_t^{M,N} \in \Gamma_{1,\delta} \cap \Gamma_{2,\delta} \text{ for all } t \in [\delta, T \wedge \tau^{M,N}) \right\} \geq 1 - 3\delta, \quad (2.146)$$

and the assertion (a) holds.

(b) Since Conditions (i) and (ii.ii) from Proposition 2.12 hold, we find for all $n \in \mathbb{N}$ a $\tilde{C}_{\delta 2^{-n}} > 0$ such that (2.123) and (2.131) hold with $\epsilon = \delta 2^{-n}$. Put

$$\Gamma_{3,\delta} := \left\{ (X, r, \nu) \in \mathbb{U}^{(2)} : \tilde{S}_{2\delta 2^{-n}}((X, r, \nu)) \leq \tilde{C}_{\delta 2^{-n}}, \text{ for all } n \in \mathbb{N} \right\}, \quad (2.147)$$

where $\tilde{S}_{2\delta 2^{-n}}((X, r, \nu))$ is the number of balls of radius $\delta 2^{-n}$ needed to cover a frequency pf $(1 - 2\delta 2^{-n})$ of $\text{supp}(M_\nu)$. By Proposition 2.5, we know that $\Gamma_{1,\delta} \cap \Gamma_{3,\delta}$ is relatively compact in $\mathbb{U}^{(2)}$. By a similar argument as for $\Gamma_{1,\delta}$, we find using (2.131) that

$$\inf_{M,N \in \mathbb{N}} \mathbb{P}^{M,N} \left\{ \mathcal{U}_t^{M,N} \in \Gamma_{3,\delta} \text{ for all } t \in [0, T \wedge \tau^{M,N}) \right\} \geq 1 - 2\delta, \quad (2.148)$$

and the assertion (b) follows. \square

The compact containment condition for the resampling dynamics. Let us now apply the Proposition 2.12 to the two-level tree-valued Moran models:

Proposition 2.27. *For each $M, N \in \mathbb{N}$, let $\mathcal{U}^{M,N} = (\mathcal{U}_t^{M,N})_{t \geq 0}$ be the tree-valued Moran dynamics with M hosts and N parasites in each host. Assume that the family $\{\mathcal{U}_0^{M,N} : M, N \in \mathbb{N}\}$ is tight in $\mathcal{M}_1(\mathbb{U}^{(2)})$. Then the family $\{\mathcal{U}^{M,N} : M, N \in \mathbb{N}\}$ satisfies the Conditions (i), (ii.i) and (ii.ii) from Proposition 2.12.*

Proof. Fix T and ϵ with $0 < \epsilon < T$. We have here that for all $M, N \in \mathbb{N}$, $\tau^{M,N} = \infty$.

To show that Condition (i) holds, consider $s \in [0, T)$ and a sequence $(\mathcal{J}^{M,N})_{M,N}$ of random subsets such that for all $M, N \in \mathbb{N}$, $\mathcal{J}^{M,N}$ is a $\mathcal{A}_s^{M,N}$ -measurable random subset of $\mathcal{I}^{M,N}$. We write $Y^{M,N} = (Y_t^{M,N})_{t \in [s, T)}$ for the process defined for $t \in [s, T)$ by

$$Y_t^{M,N} := M_{\nu_t^{M,N}}(D_t(\mathcal{J}^{M,N}, s)) = \frac{\#D_t(\mathcal{J}^{M,N}, s)}{MN}. \quad (2.149)$$

It is a birth-death process with values in $\{0, \frac{1}{MN}, \dots, 1\}$ and with following jumps and rates:

- with rate $\frac{1}{2}M^2N^2\gamma_P y(1-y)$, the process jumps from its current state y to $y + \frac{1}{MN}$,
- with rate $\frac{1}{2}M^2N^2\gamma_P y(1-y)$, the process jumps from its current state y to $y - \frac{1}{MN}$,
- with rate $\frac{1}{2}M^2\gamma_H y(1-y)$, the process jumps from its current state y to $y + \frac{1}{M}$,
- with rate $\frac{1}{2}M^2\gamma_H y(1-y)$, the process jumps from its current state y to $y - \frac{1}{M}$.

The process $Y^{M,N}$ is a martingale so by Doob's maximum inequality, we have

$$\mathbb{P}^{M,N} \left\{ \sup_{t \in [s,T]} Y_t^{M,N} \geq \epsilon \right\} \leq \frac{\mathbb{E}[Y_T^{M,N}]}{\epsilon} = \frac{\mathbb{E}[Y_s^{M,N}]}{\epsilon}. \quad (2.150)$$

So taking $\delta = \epsilon^2$, Condition (i) follows.

For Condition (ii.i), fix $0 < t < T$. We claim that the family $\{S_{2\epsilon}(\mathcal{U}_t^{M,N}) : M, N \in \mathbb{N}\}$ is stochastically uniformly bounded by $\#\kappa_P(\epsilon)$ where $(\kappa_H, \kappa_P) = ((\kappa_H(t), \kappa_P(t)))_{t \geq 0}$ denotes the two-partitions-valued nested Kingman coalescent with rates γ_H and γ_P as defined in Subsection 2.4.3. To show this, fix $M, N \in \mathbb{N}$ and let $\#\kappa_P^{M,N}$ be the process for the number of blocks in a finite nested Kingman coalescent with initial state space $\{1, \dots, M\} \times \{1, \dots, N\}$. We build a \mathbb{N} -valued stochastic process $(K_\epsilon)_{\epsilon \geq 0}$ with the same distribution as $\#\kappa_P$ and such that for all $\epsilon \geq 0$, $S_{2\epsilon}(\mathcal{U}_t^{M,N}) \leq K_\epsilon$.

Consider $\eta_H = \{\eta_H^{k,l} : 1 \leq k, l \leq M\}$ an independent family of rate $\frac{\gamma_H}{2}$ -Poisson processes, and $\eta_P = \{\eta_P^{(i,k),(i,l)} : 1 \leq i \leq M, 1 \leq k, l \leq N\}$ an independent family of rate $\frac{\gamma_P}{2}$ -Poisson processes, independent of the family η_H that define the two-level tree-valued Moran dynamics $\mathcal{U}^{M,N}$. For $\epsilon = 0$, $S_0(\mathcal{U}_t^{M,N}) = MN$ so put

$$K_0 := MN. \quad (2.151)$$

We define K as a death process that jumps down by 1 at each ϵ such that $t - \epsilon$ is the time of a resampling of two parasites that are ancestors of \mathcal{I}_t at time $t - \epsilon$. More precisely, for all ϵ such that $t - \epsilon = \eta_P^{(i,k),(i,l)}$ for some $(i,k), (i,l) \in \{A_{t-\epsilon}(\alpha) : \alpha \in \mathcal{I}_t\}$, the process K jumps from $K_{\epsilon-}$ to

$$K_\epsilon := K_{\epsilon-} - 1. \quad (2.152)$$

At the time of a resampling between two hosts, note that $(S_{2\epsilon}(\mathcal{U}_t^{M,N}))$ might decrease, but we keep K constant. For this reason, $S_{2\epsilon}(\mathcal{U}_t^{M,N}) \leq K_\epsilon$ for all $\epsilon \geq 0$ and it is easy to see that for all $\epsilon \geq 0$, K_ϵ has the same distribution as $\#\kappa_P^{M,N}$. Thus, $S_{2\epsilon}(\mathcal{U}_t^{M,N})$ is stochastically bounded by $\#\kappa_P^{M,N}$, which is bounded itself by $\#\kappa_P$. To conclude we know from [BDLS18, Section 6] that for every $t > 0$, $\kappa_P(t)$ almost surely consists of only finitely many blocks, and thus the family $\{S_{2\epsilon}(\mathcal{U}_t^{M,N}) : M, N \in \mathbb{N}\}$ is tight.

Condition (ii.ii) easily follows from Conditions (ii.i). \square

We can now conclude the proof of the tightness of the two-level tree-valued Moran dynamics.

Proof of Theorem 2.11. Let, for each $M, N \in \mathbb{N}$, $(\mathcal{U}_t^{M,N})_{t \geq 0}$ be the two-level tree-valued Moran dynamics with population sizes (M, N) such that the family $\{\mathcal{U}_0^{M,N} : M, N \in \mathbb{N}\}$ is tight in $\mathbb{U}^{(2)}$. We use Remark 4.5.2 in [EK86] that states that the family $\{\mathcal{U}^{M,N} : M, N \in \mathbb{N}\}$ is relatively compact if

- (i) the compact containment condition (2.60) holds,
- (ii) the closure of Π^1 contains an algebra that separates points and vanishes nowhere, and
- (iii) for all $\Phi \in \Pi^1$, (2.71) and (2.72) hold.

Propositions 2.12 and 2.27 give (i), and Proposition 2.16 gives (iii). For (ii), we know from Remark 2.13 that Π^1 is an algebra that separates points. It vanishes nowhere because it contains constant functions. \square

2.6.2 Proof of Theorem 2.14

Before finishing up the proof of Theorem 2.14, we show the rewriting of the resampling generators for the two-level tree-valued Moran dynamics of Lemma 2.10. The key idea is that if any of the hosts and parasites chosen for a resampling event do not appear in the integral sample, the average over all possible samples after the event will be compensated by the average before the event.

Proof of Lemma 2.10. (2.51) is just a rewriting of (2.45). For the resampling of hosts, we use (2.54) to write that

$$\Omega_{M,N}^{\uparrow,H} \Phi_{M,N}^{\phi}(\chi) = \frac{\gamma_H}{2} \sum_{1 \leq k, l \leq M} \left(\left\langle \left(\vartheta^{k,l}(\nu) \right)^{\otimes \downarrow m}, \psi_{\downarrow} \right\rangle - \left\langle \nu^{\otimes \downarrow m}, \psi_{\downarrow} \right\rangle \right). \quad (2.153)$$

Furthermore, note that if $\nu = \frac{1}{M} \sum_{i=1}^M \delta_{\mu'_i}$ for some measures $\mu'_i \in \mathcal{M}_1(T)$, then for all function $\psi_{\downarrow}: (\mathcal{M}_1(T))^m \rightarrow \mathbb{R}$,

$$\left\langle \nu^{\otimes \downarrow m}, \psi_{\downarrow} \right\rangle := \int_{(\mathcal{M}_1(T))^m} \nu^{\otimes \downarrow m}(\underline{\mu}) \psi_{\downarrow}(\underline{\mu}) = \frac{1}{M!} \sum_{\sigma \in \mathcal{S}_M} \psi_{\downarrow}(\mu'_{\sigma(1)}, \dots, \mu'_{\sigma(m)}), \quad (2.154)$$

where \mathcal{S}_M is the set of all permutations of $\{1, \dots, M\}$. This is true because we sample without repetition. Now, using again the notation $\tilde{\mu}_i = \frac{1}{N} \sum_{j=1}^N \delta_{x_{ij}}$ the sampling measure within the host i , we have:

$$\nu = \frac{1}{M} \sum_{i=1}^M \delta_{\tilde{\mu}_i} \quad \text{and} \quad \vartheta^{k,l}(\nu) = \frac{1}{M} \sum_{\substack{i=1 \\ i \neq l}}^M \delta_{\tilde{\mu}_i} + \frac{1}{M} \delta_{\tilde{\mu}_k}. \quad (2.155)$$

Therefore

$$\begin{aligned} \left\langle \left(\vartheta^{k,l}(\nu) \right)^{\otimes \downarrow m}, \psi_{\downarrow} \right\rangle &= \frac{1}{M!} \sum_{p=1}^m \sum_{\substack{\sigma \in \mathcal{S}_M \\ \sigma(p)=l}} \psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(p-1)}, \tilde{\mu}_k, \tilde{\mu}_{\sigma(p+1)}, \dots, \tilde{\mu}_{\sigma(m)}) \\ &\quad + \frac{1}{M!} \sum_{p=m+1}^M \sum_{\substack{\sigma \in \mathcal{S}_M \\ \sigma(p)=l}} \psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(m)}). \end{aligned} \quad (2.156)$$

We can now calculate the sum of the differences for all $1 \leq l \leq M$

$$\begin{aligned} &\sum_{l=1}^M \left(\left\langle \left(\vartheta^{k,l}(\nu) \right)^{\otimes \downarrow m}, \psi_{\downarrow} \right\rangle - \left\langle \nu^{\otimes \downarrow m}, \psi_{\downarrow} \right\rangle \right) \\ &= \sum_{l=1}^M \frac{1}{M!} \sum_{p=1}^m \sum_{\substack{\sigma \in \mathcal{S}_M \\ \sigma(p)=l}} \left(\psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(p-1)}, \tilde{\mu}_k, \tilde{\mu}_{\sigma(p+1)}, \dots, \tilde{\mu}_{\sigma(m)}) - \psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(m)}) \right) \\ &= \sum_{p=1}^m \frac{1}{M!} \sum_{\sigma \in \mathcal{S}_M} \left(\psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(p-1)}, \tilde{\mu}_k, \tilde{\mu}_{\sigma(p+1)}, \dots, \tilde{\mu}_{\sigma(m)}) - \psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(m)}) \right). \end{aligned} \quad (2.157)$$

Summing over all $1 \leq k \leq M$, we have

$$\begin{aligned} &\sum_{1 \leq k, l \leq M} \left(\left\langle \left(\vartheta^{k,l}(\nu) \right)^{\otimes \downarrow m}, \psi_{\downarrow} \right\rangle - \left\langle \nu^{\otimes \downarrow m}, \psi_{\downarrow} \right\rangle \right) \\ &= \sum_{k=1}^M \sum_{p=1}^m \frac{1}{M!} \sum_{r=1}^M \sum_{\substack{\sigma \in \mathcal{S}_M \\ \sigma(r)=k}} \left(\psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(p-1)}, \tilde{\mu}_{\sigma(r)}, \tilde{\mu}_{\sigma(p+1)}, \dots, \tilde{\mu}_{\sigma(m)}) - \psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(m)}) \right) \\ &= \sum_{p=1}^m \sum_{r=1}^M \frac{1}{M!} \sum_{\sigma \in \mathcal{S}_M} \left(\psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(p-1)}, \tilde{\mu}_{\sigma(r)}, \tilde{\mu}_{\sigma(p+1)}, \dots, \tilde{\mu}_{\sigma(m)}) - \psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(m)}) \right). \end{aligned} \quad (2.158)$$

Then denoting by $\tau_{r,p}$ the transposition (r, p) , we have, for $m+1 \leq r \leq M$,

$$\begin{aligned} &\sum_{\sigma \in \mathcal{S}_M} \psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(p-1)}, \tilde{\mu}_{\sigma(r)}, \tilde{\mu}_{\sigma(p+1)}, \dots, \tilde{\mu}_{\sigma(m)}) \\ &= \sum_{\sigma \in \mathcal{S}_M} \psi_{\downarrow}(\tilde{\mu}_{\sigma \circ \tau_{r,p}(1)}, \dots, \tilde{\mu}_{\sigma \circ \tau_{r,p}(p-1)}, \tilde{\mu}_{\sigma \circ \tau_{r,p}(p)}, \tilde{\mu}_{\sigma \circ \tau_{r,p}(p+1)}, \dots, \tilde{\mu}_{\sigma \circ \tau_{r,p}(m)}) \\ &= \sum_{\sigma \in \mathcal{S}_M} \psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(m)}). \end{aligned} \quad (2.159)$$

All in all, we get

$$\Omega_{M,N}^{\uparrow,H} \Phi_{M,N}^{\phi,m,\underline{n}}(\chi) = \frac{\gamma_H}{2} \sum_{p=1}^m \sum_{r=1}^m \frac{1}{M!} \sum_{\sigma \in \mathcal{S}_M} \left(\psi_{\downarrow} \circ \theta^{r,p}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(m)}) - \psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(m)}) \right), \quad (2.160)$$

and (2.53) holds. For the resampling of parasites, we can use the beginning of the previous derivation to get an expression similar to (2.157):

$$\begin{aligned} & \sum_{i=1}^M \sum_{1 \leq k, l \leq N} \left(\left\langle (\vartheta_{(i,k),(i,l)}(\nu))^{\otimes \downarrow m}, \psi_{\downarrow} \right\rangle - \left\langle \nu^{\otimes \downarrow m}, \psi_{\downarrow} \right\rangle \right) \\ &= \sum_{i=1}^M \sum_{1 \leq k, l \leq N} \frac{1}{M!} \sum_{p=1}^m \sum_{\substack{\sigma \in \mathcal{S}_M \\ \sigma(p)=i}} \left(\psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(p-1)}, \tilde{\mu}_{i,k,l}, \tilde{\mu}_{\sigma(p+1)}, \dots, \tilde{\mu}_{\sigma(m)}) - \psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(m)}) \right) \\ &= \sum_{p=1}^m \sum_{1 \leq k, l \leq N} \frac{1}{M!} \sum_{\sigma \in \mathcal{S}_M} \left(\psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(p-1)}, \tilde{\mu}_{\sigma(p),k,l}, \tilde{\mu}_{\sigma(p+1)}, \dots, \tilde{\mu}_{\sigma(m)}) - \psi_{\downarrow}(\tilde{\mu}_{\sigma(1)}, \dots, \tilde{\mu}_{\sigma(m)}) \right), \end{aligned} \quad (2.161)$$

where $\tilde{\mu}_{i,k,l} := \tilde{\mu}_i - \frac{1}{N} \delta_{x_{il}} + \frac{1}{N} \delta_{x_{ik}}$. Then, by definition of ψ_{\downarrow} ,

$$\begin{aligned} \Omega_{M,N}^{\uparrow,P} \Phi_{M,N}^{\phi,m,\underline{n}}(\chi) &= \frac{\gamma_P}{2} \sum_{p=1}^m \int \nu^{\otimes \downarrow m}(\underline{d}\mu) \int \mu_1^{\otimes \downarrow n_1}(\underline{d}\underline{u}_1) \otimes \dots \otimes \mu_{p-1}^{\otimes \downarrow n_{p-1}}(\underline{d}\underline{u}_{p-1}) \otimes \mu_{p+1}^{\otimes \downarrow n_{p+1}}(\underline{d}\underline{u}_{p+1}) \\ &\quad \otimes \dots \otimes \mu_m^{\otimes \downarrow n_m}(\underline{d}\underline{u}_m) \sum_{1 \leq k, l \leq N} \left(\int (\mu_{p,k,l})^{\otimes \downarrow n_p}(\underline{d}\underline{u}_p) \phi \circ R(\underline{u}) - \int \mu_p^{\otimes \downarrow n_p}(\underline{d}\underline{u}_p) \phi \circ R(\underline{u}) \right). \end{aligned} \quad (2.162)$$

For the sum over $1 \leq k, l \leq N$, we can use similar arguments to the one we used with the generator for the resampling of hosts (with $\mu_{p,k,l}$ instead of $\vartheta^{k,l}(\nu)$, μ_p instead of ν , $\phi \circ R$ instead of ψ_{\downarrow} ...) so that

$$\begin{aligned} & \sum_{1 \leq k, l \leq N} \left(\int (\mu_{p,k,l})^{\otimes \downarrow n_p}(\underline{d}\underline{u}_p) \phi \circ R(\underline{u}) - \int \mu_p^{\otimes \downarrow n_p}(\underline{d}\underline{u}_p) \phi \circ R(\underline{u}) \right) \\ &= \sum_{1 \leq k, l \leq n_p} \left(\int \mu_p^{\otimes \downarrow n_p}(\underline{d}\underline{u}_p) \phi \circ R \circ \theta^{(p,k),(p,l)}(\underline{u}) - \int \mu_p^{\otimes \downarrow n_p}(\underline{d}\underline{u}_p) \phi \circ R(\underline{u}) \right), \end{aligned} \quad (2.163)$$

where the map $\theta^{(p,k),(p,l)}: T^{\downarrow \underline{n}} \rightarrow T^{\downarrow \underline{n}}$ is defined by

$$(\theta^{(p,k),(p,l)}(\underline{u}))_{ij} = \begin{cases} u_{pk} & \text{if } (i, j) = (p, l) \\ u_{ij} & \text{otherwise,} \end{cases} \quad (2.164)$$

and (2.52) holds. \square

To show that the martingale problem for the diffusion limit is well-posed, we use the tightness of the particle approximation dynamics and the uniform convergence of the generators to prove the existence of a solution, and the uniqueness follows from the duality to the nested Kingman coalescent.

Proof of Theorem 2.14. Let P_0 be a probability measure on $\mathbb{U}^{(2)}$. For each $M, N \in \mathbb{N}$, let $\mathcal{U}^{M,N} := (\mathcal{U}_t^{M,N})_{t \geq 0}$ be the two-level tree-valued Moran dynamics of size (M, N) with initial distribution $P_0^{M,N}$. Assume that $(\mathcal{U}_0^{M,N})_{M,N}$ converges weakly with respect to the two-level Gromov-weak topology, as $N \rightarrow \infty$, to \mathcal{U}_0 distributed according to P_0 .

Uniqueness. Recall the duality result from Proposition 2.18 and that the set Π^1 of polynomials separates points in $\mathbb{M}^{(2)}$. Therefore, we can apply Proposition 4.4.7 in [EK86] to prove the uniqueness of the $(\Omega^{\uparrow}, \Pi^1, P_0)$ -martingale problem.

Existence. For each $M, N \in \mathbb{N}$, $\mathcal{U}^{M,N}$ solves the $(\Omega_{M,N}^{\uparrow}, \Pi_{M,N}^1, P_0^{M,N})$ -martingale problem. Furthermore, the family $\{\mathcal{U}^{M,N} : M, N \in \mathbb{N}\}$ is tight in $\mathcal{D}_{\mathbb{U}^{(2)}}[0, \infty)$ by Theorem 2.11, and the convergence

results of Proposition 2.16 imply by Lemma 4.5.1 in [EK86] that any limit point of $\{\mathcal{U}^{M,N} : M, N \in \mathbb{N}\}$ is solution of the $(\Omega^\uparrow, \Pi^1, P_0)$ -martingale problem. This also implies convergence of $\{\mathcal{U}^{M,N} : M, N \in \mathbb{N}\}$ because of the uniqueness shown above. \square

Proof of Proposition 2.15. (i) Recall the construction of the two-level tree-valued Moran dynamics $\mathcal{U}^{M,N}$ from Section 2.3.1. For all $t \geq 0$, $\mathcal{U}_t^{M,N} = (\mathcal{I}^{M,N}, r_t^{M,N}, \nu_t^{M,N})$ with $\mathcal{I}^{M,N} = \{(i, j) : 1 \leq i \leq M, 1 \leq j \leq N\}$. In order to apply [EK86, Theorem 3.10.2], we bound the jump sizes of $\mathcal{U}^{M,N}$. Denote by $\eta_H := \{\eta_H^{k,l} : 1 \leq k, l \leq M\}$ and $\eta_P := \{\eta_P^{(i,k),(i,l)} : 1 \leq i \leq M, 1 \leq k, l \leq N\}$ the Poisson processes giving the resampling events.

Let $t > 0$. We control the distance $d_{2GP}(\mathcal{U}_{t-}^{M,N}, \mathcal{U}_t^{M,N})$, where d_{2GP} is the two-level Gromov-Prokhorov metric (see [Mei19, Definition 4.1]) which was shown to induce the two-level Gromov-weak topology on $\mathbb{M}^{(2)}$ (see [Mei19, Theorem 8.1]). It is defined by

$$d_{2GP}(\mathcal{X}, \mathcal{Y}) := \inf_{Z, \phi_X, \phi_Y} d_P^{\mathcal{M}_1(Z)}(\phi_X ** \nu_X, \phi_Y ** \nu_Y) \quad (2.165)$$

where the infimum ranges over all isometric embeddings $\phi_X : X \rightarrow Z$, $\phi_Y : Y \rightarrow Z$ into a common Polish metric space (Z, r_Z) and where $d_P^{\mathcal{M}_1(Z)}$ denotes the Prokhorov metric for measures on the space $\mathcal{M}_1(Z)$, which can be written, by [EK86, Theorem 3.1.2],

$$d_P^{\mathcal{M}_1(Z)}(\nu, \nu') = \inf_{\beta} \inf \left\{ \epsilon > 0 : \beta\{(\mu, \mu') \in (\mathcal{M}_1(Z))^2 : d_P^Z(\mu, \mu') \geq \epsilon\} \leq \epsilon \right\}, \quad (2.166)$$

where the infimum is taken over all couplings β of ν_1 and ν_2 .

- Resampling of hosts: suppose $t = \eta_H^{k,l}$ for some $1 \leq k, l \leq M$. Recall that we can express the jump by keeping the distance constant and changing the two-level measure from $\nu_{t-}^{M,N} = \frac{1}{M} \sum_{i=1}^M \delta_{\tilde{\mu}_i}$ to

$$\nu_t^{M,N} = \vartheta^{k,l}(\nu_{t-}^{M,N}) = \frac{1}{M} \sum_{i=1}^M \delta_{\tilde{\mu}_i} - \frac{1}{M} \delta_{\tilde{\mu}_l} + \frac{1}{M} \delta_{\tilde{\mu}_k}, \quad (2.167)$$

where $\tilde{\mu}_i = \frac{1}{N} \sum_{j=1}^N \delta_{(i,j)}$. Fix $Z := \mathcal{I}^{M,N}$, $r_Z := r_{t-}^{M,N}$ and $\phi_X = \phi_Y := \text{id}_Z$. Define now β a coupling of $\nu_{t-}^{M,N}$ and $\nu_t^{M,N}$ by

$$\beta := \frac{1}{M} \sum_{i \neq l} \delta_{\tilde{\mu}_i \otimes \tilde{\mu}_i} + \frac{1}{M} \delta_{\tilde{\mu}_l \otimes \tilde{\mu}_k}. \quad (2.168)$$

Then, $\beta\{(\mu, \mu') \in (\mathcal{M}_1(Z))^2 : d_P^Z(\mu, \mu') \neq 0\} \leq \frac{1}{M}$ so that

$$d_{2GP}(\mathcal{U}_{t-}^{M,N}, \mathcal{U}_t^{M,N}) \leq d_P^{\mathcal{M}_1(Z)}(\nu_{t-}^{M,N}, \nu_t^{M,N}) \leq \frac{1}{M}. \quad (2.169)$$

- Resampling of parasites: suppose $t = \eta_H^{(i,k),(i,l)}$ for some $1 \leq i \leq M$ and $1 \leq k, l \leq N$. Then it is easy to see by the same argument that the bound (2.169) holds for $d_{2GP}(\mathcal{U}_{t-}^{M,N}, \mathcal{U}_t^{M,N})$.
- If there is no resampling event at time t , then $\mathcal{U}_{t-}^{M,N} = \mathcal{U}_t^{M,N}$.

Therefore, almost surely,

$$\int_0^\infty e^{-T} \left(\sup_{0 \leq t \leq T} d_{2GP}(\mathcal{U}_{t-}^{M,N}, \mathcal{U}_t^{M,N}) \right) dT \leq \frac{1}{M}, \quad (2.170)$$

and the result follows by [EK86, Theorem 3.10.2].

(ii) The fact that the potential limit points take values in the space of compact ultra-metric m2m spaces directly follows from Propositions 2.12 and 2.27. \square

Algebraic two-level measure trees

3.1 Introduction

Motivated by their applications to biology and computer science, trees have received significant interest in the mathematical literature of the last decades. In probability theory in particular, many random tree structures have been introduced to model genealogical (or phylogenetic) trees and their evolution over time. The simplest of these models are defined on state spaces of trees with a finite number of vertices. However, the size of the space of all trees with a given number of vertices grows exponentially, and it becomes hard to study qualitative statistics of the structures. To overcome this issue, it seems reasonable to consider continuum limits of tree models and study their properties. We are particularly interested in a setup which unifies discrete and continuum trees.

A common approach by now is to encode trees as metric spaces. For trees defined in the sense of graph theory, this is done by equipping them with the graph distance. The distance has then to be properly re-scaled as to obtain some limit metric spaces, like the \mathbb{R} -trees introduced in [Tit77], whose topological properties have since been studied in details (see for example [MO90, MNO92]). One way of defining \mathbb{R} -trees is as follows. A metric space (T, r) is called *\mathbb{R} -tree* if it is connected and satisfies the so-called *four-point condition*: for all $x_1, x_2, x_3, x_4 \in T$,

$$r(x_1, x_2) + r(x_3, x_4) \leq \max \{r(x_1, x_3) + r(x_2, x_4), r(x_1, x_4) + r(x_2, x_3)\}. \quad (3.1)$$

It ensures that T has a tree structure as there should only be one possible shape for the subtree spanned by four points (see Figure 3.1). However, the assumption of connectedness does not allow for discrete trees. In [ALW17], this condition was relaxed with the notion of *metric tree*, which are metric spaces (T, r) satisfying the four-point condition (3.1) and admitting branch points, i.e., for all x_1, x_2, x_3 , there exists a (necessarily unique) $c_{(T,r)}(x_1, x_2, x_3) \in T$ such that

$$r(x_i, c_{(T,r)}(x_1, x_2, x_3)) + r(c_{(T,r)}(x_1, x_2, x_3), x_j) = r(x_i, x_j) \quad \forall i, j \in \{1, 2, 3\}, i \neq j. \quad (3.2)$$

In particular, \mathbb{R} -trees admit branch points and each metric tree can be embedded isometrically into an \mathbb{R} -tree.

More recently, a new notion of potentially continuum trees was introduced in [LW21] with the algebraic trees. The focus is now shifted from the metric to the tree structure given by the so-called *branch point map* which assigns to each triple of points their branch point. The algebraic trees are defined axiomatically by several conditions on the branch point map:

Definition 3.1 (Algebraic tree). An *algebraic tree* is a non-empty set T together with a symmetric map $c: T^3 \rightarrow T$ satisfying the following:

- (2pc) For all $x_1, x_2 \in T$, $c(x_1, x_2, x_2) = x_2$.
- (3pc) For all $x_1, x_2, x_3 \in T$, $c(x_1, x_2, c(x_1, x_2, x_3)) = c(x_1, x_2, x_3)$.
- (4pc) For all $x_1, x_2, x_3, x_4 \in T$,

$$c(x_1, x_2, x_3) \in \{c(x_1, x_2, x_4), c(x_1, x_3, x_4), c(x_2, x_3, x_4)\}. \quad (3.3)$$

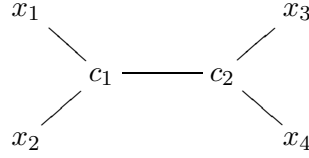


Figure 3.1: The only possible tree shape spanned by four points separates them into two pairs. Here, $c_1 = c(x_1, x_2, x_3) = c(x_1, x_2, x_4)$ and $c_2 = c(x_1, x_3, x_4) = c(x_2, x_3, x_4)$.

In order to sample leaves from an algebraic tree, we equip it with a measure. An *algebraic measure tree* (T, c, μ) consists of a separable algebraic tree (T, c) together with a probability measure μ on the Borel σ -algebra $\mathcal{B}(T, c)$. Associating each algebraic measure tree to the metric measure space given by the distance arising from the distribution of branch points, we can use the Gromov-weak topology (introduced in [GPW09]) to define a metrizable topology on the set \mathbb{T} of (equivalence classes of) algebraic measure trees. The main result of [LW21] states that this topology is compact on the subspace

$$\mathbb{T}_2 := \{(T, c, \mu) \in \mathbb{T} : \text{degrees at most 3, atoms of } \mu \text{ only at leaves}\} \quad (3.4)$$

of binary algebraic measure trees with no atoms on the skeleton. Furthermore, this topology is equivalent on \mathbb{T}_2 to the *sample shape convergence* which is based on the weak convergence of all random tree shapes spanned by finite samples.

This state space has served to construct and study the Aldous diffusion in [LMW20], and more generally the α -Ford diffusion in [NW], which are Markov processes on binary trees without edge lengths. The compactness of \mathbb{T}_2 allows to get around tightness issues in these construction. Moreover, the sample shape convergence gives rise to a family of convergence determining classes of functions which are very useful when one wants to study tree-valued stochastic processes.

More recently, new models of genealogical trees with two-level dynamics have been investigated. They are motivated by the study of two-level systems in biology, such as host-parasite or cell-virus systems, where individuals of the first level are grouped together in clusters to form the second level and both levels are subject to resampling mechanisms. For that purpose, the space of metric two-level measure spaces was introduced in [Mei19], where metric spaces are equipped with a measure on the set of measures. The idea of representing a population with hierarchical structure by such a two-level measure is not new (see e.g. [DHW90, Wu91, GHW95, GH00, DGW04, Daw18]). However, the space of metric two-level measure spaces allows for two-level tree-valued processes describing the evolution of the ancestral relationships, by encodes the genealogical distances between individuals with the metric. We applied this theory to define the two-level tree-valued Fleming-Viot dynamics in [Nus].

Our goal here is to adapt the approach of [Mei19] for two-level trees in the metric setup to the algebraic one. An *algebraic two-level measure tree* (T, c, ν) is thus defined as a separable algebraic tree (T, c) together with a two-level measure $\nu \in \mathcal{M}_1(\mathcal{M}_1(T, c))$, i.e. a Borel probability measure on the set of Borel probability measures on (T, c) . In particular, we are interested in extended the results in [LW21] to the space $\mathbb{T}_2^{(2)}$ of (equivalence classes of) algebraic two-level measure trees. For this, a crucial ingredient for us is the intensity measure M_ν of a two-level measure ν defined by

$$M_\nu(\cdot) := \int \nu(d\mu) \mu(\cdot). \quad (3.5)$$

If (T, c, ν) is an algebraic two-level measure tree, (T, c, M_ν) is an algebraic measure tree and it is then quite straightforward to adapt most of the results in [LW21]. However, this method does not apply to the main result, namely the compactness of the subspace of binary trees.

Remark 3.2 (Proving the compactness of $\mathbb{T}_2^{(2)}$). It seems that the compactness of \mathbb{T}_2 cannot be directly used to prove the compactness of $\mathbb{T}_2^{(2)}$. We explored the two following ideas which do not allow to conclude:

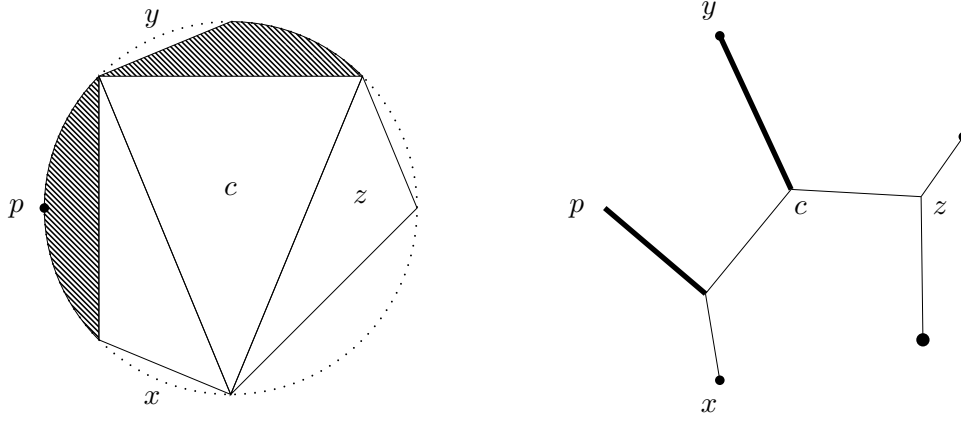


Figure 3.2: *On the left:* A sub-triangulation of the circle with three empty triangles, four empty circular segments and two filled areas. *On the right:* The tree coded by the sub-triangulation. The leaf p does not carry an atom, but the four other leaves carry a weight given by the lengths of the corresponding arcs. The two thick segments carry a non-atomic mass.

1. The compactness of \mathbb{T}_2 cannot be passed on by considering the map $(T, c, \nu) \mapsto (T, c, M_\nu)$. Though it is continuous from $\mathbb{T}_2^{(2)}$ to \mathbb{T}_2 , it is not injective. Consider for example, $T = \{x, y\}$, $c_{\{x,y\}}$ the only branch point map on T and $\nu_1 := \frac{1}{2}(\delta_{\delta_x} + \delta_{\delta_y})$ and $\nu_2 := \delta_{\frac{1}{2}(\delta_x + \delta_y)}$ which both have the same intensity measure $\frac{1}{2}(\delta_x + \delta_y)$.
2. An algebraic two-level measure tree naturally defines a probability measure on \mathbb{T}_2 and we know that $\mathcal{M}_1(\mathbb{T}_2)$ is compact since \mathbb{T}_2 is compact. However the map that sends an a2m tree to the corresponding measure on \mathbb{T}_2 is not injective either. Indeed, $(\{x, y\}, c, \delta_x)$, $(\{x, y\}, c, \delta_y)$ and $(\{x\}, c, \delta_x)$ are equivalent in \mathbb{T}_2 , so the measures on \mathbb{T}_2 associated to $(\{x, y\}, c, \nu_1)$ and $(\{x\}, c, \delta_{\delta_x})$ are equal.

A key ingredient to prove the compactness of \mathbb{T}_2 in [LW21] is the coding of binary algebraic measure trees by *sub-triangulations* of the circle. A similar encoding was first introduced by Aldous [Ald94a, Ald94b]. In [LW21], a sub-triangulation of the circle \mathbb{S} is defined as a closed, non-empty subset C of the disc satisfying the following two conditions:

1. The complement of the convex hull of C consists of open interiors of triangles.
2. C is the union of non-crossing (non-intersecting except at endpoints), possibly degenerate closed straight line segments with endpoints in \mathbb{S} .

In this coding, branch points correspond to empty triangles, leaves carrying atoms to empty circular segments, and line segments with non-atomic mass to “filled areas” (see Figure 3.2). Moreover, the arc lengths play an important role as they encode the way the mass is distributed in the algebraic measure tree. For example, a triangulation of an n -gon encodes the dual graph, equipped with the measure on the leaves given by the Lebesgue measure of the corresponding arcs of the circle line (see Figure 3.3).

We extend this coding in the two-level case. For that, one cannot simply rely on the Lebesgue measure of arc lengths anymore to encode all the information on the distribution of the random μ -mass in the tree. We rather need to replace the Lebesgue measure by a two-level measure $K \in \mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$ on the circle and we formally construct the *coding map* that associates an algebraic two-level measure tree in $\mathbb{T}_2^{(2)}$ to a pair (C, K) where C is a sub-triangulation of the circle, and K is a two-level measure on the circle line.

We showed that the coding map is continuous and surjective when the set of sub-triangulations is equipped with the Hausdorff metric topology and the set of two-level measures on the circle line with the weak topology. Using that both of these topologies are compact, our main result states that the space $\mathbb{T}_2^{(2)}$ is again compact. As an application, we finish with the construction of the Kingman algebraic two-level measure tree, which is the nested Kingman coalescent measure tree without branch length (see [Mei19]).

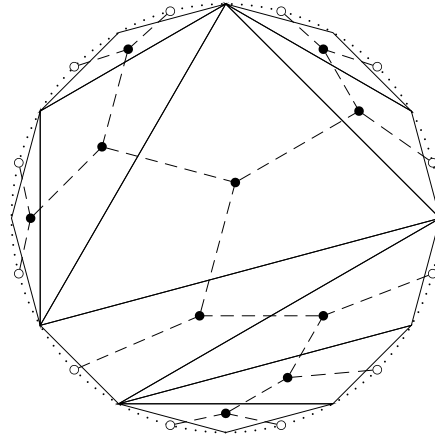


Figure 3.3: A triangulation of the 12-gon. Here, the coded tree is the dual graph, with uniform distribution on the leaves.

Outline. We recall in Section 3.2 some definitions and results on algebraic trees and on two-level measure spaces. In Section 3.3, we introduce the space of (equivalence classes of) order separable algebraic measure trees, and equip it with a separable, metrizable topology based on the two-level Gromov-weak convergence of metric repressives. Section 3.4 is devoted to the coding of binary algebraic two-level measure trees by sub-triangulations of the circle together with a two-level measure on the circle line. In Section 3.5, we introduce with the two-level sample shape another topology on the subspace of binary algebraic two-level measure trees and we use the encoding of Section 3.4 to show that both topologies are equivalent and compact. Finally, we apply this theory in Section 3.6 to define the Kingman algebraic two-level measure tree.

3.2 Preliminaries

In this section, we recall tools that will be useful in the sequel. Section 3.2.1 presents notions and results on algebraic trees from [LW21] and Section 3.2.2 on metric two-level measure spaces from [Mei19].

3.2.1 Algebraic trees

We start with the definition of algebraic trees through the branch point map, and we use it to introduce the concepts of leaves, branch points, edges, intervals, subtrees, etc. We then recall the notion of struture preserving morphisms to define a notion of equivalence on the space of a2m trees in Section 3.3. We describe the natural topology that exists on an algebraic tree. We finish with relations between algebraic trees and \mathbb{R} -trees that we will use to exploit results on metric two-level measure spaces.

Definition 3.3 (Algebraic tree). An *algebraic tree* is a non-empty set T together with a symmetric map $c: T^3 \rightarrow T$ satisfying the following:

- (2pc) For all $x_1, x_2 \in T$, $c(x_1, x_2, x_2) = x_2$.
- (3pc) For all $x_1, x_2, x_3 \in T$, $c(x_1, x_2, c(x_1, x_2, x_3)) = c(x_1, x_2, x_3)$.
- (4pc) For all $x_1, x_2, x_3, x_4 \in T$,

$$c(x_1, x_2, x_3) \in \{c(x_1, x_2, x_4), c(x_1, x_3, x_4), c(x_2, x_3, x_4)\}. \quad (3.6)$$

We call c the *branch point map*.

The properties of the map c are rich enough to define common concepts on trees. To this end, we introduce, for each point $x \in T$, an equivalence relation \sim_x on $T \setminus \{x\}$ such that for all $y, z \in T \setminus \{x\}$, $y \sim_x z$ if and only if $c(x, y, z) \neq x$. For $y \in T \setminus \{x\}$, we denote by

$$\mathcal{S}_x(y) := \{z \in T \setminus \{x\} : z \sim_x y\} \quad (3.7)$$

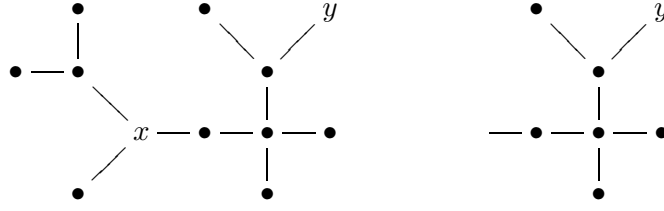


Figure 3.4: On the left: x and y are two points of the tree. On the right: the component $\mathcal{S}_x(y)$.

the equivalence class of y for this equivalence relation \sim_x . We also call $\mathcal{S}_x(y)$ the *component* of $T \setminus \{x\}$ containing y . To describe the tree structure of an algebraic tree (T, c) , we also say that:

- a *subtree* of T is a set $A \subseteq T$ such that $c(A^3) = A$,
- the *degree* of $x \in T$ is the number of components of $T \setminus \{x\}$, and we write $\deg(x) := \#\{\mathcal{S}_x(y) : y \in T \setminus \{x\}\}$,
- a *leaf* is a point $u \in T$ such that $\deg(u) = 1$, and we write $\text{lf}(T, c)$ for the set of leaves,
- a *branch point* is a point $v \in T$ such that $\deg(v) \geq 3$, or equivalently such that $v = c(x_1, x_2, x_3)$ for some $x_1, x_2, x_3 \in T \setminus \{v\}$, and we denote by $\text{br}(T)$ the set of branch points,
- for $x, y \in T$, the *interval* $[x, y]$ is

$$[x, y] := \{z \in T : c(x, y, z) = z\}, \quad (3.8)$$

- $\{x, y\}$ is an *edge* if $x \neq y$ and $[x, y] = \{x, y\}$, that is, there is “nothing between x and y ”, and we denote by

$$\text{edge}(T, c) \quad (3.9)$$

the set of edges of (T, c) .

As first examples, there is a one-to-one correspondence between (undirected) graph-theoretical trees and finite algebraic trees.

Example 3.4. Let (T, E) be a graph-theoretic tree. Define $c_E(u, v, w)$ as the unique vertex that is on the path between any two of u, v, w . Then (T, c_E) is an algebraic tree. Conversely, if (T, c) is an algebraic tree with T finite, then (T, c) corresponds to the graph-theoretic tree $(T, \text{edge}(T, c))$. Obviously, $c_{E_c} = c$.

Example 3.5 (Totally ordered spaces as algebraic trees). Let (T, \leq) be a totally ordered space. For all $x, y, z \in T$ such that $x \leq y \leq z$, we define $c_{\leq}(x, y, z) := y$. It is easy to see that c_{\leq} is symmetric and satisfies the conditions (2pc)–(4pc), so that (T, c_{\leq}) is an algebraic tree. Moreover, the interval $[x, y]$ in (T, c_{\leq}) coincides with the order interval $\{z \in T : x \leq y \leq z\}$.

Conversely, if (T, c) is an algebraic tree and $\rho \in T$ a distinguished point (often called the *root*), we can define a *partial order* \leq_ρ by letting for $x, y \in T$,

$$x \leq_\rho y \iff x \in [\rho, y]. \quad (3.10)$$

This partial order allows us to define a notion of completeness of algebraic trees.

Definition 3.6 (Directed order complete trees). Let (T, c) be an algebraic tree. We call (T, c) (*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, \leq_ρ) .

The following definition gives the analogs of complete \mathbb{R} -trees, i.e., \mathbb{R} -trees that are complete as metric spaces.

Definition 3.7 (Algebraic continuum tree). We call an algebraic tree (T, c) algebraic continuum tree if the following two conditions hold:

1. (T, c) is order complete.
2. $\text{edge}(T, c) = \emptyset$.

Even though we mainly work with unrooted trees in this paper, we shortly explain how algebraic trees are extended to rooted algebraic trees. We will use this idea to define the nested Kingman algebraic tree in Section 3.6.

Definition 3.8 (Rooted and partially ordered algebraic tree). A *partially ordered algebraic tree* is a non-empty set T together with a symmetric map $c_\wedge : T \times T \rightarrow T$ such that:

- (M1) For all $x \in T$, $c_\wedge(x, x) = x$.
 (M2) For all $x_1, x_2, x_3 \in T$, $c_\wedge(x_1, c_\wedge(x_2, x_3)) = c_\wedge(c_\wedge(x_1, x_2), x_3)$.
 (M3) For all $x_1, x_2, x_3 \in T$, $\#\{c_\wedge(x_1, x_2), c_\wedge(x_1, x_3), c_\wedge(x_2, x_3)\} \leq 2$ and if $c_\wedge(x_1, x_2) = c_\wedge(x_1, x_3)$, then
- $$c_\wedge(x_1, x_2) = c_\wedge(c_\wedge(x_1, x_2), c_\wedge(x_2, x_3)). \quad (3.11)$$

We refer to c_\wedge as the *minimum map*.

A *rooted algebraic tree* (T, c_\wedge) is a partially ordered algebraic tree for which there exists a point $\rho \in T$ with $c_\wedge(\rho, x) = \rho$ for all $x \in T$. We will refer to (this unique) ρ as the *root* of (T, c_\wedge) .

Remark 3.9 ((M1) and (M2) define a partial order). Let (T, c_\wedge) be a partially ordered algebraic tree. In what follows, we write for $x, y \in T$, $x \leq y$ if and only if $x = c_\wedge(x, y)$ respectively, $x < y$ if and only if $x = c_\wedge(x, y) \neq y$. Notice that the first two conditions (M1) and (M2) ensure that \leq defines a partial order relation. Indeed, reflexivity follows from (M1), antisymmetry follows from the fact that c_\wedge is a symmetric map and transitivity follows from (M2), i.e., if $x, y, z \in T$ are such that $x \leq y$ and $y \leq z$, or equivalently, if $x = c_\wedge(x, y)$ and $y = c_\wedge(y, z)$, then $x = c_\wedge(x, y) = c_\wedge(x, c_\wedge(y, z)) = c_\wedge(c_\wedge(x, y), z) = c_\wedge(x, z)$.

When we add the third condition (M3) to the partially ordered set defined by (M1) and (M2), we ensure that there are no loops, so that a rooted algebraic tree defines an (unrooted) algebraic tree. Reciprocally, by distinguishing a point in an algebraic tree, we can define a rooted algebraic tree (see [NW, Section 2.2] for more details).

Proposition 3.10 (Rooted versus unrooted algebraic trees). *Let $T \neq \emptyset$.*

(i) *If (T, c_\wedge) is a partially ordered, algebraic tree and $c : T^3 \rightarrow T$ the symmetric map defined as*

$$c(x, y, z) := \max\{c_\wedge(x, y), c_\wedge(x, z), c_\wedge(y, z)\}, \quad x, y, z \in T, \quad (3.12)$$

then (T, c) is an algebraic tree.

(ii) *If (T, c) is an algebraic tree, $\rho \in T$, and $c_\wedge : T^2 \rightarrow T$ the symmetric map defined as*

$$c_\wedge(x, y) := c(x, y, \rho), \quad x, y \in T, \quad (3.13)$$

then (T, c_\wedge, ρ) is a rooted algebraic tree.

We also recall a definition of structure-preserving morphisms of algebraic trees which will allow to define equivalence classes of trees in Section 3.3.

Definition 3.11 (Morphisms). Let (T, c) and $(\widehat{T}, \widehat{c})$ be algebraic trees. A map $f : T \rightarrow \widehat{T}$ is called a *tree homomorphism* (from T into \widehat{T}) if for all $x, y, z \in T$,

$$f(c(x, y, z)) = \widehat{c}(f(x), f(y), f(z)), \quad (3.14)$$

or equivalently, if for all $x, y \in T$,

$$f([x, y]) \subseteq [f(x), f(y)]. \quad (3.15)$$

We refer to a bijective tree homomorphism as *tree isomorphism*.

We can then equip an algebraic tree with a natural topology. Recall from (3.7) the components $\mathcal{S}_x(y)$ of $T \setminus \{x\}$ for $x \neq y$. The *component topology* is defined as the one generated by the set of all components $\mathcal{S}_x(y)$ in (3.7) with $x \neq y, x, y \in T$. In the following, we will suppose that algebraic trees are equipped with this topology.

Equipped with this topology, order completeness of the tree is equivalent to compactness [LW21, Proposition 2.19].

Proposition 3.12 (Compactness characterization). *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.*

In the sequel, we will often assume the topology to be separable. But for many of our results, a condition on the number of edges will be crucial, so we define the following notion of separability (see also [LW21, Example 2.23]).

Definition 3.13 (Order separability). We call an algebraic tree (T, c) *order separable* if it is separable w.r.t. the component topology and has at most countably many edges.

Order separability can also be characterized as follows [LW21, Proposition 2.20]:

Proposition 3.14 (Characterization of order separability). *Let (T, c) be an algebraic tree. Then the following are equivalent:*

1. (T, c) is order separable.
2. The component topology of (T, c) has a countable base and has at most countably many edges.
3. There exists a countable set D such that for all $x, y \in T$ with $x \neq y$,

$$D \cap [x, y] \neq \emptyset. \quad (3.16)$$

We call a set satisfying (3.16) *order dense*.

Remark 3.15 (Sufficient condition for metrizability). The component topology of any algebraic tree is Hausdorff [LW21, Lemma 2.18]. Thus, by Propositions 3.12 and 3.14, any order complete, order separable algebraic tree is a compact, second countable Hausdorff space. In particular, the component topology is metrizable and thus Polish.

There exists a connection between algebraic trees and *metric trees*, which are defined in [ALW17] as metric spaces (T, r) satisfying the 4-point condition (3.1) and admitting branch points, i.e., for all x_1, x_2, x_3 , there exists a (necessarily unique) $c_{(T, r)}(x_1, x_2, x_3) \in T$ such that

$$r(x_i, c_{(T, r)}(x_1, x_2, x_3)) + r(c_{(T, r)}(x_1, x_2, x_3), x_j) = r(x_i, x_j) \quad \forall i, j \in \{1, 2, 3\}, i \neq j. \quad (3.17)$$

Obviously, $c_{(T, r)}$ satisfies the conditions of Definition 3.3 and we call $(T, c_{(T, r)})$ the algebraic tree *induced by* (T, r) , and (T, r) a *metric representation* of $(T, c_{(T, r)})$.

Remark 3.16 (Homeomorphisms are tree homomorphisms). Let (T, r) and $(\widehat{T}, \widehat{r})$ be two \mathbb{R} -trees. Since the branch point map can be expressed in terms of intervals, a homeomorphism f between (T, c) and $(\widehat{T}, \widehat{r})$ is also a tree homomorphism between the corresponding induced algebraic trees (see [LW21, Lemma 2.35]).

Conversely, under the assumption of order separability, one can build a metric representation of an algebraic tree (T, c) as follows. Equip (T, c) with the Borel σ -algebra $\mathcal{B}(T, c)$ of the component topology. For any measure λ on $(T, \mathcal{B}(T, c))$ such that λ is finite on every interval, we consider the following pseudometric

$$r_\lambda(x, y) := \lambda([x, y]) - \frac{1}{2}\lambda(\{x\}) - \frac{1}{2}\lambda(\{y\}), \quad x, y \in T. \quad (3.18)$$

For such a measure λ , denote by T_λ the set of equivalence classes of points in T that are at distance zero for r_λ . More precisely, x and y are equivalent if λ puts weight zero on the interval $[x, y]$. Denoting again the quotient metric on T_λ by r_λ , (T_λ, r_λ) is thus the quotient metric space. Let $\pi_\lambda : T \rightarrow T_\lambda$ be the canonical projection. We have the following [LW21, Lemma 2.29].

Lemma 3.17 ($((T_\lambda, r_\lambda)$ is a metric tree). *Let (T, c) be an algebraic tree, and λ a measure on (T, c) with $\lambda([x, y]) < \infty$ for all $x, y \in T$. Then the quotient space (T_λ, r_λ) is a metric tree, and the canonical projection π_λ is a tree homomorphism.*

Furthermore, if the measure λ satisfies $\lambda[x, y] > 0$ for all $x, y \in T$, then r_λ is a metric on T . Such a measure λ always exists in the case of order separable algebraic trees (see [LW21, Lemma 2.32]).

Finally, the following result [LW21, Theorem 1] 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.

Proposition 3.18 (Characterization of order separable algebraic trees). *Let T be a non-empty set and $c: T^3 \rightarrow T$.*

1. *(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*

$$c = c_{(T, r)}. \quad (3.19)$$

2. *(T, c) is an order separable algebraic tree if and only if there is an order separable algebraic continuum tree (\bar{T}, \bar{c}) such that (T, c) is a subtree of (\bar{T}, \bar{c}) . In particular, every order separable algebraic measure tree is induced by a totally bounded metric tree.*

3.2.2 Metric two-level measure spaces

We introduce in this section the set of equivalence classes of metric two-level measure spaces (see [Mei19]). Equipped with the topology induced by so-called two-level distance polynomials, it is a Polish space. Denoting by $\mathcal{M}_1(X)$ the set of Borel probability measure on a metric space X , we start with the definition of a metric two-level measure space.

Definition 3.19 (Metric two-level measure trees). *A metric two-level measure tree (m2m space) (X, r, ν) is a non-empty Polish space (X, r) together with a two-level measure $\nu \in \mathcal{M}_1(\mathcal{M}_1(X, r))$, i.e. a Borel probability measure on the set of Borel probability measures on (X, r) .*

Equivalence classes of metric (one-level) measure spaces are defined using the notion of push-forward measures. In the case of two-level measures, we need to introduce the so-called two-level push-forward operator. Let (X, r) and (\hat{X}, \hat{r}) be Polish metric spaces and g be a Borel measurable function from X to \hat{X} . As usual, $g_*\mu$ denotes the push-forward measure $\mu \circ g^{-1}$ for a Borel probability measure $\mu \in \mathcal{M}_1(X)$. We regard g_* as an operator

$$\begin{aligned} g_* : \mathcal{M}_1(X) &\longrightarrow \mathcal{M}_1(\hat{X}) \\ \mu &\longmapsto g_*\mu. \end{aligned} \quad (3.20)$$

and call g_* the (one-level) push-forward operator of g . This enables us to define the two-level push-forward operator g_{**} of g by

$$\begin{aligned} g_{**} : \mathcal{M}_1(\mathcal{M}_1(X)) &\longrightarrow \mathcal{M}_1(\mathcal{M}_1(\hat{X})) \\ \nu &\longmapsto g_{**}\nu := \nu \circ (g_*)^{-1}. \end{aligned} \quad (3.21)$$

We need yet another notion to define equivalence classes of m2m spaces. For two metric (one-level) measure spaces to be equivalent, a measure-preserving isometry between the supports of their respective measures is enough. Thus, by considering the equivalence classe of a metric measure space (X, r, μ) , one focuses on the structure of μ and on the restriction of r to $\text{supp}(\mu)$. For an m2m space, μ is replaced by a two-level measure ν and the analog of $\text{supp}(\mu)$ is the support of the *intensity measure* $M_\nu \in \mathcal{M}_1(X)$, also called *first moment measure*, which is defined by

$$M_\nu(\cdot) = \int \nu(d\mu)\mu(\cdot). \quad (3.22)$$

This notion allows to easily adapt some results on metric measure spaces to m2m spaces by replacing the (one-level) measure μ by the intensity measure of the two-level measure ν (compare [GPW09, Proposition 7.1] to [Mei19, Theorem 7.2] in the case $\nu \in \mathcal{M}_1(\mathcal{M}_1(X))$). We will also exploit this idea in the case of algebraic trees below (e.g. in Definition 3.27).

We are now able to define a notion of equivalence for m2m spaces.

Definition 3.20 (Equivalence of m2m spaces). 1. Two m2m spaces (X_i, r_i, ν_i) , $i = 1, 2$, are called *m2m-isomorphic* if there exists a measurable function $f: X_1 \rightarrow X_2$ such that $\nu_2 = f_{**}\nu_1$ and f is isometric on the set $\text{supp}(M_{\nu_1})$ (but not necessarily on the whole space X_1). The function f is called an *m2m-isomorphism*.

2. The relation of being m2m-isomorphic is an equivalence relation on the set of m2m spaces. The set of equivalence classes of m2m trees is denoted by $\mathbb{M}^{(2)}$.

The following notion of test functions is based on the idea of sampling finite spaces of (X, r) by means of ν , i.e. we first sample measures from $\mathcal{M}_1(X)$ according to ν and then with each sampled measure, we sample finitely many points in X .

Definition 3.21 (Two-level distance polynomials). A *two-level distance polynomial* is a function $\Phi: \mathbb{M}_2^{(2)} \rightarrow \mathbb{R}$ of the form

$$\Phi(\chi) := \int \nu^{\otimes m}(\underline{d}\underline{\mu}) \int \bigotimes_{i=1}^m \mu_i(\underline{d}\underline{u}_i) \varphi\left(\left(r(u_{ij}, u_{i'j'})\right)_{(i,j),(i',j')}\right), \quad (3.23)$$

where $\chi = (X, r, \nu)$, $m \in \mathbb{N}$, $\underline{n} \in \mathbb{N}^m$ and $\varphi \in \mathcal{C}_b(\mathbb{R}^{|\underline{n}|^2})$. We write $\Pi_r^{(2)}$ for the set of all two-level distance polynomials.

The space $\mathbb{M}^{(2)}$ is then equipped with the coarsest topology such that all two-level distance polynomials are continuous.

Definition 3.22 (Two-level Gromov-weak topology). The *two-level Gromov-weak topology* is the initial topology on $\mathbb{M}^{(2)}$ induced by the test functions in $\Pi_r^{(2)}$. A sequence of m2m spaces $(\chi_n)_{n \in \mathbb{N}}$ is said to *converge two-level Gromov-weakly* to χ in $\mathbb{M}^{(2)}$ if and only if $\Phi(\chi_n)$ converges to $\Phi(\chi)$ in \mathbb{R} , for all polynomials $\Phi \in \Pi_r^{(2)}$.

The following result is given by Proposition 4.6 and Theorem 8.1 in [Mei19]:

Proposition 3.23 ($\mathbb{M}^{(2)}$ is Polish). *Equipped with the two-level Gromov-weak topology, $\mathbb{M}^{(2)}$ is a Polish space.*

3.3 The space of algebraic two-level measure trees

In this section, we define algebraic two-level measure trees (a2m trees for short) and equip the space of (equivalence classes of) a2m trees with a topology related to the two-level Gromov-weak topology for m2m spaces. We then give a result that allows to construct a2m trees from a partial knowledge of the two-level mass distribution.

Algebraic measure trees are introduced in [LW21] as order separable algebraic trees equipped with probability measures. We extend this idea by equipping algebraic trees with two-level probability measures. The order separability condition is crucial in the sequel.

Definition 3.24 (Algebraic two-level measure tree). An *algebraic two-level measure tree* (a2m tree) (T, c, ν) is an order separable algebraic tree (T, c) together with a *two-level measure* $\nu \in \mathcal{M}_1(\mathcal{M}_1(T, c))$, i.e. a Borel probability measure on the set of Borel probability measures on (T, c) .

Similarly to m2m spaces, we will consider equivalence classes of a2m trees. Recall from (3.22) the intensity measure M_ν .

Definition 3.25 (Equivalence of a2m trees). 1. Two a2m trees (T_i, c_i, ν_i) , $i = 1, 2$, are called *a2m-isomorphic* if there exist subtrees A_i of T_i with $M_{\nu_i}(A_i) = 1$ and a tree isomorphism $f: A_1 \rightarrow A_2$ such that $\nu_2 = f_{**}\nu_1$. The function f is called an *a2m-isomorphism*.

2. The relation of being a2m-isomorphic is an equivalence relation on the set of a2m trees. The set of equivalence classes of a2m trees is denoted by $\mathbb{T}^{(2)}$.

For an a2m tree (T, c, ν) , let $A := c((\text{supp}(M_\nu))^3)$. By [Mei19, Corollary 2.3], the support of ν is a subset of $\{\mu \in \mathcal{M}_1(T) \mid \text{supp}(\mu) \subseteq \text{supp}(M_\nu)\}$. Thus the subtree A of (T, c) is such that (T, c, ν) is equivalent to (A, c', ν') , where c' is the restriction of c to A^3 and ν' is the restriction of ν to $\mathcal{M}_1(A, c')$. Therefore, we define for an a2m tree $\chi := (T, c, \nu)$,

$$\text{supp}(\chi) := c((\text{supp}(M_\nu))^3), \quad (3.24)$$

and

$$\text{br}(\chi) := \text{br}(T, c) \cap \text{supp}(\chi). \quad (3.25)$$

With this in mind, note that by considering equivalence classes of a2m trees in the latter, we will only focus on the restriction of c and ν to the support of M_ν .

Example 3.26 (A2m trees without branch points nor atoms). Let $\chi := (T, c, \nu)$ be an a2m tree such that $\text{br}(\chi) = \emptyset$ and $\text{at}(M_\nu) = \emptyset$. By Proposition 3.18, there is a tree isomorphism from T into $[0, 1]$ and we may thus assume $T \subseteq [0, 1]$. Let $F_{M_\nu} : [0, 1] \rightarrow [0, 1]$ be the cumulative distribution function of the intensity measure M_ν . Define $\tilde{\nu} := (F_{M_\nu})_{**}\nu$. Then, $M_{\tilde{\nu}} = F_{M_\nu*}M_\nu$ is the Lebesgue measure on $[0, 1]$. Indeed since $\text{at}(M_\nu) = \emptyset$, the function F_{M_ν} is continuous and for $a \in [0, 1]$, the set $\{x : F_{M_\nu} \leq a\}$ is of the form $[0, x_a]$ with $F_{M_\nu}(x_a) = a$, so that $F_{M_{\tilde{\nu}}}(a) = M_\nu(\{x : F_{M_\nu} \leq a\}) = M_\nu([0, x_a]) = F_\nu(x_a) = a$.

Let $A := \{x \in \text{supp}(M_\nu) \mid \text{there is no } (y_n)_n \in ([0, 1] \setminus \text{supp}(M_\nu))^\mathbb{N} : y_n < x, y_n \rightarrow x\}$ be the support of M_ν with left boundary points removed. Then F_{M_ν} restricted to A is bijective and thus an a2m-isomorphism from (A, c, ν) onto $([0, 1], c_\leq, \tilde{\nu})$ where c_\leq is defined in Example 3.5.

We showed that if we consider the equivalence class of an a2m tree (T, c, ν) such that $\text{br}(\chi) = \text{at}(M_\nu) = \emptyset$, we can always assume that $T = [0, 1]$ and $M_\nu = \lambda_{[0, 1]}$. More generally, if there exists an interval $(v, w) \subseteq \text{supp}(M_\nu)$ such that $(v, w) \cap \text{at}(M_\nu) = \emptyset$ and $(v, w) \cap \text{br}(T, c) = \emptyset$, we can assume without loss of generality that M_ν restricted to (v, w) is $\lambda_{(v, w)}$.

We now equip the space $\mathbb{T}^{(2)}$ with a topology that relies on the two-level Gromov-weak topology on $\mathbb{M}^{(2)}$. To do so, we use that due to the order separability assumption, a2m trees allow for metric representations (see Proposition 3.18). But in order to get a useful topology on $\mathbb{T}^{(2)}$, we consider a particular metric representation of an a2m tree (T, c, ν) by using the metric r_λ defined in (3.18) with λ being the so-called *branch point distribution*.

Definition 3.27 (Branch point distribution). We call *branch point distribution* of an a2m tree (T, c, ν) the push-forward of $M_\nu^{\otimes 3}$ under the branch point map,

$$\lambda := c_* M_\nu^{\otimes 3}. \quad (3.26)$$

Recall from Lemma 3.17 that the quotient space (T_λ, r_λ) is a metric space. Therefore the following map associates a particular metric representation to each a2m tree.

Definition 3.28 (Selection map ι). Define the *selection map* $\iota : \mathbb{T}^{(2)} \rightarrow \mathbb{M}^{(2)}$ by

$$\iota(T, c, \nu) := (T_\lambda, r_\lambda, \nu_\lambda), \quad (3.27)$$

where $\lambda = c_* M_\nu^{\otimes 3}$ is the branch point distribution of (T, c, ν) , (T_λ, r_λ) is the quotient metric space, and $\nu_\lambda := \pi_{\lambda**}\nu$ is the two-level push-forward measure of ν under the canonical projection π_λ .

It is easy to see that if two a2m trees are equivalent, then their images under ι are also equivalent, with the same isomorphism (see Remark 3.16). Thus the selection map is well-defined. The following result states that ι indeed selects metric representations, and is injective.

Proposition 3.29 (ι is an embedding). *For all $\chi \in \mathbb{T}^{(2)}$, $\iota(\chi)$ is a metric representation of χ . Moreover, the selection map $\iota : \mathbb{T}^{(2)} \rightarrow \mathbb{M}^{(2)}$ is injective.*

Proof. It is enough to show that for all $\chi \in \mathbb{T}^{(2)}$, the algebraic tree induced by $\iota(\chi)$ is χ , because by Remark 3.16, if $\iota(\chi)$ and $\iota(\chi')$ are equivalent in $\mathbb{M}^{(2)}$, the measure preserving bijective homeomorphism

$f: \iota(\chi) \rightarrow \iota(\chi')$ is a tree homomorphism and thus yields an a2m-isomorphism on the corresponding a2m trees χ and χ' .

Fix $\chi = (T, c, \nu) \in \mathbb{T}^{(2)}$. We can assume w.l.o.g that for all $v \in \text{br}(T)$, $\lambda\{v\} > 0$. By Lemma 3.17, the canonical projection $\pi_\lambda: T \rightarrow T_\lambda$ is a (surjective) tree homomorphism. Therefore, to show equivalence of (T, c, ν) and $(T_\lambda, c_{(T_\lambda, r_\lambda)}, \nu_\lambda)$, it is sufficient to show that π_λ is injective on a subtree $A \subseteq T$ such that $M_\nu(A) = 1$. We take $A := T \setminus N$ with $N := \{v \in T : \pi_\lambda(v) \neq \{v\}\}$, and π_λ is injective on A .

Let us first show that $M_\nu(A) = 1$. If $\pi_\lambda(v) \neq \{v\}$ for some $v \in T$, then we can find some $u \neq v$ in T such that $r_\lambda(u, v) = 0$, i.e., $\lambda([u, v]) - \frac{1}{2}\lambda\{u\} - \frac{1}{2}\lambda\{v\} = 0$. Thus, $\lambda\{v\} = 0$ and $M_\nu\{v\} = 0$. Moreover, since π_λ is a tree isomorphism, $w \in \pi_\lambda(v)$ implies $[v, w] \subseteq \pi_\lambda(v)$. But due to order separability, there are at most countably many non-degenerate, disjoint closed intervals in T , which implies that this implies that $\pi_\lambda(N)$ is countable, and thus $M_\nu(N) = 0$.

Finally, to see that A is a subtree, let $x, y, z \in A$. If $v := c(x, y, z) \in \{x, y, z\}$, then $v \in A$. Otherwise, $v \in \text{br}(T, c)$, and hence $M_\nu\{v\} > 0$ by assumption, which implies that $\pi_\lambda(v) = \{v\}$, i.e. $v \in A$. \square

Since the selection map is an embedding, it is suitable to define a useful topology on $\mathbb{T}^{(2)}$.

Definition 3.30 (Two-level bpdd-Gromov-weak topology). Let $\mathbb{M}^{(2)}$ be equipped with the two-level Gromov-weak topology. We call the topology induced on $\mathbb{T}^{(2)}$ by the selection map ι *two-level branch-point distribution distance Gromov-weak topology* (*two-level bpdd-Gromov-weak topology*).

An immediate consequence of Proposition 2.6 is the following:

Corollary 3.31 (Separability and metrizability). $\mathbb{T}^{(2)}$ equipped with the two-level bpdd-Gromov-weak topology is a separable, metrizable space.

Proof. The two-level Gromov-weak topology on $\mathbb{M}^{(2)}$ is separable and metrizable by the two-level Gromov-Prokhorov metric $d_{2\text{GP}}$ (see [Mei19, Proposition 4.6, Theorem 8.1]). Now define for $\chi_1, \chi_2 \in \mathbb{T}^{(2)}$,

$$d_{2\text{BGP}}(\chi_1, \chi_2) := d_{2\text{GP}}(\iota(\chi_1), \iota(\chi_2)). \quad (3.28)$$

Since ι is injective, $d_{2\text{BGP}}$ is a metric on $\mathbb{T}^{(2)}$ and it induces the two-level bpdd-Gromov-weak topology. \square

We finish this section with an extension result from [LW21] for algebraic measure trees. This result will be a key ingredient in connecting a2m trees to sub-triangulations of the circle in Section 3.4.

For $x \in T$, we first extend the notation for components to

$$\mathcal{S}_x(x) := \{x\}, \quad (3.29)$$

so that T is the disjoint union of the $\deg(x) + 1$ sets in

$$\mathcal{C}_x := \{\mathcal{S}_x(y) : y \in T\}. \quad (3.30)$$

For $y \in T$ and $V \subseteq T$, we call a function $f: V \rightarrow \mathbb{R}$ *order-left continuous* on V with respect to \leq_y if for all $x, x_n \in V$ such that $x_n \uparrow_y x$, we have $\lim_{n \rightarrow \infty} f(x_n) = f(x)$, where $x_n \uparrow_y x$ means that $x_1 \leq_y x_2 \leq_y \dots$ and $x = \sup_{n \in \mathbb{N}} x_n$ with respect to \leq_y .

The following result [LW21, Proposition 3.12] allows to extend a set of functions on subtree components giving masses to a measure on the algebraic tree.

Proposition 3.32 (Extension to a measure). *Let (T, c) be an order separable algebraic continuum tree, and $V \subseteq T$ order dense. Then a set-function $\mu_0: \mathcal{C}_V := \bigcup_{x \in V} \mathcal{C}_x \rightarrow [0, 1]$ 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 $x \neq y$,

$$\mu_0(\mathcal{S}_x(y)) + \mu_0(\mathcal{S}_y(x)) \geq 1. \quad (3.31)$$

3. For every $y \in V$, the function $\psi_y: x \mapsto \mu_0(\mathcal{S}_x(y))$ is order left-continuous on V with respect to \leq_y .

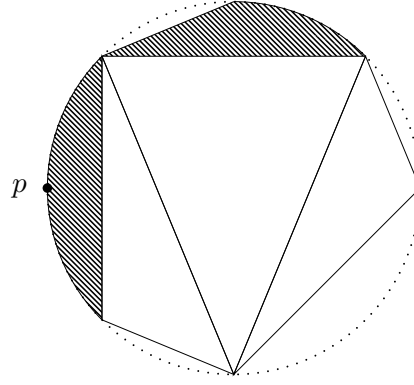


Figure 3.5: A sub-triangulation of the circle such that $\#\Delta(C) = 3$, $\#\nabla(C) = 4$, $\#\blacksquare(C) = 2$ and $\#\square(C) = 1$, (more precisely, $\square(C) = \{p\}$).

3.4 Triangulations of the circle

In [LW21, Section 4], binary algebraic measure trees were encoded by sub-triangulations of the circle, where triangles correspond to branch points in the algebraic tree and the lengths of arcs encode the mass of the corresponding subtree components. We are interested in formulating a similar result in the case of binary a2m trees. In order to encode the random masses of the subtree components, we add a two-level measure on the circle to the sub-triangulation.

We first define what we mean by sub-triangulations of the circle and in Subsection 3.4.2, we construct the map that associates an a2m tree to every sub-triangulation together with a two-level measure on the circle line and show that, as in the one-level case, this coding map is continuous and surjective.

3.4.1 The space of sub-triangulations of the circle

In the whole Section 3.4, we fix \mathbb{D} a closed disc of circumference 1, and denote by $\mathbb{S} := \partial\mathbb{D}$ the circle. We will repeatedly identify \mathbb{S} with $[0, 1]$ where the endpoints are glued. In this section, we denote by λ_I the Lebesgue measure on an interval $I \subseteq \mathbb{S} \simeq [0, 1]$. For a subset $A \subseteq \mathbb{D}$, we denote by \overline{A} , A° , ∂A and $\text{conv}(A)$ the closure, the interior, the boundary and the convex hull of A , respectively. Furthermore, we define

$$\Delta(A) := \{\text{connected components of } \text{conv}(A) \setminus A\}, \quad (3.32)$$

and

$$\nabla(A) := \{\text{connected components of } \mathbb{D} \setminus \text{conv}(A)\}, \quad (3.33)$$

so that we have the partition of the disc

$$\mathbb{D} = A \uplus \bigsqcup_{a \in \Delta(A)} a \uplus \bigsqcup_{b \in \nabla(A)} b. \quad (3.34)$$

Definition 3.33 (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} .

Note that given (Tri1), (Tri2) implies that $\nabla(C)$ consists of circular segments with the bounding straight line excluded and the rest of the bounding arc included (see Figure 3.5).

The set \mathcal{T} can be equipped with a compact metrizable topology in the following way. Let

$$\mathcal{F}(\mathbb{D}) := \{F \subseteq \mathbb{D} : F \neq \emptyset\}. \quad (3.35)$$

We equip $\mathcal{F}(\mathbb{D})$ with the *Hausdorff metric topology*. That is, we say that a sequence $(F_n)_{n \in \mathbb{N}}$ converges to F in $\mathcal{F}(\mathbb{D})$ if and only if for all $\epsilon > 0$ and all large enough $n \in \mathbb{N}$,

$$F \subseteq F_n^\epsilon \quad \text{and} \quad F_n \subseteq F^\epsilon, \quad (3.36)$$

where for all $A \in \mathcal{F}(\mathbb{D})$, $A^\epsilon := \{x \in \mathbb{D} : d(x, A) < \epsilon\}$. As \mathbb{D} is compact, $\mathcal{F}(\mathbb{D})$ is a compact metrizable space which contains \mathcal{T} and it can be shown that \mathcal{T} is actually a closed subset of $\mathcal{F}(\mathbb{D})$ (see [LW21, Lemma 4.2]).

In order to construct the branch point map from a sub-triangulation in the next subsection, we need another characterization of sub-triangulations of the circle. Roughly speaking, condition (Tri2) can be replaced by the existence of triangles that separate triples of connected components of $\mathbb{D} \setminus C$.

Proposition 3.34 (Characterization of sub-triangulations). *Let C be a closed non-empty subset of \mathbb{D} . Then C is a sub-triangulation of the circle if and only if condition (Tri1) holds, all extreme points of $\text{conv}(C)$ 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 $\mathbb{D} \setminus \partial c_{xyz}$.*

3.4.2 Coding of binary algebraic two-level measure trees

We define here a map that associates a binary a2m tree to a sub-triangulation together with a two-level measure on the circle line whose intensity measure is the Lebesgue measure. We show that this coding map is surjective and continuous.

We start by defining the set of binary algebraic two-level measure trees with no atoms on the skeleton:

$$\mathbb{T}_2^{(2)} := \left\{ \chi = \overline{(T, c, \nu)} \in \mathbb{T}^{(2)} : \deg(v) \leq 3 \ \forall v \in T, \text{ at}(\mu) \subseteq \text{lf}(T, c) \ \forall \mu \in \text{supp}(\nu) \right\}. \quad (3.37)$$

Note that for all $(T, c, \nu) \in \mathbb{T}_2^{(2)}$, we also have $\text{at}(M_\nu) \subseteq \text{lf}(T, c)$.

In Theorem 3.39, we introduce a coding map which associates an a2m tree $(T, c, \nu) \in \mathbb{T}_2^{(2)}$ to each two-level sub-triangulation.

Definition 3.35 (Two-level sub-triangulation). A *two-level sub-triangulation of the circle* (C, K) consists of sub-triangulation of the circle C together with a two-level measure $K \in \mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$ on \mathbb{S} such that its intensity measure M_K is $\lambda_{\mathbb{S}}$ the Lebesgue on the circle line.

We denote by

$$\mathfrak{D} := \{(C, K) \in \mathcal{T} \times \mathcal{M}_1(\mathcal{M}_1(\mathbb{S})) : M_K = \lambda_{\mathbb{S}}\} \quad (3.38)$$

the set of all two-level sub-triangulations of the circle.

In the construction we give below, the algebraic measure tree (T, c, M_ν) is the one associated to the sub-triangulation C given by the coding map in [LW21, Theorem 2]. In particular, $x \in \Delta(C)$ will correspond to a branch point of the associated a2m tree and $x \in \nabla(C)$ to an atom of the intensity measure M_ν of the a2m tree such that the arc $x \cap \mathbb{S}$ has length the M_ν -mass of the corresponding atom.

Example 3.36 (A simple case). Consider a triangulation C of the regular n -gon into $n - 2$ triangles (see Figure 3.6). In this case, the coded tree is the dual graph. That is, a triangle corresponds to a branch point of the tree, and two branch points of the tree are connected by an edge if and only if the triangles are adjacent. Each circular segment corresponds to a leaf and M_ν will assign to each leaf the length of the circular segment, i.e. n^{-1} since the n -gon is regular.

We then add on the circle line a two-level measure K such that $M_K = \lambda_{\mathbb{S}}$ by defining

$$K = \frac{1}{2}\delta_{\kappa_1} + \frac{1}{2}\delta_{\kappa_2}, \quad (3.39)$$

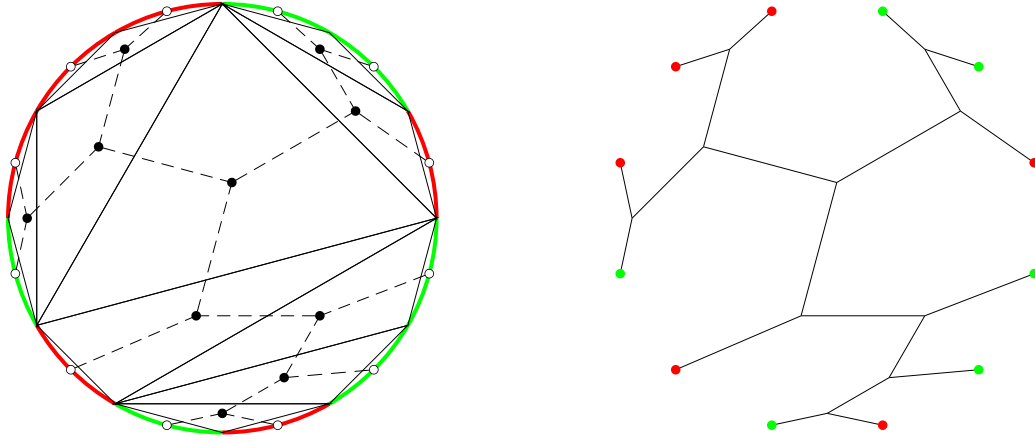


Figure 3.6: A triangulation of the 12-gon and the tree coded by it. Suppose that we add on the triangulation the two-level measure $K = \frac{1}{2}\delta_{\kappa_1} + \frac{1}{2}\delta_{\kappa_2}$, where κ_1 is the renormalized Lebesgue measure on the union of the red circular segments and κ_2 on the union of the green circular segments. Then the corresponding two-level measure on the tree is given by $\frac{1}{2}\delta_{\mu_1} + \frac{1}{2}\delta_{\mu_2}$ where μ_1 assigns mass $\frac{1}{6}$ to each red leaf and μ_2 to each green leaf.

where $\kappa_1, \kappa_2 \in \mathcal{M}_1(\mathbb{S})$ are defined in Figure 3.6. Then the associated two-level measure ν on the tree is given by $\frac{1}{2}\delta_{\mu_1} + \frac{1}{2}\delta_{\mu_2}$ where μ_i assigns to each leaf the κ_i -mass carried by the corresponding circular segment.

Remark 3.37 (About the condition $M_K = \lambda_{\mathbb{S}}$). Consider the sub-triangulation C and $K := \delta_{\delta_0}$ as in Figure 3.7 on the left. We needed a condition on K to avoid such a case where M_K puts an atom on an endpoint of a line segment of C separating two circular segments. We could have overcome this issue by deciding a priori to which of the two circular segment the atom adds mass. However, with this solution, the coding map would not be continuous when \mathcal{T} is equipped with the Hausdorff metric topology and $\mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$ with the weak topology. Indeed, suppose we decide to add the atom on endpoints to the mass of the circular segment after the endpoint and let $K_n := \delta_{\frac{1}{2}\delta_0 + \frac{1}{2}\delta_{n-1}}$ (see Figure 3.7 on the right). Then the sequence (C, K_n) converges to (C, K) but it is not true of the corresponding a2m trees since the one associated with (C, K_n) has two leaves carrying mass but not the one associated with (C, K) .

The condition $M_K = \lambda_{\mathbb{S}}$ prevents such cases from occurring and it seems natural for the specific reason that it also ensures that the algebraic measure tree (T, c, M_ν) corresponding to (C, K) is the one associated to the sub-triangulation C in the one-level case.

Before stating the result, we need further notation. We first define on the sub-triangulation the points that will correspond through the coding construction to leaves without mass. 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

$$\square(C) := \{ \{p_i(x)\} : x \in \Delta(C), i \in \{1, 2, 3\}, \text{comp}_x(p_i(x)) \subseteq C \} \quad (3.40)$$

which is the set of mid-points of “filled” circular segments for C (see Figure 3.5). In the tree, the corresponding leaves will be connected to the rest of the tree through line segments in (T, c, ν) with non-atomic measure. For (T, c, ν) an algebraic measure tree, we call an interval $(v, w) \subseteq \text{supp}(M_\nu)$ such that $(v, w) \cap \text{at}(M_\nu) = \emptyset$ and $(v, w) \cap \text{br}(T, c) = \emptyset$ a *line segment with non-atomic measure*. We denote by

$$\text{seg}(T, c, \nu) \quad (3.41)$$

the set of maximal (w.r.t. inclusion) line segments with non-atomic measure. Note that if (T, c) is order complete, every line segment with non-atomic measure is included in some $(v, w) \in \text{seg}(T, c, \nu)$. The analog of these line segments in the sub-triangulation are “filled” areas, that is, elements of

$$\blacksquare(C) := \{ \bar{b} \mid b \text{ is a connected component of } C^\circ \}. \quad (3.42)$$

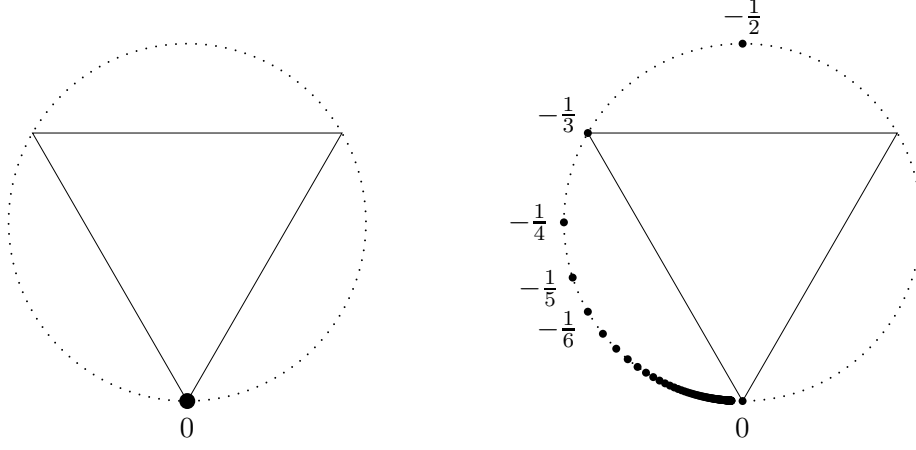


Figure 3.7: Examples of two-level measures K on the circle such that $M_K \neq \lambda_{\mathbb{S}}$. On the left, $K = \delta_{\delta_0}$ and on the right, we represent $K_n = \delta_{\frac{1}{2}\delta_0 + \frac{1}{2}\delta_{n-1}}$ for several $n \in \mathbb{N}$. Then K_n converges weakly to K . If we add the atom on endpoints to the mass of the circular segment after the endpoint, the tree associated with (C, K_n) has two leaves carrying mass but the one associated with (C, K) is a single point carrying mass. Therefore the sequence of trees corresponding to (C, K_n) do not converge to the one associated with (C, K) .

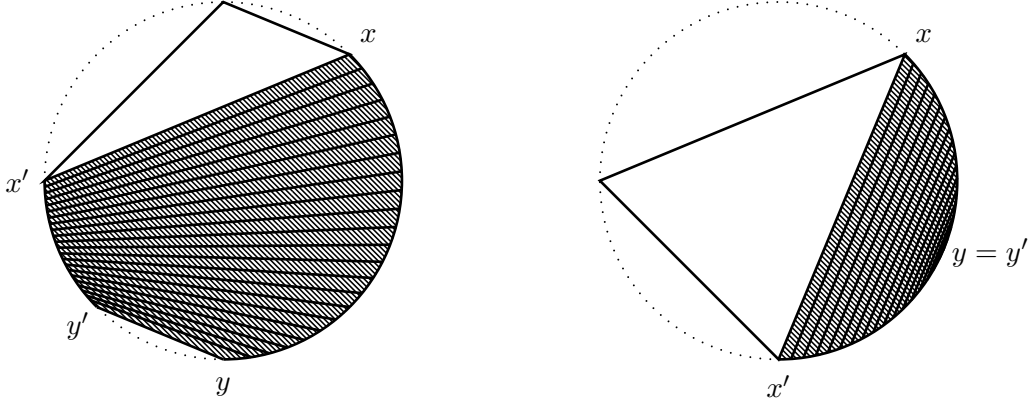


Figure 3.8: Two examples of “filled” areas $b \in \blacksquare(C)$ delimited by the circular segments $[x, y]$ and $[x', y']$. They are partitioned into straight line segments with endpoints in $[x, y]$ and $[x', y']$.

For such a filled area $b \in \blacksquare(C)$, it will be important to be able to know the mass distribution of K along $b \cap \mathbb{S}$, and not only the total mass carried by $b \cap \mathbb{S}$. For this reason, we use that by definition of a sub-triangulation, $b \in \blacksquare(C)$ is the union of non-crossing closed straight line segments with endpoints in \mathbb{S} . However, there are several possible ways to decompose b into infinitely many line segments. For the sake of simplicity, we choose to partition b “linearly” in the following sense. Recall that we identify \mathbb{S} with $[0, 1]$ where the endpoints are glued. If $b \cap \mathbb{S}$ is not connected, then it is the union of two circular segments $[x, y]$ and $[y', x']$ (with $x = y$ or $x' = y'$ possibly, see Figure 3.8). If $b \cap \mathbb{S}$ is connected, it is a circular segment $[x, x']$ and we take $y = y'$ to be its mid-point (which belongs to $\square(C)$), so that $b \cap \mathbb{S}$ is still the union of two circular segments. We then define $\|_b(C)$ to be the set of all straight line segments connecting a point of $[x, y]$ to a point of $[y', x']$ in a linear way (see Figure 3.8). For example, if $x \leq y \leq y' \leq x'$,

$$\|_b(C) := \{[(1-t)x + ty, (1-t)x' + ty'] \mid t \in [0, 1]\}. \quad (3.43)$$

Define

$$\|(C) := \bigcup_{b \in \blacksquare(C)} \|_b(C). \quad (3.44)$$

Finally, we define for the sub-triangulation the analog of the components in an algebraic tree. 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

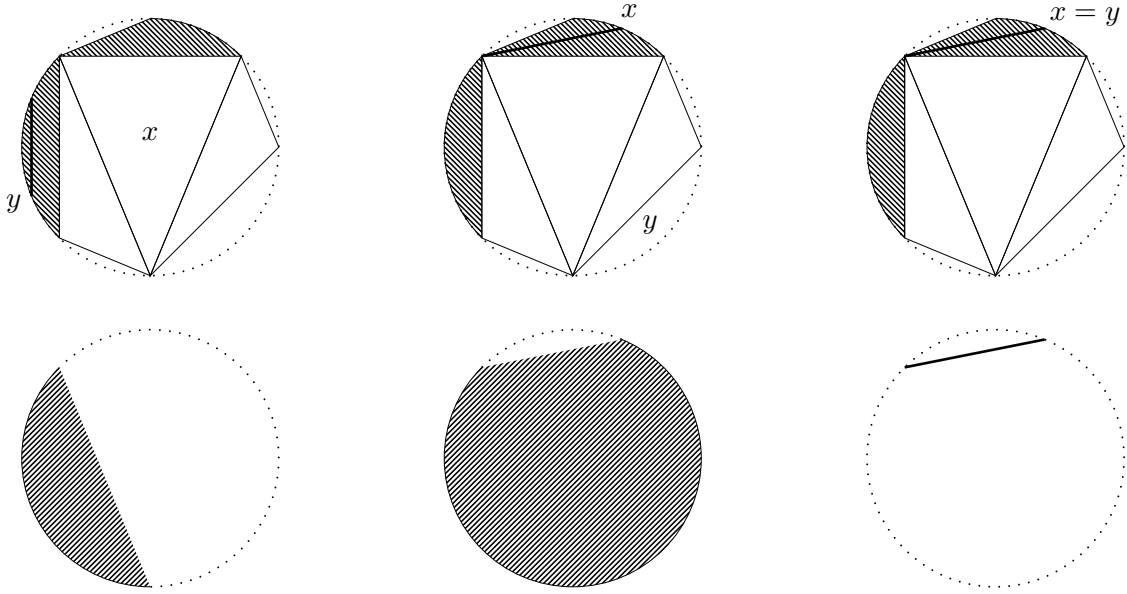


Figure 3.9: Three examples of components $\text{comp}_x(y)$ in a sub-triangulation of the circle C . On the left, $x \in \Delta(C)$ and $y \in \|(C)$. In the middle, $x \in \|(C)$ and $y \in \nabla(C)$. On the right, $x = y \in \|(C)$ and $\text{comp}_x(x) = x$. First the two other cases, $\text{comp}_x(y)$ is a circular segment with the boundary straight line excluded and the rest of the boundary arc included.

space \mathbb{D} , let

$$\text{comp}_x(y) := \text{the connected component of } \mathbb{D} \setminus \partial_{\mathbb{D}} x \text{ which contains } y. \quad (3.45)$$

Similarly, for $x \in \|(C)$ and $y \subseteq \mathbb{D}$ connected and disjoint from x , let

$$\text{comp}_x(y) := \text{the connected component of } \mathbb{D} \setminus x \text{ which contains } y. \quad (3.46)$$

We also call them *components* (of the sub-triangulation) (see Figure 3.9). Define also $\text{comp}_x(x) := x$ for $x \in \|(C) \cup \square(C)$.

The following lemma gives the tree associated to a triangulation C . Contrary to [LW21, Lemma 4.7], we add the set $\|(C)$ to the “skeleton” of the tree as it is important in the two-level case to know how the mass is distributed along a line segment of the tree.

Lemma 3.38 (induced branch-point map). *For $C \in \mathcal{T}$, let $V_C := \Delta(C) \cup \nabla(C) \cup \square(C) \cup \|(C)$. If $V_C \neq \emptyset$, then there exists a unique branch-point map $c_V: V_C^3 \rightarrow V_C$, such that (V_C, c_V) is an algebraic tree with*

$$\mathcal{S}_x^{(V_C, c_V)}(y) = \{v \in V_C : \text{comp}_x(y) = \text{comp}_x(v)\} \quad (3.47)$$

for $x, y \in V_C$. In particular, $\deg(x) = 3$ for all $x \in \Delta(C)$, $\deg(x) = 1$ for all $x \in \nabla(C) \cup \square(C)$ and $\deg(x) = 2$ for all $x \in \|(C) \setminus \square(C)$.

Proof. Condition (Tri2)’ of Proposition 3.34 gives that for pairwise distinct $x, y, z \in \Delta(C) \cup \nabla(C)$, there is a unique triangle $c_{xyz} \in \Delta(C)$ such that x, y, z are subsets of pairwise different connected components of $\mathbb{D} \setminus \partial_{c_{xyz}}$. This can be extended to triples of points in V_C in an obvious way. It is then easy to see that this defines a branch point map on V_C . \square

The following theorem is the analog of [LW21, Theorem 2]. It states that each pair (C, K) where C is a sub-triangulation of the circle and K a two-level measure on \mathbb{S} with $M_K = \lambda_{\mathbb{S}}$ can be associated with a binary a2m tree such that $\Delta(C)$ corresponds to the set of branch points and $\nabla(C)$ corresponds to the set of atoms of the intensity measure of ν in the tree. Furthermore, $\text{comp}_v(w)$ corresponds to the component $\mathcal{S}_v(w)$ and its random ν -mass is given by the random K -mass carried by $\text{comp}_v(w)$.

Theorem 3.39 (Coding map). (i) For all $\Gamma = (C, K) \in \mathfrak{D}$, there is a unique (up to equivalence) a2m tree $\chi_\Gamma = (T_\Gamma, c_\Gamma, \nu_\Gamma) \in \mathbb{T}_2^{(2)}$ such that:

(CM1) $V_C \subseteq T_\Gamma$, $\text{br}(T_\Gamma, c_\Gamma) = \Delta(C)$, and c_Γ is an extension of c_V , where (V_C, c_V) is defined in Lemma 3.38.

(CM2) For all $x, y \in V_C$ and $\sigma \geq 0$,

$$\int \nu(d\mu) e^{-\sigma \mu(\mathcal{S}_x(y))} = \int K(d\kappa) e^{-\sigma \kappa(\text{comp}_x(y) \cap \mathbb{S})}. \quad (3.48)$$

(CM3) $\text{at}(M_{\nu_\Gamma}) = \nabla(C)$.

(ii) The coding map $\tau : \mathfrak{D} \rightarrow \mathbb{T}_2^{(2)}$, $\Gamma \mapsto \chi_\Gamma$ is surjective.

(iii) Let \mathcal{T} be equipped with the Hausdorff metric topology, $\mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$ with the weak topology, $\mathcal{T} \times \mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$ with the product topology and $\mathbb{T}_2^{(2)}$ with the two-level bpdd-Gromov-weak topology. Then the coding map τ is continuous.

Let us first sketch the proof. (i) We start by extending the tree (V_C, c_V) to an algebraic continuum tree and apply Proposition 3.32 “almost surely” to construct the two-level measure ν . (ii) We give the proof of the surjectivity of the coding map τ in three steps. For $\chi = (T, c, \nu) \in \mathbb{T}_2^{(2)}$,

1. we first construct the sub-triangulation C associated with the algebraic measure tree (T, c, M_ν) as in the proof of [LW21, Theorem 2]. That is, for each branch point we remove an open triangle from the set \mathbb{D} and for each atom of M_ν , we remove a circular segment.
2. we then build K by weak approximation. For this, we rely on a result from [Mei19] that the two-level measure ν can be reconstructed from an infinite set of points randomly sampled. The two-level measure K is then built in a similar fashion by using a correspondance between points in the sample and subsets of the circle line.
3. finally, we prove that the constructed pair (C, K) is in \mathfrak{D} by partitioning the circle line into different types of intervals and showing that the restriction of M_K to each interval is the Lebesgue measure.

(iii) The proof of the continuity of τ is deferred to the next section, Proposition 3.47.

Proof of Theorem 3.39. (i) Let $\Gamma = (C, K) \in \mathfrak{D}$. For $\kappa \in \text{supp}(K)$, we define the set-function $\mu_0^\kappa : \bigcup_{x \in V_C} \mathcal{C}_x \rightarrow [0, 1]$

$$\mu_0^\kappa(\mathcal{S}_x(y)) := \kappa(\text{comp}_x(y) \cap \mathbb{S}), \quad x, y \in V_C. \quad (3.49)$$

We want to be in a position to apply Proposition 3.32 to each μ_0^κ . Since $M_K = \lambda_\mathbb{S}$, for K -almost every κ , for all $x \in V_C$, $\kappa(\partial_{\mathbb{D}} x \cap \mathbb{S}) = 0$. Thus, for K -almost every κ and all $x \in V_C$, $\sum_{A \in \mathcal{C}_x} \mu_0^\kappa(A) = \kappa(\mathbb{S}) = 1$ and $\mu_0^\kappa(\mathcal{S}_x(y)) + \mu_0^\kappa(\mathcal{S}_y(x)) \geq 1$ for $y \neq x$. Furthermore, the function $x \mapsto \mu_0^\kappa(\mathcal{S}_x(y))$ is order left-continuous on V_C w.r.t. \leq_y for all $y \in V_C$ for K -almost every κ . This can be shown by using (3.49) and the continuity from above of the measure κ , together with the condition $M_K = \lambda_\mathbb{S}$.

The tree (V_C, c_V) is order separable. Indeed, the set $\Delta(C) \cup \nabla(C)$ is countable and one can construct a countable order dense set of (V_C, c_V) by adding the straight line segments in $\|(C)$ that have at least one rational endpoint in $\mathbb{S} \simeq [0, 1]$. However, (V_C, c_V) is not necessarily order complete. Since we defined $\blacksquare(C)$ to be the set of all closures of connected components of C° , the endpoints of line segments are included in (V_C, c_V) . But we still need to add leaves which are limits of an increasing sequence (w.r.t. \leq_y for some $y \in V_C$) of branch points. Thus we define \overline{V}_C as V_C together with an uncountable set of leaves given by the limit points of these increasing sequences, and we can easily extend c_V to \overline{c}_V so that the tree $(\overline{V}_C, \overline{c}_V)$ is order complete and order separable.

Furthermore, $(\overline{V}_C, \overline{c}_V)$ need not to be an algebraic continuum tree because $\text{edge}(\overline{V}_C, \overline{c}_V)$ might be non-empty. Therefore, to apply Proposition 3.32 we also extend the tree $(\overline{V}_C, \overline{c}_V)$ to make it an order separable algebraic continuum tree. Let $\{v, w\}$ be an edge of $(\overline{V}_C, \overline{c}_V)$. Define $\tilde{V} := \overline{V}_C \uplus (\{(v, w)\} \times (0, 1))$. We can extend \overline{c}_V to a branch point map \tilde{c} on \tilde{V} in a canonical way such that

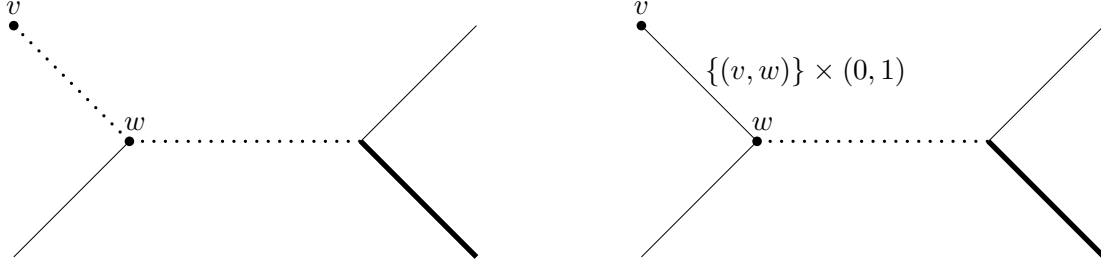


Figure 3.10: The tree is not a continuum tree because $[v, w] = \{v, w\}$ is an edge. We extend the tree in a natural way so that $[v, w]$ is no longer an edge.

$[v, w] = \{v, w\} \cup (\{(v, w)\} \times (0, 1))$ (see Figure 3.10). \bar{V}_C is not anymore order dense in (\tilde{V}, \tilde{c}) so we also define μ_0^κ on any \mathcal{C}_x with $x \in \{(v, w)\} \times (0, 1)$. Fix $\kappa \in \text{supp}(K)$. For $x \in \{(v, w)\} \times (0, 1)$, let $\mu_0^\kappa(\{x\}) = 0$ and for $y \neq x$,

$$\mu_0^\kappa(\mathcal{S}_x(y)) := \begin{cases} \mu_0^\kappa(\mathcal{S}_x(v)) & \text{if } v \in \mathcal{S}_x(y), \\ \mu_0^\kappa(\mathcal{S}_x(w)) & \text{if } w \in \mathcal{S}_x(y). \end{cases} \quad (3.50)$$

Then for K -almost every κ , μ_0^κ still satisfies the conditions (1)-(3) of Proposition 3.32 on (\tilde{V}, \tilde{c}) . We repeat this extension for each edge in $\text{edge}(\bar{V}_C, \bar{c}_V)$ which is countable as the set $\Delta(C) \cup \nabla(C)$ is itself countable. We denote by (T, c) the extended tree. By construction, it is still order complete and order separable, and $\text{edge}(T, c) = \emptyset$. Since μ_0^κ was extended to $\bigcup_{x \in T} \mathcal{C}_x$ and T is obviously dense in itself, we can now apply Proposition 3.32. That is, for K -almost every κ , μ_0^κ has a unique extension to a probability measure μ^κ on $\mathcal{B}(T, c)$.

Define ν on $\mathcal{M}_1(T, c)$ as the pushforward of K under the map $\kappa \mapsto \mu^\kappa$. Let us verify the a2m tree (T, c, ν) satisfies conditions (CM1)-(CM3). It is easy to see that (CM1) holds, and by (3.49), for all $x, y \in V_C$,

$$\int \nu(d\mu) e^{-\sigma \mu(\mathcal{S}_x(y))} = \int K(d\kappa) e^{-\sigma \mu^\kappa(\mathcal{S}_x(y))} = \int K(d\kappa) e^{-\sigma \kappa(\text{comp}_x(y) \cap \mathbb{S})}. \quad (3.51)$$

Moreover, points in $T \setminus V_C$ carry no atoms, so $\text{at}(M_\nu) \subseteq V_C$. With (CM2), for each $x \in V_C$,

$$M_{\nu_T}\{x\} = M_K(\text{comp}_x(x) \cap \mathbb{S}) = \lambda(\text{comp}_x(x) \cap \mathbb{S}), \quad (3.52)$$

which is zero if $x \in \Delta(C) \cup \square(C)$ and strictly positive if $x \in \nabla(C)$, which yields (CM3). To see that it is unique (up to equivalence), notice that in the construction above, the extensions of edges of (\bar{V}_C, \bar{c}_V) carry no mass and that the extension of μ_0^κ to μ^κ is unique for K -almost every κ by Proposition 3.32.

(ii) We give the proof of the surjectivity in three steps. Let $\chi = (T, c, \nu) \in \mathbb{T}_2^{(2)}$. We construct a sub-triangulation C and a two-level measure K on \mathbb{S} such that $\tau(C, K) = \chi$. We can suppose w.l.o.g. that (T, c) is order complete and that for all $v \in \text{br}(T, c)$, $c_*(M_\nu^{\otimes 3})\{v\} > 0$. To make the construction of K easier, we also assume that for all $(v, w) \in \text{seg}(T, c, \nu)$, M_ν restricted to (v, w) is the Lebesgue measure (see Example 3.26).

Step 1: construction of C . Fix $\rho \in \text{lf}(T, c)$ and recall that ρ induces a partial order relation \leq_ρ . We can extend this partial order to a total order \leq in the following way. For $v \in \text{br}(T, c)$, denote by $S_0(v)$, $S_1(v)$ and $S_2(v)$ the three components of $T \setminus \{v\}$ such that $S_0(v) = \mathcal{S}_v(\rho)$ and with $S_1(v)$ and $S_2(v)$ chosen by picking an order for the two remaining components. Define now for all v, w ,

$$v \leq w \iff [v \leq_\rho w] \text{ or } [v \in S_1(c(v, w, \rho)) \text{ and } w \in S_2(c(v, w, \rho))]. \quad (3.53)$$

Using this total order on the tree, the construction of C below can be understood as follows. Starting from the root ρ which corresponds to 0 in the identification $\mathbb{S} \simeq [0, 1]$, we read through T according to \leq . For each branch point v , we draw a triangle whose vertices on \mathbb{S} are given by the M_ν -mass of

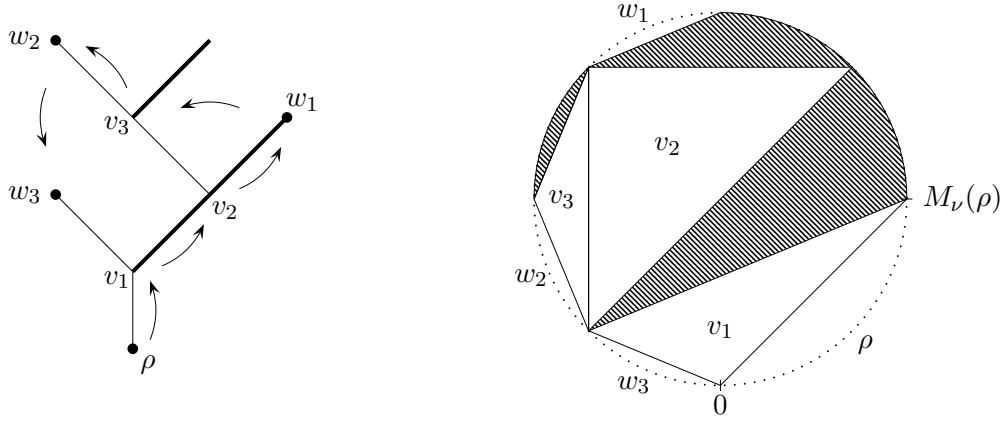


Figure 3.11: An algebraic tree (T, c) and the corresponding sub-triangulation C as constructed in Step 1 of the proof of Theorem 3.39(ii). v_1, v_2, v_3 are branch points and ρ, w_1, w_2, w_3 are leaves carrying an M_ν -atom. More precisely, $M_\nu(\rho) = \frac{1}{4}$ and $M_\nu(w_1) = M_\nu(w_2) = M_\nu(w_3) = \frac{1}{8}$. We extend the partial order \leq_ρ to a total order such that $\rho \leq v_1 \leq v_2 \leq w_1 \leq v_3 \leq w_2 \leq w_3$.

points smaller than v and the M_ν -masses of the two components above v . For each leaf carrying an atom w of M_ν , we add a straight line segment to C according to the M_ν -mass of points smaller than w and the M_ν -mass of w (see Figure 3.11).

We first introduce some more notations. 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 $a, a + b, a + b + c \in \mathbb{S}$, and let $\ell(a, b) \subseteq \mathbb{D}$ be the straight line from a to $a + b$, and $L(a, b)$ the connected component of $\mathbb{D} \setminus \ell(a, b)$ containing $a + \frac{b}{2} \in \mathbb{S}$. Now for all $v \in \text{br}(T, c) \cup \text{at}(M_\nu)$, denote by

$$\alpha(v) := M_\nu(\{u \in T : u < v\}) \quad (3.54)$$

the total M_ν -mass of points smaller than v (with respect to \leq defined in (3.53)). In the sub-triangulation C defined below, $\alpha(v)$ gives the first vertex of the triangle or circular segment corresponding to v . We define C by

$$\mathbb{D} \setminus C := \biguplus_{v \in \text{br}(T, c)} \Delta(\alpha(v), M_\nu(S_1(v)), M_\nu(S_2(v))) \uplus \biguplus_{w \in \text{at}(M_\nu)} L(\alpha(w), M_\nu\{w\}). \quad (3.55)$$

Let us show that C is a sub-triangulation of the circle. By definition of C , $\text{conv}(C) \setminus C$ is the disjoint union of open triangles, i.e. condition (Tri1) is satisfied. Furthermore, the extreme points of $\text{conv}(C)$ 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 3.34, $C \in \mathcal{T}$. Note that C is the sub-triangulation associated with the algebraic measure tree (T, c, M_ν) in the proof of Theorem 4.8 in [LW21].

To make the notations coincide with the ones from Lemma 3.38, we will use the following correspondences in the rest of the proof. That is, a branch point $v \in \text{br}(T, c)$ will correspond to the triangle $\bar{v} = \Delta(\alpha(v), M_\nu(S_1(v)), M_\nu(S_2(v)))$ in $\Delta(C)$ and an atom $w \in \text{at}(M_\nu)$ to the circular segment $\bar{w} = L(\alpha(w), M_\nu\{w\})$ in $\nabla(C)$. A leaf with zero mass $u \in \text{lf}(T, c)$, but connected to the rest of the tree by a line segment carrying mass, will correspond to the midpoint $\bar{u} \in \square(C)$ of the corresponding “filled” circular segment. Finally, we associate, in a “linear” way, a point $z \in T$ belonging to a line segment with non-atomic mass in $\text{seg}(T, c, \nu)$ to a straight line segment $\bar{z} \in \parallel(C)$. That is, if the segment $(v, w) \in \text{seg}(T, c, \nu)$ corresponds to the circular segments $]x, y[$ and $]x', y'[$, the point $y = (1 - t)v + tw$ with $t \in (0, 1)$ is associated with the straight line segment $\bar{y} := [(1 - t)x + ty, (1 - t)x' + ty']$ (see (3.43)). Denote by

$$V_T \subseteq T \quad (3.56)$$

the set of all points x in the tree that have a corresponding subset $\bar{x} \in V_C$.

With these notations, V_T is a subtree of (T, c) and by construction, we have that for all $x, y \in V_T$,

$$M_\nu(S_x(y)) = \lambda_{\mathbb{S}}(\text{comp}_{\bar{x}}(\bar{y}) \cap \mathbb{S}). \quad (3.57)$$

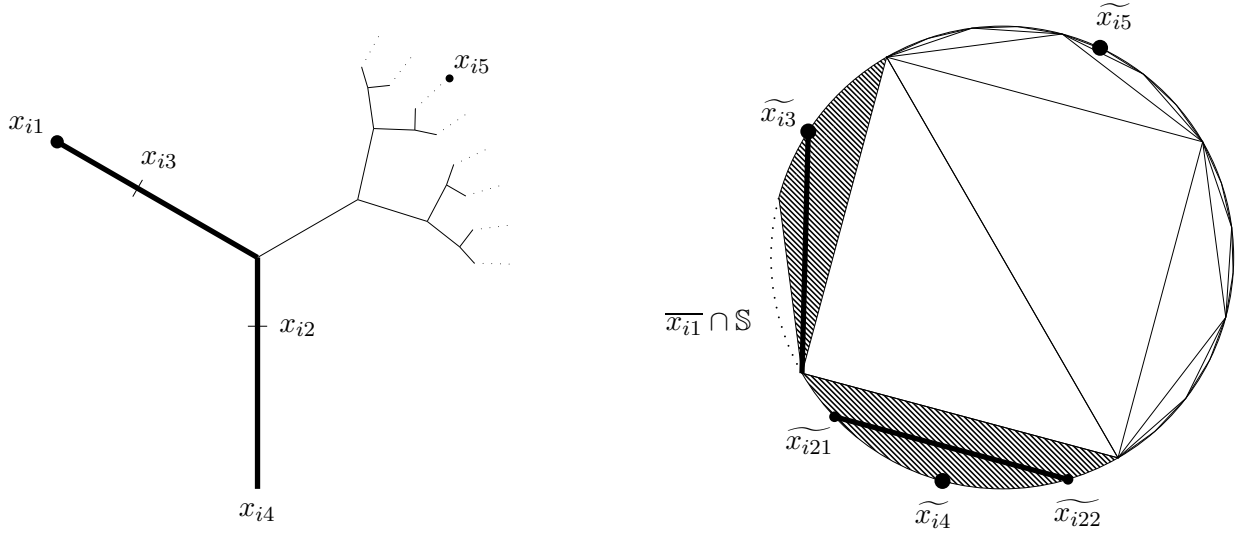


Figure 3.12: For each atom x_{i1}, \dots, x_{i5} of μ_i^n , κ_i^n assigns weight $\frac{1}{n}$ on a subset of the circle, which depends on the position of the atom in the tree. For the M_ν -atom leaf x_{i1} , κ_i^n assigns the rescaled Lebesgue measure on the corresponding circular segment $\overline{x_{i1}} \cap \mathbb{S}$. For x_{i2} , κ_i^n splits the mass $\frac{1}{n}$ between the endpoints $\widetilde{x_{i21}}$ and $\widetilde{x_{i22}}$ of the segment $\overline{x_{i2}}$. For the three other ones, $\widetilde{x_k}$ carries weight $\frac{1}{n}$.

Let $b \in \blacksquare(C)$ such that $b \cap \square(C) = \emptyset$, that is, b corresponds to a segment line of the tree carrying mass and that is not connected to a leaf without mass. In this case, $b \cap \mathbb{S}$ is the union of two disjoint intervals (see Figure 3.8, left). It is important for the next step of the proof to note that from the construction of C , one (and only one) of the two intervals is a singleton. Thus, the random K -mass we will assign to b needs to be carried by the interval with non-empty interior to ensure that $M_K = \lambda_{\mathbb{S}}$.

Step 2: construction of K . Let $(x_{ij})_{(i,j)}$ be a random infinite matrix with distribution

$$\int \nu^{\otimes \infty}(\mathrm{d}\underline{\mu}) \int \bigotimes_{i \geq 1} \mu_i^{\otimes \infty}(\cdot). \quad (3.58)$$

We assumed the tree (T, c) to be order separable and order complete. Thus the component topology on (T, c) is Polish (see Remark 3.15) and we know from [Mei19, Proposition 3.7] that, almost surely,

- (i) for every $i \in \mathbb{N}$, the weak limit $\mu_i := \text{w-lim}_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \delta_{x_{ij}}$ exists and has law ν ,
- (ii) the two-level measure $\frac{1}{m} \sum_{i=1}^m \delta_{\mu_i}$ converges weakly to ν , and
- (iii) for every $j \in \mathbb{N}$, the sequence $(x_{ij})_i$ is dense in $\text{supp}(M_\nu)$.

Fix $i, n \in \mathbb{N}$. We write $\mu_i^n := \frac{1}{n} \sum_{j=1}^n \delta_{x_{ij}}$ and we first define κ_i^n a probability measure on the circle line corresponding to μ_i^n . For each x_{ij} , we put weight $\frac{1}{n}$ on a subset of the circle depending on the position of x_{ij} in (T, c) . We will make use of the correspondence between points of $V_T \subseteq T$ and subsets of \mathbb{D} discussed at the end of Step 1. Since $(T, c, \nu) \in \mathbb{T}_2^{(2)}$, $\text{at}(\mu) \subseteq \text{lf}(T, c)$ for all $\mu \in \text{supp}(\nu)$ so we have the following possible cases (see Figure 3.12):

1. Set A_1 : if $x_{ij} \in \text{at}(M_\nu)$, $\overline{x_{ij}}$ is the circular segment $L(\alpha(x_{ij}), M_\nu\{x_{ij}\})$ and we will put the rescaled Lebesgue measure on the arc $\overline{x_{ij}} \cap \mathbb{S}$.
2. Suppose now that x_{ij} belongs to a line segment with non-atomic mass, or equivalently that $\deg(x_{ij}) = 2$. Since (T, c) is order complete, there exist $v, w \in T$ such that $(v, w) \in \text{seg}(T, c, \nu)$ and $x_{ij} \in (v, w)$.
 - (a) Set A_2 : if (v, w) is adjacent to a leaf, i.e. $\overline{x_{ij}}$ is included in a filled circular segment, then we put weight $\frac{1}{2n}$ on the endpoints $\widetilde{x_{ij1}}$ and $\widetilde{x_{ij2}}$ of the line segment $\overline{x_{ij}}$.
 - (b) Set A_3 : if not, then $\overline{x_{ij}} \subset b$ where $b \in \blacksquare(C)$ is such that $b \cap \mathbb{S}$ is the union of an interval and a singleton. In this case, we put weight $\frac{1}{n}$ on the endpoint $\widetilde{x_{ij}}$ of the line segment $\overline{x_{ij}}$ that is not the singleton.

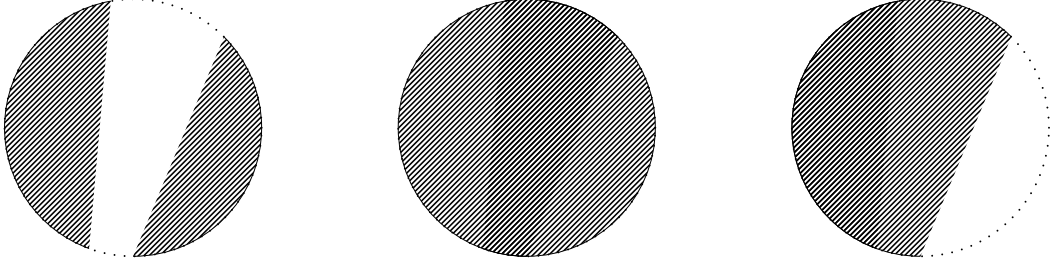


Figure 3.13: Thus union of two components is either disjoint (*left*), the whole circle \mathbb{S} (*middle*) or one component if one is included in the other (*right*).

3. Finally, suppose that $x_{ij} \in \text{lf}(T, c) \setminus \text{at}(M_\nu)$.

- (a) Set A_4 : if x_{ij} is adjacent to a line segment with non-atomic mass, then $\overline{x_{ij}} \in \square(C)$ and we put weight $\frac{1}{n}$ on $\widetilde{x_{ij}} = \overline{x_{ij}}$.
- (b) Set A_5 : if not, $x_{ij} \notin V_T$, i.e. $\overline{x_{ij}}$ is not defined. However, it naturally corresponds to a specific point of the circle. To see this, let $y \in \text{br}(T, c)$. Then there exists $(z_n)_{n \in \mathbb{N}}$ a sequence of distinct branch points in T such that $z_n \leq_y z_{n+1} \leq_y x_{ij}$ for all $n \in \mathbb{N}$. For all $n \in \mathbb{N}$, let $a_n, b_n \in \mathbb{S} \simeq [0, 1]$ such that $\text{comp}_{\overline{z_n}}(\overline{z_{n+1}}) \cap \mathbb{S} = [a_n, b_n]$. We have $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$. By (3.57), $\lambda_{\mathbb{S}}([a_n, b_n]) = M_\nu(\mathcal{S}_{z_n}(z_{n+1})) = M_\nu(\mathcal{S}_{z_n}(x_{ij}))$. But since $x_{ij} \notin \text{at}(M_\nu)$ and x_{ij} is not adjacent to a line segment with non-atomic mass, $\lim_{n \rightarrow \infty} M_\nu(\mathcal{S}_{z_n}(x_{ij})) = 0$. Therefore, there exists $\widetilde{x_{ij}} \in \mathbb{S}$ unique such that $\widetilde{x_{ij}} = \bigcap_{n \in \mathbb{N}} [a_n, b_n]$. We put weight $\frac{1}{n}$ on $\widetilde{x_{ij}}$.

Define the random probability measure κ_i^n on \mathbb{S} by

$$\kappa_i^n := \sum_{x_{ij} \in A_1} \frac{\lambda_{\mathbb{S}}(\cdot \cap \overline{x_{ij}})}{n \lambda_{\mathbb{S}}(\mathbb{S} \cap \overline{x_{ij}})} + \sum_{x_{ij} \in A_2} \frac{\delta_{\widetilde{x_{ij}1}}(\cdot) + \delta_{\widetilde{x_{ij}2}}(\cdot)}{2n} + \sum_{x_{ij} \in A_3 \cup A_4 \cup A_5} \frac{\delta_{\widetilde{x_{ij}}}(\cdot)}{n}. \quad (3.59)$$

Now for each $i, n \in \mathbb{N}$, κ_i^n is a random variable in $\mathcal{M}_1(\mathbb{S})$ which is compact when equipped with the weak topology. Therefore, the sequence $(\kappa_i^n)_n$ is tight and has a subsequence $(\kappa_i^{\varphi_i(n)})_n$ that converges to some random measure κ_i , for each $i \in \mathbb{N}$. Before taking the weak limit of the empirical distribution of the sequence $(\kappa_i)_i$, note first that a.s. μ_i and κ_i agree on (union of) corresponding components on the tree T and on the circle line \mathbb{S} . Indeed, by construction, a.s. for all $x, y \in V_T$,

$$\kappa_i^n(\text{comp}_{\overline{x}}(\overline{y}) \cap \mathbb{S}) = \mu_i^n(\mathcal{S}_x(y)). \quad (3.60)$$

Furthermore, consider $x, y, x', y' \in V_C$. Then $\text{comp}_x(y)$ and $\text{comp}_{x'}(y')$ are two circular segments delimited by two straight line segments that are disjoint or equal. Thus $\text{comp}_x(y) \cup \text{comp}_{x'}(y')$ is either disjoint, the whole circle \mathbb{S} or one component if one is included in the other (see Figure 3.13). Therefore, we have a.s. for all $x_k, y_k \in V_T$,

$$\kappa_i^n \left(\bigcup_{k \in \mathbb{N}} \text{comp}_{\overline{x_k}}(\overline{y_k}) \cap \mathbb{S} \right) = \mu_i^n \left(\bigcup_{k \in \mathbb{N}} \mathcal{S}_{x_k}(y_k) \right). \quad (3.61)$$

Then we can generalize it by saying that for n large enough, a.s. for all $x_k, y_k \in V_C$,

$$(\kappa_1^n, \dots, \kappa_m^n) \left(\bigcup_{k \in \mathbb{N}} \text{comp}_{\overline{x_k}}(\overline{y_k}) \cap \mathbb{S} \right) = (\mu_1^n, \dots, \mu_m^n) \left(\bigcup_{k \in \mathbb{N}} \mathcal{S}_{x_k}(y_k) \right). \quad (3.62)$$

Taking the limit of the subsequences $(\kappa_i^{\varphi_i(n)})_n$ when n goes to infinity for $1 \leq i \leq m$,

$$(\kappa_1, \dots, \kappa_m) \left(\bigcup_{k \in \mathbb{N}} \text{comp}_{\overline{x_k}}(\overline{y_k}) \cap \mathbb{S} \right) \quad \text{and} \quad (\mu_1, \dots, \mu_m) \left(\bigcup_{k \in \mathbb{N}} \mathcal{S}_{x_k}(y_k) \right). \quad (3.63)$$

have same distribution for all $x_k, y_k \in V_T$.

We now denote by K_m the empirical distribution of the sequence $(\kappa_i)_i$, that is,

$$K_m := \frac{1}{m} \sum_{i=1}^m \delta_{\kappa_i}, \quad (3.64)$$

which is a random variable in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$. Since $\mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$ is also compact, the sequence $(K_m)_m$ is tight, so that there exists a subsequence $(K_{\varphi(m)})_m$ that converges to some random two-level measure K . For all $m \in \mathbb{N}$, $x_k, y_k \in V_T$, and $\sigma \geq 0$,

$$\frac{1}{m} \sum_{i=1}^m e^{-\sigma \kappa_i} (\bigcup_{k \in \mathbb{N}} \text{comp}_{\overline{x_k}}(\overline{y_k}) \cap \mathbb{S}) \quad \text{and} \quad \frac{1}{m} \sum_{i=1}^m e^{-\sigma \mu_i} (\bigcup_{k \in \mathbb{N}} \mathcal{S}_{x_k}(y_k)) \quad (3.65)$$

have same distribution. Recall that the two-level measure $\frac{1}{m} \sum_{i=1}^m \delta_{\mu_i}$ converges weakly to ν . Thus, taking the limit of the subsequence $(K_{\varphi(m)})_m$, for all $x_k, y_k \in V_T$, and $\sigma \geq 0$,

$$\int K(d\kappa) e^{-\sigma \kappa} (\bigcup_{k \in \mathbb{N}} \text{comp}_{\overline{x_k}}(\overline{y_k}) \cap \mathbb{S}) \quad \text{and} \quad \int \nu(d\mu) e^{-\sigma \mu} (\bigcup_{k \in \mathbb{N}} \mathcal{S}_{x_k}(y_k)) \quad (3.66)$$

have same distribution. But the term on the right above is deterministic, so we have that a.s. for all $x, y \in V_T$, and $\sigma \geq 0$,

$$\int K(d\kappa) e^{-\sigma \kappa} (\bigcup_{k \in \mathbb{N}} \text{comp}_{\overline{x_k}}(\overline{y_k}) \cap \mathbb{S}) = \int \nu(d\mu) e^{-\sigma \mu} (\bigcup_{k \in \mathbb{N}} \mathcal{S}_{x_k}(y_k)), \quad (3.67)$$

which yields in particular (CM2).

Step 3: $M_K = \lambda_{\mathbb{S}}$ a.s. By taking the derivative of (3.67) at $\sigma = 0$ and with (3.57), a.s. for all $x_k, y_k \in V_C$,

$$M_K \left(\bigcup_{k \in \mathbb{N}} \text{comp}_{\overline{x_k}}(\overline{y_k}) \cap \mathbb{S} \right) = M_{\nu} \left(\bigcup_{k \in \mathbb{N}} \mathcal{S}_{x_k}(y_k) \right) = \lambda_{\mathbb{S}} \left(\bigcup_{k \in \mathbb{N}} \text{comp}_{\overline{x_k}}(\overline{y_k}) \cap \mathbb{S} \right). \quad (3.68)$$

In other words, $M_K = \lambda_{\mathbb{S}}$ on the σ -algebra generated by $\{\text{comp}_{\overline{x}}(\overline{y}) \cap \mathbb{S} : x, y \in V_T\}$. To prove the equality on the Borel σ -algebra of \mathbb{S} , we partition the circle line \mathbb{S} in the following way. Since $\Delta(C)$ is countable, there exist countably many disjoint intervals of non-empty interior I_p , $p \in \mathbb{N}$, such that

$$\mathbb{S} = \bigsqcup_{p \in \mathbb{N}} I_p, \quad (3.69)$$

and for each $p \in \mathbb{N}$, I_p satisfies one of the following (see Figure 3.12 for a sub-triangulation with the different types of boundaries):

- (a) I_p corresponds to an atom leaf of (T, c) , that is, there exists $w \in \text{at}(M_{\nu})$ such that $(I_p)^{\circ} = \overline{w} \cap \mathbb{S}$,
- (b) I_p corresponds to a line segment in $\text{seg}(T, c, \nu)$, i.e. there exists $b \in \blacksquare(C)$ such that $(I_p)^{\circ} = (b \cap \mathbb{S})^{\circ}$,
- (c) $I_p \subseteq C$ and for all distinct $x, y \in I_p$, there exists $c_{xy} \in \Delta(C)$ such that x, y belong to different connected components of $\mathbb{D} \setminus \partial c_{xy}$.

We show that a.s. $M_K = \lambda_{\mathbb{S}}$ by showing that a.s. for any $I := I_p$, $M_{K|I} = \lambda_I$.

- (a) Suppose $I^{\circ} = \overline{w} \cap \mathbb{S}$ for $w \in \text{at}(M_{\nu})$. For all $n \leq i$, we defined κ_i^n such that

$$\kappa_{i|I}^n = C_I^{i,n} \lambda_I \quad (3.70)$$

where $C_I^{i,n}$ is a real-valued random-variable. Taking the weak limit of $(\kappa_i^{\varphi_i(n)})_n$ when n goes to infinity,

$$\kappa_{i|I} = C_I^i \lambda_I \quad (3.71)$$

a.s. for some real-valued random-variable C_I^i . Therefore, a.s. for all $m \in \mathbb{N}$,

$$M_{K_m|I} = \left(\frac{1}{m} \sum_{i=1}^m C_I^i \right) \lambda_I. \quad (3.72)$$

Taking the weak limit of $(K_{\varphi(m)})_m$ when m goes to infinity,

$$M_{K|I} = C_I \lambda_I \quad (3.73)$$

a.s. for some real-valued random-variable C_I . Noticing that I belongs to the σ -algebra generated by $\{\text{comp}_{\overline{x}}(\overline{y}) \cap \mathbb{S} : x, y \in V_C\}$, we have $M_K(I) = \lambda_I(I)$ so that $C_I = 1$ a.s.

(b) Suppose that $I^\circ = (b \cap \mathbb{S})^\circ$ for some $b \in \blacksquare(C)$. We will use here the assumption that for the corresponding $(v, w) \in \text{seg}(T, c, \nu)$, M_ν restricted to (v, w) is the Lebesgue measure.

- If $b \cap \mathbb{S}$ is the union of two disjoint intervals (one of which is a singleton), then the map that associates a point in (v, w) to the endpoint of the corresponding segment in b is linear. Therefore, with the assumption on ν , $M_{K|I} = \lambda_I$.
- If not, let $x, x' \in \mathbb{S}$ such that $b \cap \mathbb{S} = [x, x']$, $y = y' = \frac{x+x'}{2}$ (see Figure 3.8). In this case, (v, w) is adjacent to a leaf and we can assume w.l.o.g. that $w \in \text{lf}(T, c)$. The point $u = (1-t)v + tw$, $t \in (0, 1)$ corresponds to the segment with endpoints $\widetilde{u}_1 = (1-t)x + ty$ and $\widetilde{u}_2 = (1-t)x' + ty'$ (see Figure 3.12). Therefore, by construction and assumption on ν , $M_{K|(x,y)}$ is proportional to $\lambda_{(x,y)}$ and $M_K(x, y) = \frac{M_K(x, x')}{2} = \frac{\lambda_{\mathbb{S}}(x, x')}{2} = \lambda_{\mathbb{S}}(x, y)$ because $[x, x']$ is the intersection of a component with \mathbb{S} . Thus $M_{K|(x,y)} = \lambda_{(x,y)}$ and since it also holds for (y', x') , $M_{K|I} = \lambda_I$.

(c) Finally, we assume that $I \subseteq C$ and for all distinct $x, y \in I$, there exists $c_{xy} \in \Delta(C)$ such that x, y belong to different connected components of $\mathbb{D} \setminus \partial c_{xy}$. Note that in this case, $[x, y] \cap c_{xy} \neq \emptyset$. In other words, the set of endpoints of (boundaries of) triangles in $\Delta(C)$ is dense in I . Let $J = [a, b] \subseteq I$. There exist two sequences $(a_n)_n$ and $(b_n)_n$ in I such that for all n , $a_n \in \partial_{\mathbb{D}} v$ and $b_n \in \partial_{\mathbb{D}} w$ for some $v, w \in \Delta(C)$ and $a_n \uparrow a$ and $b_n \downarrow b$. Since $a_n \in \partial_{\mathbb{D}} v$ and $b_n \in \partial_{\mathbb{D}} w$, $[a_n, b_n]$ is the union of countably many disjoint components of C intersected with \mathbb{S} , so that $M_K([a_n, b_n]) = \lambda_{\mathbb{S}}([a_n, b_n])$. Therefore, $M_K(J) = \lim_{n \rightarrow \infty} M_K([a_n, b_n]) = \lambda_{\mathbb{S}}(J)$. Since it holds for all interval $J \subseteq I$, $M_{K|I} = \lambda_I$.

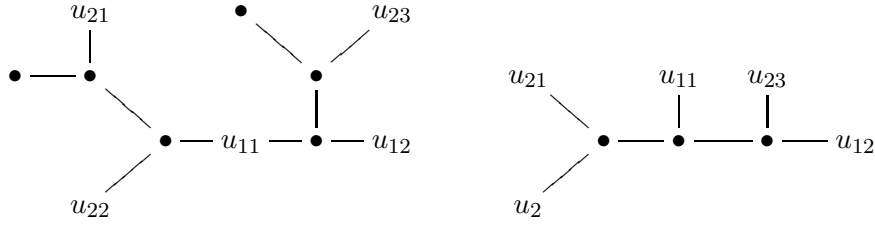
We have shown that a.s. the random two-level measure K together with the sub-triangulation C is such that $\tau(C, K) = (T, c, \nu)$ and $M_K = \lambda_{\mathbb{S}}$. Therefore, by taking a realization of K such that it holds, we have shown surjectivity of τ . \square

3.5 Topologies on the subspace of binary algebraic two-level measure trees

In this section we introduce on the space of binary a2m trees another notion of convergence called *two-level sample shape convergence*. It exploits the idea of the Gromov-weak topology to sample finite sub-spaces and then require these to converge in distribution. Whereas one samples metric sub-spaces in the Gromov-weak topology, we now consider subtrees of a2m trees as combinatorial objects, which will make it useful to show convergence of tree-valued Markov chains in the future. We then show that this topology is equivalent to the two-level bpdd-Gromov-weak topology on $\mathbb{T}_2^{(2)}$, and that both topologies are compact.

3.5.1 Two-level sample shape convergence

Obviously the two-level sample shape convergence is similar to the notion sample shape convergence defined on the space of algebraic measure trees. A sequence of trees converges to a limit tree if all random tree shapes spanned by finite samples converge weakly to the corresponding limit shapes.


 Figure 3.14: A tree T and the shape $\mathfrak{s}_{(T,c)}(u_{11}, u_{12}, u_{21}, u_{22}, u_{23})$.

On $\mathbb{T}_2^{(2)}$, there is a two-level sampling: we first sample a finite number of measures on the tree from the two-level measure and then sample a finite number of points according to each sampled measure. Therefore, the notion of tree shape we define keeps track of this two-level sampling by using double indices.

For $m \in \mathbb{N}$ and $\underline{n} \in \mathbb{N}^m$, we denote by

$$\mathbb{N}_{m,\underline{n}} = \{(i, j) : 1 \leq i \leq m, 1 \leq j \leq n_i\} \quad (3.74)$$

the set of indices. The sampled subtrees from an a2m tree will be cladograms, which are binary finite trees with labelled leaves. Since the two-level measure on the tree may have atoms on leaves, a given leaf may be sampled several times so we need to allow multi-labels in cladograms.

Definition 3.40 ((m, \underline{n}) -cladogram). For $m \in \mathbb{N}$ and $\underline{n} \in \mathbb{N}^m$, an (m, \underline{n}) -labelled cladogram is a binary, finite algebraic tree (C, c) consisting only of leaves and branch points together with a surjective labelling map $\zeta: \mathbb{N}_{m,\underline{n}} \rightarrow \text{lf}(C)$. An (m, \underline{n}) -cladogram (C, c, ζ) is an (m, \underline{n}) -labelled cladogram such that ζ is also injective.

We call two (m, \underline{n}) -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 $\zeta_2 = \phi \circ \zeta_1$. We then write

$$\overline{\mathfrak{C}}_{m,\underline{n}} := \{\text{isomorphism classes of } (m, \underline{n})\text{-labelled cladograms}\} \quad (3.75)$$

and

$$\mathfrak{C}_{m,\underline{n}} := \{(C, c, \zeta) \in \overline{\mathfrak{C}}_{m,\underline{n}} : \zeta \text{ injective}\}. \quad (3.76)$$

The shape function encodes as cladograms the shape of the subtree spanned by finite samples of points (see Figure 3.14).

Definition 3.41 (Shape function). Let (T, c) be a binary algebraic tree, $m \in \mathbb{N}$, $\underline{n} \in \mathbb{N}^m$, and $u_{ij} \in T \setminus \text{br}(T)$ for $(i, j) \in \mathbb{N}_{m,\underline{n}}$. Then there exists a unique (up to isomorphism) (m, \underline{n}) -labelled cladogram

$$\mathfrak{s}_{(T,c)}(\underline{u}) = (C, c_C, \zeta) \quad (3.77)$$

with $\text{lf}(C) = \{u_{ij}\}_{(i,j) \in \mathbb{N}_{m,\underline{n}}}$ and $\zeta(i, j) = u_{ij}$, such that the identity on $\text{lf}(C)$ extends to a tree homomorphism π from C onto $c\left(\left(\{u_{ij}\}_{(i,j) \in \mathbb{N}_{m,\underline{n}}}\right)^3\right)$, i.e. for all $(i_1, j_1), (i_2, j_2), (i_3, j_3) \in \mathbb{N}_{m,\underline{n}}$,

$$\pi(c_C(u_{i_1 j_1}, u_{i_2 j_2}, u_{i_3 j_3})) = c(u_{i_1 j_1}, u_{i_2 j_2}, u_{i_3 j_3}). \quad (3.78)$$

We will refer to $\mathfrak{s}_{(T,c)}(\underline{u}) \in \overline{\mathfrak{C}}_m$ as the *shape* of \underline{u} in (T, c) .

We introduce a notion of convergence on $\mathbb{T}_2^{(2)}$ based on the weak convergence of random tree shapes spanned by finite samples. This topology is defined as the topology induced by the two-level shape polynomials, which are test functions evaluating the tree shape distributions.

Definition 3.42 (Two-level shape polynomials). A *two-level shape polynomial* is a function $\Phi: \mathbb{T}_2^{(2)} \rightarrow \mathbb{R}$ of the form

$$\Phi(\chi) = \Phi^{m,\underline{n},\varphi}(\chi) := \int_{(\mathcal{M}_1(T))^m} \nu^{\otimes m}(\mathrm{d}\underline{\mu}) \int_{T^{\lfloor \underline{n} \rfloor}} \bigotimes_{i=1}^m \mu_i(\mathrm{d}u_i) \varphi(\mathfrak{s}_{(T,c)}(\underline{u})), \quad (3.79)$$

where $\chi = (T, c, \nu)$, $m \in \mathbb{N}$, $\underline{n} \in \mathbb{N}^m$ and $\varphi: \mathfrak{C}_{m,\underline{n}} \rightarrow \mathbb{R}$. We write $\Pi_s^{(2)}$ for the set of all two-level shape polynomials.

Definition 3.43 (Two-level sample shape topology). The *two-level sample shape topology* on $\mathbb{T}_2^{(2)}$ is defined as the initial topology induced by $\Pi_s^{(2)}$.

Remark 3.44 (Tree shape distribution). Fix $m \in \mathbb{N}$ and $\underline{n} \in \mathbb{N}^m$. For all $\chi = (T, c, \nu)$, we define the (m, \underline{n}) -tree shape distribution $\mathfrak{S}_{m, \underline{n}}(\chi)$ as the probability measure on $\mathfrak{C}_{m, \underline{n}}$ such that for all $\varphi: \mathfrak{C}_{m, \underline{n}} \rightarrow \mathbb{R}$,

$$\int_{\mathfrak{C}_{m, \underline{n}}} d\mathfrak{S}_{m, \underline{n}}(\chi) \varphi = \int_{\mathcal{M}_1(T)} \nu^{\otimes m}(d\mu) \int_{T^{|\underline{n}|}} \bigotimes_{i=1}^m \mu_i(d\mathbf{u}_i) \varphi(\mathfrak{s}_{(T, c)}(\underline{u})). \quad (3.80)$$

Then the two-level sample shape topology is induced by the set of functions $\{\mathfrak{S}_{m, \underline{n}} : m \in \mathbb{N}, \underline{n} \in \mathbb{N}^m\}$.

The following result gives that on $\mathbb{T}_2^{(2)}$, two-level sample shape convergence implies two-level bpdd-Gromov-weak convergence.

Proposition 3.45. *On $\mathbb{T}_2^{(2)}$, the two-level sample shape topology is stronger than the two-level bpdd-Gromov-weak topology.*

Proof. By definition, the two-level bpdd-Gromov-weak topology is induced by the set $\Pi_l^{(2)}$ of polynomials of the form

$$\Phi(\chi) := \int \nu^{\otimes m}(d\mu) \int \bigotimes_{i=1}^m \mu_i(d\mathbf{u}_i) \varphi((r_\lambda(u_{ij}, u_{i'j'}))_{(i,j),(i',j')}), \quad (3.81)$$

where $\chi = (T, c, \nu)$, $\lambda := c_* M_\nu^{\otimes 3}$, $m \in \mathbb{N}$, $\underline{n} \in \mathbb{N}^m$ and $\varphi \in \mathcal{C}_b(\mathbb{R}^{|\underline{n}|^2})$ (see Definition 3.22 and Definition 3.30). Now, the set of $\phi \in \mathcal{C}_b(\mathbb{R}^{|\underline{n}|^2})$ that are Lipschitz continuous is convergence determining for probability measure on $\mathbb{R}^{|\underline{n}|^2}$. Hence, the subset of $\Psi \in \Pi_l^{(2)}$ with

$$\Psi(T, c, \nu) = \int \nu^{\otimes m}(d\mu) \int \bigotimes_{k=1}^m \mu_k^{\otimes n_k}(d\mathbf{u}_k) \phi((r_\lambda(u_{ij}, u_{i'j'}))_{(i,j),(i',j')}) \quad (3.82)$$

for some $m \in \mathbb{N}$, $\underline{n} \in \mathbb{N}^m$ and a Lipschitz continuous function $\phi \in \mathcal{C}_b(\mathbb{R}^{|\underline{n}|^2})$ also induces the two-level bpdd-Gromov-weak topology. Therefore, it is enough to show that such a Ψ is continuous on $\mathbb{T}_2^{(2)}$ with respect to the sample shape topology. To do this, we show that the restriction of Ψ to $\mathbb{T}_2^{(2)}$ can be uniformly approximated by polynomials in $\Pi_s^{(2)}$.

For $p \in \mathbb{N}$ with $3p \geq m$, we define

$$\Phi_p(T, c, \nu) = \int \nu^{\otimes 3p}(d\mu) \int \bigotimes_{k=1}^m \mu_k^{\otimes n_k}(d\mathbf{u}_k) \bigotimes_{k=m+1}^{3p} \mu_k(d\mathbf{u}_k) \phi((r_{\lambda_{p, \underline{u}}}(u_{ij}, u_{i'j'}))_{(i,j),(i',j')}) \quad (3.83)$$

where

$$\lambda_{p, \underline{u}} := \frac{1}{p} \sum_{l=0}^p \delta_{c(u_{3l+1,1}, u_{3l+2,1}, u_{3l+3,1})}. \quad (3.84)$$

Note that whether or not $c(u_{3l+1,1}, u_{3l+2,1}, u_{3l+3,1})$ lies on $[u_{(i,j)}, u_{(i',j')}]$, for some $l \in \{0, \dots, p-1\}$ and $(i, j), (i', j') \in \mathbb{N}_{m, \underline{n}}$ only depends on the shape $\mathfrak{s}_{(T, c)}(\underline{u})$ and thus the restriction of Φ_p to $\mathbb{T}_2^{(2)}$ belongs to $\Pi_s^{(2)}$.

To show that Φ_p approximates Ψ , we uniformly bound the distance of the empirical branch point distribution to the branch point distribution. Recall that for $x, y \in T$,

$$r_\lambda(x, y) := \lambda([x, y]) - \frac{1}{2}\lambda(\{x\}) - \frac{1}{2}\lambda(\{y\}). \quad (3.85)$$

Thus, denoting by L the Lipschitz constant of ϕ w.r.t. the l_∞ -norm on $\mathbb{R}^{|\underline{n}|^2}$,

$$\begin{aligned} \|\Psi - \Phi_p\|_\infty &\leq \sup_{(T, c, \nu) \in \mathbb{T}_2^{(2)}} \int \nu^{\otimes 3p}(d\mu) \int \bigotimes_{k=1}^m \mu_k^{\otimes n_k}(d\mathbf{u}_k) \bigotimes_{k=m+1}^{3p} \mu_k(d\mathbf{u}_k) L \sup_{I \in \mathcal{I}_T} |\lambda(I) - \lambda_{p, \underline{u}}(I)| \\ &\leq 2L \sup_{(T, c, \nu) \in \mathbb{T}_2^{(2)}} \int (M_\nu)^{\otimes 3p}(d\mathbf{u}) \sup_{I \in \mathcal{I}_T} |\lambda(I) - \lambda_{p, \underline{u}}(I)|, \end{aligned} \quad (3.86)$$

where $\mathcal{I}_T := \{[x, y] : x, y \in T\}$ is the collection of all intervals of the algebraic tree T .

Then [LW21, Lemma A.4] gives the following estimates of the rate of convergence in the approximation of the branch point distribution by empirical distribution:

$$\int (M_\nu)^{\otimes 3p}(\underline{d}\underline{u}) \sup_{I \in \mathcal{I}_T} |\lambda(I) - \lambda_{p, \underline{u}}(I)| \leq 96 \sqrt{\frac{\dim_{\text{VC}}(\mathcal{I}_T)}{p}} = 96 \sqrt{\frac{2}{p}}, \quad (3.87)$$

where $\dim_{\text{VC}}(\mathcal{I}_T)$ is the Vapnik-Chervonenkis dimension of \mathcal{I}_T which can easily be shown to be 2 (see [LW21, Example A.2] for more details). This concludes the proof. \square

Corollary 3.46 (Metrizability). *The two-level sample shape topology is metrizable on $\mathbb{T}_2^{(2)}$.*

Proof. By Proposition 3.45, the two-level sample shape topology is stronger than the two-level bpdd Gromov-weak topology, which is Hausdorff by Proposition 3.31. Hence the two-level sample shape topology is also Hausdorff. Moreover, by Remark 3.44, it is induced by the set $\{\mathfrak{S}_{m, \underline{n}} : m \in \mathbb{N}, \underline{n} \in \mathbb{N}^m\}$, which is a countable family of functions with values in metrizable spaces. Thus we can define a pseudo-metric on $\mathbb{T}_2^{(2)}$ that induces the two-level sample shape topology as follows: for $\chi, \chi' \in \mathbb{T}_2^{(2)}$,

$$d_s(\chi, \chi') := \sum_{m \in \mathbb{N}} \frac{1}{2^m} \sum_{\underline{n} \in \mathbb{N}^m} \frac{1}{2^{|\underline{n}|}} \min \left(d_{\text{Pr}}(\mathfrak{S}_{m, \underline{n}}(\chi), \mathfrak{S}_{m, \underline{n}}(\chi')), 1 \right), \quad (3.88)$$

where d_{Pr} denotes the Prokhorov distance. Therefore, the two-level sample shape topology is Hausdorff and pseudo-metrizable. In particular, if $d_s(\chi, \chi') = 0$ then every open set that contains χ also contains χ' , so that d_s is actually a metric. \square

3.5.2 Equivalence and compactness of topologies

In this section, we first show that the coding map τ is continuous when $\mathbb{T}_2^{(2)}$ is equipped with the two-level sample shape topology. This implies that it is a compact topology. Finally, we prove that the two-level sample shape convergence and the two-level bpdd Gromov-weak convergence are equivalent on $\mathbb{T}_2^{(2)}$.

Recall the space \mathfrak{D} of pairs $(C, K) \in \mathcal{T} \times \mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$ such that $M_K = \lambda_{\mathbb{S}}$.

Proposition 3.47. *Let \mathcal{T} be equipped with the Hausdorff metric topology, $\mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$ with the weak topology, $\mathcal{T} \times \mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$ with the product topology and $\mathbb{T}_2^{(2)}$ with the two-level sample shape topology. Then the coding map $\tau : \mathfrak{D} \rightarrow \mathbb{T}_2^{(2)}$ is continuous.*

We will use the following lemma in the proof.

Lemma 3.48. *Let (S, d) be separable and let $P, Q \in \mathcal{M}_1(S)$, $\epsilon > 0$ such that $d_{\text{Pr}}(P, Q) \leq \epsilon$ where d_{Pr} denotes the Prokhorov distance. Then we can define two random variables X , resp. Y , distributed according to P , resp. Q , on the same probability space (Ω, \mathbb{P}) such that*

$$\mathbb{P}\{d(X, Y) \geq \epsilon\} \leq \epsilon. \quad (3.89)$$

Proof. The lemma directly follows from [EK86, Theorem 3.1.2]:

$$d_{\text{Pr}}(P, Q) = \inf_{\gamma} \inf \{ \epsilon > 0 \mid \gamma\{(x, y) \mid d(x, y) \geq \epsilon\} \leq \epsilon \}, \quad (3.90)$$

where the first infimum is taken over the set of measures $\gamma \in \mathcal{M}_1(S \times S)$ with marginals P and Q . \square

We can now proceed with the proof of the continuity of the coding map.

Proof of Proposition 3.47. Fix $\Gamma = (C, K) \in \mathfrak{D}$, $m \in \mathbb{N}$ and $\underline{n} \in \mathbb{N}^m$. By Remark 3.44, it is enough to show that $\mathfrak{S}_{m, \underline{n}} \circ \tau : \mathcal{T} \rightarrow \mathcal{M}_1(\mathfrak{C}_{m, \underline{n}})$ is continuous at (C, K) .

Let $\kappa_1, \dots, \kappa_m$ be independent, identically distributed measures on $\mathcal{M}_1(\mathbb{S})$ with distribution K . For all $1 \leq i \leq m$, let U_{i1}, \dots, U_{in_i} be independent, identically distributed points on \mathbb{S} with distribution κ_i . Fix $\epsilon > 0$. Since $M_K = \lambda_{\mathbb{S}}$, there exist $N \in \mathbb{N}$ and $v_1, \dots, v_N \in \Delta(C) \cap \nabla(C)$ distinct, such that with probability at least $1 - \epsilon$ the following holds:

- if $\{U_{ij} : (i, j) \in \mathbb{N}_{m, \underline{n}}\} \cap \{v\} \neq \emptyset$ for some $v \in \nabla(C)$, then $v \in \{v_1, \dots, v_N\}$, and
- if $\{U_{ij} : (i, j) \in \mathbb{N}_{m, \underline{n}}\} \cap \text{comp}_v(w) \neq \emptyset$ for some $v \in \Delta(C)$ and all $w \in \Delta(C) \cup \nabla(C) \cup \square(C)$ with $w \neq v$, then $v \in \{v_1, \dots, v_N\}$.

Put $\epsilon' := \epsilon(12N|\underline{n}|)^{-1}$. Then, since $M_K = \lambda_{\mathbb{S}}$, each U_{ij} is distributed according to the Lebesgue measure on \mathbb{S} , so that

$$\begin{aligned} & \mathbb{P}(\{d(U_{ij}, \partial v_k) \geq \epsilon', \forall (i, j) \in \mathbb{N}_{m, \underline{n}}, k = 1, \dots, N\}) \\ &= 1 - \sum_{(i, j) \in \mathbb{N}_{m, \underline{n}}} \mathbb{P}(\{d(U_{ij}, \partial v_k) \leq \epsilon', \forall k = 1, \dots, N\}) \\ &\geq 1 - |\underline{n}| \mathbb{P}(\{d(U_{11}, \partial v_k) \leq \epsilon', \forall k = 1, \dots, N\}) \\ &\geq 1 - \epsilon. \end{aligned} \tag{3.91}$$

Now, there is a $\delta = \delta(\epsilon) > 0$ small enough such that for any $C' \in \mathcal{T}$ satisfying $d_H(C, C') < \delta$ there are distinct $v'_1, \dots, v'_N \in \Delta(C') \cup \nabla(C')$ such that $d_H(v_k, v'_k) \leq \frac{\epsilon'}{2}$ for $k = 1, \dots, N$. Let $K' \in \mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$ such that $M_{K'} = \lambda_{\mathbb{S}}$. Suppose that $d_{\text{Pr}}(K, K') \leq \epsilon'' := \min(\frac{\epsilon}{2|\underline{n}|}, \frac{\epsilon'}{2})$. By Lemma 3.48, there exist $\kappa'_1, \dots, \kappa'_m$ independent, identically distributed measures on $\mathcal{M}_1(\mathbb{S})$ with distribution K' coupled to $\kappa_1, \dots, \kappa_m$ such that for all $i = 1, \dots, m$, $\mathbb{P}\{d_{\text{Pr}}(\kappa_i, \kappa'_i) \geq \epsilon''\} \leq \epsilon''$. Applying again Lemma 3.48 for each $1 \leq i \leq m$, there exist $U'_{i1}, \dots, U'_{in_i}$ independent, identically distributed points on \mathbb{S} with distribution κ'_i coupled to U_{i1}, \dots, U_{in_i} such that for all $j = 1, \dots, n_i$,

$$\mathbb{P}\{d(U_{ij}, U'_{ij}) \leq \epsilon''\} = \mathbb{P}\{d(U_{ij}, U'_{ij}) \leq \epsilon'' \mid d_{\text{Pr}}(\kappa_i, \kappa'_i) \leq \epsilon''\} \mathbb{P}\{d_{\text{Pr}}(\kappa_i, \kappa'_i) \leq \epsilon''\} \geq 1 - 2\epsilon''. \tag{3.92}$$

Therefore,

$$\mathbb{P}(\{d(U_{ij}, U'_{ij}) \leq \epsilon'', \forall (i, j) \in \mathbb{N}_{m, \underline{n}}\}) \geq 1 - 2|\underline{n}|\epsilon''. \tag{3.93}$$

Thus, since $\epsilon'' \leq \frac{\epsilon'}{2}$, and using (3.91),

$$\mathbb{P}\left(\left\{d(U'_{ij}, \partial v_k) \geq \frac{\epsilon'}{2}, \forall (i, j) \in \mathbb{N}_{m, \underline{n}}, k = 1, \dots, N\right\}\right) \geq 1 - 2\epsilon. \tag{3.94}$$

Let $\chi = (T, c, \nu) := \tau(C, K)$ and $(V_{ij})_{(i, j) \in \mathbb{N}_{m, \underline{n}}}$ be distributed according to

$$M_{\nu}^{m, \underline{n}}(\cdot) := \int \nu^{\otimes m}(\mathrm{d}\underline{\mu}) \int \bigotimes_{i=1}^m \mu_i^{\otimes n_i}(\cdot), \tag{3.95}$$

coupled to $(U_{ij})_{(i, j) \in \mathbb{N}_{m, \underline{n}}}$ such that $V_{ij} \in \mathcal{S}_v(w)$ if and only if $U_{ij} \in \text{comp}_v(w)$. This is possible due to the properties of the coding map τ stated in Theorem 3.39. Define χ' and $(V'_{ij})_{(i, j) \in \mathbb{N}_{m, \underline{n}}}$ similarly with (C, K) replaced by (C', K') . Then

$$\mathbb{P}(\{\mathfrak{s}_{(T, c)}(\underline{V}) = \mathfrak{s}_{(T', c')}(\underline{V}')\}) \geq 1 - 2\epsilon. \tag{3.96}$$

Therefore,

$$d_{\text{Pr}}(\mathfrak{S}_{m, \underline{n}}(\tau(C, K)), \mathfrak{S}_{m, \underline{n}}(\tau(C', K'))) \leq 2\epsilon. \tag{3.97}$$

We proved that $\mathfrak{S}_{m, \underline{n}} \circ \tau$ is continuous at (C, K) , with m, \underline{n} and (C, K) arbitrary. Therefore, the coding map τ is continuous. \square

We finish this section with our second main result.

Theorem 3.49 (Equivalence of topologies and compactness). *The two-level sample shape topology and the two-level bpdd-Gromov-weak topology coincide on $\mathbb{T}_2^{(2)}$. Furthermore, $\mathbb{T}_2^{(2)}$ is compact and metrizable in this topology.*

Proof. We start by showing that the two-level sample shape topology on $\mathbb{T}_2^{(2)}$ is compact. The set \mathcal{T} equipped with the Hausdorff metric topology is compact (see [LW21, Lemma 4.2]). Moreover, the circle line \mathbb{S} is compact so that $\mathcal{M}_1(\mathbb{S})$ is compact and $\mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$ as well. Since the subset of two-level measures K such that $M_K = \lambda_{\mathbb{S}}$ is closed in $\mathcal{M}_1(\mathcal{M}_1(\mathbb{S}))$, \mathfrak{D} is a compact space. By Theorem 3.39, the coding map is surjective and by Proposition 3.47, it is continuous when $\mathbb{T}_2^{(2)}$ is equipped with the two-level sample shape topology. Therefore, the sample shape topology is a compact topology.

Furthermore, the two-level bpdd-Gromov-weak topology is a Hausdorff topology by Proposition 3.31 and weaker than the two-level sample shape topology by Proposition 3.45. Thus, both topologies coincide on $\mathbb{T}_2^{(2)}$, and we know from Corollary 3.46 that it is metrizable. \square

3.6 Example: the Kingman algebraic two-level measure tree

We introduce here the Kingman algebraic two-level measure tree, which corresponds to the nested Kingman coalescent measure tree, as defined in [Mei19], without branch length. For this, we rely on a sampling consistency of the nested Kingman coalescent and the compactness of $\mathbb{T}_2^{(2)}$. We first recall a definition of the nested Kingman coalescent for a host-parasite population indexed in \mathbb{N}^2 , that is, (i, j) is the j -th parasite in the i -th host.

For $I \subseteq \mathbb{N}$ let $\mathcal{E}(I) \subset I^2$ be the set of equivalence relations on I . The equivalence classes of an equivalence relation are called *blocks*. Let

$$\mathcal{P}_{nest}(I) \subset \mathcal{E}(I) \times \mathcal{E}(I) \quad (3.98)$$

denote the set of all equivalence relations such that

- if $i_1 = i_2$, then (i_1, j_1) and (i_2, j_2) belong to the same block of \mathfrak{p}_H , that is, all parasites of the same host belong to the same block of \mathfrak{p}_H ,
- if (i_1, j_1) and (i_2, j_2) belong to the same block of \mathfrak{p}_P , then they belong to the same block of \mathfrak{p}_H .
In other words, each block of \mathfrak{p}_P is contained in a single block of \mathfrak{p}_H .

Therefore, \mathfrak{p}_H represents the population of hosts and \mathfrak{p}_P the population of parasites.

We also define the following equivalence relations on \mathbb{N}^2 :

$$\begin{aligned} P_0 &:= \{((i, j), (i, j)) \mid (i, j) \in \mathbb{N}^2\} \\ H_0 &:= \{((i, j), (i, k)) \mid i, j, k \in \mathbb{N}\}, \end{aligned} \quad (3.99)$$

which will be the initial states of the nested Kingman coalescent.

Definition 3.50 (Finite nested Kingman coalescent). Let I be a finite subset of \mathbb{N}^2 and $\gamma_H, \gamma_P > 0$. The *finite nested Kingman coalescent*

$$(\mathcal{K}^I(t))_{t \geq 0} = (\kappa_H^I(t), \kappa_P^I(t))_{t \geq 0} \quad (3.100)$$

on I with rates (γ_H, γ_P) is a continuous-time Markov process with values in $\mathcal{P}_{nest}(I)$ such that:

1. The initial state is $\mathcal{K}^I(0) = (H_0 \cap I^2, P_0 \cap I^2)$.
2. $(\kappa_H^I(t))_{t \geq 0}$ is a Kingman coalescent with rate γ_H , i.e. any pair of blocks in $\kappa_H^I(t)$ merge at rate γ_H .
3. $(\kappa_P^I(t))_{t \geq 0}$ behaves in the following way: any pair of blocks π_1, π_2 of $\kappa_P^I(t)$ such that $\pi_1 \cup \pi_2$ is contained in a single block of $\kappa_H^I(t)$ merge at rate γ_P . Other blocks of $\kappa_P^I(t)$ cannot merge.

Roughly speaking, hosts merge as in a Kingman coalescent with rate γ_H , and parasites within the same host merge as in a Kingman coalescent with rate γ_P . Since this process has only finitely many states, it is well defined and unique. Furthermore, it satisfies an important property:

Proposition 3.51 (Sampling consistency). Let I and J be two finite subsets of \mathbb{N}^2 such that $J \subset I$ and let $\gamma_H, \gamma_P > 0$. The restriction of \mathcal{K}^I to J has same distribution as \mathcal{K}^J , that is, the restriction of \mathcal{K}^I to J is a finite nested Kingman coalescent on J with rates (γ_H, γ_P) .

Using this property, we can show the existence of the nested Kingman coalescent for an infinite set of hosts and of parasites (see [BDLS18, Section 5] for a construction of more general nested coalescents).

Definition 3.52 (Nested Kingman coalescent). Let $\gamma_H, \gamma_P > 0$. The *nested Kingman coalescent* \mathcal{K} with rates (γ_H, γ_P) is a continuous-time Markov process with values in $\mathcal{P}_{\text{nest}}(\mathbb{N}^2)$ such that for any finite subset I of \mathbb{N}^2 , the restriction of \mathcal{K} to I is a finite nested Kingman coalescent on I with rates (γ_H, γ_P) .

We now construct the Kingman a2m tree with rates (γ_H, γ_P) . Recall from Definition 3.8 the notion of rooted algebraic tree and from Proposition 3.10 its relation to (unrooted) algebraic trees. Given $\mathcal{K} = (\kappa_H, \kappa_P)$, we define the *nested Kingman rooted algebraic tree* as the random rooted algebraic tree (T, c_\wedge, ρ) with the vertex set

$$T := \{\varpi_\rho\} \uplus \bigcup_{t \geq 0} \bigcup_{\varpi \in \kappa_P(t)} \varpi, \quad (3.101)$$

with $\varpi_\rho = \mathbb{N}^2 \uplus \{\rho\}$ for a point $\rho \notin \mathbb{N}^2$ and the minimal map c_ρ which sends two elements $\varpi, \varpi' \in T$ to the smallest $\tilde{\varpi} \in T$ which contains both ϖ and ϖ' , i.e.,

$$c_\wedge(\varpi, \varpi') := \bigcap_{\tilde{\varpi} \in T: \varpi, \varpi' \subseteq \tilde{\varpi}} \tilde{\varpi}. \quad (3.102)$$

Further, we define the *nested Kingman algebraic tree* as the random algebraic space (T, c) obtained from the rooted nested Kingman algebraic tree $(T, c_\wedge, \varpi_\rho)$ as in (3.10).

For $M \in \mathbb{N}$ and $\underline{N} \in \mathbb{N}^M$, define also the two-level measure

$$\nu^{M, \underline{N}} := \frac{1}{M} \sum_{i=1}^M \delta_{\frac{1}{N_i} \sum_{j=1}^{N_i} \delta_{\{(i,j)\}}}, \quad (3.103)$$

so that $(T, c, \nu^{M, \underline{N}})$ is a random binary a2m tree in $\mathbb{T}_2^{(2)}$. We define the Kingman a2m tree as the weak limit (with respect to the two-level sample shape convergence) of these random trees as $M, \underline{N} \rightarrow \infty$, that is, M and $\inf_{i \in M} N_i$ simultaneously go to infinity.

Proposition 3.53 (Kingman algebraic two-level measure tree). *For $M \in \mathbb{N}$ and $\underline{N} \in \mathbb{N}^M$, let $\chi^{M, \underline{N}} = (T, c, \nu^{M, \underline{N}})$. Then there exists a random binary a2m tree $\chi \in \mathbb{T}_2^{(2)}$ such that*

$$\chi^{M, \underline{N}} \xrightarrow[M, \underline{N} \rightarrow \infty]{} \chi, \quad (3.104)$$

where \Rightarrow stands for weak convergence on $\mathbb{T}_2^{(2)}$ equipped with the two-level sample shape convergence.

Proof. By Proposition 3.49, the space $\mathbb{T}_2^{(2)}$ equipped with the two-level sample shape convergence is compact. Therefore the sequence $\{\chi^{M, \underline{N}} : M \in \mathbb{N}, \underline{N} \in \mathbb{N}^M\}$ is clearly tight.

The *uniqueness* of the limit follows from the sampling consistency of the family of finite nested Kingman coalescents. Fix $M \in \mathbb{N}$ and $\underline{N} \in \mathbb{N}^M$, and $m \in \mathbb{N}$, $\underline{n} \in \mathbb{N}^m$ such that $m \leq M$ and for each $1 \leq i \leq m$, $n_i \leq N_i$. Consider $\{U_{ij} : (i, j) \in \mathbb{N}_{m, \underline{n}}\}$ sampled (without repetition) at random from $\chi^{M, \underline{N}}$. By Proposition 3.51, the restriction of $\mathcal{K}^{\mathbb{N}_{M, \underline{N}}}$ to $\{U_{ij} : (i, j) \in \mathbb{N}_{m, \underline{n}}\}$ is a finite nested Kingman coalescent on $\{U_{ij} : (i, j) \in \mathbb{N}_{m, \underline{n}}\}$. Therefore, the shape $\mathbf{s}_{(T, c)}(\underline{U}) \in \mathfrak{C}_{m, \underline{n}}$ of the subtree spanned by the leaves $\{U_{ij} : (i, j) \in \mathbb{N}_{m, \underline{n}}\}$ has the distribution of the shape of $\chi^{m, \underline{n}}$. Since sampling with and without repetition is asymptotically equivalent when $M, \underline{N} \rightarrow \infty$, the claim follows by definition of the two-level sample shape convergence. \square

The construction of the Kingman algebraic two-level measure tree was one of the very original goals of this thesis.

Definition 3.54 (Kingman algebraic two-level measure tree). The *Kingman algebraic two-level measure tree* χ with rates (γ_H, γ_P) is the unique limit in $\mathbb{T}_2^{(2)}$ of the sequence $(\chi^{M, \underline{N}})_{M, \underline{N}}$, where $\chi^{M, \underline{N}}$ is the random algebraic two-level measure tree obtained from the finite nested Kingman coalescent on $\mathbb{N}_{M, \underline{N}}$ with rates (γ_H, γ_P) .

Results on the subtree mass distribution of the rooted α -Ford trees

In this paper, we have mainly worked with the subtree mass distribution of (unrooted) algebraic measure trees by looking at the quenched law of the vector $\underline{\eta}(\underline{u})$ of the three masses of the components connected to the branch point $c(\underline{u})$ of a sample of size three $\underline{u} = (u_1, u_2, u_3)$. Here we state analogous results for *rooted* algebraic measure trees.

Recall from Definition 1.13 the definition of rooted algebraic measure tree (T, c_\wedge, ρ, μ) . For such a tree, we can still evaluate the subtree mass distribution by sampling three leaves according to μ . Their branch point (in the sense of unrooted trees) is given by $v = \max\{c_\wedge(u_1, u_2), c_\wedge(u_1, u_3), c_\wedge(u_2, u_3)\}$ and we can consider the quenched law of the vector $\underline{\eta}(\underline{u}) = (\eta_0(v), \eta_1(v), \eta_2(v))$ of the three component masses, where η_0 is for the component containing the root. Nevertheless, the minimum map also associates a branch point $w = c_\wedge(u_1, u_2)$ to only two sampled points u_1, u_2 . Therefore, we can also consider the quenched law of the vector $\underline{\eta}(u_1, u_2) = (\eta_0(w), \eta_1(w), \eta_2(w))$. For this reason, we state here results in both cases, that is, when we sample two or three leaves.

We denote by $\chi^{\alpha, \rho}$ the rooted α -Ford algebraic measure tree defined in Remark 1.19.

Sample of 2 leaves: we first consider polynomials that evaluate the vector

$$\underline{\eta}(u_1, u_2) := (\eta_0(w), \eta_1(w), \eta_2(w)), \quad (\text{A.1})$$

where $w = c_\wedge(u_1, u_2)$ is the minimum of two leaves sampled according to μ . For $\alpha = 0$ and $\alpha = \frac{1}{2}$, the subtree mass distribution of the algebraic measure α -Ford tree can be explicitly expressed through products of independent random variables.

Proposition A.1 (Subtree mass distribution in the rooted Kingman algebraic measure tree). *Let Y and $B_{2,2}$ be two independent random variables, such that Y has law $\frac{1}{3}\delta_1 + \frac{2}{3}U_{[0,1]}$ with $U_{[0,1]}$ uniformly distributed on $[0, 1]$, and $B_{2,2}$ has law $\text{Beta}(2, 2)$. For all $f: \Delta_2 \rightarrow \mathbb{R}$ continuous and bounded,*

$$\mathbb{E} \left[\int_{(T_{\chi^{0, \rho}})^2} (\mu_{\chi^{0, \rho}})^{\otimes 2}(\underline{du}) f \left(\underline{\eta}_{\chi^{0, \rho}}(\underline{u}) \right) \right] = \mathbb{E} \left[f(1 - Y, Y B_{2,2}, Y(1 - B_{2,2})) \right]. \quad (\text{A.2})$$

Proposition A.2 (Subtree mass distribution in the rooted algebraic Brownian CRT). *Let $B_{1, \frac{1}{2}}$ and $B_{\frac{1}{2}, \frac{1}{2}}$ be two independent beta random variables, such that $B_{1, \frac{1}{2}}$ has law $\text{Beta}(1, \frac{1}{2})$ and $B_{\frac{1}{2}, \frac{1}{2}}$ has law $\text{Beta}(\frac{1}{2}, \frac{1}{2})$. For all $f: \Delta_2 \rightarrow \mathbb{R}$ continuous and bounded,*

$$\mathbb{E} \left[\int_{(T_{\chi^{\frac{1}{2}, \rho}})^2} (\mu_{\chi^{\frac{1}{2}, \rho}})^{\otimes 2}(\underline{du}) f \left(\underline{\eta}_{\chi^{\frac{1}{2}, \rho}}(\underline{u}) \right) \right] = \mathbb{E} \left[f(1 - B_{1, \frac{1}{2}}, B_{1, \frac{1}{2}} B_{\frac{1}{2}, \frac{1}{2}}, B_{1, \frac{1}{2}}(1 - B_{\frac{1}{2}, \frac{1}{2}})) \right]. \quad (\text{A.3})$$

In the general case $\alpha \in [0, 1]$, we have a recurrence relation for the first order moments. For $\underline{k} = (k_0, k_1, k_2) \in \mathbb{N}_0^3$, define $f^{\underline{k}} : \Delta_2 \rightarrow [0, 1]$ by

$$f^{\underline{k}}(\underline{\eta}) = \eta_1^{k_0} \eta_2^{k_1} \eta_3^{k_2}. \quad (\text{A.4})$$

Obviously, $\mathbb{E}[\Phi^{f^{(0,0,0)}}(\chi^\alpha)] = 1$ and for all $k \in \mathbb{N}$,

$$\mathbb{E} \left[\Phi^{f^{(k,0,0)}}(\chi^\alpha) \right] = \frac{1}{(k+2)(k+1-\alpha)} \left(k(k-1+\alpha) \mathbb{E} \left[\Phi^{f^{(k-1,0,0)}}(\chi^\alpha) \right] + 2(1-\alpha) \right) \quad (\text{A.5})$$

Therefore, for all $\alpha \in (0, 1)$,

$$\mathbb{E} \left[\Phi^{f^{(k,0,0)}}(\chi^\alpha) \right] = \frac{2(1-\alpha)k + \frac{\Gamma(1-\alpha)\Gamma(k+\alpha)}{\Gamma(\alpha-2)\Gamma(k+2-\alpha)} + 4 - 3\alpha}{(3-2\alpha)(k+1)(k+2)}. \quad (\text{A.6})$$

Sample of 3 leaves: we now consider polynomials that evaluate the vector

$$\underline{\eta}(u_1, u_2, u_3) := (\eta_0(v), \eta_1(v), \eta_2(v)), \quad (\text{A.7})$$

where $v = \max\{c_\wedge(u_1, u_2), c_\wedge(u_1, u_3), c_\wedge(u_2, u_3)\}$ is the branch point of three leaves sampled according to μ . For $\alpha = 0$ and $\alpha = \frac{1}{2}$, the subtree mass distribution of the algebraic measure α -Ford tree can be explicitly expressed through products of independent Beta random variables.

Proposition A.3 (Subtree mass distribution in the rooted Kingman algebraic measure tree). *Let $B_{1,2}$ and $B_{2,2}$ be two independent beta random variables, such that $B_{1,2}$ has law $\text{Beta}(1, 2)$ and $B_{2,2}$ has law $\text{Beta}(2, 2)$. For all $f : \Delta_2 \rightarrow \mathbb{R}$ continuous and bounded,*

$$\mathbb{E} \left[\int_{(T_{\chi^{0,\rho}})^3} (\mu_{\chi^{0,\rho}})^{\otimes 3}(\underline{du}) f \left(\underline{\eta}_{\chi^{0,\rho}}(\underline{u}) \right) \right] = \mathbb{E} \left[f(1 - B_{1,2}, B_{1,2}B_{2,2}, B_{1,2}(1 - B_{2,2})) \right]. \quad (\text{A.8})$$

Proposition A.4 (Subtree mass distribution in the rooted algebraic Brownian CRT). *Let $B_{1,\frac{3}{2}}$ and $B_{\frac{1}{2},\frac{1}{2}}$ be two independent beta random variables, such that $B_{1,\frac{3}{2}}$ has law $\text{Beta}(1, \frac{3}{2})$ and $B_{\frac{1}{2},\frac{1}{2}}$ has law $\text{Beta}(\frac{1}{2}, \frac{1}{2})$. For all $f : \Delta_2 \rightarrow \mathbb{R}$ continuous and bounded,*

$$\mathbb{E} \left[\int_{(T_{\chi^{\frac{1}{2},\rho}})^3} (\mu_{\chi^{\frac{1}{2},\rho}})^{\otimes 3}(\underline{du}) f \left(\underline{\eta}_{\chi^{\frac{1}{2},\rho}}(\underline{u}) \right) \right] = \mathbb{E} \left[f(1 - B_{1,\frac{3}{2}}, B_{1,\frac{3}{2}}B_{\frac{1}{2},\frac{1}{2}}, B_{1,\frac{3}{2}}(1 - B_{\frac{1}{2},\frac{1}{2}})) \right]. \quad (\text{A.9})$$

In the general case $\alpha \in [0, 1]$, we again have a recurrence relation for the first order moments. For $\underline{k} = (k_0, k_1, k_2) \in \mathbb{N}_0^3$, define $f^{\underline{k}} : \Delta_2 \rightarrow [0, 1]$ by

$$f^{\underline{k}}(\underline{\eta}) = \eta_1^{k_0} \eta_2^{k_1} \eta_3^{k_2}. \quad (\text{A.10})$$

Obviously, $\mathbb{E}[\Phi^{f^{(0,0,0)}}(\chi^\alpha)] = 1$ and for all $k \in \mathbb{N}$,

$$\mathbb{E} \left[\Phi^{f^{(k,0,0)}}(\chi^\alpha) \right] = \frac{1}{(k+3)(k+2-\alpha)} \left((k+1)(k+\alpha) \mathbb{E} \left[\Phi^{f^{(k-1,0,0)}}(\chi^\alpha) \right] + 6(1-\alpha) \right) \quad (\text{A.11})$$

Hence, for all $\alpha \in (0, 1)$,

$$\mathbb{E} \left[\Phi^{f^{(k,0,0)}}(\chi^\alpha) \right] = \frac{3 \left(2(1-\alpha)k + \frac{\Gamma(1-\alpha)\Gamma(k+1+\alpha)}{\Gamma(\alpha-2)\Gamma(k+3-\alpha)} + 6 - 5\alpha \right)}{(3-2\alpha)(k+2)(k+3)}. \quad (\text{A.12})$$

One can prove the above results by constructing the rooted analog of the α -Ford diffusion, for which the rooted α -Ford algebraic measure tree is an invariant distribution.

Bibliography

- [Ald93] David Aldous. The continuum random tree III. *Ann. Probab.*, 21:248–289, 1993. [x](#), [xi](#), [12](#)
- [Ald94a] David Aldous. Recursive self-similarity for random trees, random triangulations and Brownian excursion. *Ann. Probab.*, 22(2):527–545, 1994. [xiv](#), [xx](#), [2](#), [12](#), [67](#)
- [Ald94b] David Aldous. Triangulating the circle, at random. *The American Mathematical Monthly*, 101(3):223–233, 1994. [xx](#), [67](#)
- [Ald96] David Aldous. Probability distributions on cladograms. *Random and Discrete Structures*, 76:1–18, 1996. [xii](#), [xxiii](#), [1](#)
- [Ald00] David Aldous. Mixing time for a Markov chain on cladograms. *Combinatorics, Probability and Computing*, 9:191–204, 2000. [xv](#), [3](#), [17](#)
- [ALW16] Siva Athreya, Wolfgang Löhr, and Anita Winter. The gap between Gromov-vague and Gromov–Hausdorff-vague topology. *Stochastic Processes and their Applications*, 126(9):2527–2553, 2016. [xi](#)
- [ALW17] Siva Athreya, Wolfgang Löhr, and Anita Winter. Invariance principle for variable speed random walks on trees. *Ann. Probab.*, 45(2):625–667, 2017. [65](#), [71](#)
- [Ban08] V. Bansaye. Proliferating parasites in dividing cells : Kimmel’s branching model revisited. *Annals of Applied Probability*, 2008. [33](#)
- [BBS08] Julien Berestycki, Nathanaël Berestycki, and Jason Schweinsberg. Small-time behavior of beta coalescents. *Annales de l’I.H.P. Probabilités et statistiques*, 44(2):214–238, 2008. [51](#)
- [BDLS18] Airam Blancas Benítez, Jean-Jil Duchamps, Amaury Lambert, and Arno Siri-Jégousse. Trees within Trees: Simple Nested Coalescents. *Electron. J. Probab.*, 23:1–27, 2018. [xix](#), [35](#), [61](#), [93](#)
- [Ber09] Nathanaël Berestycki. *Recent progress in coalescent theory*, volume 16 of *Ensaaios Matemáticos [Mathematical Surveys]*. Sociedade Brasileira de Matemática, Rio de Janeiro, 2009. [51](#)
- [BGK⁺21] A. Blancas, S. Gufler, S. Kliem, V.C. Tran, and A. Wakolbinger. Evolving genealogies for branching populations under selection and competition. *arXiv:2106.16201*, 2021. [33](#)
- [BJ10] M.G.B. Blum and M. Jakobsson. Deep Divergences of Human Gene Trees and Models of Human Origins. *Molecular Biology and Evolution*, 28(2):889–898, 10 2010. [xvii](#), [33](#)
- [BRSSJ19] Airam Blancas Benítez, Tim Rogers, Jason Schweinsberg, and Arno Siri-Jégousse. The Nested Kingman Coalescent: Speed of Coming Down from Infinity. *Annals of Applied Probability*, 29(3):1808–1836, February 2019. 24 pages. [xix](#), [35](#)

- [BS15] Nicolas Broutin and Henning Sulzbach. The dual tree of a recursive triangulation of the disk. *The Annals of Probability*, 43(2):738–781, 2015. [xx](#)
- [BT11] V. Bansaye and V.C. Tran. Branching Feller diffusion for cell division with parasite infection. *ALEA*, 8:95–127, 2011. [xvii](#), [33](#)
- [BvH99] Ellen Baake and Arndt von Haeseler. Distance measures in terms of substitution processes. *Theor. Popul. Biol.*, 55:166–175, 1999. [2](#)
- [CFW09] Bo Chen, Daniel J. Ford, and Matthias Winkel. A new family of Markov branching trees: the alpha-gamma model. *Electronic Journal of Probability*, 14:400–430, 2009. [xii](#), [1](#)
- [CHK14] Nicolas Curien, Bénédicte Haas, and Igor Kortchemski. The CRT is the scaling limit of random dissections. *Random Structures and Algorithms*, 47, 06 2014. [xx](#)
- [CK13] Nicolas Curien and Igor Kortchemski. Percolation on random triangulations and stable looptrees. *Probability Theory and Related Fields*, 163, 07 2013. [xx](#)
- [CLG11] Nicolas Curien and Jean-François Le Gall. Random recursive triangulations of the disk via fragmentation theory. *Ann. Prob.*, 39, 06 2011. [xx](#)
- [CMR18] Tomás M. Coronado, Arnau Mir, and Francesc Rosselló. The probabilities of trees and cladograms under Ford’s α -model. *The Scientific World Journal*, (1916094), 2018. [1](#)
- [Cur14] Nicolas Curien. Dissecting the circle, at random. *ESAIM: Proceedings*, 44:129–139, 01 2014. [xx](#)
- [Daw93] D. A. Dawson. Measure-valued Markov processes. In P.L. Hennequin, editor, *École d’Été de probabilités de Saint-Flour XXI—1991*, volume 1541 of *Lecture Notes in Mathematics*, pages 1–260, Berlin, 1993. Springer. [xvii](#), [39](#)
- [Daw18] D. Dawson. Multilevel mutation-selection systems and set-valued duals. *J. Math. Biol.*, 76:295–378, 2018. [xvii](#), [33](#), [34](#), [48](#), [66](#)
- [DDSJ08] J.-F. Delmas, J.-S. Dhersin, and A. Siri-Jégousse. Asymptotic results on the length of coalescent trees. *Ann. Appl. Probab.*, 18(3):997–1025, 2008. [51](#)
- [DDV95] K.J. Hochberg D.A. Dawson and V. Vinogradov. On path properties of super-2 processes II. In M.G. Cranston and M.A. Pinsky, editors, *Proceedings of Symposia in Pure Mathematics Series*, volume 57, pages 385–403. AMS, Providence, 1995. [33](#)
- [DFSJY13] J.-S. Dhersin, F. Freund, A. Siri-Jégousse, and L. Yuan. On the length of an external branch in the beta-coalescent. *Stochastic Process. Appl.*, 123(5):1691–1715, 2013. [51](#)
- [DGP11] Andrej Depperschmidt, Andreas Greven, and Peter Pfaffelhuber. Marked metric measure spaces. *Electronic Communications in Probability*, 16(none):174 – 188, 2011. [xi](#)
- [DGW04] Donald Dawson, Luis Gorostiza, and Anton Wakolbinger. Hierarchical Equilibria of Branching Populations. *Electronic Journal of Probability*, 9(none):316 – 381, 2004. [xvii](#), [33](#), [66](#)
- [DH91] D. Dawson and K.J. Hochberg. A multilevel branching model. *Adv. Appl. Prob.*, 23:701–715, 1991. [33](#)
- [DHV96] Donald A. Dawson, Kenneth J. Hochberg, and Vladimir Vinogradov. High-density limits of hierarchically structured branching-diffusing populations. *Stochastic Processes and their Applications*, 62(2):191–222, July 1996. [xvii](#), [33](#)
- [DHW90] D.A. Dawson, K.J. Hochberg, and Y. Wu. *White Noise Analysis: Mathematics and Applications*, chapter Multilevel branching systems. World Scientific Publ., 1990. [xvii](#), [33](#), [66](#)

-
- [DIMR07] M. Drmota, A. Iksanov, M. Moehle, and U. Roesler. Asymptotic results concerning the total branch length of the Bolthausen-Sznitman coalescent. *Stochastic Process. Appl.*, 117(10):1404–1421, 2007. [51](#)
- [DK96] Peter Donnelly and Thomas G. Kurtz. A countable representation of the Fleming-Viot processes. *Ann. Probab.*, 24(2):698–742, 1996. [33](#), [48](#)
- [DK99a] P. Donnelly and T.G. Kurtz. Genealogical processes for Fleming-Viot models with selection and recombination. *Annals of Applied Probability*, 9(4):1091–1148, 1999. [33](#), [41](#)
- [DK99b] P. Donnelly and T.G. Kurtz. Particle representations for measure-valued population models. *Annals of Probability*, 27(1):166–205, 1999. [33](#)
- [DKL20] Giorgin Daphné, Sarah Kaakai, and Vincent Lemaire. IBMPopSim package description. <https://daphnegiorgi.github.io/IBMPopSim/>, 2020. [42](#)
- [DP91] D. A. Dawson and E. A. Perkins. Historical processes. *Mem. Amer. Math. Soc.*, 93(454):iv+179, 1991. [34](#)
- [Dre84] Andreas W.M. Dress. Trees, tight extensions of metric spaces, and the cohomological dimension of certain groups: A note on combinatorial properties of metric spaces. *Adv. Math.*, 53:321–402, 1984. [39](#)
- [Duc20] J.-J. Duchamps. Trees within trees II: nested fragmentations. *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques*, 56(2):1203–1229, 2020. [35](#)
- [EK86] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence*. John Wiley, New York, 1986. [18](#), [20](#), [23](#), [24](#), [50](#), [61](#), [63](#), [64](#), [90](#)
- [EK93] S.N. Ethier and Thomas G. Kurtz. Fleming-Viot processes in population genetics. *SIAM J. Contr. Optim.*, 31:345–386, 1993. [xvii](#)
- [Eth93] A.M. Etheridge. Limiting behaviour of two-level measure-branching. *Adv. Appl. Prob.*, 25(4):773–782, 1993. [33](#)
- [Eva00] Steven N. Evans. Kingman’s coalescent as a random metric space. In *Stochastic Models: Proceedings of the International Conference on Stochastic Models in Honour of Professor Donald A. Dawson, Ottawa, Canada, June 10-13, 1998 (L.G Gorostiza and B.G. Ivanoff eds.)*, Canadian Mathematical Society, 2000. [ix](#), [34](#), [41](#)
- [Eva08] Steven N. Evans. *Probability and real trees*, volume 1920 of *Lecture Notes in Mathematics*. Springer, Berlin, 2008. Lectures from the 35th Summer School on Probability Theory held in Saint-Flour, July 6–23, 2005. [39](#)
- [EW06] Steven N. Evans and Anita Winter. Subtree prune and re-graft: A reversible real-tree valued Markov chain. *Ann. Prob.*, 34(3):918–961, 2006. [38](#)
- [Fel71] William Feller. *An introduction to probability theory and its applications. Vol. II*. John Wiley & Sons, Inc., New York-London-Sydney, second edition, 1971. [51](#)
- [Fel05] Joseph Felsenstein. *Inferring Phylogenies*. Sinauer Assoc., 2005. [2](#)
- [For] Daniel J. Ford. Probabilities on cladograms: introduction to the alpha model. [arXiv:math/0511246](https://arxiv.org/abs/math/0511246). [iii](#), [v](#), [vii](#), [xi](#), [xii](#), [1](#), [11](#), [14](#), [19](#)
- [FPRWa] Noah Forman, Soumik Pal, Douglas Rizzolo, and Matthias Winkel. Aldous diffusion I: a projective system of continuum k -tree evolutions. [arXiv:1809.07756](https://arxiv.org/abs/1809.07756). [xv](#), [3](#)
- [FPRWb] Noah Forman, Soumik Pal, Douglas Rizzolo, and Matthias Winkel. Interval partition evolutions with emigration related to the Aldous diffusion. [arXiv:1804.01205v1](https://arxiv.org/abs/1804.01205v1). [xv](#), [3](#)
-

- [FPRW20] Noah Forman, Soumik Pal, Douglas Rizzolo, and Matthias Winkel. Projections of the Aldous chain on binary trees: Intertwining and consistency. *Random Structures & Algorithms*, 57(3):745–769, 2020. [xv](#), [3](#)
- [FPRW21] Noah Forman, Soumik Pal, Douglas Rizzolo, and Matthias Winkel. Diffusions on a space of interval partitions: Poisson–Dirichlet stationary distributions. *The Annals of Probability*, 49(2):793 – 831, 2021. [xv](#), [3](#)
- [Fre77] David Freedman. A Remark on the Difference between Sampling with and without Replacement. *Journal of the American Statistical Association*, 72:681, 1977. [47](#)
- [FT09] R. Ferrière and V.C. Tran. Stochastic and deterministic models for age-structured populations with genetically variable traits. *ESAIM: Proceedings*, 27:289–310, 2009. [42](#)
- [Fuk87] K. Fukaya. Collapsing of Riemannian manifolds and eigenvalues of Laplace operator. *Inventiones mathematicae*, 87:517–548, 1987. [x](#)
- [FV79] Wendell H. Fleming and Michel Viot. Some measure-valued Markov processes in population genetics theory. *Indiana University Mathematics Journal*, 28(5):817–843, 1979. [xvii](#), [39](#)
- [GdHKK14] A. Greven, F. den Hollander, S. Kliem, and A. Klimovsky. Renormalisation of hierarchically interacting Lambda-Cannings processes. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 11(1):43–140, 2014. [33](#)
- [GH00] Andreas Greven and Kenneth Hochberg. New behavioral patterns for two-level branching systems. *Canadian Mathematical Society Conference Proceedings*, 6:205–215, 01 2000. [xvii](#), [33](#), [66](#)
- [GHW95] Luis G. Gorostiza, Kenneth J. Hochberg, and Anton Wakolbinger. Persistence of a critical super-2 process. *Journal of Applied Probability*, 32(2):534–540, 1995. [xvii](#), [33](#), [66](#)
- [GJB13] L. M. Gattepaille, M. Jakobsson, and M. G. B. Blum. Inferring population size changes with sequence and SNP data: lessons from human bottlenecks. *Heredity*, 110(5):409–419, May 2013. [xxiii](#)
- [Glö12] Patrick Karl Glöde. *Dynamics of genealogical trees for autocatalytic branching processes*. PhD thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg, 2012. [xi](#)
- [GLW05] Andreas Greven, Vlada Limic, and Anita Winter. Representation Theorems for Interacting Moran Models, Interacting Fisher-Wrighter Diffusions and Applications. *Electronic Journal of Probability*, 10(none):1286 – 1358, 2005. [34](#)
- [GPW09] Andreas Greven, Peter Pfaffelhuber, and Anita Winter. Convergence in distribution of random metric measure spaces (Λ -coalescent measure trees). *Probab. Theo. Rel. Fields*, 145:285–322, 2009. [x](#), [xi](#), [7](#), [34](#), [38](#), [41](#), [66](#), [73](#)
- [GPW13] Andreas Greven, Peter Pfaffelhuber, and Anita Winter. Tree-valued resampling dynamics: Martingale problems and applications. *Probab. Theo. Rel. Fields*, 155(3–4):789–838, 2013. [iii](#), [v](#), [vii](#), [xvii](#), [xix](#), [xxiii](#), [34](#), [35](#), [39](#), [48](#), [51](#), [56](#)
- [Gro99] Misha Gromov. *Metric structures for Riemannian and non-Riemannian spaces*, volume 152 of *Progress in Mathematics*. Birkhäuser Boston Inc., Boston, MA, 1999. [x](#), [38](#)
- [GSW16] Andreas Greven, Rongfeng Sun, and Anita Winter. Continuum space limit of the genealogies of interacting Fleming-Viot processes on \mathbb{Z} . *Electronic Journal of Probability*, 21(58):1–64, 2016. [xi](#)
- [Guf18] S. Gufler. Pathwise construction of tree-valued Fleming-Viot processes. *Electronic J. Probab.*, 23:1–58, 2018. [41](#)

-
- [HMPW08] Bénédicte Haas, Grégory Miermont, Jim Pitman, and Matthias Winkel. Continuum tree asymptotics of discrete fragmentations and applications to phylogenetic models. *The Annals of Probability*, 36:1790–1837, 2008. [xii](#), [1](#)
- [HS21] Bénédicte Haas and Robin Stephenson. Scaling limits of multi-type Markov Branching trees. *Probability Theory and Related Fields*, 180(3):727–797, 2021. [xxiv](#)
- [JBA19] F. Jay, S. Boitard, and F. Austerlitz. An ABC method for whole-genome sequence data: inferring Paleolithic and Neolithic human expansions. *Molecular Biology and Evolution*, 36:1565–1579, 2019. [xvii](#), [33](#)
- [Kim69] M. Kimura. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics*, 61:893–903, 1969. [51](#)
- [Kim97] M. Kimmel. Quasistationarity in a branching model of division-within-division. In *Classical and modern branching processes (Minneapolis, MN, 1994)*, volume 84 of *IMA Vol. Math. Appl.*, pages 157–164, New-York, 1997. Springer. [xvii](#), [33](#)
- [Kin82] J. F. C. Kingman. The coalescent. *Stochastic Process. Appl.*, 13(3):235–248, 1982. [ix](#), [14](#)
- [Kle08] Achim Klenke. *Probability theory*. Universitext. Springer-Verlag London, Ltd., London, 2008. A comprehensive course, translated from the 2006 German original. [38](#)
- [KS03] Kazuhiro Kuwae and Takashi Shioya. Convergence of spectral structures: A functional analytic theory and its applications to spectral geometry. *Communications in Analysis and Geometry*, 11:599–673, 09 2003. [x](#)
- [KW19] Sandra Kliem and Anita Winter. Evolving phylogenies of trait-dependent branching with mutation and competition, Part I: Existence. *Stochastic Processes and their Applications*, 129(12):4837–4877, 2019. [xi](#)
- [LBP⁺21] C. Lepers, S. Billiard, M. Porte, S. Méléard, and V.C. Tran. Inference with selection, varying population size and evolving population structure: Application of abc to a forward-backward. *Heredity*, 126:335–350, 2021. [xvii](#), [33](#)
- [LMW20] Wolfgang Löhr, Leonid Mytnik, and Anita Winter. The Aldous chain on cladograms in the diffusion limit. *The Annals of Probability*, 48:2565–2590, 09 2020. [iii](#), [v](#), [vii](#), [xiv](#), [xv](#), [xvi](#), [xxiii](#), [2](#), [3](#), [12](#), [17](#), [20](#), [23](#), [25](#), [26](#), [66](#)
- [Löh13] Wolfgang Löhr. Equivalence of Gromov-Prohorov- and Gromov’s \square_λ -metric on the space of metric measure spaces. *Electronic Communications in Probability*, 18(none):1 – 10, 2013. [xi](#)
- [LS20] Amaury Lambert and Emmanuel Schertzer. Coagulation-transport equations and the nested coalescents. *Probab. Theory Relat. Fields*, 176:77–147, 2020. [xix](#), [35](#)
- [LW21] Wolfgang Löhr and Anita Winter. Spaces of algebraic measure trees and triangulations of the circle. *Bulletin de la Société Mathématique de France*, 149(1):1–63, 2021. [iii](#), [v](#), [vii](#), [xi](#), [xii](#), [xiv](#), [xix](#), [xx](#), [xxiii](#), [1](#), [2](#), [5](#), [6](#), [8](#), [9](#), [10](#), [11](#), [29](#), [65](#), [66](#), [67](#), [68](#), [71](#), [72](#), [73](#), [75](#), [76](#), [77](#), [80](#), [81](#), [83](#), [90](#), [92](#)
- [Mö06] M. Möhle. On the number of segregating sites for populations with large family sizes. *Adv. in Appl. Probab.*, 38(3):750–767, 2006. [51](#)
- [Mei19] Roland Meizis. Convergence of metric two-level measure spaces. *Stochastic Processes and their Applications*, 130, 10 2019. [iii](#), [v](#), [vii](#), [xi](#), [xvii](#), [xix](#), [xx](#), [xxii](#), [34](#), [35](#), [36](#), [38](#), [45](#), [52](#), [57](#), [64](#), [66](#), [67](#), [68](#), [72](#), [73](#), [74](#), [75](#), [81](#), [84](#), [92](#)
- [MNO92] John C. Mayer, Jacek Nikiel, and Lex G. Oversteegen. Universal spaces for \mathbb{R} -trees. *Transactions of the American Mathematical Society*, 334(1):411–432, 1992. [65](#)
-

- [MO90] John Mayer and Lex Oversteegen. A topological characterization of \mathbb{R} -trees. *Trans. Amer. Math. Soc.*, 320:395–415, 07 1990. [65](#)
- [Mor58] P. A. P. Moran. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society*, 54(1):60–71, 1958. [xvi](#), [34](#)
- [MR13] Sylvie Méléard and Sylvie Roelly. Evolutive two-level population process and large population approximations. *Ann. Univ. Buchar. Math. Ser.*, 4(LXII)(1):37–70, 2013. [xvii](#), [33](#)
- [NTW] Josué Nussbaumer, Viet Chi Tran, and Anita Winter. Algebraic two-level measure trees. Manuscript in prep. [xix](#)
- [Nus] Josué Nussbaumer. Resampling dynamics on metric two-level measure trees. Manuscript in prep. [xvi](#), [66](#)
- [NW] Josué Nussbaumer and Anita Winter. The algebraic alpha-Ford tree under evolution. arXiv:2006.09316. [xii](#), [66](#), [70](#)
- [OW20] L. Osorio and A. Winter. Two level branching model for virus population under cell division. arXiv:2004.14352, 2020. [33](#)
- [PW09] Jim Pitman and Matthias Winkel. Regenerative tree growth: binary self-similar continuum random trees and Poisson-Dirichlet compositions. *The Annals of probability*, 37:1999–2041, 2009. [xii](#), [1](#)
- [Sch01] Jason Schweinsberg. An $O(N^2)$ bound for the relaxation time of a markov chain on cladograms. *Random Struct. Alg.*, 20:59–70, 2001. [xv](#), [3](#), [17](#)
- [SSV15] Raazesh Sainudiin, Tanja Stadler, and Amandine Véber. Finding the best resolution for the Kingman–Tajima coalescent: theory and applications. *Journal of Mathematical Biology*, 70(6):1207–1247, May 2015. [xxiii](#)
- [Ste09] Sigurður Örn Stefánsson. The infinite volume limit of Ford’s alpha model. *Acta Phys. Polon. B*, 3:555–562, 2009. [1](#)
- [STH⁺11] Raazesh Sainudiin, Kevin Thornton, Jennifer Harlow, James Booth, Michael Stillman, Ruriko Yoshida, Robert Griffiths, Gil McVean, and Peter Donnelly. Experiments with the Site Frequency Spectrum. *Bulletin of Mathematical Biology*, 73(4):829–872, Apr 2011. [xxiii](#)
- [SV18] Raazesh Sainudiin and Amandine Véber. Full likelihood inference from the site frequency spectrum based on the optimal tree resolution. *Theoretical Population Biology*, 124:1–15, 2018. [xxiii](#)
- [Ter97] W.F. Terhalle. \mathbb{R} -trees and symmetric differences of sets. *Europ. J. Combinatorics*, 18:825–833, 1997. [39](#)
- [Tit77] Jacques Tits. A “theorem of Lie-Kolchin” for trees. *Contributions to Algebra: A Collection of Papers Dedicated to Ellis Kolchin*, 12 1977. [65](#)
- [VAE⁺09] P. Verdu, F. Austerlitz, A. Estoup, R. Vitalis, M. Georges, S. Théry, A. Froment, S. Le Bomin, A. Gessain, J.-M. Hombert, L. Van der Veen, L. Quintana-Murci, S. Bahuchet, and E. Heyer. Origins and genetic diversity of Pygmy hunter-gatherers from Western Central Africa. *Current Biology*, 19(4):312–318, 2009. [xvii](#), [33](#)
- [Wat75] G. A. Watterson. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.*, 7:256–276, 1975. [x](#), [xxiii](#), [2](#), [51](#)

- [Wu91] Yadong Wu. *Dynamic Particle Systems and Multilevel Measure Branching Processes*. PhD thesis, Carleton University, Ottawa, Canada, 1991. [xvii](#), [33](#), [66](#)
- [Wu94] Y. Wu. Asymptotic behavior of the two-level measure branching process. *Annals of Probab.*, 22(2):854–874, 1994. [33](#)