Electron. Commun. Probab. **18** (2013), no. 17, 1–10.
DOI: 10.1214/ECP.v18-2268
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Equivalence of Gromov-Prohorov- and Gromov's $\underline{\Box}_\lambda$-metric on the space of metric measure spaces

Wolfgang Löhr*

Abstract

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

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

Keywords: space of metric measure spaces ; Gromov-Prohorov metric ; Gromov's box-metric ; Gromov-weak topology ; real tree ; coding trees by excursions ; Lusin topology on excursions.
AMS MSC 2010: NA.
Submitted to ECP on August 23, 2012, final version accepted on February 28, 2013.
Supersedes arXiv:1111.5837v3.

1 Introduction

Tree-valued stochastic processes frequently appear in probability theory and its application areas, such as theoretical biology. For instance, in an evolutionary model, the development of the genealogical tree is of interest. In the continuum limit of infinite population size, the finite tree becomes a continuum tree ($\mathbb{R}$-tree) and the normalised counting measure of individuals becomes a probability measure on it. This measure is needed to describe the population density on the tree and to sample individuals from it. See Aldous' seminal paper [3] for the convergence of finite variance Galton-Watson trees to a (Brownian) continuum measure tree, and results of Duquesne and Le Gal ([13, 11]) for the convergence of infinite variance Galton-Watson trees to Lévy trees.

More generally than $\mathbb{R}$-trees, we can considers random metric (probability) measure spaces, an approach introduced by Greven, Pfaffelhuber and Winter in [17] and applied by the authors and Depperschmidt to obtain tree-valued Fleming-Viot dynamics in [18,

*University of Duisburg-Essen. E-mail: wolfgang.loehr@uni-due.de

9]. Here, $\mathcal{X} = (X, d, \mu)$ is a **metric measure space** (**mm-space**) if (X, d) is a complete, separable metric space and μ a probability measure on the Borel σ-algebra of X. To work with mm-space valued processes, it is crucial to have an appropriate topology on the set of mm-spaces, or rather the set $\mathfrak{X}$ of isometry classes of mm-spaces. A fruitful topology is given by the Gromov-weak topology introduced in [17]. In the same paper, the authors conjectured that it coincides with the topology induced by Gromov's metric $\underline{\square}_1$, which is defined in [19, Chapter $3\frac{1}{2}$]. They also introduced a complete metric, the Gromov-Prohorov metric d_{GP}, that metrises the Gromov-weak topology.

Here, we show that $\underline{\square}_1$ and d_{GP} are bi-Lipschitz equivalent, which in particular implies that the conjecture is true and $\underline{\square}_1$ indeed metrises Gromov-weak topology. Furthermore, we use this result to prove that the measure $\mathbb{R}$-tree coded by an excursion depends continuously on the excursion. To this end, we consider two topologies on the space of lower semi-continuous excursions. For the uniform topology, Lipschitz continuity is already shown by Abraham, Delmas and Hoscheit in [1, Prop. 2.9] (with their metric on trees, which implies the result for ours), but we obtain a much shorter proof using the equivalence of d_{GP} and $\underline{\square}_1$. The uniform topology has the disadvantage of being non-separable, therefore we introduce a new, weaker, separable, metrisable topology, which is Lusin on the subset of bounded excursions. We also show continuous dependence of the tree on the excursion in this weaker topology.

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

2 Metrics and topologies on the space of mm-spaces

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

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

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

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

Gromov-Prohorov metric

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

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

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

where $A^\varepsilon := \big\{ x \in X \ \big| \ d(A, x) < \varepsilon \big\}$ and $\mathfrak{B}$ denotes the Borel σ-algebra.

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

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

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

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

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

Gromov-weak topology

The idea of Gromov-weak topology is to use convergence in distribution of finite metric subspaces, which are sampled from X with the measure μ. A very nice property of the Gromov-Prohorov metric is that it induces precisely the Gromov-weak topology, as shown in [17]. This alternative characterisation of convergence provides us with a subalgebra of $\mathcal{C}_b(\mathfrak{X})$, called algebra of polynomials. The usefulness of this algebra stems from the fact that it is rich enough to determine convergence of measures on $\mathfrak{X}$. To emphasize that polynomials are an essential tool for working with convergence in distribution of $\mathfrak{X}$-valued random variables, we remark that one cannot use the space $\mathcal{C}_c(\mathfrak{X})$ of continuous functions with compact support, because no point in $\mathfrak{X}$ has a compact neighbourhood, and hence $\mathcal{C}_c(\mathfrak{X}) = \{0\}$ is trivial.

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

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

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

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

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

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

Since $\mathcal{C}_b(\mathfrak{X})$ is difficult to describe, it is important to have such a set with a more tractable description. That Π is indeed convergence determining is shown with some effort by Depperschmidt, Greven and Pfaffelhuber in [8]. We can also deduce it from an apparently not so well-known general theorem due to Le Cam.

Theorem 2.7 (Le Cam, [21]; see also [20, Lem. 4.1])**.** *Let X be a completely regular Hausdorff space, and $\mathcal{F} \subseteq \mathcal{C}_b(X)$ multiplicatively closed. Then $\mathcal{F}$ is convergence determining for Radon probability measures if and only if $\mathcal{F}$ generates the topology of X.*

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

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

Gromov's metric $\underline{\square}_\lambda$

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

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

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

Obviously, we have

$$\square_\lambda \;\leq\; \square_{\lambda'} \;\leq\; \tfrac{\lambda}{\lambda'}\square_\lambda \qquad \forall \lambda > \lambda'.$$

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

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

Remark 2.11. *Because (X,d) is a Polish space, the set $\mathcal{F}(\mathcal{X})$ of (measure preserving) parametrisations is non-empty. This follows for example from the version of the Skorohod representation on $[0,1]$ given in [6, Thm. 8.5.4].*

Related topologies

1. In [16], Fukaya introduced the **measured Hausdorff topology** (often cited as measured Gromov-Hausdorff topology) for compact mm-spaces. The same topology is called **weighted Gromov-Hausdorff topology**, and a complete metric inducing it is constructed by Evans and Winter in [15]. The idea is that spaces are close if there is an ε-isometry mapping one measure Prohorov-close to the other. Convergence in measured Hausdorff topology implies Gromov-weak convergence, but not vice versa, because the former implies Gromov-Hausdorff convergence of the underlying metric spaces, which is not the case for Gromov-weak topology. Note that the underlying equivalence classes are also different: For two mm-spaces to be equivalent in the measured Hausdorff topology, the whole spaces have to be isometric, while in a Gromov-weak sense, this is required only for the supports of the measures.

2. Recently, Abraham, Delmas and Hoscheit ([2]) extended the measured Hausdorff topology to complete, locally compact, rooted length spaces with locally finite measures. Note that these measures are finite on all balls, because closed balls are compact in such spaces. The authors introduced the **Gromov-Hausdorff-Prohorov metric**, first on compact spaces using an embedding and measuring the sum of Hausdorff and Prohorov distance. That this metrises measured Hausdorff topology is easy to see from the definitions, using the same connection between ε-isometries and Hausdorff-close embeddings that is frequently applied in the context of Gromov-Hausdorff convergence. In the locally compact setting, they integrate the weighted distances of the measures restricted to balls. Note that this extended topology is vague in the sense that the total mass is not preserved. Thus, on spaces with finite (not necessarily probability) measures, it is not stronger than the natural extension of Gromov-weak topology, where the measures in Definition 2.5 are no longer required to be probabilities.

3. In [24], Sturm defines the L_2**-transportation distance** analogously to d_{GP}, but with the (2-)Wasserstein metric instead of the Prohorov metric. It induces a topology on $\mathfrak{X}$ that is strictly stronger than Gromov-weak topology, but coincides with it on subspaces of $\mathfrak{X}$ consisting of spaces with uniformly bounded (essential) diameter. Its restriction to the space of compact mm-spaces is strictly weaker than measured Hausdorff topology.

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

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

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

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

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

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

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

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

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

$$d\restriction_{X_i^2} := d_i \qquad \text{and} \qquad d(x,y) := \inf_{s \in S} d_1\big(x, \varphi_1(s)\big) + d_2\big(\varphi_2(s), y\big) + \varepsilon \;\; \forall x \in X_1,\, y \in X_2. \tag{3.1}$$

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

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

Therefore,

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

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

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

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

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

Proof. For $\lambda\ge\frac12$, the equation $\underline{\Box}_{\frac12} \,\le\, 2\lambda\underline{\Box}_\lambda \,\le\, 2\lambda\underline{\Box}_{\frac12}$ is obvious from the definition of $\underline{\Box}_\lambda$. For $\lambda\le\frac12$, we get the same inequality with "$\ge$" instead of "$\le$". Now the theorem implies the claim. □

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

Lemma 3.3. *The d defined in* (3.1) *satisfies the $\triangle$-inequality. Thus it is a metric.*

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

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

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

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

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

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

An $\mathbb{R}$-tree (see [10]) is a complete, connected 0-hyperbolic metric space (T,d). One of the possible definitions of 0-hyperbolicity is that it satisfies the four point condition, i.e.

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

Note that every 0-hyperbolic space can be embedded isometrically into a unique smallest $\mathbb{R}$-tree (see [14, Thm. 3.38]), which is separable whenever the original space was separable. Because d_{GP} (unlike the measured Hausdorff topology) identifies a metric measure space with every subspace containing the support of the measure, the equivalence class of every 0-hyperbolic space contains an $\mathbb{R}$-tree.

One possibility to construct 0-hyperbolic spaces is to code them by excursions, see [3, 22, 13]. To this end, let $h\colon[0,1]\to\mathbb{R}_+$ be a positive function with $h(0)=0$, and consider the semi-metric

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

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

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

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

Remark 4.1. *1. If the graph of h is connected, then T_h is complete and connected to begin with. We do not, however, make this restriction.*

2. If h is continuous, π_h is continuous and T_h is compact. Conversely, every compact $\mathbb{R}$-tree can be coded by a (non-unique) continuous excursion ([15, Rem. 3.2]). To code compact measured *trees, continuous excursions are not sufficient. See [12] for a detailed account on coding compact, rooted, ordered, measured $\mathbb{R}$-trees in a unique way by upper semi-continuous càglàd excursions.*

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

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

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

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

It is shown in [1, Prop. 2.9] that the coding function $\mathfrak{C}$ is Lipschitz continuous when the space of excursions is equipped with the uniform metric and the space of trees with the Gromov-Hausdorff-Prohorov metric. For the Gromov-Prohorov metric, this is a slightly weaker statement. The proof, however, becomes trivial in this case if we use Theorem 3.1, because the trees are already given in a parameterised form.

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

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

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

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

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

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

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

Note that the epigraph of a function is closed if and only if the function is lower semi-continuous. Epigraph convergence is usually defined as convergence in Fell topology (or equivalently Kuratowski convergence) of the epigraphs, see e.g. [4]. It is a compact, metrisable topology on the set $\overline{\mathcal{E}}$ of $(\mathbb{R}_+ \cup \{\infty\})$-valued, lower semi-continuous functions on $[0,1]$. On $\overline{\mathcal{E}}$, the topology induced by d_Γ is strictly stronger. Restricted to $\mathcal{E}$, however, the topologies coincide, which follows from [5, Thm. 1] using compactness of $[0,1]$ and $\mathbb{R}$-valuedness of excursions. Epigraph convergence also coincides with Γ-convergence (see e.g. [23]), whence the name d_Γ.

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

Recall that a metrisable topological space X is called *Lusin space* if it is the continuous, injective image of a Polish space, i.e. if there exists a Polish space Y and a continuous bijection $f\colon Y \to X$. X is Lusin if and only if it is homeomorphic to a Borel subset of a Polish space (see [7, Sec. 8.6] for details).

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

Proof. $d_{\mathcal{E}}$ is obviously a metric, and the continuous excursions are both d_Γ-dense (increasing pointwise convergence implies d_Γ-convergence) and d_λ-dense in $\mathcal{E}$. Hence $\mathcal{E}$ is separable, and it remains to show that $\mathcal{E}_b$ is a Borel subset of a Polish space. First note that this is the case for $(\mathcal{E}_b, d_\Gamma)$, because the set of excursions bounded by a fixed $M \in \mathbb{N}$ is closed in the compact metric space $\overline{\mathcal{E}}$ with epigraph topology. Now we can identify $(\mathcal{E}_b, d_{\mathcal{E}})$ with the graph of the function $\pi\colon (\mathcal{E}_b, d_\Gamma) \to L^0 := \big(L^0(\lambda), d_\lambda\big)$, which maps an excursion to its λ-a.e. equivalence class. It is enough to show that π is measurable, because then $(\mathcal{E}_b, d_{\mathcal{E}}) \cong \mathrm{gr}(\pi)$ is an injective measurable image of a Lusin space, hence Lusin itself by [7, Thm. 8.3.7].

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

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

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

That such a set cannot be Polish can be seen as follows. Let A_n be closed with dense complement in $\mathcal{E}$. Then its closure $\overline{A}_n$ in $\overline{\mathcal{E}}$ is closed with empty interior in the Polish space $\overline{\mathcal{E}}$. Assume that $A := \bigcup_{n\in\mathbb{N}} A_n$ is Polish. By the Mazurkiewicz theorem ([7, Thm. 8.1.4]), A is a $\mathcal{G}_\delta$-set in $\overline{\mathcal{E}}$, i.e. $A = \bigcap_{n\in\mathbb{N}} U_n$ for some open sets $U_n \subseteq \overline{\mathcal{E}}$. Let $A'_n := \overline{\mathcal{E}} \setminus U_n$. Then $\overline{\mathcal{E}} = \bigcup_{n\in\mathbb{N}} \big(\overline{A}_n \cup A'_n\big)$ and by the Baire category theorem ([7, Thm. D.37]), at least one A'_n has to have non-empty interior. This means that A is not dense.

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

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

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

2. "$I_g \le I_h + 2\varepsilon$": Because h is lower semi-continuous, the infimum $I_h(s,t)$ is attained and there is a $u \in [s,t]$ with $h(u) = I_h(s,t)$. From $d_\Gamma(h,g) \le \delta$, we obtain the existence of $u' \in [u-\delta, u+\delta]$ with $g(u') \le h(u) + \delta$. If $u' \in [s,t]$, then $I_g(s,t) \le g(u') \le$

$h(u) + \delta \le I_h(s,t) + \varepsilon$. For the case $u' \notin [s,t]$, assume w.l.o.g. $u' < s$, and therefore $u \in [s, s+\delta]$. Then, because s is not in A_δ, we have $I_h(s,t) = h(u) \ge h(s) - \varepsilon \ge g(s) - 2\varepsilon \ge I_g(s,t) - 2\varepsilon$.

3. "$I_h \le I_g + 2\varepsilon$": Choose $u \in [s,t]$ with $g(u) = I_g(s,t)$ and $u' \in [u-\delta, u+\delta]$ with $h(u') \le g(u) + \delta$. As above we can assume $u \in [s, s+\delta]$, $u' \in [s-\delta, s]$ and obtain $I_h(s,t) \le h(s) \le h(u') + \varepsilon \le g(u) + 2\varepsilon = I_g(s,t) + 2\varepsilon$. □

References

[1] Romain Abraham, Jean-François Delmas, and Patrick Hoscheit, *Exit times for an increasing Lévy tree-valued process*, 2012, arXiv:1202.5463.

[2] ______, *A note on Gromov-Hausdorff-Prokhorov distance between (locally) compact measure spaces*, 2012, arXiv:1202.5464.

[3] David Aldous, *The continuum random tree III*, Annals of Prob. **21** (1993), no. 1, 248–289. MR-1207226

[4] Gerald Beer, *Topologies on closed and closed convex sets*, Kluwer Acad. Publ., 1993. MR-1269778

[5] ______, *A note on epi-convergence*, Canad. Math. Bull. **37** (1994), no. 3, 294–239. MR-1289763

[6] V. I. Bogachev, *Measure theory, volume II*, Springer, 2007. MR-2267655

[7] Donald L. Cohn, *Measure theory*, Birkhäuser, 1980. MR-0578344

[8] Andrej Depperschmidt, Andreas Greven, and Peter Pfaffelhuber, *Marked metric measure spaces*, Electron. Commun. Prob. **16** (2011), 174–188. MR-2783338

[9] ______, *Tree-valued Fleming-Viot dynamics with mutation and selection*, Annals of Applied Prob. **22** (2012), no. 6, 2560–2615.

[10] Andreas W.M. Dress, V. Moulton, and W.F. Terhalle, *T-theory: An overview*, Europ. J. Combinatorics **17** (1996), no. 2-3, 161–175. MR-1379369

[11] Thomas Duquesne, *A limit theorem for the contour process of conditioned Galton-Watson trees*, Annals of Prob. **31** (2003), no. 2, 996–1027. MR-1964956

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

[13] Thomas Duquesne and Jean-François Le Gall, *Random trees, Lévy processes and spatial branching processes*, Astérisque **281** (2002), vi+147. MR-1954248

[14] Steven N. Evans, *Probability and real trees*, École d'Été de Probabilités de Saint Flour XXXV-2005, Lecture Notes in Mathematics, vol. 1920, Springer, 2007, pp. 1–193. MR-2351587

[15] Steven N. Evans and Anita Winter, *Subtree prune and regraft: a reversible real tree-valued markov process*, Annals of Prob. **34** (2006), no. 3, 918–961. MR-2243874

[16] Kenji Fukaya, *Collapsing of Riemannian manifolds and eigenvalues of Laplace operator*, Inventiones Math. **87** (1987), no. 3, 517–547. MR-0874035

[17] Andreas Greven, Peter Pfaffelhuber, and Anita Winter, *Convergence in distribution of random metric measure spaces (Λ-coalescent measure trees)*, Prob. Theo. Rel. Fields **145** (2009), no. 1-2, 285–322. MR-2520129

[18] ______, *Tree-valued resampling dynamics. Martingale problems and applications*, Prob. Theo. Rel. Fields **in press** (2011).

[19] Misha Gromov, *Metric structures for riemannian and non-riemannian spaces*, Birkhäuser, 1999. MR-1699320

[20] J. Hoffmann-Jørgensen, *Probability in banach spaces*, École d'Été de Probabilités de Saint Flour VI-1976, Lecture Notes in Mathematics, vol. 598, Springer, 1977. MR-0461610

[21] L. Le Cam, *Convergence in distribution of stochastic processes*, University of California Publications in Statistics **2** (1957), 207–236. MR-0086117

[22] Jean-François Le Gall, *The uniform random tree in a Brownian excursion*, Prob. Theo. Rel. Fields **96** (1993), no. 3, 369–383. MR-1231930

[23] Gianni Dal Maso, *An introduction to* Γ*-convergence*, Birkhäuser, 1993. MR-1201152

[24] Karl-Theodor Sturm, *On the geometry of metric measure spaces I*, Acta Math. **196** (2006), no. 1, 65–131. MR-2237206

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