Gromov-Hausdorff distances on p-metric spaces and ultrametric spaces

Facundo Mémoli[1], Zane Smith[2], and Zhengchao Wan[3]

[1]Department of Mathematics and Department of Computer Science and Engineering, The Ohio State University, memoli@math.osu.edu
[2]Department of Computer Science and Engineering, University of Minnesota, smit9474@umn.edu
[3]Department of Mathematics, The Ohio State University, wan.252@osu.edu

December 30, 2019

Abstract

We investigate certain subsets $\mathcal{M}_p$ of the collection of all compact metric spaces $\mathcal{M}$ which are characterized by satisfying a strengthened form of the triangle inequality which encompasses, for example, the strong triangle inequality satisfied by ultrametric spaces. We identify a family of Gromov-Hausdorff like distances on $\mathcal{M}_p$ and study geometric and computational properties of these distances as well as the stability of certain canonical projections $\mathfrak{S}_p : \mathcal{M} \to \mathcal{M}_p$. For the collection $\mathcal{U}$ of all ultrametric spaces, as a special example of $\mathcal{M}_p$, we explore an interleaving-type distance and reveal its relationship with the Gromov-Hausdorff distance. We study the geodesic property of $\mathcal{M}_p$ equipped with different distances. We exploit special properties of ultrametric spaces and devise efficient algorithms for computing the family of Gromov-Hausdorff distances which we prove run in polynomial time when restricted to special classes of ultrametric spaces. We generalize some of our results to the case of ultra-dissimilarity spaces.

Contents

1 Introduction

The notion of metric space is a fundamental concept in mathematics, computer science, and applied disciplines such as data science, where metric spaces serve as a model for datasets [DD09]. A metric space is a pair (X, d_X) consisting of a set X and a function $d_X : X \times X \to \mathbb{R}$ satisfying the following three conditions, for any $x, x', x'' \in X$:

1. $d_X(x, x') \geq 0$ and $d_X(x, x') = 0$ if and only if $x = x'$.
2. $d_X(x, x') = d_X(x', x)$.
3. $d_X(x, x') + d_X(x', x'') \geq d_X(x, x'')$.

The function d_X is referred to as the metric (or distance function) on X. The simplest example of a metric space is the one point metric space $*$. Common examples include subsets of Euclidean spaces, Riemannian manifolds, and metric graphs. In this paper, we are mostly interested in compact metric spaces. We denote by $\mathcal{M}$ the collection of all compact metric spaces.

An important notion regarding metric spaces is that of isometric embedding.

Definition 1.1. *A set map $f : X \to Y$ between two metric spaces is called an isometric embedding if for any $x, x' \in X$, $d_Y(f(x), f(x')) = d_X(x, x')$. We use the notation $f : X \hookrightarrow Y$ to denote isometric embeddings. If moreover f is bijective, we then say that f is an isometry. Whenever an isometry exists between X and Y we say that X is isometric to Y and denote this as $X \cong Y$.*

One natural question in metric geometry and in data analysis is how to compare two given metric spaces, or more precisely, how to define a metric structure on $\mathcal{M}$ that quantifies how far two spaces are from being isometric. In 1981, Gromov [Gro81] introduced a notion called the Gromov-Hausdorff distance to compare metric spaces. This distance is based on the Hausdorff distance.

Definition 1.2 (Hausdorff distance)**.** *Given a metric space Z, the Hausdorff distance d_H^Z between two subsets $A, B \subset Z$ is defined as*

$$d_{\mathrm{H}}^Z(A, B) = \inf\{r > 0 : \; B \subset A^r, A \subset B^r\},$$

where $A^r \coloneqq \{x \in X : \; d_X(x, A) \leq r\}$ is the r-neighborhood of A.

To compare two metric spaces, we then first isometrically embed them into a common ambient metric space, compute the Hausdorff distance, and then infimize over all such ambient spaces and embeddings. More precisely, we have the following definition.

Definition 1.3 (Gromov-Hausdorff distance)**.** *The Gromov-Hausdorff distance d_{GH} between two compact metric spaces X and Y is defined as*

$$d_{\mathrm{GH}}(X, Y) = \inf d_{\mathrm{H}}^Z(\varphi(X), \psi(Y)),$$

where the infimum is taken over all $Z \in \mathcal{M}$ and isometric embeddings $\varphi : X \hookrightarrow Z$ and $\psi : Y \hookrightarrow Z$.

Remark 1.4. *It is well known that one may replace Z above with the disjoint union $X \sqcup Y$ and infimize over all possible metrics d on the disjoint union such that when restricted to $X \times X$ (respectively $Y \times Y$) they equal d_X (respectively d_Y) [BBI01].*

Example 1.5. *An ε-net S of a compact metric space X for $\varepsilon > 0$ is a set such that for any $x \in X$, there exists $s \in S$ with $d_X(x, s) \leq \varepsilon$. In other words, $d_H^X(S, X) \leq \varepsilon$ and thus $d_{\mathrm{GH}}(S, X) \leq \varepsilon$.*

Remark 1.6. *In the definition above, it is enough to restrict Z to the disjoint union $X \sqcup Y$, and then infimize over all metrics d on the disjoint union such that $d|_{X\times X} = d_X$ and $d|_{Y\times Y} = d_Y$. From this observation, one can then see that always*

$$d_{\mathrm{GH}}(X, Y) \leq \frac{1}{2} \max\left(\mathrm{diam}(X), \mathrm{diam}(Y)\right). \tag{1}$$

Indeed, it is enough to consider the metric d such that $d(x, y) = \frac{1}{2} \max\left(\mathrm{diam}(X), \mathrm{diam}(Y)\right)$ for $x \in X$ and $y \in Y$. That the resulting d is a proper metric on the disjoint union is easy to see. That the claim in Equation (1) above is true follows now from Definition 1.2.

It is not hard to check that $d_{\mathrm{GH}}(X, Y) = 0$ if and only if X is isometric to Y. Moreover, d_{GH} is a legitimate metric on the collection of isometric classes of $\mathcal{M}$.

Theorem 1.7 (Theorem 7.3.30 in [BBI01])**.** *d_{GH} defines a finite metric on the space of isometry classes of compact metric spaces.*

It turns out that the Gromov-Hausdorff distance admits a characterization in terms of distortion of correspondences [BBI01] as follows. Given two metric spaces (X, d_X) and (Y, d_Y), a correspondence R between the underlying sets X and Y is any subset of $X \times Y$ such that the images of R under the canonical projections are full: $p_X(R) = X$ and $p_Y(R) = Y$. We define the distortion of R with respect to d_X and d_Y as follows:

$$\mathrm{dis}\left(R, d_X, d_Y\right) := \sup_{(x,y),(x',y')\in R} |d_X(x, x') - d_Y(y, y')|. \tag{2}$$

We will abbreviate $\mathrm{dis}\left(R, d_X, d_Y\right)$ to $\mathrm{dis}(R)$ whenever the metric structures are clear from the context. Then, the Gromov-Hausdorff distance can be characterized via distortion of correspondences as follows:

$$d_{\mathrm{GH}}(X, Y) = \frac{1}{2} \inf_R \mathrm{dis}(R). \tag{3}$$

It is shown in [CM16] that the infimum is always realized by a closed correspondence.

Remark 1.8. *With this characterization, one can prove that for all $X \in \mathcal{M}$, $d_{\mathrm{GH}}(X, *) = \frac{1}{2}\mathrm{diam}(X)$. Furthermore, for all X and Y in $\mathcal{M}$ one has the bound*

$$\frac{1}{2}\big|\mathrm{diam}(X) - \mathrm{diam}(Y)\big| \leq d_{\mathrm{GH}}(X, Y).$$

To prove the first claim note that the unique correspondence between X and $*$ is $R_* = X \times \{*\}$. Its distortion is $\mathrm{dis}(R_*) = \sup_{x,x'\in X} d_X(x, x') = \mathrm{diam}(X)$ hence the first claim. To prove the second claim note that, since d_{GH} satisfies the triangle inequality, then $|d_{\mathrm{GH}}(X, *) - d_{\mathrm{GH}}(Y, *)| \leq d_{\mathrm{GH}}(X, Y)$. The second claim then follows by invoking the first claim.

Though being theoretically interesting, it is known that computing d_{GH} is equivalent to solving a quadratic assignment problem which is an NP-hard problem [Mém07]. In his PhD thesis [Sch15] Schmiedl showed that this hardness extends to the case when the metric spaces are tree metric spaces [Gro87]. This leads us to consider the question whether d_{GH} would be computable in polynomial time when restricted to certain special subsets of $\mathcal{M}^{\mathrm{fin}}$, the collection of all finite metric spaces. If not, one may want to know if there exists any other distance on subsets of $\mathcal{M}$ that may be computationally tractable.

In this paper, we investigate a family $\{\mathcal{M}_p\}_{p\in[1,\infty]}$ of subsets of $\mathcal{M}$, where each $\mathcal{M}_p$ is the collection of compact metric spaces that satisfy a special triangle inequality, the *p-triangle inequality*: a compact metric space $(X, d_X) \in \mathcal{M}_p$ if and only for all $x, x', x'' \in X$ one has

$$(d_X(x,x'))^p + (d_X(x',x''))^p \geq (d_X(x,x''))^p\,.$$

Of course, $\mathcal{M}_1 = \mathcal{M}$. We call an element of $\mathcal{M}_p$ a *p-metric space* and refer to its metric as a p-metric. More precisely, we generalize the usual addition operator on $\mathbb{R}_{\geq 0}$ to the p-sum as follows.

$$a \boxplus_p b := \begin{cases} (a^p + b^p)^{\frac{1}{p}}, & p \in [1,\infty) \\ \max(a,b), & p = \infty \end{cases} \qquad \forall a, b \geq 0.$$

In fact, $\mathbb{R}_{\geq 0}$ becomes a commutative monoid (see [How95] for general background on monoids) with p-sum as its operator, that is, for any $a, b, c \geq 0$, there is an identity element 0 such that $0 \boxplus_p a = a \boxplus_p 0 = a$; $\left(a \boxplus_p b\right) \boxplus_p c = a \boxplus_p \left(b \boxplus_p c\right)$; and $a \boxplus_p b = b \boxplus_p a$. By associativity, we can add several numbers simultaneously and use the symbol $\boxplus_p$ in the same way as the summation symbol Σ:

$$\mathop{\boxplus_p}_{i=1}^{n} a_i = a_1 \boxplus_p a_2 \boxplus_p \cdots \boxplus_p a_n.$$

An immediate computation will give us that for $a > 0$ and $n \in \mathbb{N}$, $\boxplus_{p\,i=1}^{\;n} a = (n)^{\frac{1}{p}} a$ for $1 \leq p \leq \infty$, where we adopt the convention that $\frac{1}{\infty} = 0$ and $a^0 = 1$ for any $a > 0$.

For convenience, we adopt the following notation to represent the absolute p-difference between non-negative numbers a and b as

$$\Lambda_p(a,b) := |a^p - b^p|^{\frac{1}{p}}, \quad \text{for } p \in [1,\infty)$$

and as follows for $p = \infty$:

$$\Lambda_\infty(a,b) := \begin{cases} \max(a,b), & a \neq b \\ 0, & a = b \end{cases}$$

Remark 1.9. *It is not hard to show that for any $a, b \geq 0$, $\lim_{p\to\infty} a \boxplus_p b = a \boxplus_\infty b$ and $\lim_{p\to\infty} \Lambda_p(a,b) = \Lambda_\infty(a,b)$.*

Note that for any $a \geq 0$ and any $p \in [0,\infty]$ one has $\Lambda_p(a,0) = a$.

The following will be used in the sequel.

Proposition 1.10. *Assume that $a, b, c \geq 0$ and $p \in [1,\infty]$. we have both $a \boxplus_p b \geq c$ and $a \boxplus_p c \geq b$. Then,*

$$a \geq \Lambda_p(b,c).$$

Proof. When $p \neq \infty$, by assumption we have that $a^p + b^p \geq c^p$ and $a^p + c^p \geq b^p$. Therefore we have that $a^p \geq |b^p - c^p|$ and thus $a \geq \Lambda_p(b, c)$.

When $p = \infty$, we have the following two cases.

1. If $b = c$, then $a \geq 0 = \Lambda_p(b, c)$.

2. If $b \neq c$, we assume without loss of generality that $b > c$. Then $\max(a, c) = a \boxplus_p c \geq b$ implies that $a \geq b = \max(b, c) = \Lambda_p(b, c)$.

$\square$

We also define an asymmetric version of p-difference which we will use later.

$$A_p(a, b) := \begin{cases} \Lambda_p(a, b), & a > b \\ 0, & a \leq b \end{cases} \tag{4}$$

Proposition 1.11. *Assume that for $a, b, c \geq 0$ and $p \in [1, \infty]$ we have $\Lambda_p(a, b) \leq c$. Then,*

$$a \geq A_p(b, c) \text{ and } b \geq A_p(a, c).$$

Proof. We only need to prove the leftmost inequality. The rightmost inequality follows from essentially the same proof.

When $b \leq c$, by Equation (4), we have $a \geq 0 = A_p(b, c)$.

When $b > c$, by Equation (4), we have $A_p(b, c) = \Lambda_p(b, c)$. We need to consider the following two cases:

1. $p = \infty$. If $a < b$, then $\Lambda_\infty(a, b) = b \leq c$ contradicts with $b > c$. So $a \geq b$ and thus $a \geq b = \Lambda_\infty(b, c)$ since $b > c$.

2. $p \in [1, \infty)$. Then, $\Lambda_p(a, b) \leq c$ results in $|a^p - b^p| \leq c^p$. Hence $a^p \geq b^p - c^p = |b^p - c^p|$ and thus, $a \geq \Lambda_p(b, c)$.

$\square$

Then p-metric spaces are defined as follows.

Definition 1.12 (p-metric space)**.** *For $1 \leq p \leq \infty$, a p-metric space is a pair (X, d_X) consisting of a set X and a function $d_X : X \times X \to \mathbb{R}$ satisfying the following three conditions, for any $x, x', x'' \in X$:*

1. *$d_X(x, x') \geq 0$ and $d_X(x, x') = 0$ if and only if $x = x'$.*

2. *$d_X(x, x') = d_X(x', x)$.*

3. *$d_X(x, x') \boxplus_p d_X(x', x'') \geq d_X(x, x'')$.*

Remark 1.13 (Product p-metric)**.** *Suppose $X, Y \in \mathcal{M}_p$. Then, $(X \times Y, d_X \boxplus_p d_Y) \in \mathcal{M}_p$.*

Remark 1.14. *Λ_p actually defines a p-metric on $\mathbb{R}_{\geq 0}$. $\mathbb{R}^n_{\geq 0}$ has a natural p-metric Λ^n_p as the product p-metric of Λ_p defined in Remark 1.13.*

Note that p-metric spaces are in fact metric spaces in their own right: this follows from the fact that for $a, b \geq 0$, $a \boxplus_p b \leq a + b$ for all $p \in [1, \infty]$. This justifies the interpretation that p-metric spaces encode a certain degree of robustness to non-linear changes of scale. It is then also meaningful to consider diameter or ε-nets of p-metric spaces.

The notion of 1-metric space coincides with the usual metric space. Thus $\mathcal{M}_1$ coincides with the usual collection $\mathcal{M}$ of compact metric spaces. When $p = \infty$, $\mathcal{M}_\infty$ coincides with $\mathcal{U}$, the collection of all compact ultrametric spaces.

Definition 1.15 (Ultrametric spaces)**.** *A metric space (X, d_X) is ultrametric if for any $x, x', x'' \in X$, one has $d_X(x, x') \leq \max(d_X(x, x''), d_X(x', x''))$. We usually use u_X instead of d_X to denote the metric of an ultrametric space.*

Ultrametric spaces arise in statistics as a metric encoding of dendrograms (see Section 6 ahead) and also in phylogenetics [SS03]. In theoretical computer science, ultrametric spaces arise in particular as building blocks for the probabilistic approximation of finite metric spaces [Bar96].

We have the following:

Proposition 1.16. *For $1 \leq q \leq p \leq \infty$ the following inclusions hold: $\mathcal{U} \subseteq \mathcal{M}_p \subseteq \mathcal{M}_q \subseteq \mathcal{M}$.*

Proof. Given $X \subset \mathcal{M}_p$, we need to show that X satisfies the q-triangle inequality for $q \leq p$. Now for any $x, x', x'' \in X$, we have $d_X(x, x') \leq d_X(x, x'') \boxplus_p d_X(x'', x')$. Then it is sufficient to show that $d_X(x, x'') \boxplus_p d_X(x'', x') \leq d_X(x, x'') \boxplus_q d_X(x'', x')$. We will show in general that $a \boxplus_p b \leq a \boxplus_q b$ for $a, b \geq 0$.

The case when $a = 0$ is trivial that both sides equal b and the equality will hold. Now we assume $a > 0$. We will consider the function $f(p) \coloneqq (1 + x^p)^{\frac{1}{p}}$, for a fixed $x > 0$, $p \geq 1$. The derivative of this function is

$$f'(p) = \frac{1}{p^2}(1 + x^p)^{\frac{1-p}{p}} \left(x^p \ln(x^p) - (1 + x^p)\ln(1 + x^p)\right) \leq 0.$$

Hence f is non-increasing. Therefore if we take $x = \frac{b}{a}$, we have $\left(1 + \left(\frac{b}{a}\right)^p\right)^{\frac{1}{p}} \leq \left(1 + \left(\frac{b}{a}\right)^q\right)^{\frac{1}{q}}$ and $a \boxplus_p b \leq a \boxplus_q b$ for $a, b \geq 0$.

$\square$

Remark 1.17. *In the setting of standard metric spaces (i.e. when $p = 1$) one has the inequality $|d_X(x, x'') - d_X(x', x'')| \leq d_X(x, x')$ for all x, x', x'' in X. By Proposition 1.10 we have the following general inequality for any $p \in [1, \infty]$ and any $(X, d_X) \in \mathcal{M}_p$:*

$$\Lambda_p(d_X(x, x''), d_X(x', x'')) \leq d_X(x, x'),$$

for all $x, x', x'' \in X$.

1.1 Overview of our results

We now provide an overview of our results and a discussion of related work.

Section 2. Associated with each $p \in [1, \infty]$, there is a natural projection $\mathfrak{S}_p : \mathcal{M} \to \mathcal{M}_p$ [Seg16] which generalizes the construction of the so-called maximal subdominant ultrametric — a well known concept in the context of phylogenetic [SS03], hierarchical clustering [JS71, CM10], and theoretical computer science [FKW95]. Projections such as $\mathfrak{S}_p$ encode a certain notion of *simplification* of a metric space. We study the relation between projections for different parameters p and explore some properties of the *kernel* of $\mathfrak{S}_p$: the kernel of $\mathfrak{S}_p$ is defined as the set of all those metric spaces which are mapped to the one point metric space under $\mathfrak{S}_p$. Thus, understanding the kernel of $\mathfrak{S}_p$ is interesting because it tells us which metric spaces will be simplified "too much".

Section 3. Mimicking the structure of d_{GH}, we identify a family of Gromov-Hausdorff type distance functions $d_{\mathrm{GH}}^{(p)}$ on $\mathcal{M}_p$. In particular, in analogy with Theorem 1.7 we show that $d_{\mathrm{GH}}^{(p)}$ is a metric between isometry classes of $\mathcal{M}_p$. It is known [KO99, BBI01] that the Gromov-Hausdorff distance can be characterized by distortion of correspondences or distortion of maps as described by Equation (3). We found similar characterizations for $d_{\mathrm{GH}}^{(p)}$. As a consequence, we are able to analyze the computation complexity associated to $d_{\mathrm{GH}}^{(p)}$. It turns out that computing $d_{\mathrm{GH}}^{(p)}$ is still NP-hard when $p \in [1, \infty)$, whereas when $p = \infty$, computing $u_{\mathrm{GH}} := d_{\mathrm{GH}}^{(\infty)}$ can be done in polynomial time.

Section 4. For $p < \infty$, $d_{\mathrm{GH}}^{(p)}$ and d_{GH} are equivalent topologically on $\mathcal{M}_p$ whereas u_{GH} and d_{GH} induce different topologies on $\mathcal{U}$. Furthermore, we study convergent sequences of $d_{\mathrm{GH}}^{(p)}$ and establish a pre-compactness theorem for $d_{\mathrm{GH}}^{(p)}$: we prove that any class $\mathfrak{X}$ of p-metric spaces satisfying mild conditions is pre-compact. This implies that $\mathfrak{X}$ is actually *totally bounded*, i.e., for any $\varepsilon > 0$, there exists a positive integer $K(\varepsilon)$ and p-metric spaces $X_1, \cdots, X_{K(\varepsilon)}$ in $\mathfrak{X}$ such that for any $X \in \mathfrak{X}$, one can find $1 \leq i \leq K(\varepsilon)$ such that $d_{\mathrm{GH}}^{(p)}(X, X_i) \leq \varepsilon$. The concept of total boundedness is interesting from the point of view of studying geometric methods for data analysis in that it guarantees that for any given scale parameter ϵ, one can shatter a given dataset into a *finite* number of pieces each with size not larger than ϵ.

The pre-compactness theorem also provides us with tools to study the topology of $(\mathcal{M}_p, d_{\mathrm{GH}}^{(p)})$. In particular, we show that $(\mathcal{M}_p, d_{\mathrm{GH}}^{(p)})$ is complete and separable for $1 \leq p < \infty$ and, once again $(\mathcal{U}, u_{\mathrm{GH}})$ exhibits singular behavior in that it is complete but not separable. This suggests that $\mathcal{U}$ is rather singular among all other $\mathcal{M}_p$. Moreover, we study the subspace topology of $\mathcal{M}_p \subset \mathcal{M}_q$ when $p > q$.

Section 5. The u_{GH} metric has some special properties that make it quite singular amongst all the metrics $d_{\mathrm{GH}}^{(p)}$. In particular, we show that dis_p admits a special form when $p = \infty$ which is later used to find a poly time algorithm for its computation. In this section, we also relate u_{GH} to the curvature sets defined by Gromov [Gro07] and study $\widehat{u}_{\mathrm{GH}}$, a modified version of u_{GH} thus extending work from [Mém12]. We prove that the usual *codistortion* terms in the Gromov-Hausdorff distance do not appear in u_{GH} and that thus $u_{\mathrm{GH}} = \widehat{u}_{\mathrm{GH}}$ (cf. Theorem 5.24).

Section 6. As mentioned above, ultrametric spaces often arise in the context of data analysis in the form of *dendrograms*: a dendrogram is a certain hierarchical representation of a dataset. It is shown in [CM10] that there exists a structure preserving bijection between the set of dendrograms and the set of ultrametrics on a given finite set. There exists a natural distance called the interleaving distance d_{I} which can be used to measure discrepancy between two dendrograms. The collection of ultrametric spaces thus inherits this interleaving distance through the bijection mentioned above. We reformulate this interleaving distance in a clear form in Theorem 6.9 which allows us to obtain a characterization of d_{I} in terms of distortions of maps. We extend this usual interleaving distance to p-interleaving distance $d_{\mathrm{I},p}$ in a manner similar to the case of $d_{\mathrm{GH}}^{(p)}$ by using the notion of p-sum. It turns out that the characterization of d_{I} in terms of distortion of maps can be extended to $d_{\mathrm{I},p}$. With the help of this characterization, we prove that when restricted to $\mathcal{U}$, $d_{\mathrm{GH}}^{(p)}$ and $d_{\mathrm{I},p}$ are bi-Lipschitz equivalent and in particular $d_{\mathrm{GH}}^{(\infty)} = u_{\mathrm{GH}}$. Another advantage of our reformulation of d_{I} given in Theorem 6.9 is that we can generalize the interleaving distance from $\mathcal{U}$ to $\mathcal{M}$ and obtain a new metric on $\mathcal{M}$. We prove that this interleaving distance is a lower bound of d_{GH}.

Section 7 We have already studied some topological properties of $\mathcal{M}_p$ equipped with different distance functions in previous section. In this section, we will study one geometric property of these metric spaces, the geodesic property. It is known [CM16] that $(\mathcal{M}, d_{\mathrm{GH}})$ is a geodesic space. We introduce a notion called p-geodesic spaces which is a generalization of geodesic spaces. We prove that $(\mathcal{M}_p, d_{\mathrm{GH}}^{(p)})$ is a p-geodesic space when $p \in [1, \infty)$. Though $(\mathcal{U}, u_{\mathrm{GH}})$ is not geodesic, as an application of stability result of the projection $\mathfrak{S}_\infty$, we show that $(\mathcal{U}, d_{\mathrm{GH}})$ is geodesic. In the end, we show that $(\mathcal{U}, d_{\mathrm{I}})$ is not geodesic.

Section 8. In the end of our paper, we propose and study several algorithms for the computation of the generalized Gromov-Hausdorff distance $d_{\mathrm{GH}}^{(p)}$ on $\mathcal{U}$. For $1 \leq p \leq \infty$, we developed a recursive algorithm and a dynamic programming algorithm to determine the precise value of the $d_{\mathrm{GH}}^{(p)}$ distance between two given ultrametric spaces. We proved that within certain subset of $\mathcal{U}$, both algorithms run in polynomial time. In this section, reinforcing the observation that u_{GH} is rather singular amongst the family of metrics $d_{\mathrm{GH}}^{(p)}$, we also exhibit a polynomial time algorithm for its computation without any restriction on the spaces.

Related work Segarra thoroughly studied finite ultrametric and finite p-metric spaces in his PhD thesis [Seg16]; see also [SCMR15]. He was particularly interested in projecting finite networks onto p-metric spaces, in the process of which he identified a canonical projection map $\mathfrak{S}_p$ which we will define in the next section. In the context of finite metric spaces, Segarra proved that such a projection is unique under certain conditions. Segarra considered generalizations of metric spaces beyond p-metric spaces and in particular he identified the so-called *dioid metric spaces*. The idea was to generalize not only the addition operator but also the multiplication operator of $\mathbb{R}$. He also studied some theoretical properties of projection maps between different classes of dioid metric spaces.

The counterpart u_{GH} of the Gromov-Hausdorff distance d_{GH} on the collection of all compact ultrametric spaces was first introduced by Zarichnyi [Zar05] in 2005. He defined u_{GH} via the Hausdorff distance formulation (Definition 1.3) which is the formulation we will adopt in order to define $d_{\mathrm{GH}}^{(p)}$, a general p-version of Gromov-Hausdorff distance. He proved that u_{GH} is an ultrametric on the collection of isometry classes of ultrametric spaces and showed that the space $(\mathcal{U}, u_{\mathrm{GH}})$ is complete but not separable.

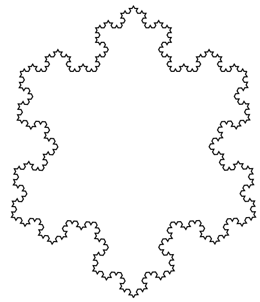

Figure 1: Koch snowflake.

Qiu further studied theoretical properties of metric structure induced by u_{GH} in his 2009 paper [Qiu09]. He found a distortion based description of u_{GH} in analogy to Equation (3), where the infimum is taken over a certain special subset of all correspondences which he called strong correspondences. Qiu also established several characterizations of u_{GH} as Burago et al. did for d_{GH} in Chapter 7 of [BBI01]. For example, Qiu modified the definition of ε-isometry and (ε,δ)-approximation to the so-called strong ε-isometry and strong ε-approximation. He proved that $u_{\mathrm{GH}}(X,Y)<\varepsilon$ if and only if there exists a strong ε-isometry between X and Y if and only if X is a strong ε-approximation of Y which are counterparts to Corollary 7.3.28 and Proposition 7.4.11 of [BBI01]. More interestingly, Qiu has also found a suitable version of Gromov's pre-compactness theorem for $(\mathcal{U},u_{\mathrm{GH}})$.

2 p-metric spaces and the projections $\mathfrak{S}_p:\mathcal{M}\to\mathcal{M}_p$

p-metric spaces are a special case of a more general notion called *p-snowflake metric spaces*. A p-snowflake metric space X is a metric space that is bi-Lipschitz equivalent to a p-metric space [TW05]. The name snowflake stems from a classical example of a fractal metric space, the Koch snowflake (see Figure 1), which turns out to be a $(\log_3 4)$-snowflake metric space.

Example 2.1 (Snowflake functor [DSSP97])**.** *For any metric space (X,d_X) and $0<\alpha<\infty$, consider the space $(X,(d_X)^\alpha)$, where*

$$(d_X)^\alpha(x,x'):=\left(d_X(x,x')\right)^\alpha,\quad \forall x,x'\in X.$$

The map that takes each (X,d_X) to $(X,(d_X)^\alpha)$ is called the α-snowflake functor. We denote $S_\alpha(X)=(X,(d_X)^\alpha)$.

It is easy to check that for $0<\alpha<1$, $(X,(d_X)^\alpha)$ is a compact $\frac{1}{\alpha}$-metric space if X is a compact metric space. If X is a compact α-metric space for some $1\le\alpha<\infty$, then $S_\alpha(X)$ is a compact metric space. Therefore, S_α can be viewed as either a map from $\mathcal{M}\to\mathcal{M}_{\frac{1}{\alpha}}$ or a map from $\mathcal{M}_\alpha\to\mathcal{M}$.

Recall that $\mathcal{M}_p$ denotes the collection of all compact p-metric spaces and $\mathcal{M}$ denotes the collection of all compact metric spaces. In this section, we will study the properties of a canonical projection $\mathfrak{S}_p:\mathcal{M}\to\mathcal{M}_p$ and a Gromov-Hausdorff type distance function $d_{\mathrm{GH}}^{(p)}:\mathcal{M}_p\times\mathcal{M}_p\to\mathbb{R}$ that makes $(\mathcal{M}_p,d_{\mathrm{GH}}^{(p)})$ a p-metric space.

2.1 The projections $\mathfrak{S}_p : \mathcal{M} \to \mathcal{M}_p$

For each $p \in [1, \infty]$ there exists a canonical projection $\mathfrak{S}_p : \mathcal{M} \to \mathcal{M}_p$ that takes a metric space to a p-metric space $(\hat{X}, \hat{d}_X^{(p)})$ which we will define below.

Given $(X, d_X) \in \mathcal{M}$, define for any $x, x' \in \tilde{X}$

$$d_X^{(p)}(x, x') := \inf \left\{ \boxplus_{p\, i=0}^{n-1} d_X(x_i, x_{i+1});\ x = x_0, x_1, \cdots, x_n = x' \right\}. \tag{5}$$

Remark 2.2. *It follows immediately that $d_X^{(p)}(x, x') \leq d_X(x, x')$ for any $x, x' \in X$.*

It is tempting to define $\mathfrak{S}_p(X, d_X)$ as $(X, d_X^{(p)})$. However, it may happen that $(X, d_X^{(p)})$ is a pseudometric space instead of a metric space, i.e., there may exist $x \neq x' \in X$ such that $d_X^{(p)}(x, x') = 0$. To circumvent this, we will transform $(X, d_X^{(p)})$ into a metric space in a canonical way: there is an equivalence relation $\sim_{d_X^{(p)}}$ where $x \sim_{d_X^{(p)}} x'$ if and only if $d_X^{(p)}(x, x') = 0$. Taking the quotient under this relation, we obtain a well-defined metric space $(\hat{X}, \hat{d}_X^{(p)})$, where $\hat{X} = X/\sim_{d_X^{(p)}} = \{[x] : x \in X\}$ is the set of equivalence classes, $[x]$ represents the equivalence class containing x, and

$$\hat{d}_X^{(p)}([x], [x']) = d_X^{(p)}(x, x'), \quad \forall x, x' \in X. \tag{6}$$

Then we define $\mathfrak{S}_p(X, d_X) := (\hat{X}, \hat{d}_X^{(p)})$, where $\hat{X} = X/\sim_{d_X^{(p)}}$.

We now verify that indeed $\mathfrak{S}_p$ maps compact metric spaces into compact p-metric spaces.

Proposition 2.3. *For every compact metric space (X, d_X), $(\hat{X}, \hat{d}_X^{(p)})$ is a compact p-metric space.*

Proof. Given $x, x', x'' \in X$, for any two chains of points $x = x_0, \cdots, x_n = x'$ and $x' = y_0, \cdots, y_m = x''$ in X, one can construct a chain between x and x'' by concatenating the previous two chains, denoted as $x = z_0, \cdots, z_{m+n+1} = x''$. Then one has

$$\left(\boxplus_{p\, i=0}^{n-1} d_X(x_i, x_{i+1})\right) \boxplus_p \left(\boxplus_{p\, j=0}^{m-1} d_X(y_j, y_{j+1})\right) = \boxplus_{p\, k=0}^{n+m} d_X(z_k, z_{k+1}) \geq d_X^{(p)}(x, x'').$$

By taking infimum on the left hand side, one can obtain that

$$d_X^{(p)}(x, x') \boxplus_p d_X^{(p)}(x', x'') \geq d_X^{(p)}(x, x'').$$

It then follows directly from Equation (6) that $\hat{d}_X^{(p)}$ satisfies the p-triangle inequality.

The map $\iota : (X, d_X) \to (X, d_X^{(p)})$ which is the identity as a set map is obviously continuous by Remark 2.2. The canonical projection map $p : (X, d_X^{(p)}) \to (\hat{X}, \hat{d}_X^{(p)})$ is also continuous. Since (X, d_X) is compact, we then have that $(\hat{X}, \hat{d}_X^{(p)}) = p \circ \iota((X, d_X))$ is compact. $\square$

Proposition 2.4 (Basic facts about $\mathfrak{S}_p$)**.** *We have the following properties about $\mathfrak{S}_p$:*

1. *For $1 \leq p \leq \infty$, when restricted to $\mathcal{M}_p$, $\mathfrak{S}_p$ coincides with the identity map.*
2. *For $q < p$, one has $\mathfrak{S}_p \circ \mathfrak{S}_q = \mathfrak{S}_p = \mathfrak{S}_q \circ \mathfrak{S}_p$.*

3. *Given $X \in \mathcal{M}_p$ and $c > 0$, $\mathfrak{S}_p(c \cdot X) = c \cdot \mathfrak{S}_p(X)$, where $c \cdot X$ denotes the metric space $(X, c \cdot d_X)$.*

4. *$\mathfrak{S}_\infty$ commutes with the snowflake functor S_p for any $1 \leq p < \infty$. More precisely, for any $X \in \mathcal{M}_p$, we have*
$$\mathfrak{S}_\infty \circ S_p(X) = S_p \circ \mathfrak{S}_\infty(X).$$

Remark 2.5. *If X is a finite space, then $(\hat{X}, \hat{d}_X^{(p)}) = (X, d_X^{(p)})$.*

The following theorem shows that p metric spaces can be viewed as a certain interpolation between metric spaces and ultrametric spaces.

Theorem 2.6. *Given a finite metric space X, the curve $\gamma : [0, \infty] \to \mathcal{M}$ defined by $p \mapsto \mathfrak{S}_p(X)$ is continuous in the sense of the Gromov-Hausdorff distance.*

Proof. Fix $x, x' \in X$. For any chain of points $x = x_0, x_1, \cdots, x_n = x'$, $\boxplus_{p\,i=0}^{\;n-1} d_X(x_i, x_{i+1})$ is continuous for $p \in [1, \infty]$. Then, we have by definition

$$d_X^{(p)}(x, x') := \inf \left\{ \mathop{\boxplus_p}_{i=0}^{n-1} d_X(x_i, x_{i+1});\, x = x_0, x_1, \cdots, x_n = x' \right\}$$

is also continuous with respect to $p \in [1, \infty]$, since X is finite and thus the infimum is taken over only finitely many chains. This implies that $(X, d_X^{(p)})$ is uniformly continuous with respect to $p \in [1, \infty]$ (i.e., $\sup_{x,x' \in X} d_X^{(p)}(x, x')$ is continuous). Thus by Example 7.4.2 in [BBI01], γ is continuous in the Gromov-Hausdorff sense. $\square$

Below, $\mathcal{M}^{\mathrm{fin}}$ and $\mathcal{M}_p^{\mathrm{fin}}$ denote the collections of finite metric spaces and p-metric spaces, respectively.

In the setting of finite spaces, Segarra et al. [SCMR15] proved that $\mathfrak{S}_p$ is actually the unique projection satisfying two reasonable conditions.

Theorem 2.7 ([Seg16]). *Let $p \in [1, \infty]$ and $\Phi_p : \mathcal{M}^{\mathrm{fin}} \to \mathcal{M}_p^{\mathrm{fin}}$ be any map satisfying the following two conditions:*

1. *Any p-metric space is a fixed point of Φ_p.*

2. *Any 1-Lipschitz map in $\mathcal{M}^{\mathrm{fin}}$ remains 1-Lipschitz in $\mathcal{M}_p^{\mathrm{fin}}$ after applying Φ_p.*

Then Φ_p exactly coincides with the restriction $\mathfrak{S}_p|_{\mathcal{M}^{\mathrm{fin}}}$.

Subdominant properties. We now provide an alternative description of $\mathfrak{S}_p$ which will reveal a certain subdominant property of $\hat{d}_X^{(p)}$. For $p = \infty$, the subdominant property of $\mathfrak{S}_\infty$ restricted to the collection of all finite metric spaces was already discussed in [CM10].

Under the equivalence relation $\sim_{d_X^{(p)}}$, not only does $d_X^{(p)}$ induce a metric $\hat{d}_X^{(p)}$, but also the original metric d_X induces a quotient metric $\hat{d}_X$ on $\hat{X}$ as follows (see Definition 3.1.12 in [BBI01] and also see Figure 2 for an illustration):

$$\hat{d}_X([x], [x']) := \inf \sum_{i=1}^{n} d_X(x_i, y_i),$$

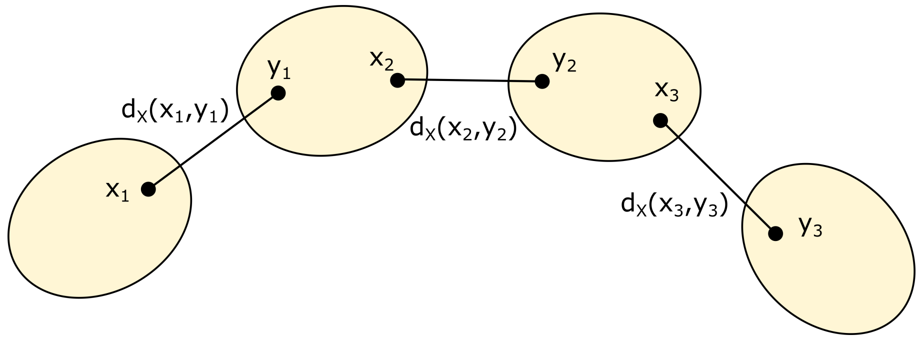

Figure 2: **Illustration of the quotient metric.** In this figure, each yellow ball represents an equivalence class of $\sim_{d_X^{(p)}}$ on a metric space X. Here, we represent one choice of chains $x = x_1, y_1, \cdots, y_n = x'$ between x and x' with $n = 3$. Then, $\hat{d}_X([x],[x'])$ is the infimum of the sum $\sum_{i=1}^n d_X(x_i, y_i)$ over all such chains .

where the infimum is taken over all chains $x = x_1, y_1, x_2, y_2, \cdots, x_n, y_n = x'$ in X with $y_i \sim_{d_X^{(p)}} x_{i+1}$ for $i = 1, \cdots, n-1$.

Remark 2.8. *It follows immediately from the definition that*

$$\hat{d}_X([x],[x']) \leq d_X(x,x').$$

Lemma 2.9. *For any compact metric space X, we have that $\hat{d}_X \geq \hat{d}_X^{(p)}$.*

Proof. By Remark 2.2, we have that for any $[x],[x'] \in \hat{X}$

$$\begin{aligned}\hat{d}_X([x],[x']) &= \inf \sum_{i=1}^n d_X(x_i, y_i) \geq \inf \sum_{i=1}^n d_X^{(p)}(x_i, y_i) \\ &= \inf \sum_{i=1}^n d_X^{(p)}(x_i, x_{i+1}) \geq d_X^{(p)}(x,x') = \hat{d}_X^{(p)}([x],[x'])\end{aligned}$$

where $x_{n+1} \coloneqq x'$ and the second equality follows from the fact that $d_X^{(p)}(y_i, x_{i+1}) = 0$. $\square$

Corollary 2.10. *Given $(X, d_X) \in \mathcal{M}$, for any $1 \leq p \leq \infty$ we have*

$$\operatorname{diam}(\mathfrak{S}_p(X)) \leq \operatorname{diam}(X).$$

Proof. This follows from Remark 2.8 and the fact that $\hat{d}_X^{(p)} \leq \hat{d}_X$. $\square$

Lemma 2.11. *$\hat{d}_X$ is a metric on $\hat{X}$.*

Proof. By Exercise 3.1.13 in [BBI01], we know that $\hat{d}_X$ is a pseudo-metric. Since $\hat{d}_X^{(p)}$ is a legitimate metric on $\hat{X}$, so is $\hat{d}_X$ by Lemma 2.9. $\square$

Proposition 2.12. *For a compact metric space X, we have that $\hat{d}_X^{(p)} = \left(\hat{d}_X\right)^{(p)}$, that is, we can obtain $\mathfrak{S}_p(X, d_X)$ by first taking the quotient of X with respect to $\sim_{d_X^{(p)}}$ and then applying the transformation defined in Equation (5). See Figure 3 for an illustration.*

$$
\begin{tikzcd}
\mathcal{M} \arrow[r, "\tilde{\mathfrak{S}}_p"] \arrow[d, "\mathfrak{T}_p"] \arrow[dr, dashed, "\mathfrak{S}_p"] & \tilde{\mathcal{M}} \arrow[d, "\mathfrak{T}"] \\
\mathcal{M} \arrow[r, "\tilde{\mathfrak{S}}_p"] & \mathcal{M}_p
\end{tikzcd}
$$

Figure 3: **Illustration of two ways of generating $\mathfrak{S}_p$.** By $\tilde{\mathcal{M}}_p$ denote the collection of p-pseudometric spaces, by $\tilde{\mathfrak{S}}_p : \mathcal{M} \to \tilde{\mathcal{M}}_p$ denote the map that takes (X, d_X) to $(X, d_X^{(p)})$, by $\mathfrak{T} : \tilde{\mathcal{M}}_p \to \mathcal{M}_p$ denote the canonical quotient map that takes a pseudometric space to its corresponding metric space, and by $\mathfrak{T}_p : \mathcal{M} \to \mathcal{M}$ denote the map that takes (X, d_X) to $(\hat{X}, \hat{d}_X)$ under the relation $\sim_{d_X^{(p)}}$. Then $\mathfrak{S}_p$ is the unique map such that the above diagram commutes.

Proof. By Lemma 2.9, we have that $\hat{d}_X \geq \hat{d}_X^{(p)}$. Then it is obvious from Equation (5) that

$$\left(\hat{d}_X^{(p)}\right)^{(p)} \leq \left(\hat{d}_X\right)^{(p)}.$$

By item 1 of Proposition 2.4, we have that $\left(\hat{d}_X^{(p)}\right)^{(p)} = \hat{d}_X^{(p)}$. Thus $\hat{d}_X^{(p)} \leq \left(\hat{d}_X\right)^{(p)}$.

On the other hand, we know by Remark 2.8 that for any $[x], [x'] \in \hat{X}$, $\hat{d}_X([x], [x']) \leq d_X(x, x')$. Then it follows again from Equation (5) that

$$\left(\hat{d}_X\right)^{(p)}([x], [x']) \leq d_X^{(p)}(x, x') = \hat{d}_X^{(p)}([x], [x'])$$

which concludes the proof. $\square$

Theorem 2.13 (Maximal subdominant p-metric)**.** *Given $(X, d_X) \in \mathcal{M}$ and $p \in [1, \infty]$, consider $(\hat{X}, \hat{d}_X^{(p)})$, the p-metric space generated by $\mathfrak{S}_p$. Define a partial order on the collection of all metrics on the set $\hat{X}$ as $d_1 \leq d_2$ if $\forall [x], [x'] \in X, d_1([x], [x']) \leq d_2([x], [x'])$. Then*

$$\hat{d}_X^{(p)} = \sup\{d : \ d \textit{ is a } p\textit{-metric on } \hat{X}, \ \textit{such that } d \leq \hat{d}_X\}.$$

Proof. By Lemma 2.9, we have $\hat{d}_X^{(p)} \leq \hat{d}_X$. Now suppose that d is a p-metric on $\hat{X}$, then $\hat{d}^{(p)} = d$ by item 1 of Proposition 2.4. Moreover, if $d \leq \hat{d}_X$, then it is easy to check that $d^{(p)} \leq \left(\hat{d}_X\right)^{(p)} = d_X^{(p)}$ where the last equality follows from Proposition 2.12. Therefore $d = d^{(p)} \leq d_X^{(p)}$. $\square$

2.2 Stability of $\mathfrak{S}_p$

For the projection $\mathfrak{S}_\infty : \mathcal{M} \to \mathcal{U}$, there is the following stability theorem in the literature.

Theorem 2.14 ([CM10])**.** *The map $\mathfrak{S}_\infty : \mathcal{M} \to \mathcal{U}$ has Lipschitz constant equal to 1, namely, for all compact metric spaces X and Y,*

$$d_{\mathrm{GH}}\left(\mathfrak{S}_\infty(X), \mathfrak{S}_\infty(Y)\right) \leq d_{\mathrm{GH}}(X, Y).$$

As a generalization of Theorem 2.14, we prove the following stability theorem for $\mathfrak{S}_p$ for all $p \in (1, \infty]$.

Theorem 2.15. *Given two finite metric spaces X and Y with $\#X = m$ and $\#Y = n$, and $p > 1$, we have*

$$d_{\mathrm{GH}}\left(\mathfrak{S}_p(X), \mathfrak{S}_p(Y)\right) \leq (\max(m,n) - 1)^{\frac{1}{p}}\ d_{\mathrm{GH}}(X, Y).$$

Remark 2.16. *Theorem 2.15 does not include the case $p = 1$ because $\mathfrak{S}_1$ on $\mathcal{M} = \mathcal{M}_1$ is just the identity map by Proposition 2.4. For $p = \infty$ Theorem 2.15 recovers Theorem 2.14.*

Proof of Theorem 2.15. The case when $m = n = 1$ reduces to comparing two one-point sets, a case in which the inequality obviously holds. If either $m = 1$ or $n = 1$, then we can obtain the inequality by invoking Corollary 2.10 and Remark 1.8.

Now we suppose $m, n > 1$. By Remark 2.5 we know that $\mathfrak{S}_p(X) = (X, d_X^{(p)})$ and $\mathfrak{S}_p(Y) = (Y, d_Y^{(p)})$, which means the underlying sets will remain unchanged.

Let R be an optimal correspondence between X and Y such that $\mathrm{dis}(R, d_X, d_Y) = \eta = 2d_{\mathrm{GH}}(X, Y)$. Then for any $(x, y), (x', y') \in R$, $|d_X(x, x') - d_Y(y, y')| \leq \eta$. Now let us bound $\left|d_X^{(p)}(x, x') - d_Y^{(p)}(y, y')\right|$. Suppose that $y = y_0, y_1, \cdots, y_k = y'$ is chain in Y such that $d_Y^{(p)}(y, y') = \mathop{\boxplus_p}\nolimits_{i=0}^{k-1} d_Y(y_i, y_{i+1})$, whose existence follows from the fact that Y is finite. Then we choose $x = x_0, x_1, \cdots, x_k = x'$ such that $(x_i, y_i) \in R$ for all $i = 0, \cdots, k$. Therefore by definition of $\mathfrak{S}_p$, we have

$$\begin{aligned} d_X^{(p)}(x, x') &\leq \mathop{\boxplus_p}\nolimits_{i=0}^{k-1} d_X(x_i, x_{i+1}) \leq \mathop{\boxplus_p}\nolimits_{i=0}^{k-1} \left(d_Y(y_i, y_{i+1}) + \eta\right) \\ &\leq \mathop{\boxplus_p}\nolimits_{i=0}^{k-1} d_Y(y_i, y_{i+1}) + \mathop{\boxplus_p}\nolimits_{i=0}^{k-1} \eta = d_Y^{(p)}(y, y') + k^{\frac{1}{p}} \eta, \end{aligned}$$

where the third inequality follows from the Minkowski inequality.

Note that an optimal chain in Y can always be chosen such that $k \leq m - 1$ hence we have

$$d_X^{(p)}(x, x') \leq d_Y^{(p)}(y, y') + (m - 1)^{\frac{1}{p}}\, \eta.$$

Similarly we can prove that $d_Y^{(p)}(y, y') \leq d_X^{(p)}(x, x') + (n - 1)^{\frac{1}{p}} \eta$. Therefore, we have

$$\left|d_X^{(p)}(x, x') - d_Y^{(p)}(y, y')\right| \leq (\max(m, n) - 1)^{\frac{1}{p}}\ \eta.$$

Then

$$\begin{aligned} &d_{\mathrm{GH}}\left(\mathfrak{S}_p(X), \mathfrak{S}_p(Y)\right) \leq \frac{1}{2} \mathrm{dis}(R, d_X^{(p)}, d_Y^{(p)}) \\ \leq &\frac{1}{2} (\max(m, n) - 1)^{\frac{1}{p}}\ \mathrm{dis}(R, d_X, d_Y) = (\max(m, n) - 1)^{\frac{1}{p}}\ d_{\mathrm{GH}}(X, Y). \end{aligned}$$

□

Example 2.17 (The coefficient in Theorem 2.15 is optimal for $p > 1$)**.** *Let (L_n, d) be the subset $\{0, 1, \cdots, n\}$ of the real line together with the Euclidean metric d_{n+1}. Let $p > 1$ and $L_n^{(p)} := \mathfrak{S}_p(L_n)$. Then $\mathrm{diam}(L_n^{(p)}) = d_{n+1}^{(p)}(0, n) = n^{\frac{1}{p}}$ and $d_{n+1}^{(p)}(i, i+1) = 1$ for $i = 0, \cdots, n - 1$.*

Let $\tilde{L}_n^{(p)} := n^{\frac{1}{p}-1} \cdot L_n$*. Then* $\mathrm{diam}(\tilde{L}_n^{(p)}) = \mathrm{diam}(L_n^{(p)})$*. By considering the diagonal correspondence* R *between* $L_n^{(p)}$ *and* $\tilde{L}_n^{(p)}$*, we have* $d_{\mathrm{GH}}(L_n^{(p)}, \tilde{L}_n^{(p)}) \leq \frac{1}{2}\mathrm{dis}(R) = \frac{1}{2}\left(1 - n^{\frac{1}{p}-1}\right)$*.*

Note that $L_n^{(p)} \in \mathcal{M}_p$*. Thus by Proposition 2.4 we have* $\mathfrak{S}_p(L_n^{(p)}) = L_n^{(p)}$*. For* $\tilde{L}_n^{(p)}$*, we have that* $\mathrm{diam}\left(\mathfrak{S}_p(\tilde{L}_n^{(p)})\right) = n^{\frac{2}{p}-1}$*. Hence by the lower bound for the Gromov-Hausdorff distance shown in Remark 1.8 we have*

$$d_{\mathrm{GH}}\left(\mathfrak{S}_p(L_n^{(p)}), \mathfrak{S}_p(\tilde{L}_n^{(p)})\right) \geq \frac{1}{2}\left|\mathrm{diam}(L_n^{(p)}) - \mathrm{diam}\left(\mathfrak{S}_p(\tilde{L}_n^{(p)})\right)\right| = \frac{1}{2}\left(n^{\frac{1}{p}} - n^{\frac{2}{p}-1}\right).$$

Therefore, we have that

$$\frac{d_{\mathrm{GH}}\left(\mathfrak{S}_p(L_n^{(p)}), \mathfrak{S}_p(\tilde{L}_n^{(p)})\right)}{d_{\mathrm{GH}}(L_n^{(p)}, \tilde{L}_n^{(p)})} \geq \frac{n^{\frac{1}{p}} - n^{\frac{2}{p}-1}}{1 - n^{\frac{1}{p}-1}} = n^{\frac{1}{p}},$$

which can be rewritten as

$$d_{\mathrm{GH}}\left(\mathfrak{S}_p(L_n^{(p)}), \mathfrak{S}_p(\tilde{L}_n^{(p)})\right) \geq n^{\frac{1}{p}}\, d_{\mathrm{GH}}(L_n^{(p)}, \tilde{L}_n^{(p)}).$$

By Theorem 2.15, we have $d_{\mathrm{GH}}\left(\mathfrak{S}_p(L_n^{(p)}), \mathfrak{S}_p(\tilde{L}_n^{(p)})\right) \leq ((n+1)-1)^{\frac{1}{p}}\ d_{\mathrm{GH}}(L_n^{(p)}, \tilde{L}_n^{(p)})$*. Hence we have that*

$$d_{\mathrm{GH}}\left(\mathfrak{S}_p(L_n^{(p)}), \mathfrak{S}_p(\tilde{L}_n^{(p)})\right) = n^{\frac{1}{p}}\, d_{\mathrm{GH}}(L_n^{(p)}, \tilde{L}_n^{(p)}).$$

Therefore the bound in Theorem 2.15 is tight. Since for $1 < p < \infty$ *the sequence* $\{n^{\frac{1}{p}}\}_{n\in\mathbb{N}}$ *is unbounded This example also shows that* $\mathfrak{S}_p : \mathcal{M} \to \mathcal{M}_p$ *is not a Lipschitz map.*

2.3 The kernel of $\mathfrak{S}_p$

In this section we study the notion of *kernel* of maps $\mathfrak{S}_p$ from $\mathcal{M}$ into $\mathcal{M}_p$. We define the *kernel* $\ker(\mathfrak{S}_p)$ of one such map to consist of all those compact metric spaces which are mapped to the one point space under the given map. It follows from the definition of $\mathfrak{S}_p$ that a compact metric space (X, d_X) lies in $\ker(\mathfrak{S}_p)$ if and only if $d_X^{(p)}(x, x') = 0$ for all $x, x' \in X$.

Recall that a metric space (X, d_X) is said to be *chain connected* if for any $x, x' \in X$ and any $\varepsilon > 0$ there exists a finite sequence $x = x_0, x_1, \ldots, x_n = x'$ such that $d_X(x_i, x_{i+1}) \leq \varepsilon$ for all i. Then it follows directly from the definition of $\mathfrak{S}_\infty$ that $\ker(\mathfrak{S}_\infty) = \mathcal{M}^{\mathrm{chain}}$, where $\mathcal{M}^{\mathrm{chain}}$ refers to the collection of all compact chain connected metric spaces. Since we are only considering compact metric spaces, a result in [AMCIL08] shows that chain connectedness is equivalent to connectedness. Therefore we have the following result for the kernel of $\mathfrak{S}_\infty$. Below, $\mathcal{M}^{\mathrm{conn}}$ denotes the collection of all compact connected metric spaces.

Proposition 2.18. $\ker(\mathfrak{S}_\infty) = \mathcal{M}^{\mathrm{conn}}$.

Remark 2.19. *Recall that any geodesic metric space [BBI01] is connected. Therefore any compact geodesic metric space lies in the kernel of* $\mathfrak{S}_\infty$*.*

Now let us turn our attention to the kernel of $\mathfrak{S}_p$ for all other $p \in (1, \infty)$. Proposition 2.18 will not hold for $p < \infty$. In fact, we have the following result.

Proposition 2.20. *Given $1 < q < p \leq \infty$, then*

$$\ker(\mathfrak{S}_q) \subsetneq \ker(\mathfrak{S}_p).$$

Proof. One can deduce from Proposition 2.4 that $\mathfrak{S}_p \circ \mathfrak{S}_q = \mathfrak{S}_p$ and thus $\ker(\mathfrak{S}_q) \subset \ker(\mathfrak{S}_p)$.

To conclude the proof, consider the following example. Let $X = ([0,1], d_X)$ be the subset of the real line with d_X being the restriction standard distance function on $\mathbb{R}$. Then, as mentioned in Example 2.1, $X_p = ([0,1], (d_X)^{\frac{1}{p}})$ becomes a p-metric space for $1 < p < \infty$. Hence, $X_p \notin \ker \mathfrak{S}_p$. But X_p is connected, thus $X_p \in \ker(\mathfrak{S}_\infty)$ which implies $\ker(\mathfrak{S}_p) \subsetneq \ker(\mathfrak{S}_\infty)$.

Now we show that $X_q \in \ker(\mathfrak{S}_p)$ for $q < p < \infty$. Take $0 \leq x < x' \leq 1$. Denote $l = x' - x$. Subdivide the interval $[x, x']$ into n equal subintervals to obtain $x = x_0, \cdots, x_n = x'$ such that $x_{i+1} - x_i = \frac{l}{n}$ for $i = 0, \cdots, n-1$. Then we have the following

$$((d_X)^{\frac{1}{q}})^{(p)}(x, x') \leq \boxplus_{p}{}_{i=0}^{n-1} (d_X)^{\frac{1}{q}}(x_i, x_{i+1}) = n^{\frac{1}{p} - \frac{1}{q}} \, l^{\frac{1}{q}}.$$

Since $\frac{1}{p} - \frac{1}{q} < 0$, by letting n go to infinity, one can derive that $((d_X)^{\frac{1}{q}})^{(p)}(x, x') = 0$. Therefore, $X_q \in \ker(\mathfrak{S}_p)$. Since $X_q \notin \ker(\mathfrak{S}_q)$, we have that $\ker(\mathfrak{S}_q) \subsetneq \ker(\mathfrak{S}_p)$. $\square$

Proposition 2.20 above leads us to consider the following object: $\bigcap_{p>1} \ker(\mathfrak{S}_p)$. We have not yet fully described this set but we conjecture that it coincides exactly with $\mathcal{M}^1$, the collection of all 1-connected compact metric spaces (defined below):

Conjecture 2.21. $\mathcal{M}^1 = \bigcap_{p>1} \ker(\mathfrak{S}_p)$.

We need to recall recall the definition of *Hausdorff dimension* [BBI01]. Let $(X, d_X) \in \mathcal{M}$, $k \geq 0$. For $A \subset X$ the k-th Hausdorff content of A is defined by

$$C_{\mathrm{H}}^k(A) \coloneqq \inf \left\{ \sum_{i \in I} r_i^k : \; A \subset \bigcup_{i \in I} B_{r_i}(x_i) \right\}.$$

The Hausdorff dimension of A is then $\dim_{\mathrm{H}}(A) \coloneqq \inf\{k \geq 0 : \; C_{\mathrm{H}}^k(A) = 0\}$.

Definition 2.22. *We say that $(X, d_X) \in \mathcal{M}$ is* 1-connected *if for all x and x' in X, there exists a closed connected subset $C \subset X$ such that $C \ni x, x'$ and $\dim_{\mathrm{H}}(C) = 1$. By $\mathcal{M}^1$ we denote the collection of all compact 1-connected metric spaces.*

Example 2.23. *Compact geodesic spaces are obviously 1-connected. An example of a compact metric space which is 1-connected yet not geodesic is the unit circle in $\mathbb{R}^2$ endowed with the Euclidean metric.*

Example 2.24 (A non-example)**.** *The space $X_p = \left([0,1], (d_X)^{\frac{1}{p}}\right)$ constructed in the proof of Proposition 2.20 has Hausdorff dimension p for $1 < p < \infty$ [Sem03] and is therefore not 1-connected.*

As a partial answer to the conjecture above, we have:

Proposition 2.25. $\mathcal{M}^1 \subset \bigcap_{p>1} \ker(\mathfrak{S}_p)$.

Proof. Let $X \in \mathcal{M}^1$. For $x, x' \in X$, let K be the 1-dimensional subset of X containing them. Then for any $p > 1$, $C_{\mathrm{H}}^p(K) = 0$. Hence we may find a finite cover of K: $\{B_{r_i}(x_i)\}$ such that $\sum_i r_i^p < \varepsilon^p$ for any $0 < \varepsilon < 1$. The 1-skeleton of the nerve of this cover is a connected graph, hence any two vertices (balls) are connected by a path on the graph. Without loss of generality, suppose $x \in B_{r_1}(x_1)$ and $x' \in B_{r_k}(x_k)$, and $\{B_{r_1}(x_1), B_{r_2}(x_2), \cdots, B_{r_k}(x_k)\}$ is a path in the nerve. Choose $y_i \in B_{r_i}(x_i) \cap B_{r_{i+1}}(x_{i+1})$ for $i = 1, \cdots, k-1$ and then construct a chain $x, x_1, y_1, x_2, y_2, \cdots, y_{k-1}, x_k, x'$. Then we have

$$\begin{aligned} &d_X(x, x_1) \boxplus_p d_X(x_1, y_1) \boxplus_p d_X(y_1, x_2) \boxplus_p \cdots \boxplus_p d_X(x_k, x') \\ \leq &r_1 \boxplus_p r_1 \boxplus_p r_2 \boxplus_p r_2 \boxplus_p \cdots \boxplus_p r_k = 2^{\frac{1}{p}} \boxplus_{p,\, i=1}^{k} r_i < 2^{\frac{1}{p}} \varepsilon. \end{aligned}$$

Since ε can be chosen arbitrarily small, we have that $d_X^{(p)}(x, x') = 0$. $\square$

3 $d_{\mathrm{GH}}^{(p)}$: generalized Gromov-Hausdorff distance

Now we define a Gromov-Hausdorff like distance on $\mathcal{M}_p$. For $X, Y \in \mathcal{M}_p$ let $\mathcal{D}_p(X, Y)$ denote the set of all p-metrics $d : X \bigsqcup Y \times X \bigsqcup Y \to \mathbb{R}_{\geq 0}$ such that $d|_{X\times X} = d_X$ and $d|_{Y\times Y} = d_Y$.

Then, consider $d_{\mathrm{GH}}^{(p)} : \mathcal{M}_p \times \mathcal{M}_p \to \mathbb{R}_{\geq 0}$, the generalized version of the Gromov-Hausdorff distance, defined by (cf. Definition 1.3 and Remark 1.4)

$$(X, Y) \longmapsto d_{\mathrm{GH}}^{(p)}(X, Y) \coloneqq \inf_{d \in \mathcal{D}_p(X,Y)} d_{\mathrm{H}}^{(X \sqcup Y, d)}(X, Y).$$

Example 3.1. *Given $\varepsilon > 0$, then for any ε-net S in X we have $d_{\mathrm{GH}}^{(p)}(S, X) \leq \varepsilon$.*

Theorem 3.2. *For each $p \in [1, \infty]$, $d_{\mathrm{GH}}^{(p)}$ is a legitimate p-metric on the collection of isometry classes of $\mathcal{M}_p$.*

We defer the proof of Theorem 3.2 until after introducing a distortion formula of $d_{\mathrm{GH}}^{(p)}$.

Remark 3.3 (Cf. Remark 1.6)**.** *Note that*

$$d_{\mathrm{GH}}^{(p)}(X, Y) \leq 2^{-\frac{1}{p}} \max(\mathrm{diam}(X), \mathrm{diam}(Y)).$$

Indeed, consider the p-metric d on $X \sqcup Y$ given by d_X on $X \times X$, d_Y on $Y \times Y$, and by $d(x, y) := 2^{-\frac{1}{p}} \max(\mathrm{diam}(X), \mathrm{diam}(Y))$ on $X \times Y$. Then, it is easy to check that d is indeed a p-metric and that the claim holds.

Remark 3.4. *For $1 \leq q < p \leq \infty$, by Proposition 1.16 we have $d_{\mathrm{GH}}^{(q)}$ is also defined on $\mathcal{M}_p \times \mathcal{M}_p$. Given X and Y in $\mathcal{M}_p$, then we have that $d_{\mathrm{GH}}^{(p)}(X, Y) \geq d_{\mathrm{GH}}^{(q)}(X, Y)$ since $\mathcal{D}_p(X, Y) \subseteq \mathcal{D}_q(X, Y)$. In particular, $d_{\mathrm{GH}}^{(p)}(X, Y) \geq d_{\mathrm{GH}}(X, Y)$ for any $p \in [1, \infty]$.*

3.1 Characterizations of $d_{\mathrm{GH}}^{(p)}$

Being a generalization of d_{GH}, $d_{\mathrm{GH}}^{(p)}$ shares similar properties with d_{GH}. In this section, we exhibit a distortion formula for $d_{\mathrm{GH}}^{(p)}$ (cf. Equation (3)).

For two p-metric spaces (X, d_X) and (Y, d_Y), the p-distortion of a correspondence R between X and Y is defined as

$$\operatorname{dis}_p(R, d_X, d_Y) := \sup_{(x,y),(x',y')\in R} \Lambda_p(d_X(x,x'), d_Y(y,y')). \tag{7}$$

When the underlying metric structures are clear, we will abbreviate $\operatorname{dis}_p(R, d_X, d_Y)$ as $\operatorname{dis}_p(R)$.

Note that for $p = 1$ the definition boils down to the usual notion of distortion of a correspondence given in Equation (2).

Theorem 3.5. *For all $p \in [1, \infty]$ and $X, Y \in \mathcal{M}_p$,*

$$d_{\mathrm{GH}}^{(p)}(X, Y) = 2^{-\frac{1}{p}} \inf_R \operatorname{dis}_p(R). \tag{8}$$

Remark 3.6 (Cf. Remark 1.8)**.** *Note that since there exists a unique correspondence $R_* = X \times \{*\}$ between a given $X \in \mathcal{M}_p$ and the one point space $*$, we have $d_{\mathrm{GH}}^{(p)}(X, *) = 2^{-\frac{1}{p}}\operatorname{dis}_p(R_*) = 2^{-\frac{1}{p}}\operatorname{diam}(X)$. Now, since $d_{\mathrm{GH}}^{(p)}$ satisfies the p-triangle inequality (Definition 1.12), we have that for all X and Y in $\mathcal{M}_p$, $\Lambda_p(d_{\mathrm{GH}}^{(p)}(X,*), d_{\mathrm{GH}}^{(p)}(Y,*)) \leq d_{\mathrm{GH}}^{(p)}(X,Y)$. Thus,*

$$2^{-\frac{1}{p}}\Lambda_p(\operatorname{diam}(X), \operatorname{diam}(Y)) \leq d_{\mathrm{GH}}^{(p)}(X, Y).$$

Proof of Theorem 3.5. To proceed with the proof, we need the following claim which is obvious from the definition of $d_{\mathrm{GH}}^{(p)}$ and [Mém11, Proposition 2.1].

Claim 1. $d_{\mathrm{GH}}^{(p)}(X, Y) := \inf_{R,d} \sup_{(x,y)\in R} d(x, y)$, *where R ranges over all correspondences between X and Y and $d \in \mathcal{D}_p(X, Y)$.*

Assume $\eta > d_{\mathrm{GH}}^{(p)}(X, Y)$ and let $d \in \mathcal{D}_p(X, Y)$ and $R \in \mathcal{R}(X, Y)$ be such that $d(x, y) < \eta$ for all $(x, y) \in R$. Then, one has for any $(x, y), (x', y') \in R$ that

$$\Lambda_p(d_X(x,x'), d_Y(y,y')) = \Lambda_p(d(x,x'), d(y,y')) \leq d(x,y) \boxplus_p d(x',y') < \eta \boxplus_p \eta = 2^{\frac{1}{p}}\eta.$$

Thus, by taking supremum over all pairs $(x, y), (x', y') \in R$, one has

$$\operatorname{dis}_p(R) \leq 2^{\frac{1}{p}}\,\eta.$$

By taking infimum of the left-hand side over all correspondences R between X and Y and letting η approach $d_{\mathrm{GH}}^{(p)}(X, Y)$, we obtain that $d_{\mathrm{GH}}^{(p)}(X, Y) \geq 2^{-\frac{1}{p}} \inf_R \operatorname{dis}_p(R)$.

For the opposite inequality, assume that R and $\eta > 0$ are such that $\operatorname{dis}_p(R) \leq 2^{\frac{1}{p}}\,\eta$. Consider $d \in \mathcal{D}_p(X, Y)$ given by

$$d(x,y) := \inf_{(x',y')\in R} d_X(x,x') \boxplus_p d_Y(y',y) \boxplus_p \eta, \quad \text{for } x \in X \text{ and } y \in Y.$$

That d is indeed a p-metric on $X \sqcup Y$ can be proved as follows. By the symmetric roles of X and Y, we only need to check the following two cases:

1. $d(x,y) \leq d(x,x') \boxplus_p d(x',y),\ x, x' \in X, y \in Y.$

2. $d(x,x') \le d(x,y) \boxplus_p d(x',y), x,x' \in X, y \in Y$.

For the first case,

$$\begin{aligned} d(x,x') \boxplus_p d(x',y) &= d(x,x') \boxplus_p \inf_{(x_1,y_1)\in R} \Big(d_X(x',x_1) \boxplus_p d_Y(y_1,y) \boxplus_p \eta \Big) \\ &= \inf_{(x_1,y_1)\in R} \Big(d(x,x') \boxplus_p d_X(x',x_1) \boxplus_p d_Y(y_1,y) \boxplus_p \eta \Big) \\ &\ge \inf_{(x_1,y_1)\in R} \Big(d(x,x_1) \boxplus_p d_Y(y_1,y) \boxplus_p \eta \Big) \\ &= d(x,y). \end{aligned}$$

For the second case,

$$\begin{aligned} &\inf_{(x_1,y_1)\in R} \Big(d_X(x,x_1) \boxplus_p d_Y(y_1,y) \boxplus_p \eta \Big) \boxplus_p \inf_{(x_2,y_2)\in R} \Big(d_X(x_2,x') \boxplus_p d_Y(y_2,y) \boxplus_p \eta \Big) \\ =& \inf_{(x_1,y_1),(x_2,y_2)\in R} \Big(d_X(x,x_1) \boxplus_p d_Y(y_1,y) \boxplus_p \eta \boxplus_p d_X(x_2,x') \boxplus_p d_Y(y_2,y) \boxplus_p \eta \Big) \\ \ge& \inf_{(x_1,y_1),(x_2,y_2)\in R} \Big(d_X(x,x_1) \boxplus_p d_X(x_2,x') \boxplus_p d_Y(y_1,y_2) \boxplus_p 2^{\frac{1}{p}}\eta \Big) \\ \ge& \inf_{(x_1,y_1),(x_2,y_2)\in R} \Big(d_X(x,x_1) \boxplus_p d_X(x_2,x') \boxplus_p d_X(x_1,x_2) \Big) \ge d(x,x'). \end{aligned}$$

The second inequality is due to the fact that $\mathrm{dis}_p(R) < 2^{\frac{1}{p}}\eta$ and the last inequality follows directly from the p-triangle inequality.

Note that $d(x,y) = \eta$ for $(x,y) \in R$. Therefore by Claim 1, $d_{\mathrm{GH}}^{(p)}(X,Y) \le \eta$. By a standard limit argument, one can then derive that $d_{\mathrm{GH}}^{(p)}(X,Y) \le 2^{-\frac{1}{p}} \inf_R \mathrm{dis}_p(R)$. $\square$

Now we can give a proof of Theorem 3.2 via the characterization given by Theorem 3.5.

Proof of Theorem 3.2. It is easy to show that $d_{\mathrm{GH}}^{(p)}(X,Y) = 0$ when $X \cong Y$. Then recall that by Remark 3.4 $d_{\mathrm{GH}}^{(p)} \ge d_{\mathrm{GH}}$. Hence, $d_{\mathrm{GH}}^{(p)}(X,Y) = 0$ implies that $d_{\mathrm{GH}}(X,Y) = 0$ and thus by Theorem 1.7 $X \cong Y$.

Given three spaces $X,Y,Z \in \mathcal{M}_p$, we need to prove the p-triangle inequality $d_{\mathrm{GH}}^{(p)}(X,Y) \le d_{\mathrm{GH}}^{(p)}(X,Z) \boxplus_p d_{\mathrm{GH}}^{(p)}(Y,Z)$. Suppose R_1 and R_2 are two correspondences between X and Z, and between Y and Z, respectively. Define the correspondence R between X and Y by

$$R := \{(x,y) \in X \times Y : \exists z \in Z, \text{ such that } (x,z) \in R_1 \text{ and } (y,z) \in R_2\}.$$

By a calculation similar to the one for $\mathrm{dis}_1 = \mathrm{dis}$ (see [BBI01, Exercise 7.3.26]), one can easily check that $\mathrm{dis}_p(R) \le \mathrm{dis}_p(R_1) \boxplus_p \mathrm{dis}_p(R_2)$, which implies that

$$d_{\mathrm{GH}}^{(p)}(X,Y) \le d_{\mathrm{GH}}^{(p)}(X,Z) \boxplus_p d_{\mathrm{GH}}^{(p)}(Y,Z)$$

by Theorem 3.5. $\square$

Another characterization of $d_{\mathrm{GH}}^{(p)}$. Besides the formula involving distortion of correspondence, there is another characterization of the Gromov-Hausdorff distance via distortion of maps due to Kalton and Ostrovskii [KO99]. Given two metric spaces X and Y, we define the distortion of any map $\varphi : X \to Y$ by

$$\mathrm{dis}(\varphi, d_X, d_Y) \coloneqq \sup_{x,x' \in X} \left| d_X(x,x') - d_Y(\varphi(x), \varphi(x')) \right|. \tag{9}$$

Given another map $\psi : Y \to X$, we define the codistortion of the pair of maps (φ, ψ) by

$$\mathrm{codis}(\varphi, \psi, d_X, d_Y) \coloneqq \sup_{x \in X, y \in Y} \left| d_X(x, \psi(y)) - d_Y(\varphi(x), y) \right|. \tag{10}$$

When the underlying metric structures are clear from the context, we will usually abbreviate $\mathrm{dis}(\varphi, d_X, d_Y)$ and $\mathrm{codis}(\varphi, \psi, d_X, d_Y)$ to $\mathrm{dis}(\varphi)$ and $\mathrm{codis}(\varphi, \psi)$, respectively.

Then one has the following formula, [KO99, Theorem 2.1]:

$$d_{\mathrm{GH}}(X,Y) = \frac{1}{2} \inf_{\substack{\varphi: X \to Y \\ \psi: Y \to X}} \max\left(\mathrm{dis}(\varphi), \mathrm{dis}(\psi), \mathrm{codis}(\varphi, \psi)\right). \tag{11}$$

In the case of p-metric spaces, a similar formula also holds. Assume X and Y are p-metric spaces. We then define p-distortion of a map $\varphi : X \to Y$ by

$$\mathrm{dis}_p(\varphi, d_X, d_Y) \coloneqq \sup_{x,x' \in X} \Lambda_p(d_X(x,x'), d_Y(\varphi(x), \varphi(x'))). \tag{12}$$

Similarly, given a map $\psi : Y \to X$, we define the p-codistortion of the pair (φ, ψ) by

$$\mathrm{codis}_p(\varphi, \psi, d_X, d_Y) \coloneqq \sup_{x \in X, y \in Y} \Lambda_p(d_X(x, \psi(y)), d_Y(\varphi(x), y)). \tag{13}$$

Similarly to what was done before, we will use abbreviations $\mathrm{dis}_p(\varphi)$ and $\mathrm{codis}_p(\varphi, \psi)$ when the underlying metric structures are clear from the context.

Then by invoking Theorem 3.5, one can easily derive the following distortion formula which is analogous to Equation (11).

Theorem 3.7. *For $X, Y \in \mathcal{M}_p$ and $p \in [1, \infty]$, one has that*

$$d_{\mathrm{GH}}^{(p)}(X,Y) = 2^{-\frac{1}{p}} \inf_{\substack{\varphi: X \to Y \\ \psi: Y \to X}} \max\left(\mathrm{dis}_p(\varphi), \mathrm{dis}_p(\psi), \mathrm{codis}_p(\varphi, \psi)\right). \tag{14}$$

3.2 Relation with d_{GH}

Isometry between $d_{\mathrm{GH}}^{(p)}$ and d_{GH} via the snowflake functor. As an application of the distortion formula (8) for $d_{\mathrm{GH}}^{(p)}$, one can directly relate $d_{\mathrm{GH}}^{(p)}$ and d_{GH} in the following way via the snowflake functor:

Theorem 3.8. *Given $1 \leq p < \infty$ and any two p-metric spaces X and Y, one has*

$$d_{\mathrm{GH}}^{(p)}(X,Y) = \left(d_{\mathrm{GH}}(S_p(X), S_p(Y))\right)^{\frac{1}{p}}.$$

Conversely, if X and Y are two metric spaces, then

$$d_{\mathrm{GH}}(X,Y) = \left(d_{\mathrm{GH}}^{(p)}\left(S_{\frac{1}{p}}(X), S_{\frac{1}{p}}(Y)\right)\right)^p.$$

Proof. Suppose X and Y are metric spaces. Now, for any correspondence R between X and Y, we have

$$\begin{aligned}\operatorname{dis}(R, d_X, d_Y) &= \sup_{(x,y),(x',y')\in R} |d_X(x,x') - d_Y(y,y')| \\ &= \sup_{(x,y),(x',y')\in R} \left| \left((d_X)^{\frac{1}{p}}(x,x')\right)^p - \left((d_Y)^{\frac{1}{p}}(y,y')\right)^p \right| \\ &= \left(\operatorname{dis}_p\left(R, (d_X)^{\frac{1}{p}}, (d_Y)^{\frac{1}{p}}\right)\right)^p\end{aligned}$$

Therefore, by the distortion characterization formula given in Theorem 3.5, we have

$$d_{\mathrm{GH}}(X,Y) = \left(d_{\mathrm{GH}}^{(p)}\left(S_{\frac{1}{p}}(X), S_{\frac{1}{p}}(Y)\right)\right)^p.$$

Similarly, if X and Y are two p-metric spaces, then for any correspondence R between X and Y, we have that

$$\operatorname{dis}_p(R, d_X, d_Y) = (\operatorname{dis}(R, (d_X)^p, (d_Y)^p))^{\frac{1}{p}}.$$

Again, by Theorem 3.5, this implies that

$$d_{\mathrm{GH}}^{(p)}(X,Y) = (d_{\mathrm{GH}}(S_p(X), S_p(Y)))^{\frac{1}{p}}.$$

□

Theorem 3.8 in particular establishes the stability of $\mathfrak{S}_\infty$ when restricted to $\mathcal{M}_p$.

Corollary 3.9. *Given $X, Y \in \mathcal{M}_p$, we have*

$$d_{\mathrm{GH}}^{(p)}(\mathfrak{S}_\infty(X), \mathfrak{S}_\infty(Y)) \leq d_{\mathrm{GH}}^{(p)}(X,Y).$$

Proof. The inequality follows directly from Proposition 2.4, Theorem 2.14 and Theorem 3.8:

$$\begin{aligned}d_{\mathrm{GH}}^{(p)}(\mathfrak{S}_\infty(X), \mathfrak{S}_\infty(Y)) &= (d_{\mathrm{GH}}(S_p(\mathfrak{S}_\infty(X)), S_p(\mathfrak{S}_\infty(Y))))^{\frac{1}{p}} \\ &= (d_{\mathrm{GH}}(\mathfrak{S}_\infty(S_p(X)), \mathfrak{S}_\infty(S_p(Y))))^{\frac{1}{p}} \\ &\leq (d_{\mathrm{GH}}(S_p(X), S_p(Y)))^{\frac{1}{p}} = d_{\mathrm{GH}}^{(p)}(X,Y).\end{aligned}$$

□

In fact, Theorem 3.8 closely relates the two spaces $(\mathcal{M}_p, d_{\mathrm{GH}}^{(p)})$ and $(\mathcal{M}, d_{\mathrm{GH}})$.

Theorem 3.10. *For $p \in [1,\infty)$, we have $(\mathcal{M}_p, d_{\mathrm{GH}}^{(p)}) \cong \left(\mathcal{M}, (d_{\mathrm{GH}})^{\frac{1}{p}}\right)$.*

Proof. Define a map $\Phi_p : \mathcal{M}_p \to \mathcal{M}$ by taking X to $S_p(X)$. By Theorem 3.8, we have that for any $X, Y \in \mathcal{M}_p$,

$$d_{\mathrm{GH}}^{(p)}(X,Y) = (d_{\mathrm{GH}}(S_p(X), S_p(Y)))^{\frac{1}{p}} = (d_{\mathrm{GH}})^{\frac{1}{p}}(\Phi_p(X), \Phi_p(Y)).$$

Define $\Psi_p : \mathcal{M} \to \mathcal{M}_p$ by taking X to $S_{\frac{1}{p}}(X)$. Similarly, we will obtain

$$(d_{\mathrm{GH}})^{\frac{1}{p}}(X,Y) = d_{\mathrm{GH}}^{(p)}(\Psi_p(X), \Psi_p(Y)).$$

It is obvious that Ψ_p is the inverse of Φ_p and thus Φ_p is an isometry between $(\mathcal{M}_p, d_{\mathrm{GH}}^{(p)})$ and $\left(\mathcal{M}, (d_{\mathrm{GH}})^{\frac{1}{p}}\right)$. □

Hölder equivalence between $d_{\mathrm{GH}}^{(p)}$ and d_{GH}. We know from Remark 3.4 that $d_{\mathrm{GH}}^{(p)} \geq d_{\mathrm{GH}}$ for every $p \in [1, \infty]$. Naturally one may wonder whether d_{GH} could somehow upperbound $d_{\mathrm{GH}}^{(p)}$. The answer is positive when $p < \infty$.

Theorem 3.11. *There exist positive constants $C(p)$, $D(p)$ and $E(p)$ depending only on $p \in [1, \infty)$ such that for any $X, Y \in \mathcal{M}_p$, we have*

$$d_{\mathrm{GH}}^{(p)}(X, Y) \leq C(p) \max\left(\mathrm{diam}(X), \mathrm{diam}(Y)\right)^{D(p)} \left(d_{\mathrm{GH}}(X, Y)\right)^{E(p)}.$$

The proof follows from the following two simple lemmas regarding properties of the p-sum and the p-difference.

Lemma 3.12. *For $a > b \geq 0$ and $1 \leq p < \infty$, $f(p) := \Lambda_p(a, b)$ is an increasing function with respect to p.*

Proof. Let $g(p) = \ln f(p) = \frac{1}{p} \ln(a^p - b^p)$. Then we have

$$\begin{aligned} g'(p) &= \frac{1}{p^2}\left(\frac{a^p \ln a^p - b^p \ln b^p}{a^p - b^p} - \ln(a^p - b^p)\right) \\ &= \frac{1}{p^2(a^p - b^p)}\left((a^p - b^p)(\ln a^p - \ln(a^p - b^p)) + b^p\left(\ln b^p - \ln(a^p - b^p)\right)\right) > 0 \end{aligned}$$

Therefore g is an increasing function and so is f. □

Lemma 3.13. *For $M > a > b \geq 0$ and $1 \leq p < \infty$, one has*

$$\Lambda_p(a, b) \leq \lceil p \rceil^{\frac{1}{\lceil p \rceil}} M^{1 - \frac{1}{\lceil p \rceil}} |a - b|^{\frac{1}{\lceil p \rceil}},$$

where $\lceil p \rceil$ is the smallest integer greater than or equal to p.

Proof. First assume that p is an integer. Then

$$\begin{aligned} a^p - b^p &= (a - b)(a^{p-1} + a^{p-2}b + \cdots + ab^{p-2} + b^{p-1}) \\ &< (a - b) \cdot p\, M^{p-1}. \end{aligned}$$

Hence $\Lambda_p(a, b) \leq p^{\frac{1}{p}} a^{1 - \frac{1}{p}} |a - b|^{\frac{1}{p}}$.

Now if p were not an integer, by the previous lemma $\Lambda_p(a, b) \leq \left|a \boxminus_{\lceil p \rceil} b\right|$. The proof now follows. □

Proof of Theorem 3.11. For any correspondence R between X and Y, we need to show that $\mathrm{dis}_p(R)$ is bounded above by some function of $\mathrm{dis}(R)$. Let $M = \max\left(\mathrm{diam}(X), \mathrm{diam}(Y)\right)$. For $(x, y), (x', y') \in R$, we have that

$$\Lambda_p(d_X(x, x'), d_Y(y, y')) \leq \lceil p \rceil^{\frac{1}{\lceil p \rceil}} M^{1 - \frac{1}{\lceil p \rceil}} |d_X(x, x') - d_Y(y, y')|^{\frac{1}{\lceil p \rceil}}$$

Therefore $\mathrm{dis}_p(R) \leq \lceil p \rceil^{\frac{1}{\lceil p \rceil}} M^{1 - \frac{1}{\lceil p \rceil}} \left(\mathrm{dis}(R)\right)^{\frac{1}{\lceil p \rceil}}$ and thus $d_{\mathrm{GH}}^{(p)}(X, Y) \leq \lceil p \rceil^{\frac{1}{\lceil p \rceil}} (2M)^{1 - \frac{1}{\lceil p \rceil}} \left(d_{\mathrm{GH}}(X, Y)\right)^{\frac{1}{\lceil p \rceil}}$. □

Combining the inequality given by Theorem 3.11 with $d_{\mathrm{GH}}^{(p)} \geq d_{\mathrm{GH}}$, one can conclude that when $p < \infty$, $d_{\mathrm{GH}}^{(p)}$ and d_{GH} induce the same topology on $\mathcal{M}_p$.

In contrast, the situation is quite different when $p = \infty$. The following example shows that u_{GH} and d_{GH} induce *different* topologies on $\mathcal{U}$.

Example 3.14. *Fix $\varepsilon > 0$. Consider the two-point metric space $\Delta_2(1)$ with interpoint distance 1 and the two-point metric space $\Delta_2(1+\varepsilon)$ with interpoint distance $1+\varepsilon$. These two spaces are obviously ultrametric spaces. Moreover $d_{\mathrm{GH}}(\Delta_2(1), \Delta_2(1+\varepsilon)) = \frac{\varepsilon}{2}$ and $u_{\mathrm{GH}}(\Delta_2(1), \Delta_2(1+\varepsilon)) = 1+\varepsilon$. Therefore, when ε approaches 0, $\Delta_2(1+\varepsilon)$ will converge to $\Delta_2(1)$ in the sense of d_{GH} but not in the sense of u_{GH}.*

In conclusion, for $p \in [1, \infty)$, $d_{\mathrm{GH}}^{(p)}$ is topologically equivalent to d_{GH} on $\mathcal{M}_p$, whereas u_{GH} induces a topology on $\mathcal{U}$ which is coarser than the one induced by d_{GH}. In Section 5, we will discuss some other singular properties of u_{GH}.

3.3 Computational complexity of $d_{\mathrm{GH}}^{(p)}$

It follows from Proposition 1.16 that an ultrametric space is a p-metric space for any $1 \leq p \leq \infty$. Therefore we can consider $d_{\mathrm{GH}}^{(p)}$ the restriction of $d_{\mathrm{GH}}^{(p)}$ to the collection $\mathcal{U}$ of all compact ultrametric spaces.

Corollary 3.15. *If X and Y are two ultrametric spaces and $1 < p < \infty$, then $S_{\frac{1}{p}}(X)$ and $S_{\frac{1}{p}}(Y)$ are still ultrametric spaces and*

$$d_{\mathrm{GH}}(X, Y) = \left(d_{\mathrm{GH}}^{(p)} \left(S_{\frac{1}{p}}(X), S_{\frac{1}{p}}(Y) \right) \right)^p.$$

Proof. That $S_{\frac{1}{p}}(X)$ and $S_{\frac{1}{p}}(Y)$ remain ultrametric is clear. The equality follows directly from Theorem 3.10. □

Corollary 3.15 above allows us to study the complexity associated to computing $d_{\mathrm{GH}}^{(p)}$. More precisely, we consider the following two computational problems: let $p \in [1, \infty)$ and $q \in [p, \infty]$

Decision Problem ($d_{\mathrm{GH}}^{(p)}$ distance computation on $\mathcal{M}_q$ ((p, q)-GHD-dec))
Input: Finite q-metric spaces (X, d_X) and (Y, d_Y), $\delta \geq 0$
Question: Is there a correspondence R between X and Y such that $\mathrm{dis}_p(R) \leq 2^{-\frac{1}{p}}\delta$?

Notice that (p, ∞)**-GHD-dec** reduces to the following problem:

Decision Problem ($d_{\mathrm{GH}}^{(p)}$ distance computation on $\mathcal{U}$ (p-GHDU-dec))
Input: Finite ultrametric spaces (X, u_X) and (Y, u_Y), $\delta \geq 0$
Question: Is there a correspondence R between X and Y such that $\mathrm{dis}_p(R) \leq 2^{-\frac{1}{p}}\delta$?

Proposition 3.16. *Problem p-**GHDU-dec** is NP-hard for $1 \leq p < \infty$. As a consequence, problem (p, q)**-GHD-dec** is NP-hard for $1 \leq p < \infty$ and $p \leq q \leq \infty$.*

Proof. Corollary 3.15 implies that an instance of Problem **1-GHDU-dec** can be reduced to an instance of Problem p**-GHDU-dec**. It is shown in the process of proving Corollary 3.8 in [Sch17] that Problem **1-GHDU-dec** is NP-hard which implies that Problem p**-GHDU-dec** is NP-hard. $\square$

The argument in the proof above cannot be applied when $p = \infty$ since the equation in Corollary 3.15 does not hold for $p = \infty$. Indeed, for an n-point ultrametric space X, $X^{\frac{1}{\infty}}$ becomes $\Delta_n(1)$, the n-point space with all interpoint distances equal to 1. Thus, $u_{\mathrm{GH}}(X^{\frac{1}{\infty}}, Y^{\frac{1}{\infty}})$ can only take values in $\{0, 1\}$, which in general prevents us from being able to recover $d_{\mathrm{GH}}(X, Y)$ from $u_{\mathrm{GH}}(X^{\frac{1}{\infty}}, Y^{\frac{1}{\infty}})$. We show in the next section that computing $d_{\mathrm{GH}}^{(\infty)} = u_{\mathrm{GH}}$ can be done in time polynomial on the cardinality of the input ultrametric spaces.

3.4 $d_{\mathrm{GH}}^{(p)}$ and approximate isometries

Aside from Theorem 3.7 above, as a counterpart to [BBI01, Corollary 7.3.28], there is another one-sided distortion characterization of $d_{\mathrm{GH}}^{(p)}$.

Definition 3.17. *Let X and Y be p-metric spaces for $1 \leq p \leq \infty$ and $\varepsilon > 0$. A map $f : X \to Y$ is called an (ε, p)-isometry if $\mathrm{dis}_p(f) \leq \varepsilon$ and $f(X)$ is an ε-net of Y.*

We then have:

Corollary 3.18. *Let X and Y be two p-metric spaces and $\varepsilon > 0$. Then*

1. *If $d_{\mathrm{GH}}^{(p)}(X, Y) < \varepsilon$, then there exists a $(2^{\frac{1}{p}}\varepsilon, p)$-isometry from X to Y.*
2. *If there exists an (ε, p)-isometry from X to Y, then $d_{\mathrm{GH}}^{(p)}(X, Y) < 2^{\frac{1}{p}}\varepsilon$.*

Proof. 1. Let R be a correspondence between X and Y with $\mathrm{dis}_p(R) < 2^{\frac{1}{p}}\varepsilon$. For every $x \in X$, choose $f(x) \in Y$ such that $(x, f(x)) \in R$. Then, obviously, $\mathrm{dis}_p(f) \leq \mathrm{dis}_p(R) < 2^{\frac{1}{p}}\varepsilon$. Now we show that $f(X)$ is a $2^{\frac{1}{p}}\varepsilon$-net for Y. For any $y \in Y$, choose $x \in X$ such that $(x, y) \in R$. Then,

$$d_Y(y, f(x)) = \Lambda_p(d_X(x, x), d_Y(y, f(x))) \leq \mathrm{dis}_p(R) < 2^{\frac{1}{p}}\varepsilon.$$

2. Let f be an (ε, p)-isometry. Define $R \subset X \times Y$ by

$$R = \{(x, y) : \, d_Y(y, f(x)) \leq \varepsilon\}.$$

R is a correspondence because $f(X)$ is an ε-net of Y. If $(x, y), (x', y') \in R$, then we have

$$\begin{aligned}\Lambda_p(d_X(x, x'), d_Y(y, y')) &\leq \Lambda_p(d_X(x, x'), d_Y(f(x), f(x'))) \boxplus_p \Lambda_p(d_Y(f(x), f(x')), d_Y(y, y')) \\ &\leq \mathrm{dis}_p(f) \boxplus_p d_Y(y, f(x)) \boxplus_p d_Y(y', f(x')) \leq 3^{\frac{1}{p}}\varepsilon.\end{aligned}$$

Hence $\mathrm{dis}_p(R) \leq 3^{\frac{1}{p}}\varepsilon$ and by Theorem 3.5 we have $d_{\mathrm{GH}}^{(p)}(X, Y) \leq \left(\frac{3}{2}\right)^{\frac{1}{p}}\varepsilon < 2^{\frac{1}{p}}\varepsilon$.

$\square$

Remark 3.19. *By taking $p = \infty$, this corollary essentially recovers Theorem 2.23 in Qiu's paper [Qiu09]. Qiu requires a slightly different condition on f called strong ε-isometry. This notion is actually a variant of (ε, ∞)-isometry which arises when one replaces $\mathrm{dis}_\infty(f) \leq \varepsilon$ by $\mathrm{dis}_\infty(f) < \varepsilon$ in Definition 3.17.*

4 Topology of $\left(\mathcal{M}_p, d_{\mathrm{GH}}^{(p)}\right)$

In this section, we study the topology induced by $d_{\mathrm{GH}}^{(p)}$ on $\mathcal{M}_p$. We will characterize convergence sequence $(\mathcal{M}_p, d_{\mathrm{GH}}^{(p)})$ and derive a pre-compactness result. We show that $(\mathcal{M}_p, d_{\mathrm{GH}}^{(p)})$ is a complete and separable space when $p < \infty$. Recall that $\mathcal{M}_p \subset \mathcal{M}_q$ when $q < p$ which leads us to also study the subspace topology of $\mathcal{M}_p$ inside $\mathcal{M}_q$.

4.1 Convergence under $d_{\mathrm{GH}}^{(p)}$

In this section, we will study convergent sequences in $\left(\mathcal{M}_p, d_{\mathrm{GH}}^{(p)}\right)$.

Definition 4.1. *Let $1 \leq p \leq \infty$. We say a sequence $\{X_n\}_{n=1}^{\infty}$ in $\mathcal{M}_p$ converges to $X \in \mathcal{M}_p$ if $d_{\mathrm{GH}}^{(p)}(X_n, X) \to 0$ as $n \to \infty$. Since $d_{\mathrm{GH}}^{(p)}$ is a metric (Theorem 3.2), the limit is unique up to isometry. We call X the $d_{\mathrm{GH}}^{(p)}$-limit of $\{X_n\}_{n=1}^{\infty}$.*

We have the following criterion for $d_{\mathrm{GH}}^{(p)}$ convergence generalizing the criterion of $d_{\mathrm{GH}} = d_{\mathrm{GH}}^{(1)}$ convergence mentioned in Section 7.4.1 of [BBI01]: a sequence $\{X_n\}$ of p-metric spaces converges to a p-metric space X if and only if there are a sequence $\{\varepsilon_n\}$ of positive numbers and a sequence of maps $f_n : X_n \to X$ (or, alternatively, $f_n : X \to X_n$) such that every f_n is an (ε_n, p)-isometry and $\varepsilon_n \to 0$.

Example 4.2. *Every compact p-metric space X is a $d_{\mathrm{GH}}^{(p)}$-limit of finite spaces. This is the counterpart to [BBI01, Example 7.4.9] and the proof is similar (it follows by considering ε-nets of X).*

This example actually indicates that convergence of compact p-metric spaces may reduce to convergence of their corresponding ε-nets. To make this precise, we define the notion of (ε, δ, p)-approximation as follows:

Definition 4.3. *Fix $1 \leq p \leq \infty$. Let X and Y be two compact p-metric spaces, and $\varepsilon, \delta > 0$. We say that X and Y are an (ε, δ, p)-approximation of each other if there exist finite collections of points $\{x_i\}_{i=1}^N$ and $\{y_i\}_{i=1}^N$ in X and Y, respectively, such that:*

1. *$\{x_i\}_{i=1}^N$ is an ε-net for X and $\{y_i\}_{i=1}^N$ is an ε-net for Y.*
2. *$\Lambda_p(d_X(x_i, x_j), d_Y(y_i, y_j)) \leq \delta$ for all $i, j = 1, \cdots, N$.*

Theorem 4.4. *Fix $1 \leq p \leq \infty$. Let X and Y be two compact p-metric spaces.*

1. *If X and Y are (ε, δ, p)-approximation of each other, then $d_{\mathrm{GH}}^{(p)}(X, Y) \leq \delta \boxplus_p 2^{\frac{1}{p}}\varepsilon$.*
2. *If $d_{\mathrm{GH}}^{(p)}(X, Y) < \varepsilon$, then Y is a $\left(5^{\frac{1}{p}}\varepsilon, 2^{\frac{1}{p}}\varepsilon, p\right)$-approximation of X (and vice versa).*

The proof follows similar steps as in the proof of Proposition 7.4.11 in [BBI01].

Proof. 1. Let $X_0 = \{x_i\}_{i=1}^N$ and $Y_0 = \{y_i\}_{i=1}^N$ be as in Definition 4.3. Then the second condition in that definition implies that the correspondence $\{(x_i, y_i) : i = 1, \cdots, N\}$ between X_0 and Y_0 has p-distortion bounded by δ. Hence, $d_{\mathrm{GH}}^{(p)}(X_0, Y_0) \leq 2^{-\frac{1}{p}}\delta \leq \delta$. By Example 3.1 we know that $\max\left(d_{\mathrm{GH}}^{(p)}(X_0, X), d_{\mathrm{GH}}^{(p)}(Y_0, Y)\right) \leq \varepsilon$. Thus

$$d_{\mathrm{GH}}^{(p)}(X, Y) \leq d_{\mathrm{GH}}^{(p)}(X, X_0) \boxplus_p d_{\mathrm{GH}}^{(p)}(X_0, Y_0) \boxplus_p d_{\mathrm{GH}}^{(p)}(Y_0, Y) \leq \delta \boxplus_p 2^{\frac{1}{p}}\varepsilon.$$

2. By Corollary 3.18 there exists a $(2^{\frac{1}{p}}\varepsilon, p)$-isometry $f : X \to Y$. Let $X_0 = \{x_i\}_{i=1}^N$ be an ε-net in X and let $y_i = f(x_i)$ for each $i = 1, \cdots, N$. Then, $\Lambda_p(d_X(x_i, x_j), d_Y(y_i, y_j)) \leq \mathrm{dis}_p(f) \leq 2^{\frac{1}{p}}\varepsilon$ for all i, j. Now, since $f(X)$ is a $2^{\frac{1}{p}}\varepsilon$-net of Y, for any $y \in Y$, there exists $x \in X$ such that $d_Y(f(x), y) \leq 2^{\frac{1}{p}}\varepsilon$. Since X_0 is an ε-net in X, we can choose $x_i \in X_0$ such that $d_X(x, x_i) \leq \varepsilon$. Then we have

$$\begin{aligned} d_Y(y, y_i) &\leq d_Y(y, f(x)) \boxplus_p d_Y(f(x), f(x_i)) \\ &\leq 2^{\frac{1}{p}}\varepsilon \boxplus_p d_X(x, x_i) \boxplus_p 2^{\frac{1}{p}}\varepsilon \leq 5^{\frac{1}{p}}\varepsilon \end{aligned}$$

Thus $\{y_i\}_{i=1}^N = f(X_0)$ is a $5^{\frac{1}{p}}\varepsilon$-net of Y.

□

In [Qiu09], Qiu introduced a notion called strong ε-approximation, which is exactly $(\varepsilon, 0, \infty)$-approximation in our language. The following corollary is a restatement of Theorem 3.5 in [Qiu09] regarding his strong ε-approximation. Though the corollary seems stronger than the result in the case $p = \infty$ of our Theorem 4.4, it turns out that they are equivalent. We will include a proof of the following corollary for completeness.

Corollary 4.5. *Let X and Y be two compact ultrametric spaces.*

1. *If X and Y are $(\varepsilon, 0, \infty)$-approximation of each other, then $u_{\mathrm{GH}}(X, Y) \leq \varepsilon$.*
2. *If $u_{\mathrm{GH}}(X, Y) < \varepsilon$, then Y is a $(\varepsilon, 0, \infty)$-approximation of X.*

Proof. Item 1 follows directly from item 1 of Theorem 4.4. Item 2 of Theorem 4.4 gives us that Y is a $(\varepsilon, \varepsilon, \infty)$-approximation of X. To conclude the proof, we only need to show that an $(\varepsilon, \varepsilon, \infty)$-approximation is automatically an $(\varepsilon, 0, \infty)$-approximation.

Since Y is a $(\varepsilon, \varepsilon, \infty)$-approximation of X, there exist ε-nets $\{x_i\}_{i=1}^N$ in X and $\{y_i\}_{i=1}^N$ in Y such that $\Lambda_\infty(u_X(x_i, x_j), u_Y(y_i, y_j)) \leq \varepsilon$.

Claim 2. *For an ultrametric space X and $x, x' \in X$, if $u_X(x, x') \leq \varepsilon$, then $B_\varepsilon(x) = B_\varepsilon(x')$ where $B_\varepsilon(x)$ is the closed ball centered at x with radius ε.*

Assuming the claim, if $u_X(x_i, x_j) \leq \varepsilon$ then by the definition of $Lambda_\infty$ we have that $u_Y(y_i, y_j) \leq \varepsilon$. Then, $B_\varepsilon(x_i) = B_\varepsilon(x_j)$ and $B_\varepsilon(y_i) = B_\varepsilon(y_j)$. This implies that after discarding x_j and y_j, $\{x_i\}_{i \neq j}$ and $\{y_i\}_{i \neq j}$ remain ε-nets of X and of Y respectively. We can continue this process to obtain two subsets $\{x_{\mathrm{n}_i}\}_{i=1}^M$ and $\{y_{\mathrm{n}_i}\}_{i=1}^M$ which are still ε-nets of X and of Y, respectively, while $u_X(x_{\mathrm{n}_i}, x_{\mathrm{n}_j}), u_Y(y_{\mathrm{n}_i}, y_{\mathrm{n}_j}) > \varepsilon$ for all $i \neq j$. Then by $\Lambda_\infty(u_X(x_{\mathrm{n}_i}, x_{\mathrm{n}_j}), u_Y(y_{\mathrm{n}_i}, y_{\mathrm{n}_j})) \leq \varepsilon$ we have that $u_X(x_{\mathrm{n}_i}, x_{\mathrm{n}_j}) = u_Y(y_{\mathrm{n}_i}, y_{\mathrm{n}_j})$ and thus $\Lambda_\infty(u_X(x_{\mathrm{n}_i}, x_{\mathrm{n}_j}), u_Y(y_{\mathrm{n}_i}, y_{\mathrm{n}_j})) \leq 0$. Then we conclude that Y is an $(\varepsilon, 0, \infty)$-approximation of X.

Proof of Claim 2. $\forall x'' \in B_\varepsilon(x)$, we have that $u_X(x', x'') \leq \max(u_X(x', x), u_X(x'', x)) \leq \varepsilon$. Therefore $x'' \in B_\varepsilon(x')$ and thus $B_\varepsilon(x) \subset B_\varepsilon(x')$. Similarly $B_\varepsilon(x') \subset B_\varepsilon(x)$ and thus $B_\varepsilon(x) = B_\varepsilon(x')$. □

□

4.2 Pre-compactness theorems

In [Gro81], Gromov proved a well known pre-compactness theorem stating that certain fairly general collections of compact metric spaces are pre-compact in the Gromov-Hausdorff sense. To be precise, we give a full description as follows.

Definition 4.6 (Definition 7.4.13 in [BBI01]). *A collection $\mathfrak{X}$ of compact metric spaces is called uniformly totally bounded, if*

1. *There exists $D > 0$ such that for any $X \in \mathfrak{X}$, $\mathrm{diam}(X) \leq D$.*
2. *For any $\varepsilon > 0$ there exists a natural number $N = N(\varepsilon) > 0$ such that for any $X \in \mathfrak{X}$, there exists an ε-net of X with cardinality bounded above by N.*

Theorem 4.7 (Gromov's pre-compactness theorem, Theorem 7.4.15 in [BBI01]). *Any uniformly totally bounded collection $\mathfrak{X}$ of compact metric spaces is pre-compact, i.e., any sequence in $\mathfrak{X}$ has a convergent subsequence in the sense of d_{GH}.*

We generalize this result to our setting of p-metric spaces and the $d_{\mathrm{GH}}^{(p)}$ distance, for $1 < p < \infty$, by invoking both Theorem 3.11 and the following lemma.

Lemma 4.8. *Fix $1 \leq p \leq \infty$. Suppose $\{X_n\}_{n=1}^{\infty}$ is a convergent sequence in $\mathcal{M}_p$. If $X \in \mathcal{M}$ is the Gromov Hausdorff limit of $\{X_n\}_{n=1}^{\infty}$, then $X \in \mathcal{M}_p$.*

Proof. We only need to check the p-triangle inequality. Let $x_1, x_2, x_3 \in X$ be three distinct points. Fix an $\varepsilon > 0$ small such that $\varepsilon \ll d(x_i, x_j)$ for $i \neq j$. Then, we can find n large enough such that $d_{\mathrm{GH}}(X_n, X) < 2\varepsilon$. Thus, there exists a correspondence R_n between X and X_n such that $\mathrm{dis}(R_n) < 4\varepsilon$. Then, there exist $y_1, y_2, y_3 \in X_n$ such that $(x_i, y_i) \in R_n$ for $i = 1, 2, 3$. Hence, we have

$$d_{X_n}(y_i, y_j) - 4\varepsilon \leq d_X(x_i, x_j) \leq d_{X_n}(y_i, y_j) + 4\varepsilon, \quad \forall i \neq j.$$

Then, we have

$$\begin{aligned} d_X(x_1, x_2) \boxplus_p d_X(x_2, x_3) &\geq (d_{X_n}(y_1, y_2) - 4\varepsilon) \boxplus_p (d_{X_n}(y_2, y_3) - 4\varepsilon) \\ &\geq d_{X_n}(y_1, y_2) \boxplus_p d_{X_n}(y_2, y_3) - 2^{\frac{1}{p}} \cdot 4\varepsilon \\ &\geq d_{X_n}(y_1, y_3) - 2^{2+\frac{1}{p}}\varepsilon \geq d_X(x_1, x_3) - (4 + 2^{2+\frac{1}{p}})\varepsilon. \end{aligned}$$

The second inequality is obvious for $p = \infty$ and follows from the Minkowski inequality for $p < \infty$.

Now, since ε can be chosen to be arbitrarily small, we conclude that $d_X(x_1, x_2) \boxplus_p d_X(x_2, x_3) \geq d_X(x_1, x_3)$. Thus, $X \in \mathcal{M}_p$. □

Corollary 4.9 ($d_{\mathrm{GH}}^{(p)}$ pre-compactness theorem)**.** *Fix $1 \leq p < \infty$. Any uniformly totally bounded collection $\mathfrak{X}$ of compact p-metric spaces is pre-compact, i.e., any sequence in $\mathfrak{X}$ has a convergent subsequence in the sense of $d_{\mathrm{GH}}^{(p)}$.*

Proof. Given any sequence $\{X_n\}_{n=1}^{\infty}$ in $\mathfrak{X}$, by Gromov's pre-compactness theorem, there exists a subsequence which converges in the sense of d_{GH}. Without loss of generality, we assume that $\{X_n\}_{n=1}^{\infty}$ is itself a convergent sequence and that X is its Gromov-Hausdorff limit. By the previous lemma, we have that $X \in \mathcal{M}_p$. Since $\mathrm{diam}(X_n) \leq D$ for any $n \in \mathbb{N}$, by Theorem 3.11, we have that $\{X_n\}_{n=1}^{\infty}$ will also converge to X in the sense of $d_{\mathrm{GH}}^{(p)}$. □

Note that the uniformly totally boundedness condition does not guarantee pre-compactness of a collection of ultrametric spaces.

Example 4.10. *Consider the collection of 2-point spaces $\left\{\Delta_2\left(1+\frac{1}{n}\right)\right\}_{n=1}^{\infty}$. This collection is obviously uniformly totally bounded. However, for any $n, m \in \mathbb{N}$, we have*

$$u_{\mathrm{GH}}\left(\Delta_2\left(1+\frac{1}{n}\right), \Delta_2\left(1+\frac{1}{m}\right)\right) = 1 + \max\left(\frac{1}{n}, \frac{1}{m}\right) > 1.$$

Therefore, there exists no Cauchy subsequence of $\left\{\Delta_2\left(1+\frac{1}{n}\right)\right\}_{n=1}^{\infty}$ and thus this sequence is not pre-compact.

Under a certain variant of the notion of uniformly totally boundedness, in [Qiu09] Qiu proved a pre-compactness theorem for u_{GH}. We include it here for completeness.

Below, for a given metric space (Y, d_Y), the spectrum of Y, $\mathrm{spec}(Y)$, is defined by $\mathrm{spec}(Y) := \{d_Y(y, y') : y, y' \in Y\}$. See Definition 5.9 for a more general concept.

Definition 4.11. *A collection $\mathfrak{X}$ of compact ultrametric spaces is called strongly uniformly totally bounded, if there exists a positive integer $N = N(\varepsilon)$ and a finite set $R(\varepsilon) \subset \mathbb{R}_{\geq 0}$ such that every $X \in \mathfrak{X}$ contains an ε-net S_X consisting of no more than N points and $\mathrm{spec}(S_X) \subset R(\varepsilon)$.*

Theorem 4.12 (u_{GH} pre-compactness theorem, [Qiu09])**.** *Any strongly uniformly totally bounded collection $\mathfrak{X}$ of compact ultrametric spaces is pre-compact.*

4.3 Separability and completeness of $(\mathcal{M}_p, d_{\mathrm{GH}}^{(p)})$

With the tools we have developed so far, we can establish the following theorem.

Theorem 4.13. *For each $1 \leq p < \infty$, $(\mathcal{M}_p, d_{\mathrm{GH}}^{(p)})$ is complete and separable.*

Proof. Fix a Cauchy sequence $\{X_n\}$ in $\mathcal{M}_p$. Then, obviously there exists $D > 0$ such that $\mathrm{diam}(X_n) \leq D$ for any $n \in \mathbb{N}$. Given $\varepsilon > 0$, suppose that when $n > N$, one has $d_{\mathrm{GH}}^{(p)}(X_N, X_n) < \varepsilon$. Then, by Theorem 4.4 we have that X_N is a $(5^{\frac{1}{p}}\varepsilon, 2^{\frac{1}{p}}\varepsilon, p)$-approximation of X_n. In fact, in the proof of item 2 of Theorem 4.4, we showed that for an ε-net of X_N, there exists a ε-net in X_n with the same cardinality. This implies that there exists $N = N(\varepsilon)$ such that for all $n \in \mathbb{N}$ there exists an ε-net in X_n with cardinality bounded by N. Applying the $d_{\mathrm{GH}}^{(p)}$-pre-compactness theorem (Corollary 4.9) we have that there exists a convergent subsequence of $\{X_n\}$, which implies that $\{X_n\}$ itself is convergent since it is Cauchy. Therefore, $(\mathcal{M}_p, d_{\mathrm{GH}}^{(p)})$ is complete.

By $\mathcal{M}_p^{(n)}$ denote the set of all n-point p-metric spaces with rational distances. Then, it is easy to check that $\bigcup_{n=1}^{\infty} \mathcal{M}_p^{(n)}$ is a countable dense set in $\mathcal{M}_p$ and thus $(\mathcal{M}_p, d_{\mathrm{GH}}^{(p)})$ is separable. □

Remark 4.14. *The proof above does not directly apply to the case when $p = \infty$. In fact, by using Qiu's pre-compactness theorem and the notion of strong approximation in [Qiu09], a slightly modification of the above proof will establish the completeness of $\mathcal{U}$. Interested readers are also referred to [Zar05] for a different method that proves completeness of $(\mathcal{U}, u_{\mathrm{GH}})$. However, it is shown in [Zar05] that $(\mathcal{U}, u_{\mathrm{GH}})$ is not a separable space, which suggests that $\mathcal{U}$ enjoys some special properties over all other $\mathcal{M}_p$.*

4.4 Subspace topology

As shown in Proposition 1.16 that $\mathcal{M}_p \subset \mathcal{M}_q$ when $1 \leq q < p$, we now study the topology of $(\mathcal{M}_p, d_{\mathrm{GH}}^{(q)})$ as a subspace of $(\mathcal{M}_q, d_{\mathrm{GH}}^{(q)})$. We need the following technical lemma about the relation between A_q and $\boxplus_p$ with $p \neq q$.

Lemma 4.15. *For $1 \leq q < p \leq \infty$ and $a, b, c \geq 0$, we have*

$$A_q(a, c) \boxplus_p A_q(b, c) \geq A_q\left(a \boxplus_p b, c \boxplus_p c\right).$$

Proof. When $p = \infty$, it is easy to see that $\max\left(A_\infty(a, c), A_\infty(b, c)\right) = A_\infty\left(\max(a, b), c\right)$, which is exactly what we want.

When $p < \infty$, we have the following cases:

1. $a, b \leq c$. Then, both sides of the inequality become 0, and thus the equality holds

2. $a, b \geq c$. Then, we need to prove the following:

$$\left((a^q - c^q)^{\frac{p}{q}} + (b^q - c^q)^{\frac{p}{q}}\right)^{\frac{1}{p}} \geq \left((a^p + b^p)^{\frac{q}{p}} - (c^p + c^p)^{\frac{q}{p}}\right)^{\frac{1}{q}},$$

   which is equivalent to

$$\left((a^q - c^q)^{\frac{p}{q}} + (b^q - c^q)^{\frac{p}{q}}\right)^{\frac{q}{p}} + (c^p + c^p)^{\frac{q}{p}} \geq (a^p + b^p)^{\frac{q}{p}}.$$

   This inequality follows directly from Minkowski inequality with the power $\frac{p}{q} > 1$.

3. $a \leq c, b > c$. It is easy to see that $A_q(c \boxplus_p b, c \boxplus_p c) \geq A_q(a \boxplus_p b, c \boxplus_p c)$. Then, it suffices to show that $A_q(b, c) = A_q(c, c) \boxplus_p A_q(b, c) \geq A_q(c \boxplus_p b, c \boxplus_p c)$, which follows from case 2.

$\square$

Proposition 4.16. *For $1 \leq q < p \leq \infty$, $(\mathcal{M}_p, d_{\mathrm{GH}}^{(q)})$ is a closed subspace of $(\mathcal{M}_q, d_{\mathrm{GH}}^{(q)})$.*

Proof. Given any $d_{\mathrm{GH}}^{(q)}$ convergent sequence $\{X_n\}_{n=1}^{\infty}$ with $X_n \in \mathcal{M}_p$ for all $n \in \mathbb{N}$, we need to show that $X = \lim_{n\to\infty} X_n$ belongs to $\mathcal{M}_p$. Take three distinct points $x_1, x_2, x_3 \in X$. Then, for any small $\varepsilon > 0$, there exists $N > 0$, such that for any $n > N$, we have $d_{\mathrm{GH}}^{(q)}(X_n, X) \leq \frac{\varepsilon}{2}$. Hence, there exists correspondence R_n between X_n and X such that $\mathrm{dis}_q(R_n) \leq \varepsilon$. Take $x_1^{(n)}, x_2^{(n)}, x_3^{(n)} \in X_n$ such that $(x_i^{(n)}, x_i) \in R_n$ for $i = 1, 2, 3$. Then, we have $\Lambda_q\left(d_X(x_i, x_j), d_{X_n}\left(x_i^{(n)}, x_j^{(n)}\right)\right) \leq \varepsilon$.

$$\begin{aligned} d_X(x_1, x_2) \boxplus_p d_X(x_2, x_3) &\geq A_q\left(d_{X_n}\left(x_1^{(n)}, x_2^{(n)}\right), \varepsilon\right) \boxplus_p A_q\left(d_{X_n}\left(x_2^{(n)}, x_3^{(n)}\right), \varepsilon\right) \\ &\geq A_q\left(d_{X_n}\left(x_1^{(n)}, x_2^{(n)}\right) \boxplus_p d_{X_n}\left(x_2^{(n)}, x_3^{(n)}\right), \varepsilon \boxplus_p \varepsilon\right) \\ &\geq A_q\left(d_{X_n}\left(x_1^{(n)}, x_3^{(n)}\right), \varepsilon \boxplus_p \varepsilon\right) \\ &\geq A_q\left(d_X(x_1, x_3), \varepsilon \boxplus_q (\varepsilon \boxplus_p \varepsilon)\right). \end{aligned}$$

The first and the last inequalities follow from Proposition 1.11. The second inequality follows from Lemma 4.15. Since $\varepsilon > 0$ is arbitrary, we have that

$$d_X(x_1, x_2) \boxplus_p d_X(x_2, x_3) \geq d_X(x_1, x_3),$$

and thus $X \in \mathcal{M}_p$. $\square$

Proposition 4.17. *Given* $1 \leq q < p \leq \infty$, $(\mathcal{M}_p, d_{\mathrm{GH}}^{(q)})$ *is a nowhere dense subset of* $(\mathcal{M}_q, d_{\mathrm{GH}}^{(q)})$, *i.e., the closure* $\overline{(\mathcal{M}_p, d_{\mathrm{GH}}^{(q)})}$ *has no interior in* $(\mathcal{M}_q, d_{\mathrm{GH}}^{(q)})$.

Proof. By Proposition 4.16, $\overline{(\mathcal{M}_p, d_{\mathrm{GH}}^{(q)})} = (\mathcal{M}_p, d_{\mathrm{GH}}^{(q)})$. Suppose $X \in \mathcal{M}_p$ is an interior point. Without loss of generality, by Example 4.2, we can assume that (X, d) is a finite space. Define a set $X_\varepsilon := X \cup \{x_1, x_2\}$ where x_1 and x_2 are two additional points. Pick an arbitrary point $x_0 \in X$ and define a distance function d_ε on X_ε as follows:

1. If $x, x' \in X$, then $d_\varepsilon(x, x') := d(x, x')$.
2. If $x \in X$ and $x \neq x_0$, then $d_\varepsilon(x, x_i) := d(x, x_0)$ for $i = 1, 2$.
3. $d(x_0, x_i) = \varepsilon$ and $d(x_1, x_2) = \varepsilon \boxplus_q \varepsilon$.

Let $\mathrm{sep}(X) = \min\{d(x, x') : x, x' \in X\}$. Then, it is easy to verify that $(X_\varepsilon, d_\varepsilon) \in \mathcal{M}_q$ when $\varepsilon \leq \mathrm{sep}(X) \boxplus_q \mathrm{sep}(X)$. Moreover, $X_\varepsilon \notin \mathcal{M}_p$ since x_0, x_1, x_2 does not satisfy the p-triangle inequality:

$$d_X(x_0, x_1) \boxplus_p d_X(x_0, x_2) = \varepsilon \boxplus_p \varepsilon = 2^{\frac{1}{p}}\varepsilon < 2^{\frac{1}{q}}\varepsilon = \varepsilon \boxplus_q \varepsilon = d_X(x_1, x_2).$$

Consider the correspondence R between X and X_ε defined as $R = \{(x, x) : x \in X\} \bigcup \{(x_0, x_i) : i = 1, 2\}$. Then, we have $\mathrm{dis}_q(R) = \varepsilon \boxplus_q \varepsilon$. Thus

$$\lim_{\varepsilon \to 0} d_{\mathrm{GH}}^{(q)}(X_\varepsilon, X) = 0.$$

This contradicts the assumption that X is an interior point of $\mathcal{M}_q$. $\square$

Proposition 4.17 indicates the following result stating that $\mathcal{M}_p$ is a very "thin" subset of $\mathcal{M}_q$ for $1 \leq q < p \leq \infty$. In fact, we have the following stronger result.

Theorem 4.18. *Let $q \in [1,\infty)$, then $\bigcup_{p\in(q,\infty]} \mathcal{M}_p \subsetneq \mathcal{M}_q$. In particular when $q = 1$, we have $\bigcup_{p\in(1,\infty]} \mathcal{M}_p \subsetneq \mathcal{M}$.*

Proof. Obviously, by Proposition 1.16, $\bigcup_{p\in(q,\infty]} \mathcal{M}_p \subset \mathcal{M}_q$.

Let $\{p_n\}_{n=1}^{\infty}$ be a strictly decreasing sequence with q being the limit point. Let $p_0 = \infty$. Then, we have the sequence $\mathcal{M}_{p_0} \subset \mathcal{M}_{p_1} \subset \cdots$. By Proposition 1.16, we know $\bigcup_{p\in(q,\infty]} \mathcal{M}_p = \bigcup_{n=0}^{\infty} \mathcal{M}_{p_n}$, which is a countable union of nowhere dense sets. Since $(\mathcal{M}_q, d_{\mathrm{GH}}^{(q)})$ is a complete metric space (Theorem 4.13), by the Baire category theorem, $\mathcal{M}_q \neq \bigcup_{p\in(q,\infty]} \mathcal{M}_p$. □

In the proof we know that $\bigcup_{p\in(q,\infty]} \mathcal{M}_p$ is actually a meager set of $\mathcal{M}_q$, which means that most elements of $\mathcal{M}_q$ cannot be captured by p-metric spaces with $p > q$.

Example 4.19. $[0,1]$ *with Euclidean metric does not belong to any $\mathcal{M}_p$ for $p \in (1,\infty]$.*

Example 4.20. *Consider the unit circle $\mathbb{S}^1 = \{(x,y) : x^2 + y^2 = 1\}$ on $\mathbb{R}^2$ with the Euclidean distance. Then, $\mathbb{S}^1 \in \mathcal{M}\backslash \bigcup_{p>1} \mathcal{M}_p$.*

5 Special structural properties of u_{GH}

Unlike the general family $\mathcal{M}_p$, there is a structural theorem for u_{GH} on $\mathcal{U}$ which gives rise to an algorithm for computing u_{GH} between finite ultrametric spaces in polynomial time.

Before delving into the structural theorem, let us have a closer look of distortion formula of u_{GH}.

Lemma 5.1. *Given two ultrametric spaces X and Y, the ∞-distortion of any correspondence R between them satisfies:*

$$\mathrm{dis}_\infty(R) = \inf\left\{ r \geq 0 \;\middle|\; \begin{array}{c} u_X(x,x') \leq \max(r, u_Y(y,y')) \\ \textit{and} \\ u_Y(y,y') \leq \max(r, u_X(x,x')) \end{array} \textit{for all } (x,y),(x',y') \in R \right\}. \tag{15}$$

Proof. This follows immediately from the definition of ∞-difference:

$$\begin{aligned} \Lambda_\infty(a,b) &= \begin{cases} \max(a,b), & a \neq b \\ 0, & a = b \end{cases} \\ &= \inf\{r : a \leq \max(b,r), b \leq \max(a,r)\}. \end{aligned}$$

□

We also need a special notion of quotient of ultrametric spaces. For $(X, u_X) \in \mathcal{U}$ and $t \geq 0$, we introduce a relation $\sim_t$ on X such that $x \sim_t x'$ if $u_X(x,x') \leq t$. It is easy to check that $\sim_t$ is an equivalence relation and we write $[x]_t$ for the equivalence class of x under the relation $\sim_t$.

Definition 5.2. *Given $(X, u_X) \in \mathcal{U}$ and $t \geq 0$, let (X_t, u_{X_t}) be the ultrametric space where $X_t = X/\sim_t$ and*

$$u_{X_t}([x]_t, [x']_t) := \begin{cases} u_X(x,x') & \textit{if } [x]_t \neq [x']_t \\ 0 & \textit{if } [x]_t = [x']_t. \end{cases} \tag{16}$$

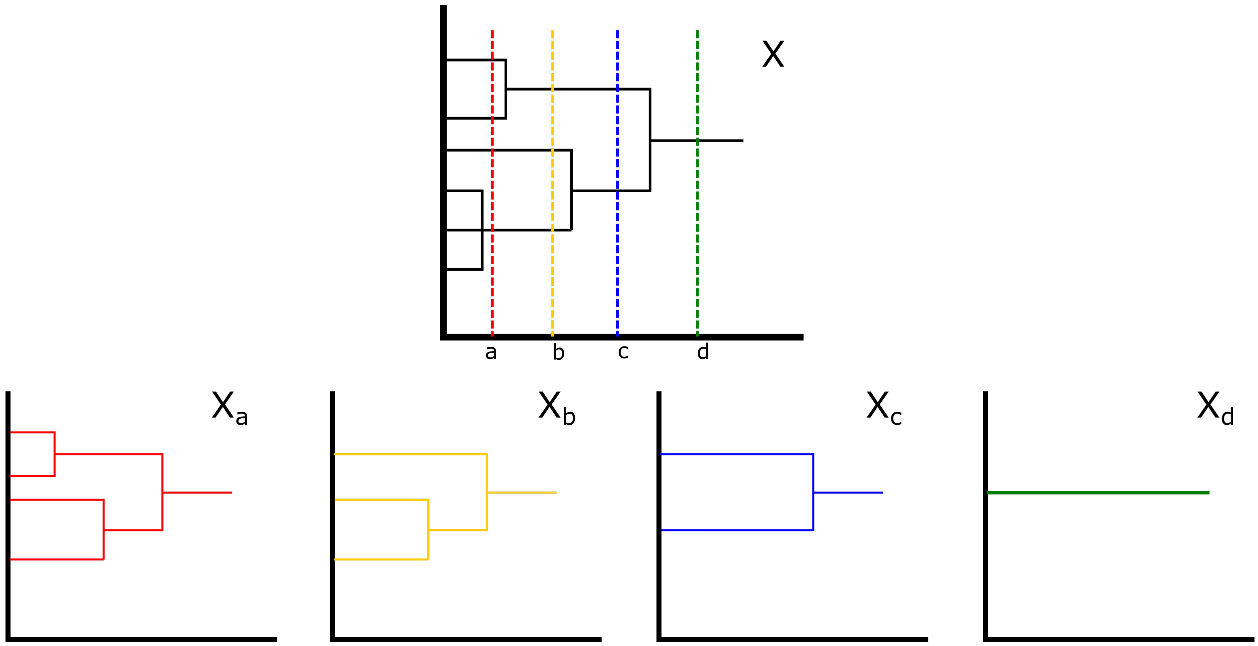

Figure 4: **Illustration of Definition 5.2.** We represent a 6-point ultrametric space X as a dendrogram in the first row of the figure. Please see Theorem 6.8 for more details. The figures on the second row show the dendrograms corresponding to X_a, X_b, X_c and X_d from left to right. As we can see from the figure, X_t forgets the structure below scale t for any $t \geq 0$.

Remark 5.3. *It follows from the ∞-triangle inequalities that u_{X_t} is a well-defined ultrametric on X_t. Since the canonical projection $X \to X_t$ is continuous, we have that (X_t, u_{X_t}) is a compact ultrametric space.*

Remark 5.4. *The equivalence relation $\sim_t$ encodes the intuitive notion of "forgetting" information below scale t. Please see Figure 4 for an illustration.*

Also, given $0 \leq t \leq s$, there is a (distance non-increasing) surjection $\iota_{t,s}^X : X_t \to X_s$ given by $[x]_t \mapsto [x]_s$.

Remark 5.5. *For any $x, x' \in X$, $[x]_t \neq [x']_t$ if and only if $u_X(x, x') > t$.*

The following lemma will be used in the sequel.

Lemma 5.6. *Let $f : X \to Y$ be a 1-Lipschitz map between two ultrametric spaces, then f induces a well defined map $f_t : X_t \to Y_t$ for any $t \geq 0$.*

Proof. Define $f_t : X_t \to Y_t$ by $f_t([x]_t^X) := [f(x)]_t^Y$. If $x' \in [x]_t^X$, then by definition we have $u_X(x, x') \leq t$. Since f is 1-Lipschitz, we have that $u_Y(f(x'), f(x)) \leq u_X(x, x') \leq t$ and thus $f(x') \in [f(x)]_t^Y$. Therefore f_t is well-defined. $\square$

5.1 A structural theorem for u_{GH}

The purpose of this section is to prove:

Theorem 5.7 (Structural theorem for u_{GH})**.** *For all $X, Y \in \mathcal{U}$ one has that*

$$u_{\mathrm{GH}}(X,Y) = \min\left\{t \geq 0 :\ (X_t, u_{X_t}) \cong (Y_t, u_{Y_t})\right\}.$$

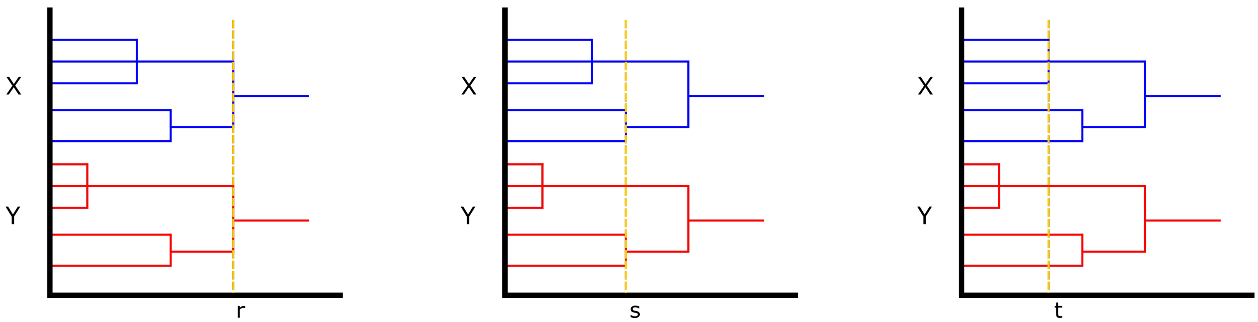

Figure 5: **(Illustration of Theorem 5.7)** We represent two ultrametric spaces X and Y as dendrograms (See Theorem 6.8 for more details.). Imagine that the yellow line is scanning from right to left to obtain quotient spaces described in Definition 5.2. It is easy to see from the figure that $X_r \cong Y_r$, $X_s \cong Y_s$, $X_t \cong Y_t$, and that t is the minimum value such that the quotients are isometric. Thus, $u_{\mathrm{GH}}(X,Y) = t$.

It follows easily from the structural theorem that one can directly determine the u_{GH} distance between two ultrametric spaces with different diameters:

Corollary 5.8. *If X and Y are compact ultrametric spaces such that* $\mathrm{diam}(X) < \mathrm{diam}(Y)$*, then* $u_{\mathrm{GH}}(X,Y) = \mathrm{diam}(Y)$.

Proof. Given $t = \mathrm{diam}(Y)$, we have $X_t \cong * \cong Y_t$. When $\mathrm{diam}(X) < t < \mathrm{diam}(Y)$, we have that $X_t \cong *$ but $Y_t \ncong *$, thus $X_t \ncong Y_t$. Therefore, by Theorem 5.7, $u_{\mathrm{GH}}(X,Y) = \mathrm{diam}(Y)$. $\square$

Proof of Theorem 5.7. We first prove a weaker version (with inf instead of min):

$$u_{\mathrm{GH}}(X,Y) = \inf\left\{t \geq 0 :\ (X_t, u_{X_t}) \cong (Y_t, u_{Y_t})\right\}. \tag{17}$$

Suppose first that $X_t \cong Y_t$ for some $t \geq 0$, i.e. there exists an isometry $f_t : X_t \to Y_t$. Then we construct a correspondence between X and Y as

$$R_t := \left\{(x,y) \in X \times Y :\ y \in f_t([x]_t^X)\right\}.$$

Equivalently, $R_t = \left\{(x,y) \in X \times Y :\ x \in f_t^{-1}([y]_t^Y)\right\}$.

It is easy to show that $\mathrm{dis}_\infty(R_t) \leq t$ which implies that $u_{\mathrm{GH}}(X,Y) \leq t$. Indeed, for $(x,y),(x',y') \in R_t$, if $u_X(x,x') \leq t$, then we already have $u_X(x,x') \leq \max(t, u_Y(y,y'))$. Otherwise, if $u_X(x,x') > t$, then by Remark 5.5 we have $[x]_t^X \neq [x']_t^X$. Thus, by definition of u_{Y_t} and u_{X_t}, and the fact that f_t is an isometry, we have $[y]_t^Y = f_t([x]_t^X) \neq f_t([x']_t^X) = [y']_t^Y$ and hence $u_Y(y,y') = u_{Y_t}([y]_t^Y,[y']_t^Y) = u_{X_t}([x]_t^X,[x']_t^X) = u_X(x,x')$. Therefore $u_X(x,x') \leq \max(t, u_Y(y,y'))$. Similarly we can show that $u_Y(y,y') \leq \max(t, u_X(x,x'))$. Hence, by Equation (15) we have $\mathrm{dis}_\infty(R_t) \leq t$. Thus $u_{\mathrm{GH}}(X,Y) \leq \inf\left\{t \geq 0 :\ X_t \cong Y_t\right\}$.

Conversely, let R be a correspondence between X an Y with $\mathrm{dis}_\infty(R) = t$. Now for any $(x,y),(x',y') \in R$ with $[x']_t^X = [x]_t^X$ (i.e. $u_X(x',x) \leq t$), then by definition of $\mathrm{dis}_\infty(R)$, we have $u_Y(y', y) \leq t$ which is equivalent to $[y]_t^Y = [y']_t^Y$. Hence, any map $f : X \to Y$ taking x to any y such that $(x, y) \in R$ is 1-Lipschitz. Then, by Lemma 5.6 f induces a well-defined map $f_t : X_t \to Y_t$, with $f_t([x]_t^X) = [f(x)]_t^Y$. There is also a well-defined map $g_t : Y_t \to X_t$ induced by a map $g : Y \to X$ with $g(y) = x$ where x is chosen such that $(x, y) \in R$. It is clear that g_t is the inverse of f_t and hence f_t is bijective. Now suppose $u_{X_t}([x]_t^X, [x']_t^X) = s > t$, which means that $u_X(x, x') = s$. Then, by the characterization of $\mathrm{dis}_\infty(R)$ given by Equation (15), $u_Y(y, y')$ is forced to be s and thus $u_{Y_t}([y]_t^Y, [y']_t^Y) = s$, where $y = f(x)$ and $y' = f(x')$. This proves that f_t is an isometry and thus $u_{\mathrm{GH}}(X, Y) \geq \inf\{t \geq 0 |\, X_t \cong Y_t\}$. □

To prove that the infimum in Equation (17) can be attained, we need the following notion of curvature sets and the corresponding metric reconstruction theorem by Gromov [Gro07].

Definition 5.9 (Curvature sets [Gro07])**.** *For a metric space X, and a positive integer n, let $\Psi_X^{(n)} : X^{\times n} \to \mathbb{R}_+^{n\times n}$ be the function given by $(x_1, \ldots, x_n) \mapsto (d_X(x_i, x_j))_{i,j=1}^n$. Then, the curvature set of order n associated to X is defined as*

$$\mathrm{K}_n(X) := \mathrm{im}\left(\Psi_X^{(n)}\right).$$

When $n = 2$, any element of $\mathrm{K}_n(X)$ is of the form of $\begin{pmatrix} 0 & d \\ d & 0 \end{pmatrix}$ for some $x, x' \in X$ and $d = d_X(x, x')$. Thus, we can equivalently represent $\mathrm{K}_2(X)$ as the set $\{d_X(x, x') : x, x' \in X\}$, which is also called the spectrum of (X, d_X): i.e. the set of all possible interpoint distances between pairs of points in X.

Theorem 5.10 (Gromov's metric space reconstruction theorem)**.** *Given two compact metric spaces X and Y, if $\mathrm{K}_n(X) = \mathrm{K}_n(Y)$ for every $n \in \mathbb{N}$, then $X \cong Y$.*

Now we can finish the proof of Theorem 5.7.

Proof that infimum in Equation (17) is a minimum. Denote $t_0 = u_{\mathrm{GH}}(X, Y)$, then for any $\delta > 0$ small, there exists $0 < \varepsilon < \delta$ such that $X_{t_0+\varepsilon} \cong Y_{t_0+\varepsilon}$. Fix a positive natural number n. Consider the curvature sets $\mathrm{K}_n(X_{t_0})$ and $\mathrm{K}_n(Y_{t_0})$ defined in Definition 5.9. For any n points $[x_1]_{t_0}^X, \cdots, [x_n]_{t_0}^X \in X_{t_0}$, without loss of generality, we can assume that $[x_i]_{t_0}^X \neq [x_j]_{t_0}^X$ for any $i \neq j$. Then, there exists $\varepsilon > 0$ small enough such that $[x_i]_{t_0+\varepsilon}^X \neq [x_j]_{t_0+\varepsilon}^X$ for all $i \neq j$ and $X_{t_0+\varepsilon} \cong Y_{t_0+\varepsilon}$. This has the following two consequences:

1. $u_{X_{t_0+\varepsilon}}\left([x_i]_{t_0+\varepsilon}^X, [x_j]_{t_0+\varepsilon}^X\right) = u_{X_{t_0}}\left([x_i]_{t_0}^X, [x_j]_{t_0}^X\right)$ for all $i \neq j$.
2. There exist $[y_1]_{t_0+\varepsilon}^Y, \cdots, [y_n]_{t_0+\varepsilon}^Y \in Y_{t_0+\varepsilon}$ such that
$$u_{Y_{t_0+\varepsilon}}\left([y_i]_{t_0+\varepsilon}^Y, [y_j]_{t_0+\varepsilon}^Y\right) = u_{X_{t_0+\varepsilon}}\left([x_i]_{t_0+\varepsilon}^X, [x_j]_{t_0+\varepsilon}^X\right)$$
for all $i \neq j$. Thus,$u_{Y_{t_0+\varepsilon}}\left([y_i]_{t_0+\varepsilon}^Y, [y_j]_{t_0+\varepsilon}^Y\right) = u_{Y_{t_0}}\left([y_i]_{t_0}^Y, [y_j]_{t_0}^Y\right)$ for all $i \neq j$.

Therefore $u_{Y_{t_0}}\left([y_i]_{t_0}^Y, [y_j]_{t_0}^Y\right) = u_{X_{t_0}}\left([x_i]_{t_0}^X, [x_j]_{t_0}^X\right)$ for all $i \neq j$ and thus

$$\left(u_{X_{t_0}}\left([x_i]_{t_0}^X, [x_j]_{t_0}^X\right)\right)_{i,j=1}^n = \left(u_{Y_{t_0}}\left([y_i]_{t_0}^Y, [y_j]_{t_0}^Y\right)\right)_{i,j=1}^n.$$

This implies that $\mathrm{K}_n(X_{t_0}) \subset \mathrm{K}_n(Y_{t_0})$. Similarly we have that $\mathrm{K}_n(Y_{t_0}) \subset \mathrm{K}_n(X_{t_0})$ and thus $\mathrm{K}_n(Y_{t_0}) = \mathrm{K}_n(X_{t_0})$. Since n is arbitrary, then by Gromov' reconstruction theorem (Theorem 5.10) we have that $X_{t_0} \cong Y_{t_0}$, which implies that $t_0 = \min\{t \geq 0 :\, X_t \cong Y_t\}$. □

Not only did we involve curvature sets for proving the structural theorem, but we now reveal that curvature sets can completely characterize u_{GH}.

First note that $\mathrm{K}_n : \mathcal{U} \to \mathbb{R}_+^{n\times n}$ is a metric invariant, i.e., if $X \cong Y$, then $\mathrm{K}_n(X) = \mathrm{K}_n(Y)$. Therefore we have

Corollary 5.11. *Given two compact ultrametric spaces X and Y, we have*

$$u_{\mathrm{GH}}(X,Y) \geq \min\{t \geq 0 : \mathrm{K}_n(X_t) = \mathrm{K}_n(Y_t)\}. \tag{18}$$

This corollary vastly generalizes [Qiu09, item (1) of Theorem 4.2] which proves an inequality similar to the case $n = 2$ of (18). The following example shows that the equality will not be obtained for general $n \in \mathbb{N}$.

Example 5.12. *Let $X = \Delta_2(1)$ and $Y = \Delta_3(1)$ be the 2-point space and 3-point space with distance 1 respectively. A simple calculation shows $\mathrm{K}_2(X) = \{0,1\} = \mathrm{K}_2(Y)$. Since $X = X_0$ and $Y = Y_0$, we have $\min\{t \geq 0 : \mathrm{K}_2(X_t) = \mathrm{K}_2(Y_t)\} = 0 < 1 = u_{\mathrm{GH}}(X,Y)$. Similarly, if we take $X = \Delta_n(1)$ and $Y = \Delta_{n+1}(1)$ for arbitrary $n \in \mathbb{N}$, then $\mathrm{K}_n(X) = \mathrm{K}_n(Y)$. Thus, $\min\{t \geq 0 : \mathrm{K}_n(X_t) = \mathrm{K}_n(Y_t)\} = 0 < 1 = u_{\mathrm{GH}}(X,Y)$.*

Taking this one step further from the case $X = \Delta_2(1)$ and $Y = \Delta_3(1)$, we will see that $\mathrm{K}_3(X_0) \neq \mathrm{K}_3(Y_0)$ and in fact $\inf\{t \geq 0 : \mathrm{K}_3(X_t) = \mathrm{K}_3(Y_t)\} = 1 = u_{\mathrm{GH}}(X,Y)$. In fact, this phenomenon is not a coincidence. If we take into account K_n for all $n \in \mathbb{N}$, then we will recover u_{GH}:

Theorem 5.13. *Given two compact ultrametric spaces X and Y, we have*

$$u_{\mathrm{GH}}(X,Y) = \sup_{n\in\mathbb{N}} \min\{t \geq 0 : \mathrm{K}_n(X_t) = \mathrm{K}_n(Y_t)\}.$$

Proof. Due to Corollary 5.11, we only need to show that $u_{\mathrm{GH}}(X,Y) \leq \sup_{n\in\mathbb{N}} \min\{t \geq 0 : \mathrm{K}_n(X_t) = \mathrm{K}_n(Y_t)\}$. We begin with a simple observation following directly from the definition of curvature sets and the quotient construction described in Definition 5.2.

Claim 3. *If $\mathrm{K}_n(X_t) = \mathrm{K}_n(Y_t)$, then $\mathrm{K}_n(X_s) = \mathrm{K}_n(Y_s)$ for $s > t$.*

Suppose to the contrary that $\sup_{n\in\mathbb{N}} \min\{t \geq 0 : \mathrm{K}_n(X_t) = \mathrm{K}_n(Y_t)\} = t_0 < u_{\mathrm{GH}}(X,Y)$. Then, by the claim above, there exists $\varepsilon > 0$ such that $t_1 := t_0 + \varepsilon < u_{\mathrm{GH}}(X,Y)$ and $\mathrm{K}_n(X_{t_1}) = \mathrm{K}_n(Y_{t_1})$ for all $n \in \mathbb{N}$. According to Gromov's reconstruction theorem (Theorem 5.10), one has that $X_{t_1} \cong Y_{t_1}$. This implies that $u_{\mathrm{GH}}(X,Y) \leq t_1$ by Theorem 5.7, which contradicts the fact that $u_{\mathrm{GH}}(X,Y) > t_1$.]

Proof of Claim 3. Given $\left(u_{X_s}([x_i]_s^X, [x_j]_s^X)\right)_{i,j=1}^n \in \mathrm{K}_n(X_s)$, consider $\left(u_{X_t}([x_i]_t^X, [x_j]_t^X)\right)_{i,j=1}^n \in \mathrm{K}_n(X_t) = \mathrm{K}_n(Y_t)$. Then, there exists a tuple $([y_1]_t^Y, \cdots, [y_n]_t^Y)$ such that for any $1 \leq i,j \leq n$ we have

$$u_{X_t}([x_i]_t^X, [x_j]_t^X) = u_{Y_t}([y_i]_t^Y, [y_j]_t^Y).$$

We have the following two cases:

1. $[x_i]_t^X \neq [x_j]_t^X$. Then, by construction of u_{X_t} we have

$$u_X(x_i, x_j) = u_{X_t}([x_i]_t^X, [x_j]_t^X) = u_{Y_t}([y_i]_t^Y, [y_j]_t^Y) = u_Y(y_i, y_j).$$

   Hence it is easy to see that for $s > t$, we have $u_{X_s}([x_i]_s^X, [x_j]_s^X) = u_{Y_s}([y_i]_s^Y, [y_j]_s^Y)$.

2. $[x_i]_t^X = [x_j]_t^X$. Then $u_{X_t}([x_i]_t^X, [x_j]_t^X) = u_{Y_t}([y_i]_t^Y, [y_j]_t^Y)$ also implies that $[y_i]_t^Y = [y_j]_t^Y$. Then obviously for $s > t$ we have $[x_i]_s^X = [x_j]_s^X$ and $[y_i]_s^Y = [y_j]_s^Y$ and thus $u_{X_s}([x_i]_s^X, [x_j]_s^X) = 0 = u_{Y_s}([y_i]_s^Y, [y_j]_s^Y)$.

The previous discussion then shows that $u_{X_s}([x_i]_s^X, [x_j]_s^X) = u_{Y_s}([y_i]_s^Y, [y_j]_s^Y)$ for any $1 \leq i, j \leq n$. Therefore $\left(u_{X_s}([x_i]_s^X, [x_j]_s^X)\right)_{i,j=1}^n = \left(u_{Y_s}([y_i]_s^Y, [y_j]_s^Y)\right)_{i,j=1}^n \in \mathrm{K}_n(Y_s)$, so $\mathrm{K}_n(X_s) \subset \mathrm{K}_n(Y_s)$. Similarly we can show $\mathrm{K}_n(Y_s) \subset \mathrm{K}_n(X_s)$ and thus $\mathrm{K}_n(X_s) = \mathrm{K}_n(Y_s)$. □

□

Hausdorff structural theorem. There exists a similar structural theorem for Hausdorff distance on ultrametric spaces.

Theorem 5.14. *Suppose X is a compact ultrametric space. For any closed subsets $A, B \subset X$, we have*

$$d_H^X(A, B) = \min\{t \geq 0 : A_t = B_t\},$$

where $A_t = \{[x]_t^X : x \in A\} \subset X_t$.

Proof. We first prove $d_H^X(A, B) = \inf\{t \geq 0 : A_t = B_t\}$.

By definition of Hausdorff distance (Definition 1.2), we know that

$$d_H^X(A, B) = \inf\{t \geq 0 : A \subset B^t, B \subset A^t\},$$

where $A^t = \{x \in X : d_X(x, A) \leq t\}$

Claim 4. $A \subset B^t, B \subset A^t$ *if and only if* $A^t = B^t$.

Assuming the claim, we have

$$d_H^X(A, B) = \inf\{t \geq 0 : A^t = B^t\}.$$

Then, we only need to show that $A^t = B^t$ if and only if $A_t = B_t$.

1. Suppose $A^t = B^t$. For any $x \in A^t$, by closeness of A, there exists $x_0 \in A$ such that $d_X(x, x_0) \leq t$. Hence, $[x]_t^X = [x_0]_t^X$ by ultrametricity. Thus, $(A^t)_t = A_t$. Similarly, $(B^t)_t = B_t$. Therefore,
$$A_t = (A^t)_t = (B^t)_t = B_t.$$

2. Suppose $A_t = B_t$. Since A is closed, one has $A^t = \bigcup_{x \in A} [x]_t^X = \bigcup_{[x]_t^X \in A_t} [x]_t^X$. Therefore,

$$A^t = \bigcup_{[x]_t^X \in A_t} [x]_t^X = \bigcup_{[x])t^X \in B_t} [x]_t^X = B^t.$$

Proof of Claim 4. First note that $(A^t)^t = A^t$ since X is an ultrametric.

The if part is obvious. As for the only if part, note that $A \subset B^t$ implies $A^t \subset (B^t)^t = B^t$. Similarly $B^t \subset A^t$ and thus $A^t = B^t$. ☐

Finally, we can use a similar argument as in the proof of Theorem 5.7 via Gromov's reconstruction theorem to show that the infimum is actually a minimum. ☐

Recall in Remark 1.14, we define a p-metric Λ_p^∞ on $\mathbb{R}^n_{\geq 0}$. When $p=\infty$, we have an ultrametric Λ_∞^n. The curvature set $\mathrm{K}_n(X)$ (Definition 5.9) is a subset of $\mathbb{R}^{n^2}_{\geq 0}$. So we can compare curvature sets $\mathrm{K}_n(X)$ and $\mathrm{K}_n(Y)$ of two ultrametric spaces X and Y via the Hausdorff distance on $(\mathbb{R}^{n^2}_{\geq 0}, \Lambda_\infty^{n^2})$.

Corollary 5.15. *For two compact ultrametric spaces X and Y we have*

$$u_{\mathrm{GH}}(X,Y) = \sup_{n\in\mathbb{N}} d_{\mathrm{H}}^{\left(\mathbb{R}^{n^2}_{\geq 0}, \Lambda_\infty^{n^2}\right)}(\mathrm{K}_n(X), \mathrm{K}_n(Y)).$$

Proof. By Theorem 5.13 and the proof of Theorem 5.14, we only need to prove that for any $t \geq 0$, $\mathrm{K}_n(X_t) = \mathrm{K}_n(Y_t)$ if and only if $(\mathrm{K}_n(X))^t = (\mathrm{K}_n(Y))^t$.

Assume $\mathrm{K}_n(X_t) = \mathrm{K}_n(Y_t)$. For any $((a_{ij}))_{i,j=1}^n \in (\mathrm{K}_n(X))^t$, there exists $x_1, \cdots, x_n \in X$ such that $\Lambda_\infty(a_{ij}, u_X(x_i,x_j)) \leq t$ by definition of $\Lambda_\infty^{n^2}$ (Remark 1.13). Then, we have

$$\begin{aligned}&\Lambda_\infty(a_{ij}, u_{X_t}([x_i]_t^X, [x_j]_t^X)) \leq \max\left(\Lambda_\infty(a_{ij}, u_X(x_i,x_j)), \Lambda_\infty(u_{X_t}([x_i]_t^X, [x_j]_t^X), u_X(x_i,x_j))\right)\\ &\leq \max\left(t, \Lambda_\infty(u_{X_t}([x_i]_t^X, [x_j]_t^X), u_X(x_i,x_j))\right) \leq t.\end{aligned}$$

The first inequality follows from the fact that Λ_∞ is an ultrametric on $\mathbb{R}$. The last inequality follows from the definition of (X_t, u_{X_t}) (Definition 5.2) and the definition of Λ_∞. Since $\mathrm{K}_n(X_t) = \mathrm{K}_n(Y_t)$, there exists $y_1, \cdots, y_n \in Y$ such that $u_{X_t}([x_i]_t^X, [x_j]_t^X) = u_{Y_t}([y_i]_t^Y, [y_j]_t^Y)$. Hence, $\Lambda_\infty(a_{ij}, u_{Y_t}([y_i]_t^Y, [y_j]_t^Y)) \leq t$ for all $i,j = 1,\cdots,n$. Then, invoking the strong inequality again, one can conclude that $\Lambda_\infty(a_{ij}, u_Y(y_i,y_j)) \leq t$ and thus $((a_{ij}))_{i,j=1}^n \in (\mathrm{K}_n(Y))^t$. Therefore, $(\mathrm{K}_n(X))^t \subset (\mathrm{K}_n(Y))^t$. Similarly $(\mathrm{K}_n(Y))^t \subset (\mathrm{K}_n(X))^t$, so $(\mathrm{K}_n(X))^t = (\mathrm{K}_n(Y))^t$.

Conversely, assume $(\mathrm{K}_n(X))^t = (\mathrm{K}_n(Y))^t$. Hence, for any sequence $x_1,\cdots,x_n \in X$, there exists a sequence $y_1,\cdots,y_n \in Y$ such that $\Lambda_\infty(u_X(x_i,x_j), u_Y(y_i,y_j)) \leq t$.

1. If both $u_X(x_i,x_j), u_Y(y_i,y_j) \leq t$, then $u_{X_t}([x_i]_t^X, [x_j]_t^X) = 0 = u_{Y_t}([y_i]_t^Y, [y_j]_t^Y)$.
2. If one of $u_X(x_i,x_j), u_Y(y_i,y_j)$ is greater than t, then by definition of Λ_∞, we have $u_X(x_i,x_j) = u_Y(y_i,y_j)$. Hence, $u_{X_t}([x_i]_t^X, [x_j]_t^X) = u_{Y_t}([y_i]_t^Y, [y_j]_t^Y)$.

Therefore, $((u_{X_t}([x_i]_t^X, [x_j]_t^X))_{i,j=1}^n \in \mathrm{K}_n(Y_t)$ and thus $\mathrm{K}_n(X_t) \subset \mathrm{K}_n(Y_t)$. Similarly, $\mathrm{K}_n(Y_t) \subset \mathrm{K}_n(X_t)$ and thus $\mathrm{K}_n(X_t) = \mathrm{K}_n(Y_t)$. ☐

5.1.1 Ultra-dissimilarity spaces

In this subsection, we will consider the collection $\mathcal{U}^w$ of ultra-dissimilarity spaces (see [SCM16] for more details), which are generalizations of ultrametric spaces. The notion of u_{GH} can be also generalized to this new collection of spaces. In the end of this subsection, we will generalize the structural theorem to u_{GH} on $\mathcal{U}^w$.

Definition 5.16 (Ultra-dissimilarity space). *A finite ultra-dissimilarity space is any pair* (X, u_X) *where* $u_X : X \times X \to \mathbb{R}_+$ *satisfies, for all* $x, x', x'' \in X$*:*

(1) ***Symmetry:*** $u_X(x, x') = u_X(x', x)$*,*

(2) ***Strong triangle inequality:*** $u_X(x, x'') \leq \max\left(u_X(x, x'), u_X(x', x'')\right)$ *,*

(3) ***Definiteness:*** $\max\left(u_X(x, x), u_X(x', x')\right) \leq u_X(x, x')$*, and the equality takes place if and only if* $x = x'$*.*

We refer to u_X *as the ultra-dissimilarity on* X*.*

Remark 5.17 (Informal interpretation). *For each* $x \in X$*, the value* $u_X(x, x)$ *can be regarded as the 'birth time' of the point* x*; when* u_X *is an actual ultrametric on* X*, all points are born at time* 0*. The value* $u_X(x, x')$ *for different points* x *and* x' *encodes the first time that the two points 'merge'. Note that then condition (3) above can be informally interpreted as encoding the property that two points can merge only after being born, and that if they merge at the same time they are born, then they are actully the same point.*

Definition 5.18. *For two ultra-dissimilarity spaces* X *and* Y*, we define* $u_{\mathrm{GH}}(X, Y)$ *by*

$$u_{\mathrm{GH}}(X, Y) = \inf_R \mathrm{dis}_\infty(R), \tag{19}$$

where $\mathrm{dis}_\infty(R)$ *is defined by Equation (15), which can be extended to the case of ultra-dissimilarity spaces without obstacle.*

Remark 5.19. *Denote the collection of all isometric classes of finite ultra-dissimilarity spaces by* $\mathcal{U}^w$*. Then,* $(\mathcal{U}^w, u_{\mathrm{GH}})$ *is an ultrametric space.*

Below, for a finite set X, $\mathrm{SubPart}(X)$ will denote the collection of all *subpartitions* of X: pairs (X', P') where X' is any subset of X and P' is a partition of X'.

Definition 5.20 (Treegrams). *A treegram* T_X *over a finite set* X *is a function* $T_X : [0, \infty) \to \mathrm{SubPart}(X)$*. For each* $t \geq 0$ *we write* $T_X(t) = (X_t, P_t)$ *and require* T_X *to satisfy the following conditions:*

1. *For* $t < s$*,* $X_t \subseteq X_s$ *and* $P_s|_{X_t}$ *is coarser than* P_t*.*
2. *There exists* $t_I > 0$ *such that* $X_{t_I} = X$ *and* $P_{t_I} = \{X\}$*.*
3. *For any* $r \geq 0$*, there exists* $\varepsilon > 0$ *such that* $T_X(r) = T_X(t)$ *for* $t \in [r, r + \varepsilon]$*.*

In analogy to Theorem 6.8, we have the following theorem.

Theorem 5.21 ([SCM16]). *Given a finite set* X*, there exists a structure preserving bijection between the collection of treegrams over* X *and the collection of ultra-dissimilarities on* X*.*

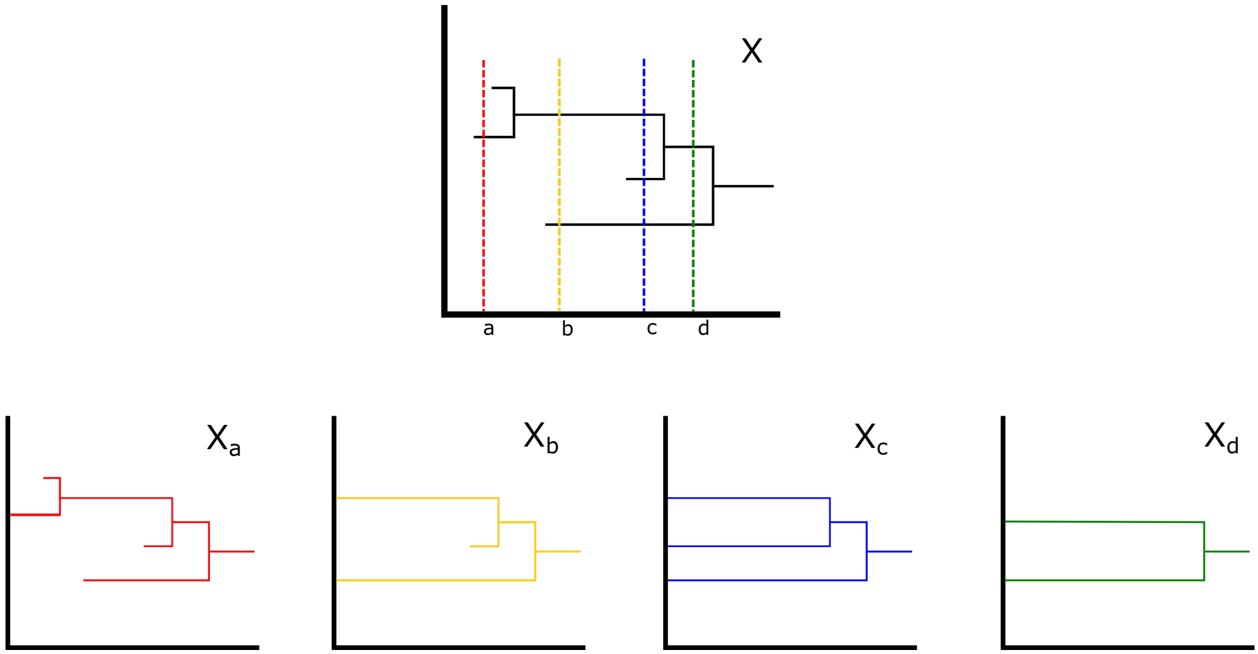

Figure 6: **Illustration of Definition 5.22.** We represent a 4-point ultra-dissimilarity space X as a treegram in the first row of the figure. The second row shows the treegrams of X_t at different times t.

Definition 5.22. *Given an ultra-dissimilarity space (X, u_X) and $t \geq 0$, let $X_t := \{[\![x]\!]_t : \forall x \in X\}$ where*

$$[\![x]\!]_t^X := \begin{cases} [x]_t & \text{if } u_X(x,x) \leq t \\ \{x\} & \text{if } u_X(x,x) > t. \end{cases} \tag{20}$$

Define by u_{X_t} an ultra-dissimilarity space on X_t^R:

$$u_{X_t}\left([\![x]\!]_t^X, [\![x']\!]_t^X\right) := \begin{cases} u_X(x,x') & \text{if } [\![x]\!]_t^X \neq [\![x']\!]_t^X, \text{ or } x = x' \text{ and } u_X(x,x) > t \\ 0 & \text{otherwise.} \end{cases} \tag{21}$$

See Figure 6 for an illustration of this process.

We use same notation X_t to denote the quotient space as in the case of ultrametric spaces (Definition 5.2) because if (X, u_X) is actually an ultrametric space, then the new definition reduces to the old one.

Theorem 5.23 (Structural theorem for u_{GH})**.** *For any two finite ultra-dissimilarity spaces X and Y one has that*

$$u_{\mathrm{GH}}(X,Y) = \min\left\{t \geq 0 : (X_t, u_{X_t}) \cong (Y_t, u_{Y_t})\right\}.$$

The proof is essentially the same with the proof of the structural theorem for u_{GH} on ultrametric spaces (Theorem 5.7).

5.2 A modified version of u_{GH}.

Theorem 5.13 actually suggests a connection with a modified version of Gromov-Hausdorff distance introduced in [Mém12], which also possesses a characterization via curvature sets. We now describe this connection.

It is known [KO99] that d_{GH} has the following distortion formula:

$$d_{\mathrm{GH}}(X,Y)=\frac{1}{2}\inf_{\substack{\varphi:X\to Y\\ \psi:Y\to X}}\max\left(\mathrm{dis}(\varphi),\mathrm{dis}(\psi),\mathrm{codis}(\varphi,\psi)\right). \tag{22}$$

By omitting the codistortion part, the computation can be reduced to solving two decoupled problems which will allow acceleration in practical application. Hence in [Mém12], the author proposed the following distance as an lower bound of d_{GH}:

$$\widehat{d}_{\mathrm{GH}}(X,Y)=\frac{1}{2}\inf_{\substack{\varphi:X\to Y\\ \psi:Y\to X}}\max\left(\mathrm{dis}(\varphi),\mathrm{dis}(\psi)\right)=\frac{1}{2}\max\left(\inf_{\varphi:X\to Y}\mathrm{dis}(\varphi),\inf_{\psi:Y\to X}\mathrm{dis}(\psi)\right).$$

It is shown that $\widehat{d}_{\mathrm{GH}}$ is a legitimate distance on the collection of isometry classes of $\mathcal{M}$ and $\widehat{d}_{\mathrm{GH}}\leq d_{\mathrm{GH}}$ whereas an inverse inequality does not exist in general. In fact, it was shown in [Mém12] that there exist finite metric spaces for which the inequality is strict.

This new distance is related to curvature sets via a structural theorem (Theorem 5.1 in [Mém12]) which shows that $\widehat{d}_{\mathrm{GH}}$ is completely characterized by curvature sets of X and Y.

Inspired by the construction of $\widehat{d}_{\mathrm{GH}}$, it is natural to consider the following modified version of u_{GH}:

$$\widehat{u}_{GH}(X,Y)=\inf_{\substack{\varphi:X\to Y\\ \psi:Y\to X}}\max\left(\mathrm{dis}_\infty(\varphi),\mathrm{dis}_\infty(\psi)\right)=\max\left(\inf_{\varphi:X\to Y}\mathrm{dis}_\infty(\varphi),\inf_{\psi:Y\to X}\mathrm{dis}_\infty(\psi)\right).$$

It is then an interesting fact that despite $d_{\mathrm{GH}}\gneq\hat{d}_{\mathrm{GH}}$ in general, the modified distance $\widehat{u}_{\mathrm{GH}}$ always coincides with u_{GH}.

Theorem 5.24. *For all X and Y in $\mathcal{U}$, we have that $\widehat{u}_{GH}(X,Y)=u_{\mathrm{GH}}(X,Y)$.*

Proof. By Theorem 3.7, we have that

$$u_{\mathrm{GH}}(X,Y)=\min_{\varphi,\psi}\max\left(\mathrm{dis}_\infty(\varphi),\mathrm{dis}_\infty(\psi),\mathrm{codis}_\infty(\varphi,\psi)\right).$$

Hence we have that $u_{\mathrm{GH}}(X,Y)\geq\widehat{u}_{GH}(X,Y)$.

Conversely, if there exist φ,ψ such that $\max\left(\mathrm{dis}_\infty(\varphi),\mathrm{dis}_\infty(\psi)\right)\leq\eta$, we need to show that $u_{\mathrm{GH}}(X,Y)\leq\eta$. Since $\mathrm{dis}_\infty(\varphi)\leq\eta$, we have that for any $x,x'\in X$,

$$\Lambda_\infty(u_X(x,x'),u_Y(\varphi(x),\varphi(x')))\leq\eta.$$

Thus, we have the following two possibilities:

1. $u_X(x,x')\neq u_Y(\varphi(x),\varphi(x'))$, and in this case neither of them is larger than η; or

2. $u_X(x,x') = u_Y(\varphi(x),\varphi(x'))$.

In either case, whenever $u_X(x,x') \leq \eta$, we have that $u_Y(\varphi(x),\varphi(x')) \leq \eta$. This is equivalent to saying that φ is 1-Lipschitz and thus φ canonically induces a map $\varphi_\eta : X_\eta \to Y_\eta$ by Lemma 5.6.

Recall from Equation (16) that we can endow X_η with a metric given by

$$u_{X_\eta}([x]_\eta,[x']_\eta) := \begin{cases} u_X(x,x') & \text{if } [x]_\eta \neq [x']_\eta \\ 0 & \text{if } [x]_\eta = [x']_\eta. \end{cases}$$

Similarly we can define u_{Y_η} on Y_η. By Remark 5.5, $[x]_\eta \neq [x']_\eta$ if and only if $u_X(x,x') > \eta$. Then, since for any x,x' such that $u_X(x,x') > \eta$ imply that $u_Y(\varphi(x),\varphi(x')) = u_X(x,x')$, it must be that

$$u_{Y_\eta}([\varphi(x)]_\eta,[\varphi(x')]_\eta) = u_Y(\varphi(x),\varphi(x')) = u_X(x,x') = u_{X_\eta}([x]_\eta,[x']_\eta).$$

Therefore φ_η is an isometric embedding. Similarly we can prove that ψ_η is an isometric embedding. This implies, by a standard argument in [BBI01, Theorem 1.6.14], that both φ_η and ψ_η are isometries, which shows $X_\eta \cong Y_\eta$. Then, by Theorem 5.7, we have that $u_{\mathrm{GH}}(X,Y) \leq \eta$.

□

6 Interleaving distances

In Section 5.1.1, we have discussed about treegrams and reviewed the relation between treegrams and ultra-dissimilarity spaces. Ultrametric spaces, as a particular type of ultra-dissimilarity spaces, possess a close relation with dendrograms, a particular type of treegrams. Let us first give a formal definition of dendrograms as follows:

Definition 6.1 (Dendrograms, [CM10])**.** *A dendrogram θ_X over a finite set X is a function $\theta_X : [0,\infty) \to \mathrm{Part}(X)$, where $\mathrm{Part}(X)$ refers to the collection of all partitions of X. We require θ_X to satisfy the following conditions:*

1. *$\theta_X(0) = \{\{x_1\}, \cdots \{x_n\}\}$.*
2. *For $t < s$, $\theta_X(t)$ is a refinement of $\theta_X(s)$.*
3. *There exists $T > 0$ such that $\theta_X(T) = \{X\}$.*
4. *For any $r \geq 0$, there exists $\varepsilon > 0$ such that $\theta_X(r) = \theta_X(t)$ for $t \in [r, r+\varepsilon]$.*

Categorically speaking, each dendrogram can be viewed as a persistent module. Let us construct a category Par of partitions as follows. The objects are pairs (X,P_X) where X is a finite set and $P_X \in \mathrm{Part}(X)$ is a partition. A morphism from (X,P_X) to (Y,P_Y) is any set map $\varphi : X \to Y$ such that $P_X \leq \varphi^* P_Y$, which means that for any element B in P_X, we have $\varphi(B) \subset C$ for some element $C \in P_Y$. Then, a dendrogram is a constructible persistent module $\theta_X : (\mathbb{R}_{\geq 0}, \leq) \to \mathrm{Par}$ in the sense of [Pat18] such that for all $t \in \mathbb{R}$, $\theta_X(t)$ has the same underlying set X. Given a point $x \in X$ and $t \geq 0$, we use $[x]_t^X$ to denote the element in $\theta_X(t)$ (a subset of X) which contains x. When the underlying set X is clear from context, we may omit the superscript X and simply use the notation $[x]_t$. It is not by coincidence that we use the same symbol as for the equivalence classes in ultrametric spaces in the last section and readers will be clear about the reason after we articulate the relation between dendrograms and ultrametric spaces.

There exists a notion of interleaving distance [BS14] between generalized persistent modules. In the particular case of dendrograms, the notion of interleaving can be defined as follows.

Definition 6.2 (Interleaving I)**.** *Given two dendrograms (X, θ_X) and (Y, θ_Y), we say they are ε-interleaved for a fixed $\varepsilon \geq 0$ if for each $t \geq 0$, there exist morphisms $\varphi_t : (X, \theta_X(t)) \to (Y, \theta_Y(t+\varepsilon))$ and $\psi_t : (Y, \theta_Y(x)) \to (X, \theta_X(t+\varepsilon))$ such that for any $x \in X$ and $y \in Y$,*

1. $[\varphi_t(x)]^Y_{s+\varepsilon} = [\varphi_s(x)]^Y_{s+\varepsilon}$ *and* $[\psi_t(y)]^X_{s+\varepsilon} = [\psi_s(y)]^X_{s+\varepsilon}$ *for any* $t \leq s$.
2. $\psi_{t+\varepsilon} \circ \varphi_t([x]^X_t) \subset [x]^X_{t+2\varepsilon}$ *and* $\varphi_{t+\varepsilon} \circ \psi_t([y]^Y_t) \subset [y]^Y_{t+2\varepsilon}$.

Remark 6.3. *In [MBW13], Morozov et al. introduced a notion of interleaving between merge trees. As dendrograms can be naturally viewed as merge trees, the two notions of interleaving agrees.*

By definition of the category Par, each morphism φ_t in the definition above is a set map $\varphi : X \to Y$ satisfying certain conditions. For $0 \leq t < s$, φ_t may be different from φ_s. It turns out that the family of set maps $\{\varphi_t\}_{t\in[0,\infty)}$ can be replaced by a single map $\varphi : X \to Y$. This observation leads us to the following definition.

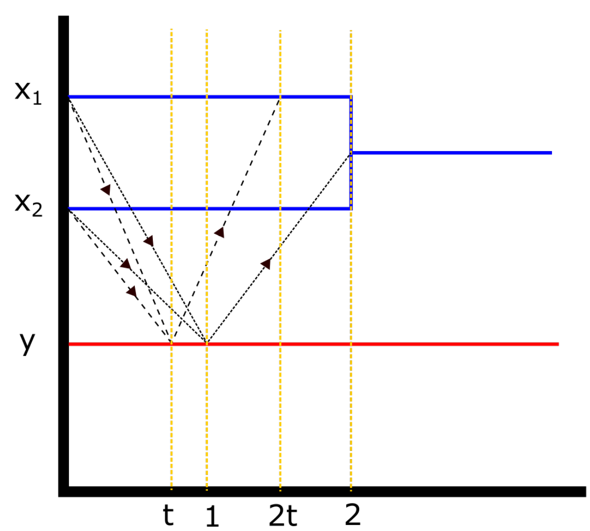

Figure 7: **(Illustration of Definition 6.4)** Here we have two dendrograms with underlying sets $X = \{x_1, x_2\}$ and $Y = \{y\}$, and $\theta_X(s) = \{\{x_1\}, \{x_2\}\}$ when $0 \leq s < 2$ and $\theta_X(s) = \{\{x_1, x_2\}\}$ when $s \geq 2$. There is only one set map $\varphi : X \to Y$ that takes both points to y, while there are two set maps from Y to X that sends y to either x_1 or x_2. Without loss of generality, we assume $\psi : Y \to X$ sending y to x_1. Then, it is easy to see from the figure that when $\varepsilon \geq 1$, φ and ψ will satisfy the conditions in Definition 6.4. However, for $t < 1$, we can see that $\psi \circ \varphi([x_2]^X_0) = \{x_1\}$ is not a subset of $[x_2]^X_{2t} = \{x_2\}$. Therefore, X and Y are not t-interleaved for $t < 1$.

Definition 6.4 (Interleaving II)**.** *Given two dendrograms (X, θ_X) and (Y, θ_Y), we say they are ε-interleaved for a fixed $\varepsilon \geq 0$ if there exist set maps $\varphi : X \to Y$ and $\psi : Y \to X$ such that $\forall t \geq 0, x \in X$ and $y \in Y$ we have*

1. $\varphi([x]_t^X) \subset [\varphi(x)]_{t+\varepsilon}^Y$ *and* $\psi([y]_t^Y) \subset [\psi(y)]_{t+\varepsilon}^X$,
2. $\psi \circ \varphi([x]_t^X) \subset [x]_{t+2\varepsilon}^X$ *and* $\varphi \circ \psi([y]_t^Y) \subset [y]_{t+2\varepsilon}^Y$.

Proposition 6.5. *Given two dendrograms (X, θ_X) and (Y, θ_Y) and $\varepsilon \geq 0$, they are ε-interleaved as in Definition 6.2 if and only if they are ε-interleaved as in Definition 6.4.*

Proof. The necessary direction is easy by taking $\varphi_t = \varphi$ and $\psi_t = \psi$ for all $t \geq 0$.

As for the sufficient direction, let $\varphi = \varphi_0$ and $\psi = \psi_0$. For any $t \geq 0$ and $x \in X$, we have that $[\varphi(x)]_{t+\varepsilon}^Y = [\varphi_0(x)]_{t+\varepsilon}^Y = [\varphi_t(x)]_{t+\varepsilon}^Y$ by item 1 in Definition 6.2. Consider the set $[x]_t^X$. It can be written as the disjoint union of subsets in $\theta_X(0)$, i.e., there exist $x_1, \cdots, x_m \in X$ such that $[x]_t^X = \cup [x_i]_0^X$. Then, we have

$$\begin{aligned}\varphi([x]_t^X) = \varphi_0([x]_t^X) = \bigcup \varphi_0([x_i]_0^X) &\subset \bigcup [\varphi_0(x_i)]_\varepsilon^Y \\ &\subset \bigcup [\varphi_0(x_i)]_{t+\varepsilon}^Y = \bigcup [\varphi_t(x_i)]_{t+\varepsilon}^Y \subset [\varphi_t(x)]_{t+\varepsilon}^Y = [\varphi(x)]_{t+\varepsilon}^Y.\end{aligned}$$

Therefore, we prove the φ part of condition 1 in Definition 6.4. The ψ part is similar.

As for the second condition in Definition 6.4, we have the following:

$$\psi \circ \varphi([x]_t^X) \subset \psi([\varphi_t(x)]_{t+\varepsilon}^Y) \subset [\psi_{t+\varepsilon} \circ \varphi_t(x)]_{t+\varepsilon}^X.$$

Then, since $\psi_{t+\varepsilon} \circ \varphi_t([x]_t^X) \subset [x]_{t+2\varepsilon}^X$, we have that $\psi_{t+\varepsilon} \circ \varphi_t(x) \in [x]_{t+2\varepsilon}^X$. Thus $[x]_{t+2\varepsilon}^X = [\psi_{t+\varepsilon} \circ \varphi_t(x)]_{t+\varepsilon}^X$. □

In the rest of our paper, we will adopt Definition 6.4 as our definition of interleaving, which is easier to analyze.

Definition 6.6. *Given two dendrograms (X, θ_X) and (Y, θ_Y), we define the interleaving distance d_{I} between them as*

$$d_{\mathrm{I}}\left((X, \theta_X), (Y, \theta_Y)\right) := \inf\{\varepsilon > 0 : (X, \theta_X) \text{ and } (Y, \theta_Y) \text{ are } \varepsilon\text{−interleaved}\}.$$

Remark 6.7. *d_{I} is a metric on the collection of all isomorphic classes of dendrograms.*

As mentioned in the introduction, there exists a close relationship between dendrograms and ultrametric spaces. Fix a finite set X, by $\mathcal{D}(X)$ denote the collection of all dendrograms over X and by $\mathcal{U}(X)$ denote the collection of all ultrametrics over X. We define a map $\Psi_X : \mathcal{D}(X) \to \mathcal{U}(X)$ by sending a dendrogram (X, θ_X) to the ultrametric u_X on X defined as follows:

$$u_X(x, x') := \inf\{t \geq 0 : [x]_t = [x']_t\}, \quad x, x' \in X.$$

Conversely, we define a map $\Phi_X : \mathcal{U}(X) \to \mathcal{D}(X)$ by sending u_X to a dendrogram θ_X as follows:

given $t \geq 0$, $x, x' \in B \in \theta_X(t)$ if $u_X(x, x') \leq t$. Thus $[x]_t = \{x' \in X : u_X(x, x') \leq t\}$ which is exactly the equivalence class of x under the relation $\sim_t$ defined right before Definition 5.2. This justifies our notation for the element in $\theta_X(t)$.

Theorem 6.8 (Dendrograms as ultrametric spaces [CM10])**.** *Given a finite set X, then $\Psi_X : \mathcal{D}(X) \to \mathcal{U}(X)$ is bijective with inverse $\Phi_X : \mathcal{U}(X) \to \mathcal{D}(X)$.*

Interleaving distance between ultrametric spaces. With the help of Theorem 6.8 above we can also define the interleaving distance d_{I} between finite ultrametric spaces (X, u_X) and (Y, u_Y) as the interleaving distance between their corresponding dendrograms generated by Φ_X and Φ_Y:

$$d_{\mathrm{I}}((X, u_X), (Y, u_Y)) \coloneqq d_{\mathrm{I}}\left((X, \Phi_X(u_X)), (Y, \Phi_Y(u_Y)\right).$$

The following theorem characterizes interleaving between ultrametric spaces completely by the distance functions.

Theorem 6.9 (Interleaving between ultrametric spaces)**.** *Two finite ultrametric spaces (X, u_X) and (Y, u_Y) are ε-interleaved if and only if there exist set maps $\varphi : X \to Y$ and $\psi : Y \to X$ such that for any $x, x' \in X$ and $y, y' \in Y$*

1. $u_Y(\varphi(x), \varphi(x')) \leq u_X(x, x') + \varepsilon$ *and* $u_X(\psi(y), \psi(y')) \leq u_Y(y, y') + \varepsilon$.
2. $u_X\left(x, \psi \circ \varphi(x)\right) \leq 2\varepsilon$ *and similarly* $u_Y\left(y, \varphi \circ \psi(y)\right) \leq 2\varepsilon$.

Remark 6.10. *If there exist $\varphi : X \to Y$ and $\psi : Y \to X$ for two compact ultrametric spaces (X, u_X) and (Y, u_Y) satisfying the two conditions in the theorem, then we say they are ε-interleaved.*

Proof of Theorem 6.9. By definition, (X, u_X) and (Y, u_Y) are ε-interleaved if and only if $(X, \theta_X) = \Phi(X, u_X)$ and $(Y, \theta_Y) = \Phi(Y, u_Y)$ are ε-interleaved. This is equivalent to the condition that there exist set maps $\varphi : X \to Y$ and $\psi : Y \to X$ such that $\forall t \geq 0, x \in X$ and $y \in Y$ we have

1. $\varphi([x]_t^X) \subset [\varphi(x)]_{t+\varepsilon}^Y$ and $\psi([y]_t^Y) \subset [\psi(y)]_{t+\varepsilon}^X$,
2. $\psi \circ \varphi([x]_t^X) \subset [x]_{t+2\varepsilon}^X$ and $\varphi \circ \psi([y]_t^Y) \subset [y]_{t+2\varepsilon}^Y$,

Recall that by construction of Φ, $x' \in [x]_t$ if and only if $u_X(x, x') \leq t$. Then the first item implies that for any $x' \in X$ such that $u_X(x, x') \leq t$, we have $u_Y(\varphi(x'), \varphi(x)) \leq t + \varepsilon$. Take $t = u_X(x, x')$ we have $u_Y(\varphi(x'), \varphi(x)) \leq u_X(x, x') + \varepsilon$. Symmetrically we have $u_X(\psi(y), \psi(y')) \leq u_Y(y, y') + \varepsilon$. As for item 2, one can derive that $u_X(x, \psi \circ \varphi(x)) \leq t + 2\varepsilon$ for any $t \geq 0$ and thus by taking $t = 0$, we obtain $u_X(x, \psi \circ \varphi(x)) \leq 2\varepsilon$ and similarly $u_Y\left(y, \varphi \circ \psi(y)\right) \leq 2\varepsilon$.

Conversely, suppose $\varphi : X \to Y$ and $g : Y \to X$ are such that the conditions in the theorem hold. Given any $t \geq 0$ and $x \in X$, suppose $u_X(x, x') \leq t$. Then

$$u_Y(\varphi(x), \varphi(x')) \leq u_X(x, x') + \varepsilon \leq t + \varepsilon,$$

which implies that $\varphi(x') \in [\varphi(x)]_{t+\varepsilon}^Y$. Therefore $\varphi([x]_t^X) \subset [\varphi(x)]_{t+\varepsilon}^Y$. Moreover,

$$\begin{aligned} u_X(\psi \circ \varphi(x'), x) &\leq \max\left(u_X(\psi \circ \varphi(x'), x'), u_X(x', x)\right) \\ &\leq \max(2\varepsilon, t) \leq t + 2\varepsilon. \end{aligned}$$

Hence $\psi \circ \varphi([x]_t^X) \subset [x]_{t+2\varepsilon}^X$. Similarly for any $y \in Y$, $\psi([y]_t^Y) \subset [\psi(y)]_{t+\varepsilon}^Y$ and $\varphi \circ \psi([y]_t^Y) \subset [y]_{t+2\varepsilon}^Y$. This shows that φ and ψ induce an ε-interleaving between (X, θ_X) and (Y, θ_Y). □

Theorem 6.9 above implies a certain structure of interleaving between ultrametric spaces which in turn provides a characterization of d_{I} in terms of I-distortion and I-codistortion of maps as follows. Given compact ultrametric spaces X and Y and a map $\varphi : X \to Y$ we define the I-distortion as follows:

$$\mathrm{dis_I}(\varphi, u_X, u_Y) \coloneqq \inf\left\{\delta \geq 0 :\ u_Y(\varphi(x), \varphi(x')) \leq u_X(x, x') + \delta,\ \forall x, x' \in X\right\}. \tag{23}$$

Given another map $\psi : Y \to X$, we define the I-codistortion of (φ, ψ) as follows:

$$\mathrm{codis_I}(\varphi, \psi, u_X, u_Y) \coloneqq \frac{1}{2}\max\left(\sup_{x\in X} u_X(x, \psi\circ\varphi(x)), \sup_{y\in Y} u_Y(y, \varphi\circ\psi(y))\right). \tag{24}$$

As before, we will use the abbreviations $\mathrm{dis_I}(\varphi)$ and $\mathrm{codis_I}(\varphi, \psi)$ when the underlying metric structures are clear.

Remark 6.11. *It is easy to check that*

$$\mathrm{dis_I}(\varphi) = \sup_{x,x'\in X}\left(u_Y(\varphi(x), \varphi(x')) - u_X(x, x')\right).$$

Hence, by Equation (9), we have that $\mathrm{dis}_I \leq \mathrm{dis}$*. Moreover,*

$$2\,\mathrm{codis_I}(\varphi, \psi) = \sup\left\{|u_X(x, \psi(y)) - u_Y(\varphi(x), y)|, x \in X, y = \varphi(x) \text{ or } y \in Y, x = \psi(x)\right\}.$$

Hence, by Equation (10), we have that $2\,\mathrm{codis_I} \leq \mathrm{codis}$*.*

Theorem 6.12. *Given* $X, Y \in \mathcal{U}$*,*

$$d_{\mathrm{I}}(X, Y) = \inf_{\varphi,\psi}\max\left(\mathrm{dis_I}(\varphi), \mathrm{dis_I}(\psi), \mathrm{codis_I}(\varphi, \psi)\right),$$

where the infimum is taken over all maps $\varphi : X \to Y$ *and* $\psi : Y \to X$*.*

Proof. We first assume that X and Y are ε-interleaved through the maps $\varphi : X \to Y$ and $\psi : Y \to X$. Then, by condition 1 of Theorem 6.9, one has $d_Y(\varphi(x), \varphi(x')) \leq d_X(x, x') + \varepsilon$ for any $x, x' \in X$ and thus $\mathrm{dis_I}(\varphi) \leq \varepsilon$. Similarly $\mathrm{dis_I}(\psi) \leq \varepsilon$. Directly from condition 2 of Theorem 6.9, we can conclude that $\mathrm{codis_I}(\varphi, \psi) \leq 2\varepsilon$. Therefore $\max\left(\mathrm{dis_I}(\varphi), \mathrm{dis_I}(\psi), \mathrm{codis_I}(\varphi, \psi)\right) \leq \varepsilon$.

Conversely, assume that $\max\left(\mathrm{dis_I}(\varphi), \mathrm{dis_I}(\psi), \mathrm{codis_I}(\varphi, \psi)\right) \leq \eta$ for $\varphi : X \to Y$ and $\psi : Y \to X$. Then by Equations (9) and (10), it is easy to check the following.

1. $u_Y(\varphi(x), \varphi(x')) \leq u_X(x, x') + \varepsilon$ and $u_X(\psi(y), \psi(y')) \leq u_Y(y, y') + \varepsilon$.
2. $u_X\left(x, \psi\circ\varphi(x)\right) \leq 2\varepsilon$ and similarly $u_Y\left(y, \varphi\circ\psi(y)\right) \leq 2\varepsilon$.

Then we can conclude that $d_{\mathrm{I}}(X, Y) \leq \varepsilon$.

□

Corollary 6.13 (Bi-Lipschitz equivalence with the Gromov-Hausdorff distance)**.** *For any two compact ultrametric spaces* X *and* Y*, we have*

$$\frac{1}{2}d_{\mathrm{I}}(X, Y) \leq d_{\mathrm{GH}}(X, Y) \leq d_{\mathrm{I}}(X, Y).$$

Example 6.14 (The bounds in the corollary are tight.)**.** *Consider the two-point spaces* $\Delta_2(2)$ *and* $\Delta_2(4)$ *with interpoint distance 2 and 4 respectively. Then it is not hard to check that* $d_{\mathrm{I}}(\Delta_2(2),*) = 1 = d_{\mathrm{GH}}(\Delta_2(2),*)$ *and* $d_{\mathrm{I}}(\Delta_2(2),\Delta_2(4)) = 2 = 2d_{\mathrm{GH}}(\Delta_2(2),\Delta_2(4))$.

Proof of Corollary 6.13. We first prove the rightmost inequality. Suppose that there exist $\varphi : X \to Y$ and $\psi : Y \to X$ inducing an ε-interleaving between X and Y. Consider the correspondence between X and Y generated by the interleaving maps φ and ψ:

$$R = \{(x,y) \in X \times Y : \varphi(x) = y \text{ or } \psi(y) = x\}.$$

Now we prove that $\mathrm{dis}(R) \leq 2\varepsilon$ which will imply that $d_{\mathrm{GH}}(X,Y) \leq \varepsilon$. Given two pairs $(x,y),(x',y') \in R$, we only need to prove that $\left|u_X(x,x') - u_Y(y,y')\right| \leq 2\varepsilon$. Due to the symmetric role of φ and ψ, we only need to check the following two cases:

1. Suppose that $y = \varphi(x)$ and $y' = \varphi(x')$. By Theorem 6.9 we have that $u_X(x,x') + \varepsilon \geq u_Y(\varphi(x),\varphi(x')) = u_Y(y,y')$. On the other hand, we have that

$$\begin{aligned} u_X(x,x') &\leq \max\left(u_X(x,\psi\circ\varphi(x)), u_X(\psi\circ\varphi(x),\psi\circ\varphi(x')), u_X(\psi\circ\varphi(x'),x')\right) \\ &\leq \max\left(2\varepsilon, u_Y(\varphi(x),\varphi(x')) + \varepsilon, 2\varepsilon\right) \leq u_Y(y,y') + 2\varepsilon. \end{aligned}$$

   Hence $\left|u_X(x,x') - u_Y(y,y')\right| \leq 2\varepsilon$.

2. Suppose that $y = \varphi(x)$ and $x' = \psi(y')$. Then

$$\begin{aligned} u_X(x,\psi(y')) &\leq \max\left(u_X(x,\psi\circ\varphi(x)), u_X(\psi\circ\varphi(x),\psi(y'))\right) \\ &\leq \max\left(2\varepsilon, u_Y(\varphi(x),y') + \varepsilon\right) \leq u_Y(y,y') + 2\varepsilon. \end{aligned}$$

   Similarly we have $u_Y(y,y') \leq t' + 2\varepsilon = u_X(x,x') + 2\varepsilon$, and thus $\left|u_X(x,x') - u_Y(y,y')\right| \leq 2\varepsilon$.

The leftmost inequality follows directly from Theorem 6.12. Assume that $d_{\mathrm{GH}}(X,Y) \leq \varepsilon$, then by Equation (22) there are two maps $\varphi : X \to Y$ and $\psi : Y \to X$ such that $\mathrm{dis}(\varphi), \mathrm{dis}(\psi), \mathrm{codis}(\varphi,\psi) \leq 2\varepsilon$. Then it is immediate that $\mathrm{dis}_{\mathrm{I}}(\varphi) \leq \mathrm{dis}(\varphi) \leq 2\varepsilon$ where the first inequality follows from Remark 6.11, and similarly $\mathrm{dis}_{\mathrm{I}}(\psi) \leq 2\varepsilon$. As for $\mathrm{codis}_{\mathrm{I}}$, we have by Remark 6.11 again that $\mathrm{codis}_{\mathrm{I}}(\varphi,\psi) \leq \frac{1}{2}\mathrm{codis}(\varphi,\psi) \leq \varepsilon$. Thus, $d_{\mathrm{I}}(X,Y) \leq 2\varepsilon$ and since $\varepsilon \geq d_{\mathrm{GH}}(X,Y)$ was arbitrary, we obtain that $d_{\mathrm{I}}(X,Y) \leq 2\, d_{\mathrm{GH}}(X,Y)$. ☐

6.1 p-interleaving distance for dendrograms and ultrametric spaces

In defining the interleaving distance between dendrograms, we used a shift operator, namely we considered an $+\varepsilon$ shift of dendrograms. Replacing $+$ by $\boxplus_p$, we will obtain the so-called p-interleaving. This new interleaving distance has an interesting relation with $d_{\mathrm{GH}}^{(p)}$.

Definition 6.15. *Given two dendrograms* (X,θ_X) *and* (Y,θ_Y)*, we say they are* (ε,p)*-interleaved for some* $\varepsilon > 0$ *and* $p \in [1,\infty]$ *if there exist set maps* $\varphi : X \to Y$ *and* $\psi : Y \to X$ *such that* $\forall t \geq 0, x \in X$ *and* $y \in Y$ *we have*

1. $\varphi([x]_t^X) \subset [\varphi(x)]^Y_{t \boxplus_p \varepsilon}$ *and* $\psi([y]_t^Y) \subset [\psi(y)]^X_{t \boxplus_p \varepsilon}$,

2. $[x]^X_{t \boxplus_p \varepsilon \boxplus_p \varepsilon} = [\psi \circ \varphi(x)]^X_{t \boxplus_p \varepsilon \boxplus_p \varepsilon}$ *and* $[y]^Y_{t \boxplus_p \varepsilon \boxplus_p \varepsilon} = [\varphi \circ \psi(y)]^Y_{t \boxplus_p \varepsilon \boxplus_p \varepsilon}$.

We then define the p-interleaving distance between (X, θ_X) and (Y, θ_Y) as

$$d_{\mathrm{I},p}\left((X, \theta_X), (Y, \theta_Y)\right) := \inf\{\varepsilon > 0 : (X, \theta_X) \text{ and } (Y, \theta_Y) \text{ are } (\varepsilon, p)\text{-interleaved.}\}$$

Remark 6.16. *Note that when $p = 1$, $(\varepsilon, 1)$-interleaving is exactly the ε-interleaving given in Definition 6.4. When $p = \infty$, the two conditions become*

1. $\varphi([x]_t^X) \subset [\varphi(x)]^Y_{\max(t,\varepsilon)}$ *and* $\psi([y]_t^Y) \subset [\psi(y)]^X_{\max(t,\varepsilon)}$,

2. $[x]^X_{\max(t,\varepsilon)} = [\psi \circ \varphi(x)]^X_{\max(t,\varepsilon)}$ *and* $[y]^Y_{\max(t,\varepsilon)} = [\varphi \circ \psi(y)]^Y_{\max(t,\varepsilon)}$.

It is easy to check that if both conditions hold for $t = \varepsilon$, then they hold for all $0 \le t \le \varepsilon$. This indicates that (ε, ∞) interleaving is performing some coarsening of dendrograms in that information corresponding to $t < \varepsilon$ is discarded. Careful readers may notice similar a phenomenon in Definition 5.2. This actually hints at a close relation between $d_{\mathrm{I}}^{(\infty)}$ and u_{GH}. See Remark 6.20.

Remark 6.17. *It easily follows Definition 6.15 that $d_{\mathrm{I},p}$ is a p-metric on the collection of all isomorphic classes of dendrograms. This is a generalization of Remark 6.7.*

A characterization result similar to Theorem 6.12 also holds for p-interleaving distance. We first define the p-I-distortion of a map $\varphi : X \to Y$:

$$\mathrm{dis}_{\mathrm{I},p}(\varphi, u_X, u_Y) := \sup_{x,x' \in X} A_p\left(u_Y(\varphi(x), \varphi(x')), u_X(x, x')\right).$$

Recall that A_p is the asymmetric p-difference defined in Equation (4).

Similarly given $\psi : Y \to X$, we define the p-I-codistortion between φ and ψ by

$$\mathrm{codis}_{\mathrm{I},p}(\varphi, \psi, u_X, u_Y) := 2^{-\frac{1}{p}} \max\left(\max_{x \in X} u_X(x, \psi \circ \varphi(x)), \max_{y \in Y} u_Y(y, \varphi \circ \psi(y))\right).$$

Same as before, we will use the abbreviations $\mathrm{dis}_{\mathrm{I},p}(\varphi)$ and $\mathrm{codis}_{\mathrm{I},p}(\varphi, \psi)$ when the underlying metric structures are clear.

Theorem 6.18. *Given $X, Y \in \mathcal{U}$ and $p \in [1, \infty]$,*

$$d_{\mathrm{I},p}(X, Y) = \inf_{\varphi,\psi} \max\left(\mathrm{dis}_{\mathrm{I},p}(\varphi), \mathrm{dis}_{\mathrm{I},p}(\psi), \mathrm{codis}_{\mathrm{I},p}(\varphi, \psi)\right),$$

where the infimum is taken over all maps $\varphi : X \to Y$ and $\psi : Y \to X$.

The proof of the theorem is essentially the same with the proof of Theorem 6.12 so we omit it here.

With this theorem, it is easy to derive the following relation between $d_{\mathrm{I},p}$ and $d_{\mathrm{GH}}^{(p)}$ in analogy with Corollary 6.13.

Corollary 6.19. *For compact ultrametric spaces (X, u_X) and (Y, u_Y), one has for $p \in [1, \infty]$*

$$2^{-\frac{1}{p}} d_{\mathrm{I},p}(X, Y) \leq d_{\mathrm{GH}}^{(p)}(X, Y) \leq d_{\mathrm{I},p}(X, Y).$$

Remark 6.20 (Relation with u_{GH} and another proof of Theorem 5.7)**.** *Note that when $p = \infty$, $1/p = 1/\infty = 0$. Then we have $2^{-\frac{1}{\infty}} = 1$ and $\delta^{(\infty)} = u_{\mathrm{GH}}$. Then Corollary 6.19 implies that*

$$d_{\mathrm{I},\infty}(X, Y) = u_{\mathrm{GH}}(X, Y)$$

for any $X, Y \in \mathcal{U}$. This statement provides us an alternative proof to Theorem 5.7:

Given two maps $\varphi : X \to Y$ and $\psi : Y \to X$ such that X and Y are (t, ∞)-interleaved, we construct $\varphi_t : X_t \to Y_t$ and $\psi_t : Y_t \to X_t$ as $\varphi_t([x]_t^X) = [\varphi(x)]_t^Y$ and $\psi_t([y]_t^Y) = [\psi(y)]_t^X$ for $x \in X$ and $y \in Y$. Then it is easy to show that these two maps are isometries and $\varphi_t = \psi_t^{-1}$. Conversely, if there are isometries $\varphi_t : X_t \to Y_t$ and $\psi_t : Y_t \to X_t$ such that $\varphi_t = \psi_t^{-1}$ at $t \geq 0$, then we construct $\varphi : X \to Y$ as follows: $\varphi(x) = y$, where y is arbitrarily chosen such that $y \in \varphi_t([x]_t^X)$. We construct $\psi : Y \to X$ similarly. Then it is routine to check that φ and ψ make X and Y be (t, ∞)-interleaved.

Remark 6.21. *If $X = *$ is the one point space, then for any $Y \in \mathcal{U}$, we have*

$$d_{\mathrm{I},p}(X, Y) = 2^{-\frac{1}{p}} \mathrm{diam}(Y).$$

Indeed, there exists only one map $\psi : Y \to X$. For any map $\varphi : X \to Y$, it is easy to check that $\mathrm{dis}_{\mathrm{I},p}(\varphi) = \mathrm{dis}_{\mathrm{I},p}(\psi) = 0$. Let $z = \varphi \circ \psi(y)$, which is invariant of choice of $y \in Y$. Since $\max_{y \in Y} u_Y(y, z) = \mathrm{diam}(Y)$, we have that $\mathrm{codis}_{\mathrm{I},p}(\varphi, \psi) = 2^{-\frac{1}{p}} \mathrm{diam}(Y)$ and thus by Theorem 6.18 we have that $d_{\mathrm{I},p}(X, Y) = 2^{-\frac{1}{p}} \mathrm{diam}(Y)$.

6.2 Interleaving distance on $\mathcal{M}$

One goal of this subsection is to generalize the interleaving distance between ultrametric spaces to a distance between general metric spaces. In fact, by the characterization theorem (Theorem 6.9) of interleaving, we can define interleaving between two ultrametric spaces directly by using their metric structures without referring to the categorical shift operator. This inspires us to make the following definition.

Definition 6.22 (Interleaving between metric spaces)**.** *Given two compact metric spaces (X, d_X) and (Y, d_Y) and $\epsilon > 0$, we say they are ε-interleaved if there exist set maps $\varphi : X \to Y$ and $\psi : Y \to X$ such that for any $x, x' \in X$ and $y, y' \in Y$*

1. *$d_Y(\varphi(x), \varphi(x')) \leq d_X(x, x') + \varepsilon$ and $d_X(\psi(y), \psi(y')) \leq d_Y(y, y') + \varepsilon$.*
2. *$u_X(x, \psi \circ \varphi(x)) \leq 2\varepsilon$ and $u_Y(y, \varphi \circ \psi(y)) \leq 2\varepsilon$, where $u_X = d_X^{(\infty)}$ and $u_Y = d_Y^{(\infty)}$.*

Definition 6.23 (Interleaving distance)**.** *We define the interleaving distance between two metric spaces X and Y as*

$$d_{\mathrm{I}}(X, Y) = \inf\{\varepsilon > 0 : X \text{ and } Y \text{ are } \varepsilon\text{-interleaved}\}.$$

Lemma 6.24. *Assume conditions in Definition 6.22. Then, we have for $u_X = d_X^{(\infty)}$ and $u_Y = d_Y^{(\infty)}$ that $u_Y(\varphi(x), \varphi(x')) \leq u_X(x, x') + \varepsilon$ and $u_X(\psi(y), \psi(y')) \leq u_Y(y, y') + \varepsilon$.*

Proof. Given any chain $x = x_0, \cdots, x_n = x'$ in X, one has

$$\begin{aligned} u_Y(\varphi(x), \varphi(x')) &\leq \max_{i=0,\cdots,n} d_Y(\varphi(x_i), \varphi(x_{i+1})) \\ &\leq \max_{i=0,\cdots,n} d_X(x_i, x_{i+1}) + \varepsilon. \end{aligned}$$

By taking infimum on the right hand side, we have that $u_Y(\varphi(x), \varphi(x')) \leq u_X(x, x') + \varepsilon$ and similarly $u_X(\psi(y), \psi(y')) \leq u_Y(y, y') + \varepsilon$. □

Theorem 6.25. *d_I is a pseudometric on the collection of all isometry classes of $\mathcal{M}$. In particular, d_I is a legitimate metric restricted on the collection of all isometry classes of finite metric spaces.*

Proof. The symmetry of d_I is obvious.

Now we prove the triangle inequality. Given finite metric spaces X, Y and Z, suppose there are $a : X \to Y$ and $b : Y \to X$ that induce an ε-interleaving between X and Y, and $f : Y \to Z$ and $g : Z \to Y$ that induce an η-interleaving between Y and Z. Then consider $f \circ a : X \to Z$ and $b \circ g : Z \to X$. Now we show that these two maps induce an $(\varepsilon + \eta)$-interleaving between X and Z.

1. First we have for any $x, x' \in X$

$$\begin{aligned} d_Z(f \circ a(x), f \circ a(x')) &\leq d_Y(a(x), a(x')) + \eta \\ &\leq d_X(x, x') + \varepsilon + \eta. \end{aligned}$$

   Similarly for any $z, z' \in Z$, we have $d_X(b \circ g(z), b \circ g(z')) \leq d_Z(z, z') + \varepsilon + \eta$.

2. For any $x \in X$,

$$\begin{aligned} u_X(x, b \circ g \circ f \circ a(x)) &\leq \max\left(u_X(x, b \circ a(x)), u_X(b \circ a(x), b \circ g \circ f \circ a(x))\right) \\ &\leq \max\left(2\varepsilon, u_Y(a(x), g \circ f \circ a(x)) + \varepsilon\right) \\ &\leq \max(2\varepsilon, 2\eta + \varepsilon) \leq 2(\varepsilon + \eta). \end{aligned}$$

   The second inequality follows from Lemma 6.24. Similarly we have that for any $z \in Z$, $u_Z(z, f \circ a \circ b \circ g(z)) \leq 2(\varepsilon + \eta)$.

Then by a standard argument of taking infimum in order, we can conclude that $d_\mathrm{I}(X, Z) \leq d_\mathrm{I}(X, Y) + d_\mathrm{I}(Y, Z)$ and thus d_I satisfies the triangle inequality.

Now, suppose $d_\mathrm{I}(X, Y) = 0$ for two finite metric spaces X and Y. Then, there exist $\varphi : X \to Y$ and $\psi : Y \to X$ such that $\mathrm{dis}_\mathrm{I}(\varphi, d_X, d_Y) = \mathrm{dis}_\mathrm{I}(\psi, d_X, d_Y) = \mathrm{codis}_\mathrm{I}(\varphi, \psi, u_X, u_Y) = 0$. By Remark 2.5 and finiteness of X and Y, we know that u_X and u_Y are legitimate metrics. Hence, $\mathrm{codis}_\mathrm{I}(\varphi, \psi, u_X, u_Y) = 0$ implies that $\psi \circ \varphi(x) = x$ and $\varphi \circ \psi(y) = y$ for any $x \in X$ and $y \in Y$. Therefore, φ and ψ are bijective and $\varphi = \psi^{-1}$. Moreover, $\mathrm{dis}_\mathrm{I}(\varphi, d_X, d_Y) = \mathrm{dis}_\mathrm{I}(\psi, d_X, d_Y) = 0$ implies that φ and ψ are 1-Lipshitz, so we have for any $x, x' \in X$

$$d_Y(\varphi(x), \varphi(x')) \leq d_X(x, x') = d_X(\psi \circ \varphi(x), \psi \circ \varphi(x)) \leq d_Y(\varphi(x), \varphi(x')).$$

This implies that $d_Y(\varphi(x), \varphi(x')) \leq d_X(x, x')$ and thus, φ is an isometry between X and Y. □

Remark 6.26. *Although the interval* $[0,1]$ *with Euclidean metric is not isometric to the one point space* $*$*, the interleaving distance between them is* 0 *(See Example 6.29). This shows that* d_{I} *is just a pseudometric on the collection of all isometric classes of* $\mathcal{M}$*.*

Remark 6.27. *The reader may wonder why we use* u_X *and* u_Y *instead of* d_X *and* d_Y *in item 2 of Definition 6.22. Suppose that we replace* u_X *and* u_Y *by* d_X *and* d_Y *and obtain a new (interleaving distance-like) discrepancy* $\widehat{d_{\mathrm{I}}}$ *such that* $\widehat{d_{\mathrm{I}}} \geq d_{\mathrm{I}}$*. Then, this discrepancy will not satisfy the triangle inequality: to see this, consider the example when* $X = \{x_1, x_2, x_3, x_4\}, Y = \{y_1, y_2\}$ *and* $Z = *$ *with* $d_X(x_1, x_4) = 6, d_X(x_i, x_{i+1}) = 2, i = 1, 2, 3$ *and* $d_Y(y_1, y_2) = 1$*.Then it is easy to check that* $\widehat{d_{\mathrm{I}}}(X,Y) = 1$*,* $\widehat{d_{\mathrm{I}}}(X,Z) = 2$ *and* $\widehat{d_{\mathrm{I}}}(Y,Z) = 1/2$*, hence* $\widehat{d_{\mathrm{I}}}(Y,Z) + \widehat{d_{\mathrm{I}}}(X,Y) = 1.5 < 2 = \widehat{d_{\mathrm{I}}}(X,Z)$ *and the triangle inequality fails. The inspiration for choosing* u_X *was drawn from [MO18], where the authors constructed an interleaving distance between filtered simplicial complexes, e.g., Vietoris-Rips complexes generated from metric spaces.*

As an analogy to Theorem 6.12, we have the following characterization of d_{I} by I-distortion of maps.

Theorem 6.28. *Given two compact metric spaces* X *and* Y*, we have that*

$$d_{\mathrm{I}}(X,Y) = \inf_{\substack{\varphi: X \to Y \\ \psi: Y \to X}} \max\left(\mathrm{dis}_{\mathrm{I}}(\varphi, d_X, d_Y), \mathrm{dis}_{\mathrm{I}}(\psi, d_X, d_Y), \mathrm{codis}_{\mathrm{I}}(\varphi, \psi, u_X, u_Y)\right),$$

where $u_X = d_X^{(\infty)}$ *and* $u_Y = d_Y^{(\infty)}$*.*

Example 6.29. *Let* $X = ([0,1], d_X)$ *be an interval with Euclidean distance and* $Y = *$ *be the one point space. Then, since* X *is geodesic, by Remark 2.19 we know* $u_X(x, x') \equiv 0$ *for any* $x, x' \in X$*. Hence, for any maps* $\varphi : X \to Y$ *and* $\psi : Y \to X$*, we have that* $\mathrm{codis}_{\mathrm{I}}^u = 0$*. It is easy to check that* $\mathrm{dis}_{\mathrm{I}}(\varphi) = \mathrm{dis}_{\mathrm{I}}(\psi) = 0$*. Therefore, by Theorem 6.28 we have that* $d_{\mathrm{I}}(X,Y) = 0$*.*

With essentially the same proof of the leftmost inequality in Corollary 6.13, we can obtain the following result.

Corollary 6.30. *Given two compact metric spaces* X *and* Y*, one has that*

$$d_{\mathrm{I}}(X,Y) \leq 2\, d_{\mathrm{GH}}(X,Y).$$

Remark 6.31. *Unlike the case with ultrametric spaces in Corollary 6.13, we do not have an inverse inequality as shown in Example 6.29.*

Remark 6.32. *If we restrict* d_{I} *to* $\mathcal{U} \times \mathcal{U}$*, then we obtain the original interleaving distance between ultrametric spaces, which justifies our usage of the same notation* d_{I}*.*

We know there exists a canonical projection $\mathfrak{S}_\infty : \mathcal{M} \to \mathcal{U}$ which is 1-Lipschitz under the Gromov-Hausdorff distance (Cf. Theorem 2.14). A natural question to to ask is what is the Lipschitz constant of this map under d_{I}. The theorem below gives a complete answer.

Theorem 6.33 ($\mathfrak{S}_\infty$ is 1-Lipschitz under d_{I})**.** *Given two finite metric spaces* X *and* Y*, we have*

$$d_{\mathrm{I}}(\mathfrak{S}_\infty(X), \mathfrak{S}_\infty(Y)) \leq d_{\mathrm{I}}(X,Y).$$

Remark 6.34. *If X and Y are ultrametric spaces, then $\mathfrak{S}_\infty(X) = X$ and $\mathfrak{S}_\infty(Y) = Y$ and thus*

$$d_{\mathrm{I}}(\mathfrak{S}_\infty(X), \mathfrak{S}_\infty(Y)) = d_{\mathrm{I}}(X, Y).$$

This shows that the coefficient in the theorem is optimal.

Remark 6.35. *In general, we will not have an opposite inequality up to some constant. Here is an example. Let $X = \{x_1, x_2, x_3\}$ and $d_X(x_1, x_3) = 2, d_X(x_i, x_{i+1}) = 1, i = 1, 2$. Let $Y = \Delta_3(1)$ be the 3-point space with interpoint distance 1. We have that $\mathfrak{S}_\infty(X) \cong \mathfrak{S}_\infty(Y) = Y$, hence $d_{\mathrm{I}}\left(\mathfrak{S}_\infty(X), \mathfrak{S}_\infty(Y)\right) = 0$. However, it is follows from Theorem 6.28 that $d_{\mathrm{I}}(X, Y) = 1$, which implies that there is no $C > 0$ such that*

$$d_{\mathrm{I}}(X, Y) \leq C \cdot d_{\mathrm{I}}\left(\mathfrak{S}_\infty(X), \mathfrak{S}_\infty(Y)\right).$$

Proof of Theorem 6.33. Let $\varepsilon > d_{\mathrm{I}}(X, Y)$. Then, there exist set maps $\varphi : X \to Y$ and $\psi : Y \to X$ such that $\mathrm{dis}_{\mathrm{I}}(\varphi, d_X, d_Y), \mathrm{dis}_{\mathrm{I}}(\psi, d_X, d_Y), \mathrm{codis}_{\mathrm{I}}(\varphi, \psi, u_X, u_Y) \leq \varepsilon$. Denote $\mathfrak{S}_\infty(X) = (\hat{X}, \hat{u}_X)$ and $\mathfrak{S}_\infty(Y) = (\hat{Y}, \hat{u}_Y)$. Recall that by definition of $\mathfrak{S}_\infty$ in Section 2.1, $\hat{X}$ and $\hat{Y}$ are quotients of X and Y by identifying points with zero distance under u_X and u_Y, respectively, and $\hat{u}_X := \hat{d}_X^{(\infty)}$ and $\hat{u}_Y := \hat{d}_Y^{(\infty)}$.

For each equivalence class in $\hat{X}$, we will fix a representative element $x \in X$. Then, we construct a set map $\hat{\varphi} : \hat{X} \to \hat{Y}$ by taking $[x]$ to $[\varphi(x)]$. Note that such a map depends on the choice of representative elements. Similarly, we construct a map $\hat{\psi} : \hat{Y} \to \hat{X}$ based on ψ and a choice of representative elements. For $\mathrm{codis}_{\mathrm{I}}$, we have that

$$2\,\mathrm{codis}_{\mathrm{I}}(\hat{\varphi}, \hat{\psi}, \hat{u}_X, \hat{u}_Y) = \max\left(\max_{[x]\in\hat{X}} \hat{u}_X\left([x], \hat{\psi}\circ\hat{\varphi}([x])\right), \max_{[y]\in\hat{Y}} \hat{u}_Y\left([y], \hat{\varphi}\circ\hat{\psi}([y])\right)\right),$$

where $[x] \in \hat{X}$ and $[y] \in \hat{Y}$ are equivalence classes with chosen representatives. Suppose $[\varphi(x)] \in \hat{Y}$ has representative $y \in Y$, which implies that $u_Y(y, \varphi(x)) = 0$, then we have

$$\begin{aligned}\hat{u}_X\left([x], \hat{\psi}\circ\hat{\varphi}([x])\right) &= u_X(x, \psi(y)) \leq \max\left(u_X(x, \psi\circ\varphi(x)), u_X(\psi\circ\varphi(x), \psi(y))\right)\\ &\leq \max\left(2\varepsilon, u_Y(\varphi(x), y) + \varepsilon\right) \leq 2\varepsilon.\end{aligned}$$

The second inequality follows from $\mathrm{codis}_{\mathrm{I}}(\varphi, \psi, d_X, d_Y) \leq 2\varepsilon$ and Lemma 6.24. Similarly we have that $\hat{u}_X\left([x], \hat{\psi}\circ\hat{\varphi}([x])\right) \leq 2\varepsilon$. Hence, $\mathrm{codis}_{\mathrm{I}}(\hat{\varphi}, \hat{\psi}, \hat{u}_X, \hat{u}_Y) \leq \varepsilon$.

To consider $\mathrm{dis}_{\mathrm{I}}(\hat{\varphi})$, choose any $[x], [x'] \in X$. By Lemma 6.24, we have that

$$u_Y\left(\varphi(x), \varphi(x')\right) \leq u_X(x, x') + \varepsilon.$$

Hence,

$$\hat{u}_Y\left(\hat{\varphi}([x]), \hat{\varphi}([x'])\right) \leq \hat{u}_X([x], [x']) + \varepsilon.$$

This implies that $\mathrm{dis}_{\mathrm{I}}(\hat{\varphi}) \leq \varepsilon$. Similarly, one has that $\mathrm{dis}_{\mathrm{I}}(\hat{\psi}) \leq \varepsilon$. Now by invoking Theorem 6.28, we have that $d_{\mathrm{I}}\left(\mathfrak{S}_\infty(X), \mathfrak{S}_\infty(Y)\right) \leq \varepsilon$. Since $\varepsilon > d_{\mathrm{I}}(X, Y)$ was arbitrary, $d_{\mathrm{I}}\left(\mathfrak{S}_\infty(X), \mathfrak{S}_\infty(Y)\right) \leq d_{\mathrm{I}}(X, Y)$. $\square$

7 Geodesic properties

In this section, we will discuss geodesic-like properties of $d_{\mathrm{GH}}^{(p)}$ and $d_{\mathrm{I},p}$. We will always assume that $1 \leq p < \infty$ unless otherwise stated.

Recall the standard concept of a geodesic (curve) [BBI01]: given a metric space (X, d_X), and $x, x' \in X$, a continuous curve $\gamma : [0,1] \to X$ with $\gamma(0) = x$ and $\gamma(1) = x'$ is called a geodesic if for all $s, t \in [0,1]$ one has $d_X(\gamma(s), \gamma(t)) = |s-t|\, d_X(x, x')$. Furthermore, one says that a metric space X is geodesic if there exists a geodesic curve connecting any two of its points.

Definition 7.1 (p-length)**.** *For a p-metric space (X, d_X) and a continuous curve $\gamma : [0,1] \to X$, we define its p-length as*

$$\mathrm{length}_p(\gamma) := \sup\left\{ \boxplus_{p\,i=1}^{n-1} d_X(\gamma(t_i), \gamma(t_{i+1})) : 0 = t_0 < t_1 < \cdots < t_n = 1 \right\}.$$

Remark 7.2. *It is clear that for any continuous curve $\gamma : [0,1] \to X$,*

$$d_X(\gamma(0), \gamma(1)) \leq \mathrm{length}_p(\gamma).$$

Definition 7.3 (p-geodesic)**.** *For a p-metric space X, a continuous curve $\gamma : [0,1] \to X$ is called a p-geodesic, if*

$$d_X(\gamma(s), \gamma(t)) = |s-t|^{\frac{1}{p}}\, d_X(\gamma(0), \gamma(1)), \forall s, t \in [0,1].$$

We say X is p-geodesic, if any two points in X can be connected by a p-geodesic.

Remark 7.4. *Note that when $p = 1$, the notion of p-geodesic reduces to the usual notion of geodesic recalled above.*

Remark 7.5. *For an ultrametric space X, it is shown in Proposition 7.12 below that every continuous curve $\gamma : [0,1] \to X$ is a trivial curve. Therefore, it is meaningless to discuss about geodesic property in ultrametric spaces.*

Lemma 7.6. *Let X be a p-metric space X and let x and x' be two distinct points in X. Then, among all curves connecting x and x', a p-geodesic has the smallest p-length.*

Proof. It is easy to show that $\mathrm{length}_p(\gamma) = d_X(\gamma(0), \gamma(1))$. Then, by Remark 7.2, we know that γ is a curve connecting x and x' with smallest p-length. □

The notions of p-geodesic and geodesic are related by the snowflake functor (Example 2.1).

Theorem 7.7. *Let X be a metric space. If X is geodesic, then $S_{\frac{1}{p}}(X)$ is p-geodesic, where $S_{\frac{1}{p}}$ denotes the $\frac{1}{p}$-snowflake functor.*

Proof. Given two point $x, x' \in X$, there exists a geodesic $\gamma : [0,1] \to X$ connecting them. Then, for any $s, t \in [0,1]$, we have

$$(d_X)^{\frac{1}{p}}(\gamma(s), \gamma(t)) = (|s-t|\, d_X(x, x'))^{\frac{1}{p}} = |s-t|^{\frac{1}{p}}\, (d_X)^{\frac{1}{p}}(x, x').$$

This implies that γ is a p-geodesic in $X^{\frac{1}{p}}$ connecting x and x'. Therefore, $S_{\frac{1}{p}}(X)$ is p-geodesic. □

Example 7.8. *$([0,l],d)$ is geodesic for any $l>0$, where d is the usual Euclidean distance on $[0,l]$. Then, by Theorem 7.7, $([0,l],d^{\frac{1}{p}})$ is p-geodesic for any $1\leq p<\infty$.*

As a generalization of Lemma 2.4.8 in [BBI01], we have the following necessary condition for p-geodesic property.

Lemma 7.9. *Let Z be a p-geodesic space. Then, for any two distinct points $z,z'\in Z$, there exists $m\in Z$ such that*

$$d_Z(z,m)=d_Z(z',m)=\left(\frac{1}{2}\right)^{\frac{1}{p}}d_Z(z,z').$$

Any such point m is called a p-midpoint between x and x'.

Proof. For any two distinct points $z,z'\in Z$, there exists a p-geodesic $\gamma:[0,1]\to Z$ connecting them. Consider $m:=\gamma\left(\frac{1}{2}\right)$. By definition of p-geodesic, we have

$$d_Z(z,m)=d_Z(z',m)=\left(\frac{1}{2}\right)^{\frac{1}{p}}d_Z(z,z').$$

□

Proposition 7.10. *Let X be a p-metric space. Suppose X is also p-geodesic. Then, for any $1\leq q<p$, X is not q-geodesic.*

Proof. Note that the proposition is trivially true when $p=1$. Suppose on the contrary that $p>1$ and that X is q-geodesic for some $1\leq q<p$. Then, by Lemma 7.9, for any two distinct points $x,x'\in X$, there exists a q-midpoint $x''\in X$ between x and x', such that

$$d_X(x,x'')=d(x',x'')=\left(\frac{1}{2}\right)^{\frac{1}{q}}d_X(x,x').$$

Therefore,

$$d_X(x,x'')\boxplus_p d_X(x'',x')=2^{\frac{1}{p}}\cdot\left(\frac{1}{2}\right)^{\frac{1}{q}}d(x,x')<d(x,x'),$$

which contradicts the fact that X is a p-metric space. □

7.1 p-metric spaces

It is shown in [CM16] that $(\mathcal{M},d_{\mathrm{GH}})$ is a geodesic space. This leads us to wondering whether $(\mathcal{M}_p,d_{\mathrm{GH}}^{(p)})$ is a geodesic space as well. The following theorem provides a complete answer.

Theorem 7.11. *$(\mathcal{M}_p,d_{\mathrm{GH}}^{(p)})$ is p-geodesic but not q-geodesic for any $q<p$. In particular, $(\mathcal{M}_p,d_{\mathrm{GH}}^{(p)})$ is not geodesic when $p>1$.*

Proof. By the previous proposition, we only need to show that $(\mathcal{M}_p,d_{\mathrm{GH}}^{(p)})$ is p-geodesic. Since by Theorem 7.7, $(\mathcal{M},d_{\mathrm{GH}})$ is geodesic, everything follows from $(\mathcal{M}_p,d_{\mathrm{GH}}^{(p)})\cong\left(\mathcal{M},(d_{\mathrm{GH}})^{\frac{1}{p}}\right)$ (Theorem 3.10). □

7.2 Ultrametric spaces

We know by Theorem 3.2 that $(\mathcal{U}, u_{\mathrm{GH}})$ is an ultrametric space. The following proposition shows that each continuous curve in an ultrametric space is trivial.

Proposition 7.12. *If X is an ultrametric space, then any continuous curve $\gamma : [0,1] \to X$ is a trivial curve, i.e., there exists $x \in X$ such that $\gamma(t) \equiv x$ for all $t \in [0,1]$.*

Proof. Let $X_0 \coloneqq \mathrm{image}(\gamma)$. We then obtain an ultrametric space $(X_0, u_X|_{X_0 \times X_0})$ by restricting u_X to $X_0 \times X_0$. Since γ is continuous, we have that X_0 is path-connected. By Proposition 2.18 we have that $\mathfrak{S}_\infty(X_0) = *$. By Proposition 2.4, $\mathfrak{S}_\infty(X_0) = X_0$. Therefore, $X_0 = *$ is a one point space and thus γ is a trivial curve. □

The proposition above precludes $(\mathcal{U}, u_{\mathrm{GH}})$ from being geodesic. However, if we consider other distance functions on $\mathcal{U}$, there may still exist geodesic structure on $\mathcal{U}$. In fact, we have:

Theorem 7.13. *$(\mathcal{U}, d_{\mathrm{GH}})$ is geodesic.*

Proof. Let X and Y be two compact ultrametric spaces. Suppose $\gamma : [0,1] \to \mathcal{M}$ is a geodesic connecting X and Y in $\mathcal{M}$. Denote $\tilde{\gamma} \coloneqq \mathfrak{S}_\infty \circ \gamma : [0,1] \to \mathcal{U}$. Then, by Proposition 2.4 we have that $\tilde{\gamma}(0) = \gamma(0) = X$ and $\tilde{\gamma}(1) = \gamma(1) = Y$. By Theorem 2.14, we have that for each $s, t \in [0,1]$,

$$d_{\mathrm{GH}}(\tilde{\gamma}(s), \tilde{\gamma}(t)) \leq d_{\mathrm{GH}}(\gamma(s), \gamma(t)) \leq |t-s|\, d_{\mathrm{GH}}(\gamma(0), \gamma(1)) = |t-s|\, d_{\mathrm{GH}}(\tilde{\gamma}(0), \tilde{\gamma}(1)).$$

Then, by triangle inequality we have that $d_{\mathrm{GH}}(\tilde{\gamma}(s), \tilde{\gamma}(t)) = |s-t|\, d_{\mathrm{GH}}(X, Y)$. This shows that $\tilde{\gamma}$ is a geodesic connecting X and Y and thus $(\mathcal{U}, d_{\mathrm{GH}})$ is geodesic. □

Remark 7.14. *More generally, we can show that $(\mathcal{U}, d_{\mathrm{GH}}^{(p)})$ is p-geodesic by modifying the previous proof slightly, e.g., replacing Theorem 2.14 in the proof by Corollary 3.9.*

7.3 Ultra-dissimilarity spaces

Theorem 7.13 can be extended to the case of ultra-dissimilarity spaces (Definition 5.1.1). In fact, ultra-dissimilarity spaces belong to a larger class of objects, called networks [Cho19]. A network (X, w_X) consists of a set X and any function $w_X : X \times X \to \mathbb{R}_{\geq 0}$. The Gromov-Hausdorff distance can be naturally generalized to compare two networks via the distortion formula (Equation 3). In [Cho19], it is shown that the collection of all finite networks equipped with d_{GH} is a geodesic space. Moreover, there exists a 1-Lipschitz projection generalizing $\mathfrak{S}_\infty$ mapping finite networks to ultra-dissimilarity spaces [SCM16]. Then, using the same technique in the proof of Theorem 7.13, one can derive the following result:

Theorem 7.15. *$(\mathcal{U}^w, d_{\mathrm{GH}})$ is geodesic.*

We can also consider the p-interleaving distance for $p \in [1, \infty]$ on $\mathcal{U}$. However, none of $d_{\mathrm{I}}^{(p)}$ will impose geodesic structure on $\mathcal{U}$. To prove this, we will use the following characterization of geodesic property.

Theorem 7.16. *$(\mathcal{U}, d_{\mathrm{I},p})$ is not geodesic for any $p \in [1, \infty]$.*

Proof. Let $X = *$ be the one point space and $Y = \Delta_2(2)$ be the two-point space with inter-point distance 2. We prove that there is no mid point between X and Y. Then, by Lemma 7.9 we have that there is no geodesic connecting X and Y.

Fix $p \in [1, \infty]$. It is easy to show that $d_{\mathrm{I},p}(X, Y) = 2^{1-\frac{1}{p}}$. Suppose there exists a 1-midpoint $Z \in \mathcal{U}$ such that $d_{\mathrm{I},p}(X, Z) = d_{\mathrm{I},p}(Y, Z) = 2^{-\frac{1}{p}}$. Then, by Remark 6.21, $\mathrm{diam}(Z) = 2^{\frac{1}{p}}\, d_{\mathrm{I},p}(X, Z) = 1$.

First consider the case when $p > 1$. By Corollary 6.19, $d_{\mathrm{I},p}(Y, Z) \geq d_{\mathrm{GH}}^{(p)}(Y, Z)$. By Remark 3.6, we have

$$d_{\mathrm{GH}}^{(p)}(Y, Z) \geq 2^{-\frac{1}{p}} \Lambda_p(\mathrm{diam}(Y), \mathrm{diam}(Z)) = 2^{-\frac{1}{p}} \Lambda_p(2, 1) > 2^{-\frac{1}{p}}.$$

Hence $d_{\mathrm{I},p}(Y, Z) \geq d_{\mathrm{GH}}^{(p)}(Y, Z) > 2^{\frac{1}{p}}$, contradiction!

Now suppose $p = 1$, then the argument above does not work since $\Lambda_p(2, 1) = 1$ when $p = 1$. Consider any two maps $\varphi : Y \to Z, \psi : Z \to Y$. If $\varphi(y_1) = \varphi(y_2)$, then

$$\mathrm{codis}_{\mathrm{I},1}(\varphi, \psi) \geq \frac{1}{2} \max(u_Y(y_1, \psi \circ \varphi(y_1)), u_Y(y_2, \psi \circ \varphi(y_2)) = 1.$$

Otherwise suppose $z_1 := \varphi(y_1) \neq \varphi(y_2) =: z_2$. Since $\mathrm{diam}(Z) = 1$, we have that $u_Z(z_1, z_2) \leq 1$. If $\psi(z_i) = y_i$ for $i = 1, 2$, then $\mathrm{dis}_{\mathrm{I},i}(\psi) \geq 1$. Otherwise,

$$\mathrm{codis}_{\mathrm{I},1}(\varphi, \psi) \geq \frac{1}{2} \max(u_Y(y_1, \psi \circ \varphi(y_1)), u_Y(y_2, \psi \circ \varphi(y_2))) = 1.$$

In conclusion, $d_{\mathrm{I}}(Y, Z) \geq 1 > \frac{1}{2}$ by Theorem 6.18, contradiction! □

Remark 7.17. *We can modify the case of $p = 1$ in the proof above to show that $(\mathcal{U}, d_{\mathrm{I},p})$ is not p-geodesic for all $p \in [1, \infty)$.*

8 Computing $d_{\mathrm{GH}}^{(p)}$ on $\mathcal{U}$

In this section, we investigate algorithms for computing $d_{\mathrm{GH}}^{(p)}$ between finite ultrametric spaces for $p \in [1, \infty]$. Matlab implementations of Algorithm 1 and Algorithm 8 below have been made available at

https://github.com/ndag/ultrametrics.

Data structure for ultrametric spaces. A dendrogram automatically induces a rooted tree [CM10] as one can see in the graphic representation (see Figure 8 for example). Each vertex of the tree corresponds to a closed ball in the underlying ultrametric space, e.g., the root corresponds to the whole space whereas a leaf corresponds to a singleton. So we represent a dendrogram and thus an ultrametric space X by a special *weighted tree* data structure. This is a self-referential tree structure [KR06, Chapter 6.5] in that each node contains a diameter value, a list of pointers that refer to its children, and a subset of X. The subset is the closed ball which the node corresponds to whereas the diameter value is exactly the diameter of this ball. This data structure of ultrametric spaces has certain computational advantages over the distance matrix representation. For example, one can easily perform quotient operations. See Appendix B for algorithms for fundamental operations on ultrametric spaces.

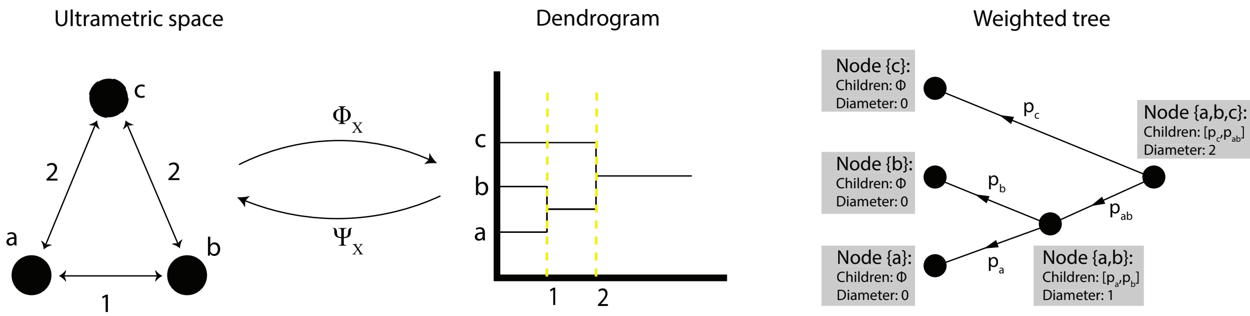

Figure 8: **(Transforming ultrametric spaces into dendrograms/weighted trees.)**

8.1 Computation of d_{GH} on $\mathcal{U}$

We consider the following two problems.

> **Decision Problem** (d_{GH} **distance computation on** $\mathcal{U}$ **(GHDU-dec))**
> **Input:** Finite ultrametric spaces (X, u_X) and (Y, u_Y) and $\varepsilon \geq 0$.
> **Question:** Is there a correspondence R between X and Y such that $\mathrm{dis}(R) \leq \varepsilon$?

> **Optimization Problem** (d_{GH} **distance computation on** $\mathcal{U}$ **(GHDU-opt))**
> **Input:** Finite ultrametric spaces (X, u_X) and (Y, u_Y).
> **Output:** The value $d_{\mathrm{GH}}(X, Y)$.

We mainly focus on solving the decision problem whereas the optimization problem can be solved based on the solution of the decison problem (see Section 8.1.4).

Base cases for Problem GHDU-dec. Based on standard properties of d_{GH} [Mém12, Theorems 3.3 and 3.4], **GHDU-dec** can be solved immediately in the following two special base cases:

1. If $|\mathrm{diam}(X) - \mathrm{diam}(Y)| > \varepsilon$, then there exists no correspondence R between X and Y with $\mathrm{dis}(R) \leq \varepsilon$.
2. If $\max(\mathrm{diam}(X), \mathrm{diam}(Y)) \leq \varepsilon$, then any correspondence R between them has $\mathrm{dis}(R) \leq \varepsilon$.

Note that the situation when one of the spaces is the one point space will automatically fall into the above two base cases.

8.1.1 Breaking GHDU-dec into subproblems

We begin with an observation (Theorem 8.2) regarding the distortion of correspondences between ultrametric spaces which will allow us to 'break' the decision problem **GHDU-dec** into subproblems.

An 'open' equivalence relation. We need the following open equivalence relation to proceed with the description of our algorithms. Given $X \in \mathcal{U}$ and $t > 0$, let $\sim_{\mathring{t}}$ be the equivalence relation on X induced by the relation $\{(x,x') : u_X(x,x') < t\}$. The difference from the equivalence relation $\sim_t$ defined before is that we now require strict inequality. Denote $X_{\mathring{t}} := X/\sim_{\mathring{t}}$, and $[x]^X_{\mathring{t}} := \{x' \in X : u_X(x,x') < t\}$ the equivalence class of x in $X_{\mathring{t}}$. We will use the abbreviation $[x]_{\mathring{t}} = [x]^X_{\mathring{t}}$ when the underlying set is clear from the context. Then, we define an ultrametric $u_{X_{\mathring{t}}}$ on $X_{\mathring{t}}$ by

$$u_{X_{\mathring{t}}}([x]_{\mathring{t}}, [x']_{\mathring{t}}) := \begin{cases} u_X(x,x') & \text{if } [x]_{\mathring{t}} \neq [x']_{\mathring{t}} \\ 0 & \text{if } [x]_{\mathring{t}} = [x']_{\mathring{t}}. \end{cases} \tag{25}$$

As a convention, when $t = 0$, we define $X_{\mathring{0}} := X$.

Remark 8.1. *Obviously, for any $x \in X$ and $t \geq 0$, we have $[x]_{\mathring{t}} \subset [x]_t$, where $[x]_t$ is an element of X_t, the 'closed' quotient defined in Definition 5.2. Be aware that we will be using both the 'open' and the 'closed' equivalence relation in this section.*

Inspired by the structural theorem for u_{GH} (Theorem 5.7), we establish the following structural theorem for d_{GH}.

Theorem 8.2 (Structural theorem for d_{GH})**.** *Let $(X,u_X),(Y,u_Y) \in \mathcal{U}$. For each $\varepsilon \in [0, \mathrm{diam}(Y))$ let $\delta_0(Y) := \mathrm{diam}(Y)$ and $\delta_\varepsilon(Y) := \delta_0(Y) - \varepsilon$ and write $X_{\mathring{\delta}_\varepsilon(Y)} = \{X_i\}_{i\in I}$ and $Y_{\mathring{\delta}_0(Y)} = \{Y_j\}_{j\in J}$.*

Assume $|\mathrm{diam}(X) - \mathrm{diam}(Y)| \leq \varepsilon$. Then, there exists a correspondence R between X and Y with $\mathrm{dis}(R) \leq \varepsilon$ if and only if there exist a surjection $\Psi : I \twoheadrightarrow J$ and for every $j \in J$ a correspondence R_j between $\left(X_{\Psi^{-1}(j)}, u_X|_{X_{\Psi^{-1}(j)} \times X_{\Psi^{-1}(j)}}\right)$ and $\left(Y_j, u_Y|_{Y_j \times Y_j}\right)$ with $\mathrm{dis}(R_j) \leq \epsilon$, where for each $j \in J$, $X_{\Psi^{-1}(j)} := \bigcup_{i \in \Psi^{-1}(j)} X_i$.

Proof. For simplicity of notation, we abbreviate $\delta_\varepsilon := \delta_\varepsilon(Y)$ and $\delta_0 := \delta_0(Y)$. Suppose that there exists a correspondence R between X and Y such that $\mathrm{dis}(R) \leq \varepsilon$. Then, with the notation in the statement, for any $x, x' \in X_i$ and $i \in I$, we have that $u_X(x,x') < \delta_\varepsilon$. Suppose $y, y' \in Y$ are such that $(x,y),(x',y') \in R$. Then $u_Y(y,y') \leq u_X(x,x') + \varepsilon < \delta_\varepsilon + \varepsilon = \delta_0$. Hence, there exists some $j \in J$ such that $y, y' \in Y_j$. Then it is easy to check that the following map is well defined: $\Psi : I \to J$ by $\Psi(i) = j$ if there exists $(x,y) \in R$ with $x \in X_i$ and $y \in Y_j$. Ψ must be surjective since R is a correspondence. Now we can restrict R to $X_{\Psi^{-1}(j)} \times Y_j$ to obtain a correspondence R_j between $X_{\Psi^{-1}(j)}$ and Y_j. The distortion of R_j is bounded above by ε, which follows from the fact $\mathrm{dis}(R) \leq \varepsilon$.

Conversely, suppose we have a surjection $\Psi : I \to J$ and correspondences R_j between $X_{\Psi^{-1}(j)}$ and Y_j with $\mathrm{dis}(R_j) \leq \varepsilon$ for each $j \in J$. Then, we construct a correspondence R between X and Y as

$$R = \bigcup_{j \in J} R_j.$$

That R is a correspondence follows because

$$p_X\left(\bigcup_{j\in J} R_j\right) = \bigcup_{j\in J} p_X(R_j) = \bigcup_{j \in J} X_{\Psi^{-1}(j)} = X$$

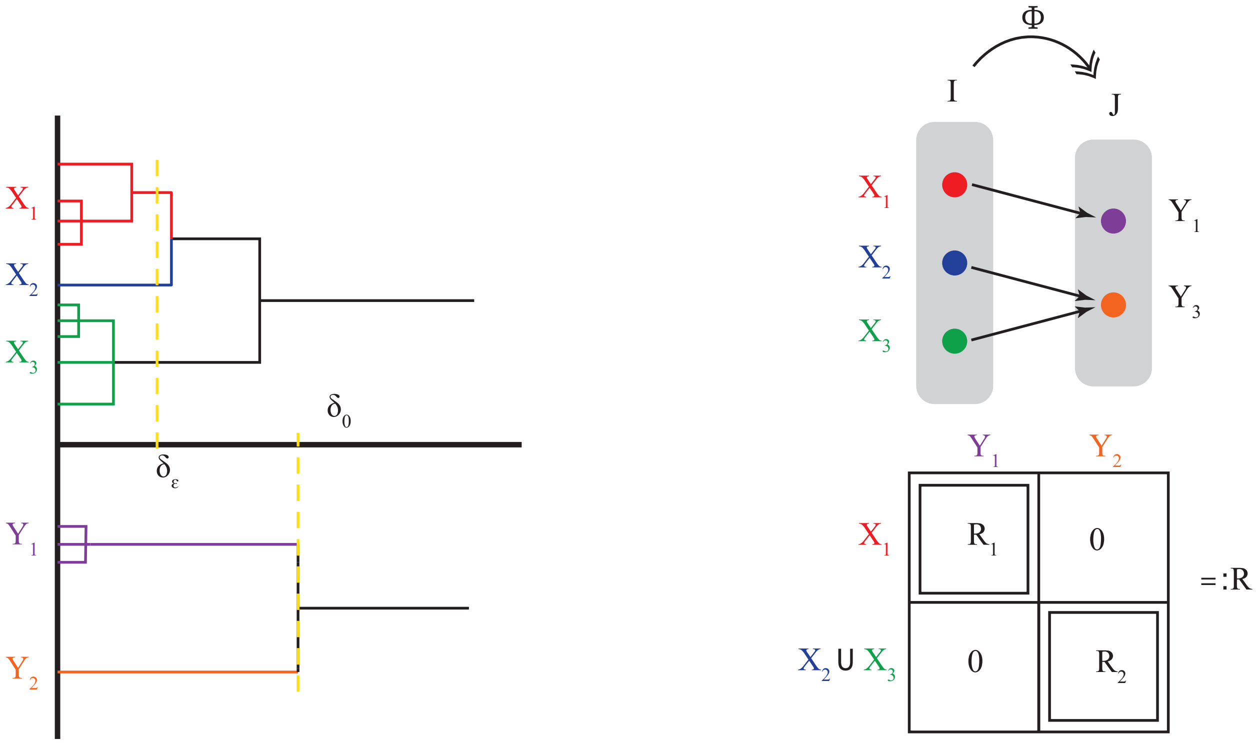

Figure 9: **(Illustration of Theorem 8.2.)** See statement of Theorem 8.2.

and

$$p_Y\left(\bigcup_{j\in J} R_j\right) = \bigcup_{j\in J} p_Y(R_j) = \bigcup_{j\in J} Y_j = Y,$$

where $p_X : X \times Y \to X$ and $p_Y : X \times Y \to Y$ are the canonical projections.

Given a pair $(x, y), (x', y') \in R$, suppose $(x, y) \in R_j$ and $(x', y') \in R_{j'}$ for some $j, j' \in J$. We can establish that $|u_X(x, x') - u_Y(y, y')| \leq \varepsilon$ by analyzing the following two cases:

1. if $j = j'$, then
$$|u_X(x, x') - u_Y(y, y')| \leq \mathrm{dis}(R_j) \leq \varepsilon.$$

2. if $j \neq j'$, then $\mathrm{diam}(Y) - \varepsilon = \delta_\varepsilon \leq u_X(x, x')$ and $u_Y(y, y') = \delta_0 = \mathrm{diam}(Y)$. By assumption that $|\mathrm{diam}(X) - \mathrm{diam}(Y)| \leq \varepsilon$, we have that $u_X(x, x') \leq \mathrm{diam}(X) \leq \mathrm{diam}(Y) + \varepsilon$. Hence, we have that
$$|u_X(x, x') - u_Y(y, y')| \leq \varepsilon.$$

Therefore $\mathrm{dis}(R) \leq \varepsilon$. □

Application of Theorem 8.2. Now, suppose we are given two ultrametric spaces X and Y and $\varepsilon \geq 0$ not falling in either of the two base cases mentioned above. Note that this implies that one of $\mathrm{diam}(X)$ or $\mathrm{diam}(Y)$ must be strictly larger than ε. The structural theorem for d_{GH} on $\mathcal{U}$ (Theorem 8.2) suggests a divide-and-conquer approach to solve Problem **GHDU-dec** with such input (note that in the assumptions of Theorem 8.2 we already require that $|\mathrm{diam}(X) - \mathrm{diam}(Y)| \leq \varepsilon$).

Suppose $\operatorname{diam}(Y) > \varepsilon$ (otherwise we swap the roles of X and Y), then we can take the open quotient of X and Y to obtain $X_{\mathring{\delta}_\varepsilon(Y)} = \{X_i\}_{i\in I}$ and $Y_{\mathring{\delta}_0(Y)} = \{Y_j\}_{j\in J}$ using the same notation as in Theorem 8.2. If there is no surjection from I to J, i.e., $\#I < \#J$, then we already conclude that there is no correspondence between X and Y with distortion bounded above by ε. Otherwise, for each surjection $\Psi : I \to J$, we solve an instance of the decision problem **GHDU-dec** with input $(X_{\Psi^{-1}(j)}, Y_j, \varepsilon)$. If for some surjection Ψ, there exist correspondences R_j between $X_{\Psi^{-1}(j)}$ and Y_j with $\operatorname{dis}(R_j) \le \varepsilon$ for all $j \in J$, we take the union of R_j to obtain an correspondence R between X and Y with $\operatorname{dis}(R) \le \varepsilon$. Otherwise, there exists no such a correspondence. As for $(X_{\Psi^{-1}(j)}, Y_j)$, it is easy to see that $\#X_{\Psi^{-1}(j)} < \#X$ and $\#Y_j < \#Y$. So if we repeatedly apply the quotient operation in Theorem 8.2, we will eventually reduce the problem to one of the base cases.

8.1.2 A recursive algorithm for the decision problem GDHU-dec

From the analysis above we directly identify a recursive algorithm $\mathbf{FindCorrespondence}$. The pseudocode for this algorithm is given Algorithm 1. The algorithm takes as input two ultrametric spaces X and Y and a parameter $\varepsilon \ge 0$. If there exists a correspondence R between X and Y with $\operatorname{dis}(R) \le \varepsilon$, then $\mathbf{FindCorrespondence}(X, Y, \varepsilon)$ returns such a correspondence. If there exists no such a correspondence, $\mathbf{FindCorrespondence}(X, Y, \varepsilon)$ returns 0.

In the pseudocode, correspondences are represented by 0–1 matrices so that non-zero entries encode pairs in the correspondence, and for natural numbers n and m, $\mathbb{1}_{n,m}$ is the all-ones $n \times m$ matrix. The function $\mathbf{PartitionOpen}(X, t)$ will partition the ultrametric space X according to the open relation $\sim_{\mathring{t}}$ (see Appendix B).

Complexity analysis. Now we show that under certain restrictions on the input ultrametric spaces, the time complexity of Algorithm 1 is polynomial in $\max(\#X, \#Y)$.

For each $0 < s \le t$ consider the structure map $\iota_{s,t} : X_{\mathring{s}} \to X_t$ given by

$$[x]_{\mathring{s}} \to [x]_t, \ x \in X.$$

For $x \in X$ and $t \ge s > 0$, the *children* of $[x]_t$ in $X_{\mathring{s}}$ are the classes in $\iota_{s,t}^{-1}([x]_t)$. Please see Figure 10 for an illustration.

Definition 8.3 (First (ε, γ)-growth condition)**.** *For $\varepsilon \ge 0$, and $\gamma > 1$, we say that an ultrametric space (X, u_X) satisfies the first (ε, γ)-growth condition (FGC) if for all $x \in X$, and $t > \varepsilon$(or $\ge$),*

$$\#[x]_t \le \gamma \cdot \#[x]_{(t\mathring{-}\varepsilon)}.$$

Note that the left hand side of the inequality we used the 'closed' equivalence class whereas on the right hand side we use the 'open' equivalence class. See Figure 11 for an illustration.

Remark 8.4 (Interpretation of FGC)**.** *The main idea behind the first (ε, γ)-growth condition is that for each $t > 0$ we want to have some degree of control over both the size and the number of children of each class in X_t. More precisely, suppose $[x]_{\mathring{t}}$ has children $[x_i]_{(t\mathring{-}\varepsilon)}$, for $i = 1, \cdots, N$. First, from the definition of children, we have $[x]_t = [x_i]_t$ and thus $\#[x_i]_{(t\mathring{-}\varepsilon)} \ge \frac{\#[x]_t}{\gamma}$. This means*

Algorithm 1: $\textbf{FindCorrespondence}(X, Y, \varepsilon)$

```
BoolSwap ← FALSE
if diam(X) > diam(Y) then
    Swap X and Y; BoolSwap ← TRUE
if max(diam(X), diam(Y)) ≤ ε then
    R ← 𝟙_{#X,#Y}
    if BoolSwap then
        Transpose R
    return R
if |diam(X) − diam(Y)| > ε then
    return 0
{X_i}_{i∈I} = PartitionOpen(X, δ_ε(Y))
{Y_j}_{j∈J} = PartitionOpen(Y, δ_0(Y))
for Surjection Ψ : I ↠ J do
    for j ∈ J do
        R_j ← FindCorrespondence(X_{Ψ^{-1}(j)}, Y_j, ε)
    if ((R_j != 0) ∀j) then
        R ← ⋃_{j=1}^{#J} R_j
        if BoolSwap then
            Transpose R
        return R
return 0
```

for any class $[x]_t$ at scale t, each of its children at scale $t-\varepsilon$ contains at least as much as a fixed proportion $\frac{1}{\gamma}$ of the number of points in its parent $[x]_t$. Moreover, we have

$$\#[x]_t = \sum_{i=1}^{N} \#[x_i]_{(t \mathring{-} \varepsilon)} \geq \frac{N}{\gamma} \#[x]_t.$$

Therefore $N \leq \gamma$, which implies that each $[x]_t$ has at most γ children at $t-\varepsilon$.

Given $\gamma > 1$ let $\mathrm{b}(\gamma) := \frac{\gamma^2}{\gamma^2-1}$.

Theorem 8.5. *Assume that (X, u_X) and (Y, u_Y) both satisfy the first (ε,γ)-growth condition for some $\varepsilon \geq 0$ and $\gamma \geq 4$. Then, the algorithm* FindCorrespondence*(X,Y,ε) runs in time $O\left(n^{\gamma \log_{\mathrm{b}(\gamma)} \gamma}\right)$, where $n = \max(\#X, \#Y)$.*

Proof. We are going to invoke the master theorem [CLRS09] to analyze the complexity of our recursive algorithm.

Suppose $n = \max(\#X, \#Y)$. If (X, Y) is one of the base cases, the algorithm stops in time $O(n^2)$. Now assume that $\mathrm{diam}(Y) \geq \mathrm{diam}(X)$ (otherwise it takes time $O(n)$ to swap X and Y). Then, it takes time $O(n)$ to partition X and Y to $X_{\mathring{\delta}_\varepsilon(Y)} = \{X_i\}_{i\in I}$ and $Y_{\mathring{\delta}_0(Y)} = \{Y_j\}_{j\in J}$ via $\textbf{PartitionOpen}$ (see Appendix B).

By FGC, we have that $\max(\#I, \#J) \leq \gamma$. There will be at most γ^γ surjections $\Psi : I \to J$ and thus γ^γ subproblems. Fix a surjection $\Psi : I \to J$. Then each $Y_j = [y]^Y_{\mathring{\delta}_0}$ for some $y \in Y$. So

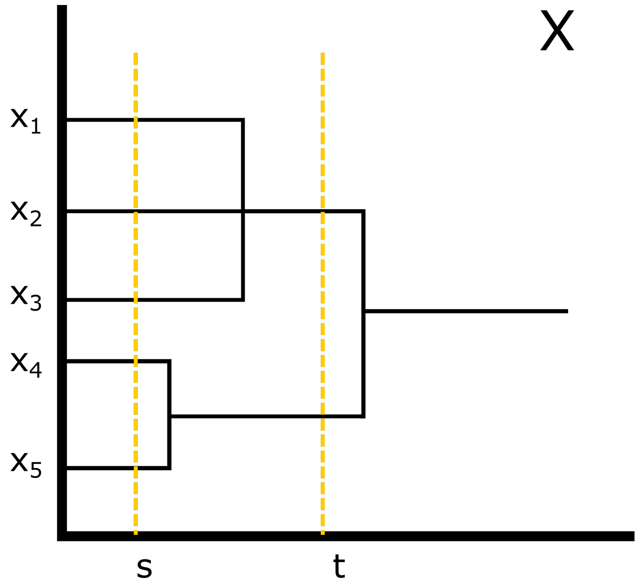

Figure 10: Here we represent a five-point ultrametric space X as a dendrogram. The children of $[x_1]_t$ in $X_{\mathring{s}}$ are $[x_1]_{\mathring{s}}$, $[x_2]_{\mathring{s}}$ and $[x_3]_{\mathring{s}}$. The children of $[x_4]_t$ in $X_{\mathring{s}}$ are $[x_4]_{\mathring{s}}$ and $[x_5]_{\mathring{s}}$.

$\#[y]^Y_{\mathring{\delta}_0} \leq \frac{\#Y}{\gamma}$ since $[y]^Y_{\mathring{\delta_0+\varepsilon}} = Y$. Now for $X_{\Psi^{-1}(j)}$, since Ψ is a surjection and $\#J \geq 2$, there exists $i \notin \Psi^{-1}(j)$ such that $X_i \bigcap X_{\Psi^{-1}(j)} = \emptyset$. Assume $X_i = [x_i]^X_{\mathring{\delta}_\varepsilon}$ for some $x_i \in X$. Then

$$\#[x_i]^X_{\mathring{\delta}_\varepsilon} \geq \frac{\#[x_i]^X_{\mathring{\delta_\varepsilon+2\varepsilon}}}{\gamma^2} = \frac{\#X}{\gamma^2},$$

since $\delta_\varepsilon + 2\varepsilon = \mathrm{diam}(Y) + \varepsilon > \mathrm{diam}(X)$. Therefore $\#X_{\Psi^{-1}(j)} \leq \left(1 - \frac{1}{\gamma^2}\right) \#X$. Inside each loop, we also need at most $O(n^2)$ time to generate R_j and R, transpose R and create unions $X_{\Psi^{-1}(j)}$.

Denote by $T(n)$ the time complexity of the algorithm with $n = \max(\#X, \#Y)$. Then, according to the above analysis and the fact that $1 - \frac{1}{\gamma^2} > \frac{1}{\gamma}$ when $\gamma \geq 4$, we have

$$T(n) \leq \gamma^\gamma \, T\left(\frac{n}{\gamma^2/(\gamma^2-1)}\right) + O\left(n^2\right).$$

The critical exponent $\log_{\frac{\gamma^2}{\gamma^2-1}} \gamma^\gamma$ is strictly greater than 2 since $\gamma \geq 4$, therefore by the master theorem we have that $T(n) = O\left(n^{\gamma \log_{\mathrm{b}(\gamma)} \gamma}\right)$. $\square$

8.1.3 A dynamic programming algorithm for the decision problem GDHU-dec

Note that in Algorithm 1, for different surjections $\Psi_1, \Psi_2 : I \to J$, there could be some $j \in J$ such that $\Psi_1^{-1}(j) = \Psi_2^{-1}(j)$. This would result in repetitive computation of $\mathbf{FindCorrespondence}(X_{\Psi_1^{-1}(j)}, Y_j, \varepsilon)$. To eliminate such repetition, we devise a dynamic programming algorithm for the decision problem **GDHU-dec** based on Theorem 8.2. The dynamic programming algorithm has less time complexity than the recursive algorithm. In fact, we will prove that the dynamic programming algorithm is fixed-parameter-tractable.

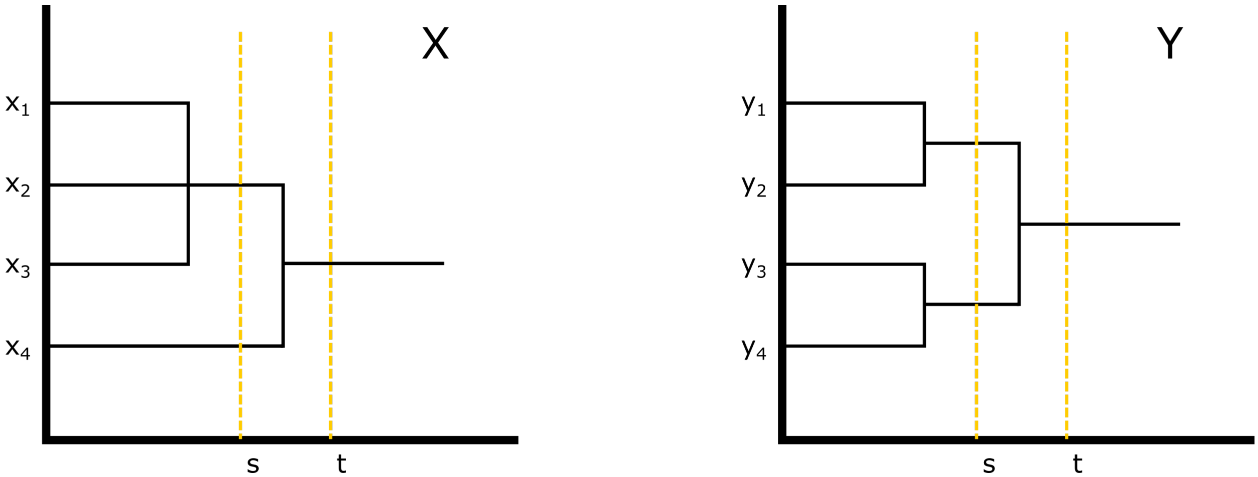

Figure 11: **(Illustration of Definition 8.3)** X and Y are two 4-point ultrametric spaces. Suppose $s = t - \varepsilon$ for some $\varepsilon > 0$. It is easy to see that Y satisfies the first $(\varepsilon, 2)$-growth condition. However that X does not satisfy the first $(\varepsilon, 2)$ growth condition can be seen from the figure clearly that $2\,\#[x_1]_{\hat{s}} = 6 > 4 = \#[x_1]^{\circ}_{t}$. It can be seen from this example that FGC prevents a given equivalence class in $X_{\hat{t}}$ from containing most of the points of X and thus the dendrogram will split more "evenly".

Let LX be the list of all closed balls of X sorted according to increasing diameter values. For each closed ball $B^X \in \mathrm{LX}$, consider the open quotient $B^X_{\hat{\rho}_\varepsilon(B^X)} = \{B^X_1, \cdots, B^X_m\}$, where $\rho_\varepsilon(B^X) := \max(\mathrm{diam}(B^X) - 2\varepsilon, 0)$. Note that for any $I \subset \{1, \cdots, m\}$, $\mathrm{diam}(\bigcup_{i \in I} B^X_i) \leq \mathrm{diam}(B^X)$. If the equality is achieved, we call $\bigcup_{i \in I} B^X_i$ an *ε-maximal union* of closed balls of B^X. Define $B^X_{(\varepsilon)}$ to be the list of all ε-maximal unions of closed balls *in* B^X sorted according to increasing cardinality. Then, we build a new list $\mathrm{LX}_{(\varepsilon)} := \bigcup_{B^X \in \mathrm{LX}} B^X_{(\varepsilon)}$ by replacing each $B^X \in \mathrm{LX}$ with the list $B^X_{(\varepsilon)}$. Note $\mathrm{LX}_{(\varepsilon)}$ is still a sorted list since the ε-maximal unions preserve diameter. We abuse the notation and still use B^X to represent an element in $\mathrm{LX}_{(\varepsilon)}$.

Remark 8.6. *The value $\rho_\varepsilon(X)$ originates from Theorem 8.2 as a lower bound for $\delta_\varepsilon(Y)$. In the case when $|\mathrm{diam}(X) - \mathrm{diam}(Y)| \leq \varepsilon$, we have $\delta_\varepsilon(Y) \geq \max(\mathrm{diam}(X) - 2\varepsilon, 0) = \rho_\varepsilon(X)$.*

Fix some input triple (X, Y, ε). It is clear that the pair (X, Y) belongs to $\mathrm{LX}_{(\varepsilon)} \times \mathrm{LY}$. We devise our DP algorithm **FindCorrespondenceDP** (Algorithm 3) so that it maintains a binary variable $\mathrm{DYN}(B^X, B^Y)$ for each pair $(B^X, B^Y) \in \mathrm{LX}_\varepsilon \times \mathrm{LY}$, such that $\mathrm{DYN}(B^X, B^Y) = 1$ if there exists an ε-correspondence between B^X and B^Y, and $\mathrm{DYN}(B^X, B^Y) = 0$ otherwise. The main idea is the following. The algorithm starts looping over all $B^Y \in \mathrm{LY}$. Inside the loop, it computes $\mathrm{DYN}(B^X, B^Y)$ by looping over all $B^X \in \mathrm{LX}_\varepsilon$. Most pairs (B^X, B^Y) fall into the base cases and $\mathrm{DYN}(B^X, B^Y)$ is determined by comparing diameters. For non base cases, we have the following two situations:

1. If $\mathrm{diam}(B^Y) > \varepsilon$, we apply Theorem 8.2 to the pair (B^X, B^Y).

2. If $\mathrm{diam}(B^Y) \leq \varepsilon$, we apply Proposition 8.8 below which induces the corresponding Algorithm 2 – see Appendix C for a proof of the proposition.

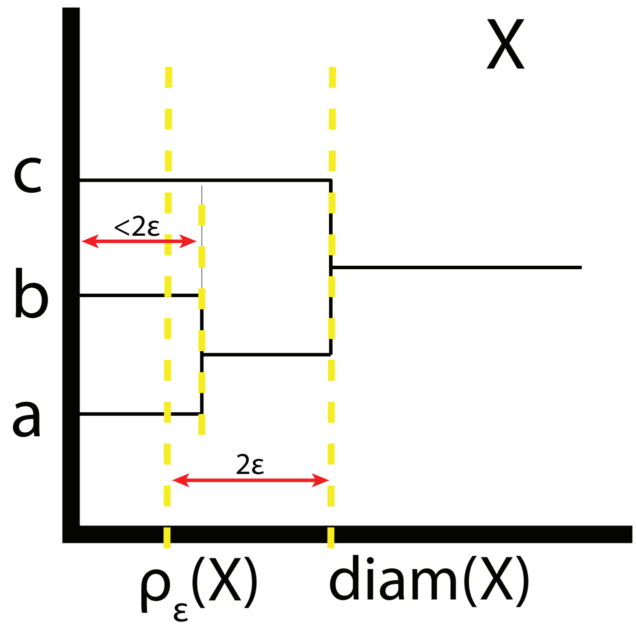

Figure 12: **(Illustration of** $\mathrm{LX}_{(\varepsilon)}$**).** $\mathrm{LX} = \langle\{a\},\{b\},\{c\},\{a,b\},\{a,b,c\}\rangle$. For the ball $\{a,b,c\}$, it is easy to see that $\{a,b,c\}_{(\varepsilon)} = \{\{a\}\bigcup\{c\},\{b\}\bigcup\{c\}\}$. For other balls B^X, we have $B^X_{(\varepsilon)} = \{B^X\}$. For example, $\{a,b\}_{\mathring{\rho}_\varepsilon(\{a,b\})} = \{\{a\},\{b\}\}$, so $\{a,b\}_{(\varepsilon)} = \{\{a,b\}\}$ since both $\{a\}$ and $\{b\}$ have zero diameter. Hence, $\mathrm{LX}_{(\varepsilon)} = \langle\{a\},\{b\},\{c\},\{a,b\},\{a,c\},\{b,c\},\{a,b,c\}\rangle$.

Remark 8.7. *To reduce redundant computation, our DP algorithm only inspects pairs in* $\mathrm{LX}_{(\varepsilon)} \times \mathrm{LY}$ *instead of a much larger symmetric set* $\mathrm{LX}_{(\varepsilon)} \times \mathrm{LY}_{(\varepsilon)}$*. Due to the asymmetry of* $\mathrm{LX}_{(\varepsilon)} \times \mathrm{LY}$*, the exceptional case in item 2 arises. Recall that in paragraph 'Application of Theorem 8.2' and in the recursive algorithm (Algorithm 1), we swap the roles of* B^X *and* B^Y *to deal with the case. Applying open quotient after swapping, we obtain* $B^X_{\mathring{\delta}_0(B^X)} = \{B^X_j\}_{j\in J}$ *and* $B^Y_{\mathring{\delta}_\varepsilon(B^X)} = \{B^Y_i\}_{i\in I}$*. We then run into decision problems with input* $(B^Y_{\Psi^{-1}(j)}, B^X_j, \varepsilon)$ *for surjections* $\Psi : I \twoheadrightarrow J$*. Being a union of closed balls in* B^Y*,* $B^Y_{\Psi^{-1}(j)}$ *does not necessarily belong to* LY*, the list of closed balls, which prevents the DP algorithm from computing* $\mathrm{DYN}(B^X, B^Y)$ *only based on* $\mathrm{LX}_{(\varepsilon)} \times \mathrm{LY}$*.*

Proposition 8.8. *Let* $X, Y \in \mathcal{U}$ *and* $\varepsilon \geq 0$ *be such that* $\mathrm{diam}(X) > \varepsilon$ *and* $\mathrm{diam}(Y) \leq \varepsilon$*. Then, there exists a correspondence* R *between* X *and* Y *with* $\mathrm{dis}(R) \leq \varepsilon$ *if and only if there exists an injection* $\varphi : X_\varepsilon \to Y$ *with* $\mathrm{dis}(\varphi) \leq \varepsilon$*.*

The function $\mathbf{QuotientClosed}(X, \varepsilon)$ in Algorithm 2 will return the closed quotient space X_ε (Definition 5.2). See Appendix B for more details.

Note that, the given pseudocode in Algorithm 3 can only determine the existence of correspondence with distortion bounded by ε without actually constructing a correspondence. However, it should be clear that one can trace back the DYN matrix to produce a correspondence if it exists.

Theorem 8.9 (Correctness of Algorithm 3)**.** *There exists a correspondence* R *between* X *and* Y *such that* $\mathrm{dis}(R) \leq \varepsilon$ *if and only if* $\mathbf{FindCorrespondenceDP}(X, Y, \varepsilon) = 1$.

Proof. By specifying the definition of closed balls in ultrametric spaces, $\mathrm{LX}_{(\varepsilon)}$ and LY can be represented in the following ways:

Algorithm 2: $\textbf{FindCorrespondenceSmall}(X, Y, \varepsilon)$

Assert $\mathrm{diam}(X) > \varepsilon$ and $\mathrm{diam}(Y) \leq \varepsilon$
if $|\mathrm{diam}(X) - \mathrm{diam}(Y)| > \varepsilon$ **then**
  return 0
$X_\varepsilon = \textbf{QuotientClosed}(X, \varepsilon)$
for Injection $\Phi : X_\varepsilon \to Y$ **do**
  DIS $\leftarrow \mathrm{dis}(\Phi)$
  if DIS $\leq \varepsilon$ **then**
    return 1
return 0

$$\mathrm{LX}_{(\varepsilon)} := \{B^X \subset X : \forall x \in X, \text{ if } \exists x' \in B^X \text{ such that } u_X(x, x') < \mathrm{diam}(B^X) - 2\varepsilon, \text{ then } x \in B^X\},$$

$$\mathrm{LY} := \{B^Y \subset Y : \forall y \in Y, \text{ if } \exists y' \in B^Y \text{ such that } u_Y(y, y') \leq \mathrm{diam}(B^Y), \text{ then } y \in B^Y\}.$$

Let $B^Y \in \mathrm{LY}$ and $B^Y_{\mathring{\delta}_0(B^Y)} = \{B^Y_j\}_{j \in J}$. Then, it is obvious that each $B^Y_j \in \mathrm{LY}$. If $B^X \in \mathrm{LX}_{(\varepsilon)}$ be such that $|\mathrm{diam}(B^X) - \mathrm{diam}(B^Y)| \leq \varepsilon$ and $\mathrm{diam}(B^Y) > \varepsilon$. We show that for any $B^X_i \in B^X_{\mathring{\delta}_\varepsilon(B^Y)} = \{B^X_1, \cdots, B^X_n\}$, one has $B^X_i \in \mathrm{LX}_{(\varepsilon)}$. For any $x \in X$, suppose that there is $x' \in B^X_i$ with $u_X(x, x') < \mathrm{diam}(B^X_i) - 2\varepsilon$. Then, $u_X(x, x') < \mathrm{diam}(B^X) - 2\varepsilon$, hence, $x \in B^X$ since $B^X \in \mathrm{LX}_{(\varepsilon)}$. Now, $u_X(x, x') < \mathrm{diam}(B^X) - 2\varepsilon \leq \mathrm{diam}(B^Y) - \varepsilon \leq \delta_\varepsilon(B^Y)$. So x and x' belong to the same block of $B^X_{\mathring{\delta}_\varepsilon(B^Y)}$ and thus $x \in B^X_i$. Therefore, $B^X_i \in \mathrm{LX}_{(\varepsilon)}$.

Next, we show that for any subset $I \subset \{1, \cdots, n\}$, the union $B^X_I = \bigcup_{i \in I} B^X_i$ belongs to $\mathrm{LX}_{(\varepsilon)}$. For any $x \in X$, suppose that there is $x' \in B^X_i \subset B^X_I$ with $u_X(x, x') < \mathrm{diam}(B^X_I) - 2\varepsilon \leq \mathrm{diam}(B^X) - 2\varepsilon$. Then, $x \in B^X$ since $B^X \in \mathrm{LX}_{(\varepsilon)}$. Now, $u_X(x, x') < \mathrm{diam}(B^X_I) - 2\varepsilon < \mathrm{diam}(B^X) - 2\varepsilon \leq \delta_\varepsilon(B^Y)$. So x and x' belong to the same block of $B^X_{\mathring{\delta})\varepsilon(B^Y)}$ and thus $x \in B^X_i$. Therefore $x \in B^X_I$ and thus $B^X_I \in \mathrm{LX}_{(\varepsilon)}$.

Then, we show that for any $(B^X, B^Y) \in \mathrm{LX}_{(\varepsilon)} \times \mathrm{LY}$, $\mathrm{DYN}[B^X][B^Y] = 1$ if and only if there exists an ε-correspondence R between B^X and B^Y. If (B^X, B^Y) belongs to one of the base cases, the statement is obviously true. For non-base cases, we prove by induction on $\mathrm{diam}(B^Y) \in \mathrm{Spec}(Y) = \{t_0 < \cdots < t_N\}$. When $\mathrm{diam}(B^Y) = t_0 = 0$, for any B^X, (B^X, B^Y) belongs to one of the base cases, so the statement holds true trivially. Suppose that the claim is true for all t_i with $i < j$ and all $\mathrm{DYN}[B^X][B^Y]$ are known for $\mathrm{diam}(B^Y) < t_j$. Then, the induction step follows directly from Theorem 8.2 and Proposition 8.8:

1. If $\mathrm{diam}(B^Y) > \varepsilon$, we quotient B^X and B^Y to $B^X_{\mathring{\delta}_\varepsilon(B^Y)} = \{B^X_i\}_{i \in I}$ and $B^Y_{\mathring{\delta}_0(B^Y)} = \{B^Y_j\}_{j \in J}$ respectively. By previous argument, we know that each $B^X_i \in \mathrm{LX}_{(\varepsilon)}$ and $B^Y_j \in \mathrm{LY}$. Since $\mathrm{diam}(B^Y_j) < \mathrm{diam}(B^Y) = t_j$, by induction assumption, the value $\mathrm{DYN}[B^X_{\Psi^{-1}(j)}][B^Y_j]$ has been computed for any surjection $\Psi : I \to J$ so we already know the existence of ε-correspondence between $B^X_{\Psi^{-1}(j)}$ and B^Y_j. Since $\mathrm{DYN}[B^X][B^Y]$ is then determined via Theorem 8.2, it is 1 if and only if there exists an ε-correspondence between B^X and B^Y.

Algorithm 3: **FindCorrespondenceDP**(X, Y, ε)

Build and sort LX and LY
Build $\mathrm{LX}_{(\varepsilon)}$ and Hash tables for $\mathrm{LX}_{(\varepsilon)}$ and LY
$\mathrm{DYN} = \mathrm{zeros}(\#\mathrm{LX}_{(\varepsilon)}, \#\mathrm{LY})$
for $B^Y \in \mathrm{LY}$ **do**
  $\mathrm{LXI} = \{B^X \in \mathrm{LX}_{(\varepsilon)} : \mathrm{diam}(B^Y) - \varepsilon \leq \mathrm{diam}(B^X) \leq \mathrm{diam}(B^Y) + \varepsilon\}$
  for $j = 0$ *to* $\#\mathrm{LXI}$ // HERE
  do
    $B^X = \mathrm{LXI[j]}$
    Find txi $= B^X$'s index in $\mathrm{LX}_{(\varepsilon)}$
    Find tyi $= B^Y$'s index in LY
    if $\max(\mathrm{diam}(B^X), \mathrm{diam}(B^Y)) \leq \varepsilon$ **then**
      $\mathrm{DYN[txi][tyi]} = 1$
    else if $\mathrm{diam}(B^X) > \varepsilon$ ***and*** $\mathrm{diam}(B^Y) \leq \varepsilon$ **then**
      $\mathrm{DYN[txi][tyi]} =$ **FindCorrespondenceSmall**(B^X, B^Y, ε)
    else
      $\{B_i^X\}_{i \in I} =$ **PartitionOpen**$(B^X, \delta_\varepsilon)$
      $\{B_j^Y\}_{j \in J} =$ **PartitionOpen**(B^Y, δ_0)
      for Each surjection $\Psi : I \twoheadrightarrow J$ **do**
        Find pxi(j) $= B^X_{\Psi^{-1}(j)}$'s index in $\mathrm{LX}_{(\varepsilon)}$ for $j = 1, \cdots, m$
        if $\mathrm{DYN[pxi(j)][tyi]} = 1 \forall j = 1, \cdots, m$ **then**
          $\mathrm{DYN[txi][tyi]} = 1$
          continue in line HERE

return DYN[END][END]

2. If $\mathrm{diam}(B^Y) \leq \varepsilon$, we have $\mathrm{DYN}[B^X][B^Y] =$ **FindCorrespondenceSmall**(B^X, B^Y, ε). Then, the statement holds true due to Proposition 8.8.

Then, since $\mathrm{LX}_{(\varepsilon)}$ and LY are sorted with increasing diameter such that the last elements are exactly X and Y respectively, $\mathrm{DYN[END][END]} = 1$ if and only if there exists a correspondence between X and Y with $\mathrm{dis}(R) \leq \varepsilon$. □

Complexity analysis. To analyze the complexity of Algorithm 3, we consider the following growth condition:

Definition 8.10 (Second (ε, γ)-growth condition)**.** *For $\varepsilon \geq 0$, and $\gamma \in \mathbb{N}$, we say that an ultrametric space (X, u_X) satisfies the second (ε, γ)-growth condition (SGC) if for all $x \in X$, and $t \geq 2\varepsilon$,*

$$\#\{[x']_{(t-2\varepsilon)}^{\circ} : x' \in [x]_t\} \leq \gamma.$$

Remark 8.11 (Interpretation of SGC and its relation with FGC)**.** *The second (ε, γ)-growth condition is equivalent to saying for any $x \in X$ and $t > 2\varepsilon$, the number of open children of $[x]_t$ at $t - 2\varepsilon$ is bounded above by γ. By Remark 8.4, if X satisfies the first (ε, γ)-growth condition, then for* *each class $[x]_t$ when $t > \varepsilon$, the number of children at $t-\varepsilon$ is bounded above by γ. This implies that X satisfies the second $\left(\frac{\varepsilon}{2},\gamma\right)$-growth condition, which indicates that the second growth condition is less rigid than the first growth condition.*

Remark 8.12. *Suppose X satisfies the second (ε,γ)-growth condition. Then, for any point $x \in X$ and any $0 \le t \le 2\varepsilon$, we have $\#[x]_t \le \gamma$. Indeed, $\#[x]_t \le \#[x]_{2\varepsilon} = \#\{x' : x' \in [x]_{2\varepsilon}\} \le \gamma$. Hence, the second (ε,γ)-growth condition can be equivalently written as for all $x \in X$ and $t \ge 0$,*

$$\#\{[x']_{\mathring{t}_\varepsilon} : x' \in [x]_t\} \le \gamma,$$

where $t_\varepsilon \coloneqq A_1(t, 2\varepsilon) = \max(0, t-2\varepsilon)$.

The following characterization lemma of the second growth condition will be used in the sequel and the proof is postponed in Appendix D.

Lemma 8.13. *Given a finite ultrametric space X and $\varepsilon \ge 0$, denote*

$$\gamma_\varepsilon(X) \coloneqq \min\{\gamma \in \mathbb{N} : \textit{X satisfies the second } (\varepsilon,\gamma)\textit{-growth condition.}\}$$

Then,

$$\gamma_\varepsilon(X) = \max\left\{\#\left\{[x']^X_{\mathring{t}_\varepsilon} : x' \in [x]^X_t\right\} : t \in \mathrm{spec}(X)\right\}.$$

Since Algorithm 3 utilizes Algorithm 2, we first analyze the complexity of Algorithm 2.

Lemma 8.14. *Algorithm 2 runs in time $O(n^2 n^n)$ where $n = \max(\#X, \#Y)$.*

Proof. We need $O(n^2)$ time to do the quotient. There are at most n^n injections and for each injection we need $O(n^2)$ time to compute the distortion. Therefore, Algorithm 2 runs in time $O(n^2 n^n)$. □

Now we start to analyze the time complexity of $\textbf{FindCorrespondenceDP}(X,Y,\varepsilon)$. We need to following lemma to estimate the size of $\mathrm{LX}_{(\varepsilon)}$ in Algorithm 3.

Lemma 8.15. *Given two finite ultrametric spaces (X,u_X) and (Y,u_Y), let $\mathrm{LX}_{(\varepsilon)}$ and LY be as in Algorithm 3. Then, $\#\mathrm{LY} = O(\#Y)$. If X satisfies the second (ε,γ)-growth condition, then $\#\mathrm{LX}_{(\varepsilon)} = O(\#X\, 2^\gamma)$.*

Proof. LY is the set of all closed balls of Y. As mentioned in the paragraph 'Data structure for ultrametric spaces' in the beginning of Section 8, each closed ball corresponds to a vertex of the underlying tree of Y. This tree has $\#Y$ many leaves, and thus has $O(\#Y)$ many vertices. Therefore, $\#\mathrm{LY} = O(\#Y)$.

$\mathrm{LX}_{(\varepsilon)}$ is defined as $\bigcup_{B^X \in \mathrm{LX}} B^X_{(\varepsilon)}$. Each $B^X_{(\varepsilon)}$ is actually a subset of the power set $2^{B^X_{\mathring{\rho}_\varepsilon(B^X)}}$. Since B^X is a closed ball in X, it can be written as an closed equivalence class $[x]^X_\rho$ for some $x \in X$ and $\rho = \mathrm{diam}(B^X)$. By the second (ε,γ)-growth condition, we have that $\#B^X_{\mathring{\rho}_\varepsilon(B^X)} = \#\{[x']^X_{\mathring{\rho}_\varepsilon} : x' \in [x]^X_\rho\} \le \gamma$ and thus $\#2^{B^X_{\mathring{\rho}_\varepsilon(B^X)}} \le 2^\gamma$. Then, by $\#\mathrm{LX} = O(X)$, we have $\#\mathrm{LX}_{(\varepsilon)} = O(\#X\, 2^\gamma)$. □

Theorem 8.16. *Assume that (X,u_X) and (Y,u_Y) both satisfy the second (ε,γ)-growth condition for some $\varepsilon \ge 0$ and $\gamma \ge 1$. Then, algorithm $\textbf{FindCorrespondenceDP}(X,Y,\varepsilon)$ runs in time $O\left(n^2 2^\gamma \gamma^{\gamma+2}\log(n)\right)$.*

Proof. By Lemma 8.15, $\#\mathrm{LX}_{(\varepsilon)} = O(n2^\gamma)$ and $\#\mathrm{LY} = O(n)$. Since we store the ultrametric spaces by weighted tree structure, building $\mathrm{LX}_{(\varepsilon)}$ and LY will take time $O(n) \times (O(n) + O(n2^\gamma)) = O(n^2 2^\gamma)$, where the first $O(n)$ results from copying a rooted weighted tree to the lists. Hence, the size of DYN is $O(n^2 2^\gamma)$. In order to access the index in $\mathrm{LX}_{(\varepsilon)}$ and LY quickly (in constant time), we will need to build Hash tables for $\mathrm{LX}_{(\varepsilon)}$ and LY respectively with total time complexity $O(n^2 2^\gamma)$ [CLRS09, Chapter 11].

For each $B^Y \in \mathrm{LY}$, we need $O(\log(n2^\gamma))$ time to build LXI through a binary search process, since $\mathrm{LX}_{(\varepsilon)}$ is sorted. Then, we have the following cases for $B^X \in \mathrm{LXI}$:

1. $\max(\mathrm{diam}(B^X), \mathrm{diam}(B^Y)) \le \varepsilon$ or $|\mathrm{diam}(B^X) - \mathrm{diam}(B^Y)| > \varepsilon$. It takes constant time to assign $\mathrm{DYN}[\mathrm{txi}][\mathrm{tyi}]$ either 1 or 0.

2. $\mathrm{diam}(B^X) > \varepsilon$ and $\mathrm{diam}(B^Y) \le \varepsilon$. We will invoke Algorithm 2 with entries B^X, B^Y and ε. In this case, both $\mathrm{diam}(B^X)$ and $\mathrm{diam}(B^Y)$ is bounded above by 2ε. Then, by SGC, it is easy to check that $\#B^X, \#B^Y \le \gamma$. Thus, by Lemma 8.14, $\mathbf{FindCorrespondenceSmall}(B^X, B^Y, \varepsilon)$ runs in time $O(\gamma^2 \gamma^\gamma)$.

3. $\mathrm{diam}(B^Y) > \varepsilon$. In this case, both B^X and B^Y will be decomposed via $\mathbf{PartitionOpen}$ into at most γ subspaces respectively, which will take time at most $O(n)$. We have at most γ^γ surjections to consider. For each surjection, we need $O(n)$ time to construct the unions of subspaces in $B^X_{\mathring{\delta}_\varepsilon(B^X)}$ and need constant time to visit the previous values in DYN since we have built Hash tables for $\mathrm{LX}_{(\varepsilon)}$ and LY to look up the index of pairs $(B^X_{\Psi^{-1}(j)}, B^Y_j)$ in matrix DYN.

Therefore, we need in total $O(n) \times O(n\gamma^{\gamma+2} \log(n2^\gamma)) = O(n^2\gamma^{\gamma+2}\log(n2^\gamma)) = O(n^2\gamma^{\gamma+2}\log(n))$ operations to fill out the matrix DYN. By adding the time for generating the Hash table and DYN, we have that the total time complexity is $O(n^2 2^\gamma \gamma^{\gamma+2}\log(n))$. □

Discussion of two dGH algorithms Both of Algorithm 1 (recursive) and Algorithm 3 (dynamic programming) follows from our key observation in Theorem 8.2. Our recursive algorithm is easier to implement than the dynamic programming algorithm, though the dynamic programming algorithm runs faster in general. Moreover, the recursive algorithm can more directly generate the correspondence matrix, whereas for the dynamic programming algorithm, one has to carefully trace back the matrix DYN to generate a correspondence matrix.

8.1.4 An algorithm for GHDU-opt

Given a metric space (X, d_X), recall that its spectrum is the set of non-negative real values $\mathrm{spec}(X) := \{d_X(x, x'),\ x, x' \in X\}$ containing all possible distances realized by pairs of points in X. Let $\Omega(X,Y) := \{|u_X(x,x') - u_Y(y,y')| : \forall x, x' \in X \text{ and } \forall y, y' \in Y\}$. Then, for any correspondence R between X and Y, $\mathrm{dis}(R)$ takes values in $\Omega(X,Y)$ by finiteness of X and Y and Equation (2). By Equation (1.3) we have that $d_{\mathrm{GH}}(X,Y) = \frac{1}{2}\inf_R \mathrm{dis}(R) = \frac{1}{2}\min_R \mathrm{dis}(R)$, thus $d_{\mathrm{GH}}(X,Y)$ takes values in $\frac{1}{2}\Omega(X,Y)$. Therefore, to compute $d_{\mathrm{GH}}(X,Y)$, we first sort the elements in Ω as $\omega_1 < \omega_2 < \cdots < \omega_M$ and run $\mathbf{FindCorrespondenceDP}$ (X, Y, ω_i) from $i = 1$ to $i = M$. If i is the smallest integer such that $\mathbf{FindCorrespondenceDP}$ $(X, Y, \omega_i) \neq 0$, then $d_{\mathrm{GH}}(X,Y) = \omega_i$. See Algorithm 4 ($\mathbf{dGH}$) for the pseudocode.

We have the following theorem regarding the complexity of algorithm $\mathbf{dGH}$.

Algorithm 4: $\mathbf{dGH}(X, Y)$

$\Omega \leftarrow$ **sort**($|\text{spec}(X) - \text{spec}(Y)|$, 'ascend')
for *i=1 to* $\#\Omega$ **do**
  if $\mathbf{FindCorrespondenceDP}(X, Y, \Omega(\mathrm{i}))$ **then**
    return $\frac{\Omega(\mathrm{i})}{2}$
return 0

Theorem 8.17. *Let (X, u_X) and (Y, u_Y) be two finite ultrametric spaces and $\varepsilon = 2d_{\mathrm{GH}}(X, Y)$. Assume that X and Y satisfy the second (ε, γ)-growth condition for some $\gamma \geq 1$. Then, the algorithm $\mathbf{dGH}(X, Y)$ runs in time $O(n^4 2^\gamma \gamma^{\gamma+2} \log(n))$, where $n = \max(\#X, \#Y)$.*

Proof. First note that $\#(\text{spec}(X) \cup \text{spec}(Y)) = O(n)$ (see [GV12]). Hence, $\#\Omega = O(n^2)$. Therefore, we will invoke algorithm $\mathbf{FindCorrespondence}$ at most $O(n^2)$ many times. Note that each invoked $\Omega(i)$ is bounded above by ε. Then, for each chosen i, it is easy to check that X and Y satisfy the second $(\Omega(i), \gamma)$-growth condition and thus the time complexity is $O(n^2) \times O(n^2 2^\gamma \gamma^{\gamma+3} \log(n)) = O(n^4 2^\gamma \gamma^{\gamma+2} \log(n))$. $\square$

Remark 8.18 (Comparison to [TW18])**.** *The second (ε, γ)-growth condition is in a similar spirit as the concept called degree-bound of merge trees defined in [TW18]: a merge tree has ε-degree-bound $\tau > 0$ if for each point in the merge tree, the sum of degree of all tree vertices inside an ε ball around the point is bounded above by τ.*

As mentioned in Remark 6.3, each finite ultrametric space can be naturally mapped into a merge tree and thus it is meaningful to talk about ε-degree-bound of an ultrametric space via this map. It is easy to see that if any ultrametric space X has ε-degree-bound τ, it automatically satisfies the second $(\frac{\varepsilon}{2}, \tau)$-growth condition. This means that our growth condition can be regarded in general as less rigid than the degree-bound condition from [TW18].

Now consider the case where two merge trees M_X and M_Y arise from finite ultrametric spaces X and Y such that $d_{\mathrm{GH}}(X, Y) = \frac{\varepsilon}{2}$. In this case, if d_{I} denotes the interleaving distance between merge trees of [MBW13], then by Remark 6.3 and Corollary 6.13, we have

$$\frac{1}{2} d_{\mathrm{I}}(M_X, M_Y) \leq d_{\mathrm{GH}}(X, Y) \leq d_{\mathrm{I}}(M_X, M_Y). \tag{26}$$

*Assume M_X and M_Y have 2ε-degree-bound τ, then X and Y satisfy the second (ε, τ)-growth condition. Since $d_{\mathrm{I}}(M_X, M_Y) \leq \varepsilon \leq 2\varepsilon$, by monotonicity of the degree bound, M_X and M_Y also have $d_{\mathrm{I}}(M_X, M_Y)$-degree bound τ. Then, it is shown in [TW18] that in time $O(n^4 2^\tau \tau^{\tau+2} \log(n))$ one can compute $d_{\mathrm{I}}(M_X, M_Y)$, which by Equation (26) is a 2-approximation of $d_{\mathrm{GH}}(X, Y)$. Note that by Theorem 8.17, with the same time complexity, our algorithm can compute the **exact** value of $d_{\mathrm{GH}}(X, Y)$.*

8.1.5 Discussion on $d_{\mathrm{GH}}^{(p)}$

One can easily modify the proofs of Theorem 8.2 and Proposition 8.8 to obtain the following results:

Theorem 8.19. *Let* $(X,u_X),(Y,u_Y)\in\mathcal{U}$ *and* $p\in[1,\infty]$. *For each* $\varepsilon\in[0,\operatorname{diam}(Y))$ *let* $\delta_0(Y)=\operatorname{diam}(Y)$ *and* $\delta_{\varepsilon,p}(Y):=A_p(\operatorname{diam}(Y),\varepsilon)$ *and write* $X_{\mathring{\delta}_{\varepsilon,p}(Y)}=\{X_i\}_{i\in I}$ *and* $Y_{\mathring{\delta}_0(Y)}=\{Y_j\}_{j\in J}$.

Assume $\Lambda_p(\operatorname{diam}(X),\operatorname{diam}(Y))\leq\varepsilon$. *Then, there exists a correspondence* R *between* X *and* Y *with* $\operatorname{dis}_p(R)\leq\varepsilon$ *if and only if there exists a surjection* $\Psi:I\to J$ *such that for every* $j\in J$ *there exists a correspondence between*

$$\left(X_{\Psi^{-1}(j)},u_X|_{X_{\Psi^{-1}(j)}\times X_{\Psi^{-1}(j)}}\right)\ \text{and}\ \left(Y_j,u_Y|_{Y_j\times Y_j}\right)$$

with p*-distortion less than or equal to* ε.

Proposition 8.20. *Assume that* $p<\infty$, *and that* $\varepsilon\geq 0$ *is such that* $\operatorname{diam}(X)>\varepsilon$ *and* $\operatorname{diam}(Y)\leq\varepsilon$. *Then, there exists a correspondence* R *between* X *and* Y *with* $\operatorname{dis}_p(R)\leq\varepsilon$ *if and only if there exists an injection* $\varphi:X_\varepsilon\to Y$ *with* $\operatorname{dis}_p(\varphi)\leq\varepsilon$.

Remark 8.21. *We exclude the case when* $p=\infty$ *in Proposition 8.20 because when* $\operatorname{diam}(Y)\leq\varepsilon<\operatorname{diam}(X)$, *there exists no correspondence* R *between* X *and* Y *such that* $\operatorname{dis}_\infty(R)\leq\varepsilon$ *(Corollary 5.8).*

Based on Proposition 8.19 and Proposition 8.20, both Algorithm 1 and Algorithm 3 can be generalized to solve the following decision problem of $d_{\mathrm{GH}}^{(p)}$ when $p\in[1,\infty]$:

Decision Problem ($d_{\mathrm{GH}}^{(p)}$ distance computation on $\mathcal{U}$ (pGHDU-dec))
Input: Finite ultrametric spaces (X,u_X) and (Y,u_Y) and $\varepsilon\geq 0$.
Question: Is there a correspondence R between X and Y such that $\operatorname{dis}_p(R)\leq\varepsilon$?

Alternatively, we can use Theorem 3.8 to solve the decision problem **pGHDU-dec** above more directly. For $p\in[1,\infty)$, we know from Theorem 3.8 that for any ultrametric spaces X and Y, there exists a correspondence R between X and Y with $\operatorname{dis}_p(R)\leq\varepsilon$ if and only if there exists a correspondence R' between $S_p(X)$ and $S_p(Y)$ with $\operatorname{dis}(R')\leq\varepsilon^p$. Thus, we have the following algorithm:

Algorithm 5: $p-\mathbf{FindCorrespondenceDP}(X,Y,\varepsilon)$

return $\mathbf{FindCorrespondenceDP}(S_p(X),S_p(Y),\varepsilon^p)$

We are interested in the limiting behavior of $p-\mathbf{FindCorrespondenceDP}(X,Y,\varepsilon)$ when p goes to infinity. In particular, we hope that the time complexity of the algorithm will decrease as p increases, which turns out to be true. Before stating the main result in Theorem 8.23, we start with the following preliminary lemma with proof given in Appendix E.

Lemma 8.22. *Given a finite ultrametric space* X *and* $\varepsilon\geq 0$, $\gamma_{\varepsilon^p}(S_p(X))$ *is a decreasing function with respect to* $p\in[1,\infty)$. *Moreover, there exists a constant* $C=C(X,\varepsilon)>0$ *depending on* X *and* ε, *such that for any* $p,q>C$, $\gamma_{\varepsilon^p}(S_p(X))=\gamma_{\varepsilon^q}(S_q(X))$. *We denote* $\gamma^\infty(X):=\lim_{p\to\infty}\gamma_{\varepsilon^p}(S_p(X))$.

Theorem 8.23. *For any two finite ultrametric spaces X and Y with $n = \max(\#X, \#Y)$, denote by $T_p(X, Y, \varepsilon)$ the time complexity of $p-\mathbf{FindCorrespondenceDP}(X, Y, \varepsilon)$. Then, $T_p(X, Y, \varepsilon)$ is a decreasing function with respect to $p \in [1, \infty)$. In particular, if X, Y satisfy the second (ε, γ)-growth condition, then*

$$\lim_{p\to\infty} T_p(X, Y, \varepsilon) = O(n^2 2^\tau \tau^{\tau+2} \log(n)),$$

where $\tau = \max(\gamma^\infty(X), \gamma^\infty(Y))$.

Proof. When applying p-$\mathbf{FindCorrespondenceDP}(X, Y, \varepsilon)$, the list $\mathrm{LX}_{(\varepsilon)}$ generated will has size $O\left(n2^{\gamma_{\varepsilon^p}(S_p(X))}\right)$. As we can see from the proof of Theorem 8.16, $T(X, Y, p) = O(n \times \#\mathrm{LX}_{(\varepsilon)}) + O(n \times \log(\#\mathrm{LX}_{(\varepsilon)})\gamma^{\gamma+2})$. Since by Lemma 8.22 $\gamma_{\varepsilon^p}(S_p(X))$ is a decreasing function, we have that $T(X, Y, p)$ is decreasing. When p is large enough, the list $\mathrm{LX}_{(\varepsilon)}$ will has size $O(n2^\tau)$ and thus $T_p(X, Y, \varepsilon) = O(n^2 2^\tau \tau^{\tau+2} \log(n))$. □

8.2 Computation of u_{GH} on $\mathcal{U}$

In this section we study the following problem.

Optimization Problem (u_{GH} **distance computation on** $\mathcal{U}$ **(UGHDU-opt))**
Input: Finite ultrametric spaces (X, u_X) and (u, d_Y).
Output: The value $u_{\mathrm{GH}}(X, Y)$.

8.2.1 A slow u_{GH} algorithm

To solve the optimization problem **UGHDU-opt**, we can adopt the same strategy of solving Problem **GHDU-opt** by solving a decision problem **GHDU-dec** for finitely many different parameters.

Let us first consider the following decision problem of u_{GH}.

Decision Problem (u_{GH} **distance computation on** $\mathcal{U}$ **(UGHDU-dec))**
Input: Finite ultrametric spaces (X, u_X) and (Y, u_Y) and $\varepsilon \geq 0$.
Question: Is there a correspondence R between X and Y such that $\mathrm{dis}_\infty(R) \leq \varepsilon$?

Theorem 3.8 excludes the case when $p = \infty$. Hence, to solve the decision problem for u_{GH}, we cannot apply the trick of utilizing the snowflake functor as we have done for the analysis of Algorithm 5. However, we can still apply Theorem 8.19 in the case of $p = \infty$ directly to build a dynamic programming algorithm to solve the decision problem **UGHDU-dec**. Note that $\delta_{\varepsilon,\infty}(Y) = \delta_0(Y)$, so we only need to consider the list LX instead of $\mathrm{LX}_{(\varepsilon)}$. The pseudocode is given in Algorithm 6. It can be proved that this algorithm is also fixed-parameter-tractable as Algorithm 3. We can also show that this algorithm can not be done in polynomial time in general.

Example 8.24. *Consider the two ultrametric spaces $X = \Delta_n(1)$ and $Y = \Delta_2(1)$, where $\Delta_m(r)$ means an m-point space with all interpoint distance equal to r. Let $\varepsilon = \frac{1}{2}$. While running $\mathbf{uGHDP}(X, Y, \varepsilon)$, we will build two lists* LX *and* LY *with sizes $\#\mathrm{LY} = 2$ and $\#\mathrm{LX} = 2^n - 1$. Hence, the time complexity is at least $O(2^n)$ and thus $\mathbf{uGHDP}$ cannot be done in polynomial time in general.*

Algorithm 6: uGHDP(X, Y, ε)

```
Build and sort LX and LY
Build Hash tables for LX and LY
DYN = zeros(#LX, #LY)
for B^Y ∈ LY do
    LXI = { B^X ∈ LX : A_∞(diam(B^Y), ε) ≤ diam(B^X) ≤ diam(B^Y) ⊞_∞ ε }
    for j = 0 to #LXI                                              // HERE
     do
        B^X = LXI[j]
        Find txi = B^X's index in LX
        Find tyi = B^Y's index in LY
        if max(diam(B^X), diam(B^Y)) ≤ ε then
            DYN[txi][tyi] = 1
        else
            {B_i^X}_{i∈I} = PartitionOpen(B^X, δ_0(Y))
            {B_j^Y}_{j∈J} = PartitionOpen(B^Y, δ_0(Y))
            for Each surjection Ψ : I ↠ J do
                Find pxi(j) = B^X_{Ψ^{-1}(j)}'s index in LX for j = 1, ··· , m
                if DYN[pxi(j)][tyi] = 1 ∀j = 1, ··· , m then
                    DYN[txi][tyi] = 1
                    continue in line HERE
return DYN[END][END]
```

By incorporating the structural theorem for u_{GH}, we can further improve Algorithm 6 to Algorithm 7 which can be done in polynomial time.

We first recall the statement of the structural theorem for u_{GH} (Theorem 5.7).

Theorem 5.7 (Structural theorem for u_{GH})**.** *For all $X, Y \in \mathcal{U}$ one has that*

$$u_{\mathrm{GH}}(X, Y) = \min\left\{t \geq 0 : \, (X_t, u_{X_t}) \cong (Y_t, u_{Y_t})\right\}.$$

Now, we state an improved version of Theorem 8.19 when $p = \infty$.

Theorem 8.25. *Let $(X, u_X), (Y, u_Y) \in \mathcal{U}$. Fix $\varepsilon \in [0, \mathrm{diam}(Y))$ and let $\delta_0 = \mathrm{diam}(Y)$. Write $X_{\mathring{\delta}_0} = \{X_i\}_{i \in I}$ and $Y_{\mathring{\delta}_0} = \{Y_j\}_{j \in J}$.*

Assume $\Lambda_\infty(\mathrm{diam}(X), \mathrm{diam}(Y)) \leq \varepsilon$. Then, there exists a correspondence R between X and Y with $\mathrm{dis}_\infty(R) \leq \varepsilon$ if and only if there exists a bijection *$\Psi : I \to J$ such that for every $j \in J$ there exists a correspondence between*

$$\left(X_{\Psi^{-1}(j)}, u_X|_{X_{\Psi^{-1}(j)} \times X_{\Psi^{-1}(j)}}\right) \ \textit{and} \ \left(Y_j, u_Y|_{Y_j \times Y_j}\right)$$

with ∞-distortion less than or equal to ε.

Proof. By Theorem 8.19, we only need to show that the surjection $\Psi : I \to J$ induced by a correspondence R between X and Y with $\mathrm{dis}_\infty(R) \leq \varepsilon$ is actually a bijection. Indeed, by Theorem 3.5, we have $u_{\mathrm{GH}}(X, Y) \leq \varepsilon$. Then, since $\delta_0 = \mathrm{diam}(Y) > \varepsilon$, by Theorem 5.7, we have that $X_{\delta_0} \cong Y_{\delta_0}$. It is easy to deduce that $X_{\mathring{\delta}_0} \cong Y_{\mathring{\delta}_0}$. Therefore, $\#I = \#J$ and thus the surjection $\Psi : I \to J$ is a bijection. $\square$

Based on Theorem 8.25, we can modify Algorithm 6 to obtain Algorithm 7 for solving the decision problem of u_{GH}. The main difference between the two algorithms is that we replace the *surjection* search in Algorithm 6 by a *bijection* search in Algorithm 7. Unlike searching the surjections, the bijection search problem can be transformed into a max-flow problem which can be solved via the Ford-Fulkerson algorithm in polynomial time [CLRS09, Section 26.2]. The function $\mathbf{contains\_bijection}$ in Algorithm 7 will determine whether there exists a bijection inside the matrix em by invoking Ford-Fulkerson algorithm, of which we will omit the detail here.

We will omit the proof of correctness of Algorithm 7 ($\mathbf{suGHDP}$) but will only show that the algorithm runs in polynomial time.

Theorem 8.26. *Given two finite ultrametric spaces X and Y, $\mathbf{suGHDP}(X, Y, \varepsilon)$ runs in time $O(n^4)$, where $n = \max(\#X, \#Y)$.*

Proof. LX is the set of all closed balls of X. As mentioned in the beginning of Section 8, each closed ball corresponds to a vertex of the underlying tree of X. This tree has $\#X$ many leaves, and thus has $O(\#X)$ many vertices. Therefore, $\#\mathrm{LX}, \#\mathrm{LY} = O(n)$. Similar as in the proof of Theorem 8.16, building LX and LY will take time $O(n) \times O(n) = O(n^2)$. The size of DYN is $O(n^2)$. In order to access the index in LX and LY quickly (in constant time), we will need to build a Hash table with time complexity $O(n^2)$.

For each $B^Y \in \mathrm{LY}$, we need $O(\log(n))$ time to build LXI. Then, we have the following cases for $B^X \in \mathrm{LXI}$:

1. $\mathrm{diam}(B^X) \leq \varepsilon$ and $\mathrm{diam}(B^Y) \leq \varepsilon$. It takes constant time to assign DYN[txi][tyi] to 1.
2. $\mathrm{diam}(B^X) = \mathrm{diam}(B^Y) > \varepsilon$. In this case, both B^X and B^Y will be decomposed as $N \leq n$ subspaces respectively, which will take time at most $O(n^2)$. We need $O(n^2)$ time to build matrix em and $O(n^3)$ time to run Ford-Fulkerson algorithm [CLRS09] to seek a bijection in em.

Therefore, we need in total $O(n) \times O(n^3) = O(n^4)$ time to fill out the matrix DYN. By adding the time for generating the Hash table and DYN, we have that the total time complexity is $O(n^4)$. $\square$

Now, we can solve the optimization problem **UGHDU-opt** by utilizing the same idea used in Algorithm 4 that we apply $\mathbf{suGHDP}(X, Y, w_i)$ over all possible values w of $\Omega_\infty(X, Y) := \{\Lambda_\infty(u_X(x, x'), u_Y(y, y')) : x, x' \in X \text{ and } y, y' \in Y\}$ following an increasing order. The smallest such value that $\mathbf{suGHDP}(X, Y, w) = 1$ is $u_{\mathrm{GH}}(X, Y)$. It is easy to show that $\Omega_\infty(X, Y) = O(n)$, and thus computing $u_{\mathrm{GH}}(X, Y)$ runs in time $O(n^5)$.

Algorithm 7: **suGHDP**(X, Y, ε)

Build and sort LX and LY
Build Hash tables for LX and LY
DYN $=$ zeros(#LX, #LY)
for $B^Y \in$ LY **do**
  LXI $= \left\{ B^X \in \mathrm{LX}_{(\varepsilon)} : A_\infty(\mathrm{diam}(B^Y), \varepsilon) \leq \mathrm{diam}(B^X) \leq \mathrm{diam}(B^Y) \boxplus_\infty \varepsilon \right\}$
  for $j = 0$ *to* #LXI **do**
    B^X= LXI[j]
    Find txi $=B^X$'s index in LX
    Find tyi $=B^Y$'s index in LY
    if $\max(\mathrm{diam}(B^X), \mathrm{diam}(B^Y)) \leq \varepsilon$ **then**
      DYN[txi][tyi] $= 1$
    else
      $\{B_i^X\}_{i \in I} =$ **PartitionOpen**$(B^X, \delta_0(Y))$
      $\{B_j^Y\}_{j \in J} =$ **PartitionOpen**$(B^Y, \delta_0(Y))$
      if $\#I \neq \#J$ **then** DYN[txi][tyi] $= 0$
      else em $= \mathrm{zeros}(\#I, \#I)$
      for $i = 1 : \#I$ **do**
        for $j = 1 : \#I$ **do**
          em$[i][j] =$ DYN$[B_i^X][B_j^Y]$
      DYN[txi][tyi] $=$ **contains_bijection**(em)
return DYN[END][END]

8.2.2 A fast u_{GH} algorithm

In fact, we can devise a faster algorithm for computing u_{GH} based solely on the structural theorem (Theorem 5.7). This algorithm can directly solve the optimization problem **UGHDU-opt** (without solving a decision problem first) and runs in time $O(n\log(n))$, which is much faster than Algorithm 7.

For a finite ultrametric space X, it is obvious that the isometry type of X_t only changes finitely many times along $0 \leq t < \infty$. In fact, the set of all points when X_t changes its isometry type is exactly the spectrum $\mathrm{spec}(X)$ of X.

Then, in order to compute u_{GH} between two finite ultrametric spaces X and Y, Theorem 5.7 suggests that we simply check whether $X_t \cong Y_t$ for all $0 < t \in \mathrm{spec}(X) \bigcup \mathrm{spec}(Y)$, starting from the largest to the smallest, and the smallest t such that $X_t \ncong Y_t$ will be $u_{\mathrm{GH}}(X, Y)$ (if otherwise $X_t \cong Y_t$ for all $0 < t \in \mathrm{spec}(X) \bigcup \mathrm{spec}(Y)$, then $u_{\mathrm{GH}}(X, Y) = 0$).

Detecting isometry of finite ultrametric spaces. Suppose (X, u_X) is a finite UMS with cardinality n, then the weighted tree generated by X (see in page 56) has $O(n)$ many vertices. **is_iso**, a slight modification of the algorithm in Example 3.2 of [AH74], can determine whether two weighted trees are isomorphic in time $O(\#vertices)$, thus in time $O(n)$.

An algorithm for UGHDU-opt. We distill an algorithmic procedure suggested by Theorem 5.7 in the pseudocode given in Algorithm 8. In the pseudocode, the **QuotientClosed** (Appendix B) function implements the quotient space construction (25) which produces the ultrametric space X_t from the ultrametric space X via quotient under the equivalence relation $\sim_t$ from Definition 5.2. In Algorithm 8, the function **is_iso** determines whether two ultrametric spaces are isometric or not.

Complexity analysis of uGH. Suppose $n = \max(\#X, \#Y)$. Then, it is easy to check (see [GV12]) that $\#\left(\mathrm{spec}(X) \bigcup \mathrm{spec}(Y)\right) = O(n)$. Sorting $\mathrm{spec}(X) \bigcup \mathrm{spec}(Y)$ will take time $O(n\log(n))$. $\mathbf{QuotientClosed}$ has time complexity $O(n)$ (see Appendix B). Since $\#X_t = O(n)$, **is_iso** runs in time $O(n)$ as well. Thus, the time complexity associated to computing $u_{\mathrm{GH}}(X, Y)$ via Algorithm 8 described above is $O(n^2)$. This can of course be reduced to $O(n\log(n))$ by using binary search instead of exhaustive search over the parameter t.

Algorithm 8: uGH(X, Y)

```
spec ← sort(spec(X) ∪ spec(Y), 'descend');
for i = 1 : length(spec) do
    t = spec(i);
    if ~ is_iso(QuotientClosed(X, t), QuotientClosed(Y, t)) then
        return t
```

Extension of our algorithmic procedures to the case of treegrams Analogously to the case of ultrametric spaces, the structural theorem (Theorem 5.23) of u_{GH} on the collection $\mathcal{U}^w$ of all finite ultra-dissimilarity spaces allows us to devise an algorithm similar to Algorithm 8 to compute u_{GH} between ultra-dissimilarity spaces. The same argument in complexity analysis of Algorithm 8 can be adapted to show that the time complexity computing u_{GH} between ultra-dissimilarity spaces is still $O(n\log(n))$.

9 Discussion

We introduced Gromov-Hausdorff like distances $d_{\mathrm{GH}}^{(p)}$ on the collection $\mathcal{M}_p$ of all p-metric spaces which make each $(\mathcal{M}_p, d_{\mathrm{GH}}^{(p)})$ into a (pseudo) p-metric space in itself. We studied the convergence of p-metric spaces under $d_{\mathrm{GH}}^{(p)}$ and proved a pre-compactness theorem for $(\mathcal{M}_p, d_{\mathrm{GH}}^{(p)})$. We elucidated distortion characterizations for $d_{\mathrm{GH}}^{(p)}$ as in Theorems 3.5 and 3.7. When $p = \infty$, $\mathcal{M}_p$ becomes the collection of all ultrametric spaces, on which there is a natural extant distance called the interleaving distance. We found a distortion characterization for this interleaving distance in Theorem 6.12. We further generalized the interleaving distance to p-interleaving distances and establish its equivalence with $d_{\mathrm{GH}}^{(p)}$. Moreover, we adapted the interleaving distance to the setting of arbitrary metric spaces and obtained a new lower bound of d_{GH}. Finally, in the computational front, we exploited properties of ultrametric spaces and created two efficient algorithms for estimating the Gromov-Hausdorff distance between two ultrametric spaces. We showed that within certain subsets of $\mathcal{U}$, the time complexity of our algorithms are of polynomial time. Both algorithms can be generalized to computing the exact value of u_{GH} between two ultrametric spaces. By exploiting the structural theorem, we further created a faster algorithm for computing u_{GH}, which we showed can be adapted to computing u_{GH} between ultra-dissimilarity spaces. It is plausible that our computational results can be generalized to the class of finite tree metric spaces.

Acknowledgements

We thank Prof. Phillip Bowers from FSU for posing questions leading to the results in Section 4. We also thank Samir Chowdhury for interesting conversations about geodesics on Gromov-Hausdorff space. This work was partially supported by the NSF through grants DMS-1723003, CCF-1740761, and CCF-1526513.

References

[AH74] Alfred V Aho and John E Hopcroft. *The design and analysis of computer algorithms*. Pearson Education India, 1974.

[AMCIL08] Manuel Alonso-Morón, Eduardo Cuchillo-Ibanez, and Ana Luzón. ε-connectedness, finite approximations, shape theory and coarse graining in hyperspaces. *Physica D: Nonlinear Phenomena*, 237(23):3109–3122, 2008.

[Bar96] Yair Bartal. Probabilistic approximation of metric spaces and its algorithmic applications. In *Proceedings of 37th Conference on Foundations of Computer Science*, pages 184–193. IEEE, 1996.

[BBI01] Dmitri Burago, Yuri Burago, and Sergei Ivanov. *A course in metric geometry*, volume 33. American Mathematical Soc., 2001.

[BS14] Peter Bubenik and Jonathan A Scott. Categorification of persistent homology. *Discrete & Computational Geometry*, 51(3):600–627, 2014.

[Cho19] Samir Chowdhury. *Metric and Topological Approaches to Network Data Analysis*. PhD thesis, The Ohio State University, 2019.

[CLRS09] Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, and Clifford Stein. *Introduction to algorithms*. MIT press, 2009.

[CM10] Gunnar Carlsson and Facundo Mémoli. Characterization, stability and convergence of hierarchical clustering methods. *Journal of machine learning research*, 11(Apr):1425–1470, 2010.

[CM16] Samir Chowdhury and Facundo Mémoli. Constructing geodesics on the space of compact metric spaces. *arXiv preprint arXiv:1603.02385*, 2016.

[CM18] Samir Chowdhury and Facundo Mémoli. The metric space of networks. *arXiv preprint arXiv:1804.02820*, 2018.

[DD09] Michel Marie Deza and Elena Deza. Encyclopedia of distances. In *Encyclopedia of distances*, pages 1–583. Springer, 2009.

[DSSP97] Guy David, Stephen W Semmes, Stephen Semmes, and Guy Rene Pierre Pierre. *Fractured fractals and broken dreams: self-similar geometry through metric and measure*, volume 7. Oxford University Press, 1997.

[FKW95] Martin Farach, Sampath Kannan, and Tandy Warnow. A robust model for finding optimal evolutionary trees. *Algorithmica*, 13(1):155–179, Feb 1995.

[Gro81] Mikhail Gromov. Groups of polynomial growth and expanding maps (with an appendix by jacques tits). *Publications Mathématiques de l'IHÉS*, 53:53–78, 1981.

[Gro87] Mikhail Gromov. Hyperbolic groups. In *Essays in group theory*, pages 75–263. Springer, 1987.

[Gro07] Mikhail Gromov. *Metric structures for Riemannian and non-Riemannian spaces*. Springer Science & Business Media, 2007.

[GV12] Vladimir Gurvich and Mikhail Vyalyi. Characterizing (quasi-) ultrametric finite spaces in terms of (directed) graphs. *Discrete Applied Mathematics*, 160(12):1742–1756, 2012.

[How95] John Mackintosh Howie. *Fundamentals of semigroup theory*, volume 12. 1995.

[JS71] Nicholas Jardine and Robin Sibson. *Mathematical Taxonomy*. Wiley series in probability and mathematical statistics. Wiley, 1971.

[KO99] Nigel J Kalton and Mikhail I Ostrovskii. Distances between banach spaces. In *Forum Mathematicum*, volume 11, pages 17–48. Walter de Gruyter, 1999.

[KR06] Brian W Kernighan and Dennis M Ritchie. *The C programming language*. 2006.

[MBW13] Dmitriy Morozov, Kenes Beketayev, and Gunther Weber. Interleaving distance between merge trees. *Discrete and Computational Geometry*, 49(22-45):52, 2013.

[Mém07] Facundo Mémoli. On the use of Gromov-Hausdorff distances for shape comparison. 2007.

[Mém11] Facundo Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417–487, 2011.

[Mém12] Facundo Mémoli. Some properties of Gromov–Hausdorff distances. *Discrete & Computational Geometry*, 48(2):416–440, 2012.

[MO18] Facundo Mémoli and Osman Berat Okutan. Quantitative simplification of filtered simplicial complexes. *arXiv preprint arXiv:1801.02812*, 2018.

[Pat18] Amit Patel. Generalized persistence diagrams. *Journal of Applied and Computational Topology*, 1(3-4):397–419, 2018.

[Qiu09] Derong Qiu. Geometry of non-archimedean Gromov-Hausdorff distance. *P-Adic Numbers, Ultrametric Analysis, and Applications*, 1(4):317, 2009.

[Sch15] Felix Schmiedl. *Shape matching and mesh segmentation*. PhD thesis, Technische Universität München, 2015.

[Sch17] Felix Schmiedl. Computational aspects of the Gromov–Hausdorff distance and its application in non-rigid shape matching. *Discrete & Computational Geometry*, 57(4):854–880, 2017.

[SCM16] Zane Smith, Samir Chowdhury, and Facundo Mémoli. Hierarchical representations of network data with optimal distortion bounds. In *2016 50th Asilomar Conference on Signals, Systems and Computers*, pages 1834–1838. IEEE, 2016.

[SCMR15] Santiago Segarra, Gunnar Carlsson, Facundo Mémoli, and Alejandro Ribeiro. Metric representations of network data. *preprint*, 2015.

[Seg16] Santiago Segarra. *Metric Representations Of Networks*. PhD thesis, University of Pennsylvania, 2016.

[Sem03] Stephen Semmes. Notes on metrics, measures, and dimensions. *arXiv preprint math/0302190*, 2003.

[SS03] Charles Semple and Mike Steel. *Phylogenetics*. Oxford lecture series in mathematics and its applications. Oxford University Press, 2003.

[TW05] Jeremy T Tyson and Jang-Mei Wu. Characterizations of snowflake metric spaces. *Ann. Acad. Sci. Fenn. Math*, 30(2):313–336, 2005.

[TW18] Elena Farahbakhsh Touli and Yusu Wang. FPT-algorithms for computing Gromov-Hausdorff and interleaving distances between trees. *arXiv preprint arXiv:1811.02425*, 2018.

[Zar05] Ihor Zarichnyi. Gromov-Hausdorff ultrametric. *arXiv preprint math/0511437*, 2005.

A Proof of Theorem 5.23

Proof. We first prove a weaker version (with $\inf$ instead of $\min$):

$$u_{\mathrm{GH}}(X,Y) = \inf\left\{t \geq 0 : (X_t, u_{X_t}) \cong (Y_t, u_{Y_t})\right\}. \tag{27}$$

Suppose first that $X_t^R \cong Y_t^R$ for some $t \geq 0$, i.e. there exists an isometry $f_t : X_t^R \to Y_t^R$. Then, we construct a correspondence between X and Y as

$$R_t := \left\{(x,y) \in X \times Y : y \in f_t([\![x]\!]_t^X)\right\}.$$

Equivalently, $R_t = \left\{(x,y) \in X \times Y : x \in f_t^{-1}([\![y]\!]_t^Y)\right\}$. That R_t is a correspondence between X and Y follows from the fact that f_t is bijective.

Then, we show that $\mathrm{dis}_\infty(R_t) \leq t$, which will imply that $u_{\mathrm{GH}}(X,Y) \leq t$. Indeed, for $(x,y),(x',y') \in R_t$, if $u_X(x,x') \leq t$, then we already have $u_X(x,x') \leq \max(t, u_Y(y,y'))$. Otherwise, if $u_X(x,x') > t$, we have the following two cases:

1. $x = x'$. Then $[\![x]\!]_t^X = [\![x']\!]_t^X = \{x\}$. Thus, $[\![y]\!]_t^Y = f_t([\![x]\!]_t^X) = f_t([\![x']\!]_t^X) = [\![y']\!]_t^Y$. Since f_t is isometry, we have $u_{Y_t}([\![y]\!]_t^Y, [\![y]\!]_t^Y) = u_{X_t}([\![x]\!]_t^X, [\![x]\!]_t^X) = u_X(x,x) > 0$. This implies $u_Y(y,y) = u_{Y_t}([\![y]\!]_t^Y, [\![y]\!]_t^Y) = u_X(x,x) > t$, and similarly $u_Y(y',y') > t$. Combining with $[\![y]\!]_t^Y = [\![y']\!]_t^Y$, we have $y = y'$ and thus $u_Y(y,y') = u_X(x,x')$.

2. $x \neq x'$. Then $[\![x]\!]_t^X \neq [\![x']\!]_t^X$. Thus, by definition of u_{Y_t} and u_{X_t}, and the fact that f_t is an isometry, we have $[\![y]\!]_t^Y = f_t([\![x]\!]_t^X) \neq f_t([\![x']\!]_t^X) = [\![y']\!]_t^Y$ and hence $u_Y(y,y') = u_{Y_t}([\![y]\!]_t^Y, [\![y']\!]_t^Y) = u_{X_t}([\![x]\!]_t^X, [\![x']\!]_t^X) = u_X(x,x')$.

Therefore $u_X(x,x') \leq \max(t, u_Y(y,y'))$. Similarly we can show that $u_Y(y,y') \leq \max(t, u_X(x,x'))$. Hence, by Equation (15) we have that $\mathrm{dis}_\infty(R_t) \leq t$. Thus $u_{\mathrm{GH}}(X,Y) \leq \inf\left\{t \geq 0 : X_t^R \cong Y_t^R\right\}$.

Conversely, let R be a correspondence between X an Y with $\mathrm{dis}_\infty(R) = t$. Now for any $(x,y),(x',y') \in R$ with $[\![x']\!]_t^X = [\![x]\!]_t^X$, we have the following two cases:

1. $u_X(x,x') \leq t$. Then, by definition of $\mathrm{dis}_\infty(R)$, we have $u_Y(y',y) \leq \max\left(t, u_X(x,x')\right) \leq t$ which implies that $[\![y]\!]_t^Y = [\![y']\!]_t^Y$.

2. $u_X(x,x) > t$ and $x = x'$. Then, by definition of $\mathrm{dis}_\infty(R)$, $u_Y(y,y') = u_X(x,x') = u_X(x,x) > t$. Similarly $u_Y(y,y) = u_X(x,x) = u_X(x',x') = u_Y(y',y')$. Therefore $y = y'$ by condition 3 of the definiton of ultra-dissimilarity spaces. Thus, $[\![y]\!]_t^Y = [\![y']\!]_t^Y$.

Hence, the map $f : X \to Y$ taking x to any y such that $(x,y) \in R$ induces a well-defined map $f_t : X_t^R \to Y_t^R$, with $f_t([\![x]\!]_t^X) = [\![f(x)]\!]_t^Y$. There is also a well-defined map $g_t : Y_t^R \to X_t^R$ induced by a map $g : Y \to X$ with $g(y) = x$ where x is chosen such that $(x,y) \in R$. It is clear that g_t is the inverse of f_t and hence f_t is bijective. Now we show that f_t is an isometry. Assume $u_{X_t}([\![x]\!]_t^X, [\![x']\!]_t^X) = s > t$, which implies that $u_X(x,x') = s$. Then, by the characterization of $\mathrm{dis}_\infty(R)$ given by Equation (15), $u_Y(y,y')$ is forced to be s and thus $u_{Y_t}([\![y]\!]_t^Y, [\![y']\!]_t^Y) = s$, where $y = f(x)$ and $y' = f(x')$. Next, if $[\![x]\!]_t^X = [\![x']\!]_t^X$, again we have the following two cases:

1. $u_X(x,x) \leq t$. Then $[\![x]\!]_t^X = [\![x']\!]_t^X$ implies that $u_X(x,x') \leq t$ and $u_{X_t}([\![x]\!]_t^X, [\![x']\!]_t^X) = 0$. As for $f(x)$, since $(x, f(x)) \in R$, we have that $u_Y(f(x), f(x)) \leq \max(t, u_X(x,x)) \leq t$. Hence, $u_{Y_t}([\![f(x)]\!]_t^Y, [\![f(x')]\!]_t^Y) = u_{Y_t}([\![f(x)]\!]_t^Y, [\![f(x)]\!]_t^Y) = 0 = u_X([\![x]\!]_t^X, [\![x']\!]_t^X)$.

2. $u_X(x,x) > t$. Then, $x = x'$ and $u_{X_t}([\![x]\!]_t^X, [\![x']\!]_t^X) = u_X(x,x) > t$. Similar with case 1, we have $u_X(x,x) \leq \max(t, u_Y(f(x), f(x)))$, which forces $u_Y(f(x), f(x)) = u_X(x,x) > t$. Then, $u_{Y_t}([\![f(x)]\!]_t^Y, [\![f(x')]\!]_t^Y) = u_{Y_t}([\![f(x)]\!]_t^Y, [\![f(x)]\!]_t^Y) = u_X(x,x) = u_{X_t}([\![x]\!]_t^X, [\![x]\!]_t^X) = u_{X_t}([\![x]\!]_t^X, [\![x']\!]_t^X)$.

This proves that f_t is an isometry and thus $u_{\mathrm{GH}}(X,Y) \geq \inf\{t \geq 0 : X_t \cong Y_t\}$. □

In [CM18, Theorem 35], the authors generalized Gromov's reconstruction theorem (Theorem 5.10) to the setting of compact networks. This allows us to complete the proof of Theorem 5.23.

Definition A.1 (Motif sets [CM18])**.** *For a ultra-dissimilarity space X, and a positive integer n, let $\Psi_X^{(n)} : X^{\times n} \to \mathbb{R}^{n\times n}$ be the function given by $(x_1, \ldots, x_n) \mapsto (u_X(x_i, x_j))_{i,j=1}^n$. Then, the motif set of order n associated to X is defined as*

$$\mathrm{M}_n(X) := \mathrm{im}\left(\Psi_X^{(n)}\right).$$

Theorem A.2 ([CM18])**.** *Given two finite ultra-dissimilarity spaces X and Y, if $\mathrm{M}_n(X) = \mathrm{M}_n(Y)$ for every $n \in \mathbb{N}$, then $X \cong Y$.*

Proof that infimum in Equation (27) is a minimum. Denote $t_0 = u_{\mathrm{GH}}(X,Y)$, then for any $\delta > 0$ small, there exists $0 < \varepsilon < \delta$ such that $X_{t_0+\varepsilon} \cong Y_{t_0+\varepsilon}$. Fix a positive natural number n. Consider the motif sets $\mathrm{M}_n(X_{t_0})$ and $\mathrm{M}_n(Y_{t_0})$. For any n points $[\![x_1]\!]_{t_0}^X, \cdots, [\![x_n]\!]_{t_0}^X \in X_{t_0}$, without loss of generality, we can assume that $[\![x_i]\!]_{t_0}^X \neq [\![x_j]\!]_{t_0}^X$ for any $i \neq j$. Then, there exists $\varepsilon > 0$ small enough such that $[\![x_i]\!]_{t_0+\varepsilon}^X \neq [\![x_j]\!]_{t_0+\varepsilon}^X$, $u_{X_t}([\![x_i]\!]_{t_0+\varepsilon}^X, [\![x_i]\!]_{t_0+\varepsilon}^X) = u_{X_t}([\![x_i]\!]_{t_0}^X, [\![x_i]\!]_{t_0}^X)$ for all $i \neq j$, and $X_{t_0+\varepsilon} \cong Y_{t_0+\varepsilon}$. This has the following two consequences:

1. $u_{X_{t_0+\varepsilon}}\left([\![x_i]\!]_{t_0+\varepsilon}^X, [\![x_j]\!]_{t_0+\varepsilon}^X\right) = u_{X_{t_0}}\left([\![x_i]\!]_{t_0}^X, [\![x_j]\!]_{t_0}^X\right)$ for all $i,j \in \{1, \cdots, n\}$.
2. There exist $[\![y_1]\!]_{t_0+\varepsilon}^Y, \cdots, [\![y_n]\!]_{t_0+\varepsilon}^Y \in Y_{t_0+\varepsilon}$ such that
$$u_{Y_{t_0+\varepsilon}}\left([\![y_i]\!]_{t_0+\varepsilon}^Y, [\![y_j]\!]_{t_0+\varepsilon}^Y\right) = u_{X_{t_0+\varepsilon}}\left([\![x_i]\!]_{t_0+\varepsilon}^X, [\![x_j]\!]_{t_0+\varepsilon}^X\right)$$
for all $i \neq j$. Thus,$u_{Y_{t_0+\varepsilon}}\left([\![y_i]\!]_{t_0+\varepsilon}^Y, [\![y_j]\!]_{t_0+\varepsilon}^Y\right) = u_{Y_{t_0}}\left([\![y_i]\!]_{t_0}^Y, [\![y_j]\!]_{t_0}^Y\right)$ for all $i \neq j$.

Therefore $u_{Y_{t_0}}\left([\![y_i]\!]_{t_0}^Y, [\![y_j]\!]_{t_0}^Y\right) = u_{X_{t_0}}\left([\![x_i]\!]_{t_0}^X, [x_j]_{t_0}^X\right)$ for all $i \neq j$. Now, we only need to show $u_{Y_{t_0}}\left([\![y_i]\!]_{t_0}^Y, [\![y_i]\!]_{t_0}^Y\right) = u_{X_{t_0}}\left([\![x_i]\!]_{t_0}^X, [\![x_i]\!]_{t_0}^X\right)$ for any $i \in \{1, \cdots, n\}$. This is not true in general since it may happen that $t_0 < u_Y(y_i, y_i) \leq t_0 + \varepsilon$ making that $u_{Y_{t_0}}\left([\![y_i]\!]_{t_0}^Y, [\![y_i]\!]_{t_0}^Y\right) \neq 0$ but $u_{Y_{t_0+\varepsilon}}\left([\![y_i]\!]_{t_0+\varepsilon}^Y, [\![y_i]\!]_{t_0+\varepsilon}^Y\right) = 0$. Let $\varepsilon = \varepsilon_1 > \cdots > \varepsilon_k > \cdots > 0$ be a decreasing sequence such that $X_{t_0+\varepsilon_k} \cong Y_{t_0+\varepsilon_k}$. For each integer k, pick n points $[\![y_1^{(k)}]\!]_{t_0+\varepsilon_k}^Y, \cdots, [\![y_n^{(k)}]\!]_{t_0+\varepsilon_k}^Y \in Y_{t_0+\varepsilon_k}$ such that item 2 above holds. Since Y is a finite space, there will be a subsequence of $\{\varepsilon_k\}_{k=1}^\infty$, without loss of generality, still denoted as $\{\varepsilon_k\}_{k=1}^\infty$, such that there exist $y_1, \cdots, y_n \in Y$ and

$$[\![y_1^{(k)}]\!]_{t_0+\varepsilon_k}^Y = [\![y_i]\!]_{t_0+\varepsilon_k}^Y, \quad \forall k = 1, \cdots, n.$$

This implies that $[\![y_i]\!]_{t_0+\varepsilon_1}^Y = [\![y_i]\!]_{t_0+\varepsilon_2}^Y = \cdots$. Thus, either $u_Y(y_i, y_i) > t_0 + \varepsilon_1 > t$ or $u_Y(y_i, y_i) \leq t_0$. In either case, we have $u_{Y_{t_0}}([\![y_i]\!]_{t_0}, [\![y_i]\!]_{t_0}) = u_{Y_{t_0+\varepsilon}}([\![y_i]\!]_{t_0+\varepsilon}, [\![y_i]\!]_{t_0})$. Then, for such choice of $[\![y_i]\!]_{t_0+\varepsilon}$, we have

$$\left(u_{X_{t_0}}\left([\![x_i]\!]_{t_0}^X, [\![x_j]\!]_{t_0}^X\right)\right)_{i,j=1}^n = \left(u_{Y_{t_0}}\left([\![y_i]\!]_{t_0}^Y, [\![y_j]\!]_{t_0}^Y\right)\right)_{i,j=1}^n.$$

This implies that $\mathrm{M}_n(X_{t_0}) \subset \mathrm{M}_n(Y_{t_0})$. Similarly we have that $\mathrm{M}_n(Y_{t_0}) \subset \mathrm{M}_n(X_{t_0})$ and thus $\mathrm{M}_n(Y_{t_0}) = \mathrm{M}_n(X_{t_0})$. Since n is arbitrary, then by Theorem A.2 we have that $X_{t_0} \cong Y_{t_0}$, which implies that $t_0 = \min\{t \geq 0 : X_t \cong Y_t\}$. □

B Algorithms for fundamental operations

In this section, we introduce algorithms for three fundamental operations on ultrametric spaces. Recall that in the beginning of Section 8, we represent each ultrametric spaces by weighted tree data structure. Note that each tree is unique determined by its root. So for an ultrametric space X, we also use X to represent the root node of the tree. We denote by $\{X_1, \cdots, X_m\}$ the list of children of X.

Closed quotient. In Algorithm 8 and Algorithm 2, we need to take the closed quotient of ultrametric spaces (Definition 5.2). We introduce a recursive algorithm for the operation in Algorithm 9. We use the one point tree to throw away information under certain scale. A one point tree is a node with empty list of children, zero diameter and a singleton.

Algorithm 9: QuotientClosed(X, t)

```
Y ← X
for Y_i ∈ children(Y) do
    if diam(Y_i) < t then
        Y_i ← QuotientClosed(Y_i, t)
    else
        replace Y_i by a one point tree with the singleton in Y_i
return Y
```

Open partition. In Algorithm 1 and Algorithm 3, we use the open equivalence relation to partition ultrametric spaces. We present a recursive algorithm for the open partition operation in Algorithm 10.

Algorithm 10: PartitionOpen(X, t)

```
Y = []
if diam(X) ≤ t then
    Y.append(X)
else
    for X_i ∈ children(X) do
        Y.append(PartitionOpen(X_i, t))
return Y
```

Union of ultrametric spaces. We also take unions of ultrametric spaces represented in weighted trees in Algorithm 1 and Algorithm 3. Since the subset information is stored in the node, it is direct to take the union of subsets of an ultrametric space X represented by weighted trees. We simply create a root node containing the list of pointers referring to the roots of these subsets, the union of these subsets and the diameter of the union.

It is obvious that both of the closed quotient and open partition algorithms can be done in time $O(n)$ with n being the number of tree nodes. As for the union, it can be done in time $O(m)$ with m being the cardinality of the union of subsets.

C Proof of Proposition 8.8

The proof of the proposition follows from a series of lemmas. In the following three lemmas, we will always assume that $X, Y \in \mathcal{U}$ and there is $\varepsilon \geq 0$ such that $\operatorname{diam}(X) > \varepsilon$ and $\operatorname{diam}(Y) \leq \varepsilon$.

Lemma C.1. *There exists a correspondence R between X and Y with $\operatorname{dis}(R) \leq \varepsilon$ if and only if there exists a correspondence R_ε between X_ε and Y with $\operatorname{dis}(R_\varepsilon) \leq \varepsilon$.*

Proof. Suppose R is a correspondence between X and Y with $\operatorname{dis}(R) \leq \varepsilon$. Then, we build a correspondence R_ε between X_ε and Y as follows:

$$R_\varepsilon \coloneqq \{([x]_\varepsilon^X, y) : (x, y) \in R\}.$$

It is easy to see that R_ε is a correspondence between X_ε and Y. Notice that for any two pairs $([x]_\varepsilon^X, y), ([x']_\varepsilon^X, y') \in R_\varepsilon$, we have

$$\left|u_{X_\varepsilon}([x]_\varepsilon^X, [x']_\varepsilon^X) - u_Y(y, y')\right| = \begin{cases} |u_X(x, x') - u_Y(y, y')| & \text{if } [x]_\varepsilon^X \neq [x']_\varepsilon^X \\ u_Y(y, y') & \text{if } [x]_\varepsilon^X = [x']_\varepsilon^X. \end{cases}$$

In each of the two cases the quantity is bounded above by ε, therefore, $\operatorname{dis}(R_\varepsilon) \leq \varepsilon$.

Now suppose there exists a correspondence R_ε between X_ε and Y with $\operatorname{dis}(R_\varepsilon) \leq \varepsilon$. Then, we build a correspondence R between X and Y as follows:

$$R \coloneqq \{(x, y) : ([x]_\varepsilon^X, y) \in R\}.$$

It is easy to check that R is a correspondence between X and Y. For any $(x, y), (x', y') \in R$, we have

$$|u_X(x, x') - u_Y(y, y')| \leq \begin{cases} |u_{X_\varepsilon}([x]_\varepsilon^X, [x']_\varepsilon^X) - u_Y(y, y')| & \text{if } u_X(x, x') > \varepsilon \\ \varepsilon - u_Y(y, y') & \text{if } u_X(x, x') \leq \varepsilon. \end{cases}$$

Therefore, $\operatorname{dis}(R) \leq \varepsilon$. $\square$

Lemma C.2. *Any correspondence R_ε between X_ε and Y with $\operatorname{dis}(R_\varepsilon) \leq \varepsilon$ is the graph of a surjection from Y to X_ε.*

Proof. We only need to show that if $([x]_\varepsilon^X, y), ([x']_\varepsilon^X, y) \in R_\varepsilon$, then $[x]_\varepsilon^X = [x']_\varepsilon^X$. Otherwise suppose $[x]_\varepsilon^X \neq [x']_\varepsilon^X$, which is equivalent to $u_X(x, x') > \varepsilon$. Then,

$$\varepsilon \geq \operatorname{dis}(R_\varepsilon) \geq \left|u_{X_\varepsilon}([x]_\varepsilon^X, [x']_\varepsilon^X) - u_Y(y, y)\right| = u_X(x, x') > \varepsilon,$$

which is a contradiction! $\square$

Recall that $\operatorname{sep}(X) = \min\{d(x, x') : x, x' \in X\}$.

Lemma C.3. *Assume* $\operatorname{sep}(X) > \varepsilon$*, then any injection* $\varphi : X \to Y$ *with* $\operatorname{dis}(\varphi) \le \varepsilon$ *will induce a correspondence* R *between* X *and* Y *with* $\operatorname{dis}(R) \le \varepsilon$*.*

Proof. Suppose $X = \{x_1, \cdots, x_n\}$ and $\operatorname{im}(\varphi) = \{y_1, \cdots, y_n\}$ where $y_i = \varphi(x_i)$. For any $y \in Y$, define

$$i_y := \min\left\{\operatorname*{argmin}_{j=1,\cdots,n} u_Y(y, y_j)\right\}.$$

Obviously, $i_{y_j} = j$. Then, we build a correspondence R between X and Y as follows:

$$R := \{(x_{i_y}, y) : \forall y \in Y\}.$$

Now we only need to check that $\operatorname{dis}(R) \le \varepsilon$. Let $(x_i, y), (x_j, y') \in R$. If $i = j$, then $|u_X(x_i, x_i) - u_Y(y, y')| = u_Y(y, y') \le \operatorname{diam}(Y) \le \varepsilon$. Now assume $i \neq j$. Since $u_X(x_i, x_j) - u_Y(y, y') > 0$ and $u_X(x_i, x_j) - u_Y(y_i, y_j) \le \operatorname{dis}(\varphi) \le \varepsilon$, the inequality $|u_X(x_i, x_j) - u_Y(y, y')| \le \varepsilon$ follows from the following observation:

Claim 5. *For* $y, y' \in Y$*, if* $i_y \neq i_{y'}$*, then* $u_Y(y, y') \ge u_Y(y_i, y_j)$*.*

Proof of Claim 5. Let $i = i_y$ and $j = i_{y'}$. Suppose otherwise $u_Y(y, y') < u_Y(y_i, y_j)$. If $u_Y(y_i, y) \le u_Y(y, y')$, then $u_Y(y', y_i) \le \max(u_Y(y, y'), u_Y(y, y_i)) \le u_Y(y, y_i)$. By definition of $j = i_{y'}$, we have that $u_Y(y_j, y') \le u_Y(y_i, y') \le u_Y(y, y')$. Hence, $u_Y(y_i, y_j) \le \max(u_Y(y_i, y'), u_Y(y', y_j)) \le u_Y(y, y')$, contradiction. Therefore, $u_Y(y_i, y) > u_Y(y, y')$ and similarly $u_Y(y_j, y') > u_Y(y, y')$. Combining with the assumption $u_Y(y, y') < u_Y(y_i, y_j)$, we have that $u_Y(y, y_i) = u_Y(y', y_i)$ and $u_Y(y, y_j) = u_Y(y', y_j)$. By definition of $i = i_y$ and $j = i_{y'}$, we have that

$$u_Y(y, y_j) = u_Y(y', y_j) \le u_Y(y', y_i) = u_Y(y, y_i),$$

which implies that $j \in \operatorname{argmin}_{k=1,\cdots,n} u_Y(y, y_k)$ and thus $j > i$. However, similarly we can prove $i > j$, contradiction! Therefore, $u_Y(y, y') \ge u_Y(y_i, y_j)$. □

□

Proof of Proposition 8.8. By Lemma C.1, we only need to show that there exists a correspondence R_ε between X_ε and Y with $\operatorname{dis}(R_\varepsilon) \le \varepsilon$ if and only if there exists an injection $\varphi : X_\varepsilon \to Y$ with $\operatorname{dis}(\varphi) \le \varepsilon$.

Assuming the existence of such a correspondence R_ε, then by Lemma C.2, there exists a surjection $\psi : Y \to X_\varepsilon$ such that $\operatorname{graph}(\psi) = R_\varepsilon$. Then, we construct an injection $\varphi : X_\varepsilon \to Y$ by taking $[x]_\varepsilon$ to y, where y is arbitrarily chosen from $\varphi^{-1}([x]_\varepsilon)$. It is easy to see that $\operatorname{dis}(\varphi) \le \operatorname{dis}(\psi) \le \varepsilon$.

Now, assume that there exists an injection $\varphi : X_\varepsilon \to Y$ with $\operatorname{dis}(\varphi) \le \varepsilon$. Obviously, we have $\operatorname{sep}(X) > \varepsilon$, and thus by Lemma C.3, there exists a correspondence R_ε between X_ε and Y with $\operatorname{dis}(R_\varepsilon) \le \varepsilon$. □

D Proof of Lemma 8.13

Proof. Denote $\tilde{\gamma}_\varepsilon(X) = \max\left\{\#\left\{[x']^X_{\hat{t}_\varepsilon} : x' \in [x]^X_t\right\} : t \in \operatorname{spec}(X)\right\}$.

We first show that X satisfies the second $(\varepsilon, \tilde{\gamma}_\varepsilon)$-growth condition. Take any $t \ge 0$, consider the set $S_t := \{[x']_{\hat{t}_\varepsilon} : x' \in [x]_t\}$. Suppose $\operatorname{spec}(X) = \{0 = w_0 < w_1 < \cdots < w_N\}$. If

$t \geq w_N = \operatorname{diam}(X)$, then $[x]_t^X = [x]_{w_N}^X$ and $[x']_{(\mathring{w}_N)_\varepsilon} \subset [x']_{\mathring{t}_\varepsilon}$. Thus, it is easy to check that $\#S_t \leq \#S_{w_N} \leq \tilde{\gamma}_\varepsilon$. Now suppose there exists $i \in \{0, \cdots, N-1\}$ such that $w_i \leq t < w_{i+1}$. Then, similarly we have $[x]_t^X = [x]_{w_i}^X$ and $[x']_{(\mathring{w}_i)_\varepsilon} \subset [x']_{\mathring{t}_\varepsilon}$. Thus, $\#S_t \leq \#S_{w_i} \leq \tilde{\gamma}_\varepsilon$. Therefore, X satisfies the second $(\varepsilon, \tilde{\gamma}_\varepsilon)$-growth condition.

Next, we show that if X satisfies the second (ε, γ)-growth condition, then $\gamma \geq \tilde{\gamma}_\varepsilon$. Indeed, for any $t \in \operatorname{spec}(X)$ and $x \in X$, we have that $\#S_t \leq \gamma$ and thus $\tilde{\gamma}_\varepsilon(X) = \min\{\#S_t : t \in \operatorname{spec}(X)\} \leq \gamma$.

Therefore, $\gamma_\varepsilon(X) = \tilde{\gamma}_\varepsilon(X)$.

□

E Proof of Lemma 8.22

Proof. Consider $\#\{[x']_{(t-\mathring{2}\varepsilon^p)}^{S_p(X)} : x' \in [x]_t^{S_p(X)}\}$ for any $x \in X$ and $t \geq 2\varepsilon^p$. It is easy to check the following:

$$\{[x']_{(t-\mathring{2}\varepsilon^p)}^{S_p(X)} : x' \in [x]_t^{S_p(X)}\} = \{[x']^X_{\overset{\circ}{(t-2\varepsilon^p)^{\frac{1}{p}}}} : x' \in [x]^X_{t^{\frac{1}{p}}}\}.$$

By Lemma 3.12,

$$(t-2\varepsilon^p)^{\frac{1}{p}} = \Lambda_p(t^{\frac{1}{p}}, 2^{\frac{1}{p}}\varepsilon) \geq t^{\frac{1}{p}} - 2^{\frac{1}{p}}\varepsilon \geq t^{\frac{1}{p}} - 2\varepsilon.$$

Hence,

$$\#\{[x']_{(t-\mathring{2}\varepsilon^p)}^{S_p(X)} : x' \in [x]_t^{S_p(X)}\} = \#\{[x']^X_{\overset{\circ}{(t-2\varepsilon^p)^{\frac{1}{p}}}} : x' \in [x]^X_{t^{\frac{1}{p}}}\} \leq \#\{[x']^X_{\overset{\circ}{(t^{\frac{1}{p}}-2\varepsilon)}} : x' \in [x]^X_{t^{\frac{1}{p}}}\}.$$

Then, since X satisfies the second $(\varepsilon, \gamma_\varepsilon(X))$-growth condition, we have that

$$\#\{[x']_{(t-\mathring{2}\varepsilon^p)}^{S_p(X)} : x' \in [x]_t^{S_p(X)}\} \leq \gamma_\varepsilon(X)$$

and thus $\gamma_{\varepsilon^p}(S_p(X)) \leq \gamma_\varepsilon(X)$. Similarly, one can show that $\gamma_{\varepsilon^p}(S_p(X)) \leq \gamma_{\varepsilon^q}(S_q(X))$ when $p > q \geq 1$. Therefore, $\gamma_{\varepsilon^p}(S_p(X))$ is a decreasing function with respect to p.

By Lemma 8.13, we have

$$\gamma_\varepsilon(X) = \max\left\{\#\left\{[x']^X_{A_1(\mathring{t},2\varepsilon)} : x' \in [x]_t^X\right\} : t \in \operatorname{spec}(X)\right\}$$

and then it follows that

$$\gamma_{\varepsilon^p}(S_p(X)) = \max\left\{\#\left\{[x']^X_{\overset{\circ}{A_p(t,2^{\frac{1}{p}}\varepsilon)}} : x' \in [x]_t^X\right\} : t \in \operatorname{spec}(X)\right\}.$$

When p goes to infinity, $A_p\left(t, 2^{\frac{1}{p}}\varepsilon\right)$ approaches t if $t \geq \varepsilon$ or 0 if $t < \varepsilon$. Thus, when p is large enough

$$\left\{[x']^X_{\overset{\circ}{A_p(t,2^{\frac{1}{p}}\varepsilon)}} : x' \in [x]_t^X\right\} = \begin{cases} \{[x']^X_{\mathring{t}} : x' \in [x]_t^X\} & \text{if } t \geq \varepsilon \\ \{[x']^X_{\mathring{0}} : x' \in [x]_t^X\} & \text{if } t < \varepsilon. \end{cases}$$

Then, since X is finite, it is easy to see that $\gamma_{\varepsilon^p}(S_p(X))$ will become fixed when p is large enough.

□