Recovering metric from full ordinal information

Thibaut Le Gouic

September 16, 2015

Abstract

Given a geodesic space (E, d), we show that full ordinal knowledge on the metric d - i.e. knowledge of the function

$$D_d : (w, x, y, z) \mapsto \mathbf{1}_{d(w,x) \leq d(y,z)},$$

determines uniquely - up to a constant factor - the metric d. For a subspace E_n of n points of E, converging in Hausdorff distance to E, we construct a metric d_n on E_n, based only on the knowledge of D_d on E_n and establish a sharp upper bound of the Gromov-Hausdorff distance between (E_n, d_n) and (E, d).

1 Introduction

Given a metric space (E, d) for which only the function

$$D_d : (w, x, y, z) \mapsto \mathbf{1}_{d(w,x) \leq d(y,z)}$$

is known (in other words, the metric itself is unknown but two distances can be compared). Is it possible to recover the metric d?

The answer is clearly *no* when the problem is formulated this way, because multiplying the metric by a constant does not change the known function D_d. More importantly, given a sub-additive positive function l (such that $l(x) = 0 \Leftrightarrow x = 0$), then the composed function $l \circ d$ is a metric that also gives the same observed function:

$$D_d = D_{l \circ d}.$$

However, one can observe that if (E, d) is a geodesic space, then $(E, l \circ d)$ is geodesic only if l is a linear function (i.e. if $f : x \mapsto cx$ for some $c > 0$). Thus, if the space (E, d) is known to be geodesic, the latter argument fails.

This question had been studied in the context of a subsets of $\mathbb{R}^k$ in [KvL14]. In this work, it is shown that under weaker knowledge on the metric (that is D_d is known only for quadruples (w, x, y, z) of the form (x, y, x, z)), the metric (up to a constant factor) is determined by D_d, on Euclidean spaces.

We first show the same result for geodesic spaces, that is D_d determines d up to a constant factor.

The main result of this paper answers how to built a metric on a finite subspace E_n of E that is known to converge in Hausdorff metric to E, when only D_d is known on E_n. Sharp bounds of this convergence are proven.

It thus solves a problem posed in [KvL14],

2 Uniqueness of the metric

In order to set the problem properly, recall the definition of a geodesic space.

Definition 1. *Let (E,d) be a complete metric space. If for any $x,y \in E$, there exists $z \in E$ such that*

$$d(x,z) = d(y,z) = \frac{1}{2}d(x,y),$$

then (E,d) is said to be a geodesic space. And z is called a middle point of (x,y).

A segment $[x,y]$ is a subset of E such that there exists a continuous mapping $\gamma : [0,1] \to E$ such that $\gamma([0,1]) = [x,y]$ and for all $t \in [0,1]$,

$$d(x,\gamma(t)) = td(x,y) \text{ and } d(\gamma(t),y) = (1-t)d(x,y).$$

Our first result can then be stated as following. Metric of geodesic spaces is determined by ordinal information on the metric.

Theorem 2. *Let (E_1,d_1) and (E_2,d_2) be two complete geodesic spaces such that there exists a one-to-one map f such that*

$$D_{d_1} = D_{d_2 \circ f \times f}, \tag{1}$$

then, there exists $c > 0$ such that f is an isometry between (E_1,d_1) and (E_2,cd_2).

Proof. We first show that the result is true when E is restricted to any segment $[w,x]$.

Let $w,x \in E_1$, then since E_1 is geodesic, there exists a middle point m, so that

$$d_1(w,m) = d_1(m,x) = \frac{1}{2}d_1(w,x).$$

Since

$$1 = D_{d_1}(w,m,m,x) = D_{d_2 \circ f \times f}(w,m,m,x) = 1, \text{ and } 1 = D_{d_1}(x,m,m,w) = D_{d_2 \circ f \times f}(x,m,m,w) = 1$$

then,

$$d_2(f(w),f(m)) = d_2(f(m),f(x)).$$

Thus, in order to show that $f(m)$ is a middle point of $[f(w),f(x)]$, is suffices to show that for any m' such that $f(m')$ is a middle point of $[f(w),f(x)]$,

$$d_2(f(w),f(m)) \leq d_2(f(w),f(m')).$$

Suppose that

$$d_2(f(w),f(m)) > d_2(f(w),f(m')),$$

then the equality

$$0 = D_{d_1}(w,m,w,m') = D_{d_2 \circ f \times f}(w,m,w,m') = 0,$$

implies

$$d_1(w,m') < d_1(w,m).$$

Similarly, we can show that

$$d_1(m',x) < d_1(m,x),$$

which contradicts that m is a middle point of $[w, x]$.

We thus showed that middle points are mapped to middle points by f.

Applying this recursively on a segment $[w, x]$, we show that for any $t \in [0, 1]$ of the form

$$t = \frac{k}{2^n}$$

with $k, n \in \mathbb{N}$, and $u_t \in [w, x]$ such that $d_1(w, u_t) = td_1(w, x)$, the following holds

$$f(u_t) \in [f(w), f(x)] \text{ and } d_2(f(w), f(u_t)) = td_2(f(w), f(x)). \tag{2}$$

Since such t are dense in $[0; 1]$, the result holds true for any $t \in [0; 1]$ by continuity. Indeed, since for a sequence $w_t \to w$ there exists a sequence $s_t \to 0$ such that $s_t \geq t$ and s_t is of the form $\frac{k}{2^n}$ with $k, n \in \mathbb{N}$, continuity of f holds using

$$d_1(w, w_t) \leq d_1(w, u_t) \implies d_2(f(w), f(w_t)) \leq d_2(f(w), f(u_t)) = td_2(f(w), f(x)).$$

Thus, we showed that the result holds for any segment (with eventually different constants c).

Take now $w, x, y, z \in E_1$ and set

$$c = \frac{d_1(w, x)}{d_2(f(w), f(x))}.$$

We want to show that constants c are the same for any other segment $[y, z]$, i.e, i.e..

$$d_1(y, z) = cd_2(f(y), f(z)).$$

Without loss of generality, we can suppose that $d_1(y, z) \leq d_1(w, x)$. Thus, there exists $u \in [w, x]$ such that $d_1(y, z) = d_1(w, u) = td_1(w, x)$ for some $t \in [0; 1]$. This equality also provides

$$\begin{aligned}
d_1(y, z) &= td_1(w, x) \\
&= tcd_2(f(w), f(x)) && \text{by definition of } c, \\
&= cd_2(f(w), f(u)) && \text{using (2),} \\
&= cd_2(f(y), f(z)) && \text{using (1) and } d_1(y, z) = d_1(w, u).
\end{aligned}$$

Thus, f is an isometry between (E_1, d_1) and (E_2, cd_2). □

3 Construction of the metric

Now that we know that we can construct - up to a constant factor - a geodesic metric d given D_d, how do we build it?

To give an answer, the problem needs to be properly posed.

Let $E_n = \{x_1, ..., x_n\}$ be a subset of a geodesic *compact* space (E, d) of diameter 1. Suppose that $(E_n)_{n \geq 1}$ converges to E in Hausdorff metric in (E, d). Can we build a metric d_n on E_n so that (E_n, d_n) converges to (E, d) in Gromov-Hausdorff distance, with d_n a function of D_d?

3.1 Main results

The idea of the proof of theorem 2 can be used to construct a consistent pseudo-metric on E_n.

Definition 3 (Pseudo metrics on E_n)**.** *Let (E,d) be a complete compact geodesic space, with diameter 1. Set $E_n = \{x_1, ..., x_n\} \subset E$. For $a, b \in E$, define - if it exists*

$$\begin{aligned}
M_{ab} &= \{z \in E; \max(d(a,z), d(b,z)) \leq d(a,b)\},\\
M_{ab}^n &= M_{ab} \cap E_n \setminus \{a,b\},\\
m_{ab} &\in \arg\min\{\max(d(a,z), d(b,z)); z \in M_{ab}\},\\
m_{ab}^n &\in \arg\min\{\max(d(a,z), d(b,z)); z \in M_{ab}^n\},
\end{aligned}$$

and set $A_0^n = (x,y)$, where $d(x,y) = \mathrm{diam}(E_n)$ and then for $p \geq 1$ and $A_p^n = (a_1^n, ..., a_k^n)$ - if all $m_{a_i^n a_j^n}^n$ exist,

$$A_{p+1}^n = (a_1^n, m_{a_1^n a_2^n}^n, a_2^n, m_{a_2^n a_3^n}^n, a_3^n, ..., m_{a_{k-1}^n a_k^n}^n, a_k^n).$$

Then, for the largest p such that A_p^n exists, define c_n on $A_p^n \times A_p^n$ by

$$c_n(a_i^n, a_j^n) = |i-j|2^{-p}.$$

and for any $p \geq 1$ such that A_p^n exists and for any $u, v \in E$, set

$$\begin{aligned}
d_{n,p}^+(u,v) &= \min\{c_n(a,b); d(a,b) \geq d(u,v), a, b \in A_p^n\}\\
d_{n,p}^-(u,v) &= \max\{c_n(a,b); d(a,b) \leq d(u,v), a, b \in A_p^n\}.
\end{aligned}$$

Finally, set $p_n = \max\{p \in \mathbb{N}^*; A_p^n \text{ exists}, \forall a, b \in A_p^n, d_{n,p}^+(a,b) = d_{n,p}^-(a,b)\}$.

Remark 4. *Given x, y in a geodesic space, the set of m_{xy} coincides with the set of middle points of (x,y).*

Intuitively, the largest A_p is longest geodesic path we can "make" from E_n, with each point being a middle point of its neighbors on A_p, and both $d_{n,p}^+$ and $d_{n,p}^-$ define a "metric" by comparing distances with the ones on this longest "segment" A_p. Then p is chosen so that $d_{n,p}^+$ and $d_{n,p}^-$ are "precise" (with a high p) and close enough.

Theorem 5. *Let (E,d) be a complete compact geodesic space, with diameter 1. Set $E_n = \{x_1, ..., x_n\} \subset E$.*

Then, for $C_0 = \frac{48}{\log 2}$,

$$\begin{aligned}
\sup_{u,v \in E_n} |d(u,v) - d_{n,p_n}^+(u,v)| &\leq C_0 d_H(E_n, E))(1 - \log d_H(E_n, E))\\
\sup_{u,v \in E_n} |d(u,v) - d_{n,p_n}^-(u,v)| &\leq C_0 d_H(E_n, E))(1 - \log d_H(E_n, E))
\end{aligned}$$

Corollary 6. *Let (E,d) be a complete compact geodesic space, with diameter 1, and E_n be a finite subset of (E,d). Then, one can construct a metric d_n on E_n, depending only on D_d such that*

$$d_{GH}((E_n, d_n), (E,d)) \leq 2C_0 d_H(E, E_n)(1 - \log(d_H(E, E_n))),$$

where $C_0 = \frac{48}{\log 2}$.

Remark 7. *This result implies that if E_n converges to E in Hausdorff metric, then the constructed (E_n, d_n) also converge to (E, d) in Gromov-Hausdorff metric. The hypotheses $\#E_n = n$ and $E_n \to E$ in Hausdorff metric implies that E is precompact. Since it is also closed, E is compact. To relax that hypothesis, one can assume that $E_n \cap B \to E \cap B$ for any closed ball B. In that case, the result states pointed Gromov-Hausdorff convergence of (E_n, d_n) to (E, d). Although, since the construction of d_n uses the fact that the diameter of (E, d) is 1, its construction have to be slightly adjusted.*

4 Ordinal multidimensional scaling (MDS)

Explain what is MDS, show where it is used, and show how this proves a result and in what conditions

Cite Ery: Some theory for ordinal embedding Cite Terada von Luxburg: Local ordinal embedding

5 Applications to statistics

Consider now that the points $E_n = \{X_1, ..., X_n\}$ of E are chosen randomly, in a i.i.d. setting. Then, if the law of X_i are smooth enough, the set E_n will converge to E in Hausdorff metric. The following proposition gives a more precise statement.

Proposition 8. *Let (E, d) be a geodesic space of diameter 1, such that*

$$\mathcal{N}(E, t) \leq \frac{C}{t^d}$$

for all $t > 0$, where $\mathcal{N}(E, t)$ denotes the minimal number of balls of radius t to cover E, C is a positive constant, and d an integer. Set μ a Borel probability measure on (E, d) such that

$$\mu(B_t) \geq \frac{c}{\mathcal{N}(E, t)}$$

for some $c > 0$ and any B_t, ball of radius $t > 0$. Set $n \in \mathbb{N}$ and let $E_n = \{X_1, ..., X_n\}$ be the set of i.i.d. random variables with common law μ. Then, there exists a constant K depending only on c and C such that,

$$\mathbb{E} d_H(E_n, E) \leq K \left(\frac{\log n}{n}\right)^{1/d}.$$

Given this random set E_n, and metric-comparison function D on this set, our theorem 5 allows us to build a metric d_n on E_n, that converges to (E, d) at a speed we can control in expectation.

Corollary 9. *Let (E, d) be a geodesic space of diameter 1, such that*

$$\mathcal{N}(E, t) \leq \frac{C}{t^d}$$

for all $t > 0$, where $\mathcal{N}(E, t)$ denotes the minimal number of balls of radius t to cover E, C is a positive constant, and d an integer. Set μ a Borel probability measure on (E, d) such that

$$\mu(B_t) \geq \frac{c}{\mathcal{N}(E, t)}$$

for some $c > 0$ and any B_t, ball of radius $t > 0$. Set $n \in \mathbb{N}$ and let $E_n = \{X_1, ..., X_n\}$ be the set of i.i.d. random variables with common law μ.

Then, one can construct a metric d_n on E_n only based on the function

$$D_d : (w, x, y, z) \in E_n^4 \mapsto \mathbf{1}_{d(w,x) \leq d(y,z)}$$

such that there exists a constant $K > 0$

$$\mathbb{E} d_{GH}(E_n, E) \leq K \left(\frac{\log n}{n} \right)^{1/d} \log n.$$

6 Simulations

In this section, we provide results of simulations of recovering of the metric. We show estimates of the $\sup d - d^+$ and of the average of $|d - d^+$. Using MDS, we also provide visual proof of efficiency of the algorithm recovering the metric from ordinal information

7 Proofs

7.1 Main theorem

The proof of theorem 5 is based on the following lemmas.

Lemma 10. *In the setting of theorem 5, denote d_H the Hausdorff metric, then,*

$$\forall n \geq 1, \forall p \geq 1, \forall a, b \in A_p^n, |d(a, b) - c_n(a, b)| \leq 6 p d_H(E_n, E).$$

Lemma 11. *In the setting of theorem 5,*

$$p_n \geq \left\lfloor \frac{1}{\log 2} \left(- \log(C_0 d_H(E_n, E)) | - \log \left(\log(e / d_H(E_n, E)) \right) \right) \right\rfloor. \tag{3}$$

Lemma 12. *In the settings of theorem 5, for $p \leq p_n$ and any $u, v \in \cup_{n \geq 1} E_n$,*

1. *$d^+_{n,p}(u, v) \leq d^-_{n,p}(u, v) + 2^{-p}$,*
2. *$d^-_{n,p}(u, v) \leq d^+_{n,p}(u, v)$.*

Proof of lemma 10. Set $\varepsilon = d_H(E_n, E)$.

First step

Remark 4 states that since E is geodesic, for all $a, b \in E$,

$$d(a, m_{ab}) \vee d(m_{ab}, b) = \frac{d(a, b)}{2}.$$

Also, by definition of the Hausdorff metric, for all $n \geq 1$, there exists $m_n \in E_n$ such that $d(m_n, m_{ab}) \leq \varepsilon$, so that

$$d(a, m_n) \vee d(m_n, b) \leq \frac{d(a, b)}{2} + \varepsilon.$$

Taking $a, b \in A_p^n$, it shows that

$$d(a, m_{ab}^n) \vee d(m_{ab}^n, b) \leq \frac{d(a,b)}{2} + \varepsilon.$$

Using, $d(a,b) \leq d(a, m_{ab}^n) \vee d(m_{ab}^n, b) + d(a, m_{ab}^n) \wedge d(m_{ab}^n, b)$, one can show that

$$d(a, m_{ab}^n) \wedge d(m_{ab}^n, b) \geq \frac{d(a,b)}{2} - \varepsilon.$$

Thus, for all $a, b \in A_p^n$,

$$|d(a, m_{ab}^n) - \frac{d(a,b)}{2}| \leq \varepsilon. \tag{4}$$

Second step

We want to show recursively on p that for all $p \geq 0$, setting $A_p^n = (a_1, ..., a_{1+2^p})$, for all $1 \leq i \leq 2^p$,

$$|d(a_i, a_{i+1}) - 2^{-p}| \leq (3 - 2^{-p})\varepsilon.$$

Triangular inequality and the fact that the diameter of E is 1 show that it is true for $p = 0$. Suppose it holds true for all $0 \leq p \leq q$. Then, set $A_{q+1}^n = (b_1, ..., b_{1+2^{q+1}})$. Thus, for any odd i (and similarly for i even), $b_{i+1} = m_{b_i b_{i+2}}^n$, so that, using (4) and the recurrence assumption,

$$\begin{aligned} d(b_i, b_{i+1}) &\leq \frac{d(b_i, b_{i+2})}{2} + \varepsilon \\ &\leq 2^{-(q+1)} + (3/2 - 2^{-(q+1)})\varepsilon + \varepsilon \\ &\leq 2^{-(q+1)} + (3 - 2^{-(q+1)})\varepsilon \end{aligned}$$

Similarly, $d(b_i, b_{i+1}) \geq 2^{-(q+1)} + (3 - 2^{-(q+1)})\varepsilon$.

So that, for all $1 \leq i \leq 2^p$,

$$|d(a_i, a_{i+1}) - 2^{-p}| \leq 3\varepsilon. \tag{5}$$

Third step

Inequality (5) proves the lemma for $p = 1$. Suppose it is true for all $1 \leq p \leq k$. Then, take $a, b \in A_{k+1}^n = (a_1, ..., a_{1+2^{k+1}})$.

- If $a, b \in A_k^n$, then it is already supposed to be true.
- If $a = a_i \in A_k^n$ and $b = a_j \notin A_k^n$, with $i < j$, then $a_{j-1}, a_{j+1} \in A_k^n$, so that

$$\begin{aligned} d(a,b) - c_n(a,b) &\leq d(a_i, a_{j-1}) - c_n(a_i, a_{j-1}) + d(a_{j-1}, a_j) - c_n(a_{j-1}, a_j) \\ &\leq 6k\varepsilon + 3\varepsilon \\ c_n(a,b) - d(a,b) &\leq c_n(a_i, a_{j+1}) - d(a_i, a_{j+1}) - c_n(a_{j+1}, a_j) + d(a_j, a_{j+1}) \\ &\leq 6k\varepsilon + 3\varepsilon \end{aligned}$$

- If $a, b \notin A_k^n$, the same ideas lead to

$$|d(a,b) - c_n(a,b)| \leq 6(k+1)\varepsilon,$$

which concludes the proof. ☐

Proof of lemma 11. First remark that if A_p^n exists and

$$\forall a, b \in A_p^n, |d(a,b) - c_n(a,b)| < 2^{-(p+1)}$$

then

$$\forall a, b \in A_p^n, d_{n,p}^+(a,b) = d_{n,p}^-(a,b).$$

Using lemma 10 and the fact that for any $a, b \in E_n$ such that $d(a,b) \geq 2^{-p}$, the set M_{ab}^n is not empty if $d_H(E_n, E) < 2^{-(p+1)}$ (as it contains the closest point of E_n to m_{ab}), one can show recursively on p that A_p^n exists for any n, p such that $6pd_H(E_n, E) < 2^{-(p+1)}$. Thus, lemma 10 and the remark above imply that if $6pd_H(E_n, E) < 2^{-(p+1)}$, then, $p_n \geq p$. Consequently, using lemma 13 (with $u = d_H(E_n, E), x = p \log 2, c = \frac{12}{\log 2}$), for $C_0 = \frac{12}{\log 2}$,

$$p_n \geq \left\lfloor \frac{1}{\log 2} \left(-\log(C_0 d_H(E_n, E)) - \log\left(\log(e/d_H(E_n, E))\right)\right) \right\rfloor.$$

☐

Proof of lemma 12. Set $n \in \mathbb{N}^*$ and $p \leq p_n$ and denote $(a_1, ..., a_{2^p+1}) = A_p^n$.

1. Take any $a_i, a_j \in A_p^n$ such that $d_{n,p}^+ = c_n(a_i, a_j)$ and

   $$d(a_i, a_j) \geq d(u, v).$$

   Then, by definition of $d_{n,p}^+(u,v)$ (as a minimum),

   $$d(a_i, a_{j-1}) < d(u, v)$$

   so that

   $$d_{n,p}^-(u,v) \geq c_n(a_i, a_{j-1}) = d_{n,p}^+(u,v) - 2^{-p}.$$

2. First, remark that since A_p^n increases with p, $d_{n,p}^-(u,v)$ increases with p and $d_{n,p}^+(u,v)$ decreases with p, so that is suffices to show $d_{n,p_n}^-(u,v) \leq d_{n,p_n}^+(u,v)$. In order to show a contradiction, suppose that there exists $u, v \in \cup_{n\geq 1} E_n$ such that $d_{n,p_n}^-(u,v) > d_{n,p_n}^+(u,v)$. Then, there exists, $a_{i_+}, a_{j_+}, a_{i_-}, a_{j_-} \in A_{p_n}^n$ such that

   $$\begin{aligned} c_n(a_{i_-}, a_{j_-}) &= d_{n,p_n}^-(u,v), \\ d(a_{i_-}, a_{j_-}) &\leq d(u,v), \\ c_n(a_{i_+}, a_{j_+}) &= d_{n,p_n}^+(u,v), \\ d(a_{i_+}, a_{j_+}) &\geq d(u,v), \end{aligned}$$

   with

   $$c_n(a_{i_-}, a_{j_-}) > c_n(a_{i_+}, a_{j_+}) \tag{6}$$
   $$d(a_{i_-}, a_{j_-}) \leq d(a_{i_+}, a_{j_+}). \tag{7}$$

   Thus, (7) gives $d_{n,p_n}^+(a_{i_-}, a_{j_-}) \leq d_{n,p_n}^+(a_{i_+}, a_{j_+})$.

So, using definitions of $d^+_{n,p}$ and $d^-_{n,p}$ (as maximum and minimum), and definition of p_n,

$$c_n(a_{i_-}, a_{j_-}) \leq d^-_{n,p_n}(a_{i_-}, a_{j_-}) = d^+_{n,p_n}(a_{i_-}, a_{j_-}) \leq d^+_{n,p_n}(a_{i_+}, a_{j_+}) \leq c_n(a_{i_+}, a_{j_+}).$$

This contradicts (6), proving that hypothesis $d^-_{n,p_n}(u,v) > d^+_{n,p_n}(u,v)$ was wrong.

□

Proof of theorem 5. Set $n \in \mathbb{N}^*$ and $p \leq p_n$. Let $u, v \in E_n$. Using lemma 10,

$$\begin{aligned} d^+_{n,p}(u,v) &= \min\{c_n(a,b); d(a,b) \geq d(u,v), a, b \in A^n_p\} \\ &\geq \min\{c_n(a,b); c_n(a,b) + 6pd_H(E_n, E) \geq d(u,v), a, b \in A^n_p\} \\ &\geq d(u,v) - 6pd_H(E_n, E). \end{aligned}$$

Similarly,

$$d^-_{n,p}(u,v) \leq d(u,v) + 6pd_H(E_n, E).$$

Thus, lemma 12 implies

$$\begin{aligned} d(u,v) - 6pd_H(E_n,E) - 2^{-p} &\leq d^+_{n,p}(u,v) - 2^{-p} \\ &\leq d^-_{n,p}(u,v) \\ &\leq d^-_{n,p_n}(u,v) \\ &\leq d^+_{n,p_n}(u,v) \\ &\leq d^+_{n,p}(u,v) \\ &\leq d^-_{n,p}(u,v) + 2^{-p} \leq d(u,v) + 6pd_H(E_n,E) + 2^{-p}. \end{aligned}$$

Taking $p = \left\lfloor \frac{1}{\log 2} \left(-\log(C_0 d_H(E_n,E)) - \log\left(\log(e/d_H(E_n,E))\right)\right) \right\rfloor \leq p_n$ as in (3), lemma 13 implies that $6pd_H(E_n,E) \leq 2^{-p-1}$, so that

$$\begin{aligned} \sup_{u,v \in E_n} |d(u,v) - d^+_{n,p_n}(u,v)| &\leq 2^{-p+1} \leq 4C_0 d_H(E_n,E))(1 - \log d_H(E_n,E)) \\ \sup_{u,v \in E_n} |d(u,v) - d^-_{n,p_n}(u,v)| &\leq 2^{-p+1} \leq 4C_0 d_H(E_n,E))(1 - \log d_H(E_n,E)) \end{aligned}$$

□

Lemma 13. *Set $u \in (0,1]$, $x \in \mathbb{R}$, and $c \geq 1$, such that*

$$x \leq \log\left(\frac{1}{cu}\right) - \log\left(1 - \log(u)\right),$$

then,

$$cxu \leq e^{-x}.$$

7.2 Corollary

Proof of corollary 6. It suffices to choose the closest metric d_n to d^+_{n,p_n} in the sup sense:

$$d_n \in \arg\min \left\{ \sup_{u,v\in E_n} |d(u,v) - d^+_{n,p_n}(u,v)|; d \text{ is a metric on } E_n \right\}.$$

Then, since $E_n \subset E$, there exists a surjective map $f : E \mapsto E_n$ such that

$$d_H((E_n,d),(E,d)) = \sup_{u,v\in E} |d(u,v) - d(f(u),f(v))|$$

so that

$$\begin{aligned} d_{GH}((E_n,d_n),(E,d)) &\le d_{GH}((E_n,d_n),(E_n,d)) + d_H(E_n,E) \\ &\le \sup_{u,v\in E_n} |d(u,v) - d_n(u,v)| + d_H(E_n,E) \\ &\le 2 \sup_{u,v\in E_n} |d(u,v) - d^+_{n,p_n}(u,v)| + d_H(E_n,E) \\ &\le C d_H(E_n,E))(1 - \log d_H(E_n,E)) \end{aligned}$$

The argmin does not actually necessarily exists, but any metric close enough satisfies it too. □

7.3 Proposition

Proof. For $t > 0$, denotes $\mathcal{N}(E,t)$ by m_t. Given balls $(B_i)_{1\le i\le m_{t/2}}$ that cover E,

$$\begin{aligned} \mathbb{E} d_H(E_n,E) &\le \mathbb{E} d_H(E_n,E)\mathbf{1}_{\{d_H(E_n,E)>t\}} + \mathbb{E} d_H(E_n,E)\mathbf{1}_{\{d_H(E_n,E)\le t\}} \\ &\le \mathbb{P}(d_H(E_n,E) > t) + t \\ &\le \mathbb{P}\left(\bigcup_{1\le i\le m_{t/2}} \bigcap_{1\le k\le n} \{X_k \notin B_i\} \right) + t \\ &\le \sum_{1\le i\le m_{t/2}} \prod_{1\le k\le n} e^{\log(1-\mu(B_k))} + t \\ &\le m_{t/2} e^{-\frac{cn}{m_t}} + t \\ &\le \frac{2^d C}{t^d} e^{-cnt^d/C} + t. \end{aligned}$$

Choosing $t = \left(\frac{C(1+1/d)}{c} \frac{\log(n)}{n} \right)^{1/d}$ leads to

$$\begin{aligned} \mathbb{E} d_H(E_n,E) &\le \frac{2^d cn}{(1+1/d)\log n} e^{-(1+1/d)\log n} + \left(\frac{C(1+1/d)}{c} \frac{\log(n)}{n} \right)^{1/d} \\ &\le K \left(\frac{\log n}{n} \right)^{1/d} \end{aligned}$$

□

8 Conclusion

We have shown that ordinal information on the metric of a geodesic space (E, d) is enough to recover the full metric. Also, given a *sample* E_n of the geodesic space E, and the ordinal information on that sample, a metric d_n can be built in such a way that the sample (E_n, d_n) equipped with this metric is as close, in Gromov-Hausdorff metric, to the geodesic space (E, d) as the sample (E_n, d) equipped with the true metric, up to a *logarithmic* factor.

This allows to quantify the information of the full ordinal information on the metric has compared to the metric itself. It is enough to recover the metric sharply (i.e. up to a log factor). An interesting question is whether a weaker ordinal information would be as efficient. For instance, knowing only D on quadruple (w, x, y, z) of the form (x, y, x, z) would be useful. It has already been solved on [KvL14] on $\mathbb{R}^d$ that this weaker notion of ordinal information is enough to recover the metric, but rates of convergence or sharp bounds are still unknown.

not used :::

Lemma 14. *Let X, Y be two subspaces of a normed vector space F, such that Y is convex. Denote $c(X \cup Y)$ the convex hull of $X \cup Y$. Then*

$$d_H(X, c(X \cup Y)) \leq 2d_H(X, Y).$$

Proof of lemma 14. For $t > 0$ and a set $A \subset F$, denote $A^t = \{a \in F; d(a, A) < t\}$. Set $t > d_H(X, Y)$. First, obviously, $X \subset c(X \cup Y)^{2t}$. Then, since $X \subset Y^t$, and Y^t is convex, then,

$$c(X \cup Y) \subset Y^t \subset \left(X^t\right)^t \subset X^{2t},$$

which proves the lemma. □

Lemma 15. *Let E_n be a metric space, and E be geodesic space. Then, for any $\varepsilon > 0$, there exists a geodesic space F such that there exists a subspace $F_n \subset F$ isometric to E_n and*

$$d_H(F_n, F) \leq 2d_{GH}(E_n, E) + \varepsilon.$$

Proof of lemma 15. It is known (see remark 7.3.12 of bbi) that for any $\varepsilon > 0$, there exists a metric d_ε on the disjoint union $\mathcal{X} = E_n \bigcup E$ such that $d_H(E_n, E) \leq d_{GH}(E_n, E) + \varepsilon$ with E_n, E considered as subspaces of $(\mathcal{X}, d_\varepsilon)$.

Denote $\mathcal{C}(\mathcal{X})$ the vector space of continuous functions on $\mathcal{X}$, equipped with the $\mathcal{L}^\infty$ norm. Define $\varphi : \mathcal{X} \to \mathcal{C}(\mathcal{X})$ by

$$\varphi : x \mapsto d_\varepsilon(x, .).$$

Then, φ is an isometry between $\mathcal{X}$ and $\varphi(\mathcal{X})$.

Set $F_n = \varphi(E_n)$ and $F = c(\varphi(\mathcal{X}))$, the closure of the convex hull of $\varphi(\mathcal{X})$. Since $\varphi(E)$ is convex (as a geodesic space in a vector space), and $\varphi(\mathcal{X}) = \varphi(E_n) \bigcup \varphi(E)$, using lemma 14,

$$d_H(F_n, F) \leq 2d_H(\varphi(E_n), \varphi(E)) = 2d_H(E_n, E) \leq 2d_{GH}(E_n, E) + \varepsilon.$$

□

:::

References

[KvL14] Matthäus Kleindessned and Ulrike von Luxburg. Uniqueness of ordinal embedding. Conference of Machine Learning (COLT), 2014.