Chapter 1

Hausdorff and Gromov–Hausdorff distances

1.1 Introduction

We begin by discussing general results related to the Gromov–Hausdorff distance. You can learn more about this theory in classical monographs [1, 2, 3]. In this theory, distance is not always a metric, since it can be infinite and also equal to zero between different points. We will stick to the following terminology.

Definition 1.1. ***Distance*** on the set X is every mapping

$$\rho\colon X \times X \to [0, \infty]$$

such that:

(i) $\rho(x, x) = 0$ for all $x \in X$;

(ii) $\rho(x, y) = \rho(y, x)$ for all $x, y \in X$ (symmetry).

If additionally the triangle inequality holds

$$\rho(x, z) \le \rho(x, y) + \rho(y, z) \quad \forall x, y, z \in X,$$

then ρ is called a ***generalized pseudometric***. If $\rho(x, y) > 0$ for $x \neq y$ (positive definiteness), then ρ is called a ***generalized metric***. If $\rho(x, y) < \infty$ for all $x, y \in X$, then the word "generalized" is omitted.

The set X together with the function ρ is called the space of the corresponding type.

1.2 Hausdorff distance

Let X be an arbitrary metric space. The distance between points $x, y \in X$ will be denoted by $|xy|$. For $x \in X$, $r > 0$ and $s \ge 0$ we define:

$$U_r(x) = \{y \in X : |xy| < r\}, \qquad B_s(x) = \{y \in X : |xy| \le s\}.$$

For non-empty $A, B \subset X$ we set

$$|xA| = \inf_{a \in A} |xa|, \qquad |AB| = \inf_{\substack{a \in A \\ b \in B}} |ab|.$$

Let us define open and closed neighborhoods of the set A:

$$U_r(A) = \{x \in X : |xA| < r\}, \qquad B_s(A) = \{x \in X : |xA| \le s\}.$$

Definition 1.2. For non-empty subsets A and B of the metric space X, the ***Hausdorff distance*** between A and B is the quantity

$$d_H(A,B) = \max\Big\{\sup_{a\in A}|aB|,\ \sup_{b\in B}|Ab|\Big\}.$$

Proposal 1.3. *The following equivalent formulas hold*:

$$d_H(A,B) = \inf\{r>0 : A\subset U_r(B),\ B\subset U_r(A)\} = \inf\{s\geq 0 : A\subset B_s(B),\ B\subset B_s(A)\}.$$

Proof. Let us denote the first expression by α, the second by β, and the original definition by γ. Let us show that $\alpha=\beta=\gamma$.

If $A\subset U_r(B)$ and $B\subset U_r(A)$, then for any $a\in A$ there is $b\in B$ with $|ab|<r$, so $|aB|<r$, hence $\sup_{a\in A}|aB|\leq r$. Similar to $\sup_{b\in B}|Ab|\leq r$. Therefore $\gamma\leq r$. Passing to the infimum along r, we obtain $\gamma\leq\alpha$.

Conversely, if $\gamma<r$, then for any $a\in A$ we have $|aB|<r$ (a strict inequality, since the supremum is less than r), which means $a\in U_r(B)$. Similar to $B\subset U_r(A)$. Therefore $\alpha\leq r$. Passing to the infimum, we obtain $\alpha\leq\gamma$.

The equality $\alpha=\beta$ follows from the fact that for any $r>0$ $U_r(A)\subset B_r(A)$ holds and $U_r(A)\supset B_{r'}(A)$ for any $r'<r$. □

Proposal 1.4. *The Hausdorff distance d_H is a generalized pseudometric on the set of all nonempty subsets of the metric space X.*

Proof. Non-negativity and symmetry follow directly from the definition. For any non-empty $A\subset X$ we have $d_H(A,A)=0$.

Let's check the triangle inequality. Let $A,B,C\subset X$ be non-empty. Let us denote $c=d_H(A,B)$, $a=d_H(B,C)$, $b=d_H(A,C)$. For any $\varepsilon>0$ we set $r=c+\varepsilon$, $s=a+\varepsilon$. Then $A\subset U_r(B)$ and $B\subset U_s(C)$.

Lemma 1.5. *For any $r,s>0$ and non-empty $A\subset X$,*

$$U_r\big(U_s(A)\big)\subset U_{r+s}(A).$$

Proof. If $x\in U_r(U_s(A))$ then there is $y\in U_s(A)$ with $|xy|<r$, and there is $a\in A$ with $|ya|<s$. Then $|xa|\leq|xy|+|ya|<r+s$, which means $x\in U_{r+s}(A)$. □

From Lemma 1.5 we get:

$$A\subset U_r(B)\subset U_r\big(U_s(C)\big)\subset U_{r+s}(C).$$

Similar to $C\subset U_{r+s}(A)$. Therefore,

$$b\leq r+s=c+a+2\varepsilon.$$

Since ε is arbitrary, we obtain $b\leq c+a$. □

Proposal 1.6. *Let X be a metric space. Then the Hausdorff distance d_H, restricted to the set $\mathcal{C}(X)$ of all non-empty closed subsets of X, is a generalized metric.*

Proof. It is enough to check positive definiteness. Let $A,B\in\mathcal{C}(X)$ be different. Then $x\in A\setminus B$ exists (without loss of generality). Since B is closed, then for some $\varepsilon>0$ $U_\varepsilon(x)\cap B=\varnothing$ holds, whence $|xB|\geq\varepsilon$. Therefore, $U_{\varepsilon/2}(B)\not\supset A$, so $d_H(A,B)\geq\varepsilon/2$. □

Consequence 1.7. *Let A and B be non-empty subsets of the metric space X. Then $d_H(A,B)=0$ if and only if the closures $\overline{A}$ and $\overline{B}$ coincide.*

Proof. For any $r > 0$, $U_r(A) \supset \overline{A}$ and $A \subset \overline{A} \subset U_r(\overline{A})$ hold, so $d_H(A, \overline{A}) = 0$. If $d_H(A, B) = 0$, then from the triangle inequality we obtain $d_H(\overline{A}, \overline{B}) = 0$. Since $\overline{A}, \overline{B} \in \mathcal{C}(X)$, and the Hausdorff distance on $\mathcal{C}(X)$ is positive definite, then $\overline{A} = \overline{B}$. The opposite is obvious. □

By $\mathcal{H}(X)$ we denote the set of all closed bounded subsets of a metric space X. Proposition 1.6 implies that d_H is a metric on $\mathcal{H}(X)$. This metrix space is often referred as ***Hausdorff space***. Transition from a metric space X to metric space $\mathcal{H}(X)$ preserves the following natural properties (a proof can be found in [1], Chapter 1, §1.6).

Theorem 1.8. *Metric spaces X and $\mathcal{H}(X)$ simultaneously have or simultaneously do not have the next properties*:

(1) *completeness (Khan Theorem),*

(2) *total boundedness,*

(3) *compactness (Hausdorff and Blaschke Theorem).*

1.3 Gromov–Hausdorff distance

In this section we give a definition of the central object of our course — the Gromov–Hausdorff distance. This distance measures how "similar" two metric spaces are to each other, allowing spaces to be compared even if they are defined on different sets.

We need the following notation. If X and Y are metric spaces, then the notation $X \approx Y$ means that X and Y **isometric**, i.e. there is a bijective mapping $f : X \to Y$ that preserves distances:

$$|f(x)f(x')|_Y = |xx'|_X \quad \forall x, x' \in X.$$

Definition 1.9. The Gromov–Hausdorff distance between non-empty metric spaces X and Y is the quantity

$$d_{GH}(X, Y) := \inf\Big\{ d_H(X', Y') \;\Big|\; \exists \text{ metric space } Z \text{ and subsets } X', Y' \subset Z \text{ such that } X' \approx X,\ Y' \approx Y \Big\}.$$

In other words, we consider all possible isometric copies of X and Y embedded in an arbitrary metric space Z and measure the Hausdorff distance between these copies. The infimum of these distances is $d_{GH}(X, Y)$.

Note 1.10. This definition is consistent with the theorem 1.16, which will be proven below and gives an equivalent expression in terms of so-called ***correspondences***:

$$d_{GH}(X, Y) = \frac{1}{2} \inf\big\{ \operatorname{dis} R : R \in \mathcal{R}(X, Y) \big\},$$

where $\mathcal{R}(X, Y)$ stands for the set of all possible correspondences between X and Y. Thus, the Gromov–Hausdorff distance can be calculated both through geometric embeddings and through combinatorial structures — correspondences, see next Section.

Example 1.11. If X and Y are isometric, then $d_{GH}(X, Y) = 0$. Indeed, it suffices to take $Z = X$, $X' = X$, $Y' = Y$ (identified with X via isometry), then $d_H(X', Y') = 0$.

1.4 Correspondences

For an arbitrary set X, let $\mathcal{P}_0(X)$ denote the set of all non-empty subsets of X. Recall that a ***binary relation*** or just a ***relation*** for sets X and Y is defined as a subset of $X \times Y$.

Definition 1.12. For any non-empty metric spaces X, Y and (non-empty) binary relation $\sigma \in \mathcal{P}_0(X \times Y)$ its ***distortion*** $\operatorname{dis}\sigma$ is the quantity

$$\operatorname{dis}\sigma = \sup\Big\{\big||xx'| - |yy'|\big| : (x,y),(x',y') \in \sigma\Big\}.$$

Proposal 1.13. *If $\sigma, \tau \in \mathcal{P}_0(X \times Y)$ and $\tau \subset \sigma$, then $\operatorname{dis}\tau \le \operatorname{dis}\sigma$.*

Proof. This follows directly from the definition of distortion: the supremum over a subset does not exceed the supremum over the entire set. □

Recall that for relations $\sigma \in \mathcal{P}_0(X \times Y)$, $\tau \in \mathcal{P}_0(Y \times Z)$ their ***composition*** $\tau \circ \sigma$ is defined as follows: $(x,z) \in \tau \circ \sigma$ iff there exists an $y \in Y$ such that $(x,y) \in \sigma$ and $(y,z) \in \tau$.

Proposal 1.14. *If $\sigma \in \mathcal{P}_0(X \times Y)$, $\tau \in \mathcal{P}_0(Y \times Z)$ and $\tau \circ \sigma \neq \varnothing$, then*

$$\operatorname{dis}(\tau \circ \sigma) \le \operatorname{dis}\sigma + \operatorname{dis}\tau.$$

Proof. For any $(x,z),(x',z') \in \tau \circ \sigma$ there exist $y, y' \in Y$ such that $(x,y),(x',y') \in \sigma$ and $(y,z),(y',z') \in \tau$. Then

$$\big||xx'| - |zz'|\big| \le \big||xx'| - |yy'|\big| + \big||yy'| - |zz'|\big|.$$

Moving on to the supremums, we get what we need. □

Recall that a ***multivalued mapping*** $f : X \rightrightarrows Y$ is a mapping $f : X \to \mathcal{P}_0(Y)$. A multivalued mapping f is ***surjective*** if $\bigcup_{x \in X} f(x) = Y$. The relation $R \subset X \times Y$, which is a surjective multivalued mapping, is called ***correspondence***. Let us denote the set of all correspondences between X and Y by $\mathcal{R}(X,Y)$.

Proposal 1.15. *Nonempty metric spaces X and Y are isometric if and only if there exists $R \in \mathcal{R}(X,Y)$ for which $\operatorname{dis} R = 0$.*

Proof. If X and Y are isometric, then R can be taken to be isometry. Conversely, if $R \in \mathcal{R}(X,Y)$ and $\operatorname{dis} R = 0$, then for any $(x,y),(x',y') \in R$ we have $|xx'| = |yy'|$, so R is a single-valued mapping preserving distances. Since R is surjective, it is an isometry. □

Theorem 1.16. *For any non-empty metric spaces X and Y,*

$$d_{GH}(X,Y) = \frac{1}{2}\inf\big\{\operatorname{dis} R : R \in \mathcal{R}(X,Y)\big\}.$$

Proof. Let $d := d_{GH}(X,Y) < \infty$. Then for any $\varepsilon > 0$ there exists a metric space Z and its subsets $X' \approx X$, $Y' \approx Y$ such that $d_H(X',Y') < d + \varepsilon$. Let's put

$$R = \big\{(x,y) \in X' \times Y' : |xy| < d + \varepsilon\big\}.$$

Since $d_H(X',Y') < d + \varepsilon$, then $R \in \mathcal{R}(X',Y')$. For any $(x,y),(x',y') \in R$ the triangle inequality gives

$$\big||xx'| - |yy'|\big| \le |xy| + |x'y'| < 2d + 2\varepsilon.$$

Therefore, $\operatorname{dis} R \le 2d + 2\varepsilon$. Identifying X with X' and Y with Y', we get

$$\frac{1}{2}\inf\bigl\{\operatorname{dis} R : R \in \mathcal{R}(X,Y)\bigr\} \le d.$$

Conversely, let $I := \inf\bigl\{\operatorname{dis} R : R \in \mathcal{R}(X,Y)\bigr\} < \infty$. For any $\varepsilon > 0$ there is $R \in \mathcal{R}(X,Y)$ such that $\operatorname{dis} R < I + \varepsilon$. If $r := \operatorname{dis} R = 0$, then X and Y are isometric, so $d_{GH}(X,Y) = 0$.

Let $r > 0$. Let us extend the metrics of the spaces X and Y to $Z := X \sqcup Y$, setting for $x \in X$, $y \in Y$:

$$|xy| = \inf\bigl\{|xx'| + \frac{r}{2} + |yy'| : (x',y') \in R\bigr\}.$$

Let us show that this is a metric. It is enough to check the triangle inequality for the points $x_1, x_2 \in X$, $y \in Y$.

For an arbitrary $\varepsilon > 0$ we choose $(x_i', y_i') \in R$ so that

$$|x_i y| \ge |x_i x_i'| + \frac{r}{2} + |y_i' y| - \varepsilon.$$

Since $\operatorname{dis} R = r$, then $|x_1'x_2'| \le |y_1'y_2'| + r$. Then

$$\begin{aligned}
|x_1x_2| &\le |x_1x_1'| + |x_1'x_2'| + |x_2'x_2| \le |x_1x_1'| + \bigl(|y_1'y_2'| + r\bigr) + |x_2'x_2| \\
&\le |x_1x_1'| + \bigl(|y_1'y| + |yy_2'| + r\bigr) + |x_2'x_2| = \Bigl(|x_1x_1'| + |y_1'y| + \frac{r}{2}\Bigr) + \Bigl(|x_2'x_2| + |yy_2'| + \frac{r}{2}\Bigr) \\
&\le (|x_1y| + \varepsilon) + (|x_2y| + \varepsilon) = |x_1y| + |x_2y| + 2\varepsilon.
\end{aligned}$$

Since ε is arbitrary, we obtain $|x_1x_2| \le |x_1y| + |x_2y|$.

Further, for the same $(x_2', y_2') \in R$ and ε,

$$|x_1y| \le |x_1x_2'| + \frac{r}{2} + |y_2'y| \le |x_1x_2| + \bigl(|x_2x_2'| + \frac{r}{2} + |y_2'y|\bigr) \le |x_1x_2| + |x_2y| + \varepsilon,$$

and since ε is arbitrary, we obtain $|x_1y| \le |x_1x_2| + |x_2y|$.

Note that for all $x \in X$, $y \in Y$, $|xy| = r/2$ holds if and only if $(x,y) \in R$. Since R is a correspondence, then $d_H(X,Y) = r/2$. Therefore, $d_{GH}(X,Y) \le I/2$. □

1.5 Elementary properties of the Gromov–Hausdorff distance

For a metric space X, let $\operatorname{diam} X$ denote its ***diameter***: $\operatorname{diam} X := \sup\bigl\{|xy| : x, y \in X\bigr\}$. For a single-point space, we reserve the notation Δ_1. If $\lambda > 0$, then by λX we denote the metric space obtained from X by multiplying all distances by λ. If the space X is bounded, then we set $0X = \Delta_1$.

Proposal 1.17. *For any $X, Y \in \mathcal{GH}$ the following relations hold.*

(1) $2d_{GH}(\Delta_1, X) = \operatorname{diam} X$.

(2) $2d_{GH}(X,Y) \le \max\{\operatorname{diam} X, \operatorname{diam} Y\}$.

(3) *If* $\operatorname{diam} X < \infty$ *or* $\operatorname{diam} Y < \infty$, *then*

$$|\operatorname{diam} X - \operatorname{diam} Y| \le 2d_{GH}(X,Y).$$

(4) *If* $\operatorname{diam} X < \infty$*, then for any* $\lambda, \mu \geq 0$*:*

$$2d_{GH}(\lambda X, \mu X) = |\lambda - \mu| \cdot \operatorname{diam} X.$$

In particular, the mapping $\gamma(t) := tX$ *is the shortest geodesic between any of its points.*

(5) *For any* $\lambda > 0$*:*

$$d_{GH}(\lambda X, \lambda Y) = \lambda d_{GH}(X, Y).$$

If X *and* Y *are bounded, then the equality holds for* $\lambda = 0$*.*

Proof. (1) The set $\mathcal{R}(\Delta_1, X)$ consists of exactly one correspondence $R := \Delta_1 \times X$. Its distortion is

$$\operatorname{dis} R = \sup_{x,y \in X} |xy| = \operatorname{diam} X.$$

By theorem 1.16 we obtain $d_{GH}(\Delta_1, X) = \frac{1}{2} \operatorname{diam} X$.

(2) Consider the correspondence $R = X \times Y \in \mathcal{R}(X, Y)$. By definition of distortion:

$$\operatorname{dis} R = \max\{\operatorname{diam} X, \operatorname{diam} Y\}.$$

Therefore, $d_{GH}(X, Y) \leq \frac{1}{2} \max\{\operatorname{diam} X, \operatorname{diam} Y\}$.

(3) Let $\operatorname{diam} X < \infty$. Consider an arbitrary correspondence $R \in \mathcal{R}(X, Y)$. For any $\varepsilon > 0$ there are $x, x' \in X$ such that $|xx'| > \operatorname{diam} X - \varepsilon$. Since R is a correspondence, there exist $y, y' \in Y$ with $(x, y), (x', y') \in R$. Then

$$\operatorname{dis} R \geq \big||xx'| - |yy'|\big| \geq |xx'| - \operatorname{diam} Y > \operatorname{diam} X - \operatorname{diam} Y - \varepsilon.$$

Since ε is arbitrary, we obtain $2d_{GH}(X, Y) \geq \operatorname{diam} X - \operatorname{diam} Y$. Swapping X and Y, we get $|\operatorname{diam} X - \operatorname{diam} Y| \leq 2d_{GH}(X, Y)$.

(4) Consider the identity mapping $\mathrm{id} : \lambda X \to \mu X$. Its distortion:

$$\operatorname{dis} \mathrm{id} = |\lambda - \mu| \cdot \operatorname{diam} X.$$

Therefore, $2d_{GH}(\lambda X, \mu X) \leq |\lambda - \mu| \cdot \operatorname{diam} X$. The reverse inequality follows from (3).

(5) For $\lambda > 0$ and a correspondence R construct correspondence R_λ between λX and λY, where R_λ is the same subset of $X \times Y$. Then $\operatorname{dis} R_\lambda = \lambda \operatorname{dis} R$. Moving on to the infimum, we get what we need. If X and Y are bounded, then $\lambda X = \lambda Y = \Delta_1$. □

1.6 Space $\mathcal{GH}$

Let $\mathcal{GH}$ be a proper class of all metric spaces considered up to isometry. The latter means that in each class of isometric spaces we choose one representative (this requires the global axiom of choice).

Theorem 1.18. *The function* d_{GH} *is a generalized pseudometric on* $\mathcal{GH}$*.*

Proof. Non-negativity and symmetry are obvious.

Triangle inequality. Let $X, Y, Z \in \mathcal{GH}$. By theorem 1.16 for any $\varepsilon > 0$ there are correspondences $R_{XY} \in \mathcal{R}(X, Y)$ and $R_{YZ} \in \mathcal{R}(Y, Z)$ such that $\operatorname{dis} R_{XY} < 2d_{GH}(X, Y) + \varepsilon$ and $\operatorname{dis} R_{YZ} < 2d_{GH}(Y, Z) + \varepsilon$.

Consider the composition $R_{XZ} := R_{YZ} \circ R_{XY} \in \mathcal{R}(X, Z)$. As suggested by 1.14:

$$\operatorname{dis} R_{XZ} \leq \operatorname{dis} R_{XY} + \operatorname{dis} R_{YZ} < 2d_{GH}(X, Y) + 2d_{GH}(Y, Z) + 2\varepsilon.$$

Applying the theorem 1.16 again, we get

$$2d_{GH}(X, Z) \leq \operatorname{dis} R_{XZ} < 2d_{GH}(X, Y) + 2d_{GH}(Y, Z) + 2\varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, then $d_{GH}(X, Z) \leq d_{GH}(X, Y) + d_{GH}(Y, Z)$. □

Note 1.19. In general, d_{GH} is not a metric on $\mathcal{GH}$: there are non-isometric spaces with zero GH-distance. For example, $d_{GH}\big([0,1],(0,1)\big) = 0$, but these spaces are not isometric. However, on the subclass of compact metric spaces d_{GH} becomes a real metric (see chapter 3).

Theorem 1.20. *The class $\mathcal{GH}$ is not separable. Moreover, it contains an $\frac{1}{2}$-separated proper class.*

Proof. For each cardinal number n, consider the metric simplex Δ_n — a set of cardinality n with metric

$$|xy| = \begin{cases} 0, & x = y, \\ 1, & x \neq y. \end{cases}$$

If $n \neq m$, then Δ_n and Δ_m are not isometric.

Let us show that for any $n \neq m$, $d_{GH}(\Delta_n, \Delta_m) = \frac{1}{2}$ holds. Indeed, let $R \in \mathcal{R}(\Delta_n, \Delta_m)$ be an arbitrary correspondence, then R is not a bijection graph, so $\operatorname{dis} R = 1$. By Theorem 1.16 we obtain $d_{GH}(\Delta_n, \Delta_m) = 1/2$.

Therefore, the family $\{\Delta_n\}$, where n runs over all cardinal numbers, is a $\frac{1}{2}$-separated proper class in $\mathcal{GH}$. □

Theorem 1.21. *The class $\mathcal{GH}$ is complete.*

Proof. Let $\{X_n\}_{n=1}^{\infty}$ be the fundamental sequence in $\mathcal{GH}$. Without loss of generality, we assume that

$$2d_{GH}(X_n, X_{n+1}) \leq \frac{1}{2^n} \quad \forall n \in \mathbb{N}.$$

This can be done by going to a subsequence.

By theorem 1.16, for each n there is a correspondence $R_n \in \mathcal{R}(X_n, X_{n+1})$ such that

$$\operatorname{dis} R_n < \frac{1}{2^n}. \tag{1}$$

Construction of the limit space. Consider the disjoint union

$$\widehat{X} := \bigsqcup_{n=1}^{\infty} X_n.$$

Threads. Let us call a **thread**, starting at the point $x_1 \in X_1$, a sequence $(x_1, x_2, \ldots) \in \prod_{n=1}^{\infty} X_n$ such that

$$(x_n, x_{n+1}) \in R_n \quad \forall n \geq 1.$$

We denote the set of all threads by $N(\widehat{X})$.

For two threads $\nu = (x_1, x_2, \ldots)$ and $\nu' = (x'_1, x'_2, \ldots)$ consider the sequence of numbers

$$|x_1x'_1|, |x_2x'_2|, \ldots$$

Let us show that this sequence is fundamental. From (1) it follows that for any n

$$\big||x_nx'_n| - |x_{n+1}x'_{n+1}|\big| < \frac{1}{2^n}.$$

Then for any $m \geq 1$:

$$\big||x_nx'_n| - |x_{n+m}x'_{n+m}|\big| \leq \sum_{k=1}^{m} \big||x_{n+k-1}x'_{n+k-1}| - |x_{n+k}x'_{n+k}|\big| < \sum_{k=1}^{m} \frac{1}{2^{n+k-1}} < \frac{1}{2^{n-1}}.$$

Consequently, the sequence is fundamental and has a limit, which we denote by

$$|\nu\nu'| := \lim_{n\to\infty} |x_n x'_n|.$$

Metric on threads. The function $(\nu, \nu') \mapsto |\nu\nu'|$ is a pseudometric on $N(\widehat{X})$. The triangle inequality for threads follows from the triangle inequality for the corresponding coordinates x_n, x'_n, x''_n and passage to the limit.

Let's define

$$X := N(\widehat{X})/\sim,$$

where $\nu \sim \nu' \iff |\nu\nu'| = 0$, with the metric induced by $|\cdot|$. This is a metric space.

Convergence check. For each n we define the relation $R'_n \subset X \times X_n$ as follows: for each class $[\nu] \in X$ and each thread $\nu = (x_1, x_2, \ldots) \in [\nu]$ we set

$$\big([\nu], x_n\big) \in R'_n.$$

Let's check that R'_n is a correspondence.

The projection of R'_n onto X is equal to the entire X: for any class $[\nu] \in X$ there is a thread $\nu \in [\nu]$, and then $\big([\nu], x_n\big) \in R'_n$ for the corresponding x_n.

Let's check that the projection of R'_n onto X_n is equal to the entire X_n. Let's take an arbitrary point $x_n \in X_n$. Since the correspondences R_k are surjective for all k, there is a thread $\nu = (x_1, x_2, \ldots)$ passing through x_n: for $k \geq n$ we choose $x_{k+1} \in R_k(x_k)$ (this is possible since R_k is a correspondence), and for $k < n$ we choose $x_{k-1} \in R_{k-1}^{-1}(x_k)$ (this is possible since R_{k-1}^{-1} is a correspondence). Then $\big([\nu], x_n\big) \in R'_n$, hence x_n lies in the projection of R'_n onto X_n.

Thus, R'_n is a correspondence between X and X_n.

Let's estimate the distortion of R'_n. For any $[\nu], [\nu'] \in X$ and corresponding threads $\nu = (x_1, x_2, \ldots)$, $\nu' = (x'_1, x'_2, \ldots)$ we have

$$\big|[\nu][\nu']\big| = \lim_{k\to\infty} |x_k x'_k|.$$

As shown above,

$$\Big||x_n x'_n| - \big|[\nu][\nu']\big|\Big| < \frac{1}{2^{n-1}}.$$

Therefore,

$$\operatorname{dis} R'_n \leq \frac{1}{2^{n-1}}.$$

Then, by theorem 1.16,

$$d_{GH}(X, X_n) \leq \frac{1}{2} \operatorname{dis} R'_n \leq \frac{1}{2^n}.$$

Thus, $X_n \to X$ in $\mathcal{GH}$. Since the original sequence $\{X_n\}$ is fundamental and the subsequence converges to X, then the entire sequence converges to X. Therefore, $\mathcal{GH}$ is complete. □

1.7 The Gromov–Hausdorff distance is intrinsic

In this Section we will show that the distance d_{GH} on the class $\mathcal{GH}$ is intrinsic.

Definition 1.22. Let d be a generalized pseudometric on the class $\mathcal{C}$. ***Curve*** in $(\mathcal{C}, d)$ is the mapping $\gamma : [a, b] \to \mathcal{C}$ (where $[a, b] \subset \mathbb{R}$ is a segment) for which the function

$$t \mapsto d\big(\gamma(t), \gamma(t_0)\big)$$

is continuous at every point $t_0 \in [a, b]$.

Definition 1.23. Let d be a generalized pseudometric on the class $\mathcal{C}$, and let $\gamma : [a, b] \to \mathcal{C}$ be a curve. ***The length*** of a curve γ is the quantity

$$L_d(\gamma) := \sup \left\{ \sum_{i=1}^{n} d\big(\gamma(t_{i-1}), \gamma(t_i)\big) \mid a = t_0 < t_1 < \cdots < t_n = b \right\}.$$

Definition 1.24. A generalized pseudometric d on the class $\mathcal{C}$ is called ***intrinsic*** if for any $X, Y \in \mathcal{C}$:

$$d(X, Y) = \inf\{L_d(\gamma) \mid \gamma \text{ — curve in } (\mathcal{C}, d),\ \gamma(a) = X,\ \gamma(b) = Y\}.$$

Theorem 1.25. *The Gromov–Hausdorff distance d_{GH} is an intrinsic generalized pseudometric on $\mathcal{GH}$.*

Proof. Let us fix $X, Y \in \mathcal{GH}$. If $d_{GH}(X, Y) = \infty$, then the inequality is trivial.

Step 1. Constructing the curve. Let $R \in \mathcal{R}(X, Y)$ be a correspondence with $\operatorname{dis} R < 2d_{GH}(X, Y) + 1$. For each $t \in [0, 1]$, consider the set R with the function

$$d_t\big((x, y), (x', y')\big) := (1 - t)|xx'|_X + t|yy'|_Y.$$

For $0 < t < 1$, the function d_t is a metric on R. For $t = 0$ we have $d_0\big((x, y), (x', y')\big) = |xx'|_X$, therefore d_0 is a pseudometric: points (x, y) and (x, y') with the same x, but different y, are at zero distance. Similarly for $t = 1$. Factoring R by the equivalence relation $(x, y) \sim_t (x', y') \iff d_t\big((x, y), (x', y')\big) = 0$, for $t = 0$ we obtain a space isometric to X (the equivalence class of the point (x, y) is identified with x), and for $t = 1$ we obtain a space isometric to Y (the equivalence class is identified with y). Let $\gamma(t) \in \mathcal{GH}$ denote the isometry class of the factor space $Z_t = (R, d_t)/\sim_t$.

For any $s, t \in [0, 1]$ we consider the identity correspondence $I_{s,t} \in \mathcal{R}(Z_s, Z_t)$. Its distortion: $\operatorname{dis} I_{s,t} = |t - s| \cdot \operatorname{dis} R$. According to the theorem 1.16,

$$d_{GH}(Z_s, Z_t) \le \frac{|t - s|}{2} \operatorname{dis} R.$$

This means that γ is a curve in $(\mathcal{GH}, d_{GH})$.

Step 2. Length estimation. For any partition $0 = t_0 < t_1 < \cdots < t_n = 1$:

$$\sum_{i=1}^{n} d_{GH}(\gamma(t_{i-1}), \gamma(t_i)) \le \frac{1}{2} \operatorname{dis} R.$$

Therefore, $L_{d_{GH}}(\gamma) \le \frac{1}{2} \operatorname{dis} R < d_{GH}(X, Y) + \frac{1}{2}$. Since R can be chosen with an arbitrarily small excess, $\inf L_{d_{GH}}(\gamma) \le d_{GH}(X, Y)$.

Step 3. Reverse inequality. For any curve γ with $\gamma(a) = X$, $\gamma(b) = Y$:

$$d_{GH}(X, Y) \le \sum_{i=1}^{n} d_{GH}(\gamma(t_{i-1}), \gamma(t_i)) \le L_{d_{GH}}(\gamma).$$

Therefore, $d_{GH}(X, Y) \le \inf L_{d_{GH}}(\gamma)$.

Combining, we get $d_{GH}(X, Y) = \inf L_{d_{GH}}(\gamma)$. □

1.8 Irreducible correspondences

In the previous section, we established that the Gromov–Hausdorff distance can be calculated as the infimum of distortions over all correspondences (theorem 1.16). However, the set of all correspondences between two spaces, especially in the case of infinite sets, is extremely large and difficult to analyze. The following construction allows us to significantly simplify the problem by showing that when calculating the GH-distance it is enough to consider only a special subclass of correspondences — the so-called ***irreducible*** correspondences.

1.8.1 Definition and motivation

Definition 1.26. Let X and Y be non-empty sets, and let $R \in \mathcal{R}(X, Y)$ be a correspondence. A correspondence R is said to be ***irreducible*** if for any correspondence $R' \in \mathcal{R}(X, Y)$ such that $R' \subset R$, $R' = R$ holds.

In other words, an irreducible correspondence is minimal in inclusion among all correspondences. This definition and the proposition 1.13 have a key corollary: for any correspondence R, its subcorrespondence R_0 has a distortion no greater than that of the original R. Thus, if we show that for any correspondence R there is an irreducible subcorrespondence $R_0 \subset R$, then the infimum of distortions over all correspondences will be achieved on the irreducible ones. This is the main reason why the concept of irreducibility is useful in calculating the Gromov–Hausdorff distance. Denote by $\mathcal{R}_0(X, Y)$ the subclass of irreduc

1.8.2 Existence of irreducible subcorrespondences

Theorem 1.27. *For any non-empty sets X, Y and any correspondence $R \in \mathcal{R}(X, Y)$ there is an irreducible correspondence $R_0 \in \mathcal{R}(X, Y)$ such that $R_0 \subset R$.*

Proof. We will look at each relation $\sigma \in \mathcal{P}(X, Y)$ as a multivalued mapping that can have a domain smaller than X.

Let R be an arbitrary correspondence between X and Y. For each $x \in X$, choose any one element (x, y) from R and define a mapping $f\colon X \to Y$ by setting $y = f(x)$. Having identified the mapping with its graph, we note that $f \subset R$. Let $Y_1 = f(X)$ and $Y_2 = Y \setminus Y_1$.

Now for each $y \in Y_2$ we choose any one element (x, y) from R and define the mapping $g\colon Y_2 \to X$ by putting $x = g(y)$. Once again, identifying the mapping with its graph, we note that $g^{-1} \subset R$. Let $X_2 = g(Y_2)$ and $X_1 = X \setminus X_2$.

Let $Y_3 = f(X_2)$. It is clear that $Y_3 \subset Y_1$.

Using f and g, we define another relation:

$$h = f \cup g^{-1}.$$

Lemma 1.28. *We have $h \in \mathcal{R}(X, Y)$.*

Proof. By the definition of f, for each $x \in X$, $\big(x, f(x)\big) \in f \subset h$ holds.

Let us now consider an arbitrary $y \in Y$. If $y \in Y_1$ then, since $Y_1 = \operatorname{im} f$, there is $x \in X$ such that $y = f(x)$, so $\big(x, y\big) \in f \subset h$. If $y \in Y_2$, then, by the definition of g, $\big(g(y), y\big) \in g^{-1} \subset h$ holds. □

Let us now define the relation R_0 by removing from h some pairs (x, y) for each $y \in Y_3$ according to the following rule:

(1) if $h^{-1}(y) \cap X_1 \neq \varnothing$, then throw out $\big(h^{-1}(y) \cap X_2\big) \times \{y\}$;

(2) if $h^{-1}(y) \cap X_1 = \varnothing$, that is, $h^{-1}(y) \subset X_2$, then we throw out all the elements that are in $h^{-1}(y) \times \{y\}$, except any one.

Lemma 1.29. *We have $R_0 \in \mathcal{R}(X, Y)$.*

Proof. For each $y \in Y \setminus Y_3$ we did not throw away anything, so for such y there are always $x \in X$ for which $(x, y) \in R_0$.

Let now $y \in Y_3$. If $h^{-1}(y) \cap X_1 \neq \varnothing$, then only (x, y) with $x \in X_2$ were thrown out, so $x \in X_1$ remained, for which we did not throw out the pair $(x, y) \in h$ and, therefore, it lies

in R_0. If $h^{-1}(y) \cap X_1 = \varnothing$, then we threw out all the elements that were in $h^{-1}(y) \times \{y\}$, except one, so for the non-thrown $(x, y) \in h$ we have $(x, y) \in R_0$.

Let us now deal with $x \in X$. If $x \in X_1$, then we did not discard the pairs $(x, y) \in h$. If $x \in X_2$, then, since $X_2 = \operatorname{im} g$, there is $y \in Y_2$ for which $x = g(y)$ and, therefore, $(x, y) \in h$, but we did not remove such pairs from h because $Y_2 \cap Y_3 = \emptyset$, so $(x, y) \in R_0$. □

Lemma 1.30. *We have $R_0 \in \mathcal{R}_0(X, Y)$.*

Proof. It is enough to show that for each pair $(x, y) \in R_0$ either x or y are not included in other pairs. If $y \in Y_2$, then y is included only in the pair $\big(g(y), y\big)$. If $y \in Y_1 \setminus Y_3$, then x is included only in the pair $\big(x, f(x)\big)$.

Finally, let now $y \in Y_3$. If $h^{-1}(y) \cap X_1 \neq \varnothing$, then we have eliminated all pairs of the form (x', y), $x' \in X_2$, so that $x \in X_1$, but each such x is included in exactly one pair, namely, $(x, f(x))$. If $h^{-1}(y) \cap X_1 = \varnothing$, then we have eliminated all pairs (x', y), $x' \in h^{-1}(y) \subset X_2$, except one, so y is included in exactly one such pair. □

This lemma completes the proof of the theorem. □

Note 1.31. Note that standard methods such as Zorn's Lemma are not applicable here. Indeed, the intersection of a chain of correspondences may not be a correspondence. Consider the following counterexample. Let

$$X = \{x_1, x_2\}, \qquad Y = \{y_i\}_{i \in \mathbb{N}},$$

and for each $i \in \mathbb{N}$ we define

$$R_i = \big(\{x_1\} \times Y\big) \cup \big(\{x_2\} \times \{y_k\}_{k \geq i}\big).$$

Each R_i is a correspondence, and the family $\{R_i\}_{i \in \mathbb{N}}$ forms a descending chain by inclusion. However

$$\bigcap_{i \in \mathbb{N}} R_i = \{x_1\} \times Y,$$

which is not a correspondence, since its projection onto X is not equal to X (the point x_2 does not have an image in Y). Thus, standard methods do not work. The above proof uses a constructive approach based on the explicit construction of the minimal element.

1.8.3 Structure of irreducible correspondences

Proposal 1.32. *The correspondence $R \in \mathcal{R}(X, Y)$ is irreducible if and only if there are partitions*

$$X = \bigsqcup_{i \in I} X_i, \qquad Y = \bigsqcup_{i \in I} Y_i$$

such that

$$R = \bigcup_{i \in I} (X_i \times Y_i),$$

and for each $i \in I$ the following condition is satisfied:

$$\min\{\#X_i, \#Y_i\} = 1.$$

Proof. Let first $R \in \mathcal{R}_0(X, Y)$. Let's put

$$R_X = \bigcup_{y \in Y} \{R^{-1}(y)\}, \qquad R_Y = \bigcup_{x \in X} \{R(x)\}.$$

Let us show that R_X and R_Y are partitions. Assume the opposite, and let, say, R_Y not be a partition. Since R is a correspondence, then R_Y is a covering of Y in which some elements of $R(x)$ and $R(x')$ for $x \neq x'$ intersect but do not coincide. Let $y \in R(x) \cap R(x')$, then $(x, y), (x', y) \in R$. Since $R(x) \neq R(x')$, one of these sets contains an element that is not contained in the other. Let, for definiteness, $y' \in R(x') \setminus R(x)$. Then $(x', y') \in R$, so if we remove the element (x', y) from the correspondence R, then in the resulting relation σ there will be $y \in \sigma(x)$ and $x' \in \sigma^{-1}(y')$, so σ is a correspondence. The latter contradicts the irreducibility of R.

So, let's write the partition R_X in the form $\{X_i\}_{i \in I}$. Note that for any $x, x' \in X_i$ we have $R(x) = R(x')$. Indeed, if $X_i = R^{-1}(y)$, then $R(x)$ and $R(x')$ contain y and, therefore, intersect. But R_Y is a partition, which means $R(x) = R(x')$.

Let us choose arbitrary $i \in I$, $x \in X_i$ and put $Y_i = R(x)$. Then, by virtue of what was said above, this definition is correct (does not depend on the choice of x). Let us show that the mapping $\varphi\colon X_i \mapsto Y_i$ is a bijection between R_X and R_Y.

If φ is not injective, then there exist $x, x' \in X$ lying in different elements of the partition R_X for which $R(x) = R(x')$. But then for $y \in R(x)$ $x, x' \in R^{-1}(y) \in R_X$ holds, a contradiction.

Finally, φ is surjective, since for any Y_i, $y \in Y_i$, the set $R^{-1}(y)$ is an element of the partition R_X. Let us choose an arbitrary $x \in R^{-1}(y)$. But then $R(x) \in R_Y$ contains y, so $\varphi(R^{-1}(y)) = Y_i$.

Since for any $x, x' \in X_i$ we have $R(x) = R(x') = Y_i$, then $X_i \times Y_i \subset R$. On the other hand, since R_X is a partition of X, then for any $x \in X$ there is an $X_i \in R_X$ such that $x \in X_i$, so each $(x, y) \in R$ is contained in some $X_i \times Y_i$.

Let us now show that $\min\{\#X_i, \#Y_i\} = 1$ for all i. If this is not so, then $\{x, x'\} \subset X_i$, $x \neq x'$, $\{y, y'\} \subset Y_i$, $y \neq y'$, and by removing the pair (x', y') from R, we obtain the relation σ in which $y' \in \sigma(x)$ and $x' \in \sigma^{-1}(y)$, so σ is again a correspondence, a contradiction with the irreducibility of R.

The converse statement is trivial. □

Example 1.33. Let us illustrate the structure of irreducible correspondence with a simple example. Let $X = \{x_1, x_2\}$ and $Y = \{y_1, y_2, y_3\}$. Compliance

$$R = \big(\{x_1\} \times \{y_1, y_2\}\big) \cup \big(\{x_2\} \times \{y_3\}\big)$$

is irreducible. Here the partition X consists of $X_1 = \{x_1\}$, $X_2 = \{x_2\}$, and the partition Y consists of $Y_1 = \{y_1, y_2\}$, $Y_2 = \{y_3\}$. In this case, $\min\{\#X_1, \#Y_1\} = \min\{1, 2\} = 1$ and $\min\{\#X_2, \#Y_2\} = \min\{1, 1\} = 1$.

1.8.4 Corollary for the Gromov–Hausdorff distance

The results obtained on the existence and structure of irreducible correspondences give us a powerful tool for calculating the Gromov–Hausdorff distance.

Consequence 1.34. *For any metric spaces X, Y:*

$$d_{GH}(X, Y) = \frac{1}{2} \inf\Big\{\operatorname{dis} R \mid R \in \mathcal{R}_0(X, Y)\Big\},$$

where $\mathcal{R}_0(X, Y)$ is the set of all irreducible correspondences between X and Y.

Proof. According to the theorem 1.16, we know that

$$d_{GH}(X, Y) = \frac{1}{2} \inf_{R \in \mathcal{R}(X, Y)} \operatorname{dis} R.$$

Now, for any correspondence $R \in \mathcal{R}(X,Y)$ by theorem 1.27 there is an irreducible $R_0 \in \mathcal{R}_0(X,Y)$ with $R_0 \subset R$. As proposed by 1.13, $\operatorname{dis} R_0 \leq \operatorname{dis} R$. Consequently, the infimum of distortions for all correspondences coincides with the infimum for irreducible correspondences. □

This result is key for the whole theory. It allows one to confine oneself to the study of correspondences whose structure is well understood (the 1.32 clause), which greatly simplifies the analysis and calculation of the GH distance for many classes of spaces, especially in finite-dimensional and combinatorial problems, which will be considered in the following lectures.

Example 1.35. Consider two three-point spaces $X = \{x_1, x_2, x_3\}$ and $Y = \{y_1, y_2, y_3\}$. Let $\rho_{ij} = |x_i x_j| = \rho_k$ and $\nu_{ij} = |y_i y_j| = \nu_k$, where $\{i,j,k\} = \{1,2,3\}$. Without loss of generality, we will assume that $\rho_1 \leq \rho_2 \leq \rho_3$ and $\nu_1 \leq \nu_2 \leq \nu_3$. We will show that

$$2d_{GH}(X,Y) = \max\{|\rho_1 - \nu_1|, |\rho_2 - \nu_2|, |\rho_3 - \nu_3|\}.$$

By Corollary 1.34, it suffices to describe what the irreducible correspondences $R \in \mathcal{R}_0(X,Y)$ look like. Let's use the proposal 1.32 for this. There are three types of three-point space partitions:

(1) to singletons,

(2) for one two-point and one single-point,

(3) for one three-point.

By virtue of the proposition 1.32, an irreducible correspondence defines a bijection between partitions, so both spaces must be partitioned in the same way. Case (3) is obviously not realized, in case (1) we have a bijection, and in case (2) the one-point and two-point subsets must correspond to each other. Thus, we have two types of correspondences: if $\{i,j,k\} = \{p,q,r\} = \{1,2,3\}$, then

$$(1) \quad R_1 = \{(x_i, y_p), (x_j, y_q), (x_k, y_r)\},$$

$$(2) \quad R_2 = \{(x_i, y_q), (x_i, y_r), (x_j, y_p), (x_k, y_p)\}.$$

We have

$$\operatorname{dis} R_1 = \max\{|\rho_{ij} - \nu_{pq}|, |\rho_{ik} - \nu_{pr}|, |\rho_{jk} - \nu_{qr}|\},$$
$$\operatorname{dis} R_2 = \max\{\rho_{jk},\ \nu_{qr},\ |\rho_{ij} - \nu_{pq}|,\ |\rho_{ik} - \nu_{pq}|,\ |\rho_{ij} - \nu_{pr}|,\ |\rho_{ik} - \nu_{pr}|\}.$$

Note that $\max\{\rho_{jk}, \nu_{qr}\} \geq |\rho_{jk} - \nu_{qr}|$, so $\operatorname{dis} R_2 \geq \operatorname{dis} R_1$, so it is sufficient to consider only the bijections R_1.

Note also that each bijection R_1 defines a bijection between the sets $\{\rho_a\}$ and $\{\nu_a\}$. We will show that the bijection $\rho_i \leftrightarrow \nu_i$ is optimal. To do this, let us check that when replacing this bijection with any other maximum

$$\max\{|\rho_1 - \nu_1|, |\rho_2 - \nu_2|, |\rho_3 - \nu_3|\}$$

cannot decrease.

Let the maximum be achieved at $|\rho_1 - \nu_1|$, and assume, without loss of generality, that $\rho_1 \leq \nu_1$. Then, when replacing the bijection with another, either $|\rho_1 - \nu_1|$ does not change, which can only lead to an increase in the distortion, or ν_1 changes to greater or equal to

ν_2 or ν_3, and the distortion can only increase. The case when the maximum is reached at $|\rho_3 - \nu_3|$ is considered similarly.

Let now the maximum be achieved at $|\rho_2 - \nu_2|$. Again, without loss of generality, assume that $\rho_2 \leq \nu_2$. If the change in bijection does not affect ρ_2 and ν_2, then the distortion can only increase. If the pair (ρ_2, ν_2) changes to (ρ_2, ν_3) or (ρ_1, ν_2), then the maximum can only increase. Therefore, it remains to consider the case when ρ_2 corresponds to ν_1, and ν_2 corresponds to ρ_3. Then ρ_1 corresponds to ν_3. But $|\nu_3 - \rho_1| \geq |\nu_2 - \rho_2|$, so the maximum again cannot decrease, which completes the proof of the formula.

Lecture summary

This Lecture proves:

(1) Hausdorff distance is a generalized pseudometric (proposition 1.4);

(2) on closed subsets the Hausdorff distance is a generalized metric (proposition 1.6);

(3) definition of the Gromov–Hausdorff distance (definition 1.9);

(4) the Gromov–Hausdorff distance is expressed through the distortion of correspondences (theorem 1.16);

(5) elementary properties of the Gromov–Hausdorff distance (proposition 1.17);

(6) $\mathcal{GH}$ is a generalized pseudometric space (theorem 1.18);

(7) $\mathcal{GH}$ is not separable (theorem 1.20);

(8) $\mathcal{GH}$ complete (theorem 1.21);

(9) distance d_{GH} is intrinsic (theorem 1.25);

(10) for any correspondence there is an irreducible subcorrespondence (theorem 1.27);

(11) the structure of irreducible correspondences is described by partitions (proposition 1.32);

(12) it is enough to consider only irreducible correspondences (a corollary of 1.34).

Accepted as known fact: fundamentals of metric geometry (see [1]). Left out of scope: constructing $\mathcal{GH}$ as its own class requires the axiom of choice.