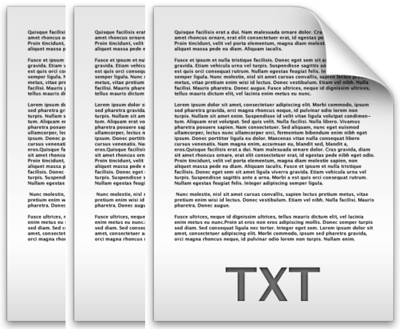
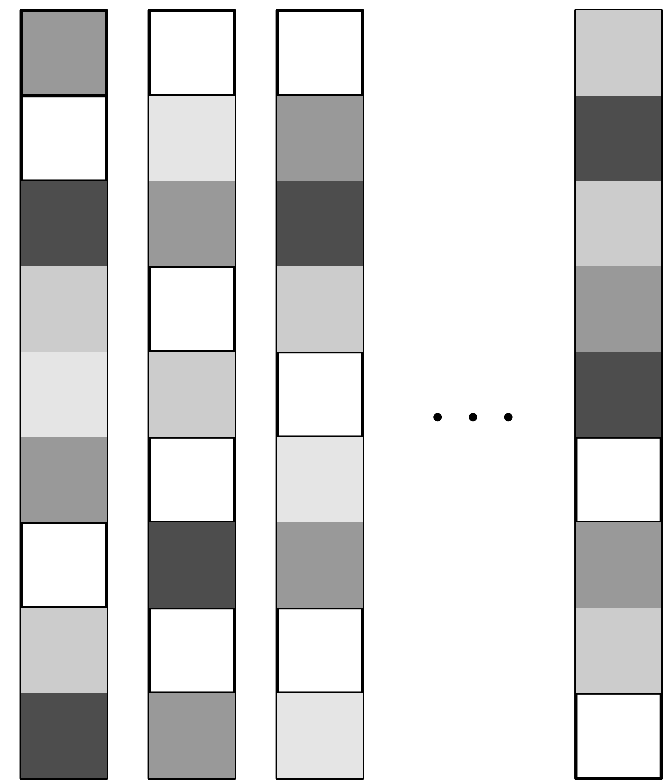
Topological Descriptors for Data

Features for data

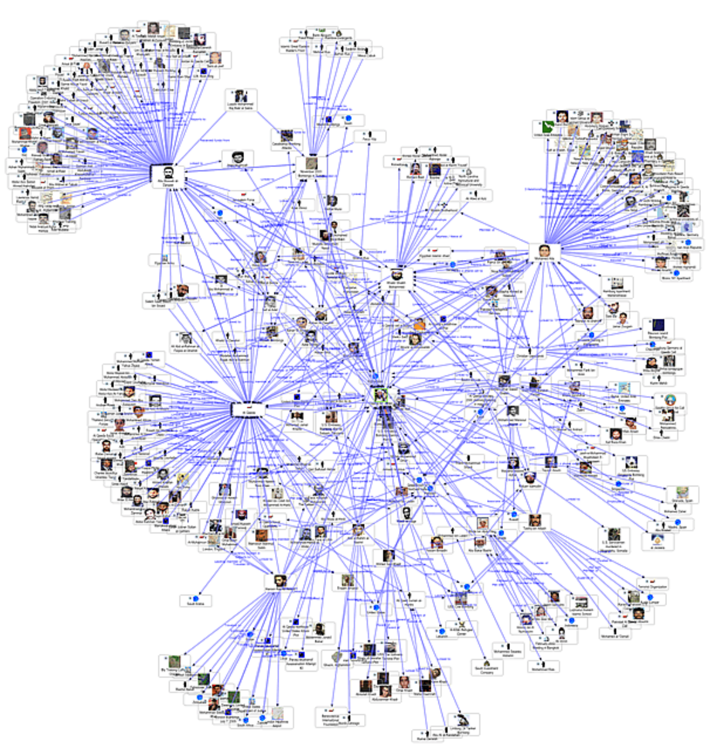


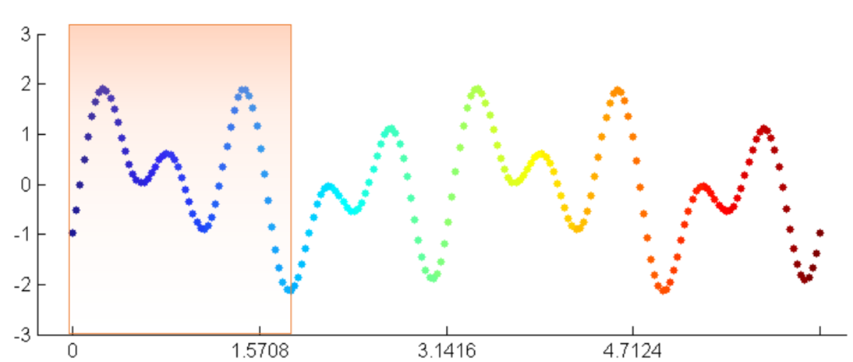
Data

Features $\in \mathbb{R}^n$



(feature design or learning)





bag of words, word2vec

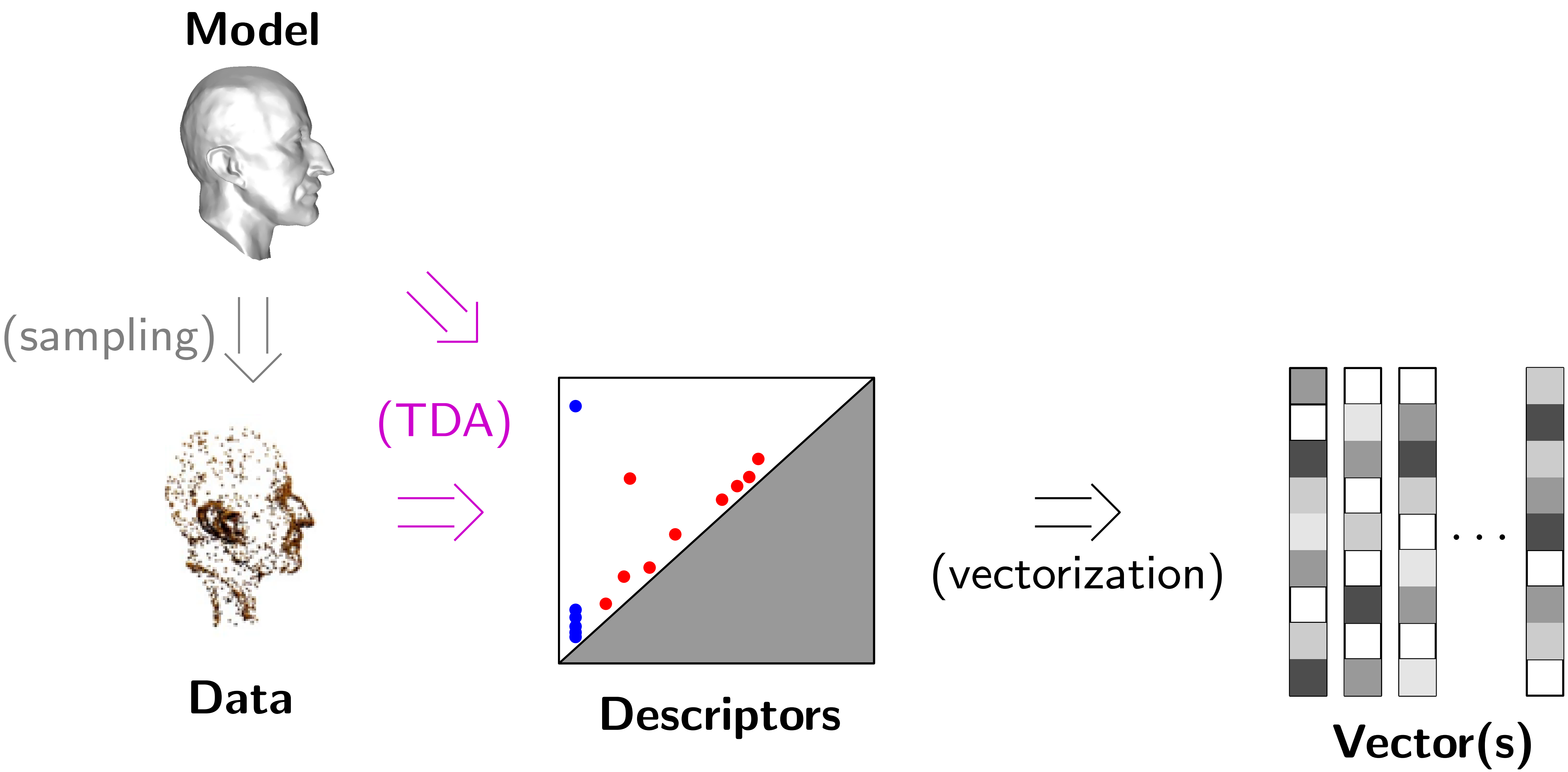
shape contexts, heat kernels

node2vec, Laplacian fact., rand. walks

sliding-window embeddings

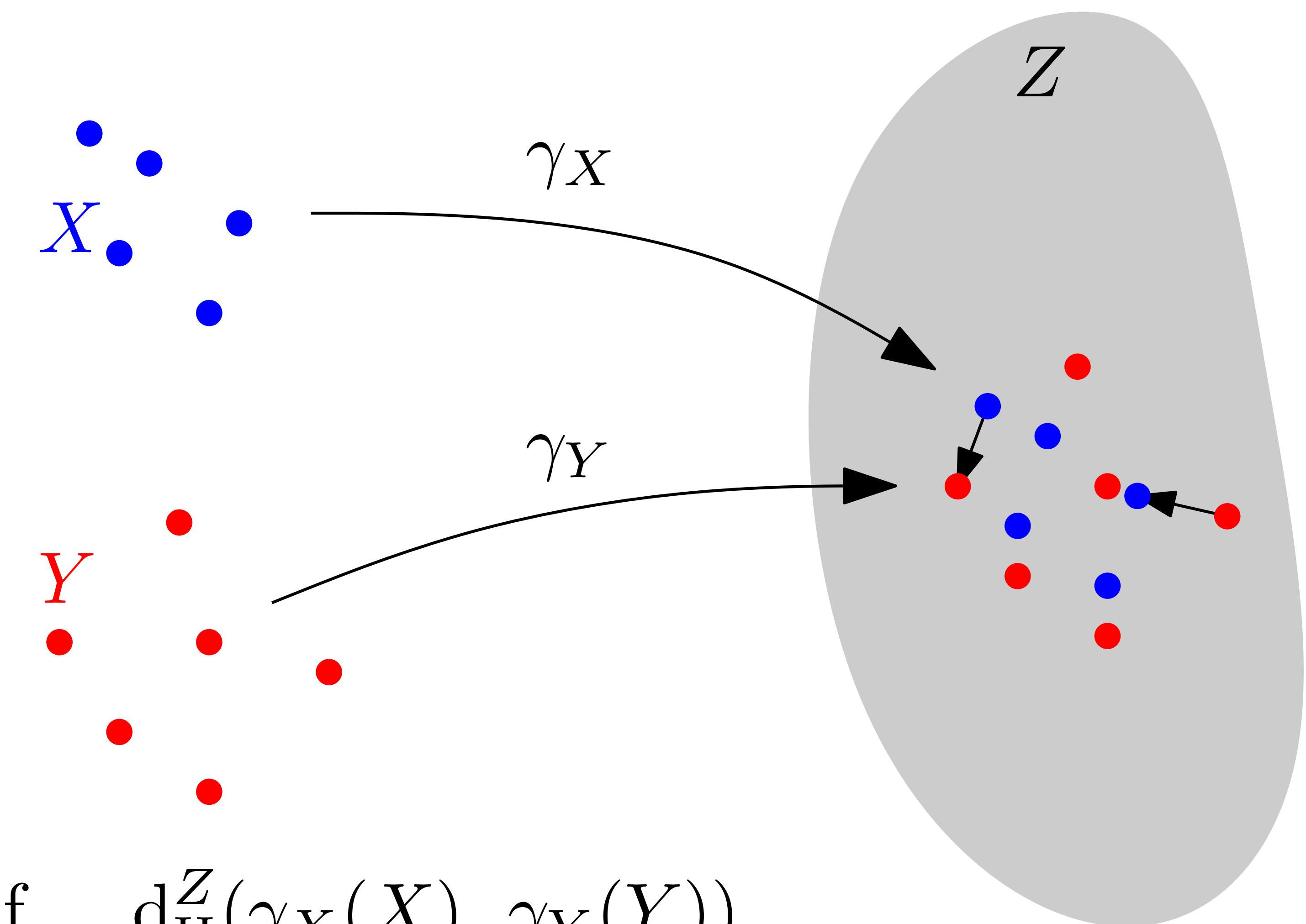
metric embeddings, auto-encoders

TDA for feature design



Mathematical framework

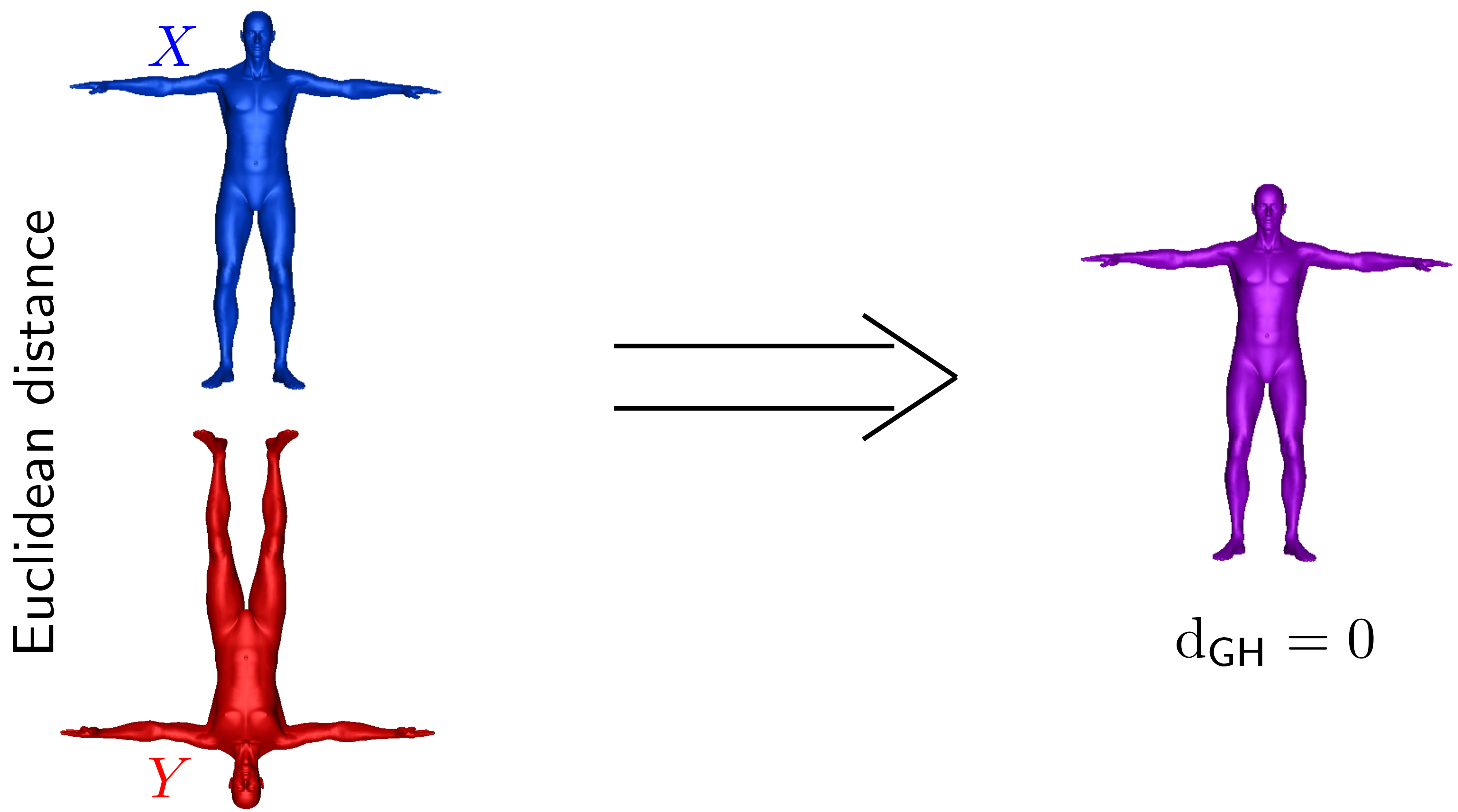
- geometric data set / underlying space $\equiv$ compact metric space
- distance between compact metric spaces $\equiv$ Gromov-Hausdorff (GH) distance



$$\mathrm{d_{GH}}(X,Y) = \inf_{\substack{\gamma_X : X \to Z \\ \gamma_Y : Y \to Z}} \mathrm{d}_\mathrm{H}^Z(\gamma_X(X),\, \gamma_Y(Y))$$

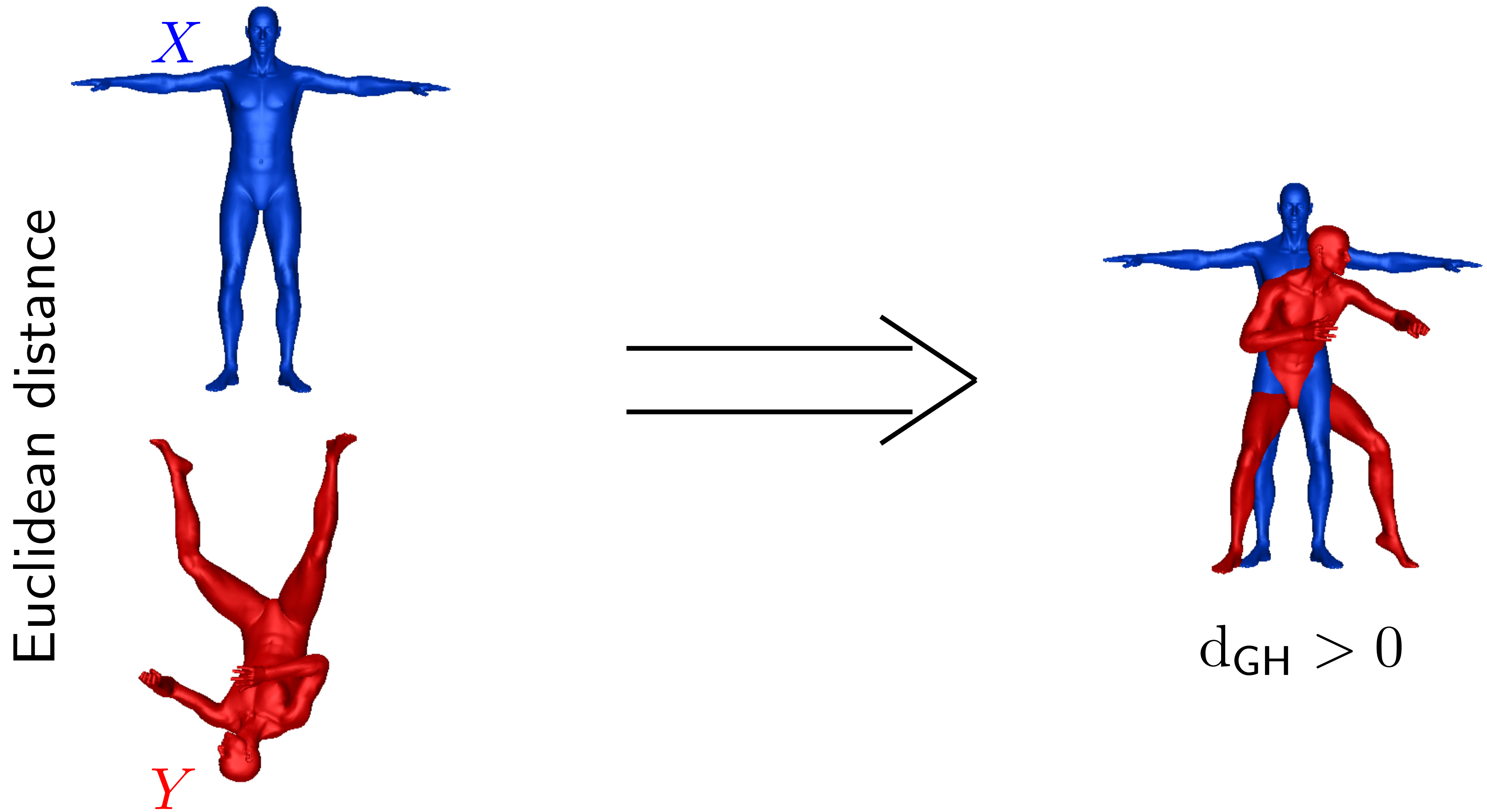
Mathematical framework

- geometric data set / underlying space $\equiv$ compact metric space
- distance between compact metric spaces $\equiv$ Gromov-Hausdorff (GH) distance



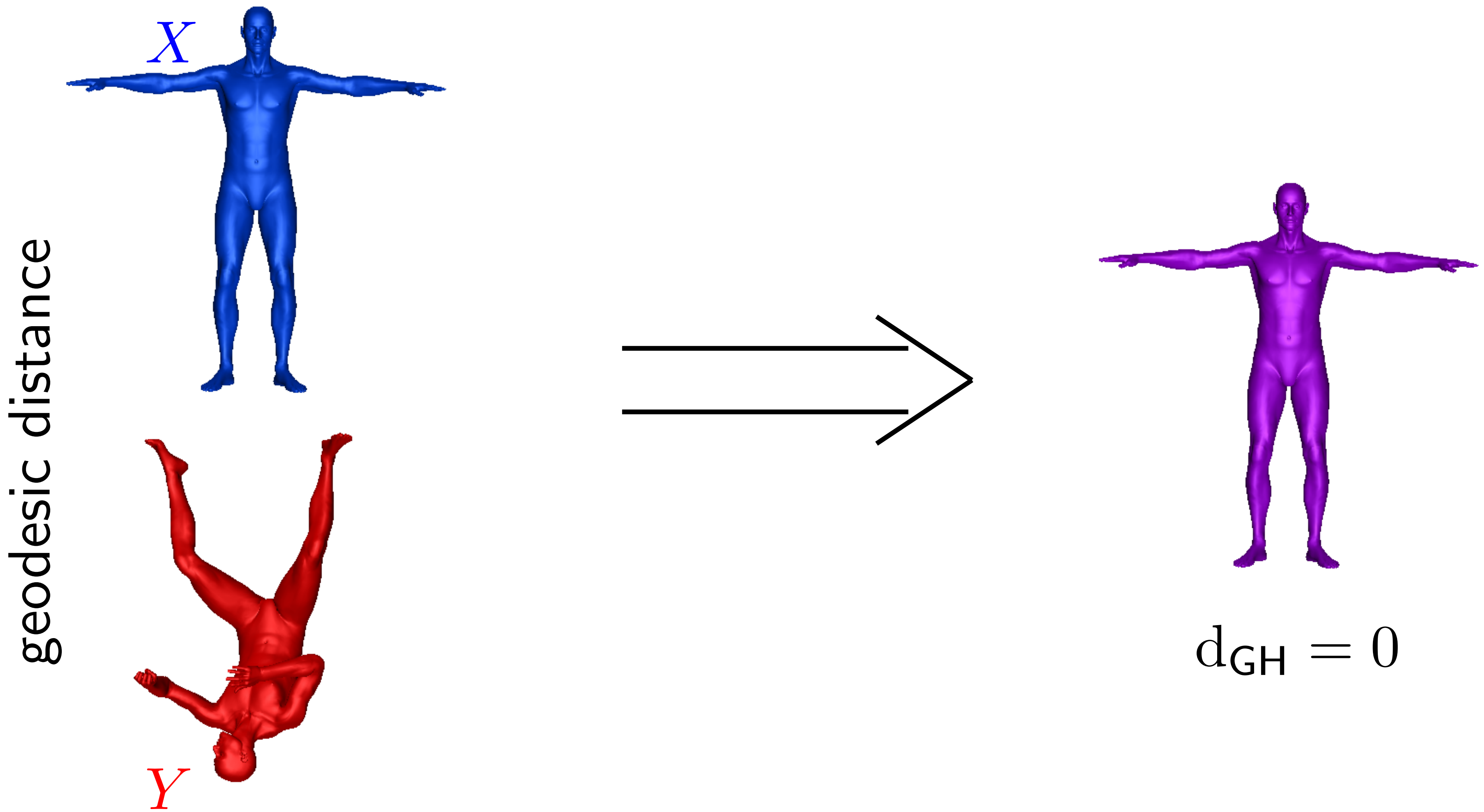
Mathematical framework

- geometric data set / underlying space $\equiv$ compact metric space
- distance between compact metric spaces $\equiv$ Gromov-Hausdorff (GH) distance



Mathematical framework

- geometric data set / underlying space $\equiv$ compact metric space
- distance between compact metric spaces $\equiv$ Gromov-Hausdorff (GH) distance



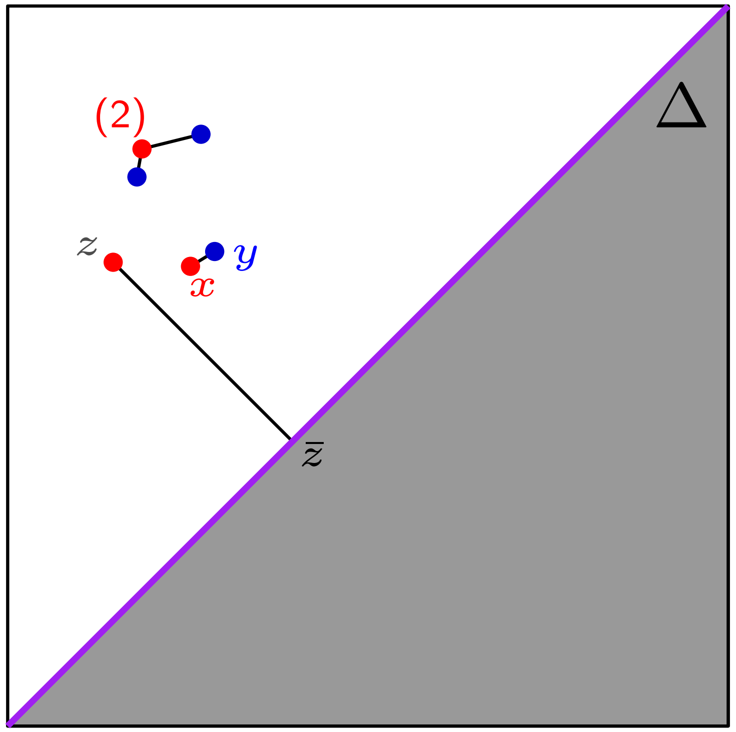
Mathematical framework

- geometric data set / underlying space $\equiv$ compact metric space

- distance between compact metric spaces $\equiv$ Gromov-Hausdorff (GH) distance

- distance between invariants (persistence diagrams) $\equiv \mathrm{d}_\infty$

$$\mathrm{d}_\infty := \inf_{\text{matchings}} \max \Big\{ \max_{(x,\,y)\text{ matched}} \|x-y\|_\infty, \max_{z\text{ unmatched}} \|z-\bar{z}\|_\infty \Big\}$$



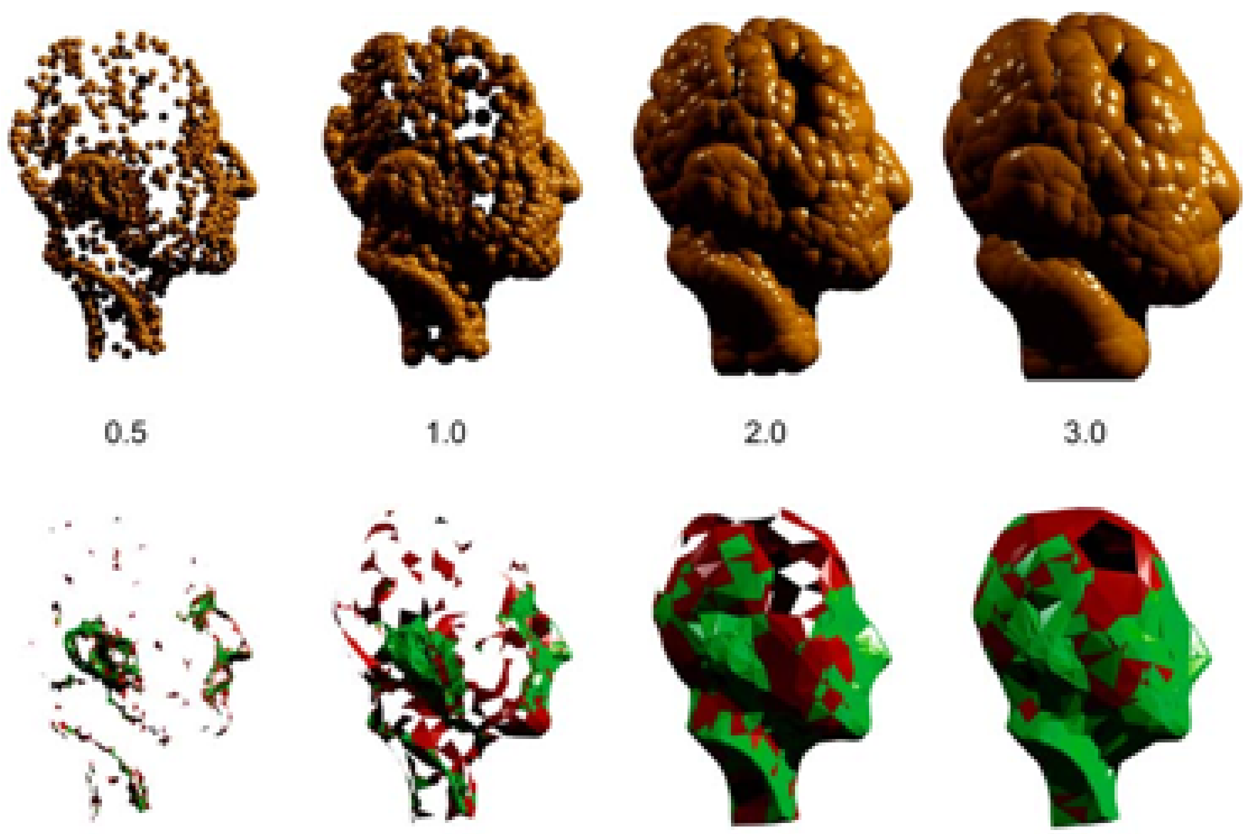
Global topological descriptors

Input: a compact metric space (X, d_X)

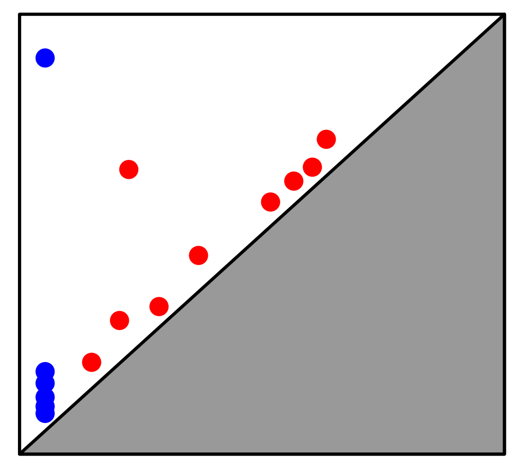
Descriptor: $\mathrm{dgm}\, \mathcal{F}(X, \mathrm{d}_X)$, where $\mathcal{F}(X, \mathrm{d}_X)$ is some simplicial filtration over X derived from d_X (proxy for union of balls)

dataset

$\Longrightarrow$



$\Longrightarrow$



persistence diagram

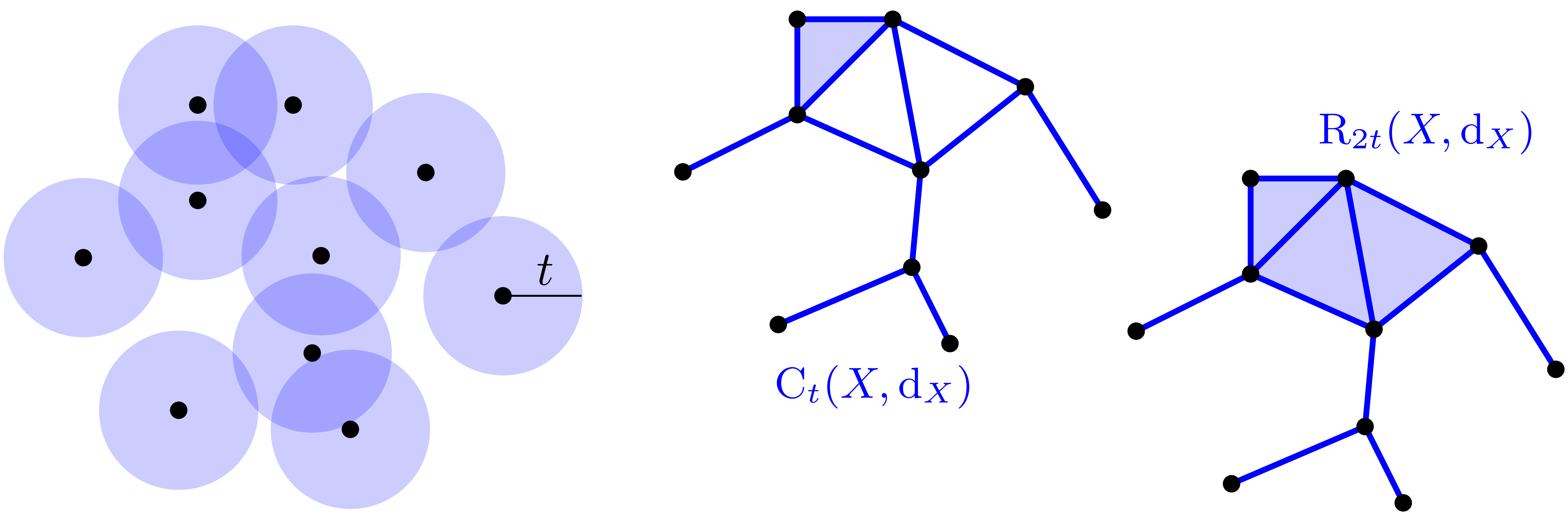
filter: distance function

simplicial approximation (/ Čech / Rips filtration)

Global topological descriptors

Input: a compact metric space (X, d_X)

Descriptor: $\mathrm{dgm}\,\mathcal{F}(X, \mathrm{d}_X)$, where $\mathcal{F}(X, \mathrm{d}_X)$ is some simplicial filtration over X derived from d_X (proxy for union of balls)



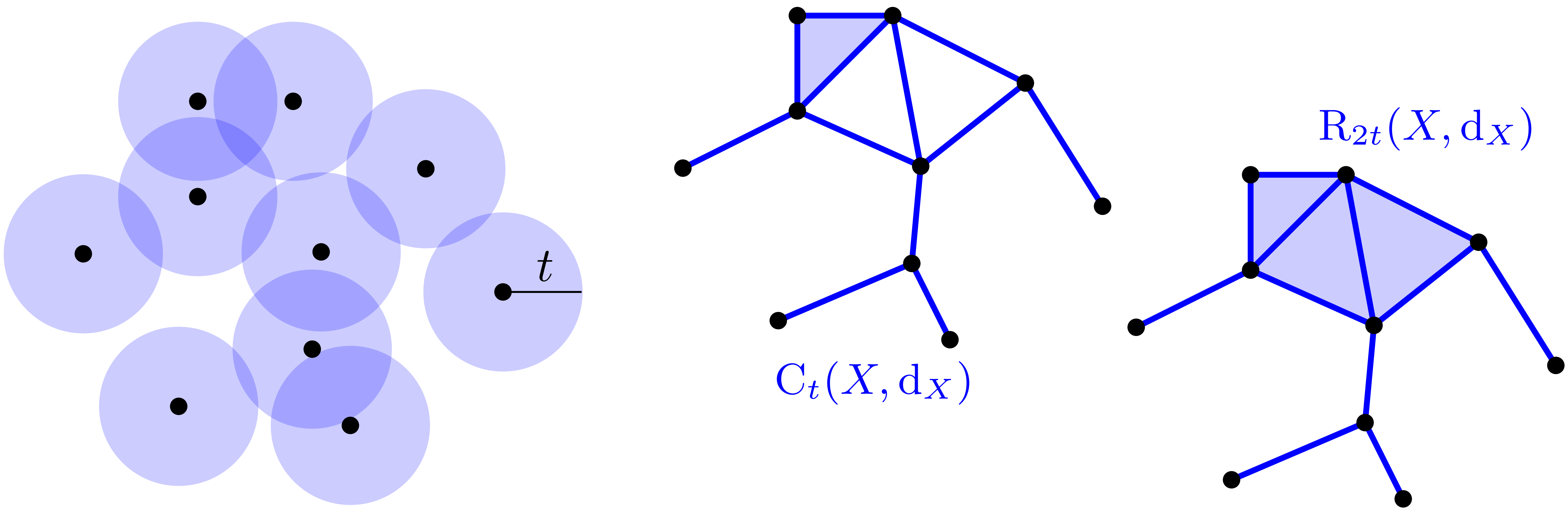
$$C_t(X, \mathrm{d}_X) := \left\{ \sigma = \{x_0, \cdots, x_k\} \in 2^X \mid \bigcap_{i=0}^{k} B_X(x_i, t) \neq \emptyset \right\}$$

$$R_t(X, \mathrm{d}_X) := \left\{ \sigma = \{x_0, \cdots, x_k\} \in 2^X \mid \max_{0 \leq i \leq j \leq k} \mathrm{d}_X(x_i, x_j) \leq t \right\}$$

Global topological descriptors

Input: a compact metric space (X, d_X)

Descriptor: $\mathrm{dgm}\, \mathcal{F}(X, \mathrm{d}_X)$, where $\mathcal{F}(X, \mathrm{d}_X)$ is some simplicial filtration over X derived from d_X (proxy for union of balls)

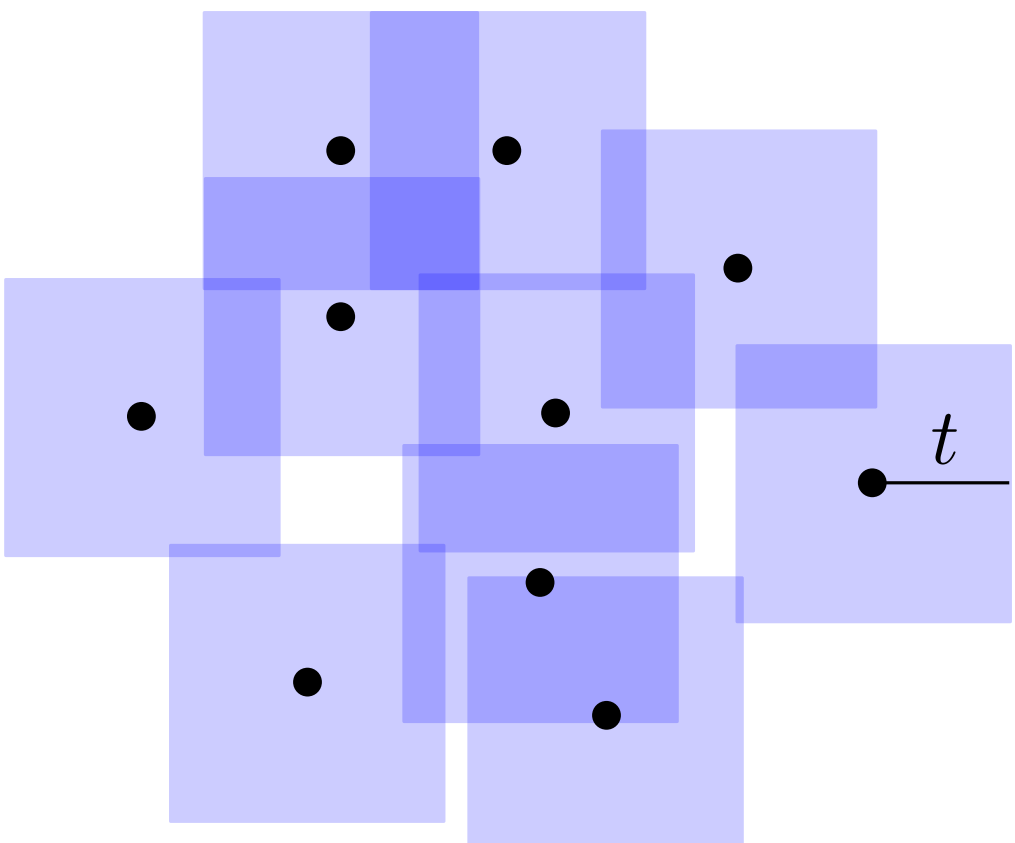


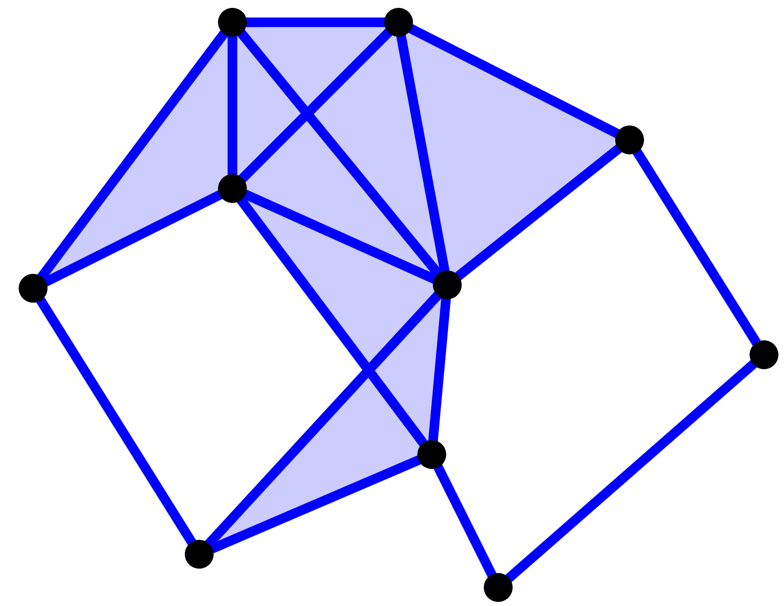
$$R_t(X, \mathrm{d}_X) \subseteq C_t(X, \mathrm{d}_X) \subseteq R_{2t}(X, \mathrm{d}_X)$$

Global topological descriptors

Input: a compact metric space (X, d_X)

Descriptor: $\mathrm{dgm}\,\mathcal{F}(X, \mathrm{d}_X)$, where $\mathcal{F}(X, \mathrm{d}_X)$ is some simplicial filtration over X derived from d_X (proxy for union of balls)

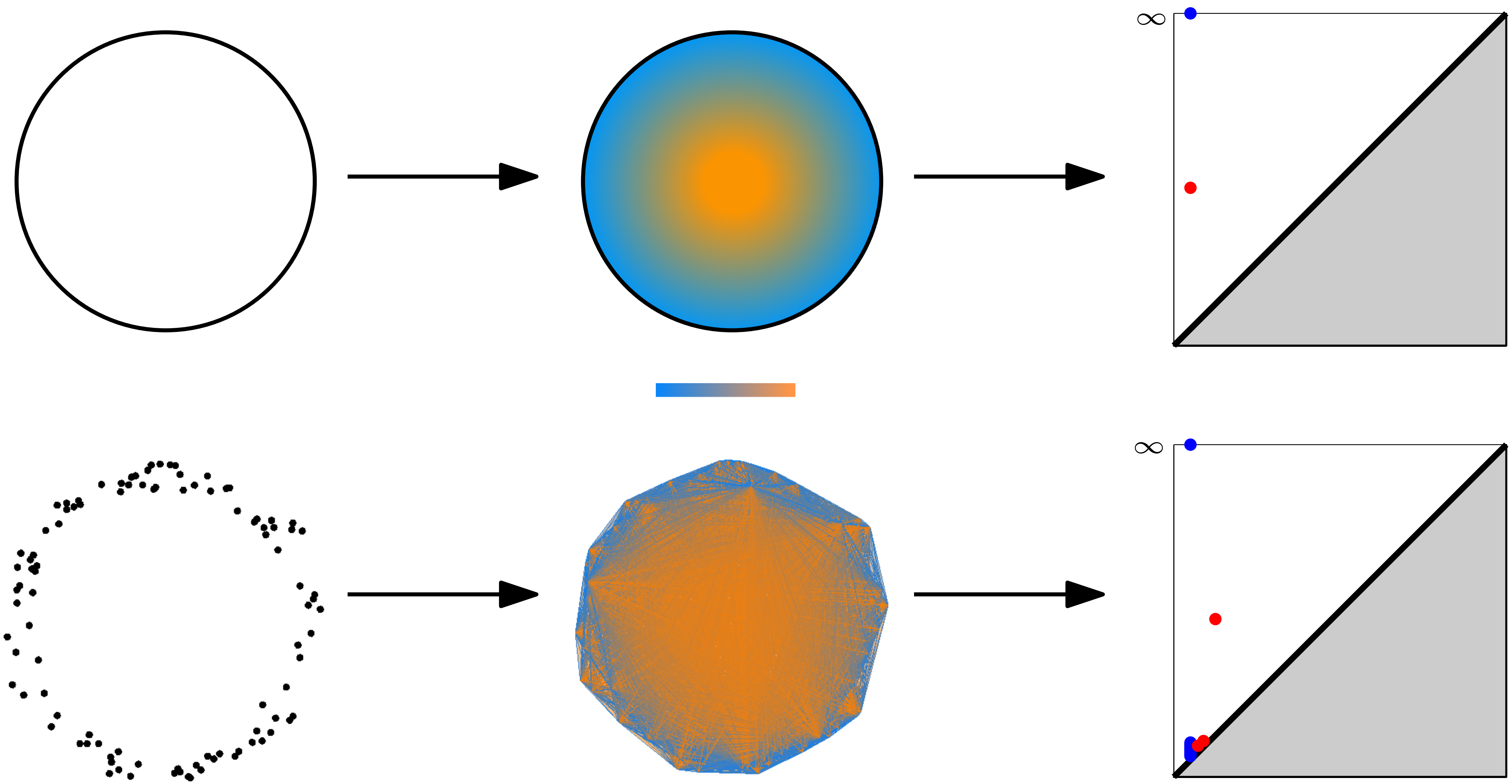




$\mathrm{C}_t(X, \mathrm{d}_X) = \mathrm{R}_{2t}(X, \mathrm{d}_X)$

Stability

Theorem: [Chazal, de Silva, O. 2013]
For any compact metric spaces (X, d_X) and (Y, d_Y),
$\mathrm{d}_\infty^\infty(\mathrm{dgm}\,\mathcal{R}(X, \mathrm{d}_X),\ \mathrm{dgm}\,\mathcal{R}(Y, \mathrm{d}_Y)) \leq 2\mathrm{d}_{\mathrm{GH}}(X, Y)$.

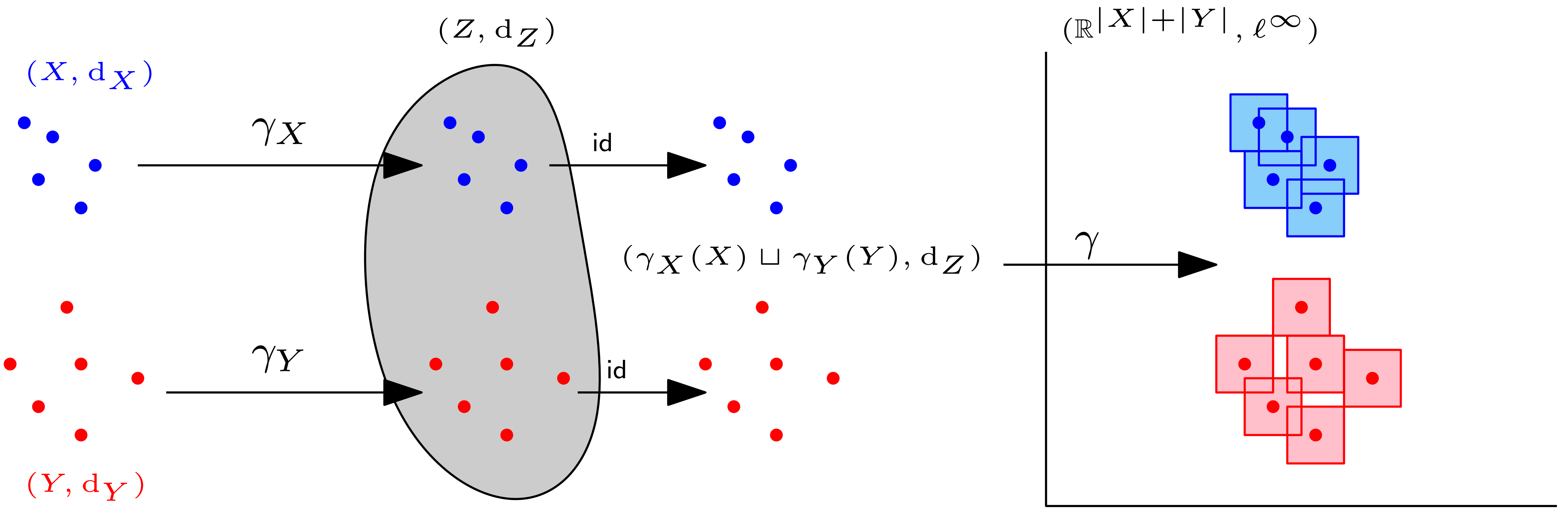


Stability

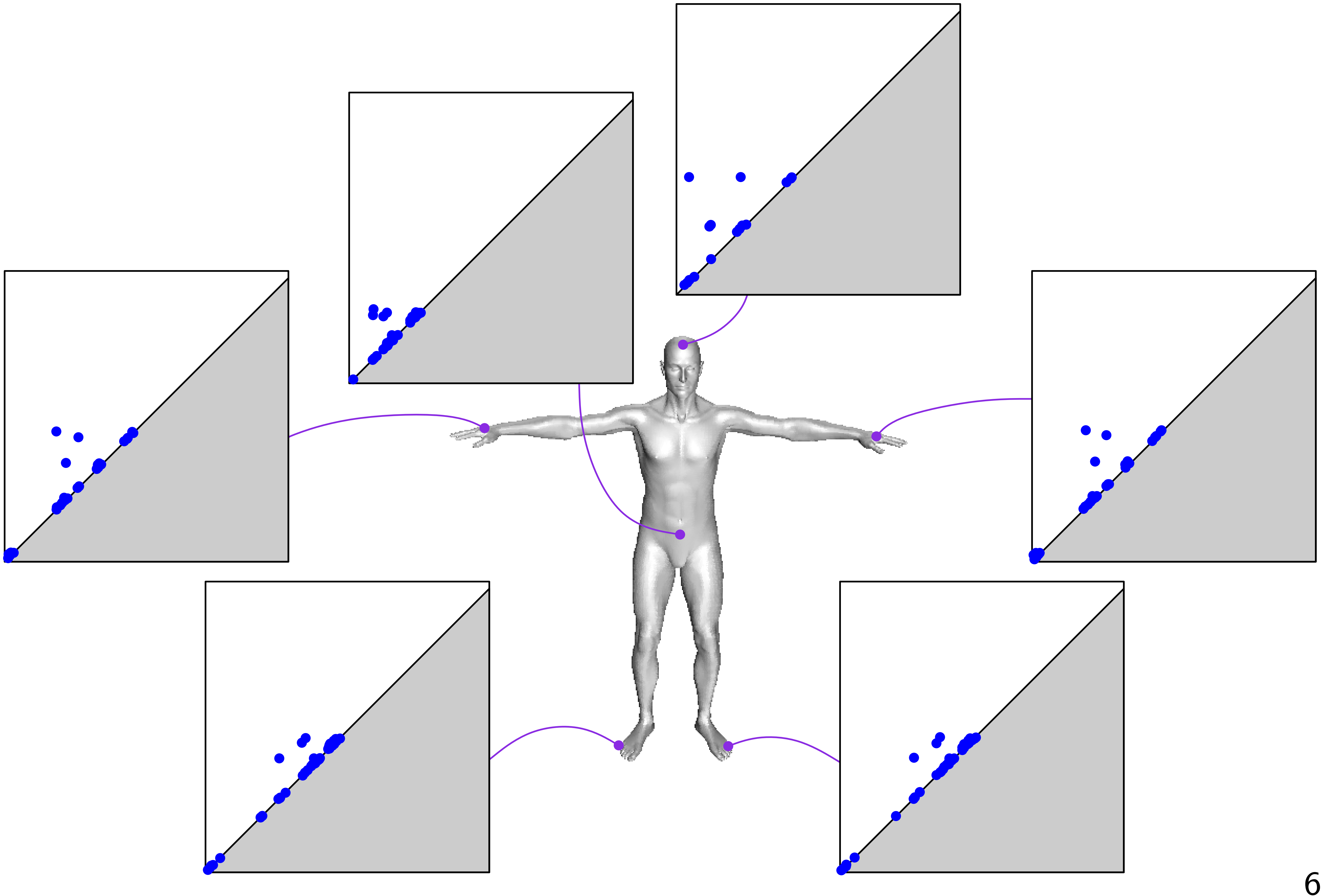
finite

Theorem: [Chazal, de Silva, O. 2013]
For any ~~compact~~ metric spaces (X, d_X) and (Y, d_Y),
$\mathrm{d}_\infty^\infty(\mathrm{dgm}\,\mathcal{R}(X, \mathrm{d}_X),\ \mathrm{dgm}\,\mathcal{R}(Y, \mathrm{d}_Y)) \leq 2\mathrm{d}_{\mathrm{GH}}(X, Y)$.

Proof outline:



Local topological descriptors

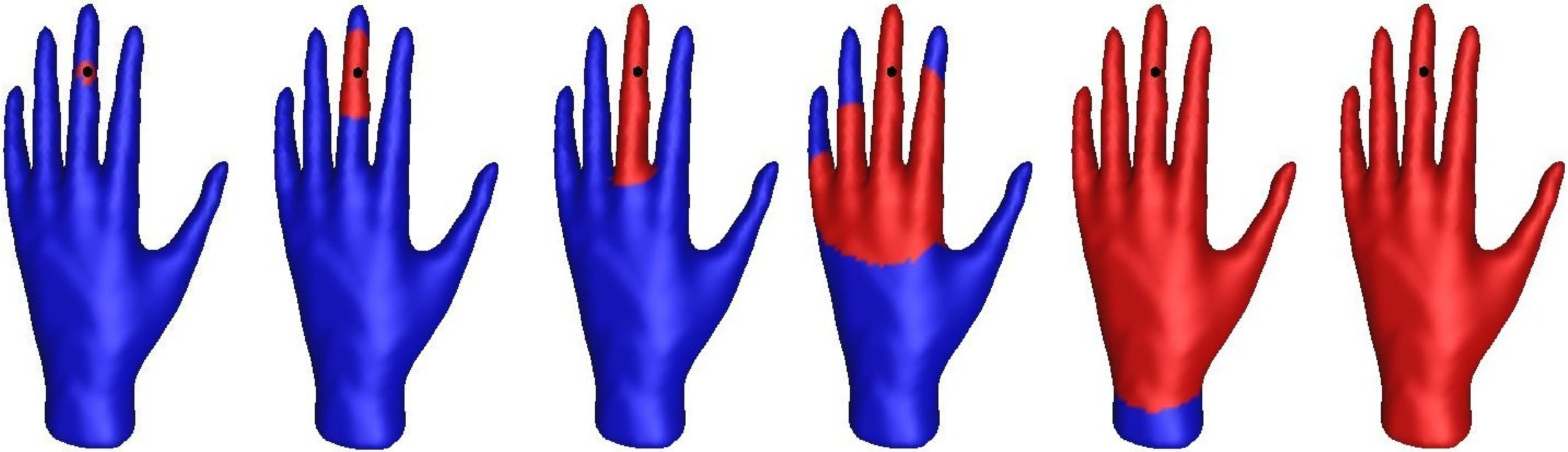


Local topological descriptors

Input: a compact *geodesic space* (X, d_X), a basepoint $x \in X$

Filter: $\operatorname{dgm} \mathrm{d}_X(x, \cdot)$

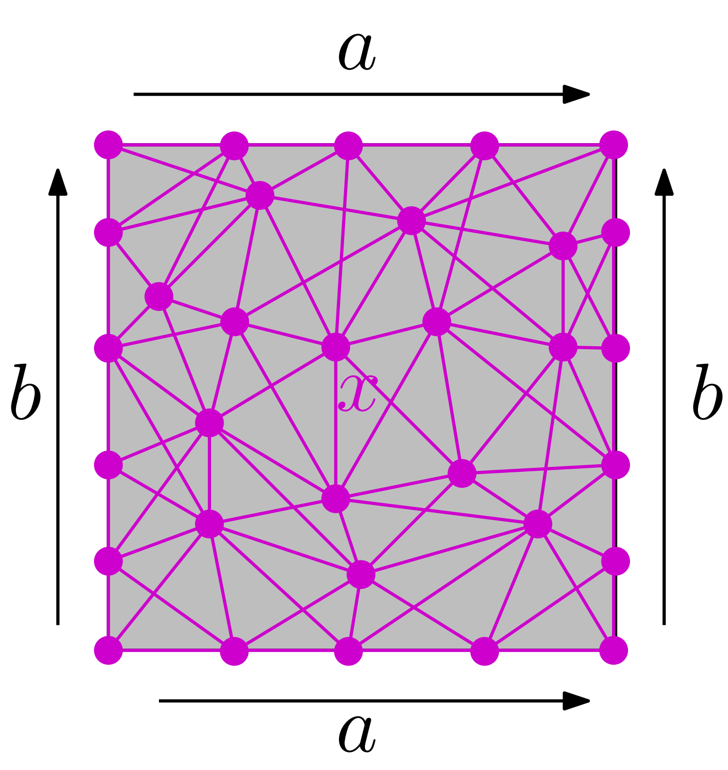
Computation: using a pair of Rips filtrations [Chazal et al. 2009]



Stability

Thm (local stability): [Carrière, O., Ovsjanikov 2015]
Let (X, d_X) and (Y, d_Y) be compact geodesic spaces with positive convexity radius ($\varrho(X), \varrho(Y) > 0$). Let $x \in X$ and $y \in Y$.
If $\mathrm{d}_{\mathrm{GH}}((X,x),(Y,y)) \leq \frac{1}{20} \min\{\varrho(X),\ \varrho(Y)\}$, then

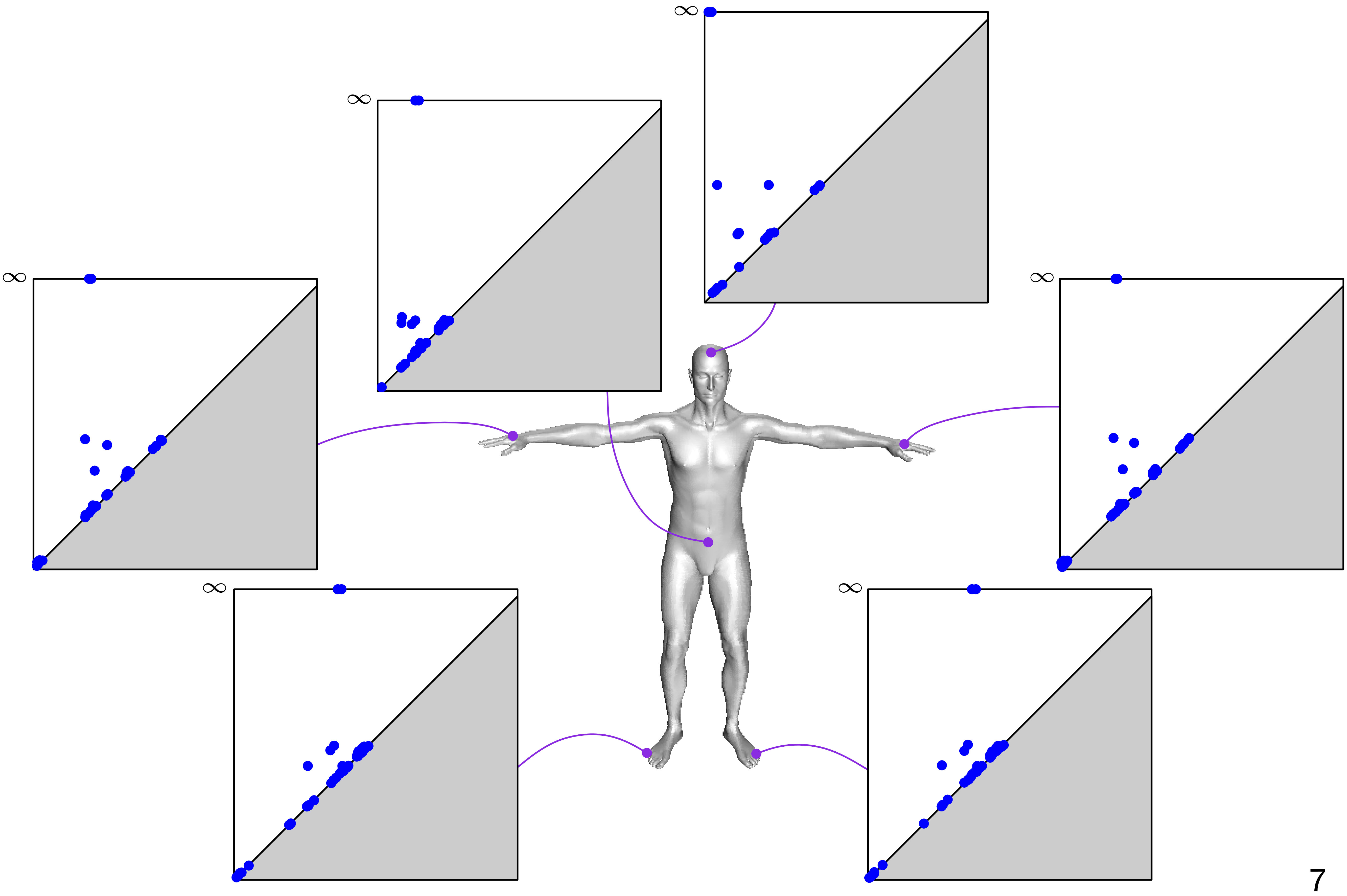
$$\mathrm{d}_\infty(\mathrm{dgm}\,\mathrm{d}_X(\cdot,x), \mathrm{dgm}\,\mathrm{d}_Y(\cdot,y)) \leq 20\ \mathrm{d}_{\mathrm{GH}}((X,x),(Y,y)).$$



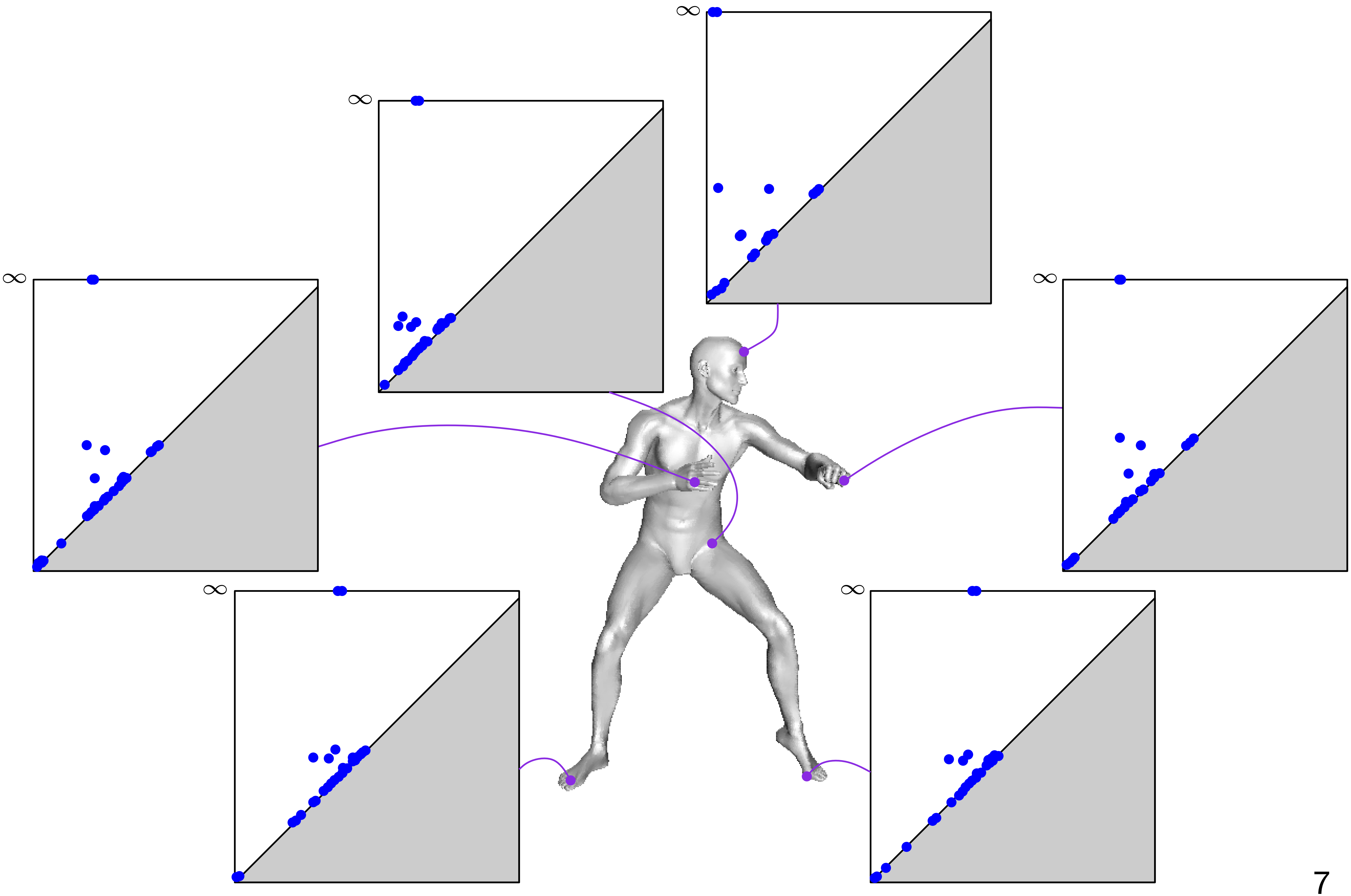
$$\mathrm{d}_{\mathrm{GH}}(T, X) \overset{\#X\to\infty}{\longrightarrow} 0$$

$$\mathrm{d}_\infty(\mathrm{dgm}\,\mathrm{d}_T(\cdot,x),\ \mathrm{dgm}\,\mathrm{d}_X(\cdot,x)) > 0$$

Stability



Stability

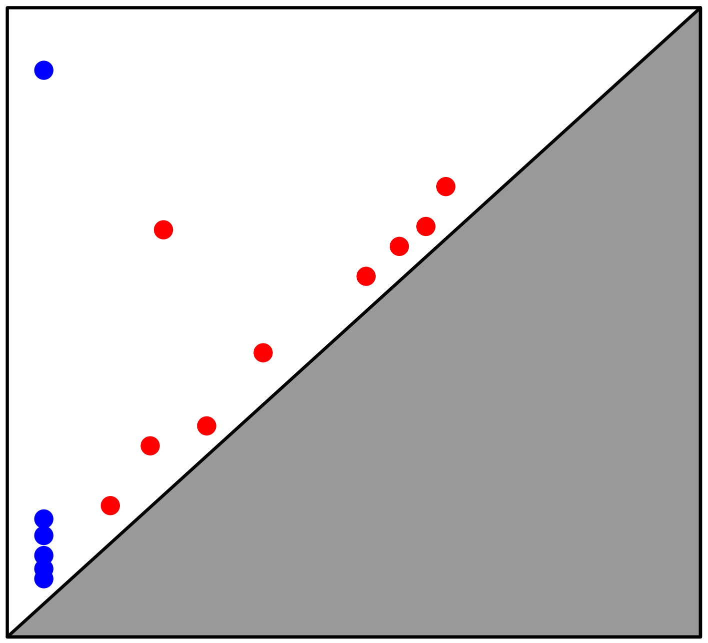


Wrap'up

Topological Persistence

⟹

Lipschitz



Data

Descriptors

Mathematical framework:

compact mspaces

mod isometries

(GH-distance)

⟹

discrete measures

(Wasserstein-type dist.)