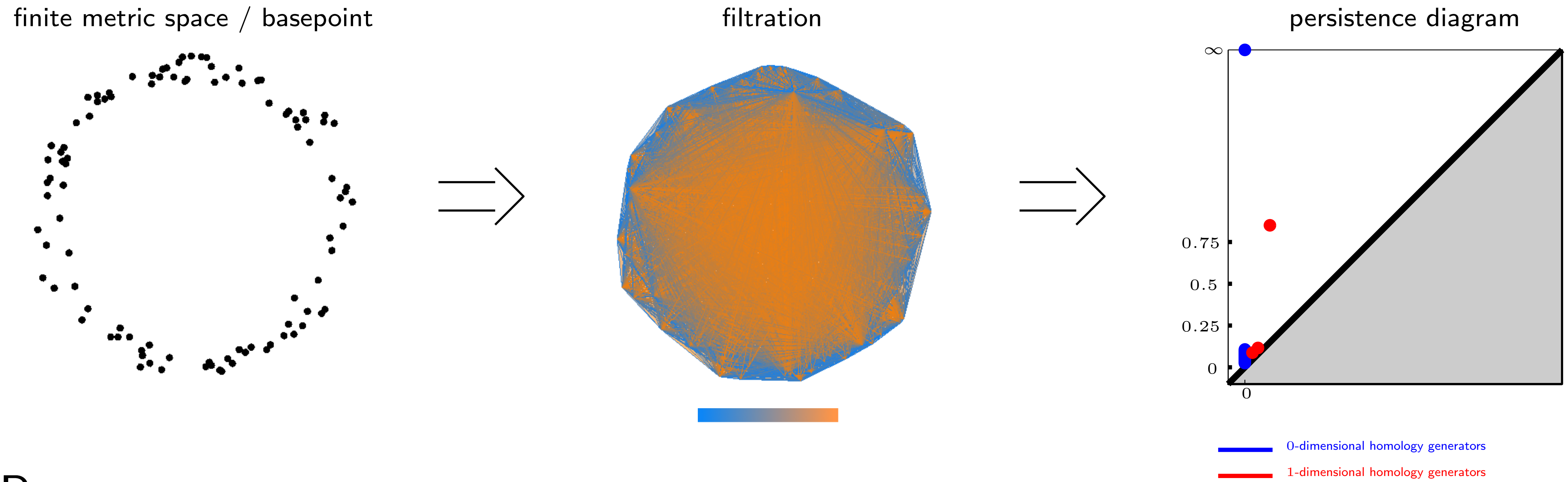
Persistence diagrams as descriptors



Pros:

- topological descriptors carry information of a different nature

- they enjoy stability properties, e.g. $\mathrm{d}_{\mathrm{B}}^{\infty}(\mathcal{R}(X), \mathcal{R}(Y)) \leq 2\mathrm{d}_{\mathrm{GH}}(X, Y)$

Cons:

- the space of persistence diagrams is not a vector/Hilbert space
  $\rightarrow$ bad for supervised learning and statistics
- descriptors can be slow to compute and (more importantly) to compare
  $\rightarrow$ bad for applications

Statistics on the space of persistence diagrams

Defining means (Fréchet means):

Given $D_1, \cdots, D_n$: persistence diagrams, is the following set empty?

$$\arg\min_{D} \sum_{i=1}^{n} \mathrm{d}(D, D_i)^2 \quad \text{for some metric } \mathrm{d} \text{ between diagrams}$$

Statistics on the space of persistence diagrams

Defining means (Fréchet means):

Definition p-th Wasserstein distance:

$$W_p(A,B) = \inf_{M:A\leftrightarrow B} c_p(M)$$

where

$$c_p(M) = \left(\sum_{(a,b)\in M} \|a-b\|_\infty^p + \sum_{\substack{s\in A\sqcup B \\ s \text{ unmatched}}} \|s-\bar{s}\|_\infty^p \right)^{1/p}$$

Statistics on the space of persistence diagrams

Defining means (Fréchet means):

Definition p-th Wasserstein distance:

$$W_p(A,B) = \inf_{M:A\leftrightarrow B} c_p(M)$$

where

$$c_p(M) = \left(\sum_{(a,b)\in M} \|a-b\|_\infty^p + \sum_{\substack{s\in A\sqcup B \\ s \text{ unmatched}}} \|s-\bar{s}\|_\infty^p \right)^{1/p}$$

Theorem [Turner et al. 2012]: For any finite $p \geq 1$, the space of persistence diagrams equipped with W_p is *Polish* (complete and separable).

Corollary: Fréchet mean is well-defined

Statistics on the space of persistence diagrams

Defining means (Fréchet means):

Definition p-th Wasserstein distance:

$$W_p(A, B) = \inf_{M: A \leftrightarrow B} c_p(M)$$

where

$$c_p(M) = \left(\sum_{(a,b)\in M} \|a - b\|_\infty^p + \sum_{\substack{s \in A \sqcup B \\ s \text{ unmatched}}} \|s - \bar{s}\|_\infty^p \right)^{1/p}$$

Theorem [Turner et al. 2012]: For any finite $p \geq 1$, the space of persistence diagrams equipped with W_p is *Polish* (complete and separable).

Corollary: Fréchet mean is well-defined... but not unique (+hard to compute)

Outline

1. Supervised learning with diagrams: the kernel trick

2. Statistics with diagrams: the push-forward trick

1. The kernel trick

$\mathcal{X}$: a space in which we want to compare/classify elements

- feature map $\phi : \mathcal{X} \to \mathcal{H}$ equipped with inner product $< \cdot, \cdot >_{\mathcal{H}}$
- lift training/testing data to $\mathcal{H}$ through ϕ then solve learning problem

1. The kernel trick

$\mathcal{X}$: a space in which we want to compare/classify elements

- feature map $\phi : \mathcal{X} \rightarrow \mathcal{H}$ equipped with inner product $< \cdot, \cdot >_{\mathcal{H}}$
- lift training/testing data to $\mathcal{H}$ through ϕ then solve learning problem
- observation: many learning methods use only inner product

  $\rightarrow$ do not lift the data, instead compute the $k(x, y) =< \phi(x), \phi(y) >_{\mathcal{H}}$

1. The kernel trick

$\mathcal{X}$: a space in which we want to compare/classify elements

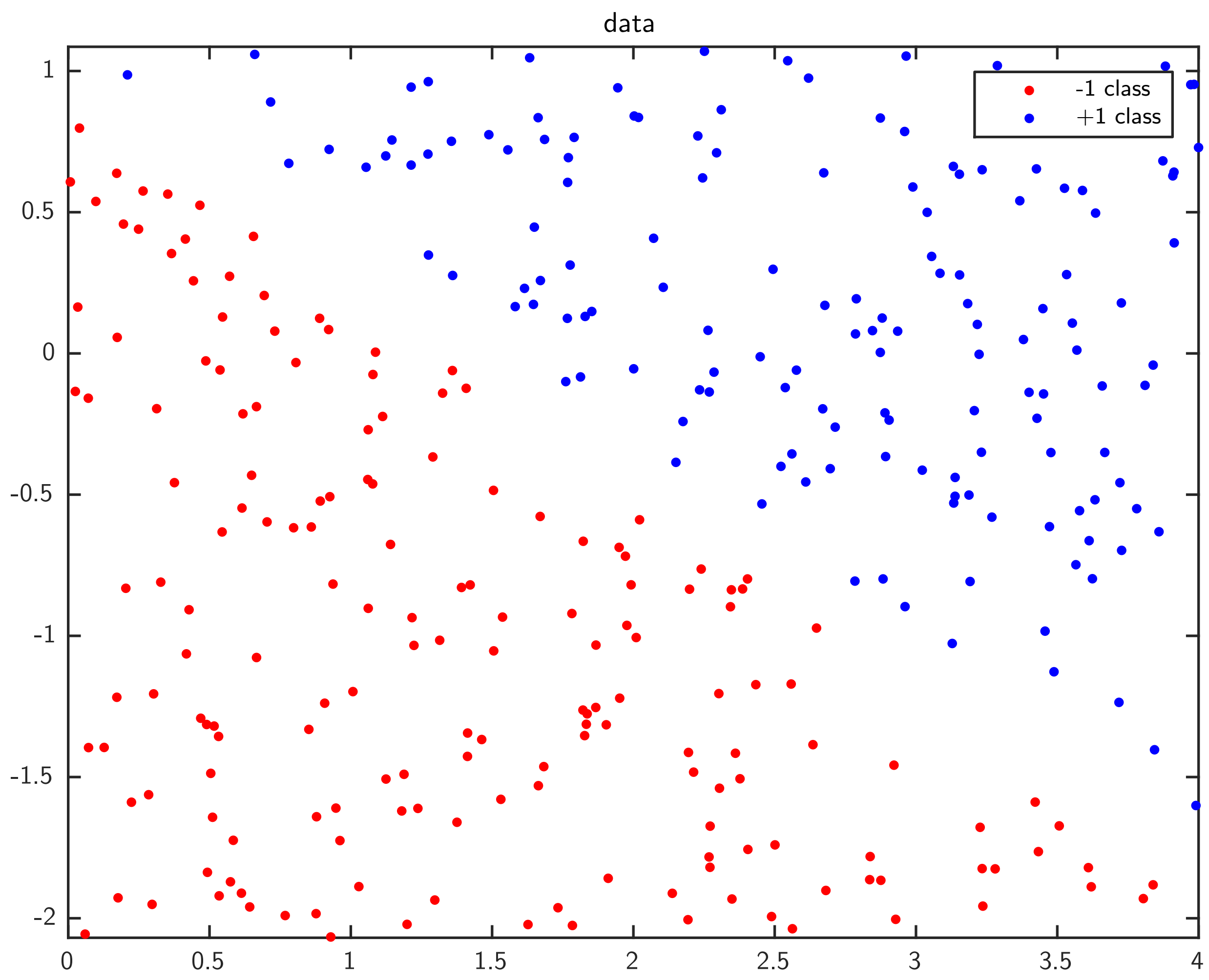
- feature map $\phi : \mathcal{X} \to \mathcal{H}$ equipped with inner product $< \cdot, \cdot >_{\mathcal{H}}$
- lift training/testing data to $\mathcal{H}$ through ϕ then solve learning problem
- observation: many learning methods use only inner product

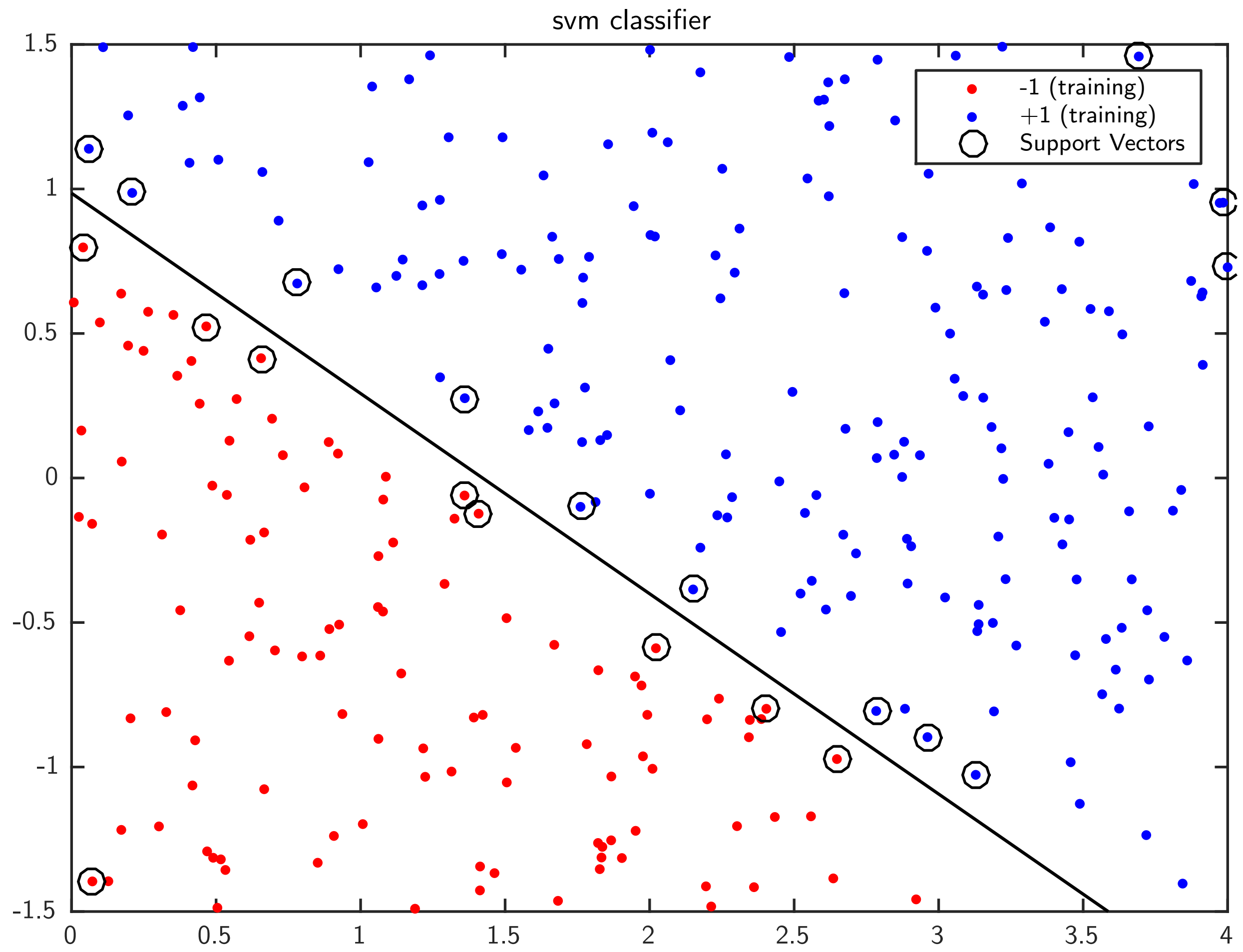
  $\to$ do not lift the data, instead compute the $k(x, y) =< \phi(x), \phi(y) >_{\mathcal{H}}$

Def.: A *reproducing kernel* is a map $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ such that $k(\cdot, \cdot) =< \phi(\cdot), \phi(\cdot) >_{\mathcal{H}}$ for some pair $(\phi, \mathcal{H})$.

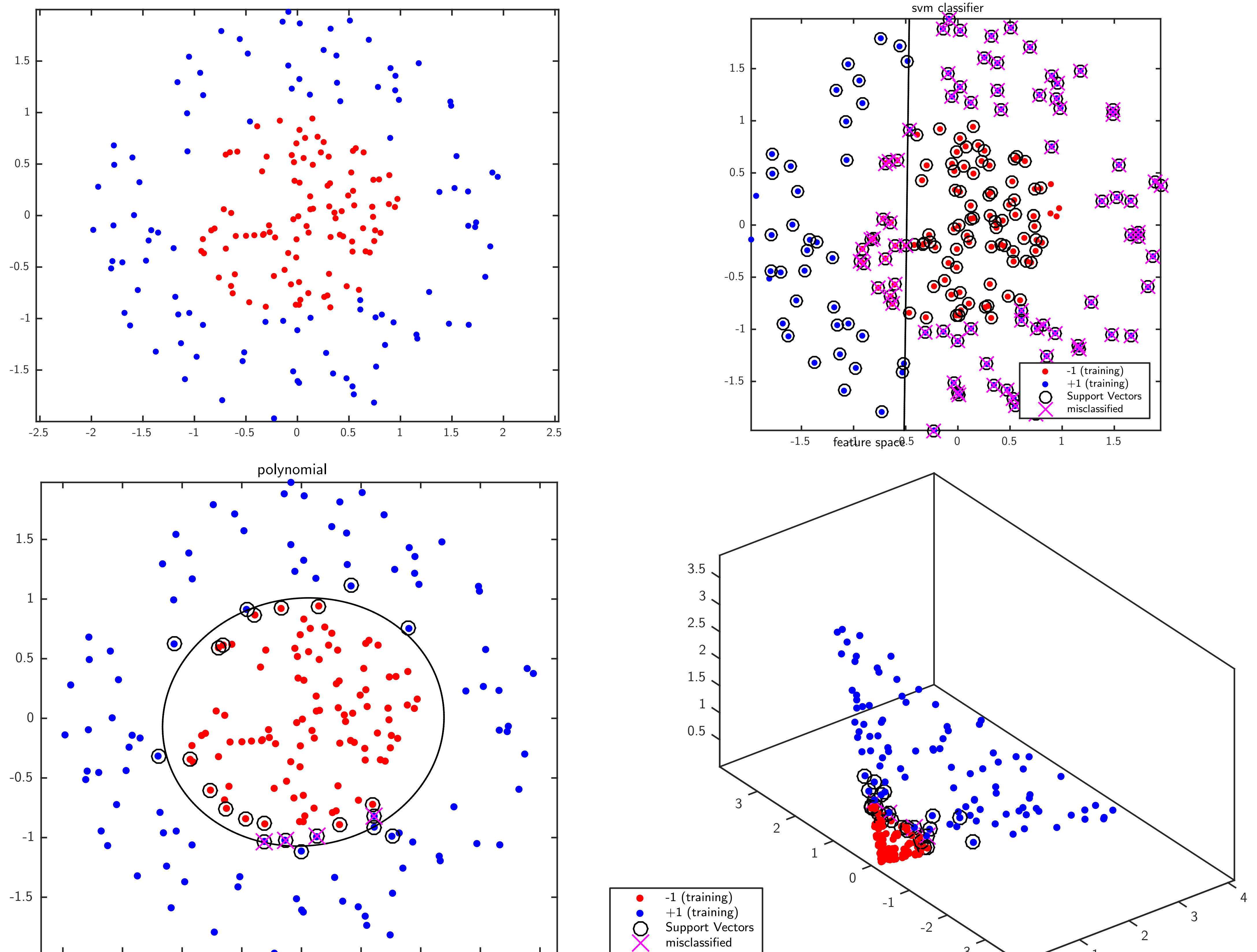
Thm.: [Moore, Aronszajn]
A pair $(\phi, \mathcal{H})$ exists whenever k is *positive semidefinite*, i.e.
$\sum_{i,j=1}^{n} c_i c_j \, k(x_i, x_j) \geq 0$ for all $n \in \mathbb{N}$, $c_1, \cdots, c_n \in \mathbb{R}$, and $x_1, \cdots, x_n \in \mathcal{X}$.

1. The kernel trick





1. The kernel trick



Desired properties (in the context of persistence diagrams)

- positive (semi-)definiteness
- stability w.r.t. $\mathrm{d_B}$ (a.k.a. W_∞) or W_p for some $p < +\infty$
- injectivity, or (even better) discriminativity w.r.t. $\mathrm{d_B}$ or some W_p
- small algorithmic cost
- finite dimensionality (Euclidean space)
- universality
- additivity

Kernels for persistence diagrams

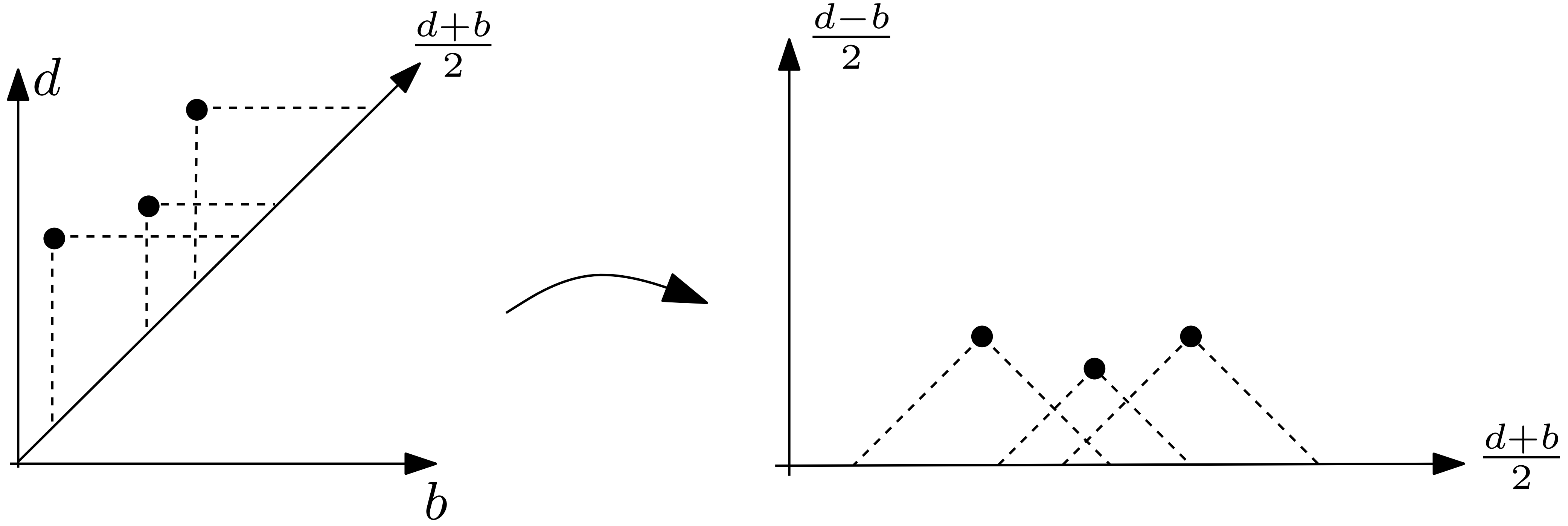
View persistence diagrams as:

- landscapes (collections of 1-d functions) [Bubenik 2012] [Bubenik, Dłotko 2015]
- discrete measures:
  - $\rightarrow$ histogram [Bendich et al. 2014]
  - $\rightarrow$ convolution with fixed kernel [Chepushtanova et al. 2015]
  - $\rightarrow$ convolution with varying kernel [Kusano, Fukumisu, Hiraoka 2016]
  - $\rightarrow$ heat diffusion [Reininghaus et al. 2015] $+$ exponential [Kwit et al. 2015]
- finite metric spaces [Carrière, O., Ovsjanikov 2015]
- roots of polynomials [Di Fabio, Ferri 2015]

Kernels for persistence diagrams

	landscapes	discrete measures	metric spaces
positive (semi-)definiteness	✓	✓	✓
RKHS	$L^2(\mathbb{N} \times \mathbb{R})$	$L^2(\mathbb{R}^2)$	$(\mathbb{R}^d, \|\cdot\|_2)$
stability w.r.t W_p	✓ $p = 2$	✓ $p = 1$	✓ $p = \infty$
injectivity	✓	✓	✗
discriminativity w.r.t. W_p	?	?	✗
algorithmic cost	$O(n^2)$	$O(n^2)$	f. map: $O(n^2)$ kernel: $O(d)$
universality	✗	✓	✗
additivity	✗	✓	✗

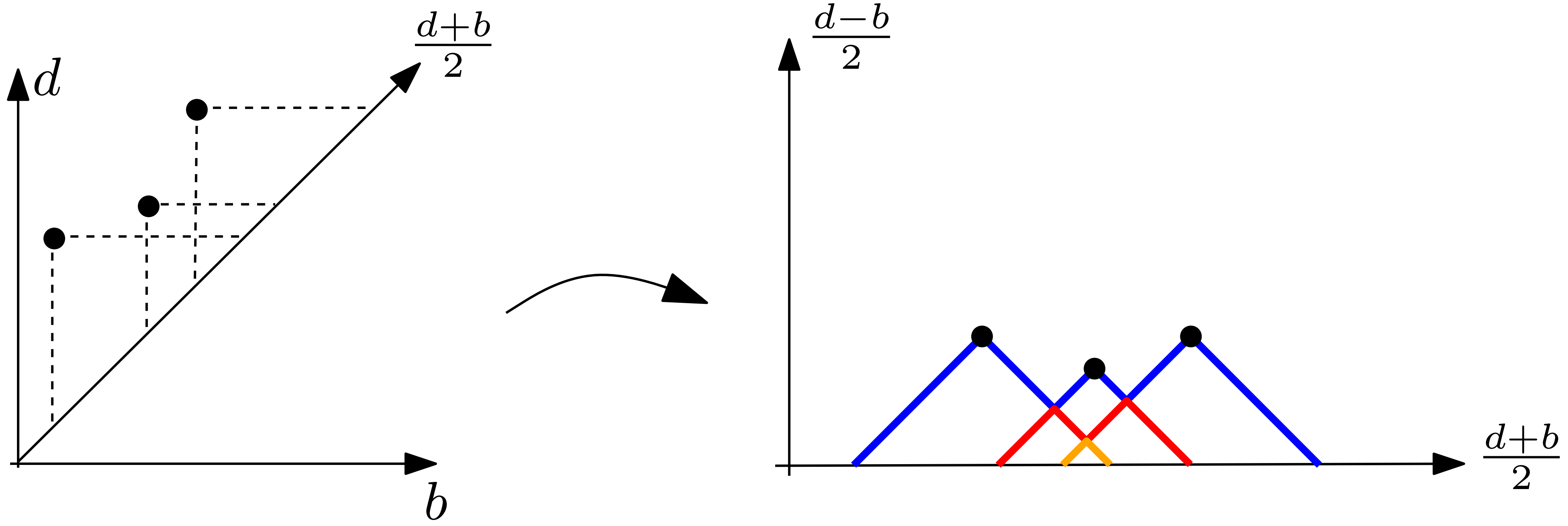
Landscapes



$$D = \{(\tfrac{d_i+b_i}{2}, \tfrac{d_i+b_i}{2})\} i \in I$$

$$(b, d) \quad \mapsto \quad (\tfrac{d+b}{2}, \tfrac{d-b}{2})$$

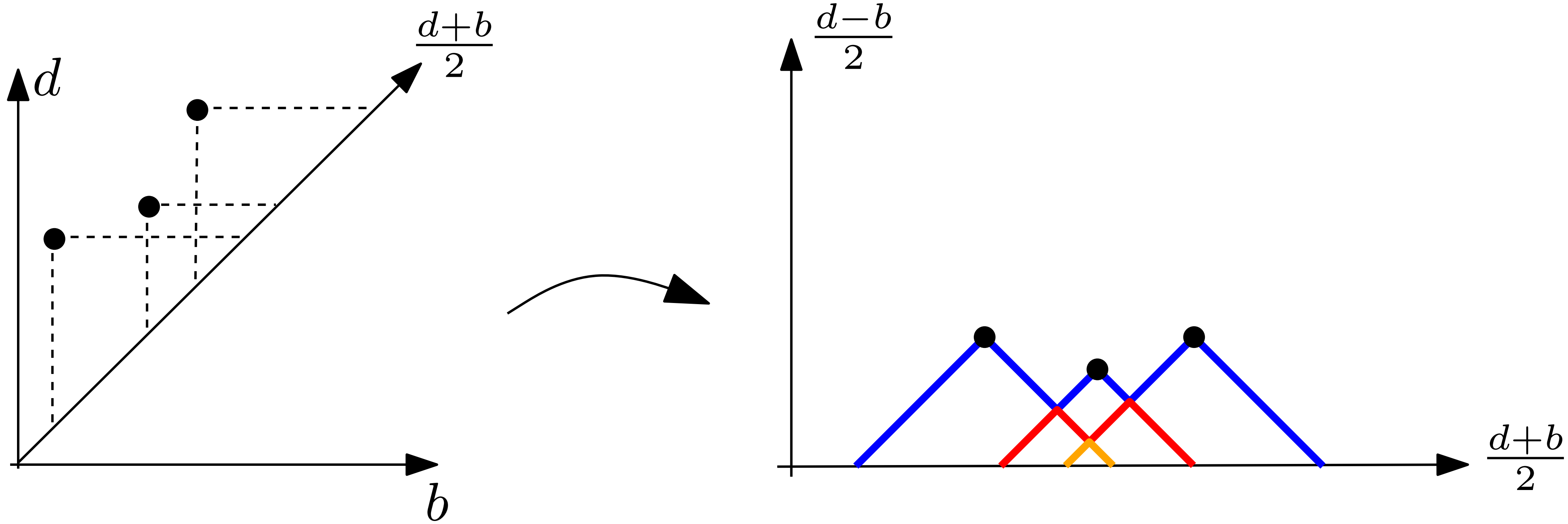
Landscapes



$$D = \{(\tfrac{d_i+b_i}{2}, \tfrac{d_i+b_i}{2})\} i \in I$$

$$(b,d) \quad \mapsto \quad (\tfrac{d+b}{2}, \tfrac{d-b}{2}) \quad \mapsto \quad \Lambda_p(t) \; = \begin{cases} t-b & t \in [b, \frac{b+d}{2}] \\ d-t & t \in (\frac{b+d}{2}, d] \\ 0 & \text{otherwise.} \end{cases}$$

Landscapes



$$D = \{(\tfrac{d_i+b_i}{2}, \tfrac{d_i+b_i}{2})\} i \in I$$

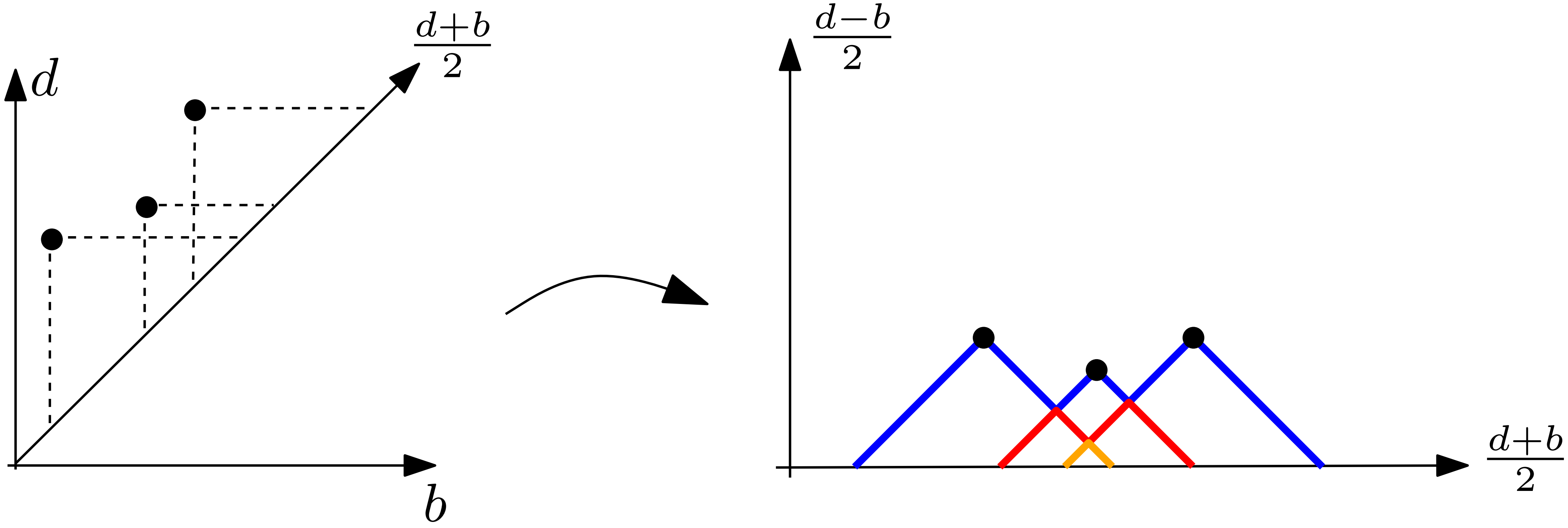
$$(b,d) \quad \mapsto \quad (\tfrac{d+b}{2}, \tfrac{d-b}{2}) \quad \mapsto \quad \Lambda_p(t) \;=\; \begin{cases} t-b & t \in [b, \frac{b+d}{2}] \\ d-t & t \in (\frac{b+d}{2}, d] \\ 0 & \text{otherwise.} \end{cases}$$

Persistence landscape [Bubenik 2012]:

$$\lambda_D(k,t) = \operatorname*{kmax}_{p \in D} \Lambda_p(t), \quad k \in \mathbb{N}, t \in \mathbb{R}$$

where kmax is the kth largest value in the set.

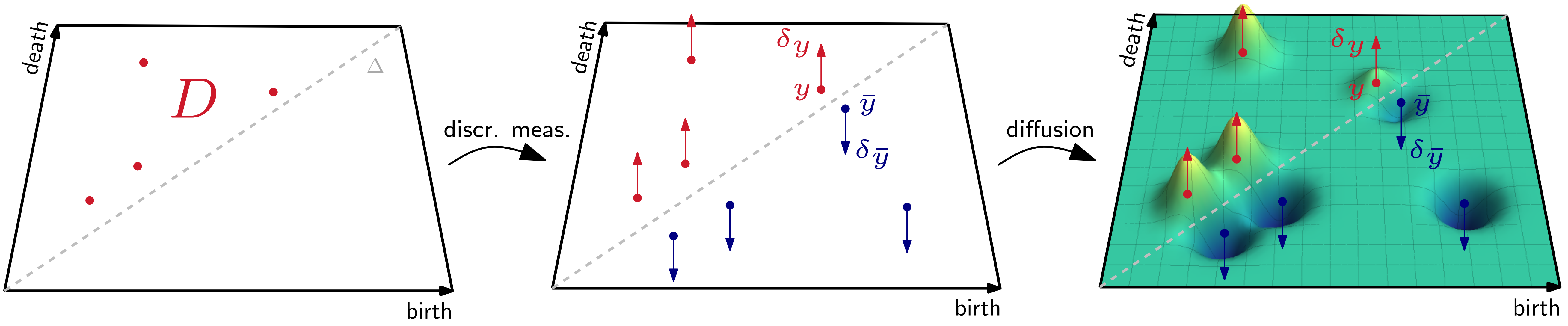
Landscapes



Properties:

- For any $t \in \mathbb{R}$ and $k \in \mathbb{N}$, $\lambda_D(k,t) \geq \lambda_D(k+1,t) \geq 0$.
- $\|\lambda_D - \lambda_{D'}\|_\infty \leq \mathrm{d_B}(D, D')$.
- $\|\lambda_D(1,t) - \lambda_{D'}(1,t)\|_2 \leq C\ W_2(D, D')$.

Discrete measures



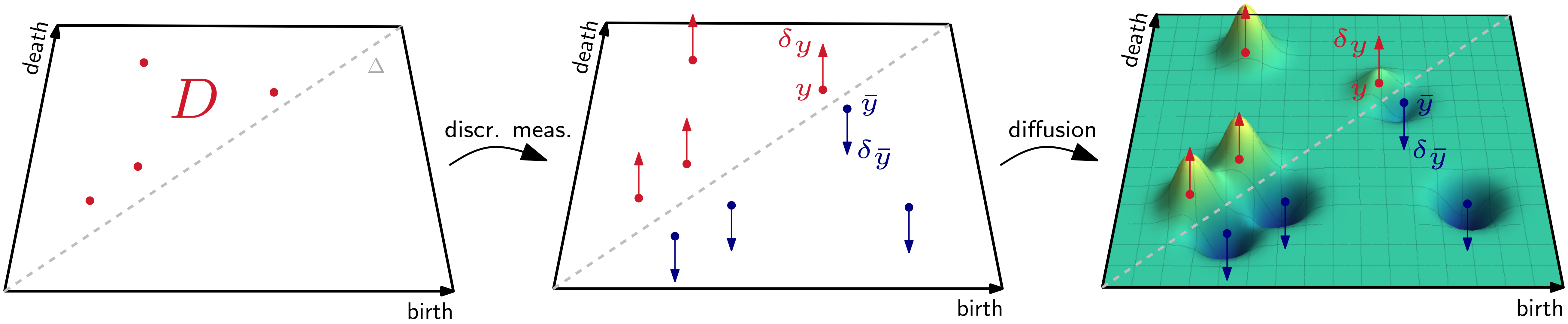
Feature map (solution of heat eq. with Dirichlet cond. on Δ):

$$\phi_k(D)(\cdot) = \frac{1}{4\pi\sigma} \sum_{p \in D} \exp\left(-\frac{\|\cdot - p\|^2}{4\sigma}\right) - \exp\left(-\frac{\|\cdot - \bar{p}\|^2}{4\sigma}\right) \in L^2(\mathbb{R}^2)$$

Persistence Scale Space Kernel [Reininghaus et al. 2015]:

$$k(D, D') = \frac{1}{8\pi\sigma} \sum_{\substack{p \in D \\ q \in D'}} \exp\left(-\frac{\|p - q\|^2}{8\sigma}\right) - \exp\left(-\frac{\|p - \bar{q}\|^2}{8\sigma}\right)$$

Discrete measures



Properties:

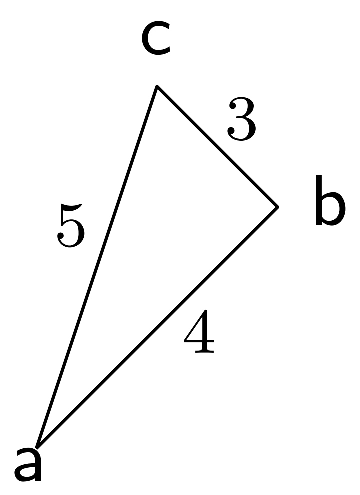
- $\|\phi_k(D) - \phi_k(D')\|_2 \leq \frac{1}{2\sigma\sqrt{2\pi}} W_1(D, D')$.
- $k(D \cup D', D'') = d(D, D'' + k(D', D'')$
- k is injective and $\exp(k)$ is universal

Persistence Scale Space Kernel [Reininghaus et al. 2015]:

$$k(D, D') = \frac{1}{8\pi\sigma} \sum_{\substack{p \in D \\ q \in D'}} \exp\left(-\frac{\|p-q\|^2}{8\sigma}\right) - \exp\left(-\frac{\|p-\bar{q}\|^2}{8\sigma}\right)$$

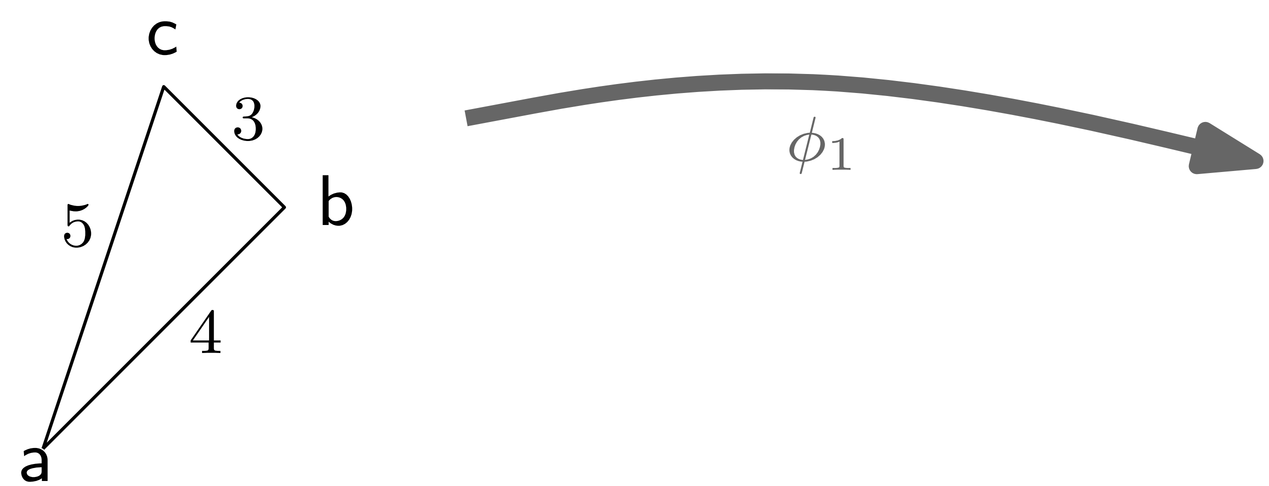
Finite metric spaces... with same cardinality

finite metric space



Finite metric spaces... with same cardinality

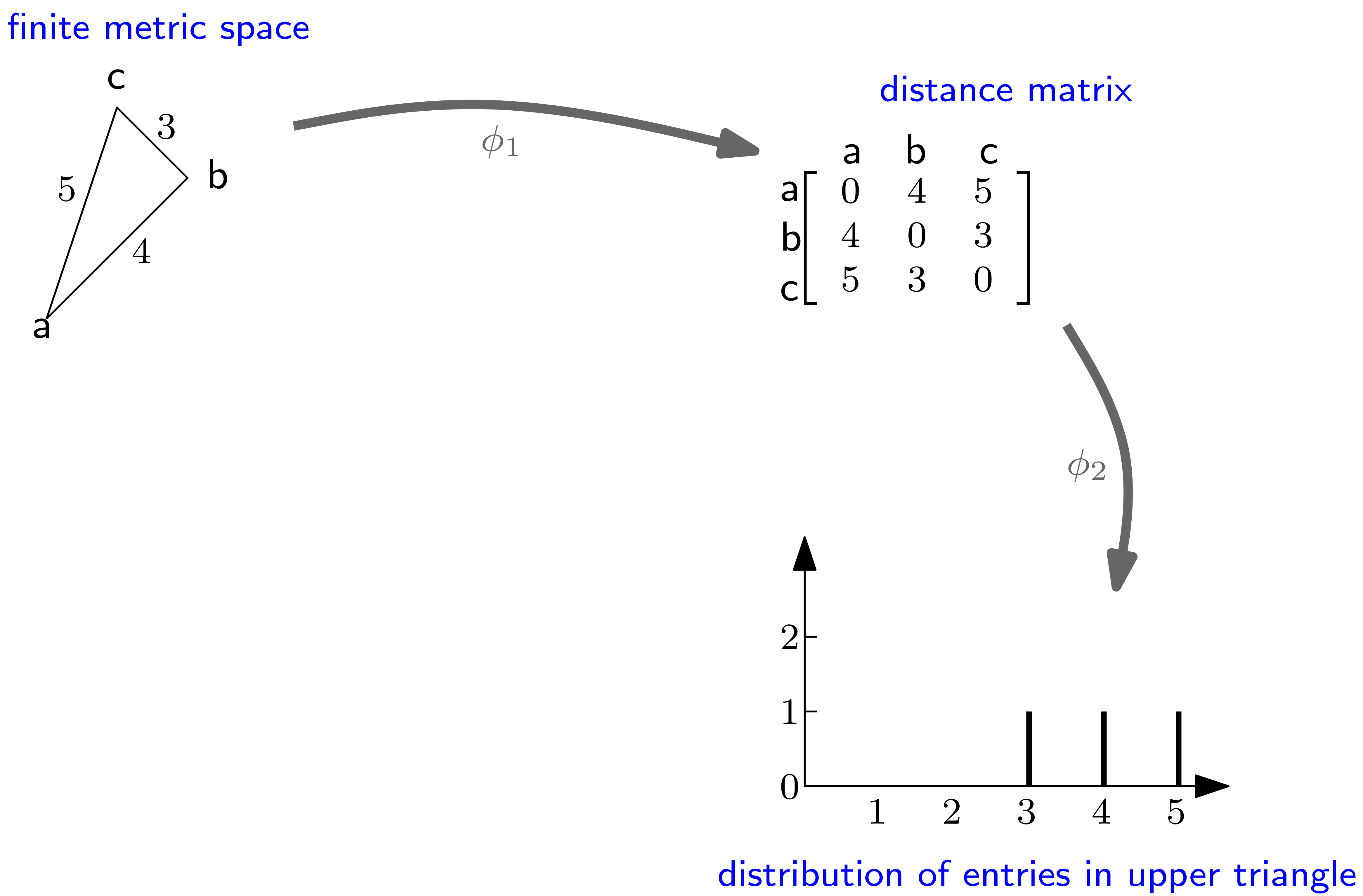
finite metric space



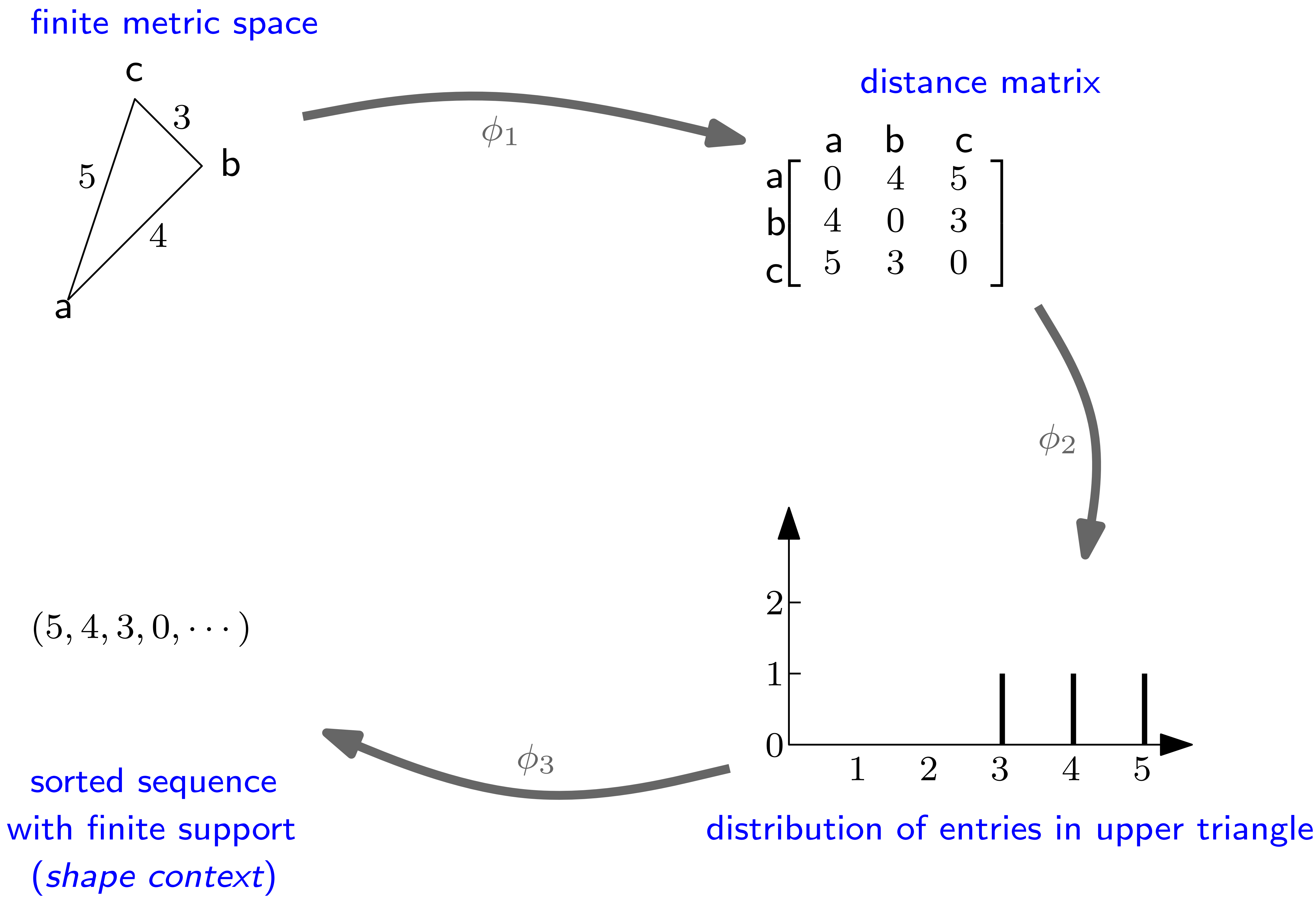
distance matrix

	a	b	c
a	0	4	5
b	4	0	3
c	5	3	0

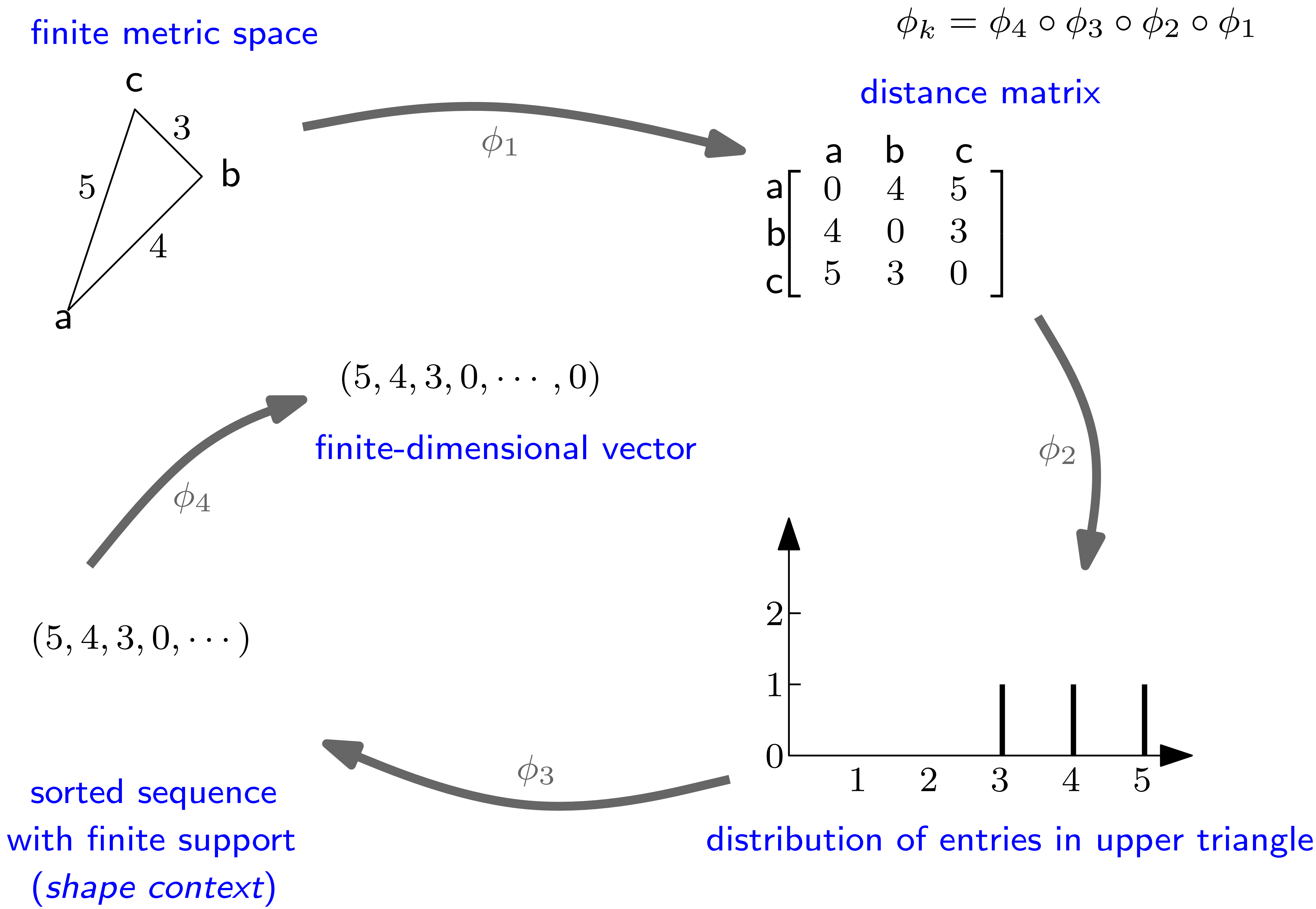
Finite metric spaces... with same cardinality



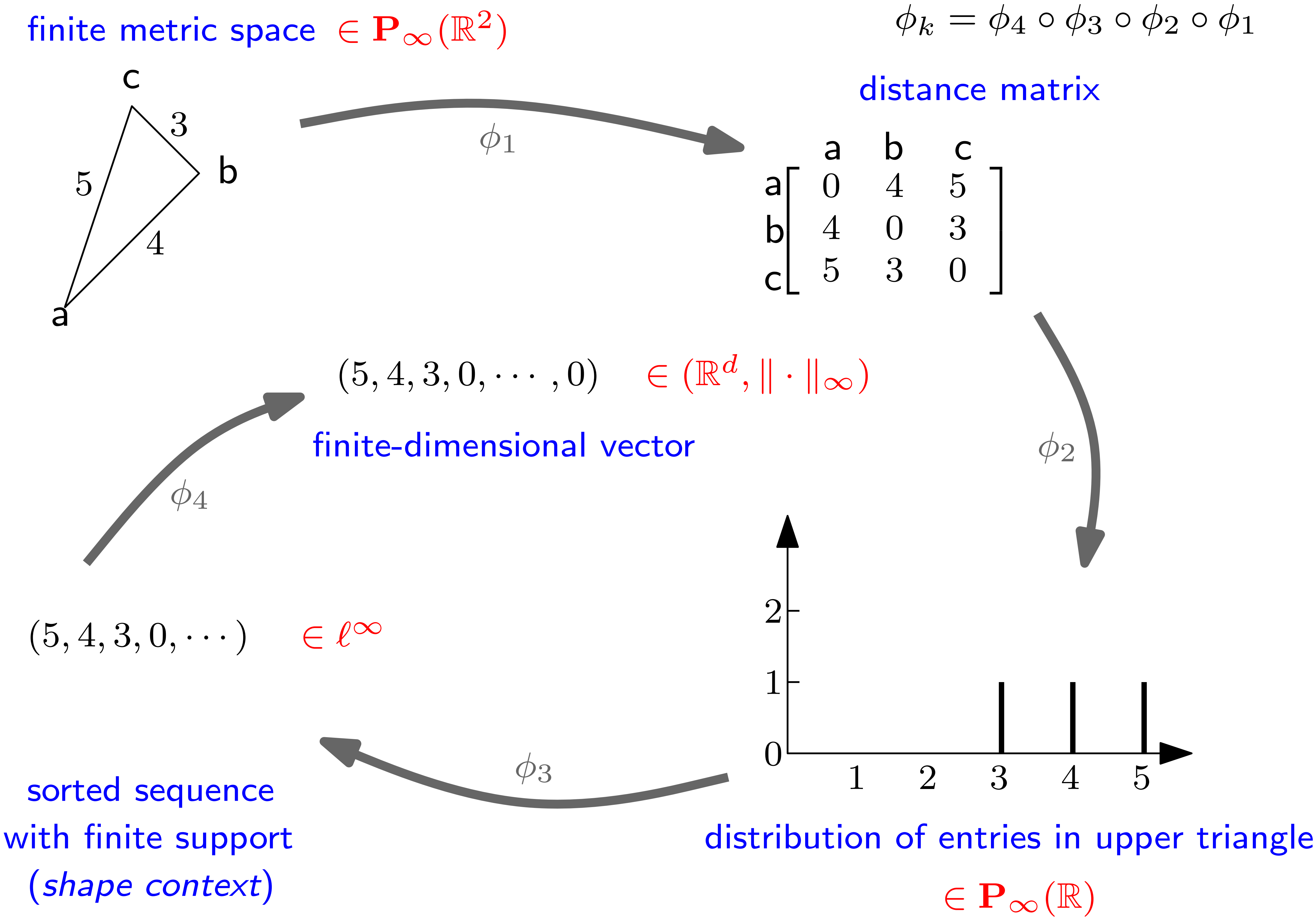
Finite metric spaces... with same cardinality



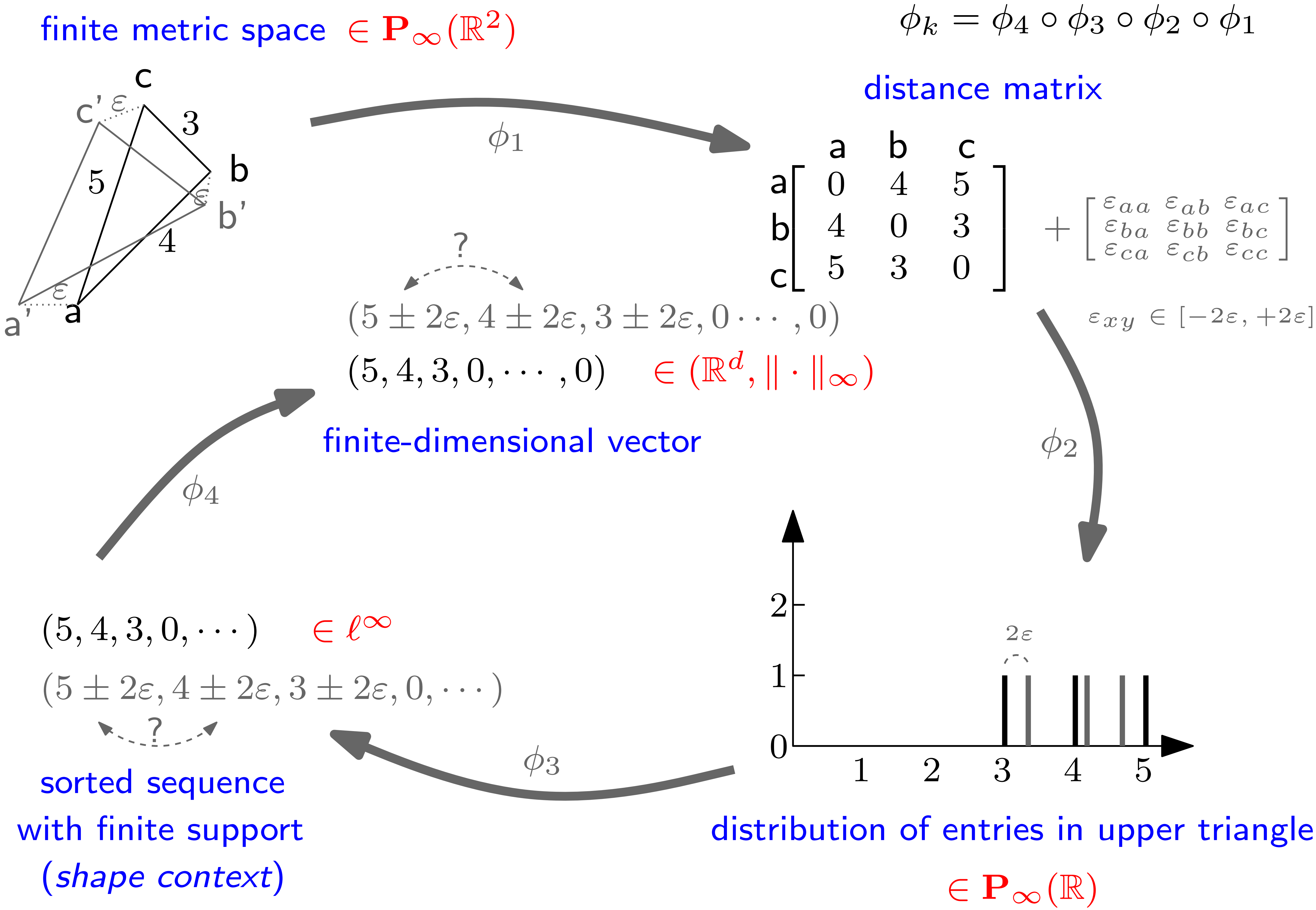
Finite metric spaces... with same cardinality



Finite metric spaces... with same cardinality

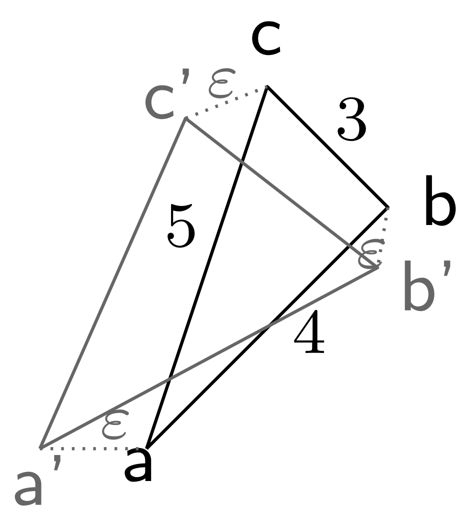


Finite metric spaces... with same cardinality



Finite metric spaces... with same cardinality

finite metric space $\in \mathbf{P}_\infty(\mathbb{R}^2)$



$$\phi_k = \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1$$

ϕ_1

distance matrix

$$\begin{array}{c|ccc} & a & b & c \\ a & 0 & 4 & 5 \\ b & 4 & 0 & 3 \\ c & 5 & 3 & 0 \end{array} + \begin{bmatrix} \varepsilon_{aa} & \varepsilon_{ab} & \varepsilon_{ac} \\ \varepsilon_{ba} & \varepsilon_{bb} & \varepsilon_{bc} \\ \varepsilon_{ca} & \varepsilon_{cb} & \varepsilon_{cc} \end{bmatrix}$$

$\varepsilon_{xy} \in [-2\varepsilon, +2\varepsilon]$

ϕ_2

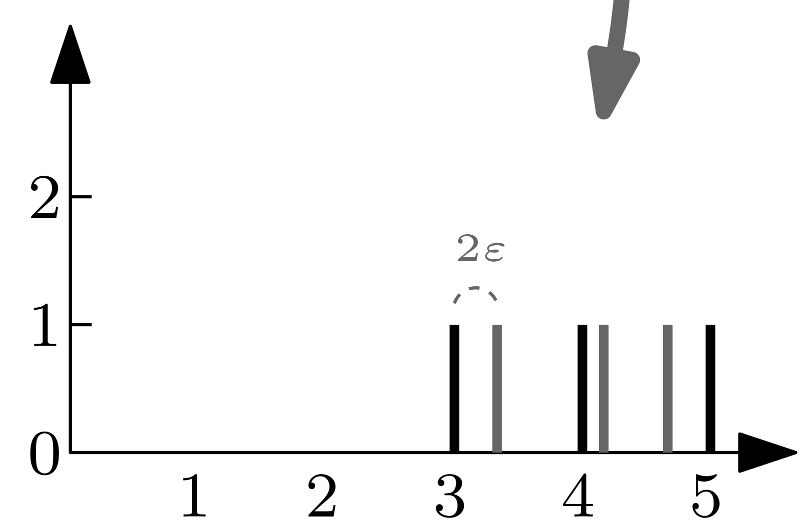
?

$(5 \pm 2\varepsilon, 4 \pm 2\varepsilon)$ (further truncation)

$(5, 4, 3, 0, \cdots, 0) \quad \in (\mathbb{R}^d, \|\cdot\|_\infty)$

finite-dimensional vector

ϕ_4



distribution of entries in upper triangle

$\in \mathbf{P}_\infty(\mathbb{R})$

ϕ_3

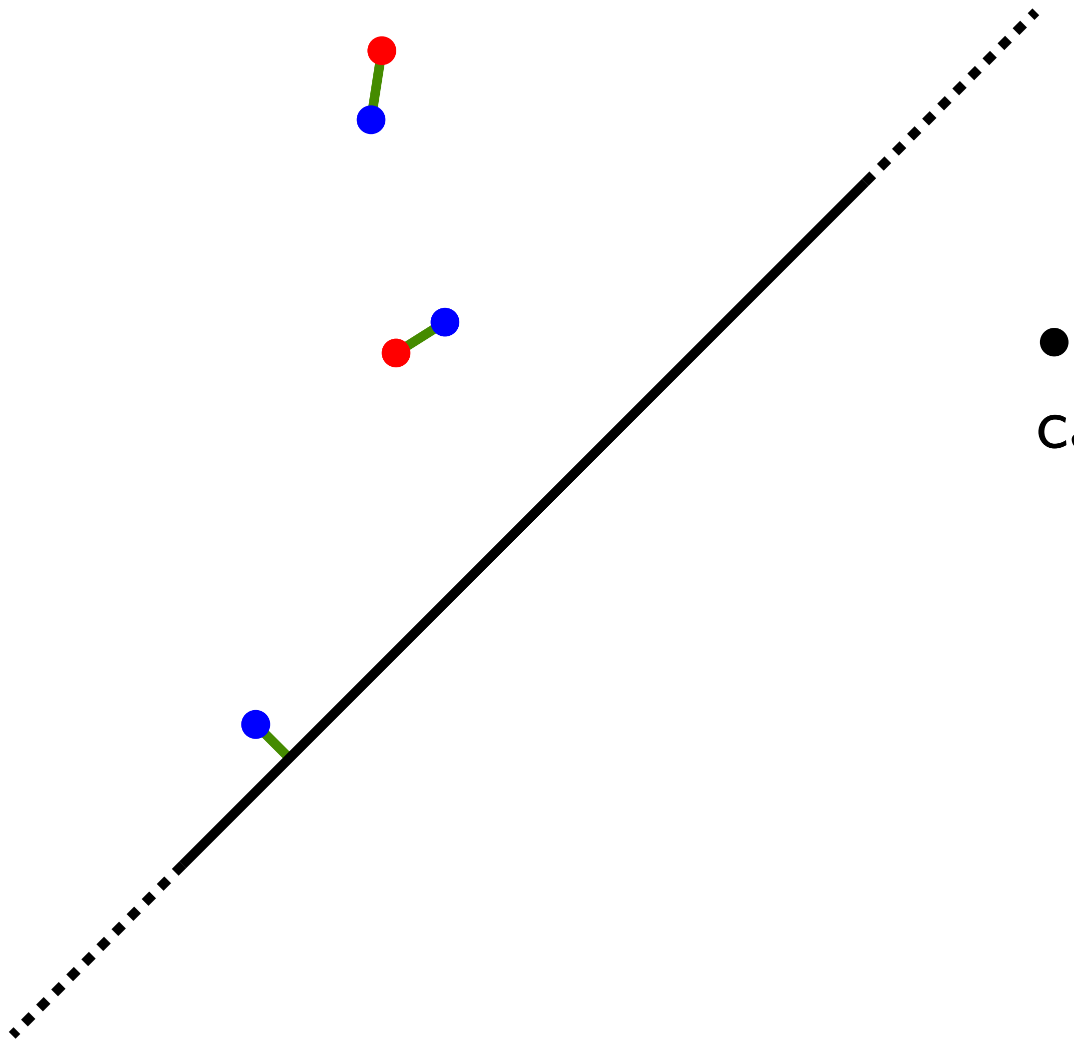
$(5, 4, 3, 0, \cdots) \quad \in \ell^\infty$

$(5 \pm 2\varepsilon, 4 \pm 2\varepsilon, 3 \pm 2\varepsilon, 0, \cdots)$

?

sorted sequence
with finite support
(*shape context*)

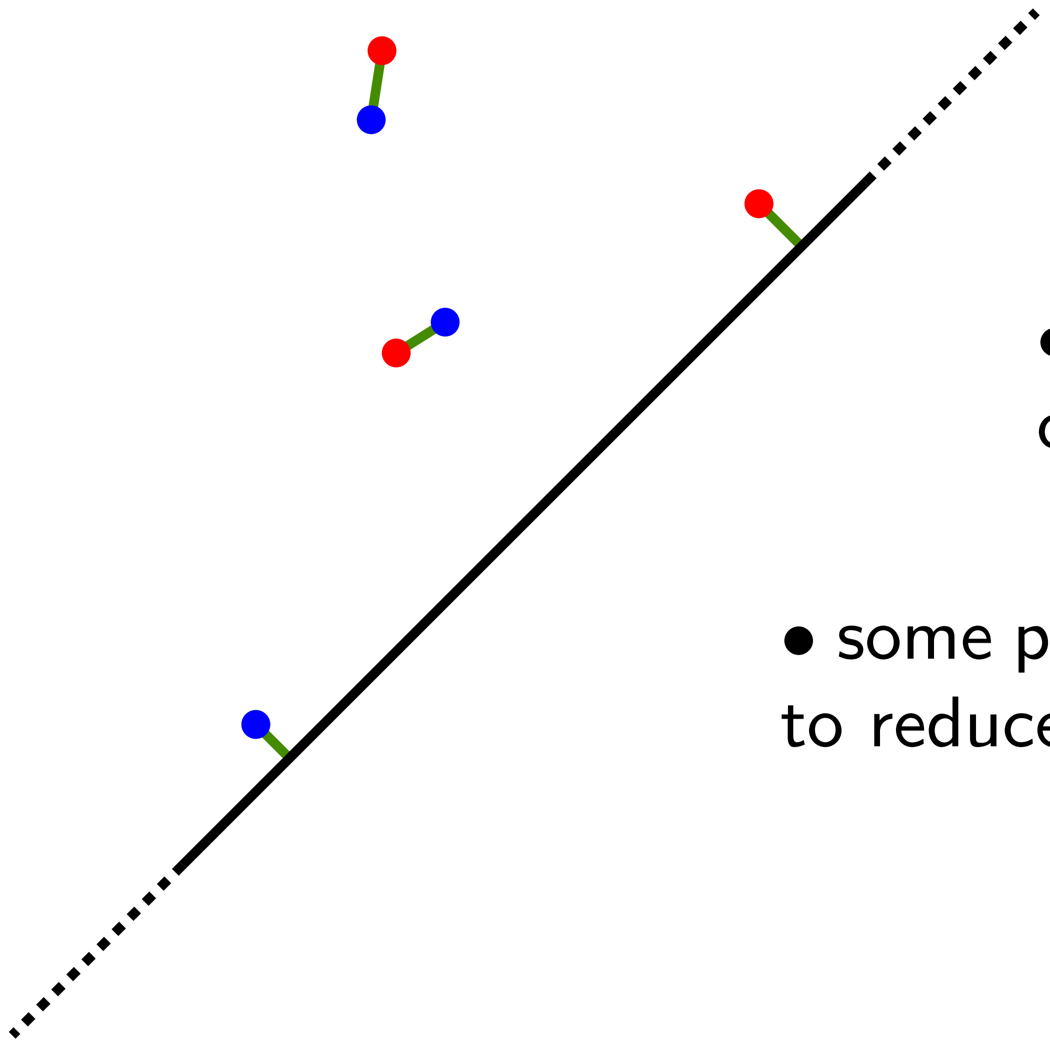
Finite metric spaces... with diagonal



- diagonal has infinite multiplicity

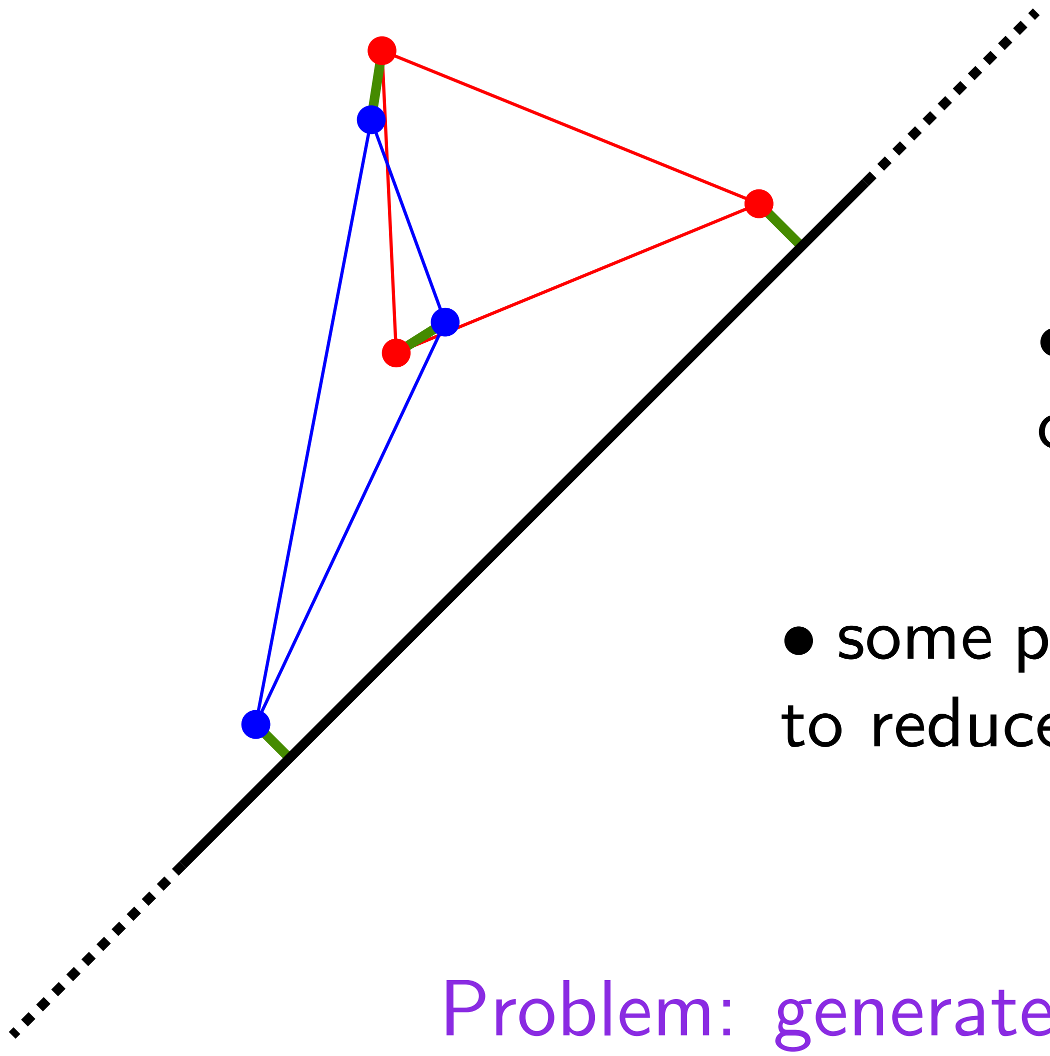
- useful for when point clouds have different cardinalities

Finite metric spaces... with diagonal



- diagonal has infinite multiplicity
- useful for when point clouds have different cardinalities
- some points may prefer the diagonal to other points to reduce the cost of the matching

Finite metric spaces... with diagonal



• diagonal has infinite multiplicity

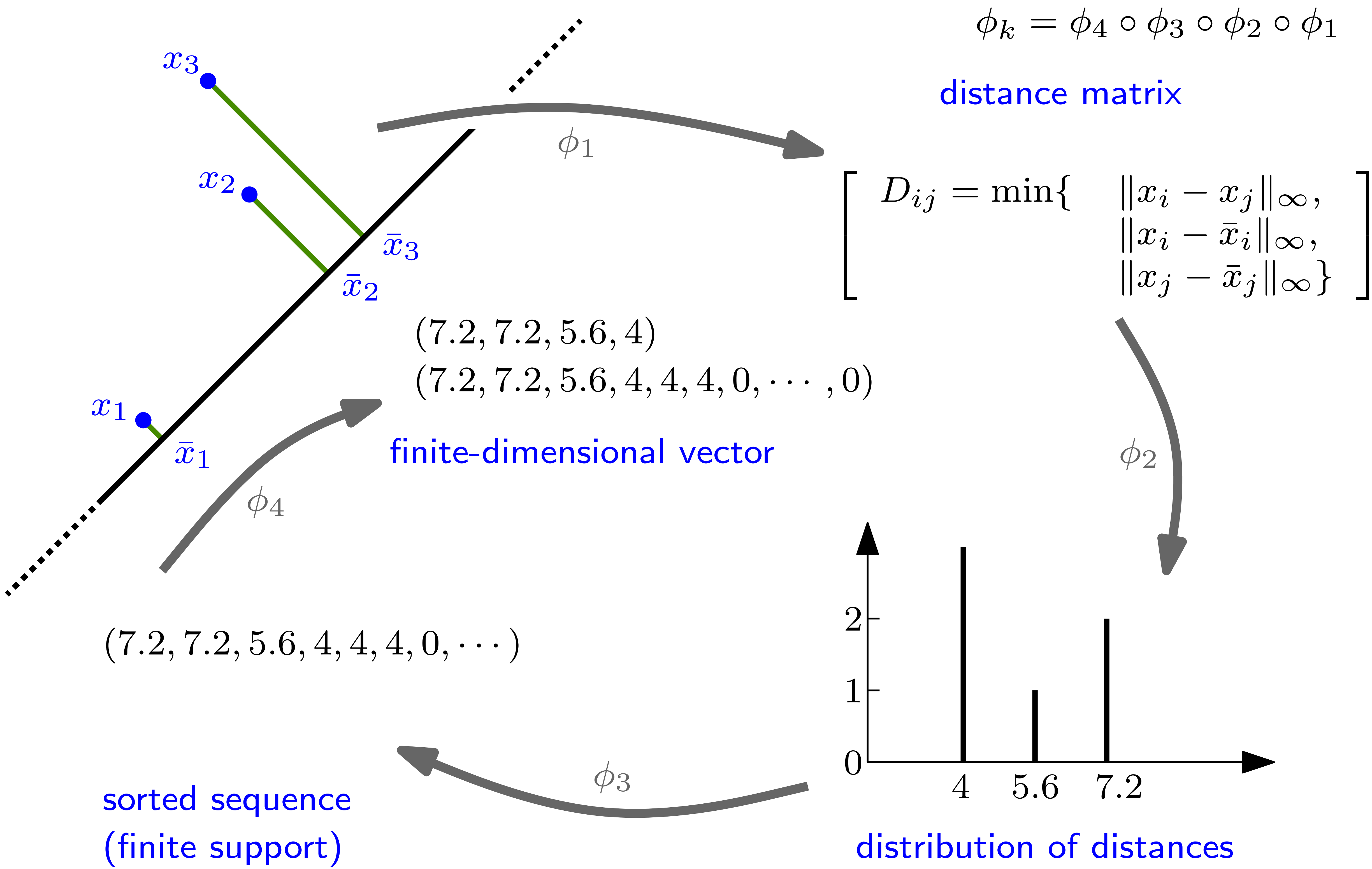
• useful for when point clouds have different cardinalities

• some points may prefer the diagonal to other points to reduce the cost of the matching

Problem: generates instability in distance matrix

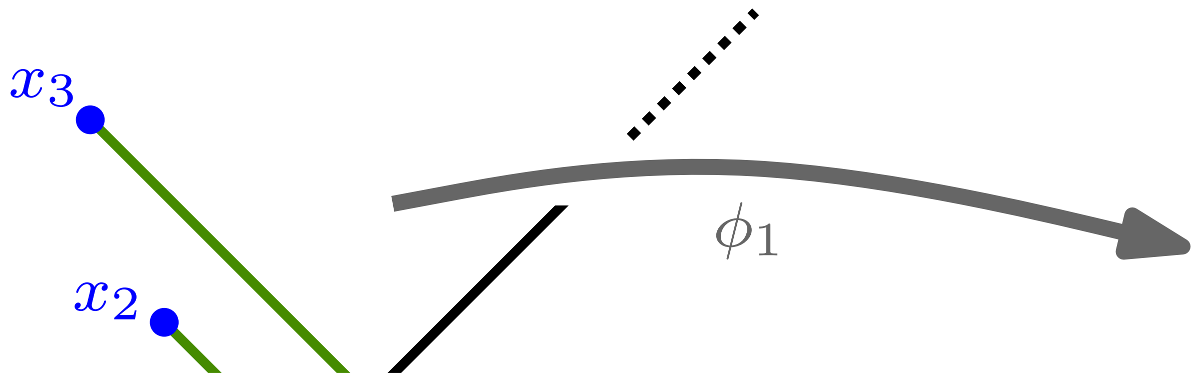
Solution: change the metric

Finite metric spaces... with diagonal



Finite metric spaces... with diagonal

$$\phi_k = \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1$$



distance matrix

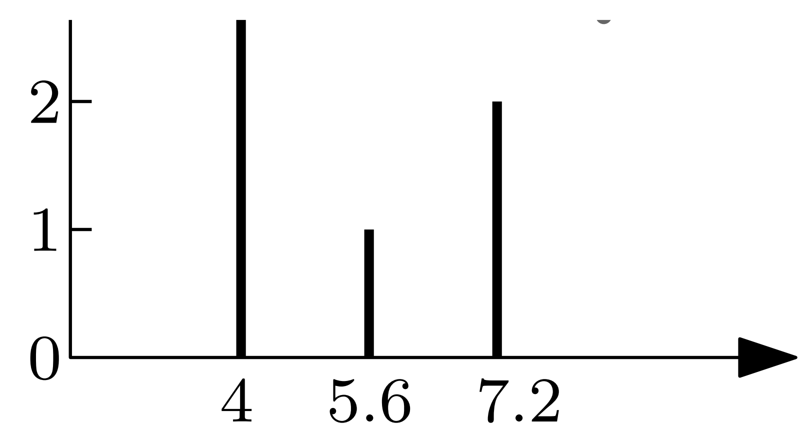
$$\left[\ D_{ij} = \min\{ \quad \|x_i - x_j\|_\infty, \quad \right]$$

Properties: [Carrière et al. 2015]

- RKHS is finite-dimensional ($d < +\infty$)
- $\|\phi_k(D) - \phi_k(D')\|_\infty \leq 2\,\mathrm{d}_\mathrm{B}^\infty(D, D')$
- $\|\phi_k(D) - \phi_k(D')\|_p \leq 2d^{-1/p}\,\mathrm{d}_\mathrm{B}^\infty(D, D')$

$$(7.2, 7.2, 5.6, 4, 4, 4, 0, \cdots)$$

sorted sequence
(finite support)

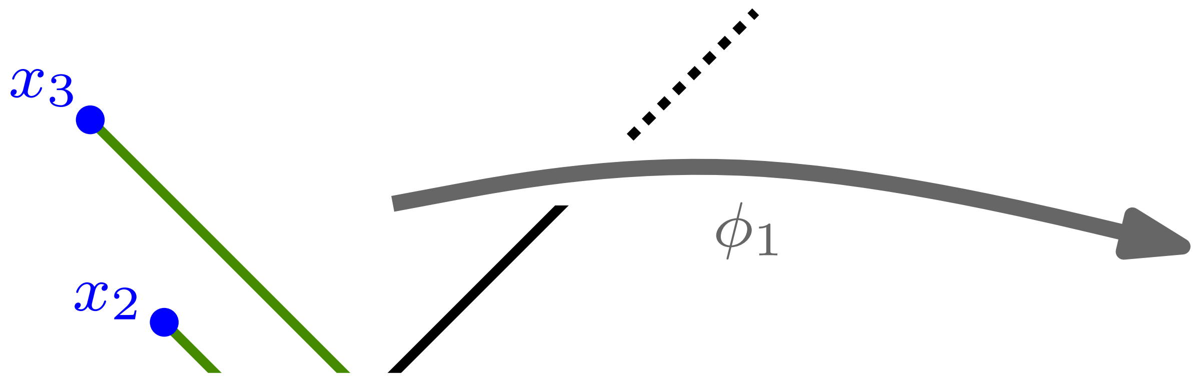


distribution of distances

Finite metric spaces... with diagonal

$$\phi_k = \phi_4 \circ \phi_3 \circ \phi_2 \circ \phi_1$$

distance matrix

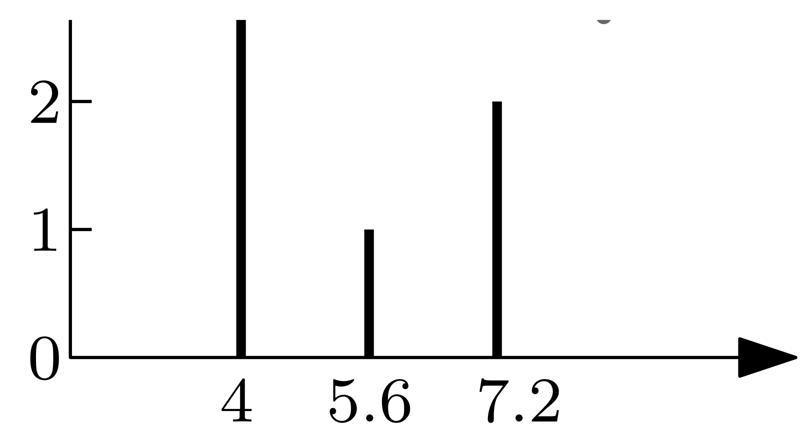


$\left[\; D_{ij} = \min\{ \quad \|x_i - x_j\|_\infty, \quad \right]$

Properties: [Carrière et al. 2015]

- RKHS is finite-dimensional ($d < +\infty$)
- $\|\phi_k(D) - \phi_k(D')\|_\infty \leq 2\, \mathrm{d}_{\mathrm{B}}^{\infty}(D, D')$ ← for NN-classif.
- $\|\phi_k(D) - \phi_k(D')\|_p \leq 2d^{-1/p}\, \mathrm{d}_{\mathrm{B}}^{\infty}(D, D')$ ← for linear classif.

$(7.2, 7.2, 5.6, 4, 4, 4, 0, \cdots)$



sorted sequence
(finite support)

distribution of distances

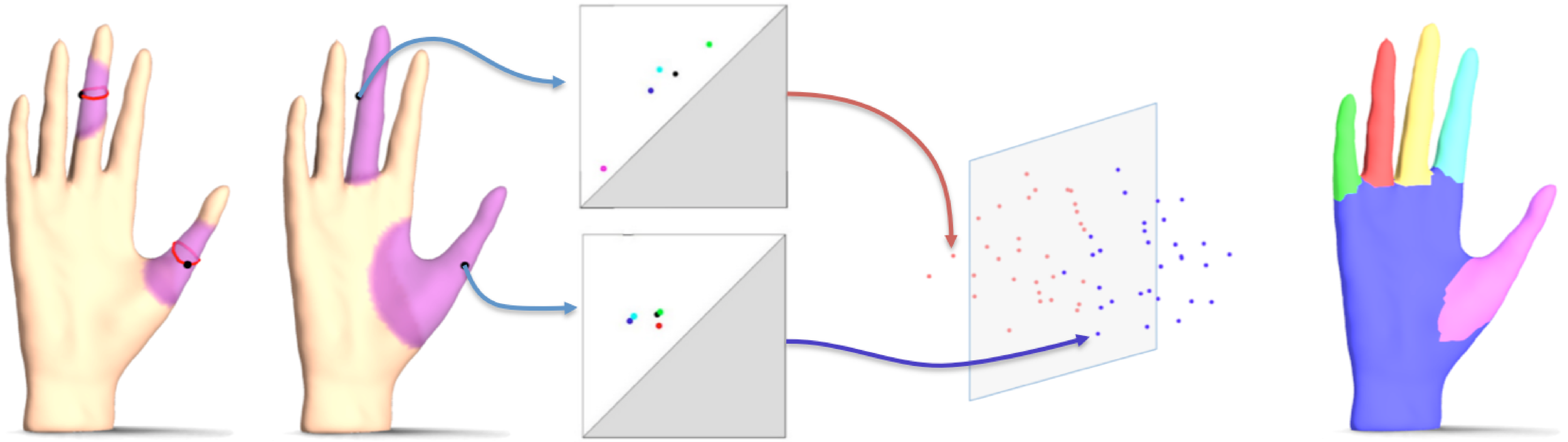
Application to a supervised learning task

Goal: segment 3d shapes based on examples

Approach:

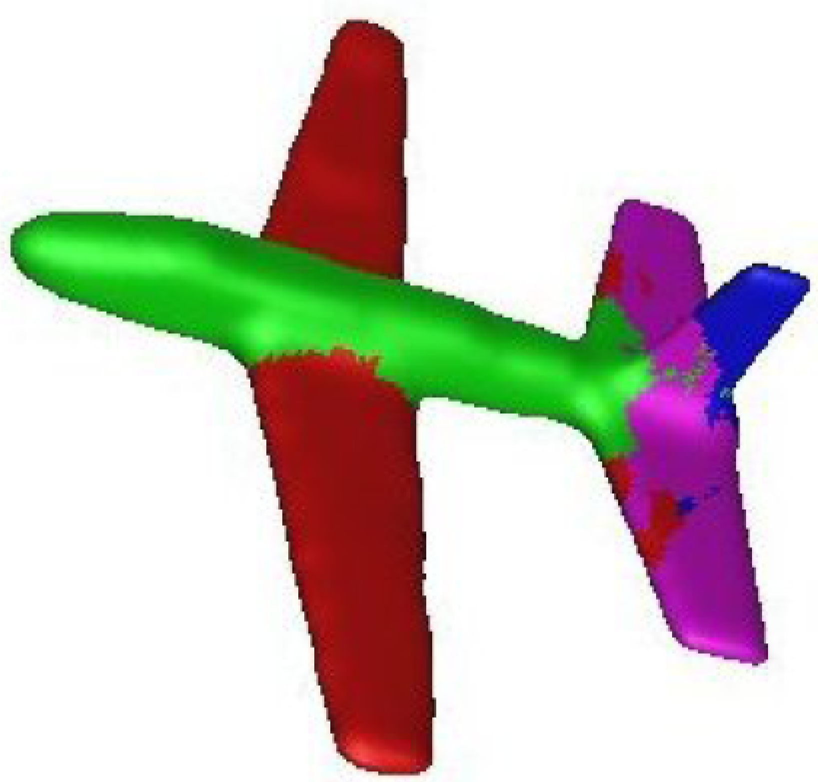
- train a (multiclass) classifier on PDs extracted from the training shapes

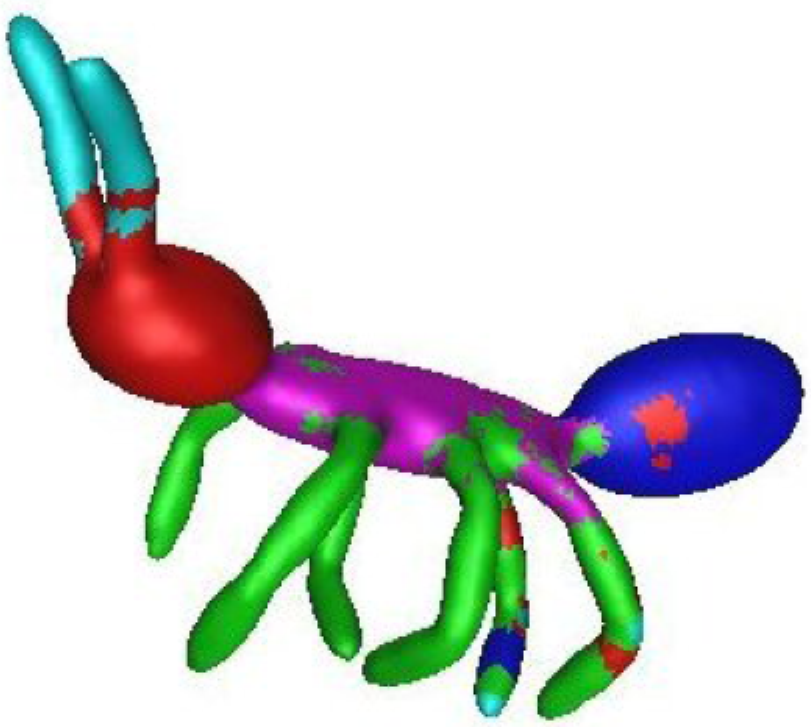
- apply classifier to PDs extracted from query shape



Application to a supervised learning task

Strategy 1: use k-NN classifier in feature space $(\mathbb{R}^d, \|\cdot\|_\infty)$



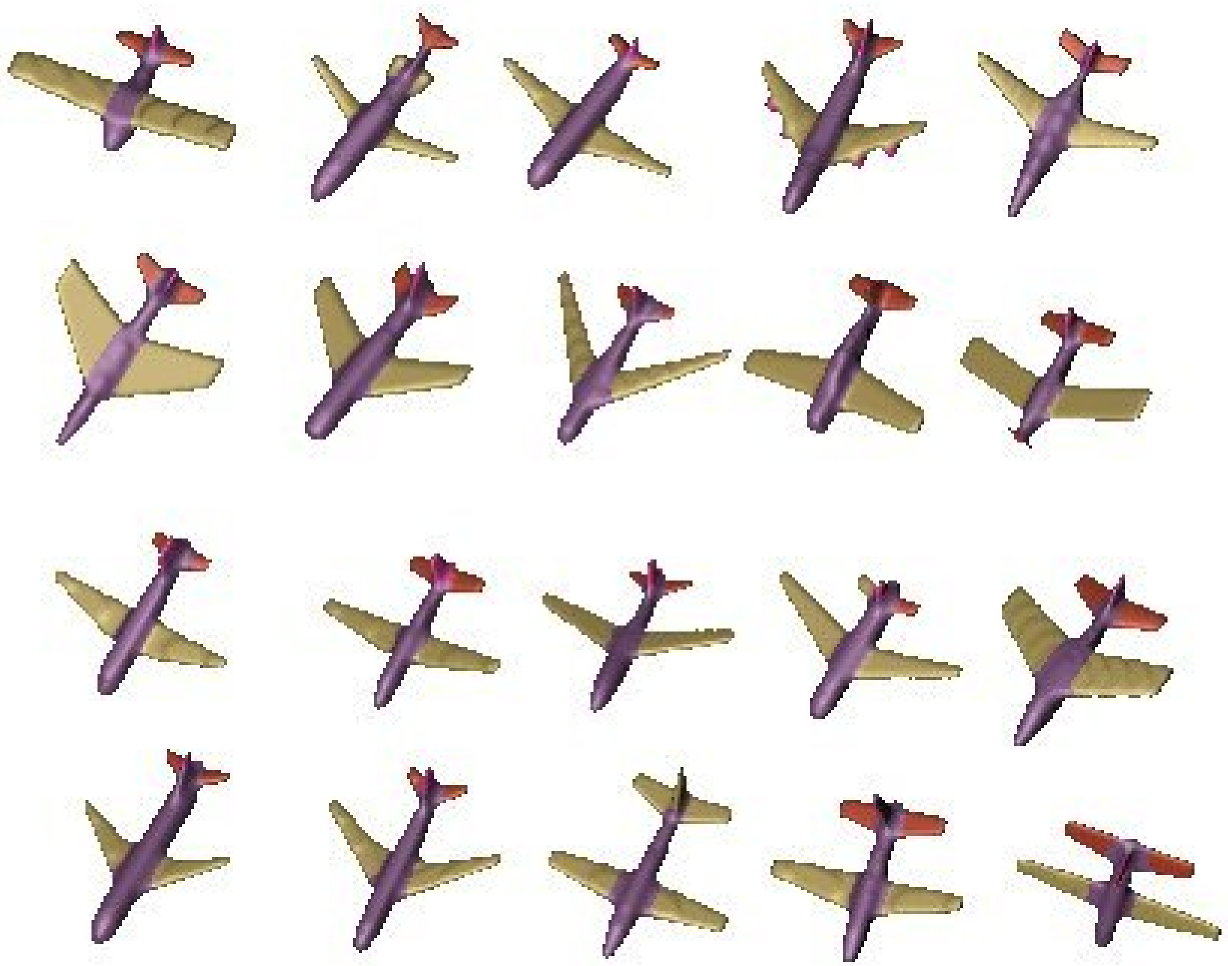


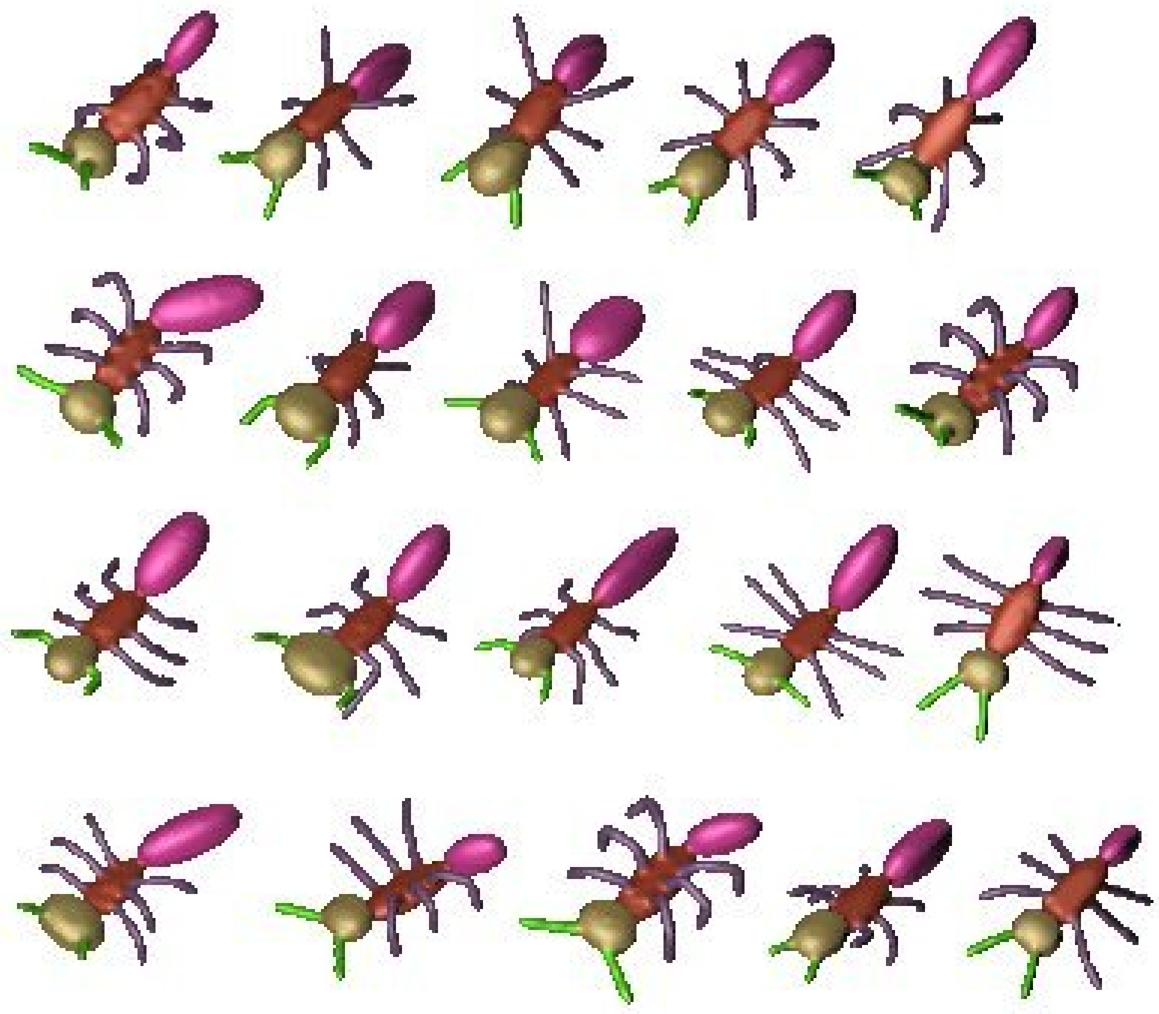
Application to a supervised learning task

Strategy 1: use k-NN classifier in feature space $(\mathbb{R}^d, \|\cdot\|_\infty)$

Strategy 2: use linear classifier (SVM) in feature space $(\mathbb{R}^d, \|\cdot\|_2)$

$+$ graph cut [Kalogerakis et al. 2010]





Application to a supervised learning task

Strategy 1: use k-NN classifier in feature space $(\mathbb{R}^d, \|\cdot\|_\infty)$

Strategy 2: use linear classifier (SVM) in feature space $(\mathbb{R}^d, \|\cdot\|_2)$

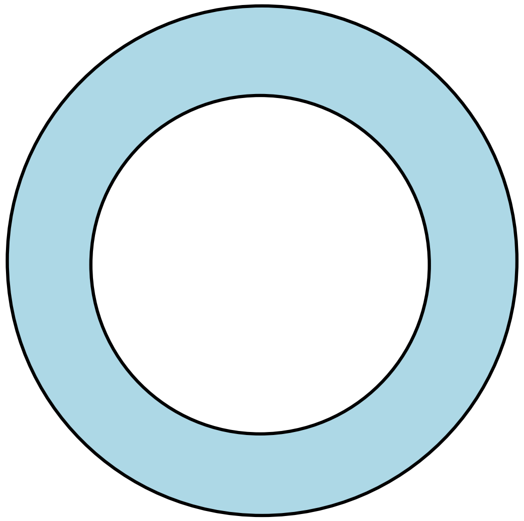
$+$ graph cut [Kalogerakis et al. 2010]

	SB5	SB5+PDs
Human	21.3	**11.3**
Cup	10.6	**10.1**
Glasses	21.8	**25.0**
Airplane	18.7	**9.3**
Ant	9.7	**1.5**
Chair	15.1	**7.3**
Octopus	5.5	**3.4**
Table	7.4	**2.5**
Teddy	6.0	**3.5**
Hand	21.1	**12.0**

	SB5	SB5+PDs
Plier	12.3	**9.2**
Fish	20.9	**7.7**
Bird	24.8	**13.5**
Armadillo	18.4	**8.3**
Bust	35.4	**22.0**
Mech	22.7	**17.0**
Bearing	25.0	**11.2**
Vase	26.4	**17.8**
FourLeg	25.6	**15.8**

percentage of mislabelling ($100-$rand index)

2. Statistics via push-forwards



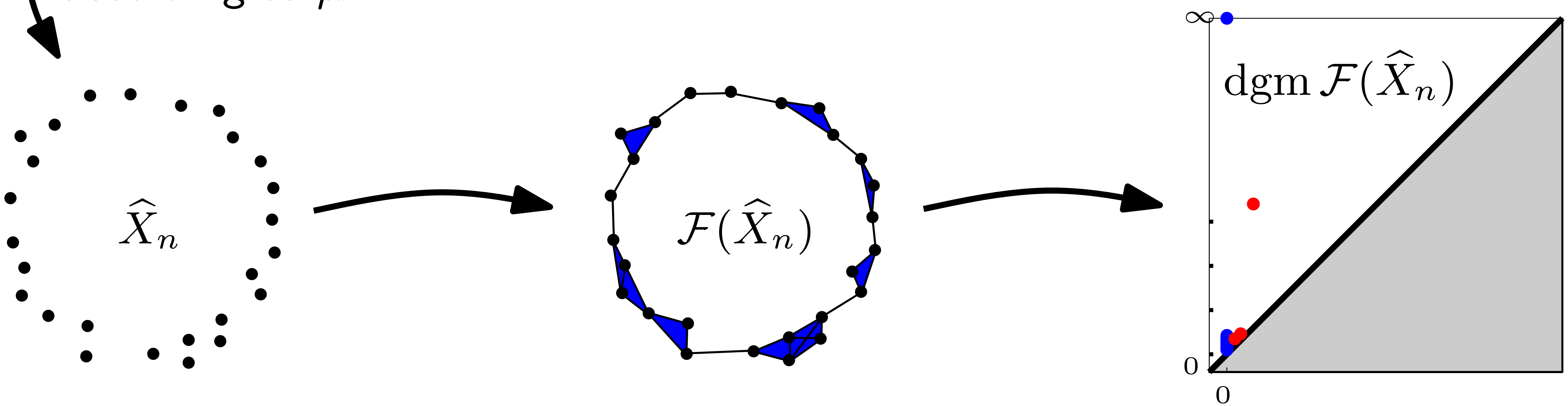
(X, d_X) compact metric space

μ probability measure supported on X $(\operatorname{supp}\mu = X)$

Sample n points iid according to μ.

Examples:

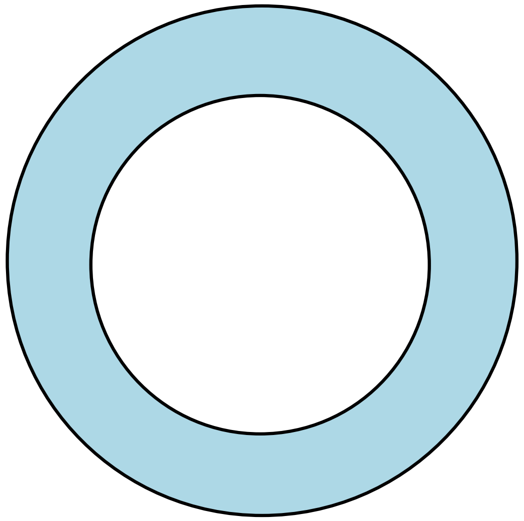
- $\mathcal{F}(\widehat{X}_n) = \mathcal{R}(\widehat{X}_n, \mathrm{d}_X)$
- $\mathcal{F}(\widehat{X}_n) = \mathcal{C}(\widehat{X}_n, \mathrm{d}_X)$
- $\cdots$



Questions:

- Statistical properties of the estimator $\operatorname{dgm}\mathcal{F}(\widehat{X}_n)$?
- Convergence to the ground truth $\operatorname{dgm}\mathcal{F}(X)$? Deviation bounds?

2. Statistics via push-forwards



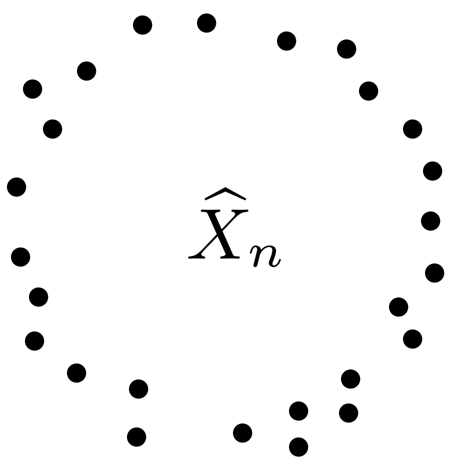
(X, d_X) compact metric space

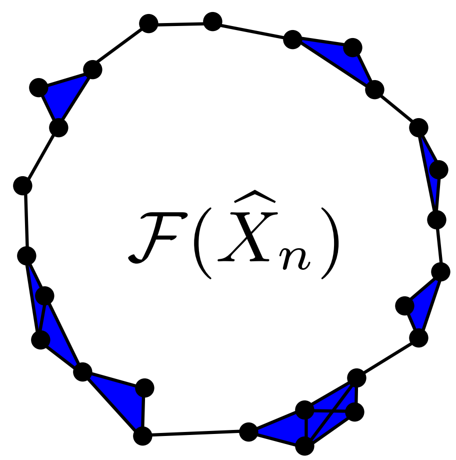
μ probability measure supported on X $(\operatorname{supp}\mu = X)$

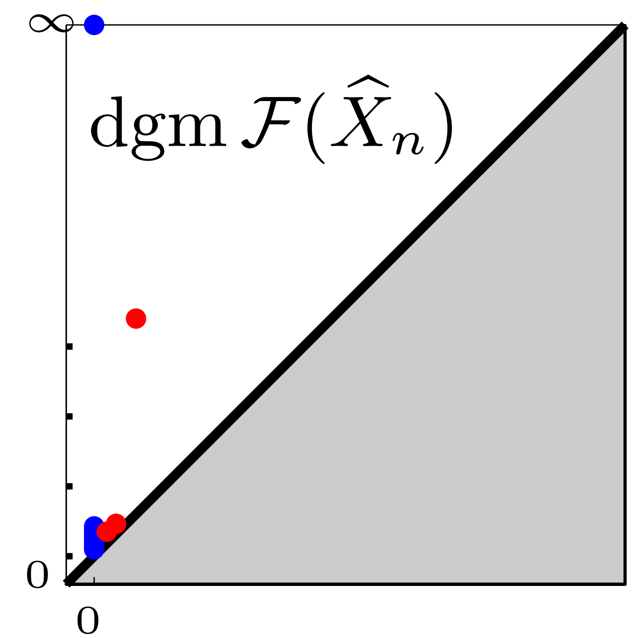
Sample n points iid according to μ.

Examples:

- $\mathcal{F}(\widehat{X}_n) = \mathcal{R}(\widehat{X}_n, \mathrm{d}_X)$
- $\mathcal{F}(\widehat{X}_n) = \mathcal{C}(\widehat{X}_n, \mathrm{d}_X)$
- $\cdots$





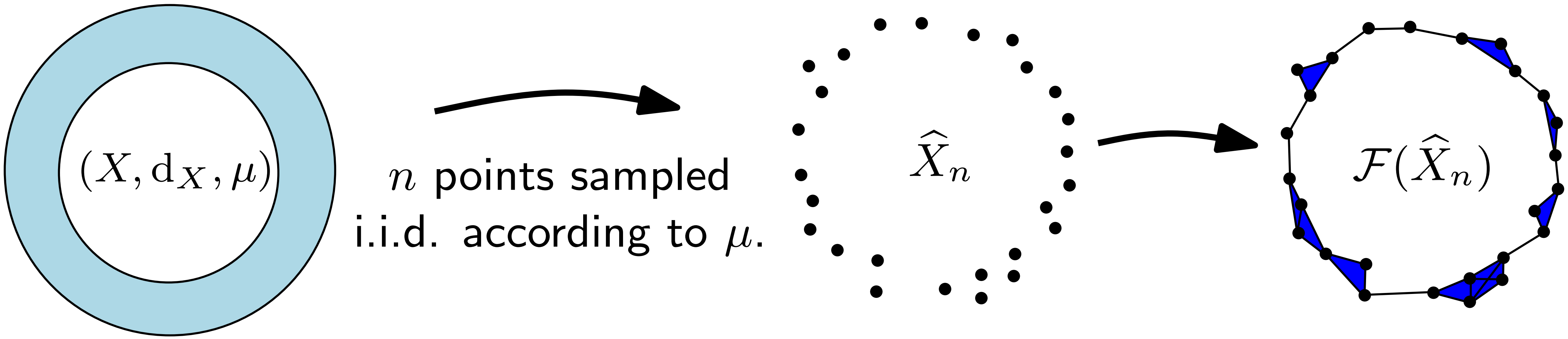


Stability thm: $\mathrm{d}_\mathrm{B}(\operatorname{dgm}\mathcal{F}(\widehat{X}_n), \operatorname{dgm}\mathcal{F}(X)) \leq 2\mathrm{d}_\mathrm{H}(\widehat{X}_n, X)$

$\Rightarrow$ for any $\varepsilon > 0$,

$$\mathbb{P}\left(\mathrm{d}_\mathrm{B}\left(\operatorname{dgm}\mathcal{F}(\widehat{X}_n), \operatorname{dgm}\mathcal{F}(X),\right) > \varepsilon\right) \leq \mathbb{P}\left(\mathrm{d}_\mathrm{H}(\widehat{X}_n, X) > \frac{\varepsilon}{2}\right)$$

Deviation inequality



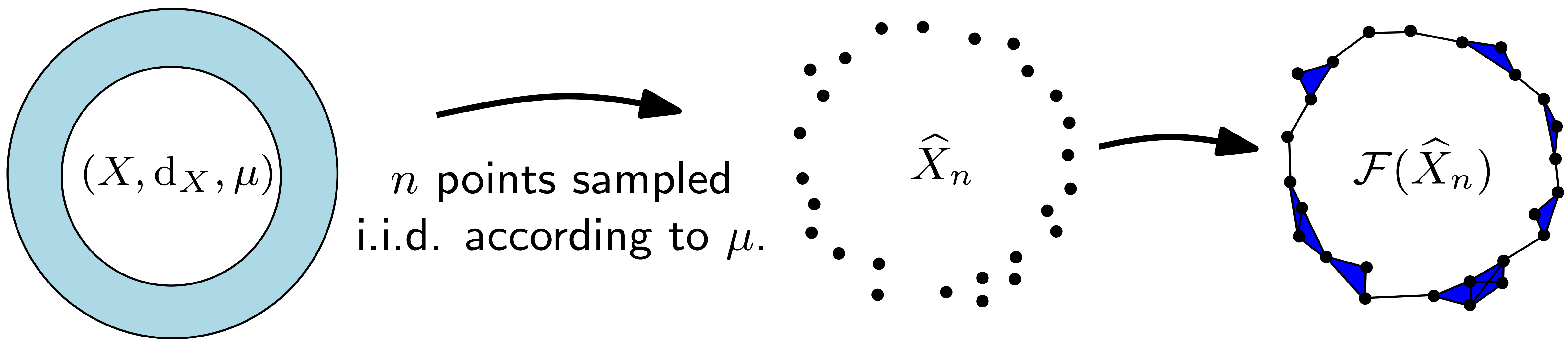
For $a, b > 0$, μ satisfies the (a,b)-**standard** assumption if for any $x \in X$ and any $r > 0$, we have $\mu(B(x,r)) \geq \min(ar^b, 1)$.

Theorem [Chazal, Glisse, Labruère, Michel 2014-15]:

If μ is (a,b)-standard then for any $\varepsilon > 0$:

$$\mathbb{P}\left(\mathrm{d_B}\left(\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n), \mathrm{dgm}\,\mathcal{F}(X)\right) > \varepsilon\right) \leq \frac{8^b}{a\varepsilon^b}\exp(-na\varepsilon^b)$$

Deviation inequality



For $a, b > 0$, μ satisfies the (a,b)-**standard** assumption if for any $x \in X$ and any $r > 0$, we have $\mu(B(x,r)) \geq \min(ar^b, 1)$.

Theorem [Chazal, Glisse, Labruère, Michel 2014-15]:

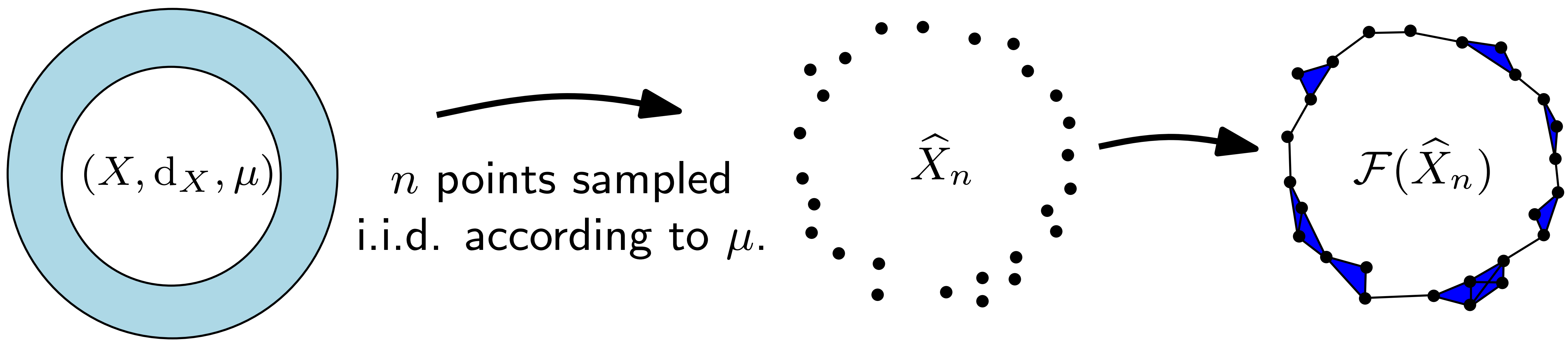
If μ is (a,b)-standard then for any $\varepsilon > 0$:

$$\mathbb{P}\left(\mathrm{d_B}\left(\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n), \mathrm{dgm}\,\mathcal{F}(X)\right) > \varepsilon\right) \leq \frac{8^b}{a\varepsilon^b}\exp(-na\varepsilon^b)$$

Proof sketch:

1. upper-bound $\mathbb{P}\left(\mathrm{d_H}(\widehat{X}_n, X) > \frac{\varepsilon}{2}\right)$.
2. (a,b) standard assumption $\Rightarrow$ explicit upper bound for the covering number of X (by balls of radius $\varepsilon/2$).
3. Apply "union bound" argument.

Deviation inequality / rate of convergence



For $a, b > 0$, μ satisfies the (a, b)-**standard** assumption if for any $x \in X$ and any $r > 0$, we have $\mu(B(x, r)) \geq \min(ar^b, 1)$.

Theorem [Chazal, Glisse, Labruère, Michel 2014-15]:

If μ is (a, b)-standard then for any $\varepsilon > 0$:

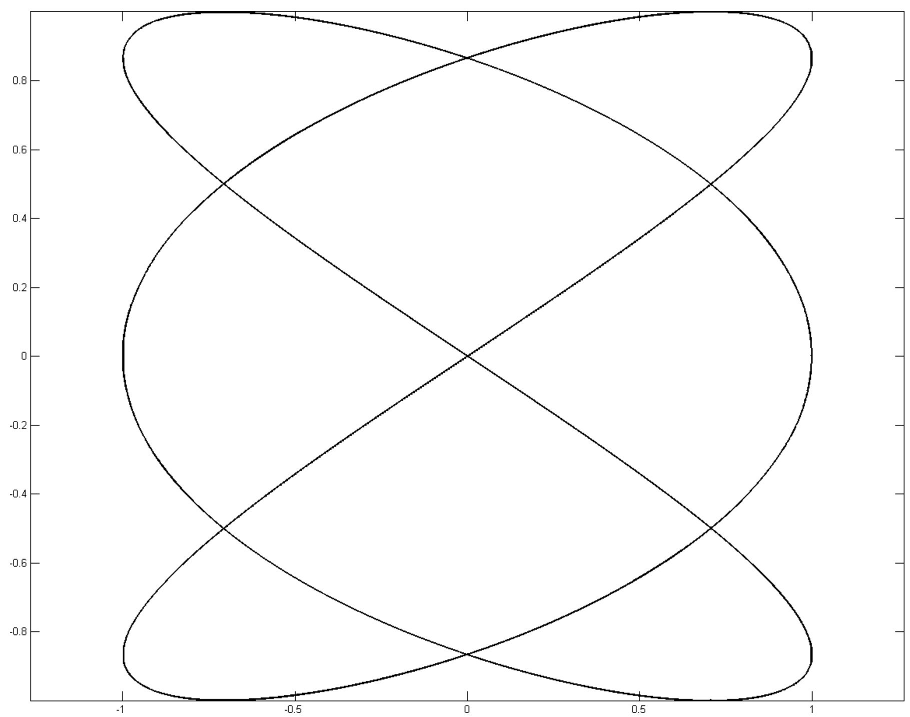
$$\mathbb{P}\left(\mathrm{d_B}\left(\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n), \mathrm{dgm}\,\mathcal{F}(X)\right) > \varepsilon\right) \leq \frac{8^b}{a\varepsilon^b} \exp(-na\varepsilon^b)$$

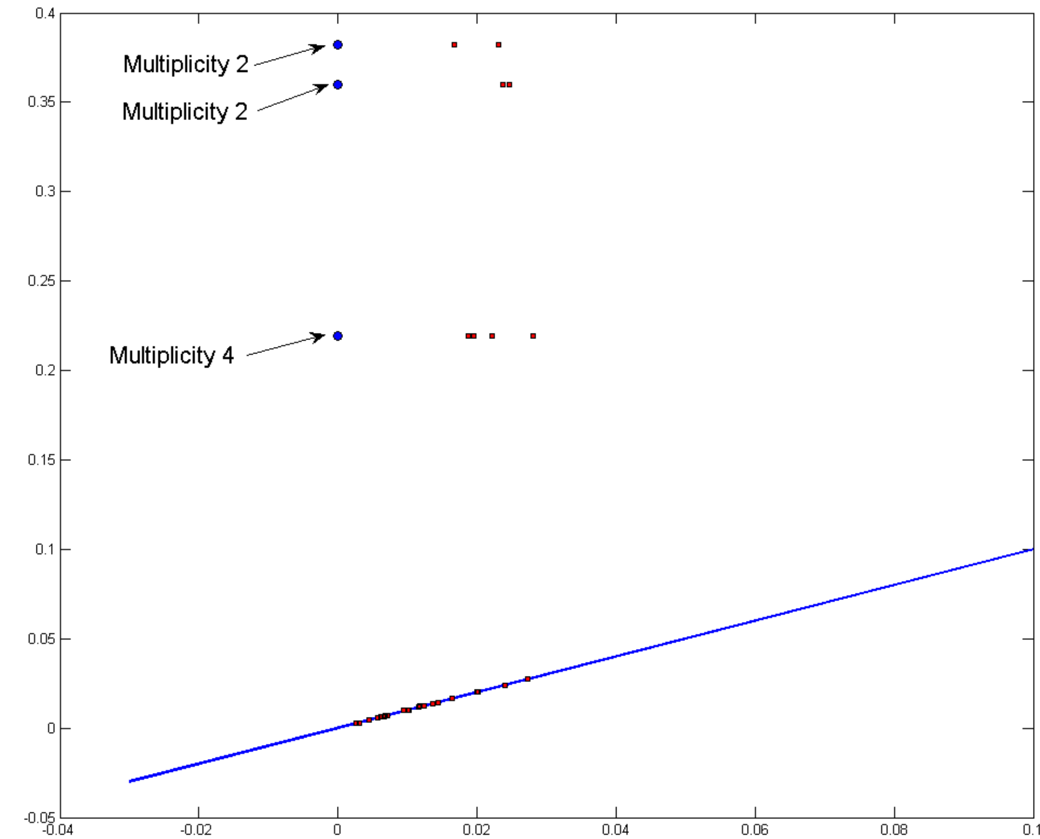
Corollary [Chazal, Glisse, Labruère, Michel 2014]:

$$\sup_{\mu \in \mathcal{P}} \mathbb{E}\left[\mathrm{d_B}\left(\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n),\ \mathrm{dgm}\,\mathcal{F}(X)\right)\right] \leq C\ \left(\frac{\log n}{n}\right)^{1/b},$$

where C depends only on a, b. Moreover, the estimator $\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n)$ is **minimax optimal** (up to a $\log n$ factor) on the space $\mathcal{P}$ of (a, b)-standard probability measures on X.

Numerical illustrations

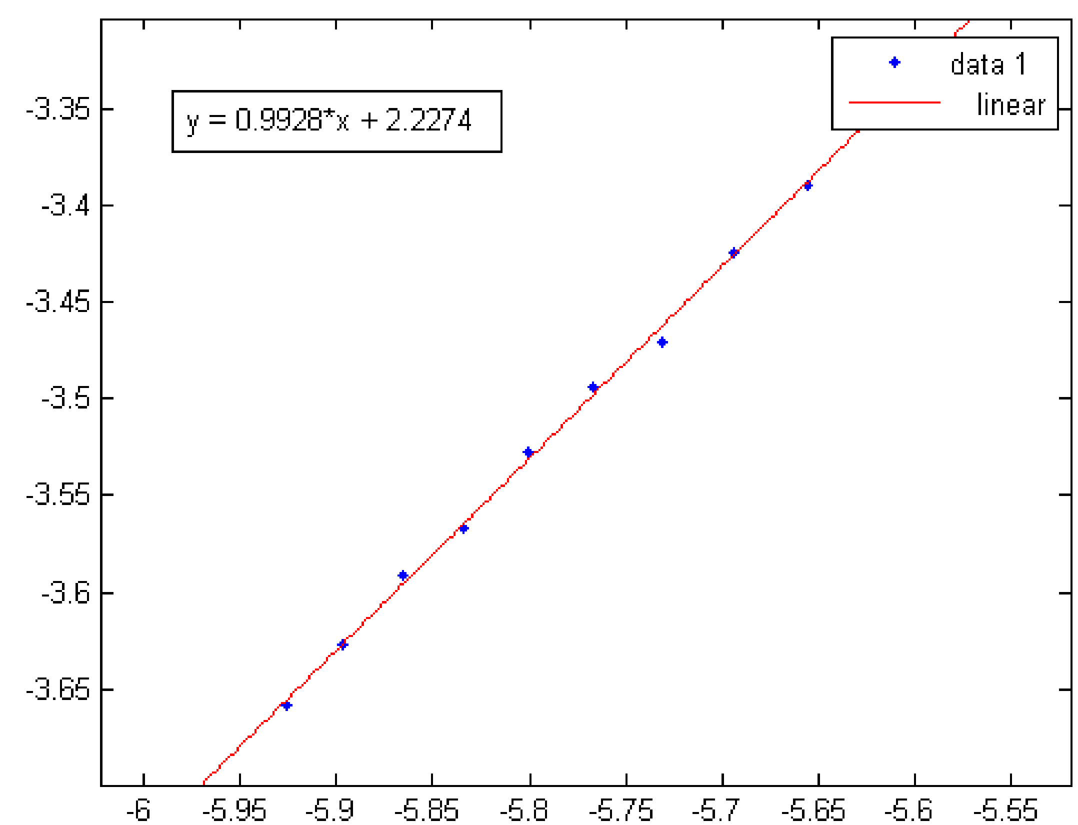




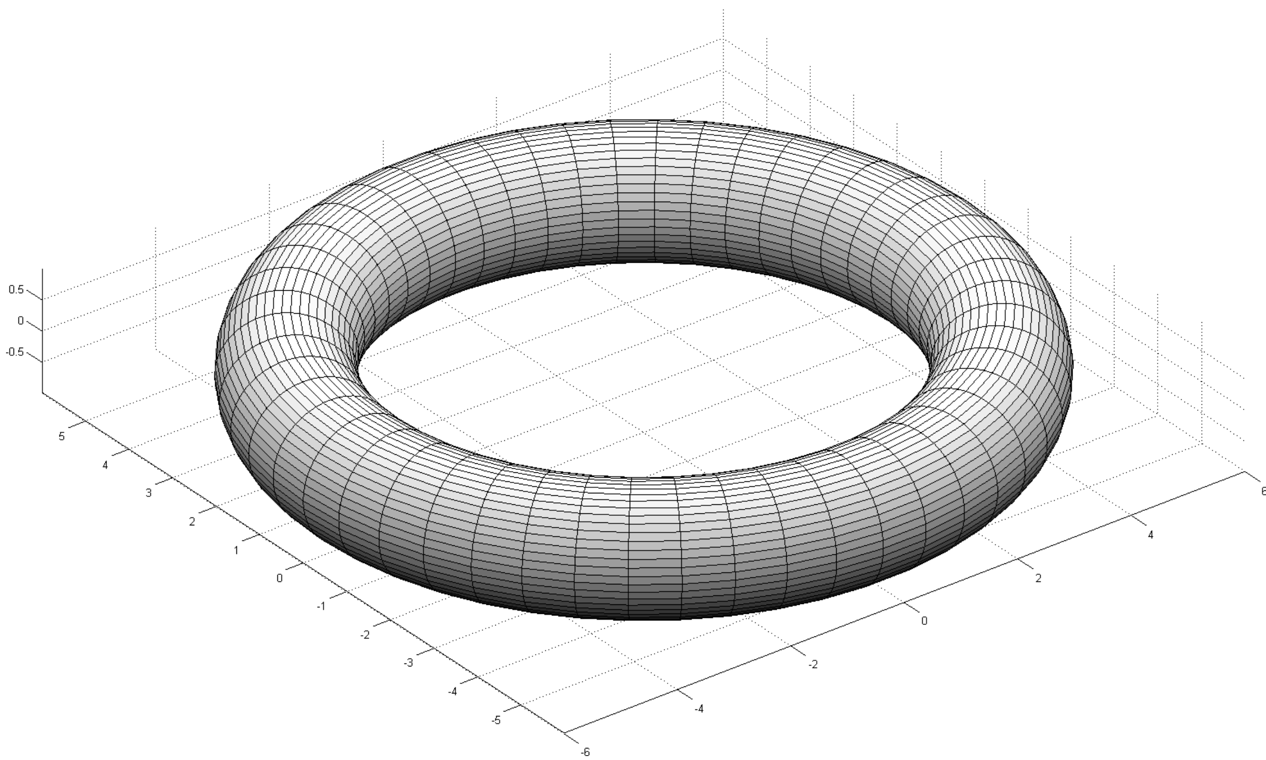
- μ: unif. measure on Lissajous curve X.
- $\mathcal{F}$: distance to X in $\mathbb{R}^2$.
- sample $k = 300$ sets of n points for $n = [2100 : 100 : 3000]$.
- compute

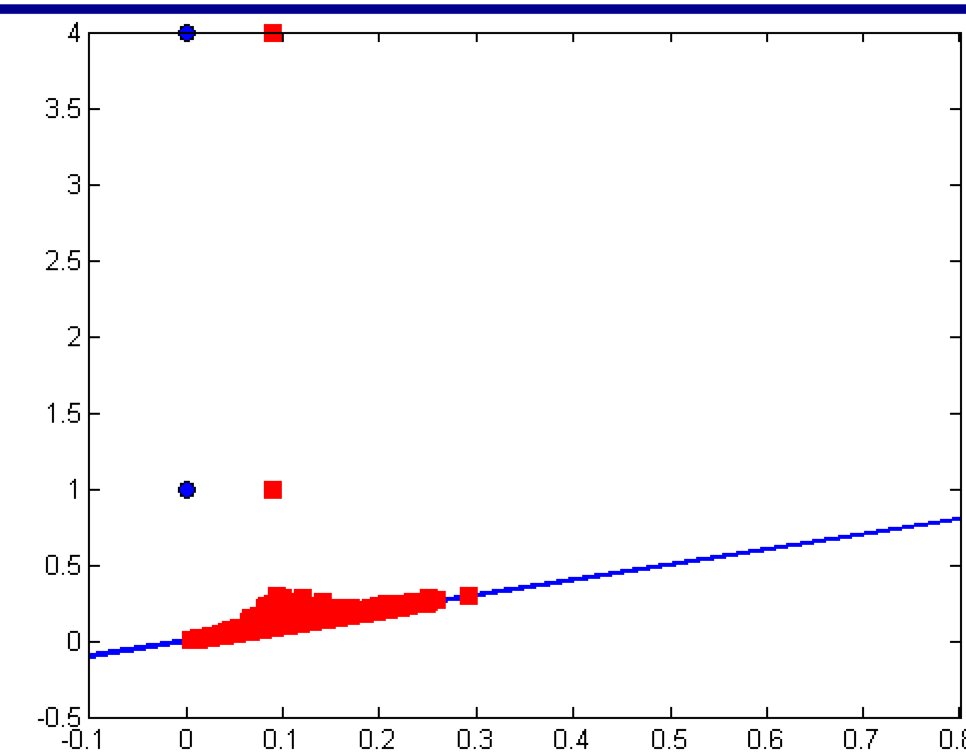
$$\widehat{\mathbb{E}}_n = \widehat{\mathbb{E}}[\mathrm{d_B}(\operatorname{dgm} \mathcal{F}(\widehat{X_n}), \operatorname{dgm} \mathcal{F}(X))].$$

- plot $\log(\widehat{\mathbb{E}}_n)$ as a function of $\log(\log(n)/n)$.



Numerical illustrations

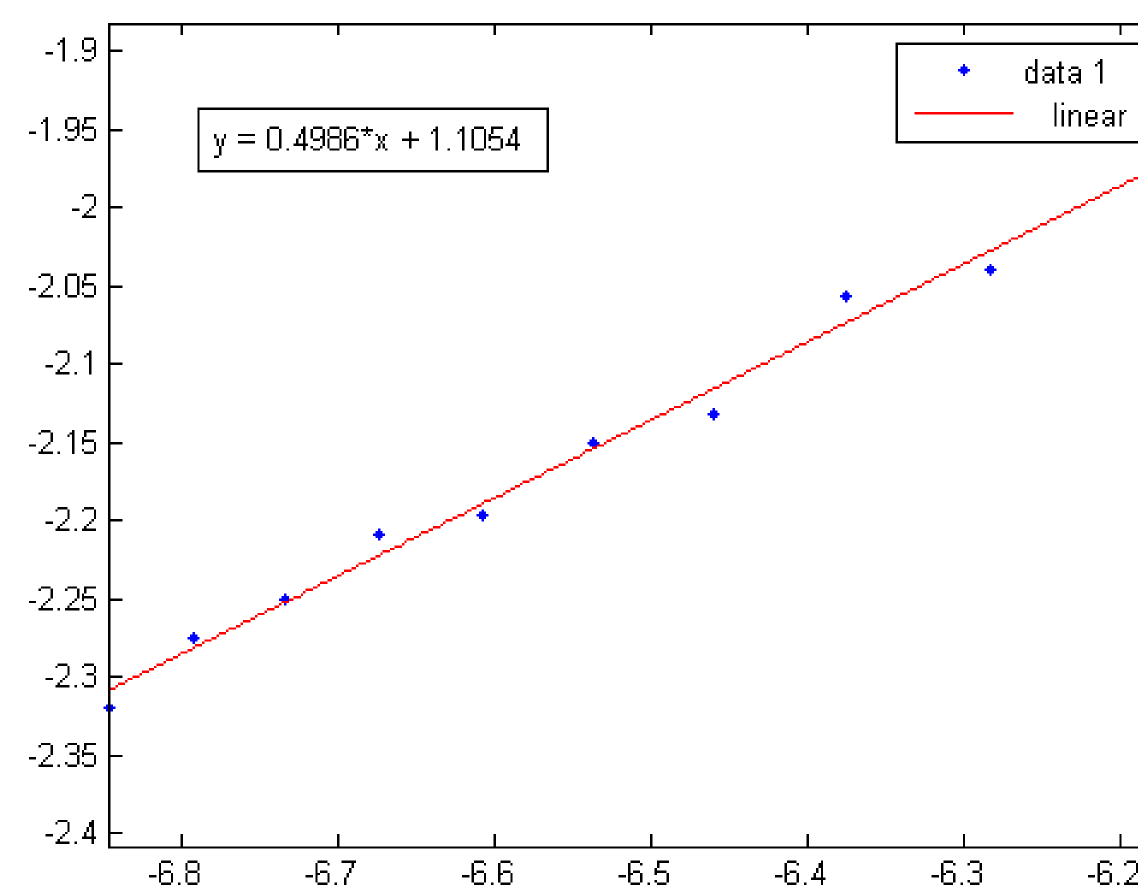




- μ: unif. measure on a torus X.
- $\mathcal{F}$: distance to X in $\mathbb{R}^3$.
- sample $k = 300$ sets of n points for $n = [12000 : 1000 : 21000]$.
- compute

$$\widehat{\mathbb{E}}_n = \widehat{\mathbb{E}}[\mathrm{d_B}(\mathrm{dgm}\,\mathcal{F}(\widehat{X_n}), \mathrm{dgm}\,\mathcal{F}(X))].$$

- plot $\log(\widehat{\mathbb{E}}_n)$ as a function of $\log(\log(n)/n)$.



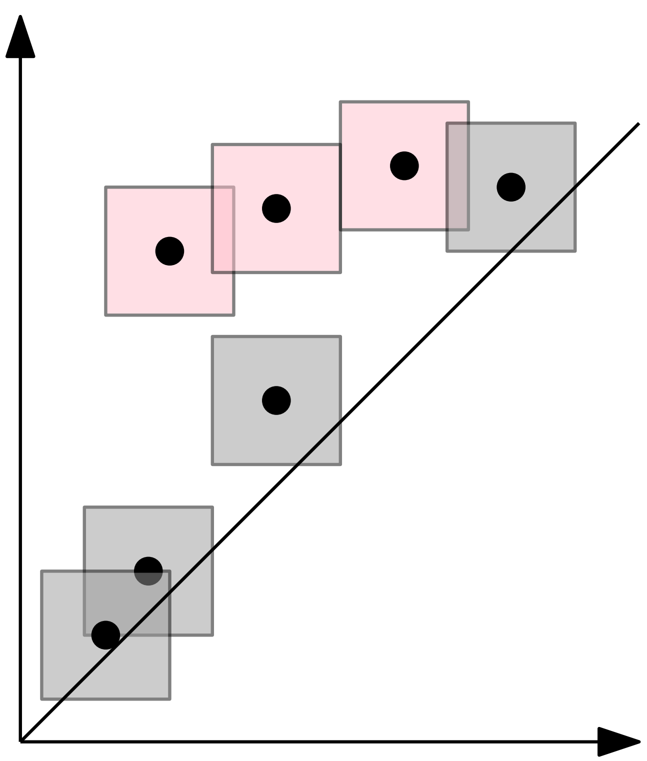
Confidence regions

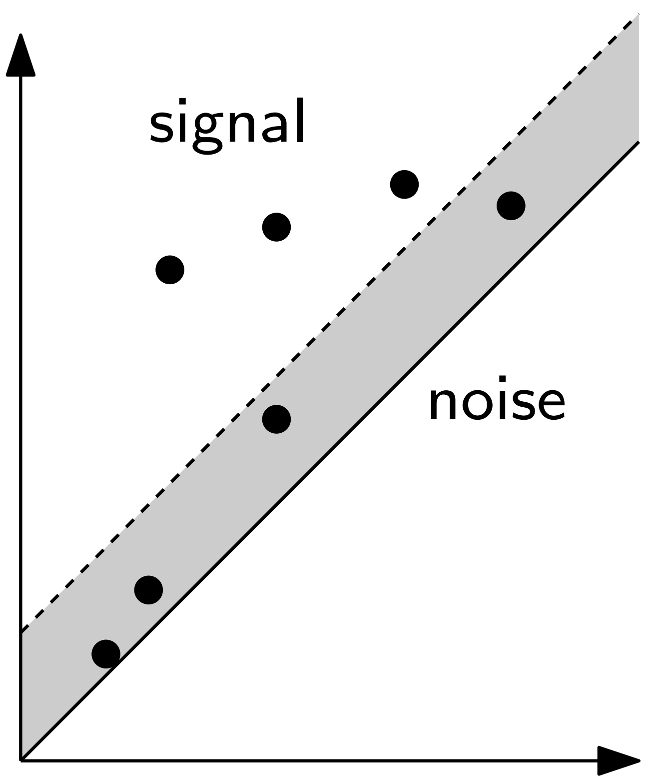
Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n \to \mathcal{F}(\widehat{X}_n) \to \operatorname{dgm} \mathcal{F}(\widehat{X}_n)$

Goal: given $\alpha \in (0,1)$, estimate $c_n(\alpha) \geq 0$ such that

$$\limsup_{n \to \infty} \mathbb{P}\left(\mathrm{d}_\mathrm{B}\left(\operatorname{dgm} \mathcal{F}(\widehat{X}_n), \operatorname{dgm} \mathcal{F}(X)\right) > c_n(\alpha)\right) \leq \alpha$$

$\to$ confidence region: d_B-ball around $\operatorname{dgm} \mathcal{F}(\widehat{X}_n)$





Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n \to \mathcal{F}(\widehat{X}_n) \to \mathrm{dgm}\, \mathcal{F}(\widehat{X}_n)$

Goal: given $\alpha \in (0,1)$, estimate $c_n(\alpha) \geq 0$ such that

$$\limsup_{n \to \infty} \mathbb{P}\left(\mathrm{d_B}\left(\mathrm{dgm}\, \mathcal{F}(\widehat{X}_n), \mathrm{dgm}\, \mathcal{F}(X)\right) > c_n(\alpha)\right) \leq \alpha$$

Note: we already have an inequality of this kind but...

$$\mathbb{P}\left(\mathrm{d_B}\left(\mathrm{dgm}\, \mathcal{F}(\widehat{X}_n), \mathrm{dgm}\, \mathcal{F}(X)\right) > \varepsilon\right) \leq \frac{8^b}{a\varepsilon^b} \exp(-na\varepsilon^b)$$

unknown

Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n \to \mathcal{F}(\widehat{X}_n) \to \mathrm{dgm}\,\mathcal{F}(\widehat{X}_n)$

Goal: given $\alpha \in (0,1)$, estimate $c_n(\alpha) \geq 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\left(\mathrm{d_B}\left(\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n), \mathrm{dgm}\,\mathcal{F}(X)\right) > c_n(\alpha)\right) \leq \alpha$$

Bootstrap:

- draw $X^* = X_1^*, \cdots, X_n^*$ iid from $\mu_{\widehat{X}_n}$ (empirical measure on $\widehat{X}_n$)
- compute $d^* = \mathrm{d_B}\left(\mathrm{dgm}\,\mathcal{F}(X^*), \mathrm{dgm}\,\mathcal{F}(\widehat{X}_n)\right)$
- repeat N times to get $d_1^*, \cdots, d_N^*$
- let q_α be the $(1-\alpha)$ quantile of $\frac{1}{N}\sum_{i=1}^N I(\sqrt{n}\, d_i^* \geq t)$

Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n \to \mathcal{F}(\widehat{X}_n) \to \mathrm{dgm}\,\mathcal{F}(\widehat{X}_n)$

Goal: given $\alpha \in (0,1)$, estimate $c_n(\alpha) \geq 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\left(\mathrm{d_B}\left(\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n), \mathrm{dgm}\,\mathcal{F}(X)\right) > c_n(\alpha)\right) \leq \alpha$$

Bootstrap:

- draw $X^* = X_1^*, \cdots, X_n^*$ iid from $\mu_{\widehat{X}_n}$ (empirical measure on $\widehat{X}_n$)
- compute $d^* = \mathrm{d_B}\left(\mathrm{dgm}\,\mathcal{F}(X^*), \mathrm{dgm}\,\mathcal{F}(\widehat{X}_n)\right)$
- repeat N times to get $d_1^*, \cdots, d_N^*$
- let q_α be the $(1-\alpha)$ quantile of $\frac{1}{N}\sum_{i=1}^N I(\sqrt{n}\, d_i^* \geq t)$

Principle [Efron 1979]: variations of $\mathrm{dgm}\,\mathcal{F}(X^*)$ around $\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n)$ are same as variations of $\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n)$ around $\mathrm{dgm}\,\mathcal{F}(X)$.

Note: requires some conditions on (X, d_X, μ), hence the $\sqrt{n}$.

Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n \to \mathcal{F}(\widehat{X}_n) \to \operatorname{dgm}\mathcal{F}(\widehat{X}_n)$

Goal: given $\alpha \in (0,1)$, estimate $c_n(\alpha) \geq 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\left(\mathrm{d_B}\left(\operatorname{dgm}\mathcal{F}(\widehat{X}_n), \operatorname{dgm}\mathcal{F}(X)\right) > c_n(\alpha)\right) \leq \alpha$$

Bootstrap:

- draw $X^* = X_1^*, \cdots, X_n^*$ iid from $\mu_{\widehat{X}_n}$ (empirical measure on $\widehat{X}_n$)
- compute $d^* = \mathrm{d_B}\left(\operatorname{dgm}\mathcal{F}(X^*), \operatorname{dgm}\mathcal{F}(\widehat{X}_n)\right)$
- repeat N times to get $d_1^*, \cdots, d_N^*$
- let q_α be the $(1-\alpha)$ quantile of $\frac{1}{N}\sum_{i=1}^N I(\sqrt{n}\, d_i^* \geq t)$

Theorem [Balakrishnan et al. 2013] + [Chazal et al. 2014]:

$$\limsup_{n\to\infty} \mathbb{P}\left(\mathrm{d_B}\left(\operatorname{dgm}\mathcal{F}(\widehat{X}_n), \operatorname{dgm}\mathcal{F}(X)\right) > \frac{q_\alpha}{\sqrt{n}}\right) \leq \alpha.$$

Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n \to \mathcal{F}(\widehat{X}_n) \to \mathrm{dgm}\, \mathcal{F}(\widehat{X}_n)$

Goal: given $\alpha \in (0,1)$, estimate $c_n(\alpha) \geq 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\left(\mathrm{d_B}\left(\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n), \mathrm{dgm}\,\mathcal{F}(X)\right) > c_n(\alpha)\right) \leq \alpha$$

Bootstrap:

- draw $X^* = X_1^*, \cdots, X_n^*$ iid from $\mu_{\widehat{X}_n}$ (empirical measure on $\widehat{X}_n$)
- compute $d^* = \mathrm{d_B}\left(\mathrm{dgm}\,\mathcal{F}(X^*), \mathrm{dgm}\,\mathcal{F}(\widehat{X}_n)\right)$
- repeat N times to get $d_1^*, \cdots, d_N^*$
- let q_α be the $(1-\alpha)$ quantile of $\frac{1}{N}\sum_{i=1}^N I(\sqrt{n}\, d_i^* \geq t)$

Theorem [Balakrishnan et al. 2013] + [Chazal et al. 2014]:

Note: extends to stable feature vectors

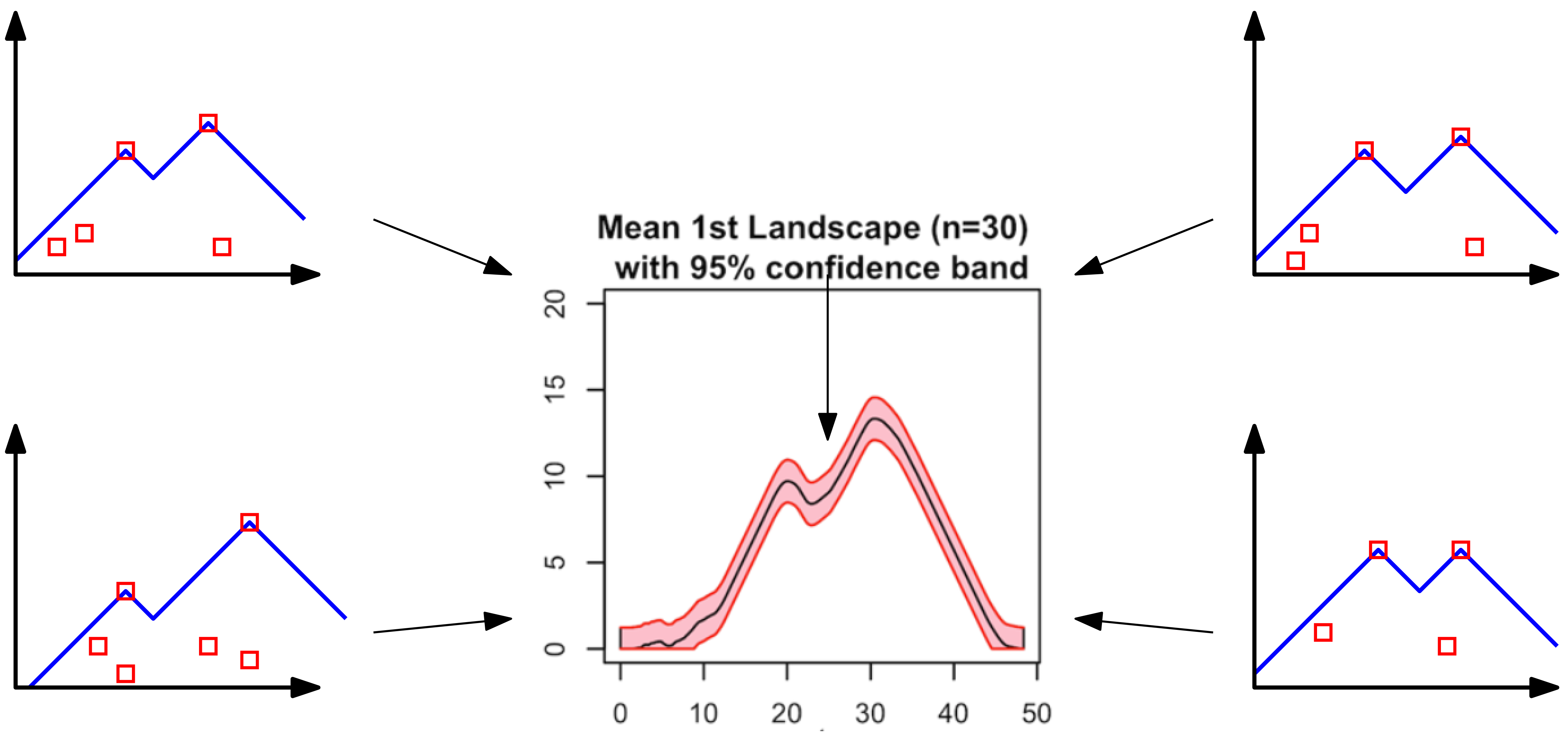
$$\limsup_{n\to\infty} \mathbb{P}\left(\mathrm{d_B}\left(\mathrm{dgm}\,\mathcal{F}(\widehat{X}_n), \mathrm{dgm}\,\mathcal{F}(X)\right) > \frac{q_\alpha}{\sqrt{n}}\right) \leq \alpha.$$

Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \rightarrow \widehat{X}_n^1, \cdots, \widehat{X}_n^m \rightarrow \phi_k(D_n^1), \cdots, \phi_k(D_n^m)$

$$\downarrow$$

empirical mean feature vector ⟶ $\bar{v} = \frac{1}{m} \sum_{i=1}^{m} \phi_k(D_n^i)$



Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n^1, \cdots, \widehat{X}_n^m \to \phi_k(D_n^1), \cdots, \phi_k(D_n^m)$

$$\downarrow$$

empirical mean feature vector ⟶ $\bar{v} = \frac{1}{m} \sum_{i=1}^{m} \phi_k(D_n^i)$

Goal: given $\alpha \in (0,1)$, estimate $c_n(\alpha) \geq 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}\left(\left\| \bar{v} - \underbrace{\mathbb{E}_{(\phi_k \circ \mathrm{dgm} \circ \mathcal{F})^*(\mu^{\otimes n})}[v]} \right\|_{\mathcal{H}_k} > c_n(\alpha) \right) \leq \alpha$$

↑

mean feature vector according to the measure induced by $\mu^{\otimes n}$

(call it $\Lambda_{\mu,n}$ for landscapes)

Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n^1, \cdots, \widehat{X}_n^m \to \lambda(D_n^1), \cdots, \lambda(D_n^m)$

$$\downarrow$$

$$\bar{\lambda} = \frac{1}{m} \sum_{i=1}^{m} \lambda(D_n^i)$$

Bootstrap with landscapes:

- draw $\lambda_1^*, \cdots, \lambda_m^*$ iid from $\frac{1}{m} \sum_{i=1}^{m} \delta_{\lambda(D_n^i)}$
- compute $\bar{\lambda}^* = \frac{1}{m} \sum_{i=1}^{m} \lambda_i^*$ and $d^* = \|\bar{\lambda}^* - \bar{\lambda}\|_\infty$
- repeat N times to get $d_1^*, \cdots, d_N^*$
- let q_α be the $(1-\alpha)$ quantile of $\frac{1}{N} \sum_{i=1}^{N} I(\sqrt{m}\, d_i^* \geq t)$

Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n^1, \cdots, \widehat{X}_n^m \to \lambda(D_n^1), \cdots, \lambda(D_n^m)$

$$\downarrow$$

$$\bar{\lambda} = \frac{1}{m}\sum_{i=1}^m \lambda(D_n^i)$$

Bootstrap with landscapes:

- draw $\lambda_1^*, \cdots, \lambda_m^*$ iid from $\frac{1}{m}\sum_{i=1}^m \delta_{\lambda(D_n^i)}$
- compute $\bar{\lambda}^* = \frac{1}{m}\sum_{i=1}^m \lambda_i^*$ and $d^* = \|\bar{\lambda}^* - \bar{\lambda}\|_\infty$
- repeat N times to get $d_1^*, \cdots, d_N^*$
- let q_α be the $(1-\alpha)$ quantile of $\frac{1}{N}\sum_{i=1}^N I(\sqrt{m}\, d_i^* \geq t)$

Theorem [Chazal et al. 2014]:

$$\limsup_{m\to\infty} \mathbb{P}\left(\left\|\bar{\lambda} - \Lambda_{\mu,n}\right\|_\infty > \frac{q_\alpha}{\sqrt{m}}\right) \leq \alpha.$$

Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n^1, \cdots, \widehat{X}_n^m \to \lambda(D_n^1), \cdots, \lambda(D_n^m)$

$$\downarrow$$

$$\bar{\lambda} = \frac{1}{m}\sum_{i=1}^m \lambda(D_n^i)$$

Bootstrap with landscapes:

- draw $\lambda_1^*, \cdots, \lambda_m^*$ iid from $\frac{1}{m}\sum_{i=1}^m \delta_{\lambda(D_n^i)}$
- compute $\bar{\lambda}^* = \frac{1}{m}\sum_{i=1}^m \lambda_i^*$ and $d^* = \|\bar{\lambda}^* - \bar{\lambda}\|_\infty$
- repeat N times to get $d_1^*, \cdots, d_N^*$ $|\bar{\lambda}^*(t) - \bar{\lambda}(t)|$
- let q_α be the $(1-\alpha)$ quantile of $\frac{1}{N}\sum_{i=1}^N I(\sqrt{m}\, d_i^* \geq t)$

Theorem [Chazal et al. 2014]:

Note: can be done for a fixed t

$$\limsup_{m\to\infty} \mathbb{P}\left(\left\|\bar{\lambda} - \Lambda_{\mu,n}\right\|_\infty > \frac{q_\alpha}{\sqrt{m}}\right) \leq \alpha.$$

$|\bar{\lambda}(t) - \Lambda_{\mu,n}(t)|$

Confidence regions

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n^1, \cdots, \widehat{X}_n^m \to \lambda(D_n^1), \cdots, \lambda(D_n^m)$

$$\downarrow$$

$$\bar{\lambda} = \frac{1}{m}\sum_{i=1}^m \lambda(D_n^i)$$

Bootstrap with landscapes:

- draw $\lambda_1^*, \cdots, \lambda_m^*$ iid from $\frac{1}{m}\sum_{i=1}^m \delta_{\lambda(D_n^i)}$
- compute $\bar{\lambda}^* = \frac{1}{m}\sum_{i=1}^m \lambda_i^*$ and $d^* = \|\bar{\lambda}^* - \bar{\lambda}\|_\infty$
- repeat N times to get $d_1^*, \cdots, d_N^*$
- let q_α be the $(1-\alpha)$ quantile of $\frac{1}{N}\sum_{i=1}^N I(\sqrt{m}\, d_i^* \geq t)$

Theorem [Chazal et al. 2015]:

$$\left\|\bar{\lambda} - \lambda(\mathrm{dgm}\,\mathcal{F}(X))\right\|_\infty \leq \underbrace{\left\|\bar{\lambda} - \Lambda_{\mu,n}\right\|_\infty}_{\text{variance term}} + \underbrace{\|\Lambda_{\mu,n} - \lambda(\mathrm{dgm}\,\mathcal{F}(X))\|_\infty}_{\text{bias term}}$$

variance term

bias term $\leq C\left(\frac{\log n}{an}\right)^{1/b}$ when μ is (a,b)-standard

Subsampling

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n$ with n large $(10^6$ to $10^9)$

Subsampling

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n$ with n large (10^6 to 10^9)

Subsampling with landscapes: Let $m \ll n$

- draw X^* from $\mu_{\widehat{X}_n}^{\otimes m}$ (m points iid from empirical measure on $\widehat{X}_n$)
- compute $\lambda^* = \lambda(\operatorname{dgm} \mathcal{F}(X^*))$
- repeat N times to get $\lambda_1^*, \cdots, \lambda_N^*$
- compute $\bar{\lambda}^* = \frac{1}{N} \sum_{i=1}^{N} \lambda_i^*$

Subsampling

Setup: $(X, \mathrm{d}_X, \mu) \to \widehat{X}_n$ with n large (10^6 to 10^9)

Subsampling with landscapes: Let $m \ll n$

- draw X^* from $\mu_{\widehat{X}_n}^{\otimes m}$ (m points iid from empirical measure on $\widehat{X}_n$)
- compute $\lambda^* = \lambda(\operatorname{dgm} \mathcal{F}(X^*))$
- repeat N times to get $\lambda_1^*, \cdots, \lambda_N^*$
- compute $\bar{\lambda}^* = \frac{1}{N} \sum_{i=1}^N \lambda_i^*$

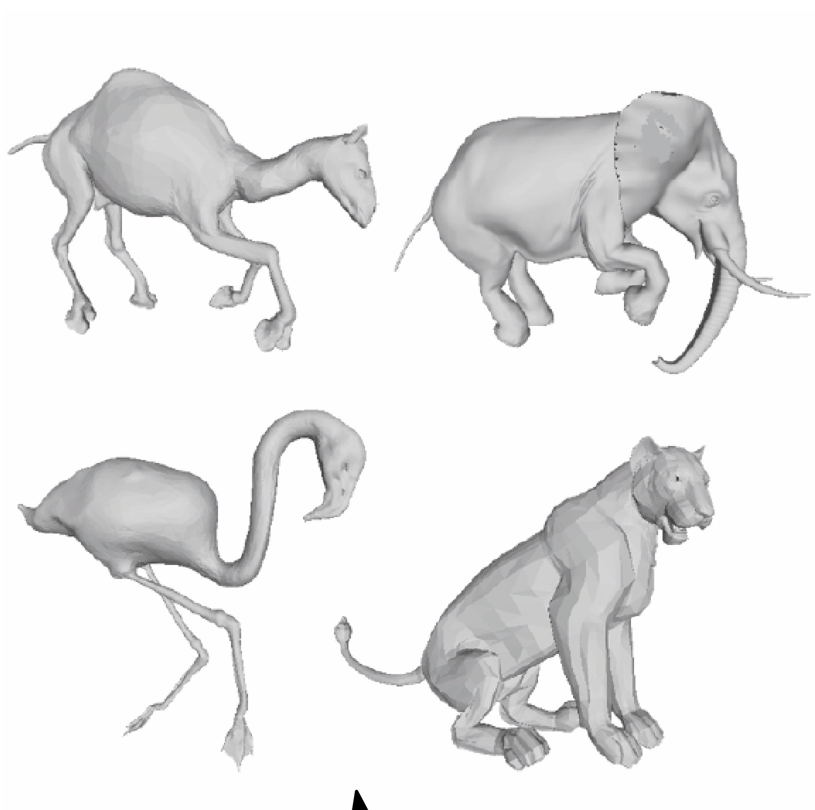
Theorem [Chazal et al. 2015]:

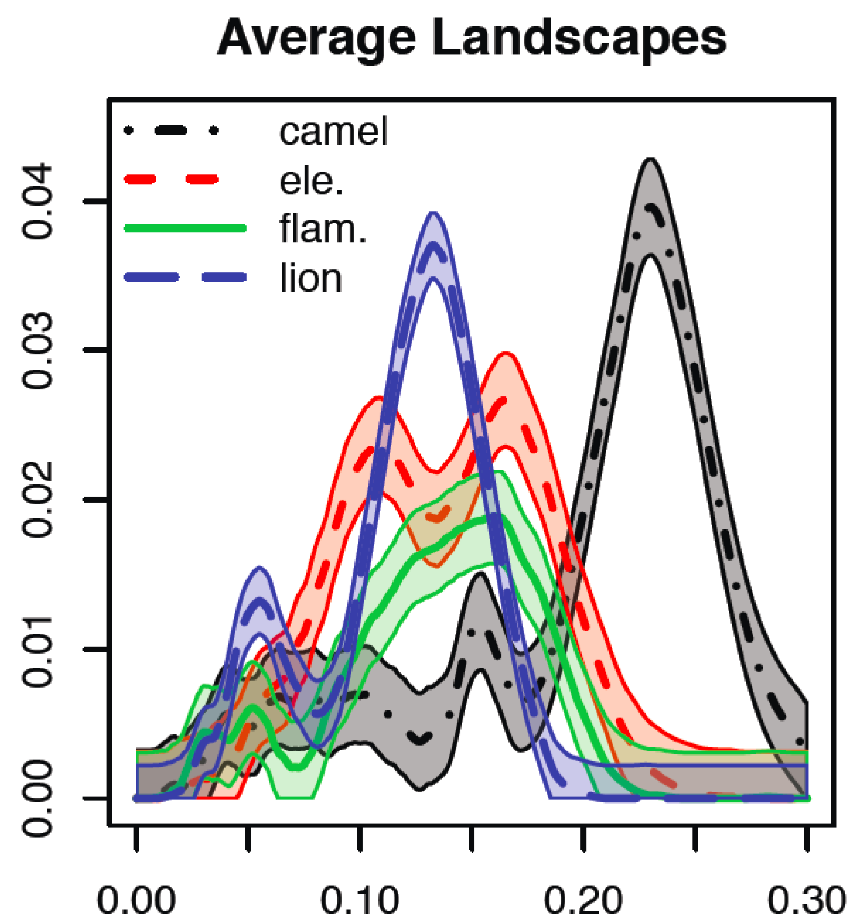
$$\left\| \Lambda_{\mu_{\widehat{X}_n}, m} - \Lambda_{\mu, m} \right\|_\infty \leq m^{1/p} \, W_p(\mu_{\widehat{X}_n}, \mu)$$

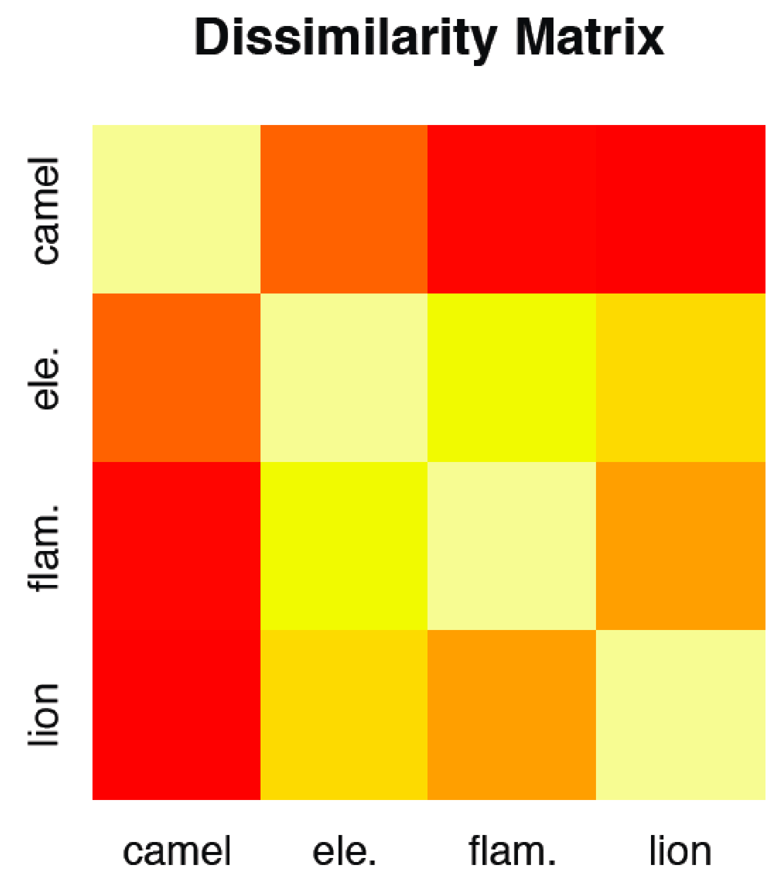
$\to$ by approximating $\Lambda_{\mu_{\widehat{X}_n}, m}$, the empirical mean $\bar{\lambda}^*$ also approximates $\Lambda_{\mu, m}$

Some applications

Application 1: 3D shapes classification





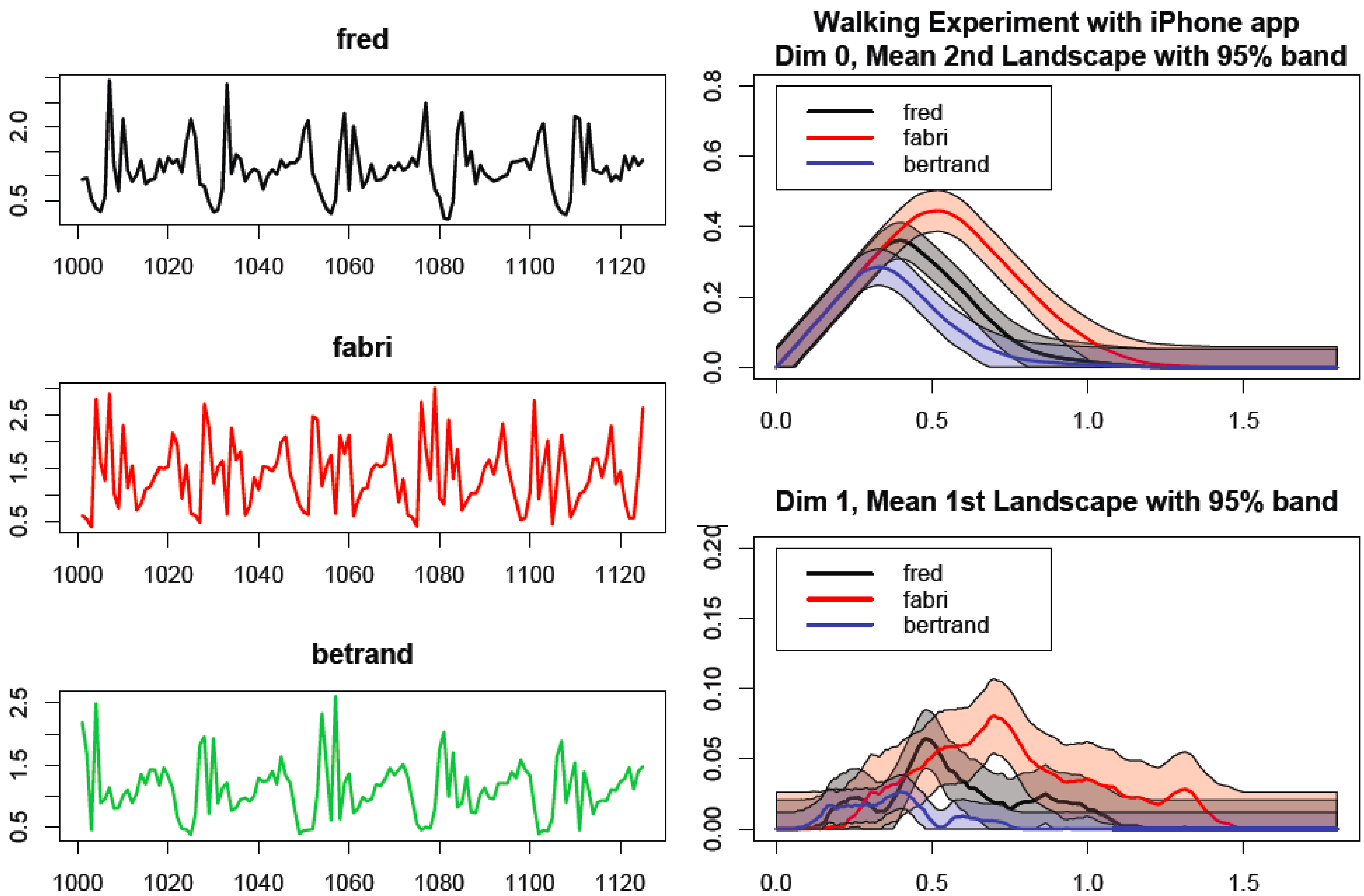


each mesh has 7K to 40K vertices

From $m = 100$ subsamples of size $n = 300$

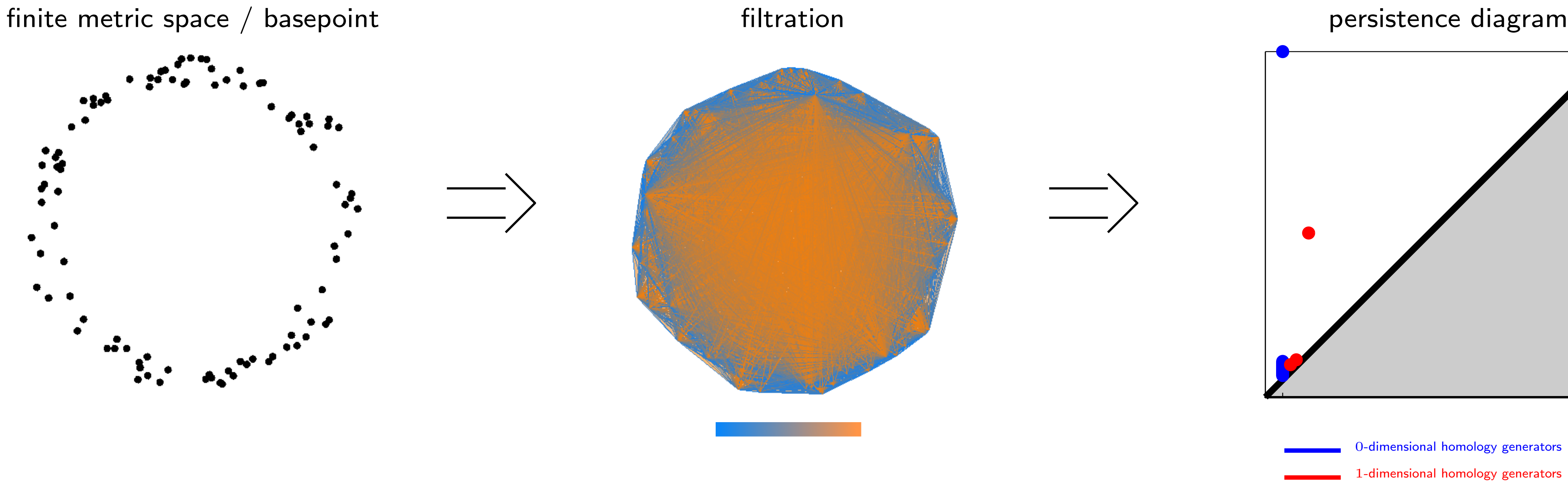
Some applications

Application 2: walking behaviors classification from smartphone accelerometer data

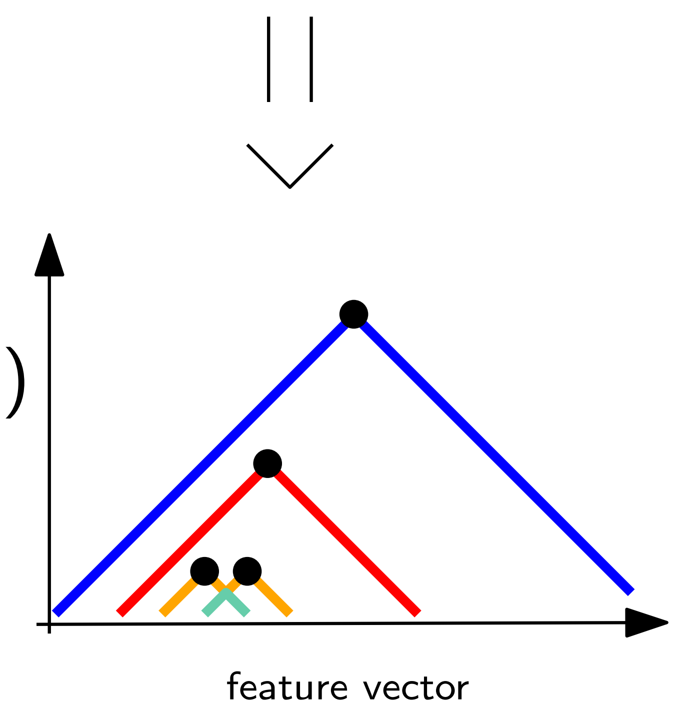


- spatial time series (accelerometer data from the smarphone of users).
- no registration/calibration preprocessing step needed to compare!

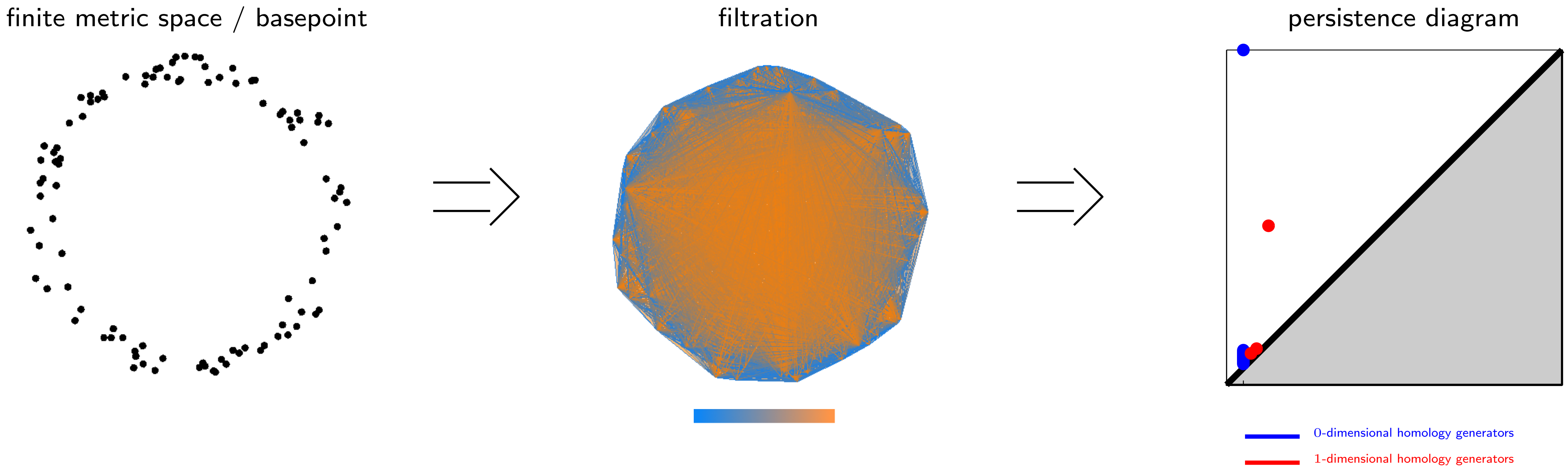
Recap'



- kernels for persistence diagrams:
  - stable
  - informative (injective $\rightarrow$ discriminative?)
  - easy to compute (closed-form expr., finite-dim. vectors)
  - additive, universal, etc.



Recap'



- kernels for persistence diagrams:
- statistical analysis based on stability theorem(s):
  - cvgence rates
  - confidence regions (bootstrap, subsampling)
  - stats. on diagrams (Fréchet means [Turner et al. 2012])
  - stats. on feature vectors (landscapes)