Unsupervised Learning

— A Brief Introduction —

Christoph Schnörr

Image & Pattern Analysis Group
Heidelberg University

STRUCTURES
EP Mathematical Data Analysis

April 9, 2019

Setting, Basic Task

$$
\begin{tikzcd}
\text{measure} \quad \mu_X \arrow[d, "\text{sample}"] & \text{compact space (“shape”)} \arrow[l, "\text{supp}"'] \\
\text{finite data set} \quad (X, d) &
\end{tikzcd}
$$

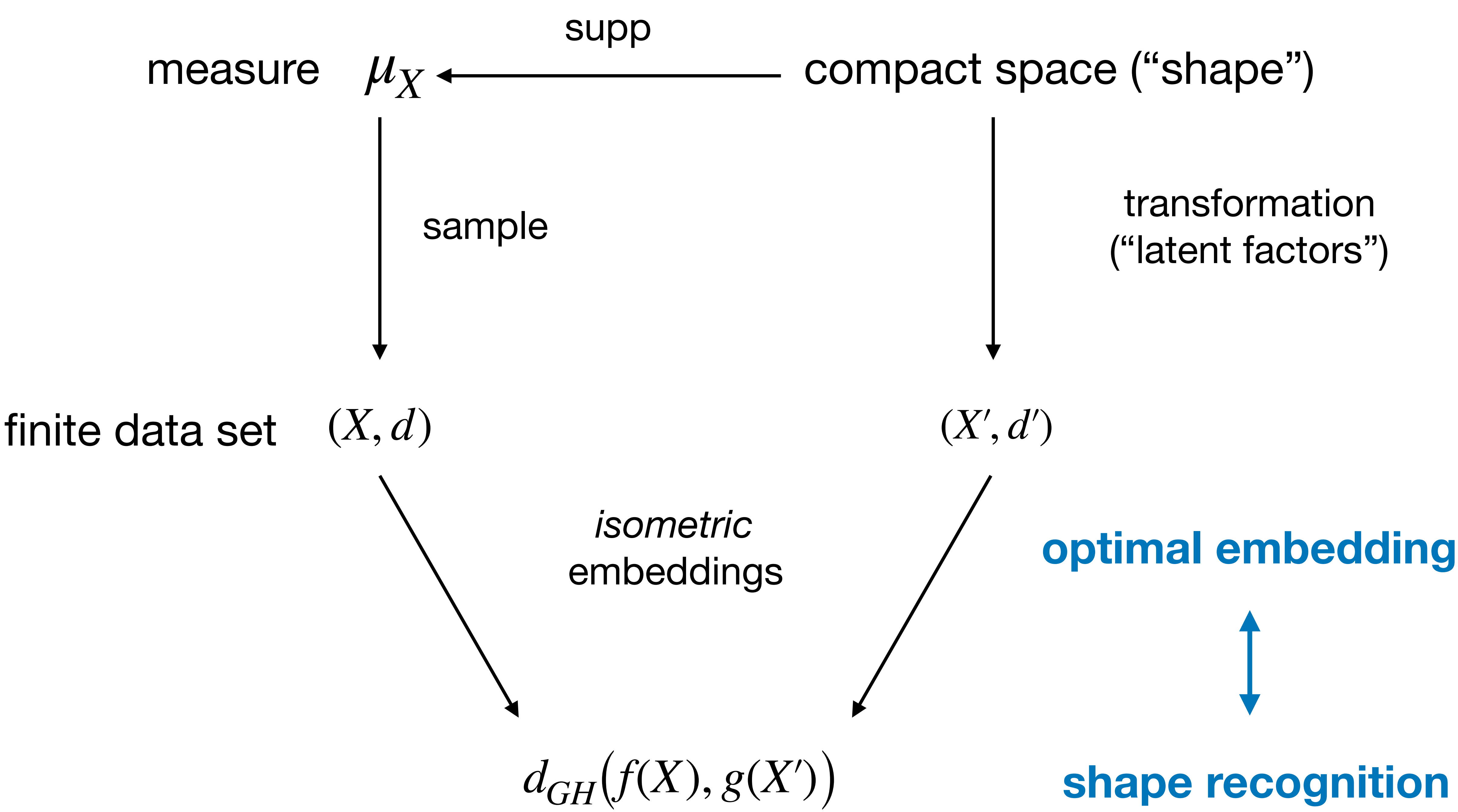
Setting, Basic Task

$$
\begin{tikzcd}
\text{measure} & \mu_X \arrow[d, "\text{sample}"] & & \text{compact space (``shape'')} \arrow[ll, "\text{supp}"'] \arrow[d, "\text{transformation (``latent factors'')}"] \\
\text{finite data set} & (X, d) \arrow[r] & \text{``structure''} & (X', d') \arrow[l]
\end{tikzcd}
$$

Setting, Basic Task

$$
\begin{tikzcd}
\text{measure} \quad \mu_X \arrow[d, "\text{sample}"] & & \text{compact space (``shape'')} \arrow[ll, "\text{supp}"'] \arrow[d, "\text{transformation (``latent factors'')}"] \\
\text{finite data set} \quad (X, d) \arrow[dr] & \text{isometric embeddings} & (X', d') \arrow[dl] \\
& d_{GH}\big(f(X), g(X')\big) &
\end{tikzcd}
$$

Setting, Basic Task



Data Representation

finite data set (X, d_X)

↓

embedding (distortion) $\ell_2^n \quad \ell_2 \quad \ell_p^n \quad (Y, d_Y)$

Euclidean
Hilbert
Banach
metric

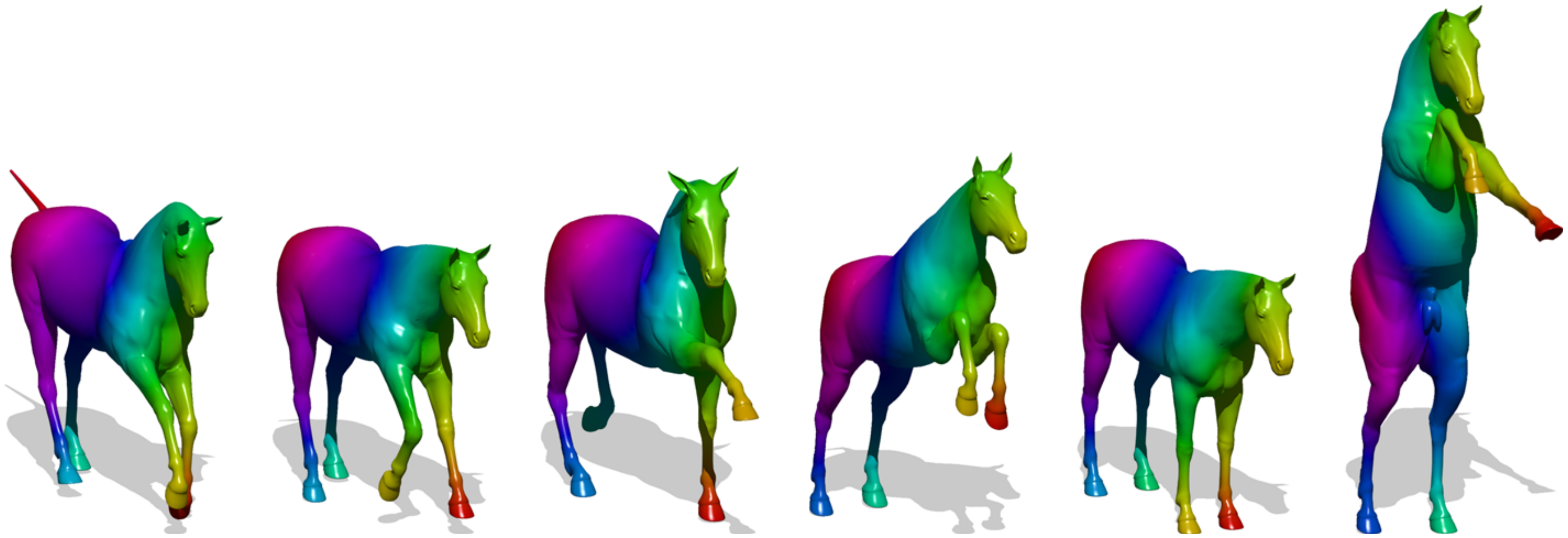
↕

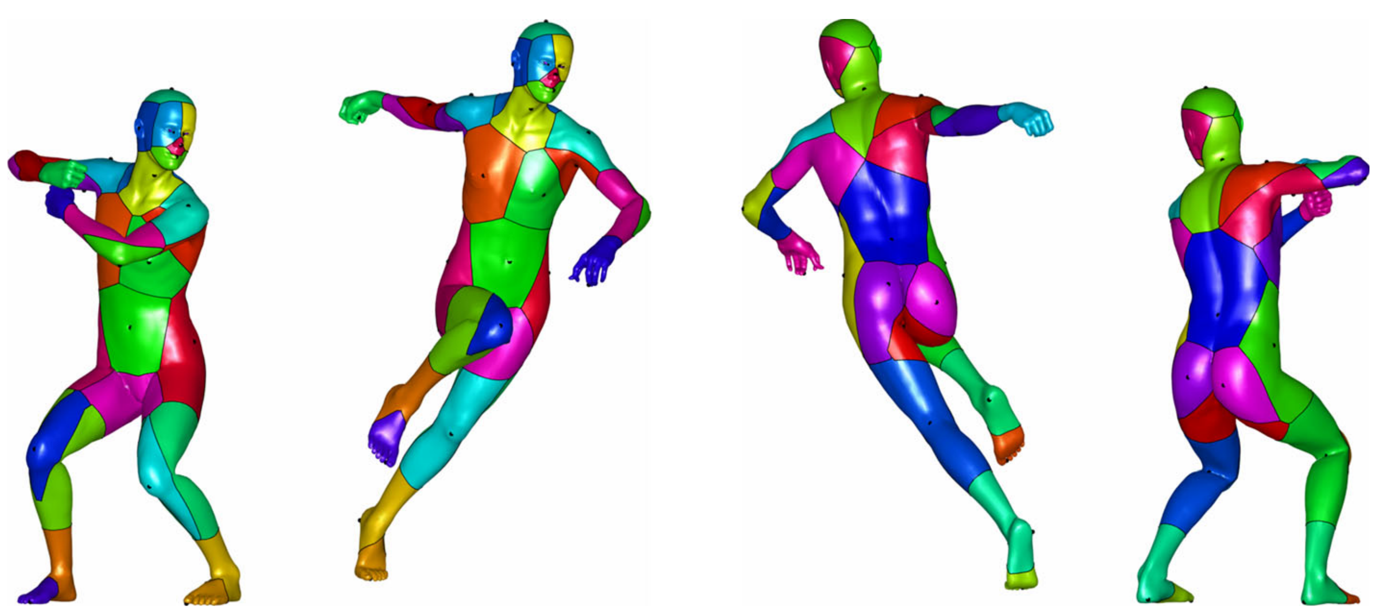
tasks

shape classification & recognition
taming combinatorial optimization
efficient distance oracles
visualisation
…

Example, Manifold Assumption

(Kimmel et al. IJCV'16)





good metric

clustering (partitioning, coding)

Manifold Assumption ?

stratified spaces
("union of subspaces model")

Manifold Assumption ?

Outline

embedding

basic clustering
(ignoring context)

metric
Euclidean, Hilbert

(representative
examples)

preliminary remark on TDA

---

clustering
(context sensitive)

Metric Embedding (I)

$$(X, d) \longrightarrow (X, d_{\mathscr{T}})$$

d_T dominates d, $\quad \forall T \in \mathscr{T}$

$\mathbb{E}_P[d_T(u, v)] \leq \alpha\, d(u, v), \quad \forall u, v \in X$

Metric Embedding (I)

$$(X, d) \longrightarrow (X, d_{\mathcal{T}})$$

d_T dominates d, $\quad \forall T \in \mathcal{T}$

$$\mathbb{E}_P[d_T(u,v)] \leq \alpha\, d(u,v), \quad \forall u, v \in X$$

clustering

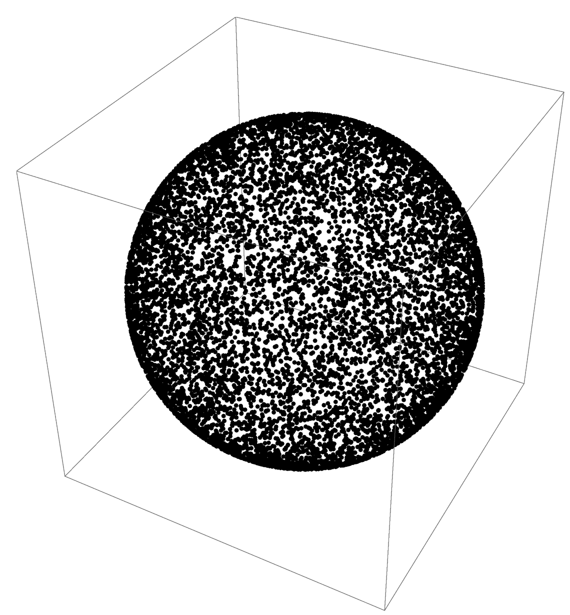
$$E_\infty^* = \min_M E_\infty(M), \qquad E_\infty(M) = \max_{x \in X} d(x, M)$$

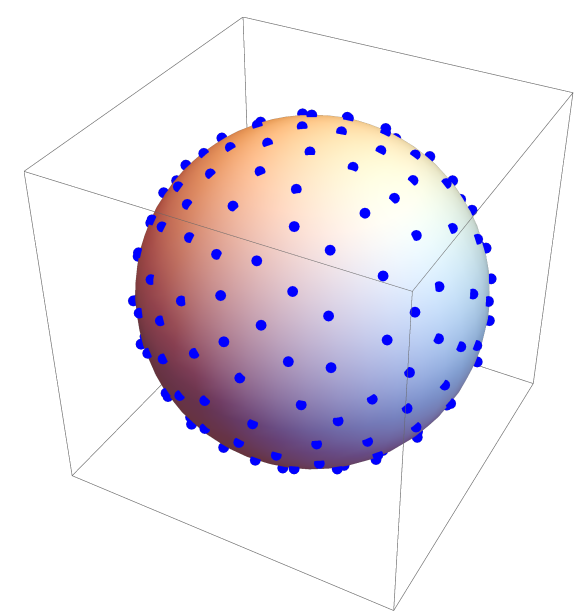
any **metric**
global **method**

$$E_\infty(M) \leq 2E_\infty^*$$

$$M = \{m^1, \ldots, m^c\} \subset X$$

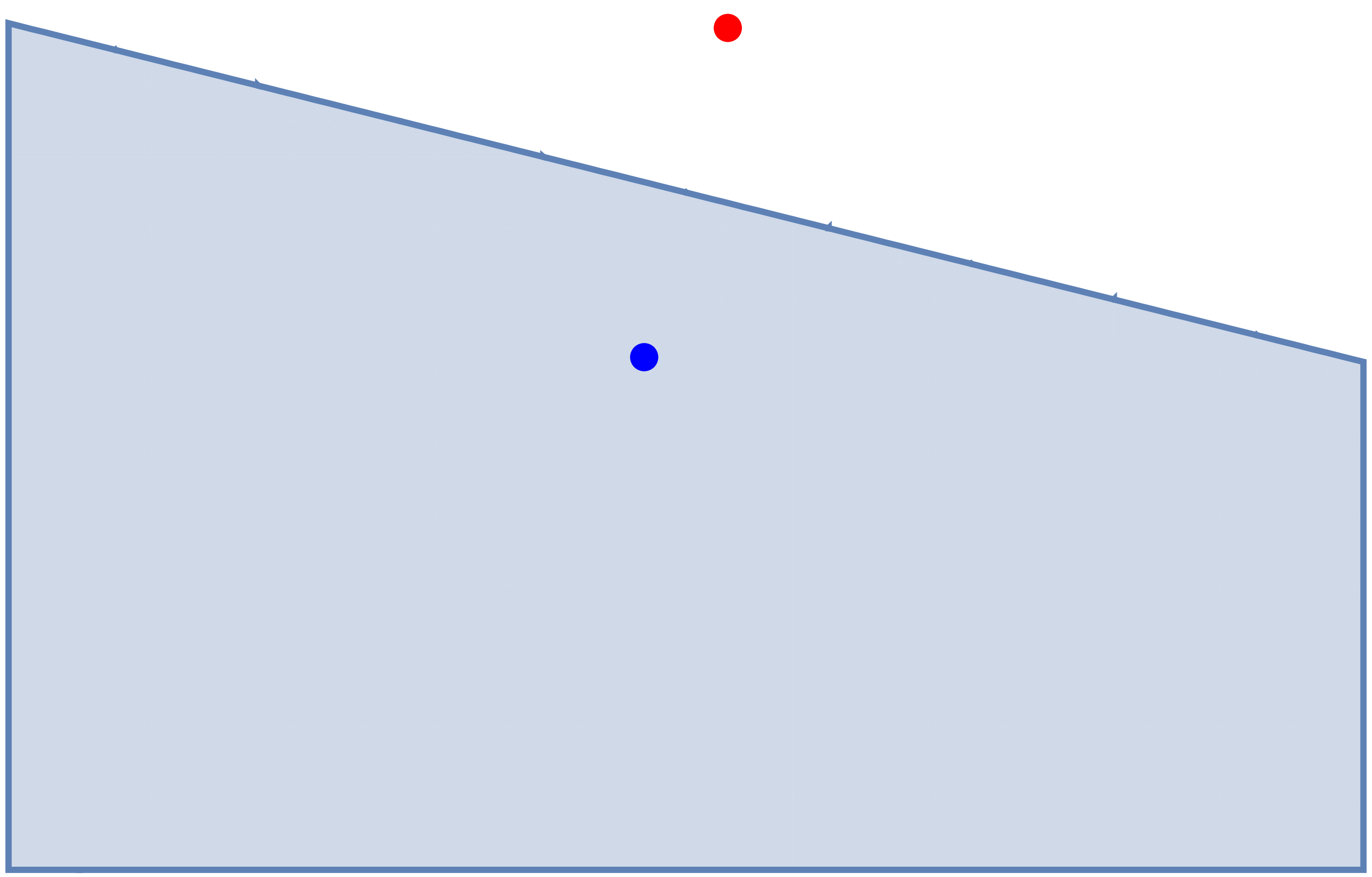
"core set"
initialization for more
advanced techniques





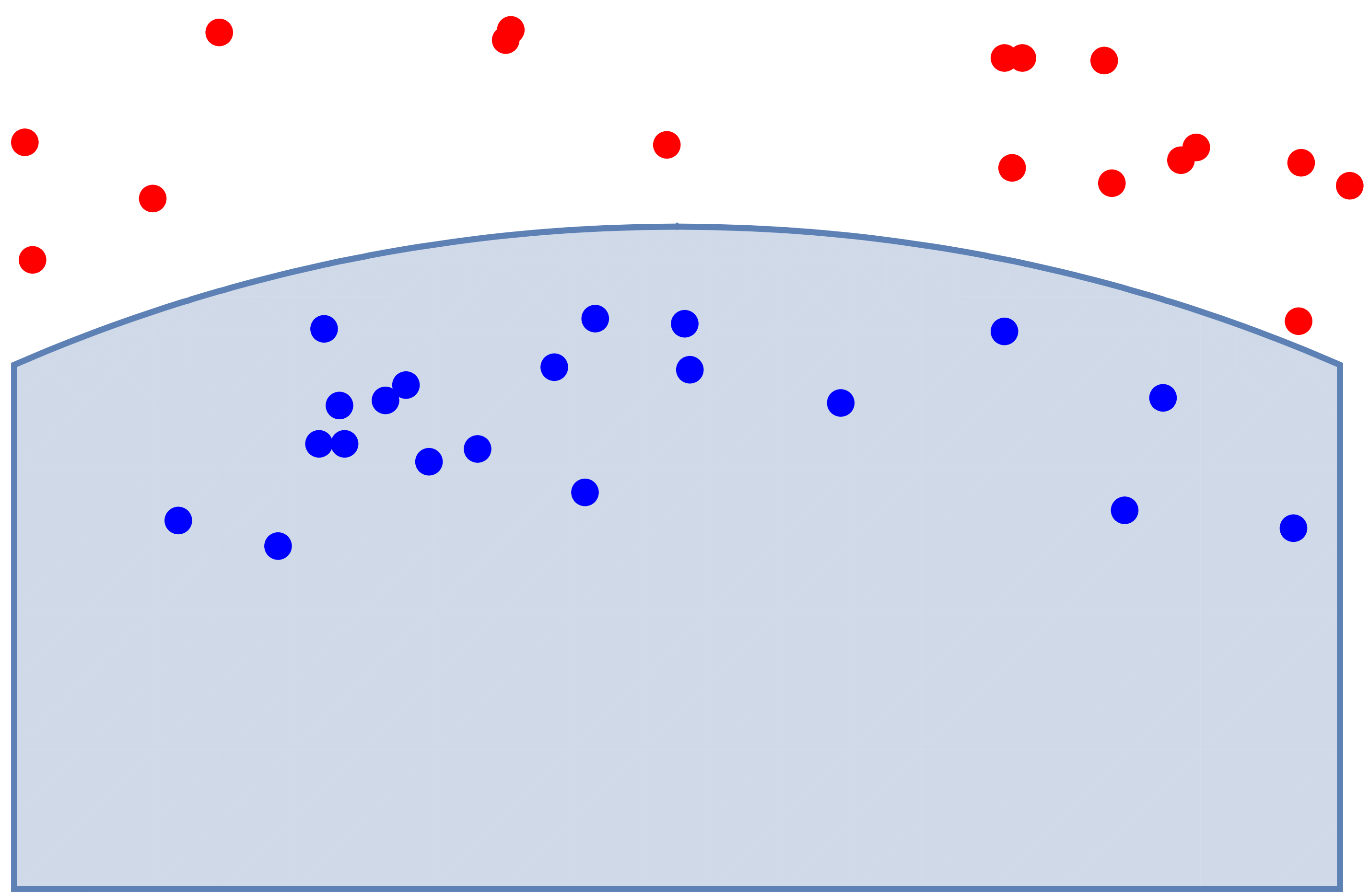
Metric Embedding (II)

data, labels, prior beliefs



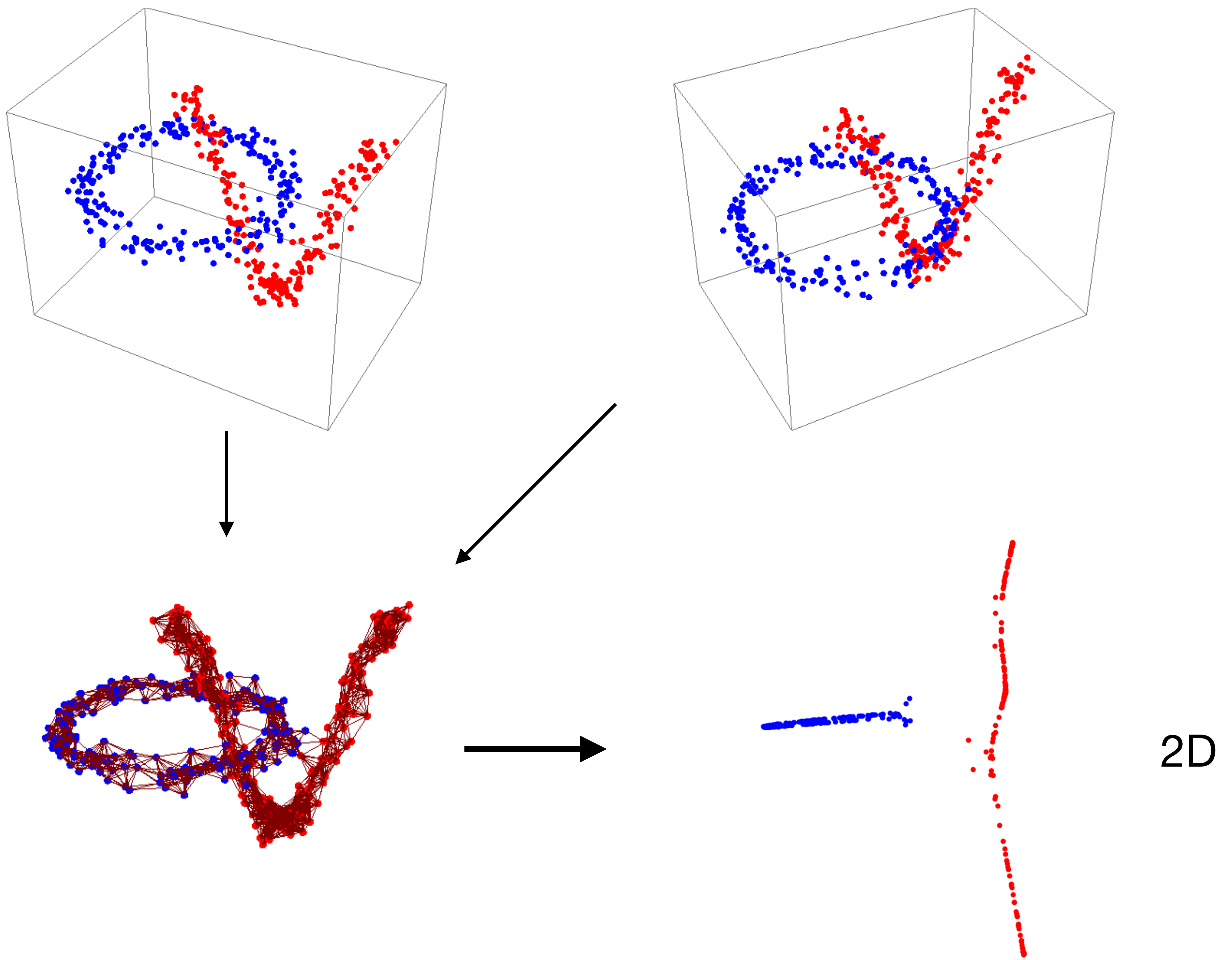
Metric Embedding (II)

more data, *noise*, labels, priors/regularization/smoothness assumption, …
complexity & learning: sample size vs. hypothesis space
online learning
….



manifold assumption

Metric Embedding (II)



Metric Embedding (II)

data “generating” compact ***manifold*** $\mathscr{M}$

data: labeled, unlabeled $X = (X_l, X_u) \subset \mathscr{M}$

labels (Y, X_l)

unsupervised learning $X = X_u, \quad Y = \emptyset$

Metric Embedding (II)

data "generating" compact *manifold* $\mathscr{M}$

data: labeled, unlabeled $X = (X_l, X_u) \subset \mathscr{M}$

labels (Y, X_l)

unsupervised learning $X = X_u, \quad Y = \varnothing$

Hilbert space embedding $\boxed{\mathscr{H}_K \ni f\colon X \to \mathbb{R}}$

Mercer kernel $K\colon X \times X \to \mathbb{R}$

$\updownarrow$ $\updownarrow$

RKHS $(\mathscr{H}_K, \|\cdot\|_K)$

Metric Embedding (II)

data “generating” compact manifold $\mathscr{M}$

data: labeled, unlabeled $X = (X_l, X_u) \subset \mathscr{M}$

labels (Y, X_l)

unsupervised learning $X = X_u, \quad Y = \varnothing$

Hilbert space embedding $\boxed{\mathscr{H}_K \ni f\colon X \to \mathbb{R}}$

Mercer kernel $K\colon X \times X \to \mathbb{R}$

$\updownarrow$ $\updownarrow$

RKHS $(\mathscr{H}_K, \|\cdot\|_K)$

$$\boxed{\begin{aligned} \forall f \in \mathscr{H}_K: \quad \langle h_x, f\rangle_K &= f(x) \\ K(x, x') &= \langle h_x, h_{x'}\rangle_K \\ \langle K(x, \cdot\,), f\rangle_K &= f(x) \end{aligned}}$$

Metric Embedding (II)

$$E(f) = \frac{1}{l}\sum_{i\in[l]} L(x_i, y_i, f) + \lambda_A \|f\|_K^2 + \lambda_I \int_{\mathscr{M}} \|\nabla_{\mathscr{M}} f\|^2 d\mu_X$$

embedding quality

labels, loss: supervision

control: hypothesis space

intrinsic geometry

key problem: how do labels (symbols) emerge from raw data?

Metric Embedding (III)

Euclidean case $\quad D_{ij} = \left(d(x_i, x_j)^2\right)$

$$\Leftrightarrow \quad -\frac{1}{2}CDC \succeq 0, \qquad C = I - \frac{1}{|X|}ee^\top$$

→ embedding, approximation etc. by *semidefinite programming*

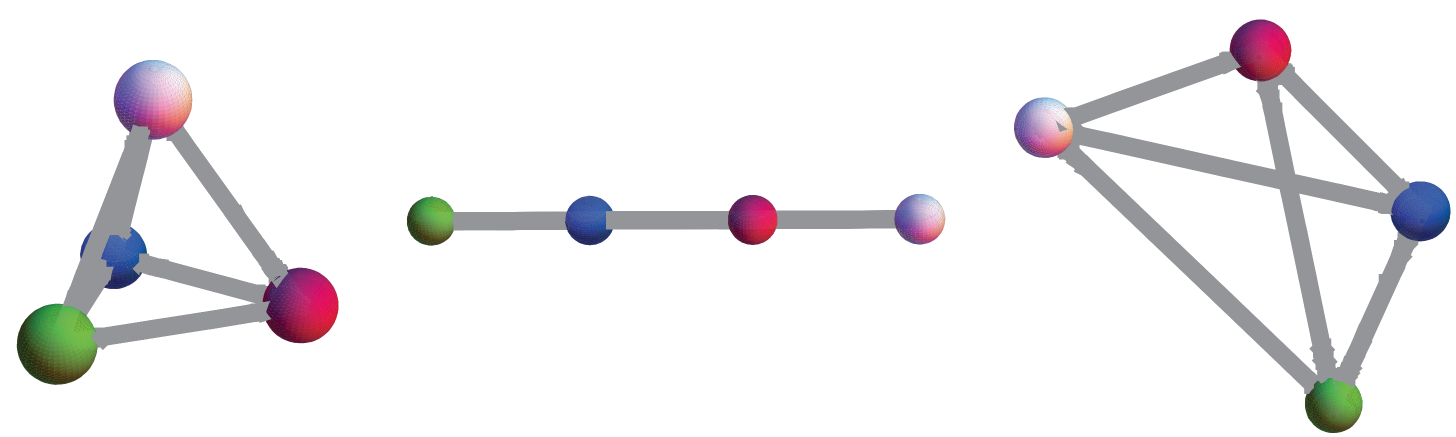
Metric Embedding (III)

Euclidean case $\quad D_{ij} = \left(d(x_i, x_j)^2\right)$

$$\Leftrightarrow \quad -\frac{1}{2}CDC \succeq 0, \qquad C = I - \frac{1}{|X|}ee^\top$$

→ embedding, approximation etc. by *semidefinite programming*

Example: Euclidean representation of label metrics
($\longleftrightarrow$ convex relaxation of variational approaches)



Potts   linear   approx.: linear-truncated

Summing Up

data analysis: 20min brainstorming

- metric geometry
- geometric functional analysis
- differential geometry
- functional analysis, PDEs
- convex analysis

Yet, our understanding is quite limited
including tools (deep networks) claimed to perform well

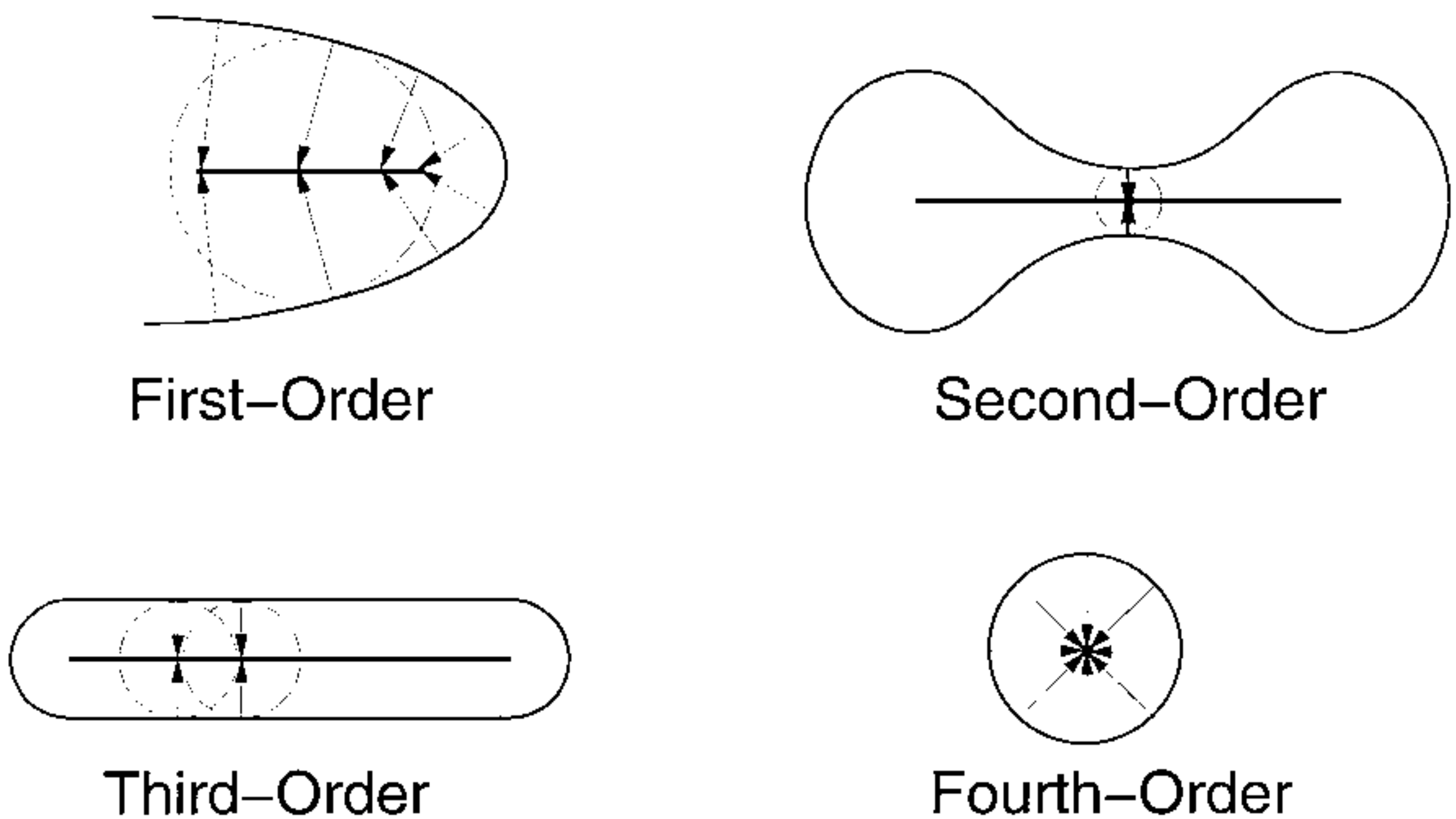
TDA *may* help!

Remark on TDA

TDA has a long history in computer vision and elsewhere

Blum'67: "grassfire transform"
distance & medial axis transform, curvature-driven shape evolution, shock graphs, …

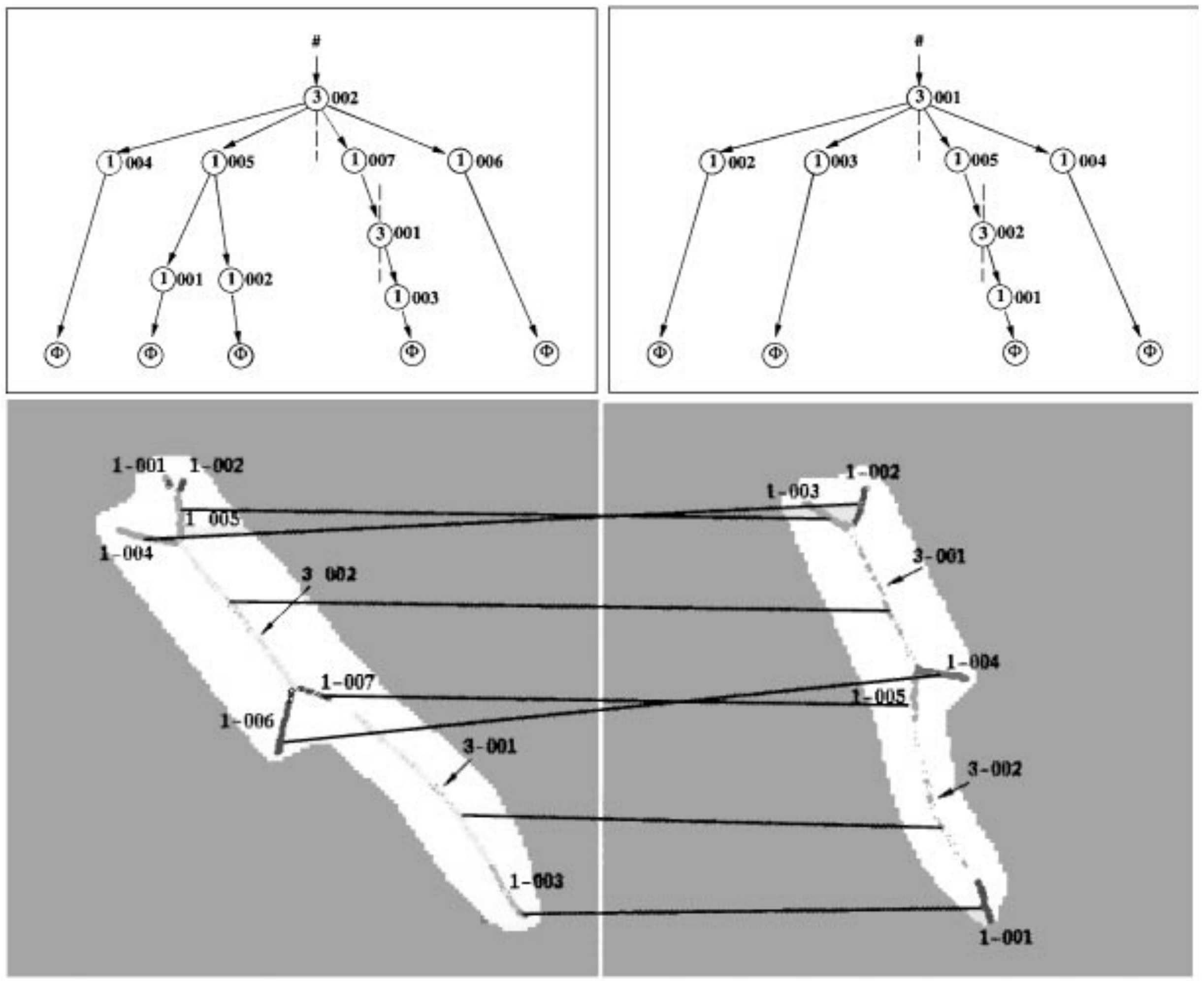
(Siddiqi et al. IJCV'99)



Remark on TDA

Blum'67: "grassfire transform"
distance & medial axis transform, curvature-driven shape evolution, shock graphs, …
maximal subgraph isomophisms, shock trees & shape matching, etc.

(Siddiqi et al. IJCV'99)



Main objection: *lack of stability !*

Remark on TDA

compact space ("shape") $\xrightarrow{\text{supp}}$ measure μ_X $\xrightarrow{\text{sample}}$ (X, d)

Remark on TDA

compact space ("shape") $\xrightarrow{\text{supp}}$ measure μ_X $\xrightarrow{\text{sample}}$ (X, d)

robust distance to measure *(Boissonnat et al. 2018)*

$$d^2_{\mu,m_0}(x) = \frac{1}{m_0}\int_0^{m_0} \delta^2_{\mu,m} dm, \quad m_0 \in (0,1)$$

$$\delta_{\mu,m}(x) = \inf\left\{r > 0 : \mu\left(\overline{B}_r(x)\right) > m\right\}$$

Remark on TDA

compact space ("shape") $\xrightarrow{\text{supp}}$ measure μ_X $\xrightarrow{\text{sample}}$ (X, d)

robust distance to measure *(Boissonnat et al. 2018)*

$$d^2_{\mu,m_0}(x) = \frac{1}{m_0}\int_0^{m_0} \delta^2_{\mu,m}\, dm, \quad m_0 \in (0,1)$$

$$\delta_{\mu,m}(x) = \inf\left\{r > 0 : \mu\left(\overline{B}_r(x)\right) > m\right\}$$

in terms of the *empirical* measure $\hat{\mu}_X = \frac{1}{|X|}\sum_{x_i \in X} \delta_{x_i}$

$$d^2_{\mu,m_0}(x) = \frac{1}{k_0}\sum_{x_i \in k_0\mathrm{NN}_X(x)} \|x_i - x\|^2, \qquad m_0 = \frac{k_0}{n}$$

embedding matters !

Remark on TDA

Consequently: **embedding into** the space of (prob.) **measures** !

$$d_{\mu,m_0}(x) = \min\left\{\frac{1}{\sqrt{m_0}} W_2(m_0\delta_x, \nu) : \nu \in \mathrm{Sub}_{m_0}(\mu)\right\}$$

$\boxed{W_2 \text{ metric !}}$

Remark on TDA

Consequently: **embedding into** the space of (prob.) **measures** !

$$d_{\mu,m_0}(x) = \min\left\{\frac{1}{\sqrt{m_0}} W_2(m_0\delta_x, \nu) : \nu \in \mathrm{Sub}_{m_0}(\mu)\right\}$$

$\boxed{W_2 \text{ metric !}}$

- metric geometry
- geometric functional analysis
- differential geometry
- functional analysis, PDEs
- convex analysis
- optimal transport
- dynamical systems
- PDEs (W-geometry)
- information geometry

Remark on TDA

Consequently: **embedding into** the space of (prob.) **measures** !

$$d_{\mu,m_0}(x) = \min\left\{\frac{1}{\sqrt{m_0}} W_2(m_0\delta_x, \nu) : \nu \in \mathrm{Sub}_{m_0}(\mu)\right\}$$

$\boxed{W_2 \text{ metric !}}$

$$\updownarrow \quad \mu_{x,m_0} = \nu^*$$

$$d_{\mu,m_0}(x) = \left(\frac{1}{m_0}\int \|x-h\|^2 d\mu_{x,m_0}(h)\right)^{1/2}$$

stability

$d_H\big(\mathrm{Sub}_{m_0}(\nu), \mathrm{Sub}_{m_0}(\nu')\big) \leq W_2(\nu, \nu') \quad \Longrightarrow \quad$ stability of d_{μ,m_0}

$\longleftrightarrow$ fundamental task, mm-spaces, $d_{GH} \leftarrow d_{GW}$

Unsupervised Learning

— A Brief Introduction —

Christoph Schnörr

Image & Pattern Analysis Group
Heidelberg University

STRUCTURES
EP Mathematical Data Analysis

April 9, 2019

Hausdorff distance

$$d_H^X(A,B) = \max\left\{\sup_{a\in A}\inf_{b\in B} d(a,b), \sup_{b\in B}\inf_{a\in A} d(a,b)\right\}$$

$A, B:$ shapes

$$= \inf_R \sup_{(a,b)\in R} d(a,b)$$

$R \subset A \times B$ correspondences

Gromov-Hausdorff distance

$$d_{GH}(X,Y) = \inf_{Z,f,g} d_H^Z\big(f(X), g(Y)\big)$$

$f, g:$ isometries

$$= \frac{1}{2}\inf_R \sup_{\substack{(x,y)\in R\\(x',y')\in R}} \left| d_X(x,x') - d_Y(y,y') \right|$$

Gromov-Wasserstein distance

$$d_{GW,p}(X,Y) = \frac{1}{2}\inf_\mu \left(\iint \left| d_X(x,x') - d_Y(y,y') \right|^p d\mu(x,y) d\mu(x',y')\right)^{1/p}$$

μ couples μ_X, μ_Y