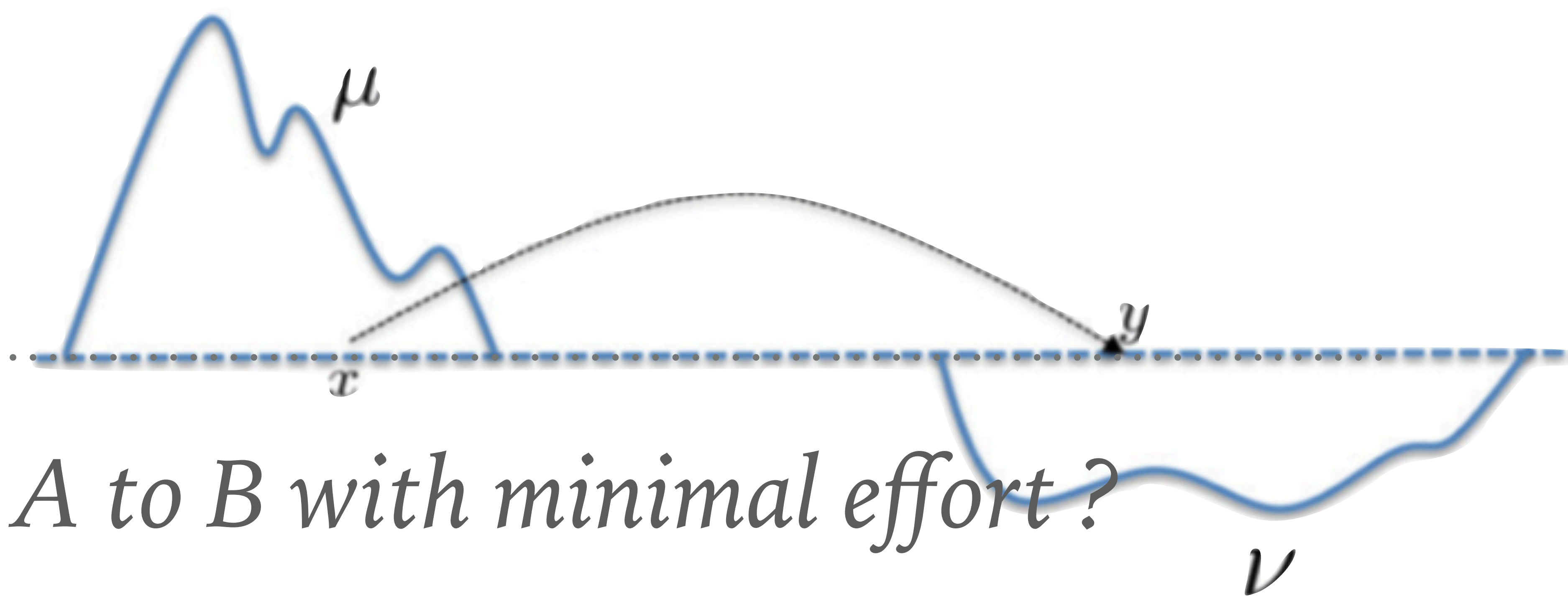
Introduction to Optimal Transport

D. Coeurjolly



Definitions



- The « Lazy dude »: *How to move a bunch a stuff from A to B with minimal effort ?*
  - ➤ effort: mass times Euclidean distance (or any cost function)

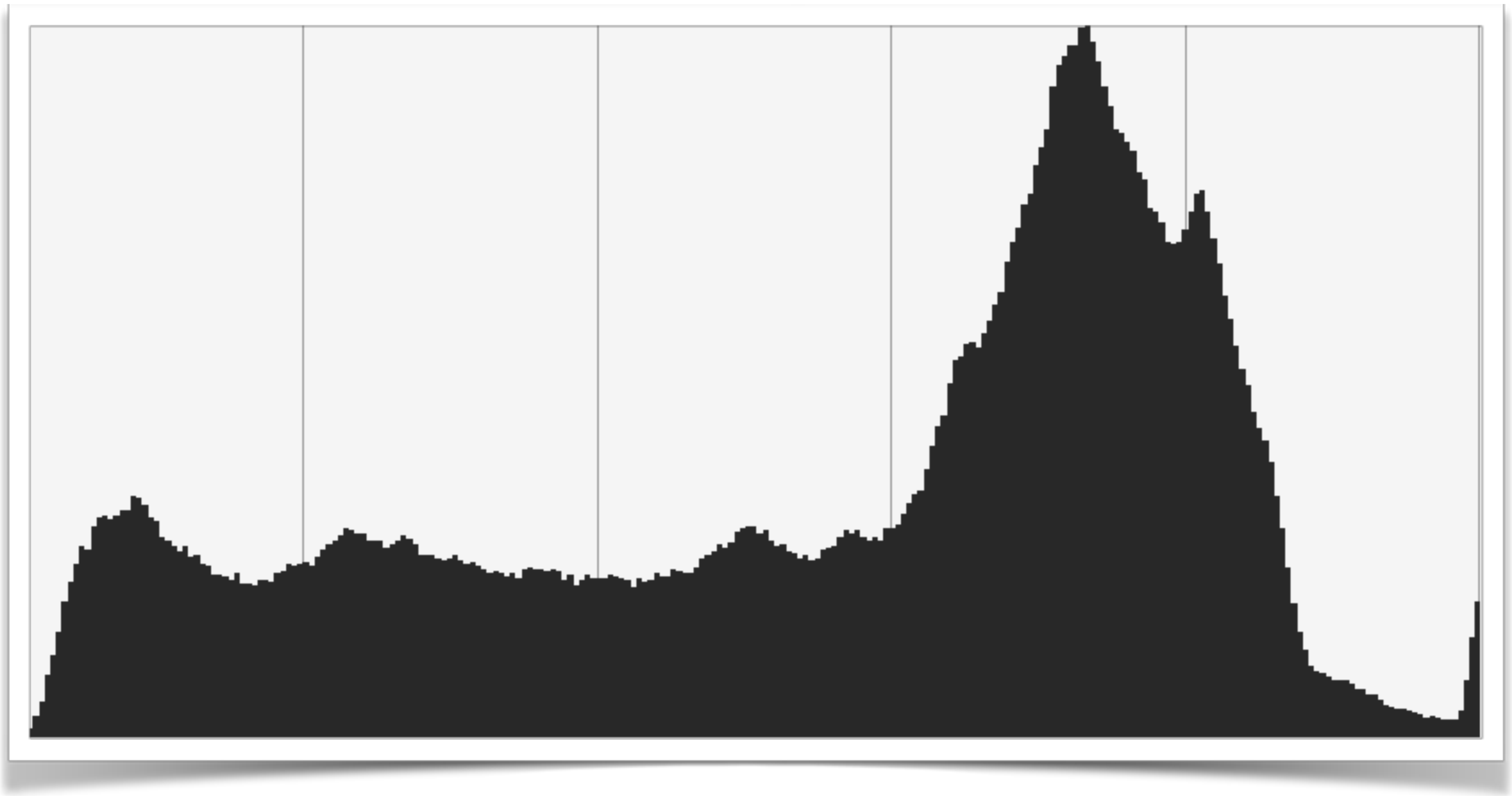
- The « Engineer »: *I need a distance function between histograms*
  - ➤ distance between (discrete) probability distributions
  - ➤ interpolated objects **MUST BE** histograms

- The « Dude »: *Given two probability measures μ and ν on some metric spaces X and Y, and a Borel-measurable cost function c(,) find the push-forward T such that :*

$$\inf \left\{ \int_X c(x, T(x))\, \mathrm{d}\mu(x) \;\middle|\; T_*(\mu) = \nu \right\}$$

Histograms

- Discrete estimate of a probability distribution: probability of a given pixel to have a value in a given range



$$\alpha = \sum_i \alpha_i \delta_{x_i}$$

- Operations on histograms (equalization, gamma correction…) = transfer functions between discrete probability distributions (result must integrate to one) => ***mass preserving transport***
- Distance between histograms => ***metric between probability distributions***

Optimal transport for short

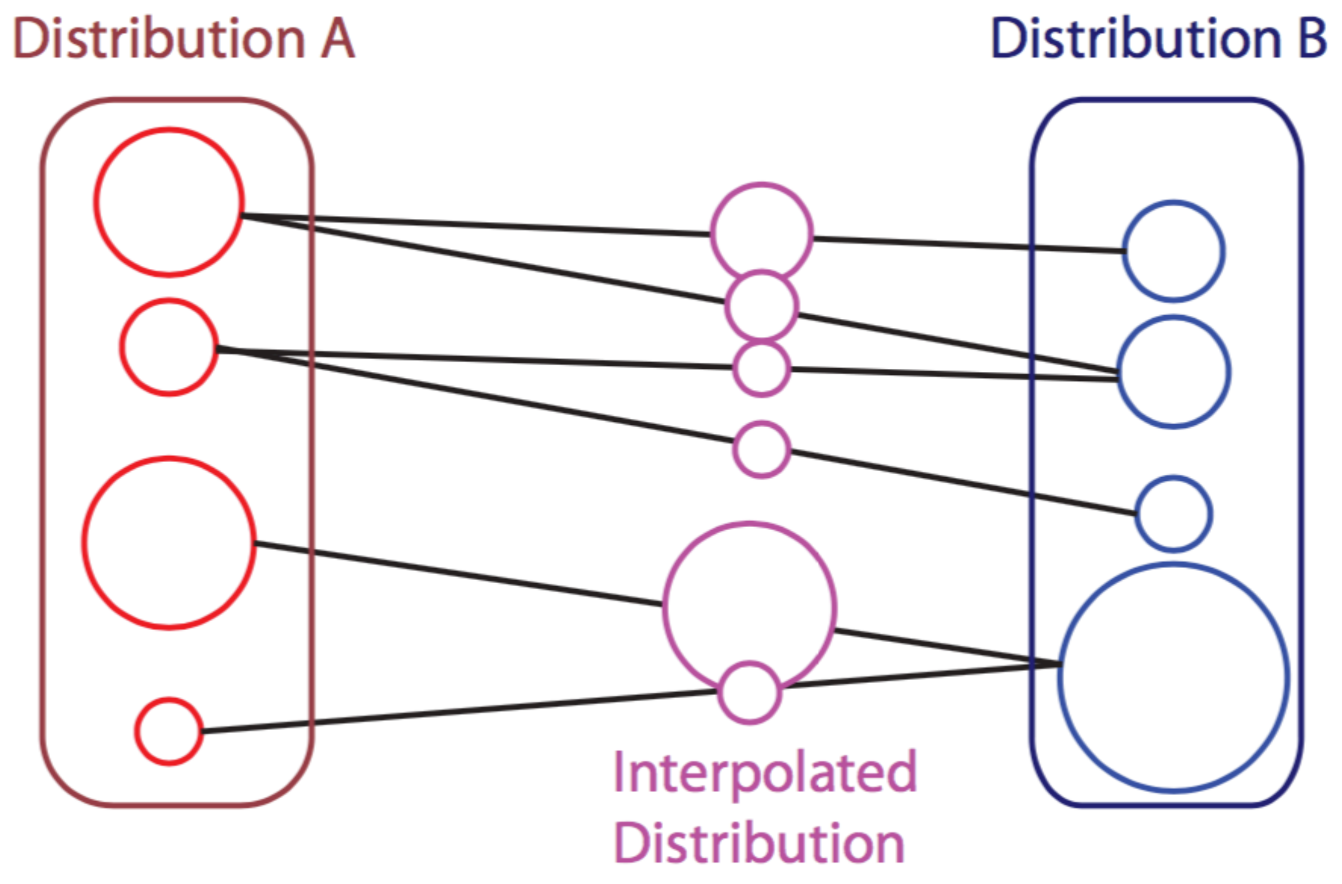
- Most efficient transport between probability measures (continuous, discrete, semi-continuous) on a domain (1-D, n-D, on surfaces, …) for a user-specified cost function

- Interpolated objects are probability measures

- The cost of an optimal transport defines a metric between distributions (Wasserstein metric), the optimal transport plan is a geodesic in the Wasserstein space

OT of Discrete Measures; assignment problem

$$\mu = \sum_{x \in X} \mu_x \delta_x \qquad \nu = \sum_{y \in Y} \nu_y \delta_y \qquad c : X \times Y \to \mathbb{R}$$



$$\min_{\pi} \sum_{x \in X} c(x, y) \pi_{xy}$$

$$\text{where } \pi : X \times Y \to \mathbb{R}^+ \text{ is s.t. } \begin{cases} \forall x \in X, \sum_{y \in Y} \pi_{xy} = \mu_x \text{ and} \\ \forall y \in Y, \sum_{x \in X} \pi_{xy} = \nu_x \end{cases}$$

Cost function : application dependent (Euclidean distance, ...)

Solving Discrete OT

- Linear programming approach (solve a linear system where the unknowns are π_{xy})
- Assignment problem (Hungarian algorithm)
- Bipartite graph matching
- Network simplex
- ….

1-D Case: easy ;)

- Closed form for 1-D measures (aka Earth mover distance) !

$$OT = \int_0^1 c(F_\mu^{-1}(s), F_\nu^{-1}(s)ds$$

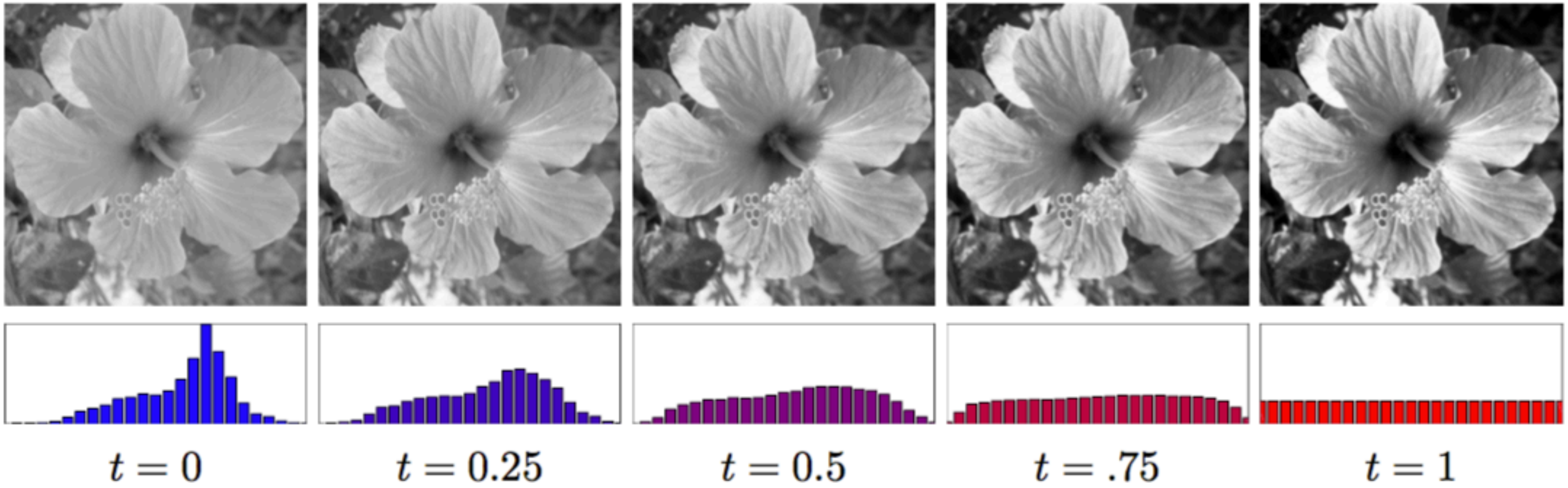
F_μ^{-1} and F_ν^{-1} *are the (pseudo) inverse of the cumulative distribution function of* μ *and* ν

- Wasserstein distance between histograms

$$W_p(\mu,\nu) = \left(\int_0^1 |F_\mu^{-1}(s) - F_\nu^{-1}(s)|^p ds\right)^{1/p}$$

- gné?

$$W_1(h_1, h_2) = \sum_{i=0}^{255} |F_{h_1}(i) - F_{h_2}(i)|$$


$t = 0$
$t = 0.25$
$t = 0.5$
$t = .75$
$t = 1$

Continuous OT and Wasserstein Space

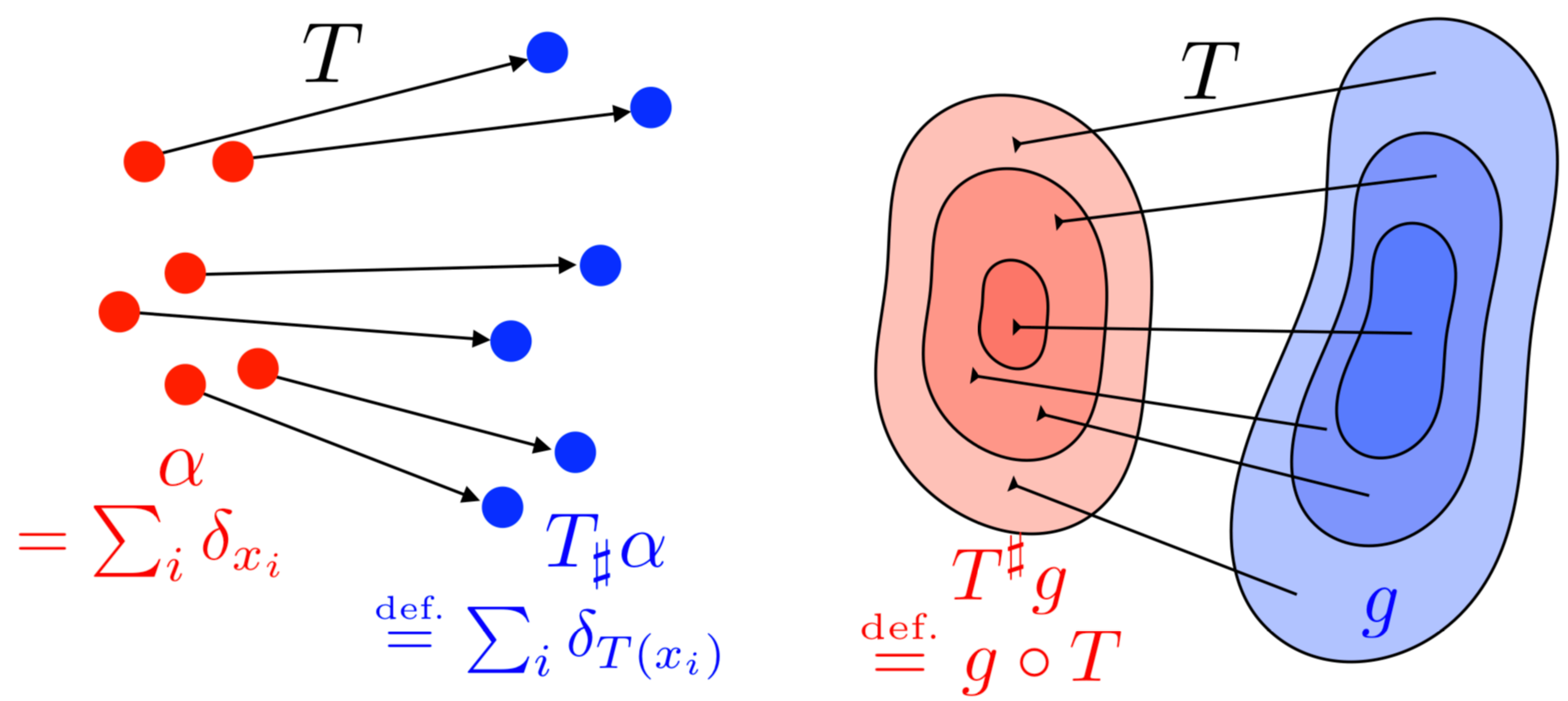
- Huge literature: Monge, Ampere, Kantorovich, Wasserstein, Benamou, Brenier, Vilani…

$$W_p(\mu,\nu) := \left(\inf_{\gamma \in \Gamma(\mu,\nu)} \int_{M \times M} d(x,y)^p \, \mathrm{d}\gamma(x,y) \right)^{1/p}$$

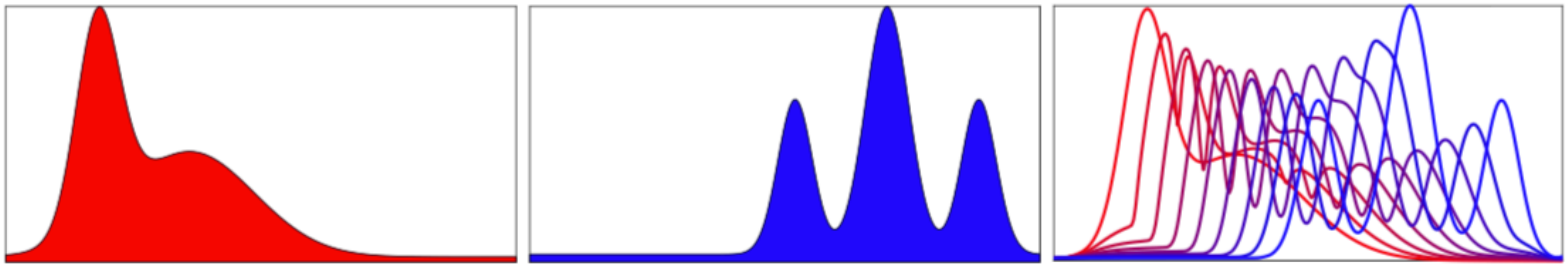
induces a metric space on probability distributions !

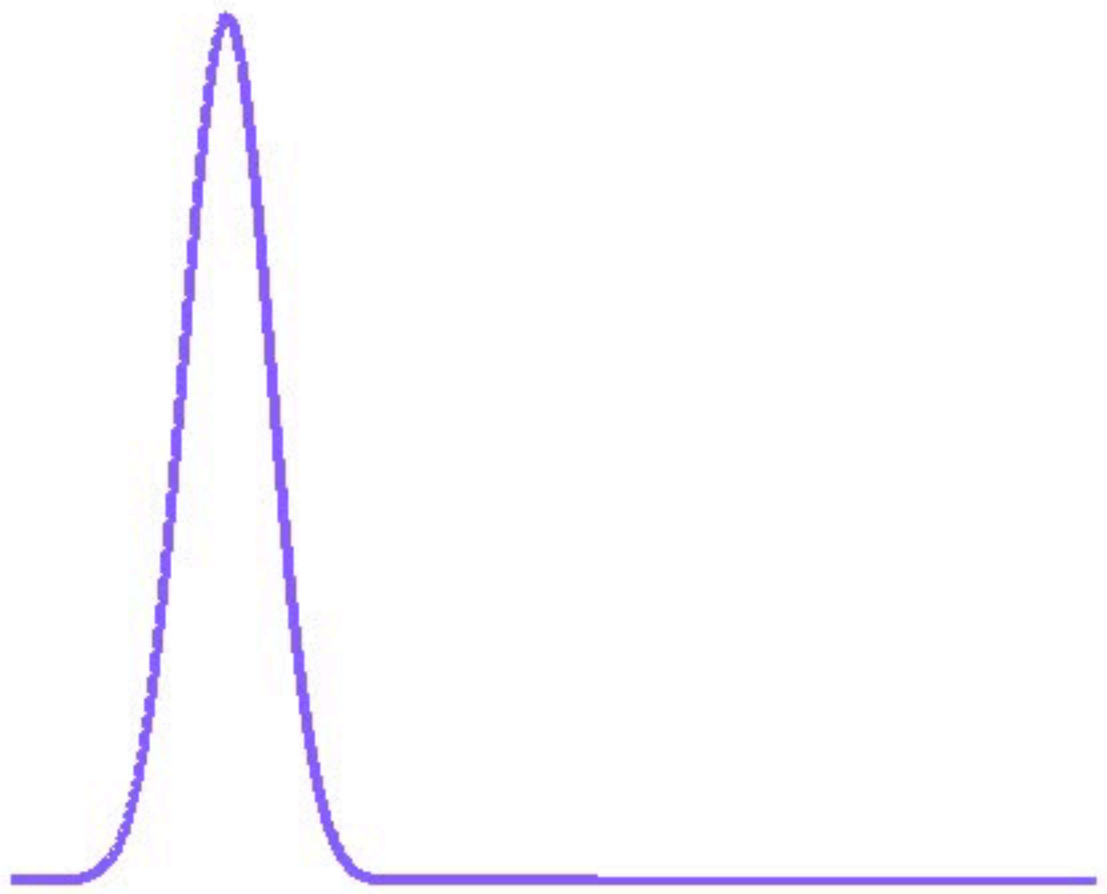
- Variational formulation (PDE,…)

$$\mathcal{W}(\mu_0,\mu_1) = \begin{cases} \inf_J \int_M \|J(x)\| \, dx \\ \text{s.t. } \nabla \cdot J(x) = \rho_1(x) - \rho_0(x) \\ \quad J(x) \cdot n(x) = 0 \ \forall x \in \partial M \end{cases}$$

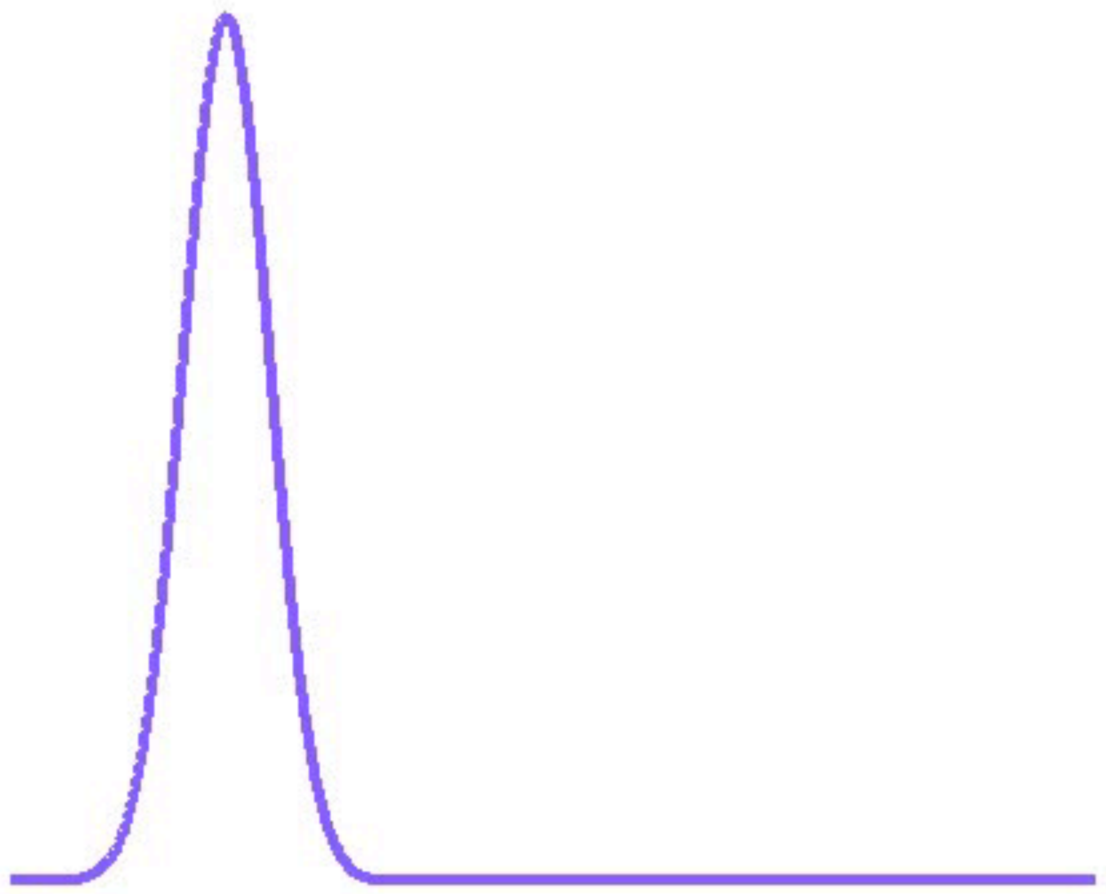


Interpolation / Advection



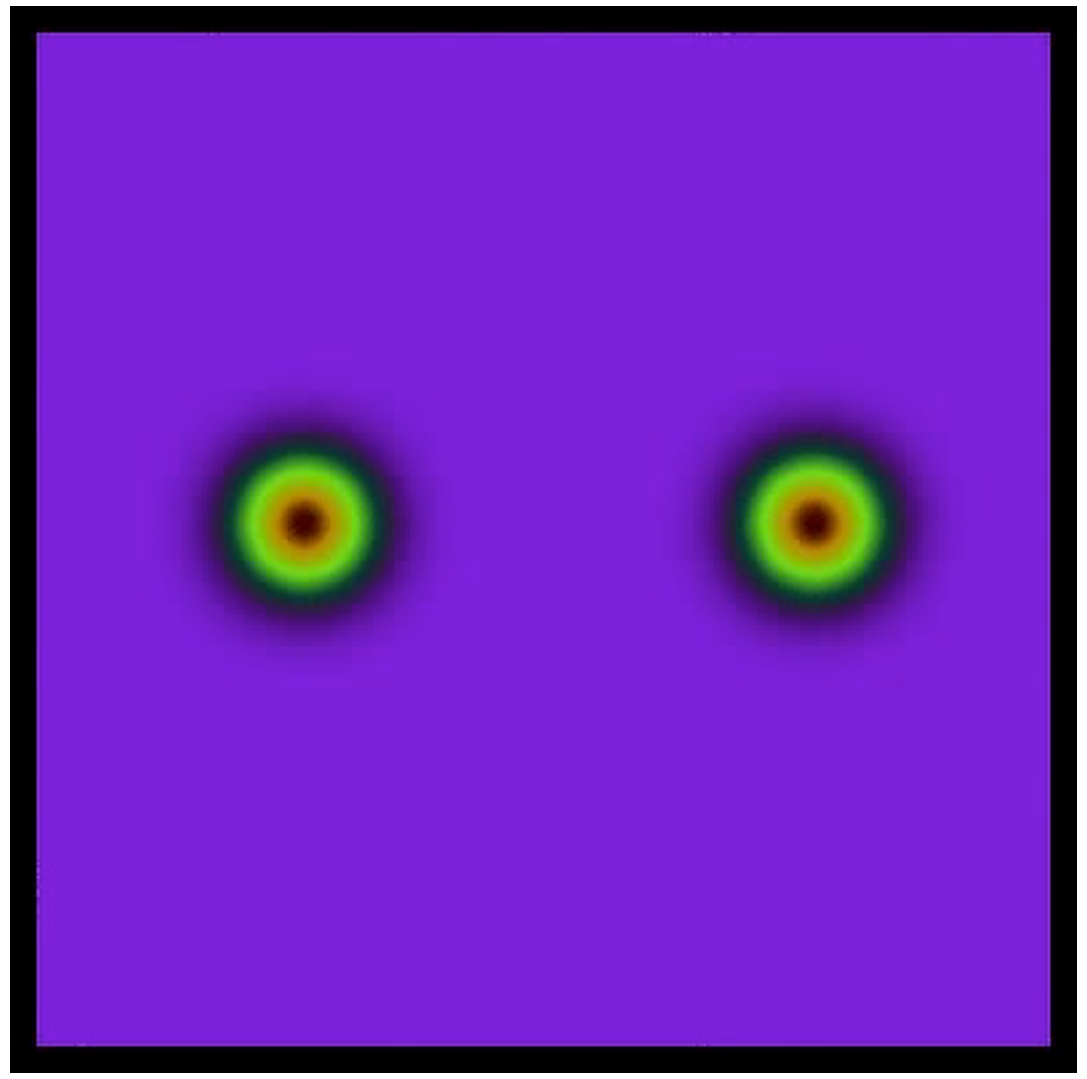


linear interpolation

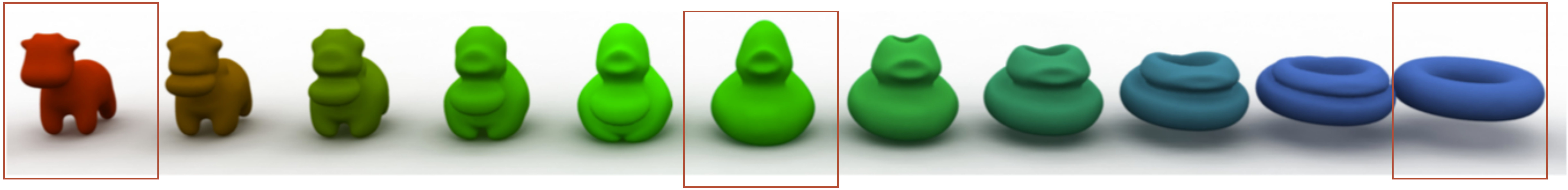


OT

Transport between two 2-D Gaussian functions



Transport between 3-D histograms

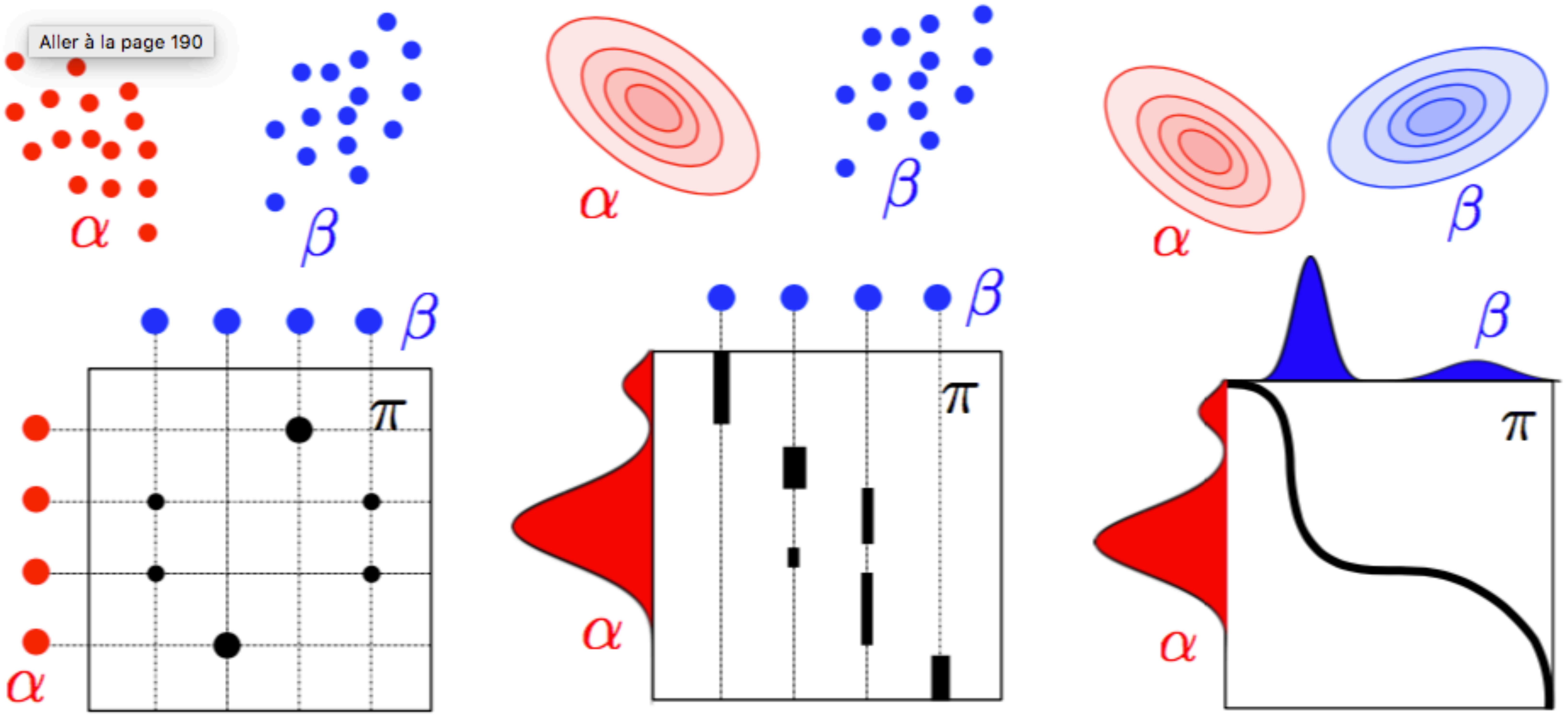


3D objects as 3-D histograms (100x100x100 volumetric grid) and regularized OT

OT on surfaces

All defs remain the same (X = Y = smooth manifold) !

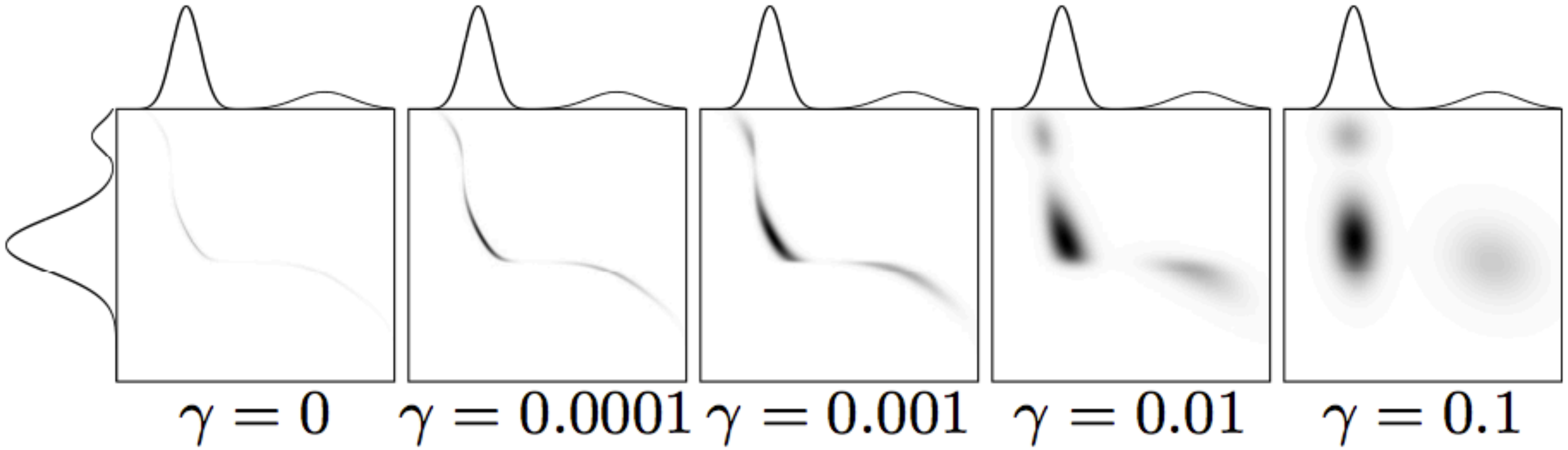
Wrap-up: Discrete / Semi-discrete / Continuous OT



Focus 1: Entropy regularization aka Sinkhorn

- OT + Entropy term to regularized the transport

$$W(p,q) \stackrel{\text{def.}}{=} \min_{T \in \mathbb{R}_+^{N \times N}} \left\{ \langle T, C \rangle + \gamma H(T) \,;\, T\mathbb{1} = p, T^\top \mathbb{1} = q \right\},$$



- Simple algorithm : iterate over sums of convolved columns / rows of π + element wise division
- fast convergence, numerically stable, and can be auto-diff !

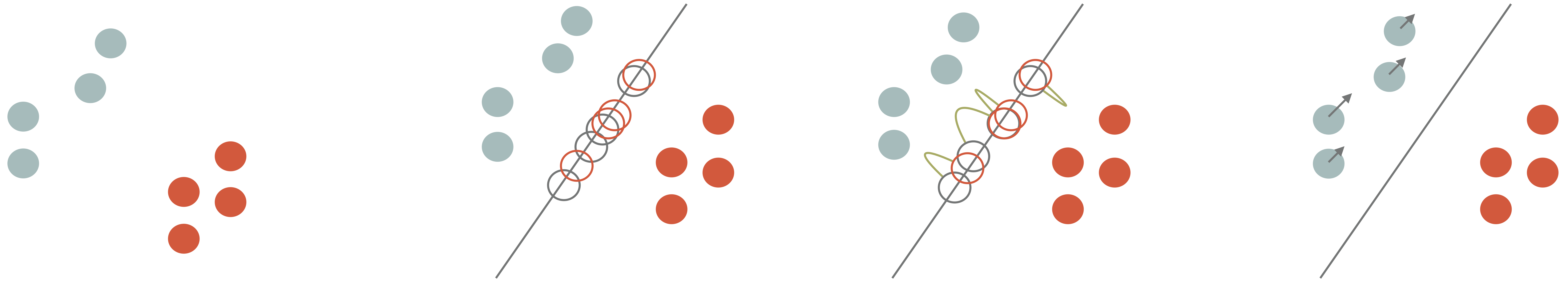
Focus 2: Sliced Optimal Transport 👍

$$SW_2(\mu_f, \mu_g)^2 = \int_{\mathbb{S}^{d-1}} W_2(\mu_{f_\theta}, \mu_{g_\theta})^2 \, d\theta$$

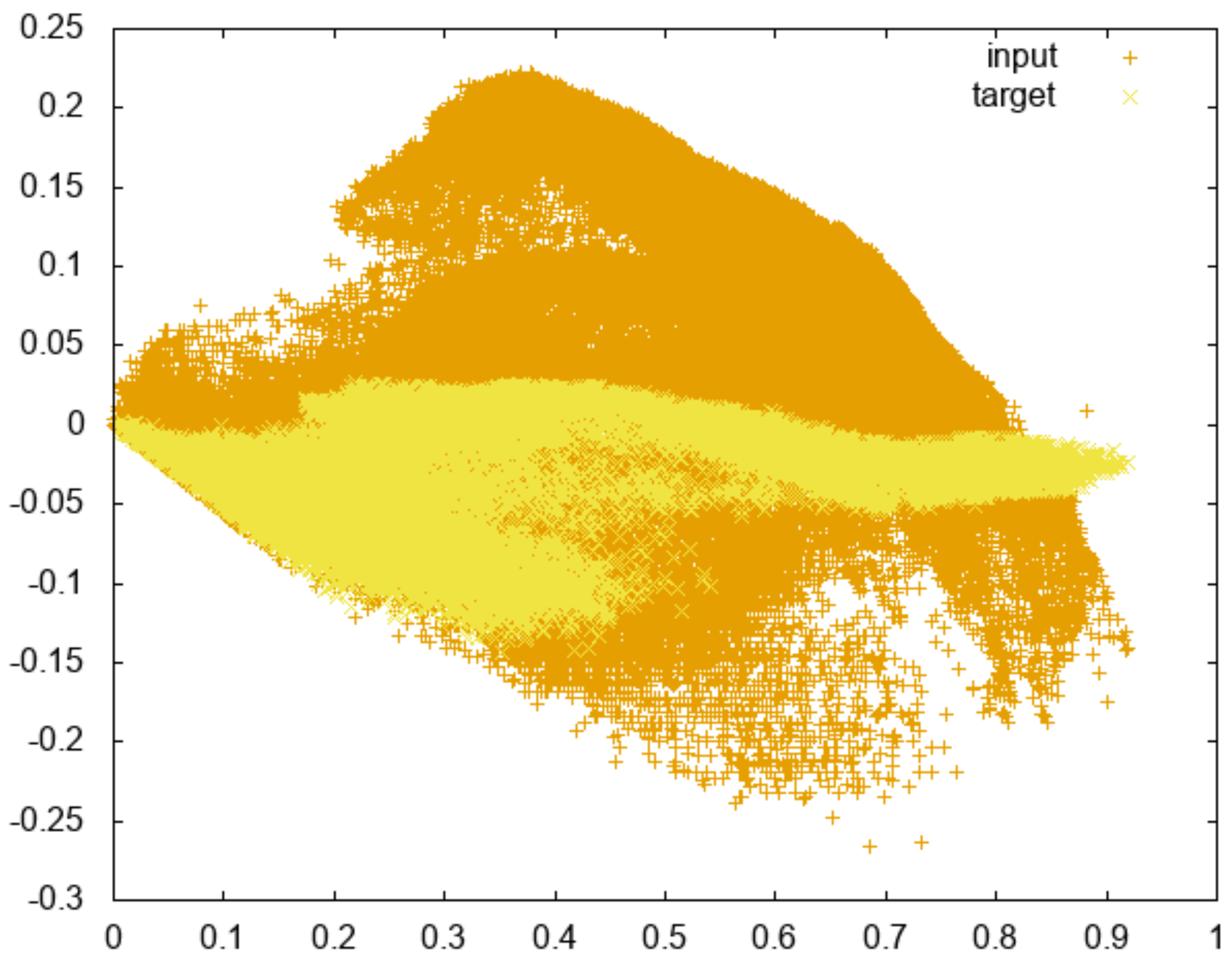
1-D OT = difference between CDFs
=
2 sorts of the Diracs + pairing

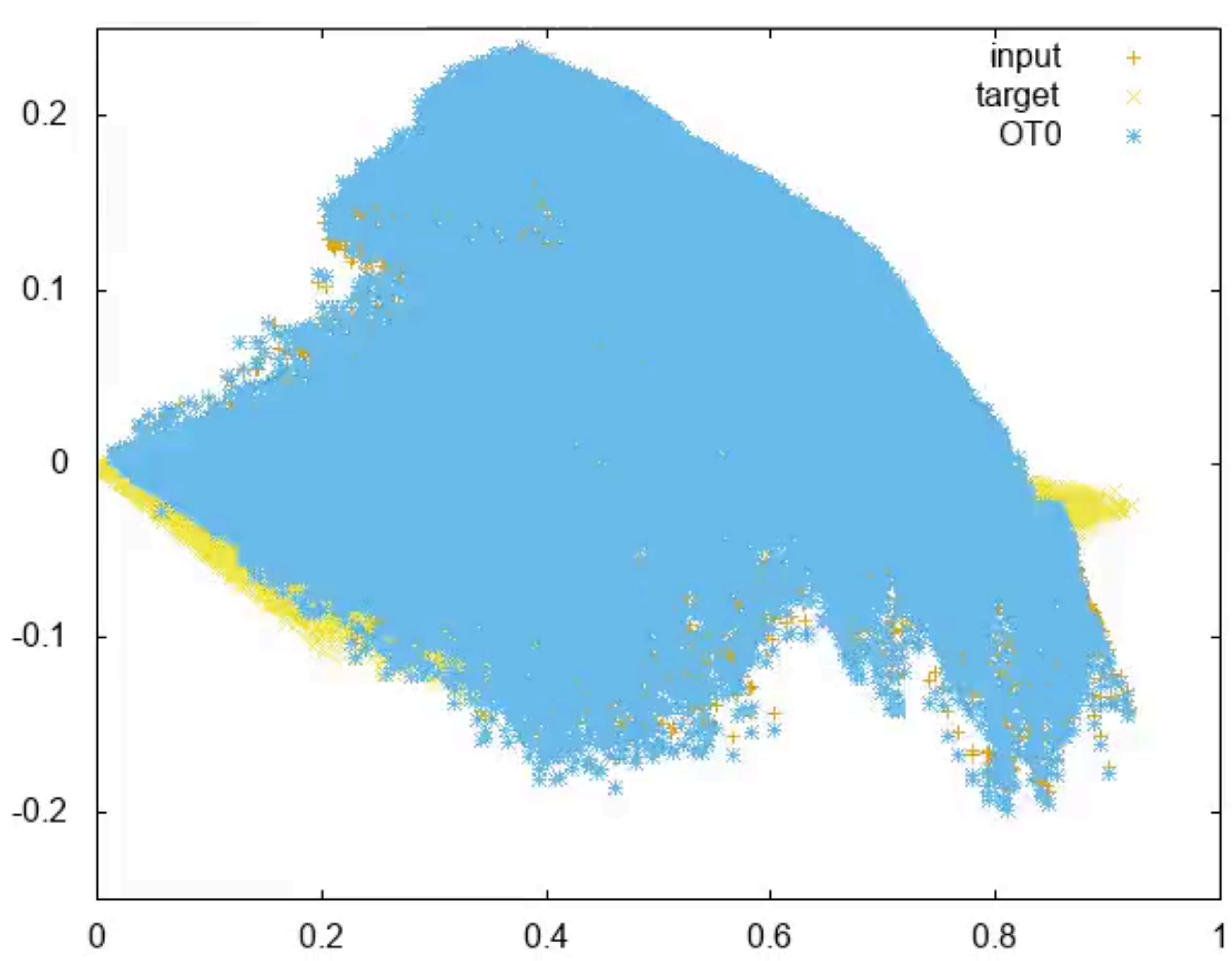
- **Simple Idea**: given discrete measures (aka sum of Diracs) μ and ν in R^d

    1. Project them onto a random 1-D line with direction d
    2. Solve 1-D OT problem and move samples in μ in the direction d to get closer to ν (aka just a sort on the projections + coupling)
    3. goto step 1



Easy impelem., GPU friendly, accurate approximation …


input
target
0.25
0.2
0.15
0.1
0.05
0
-0.05
-0.1
-0.15
-0.2
-0.25
-0.3
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


input
target
OT0
0.2
0.1
0
-0.1
-0.2
0
0.2
0.4
0.6
0.8
1

Next steps

- sorts + *std::tranform à la* STL on GPU (boost::compute)
- k-means clustering + OT on the k seeds + advections of clusters
    - CPU demo: 2.6M samples sorted, 200ms per iteration —this laptop—
- Regularized 1-D OT

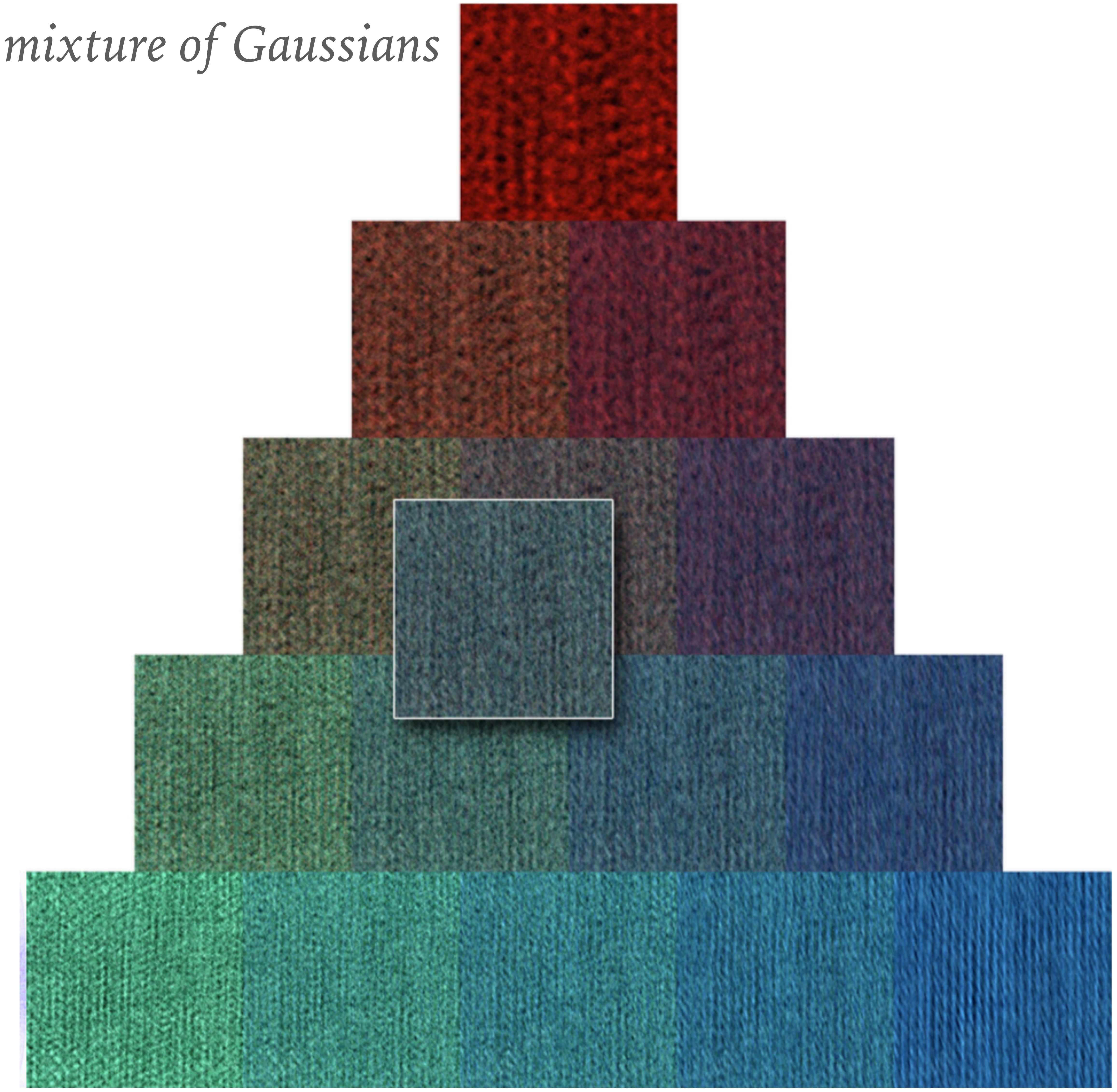
Applications

Color transfert (demo)

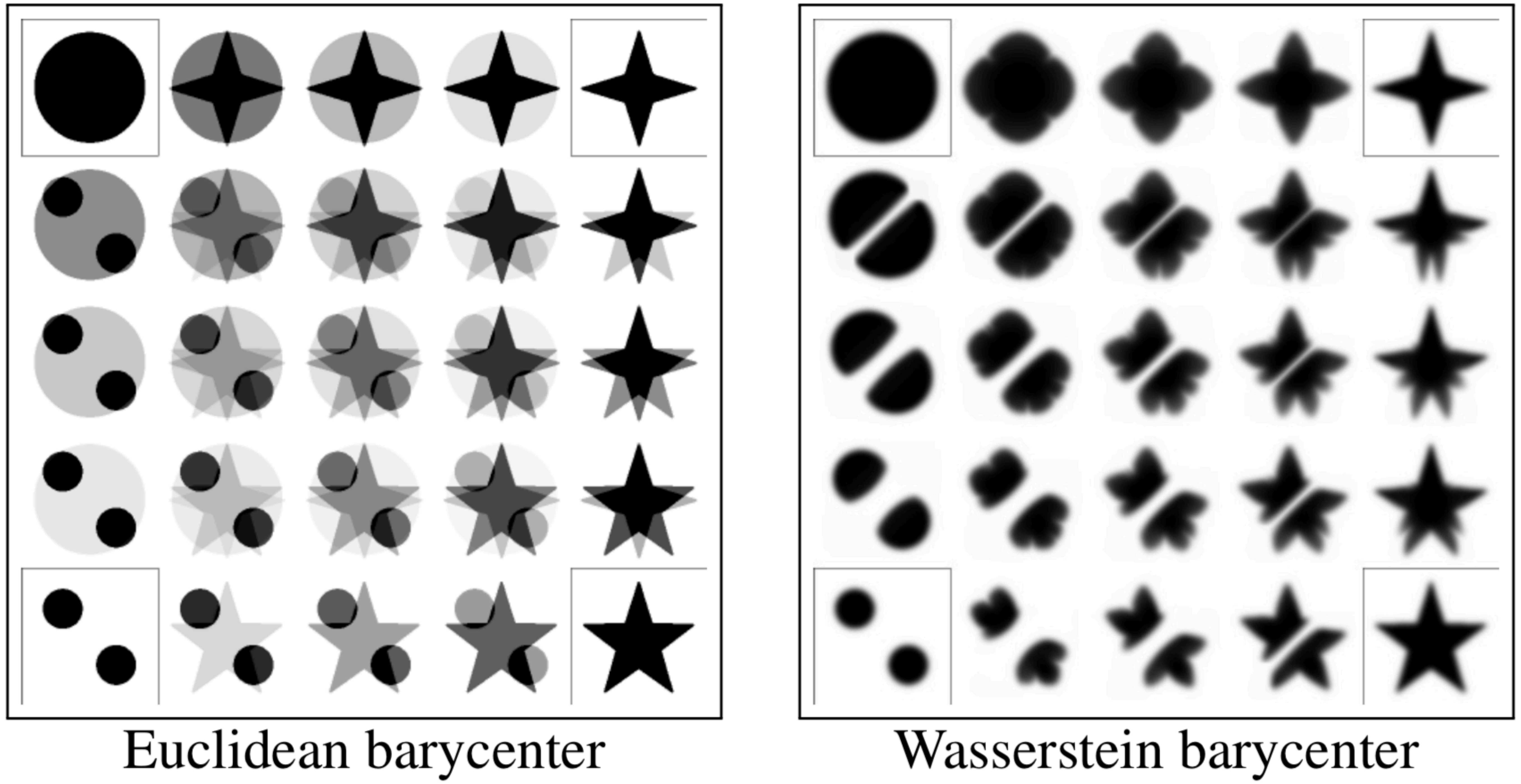
Color transfert + Wasserstein Barycenters (interpolate between several measures)

Texture transfert (micro-texture with sparse Fourier content)

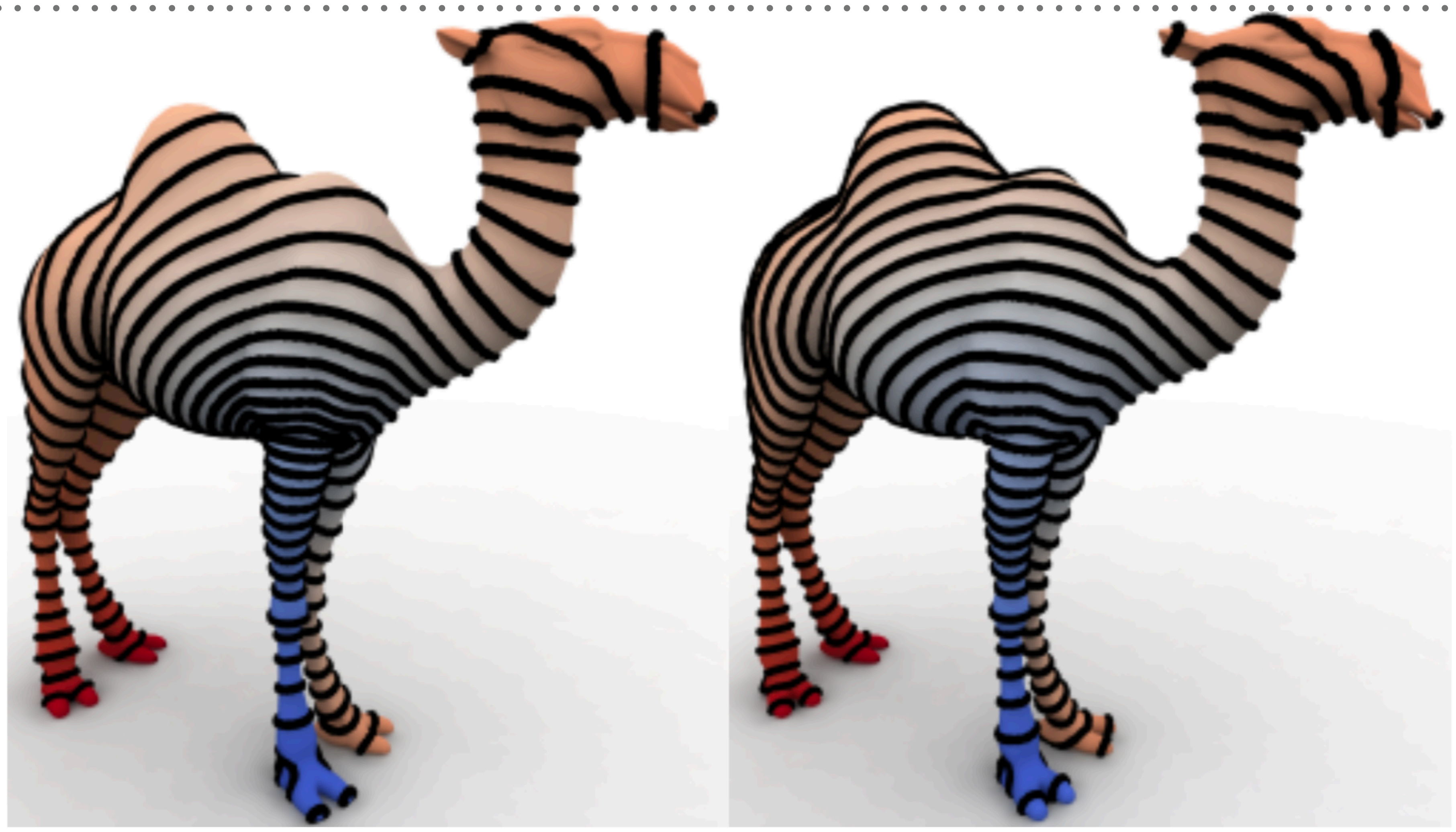
Input: Power spectrum as mixture of Gaussians



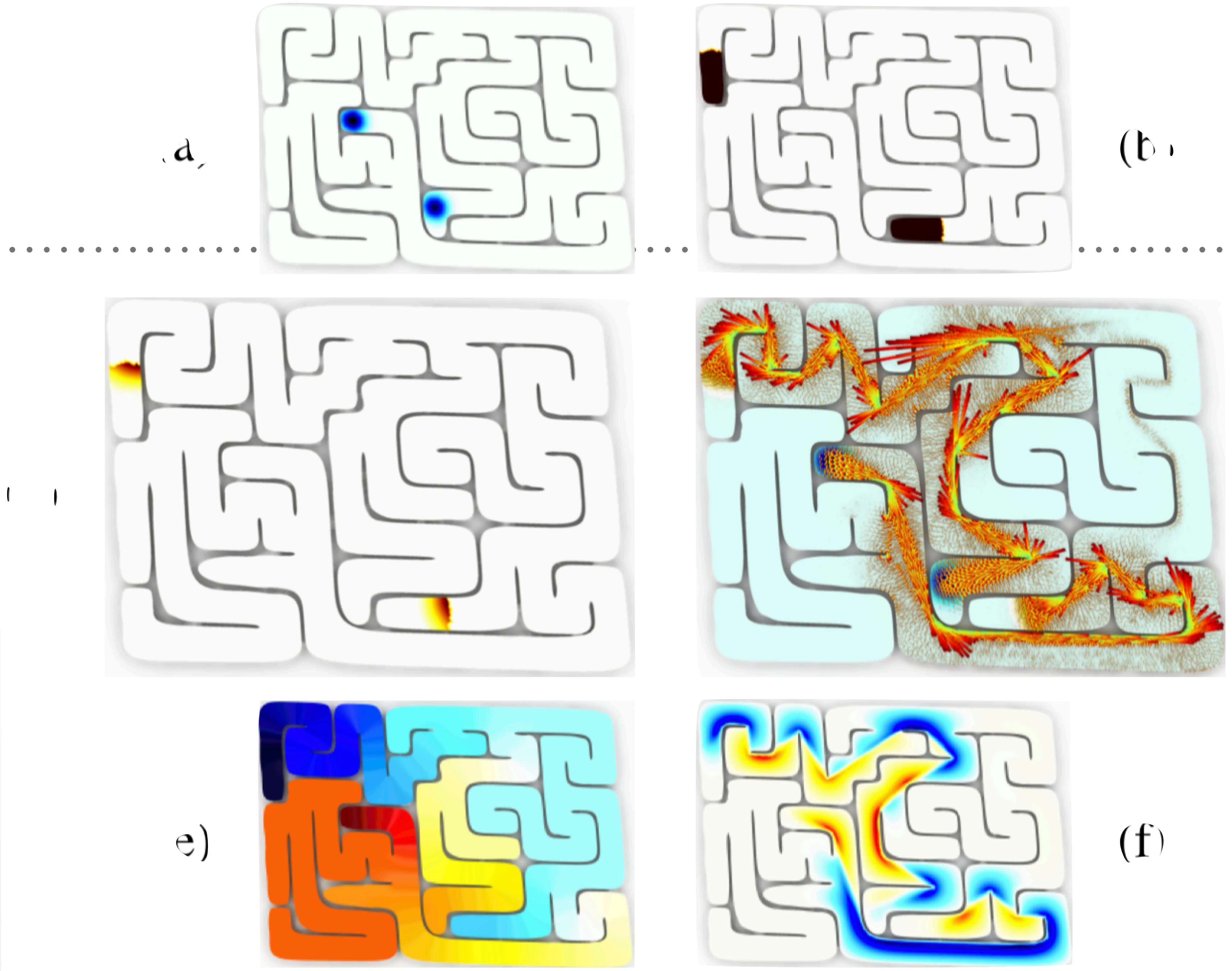
Maybe a bit strange for geometry interpolation…

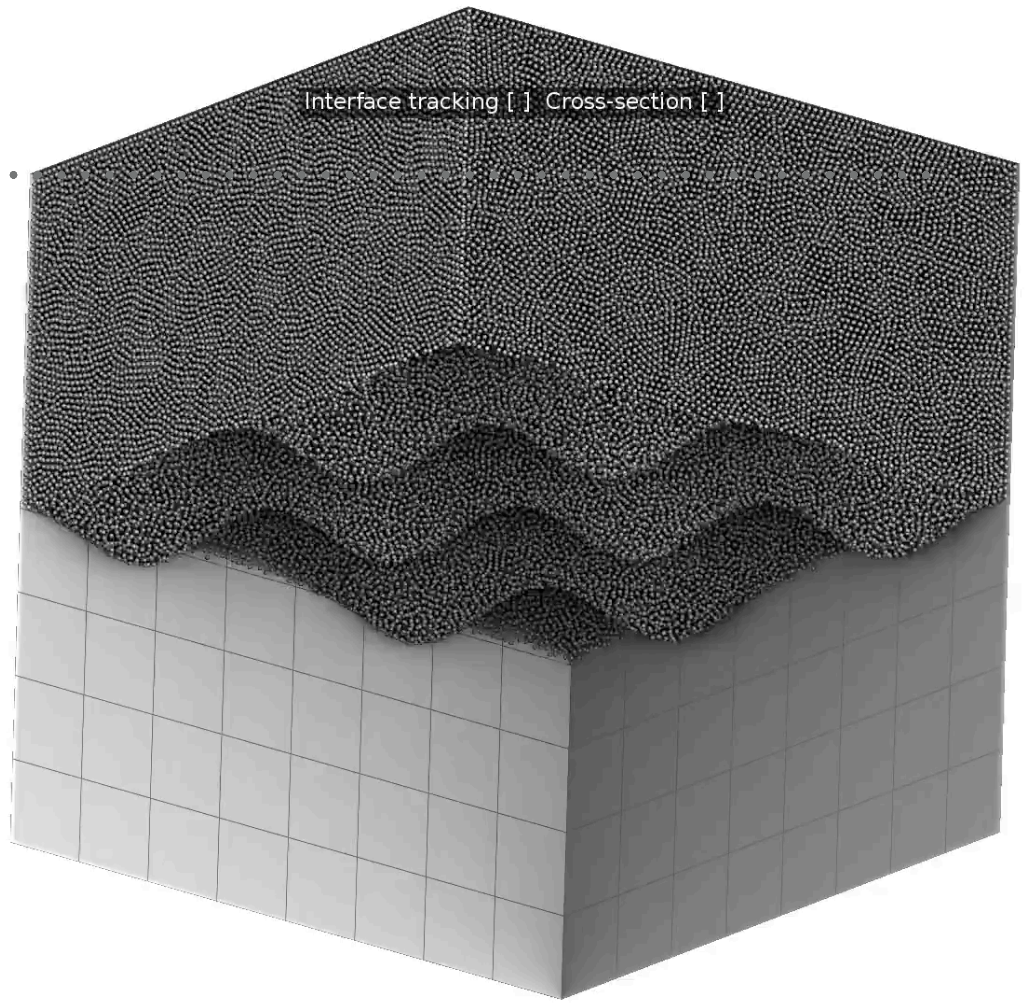


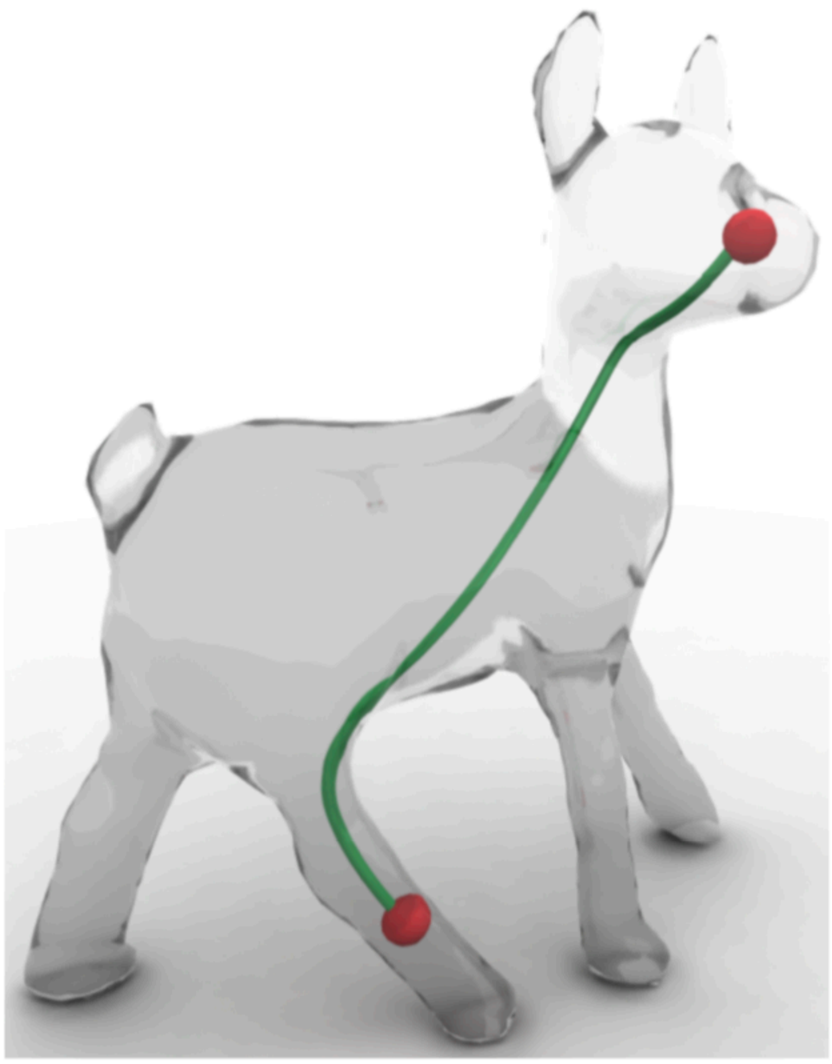
. . . but very useful

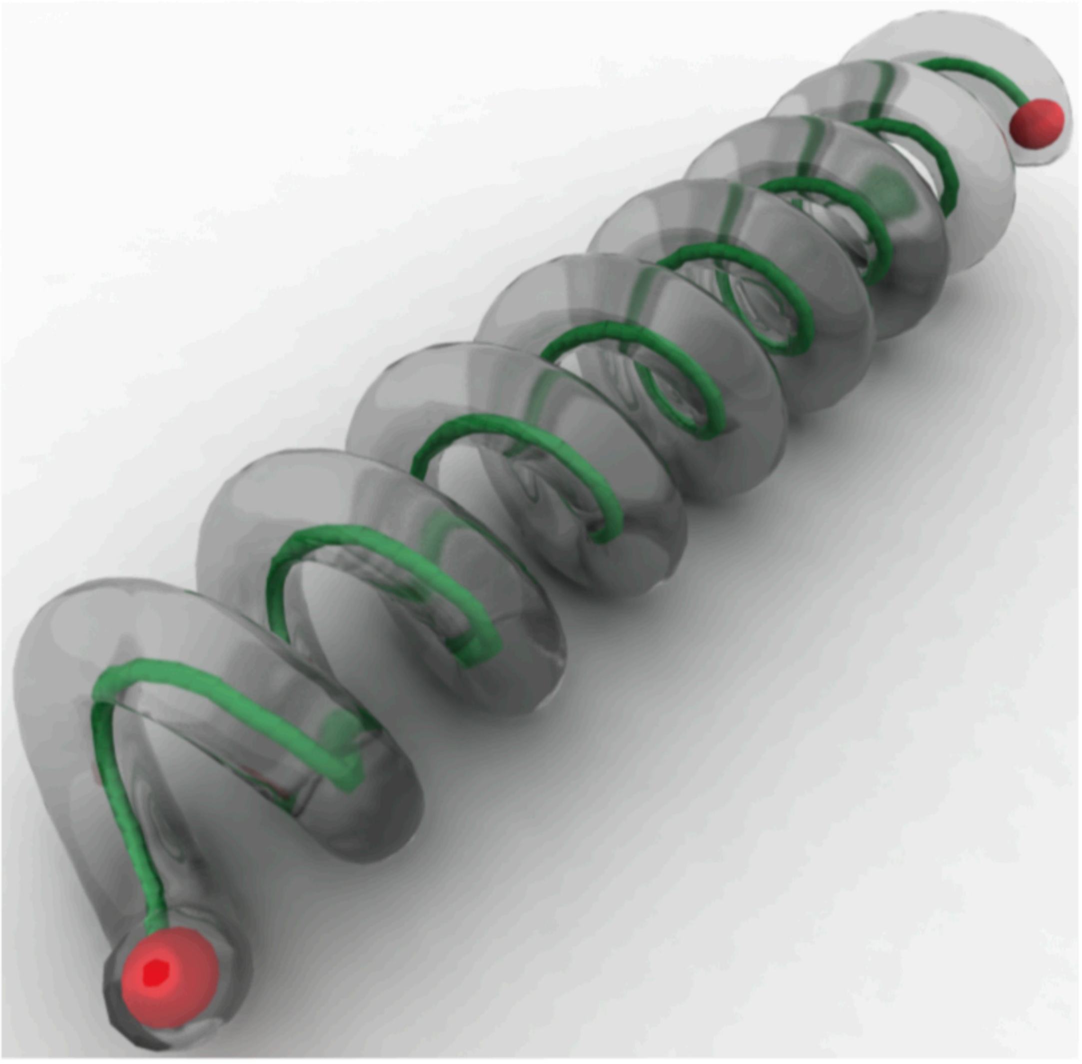


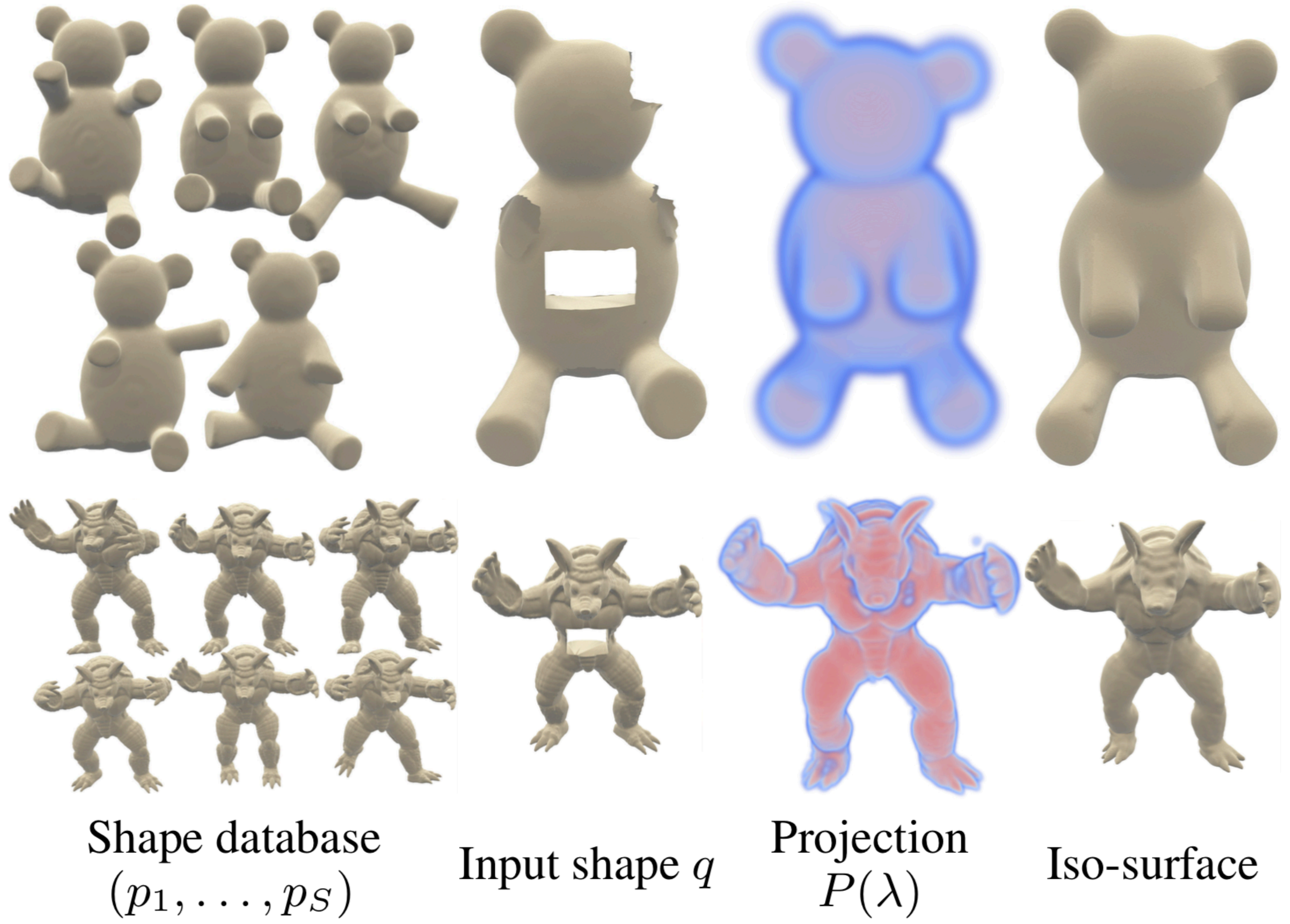
$d_{\mathcal{W}}^{0}$ $d_{\mathcal{W}}^{100}$

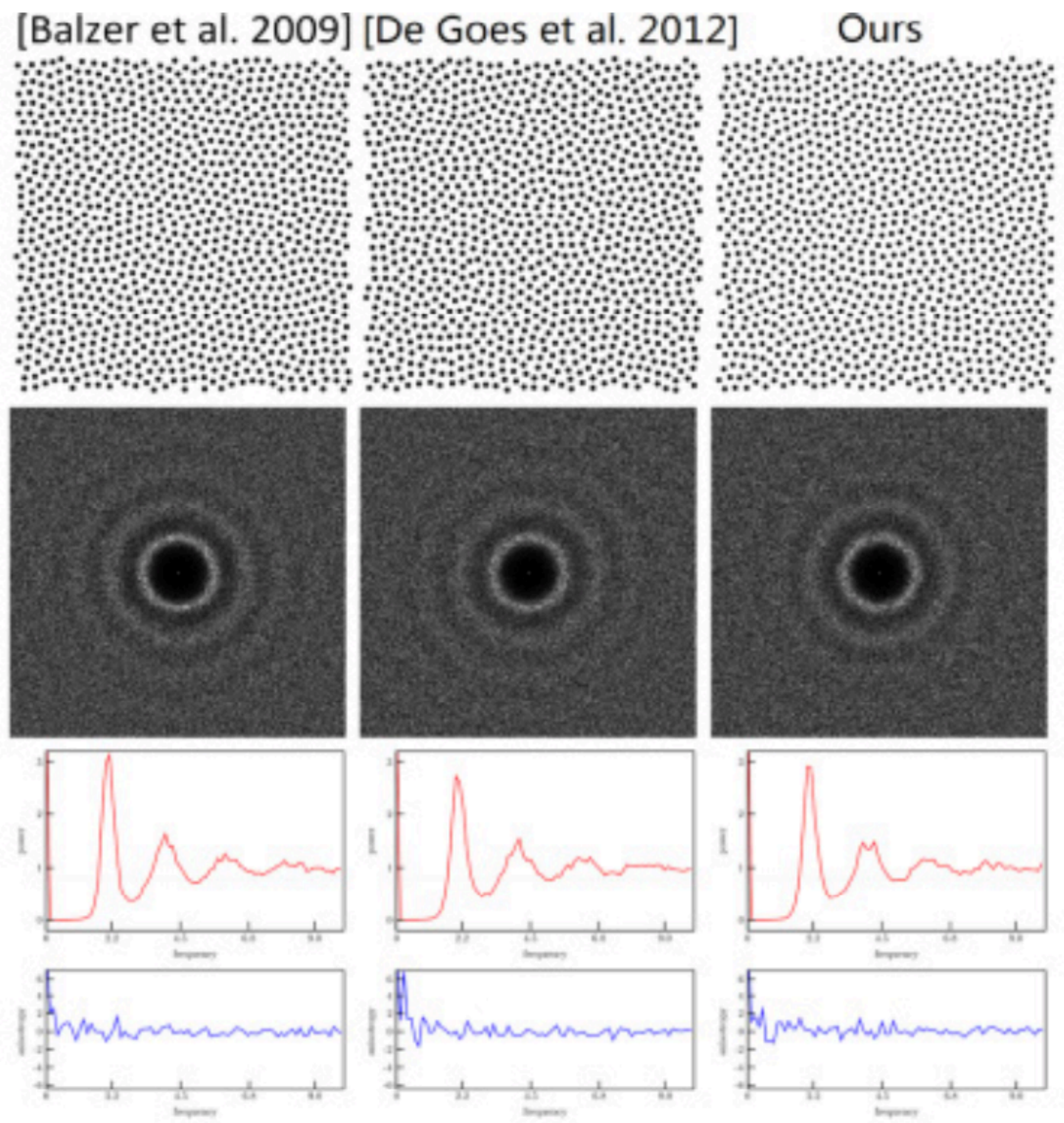




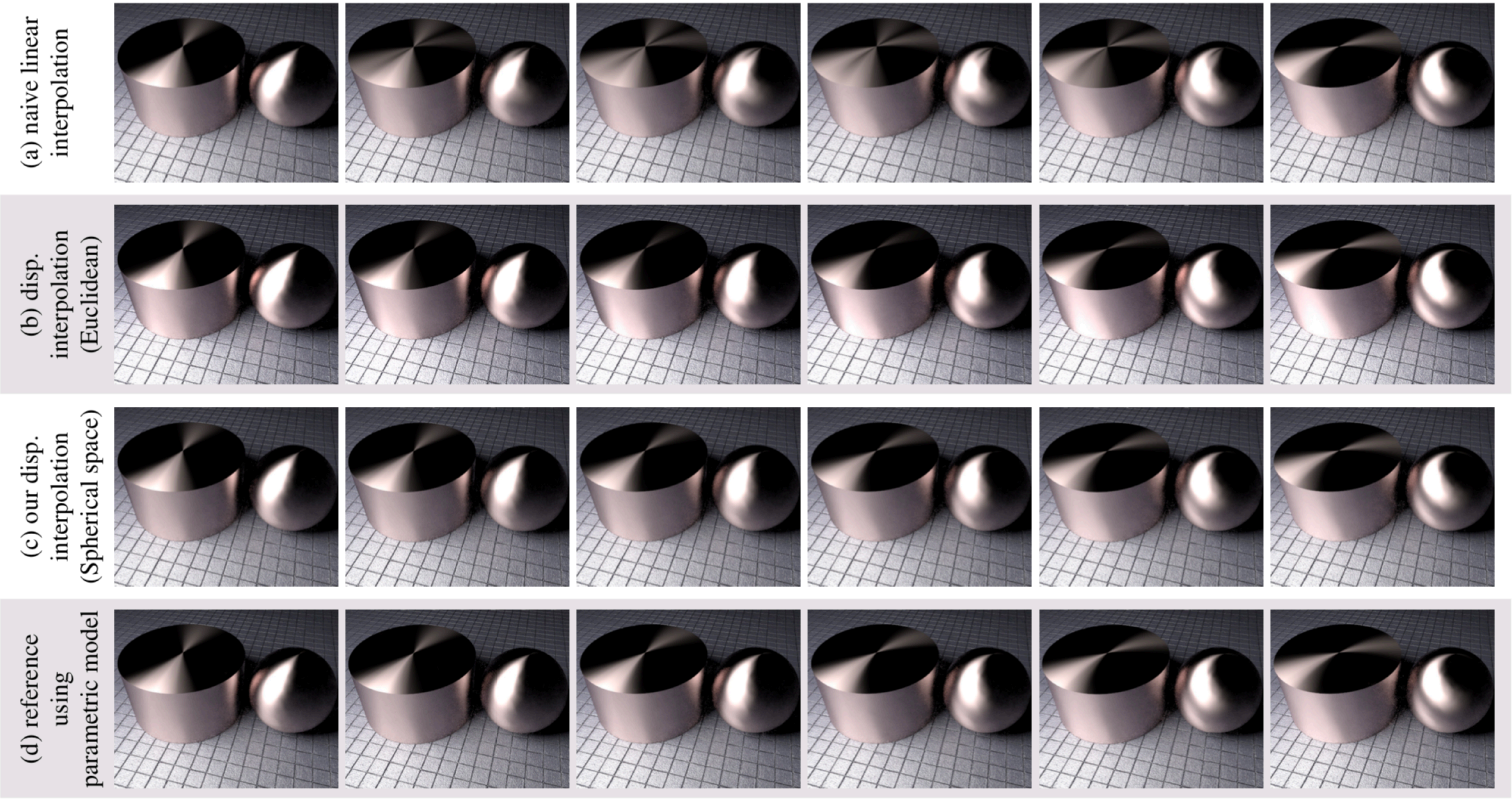








BRDF interpolation



One more things

Wasserstein based non-linear dictionary learning

- Wasserstein barycenter: given a set of atoms and an input image, find the barycentric coordinates minimizing the OT distance between the input image and the barycentric one

$$P : \lambda \mapsto P(\lambda) \stackrel{\text{def.}}{=} \underset{p \in \Sigma_N}{\operatorname{argmin}} \sum_s \lambda_s W(p, p_s).$$

$$\underset{\lambda \in \Sigma_S}{\operatorname{argmin}} \, \mathcal{E}(\lambda), \quad \textit{where } \mathcal{E}(\lambda) \stackrel{\text{def.}}{=} \mathcal{L}(P(\lambda), q).$$

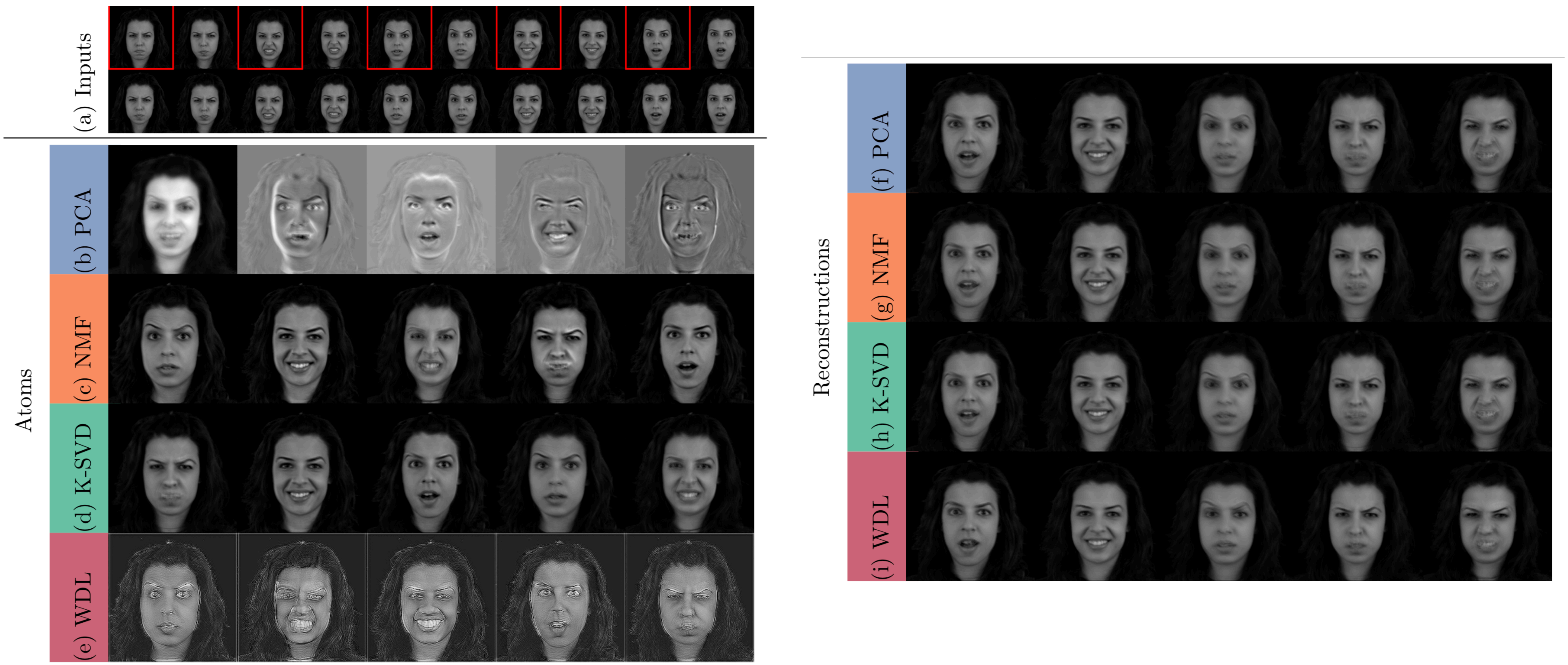
$$\nabla \mathcal{E}_L(\lambda) :$$

- Dictionary learning: given a set of N distributions, find the K atoms that represent the N distributions in the best « Wasserstein » way

$$\min_{D,\Lambda} \mathcal{E}(D, \Lambda) := \sum_{i=1}^{M} \mathcal{L}\left(P(D, \lambda_i), x_i\right).$$

$$\nabla_D \mathcal{E}_L(D, \Lambda) = \left[\partial_D P^{(L)}(D, \lambda)\right]^\top \nabla \mathcal{L}(P^{(L)}(D, \lambda), x)$$

$$\nabla_\lambda \mathcal{E}_L(D, \Lambda) = \left[\partial_\lambda P^{(L)}(D, \lambda)\right]^\top \nabla \mathcal{L}(P^{(L)}(D, \lambda), x).$$


(a) Inputs
Atoms
(b) PCA
(c) NMF
(d) K-SVD
(e) WDL
Reconstructions
(f) PCA
(g) NMF
(h) K-SVD
(i) WDL

Conclusion

- Many computer graphics problems can be expressed as
    - distance between distributions,
    - geodesics between distributions in Wasserstein spaces
    - ..

Color, mixture of Gaussian (spatial or spectral domain), …

- Optimal Transport is **THE** tool
- A lot of very nice algorithms (linear programming, graph based, Radon, Sliced, Sinkhorn,…)
- Many extensions (regularized OT, unbalanced, heavyball, …)

- Gabriel Peyré and Marco Cuturi, Computational Optimal Transport, ArXiv:1803.00567, 2018.
- M. A. Schmitz, M. Heitz, N. Bonneel, F. M. Ngolè Mboula, D. Coeurjolly, M. Cuturi, G. Peyré, J.-L. Starck., Wasserstein Dictionary Learning: Optimal Transport-based unsupervised non-linear dictionary learningSIAM Journal on Imaging Sciences, 11(1), 2018. (4)
- Nicolas Bonneel, Gabriel Peyré, Marco Cuturi, Wasserstein Barycentric Coordinates: Histogram Regression Using Optimal Transport, ACM Transactions on Graphics (SIGGRAPH), July 2016
- Solomon, Justin, Fernando de Goes, Gabriel Peyré, Marco Cuturi, Adrian Butscher, Andy Nguyen, Tao Du, and Leonidas Guibas. "Convolutional Wasserstein Distances: Efficient Optimal Transportation on Geometric Domains." SIGGRAPH 2015,
- Solomon, Justin, Raif Rustamov, Leonidas Guibas, and Adrian Butscher. "Earth Mover's Distances on Discrete Surfaces." SIGGRAPH 2014,
- Nicolas Bonneel, Julien Rabin, Gabriel Peyré, Hanspeter Pfister , Sliced and Radon Wasserstein Barycenters of Measures, Journal of Mathematical Imaging and Vision (JMIV), 2015
- Bonneel, N., van de Panne, M., Paris, S., Heidrich, W.: Displacement interpolation using lagrangian mass transport. ACM Trans. Graph. (SIGGRAPH ASIA'11) 30(6), 1–12 (2011)
- Pitié, F., Kokaram, A.C., Dahyot, R.: N-Dimensional probability density function transfer and its application to color transfer. In: Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on, vol. 2, pp. 1434–1439. (2005)
- Qin, Hongxing, et al. "Wasserstein blue noise sampling." ACM Transactions on Graphics (TOG) 36.5 (2017): 168.
- Thomas Gallouët, Quentin Mérigot, A Lagrangian scheme for the incompressible Euler equation using optimal transport, ArXiv, arXiv:1605.00568, 2016