Numerical Optimal Transport

http://optimaltransport.github.io

Entropic Regularization

Gabriel Peyré

www.numerical-tours.com

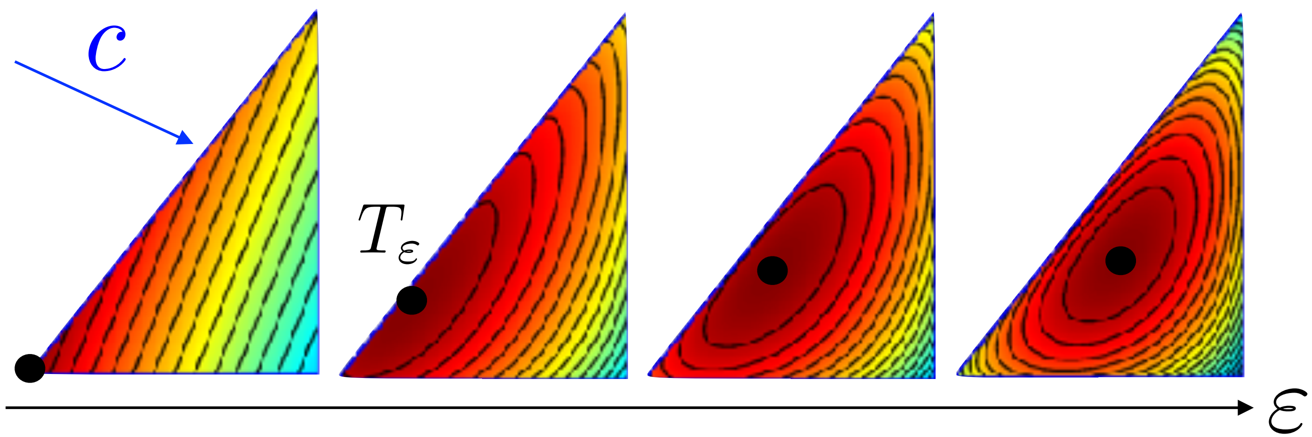
Overview

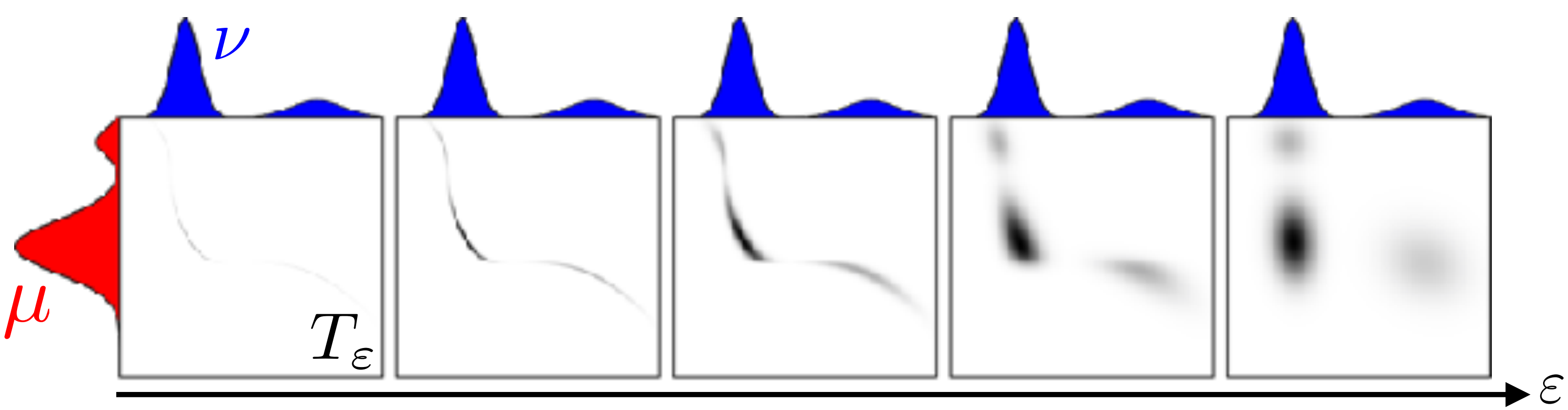
- Entropic Regularization and Sinkhorn
- Convergence Analysis
- Barycenters
- Generalized Sinkhorn and Applications (dedicated section)

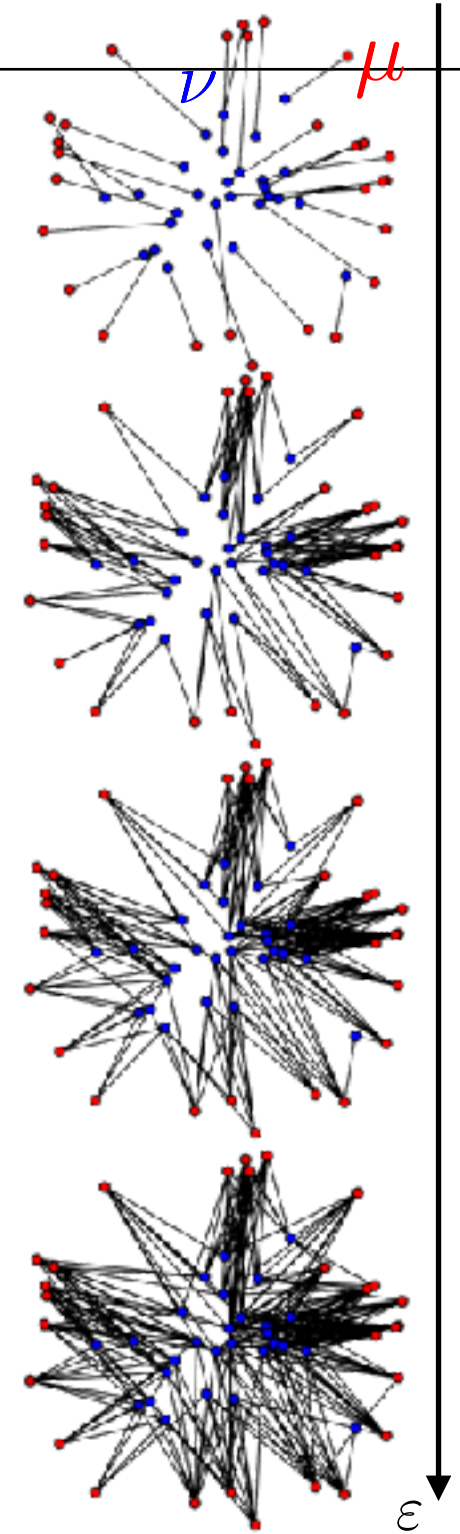
Entropic Regularization

Entropy: $\mathbf{H}(\mathbf{P}) \overset{\text{def.}}{=} -\sum_{i,j} \mathbf{P}_{i,j}(\log(\mathbf{P}_{i,j}) - 1)$

$$\mathrm{L}^{\varepsilon}_{\mathbf{C}}(\mathbf{a}, \mathbf{b}) \overset{\text{def.}}{=} \min_{\mathbf{P} \in \mathbf{U}(\mathbf{a},\mathbf{b})} \langle \mathbf{P}, \mathbf{C} \rangle - \varepsilon \mathbf{H}(\mathbf{P})$$



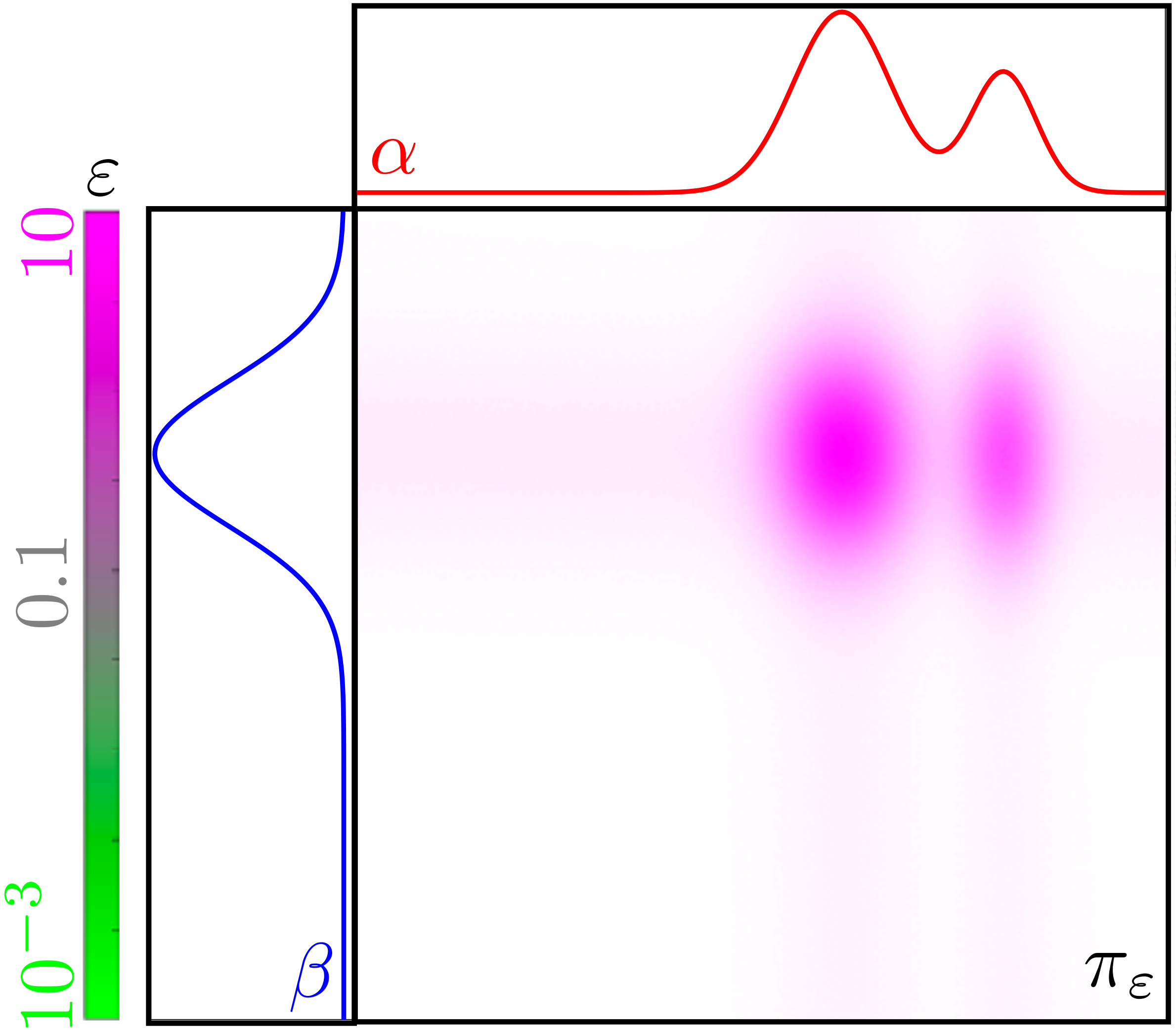


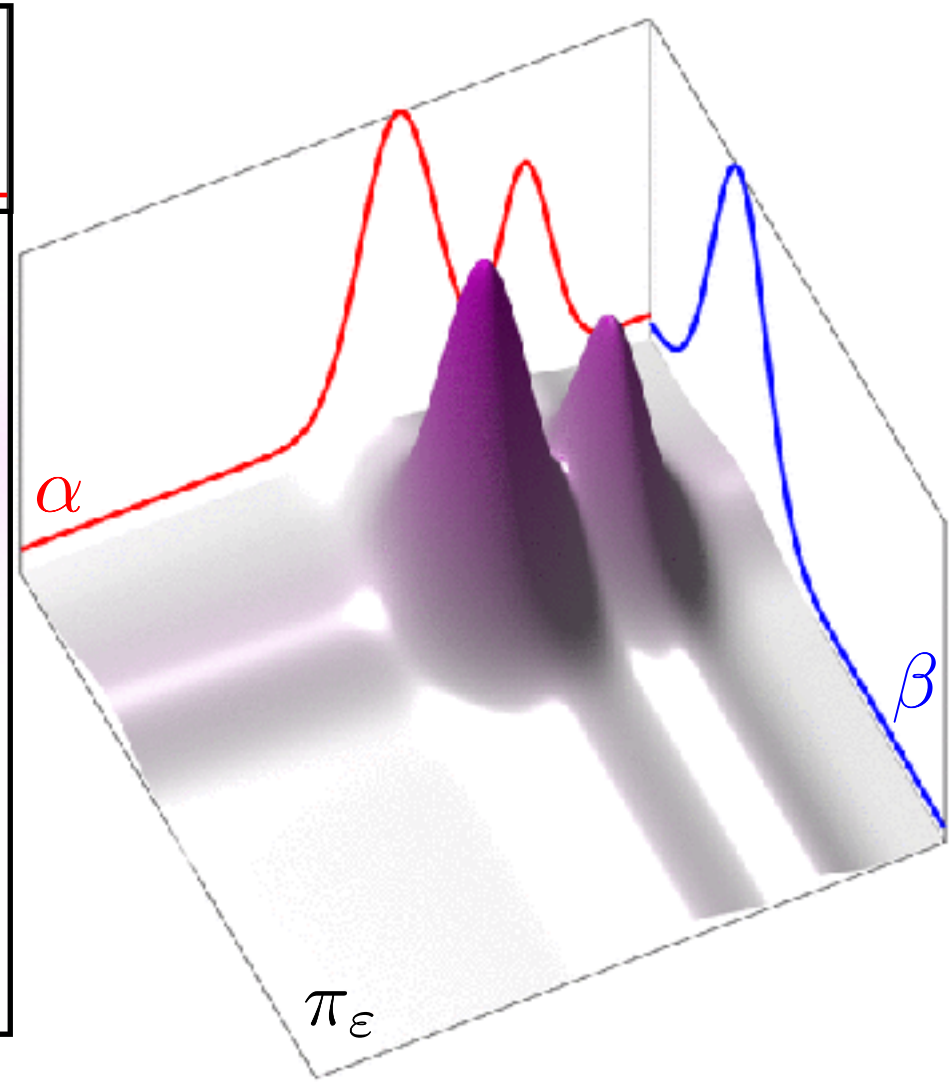


$$\mathcal{L}_c^\varepsilon(\alpha,\beta) \stackrel{\text{def.}}{=} \min_{\pi\in\mathcal{U}(\alpha,\beta)} \int_{X\times Y} c(x,y)\mathrm{d}\pi(x,y) + \varepsilon\,\mathrm{KL}(\pi|\alpha\otimes\beta)$$

$$\mathrm{KL}(\pi|\xi) \stackrel{\text{def.}}{=} \int_{\mathcal{X}\times\mathcal{Y}} \log\Big(\frac{\mathrm{d}\pi}{\mathrm{d}\xi}(x,y)\Big)\mathrm{d}\pi(x,y) + \int_{\mathcal{X}\times\mathcal{Y}} (\mathrm{d}\xi(x,y) - \mathrm{d}\pi(x,y))$$

Impact of Regularization





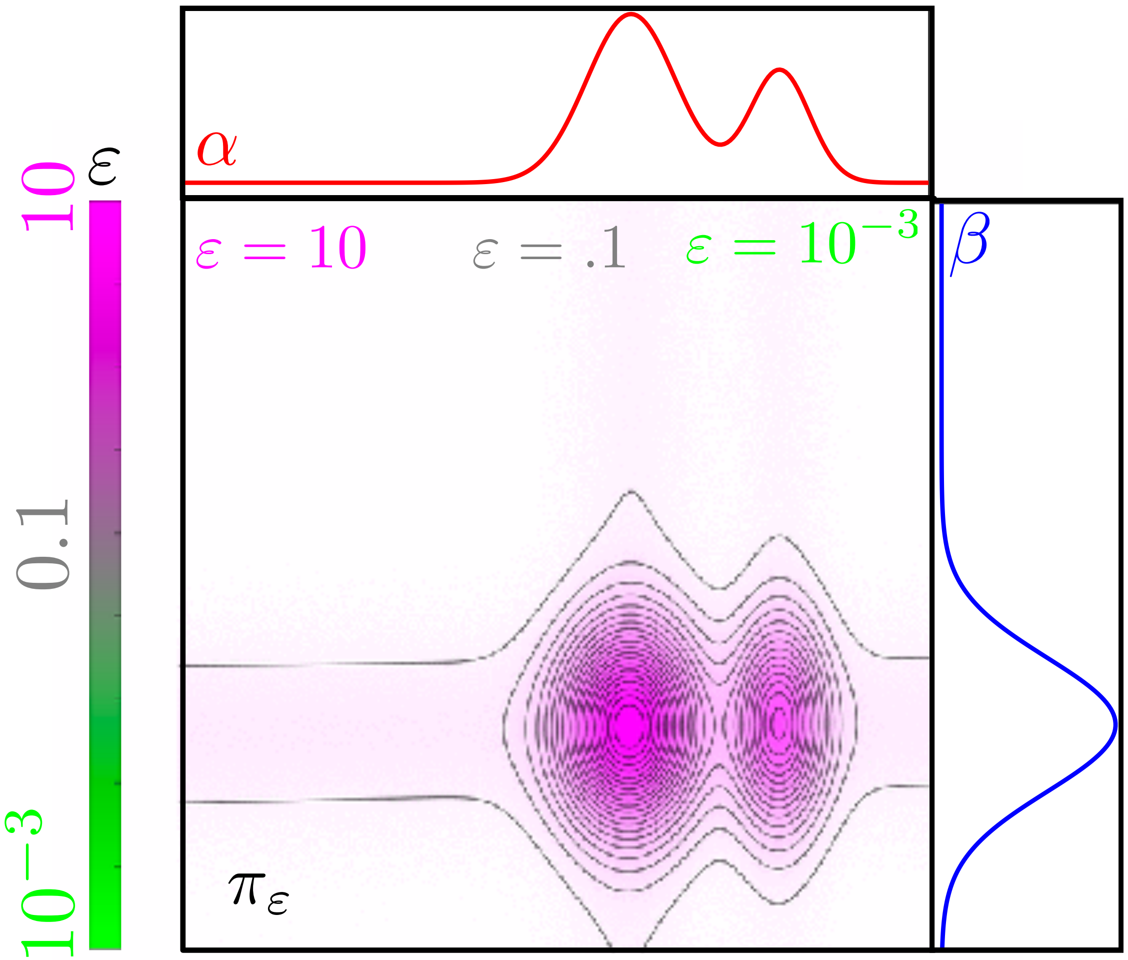
$$\pi_\varepsilon = \mathrm{argmin}_\pi \left\{ \int_{\mathbb{R}^2} \left(\|x-y\|^2 + \varepsilon \log\left(\frac{\mathrm{d}\pi}{\mathrm{d}\alpha \mathrm{d}\beta}(x,y) \right) \right) \mathrm{d}\pi(x,y) + \; ; \; \pi_1 = \alpha, \pi_2 = \beta \right\}$$

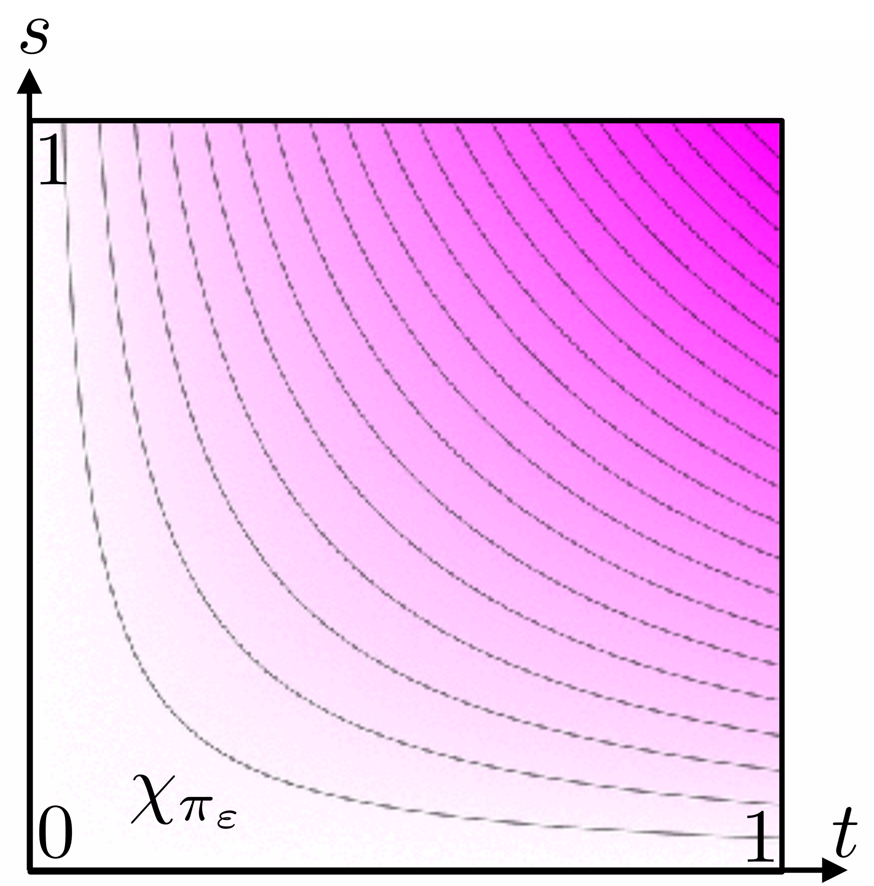
Theorem: $\pi_\varepsilon \xrightarrow{\varepsilon \to +\infty} \alpha \otimes \beta \qquad \pi_\varepsilon \xrightarrow{\varepsilon \to 0} \pi_{\mathrm{OT}}$

Impact of Regularization

Cumulative: $C_\pi(x,y) \overset{\text{def.}}{=} \int_{-\infty}^{x}\int_{-\infty}^{y} \mathrm{d}\pi(x,y)$

Copula: $\chi_\pi(s,t) \overset{\text{def.}}{=} C_\pi(C_\alpha^{-1}(s), C_\beta^{-1}(t))$





Theorem: $\chi_{\pi_\varepsilon}(s,t) \xrightarrow{\varepsilon \to 0} \min(s,t)$ (dependence)

$\chi_{\pi_\varepsilon}(s,t) \xrightarrow{\varepsilon \to +\infty} st$ (independence)

Sinkhorn's Algorithm

$$\mathrm{L}_{\mathbf{C}}^{\varepsilon}(\mathbf{a}, \mathbf{b}) \stackrel{\text{def.}}{=} \min_{\mathbf{P} \in \mathbf{U}(\mathbf{a},\mathbf{b})} \langle \mathbf{P},\, \mathbf{C} \rangle - \varepsilon \mathbf{H}(\mathbf{P})$$

$$\forall\,(i,j) \in [\![n]\!] \times [\![m]\!], \quad \mathbf{P}_{i,j} = \mathbf{u}_i \mathbf{K}_{i,j} \mathbf{v}_j$$

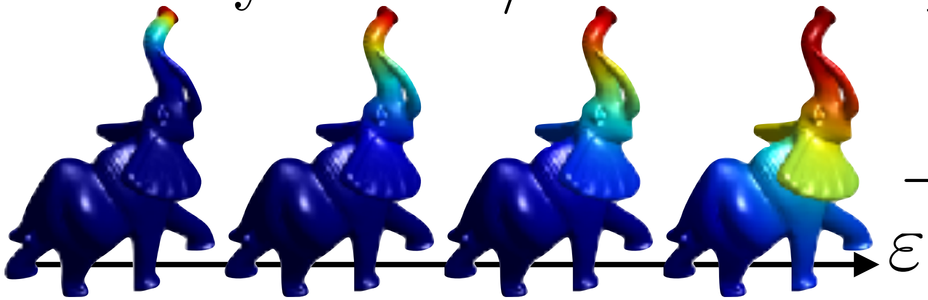
Row constraint: $\mathbf{u} \odot (\mathbf{K}\mathbf{v}) = \mathbf{a}$

Col. constraint: $\mathbf{v} \odot (\mathbf{K}^{\mathrm{T}}\mathbf{u}) = \mathbf{b}$

Sinkhorn iterations: $\mathbf{u}^{(\ell+1)} \stackrel{\text{def.}}{=} \dfrac{\mathbf{a}}{\mathbf{K}\mathbf{v}^{(\ell)}}$ and $\mathbf{v}^{(\ell+1)} \stackrel{\text{def.}}{=} \dfrac{\mathbf{b}}{\mathbf{K}^{\mathrm{T}}\mathbf{u}^{(\ell+1)}}$

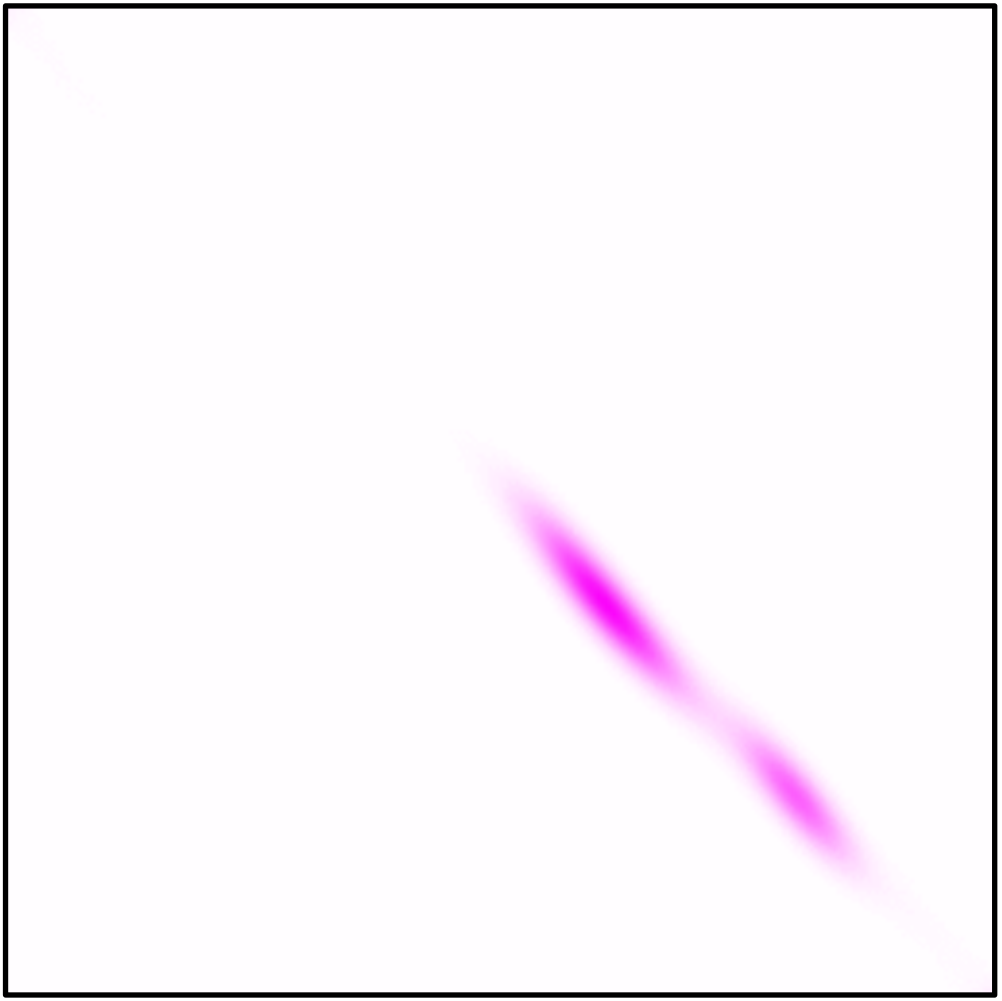
Only matrix/vector multiplications. $\rightarrow$ Parallelizable.

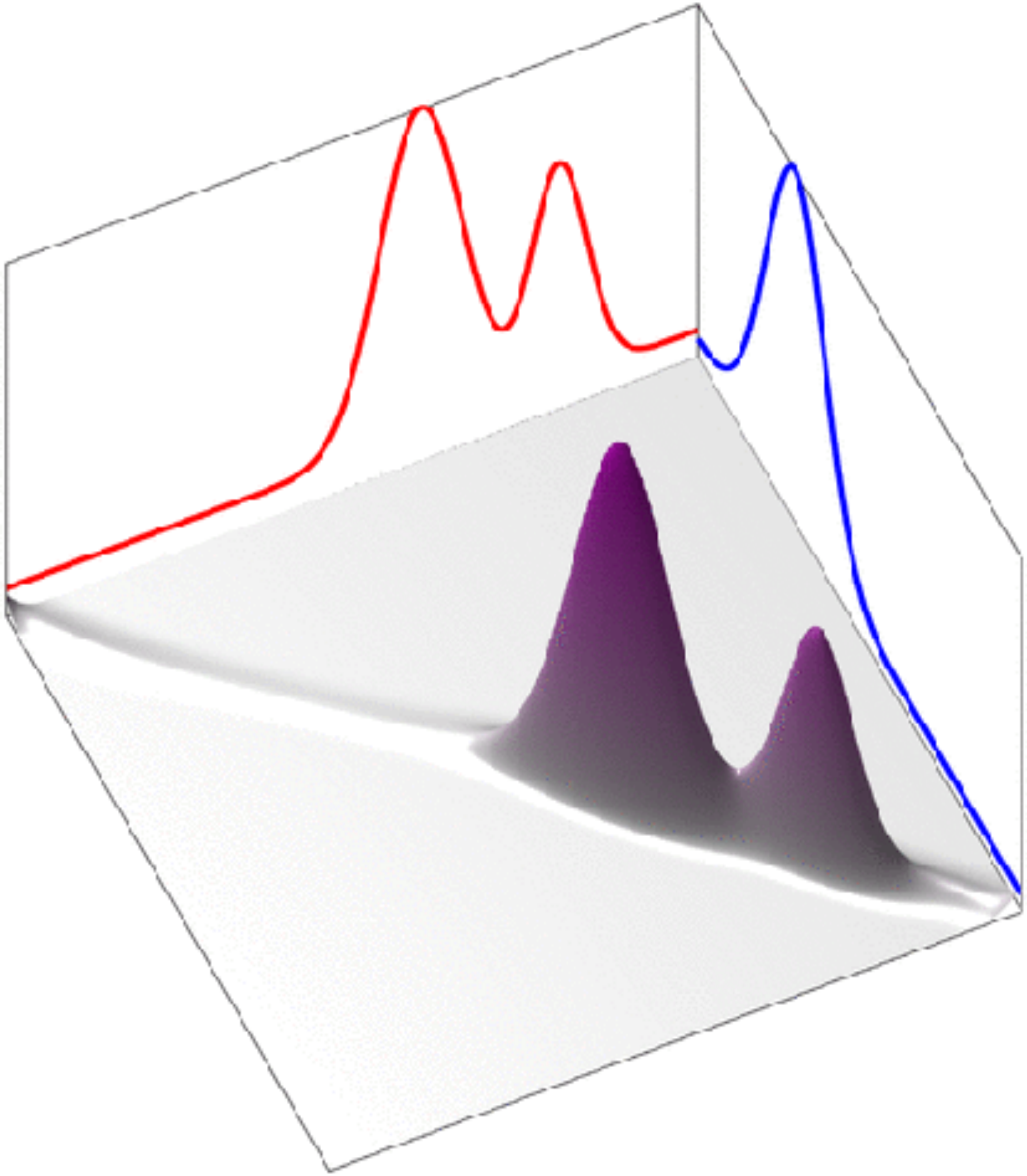
$\rightarrow$ Streams well on GPU.



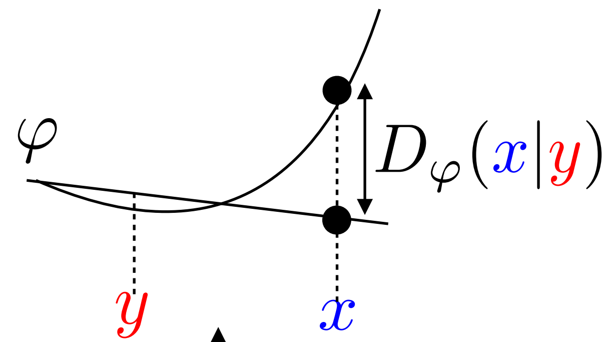
$\rightarrow$ convolutive/heat structure for K

Sinkhorn Convergence



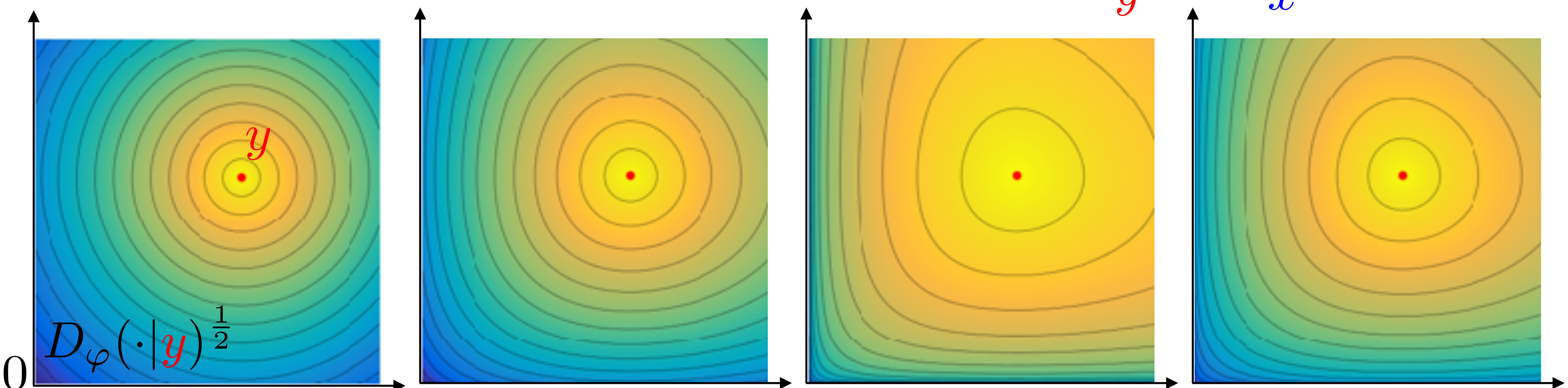


Bregman Divergences



Bregman divergence:

$$D_\varphi(x|y) \stackrel{\text{def.}}{=} \varphi(x) - \varphi(y) - \langle x - y,\, \nabla\varphi(y)\rangle$$



$\varphi(x) = \|x\|^2$ $\quad$ $\varphi(x) = \sum_i x_i \log(x_i)$ $\quad$ $\varphi(x) = \sum_i -\log(x_i)$ $\quad$ $\varphi(x) = \sum_i -\sqrt{x_i}$

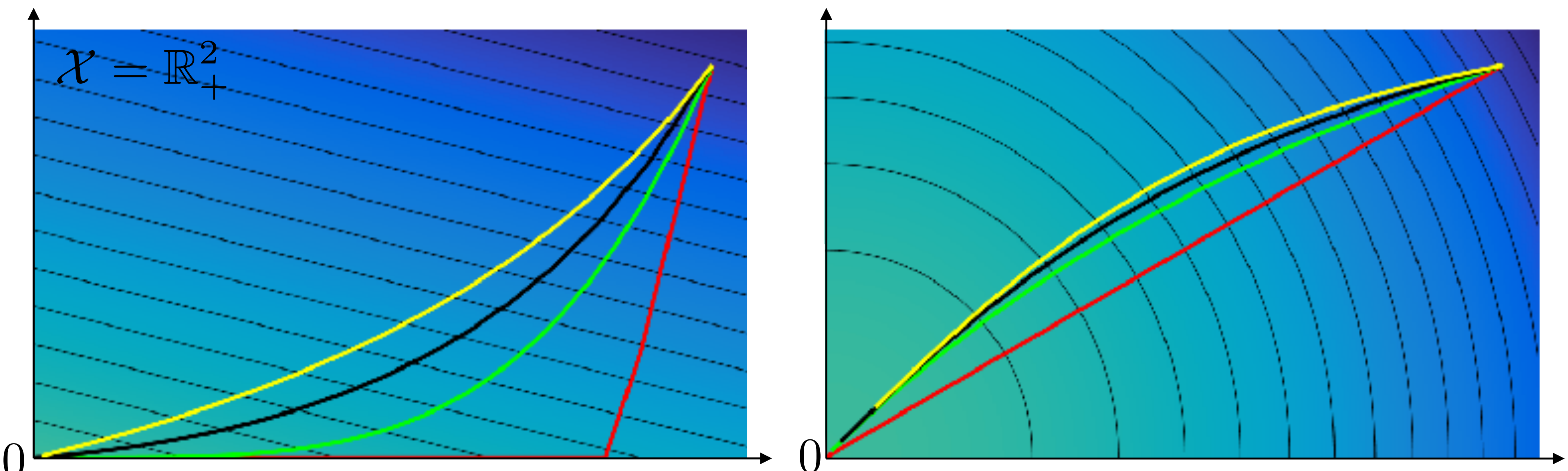
$$\left.\begin{array}{l} D_\varphi(x|x+\varepsilon) \\ D_\varphi(x+\varepsilon|x) \end{array}\right\} = \frac{1}{2}\langle \partial^2\varphi(x)\varepsilon,\, \varepsilon\rangle + o(\|\varepsilon\|^2)$$

Cannot be used for Radon Measure. Need to use f-divergences

Bregman Mirror Descent

Bregman divergence: $D_\varphi(x|y) \overset{\text{def.}}{=} \varphi(x) - \varphi(y) - \langle x - y,\, \nabla\varphi(y)\rangle$

Mirror descent:
$$x_{k+1} = \underset{x\in\mathcal{X}}{\operatorname{argmin}}\; D_\varphi(x|x_k) + \tau\langle \nabla f(x_k),\, x\rangle$$
$$= (\nabla\varphi)^{-1}\left(\nabla\varphi(x_k) - \tau\nabla f(x_k)\right)$$



$\varphi(x) = \|x\|^2$ $\varphi(x) = \sum_i -\log(x_i)$ $\varphi(x) = \sum_i x_i \log(x_i)$ $\varphi(x) = \sum_i -\sqrt{x_i}$

Bregman Iterative Projections

$$\mathcal{C}^1_{\mathbf{a}} \stackrel{\text{def.}}{=} \{\mathbf{P} \;:\; \mathbf{P}\mathbb{1}_m = \mathbf{a}\} \quad \text{and} \quad \mathcal{C}^2_{\mathbf{b}} \stackrel{\text{def.}}{=} \left\{\mathbf{P} \;:\; \mathbf{P}^{\mathrm{T}}\mathbb{1}_m = \mathbf{b}\right\}$$

$$\mathbf{KL}(\mathbf{P}|\mathbf{K}) \stackrel{\text{def.}}{=} \sum_{i,j} \mathbf{P}_{i,j} \log\left(\frac{\mathbf{P}_{i,j}}{\mathbf{K}_{i,j}}\right) - \mathbf{P}_{i,j} + \mathbf{K}_{i,j} \qquad \mathbf{K}_{i,j} \stackrel{\text{def.}}{=} e^{-\frac{\mathbf{C}_{i,j}}{\varepsilon}}$$

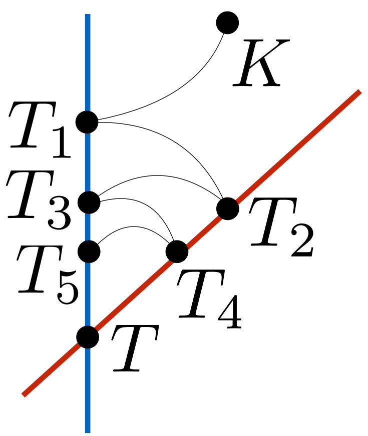
$$\mathbf{P}_\varepsilon = \mathrm{Proj}^{\mathbf{KL}}_{\mathbf{U}(\mathbf{a},\mathbf{b})}(\mathbf{K}) \stackrel{\text{def.}}{=} \underset{\mathbf{P}\in\mathbf{U}(\mathbf{a},\mathbf{b})}{\mathrm{argmin}}\ \mathbf{KL}(\mathbf{P}|\mathbf{K})$$

$$\mathbf{P}^{(\ell+1)} \stackrel{\text{def.}}{=} \mathrm{Proj}^{\mathbf{KL}}_{\mathcal{C}^1_{\mathbf{a}}}(\mathbf{P}^{(\ell)}) \quad \text{and} \quad \mathbf{P}^{(\ell+2)} \stackrel{\text{def.}}{=} \mathrm{Proj}^{\mathbf{KL}}_{\mathcal{C}^2_{\mathbf{b}}}(\mathbf{P}^{(\ell+1)})$$

$$\begin{aligned} \mathbf{P}^{(2\ell)} &\stackrel{\text{def.}}{=} \mathrm{diag}(\mathbf{u}^{(\ell)})\mathbf{K}\,\mathrm{diag}(\mathbf{v}^{(\ell)}), \\ \mathbf{P}^{(2\ell+1)} &\stackrel{\text{def.}}{=} \mathrm{diag}(\mathbf{u}^{(\ell+1)})\mathbf{K}\,\mathrm{diag}(\mathbf{v}^{(\ell)}) \\ \mathbf{P}^{(2\ell+2)} &\stackrel{\text{def.}}{=} \mathrm{diag}(\mathbf{u}^{(\ell+1)})\mathbf{K}\,\mathrm{diag}(\mathbf{v}^{(\ell+1)}) \end{aligned}$$

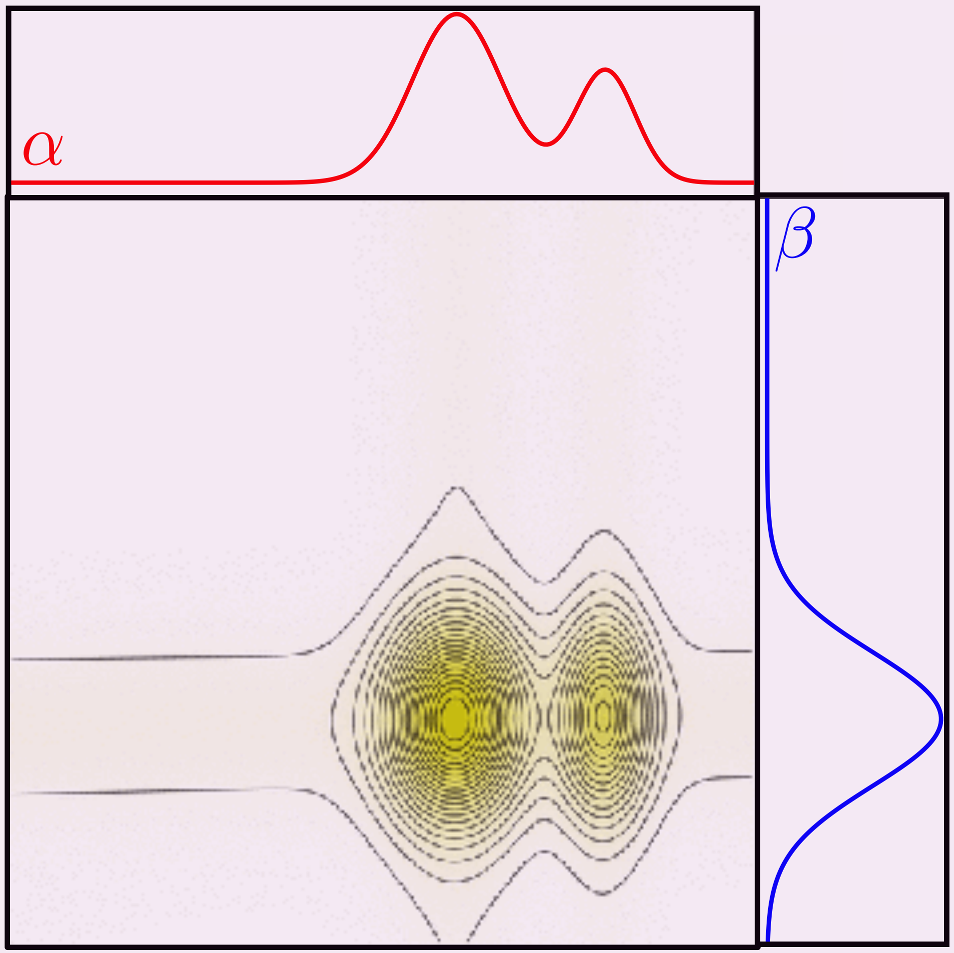
Prop. $(\star) \iff \min_T \{\mathrm{KL}(T|K)\ ;\ T \in \mathcal{C}_{\mu,\nu}\}$

Sinkhorn $\iff$ iterative projections.



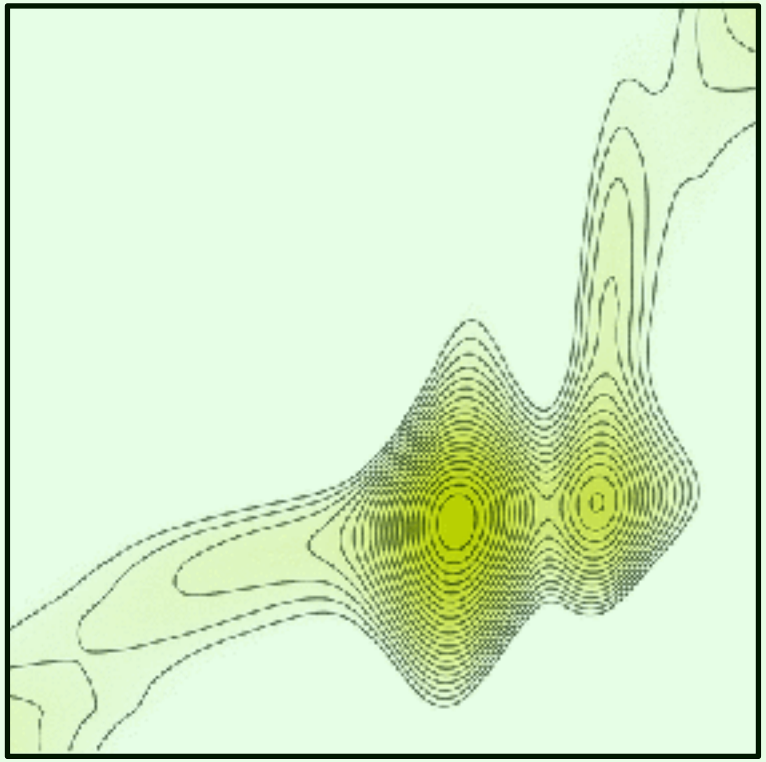
Other Regularizations

$$\min_{\pi} \left\{ \int_{\mathbb{R}^2} \|x-y\|^2 \mathrm{d}\pi(x,y) + \varepsilon R(\pi) \; ; \; \pi_1 = \alpha, \pi_2 = \beta \right\}$$



$$R(\pi) = \int \log\left(\frac{\mathrm{d}\pi}{\mathrm{d}x\mathrm{d}y}\right) \mathrm{d}\pi(x,y)$$

Dykstra's algorithm



$$R(\pi) = \int \left(\frac{\mathrm{d}\pi}{\mathrm{d}x\mathrm{d}y}\right)^2 \mathrm{d}x\mathrm{d}y$$

Overview

- Entropic Regularization and Sinkhorn
- Convergence Analysis
- Barycenters
- Generalized Sinkhorn and Applications (dedicated section)

Convergence of Sinkhorn

Sinkhorn/IPFP algorithm: [Sinkhorn 1967][Deming,Stephan 1940]

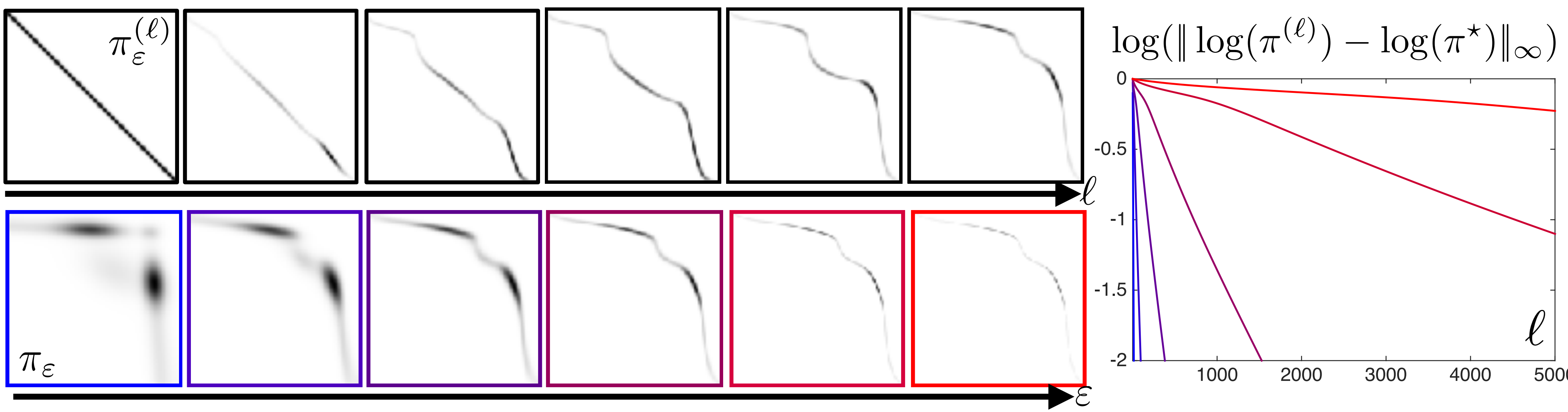
$$a^{(\ell+1)} \overset{\text{def.}}{=} \frac{\mu}{K b^{(\ell)}} \quad \text{and} \quad b^{(\ell+1)} \overset{\text{def.}}{=} \frac{\nu}{K^* a^{(\ell+1)}}$$

Proposition: $\| \log(\pi^{(\ell)}) - \log(\pi^\star) \|_\infty = O(1-\delta)^\ell, \quad \delta \sim \kappa_c^{-1/\varepsilon}$

$$\pi^{(\ell)} \overset{\text{def.}}{=} \operatorname{diag}(a^{(\ell)}) K \operatorname{diag}(b^{(\ell)})$$

[Franklin,Lorenz 1989]

Local rate: [Knight 2008]

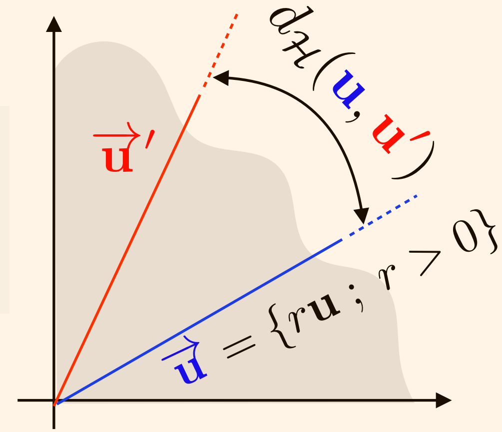


Hilbert Projective Metric

Hilbert's projective metric:

$$\forall\,(\mathbf{u},\mathbf{u}') \in (\mathbb{R}^n_{+,*})^2, \quad d_{\mathcal{H}}(\mathbf{u},\mathbf{u}') \overset{\text{def.}}{=} \log \max_{i,i'} \frac{\mathbf{u}_i \mathbf{u}'_{i'}}{\mathbf{u}_{i'} \mathbf{u}'_i}.$$

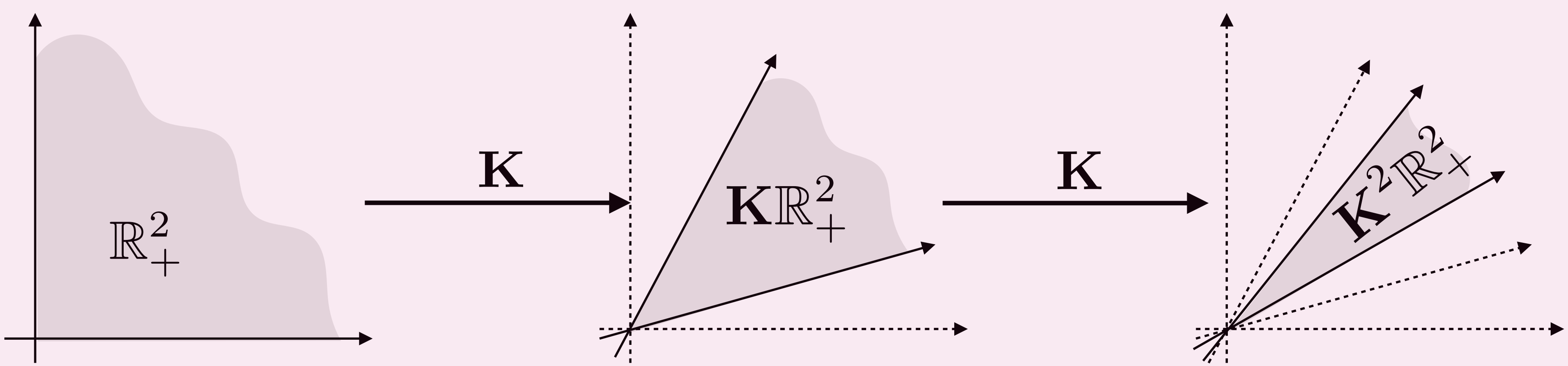
$d_{\mathcal{H}}$ is a distance on the set of rays $\overrightarrow{\mathbf{u}}$.



Birkhoff's contraction theorem:

Theorem 1.1. Let $\mathbf{K} \in \mathbb{R}^{n\times m}_{+,*}$, then for $(\mathbf{v},\mathbf{v}') \in (\mathbb{R}^m_{+,*})^2$

$$d_{\mathcal{H}}(\mathbf{K}\mathbf{v},\mathbf{K}\mathbf{v}') \leq \lambda(\mathbf{K}) d_{\mathcal{H}}(\mathbf{v},\mathbf{v}') \text{ where } \begin{cases} \lambda(\mathbf{K}) \overset{\text{def.}}{=} \frac{\sqrt{\eta(\mathbf{K})}-1}{\sqrt{\eta(\mathbf{K})}+1} < 1 \\ \eta(\mathbf{K}) \overset{\text{def.}}{=} \max_{i,j,k,\ell} \frac{\mathbf{K}_{i,k}\mathbf{K}_{j,\ell}}{\mathbf{K}_{j,k}\mathbf{K}_{i,\ell}}. \end{cases}$$



Perron Frobenius

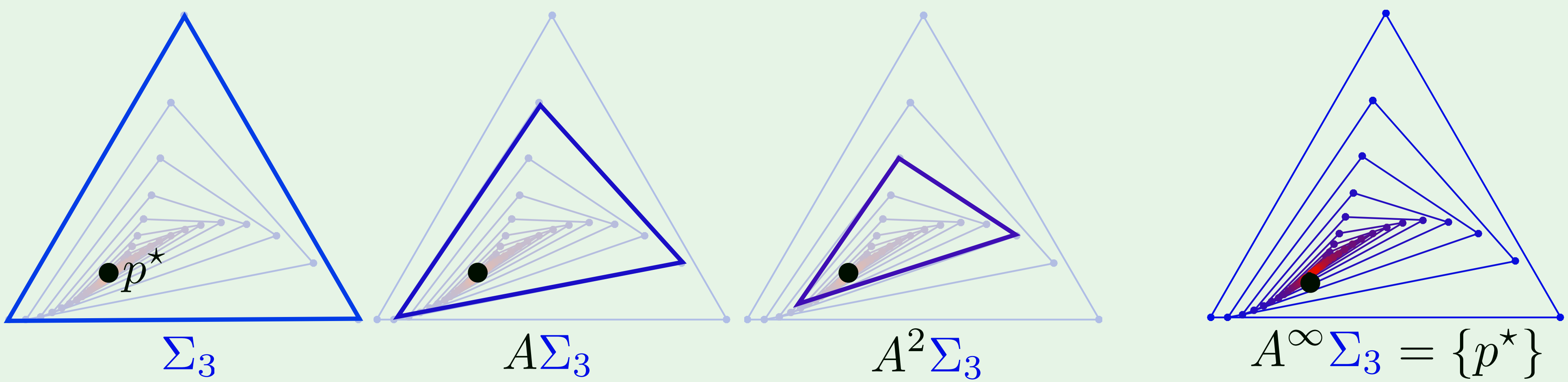
Simplex: $\Sigma_k = \{p \in \mathbb{R}^k_+ \; ; \; \sum_i p_i = 1\}$

Stochastic matrix: $A \in \mathbb{R}^n_+$, $A^\top \mathbb{1}_k = \mathbb{1}_k$

$A : \Sigma_k \to \Sigma_k$

Theorem: [Perron-Frobenius] If $A > 0$, $\exists! p^\star$, $Ap^\star = p^\star$.

$\exists \rho \in [0, 1[$, $\|A^k p - p^\star\| \leqslant \rho^k$



Sinkhorn under Hilbert's Metric

One has $(\mathbf{u}^{(\ell)}, \mathbf{v}^{(\ell)}) \rightarrow (\mathbf{u}^\star, \mathbf{v}^\star)$

$$d_{\mathcal{H}}(\mathbf{u}^{(\ell)}, \mathbf{u}^\star) = O(\lambda(\mathbf{K})^{2\ell}), \quad d_{\mathcal{H}}(\mathbf{v}^{(\ell)}, \mathbf{v}^\star) = O(\lambda(\mathbf{K})^{2\ell}).$$

$$d_{\mathcal{H}}(\mathbf{u}^{(\ell)}, \mathbf{u}^\star) \leq \frac{d_{\mathcal{H}}(\mathbf{P}^{(\ell)}\mathbb{1}_m, \mathbf{a})}{1 - \lambda(\mathbf{K})^2}$$

$$d_{\mathcal{H}}(\mathbf{v}^{(\ell)}, \mathbf{v}^\star) \leq \frac{d_{\mathcal{H}}(\mathbf{P}^{(\ell),\top}\mathbb{1}_n, \mathbf{b})}{1 - \lambda(\mathbf{K})^2}$$

$$\|\log(\mathbf{P}^{(\ell)}) - \log(\mathbf{P}^\star)\|_\infty \leq d_{\mathcal{H}}(\mathbf{u}^{(\ell)}, \mathbf{u}^\star) + d_{\mathcal{H}}(\mathbf{v}^{(\ell)}, \mathbf{v}^\star)$$

Local Analysis of Sinkhorn

$$\mathbf{f}^{(\ell+1)} = \Phi(\mathbf{f}^{(\ell)})$$

$$\Phi \overset{\text{def.}}{=} \Phi_2 \odot \Phi_1 \quad \text{where} \quad \begin{cases} \Phi_1(\mathbf{f}) = \varepsilon \log \mathbf{K}^{\mathrm{T}}(e^{\mathbf{f}/\varepsilon}) - \log(\mathbf{b}), \\ \Phi_2(\mathbf{g}) = \varepsilon \log \mathbf{K}(e^{\mathbf{g}/\varepsilon}) - \log(\mathbf{a}). \end{cases}$$

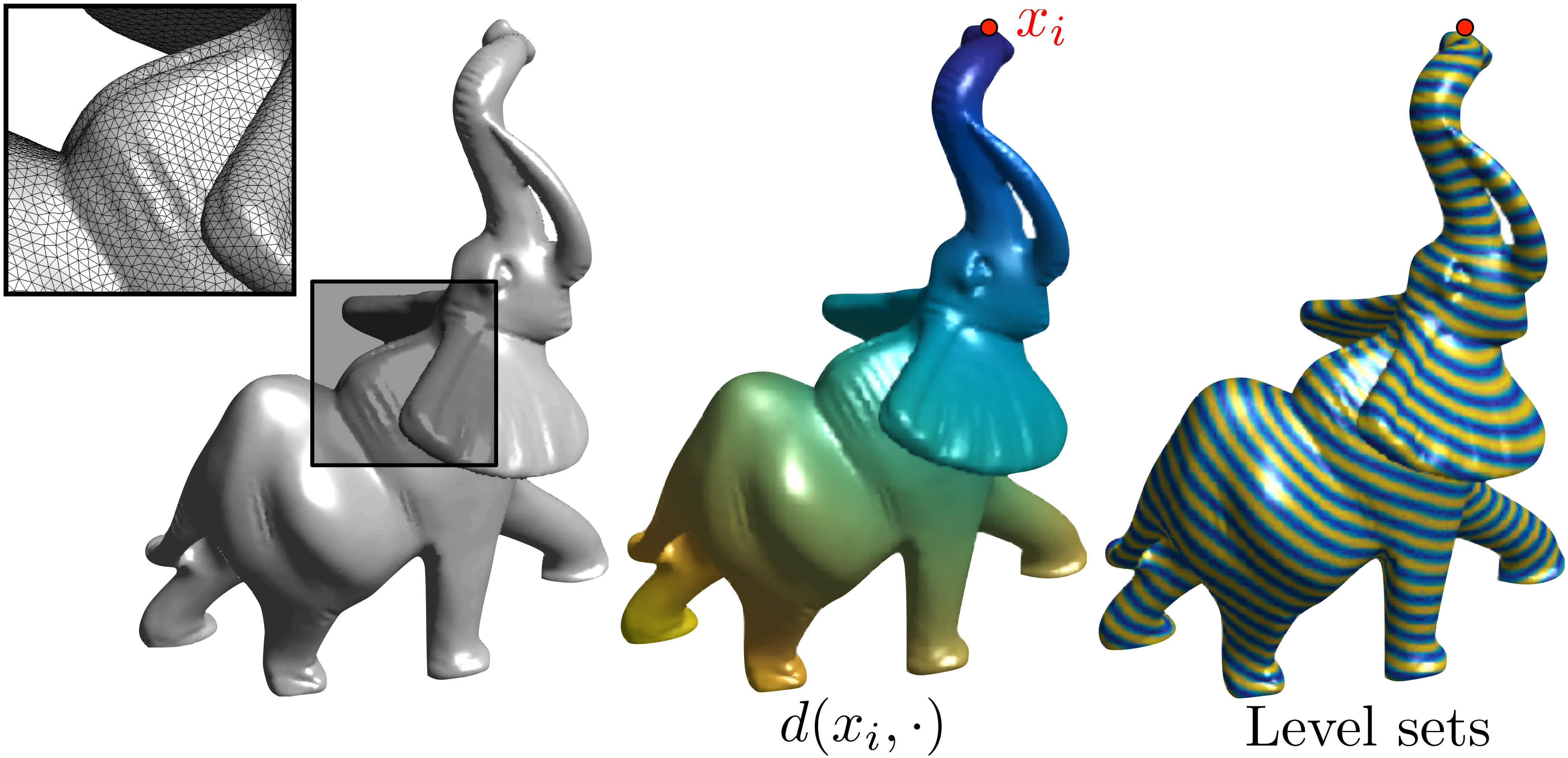
$$\partial\Phi(\mathbf{f}) = \mathrm{diag}(\mathbf{a})^{-1} \odot \mathbf{P} \odot \mathrm{diag}(\mathbf{b})^{-1} \odot \mathbf{P}^{\mathrm{T}}.$$

$$\|\mathbf{f}^{(\ell)} - \mathbf{f}\| = O((1-\kappa)^{\ell})$$

Optimal Transport on Surfaces

Triangulated mesh M. Geodesic distance d_M.

Ground cost: $c(x,y) = d_M(x,y)^\alpha$.



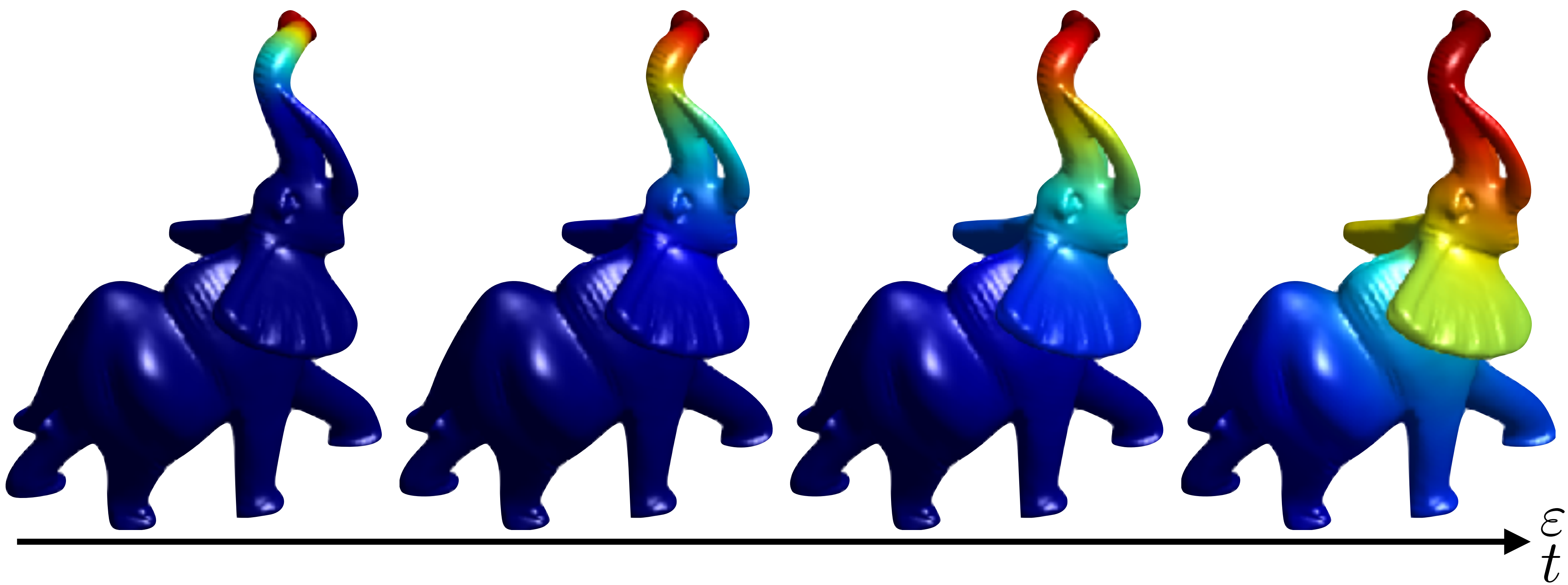
Computing c (Fast-Marching): $N^2 \log(N) \rightarrow$ too costly.

Entropic Transport on Surfaces

Heat equation on M: $\partial_t u_t(x,\cdot) = \Delta_M u_t(x,\cdot),\ u_{t=0}(x,\cdot) = \delta_x$

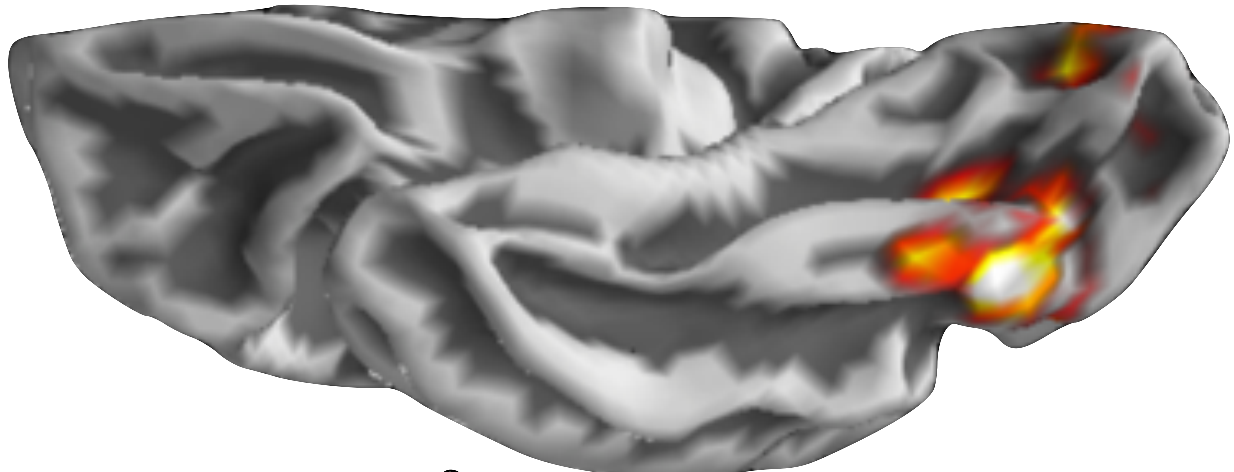
$$\textit{Theorem:}\ \text{[Varadhan]} \quad -\varepsilon \log(u_\varepsilon) \xrightarrow{\varepsilon \to 0} d_M^2$$

Sinkhorn kernel: $K \overset{\text{def.}}{=} e^{-\frac{d_M^2}{\varepsilon}} \approx u_\varepsilon \approx \left(\mathrm{Id} - \frac{\varepsilon}{\ell}\Delta_M\right)^{-\ell}$

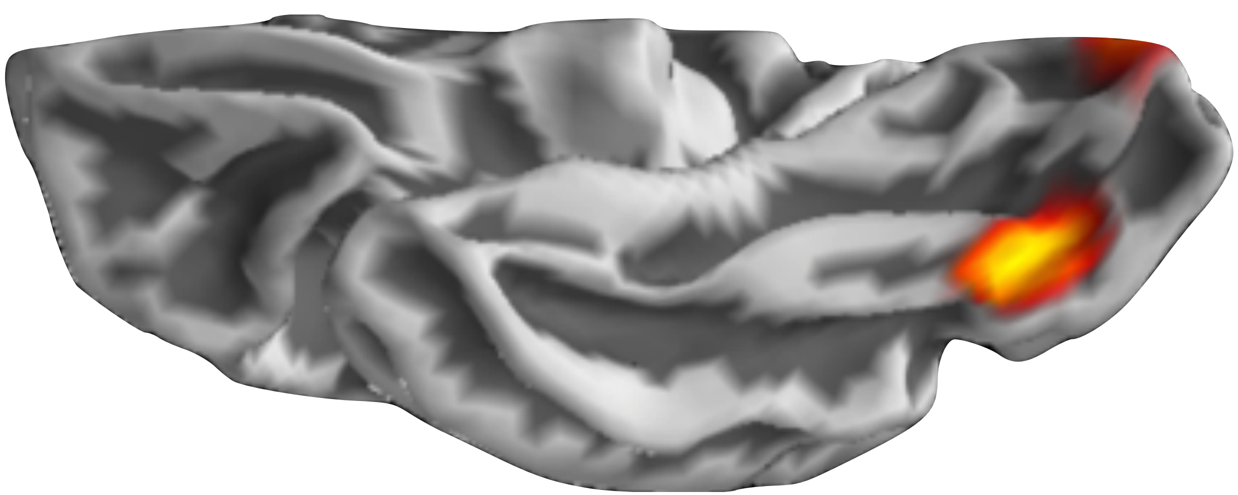


MRI Data Procesing [with A. Gramfort]

Ground cost $c = d_M$: geodesic on cortical surface M.



L^2 barycenter



W_2^2 barycenter

Regularization for General Measures

$$\pi_\varepsilon \stackrel{\text{def.}}{=} \underset{\pi}{\text{argmin}} \; \{\langle d^p, \pi\rangle + \varepsilon \text{KL}(\pi|\pi_0) \; ; \; \pi \in \Pi(\mu, \nu)\}$$

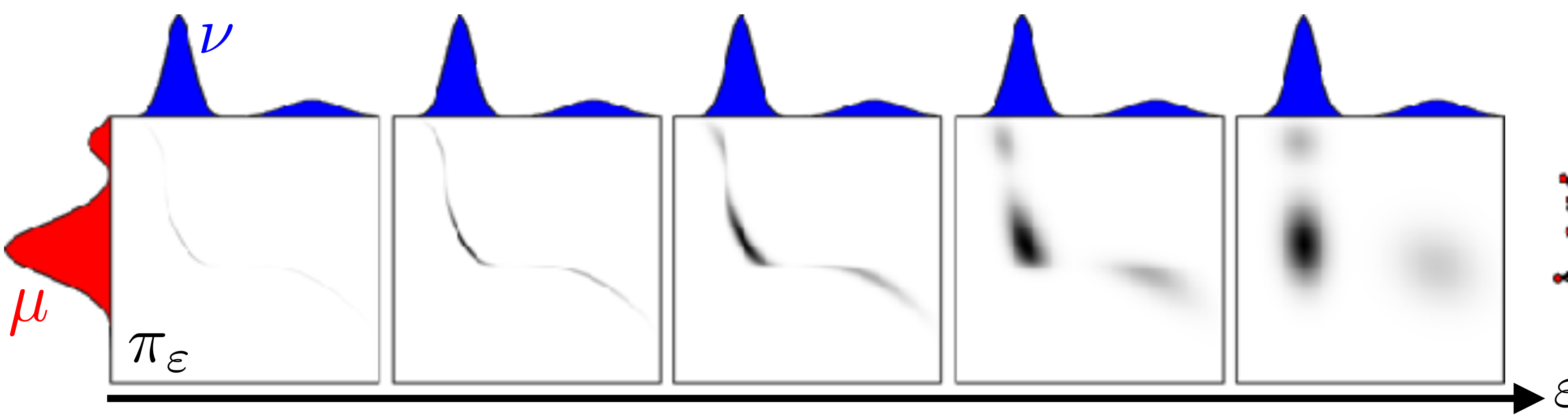
Schrödinger's problem: $\pi_\varepsilon = \underset{\pi\in\Pi(\mu,\nu)}{\text{argmin}} \; \text{KL}(\pi|K)$

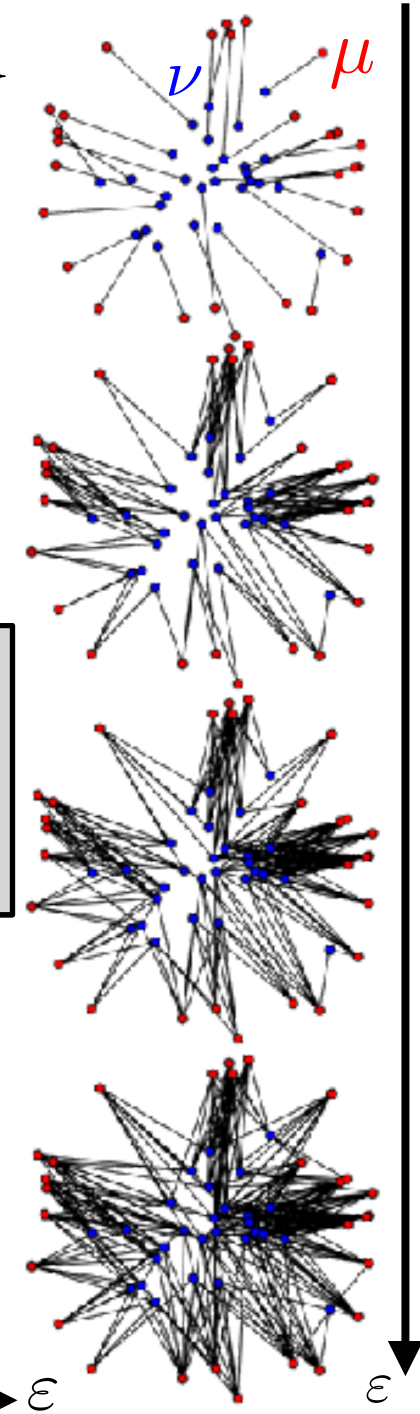
$$K(x,y) \stackrel{\text{def.}}{=} e^{-\frac{d^p(x,y)}{\varepsilon}} \pi_0(x,y)$$

Landmark computational paper: [Cuturi 2013].

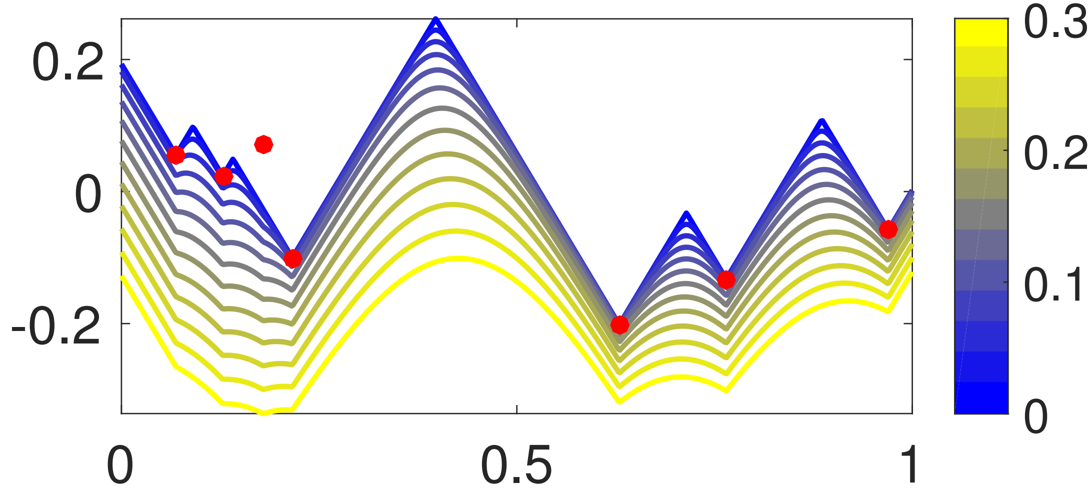
Proposition: [Carlier, Duval, Peyré, Schmitzer 2015]

$$\pi_\varepsilon \xrightarrow{\varepsilon \to 0} \underset{\pi\in\Pi(\mu,\nu)}{\text{argmin}} \; \langle d^p, \pi\rangle \qquad \pi_\varepsilon \xrightarrow{\varepsilon\to+\infty} \mu(x)\nu(y)$$

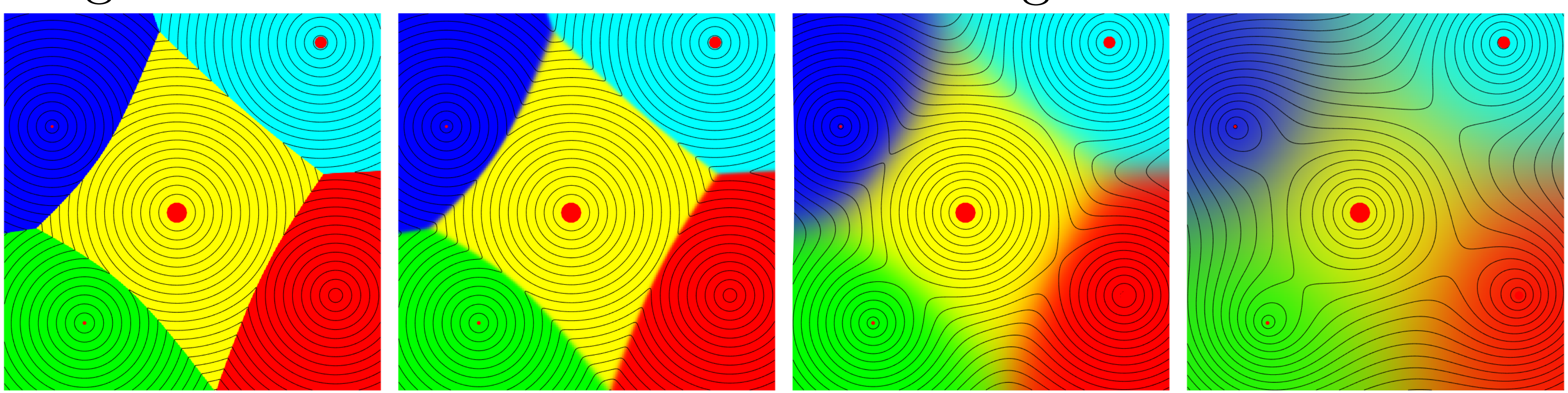




Semi-discrete OT and Entropy



Laguerre cells "Sinkhorn" Laguerre cells



$\varepsilon = 0$ $\varepsilon = 0.01$ $\varepsilon = 0.1$ $\varepsilon = 0.3$

Semi-discrete Optimal Transport is a multiclass SVM. Sinkhorn version is logistic regression. Decomposes space in entropic Laguerre diagrams

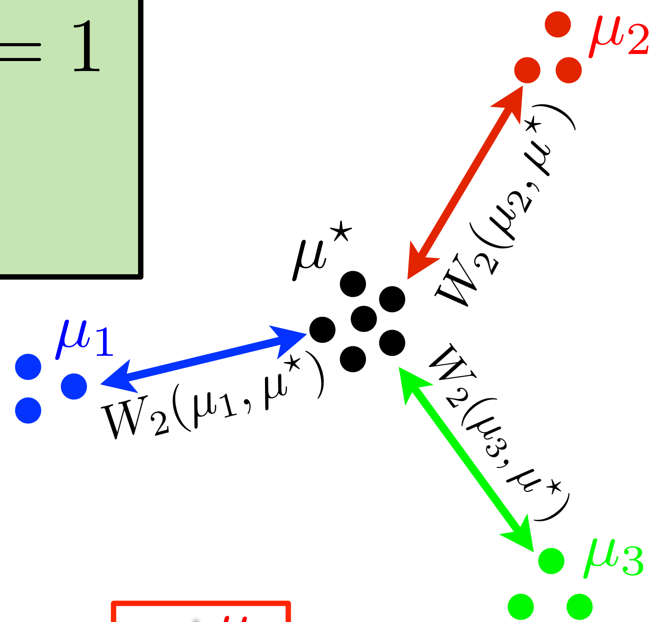
Overview

- Entropic Regularization and Sinkhorn
- Convergence Analysis
- **Barycenters**
- Generalized Sinkhorn and Applications (dedicated section)

Wasserstein Barycenters

Barycenters of measures $(\mu_k)_k$: $\quad \sum_k \lambda_k = 1$

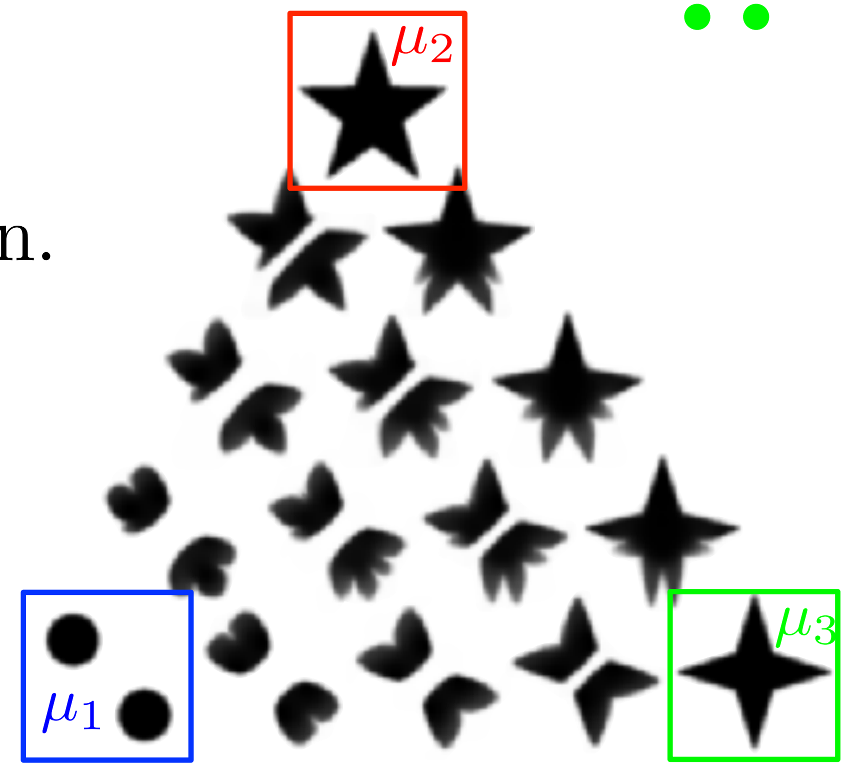
$$\mu^\star \in \underset{\mu}{\operatorname{argmin}} \ \sum_k \lambda_k W_2^2(\mu_k, \mu)$$



Generalizes Euclidean barycenter:

If $\mu_k = \delta_{x_k}$ then $\mu^\star = \delta_{\sum_k \lambda_k x_k}$

Mc Cann's displacement interpolation.



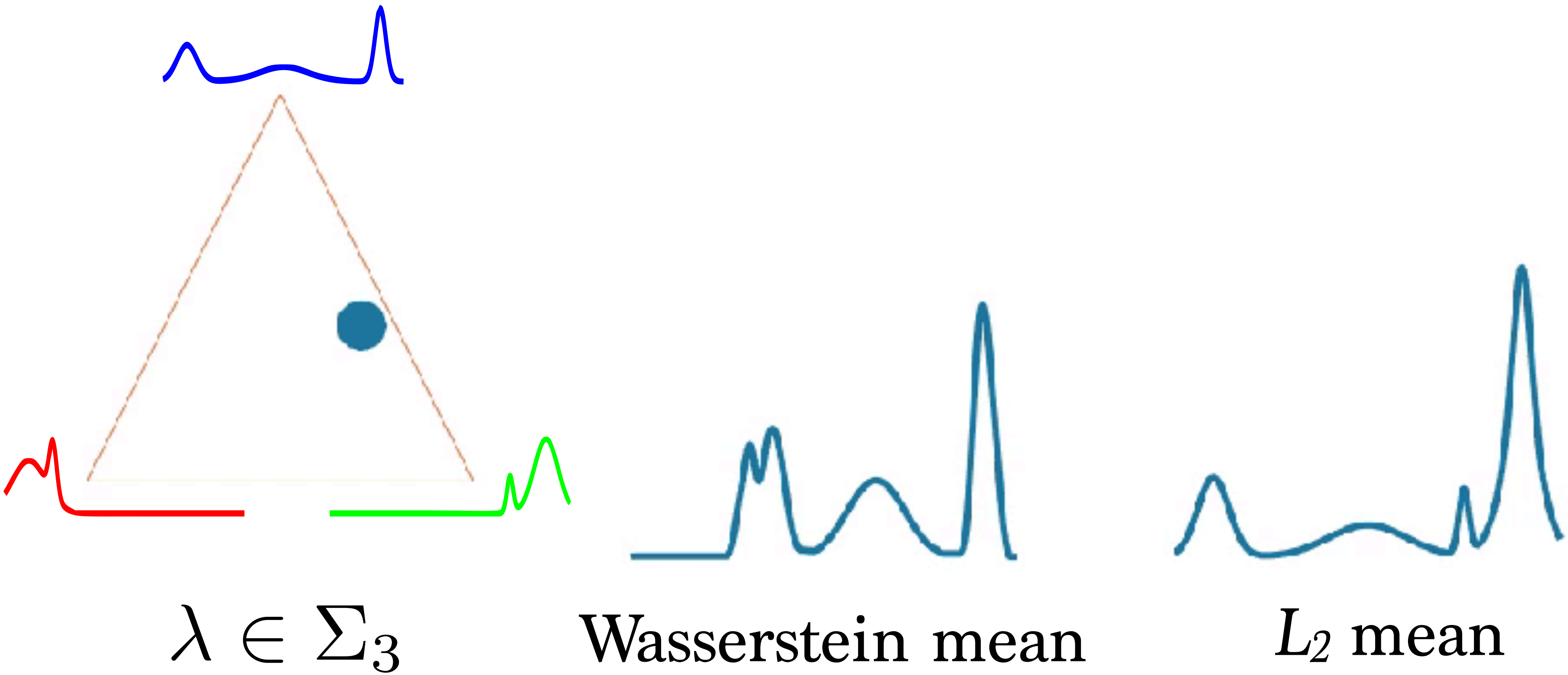
Theorem: [Agueh, Carlier, 2010]

(for $c(x,y) = \|x - y\|^2$)

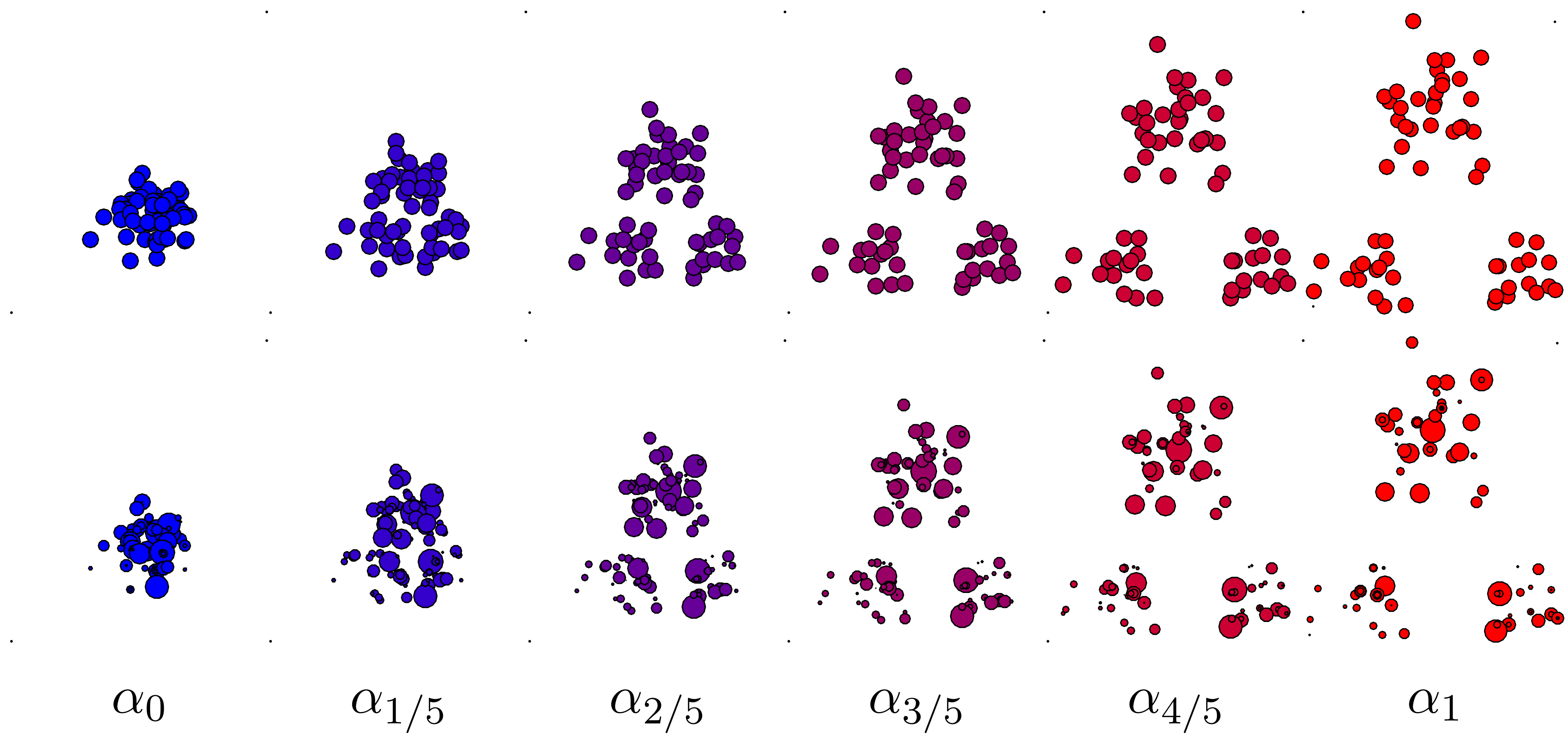
if μ_1 does not vanish on small sets,

$\mu^\star$ exists and is unique.

Wasserstein Barycenters



Displacement Interpolation



Displacement Interpolation

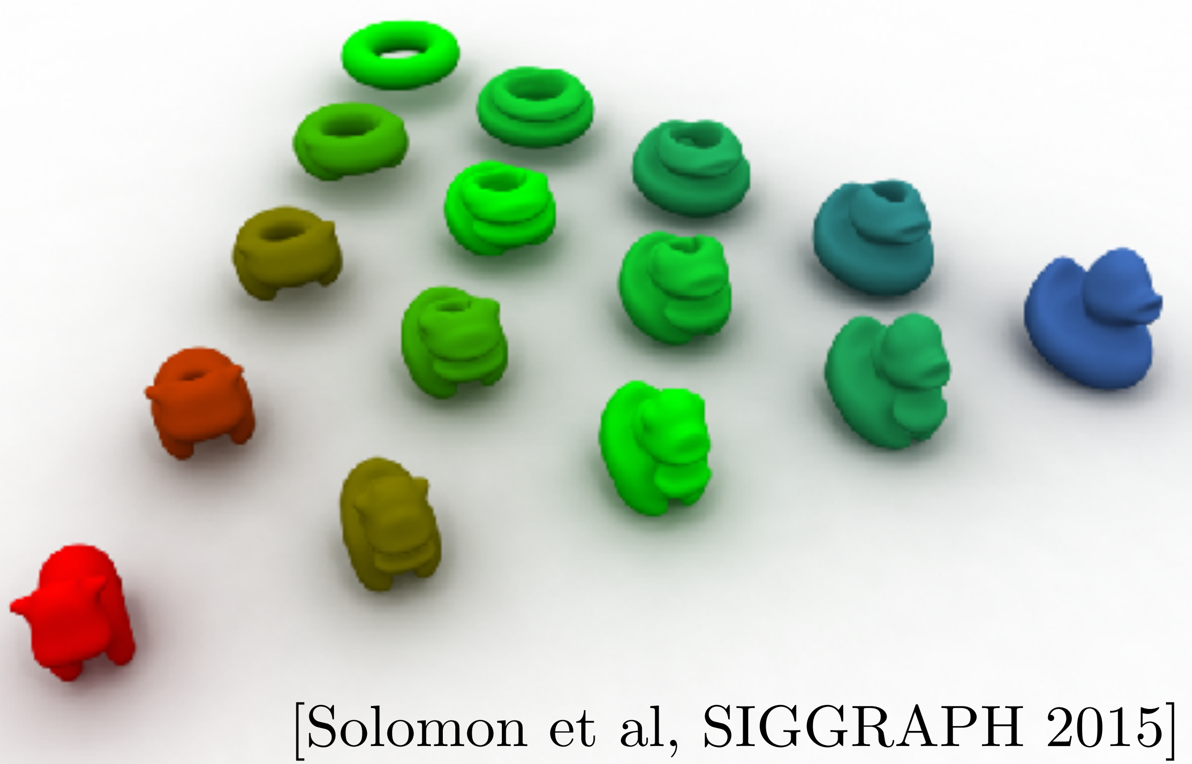
 

Regularized Barycenters

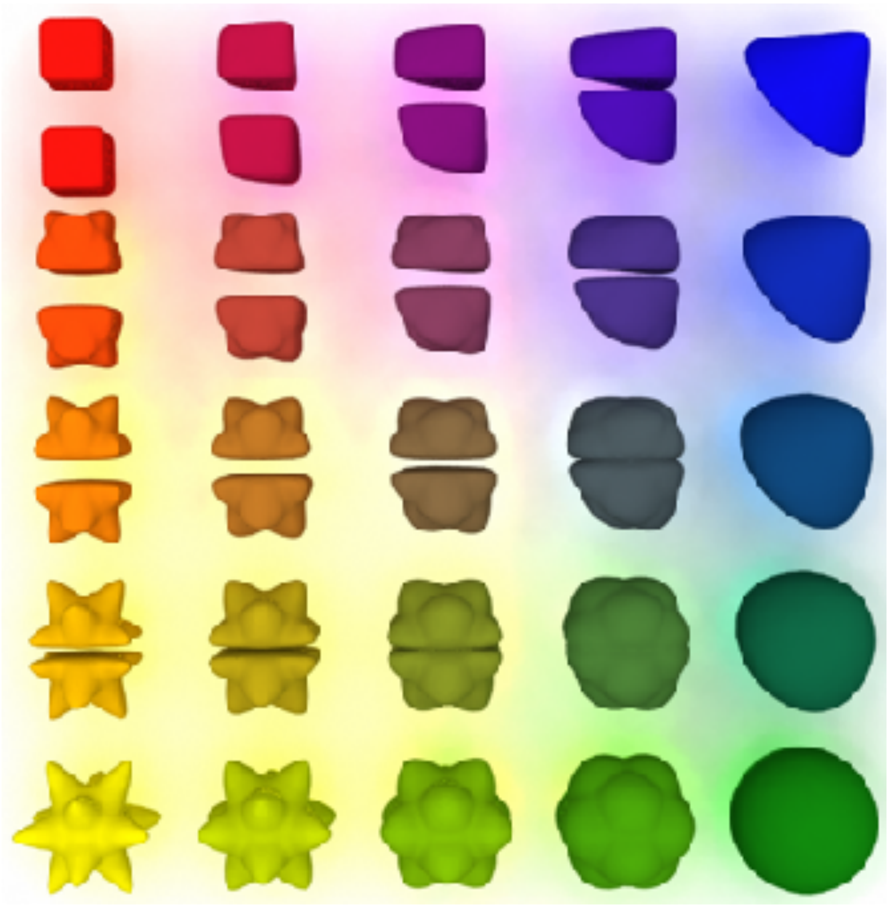
$$\min_{(\pi_k)_k,\mu} \left\{ \sum_k \lambda_k \left(\langle c,\, \pi_k \rangle + \varepsilon \mathrm{KL}(\pi_k | \pi_{0,k}) \right) \;\; ; \;\; \forall k, \pi_k \in \Pi(\mu_k, \mu) \right\}$$

$\longrightarrow$ Need to fix a discretization grid for μ, i.e. choose $(\pi_{0,k})_k$

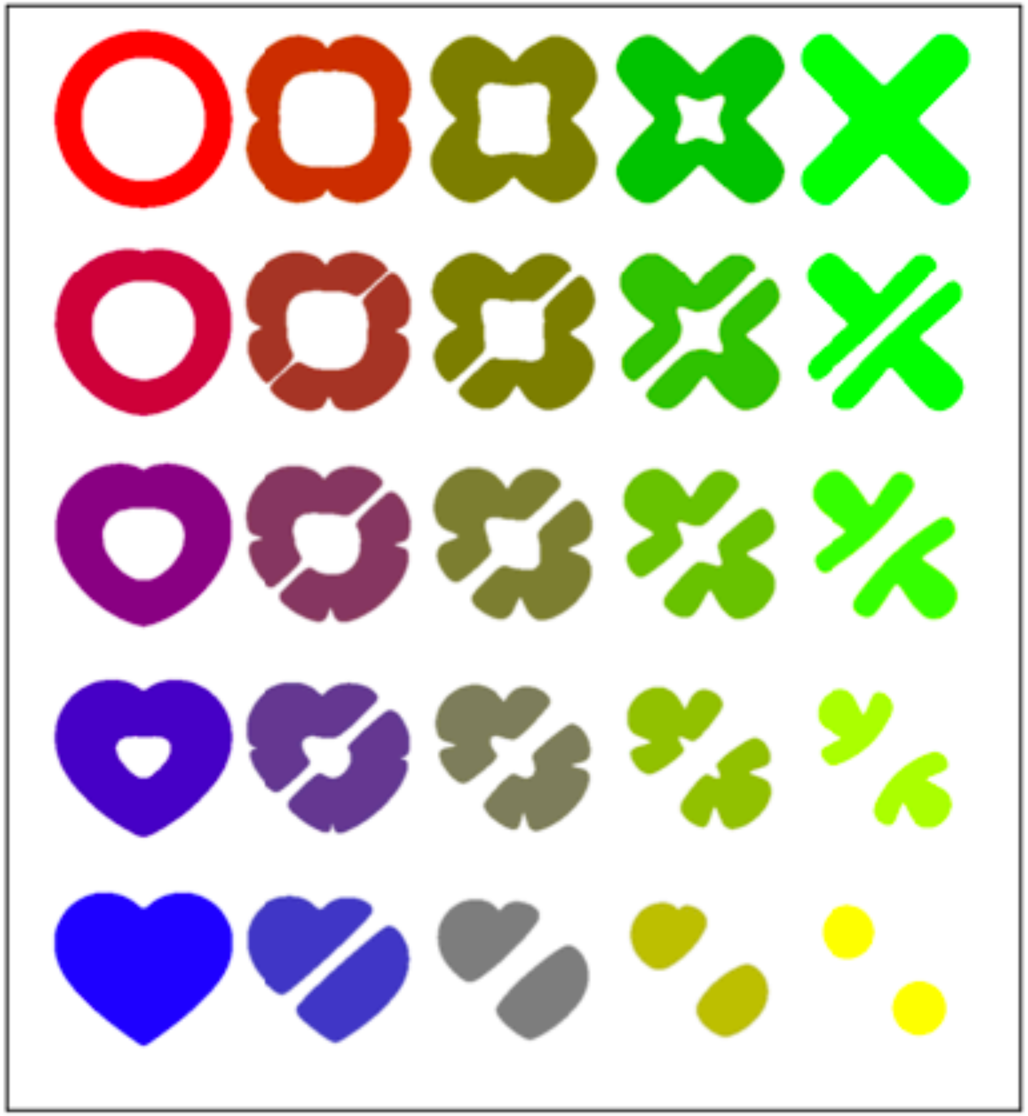
$\longrightarrow$ Sinkhorn-like algorithm [Benamou, Carlier, Cuturi, Nenna, Peyré, 2015].

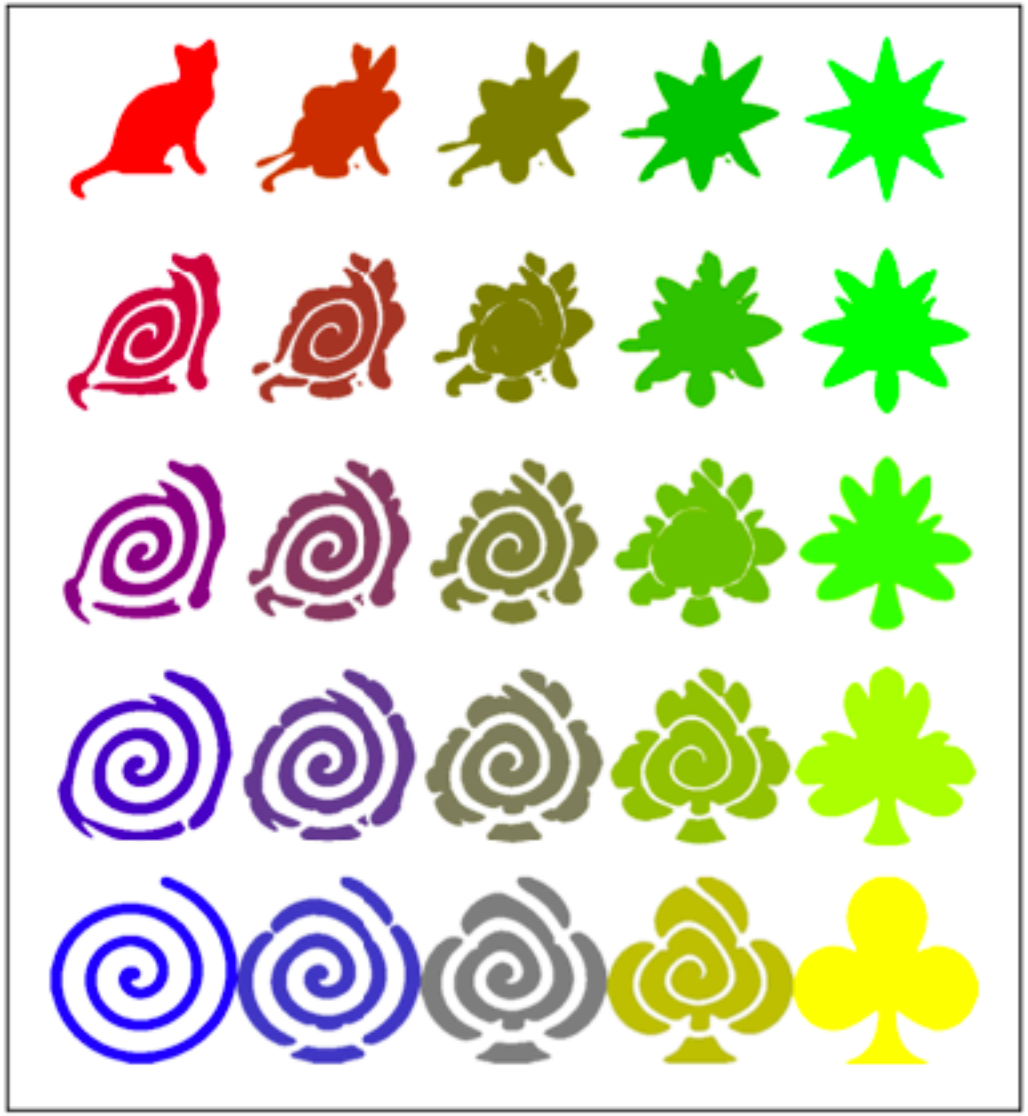


[Solomon et al, SIGGRAPH 2015]

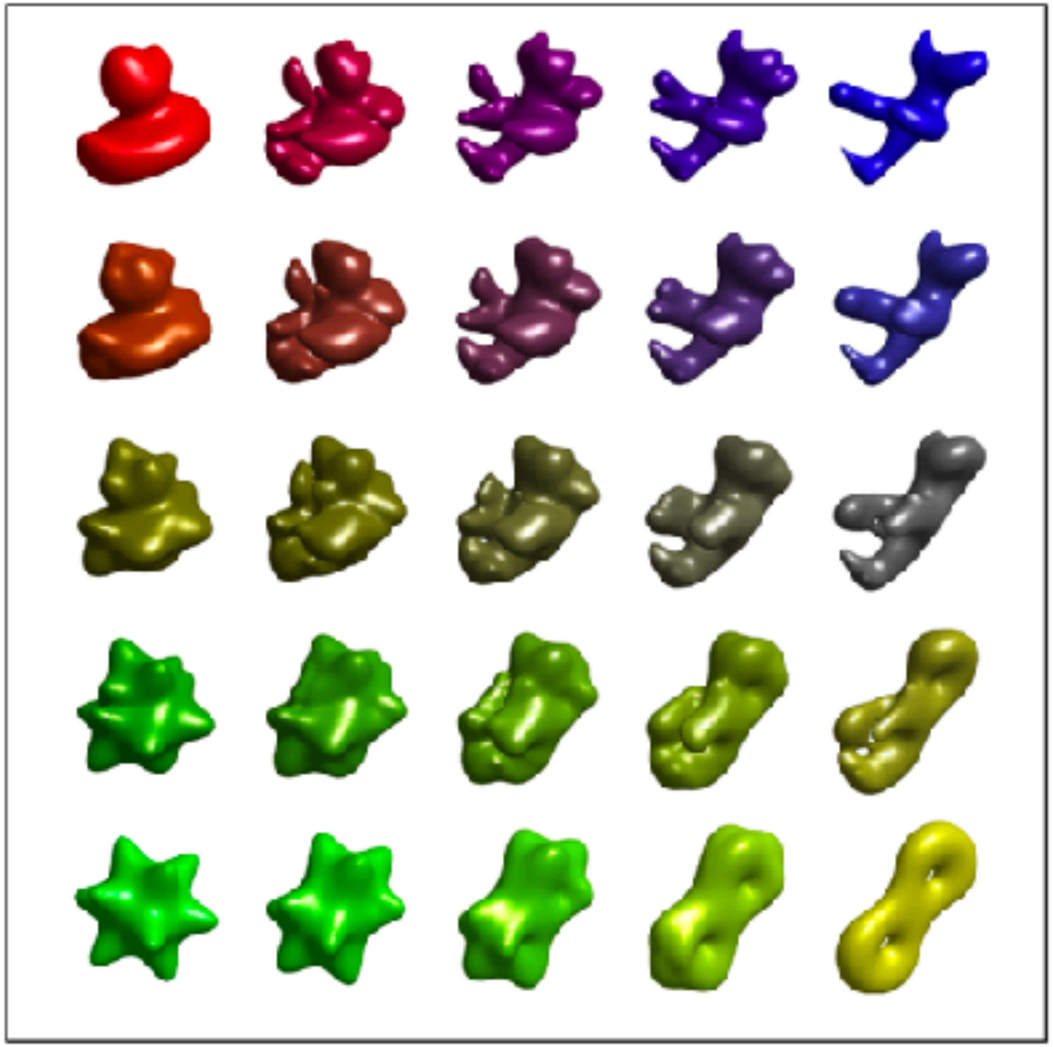


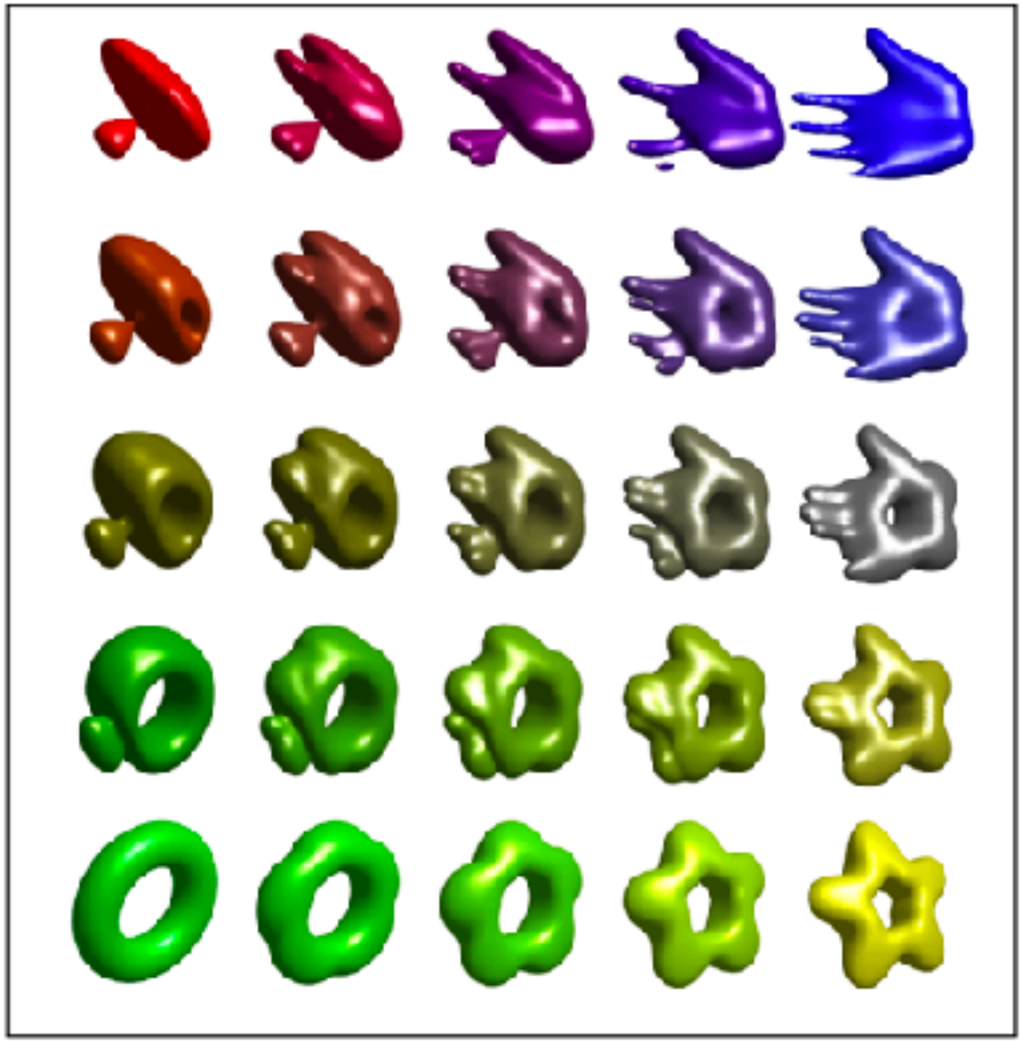
Barycenters of 2D Shapes



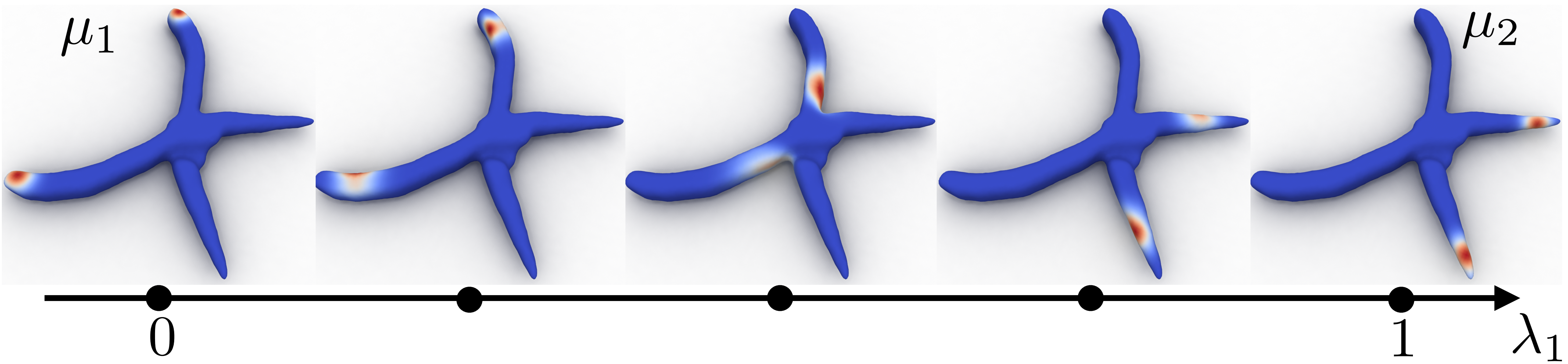


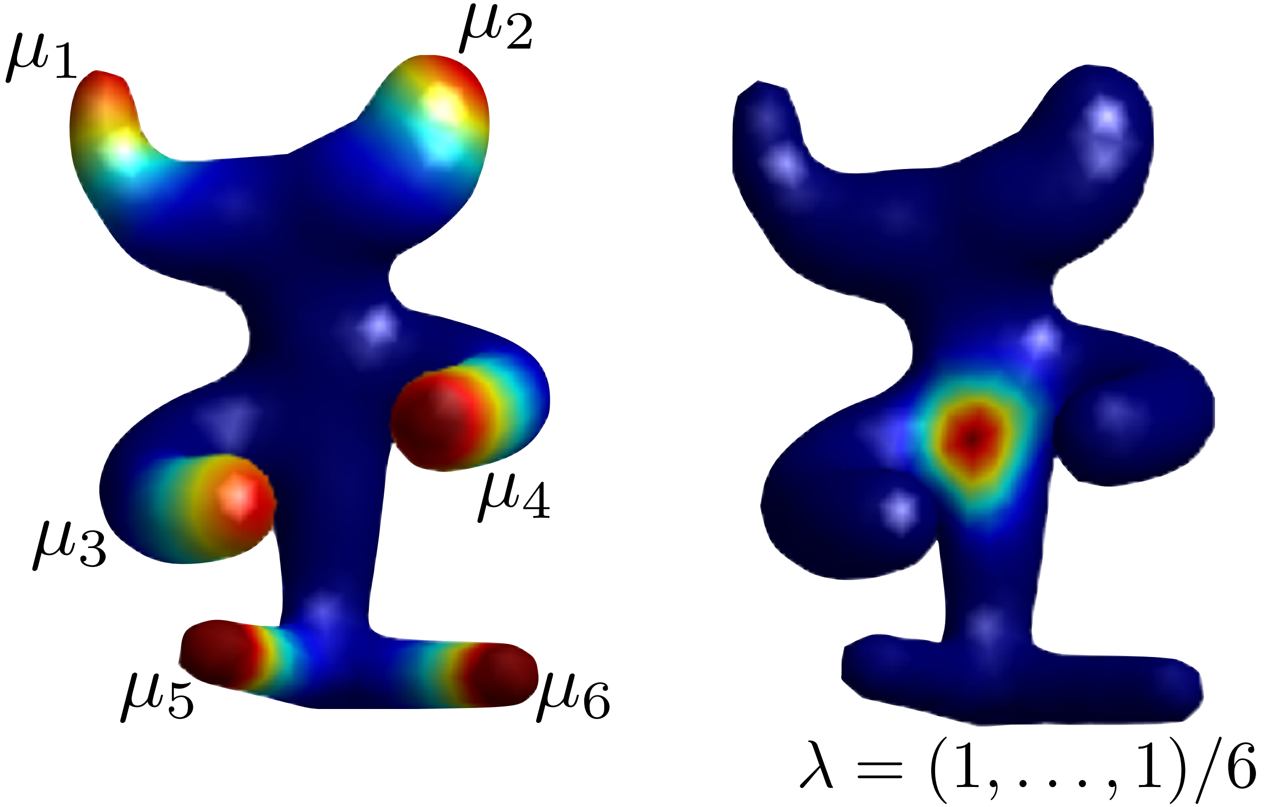
Barycenters of 3D Shapes





Barycenters on a Surface





Application: Color Harmonization

Overview

- Entropic Regularization and Sinkhorn
- Convergence Analysis
- Barycenters
- Generalized Sinkhorn and Applications (dedicated section)