Numerical Optimal Transport and Applications

Gabriel Peyré

Joint works with:
Jean-David Benamou, Guillaume Carlier,
Marco Cuturi, Luca Nenna, Justin Solomon

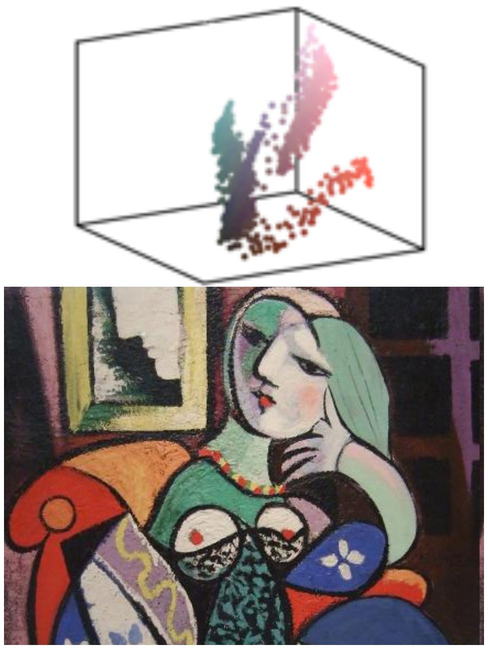
CEREMADE

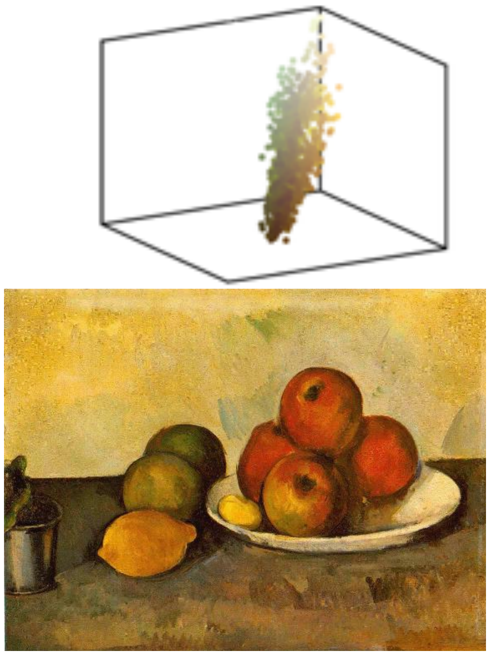
DAUPHINE UNIVERSITÉ PARIS

www.numerical-tours.com

Histograms in Imaging and Machine Learning

Color histograms:

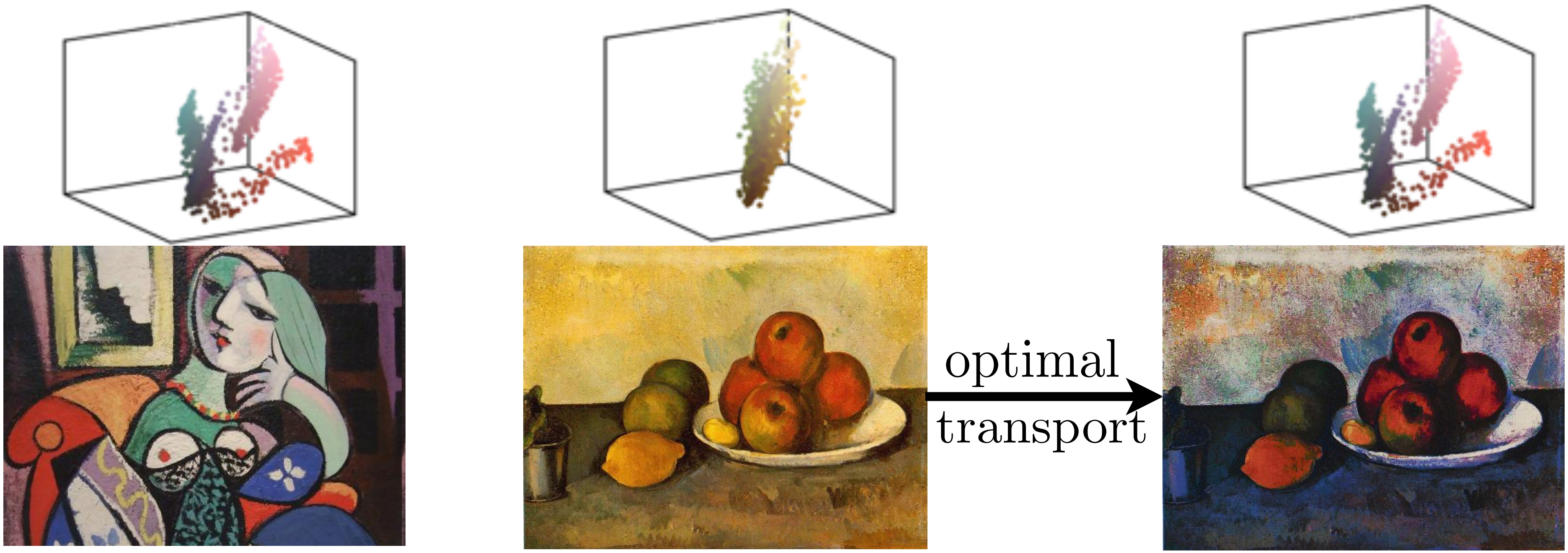




Input image

Histograms in Imaging and Machine Learning

Color histograms:

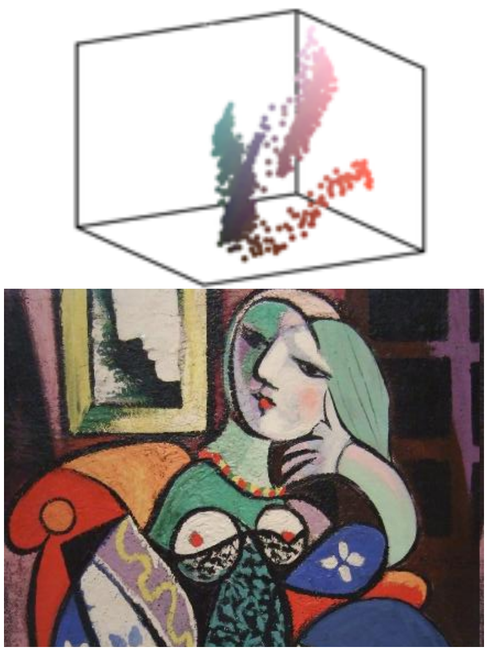


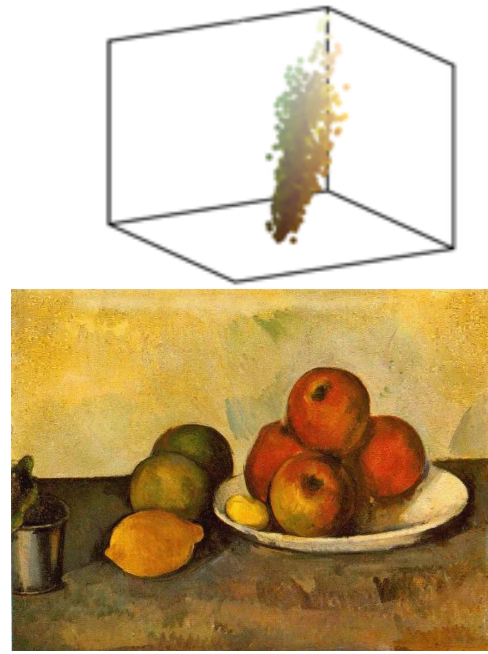
Input image

Modified image

Histograms in Imaging and Machine Learning

Color histograms:

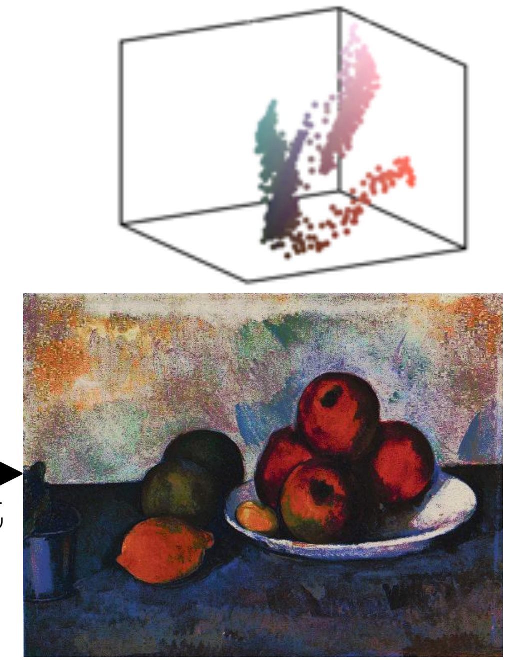




Input image

optimal

transport



Modified image

Bag of words:

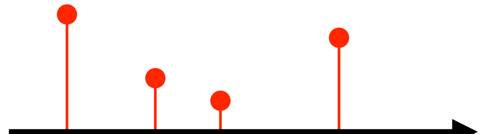
Overview

- Transportation-like Problems
- Regularized Transport
- Optimal Transport Barycenters
- Heat Kernel Approximation

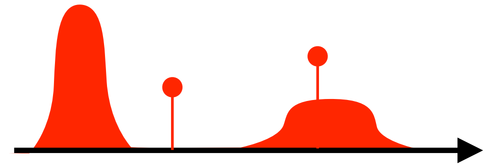
Radon Measures and Couplings

Positive Radon measures $\mathcal{M}_+(X)$: On a metric space (X, d).

$\mathrm{d}\mu(x) = f(x)\mathrm{d}x$



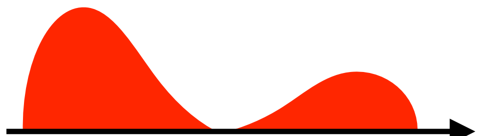
$\mu = \sum_i p_i \delta_{x_i}$



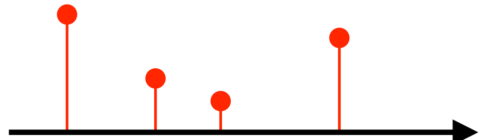
$X = \mathbb{R}$

Radon Measures and Couplings

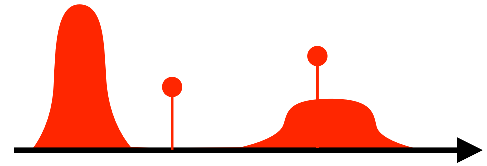
Positive Radon measures $\mathcal{M}_+(X)$: On a metric space (X,d).



$\mathrm{d}\mu(x) = f(x)\mathrm{d}x$



$\mu = \sum_i p_i \delta_{x_i}$

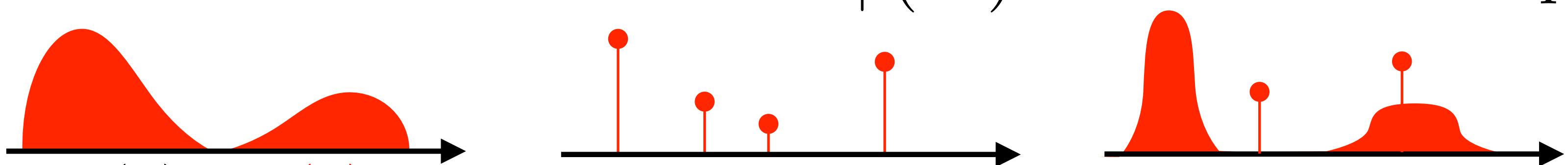


$X = \mathbb{R}$

Probability measures. $\mu(X) = 1$ $\int f = 1$ $\sum_i p_i = 1$

Radon Measures and Couplings

Positive Radon measures $\mathcal{M}_+(X)$: On a metric space (X, d).

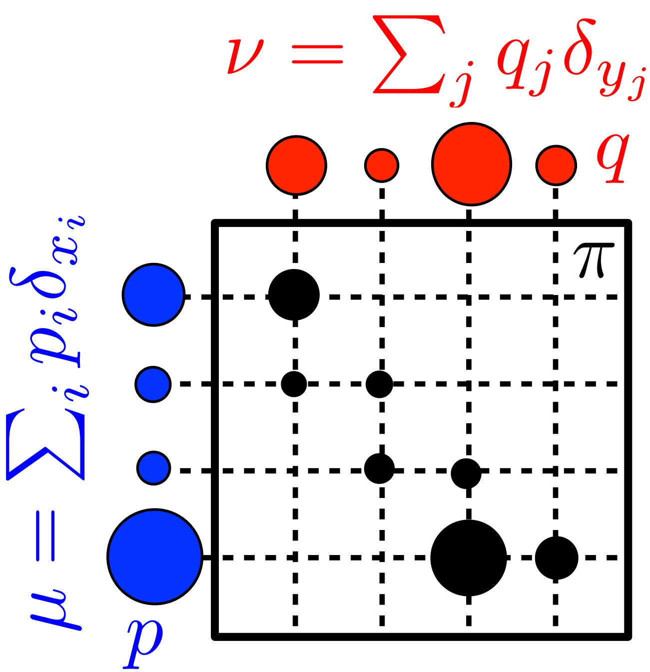


$\mathrm{d}\mu(x) = f(x)\mathrm{d}x$ $\quad$ $\mu = \sum_i p_i \delta_{x_i}$ $\quad$ $X = \mathbb{R}$

Probability measures. $\mu(X) = 1$ $\quad$ $\int f = 1$ $\quad$ $\sum_i p_i = 1$

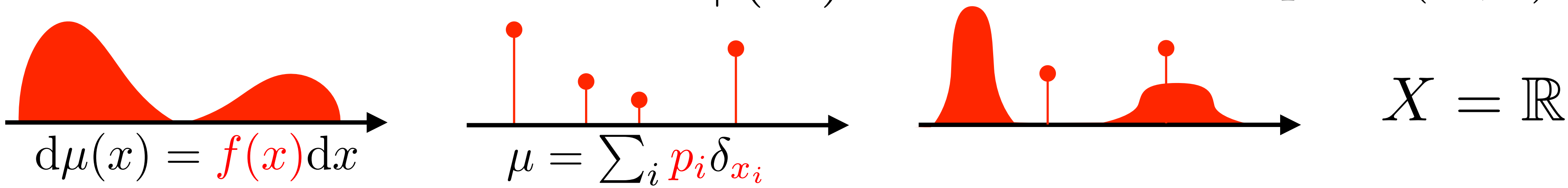
Couplings: $\Pi(\mu, \nu) \overset{\text{def.}}{=} \{\pi \in \mathcal{M}_+(X \times X) \; ; \; P_{1\sharp}\pi = \mu, P_{2\sharp}\pi = \nu\}$

Marginals: $P_{1\sharp}\pi(S) \overset{\text{def.}}{=} \pi(S, X)$ $\quad$ $P_{2\sharp}\pi(S) \overset{\text{def.}}{=} \pi(X, S)$



Radon Measures and Couplings

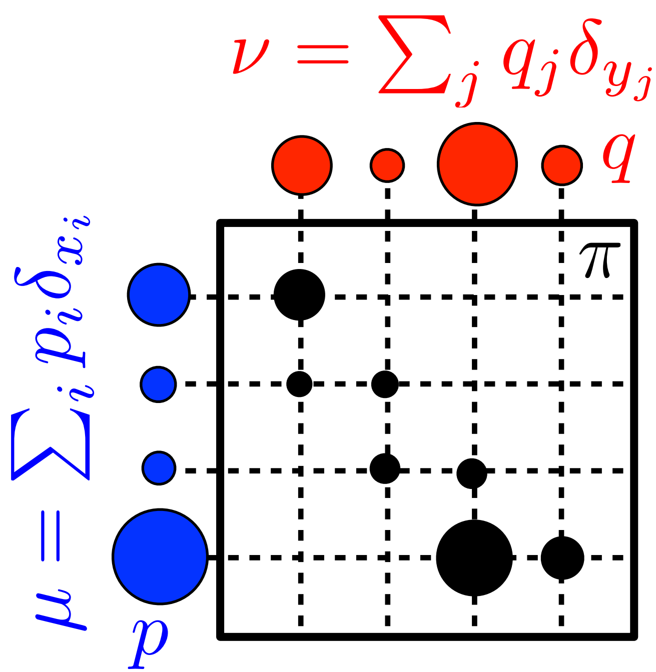
Positive Radon measures $\mathcal{M}_+(X)$: On a metric space (X,d).

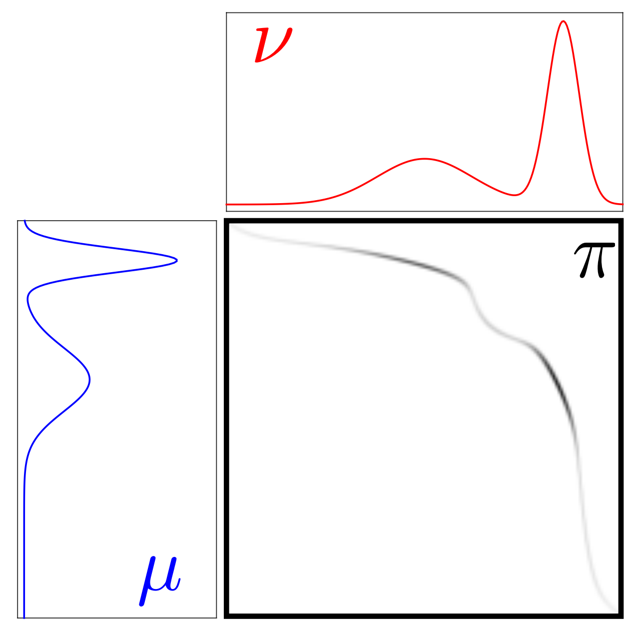


Probability measures. $\mu(X) = 1$ $\quad \int f = 1$ $\quad \sum_i p_i = 1$

Couplings: $\Pi(\mu,\nu) \overset{\text{def.}}{=} \{\pi \in \mathcal{M}_+(X \times X) \; ; \; P_{1\sharp}\pi = \mu, P_{2\sharp}\pi = \nu\}$

Marginals: $P_{1\sharp}\pi(S) \overset{\text{def.}}{=} \pi(S,X)$ $\quad P_{2\sharp}\pi(S) \overset{\text{def.}}{=} \pi(X,S)$





Optimal Transport

Ground cost $c(x, y)$ on $X \times X$.

Optimal transport: [Kantorovitch 1942]

$$\min_{\pi} \left\{ \langle c,\, \pi \rangle = \int_{X \times X} c(x,y) \mathrm{d}\pi(x,y) \;;\; \pi \in \Pi(\mu, \nu) \right\}$$

Optimal Transport

Ground cost $c(x, y)$ on $X \times X$.

Optimal transport: [Kantorovitch 1942]

$$W_\alpha(\mu, \nu)^\alpha \stackrel{\text{def.}}{=} \min_\pi \left\{ \langle c,\, \pi \rangle = \int_{X \times X} c(x, y) \mathrm{d}\pi(x, y) \;;\; \pi \in \Pi(\mu, \nu) \right\}$$

For $c(x, y) = d(x, y)^\alpha$, α-Wasserstein distance W_α.

Optimal Transport

Ground cost $c(x,y)$ on $X \times X$.

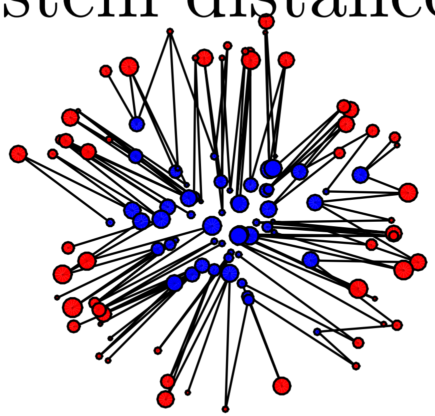
Optimal transport: [Kantorovitch 1942]

$$W_\alpha(\mu,\nu)^\alpha \overset{\text{def.}}{=} \min_\pi \left\{ \langle c,\, \pi \rangle = \int_{X\times X} c(x,y)\mathrm{d}\pi(x,y) \;;\; \pi \in \Pi(\mu,\nu) \right\}$$

For $c(x,y) = d(x,y)^\alpha$, α-Wasserstein distance W_α.

Linear programming:

$$\mu = \sum_{i=1}^{N_1} p_i \delta_{x_i}, \nu = \sum_{j=1}^{N_2} p_j \delta_{y_i}$$



Optimal Transport

Ground cost $c(x,y)$ on $X \times X$.

Optimal transport: [Kantorovitch 1942]

$$W_\alpha(\mu,\nu)^\alpha \stackrel{\text{def.}}{=} \min_\pi \left\{ \langle c,\, \pi \rangle = \int_{X\times X} c(x,y)\mathrm{d}\pi(x,y) \;;\; \pi \in \Pi(\mu,\nu) \right\}$$

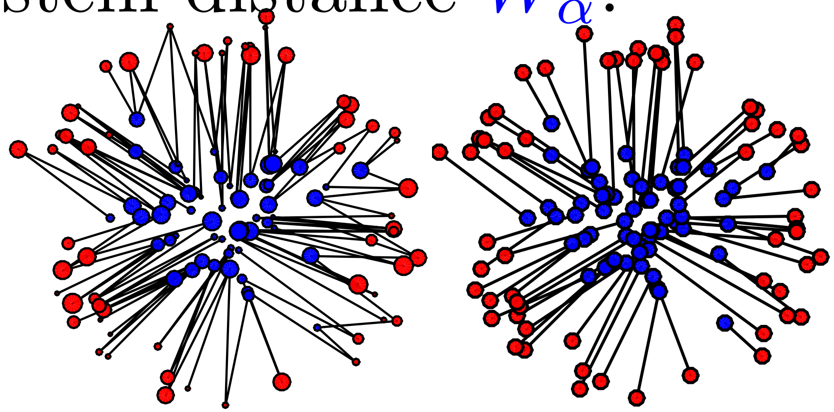
For $c(x,y) = d(x,y)^\alpha$, α-Wasserstein distance W_α.

Linear programming:

$$\mu = \sum_{i=1}^{N_1} p_i \delta_{x_i}, \nu = \sum_{j=1}^{N_2} p_j \delta_{y_i}$$

Hungarian/Auction: $\sim O(N^3)$

$$\mu = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}, \nu = \frac{1}{N} \sum_{j=1}^{N} \delta_{y_j}$$



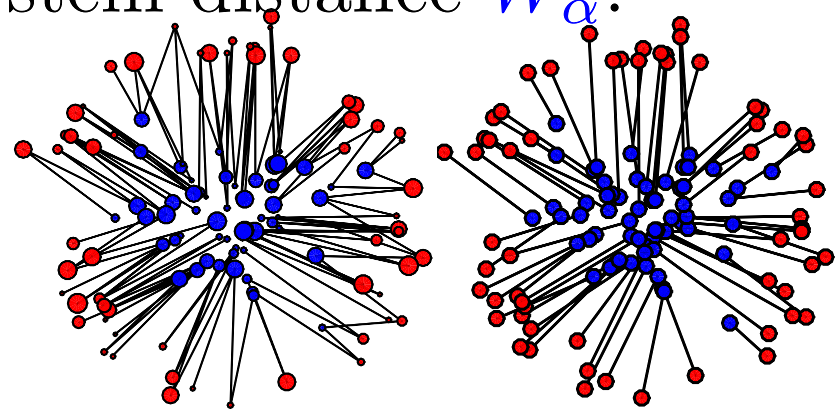
Optimal Transport

Ground cost $c(x,y)$ on $X \times X$.

Optimal transport: [Kantorovitch 1942]

$$W_\alpha(\mu,\nu)^\alpha \overset{\text{def.}}{=} \min_\pi \left\{ \langle c,\, \pi \rangle = \int_{X\times X} c(x,y)\mathrm{d}\pi(x,y) \;;\; \pi \in \Pi(\mu,\nu) \right\}$$

For $c(x,y) = d(x,y)^\alpha$, α-Wasserstein distance W_α.

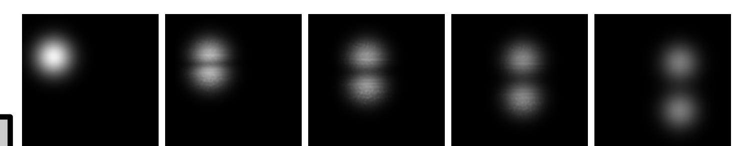


Linear programming:

$$\mu = \sum_{i=1}^{N_1} p_i \delta_{x_i}, \nu = \sum_{j=1}^{N_2} p_j \delta_{y_i}$$

Hungarian/Auction: $\sim O(N^3)$

$$\mu = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_i}, \nu = \frac{1}{N} \sum_{j=1}^{N} \delta_{y_j}$$



Monge-Ampère/Benamou-Brenier, $d = \|\cdot\|_2^2$.

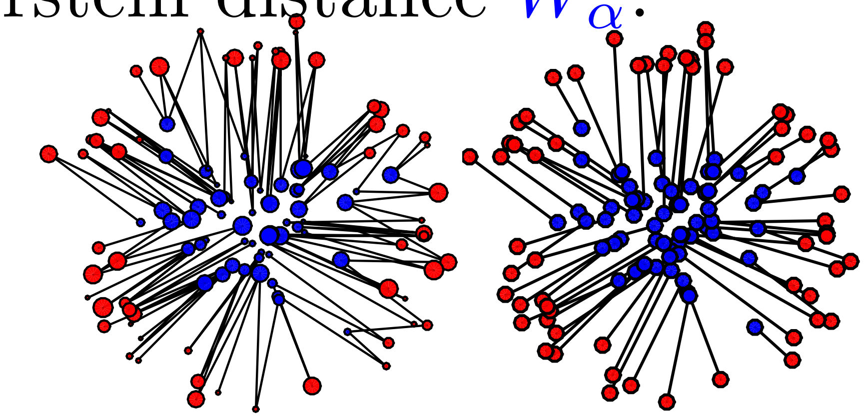
Optimal Transport

Ground cost $c(x,y)$ on $X \times X$.

Optimal transport: [Kantorovitch 1942]

$$W_\alpha(\mu,\nu)^\alpha \overset{\text{def.}}{=} \min_\pi \left\{ \langle c,\, \pi \rangle = \int_{X\times X} c(x,y)\mathrm{d}\pi(x,y) \;;\; \pi \in \Pi(\mu,\nu) \right\}$$

For $c(x,y) = d(x,y)^\alpha$, α-Wasserstein distance W_α.

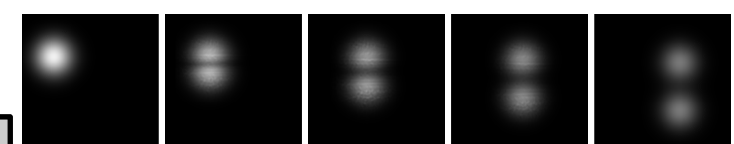


Linear programming:

$$\mu = \sum_{i=1}^{N_1} p_i \delta_{x_i}, \nu = \sum_{j=1}^{N_2} p_j \delta_{y_i}$$

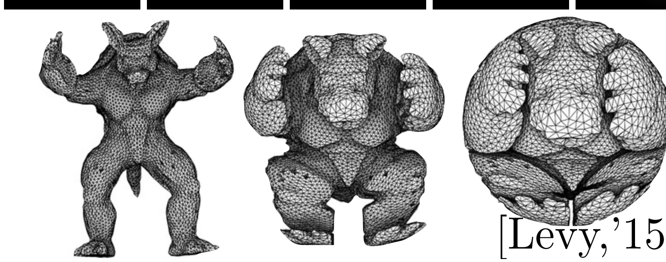
Hungarian/Auction: $\sim O(N^3)$

$$\mu = \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}, \nu = \frac{1}{N}\sum_{j=1}^{N} \delta_{y_j}$$



Monge-Ampère/Benamou-Brenier, $d = \|\cdot\|_2^2$.

Semi-discrete: Laguerre cells, $d = \|\cdot\|_2^2$.



[Levy,'15]

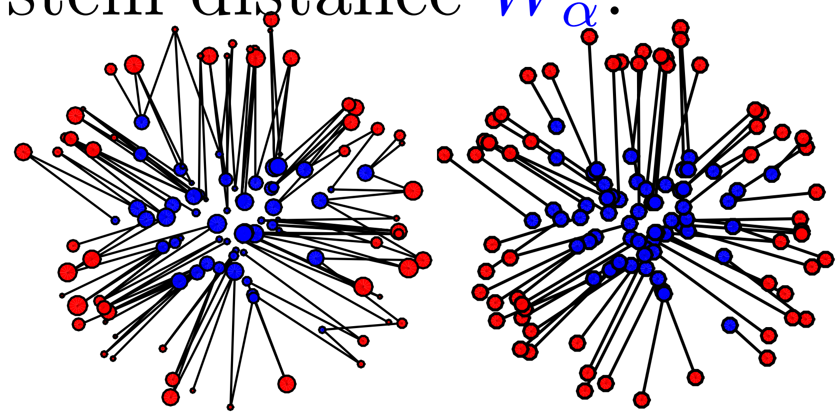
Optimal Transport

Ground cost $c(x,y)$ on $X \times X$.

Optimal transport: [Kantorovitch 1942]

$$W_\alpha(\mu,\nu)^\alpha \overset{\text{def.}}{=} \min_\pi \left\{ \langle c,\, \pi \rangle = \int_{X\times X} c(x,y)\mathrm{d}\pi(x,y) \;;\; \pi \in \Pi(\mu,\nu) \right\}$$

For $c(x,y) = d(x,y)^\alpha$, α-Wasserstein distance W_α.

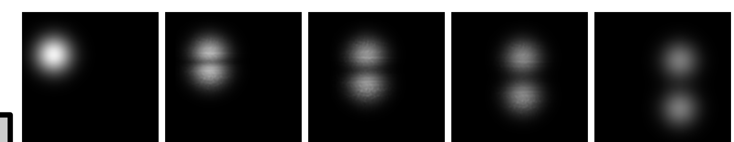


Linear programming:

$$\mu = \sum_{i=1}^{N_1} p_i \delta_{x_i}, \nu = \sum_{j=1}^{N_2} p_j \delta_{y_i}$$

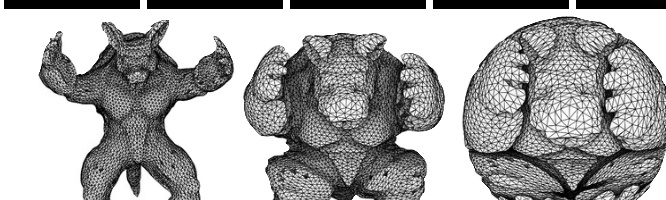
Hungarian/Auction: $\sim O(N^3)$

$$\mu = \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}, \nu = \frac{1}{N}\sum_{j=1}^{N} \delta_{y_j}$$



Monge-Ampère/Benamou-Brenier, $d = \|\cdot\|_2^2$.

Semi-discrete: Laguerre cells, $d = \|\cdot\|_2^2$.



[Levy,'15]

Need for fast approximate algorithms for generic c.

Unbalanced Transport

$$(\mu, \nu) \in \mathcal{M}_+(X) \qquad \mathrm{KL}(\nu|\mu) \overset{\text{def.}}{=} \int_X \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right) \mathrm{d}\mu + \int_X (\mathrm{d}\mu - \mathrm{d}\nu)$$

Unbalanced Transport

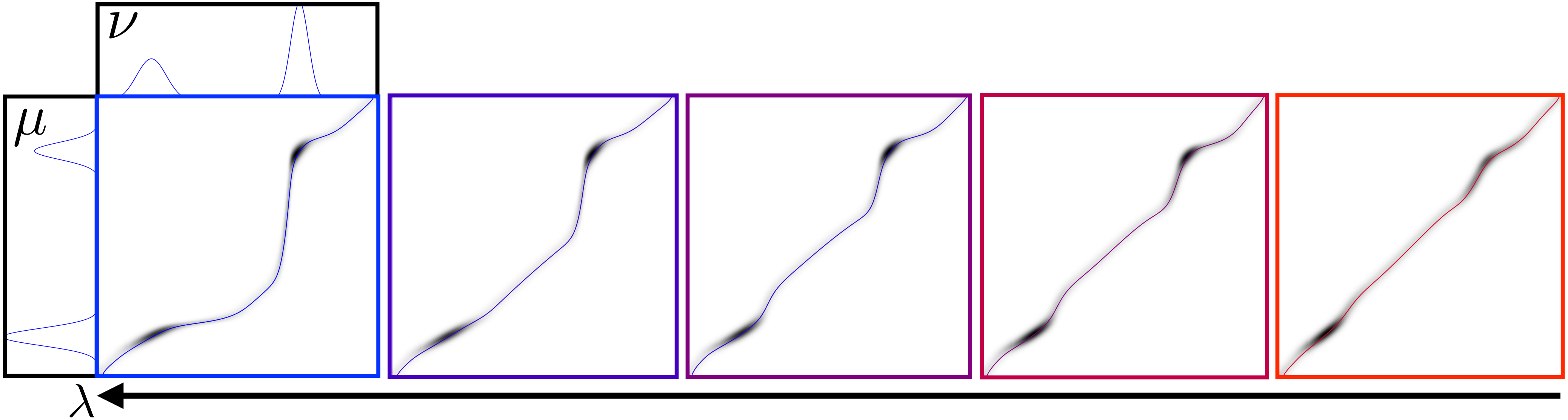
$$(\mu, \nu) \in \mathcal{M}_+(X) \qquad \mathrm{KL}(\nu|\mu) \stackrel{\text{def.}}{=} \int_X \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right) \mathrm{d}\mu + \int_X (\mathrm{d}\mu - \mathrm{d}\nu)$$

$$\boxed{WF_c(\mu, \nu) \stackrel{\text{def.}}{=} \min_{\pi} \langle c,\, \pi \rangle + \lambda \mathrm{KL}(P_{1\sharp}\pi|\mu) + \lambda \mathrm{KL}(P_{2\sharp}\pi|\nu)}$$

Unbalanced Transport

$$(\mu, \nu) \in \mathcal{M}_+(X) \qquad \mathrm{KL}(\nu|\mu) \stackrel{\text{def.}}{=} \int_X \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right) \mathrm{d}\mu + \int_X (\mathrm{d}\mu - \mathrm{d}\nu)$$

$$\boxed{WF_c(\mu, \nu) \stackrel{\text{def.}}{=} \min_\pi \, \langle c, \, \pi \rangle + \lambda \mathrm{KL}(P_{1\sharp}\pi|\mu) + \lambda \mathrm{KL}(P_{2\sharp}\pi|\nu)}$$



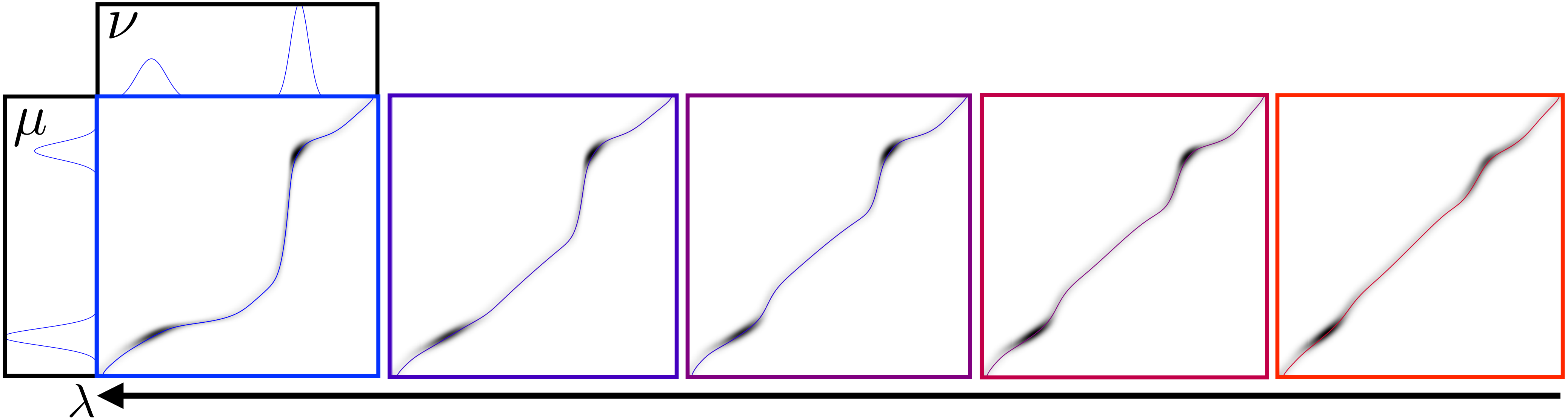
Unbalanced Transport

$$(\mu, \nu) \in \mathcal{M}_+(X) \qquad \mathrm{KL}(\nu|\mu) \overset{\text{def.}}{=} \int_X \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right) \mathrm{d}\mu + \int_X (\mathrm{d}\mu - \mathrm{d}\nu)$$

$$WF_c(\mu, \nu) \overset{\text{def.}}{=} \min_{\pi} \langle c,\, \pi \rangle + \lambda \mathrm{KL}(P_{1\sharp}\pi|\mu) + \lambda \mathrm{KL}(P_{2\sharp}\pi|\nu)$$

Proposition: If $c(x, y) = -\log(\cos(\min(\frac{d(x,y)}{\delta}, \frac{\pi}{2}))$
then $WF_c^{1/2}$ is a distance on $\mathcal{M}_+(X)$.

[Liereo, Mielke, Savaré 2015] [Chizat, Schmitzer, Peyré, Vialard 2015]



Unbalanced Transport

$$(\mu, \nu) \in \mathcal{M}_+(X) \qquad \mathrm{KL}(\nu|\mu) \stackrel{\text{def.}}{=} \int_X \log\left(\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right) \mathrm{d}\mu + \int_X (\mathrm{d}\mu - \mathrm{d}\nu)$$

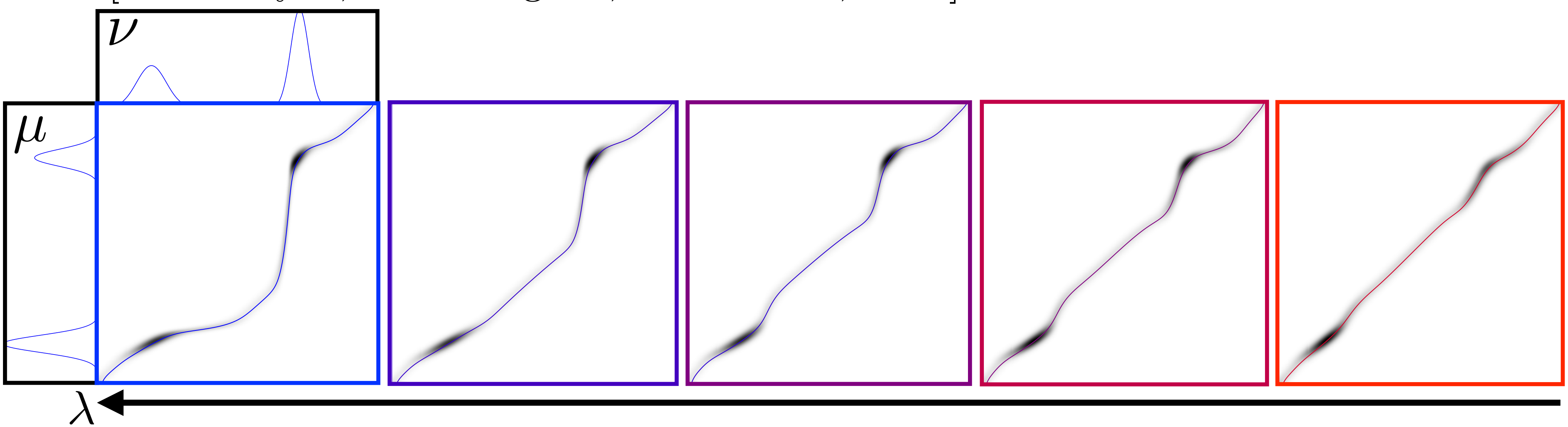
$$WF_c(\mu, \nu) \stackrel{\text{def.}}{=} \min_\pi \langle c,\, \pi\rangle + \lambda \mathrm{KL}(P_{1\sharp}\pi|\mu) + \lambda \mathrm{KL}(P_{2\sharp}\pi|\nu)$$

Proposition: If $c(x, y) = -\log(\cos(\min(\frac{d(x,y)}{\delta}, \frac{\pi}{2}))$
then $WF_c^{1/2}$ is a distance on $\mathcal{M}_+(X)$.

[Liereo, Mielke, Savaré 2015] [Chizat, Schmitzer, Peyré, Vialard 2015]

$\rightarrow$ "Dynamic" Benamou-Brenier formulation.

[Liereo, Mielke, Savaré 2015] [Chizat, Schmitzer, Peyré, Vialard 2015]
[Kondratyev, Monsaingeon, Vorotnikov, 2015]



Wasserstein Gradient Flows

Implicit Euler step: [Jordan, Kinderlehrer, Otto 1998]

$$\mu_{t+1} = \mathrm{Prox}^{W}_{\tau f}(\mu_t) \stackrel{\text{def.}}{=} \underset{\mu \in \mathcal{M}_+(X)}{\operatorname{argmin}} \; W_\alpha^\alpha(\mu_t, \mu) + \tau f(\mu)$$

Wasserstein Gradient Flows

Implicit Euler step: [Jordan, Kinderlehrer, Otto 1998]

$$\mu_{t+1} = \mathrm{Prox}^{W}_{\tau f}(\mu_t) \overset{\text{def.}}{=} \underset{\mu \in \mathcal{M}_+(X)}{\mathrm{argmin}} \; W_\alpha^\alpha(\mu_t, \mu) + \tau f(\mu)$$

Formal limit $\tau \to 0$: $\quad \partial_t \mu = \mathrm{div}\left(\mu \nabla (f'(\mu))\right)$

Wasserstein Gradient Flows

Implicit Euler step: [Jordan, Kinderlehrer, Otto 1998]

$$\boxed{\mu_{t+1} = \mathrm{Prox}^{W}_{\tau f}(\mu_t) \overset{\text{def.}}{=} \underset{\mu \in \mathcal{M}_+(X)}{\mathrm{argmin}} \; W_\alpha^\alpha(\mu_t, \mu) + \tau f(\mu)}$$

Formal limit $\tau \to 0$: $\partial_t \mu = \mathrm{div}\left(\mu \nabla (f'(\mu))\right)$

$$f(\mu) = \int \log(\tfrac{\mathrm{d}\mu}{\mathrm{d}x})\mathrm{d}\mu \longrightarrow \partial_t \mu = \Delta \mu \qquad \text{(heat diffusion)}$$

Wasserstein Gradient Flows

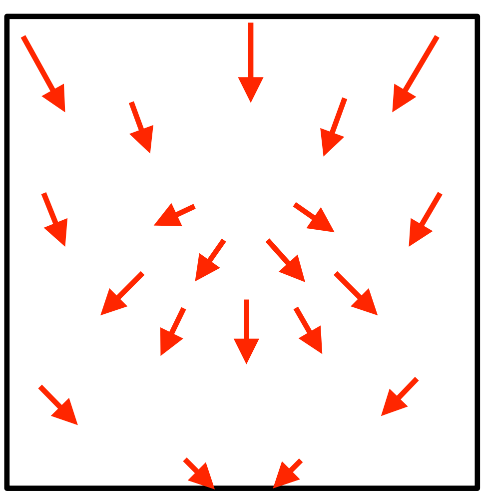
Implicit Euler step: [Jordan, Kinderlehrer, Otto 1998]

$$\mu_{t+1} = \mathrm{Prox}^{W}_{\tau f}(\mu_t) \overset{\text{def.}}{=} \underset{\mu \in \mathcal{M}_+(X)}{\mathrm{argmin}} \; W_\alpha^\alpha(\mu_t, \mu) + \tau f(\mu)$$

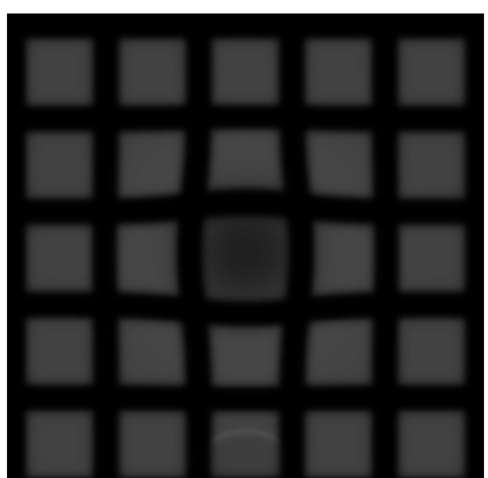
Formal limit $\tau \to 0$: $\quad \partial_t \mu = \mathrm{div}\left(\mu \nabla (f'(\mu))\right)$

$$f(\mu) = \int \log(\tfrac{\mathrm{d}\mu}{\mathrm{d}x})\mathrm{d}\mu \longrightarrow \partial_t \mu = \Delta \mu \qquad \text{(heat diffusion)}$$

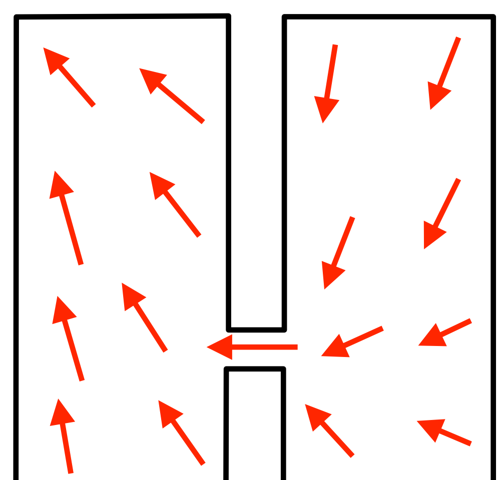
$$f(\mu) = \int w \mathrm{d}\mu \longrightarrow \partial_t \mu = \mathrm{div}(\mu \nabla w) \qquad \text{(advection)}$$



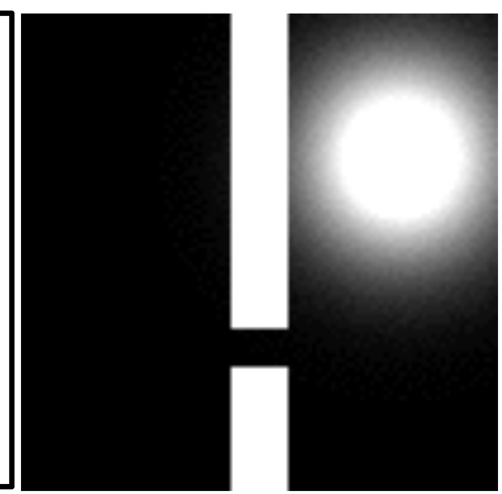
∇w



Evolution μ_t



∇w



Evolution μ_t

Wasserstein Gradient Flows

Implicit Euler step: [Jordan, Kinderlehrer, Otto 1998]

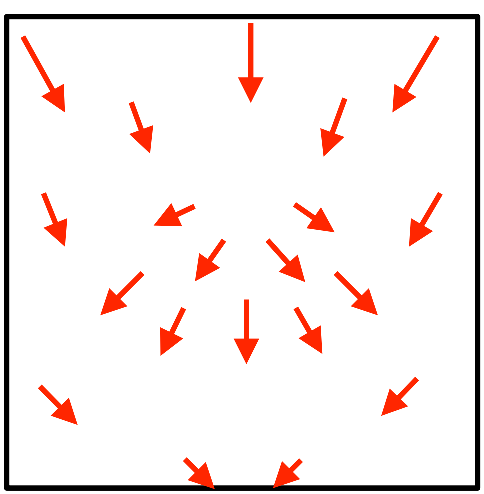
$$\mu_{t+1} = \mathrm{Prox}^{W}_{\tau f}(\mu_t) \overset{\text{def.}}{=} \underset{\mu \in \mathcal{M}_+(X)}{\operatorname{argmin}} \; W_\alpha^\alpha(\mu_t, \mu) + \tau f(\mu)$$

Formal limit $\tau \to 0$: $\partial_t \mu = \mathrm{div}\left(\mu \nabla (f'(\mu))\right)$

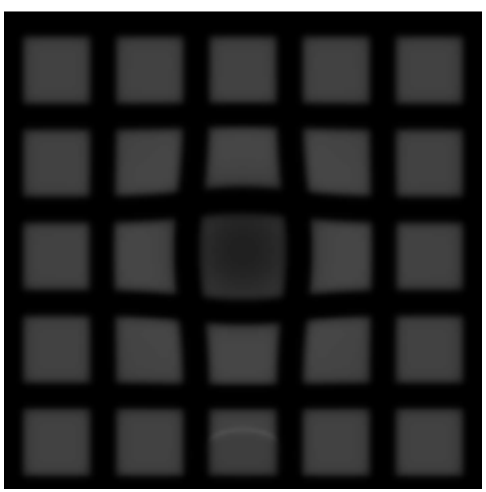
$f(\mu) = \int \log(\frac{\mathrm{d}\mu}{\mathrm{d}x})\mathrm{d}\mu \longrightarrow \partial_t \mu = \Delta \mu$ (heat diffusion)

$f(\mu) = \int w \mathrm{d}\mu \longrightarrow \partial_t \mu = \mathrm{div}(\mu \nabla w)$ (advection)

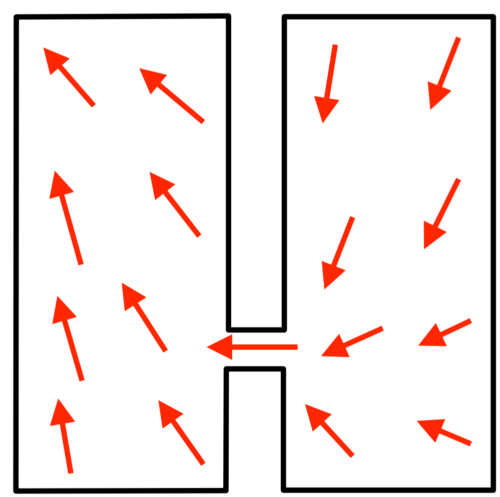
$f(\mu) = \frac{1}{m-1} \int (\frac{\mathrm{d}\mu}{\mathrm{d}x})^{m-1}\mathrm{d}\mu \longrightarrow \partial_t \mu = \Delta \mu^m$ (non-linear diffusion)



∇w



Evolution μ_t



∇w

Evolution μ_t

Overview

- Transportation-like Problems
- **Regularized Transport**
- Optimal Transport Barycenters
- Heat Kernel Approximation

Transportation-like Problems

$$\min_{\pi \in \mathcal{M}_+(X \times X)} \mathcal{E}(\pi) \stackrel{\text{def.}}{=} \langle c,\, \pi \rangle + f_1(P_{1\sharp}\pi) + f_2(P_{2\sharp}\pi)$$

Transportation-like Problems

$$\min_{\pi \in \mathcal{M}_+(X \times X)} \mathcal{E}(\pi) \stackrel{\text{def.}}{=} \langle c, \pi \rangle + f_1(P_{1\sharp}\pi) + f_2(P_{2\sharp}\pi)$$

Optimal transport: $f_1 = \iota_\mu$ $f_2 = \iota_\nu$

Transportation-like Problems

$$\min_{\pi \in \mathcal{M}_+(X \times X)} \mathcal{E}(\pi) \stackrel{\text{def.}}{=} \langle c,\, \pi \rangle + f_1(P_{1\sharp}\pi) + f_2(P_{2\sharp}\pi)$$

Optimal transport: $f_1 = \iota_\mu$ $f_2 = \iota_\nu$

Unbalanced transport: $f_1 = \lambda \mathrm{KL}(\cdot|\mu)$ $f_2 = \lambda \mathrm{KL}(\cdot|\nu)$

Transportation-like Problems

$$\min_{\pi \in \mathcal{M}_+(X \times X)} \mathcal{E}(\pi) \overset{\text{def.}}{=} \langle c,\, \pi \rangle + f_1(P_{1\sharp}\pi) + f_2(P_{2\sharp}\pi)$$

Optimal transport:	$f_1 = \iota_\mu$	$f_2 = \iota_\nu$
Unbalanced transport:	$f_1 = \lambda \mathrm{KL}(\cdot \vert \mu)$	$f_2 = \lambda \mathrm{KL}(\cdot \vert \nu)$
Gradient flows:	$f_1 = \iota_{\mu_t}$	$f_2 = f$

Transportation-like Problems

$$\min_{\pi \in \mathcal{M}_+(X \times X)} \mathcal{E}(\pi) \stackrel{\text{def.}}{=} \langle c,\, \pi \rangle + f_1(P_{1\sharp}\pi) + f_2(P_{2\sharp}\pi)$$

Optimal transport: $f_1 = \iota_\mu$ $f_2 = \iota_\nu$

Unbalanced transport: $f_1 = \lambda \mathrm{KL}(\cdot|\mu)$ $f_2 = \lambda \mathrm{KL}(\cdot|\nu)$

Gradient flows: $f_1 = \iota_{\mu_t}$ $f_2 = f$

Regularization: $\min_\pi \mathcal{E}(\pi) + \varepsilon \mathrm{KL}(\pi|\pi_0)$

$\varepsilon \mathrm{KL}(\pi|\pi_0)$ Regularization and positivity barrier.

Transportation-like Problems

$$\min_{\pi \in \mathcal{M}_+(X \times X)} \mathcal{E}(\pi) \overset{\text{def.}}{=} \langle c,\, \pi \rangle + f_1(P_{1\sharp}\pi) + f_2(P_{2\sharp}\pi)$$

Optimal transport:	$f_1 = \iota_\mu$	$f_2 = \iota_\nu$
Unbalanced transport:	$f_1 = \lambda \mathrm{KL}(\cdot \vert \mu)$	$f_2 = \lambda \mathrm{KL}(\cdot \vert \nu)$
Gradient flows:	$f_1 = \iota_{\mu_t}$	$f_2 = f$

Regularization: $$\min_{\pi} \mathcal{E}(\pi) + \varepsilon \mathrm{KL}(\pi|\pi_0)$$

$\varepsilon \mathrm{KL}(\pi|\pi_0)$ → Regularization and positivity barrier.

$\varepsilon \mathrm{KL}(\pi|\pi_0)$ → Discretization grid (prescribed support).

Transportation-like Problems

$$\min_{\pi \in \mathcal{M}_+(X \times X)} \mathcal{E}(\pi) \stackrel{\text{def.}}{=} \langle c,\, \pi \rangle + f_1(P_{1\sharp}\pi) + f_2(P_{2\sharp}\pi)$$

Optimal transport:	$f_1 = \iota_\mu$	$f_2 = \iota_\nu$
Unbalanced transport:	$f_1 = \lambda \mathrm{KL}(\cdot\|\mu)$	$f_2 = \lambda \mathrm{KL}(\cdot\|\nu)$
Gradient flows:	$f_1 = \iota_{\mu_t}$	$f_2 = f$

Regularization: $\mathrm{Prox}^{\mathrm{KL}}_{\mathcal{E}/\varepsilon}(\pi_0) \stackrel{\text{def.}}{=} \underset{\pi}{\mathrm{argmin}}\, \mathcal{E}(\pi) + \varepsilon \mathrm{KL}(\pi|\pi_0)$

$\varepsilon \mathrm{KL}(\pi|\pi_0)$ →
- Regularization and positivity barrier.
- Discretization grid (prescribed support).
- Implicit KL stepping.

π_0

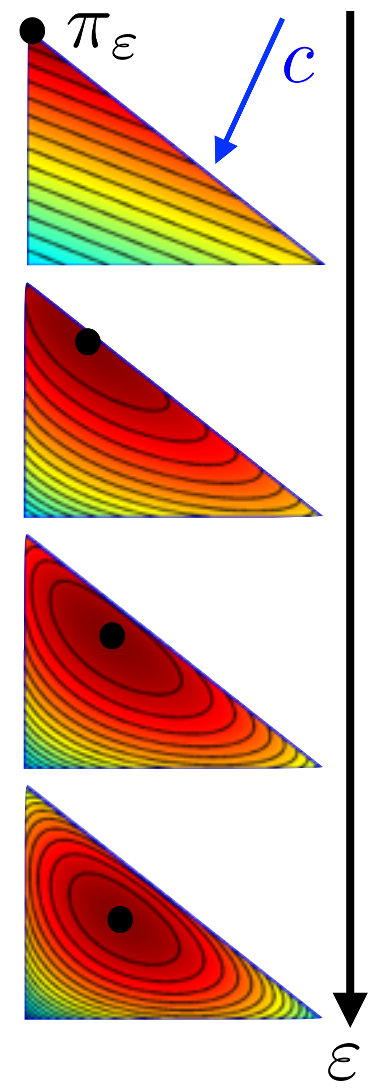
$\pi_1 = \mathrm{Prox}_{\frac{1}{\varepsilon}\mathcal{E}}(\pi_0)$

$\pi_2 = \mathrm{Prox}_{\frac{1}{\varepsilon}\mathcal{E}}(\pi_1)$

$\mathrm{argmin}\, \mathcal{E}(\pi)$

Entropy Regularized Transport

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



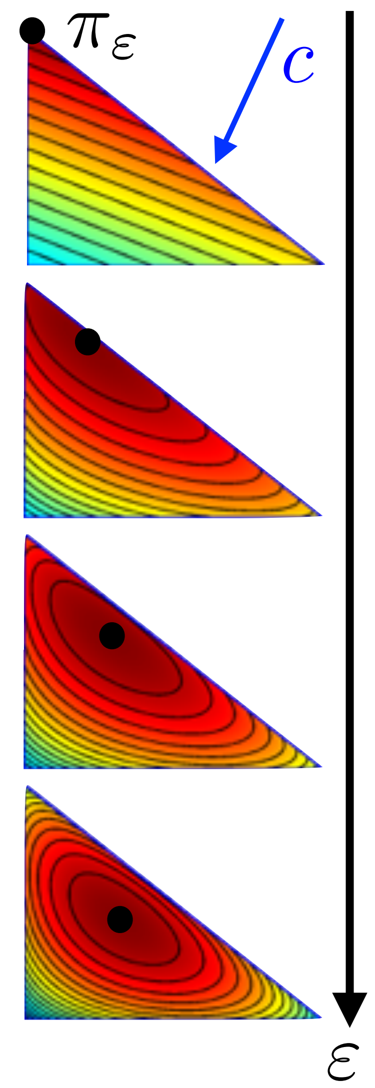
Entropy Regularized Transport

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

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

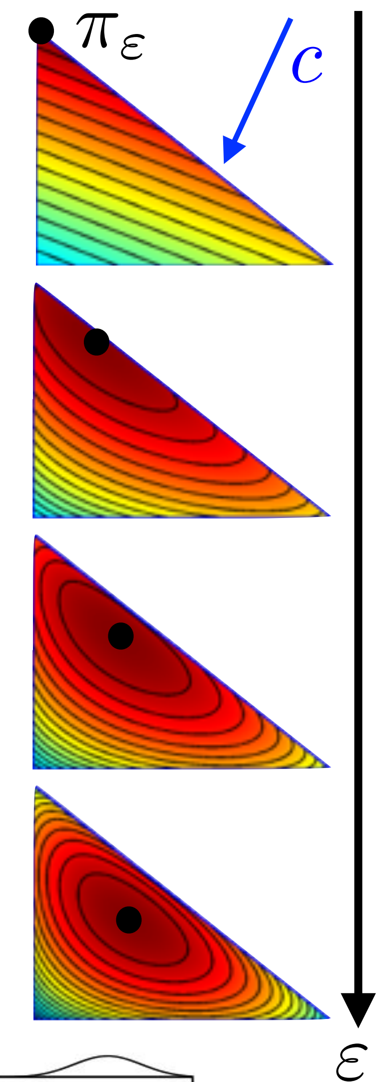
Gibbs kernel: $K(x,y) \overset{\text{def.}}{=} e^{-\frac{c(x,y)}{\varepsilon}} \pi_0(x,y)$

Landmark computational paper: [Cuturi 2013].



Entropy Regularized Transport

$$\pi_\varepsilon \overset{\text{def.}}{=} \underset{\pi}{\operatorname{argmin}} \; \{\langle c,\, \pi\rangle + \varepsilon \mathrm{KL}(\pi|\pi_0) \; ; \; \pi \in \Pi(\mu,\nu)\}$$



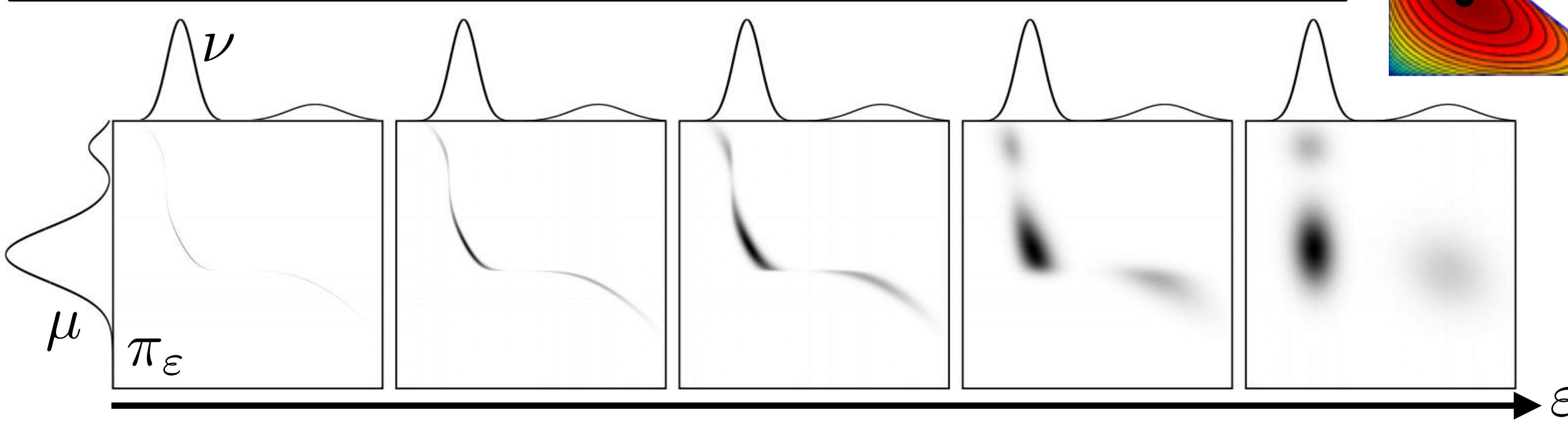
Schrödinger's problem: $\pi_\varepsilon = \underset{\pi \in \Pi(\mu,\nu)}{\operatorname{argmin}} \; \mathrm{KL}(\pi|K)$

Gibbs kernel: $K(x,y) \overset{\text{def.}}{=} e^{-\frac{c(x,y)}{\varepsilon}} \pi_0(x,y)$

Landmark computational paper: [Cuturi 2013].

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

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



Dykstra-like Iterations

$$\textit{Primal:}\ \min_{\pi}\ \langle c,\, \pi\rangle + f_1(P_{1\sharp}\pi) + f_2(P_{2\sharp}\pi) + \varepsilon \mathrm{KL}(\pi|\pi_0)$$

Dykstra-like Iterations

Primal: $\min_{\pi} \langle c,\, \pi\rangle + f_1(P_{1\sharp}\pi) + f_2(P_{2\sharp}\pi) + \varepsilon \mathrm{KL}(\pi|\pi_0)$

Dual: $\max_{u,v} \; - f_1^*(u) - f_2^*(u) - \varepsilon\langle e^{\frac{u}{\varepsilon}},\, K e^{-\frac{v}{\varepsilon}}\rangle$

Dykstra-like Iterations

Primal: $\min_{\pi} \langle c, \pi \rangle + f_1(P_{1\sharp}\pi) + f_2(P_{2\sharp}\pi) + \varepsilon \mathrm{KL}(\pi|\pi_0)$

Dual: $\max_{u,v} \; - f_1^*(u) - f_2^*(u) - \varepsilon \langle e^{\frac{u}{\varepsilon}}, \, K e^{-\frac{v}{\varepsilon}} \rangle$

$$\pi(x,y) = a(x)K(x,y)b(y) \qquad (a,b) = (e^{-\frac{u}{\varepsilon}}, e^{-\frac{v}{\varepsilon}})$$

Dykstra-like Iterations

Primal: $\min_{\pi} \langle c, \pi \rangle + f_1(P_{1\sharp}\pi) + f_2(P_{2\sharp}\pi) + \varepsilon \mathrm{KL}(\pi|\pi_0)$

Dual: $\max_{u,v} \; - f_1^*(u) - f_2^*(u) - \varepsilon \langle e^{\frac{u}{\varepsilon}}, K e^{-\frac{v}{\varepsilon}} \rangle$

$$\pi(x,y) = a(x)K(x,y)b(y) \qquad (a,b) = (e^{-\frac{u}{\varepsilon}}, e^{-\frac{v}{\varepsilon}})$$

Block coordinates relaxation: $\max_{u} - f_1^*(u) - \varepsilon \langle e^{\frac{u}{\varepsilon}}, K e^{-\frac{v}{\varepsilon}} \rangle \qquad (\mathcal{I}_u)$

Dykstra-like Iterations

Primal: $\min_{\pi} \langle c, \pi \rangle + f_1(P_{1\sharp}\pi) + f_2(P_{2\sharp}\pi) + \varepsilon \mathrm{KL}(\pi|\pi_0)$

Dual: $\max_{u,v} \; -f_1^*(u) - f_2^*(u) - \varepsilon \langle e^{\frac{u}{\varepsilon}}, K e^{-\frac{v}{\varepsilon}} \rangle$

$$\pi(x,y) = a(x)K(x,y)b(y) \qquad (a,b) = (e^{-\frac{u}{\varepsilon}}, e^{-\frac{v}{\varepsilon}})$$

Block coordinates relaxation:

$$\max_{u} - f_1^*(u) - \varepsilon \langle e^{\frac{u}{\varepsilon}}, K e^{-\frac{v}{\varepsilon}} \rangle \qquad (\mathcal{I}_u)$$

$$\max_{v} - f_2^*(v) - \varepsilon \langle e^{\frac{v}{\varepsilon}}, K^* e^{-\frac{u}{\varepsilon}} \rangle \qquad (\mathcal{I}_v)$$

Dykstra-like Iterations

Primal: $\min_{\pi} \langle c,\, \pi\rangle + f_1(P_{1\sharp}\pi) + f_2(P_{2\sharp}\pi) + \varepsilon\mathrm{KL}(\pi|\pi_0)$

Dual: $\max_{u,v} \; - f_1^*(u) - f_2^*(u) - \varepsilon\langle e^{\frac{u}{\varepsilon}},\, K e^{-\frac{v}{\varepsilon}}\rangle$

$$\pi(x,y) = a(x)K(x,y)b(y) \qquad (a,b) = (e^{-\frac{u}{\varepsilon}}, e^{-\frac{v}{\varepsilon}})$$

Block coordinates relaxation:

$$\max_{u} - f_1^*(u) - \varepsilon\langle e^{\frac{u}{\varepsilon}},\, K e^{-\frac{v}{\varepsilon}}\rangle \qquad (\mathcal{I}_u)$$

$$\max_{v} - f_2^*(v) - \varepsilon\langle e^{\frac{v}{\varepsilon}},\, K^* e^{-\frac{u}{\varepsilon}}\rangle \qquad (\mathcal{I}_v)$$

Proposition: the solutions of $(\mathcal{I}_u)$ and $(\mathcal{I}_v)$ read:

$$a = \frac{\mathrm{Prox}^{\mathrm{KL}}_{f_1/\varepsilon}(Kb)}{Kb} \qquad b = \frac{\mathrm{Prox}^{\mathrm{KL}}_{f_2/\varepsilon}(K^*a)}{K^*a}$$

$$\mathrm{Prox}^{\mathrm{KL}}_{f_1/\varepsilon}(\mu) \overset{\text{def.}}{=} \mathrm{argmin}_{\,\nu} f_1(\nu) + \varepsilon\mathrm{KL}(\nu|\mu)$$

Dykstra-like Iterations

Primal: $\min_{\pi} \langle c, \pi \rangle + f_1(P_{1\sharp}\pi) + f_2(P_{2\sharp}\pi) + \varepsilon \mathrm{KL}(\pi|\pi_0)$

Dual: $\max_{u,v} \; - f_1^*(u) - f_2^*(u) - \varepsilon \langle e^{\frac{u}{\varepsilon}}, K e^{-\frac{v}{\varepsilon}} \rangle$

$$\pi(x,y) = a(x) K(x,y) b(y) \qquad (a,b) = (e^{-\frac{u}{\varepsilon}}, e^{-\frac{v}{\varepsilon}})$$

Block coordinates relaxation:

$$\max_{u} - f_1^*(u) - \varepsilon \langle e^{\frac{u}{\varepsilon}}, K e^{-\frac{v}{\varepsilon}} \rangle \qquad (\mathcal{I}_u)$$

$$\max_{v} - f_2^*(v) - \varepsilon \langle e^{\frac{v}{\varepsilon}}, K^* e^{-\frac{u}{\varepsilon}} \rangle \qquad (\mathcal{I}_v)$$

Proposition: the solutions of $(\mathcal{I}_u)$ and $(\mathcal{I}_v)$ read:

$$a = \frac{\mathrm{Prox}^{\mathrm{KL}}_{f_1/\varepsilon}(Kb)}{Kb} \qquad b = \frac{\mathrm{Prox}^{\mathrm{KL}}_{f_2/\varepsilon}(K^*a)}{K^*a}$$

$$\mathrm{Prox}^{\mathrm{KL}}_{f_1/\varepsilon}(\mu) \stackrel{\text{def.}}{=} \mathrm{argmin}_{\nu} f_1(\nu) + \varepsilon \mathrm{KL}(\nu|\mu)$$

→ Only matrix-vector multiplications. → Highly parallelizable.

→ On regular grids: only convolutions! Linear time iterations.

Sinkhorn's Algorithm

Optimal transport problem:

$$f_1 = \iota_\mu \longrightarrow \mathrm{Prox}^{\mathrm{KL}}_{f_1/\varepsilon}(\tilde{\mu}) = \mu$$

$$f_2 = \iota_\nu \longrightarrow \mathrm{Prox}^{\mathrm{KL}}_{f_2/\varepsilon}(\tilde{\nu}) = \nu$$

Sinkhorn's Algorithm

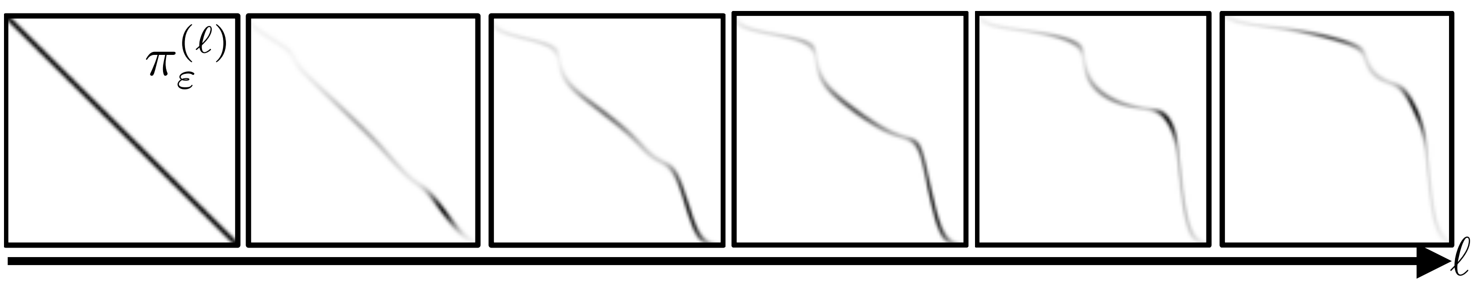
Optimal transport problem:

$$f_1 = \iota_\mu \longrightarrow \mathrm{Prox}^{\mathrm{KL}}_{f_1/\varepsilon}(\tilde{\mu}) = \mu$$

$$f_2 = \iota_\nu \longrightarrow \mathrm{Prox}^{\mathrm{KL}}_{f_2/\varepsilon}(\tilde{\nu}) = \nu$$

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

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



Sinkhorn's Algorithm

Optimal transport problem:

$$f_1 = \iota_\mu \longrightarrow \mathrm{Prox}^{\mathrm{KL}}_{f_1/\varepsilon}(\tilde{\mu}) = \mu$$

$$f_2 = \iota_\nu \longrightarrow \mathrm{Prox}^{\mathrm{KL}}_{f_2/\varepsilon}(\tilde{\nu}) = \nu$$

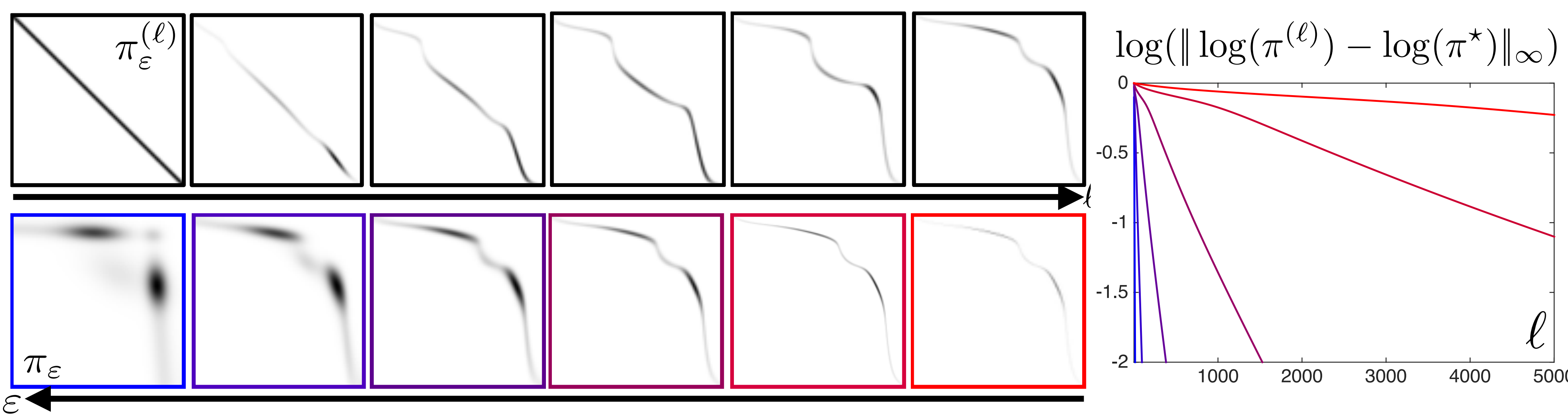
Sinkhorn/IPFP algorithm: [Sinkhorn 1967][Deming,Stephan 1940]

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

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

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

[Franklin,Lorenz 1989]
Local rate: [Knight 2008]



Gradient Flows: Crowd Motion

$$\mu_{t+1} \overset{\text{def.}}{=} \operatorname{argmin}_{\mu} W_\alpha^\alpha(\mu_t, \mu) + \tau f(\mu)$$

Congestion-inducing function:

$$f(\mu) = \iota_{[0,\kappa]}(\mu) + \langle w,\, \mu \rangle$$

[Maury, Roudneff-Chupin, Santambrogio 2010]

Gradient Flows: Crowd Motion

$$\mu_{t+1} \overset{\text{def.}}{=} \operatorname{argmin}_\mu W_\alpha^\alpha(\mu_t, \mu) + \tau f(\mu)$$

Congestion-inducing function:

$$f(\mu) = \iota_{[0,\kappa]}(\mu) + \langle w,\, \mu \rangle$$

[Maury, Roudneff-Chupin, Santambrogio 2010]

Proposition: $\operatorname{Prox}_{\frac{1}{\varepsilon} f}(\mu) = \min(e^{-\varepsilon w}\mu, \kappa)$

Gradient Flows: Crowd Motion

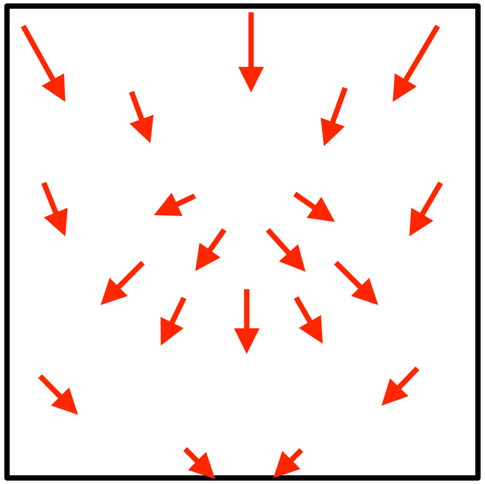
$$\mu_{t+1} \overset{\text{def.}}{=} \operatorname{argmin}_\mu W_\alpha^\alpha(\mu_t, \mu) + \tau f(\mu)$$

Congestion-inducing function:

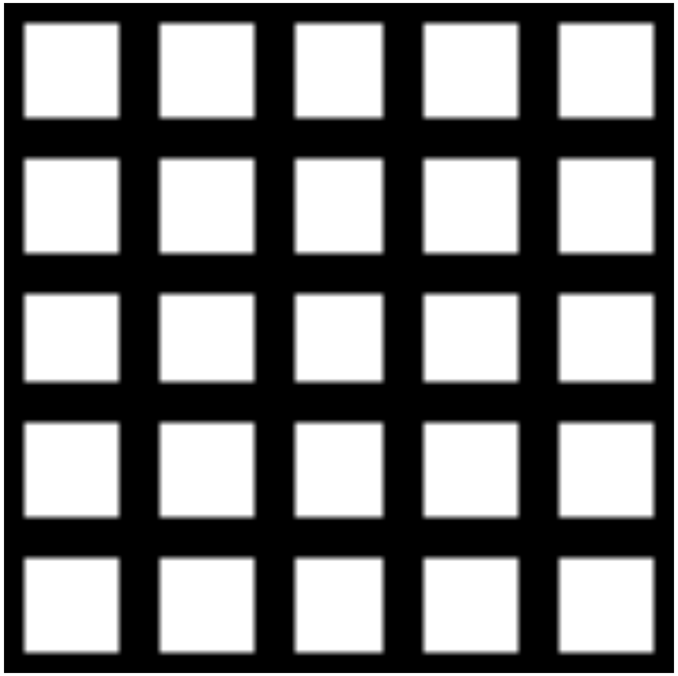
$$f(\mu) = \iota_{[0,\kappa]}(\mu) + \langle w, \mu \rangle$$

[Maury, Roudneff-Chupin, Santambrogio 2010]

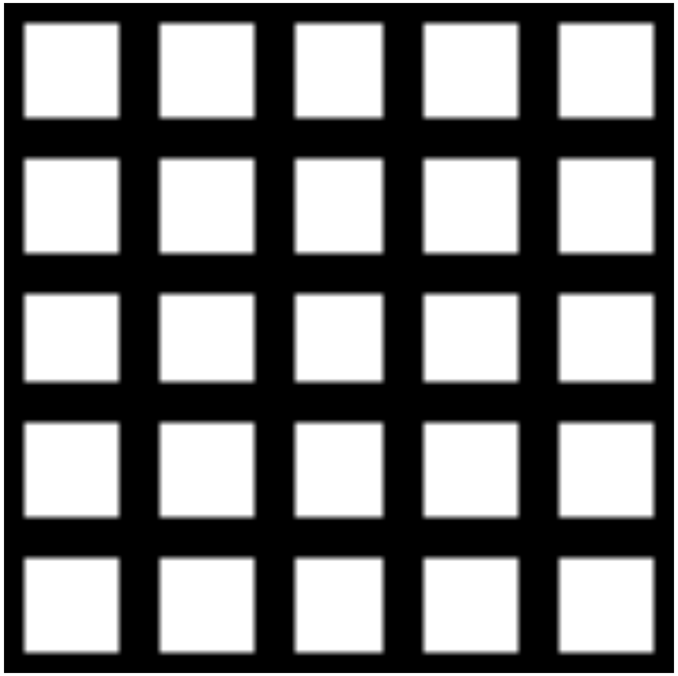
Proposition: $\operatorname{Prox}_{\frac{1}{\varepsilon} f}(\mu) = \min(e^{-\varepsilon w}\mu, \kappa)$



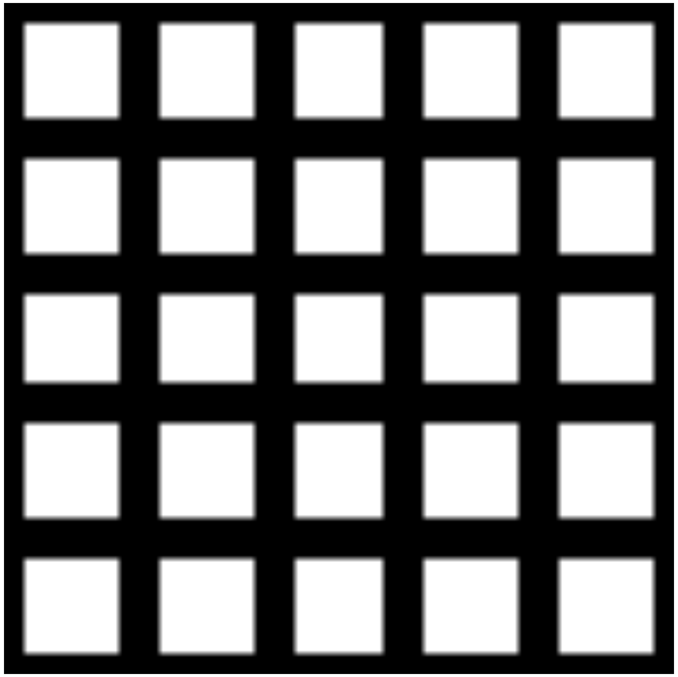
∇w



$\kappa = \|\mu_{t=0}\|_\infty$



$\kappa = 2\|\mu_{t=0}\|_\infty$



$\kappa = 4\|\mu_{t=0}\|_\infty$

Multiple-Density Gradient Flows

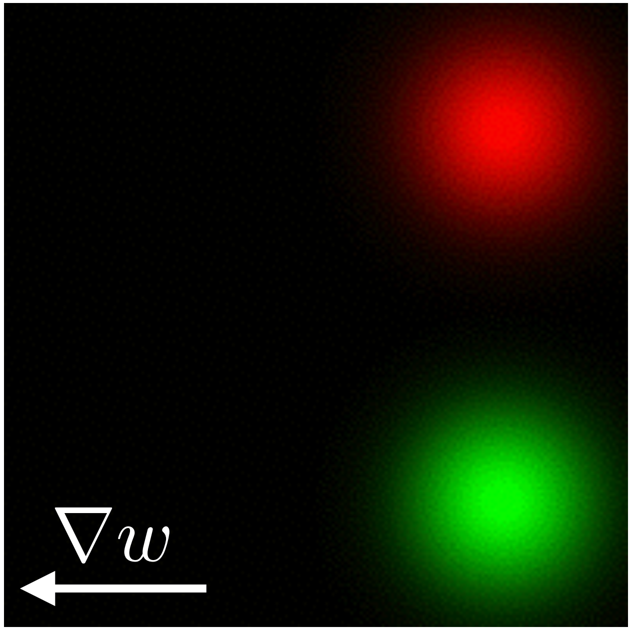
$$(\mu_{1,t+1}, \mu_{2,t+1}) \overset{\text{def.}}{=} \underset{(\mu_1,\mu_2)}{\operatorname{argmin}}\, W_\alpha^\alpha(\mu_{1,t}, \mu_1) + W_\alpha^\alpha(\mu_{2,t}, \mu_2) + \tau f(\mu_1, \mu_2)$$

Multiple-Density Gradient Flows

$$(\mu_{1,t+1}, \mu_{2,t+1}) \stackrel{\text{def.}}{=} \underset{(\mu_1,\mu_2)}{\operatorname{argmin}}\; W_\alpha^\alpha(\mu_{1,t}, \mu_1) + W_\alpha^\alpha(\mu_{2,t}, \mu_2) + \tau f(\mu_1, \mu_2)$$

Wasserstein attraction:

$$f(\mu_1, \mu_2) = W_\alpha^\alpha(\mu_1, \mu_2) + h_1(\mu_1) + h_2(\mu_2)$$



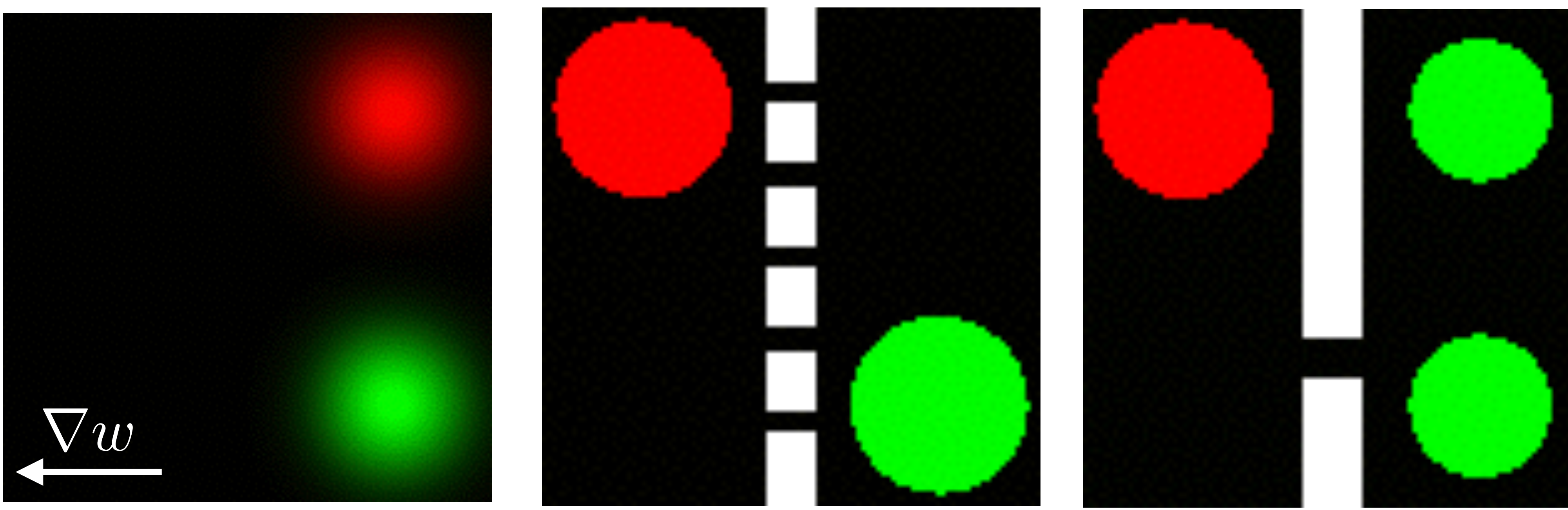
Example: $h_i(\mu) = \langle w,\, \mu \rangle.$

Multiple-Density Gradient Flows

$$(\mu_{1,t+1}, \mu_{2,t+1}) \stackrel{\text{def.}}{=} \underset{(\mu_1,\mu_2)}{\operatorname{argmin}}\; W_\alpha^\alpha(\mu_{1,t}, \mu_1) + W_\alpha^\alpha(\mu_{2,t}, \mu_2) + \tau f(\mu_1, \mu_2)$$

Wasserstein attraction:

$$f(\mu_1, \mu_2) = W_\alpha^\alpha(\mu_1, \mu_2) + h_1(\mu_1) + h_2(\mu_2)$$



Example: $h_i(\mu) = \langle w,\, \mu \rangle.$

Overview

- Transportation-like Problems
- Regularized Transport
- **Optimal Transport Barycenters**
- Heat Kernel Approximation

Wasserstein Barycenters

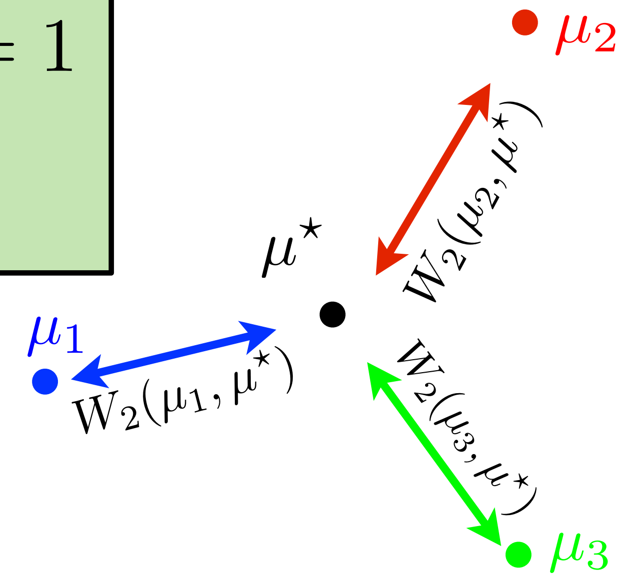
Barycenters of measures $(\mu_k)_k$: $\quad \sum_k \lambda_k = 1$

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

Wasserstein Barycenters

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

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



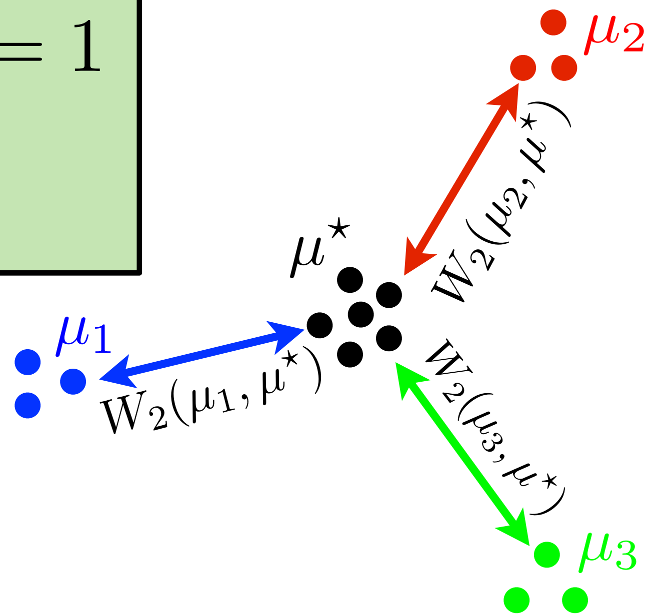
Generalizes Euclidean barycenter:

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

Wasserstein Barycenters

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

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



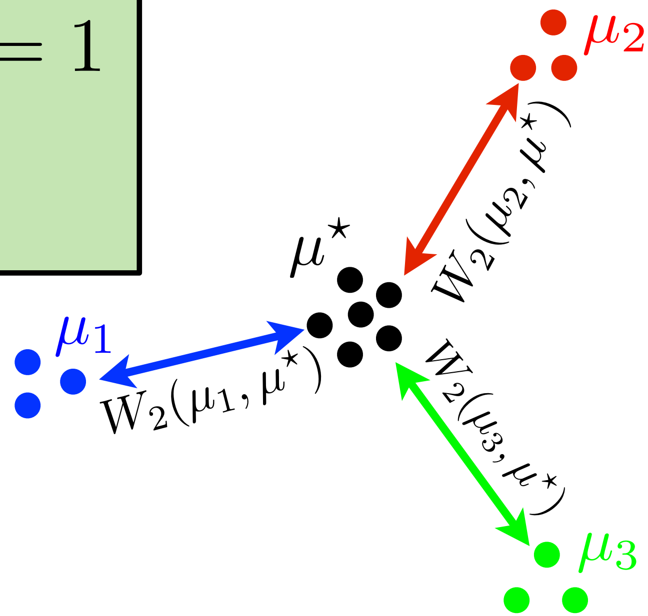
Generalizes Euclidean barycenter:

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

Wasserstein Barycenters

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

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



Generalizes Euclidean barycenter:

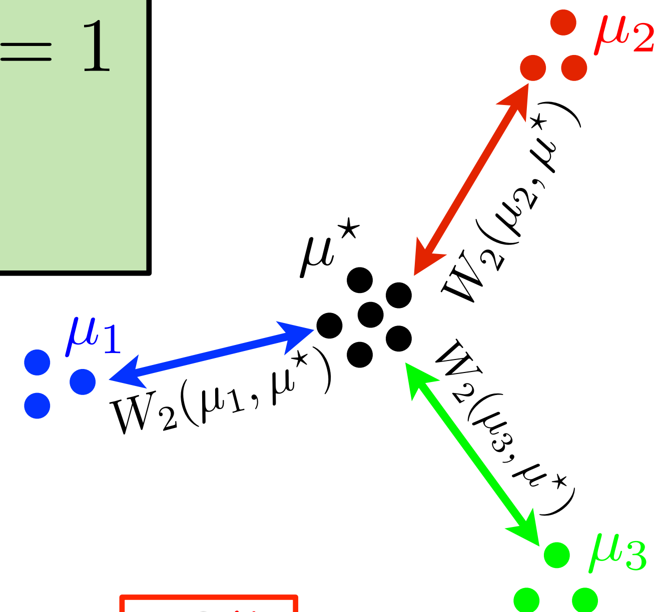
If $\mu_k = \delta_{x_k}$ then $\mu^\star = \delta_{\sum_k \lambda_k x_k}$

Mc Cann's displacement interpolation.

Wasserstein Barycenters

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

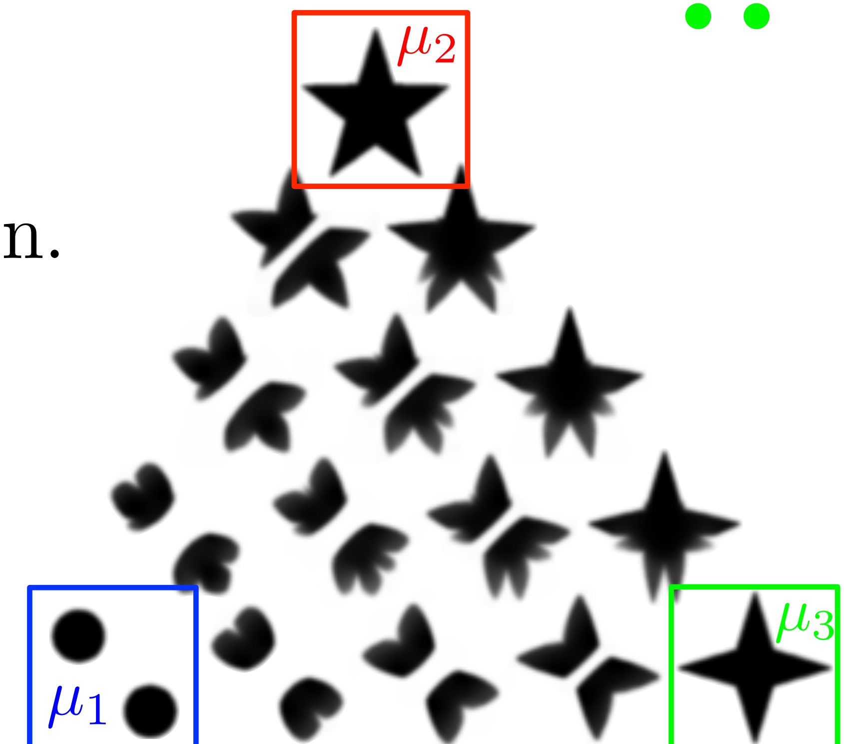
$$\mu^\star \in \underset{\mu}{\operatorname{argmin}} \sum_k \lambda_k W_\alpha^\alpha(\mu_k, \mu)$$



Generalizes Euclidean barycenter:

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

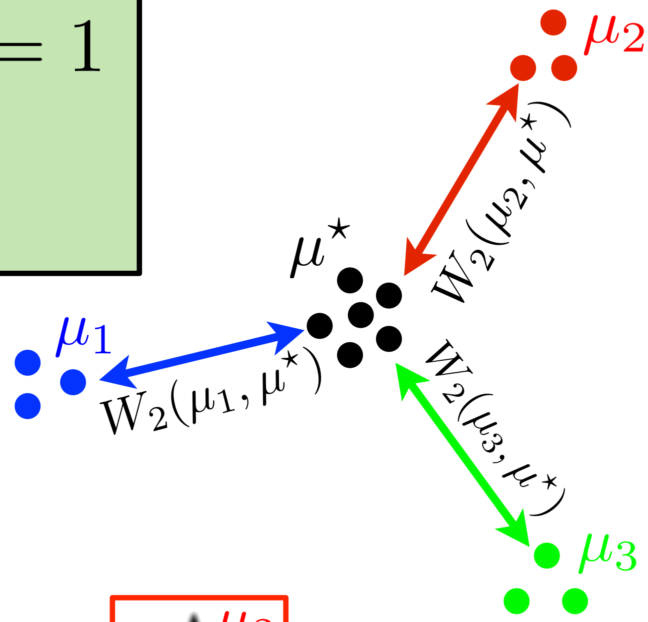
Mc Cann's displacement interpolation.



Wasserstein Barycenters

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

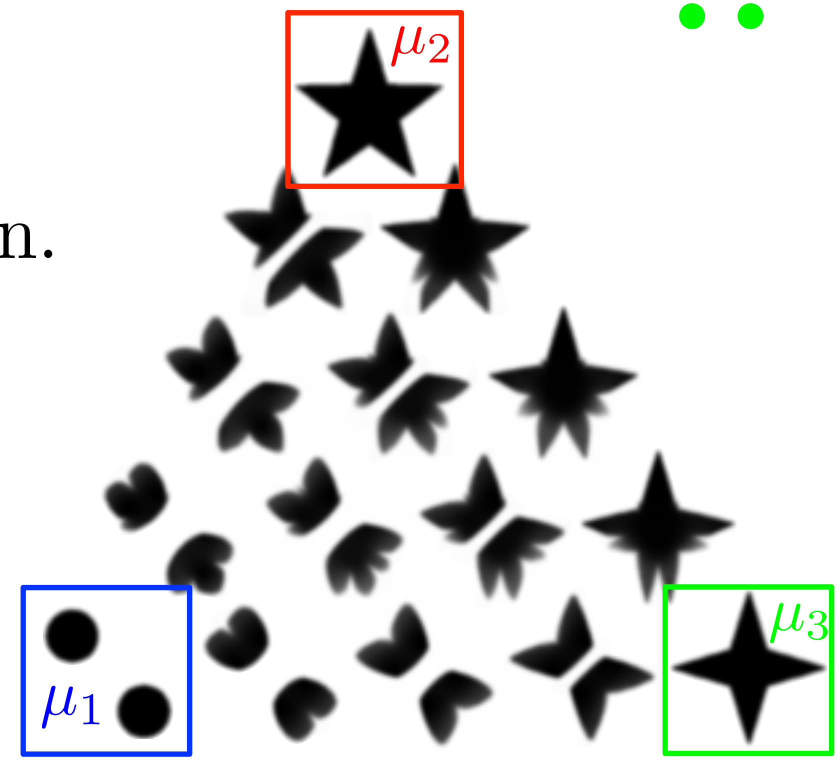
$$\mu^\star \in \underset{\mu}{\operatorname{argmin}} \sum_k \lambda_k W_\alpha^\alpha(\mu_k, \mu)$$



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.

Regularized Barycenters

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

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 $(\pi_{0,k})_k$

Regularized Barycenters

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

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

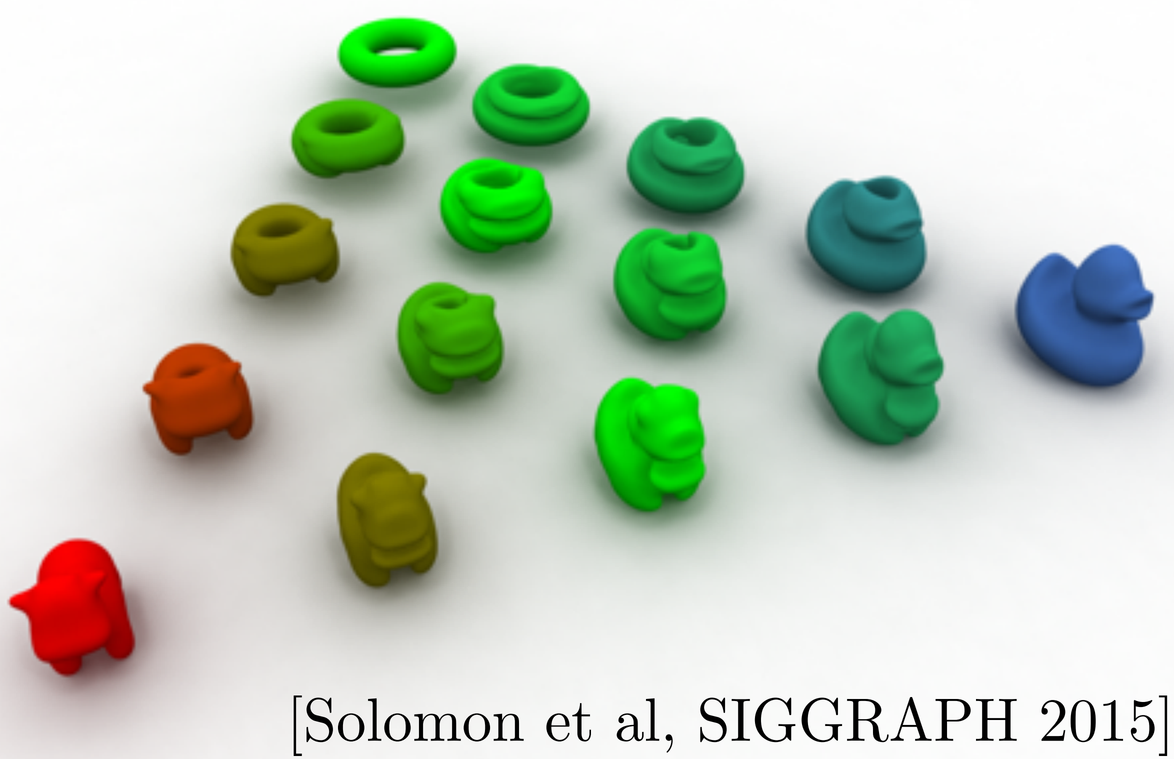
$\longrightarrow$ Sinkhorn-like algorithm [Benamou, Carlier, Cuturi, Nenna, Peyré, 2015].

Regularized Barycenters

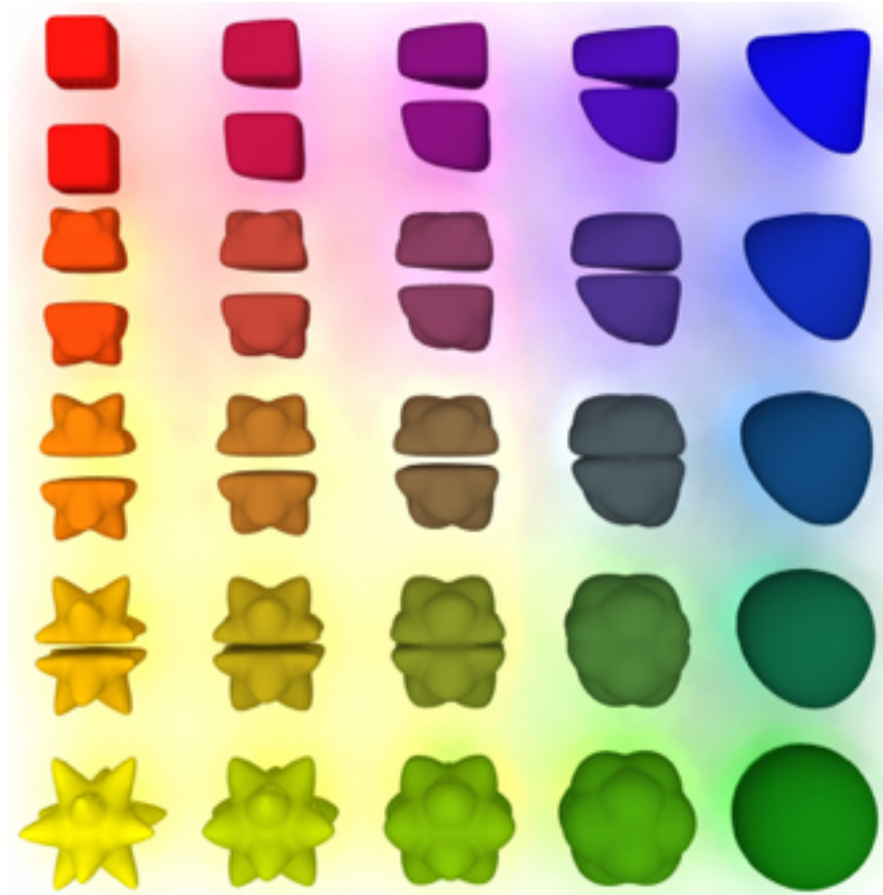
$$\min_{(\pi_k)_k,\mu} \left\{ \sum_k \lambda_k \left(\langle c,\, \pi_k \rangle + \varepsilon \mathrm{KL}(\pi_k | \pi_{0,k}) \right) \ ; \ \forall k, \pi_k \in \Pi(\mu_k, \mu) \right\}$$

$\longrightarrow$ 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]

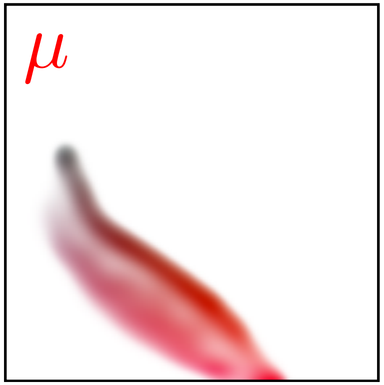


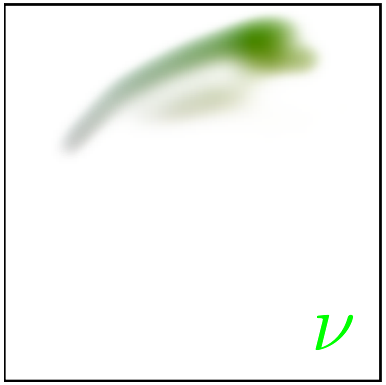
Color Transfer

Input images: (f, g) (chrominance components)

Input measures: $\mu(A) = \mathcal{U}(f^{-1}(A)), \nu(A) = \mathcal{U}(g^{-1}(A))$

f





g

Color Transfer

Input images: (f, g) (chrominance components)

Input measures: $\mu(A) = \mathcal{U}(f^{-1}(A)), \nu(A) = \mathcal{U}(g^{-1}(A))$

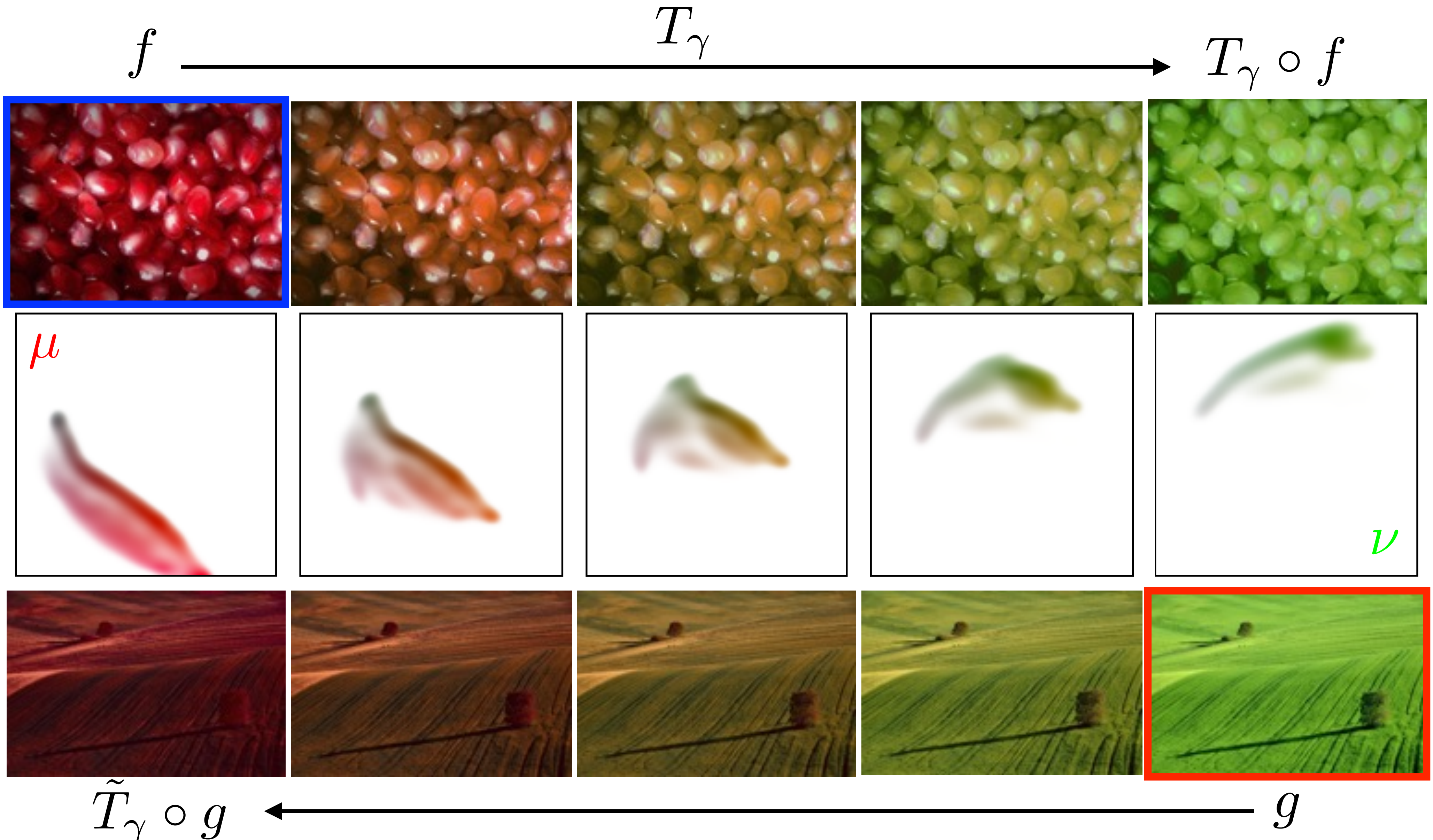
f

g

Color Transfer

Input images: (f, g) (chrominance components)

Input measures: $\mu(A) = \mathcal{U}(f^{-1}(A)), \nu(A) = \mathcal{U}(g^{-1}(A))$



Color Harmonization

Raw image sequence

↓

Color Harmonization

Raw image sequence

↓

Compute Wasserstein barycenter

↓

Project on the barycenter

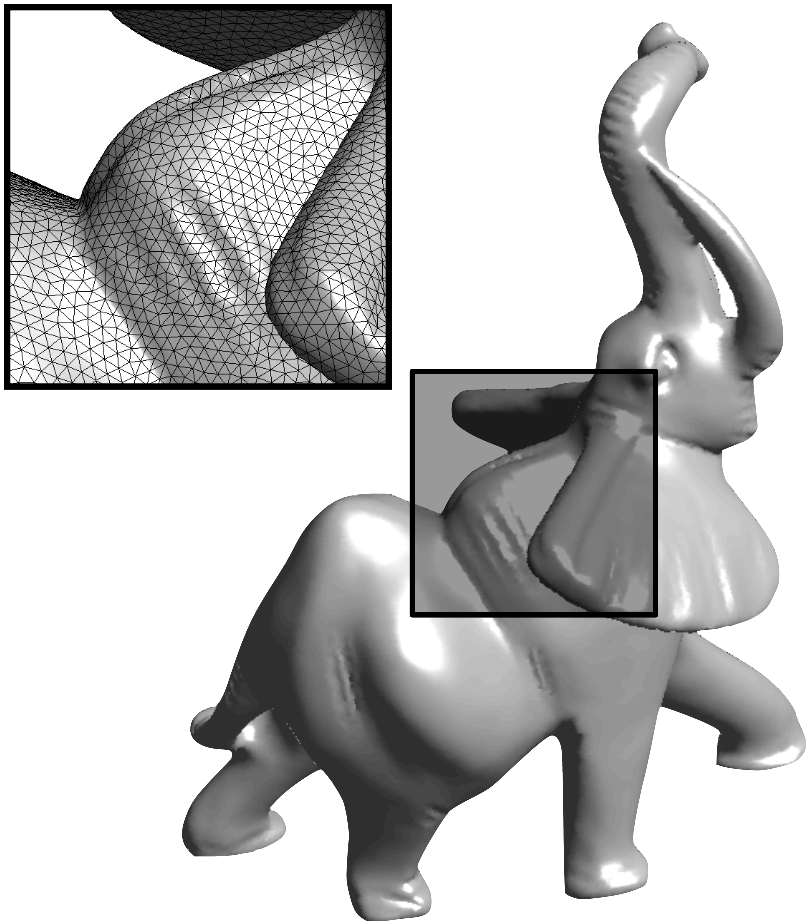
Overview

- Transportation-like Problems
- Regularized Transport
- Optimal Transport Barycenters
- **Heat Kernel Approximation**

Optimal Transport on Surfaces

Triangulated mesh M.

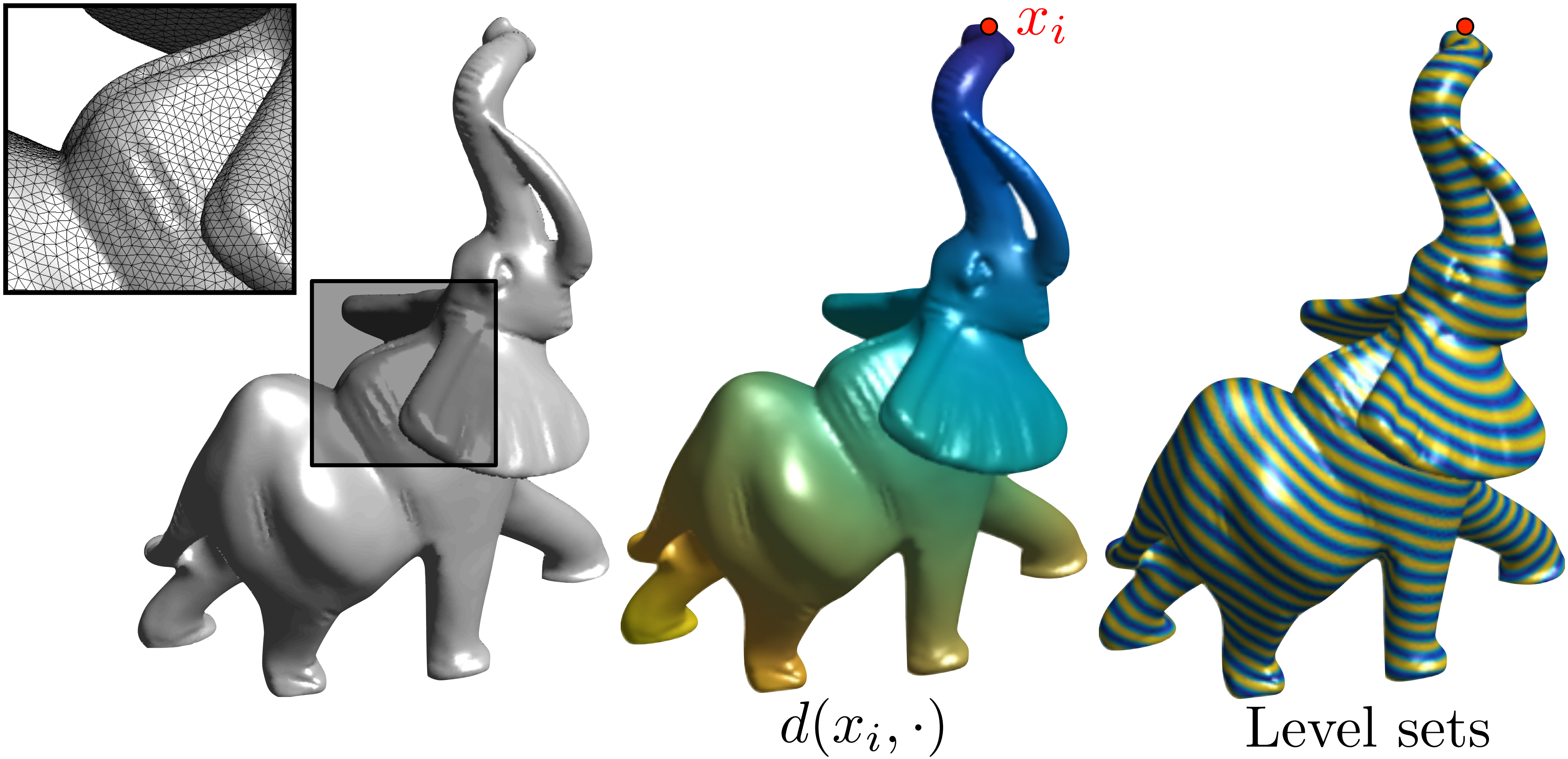
Geodesic distance d_M.



Optimal Transport on Surfaces

Triangulated mesh M. Geodesic distance d_M.

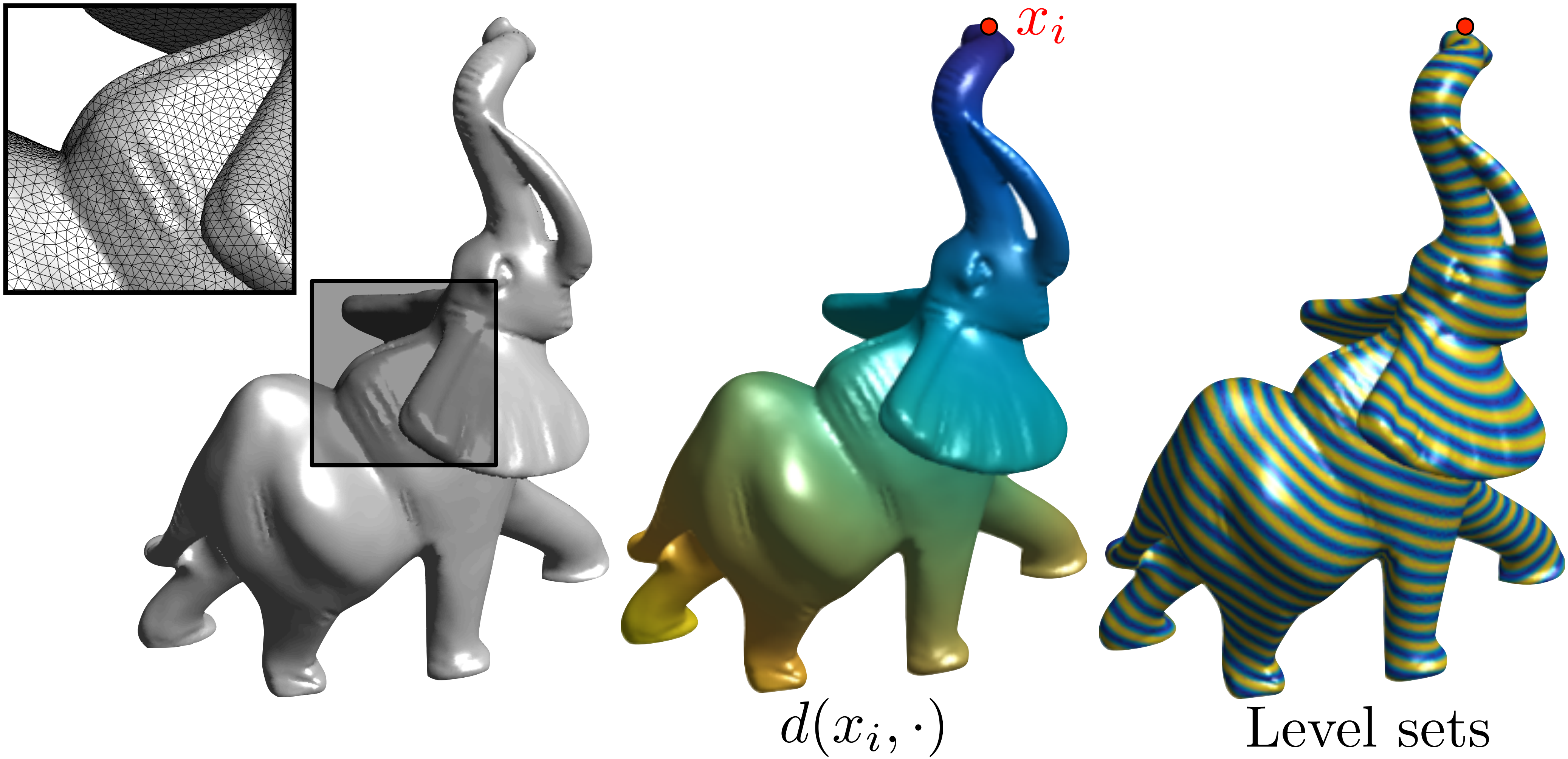
Ground cost: $c(x, y) = d_M(x, y)^\alpha$.



Optimal Transport on Surfaces

Triangulated mesh M. Geodesic distance d_M.

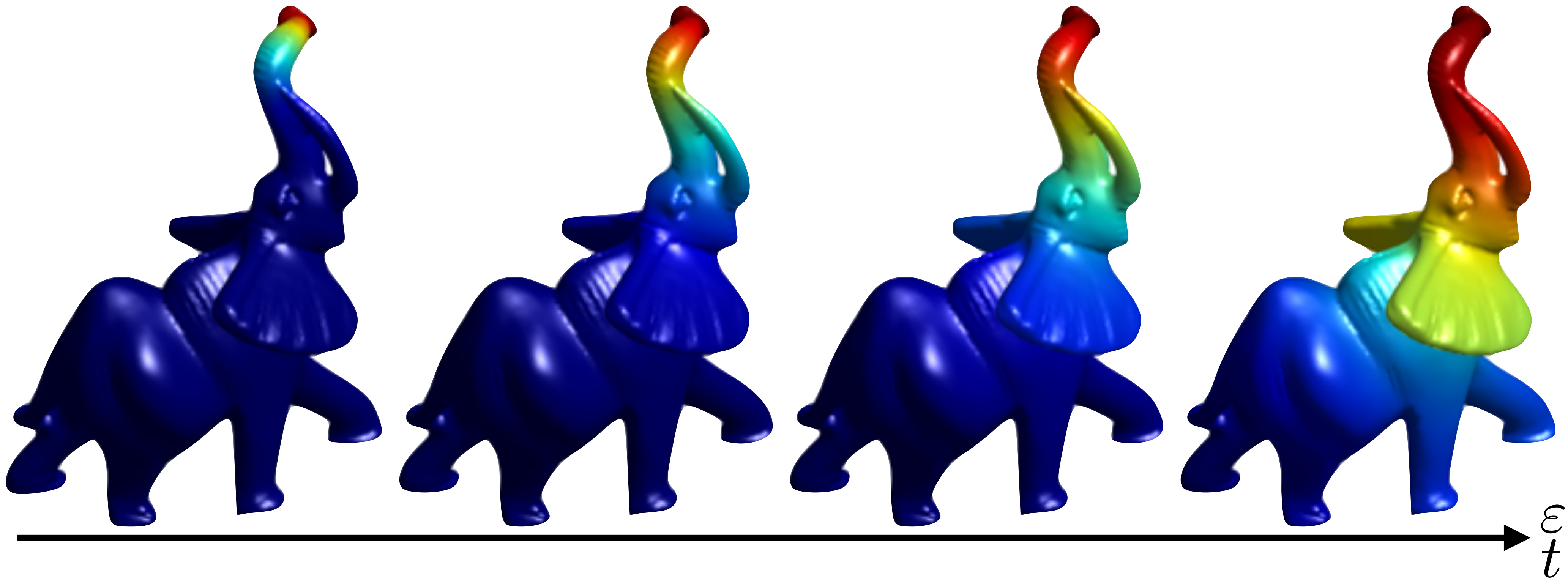
Ground cost: $c(x,y) = d_M(x,y)^\alpha$.



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

Entropic Transport on Surfaces

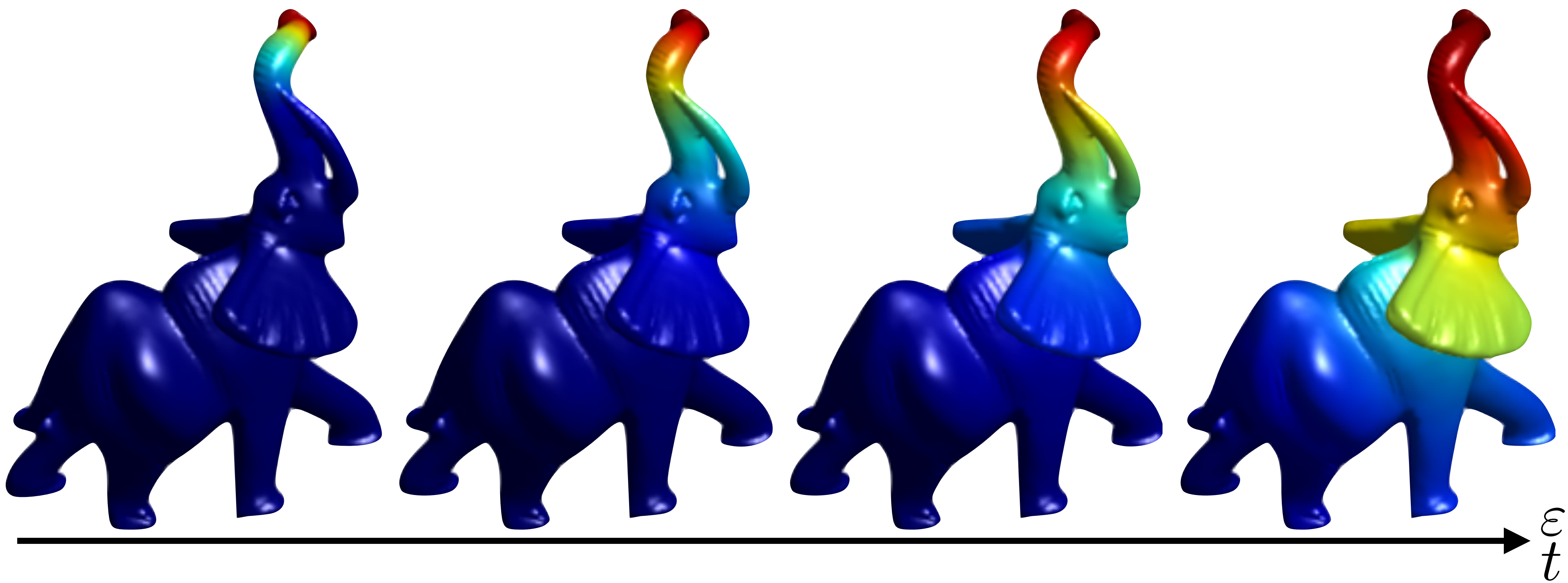
Heat equation on M: $\partial_t u_t(x,\cdot) = \Delta_M u_t(x,\cdot)$, $u_{t=0}(x,\cdot) = \delta_x$



Entropic Transport on Surfaces

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

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

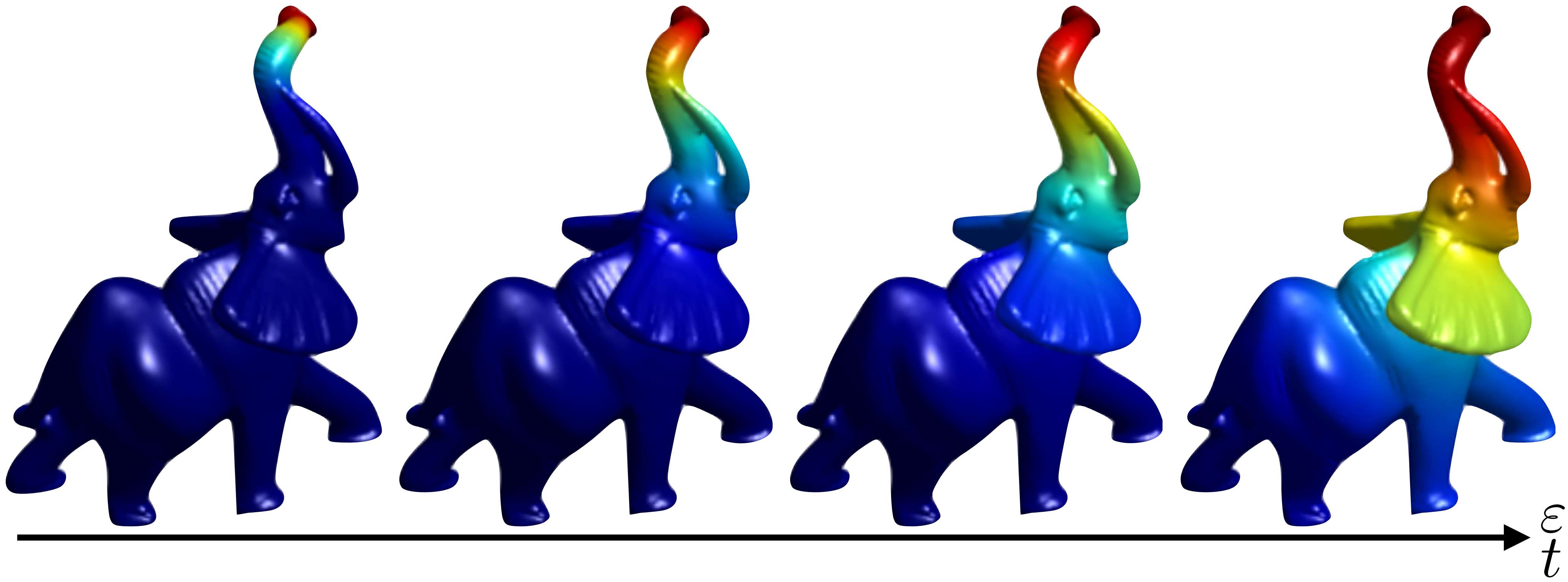


Entropic Transport on Surfaces

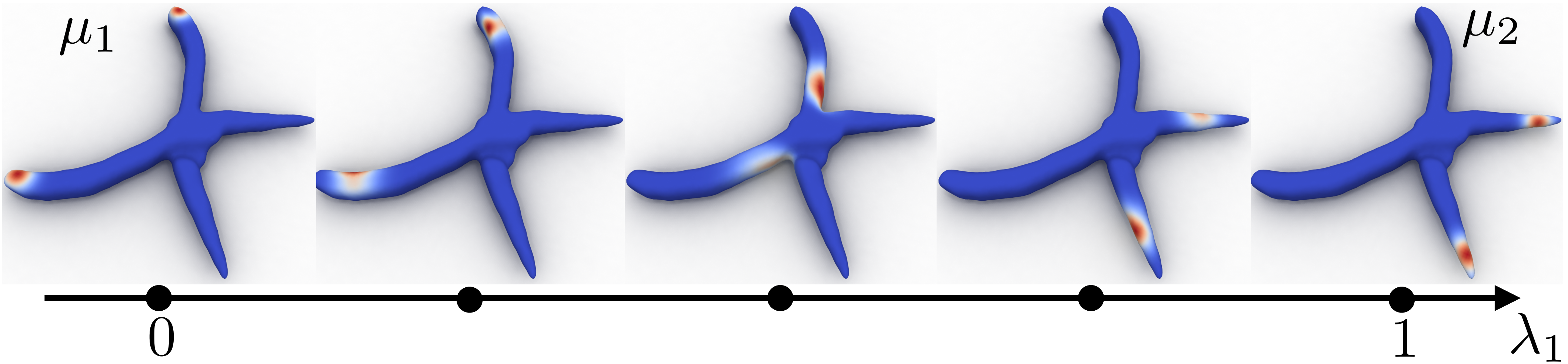
Heat equation on M: $\partial_t u_t(x,\cdot) = \Delta_M u_t(x,\cdot)$, $u_{t=0}(x,\cdot) = \delta_x$

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

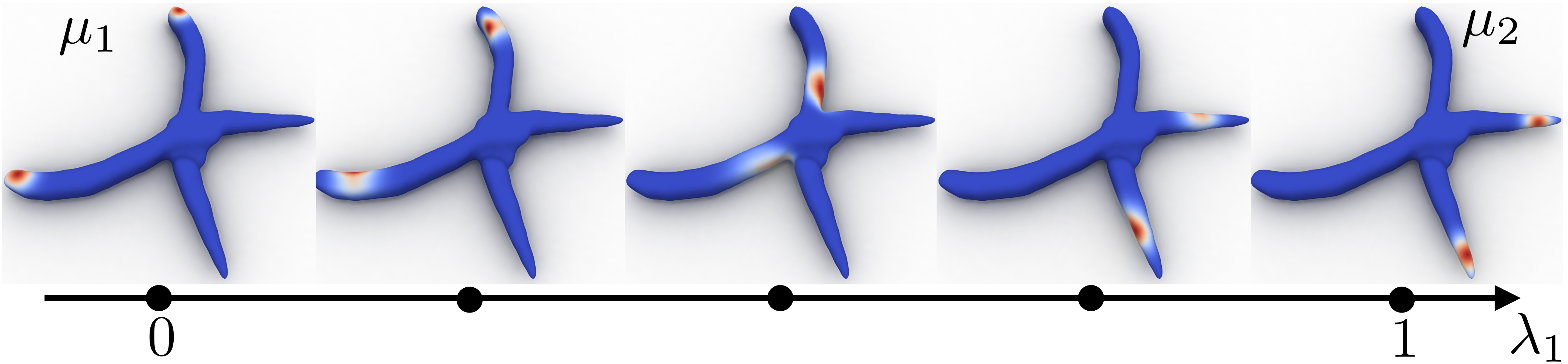
Sinkhorn kernel: $\quad K \overset{\text{def.}}{=} e^{-\frac{d_M^2}{\varepsilon}} \approx u_\varepsilon \approx \left(\mathrm{Id} - \frac{\varepsilon}{\ell}\Delta_M\right)^{-\ell}$

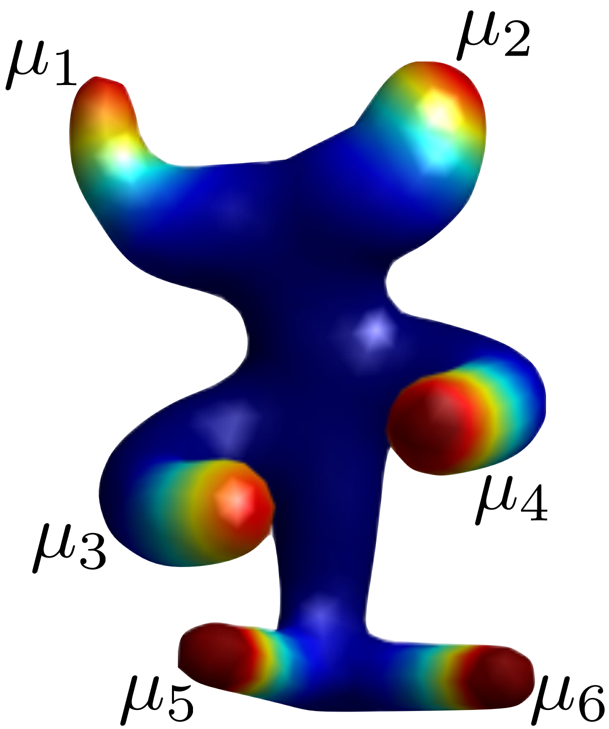


Barycenter on a Surface

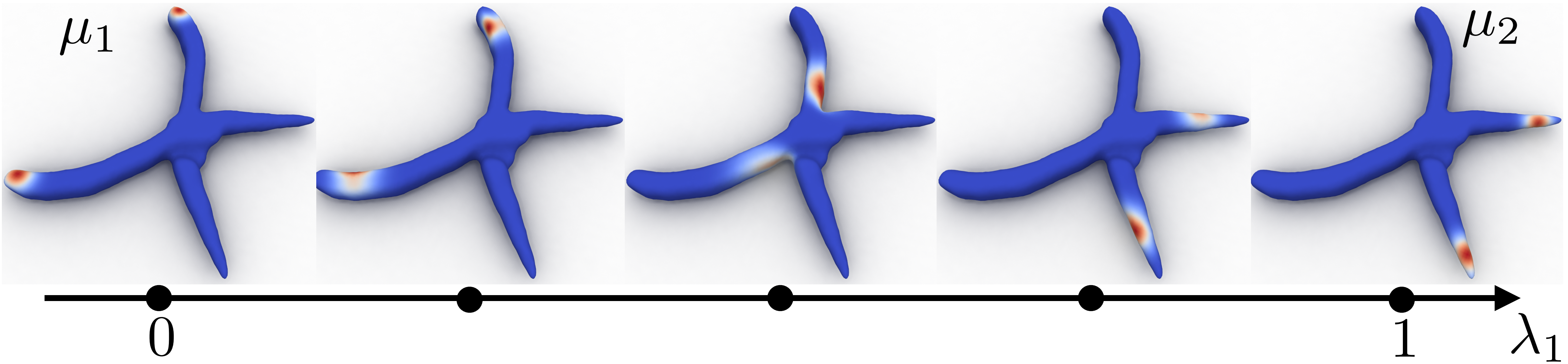


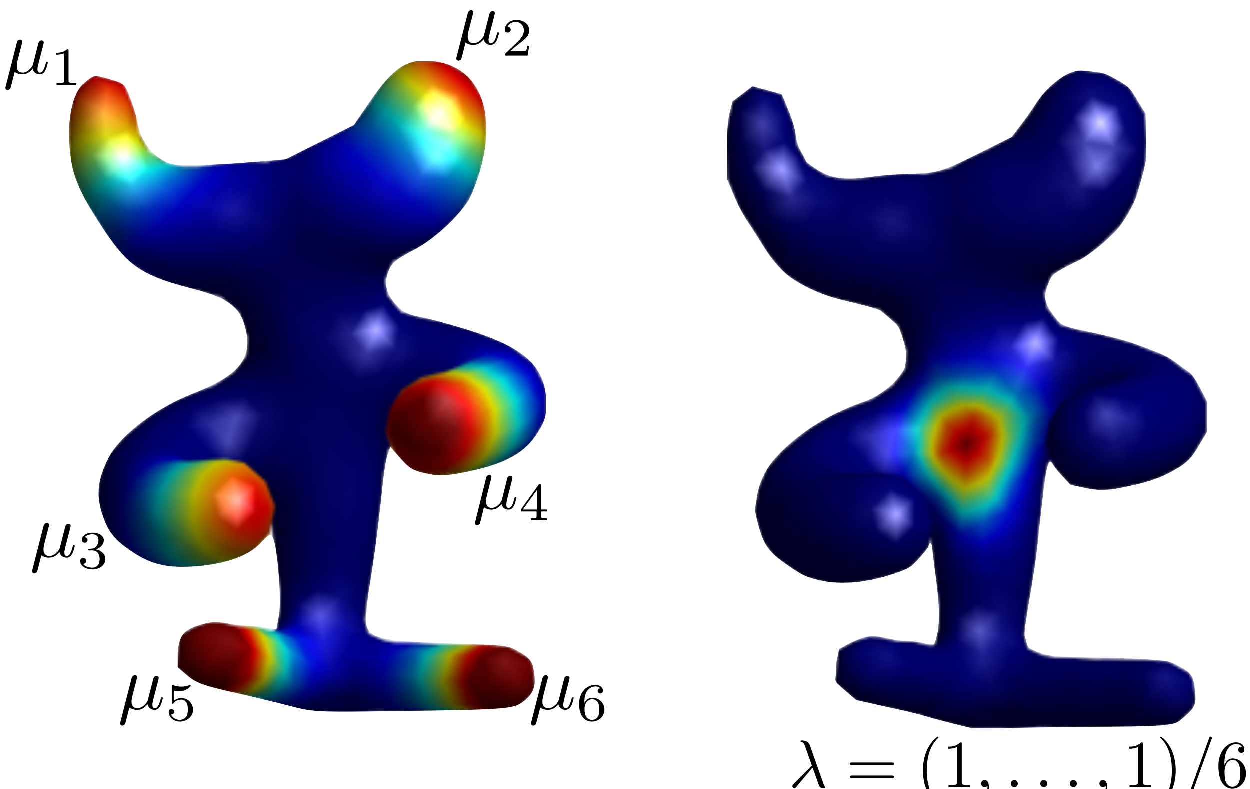
Barycenter on a Surface





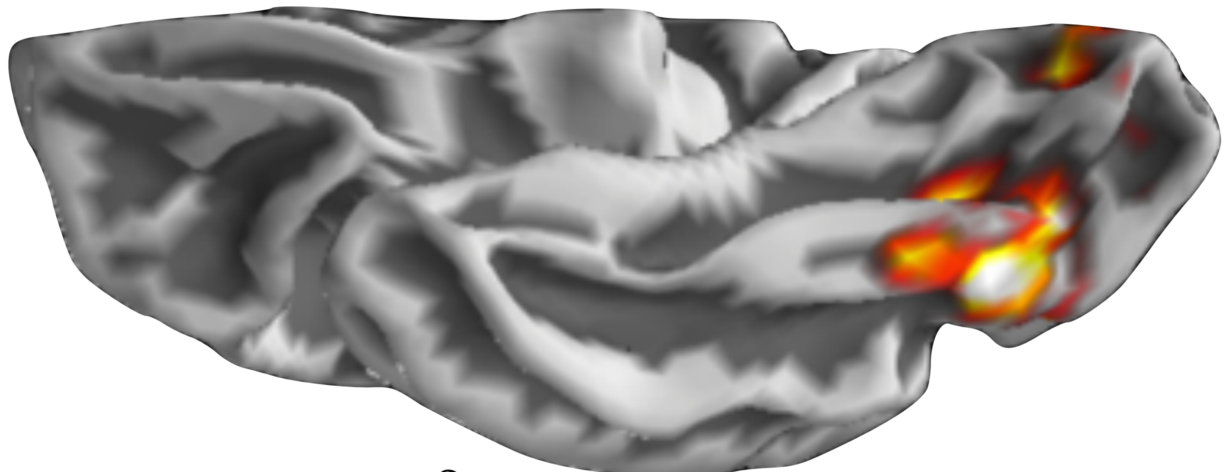
Barycenter on a Surface



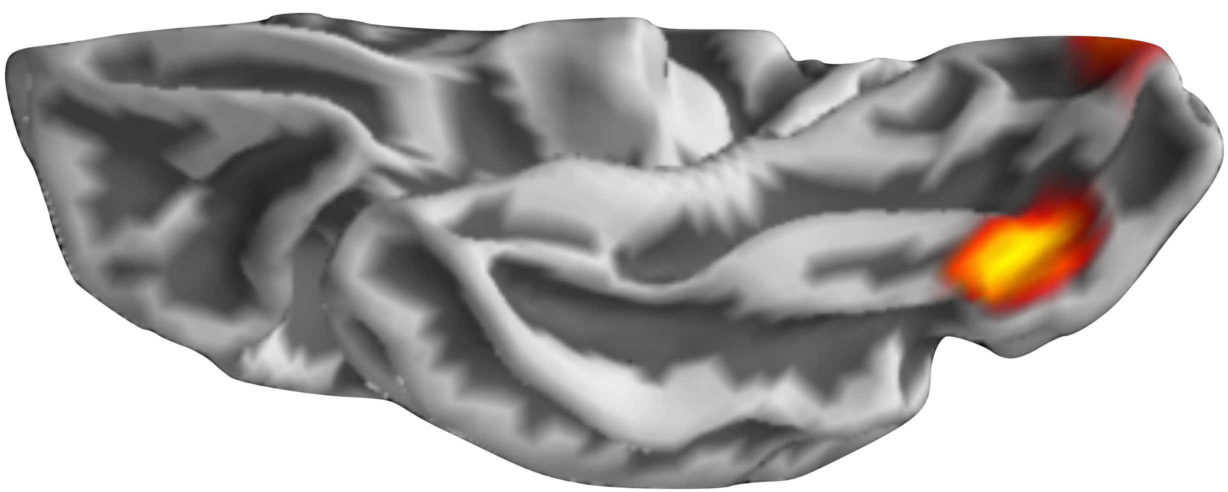


MRI Data Procesing [with A. Gramfort]

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



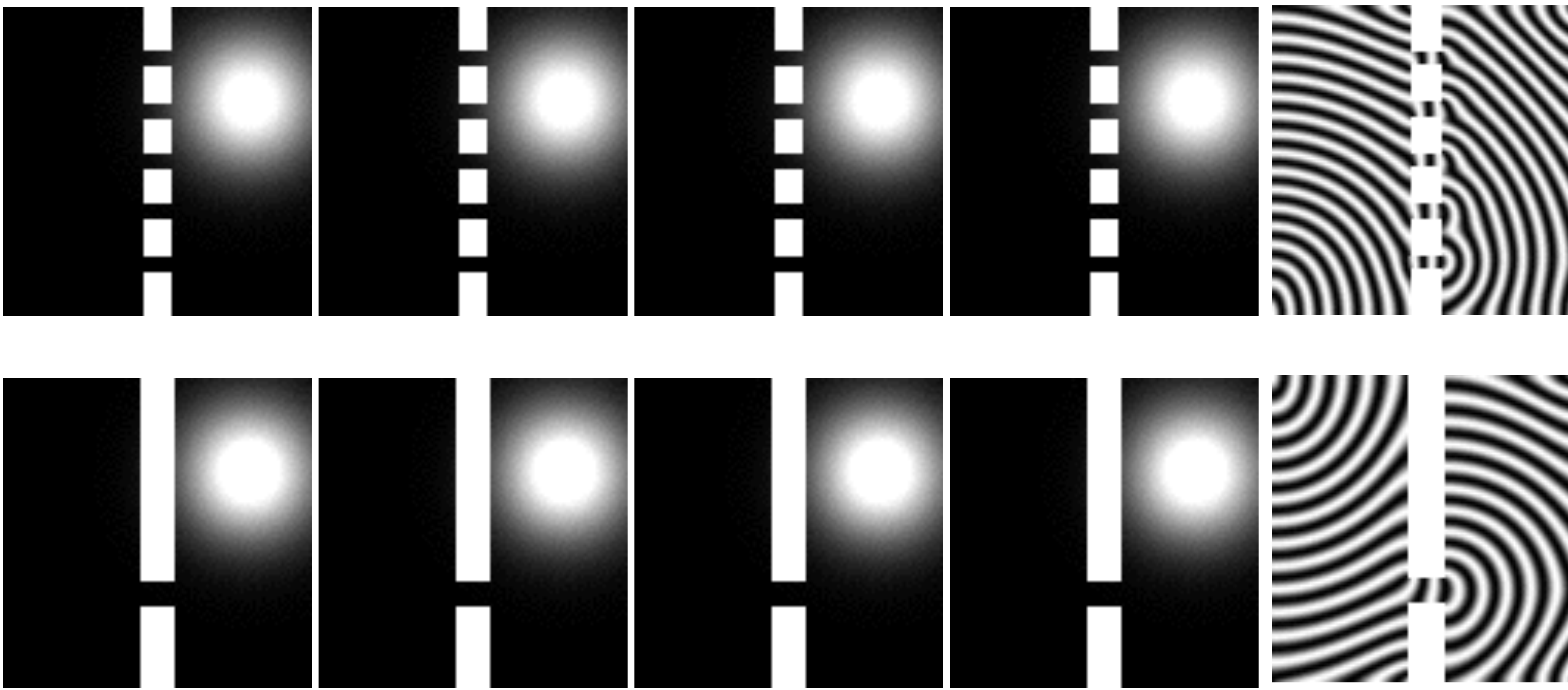
L^2 barycenter



W_2^2 barycenter

Gradient Flows: Crowd Motion with Obstacles

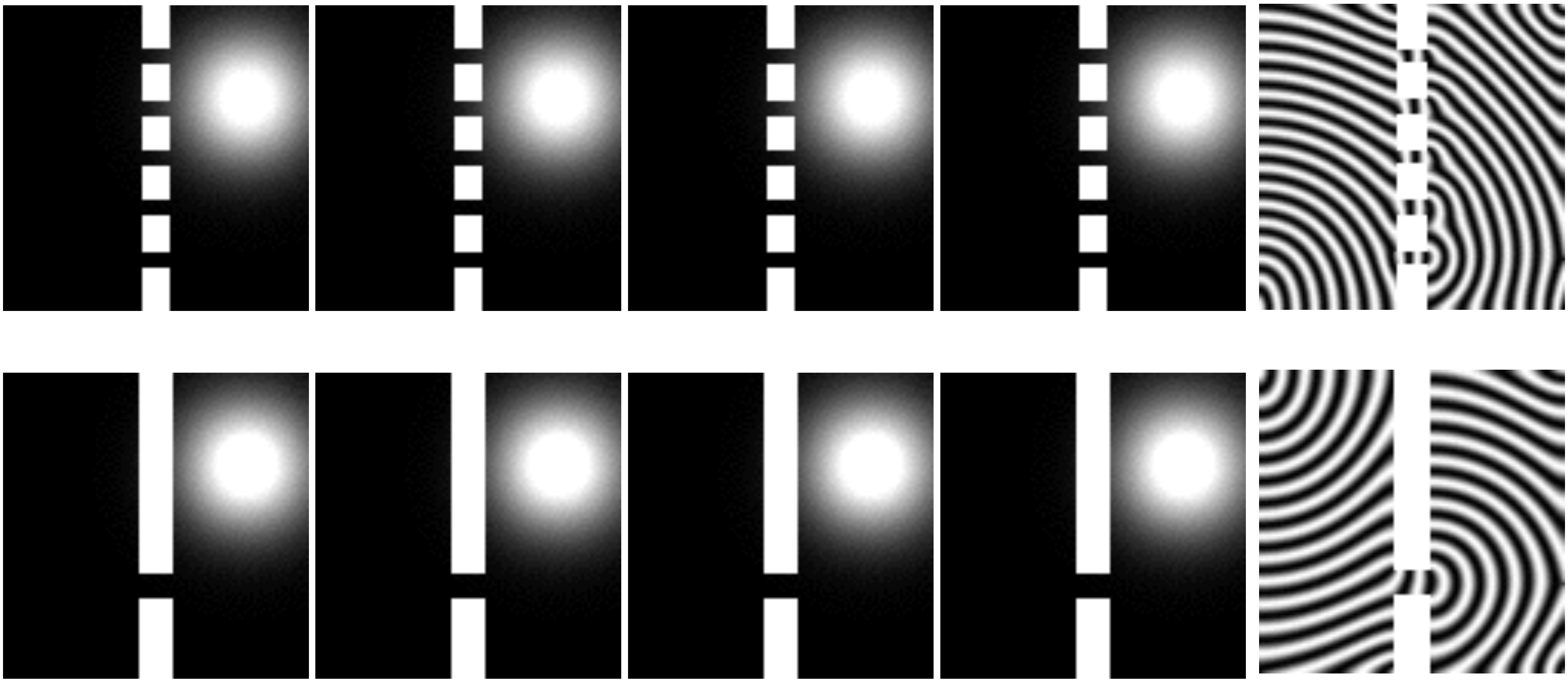
$$M = \text{sub-domain of } \mathbb{R}^2.$$



$\kappa = \|\mu_{t=0}\|_\infty$ $\kappa = 2\|\mu_{t=0}\|_\infty$ $\kappa = 4\|\mu_{t=0}\|_\infty$ $\kappa = 6\|\mu_{t=0}\|_\infty$ Potential $\cos(w)$

Gradient Flows: Crowd Motion with Obstacles

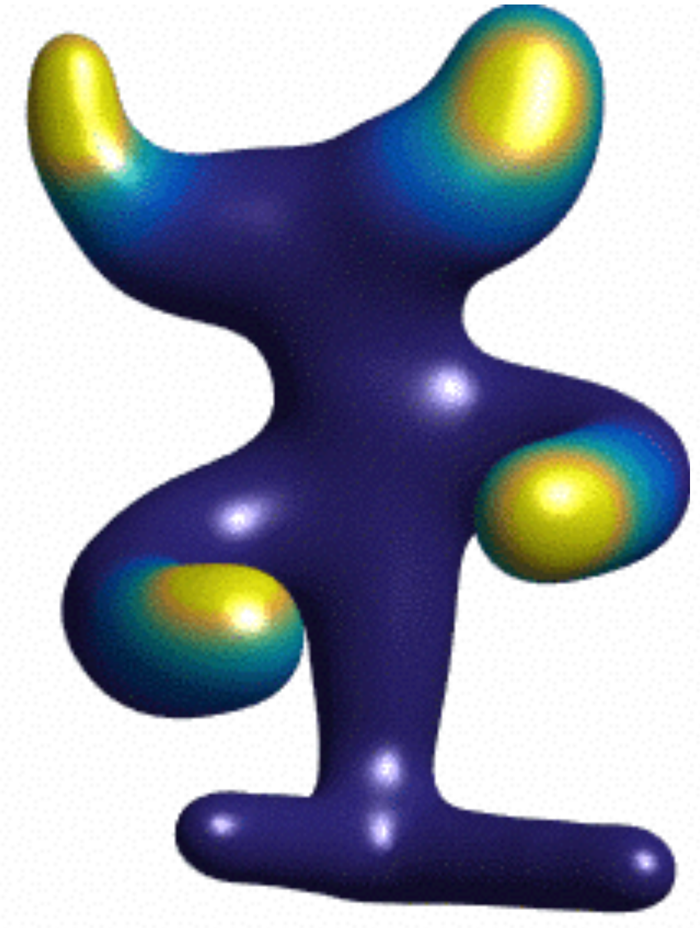
$$M = \text{sub-domain of } \mathbb{R}^2.$$



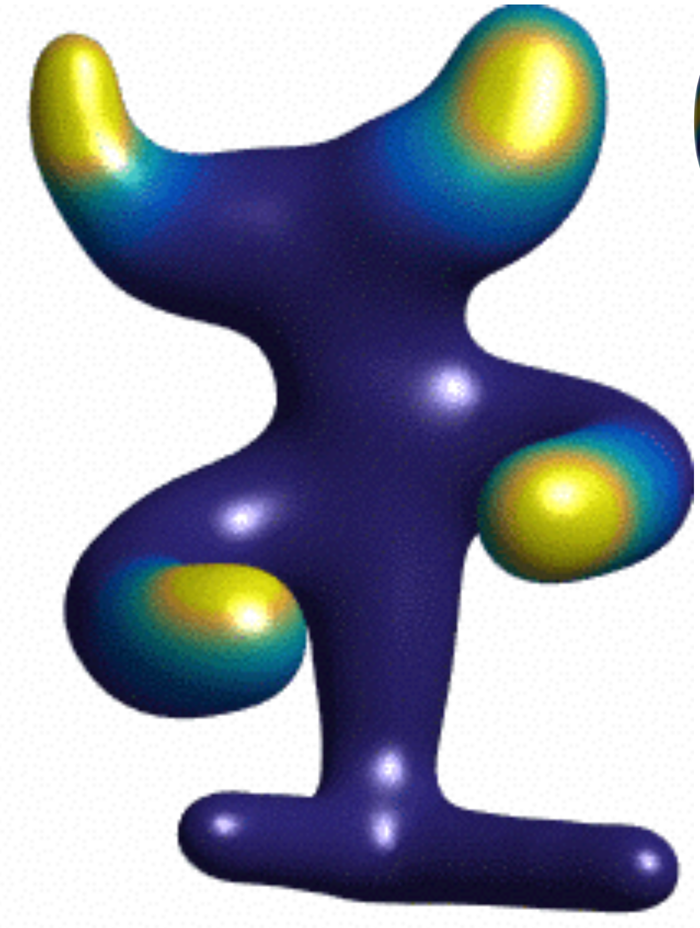
$\kappa = \|\mu_{t=0}\|_\infty$ $\kappa = 2\|\mu_{t=0}\|_\infty$ $\kappa = 4\|\mu_{t=0}\|_\infty$ $\kappa = 6\|\mu_{t=0}\|_\infty$ Potential $\cos(w)$

Crowd Motion on a Surface

M = triangulated mesh.



$\kappa = \|\mu_{t=0}\|_\infty$

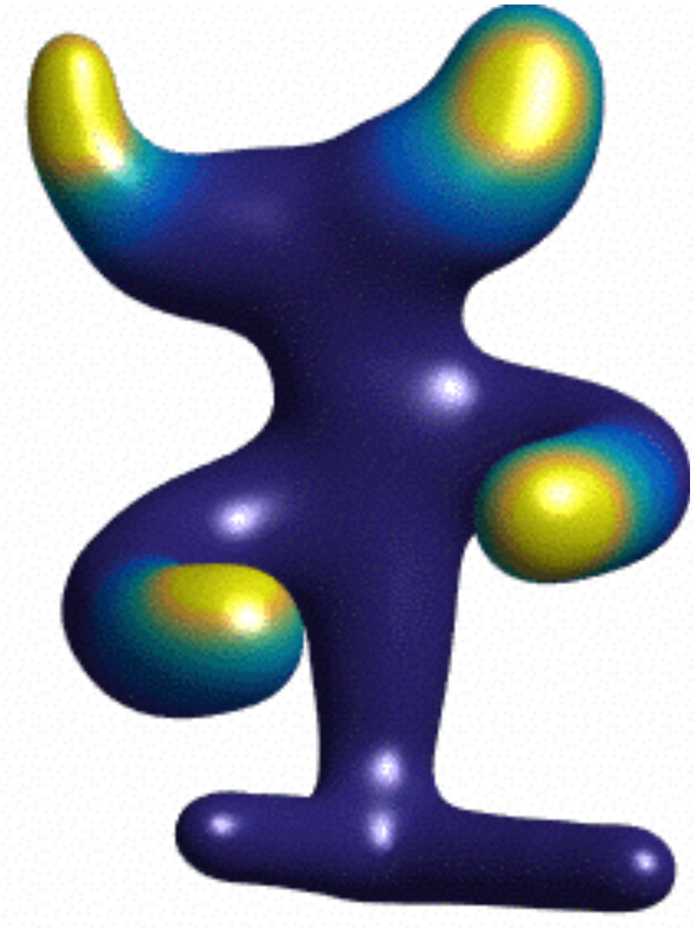


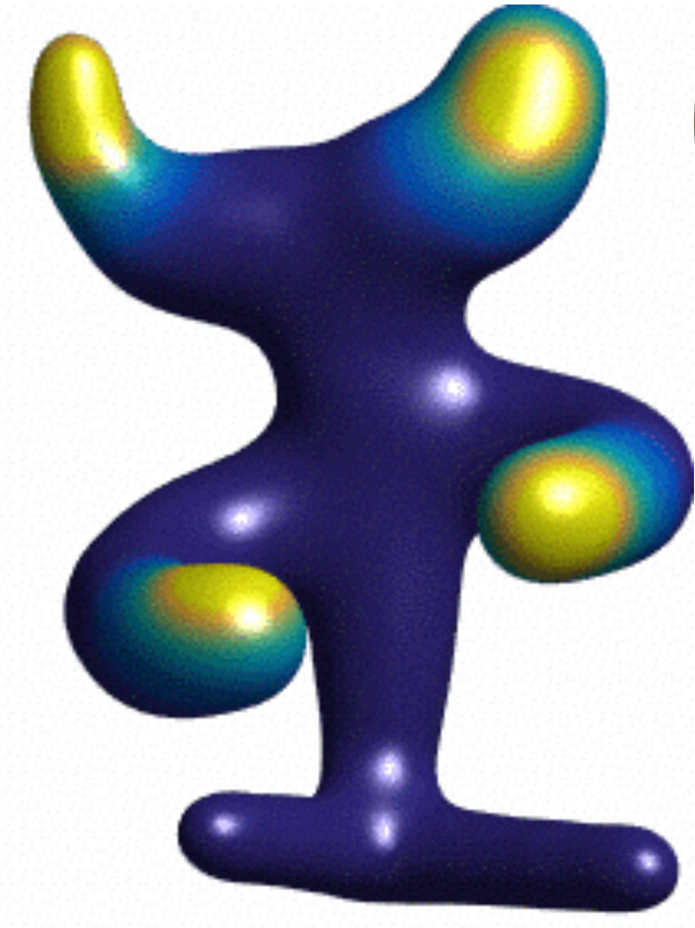
$\kappa = 6\|\mu_{t=0}\|_\infty$

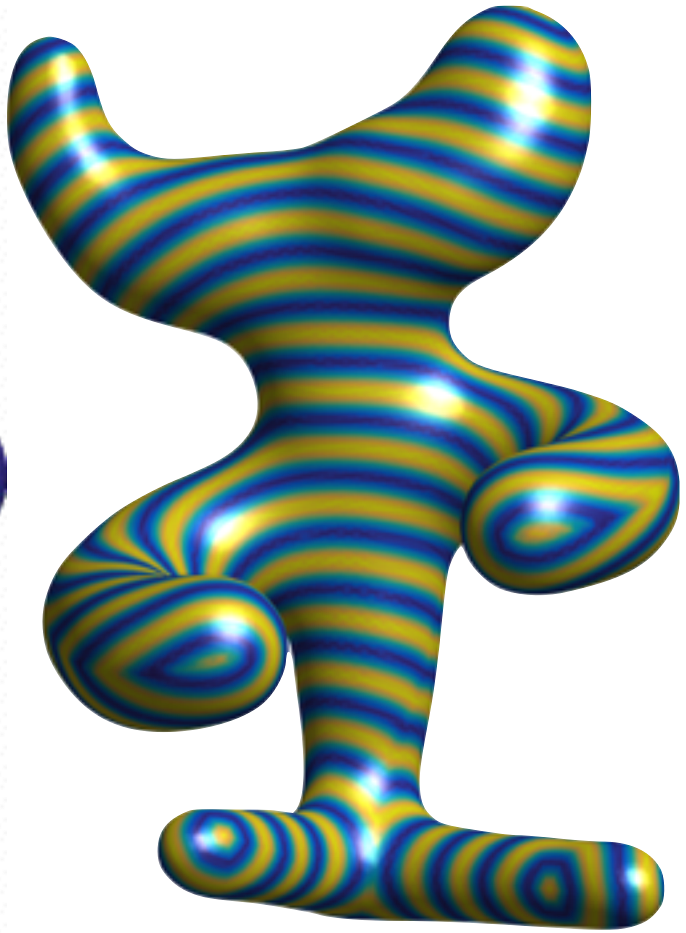
Potential $\cos(w)$

Crowd Motion on a Surface

M = triangulated mesh.



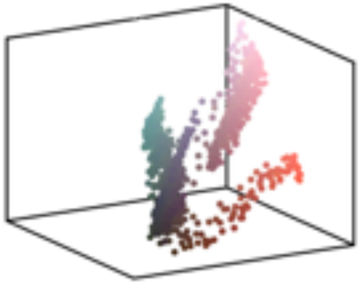




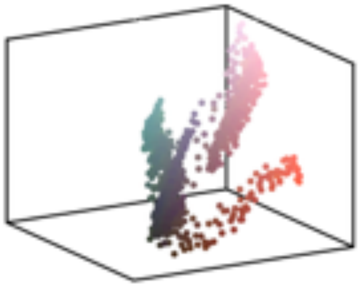
$\kappa = \|\mu_{t=0}\|_\infty$ $\qquad$ $\kappa = 6\|\mu_{t=0}\|_\infty$ $\qquad$ Potential $\cos(w)$

Conclusion

Histogram features in imaging and machine learning.

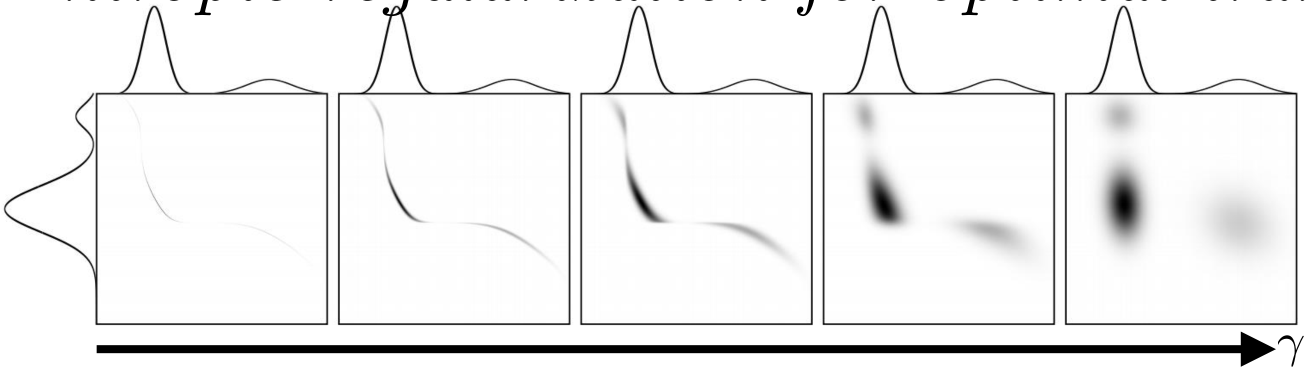


Conclusion



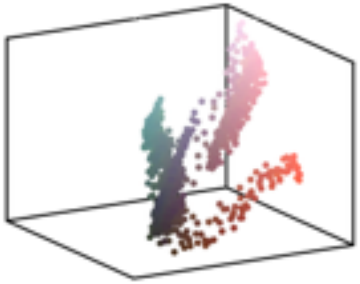
Histogram features in imaging and machine learning.

Entropic regularization for optimal transport.

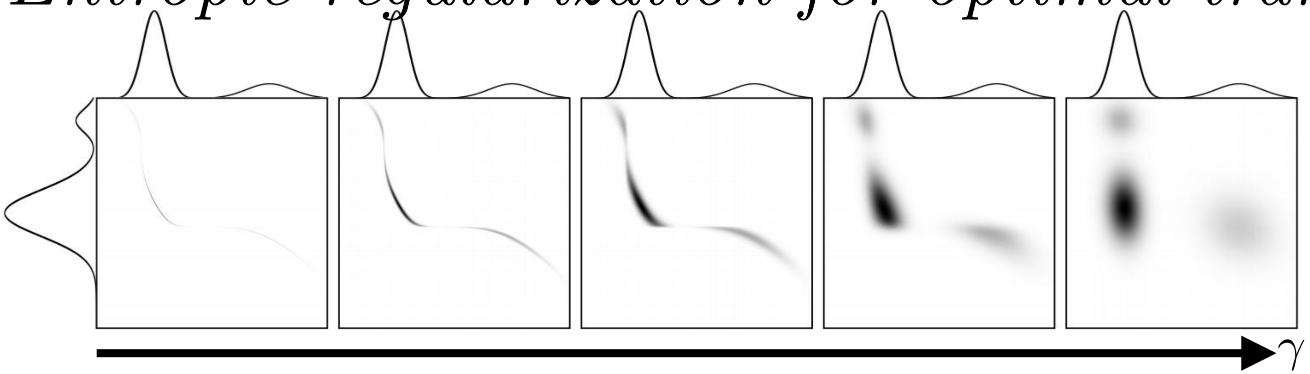


Conclusion

Histogram features in imaging and machine learning.



Entropic regularization for optimal transport.



Barycenters, unbalanced OT, gradient flows, . . .