Numerical Optimal Transport

http://optimaltransport.github.io

Entropic Regularization

Gabriel Peyré

www.numerical-tours.com

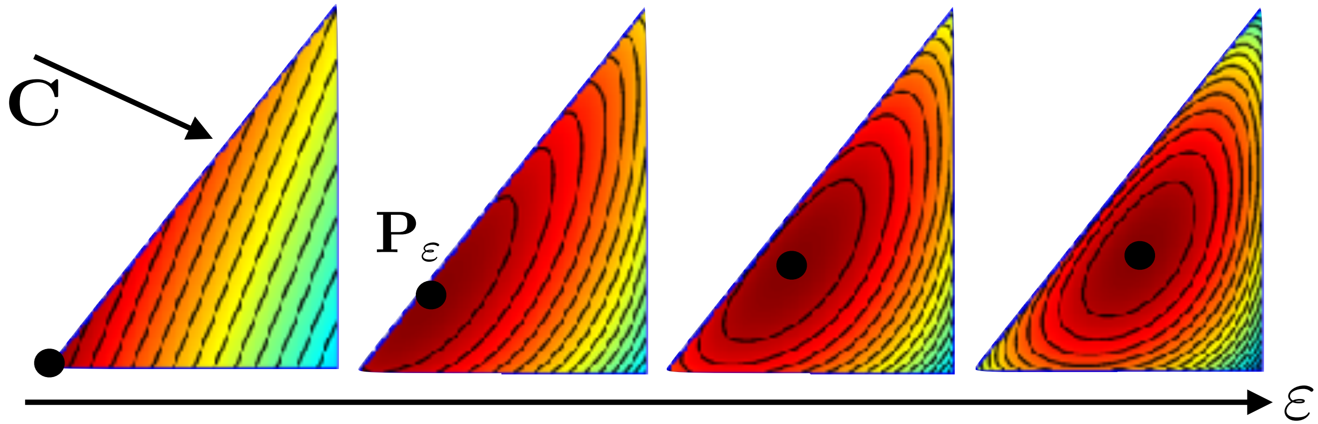
Overview

- **Entropic Regularization and Sinkhorn**
- Convergence Analysis
- Barycenters

Entropic Regularization

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

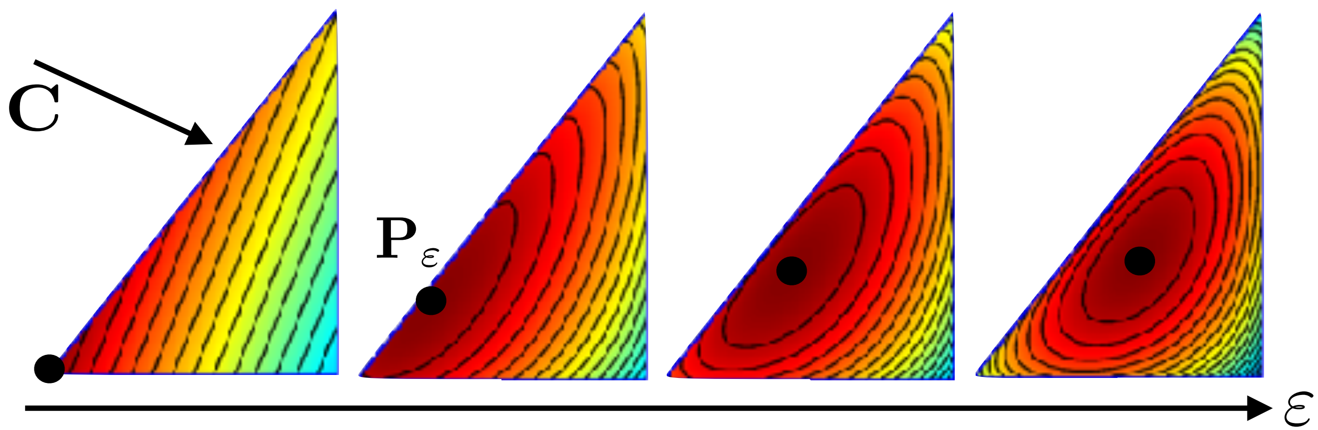
$$\mathrm{L}_{\mathbf{C}}^{\varepsilon}(\mathbf{a}, \mathbf{b}) \stackrel{\text{def.}}{=} \min_{\mathbf{P} \in \mathbf{U}(\mathbf{a},\mathbf{b})} \langle \mathbf{P}, \mathbf{C} \rangle - \varepsilon \mathbf{H}(\mathbf{P})$$

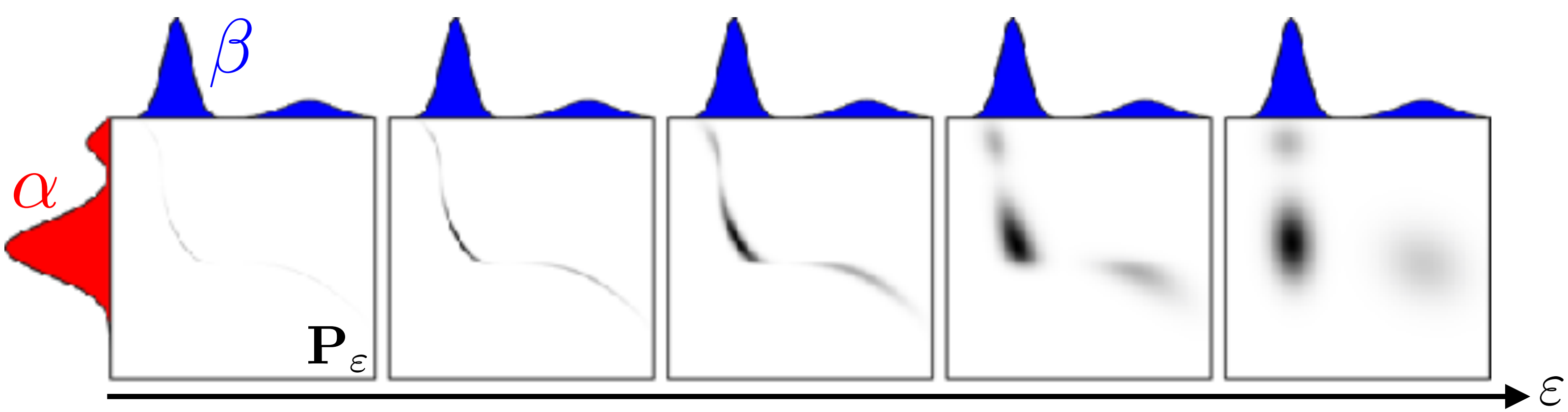


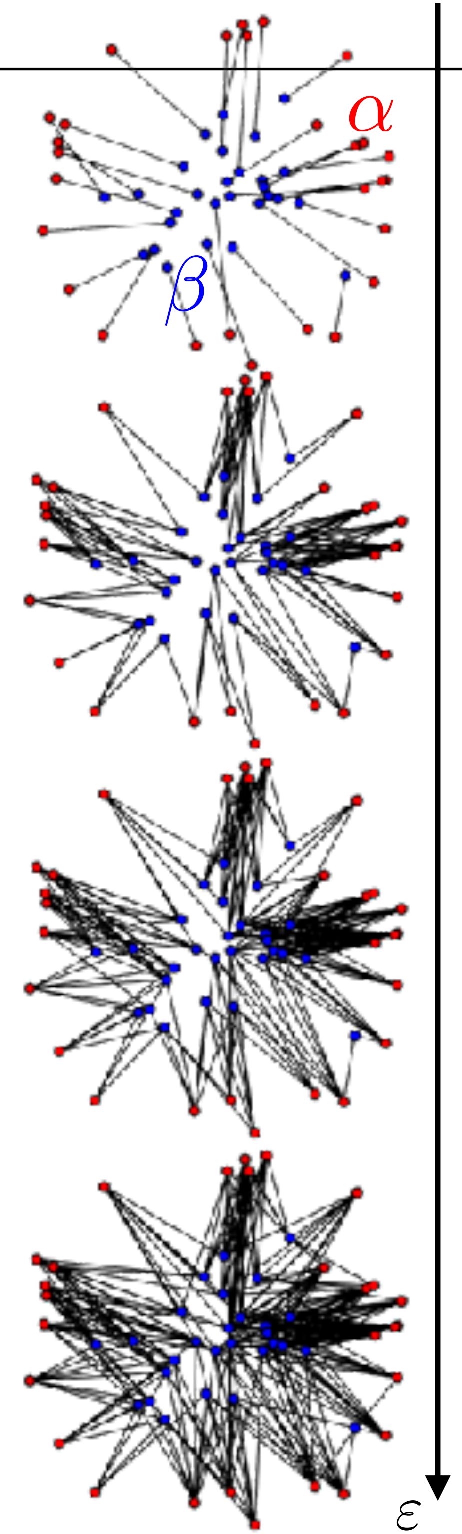
Entropic Regularization

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

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



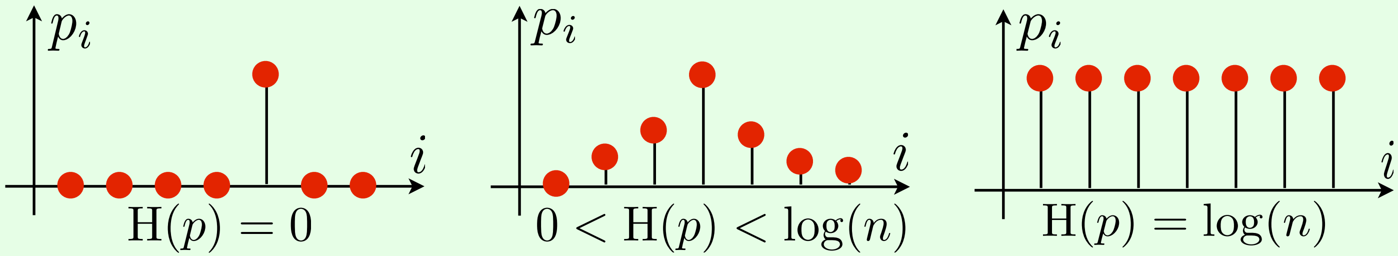




Entropies

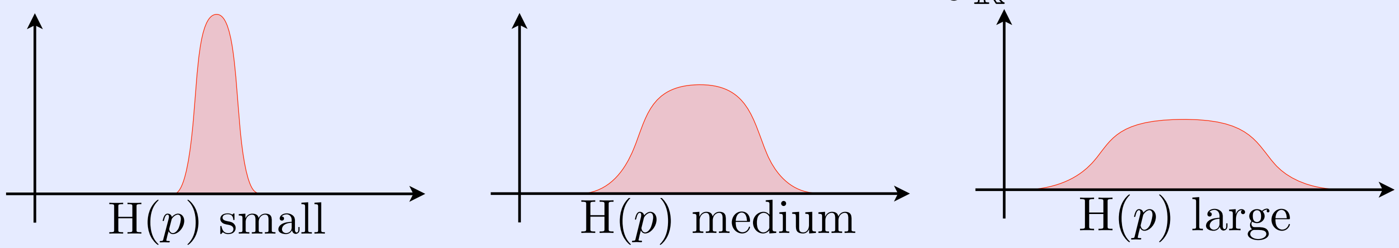
Discrete

$p_i \geqslant 0, \sum_{i=1}^n p_i = 1 \qquad \mathrm{H}(p) \overset{\text{def.}}{=} -\sum_i p_i \log(p_i)$



Continuous

$p(x) \geqslant 0, \int_{\mathbb{R}^d} p(x) = 1 \quad \mathrm{H}(p) \overset{\text{def.}}{=} -\int_{\mathbb{R}^d} p(x) \log(p(x)) \mathrm{d}x$

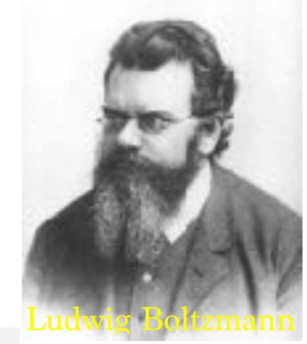


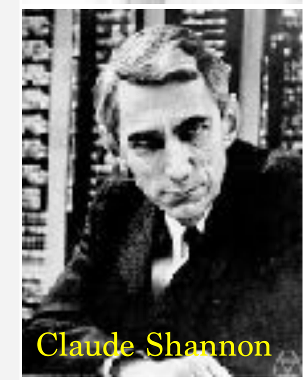
General

Relative entropy (Kullback-Leibler)

Measures (μ, ν): $\qquad \mathrm{KL}(\mu|\nu) \overset{\text{def.}}{=} \int_{\mathcal{X}} \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x)\right) \mathrm{d}\mu(x)$

$\mathrm{H}(p) = -\mathrm{KL}(p\,\mathrm{d}x|\mathrm{d}x)$





Arbitrary Measures

Relative-entropy:

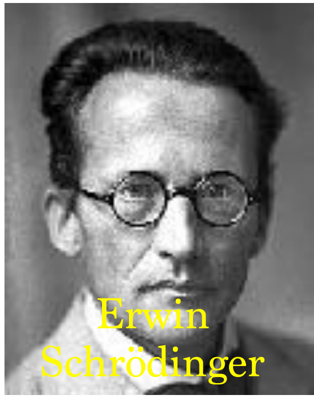
$$\mathrm{KL}(\pi|\xi) \overset{\text{def.}}{=} \int_{\mathcal{X}\times\mathcal{Y}} \log\left(\frac{\mathrm{d}\pi}{\mathrm{d}\xi}(x,y)\right)\mathrm{d}\pi(x,y) + \int_{\mathcal{X}\times\mathcal{Y}} (\mathrm{d}\xi(x,y) - \mathrm{d}\pi(x,y))$$

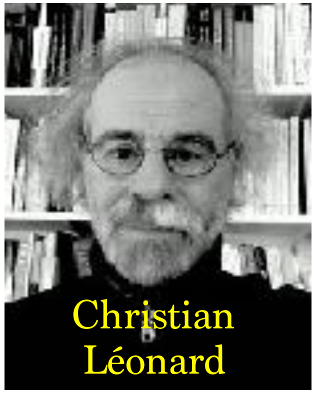
Schrödinger's problem: [1931]

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

$$\min_{(X,Y)} \{\mathbb{E}(c(X,Y)) + \varepsilon \mathrm{I}(X,Y) \;;\; X\sim\alpha, Y\sim\beta\}$$

Mutual information





Arbitrary Measures

Relative-entropy:

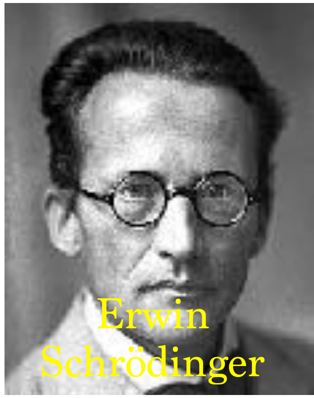
$$\mathrm{KL}(\pi|\xi) \overset{\text{def.}}{=} \int_{\mathcal{X}\times\mathcal{Y}} \log\left(\frac{\mathrm{d}\pi}{\mathrm{d}\xi}(x,y)\right)\mathrm{d}\pi(x,y) + \int_{\mathcal{X}\times\mathcal{Y}} (\mathrm{d}\xi(x,y) - \mathrm{d}\pi(x,y))$$

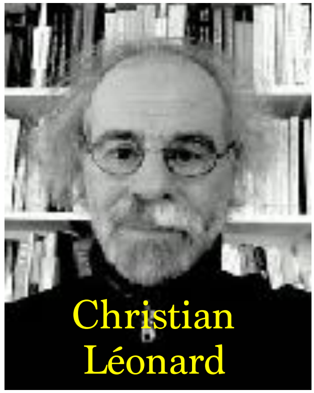
Schrödinger's problem: [1931]

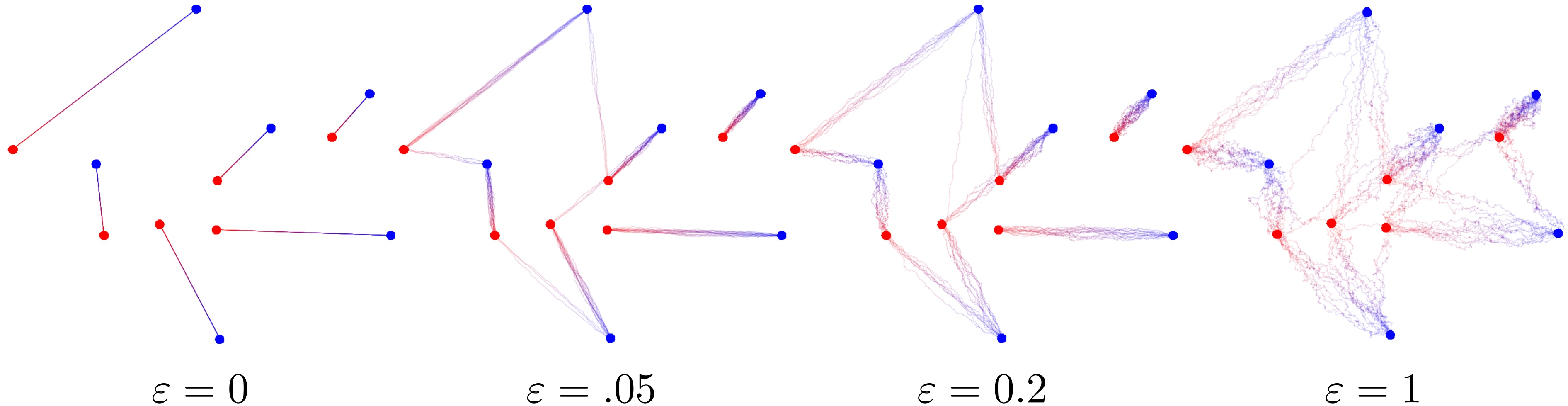
$$\mathcal{L}_c^\varepsilon(\alpha,\beta) \overset{\text{def.}}{=} \min_{\pi\in\mathcal{U}(\alpha,\beta)} \int_{X\times Y} c(x,y)\mathrm{d}\pi(x,y) + \varepsilon\,\mathrm{KL}(\pi|\alpha\otimes\beta)$$

$$\min_{(X,Y)} \{\mathbb{E}(c(X,Y)) + \varepsilon \mathrm{I}(X,Y)\ ;\ X\sim\alpha, Y\sim\beta\}$$

Mutual information







Brownian Bridge

Random walk: $x_{k+1} = x_k + \frac{\sigma}{\sqrt{K}} \varepsilon_k$

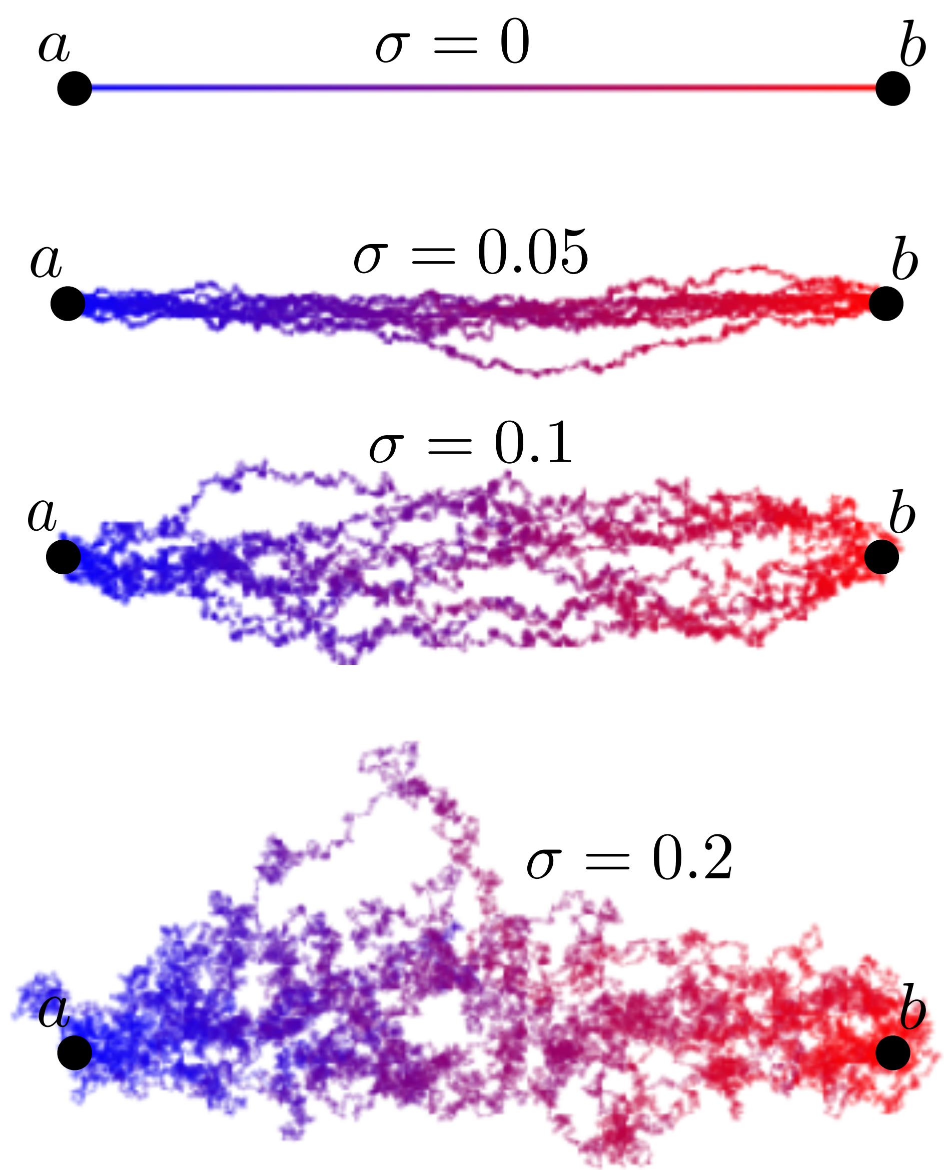
$\varepsilon_k \sim \mathcal{N}(0, \mathrm{Id}_{\mathbb{R}^2})$

Brownian motion / Wiener process:

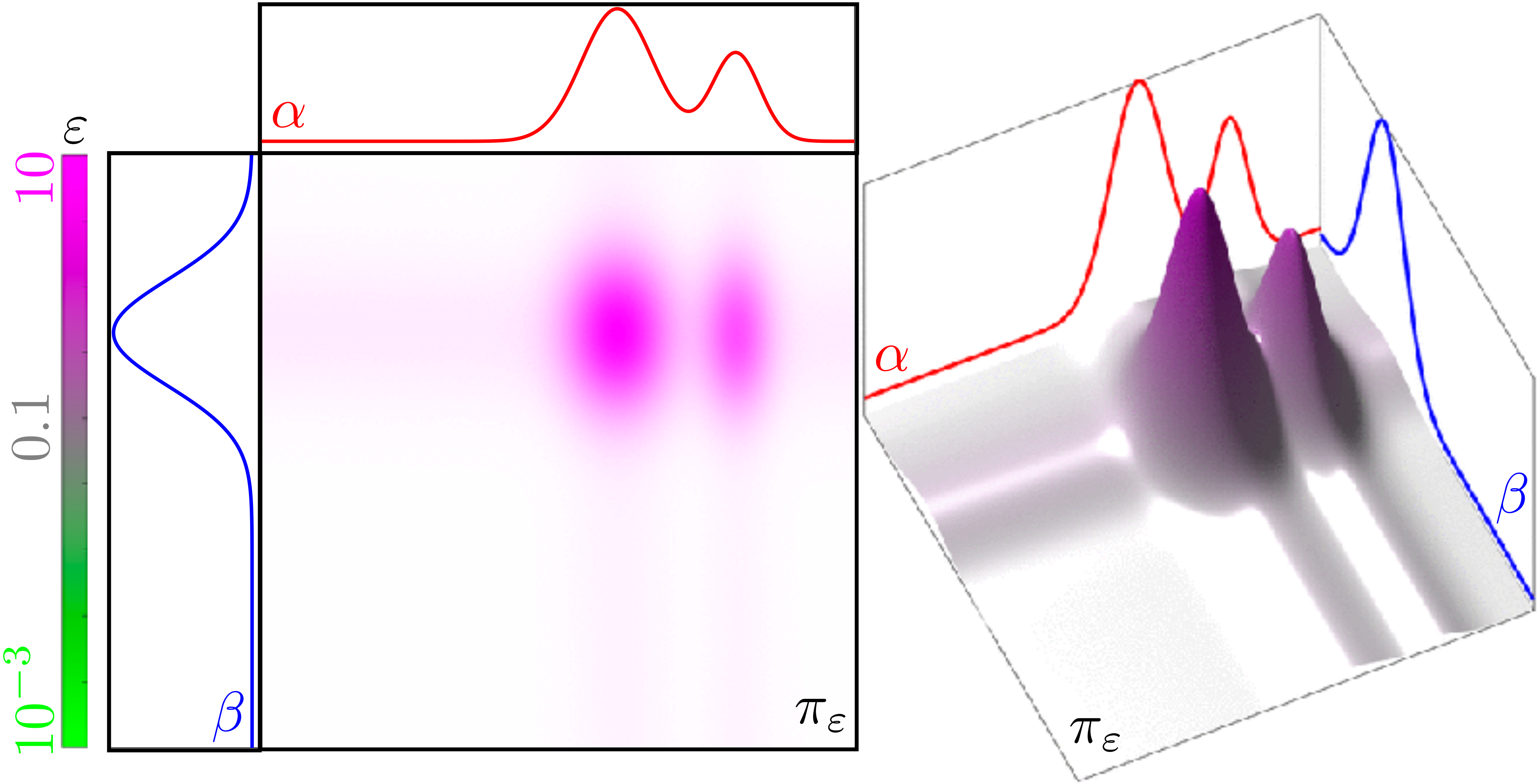
$$x_k \xrightarrow[k/K \to t]{K \to +\infty} W_t$$

Brownian bridge between $(a, b) \in \mathbb{C}^2$:

$$a + (b - a) \frac{x(t) - x(0)}{x(1) - x(0)}$$



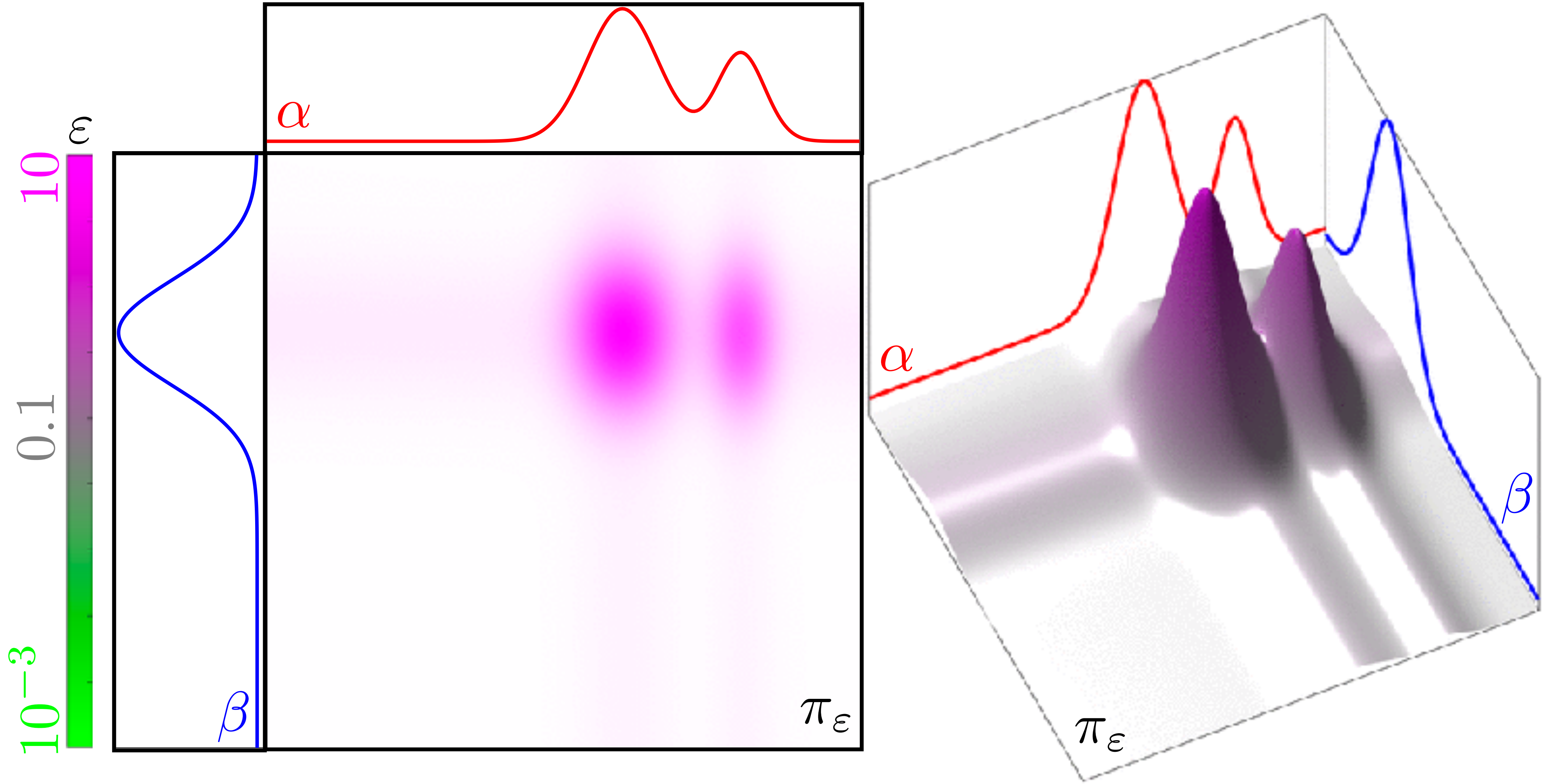
Impact of Regularization



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

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

Impact of Regularization



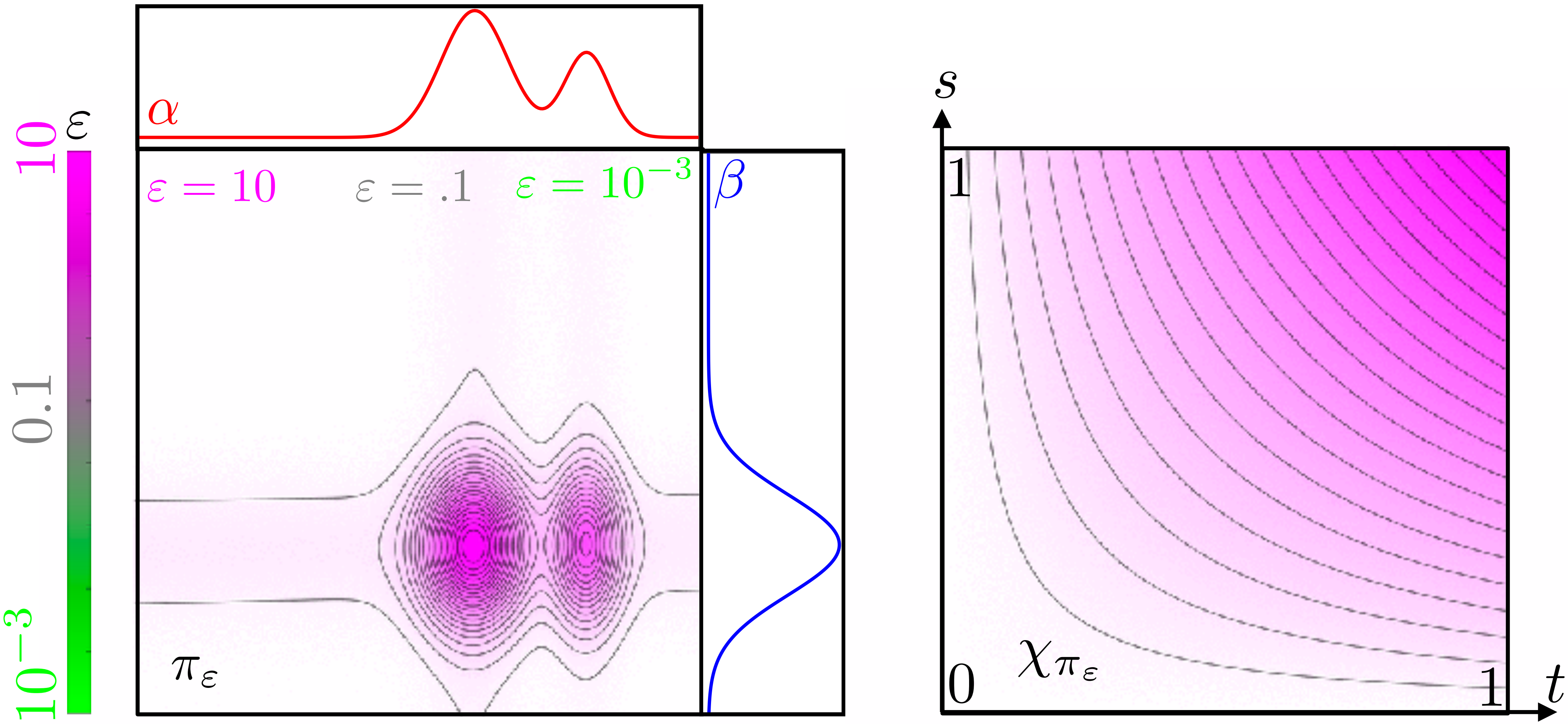
$$\pi_\varepsilon = \operatorname{argmin}_\pi \left\{ \int_{\mathbb{R}^2} \left(\|x-y\|^2 + \varepsilon \log\left(\frac{\mathrm{d}\pi}{\mathrm{d}\alpha \mathrm{d}\beta}(x,y) \right) \right) \mathrm{d}\pi(x,y) + \; ; \; \pi_1 = \alpha, \pi_2 = \beta \right\}$$

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



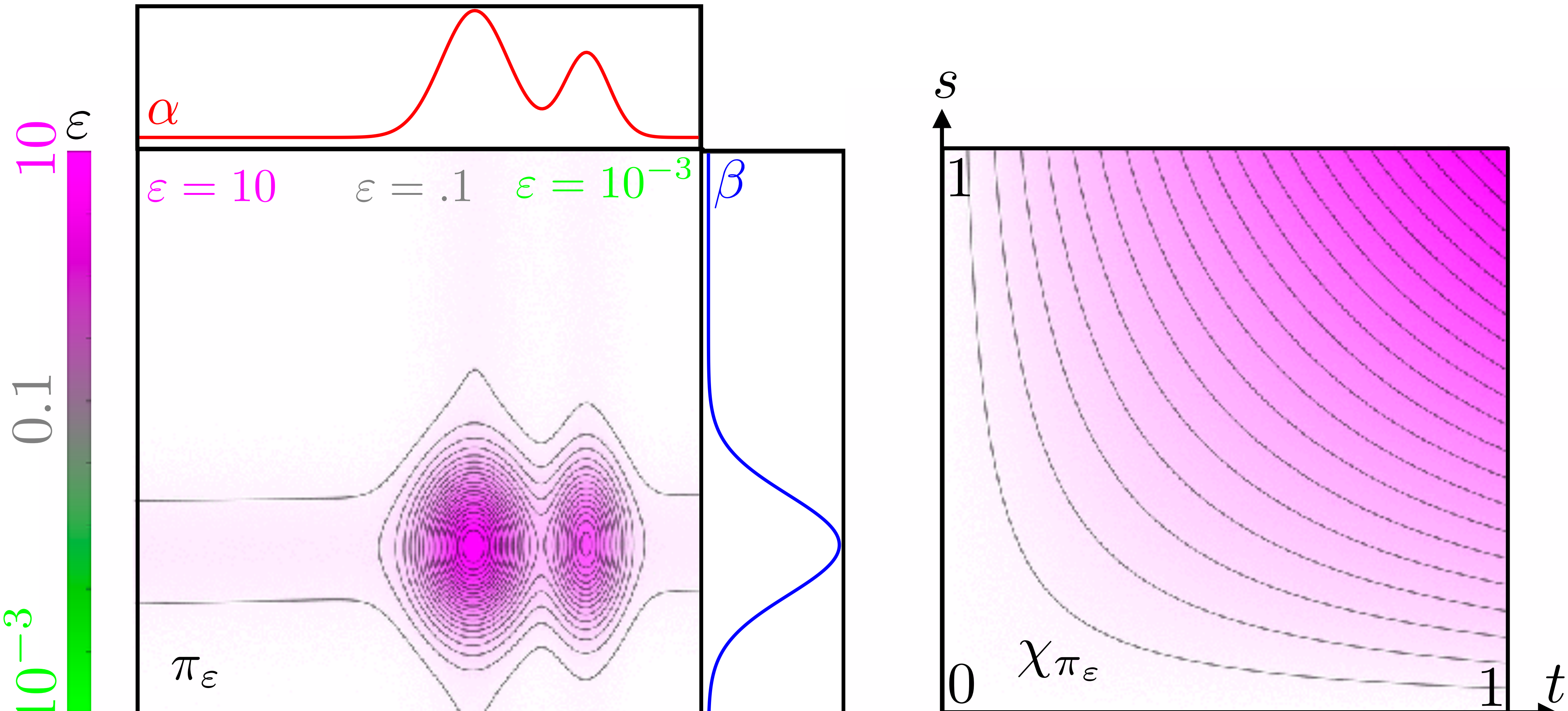
Theorem: $\chi_{\pi_\varepsilon}(s,t)$ $\xrightarrow{\varepsilon \to 0}$ $\min(s,t)$ (dependence)

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

Impact of Regularization

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

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



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

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

Sinkhorn's Algorithm

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

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

Sinkhorn's Algorithm

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

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

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's Algorithm

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

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

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

Sinkhorn's Algorithm

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

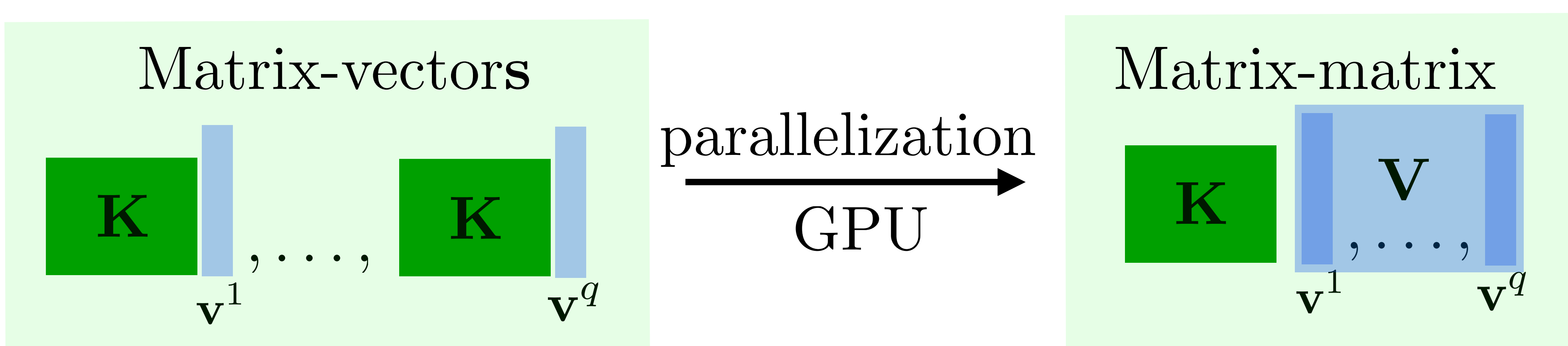
Proposition: $\forall\,(i,j) \in [\![n]\!] \times [\![m]\!], \quad \mathbf{P}_{i,j} = \mathbf{u}_i \mathbf{K}_{i,j} \mathbf{v}_j$

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

Only matrix/vector multiplications.



Sinkhorn's Algorithm

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

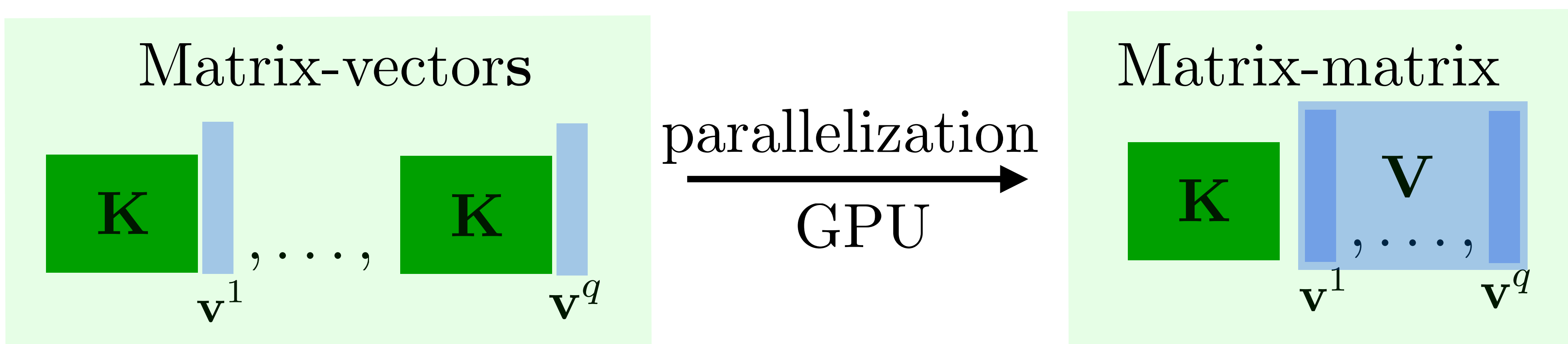
Proposition: $\forall\,(i,j) \in [\![n]\!] \times [\![m]\!], \quad \mathbf{P}_{i,j} = \mathbf{u}_i \mathbf{K}_{i,j} \mathbf{v}_j$

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

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

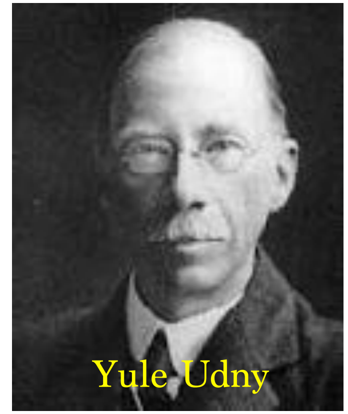
Sinkhorn iterations: $\mathbf{u}^{(\ell+1)} \overset{\text{def.}}{=} \dfrac{\mathbf{a}}{\mathbf{K}\mathbf{v}^{(\ell)}}$ and $\mathbf{v}^{(\ell+1)} \overset{\text{def.}}{=} \dfrac{\mathbf{b}}{\mathbf{K}^{\mathrm{T}}\mathbf{u}^{(\ell+1)}}$

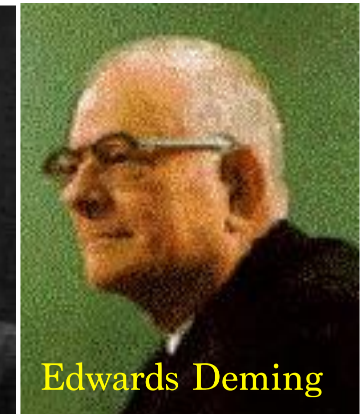
Only matrix/vector multiplications.



$\rightarrow$ Convolution on regular grids, separable kernels.

Sinkhorn, IPFP, RAS, ...





Many names:

- Sinkhorn algorithm
- DAD scaling
- Iterative proportional fitting
- Biproportional fitting
- RAS algorithm
- Matrix scaling

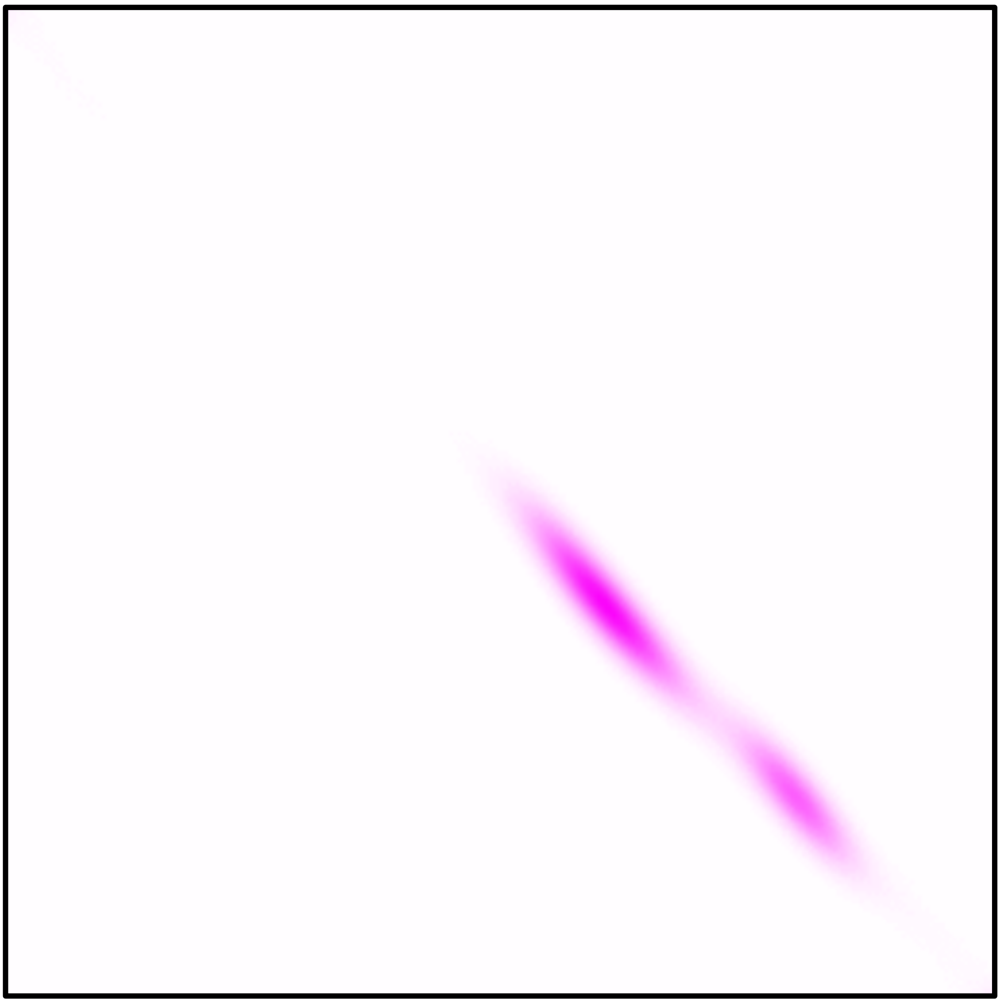
Udny 1912

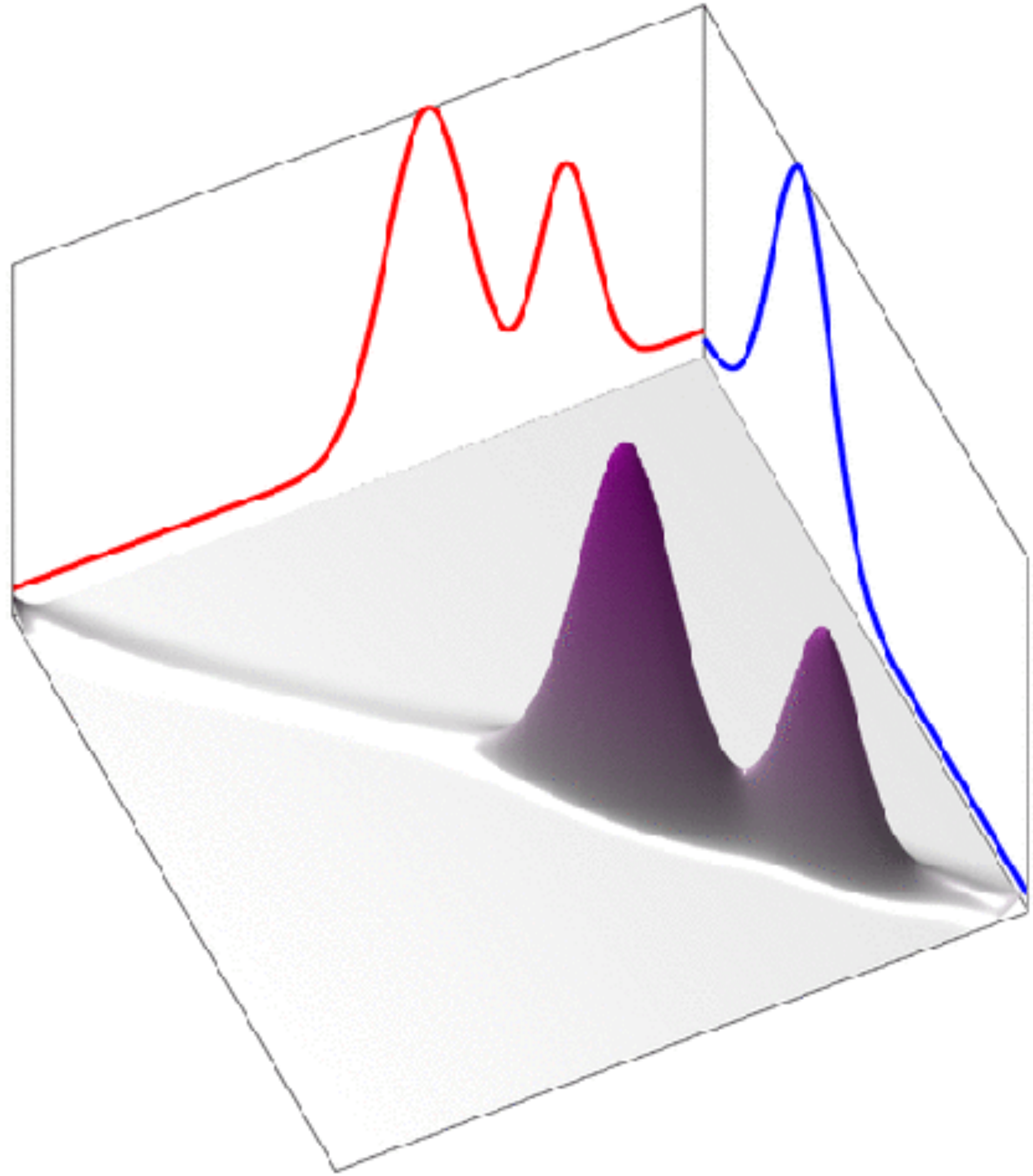
Kruithof, 1937

Deming and Stephan in 1940

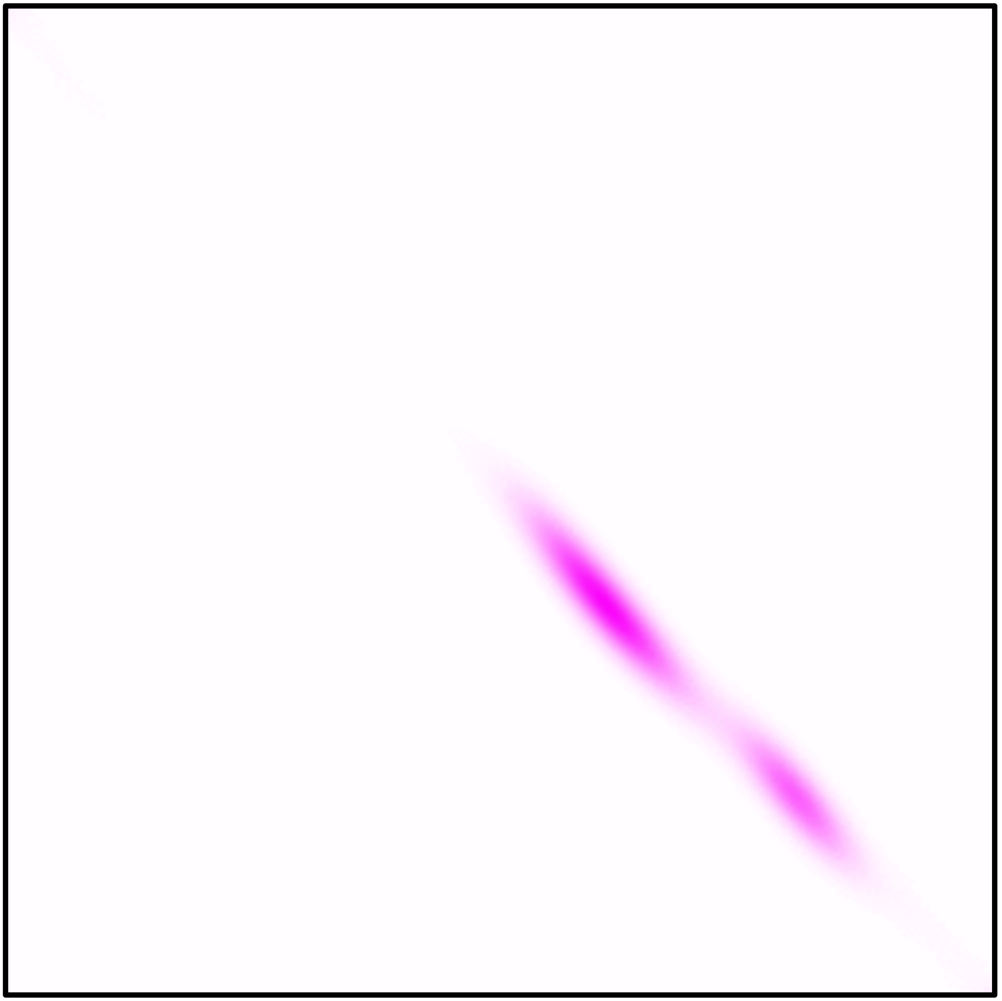
Sinkhorn 1964

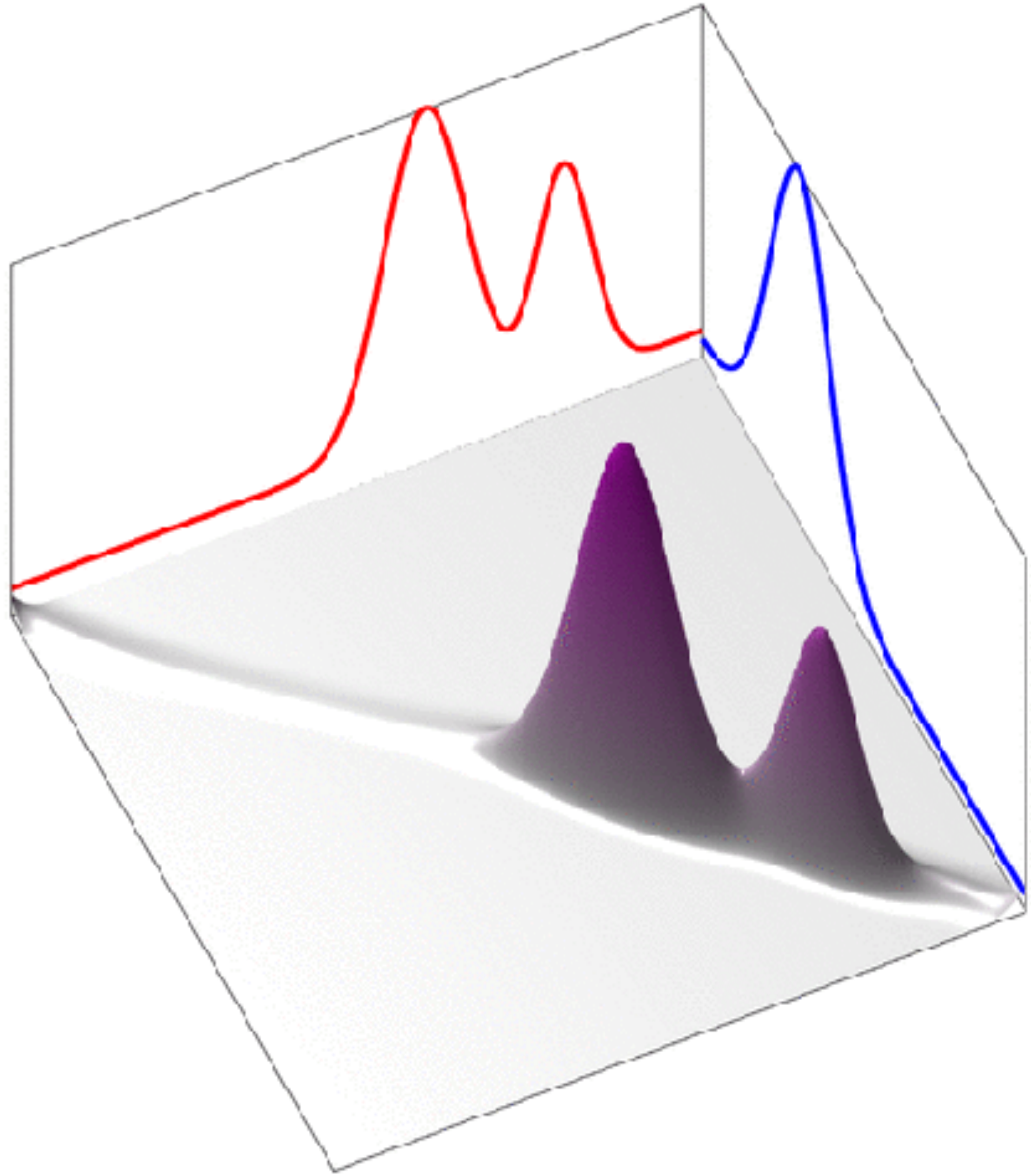
Sinkhorn Evolution





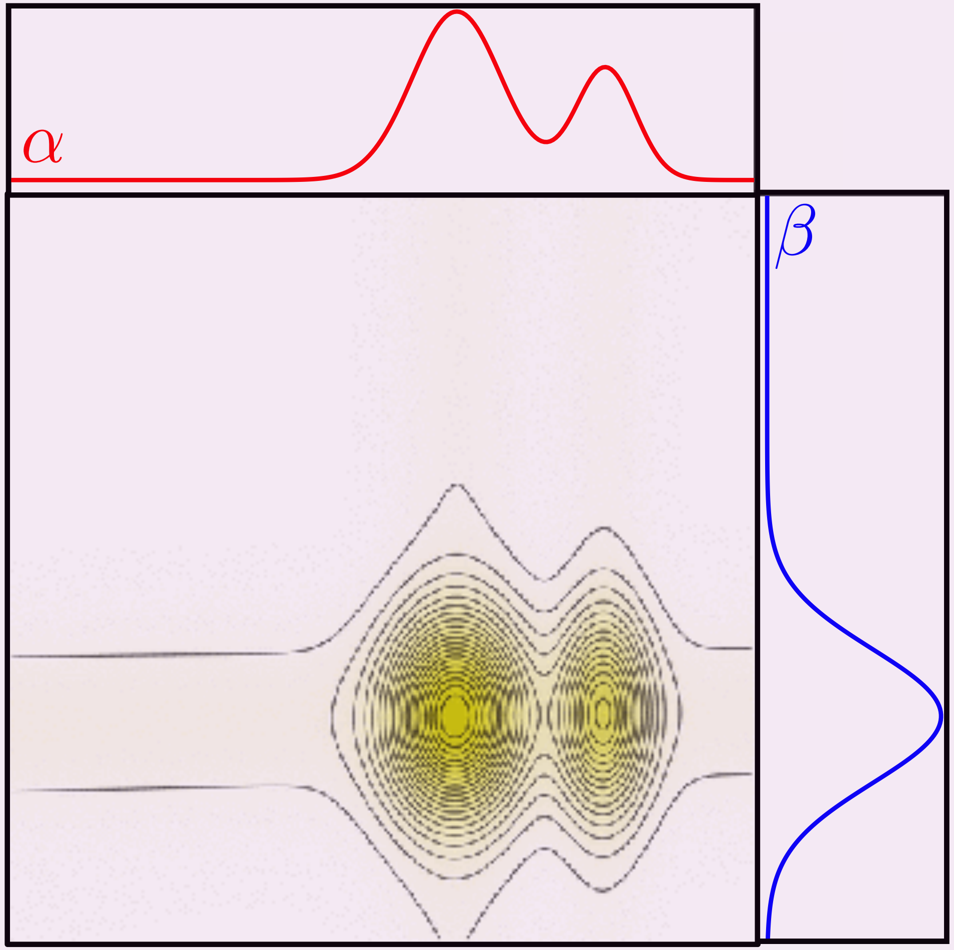
Sinkhorn Evolution





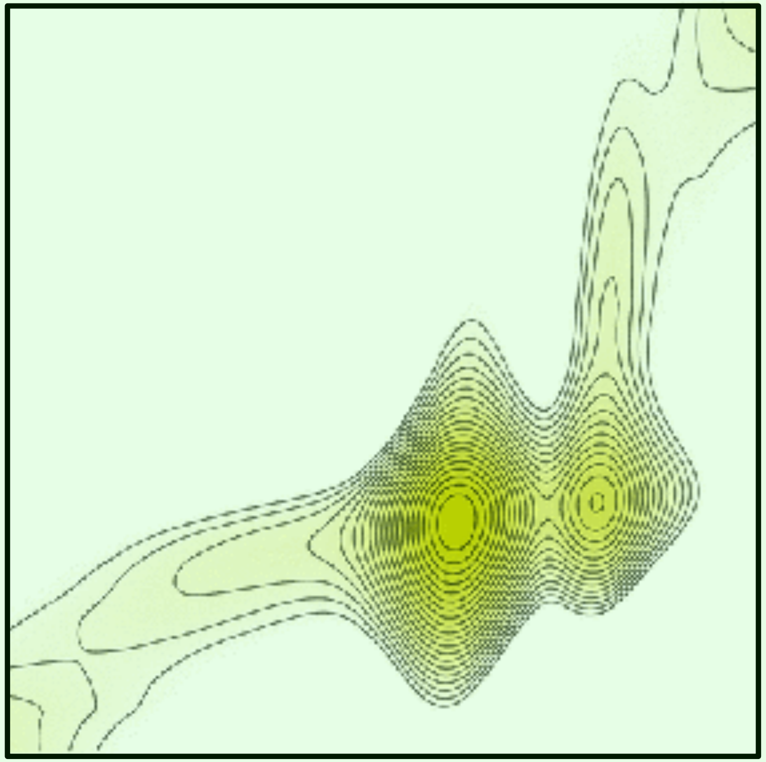
Other Regularizations

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



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

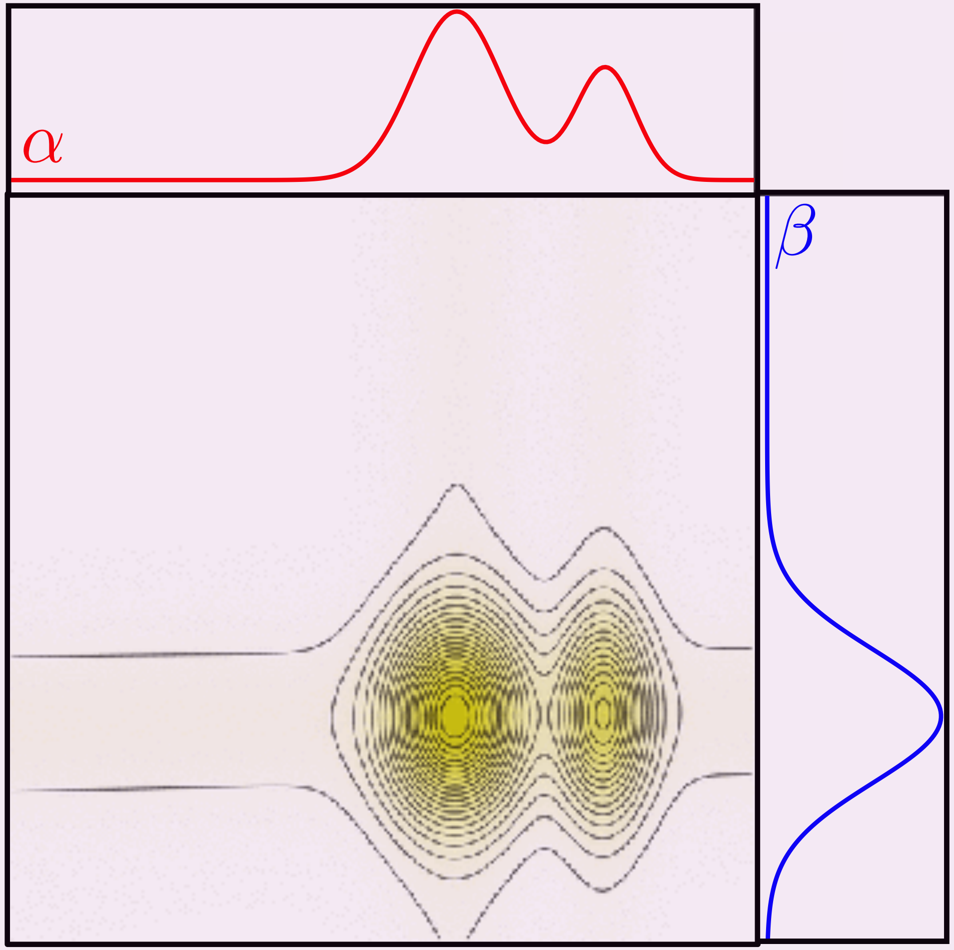
Dykstra's algorithm



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

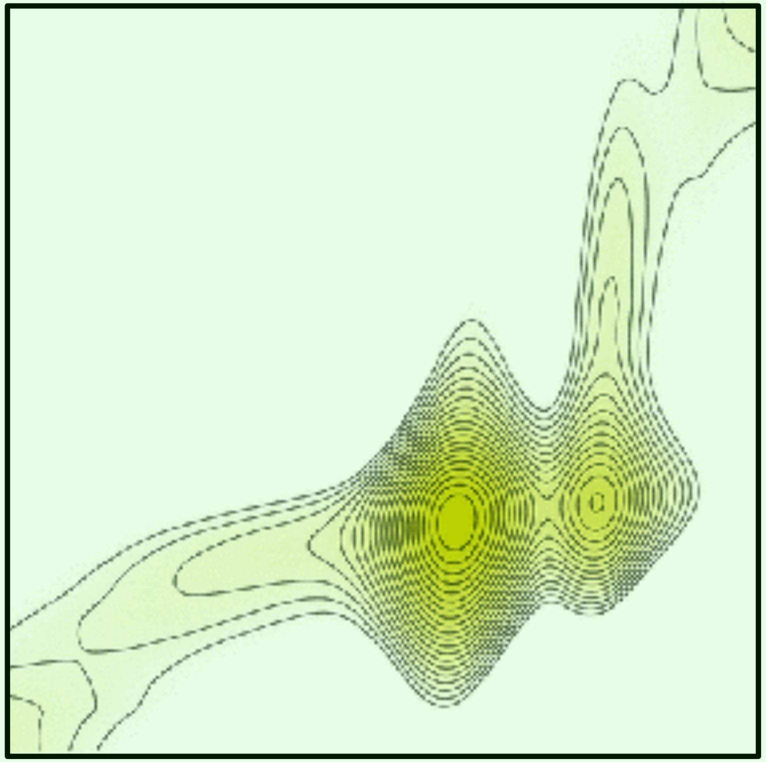
Other Regularizations

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



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

Dykstra's algorithm



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

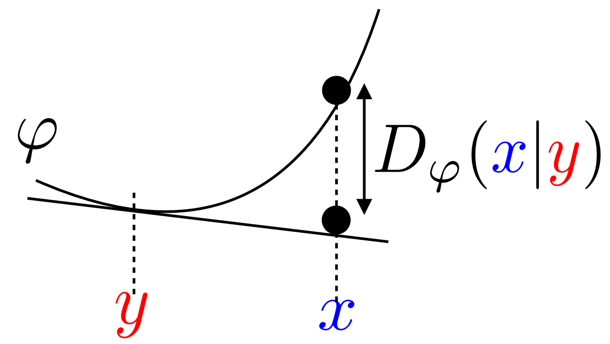
Overview

- Entropic Regularization and Sinkhorn
- **Convergence Analysis**
- Barycenters

Bregman Divergences

Bregman divergence: φ convex

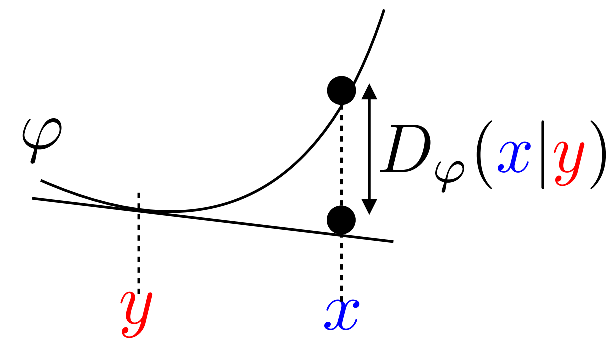
$$D_\varphi(x|y) \overset{\text{def.}}{=} \varphi(x) - \varphi(y) - \langle x - y,\, \nabla\varphi(y)\rangle$$

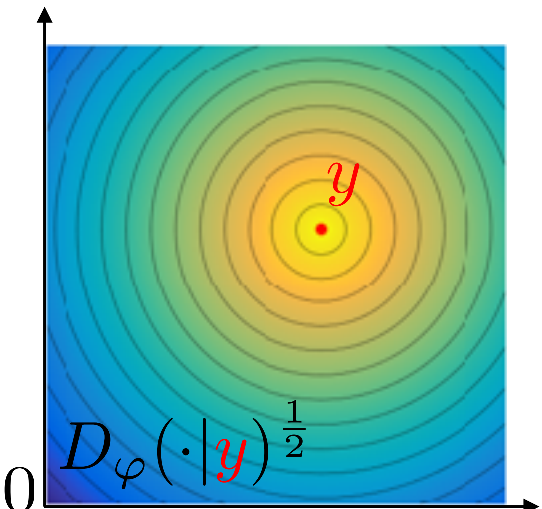


Bregman Divergences

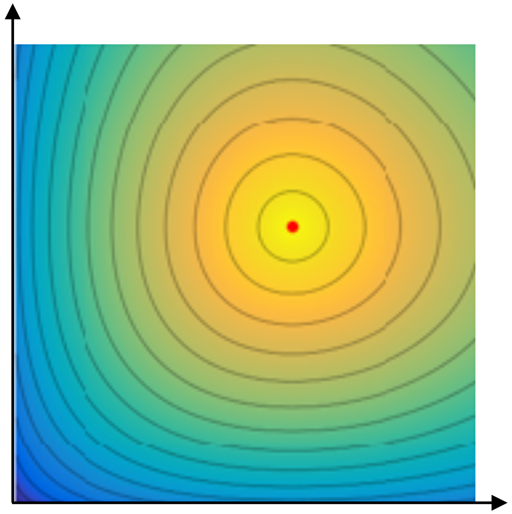
Bregman divergence: φ convex

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

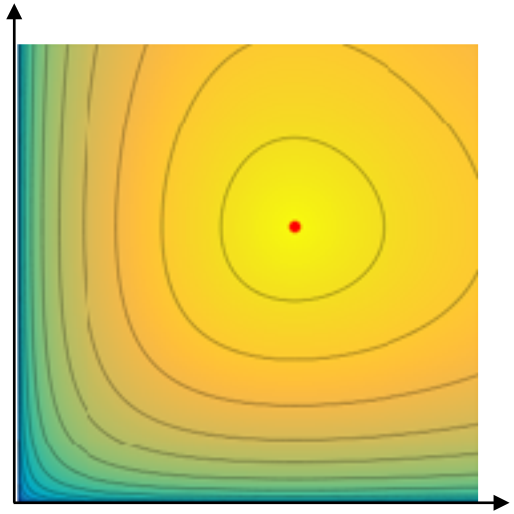




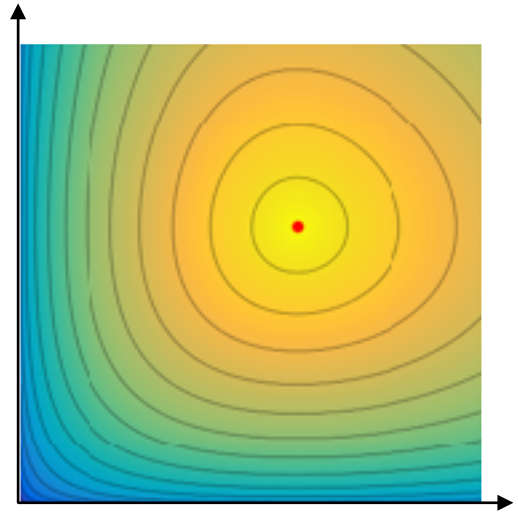
$\varphi(x) = \|x\|^2$



$\varphi(x) = \sum_i x_i \log(x_i)$



$\varphi(x) = \sum_i -\log(x_i)$

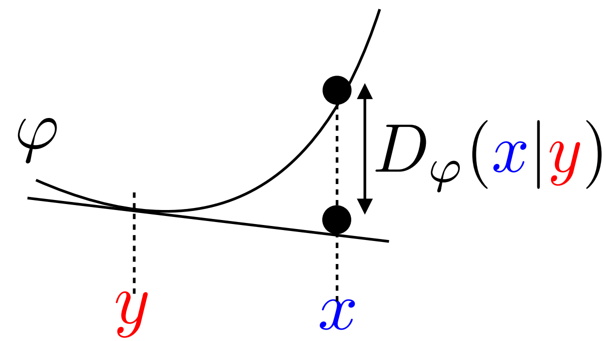


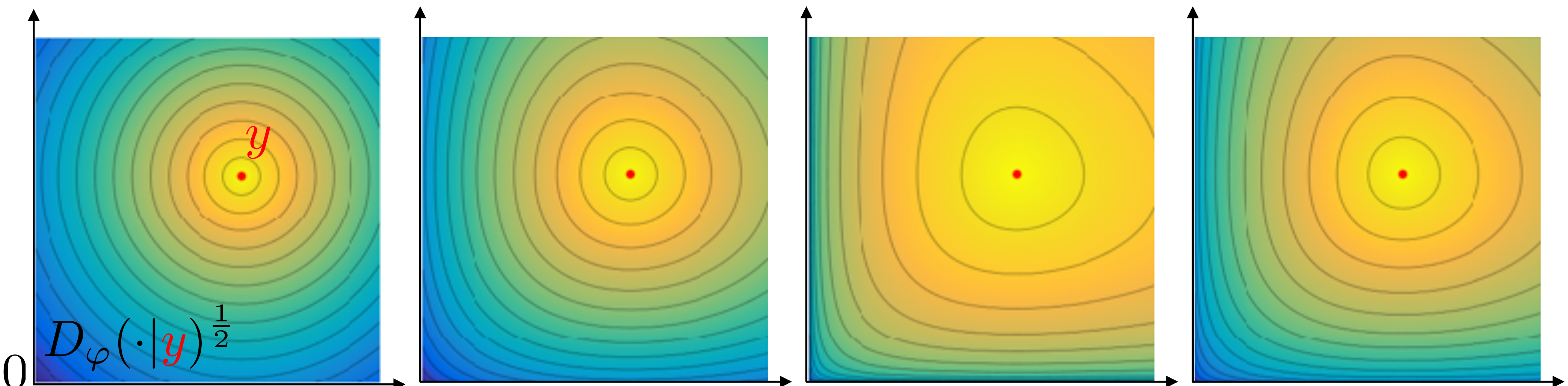
$\varphi(x) = \sum_i -\sqrt{x_i}$

Bregman Divergences

Bregman divergence: φ convex

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





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

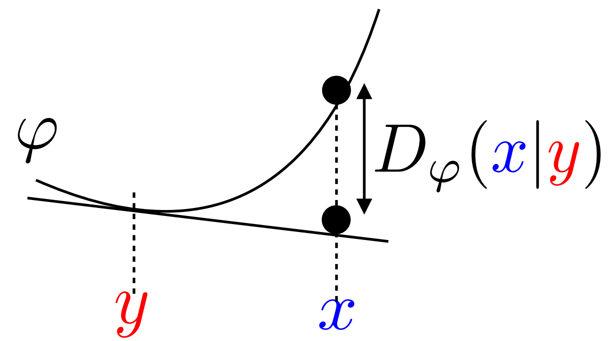
Locally Euclidean:

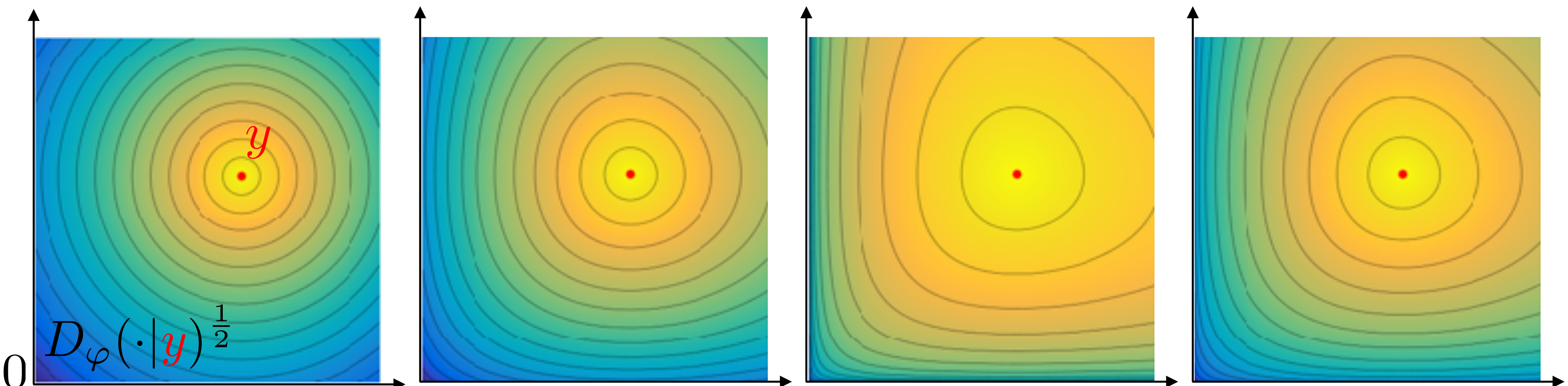
$$\left.\begin{array}{l} D_\varphi(x|x+\varepsilon) \\ D_\varphi(x+\varepsilon|x) \end{array}\right\} = \frac{1}{2}\langle \partial^2\varphi(x)\varepsilon,\, \varepsilon\rangle + o(\|\varepsilon\|^2)$$

Bregman Divergences

Bregman divergence: φ convex

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





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

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

Usually do not work for Radon Measures. Need to use f-divergences

Bregman Iterative Projections

KL divergence:

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

$$\mathrm{KL}(\mathbf{P}|\mathbf{K}) = D_\varphi(\mathbf{P}|\mathbf{K}) \quad \text{for} \quad \varphi(\mathbf{P}) = \sum_{i,j} \mathbf{P}_{i,j} \log(\mathbf{P}_{i,j})$$

$Proposition:$

$$\min_{\mathbf{P}\in\mathbf{U}(\mathbf{a},\mathbf{b})} \langle \mathbf{P},\, \mathbf{C}\rangle - \varepsilon \mathbf{H}(\mathbf{P}) \quad \Leftrightarrow \min_{\mathbf{P}\in\mathbf{U}(\mathbf{a},\mathbf{b})} \varepsilon\, \mathrm{KL}(\mathbf{P}|\mathbf{K}) \qquad \mathbf{K}_{i,j} \stackrel{\text{def.}}{=} e^{-\frac{\mathbf{C}_{i,j}}{\varepsilon}}$$

Bregman Iterative Projections

KL divergence:

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

$$\mathrm{KL}(\mathbf{P}|\mathbf{K}) = D_\varphi(\mathbf{P}|\mathbf{K}) \quad \text{for} \quad \varphi(\mathbf{P}) = \sum_{i,j} \mathbf{P}_{i,j} \log(\mathbf{P}_{i,j})$$

Proposition :

$$\min_{\mathbf{P}\in\mathbf{U}(\mathbf{a},\mathbf{b})} \langle \mathbf{P},\, \mathbf{C} \rangle - \varepsilon \mathbf{H}(\mathbf{P}) \;\Leftrightarrow\; \min_{\mathbf{P}\in\mathbf{U}(\mathbf{a},\mathbf{b})} \varepsilon\, \mathrm{KL}(\mathbf{P}|\mathbf{K}) \qquad \mathbf{K}_{i,j} \stackrel{\text{def.}}{=} e^{-\frac{\mathbf{C}_{i,j}}{\varepsilon}}$$

$$\mathbf{U}(\mathbf{a},\mathbf{b}) = \mathcal{C}^1_{\mathbf{a}} \cup \mathcal{C}^2_{\mathbf{b}}$$

$$\mathcal{C}^1_{\mathbf{a}} \stackrel{\text{def.}}{=} \{\mathbf{P} \;:\; \mathbf{P}\mathbb{1}_m = \mathbf{a}\}$$

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

Bregman Iterative Projections

KL divergence:

$$\mathbf{KL}(\mathbf{P}|\mathbf{K}) \overset{\text{def.}}{=} \sum_{i,j} \mathbf{P}_{i,j} \log\left(\frac{\mathbf{P}_{i,j}}{\mathbf{K}_{i,j}}\right) - \mathbf{P}_{i,j} + \mathbf{K}_{i,j}$$

$$\mathrm{KL}(\mathbf{P}|\mathbf{K}) = D_\varphi(\mathbf{P}|\mathbf{K}) \quad \text{for} \quad \varphi(\mathbf{P}) = \sum_{i,j} \mathbf{P}_{i,j} \log(\mathbf{P}_{i,j})$$

Proposition :

$$\min_{\mathbf{P}\in\mathbf{U}(\mathbf{a},\mathbf{b})} \langle \mathbf{P},\, \mathbf{C}\rangle - \varepsilon \mathbf{H}(\mathbf{P}) \iff \min_{\mathbf{P}\in\mathbf{U}(\mathbf{a},\mathbf{b})} \varepsilon\, \mathrm{KL}(\mathbf{P}|\mathbf{K}) \qquad \mathbf{K}_{i,j} \overset{\text{def.}}{=} e^{-\frac{\mathbf{C}_{i,j}}{\varepsilon}}$$

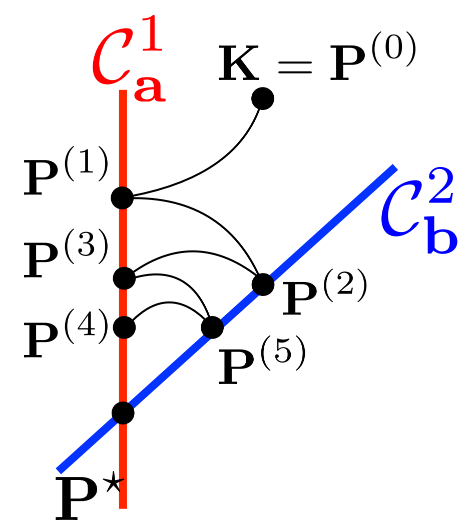
$$\mathbf{U}(\mathbf{a},\mathbf{b}) = \mathcal{C}^1_{\mathbf{a}} \cup \mathcal{C}^2_{\mathbf{b}} \qquad \mathcal{C}^1_{\mathbf{a}} \overset{\text{def.}}{=} \{\mathbf{P} \,:\, \mathbf{P}\mathbb{1}_m = \mathbf{a}\} \qquad \mathcal{C}^2_{\mathbf{b}} \overset{\text{def.}}{=} \left\{\mathbf{P} \,:\, \mathbf{P}^{\mathrm{T}}\mathbb{1}_m = \mathbf{b}\right\}$$

[Bregman, 1967]

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

Theorem: For affine $(\mathcal{C}^1_{\mathbf{a}}, \mathcal{C}^2_{\mathbf{b}})$,

$$\mathbf{P}^{(\ell)} \to \mathbf{P}^\star = \underset{\mathbf{P}\in\mathcal{C}^1_{\mathbf{a}}\cap\mathcal{C}^1_{\mathbf{b}}}{\operatorname{argmin}} \ \mathrm{KL}(\mathbf{P}|\mathbf{K})$$



Bregman Iterative Projections

KL divergence:

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

$$\mathrm{KL}(\mathbf{P}|\mathbf{K}) = D_\varphi(\mathbf{P}|\mathbf{K}) \quad \text{for} \quad \varphi(\mathbf{P}) = \sum_{i,j} \mathbf{P}_{i,j} \log(\mathbf{P}_{i,j})$$

Proposition :

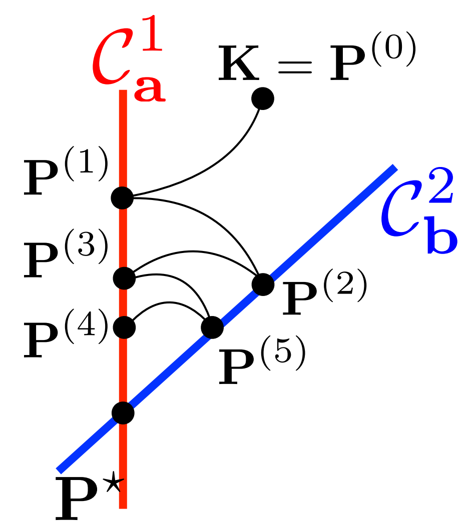
$$\min_{\mathbf{P}\in\mathbf{U}(\mathbf{a},\mathbf{b})} \langle \mathbf{P},\, \mathbf{C}\rangle - \varepsilon \mathbf{H}(\mathbf{P}) \iff \min_{\mathbf{P}\in\mathbf{U}(\mathbf{a},\mathbf{b})} \varepsilon\, \mathrm{KL}(\mathbf{P}|\mathbf{K}) \qquad \mathbf{K}_{i,j} \stackrel{\text{def.}}{=} e^{-\frac{\mathbf{C}_{i,j}}{\varepsilon}}$$

$$\mathbf{U}(\mathbf{a},\mathbf{b}) = \mathcal{C}^1_{\mathbf{a}} \cup \mathcal{C}^2_{\mathbf{b}} \qquad \mathcal{C}^1_{\mathbf{a}} \stackrel{\text{def.}}{=} \{\mathbf{P} \;:\; \mathbf{P}\mathbb{1}_m = \mathbf{a}\} \qquad \mathcal{C}^2_{\mathbf{b}} \stackrel{\text{def.}}{=} \left\{\mathbf{P} \;:\; \mathbf{P}^{\mathrm{T}}\mathbb{1}_m = \mathbf{b}\right\}$$

[Bregman, 1967]

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

Theorem: For affine $(\mathcal{C}^1_{\mathbf{a}}, \mathcal{C}^2_{\mathbf{b}})$, $\quad \mathbf{P}^{(\ell)} \to \mathbf{P}^\star = \underset{\mathbf{P}\in\mathcal{C}^1_{\mathbf{a}}\cap\mathcal{C}^1_{\mathbf{b}}}{\mathrm{argmin}}\ \mathrm{KL}(\mathbf{P}|\mathbf{K})$



Sinkhorn $\iff$ iterative projections.

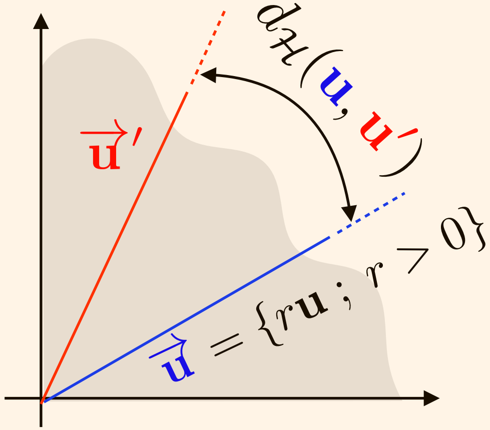
$$\mathbf{P}^{(2\ell)} \stackrel{\text{def.}}{=} \mathrm{diag}(\mathbf{u}^{(\ell)})\mathbf{K}\,\mathrm{diag}(\mathbf{v}^{(\ell)}), \qquad \mathbf{P}^{(2\ell+1)} \stackrel{\text{def.}}{=} \mathrm{diag}(\mathbf{u}^{(\ell+1)})\mathbf{K}\,\mathrm{diag}(\mathbf{v}^{(\ell)})$$

Hilbert Projective Metric

Hilbert's projective metric:

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

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

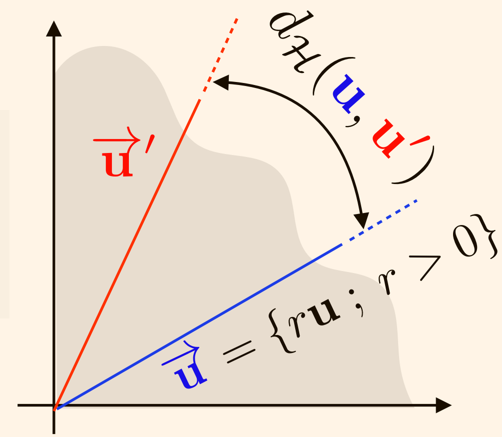


Hilbert Projective Metric

Hilbert's projective metric:

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

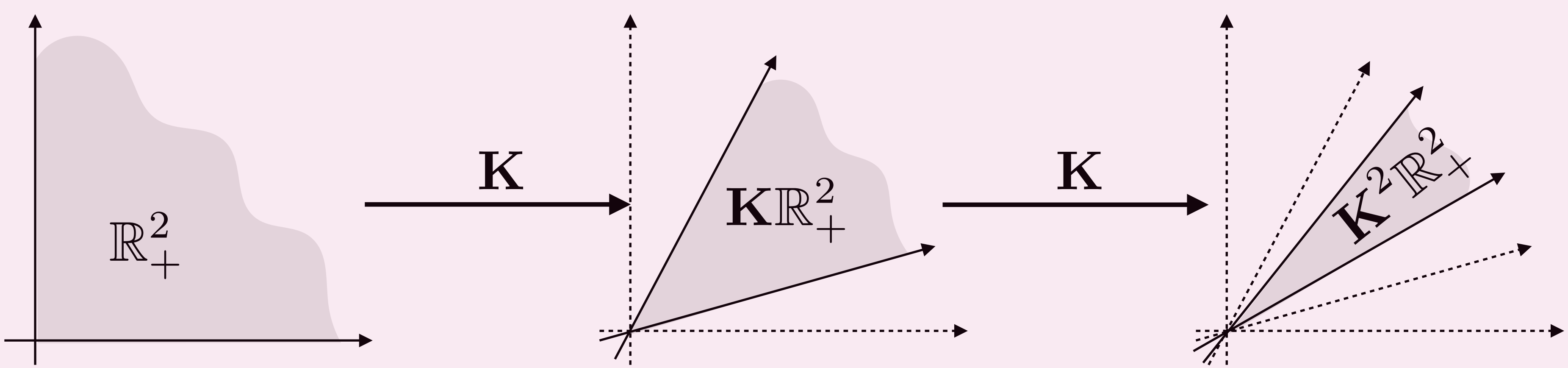
$d_{\mathcal{H}}$ is a distance on the set of rays $\vec{\mathbf{u}}$.



Birkhoff's contraction theorem:

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

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



Perron Frobenius

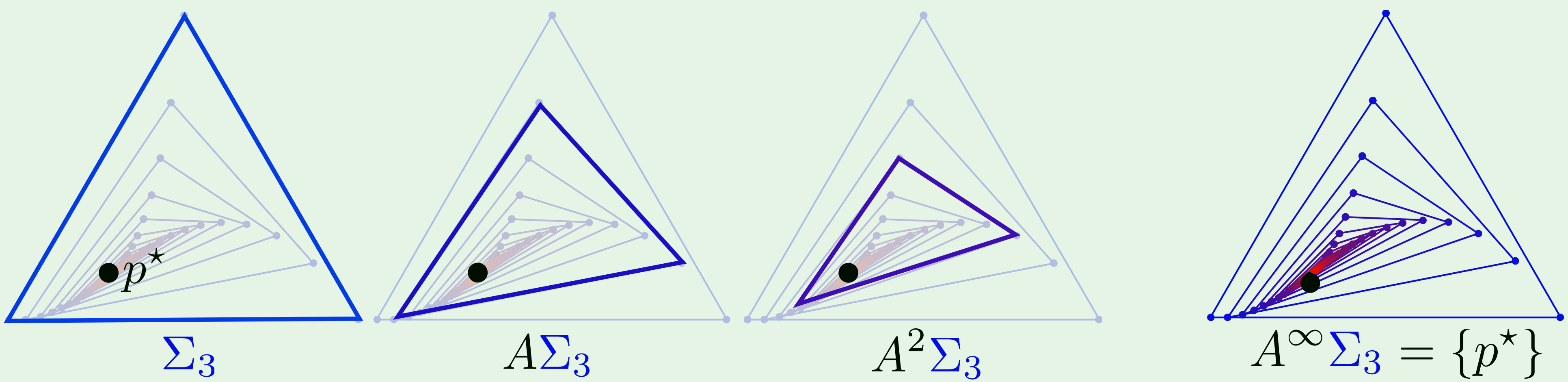
Simplex: $\Sigma_k = \{p \in \mathbb{R}_+^k \; ; \; \sum_i p_i = 1\}$

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

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

[Franklin and Lorenz, 1989]

Theorem: One has $(\mathbf{u}^{(\ell)}, \mathbf{v}^{(\ell)}) \to (\mathbf{u}^\star, \mathbf{v}^\star)$

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

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

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

Local Analysis of Sinkhorn

Sinkhorn fixed point: $\mathbf{f}^{(\ell+1)} = \Phi(\mathbf{f}^{(\ell)})$

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

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

For ℓ large enough, $\|\mathbf{f}^{(\ell)} - \mathbf{f}\| = O((1-\kappa)^{\ell})$

Local Analysis of Sinkhorn

Sinkhorn fixed point: $\mathbf{f}^{(\ell+1)} = \Phi(\mathbf{f}^{(\ell)})$

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

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

For ℓ large enough, $\|\mathbf{f}^{(\ell)} - \mathbf{f}\| = O((1-\kappa)^{\ell})$

Global rate: $\kappa \sim e^{-\frac{1}{\varepsilon}}$
[Franklin and Lorenz, 1989]

$\longleftrightarrow$

Local rate: $\kappa \sim \varepsilon$
[Robert Berman 2017]

Local Analysis of Sinkhorn

Sinkhorn fixed point: $\mathbf{f}^{(\ell+1)} = \Phi(\mathbf{f}^{(\ell)})$

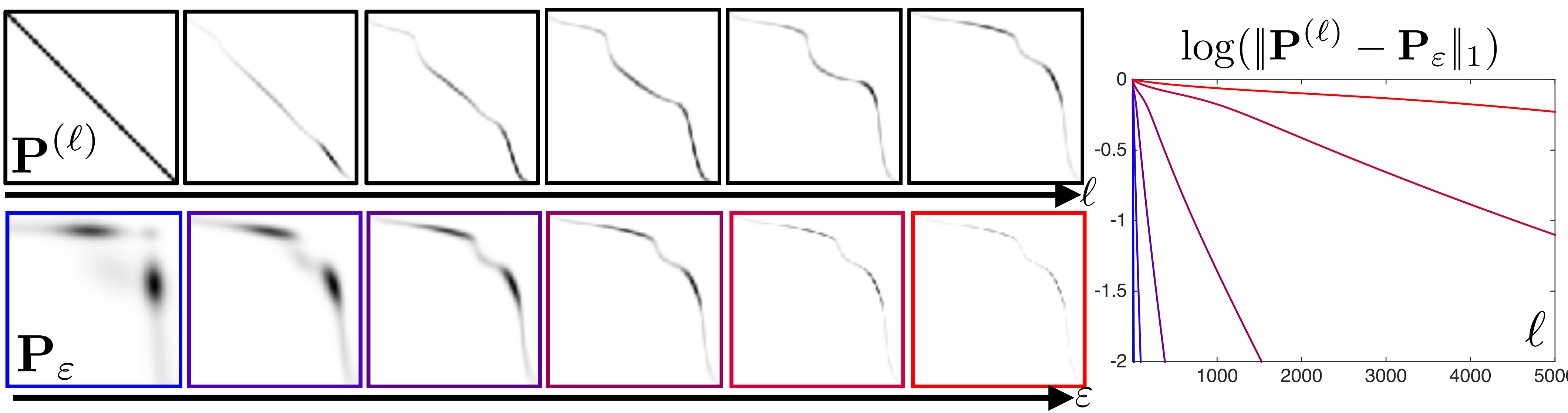
$$\Phi \overset{\text{def.}}{=} \Phi_2 \odot \Phi_1 \quad \text{where} \quad \begin{cases} \Phi_1(\mathbf{f}) = \varepsilon \log \mathbf{K}^{\mathrm{T}}(e^{\mathbf{f}/\varepsilon}) - \log(\mathbf{b}), \\ \Phi_2(\mathbf{g}) = \varepsilon \log \mathbf{K}(e^{\mathbf{g}/\varepsilon}) - \log(\mathbf{a}).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

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

For ℓ large enough, $\|\mathbf{f}^{(\ell)} - \mathbf{f}\| = O((1-\kappa)^{\ell})$

Global rate: $\kappa \sim e^{-\frac{1}{\varepsilon}}$
[Franklin and Lorenz, 1989]

Local rate: $\kappa \sim \varepsilon$
[Robert Berman 2017]



Overview

- Entropic Regularization and Sinkhorn
- Convergence Analysis
- **Barycenters**

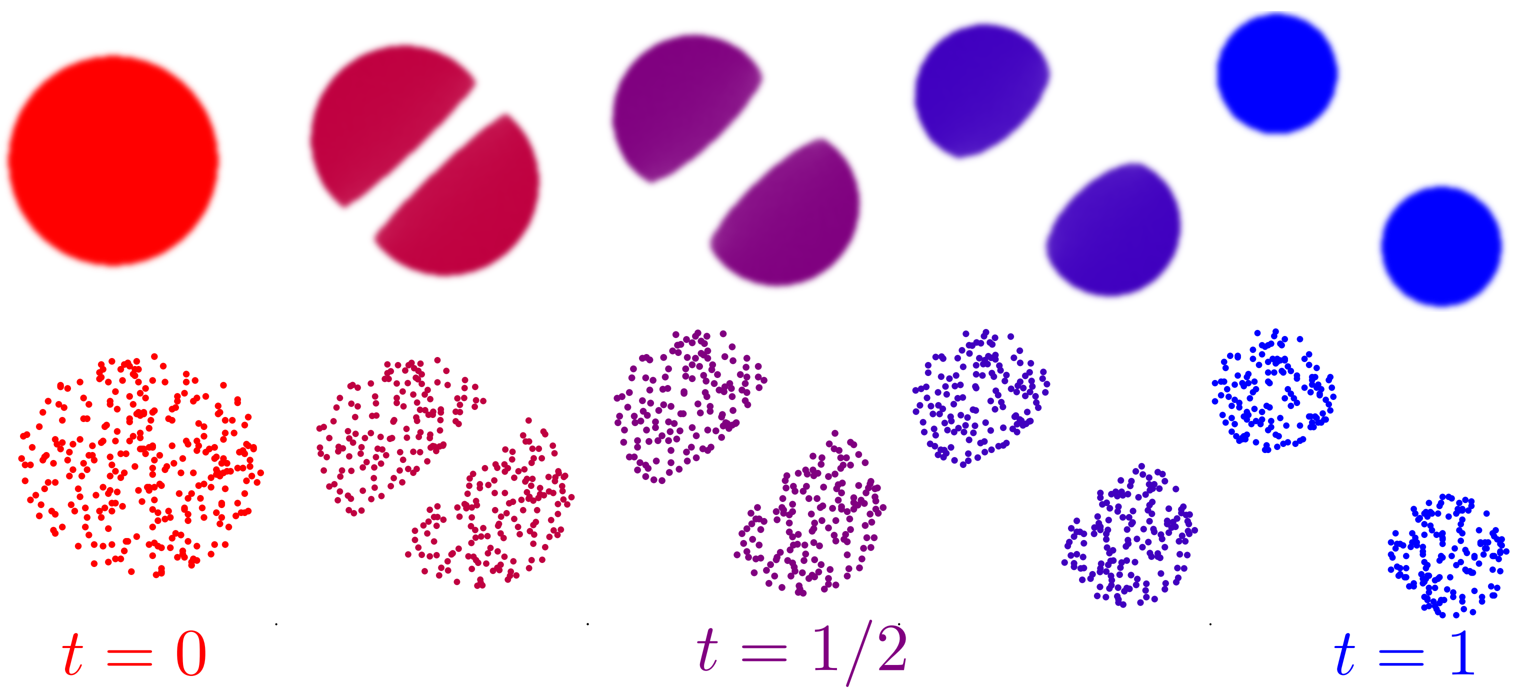
Displacement Interpolation

For $c(x,y) = \|x-y\|^2$, displacement interpolation geodesic:

$$\alpha_t \in \underset{\alpha}{\operatorname{argmin}} \; (1-t)\mathrm{W}_2^2(\alpha, \alpha_0) + t\mathrm{W}_2^2(\alpha, \alpha_1)$$

Optimal coupling $\pi \in \mathcal{U}(\alpha_0, \alpha_1)$:

$$\alpha_t \overset{\text{def.}}{=} ((1-t)P_{\mathcal{X}} + tP_{\mathcal{Y}})_\sharp \pi$$



Displacement Interpolation

For $c(x,y) = \|x-y\|^2$, displacement interpolation geodesic:

$$\alpha_t \in \underset{\alpha}{\operatorname{argmin}}\ (1-t)\mathrm{W}_2^2(\alpha,\alpha_0) + t\mathrm{W}_2^2(\alpha,\alpha_1)$$

Optimal coupling $\pi \in \mathcal{U}(\alpha_0, \alpha_1)$:

$$\alpha_t \overset{\text{def.}}{=} ((1-t)P_{\mathcal{X}} + tP_{\mathcal{Y}})_\sharp \pi$$

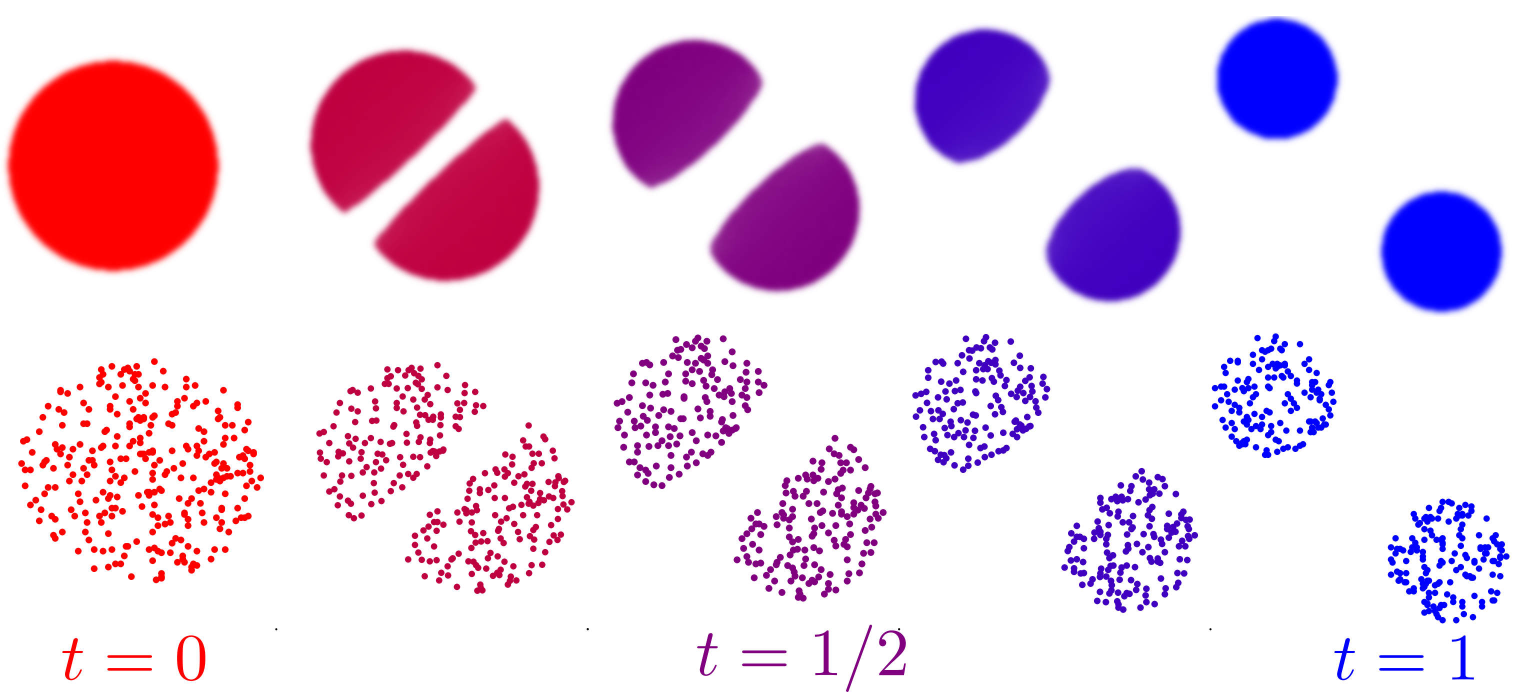


Image Color Palette Equalization

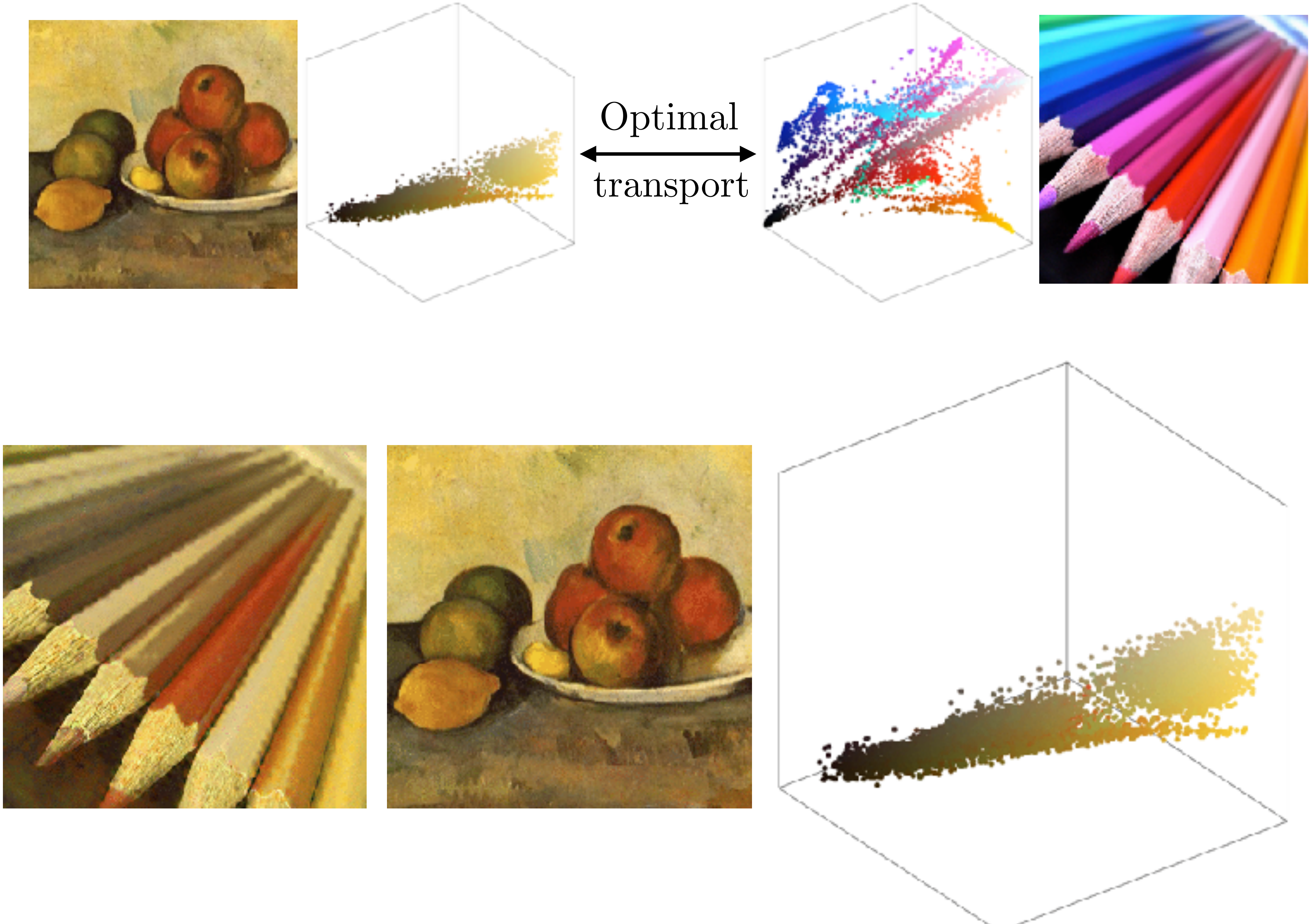
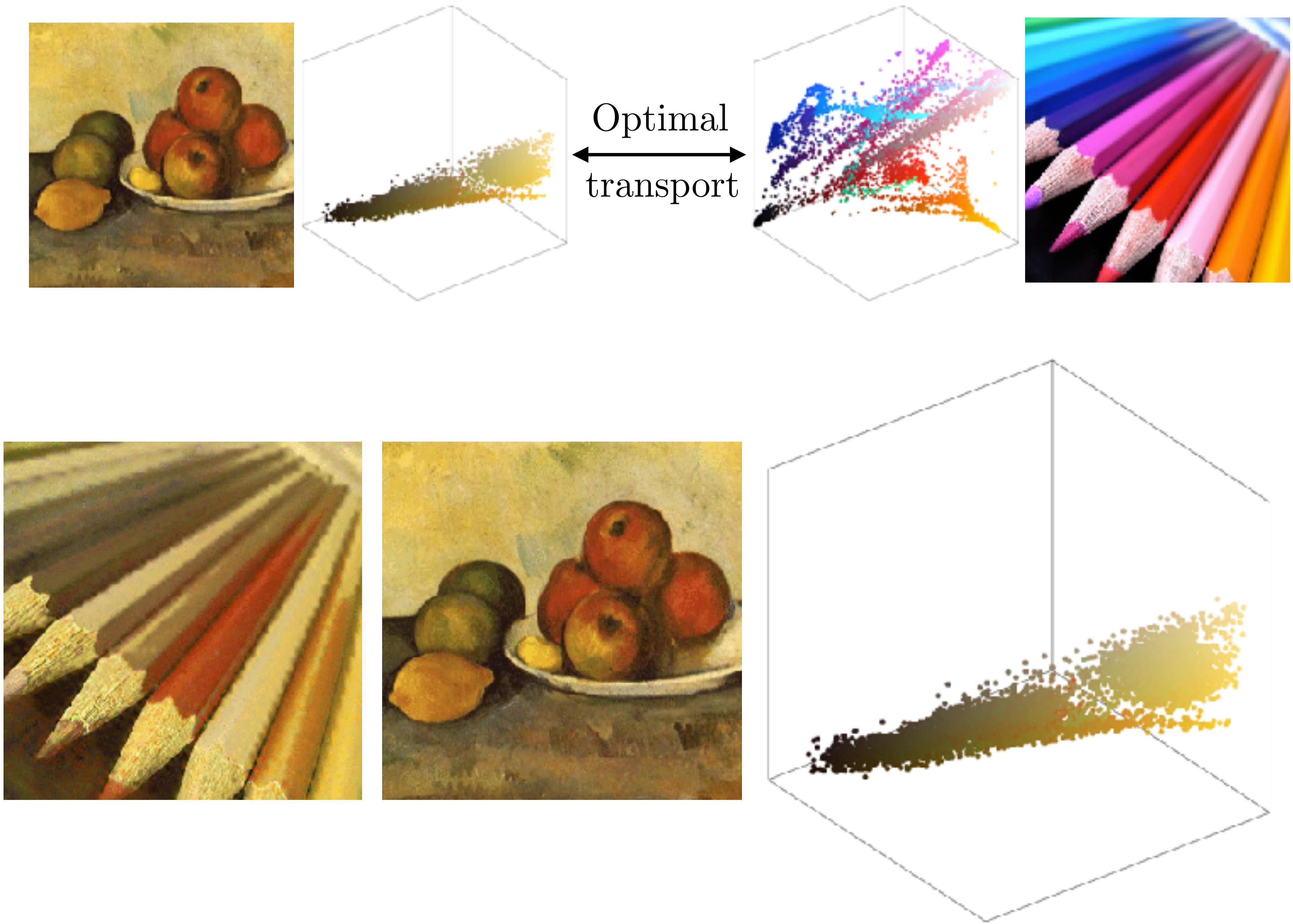


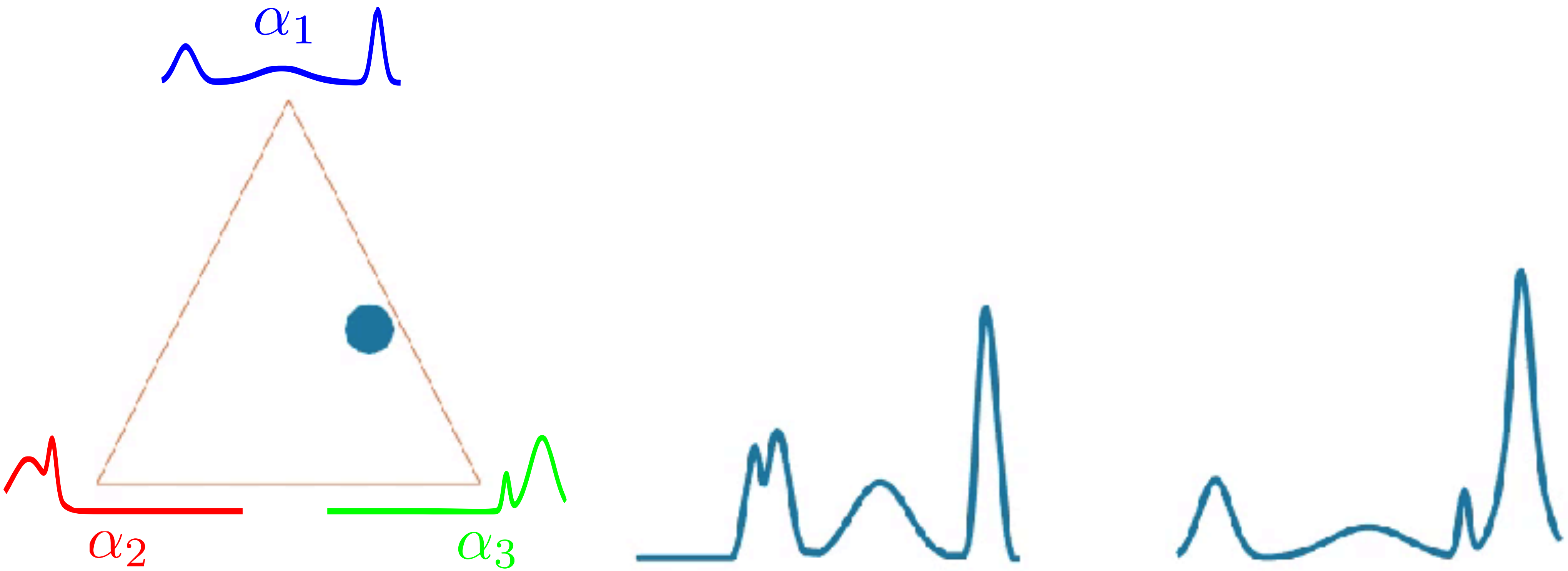
Image Color Palette Equalization



Wasserstein Barycenters

Barycenters of measures $(\alpha_s)_{s=1}^S$: $\quad \sum_s \lambda_s = 1$

$$\alpha^\star \in \underset{\alpha}{\text{argmin}} \; \sum_s \lambda_s \mathrm{W}_p^p(\alpha, \alpha_s)$$



$\lambda \in \Sigma_3$

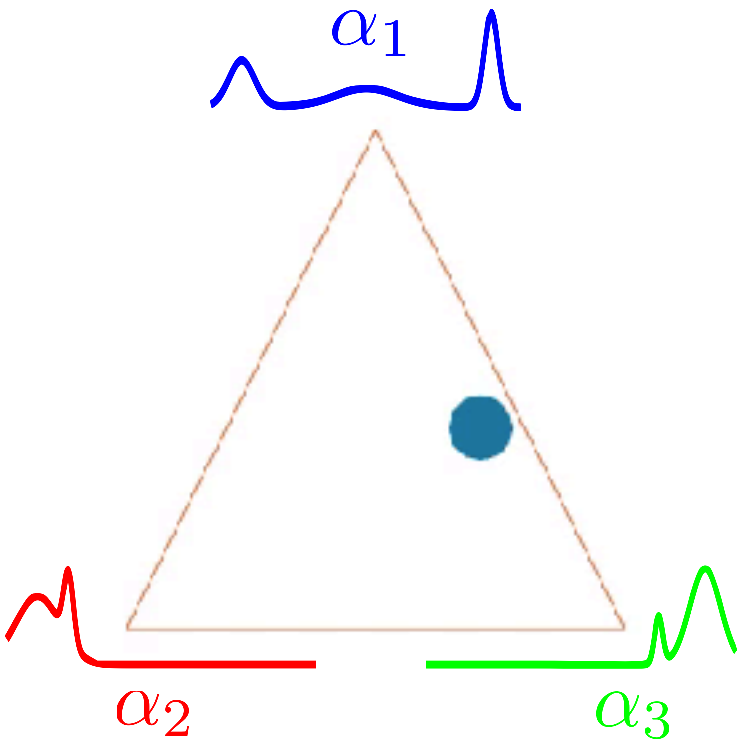
$\min_\alpha \sum_s \lambda_s \mathrm{W}_p^p(\alpha, \alpha_s)$
Wasserstein

$\sum_s \lambda_s \alpha_s$
Euclidean

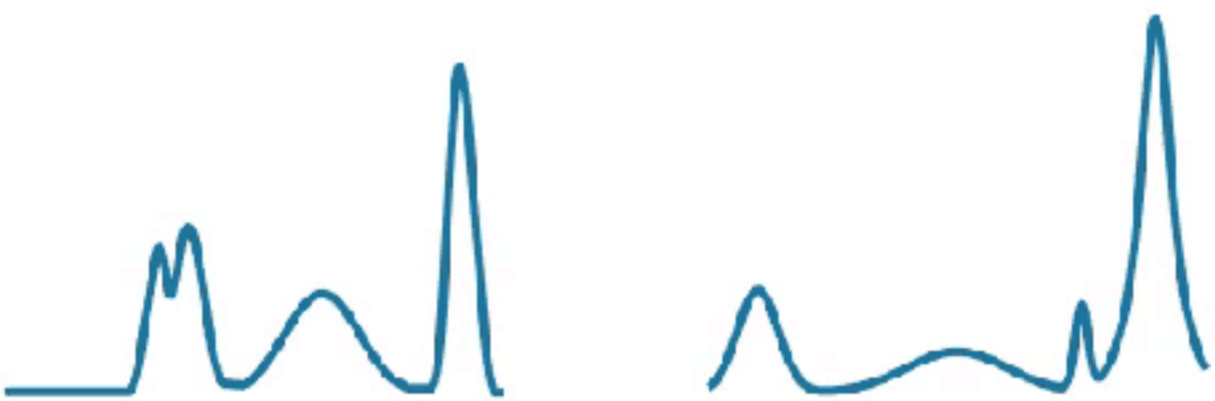
Wasserstein Barycenters

Barycenters of measures $(\alpha_s)_{s=1}^{S}$: $\quad \sum_s \lambda_s = 1$

$$\alpha^\star \in \underset{\alpha}{\operatorname{argmin}} \ \sum_s \lambda_s \mathrm{W}_p^p(\alpha, \alpha_s)$$



$\lambda \in \Sigma_3$



$\min_\alpha \sum_s \lambda_s \mathrm{W}_p^p(\alpha, \alpha_s)$

Wasserstein

$\sum_s \lambda_s \alpha_s$

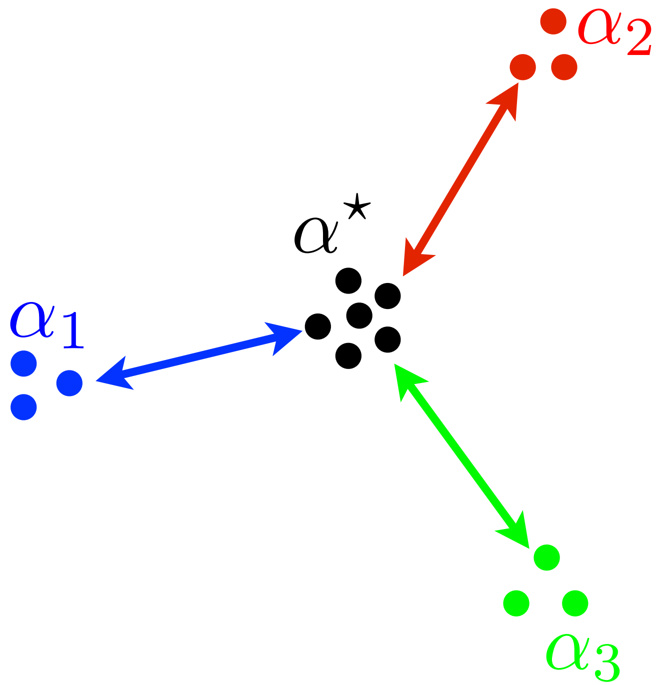
Euclidean

Wasserstein Barycenters

Generalizes usual barycenter:

If $\alpha_s = \delta_{x_s}$ then $\alpha^\star = \delta_{x^\star}$

$$x^\star \in \underset{x}{\operatorname{argmin}} \sum_s \lambda_s d(x, x_s)^p$$



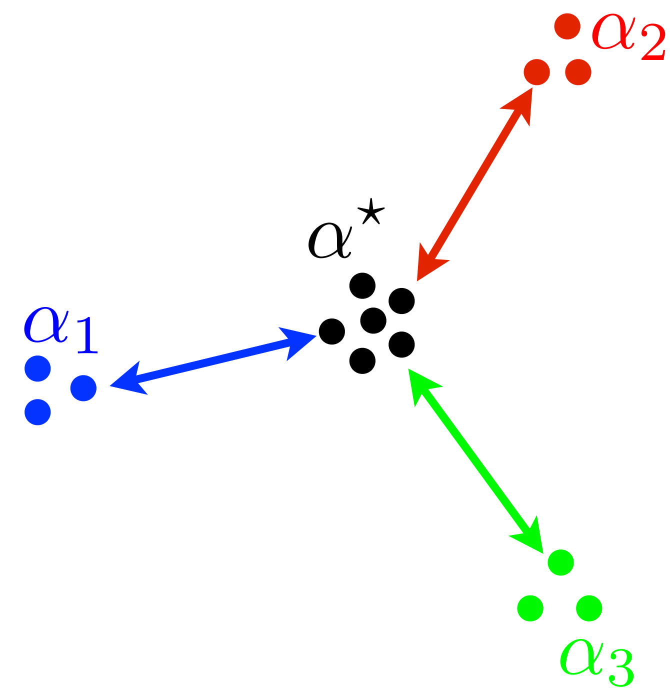
Wasserstein Barycenters

Generalizes usual barycenter:

If $\alpha_s = \delta_{x_s}$ then $\alpha^\star = \delta_{x^\star}$

$$x^\star \in \underset{x}{\operatorname{argmin}} \sum_s \lambda_s d(x, x_s)^p$$

If $|\,\mathrm{supp}(\alpha_s)| \leqslant n$, then $|\,\mathrm{supp}(\alpha^\star)| \leqslant Sn - 1$.



Wasserstein Barycenters

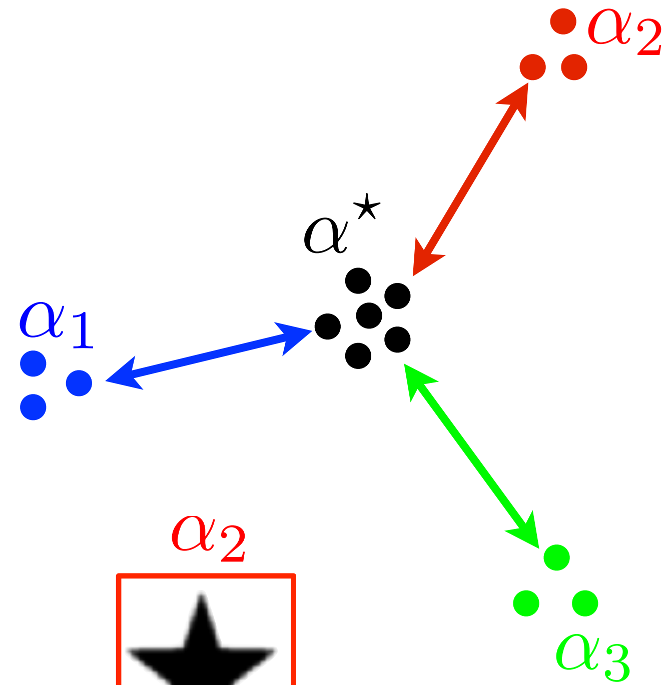
Generalizes usual barycenter:

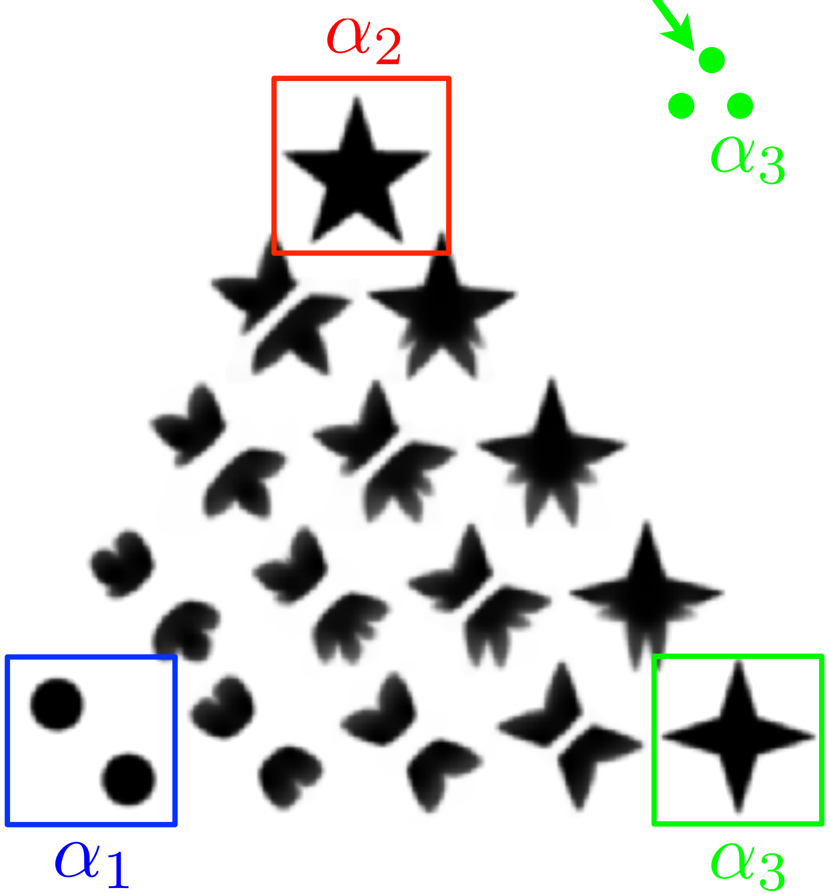
If $\alpha_s = \delta_{x_s}$ then $\alpha^\star = \delta_{x^\star}$

$$x^\star \in \underset{x}{\operatorname{argmin}} \sum_s \lambda_s d(x, x_s)^p$$

If $|\operatorname{supp}(\alpha_s)| \leqslant n$, then
$|\operatorname{supp}(\alpha^\star)| \leqslant Sn - 1$.

Theorem: [Agueh, Carlier, 2010]
(for $c(x,y) = \|x-y\|^2$)
if α_1 has a density,
$\alpha^\star$ exists and is unique.





Barycenters of Gaussians

OT between $\alpha = \mathcal{N}(\mathbf{m}_\alpha, \Sigma_\alpha)$ $\beta = \mathcal{N}(\mathbf{m}_\beta, \Sigma_\beta)$

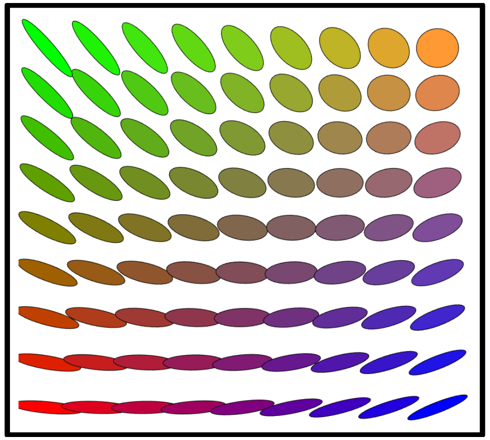
$$\mathcal{W}_2^2(\alpha, \beta) = \|\mathbf{m}_\alpha - \mathbf{m}_\beta\|^2 + \mathcal{B}(\mathbf{\Sigma}_\alpha, \mathbf{\Sigma}_\beta)^2$$

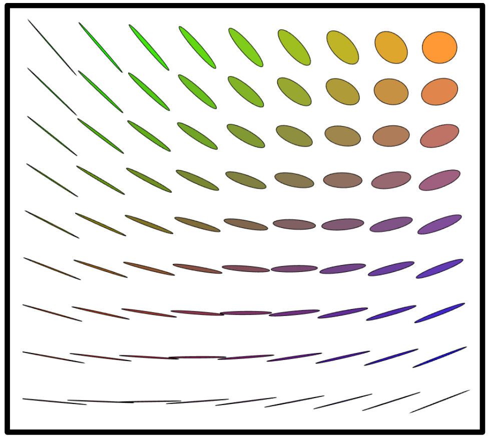
$$\mathcal{B}(\mathbf{\Sigma}_\alpha, \mathbf{\Sigma}_\beta)^2 \stackrel{\text{def.}}{=} \operatorname{tr}\left(\mathbf{\Sigma}_\alpha + \mathbf{\Sigma}_\beta - 2(\mathbf{\Sigma}_\alpha^{1/2}\mathbf{\Sigma}_\beta\mathbf{\Sigma}_\alpha^{1/2})^{1/2}\right)$$

Proposition: barycenter is $\mathcal{N}(\mathbf{m}^\star, \Sigma^\star)$ and solves

[Carlier, Agueh, 2011]

$$\mathbf{m}^\star = \sum_s \lambda_s \mathbf{m}_s \qquad \min_{\mathbf{\Sigma}} \sum_s \lambda_s \mathcal{B}(\mathbf{\Sigma}, \mathbf{\Sigma}_s)^2$$





Barycenters of Gaussians

OT between $\alpha = \mathcal{N}(\mathbf{m}_\alpha, \Sigma_\alpha) \quad \beta = \mathcal{N}(\mathbf{m}_\beta, \Sigma_\beta)$

$$\mathcal{W}_2^2(\alpha,\beta) = \|\mathbf{m}_\alpha - \mathbf{m}_\beta\|^2 + \mathcal{B}(\mathbf{\Sigma}_\alpha, \mathbf{\Sigma}_\beta)^2$$

$$\mathcal{B}(\mathbf{\Sigma}_\alpha, \mathbf{\Sigma}_\beta)^2 \overset{\text{def.}}{=} \operatorname{tr}\left(\mathbf{\Sigma}_\alpha + \mathbf{\Sigma}_\beta - 2(\mathbf{\Sigma}_\alpha^{1/2}\mathbf{\Sigma}_\beta\mathbf{\Sigma}_\alpha^{1/2})^{1/2}\right)$$

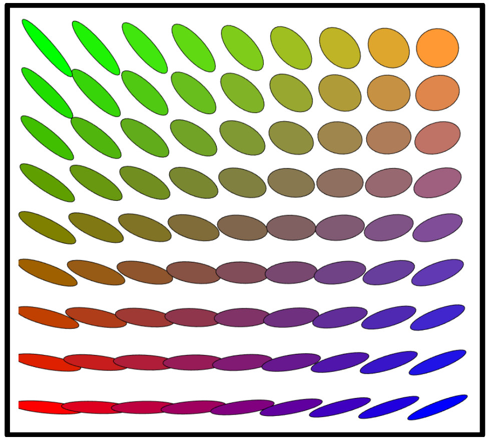
Proposition: barycenter is $\mathcal{N}(\mathbf{m}^\star, \Sigma^\star)$ and solves

[Carlier, Agueh, 2011]

$$\mathbf{m}^\star = \sum_s \lambda_s \mathbf{m}_s \qquad \min_{\mathbf{\Sigma}} \sum_s \lambda_s \mathcal{B}(\mathbf{\Sigma}, \mathbf{\Sigma}_s)^2$$

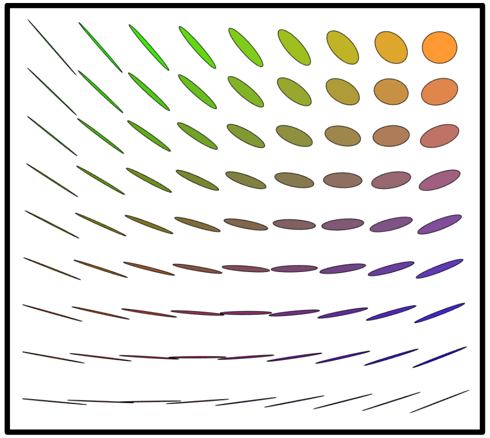
Fixed point equation:

$$\mathbf{\Sigma}^\star = \Psi(\mathbf{\Sigma}^\star) \quad \text{where} \quad \Psi(\mathbf{\Sigma}) \overset{\text{def.}}{=} \sum_s \lambda_s (\mathbf{\Sigma}^{\frac{1}{2}} \mathbf{\Sigma}_s \mathbf{\Sigma}^{\frac{1}{2}})^{\frac{1}{2}}$$



$\rightarrow$ empirically convergent.

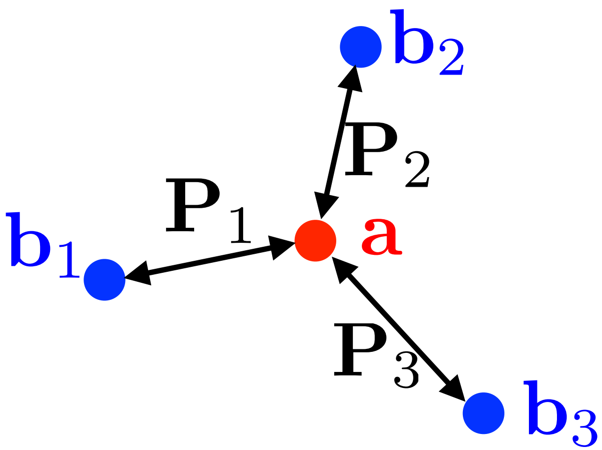
$\rightarrow$ related provably convergent scheme by [Alvarez-Esteban et al 2016].



Regularized Barycenters

$\rightarrow$ Need to fix a grid (Eulerian scheme).

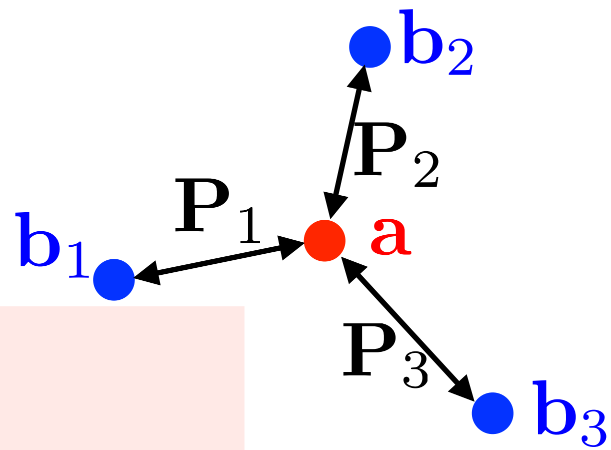
$$\mathrm{L}_{\mathbf{C}}^{\varepsilon}(\mathbf{a}, \mathbf{b}) \overset{\text{def.}}{=} \min_{\mathbf{P} \in \mathbf{U}(\mathbf{a}, \mathbf{b})} \langle \mathbf{P}, \mathbf{C} \rangle - \varepsilon \mathbf{H}(\mathbf{P})$$



Regularized Barycenters

$\rightarrow$ Need to fix a grid (Eulerian scheme).

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



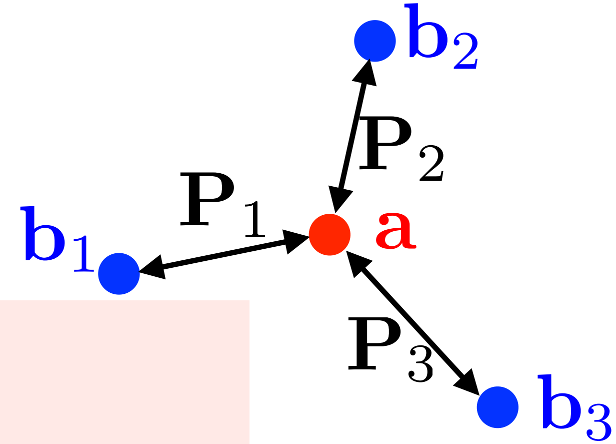
Coupling re-writing: $$\min_{\mathbf{a} \in \Sigma_n} \sum_{s=1}^{S} \lambda_s \mathrm{L}_{\mathbf{C}_s}^{\varepsilon}(\mathbf{a}, \mathbf{b}_s)$$

$$= \min_{(\mathbf{P}_s)_s} \left\{ \sum_s \lambda_s \varepsilon \mathbf{KL}(\mathbf{P}_s | \mathbf{K}_s) \ : \ \forall s, \mathbf{P}_s{}^{\mathrm{T}} \mathbb{1}_m = \mathbf{b}_s, \ \mathbf{P}_1 \mathbb{1}_1 = \ldots = \mathbf{P}_S \mathbb{1}_S \right\}$$

Regularized Barycenters

$\rightarrow$ Need to fix a grid (Eulerian scheme).

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



Coupling re-writing: $$\min_{\mathbf{a}\in\Sigma_n} \sum_{s=1}^{S} \lambda_s \mathrm{L}_{\mathbf{C}_s}^{\varepsilon}(\mathbf{a},\mathbf{b}_s)$$

$$= \min_{(\mathbf{P}_s)_s} \left\{ \sum_s \lambda_s \varepsilon \mathbf{KL}(\mathbf{P}_s|\mathbf{K}_s) \;:\; \forall\, s, \mathbf{P}_s{}^{\mathrm{T}} \mathbb{1}_m = \mathbf{b}_s,\; \mathbf{P}_1 \mathbb{1}_1 = \ldots = \mathbf{P}_S \mathbb{1}_S \right\}$$

Coupling factorization: $\mathbf{P}_s = \mathrm{diag}(\mathbf{u}_s)\mathbf{K}\,\mathrm{diag}(\mathbf{v}_s)$

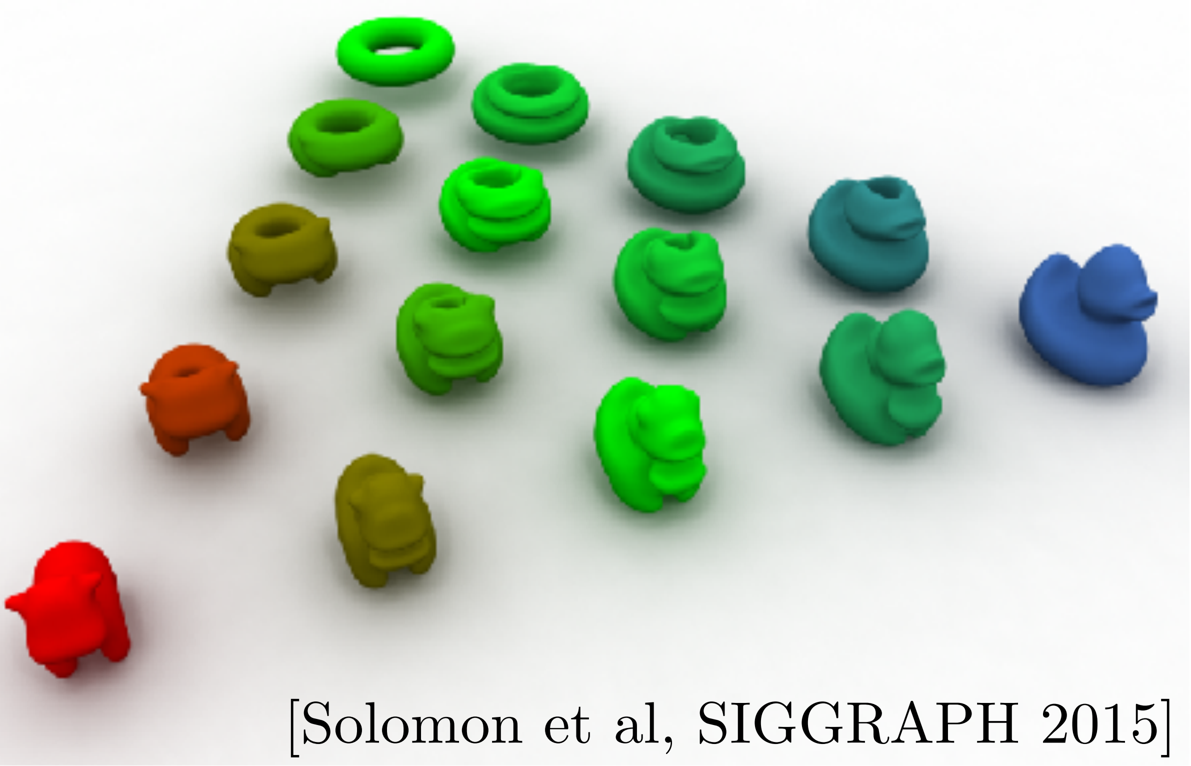
Sinkhorn-like

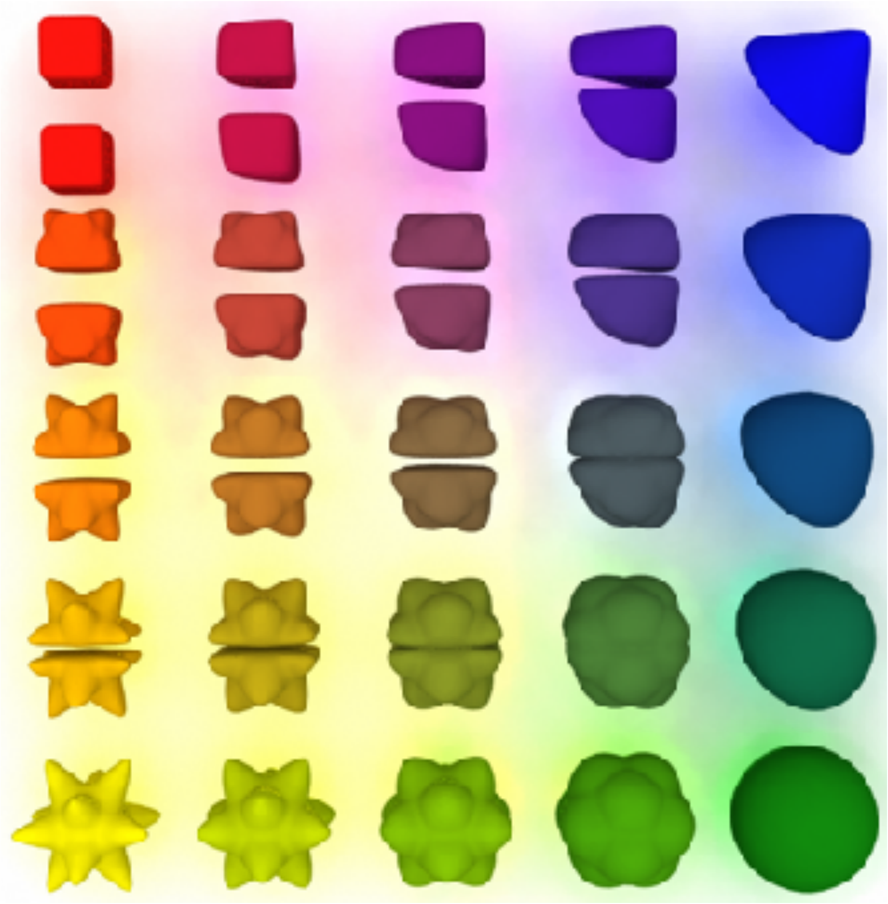
$$\forall\, s \in [\![1, S]\!], \quad \mathbf{v}_s^{(\ell+1)} \overset{\text{def.}}{=} \frac{\mathbf{b}_s}{\mathbf{K}_s^{\mathrm{T}} \mathbf{u}_s^{(\ell)}}$$

$$\mathbf{a}^{(\ell+1)} \overset{\text{def.}}{=} \prod_s (\mathbf{K}_s \mathbf{v}_s^{(\ell+1)})^{\lambda_s}$$

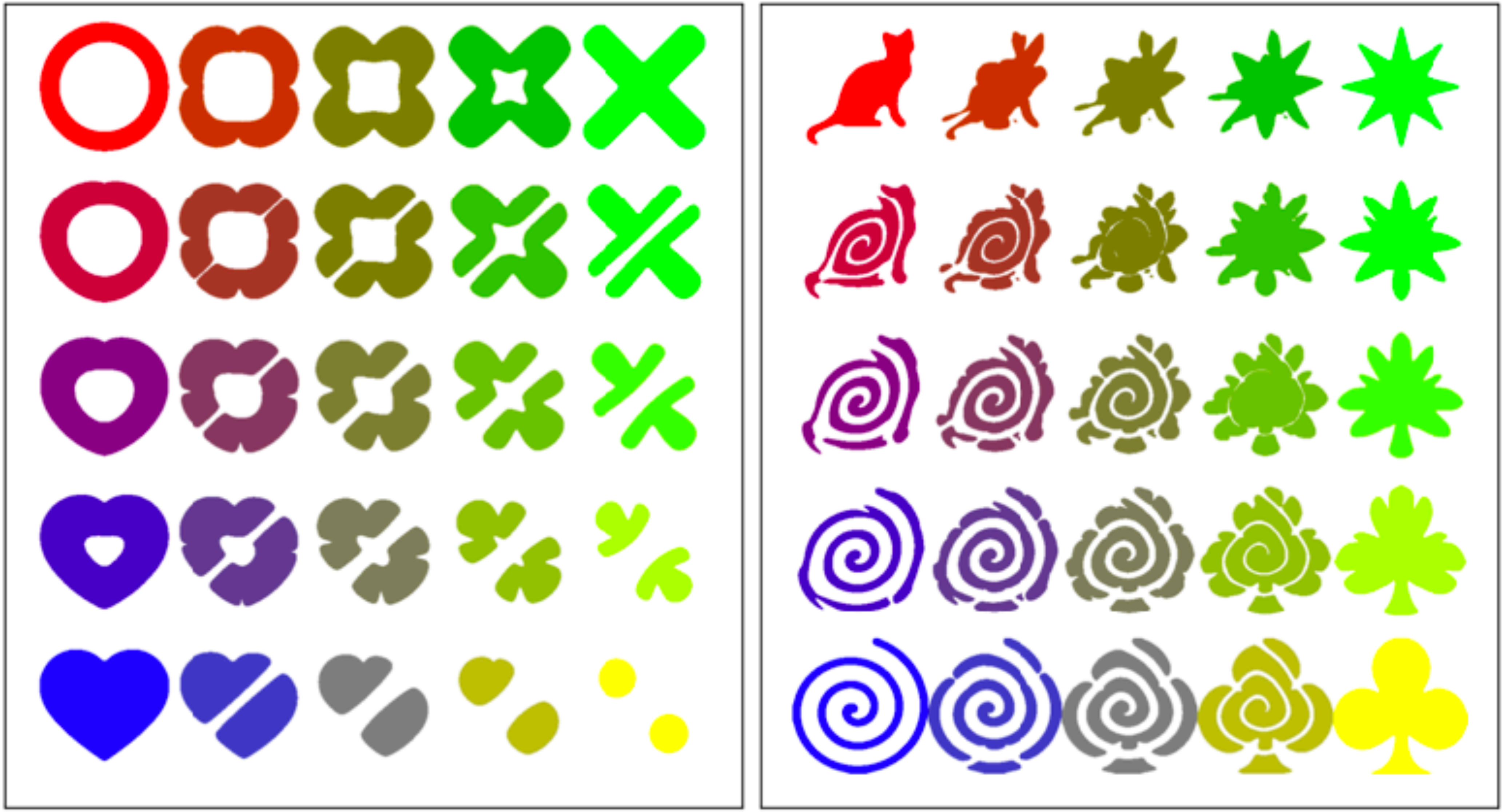
$$\forall\, s \in [\![1, S]\!], \quad \mathbf{u}_s^{(\ell+1)} \overset{\text{def.}}{=} \frac{\mathbf{a}^{(\ell+1)}}{\mathbf{K}_s \mathbf{v}_s^{(\ell+1)}}$$

Examples

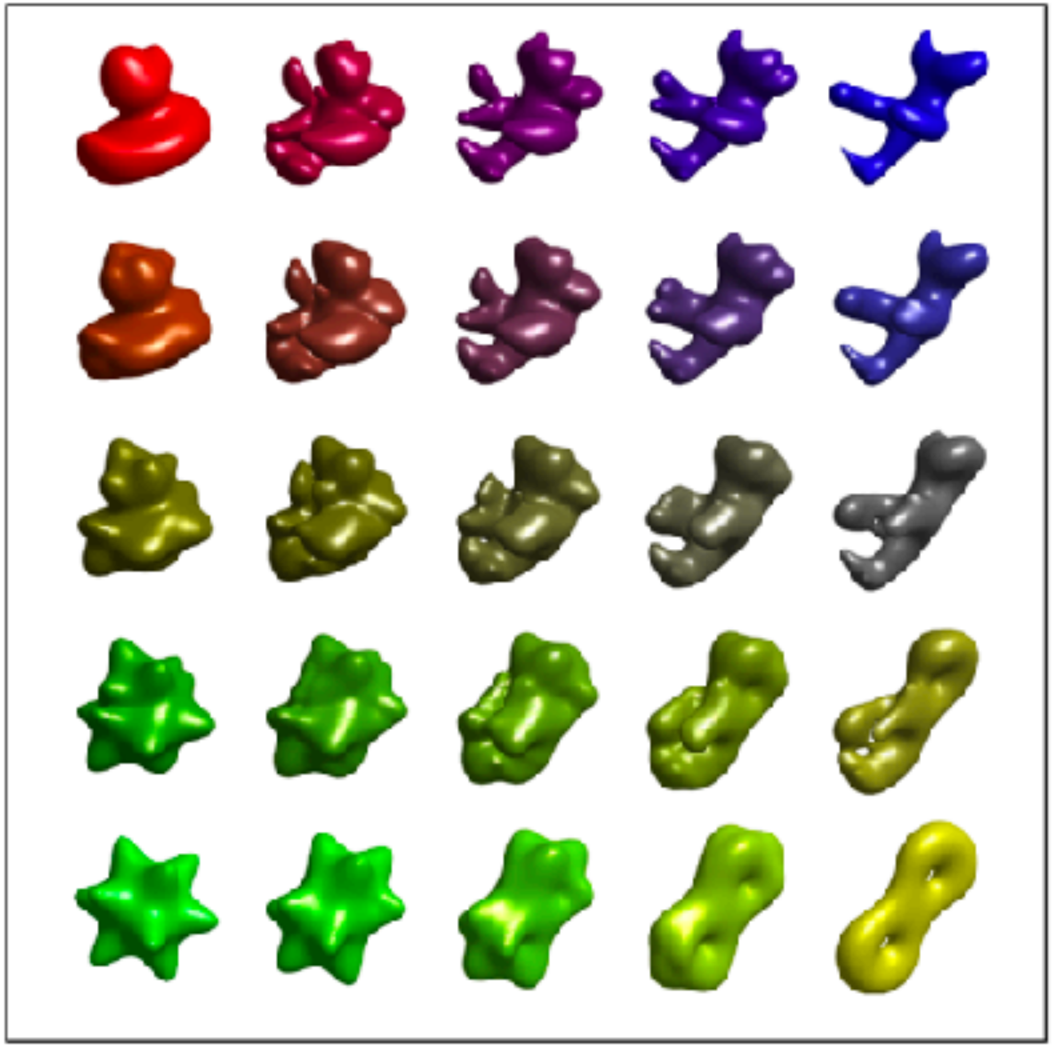


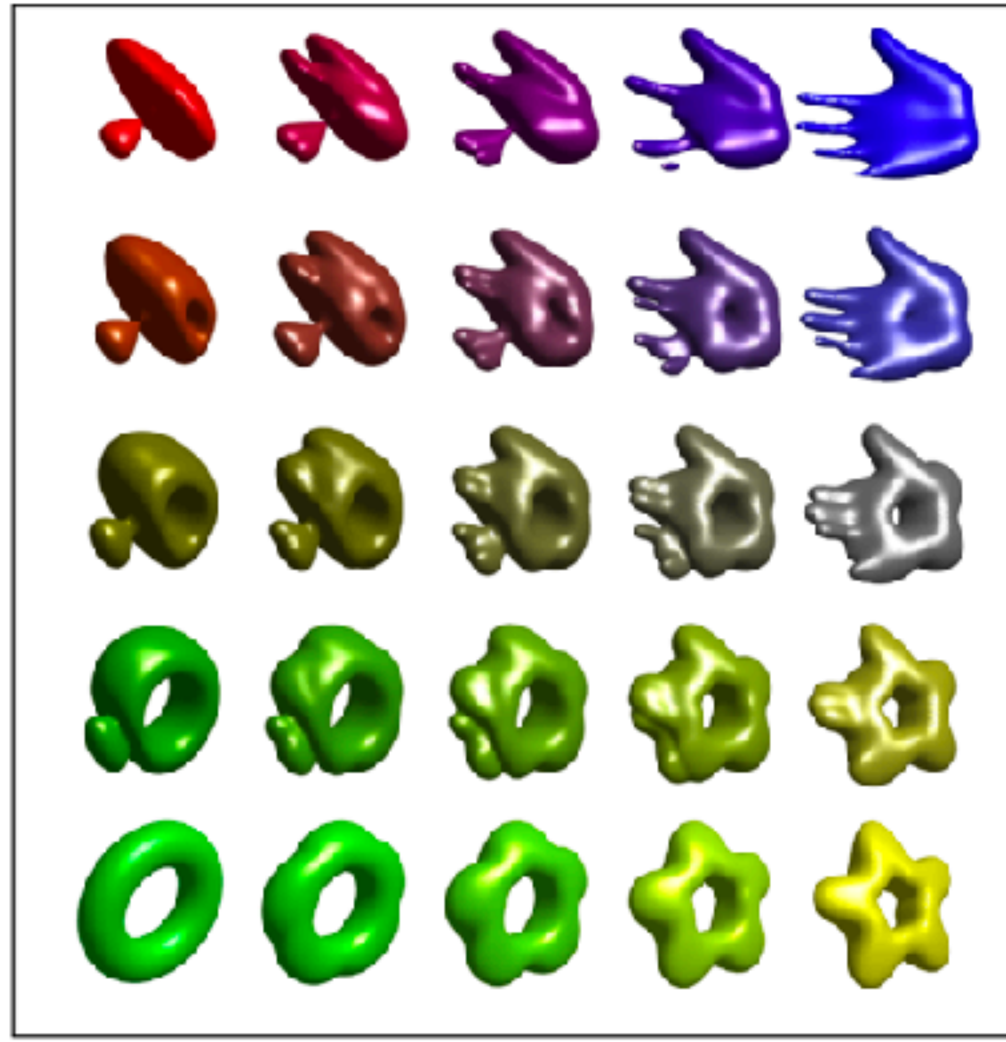


Barycenters of 2D Shapes



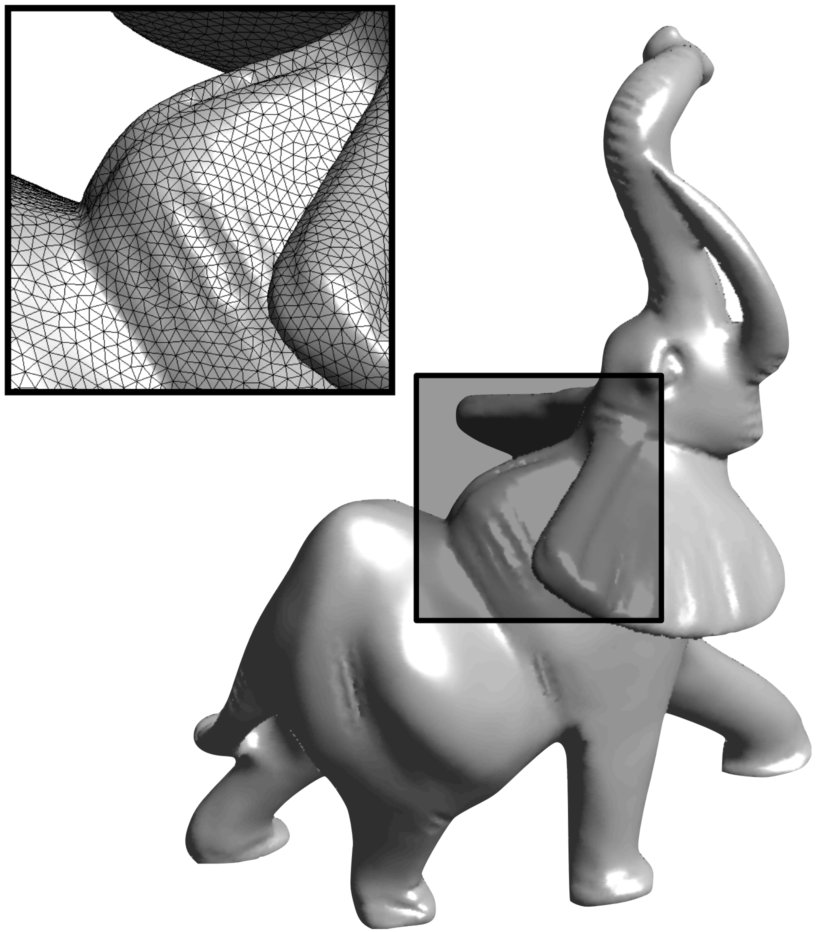
Barycenters of 3D Shapes





Optimal Transport on Surfaces

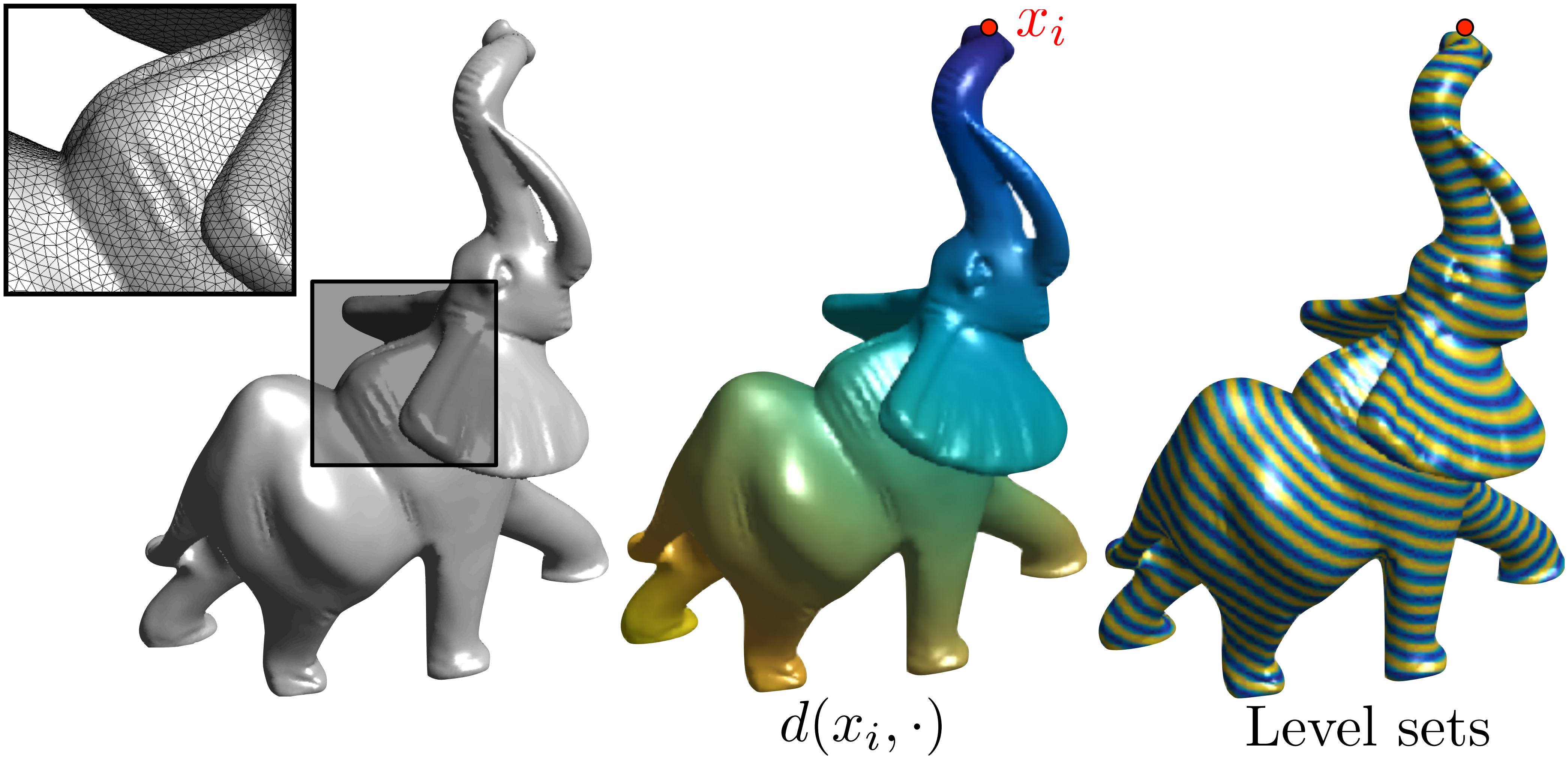
Triangulated mesh M. Geodesic distance d_M.



Optimal Transport on Surfaces

Triangulated mesh M. Geodesic distance d_M.

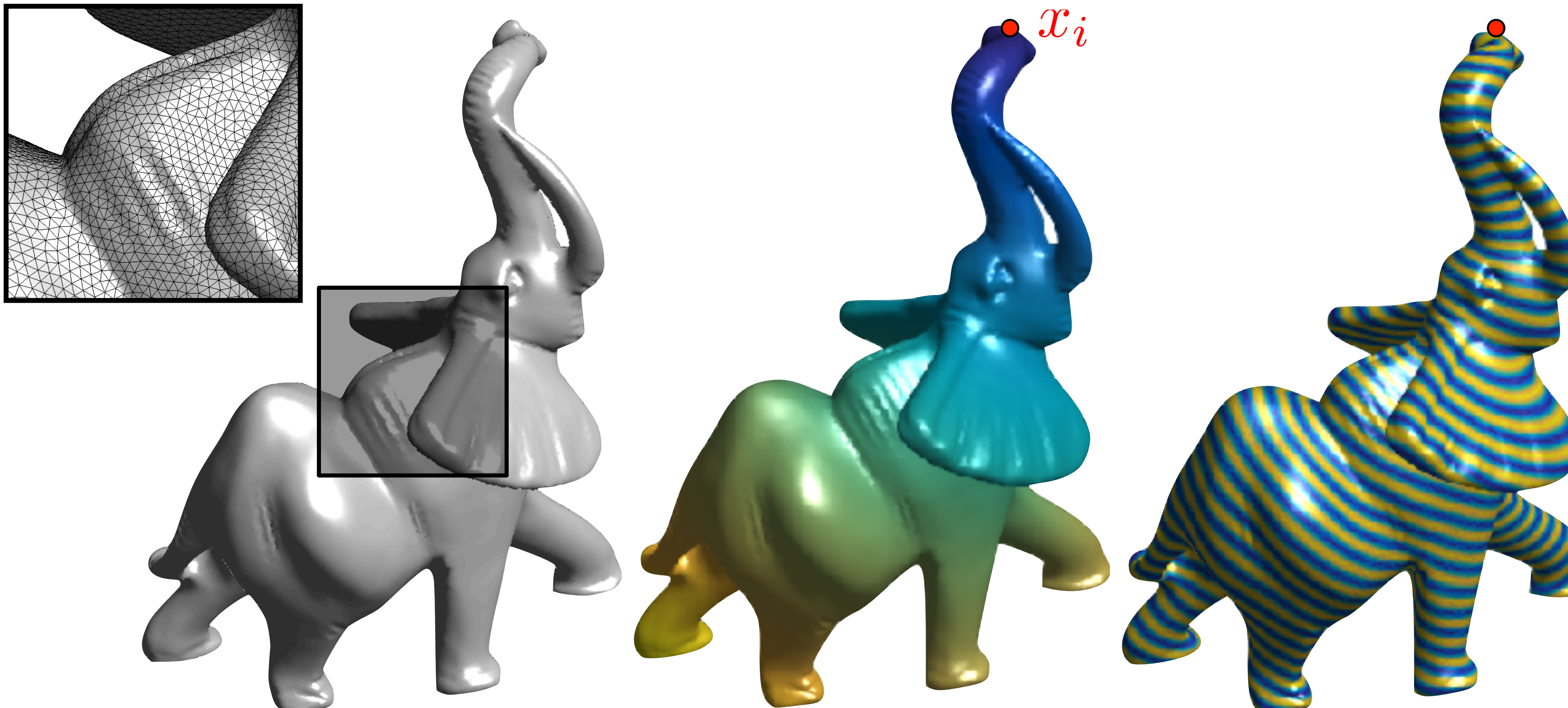
Ground cost: $c(x,y) = d_M(x,y)^\alpha$.



Optimal Transport on Surfaces

Triangulated mesh M. Geodesic distance d_M.

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



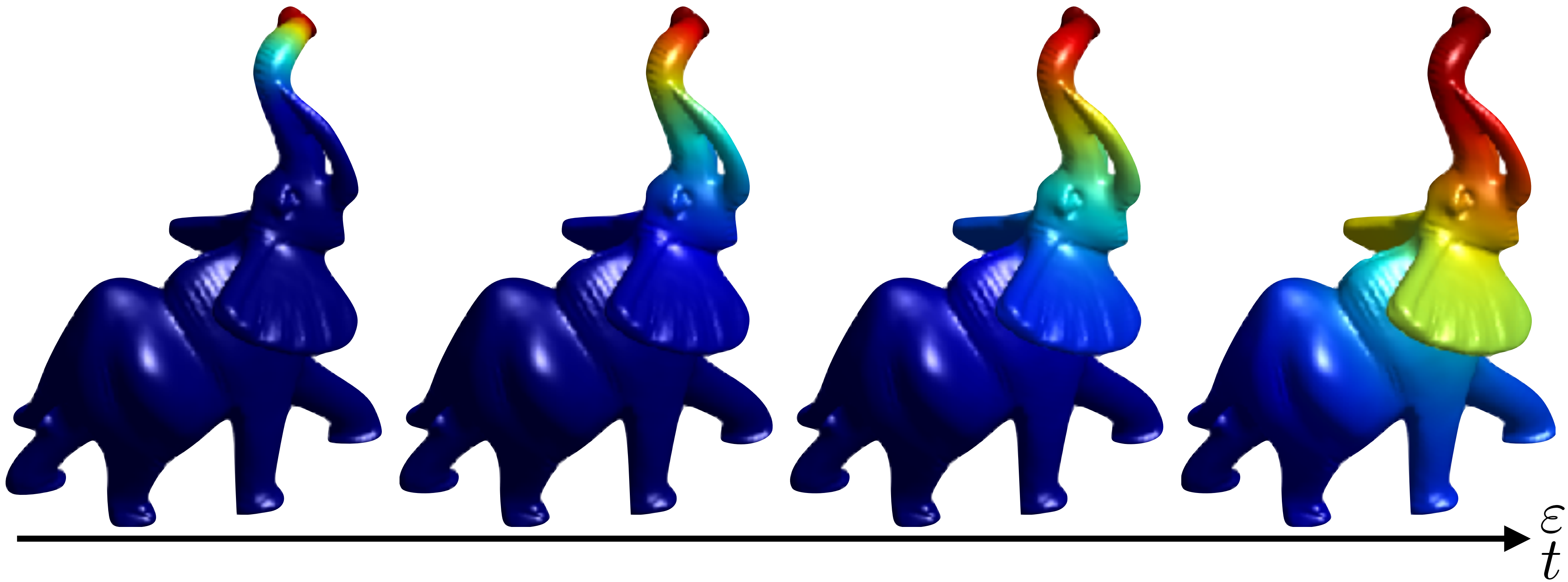
$d(x_i, \cdot)$ Level sets

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

Entropic Transport on Surfaces

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

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

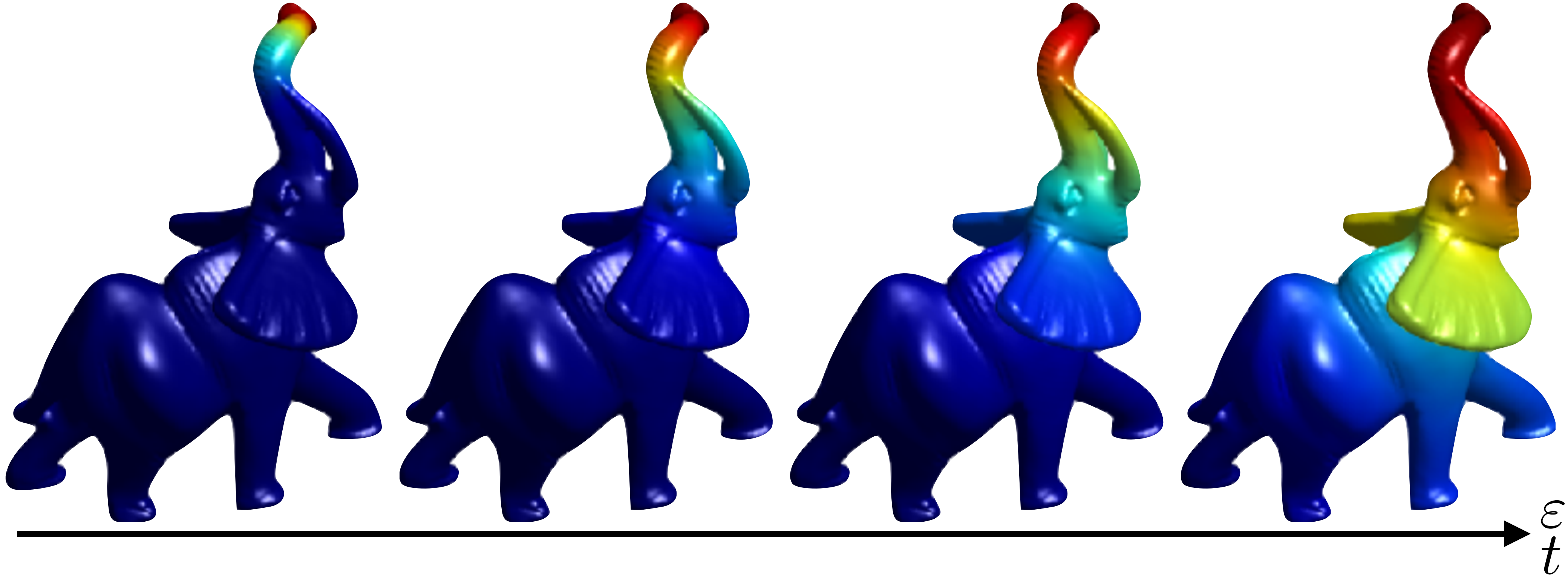


Entropic Transport on Surfaces

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

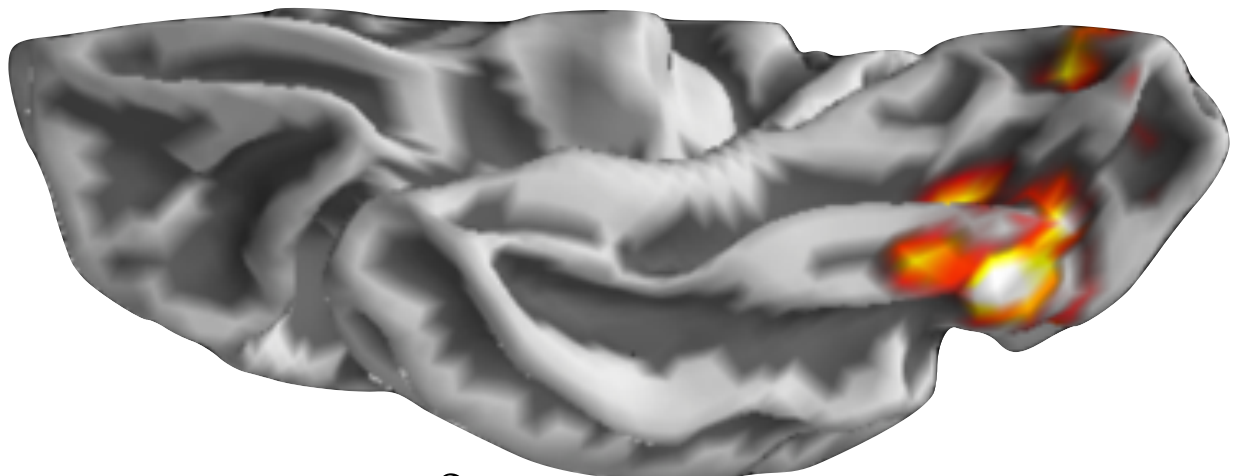
Theorem: [Varadhan] $-\varepsilon \log(u_\varepsilon) \xrightarrow{\varepsilon \to 0} d_M^2$

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

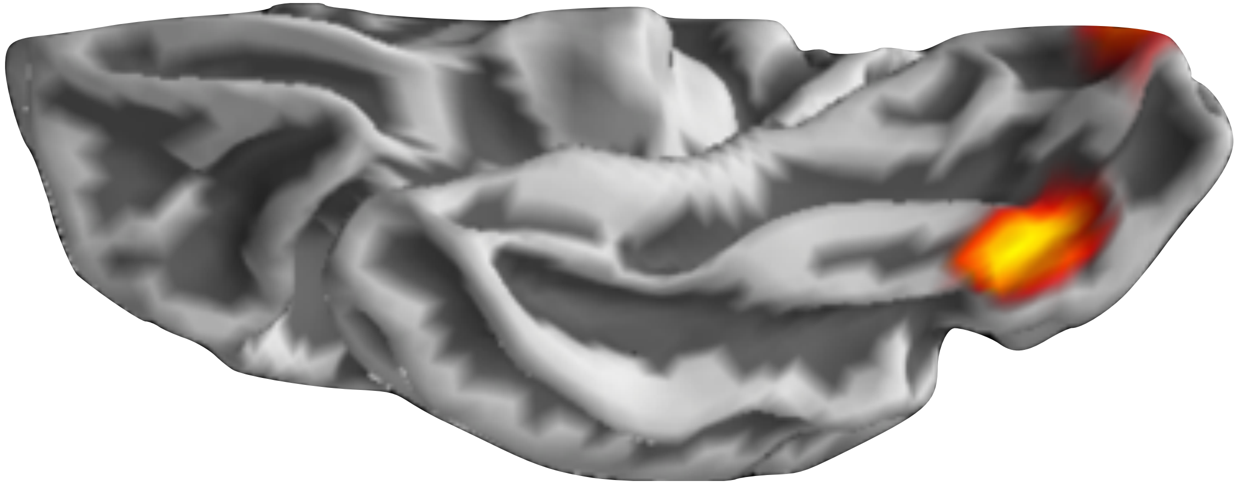


MRI Data Procesing [with A. Gramfort]

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



L^2 barycenter



W_2^2 barycenter

Barycenters on a Surface