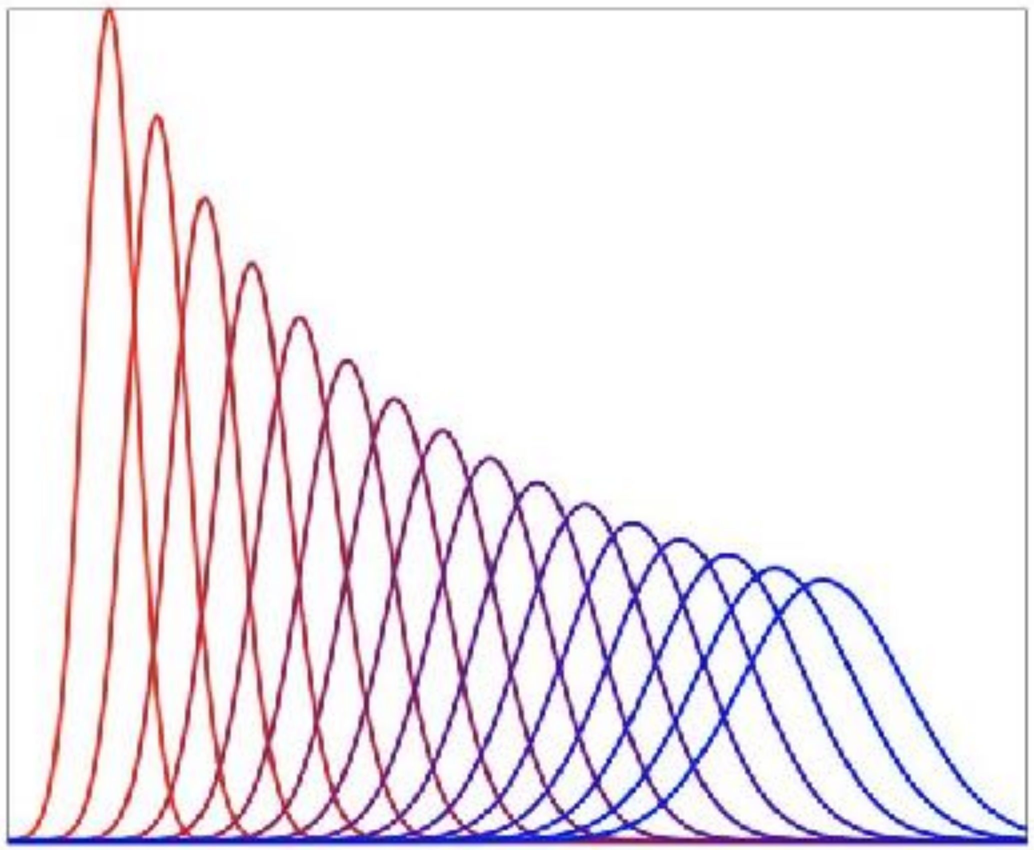
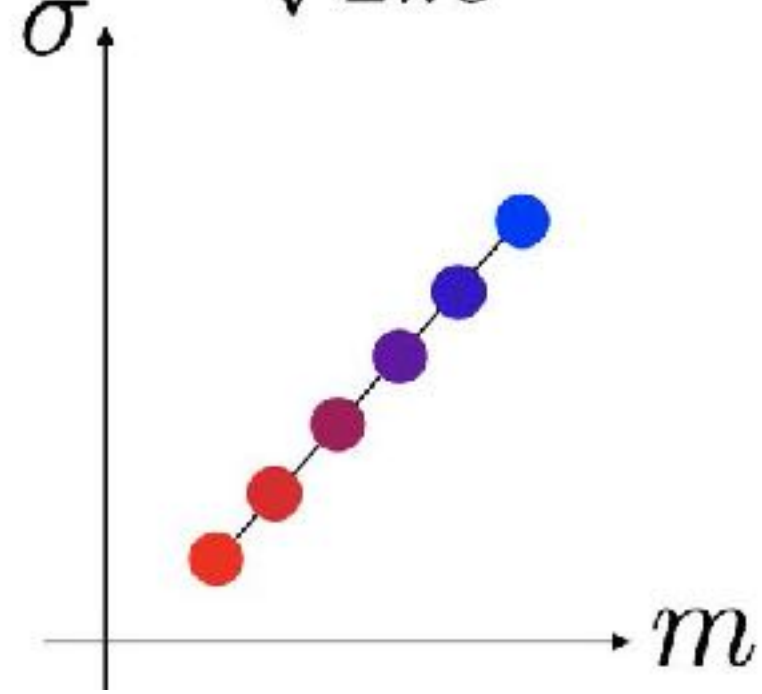
Optimal Transport



$$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$$



Dual norms: (aka Integral Probability Metrics)

$$\|\alpha - \beta\|_B \stackrel{\text{def.}}{=} \max\left\{ \int_{\mathcal{X}} f(x)(\mathrm{d}\alpha(x) - \mathrm{d}\beta(x)) \;;\; f \in B \right\}$$

Wasserstein 1: $B = \{f \;;\; \|\nabla f\|_\infty \leqslant 1\}$.

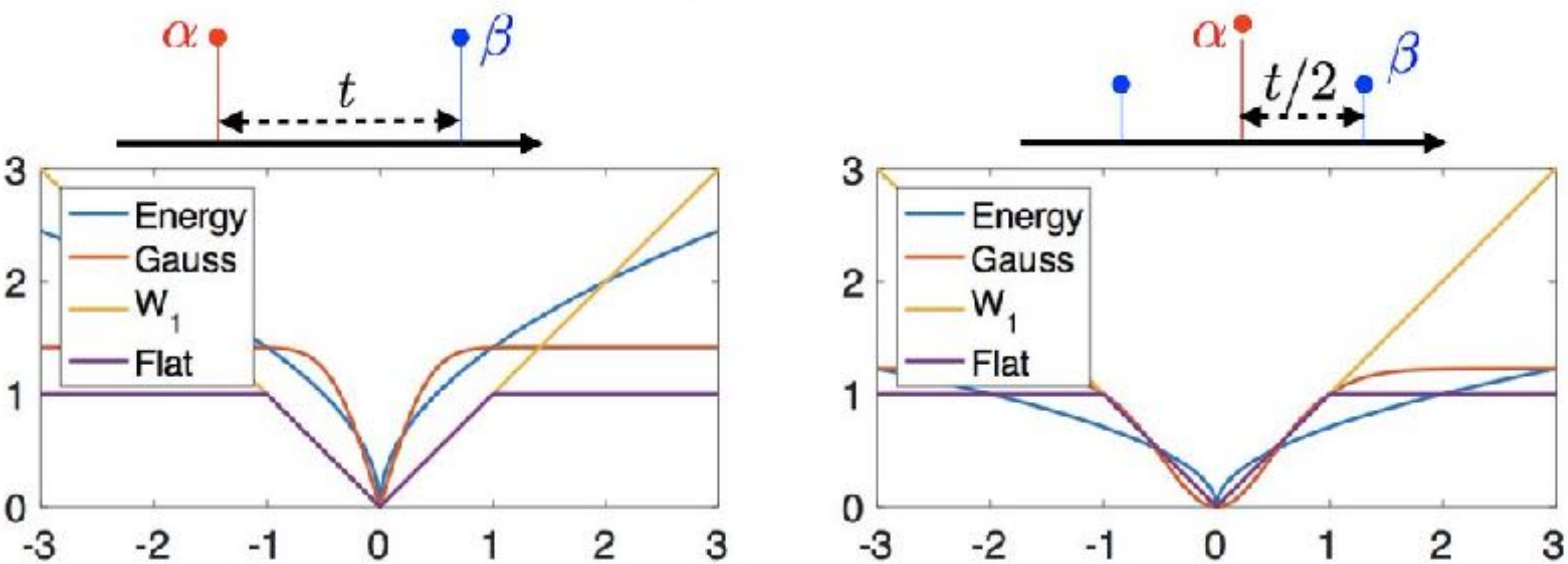
Flat norm: $B = \{f \;;\; \|f\|_\infty \leqslant 1, \|\nabla f\|_\infty \leqslant 1\}$.

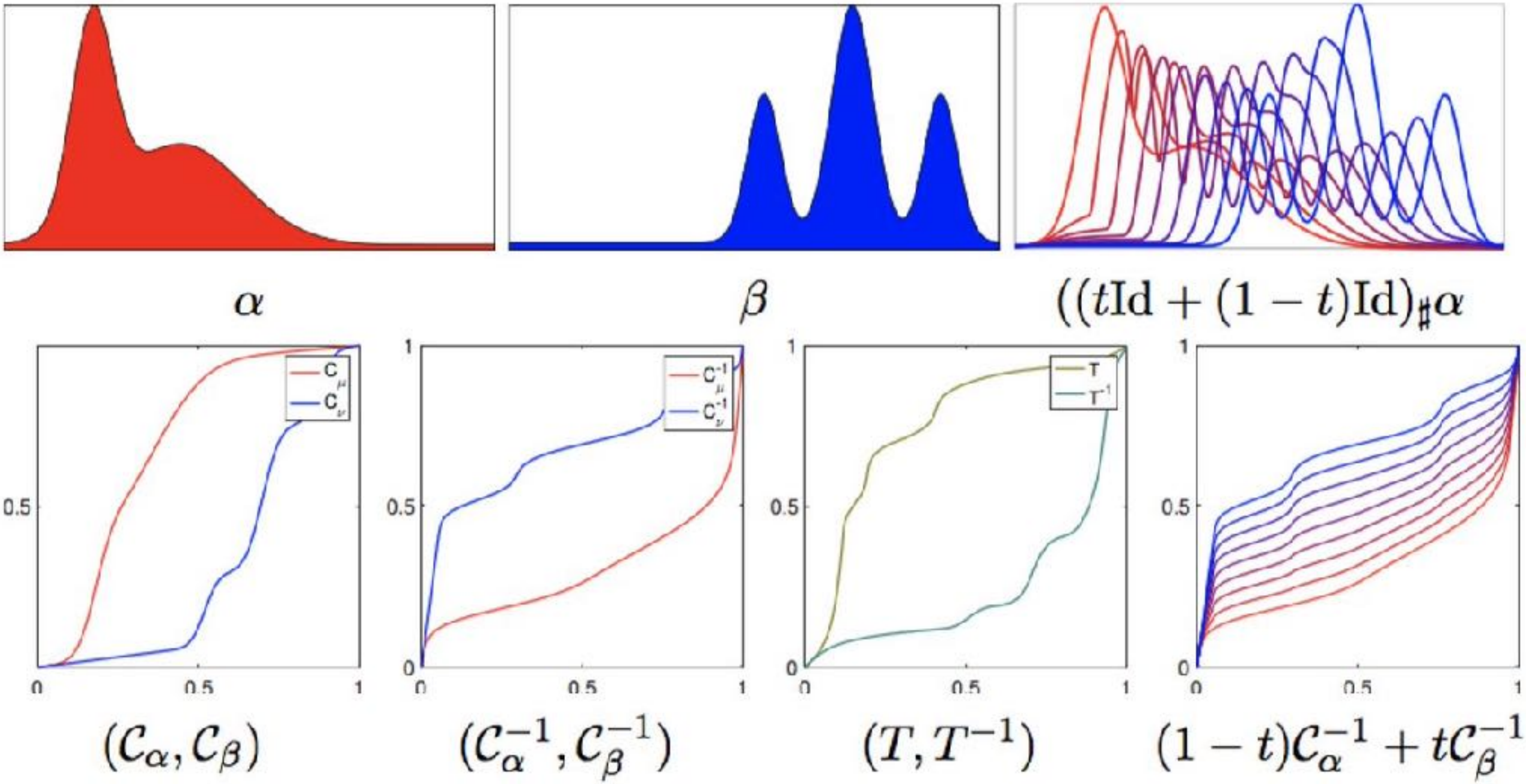
RKHS: $B = \left\{f \;;\; \|f\|_k^2 \leqslant 1\right\}$.

$$\|\alpha - \beta\|_B^2 = \int k(x,x')\mathrm{d}\alpha(x)\mathrm{d}\alpha(x') + \int k(x,x')\mathrm{d}\beta(y)\mathrm{d}\beta(y') - 2\int k(x,y)\mathrm{d}\alpha(x)\mathrm{d}\beta(y)$$

Energy distance: $k(x,y) = -\|x-y\|_2$

Gaussian: $k(x,y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$





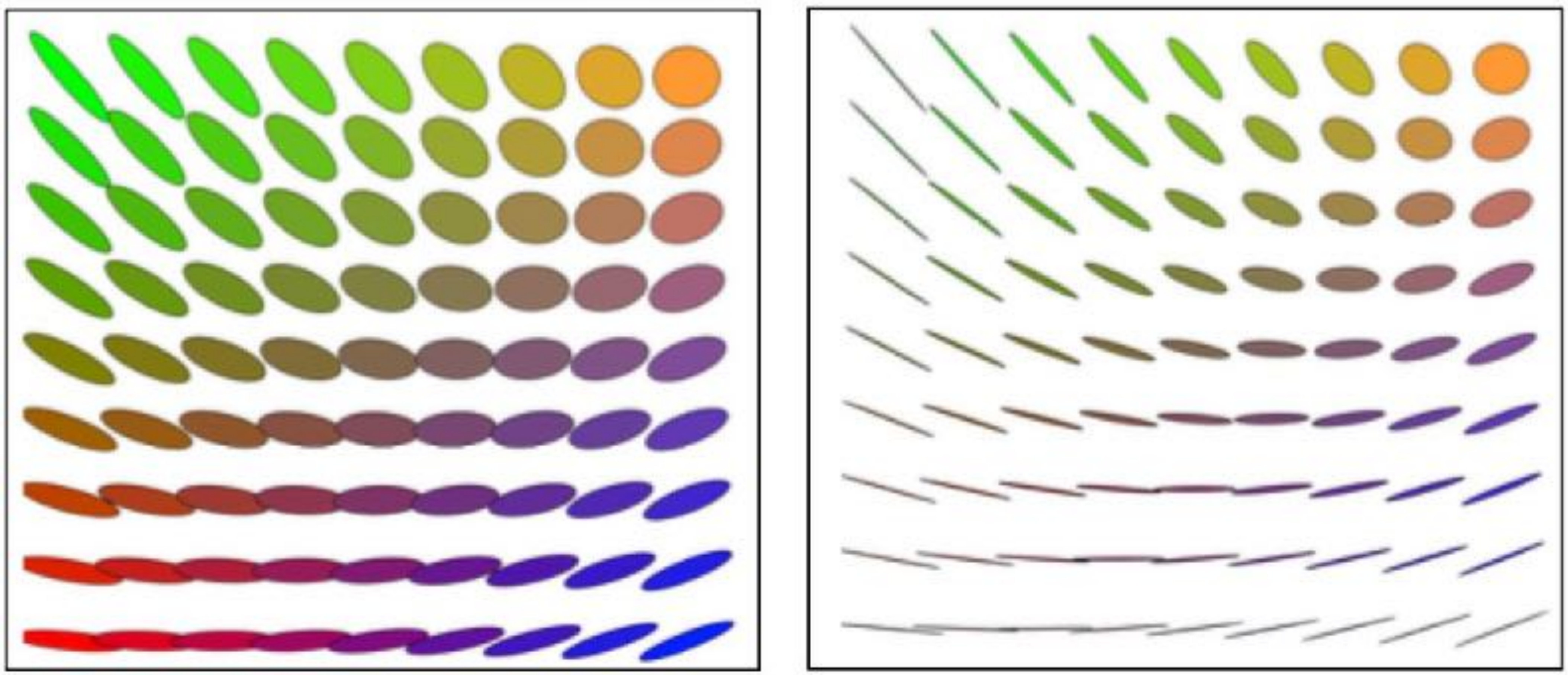
α
β
$((t\mathrm{Id} + (1-t)\mathrm{Id})_\sharp\alpha$
C_μ
C_ν
C_μ^{-1}
C_ν^{-1}
T
T^{-1}
0
0.5
1
$(\mathcal{C}_\alpha, \mathcal{C}_\beta)$
$(\mathcal{C}_\alpha^{-1}, \mathcal{C}_\beta^{-1})$
(T, T^{-1})
$(1-t)\mathcal{C}_\alpha^{-1} + t\mathcal{C}_\beta^{-1}$

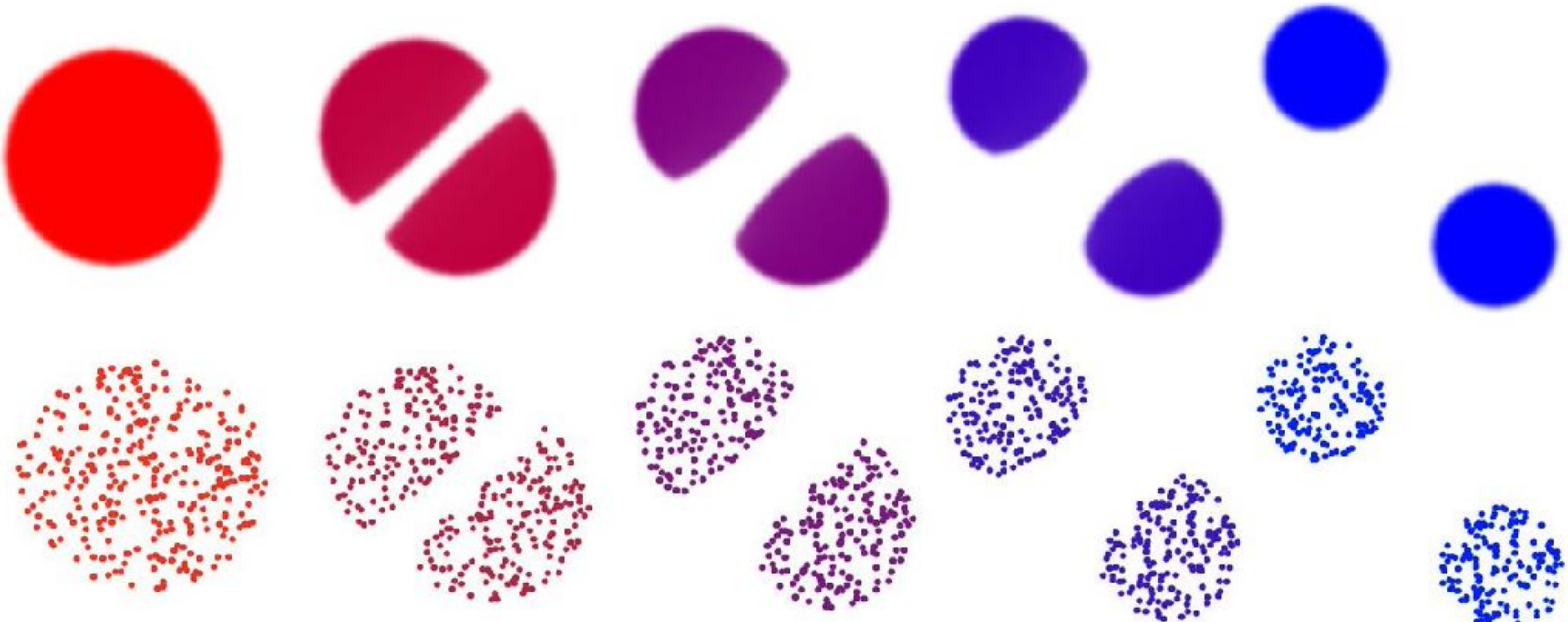
Remark 2.11 (Distance between Gaussians). If $\alpha = \mathcal{N}(m_\alpha, C_\alpha)$ and $\beta = \mathcal{N}(m_\beta, C_\beta)$, then one can show that

$$\mathcal{W}_2^2(\alpha, \beta) = \|m_\alpha - m_\beta\|^2 + \mathcal{B}(C_\alpha, C_\beta)^2 \tag{2.19}$$

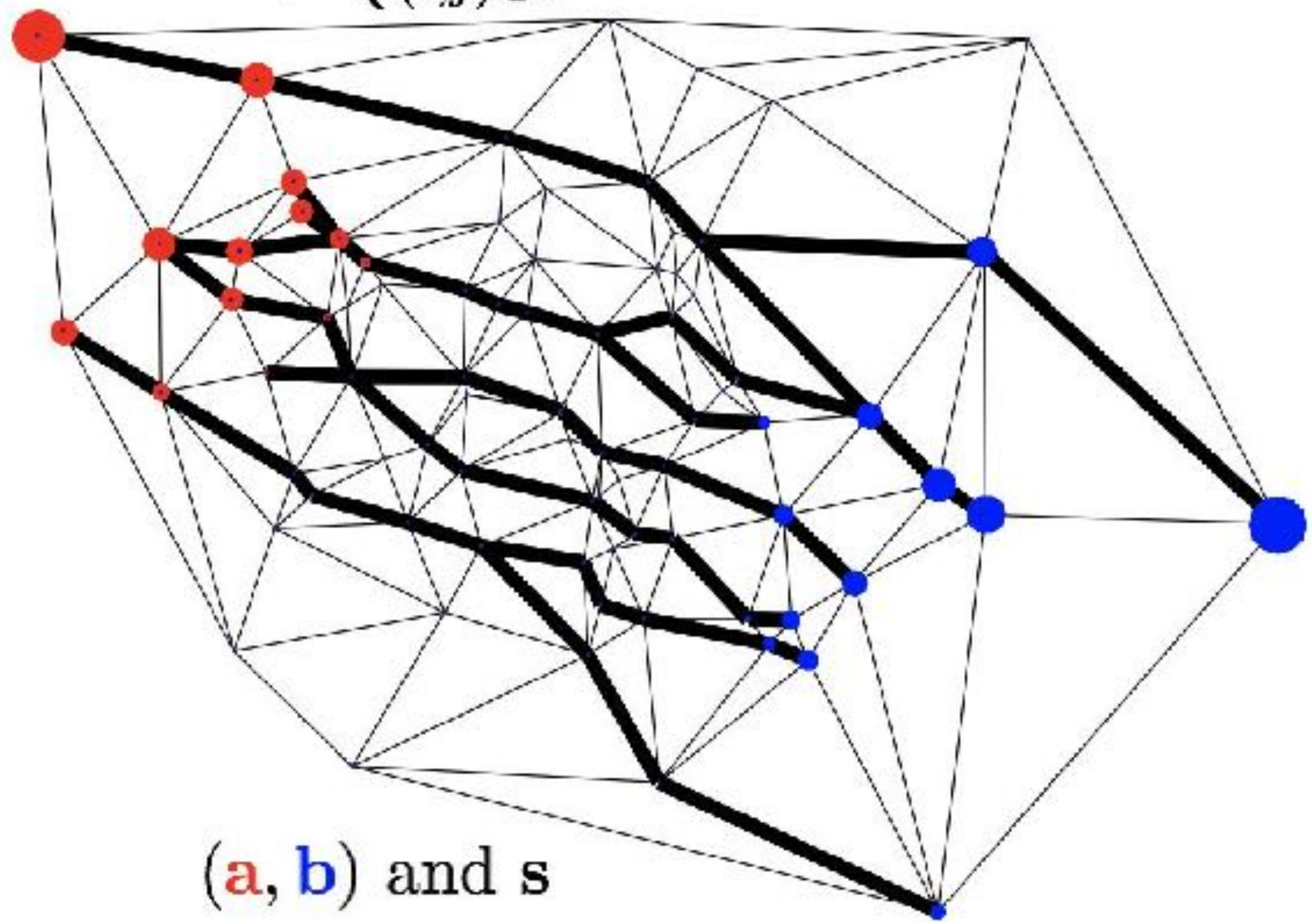
where $\mathcal{B}$ is the so-called Bures metric

$$\mathcal{B}(C_\alpha, C_\beta)^2 \stackrel{\text{def.}}{=} \operatorname{tr}\left(C_\alpha + C_\alpha - 2(C_\alpha^{1/2} C_\beta C_\alpha^{1/2})^{1/2}\right) \tag{2.20}$$

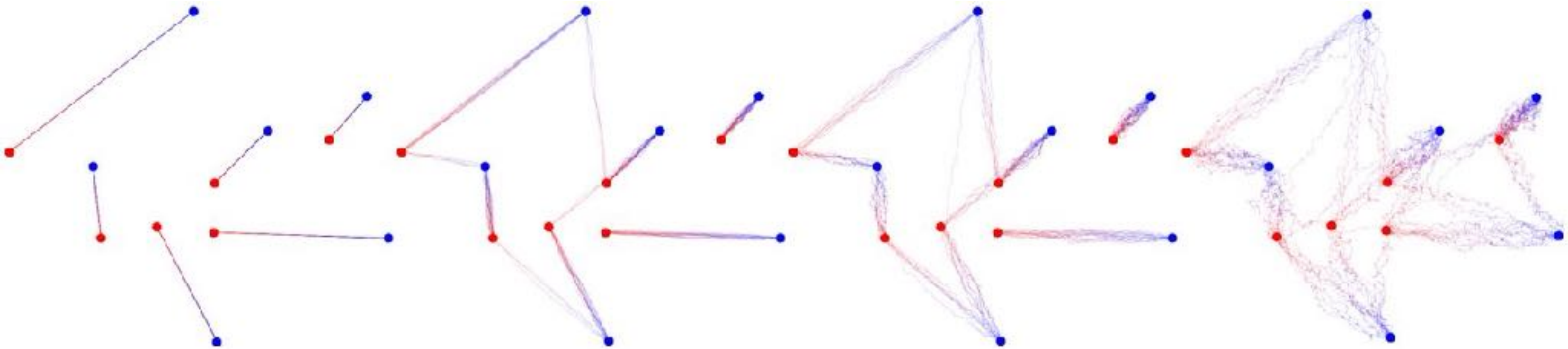




$$\mathrm{W}_1(\mathbf{a},\mathbf{b}) = \min_{\mathbf{s}\in\mathbb{R}_+^{\mathcal{E}}} \left\{ \sum_{(i,j)\in\mathcal{E}} \mathbf{w}_{i,j}\mathbf{s}_{i,j} \;:\; \mathrm{div}(\mathbf{s}) = \mathbf{a}-\mathbf{b} \right\}$$



$(\mathbf{a}, \mathbf{b})$ and $\mathbf{s}$



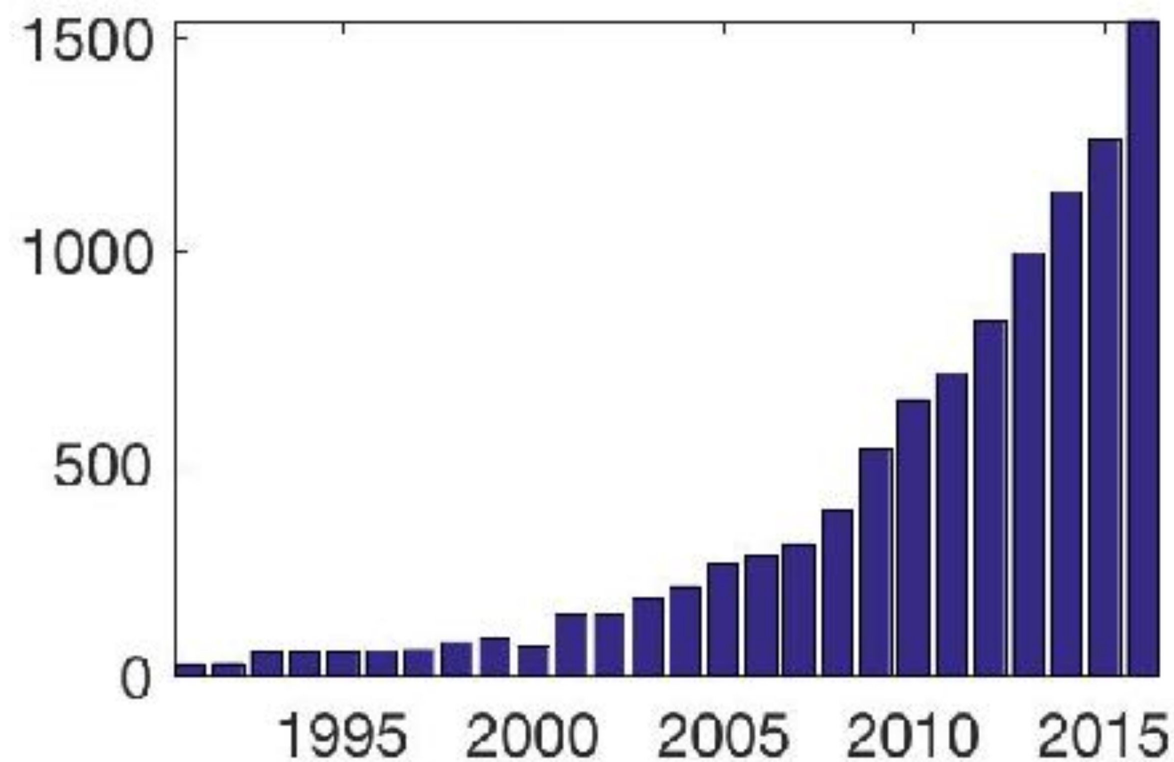
$\varepsilon = 0$ $\varepsilon = .05$ $\varepsilon = 0.2$ $\varepsilon = 1$

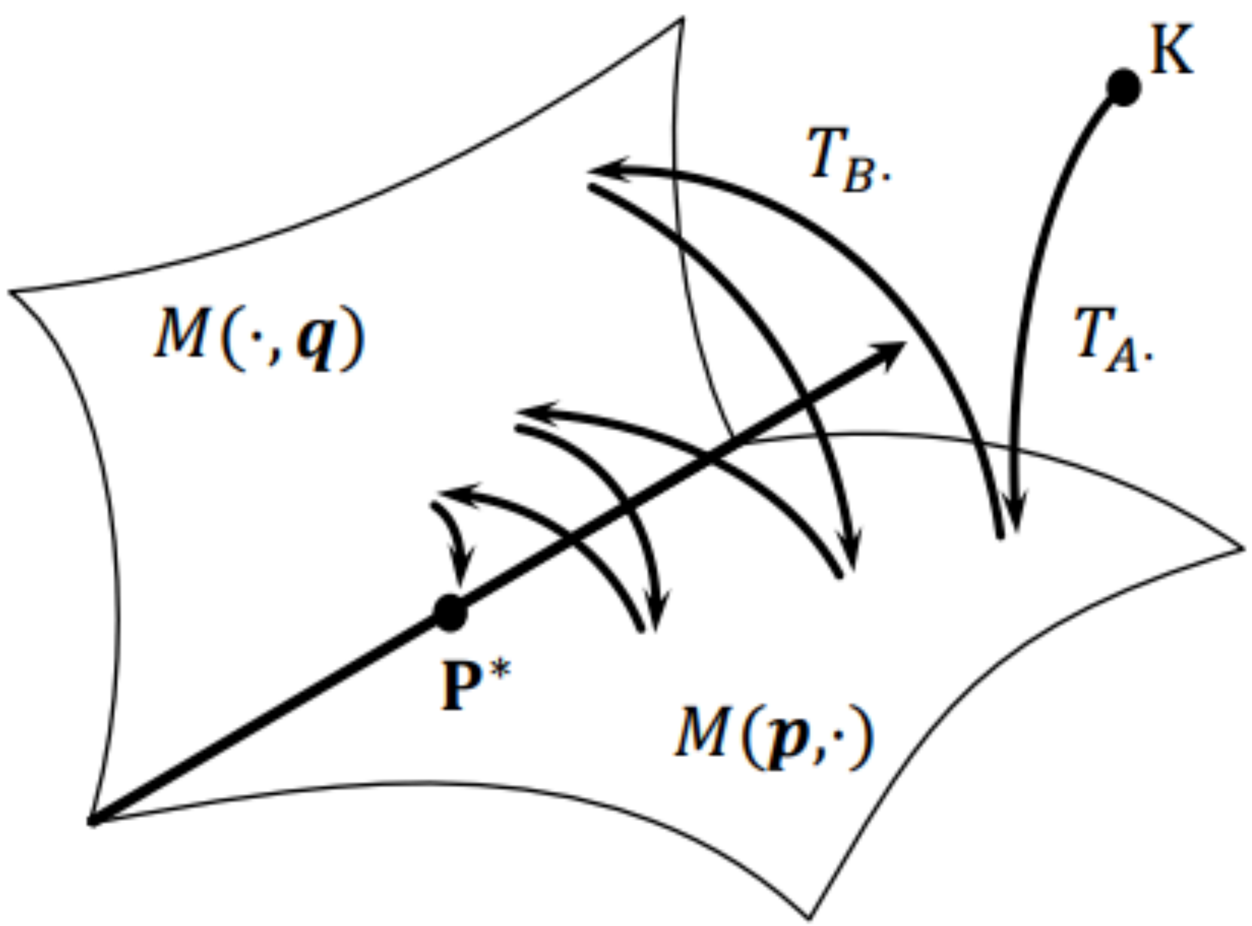
$$W_c(\mu,\nu) = \sup_{f,g} \left\{ \int_X f \mathrm{d}\mu + \int_Y g \mathrm{d}\nu \; ; \; \forall (x,y), f(x) + g(y) \leqslant c(x,y) \right\}$$

$$W_c^*(f,g) = \inf_{x,y} \; c(x,y) - f(x) + g(y)$$

If (μ, ν) have densities bounded from below/above by $0 < a < b < +\infty$,

$$b^{-1/2}\|\mu - \nu\|_{H^{-1}} \leqslant W_2(\mu, \nu) \leqslant a^{-1/2}\|\mu - \nu\|_{H^{-1}}$$


1500
1000
500
0
1995
2000
2005
2010
2015



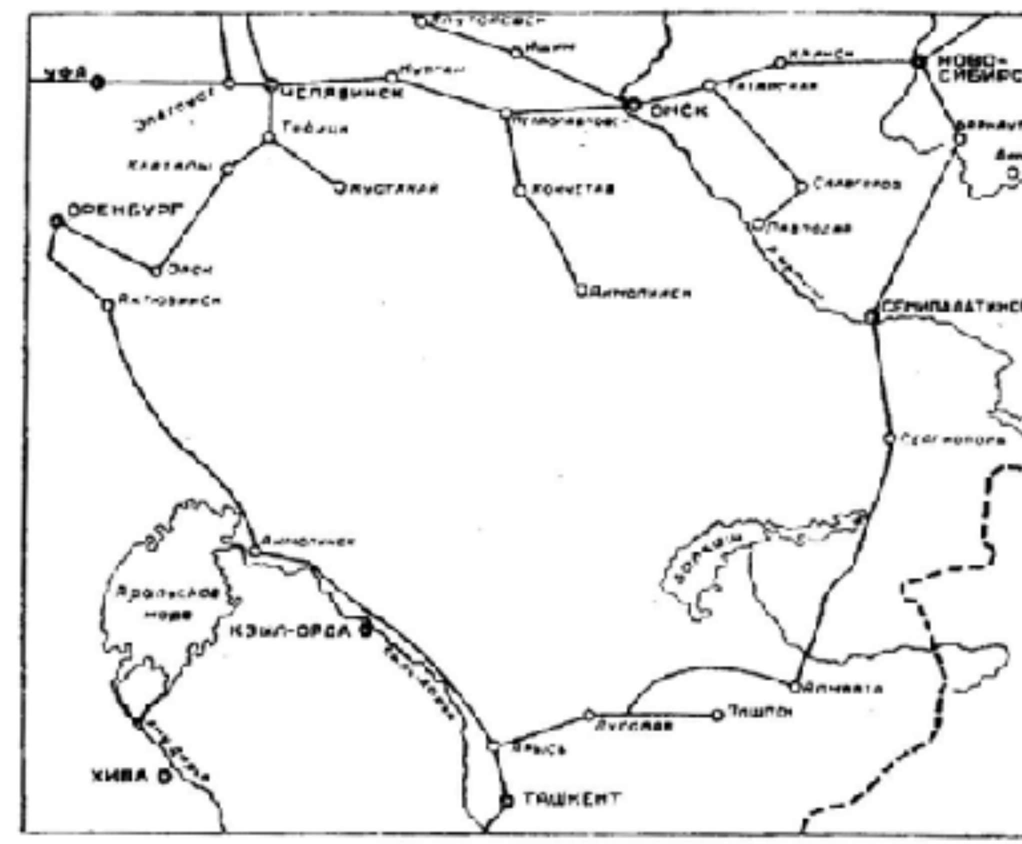
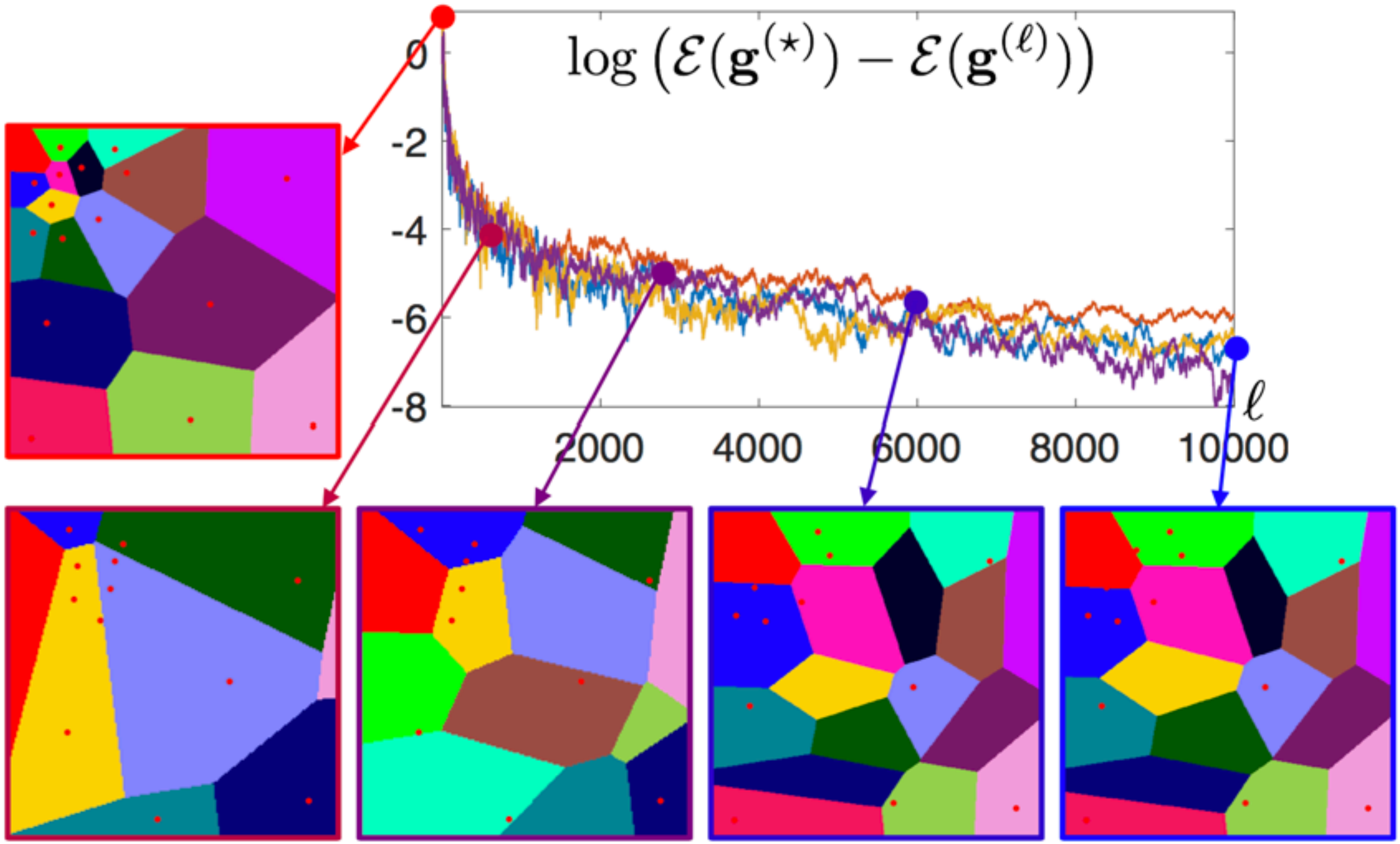
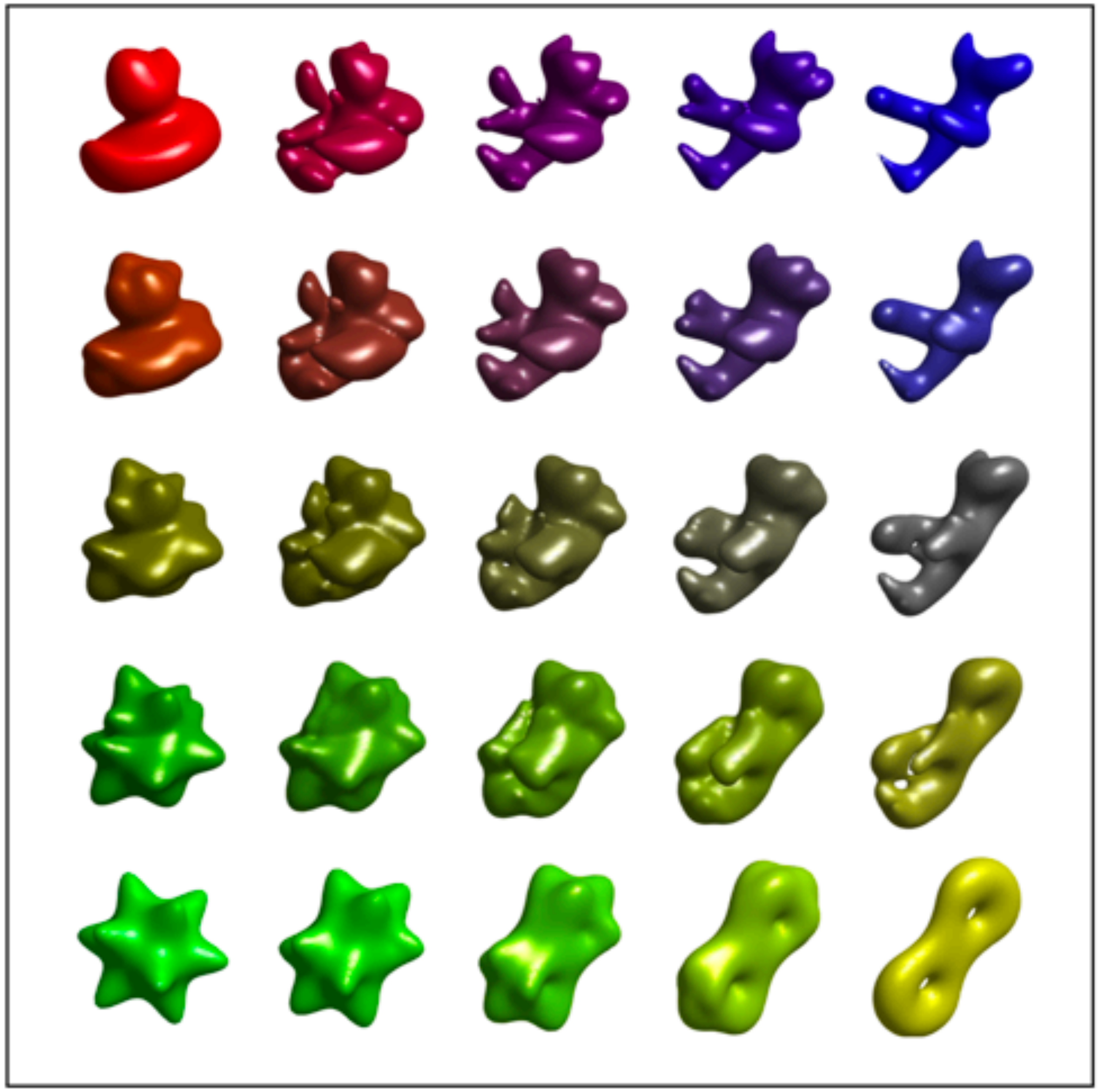
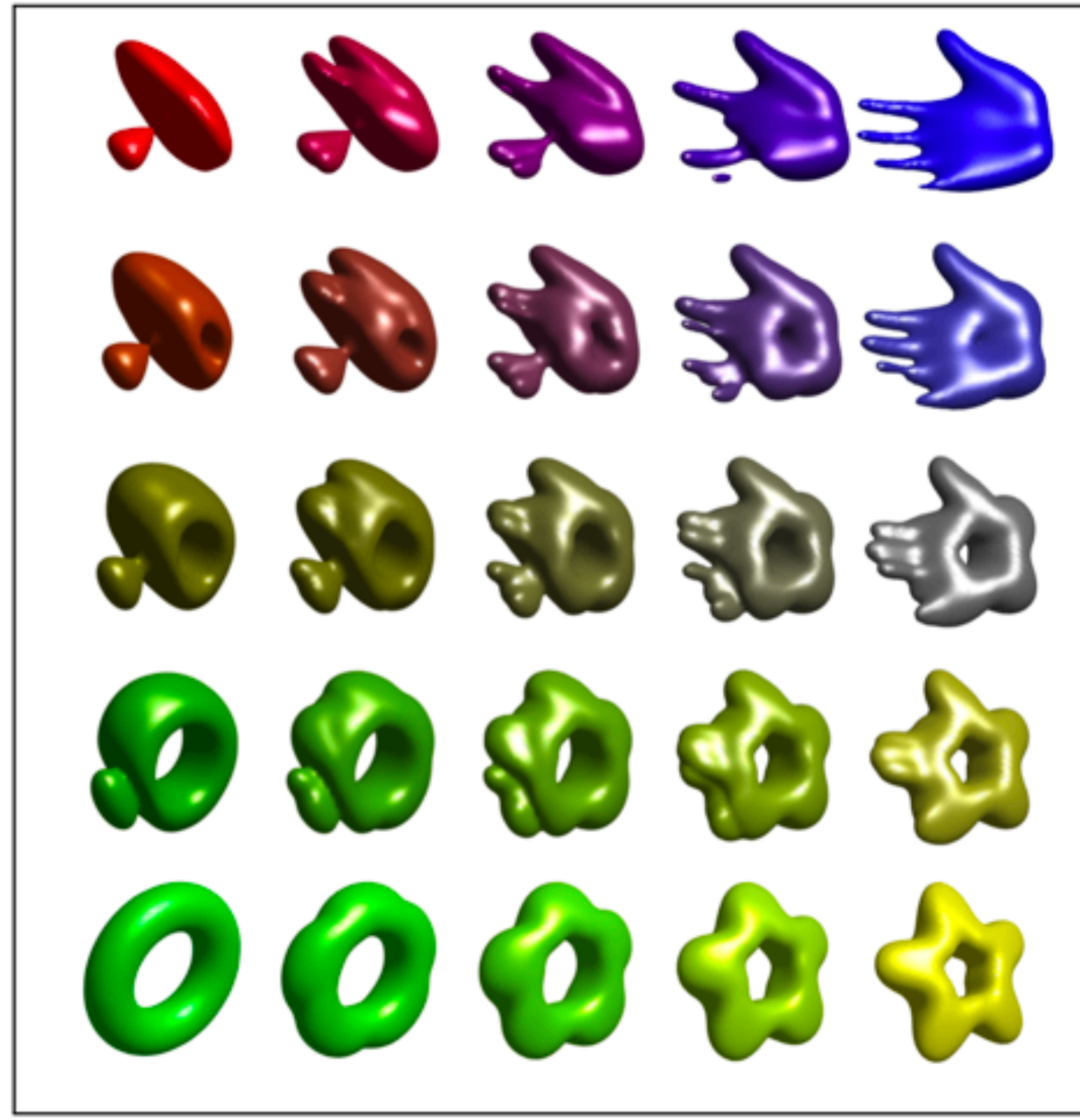
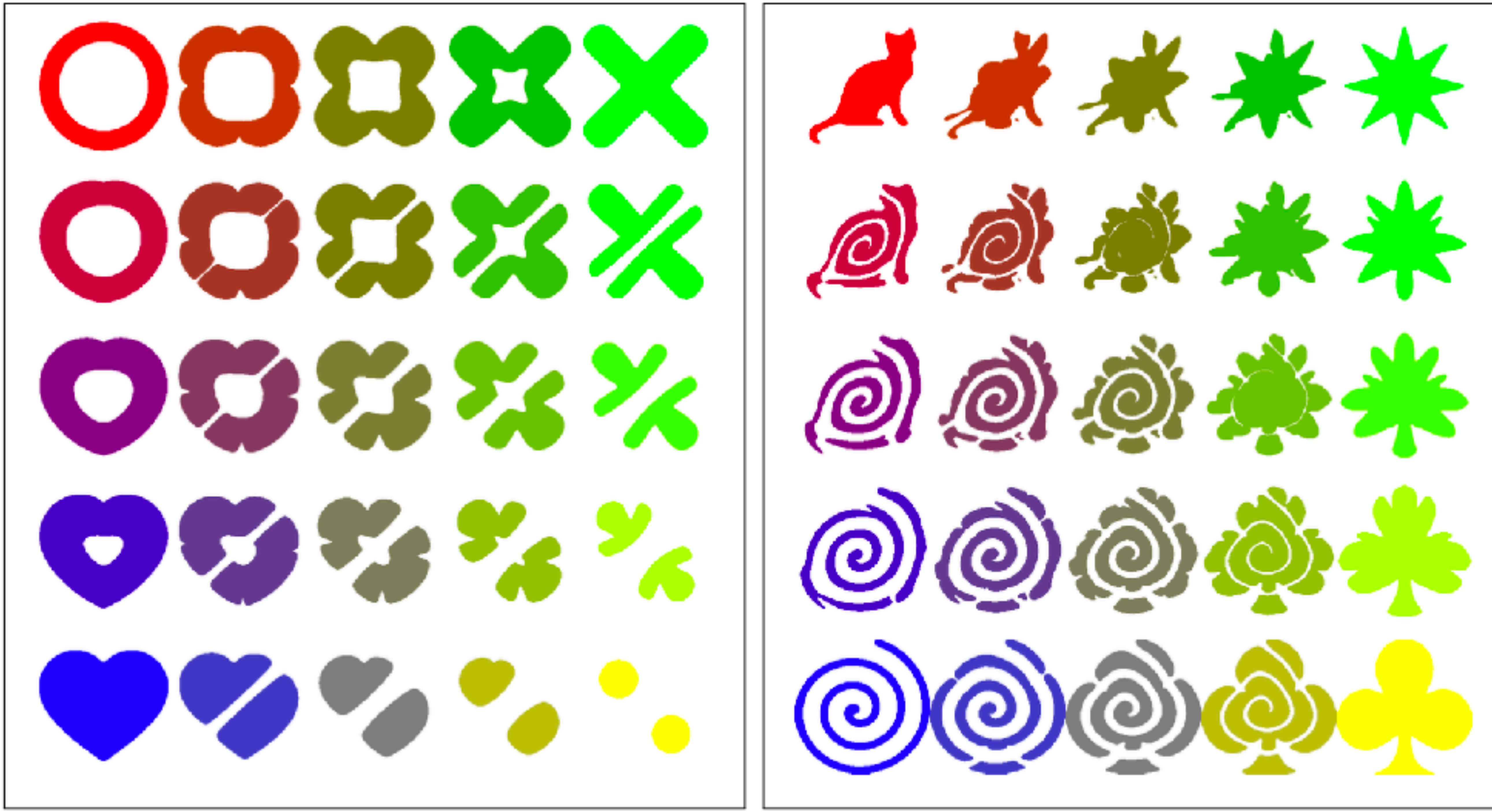


Figure 1: Figure from Tolstoĭ [1930] to illustrate a negative cycle

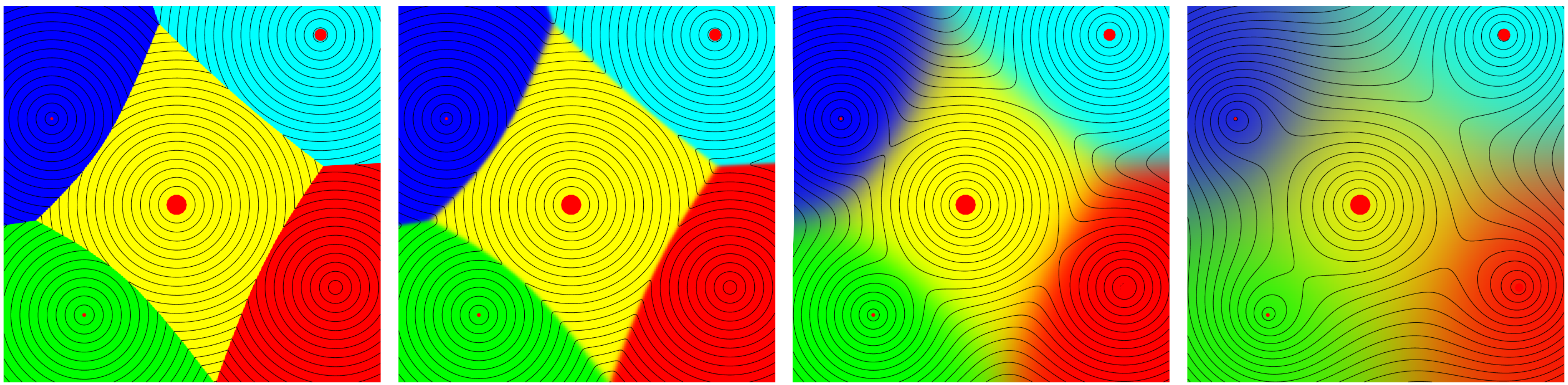

$\log\left(\mathcal{E}(\mathbf{g}^{(\star)}) - \mathcal{E}(\mathbf{g}^{(\ell)})\right)$
0
-2
-4
-6
-8
ℓ
2000
4000
6000
8000
10000



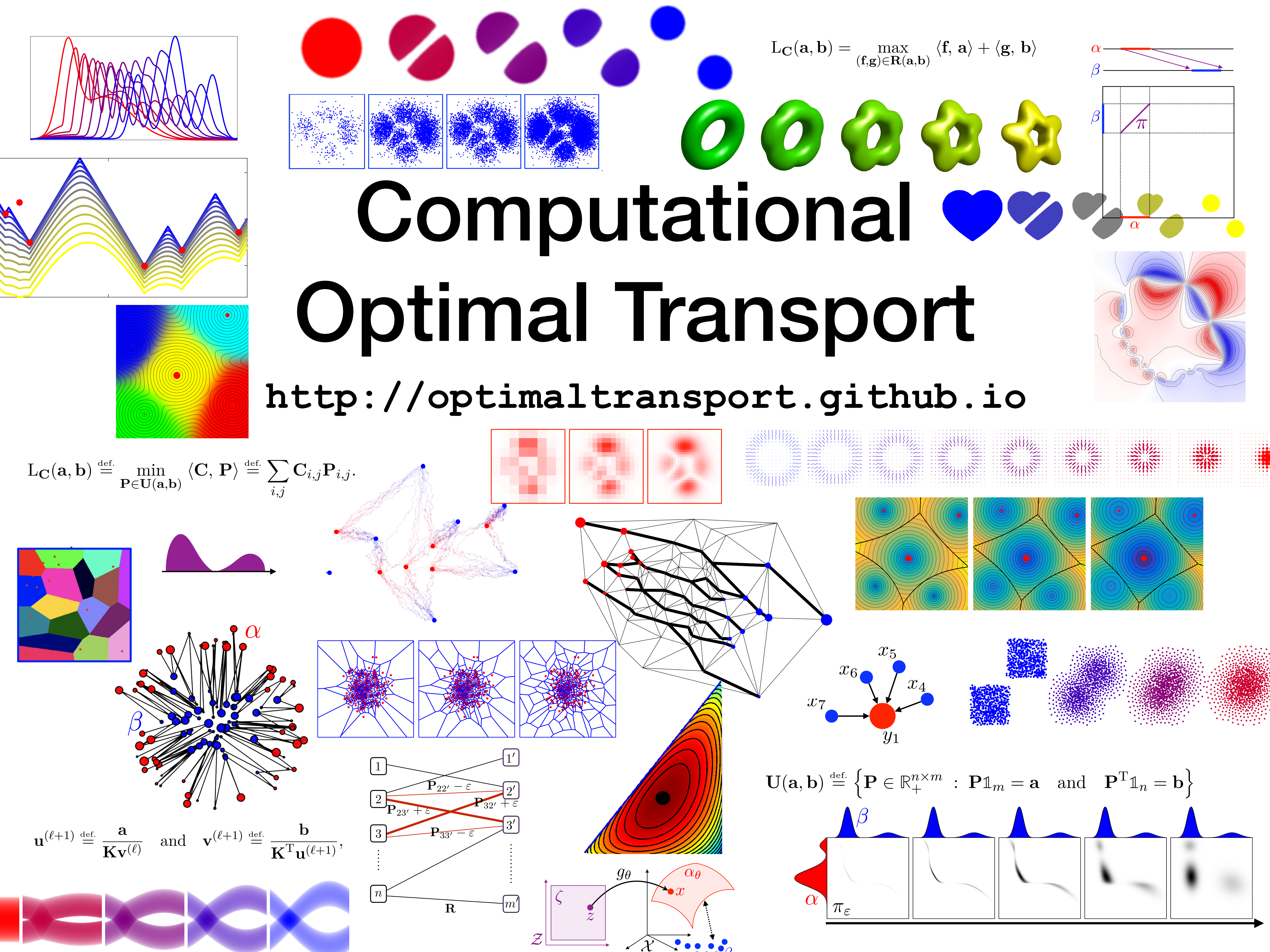


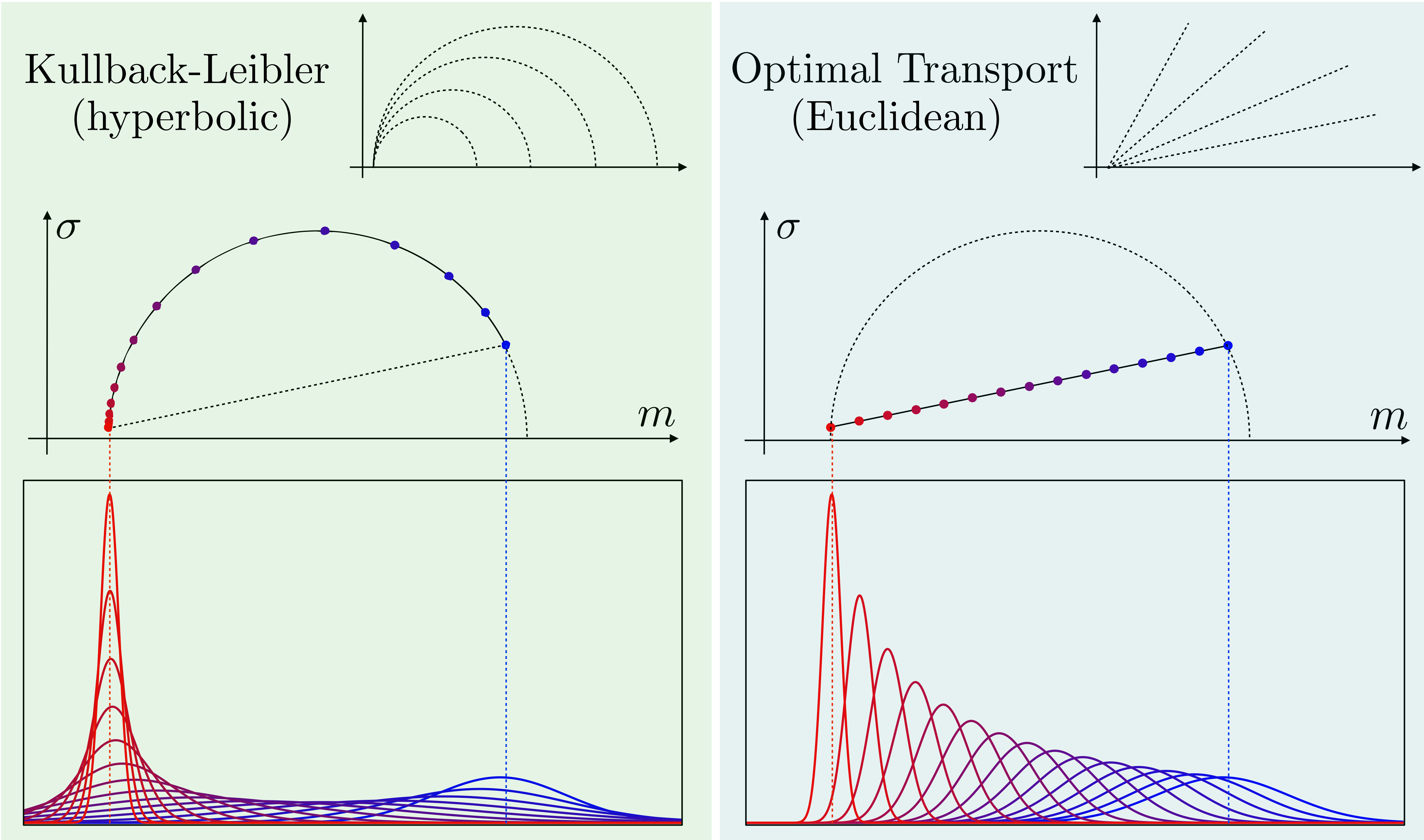


Laguerre cells | “Sinkhorn” Laguerre cells



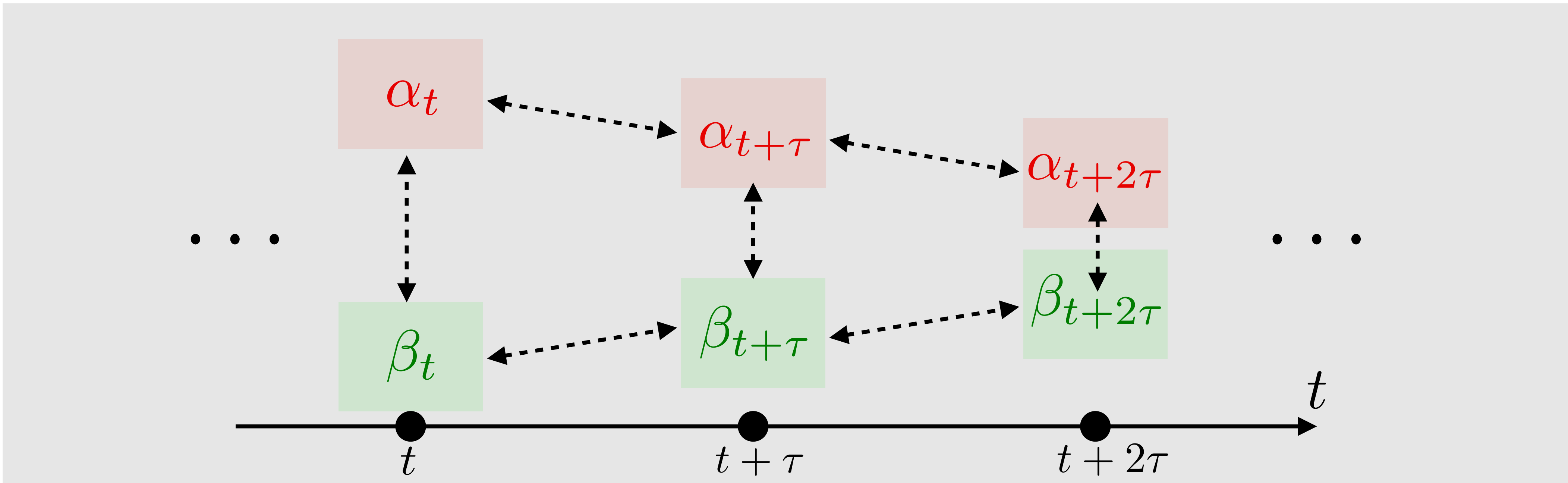
$\varepsilon = 0$ | $\varepsilon = 0.01$ | $\varepsilon = 0.1$ | $\varepsilon = 0.3$

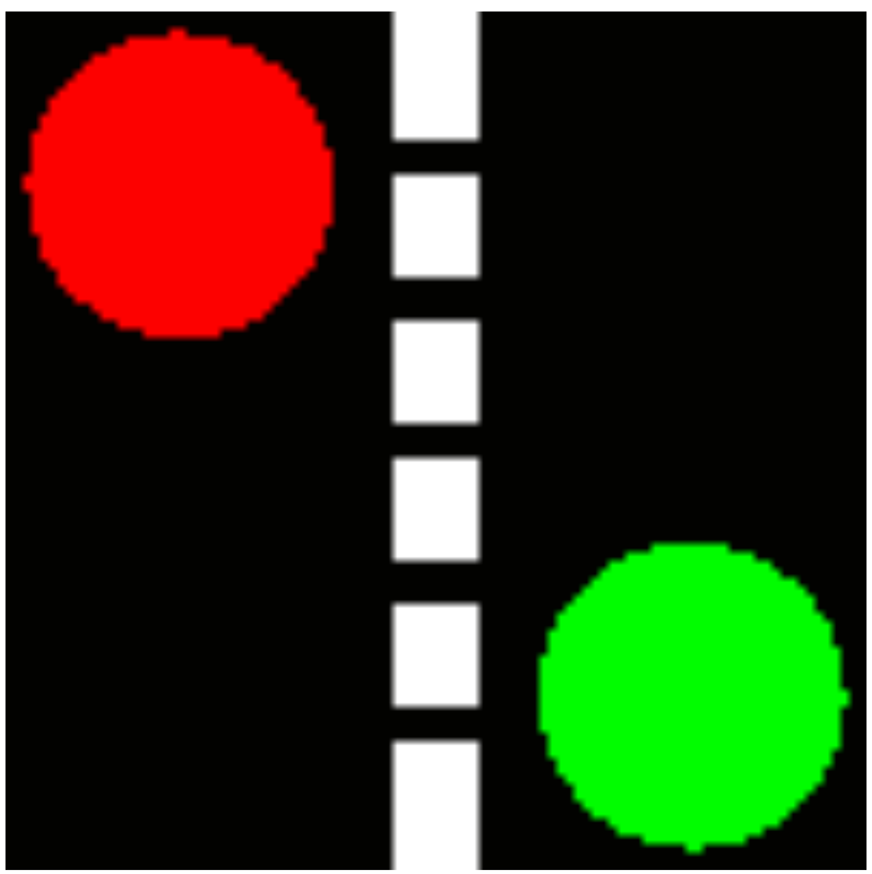

Computational
Optimal Transport
http://optimaltransport.github.io
$\mathrm{L}_{\mathbf{C}}(\mathbf{a},\mathbf{b}) = \max_{(\mathbf{f},\mathbf{g})\in\mathbf{R}(\mathbf{a},\mathbf{b})} \langle \mathbf{f},\, \mathbf{a}\rangle + \langle \mathbf{g},\, \mathbf{b}\rangle$
$\mathrm{L}_{\mathbf{C}}(\mathbf{a},\mathbf{b}) \overset{\text{def.}}{=} \min_{\mathbf{P}\in\mathbf{U}(\mathbf{a},\mathbf{b})} \langle \mathbf{C},\, \mathbf{P}\rangle \overset{\text{def.}}{=} \sum_{i,j} \mathbf{C}_{i,j}\mathbf{P}_{i,j}.$
$\mathbf{U}(\mathbf{a},\mathbf{b}) \overset{\text{def.}}{=} \left\{\mathbf{P}\in\mathbb{R}_+^{n\times m} \;:\; \mathbf{P}\mathbb{1}_m = \mathbf{a} \quad \text{and} \quad \mathbf{P}^{\mathrm{T}}\mathbb{1}_n = \mathbf{b}\right\}$
$\mathbf{u}^{(\ell+1)} \overset{\text{def.}}{=} \frac{\mathbf{a}}{\mathbf{K}\mathbf{v}^{(\ell)}} \quad \text{and} \quad \mathbf{v}^{(\ell+1)} \overset{\text{def.}}{=} \frac{\mathbf{b}}{\mathbf{K}^{\mathrm{T}}\mathbf{u}^{(\ell+1)}},$
α
β
π
π_ε
x_4 x_5 x_6 x_7 y_1
1 2 3 n
1′ 2′ 3′ m′
$\mathbf{P}_{22'} - \varepsilon$
$\mathbf{P}_{23'} + \varepsilon$
$\mathbf{P}_{32'} + \varepsilon$
$\mathbf{P}_{33'} - \varepsilon$
R
ζ z $\mathcal{Z}$ g_θ α_θ x $\mathcal{X}$

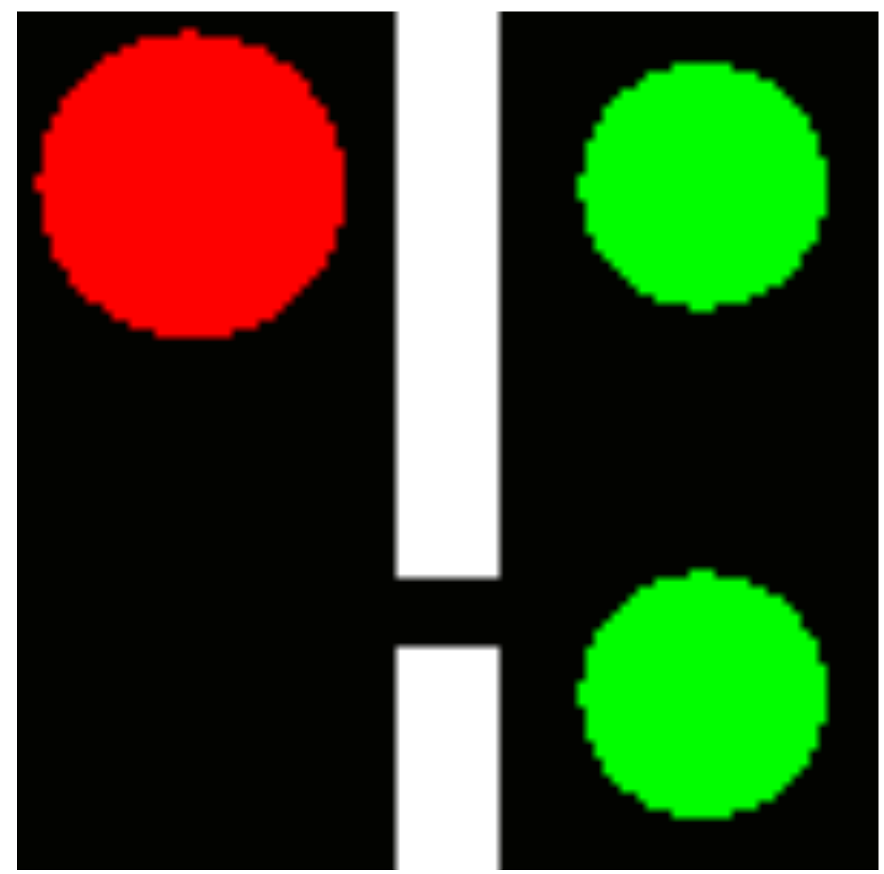


Kullback-Leibler
(hyperbolic)
σ
m
Optimal Transport
(Euclidean)
σ
m

$$(\alpha_{t+\tau}, \beta_{t+\tau}) = \operatorname*{argmin}_{(\alpha,\beta)\leqslant C} W_2^2(\alpha_t, \alpha) + W_2^2(\beta_t, \beta) + \tau W_2^2(\alpha, \beta)$$





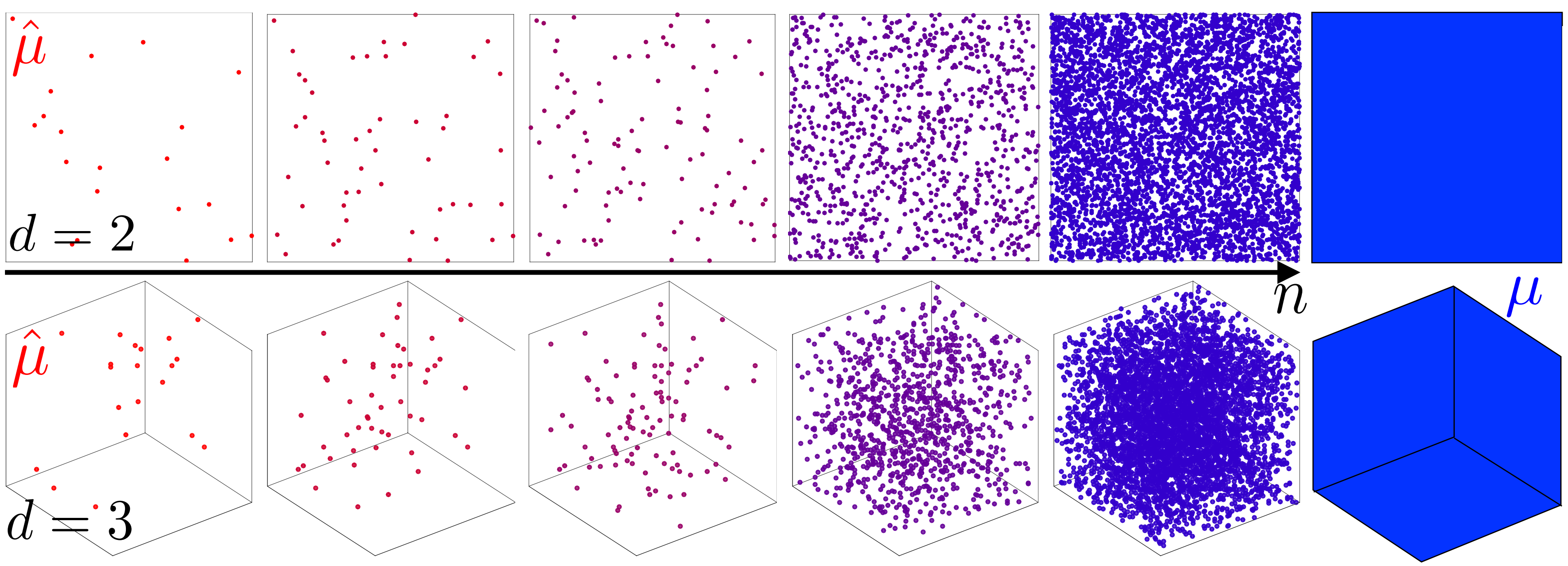


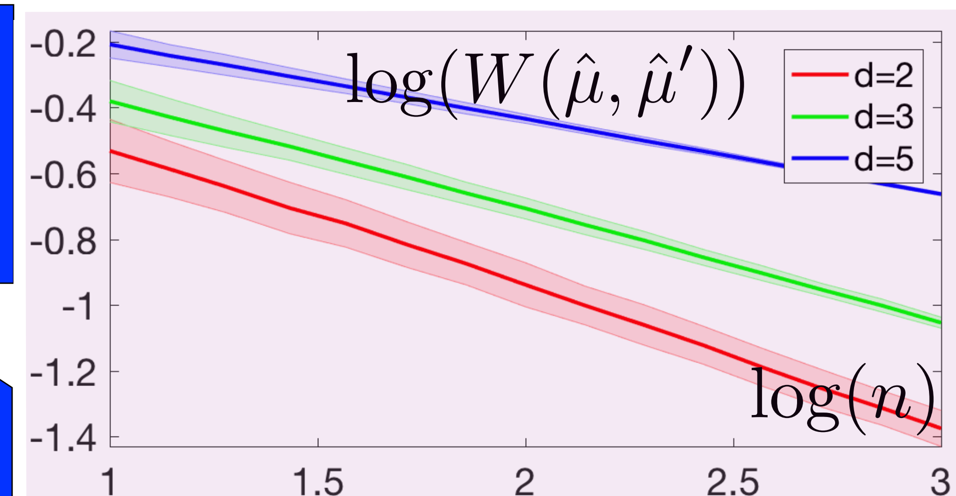
Optimal transport:

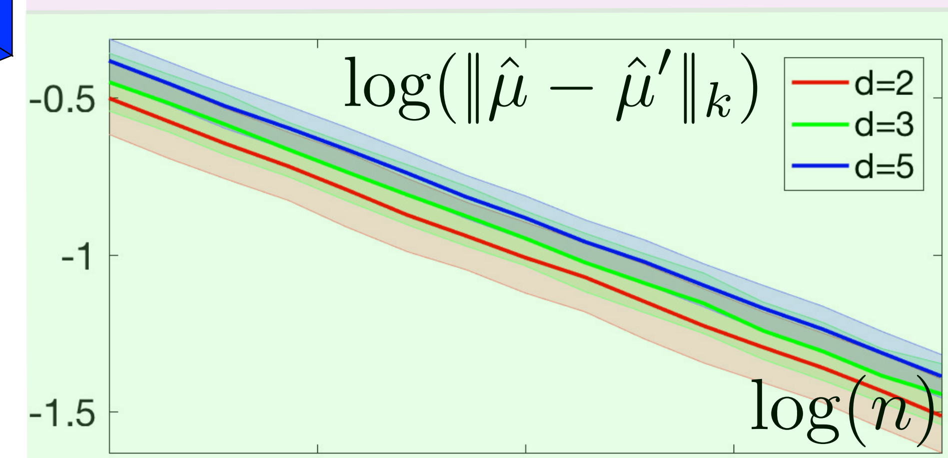
$$W(\mu,\nu)^p \overset{\text{def.}}{=} \left\{ \max_{f,g} \int f \mathrm{d}\mu + \int g \mathrm{d}\nu \;;\; \forall (x,y), f(x)+g(y) \leqslant \|x-y\|^p \right\}$$

Euclidean RKHS (MMD): $\|\mu-\nu\|_k^2 \overset{\text{def.}}{=} \mathbb{E}_{\mu-\nu}(k(x,y))$

$$\text{e.g. } k(x,y) = e^{-\|x-y\|^2},\; k(x,y) = -\|x-y\|,\; \ldots$$





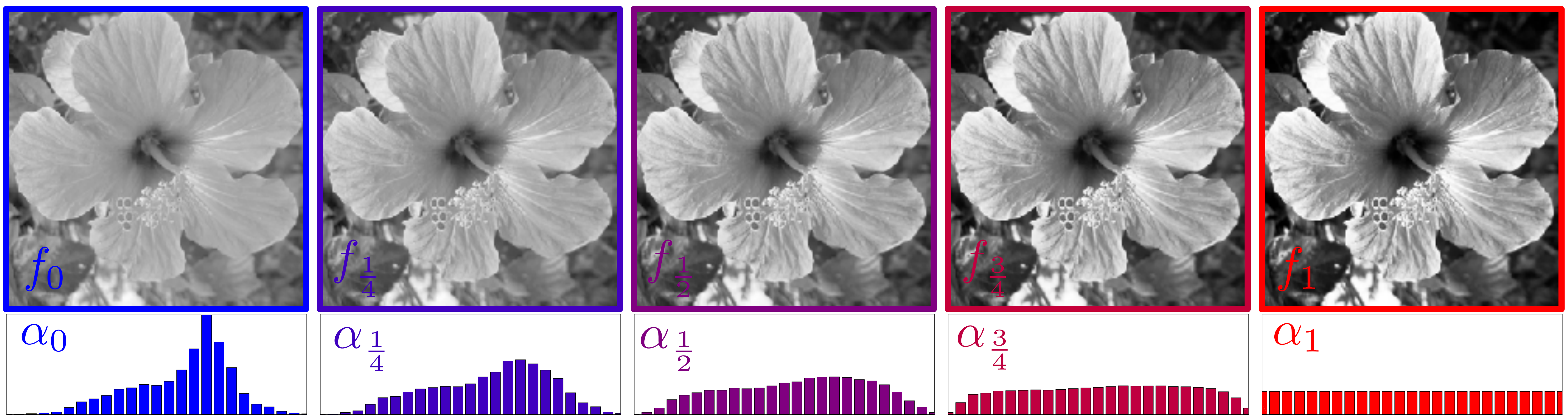


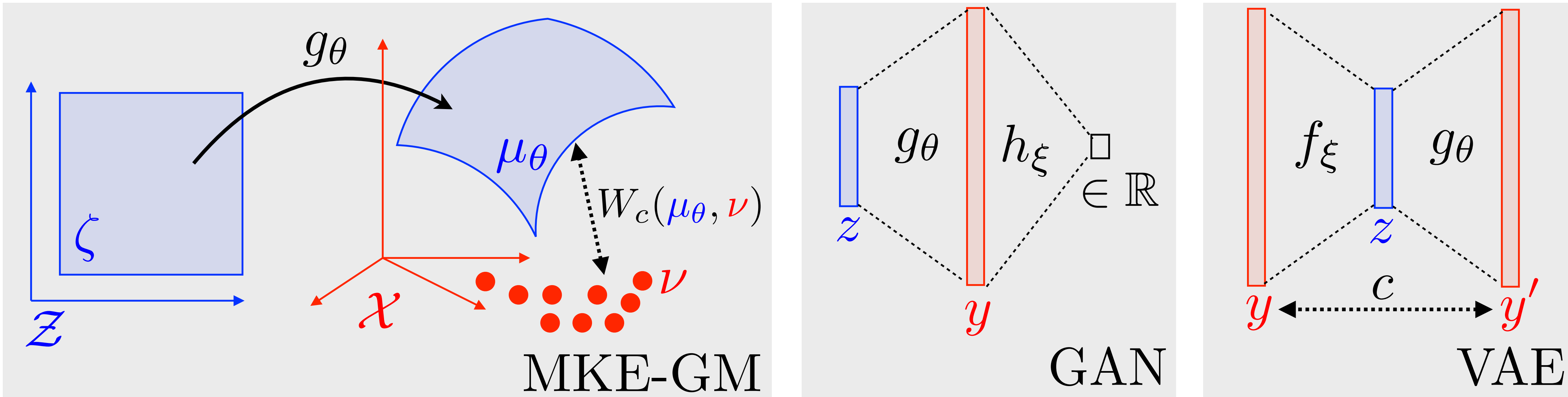
Theorem: $\mathbb{E}(|W(\hat{\mu},\hat{\nu}) - W(\mu,\nu)|) = O(n^{-\frac{1}{d}})$

$\mathbb{E}(|\|\hat{\mu}-\hat{\nu}\|_k - \|\mu-\nu\|_k|) = O(n^{-\frac{1}{2}})$

```
[~,I] = sort(f(:));
f(I) = linspace(0,1,length(f(:)));
```

```
f[argsort(f.flatten())] = np.linspace(0,1,n*n)
```





Minimum Kantorovich estimator:

$$\min_{\theta} E(\theta) \stackrel{\text{def.}}{=} W_c(g_{\theta\sharp}\zeta, \nu)$$

(MKE-GM)

Deep-net approximation

Dual

Primal

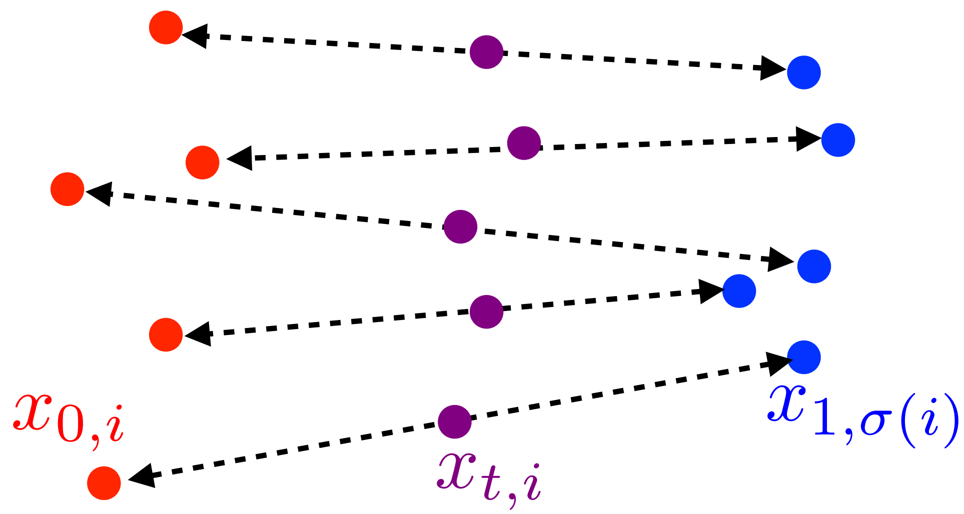
$$\min_{\theta} \max_{\xi} \int_{\mathcal{Z}} h_\xi \circ g_\xi(z) \mathrm{d}\zeta(z) + \sum_j h_\xi^c(y_j)$$

(WGAN)

$$\min_{(\theta,\xi)} \Delta_\nu(g_\theta \circ f_\xi, \mathrm{Id}_{\mathcal{X}}) + \lambda D(f_{\xi\sharp}\nu|\zeta)$$

(WVAE)

$$E(\theta_{\text{WGAN}}) \leqslant E(\theta_{\text{MKE}}) \leqslant E(\theta_{\text{WVAE}})$$

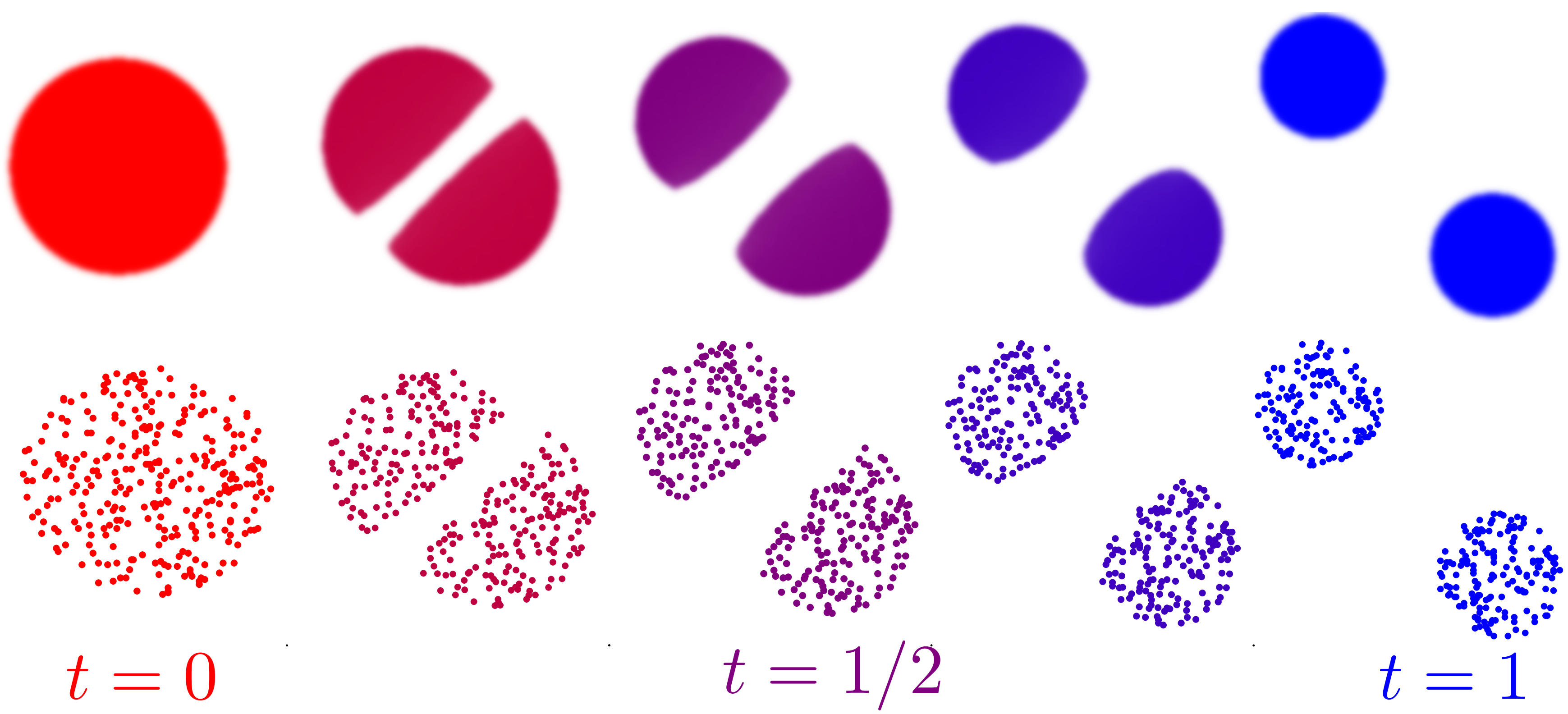


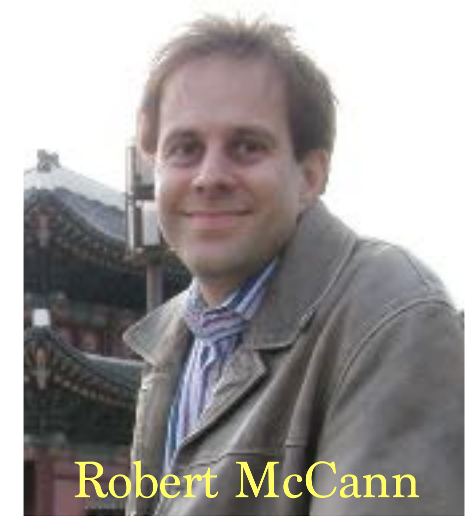
Optimal assignement:

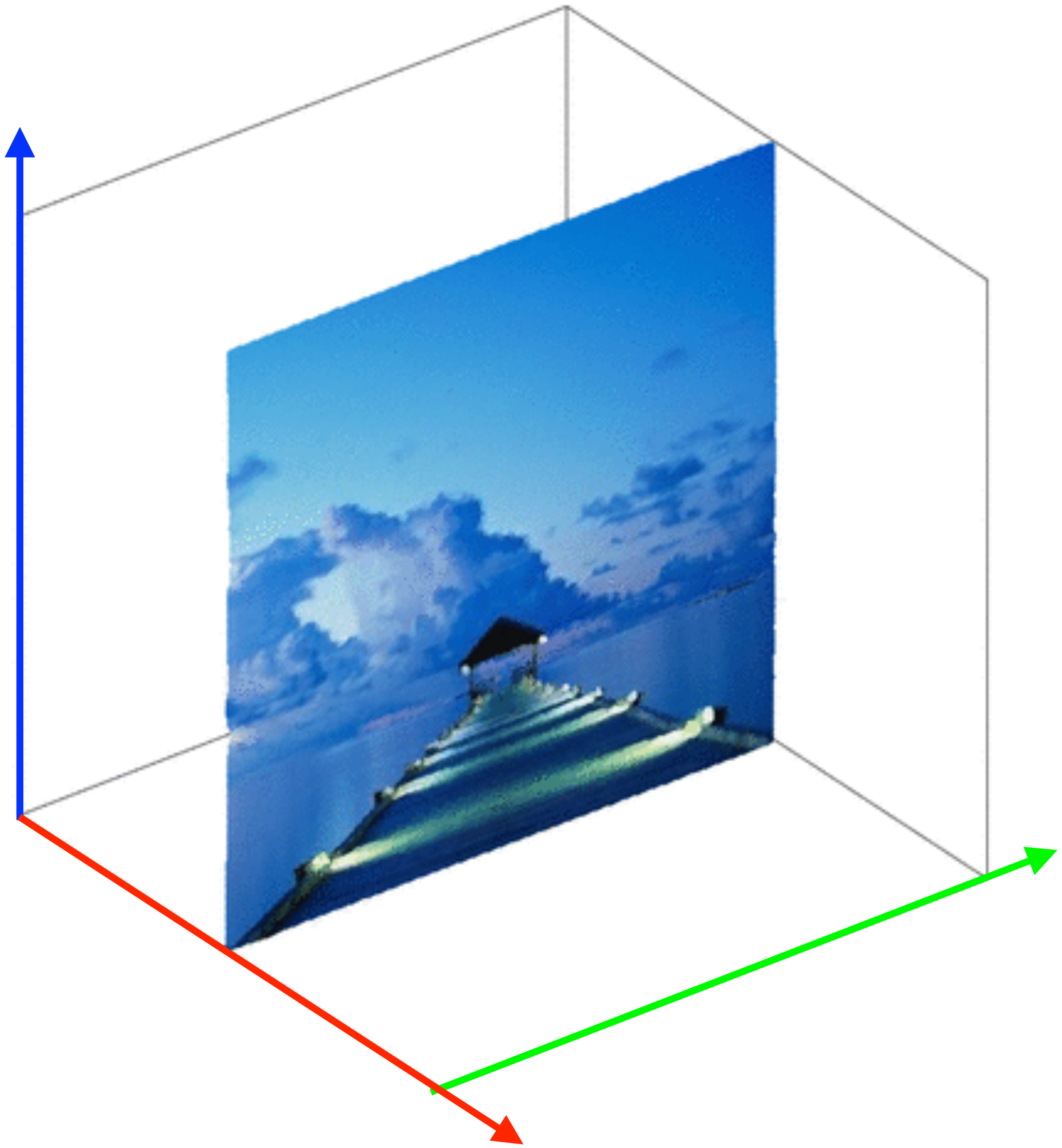
$$\min_{\sigma} \|x_0 - x_1 \circ \sigma\|$$

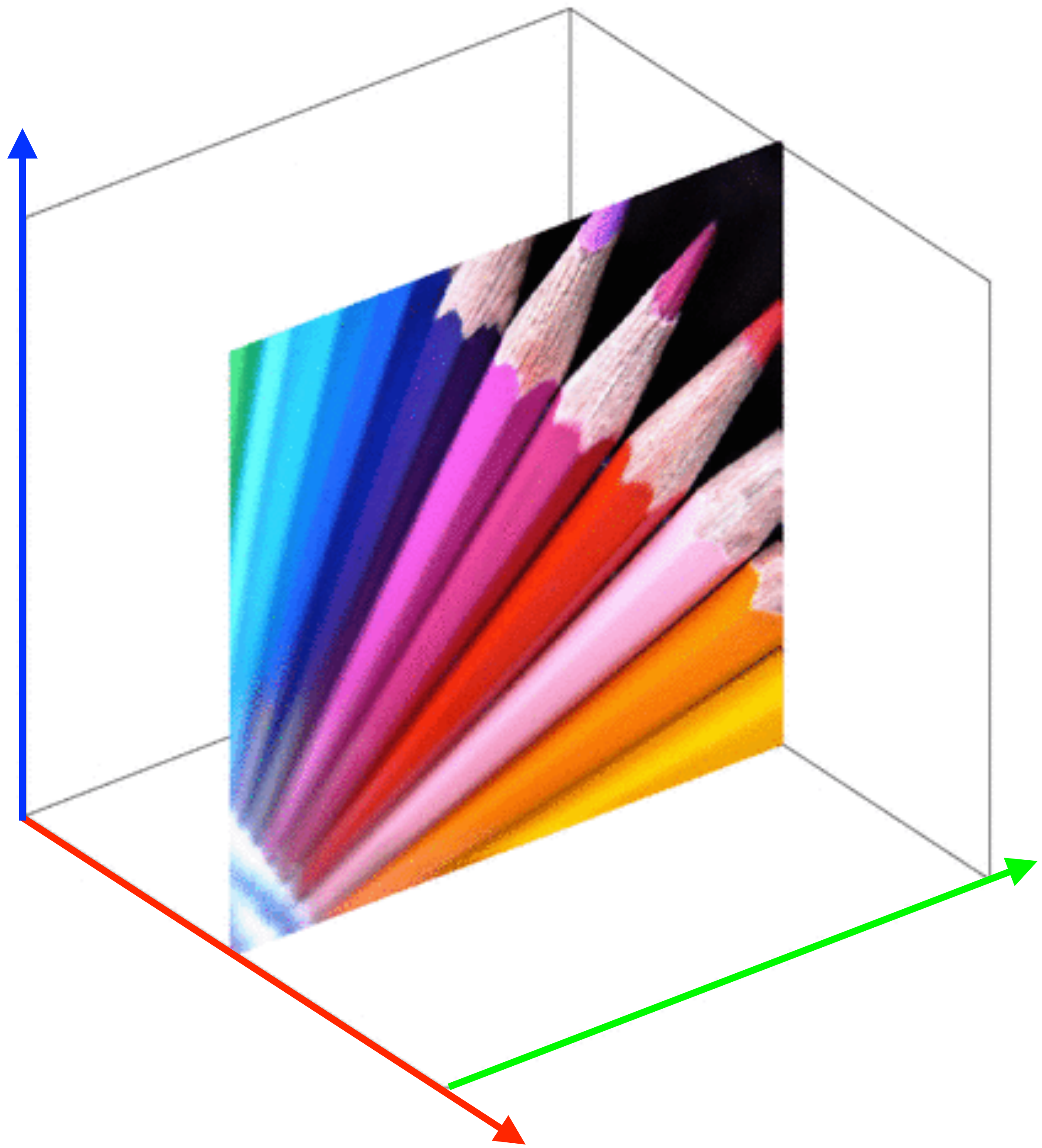
Displacement interpolation:

$$x_t = (1-t)x_0 + t x_1 \circ \sigma$$









Euclidean L^2 flow

$$\frac{\partial f}{\partial t} = -\mathcal{E}'(f)$$

$$\mathcal{E}(f) = \int \|\nabla f\|^2$$

$$\min_f \mathcal{E}(f)$$

Optimal transport flow

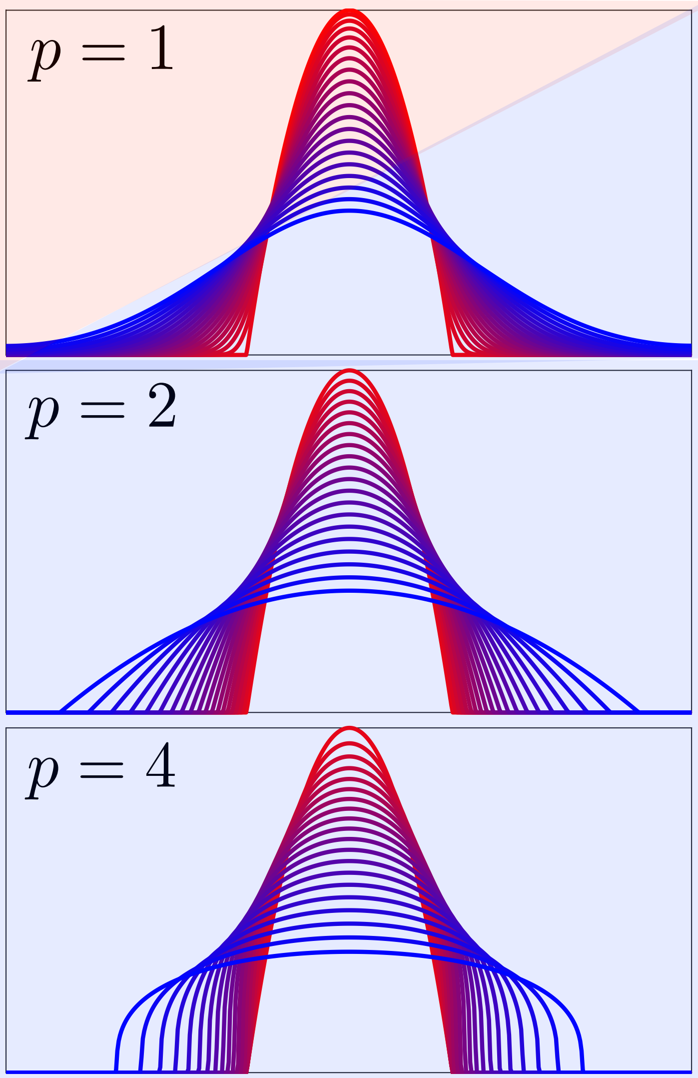
$$\frac{\partial f}{\partial t} = \mathrm{div}(f\nabla(\mathcal{E}'(f)))$$

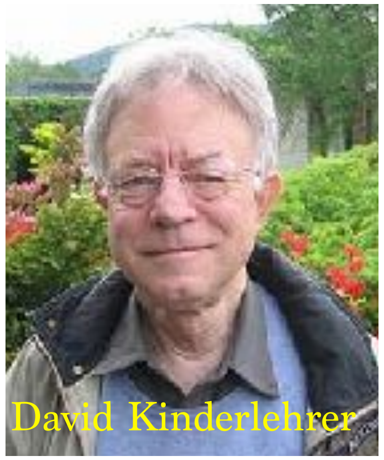
$$\mathcal{E}(f) = -\int f(\log(f) - 1)$$

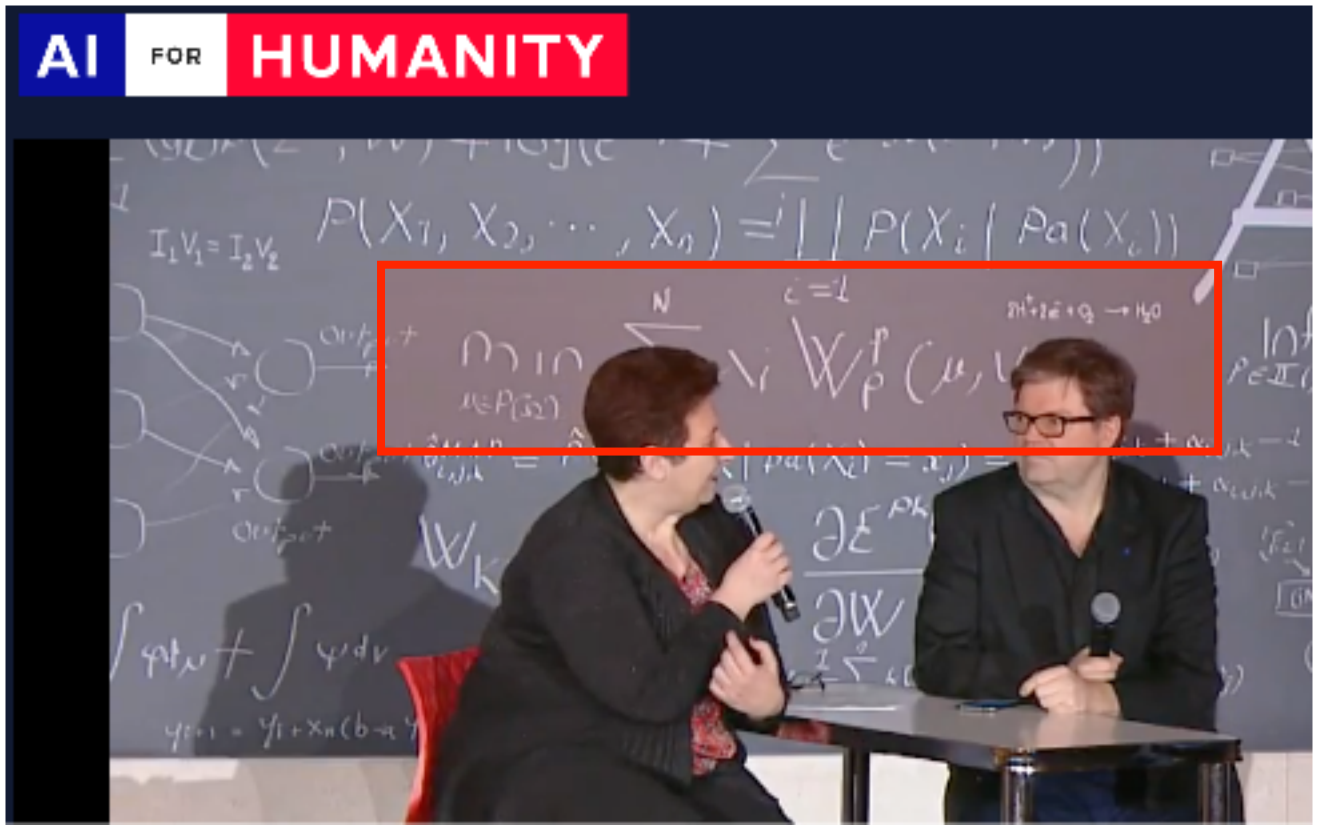
Heat equation: $\dfrac{\partial f}{\partial t} = \Delta f$

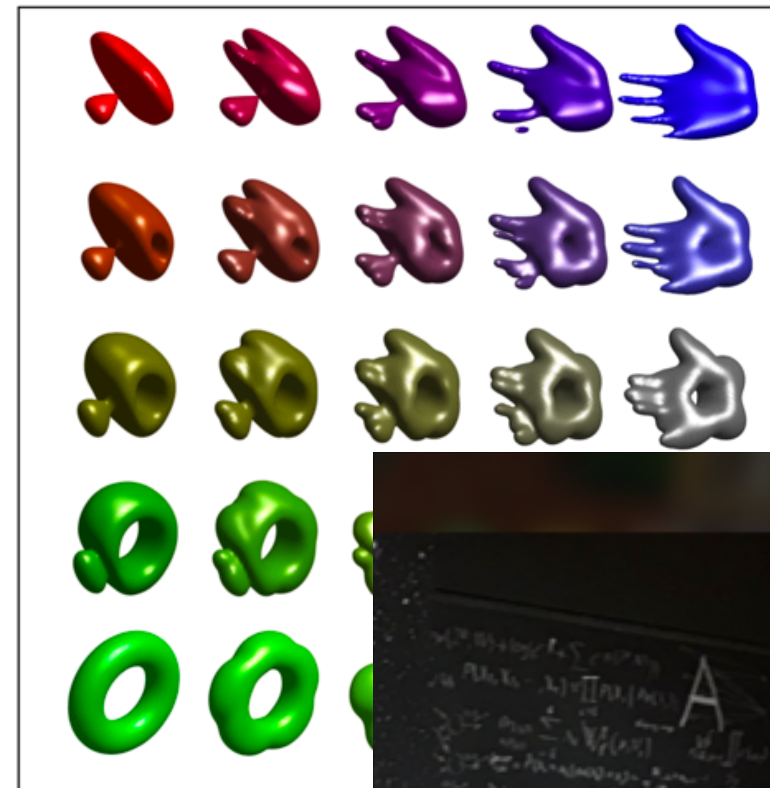
$$\mathcal{E}(f) = \int f \frac{f^{p-1} - p}{p-1}$$

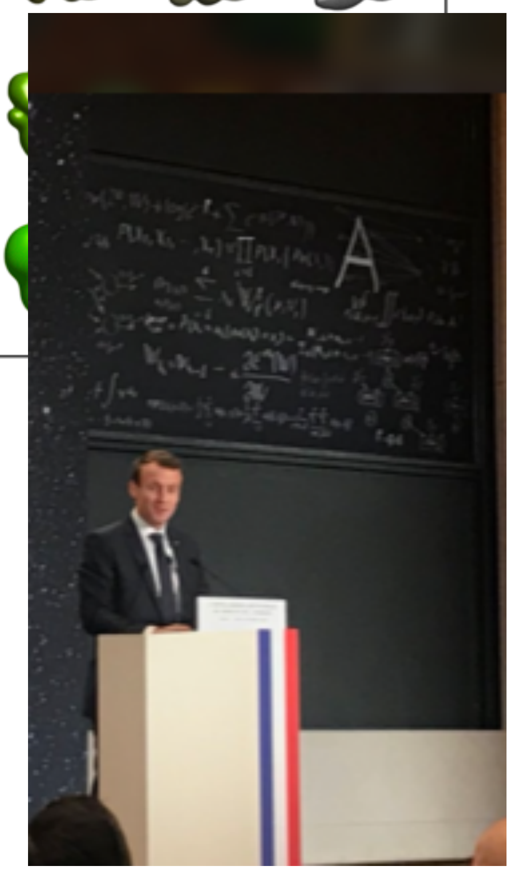
Porous medium: $\dfrac{\partial f}{\partial t} = \Delta f^p$





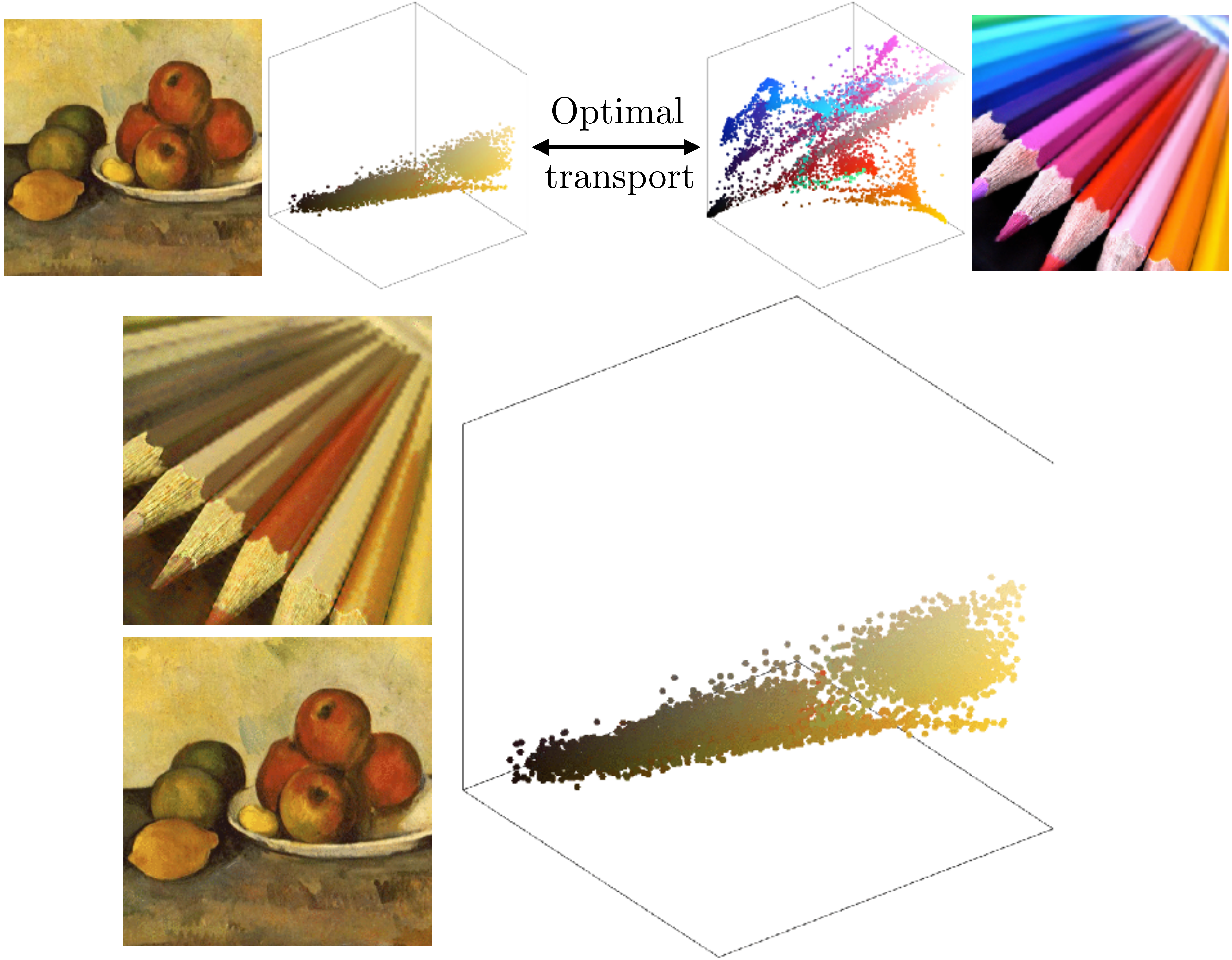






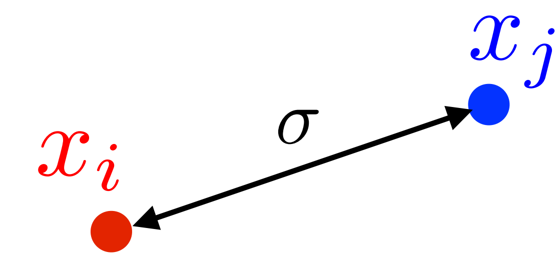
$$\boxed{\min_{\mu} \sum_{i} \lambda_i W_p^p(\mu, \nu_i)}$$

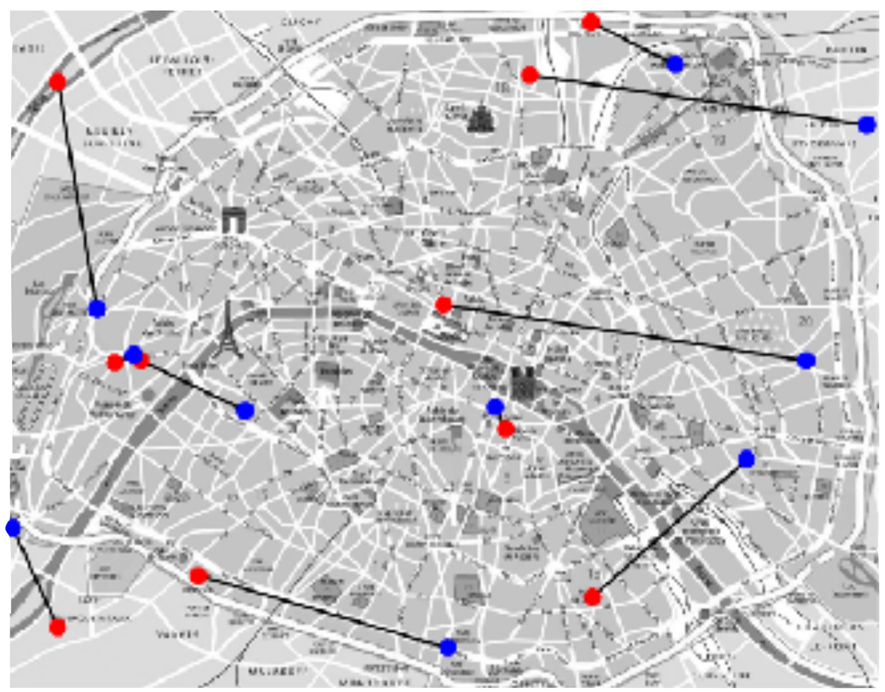
$$W(\mu,\nu) \overset{\text{def.}}{=} \max_{(\varphi,\psi)\in\mathcal{C}(\mathcal{X})\times\mathcal{C}(\mathcal{Y})} \left\{ \int_{\mathcal{X}} \varphi \mathrm{d}\mu + \int_{\mathcal{Y}} \psi \mathrm{d}\nu \;;\; \varphi(x) + \psi(y) \leqslant c(x,y) \right\}$$

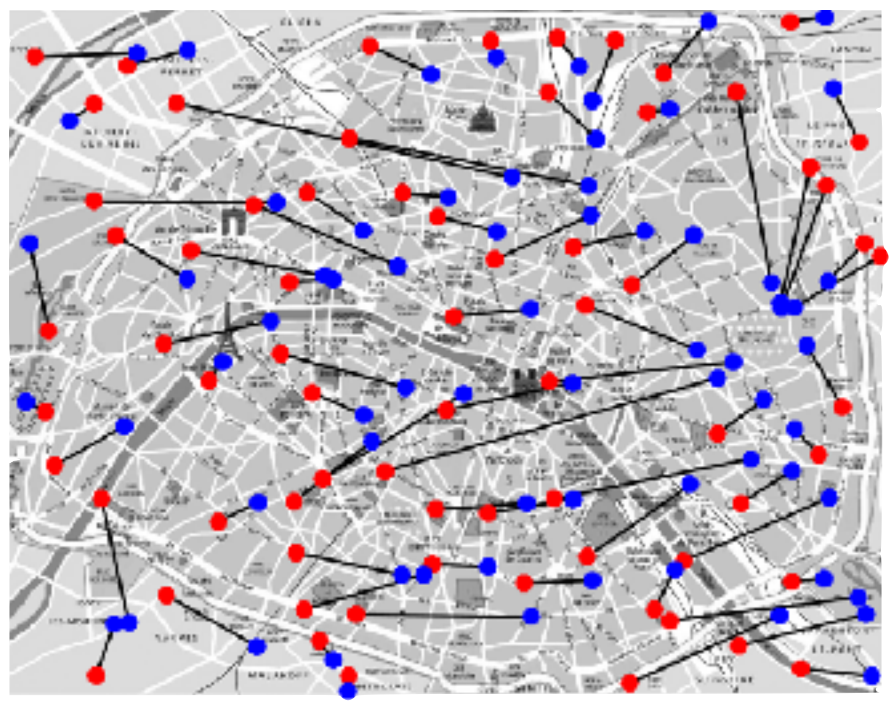


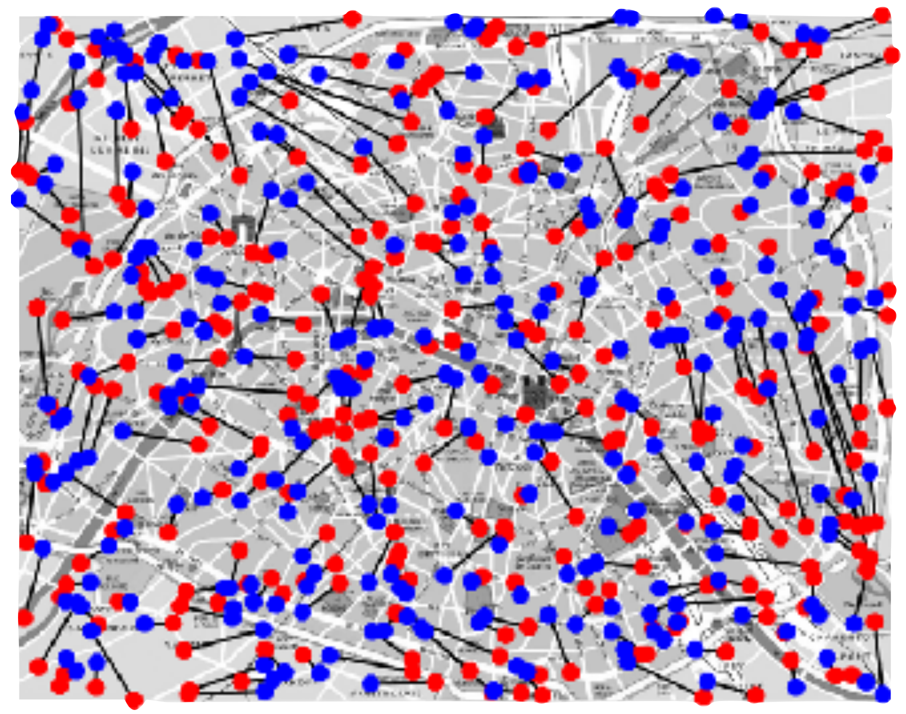
Optimal
transport

Monge problem: $\min_{\sigma \in \Sigma_n} \sum_{i=1}^{n} \|x_i - y_{\sigma(i)}\|$




$n = 10$


$n = 70$


$n = 300$

Proposition: Two optimal segments $[x_i, y_{\sigma(i)}]$ and $[x_{i'}, y_{\sigma(i')}]$ cannot cross.