Numerical Optimal Transport

http://optimaltransport.github.io

Gradient Flows

Gabriel Peyré

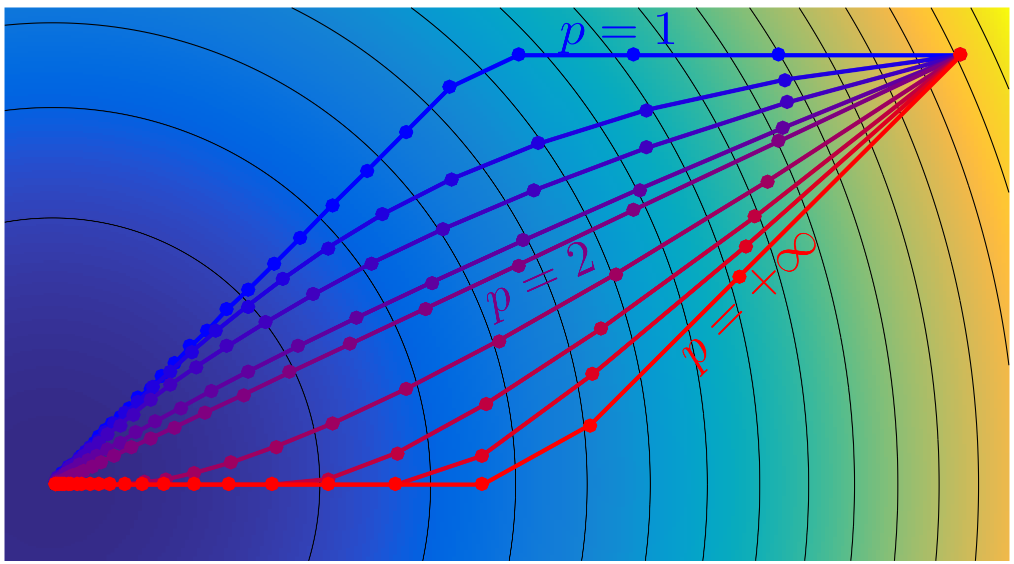
www.numerical-tours.com

Overview

- Gradient Flows
- Lagrangian Discretization
- Eulerian Discretization and Entropic Regularization
- Unbalanced OT
- Generalized Sinkhorn

Implicit Euler Stepping

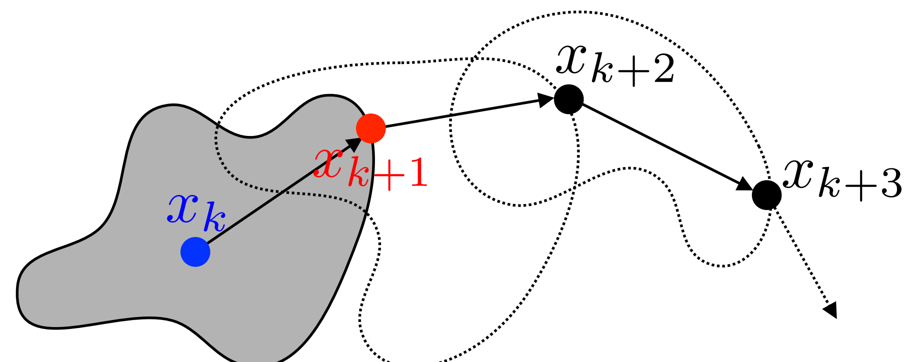
Metric space $(\mathcal{X}, d)$, minimize $F(x)$ on $\mathcal{X}$.



$F(x) = \|x\|^2$ on $(\mathcal{X} = \mathbb{R}^2, \|\cdot\|_p)$

Implicit Euler step:

$$x_{k+1} \overset{\text{def.}}{=} \underset{x \in \mathcal{X}}{\operatorname{argmin}} \; d(x_k, x)^2 + \tau F(x)$$



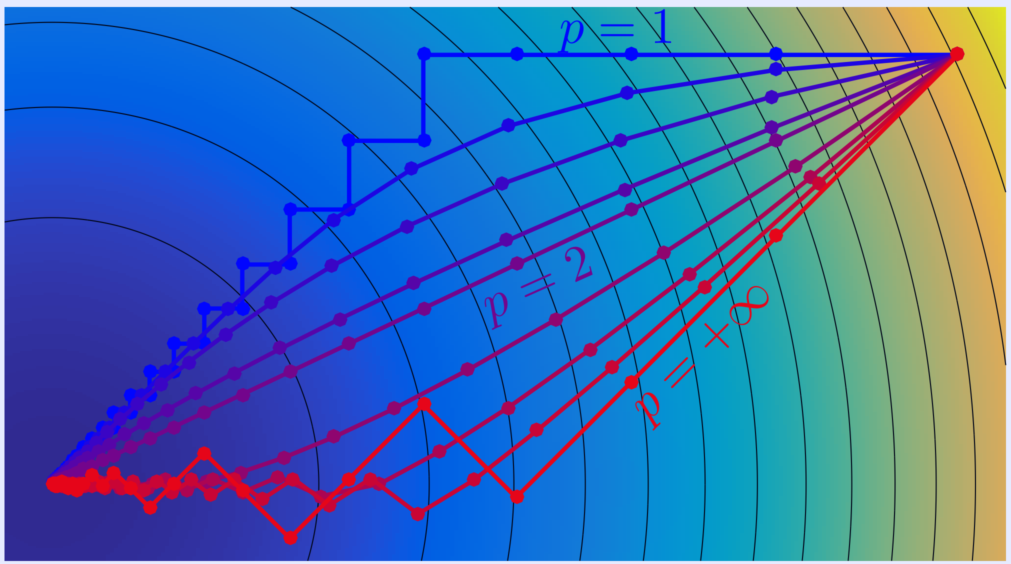
$\{x \; ; \; d(x_k, x) \sim \tau\}$

Implicit vs. Explicit Stepping

Metric space $(\mathcal{X}, d)$, minimize $F(x)$ on $\mathcal{X}$.

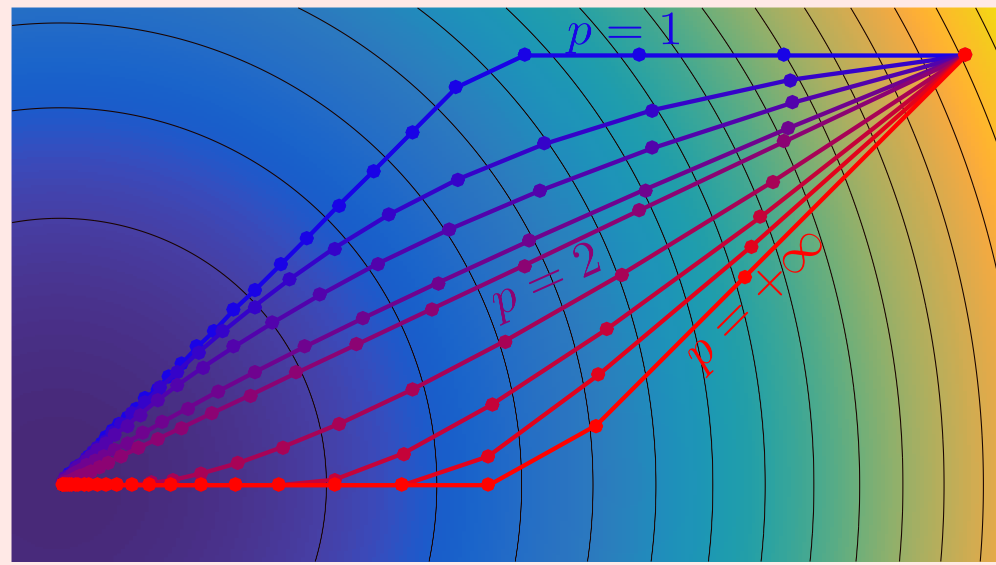
Explicit

$$x_{k+1} = \underset{x\in\mathcal{X}}{\text{argmin}}\ d(x_k, x)^2 + \tau\langle\nabla F(x_k),\, x\rangle$$



Implicit

$$x_{k+1} = \underset{x\in\mathcal{X}}{\text{argmin}}\ d(x_k, x)^2 + \tau F(x)$$



$F(x) = \|x\|^2$ on $(\mathcal{X} = \mathbb{R}^2, \|\cdot\|_p)$

Wasserstein Gradient Flows

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

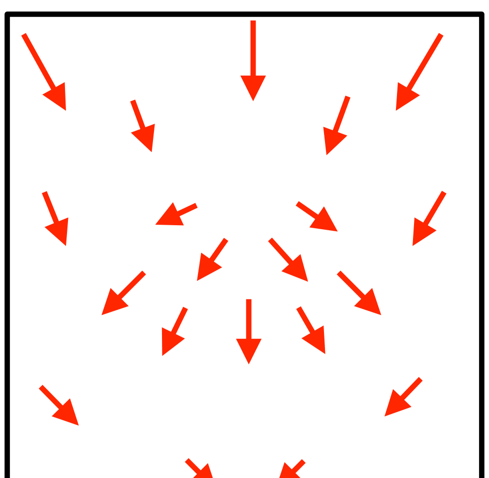
$$\boxed{\mu_{t+1} = \mathrm{Prox}^{W}_{\tau f}(\mu_t) \overset{\text{def.}}{=} \underset{\mu \in \mathcal{M}_+(X)}{\operatorname{argmin}} \; W_2^2(\mu_t, \mu) + \tau f(\mu)}$$

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

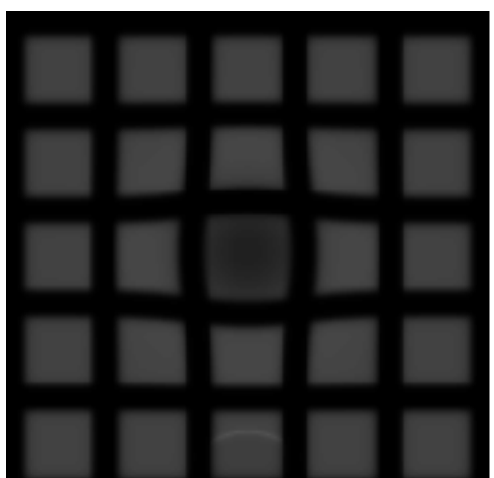
$f(\mu) = \int \log(\frac{\mathrm{d}\mu}{\mathrm{d}x})\mathrm{d}\mu \longrightarrow \partial_t \mu = \Delta \mu$ (heat diffusion)

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

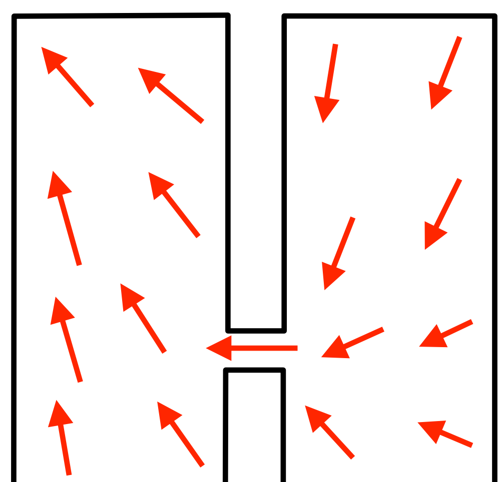
$f(\mu) = \frac{1}{m-1} \int (\frac{\mathrm{d}\mu}{\mathrm{d}x})^{m-1}\mathrm{d}\mu \longrightarrow \partial_t \mu = \Delta \mu^m$ (non-linear diffusion)



∇w



Evolution μ_t



∇w

Evolution μ_t

Euclidean L^2 flow

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

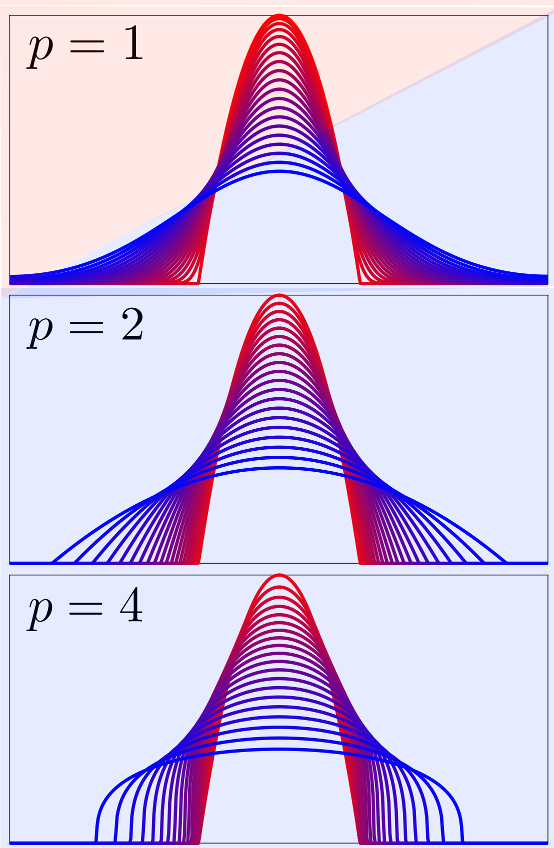
$$\min_f \mathcal{E}(f)$$

Optimal transport flow

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

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

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

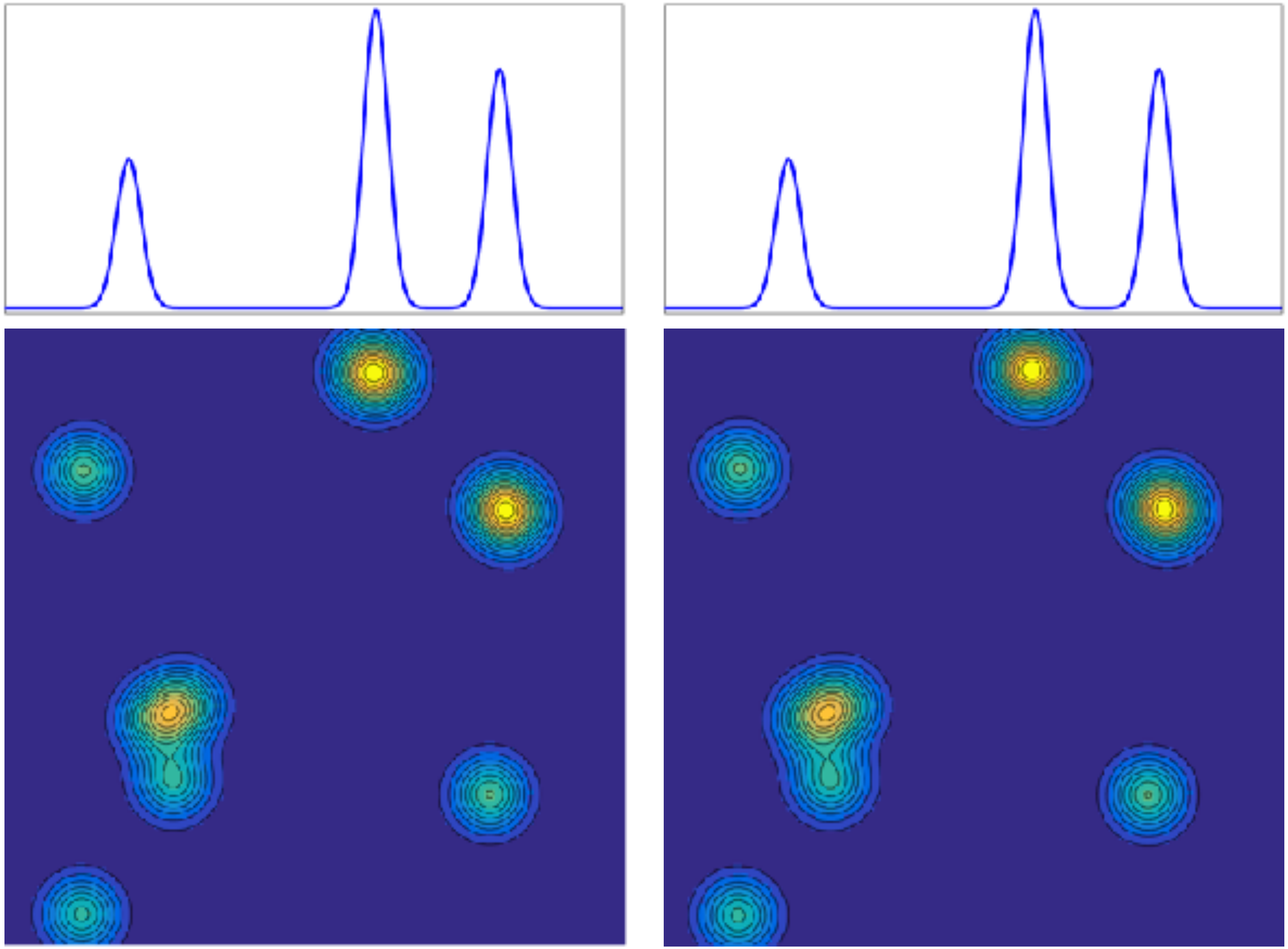


Heat equation: $\dfrac{\partial f}{\partial t} = \Delta f$

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

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

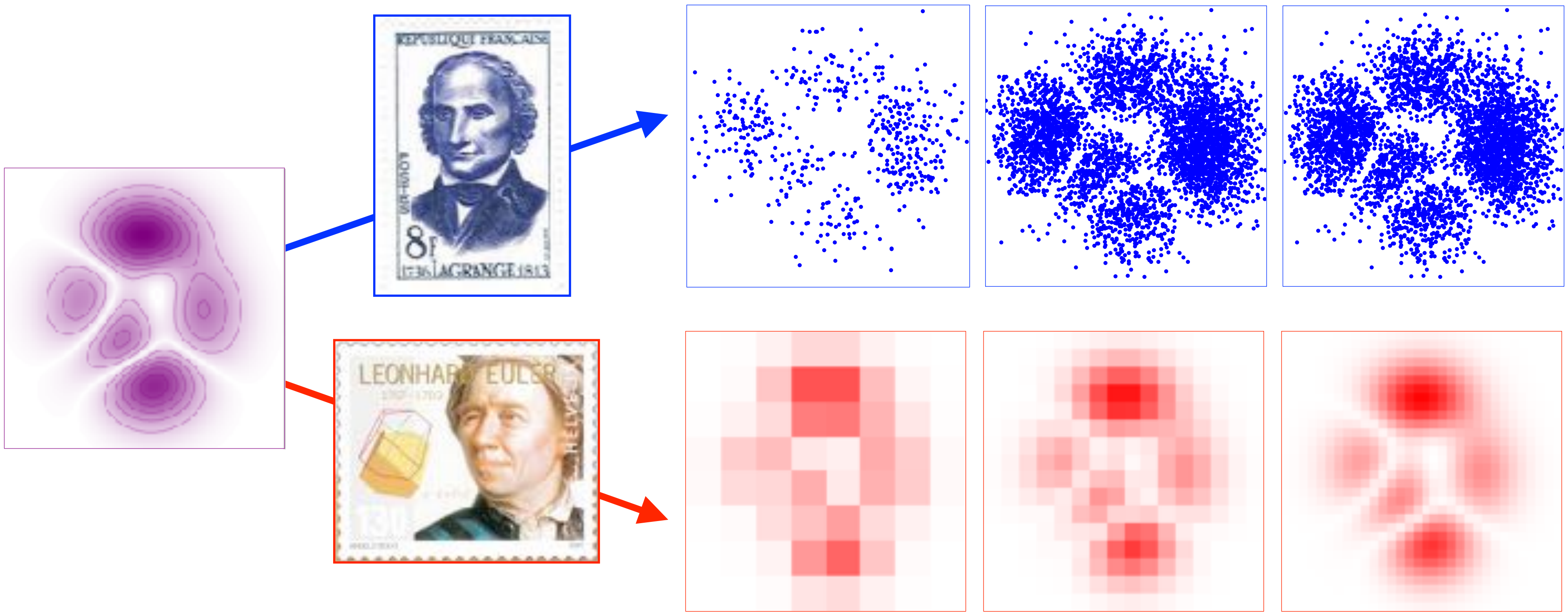
Linear vs Non-linear Diffusions



Overview

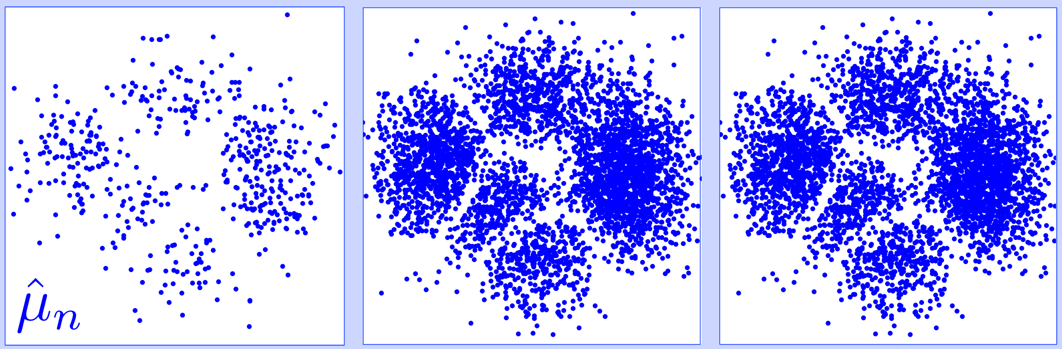
- Gradient Flows
- **Lagrangian Discretization**
- Eulerian Discretization and Entropic Regularization
- Unbalanced OT
- Generalized Sinkhorn

Eulerian vs. Lagrangian Discretization

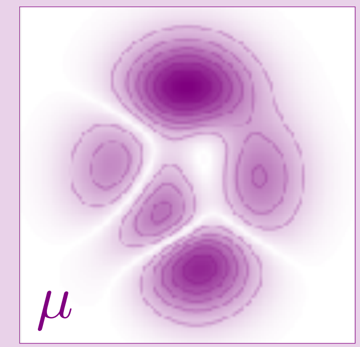


Lagrangian Discretization of Entropy

$$\hat{\mu}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$$



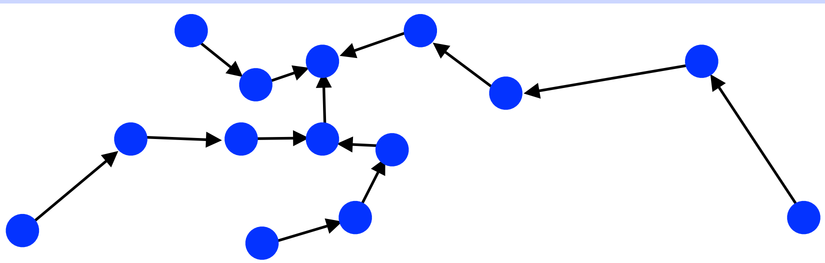
$n \to +\infty$



$$\hat{H}(\hat{\mu}_n) \overset{\text{def.}}{=} \sum_i \log(\min_{j \neq i} \|x_i - x_j\|)$$

$n \to +\infty$

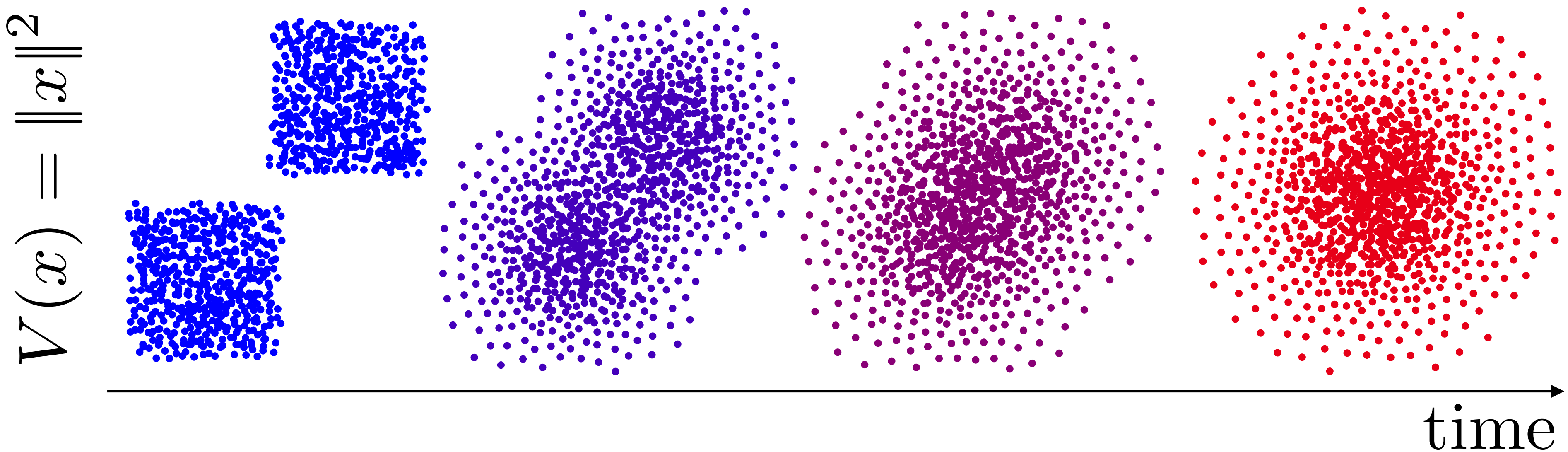
$$H(\mu) \overset{\text{def.}}{=} -\int \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}x}(x)\right)\mathrm{d}\mu(x)$$



Lagrangian Discretization of Gradient Flows

$$\min_{\rho} E(\rho) \stackrel{\text{def.}}{=} \int V(x)\rho(x)\mathrm{d}x + \int \rho(x)\log(\rho(x))\mathrm{d}x$$

Wasserstein flow of E: $\quad \dfrac{\mathrm{d}\rho_t}{\mathrm{d}t} = \Delta\rho_t + \nabla(V\rho_t)$



Overview

- Gradient Flows
- Lagrangian Discretization
- **Eulerian Discretization and Entropic Regularization**
- Unbalanced OT
- Generalized Sinkhorn

Gradient Flows: Crowd Motion

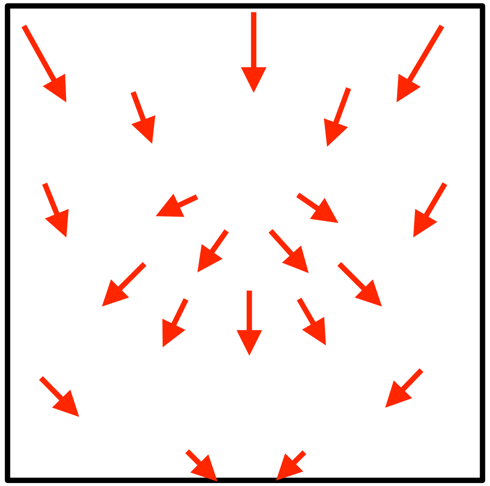
$$\mu_{t+1} \overset{\text{def.}}{=} \text{argmin}_{\mu}\, W_\alpha^\alpha(\mu_t, \mu) + \tau f(\mu)$$

Congestion-inducing function:

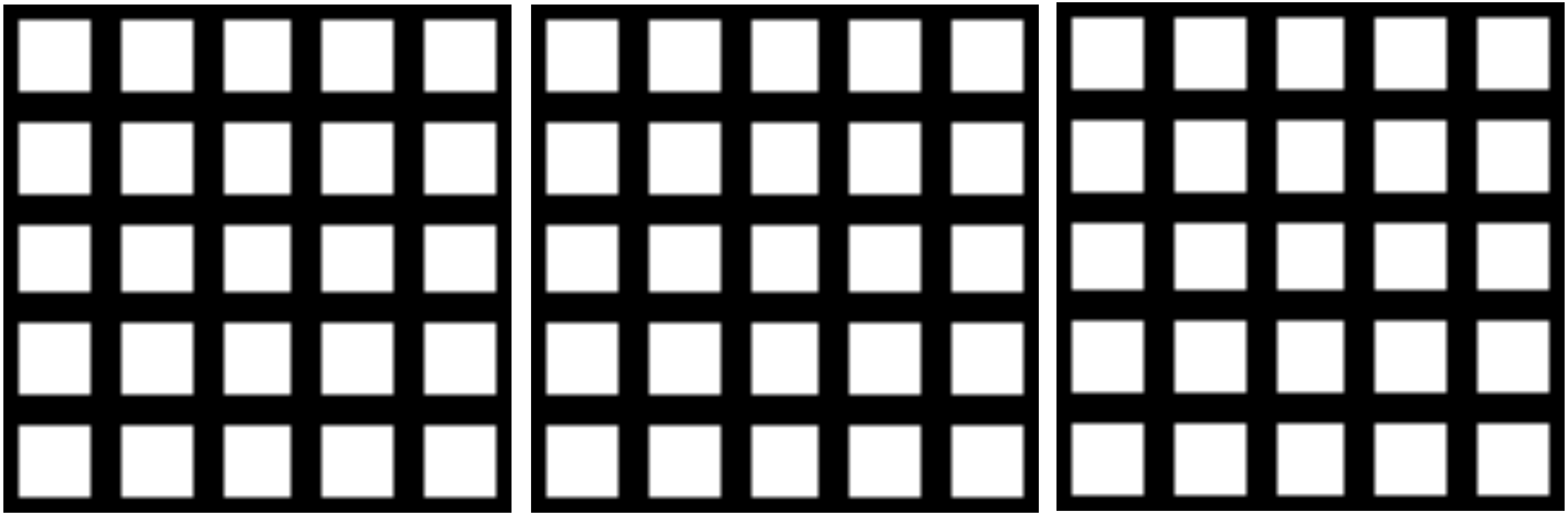
$$f(\mu) = \iota_{[0,\kappa]}(\mu) + \langle w,\, \mu \rangle$$

[Maury, Roudneff-Chupin, Santambrogio 2010]

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



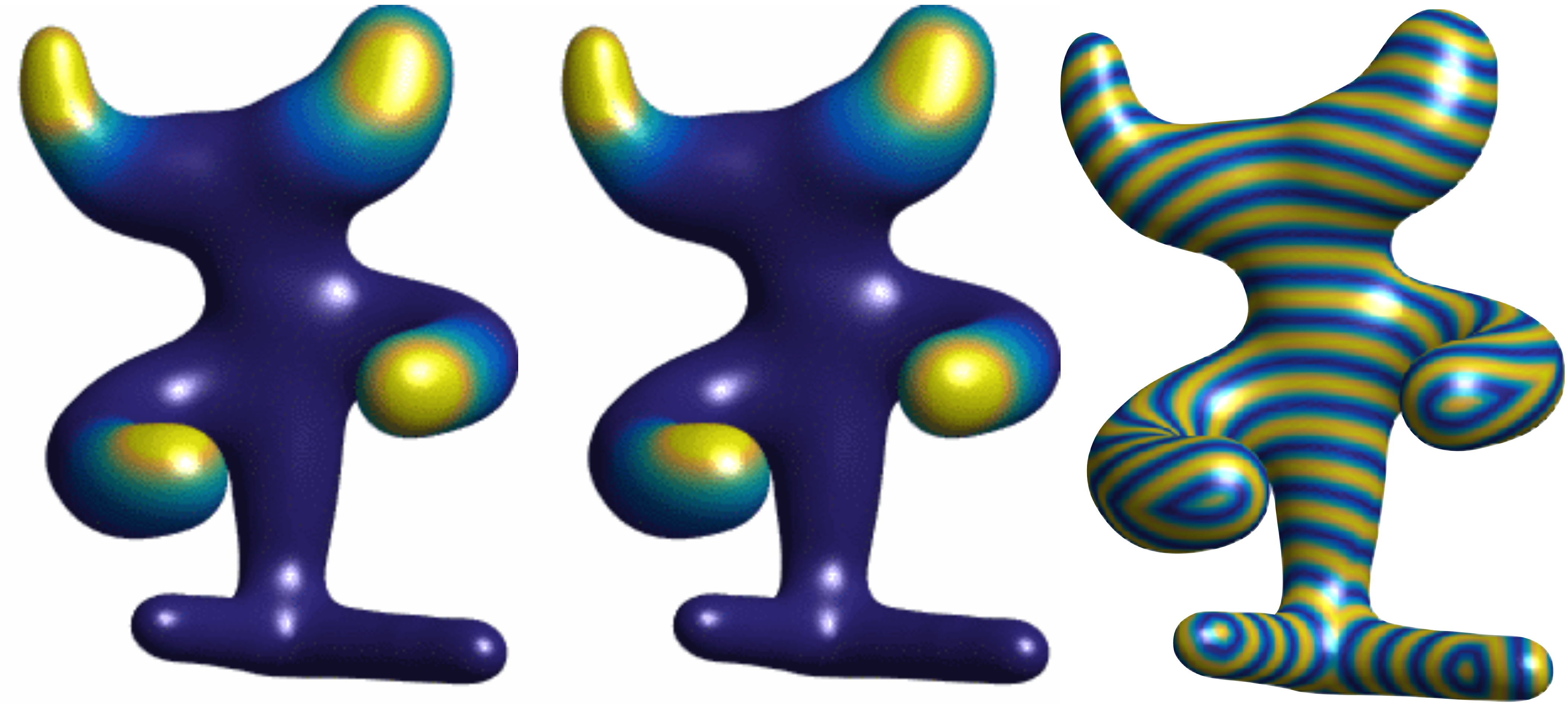
∇w



$\kappa = \|\mu_{t=0}\|_\infty$ $\quad$ $\kappa = 2\|\mu_{t=0}\|_\infty$ $\quad$ $\kappa = 4\|\mu_{t=0}\|_\infty$

Crowd Motion on a Surface

X = triangulated mesh.



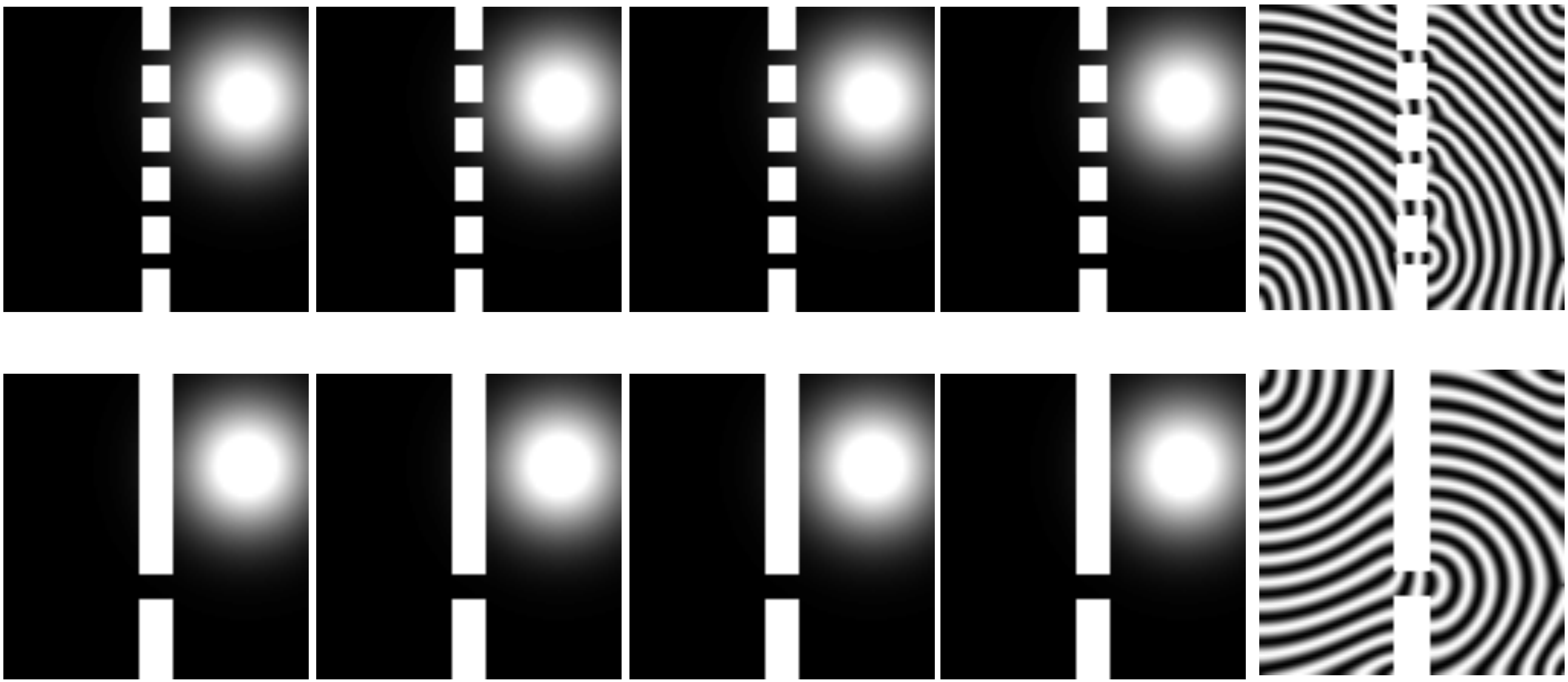
$\kappa = \|\mu_{t=0}\|_\infty$

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

Potential $\cos(w)$

Gradient Flows: Crowd Motion with Obstacles

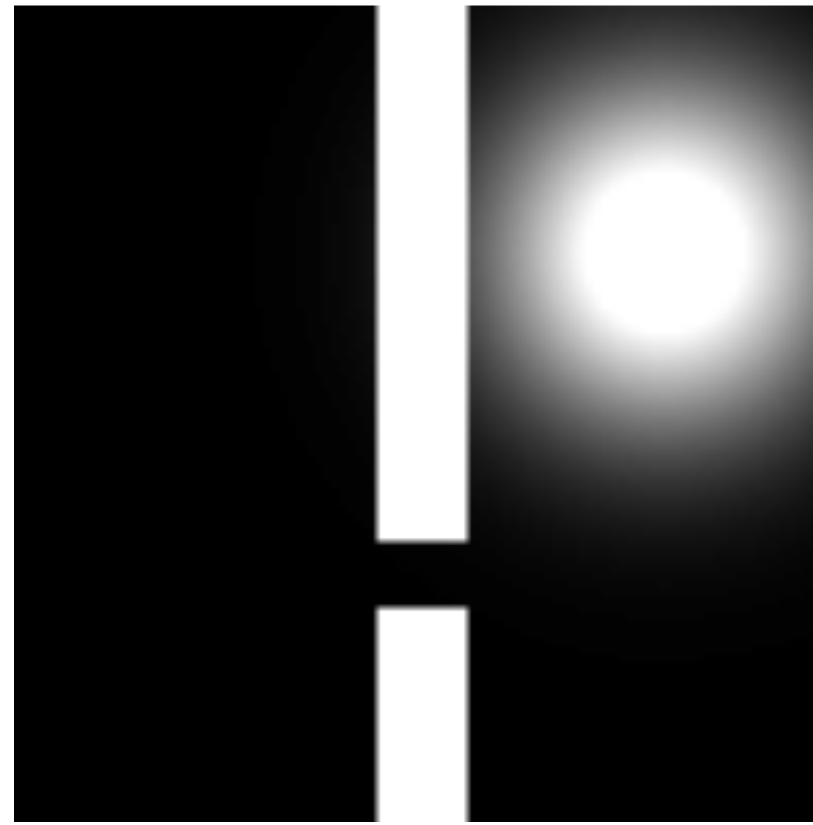
$$X = \text{sub-domain of } \mathbb{R}^2.$$



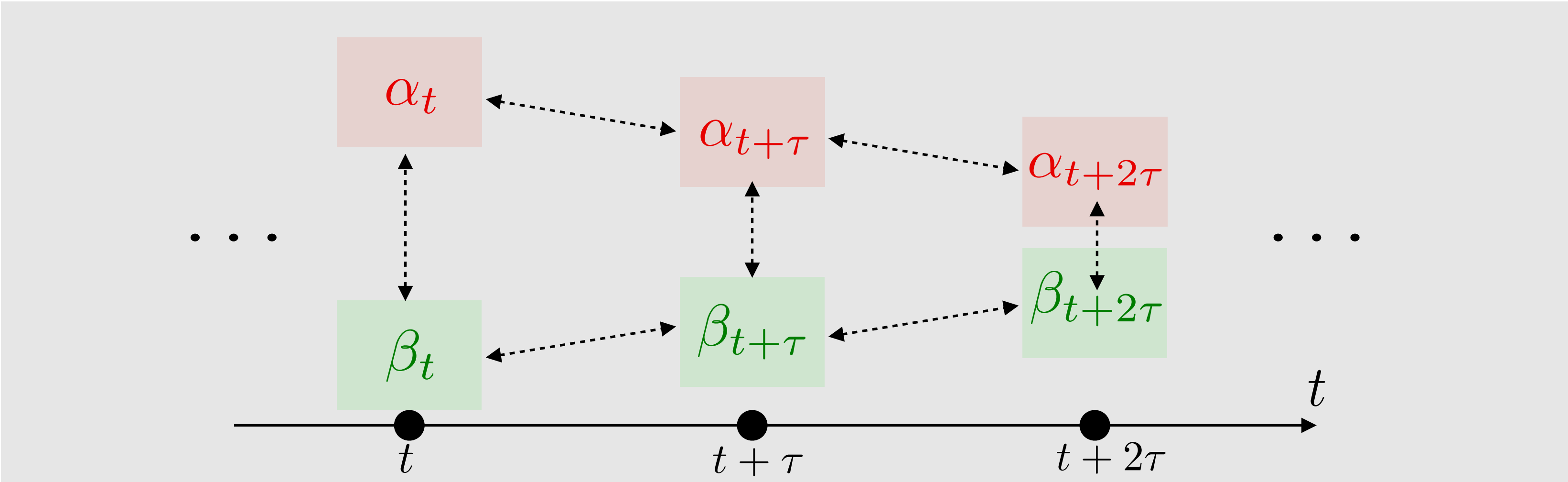
$\kappa = \|\mu_{t=0}\|_\infty$ $\kappa = 2\|\mu_{t=0}\|_\infty$ $\kappa = 4\|\mu_{t=0}\|_\infty$ $\kappa = 6\|\mu_{t=0}\|_\infty$ Potential $\cos(w)$

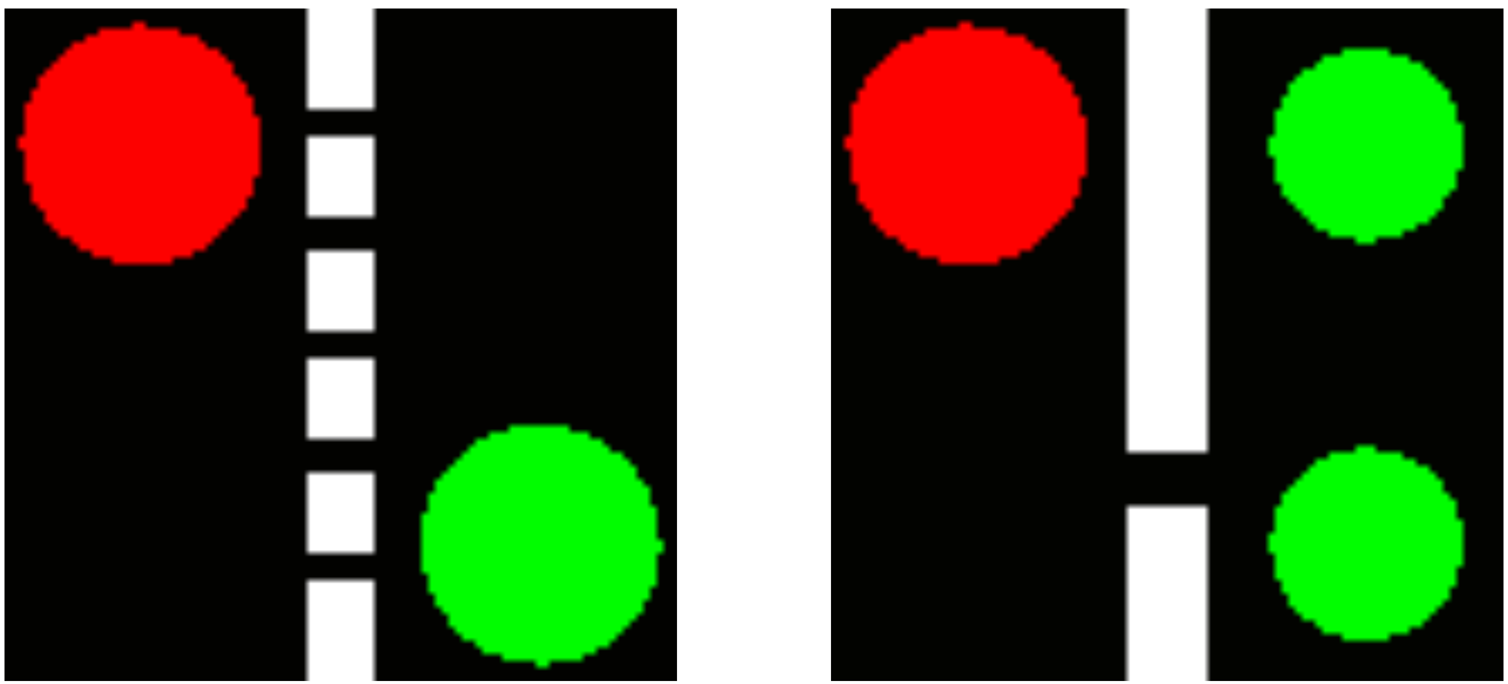
Crowd of Sheeps

https://www.youtube.com/watch?v=tDQw21ntR64
Tim Whittaker (New Zealand)



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





Overview

- Gradient Flows
- Lagrangian Discretization
- Eulerian Discretization and Entropic Regularization
- **Unbalanced OT**
- Generalized Sinkhorn

Unbalanced Transport

$$(\xi, \mu) \in \mathcal{M}_+(X)^2, \quad \mathrm{KL}(\xi|\mu) \overset{\text{def.}}{=} \int_X \log\left(\frac{\mathrm{d}\xi}{\mathrm{d}\mu}\right) \mathrm{d}\mu + \int_X (\mathrm{d}\mu - \mathrm{d}\xi)$$

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

[Liereo, Mielke, Savaré 2015]

Proposition: If $c(x, y) = -\log(\cos(\min(d(x, y), \frac{\pi}{2}))$ 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] [Kondratyev, Monsaingeon, Vorotnikov, 2015]

[Chizat, Schmitzer, P, Vialard 2015]

Overview

- Gradient Flows
- Lagrangian Discretization
- Eulerian Discretization and Entropic Regularization
- Unbalanced OT
- **Generalized Sinkhorn**

Generalized Entropic Regularization

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

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

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

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:

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

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

$\rightarrow$ Only matrix-vector multiplications. $\rightarrow$ Highly parallelizable.

$\rightarrow$ On regular grids: only convolutions! Linear time iterations.

Generalized Sinkhorn

$$\min_{\mathbf{P}} \sum_{i,j} \mathbf{C}_{i,j}\mathbf{P}_{i,j} - \varepsilon\mathbf{H}(\mathbf{P}) + F(\mathbf{P}\mathbb{1}_m) + G(\mathbf{P}^{\mathrm{T}}\mathbb{1}_n)$$

$$\mathbf{u} \leftarrow \frac{\mathrm{Prox}_F^{\mathbf{KL}}(\mathbf{K}\mathbf{v})}{\mathbf{K}\mathbf{v}} \quad \text{and} \quad \mathbf{v} \leftarrow \frac{\mathrm{Prox}_G^{\mathbf{KL}}(\mathbf{K}^{\mathrm{T}}\mathbf{u})}{\mathbf{K}^{\mathrm{T}}\mathbf{u}}$$

$$\forall\, \mathbf{u} \in \mathbb{R}_+^N, \quad \mathrm{Prox}_F^{\mathbf{KL}}(\mathbf{u}) = \underset{\mathbf{u}' \in \mathbb{R}_+^N}{\mathrm{argmin}}\; \mathbf{KL}(\mathbf{u}'|\mathbf{u}) + F(\mathbf{u}')$$