Numerical Optimal Transport

http://optimaltransport.github.io

Algorithmic Foundations

Gabriel Peyré

www.numerical-tours.com

Overview

- **Linear Programming**
- PDE-based
- Semi-discrete
- Entropic Regularization

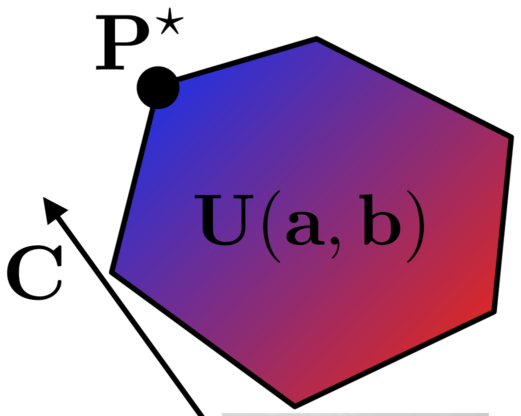
Linear Programming

Transportation polytope:

$$\mathbf{U}(\mathbf{a},\mathbf{b}) \overset{\text{def.}}{=} \left\{ \mathbf{P} \in \mathbb{R}_+^{n\times m} \; ; \; \mathbf{P}\mathbb{1}_n = \mathbf{a}, \mathbf{P}^\top \mathbb{1}_m = \mathbf{b} \right\}$$

Linear program:

$$\min \left\{ \sum_{i,j} \mathbf{P}_{i,j}\mathbf{C}_{i,j} \; ; \; \mathbf{P} \in \mathbf{U}(\mathbf{a},\mathbf{b}) \right\}$$



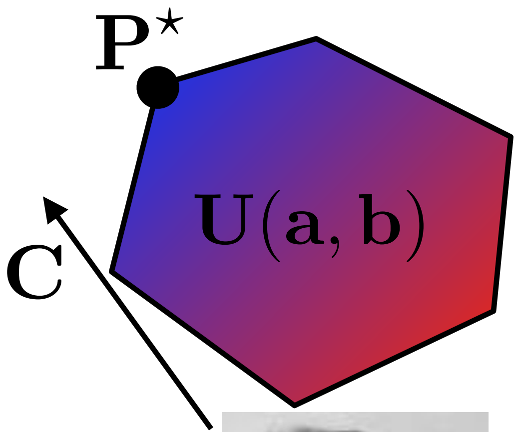
Linear Programming

Transportation polytope:

$$\mathbf{U}(\mathbf{a},\mathbf{b}) \overset{\text{def.}}{=} \left\{ \mathbf{P} \in \mathbb{R}_+^{n\times m} \; ; \; \mathbf{P}\mathbb{1}_n = \mathbf{a}, \mathbf{P}^\top \mathbb{1}_m = \mathbf{b} \right\}$$

Linear program:

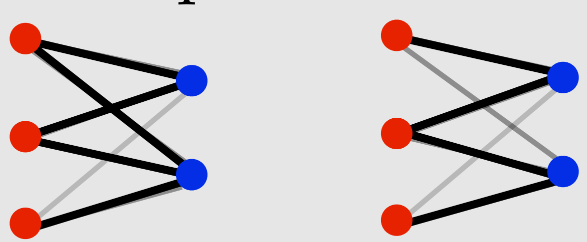
$$\min \left\{ \textstyle\sum_{i,j} \mathbf{P}_{i,j}\mathbf{C}_{i,j} \; ; \; \mathbf{P} \in \mathbf{U}(\mathbf{a},\mathbf{b}) \right\}$$



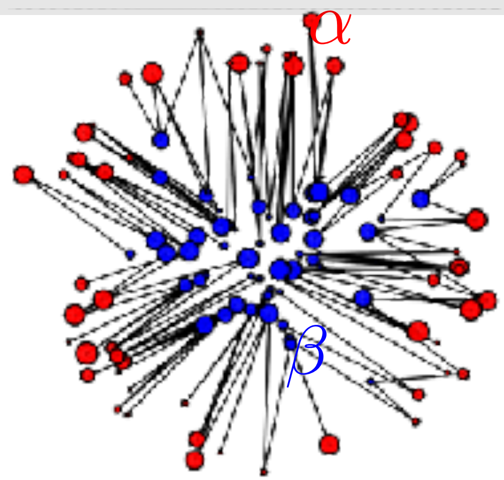
Theorem: $\exists \mathbf{P}^\star$ solution extremal point of $\mathbf{U}(\mathbf{a},\mathbf{b})$

$$|\{(i,j) \; ; \; \mathbf{P}^\star_{i,j} \neq 0\}| \leqslant n+m-1$$

Extremal points $\Leftrightarrow$ no cycle



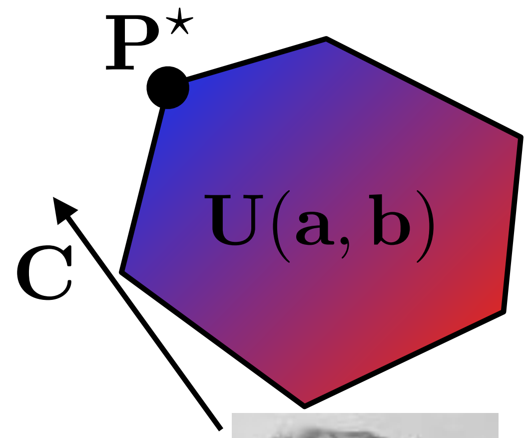
Non-extremal Extremal



Linear Programming

Transportation polytope:

$$\mathbf{U}(\mathbf{a},\mathbf{b}) \overset{\text{def.}}{=} \left\{ \mathbf{P} \in \mathbb{R}_+^{n\times m} \;;\; \mathbf{P}\mathbb{1}_n = \mathbf{a}, \mathbf{P}^\top \mathbb{1}_m = \mathbf{b} \right\}$$



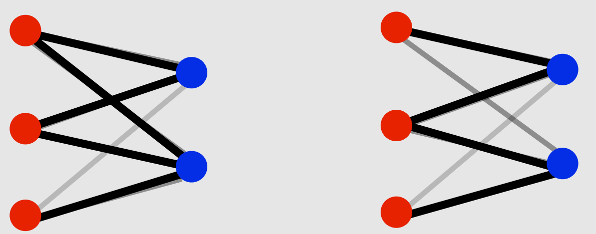
Linear program:

$$\min \left\{ \textstyle\sum_{i,j} \mathbf{P}_{i,j}\mathbf{C}_{i,j} \;;\; \mathbf{P} \in \mathbf{U}(\mathbf{a},\mathbf{b}) \right\}$$

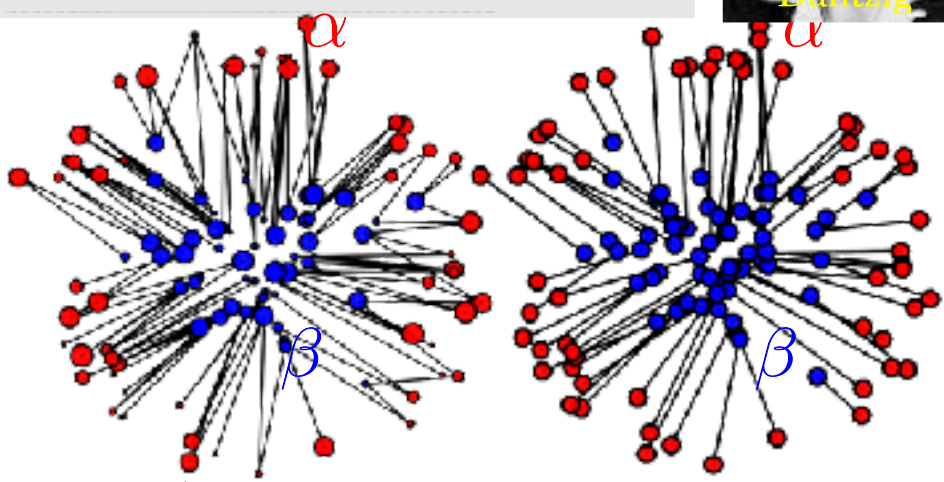
Theorem: $\exists \mathbf{P}^\star$ solution extremal point of $\mathbf{U}(\mathbf{a},\mathbf{b})$

$$|\{(i,j) \;;\; \mathbf{P}^\star_{i,j} \neq 0\}| \leqslant n+m-1$$

Extremal points $\Leftrightarrow$ no cycle



Non-extremal Extremal



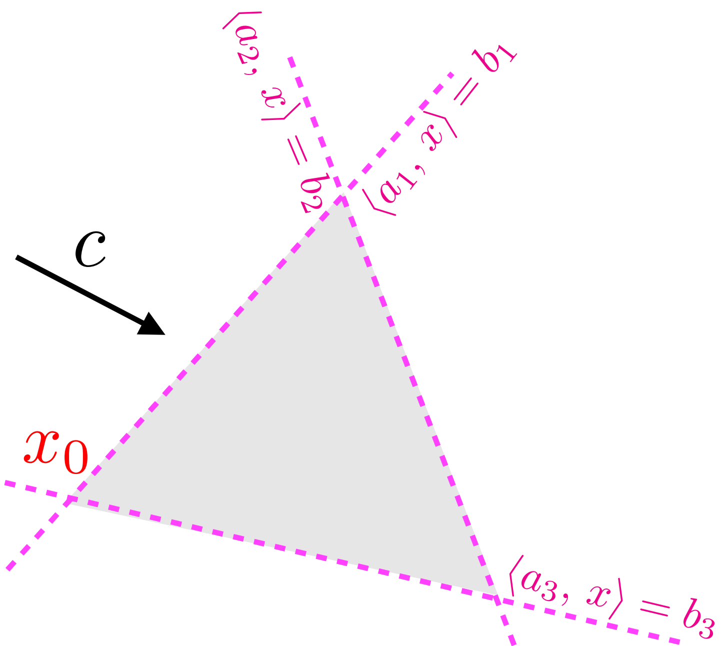
Example: if $n = n$, $\mathbf{a} = \mathbf{b} = \mathbb{1}/n$, $\mathbf{P}^\star$ permutation matrix.

$\rightarrow \sim n!$ extremal points.

Interior Point Methods

Linear programming:

$$x_0 \in \underset{x}{\operatorname{argmin}} \ \{\langle x,\, c\rangle \ ; \ i = 1, \ldots, m, \langle a_i,\, x\rangle \leqslant b_i\}$$



Interior Point Methods

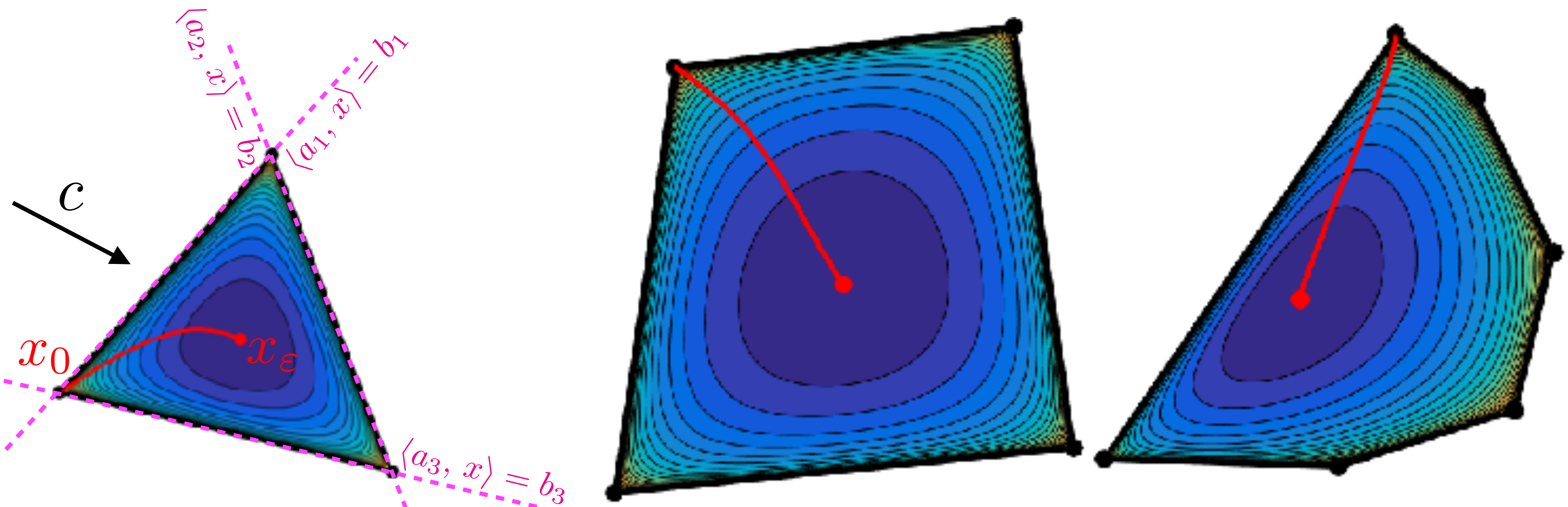
Linear programming:

$$x_0 \in \underset{x}{\operatorname{argmin}} \; \{\langle x,\, c\rangle \;;\; i = 1, \ldots, m, \langle a_i,\, x\rangle \leqslant b_i\}$$

$\varepsilon \to 0$

Log-barrier approximation:

$$x_\varepsilon \overset{\text{def.}}{=} \underset{x}{\operatorname{argmin}} \; \langle x,\, c\rangle - \varepsilon \textstyle\sum_i \log(b_i - \langle a_i,\, x\rangle)$$



Interior Point Methods

Linear programming:

$$x_0 \in \underset{x}{\text{argmin}} \ \{\langle x,\, c\rangle \ ; \ i = 1, \ldots, m, \langle a_i,\, x\rangle \leqslant b_i\}$$

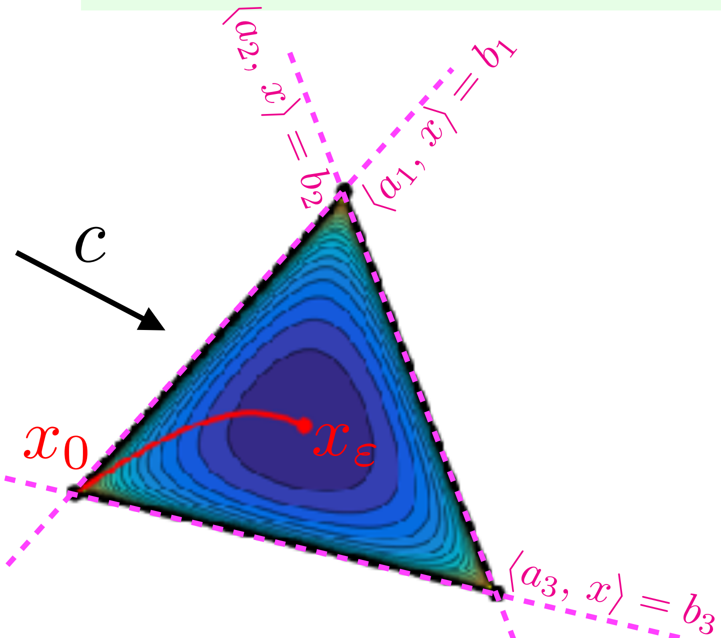
$\varepsilon \to 0$

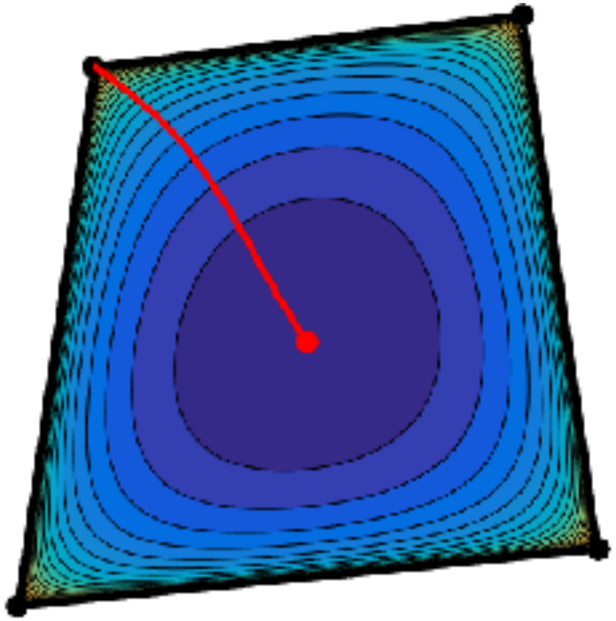
Log-barrier approximation:

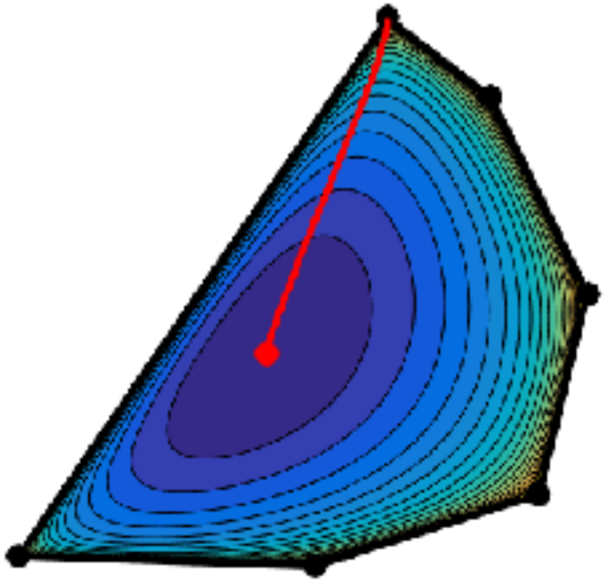
$$x_\varepsilon \overset{\text{def.}}{=} \underset{x}{\text{argmin}} \ \langle x,\, c\rangle - \varepsilon \textstyle\sum_i \log(b_i - \langle a_i,\, x\rangle)$$

Interior point method:

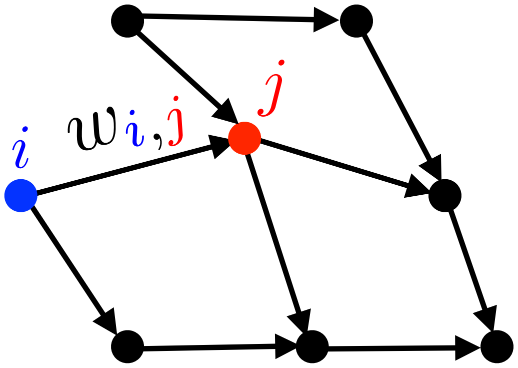
$O(\sqrt{m}\log(\frac{m}{\tau}))$ Newton iterations computes feasible $\hat{x}_\varepsilon$ with $\langle \hat{x}_\varepsilon - x_0,\, c\rangle \leqslant \tau$







Network Flow



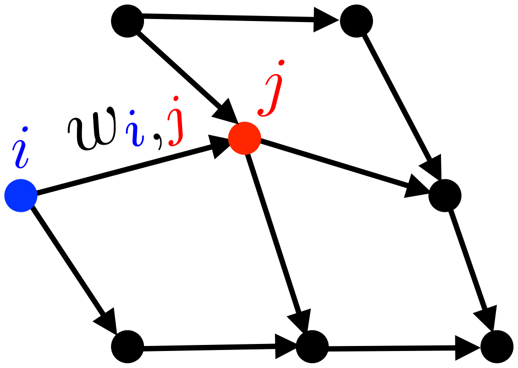
Divergence on a graph:

$$\text{div}(\mathbf{s})_i \overset{\text{def.}}{=} \sum_{(i,k)\in G} \mathbf{s}_{i,k} - \sum_{(k,i)\in G} \mathbf{s}_{k,i}$$

Min-cost flow:

$$\min_{\mathbf{s}\geqslant 0} \left\{ \langle \mathbf{s},\, w \rangle \;;\; \text{div}(\mathbf{s}) = h \right\}$$

Network Flow



Divergence on a graph:

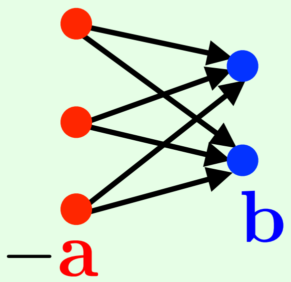
$$\text{div}(\mathbf{s})_i \overset{\text{def.}}{=} \sum_{(i,k)\in G} \mathbf{s}_{i,k} - \sum_{(k,i)\in G} \mathbf{s}_{k,i}$$

Min-cost flow:

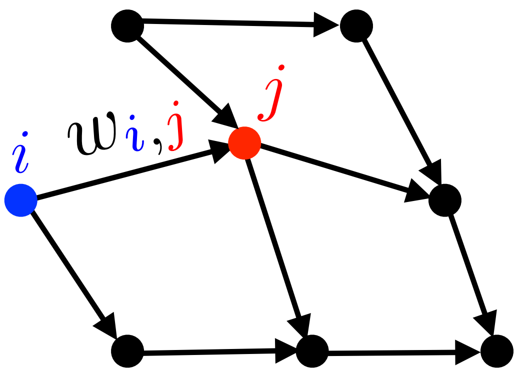
$$\min_{\mathbf{s}\geqslant 0} \ \{\langle \mathbf{s},\, w\rangle \ ; \ \text{div}(\mathbf{s}) = h\}$$

Optimal transport: bi-partite graph.

$$\mathbf{s} = \mathbf{P} \qquad h = (-\mathbf{a}, \mathbf{b}) \qquad w_{i,j} = \mathbf{C}_{i,j}$$



Network Flow



Divergence on a graph:

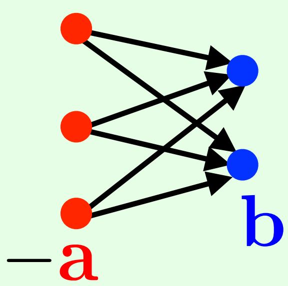
$$\text{div}(\mathbf{s})_i \overset{\text{def.}}{=} \sum_{(i,k)\in G} \mathbf{s}_{i,k} - \sum_{(k,i)\in G} \mathbf{s}_{k,i}$$

Min-cost flow:

$$\min_{\mathbf{s}\geqslant 0} \; \{\langle \mathbf{s},\, w\rangle \; ; \; \text{div}(\mathbf{s}) = h\}$$

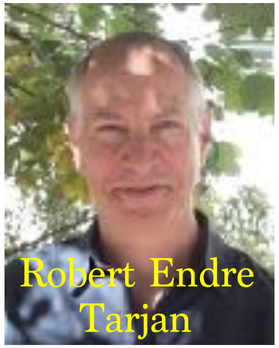
Optimal transport: bi-partite graph.

$$\mathbf{s} = \mathbf{P} \qquad h = (-\mathbf{a}, \mathbf{b}) \qquad w_{i,j} = \mathbf{C}_{i,j}$$



Theorem: on a graph with E edges and V vertices, $\exists$ a network simplex algorithm of complexity $O(VE \log V \log(V\|\mathbf{C}\|_\infty))$ if $\mathbf{C}_{i,j} \in \mathbb{Z}$.

OT simplex: $n = m$, complexity $O(n^3 \log(n)^2)$.

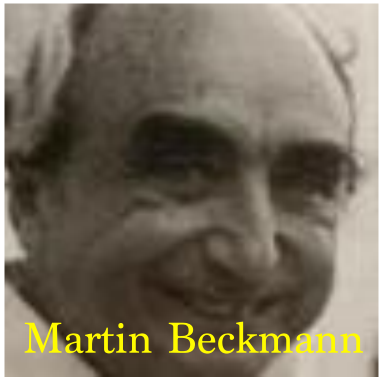


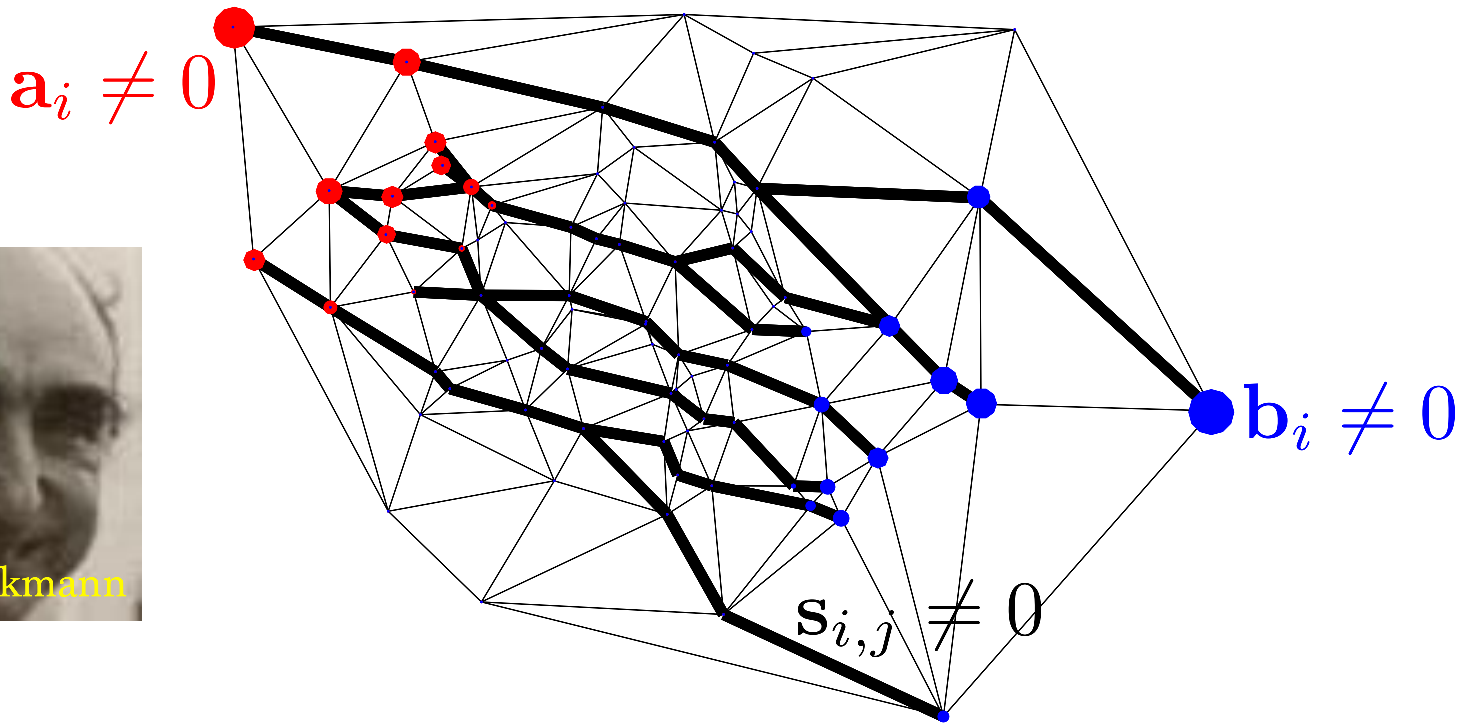
W1 as a Reduced Min-cost Flows

$$\mathbf{C}_{i,j} = \mathrm{GeodDist}_{\mathbf{w}}(i,j)$$

Proposition: [Beckmann]

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



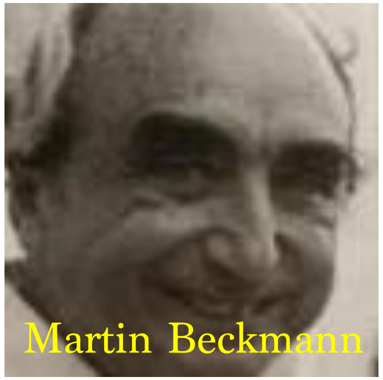


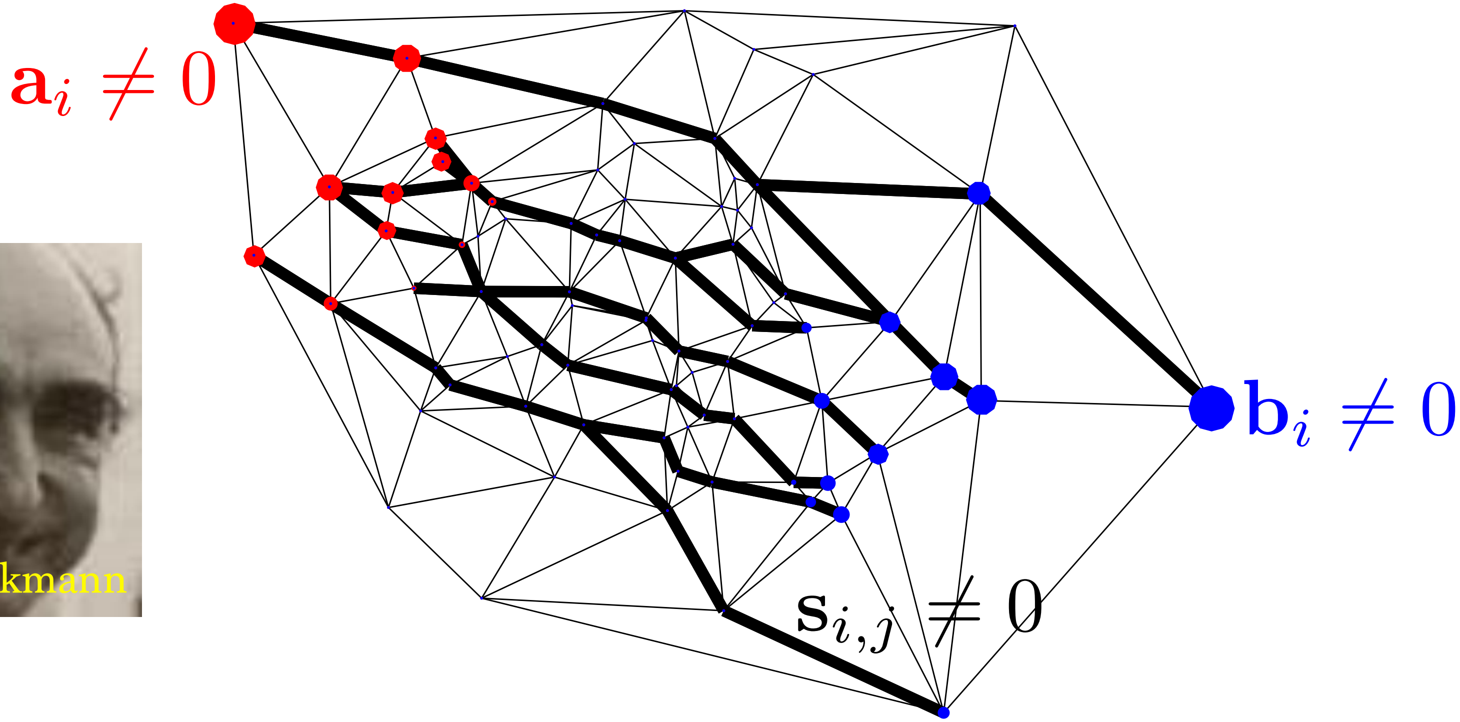
W1 as a Reduced Min-cost Flows

$$\mathbf{C}_{i,j} = \mathrm{GeodDist}_{\mathbf{w}}(i,j)$$

Proposition: [Beckmann]

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





Network simplex, $E, V = O(n)$ (e.g. regular graph):

W_p in $O(n^3 \log(n)^2)$ ⟶ W_1 in $O(n^2 \log(n)^2)$

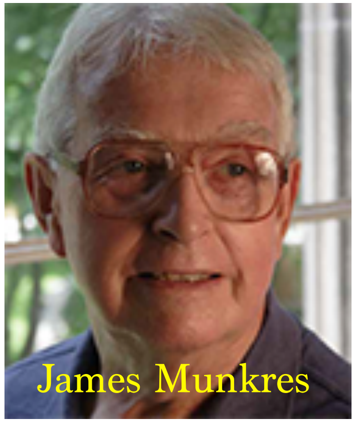
Hungarian and Auction Algorithms

→ Primal-dual algorithms.

→ Hungarian only works for $n = m$, $\mathbf{a} = \mathbf{b} = \mathbb{1}$.

→ Auction is approximate alternate c-transforms.

→ Complexity $O(n^3 \log(\|\mathbf{C}\|_\infty))$.

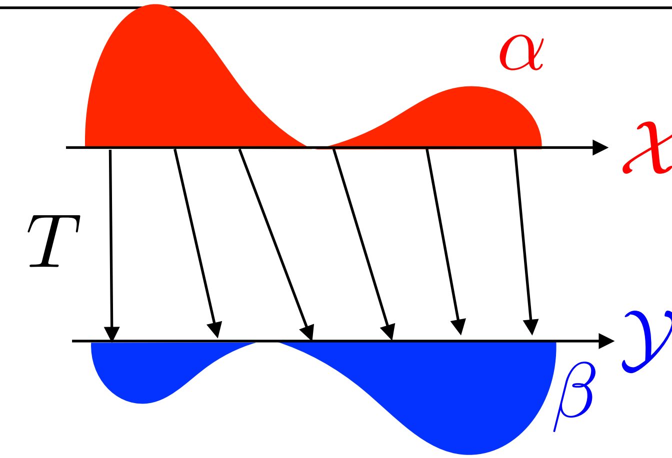


Overview

- Linear Programming
- **PDE-based**
- Semi-discrete
- Entropic Regularization

Monge-Ampère equation

$$\min_{\beta = T_\sharp \alpha} \int_{\mathcal{X}} \|x - T(x)\|^2 \mathrm{d}\alpha(x)$$



Densities: $\frac{\mathrm{d}\alpha}{\mathrm{d}x} = \rho_\alpha, \frac{\mathrm{d}\beta}{\mathrm{d}x} = \rho_\beta$

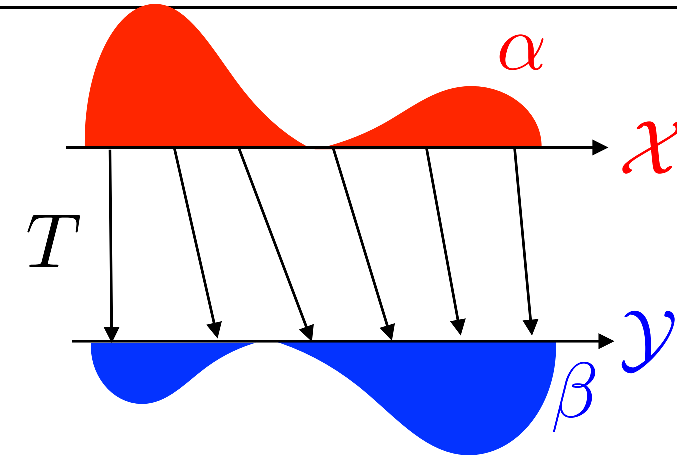
Theorem: [Brenier] Unique $T = \nabla\varphi$ solving

$$\det(\partial^2 \varphi(x)) \rho_\beta(\nabla\varphi(x)) = \rho_\alpha(x)$$

φ convex

Monge-Ampère equation

$$\min_{\beta = T_\sharp \alpha} \int_{\mathcal{X}} \|x - T(x)\|^2 \mathrm{d}\alpha(x)$$



Densities: $\frac{\mathrm{d}\alpha}{\mathrm{d}x} = \rho_\alpha, \frac{\mathrm{d}\beta}{\mathrm{d}x} = \rho_\beta$

Theorem: [Brenier] Unique $T = \nabla\varphi$ solving

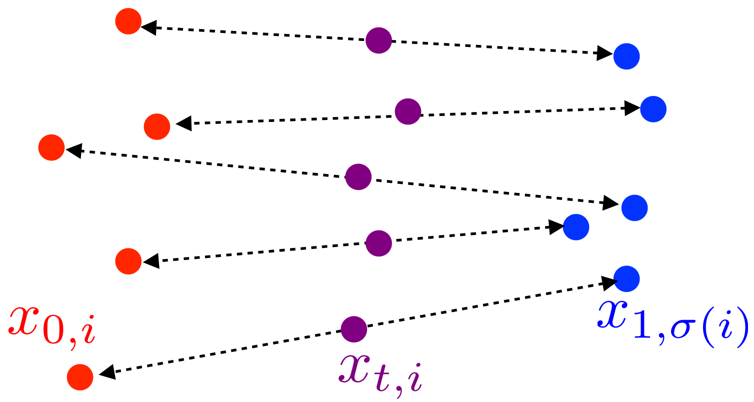
$$\det(\partial^2 \varphi(x))\rho_\beta(\nabla\varphi(x)) = \rho_\alpha(x) \qquad \varphi \text{ convex}$$

→ Finite-elements / finite-differences discretization of the cone of convex functions.

→ non-classical boundary conditions.

...

Displacement Interpolation

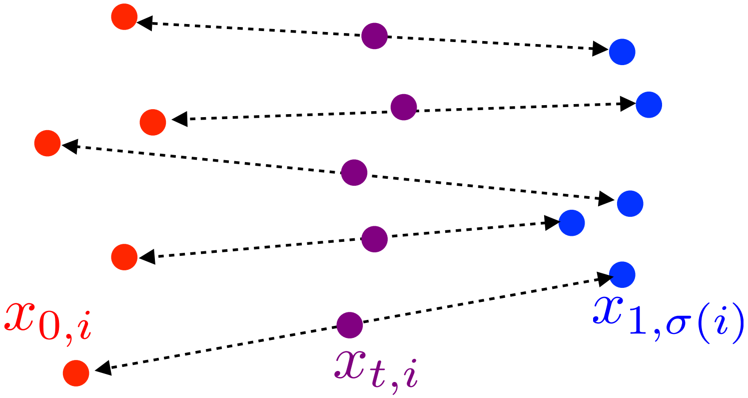


Optimal assignement: $\min_{\sigma} \|x_0 - x_1 \circ \sigma\|$

Displacement interpolation: $\alpha_t \overset{\text{def.}}{=} \frac{1}{n} \sum_i \delta_{x_{t,i}}$

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

Displacement Interpolation



Optimal assignement: $\min_{\sigma} \|x_0 - x_1 \circ \sigma\|$

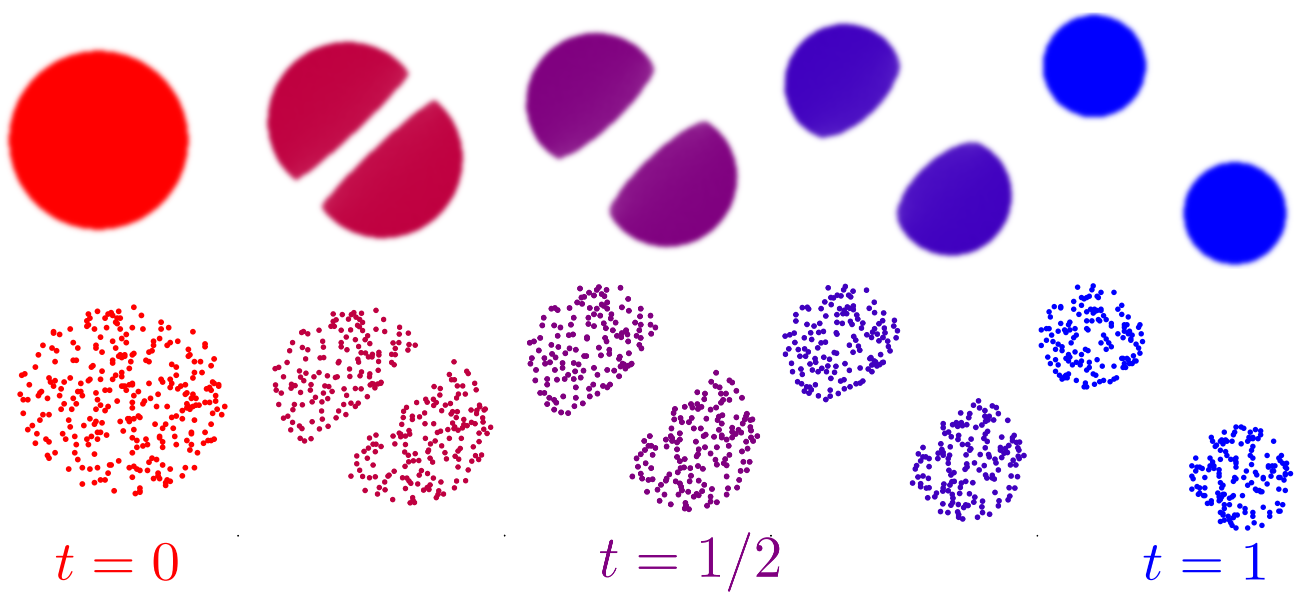
Displacement interpolation: $\alpha_t \stackrel{\text{def.}}{=} \frac{1}{n} \sum_i \delta_{x_{t,i}}$

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

Monge map $\psi_\sharp \alpha = \beta$:

$$\alpha_t \stackrel{\text{def.}}{=} ((1-t)\mathrm{Id} + t\psi)_\sharp \alpha = (t\mathrm{Id} + (1-t)\psi^{-1})_\sharp \beta$$

Optimal coupling $\pi \in \mathcal{U}(\alpha, \beta)$: $\alpha_t \stackrel{\text{def.}}{=} ((1-t)P_{\mathcal{X}} + tP_{\mathcal{Y}})_\sharp \pi$



Benamou-Brenier Formulation

Geodesic formulation:

$$\mathcal{W}_2^2(\alpha_0, \alpha_1) = \min \int_0^1 \int_{\mathbb{R}^d} \|v_t(x)\|^2 \, \mathrm{d}\alpha_t(x) \mathrm{d}t$$

$$\frac{\partial \alpha_t}{\partial t} + \mathrm{div}(\alpha_t v_t) = 0 \quad \text{and} \quad \alpha_{t=0} = \alpha_0, \alpha_{t=1} = \alpha_1$$

Benamou-Brenier Formulation

Geodesic formulation:

$$\mathcal{W}_2^2(\alpha_0, \alpha_1) = \min \int_0^1 \int_{\mathbb{R}^d} \|v_t(x)\|^2 \, \mathrm{d}\alpha_t(x)\mathrm{d}t$$

$$\frac{\partial \alpha_t}{\partial t} + \mathrm{div}(\alpha_t v_t) = 0 \quad \text{and} \quad \alpha_{t=0} = \alpha_0, \alpha_{t=1} = \alpha_1$$

Benamou-Brenier Formulation

Geodesic formulation:

$$\mathcal{W}_2^2(\alpha_0,\alpha_1) = \min \int_0^1 \int_{\mathbb{R}^d} \|v_t(x)\|^2 \, \mathrm{d}\alpha_t(x)\mathrm{d}t$$

$$\frac{\partial \alpha_t}{\partial t} + \mathrm{div}(\alpha_t v_t) = 0 \quad \text{and} \quad \alpha_{t=0} = \alpha_0, \alpha_{t=1} = \alpha_1$$

Momentum change of variable: $J_t \overset{\text{def.}}{=} \alpha_t v_t$

$$\mathcal{W}_2^2(\alpha_0,\alpha_1) = \min_{(\alpha_t, J_t)_t \in \mathcal{C}(\alpha_0,\alpha_1)} \int_0^1 \int_{\mathbb{R}^d} \theta(\alpha_t(x), J_t(x)) \mathrm{d}x\mathrm{d}t$$

$$\mathcal{C}(\alpha_0,\alpha_1) \overset{\text{def.}}{=} \left\{ (\alpha_t, J_t) \ : \ \frac{\partial \alpha_t}{\partial t} + \mathrm{div}(J_t) = 0, \alpha_{t=0} = \alpha_0, \alpha_{t=1} = \alpha_1 \right\}$$

$$\forall\,(a,b) \in \mathbb{R}_+ \times \mathbb{R}^d, \quad \theta(a,b) = \begin{cases} \frac{\|b\|^2}{a} & \text{if} \quad a > 0, \\ 0 \quad \text{if} & (a,b) = 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Benamou-Brenier Formulation

Geodesic formulation:

$$\mathcal{W}_2^2(\alpha_0,\alpha_1) = \min \int_0^1 \int_{\mathbb{R}^d} \|v_t(x)\|^2 \,\mathrm{d}\alpha_t(x)\mathrm{d}t$$

$$\frac{\partial \alpha_t}{\partial t} + \mathrm{div}(\alpha_t v_t) = 0 \quad \text{and} \quad \alpha_{t=0} = \alpha_0, \alpha_{t=1} = \alpha_1$$

Momentum change of variable: $J_t \overset{\text{def.}}{=} \alpha_t v_t$

$$\mathcal{W}_2^2(\alpha_0,\alpha_1) = \min_{(\alpha_t,J_t)_t \in \mathcal{C}(\alpha_0,\alpha_1)} \int_0^1 \int_{\mathbb{R}^d} \theta(\alpha_t(x), J_t(x))\mathrm{d}x\mathrm{d}t$$

$$\mathcal{C}(\alpha_0,\alpha_1) \overset{\text{def.}}{=} \left\{(\alpha_t, J_t) \;:\; \frac{\partial \alpha_t}{\partial t} + \mathrm{div}(J_t) = 0, \alpha_{t=0} = \alpha_0, \alpha_{t=1} = \alpha_1\right\}$$

$$\forall\,(a,b) \in \mathbb{R}_+ \times \mathbb{R}^d, \quad \theta(a,b) = \begin{cases} \frac{\|b\|^2}{a} & \text{if} \quad a > 0, \\ 0 \quad \text{if} & (a,b) = 0, \\ +\infty & \text{otherwise.} \end{cases}$$

Non-smooth convex optimization.

Finite elements/differences discretization.

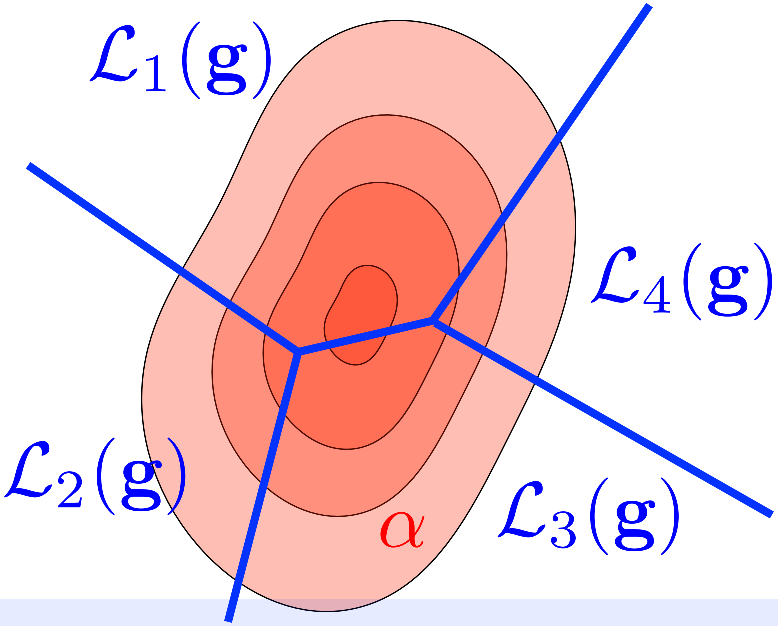
→ Quadratic cone interior point.

→ First order proximal methods (ADMM/DR).

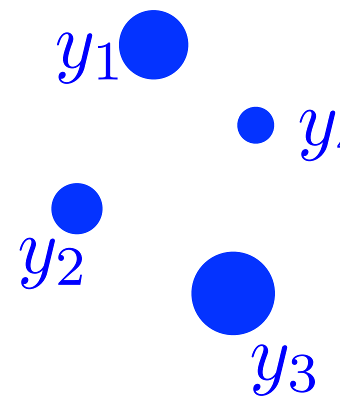
Overview

- Linear Programming
- PDE-based
- **Semi-discrete**
- Entropic Regularization

Semi-discrete Method



$\beta = \sum_{j=1}^m \mathbf{b}_j \delta_{y_j}$

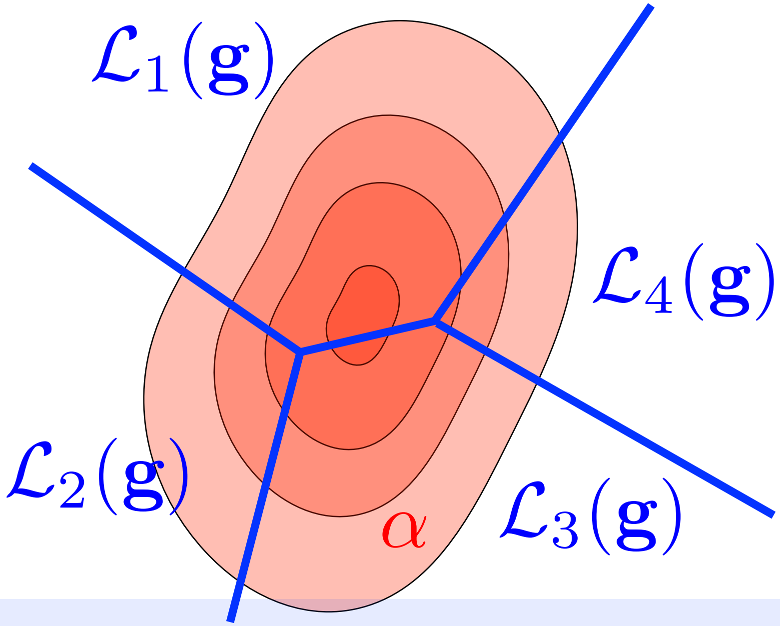


Proposition:
Optimal transport: $y_j \mapsto \mathcal{L}_j(\mathbf{g})$

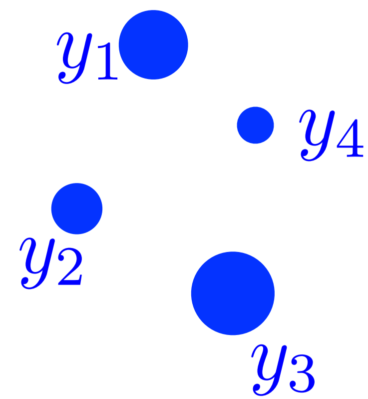
Laguerre cell: $\mathcal{L}_j(\mathbf{g}) \overset{\text{def.}}{=} \{x \; ; \; \forall \ell, \; \|x - y_j\|^2 - \mathbf{g}_j \leqslant \|x - y_\ell\|^2 - \mathbf{g}_\ell\}$

$\rightarrow$ computation in $O(m \log(m))$ in 2-D and 3-D.

Semi-discrete Method



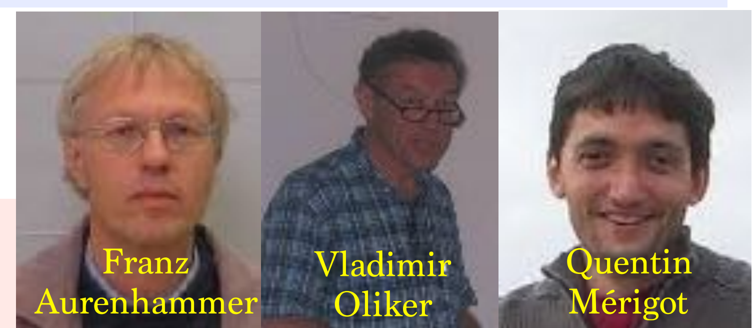
$\beta = \sum_{j=1}^m \mathbf{b}_j \delta_{y_j}$



Proposition:
Optimal transport:
$y_j \mapsto \mathcal{L}_j(\mathbf{g})$

Laguerre cell: $\mathcal{L}_j(\mathbf{g}) \overset{\text{def.}}{=} \{x \,;\, \forall \ell,\, \|x - y_j\|^2 - \mathbf{g}_j \leqslant \|x - y_\ell\|^2 - \mathbf{g}_\ell\}$

$\rightarrow$ computation in $O(m\log(m))$ in 2-D and 3-D.

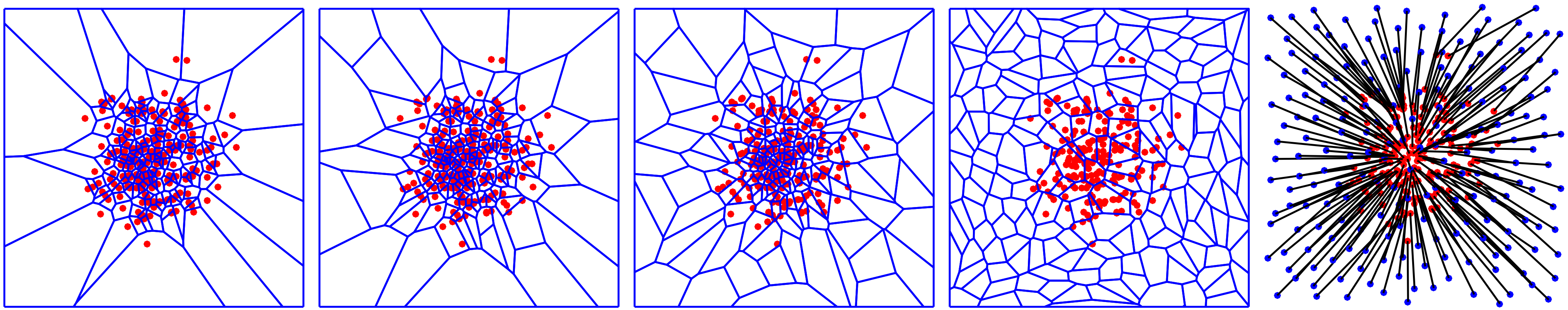


+ Laird Prussner

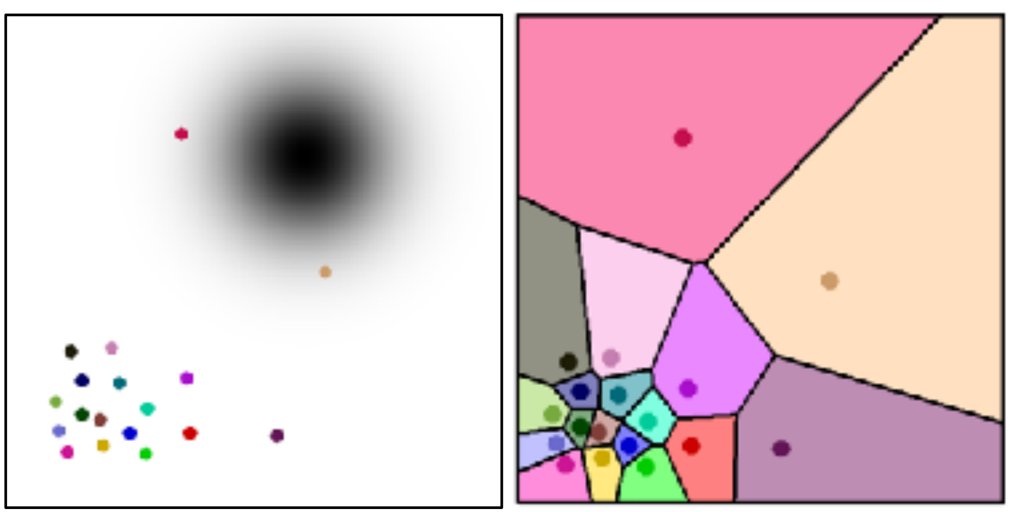
Mass conservation: $\forall\, j,\ \int_{\mathcal{L}_j(\mathbf{g})} \mathrm{d}\alpha = \mathbf{b}_j$

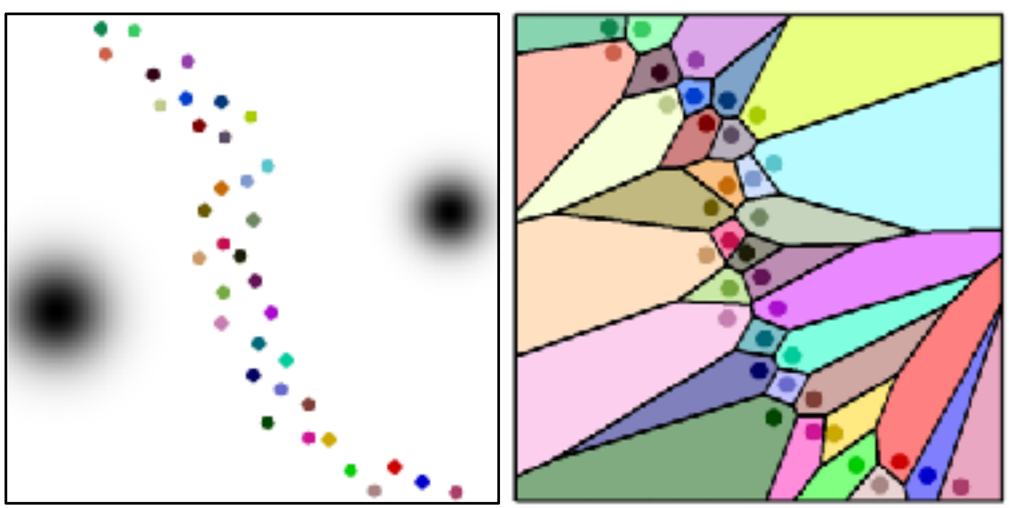
Gradient descent: $\mathbf{g} \leftarrow (1-\tau)\mathbf{g} + \tau \int_{\mathcal{L}_j(\mathbf{g})} \mathrm{d}\alpha$

Evolution of the Semi-Discrete Optimization

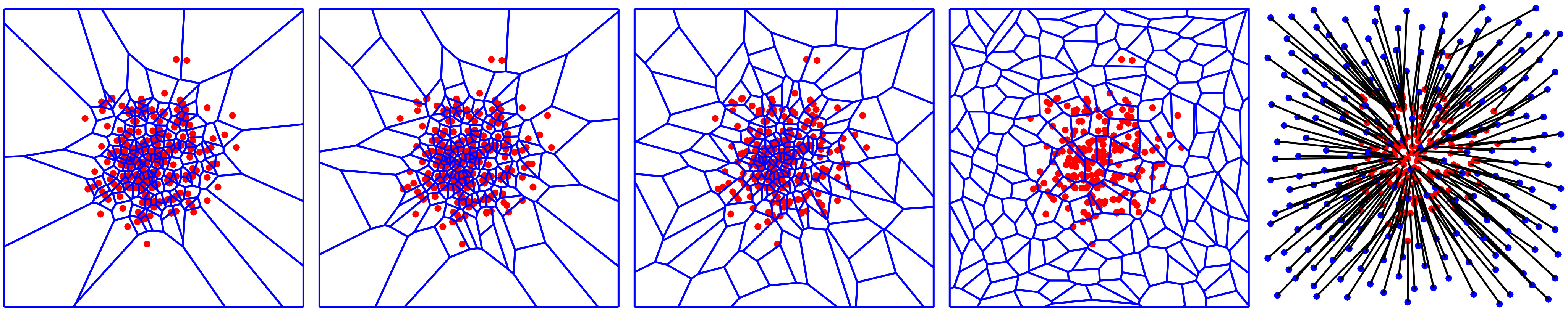


$\ell = 1$ $\quad$ $\ell = 3$ $\quad$ $\ell = 50$ $\quad$ $\ell = 100$ $\quad$ Matching

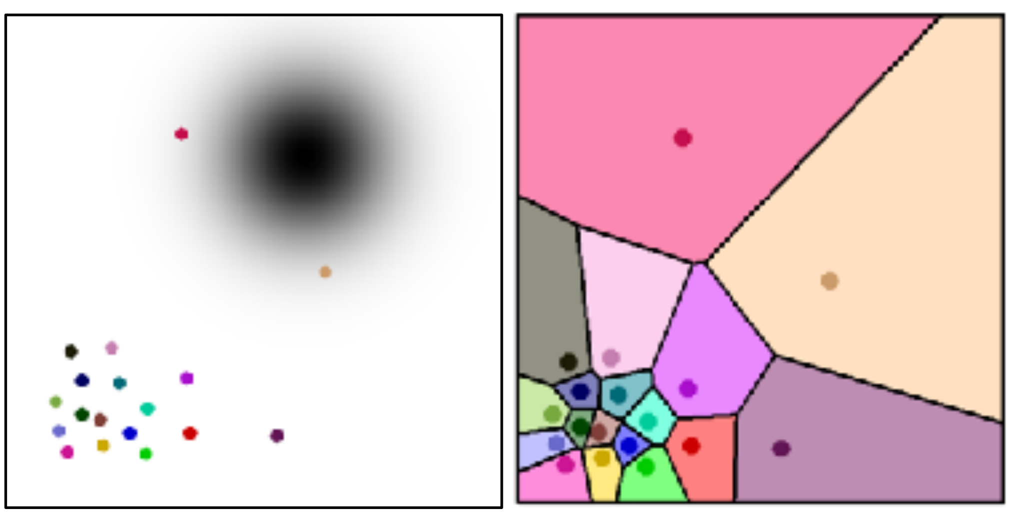


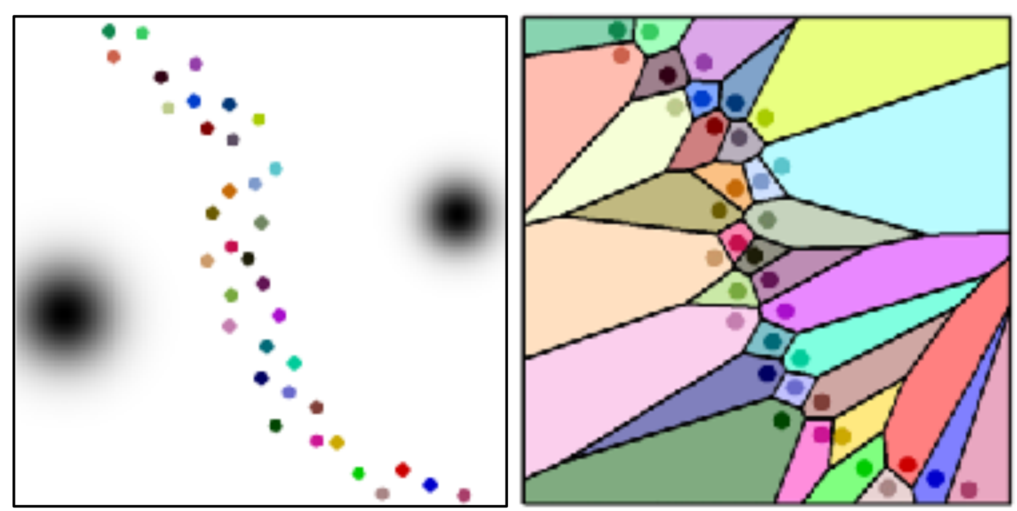


Evolution of the Semi-Discrete Optimization



$\ell = 1$ $\ell = 3$ $\ell = 50$ $\ell = 100$ Matching





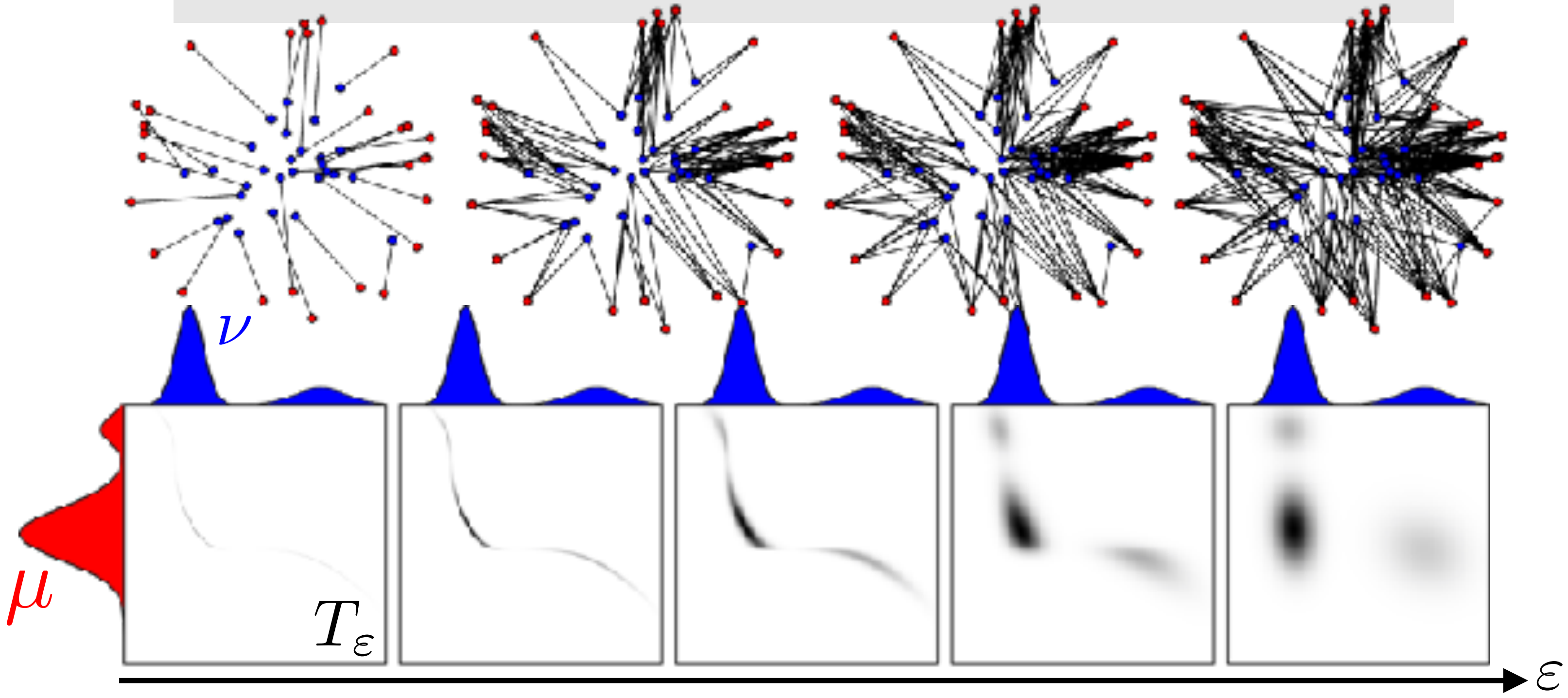
Overview

- Linear Programming
- PDE-based
- Semi-discrete
- **Entropic Regularization**

Entropic Regularization

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

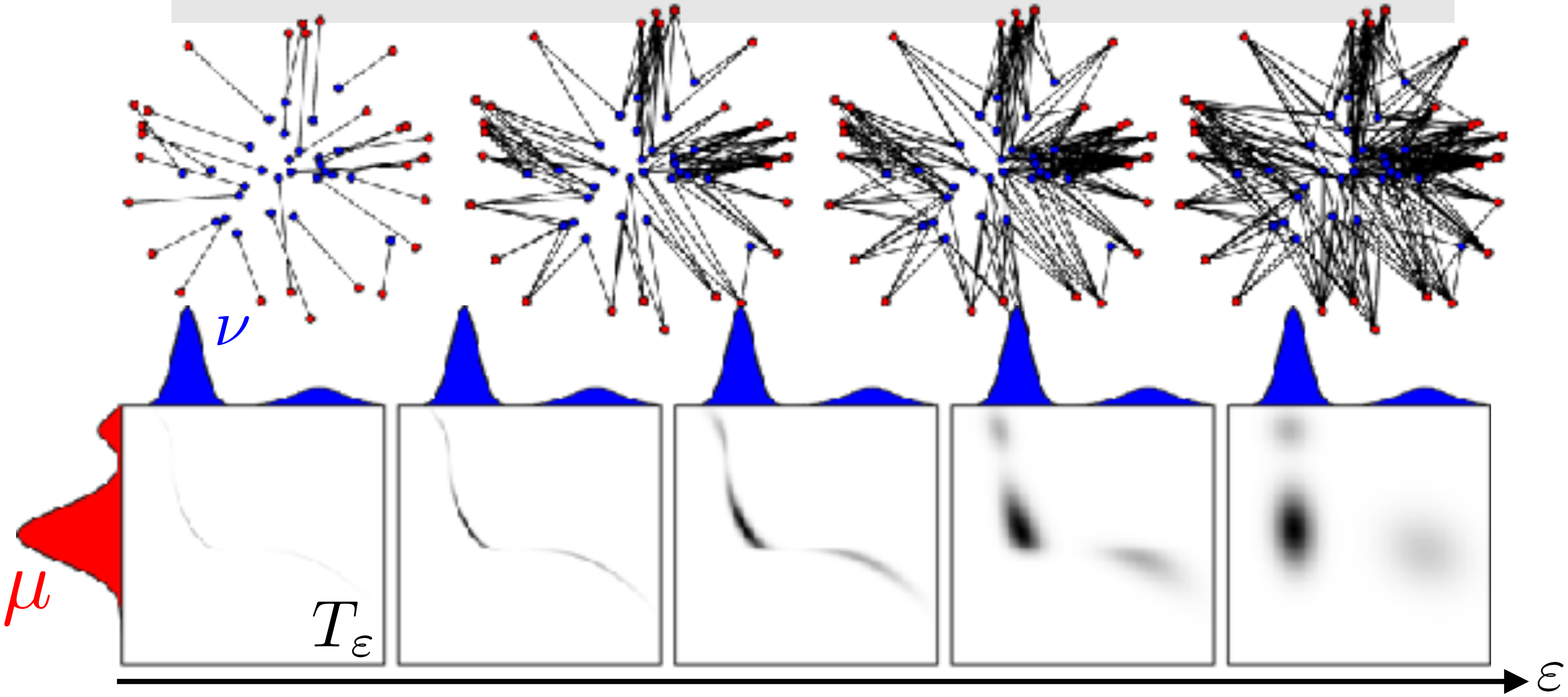
$$\mathbf{H}(\mathbf{P}) \overset{\text{def.}}{=} -\sum_{i,j} \mathbf{P}_{i,j}(\log(\mathbf{P}_{i,j}) - 1)$$



Entropic Regularization

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

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



Sinkhorn algorithm: τ-approximate solution in $O(n^2\tau^{-3})$.

Interior points: $O(n^{\frac{7}{2}}\log(\tau))$. Network simplex: $O(n^3)$ (exact).

$\rightarrow$ Regularization is crucial in high dimension.