L^2 Optimal Transport via Dual Convex Programming

Quentin Mérigot

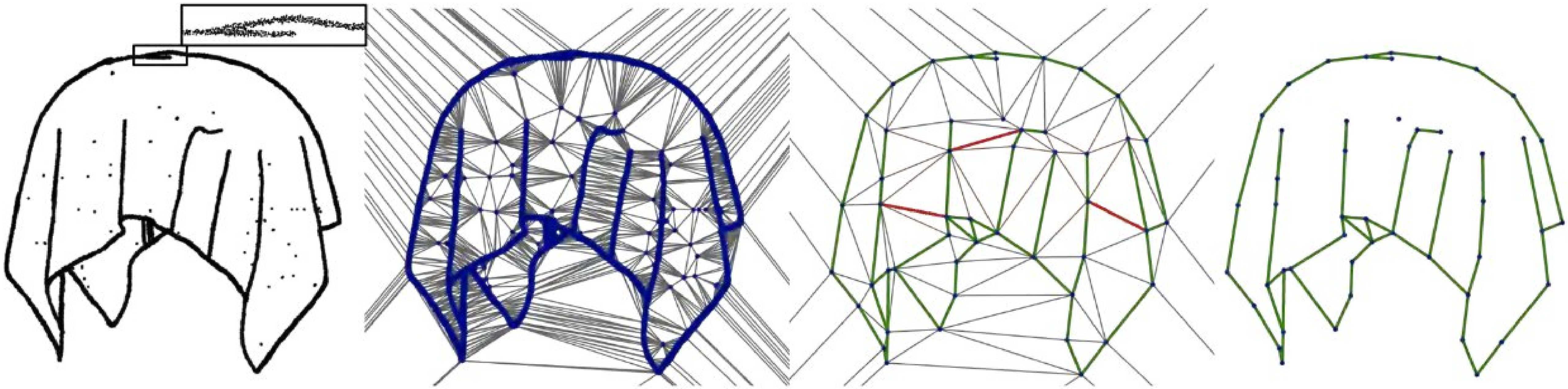
Laboratoire Jean Kuntzmann, Université de Grenoble

Journées de traitement d'image
Marseille
24 décembre 2011

Motivations

- Reconstructions with sharp corners and boundaries.

De Goes-Cohen-Steiner-Alliez-Desbrun (SGP '11)

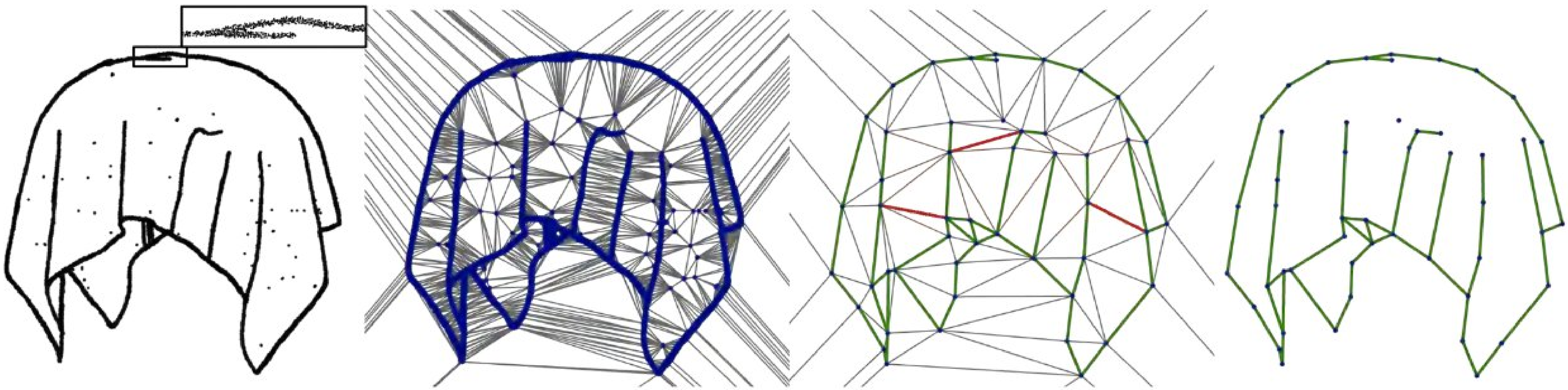


Given a point sample, reconstruct the underlying object as a sub-graph of a triangulation minimizing an energy related to OT.

Motivations

- Reconstructions with sharp corners and boundaries.

De Goes-Cohen-Steiner-Alliez-Desbrun (SGP '11)

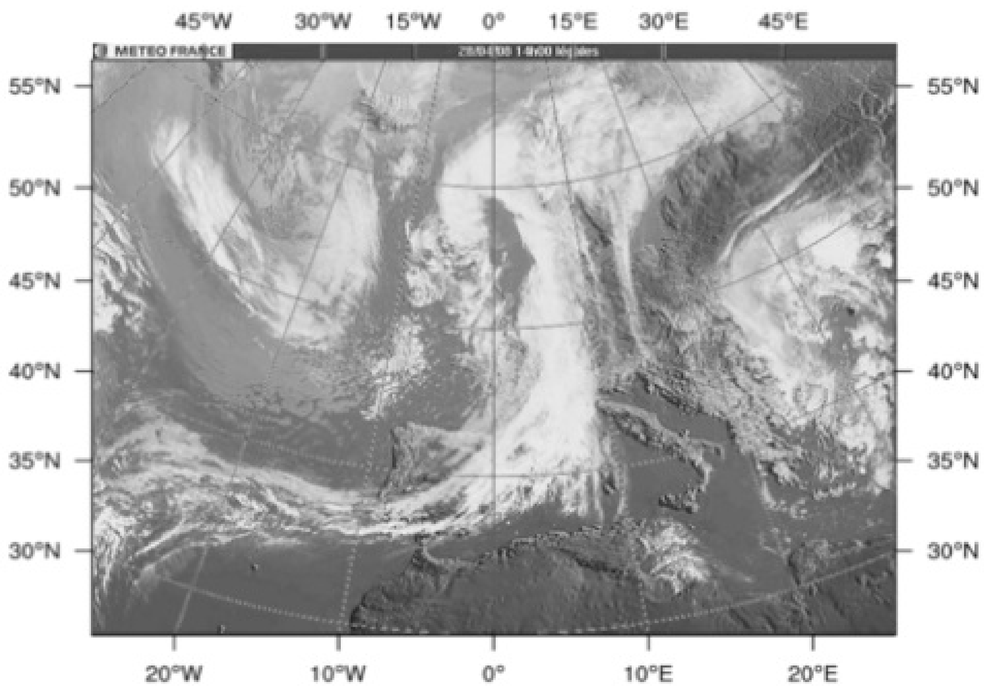


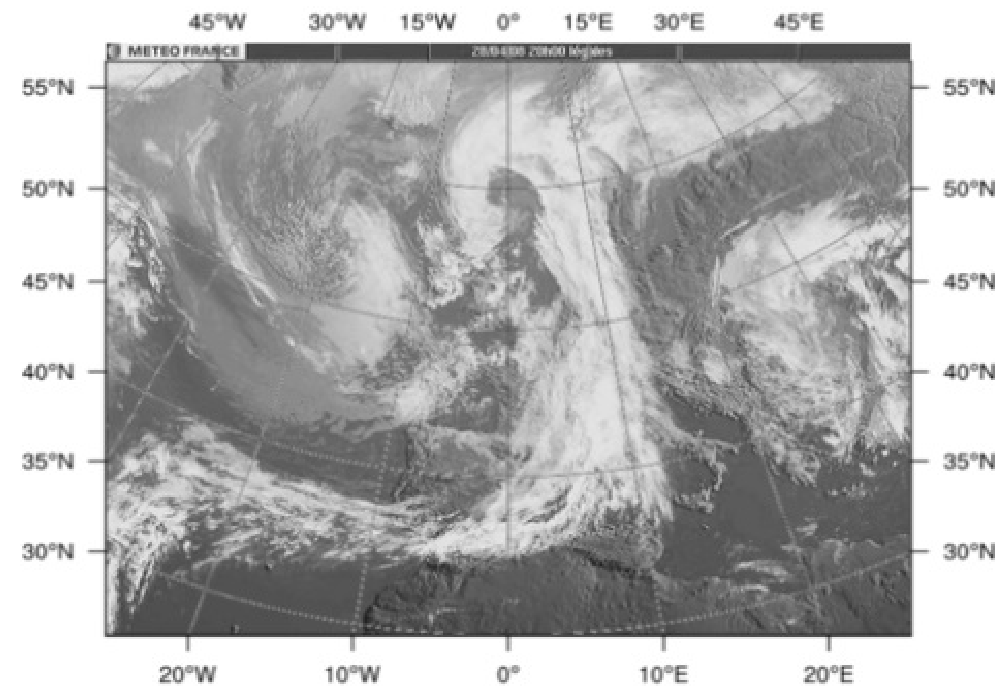
Given a point sample, reconstruct the underlying object as a sub-graph of a triangulation minimizing an energy related to OT.

Main ingredient for a 3D version: compute the L^2 optimal transport between the uniform measures ν on N points in the plane and μ on a triangle.

Motivations

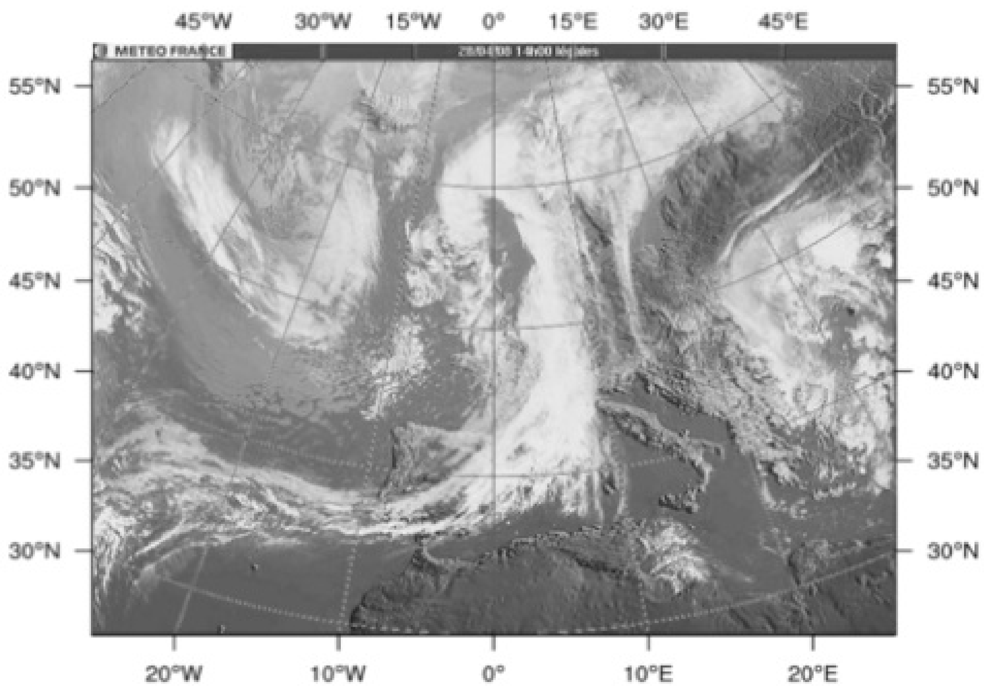
- Distance between grayscale images representing a density.

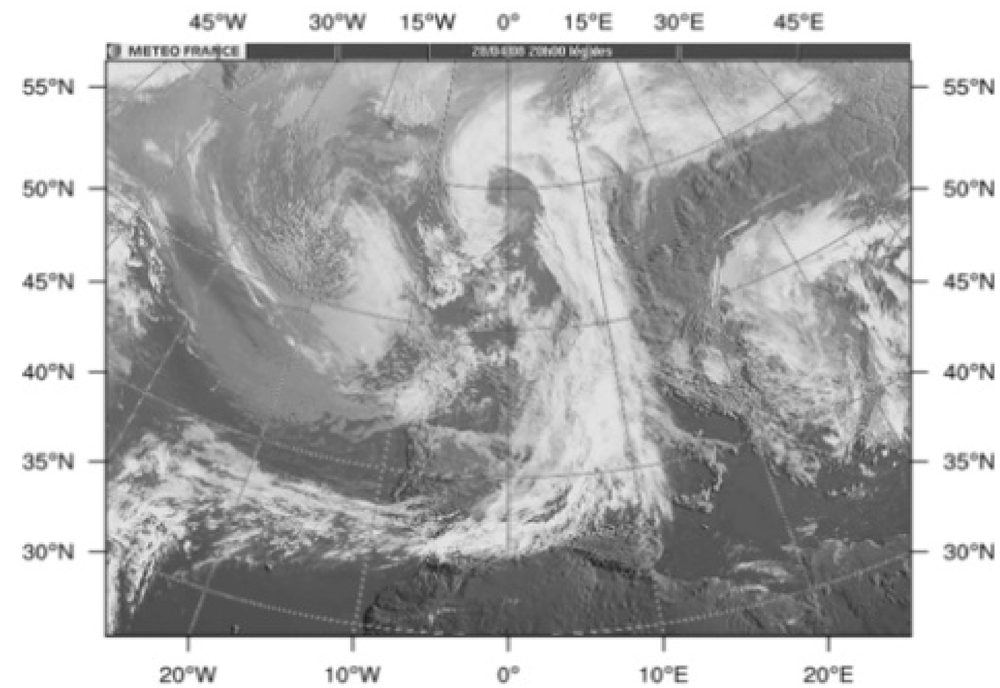




Motivations

- Distance between grayscale images representing a density.



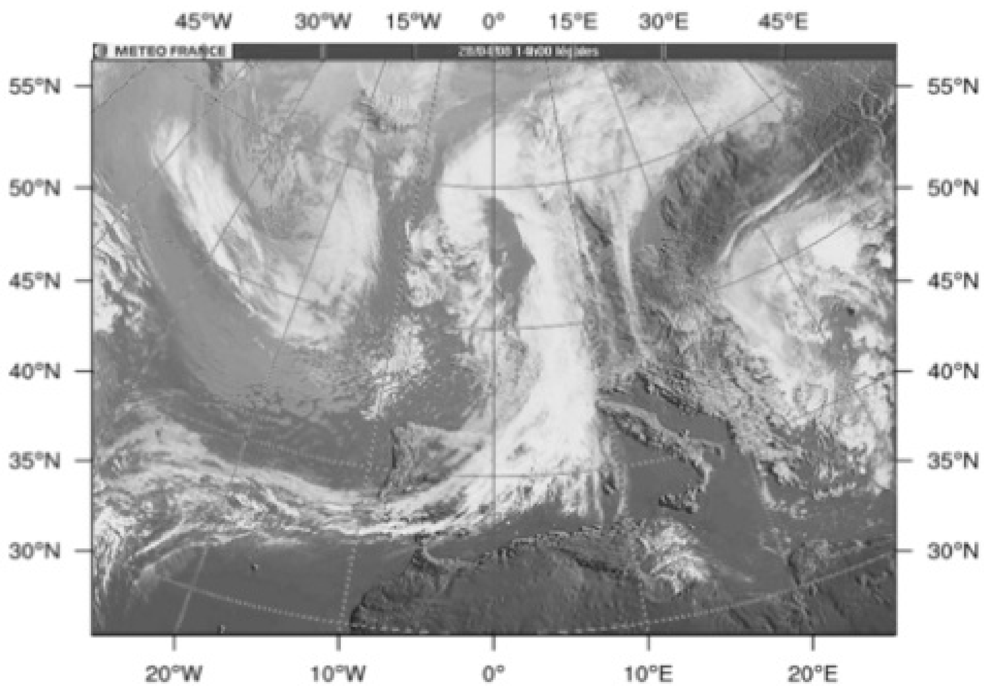


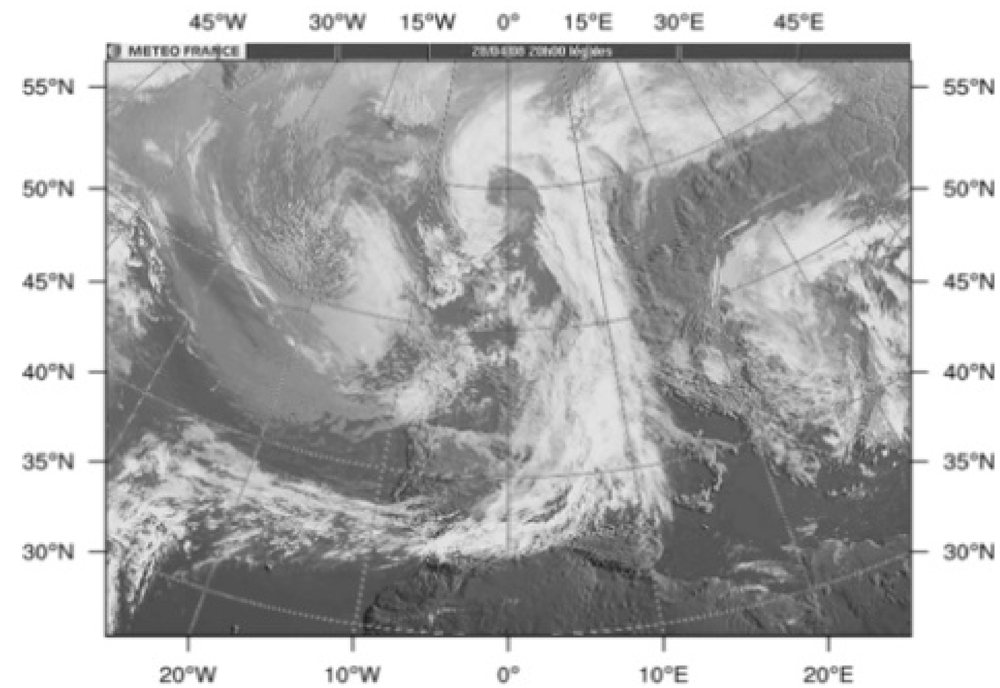
Meaningful distances between such images can be of the form:

$$E(\rho,\sigma) = \min_T \int \|x - T(x)\|^2 \rho(x)\,\mathrm{d}\,x + \mathrm{E}_{\mathrm{regu}}(T) \text{ where } T_{\#}\rho = \sigma.$$

Motivations

- Distance between grayscale images representing a density.





Meaningful distances between such images can be of the form:

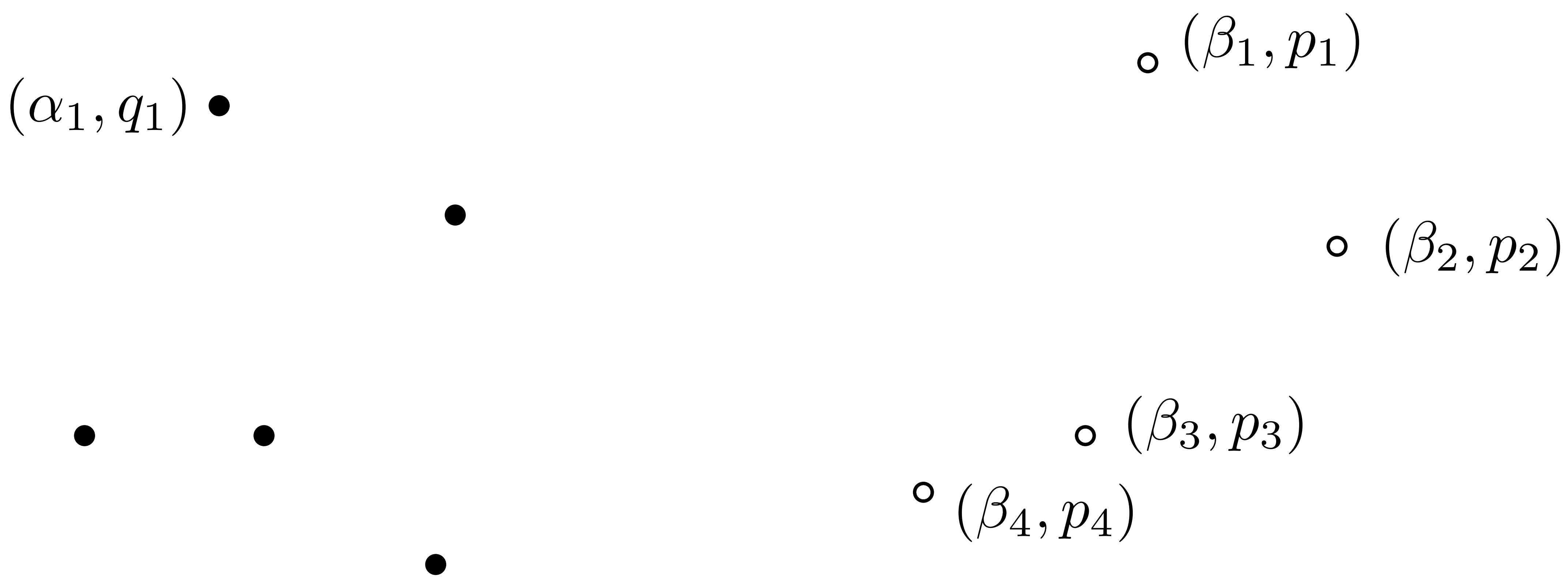
$$E(\rho, \sigma) = \min_T \int \|x - T(x)\|^2 \rho(x) \, \mathrm{d}\, x + \mathrm{E}_{\mathrm{regu}}(T) \text{ where } T_{\#}\rho = \sigma.$$

ANR project TOMMI (LJK / MAP5)

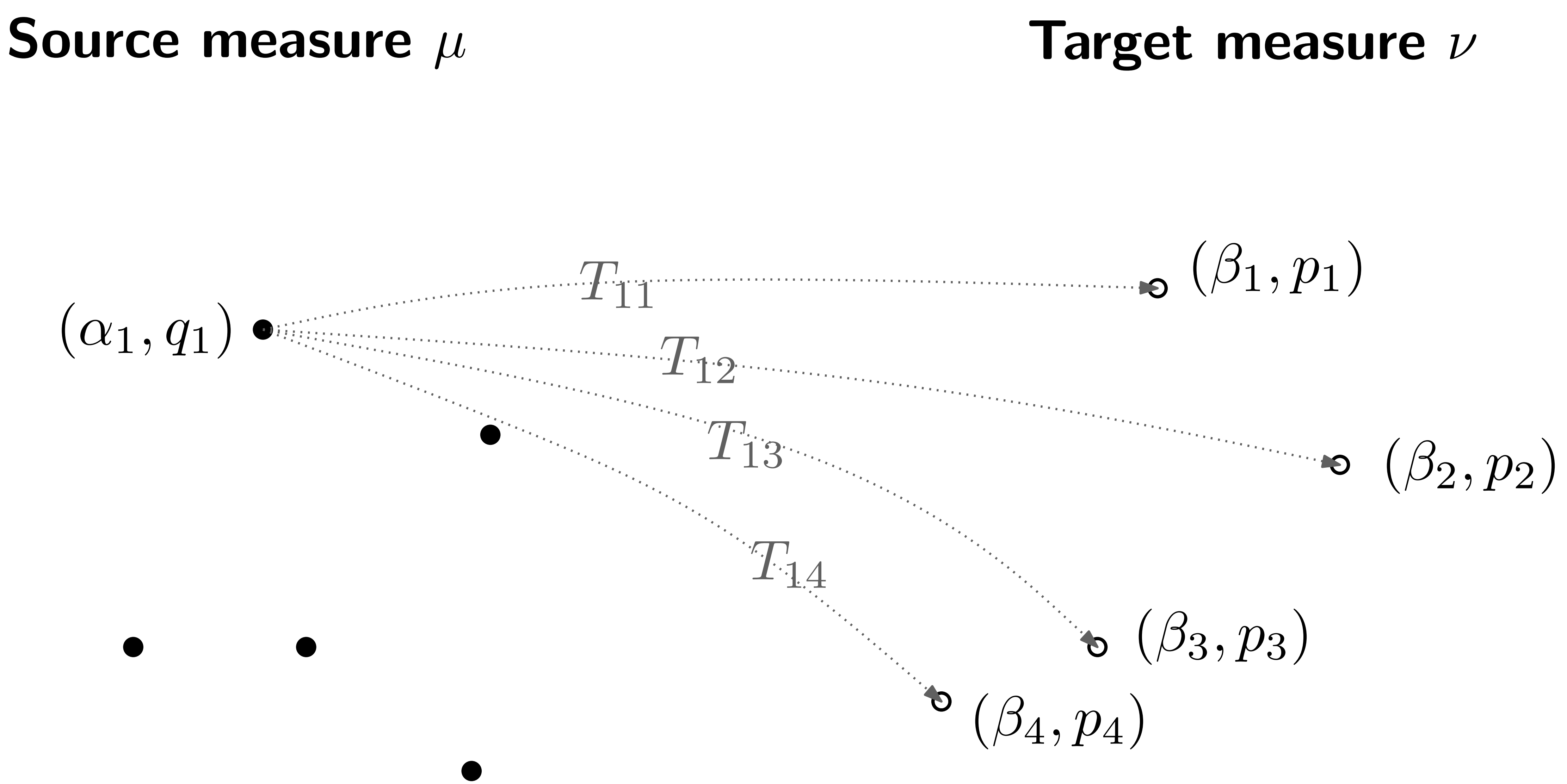
0. L^2 Optimal Transport

L^2 Optimal Transport

Source measure μ **Target measure ν**



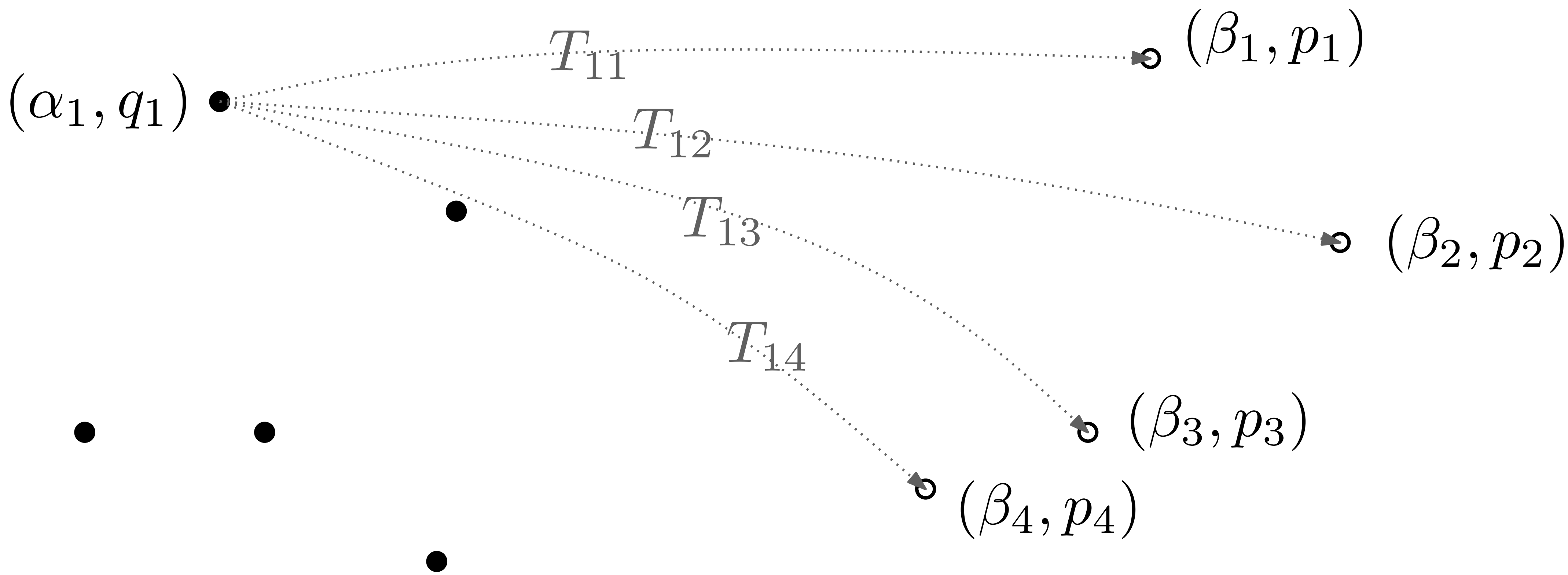
L^2 Optimal Transport



Transport plan: a matrix (T_{ij}) satisfiying $\sum_i T_{ij} = \alpha_i$ and $\sum_j T_{ij} = \beta_i$.

L^2 Optimal Transport

Source measure μ **Target measure** ν



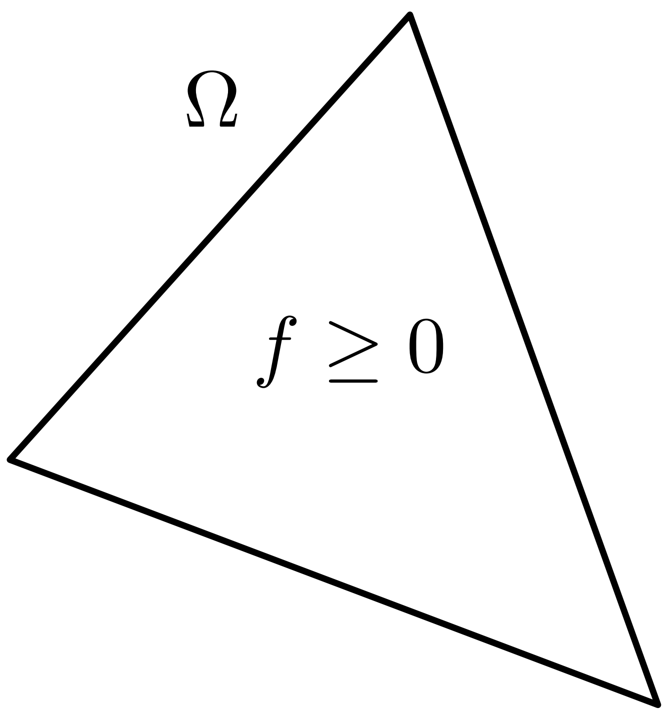
Transport plan: a matrix (T_{ij}) satisfiying $\sum_i T_{ij} = \alpha_i$ and $\sum_j T_{ij} = \beta_i$.

Cost: $c(T) = \sum_{i,j} T_{ij} \|q_i - p_j\|^2$.

Wasserstein: $\mathrm{W}_2(\mu, \nu) := (\min_T c(T))^{1/2}$

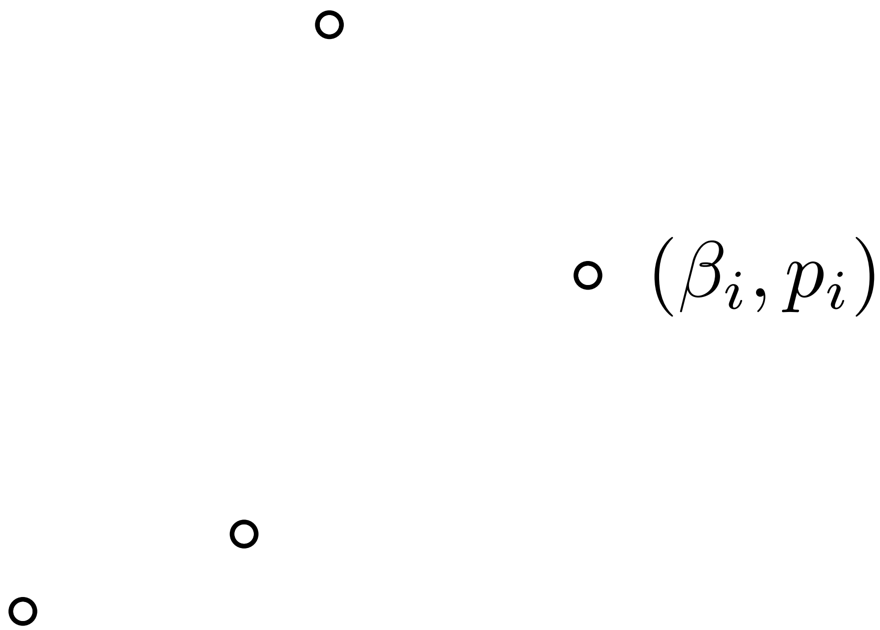
L^2 Optimal Transport

Source measure μ

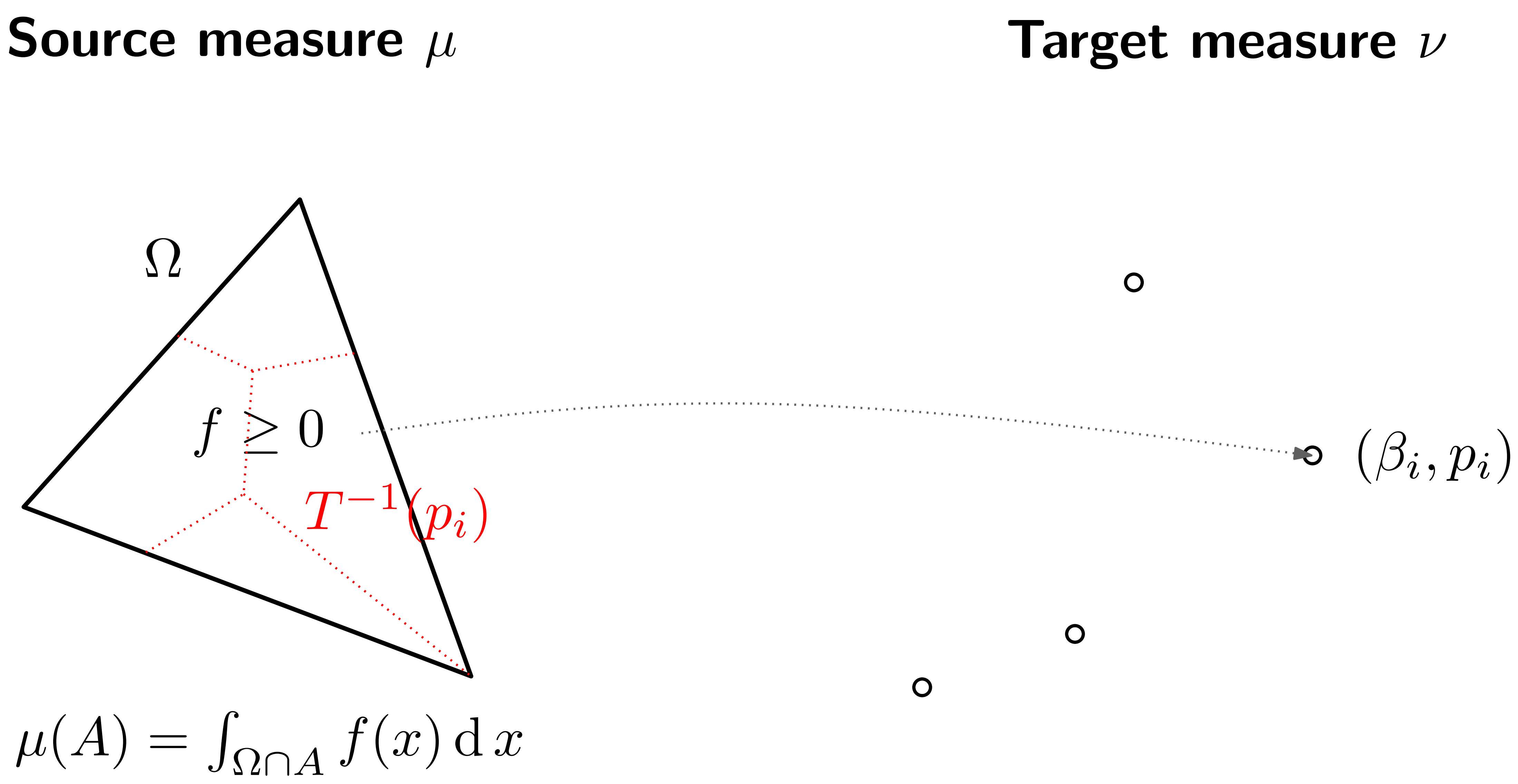


$\mu(A) = \int_{\Omega \cap A} f(x) \, \mathrm{d}\, x$

Target measure ν



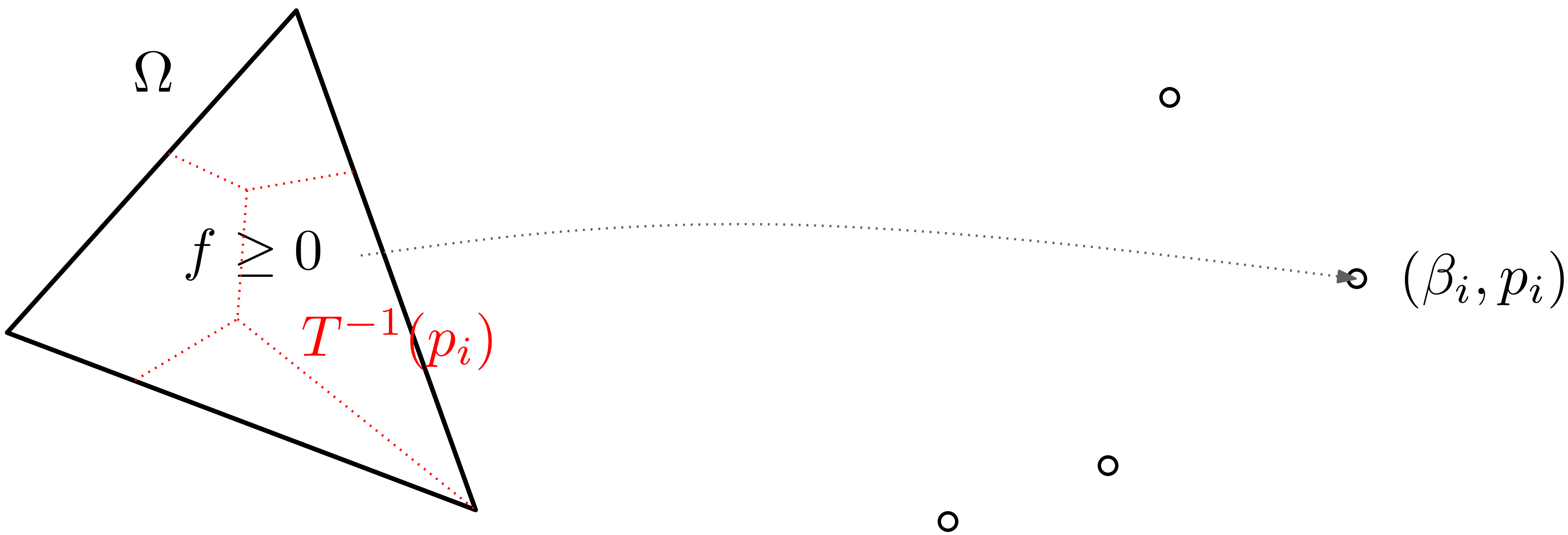
L^2 Optimal Transport



Transport plan: a map $T : \Omega \to \{p_i\}$ such that $\mu(T^{-1}(p_i)) = \beta_i$.

L^2 Optimal Transport

Source measure μ **Target measure** ν



$\mu(A) = \int_{\Omega \cap A} f(x)\,\mathrm{d}\,x$

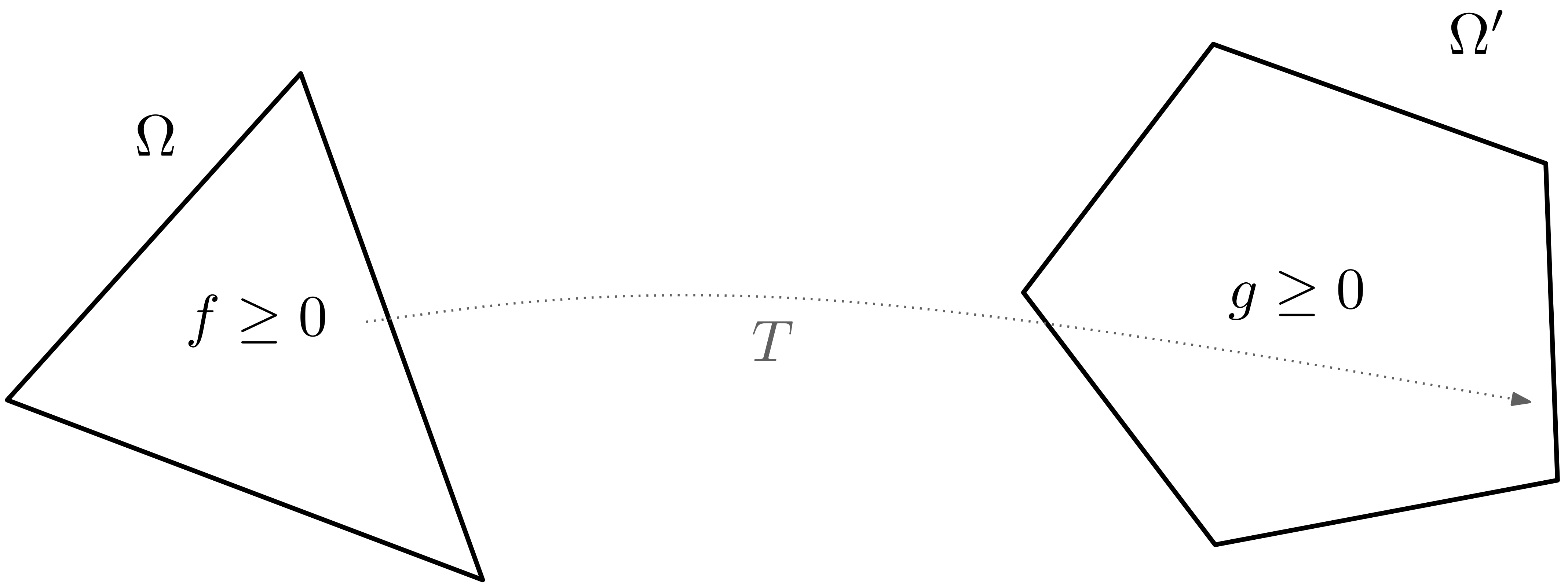
Transport plan: a map $T : \Omega \to \{p_i\}$ such that $\mu(T^{-1}(p_i)) = \beta_i$.

Cost: $c(T) = \int_\Omega \|x - T(x)\|^2 \,\mathrm{d}\,x$.

Wasserstein: $\mathrm{W}_2(\mu, \nu) := (\min_T c(T))^{1/2}$

L^2 Optimal Transport

Source measure μ **Target measure** ν

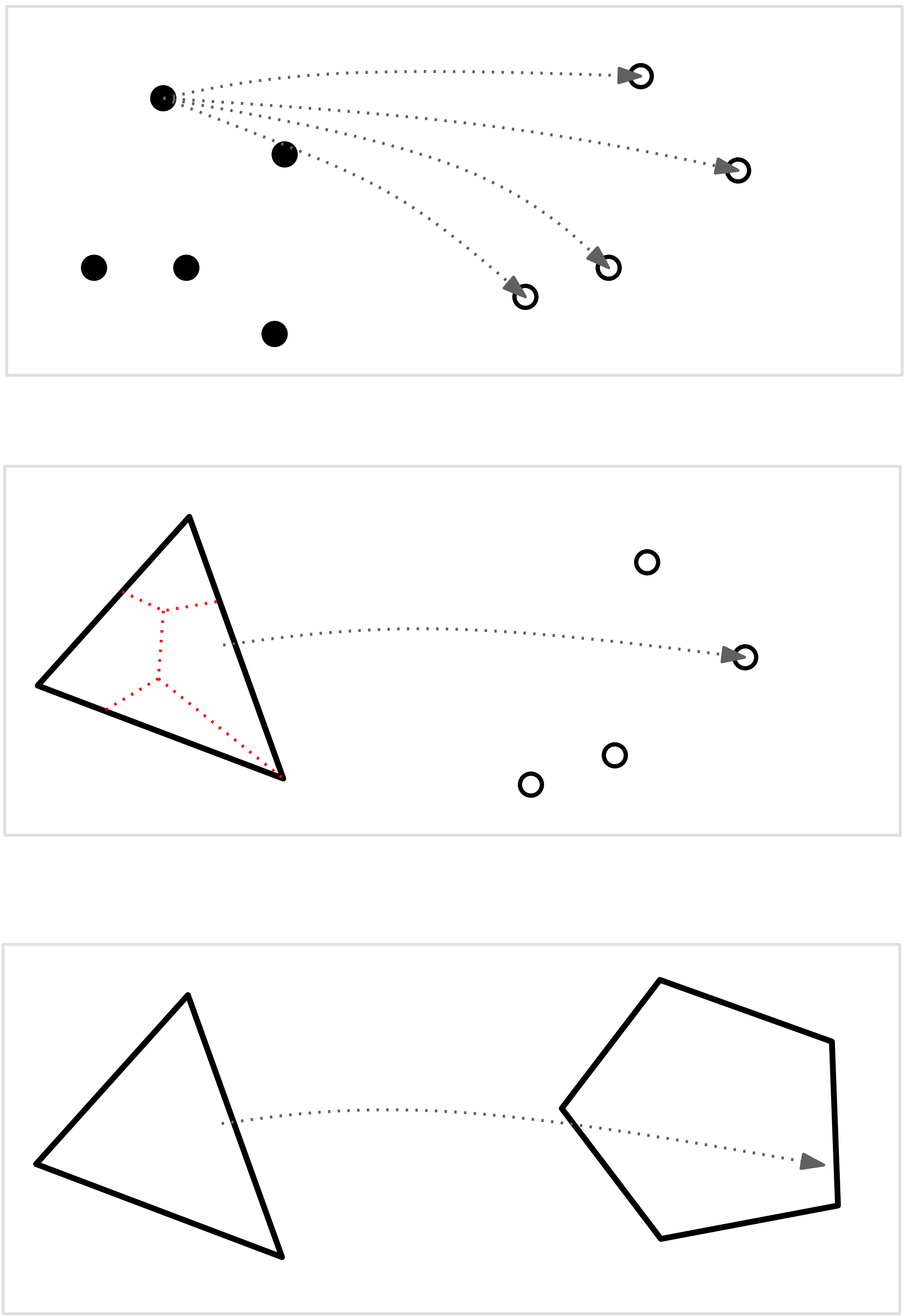


Transport plan: a map $T : \Omega \to \Omega'$ such that $\det(\mathrm{d}\, T(x)) = g(T(x))/f(x)$.

Cost: $c(T) = \int_\Omega \|x - T(x)\|^2 \,\mathrm{d}\, x$.

Wasserstein: $\mathrm{W}_2(\mu, \nu) := (\min_T c(T))^{1/2}$

L^2 Optimal Transport

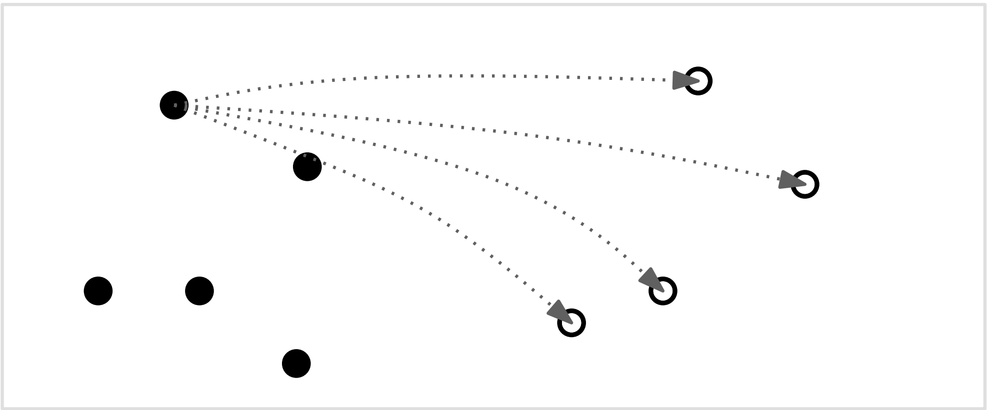


General α_i, β_j: linear programming

For $\alpha_i, \beta_j = 1$ and $p_i, q_j \in \mathbb{Z}^d$:

Hungarian algorithm, Bertsekas 'auction' algorithm

L^2 Optimal Transport

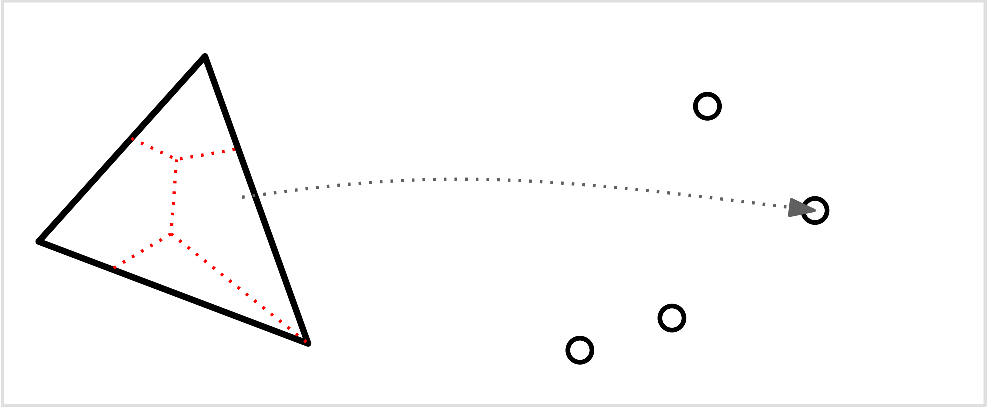


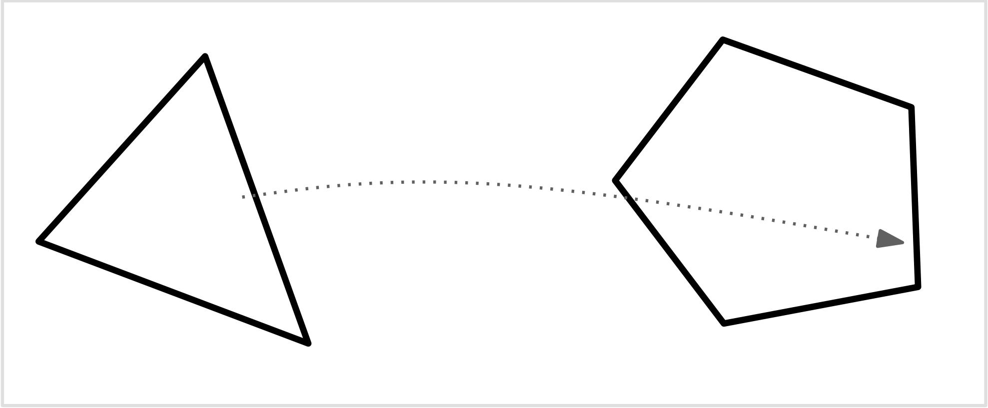
General α_i, β_j:

linear programming

For $\alpha_i, \beta_j = 1$ and $p_i, q_j \in \mathbb{Z}^d$:

Hungarian algorithm, Bertsekas 'auction' algorithm





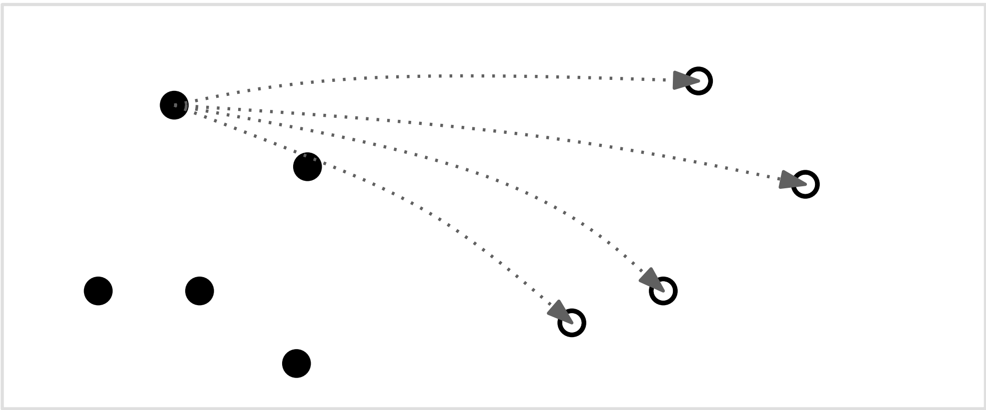
Smooth f, g with positive lower bound:

Benamou-Brenier '00

Loeper '05

Angenent-Haker-Tannenbaum '03

L^2 Optimal Transport

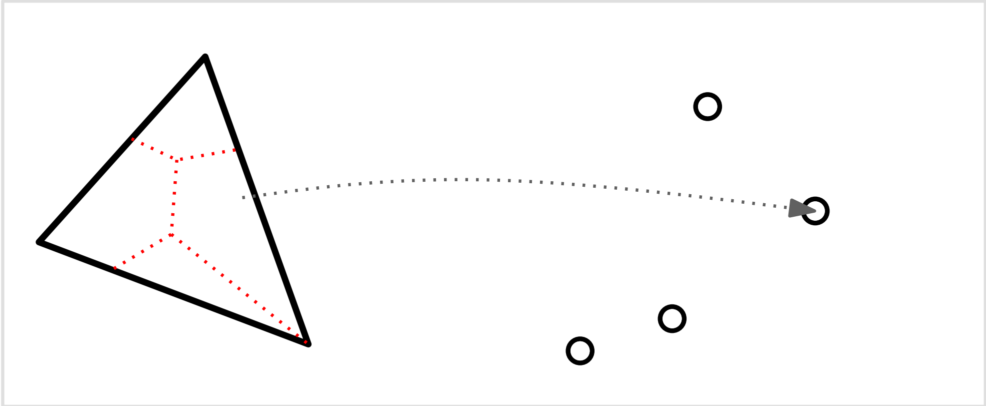


General α_i, β_j:

linear programming

For $\alpha_i, \beta_j = 1$ and $p_i, q_j \in \mathbb{Z}^d$:

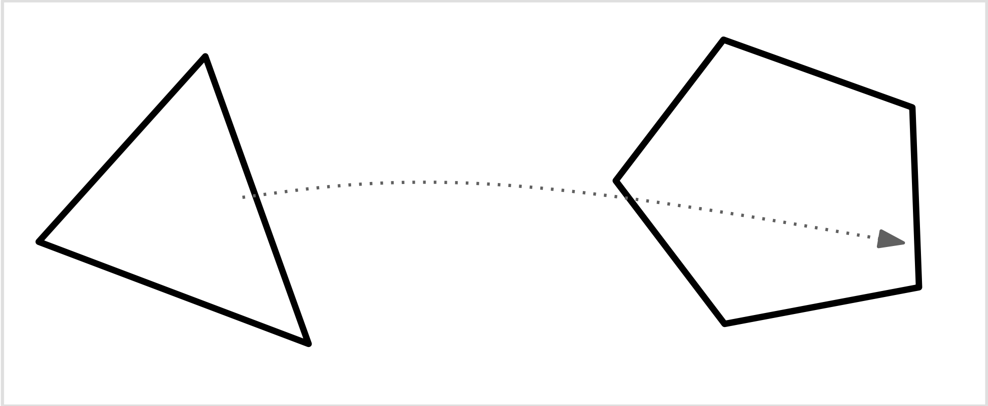
Hungarian algorithm, Bertsekas 'auction' algorithm



Source with density, discrete target:

Aurenhammer, Hoffmann, Aronov '98

McCann, Gangbo 98



Smooth f, g with positive lower bound:

Benamou-Brenier '00

Loeper '05

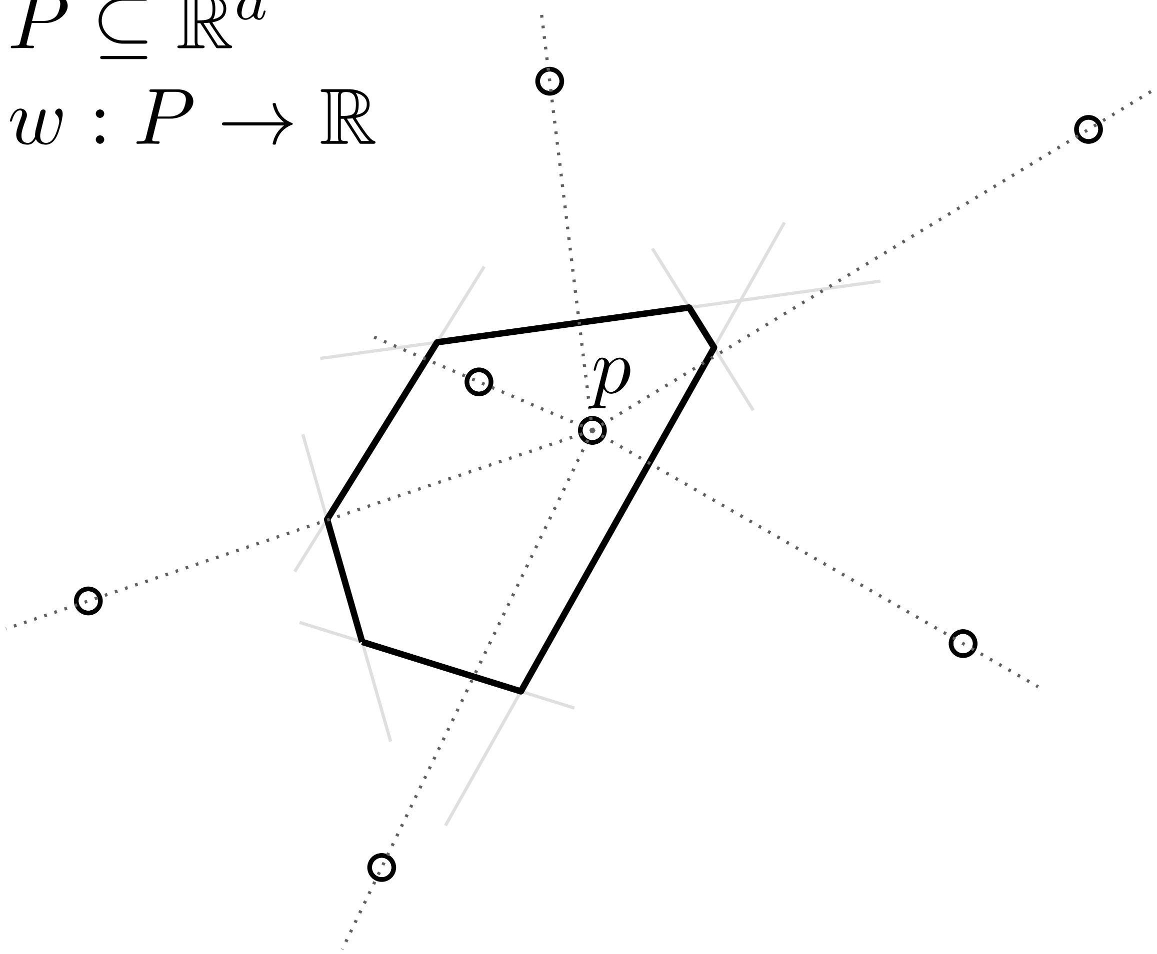
Angenent-Haker-Tannenbaum '03

1. Optimal Transport via Convex Programming

Power Diagrams and Optimal Transport

$P \subseteq \mathbb{R}^d$

$w : P \to \mathbb{R}$

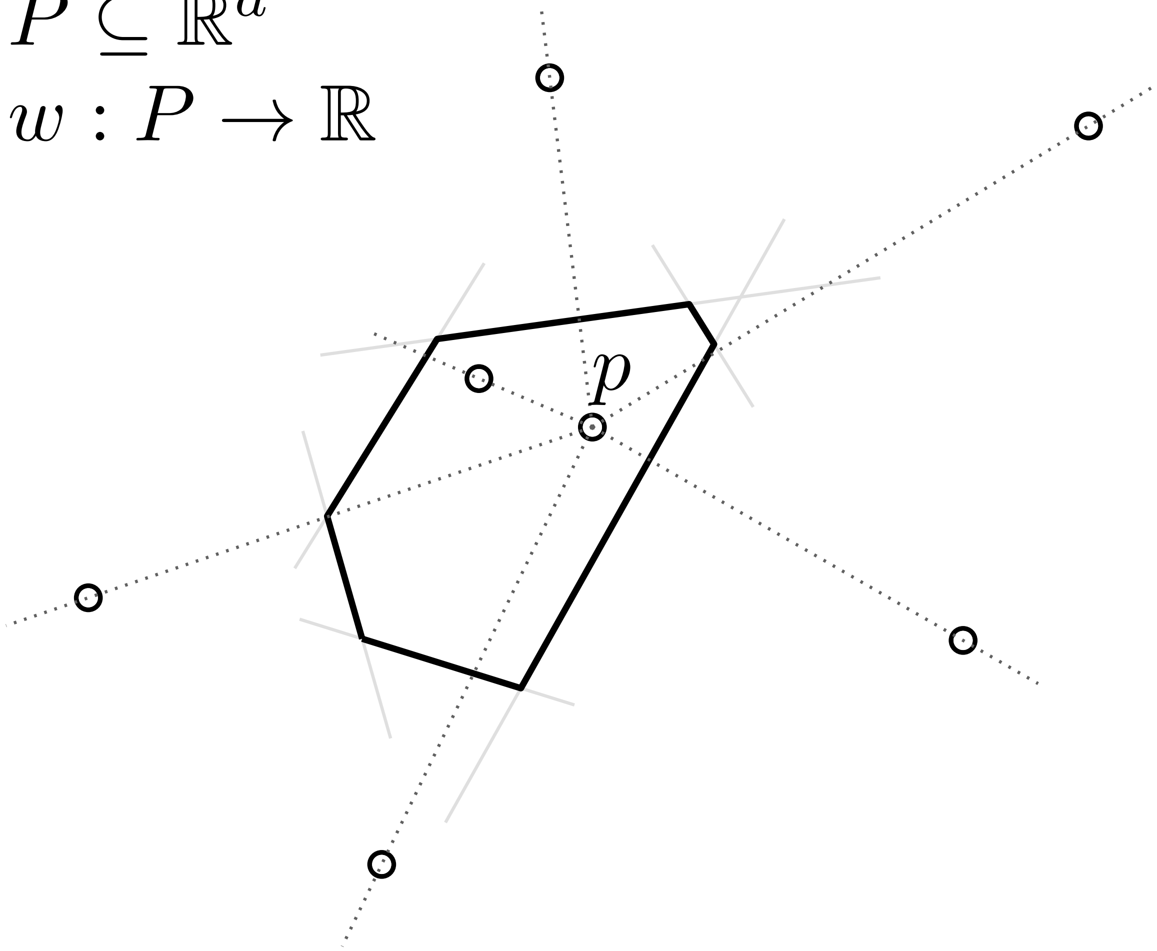


Transport map:

$$T_P^w(x) := \arg\min_{p \in P} \|x - p\|^2 + w(p)$$

Power Diagrams and Optimal Transport

$P \subseteq \mathbb{R}^d$
$w : P \to \mathbb{R}$



Transport map:

$$T_P^w(x) := \arg\min_{p \in P} \|x - p\|^2 + w(p)$$

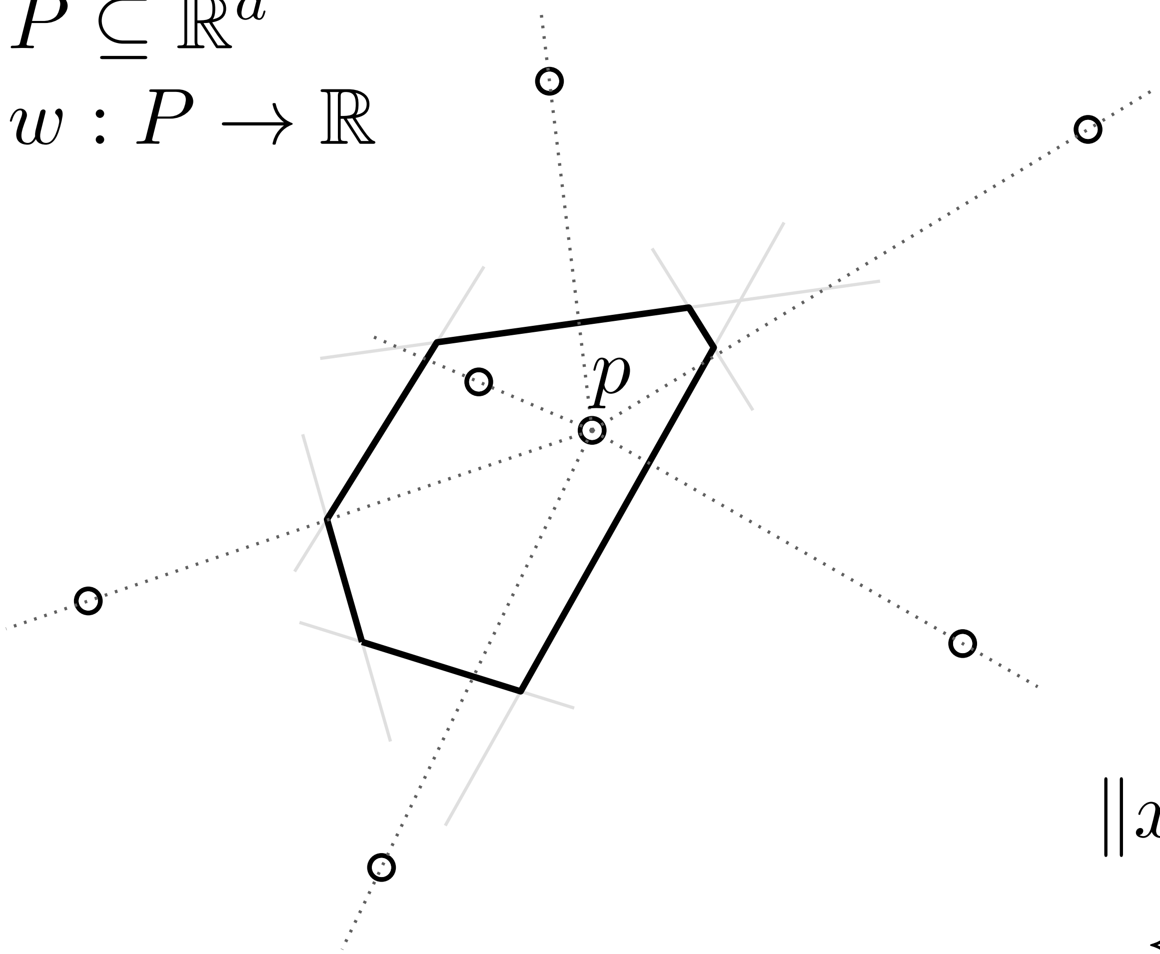
Power cell of p:

$$\mathrm{Vor}_P^w(p) := \{x \in \mathbb{R}^d ; T_P^w(x) = p\}$$

Power Diagrams and Optimal Transport

$P \subseteq \mathbb{R}^d$

$w : P \to \mathbb{R}$



Transport map:

$$T_P^w(x) := \arg\min_{p \in P} \|x - p\|^2 + w(p)$$

Power cell of p:

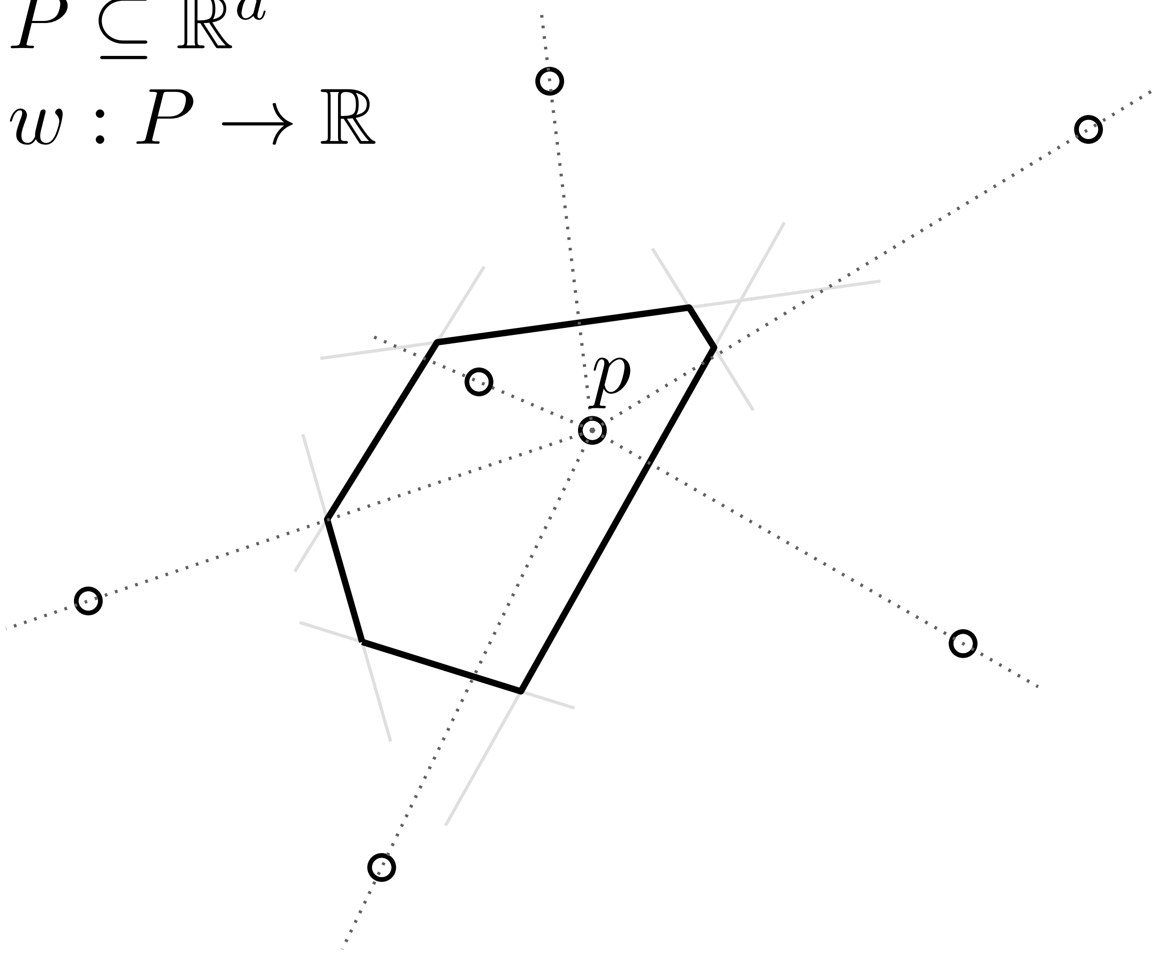
$$\mathrm{Vor}_P^w(p) := \{x \in \mathbb{R}^d ; T_P^w(x) = p\}$$

$$\|x - p\|^2 + w(p) \leq \|x - q\|^2 + w(q)$$

$$\iff 2\langle x | q - p \rangle \leq w(q) - w(p)$$

Power Diagrams and Optimal Transport

$P \subseteq \mathbb{R}^d$
$w : P \to \mathbb{R}$



Transport map:

$$T_P^w(x) := \arg\min_{p \in P} \|x - p\|^2 + w(p)$$

Power cell of p:

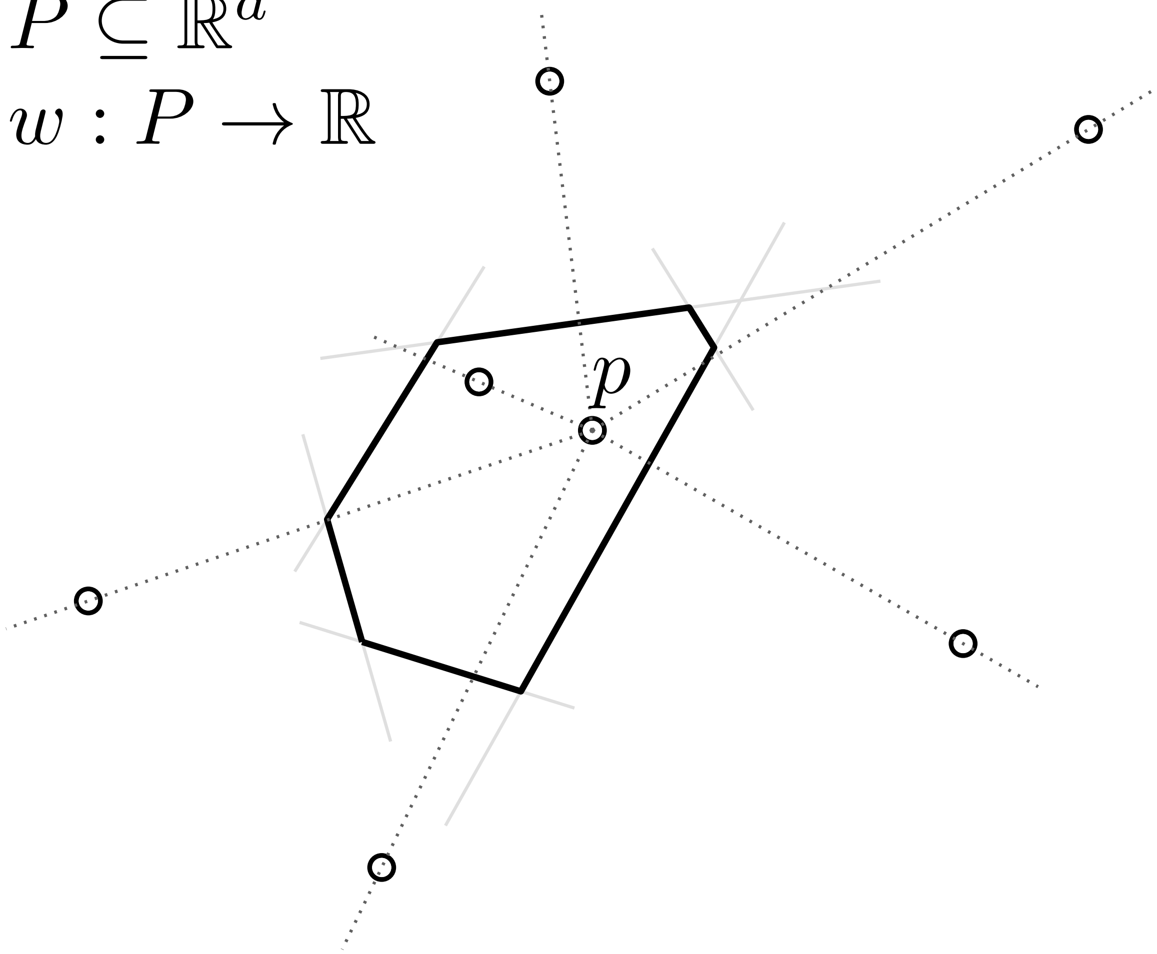
$$\mathrm{Vor}_P^w(p) := \{x \in \mathbb{R}^d;\, T_P^w(x) = p\}$$

Lemma: Given a measure μ with density and (P, w), the map T_P^w is an optimal transport between μ and

$$\nu := \sum_{p \in P} \mu(\mathrm{Vor}_P^w(p))\delta_p \qquad \text{(i.e. } \nu = T_{P\#}^w \mu\text{)}$$

Power Diagrams and Optimal Transport

$P \subseteq \mathbb{R}^d$
$w : P \to \mathbb{R}$



Transport map:

$$T_P^w(x) := \arg\min_{p \in P} \|x - p\|^2 + w(p)$$

Power cell of p:

$$\mathrm{Vor}_P^w(p) := \{x \in \mathbb{R}^d ; T_P^w(x) = p\}$$

Theorem: Given a measure μ with density and a discrete measure $\nu = \sum_{p \in P} \alpha_p \delta_p$, there exists $w : P \to \mathbb{R}$ s.t.

$$\forall p \in P, \ \alpha_p = \mu(\mathrm{Vor}_P^w(p)) \qquad \text{(i.e. } \nu = T_{P\#}^w \mu\text{)}$$

Aurenhammer, Hoffman, Aronov '94 Brenier '92

Optimal Transport via Convex Programming

Kantorovich Duality: Given two probability measure μ, ν

$$\text{Wass}_2(\mu, \nu) = \max_{\mathbb{R}^d} \int v(x) \,\mathrm{d}\, \mu(x) - \int w(p) \,\mathrm{d}\, \nu(p)$$

where v, w are such that $v(x) - w(p) \leq \|x - p\|^2$.

Optimal Transport via Convex Programming

Kantorovich Duality: Given two probability measure μ, ν

$$\mathrm{Wass}_2(\mu,\nu) = \max_{\mathbb{R}^d} \int \min_p (\|x-p\|^2 + w(p)) \,\mathrm{d}\,\mu(x) - \int w(p)\,\mathrm{d}\,\nu(p).$$

Optimal Transport via Convex Programming

Kantorovich Duality: Given two probability measure μ, ν

$$\mathrm{Wass}_2(\mu,\nu) = \max_{\mathbb{R}^d} \int \min_p (\|x - p\|^2 + w(p)) \,\mathrm{d}\,\mu(x) - \int w(p) \,\mathrm{d}\,\nu(p).$$

Discrete case: μ with density and $\nu = \sum_{p\in P} \alpha_p \delta_p$,

$$\Phi(w) := -\sum_{p\in P} \int_{\mathrm{Vor}_P^w(p)} [\|x - p\|^2 + w(p)] \,\mathrm{d}\,\mu(x) + \sum_{p\in P} \alpha_p w(p)$$

$$\boxed{\mathrm{Wass}_2(\mu,\nu) = \min_w \Phi(w)}$$

Optimal Transport via Convex Programming

Kantorovich Duality: Given two probability measure μ, ν

$$\mathrm{Wass}_2(\mu,\nu) = \max_{\mathbb{R}^d} \int \min_p (\|x-p\|^2 + w(p)) \,\mathrm{d}\,\mu(x) - \int w(p)\,\mathrm{d}\,\nu(p).$$

Discrete case: μ with density and $\nu = \sum_{p\in P} \alpha_p \delta_p$,

$$\Phi(w) := -\sum_{p\in P} \int_{\mathrm{Vor}_P^w(p)} [\|x-p\|^2 + w(p)]\,\mathrm{d}\,\mu(x) + \sum_{p\in P} \alpha_p w(p)$$

$$\boxed{\mathrm{Wass}_2(\mu,\nu) = \min_w \Phi(w)}$$

Gradient: $\Phi(w+\varepsilon h) - \Phi(w) = \sum_{p\in P} h(p)\,(\mu(\mathrm{Vor}_P^w(p)) - \alpha_p)\,\varepsilon + \mathrm{O}(\varepsilon^2)$

|| changes in Power cells

I.e. $\nabla\Phi(w) = (\alpha_p - \mu(\mathrm{Vor}_P^w(p)))_{p\in P}$

Optimal Transport via Convex Programming

Kantorovich Duality: Given two probability measure μ, ν

$$\mathrm{Wass}_2(\mu,\nu) = \max_{\mathbb{R}^d} \int \min_p(\|x-p\|^2 + w(p))\,\mathrm{d}\,\mu(x) - \int w(p)\,\mathrm{d}\,\nu(p).$$

Discrete case: μ with density and $\nu = \sum_{p\in P} \alpha_p \delta_p$,

$$\Phi(w) := -\sum_{p\in P} \int_{\mathrm{Vor}_P^w(p)} [\|x-p\|^2 + w(p)]\,\mathrm{d}\,\mu(x) + \sum_{p\in P} \alpha_p w(p)$$

$$\boxed{\mathrm{Wass}_2(\mu,\nu) = \min_w \Phi(w)}$$

Gradient: $\Phi(w+\varepsilon h) - \Phi(w) = \sum_{p\in P} h(p)\,(\mu(\mathrm{Vor}_P^w(p)) - \alpha_p)\,\varepsilon + \mathrm{O}(\varepsilon^2)$

|| changes in Power cells

I.e. $\nabla\Phi(w) = (\alpha_p - \mu(\mathrm{Vor}_P^w(p)))_{p\in P}$

$\nabla\Phi(w)$ is actually a subgradient, i.e. the function Φ is convex

Implementation of Convex Programming

1. Computation of Φ and $\nabla\Phi$:

$$\int_{\Omega\cap\mathrm{Vor}_P^w(p)} f(x)\,\mathrm{d}\,x$$

$$\int_{\Omega\cap\mathrm{Vor}_P^w(p)} \|x-p\|^2 f(x)\,\mathrm{d}\,x$$

Implementation of Convex Programming

1. Computation of Φ and $\nabla\Phi$:

$$\int_{\Omega\cap \mathrm{Vor}_P^w(p)} f(x)\,\mathrm{d}\,x$$

$$\int_{\Omega\cap \mathrm{Vor}_P^w(p)} \|x-p\|^2 f(x)\,\mathrm{d}\,x$$

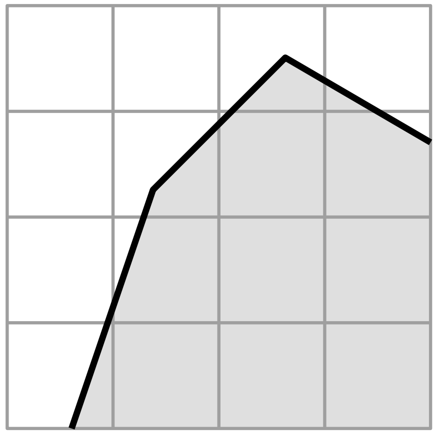
$f = 1$: Power diagram, Fast intersection of polygons

CGAL

O'Rourke, Chien, Olson, Naddor '82

$f =$ grayscale image: Piecewise constant on pixels

Modification of Bresenham algorithm to compute exact pixel coverage



Implementation of Convex Programming

1. **Computation of Φ and $\nabla\Phi$:**

$$\int_{\Omega \cap \mathrm{Vor}_P^w(p)} f(x)\,\mathrm{d}\,x$$

$$\int_{\Omega \cap \mathrm{Vor}_P^w(p)} \|x - p\|^2 f(x)\,\mathrm{d}\,x$$

2. **Iterative unconstrained convex programming:**

- Choice of an initial weight vector, e.g. $w_0(p) := 0$ for all p.

Implementation of Convex Programming

1. Computation of Φ and $\nabla\Phi$:

$$\int_{\Omega\cap\mathrm{Vor}_P^w(p)} f(x)\,\mathrm{d}\,x$$

$$\int_{\Omega\cap\mathrm{Vor}_P^w(p)} \|x-p\|^2 f(x)\,\mathrm{d}\,x$$

2. Iterative unconstrained convex programming:

- Choice of an initial weight vector, e.g. $w_0(p) := 0$ for all p.
- Computation of descent direction d_k

  steepest descent $-\nabla\Phi(w_k)$, Newton $-[\mathrm{D}^2\Phi(w_k)]^{-1}(\nabla\Phi(x_k))$, quasi-Newton...

  L-BFGS: low-storage version of the BFGS quasi-Newton scheme

Implementation of Convex Programming

1. Computation of Φ and $\nabla\Phi$:

$$\int_{\Omega \cap \mathrm{Vor}_P^w(p)} f(x)\,\mathrm{d}\,x$$

$$\int_{\Omega \cap \mathrm{Vor}_P^w(p)} \|x-p\|^2 f(x)\,\mathrm{d}\,x$$

2. Iterative unconstrained convex programming:

- Choice of an initial weight vector, e.g. $w_0(p) := 0$ for all p.
- Computation of descent direction d_k
- Computation of time step s_k

  optimal $s_k = \arg\min_s \Phi(w_k + s d_k)$, fixed $s_k = \mathrm{cst}$

  in practice: backtracking line-search (e.g. Wolfe condition)

Implementation of Convex Programming

1. Computation of Φ and $\nabla\Phi$:

$$\int_{\Omega\cap \mathrm{Vor}_P^w(p)} f(x)\,\mathrm{d}\,x$$

$$\int_{\Omega\cap \mathrm{Vor}_P^w(p)} \|x-p\|^2 f(x)\,\mathrm{d}\,x$$

2. Iterative unconstrained convex programming:

- Choice of an initial weight vector, e.g. $w_0(p) := 0$ for all p.
- Computation of descent direction d_k
- Computation of time step s_k
- $w_{k+1} = w_k + s_k d_k$

Implementation of Convex Programming

1. Computation of Φ and $\nabla\Phi$:

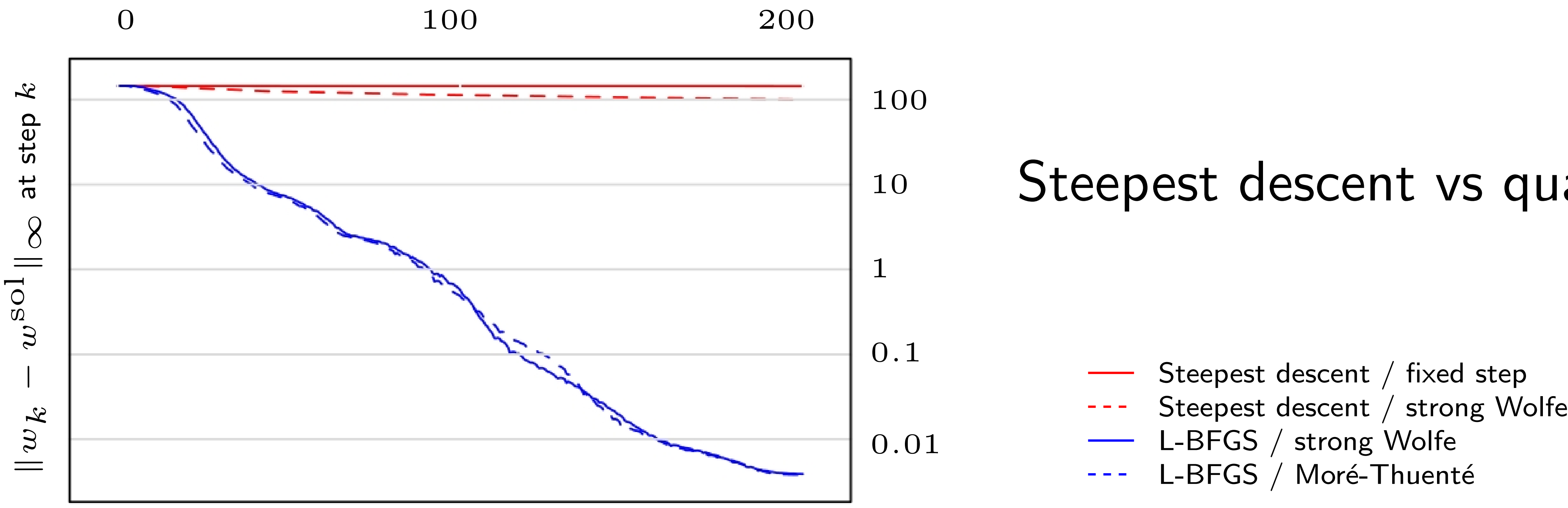
$$\int_{\Omega \cap \mathrm{Vor}_P^w(p)} f(x)\,\mathrm{d}\,x$$

$$\int_{\Omega \cap \mathrm{Vor}_P^w(p)} \|x-p\|^2 f(x)\,\mathrm{d}\,x$$

2. Iterative unconstrained convex programming:

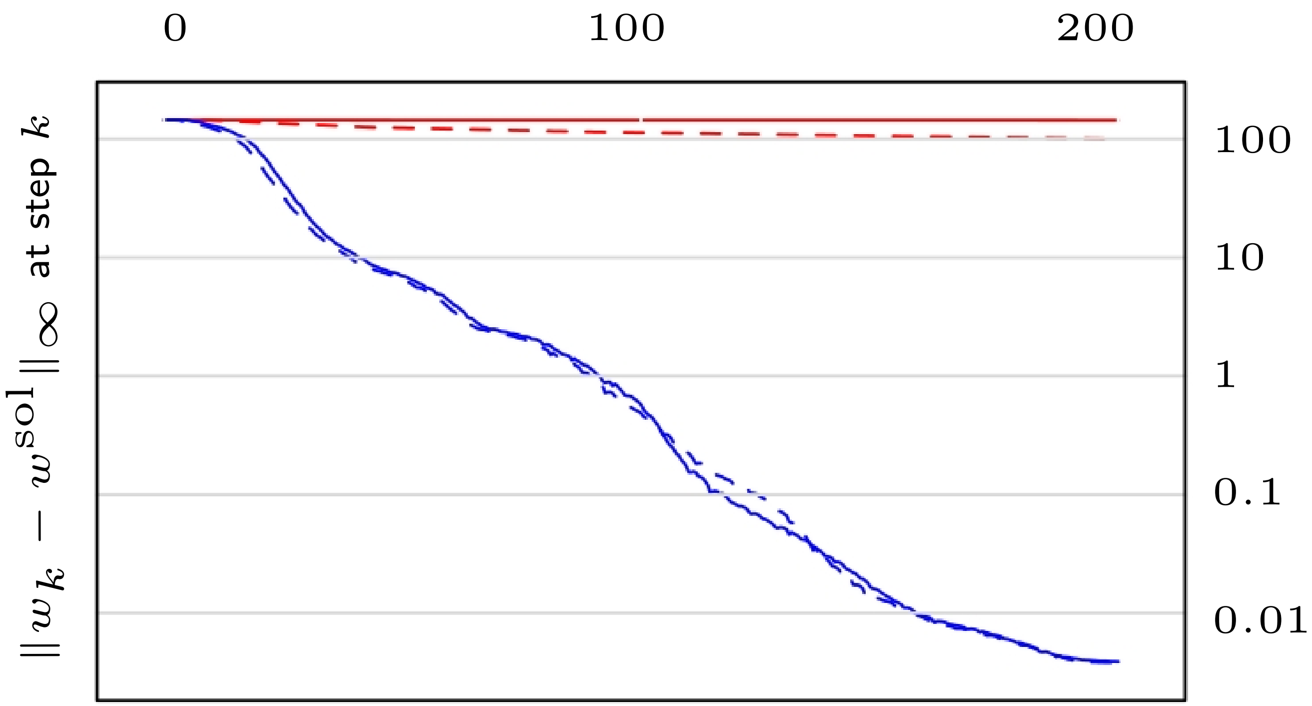
- Choice of an initial weight vector, e.g. $w_0(p) := 0$ for all p.
- Computation of descent direction d_k
- Computation of time step s_k
- $w_{k+1} = w_k + s_k d_k$

Comparison of Convex Optimization Methods



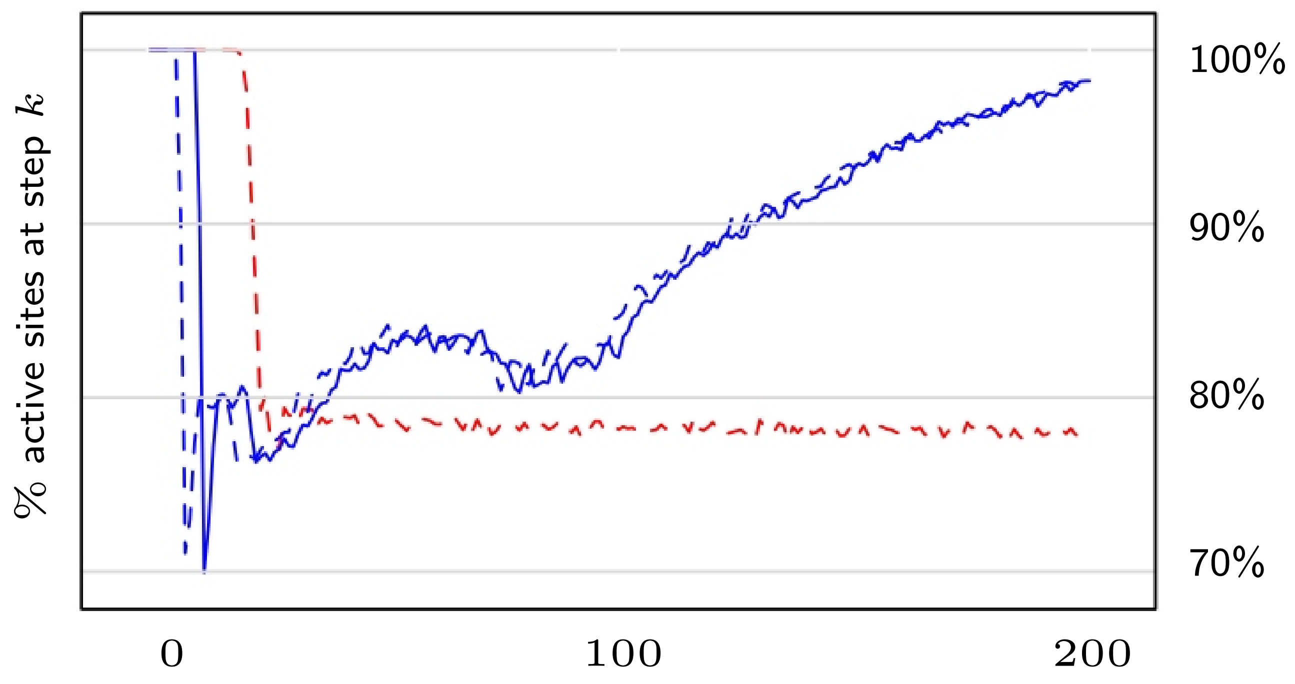
Steepest descent vs quasi-Newton

Comparison of Convex Optimization Methods

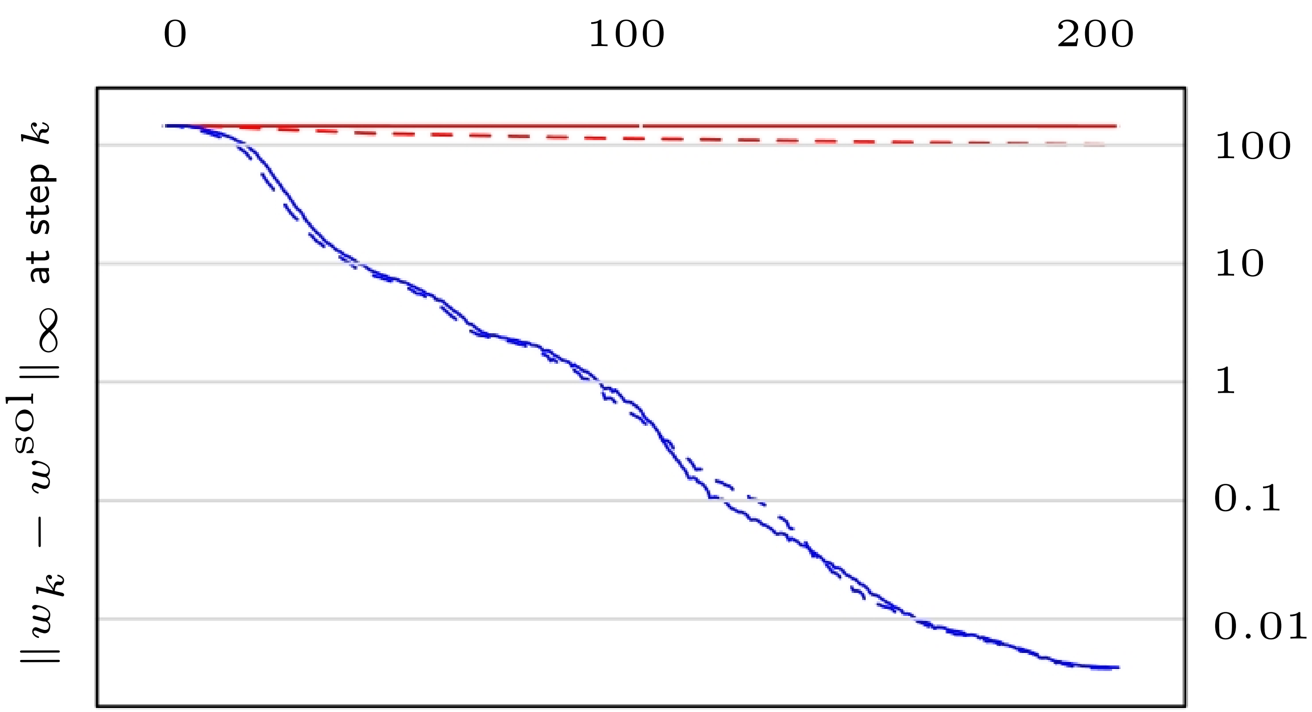


Steepest descent vs quasi-Newton

Number of sites with non-empty Power cell



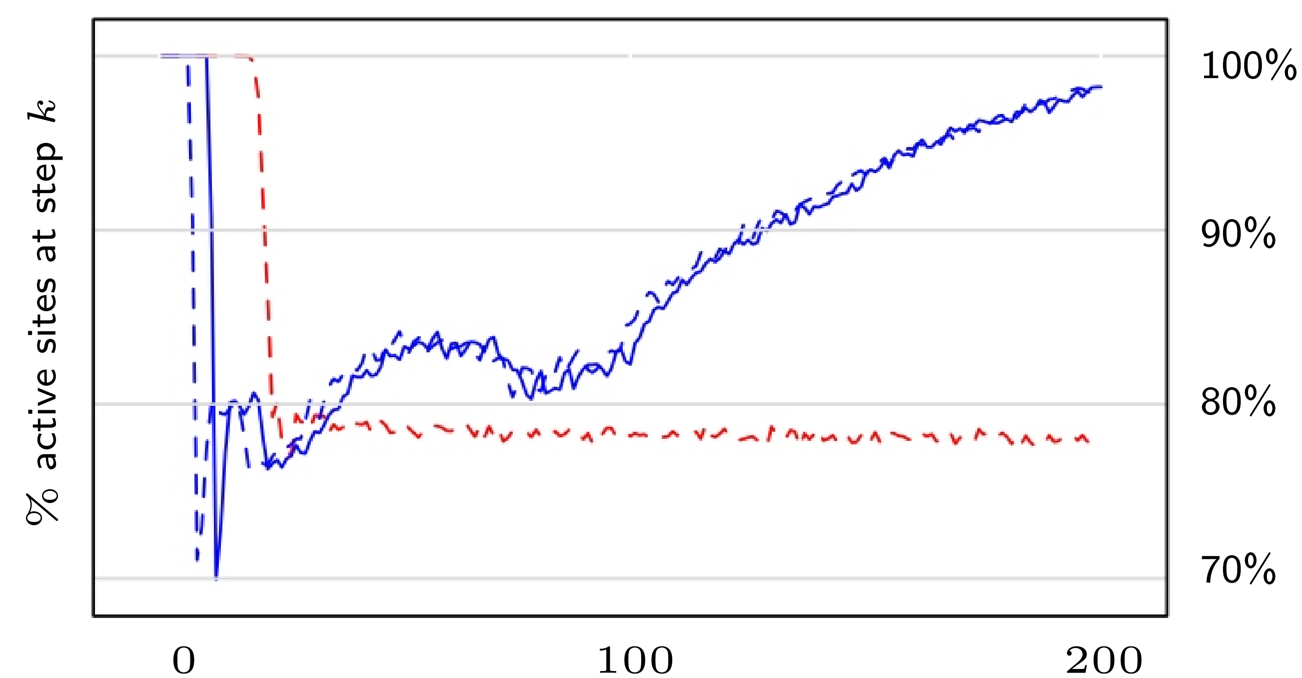
Comparison of Convex Optimization Methods



Steepest descent vs quasi-Newton

Number of sites with non-empty Power cell

$\Longrightarrow$ Need to recompute completely the Power diagram at every step



2. Multiscale approach

An Approximation Theorem

Proposition: Suppose the following:

- μ probability with density $f \geq m > 0$ on a bounded connected domain Ω with piecewise smooth boundary.
- (ν_n) and ν_∞ are are supported on finite sets $P_n \subseteq \Omega$, and $\lim \mathrm{W}_2(\nu_n, \nu_\infty) = 0.$

Let w_n be weights that solve OT between μ and ν_n. Then,

$$\forall p_n \in P_n, \quad \lim p_n = p \in P_\infty \Longrightarrow w_\infty(p) = \lim w_n(p_n)$$

An Approximation Theorem

Proposition: Suppose the following:

- μ probability with density $f \geq m > 0$ on a bounded connected domain Ω with piecewise smooth boundary.
- (ν_n) and ν_∞ are are supported on finite sets $P_n \subseteq \Omega$, and $\lim \mathrm{W}_2(\nu_n, \nu_\infty) = 0.$

Let w_n be weights that solve OT between μ and ν_n. Then,

$$\forall p_n \in P_n, \quad \lim p_n = p \in P_\infty \Longrightarrow w_\infty(p) = \lim w_n(p_n)$$

- Weights are defined up to an additive constant.
- Open question: a quantitative version of this theorem.

An Approximation Theorem – Sketch of Proof

Proposition: Suppose [...] $\lim \mathrm{W}_2(\nu_n, \nu_\infty) = 0$. Then,

$$\forall p_n \in P_n, \quad \lim p_n = p \in P_\infty \implies w_\infty(p) = \lim w_n(p_n)$$

An Approximation Theorem – Sketch of Proof

Proposition: Suppose [...] $\lim \mathrm{W}_2(\nu_n, \nu_\infty) = 0$. Then,

$$\forall p_n \in P_n, \quad \lim p_n = p \in P_\infty \Longrightarrow w_\infty(p) = \lim w_n(p_n)$$

Convex potential:

$$\phi_P^w(x) = \|x\|^2 - \min_{p \in P}(\|x - p\|^2 - w(p))$$
$$= \max_{p \in P} \langle x | p \rangle + \tfrac{1}{2}(w(p) - \|p\|^2)$$
$$\nabla \phi_P^w(x) = T_S^w$$

Zero-mean: we assume w.l.o.g. that $\int_\Omega \phi_S^w(x) \, \mathrm{d}\, \mu(x) = 0$

An Approximation Theorem – Sketch of Proof

Proposition: Suppose [...] $\lim \mathrm{W}_2(\nu_n, \nu_\infty) = 0$. Then,

$$\forall p_n \in P_n, \quad \lim p_n = p \in P_\infty \Longrightarrow w_\infty(p) = \lim w_n(p_n)$$

Zero-mean Convex potential: $\nabla \phi_n = T_n$ and $\int_\Omega \phi_n(x) f(x) \,\mathrm{d}\,\mu(x) = 0$.

- By stability of optimal transport plans, $\lim \|T_n - T_\infty\|_{\mathrm{L}^2(\mu)} = 0$.

An Approximation Theorem – Sketch of Proof

Proposition: Suppose [...] $\lim \mathrm{W}_2(\nu_n, \nu_\infty) = 0$. Then,

$$\forall p_n \in P_n, \quad \lim p_n = p \in P_\infty \Longrightarrow w_\infty(p) = \lim w_n(p_n)$$

Zero-mean Convex potential: $\nabla \phi_n = T_n$ and $\int_\Omega \phi_n(x) f(x) \,\mathrm{d}\,\mu(x) = 0$.

- By stability of optimal transport plans, $\lim \|T_n - T_\infty\|_{\mathrm{L}^2(\mu)} = 0$.

- By Poincaré inequality (assumptions on Ω and f),

$$\|\phi\|_{\mathrm{L}^2(\mu)} \leq \mathrm{cst} \times \|\nabla \phi\|_{\mathrm{L}^2(\mu)} \quad \text{provided that} \quad \int_\Omega \phi(x) \,\mathrm{d}\,\mu(x) = 0.$$

An Approximation Theorem – Sketch of Proof

Proposition: Suppose [...] $\lim \mathrm{W}_2(\nu_n, \nu_\infty) = 0$. Then,

$$\forall p_n \in P_n, \quad \lim p_n = p \in P_\infty \Longrightarrow w_\infty(p) = \lim w_n(p_n)$$

Zero-mean Convex potential: $\nabla \phi_n = T_n$ and $\int_\Omega \phi_n(x) f(x) \,\mathrm{d}\,\mu(x) = 0$.

- By stability of optimal transport plans, $\lim \|T_n - T_\infty\|_{\mathrm{L}^2(\mu)} = 0$.

- By Poincaré inequality: $\|\phi_n - \phi_\infty\|_{\mathrm{L}^2(\mu)} \leq \mathrm{cst} \times \|T_n - T_\infty\|$.

An Approximation Theorem – Sketch of Proof

Proposition: Suppose [...] $\lim \mathrm{W}_2(\nu_n, \nu_\infty) = 0$. Then,

$$\forall p_n \in P_n, \quad \lim p_n = p \in P_\infty \implies w_\infty(p) = \lim w_n(p_n)$$

Zero-mean Convex potential: $\nabla \phi_n = T_n$ and $\int_\Omega \phi_n(x) f(x) \,\mathrm{d}\,\mu(x) = 0.$

- By stability of optimal transport plans, $\lim \|T_n - T_\infty\|_{\mathrm{L}^2(\mu)} = 0$.
- By Poincaré inequality: $\|\phi_n - \phi_\infty\|_{\mathrm{L}^2(\mu)} \leq \mathrm{cst} \times \|T_n - T_\infty\|$.
- Since ϕ_n and ϕ_∞ are Lipschitz, $\lim \|\phi_n - \phi_\infty\|_{\mathrm{L}^\infty(\Omega)} = 0$.

An Approximation Theorem – Sketch of Proof

Proposition: Suppose [...] $\lim \mathrm{W}_2(\nu_n, \nu_\infty) = 0$. Then,

$$\forall p_n \in P_n, \quad \lim p_n = p \in P_\infty \implies w_\infty(p) = \lim w_n(p_n)$$

Zero-mean Convex potential: $\nabla \phi_n = T_n$ and $\int_\Omega \phi_n(x) f(x) \,\mathrm{d}\,\mu(x) = 0$.

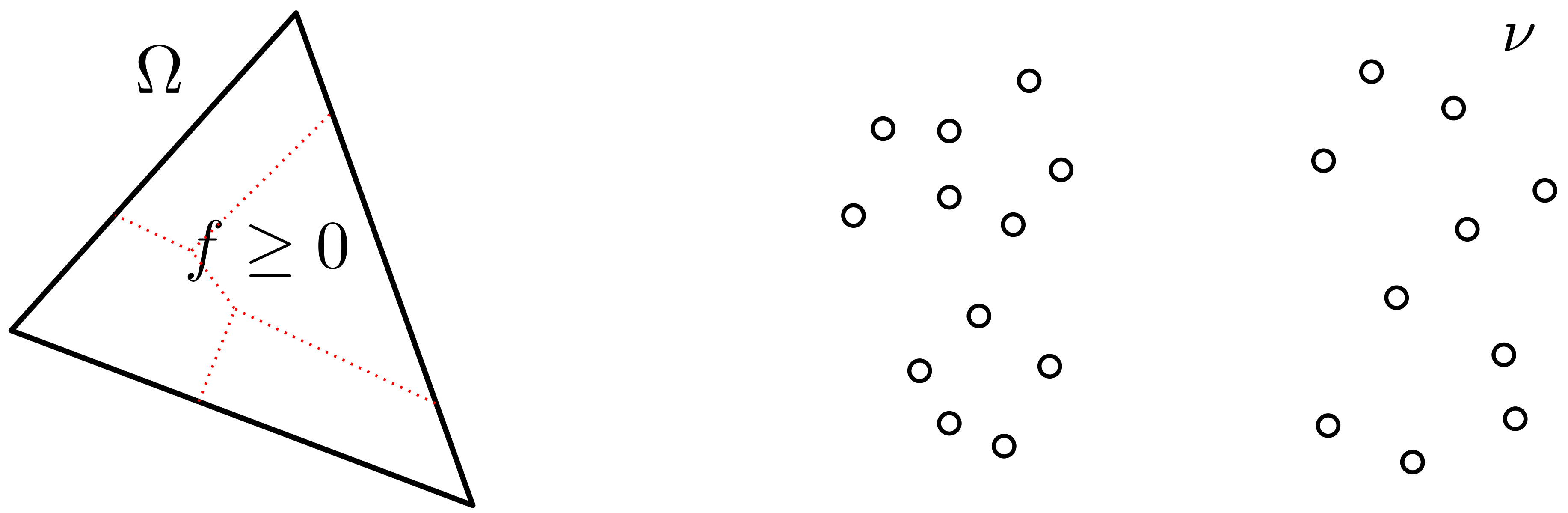
- By stability of optimal transport plans, $\lim \|T_n - T_\infty\|_{\mathrm{L}^2(\mu)} = 0$.
- By Poincaré inequality: $\|\phi_n - \phi_\infty\|_{\mathrm{L}^2(\mu)} \leq \mathrm{cst} \times \|T_n - T_\infty\|$.
- Since ϕ_n and ϕ_∞ are Lipschitz, $\lim \|\phi_n - \phi_\infty\|_{\mathrm{L}^\infty(\Omega)} = 0$.

With a bit more work, this result implies the conclusion of the theorem.

Two-scale Approach for Optimization

Goal: Given a measure μ with density, and ν supported on P, $|P| = N$,

minimise $\Phi : w \mapsto \ldots$



Two-scale Approach for Optimization

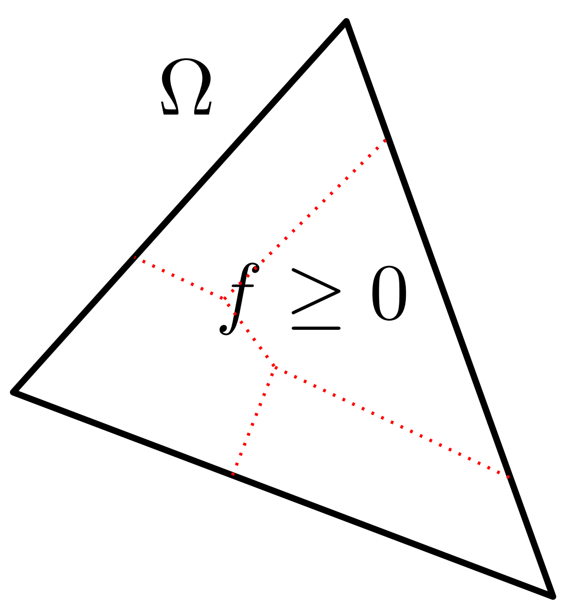
Goal: Given a measure μ with density, and ν supported on P, $|P| = N$,

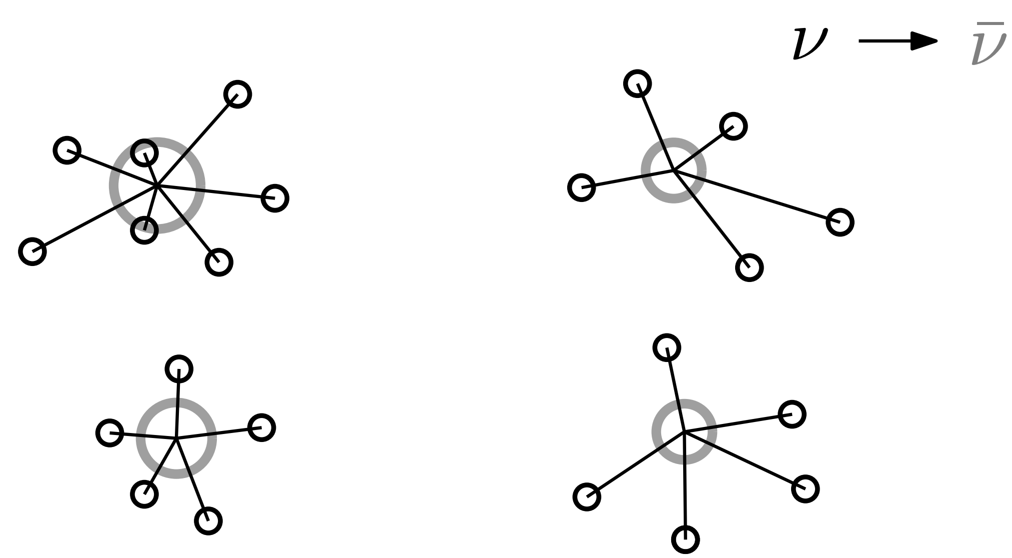
minimise $\Phi : w \mapsto \ldots$

Approach:

- Replace ν by $\bar{\nu}$ supported on $\bar{P}$, $|\bar{P}| = n \ll N$ points.

$$\bar{\nu} = \arg\min\{\mathrm{W}_2(\nu, \bar{\nu});\ |\mathrm{spt}(\bar{\nu})| \leq n\}$$





Two-scale Approach for Optimization

Goal: Given a measure μ with density, and ν supported on P, $|P| = N$,

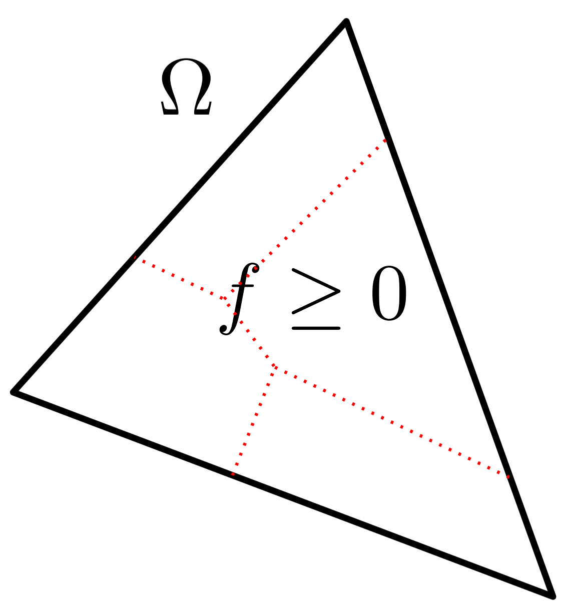
minimise $\Phi : w \mapsto \ldots$

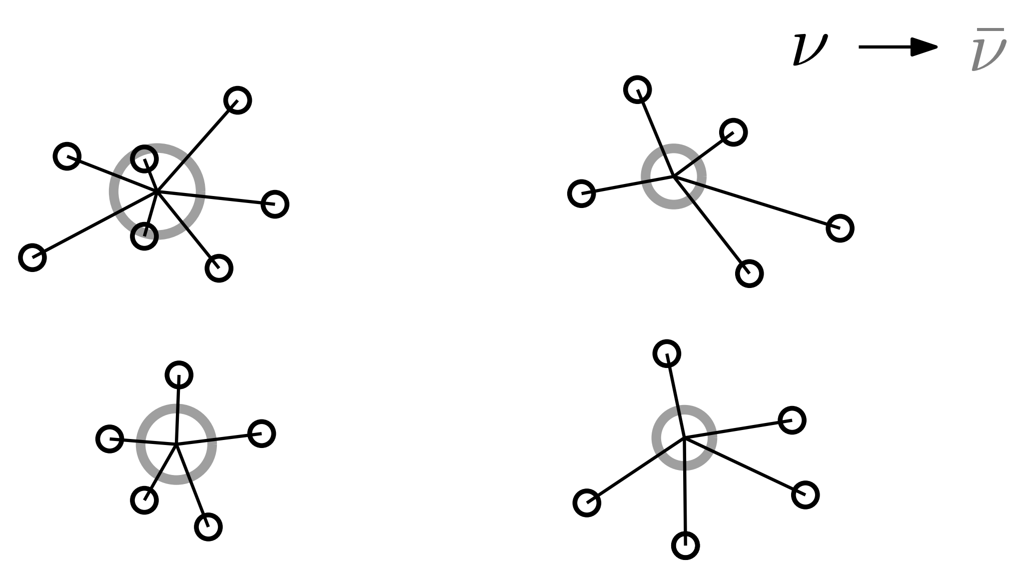
Approach:

- Replace ν by $\bar{\nu}$ supported on $\bar{P}$, $|\bar{P}| = n \ll N$ points.

$$\bar{\nu} = \arg\min\{W_2(\nu, \bar{\nu});\ |\mathrm{spt}(\bar{\nu})| \leq n\}$$

k-means / Lloyd's algorithm (local minimum)





Two-scale Approach for Optimization

Goal: Given a measure μ with density, and ν supported on P, $|P| = N$,

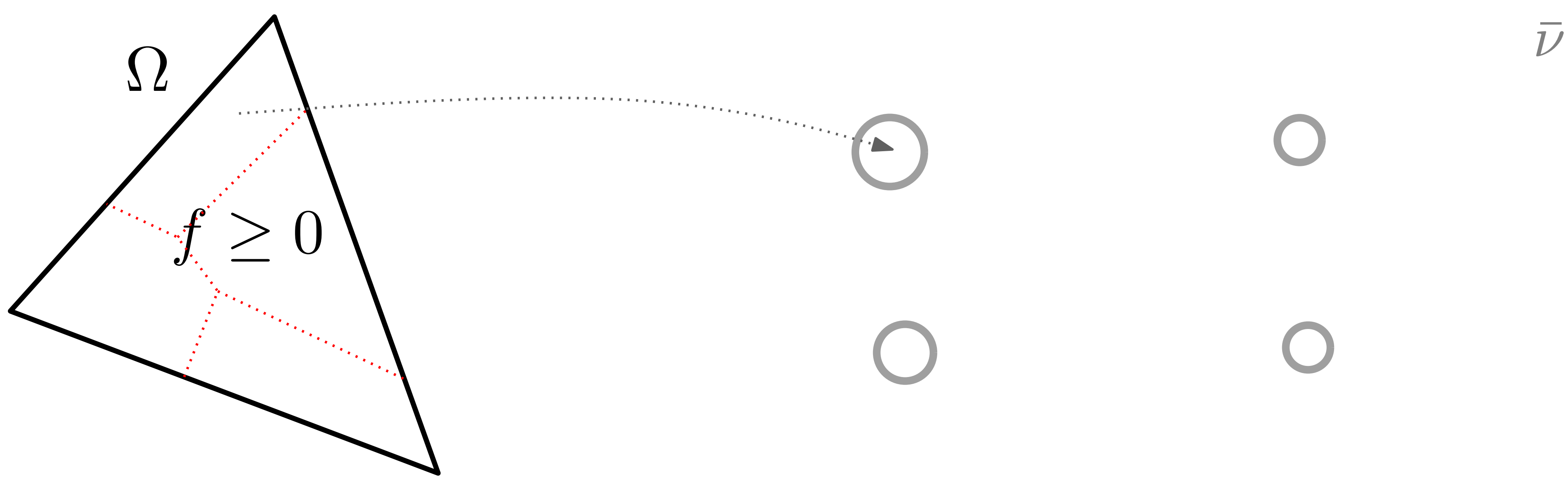
minimise $\Phi : w \mapsto \ldots$

Approach:

- Replace ν by $\bar{\nu}$ supported on $\bar{P}$, $|\bar{P}| = n \ll N$ points.

$$\bar{\nu} = \arg\min\{\mathrm{W}_2(\nu, \bar{\nu});\ |\mathrm{spt}(\bar{\nu})| \leq n\}$$

- Solve the OT from μ to $\bar{\nu}$, constructing $\bar{w} : \bar{P} \to \mathbb{R}$.



Two-scale Approach for Optimization

Goal: Given a measure μ with density, and ν supported on P, $|P| = N$,

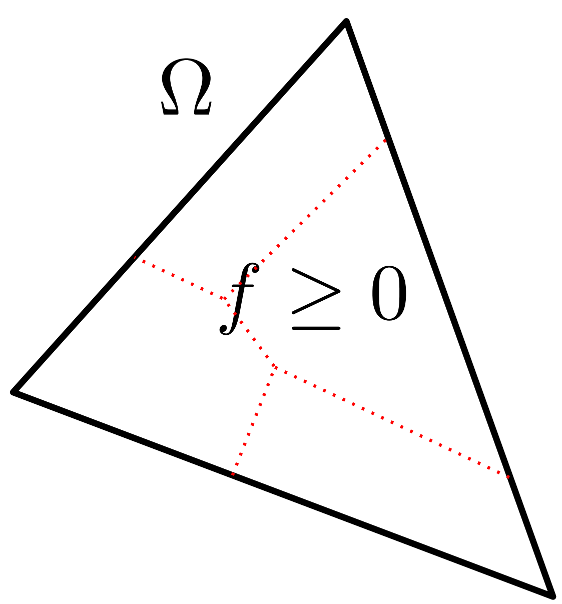
minimise $\Phi : w \mapsto \ldots$

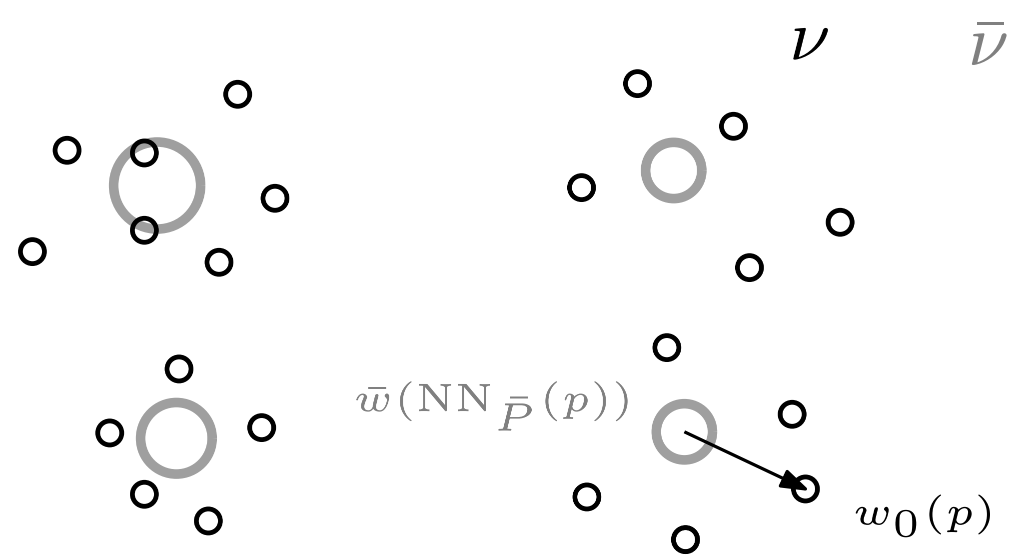
Approach:

- Replace ν by $\bar{\nu}$ supported on $\bar{P}$, $|\bar{P}| = n \ll N$ points.

$$\bar{\nu} = \arg\min\{\mathrm{W}_2(\nu, \bar{\nu});\ |\mathrm{spt}(\bar{\nu})| \leq n\}$$

- Solve the OT from μ to $\bar{\nu}$, constructing $\bar{w} : \bar{P} \to \mathbb{R}$.
- Minimize Φ starting from $w_0 : p \in P \mapsto \bar{w}(\mathrm{NN}_{\bar{P}}(p))$.





Two-scale Approach for Optimization

Goal: Given a measure μ with density, and ν supported on P, $|P| = N$,

minimise $\Phi : w \mapsto \ldots$

Approach:

- Replace ν by $\bar{\nu}$ supported on $\bar{P}$, $|\bar{P}| = n \ll N$ points.

$$\bar{\nu} = \arg\min\{\mathrm{W}_2(\nu, \bar{\nu});\ |\mathrm{spt}(\bar{\nu})| \leq n\}$$

- Solve the OT from μ to $\bar{\nu}$, constructing $\bar{w} : \bar{P} \to \mathbb{R}$.
- Minimize Φ starting from $w_0 : p \in P \mapsto \bar{w}(\mathrm{NN}_{\bar{P}}(p))$.

$$\lim \mathrm{W}_2(\nu_n, \nu_\infty) = 0$$

$$\lim p_n = p \in P_\infty \Longrightarrow \lim w_\infty(p) = \lim w_n(p_n)$$

Summary of the multiscale-scale algorithm

Input: a measure μ with density and a discrete measure ν on $\mathbb{R}^2$.

- Compute a sequence of discretizations of the target measure: $\nu_0 := \nu, \dots, \nu_L$, s.t. ν_ℓ is supported P_ℓ with $Nk^{-\ell}$ points.

Summary of the multiscale-scale algorithm

Input: a measure μ with density and a discrete measure ν on $\mathbb{R}^2$.

- Compute a sequence of discretizations of the target measure: $\nu_0 := \nu, \dots, \nu_L$, s.t. ν_ℓ is supported P_ℓ with $Nk^{-\ell}$ points.

- Solve OT from μ to ν_L starting with $w_L := 0$.

Summary of the multiscale-scale algorithm

Input: a measure μ with density and a discrete measure ν on $\mathbb{R}^2$.

- Compute a sequence of discretizations of the target measure: $\nu_0 := \nu, \ldots, \nu_L$, s.t. ν_ℓ is supported P_ℓ with $Nk^{-\ell}$ points.

- Solve OT from μ to ν_L starting with $w_L := 0$.

$\ldots$

- Solve OT from μ to ν_ℓ starting from $w_\ell(p) := w_{\ell+1}(\mathrm{NN}_{P_\ell}(p))$.

Summary of the multiscale-scale algorithm

Input: a measure μ with density and a discrete measure ν on $\mathbb{R}^2$.

- Compute a sequence of discretizations of the target measure: $\nu_0 := \nu, \dots, \nu_L$, s.t. ν_ℓ is supported P_ℓ with $Nk^{-\ell}$ points.
- Solve OT from μ to ν_L starting with $w_L := 0$.

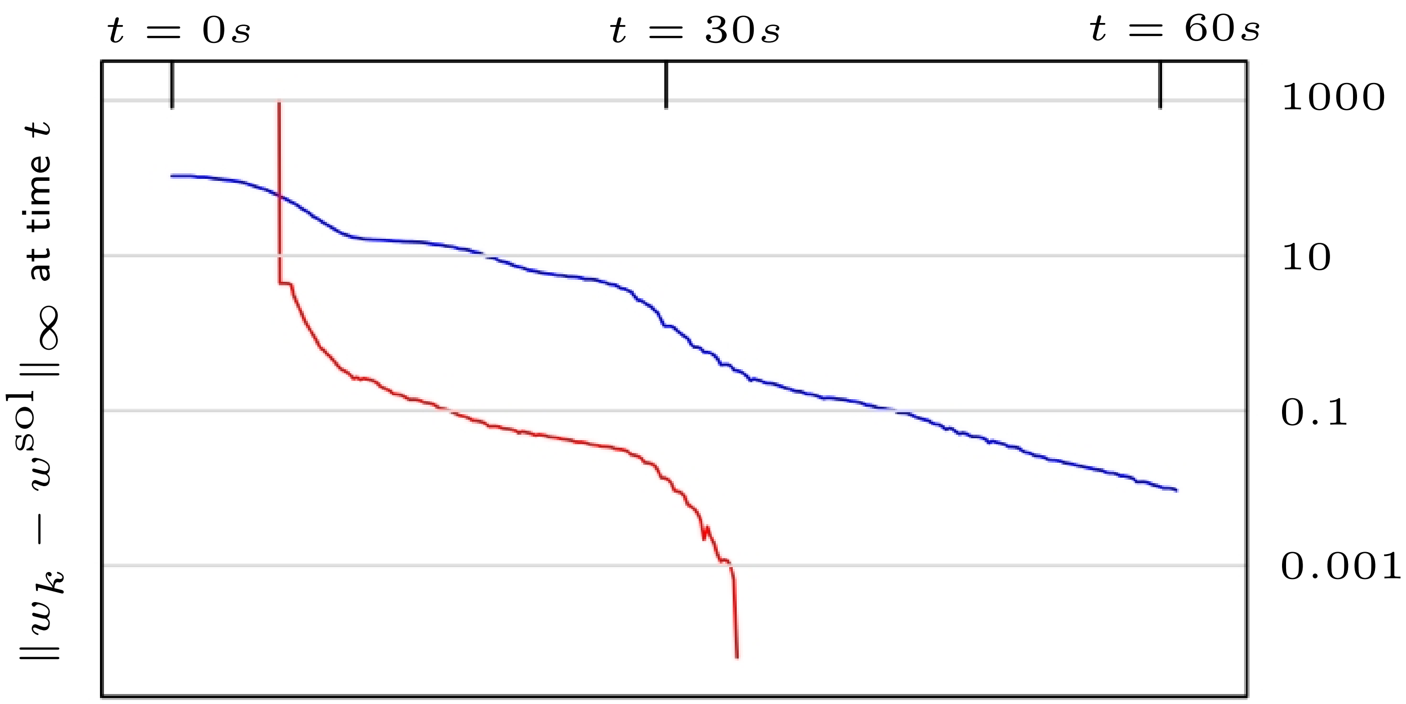
$\dots$

- Solve OT from μ to ν_ℓ starting from $w_\ell(p) := w_{\ell+1}(\mathrm{NN}_{P_\ell}(p))$.

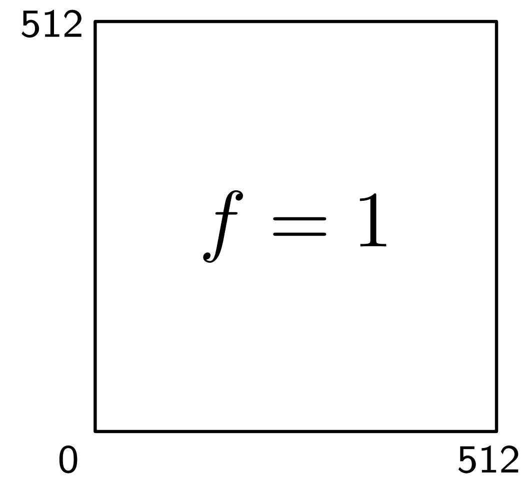
Remark: If the target measure is not discrete, one can obtain a first discretisation by an application of Lloyd's algorithm.

3. Experiments

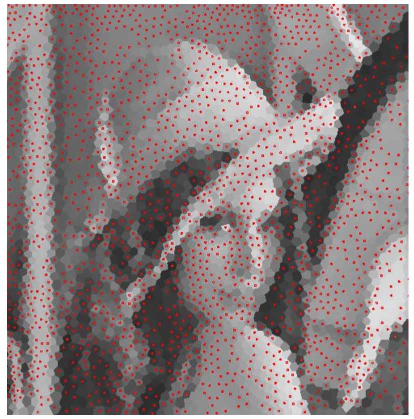
Multiscale vs Original — Convergence Speed



Source

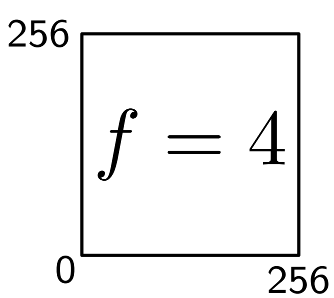


Target

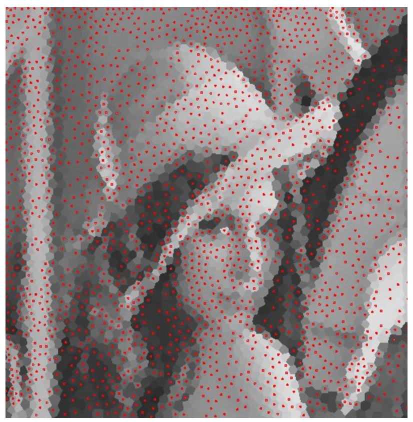


3k Lloyd sampling of Lena

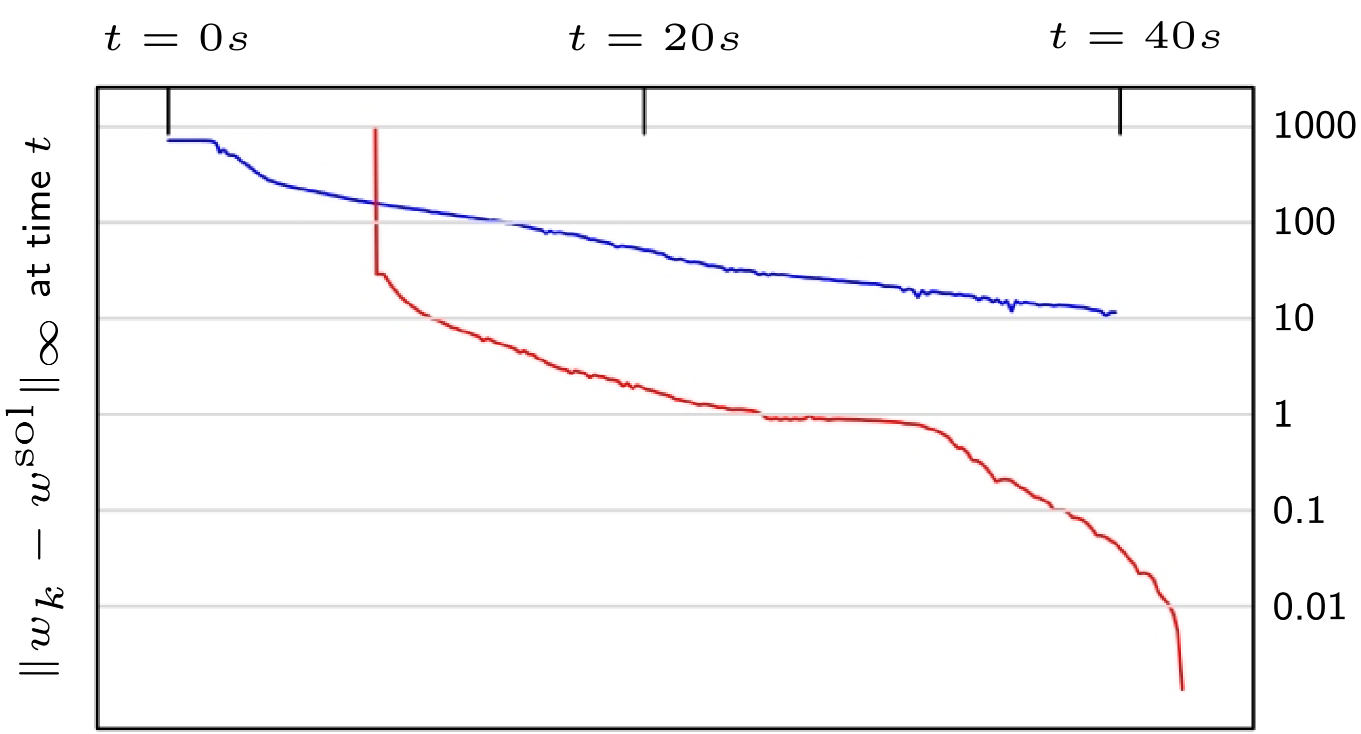
Source



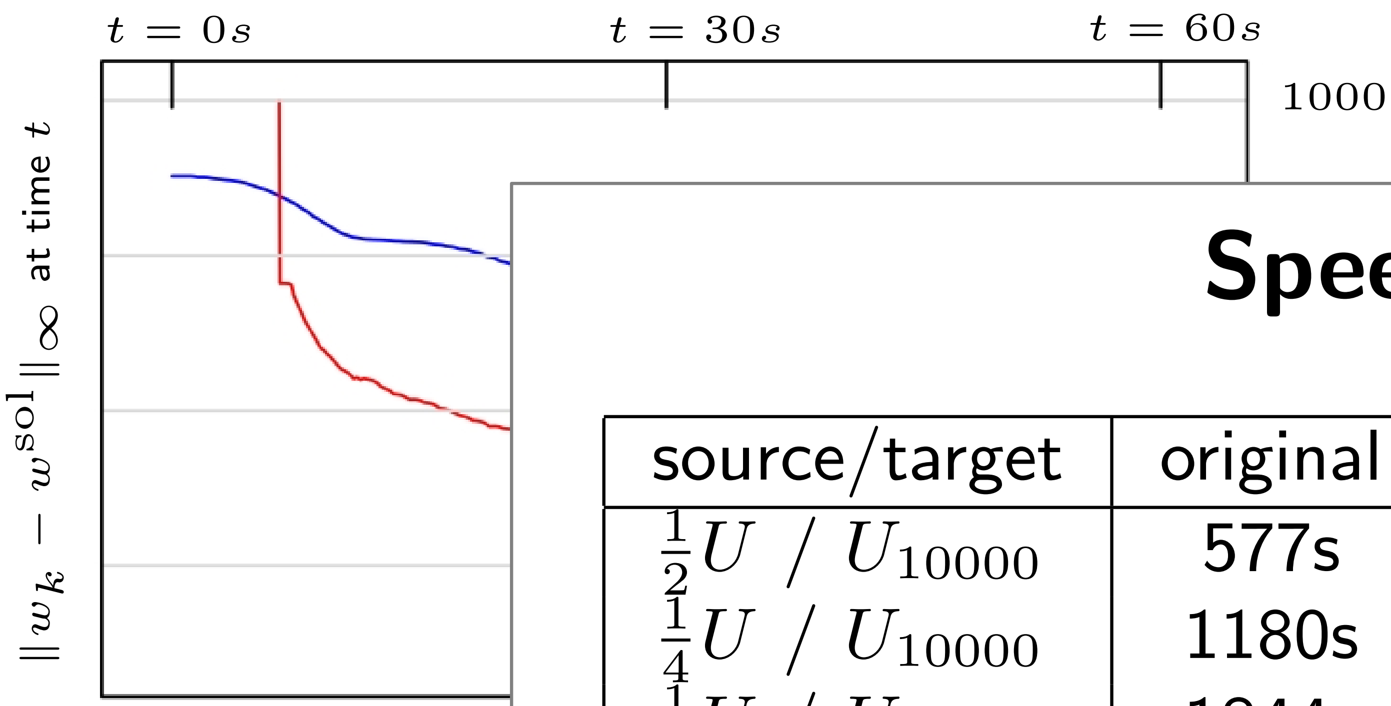
Target

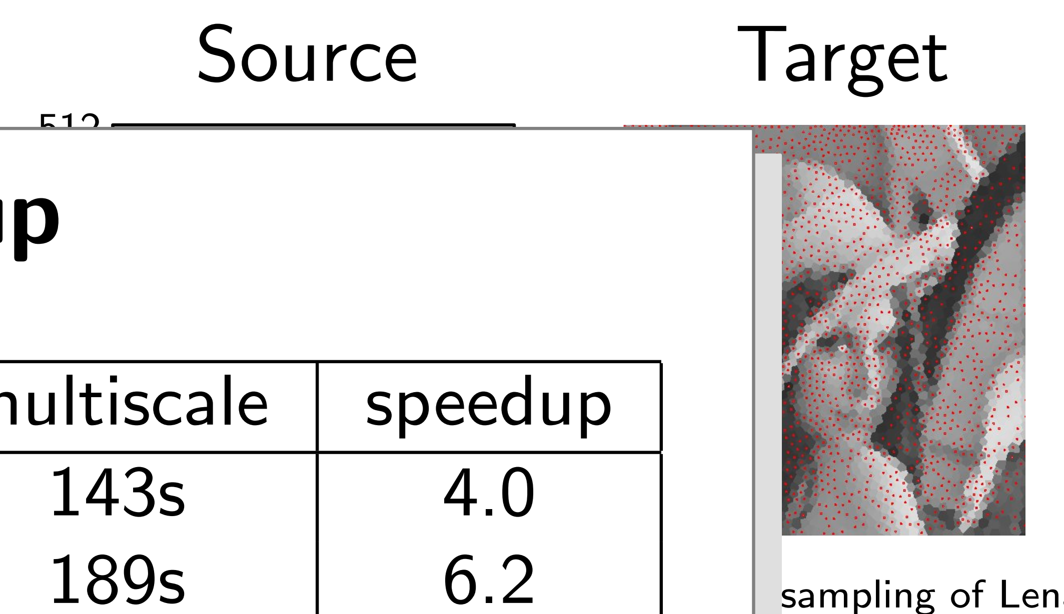


3k Lloyd sampling of Lena



Multiscale vs Original — Convergence Speed





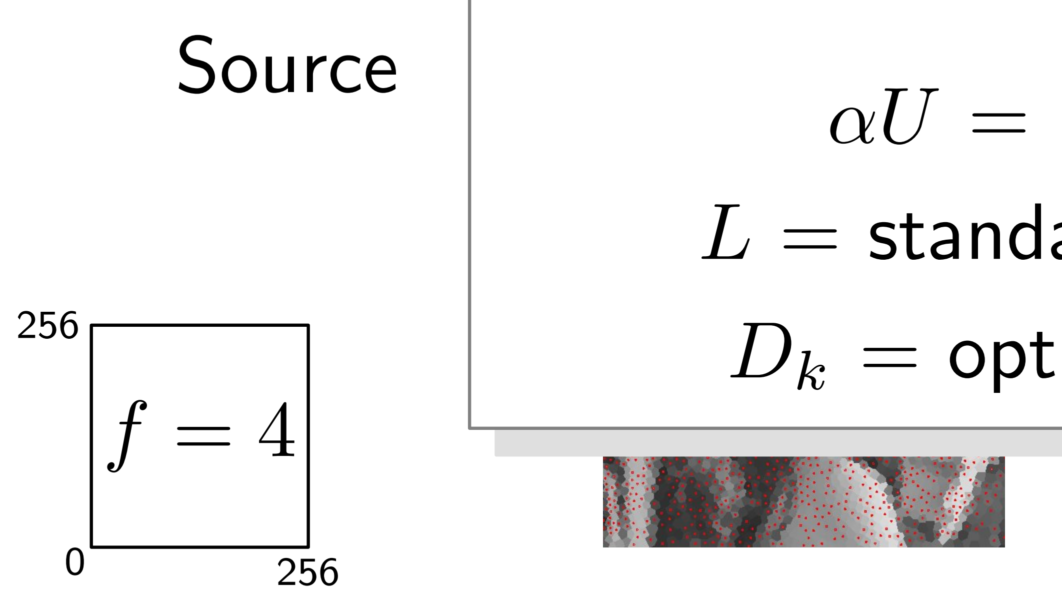
Speedup

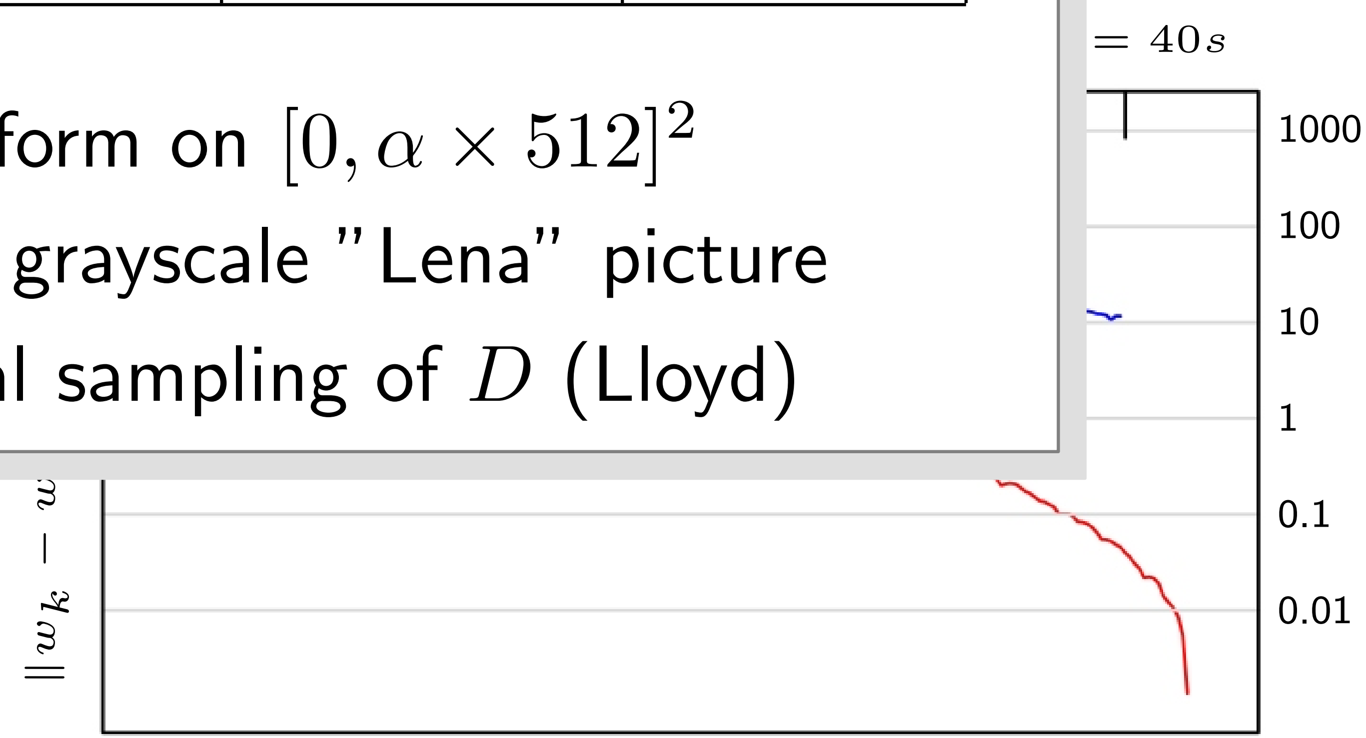
source/target	original	multiscale	speedup
$\frac{1}{2}U$ / U_{10000}	577s	143s	4.0
$\frac{1}{4}U$ / U_{10000}	1180s	189s	6.2
$\frac{1}{8}U$ / U_{10000}	1844s	241s	7.6
U / L_{10000}	216s	52s	4.1

$\alpha U =$ uniform on $[0, \alpha \times 512]^2$

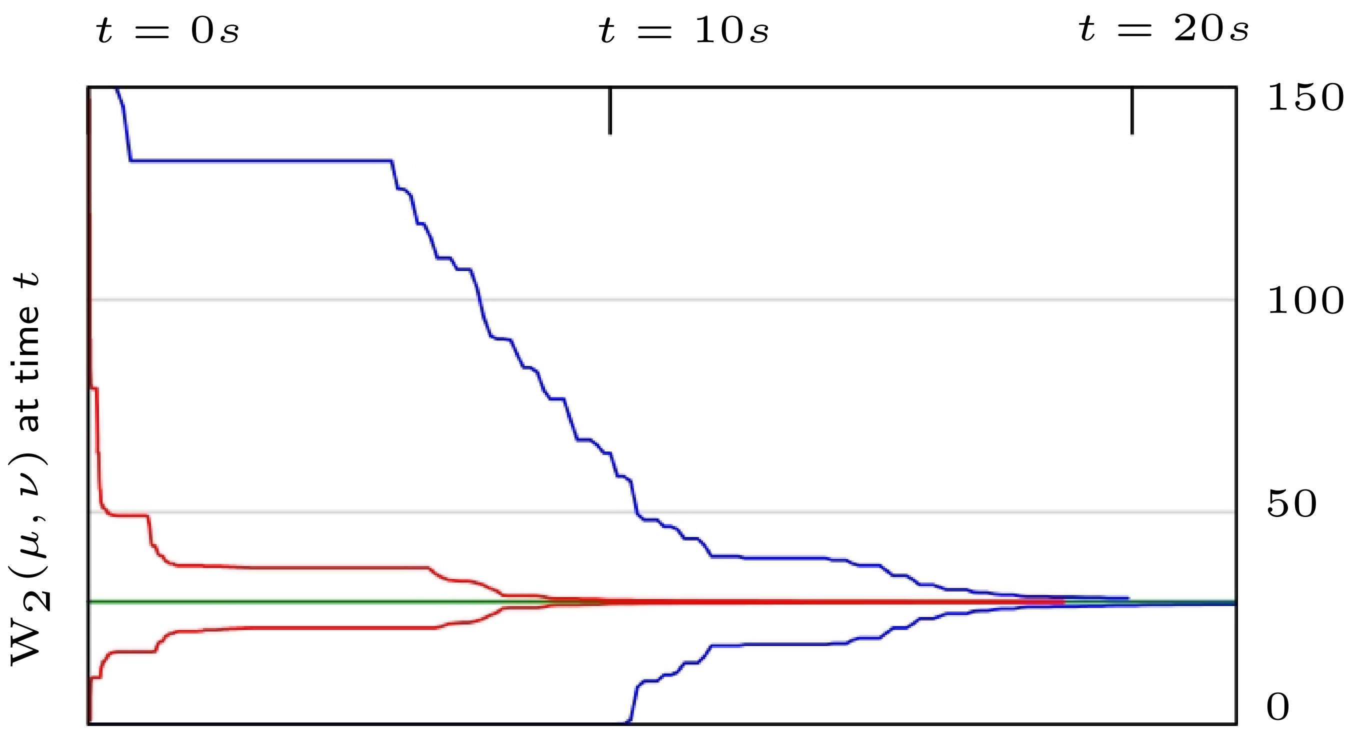
$L =$ standard grayscale "Lena" picture

$D_k =$ optimal sampling of D (Lloyd)

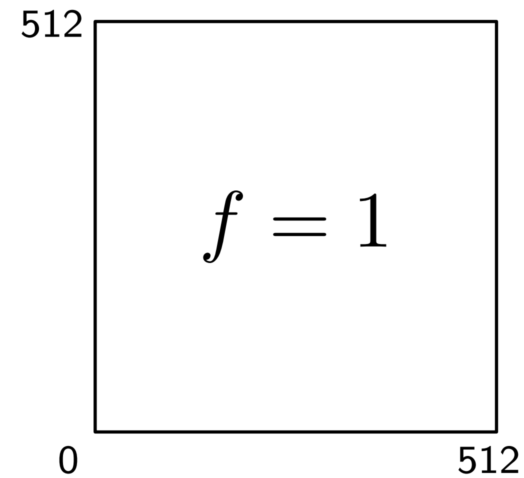




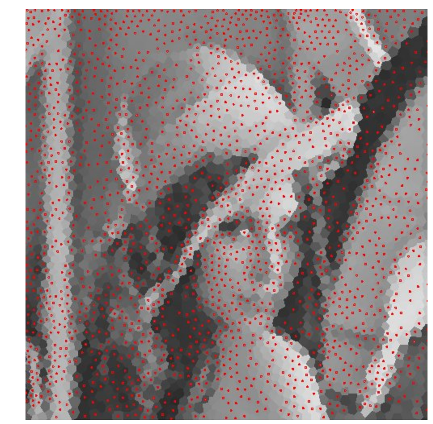
Multiscale vs Original — Wasserstein



Source

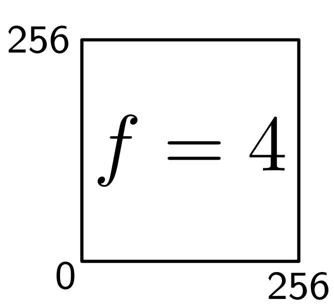


Target

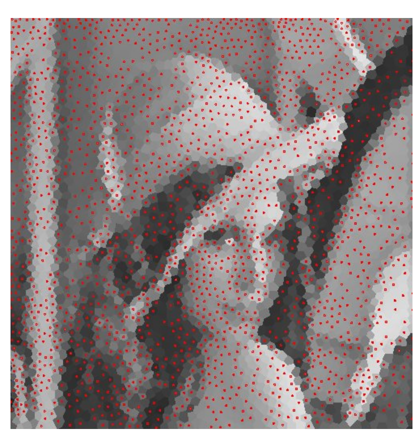


3k Lloyd sampling of Lena

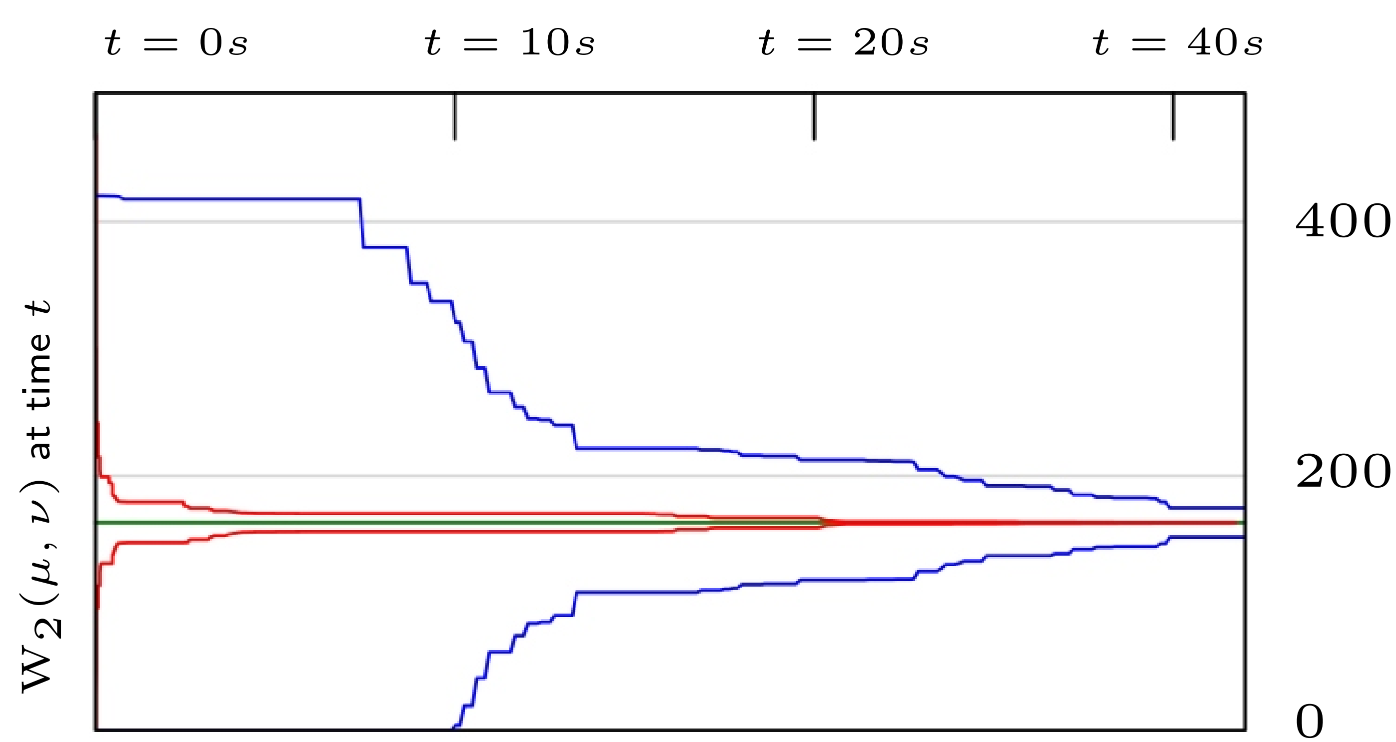
Source



Target



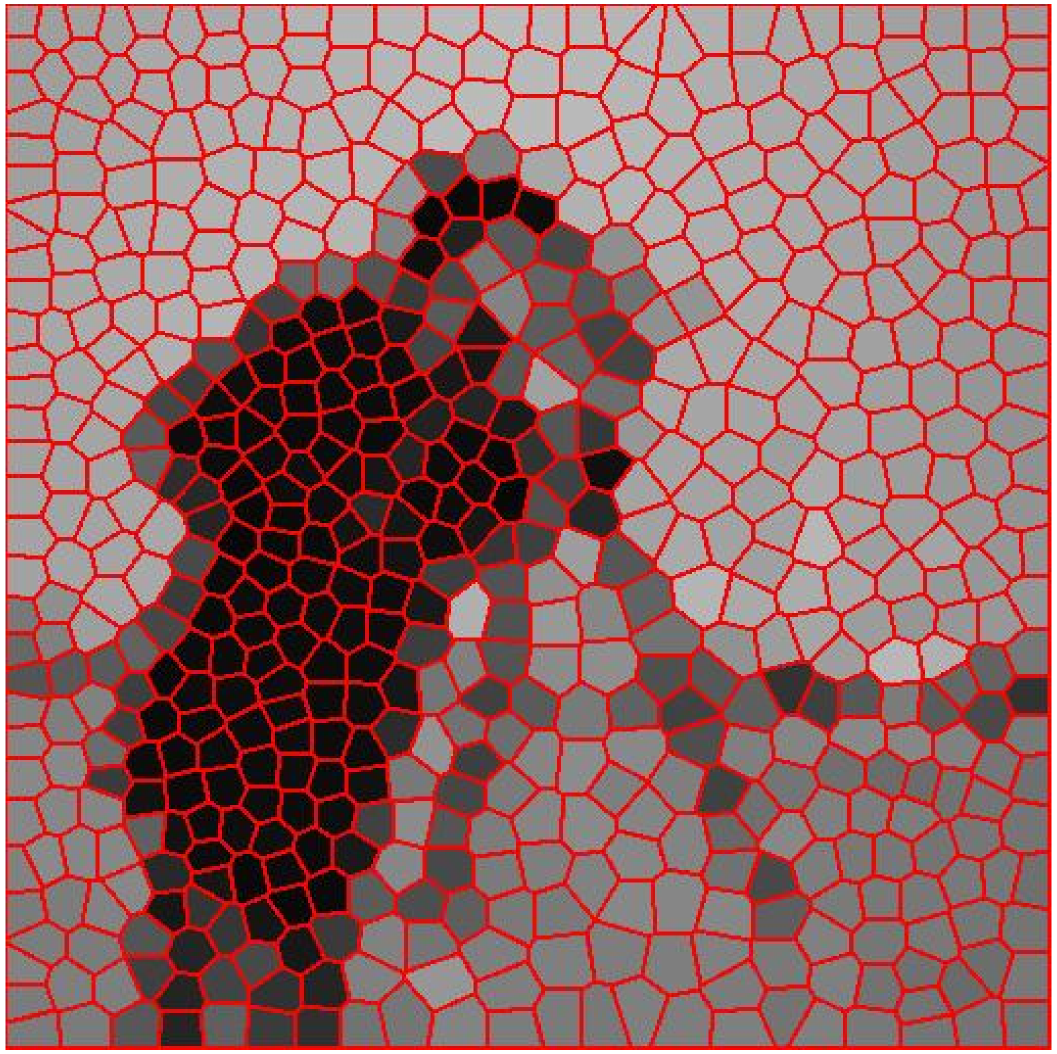
3k Lloyd sampling of Lena



Some Pictures of Optimal Transport Plans

Source: picture "Cameraman"
Target: Lloyd sampling of picture "Peppers" ($k = 625$)

"Displacement interpolation"
McCann '97



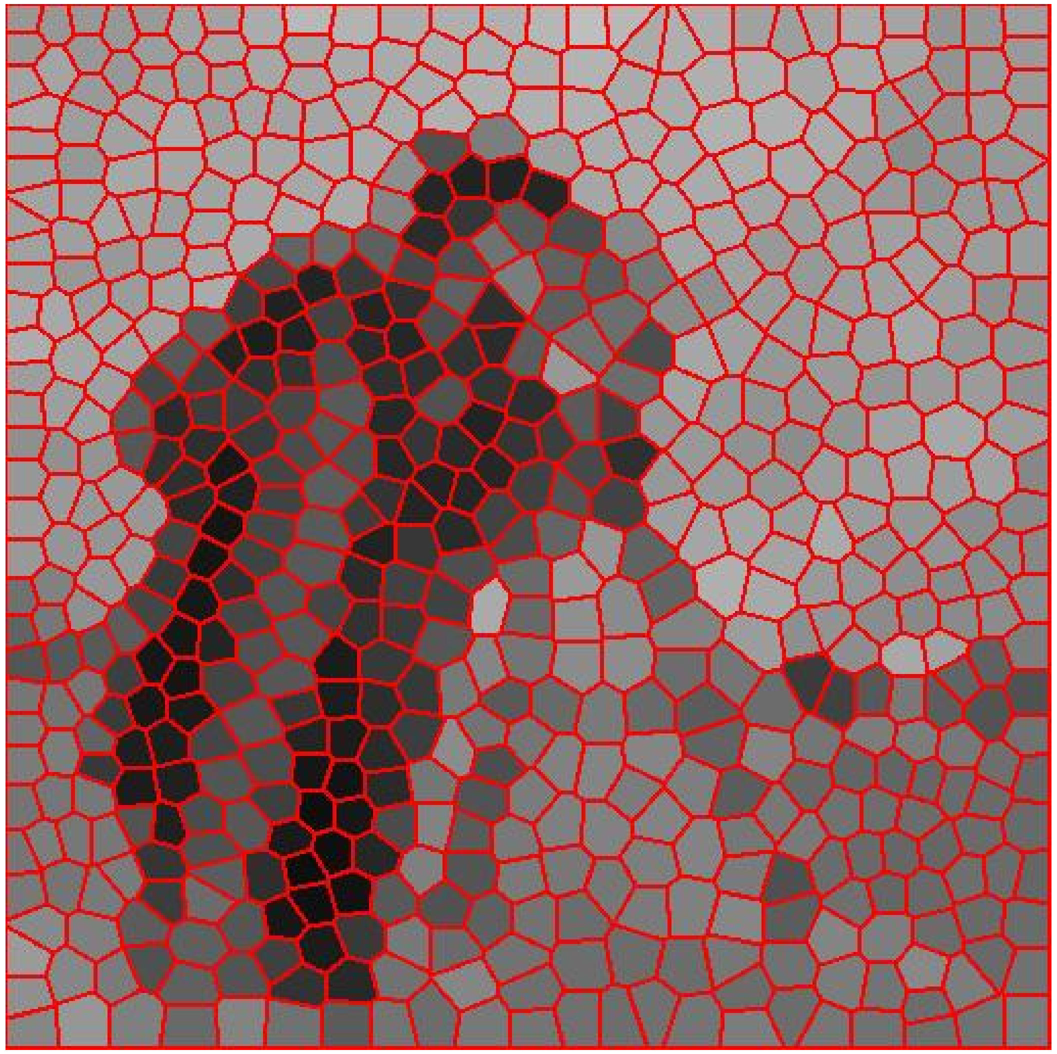
The mass of Dirac at p is spread onto $\mathrm{Vor}_P^w(p)$

$$w = w^{\mathrm{sol}}$$

Some Pictures of Optimal Transport Plans

Source: picture "Cameraman"
Target: Lloyd sampling of picture "Peppers" $(k = 625)$

"Displacement interpolation"
McCann '97



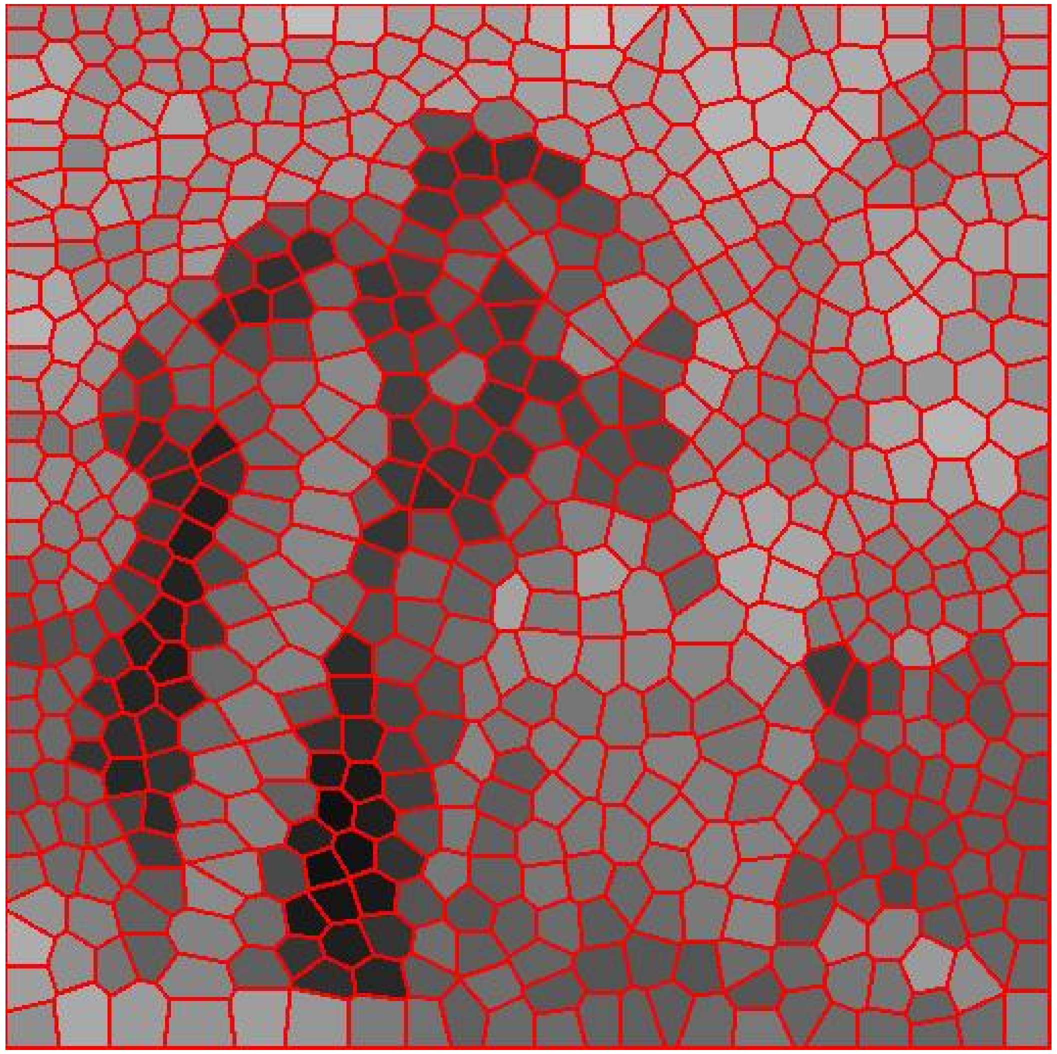
The mass of Dirac at p is spread onto $\mathrm{Vor}_P^w(p)$

$$w = w^{\mathrm{sol}} - \frac{1}{4} w^{\mathrm{sol}}$$

Some Pictures of Optimal Transport Plans

Source: picture "Cameraman"
Target: Lloyd sampling of picture "Peppers" ($k = 625$)

"Displacement interpolation"
McCann '97



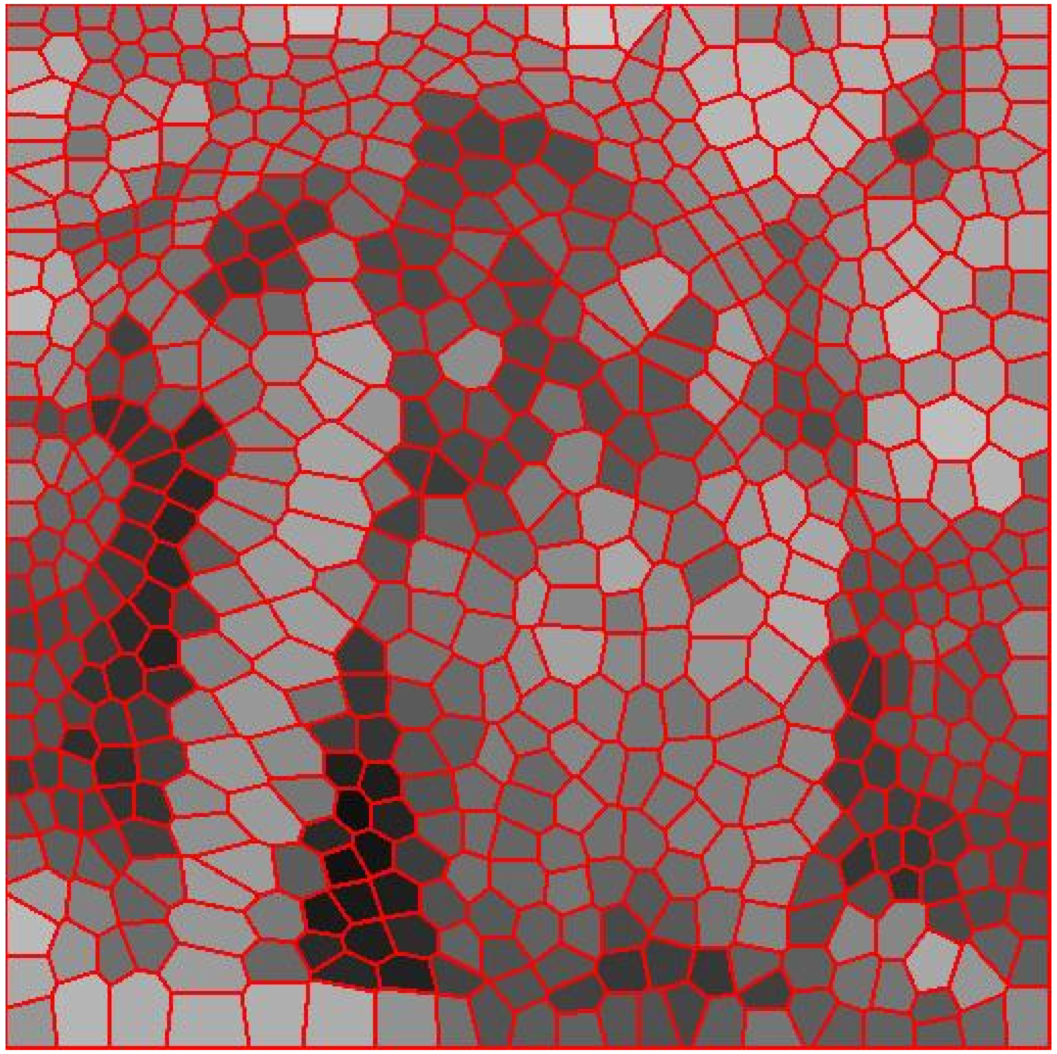
The mass of Dirac at p is spread onto $\mathrm{Vor}_P^w(p)$

$$w = w^{\mathrm{sol}} - \frac{1}{2} w^{\mathrm{sol}}$$

Some Pictures of Optimal Transport Plans

Source: picture "Cameraman"
Target: Lloyd sampling of picture "Peppers" ($k = 625$)

"Displacement interpolation"
McCann '97



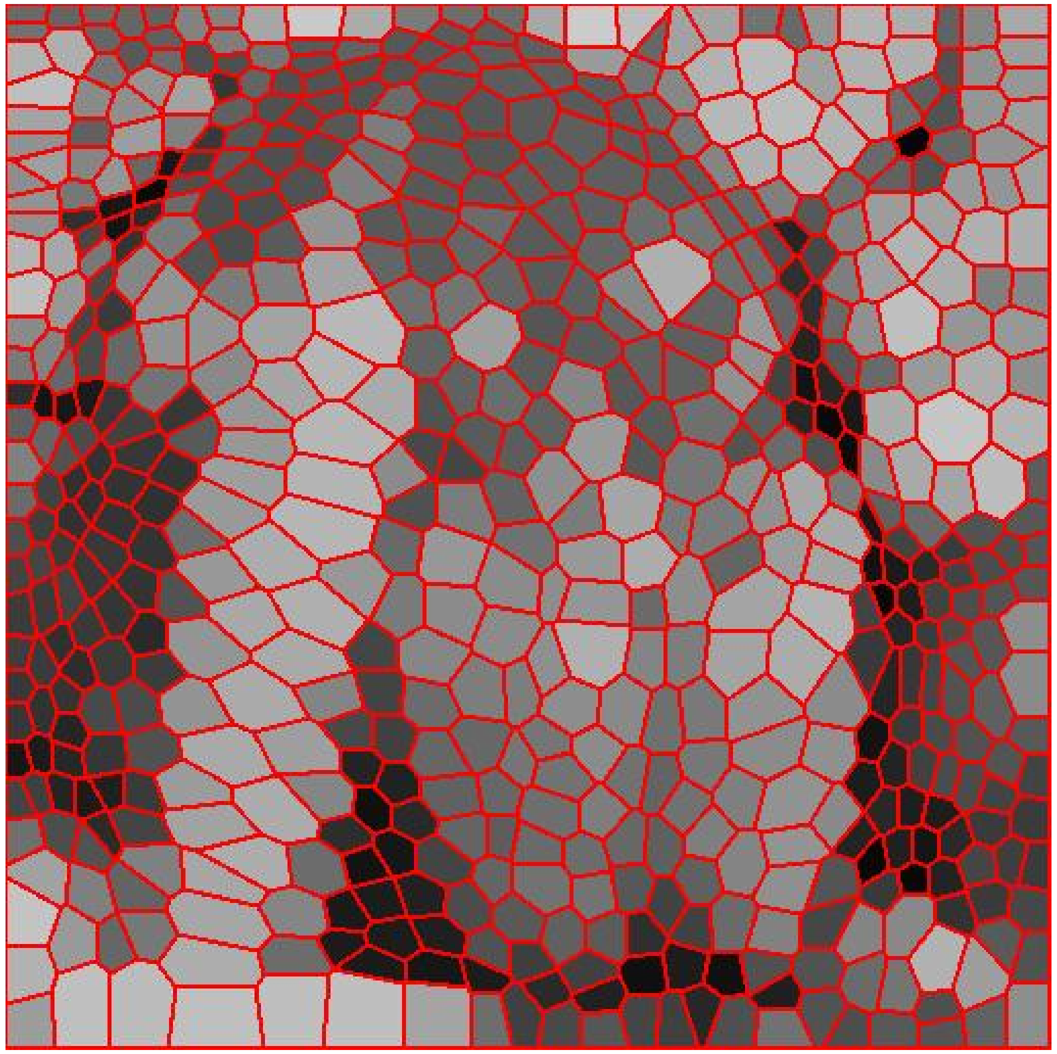
The mass of Dirac at p is spread onto $\mathrm{Vor}_P^w(p)$

$$w = w^{\mathrm{sol}} - \frac{3}{4} w^{\mathrm{sol}}$$

Some Pictures of Optimal Transport Plans

Source: picture "Cameraman"
Target: Lloyd sampling of picture "Peppers" ($k = 625$)

"Displacement interpolation"
McCann '97

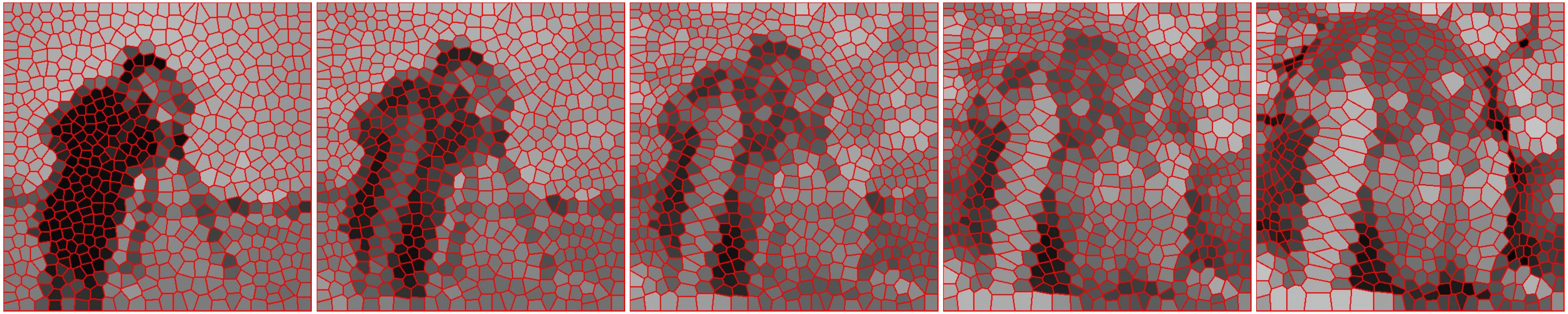


The mass of Dirac at p is spread onto $\mathrm{Vor}_P^w(p)$

$$w = 0$$

Some Pictures of Optimal Transport Plans

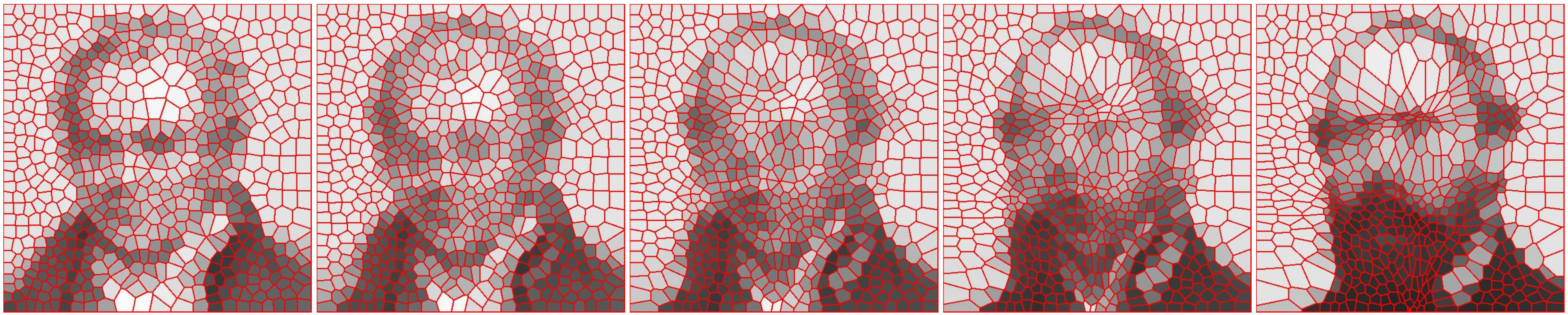
$$k = 625$$



$$k = 15000$$

Some Pictures of Optimal Transport Plans

$$k = 625$$



$$k = 15000$$

4. Assignment problem

ongoing work with Édouard Oudet

L^2 Assignment Problem

Problem: Given $P, Q \subseteq \mathbb{R}^d$ with $|P| = |Q| = N$,
find one-to-one $\sigma : Q \to P$ minimizing $\sum_{q \in Q} \|q - \sigma(q)\|^2$.

L^2 Assignment Problem

Problem: Given $P, Q \subseteq \mathbb{R}^d$ with $|P| = |Q| = N$,
find one-to-one $\sigma : Q \to P$ minimizing $\sum_{q \in Q} \|q - \sigma(q)\|^2$.

$\iff$ minimising $\Psi(v, w) = -\sum_{q \in Q} v(q) + \sum_{p \in P} w(p)$

under the constraint $v(q) - w(p) \leq \|p - q\|^2$

L^2 Assignment Problem

Problem: Given $P, Q \subseteq \mathbb{R}^d$ with $|P| = |Q| = N$,
find one-to-one $\sigma : Q \to P$ minimizing $\sum_{q \in Q} \|q - \sigma(q)\|^2$.

$\iff$ minimising $\Psi(v, w) = -\sum_{q \in Q} v(q) + \sum_{p \in P} w(p)$

under the constraint $v(q) - w(p) \leq \|p - q\|^2$

$\iff$ minimising $\Phi(w) = -\sum_{q \in Q} \min_{p \in P}(\|q - p\|^2 + w(p)) + \sum_{p \in P} w(p)$

L^2 Assignment Problem

Problem: Given $P, Q \subseteq \mathbb{R}^d$ with $|P| = |Q| = N$,
find one-to-one $\sigma : Q \to P$ minimizing $\sum_{q \in Q} \|q - \sigma(q)\|^2$.

$\iff$ minimising $\Psi(v, w) = -\sum_{q \in Q} v(q) + \sum_{p \in P} w(p)$

under the constraint $v(q) - w(p) \leq \|p - q\|^2$

$\iff$ minimising $\Phi(w) = -\sum_{q \in Q} \min_{p \in P}(\|q - p\|^2 + w(p)) + \sum_{p \in P} w(p)$

Dual formulation yields a non-smooth convex function

Φ is smooth at $w \iff \min_{p \in P}(\|q - p\|^2 + w(p))$ is unique for every $q \in Q$

L^2 Assignment Problem: A useful trick

Problem: min. $\Phi(w) = -\sum_{q\in Q} \min_{p\in P}(\|q-p\|^2 + w(p)) + \sum_{p\in P} w(p)$

- Suppose that $w \geq 0$ (Φ is invariant to shifts of w by a constant),

$$\min_{p\in P}(\|q-p\|^2 + w(p)) = \min_{p\in P} \|(q,0) - (p, \sqrt{w(p)})\|_{\mathbb{R}^{d+1}}$$

L^2 Assignment Problem: A useful trick

Problem: min. $\Phi(w) = -\sum_{q\in Q} \min_{p\in P}(\|q-p\|^2 + w(p)) + \sum_{p\in P} w(p)$

- Suppose that $w \geq 0$ (Φ is invariant to shifts of w by a constant),

$$\min_{p\in P}(\|q-p\|^2 + w(p)) = \min_{p\in P} \|(q,0) - (p, \sqrt{w(p)})\|_{\mathbb{R}^{d+1}}$$

- **kD-tree:** A data-structure for finding nearest neighbors.

Build time: $\mathrm{O}(N \log N)$

Query time: $\mathrm{O}(\log N)$ (for uniformly distributed points)

L^2 Assignment Problem: A useful trick

Problem: min. $\Phi(w) = -\sum_{q\in Q} \min_{p\in P}(\|q-p\|^2 + w(p)) + \sum_{p\in P} w(p)$

- Suppose that $w \geq 0$ (Φ is invariant to shifts of w by a constant),

$$\min_{p\in P}(\|q-p\|^2 + w(p)) = \min_{p\in P} \|(q,0) - (p, \sqrt{w(p)})\|_{\mathbb{R}^{d+1}}$$

- **kD-tree:** A data-structure for finding nearest neighbors.

Build time: $\mathrm{O}(N \log N)$

Query time: $\mathrm{O}(\log N)$ (for uniformly distributed points)

$\longrightarrow$ Efficient evaluation of $\Phi(w)$, $\nabla\Phi(w)$ and $\partial\Phi(w)$ (equality cases)

L^2 Assignment Problem: Auction algorithm

Auction is an iterative algorithm of time complexity $\mathrm{O}(N^2 + C \log N)$ where $C = \max \|p_i - q_j\|^2$. All distances need to be integers.

L^2 Assignment Problem: Auction algorithm

Auction is an iterative algorithm of time complexity $\mathrm{O}(N^2 + C \log N)$ where $C = \max \|p_i - q_j\|^2$. All distances need to be integers.

- It is not clear how to correctly "floor" the two point sets $P, Q \subseteq \mathbb{R}^d$ while keeping C small (e.g. concentration).

L^2 Assignment Problem: Auction algorithm

Auction is an iterative algorithm of time complexity $\mathrm{O}(N^2 + C \log N)$ where $C = \max \|p_i - q_j\|^2$. All distances need to be integers.

- It is not clear how to correctly "floor" the two point sets $P, Q \subseteq \mathbb{R}^d$ while keeping C small (e.g. concentration).

- Auction uses only 2 nearest neighbor, i.e. only first order information about Φ.

L^2 Assignment Problem: Auction algorithm

Auction is an iterative algorithm of time complexity $\mathrm{O}(N^2 + C \log N)$ where $C = \max \|p_i - q_j\|^2$. All distances need to be integers.

- It is not clear how to correctly "floor" the two point sets $P, Q \subseteq \mathbb{R}^d$ while keeping C small (e.g. concentration).

- Auction uses only 2 nearest neighbor, i.e. only first order information about Φ.

- The algorithm does not build the dual variable w. Hence, it is not possible to use LP in the final phase.

L^2 Assignment Problem: Auction algorithm

Auction is an iterative algorithm of time complexity $\mathrm{O}(N^2 + C \log N)$ where $C = \max \|p_i - q_j\|^2$. All distances need to be integers.

- It is not clear how to correctly "floor" the two point sets $P, Q \subseteq \mathbb{R}^d$ while keeping C small (e.g. concentration).
- Auction uses only 2 nearest neighbor, i.e. only first order information about Φ.
- The algorithm does not build the dual variable w. Hence, it is not possible to use LP in the final phase.

Many improvements since Bertsekas' original Auction algorithm. We use the fastest to date: Bus-Tvrdik '11

L^2 Assignment Problem: Another approach?

LBFGS is a quasi-Newton algorithm to compute the minimum of smooth functions.

L^2 Assignment Problem: Another approach?

LBFGS is a quasi-Newton algorithm to compute the minimum of smooth functions.

- Many authors have observed the good behaviour of LBFGS methods in minimizing non-smooth functions.

e.g. Lewis-Overton ('10)

L^2 Assignment Problem: Another approach?

LBFGS is a quasi-Newton algorithm to compute the minimum of smooth functions.

- Many authors have observed the good behaviour of LBFGS methods in minimizing non-smooth functions.

e.g. Lewis-Overton ('10)

- LBFGS copes well with the non-smoothness of Φ at the beginning. However, it becomes eventually not possible to find a good descent direction.

L^2 Assignment Problem: Another approach?

LBFGS is a quasi-Newton algorithm to compute the minimum of smooth functions.

- Many authors have observed the good behaviour of LBFGS methods in minimizing non-smooth functions.

  e.g. Lewis-Overton ('10)

- LBFGS copes well with the non-smoothness of Φ at the beginning. However, it becomes eventually not possible to find a good descent direction.

- Proposal: when this happens, turn to a *local linearization* of Ψ and use a LP solver.

L^2 Assignment Problem: Another approach?

Local linearization: we replace the $N \times N$ constraints of the dual program by $k \times N$ constraints + box constraints.

L^2 Assignment Problem: Another approach?

Local linearization: we replace the $N \times N$ constraints of the dual program by $k \times N$ constraints + box constraints.

$$\Psi(v,w) = -\sum_{q \in Q} v(q) + \sum_{p \in P} w(p)$$

under the constraint $v(q) - w(p) \leq \|p-q\|^2$

L^2 Assignment Problem: Another approach?

Local linearization: we replace the $N \times N$ constraints of the dual program by $k \times N$ constraints + box constraints.

$$\Psi(v, w) = -\sum_{q \in Q} v(q) + \sum_{p \in P} w(p)$$

under the constraint $v(q) - w(p) \leq \|p - q\|^2$

Given $w_0 : P \to \mathbb{R}$ and $q \in Q$, define $p_i(q)$ $(1 \leq i \leq N)$ by

$$\|q - p_1(q)\|^2 + w_0(p_1(q)) \leq \ldots \leq \|q - p_N(q)\|^2 + w_0(p_N(q))$$

L^2 Assignment Problem: Another approach?

Local linearization: we replace the $N \times N$ constraints of the dual program by $k \times N$ constraints + box constraints.

$$\Psi(v, w) = -\sum_{q \in Q} v(q) + \sum_{p \in P} w(p)$$

under the constraint $v(q) - w(p) \leq \|p - q\|^2$

Given $w_0 : P \to \mathbb{R}$ and $q \in Q$, define $p_i(q)$ ($1 \leq i \leq N$) by

$$\|q - p_1(q)\|^2 + w_0(p_1(q)) \leq \ldots \leq \|q - p_N(q)\|^2 + w_0(p_N(q))$$

Set $w = w_0 + \delta$. If $\|\delta\| \leq \delta_0$, the nearest neighbor for w remains among the first k ones for w_0, i.e.

L^2 Assignment Problem: Another approach?

Local linearization: we replace the $N \times N$ constraints of the dual program by $k \times N$ constraints + box constraints.

$$\Psi(v,w) = -\sum_{q\in Q} v(q) + \sum_{p \in P} w(p)$$

under the constraint $v(q) - w(p) \leq \|p-q\|^2$

Given $w_0 : P \to \mathbb{R}$ and $q \in Q$, define $p_i(q)$ $(1 \leq i \leq N)$ by

$$\|q - p_1(q)\|^2 + w_0(p_1(q)) \leq \ldots \leq \|q - p_N(q)\|^2 + w_0(p_N(q))$$

Set $w = w_0 + \delta$. If $\|\delta\| \leq \delta_0$, the nearest neighbor for w remains among the first k ones for w_0, i.e.

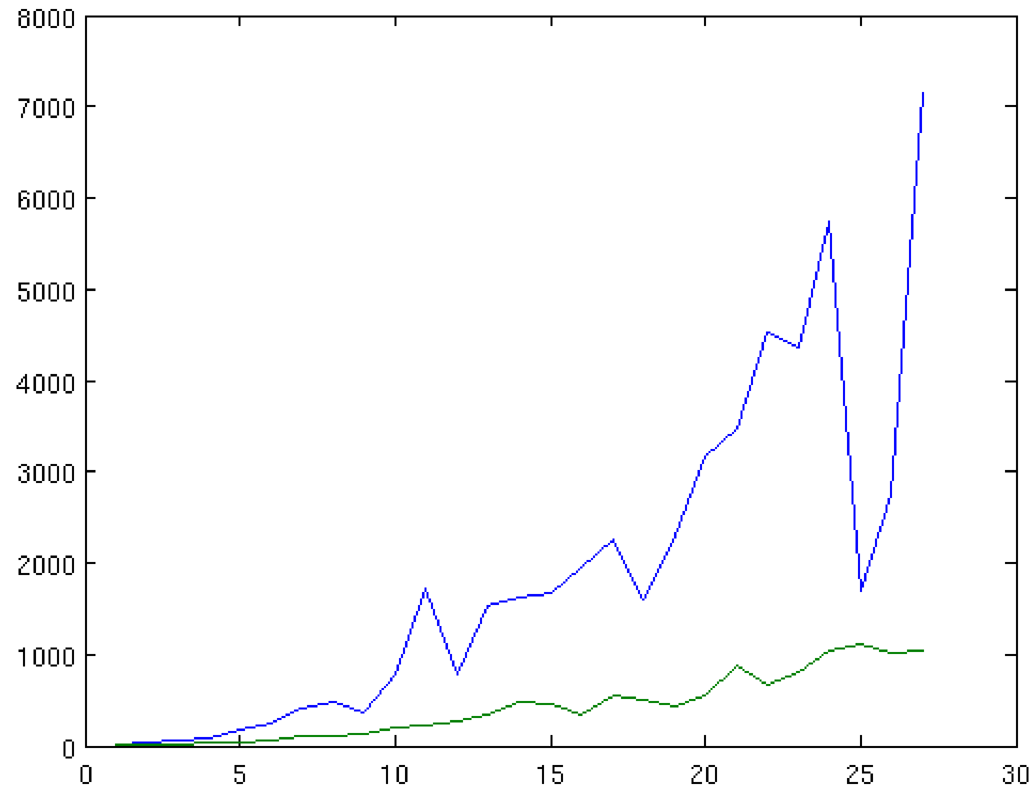
$$\forall p, q, \quad v(q) - w(p) \leq \|p-q\|^2$$

$$\Longleftrightarrow$$

$$\forall q, \forall 1 \leq i \leq k, \quad v(q) - w(p_i(q)) \leq \|p_i(q) - q\|^2 \quad \textbf{and} \quad \|\delta\| \leq \delta_0$$

L^2 Assignment Problem: Another approach?

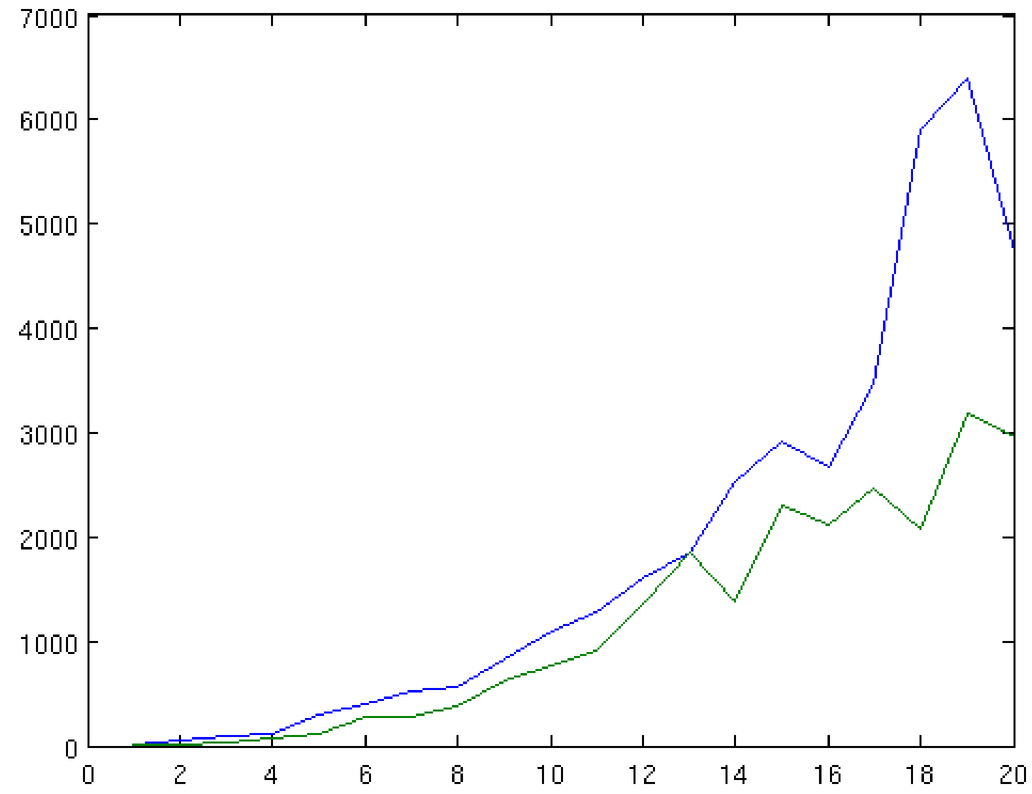
Running time in seconds of **Auction** (blue) vs **LBFGS** and linearisation (green)



Data: P and Q are two random sample of N points in the cube $[0, 10^5]^3 \cap \mathbb{Z}^3$, for $N = 1k, \ldots, 30k$.

L^2 Assignment Problem: Another approach?

Running time in seconds of **Auction** (blue) vs **LBFGS** and linearisation (green)



Data: P is a random sample of N points in the cube $[0, 10^5]^3 \cap \mathbb{Z}^3$, Q is obtained from a mixture of 15 isotropic Gaussian distributions for $N = 1k, \ldots, 20k$.

Conclusion

Open questions:

- Dependency of convergence speed on the "geometry" of μ and ν, i.e. quantitative stability theorem for OT plans ?
- Theory when both measures are discrete ? Complexity ?

Conclusion

Open questions:

- Dependency of convergence speed on the "geometry" of μ and ν, i.e. quantitative stability theorem for OT plans ?
- Theory when both measures are discrete ? Complexity ?

Other applications of the multiscale approach:

- Minkowski's problem: designing a convex polyhedron with given facets normals and areas.
- Design of reflector antennas with prescribed far-field image.

Conclusion

Open questions:

- Dependency of convergence speed on the "geometry" of μ and ν, i.e. quantitative stability theorem for OT plans ?
- Theory when both measures are discrete ? Complexity ?

Other applications of the multiscale approach:

- Minkowski's problem: designing a convex polyhedron with given facets normals and areas.
- Design of reflector antennas with prescribed far-field image.

C++ code: http://quentin.mrgt.fr/

Conclusion

Open questions:

- Dependency of convergence speed on the "geometry" of μ and ν, i.e. quantitative stability theorem for OT plans ?
- Theory when both measures are discrete ? Complexity ?

Other applications of the multiscale approach:

- Minkowski's problem: designing a convex polyhedron with given facets normals and areas.
- Design of reflector antennas with prescribed far-field image.

C++ code: http://quentin.mrgt.fr/

Thank you for your attention!