Damped Newton Algorithm for Semi-discrete optimal transport

Boris Thibert

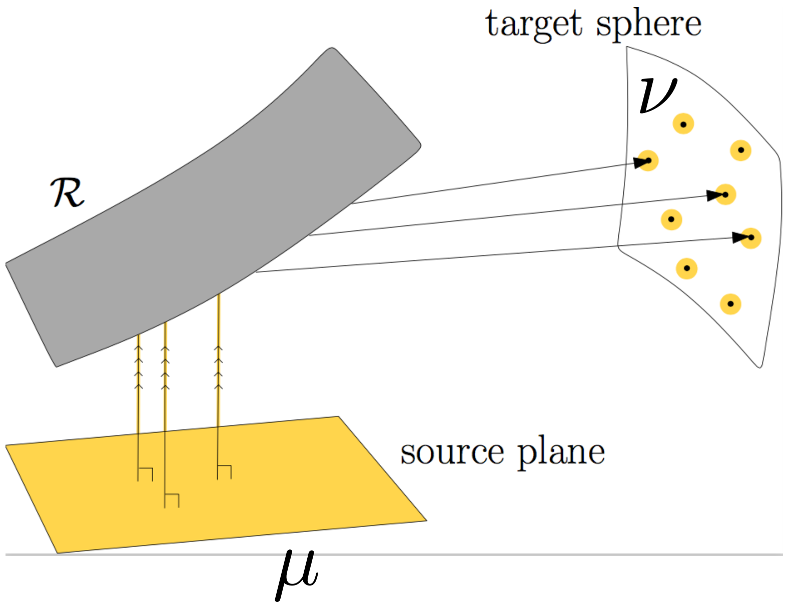
with Jun Kitagawa, Jocelyn Meyron and Quentin Mérigot

Banff – April 10-14, 2017

Motivations with optimal transport

Inverse problems in optics

$\rightsquigarrow$ reflector surfaces $\mathcal{R}$



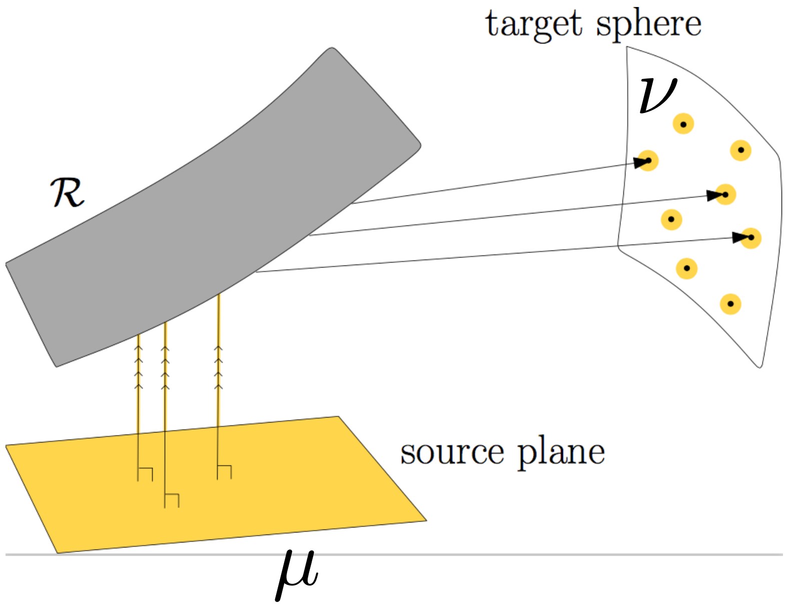
Collimated source / Far-field target

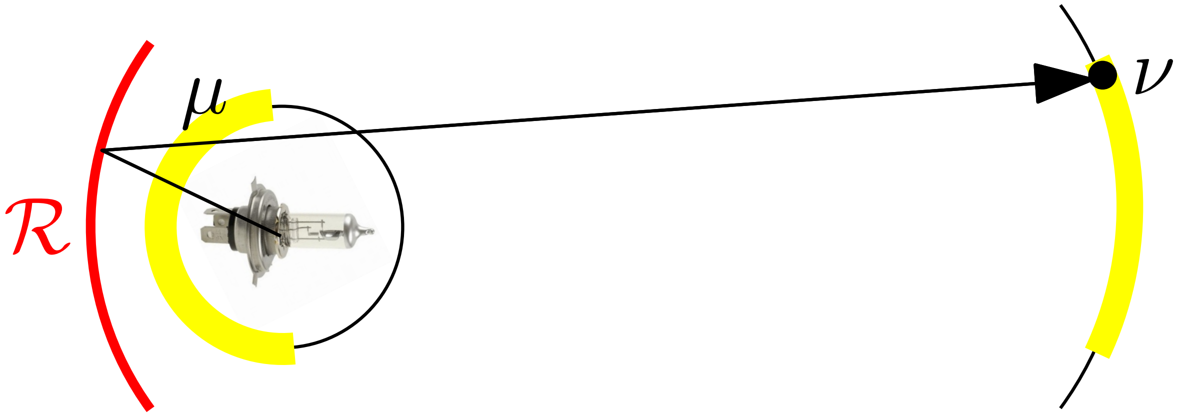
OT in $\mathbb{R}^2$, $c(x,y) = \|x-y\|^2$

Motivations with optimal transport

Inverse problems in optics

$\rightsquigarrow$ reflector surfaces $\mathcal{R}$





Collimated source / Far-field target

OT in $\mathbb{R}^2$, $c(x,y) = \|x-y\|^2$

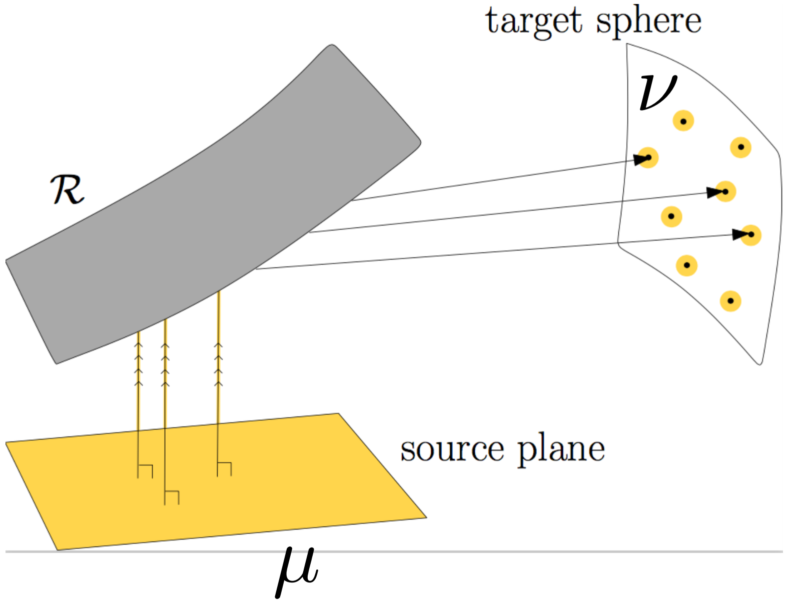
Punctual source / Far-field target

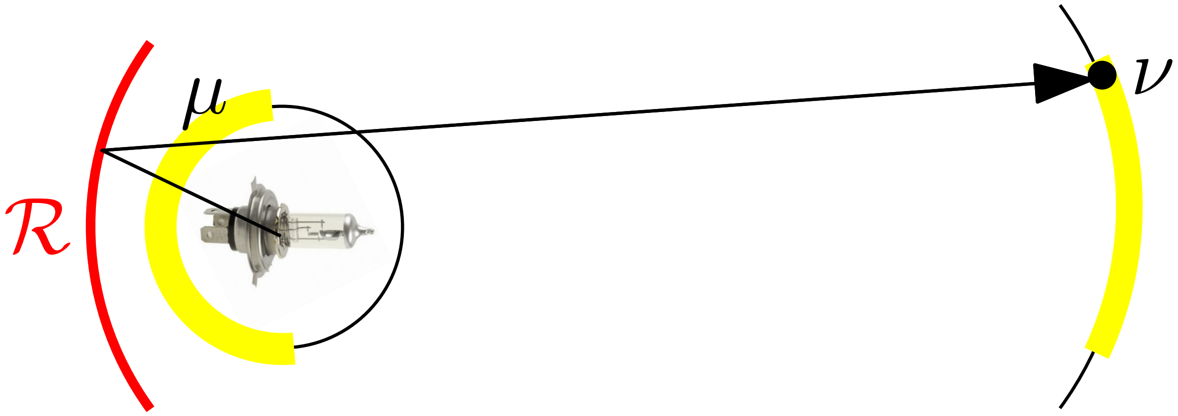
OT in $\mathcal{S}^2$, $c(x,y) = -\ln(1-\langle x|y\rangle)$

Motivations with optimal transport

Inverse problems in optics

$\rightsquigarrow$ reflector surfaces $\mathcal{R}$





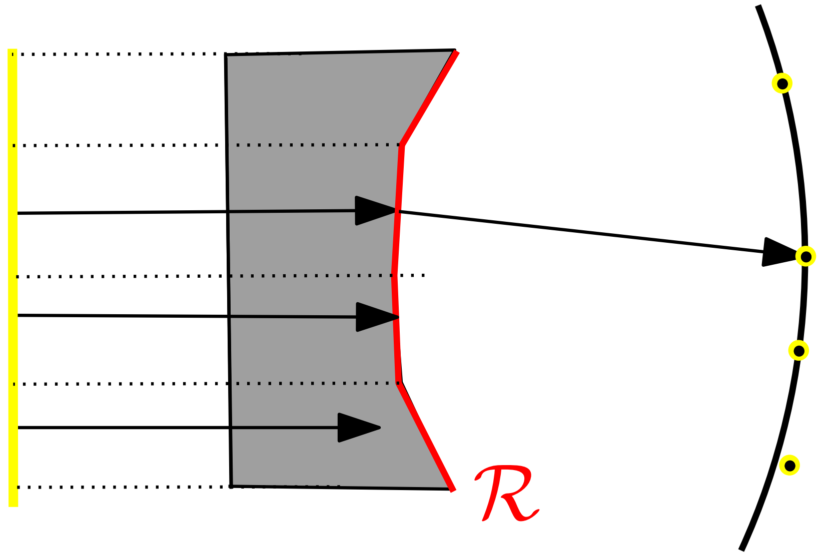
Collimated source / Far-field target

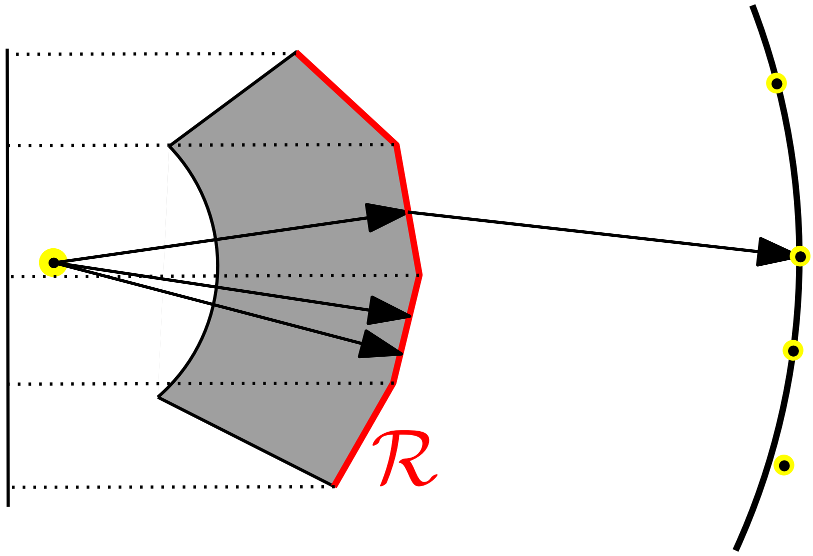
OT in $\mathbb{R}^2$, $c(x,y) = \|x-y\|^2$

Punctual source / Far-field target

OT in $\mathcal{S}^2$, $c(x,y) = -\ln(1-\langle x|y\rangle)$

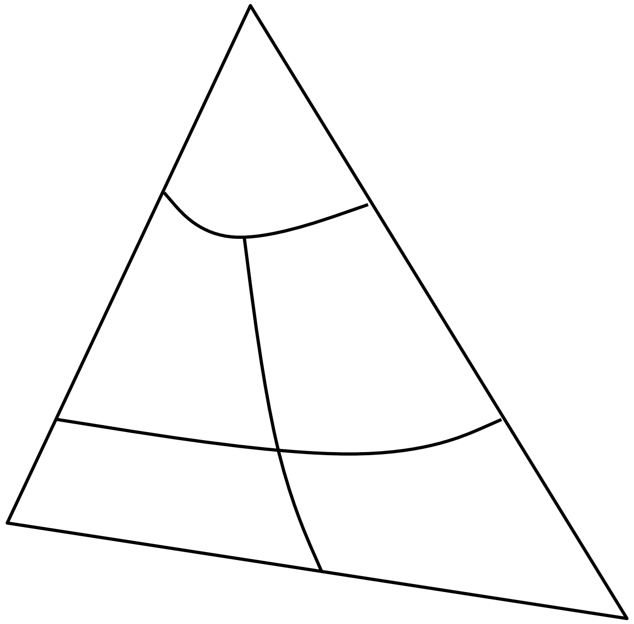
$\rightsquigarrow$ refractor surfaces $\mathcal{R}$



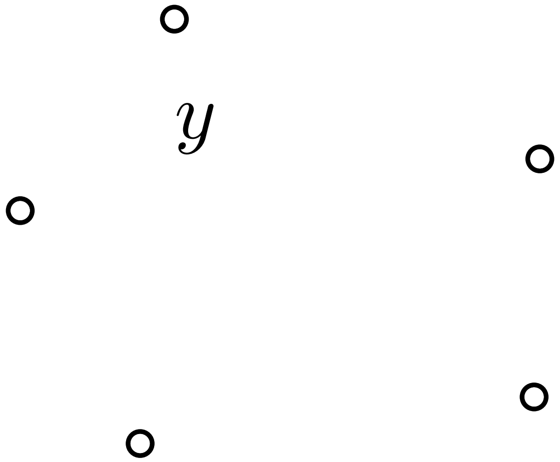


Semi-discrete optimal transport

$\mu =$ probability measure on X
with density ρ, $X =$ manifold



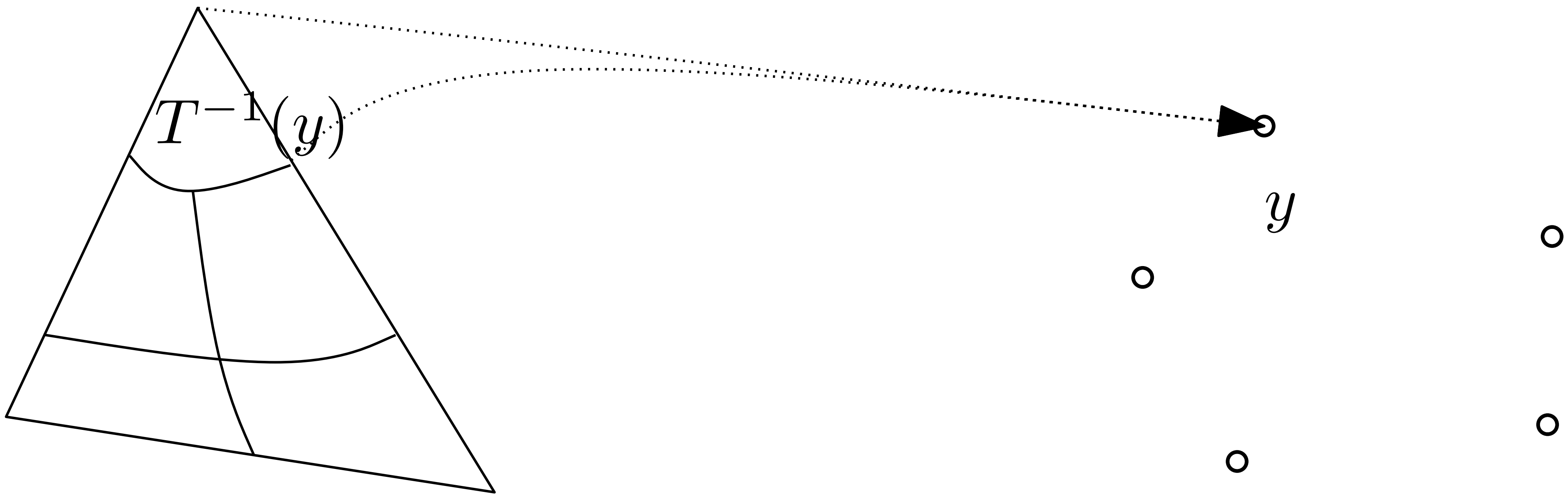
$\nu =$ prob. measure on finite Y
$= \sum_{y \in Y} \nu_y \delta_y$



Semi-discrete optimal transport

$\mu =$ probability measure on X
with density ρ, $X =$ manifold

$\nu =$ prob. measure on finite Y
$= \sum_{y \in Y} \nu_y \delta_y$



Transport map: $T : X \to Y$ s.t.

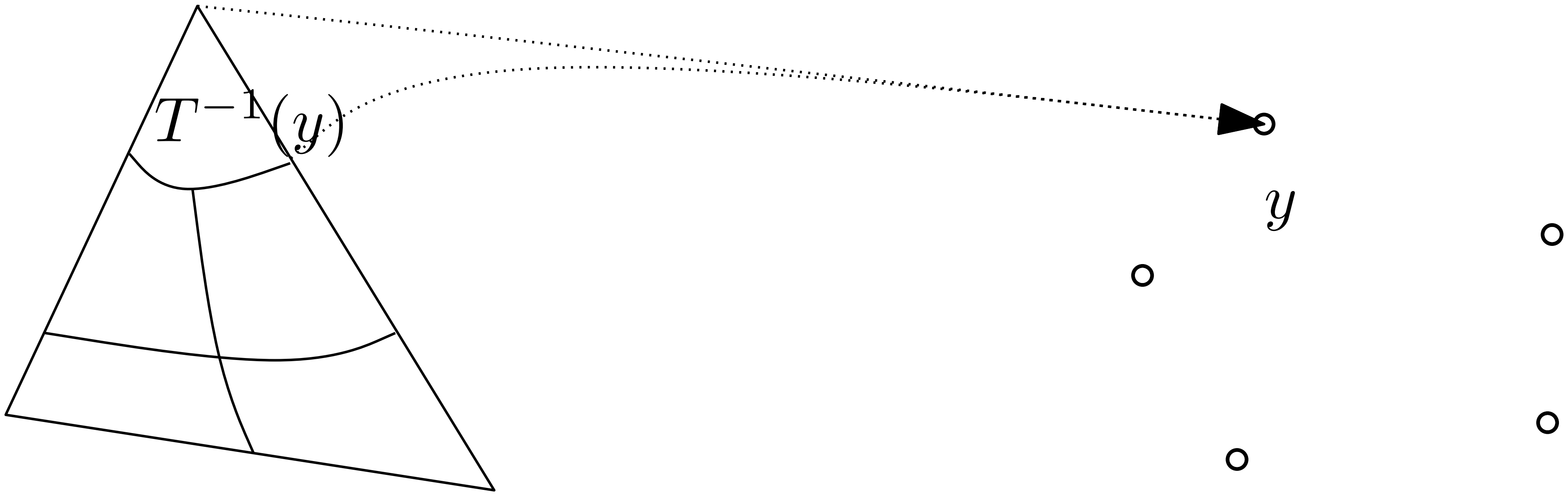
$$\forall y \in Y, \ \mu(T^{-1}(\{y\})) = \nu_y$$

(i.e. $T_{\#}\mu = \nu$)

Semi-discrete optimal transport

$\mu =$ probability measure on X
with density ρ, $X =$ manifold

$\nu =$ prob. measure on finite Y
$= \sum_{y \in Y} \nu_y \delta_y$



Transport map: $T : X \to Y$ s.t.

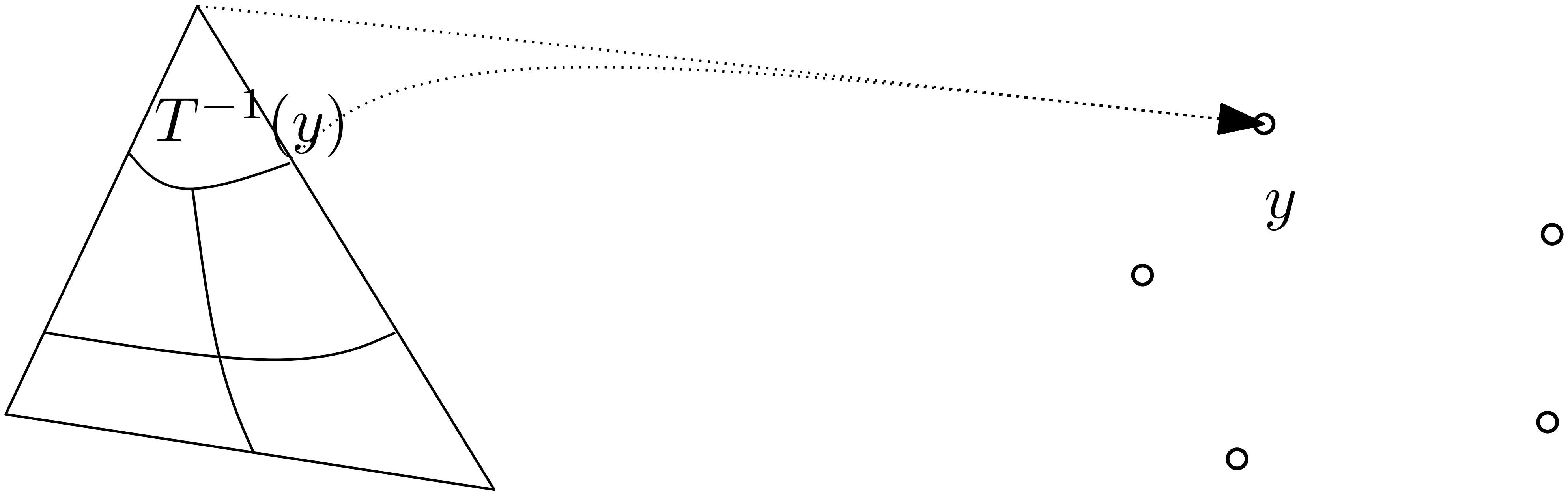
$\forall y \in Y,\ \mu(T^{-1}(\{y\})) = \nu_y$
(i.e. $T_\# \mu = \nu$)

Cost function: $c : X \times Y \to \mathbb{R}$

Semi-discrete optimal transport

$\mu =$ probability measure on X
with density ρ, $X =$ manifold

$\nu =$ prob. measure on finite Y
$= \sum_{y\in Y} \nu_y \delta_y$



Transport map: $T : X \to Y$ s.t.

$$\forall y \in Y,\ \mu(T^{-1}(\{y\})) = \nu_y$$

(i.e. $T_{\#}\mu = \nu$)

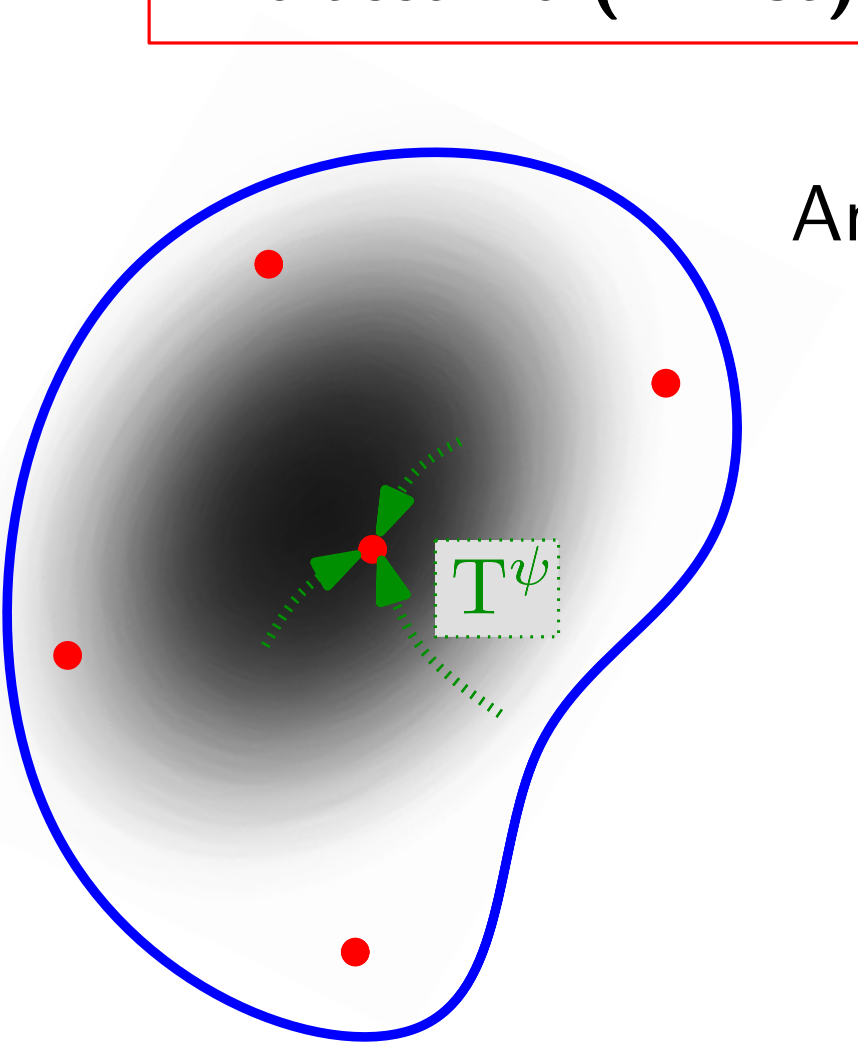
Cost function: $c : X \times Y \to \mathbb{R}$

Monge problem:

$$\min\{\int_X c(x, T(x))\, \mathrm{d}\, \mu(x); T_{\#}\mu = \nu\}$$

Semi-discrete optimal transport

We assume **(Twist):** $\forall x \in X$, the map $y \in Y \mapsto \nabla_x c(x,y)$ is injective.

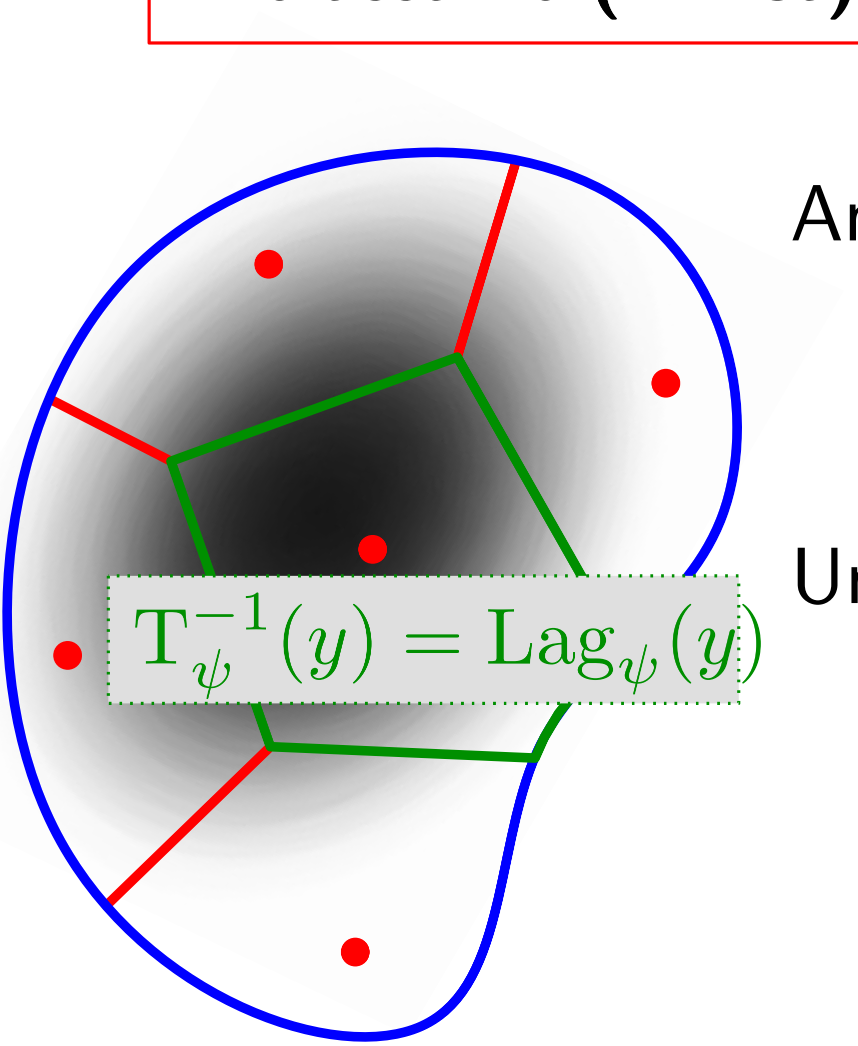


Any function ψ on Y defines a transport map:

$$\boxed{\mathrm{T}_{\psi}(x) = \arg\min_{y\in Y} c(x,y) + \psi(y)}$$

Semi-discrete optimal transport

We assume **(Twist):** $\forall x \in X$, the map $y \in Y \mapsto \nabla_x c(x,y)$ is injective.



Any function ψ on Y defines a transport map:

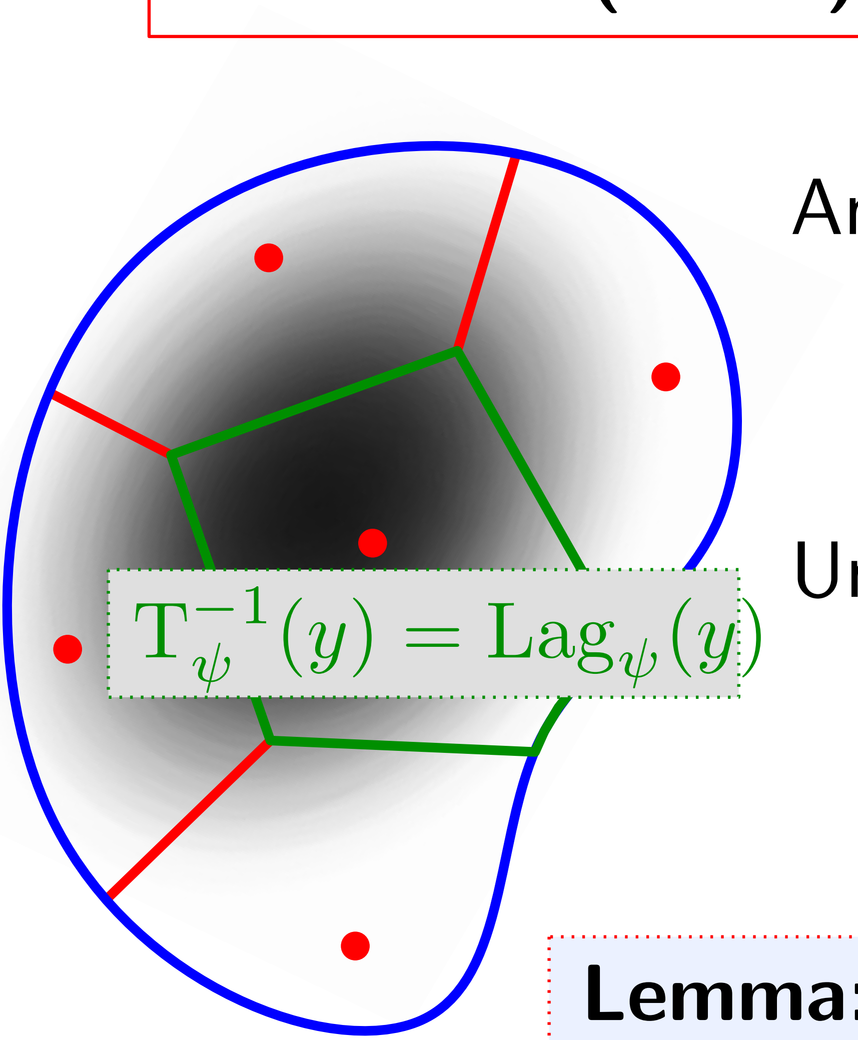
$$\boxed{\mathrm{T}_\psi(x) = \arg\min_{y \in Y} c(x,y) + \psi(y)}$$

Under **(Twist)**, T_ψ is well-defined a.e. and

$$\boxed{\mathrm{T}_\psi^{-1}(y) = \mathrm{Lag}_\psi(y) \; T_{\psi\#}\rho = \sum_y \rho(\mathrm{Lag}_\psi(y))\delta_y}$$

Semi-discrete optimal transport

We assume **(Twist):** $\forall x \in X$, the map $y \in Y \mapsto \nabla_x c(x,y)$ is injective.



Any function ψ on Y defines a transport map:

$$\boxed{\mathrm{T}_\psi(x) = \arg\min_{y \in Y} c(x,y) + \psi(y)}$$

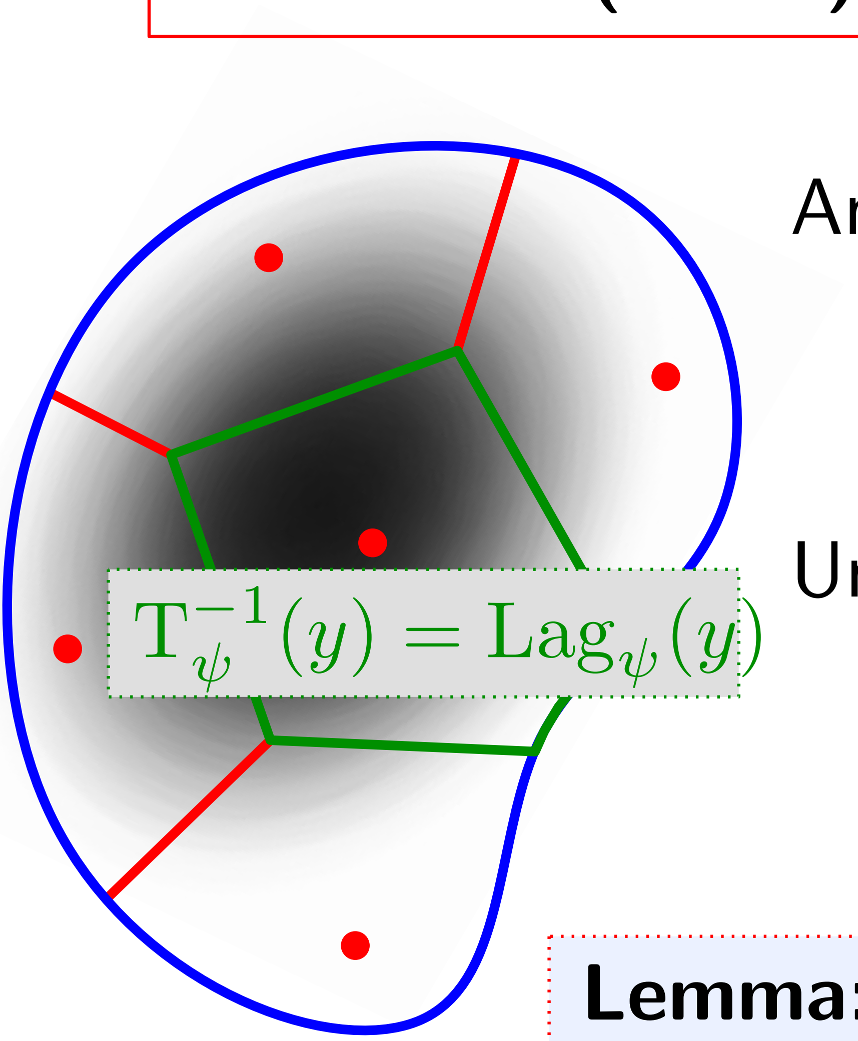
Under **(Twist)**, T_ψ is well-defined a.e. and

$$\boxed{\mathrm{T}_\psi^{-1}(y) = \mathrm{Lag}_\psi(y) \; T_{\psi\#}\rho = \sum_y \rho(\mathrm{Lag}_\psi(y))\delta_y}$$

Lemma: T_ψ is an optimal transport map between ρ and $T_{\psi\#}\rho$.

Semi-discrete optimal transport

We assume **(Twist):** $\forall x \in X$, the map $y \in Y \mapsto \nabla_x c(x,y)$ is injective.



$$\mathrm{T}_\psi^{-1}(y) = \mathrm{Lag}_\psi(y)$$

Any function ψ on Y defines a transport map:

$$\boxed{\mathrm{T}_\psi(x) = \arg\min_{y \in Y} c(x,y) + \psi(y)}$$

Under **(Twist)**, T_ψ is well-defined a.e. and

$$\boxed{\mathrm{T}_\psi^{-1}(y) = \mathrm{Lag}_\psi(y) \; T_{\psi\#}\rho = \sum_y \rho(\mathrm{Lag}_\psi(y))\delta_y}$$

Lemma: T_ψ is an optimal transport map between ρ and $T_{\psi\#}\rho$.

Optimal transport problem:

Find $\psi = (\psi_y)_y$ such that $T_{\psi\#}\rho = \nu$,

source density measure

target discrete constraint

A damped Newton algorithm

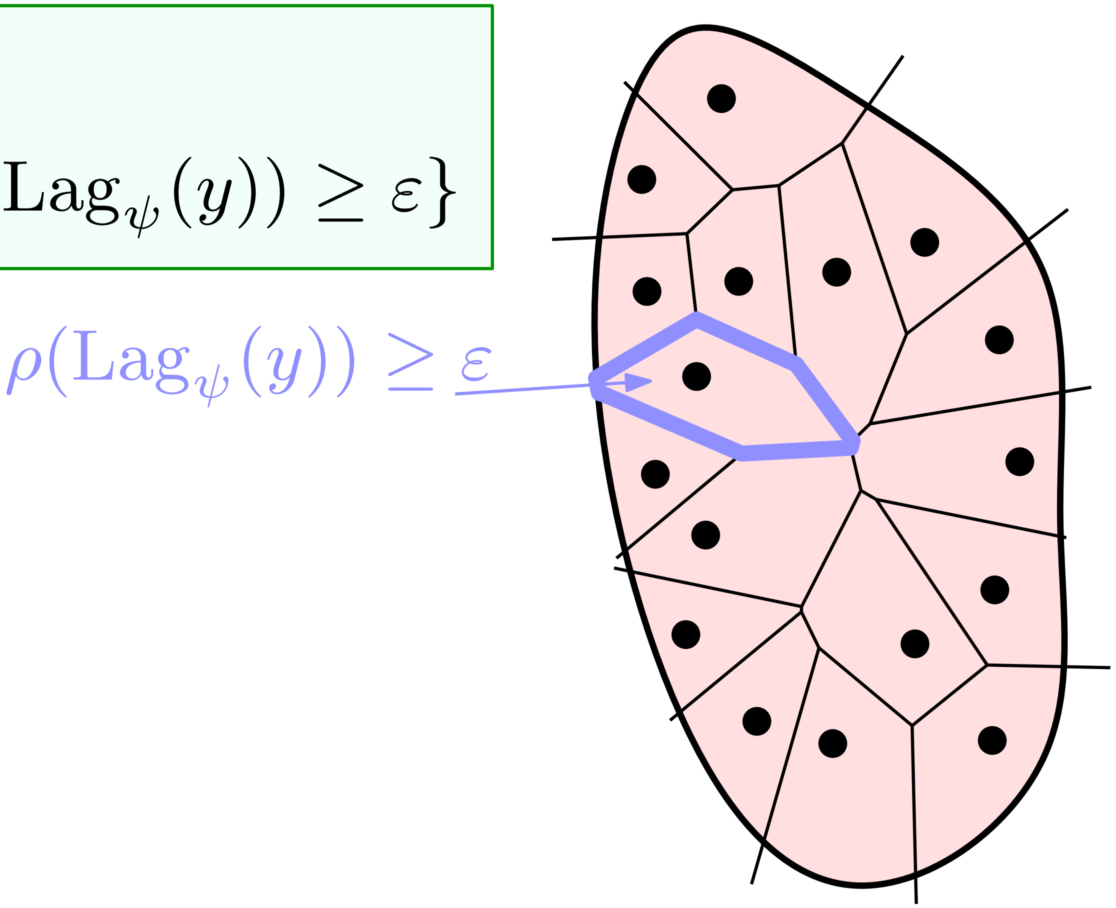
with Jun Kitagawa and Quentin Mérigot

Damped Newton Algorithm cf [Mirebeau '15]

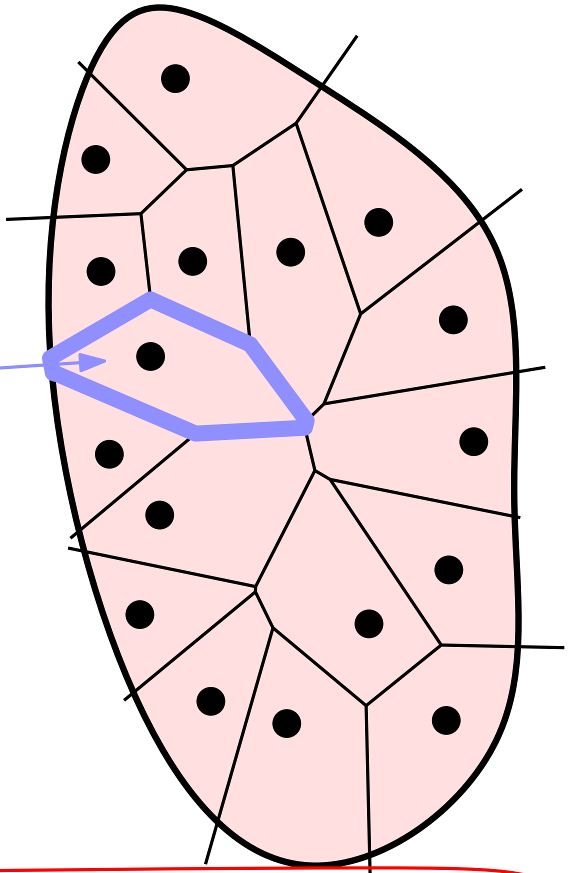
Equation $(\rho(\mathrm{Lag}_y(\psi)) - \nu_y) = 0$ for all y

Admissible domain: $E_\varepsilon := \{\psi \in Y^{\mathbb{R}}; \forall y \in Y, \rho(\mathrm{Lag}_\psi(y)) \geq \varepsilon\}$

We put $G_y(\psi) = \rho(\mathrm{Lag}_y(\psi))$



Damped Newton Algorithm cf [Mirebeau '15]



Equation $(\rho(\mathrm{Lag}_y(\psi)) - \nu_y) = 0$ for all y

Admissible domain: $E_\varepsilon := \{\psi \in Y^{\mathbb{R}}; \forall y \in Y, \rho(\mathrm{Lag}_\psi(y)) \geq \varepsilon\}$

We put $G_y(\psi) = \rho(\mathrm{Lag}_y(\psi))$

Damped Newton algorithm: for solving $G(\psi) = \nu$

Input: $\psi_0 \in Y^{\mathbb{R}}$ s.t. $\varepsilon := \frac{1}{2} \min_{y \in Y} \min(G(\psi_0)_y, \nu_y) > 0$

Loop: $\longrightarrow$ Define $\psi_k^\tau = \psi_k - \tau \mathrm{D}G(\psi_k)^{-1}(G(\psi_k) - \nu)$

$\longrightarrow$ $\tau_k := \max\{\tau \in 2^{-\mathbb{N}} \mid \psi_k^\tau \in E_\varepsilon$ and $\|G(\psi_k^\tau) - \nu\| \leq (1 - \frac{\tau}{2})\|G(\psi_k) - \nu\|\}$

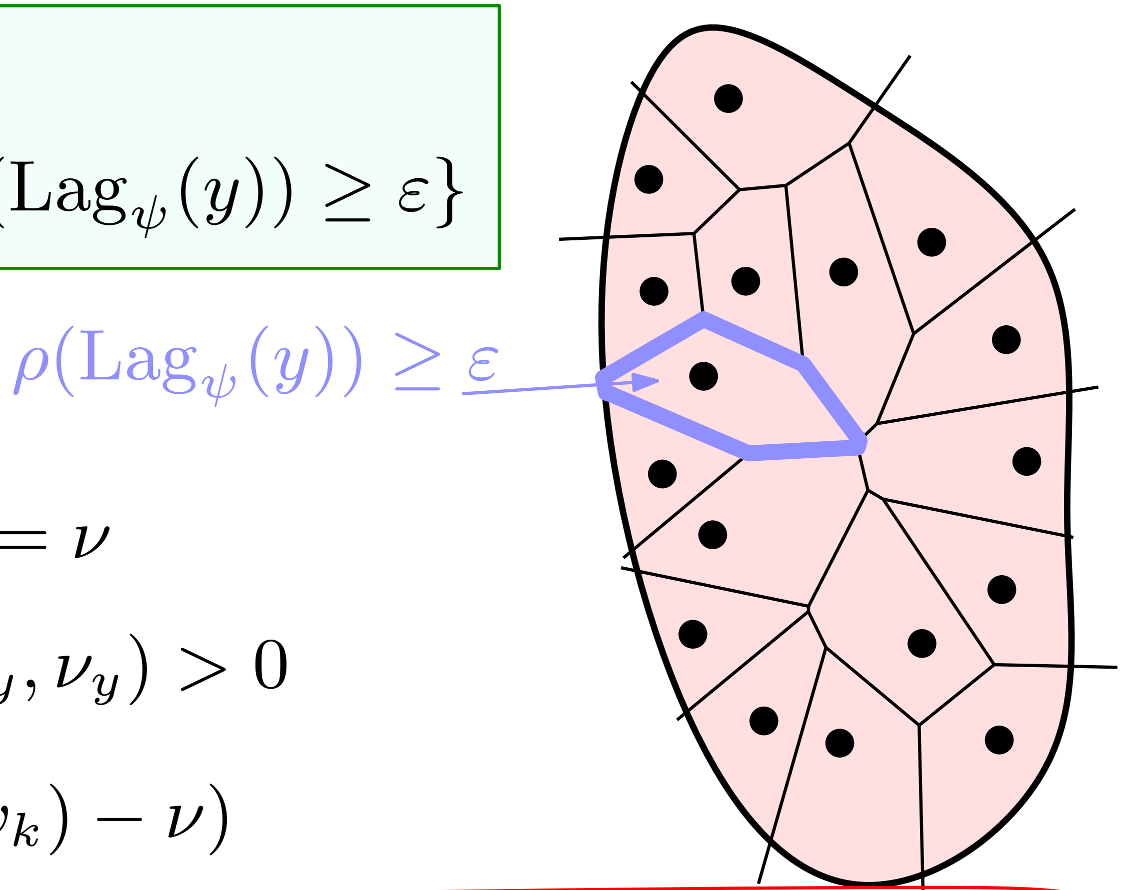
$\longrightarrow$ $\psi_{k+1} := \psi_k^{\tau_k}$

Damped Newton Algorithm cf [Mirebeau '15]

Equation $(\rho(\mathrm{Lag}_y(\psi)) - \nu_y) = 0$ for all y

Admissible domain: $E_\varepsilon := \{\psi \in Y^{\mathbb{R}}; \forall y \in Y, \rho(\mathrm{Lag}_\psi(y)) \geq \varepsilon\}$

We put $G_y(\psi) = \rho(\mathrm{Lag}_y(\psi))$



Damped Newton algorithm: for solving $G(\psi) = \nu$

Input: $\psi_0 \in Y^{\mathbb{R}}$ s.t. $\varepsilon := \frac{1}{2}\min_{y\in Y}\min(G(\psi_0)_y, \nu_y) > 0$

Loop: $\longrightarrow$ Define $\psi_k^\tau = \psi_k - \tau \mathrm{D}G(\psi_k)^{-1}(G(\psi_k) - \nu)$

$\longrightarrow \tau_k := \max\{\tau \in 2^{-\mathbb{N}} \mid \psi_k^\tau \in E_\varepsilon$ and $\|G(\psi_k^\tau) - \nu\| \leq (1 - \frac{\tau}{2})\|G(\psi_k) - \nu\|\}$

$\longrightarrow \psi_{k+1} := \psi_k^{\tau_k}$

Remark: The damped Newton's algorithm converges **globally** provided that:

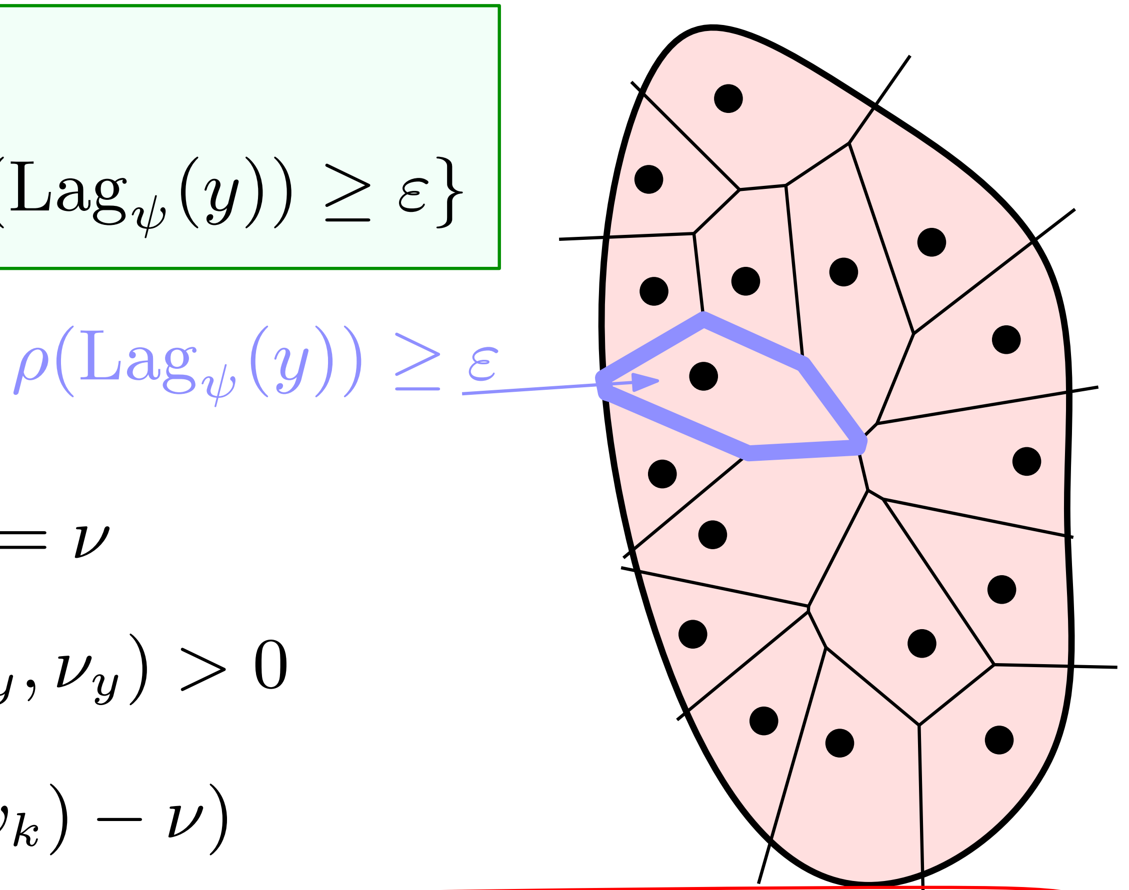
(Smoothness): G is $\mathcal{C}^1$ on E_ε.

Damped Newton Algorithm cf [Mirebeau '15]

Equation $(\rho(\mathrm{Lag}_y(\psi)) - \nu_y) = 0$ for all y

Admissible domain: $E_\varepsilon := \{\psi \in Y^{\mathbb{R}}; \forall y \in Y, \rho(\mathrm{Lag}_\psi(y)) \geq \varepsilon\}$

We put $G_y(\psi) = \rho(\mathrm{Lag}_y(\psi))$



Damped Newton algorithm: for solving $G(\psi) = \nu$

Input: $\psi_0 \in Y^{\mathbb{R}}$ s.t. $\varepsilon := \frac{1}{2} \min_{y \in Y} \min(G(\psi_0)_y, \nu_y) > 0$

Loop: $\longrightarrow$ Define $\psi_k^\tau = \psi_k - \tau \mathrm{D}G(\psi_k)^{-1}(G(\psi_k) - \nu)$

$\longrightarrow$ $\tau_k := \max\{\tau \in 2^{-\mathbb{N}} \mid \psi_k^\tau \in E_\varepsilon$ and $\|G(\psi_k^\tau) - \nu\| \leq (1 - \frac{\tau}{2})\|G(\psi_k) - \nu\|\}$

$\longrightarrow$ $\psi_{k+1} := \psi_k^{\tau_k}$

Remark: The damped Newton's algorithm converges **globally** provided that:

(Smoothness): G is $\mathcal{C}^1$ on E_ε.

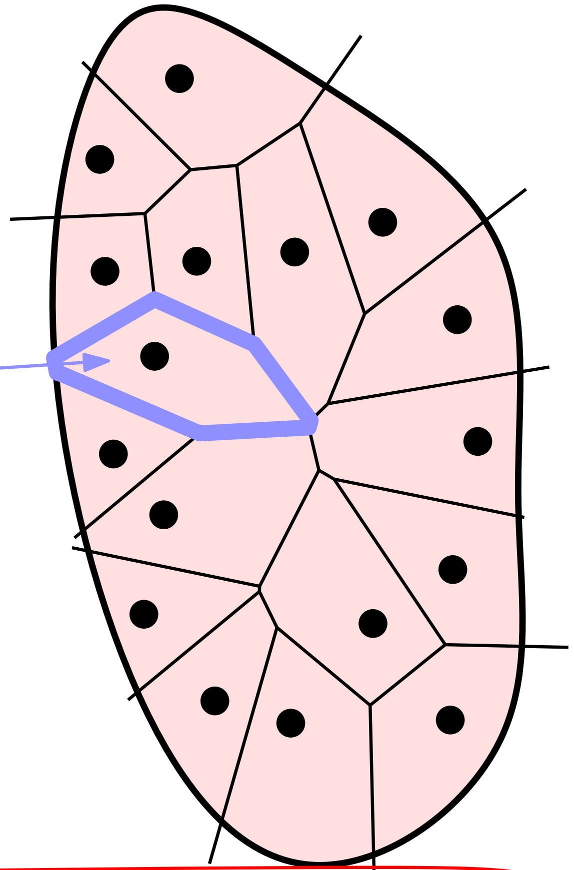
(Strict monotonicity): $\forall \psi \in E_\varepsilon$, $DG(\psi)$ is neg. definite on $E_\varepsilon \cap \{cst\}^\perp$

Damped Newton Algorithm$_{\text{cf [Mirebeau '15]}}$

Equation $(\rho(\mathrm{Lag}_y(\psi)) - \nu_y) = 0$ for all y

Admissible domain: $E_\varepsilon := \{\psi \in Y^{\mathbb{R}} ; \forall y \in Y, \rho(\mathrm{Lag}_\psi(y)) \geq \varepsilon\}$

We put $G_y(\psi) = \rho(\mathrm{Lag}_y(\psi))$



$\rho(\mathrm{Lag}_\psi(y)) \geq \varepsilon$

Damped Newton algorithm: for solving $G(\psi) = \nu$

Input: $\psi_0 \in Y^{\mathbb{R}}$ s.t. $\varepsilon := \frac{1}{2} \min_{y \in Y} \min(G(\psi_0)_y, \nu_y) > 0$

Loop: $\longrightarrow$ Define $\psi_k^\tau = \psi_k - \tau \mathrm{D}G(\psi_k)^{-1}(G(\psi_k) - \nu)$

$\longrightarrow \tau_k := \max\{\tau \in 2^{-\mathbb{N}} \mid \boxed{\psi_k^\tau \in E_\varepsilon}$ and $\boxed{\|G(\psi_k^\tau) - \nu\| \leq (1 - \frac{\tau}{2})\|G(\psi_k) - \nu\|}\}$

$\longrightarrow \psi_{k+1} := \psi_k^{\tau_k}$

Remark: The damped Newton's algorithm converges **globally** provided that:

(Smoothness): G is $\mathcal{C}^1$ on E_ε.

(Strict monotonicity): $\forall \psi \in E_\varepsilon$, $DG(\psi)$ is neg. definite on $E_\varepsilon \cap \{cst\}^\perp$

$\Rightarrow$ We have to show smoothness and strict monotonicity

Goal: prove the CV of the algorithm

- Remarks in the quadratic case, with a measure with density
- CV for cost satisfying MTW
- CV for measure supported on sets with codimension ≥ 1 (and quadratic cost)

Quadratic cost : smoothness of $\mathcal{K}$

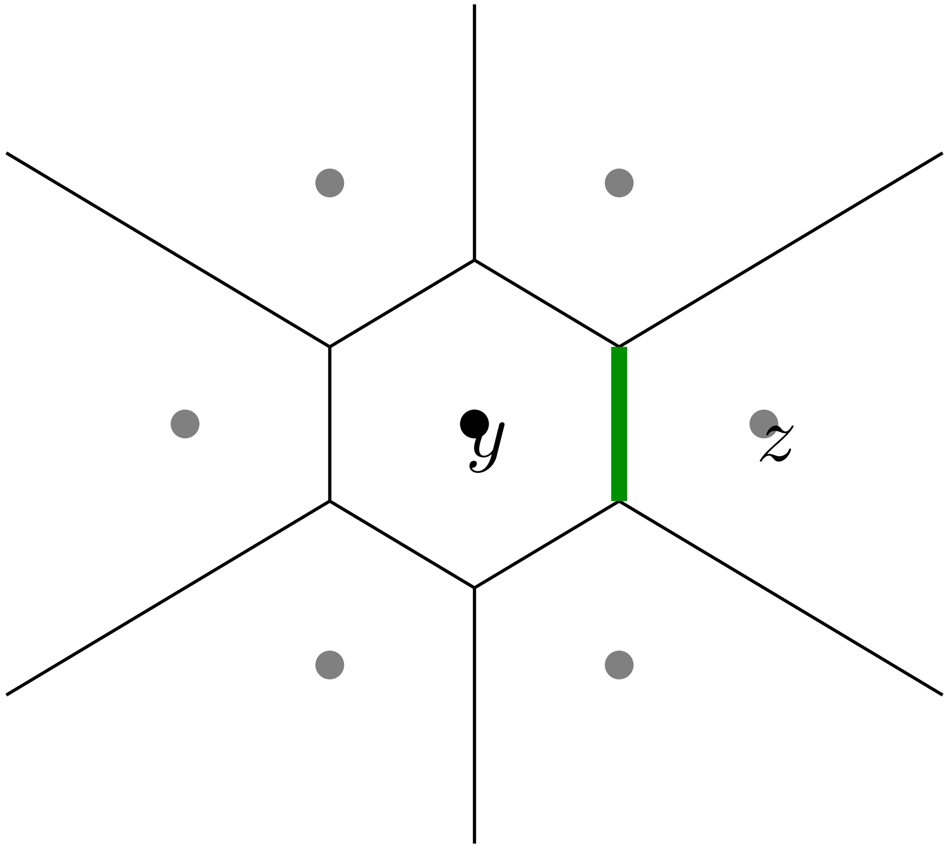
we have $G_y(\psi) = \rho(\mathrm{Lag}_\psi(y))$ $\quad c(x,y) := \|x-y\|^2$

Proposition: For $\psi \in E_\varepsilon$, and assuming that $\rho \in \mathcal{C}^0_c(\mathbb{R}^d)$ one has

(A) $\frac{\partial G_y}{\partial z}(\psi) = \frac{1}{2\|y-z\|} \int_{\boxed{\mathrm{Lag}_{yz}(\psi)}} \rho(x)\,\mathrm{d}\,x$ (B) $\frac{\partial G_y}{\partial y}(\psi) = -\sum_{z\neq y} \frac{\partial G_y}{\partial z}(\psi)$

$z \neq y$

$\mathrm{Lag}_{yz}(\psi) := \mathrm{Lag}_y(\psi) \cap \mathrm{Lag}_z(\psi)$



Let $\psi_t := \psi + t\mathbf{1}_z$.

Quadratic cost : smoothness of $\mathcal{K}$

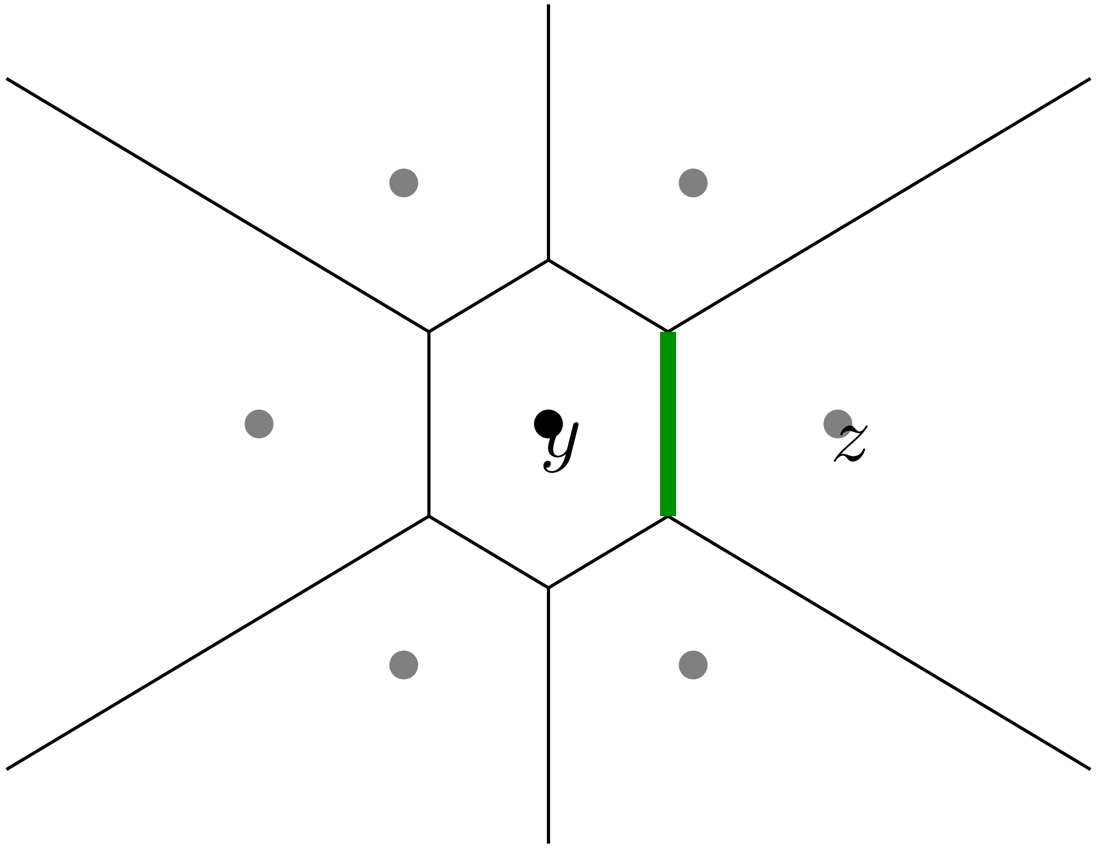
we have $G_y(\psi) = \rho(\mathrm{Lag}_\psi(y))$ $\quad c(x,y) := \|x-y\|^2$

Proposition: For $\psi \in E_\varepsilon$, and assuming that $\rho \in \mathcal{C}^0_c(\mathbb{R}^d)$ one has

(A) $\frac{\partial G_y}{\partial z}(\psi) = \frac{1}{2\|y-z\|}\int_{\mathrm{Lag}_{yz}(\psi)} \rho(x)\,\mathrm{d}\,x$ (B) $\quad \frac{\partial G_y}{\partial y}(\psi) = -\sum_{z\neq y} \frac{\partial G_y}{\partial z}(\psi)$

$z \neq y$

$\mathrm{Lag}_{yz}(\psi) := \mathrm{Lag}_y(\psi) \cap \mathrm{Lag}_z(\psi)$



Let $\psi_t := \psi + t\mathbf{1}_z$.When t varies, $\frac{\partial G_y}{\partial z}(\psi_t)$ increases ...

Quadratic cost : smoothness of $\mathcal{K}$

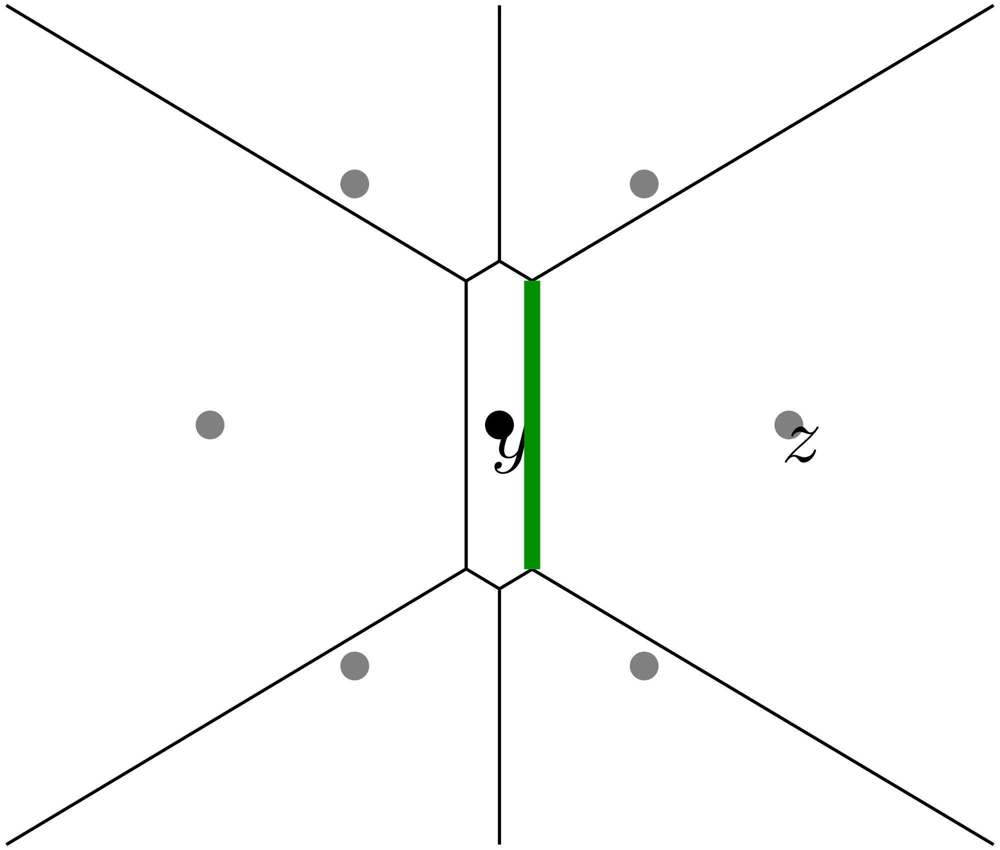
we have $G_y(\psi) = \rho(\mathrm{Lag}_\psi(y))$ $c(x,y) := \|x-y\|^2$

Proposition: For $\psi \in E_\varepsilon$, and assuming that $\rho \in \mathcal{C}^0_c(\mathbb{R}^d)$ one has

(A) $\frac{\partial G_y}{\partial z}(\psi) = \frac{1}{2\|y-z\|} \int_{\boxed{\mathrm{Lag}_{yz}(\psi)}} \rho(x)\,\mathrm{d}\,x$ (B) $\frac{\partial G_y}{\partial y}(\psi) = -\sum_{z\neq y} \frac{\partial G_y}{\partial z}(\psi)$

$z \neq y$

$\mathrm{Lag}_{yz}(\psi) := \mathrm{Lag}_y(\psi) \cap \mathrm{Lag}_z(\psi)$



Let $\psi_t := \psi + t\mathbf{1}_z$.When t varies, $\frac{\partial G_y}{\partial z}(\psi_t)$ increases ...

Quadratic cost : smoothness of $\mathcal{K}$

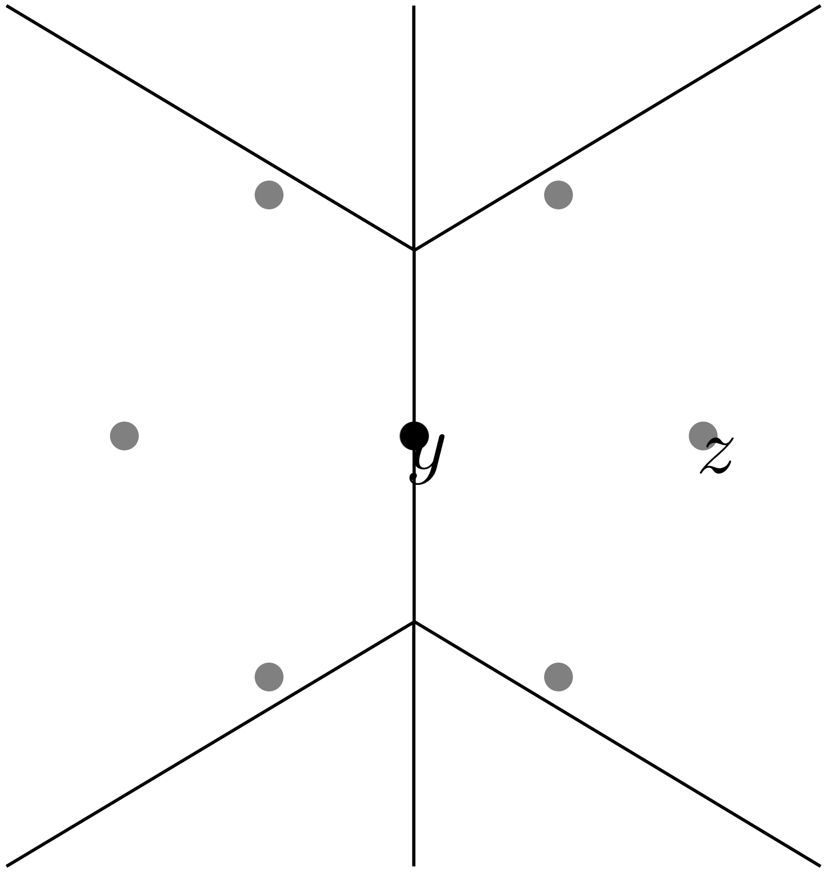
we have $G_y(\psi) = \rho(\mathrm{Lag}_\psi(y))$ $\quad c(x,y) := \|x-y\|^2$

Proposition: For $\psi \in E_\varepsilon$, and assuming that $\rho \in \mathcal{C}^0_c(\mathbb{R}^d)$ one has

(A) $\frac{\partial G_y}{\partial z}(\psi) = \frac{1}{2\|y-z\|} \boxed{\int_{\mathrm{Lag}_{yz}(\psi)}} \rho(x)\,\mathrm{d}\,x$ (B) $\quad \frac{\partial G_y}{\partial y}(\psi) = -\sum_{z\neq y} \frac{\partial G_y}{\partial z}(\psi)$

$z \neq y$

$\mathrm{Lag}_{yz}(\psi) := \mathrm{Lag}_y(\psi) \cap \mathrm{Lag}_z(\psi)$



Let $\psi_t := \psi + t\mathbf{1}_z$.When t varies, $\frac{\partial G_y}{\partial z}(\psi_t)$ increases ...
and then suddenly vanishes.

Quadratic cost : smoothness of $\mathcal{K}$

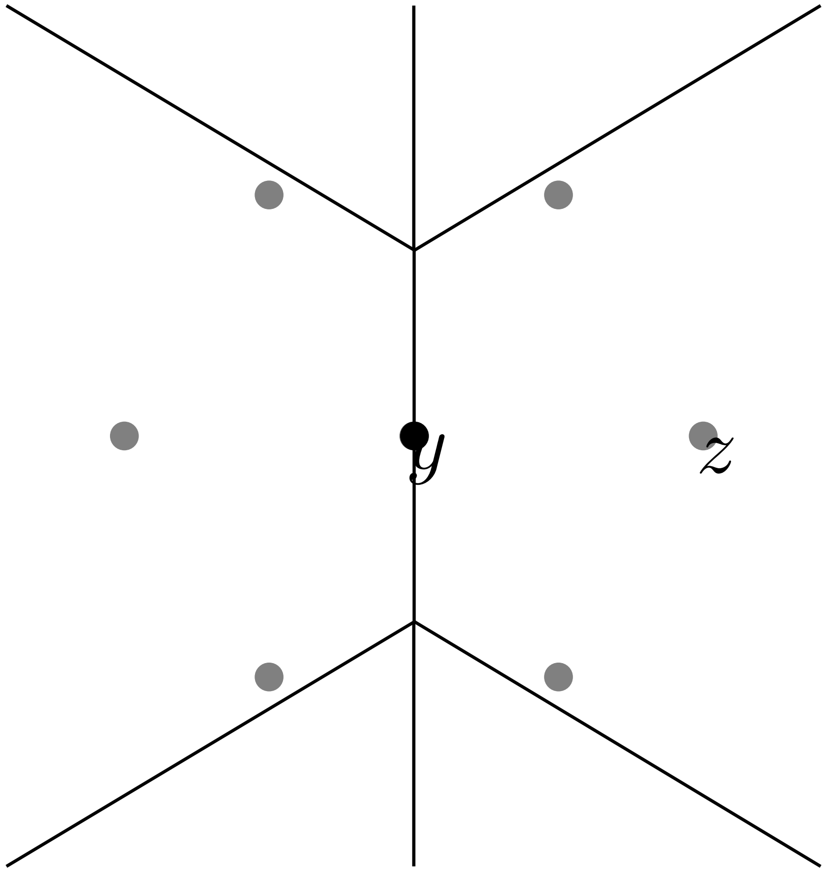
we have $G_y(\psi) = \rho(\mathrm{Lag}_\psi(y))$ $c(x,y) := \|x-y\|^2$

Proposition: For $\psi \in E_\varepsilon$, and assuming that $\rho \in \mathcal{C}^0_c(\mathbb{R}^d)$ one has

(A) $\frac{\partial G_y}{\partial z}(\psi) = \frac{1}{2\|y-z\|} \boxed{\int_{\mathrm{Lag}_{yz}(\psi)}} \rho(x)\,\mathrm{d}\,x$ (B) $\frac{\partial G_y}{\partial y}(\psi) = -\sum_{z\neq y} \frac{\partial G_y}{\partial z}(\psi)$

$z \neq y$

$\mathrm{Lag}_{yz}(\psi) := \mathrm{Lag}_y(\psi) \cap \mathrm{Lag}_z(\psi)$



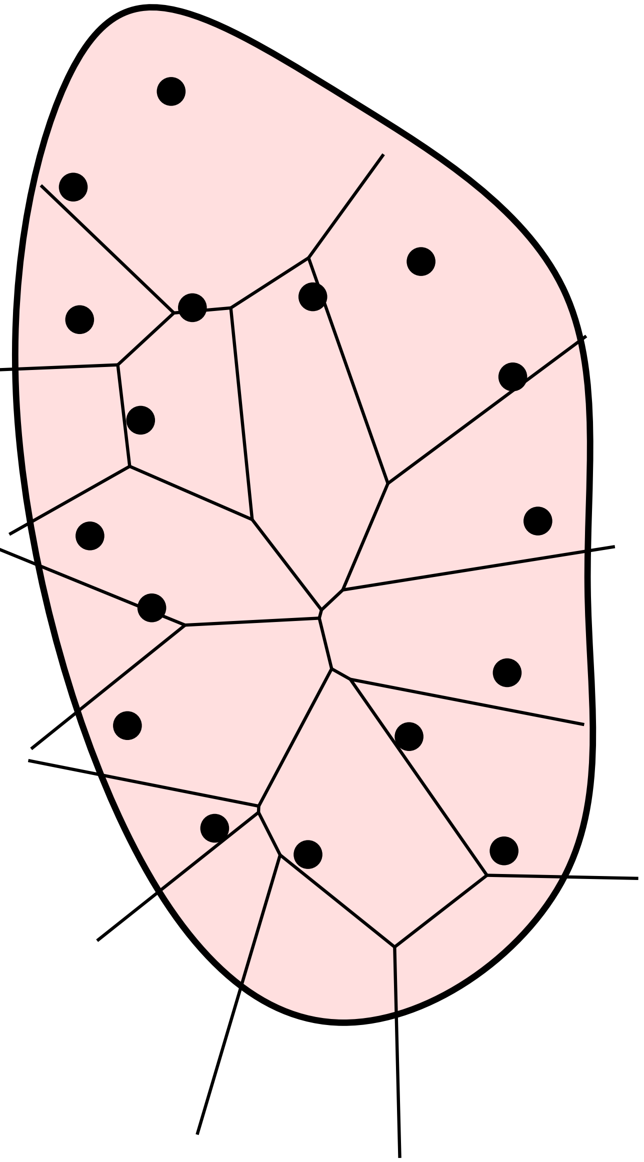
Let $\psi_t := \psi + t\mathbf{1}_z$.When t varies, $\frac{\partial G_y}{\partial z}(\psi_t)$ increases ...
and then suddenly vanishes. $\rightsquigarrow$ we require $\rho(\mathrm{Lag}_\psi(y)) > 0$ at all times

Quadratic cost: strict monotonicity of G

we have $G_y(\psi) = \rho(\mathrm{Lag}_\psi(y))$

Recall: $\frac{\partial G_y}{\partial z}(\psi) = \int_{\boxed{\mathrm{Lag}_{yz}(\psi)}} \frac{\rho(x)\,\mathrm{d}\,x}{2\|y-z\|}$ $\qquad$ $\frac{\partial G_y}{\partial y}(\psi) = -\sum_{z\neq y} \frac{\partial G_y}{\partial z}(\psi)$

$\mathrm{Lag}_{yz}(\psi) := \mathrm{Lag}_y(\psi) \cap \mathrm{Lag}_z(\psi)$

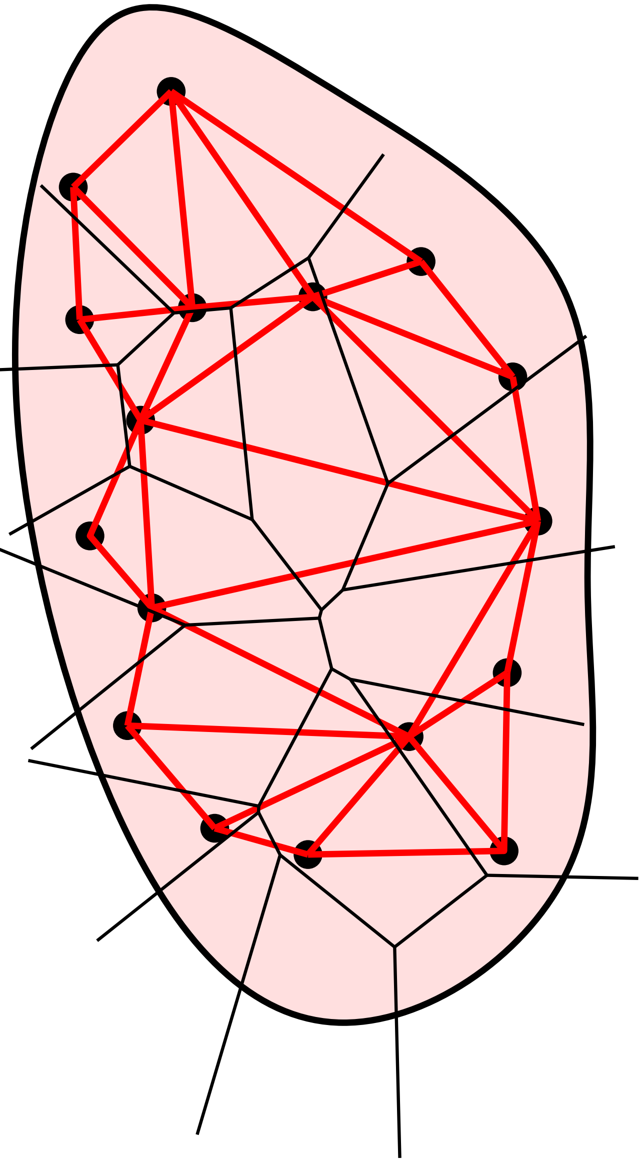


Quadratic cost: strict monotonicity of G

we have $G_y(\psi) = \rho(\mathrm{Lag}_\psi(y))$

Recall: $\frac{\partial G_y}{\partial z}(\psi) = \int_{\boxed{\mathrm{Lag}_{yz}(\psi)}} \frac{\rho(x)\,\mathrm{d}\,x}{2\|y-z\|}$ $\qquad$ $\frac{\partial G_y}{\partial y}(\psi) = -\sum_{z\neq y} \frac{\partial G_y}{\partial z}(\psi)$

$\mathrm{Lag}_{yz}(\psi) := \mathrm{Lag}_y(\psi) \cap \mathrm{Lag}_z(\psi)$



- Consider the matrix $(L_{yz}) := \frac{\partial G_y}{\partial z}(\psi)$ and the graph H:

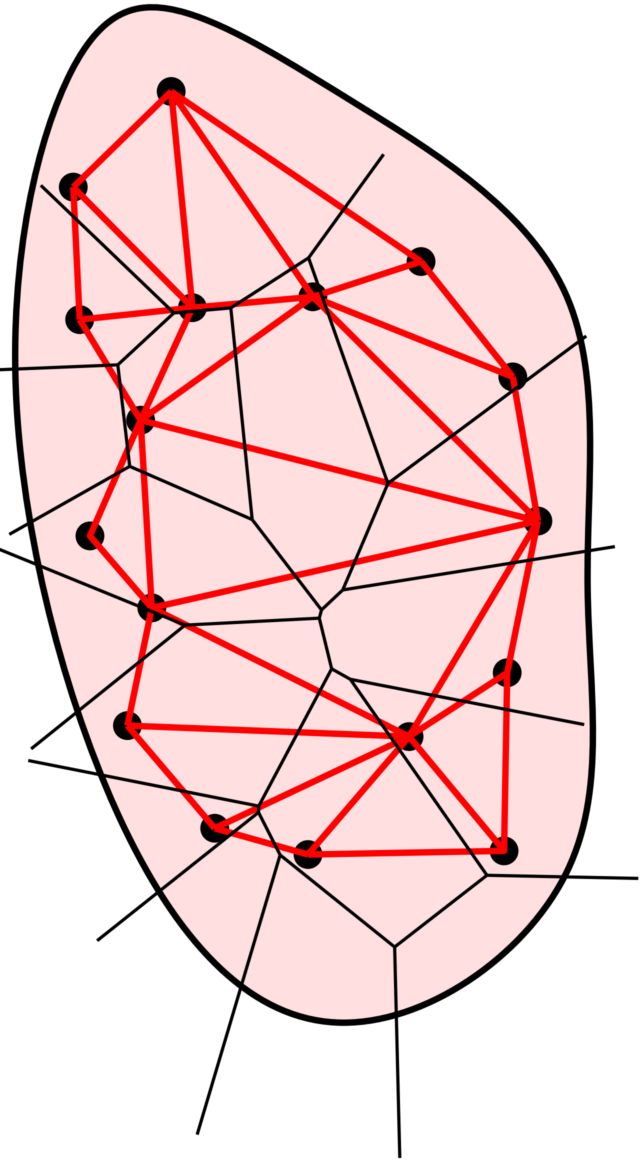
$$\boxed{(y, z) \in H \iff L_{zy} > 0 \iff \mathrm{Lag}_{yz}(\psi) \cap \{\rho > 0\} \neq \emptyset.}$$

Quadratic cost: strict monotonicity of G

we have $G_y(\psi) = \rho(\mathrm{Lag}_\psi(y))$

Recall: $\frac{\partial G_y}{\partial z}(\psi) = \int_{\boxed{\mathrm{Lag}_{yz}(\psi)}} \frac{\rho(x)\,\mathrm{d}\,x}{2\|y-z\|}$ $\qquad$ $\frac{\partial G_y}{\partial y}(\psi) = -\sum_{z \neq y} \frac{\partial G_y}{\partial z}(\psi)$

$\mathrm{Lag}_{yz}(\psi) := \mathrm{Lag}_y(\psi) \cap \mathrm{Lag}_z(\psi)$



- Consider the matrix $(L_{yz}) := \frac{\partial G_y}{\partial z}(\psi)$ and the graph H:

$$\boxed{(y, z) \in H \iff L_{zy} > 0 \iff \mathrm{Lag}_{yz}(\psi) \cap \{\rho > 0\} \neq \emptyset.}$$

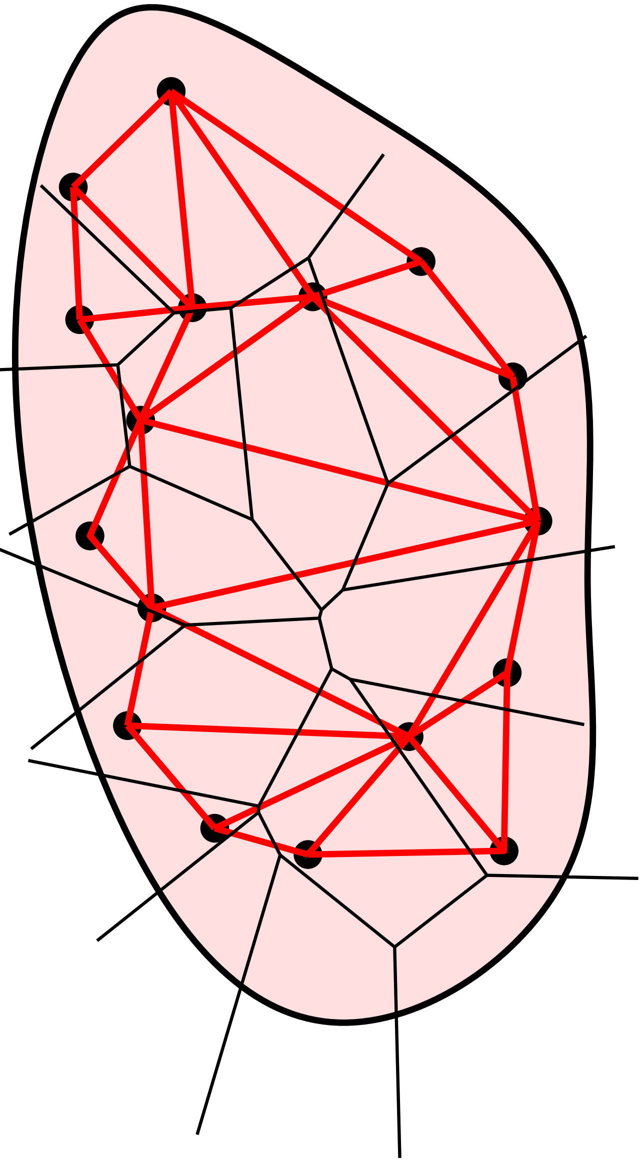
- If $\{\rho > 0\}$ is connected and $\psi \in E_\varepsilon$, then H is connected.

Quadratic cost: strict monotonicity of G

we have $G_y(\psi) = \rho(\mathrm{Lag}_\psi(y))$

Recall: $\frac{\partial G_y}{\partial z}(\psi) = \int_{\boxed{\mathrm{Lag}_{yz}(\psi)}} \frac{\rho(x)\,\mathrm{d}\,x}{2\|y-z\|}$ $\qquad$ $\frac{\partial G_y}{\partial y}(\psi) = -\sum_{z\neq y} \frac{\partial G_y}{\partial z}(\psi)$

$\mathrm{Lag}_{yz}(\psi) := \mathrm{Lag}_y(\psi) \cap \mathrm{Lag}_z(\psi)$



- Consider the matrix $(L_{yz}) := \frac{\partial G_y}{\partial z}(\psi)$ and the graph H:

$$\boxed{(y,z) \in H \iff L_{zy} > 0 \iff \mathrm{Lag}_{yz}(\psi) \cap \{\rho > 0\} \neq \emptyset.}$$

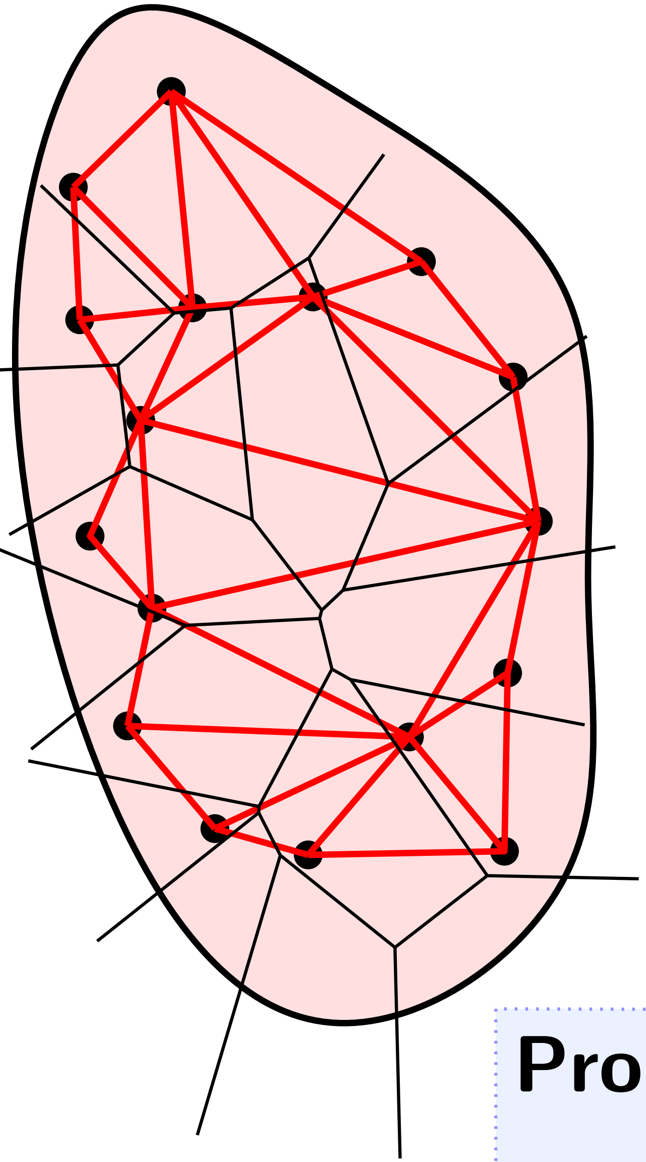
- If $\{\rho > 0\}$ is connected and $\psi \in E_\varepsilon$, then H is connected.
- The second eigenvector of L is strictly negative

Quadratic cost: strict monotonicity of G

we have $G_y(\psi) = \rho(\mathrm{Lag}_\psi(y))$

Recall: $\frac{\partial G_y}{\partial z}(\psi) = \boxed{\int_{\mathrm{Lag}_{yz}(\psi)}} \frac{\rho(x)\,\mathrm{d}\,x}{2\|y-z\|}$ $\qquad$ $\frac{\partial G_y}{\partial y}(\psi) = -\sum_{z\neq y} \frac{\partial G_y}{\partial z}(\psi)$

$\mathrm{Lag}_{yz}(\psi) := \mathrm{Lag}_y(\psi) \cap \mathrm{Lag}_z(\psi)$



- Consider the matrix $(L_{yz}) := \frac{\partial G_y}{\partial z}(\psi)$ and the graph H:

$$\boxed{(y,z) \in H \iff L_{zy} > 0 \iff \mathrm{Lag}_{yz}(\psi) \cap \{\rho > 0\} \neq \emptyset.}$$

- If $\{\rho > 0\}$ is connected and $\psi \in E_\varepsilon$, then H is connected.
- The second eigenvector of L is strictly negative

Proposition: Assume $\rho \in \mathcal{C}^0_c(\mathbb{R}^d)$ and $\{\rho > 0\}$ connected. Then,
$\forall \psi \in E_\varepsilon$, $\;DG(\psi)$ is neg. definite on $E_\varepsilon \cap \{cst\}^\perp$

$\rightsquigarrow$ we require connectedness condtions on ρ

Ma Trudinger Wang cost

Cost satisfying Loeper's MTW condition

$\rightarrow$ MTW: non-local 4th order inequality appearing in the regularity theory for OT

$\rightarrow$ we rely on a (slightly modified) geometric reformulation due to Loeper.

Cost satisfying Loeper's MTW condition

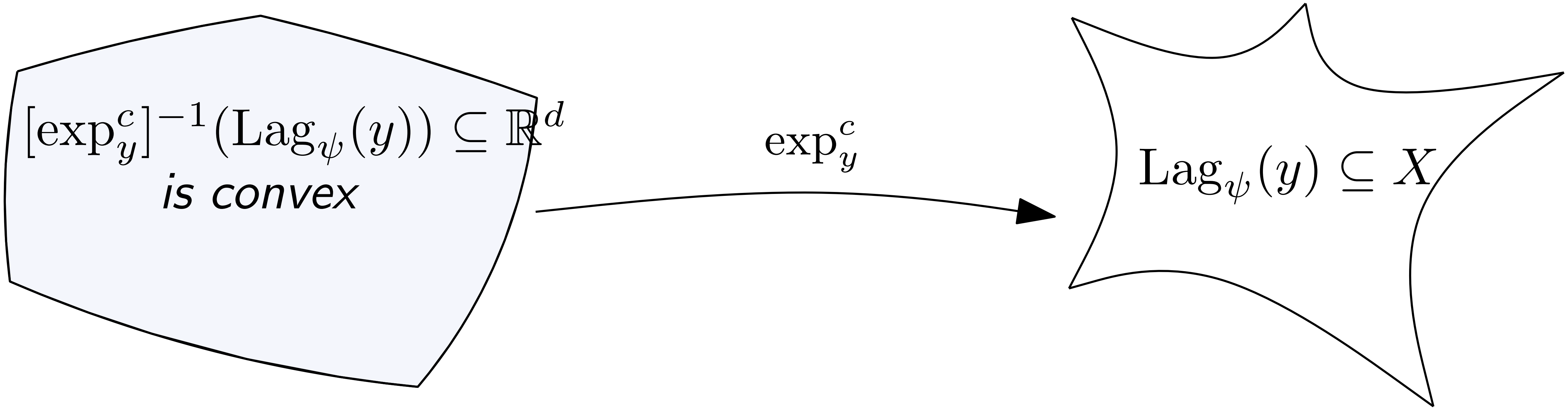
→ MTW: non-local 4th order inequality appearing in the regularity theory for OT

→ we rely on a (slightly modified) geometric reformulation due to Loeper.

Def: The cost function $c : X \times Y$ satisfies Loeper's condition if for every $y \in Y$, there exists a diffeomorphism $\exp_y^c : X_y \subseteq \mathbb{R}^d \to X$ s.t.

$$v \in X_y \mapsto c(\exp_y^c(v), y) - c(\exp_y^c(v), z) \text{ is quasi-convex } \forall z$$

$\rightsquigarrow$ for all $\psi \in Y^{\mathbb{R}}$, $[\exp_y^c]^{-1}(\mathrm{Lag}_\psi(y))$ is convex



MTW cost: Convergence result

Theorem: Let X be a (closed) bounded domain of $\mathbb{R}^d$ with smooth boundar Y be a finite set and $c \in \mathcal{C}^2(X \times Y)$. Assume:

(A) c satisfies **(Twist)**, **(MTW)** and X is c-convex

(B) $\rho \in \mathcal{C}^\alpha(X)$ and satisfies a weighted L^1-Poincaré inequality, i.e.

$$\forall f \in \mathcal{C}^1(X), \quad \|f - \mathbb{E}_\rho(f)\|_{\mathrm{L}^1(\rho)} \leq \mathrm{cst} \cdot \|\nabla f\|_{\mathrm{L}^1(\rho)}$$

Then, the damped Newton algorithm for SD-OT converges **globally** with linear rate and locally with $1 + \alpha$ rate.

[Kitagawa, Mérigot, T., JEMS '17]

MTW cost: Convergence result

Theorem: Let X be a (closed) bounded domain of $\mathbb{R}^d$ with smooth boundary, Y be a finite set and $c \in \mathcal{C}^2(X \times Y)$. Assume:

(A) c satisfies **(Twist)**, **(MTW)** and X is c-convex

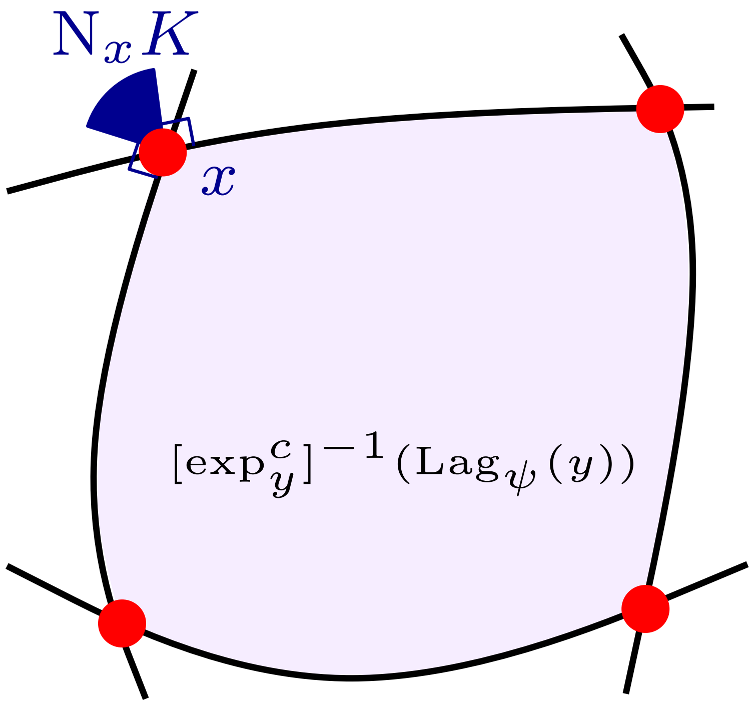
(B) $\rho \in \mathcal{C}^\alpha(X)$ and satisfies a weighted L^1-Poincaré inequality, i.e.

$$\forall f \in \mathcal{C}^1(X), \quad \|f - \mathbb{E}_\rho(f)\|_{\mathrm{L}^1(\rho)} \leq \mathrm{cst} \cdot \|\nabla f\|_{\mathrm{L}^1(\rho)}$$

Then, the damped Newton algorithm for SD-OT converges **globally** with linear rate and locally with $1+\alpha$ rate.

[Kitagawa, Mérigot, T., JEMS '17]

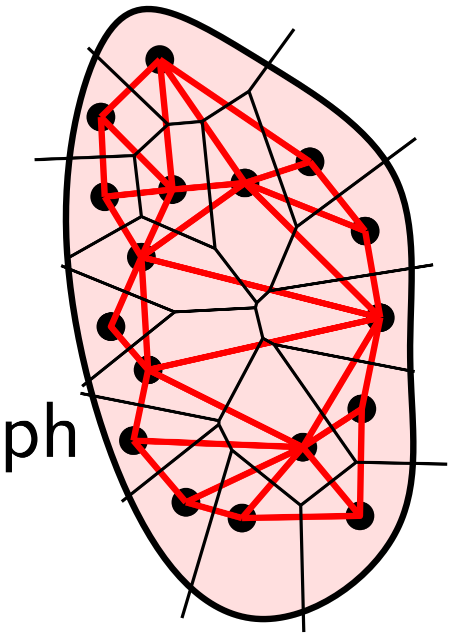
Proof:



→ convexity

→ transversality

→ connectedness of the graph



MTW cost: Convergence result

Theorem: Let X be a (closed) bounded domain of $\mathbb{R}^d$ with smooth boundar
Y be a finite set and $c \in \mathcal{C}^2(X \times Y)$. Assume:

(A) c satisfies **(Twist)**, **(MTW)** and X is c-convex

(B) $\rho \in \mathcal{C}^\alpha(X)$ and satisfies a weighted L^1-Poincaré inequality, i.e.

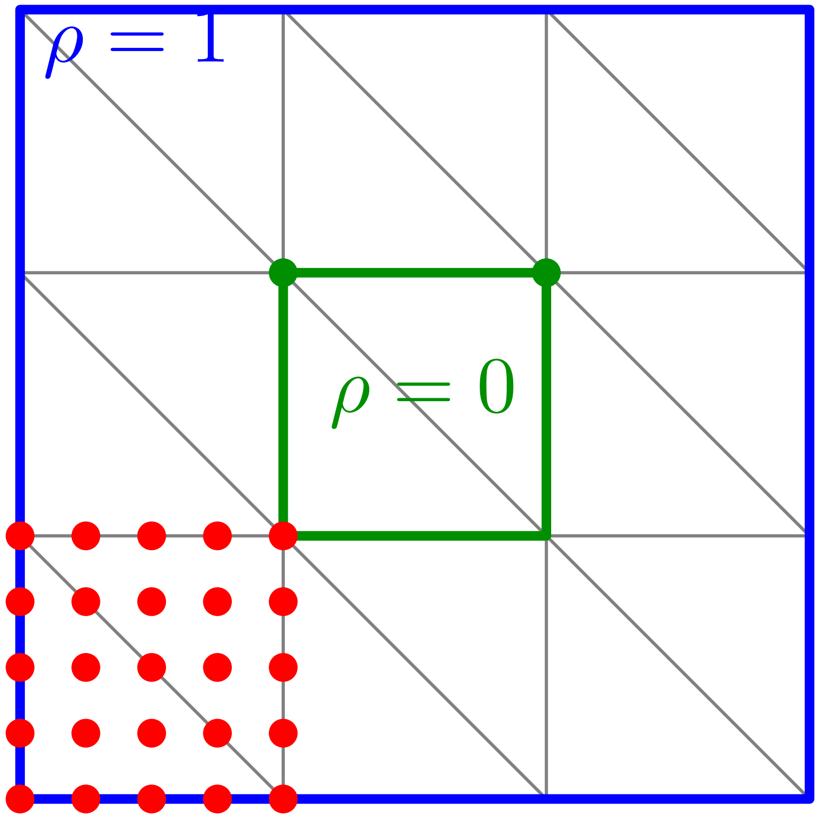
$$\forall f \in \mathcal{C}^1(X), \quad \|f - \mathbb{E}_\rho(f)\|_{\mathrm{L}^1(\rho)} \leq \mathrm{cst} \cdot \|\nabla f\|_{\mathrm{L}^1(\rho)}$$

Then, the damped Newton algorithm for SD-OT converges **globally** with linear rate and locally with $1 + \alpha$ rate.

[Kitagawa, Mérigot, T., JEMS '17]

- The condition **(B)** seems to allow vanishing densities on X.
- Condition **(A)** applies to reflector problems.

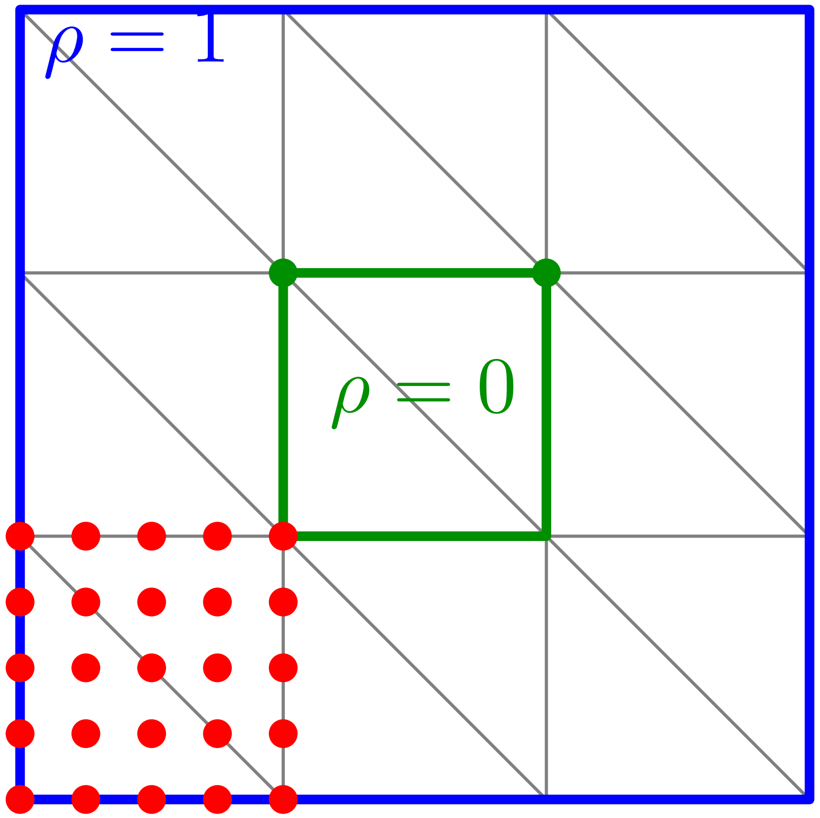
Quadratic cost: numerics

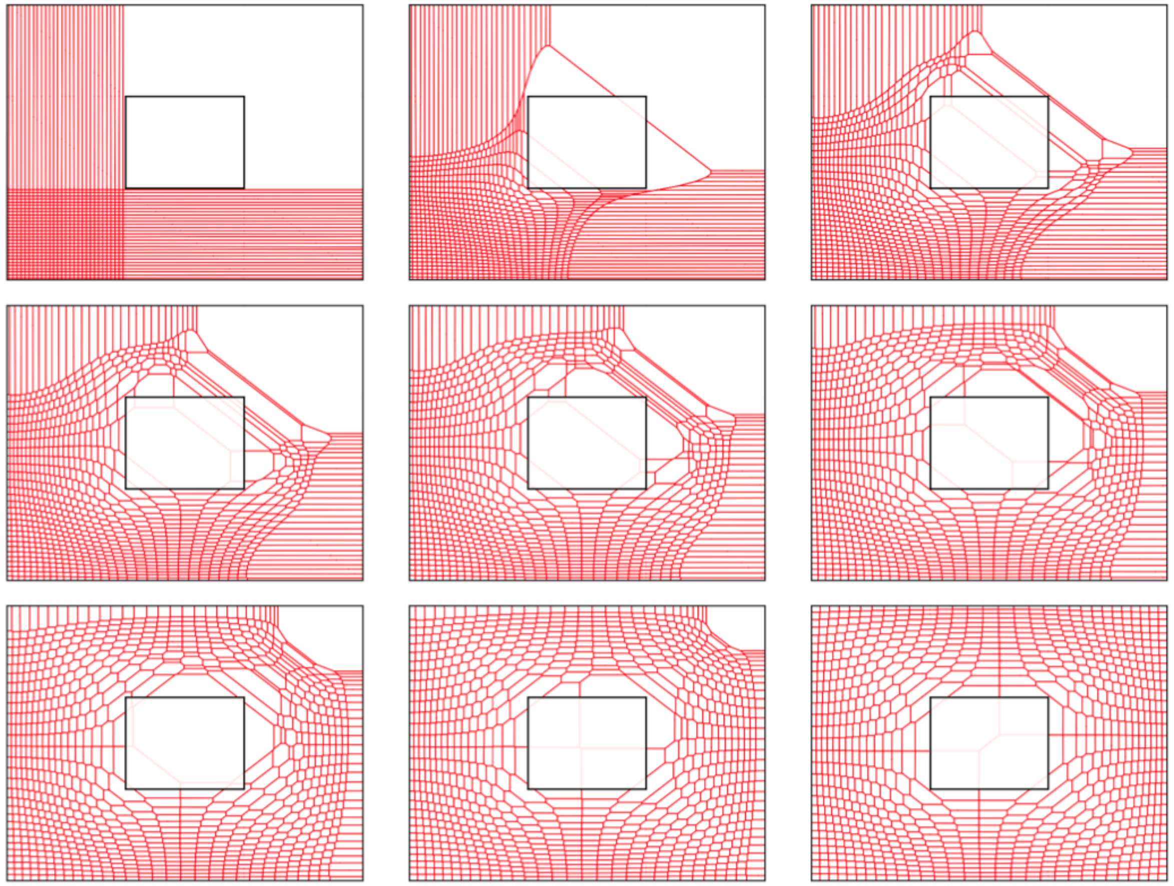


Source: PL density on $X = [0,3]^2$

Target: Uniform grid Y in $[0,1]^2$.

Quadratic cost: numerics



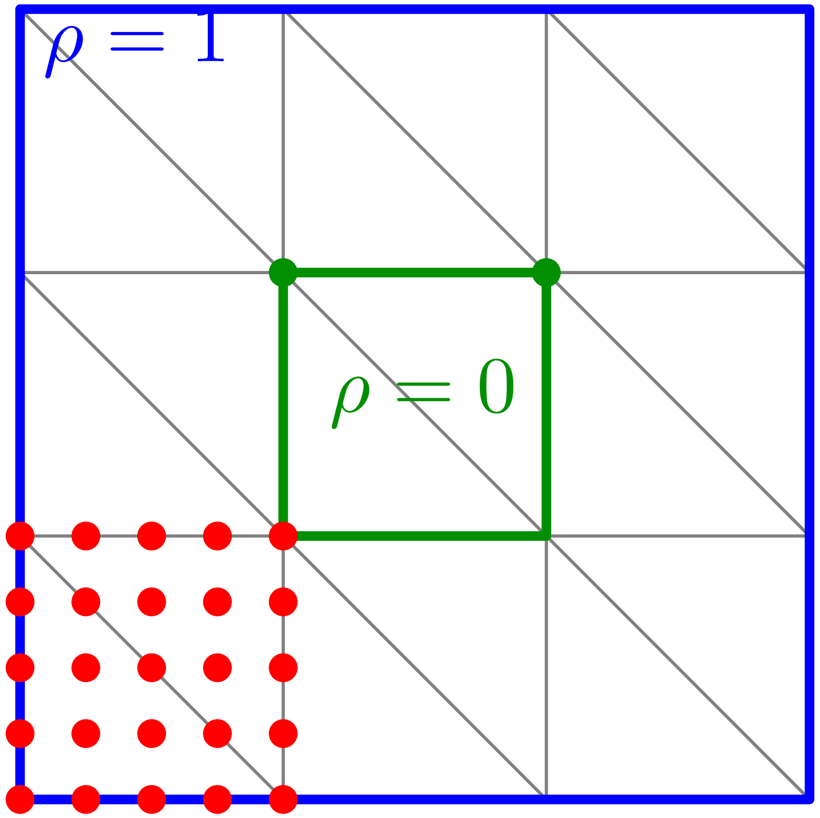


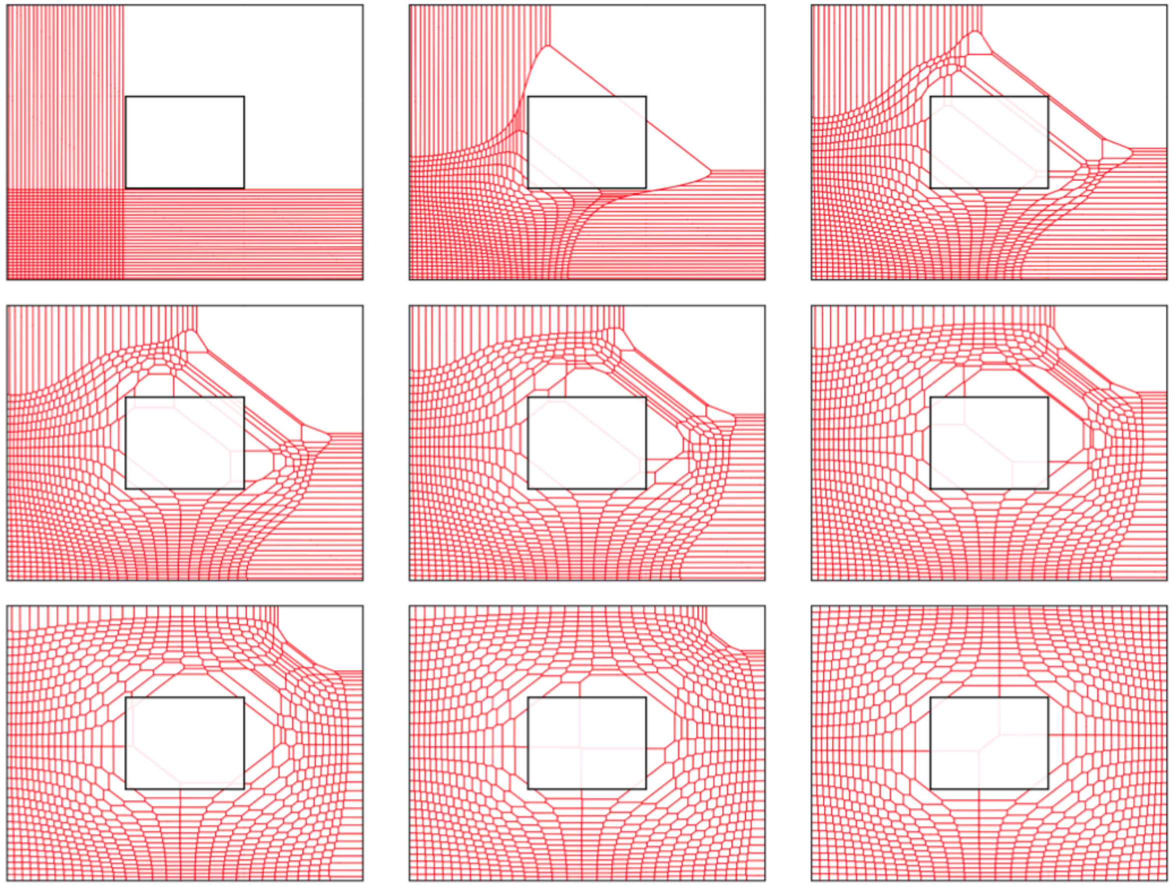
Source: PL density on $X = [0,3]^2$

Target: Uniform grid Y in $[0,1]^2$.

- The damped Newton's algorithm converges even when ρ vanishes.

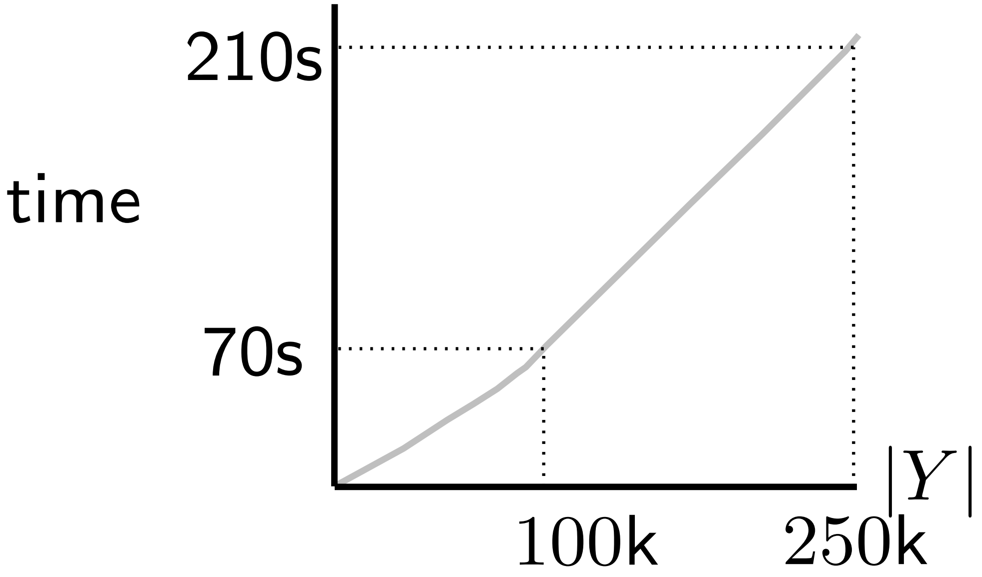
Quadratic cost: numerics





Source: PL density on $X = [0,3]^2$

Target: Uniform grid Y in $[0,1]^2$.



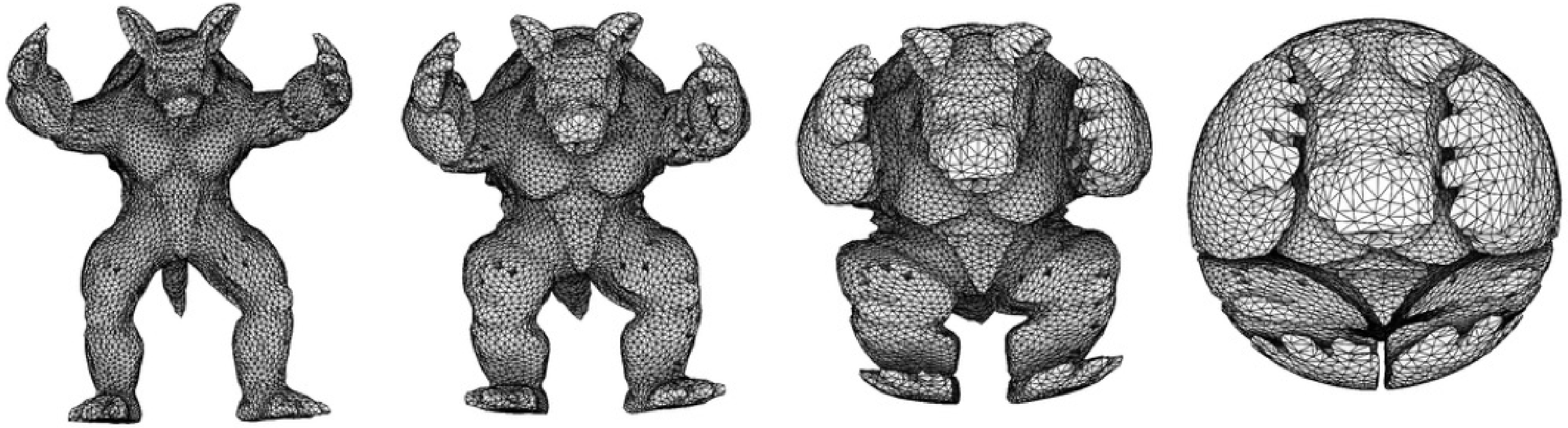
- The damped Newton's algorithm converges even when ρ vanishes.
- Computational cost seems nearly linear in number of Diracs.

Quadratic cost: numerics

2D

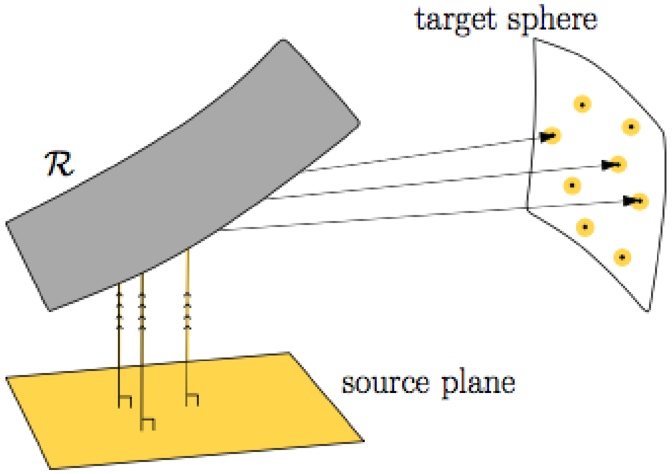
[Mérigot, SGP 2010]

3D

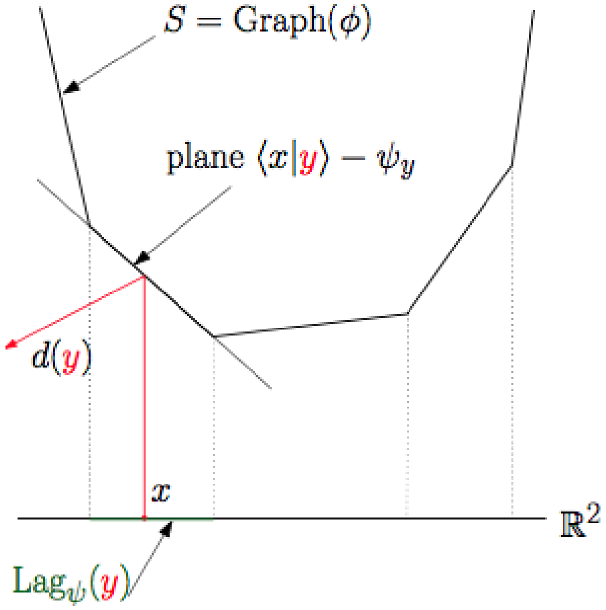


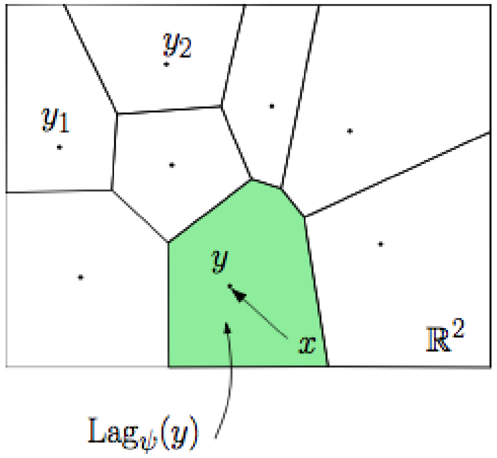
[Levy 2014]
$N = 1$ million, even $N = 10$ millions

Quadratic cost: numerics



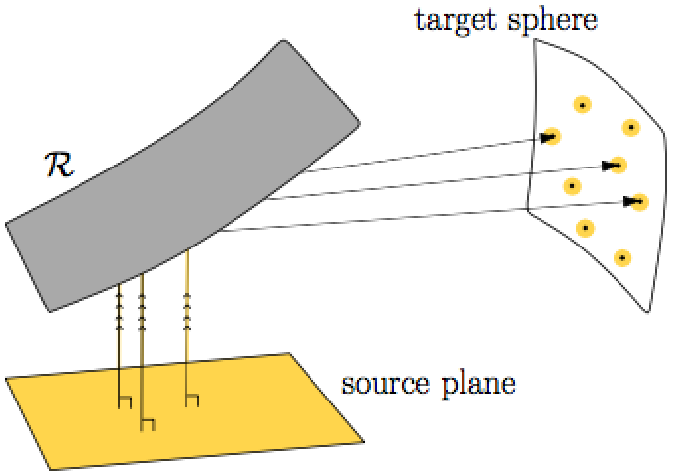
targeted image $N = 400 \times 480$





Reflector : punctual / Far Field

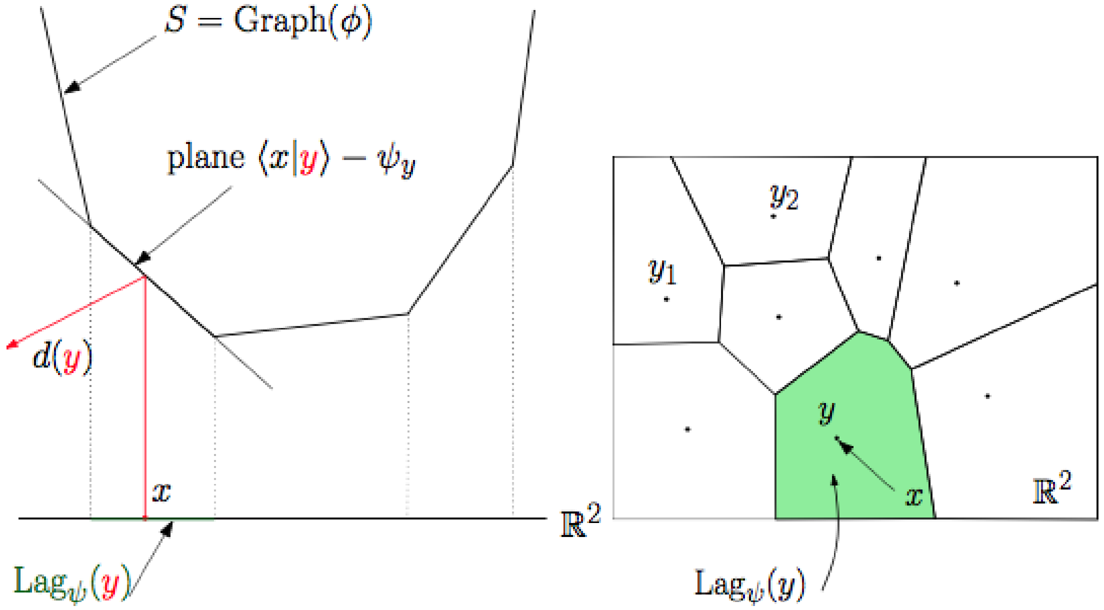
Quadratic cost: numerics



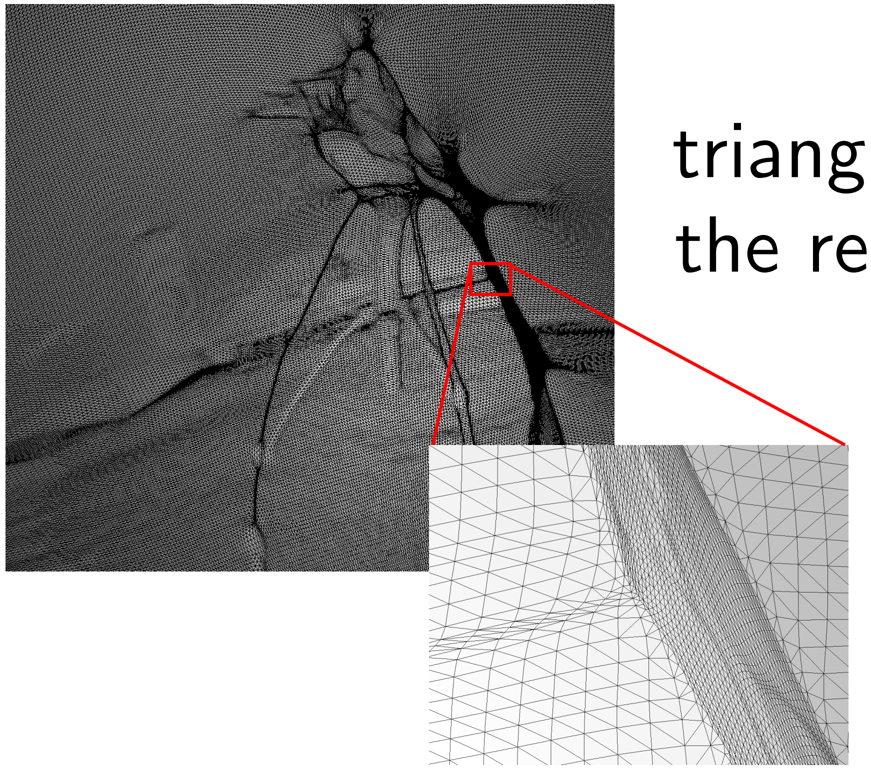
targeted image $N = 400 \times 480$

Experiments by Jocelyn Meyron

rendered image



Reflector : punctual / Far Field

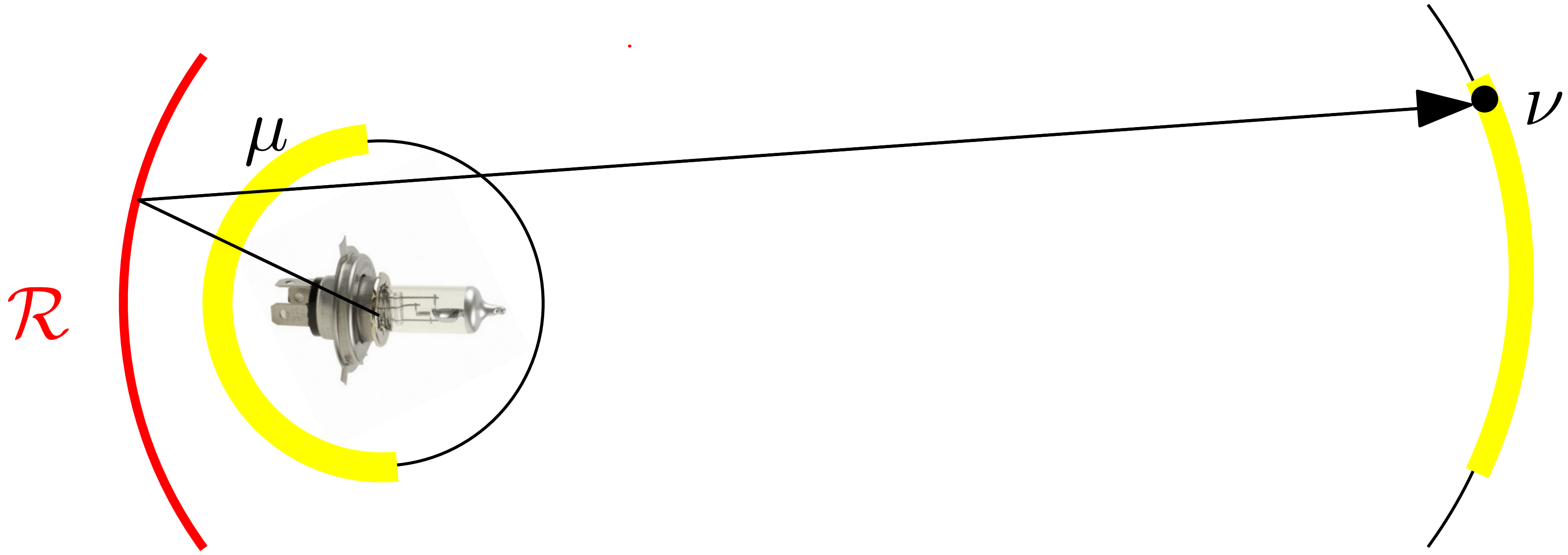


triangulation of the reflector

Reflector problem: Punctual / Far Field

$\nu = \sum_{i=1}^{N} \nu_i \delta_{x_i}$ obtained by discretizing a picture of G. Monge.

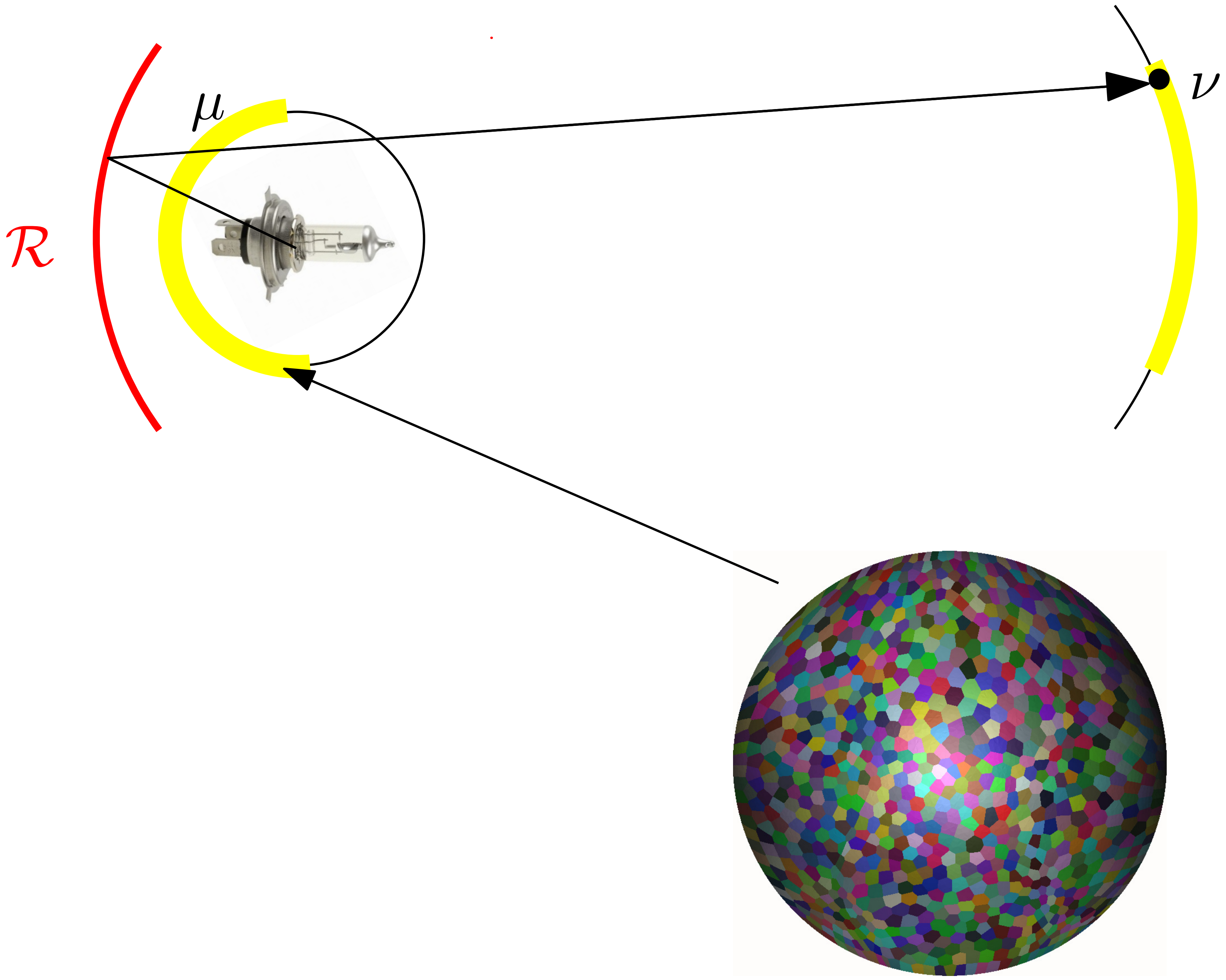
$\mu =$ uniform measure on half-sphere $\mathcal{S}^2_+$ $N = 1000$



Reflector problem: Punctual / Far Field

$\nu = \sum_{i=1}^{N} \nu_i \delta_{x_i}$ obtained by discretizing a picture of G. Monge.

$\mu =$ uniform measure on half-sphere $\mathcal{S}^2_+$ $N = 1000$

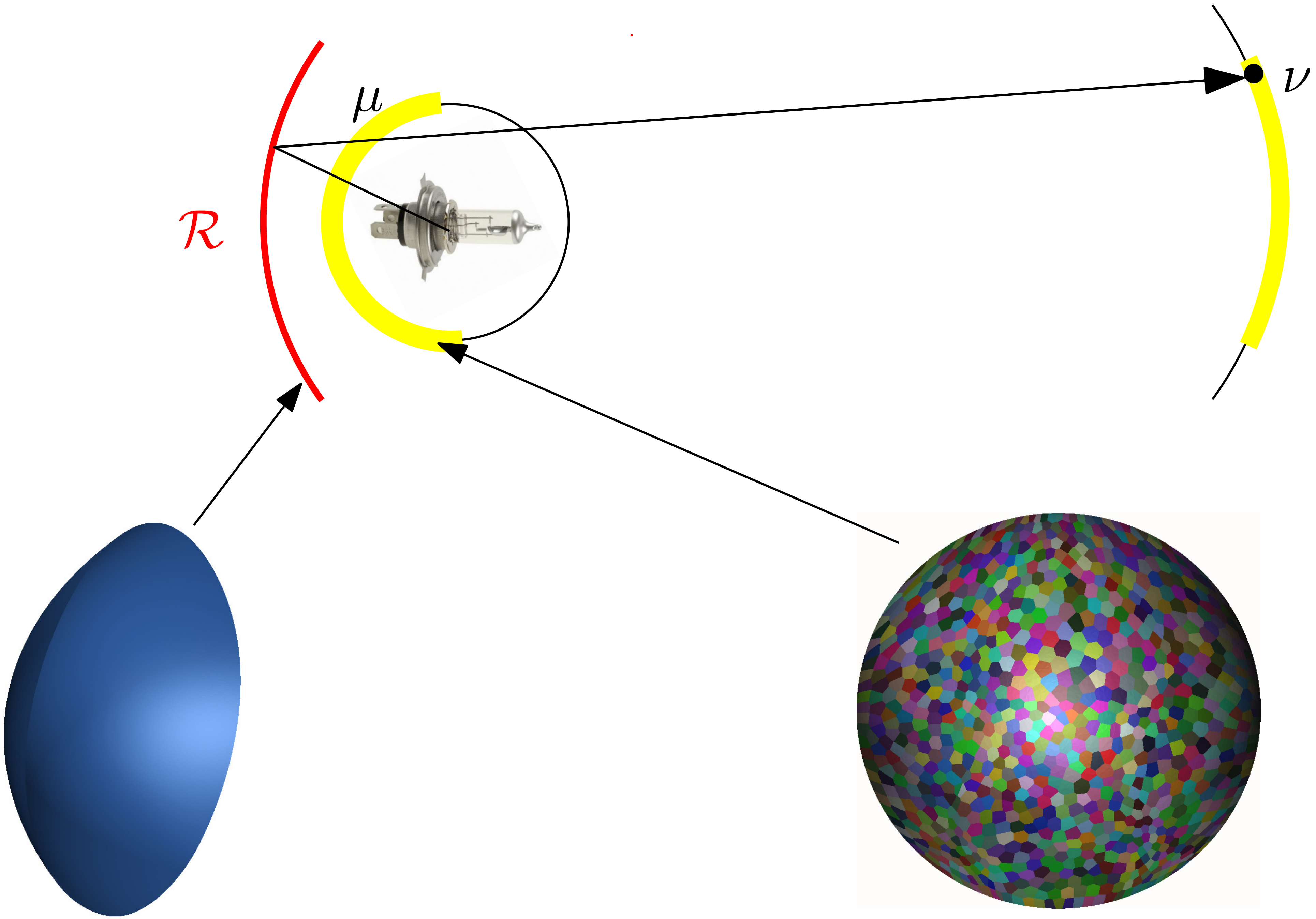


Final Laguerre cells

Reflector problem: Punctual / Far Field

$\nu = \sum_{i=1}^{N} \nu_i \delta_{x_i}$ obtained by discretizing a picture of G. Monge.

$\mu =$ uniform measure on half-sphere $\mathcal{S}^2_+$ $N = 1000$



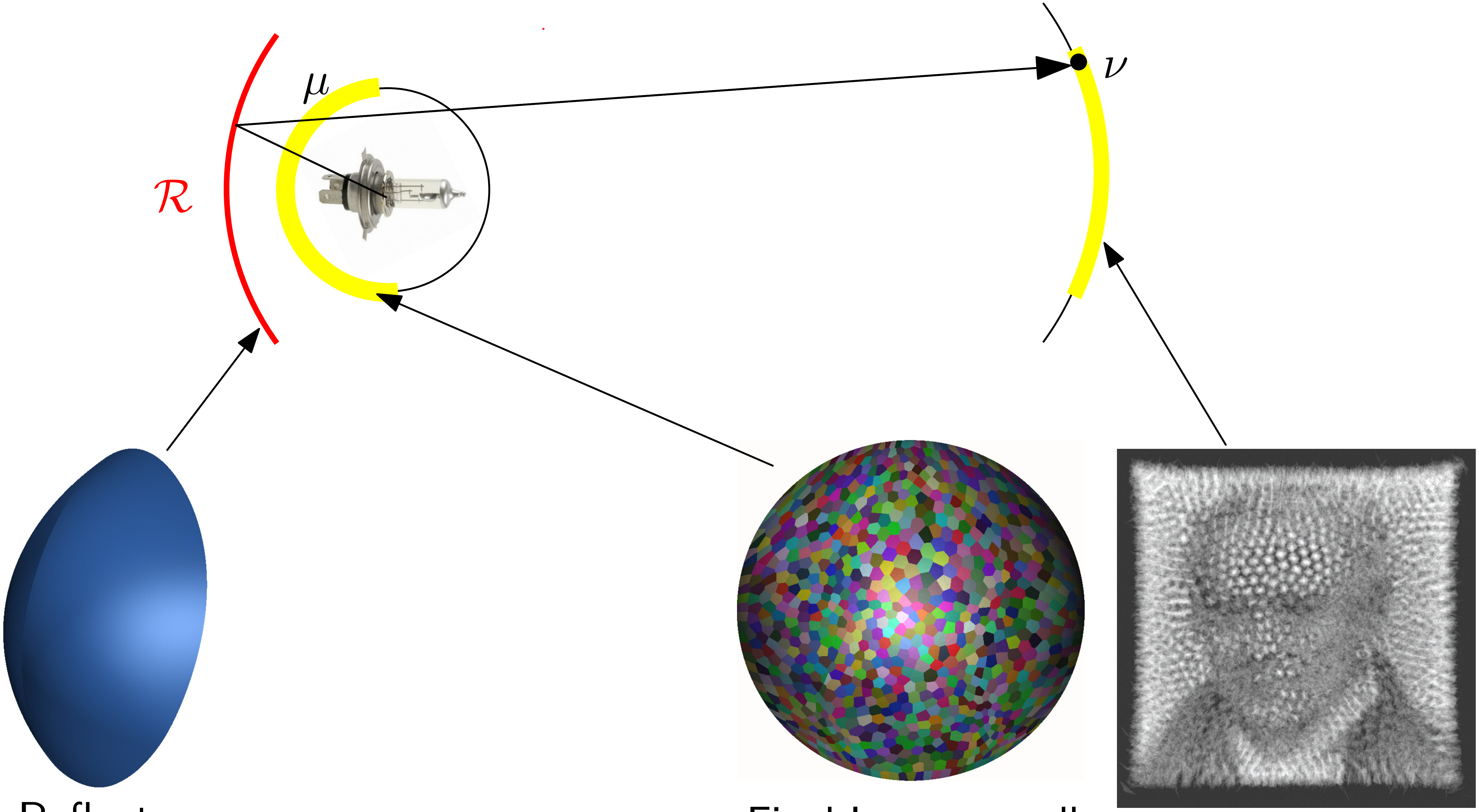
Reflector

Final Laguerre cells

Reflector problem: Punctual / Far Field

$\nu = \sum_{i=1}^{N} \nu_i \delta_{x_i}$ obtained by discretizing a picture of G. Monge.

$\mu =$ uniform measure on half-sphere $\mathcal{S}^2_+$ $\qquad N = 1000$



Reflector

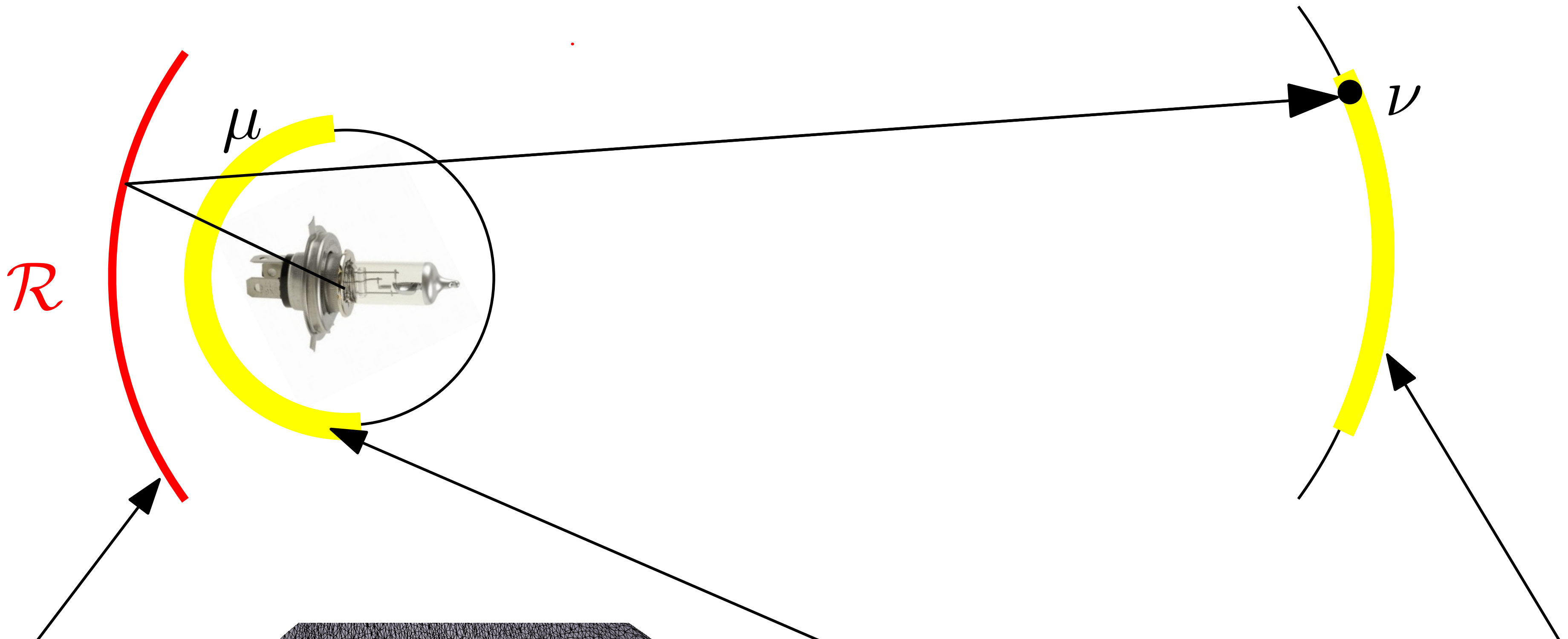
Final Laguerre cells

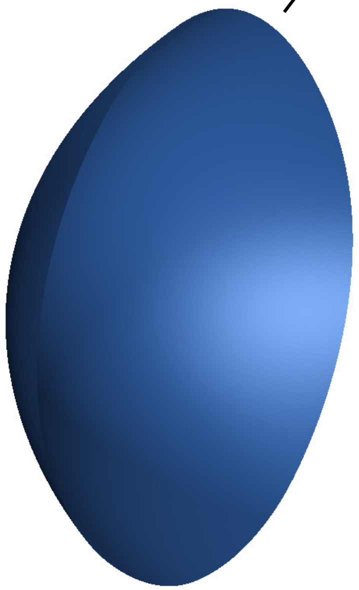
Rendering

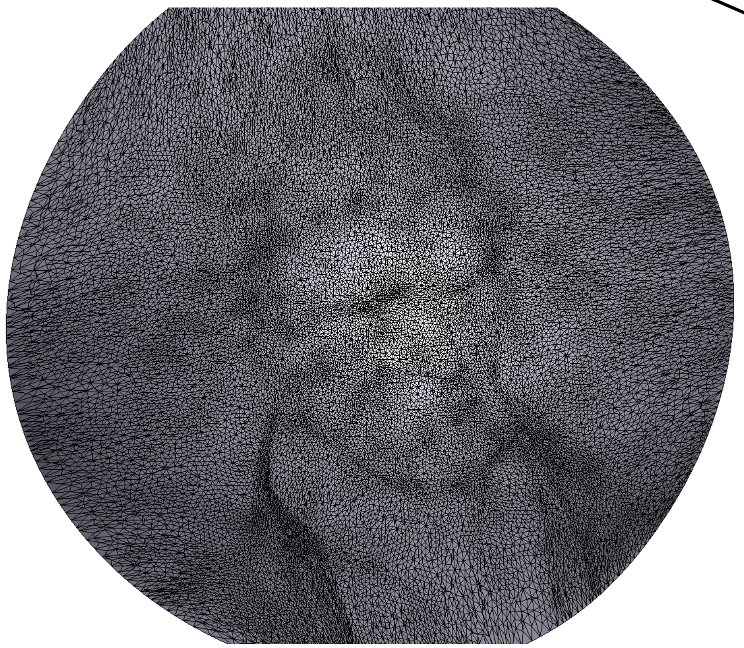
Reflector problem: Punctual / Far Field

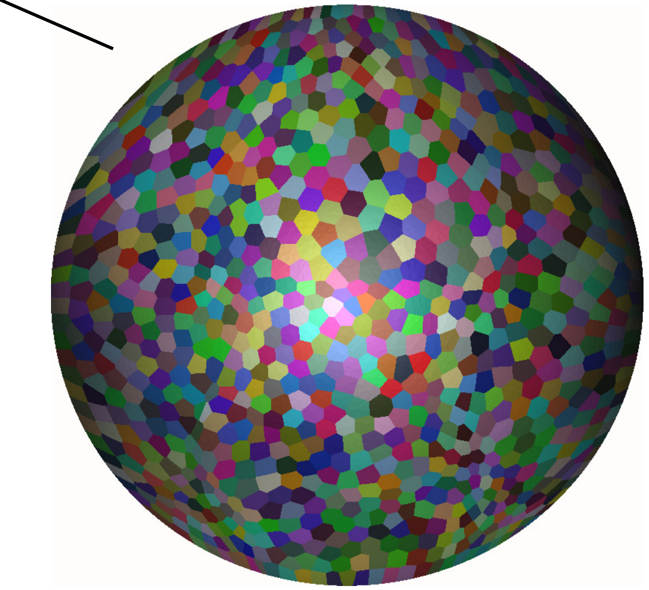
$\nu = \sum_{i=1}^{N} \nu_i \delta_{x_i}$ obtained by discretizing a picture of G. Monge.

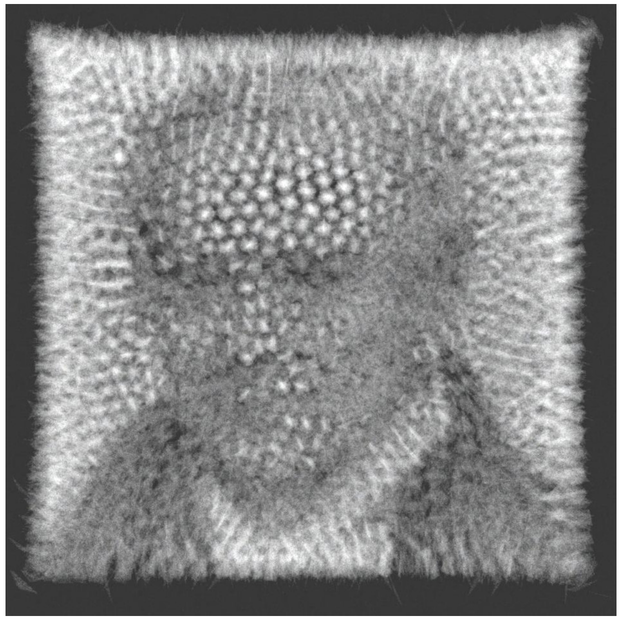
$\mu =$ uniform measure on half-sphere $\mathcal{S}^2_+$ $N = 1000$




Reflector


Triangulated reflector


Final Laguerre cells

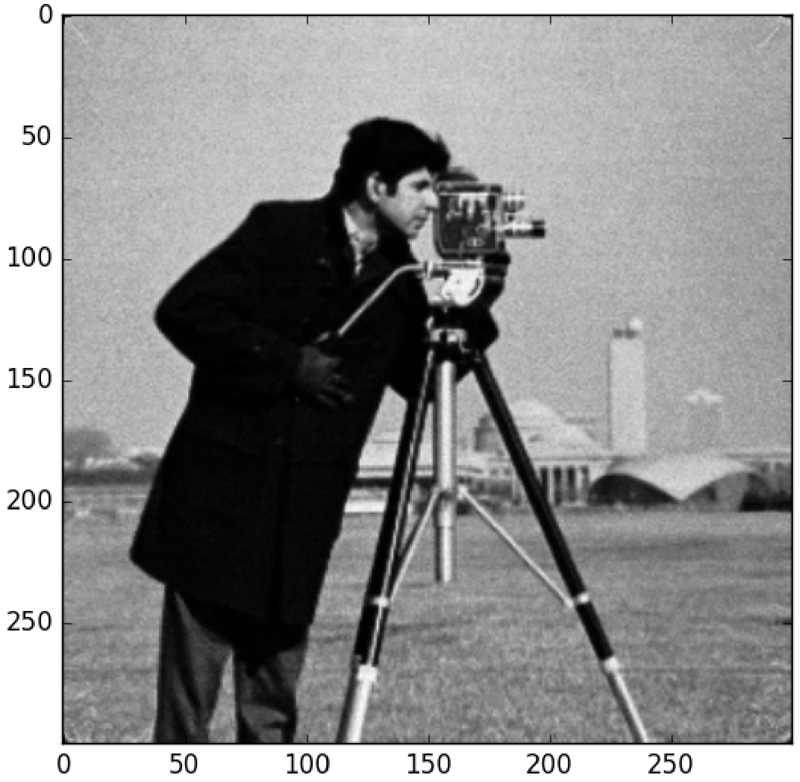

Rendering

Reflector problem: Punctual / Far Field

$\nu = \sum_{i=1}^{N} \nu_i \delta_{x_i}$ obtained by discretizing a picture of G. Monge.

$\mu =$ uniform measure on half-sphere $\mathcal{S}^2_+$ $N = 90,000$

Initial image



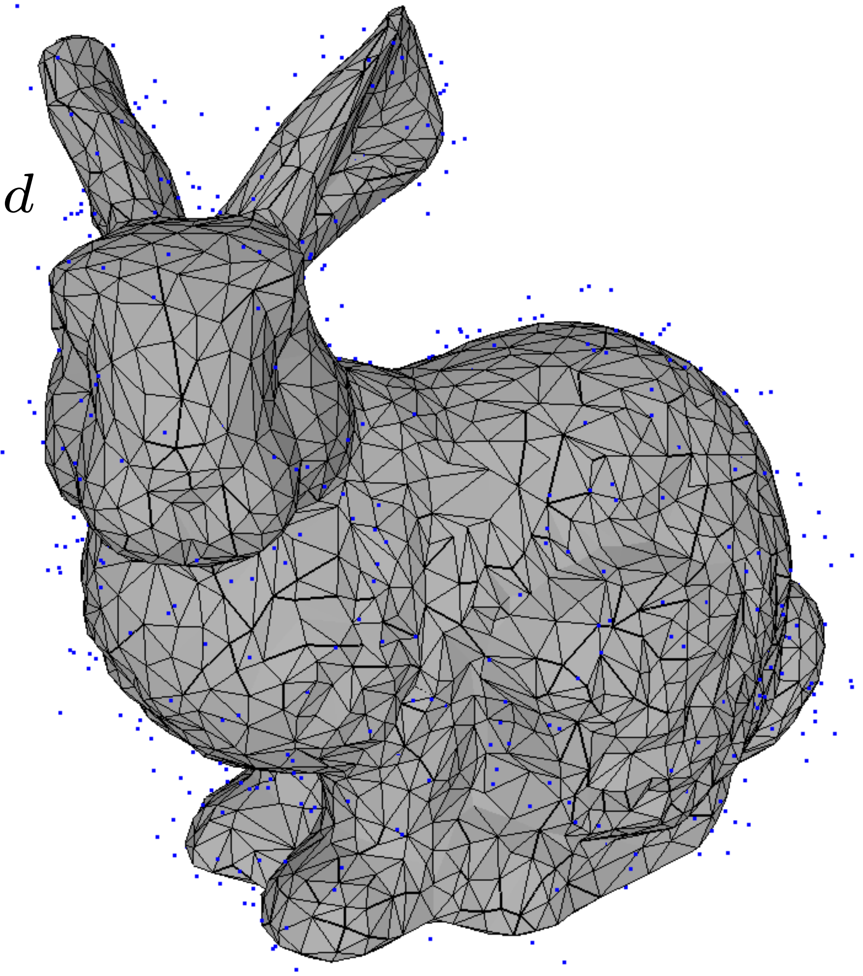
rendered image

Experiments by Jocelyn Meyron

OT between a simplex soup and a point cloud

with Quentin Mérigot and Jocelyn Meyron

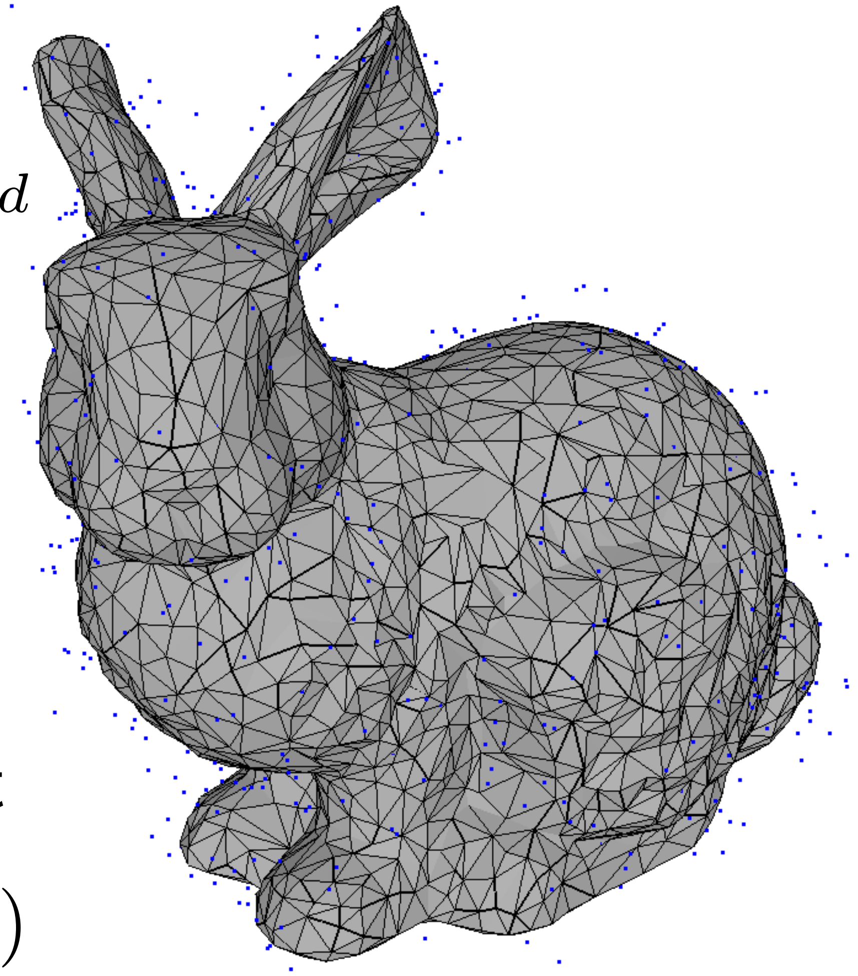
Problematic:



Input:

- A (probability) measure on a simplex soup K in $\mathbb{R}^d$
  $\mu = \sum_\sigma \mu_\sigma$, with σ simplex of any dimension.
- A (probability) measure on a point cloud $Y \subset \mathbb{R}^d$
  $\nu = \sum_y \nu_y \delta_y$.

Problematic:



Input:

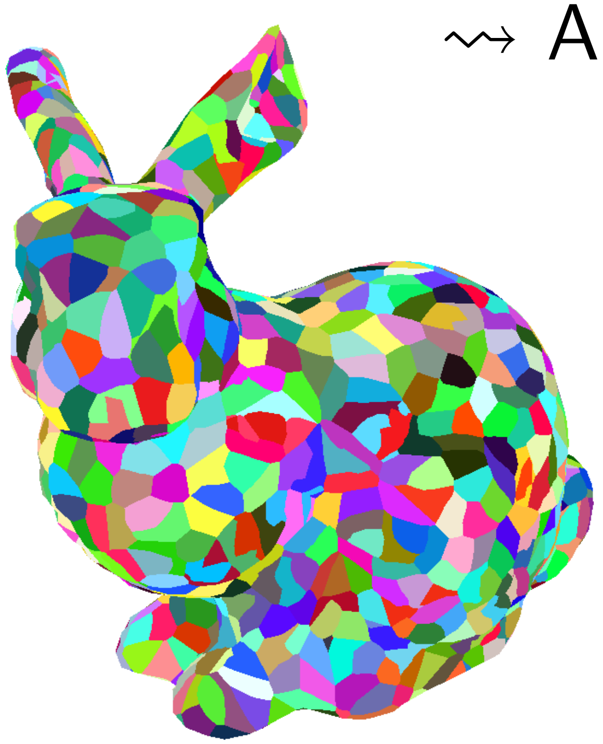
- A (probability) measure on a simplex soup K in $\mathbb{R}^d$
  $\mu = \sum_\sigma \mu_\sigma$, with σ simplex of any dimension.

- A (probability) measure on a point cloud $Y \subset \mathbb{R}^d$
  $\nu = \sum_y \nu_y \delta_y$.

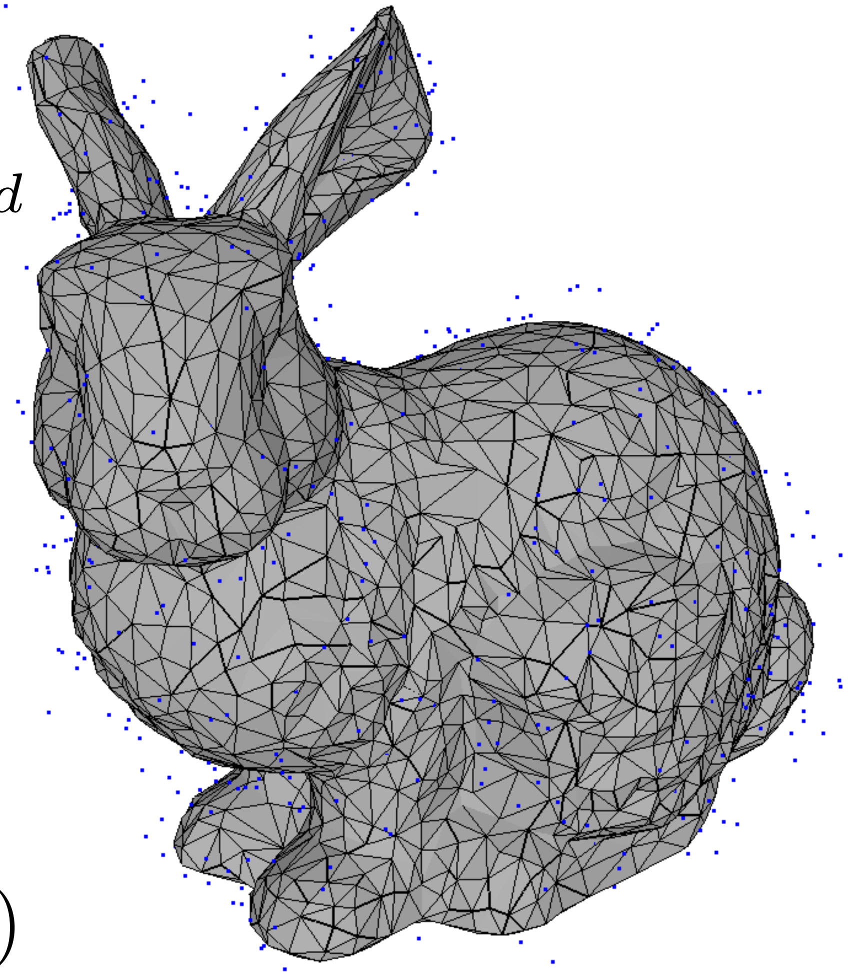
Output:

- Transport plan between μ and ν for quadratic cost

$\rightsquigarrow$ A family of Laguerre cells $\mathrm{Lag}_y(\psi)$



Problematic:



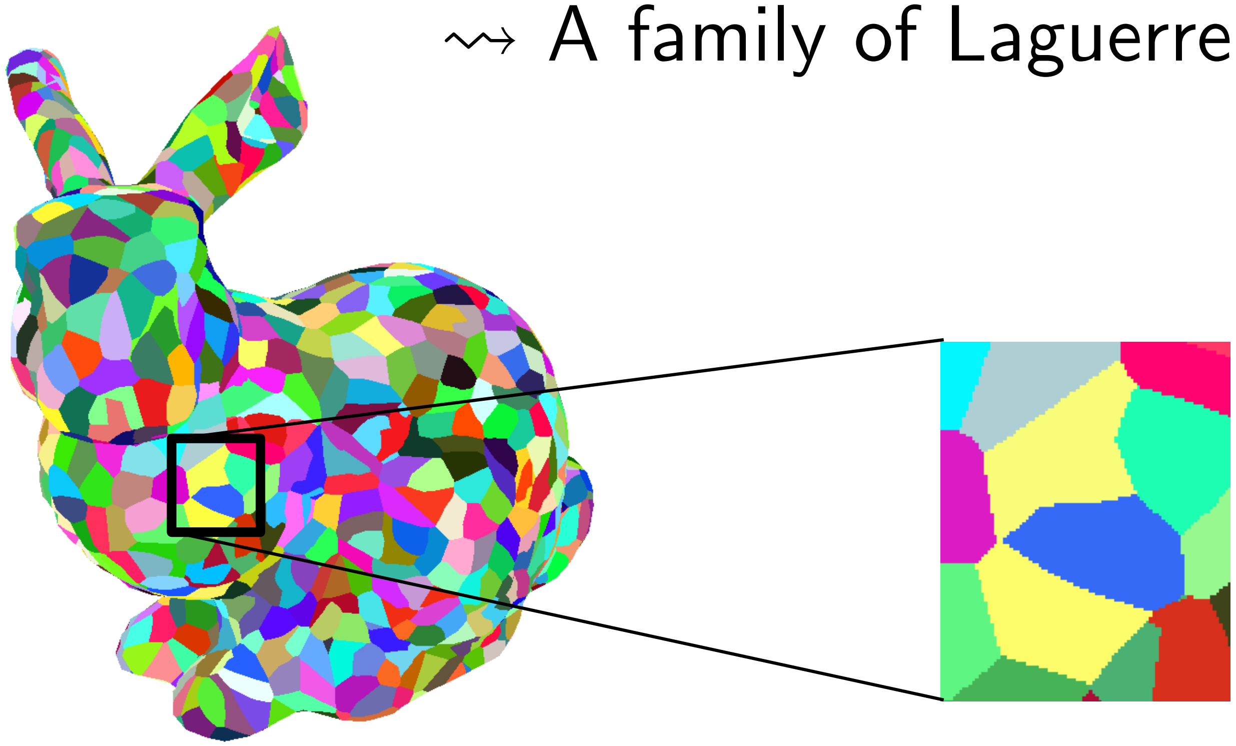
Input:

- A (probability) measure on a simplex soup K in $\mathbb{R}^d$
  $\mu = \sum_\sigma \mu_\sigma$, with σ simplex of any dimension.
- A (probability) measure on a point cloud $Y \subset \mathbb{R}^d$
  $\nu = \sum_y \nu_y \delta_y$.

Output:

- Transport plan between μ and ν for quadratic cost

$\rightsquigarrow$ A family of Laguerre cells $\mathrm{Lag}_y(\psi)$



However does not satisfy MTW:

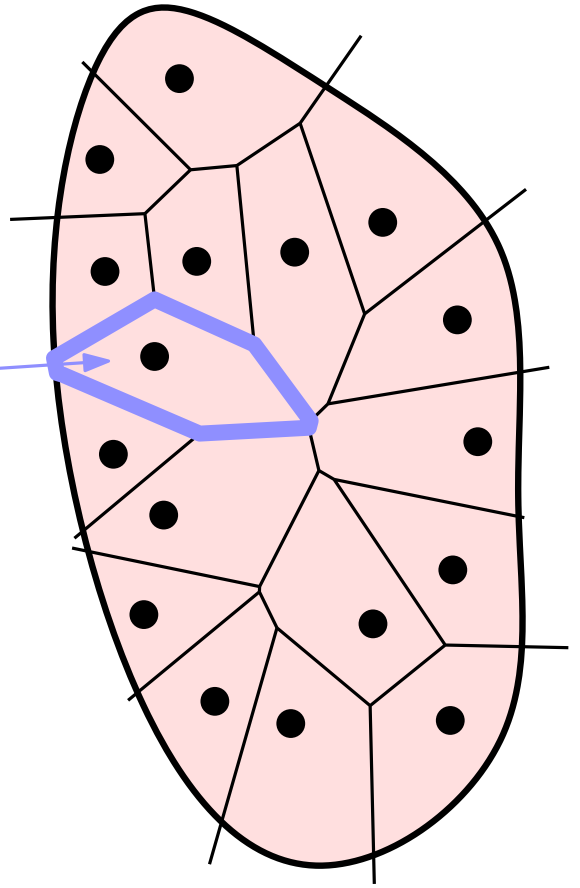
- Not c-convex in general
- Not connected in general

Damped Newton Algorithm

Equation $(\rho(\mathrm{Lag}_y(\psi)) - \nu_y) = 0$

Admissible domain: $E_\varepsilon := \{\psi \in Y^{\mathbb{R}} ; \forall y \in Y, \rho(\mathrm{Lag}_\psi(y)) \geq \varepsilon\}$

We put $G_y(\psi) = \rho(\mathrm{Lag}_y(\psi))$



Damped Newton algorithm: for solving $G(\psi) = \nu$

Input: $\psi_0 \in Y^{\mathbb{R}}$ s.t. $\varepsilon := \frac{1}{2} \min_{y \in Y} \min(G(\psi_0)_y, \nu_y) > 0$

Loop: $\longrightarrow$ Define $\psi_k^\tau = \psi_k - \tau \mathrm{D}G(\psi_k)^{-1}(G(\psi_k) - \nu)$

$\longrightarrow \tau_k := \max\{\tau \in 2^{-\mathbb{N}} \mid \psi_k^\tau \in E_\varepsilon$ and $\|G(\psi_k^\tau) - \nu\| \leq (1 - \frac{\tau}{2})\|G(\psi_k) - \nu\|\}$

$\longrightarrow \psi_{k+1} := \psi_k^{\tau_k}$

Remark: The damped Newton's algorithm converges **globally** provided that:

(Smoothness): $\nabla \mathcal{K} = G - \nu$ is $\mathcal{C}^1$ on E_ε.

(Strict concavity): $\forall \psi \in E_\varepsilon, \ \mathrm{D}^2\mathcal{K}(\psi) = DG(\psi)$ is neg. definite on $E_\varepsilon \cap \{cst\}^\perp$

$\Rightarrow$ We have to show smoothness and strict monotonicity

Convergence

Theorem: [Mérigot, Meyron, T. '17]

Assume μ is regular simplicial measure

$y_1, \cdots, y_N$ are in generic position

Then:

- G has class C^1 on $\mathbb{R}^N$.
- G is strictly monotone

$$\forall \psi \in \mathcal{K}^+, \forall v \in \{cst\} \perp \setminus \{0\}, \quad \langle DG(\psi)v|v\rangle < 0.$$

Convergence

Theorem: [Mérigot, Meyron, T. '17]

Assume μ is regular simplicial measure

$y_1, \cdots, y_N$ are in generic position

Then:

- G has class C^1 on $\mathbb{R}^N$.
- G is strictly monotone

$$\forall \psi \in \mathcal{K}^+, \forall v \in \{cst\} \perp \setminus \{0\}, \quad \langle DG(\psi)v|v\rangle < 0.$$

Convergence

Theorem: [Mérigot, Meyron, T. '17]

Assume μ is regular simplicial measure

$y_1, \cdots, y_N$ are in generic position

Then:

- G has class C^1 on $\mathbb{R}^N$.
- G is strictly monotone

$$\forall \psi \in \mathcal{K}^+, \forall v \in \{cst\} \perp \backslash\{0\}, \quad \langle DG(\psi)v|v\rangle < 0.$$

Corollary: [Mérigot, Meyron, T. '17]

Assume μ is regular simplicial measure

$y_1, \cdots, y_N$ are in generic position

Then the damped Newton algorithm converges with linear rate globally, *i.e.*

$$\|G(\psi_k) - \nu\| \leq (1 - \frac{\tau^*}{2})^k \|G(\psi_0) - \nu\|\}$$

Regular simplicial measure

Definition A *simplex soup* is a finite family Σ of simplices of $\mathbb{R}^d$.

- d_σ: dimension of a simplex σ is denoted
- $K = \cup_{\sigma \in \Sigma} \sigma$: support of Σ

Regular simplicial measure

Definition A *simplex soup* is a finite family Σ of simplices of $\mathbb{R}^d$.

- d_σ: dimension of a simplex σ is denoted
- $K = \cup_{\sigma \in \Sigma} \sigma$: support of Σ

Definition: $\mu = \sum_{\sigma \in \Sigma} \mu_\sigma$ is a *regular simplicial measure* if

- μ_σ has density ρ_σ
- the dimension d is ≥ 2
- $\rho_\sigma : \sigma \to \mathbb{R}$ is continuous and $\min \rho_\sigma > 0$
- $K \setminus S$ is connected, for every S finite.

Regular simplicial measure

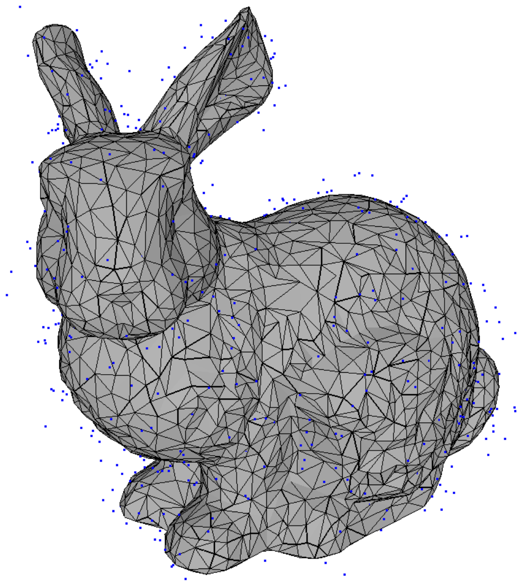
Definition A *simplex soup* is a finite family Σ of simplices of $\mathbb{R}^d$.

- d_σ: dimension of a simplex σ is denoted
- $K = \cup_{\sigma\in\Sigma}\sigma$: support of Σ

Definition: $\mu = \sum_{\sigma\in\Sigma}\mu_\sigma$ is a *regular simplicial measure* if

- μ_σ has density ρ_σ
- the dimension d is ≥ 2
- $\rho_\sigma : \sigma \to \mathbb{R}$ is continuous and $\min\rho_\sigma > 0$
- $K \setminus S$ is connected, for every S finite.

e.g. uniform measure on a connected triangulated surface of $\mathbb{R}^3$.



Genericity condition

Definition: $\{y_1, \cdots, y_N\}$ is in generic position with respect to σ if $\forall p < k \; \forall l \leq \min(d, N-1)$

$$\dim(\mathrm{vect}(y_{i_1} - y_{i_0}, \ldots, y_{i_\ell} - y_{i_0})^{\perp} \cap \mathrm{vect}(x_{j_1} - x_{j_0}, \ldots, x_{j_p} - x_{j_0})) = \max(p - \ell, 0)$$

Genericity condition

Definition: $\{y_1, \cdots, y_N\}$ is in generic position with respect to σ if

$\forall p < k \; \forall l \leq \min(d, N-1)$

$$\dim(\mathrm{vect}(y_{i_1} - y_{i_0}, \ldots, y_{i_\ell} - y_{i_0})^{\perp} \cap \mathrm{vect}(x_{j_1} - x_{j_0}, \ldots, x_{j_p} - x_{j_0})) = \max(p - \ell, 0)$$

Generically

dim $d - l$

dim p

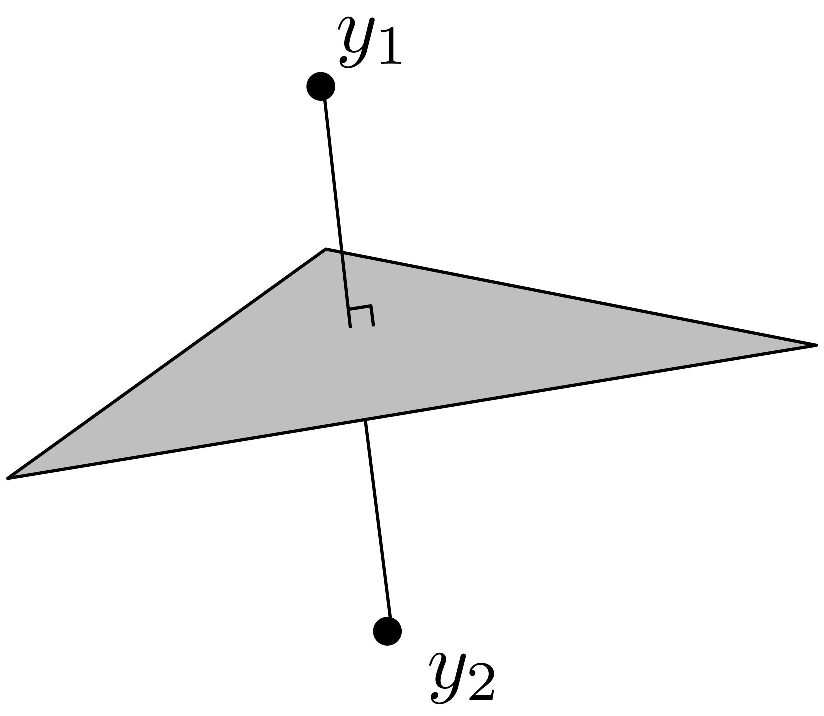
minimum dimension

Genericity condition

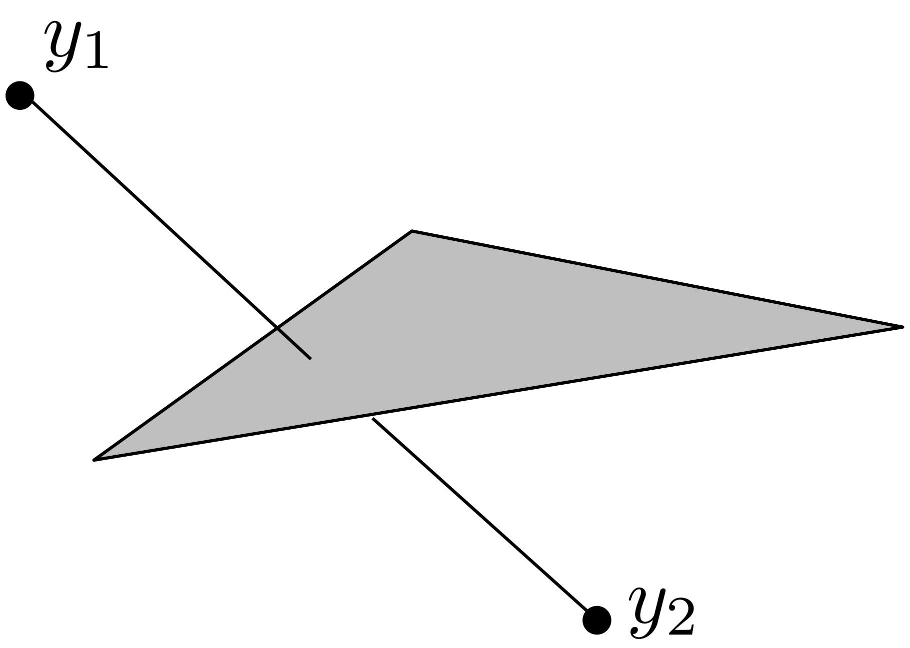
Definition: $\{y_1, \cdots, y_N\}$ is in generic position with respect to σ if
$\forall p < k \; \forall l \leq \min(d, N-1)$

$$\dim(\mathrm{vect}(y_{i_1} - y_{i_0}, \ldots, y_{i_\ell} - y_{i_0})^\perp \cap \mathrm{vect}(x_{j_1} - x_{j_0}, \ldots, x_{j_p} - x_{j_0})) = \max(p - \ell, 0)$$

Generically dim $d - l$ dim p minimum dimension



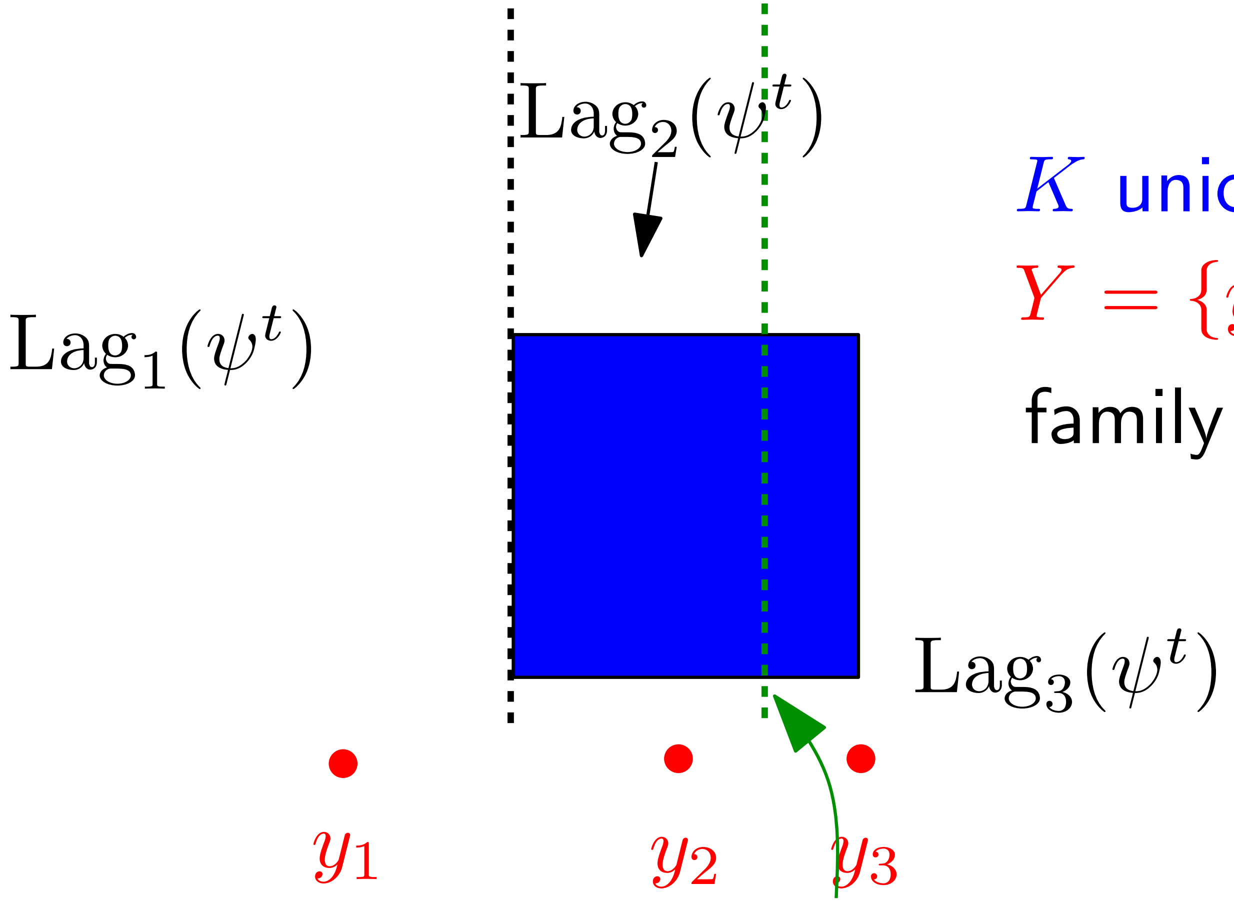
Not generic



Generic

Smoothness of G

Example 1: not a generic case



K union of two triangles

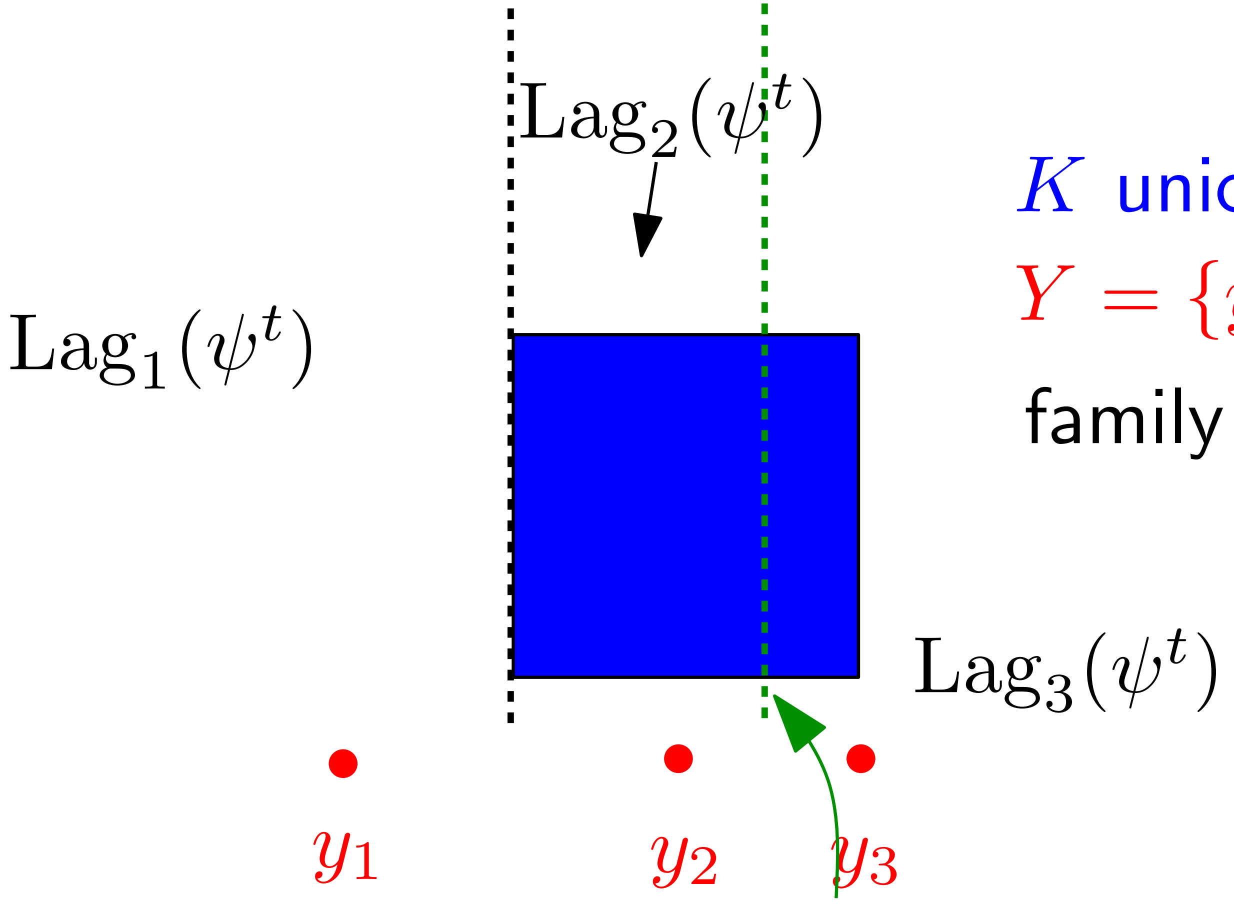
$Y = \{y_1, y_2, y_3\}$

family of weight $\phi^t = (t, 0, 0)$

$$\frac{\partial G_2}{\partial \psi_3}(\psi^t) = \mathcal{H}^1(K \cap Lag_{2,3}(\psi^t))$$

Smoothness of G

Example 1: not a generic case



K union of two triangles

$Y = \{y_1, y_2, y_3\}$

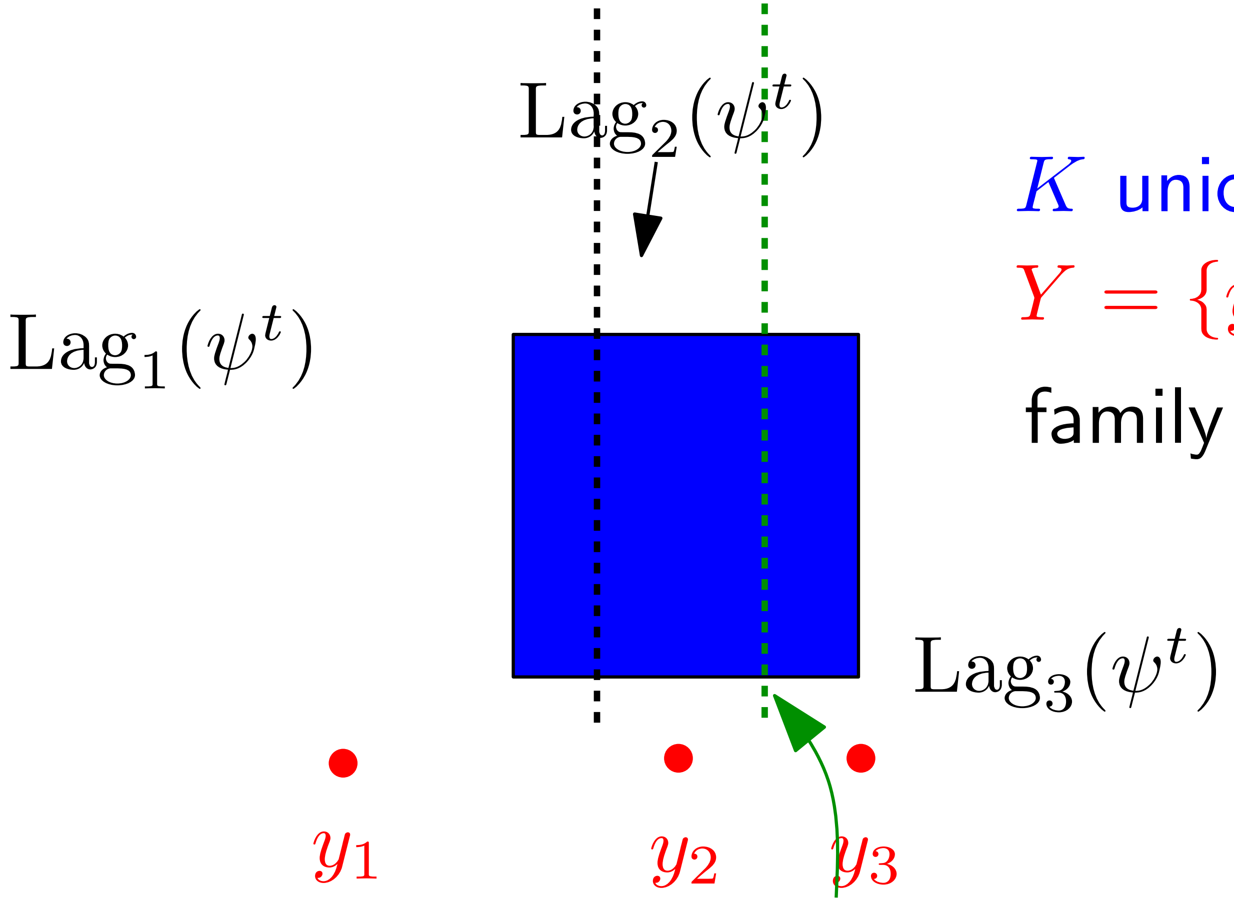
family of weight $\phi^t = (t, 0, 0)$

$$\frac{\partial G_2}{\partial \psi_3}(\psi^t) = \mathcal{H}^1(K \cap Lag_{2,3}(\psi^t))$$

If $t = 0$ $\frac{\partial G_2}{\partial \psi_3}(\psi^t) = 1$

Smoothness of G

Example 1: not a generic case



K union of two triangles

$Y = \{y_1, y_2, y_3\}$

family of weight $\phi^t = (t, 0, 0)$

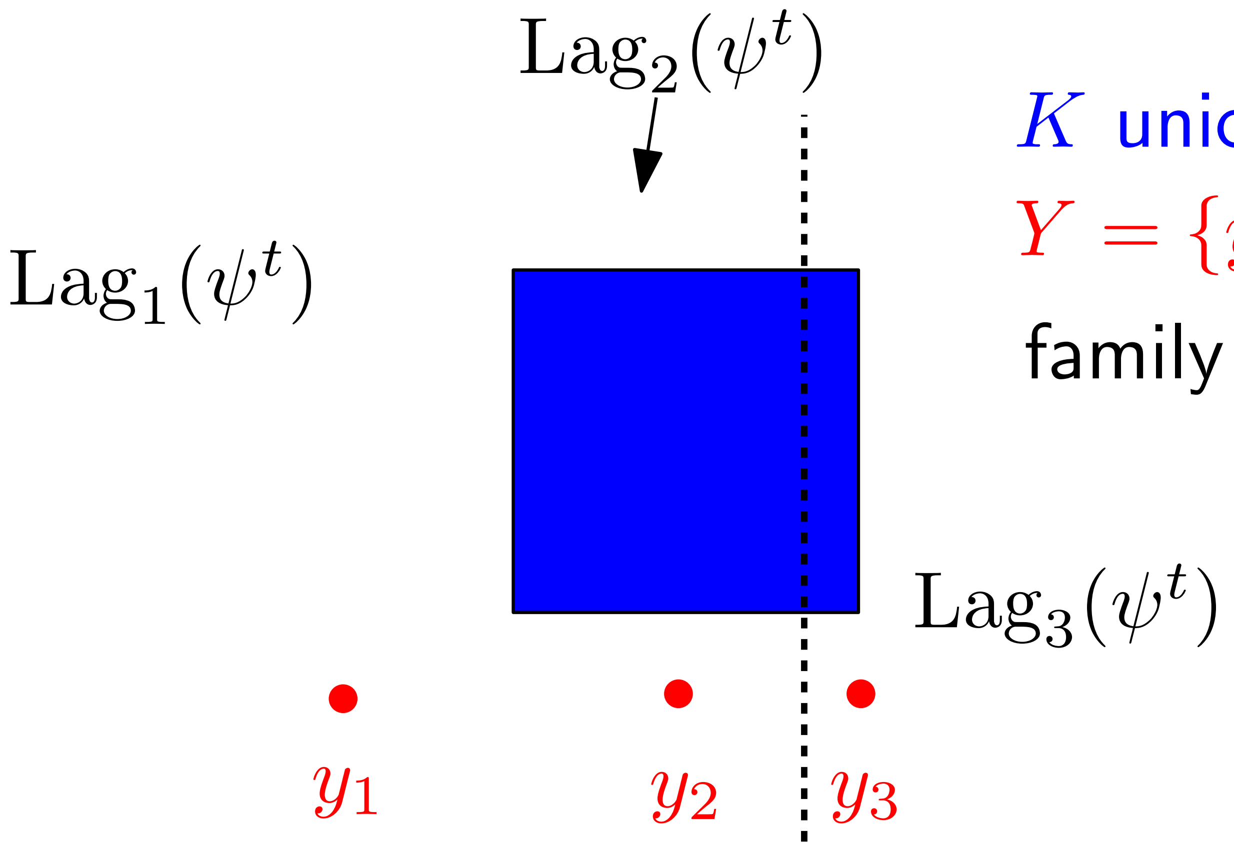
$$\frac{\partial G_2}{\partial \psi_3}(\psi^t) = \mathcal{H}^1(K \cap Lag_{2,3}(\psi^t))$$

If $t = 0$ $\frac{\partial G_2}{\partial \psi_3}(\psi^t) = 1$

If t decreases, $\frac{\partial G_2}{\partial \psi_3}(\psi^t) = 1$

Smoothness of G

Example 1: not a generic case



K union of two triangles

$Y = \{y_1, y_2, y_3\}$

family of weight $\phi^t = (t, 0, 0)$

$$\frac{\partial G_2}{\partial \psi_3}(\psi^t) = \mathcal{H}^1(K \cap Lag_{2,3}(\psi^t))$$

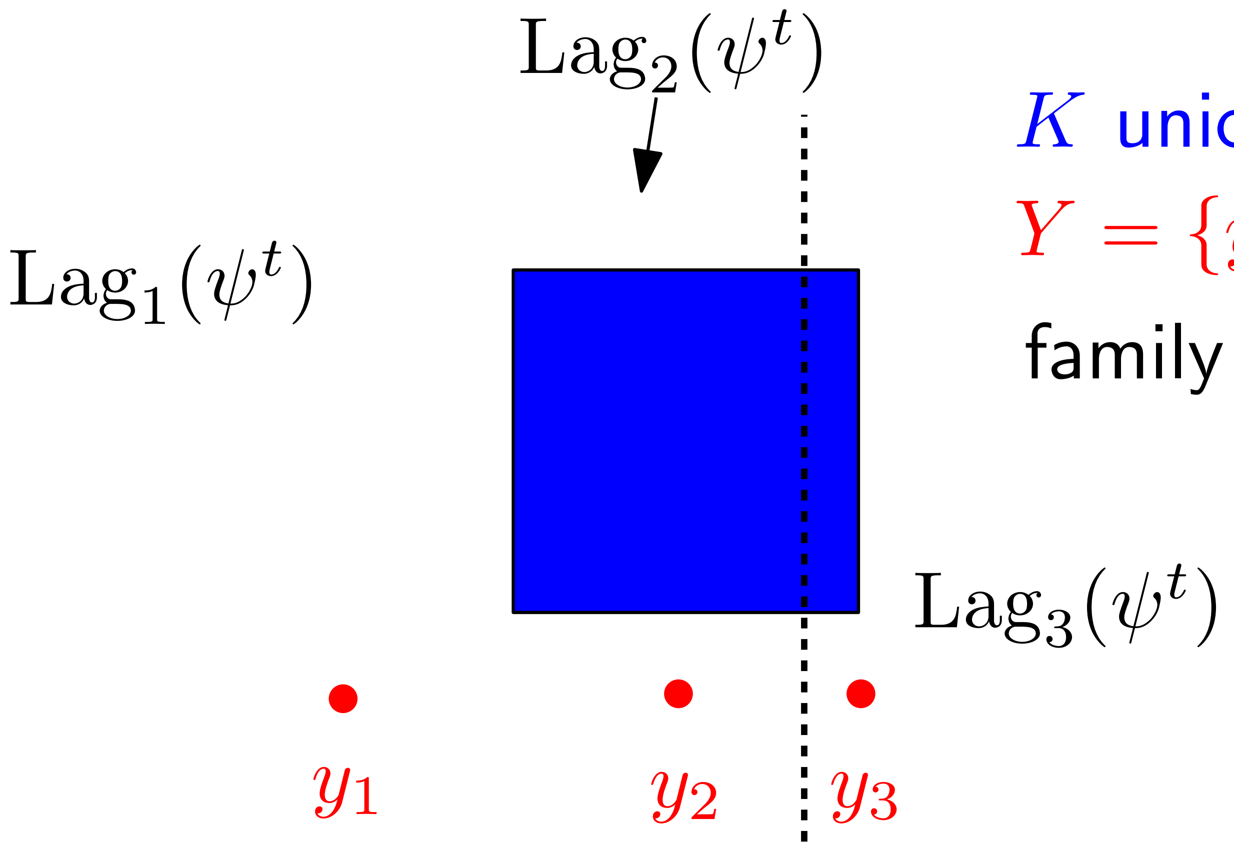
If $t = 0$ $\frac{\partial G_2}{\partial \psi_3}(\psi^t) = 1$

If t decreases, $\frac{\partial G_2}{\partial \psi_3}(\psi^t) = 1$

If t still decreases, suddenly $\frac{\partial G_2}{\partial \psi_3}(\psi^t) = 0$

Smoothness of G

Example 1: not a generic case



K union of two triangles

$Y = \{y_1, y_2, y_3\}$

family of weight $\phi^t = (t, 0, 0)$

$$\frac{\partial G_2}{\partial \psi_3}(\psi^t) = \mathcal{H}^1(K \cap Lag_{2,3}(\psi^t))$$

If $t = 0$ $\frac{\partial G_2}{\partial \psi_3}(\psi^t) = 1$

If t decreases, $\frac{\partial G_2}{\partial \psi_3}(\psi^t) = 1$

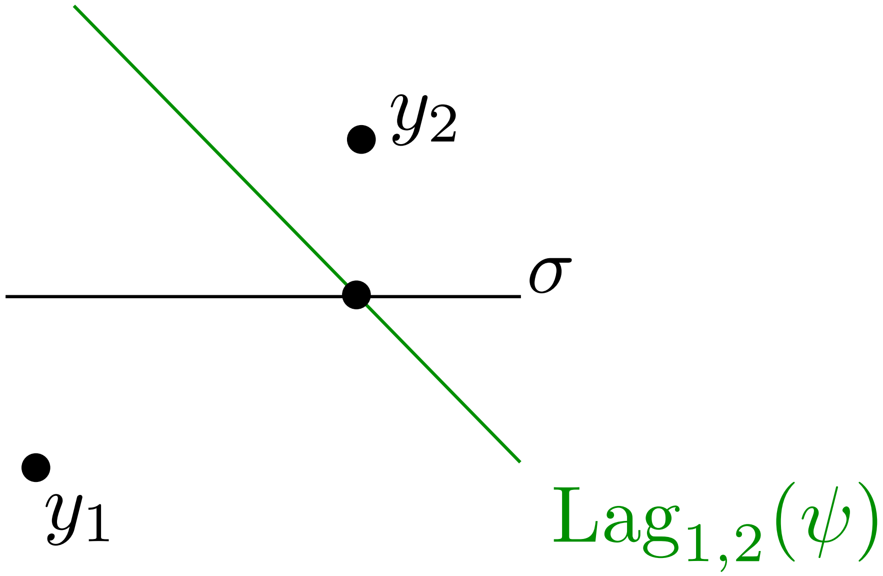
If t still decreases, suddenly $\frac{\partial G_2}{\partial \psi_3}(\psi^t) = 0$

⇝ G is not continuous ⇝ need genericity

Smoothness of G

Example 2: not a regular measure $(dim(\sigma) = 1)$

σ is a simplex of dim 1

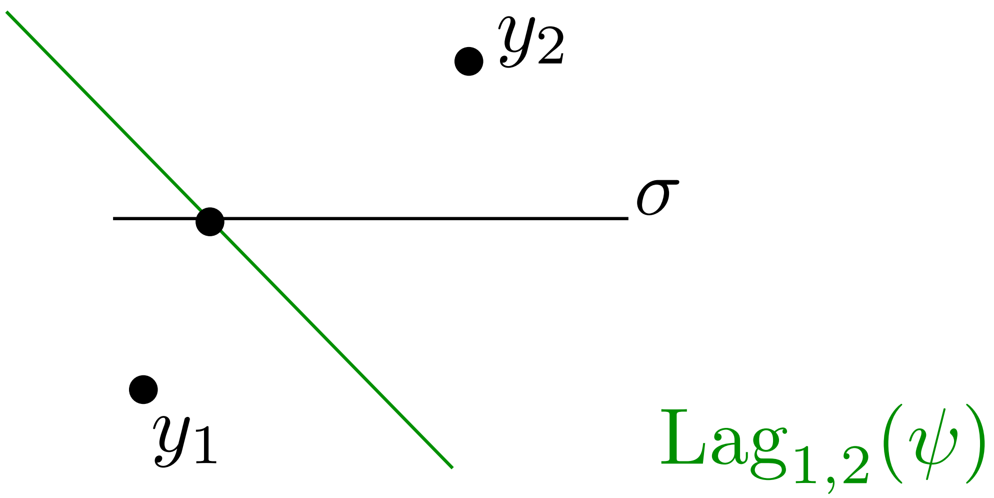


$$\frac{\partial G_1}{\partial \psi_2}(\psi^t) = \mathcal{H}^0(K \cap Lag_{1,2}(\psi^t)) = 1$$

Smoothness of G

Example 2: not a regular measure $(dim(\sigma) = 1)$

σ is a simplex of dim 1

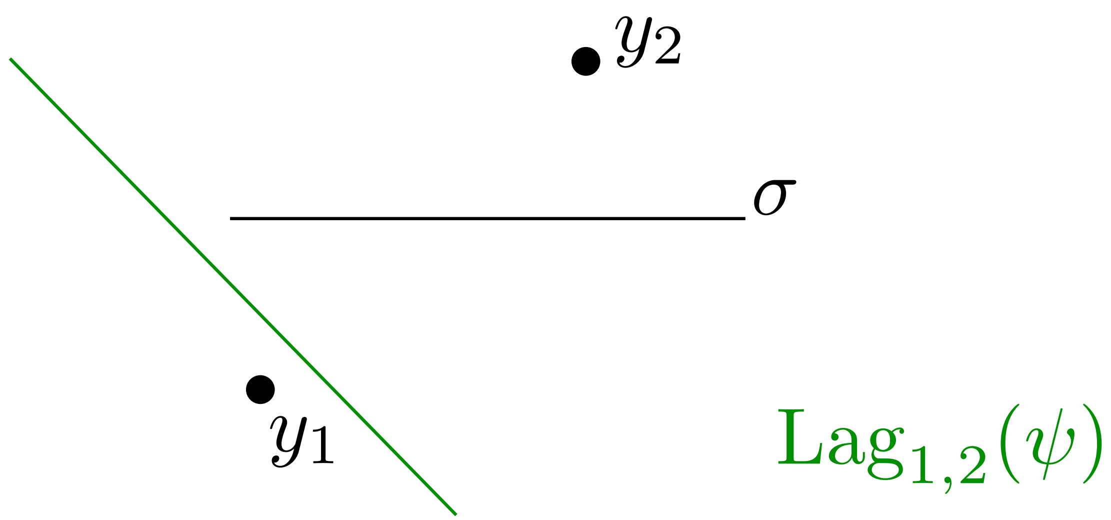


$$\frac{\partial G_1}{\partial \psi_2}(\psi^t) = \mathcal{H}^0(K \cap Lag_{1,2}(\psi^t)) \ = 1$$

Smoothness of G

Example 2: not a regular measure $(dim(\sigma) = 1)$

σ is a simplex of dim 1

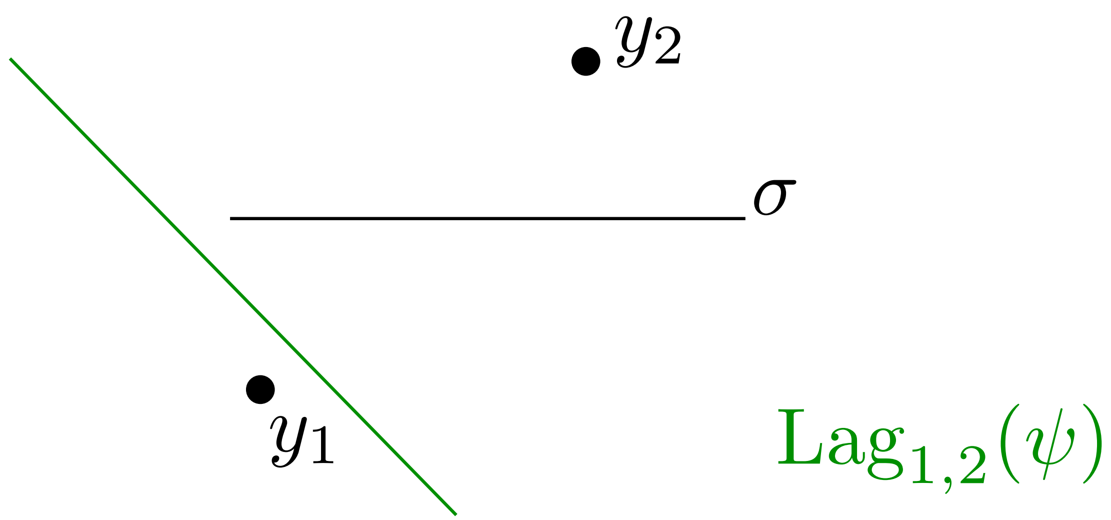


$$\frac{\partial G_1}{\partial \psi_2}(\psi^t) = \mathcal{H}^0(K \cap Lag_{1,2}(\psi^t)) = 0$$

Smoothness of G

Example 2: not a regular measure $(dim(\sigma) = 1)$

σ is a simplex of dim 1



$$\frac{\partial G_1}{\partial \psi_2}(\psi^t) = \mathcal{H}^0(K \cap Lag_{1,2}(\psi^t)) = 0$$

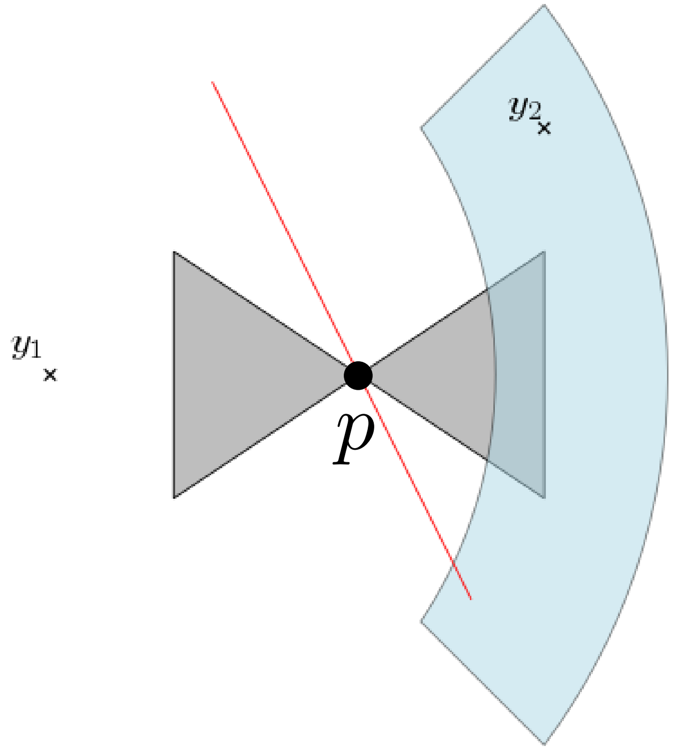
$\rightsquigarrow$ G is not continuous

Strict monotonicity of G

Example 3: K connected, but $K \setminus \{p\}$ not connected.

$Y = \{y_1, y_2\}$

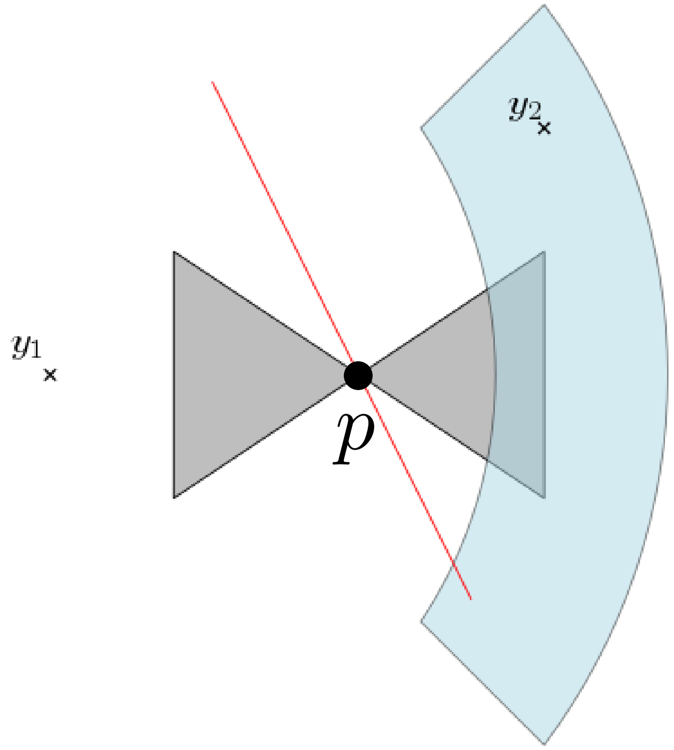
K= union of two triangles



$$\mathrm{D}G(\psi) = \begin{pmatrix} a & -a \\ -a & a \end{pmatrix} \text{ where } a = \frac{1}{2\|y_1 - y_2\|}\mathcal{H}^1(\mathrm{Lag}_{1,2}(\psi) \cap K).$$

For every y_2 in blue domain, there exists ψ_1 and ψ_2 s.t. $DG(\psi) = 0$

Strict monotonicity of G



Example 3: K connected, but $K \setminus \{p\}$ not connected.

$Y = \{y_1, y_2\}$

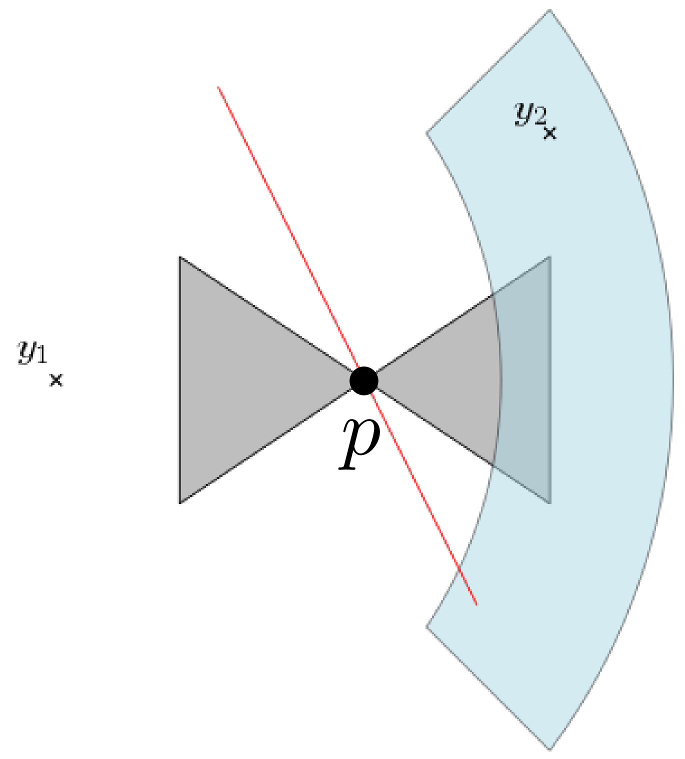
K= union of two triangles

$$\mathrm{D}G(\psi) = \begin{pmatrix} a & -a \\ -a & a \end{pmatrix} \text{ where } a = \frac{1}{2\|y_1 - y_2\|}\mathcal{H}^1(\mathrm{Lag}_{1,2}(\psi) \cap K).$$

For every y_2 in blue domain, there exists ψ_1 and ψ_2 s.t. $DG(\psi) = 0$

⇝ Even generically, $DG(\psi)$ is not strictly monotone

Strict monotonicity of G



Example 3: K connected, but $K \setminus \{p\}$ not connected.

$$Y = \{y_1, y_2\}$$

K= union of two triangles

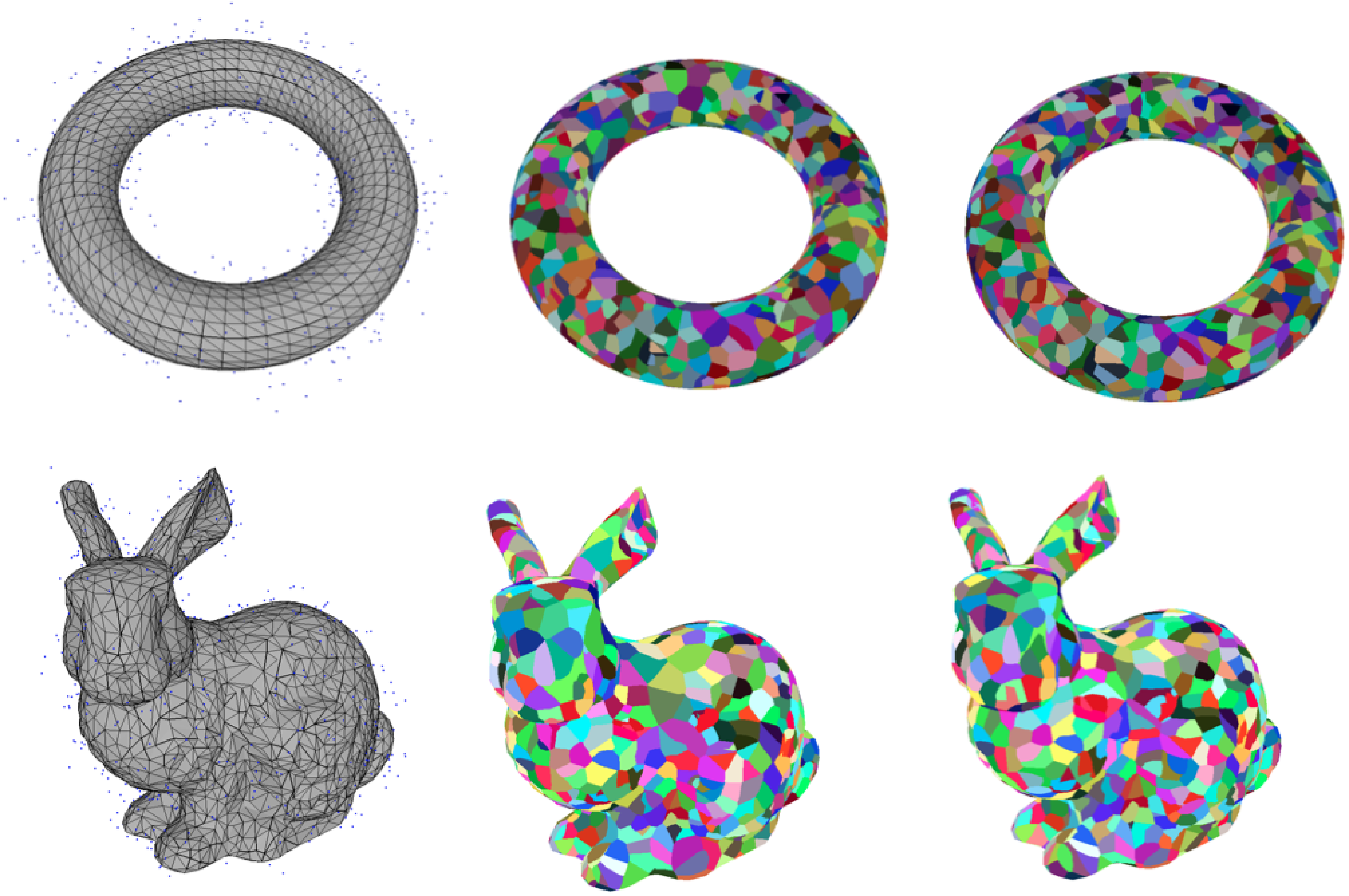
$$\mathrm{D}G(\psi) = \begin{pmatrix} a & -a \\ -a & a \end{pmatrix} \text{ where } a = \frac{1}{2\|y_1-y_2\|}\mathcal{H}^1(\mathrm{Lag}_{1,2}(\psi)\cap K).$$

For every y_2 in blue domain, there exists ψ_1 and ψ_2 s.t. $DG(\psi) = 0$

$\rightsquigarrow$ Even generically, $DG(\psi)$ is not strictly monotone

$\rightsquigarrow$ we need this connectedness condition.

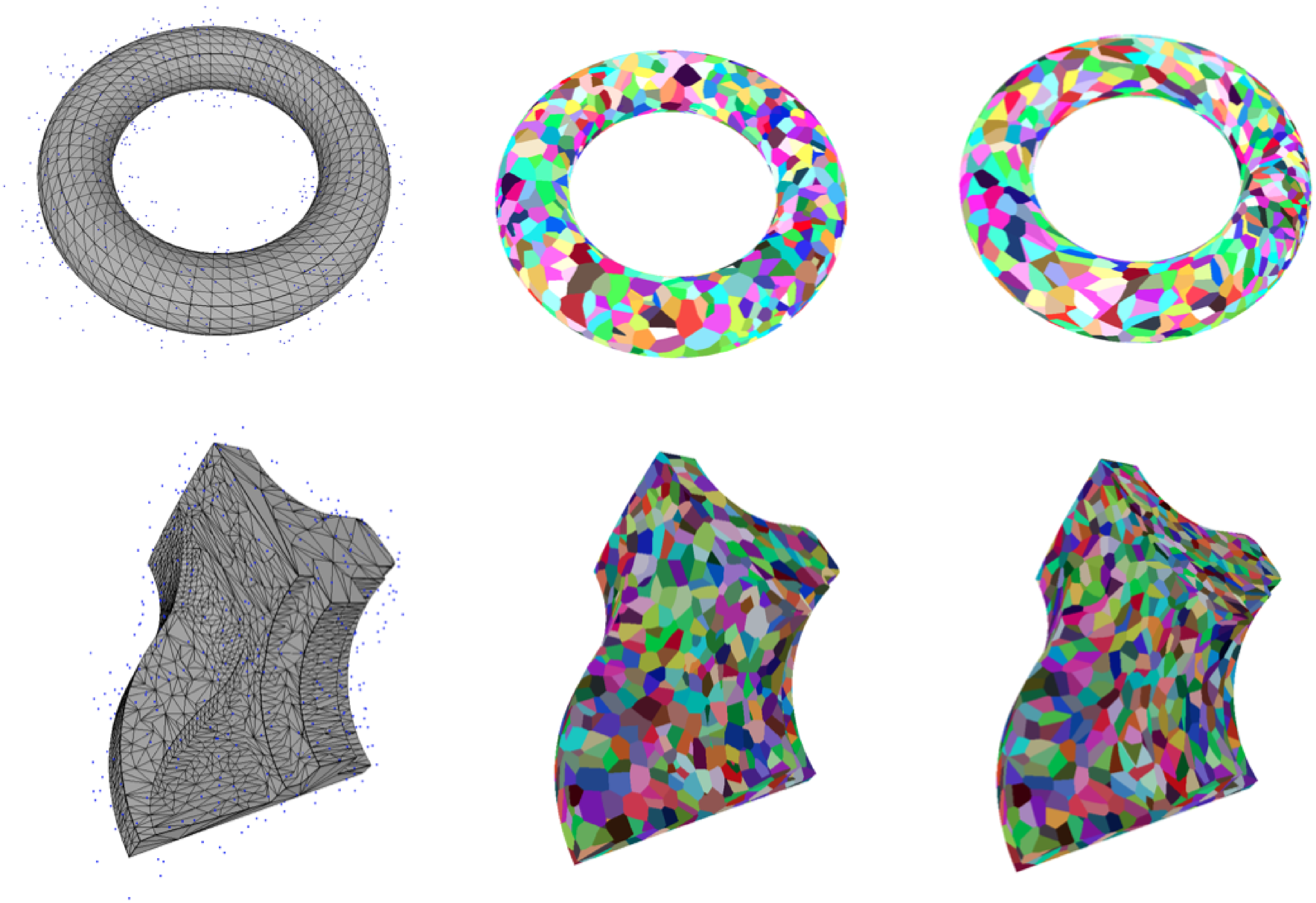
Application



Uniform measure

$N = 1000$, $< 60s$, less than 9 iterations, error $< 10^{-6}$.

Application



Target measure not uniform (decreases from left to right)

$N = 1000$, $< 60s$, less than 9 iterations, error $< 10^{-6}$.

Conclusion

A damped Newton algorithm can be used to solve large geometric instances of optimal transport.

- For cost satisfying MTW and source measure with density
- For measure supported on sets with codimension and quadratic cost.

Conclusion

A damped Newton algorithm can be used to solve large geometric instances of optimal transport.

- For cost satisfying MTW and source measure with density
- For measure supported on sets with codimension and quadratic cost.

$\rightsquigarrow$ Generalization to generated jacobian equations (application to optics, near field target)

$\rightsquigarrow$ Applications to optimal transport beween measures supported on graphs.

Conclusion

A damped Newton algorithm can be used to solve large geometric instances of optimal transport.

- For cost satisfying MTW and source measure with density
- For measure supported on sets with codimension and quadratic cost.

$\rightsquigarrow$ Generalization to generated jacobian equations (application to optics, near field target)

$\rightsquigarrow$ Applications to optimal transport beween measures supported on graphs.

Looking for post-docs (French ANR project MAGA)

Thank you!