Optimal transport and its (semi-)discretization

Quentin Mérigot
Université Paris-Saclay

Finite volume and optimal transport, LMO, 19-21 Novembre 2024

Outline

Lagrangian Formulation of (Some) PDEs

- We follow the evolution of a population of particles living in $\mathbb{R}^d$, described by its initial distribution $\rho_0 \in \mathcal{P}(\Omega)$ and by its **displacement**

$$X : [0, T] \to \mathrm{L}^2(\rho_0, \mathbb{R}^d).$$

Lagrangian Formulation of (Some) PDEs

- We follow the evolution of a population of particles living in $\mathbb{R}^d$, described by its initial distribution $\rho_0 \in \mathcal{P}(\Omega)$ and by its **displacement**

$$X : [0, T] \to \mathrm{L}^2(\rho_0, \mathbb{R}^d).$$

- $\rho_t =$ distribution of particles $= X_{t\#}\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$.

Lagrangian Formulation of (Some) PDEs

- We follow the evolution of a population of particles living in $\mathbb{R}^d$, described by its initial distribution $\rho_0 \in \mathcal{P}(\Omega)$ and by its **displacement**

$$X : [0, T] \to \mathrm{L}^2(\rho_0, \mathbb{R}^d).$$

- $\rho_t =$ distribution of particles $= X_{t\#}\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$.
- The displacement follows $\boxed{\dot{X} \in -\partial E(X) \text{ or } \ddot{X} \in -\partial E(X).}$

Lagrangian Formulation of (Some) PDEs

- We follow the evolution of a population of particles living in $\mathbb{R}^d$, described by its initial distribution $\rho_0 \in \mathcal{P}(\Omega)$ and by its **displacement**

$$X : [0, T] \to \mathrm{L}^2(\rho_0, \mathbb{R}^d).$$

- $\rho_t =$ distribution of particles $= X_{t\#}\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$.
- The displacement follows $\dot{X} \in -\partial E(X)$ or $\ddot{X} \in -\partial E(X)$.
- **Main assumption:** The energy E only depends on the distribution of particles.

i.e. $E(X) = \mathcal{E}(X\#\rho_0)$ with $\mathcal{E} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}$.

Lagrangian Formulation of (Some) PDEs

- We follow the evolution of a population of particles living in $\mathbb{R}^d$, described by its initial distribution $\rho_0 \in \mathcal{P}(\Omega)$ and by its **displacement**

$$X : [0, T] \to \mathrm{L}^2(\rho_0, \mathbb{R}^d).$$

- ρ_t = distribution of particles $= X_{t\#}\rho_0 \in \mathcal{P}_2(\mathbb{R}^d)$.
- The displacement follows $\dot{X} \in -\partial E(X)$ or $\ddot{X} \in -\partial E(X)$.
- **Main assumption:** The energy E only depends on the distribution of particles.

$$\text{i.e. } E(X) = \mathcal{E}(X\#\rho_0) \text{ with } \mathcal{E} \colon \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R} \cup \{+\infty\}.$$

- Example of energies:

$$\mathcal{E}_{\mathrm{ent}}(\rho) = \begin{cases} \int \rho \log \rho & \text{if } \rho \in \mathcal{P}_2^{\mathrm{ac}}(\Omega), \\ +\infty & \text{otherwise.} \end{cases} \qquad \mathcal{E}_{\mathrm{inc}}(\rho) = \begin{cases} 0 & \text{if } \rho = \mathrm{Leb}_\Omega, \\ +\infty & \text{otherwise.} \end{cases}$$

$$\mathcal{E}_{\mathrm{cong}}(\rho) = \begin{cases} 0 & \text{if } \rho \leq \mathrm{Leb}_\Omega, \\ +\infty & \text{otherwise.} \end{cases} \qquad \mathcal{E}_{\mathrm{pot}}(\rho) = \int V \, d\rho.$$

Lagrangian-Eulerian dictionary

	$\dot{X} \in -\partial E(X)$	$\ddot{X} \in -\partial E(X)$
$\mathcal{E}_{\mathrm{inc}}$		incompressible Euler equation
$\mathcal{E}_{\mathrm{cong}} + \mathcal{E}_{\mathrm{pot}}$	crowd motion	pressureless Euler equation
$\mathcal{E}_{\mathrm{ent}} + \mathcal{E}_{\mathrm{pot}}$	linear Fokker-Planck equation	isentropic Euler equation

Lagrangian-Eulerian dictionary

	$\dot{X} \in -\partial E(X)$	$\ddot{X} \in -\partial E(X)$
$\mathcal{E}_{\mathrm{inc}}$		incompressible Euler equation
$\mathcal{E}_{\mathrm{cong}} + \mathcal{E}_{\mathrm{pot}}$	crowd motion	pressureless Euler equation
$\mathcal{E}_{\mathrm{ent}} + \mathcal{E}_{\mathrm{pot}}$	linear Fokker-Planck equation	isentropic Euler equation

- The Euler side was started by [Arnold '66] and [Brenier 87], while the the gradient flow side was initiated by [Otto '99]–[Jordan-Kinderlehrer-Otto '98].

Lagrangian-Eulerian dictionary

	$\dot{X} \in -\partial E(X)$	$\ddot{X} \in -\partial E(X)$
$\mathcal{E}_{\mathrm{inc}}$		incompressible Euler equation
$\mathcal{E}_{\mathrm{cong}} + \mathcal{E}_{\mathrm{pot}}$	crowd motion	pressureless Euler equation
$\mathcal{E}_{\mathrm{ent}} + \mathcal{E}_{\mathrm{pot}}$	linear Fokker-Planck equation	isentropic Euler equation

- The Euler side was started by [Arnold '66] and [Brenier 87], while the the gradient flow side was initiated by [Otto '99]–[Jordan-Kinderlehrer-Otto '98].
- *Many* mathematicians have studied connections between OT and evolution PDEs.

Lagrangian-Eulerian dictionary

	$\dot{X} \in -\partial E(X)$	$\ddot{X} \in -\partial E(X)$
$\mathcal{E}_{\mathrm{inc}}$		incompressible Euler equation
$\mathcal{E}_{\mathrm{cong}} + \mathcal{E}_{\mathrm{pot}}$	crowd motion	pressureless Euler equation
$\mathcal{E}_{\mathrm{ent}} + \mathcal{E}_{\mathrm{pot}}$	linear Fokker-Planck equation	isentropic Euler equation

- The Euler side was started by [Arnold '66] and [Brenier 87], while the the gradient flow side was initiated by [Otto '99]–[Jordan-Kinderlehrer-Otto '98].
- *Many* mathematicians have studied connections between OT and evolution PDEs.
- The energy E is typically non-convex, with values in $\mathbb{R} \cup \{+\infty\}$: in this introduction we will remain at a very formal level.

Fokker-Planck Case

- We consider $\mathcal{E}(\rho) = \int f(\rho) + \rho V$ and $E(X) = \mathcal{E}(X\#\rho_0)$ with $X \in L^2(\rho_0, \mathbb{R}^d)$

Fokker-Planck Case

- We consider $\mathcal{E}(\rho) = \int f(\rho) + \rho V$ and $E(X) = \mathcal{E}(X\#\rho_0)$ with $X \in L^2(\rho_0, \mathbb{R}^d)$
- **What is the gradient of E ?** Consider a curve X_t satisfying $\dot{X}_t = v_t(X_t)$ and its image density $\rho_t := X_{t\#}\rho_0$, which satisfies $\dot{\rho}_t + \mathrm{div}(\rho_t v_t) = 0$.

Fokker-Planck Case

- We consider $\mathcal{E}(\rho) = \int f(\rho) + \rho V$ and $E(X) = \mathcal{E}(X\#\rho_0)$ with $X \in L^2(\rho_0, \mathbb{R}^d)$
- **What is the gradient of E ?** Consider a curve X_t satisfying $\dot{X}_t = v_t(X_t)$ and its image density $\rho_t := X_{t\#}\rho_0$, which satisfies $\dot{\rho}_t + \mathrm{div}(\rho_t v_t) = 0$. Then,

$$\frac{d}{dt}E(X_t) = \int (f'(\rho_t) + V)\dot{\rho}_t \mathrm{d}x$$

Fokker-Planck Case

- We consider $\mathcal{E}(\rho) = \int f(\rho) + \rho V$ and $E(X) = \mathcal{E}(X\#\rho_0)$ with $X \in L^2(\rho_0, \mathbb{R}^d)$
- **What is the gradient of E ?** Consider a curve X_t satisfying $\dot{X}_t = v_t(X_t)$ and its image density $\rho_t := X_{t\#}\rho_0$, which satisfies $\dot{\rho}_t + \mathrm{div}(\rho_t v_t) = 0$. Then,

$$\frac{d}{dt}E(X_t) = \int (f'(\rho_t) + V)\dot{\rho}_t \mathrm{d}x = -\int (f'(\rho_t) + V)\mathrm{div}(\rho_t v_t)\mathrm{d}x$$

Fokker-Planck Case

- We consider $\mathcal{E}(\rho) = \int f(\rho) + \rho V$ and $E(X) = \mathcal{E}(X\#\rho_0)$ with $X \in L^2(\rho_0, \mathbb{R}^d)$
- **What is the gradient of E ?** Consider a curve X_t satisfying $\dot{X}_t = v_t(X_t)$ and its image density $\rho_t := X_{t\#}\rho_0$, which satisfies $\dot{\rho}_t + \mathrm{div}(\rho_t v_t) = 0$. Then,

$$\frac{d}{dt}E(X_t) = \int (f'(\rho_t) + V)\dot{\rho}_t \mathrm{d}x = -\int (f'(\rho_t) + V)\mathrm{div}(\rho_t v_t)\mathrm{d}x$$
$$= \int \langle \nabla(f'(\rho) + V) | v_t \rangle \, \mathrm{d}\rho_t =$$

Fokker-Planck Case

- We consider $\mathcal{E}(\rho) = \int f(\rho) + \rho V$ and $E(X) = \mathcal{E}(X\#\rho_0)$ with $X \in L^2(\rho_0, \mathbb{R}^d)$
- **What is the gradient of E ?** Consider a curve X_t satisfying $\dot{X}_t = v_t(X_t)$ and its image density $\rho_t := X_{t\#}\rho_0$, which satisfies $\dot{\rho}_t + \mathrm{div}(\rho_t v_t) = 0$. Then,

$$\frac{d}{dt}E(X_t) = \int (f'(\rho_t) + V)\dot{\rho}_t \mathrm{d}x = -\int (f'(\rho_t) + V)\mathrm{div}(\rho_t v_t)\mathrm{d}x$$

$$= \int \langle \nabla(f'(\rho) + V) | v_t \rangle \, \mathrm{d}\rho_t = \int \left\langle \nabla(f'(\rho) + V) \circ X_t | \dot{X}_t \right\rangle \mathrm{d}\rho_0$$

Fokker-Planck Case

- We consider $\mathcal{E}(\rho) = \int f(\rho) + \rho V$ and $E(X) = \mathcal{E}(X\#\rho_0)$ with $X \in L^2(\rho_0, \mathbb{R}^d)$
- **What is the gradient of E ?** Consider a curve X_t satisfying $\dot{X}_t = v_t(X_t)$ and its image density $\rho_t := X_{t\#}\rho_0$, which satisfies $\dot{\rho}_t + \mathrm{div}(\rho_t v_t) = 0$. Then,

$$\frac{d}{dt}E(X_t) = \int (f'(\rho_t) + V)\dot{\rho}_t \mathrm{d}x = -\int (f'(\rho_t) + V)\mathrm{div}(\rho_t v_t)\mathrm{d}x$$
$$= \int \langle \nabla(f'(\rho) + V) | v_t \rangle \, \mathrm{d}\rho_t = \int \left\langle \nabla(f'(\rho) + V) \circ X_t | \dot{X}_t \right\rangle \mathrm{d}\rho_0$$

Identifying terms, we find $\boxed{\nabla E(X) = \nabla(f'(\rho) + V) \circ X}$

Fokker-Planck Case

- We consider $\mathcal{E}(\rho) = \int f(\rho) + \rho V$ and $E(X) = \mathcal{E}(X\#\rho_0)$ with $X \in L^2(\rho_0, \mathbb{R}^d)$
- **What is the gradient of E ?** Consider a curve X_t satisfying $\dot{X}_t = v_t(X_t)$ and its image density $\rho_t := X_{t\#}\rho_0$, which satisfies $\dot{\rho}_t + \mathrm{div}(\rho_t v_t) = 0$. Then,

$$\frac{d}{dt}E(X_t) = \int (f'(\rho_t) + V)\dot{\rho}_t \mathrm{d}x = -\int (f'(\rho_t) + V)\mathrm{div}(\rho_t v_t)\mathrm{d}x$$
$$= \int \langle \nabla(f'(\rho) + V) | v_t \rangle \,\mathrm{d}\rho_t = \int \left\langle \nabla(f'(\rho) + V) \circ X_t | \dot{X}_t \right\rangle \mathrm{d}\rho_0$$

  Identifying terms, we find $\boxed{\nabla E(X) = \nabla(f'(\rho) + V) \circ X}$

- The equation $\dot{X}_t = -\nabla E(X_t)$ thus leads to

$$\boxed{\begin{cases} \dot{X}_t = v_t(X_t) \\ \rho_t = X_{t\#}\rho_0 \\ v_t = -(\nabla f'(\rho_t) + \nabla V) \end{cases}} \qquad \textit{(Lagrangian)}$$

Fokker-Planck Case

- We consider $\mathcal{E}(\rho) = \int f(\rho) + \rho V$ and $E(X) = \mathcal{E}(X\#\rho_0)$ with $X \in L^2(\rho_0, \mathbb{R}^d)$
- **What is the gradient of E ?** Consider a curve X_t satisfying $\dot{X}_t = v_t(X_t)$ and its image density $\rho_t := X_{t\#}\rho_0$, which satisfies $\dot{\rho}_t + \mathrm{div}(\rho_t v_t) = 0$. Then,

$$\frac{d}{dt}E(X_t) = \int (f'(\rho_t) + V)\dot{\rho}_t \mathrm{d}x = -\int (f'(\rho_t) + V)\mathrm{div}(\rho_t v_t)\mathrm{d}x$$

$$= \int \langle \nabla(f'(\rho) + V) | v_t \rangle \, \mathrm{d}\rho_t = \int \left\langle \nabla(f'(\rho) + V) \circ X_t | \dot{X}_t \right\rangle \mathrm{d}\rho_0$$

Identifying terms, we find $\boxed{\nabla E(X) = \nabla(f'(\rho) + V) \circ X}$

- The equation $\dot{X}_t = -\nabla E(X_t)$ thus leads to

$$\boxed{\begin{cases} \dot{X}_t = v_t(X_t) \\ \rho_t = X_{t\#}\rho_0 \\ v_t = -(\nabla f'(\rho_t) + \nabla V) \end{cases}} \qquad \textit{(Lagrangian)}$$

$$\boxed{\dot{\rho}_t = \mathrm{div}[\rho_t(\nabla f'(\rho_t) + \nabla V))} \qquad \textit{(Eulerian)}$$

Fokker-Planck Case

- We consider $\mathcal{E}(\rho) = \int f(\rho) + \rho V$ and $E(X) = \mathcal{E}(X\#\rho_0)$ with $X \in L^2(\rho_0, \mathbb{R}^d)$
- **What is the gradient of E ?** Consider a curve X_t satisfying $\dot{X}_t = v_t(X_t)$ and its image density $\rho_t := X_{t\#}\rho_0$, which satisfies $\dot{\rho}_t + \mathrm{div}(\rho_t v_t) = 0$. Then,

$$\frac{d}{dt}E(X_t) = \int (f'(\rho_t) + V)\dot{\rho}_t \mathrm{d}x = -\int (f'(\rho_t) + V)\mathrm{div}(\rho_t v_t)\mathrm{d}x$$
$$= \int \langle \nabla(f'(\rho) + V) | v_t \rangle \, \mathrm{d}\rho_t = \int \left\langle \nabla(f'(\rho) + V) \circ X_t | \dot{X}_t \right\rangle \mathrm{d}\rho_0$$

  Identifying terms, we find $\boxed{\nabla E(X) = \nabla(f'(\rho) + V) \circ X}$

- The equation $\dot{X}_t = -\nabla E(X_t)$ thus leads to

$$\boxed{\begin{cases} \dot{X}_t = v_t(X_t) \\ \rho_t = X_{t\#}\rho_0 \\ v_t = -(\nabla f'(\rho_t) + \nabla V) \end{cases}} \qquad \textit{(Lagrangian)}$$

$$\boxed{\dot{\rho}_t = \mathrm{div}[\rho_t(\nabla f'(\rho_t) + \nabla V))} \qquad \textit{(Eulerian)}$$

Example: $f(r) = r \log r \longrightarrow \dot{\rho}_t = \Delta\rho_t + \mathrm{div}(\rho_t \nabla V)$

Fokker-Planck Case

- We consider $\mathcal{E}(\rho) = \int f(\rho) + \rho V$ and $E(X) = \mathcal{E}(X \# \rho_0)$ with $X \in L^2(\rho_0, \mathbb{R}^d)$
- **What is the gradient of E ?** Consider a curve X_t satisfying $\dot{X}_t = v_t(X_t)$ and its image density $\rho_t := X_{t\#}\rho_0$, which satisfies $\dot{\rho}_t + \mathrm{div}(\rho_t v_t) = 0$. Then,

$$\frac{d}{dt} E(X_t) = \int (f'(\rho_t) + V)\dot{\rho}_t \mathrm{d}x = -\int (f'(\rho_t) + V)\mathrm{div}(\rho_t v_t)\mathrm{d}x$$

$$= \int \langle \nabla (f'(\rho) + V) | v_t \rangle \, \mathrm{d}\rho_t = \int \left\langle \nabla (f'(\rho) + V) \circ X_t | \dot{X}_t \right\rangle \mathrm{d}\rho_0$$

  Identifying terms, we find $\boxed{\nabla E(X) = \nabla (f'(\rho) + V) \circ X}$

- The equation $\dot{X}_t = -\nabla E(X_t)$ thus leads to

$$\boxed{\begin{cases} \dot{X}_t = v_t(X_t) \\ \rho_t = X_{t\#}\rho_0 \\ v_t = -(\nabla f'(\rho_t) + \nabla V) \end{cases}} \qquad \textit{(Lagrangian)}$$

$$\boxed{\dot{\rho}_t = \mathrm{div}[\rho_t(\nabla f'(\rho_t) + \nabla V))} \qquad \textit{(Eulerian)}$$

  Example: $f(r) = r \log r \longrightarrow \dot{\rho}_t = \Delta \rho_t + \mathrm{div}(\rho_t \nabla V)$

- **Regular** lagrangian trajectory associated to heat flow ($\neq$ Brownian motion!)

From Lagrangian dynamic to the JKO scheme

- **Minimizing movement scheme:** for $\tau > 0$, define a time-discretization by

$$X_\tau^{k+1} \in \arg\min_X \frac{1}{2\tau} \left\| X_\tau^k - X \right\|_{L^2(\rho_0)}^2 + E(X)$$

From Lagrangian dynamic to the JKO scheme

- **Minimizing movement scheme:** for $\tau > 0$, define a time-discretization by

$$X_\tau^{k+1} \in \arg\min_X \frac{1}{2\tau} \left\| X_\tau^k - X \right\|^2_{\mathrm{L}^2(\rho_0)} + E(X)$$

- The minimum can be rewritten as :

$$\min_\rho \frac{1}{2\tau} \left(\min_{X \text{ s.t. } X_{\#\rho_0} = \rho} \left\| X_\tau^k - X \right\|^2_{\mathrm{L}^2(\rho_0)} \right) + \mathcal{E}(\rho)$$

From Lagrangian dynamic to the JKO scheme

- **Minimizing movement scheme:** for $\tau > 0$, define a time-discretization by

$$X_\tau^{k+1} \in \arg\min_X \frac{1}{2\tau} \left\| X_\tau^k - X \right\|^2_{\mathrm{L}^2(\rho_0)} + E(X)$$

- The minimum can be rewritten as :

$$\min_\rho \frac{1}{2\tau} \left(\min_{X \text{ s.t. } X_{\#\rho_0} = \rho} \left\| X_\tau^k - X \right\|^2_{\mathrm{L}^2(\rho_0)} \right) + \mathcal{E}(\rho)$$

Setting $\rho_\tau^k = X^k_{\tau\#}\rho_0$, the inner minimum is a quadratic optimal transport cost (a consequence of Brenier's polar factorization theorem):

$$\min_{X \text{ s.t. } X_{\#\rho_0} = \rho} \left\| X_\tau^k - X \right\|^2_{\mathrm{L}^2(\rho_0)} = \min_{X \text{ s.t. } X_{\#\rho_0} = \rho} \left\| \mathrm{Id} - X \circ (X_\tau^k)^{-1} \right\|^2_{\mathrm{L}^2(\rho_\tau^k)}$$

From Lagrangian dynamic to the JKO scheme

- **Minimizing movement scheme:** for $\tau > 0$, define a time-discretization by

$$X_\tau^{k+1} \in \arg\min_X \frac{1}{2\tau} \left\| X_\tau^k - X \right\|^2_{\mathrm{L}^2(\rho_0)} + E(X)$$

- The minimum can be rewritten as :

$$\min_\rho \frac{1}{2\tau} \left(\min_{X \text{ s.t. } X_{\#\rho_0} = \rho} \left\| X_\tau^k - X \right\|^2_{\mathrm{L}^2(\rho_0)} \right) + \mathcal{E}(\rho)$$

Setting $\rho_\tau^k = X^k_{\tau\#}\rho_0$, the inner minimum is a quadratic optimal transport cost (a consequence of Brenier's polar factorization theorem):

$$\min_{X \text{ s.t. } X_{\#\rho_0} = \rho} \left\| X_\tau^k - X \right\|^2_{\mathrm{L}^2(\rho_0)} = \min_{X \text{ s.t. } X_{\#\rho_0} = \rho} \left\| \mathrm{Id} - X \circ (X_\tau^k)^{-1} \right\|^2_{\mathrm{L}^2(\rho_\tau^k)}$$

$$= \min_{T \text{ s.t. } T_{\#}\rho_\tau^k = \rho} \left\| \mathrm{Id} - T \right\|^2_{\mathrm{L}^2(\rho^k)} =: \mathrm{W}_2^2(\rho_\tau^k, \rho)$$

From Lagrangian dynamic to the JKO scheme

- **Minimizing movement scheme:** for $\tau > 0$, define a time-discretization by

$$X_\tau^{k+1} \in \arg\min_X \frac{1}{2\tau} \left\| X_\tau^k - X \right\|^2_{\mathrm{L}^2(\rho_0)} + E(X)$$

- The minimum can be rewritten as :

$$\min_\rho \frac{1}{2\tau} \left(\min_{X \text{ s.t. } X_{\#\rho_0} = \rho} \left\| X_\tau^k - X \right\|^2_{\mathrm{L}^2(\rho_0)} \right) + \mathcal{E}(\rho)$$

Setting $\rho_\tau^k = X^k_{\tau\#}\rho_0$, the inner minimum is a quadratic optimal transport cost (a consequence of Brenier's polar factorization theorem):

$$\min_{X \text{ s.t. } X_{\#\rho_0} = \rho} \left\| X_\tau^k - X \right\|^2_{\mathrm{L}^2(\rho_0)} = \min_{X \text{ s.t. } X_{\#\rho_0} = \rho} \left\| \mathrm{Id} - X \circ (X_\tau^k)^{-1} \right\|^2_{\mathrm{L}^2(\rho_\tau^k)}$$

$$= \min_{T \text{ s.t. } T_{\#}\rho_\tau^k = \rho} \left\| \mathrm{Id} - T \right\|^2_{\mathrm{L}^2(\rho^k)} =: \mathrm{W}_2^2(\rho_\tau^k, \rho)$$

- We recover the JKO scheme: $\rho_\tau^{k+1} \in \arg\min_\rho \frac{1}{2\tau} \mathrm{W}_2^2(\rho_\tau^k, \rho) + \mathcal{E}(X).$

Outline

Optimal transport

- Let $\mathcal{P}_p(\mathbb{R}^d)$ be the set of probability measures with finite pth moment.

Wasserstein distance between $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$:

$$\mathrm{W}_p(\mu, \nu) = \left(\min_{\gamma \in \Gamma(\mu,\nu)} \int \|x - y\|^p \, \mathrm{d}\gamma(x, y) \right)^{1/p},$$

where $\Gamma(\mu, \nu)$ is the set of *couplings* between μ, ν.

Optimal transport

- Let $\mathcal{P}_p(\mathbb{R}^d)$ be the set of probability measures with finite pth moment.

Wasserstein distance between $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$:

$$\mathrm{W}_p(\mu,\nu) = \left(\min_{\gamma \in \Gamma(\mu,\nu)} \int \|x - y\|^p \, \mathrm{d}\gamma(x,y) \right)^{1/p},$$

where $\Gamma(\mu,\nu)$ is the set of *couplings* between μ, ν.

- Kantorovich-Rubinstein formula: $\boxed{\mathrm{W}_1(\mu,\nu) = \sup_{\mathrm{Lip}(\phi) \le 1} \int \phi \mathrm{d}\mu - \int \phi \mathrm{d}\nu.}$

Optimal transport

- Let $\mathcal{P}_p(\mathbb{R}^d)$ be the set of probability measures with finite pth moment.

Wasserstein distance between $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$:

$$\mathrm{W}_p(\mu,\nu) = \left(\min_{\gamma \in \Gamma(\mu,\nu)} \int \|x-y\|^p \,\mathrm{d}\gamma(x,y) \right)^{1/p},$$

where $\Gamma(\mu,\nu)$ is the set of *couplings* between μ, ν.

- Kantorovich-Rubinstein formula: $\boxed{\mathrm{W}_1(\mu,\nu) = \sup_{\mathrm{Lip}(\phi)\leq 1} \int \phi \mathrm{d}\mu - \int \phi \mathrm{d}\nu.}$

In particular, on $\mathcal{P}(X)$, with X compact, W_p metrizes the **weak$*$ convergence**

Optimal transport

- Let $\mathcal{P}_p(\mathbb{R}^d)$ be the set of probability measures with finite pth moment.

  Wasserstein distance between $\mu, \nu \in \mathcal{P}_p(\mathbb{R}^d)$:

  $$\mathrm{W}_p(\mu,\nu) = \left(\min_{\gamma \in \Gamma(\mu,\nu)} \int \|x - y\|^p \, \mathrm{d}\gamma(x,y) \right)^{1/p},$$

  where $\Gamma(\mu,\nu)$ is the set of *couplings* between μ, ν.

- Kantorovich-Rubinstein formula: $\boxed{\mathrm{W}_1(\mu,\nu) = \sup_{\mathrm{Lip}(\phi)\leq 1} \int \phi \mathrm{d}\mu - \int \phi \mathrm{d}\nu.}$

  In particular, on $\mathcal{P}(X)$, with X compact, W_p metrizes the **weak$*$ convergence**

- Optimal transport and Wasserstein distances W_p have found applications in geometry, functional inequalities, probabilities, PDEs, statistical learning.

1D Wasserstein space and quantiles

- Given $\mu \in \mathcal{P}(\mathbb{R})$, there exists a unique nondecreasing $T_\mu \in \mathrm{L}^1([0,1])$ satisfying $T_{\mu\#}\lambda = \mu$, with λ = Lebesgue measure on $[0,1]$.

1D Wasserstein space and quantiles

- Given $\mu \in \mathcal{P}(\mathbb{R})$, there exists a unique nondecreasing $T_\mu \in \mathrm{L}^1([0,1])$ satisfying $T_{\mu\#}\lambda = \mu$, with λ = Lebesgue measure on $[0,1]$.
- T_μ is the *inverse cumulative distribution function*, also called *quantile function*.

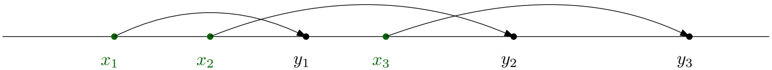
$\mu = \frac{1}{N}\sum_{1 \le i \le N} \delta_{x_i}$ with $x_1 \le \ldots \le x_N \Longrightarrow T_\mu = x_i$ on $\left[\frac{i-1}{N}, \frac{i}{N}\right]$

1D Wasserstein space and quantiles

- Given $\mu \in \mathcal{P}(\mathbb{R})$, there exists a unique nondecreasing $T_\mu \in \mathrm{L}^1([0,1])$ satisfying $T_{\mu\#}\lambda = \mu$, with $\lambda =$ Lebesgue measure on $[0,1]$.
- T_μ is the *inverse cumulative distribution function*, also called *quantile function*.

$$\mu = \tfrac{1}{N}\textstyle\sum_{1\le i\le N}\delta_{x_i} \text{ with } x_1 \le \ldots \le x_N \Longrightarrow T_\mu = x_i \text{ on } \left[\tfrac{i-1}{N}, \tfrac{i}{N}\right]$$
$$\nu = \tfrac{1}{N}\textstyle\sum_{1\le i\le N}\delta_{y_i} \text{ with } y_1 \le \ldots \le y_N \Longrightarrow T_\nu = y_i \text{ on } \left[\tfrac{i-1}{N}, \tfrac{i}{N}\right]$$

$$\mathrm{W}_p^p(\mu,\nu) = \frac{1}{N}\sum_{1\le i\le N}\|x_i - y_i\|^p = \|T_\mu - T_\nu\|^p_{\mathrm{L}^p([0,1])}.$$



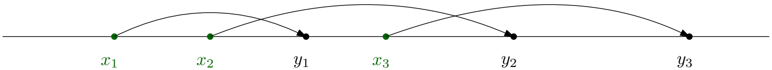
1D Wasserstein space and quantiles

- Given $\mu \in \mathcal{P}(\mathbb{R})$, there exists a unique nondecreasing $T_\mu \in \mathrm{L}^1([0,1])$ satisfying $T_{\mu\#}\lambda = \mu$, with $\lambda =$ Lebesgue measure on $[0,1]$.
- T_μ is the *inverse cumulative distribution function*, also called *quantile function*.

$\mu = \frac{1}{N}\sum_{1\leq i\leq N}\delta_{x_i}$ with $x_1 \leq \ldots \leq x_N \Longrightarrow T_\mu = x_i$ on $\left[\frac{i-1}{N}, \frac{i}{N}\right]$

$\nu = \frac{1}{N}\sum_{1\leq i\leq N}\delta_{y_i}$ with $y_1 \leq \ldots \leq y_N \Longrightarrow T_\nu = y_i$ on $\left[\frac{i-1}{N}, \frac{i}{N}\right]$

$$\mathrm{W}_p^p(\mu,\nu) = \frac{1}{N}\sum_{1\leq i\leq N}\|x_i - y_i\|^p = \|T_\mu - T_\nu\|^p_{\mathrm{L}^p([0,1])}.$$



- The above formula remains true for any probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R})$, i.e.

Theorem: $\mu \mapsto T_\mu$ is an isometry: $\mathrm{W}_p^p(\mu,\nu) = \|T_\mu - T_\nu\|_{\mathrm{L}^2(\lambda)}$.

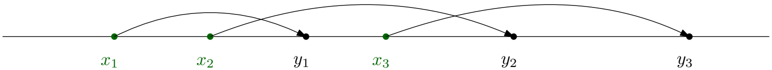
1D Wasserstein space and quantiles

- Given $\mu \in \mathcal{P}(\mathbb{R})$, there exists a unique nondecreasing $T_\mu \in \mathrm{L}^1([0,1])$ satisfying $T_{\mu\#}\lambda = \mu$, with $\lambda =$ Lebesgue measure on $[0,1]$.
- T_μ is the *inverse cumulative distribution function*, also called *quantile function*.

$\mu = \frac{1}{N}\sum_{1\le i\le N}\delta_{x_i}$ with $x_1 \le \ldots \le x_N \Longrightarrow T_\mu = x_i$ on $\left[\frac{i-1}{N}, \frac{i}{N}\right]$

$\nu = \frac{1}{N}\sum_{1\le i\le N}\delta_{y_i}$ with $y_1 \le \ldots \le y_N \Longrightarrow T_\nu = y_i$ on $\left[\frac{i-1}{N}, \frac{i}{N}\right]$

$$\mathrm{W}_p^p(\mu,\nu) = \frac{1}{N}\sum_{1\le i\le N}\|x_i - y_i\|^p = \|T_\mu - T_\nu\|^p_{\mathrm{L}^p([0,1])}.$$



- The above formula remains true for any probability measures $\mu,\nu \in \mathcal{P}(\mathbb{R})$, i.e.

Theorem: $\mu \mapsto T_\mu$ is an isometry: $\mathrm{W}_p^p(\mu,\nu) = \|T_\mu - T_\nu\|_{\mathrm{L}^2(\lambda)}$.

- When μ is a density and $\nu = \sum_{i=1}^N \alpha_i\delta_{y_i}$, the optimal transport plan induces a partition (a mesh?) of $\mathrm{spt}(\mu)$ into intervals $(V_i)_{1\le i\le N}$ with $\mu(V_i) = \alpha_i$.

Wasserstein geodesics

Proposition: Let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$, and let $\gamma \in \Gamma_{\mathrm{opt}}(\mu_0, \mu_1)$. Then, the *curve*

$$\mu_t = \Pi_{t\#}\gamma, \text{ with } \Pi_t(x, y) = (1-t)x + ty$$

is a minimizing geodesic between μ_0 and μ_1.

Wasserstein geodesics

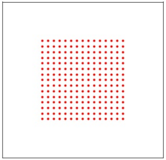
Proposition: Let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$, and let $\gamma \in \Gamma_{\mathrm{opt}}(\mu_0, \mu_1)$. Then, the *curve*

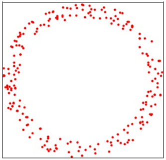
$$\mu_t = \Pi_{t\#}\gamma, \text{ with } \Pi_t(x, y) = (1-t)x + ty$$

is a minimizing geodesic between μ_0 and μ_1.

- If $\mu_0 = \frac{1}{N}\sum_i \delta_{x_i}$, $\mu_1 = \frac{1}{N}\sum_i \delta_{y_i}$ and $\sigma \in \mathfrak{S}_N$ is an optimal bijection, then

$$\boxed{\mu_t = \frac{1}{N}\sum_i \delta_{(1-t)x_i + ty_{\sigma(i)}}}$$





Wasserstein geodesics

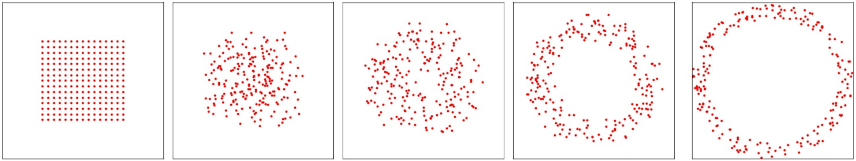
Proposition: Let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$, and let $\gamma \in \Gamma_{\mathrm{opt}}(\mu_0, \mu_1)$. Then, the *curve*

$$\mu_t = \Pi_{t\#}\gamma, \text{ with } \Pi_t(x,y) = (1-t)x + ty$$

is a minimizing geodesic between μ_0 and μ_1.

- If $\mu_0 = \frac{1}{N}\sum_i \delta_{x_i}$, $\mu_1 = \frac{1}{N}\sum_i \delta_{y_i}$ and $\sigma \in \mathfrak{S}_N$ is an optimal bijection, then

$$\mu_t = \frac{1}{N}\sum_i \delta_{(1-t)x_i + ty_{\sigma(i)}}$$



Wasserstein geodesics

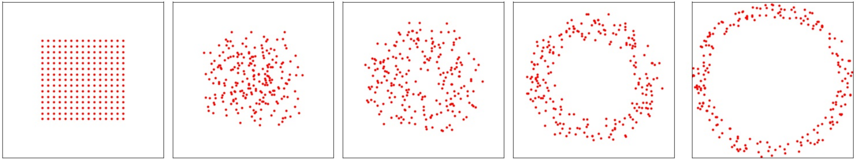
Proposition: Let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$, and let $\gamma \in \Gamma_{\mathrm{opt}}(\mu_0, \mu_1)$. Then, the *curve*

$$\mu_t = \Pi_{t\#}\gamma, \text{ with } \Pi_t(x, y) = (1-t)x + ty$$

is a minimizing geodesic between μ_0 and μ_1.

- If $\mu_0 = \frac{1}{N}\sum_i \delta_{x_i}$, $\mu_1 = \frac{1}{N}\sum_i \delta_{y_i}$ and $\sigma \in \mathfrak{S}_N$ is an optimal bijection, then

$$\mu_t = \frac{1}{N}\sum_i \delta_{(1-t)x_i + ty_{\sigma(i)}}$$



- All minimizing geodesics are of this form.

Wasserstein geodesics

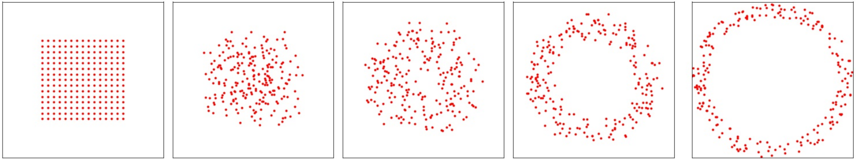
Proposition: Let $\mu_0, \mu_1 \in \mathcal{P}_2(\mathbb{R}^d)$, and let $\gamma \in \Gamma_{\mathrm{opt}}(\mu_0, \mu_1)$. Then, the *curve*

$$\mu_t = \Pi_{t\#}\gamma, \text{ with } \Pi_t(x, y) = (1 - t)x + ty$$

is a minimizing geodesic between μ_0 and μ_1.

- If $\mu_0 = \frac{1}{N}\sum_i \delta_{x_i}$, $\mu_1 = \frac{1}{N}\sum_i \delta_{y_i}$ and $\sigma \in \mathfrak{S}_N$ is an optimal bijection, then

$$\mu_t = \frac{1}{N}\sum_i \delta_{(1-t)x_i + ty_{\sigma(i)}}$$

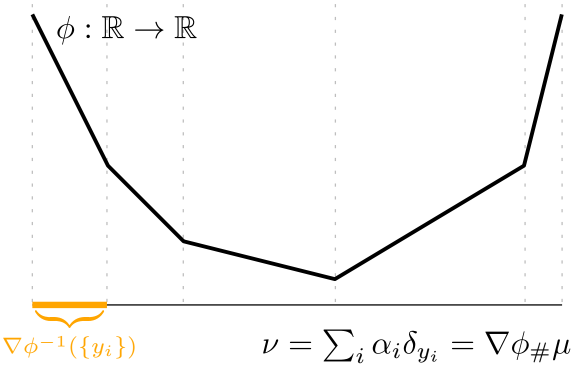


- All minimizing geodesics are of this form.
- There may exist **uncountably many geodesics**, e.g. if $\mathrm{spt}(\mu_0) \subseteq \mathbb{R} \times \{0\}$ and $\mathrm{spt}(\mu_0) \subseteq \{0\} \times \mathbb{R}$ on $\mathbb{R}^2$, any $\gamma \in \Gamma(\mu_0, \mu_1)$ is optimal!

Characterization of quadratic optimal transport

Theorem (Brenier) Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \ll \mathrm{Leb}$. Then

(i) there exists a unique map $T = \nabla\phi$, with ϕ convex, transporting μ to ν,

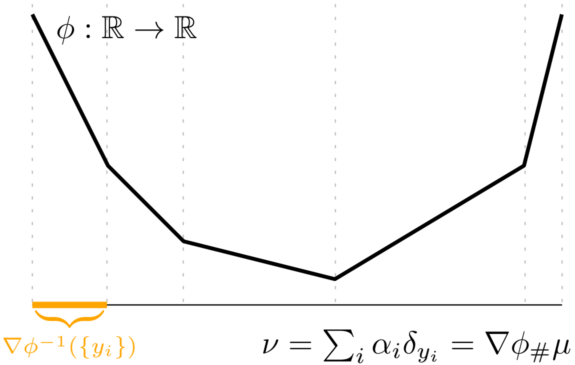


Characterization of quadratic optimal transport

Theorem (Brenier) Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \ll \mathrm{Leb}$. Then

(i) there exists a unique map $T = \nabla\phi$, with ϕ convex, transporting μ to ν,

(ii) the unique OT plan between μ and ν is $\gamma_T = (\mathrm{id}, T)_\# \mu$



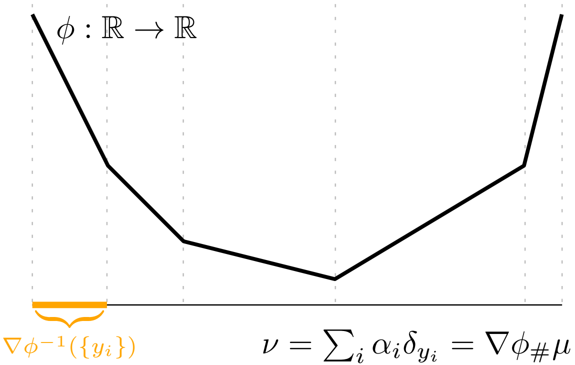
Characterization of quadratic optimal transport

Theorem (Brenier) Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \ll \mathrm{Leb}$. Then

(i) there exists a unique map $T = \nabla\phi$, with ϕ convex, transporting μ to ν,

(ii) the unique OT plan between μ and ν is $\gamma_T = (\mathrm{id}, T)_\# \mu$

(iii) the unique geodesic between μ and ν is $\mu_t = ((1-t)\,\mathrm{id} + tT)_\# \mu$ **(McCann)**



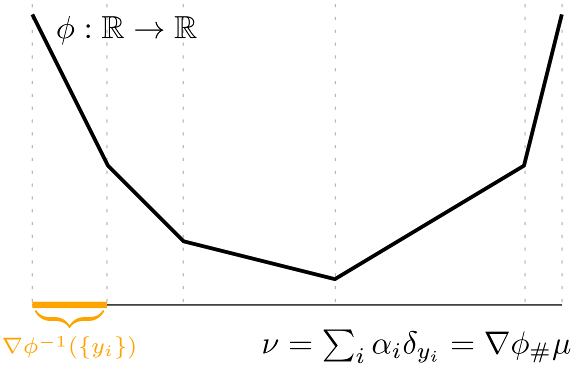
Characterization of quadratic optimal transport

Theorem (Brenier) Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ with $\mu \ll \mathrm{Leb}$. Then

(i) there exists a unique map $T = \nabla\phi$, with ϕ convex, transporting μ to ν,

(ii) the unique OT plan between μ and ν is $\gamma_T = (\mathrm{id}, T)_\# \mu$

(iii) the unique geodesic between μ and ν is $\mu_t = ((1-t)\,\mathrm{id} + tT)_\# \mu$ **(McCann)**



Proof relies on the dual problem to optimal transport, due to Kantorovich.

Kantorovich duality, quadratic case

- Quadratic optimal transport is equivalent to the maximal correlation problem:

$$\min_{\gamma\in\Gamma(\mu,\nu)} \int \|x-y\|^2 \,\mathrm{d}\gamma(x,y) = M_2(\mu) + M_2(\nu) - 2 \max_{\gamma\in\Gamma(\mu,\nu)} \int \langle x|y\rangle \,\mathrm{d}\gamma(x,y).$$

Kantorovich duality, quadratic case

- Quadratic optimal transport is equivalent to the maximal correlation problem:

$$\min_{\gamma\in\Gamma(\mu,\nu)} \int \|x-y\|^2 \,\mathrm{d}\gamma(x,y) = M_2(\mu) + M_2(\nu) - 2 \max_{\gamma\in\Gamma(\mu,\nu)} \int \langle x|y\rangle \,\mathrm{d}\gamma(x,y).$$

- **Kantorovich duality:** If $\mathrm{spt}(\mu) \subseteq X, \mathrm{spt}(\nu) \subseteq Y$ compact

$$\max_{\gamma\in\Gamma(\mu,\nu)} \int \langle x|y\rangle \,\mathrm{d}\gamma(x,y) = \min_{\substack{\phi\in\mathcal{C}^0(X),\psi\in\mathcal{C}^0(Y)\\ \phi(x)+\psi(y)\geq\langle x|y\rangle}} \int \phi \mathrm{d}\mu + \int \psi \mathrm{d}\nu.$$

Nb: the inequality min $\geq$ max follows from the constraint $\psi \oplus \psi \geq \langle\cdot|\cdot\rangle$.

Kantorovich duality, quadratic case

- Quadratic optimal transport is equivalent to the maximal correlation problem:

$$\min_{\gamma \in \Gamma(\mu,\nu)} \int \|x-y\|^2 \, \mathrm{d}\gamma(x,y) = M_2(\mu) + M_2(\nu) - 2 \max_{\gamma \in \Gamma(\mu,\nu)} \int \langle x|y \rangle \, \mathrm{d}\gamma(x,y).$$

- **Kantorovich duality:** If $\mathrm{spt}(\mu) \subseteq X, \mathrm{spt}(\nu) \subseteq Y$ compact

$$\boxed{\max_{\gamma \in \Gamma(\mu,\nu)} \int \langle x|y \rangle \, \mathrm{d}\gamma(x,y) = \min_{\substack{\phi \in \mathcal{C}^0(X), \psi \in \mathcal{C}^0(Y) \\ \phi(x)+\psi(y) \geq \langle x|y \rangle}} \int \phi \mathrm{d}\mu + \int \psi \mathrm{d}\nu.}$$

  Nb: the inequality min $\geq$ max follows from the constraint $\psi \oplus \psi \geq \langle \cdot | \cdot \rangle$.
- Many (most?) numerical methods are based on the dual problem.

Kantorovich duality, quadratic case

- Quadratic optimal transport is equivalent to the maximal correlation problem:

$$\min_{\gamma\in\Gamma(\mu,\nu)} \int \|x-y\|^2 \,\mathrm{d}\gamma(x,y) = M_2(\mu) + M_2(\nu) - 2 \max_{\gamma\in\Gamma(\mu,\nu)} \int \langle x|y\rangle \,\mathrm{d}\gamma(x,y).$$

- **Kantorovich duality:** If $\mathrm{spt}(\mu) \subseteq X, \mathrm{spt}(\nu) \subseteq Y$ compact

$$\max_{\gamma\in\Gamma(\mu,\nu)} \int \langle x|y\rangle \,\mathrm{d}\gamma(x,y) = \min_{\substack{\phi\in\mathcal{C}^0(X),\psi\in\mathcal{C}^0(Y)\\ \phi(x)+\psi(y)\geq\langle x|y\rangle}} \int \phi \mathrm{d}\mu + \int \psi \mathrm{d}\nu.$$

  Nb: the inequality min $\geq$ max follows from the constraint $\psi \oplus \psi \geq \langle\cdot|\cdot\rangle$.
- Many (most?) numerical methods are based on the dual problem.

Def: We will call **Kantorovich functional** $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* denotes the convex conjugate of ψ, i.e. $\psi^*(x) = \max_{y\in Y} \langle x|y\rangle - \psi(y)$

- Maximum correlation problem $\longleftrightarrow$ unconstrained convex minimization problem:

$$\max_{\gamma\in\Gamma(\mu,\nu)} \int \langle x|y\rangle \,\mathrm{d}\gamma(x,y) = \min_{\psi\in\mathcal{C}^0(Y)} \mathcal{K}_\mu(\psi) + \int \psi \mathrm{d}\nu.$$

Kantorovich functional: first derivative

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

Proposition: $\nabla \mathcal{K}_\mu(\psi) = -\nabla \psi^*_{\#}\mu.$

Kantorovich functional: first derivative

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

Proposition: $\nabla \mathcal{K}_\mu(\psi) = -\nabla\psi^*_\# \mu.$

- **Proof:** Let $\psi_t = (1-t)\psi_0 + t\psi_1$ and let $v = \psi_1 - \psi_0$. Differentiating the Fenchel-Young equality $\psi_t^* + \psi_t(\nabla\psi_t^*) = \langle \mathrm{id} \,|\, \nabla\psi_t^* \rangle$ gives $\frac{\mathrm{d}}{\mathrm{d}t}\psi_t^* = -v(\nabla\psi_t^*)$,

Kantorovich functional: first derivative

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

Proposition: $\nabla \mathcal{K}_\mu(\psi) = -\nabla\psi^*_{\#}\mu.$

- **Proof:** Let $\psi_t = (1-t)\psi_0 + t\psi_1$ and let $v = \psi_1 - \psi_0$. Differentiating the Fenchel-Young equality $\psi_t^* + \psi_t(\nabla\psi_t^*) = \langle \mathrm{id} \,|\, \nabla\psi_t^* \rangle$ gives $\frac{\mathrm{d}}{\mathrm{d}t}\psi_t^* = -v(\nabla\psi_t^*)$,and

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}_\mu(\psi_t) = \int \frac{\mathrm{d}}{\mathrm{d}t}\psi_t^* \mathrm{d}\mu = -\int v(\nabla\psi_t^*)\mathrm{d}\mu = -\int v \mathrm{d}[\nabla\psi^*_{t\#}\mu].$$

Kantorovich functional: first derivative

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

Proposition: $\boxed{\nabla\mathcal{K}_\mu(\psi) = -\nabla\psi^*_{\#}\mu.}$ In particular, ψ minimizes $\mathcal{K}_\mu(\cdot) + \langle\cdot|\nu\rangle \iff \nabla\psi^*_{\#}\mu = \nu \iff \phi := \psi^*$ is the Brenier potential between μ and ν.

- **Proof:** Let $\psi_t = (1-t)\psi_0 + t\psi_1$ and let $v = \psi_1 - \psi_0$. Differentiating the Fenchel-Young equality $\psi_t^* + \psi_t(\nabla\psi_t^*) = \langle \mathrm{id}\,|\nabla\psi_t^*\rangle$ gives $\frac{\mathrm{d}}{\mathrm{d}t}\psi_t^* = -v(\nabla\psi_t^*)$,and

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}_\mu(\psi_t) = \int \frac{\mathrm{d}}{\mathrm{d}t}\psi_t^*\mathrm{d}\mu = -\int v(\nabla\psi_t^*)\mathrm{d}\mu = -\int v\mathrm{d}[\nabla\psi^*_{t\#}\mu].$$

Kantorovich functional: first derivative

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

Proposition: $\boxed{\nabla \mathcal{K}_\mu(\psi) = -\nabla \psi^*_{\#}\mu.}$ In particular, ψ minimizes $\mathcal{K}_\mu(\cdot) + \langle \cdot | \nu \rangle \iff \nabla \psi^*_{\#}\mu = \nu \iff \phi := \psi^*$ is the Brenier potential between μ and ν.

- **Proof:** Let $\psi_t = (1-t)\psi_0 + t\psi_1$ and let $v = \psi_1 - \psi_0$. Differentiating the Fenchel-Young equality $\psi_t^* + \psi_t(\nabla \psi_t^*) = \langle \mathrm{id} \,|\nabla \psi_t^* \rangle$ gives $\frac{\mathrm{d}}{\mathrm{d}t}\psi_t^* = -v(\nabla \psi_t^*)$,and

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}_\mu(\psi_t) = \int \frac{\mathrm{d}}{\mathrm{d}t}\psi_t^* \mathrm{d}\mu = -\int v(\nabla \psi_t^*)\mathrm{d}\mu = -\int v \mathrm{d}[\nabla \psi^*_{t\#}\mu].$$

- When μ, ν are densities and $\nabla \psi^*$ is a diffeomorphism between their support, the pushforward condition ($\nabla \psi^*_{\#}\mu = \nu$) becomes the **Monge-Ampère equation**

$$\boxed{\nu(\nabla \psi^*) \det(\mathrm{D}^2 \psi^*) = \mu.}$$

Kantorovich functional: first derivative

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

Proposition: $\boxed{\nabla \mathcal{K}_\mu(\psi) = -\nabla\psi^*_\# \mu.}$ In particular, ψ minimizes $\mathcal{K}_\mu(\cdot) + \langle \cdot | \nu \rangle \iff \nabla\psi^*_\#\mu = \nu \iff \phi := \psi^*$ is the Brenier potential between μ and ν.

- **Proof:** Let $\psi_t = (1-t)\psi_0 + t\psi_1$ and let $v = \psi_1 - \psi_0$. Differentiating the Fenchel-Young equality $\psi_t^* + \psi_t(\nabla\psi_t^*) = \langle \mathrm{id} \,|\nabla\psi_t^*\rangle$ gives $\frac{\mathrm{d}}{\mathrm{d}t}\psi_t^* = -v(\nabla\psi_t^*)$,and

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}_\mu(\psi_t) = \int \frac{\mathrm{d}}{\mathrm{d}t}\psi_t^* \mathrm{d}\mu = -\int v(\nabla\psi_t^*)\mathrm{d}\mu = -\int v \mathrm{d}[\nabla\psi^*_{t\#}\mu].$$

- When μ, ν are densities and $\nabla\psi^*$ is a diffeomorphism between their support, the pushforward condition $(\nabla\psi^*_\#\mu = \nu)$ becomes the **Monge-Ampère equation**

$$\boxed{\nu(\nabla\psi^*)\det(\mathrm{D}^2\psi^*) = \mu.}$$

- The Monge-Ampère operator is linearized (in $\phi = \psi^*$!) into :

$$\det(D^2(\phi + v)) \simeq \det(D^2\phi)(1 + \mathrm{Tr}(D^2\phi^{-1}D^2 v))$$

Kantorovich functional: second derivative

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

- Let $\psi_t = (1-t)\psi_0 + t\psi_1$, with ψ_0, ψ_1 strongly convex and $\mathcal{C}^2$, and let $v = \psi_1 - \psi_0$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}_\mu(\psi_t) = -\int v(\nabla \psi_t^*)\mathrm{d}\mu.$$

Kantorovich functional: second derivative

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

- Let $\psi_t = (1-t)\psi_0 + t\psi_1$, with ψ_0, ψ_1 strongly convex and $\mathcal{C}^2$, and let $v = \psi_1 - \psi_0$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}_\mu(\psi_t) = -\int v(\nabla\psi_t^*)\mathrm{d}\mu.$$

  Taking the time-derivative of $\nabla\psi_t^* \circ \nabla\psi_t = \mathrm{id}$, we see that

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{K}_\mu(\psi_t) = \int \left\langle \nabla v(\nabla\psi_t^*) | D^2\psi_t^* \cdot \nabla v(\nabla\psi_t^*) \right\rangle \mathrm{d}\mu.$$

Kantorovich functional: second derivative

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

- Let $\psi_t = (1-t)\psi_0 + t\psi_1$, with ψ_0, ψ_1 strongly convex and $\mathcal{C}^2$, and let $v = \psi_1 - \psi_0$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}_\mu(\psi_t) = -\int v(\nabla\psi_t^*)\mathrm{d}\mu.$$

Taking the time-derivative of $\nabla\psi_t^* \circ \nabla\psi_t = \mathrm{id}$, we see that

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{K}_\mu(\psi_t) = \int \left\langle \nabla v(\nabla\psi_t^*) | D^2\psi_t^* \cdot \nabla v(\nabla\psi_t^*) \right\rangle \mathrm{d}\mu.$$

i.e. $\boxed{\left\langle D^2\mathcal{K}_\mu(\psi) v | v \right\rangle = \int \left\langle \nabla v(\nabla\psi^*) | D^2\psi^* \cdot \nabla v(\nabla\psi^*) \right\rangle \mathrm{d}\mu}$

Kantorovich functional: second derivative

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

- Let $\psi_t = (1-t)\psi_0 + t\psi_1$, with ψ_0, ψ_1 strongly convex and $\mathcal{C}^2$, and let $v = \psi_1 - \psi_0$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}_\mu(\psi_t) = -\int v(\nabla\psi_t^*)\mathrm{d}\mu.$$

  Taking the time-derivative of $\nabla\psi_t^* \circ \nabla\psi_t = \mathrm{id}$, we see that

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{K}_\mu(\psi_t) = \int \left\langle \nabla v(\nabla\psi_t^*) | D^2\psi_t^* \cdot \nabla v(\nabla\psi_t^*)\right\rangle \mathrm{d}\mu.$$

  i.e. $\boxed{\left\langle D^2\mathcal{K}_\mu(\psi) v | v\right\rangle = \int \left\langle \nabla v(\nabla\psi^*) | D^2\psi^* \cdot \nabla v(\nabla\psi^*)\right\rangle \mathrm{d}\mu}$

- Setting $\phi = \psi^*$ and $\tilde{v} = v(\nabla\phi)$, we get

$$\left\langle D^2\mathcal{K}_\mu(\psi) v | v\right\rangle = \int \left\langle \mathrm{D}^2\phi^{-1} \cdot \nabla\tilde{v} | \nabla\tilde{v}\right\rangle \mathrm{d}\mu$$

Kantorovich functional: second derivative

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

- Let $\psi_t = (1-t)\psi_0 + t\psi_1$, with ψ_0, ψ_1 strongly convex and $\mathcal{C}^2$, and let $v = \psi_1 - \psi_0$. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{K}_\mu(\psi_t) = -\int v(\nabla\psi_t^*)\mathrm{d}\mu.$$

Taking the time-derivative of $\nabla\psi_t^* \circ \nabla\psi_t = \mathrm{id}$, we see that

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathcal{K}_\mu(\psi_t) = \int \left\langle \nabla v(\nabla\psi_t^*) | D^2\psi_t^* \cdot \nabla v(\nabla\psi_t^*)\right\rangle \mathrm{d}\mu.$$

i.e. $\left\langle D^2\mathcal{K}_\mu(\psi)v | v\right\rangle = \int \left\langle \nabla v(\nabla\psi^*) | D^2\psi^* \cdot \nabla v(\nabla\psi^*)\right\rangle \mathrm{d}\mu$

- Setting $\phi = \psi^*$ and $\tilde{v} = v(\nabla\phi)$, we get

$$\left\langle D^2\mathcal{K}_\mu(\psi)v | v\right\rangle = \int \left\langle \mathrm{D}^2\phi^{-1} \cdot \nabla\tilde{v} | \nabla\tilde{v}\right\rangle \mathrm{d}\mu$$

- For $\mu \equiv 1$, a Newton method for minimizing $\mathcal{K}(\cdot) + \langle\nu|\cdot\rangle$ at ψ will involve the linear operator $\tilde{v} \mapsto \mathrm{div}((\mathrm{D}^2\psi^*)^{-1}\nabla\tilde{v})$

Outline

Kantorovich functional and computational geometry

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

Kantorovich functional and computational geometry

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

- If $\nu = \sum_i \alpha_i \delta_{y_j}$, we parametrize ψ by the vector $\Psi = (\psi_i)_{1 \leq i \leq N}$ with $\psi_i = \psi(y_i)$.

Kantorovich functional and computational geometry

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

- If $\nu = \sum_i \alpha_i \delta_{y_j}$, we parametrize ψ by the vector $\Psi = (\psi_i)_{1 \leq i \leq N}$ with $\psi_i = \psi(y_i)$. Then, $\psi^* = \max_i \langle \cdot | y_i \rangle - \psi_i$ is affine on each **power (or Laguerre) cell**

$$V_i(\Psi) = \{x \in \mathbb{R}^d \mid \forall j, \langle x | y_i \rangle - \psi_i \geq \langle x | y_j \rangle - \psi_j\}.$$

Kantorovich functional and computational geometry

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

- If $\nu = \sum_i \alpha_i \delta_{y_j}$, we parametrize ψ by the vector $\Psi = (\psi_i)_{1 \le i \le N}$ with $\psi_i = \psi(y_i)$. Then, $\psi^* = \max_i \langle \cdot | y_i \rangle - \psi_i$ is affine on each **power (or Laguerre) cell**
$$V_i(\Psi) = \{x \in \mathbb{R}^d \mid \forall j, \langle x | y_i \rangle - \psi_i \ge \langle x | y_j \rangle - \psi_j\}.$$
- In general, y_i does **NOT** belong to its Laguerre cell $V_i(\Psi)$.

Kantorovich functional and computational geometry

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

- If $\nu = \sum_i \alpha_i \delta_{y_j}$, we parametrize ψ by the vector $\Psi = (\psi_i)_{1 \leq i \leq N}$ with $\psi_i = \psi(y_i)$. Then, $\psi^* = \max_i \langle \cdot | y_i \rangle - \psi_i$ is affine on each **power (or Laguerre) cell**

$$V_i(\Psi) = \{x \in \mathbb{R}^d \mid \forall j, \langle x|y_i\rangle - \psi_i \geq \langle x|y_j\rangle - \psi_j\}.$$

- The dual mesh associated to this tesselation is the **regular triangulation** of the points $\{y_i\}_{1 \leq i \leq N}$. Two points y_i and y_j are connected iff the interface

$$\Gamma_{ij}(\Psi) = V_i(\Psi) \cap V_j(\Psi)$$

is non-empty.

Kantorovich functional and computational geometry

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

- If $\nu = \sum_i \alpha_i \delta_{y_j}$, we parametrize ψ by the vector $\Psi = (\psi_i)_{1 \leq i \leq N}$ with $\psi_i = \psi(y_i)$. Then, $\psi^* = \max_i \langle \cdot | y_i \rangle - \psi_i$ is affine on each **power (or Laguerre) cell**

$$V_i(\Psi) = \{x \in \mathbb{R}^d \mid \forall j, \langle x | y_i \rangle - \psi_i \geq \langle x | y_j \rangle - \psi_j\}.$$

- The dual mesh associated to this tesselation is the **regular triangulation** of the points $\{y_i\}_{1 \leq i \leq N}$. Two points y_i and y_j are connected iff the interface

$$\Gamma_{ij}(\Psi) = V_i(\Psi) \cap V_j(\Psi)$$

is non-empty. In addition, $\Gamma_{ij}(\Psi)$ is orthogonal to $y_j - y_i$.

Kantorovich functional and computational geometry

Kantorovich functional: $\mathcal{K}_\mu(\psi) = \int \psi^* \mathrm{d}\mu$, where ψ^* is the convex conjugate of ψ

- If $\nu = \sum_i \alpha_i \delta_{y_j}$, we parametrize ψ by the vector $\Psi = (\psi_i)_{1 \le i \le N}$ with $\psi_i = \psi(y_i)$. Then, $\psi^* = \max_i \langle \cdot | y_i \rangle - \psi_i$ is affine on each **power (or Laguerre) cell**

$$V_i(\Psi) = \{x \in \mathbb{R}^d \mid \forall j, \langle x | y_i \rangle - \psi_i \ge \langle x | y_j \rangle - \psi_j\}.$$

- The dual mesh associated to this tesselation is the **regular triangulation** of the points $\{y_i\}_{1 \le i \le N}$. Two points y_i and y_j are connected iff the interface

$$\Gamma_{ij}(\Psi) = V_i(\Psi) \cap V_j(\Psi)$$

is non-empty. In addition, $\Gamma_{ij}(\Psi)$ is orthogonal to $y_j - y_i$.

Theorem Finding an *optimal transport* between a probability density μ on $\mathbb{R}^d$ and $\nu = \sum_i \alpha_i \delta_{y_i}$ amounts to maximizing $\Psi \in \mathbb{R}^N \mapsto \mathrm{K}_\mu(\Psi) + \sum_i \alpha_i \Psi_i$, where

$$\mathrm{K}_\mu(\Psi) = \sum_i \int_{V_i(\Psi)} \langle x | y_i \rangle - \psi_i \mathrm{d}\mu.$$

Moreover, Ψ is $\mathcal{C}^1$ and $\partial_{y_i} \mathcal{K}_\mu(\Psi) = -\mu(V_i(\Psi))$. [Aurenhammer, Hoffman, Aronov]

$\nabla \mathrm{K}(\Psi)$ is consistent with continuous case: $\nabla \mathcal{K}(\psi) = -\nabla \psi^*_\# \mu = \sum_i \mu(V_i(\Psi)) \delta_{y_i}$.

Optimal transport and mass-constrained power diagrams

Theorem Finding an *optimal transport* between a probability density μ on $\mathbb{R}^d$ and $\nu = \sum_i \alpha_i \delta_{y_i}$ amounts to maximizing $\Psi \in \mathbb{R}^N \mapsto \mathrm{K}_\mu(\Psi) + \sum_i \alpha_i \Psi_i$, where

$$\mathrm{K}_\mu(\Psi) = \sum_i \int_{V_i(\Psi)} \langle x | y_i \rangle - \psi_i \mathrm{d}\mu.$$

Moreover, Ψ is $\mathcal{C}^1$ and $\partial_{y_i} \mathcal{K}_\mu(\Psi) = -\mu(V_i(\Psi))$. [Aurenhammer, Hoffman, Aronov]

Optimal transport and mass-constrained power diagrams

Theorem Finding an *optimal transport* between a probability density μ on $\mathbb{R}^d$ and $\nu = \sum_i \alpha_i \delta_{y_i}$ amounts to maximizing $\Psi \in \mathbb{R}^N \mapsto \mathrm{K}_\mu(\Psi) + \sum_i \alpha_i \Psi_i$, where

$$\mathrm{K}_\mu(\Psi) = \sum_i \int_{V_i(\Psi)} \langle x|y_i\rangle - \psi_i \mathrm{d}\mu.$$

Moreover, Ψ is $\mathcal{C}^1$ and $\partial_{y_i}\mathcal{K}_\mu(\Psi) = -\mu(V_i(\Psi))$. [Aurenhammer, Hoffman, Aronov]

- Denote $G(\Psi) = -\nabla\mathcal{K}(\Psi) = (\mu(V_i(\Psi)))_{1\leq i\leq N}$. Then

$$\begin{aligned}\Psi \text{ solves OT} &\iff G(\Psi) = \alpha \\ &\iff \psi = \Psi^{**} \text{ is an Alexandrov solution to } \mu(\nabla\psi)\det(\mathrm{D}^2\psi) = \nu\end{aligned}$$

demo

Optimal transport and mass-constrained power diagrams

Theorem Finding an *optimal transport* between a probability density μ on $\mathbb{R}^d$ and $\nu = \sum_i \alpha_i \delta_{y_i}$ amounts to maximizing $\Psi \in \mathbb{R}^N \mapsto \mathrm{K}_\mu(\Psi) + \sum_i \alpha_i \Psi_i$, where

$$\mathrm{K}_\mu(\Psi) = \sum_i \int_{V_i(\Psi)} \langle x|y_i\rangle - \psi_i \mathrm{d}\mu.$$

Moreover, Ψ is $\mathcal{C}^1$ and $\partial_{y_i}\mathcal{K}_\mu(\Psi) = -\mu(V_i(\Psi))$. [Aurenhammer, Hoffman, Aronov]

- Denote $G(\Psi) = -\nabla\mathcal{K}(\Psi) = (\mu(V_i(\Psi)))_{1\leq i\leq N}$. Then

$$\begin{aligned} \Psi \text{ solves OT } &\iff G(\Psi) = \alpha \\ &\iff \psi = \Psi^{**} \text{ is an Alexandrov solution to } \mu(\nabla\psi)\det(\mathrm{D}^2\psi) = \nu \end{aligned}$$

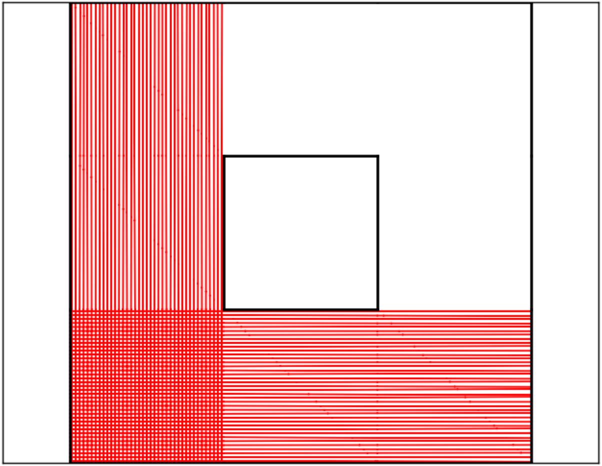
demo

- Computing $\mathrm{D}G(\Psi)$ is as costly as computing $G \rightsquigarrow$ Newton's method

$$\boxed{\Psi^{k+1} = \Psi^k - \mathrm{D}G(\Psi^k)^{-1}(G(\Psi^k) - \nu).}$$

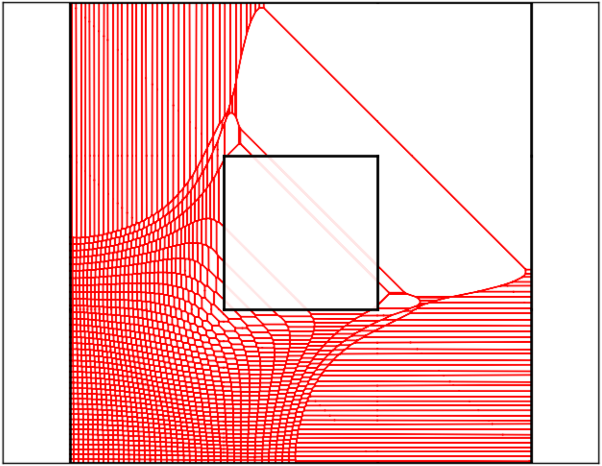
Difficulty: $\mathrm{D}G(\Psi^k)$ is in general not invertible $\rightsquigarrow$ damping.

Numerical example: Newton iterations



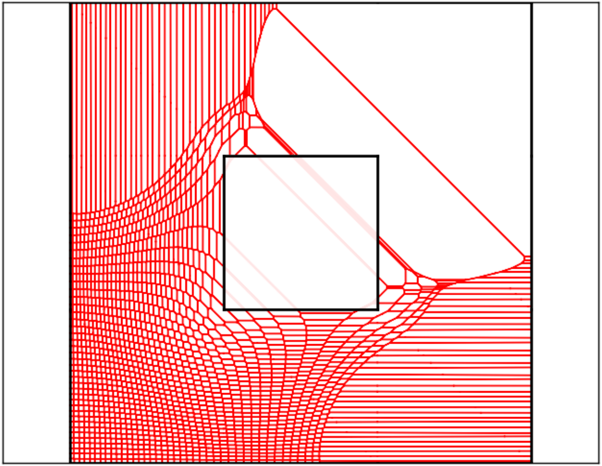
Source: Piecewise linear density on $X = [0,3]^2$ / **Target:** Uniform grid in $[0,1]^2$.

Numerical example: Newton iterations



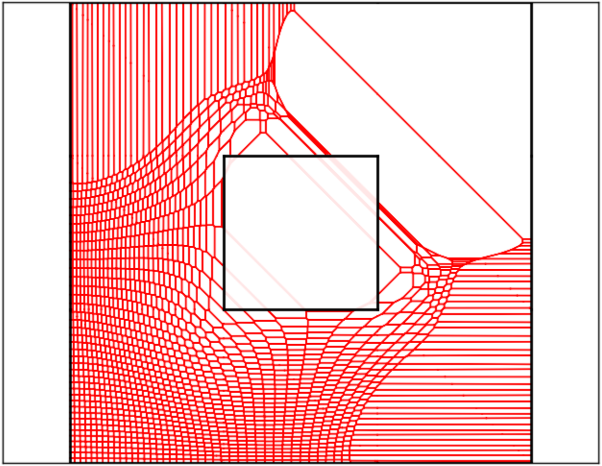
Source: Piecewise linear density on $X = [0,3]^2$ / **Target:** Uniform grid in $[0,1]^2$.

Numerical example: Newton iterations



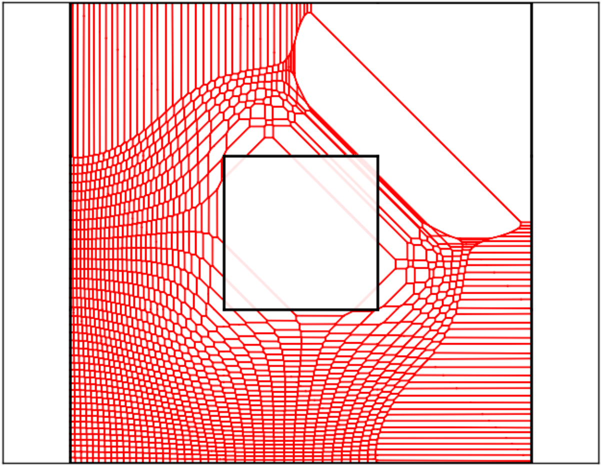
Source: Piecewise linear density on $X = [0,3]^2$ / **Target:** Uniform grid in $[0,1]^2$.

Numerical example: Newton iterations



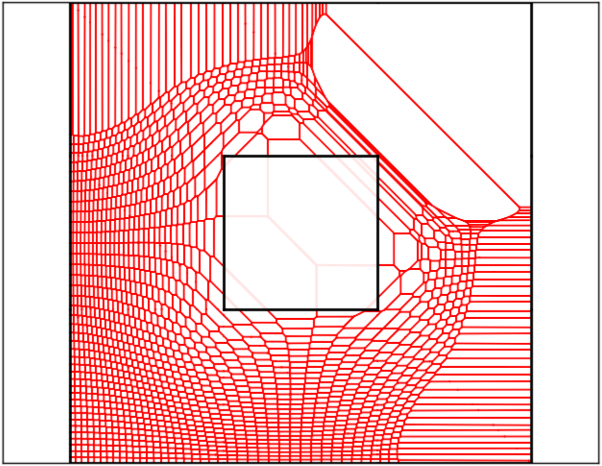
Source: Piecewise linear density on $X = [0,3]^2$ / **Target:** Uniform grid in $[0,1]^2$.

Numerical example: Newton iterations



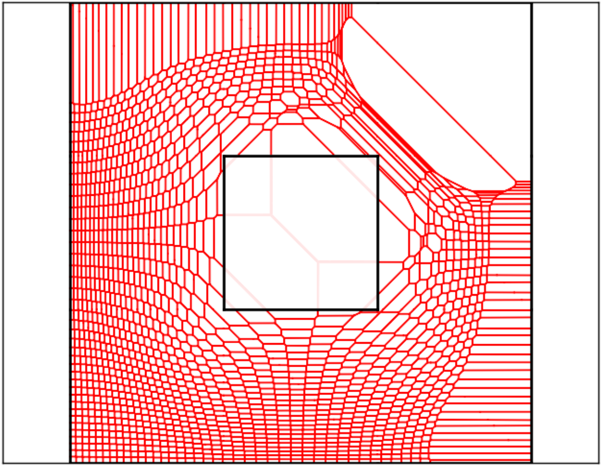
Source: Piecewise linear density on $X = [0, 3]^2$ / **Target:** Uniform grid in $[0, 1]^2$.

Numerical example: Newton iterations



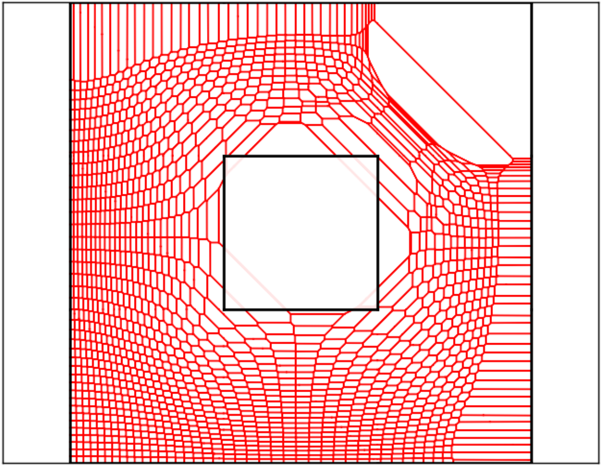
Source: Piecewise linear density on $X = [0, 3]^2$ / **Target:** Uniform grid in $[0, 1]^2$.

Numerical example: Newton iterations



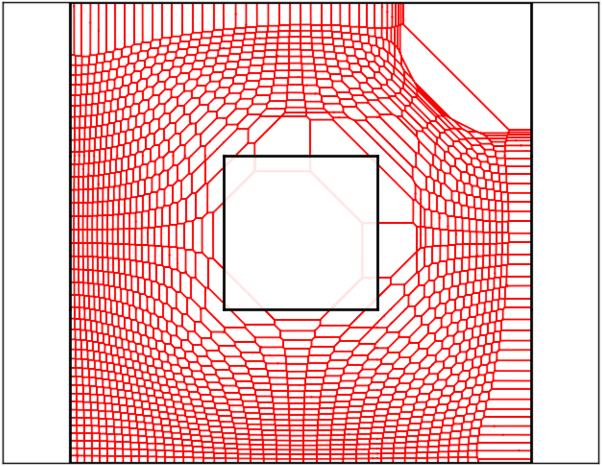
Source: Piecewise linear density on $X = [0,3]^2$ / **Target:** Uniform grid in $[0,1]^2$.

Numerical example: Newton iterations



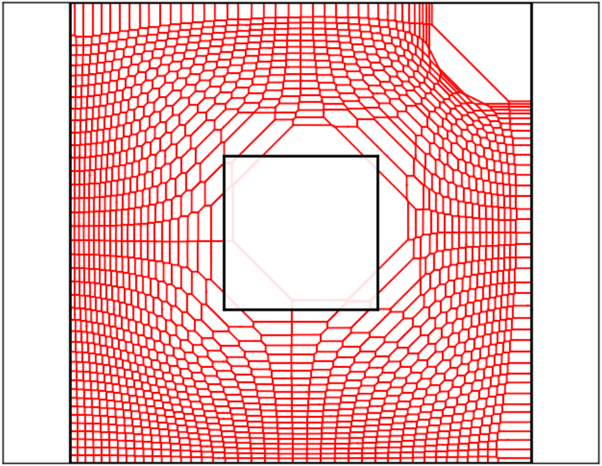
Source: Piecewise linear density on $X = [0,3]^2$ / **Target:** Uniform grid in $[0,1]^2$.

Numerical example: Newton iterations



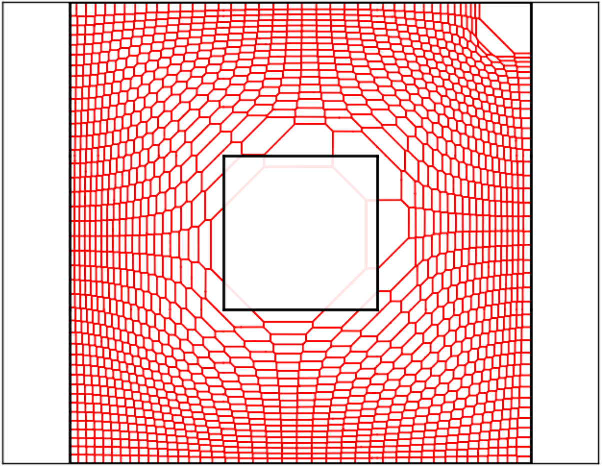
Source: Piecewise linear density on $X = [0,3]^2$ / **Target:** Uniform grid in $[0,1]^2$.

Numerical example: Newton iterations



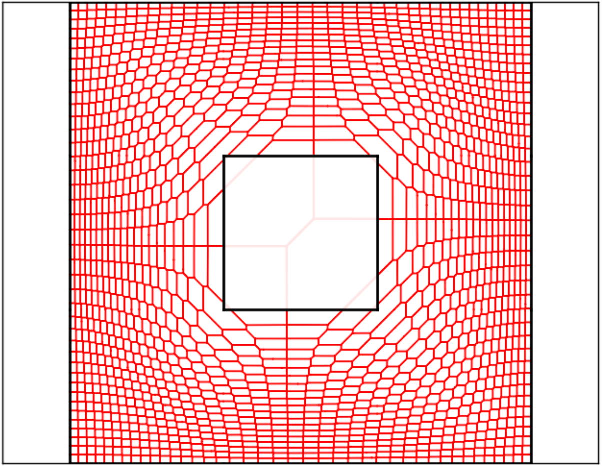
Source: Piecewise linear density on $X = [0, 3]^2$ / **Target:** Uniform grid in $[0, 1]^2$.

Numerical example: Newton iterations



Source: Piecewise linear density on $X = [0,3]^2$ / **Target:** Uniform grid in $[0,1]^2$.

Numerical example: Newton iterations



Source: Piecewise linear density on $X = [0,3]^2$ / **Target:** Uniform grid in $[0,1]^2$.

Newton's method for semi-discrete OT is able to handle huge problems in 2D/3D, up to $N \sim 10^8$ in 3D. See e.g. Geogram (Bruno Lévy) or SDOT (Hugo Leclerc).

Kantorovich's functional: second derivative

Denote $G(\Psi) = -\nabla \mathcal{K}(\Psi) = (\mu(V_i(\Psi)))_{1 \leq i \leq N}$.

Prop.: Assume $\mu \in \mathcal{C}^0(X)$, X convex and $\Psi \in \mathbb{R}^N$ is such that $\forall i, G_i(\Psi) > 0$. Then,

$$\forall i \neq j, \quad \frac{\partial G_i}{\partial \psi_j}(\Psi) = \frac{1}{\|y_i - y_j\|} \int_{\Gamma_{ij}(\psi)} \rho(x) \mathrm{d}\mathcal{H}^{d-1}(x) \quad \left[= -\frac{\partial^2 \mathcal{K}}{\partial \psi_j \psi_i}(\Psi) \right]$$

$$\forall i, \quad \frac{\partial G_i}{\partial \psi_i}(\Psi) = -\sum_{j \neq i} \frac{\partial G_i}{\partial \psi_j}(\Psi) \qquad \left[= -\frac{\partial^2 \mathcal{K}}{\partial \psi_i^2}(\Psi) \right]$$

Kantorovich's functional: second derivative

Denote $G(\Psi) = -\nabla\mathcal{K}(\Psi) = (\mu(V_i(\Psi)))_{1\leq i\leq N}$.

Prop.: Assume $\mu \in \mathcal{C}^0(X)$, X convex and $\Psi \in \mathbb{R}^N$ is such that $\forall i$, $G_i(\Psi) > 0$. Then,

$$\forall i \neq j, \quad \frac{\partial G_i}{\partial \psi_j}(\Psi) = \frac{1}{\|y_i - y_j\|} \int_{\Gamma_{ij}(\psi)} \rho(x) \mathrm{d}\mathcal{H}^{d-1}(x) \quad \left[= -\frac{\partial^2 \mathcal{K}}{\partial \psi_j \psi_i}(\Psi)\right]$$

$$\forall i, \quad \frac{\partial G_i}{\partial \psi_i}(\Psi) = -\sum_{j\neq i} \frac{\partial G_i}{\partial \psi_j}(\Psi) \quad \left[= -\frac{\partial^2 \mathcal{K}}{\partial \psi_i^2}(\Psi)\right]$$

- We recover convexity of $\mathcal{K}$ ($\mathrm{D}^2\mathcal{K}(\Psi) \geq 0$), and we see that constant vectors $\mathbb{R}(1, \ldots, 1)$ belong to $\mathrm{Ker}(\mathrm{D}^2\mathcal{K})$.

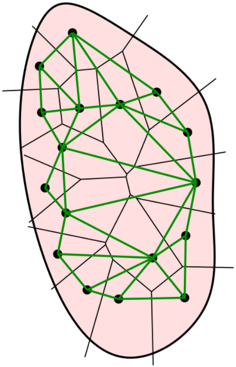
Kantorovich's functional: second derivative

Denote $G(\Psi) = -\nabla\mathcal{K}(\Psi) = (\mu(V_i(\Psi)))_{1\leq i\leq N}$.

Prop.: Assume $\mu \in \mathcal{C}^0(X)$, X convex and $\Psi \in \mathbb{R}^N$ is such that $\forall i$, $G_i(\Psi) > 0$. Then,

$$\forall i \neq j, \quad \frac{\partial G_i}{\partial \psi_j}(\Psi) = \frac{1}{\|y_i - y_j\|}\int_{\Gamma_{ij}(\psi)} \rho(x)\mathrm{d}\mathcal{H}^{d-1}(x) \quad \left[= -\frac{\partial^2\mathcal{K}}{\partial\psi_j\psi_i}(\Psi)\right]$$

$$\forall i, \quad \frac{\partial G_i}{\partial \psi_i}(\Psi) = -\sum_{j\neq i}\frac{\partial G_i}{\partial \psi_j}(\Psi) \quad \left[= -\frac{\partial^2\mathcal{K}}{\partial\psi_i^2}(\Psi)\right]$$



- We recover convexity of $\mathcal{K}$ ($\mathrm{D}^2\mathcal{K}(\Psi) \geq 0$), and we see that constant vectors $\mathbb{R}(1,\dots,1)$ belong to $\mathrm{Ker}(\mathrm{D}^2\mathcal{K})$.
- The matrix $\mathrm{D}G$ is the Laplacian of a weighted graph.

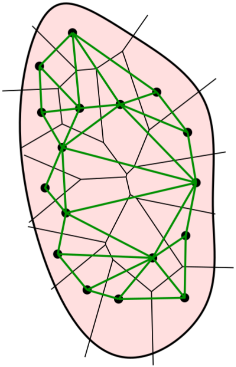
Kantorovich's functional: second derivative

Denote $G(\Psi) = -\nabla\mathcal{K}(\Psi) = (\mu(V_i(\Psi)))_{1\leq i\leq N}$.

Prop.: Assume $\mu \in \mathcal{C}^0(X)$, X convex and $\Psi \in \mathbb{R}^N$ is such that $\forall i, G_i(\Psi) > 0$. Then,

$$\forall i \neq j, \quad \frac{\partial G_i}{\partial \psi_j}(\Psi) = \frac{1}{\|y_i - y_j\|}\int_{\Gamma_{ij}(\psi)} \rho(x) \mathrm{d}\mathcal{H}^{d-1}(x) \quad \left[= -\frac{\partial^2 \mathcal{K}}{\partial \psi_j \psi_i}(\Psi)\right]$$

$$\forall i, \quad \frac{\partial G_i}{\partial \psi_i}(\Psi) = -\sum_{j\neq i} \frac{\partial G_i}{\partial \psi_j}(\Psi) \quad \left[= -\frac{\partial^2 \mathcal{K}}{\partial \psi_i^2}(\Psi)\right]$$



- We recover convexity of $\mathcal{K}$ ($\mathrm{D}^2\mathcal{K}(\Psi) \geq 0$), and we see that constant vectors $\mathbb{R}(1,\ldots,1)$ belong to $\mathrm{Ker}(\mathrm{D}^2\mathcal{K})$.
- The matrix $\mathrm{D}G$ is the Laplacian of a weighted graph.
- If $\rho \equiv 1$ on a bounded convex set X, the operator $\mathrm{D}^2\mathcal{K}(\Psi)$ satisfies a Poincaré inequality. With $\mu = G(\Psi)$,

$$\mathrm{Var}_\mu(v) \leq C(X)\left\langle \mathrm{D}^2\mathcal{K}(\Psi)v|v\right\rangle.$$

(Estimation follows from [Eymard, Gallouët, Herbin '00])

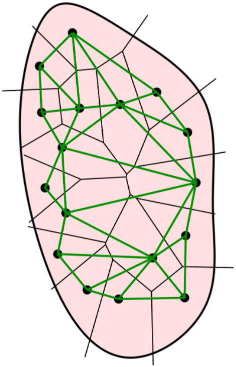
Kantorovich's functional: second derivative

Denote $G(\Psi) = -\nabla\mathcal{K}(\Psi) = (\mu(V_i(\Psi)))_{1\leq i\leq N}$.

Prop.: Assume $\mu \in \mathcal{C}^0(X)$, X convex and $\Psi \in \mathbb{R}^N$ is such that $\forall i, G_i(\Psi) > 0$. Then,

$$\forall i \neq j, \quad \frac{\partial G_i}{\partial \psi_j}(\Psi) = \frac{1}{\|y_i - y_j\|} \int_{\Gamma_{ij}(\psi)} \rho(x)\mathrm{d}\mathcal{H}^{d-1}(x) \quad \left[= -\frac{\partial^2 \mathcal{K}}{\partial \psi_j \psi_i}(\Psi)\right]$$

$$\forall i, \quad \frac{\partial G_i}{\partial \psi_i}(\Psi) = -\sum_{j\neq i} \frac{\partial G_i}{\partial \psi_j}(\Psi) \quad \left[= -\frac{\partial^2 \mathcal{K}}{\partial \psi_i^2}(\Psi)\right]$$



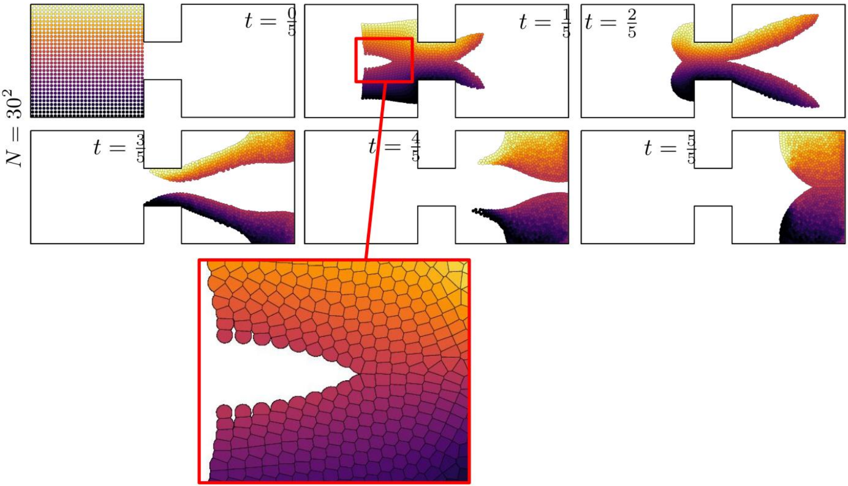
- We recover convexity of $\mathcal{K}$ ($\mathrm{D}^2\mathcal{K}(\Psi) \geq 0$), and we see that constant vectors $\mathbb{R}(1, \dots, 1)$ belong to $\mathrm{Ker}(\mathrm{D}^2\mathcal{K})$.
- The matrix $\mathrm{D}G$ is the Laplacian of a weighted graph.
- If $\rho \equiv 1$ on a bounded convex set X, the operator $\mathrm{D}^2\mathcal{K}(\Psi)$ satisfies a Poincaré inequality. With $\mu = G(\Psi)$,

$$\mathrm{Var}_\mu(v) \leq C(X) \left\langle \mathrm{D}^2\mathcal{K}(\Psi)v | v \right\rangle .$$

  (Estimation follows from [Eymard, Gallouët, Herbin '00])
- Finite-volume discretization of $\tilde{v} \mapsto \mathrm{div}((\mathrm{D}^2\psi^*)^{-1}\nabla\tilde{v})$?

Outline

Crowd motion



- Macroscopic crowd motion model [Maury, Roudneff-Chupin, Santambrogio]
- Particle discretization relies on a Moreau-Yosida regularization of the Lagrangian energy $\mathcal{E}_{\mathrm{cong}}$ ($\rightsquigarrow$ partial optimal transport) [M., Santambrogio, Stra]
- By OT, a moving mesh is automatically associated to the moving particle cloud!

Summary/questions

- OT can be used to interpret/reformulate some PDEs from fluid mechanics.
- This leads to new Eulerian numerical schemes (which will be the main object of this workshop), but also to Lagrangian (particle) discretization of evolution PDE.

`https://github.com/sd-ot` `https://github.com/BrunoLevy/geogram`

- Can OT be useful to construct (moving) meshes for finite volumes ?