The Principle of Least Action and its Geometric Aspects

... What I understood ...

UCLA IPAM, February 2016
OVERVIEW

Part. 1. The Least Action Principle
Part. 2. Optimal Transport – Elementary intro.
Part. 3. The Semi-Discrete Case
Part. 4. Understanding What’s Going On…
Part. 5. Concluding Words
Yann Brenier
The polar factorization theorem
(Brenier Transport)

“Each time the Laplace operator is used in a PDE, it can be replaced with the Monte-Ampère operator, and then interesting things occur”
OVERVIEW

**Yann Brenier**
The polar factorization theorem
(Brenier Transport)

**Cédric Villani**
Optimal Transport Old & New
Topics on Optimal Transport

“Each time the Laplace operator is used in a PDE, it can be replaced with the Monte-Ampère operator, and then interesting things occur”
OVERVIEW

A computer programmer’s adventure in MathLand

June 2015
Institut Fourier

March 2015
Bonn

Febr 2015
BIRS (Canada)

Febr 2015
LJLL

Discuss. with Quentin Mérigot (computational geometry – geometric measure theory)
1

The Least Action Principle
How to “morph” a shape into another one of same mass while minimizing the “effort”? 
Part. 1 The Least Action Principle

How to “morph” a shape into another one of same mass while minimizing the “effort”? 

The “effort” of the best $T$ defines a distance between the shapes.
The Least Action Principle

Axiom 1: There exists a function $L(x, \dot{x}, t)$ that describes the state of a physical system.

Short summary of the 1st chapter of Landau, Lifshitz Course of Theoretical Physics
Axiom 1: There exists a function $L(x, \dot{x}, t)$ that describes the state of a physical system.

The Least Action Principle

Euler

Hamilton, Legendre, Maupertuis

Lagrange
The Least Action Principle

**Axiom 1:** There exists a function \( L(x, \dot{x}, t) \) that describes the state of a physical system.
The Least Action Principle

Axiom 1: There exists a function \( L(x, \dot{x}, t) \) that describes the state of a physical system.
The Least Action Principle

Axiom 1: There exists a function \( L(x, \dot{x}, t) \) that describes the state of a physical system

\[
\int_{t_1}^{t_2} L(x, \dot{x}, t) \, dt
\]

Axiom 2: The movement (time evolution) of the physical system minimizes the following integral

Euler

Hamilton, Legendre, Maupertuis

Lagrange
The Least Action Principle

**Axiom 1:** There exists a function $L(x, \dot{x}, t)$ that describes the state of a physical system

$$\int_{t_1}^{t_2} L(x, \dot{x}, t) \, dt$$

**Axiom 2:** The movement (time evolution) of the physical system minimizes the following integral
The Least Action Principle

Axiom 1: There exists a function $L(x, \dot{x}, t)$ that describes the state of a physical system

\[
\int_{t_1}^{t_2} L(x, \dot{x}, t) \, dt
\]

Axiom 2: The movement (time evolution) of the physical system minimizes the following integral

Theorem 1: (Lagrange equation):

\[
\frac{\partial L}{\partial x} = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}}
\]
The Least Action Principle

Axiom 1: There exists $L$

Axiom 2: The movement minimizes

$$\int_{t_1}^{t_2} L(x, \dot{x}, t) \, dt$$

Axiom 3:

Invariance w.r.t. change of Gallileo frame + hom. + isotrop.:

$$x' = x + vt$$

$$t' = t$$

Theorem 1: (Lagrange equation):

$$\frac{\partial L}{\partial x} = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}}$$
The Least Action Principle

**Axiom 1**: There exists $L$

**Axiom 2**: The movement minimizes

$$\int_{t_1}^{t_2} L(x, \dot{x}, t) \, dt$$

**Theorem 1**: (Lagrange equation):

$$\frac{\partial L}{\partial x} = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}}$$

**Axiom 3**:

Invariance w.r.t. change of Gallileo frame + hom. + isotrop. :

$$x' = x + vt$$

$$t' = t$$

**Theorem 2**:
The Least Action Principle

Axiom 1: There exists $L$

Axiom 2: The movement minimizes

$$\int_{t_1}^{t_2} L(x, \dot{x}, t) \, dt$$

Axiom 3:

Invariance w.r.t. change of Gallileo frame + hom. + isotrop. :

$$x' = x + vt$$
$$t' = t$$

Theorem 2:

$$\dot{x} \frac{\partial L}{\partial \dot{x}} - L = \text{cte}$$

Homogeneity of time $\rightarrow$
Preservation of energy

Theorem 1: (Lagrange equation):

$$\frac{\partial L}{\partial \dot{x}} = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}}$$
The Least Action Principle

**Axiom 1:** There exists $L$

**Axiom 2:** The movement minimizes

$$\int_{t_1}^{t_2} L(x, \dot{x}, t) \, dt$$

**Theorem 1:** (Lagrange equation):

$$\frac{\partial L}{\partial \dot{x}} = \frac{d}{dt} \frac{\partial L}{\partial x}$$

**Axiom 3:**
Invariance w.r.t. change of Gallileo frame + hom. + isotrop. :

$$x' = x + vt$$

$$t' = t$$

**Theorem 2:**

$$\dot{x} \frac{\partial L}{\partial \dot{x}} - L = \text{cte}$$

Homogeneity of time $\rightarrow$ Preservation of energy

Homogeneity of space $\rightarrow$ Preservation of momentum
The Least Action Principle

Axiom 1: There exists $L$

Axiom 2: The movement minimizes
\[ \int_{t_1}^{t_2} L(x, \dot{x}, t) \, dt \]

Theorem 1: (Lagrange equation):
\[ \frac{\partial L}{\partial x} = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \]

Axiom 3: Invariance w.r.t. change of Gallileo frame + hom. + isotrop.:
- $x' = x + vt$
- $t' = t$

Theorem 2:
\[ \dot{x} \frac{\partial L}{\partial \dot{x}} - L = \text{cte} \]

- Homogeneity of time → Preservation of energy
- Homogeneity of space → Preservation of momentum
- Isotropy of space → Preservation of angular momentum
The Least Action Principle

**Axiom 1:** There exists $L$

**Axiom 2:** The movement minimizes

$$\int_{t_1}^{t_2} L(x, \dot{x}, t) \, dt$$

**Theorem 1:** (Lagrange equation):

$$\frac{\partial L}{\partial x} = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}}$$

**Axiom 3:**
Invariance w.r.t. change of Gallileo frame + hom. + isotrop. :

$$x' = x + vt$$

$$t' = t$$

**Theorem 2:**

$$\dot{x} \frac{\partial L}{\partial \dot{x}} - L = \text{cte}$$

Preserved quantities

“Integrals of Motion”

Noether’s theorem

Homogeneity of space → Preservation of energy

Homogeneity of time → Preservation of momentum

Isotropy of space → Preservation of angular momentum
The Least Action Principle

Axiom 1: There exists $L$

Axiom 2: The movement minimizes $\int L$

Theorem 1: (Lagrange equation):

$$\frac{\partial L}{\partial x} = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}}$$

Free particle:

Theorem 3: $v = \text{cte}$ (*Newton’s law I*)

Axiom 3: Invariance w.r.t. change of Gallileo frame + hom. + isotrop.:

$$x' = x + vt$$

$$t' = t$$

Theorem 2:

$$\dot{x} \frac{\partial L}{\partial \dot{x}} - L = \text{cte}$$

Homogeneity of time $\rightarrow$

Preservation of *energy*

Homogeneity of space $\rightarrow$

Preservation of *momentum*

Isotropy of space $\rightarrow$

Preservation of *angular momentum*
The Least Action Principle

**Axiom 1:** There exists $L$

**Axiom 2:** The movement minimizes $\int L$

**Theorem 1:** (Lagrange equation):

$$\frac{\partial L}{\partial x} = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}}$$

**Free particle:**

**Theorem 3:** $v = \text{cte}$ (*Newton’s law I*)

Expression of the Lagrangian:

$L = \frac{1}{2} m v^2$

**Axiom 3:**
Invariance w.r.t. change of Gallileo frame + hom. + isotrop. :

$$x' = x + vt$$  
$$t' = t$$

**Theorem 2:**

$$\dot{x} \frac{\partial L}{\partial \dot{x}} - L = \text{cte}$$

Homogeneity of time $\rightarrow$

Preservation of energy

Homogeneity of space $\rightarrow$

Preservation of momentum

Isotropy of space $\rightarrow$

Preservation of angular momentum
The Least Action Principle

Axiom 1: There exists $L$

Axiom 2: The movement minimizes $\int L$

Theorem 1: (Lagrange equation):

$$\frac{\partial L}{\partial x} = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}}$$

Free particle:

Theorem 3: $v = \text{cte}$ (Newton’s law I)

Expression of the Lagrangian:

$$L = \frac{1}{2} m \: v^2$$

Expression of the Energy:

$$E = \frac{1}{2} m \: v^2$$

Axiom 3:

Invariance w.r.t. change of Gallileo frame + hom. + isotrop.:

$$x' = x + vt$$
$$t' = t$$

Theorem 2:

$$\dot{x} \frac{\partial L}{\partial \dot{x}} - L = \text{cte}$$

Homogeneity of time $\rightarrow$
Preservation of energy

Homogeneity of space $\rightarrow$
Preservation of momentum

Isotropy of space $\rightarrow$
Preservation of angular momentum
The Least Action Principle

Axiom 1: There exists $L$

Axiom 2: The movement minimizes $\int L$

Theorem 1: (Lagrange equation):

$$\frac{\partial L}{\partial x} = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}}$$

Free particle:

Theorem 3: $v = \text{cte}$ *(Newton’s law I)*

Expression of the Lagrangian:

$$L = \frac{1}{2} m \, v^2$$

Expression of the Energy:

$$E = \frac{1}{2} m \, v^2$$

Axiom 3:

Invariance w.r.t. change of Gallileo frame + hom. + isotrop. :

$$x' = x + vt$$

$$t' = t$$

Particle in a field:

Expression of the Lagrangian:

$$L = \frac{1}{2} m \, v^2 - U(x)$$
The Least Action Principle

Axiom 1: There exists $L$

Axiom 2: The movement minimizes $\int L$

Theorem 1: (Lagrange equation):
$$\frac{\partial L}{\partial x} = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}}$$

Free particle:

Theorem 3: $v = \text{cte}$ (Newton's law I)

Expression of the Lagrangian:
$$L = \frac{1}{2} m \ v^2$$

Expression of the Energy:
$$E = \frac{1}{2} m \ v^2$$

Axiom 3:

Invariance w.r.t. change of Gallileo frame + hom. + isotrop. :

$$x' = x + vt$$
$$t' = t$$

Particle in a field:

Expression of the Lagrangian:
$$L = \frac{1}{2} m \ v^2 - U(x)$$

Expression of the Energy:
$$E = \frac{1}{2} m \ v^2 + U(x)$$
The Least Action Principle

Axiom 1: There exists $L$

Axiom 2: The movement minimizes $\int L$

Theorem 1: (Lagrange equation):
\[
\frac{\partial L}{\partial x} = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}}
\]

Free particle:

Theorem 3: $v = \text{cte} \ (\text{Newton’s law I})$

Expression of the Lagrangian:
\[ L = \frac{1}{2} m \ v^2 \]
Expression of the Energy:
\[ E = \frac{1}{2} m \ v^2 \]

Axiom 3:
Invariance w.r.t. change of Gallileo frame + hom. + isotrop.:
\[
\begin{align*}
\dot{x}' &= x + vt \\
\dot{t}' &= t
\end{align*}
\]

Particle in a field:

Expression of the Lagrangian:
\[ L = \frac{1}{2} m \ v^2 - U(x) \]
Expression of the Energy:
\[ E = \frac{1}{2} m \ v^2 + U(x) \]

Theorem 4:
\[ m \ddot{x} = -\nabla U \ (\text{Newton’s law II}) \]
The Least Action Principle

*(relativistic setting)*

**Axiom 1**: There exists \( L \)

**Axiom 2**: The movement minimizes \( \int L \)

**Theorem 1**: (Lagrange equation): \[ \frac{\partial L}{\partial x} = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \]

**Axiom 3**: Invariance w.r.t. Lorentz change of frame

\[
\begin{align*}
x' &= (x-\nu t) \times \gamma \\
t' &= (t - \nu x/c^2) \times \gamma \\
\gamma &= 1 / \sqrt{1 - v^2 / c^2}
\end{align*}
\]
The Least Action Principle

*(relativistic setting – just for fun…)*

**Axiom 1:** There exists \( L \)

**Axiom 2:** The movement minimizes \( \int L \)

**Theorem 1:** (Lagrange equation):

\[
\frac{\partial L}{\partial x} = \frac{d}{dt} \frac{\partial L}{\partial \dot{x}}
\]

**Axiom 3:**

Invairance w.r.t. Lorentz change of frame

\[
x' = (x - vt) \times \gamma
\]

\[
t' = (t - vx/c^2) \times \gamma
\]

\[
\gamma = 1 / \sqrt{1 - v^2 / c^2}
\]

**Theorem 5:**

\[
E = \frac{1}{2} \gamma m v^2 + mc^2
\]
The Least Action Principle

*(quantum physics setting – just for fun…)*

In quantum mechanics non just the extreme path contributes to the probability amplitude

\[ K(B,A) = \sum \phi[x(t)] \text{ over all possible paths} \]

where

\[ \phi[x(t)] = A \exp \left( \frac{i}{\hbar} S[x(t)] \right) \]

Feynman's path integral formula

\[ K(B,A) = \int_{A}^{B} \exp \left( \frac{i}{\hbar} S[B,A] \right) D x(t) \]
Fluids
“Lagrange” point of view

Fluids

\[
\begin{bmatrix}
\dot{x}_i \\
\dot{y}_i
\end{bmatrix}
\]

\[
L(x_i, y_i, \dot{x}_i, \dot{y}_i, \ldots, \epsilon)
\]

\[
\mathbf{f}_i = m \frac{\dot{\mathbf{x}}_i}{\dot{t}_i}
\]
Fluids

“Lagrange” point of view

“Euler” point of view

\[ \begin{bmatrix} x_i \end{bmatrix} [x_i] \]

\[ \begin{bmatrix} y_i \end{bmatrix} [y_i] \]

\[ L(x, y, x_i, y_i, \ldots, t) \]

\[ f_i = \rho \left( \frac{\partial x}{\partial t} \right) \]

\[ \rho \text{ “nb particles per square”} \]
"Lagrange" point of view

\[ \begin{bmatrix} x_i \\ y_i \end{bmatrix} \]

\( v(x_i, y_i, t) \) speed of the particle under \( (x_i, y_i) \) at time \( t \)

\[ L(x_i, y_i, x_i', y_i', \ldots, t') \]

\[ f_i = \rho \begin{bmatrix} x_i' \\ y_i' \end{bmatrix} \]

\( \rho(x,y,t) \) "nb particles per square"

\( v(x,y,t) \) speed of the particle under "grid point" \((x,y)\) at time \( t \)
Q1: how to compute the acceleration of the particles from $v(x,y,t)$?
Q1: how to compute the acceleration of the particles from $v(x,y,t)$?

$$\frac{dv}{dt} = \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} \frac{dx}{dt} + \frac{\partial v}{\partial y} \frac{dy}{dt}$$
Q1: how to compute the acceleration of the particles from \( v(x,y,t) \)?

\[
\frac{dv}{dt} = \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} \frac{dx}{dt} + \frac{\partial v}{\partial y} \frac{dy}{dt}
\]

\[
= \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} v_x + \frac{\partial v}{\partial y} v_y
\]
Q1: how to compute the acceleration of the particles from $v(x,y,t)$?

$$\frac{dv}{dt} = \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} \frac{dx}{dt} + \frac{\partial v}{\partial y} \frac{dy}{dt}$$

$$= \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} v_x + \frac{\partial v}{\partial y} v_y$$

$$= \frac{\partial v}{\partial t} + v \cdot \nabla v$$
Q1: how to compute the acceleration of the particles from $v(x,y,t)$?

\[
\frac{dv}{dt} = \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} \frac{dx}{dt} + \frac{\partial v}{\partial y} \frac{dy}{dt}
\]

\[
= \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x} v_x + \frac{\partial v}{\partial y} v_y
\]

\[
= \frac{\partial v}{\partial t} + v \cdot \nabla v
\]

$\rho(x,y,t)$ “nb particles per square”

$v(x,y,t)$ speed of the particle under “grid point” $(x,y)$ at time $t$
Fluids

“Euler” point of view

Q1: how to compute the acceleration of the particles from \( v(x,y,t) \) ?

\[
\frac{dv}{dt} = \frac{\partial v}{\partial t} + v \cdot \nabla v
\]

Q2: incompressible fluids?

\( \rho(x,y,t) \) “nb particles per square”

\( v(x,y,t) \) speed of the particle under “grid point” \((x,y)\) at time \(t\)
Q1: how to compute the acceleration of the particles from \( v(x,y,t) \)?

\[
\frac{dv}{dt} = \frac{\partial v}{\partial t} + v \cdot \nabla v
\]

Q2: incompressible fluids?

\( \rho(x,y,t) \) “nb particles per square”

\( v(x,y,t) \) speed of the particle under “grid point” \((x,y)\) at time \(t\)
Q1: how to compute the acceleration of the particles from \( v(x,y,t) \) ?

\[
\frac{dv}{dt} = \frac{\partial v}{\partial t} + v \cdot \nabla v
\]

Q2: incompressible fluids?

\( \rho(x,y,t) \) “nb particles per square”

\( v(x,y,t) \) speed of the particle under “grid point” \((x,y)\) at time \(t\)

Fluids

“Euler” point of view

what goes in = what goes out
Q1: how to compute the acceleration of the particles from $v(x,y,t)$?

$$\frac{dv}{dt} = \frac{\partial v}{\partial t} + v \cdot \nabla v$$

Q2: incompressible fluids?

$\text{div}(v) = 0$

what goes in = what goes out

$\rho(x,y,t)$ “nb particles per square”

$v(x,y,t)$ speed of the particle under “grid point” (x,y) at time t
Q1: how to compute the acceleration of the particles from \( v(x,y,t) \) ?

\[
\frac{dv}{dt} = \frac{\partial v}{\partial t} + v \nabla v
\]

Q2: incompressible fluids?

\[ \text{div}(v) = 0 \]

Q3: mass preservation?

\( \rho(x,y,t) \) “nb particles per square”

\( v(x,y,t) \) speed of the particle under “grid point” \((x,y)\) at time \(t\)
Fluids

“Euler” point of view

Q1: how to compute the acceleration of the particles from \( v(x,y,t) \)?

\[
\frac{dv}{dt} = \frac{\partial v}{\partial t} + v \cdot \nabla v
\]

Q2: incompressible fluids?

\[
\text{div}(v) = 0
\]

Q3: mass preservation?

\[
\frac{d \rho}{dt} = - \text{div}(\rho v)
\]

(Continuity equation)

\( \rho(x,y,t) \) “nb particles per square”

\( v(x,y,t) \) speed of the particle under “grid point” (x,y) at time t
Fluids

Start with Lagrange coordinates:

“particle trajectories”: \( X(t,x) \)

\[ (\rho = \text{cte}) \]

Minimize 

\[ \int_{t_1}^{t_2} \int_{\Omega} \left\| \frac{\partial X(t,x)}{\partial t} \right\|^2 \, dx \, dt \]

s.t. \( X \) satisfies mass preservation

(\( X \) is measure-preserving, more on this later...)
Fluids

Start with Lagrange coordinates: "particle trajectories": $X(t,x)$

$(\rho = \text{cte})$

Minimize Action: $\frac{1}{2} \int_{t_1}^{t_2} \int_{\Omega} \left\| \frac{\partial X(t,x)}{\partial t} \right\|^2 \, dx \, dt$

s.t. $X$ satisfies mass preservation
(X is measure-preserving, more on this later...)
Start with Lagrange coordinates: “particle trajectories”: \( X(t,x) \)

\( (p = \text{cte}) \)

Minimize Action:

\[
\int_{t_1}^{t_2} \int_{\Omega} \left\| \frac{\partial X(t,x)}{\partial t} \right\|^2 \, dx \, dt
\]

s.t. \( X \) satisfies mass preservation

(\( X \) is measure-preserving, more on this later...)
Fluids

Start with Lagrange coordinates: "particle trajectories": \( X(t, x) \)

\[(\rho = \text{cte}) \]

Minimize

\[
\frac{1}{2} \int_{t_1}^{t_2} \int_{\Omega} \left\| \frac{\partial X(t, x)}{\partial t} \right\|^2 \, dx \, dt
\]

s.t. \( X \) satisfies mass preservation

(X is measure-preserving, more on this later…)
Fluids

Start with Lagrange coordinates: “particle trajectories”: $X(t,x)$

$(\rho = \text{cte})$

Minimize Action: $\frac{1}{2} \int_{t_1}^{t_2} \int_{\Omega} \left\| \frac{\partial X(t,x)}{\partial t} \right\|^2 \, dx \, dt$

s.t. $X$ satisfies mass preservation ($X$ is measure-preserving, more on this later...)

Acceleration of the particle “under the grid”
Fluids

Start with Lagrange coordinates: “particle trajectories”: $X(t, x)$

$(\rho = \text{cte})$

Minimize Action:

$$\frac{1}{2} \int_{t_1}^{t_2} \int_{\Omega} \left\| \frac{\partial X(t, x)}{\partial t} \right\|^2 \, dx \, dt$$

s.t. $X$ satisfies mass preservation ($X$ is measure-preserving, more on this later…)

The “Lagrange multiplier” for the constraint = pressure
Fluids – Benamou Brenier

\[ \rho_1 \xrightarrow{?} T \xrightarrow{} \rho_2 \]
Minimize

\[ A_{(\rho,v)} = (t_2 - t_1) \int_{t_1}^{t_2} \int_{\Omega} \rho(x,t) \|v(t,x)\|^2 dx dt \]

s.t. \( \rho(t_1,.) = \rho_1 \); \( \rho(t_2,.) = \rho_2 \); \( \frac{d\rho}{dt} = - \text{div}(\rho v) \)

**Fluids – Benamou Brenier**
Fluids – Benamou Brenier

Minimize
\[ A(\rho,v) = \int_{t_1}^{t_2} \int_{\Omega} \rho(x,t) \|v(t,x)\|^2 \, dx \, dt \]

s.t. \( \rho(t_1,\cdot) = \rho_1 \); \( \rho(t_2,\cdot) = \rho_2 \); \( \frac{d\rho}{dt} = -\text{div}(\rho v) \)

Minimize \( C(T) = \int_{\Omega} \rho_1(x) \| x - T(x) \|^2 \, dx \)

s.t. \( T \) is measure-preserving
2

Optimal Transport
an elementary introduction
Part. 2 Optimal Transport – Monge’s problem

Two measures $\mu$, $\nu$ such that $\int_X d\mu(x) = \int_Y d\nu(x)$
Part. 2 Optimal Transport – Monge’s problem

A map $T$ is a transport map between $\mu$ and $\nu$ if

$$\mu(T^{-1}(B)) = \nu(B)$$

for any Borel subset $B$ of $Y$. 

$$(X;\mu)$$

$$(Y;\nu)$$
A map $T$ is a \textit{transport map} between $\mu$ and $\nu$ if $\mu(T^{-1}(B)) = \nu(B)$ for any Borel subset $B$. 
Part. 2 Optimal Transport – Monge’s problem

A map $T$ is a transport map between $\mu$ and $\nu$ if $\mu(T^{-1}(B)) = \nu(B)$ for any Borel subset $B$. 

$(X; \mu)$ $\xrightarrow{T}$ $(Y; \nu)$
A map $T$ is a transport map between $\mu$ and $\nu$ if $\mu(T^{-1}(B)) = \nu(B)$ for any Borel subset $B$ (or $\nu = T#\mu$ the pushforward of $\mu$)
Monge’s problem (1787):

Find a transport map $T$ that minimizes $C(T) = \int_X \| x - T(x) \|^2 \, d\mu(x)$
Part. 2 Optimal Transport – Monge’s problem

Monge’s problem:

Find a transport map $T$ that minimizes $C(T) = \int_X \| x - T(x) \|^2 \, d\mu(x)$

- Difficult to study
- If $\mu$ has an atom (isolated Dirac), it can only be mapped to another Dirac ($T$ needs to be a map)
Part. 2 Optimal Transport – Kantorovich

Monge’s problem:
Find a transport map $T$ that minimizes $C(T) = \int_X ||x - T(x)||^2 \, d\mu(x)$

Kantorovich’s problem (1942):
Find a measure $\gamma$ defined on $X \times Y$ such that $\int_{x \in X} d\gamma(x,y) = d\nu(y)$
and $\int_{y \in Y} d\gamma(x,y) = d\mu(x)$
that minimizes $\int\int_{X \times Y} ||x - y||^2 \, d\gamma(x,y)$
Part. 2 Optimal Transport – Kantorovich

**Monge’s problem:**
Find a transport map $T$ that minimizes $C(T) = \int_X || x - T(x) ||^2 d\mu(x)$

**Kantorovich’s problem:**
Find a measure $\gamma$ defined on $X \times Y$
such that $\int_{X \text{ in } X} d\gamma(x,y) = d\nu(y)$
and $\int_{Y \text{ in } Y} d\gamma(x,y) = d\mu(x)$
that minimizes $\iint_{X \times Y} || x - y ||^2 d\gamma(x,y)$

“How much sand goes from $x$ to $y$”
Monge’s problem:
Find a transport map $T$ that minimizes $C(T) = \int_X \| x - T(x) \|^2 \, d\mu(x)$

Kantorovich’s problem:
Find a measure $\gamma$ defined on $X \times Y$
such that $\int_X d\gamma(x,y) = d\nu(y)$
and $\int_Y d\gamma(x,y) = d\mu(x)$
that minimizes $\int\int_{X \times Y} \| x - y \|^2 \, d\gamma(x,y)$

Everything that is transported from $x$ sums to “$\mu(x)$”
Part. 2 Optimal Transport – Kantorovich

**Monge’s problem:**

Find a transport map $T$ that minimizes $C(T) = \int_X ||x - T(x)||^2 \, d\mu(x)$

**Kantorovich’s problem:**

Find a measure $\gamma$ defined on $X \times Y$

such that $\int_{x \in X} d\gamma(x,y) = d\nu(y)$

and $\int_{y \in Y} d\gamma(x,y) = d\mu(x)$

that minimizes $\iint_{X \times Y} ||x - y||^2 \, d\gamma(x,y)$

Everything that is transported to $y$ sums to “$\nu(y)$”
Transport plan – example 1/2 : translation of a segment
Transport plan – example 1/2: translation of a segment
Transport plan – example 2/2: spitting a segment
Part. 2 Optimal Transport – Duality

\[
\inf \int c(x, y) \, d\gamma
\]

\[\forall \gamma \in \mathcal{P}(X, Y), \quad \int_B d\mu = \int_{\gamma \circ x} d\gamma \quad (\rho \# \delta = \mu)\]

\[\forall \gamma \in \mathcal{P}(Y, X), \quad \int_B d\gamma = \int_{x \circ \gamma} d\gamma \quad (\rho \# \delta = \mu)\]
Duality is easier to understand with a discrete version
Then we’ll go back to the continuous setting.
Part. 2 Optimal Transport – Duality

(DMK):
Min $\langle c, \gamma \rangle$

$s.t.
\begin{align*}
P_1 \gamma &= u \\
P_2 \gamma &= v \\
\gamma &\geq 0
\end{align*}$
Part. 2 Optimal Transport – Duality

(DMK):
Min $<c, \gamma>$

\[ \begin{align*}
\quad & P_1 \gamma = u \\
\text{s.t.} & \\
\quad & P_2 \gamma = v \\
\quad & \gamma \geq 0
\end{align*} \]

$\gamma = \begin{bmatrix}
\gamma_{11} \\
\gamma_{12} \\
\vdots \\
\gamma_{1n} \\
\gamma_{22} \\
\vdots \\
\gamma_{2n} \\
\vdots \\
\gamma_{mn}
\end{bmatrix} \in \mathbb{R}^{mn}$
Part. 2 Optimal Transport – Duality

(DMK):
Min $\langle c, \gamma \rangle$

\[
\begin{align*}
P_1 \gamma &= u \\
P_2 \gamma &= v \\
\gamma &\geq 0
\end{align*}
\]

\[
c = \begin{bmatrix} c_{11} \\ c_{12} \\ \vdots \\ c_{1n} \\ c_{21} \\ \vdots \\ c_{2n} \\ \vdots \\ c_{mn} \end{bmatrix} \in \mathbb{IR}^{mn}
\]

\[
\gamma = \begin{bmatrix} \gamma_{11} \\ \gamma_{12} \\ \vdots \\ \gamma_{1n} \\ \gamma_{21} \\ \vdots \\ \gamma_{2n} \\ \vdots \\ \gamma_{mn} \end{bmatrix} \in \mathbb{IR}^{mn}
\]
Part. 2 Optimal Transport – Duality

(DMK):
\[
\begin{align*}
\text{Min} \quad & \langle c, \gamma \rangle \\
\text{s.t.} \quad & P_1 \gamma = u \\
& P_2 \gamma = v \\
& \gamma \geq 0
\end{align*}
\]

\[
c = \begin{bmatrix}
c_{11} \\
\vdots \\
c_{1n} \\
\vdots \\
c_{22} \\
\vdots \\
\vdots \\
c_{mn}
\end{bmatrix} \in \mathbb{R}^{mn}
\]

\[
\gamma = \begin{bmatrix}
\gamma_{11} \\
\vdots \\
\gamma_{1n} \\
\vdots \\
\gamma_{22} \\
\vdots \\
\vdots \\
\gamma_{mn}
\end{bmatrix} \in \mathbb{R}^{mn}
\]

\[
c_{ij} = \| x_i - y_j \|^2
\]
Part. 2 Optimal Transport – Duality

(DMK):
\[
\begin{align*}
\text{Min } & \langle c, \gamma \rangle \\
\text{s.t.} & \\
P_1 \gamma = u \\
\gamma \geq 0
\end{align*}
\]

\[
c = \begin{bmatrix}
c_{11} \\
c_{12} \\
\vdots \\
c_{1n} \\
c_{22} \\
\vdots \\
c_{2n} \\
\vdots \\
\vdots \\
c_{mn}
\end{bmatrix} \in \mathbb{IR}^{mn}
\]

\[
\gamma = \begin{bmatrix}
\gamma_{11} \\
\gamma_{12} \\
\vdots \\
\gamma_{1n} \\
\gamma_{22} \\
\vdots \\
\vdots \\
\vdots \\
\gamma_{mn}
\end{bmatrix} \in \mathbb{IR}^{mn}
\]

\[
c_{ij} = \| x_i - y_j \|^2
\]
Part. 2 Optimal Transport – Duality

(DMK):
Min $<c, \gamma>$

$s.t.$
$P_1 \gamma = u$

$P_2 \gamma = v$

$\gamma \geq 0$

$c_{ij} = ||x_i - y_j||^2$

$c = \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ c_{12} & \cdots & c_{2n} \\ \vdots & \ddots & \vdots \\ c_{1n} & \cdots & c_{nn} \end{bmatrix} \in \mathbb{IR}^{mn}$

$\gamma = \begin{bmatrix} \gamma_{11} & \cdots & \gamma_{1n} \\ \gamma_{12} & \cdots & \gamma_{2n} \\ \vdots & \ddots & \vdots \\ \gamma_{1n} & \cdots & \gamma_{nn} \end{bmatrix} \in \mathbb{IR}^{mn}$
Part. 2 Optimal Transport – Duality

\[(DMK)\]:

\[
\begin{align*}
\text{Min } & \langle c, \gamma \rangle \\
\text{s.t. } & \begin{cases} 
P_1 \gamma = u \\
P_2 \gamma = v \\
\gamma \geq 0
\end{cases}
\end{align*}
\]

\[
c = \begin{bmatrix} 
\gamma_{11} \\
\gamma_{12} \\
\vdots \\
\gamma_{1n} \\
\gamma_{22} \\
\vdots \\
\vdots \\
\gamma_{mn}
\end{bmatrix} \in \mathbb{R}^{mn},
\]

\[
c = \begin{bmatrix} 
c_{11} \\
c_{12} \\
\vdots \\
c_{1n} \\
c_{22} \\
\vdots \\
\vdots \\
c_{mn}
\end{bmatrix} \in \mathbb{R}^{mn}.
\]
Part. 2 Optimal Transport – Duality

\[ <u, v> \text{ denotes the dot product between } u \text{ and } v \]

\[
\begin{align*}
\text{(DMK):} \\
\text{Min } <c, \gamma> \\
\text{s.t. } \\
P_1 \gamma = u \\
P_2 \gamma = v \\
\gamma \geq 0
\end{align*}
\]

Consider \( L(\varphi, \psi) = <c, \gamma> - <\varphi, P_1 \gamma - u> - <\psi, P_2 \gamma - v> \)
Part. 2 Optimal Transport – Duality

Consider \( L(\varphi, \psi) = \langle c, \gamma \rangle - \langle \varphi, P_1 \gamma - u \rangle - \langle \psi, P_2 \gamma - v \rangle \)

Remark: \( \text{Sup} \left[ L(\varphi, \psi) \right] = \langle c, \gamma \rangle \) if \( P_1 \gamma = u \) and \( P_2 \gamma = v \)

\( \varphi \in \mathbb{IR}^m \)
\( \psi \in \mathbb{IR}^n \)

(DMK):
\[
\begin{align*}
\text{Min} & \quad \langle c, \gamma \rangle \\
\text{s.t.} & \quad P_1 \gamma = u \\
& \quad P_2 \gamma = v \\
& \quad \gamma \geq 0
\end{align*}
\]
Part. 2 Optimal Transport – Duality

(DMK):
Min  \langle c, \gamma \rangle
\begin{align*}
P_1 \gamma &= u \\
P_2 \gamma &= v \\
\gamma &\geq 0
\end{align*}

Consider \( L(\phi, \psi) = \langle c, \gamma \rangle - \langle \phi, P_1 \gamma - u \rangle - \langle \psi, P_2 \gamma - v \rangle \)

Remark: \( \text{Sup}[ L(\phi, \psi) ] = \langle c, \gamma \rangle \) if \( P_1 \gamma = u \) and \( P_2 \gamma = v \)
\[
\begin{align*}
\phi &\in \mathbb{IR}^m \\
\psi &\in \mathbb{IR}^n
\end{align*}
\]
\[= +\infty \text{ otherwise} \]
Part. 2 Optimal Transport – Duality

Consider \( L(\varphi, \psi) = \langle c, \gamma \rangle - \langle \varphi, P_1 \gamma - u \rangle - \langle \psi, P_2 \gamma - v \rangle \)

Remark: \( \text{Sup}[ L(\varphi, \psi) ] = \langle c, \gamma \rangle \) if \( P_1 \gamma = u \) and \( P_2 \gamma = v \)
\begin{align*}
\varphi &\in \mathbb{R}^m \\
\psi &\in \mathbb{R}^n \\
&= +\infty \text{ otherwise}
\end{align*}

Consider now: \( \text{Inf} \left[ \text{Sup}[ L(\varphi, \psi) ] \right] \)
\begin{align*}
\gamma &\geq 0 \\
\varphi &\in \mathbb{R}^m \\
\psi &\in \mathbb{R}^n
\end{align*}
Part. 2 Optimal Transport – Duality

Consider \( L(\varphi, \psi) = \langle c, \gamma \rangle - \langle \varphi, P_1 \gamma - u \rangle - \langle \psi, P_2 \gamma - v \rangle \)

Remark: \( \text{Sup}[ L(\varphi, \psi) ] = \langle c, \gamma \rangle \) if \( P_1 \gamma = u \) and \( P_2 \gamma = v \)

\[ \varphi \in \mathbb{IR}^m \]
\[ \psi \in \mathbb{IR}^n \]

\[ = +\infty \text{ otherwise} \]

Consider now: \( \text{Inf} \left[ \text{Sup}[ L(\varphi, \psi) ] \right] = \text{Inf} \left[ \langle c, \gamma \rangle \right] \)

\[ \gamma \geq 0 \]
\[ \varphi \in \mathbb{IR}^m \]
\[ \psi \in \mathbb{IR}^n \]

\[ \gamma \geq 0 \]
\[ P_1 \gamma = u \]
\[ P_2 \gamma = v \]
Part. 2 Optimal Transport – Duality

Consider $L(\varphi, \psi) = \langle c, \gamma \rangle - \langle \varphi, P_1 \gamma - u \rangle - \langle \psi, P_2 \gamma - v \rangle$

Remark: $\sup\{ L(\varphi, \psi) \} = \langle c, \gamma \rangle$ if $P_1 \gamma = u$ and $P_2 \gamma = v$

Consider now: $\inf\{ \sup\{ L(\varphi, \psi) \} \} = \inf\{ \langle c, \gamma \rangle \}$

(DMK): $\begin{align*}
\min \langle c, \gamma \rangle \\
P_1 \gamma = u \\
\text{s.t.} \\
P_2 \gamma = v \\
\gamma \geq 0
\end{align*}$

$\varphi \in \mathbb{IR}^m$
$\psi \in \mathbb{IR}^n$

$= +\infty$ otherwise
Part. 2 Optimal Transport – Duality

(DMK):
Min $\langle c, \gamma \rangle$

\[
\begin{cases}
P_1 \gamma = u \\
P_2 \gamma = v \\
\gamma \geq 0
\end{cases}
\]

\[\inf \left[ \sup \left[ \langle c, \gamma \rangle - \langle \varphi, P_1 \gamma - u \rangle - \langle \psi, P_2 \gamma - v \rangle \right] \right]
\]

\[
\gamma \geq 0 \quad \varphi \in \mathbb{R}^m \quad \psi \in \mathbb{R}^n
\]
Part. 2 Optimal Transport – Duality

(DMK):
Min $\langle c, \gamma \rangle$

\[
\begin{align*}
    &\text{s.t.} \\
    &P_1 \gamma = u \\
    &P_2 \gamma = v \\
    &\gamma \geq 0
\end{align*}
\]

\[
\begin{align*}
    &\text{Inf} \left[ \text{Sup} [ \langle c, \gamma \rangle - \langle \phi, P_1 \gamma - u \rangle - \langle \psi, P_2 \gamma - v \rangle ] \right] \\
    &\gamma \geq 0 \\
    &\phi \in \mathbb{R}^m \\
    &\psi \in \mathbb{R}^n
\end{align*}
\]

\[
\begin{align*}
    &\text{Sup} [ \text{Inf} [ \langle c, \gamma \rangle - \langle \phi, P_1 \gamma - u \rangle - \langle \psi, P_2 \gamma - v \rangle ] ] \\
    &\phi \in \mathbb{R}^m \\
    &\gamma \geq 0 \\
    &\psi \in \mathbb{R}^n
\end{align*}
\]
Part. 2 Optimal Transport – Duality

(DMK):
\[
\begin{align*}
\text{Min } & \langle c, \gamma \rangle \\
\text{s.t. } & P_1 \gamma = u \\
& P_2 \gamma = v \\
& \gamma \geq 0
\end{align*}
\]

\[
\begin{aligned}
\inf \left[ \sup \left[ \langle c, \gamma \rangle - \langle \varphi, P_1 \gamma - u \rangle - \langle \psi, P_2 \gamma - v \rangle \right] \right] \\
\quad \gamma \geq 0 & \quad \varphi \in \mathbb{R}^m \\
\quad & \quad \psi \in \mathbb{R}^n
\end{aligned}
\]

Exchange Inf Sup

\[
\begin{aligned}
\sup \left[ \inf \left[ \langle c, \gamma \rangle - \langle \varphi, P_1 \gamma - u \rangle - \langle \psi, P_2 \gamma - v \rangle \right] \right] \\
\quad \varphi \in \mathbb{R}^m & \quad \gamma \geq 0 \\
\quad & \quad \psi \in \mathbb{R}^n
\end{aligned}
\]

Expand/Reorder/Collect

\[
\begin{aligned}
\sup \left[ \inf \left[ \langle \gamma, c-P_1^t \varphi - P_2^t \psi \rangle + \langle \varphi, u \rangle + \langle \psi, v \rangle \right] \right] \\
\quad \varphi \in \mathbb{R}^m & \quad \gamma \geq 0 \\
\quad & \quad \psi \in \mathbb{R}^n
\end{aligned}
\]
**Part. 2  Optimal Transport – Duality**

\[(DMK):\]
\[
\begin{align*}
\text{Min } & \langle c, \gamma \rangle \\
\text{s.t. } & P_1 \gamma = u \\
& P_2 \gamma = v \\
& \gamma \geq 0
\end{align*}
\]

\[
\begin{align*}
\text{Inf } \left[ \text{Sup} \left[ \langle c, \gamma \rangle - \langle \varphi, P_1 \gamma - u \rangle - \langle \psi, P_2 \gamma - v \rangle \right] \right] \\
\end{align*}
\]
\[
\begin{align*}
\gamma \geq 0 & \quad \varphi \in \mathbb{R}^m \\
\psi \in \mathbb{R}^n
\end{align*}
\]

**Exchange Inf Sup**

\[
\begin{align*}
\text{Sup} \left[ \text{Inf} \left[ \langle c, \gamma \rangle - \langle \varphi, P_1 \gamma - u \rangle - \langle \psi, P_2 \gamma - v \rangle \right] \right] \\
\end{align*}
\]
\[
\begin{align*}
\varphi \in \mathbb{R}^m & \quad \gamma \geq 0 \\
\psi \in \mathbb{R}^n
\end{align*}
\]

**Expand/Reorder/Collect**

\[
\begin{align*}
\text{Sup} \left[ \text{Inf} \left[ \langle \gamma, c-P_1^t \varphi - P_2^t \psi \rangle + \langle \varphi, u \rangle + \langle \psi, v \rangle \right] \right] \\
\end{align*}
\]
\[
\begin{align*}
\varphi \in \mathbb{R}^m & \quad \gamma \geq 0 \\
\psi \in \mathbb{R}^n
\end{align*}
\]

**Interpret**
Part. 2 Optimal Transport – Duality

\[(DMK):\]
\[
\begin{align*}
\text{Min } & \langle c, \gamma \rangle \\
\text{s.t. } & P_1 \gamma = u \\
& P_2 \gamma = v \\
& \gamma \geq 0
\end{align*}
\]

\[
\text{Sup}\left[\text{Inf}\left[ \langle \gamma, c-P_1^t \phi - P_2^t \psi \rangle + \langle \phi, u \rangle + \langle \psi, v \rangle \right]\right]
\]

\[
\phi \in \mathbb{IR}^m \quad \gamma \geq 0 \\
\psi \in \mathbb{IR}^n
\]

Interpret

\[
\text{Sup}\left[ \langle \phi, u \rangle + \langle \psi, v \rangle \right]
\]

\[(DDMK)\]

\[
\phi \in \mathbb{IR}^m \\
\psi \in \mathbb{IR}^n \\
P_1^t \phi + P_2^t \psi \leq c
\]
Part. 2 Optimal Transport – Duality

(DMK):
Min $<c, \gamma>$

\[
P_1 \gamma = u
\]
\[
P_2 \gamma = v
\]
\[
\gamma \geq 0
\]

Sup\[ \text{Inf}\[ <\gamma, c-P_1^t \varphi - P_2^t \psi > + <\varphi, u> + <\psi, v> \] ]

Sup\[ <\varphi, u> + <\psi, v> \] \quad (DDMK)

$\varphi \in \mathbb{R}^m \quad \gamma \geq 0$
$\psi \in \mathbb{R}^n$

$\varphi_i + \psi_j \leq c_{ij} \quad \forall (i,j)$
Part. 2 Optimal Transport – Kantorovich dual

Kantorovich’s problem:

Find a measure $\gamma$ defined on $X \times Y$ such that $\int_{x \in X} \gamma(x,y) = d\mu(x)$ and $\int_{y \in Y} \gamma(x,y) = d\nu(x)$ such that $\int_{x \in X} \int_{y \in Y} ||x - y||^2 \, d\gamma(x,y)$ minimizes.

Dual formulation of Kantorovich’s problem (Continuous):

Find two functions $\phi$ in $L^1(\mu)$ and $\psi$ in $L^1(\nu)$ such that for all $x,y$, $\phi(x) + \psi(y) \leq \frac{1}{2}||x - y||^2$ that maximize $\int_X \phi \, d\mu + \int_Y \psi \, d\nu$. 
Part. 2 Optimal Transport – Kantorovich dual

Kantorovich’s problem:

Find a measure \( \gamma \) defined on \( X \times Y \) such that
\[
\int_{x \in X} d\gamma(x,y) = d\mu(x)
\]
and
\[
\int_{y \in Y} d\gamma(x,y) = d\nu(x)
\]
that minimizes
\[
\iint_{X \times Y} \| x - y \|^2 \, d\gamma(x,y)
\]

Dual formulation of Kantorovich’s problem:

Find two functions \( \varphi \) in \( L^1(\mu) \) and \( \psi \) in \( L^1(\nu) \) such that for all \( x, y \),
\[
\varphi(x) + \psi(y) \leq \frac{1}{2} \| x - y \|^2
\]
that maximize
\[
\int_X \varphi \, d\mu + \int_Y \psi \, d\nu
\]
**Part. 2 Optimal Transport – Kantorovich dual**

*Kantorovich’s problem:*

Find a measure \( \gamma \) defined on \( X \times Y \) such that
\[
\int_{x \in X} d\gamma(x,y) = d\mu(x)
\]
and
\[
\int_{y \in Y} d\gamma(x,y) = d\nu(x)
\]
that minimizes
\[
\iint_{X \times Y} ||x - y||^2 d\gamma(x,y)
\]

*Dual formulation of Kantorovich’s problem:*

Find two functions \( \phi \) in \( L^1(\mu) \) and \( \psi \) in \( L^1(\nu) \) such that for all \( x,y \),
\[
\phi(x) + \psi(y) \leq \frac{1}{2} ||x - y||^2
\]
that maximize
\[
\int_X \phi d\mu + \int_Y \psi d\nu
\]

**Your point of view:**
Try to minimize transport cost

**Point of view of a “transport company”:**
Try to maximize transport price
Part. 2 Optimal Transport – Kantorovich dual

Kantorovich’s problem:

Find a measure $\gamma$ defined on $X \times Y$
such that $\int_{x \in X} d\gamma(x,y) = d\mu(x)$
and $\int_{y \in Y} d\gamma(x,y) = d\nu(x)$

that minimizes $\iint_{X \times Y} \|x - y\|^2 \, d\gamma(x,y)$

Dual formulation of Kantorovich’s problem:

Find two functions $\phi$ in $L^1(\mu)$ and $\psi$ in $L^1(\nu)$
Such that for all $x,y$, $\phi(x) + \psi(y) \leq \frac{1}{2} \|x - y\|^2$

that maximize $\int_X \phi(x) \, d\mu + \int_Y \psi(y) \, d\nu$

What they charge for loading at $x$

Your point of view:
Try to minimize transport cost
Part. 2  Optimal Transport – Kantorovich dual

**Kantorovich’s problem:**

Find a measure $\gamma$ defined on $X \times Y$ such that $\int_{x \in X} d\gamma(x,y) = d\mu(x)$ and $\int_{y \in Y} d\gamma(x,y) = d\nu(x)$

that minimizes $\iint_{X \times Y} ||x - y||^2 d\gamma(x,y)$

**Dual formulation of Kantorovich’s problem:**

Find two functions $\varphi$ in $L^1(\mu)$ and $\psi$ in $L^1(\nu)$ such that for all $x,y$, $\varphi(x) + \psi(y) \leq \frac{1}{2} ||x - y||^2$

that maximize $\int_X \varphi(x)d\mu + \int_Y \psi(y)d\nu$

Your point of view: Try to minimize transport cost

What they charge for loading at $x$

What they charge for unloading at $y$
Part. 2 Optimal Transport – Kantorovich dual

**Kantorovich’s problem:**

Find a measure $\gamma$ defined on $X \times Y$
such that $\int_{x \in X} d\gamma(x, y) = d\mu(x)$
and $\int_{y \in Y} d\gamma(x, y) = d\nu(x)$

that minimizes $\iint_{X \times Y} || x - y ||^2 d\gamma(x, y)$

**Dual formulation of Kantorovich’s problem:**

Find two functions $\varphi$ in $L^1(\mu)$ and $\psi$ in $L^1(\nu)$
Such that for all $x, y$, $\varphi(x) + \psi(y) \leq \frac{1}{2} || x - y ||^2$

that maximize $\int_X \varphi(x) d\mu + \int_Y \psi(y) d\nu$

Price (loading + unloading) cannot be greater than transport cost (else you do the job yourself)

Your point of view:
Try to minimize transport cost

What they charge for loading at $x$
What they charge for unloading at $y$
Dual formulation of Kantorovich’s problem:

Find two functions $\varphi$ in $L^1(\mu)$ and $\psi$ in $L^1(\nu)$
Such that for all $x, y$, $\varphi(x) + \psi(y) \leq \frac{1}{2}\|x - y\|^2$
that maximize $\int_X \varphi(x)d\mu + \int_Y \psi(y)d\nu$
Part. 2 Optimal Transport – c-conjugate functions

**Dual formulation of Kantorovich’s problem:**

Find two functions $\varphi$ in $L^1(\mu)$ and $\psi$ in $L^1(\nu)$

Such that for all $x,y$, $\varphi(x) + \psi(y) \leq \frac{1}{2}\|x - y\|^2$

that maximize $\int_X \varphi(x)d\mu + \int_Y \psi(y)d\nu$

If we got two functions $\varphi$ and $\psi$ that satisfy the constraint

Then it is possible to obtain a better solution by replacing $\psi$ with $\varphi^c$ defined by:

For all $y$, $\varphi^c(y) = \inf_{x \in X} \frac{1}{2}\|x - y\|^2 - \varphi(y)$
Part. 2 Optimal Transport – c-conjugate functions

Dual formulation of Kantorovich’s problem:

Find two functions $\phi$ in $L^1(\mu)$ and $\psi$ in $L^1(\nu)$ such that for all $x,y$, $\phi(x) + \psi(y) \leq \frac{1}{2}\|x - y\|^2$

that maximize $\int_X \phi(x)d\mu + \int_Y \psi(y)d\nu$

If we got two functions $\phi$ and $\psi$ that satisfy the constraint

Then it is possible to obtain a better solution by replacing $\psi$ with $\phi^c$ defined by:

For all $y$, $\phi^c(y) = \inf_{x \in X} \frac{1}{2}\|x - y\|^2 - \phi(y)$

- $\phi^c$ is called the **c-conjugate** function of $\phi$
- If there is a function $\phi$ such that $\psi = \phi^c$ then $\psi$ is said to be **c-concave**
- If $\psi$ is c-concave, then $\psi^{cc} = \psi$
Part. 2 Optimal Transport – c-conjugate functions

Dual formulation of Kantorovich’s problem:

Find a c-concave function $\psi$

that maximizes $\int_X \psi(x) d\mu + \int_Y \psi^c(y) d\nu$
**Part. 2** Optimal Transport – c-conjugate functions

*Dual formulation of Kantorovich’s problem:*

Find a c-concave function $\psi$

that maximizes $\int_X \psi(x) d\mu + \int_Y \psi^c(y) d\nu$

$\psi$ is called a “Kantorovich potential”
Dual formulation of Kantorovich’s problem:

Find a c-concave function $\psi$

that maximizes $\int_{X} \psi(x)d\mu + \int_{Y} \psi^c(y)d\nu$

$\psi$ is called a “Kantorovich potential”

What about our initial problem?
**Dual formulation of Kantorovich’s problem:**

Find a c-concave function $\psi$

that maximizes $\int_X \psi(x)d\mu + \int_Y \psi^c(y)d\nu$

$\psi$ is called a “**Kantorovich potential**”

What about our initial problem? (i.e., this is T() that we want to find …)
Theorem 1.

$$\forall (x, y) \in \partial_c \psi, \nabla \psi(x) - \nabla_x c(x, y) = 0$$

where $\partial_c \psi = \{(x, y)|\phi(x) + \psi(y) = c(x, y)\}$ denotes the so-called c-subdifferential of $\psi$. 
Theorem 1.

\[ \forall (x, y) \in \partial_c \psi, \nabla \psi(x) - \nabla_x c(x, y) = 0 \]

where \( \partial_c \psi = \{(x, y)|\phi(x) + \psi(y) = c(x, y)\} \) denotes the so-called c-subdifferential of \( \psi \).

Proof: see OTON, chap. 10.
Theorem 1.

\[ \forall (x, y) \in \partial_c \psi, \nabla \psi(x) - \nabla_x c(x, y) = 0 \]

where \( \partial_c \psi = \{(x, y) | \phi(x) + \psi(y) = c(x, y)\} \) denotes the so-called c-subdifferential of \( \psi \).

Proof: see OTON, chap. 10.

Heuristic argument (at the beginning of the same chapter):
Theorem 1.

\[ \forall (x, y) \in \partial c\psi, \nabla \psi(x) - \nabla_x c(x, y) = 0 \]

where \( \partial c\psi = \{(x, y) | \phi(x) + \psi(y) = c(x, y)\} \) denotes the so-called c-subdifferential of \( \psi \).

**Proof:** see OTON, chap. 10.

Heuristic argument (at the beginning of the same chapter):

Consider a point \((x, y)\) on the c-subdifferential \( \partial c\psi \), that satisfies \( \phi(y) + \psi(x) = c(x, y) \) (1).
Theorem 1.

\[ \forall (x, y) \in \partial_c \psi, \nabla \psi(x) - \nabla_x c(x, y) = 0 \]

where \( \partial_c \psi = \{(x, y) | \phi(x) + \psi(y) = c(x, y)\} \) denotes the so-called c-subdifferential of \( \psi \).

Proof: see OTON, chap. 10.

Heuristic argument (at the beginning of the same chapter):

Consider a point \((x, y)\) on the c-subdifferential \(\partial_c \psi\), that satisfies \(\phi(y) + \psi(x) = c(x, y)\) \((1)\).

By definition, \(\phi(y) = \psi_c(y) = \inf_x c(x, y) - \psi(x)\), thus \(\forall \tilde{x}, \phi(y) \leq c(\tilde{x}, y) - \psi(\tilde{x})\), or \(\phi(y) + \psi(\tilde{x}) \leq c(\tilde{x}, y)\) \((2)\).

By substituting \((1)\) into \((2)\), one gets \(\psi(\tilde{x}) - \psi(x) \leq c(\tilde{x}, y) - c(x, y)\) for all \(\tilde{x}\). 

\[ \tilde{x} \quad x \]
Part. 2 Optimal Transport – c-subdifferential

Theorem 1.

\[ \forall (x, y) \in \partial_c \psi, \nabla \psi(x) - \nabla_x c(x, y) = 0 \]

where \( \partial_c \psi = \{(x, y)|\phi(x) + \psi(y) = c(x, y)\} \) denotes the so-called c-subdifferential of \( \psi \).

**Proof:** see OTON, chap. 10.

Heuristic argument (at the beginning of the same chapter):

Consider a point \((x, y)\) on the c-subdifferential \( \partial_c \psi \), that satisfies \( \phi(y) + \psi(x) = c(x, y) \) (1).

By definition, \( \phi(y) = \psi^c(y) = \inf_x c(x, y) - \psi(x) \), thus \( \forall \tilde{x}, \phi(y) \leq c(\tilde{x}, y) - \psi(\tilde{x}) \), or \( \phi(y) + \psi(\tilde{x}) \leq c(\tilde{x}, y) \) (2).

By substituting (1) into (2), one gets \( \psi(\tilde{x}) - \psi(x) \leq c(\tilde{x}, y) - c(x, y) \) for all \( \tilde{x} \).

Imagine now that \( \tilde{x} \) follows a trajectory parameterized by \( \epsilon \) and starting at \( x \). One can compute the gradient along an arbitrary direction \( w \) by taking the limit when \( \epsilon \) tends to zero in the relation \( \frac{\psi(\tilde{x}) - \psi(x)}{\epsilon} \leq \frac{c(\tilde{x}, y) - c(x, y)}{\epsilon} \).
Part. 2 Optimal Transport – c-subdifferential

Theorem 1.

\[ \forall (x, y) \in \partial_c \psi, \nabla \psi(x) - \nabla_c c(x, y) = 0 \]

where \( \partial_c \psi = \{(x, y)|\phi(x) + \psi(y) = c(x, y)\} \) denotes the so-called c-subdifferential of \( \psi \).

**Proof:** see OTON, chap. 10.

Heuristic argument (at the beginning of the same chapter):

Consider a point \((x, y)\) on the c-subdifferential \( \partial_c \psi \), that satisfies \( \phi(y) + \psi(x) = c(x, y) \) (1).

By definition, \( \phi(y) = \psi^c(y) = \inf_x c(x, y) - \psi(x) \), thus \( \forall \tilde{x}, \phi(y) \leq c(\tilde{x}, y) - \psi(\tilde{x}) \), or \( \phi(y) + \psi(\tilde{x}) \leq c(\tilde{x}, y) \) (2). By substituting (1) into (2), one gets \( \psi(\tilde{x}) - \psi(x) \leq c(\tilde{x}, y) - c(x, y) \) for all \( \tilde{x} \).

Imagine now that \( \tilde{x} \) follows a trajectory parameterized by \( \epsilon \) and starting at \( x \). One can compute the gradient along an arbitrary direction \( w \) by taking the limit when \( \epsilon \) tends to zero in the relation \( \frac{\psi(\tilde{x}) - \psi(x)}{\epsilon} \leq \frac{c(\tilde{x}, y) - c(x, y)}{\epsilon} \).
Part. 2 Optimal Transport – c-subdifferential

Theorem 1.

\[ \forall (x, y) \in \partial_c \psi, \nabla \psi(x) - \nabla_x c(x, y) = 0 \]

where \( \partial_c \psi = \{(x, y)|\phi(x) + \psi(y) = c(x, y)\} \) denotes the so-called c-subdifferential of \( \psi \).

Proof: see OTON, chap. 10.

Heuristic argument (at the beginning of the same chapter):

Consider a point \((x, y)\) on the c-subdifferential \(\partial_c \psi\), that satisfies \(\phi(y) + \psi(x) = c(x, y)\) (1).

By definition, \(\phi(y) = \psi^c(y) = \inf_x c(x, y) - \psi(x)\), thus \(\forall \tilde{x}, \phi(y) \leq c(\tilde{x}, y) - \psi(\tilde{x})\), or \(\phi(y) + \psi(\tilde{x}) \leq c(\tilde{x}, y)\) (2).

By substituting (1) into (2), one gets \(\psi(\tilde{x}) - \psi(x) \leq c(\tilde{x}, y) - c(x, y)\) for all \(\tilde{x}\).

Imagine now that \(\tilde{x}\) follows a trajectory parameterized by \(\epsilon\) and starting at \(x\). One can compute the gradient along an arbitrary direction \(w\) by taking the limit when \(\epsilon\) tends to zero in the relation \( \frac{\psi(\tilde{x}) - \psi(x)}{\epsilon} \leq \frac{c(\tilde{x}, y) - c(x, y)}{\epsilon} \).
Part. 2 Optimal Transport – c-subdifferential

Theorem 1.

\[ \forall (x, y) \in \partial_c \psi, \nabla \psi(x) - \nabla_x c(x, y) = 0 \]

where \( \partial_c \psi = \{(x, y)|\phi(x) + \psi(y) = c(x, y)\} \) denotes the so-called c-subdifferential of \( \psi \).

Proof: see OTON, chap. 10.

Heuristic argument (at the beginning of the same chapter):

Consider a point \((x, y)\) on the c-subdifferential \(\partial_c \psi\), that satisfies \(\phi(y) + \psi(x) = c(x, y)\) (1). By definition, \(\phi(y) = \psi^c(y) = \inf_x c(x, y) - \psi(x)\), thus \(\forall \tilde{x}, \phi(y) \leq c(\tilde{x}, y) - \psi(\tilde{x})\), or \(\phi(y) + \psi(\tilde{x}) \leq c(\tilde{x}, y)\) (2). By substituting (1) into (2), one gets \(\psi(\tilde{x}) - \psi(x) \leq c(\tilde{x}, y) - c(x, y)\) for all \(\tilde{x}\).

Imagine now that \(\tilde{x}\) follows a trajectory parameterized by \(\epsilon\) and starting at \(x\). One can compute the gradient along an arbitrary direction \(w\) by taking the limit when \(\epsilon\) tends to zero in the relation \(\frac{\psi(\tilde{x}) - \psi(x)}{\epsilon} \leq \frac{c(\tilde{x}, y) - c(x, y)}{\epsilon}\). Thus we have \(\nabla \psi(x) \cdot w \leq \nabla_x c(x, y) \cdot w\)
Part. 2 Optimal Transport – c-subdifferential

Theorem 1.

\[ \forall (x, y) \in \partial_c \psi, \nabla \psi(x) - \nabla_x c(x, y) = 0 \]

where \( \partial_c \psi = \{(x, y)|\phi(x) + \psi(y) = c(x, y)\} \) denotes the so-called c-subdifferential of \( \psi \).

Proof: see OTON, chap. 10.

Heuristic argument (at the beginning of the same chapter):

Consider a point \((x, y)\) on the c-subdifferential \( \partial_c \psi \), that satisfies \( \phi(y) + \psi(x) = c(x, y) \) (1). By definition, \( \phi(y) = \psi^c(y) = \inf_x c(x, y) - \psi(x) \), thus \( \forall \tilde{x}, \phi(y) \leq c(\tilde{x}, y) - \psi(\tilde{x}) \), or \( \phi(y) + \psi(\tilde{x}) \leq c(\tilde{x}, y) \) (2).

By substituting (1) into (2), one gets \( \psi(\tilde{x}) - \psi(x) \leq c(\tilde{x}, y) - c(x, y) \) for all \( \tilde{x} \).

Imagine now that \( \tilde{x} \) follows a trajectory parameterized by \( \epsilon \) and starting at \( x \). One can compute the gradient along an arbitrary direction \( w \) by taking the limit when \( \epsilon \) tends to zero in the relation

\[
\frac{\psi(\tilde{x}) - \psi(x)}{\epsilon} \leq \frac{c(\tilde{x}, y) - c(x, y)}{\epsilon}.
\]

Thus we have \( \nabla \psi(x) \cdot w \leq \nabla_x c(x, y) \cdot w \)

The same derivation can be done with \(-w\) instead of \( w \), and one gets:

\[
\forall w, \nabla \psi(x) \cdot w = \nabla_x c(x, y) \cdot w, \text{ thus } \forall (x, y) \in \partial_c \psi, \nabla \psi(x) - \nabla_x c(x, y) = 0.
\]
Dual formulation of Kantorovich’s problem:

Find a $c$-concave function $\psi$

that maximizes $\int_X \psi(x) d\mu + \int_Y \psi^c(y) d\nu$

In the $L_2$ case, i.e. $c(x, y) = 1/2 \|x - y\|^2$, we have $\forall (x, y) \in \partial c \psi, \nabla \psi(x) + y - x = 0$, thus, whenever the optimal transport map $T$ exists, we have $T(x) = x - \nabla \psi(x) = \nabla (\|x\|^2/2 - \psi(x))$. 
Dual formulation of Kantorovich’s problem:

Find a $c$-concave function $\psi$ that maximizes $\int_X \psi(x)d\mu + \int_Y \psi^c(y)d\nu$

In the $L_2$ case, i.e. $c(x, y) = \frac{1}{2}\|x - y\|^2$, we have $\forall (x, y) \in \partial_c \psi, \nabla \psi(x) + y - x = 0$, thus, whenever the optimal transport map $T$ exists, we have $T(x) = x - \nabla \psi(x) = \nabla(\|x\|^2/2 - \psi(x))$.

$\nabla \overline{\psi}(x)$ with $\overline{\psi}(x) := (\frac{1}{2} x^2 - \psi(x))$
Part. 2 Optimal Transport – convexity

Dual formulation of Kantorovich’s problem:

Find a $c$-concave function $\psi$

that maximizes $\int_X \psi(x) d\mu + \int_Y \psi^c(y) d\nu$

In the $L_2$ case, i.e. $c(x, y) = 1/2 \|x - y\|^2$, we have $\forall (x, y) \in \partial_c \psi, \nabla \psi(x) + y - x = 0$, thus, whenever the optimal transport map $T$ exists, we have $T(x) = x - \nabla \psi(x) = \nabla (\|x\|^2/2 - \psi(x))$.

$\nabla \overline{\psi}(x)$ with $\overline{\psi}(x) := (1/2 \ x^2 - \psi(x))$

$\overline{\psi}$ is convex
Part. 2 Optimal Transport – convexity

Dual formulation of Kantorovich’s problem:

Find a $c$-concave function $\psi$ that maximizes $\int_X \psi(x) d\mu + \int_Y \psi^c(y) d\nu$

In the $L_2$ case, i.e. $c(x, y) = 1/2\|x - y\|^2$, we have $\forall (x, y) \in \partial_c \psi, \nabla \psi(x) + y - x = 0$, thus, whenever the optimal transport map $T$ exists, we have $T(x) = x - \nabla \psi(x) = \nabla(\|x\|^2/2 - \psi(x))$.

$$\nabla \psi(x) \text{ with } \overline{\psi}(x) := (1/2 \ x^2 - \psi(x))$$

$\overline{\psi}$ is convex

Proof.

$$\psi(x) = \inf_y \frac{|x-y|^2}{2} - \phi(y)$$
$$= \inf_y \frac{\|x\|^2}{2} - x \cdot y + \frac{\|y\|^2}{2} - \phi(y)$$

$$-\overline{\psi}(x) = \phi(x) - \frac{\|x\|^2}{2} = \inf_y -x \cdot y + \left(\frac{\|y\|^2}{2} - \phi(y)\right)$$

$$\overline{\psi}(x) = \sup_y x \cdot y - \left(\frac{\|y\|^2}{2} - \phi(y)\right)$$
Part. 2 Optimal Transport – convexity

Dual formulation of Kantorovich’s problem:

Find a c-concave function $\psi$ that maximizes $\int_X \psi(x) d\mu + \int_Y \psi^c(y) d\nu$

In the $L_2$ case, i.e. $c(x, y) = 1/2 \|x - y\|^2$, we have $\forall (x, y) \in \partial_c \psi, \nabla \psi(x) + y - x = 0$, thus, whenever the optimal transport map $T$ exists, we have $T(x) = x - \nabla \psi(x) = \nabla(\|x\|^2/2 - \psi(x))$.

$\nabla \overline{\psi}(x)$ with $\overline{\psi}(x) := (1/2 x^2 - \psi(x))$ $\overline{\psi}$ is convex

$$
\psi(x) = \inf_y \frac{|x-y|^2}{2} - \phi(y) \\
= \inf_y \frac{\|x\|^2}{2} - x \cdot y + \frac{\|y\|^2}{2} - \phi(y) \\
-\overline{\psi}(x) = \phi(x) - \frac{\|x\|^2}{2} = \inf_y -x \cdot y + \left(\frac{\|y\|^2}{2} - \phi(y)\right) \\
\overline{\psi}(x) = \sup_y x \cdot y - \left(\frac{\|y\|^2}{2} - \phi(y)\right)
$$
Part. 2 Optimal Transport – no collision

*Dual formulation of Kantorovich’s problem:*

Find a c-concave function $\psi$

that maximizes $\int_X \psi(x) d\mu + \int_Y \psi^c(y) d\nu$

If $T(.)$ exists, then

$T(x) = x - \text{grad } \psi(x) = \text{grad } (\frac{1}{2} x^2 - \psi(x))$

$\overline{\psi}$ is convex

$\text{grad } \overline{\psi} (x)$

*Two transported particles cannot “collide”*
Part. 2 Optimal Transport – no collision

**Dual formulation of Kantorovich’s problem:**

Find a c-concave function $\psi$

that maximizes $\int_X \psi(x) d\mu + \int_Y \psi^c(y) d\nu$

If $T(.)$ exists, then

$T(x) = x - \text{grad } \psi(x) = \text{grad } (\frac{1}{2} x^2 - \psi(x))$

$\overline{\psi}$ is convex

$\text{grad } \overline{\psi}(x)$

**Two transported particles cannot “collide”**

*Proof.* By contradiction, suppose that you have $t \in (0, 1)$ and $x_1 \neq x_2$ such that:

$$(1 - t)x_1 + tT(x_1) = (1 - t)x_2 + tT(x_2)$$
Part. 2 Optimal Transport – no collision

**Dual formulation of Kantorovich’s problem:**

Find a $c$-concave function $\psi$

that maximizes $\int_X \psi(x) d\mu + \int_Y \psi^c(y) d\nu$

If $T(.)$ exists, then

$$T(x) = x - \text{grad } \psi(x) = \text{grad } (\frac{1}{2} x^2 - \psi(x))$$

$\text{grad } \psi(x)$

**Two transported particles cannot “collide”**

*Proof.* By contradiction, suppose that you have $t \in (0, 1)$ and $x_1 \neq x_2$ such that:

$$(1 - t)x_1 + tT(x_1) = (1 - t)x_2 + tT(x_2)$$

$$(1 - t)x_1 + t\nabla \psi(x_1) = (1 - t)x_2 + t\nabla \psi(x_2)$$

$$(1 - t)(x_1 - x_2) + t(\nabla \psi(x_1) - \nabla \psi(x_2)) = 0$$

$\forall v, (1 - t)v \cdot (x_1 - x_2) + tv \cdot (\nabla \psi(x_1) - \nabla \psi(x_2)) = 0$

take $v = (x_1 - x_2)$

$$(1 - t)\|x_1 - x_2\|^2 + t(x_1 - x_2) \cdot (\nabla \psi(x_1) - \nabla \psi(x_2)) = 0$$
Part. 2 Optimal Transport – Monge-Ampere

Dual formulation of Kantorovich’s problem:

Find a c-concave function $\psi$

that maximizes $\int_X \psi(x) d\mu + \int_Y \psi^c(y) d\nu$

What about our initial problem? If $T(.)$ exists, then one can show that:

$T(x) = x - \text{grad } \psi(x) = \text{grad } \left( \frac{1}{2} x^2 - \psi(x) \right)$

grad $\overline{\psi}(x)$ with $\overline{\psi}(x) := \left( \frac{1}{2} x^2 - \psi(x) \right)$

for all borel set $A$, $\int_A d\mu = \int_{T(A)} (|JT|) d\nu$ (change of variable)

Jacobian of $T$ (1st order derivatives)
**Part. 2 Optimal Transport – Monge-Ampere**

*Dual formulation of Kantorovich’s problem:*

Find a c-concave function $\psi$

that maximizes $\int_X \psi(x) d\mu + \int_Y \psi^c(y) dv$

What about our initial problem? If $T(.)$ exists, then one can show that:

$T(x) = x - \text{grad } \psi(x) = \text{grad } (\frac{1}{2} x^2 - \psi(x))$

\[\text{grad } \overline{\psi}(x) \text{ with } \overline{\psi}(x) := \left(\frac{1}{2} x^2 - \psi(x)\right)\]

for all borel set $A$, $\int_A d\mu = \int_{T(A)} (|JT|) d\nu = \int_{T(A)} (H \overline{\psi}) d\nu$

Det. of the Hessian of $\overline{\psi}$ (2nd order derivatives)
Part. 2 Optimal Transport – Monge-Ampère

**Dual formulation of Kantorovich’s problem:**

Find a c-concave function $\psi$

that maximizes $\int_X \psi(x) d\mu + \int_Y \psi^c(y) d\nu$

What about our initial problem?

$T(x) = x - \text{grad } \psi(x) = \text{grad } (\frac{1}{2} x^2 - \psi(x))$

$\text{grad } \Phi(x)$ with $\Phi(x) := (\frac{1}{2} x^2 - \psi(x))$

for all borel set $A$, $\int_A d\mu = \int_{T(A)} (|JT|) d\nu = \int_{T(A)} (H \bar{\psi}) d\nu$

When $\mu$ and $\nu$ have a density $u$ and $v$, $(H \bar{\psi}(x)) \cdot v(\text{grad } \bar{\psi}(x)) = u(x)$

*Monge-Ampère equation*
Part. 2 Optimal Transport – summary

Find a transport map $T$ that minimizes $C(T) = \int_X ||x - T(x)||^2 \, d\mu(x)$.
Part. 2 Optimal Transport – summary

Find a transport map $T$ that minimizes $C(T) = \int_X \| x - T(x) \|^2 \, d\mu(x)$

After several rewrites and under some conditions....
(Kantorovich formulation, dual, c-convex functions)
Part. 2 Optimal Transport – summary

Find a transport map $T$ that minimizes $C(T) = \int_X \|x - T(x)\|^2 \, d\mu(x)$

After several rewrites and under some conditions....
(Kantorovich formulation, dual, c-convex functions)

Solve $(H \overline{\psi}(x)) \cdot \nu(\text{grad } \overline{\psi}(x)) = u(x)$  
Monge-Ampère equation
(When $\mu$ and $\nu$ have a density $u$ and $v$ resp.)
Part. 2 Optimal Transport – summary

Find a transport map \( T \) that minimizes \( C(T) = \int_Z || x - T(x) ||^2 \, d\mu(x) \)

After several rewrites and under some conditions .... (Kantorovich formulation, dual, c-convex functions)

Solve \( (H \bar{\psi}(x)). \nabla (\grad \bar{\psi}(x)) = u(x) \) Monge-Ampère equation
(When \( \mu \) and \( v \) have a density \( u \) and \( v \) resp.)

Brenier, Mc Cann, Trudinger: *The optimal transport map is then given by:*

\[ T(x) = \grad \bar{\psi}(x) \]
Part. 2 Optimal Transport – Isoperimetric inequality

For a given volume, ball is the shape that minimizes border area
Part. 2 Optimal Transport – Isoperimetric inequality

$L_1$ Sobolev inequality: Given $f: \mathbb{R}^n \rightarrow \mathbb{R}$ sufficiently regular

\[
\int |\nabla f| \geq n \text{Vol}(B_2^n)^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n}
\]

Explanation in [Dario Cordero Eruauquin] course notes
L₁ Sobolev inequality: Given $f: \mathbb{R}^n \rightarrow \mathbb{R}$ sufficiently regular,

$$\int |\nabla f| \geq n \, \text{Vol}(B_2^n)^{1/n} \left( \int |f|^{n/(n-1)} \right)^{(n-1)/n}$$

Explanation in [Dario Cordero Erauquin] course notes.
L_1 Sobolev inequality: Given f: \mathbb{R}^n \to \mathbb{R} sufficiently regular

\[ \int | \nabla f | \geq n \text{Vol}(B_2^n)^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n} \]

Explanation in [Dario Cordero Erauquin] course notes
Part. 2 Optimal Transport – Isoperimetric inequality

$L_1$ Sobolev inequality: Given $f: \mathbb{R}^n \to \mathbb{R}$ sufficiently regular

Consider a compact set $\Omega$ such that $\text{Vol}(\Omega) = \text{Vol}(B_2^3)$ and $f = \text{the indicatrix function of } \Omega$

$$\int |\text{grad } f| \geq n \text{Vol}(B_2^n)^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n}$$
**Part. 2 Optimal Transport – Isoperimetric inequality**

**L₁ Sobolev inequality:** Given $f: \mathbb{R}^n \rightarrow \mathbb{R}$ sufficiently regular

Consider a compact set $\Omega$ such that $\text{Vol}(\Omega) = \text{Vol}(B_2^3)$ and $f = \text{the indicatrix function of } \Omega$

\[
\int |\text{grad } f| \geq n \text{Vol}(B_2^n)^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n}
\]

\[
\text{Vol}(\partial \Omega) \geq n \text{Vol}(B_2^3)^{1/3} \text{Vol}(B_2^3)^{2/3}
\]
Part. 2 Optimal Transport – Isoperimetric inequality

$L_1$ Sobolev inequality: Given $f: \mathbb{R}^n \to \mathbb{R}$ sufficiently regular

Consider a compact set $\Omega$ such that $\text{Vol}(\Omega) = \text{Vol}(B_{2^3})$

and $f = \text{the indicatrix function of } \Omega$

$$\int |\nabla f| \geq n \text{Vol}(B_{2^n})^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n}$$

$$\text{Vol}(\partial \Omega) \geq n \text{Vol}(B_{2^3})^{1/3} \text{Vol}(B_{2^3})^{2/3}$$

$$\text{Vol}(\partial \Omega) \geq 4 \pi = \text{Vol}(\partial B_{2^3})$$
Part. 2 Optimal Transport – Isoperimetric inequality

L₁ Sobolev inequality: a proof with OT [Gromov]

\[ \int |\nabla f| \geq n \, \text{Vol}(B_2^n)^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n} \]
L₁ Sobolev inequality: a proof with OT [Gromov]

\[ \int |\text{grad } f| \geq n \text{ Vol}(B₂^n)^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n} \]

We suppose w.l.o.g. that \( \int f^{n/(n-1)} = 1 \)
Part. 2 Optimal Transport – Isoperimetric inequality

L₁ Sobolev inequality: a proof with OT [Gromov]

\[ \int |\nabla f| \geq n \text{Vol}(B_2^n)^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n} \]

We suppose w.l.o.g. that \( \int f^{n/(n-1)} = 1 \)

There exists an optimal transport \( T = \nabla \Psi \) between 
\( f^{n/(n-1)}(x)dx \) and \( 1B_2^n/\text{Vol}(B_2^n)dx \)
Part. 2 Optimal Transport – Isoperimetric inequality

$L_1$ Sobolev inequality: a proof with OT [Gromov]

$$\int |\text{grad } f| \geq n \ Vol(B_2^n)^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n}$$

We suppose w.l.o.g. that $\int f^{n/(n-1)} = 1$

There exists an optimal transport $T = \text{grad} \Psi$ between $f^{n/(n-1)}(x) \, dx$ and $1B_2^n/Vol(B_2^n) \, dx$

Monge-Ampère equation: $Vol(B_2^n) \ f^{n/(n-1)}(x) = \det \text{Hess } \Psi$
Part. 2 Optimal Transport – Isoperimetric inequality

$L_1$ Sobolev inequality: a proof with OT [Gromov]

$$\int |\text{grad } f| \geq n \text{Vol}(B_2^n)^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n}$$

We suppose w.l.o.g. that $\int f^{n/(n-1)} = 1$

There exists an optimal transport $T = \text{grad} \Psi$ between $f^{n/(n-1)}(x)\,dx$ and $1B_2^n/\text{Vol}(B_2^n)\,dx$

Monge-Ampère equation: $\text{Vol}(B_2^n) f^{n/(n-1)}(x) = \det \text{Hess} \Psi$

Arithmetico-geometric ineq: $\det (H)^{1/n} \leq \text{trace}(H)/n$ if $H$ positive
Part. 2 Optimal Transport – Isoperimetric Inequality

$L_1$ Sobolev inequality: a proof with OT [Gromov]

\[ \int |\nabla f| \geq n \frac{\text{Vol}(B_2^n)^{1/n}}{\left( \int f^{n/(n-1)} \right)^{(n-1)/n}} \]

We suppose w.l.o.g. that $\int f^{n/(n-1)} = 1$

There exists an optimal transport $T = \nabla \Psi$ between $f^{n/(n-1)}(x)dx$ and $1B_2^n/\text{Vol}(B_2^n)dx$

Monge-Ampère equation: $\text{Vol}(B_2^n) f^{n/(n-1)}(x) = \det \text{Hess} \Psi$

Arithmetico-geometric inequality: $\det (H)^{1/n} \leq \text{trace}(H)/n$ if $H$ positive

$\det (\text{Hess} \Psi)^{1/n} \leq \text{trace}(\text{Hess} \Psi)/n$
Part. 2 Optimal Transport – Isoperimetric inequality

$L_1$ Sobolev inequality: a proof with OT [Gromov]

$$\int |\nabla f| \geq n \operatorname{Vol}(B_{2^n})^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n}$$

We suppose w.l.o.g. that $\int f^{n/(n-1)} = 1$

There exists an optimal transport $T = \nabla \Psi$ between $f^{n/(n-1)}(x)dx$ and $1B_{2^n}/\operatorname{Vol}(B_{2^n})dx$

Monge-Ampère equation: $\operatorname{Vol}(B_{2^n}) f^{n/(n-1)}(x) = \det \operatorname{Hess} \Psi$

Arithmetico-geometric ineq: $\det (H)^{1/n} \leq \operatorname{trace}(H)/n$ if $H$ positive

$\det (\operatorname{Hess} \Psi)^{1/n} \leq \operatorname{trace}(\operatorname{Hess} \Psi)/n$

$\det (\operatorname{Hess} \Psi)^{1/n} \leq \Delta \Psi / n$
Part. 2 Optimal Transport – Isoperimetric inequality

**$L_1$ Sobolev inequality:** a proof with OT [Gromov]

$$\int |\text{grad } f| \geq n \operatorname{Vol}(B_2^n)^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n}$$

We suppose w.l.o.g. that $\int f^{n/(n-1)} = 1$

$$\det (\text{Hess } \Psi)^{1/n} \leq (\Delta \Psi)/n$$

Monge-Ampère equation:

$$\operatorname{Vol}(B_2^n) f^{n/(n-1)}(x) = \det \text{Hess } \Psi$$
Part. 2 Optimal Transport – Isoperimetric inequality

$L_1$ Sobolev inequality: a proof with OT [Gromov]

$$\int |\text{grad } f| \geq n \, \text{Vol}(B_2^n)^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n}$$

We suppose w.l.o.g. that \( \int f^{n/(n-1)} = 1 \)

\[ \det (\text{Hess } \Psi)^{1/n} \leq (\Delta \Psi)/n \]

Monge-Ampère equation:

\[ \text{Vol}(B_2^n) f^{n/(n-1)}(x) = \det \text{Hess } \Psi \]

\[ \text{Vol}(B_2^n) = \text{Vol}(B_2^n) \int f^{n/(n-1)} = \int f \, \text{Vol}(B_2^n) f^{1/(n-1)} \]
Part. 2 Optimal Transport – Isoperimetric inequality

L₁ Sobolev inequality: a proof with OT [Gromov]

\[ \int |\nabla f| \geq n \, \text{Vol}(B_2^n)^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n} \]

We suppose w.l.o.g. that \( \int f^{n/(n-1)} = 1 \)

\[ \det (\text{Hess } \bar{\Psi})^{1/n} \leq (\Delta \bar{\Psi})/n \]

Monge-Ampère equation:

\[ \text{Vol}(B_2^n) \, f^{n/(n-1)}(x) = \det \text{Hess } \bar{\Psi} \]

\[ \text{Vol}(B_2^n) = \text{Vol}(B_2^n) \int f^{n/(n-1)} = \int f \, \text{Vol}(B_2^n) \, f^{1/(n-1)} \leq 1/n \int f \, \Delta \bar{\Psi} \]
Part. 2 Optimal Transport – Isoperimetric Inequality

L₁ Sobolev Inequality: a proof with OT [Gromov]

\[
\int \left| \nabla f \right| \geq n \, \text{Vol}(B_2^n)^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n}
\]

We suppose w.l.o.g. that \( \int f^{n/(n-1)} = 1 \)

\[
\det(\text{Hess } \Psi)^{1/n} \leq (\Delta \Psi)/n
\]

Monge-Ampère equation:

\[
\text{Vol}(B_2^n) \, f^{n/(n-1)}(x) = \det \text{Hess } \Psi
\]

\[
\text{Vol}(B_2^n) = \text{Vol}(B_2^n) \, \int f^{n/(n-1)} = \int f \, \text{Vol}(B_2^n) \, f^{1/(n-1)} \leq 1/n \int f \, \Delta \Psi
\]

\[
\int f \, \Delta \Psi = -\int \nabla f \cdot \nabla \Psi
\]
Part. 2 Optimal Transport – Isoperimetric inequality

$L_1$ Sobolev inequality: a proof with OT [Gromov]

$$\int |\nabla f| \geq n \ Vol(B_2^n)^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n}$$

We suppose w.l.o.g. that $\int f^{n/(n-1)} = 1$

$$\det (\text{Hess } \overline{\Psi})^{1/n} \leq (\Delta \overline{\Psi})/n$$

Monge-Ampère equation:

$$\text{Vol}(B_2^n) f^{n/(n-1)}(x) = \det \text{Hess } \overline{\Psi}$$

$$\text{Vol}(B_2^n) = \text{Vol}(B_2^n) \int f^{n/(n-1)} = \int f \ \text{Vol}(B_2^n) f^{1/(n-1)} \leq 1/n \int f \Delta \overline{\Psi}$$

$$\int f \Delta \overline{\Psi} = - \int \nabla f \cdot \nabla \overline{\Psi} \leq \int |\nabla f| \quad (T = \nabla \overline{\Psi} \in B_2^n)$$
Part. 2 Optimal Transport – Isoperimetric inequality

$L_1$ Sobolev inequality: a proof with OT [Gromov]

\[ \int | \text{grad } f | \geq n \ Vol(B_2^n)^{1/n} \left( \int f^{n/(n-1)} \right)^{(n-1)/n} \]

We suppose w.l.o.g. that $\int f^{n/(n-1)} = 1$

\[ \det(\text{Hess } \Psi)^{1/n} \leq (\Delta \Psi)/n \]

Monge-Ampère equation:

\[ \text{Vol}(B_2^n) f^{n/(n-1)}(x) = \det \text{Hess } \Psi \]

\[ \text{Vol}(B_2^n) = \text{Vol}(B_2^n) \int f^{n/(n-1)} = \int f \ \text{Vol}(B_2^n) f^{1/(n-1)} \leq 1/n \int f \Delta \Psi \]

\[ \int f \Delta \Psi = -\int \text{grad } f \cdot \text{grad } \Psi \leq \int | \text{grad } f | \quad (T = \text{grad } \Psi \in B_2^n) \]

\[ \int | \text{grad } f | \geq n \ Vol(B_2^n)^{1/n} \]
3

Semi-Discrete Optimal Transport
Part. 3 Optimal Transport – how to program?

Continuous

$(X; \mu)$

$(Y; \nu)$
Part. 3 Optimal Transport – how to program?

Continuous

Semi-discrete

$(X;\mu)$  $(Y;\nu)$
Part. 3 Optimal Transport – how to program?

Continuous

Semi-discrete

Discrete

$$(X;\mu)$$

$$\rightarrow$$

$$(Y;\nu)$$
Part. 3  Optimal Transport – how to program?

Continuous

Semi-discrete

Discrete

$(X; \mu)$

$(Y; \nu)$
Part. 3 Optimal Transport – semi-discrete

\[(X;\mu) \quad \rightarrow \quad (Y;\nu)\]

\[(DMK) \quad \sup_{\psi \in \psi^c} \int_X \psi^c(x) \, d\mu + \int_Y \psi(y) \, d\nu\]
Part. 3 Optimal Transport – semi-discrete

\[(X; \mu) \quad \rightarrow \quad (Y; \nu)\]

\[(\text{DMK}) \quad \sup_{\psi \in \Psi^c} \int_X \psi^c(x) d\mu + \int_Y \psi(y) d\nu\]

\[\sum_j \psi(y_j) \nu_j\]
Part. 3 Optimal Transport – semi-discrete

\[
\text{(DMK)} \quad \sup_{\psi \in \psi^c} \int_X \psi^c(x) d\mu + \int_Y \psi(y) dv \\
\sum_j \psi(y_j) v_j
\]
Part. 3 Optimal Transport – semi-discrete

\[(DMK) \quad \sup_{\psi \in \Psi^c} \int_X \psi^c(x) \, d\mu + \int_Y \psi(y) \, dv\]

\[
\int_X \inf_{y_j \in Y} \left[ ||x - y_j||^2 - \psi(y_j) \right] \, d\mu
\]

\[
\sum_j \psi(y_j) \, v_j
\]
Part. 3 Optimal Transport – semi-discrete

\[(DMK) \quad \sup_{\psi \in \Psi^c} \int_X \psi^c(x) d\mu + \int_Y \psi(y) d\nu\]

\[\int_X \inf_{y_j \in Y} [ \| x - y_j \|^2 - \psi(y_j) ] d\mu\]

\[\sum_j \int_{\text{Lag}_{\psi}(y_j)} \| x - y_j \|^2 - \psi(y_j) d\mu\]

\[\sum_j \psi(y_j) v_j\]
Part. 3 Optimal Transport – semi-discrete

\[ (DMK) \quad \sup_{\psi \in \psi^c} G(\psi) = \sum_j \int_{\text{Lag} \ \psi(y_j)} \| x - y_j \|^2 - \psi(y_j) \ d\mu \ + \ \sum_j \psi(y_j) \ v_j \]

Where: \( \text{Lag} \ \psi(y_j) = \{ x \mid \| x - y_j \|^2 - \psi(y_j) < \| x - y_j \|^2 - \psi(y_{j'}) \} \) for all \( j' \neq j \)
Part. 3 Optimal Transport – semi-discrete

\[
\text{Sup}_{\psi \in \Psi^c} G(\psi) = \sum_j \int_{\text{Lag} \, \psi(y_j)} \| x - y_j \|^2 - \psi(y_j) \, d\mu + \sum_j \psi(y_j) \, v_j
\]

Where: \( \text{Lag} \, \psi(y_j) = \{ x \mid \| x - y_j \|^2 - \psi(y_j) < \| x - y_j \|^2 - \psi(y_j') \} \) for all \( j' \neq j \)

Laguerre diagram of the \( y_j \)'s
(with the \( L_2 \) cost \( \| x - y \|^2 \) used here, Power diagram)
Part. 3 Optimal Transport – semi-discrete

\[
\sup_{\psi \in \psi^c} G(\psi) = \sum_j \int_{\text{Lag } \psi(y_j)} ||x - y_j||^2 - \psi(y_j) \, d\mu + \sum_j \psi(y_j) \, v_j
\]

Where: \( \text{Lag } \psi(y_j) = \{ x \mid ||x - y_j||^2 - \psi(y_j) < ||x - y_j||^2 - \psi(y_j) \} \) for all \( j' \neq j \)

Laguerre diagram of the \( y_j \)'s
(with the \( L_2 \) cost \( ||x - y||^2 \) used here, Power diagram)

Weight of \( y_j \) in the power diagram
Part. 3 Optimal Transport – semi-discrete

\[ \text{(DMK)} \quad \sup_{\psi \in \psi_c} G(\psi) = \sum_j \int_{\text{Lag} \psi(y_j)} \| x - y_j \|^2 - \psi(y_j) \, d\mu + \sum_j \psi(y_j) \, v_j \]

Where: \( \text{Lag} \psi(y_j) = \{ x \mid \| x - y_j \|^2 - \psi(y_j) < \| x - y_j \|^2 - \psi(y_{j'}) \} \) for all \( j' \neq j \)

- **Laguerre diagram** of the \( y_j \)'s
- **Weight** of \( y_j \) in the power diagram

\( \psi \) is determined by the **weight vector** \( [\psi(y_1), \psi(y_2), \ldots, \psi(y_m)] \)
Part. 3 Optimal Transport – semi-discrete

\[
(\text{DMK}) \quad \sup_{\psi \in \psi^c} G(\psi) = \sum_j \int_{\text{Lag} \psi(y_j)} \| x - y_j \|^2 - \psi(y_j) \, d\mu \quad + \quad \sum_j \psi(y_j) \, v_j
\]

Where: \( \text{Lag} \psi(y_j) = \{ \, x \mid \| x - y_j \|^2 - \psi(y_j) < \| x - y_j \|^2 - \psi(y_{j'}) \, \} \) for all \( j' \neq j \)

\[\psi\] is determined by the \textbf{weight vector} \([\psi(y_1) \, \psi(y_2) \, \ldots \, \psi(y_m)]\)

For all weight vector, \( \psi \) is c-concave
Part. 3 Power Diagrams

Voronoï diagram: \( \text{Vor}(x_i) = \{ x \mid d^2(x, x_i) < d^2(x, x_j) \} \)
Voronoi diagram: \( \text{Vor}(x_i) = \{ x \mid d^2(x, x_i) < d^2(x, x_j) \} \)

Power diagram: \( \text{Pow}(x_i) = \{ x \mid d^2(x, x_i) - \psi_i < d^2(x, x_j) - \psi_j \} \)
Part. 3 Power Diagrams
**Theorem:** (direct consequence of MK duality

alternative proof in [Aurenhammer, Hoffmann, Aronov 98]):

Given a measure $\mu$ with density, a set of points $(y_j)$, a set of positive coefficients $v_j$ such that $\sum v_j = \int d\mu(x)$, it is possible to find the weights $W = [\psi(y_1) \; \psi(y_2) \ldots \; \psi(y_m)]$ such that the map $T_S^W$ is the unique optimal transport map between $\mu$ and $\nu = \sum v_j \delta(y_j)$.
\textbf{Theorem:} (direct consequence of MK duality alternative proof in [Aurenhammer, Hoffmann, Aronov 98]):

Given a measure $\mu$ with density, a set of points $(y_j)$, a set of positive coefficients $v_j$ such that $\sum v_j = \int d\mu(x)$, it is possible to find the weights $W = [\psi(y_1) \, \psi(y_2) \, \ldots \, \psi(y_m)]$ such that the map $T_S^W$ is the unique optimal transport map between $\mu$ and $v = \sum v_j \delta(y_j)$.

\textbf{Proof:} $G(\psi) = \sum_j \int \text{Lag } \psi(y_j) \parallel x - y_j \parallel^2 - \psi(y_j) \, d\mu + \sum_j \psi(y_j) \, v_j$

Is a concave function of the weight vector $[\psi(y_1) \, \psi(y_2) \, \ldots \, \psi(y_m)]$
**Theorem:** (direct consequence of MK duality
alternative proof in [Aurenhammer, Hoffmann, Aronov 98]):

Given a measure $\mu$ with density, a set of points $(y_j)$, a set of positive coefficients $v_j$ such that $\sum v_j = \int d\mu(x)$, it is possible to find the weights $W = [\psi(y_1) \psi(y_2) \ldots \psi(y_m)]$ such that the map $T_S^W$ is the unique optimal transport map between $\mu$ and $\nu = \sum v_j \delta(y_j)$

**Proof:**

$$G(\psi) = \sum_j \int \text{Lag}_{\psi(y_j)} \| x - y_j \|^2 - \psi(y_j) \, d\mu + \sum_j \psi(y_j) \, v_j$$

Is a concave function of the weight vector $[\psi(y_1) \psi(y_2) \ldots \psi(y_m)]$
**Part. 3** Optimal Transport – the AHA paper

*Idea of the proof*

Consider the function

$$ f_T(W) = \int \left( \| x - T(x) \|^2 - \psi(T(X)) \right) d\mu(x) $$

The (unknown) weights $W = [\psi(y_1) \, \psi(y_2) \, \ldots \, \psi(y_m)]$
**Part. 3** Optimal Transport – the AHA paper

**Idea of the proof**

Consider the function

\[
    f_T(W) = \int \left( \| x - T(x) \|^2 - \psi(T(x)) \right) d\mu(x)
\]

\(T\) : an arbitrary but fixed assignment.
Part. 3 Optimal Transport – the AHA paper

Idea of the proof

Consider the function \( f_T(W) = \int \left( \| x - T(x) \|^2 - \psi(T(X)) \right) d\mu(x) \)

where \( T \) is an arbitrary but fixed assignment.
Part. 3 Optimal Transport – the AHA paper

Idea of the proof

Consider the function

$$f_T(W) = \int \left( \| x - T(x) \|^2 - \psi(T(X)) \right) d\mu(x)$$

$T$: an arbitrary but fixed assignment.
Idea of the proof

Consider the function

\[ f_T(W) = \int \left( \| x - T(x) \|^2 - \psi(T(x)) \right) d\mu(x) \]

\( T \) : an arbitrary but fixed assignment.
Part. 3 Optimal Transport – the AHA paper

Idea of the proof

Consider the function

\[ f_T(W) = \int \left( \left\| x - T(x) \right\|^2 - \psi(T(X)) \right) d\mu(x) \]
Part. 3 Optimal Transport – the AHA paper

Idea of the proof

Consider the function

$$f_T(W) = \int \left( \| x - T(x) \|^2 - \psi(T(X)) \right) d\mu(x)$$
**Part. 3** Optimal Transport – the AHA paper

**Idea of the proof**

Consider the function

\[ f_T(W) = \int \left( \| x - T(x) \|^2 - \psi(T(x)) \right) d\mu(x) \]

**Fixed T**

\[ f_T(W) \text{ is linear in } W \]
Part. 3  Optimal Transport – the AHA paper

Idea of the proof

Consider the function

\[ f_T(W) = \int \left( \| x - T(x) \|^2 - \psi(T(x)) \right) d\mu(x) \]

\( f_T(W) \) is linear in \( W \)

\( f_{T_W}(W) \): defined by power diagram
**Part. 3** Optimal Transport – the AHA paper

**Idea of the proof**

Consider the function

\[
 f_T(W) = \int \left( \| x - T(x) \|^2 - \psi(T(x)) \right) d\mu(x)
\]

- \( f_T(W) \) is linear in \( W \)

- \( f_{TW}(W) = \min_T f_T(W) \)
Part. 3 Optimal Transport – the AHA paper

Idea of the proof

Consider the function

\[ f_T(W) = \int \left( \| x - T(x) \|^2 - \psi(T(x)) \right) d\mu(x) \]

\( f_T(W) \) is linear in \( W \)

\( f: W \to f_{T_W}(W) \) is concave !!

(because its graph is the lower enveloppe of linear functions)
**Part. 3** Optimal Transport – the AHA paper

**Idea of the proof**

Consider the function

\[
 f_T(W) = \int \left( \| x - T(x) \|^2 - \psi(T(x)) \right) d\mu(x)
\]

- \( f_T(W) \) is linear in \( W \)
- \( f: W \rightarrow f_{TW}(W) \) is concave
  (because its graph is the lower envelope of linear functions)

Consider \( g(W) = f_{TW}(W) + \sum v_j \psi_j \)
Idea of the proof

Consider the function

\[ f_T(W) = \int \left( \| x - T(x) \|^2 - \psi(T(x)) \right) d\mu(x) \]

\( f_T(W) \) is linear in \( W \)

\( f: W \rightarrow f_{T,W}(W) \) is concave (because its graph is the lower enveloppe of linear functions)

Consider \( g(W) = f_{T,W}(W) + \sum v_j \psi_j \)

\[ \frac{\partial g}{\partial \psi_j} = V_j - \int_{\text{pow}(y_j)} \| x - y_j \|^2 d\mu(x) \]

and \( g \) is concave.
Semi-discrete OT Summary:

(DMK) \[
\sup_{\psi \in \psi^c} G(\psi) = \int_X \psi^c(x) d\mu + \int_Y \psi(y) d\nu
\]
Part. 3 Optimal Transport – the algorithm

Semi-discrete OT Summary:

\[
\text{(DMK)} \quad \sup_{\psi \in \psi^c} G(\psi) = \int_X \psi^c(x) d\mu + \int_Y \psi(y) d\nu
\]

\[
G(\psi) = g(W) = \sum_j \int_{\text{Lag} \psi(y_j)} \| x - y_j \|^2 - \psi(y_j) \, d\mu + \sum_j \psi(y_j) \, v_j \quad \text{is concave}
\]
Part. 3 Optimal Transport – the algorithm

Semi-discrete OT Summary:

\[
\begin{align*}
\text{(DMK)} \\ 
\sup_{\psi \in \psi^c} \ G(\psi) &= \int_X \psi^c(x) d\mu + \int_Y \psi(y) d\nu \\
G(\psi) &= g(W) = \sum_j \int \text{Lag} \ \psi(y_j) \| x - y \|^2 - \psi(y_j) \ d\mu + \sum_j \psi(y_j) v_j \text{ is concave} \\
\partial G / \partial \psi_j &= V_j - \int \text{pow}(y_j) \| x - y_j \|^2 d\mu(x) \ (= 0 \text{ at the maximum})
\end{align*}
\]
Part. 3 Optimal Transport – the algorithm

Semi-discrete OT Summary:

\[ G(\psi) = \frac{\sup_{\psi \in \psi^c}}{} G(\psi) = \int_X \psi_c(x) d\mu + \int_Y \psi(y) d\nu \]

\[ G(\psi) = g(W) = \sum_j \int_{\text{Lag}} \psi(y_j) \|x - y_j\|^2 - \psi(y_j) \ d\mu + \sum_j \psi(y_j) \ v_j \text{ is concave} \]

\[ \frac{\partial G}{\partial \psi_j} = V_j - \int_{\text{pow}(y_j)} \|x - y_j\|^2 d\mu(x) \ (= 0 \text{ at the maximum}) \]

Desired mass at \( y_j \)
Mass transported to \( y_j \)
The [AHA] paper summary:
• The optimal weights minimize a convex function
• The gradient of this convex function is easy to compute
Note: the weight $w(s)$ correspond to the Kantorovich potential $\psi(x)$
  (solves a “discrete Monge-Ampere” equation)

The algorithm:
Input: two tetrahedral meshes $M_1$ and $M_2$
Output: a morphing between $M_1$ and $M_2$
The [AHA] paper summary:
- The optimal weights minimize a convex function
- The gradient of this convex function is easy to compute

Note: the weight $w(s)$ correspond to the Kantorovich potential $\psi(x)$
  (solves a “discrete Monge-Ampere” equation)

The algorithm:
Input: two tetrahedral meshes $M_1$ and $M_2$
Output: a morphing between $M_1$ and $M_2$

Step 1: sample $M_2$ with $N$ points $(s_1 \ldots s_N)$
Part. 3 Optimal Transport – the algorithm

The [AHA] paper summary:
- The optimal weights minimize a convex function
- The gradient of this convex function is easy to compute

Note: the weight $w(s)$ correspond to the Kantorovich potential $\psi(x)$
(solves a “discrete Monge-Ampere” equation)

The algorithm:
Input: two tetrahedral meshes $M_1$ and $M_2$
Output: a morphing between $M_1$ and $M_2$

Step 1: sample $M_2$ with N points ($s_1 \ldots s_N$)
Step 2: initialize the weights $(w_1 \ldots w_N) = (0 \ldots 0)$
Part. 3 Optimal Transport – the algorithm

The [AHA] paper summary:
• The optimal weights minimize a convex function
• The gradient of this convex function is easy to compute
Note: the weight $w(s)$ correspond to the Kantorovich potential $\psi(x)$
  (solves a “discrete Monge-Ampere” equation)

The algorithm:
Input: two tetrahedral meshes $M_1$ and $M_2$
Output: a morphing between $M_1$ and $M_2$

Step 1: sample $M_2$ with $N$ points $(s_1 \ldots s_N)$
Step 2: initialize the weights $(w_1 \ldots w_N) = (0 \ldots 0)$
Step 3: minimize $g(w_1 \ldots w_N)$ with a quasi-Newton algorithm:
The [AHA] paper summary:
- The optimal weights minimize a convex function
- The gradient of this convex function is easy to compute

Note: the weight $w(s)$ correspond to the Kantorovich potential $\psi(x)$
  (solves a “discrete Monge-Ampere” equation)

**The algorithm:**
**Input:** two tetrahedral meshes $M_1$ and $M_2$
**Output:** a morphing between $M_1$ and $M_2$

**Step 1:** sample $M_2$ with $N$ points $(s_1 \ldots s_N)$
**Step 2:** initialize the weights $(w_1 \ldots w_N) = (0 \ldots 0)$
**Step 3:** minimize $g(w_1 \ldots w_N)$ with a quasi-Newton algorithm:
  For each iterate $(s_1 \ldots s_N)^{(k)}$: 
Part. 3 Optimal Transport – the algorithm

The [AHA] paper summary:
• The optimal weights minimize a convex function
• The gradient of this convex function is easy to compute
Note: the weight w(s) correspond to the Kantorovich potential $\psi(x)$
  (solves a “discrete Monge-Ampere” equation)

The algorithm:
Input: two tetrahedral meshes $M_1$ and $M_2$
Output: a morphing between $M_1$ and $M_2$

Step 1: sample $M_2$ with N points $(s_1 \ldots s_N)$
Step 2: initialize the weights $(w_1 \ldots w_N) = (0 \ldots 0)$
Step 3: minimize $g(w_1 \ldots w_N)$ with a quasi-Newton algorithm:
  For each iterate $(s_1 \ldots s_N)^{(k)}$:
  Compute $\text{Pow}( (w_i, s_i) ) \cap M_1$ [Nivoliers, L 2014, Curves and Surfaces]
Part. 3 Optimal Transport – the algorithm

Compute $\text{Pow}( (w_i, s_i) ) \cap M_1$ [Nivoliers, L 2014, Curves and Surfaces]
Part. 3 Optimal Transport – the algorithm

Compute $\text{Pow}( (w_i, s_i) ) \cap M_1$ [Nivoliers, L 2014, Curves and Surfaces]

Implementation in GEOGRAM (http://alice.loria.fr/software/geogram)

Predicates + Predicate Construction Kit [L 2015] – available for download
Part. 3 Optimal Transport – the algorithm

The [AHA] paper summary:
- The optimal weights minimize a convex function
- The gradient of this convex function is easy to compute

Note: the weight $w(s)$ correspond to the Kantorovich potential $\psi(x)$
(solves a “discrete Monge-Ampere” equation)

The algorithm:
**Input:** two tetrahedral meshes $M_1$ and $M_2$
**Output:** a morphing between $M_1$ and $M_2$

**Step 1:** sample $M_2$ with $N$ points $(s_1 \ldots s_N)$
**Step 2:** initialize the weights $(w_1 \ldots w_N) = (0 \ldots 0)$
**Step 3:** minimize $g(w_1 \ldots w_N)$ with a quasi-Newton algorithm:
  - For each iterate $(s_1 \ldots s_N)^{(k)}$:
    - Compute $\text{Pow}\left( (w_i, s_i) \right) \cap M_1$ [L 2014, Curves and Surfaces]
  - Compute $g$ and $\text{grad } g$
Part. 3 Optimal Transport – the algorithm

The [AHA] paper summary:
- The optimal weights minimize a convex function
- The gradient of this convex function is easy to compute

Note: the weight \( w(s) \) correspond to the Kantorovich potential \( \psi(x) \)
(solves a “discrete Monge-Ampere” equation)

The algorithm:
Input: two tetrahedral meshes \( M_1 \) and \( M_2 \)
Output: a morphing between \( M_1 \) and \( M_2 \)

Step 1: sample \( M_2 \) with \( N \) points \( (s_1 \ldots s_N) \)
Step 2: initialize the weights \( (w_1 \ldots w_N) = (0 \ldots 0) \)
Step 3: minimize \( g(w_1 \ldots w_N) \) with a quasi-Newton algorithm:
  For each iterate \( (s_1 \ldots s_N)^{(k)} \):
  Compute \( \text{Pow} \left( (w_i, s_i) \right) \cap M_1 \) \[L 2014, Curves and Surfaces]\)
  Compute \( g \) and \( \text{grad} \ g \)

+ Multilevel version [Merigot 2011] (2D),
[\( L 2014 \) arXiv, M2AN 2015] (3D & relation with Centroidal Voronoi Tesselation)
The [AHA] paper summary:
- The optimal weights minimize a convex function
- The gradient of this convex function is easy to compute

Note: the weight $w(s)$ correspond to the Kantorovich potential $\psi(x)$
(solves a “discrete Monge-Ampere” equation)

The algorithm:

Summary:
The algorithm computes the weights $w_i$ such that the power cells associated with the Diracs correspond to the preimages of the Diracs.
The [AHA] paper summary:
- The optimal weights minimize a convex function
- The gradient of this convex function is easy to compute

Note: the weight $w(s)$ correspond to the Kantorovich potential $\psi(x)$
(solves a “discrete Monge-Ampere” equation)

The algorithm:

Summary:
The algorithm computes the weights $w_i$ such that the power cells associated with the Diracs correspond to the preimages of the Diracs.
The [AHA] paper summary:
- The optimal weights minimize a convex function
- The gradient of this convex function is easy to compute

Note: the weight \( w(s) \) correspond to the Kantorovich potential \( \psi(x) \)
  (solves a “discrete Monge-Ampere” equation)

The algorithm:

Summary:
The algorithm computes the weights \( w_i \) such that the power cells associated with the Diracs correspond to the preimages of the Diracs.
The [AHA] paper summary:

- The optimal weights minimize a convex function
- The gradient of this convex function is easy to compute

Note: the weight \( w(s) \) correspond to the Kantorovich potential \( \psi(x) \)
  (solves a “discrete Monge-Ampere” equation)

The algorithm:

Summary:
The algorithm computes the weights \( w_i \) such that the power cells associated with the Diracs correspond to the preimages of the Diracs.
Part. 3 Optimal Transport – the algorithm

The [AHA] paper summary:
- The optimal weights minimize a convex function
- The gradient of this convex function is easy to compute

Note: the weight $w(s)$ correspond to the Kantorovich potential $\psi(x)$
(solves a “discrete Monge-Ampere” equation)

The algorithm:

Summary:
The algorithm computes the weights $w_i$ such that the power cells associated with the Diracs correspond to the preimages of the Diracs.
Part. 4 Power Diagrams & Transport
Part. 4 Power Diagrams & Transport
Part. 4 Power Diagrams & Transport
Part. 4 Power Diagrams & Transport
Part. 4 Power Diagrams & Transport
Part. 4 Power Diagrams & Transport
Part. 4 Power Diagrams & Transport

Translating a Voronoi diagram.

1st Try: linear lifting

( Fail: scale by $1/\cos(x)$ )
Part. 4 Power Diagrams & Transport

2nd Try: Curved lifting
"converging beams" can compensate the 
\frac{1}{\cos(x)} expansion by "re-concentrating" the points
Part. 4 Power Diagrams & Transport

\[ d^2(p_i, q) - w_i < d^2(p_j, q) - w_j \quad \forall j \]

\[ d^2(p_i, q - T) < d^2(p_j, q - T) \quad \forall j \]

\[ (p_i - q + T)^2 < (p_j - q + T)^2 \quad \forall j \]

\[ d^2(p_i, q) + 2T \cdot (p_i - q) + T^2 < d^2(p_j, q) + 2T \cdot (p_j - q) + T^2 \quad \forall j \]

\[ d^2(p_i, q) + 2T \cdot p_i < d^2(p_j, q) + 2T \cdot p_j \]

\[ w_i = -2T \cdot p_i + \text{const} \]

\[ h_i = \sqrt{2T \cdot p_i - \min(T, p_i)} \]
C'est quoi l'équation au continu?

\[
\frac{1}{\sqrt{\left(\frac{\partial h}{\partial x}\right)^2 + \left(\frac{\partial h}{\partial y}\right)^2}} u(y) = \varphi(x)
\]

On suppose que :

\[
(n \cdot \mathbf{Z}) \mu(y) = \varphi(x)
\]

\[
T^{-1}(x) = \{ y \mid h^2(y) + d^2(x, y) \text{ minimum} \}
\]
Part. 4 Power Diagrams and Transport

[B\delta is discretized (Sum of Diracs)]

[Voronoi diagram of B samples. How to "back displace" it onto \( A \)?]

[Lifting on two "square root wings" translates both halves of B points into the two blobs of A]

[Solving for the OTM \((T(z,v))\) vector field is equivalent to solve for the "square root wings" \((h(x,v))\) scalar function. Simplest Unconstrained]
Part. 4 Optimal Transport – 2D examples
Numerical Experiment: A disk becomes two disks
Numerical Experiment: A sphere becomes a cube
Numerical Experiment: A sphere becomes two spheres
Part. 4 Optimal Transport – 3D examples

Numerical Experiment: Armadillo to sphere
Part. 4 Optimal Transport – 3D examples
Numerical Experiment: *Other examples*
Part. 4 Optimal Transport – 3D examples

Numerical Experiment: *Varying density*
Part. 4 Optimal Transport – 3D examples

Numerical Experiment: *Performances*

<table>
<thead>
<tr>
<th>nb masses</th>
<th>1000</th>
<th>2000</th>
<th>5000</th>
<th>10000</th>
<th>50000</th>
<th>$10^5$</th>
<th>$3 \times 10^5$</th>
<th>$5 \times 10^5$</th>
<th>$10^6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>time (s)</td>
<td>1.45</td>
<td>3.2</td>
<td>7.3</td>
<td>17.3</td>
<td>55</td>
<td>154</td>
<td>187</td>
<td>671</td>
<td>1262</td>
</tr>
</tbody>
</table>

TABLE 4. Statistics for the Armadillo $\rightarrow$ sphere optimal transport with varying number of masses (see third row of Figure 12). Timings are given in seconds. The multi-level algorithm with BRIO pre-ordering and degree 2 regressions is used.
Numerical Experiment: *Performances*

<table>
<thead>
<tr>
<th>nb masses</th>
<th>1000</th>
<th>2000</th>
<th>5000</th>
<th>10000</th>
<th>30000</th>
<th>50000</th>
<th>$10^5$</th>
<th>$3 \times 10^5$</th>
<th>$5 \times 10^5$</th>
<th>$10^6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>time (s)</td>
<td>1.45</td>
<td>3.2</td>
<td>7.3</td>
<td>17.3</td>
<td>55</td>
<td>154</td>
<td>187</td>
<td>671</td>
<td>1262</td>
<td>2649</td>
</tr>
</tbody>
</table>

TABLE 4. Statistics for the Armadillo $\rightarrow$ sphere optimal transport with varying number of masses (see third row of Figure 12). Timings are given in seconds. The multi-level algorithm with BRIO pre-ordering and degree 2 regressions is used.

Note that a few years ago, several hours of supercomputer time were needed for computing OT with a few thousand Dirac masses, with a combinatorial algorithm in $O(n^3)$.
Note that a few years ago, several hours of supercomputer time were needed for computing OT with a few thousand Dirac masses, with a combinatorial algorithm in $O(n^3)$.

With the semi-discrete algorithm, it takes **less than 10 seconds** on my laptop.
Part. 4 Optimal Transport – 3D examples

Numerical Experiment: Performances

Note that a few years ago, several hours of supercomputer time were needed for computing OT with a few thousand Dirac masses, with a combinatorial algorithm in $O(n^3)$

In the semi-discrete setting, my 3D version of Merigot’s multigrid algorithm computes OT for **1 million Dirac masses** in less than 1 hour on a laptop PC
Numerical Experiment: Performances

<table>
<thead>
<tr>
<th>nb masses</th>
<th>1000</th>
<th>2000</th>
<th>5000</th>
<th>10000</th>
<th>30000</th>
<th>50000</th>
<th>$10^5$</th>
<th>$3 \times 10^5$</th>
<th>$5 \times 10^5$</th>
<th>$10^6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>time (s)</td>
<td>1.45</td>
<td>3.2</td>
<td>7.3</td>
<td>17.3</td>
<td>55</td>
<td>154</td>
<td>187</td>
<td>671</td>
<td>1262</td>
<td>2649</td>
</tr>
</tbody>
</table>

TABLE 4. Statistics for the Armadillo → sphere optimal transport with varying number of masses (see third row of Figure 12). Timings are given in seconds. The multi-level algorithm with BRIO pre-ordering and degree 2 regressions is used.

Note that a few years ago, several hours of supercomputer time were needed for computing OT with a few thousand Dirac masses, with a combinatorial algorithm in $O(n^3)$

In the semi-discrete setting, my 3D version of Merigot’s multigrid algorithm computes OT for **1 million Dirac masses** in less than 1 hour on a laptop PC

Even much faster convergence can probably be reached with a true Newton solver (and several acceleration tricks), still investigating…
Other topics

• Euler equation in more complicated setting: [Merigot & Mirebeau]

• Using semi-discrete OT to solve other PDEs [Benamou, Carlier, Merigot, Oudet]

• New fluid simulation methods “power particles” [DeGoes et.al]
Conclusions – Open questions

* Connections with physics, Legendre transform and entropy ?
  [Cuturi & Peyré] – regularized discrete optimal transport – why does it work ?
  Hint 1: Minimum action principle subject to conservation laws
  Hint 2: Entropy = dual of temperature ; Legendre = Fourier[(+,*) → (Max,+)]…

* More continuous numerical algorithms ?
  [Benamou & Brenier] fluid dynamics point of view – very elegant, but 4D problem !
  FEM-type adaptive discretization of the subdifferential (graph of T) ?

* Can we characterize OT in other semi-discrete settings ?
  measures supported on unions of spheres
  piecewise linear densities

* Connections with computational geometry ?
  Singularity set [Figalli] = set of points where T is discontinuous
  Looks like a “mutual power diagram”, anisotropic Voronoi diagrams
Conclusions - References

Some references (that this presentation is based on)

A Multiscale Approach to Optimal Transport, 

Variational Principles for Minkowski Type Problems, Discrete Optimal Transport, 
and Discrete Monge-Ampere Equations 
**Xianfeng Gu, Feng Luo, Jian Sun, S.-T. Yau**, ArXiv 2013

Minkowski-type theorems and least-squares clustering 
**AHA! (Aurenhammer, Hoffmann, and Aronov)**, SIAM J. on math. ana. 1998

Topics on Optimal Transportation, 2003 
Optimal Transport Old and New, 2008 
**Cédric Villani**

**Jean-David Benamou & Yann Brenier** 
a fluid formulation of Optimal Transportation, 2000

**Laudau and Lifschitz** – Course of Theoretical Physics – Volumes I and III
Conclusions - References

Other references

Polar factorization and monotone rearrangement of vector-valued functions


Pogorelov, Alexandrov – Gradient maps, Minkovsky problem (older than AHA paper, some overlap, in slightly different context, formalism used by Gu & Yau)

Rockafeller – Convex optimization – Theorem to switch inf(sup()) – sup(inf()) with convex functions (used to justify Kantorovich duality)

Online resources

All the source code/documentation available from: alice.loria.fr/software/geogram

Computes semi-discrete OT in 3D
Scales up to millions Dirac masses on a laptop

L., A numerical algorithm for semi-discrete L2 OT in 3D, ESAIM Math. Modeling and Analysis, accepted
(draft: http://arxiv.org/abs/1409.1279 <= to be fixed: bug in MA equation in this version, fixed in M2AN journal version)
Downloads: alice.loria.fr/software GEOGRAM & GRAPHITE
Video of course on OT: www.loria.fr/~levy

Acknowledgements

Funding: European Research Council & ANR (“french NSF”)
GOODSHAPE ERC-StG-205693
VORPALINE ERC-PoC-334829
ANR MORPHO, ANR BECASIM
New project: EXPLORAGRAM (Inria) with Q. Mérigot and J.-D. Benamou

Quentin Merigot, Yann Brenier, Boris Thibert, Emmanuel Maitre,
Jean-David Benamou, Filippo Santambrogio, Edouard Oudet, Hervé Pajot.
ANR TOMMI, ANR GEOMETRYA