Optimal Transportation: Convex Geometric View

David Gu

Computer Science Department
Stony Brook University

gu@cs.stonybrook.edu

September 1, 2022
Overview

There are three views of optimal transportation theory:

1. Duality view
2. Fluid dynamics view
3. Differential geometric view

Different views give different insights and induce different computational methods; but all three theories are coherent and consistent.
Figure: Buddha surface.
Figure: Optimal transportation map.
**Figure**: Brenier potential.
Figure: Brenier potential.
Figure: Brenier potential.
Convex Geometric View
Problem (Brenier)

Given $(\Omega, \mu)$ and $(\Sigma, \nu)$ and the cost function $c(x, y) = \frac{1}{2}|x - y|^2$, the optimal transportation map $T : \Omega \rightarrow \Sigma$ is the gradient map of the Brenier potential $u : \Omega \rightarrow \mathbb{R}$, which satisfies the Monge-Ampère equation,

$$\det \left( \frac{\partial^2 u(x)}{\partial x_i \partial x_j} \right) = \frac{f(x)}{g \circ \nabla u(x)}$$
Monge-Ampère Equation

Problem (prescribed Gauss curvature)

Suppose that a real-valued function $K$ is specified on a domain $\Omega$ in $\mathbb{R}^d$, the problem seeks to identify a hypersurface of $\mathbb{R}^{d+1}$ as a graph $z = u(x)$ over $x \in \Omega$ so that at each point of the surface the Gauss curvature is given by $K(x)$.

$r(x, y) = (x, y, u(x, y))$, $r_x = (1, 0, u_x)$, $r_y = (0, 1, u_y)$, $n = \frac{(-u_x, -u_y, 1)}{\sqrt{1+|\nabla u|^2}}$,

$$E = 1 + u_x^2, \quad F = u_x u_y, \quad G = 1 + u_y^2$$

$$L = \frac{u_{xx}}{\sqrt{1+|\nabla u|^2}}, \quad M = \frac{u_{xy}}{\sqrt{1+|\nabla u|^2}}, \quad N = \frac{u_{yy}}{\sqrt{1+|\nabla u|^2}}$$

$$K(x, y) = \frac{u_{xx} u_{yy} - u_{xy}^2}{(1 + |\nabla u|^2)^2}, \quad \text{Geneal case} \quad K(x)(1 + |\nabla u|^2)^{\frac{n+2}{2}} = \det D^2 u$$
Semi-Discrete Optimal Transportation Problem

Problem (Semi-discrete OT)

Given a compact convex domain $\Omega$ in $\mathbb{R}^d$, and $p_1, p_2, \cdots, p_k$ and weights $w_1, w_2, \cdots, w_k > 0$, find a transport map $T : \Omega \rightarrow \{p_1, \ldots, p_k\}$, such that $\text{vol}(T^{-1}(p_i)) = w_i$, so that $T$ minimizes the transportation cost:

$$C(T) := \frac{1}{2} \int_{\Omega} |x - T(x)|^2 dx$$
According to Brenier theorem, there will be a piecewise linear convex function $u : \Omega \rightarrow \mathbb{R}$, the gradient map gives the optimal transport map.
Suppose the projection of Brenier potential $u_h$ induces a cell decomposition $\Omega = \bigcup_{i=1}^k W_i$, and the map $T : W_i \rightarrow p_i$. Given another cell decomposition $\Omega = \bigcup_{i=1}^k W'_i$, $\text{vol}(W_i) = \text{vol}(W'_i)$ and $T' : W'_i \rightarrow p_i$, then $C(T) \leq C(T')$.

Proof.

Since $\text{vol}(W_i) = \text{vol}(W'_j)$, we have $\sum_{i=1}^k \text{vol}(W_i) h_i = \sum_{j=1}^k \text{vol}(W'_j) h_j$, namely

$$\sum_{i=1}^k \sum_{j=1}^k \text{vol}(W_i \cap W'_j) h_i = \sum_{i=1}^k \sum_{j=1}^k \text{vol}(W_i \cap W'_j) h_j,$$

$$\sum_{i,j=1}^k \int_{W_i \cap W'_j} (h_i - h_j) \, dx = 0 \quad \sum_{i,j=1}^k \int_{W_i \cap W'_j} (|p_i|^2 - |p_j|^2) \, dx = 0$$

therefore
\[ C(T) - C(T') \]
\[ = \frac{1}{2} \sum_{i,j} \int_{W_i \cap W_j'} |x - p_i|^2 - |x - p_j|^2 \, dx \]
\[ = \frac{1}{2} \sum_{i,j} \int_{W_i \cap W_j'} |x|^2 - 2\langle x, p_i \rangle + |p_i|^2 - (|x|^2 - 2\langle x, p_j \rangle + |p_j|^2) \, dx \]
\[ = - \sum_{i,j} \int_{W_i \cap W_j'} \langle x, p_i \rangle - \langle x, p_j \rangle \, dx \]
\[ = - \sum_{i,j} \int_{W_i \cap W_j'} (\langle x, p_i \rangle - h_i) - (\langle x, p_j \rangle - h_j) \, dx \]
\[ \leq 0. \]
Each target point $p_i$ corresponds to a supporting plane

$$\pi_{h,i}(x) = \langle x, p_i \rangle - h_i.$$ 

The Brenier potential is the upper envelope of the supporting planes,

$$u_h(x) := \max_{i=1}^{k} \{ \pi_{h,i}(x) \} = \max_{i=1}^{k} \{ \langle x, p_i \rangle - h_i \}.$$
Theorem

Minkowski Given $k$ unit vectors $\mathbf{n}_1, \cdots, \mathbf{n}_k$ not contained in a half-space in $\mathbb{R}^n$ and $A_1, \cdots, A_k > 0$, such that

$$\sum_i A_i \mathbf{n}_i = 0,$$

there is a compact convex polytope $P$ with exactly $k$ codimension-1 faces $F_1, \cdots, F_k$, such that

1. $\text{area}(F_i) = A_i$,
2. $\mathbf{n}_i \perp F_i$.

All such polytopes differ by a translation.
Brunn-Minkowski inequality

**Definition (Minkowski Sum)**

Given $A, B \subset \mathbb{R}^n$, their Minkowski sum is defined as

\[ A \oplus B := \{ p + q | p \in A, q \in B \} . \]

**Theorem (Brunn-Minkowski)**

For every pair of nonempty compact subsets $A$ and $B$ of $\mathbb{R}^n$ and every $0 \leq t \leq 1$,

\[ [\text{Vol}(tA \oplus (1 - t)B)]^{\frac{1}{n}} \geq t[\text{vol}(A)]^{\frac{1}{n}} + (1 - t)[\text{vol}(B)]^{\frac{1}{n}}. \]

For convex sets $A$ and $B$, the inequality is strict for $0 < t < 1$ unless $A$ and $B$ are homothetic i.e. are equal up to translation and dilation.
Minkowski Theorem

Proof.

Construct hyper-planes $\langle x, n_i \rangle = h_i$, the hyper-planes support a convex polytope $P(h_1, h_2, \ldots, h_k)$, we maximize the volume of $P(h)$,

$$\max_h \text{Vol}(P(h_1, h_2, \ldots, h_k))$$

under the constraint

$$h_1 A_1 + h_2 A_2 + \cdots + h_k A_k = 1.$$

We use Lagrange multiplier method,

$$\max_{h,\lambda} \text{Vol}(P(h)) - \lambda \left( \sum_{i=1}^{k} h_i A_i - 1 \right),$$
We define admissible space of the heights

\[ \mathcal{H} := \{ h | w_i(h) > 0, \ i = 1, 2, \ldots, k \} \]

By Brunn-Minkowski inequality, \( \mathcal{H} \) is convex. At the boundary of \( \mathcal{H} \), some face \( F_i \) has zero volume, \( w_i(h) = 0 \). The functional is \( C^1 \), hence we get the gradient

\[ \frac{\partial \text{Vol}(P(h))}{\partial h_i} - \lambda A_i = w_i(h) - \lambda A_i < 0, \]

hence the maximal point \( h^* \) is the interior point of \( \mathcal{H} \). At the maximal point, the gradient equals to zero, then we obtain

\[ (w_1(h^*), w_2(h^*), \ldots, w_k(h^*)) = \lambda (A_1, A_2, \ldots, A_k). \]
Alexandrov Theorem

**Theorem (Alexandrov 1950)**

Given \( \Omega \) compact convex domain in \( \mathbb{R}^n \), \( p_1, \cdots, p_k \) distinct in \( \mathbb{R}^n \), \( A_1, \cdots, A_k > 0 \), such that \( \sum A_i = \text{Vol}(\Omega) \), there exists PL convex function

\[
f(x) := \max\{\langle x, p_i \rangle + h_i | i = 1, \cdots, k\}
\]

unique up to translation such that

\[
\text{Vol}(W_i) = \text{Vol}(\{x | \nabla f(x) = p_i\}) = A_i.
\]

Alexandrov’s proof is topological, not variational. It has been open for years to find a constructive proof.
**Variational Proof**

**Theorem (Gu-Luo-Sun-Yau 2013)**

\( \Omega \) is a compact convex domain in \( \mathbb{R}^n \), \( y_1, \cdots, y_k \) distinct in \( \mathbb{R}^n \), \( \mu \) a positive continuous measure on \( \Omega \). For any \( \nu_1, \cdots, \nu_k > 0 \) with \( \sum \nu_i = \mu(\Omega) \), there exists a vector \((h_1, \cdots, h_k)\) so that

\[
  u(x) = \max \{ \langle x, p_i \rangle + h_i \}
\]

satisfies \( \mu(W_i \cap \Omega) = \nu_i \), where \( W_i = \{ x | \nabla f(x) = p_i \} \). Furthermore, \( h \) is the maximum point of the concave function

\[
  E(h) = \sum_{i=1}^{k} \nu_i h_i - \int_0^h \sum_{i=1}^{k} w_i(\eta) d\eta_i,
\]

where \( w_i(\eta) = \mu(W_i(\eta) \cap \Omega) \) is the \( \mu \)-volume of the cell.
Outline of a variational Proof

Definition (Admissible Height Space)
Define admissible height space

\[ \mathcal{H} := \{(h_1, h_2, \cdots, h_k) | w_i(h) > 0, \forall i = 1, 2, \cdots, k \} \].

Lemma
The admissible height space \( \mathcal{H} \) is convex.

Proof.
Suppose \( h_0, h_1 \in \mathcal{H} \), construct the minkowski sum

\[ P((1 - t)h_0) \oplus P(th_1) = P((1 - t)h_0 + th_1), \]

By Brunn-Minkowski inequality, the volume of each face is positive, hence \( (1 - t)h_0 + th_1 \in \mathcal{H} \). \( \mathcal{H} \) is convex.
Variational Proof

Lemma

The following symmetric relation holds, \( w_i(h) \) is the area of face \( F_i \):

\[
\frac{\partial w_i(h)}{\partial h_j} = \frac{\partial w_j(h)}{\partial h_i} = -\frac{|e_{ij}|}{|\bar{e}_{ij}|} \leq 0.
\]

Proof.

\( \forall x \in e_{ij}, \langle p_i, x \rangle - h_i = \langle p_j, x \rangle - h_j \), hence

\( \langle p_i - p_j, x \rangle = h_i - h_j \). Change

\( h_i \to h_i + \delta h_i \), then \( x \to x + d, \quad |d| = \frac{\delta h_i}{|p_i - p_j|} \),

\[
\delta w_j = -|e_{ij}| |d| + o(\delta h_i^2) = -\frac{|e_{ij}|}{|p_i - p_j|} \delta h_i
\]

\( \bar{e}_{ij} = |p_i - p_j| \). 

David Gu (Stony Brook University) Computational Conformal Geometry September 1, 2022 23 / 48
Lemma

The energy

\[ E(\mathbf{h}) = \int_\mathbf{h} \sum_{i=1}^{k} w_i(\eta_i) d\eta_i \]

is well defined and strictly convex in the space

\[ \mathcal{H} \cap \{ \mathbf{h} | h_1 + h_2 + \cdots + h_k = 1 \}. \]

Proof.

Define a differential form, \( \omega = w_1(\mathbf{h}) dh_1 + w_2(\mathbf{h}) dh_2 + \cdots + w_k(\mathbf{h}) dh_k, \)

\[ d\omega = \sum_{i,j} \left( \frac{\partial w_j}{\partial h_i} - \frac{\partial w_i}{\partial h_j} \right) dh_i \wedge dh_j = 0. \]

\( \mathcal{H} \) is simply connected, \( \omega \) is closed, hence exact. \( \int_\mathbf{h} \omega \) is well defined.
The total area is fixed, $\sum_{i=1}^{k} w_i(h) = \text{vol}(\Omega)$, hence

$$\frac{\partial w_i}{\partial h_i} = - \sum_{j=1}^{k} \frac{\partial w_j}{\partial h_i} = - \sum_{j=1}^{k} \frac{\partial w_i}{\partial h_j} > 0,$$

all the off-diagonal elements are non-positive, the diagonal elements are positive. The Hessian matrix is diagonal dominant, with a null space $\{\lambda(1, 1, \cdots, 1)\}$. Hence the energy is strictly convex, the Hessian is positive definite on $\{\sum_{i=1}^{k} h_i = 1\} \cap \mathcal{H}$. 
One can define a cylinder through $\partial \Omega$, the cylinder is truncated by the xy-plane and the convex polyhedron. The energy term $\int h \sum w_i(\eta) d\eta_i$ equals to the volume of the truncated cylinder.
Definition (Alexandrov Potential)

The concave energy is

\[
E(h_1, h_2, \cdots, h_k) = \sum_{i=1}^{k} \nu_i h_i - \int_{0}^{h} \sum_{j=1}^{k} w_j(\eta) d\eta_j,
\]
Now we can prove Alexandrov’s theorem.

Proof.

The energy $E(h)$ is strictly concave. On the boundary $\Omega \cap \{h| \sum_{i=1}^{k} h_i = 1\}$, the gradient is given by

$$E(h) = (\nu_1 - w_1(h), \nu_2 - w_2(h), \cdots, \nu_k - w_k(h)),$$

The gradient points to the interior of the admissible space, hence the energy reaches maximum on an interior point $h^*$, where the gradient vanishes, namely $\nu_i = w_i(h^*)$. \qed
The gradient of the Alexanrov potential is the differences between the target measure and the current measure of each cell

\[ \nabla E(h_1, h_2, \cdots, h_k) = (\nu_1 - w_1, \nu_2 - w_2, \cdots, \nu_k - w_k) \]
The Hessian of the energy is the length ratios of edge and dual edges,

$$
\frac{\partial w_i}{\partial h_j} = -\frac{|e_{ij}|}{|\bar{e}_{ij}|}
$$
Convex Hull Algorithm

Input: A set of distinct points \( P = \{p_1, p_2, \cdots, p_k\} \subset \mathbb{R}^3 \);
Output: Convex hull of \( P \), \( \text{Conv}(P) \);

1. Use the first 4 points to construct a tetrahedron, adjust the order of the points, such that the volume of the tetrahedron is positive. Initialize \( \text{Conv}(P) \) as the tetrahedron;
2. Select the next point \( p_i \in P, p_i \not\in \text{Conv}(P) \);
3. Compute the visibility of all faces of \( \text{Conv}(P) \); remove all visible faces;
4. For all edges on the silhouette, connect the edge with \( p_i \) to form a new face. All the new faces with the invisible faces form the updated \( \text{Conv}(P) \);
5. Repeat step 2 through 4 until all points in \( P \) are processed.
Upper Envelope Algorithm

Input: A set of planes $\Pi = \{\pi_1, \pi_2, \cdots, \pi_k\}$;
Output: The upper envelope of $\Pi$, $\text{Env}(\Pi)$;

1. For each plane $\pi_i(x) = \langle x, y_i \rangle - h_i$, $y_i \in \mathbb{R}^2$, construct a dual point $\pi_i^* = (y_i, h_i)$;
2. Construct the convex hull of $\Pi^* := \{\pi_i^*\}$, $\text{Conv}(\Pi^*)$;
3. Remove all faces of $\text{Conv}(\Pi^*)$, whose normals are upwards;
4. Compute the Poincaré dual of $\text{Conv}(\Pi^*)$, each face $[\pi_i^*, \pi_j^*, \pi_k^*]$ corresponds to a vertex $\pi_i \cap \pi_j \cap \pi_k$; every edge $[\pi_i^*, \pi_j^*]$ corresponds to an edge $\pi_i \cap \pi_j$; every vertex $\pi_i^*$ corresponds to a face $\pi_i$. 
Optimal Transport Map

Input: A set of distinct points \( P = \{p_1, p_2, \cdots, p_k\} \), and the weights \( \{A_1, A_2, \cdots, A_k\} \); A convex domain \( \Omega \), \( \sum A_j = \text{Vol}(\Omega) \);
Output: The optimal transport map \( T : \Omega \rightarrow P \)

1. Scale and translate \( P \), such that \( P \subset \Omega \);
2. Initialize \( h^0 \leftarrow \frac{1}{2} (|p_1|^2, |p_2|^2, \cdots, |p_k|^2)^T \);
3. Compute the Brenier potential \( u(h^k) \) (envelope of \( \pi_i \)'s) and its Legendre dual \( u^*(h^k) \) (convex hull of \( \pi_i^* \)'s);
4. Project the Brenier potential and Legendre dual to obtain weighted Delaunay triangulation \( T(h^k) \) and power diagram \( D(h^k) \);
5. Compute the gradient of the energy

\[ \nabla E(h) = (A_1 - w_1(h), A_2 - w_2(h), \ldots, A_k - w_k(h))^T. \]

6. If \( \| E(h^k) \| \) is less than \( \varepsilon \), then return \( T = \nabla u(h^k) \);

7. Compute the Hessian matrix of the energy

\[ \frac{\partial w_i(h)}{\partial h_j} = -\frac{|e_{ij}|}{|\bar{e}_{ij}|}, \quad \frac{\partial w_i}{\partial h_i} = -\sum_j \frac{\partial w_i(h)}{\partial h_j}. \]

8. Solve linear system

\[ \nabla E(h) = \text{Hess}(h^k)d; \]
11 Set the step length $\lambda \leftarrow 1$;
12 Construct the convex hull $\text{Conv}(h^k + \lambda d)$;
13 if there is any empty power cell, $\lambda \leftarrow \frac{1}{2} \lambda$, repeat step 3 and 4, until all power cells are non-empty;
14 set $h^{k+1} \leftarrow h^k + \lambda d$;
15 Repeat step 3 through 14.
Regularity of Optimal Transportation Map

Theorem (Ma-Trudinger-Wang)

The potential function \( u \) is \( C^3 \) smooth if the cost function \( c \) is smooth, \( f, g \) are positive, \( f \in C^2(\Omega), \ g \in C^2(\Omega^*) \), and

- **A1** \( \forall x, \xi \in \mathbb{R}^n, \exists! y \in \mathbb{R}^n, \ s.t. \ \xi = D_x c(x, y) \) (for existence)
- **A2** \( |D^2_{xy} c| \neq 0 \).
- **A3** \( \exists c_0 > 0 \ s.t. \ \forall \xi, \eta \in \mathbb{R}^n, \ \xi \perp \eta \)

\[
\sum (c_{ij,rs} - c^{p,q} c_{ij,p} c_{rs,q}) c^{r,k} c^{s,l} \xi_i \xi_j \eta_k \eta_l \geq c_0 |\xi|^2 |\eta|^2.
\]

- **B1** \( \Omega^* \) is \( c \)-convex w.r.t. \( \Omega \), namely \( \forall x_0 \in \Omega \),

\[
\Omega^*_{x_0} := D_x c(x_0, \Omega^*)
\]

is convex.
Definition (subgradient)

Given an open set $\Omega \subset \mathbb{R}^d$ and $u : \Omega \to \mathbb{R}$ a convex function, for $x \in \Omega$, the subgradient (subdifferential) of $u$ at $x$ is defined as

$$\partial u(x) := \{ p \in \mathbb{R}^n : u(z) \geq u(x) + \langle p, z - x \rangle \ \forall z \in \Omega \}.$$ 

The Brenier potential $u$ is differentiable at $x$ if its subgradient $\partial u(x)$ is a singleton. We classify the points according to the dimensions of their subgradients, and define the sets

$$\Sigma_k(u) := \left\{ x \in \mathbb{R}^d \mid \dim(\partial u(x)) = k \right\}, \quad k = 0, 1, 2 \ldots, d.$$
Theorem (Figalli Regularity)

Let $\Omega, \Lambda \subset \mathbb{R}^d$ be two bounded open sets, let $f, g : \mathbb{R}^d \to \mathbb{R}^+ \to R^+$ be two probability densities, that are zero outside $\Omega, \Lambda$ and are bounded away from zero and infinity on $\Omega, \Lambda$, respectively. Denote by $T = \nabla u : \Omega \to \Lambda$ the optimal transport map provided by Brenier theorem. Then there exist two relatively closed sets $\Sigma_{\Omega} \subset \Omega$ and $\Sigma_{\Lambda} \subset \Lambda$ with $|\Sigma_{\Omega}| = |\Sigma_{\Lambda}| = 0$ such that $T : \Omega \setminus \Sigma_{\Omega} \to \Lambda \setminus \Sigma_{\Lambda}$ is a homeomorphism of class $C^{0, \alpha}_{loc}$ for some $\alpha > 0$. 

David Gu (Stony Brook University)
Singularity Set of OT Maps

We call $\Sigma_\Omega$ as singular set of the optimal transportation map $\nabla u : \Omega \to \Lambda$. 

Figure: Singularity structure of an optimal transportation map.
Figure: Discontinuous Optimal transportation map, produced by a GPU implementation of algorithm based on our theorem. The middle line is the singularity set $\Sigma_1$. 
Figure: Discontinuous Optimal transportation map, produced by a GPU implementation of algorithm based on regularity theorem. $\gamma_1$ and $\gamma_2$ are two singularity sets.
Discontinuity of Optimal Transportation Map

Figure: Optimal transportation between a solid ball to the Stanford bunny. The singular sets are the foldings on the boundary surface.
Figure: Optimal transportation between a solid ball to the Stanford bunny. The singular sets are the foldings on the boundary surface.
Figure: Optimal transportation map is discontinuous, but the Brenier potential itself is continuous. The projection of ridges are the discontinuity singular sets.
Figure: Optimal transportation map.
Figure: Optimal transportation map.
Figure: Optimal transportation map is discontinuous.
Figure: Optimal transportation map is discontinuous.