Optimal Transportation: Convex Geometric View

David Gu

Yau Mathematics Science Center
Tsinghua University
Computer Science Department
Stony Brook University

gu@cs.stonybrook.edu

August 6, 2020

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.

Optimal Transportation Map

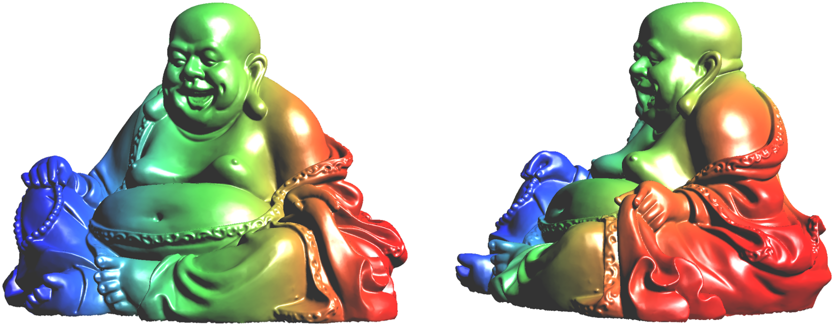


Figure: Buddha surface.

Optimal Transportation Map

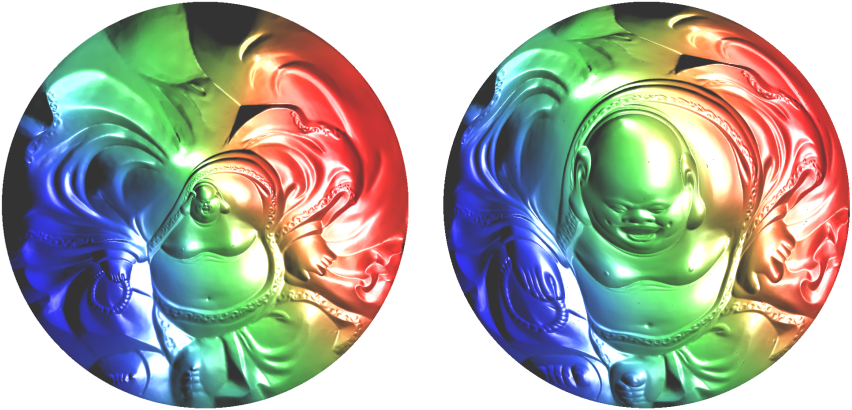


Figure: Optimal transportation map.

Optimal Transportation Map

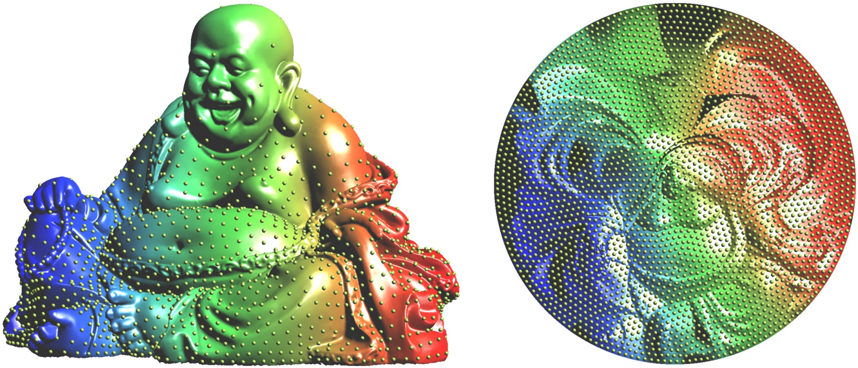


Figure: Brenier potential.

Optimal Transportation Map

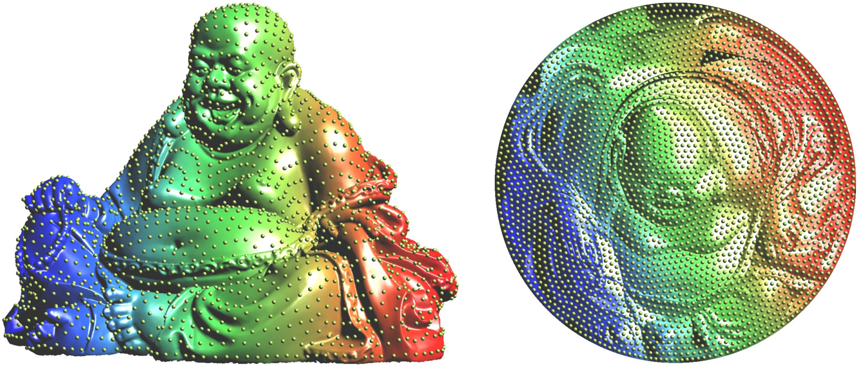


Figure: Brenier potential.

Optimal Transportation Map

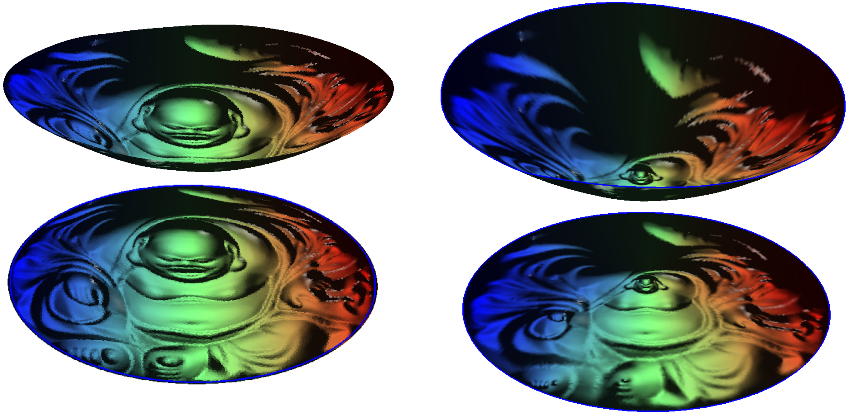


Figure: Brenier potential.

Convex Geometric View

Monge-Ampére Equation

Problem (Brenier)

Given (Ω, μ) *and* (Σ, ν) *and the cost function* $c(x,y) = \frac{1}{2}|x-y|^2$, *the optimal transportation map* $T : \Omega \to \Sigma$ *is the gradient map of the Brenier potential* $u : \Omega \to \mathbb{R}$, *which satisfies the Monge-Ampére equation,*

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

Monge-Ampére Equation

Problem (prescribed Gauss curvature)

Suppose that a real-valued function K is specified on a domain Ω 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 $K(x)$.

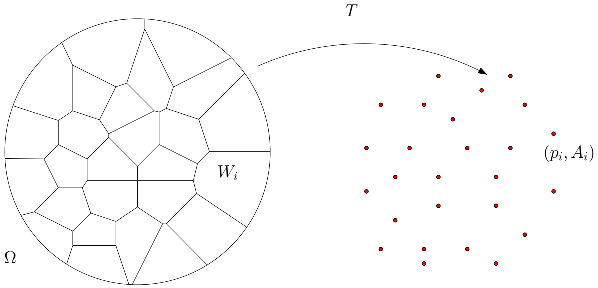
$\mathbf{r}(x,y) = (x, y, u(x,y))$, $\mathbf{r}_x = (1, 0, u_x)$, $\mathbf{r}_y = (0, 1, u_y)$, $\mathbf{n} = \frac{(-u_x, -u_y, 1)}{\sqrt{1+|\nabla u|^2}}$,

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

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

$K(x,y) = \frac{u_{xx}u_{yy} - u_{xy}^2}{(1+|\nabla u|^2)^2}$, Geneal case $\boxed{K(x)(1+|\nabla u|^2)^{\frac{n+2}{2}} = \det D^2 u}$

Semi-Discrete Optimal Transportation Problem

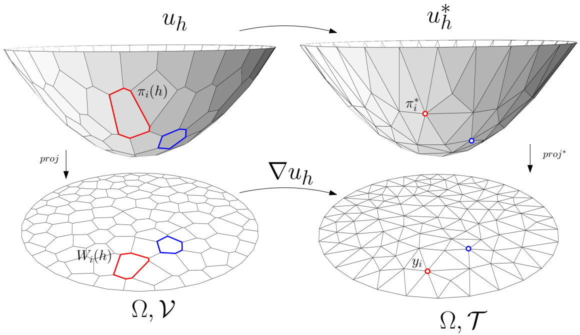


Problem (Semi-discrete OT)

Given a compact convex domain Ω 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 $T : \Omega \to \{p_1, \ldots, p_k\}$, such that $vol(T^{-1}(p_i)) = w_i$, so that T minimizes the transportation cost:

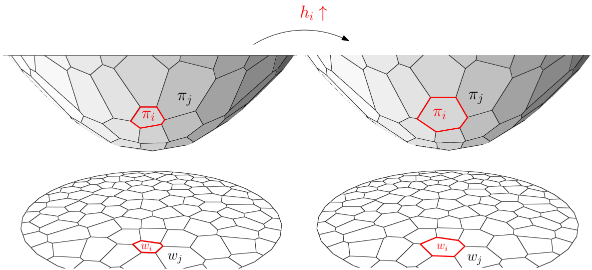
$$\mathcal{C}(T) := \frac{1}{2} \int_{\Omega} |x - T(x)|^2 dx$$

Semi-Discrete Optimal Transportation Problem



According to Brenier theorem, there will be a piecewise linear convex function $u : \Omega \to \mathbb{R}$, the gradient map gives the optimal transport map.

Semi-Discrete Optimal Transportation Problem



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,

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

Minkowski problem - General Case

Theorem

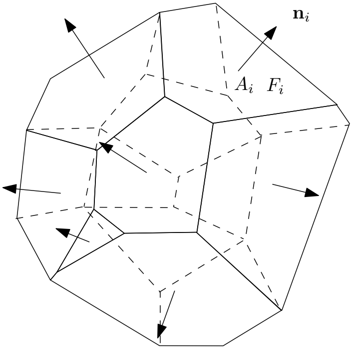
Minkowski Given k unit vectors $\mathbf{n}_1, \cdots, \mathbf{n}_k$ not contained in a half-space in $\mathbb{R}^n$ and $A_1, \cdots, A_k > 0$, such that

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

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

1. *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 $0 \leq t \leq 1$,

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

For convex sets A and B, the inequality is strick 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 $\langle x, \mathbf{n}_i \rangle = h_i$, the hyper-planes support a convex polytope $P(h_1, h_2, \dots, h_k)$, we maximize the volume of $P(h)$,

$$\max_{\mathbf{h}} \mathrm{Vol}(P(h_1, h_2, \dots, h_k))$$

under the constraint

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

We use Lagrange multiplier method,

$$\max_{\mathbf{h},\lambda} Vol(P(\mathbf{h})) - \lambda \left(\sum_{i=1}^{k} h_i A_i - 1 \right),$$

$\square$

Minkowski Theorem

continued

We define admissible space of the heights

$$\mathcal{H} := \{\mathbf{h} | w_i(\mathbf{h}) > 0, i = 1, 2, \cdots, k\}$$

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

$$\frac{\partial Vol(P(\mathbf{h}))}{\partial h_i} - \lambda A_i = w_i(\mathbf{h}) - \lambda A_i < 0,$$

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

$$(w_1(\mathbf{h}^*), w_2(\mathbf{h}^*), \cdots, w_k(\mathbf{h}^*)) = \lambda(A_1, A_2, \cdots, A_k).$$

□

Alexandrov Theorem

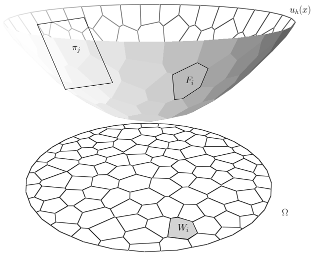
Theorem (Alexandrov 1950)

Given Ω 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 $\sum A_i = Vol(\Omega)$, there exists PL convex function

$$f(\mathbf{x}) := \max\{\langle \mathbf{x}, \mathbf{p}_i \rangle + h_i | i = 1, \cdots, k\}$$

unique up to translation such that

$$Vol(W_i) = Vol(\{\mathbf{x} | \nabla f(\mathbf{x}) = \mathbf{p}_i\}) = A_i.$$



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)

Ω is a compact convex domain in $\mathbb{R}^n$, $y_1, \cdots, y_k$ distinct in $\mathbb{R}^n$, μ a positive continuous measure on Ω. 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

$$u(\mathbf{x}) = \max\{\langle \mathbf{x}, \mathbf{p}_i \rangle + h_i\}$$

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

$$E(\mathbf{h}) = \sum_{i=1}^{k} \nu_i h_i - \int_{\mathbf{0}}^{\mathbf{h}} \sum_{i=1}^{k} w_i(\eta) d\eta_i,$$

where $w_i(\eta) = \mu(W_i(\eta) \cap \Omega)$ is the μ-volume of the cell.

Outline of a variational Proof

Definition (Admissible Height Space)

Define admissible height space

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

Lemma

The admissible height space $\mathcal{H}$ is convex.

Proof.

Suppse $\mathbf{h}_0, \mathbf{h}_1 \in \mathcal{H}$, construct the minkowski sum

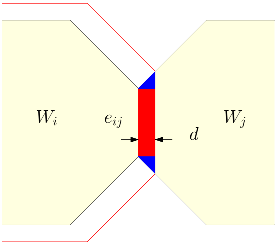
$$P((1-t)\mathbf{h}_0) \oplus P(t\mathbf{h}_1) = P((1-t)\mathbf{h}_0 + t\mathbf{h}_1),$$

By Brunn-Minkowski inequality, the volume of each face is positive, hence $(1-t)\mathbf{h}_0 + t\mathbf{h}_1 \in \mathcal{H}$. $\mathcal{H}$ is convex. $\square$

Lemma

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

$$\frac{w_i(\mathbf{h})}{\partial h_j} = \frac{w_j(\mathbf{h})}{\partial h_i} = -\frac{|e_{ij}|}{|\bar{e}_{ij}|}.$$



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 $x \to x + d$, $|d| = \frac{\delta h_i}{|p_i - p_j|}$,

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

□

Variational Proof

Lemma

The energy

$$E(\mathbf{h}) = \int^{\mathbf{h}} \sum_{i=1}^{k} w_i(\eta) d\eta_i$$

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, ω is closed, hence exact. $\int^{\mathbf{h}} \omega$ is well defined. □

Variational Proof

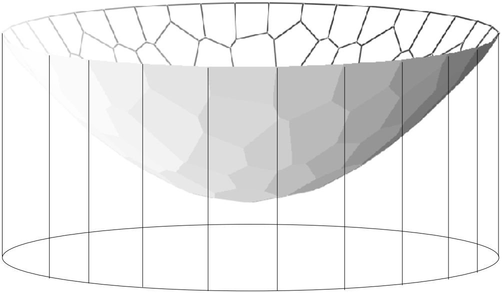
continued

The total area is fixed, $\sum_{i=1}^{k} w_i(\mathbf{h})$, hence

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

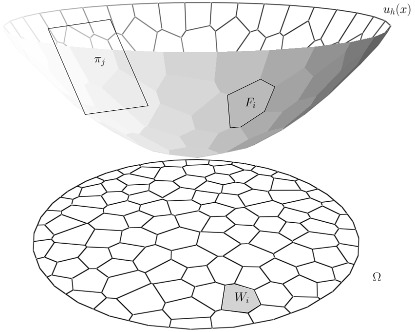
all the off-diagonal elements are non-negative. The Hessian matrix is diagonal dominant, with a null space $\{\lambda(1,1,\cdots,1)\}$. Hence the energy is positive definite on $\{\sum_{i=1}^{k} h_i = 1\} \cap \mathcal{H}$.

Geometric Interpretation



One can define a cylinder through $\partial\Omega$, the cylinder is truncated by the xy-plane and the convex polyhedron. The energy term $\int^{\mathbf{h}} \sum w_i(\eta) d\eta_i$ equals to the volume of the truncated cylinder.

Computational Algorithm



Definition (Alexandrov Potential)

The concave energy is

$$E(h_1, h_2, \cdots, h_k) = \sum_{i=1}^{k} \nu_i h_i - \int_{\mathbf{0}}^{\mathbf{h}} \sum_{j=1}^{k} w_j(\eta) d\eta_j,$$

Existence Proof

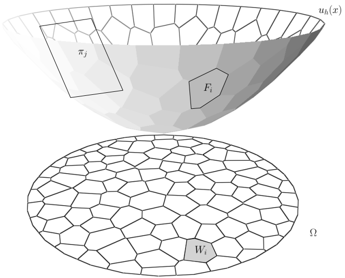
Proof.

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

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

The gradient points to the interior of the admissible space, hence the energy reaches maximum on an interior point $\mathbf{h}^*$, where the gradient vanishes, namely $\nu_i = w_i(\mathbf{h}^*)$. $\square$

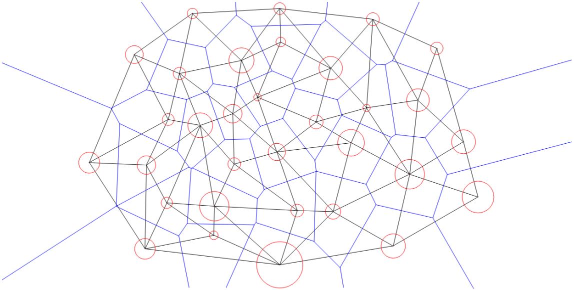
Computational Algorithm



The gradient of the Alexanrov potential is the differences between the target measure and the current measure of each cell

$$\nabla E(h_1, h_2, \cdots, h_k) = (\nu_1 - w_1, \nu_2 - w_2, \cdots, \nu_k - w_k)$$

Computational Algorithm



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 $p_i \in P$, $p_i \notin \text{Conv}(P)$;
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 Π, $\text{Env}(\Pi)$;

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 π_i^* corresponds to a face π_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 Ω, $\sum A_j = \mathrm{Vol}(\Omega)$;
Output: The optimal transport map $T : \Omega \to P$

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

Optimal Transport Map

5. Compute the gradient of the energy

$$\nabla E(\mathbf{h}) = (A_1 - w_1(\mathbf{h}), A_2 - w_2(\mathbf{h}), \cdots, A_k - w_k(\mathbf{h}))^T.$$

6. If $\|E(\mathbf{h}^k)\|$ is less than ε, then return $T = \nabla u(\mathbf{h}^k)$;
7. Compute the Hessian matrix of the energy

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

8. Solve linear system

$$\nabla E(\mathbf{h}) = \text{Hess}(\mathbf{h}^k)\mathbf{d};$$

11. Set the step length $\lambda \leftarrow 1$;
12. Construct the convex hull $\text{Conv}(\mathbf{h}^k + \lambda \mathbf{d})$;
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 $\mathbf{h}^{k+1} \leftarrow \mathbf{h}^k + \lambda \mathbf{d}$;
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.* $\forall \xi, \eta \in \mathbb{R}^n$, $\xi \perp \eta$

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

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

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

is convex.

Subgradient

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

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

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

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

Theorem (Figalli Regularity)

Let $\Omega, \Lambda \subset \mathbb{R}^d$ be two bounded open sets, let $f, g : \mathbb{R}^d \to \mathbb{R}^+$ be two probability densities, that are zero outside Ω, Λ and are bounded away from zero and infinity on Ω, Λ, 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 $\alpha > 0$.

Singularity Set of OT Maps

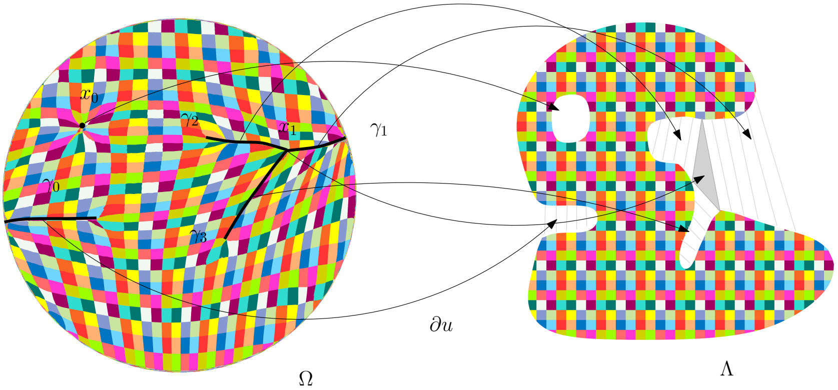


Figure: Singularity structure of an optimal transportation map.

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

Discontinuity of Optimal Transportation Map

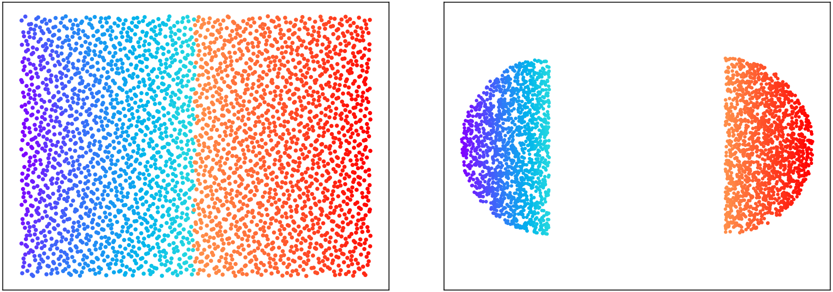


Figure: Discontinuous Optimal transportation map, produced by a GPU implementation of algorithm based on our theorem. The middle line is the singularity set Σ_1.

Discontinuity of Optimal Transportation Map

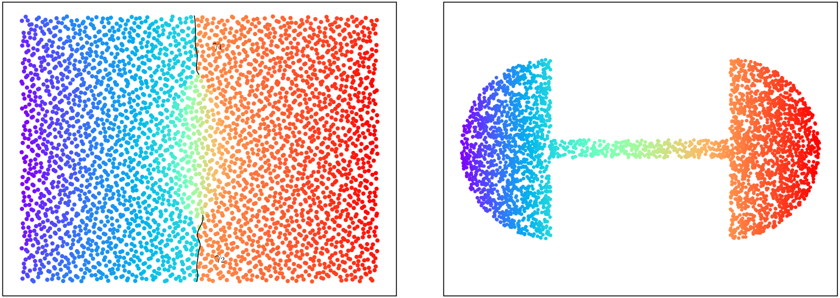


Figure: Discontinuous Optimal transportation map, produced by a GPU implementation of algorithm based on regularity theorem. γ_1 and γ_2 are two singularity sets.

Discontinuity of Optimal Transportation Map

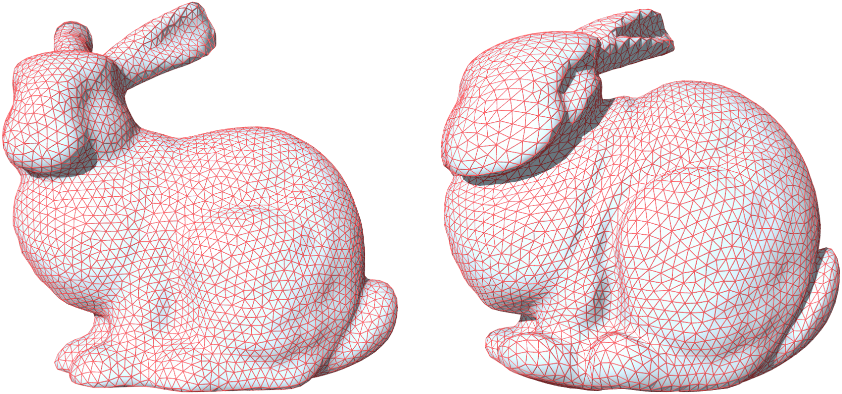


Figure: Optimal transportation between a solid ball to the Stanford bunny. The singular sets are the foldings on the boundary surface.

Discontinuity of Optimal Transportation Map

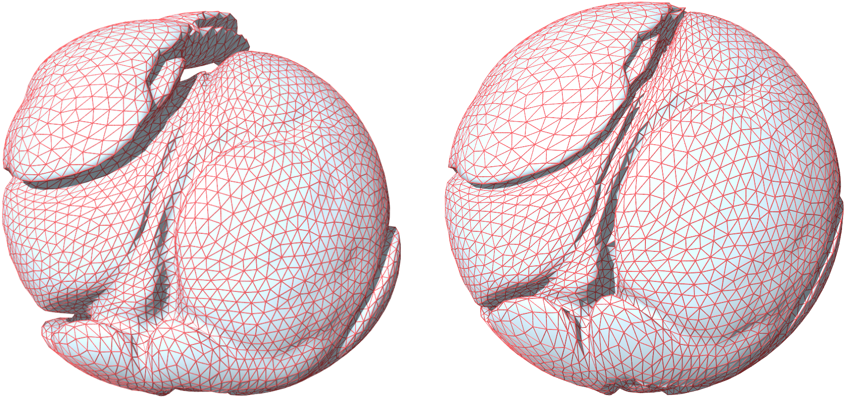


Figure: Optimal transportation between a solid ball to the Stanford bunny. The singular sets are the foldings on the boundary surface.

Discontinuity of Optimal Transportation Map

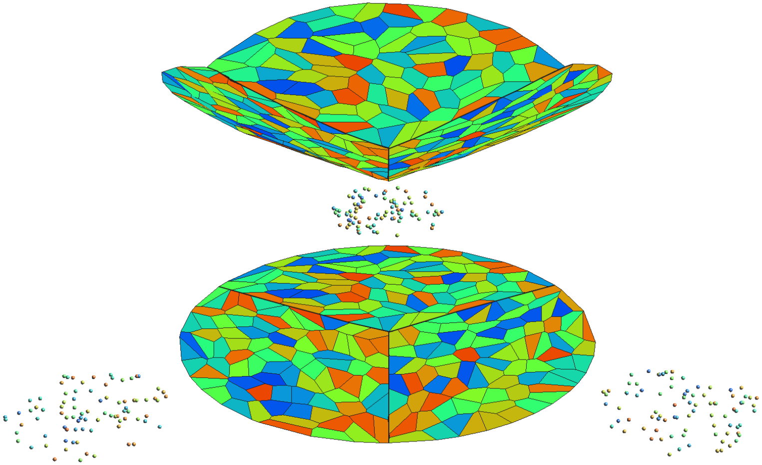


Figure: Optimal transportation map is discontinuous, but the Brenier potential itself is continuous. The projection of ridges are the discontinuity singular sets.

Optimal Transportation Map

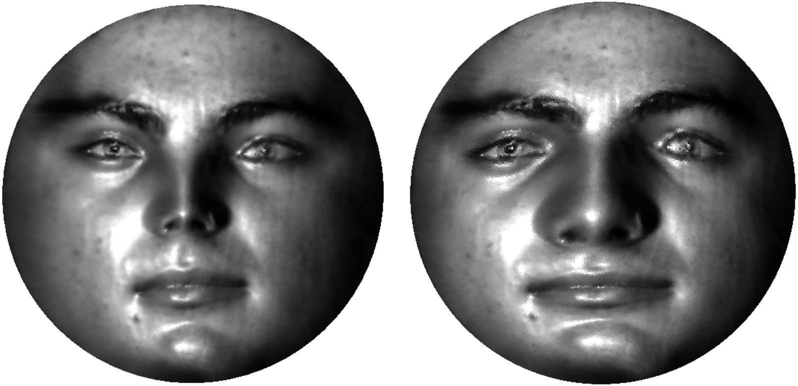


Figure: Optimal transportation map.

Optimal Transportation Map

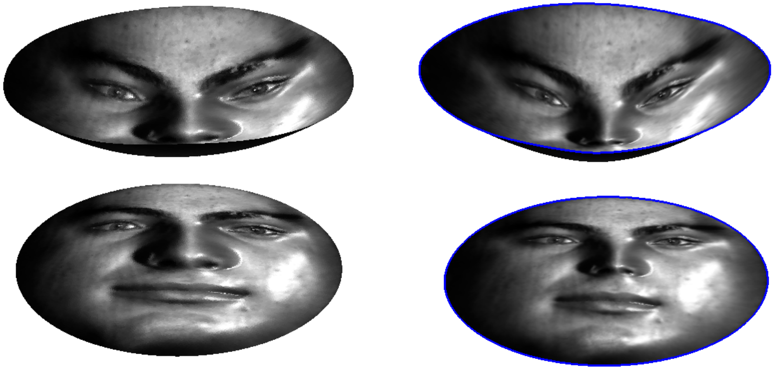


Figure: Optimal transportation map.

Discontinuity of Optimal Transportation Map

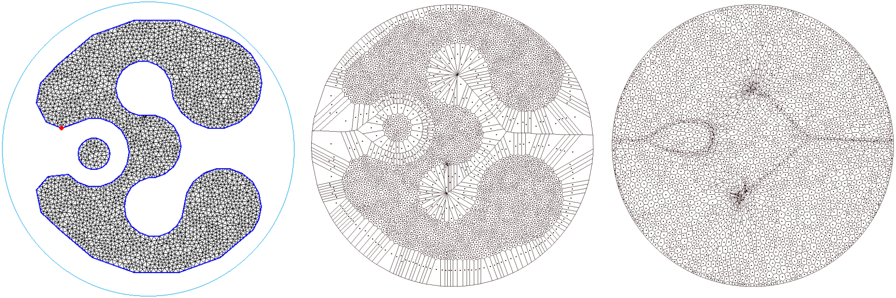


Figure: Optimal transportation map is discontinuous.

Discontinuity of Optimal Transportation Map

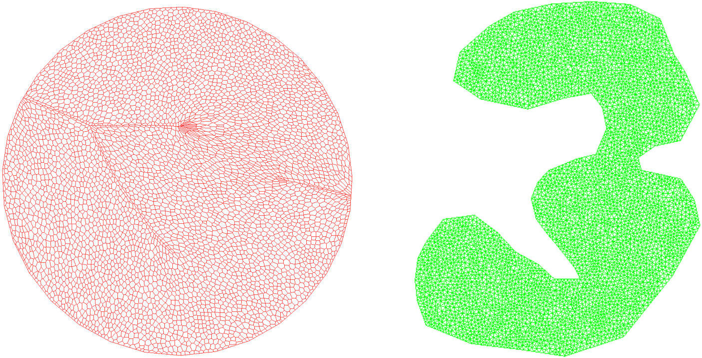


Figure: Optimal transportation map is discontinuous.

Optimal Transportation

Solution

- Optimal transport theory discovers a natural Riemannian metric of $\mathcal{P}(X)$, called Wasserstein metric;
- the covariant calculus on $\mathcal{P}(X)$ can be defined accordingly;
- the optimization in $\mathcal{P}(X)$ can be carried out.

Optimal Transportation

- The geodesic distance between $d\mu = f(x)dx$ and $d\nu(y) = g(y)dy$ is given by the optimal transport map $T : X \to X$, $T = \nabla u$,

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

- The geodesic between them is McCann's displacement,

$$\gamma(t) := ((1-t)I + t\nabla u)_{\#}\mu$$

- The tangent vectors of a probability measure is a gradient field on X, the Riemannian metric is given by

$$\langle d\varphi_1, d\varphi_2 \rangle = \int_X \langle d\varphi_1, d\varphi_2 \rangle_{\mathbf{g}} f(x)dx.$$

Optimal Transportation

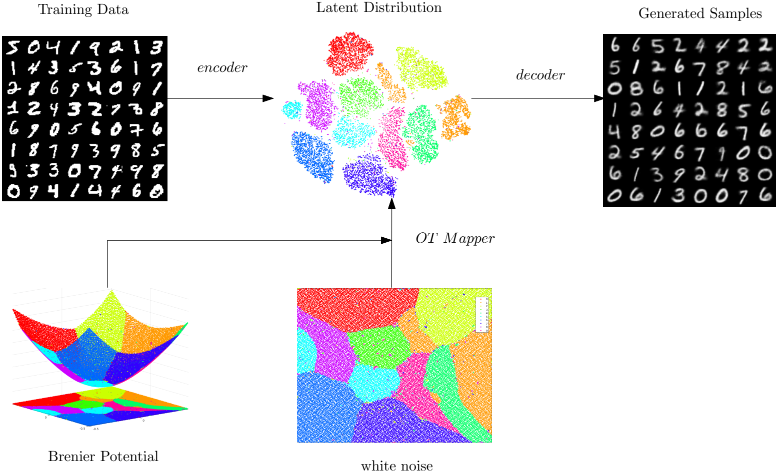


Figure: Geometric Generative Model.