

Optimal Transportation Theory and Computation

Computational Methods



David Gu
Computer Science Department
Stony Brook University

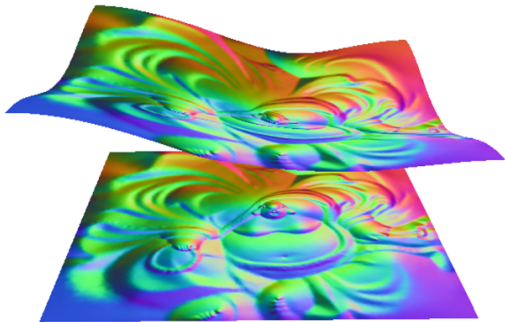


图: Textbook: Optimal Transportation Theory and Computation

FFT-OT Algorithm



(a). Buddha Surface (S, \mathbf{g})



(b). Potential φ

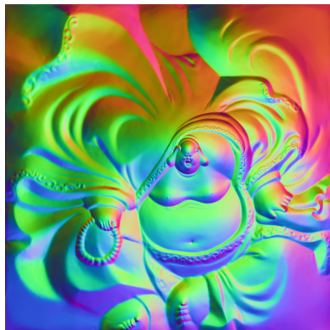
(c). Conformal Mapping ψ (d). Optimal Transportation map T

图: FFT-OT algorithm.

问题 (Monge-Ampère Equation)

Given probability measures $f(x)dx$, $g(y)dy$ defined on planar domains Ω and Ω^* ,

$$\int_{\Omega} f(x) dx = \int_{\Omega^*} g(y) dy.$$

find a convex function $u : \Omega \rightarrow \mathbb{R}$, satisfying the Monge-Ampère equation:

$$\det D^2 u(x) = \frac{f(x)}{g \circ Du(x)}, \quad (1)$$

with the natural boundary condition: $Du(\Omega) = \Omega^*$.

Two dimensional Monge-Ampère equation can be written as

$$u_{xx}u_{yy} - u_{xy}^2 = f/g \circ Du.$$

hence

$$(u_{xx} + u_{yy})^2 = u_{xx}^2 + u_{yy}^2 + 2u_{xy}^2 + 2\frac{f}{g \circ Du}, \quad (2)$$

convert to Poisson equation,

$$\Delta u = \sqrt{u_{xx}^2 + u_{yy}^2 + 2u_{xy}^2 + 2f/g \circ Du}, \quad (3)$$

where $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$. Brenier potential function is convex, Laplacian is non-negative, take the positive square root.

Define the operator $\mathcal{T} : H^2(\Omega) \rightarrow H^2(\Omega)$,

$$\mathcal{T}[u] = \Delta^{-1} \left\{ \sqrt{u_{xx}^2 + u_{yy}^2 + 2u_{xy}^2 + 2f/g \circ Du} \right\} \quad (4)$$

The solution to the Monge-Ampère is the fixed point of \mathcal{T} , $\mathcal{T}(u^*) = u^*$. Use iteration method to obtain the fixed point $u^{(n+1)} = \mathcal{T}[u^{(n)}]$.

引理 (Oblique Boundary Condition)

Assume $\Omega, \Omega^ \subset \mathbb{R}^n$ is a bounded domain, Ω is convex, $\partial\Omega^*$ is C^1 smooth. The density functions f and g satisfy the balance condition $\int_{\Omega} f = \int_{\Omega^*} g$, and are bounded, $0 < c_0 < f, g < c_1 < \infty$, the Brenier potential is $u : \Omega \rightarrow \mathbb{R}$. Suppose $x \in \partial\Omega$ and $y \in \partial\Omega^*$, $Du(x) = y$, then $\langle \mathbf{n}(x), \mathbf{n}(y) \rangle > 0$, where \mathbf{n} represents the inner normal vector.*

Suppose $\Omega = \Omega^* = [-1, +1] \times [-1, +1]$, the four corner points are p_k , $k = 0, 1, 2, 3$, divide the boundary into four segments γ_k . Oblique boundary condition implies: for each boundary point $x \in \gamma_k$,

$$\mathbf{n}(x) = \mathbf{n}(T(x)), \quad (5)$$

at the corner points p_k 's, $T(p_k) = p_k$.

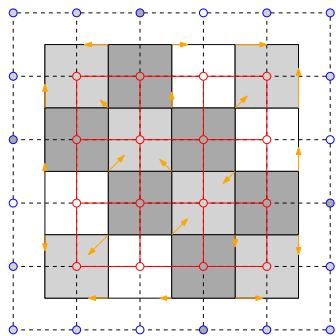
Let $u(x, y) = \varphi(x, y) + (x^2 + y^2)/2$, φ is called the *potential*. We obtain $Du = D\varphi + \text{Id}$ and $\Delta u = \Delta\varphi + 2$. The operator $\mathcal{T}[u]$ is changed to $\mathcal{P}[\varphi]$,

$$\mathcal{P}[\varphi^{(n+1)}] := \Delta^{-1} \mathcal{F}[\varphi^{(n)}]$$

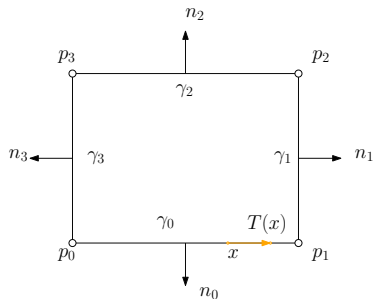
where

$$\mathcal{F}(\varphi) := \left\{ \sqrt{(\varphi_{xx} + 1)^2 + (\varphi_{yy} + 1)^2 + 2\varphi_{xy}^2 + 2f/g \circ (\text{Id} + D\varphi)} - 2 \right\}.$$

the obliqueness boundary condition is converted to the Neumann boundary condition $\partial\varphi/\partial\mathbf{n} = 0$.



(a) grid structure



(b) Obliqueness condition

图: The samples (red) and the shadow points (blue) are centers of cells. The obliqueness condition is equivalent to the Neumann boundary condition.

Finite Difference Operator:

$$\begin{aligned}\mathcal{D}_{xx}^2 u_{ij} &= \frac{1}{h_x^2} (u_{i+1,j} + u_{i-1,j} - 2u_{i,j}) \\ \mathcal{D}_{yy}^2 u_{ij} &= \frac{1}{h_y^2} (u_{i,j+1} + u_{i,j-1} - 2u_{i,j}) \\ \mathcal{D}_{xy}^2 u_{ij} &= \frac{1}{4h_x h_y} (u_{i+1,j+1} + u_{i-1,j-1} \\ &\quad - u_{i-1,j+1} - u_{i+1,j-1})\end{aligned}\tag{6}$$

Discrete Poisson Equation:

$$u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1} - 4u_{ij} = \rho_{i,j},\tag{7}$$

Given a two dimensional array $u(i, j)$, the two dimensional Discrete Cosine Transformation is defined as:

$$\tilde{u}(m, n) = c(m, n) \sum_{i,j} u(i, j) \cos \frac{(2i+1)m\pi}{2M} \cos \frac{(2j+1)n\pi}{2N},$$

The inverse DCT is defined as:

$$u(i, j) = \sum_{m,n} c(m, n) \tilde{u}(m, n) \cos \frac{(2i+1)m\pi}{2M} \cos \frac{(2j+1)n\pi}{2N},$$

where $m, i = 0, 1, \dots, M-1$ and $n, j = 0, 1, \dots, N-1$,

$$c(m, n) = \begin{cases} \frac{\sqrt{2}}{\sqrt{MN}} & m = 0, n = 0 \\ \frac{2}{\sqrt{MN}} & \text{otherwise} \end{cases}$$

引理

Given a discrete Poisson equation with Neumann boundary condition:

$$\Delta u = \rho, \quad \frac{\partial u}{\partial \mathbf{n}} = 0.$$

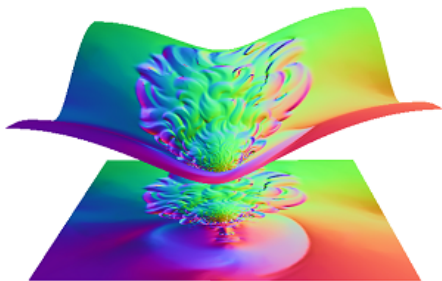
Let $\tilde{\rho} = DCT(\rho)$, $\tilde{u} = DCT(u)$, we have

$$\tilde{u}(m, n) = \frac{\tilde{\rho}(m, n)}{2[\cos \frac{m\pi}{M} + \cos \frac{n\pi}{N} - 2]}. \quad (8)$$

Different solutions differ by a constant. Let $\tilde{u}(0, 0)$ be 0, we obtain the unique solution. DCT and IDCT can be calculated by FFT on GPU.

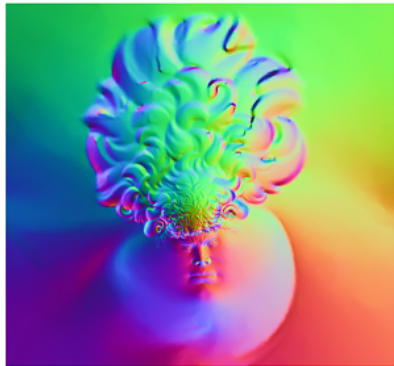


(a). David Head

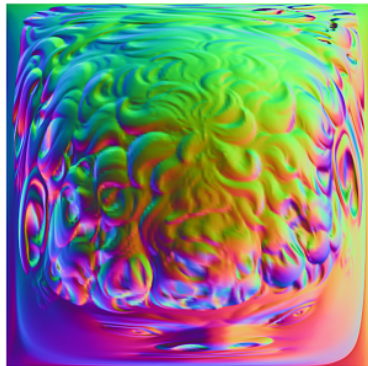


(b). Potential φ

图: FFT-OT David head.



(c). Conformal Map



(d). Optimal Transportation map

图: FFT-OT David head sculpture.

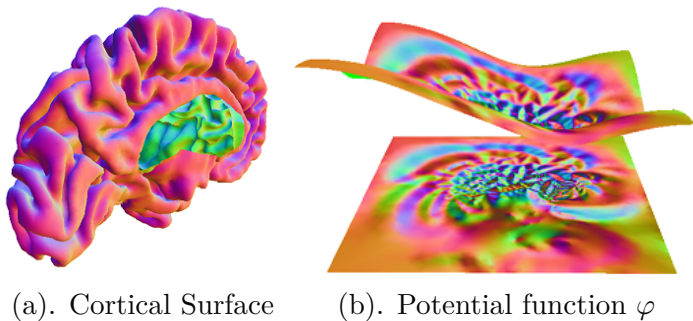
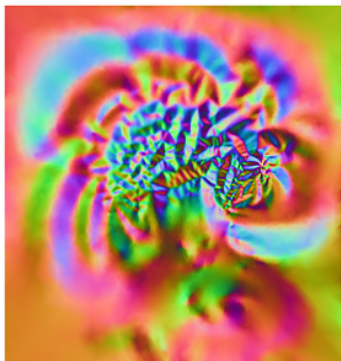


图: FFT-OT cortical surface.



(c). Conformal map

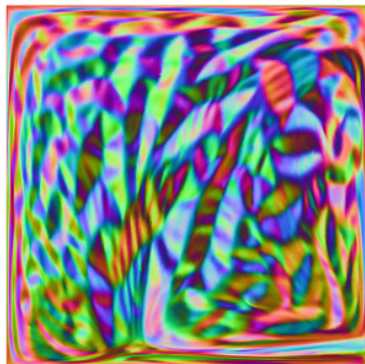
(d). Optimal Transportation map T

图: FFT-OT cortical surface.

Linearized Monge-Ampère Equation

In the neighborhood of the identity matrix, the linearization of the determinant is

$$\det(I + \varepsilon N) = 1 + \varepsilon \operatorname{tr}[N] + O(\varepsilon^2).$$

In a neighborhood of a matrix M , linearization of determinant is

$$\begin{aligned}\det(M + \varepsilon N) &= \det(M) \det(I + \varepsilon M^{-1} N) \\ &= \det(1 + \varepsilon \operatorname{tr}[M^{-1} N] + O(\varepsilon^2)) \\ &= \det(M) + \varepsilon \operatorname{tr}[\det(M) M^{-1} N] + O(\varepsilon^2) \\ &= \det(M) + \varepsilon \operatorname{tr}[M_{adj} N] + O(\varepsilon^2),\end{aligned}$$

where $M_{adj} := \det(M) M^{-1}$. Linearization of determinant is:

$$\nabla_M \det(M)[N] := \operatorname{tr}(M_{adj} N).$$

When $u \in C^2(\Omega; \mathbb{R})$, the linearization of the Monge-Ampère operator:

$$\nabla_u \det(D^2 u)[v] = \operatorname{tr}((D^2 u)_{adj} D^2 v) = \sum_{p,q=1}^n u^{pq}(x) \partial_p \partial_q v(x).$$

where the adjoint matrix of the Hessian matrix is $(u^{pq}) = (D^2 u)_{adj}$ where $\partial_p \partial_q := \frac{\partial^2}{\partial x_p \partial x_q}$. Because u is strictly convex, the Hessian matrix is positive definite, the operator

$$\mathcal{L}[v] = \sum_{p,q=1}^n u^{pq}(x) \partial_p \partial_q v(x)$$

is a non-degenerated elliptic operator.

In order to solve Monge-Ampère equation $\det D^2 u(x) = f(x)$, define the density flow

$$\rho(x, t) = (1 - t) + tf(x), \quad t \in [0, 1].$$

the corresponding Brenier potential is $u(x, t) : \Omega \times [0, 1] \rightarrow \mathbb{R}$,

$$\det D_x^2 u(x, t) = \rho(x, t), \quad \nabla_x u(x, t)(\Omega) = \Omega.$$

Let $v(x, t) = \dot{u}(x, t)$, the velocity of the Brenier potential,

$$\sum_{p, q=1}^n u^{pq}(x, t) \partial_p \partial_q v(x, t) = \frac{\partial}{\partial t} \rho(x, t) = f(x) - 1$$

with the oblique boundary condition, at the time $t = 0$,
 $u(x, 0) = 1/2 \|x\|^2$.

Angenent-Haker-Tannenbaum Algorithm

Given a compact set Ω with regular boundary, a smooth density function $\rho : \Omega \rightarrow \mathbb{R}^+$, target measure ν , a smooth cost function c . Find the Monge-Kantorovich Optimal Transportation map. Starting from an measure preserving map T , evolving with time t to reduce the transportation cost gradually

$$M(T) = \int_{\Omega} c(x, T(x))\rho(x)dx.$$

for each time t , construct a diffeomorphism $g_t : \Omega \rightarrow \Omega$, preserving the measure ρ , $(g_t)_\# \rho = \rho$. Compose T with g_t , replace T by $T \circ (g_t)^{-1}$, such that the transportation cost of $T \circ (g_t)^{-1}$ decreases monotonously.

Construct time dependent velocity field v_t to obtain a family of diffeomorphisms g_t , the orbit of a particle is $\gamma_x(t)$, satisfying the equation

$$\begin{cases} \frac{d}{dt}\gamma_x(t) &= v_t(\gamma_x(t)), \\ \gamma_x(0) &= x, \end{cases}$$

where $g_t(x) := \gamma_x(t)$, by continuity equation

$$\partial_t \rho_t + \nabla \cdot (\rho_t v_t) = 0, \quad \rho_0 = \rho.$$

The velocity field satisfies the equation $\nabla \cdot (\rho_t v_t) = 0$.

Let $x' = \gamma_x(-t)$, then $x = \gamma_{x'}(t)$, hence

$$c(x, T(\gamma_x(-t))) = c(\gamma_{x'}(t), T(x')).$$

Since g_{-t} is measure preserving, we have $\rho(x)dx = \rho(x')dx'$, hence

$$\begin{aligned} & \int_{\Omega} c(x, T(\gamma_x(-t)))\rho(x)dx \\ &= \int_{\Omega} c(\gamma_{x'}(t), T(x'))\rho(x')dx' \\ &= \int_{\Omega} c(\gamma_x(t), T(x))\rho(x)dx \end{aligned}$$

Therefore

$$\begin{aligned} & \frac{d}{dt} \int_{\Omega} c(x, T(\gamma_x(-t))) \rho(x) dx \\ &= \frac{d}{dt} \int_{\Omega} c(\gamma_x(t), T(x)) \rho(x) dx \\ &= \int_{\Omega} \nabla_x c(\gamma_x(t), T(x)) \cdot \gamma'_x(t) \rho(x) dx \\ &= \int_{\Omega} \nabla_x c(\gamma_x(t), T(x)) \cdot v(\gamma_x(t)) \rho(x) dx \end{aligned}$$

when $t = 0$, the derivative equals to

$$\int_{\Omega} \nabla_x c(x, T(x)) \cdot v(x) \rho(x) dx = \int_{\Omega} \xi(x) \cdot w(x) dx,$$

The derivative of the total transportation cost equals to

$$\langle \xi, w \rangle_{L^2} = \int_{\Omega} \xi \cdot w,$$

therefore the best choice of w is the projection of $-\xi$ to the space of divergence-free vector field. w satisfies the Neumann boundary condition $w \cdot \mathbf{n} = 0$, so that no mass is lost outside Ω . By Hodge-Helmholtz decomposition, ξ can be decomposed into a gradient field and a divergence-free field $\xi = \nabla u + w$,

$$\nabla \cdot \xi = \nabla \cdot w + \Delta u = \Delta u.$$

hence we obtain

$$w = P[\xi] = \xi - \nabla(\Delta^{-1}(\nabla \cdot \xi)).$$

Let $T_t = T \circ (g_t)^{-1}$, $g_t(x) = \gamma_x(t)$, therefore $T_t(\gamma_x(t)) = T(x)$,
obtain

$$0 = \frac{d}{dt} T(x) = \frac{d}{dt} T_t(\gamma_x(t)) = \partial_t T_t + \gamma'_x(t) \cdot \nabla T_t(\gamma_x(t)).$$

We obtain the differential equation of T_t ,

$$\partial_t T_t + v_t \cdot \nabla T_t = 0.$$

We calculate the time derivative of the transportation cost:

$$\begin{aligned} \frac{d}{dt} M(T_t) &= \frac{d}{dt} \int c(x, T_t) \rho = \int \nabla_y c(x, T_t) \cdot (\partial_t T_t) \rho \\ &= - \int \nabla_y c(x, T_t) \cdot (\nabla T_t \cdot v_t) \rho \\ &= - \int \nabla_y c(x, T_t) \cdot \nabla T_t \cdot w_t \end{aligned}$$

By Stokes theorem, and $\nabla \cdot w_t = 0$, we have:

$\nabla \cdot (cw) = \nabla c \cdot w + c \nabla \cdot w = \nabla c \cdot w$, By Neumann boundary condition, we have

$$\int_{\Omega} \nabla c \cdot w = \int_{\Omega} \nabla \cdot (cw) = \int_{\partial\Omega} cw \cdot \mathbf{n} = 0.$$

Continue using

$$\nabla c(x, T_t) = \nabla_x c(x, T_t) + \nabla_y c(x, T_t) \cdot \nabla T_t,$$

we get

$$\begin{aligned} 0 &= \int \nabla c \cdot w = \int \nabla_x c \cdot w + \nabla_y c(x, T_t) \cdot \nabla T_t \cdot w_t \\ &\int \nabla_x c \cdot w = - \int \nabla_y c(x, T_t) \cdot \nabla T_t \cdot w_t \end{aligned}$$

Then we obtain:

$$\frac{d}{dt}M(T_t) = - \int \nabla_y c(x, T_t) \cdot \nabla T_t \cdot w_t = \int \nabla_x c(x, T_t) \cdot w_t = \int \xi_t \cdot w_t$$

We choose $w_t = -P[\xi_t]$, then we have

$$\frac{d}{dt}M(T_t) = - \int \xi_t \cdot P[\xi_t] = -\langle \xi_t, P[\xi_t] \rangle_{L^2} = -\|P[\xi_t]\|_{L^2}^2 \leq 0.$$

flow T_t makes $M(T_t)$ to reduce monotonously with time, reach the steady state when $P[\xi_t] = 0$, then $\xi_t = \nabla_x c(x, T_t)$ is a gradient field.

Angenent-Haker-Tannenbaum algorithm can be formulated as

$$\left\{ \begin{array}{l} \text{time independent density } \rho(x), \text{ initial map } T_0 \text{ given .} \\ -\nabla_x c(x, T_t) = \xi_t \\ \xi_t = \rho v_t + \nabla u_t, P[\xi_t] = \rho v_t \\ \nabla \cdot (\rho v_t) = 0, v_t \cdot \mathbf{n} = 0 \\ \partial_t T_t + v_t \cdot \nabla T_t = 0 \end{array} \right.$$

Geometric Variational Algorithm

定理 (Minkowski)

Given A_1, A_2, \dots, A_k and $\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_k$, such that $\sum_{i=1}^k A_i \mathbf{n}_i = \mathbf{0}$. There exists a convex polyhedron P , unique up to a translation, the area of the i -th face F_i is A_i , the normal to F_i is \mathbf{n}_i .

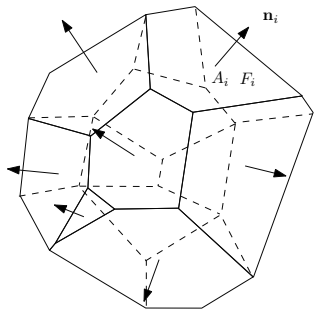


图: Minkowski problem.

定理 (Alexandrov 1950)

Ω is a compact convex domain in \mathbb{R}^n , p_1, \dots, p_k are distinct vectors in \mathbb{R}^n , $A_1, \dots, A_k > 0$, satisfying $\sum A_i = \text{vol}(\Omega)$, then there is a convex piecewise-linear function, unique up to a constant,

$$u(x) = \max_{i=1}^k \{ \langle p_i, x \rangle - h_i \},$$

satisfying

$$\text{vol}(W_i) = A_i, \quad W_i = \{x \mid \nabla u(x) = p_i\}.$$

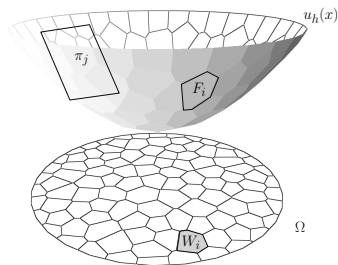


图: Alexandrov Theorem.

Alexandrov theorem is equivalent to the semi-discrete Optimal Transportation map. Let $\Omega \subset \mathbb{R}^d$ be a compact convex set, the measure μ is absolutely continuous, the target measure is the summation of Dirac measures,

$$\nu = \sum_{i=1}^n \nu_i \delta(y - y_i),$$

satisfying the condition $\mu(\Omega) = \sum_{i=1}^n \nu_i$, the transportation cost is the square of Euclidean distance $c(x, y) = \frac{1}{2}|x - y|^2$, then there is a unique Optimal Transportation map $T: (\Omega, \mu) \rightarrow (\{y_i\}_{i=1}^n, \nu)$, $T = \nabla u$, where the Briener potential function $u: \Omega \rightarrow \mathbb{R}$ is a PL convex function, u unique up to a constant:

$$u(x) = \max_{i=1}^n \{\langle x, y_i \rangle - h_i\},$$

the graph of u is the upper envelope of the supporting planes

The graph of the Brenier potential induces a cell decomposition of \mathbb{R}^d

$$\Omega = \bigcup_{i=1}^n W_i(u), \quad W_i(u) = \{x \in \Omega \mid \nabla u = y_i\}.$$

The Optimal Transportation map transforms each cell $W_i(u)$ to a target point y_i , $T: W_i(u) \mapsto y_i$. The Legendre dual u^* is the convex hull of the dual points $\{(y_i, h_i)\}_{i=1}^n$, each point (y_i, h_i) is dual to a supporting plane $\langle x, y_i \rangle - h_i$.

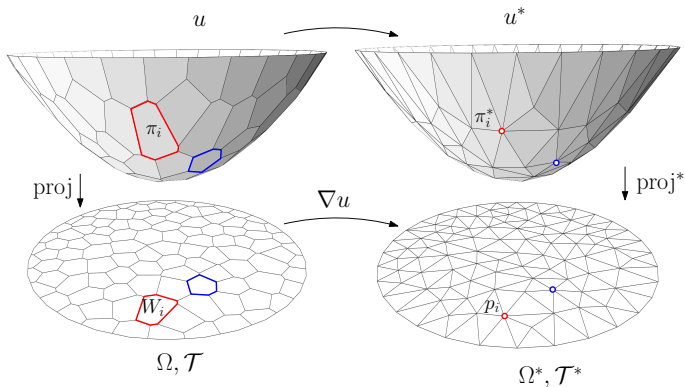


图: Semi-discrete OT map (from left to right): maps W_i to p_i .
 Discrete Monge-Ampère equation (from right to left): $\mu_\sigma(W_i)$ is the discrete Hessian determinant of p_i .

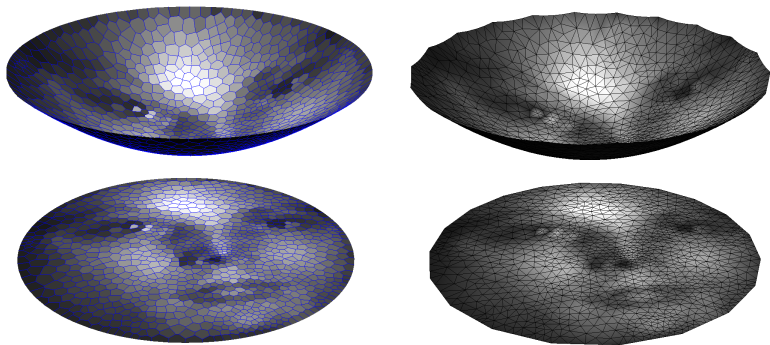


图: Semi-discrete Optimal Transportation Map

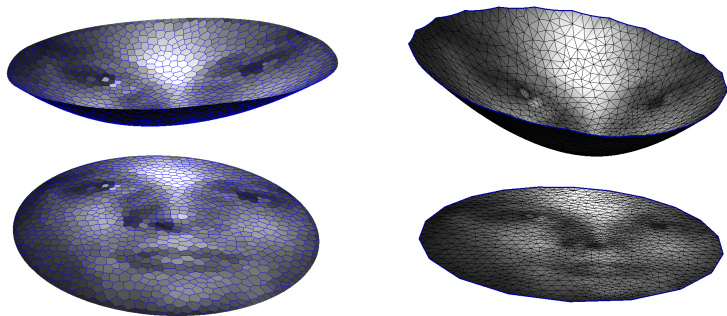


图: Semi-discrete Optimal Transportation Map

Input: Discrete point set $P = \{p_1, p_2, \dots, p_k\}$, target measure $\{\nu_1, \nu_2, \dots, \nu_k\}$; planar convex domain Ω , satisfying $\sum \nu_i = \text{Area}(\Omega)$;

Output: Optimal Transportation map $T: \Omega \rightarrow P$;

1. Translate, scale P , such that $P \subset \Omega$;
2. Initialize the height vector

$$\mathbf{h}^0 \leftarrow \frac{1}{2}(|p_1|^2, |p_2|^2, \dots, |p_k|^2)^T;$$

3. Construct the supporting planes $\{\pi_i(\mathbf{h}^n)\}_{i=1}^k$

$$\pi_i(\mathbf{h}^n, x) = \langle p_i, x \rangle - h_i, \quad i = 1, 2, \dots, k.$$

- Construct the dual points of the supporting planes $\{\pi_i^*(\mathbf{h}^n)\}_{i=1}^k$,

$$\pi_i^*(\mathbf{h}^n) = (p_i, h_i), \quad i = 1, 2, \dots, k.$$

- Compute the convex hull of the dual points $\text{Conv}(\{\pi_i^*(\mathbf{h}^n)\}_{i=1}^k)$, to get the Legendre dual of the potential $u^*(\mathbf{h}^n)$;
- Compute the dual of the convex hull, get the upper envelope of the supporting planes $\text{Env}(\{\pi_i(\mathbf{h}^n)\}_{i=1}^k)$, get the Brenier potential $u(\mathbf{h}^n)$,

$$u(\mathbf{h}^n, x) = \max_{i=1}^k \pi_i(\mathbf{h}^n, x) = \max_{i=1}^k \{ \langle p_i, x \rangle - h_i \}$$

7. Project the Legendre dual of the potential to get a weighted Delaunay triangulation of P , $\mathcal{T}(\mathbf{h}^n)$
8. Project the Brenier potential to get the Power diagram of Ω , $\mathcal{D}(\mathbf{h}^n)$, compute the intersection between each cell and Ω ,

$$\Omega = \bigcup_{i=1}^k W_i(\mathbf{h}^n) = \bigcup_{i=1}^k \{x \in \mathbb{R}^2 \mid \nabla u(\mathbf{h}^n, x) = p_i\} \cap \Omega.$$

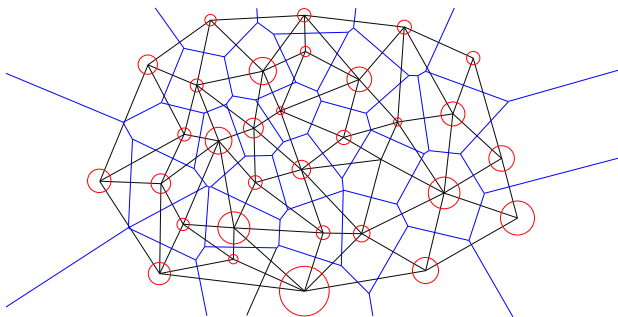
9. Compute the area of each cell, $w_i(\mathbf{h}^n)$, $i = 1, 2, \dots, k$,

10. Compute the gradient of the energy $E(\mathbf{h})$:

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

$$\nabla E(\mathbf{h}^n) = w_i(\mathbf{h}^n) - \nu_i.$$

11. If $|\nabla E(\mathbf{h}^n)|$ is less than a threshold ε , return the map $T = \nabla u(\mathbf{h}^n)$, $W_i(\mathbf{h}^n) \mapsto p_i$, $i = 1, 2, \dots, k$.



12. Compute the Hessian matrix of the energy $E(\mathbf{h}^n)$

$$\frac{\partial^2 E(\mathbf{h}^n)}{\partial h_i \partial h_j} = \frac{\partial w_i(\mathbf{h}^n)}{\partial h_j} = -\frac{|e_{ij}|}{|\bar{e}_{ij}|},$$
$$\frac{\partial^2 E(\mathbf{h}^n)}{\partial h_i^2} = \frac{\partial w_i(\mathbf{h}^n)}{\partial h_i} = -\sum_{i \neq j} \frac{\partial w_i(\mathbf{h}^n)}{\partial h_j},$$

13. Solve the linear system

$$\text{Hess}(\mathbf{h}^n)\mathbf{d} = \nabla E(\mathbf{h}^n),$$

with linear constraint

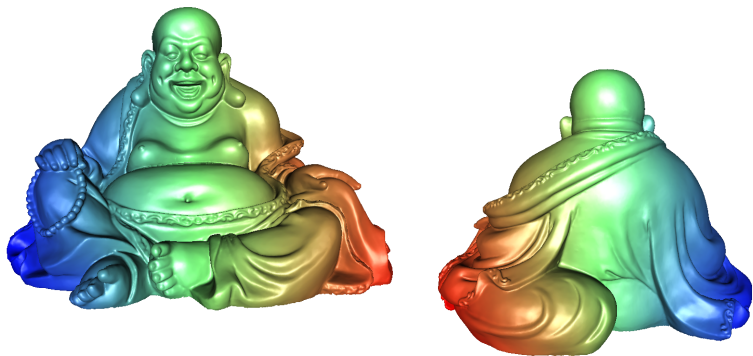
$$d_1 + d_2 + \cdots + d_k = 0.$$

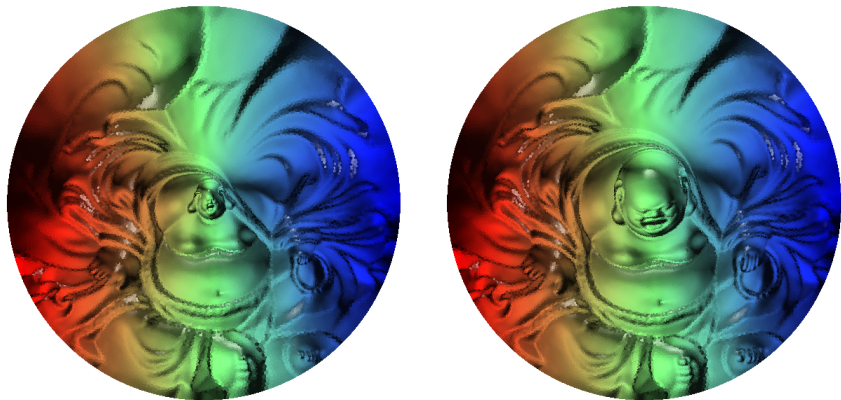
14. Set the initial step length $\lambda \leftarrow 1$,
15. Construct supporting planes $\{\pi_i(\mathbf{h}^n + \lambda\mathbf{d})\}_{i=1}^k$, dual points $\{\pi_i(\mathbf{h}^n + \lambda\mathbf{d})^*\}_{i=1}^k$
16. Construct the convex hull $\text{Conv}(\{\pi_i(\mathbf{h}^n + \lambda\mathbf{d})^*\}_{i=1}^k)$
17. If there is a dual point $\pi_i(\mathbf{h}^n + \lambda\mathbf{d})^*$ which is not on the convex hull, let $\lambda \leftarrow \lambda/2$, repeat steps 15, 16, until all the dual points are on the convex hull;

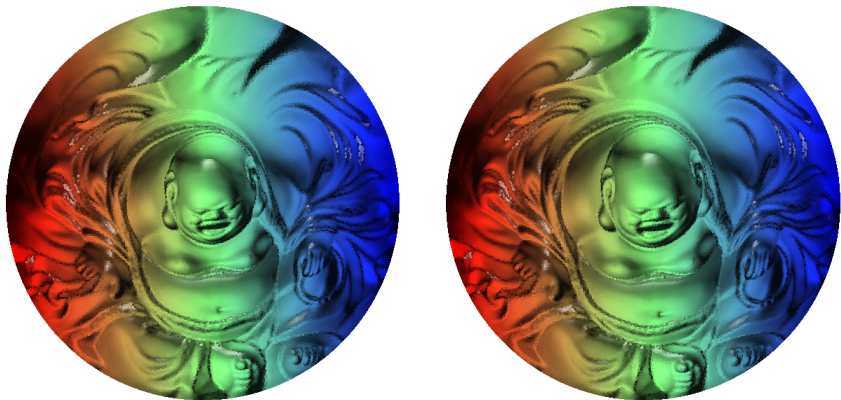
18. Update the height vector

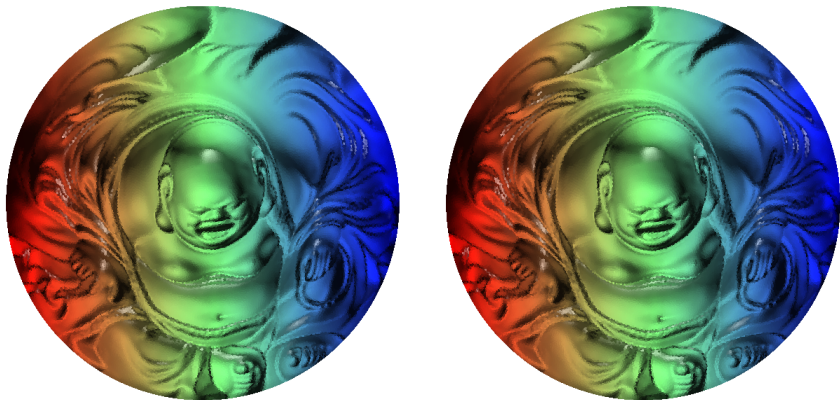
$$\mathbf{h}^{n+1} \leftarrow \mathbf{h}^n + \lambda \mathbf{d};$$

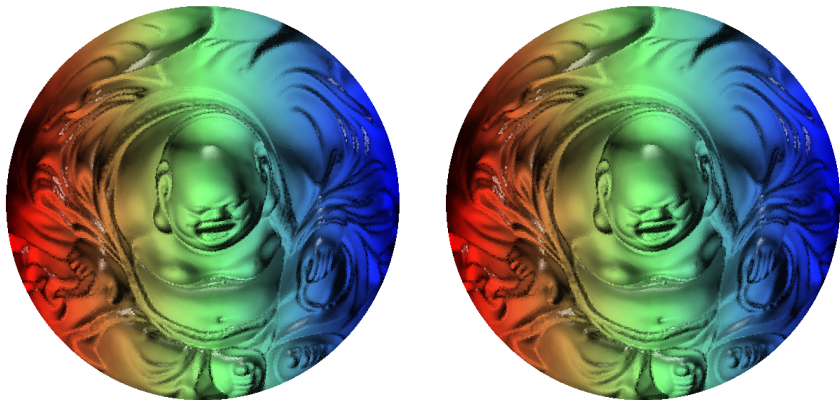
19. Repeat step 3 to step 18.

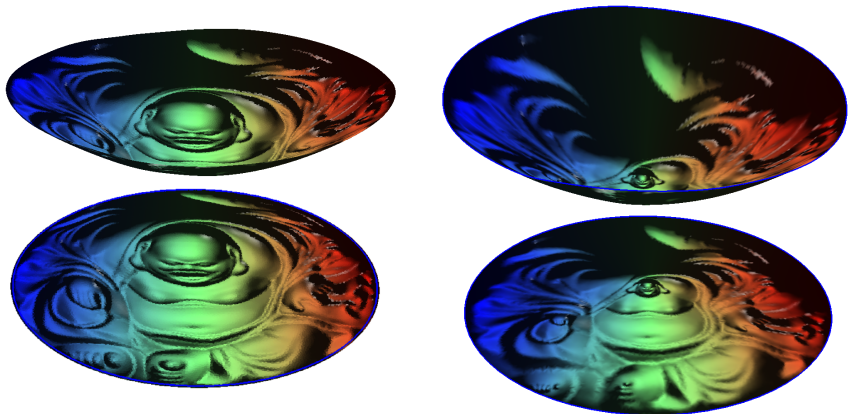












for more information, please contact gu@cs.stonybrook.edu.

Thank You!