Mathematics, Computation and Geometry of Data
Volume 1, Number 2, 207–253, 2021

A geometric variational framework for computing optimal transportation maps I

Dongsheng An, Na Lei, Li Cui, Kehua Su, Xiaoyin Xu,
Feng Luo, Xianfeng Gu*, and Shing-Tung Yau

Optimal transportation (OT) maps play fundamental roles in many engineering and medical fields. The computation of optimal transportation maps can be reduced to solve highly non-linear Monge-Ampère equations. In this work, we summarize the geometric variational framework to solve optimal transportation maps in Euclidean spaces. We generalize the method to solve worst transportation maps and discuss about the symmetry between the optimal and the worst transportation maps. Many algorithms from computational geometry are incorporated into the method to improve the efficiency, the accuracy and the robustness of computing optimal transportation.

1. Introduction

Optimal transportation (OT) map finds the most economical way to transfer one probability measure to another. The transportation cost gives a metric between measures – the so-called Wasserstein metric. OT method has become an important tool in optics [19], economy [17], and life science [39]. Recently, OT has been successfully applied in different areas of computer science, such as parameter estimation in Bayesian nonparametric models [35], computer vision [8, 13, 46], natural language processing [26, 51], medical image registration [25], 3D surface registration [43, 27, 54, 24], color transfer [15, 37, 38] and so on. The Wasserstein metric has also been broadly applied in generative models in deep learning, including the WGAN model [8, 33], WAE [46], AE-OT [4, 5], and energy-based models [3].

The origin of the OT problem can be traced back to 1781, when Monge asked if there existed an OT map between two measures for a given cost function. Depending on the cost function and the measures, an OT map may not exist. In the 1950s, Kantorovich relaxed the OT map to OT plan,

---

*Corresponding author.

and showed the existence and the uniqueness of the plan under mild conditions [48]. In the 1980s, Brenier [12] discovered that, when the density of the source measure is absolutely continuous and the cost function is the squared Euclidean distance, the OT map is given by the gradient of a convex function, the so-called Brenier potential.

Recently, the equivalence between the Brenier potential and Alexandrov's convex polytope has been proved in [21], as both of them can be obtained by solving the non-linear Monge-Ampère equation. This connection leads to a practical algorithm to solve the semi-discrete OT problem using convex geometry. According to the Brenier theorem, the Brenier potential can be represented as the upper envelope of a set of hyperplanes, and its projection induces a power diagram of the source domain, which gives the semi-discrete OT map. The geometric variational method proposed by Su et al. [43] is based on classical computational geometry [10] and solves the semi-discrete OT problem by minimizing a convex energy. Though with high accuracy, the method is very slow and inefficient. At each step, it constructs a new power diagram, or a weighted Delaunay triangulation of the samples. During the optimization, the combinatorial structures of the triangulation are changed dynamically, which makes the algorithm complicate. Moreover, as the method assumes that the source distribution is uniform, it cannot handle complex source measures. To circumvent this limitation, we can generalize the Lawson's edge flip algorithm [28] to update the power diagram during the optimization process, instead of constructing a new convex hull at each iteration like Su et al. did in [43]. This improves the computational efficiency by a factor of more than five. To handle the piecewise linearly defined source measures, we can use the sweep convex polygon algorithm, which is a generalization of the classical Bentley-Ottman's sweep line algorithm [9], to compute the subdivision of two cell decompositions. Thus we can compute the semi-discrete OT map from a piecewise linear source measure to the target measure. This improves the efficiency and the robustness of the algorithm.

Within the same framework of the OT solution, we can extend the OT problem to solve the worst transportation (WT) problem [53], which tries to maximize the cost between two distributions, given a cost function between them. Different from finding the convex Brenier potential for the OT problem, the WT problem finds the concave Brenier potential whose gradient gives the WT map. Moreover, we can unify the OT solution and WT solution into a single framework and solve them by Newton's method. We also discover the intrinsic symmetry between the OT map and the WT map.

The geometric variational framework is generalizable for any dimensions. For example, volumetric OT maps can also be directly computed based on the same principle. However, the volumetric computational algorithms are more complicated than planar ones.

2. Related work

OT maps [47, 48] play fundamental roles in many engineering and medical fields and there is a huge body of research in this area. For a thorough review, we refer readers to [36]. In the following we will concentrate on the semi-discrete OT algorithm and its direct applications.

2.1. Semi-discrete OT

Haber et al. [24] propose to solve the semi-discrete OT problem based on the connection between minimizing flow and OT, but this method is relatively inefficient. Arjovsky et al. [8] propose an approach that specializes to 1-Wasserstein distance, where the Lipschitz constraint is replaced by the weight clipping at each iteration. This limits the approximation accuracy of Wasserstein distance. Genevay et al. [18] extend the entropic regularized OT to semi-discrete cases through stochastic gradient descent, which improves the speed but sacrifices the accuracy. Seguy et al. [40] propose a regularization relaxation of the OT problem and approximate the Alexandrov potential with a deep neural network (DNN). However, their approximation cannot guarantee the convexity of the energy, and the algorithm may stuck at a local optimum. An et al. [6] use the Pogrelov map to approximate the OT map by solving the Monge-Ampère equation with a Dirichlet boundary condition. Some other methods [34, 32] choose to approximate the Monge-Ampère equation and solve the OT problem by discretizing the Hessian matrix. One major problem of this kind of methods is that they assume one measure to be uniformly defined on a rectangular space. As a result, they cannot handle complex source measures with arbitrary convex support or complex density functions. Their accuracy is also lower than those of the geometric variational methods. Gu et al. [21] build the theoretical connection between the OT problem and convex geometry. Similar OT maps have been introduced in [42, 43] for surfaces and in [30, 31] for volumes, but these algorithms compute the power diagrams from scratch at each iteration, which is highly inefficient. *Our proposed algorithm updates the existing power diagrams by a few edge flips, which greatly improve the computational speed.*

2.2. Direct applications

Semi-discrete OT maps have been applied in various areas of computer vision and medical imaging. In [25], OT maps are used to register grayscale medical images with high accuracy. In [42], human cortical surfaces are matched using spherical OT maps. In [43], surfaces with large non-rigid deformations are reliably registered, where surfaces are first conformally mapped to the planar domains with large area distortions and then the mapping is composed with an OT map to balance the area distortion and this greatly improves the matching accuracy. The method has been generalized to registering surfaces with complicated topologies, such as human skull in [54] and multiply-connected surface registration in [44]. Besides, the semi-discrete OT can be used to generate adaptive mesh in computational geometry [7]. On the other hand, the WT problem [53] can help to analyze difference in cortical shape caused by the dementia due to Alzheimer's disease and the normal aging process. All of these applications require computation of low dimensional but accurate OT maps, and most of the approximate methods [14, 18, 40, 4] are not suitable since they approximate the OT maps by transportation plans with relatively large approximation errors.

3. Theoretic foundation

This section introduces the theoretic foundation of our geometric variational framework for OT maps. For more details, we refer readers to [47, 48, 29].

3.1. Monge-Kantorovich theory

Suppose Ω, Ω^* are complete, separable metric spaces, such as domains in the Euclidean space $\mathbb{R}^n$, with probability measures μ and ν, respectively, with corresponding density functions $d\mu = f(x)dx$ and $d\nu = g(y)dy$, satisfying the equal total mass condition:

$$\mu(\Omega) = \nu(\Omega^*).$$

A transportation map $T : \Omega \to \Omega^*$ is *measure preserving* if, for any Borel set $B \subset \Omega^*$,

$$\int_{T^{-1}(B)} d\mu(x) = \int_B d\nu(y),$$

denoted as $T_{\#}\mu = \nu$. Monge raised the following problem [48]:

Problem 3.1 (Monge)**.** Given the cost function $c : \Omega \times \Omega^* \to \mathbb{R}^+$, find the transportation map $T : \Omega \to \Omega^*$ that minimizes the *total transportation cost*

$$(MP) \quad \min \left\{ \int_\Omega c(x, T(x)) : T : \Omega \to \Omega^*, T_\# \mu = \nu \right\}.$$

The minimizer is call the *optimal transportation map* and the corresponding transportation cost is called the *optimal transportation cost.* Kantorovich generalized the transportation maps to transportation schemes, namely the joint distribution γ defined on $\Omega \times \Omega^*$, $\gamma \in \mathcal{P}(\Omega \times \Omega^*)$, where $\mathcal{P}(\Omega \times \Omega^*)$ represents the space of all probability measures on $\Omega \times \Omega^*$, its marginal distributions equals to μ and ν, respectively,

$$(\pi_x)_\# \gamma = \mu, \quad (\pi_y)_\# \gamma = \nu,$$

where $\pi_x(x,y) = x$ and $\pi_y(x,y) = y$ are projections.

Problem 3.2 (Kantorovich)**.** Given two probability measures $\mu \in \mathcal{P}(\Omega)$, $\nu \in \mathcal{P}(\Omega^*)$ and the cost function $c : \Omega \times \Omega^* \to [0, +\infty]$, find

$$(KP) \quad \inf \left\{ K(\gamma) := \int_{\Omega \times \Omega^*} c(x,y) d\gamma(x,y) : \gamma \in \Pi(\mu, \nu) \right\},$$

where the joint probability measure γ is in the transportation scheme space

$$\Pi(\mu, \nu) := \{\gamma \in \mathcal{P}(\Omega \times \Omega^*) : (\pi_x)_\# \gamma = \mu, (\pi_y)_\# \gamma = \nu\}.$$

By using the generalized Lagrange multiplier method, we can convert Kantorovich problem to the dual problem:

Problem 3.3 (Dual)**.** Given $\mu \in \mathcal{P}(\Omega)$, $\nu \in \mathcal{P}(\Omega^*)$ and the cost funtion $c : \Omega \times \Omega^* \to [0, +\infty)$, consider the following problem

$$(DP) \quad \max \left\{ \int_\Omega \varphi d\mu + \int_{\Omega^*} \psi d\nu : \varphi \in C_b(\Omega), \psi \in C_b(\Omega^*), \varphi \oplus \psi \le c \right\},$$

where $\varphi \oplus \psi := \varphi(x) + \psi(y)$.

Definition 3.4 (c-transform)**.** Given a function $\varphi : \Omega \to \bar{\mathbb{R}}$, its c-transform (c-conjugate) $\varphi^c : \Omega^* \to \bar{\mathbb{R}}$ is defined as

$$\varphi^c(y) := \inf_{x \in \Omega} c(x,y) - \varphi(x), \tag{3.1}$$

where $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$.

By using the concept of c-transform, the dual problem can be reformulated as

Problem 3.5 (Dual-2)**.** Given $\mu \in \mathcal{P}(\Omega)$, $\nu \in \mathcal{P}(\Omega^*)$ and the cost function $c : \Omega \times \Omega^* \to [0, +\infty)$, consider the following problem

$$(DP) \quad \max\left\{ \int_{\Omega} \varphi d\mu + \int_{\Omega^*} \varphi^c d\nu : \varphi \in C_b(\Omega) \right\}. \tag{3.2}$$

3.2. Brenier theory

When the cost function is the square of Euclidean distance, the OT map is the gradient of a convex function, claimed by the below:

Theorem 3.6 (Brenier [11])**.** *Given the probability measures μ and ν on compact supports $\Omega, \Omega^* \subset \mathbb{R}^d$ with density functions $f, g \in L^1(\mathbb{R}^d)$ and total equal mass $\mu(\Omega) = \nu(\Omega^*)$, if the cost function is $c(x,y) = \frac{1}{2}|x-y|^2$, then the OT map exists and is unique. It is the gradient of a convex function $u : \Omega \to \mathbb{R}$, where u is called the Brenier potential.*

Suppose the Brenier potential u is C^2, by the measure preserving condition, it satisfies the Monge-Ampère equation,

$$\det D^2 u(x) = \frac{f(x)}{g \circ \nabla u(x)}, \tag{3.3}$$

and the unique OT map is given by $T = \nabla u$.

In the following, we show that the OT map with the L^2 cost is equivalent to the WT map with the inner product cost function, which greatly simplifies the computation.

Lemma 3.7. *The OT map with the cost $c(x,y) = |x-y|^2/2$ is equivalent to the WT map with the cost $c(x,y) = \langle x, y \rangle$.*

Proof. Assume $T_{\#}\mu = \nu$, then $f(x)dx = g(y)dy$ for $y = T(x)$. By direct computation:

$$\begin{aligned}
&\frac{1}{2}\int_{\Omega} |x - T(x)|^2 f(x)dx \\
=&\frac{1}{2}\int_{\Omega} |x|^2 f(x)dx - \int_{\Omega} \langle x, T(x)\rangle f(x)dx + \frac{1}{2}\int_{\Omega} |T(x)|^2 f(x)dx \\
=&\frac{1}{2}\int_{\Omega} |x|^2 f(x)dx - \int_{\Omega} \langle x, T(x)\rangle f(x)dx + \frac{1}{2}\int_{\Omega} |y|^2 g(y)dy
\end{aligned}$$

The first and the third terms are independent of T, therefore

$$\min_{T_{\#}\mu=\nu}\int_{\Omega}|x-T(x)|^2 f(x)dx \Longleftrightarrow \max_{T_{\#}\mu=\nu}\int_{\Omega}\langle x, T(x)\rangle f(x)dx.$$

$\square$

In this situation, the c-transform becomes the conventional Legendre dual.

Definition 3.8 (Legendre Dual)**.** The Legendre dual of a function $\varphi : \Omega \to \mathbb{R}$ is

$$\varphi^*(y) := \sup_{x\in\Omega}\langle x, y\rangle - \varphi(x), \tag{3.4}$$

If the Brenier u is not smooth, we can still define the Alexandrov solution to the Monge-Ampère Eqn. (3.3).

Definition 3.9 (sub-gradient)**.** The sub-gradient of a function u at x is defined as

$$\partial u(x) := \left\{p \in \mathbb{R}^d : u(z) \geq \langle p, z-x\rangle + u(x), \forall z \in \Omega\right\} \tag{3.5}$$

which induces a set-valued map: $\partial u : \Omega \to \Omega^*, x \mapsto \partial u(x)$.

We can use the sub-gradient to replace gradient map in the Monge-Ampère Eqn. (3.3), and define

Definition 3.10 (Alexandrov Solution [2])**.** Suppose a convex function $u : \Omega \to \mathbb{R}$ satisfies the equation

$$(\partial u)_{\#}\mu = \nu, \tag{3.6}$$

or equivalently $\mu((\partial u)^{-1}(B)) = \nu(B)$, $\forall \mathit{Borel}\ B \subset \Omega^*$, then u is called an Alexandrov solution to the Monge-Ampère Eqn. (3.3).

3.3. Theory for worst transportation

With the same setup, the WT problem can be formulated as follows:

Problem 3.11 (Worst Transportation Map)**.** Given a transportation cost function, $c : \Omega \times \Omega^* \to \mathbb{R}^+$, find a transportation map $T : \Omega \to \Omega^*$ that maximizes the total transportation cost,

$$(WP) \quad \max\left\{\int_{\Omega} c(x, T(x)) : T : \Omega \to \Omega^*, T_{\#}\mu = \nu\right\}.$$

The maximizer is called the *worst transportation map* (WT map). The transportation cost of the WT map is called the *worst Wasserstein distance* between measures.

We can generalize the Brenier theorem to the worst transportation map.

Theorem 3.12 (Worst Transportation Map)**.** *Given the probability measures* μ *and* ν *with compact supports* $\Omega, \Omega^* \subset \mathbb{R}^{2d}$*, respectively, and equal total mass* $\mu(\Omega) = \nu(\Omega^*)$*. Assume the density functions* $f, g \in L^1(\mathbb{R}^{2d})$ *and the cost function* $c(x,y) = \frac{1}{2}|x-y|^2$*, then the worst transportation map exists and is unique. It is the gradient of a concave function* $u : \Omega \to \mathbb{R}$*,* $T = \nabla u$*, where* u *is the worst Brenier potential function, unique up to a constant. Furthermore, if* u *is* C^2*, then it satisfies the Monge-Ampère Eqn. (3.3).*

Proof. Suppose $T : \Omega \to \Omega^*$ is a measure-preserving map, $T_\#\mu = \nu$. Consider the total transportation cost,

$$\begin{aligned}\int_\Omega |x - T(x)|^2 d\mu &= \int_\Omega |x|^2 d\mu + \int_\Omega |T(x)|^2 d\mu - 2\int_\Omega \langle x, T(x)\rangle d\mu \\ &= \int_\Omega |x|^2 d\mu + \int_{\Omega^*} |y|^2 d\nu - 2\int_\Omega \langle x, y\rangle d\mu, \quad T(x) \text{ replaced by } y.\end{aligned}$$

Therefore maximizing the transportation cost is equivalent to

$$\min_{T_\#\mu=\nu} \int_\Omega \langle x, T(x)\rangle d\mu.$$

Use Kantorovich formula, this is equivalent to finding a transportation plan $\gamma : \Omega \times \Omega^* \to \mathbb{R}$,

$$\min\left\{\int_{\Omega\times\Omega^*} \langle x, y\rangle d\gamma, (\pi_x)_\#\gamma = \mu, (\pi_y)_\#\gamma = \nu\right\},$$

where π_x, π_y are the projections from $\Omega \times \Omega^*$ to Ω and Ω^*, respectively. By duality, this is equivalent to

$$\max\{J(u,v), (u,v) \in K\}, \quad J(u,v) := \int_\Omega u(x)f(x)dx + \int_{\Omega^*} v(y)g(y)dy,$$

and the function space

$$K := \{(u,v) : u(x) + v(y) \le \langle x, y\rangle\}.$$

Now we define c-transform,

$$u^c(y) := \inf_{x\in\bar{\Omega}} \langle x, y\rangle - u(x). \tag{3.7}$$

Fixing x, $\langle x,y\rangle - u(x)$ is a linear function, hence $u^c(y)$ is the lower envelope of linear functions and it is a concave function, and Lipschitz (since the gradients $x \in \bar{\Omega}$). We construct a sequence of function pairs $\{(u_k,v_k)\}$, where $u_k = v_{k-1}^c$, $v_k = u_k^c$. Then $J(u_k,v_k)$ increases monotonically, Lipschitz function pairs (u_k,v_k) converge to the limit (u,v), which is the maximizer of J. Since u and v are c-transforms of each other,

$$u(x) + v(T(x)) = \langle x, T(x)\rangle. \tag{3.8}$$

This shows the existence of the solution.

From the definition of c-transform Eqn. (3.7), we obtain

$$v(y) = \inf_{x\in\bar{\Omega}} \langle x,y\rangle - u(x).$$

where $u(x)$ is concave, therefore almost everywhere differentiable, hence $\nabla_x\langle x,y\rangle - \nabla u(x) = 0$, implying $y = T(x) = \nabla u(x)$. This shows the WT map is the gradient of the worst Brenier potential u.

Suppose there are two maximizers $(\varphi,\psi) \in K$ and $(u,v) \in K$, because $J(u,v)$ is linear, then $\frac{1}{2}(\varphi+u,\psi+v) \in K$ is also a maximizer. Assume

$$\begin{aligned} \varphi(x_0)+\psi(y_0) = \langle x_0,y_0\rangle, \varphi(x_0)+\psi(y) < \langle x_0,y\rangle, \forall y \neq y_0 \\ u(x_0)+v(z_0) = \langle x_0,z_0\rangle, u(x_0)+v(z) < \langle x_0,z\rangle, \forall z \neq z_0. \end{aligned}$$

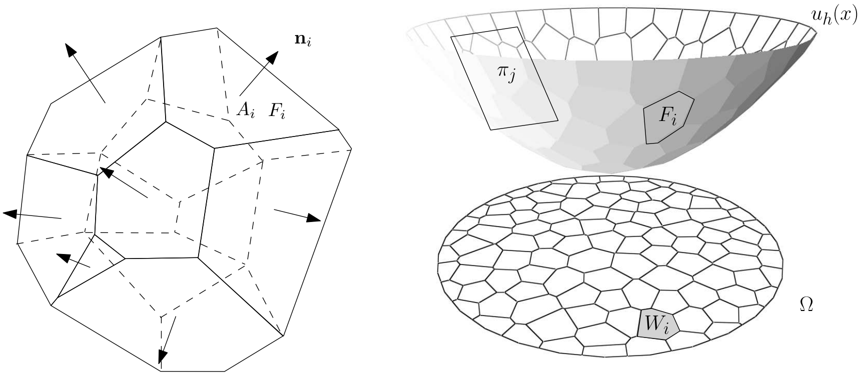
If $y_0 \neq z_0$, then $\forall y$, $1/2(\varphi+u)(x_0) + 1/2(\psi+v)(y) < \langle x_0,y\rangle$. But $(\frac{1}{2}(\varphi+u), \frac{1}{2}(\psi+v))$ is also a maximizer, this contradicts Eqn. (3.8). This shows the uniqueness of the WT map.

Since u is concave, by Alexandrov's theorem [1], it is almost everywhere C^2. The WT map $T = \nabla u$ is measure-preserving, $T_\# \mu = \nu$, thus

$$\det(DT)(x) = \frac{f(x)}{g\circ T(x)} \implies \det(D^2u)(x) = \frac{f(x)}{g\circ \nabla u(x)}$$

This completes the proof. □

Note that, this theorem cannot be proved by simply negating the cost function $c(x,y)$ and using the classical Brenier's theorem, it is closely related to the duality between upper and lower envelopes. This is more evident by the spherical OT theory [49, 50], which focuses on the duality between the inner and outer envelopes.



(a). Minkowski theorem (b). Alexandrov theorem

Figure 1: Minkowski theorem and Alexandrov theorem.

3.4. Convex geometry

The OT theory has intrinsic relation with the convex geometry, especially Minkowski-Alexandrov theory [2].

As shown in the left frame of Fig. 1, Minkowski asked about how to recover a convex polyhedron from its face normals and face areas:

Problem 3.13 (Minkowski)**.** Given k unit vectors in $\mathbb{R}^n$, $\mathbf{n}_1, \mathbf{n}_2, \dots, \mathbf{n}_k$, not contained in any half-space of $\mathbb{R}^n$, and positive numbers $A_1, A_2, \dots, A_k$, satisfying the condition

$$\sum_{i=1}^{k} A_i \mathbf{n}_i = \mathbf{0}. \tag{3.9}$$

Find a compact convex polyhedron P, with k codimension 1 faces, $F_1, F_2, \dots, F_k$, such that the volume of F_i is A_i and the normal to F_i is $\mathbf{n}_1$.

Minkowski proved the existence and the uniqueness of the solution. The existence is based on a variational method. The uniqueness is based on Brunn-Minkowski inequality.

Theorem 3.14 (Minkowski)**.** *The solution to the Minkowski problem exists, and is unique up to a translation.*

Alexandrov generalized Minkowski's theorem from compact convex polyhedron to open convex polyhedron, as shown in the right frame of Fig. 1.

Theorem 3.15 (Alexandrov 1950)**.** *Suppose* $\Omega \subset \mathbb{R}^n$ *is a compact and convex domain in* $\mathbb{R}^n$*,* $p_1, \dots, p_k$ *are distinct vectors in* $\mathbb{R}^n$*,* $A_1, A_2, \dots, A_k > 0$ *are positive numbers, satisfying the condition*

$$\sum_{i=1}^{k} A_i = vol(\Omega),$$

then there exists a convex polyhedron P *in* $\mathbb{R}^{n+1}$*, which has* k *codimension* 1 *faces,* $F_1, F_2, \dots, F_k$*, such that the volume of the intersection between the projection of* F_i *and* Ω *equals to* A_i *and the gradient of the supporting plane of* F_i *is* p_i*. Furthermore, such kind of polyhedron* P *is unique up to a vertical translation.*

Equivalently, P can be treated as the graph of a piecewise-linear function

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

satisfying $\text{vol}(w_i \cap \Omega) = A_i$, where the cell W_i equals to

$$w_i := \{x \in \mathbb{R}^n | \nabla u(x) = p_i\}.$$

In fact, Alexandrov theorem is equivalent to Brenier theorem, both of them satisfy the Monge-Ampère equation.

Alexandrov proved the theorem using an algebraic topology method, which is not constructive and difficult to translate to algorithms directly.

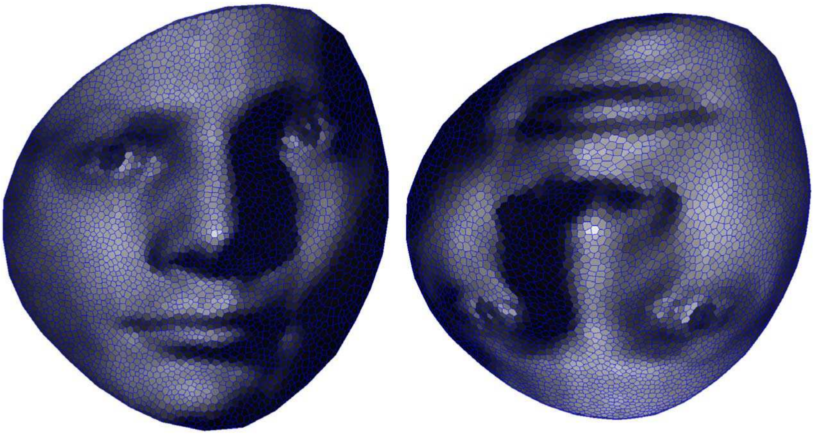
3.5. Semi-discrete OT maps

The Brenier theorem 3.6 can be directly generalized to the discrete situation. In discrete cases, it is convenient to use the concepts and methods from computational geometry.

Basic Concepts from Computational Geometry A hyperplane in $\mathbb{R}^{d+1}$ is represented as $\pi(x) := \langle p, x\rangle - h$. Given a family of hyperplanes $\{\pi_i(x) = \langle p_i, x\rangle - h_i\}_{i=1}^n$, their *upper envelope* of $\{\pi_i\}_{i=1}^n$ is the graph of the function $u(x) := \max_{i=1}^n \{\langle p_i, x\rangle - h_i\}$; The *lower envelope* is the graph of the function $u(x) := \min_{i=1}^n \{\langle p_i, x\rangle - h_i\}$. The Legendre dual of u is defined as $u^*(y) := \max_{x\in\mathbb{R}^d} \langle x, y\rangle - u(x)$. The c-transform of u is defined as

$$u^c(y) := \min_{x\in\mathbb{R}^d} \langle x, y\rangle - u(x). \tag{3.10}$$

Each hyperplane $\pi_i(x)$ has a dual point in $\mathbb{R}^{2d+1}$, $\pi_i^* := (p_i, h_i)$, the graph of u^* is the *lower convex hull* of the dual points $\{\pi_i^*\}_{i=1}^n$, the graph of u^c



(a). OT map image (b). WT map image

Figure 2: The image of the OT map (left frame), and the WT map (right frame).

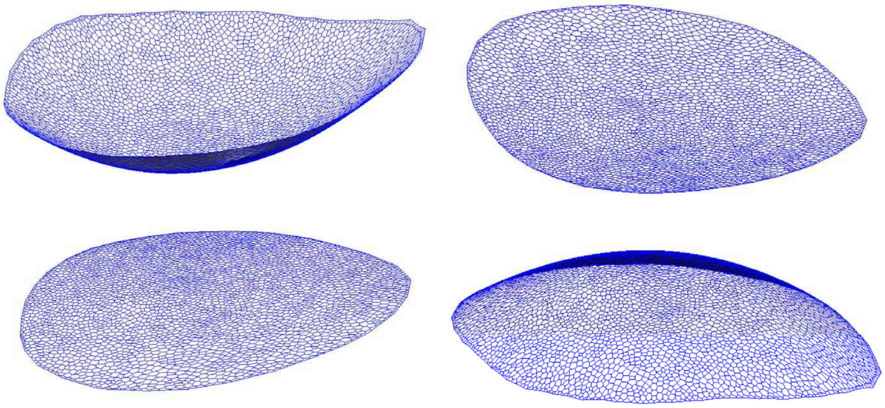
is the *upper convex hull* of the dual points $\{p_i^*\}_{i=1}^n$. The projection of the upper envelope induces a *nearest power diagram* $\mathcal{D}(\Omega)$ of Ω,

$$\Omega = \bigcup_{i=1}^n W_i(u),\ W_i(u) := \{x \in \Omega | \nabla u(x) = p_i\}.$$

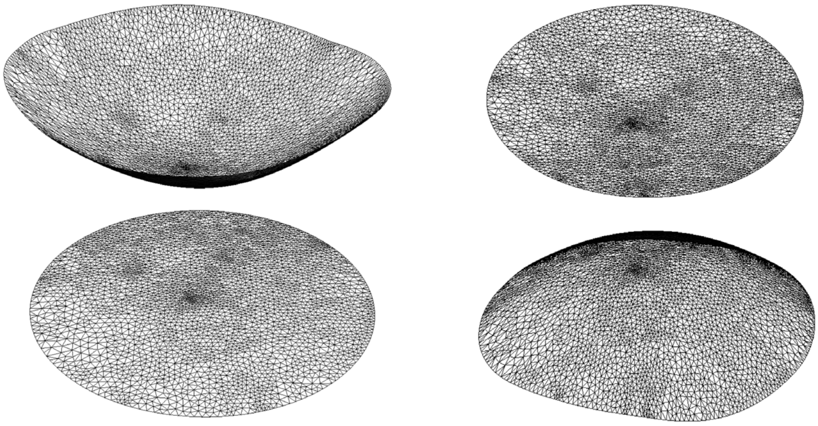
The projection of the lower envelope induces a *farthest power diagram* $\mathcal{D}^c$ of Ω. The projection of the lower convex hull u^* induces a *nearest weighted Delaunay triangulation* $\mathcal{T}(\Omega^*)$ of Ω^*. The projection of the upper convex hull u^c induces a *farthest weighted Delaunay triangulation* $\mathcal{T}^c(\Omega^*)$. $\mathcal{D}(\Omega)$ and $\mathcal{T}(\Omega^*)$ are dual to each other, namely p_i connects p_j in $\mathcal{T}(\Omega^*)$ if and only if $W_i(u)$ is adjacent to $W_j(u)$. Similarly, $\mathcal{D}^c$ and $\mathcal{T}^c$ are also dual to each other. Fig. 2 and Fig. 3 show these basic concepts.

Problem 3.16 (Semi-discrete Optimal Transportation Map)**.** Suppose $\Omega \in \mathbb{R}^d$ is a compact convex domain, the source density function is $d\mu(x) = f(x)dx$, $f \in L^1(\Omega)$. The target domain $\Omega^* = \{p_1, p_2, \dots, p_n\} \subset \mathbb{R}^d$ is a set of n distinct points. The target measure is a summation of Dirac measures $\nu = \sum_{i=1}^n \nu_i \delta(y - p_i)$, where $\nu_1, \dots, \nu_n$ are positive numbers and satisfy

$$\mu(\Omega) = \sum_{i=1}^n \nu_i.$$



(a). OT upper envelope (b). WT lower envelope



(c). OT lower convex hull (d). WT upper convex hull

Figure 3: The Brenier potential for OT map (a) and WT map (b), equivalently the upper and lower envelopes. The Legendre dual to the potential of the OT map (c) and the WT map (d), equivalently the lower and upper convex hulls.

Find the transportation map $T : \Omega \to \Omega^*$ that minimizes the total transportation cost

(SD)

$$\min\left\{\frac{1}{2}\int_\Omega |x - T(x)|^2 d\mu(x) : T : \Omega \to \Omega^*, \mu(T^{-1}(p_i)) = \nu_i, i = 1, \dots, n\right\}.$$

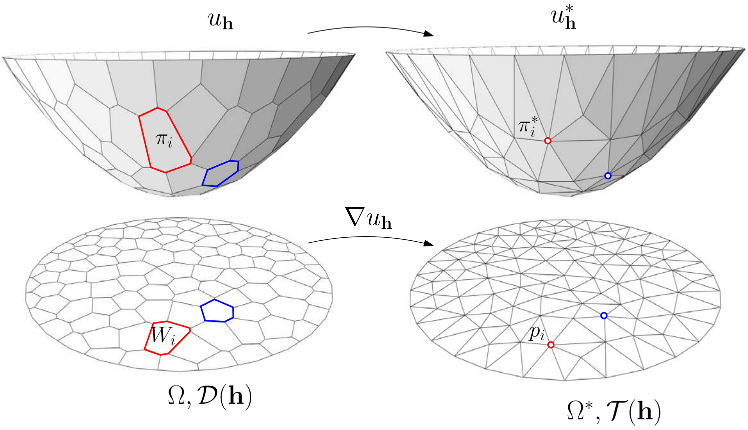


Figure 4: Semi-discrete OT: top left the Brenier potential $u_{\mathbf{h}}$, top right the Legendre dual $u^*_{\mathbf{h}}$, bottom left the Power diagram, bottom right the weighted Delaunay triangulation.

The Brenier theorem can be generalized to the semi-discrete cases as shown in Fig. 4. The Brenier potential becomes a piecewise-linear convex function $u : \mathbb{R}^n \to \mathbb{R}$, each point $p_i \in \Omega^*$ corresponds to a support plane

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

where vector $\mathbf{h}$ represents all the heights $(h_1, h_2, \dots, h_n)$. Each support plane $\pi_i(x)$ has a dual point $\pi^*_i \in \mathbb{R}^{d+1}$

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

The Brenier potential is the *upper envelope* of the support planes $\pi_i(\mathbf{h})$'s as below:

$$u_{\mathbf{h}}(x) := \max_{i=1}^{n}\{\pi_i(\mathbf{h}, x)\} = \max_{i=1}^{n}\{\langle p_i, x\rangle - h_i\}. \tag{3.11}$$

The Legendre dual of the Brenier potential is the *lower convex hull* of the dual points $\pi^*_i(\mathbf{h})$'s

$$u^*_{\mathbf{h}}(h) := \min\left\{\sum_{j=k}^{d+1}\lambda_k h_{i_k} | y = \sum_{k=1}^{d+1}\lambda_k p_{i_k}, \sum_{k=1}^{d+1}\lambda_k = 1, \lambda_k \geq 0, \forall k\right\}. \tag{3.12}$$

The projection of the Brenier potential induces a cell decomposition of $\mathbb{R}^d$, which is called the *power diagram*, denoted as $\mathcal{D}(\mathbf{h})$, where each cell is called a *power cell*,

$$\mathbb{R}^d = \bigcup_{i=1}^{n} W_i(\mathbf{h}), \quad W_i(\mathbf{h}) := \{x \in \mathbb{R}^d : \nabla u_{\mathbf{h}}(x) = p_i\}. \tag{3.13}$$

The power cell can be equivalently defined as

$$W_i(\mathbf{h}) := \{x \in \mathbb{R}^d : \pi_i(\mathbf{h}, x) \geq \pi_j(\mathbf{h}, x), \forall j\}. \tag{3.14}$$

The projection of the Legendre dual of the Brenier potential induces a triangulation of $\mathbb{R}^d$, which is called the *weighted Delaunay triangulation*, denoted as $\mathcal{T}(\mathbf{h})$. The OT map $T : \Omega \to \Omega^*$ is given by the gradient map of the Brenier potential:

$$T = \nabla u_{\mathbf{h}} : \quad W_i(\mathbf{h}) \cap \Omega \mapsto p_i, i = 1, 2, \ldots, n. \tag{3.15}$$

The inverse of the OT map $T : \Omega \to \Omega^*$ is also an OT map $T^{-1} : \Omega^* \to \Omega$, where the Brenier potential is $u^*_{\mathbf{h}}$, and $u_{\mathbf{h}} = (u^*_{\mathbf{h}})^*$, the classical gradient is replaced by sub-gradient in Eqn. (3.5),

$$T^{-1} = \partial u^*_{\mathbf{h}} : \quad p_i \mapsto W_i(\mathbf{h}) \cap \Omega, i = 1, 2, \ldots, n. \tag{3.16}$$

The existence and the uniqueness of the solution is proved in [21] using geometric variational approach. In the following we give a brief proof based on Kantorovich dual theorem.

Theorem 3.17 (Gu et al. [21]). *Let Ω be a compact convex domain in $\mathbb{R}^d$, $\Omega^* = \{p_1, ..., p_n\}$ be a set of distinct points in $\mathbb{R}^d$, the source measure density function be $d\mu(x) = f(x)dx$, $f : \Omega \to \mathbb{R}$ be positive and continuous. Then for any $\nu_1, \ldots, \nu_n > 0$ with $\sum_{i=1}^n \nu_i = \mu(\Omega)$, there exists $\mathbf{h} = (h_1, h_2, \ldots, h_n) \in \mathbb{R}^n$, unique up to adding a constant $(c, c, \ldots, c)$, so that the projection of $u_{\mathbf{h}}$*

$$\mu(W_i(\mathbf{h}) \cap \Omega) = \int_{W_i(\mathbf{h}) \cap \Omega} f(x)dx = \nu_i \quad \forall i = 1, 2, \ldots, n.$$

The height vector $\mathbf{h}$ is exactly the optimal solver of the following convex function

$$E(\mathbf{h}) = \int_{\mathbf{0}}^{\mathbf{h}} \sum_{i=1}^{n} h_i \nu_i - \sum_{i=1}^{n} \mu(W_i(\mathbf{h}) \cap \Omega) dh_i \tag{3.17}$$

on the open convex set (the admissible solution space)

$$\mathcal{H} = \{\mathbf{h} \in \mathbb{R}^n | \mu(W_i(\mathbf{h}) \cap \Omega) > 0, \,\forall i\} \bigcap \left\{ \sum_{i=1}^n h_i = 0 \right\}. \tag{3.18}$$

The gradient map $\nabla u_{\mathbf{h}} : W_i \to p_i \;\forall i = 1, 2, \ldots, n$ *minimizes the quadratic cost* $\frac{1}{2}\int_\Omega |x - T(x)|^2 f(x) dx$ *among all the measure preserving transportation maps* $T : (\Omega, \mu) \to (\mathbb{R}^d, \, \nu = \sum_{i=1}^n \nu_i \delta_{p_i})$ *with* $T_\# \mu = \nu$*.*

Proof. In order to find the OT map,

$$\max_{T_\# \mu = \nu} \int_\Omega \langle x, T(x) \rangle d\mu(x),$$

by the setup of the dual problem Eqn. (3.2), we need to find the Kantorovich potential $u_{\mathbf{h}}$ and its Legendre dual $u^*_{\mathbf{h}}$ to minimize the potential,

$$\begin{aligned}
\min_{\mathbf{h}} E(\mathbf{h}) &= \min_{\mathbf{h}} \int_\Omega u_{\mathbf{h}}(x) d\mu(x) + \int_{\Omega^*} u^*_{\mathbf{h}}(y) d\nu(y) \\
&= \min_{\mathbf{h}} \int_\Omega \sup_{y \in \Omega^*} (\langle x, y \rangle - u^*_{\mathbf{h}}(y)) d\mu(x) + \int_{\Omega^*} u^*_{\mathbf{h}}(y) d\nu(y) \\
&= \min_{\mathbf{h}} \int_\Omega \max_{i=1}^n \{\langle x, p_i \rangle - h_i\} d\mu(x) + \sum_{i=1}^n h_i \nu_i \\
&= \min_{\mathbf{h}} \sum_{i=1}^n \int_{W_i(\varphi) \cap \Omega} (\langle x, p_i \rangle - h_i) d\mu(x) + \sum_{i=1}^n h_i \nu_i.
\end{aligned}$$

We set the μ-volume of each cell as

$$w_i(\mathbf{h}) := \mu(W_i(\mathbf{h}) \cap \Omega). \tag{3.19}$$

By direct computation, the gradient is

$$\nabla E(\mathbf{h}) = (\nu_1 - w_1(\mathbf{h}), \nu_2 - w_2(\mathbf{h}), \cdots, \nu_n - w_n(\mathbf{h}))^T. \tag{3.20}$$

Suppose the Hessian matrix of the energy is $\mathrm{Hess}(E(\mathbf{h})) = (h_{ij})$, its off-diagonal element h_{ij} is the ratio between the length of the corresponding Voronoi and Delaunay edges,

$$h_{ij} = -\frac{\partial w_i(\mathbf{h})}{\partial h_j} = \frac{1}{|p_i - p_j|} \mu(W_i(\mathbf{h}) \cap W_j(\mathbf{h})) = h_{ji}. \tag{3.21}$$

Because the total volume of the cells is constant $\sum_{i=1}^n w_i(\mathbf{h}) = \mu(\Omega)$, therefore

$$h_{ii} = \frac{\partial w_i(\mathbf{h})}{\partial h_i} = -\sum_{j\neq i} \frac{\partial w_j(\mathbf{h})}{\partial h_i} = -\sum_{j\neq i} h_{ji} = -\sum_{j\neq i} h_{ij}. \tag{3.22}$$

This shows the Hessian matrix has one dimensional null space $(1, 1, \dots, 1)^T$. In the complement space $\sum_{i=1}^n h_i = 0$, the Hessian matrix is diagonal dominant, therefore it is negative definite and the energy is strictly convex.

The admissible solution space $\mathcal{H}$ in Eqn. (3.18) is convex according to Brunn-Minkowski inequality. The OT map is the optimum interior point in $\mathcal{H}$, therefore it exists and is unique. ☐

3.6. Semi-discrete WT maps

The WT theorem 3.12 can be directly generalized to the discrete situation as follows.

Theorem 3.18 (Semi-discrete Worst Transportation Map)**.** *Let* Ω *be a compact convex domain in* $\mathbb{R}^{2d}$, $\{p_1, ..., p_n\}$ *be a set of distinct points in* $\mathbb{R}^{2d}$ *and* $f : \Omega \to \mathbb{R}$ *be a positive continuous function. Then for any* $\nu_1, \dots, \nu_n > 0$ *with* $\sum_{i=1}^n \nu_i = \int_\Omega f(x)dx$, *there exists* $\mathbf{h} = (h_1, h_2, \dots, h_n) \in \mathbb{R}^n$, *unique up to adding a constant* $(c, c, \dots, c)$, *so that*

$$\mu(W_i(\mathbf{h}) \cap \Omega) = \int_{W_i(\mathbf{h})\cap\Omega} f(x)dx = \nu_i \quad \forall i.$$

The height vectors $\mathbf{h}$*'s are exactly the maximum points of the concave function*

$$E(\mathbf{h}) = \sum_{i=1}^n h_i \nu_i - \int_0^h \sum_{i=1}^n \mu(W_i(h) \cap \Omega) dh_i \tag{3.23}$$

On the open convex set (admissible solution space)

$$\mathcal{H} = \{\mathbf{h} \in \mathbb{R}^n | \mu(W_i(\mathbf{h}) \cap \Omega) > 0, \ \forall i\} \bigcap \left\{ \sum_{i=1}^n h_i = 0 \right\}. \tag{3.24}$$

Furthermore, the gradient map $\nabla u_{\mathbf{h}}$ *maximizes the quadratic cost* $\frac{1}{2}\int_\Omega |x - T(x)|^2 f(x)dx$ *among all the measure preserving transport maps* $T : (\Omega, \mu) \to (\mathbb{R}^{2d}, \nu = \sum_{i=1}^n \nu_i \delta_{p_i})$, $T_\# \mu = \nu$.

Proof. Given the height vector $\mathbf{h} = (h_1, h_2, \cdots, h_n)$, $\mathbf{h} \in \mathcal{H}$, we construct the upper convex hull of $v_i(\mathbf{h}) = (p_i, h_i)$'s, see Fig. 3(d), each vertex corresponds to a plane $\pi_i(\mathbf{h}, x) := \langle p_i, x\rangle - h_i$. The convex hull is dual to the lower envelope of the plane $\pi_i(\mathbf{h}, \cdot)$'s, see Fig. 3(b), which is the graph of the concave function,

$$u_{\mathbf{h}}(x) := \min_{i=1}^{n} \langle p_i, x\rangle - h_i. \tag{3.25}$$

The projection of the lower envelope induces a farthest power diagram $\mathcal{D}(\mathbf{h})$, see Fig. 3 top right frame,

$$\Omega = \bigcup_{i=1}^{n} W_i(\mathbf{h}) \cap \Omega, \quad W_i(\mathbf{h}) := \{x \in \Omega, \nabla u_{\mathbf{h}}(x) = p_i\}.$$

The μ-volume of each cell is defined as

$$w_i(\mathbf{h}) := \mu(W_i(\mathbf{h}) \cap \Omega) = \int_{W_i(\mathbf{h})\cap\Omega} f(x)dx. \tag{3.26}$$

By direct computation similar to lemma in [22], we can show the symmetric relation holds:

$$\frac{\partial w_i(\mathbf{h})}{\partial h_j} = \frac{\partial w_j(\mathbf{h})}{\partial h_i} = \frac{1}{|p_i - p_j|} \int_{W_i(\mathbf{h})\cap W_j(\mathbf{h})\cap\Omega} f(x)ds. \tag{3.27}$$

This shows the differential form $\omega = \sum_{i=1}^{n} w_i(\mathbf{h})dh_i$ is a closed one-form. By Brunn-Minkowski inequality, we can show that the admissible height space $\mathcal{H}$ in Eqn. (3.24) is convex, therefore simply connected. Hence ω is exact. So the energy $E(\mathbf{h}) := \int_0^{\mathbf{h}} \omega$ is well defined. Furthermore, Hessian matrix of $E(\mathbf{h})$ is given by

$$\frac{\partial^2 E(\mathbf{h})}{\partial h_i \partial h_j} = \frac{w_i(\mathbf{h})}{\partial h_j} \geq 0, \tag{3.28}$$

since the total volume of all the cells is the constant $\mu(\Omega)$, we obtain

$$\frac{\partial^2 E(\mathbf{h})}{\partial h_i^2} = -\sum_{j\neq i} \frac{w_i(\mathbf{h})}{\partial h_j} < 0. \tag{3.29}$$

Therefore the Hessian matrix is negative definite in $\mathcal{H}$, the energy is strictly concave in $\mathcal{H}$. By adding a linear term, the energy is still strictly concave,

$$E(\mathbf{h}) = \int_0^{\mathbf{h}} \sum_{i=1}^n w_i(\mathbf{h}) dh_i - \sum_{i=1}^n \nu_i h_i.$$

The gradient of $E(\mathbf{h})$ is given by

$$\nabla E(\mathbf{h}) = (w_1(\mathbf{h}) - \nu_1, w_2(\mathbf{h}) - \nu_2, \cdots, w_n(\mathbf{h}) - \nu_n). \tag{3.30}$$

On the boundary of $\mathcal{H}$, there is an empty cell $W_k(\mathbf{h})$, and the k-th component of the gradient is $-\nu_k$, pointing to the interior of $\mathcal{H}$. This shows the global unique maximum of the energy is the interior point of $\mathcal{H}$. At the maximum point $\mathbf{h}^*$, $\nabla E(\mathbf{h}^*)$ is zero, $w_i(\mathbf{h}^*) = \nu_i$, hence $\mathbf{h}^*$ is the unique solution to the semi-discrete WT problem. $\square$

3.7. Symmetry between the optimal and the worst transportation maps

The OT map and WT map have intrinsic symmetry. We illustrate it by an example. Fig. 5 shows the source measure (Ω, μ), a facial surface is conformally mapped to the unit disk, the surface area element is push forwarded to the disk, and defines μ.

The target measure (Ω^*, ν) is the uniform distribution on the unit disk, namely $g(y) \equiv 1$. Let $\varphi : \Omega^* \to \Omega^*$ be the reflection in the origin, $\varphi(x) = -x$. Then the target measure has the symmetry, $\varphi_\# \nu = \nu$, namely $g(-y) = g(y)$. The shape of the support Ω^* also has the symmetry, $\varphi(\Omega^*) = \Omega^*$. In this case, the OT map and WT map are symmetric with respect to each other. The OT map $T : (\Omega, \mu) \to (\Omega^*, \nu)$ is shown in the left frame of the bottom row of Fig. 5, the WT map $t : (\Omega, \mu) \to (\Omega^*, \nu)$ is shown in the right frame of the bottom row of Fig. 5. They differ by a reflection in the origin, namely $T \circ t^{-1} = \varphi$.

If (Ω^*, ν) is not symmetric, $\varphi_\# \nu \neq \nu$, then the OT map and WT map are not symmetric either. As shown in Fig. 6, when the target domain Ω^* is not symmetric about the origin, the images of OT map and WT map are not symmetric anymore.

Theorem 3.19 (Continuous Symmetry). *Given the probability measures μ and ν with compact supports $\Omega, \Omega^* \subset \mathbb{R}^{2d}$ respectively, and equal total mass $\mu(\Omega) = \nu(\Omega^*)$. Assume the density functions $f, g \in L^1(\mathbb{R}^d, \Omega, \Omega^*)$, furthermore, Ω^* is symmetric about the origin, $\varphi_\# \nu = \nu$, where $\varphi(y) = -y$*

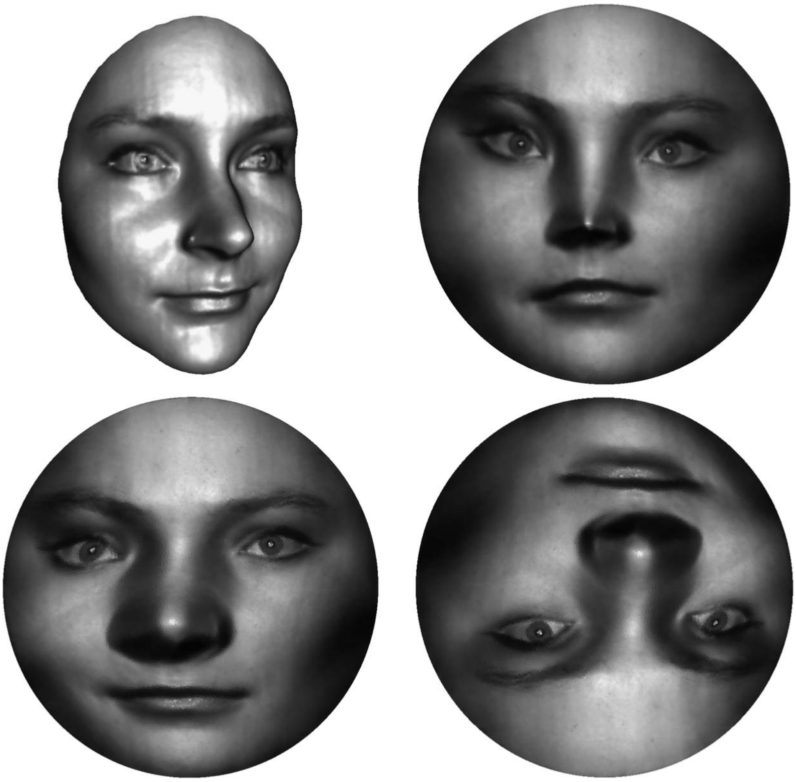


Figure 5: The top row is the target measure (Ω^*, ν), which is induced by a conformal mapping from a female face to the planar disk. The push-forwarded surface area element is ν. The bottom row shows the source measure (Ω, μ), which is the uniform distribution on the disk. The OT map image (left) and the WT map image (right) differ by a rotation of angle π, namely they are symmetric under the reflection in the origin.

is the reflection in the origin. Then the OT map $T : \Omega \to \Omega^*$ and the WT map $t : \Omega \to \Omega^*$ differ by φ, $T = \varphi \circ t$, $t = \varphi \circ T$.

Proof. By Brenier Thm. 3.6, there is a convex Brenier potential $u : \Omega \to \mathbb{R}$, the OT map $T = \nabla u$. By Thm. 3.12, there is a concave function $v : \Omega \to \mathbb{R}$, such that the WT map $t = \nabla v$. $-u$ is a concave function, $-T = -\nabla u$,

$$(-\nabla u)_{\#}\mu = (\varphi \circ T)_{\#}\mu = \varphi_{\#}(T_{\#}\mu) = \varphi_{\#}\nu = \nu,$$

therefore $(-T)_{\#}\mu = \nu$, and $-T$ is the gradient map of a concave function $-u$. By the uniqueness of the WT map, we have $-T = t$ and the worst

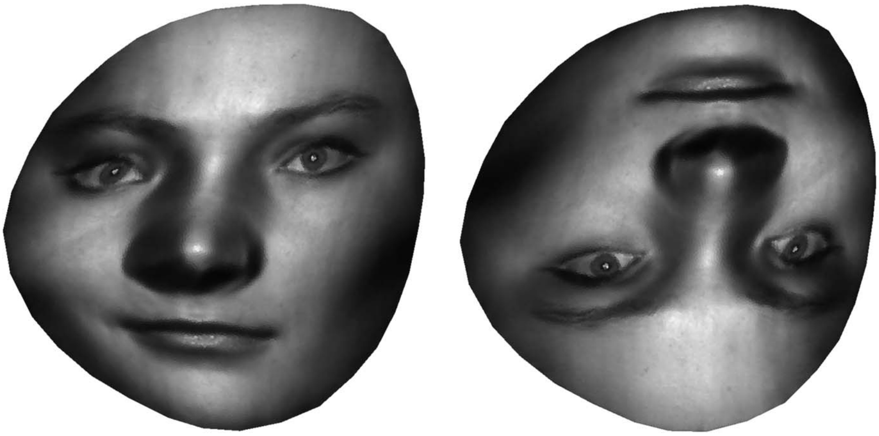


Figure 6: The target measure (Ω^*, ν) is the uniform distribution on an irregular domain, the OT map image (left), the WT map image (right) are not symmetric.

Brenier potential equals to $-u$. This shows $t = \varphi \circ T$. Similarly, we can show $t \circ \varphi = T$. ☐

Theorem 3.20 (Semi-discrete Symmetry)**.** *Let Ω be a compact convex domain in $\mathbb{R}^{2d}$, the measure μ is with continuous density function $d\mu = f(x)dx$, $f : \Omega \to \mathbb{R}$. $\varphi_{\#}\mu = \mu$, where $\varphi(x) = -x$ is the reflection in the origin. The target measure $\nu = \sum_{i=1}^n \nu_i \delta(y - p_i)$, where $\{p_1, \ldots, p_n\}$ be a set of distinct points in $\mathbb{R}^{2d}$, and $\nu_1, \ldots, \nu_n > 0$ with $\sum_{i=1}^n \nu_i = \mu(\Omega)$. Then the OT map $T : \Omega \to \{p_i\}_{i=1}^n$ and the WT map $t : \Omega \to \{p_i\}_{i=1}^n$ differ by a reflection in the origin, $T \circ \varphi = t$, $t \circ \varphi = T$.*

Proof. By Thm. 3.17, there is a convex Brenier potential $u : \Omega \to \{p_i\}_{i=1}^n$, $T = \nabla u$, $T_{\#}\mu = \nu$.

$$(-\nabla u)_{\#}\mu = (\nabla u(-x))_{\#}\mu = (\nabla u)_{\#}(\varphi_{\#}\mu) = (\nabla u)_{\#}\mu = \nu,$$

according to the uniqueness of the WT map, we obtain $T \circ \varphi = t$. Similarly, we can show $t \circ \varphi = T$. ☐

In summary, if the source measure is symmetric, $\varphi_{\#}\mu = \mu$, then $T \circ \varphi = t$ and $t \circ \varphi = T$; if the target measure is symmetric, $\varphi_{\#}\nu = \nu$, then $\varphi \circ T = t$ and $\varphi \circ t = T$.

4. Computational algorithms

4.1. OT map algorithm

The algorithm for computing the semi-discrete OT map is to optimize the convex energy Eqn. (3.17) using Newton's method in the admissible space $\mathcal{H}$ in Eqn. (3.24). By Newton's method, the searching direction $\mathbf{d}$ is the solution of the linear system,

$$\text{Hess}(E(\mathbf{h}))\mathbf{d} = \nabla E(\mathbf{h}). \tag{4.1}$$

In each step, we update the current height vector $\mathbf{h}$ by $\mathbf{h} - \lambda\mathbf{d}$ with λ being a carefully chosen step length.

To compute the gradient and the Hessian matrix of the energy Eqn. (3.17), we need to compute the Brenier potential $u_{\mathbf{h}}$ and its corresponding cell decomposition $\mathcal{D}(\mathbf{h})$, which is realized through computing the dual convex hull $u_{\mathbf{h}}^*$ and its corresponding weighted Delaunay triangulation $\mathcal{T}(\mathbf{h})$. To keep the Hessian matrix nondegenerate, we need to carefully select the step length λ such that $\mathbf{h} - \lambda\mathbf{d}$ is still in the admissible space $\mathcal{H}$.

During the optimization, the Brenier potential $u_{\mathbf{h}}$ and its Legendre dual $u_{\mathbf{h}}^*$ are updated step by step. Su et al. [43] construct the new convex hull $u_{\mathbf{h}}^*$ at each step with time complexity $O(n\log(n))$ (such as the incremental algorithm, or the divide and conquer algorithm [10]). This is the most time consuming step in the pipeline. We observe that in the later stages of the optimization, the combinatorial structure of the convex hull $u_{\mathbf{h}}^*$ keeps invariant, thus it is unnecessary to construct the new triangulation structure $\mathcal{T}(\mathbf{h})$ in each step [43]. Here, we propose to modify the previous triangulated convex hull $u_{\mathbf{h}}^*$ to make the current $\mathcal{T}(\mathbf{h} - \lambda\mathbf{d})$ weighted Delaunay triangulation. To do so, we generalize the Lawson's edge flip algorithm [28] from conventional Delaunay triangulations to the weighted Delaunay triangulations (section 4.1.1). With the convex hull $u_{\mathbf{h}}^*$ and the weighted Delaunay triangulation $\mathcal{T}(\mathbf{h})$, we can then compute the Brenier potential $u_{\mathbf{h}}$ (section. 4.1.2) and its corresponding cell decomposition $\mathcal{D}(\mathbf{h})$ (section 4.1.3). This greatly improves the computational efficiency.

By translation and scaling, we can assume $\Omega^* \subset \Omega$, namely all the samples p_is are contained in the source domain Ω. To keep the height vector $\mathbf{h}$ in the admissible space during the optimization, we initialize $\mathbf{h} = 1/2(|p_1|^2, |p_2|^2, \cdots, |p_n|^2)$ and $\mathbf{h} = \mathbf{h} - \text{mean}(\mathbf{h})$. Its corresponding power diagram $\mathcal{D}(\mathbf{h})$ is the classical power diagram, and the corresponding weighted Delaunay triangulation $\mathcal{T}(\mathbf{h})$ is the traditional Delaunay triangulation of

Algorithm 1: Optimal Transportation Map

```
Input: $\Omega$, $\mu$, $\nu$, threshold $\varepsilon$
Output: The Brenier potential $u_{\mathbf{h}}$
Normalize $\{p_1, p_2, \ldots, p_n\}$ to be inside $\Omega$ by translation and scaling;
Initialize $h_i = \langle p_i, p_i \rangle / 2$;
while true do
    Call Alg. 3 to compute the power diagram $\mathcal{D}(\mathbf{h})$;
    Compute $\mu(W_i)$ $i = 1, 2, \ldots, n$ using Eqn. (4.2);
    Compute $\nabla E(\mathbf{h})$ using Eqn. (3.20);
    if $\|\nabla E(\mathbf{h})\| < \varepsilon$ then
        break;
    else
        Compute the $\mu$-lengths of the power Voronoi edges using Eqn. (4.3);
        Construct the Hessian matrix of the energy $E(\mathbf{h})$ using Eqn. (3.21);
        Find the searching direction $\mathbf{d}$ by solving the linear system in Eqn. (4.1);
        Set $\lambda = 1$ ;
        repeat
            Compute the power diagram $\mathcal{D}(\mathbf{h} - \lambda \mathbf{d})$;
            if $\mathbf{h} - \lambda \mathbf{d} \notin \mathcal{H}$ then
                $\lambda \leftarrow \frac{1}{2}\lambda$;
                continue;
            end
        until no empty power cell;
        Update the height vector $\mathbf{h} \leftarrow \mathbf{h} - \lambda \mathbf{d}$;
    end
end
```

$\{p_i\}_{i=1}^n$. Since all the samples are inside Ω, all the Voronoi cells $W_i(\mathbf{h})$'s are non-empty. Thus, the initialized $\mathbf{h}$ is in the admissible space $\mathcal{H}$ in Eqn. (3.24). Given the previous height vector $\mathbf{h} \in \mathcal{H}$ and the search direction $\mathbf{d}$, we firstly set $\lambda = 1$ and compute $\mathcal{D}(\mathbf{h} - \lambda \mathbf{d})$. If some cells of $\mathcal{D}(\mathbf{h} - \lambda \mathbf{d})$ are empty, namely $\mathbf{h} - \lambda \mathbf{d}$ is not in the admissible space $\mathcal{H}$, we then set $\lambda = \frac{1}{2}\lambda$ and repeat the above procedure until there is no empty cell in $\mathcal{D}(\mathbf{h} - \lambda \mathbf{d})$. In such a way, we find the proper step length λ to keep the updated height vector $\mathbf{h} := \mathbf{h} - \lambda \mathbf{d}$ in the admissible space $\mathcal{H}$.

Furthermore, Su et al. [43] assume the source distribution to be uniform and cannot handle more complex source measures. In Sec. 4.1.4, we propose the sweep convex polygon algorithm that can help compute the OT map from a piecewise linearly defined source measure to the target discrete measure.

We put all the pieces together, and the whole algorithmic pipeline is summarized in Alg. 1. The evaluations of μ-areas of the cells and the μ-lengths of the power Voronoi edges use Eqn. (4.2) and Eqn. (4.3), respectively.

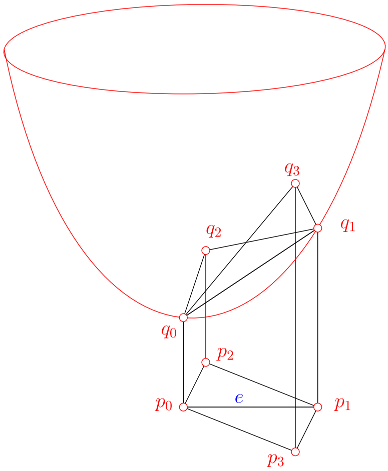


Figure 7: Edge local weighted Delaunay condition.

The computational process is to optimize the convex energy Eqn. (3.17) in the admissible solution space Eqn. (3.24). At each step, we compute (i) the convex hull $u^*_{\mathbf{h}}$ and its corresponding weighted Delaunay triangulation $\mathcal{T}(\mathbf{h})$ with our generalized Lawson's edge flip, (ii) the upper envelope $u_{\mathbf{h}}$ by Legendre dual and (iii) the power diagram $\mathcal{D}(\mathbf{h})$ through the Sutherland-Hodgman clipping. We also introduce the sweep convex polygon algorithm to handle the piecewise linearly defined source measures.

4.1.1. Generalized Lawson's edge flip The generalized Lawson's edge flip is used to compute the weighted Delaunay triangulation $\mathcal{T}(\mathbf{h} - \lambda\mathbf{d})$ through updating $\mathcal{T}(\mathbf{h})$. In general, the Lawson's edge flip algorithm does not guarantee the recovery of the weighted Delaunay triangulations. An elementary example of when the algorithm will fail to terminate is shown in [16]. In our case, if the Lawson's edge flip fails to terminate, there must exist some vertices p_is that do not belong to the weighted Delaunay triangulation $\mathcal{T}(\mathbf{h}-\lambda\mathbf{d})$. Equivalently, the dual Voronoi cells W_is of such vertices in $\mathcal{D}(\mathbf{h} - \lambda\mathbf{d})$ are empty. In such a case, we claim that $\mathbf{h} - \lambda\mathbf{d} \notin \mathcal{H}$, and λ needs to be further decreased to make $\mathbf{h}-\lambda\mathbf{d} \in \mathcal{H}$. If the following Lawson's edge flip works well, then we can finally get the weighted Delaunay triangulation $\mathcal{T}(\mathbf{h}-\lambda\mathbf{d})$.

The idea of Lawson's edge flip algorithm is to check whether each interior edge of a triangulation is *locally weighted Delaunay*. If not, then the algorithm tries to flip it to become locally weighted Delaunay. The locally weighted Delaunay is shown in Fig. 7. Given an edge $e = (p_0, p_1)$ in a planar triangulation $\mathcal{T}$, we 1) find its two neighboring faces, 2) lift the four

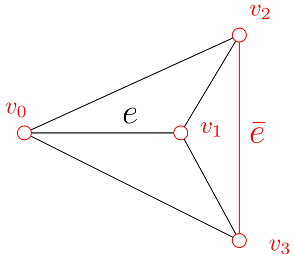


Figure 8: A non-flippable edge.

vertices to the convex hull, and 3) compute the volume of the tetrahedron $[q_0, q_1, q_2, q_3]$ with $q_i = (p_i, h_i)$. If the volume is positive, then e is locally weighted Delaunay. If the volume is negative, then e is not locally weighted Delaunay. Before flipping an edge, we need to check whether it is flippable.

As shown in the Fig. 8, if we flip the edge e to edge $\bar{e}$, one triangle becomes clockwise, thus the edge e is not flippable. *Our generalized Lawson's edge flip algorithm* marks all the non-local Delaunay edges and pushes them into a stack. When the stack is non-empty, the top edge e is popped and unmarked. If e is still non-local weighted Delaunay and flippable, then e is flipped and its unmarked neighboring edges are pushed into the stack. If the stack is empty but there are still some non local weighted Delaunay edges, it means that not all the points (p_i, h_i) are on the convex hull and we need to decrease λ. It has been proven that if all the points (p_i, h_i) are on the convex hull, then the Lawson's edge flip algorithm converges and gives the desired weighted Delaunay triangulation $\mathcal{T}(\mathbf{h} - \lambda \mathbf{d})$ of p_i's [10], which is the same as the triangulation given by the convex hull of (p_i, h_i)s. The bottom row of Fig. 3 shows a convex hull obtained by the generalized Lawson edge flip algorithm, whose pipeline is given by Alg. 2.

4.1.2. Legendre dual The convex hull is the graph of $u^*_{\mathbf{h}}$, and the Brenier potential $u_{\mathbf{h}}$ is the Legendre dual of $u^*_{\mathbf{h}}$ and is represented as the upper envelope of planes $\{\pi_i(x) = \langle x, p_i \rangle - h_i\}_{i=1}^n$. The upper envelope $u_{\mathbf{h}}$ is constructed as follows: given a vertex $q_i = (p_i, h_i)$ on the convex hull $u^*_{\mathbf{h}}$, it corresponds to a planar face of the upper envelope $u_{\mathbf{h}}$ with the supporting plane $\pi_i(x) = \langle p_i, x \rangle - h_i$. Given an edge $[q_i, q_j]$ on the convex hull, its dual is an edge on the upper envelope $u_{\mathbf{h}}$ along the intersection line between π_i and π_j. Given a triangle $[q_i, q_j, q_k]$ on the convex hull, it corresponds to a vertex on the upper envelope $u_{\mathbf{h}}$ that is the intersection of three planes π_i, π_j and π_k.

Since we only need the projection of the Brenier potential on the regions around Ω, we add an infinity vertex $(0, 0, -\infty)$, which represents a horizontal

Algorithm 2: Generalized Lawson's Edge Flip

```
Input: The planar point set $\{p_1, \dots, p_n\}$ with heights $\{h_1, \dots, h_n\}$, the initial
       triangulation $\mathcal{T}$ of $p_i$s
Output: The convex hull of $\{(p_i, h_i)\}$, the weighted Delaunay triangulation of
        $\{p_i\}$, all $(p_i, h_i)$s are on the convex hull or not
Push all non-locally interior edges of $\mathcal{T}$ into a stack and mark them;
while the stack is non-empty do
    $e \leftarrow pop()$, unmark $e$;
    if $e$ is not locally weighted Delaunay then
        if $e$ can be flipped then
            flip edge $e$;
            push the other four edges of the two triangles adjacent to $e$ into the
              stack if they are unmarked;
        end
    end
end
return all edges are local weighted Delaunay
```

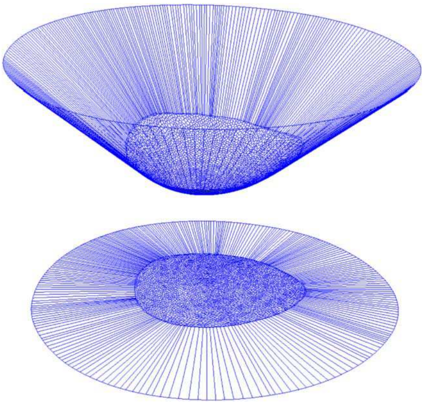


Figure 9: Legendre dual of the augmented convex hull.

plane $z = \infty$ in the upper envelope $u^*_{\mathbf{h}}$, to the convex hull and get an augmented convex hull. The Legendre dual of the augmented convex hull gives the upper envelope truncated by the horizontal plane $z = \infty$. Fig. 9 shows the Legendre dual of the augmented convex hull (corresponds to the bottom row of Fig. 3 obtained by the algorithm, where the height of the horizontal plane is chosen to be large enough.

4.1.3. Sutherland-Hodgman clipping The projection of the Legendre dual of the augmented convex hull is a planar power diagram, denoted as

$\bar{\mathcal{D}}(\mathbf{h})$ defined on a region including Ω. It is then clipped by the boundary of the source domain $\partial\Omega$. In our algorithm Ω is required to be convex. Since each cell of the power diagram $\bar{\mathcal{D}}(\mathbf{h})$ is a convex planar polygon, and $\partial\Omega$ is also convex, the clipping can be carried out using the classical Sutherland-Hodgman algorithm [45], which goes through all the edges of $\partial\Omega$. Basically, suppose $\bar{W}_i$ is a cell of $\bar{\mathcal{D}}(\mathbf{h})$, then the line through one edge of $\partial\Omega$ divides the plane into two halves and only the left half where Ω belongs to is kept. This may cut off a corner portion of $\bar{W}_i$. The remaining part of $\bar{W}_i$ after the corner cuttings of all the lines in $\partial\Omega$ is the intersection between $\bar{W}_i$ and $\partial\Omega$, namely W_i of $\mathcal{D}(\mathbf{h})$.

The algorithm for constructing the power diagram $\mathcal{D}(\mathbf{h})$ is obtained by combining all the procedures explained so far. As shown in Alg. 3, for a given set of samples $\{p_i\}_{i=1}^n$ and the corresponding height vector $\mathbf{h}$, it produces the power diagram of Ω induced by the upper envelope $u_{\mathbf{h}}$.

Algorithm 3: Power Diagram and Upper Envelope

```
Input: $\Omega$, samples $\{p_i\}_{i=1}^n$, heights $\{h_i\}_{i=1}^n$
Output: The power diagram of $\Omega$ induced by the samples and its corresponding
        heights
Use the generalized Lawson's edge flip Alg. 2 to compute the convex hull;
Add the infinity vertex $(0, 0, -\infty)$ to the convex hull and get the augmented
  convex hull;
Compute the Legendre dual of the augmented convex hull to obtain the
  unclipped upper envelope;
Project the unclipped upper envelope to obtain the unclipped power diagram
  $\bar{\mathcal{D}}(\mathbf{h})$ ;
Clip the unclipped power diagram $\bar{\mathcal{D}}(\mathbf{h})$ by $\partial\Omega$ using the Sutherland-Hodgman
  algorithm [45] to obtain the power diagram $\mathcal{D}(\mathbf{h})$;
Lift $\mathcal{D}(\mathbf{h})$ to obtain the clipped upper envelope $u_{\mathbf{h}}$ defined on $\Omega$.
```

4.1.4. Sweep convex polygon algorithm In practice, the source domain Ω is triangulated as $\mathcal{S}$ and $\Omega = \bigcup_{j=1}^m \Delta_j$. The density function $d\mu = f(x)dx$ can be represented as a piecewise linear function defined on the triangulation $\mathcal{S}$ and $f(x)$ is linear on each triangle face Δ_j. The power diagram $\mathcal{D}$ partitions Ω into cells, $\Omega = \bigcup_{i=1}^n W_i(\mathbf{h})$. In order to evaluate the μ-area of each cell $w_i(\mathbf{h}) = \int_{W_i(\mathbf{h})} f(x)dx$ and compute the μ-length of each power edge, we need to compute the subdivision of $\mathcal{S}$ and $\mathcal{D}$,

$$\Omega = \bigcup_{i=1}^{n}\bigcup_{j=1}^{m} W_i(\mathbf{h}) \cap \Delta_j,$$

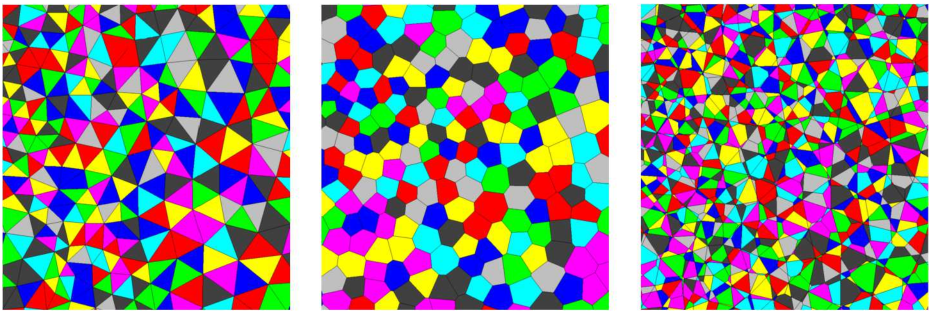


Figure 10: The sweep convex polygon algorithm to compute the overlay (right) of two cell decompositions (left $\mathcal{S}$ and middle $\mathcal{D}$).

as shown in Fig. 10. The left frame shows the triangulation $\mathcal{S}$ of Ω, the middle frame the power diagram $\mathcal{D}$, and the right frame demonstrates the subdivision of $\mathcal{S}$ and $\mathcal{D}$.

A naive way to compute the subdivision is to compute the intersection of each pair of cell and triangle $(W_i(\mathbf{h}), \Delta_j)$ using the Sutherland-Hodgman algorithm [45], and the complexity will be $O(mn)$, where n, m are the number of cells and triangles, respectively. Since most of the pairs have empty intersection, this method is highly inefficient. An efficient solution is to use the Bentley-Ottman's sweepline algorithm [9]. Unfortunately, the original sweepline algorithm is notoriously vulnerable to arithmetic errors. In order to ensure the stability of the algorithm, people usually use exact arithmetic methods, such as rational computation. But this will greatly slow down the computation. To solve this problem, we propose a method called *the sweep convex polygon algorithm* that combines the simplicity and stability of the Sutherland-Hodgman method with the efficiency of the Bently-Ottman method, as shown in Alg. 4. At first, we sort the planar points from left to right, then from bottom to top. For the planar polygons, we sort them by their minimal vertices. If two polygons share the same minimal vertex, we sort them by the slope of the edge starting from the minimal vertex. If their slopes are the same, we consider the second minimal vertex and so on. Similar to the classical Bentley-Ottman algorithm, the algorithm preserves an event queue (X-structure) and a sweepline data structure (Y-structure).

The algorithm sweeps a vertical line from left to right across a polygon, if the sweepline crosses the minimal vertex of a polygon, the polygon is born; if the sweepline goes through the maximal vertex of an alive polygon, then the polygon is killed. The births and the deaths of all the polygons are treated as events. All the events are put into the event queue and sorted

Algorithm 4: The Sweep Convex Polygon Method

```
Input: The triangulation $\mathcal{S}$ and the power diagram $\mathcal{D}$ of $\Omega$
Output: The subdivision of $\mathcal{S}$ and $\mathcal{D}$
For each polygon in $\mathcal{S}$ and $\mathcal{D}$, compute its minimal and maximal vertex;
Enqueue all the birth and death events to the event queue;
while the event queue is not empty do
    Pop the event queue to get an event $et$;
    if $et$ is a birth event then
        Find all the newborn polygons, whose minimal vertices equal to the event point;
        Insert all the newborn polygons to the Y-structure;
        for each newborn $\sigma \in \mathcal{S}$ and each $\tau \in \mathcal{D}$ in the Y-structure do
            if $\tau$ and $\sigma$ have not been tested for intersection then
                use the Sutherland-Hodgman algorithm to compute $\sigma \cap \tau$;
            end
        end
    else
        Find all the dead polygons in the Y-structure, whose maximal vertices equal to the event point;
        Delete all the dead polygons from the Y-structure;
    end
end
```

by the corresponding minimal and maximal vertex positions. At each step, the algorithm handles the top event in the event-queue. If it is a birth event, then all the polygons starting from the event point are added to the Y-structure. For each newly inserted polygon $\sigma \in \mathcal{S}$ and each polygon in the Y-structure $\tau \in \mathcal{D}$, the algorithm performs the intersection test, and finds the overlapping $\sigma \cap \tau$ between them. Similarly, for each newly inserted polygon $\tau \in \mathcal{D}$ and each polygon in the Y-structure $\sigma \in \mathcal{S}$, the algorithm finds the intersection. In practice, the bounding box of each polygon is used to speed up the intersection testing.

Finally we obtain the subdivision of $\mathcal{D}$ and $\mathcal{S}$, formed by the convex polygons $\{W_i(\mathbf{h}) \cap \Delta_j\}_{i=1,j=1}^{n,m}$, as shown in the right frame of Fig. 10. Then the μ-area of a cell $W_i(\mathbf{h})$ is calculated as:

$$\mu(W_i(\mathbf{h})) = \sum_j \int_{W_i(\mathbf{h}) \cap \Delta_j} f(x)\big|_{\Delta_j} dx. \tag{4.2}$$

Since the restriction of the density function $f(x)$ of each Δ_j, namely $f(x)\big|_{\Delta_j}$, is linear, the computation is straightforward. Each power Voronoi edge $W_i(\mathbf{h}) \cap W_k(\mathbf{h})$ is also segmented by Δ_j's. The μ-length of each segment

$W_i(\mathbf{h}) \cap W_k(\mathbf{h}) \cap \Delta_j$ is evaluated, and summed up to obtain the μ-length of $W_i(\mathbf{h}) \cap W_k(\mathbf{h})$,

$$\mu(W_i(\mathbf{h}) \cap W_k(\mathbf{h})) = \sum_j \int_{W_i(\mathbf{h}) \cap W_k(\mathbf{h}) \cap \Delta_j} f(x)\big|_{\Delta_j} ds. \tag{4.3}$$

4.2. WT map algorithm

This section explains a unified algorithm to compute both the OT and WT maps.

The algorithm mainly optimizes the energy $E(\mathbf{h})$ in the admissible solution space $\mathcal{H}$ using Newton's method. At the beginning, for the OT (WT) map, the height vector $\mathbf{h}_0$ is initialized as $h_i = \frac{1}{2}|p_i|^2$ ($h_i = -\frac{1}{2}|p_i|^2$). At each step, the convex hull of $\{(p_i, h_i)\}_{i=1}^n$ is constructed. For OT (WT) map, the lower (upper) convex hull is projected to induce a nearest (farthest) weighted Delaunay triangulation $\mathcal{T}$ of $\{p_i\}$'s. Each vertex on the convex hull $v_i(\mathbf{h}) = (p_i, h_i)$ corresponds to a supporting plane $\pi_i(\mathbf{h}, x) = \langle p_i, x\rangle - h_i$, each face $[v_i, v_j, v_k]$ in the convex hull is dual to the vertex in the envelope, which is the intersection point of π_i, π_j and π_k. For the OT (WT) map, the lower (upper) convex hull is dual to the upper (lower) envelope. The projection of the envelopes induces the power diagrams. For the OT (WT) map, the upper (lower) envelope induces the nearest (farthest) power diagram.

Then we compute the μ-volume of each power cell using Eqn. (3.26), the gradient of the energy Eqn. (3.17) is given by Eqn. (3.20). The Hessian matrix $\mathrm{Hess}(E(\mathbf{h}))$ can be constructed using Eqn. (3.28) for off diagonal elements and Eqn. (3.29) for diagonal elements. The Hessian matrices of the OT and WT maps differ by a sign. Then we solve a linear system to find the update direction,

$$\mathrm{Hess}(E(\mathbf{h}))\mathbf{d} = \nabla E(\mathbf{h}). \tag{4.4}$$

Then we need to determine the step length λ, such that $\mathbf{h} - \lambda\mathbf{d}$ is still in the admissible solution space $\mathcal{H}$. First, we set λ to be $+1$ for OT map, -1 for WT map. Then compute the power diagram $\mathcal{D}(\mathbf{h} - \lambda\mathbf{d})$ using $\mathbf{h} - \lambda\mathbf{d}$. If some cells disappear in the power diagram, then it means $\mathbf{h} - \lambda\mathbf{d}$ falls out of the admissible space. Then we shrink λ by half, $\lambda \leftarrow \frac{1}{2}\lambda$, and reconstruct the power diagram $\mathbf{h} - \lambda\mathbf{d}$. We repeat this process, until we find an appropriate step length λ and update $\mathbf{h}$ as $\mathbf{h} - \lambda\mathbf{d}$. We repeat this procedure until the norm of the gradient of the energy is less than a prescribed threshold ε. Then the upper (lower) envelope is the Brenier potential for the OT (WT) map. Each nearest (farthest) power cell $W_i(\mathbf{h})$ is mapped to the corresponding sample point p_i, which is the desired OT (WT) map.

Algorithm 5: Optimal/Worst Transportation Map

```
Input: $(\Omega,\mu)$, $\{(p_i,\nu_i)\}_{i=1}^n$, OT or WT map
Output: The optimizer $\mathbf{h}$ the Brenier potential $u_\mathbf{h}$
Normalize $\{p_1,p_2,\dots,p_n\}$ to be inside $\Omega$ by translation and scaling;
Initialize $h_i=\pm\langle p_i,p_i\rangle/2$ for OT/WT;
while true do
    Compute the lower (upper) convex hull of $\{(p_i,h_i)\}_{i=1}^n$ for OT/WT map;
    Compute the upper (lower) envelope of the planes $\{\langle p_i,x\rangle-h_i\}_{i=1}^n$ for OT/WT map;
    Project the upper (lower) envelope to $\Omega$ to get a nearest (farthest) power diagram $\Omega=\bigcup_{i=1}^n W_i(\mathbf{h})$ for OT/WT map ;
    Compute the $\mu$-volume of each cell $w_i(\mathbf{h})=\mu(W_i(\mathbf{h}))$ using Eqn. (3.26);
    Compute the gradient of the energy $E(\mathbf{h})$, $\nabla E(\mathbf{h})=(w_i(\mathbf{h})-\nu_i)$;
    if $\|\nabla E(\mathbf{h})\|<\varepsilon$ then
        return h;
    end
    Compute the $\mu$-lengths of the power Voronoi edges $W_i(\mathbf{h})\cap W_j(\mathbf{h})\cap\Omega$ using Eqn. (3.27);
    Construct the Hessian matrix of the energy $E(\mathbf{h})$ for OT/WT map:
        $$\mathrm{Hess}(E(\mathbf{h})):=\frac{\partial^2E(\mathbf{h})}{\partial h_i\partial h_j}=\pm\frac{\mu(W_i(\mathbf{h})\cap W_j(\mathbf{h}))}{|y_i-y_j|}$$
    Solve linear system: $\mathrm{Hess}(E(\mathbf{h}))\mathbf{d}=\nabla E(\mathbf{h})$;
    $\lambda\leftarrow\pm1$ for OT/WT map;
    repeat
        Compute the nearest (farthest) power diagram $\mathcal{D}(\mathbf{h}-\lambda\mathbf{d})$ for OT/WT map;
        $\lambda\leftarrow\frac{1}{2}\lambda$;
    until no empty power cell;
    Update the height vector $\mathbf{h}\leftarrow\mathbf{h}-\lambda\mathbf{d}$;
end
```

4.3. Volumetric OT map algorithm

The algorithm for volumetric OT maps is very similar. The computation of power cell is based on cell clipping algorithm. Suppose at the current step, the power diagram is given by

$$\mathbb{R}^d=\bigcup_{i=1}^n W_i(\mathbf{h}), \tag{4.5}$$

where $W_i(\mathbf{h})$ is the projection of the supporting plane $\pi_i(\mathbf{h})$ on the graph of the Brenier potential $u_\mathbf{h}$, we need to clip $W_i(\mathbf{h})$ by Ω to obtain the cell

$W_i(\mathbf{h}) \cap \Omega$. The domain Ω is represented by a convex polyhedron, with supporting planes τ_j's, $j = 1, 2, \ldots, k$. Each plane τ_j divides $\mathbb{R}^d$ into an upper half-space and a lower half-space, Ω is the intersection of all the lower half-spaces. The power cell $W_i(\mathbf{h})$ is clipped by each supporting plane τ_j sequentially to obtain $W_i(\mathbf{h}) \cap \Omega$. At the j-th step, we test whether the current cell $W_i(\mathbf{h})$ is above the plane τ_j, if so, then the power cell $W_i(\mathbf{h})$ is empty; otherwise, if $W_i(\mathbf{h})$ is below the plane τ_j, then we skip the plane τ_j and proceed to the next plane; if $W_i(\mathbf{h})$ intersects τ_j, we cut off the component of $\tilde{W}_i(\mathbf{h})$ above τ_j.

5. Experimental results

In this section, we present our experimental results to show the accuracy, efficiency and robustness of the proposed methods, including the planar OT map, the planar WT map, and the volumetric OT map.

Setup All the algorithms are developed using generic C++ compatible with Windows and Linux platforms. The code uses half-edge data structure to represent meshes, Eigen [23] for numerical computation and OpenGL for interface. All the experiments are conducted on a Windows laptop with Intel Core i7-7700HQ CPU (planar and volume OT maps) or Intel Core i7-6700HQ CPU (planar WT map) using single thread and 16 GB memory.

5.1. Planar OT map

In order to show the *accuracy* and *efficiency* of the proposed algorithm, we compute the area preserving mappings from 3D models to the planar disk and compare with the method by Su et al. [43]. Each triangulated 3D model M in Fig. 11(a) is firstly conformally mapped to the unit disk $\mathbb{D}$ by the Ricci flow method [52], as shown in Fig. 11(b). It is obvious that the conformal mapping has large area distortions compared with the 3D surfaces. The conformal mapping is denoted by ϕ, and each vertex $v_i \in M$ is mapped to $p_i = \phi(v_i)$ on $\mathbb{D}$. Its corresponding discrete measure is given by $\nu_i = \frac{1}{3}\sum_{[v_i,v_j,v_k]\in M} area([v_i, v_j, v_k])$, where $[v_i, v_j, v_k]$ represents a face adjacent to v_i on M. After normalization, the summation of the discrete measures will be equal to the area of the planar disk, namely π. Then we compute the semi-discrete OT map T from the uniform distribution defined on the planar disk to the target measure $\nu = \sum_{i=1}^{n} \nu_i \delta(x - p_i)$. Finally, the composition $T^{-1} \circ \phi$ gives the discrete area-preserving map from the discrete mesh M to the planar disk $\mathbb{D}$, as shown in Fig. 11(c). (i) To show

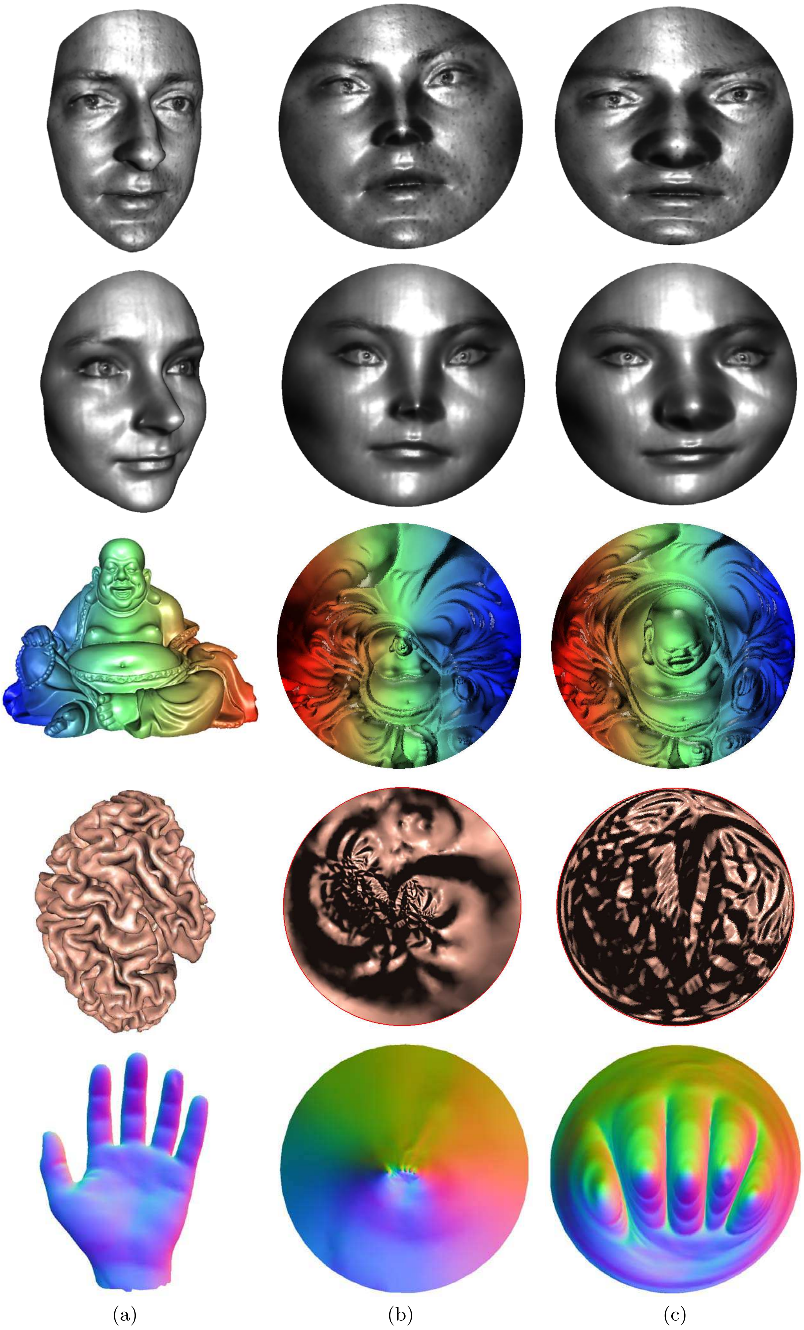


Figure 11: The conformal mapping (b) and the computed area-preserving mapping (c) for various 3D surface models (a).

Table 1: Running time (s) comparison

Model	Vertex	Method [43]	Our method	Speedup
Alex	14358	11.189	1.390	8.05
David	21466	16.656	2.380	7.00
Sophie	24891	22.462	2.373	9.47
Hand	3706	5.390	1.287	4.19
Buddha	18237	14.549	2.264	6.43
	32765	29.500	5.422	5.44
Brain	14971	14.960	2.592	5.77
	25909	24.892	5.047	4.93
Luke	17830	13.849	1.841	7.52
	24977	23.016	2.653	8.68

the efficiency of the proposed method, we set the stopping condition for Alg. 1 to be $\max_i \mu(W_i)/\nu_i < 10^{-8}$ and report the computing time for the proposed algorithm and Su et al. [43] in Tab. 1. For our method, the reduced area-distortion within each iteration is two to three times faster than [43], and the total iteration number required by the proposed method is three to four times less. Therefore, the new method is 6 to 10 times faster than that of [43]. (ii) To show the accuracy of the proposed method, we plot the histograms of $\log(\mu(W_i)/\nu_i), i = 1, 2 \ldots, n$ in different iterations and compare with [43] in Fig. 12 for the first model of Fig. 11. Upon convergence, we have $\mu(W_i) = \nu_i$, namely $\log(\mu(W_i)/\nu_i) = 0$ for all $i = 1, 2, \ldots, n$. It is obvious that our method (the second row of Fig. 12) converges much faster than Su et al. (the first row), and it finds the global minimum nearly after five iterations.

5.2. Planar WT map

The testing surfaces are either scanned from the real world, Fig. 5, Fig. 13, Fig. 14, or reconstructed from medical images, Fig. 15 with multiple resolutions. The surfaces are conformally mapped to the planar disk, and the push-forwarded surface area elements are treated as the source measure. The target measures are the uniform distribution on different planar domains.

Symmetry We verify our symmetry theorems Thm 3.19 and Thm 3.20 using the Buddha model in Fig. 14 and the brain model in Fig. 15. Fig. 14(c-d) and Fig. 15 (c-d) show the OT and WT map results to the unit disk with uniform distribution, visually they differ by the reflection in the origin.

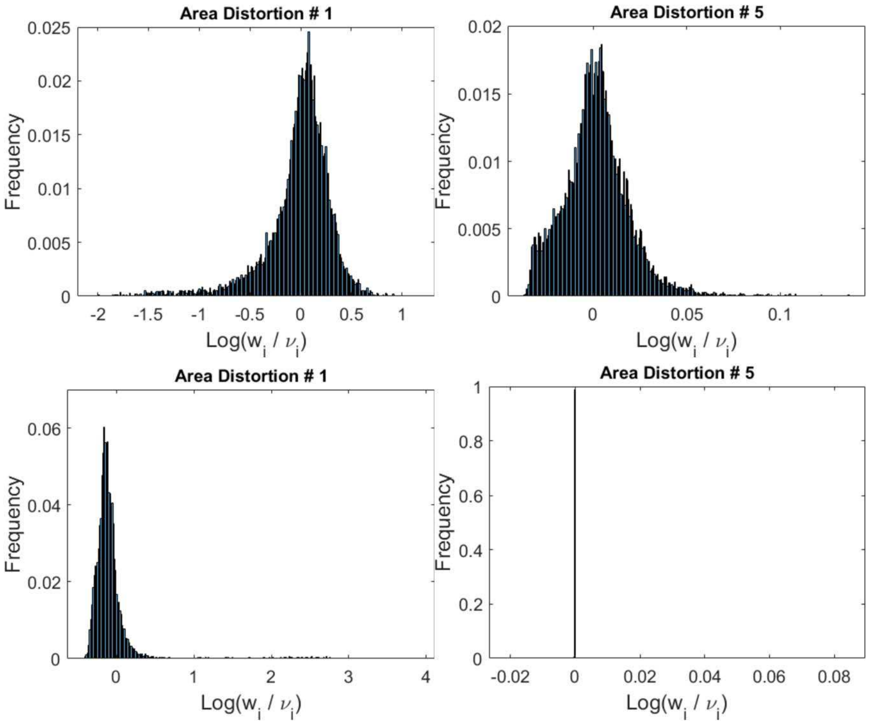


Figure 12: Histogram comparison in different iterations for Su et al. [43] (the first row) and the proposed method (the second row). The iterations are given in the title of the figures.

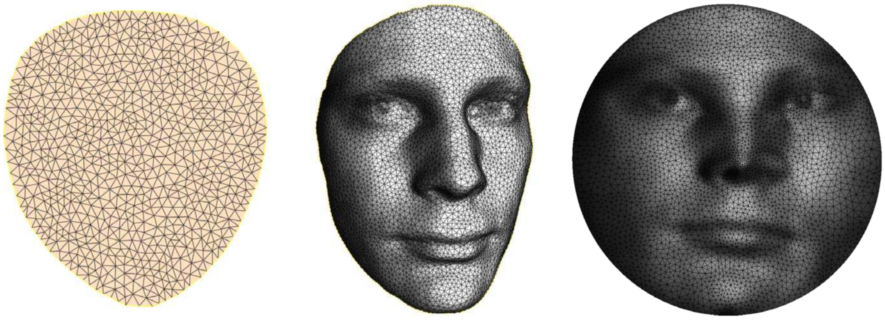

(a). (Ω, μ) (b). 3D surface (c). (Ω^*, ν)

Figure 13: The input probability measures.

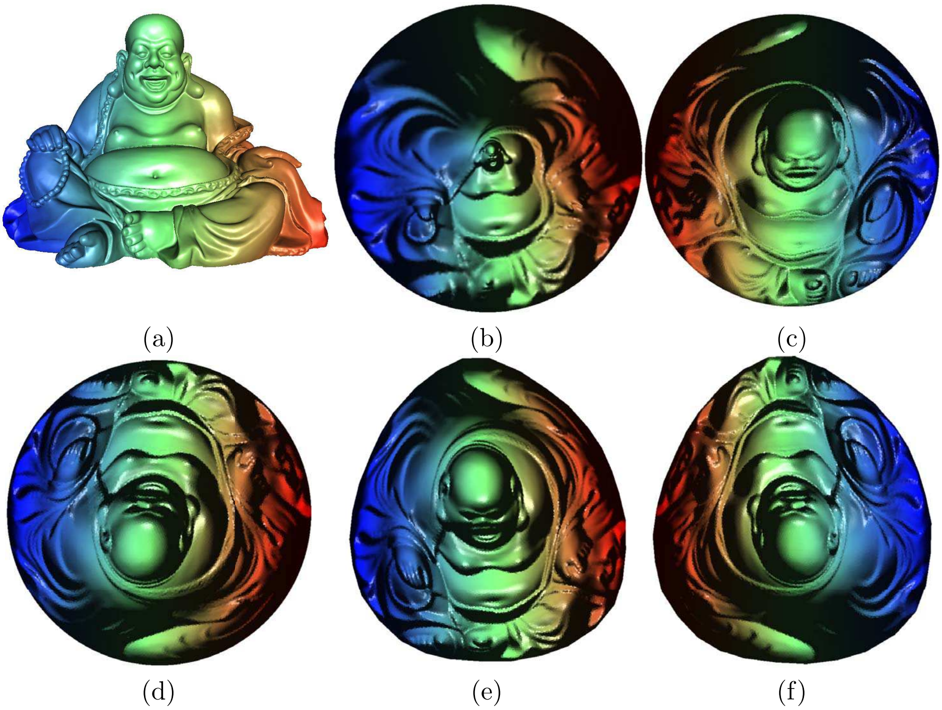


Figure 14: From top left to bottom right, the original Buddha surface (a), conformal mapping image (b), the OT and WT maps to the unit disk with uniform distribution (c) and (d), the OT and WT map to an irregular convex shape with uniform distribution (e) and (f).

We also quantitatively measure the symmetry. For each vertex v_i on the surface, we compute the Euclidean distance between its OT and WT images, $d_i := |T(v_i) - t(v_i)|$, then plot the histograms of the matching error $\{d_i\}$'s in Fig. 16 (a-b). It is obvious that the error is concentrated at the origin. If we examine frames (e) and (f), since the target measures are not symmetric, the OT and WT images are not symmetric anymore.

Accuracy In order to demonstrate the accuracy of the algorithm, for each triangular face on the input surface, we calculate its area on the surface and its area of the OT/WT image, then plot the histograms of the logarithms of the area ratios for OT/WT map. As shown in Fig. 16 (c-f), the histograms highly concentrated at the origin, this shows that the OT/WT mapping results are accurately measure-preserving.

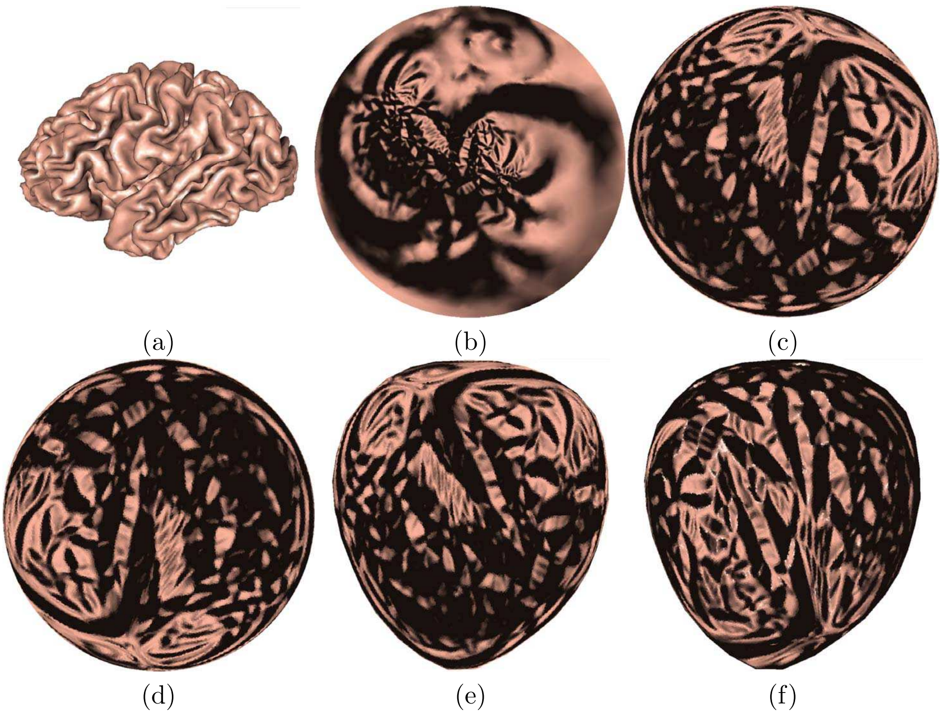
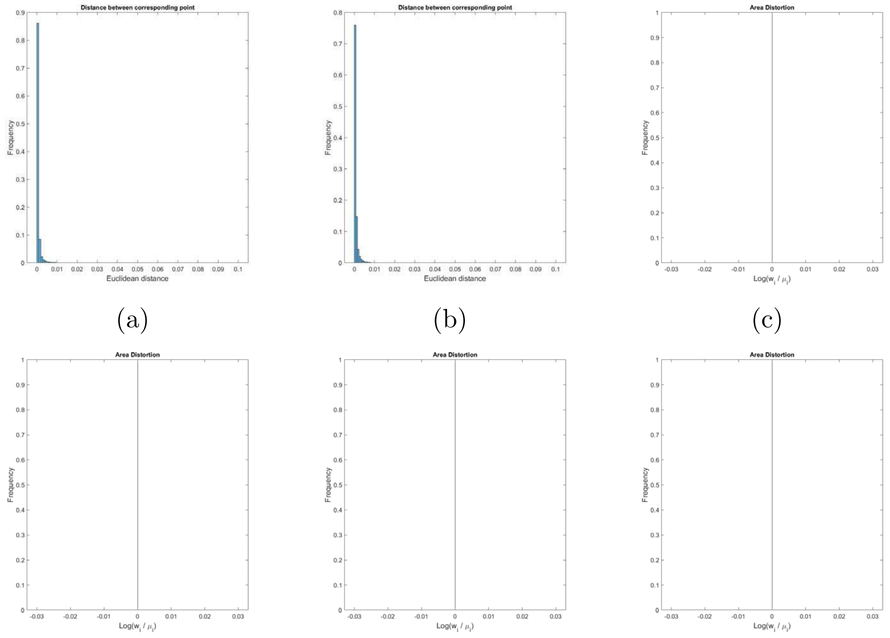


Figure 15: From top left to bottom right, the original Brain surface (a), conformal mapping image (b), the OT and WT maps to the unit disk with uniform distribution (c) and (d), the OT and WT map to an irregular convex shape with uniform distribution (e) and (f).

Efficiency The model complexities and running times are reported in Tab 2, we can see that the algorithm is highly efficient, for meshes with $32.7k$ vertices, the running time is less than 6 seconds.

5.3. Volumetric OT map

We tested our volumetric OT map algorithm on the geometric model shown in Fig. 17. The top row shows the cortical surface reconstructed from MRI images, the bottom row shows the tetrahedral mesh obtained using Tetgen [41]. The volumetric brain has been normalized such that its total volume equals to $\frac{4}{3}\pi$.



(a) (b) (c)

(d) (e) (f)

Figure 16: Statistical results, from top left to bottom right, symmetry verification for Buddha (a) and brain (b); area distortion histograms for Buddha OT (c) and WT (d), for brain OT (e) and WT (f).

Table 2: The processing time (in seconds) of OT and WT maps

Model	Vertex	OT	WT	Model	Vertex	OT	WT
Brain	3842	0.59	0.47	Budd.	3762	0.360	0.325
	11341	1.87	1.97		11038	1.17	1.43
	18564	3.48	2.81		18237	2.37	2.34
	22275	4.05	3.79		25473	3.99	4.24
	25909	5.03	5.14		32765	5.70	5.53
Alex	14358	1.63	1.643	Soph	24891	3.01	2.99
David	21466	2.58	2.82				

Fig. 18 shows the harmonic mapping results. The top row shows the harmonic map image of the cortical surface to the unit sphere using the algorithm in [20]. The bottom row shows the harmonic map image of the volumetric brain mesh to the unit ball with the cortical surface mapping result as the Dirichlet boundary condition. The volumetric harmonic map

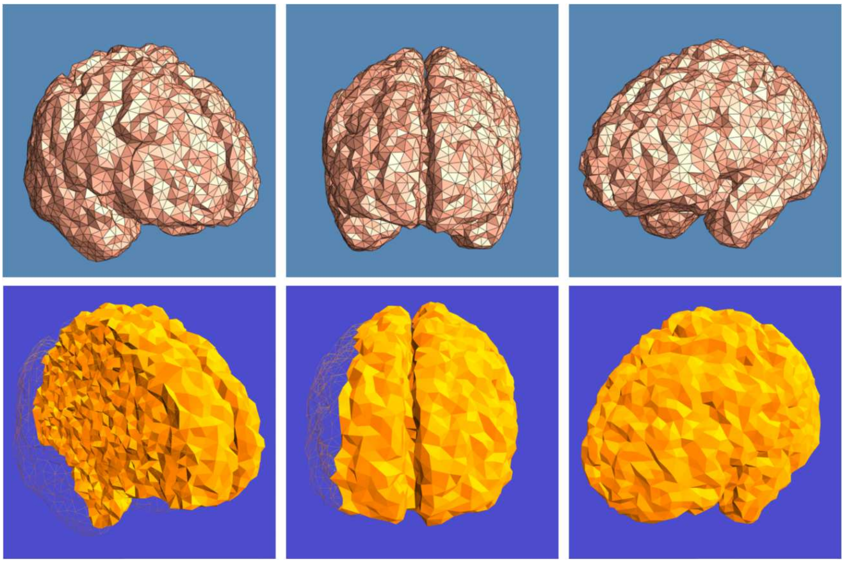


Figure 17: The cortical surface (top row) and the tetrahedral mesh (bottom row).

image of each vertex v_i defines a sample p_i, one third of the total volume of all the tetrahedra adjacent to v_i defines the target measure ν_i. The central row shows the volumetric vonoroi diagram.

The OT map result is shown in Fig. 19, the top row is the power diagram. Each sample p_i is mapped to the mass center of the power cell W_i, the OT map is approximated by the piece-wise linear map as shown in the bottom row.

Robustness During the optimization process, if we skip the admissibility test procedure and if the step length is large, the algorithm collapses; if the step length is chosen to be too small, the process converges but is very slow. With the admissibility test procedure, the algorithm is very stable and converges fast. This shows the proposed method greatly improves the robustness of finding OT maps.

6. Conclusions and future works

In this work, we summarize the geometric variational framework to solve OT maps in Euclidean spaces. We generalize the method to solve WT maps and discuss about the symmetry between the OT and WT maps. Computational geometric algorithms, such as Lawson edge flip, Sutherland-Hodgman

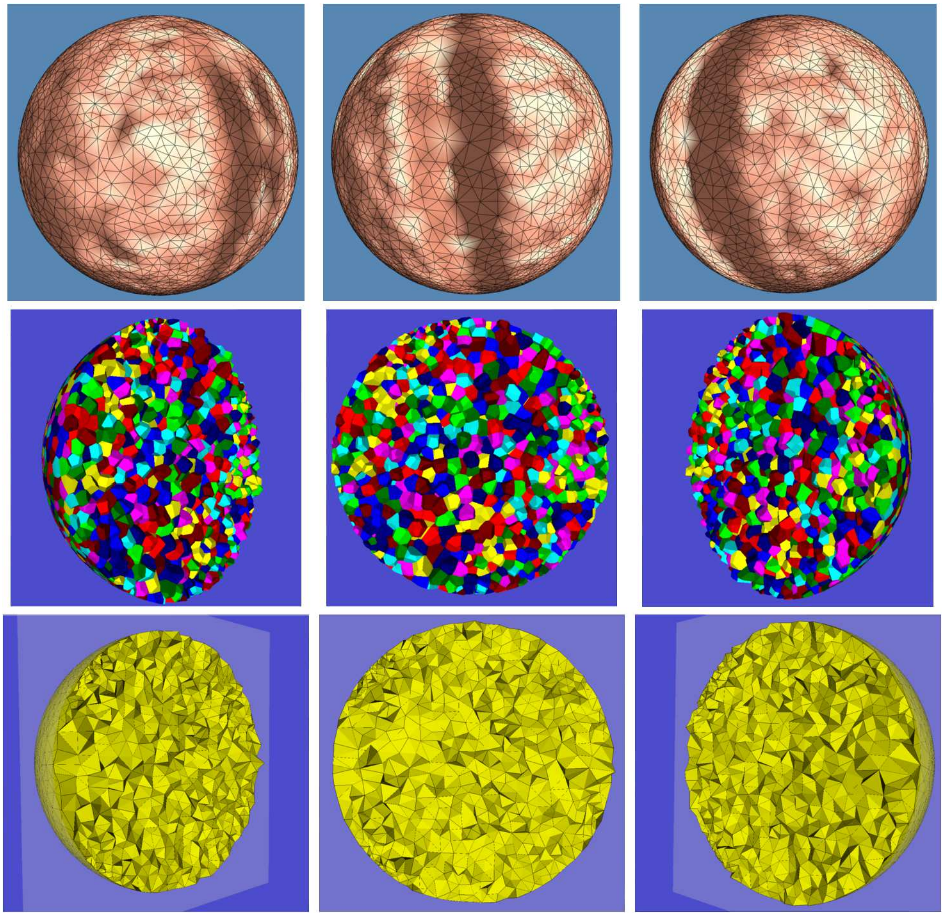


Figure 18: The cortical surface is harmonically mapped onto the unit sphere (top row) and the tetrahedral mesh is harmonically mapped onto the unit ball (bottom row), the corresponding volumetric Voronoi diagram is computed (central row).

clipping, sweep convex polygon, power diagram and the weighted Delaunay triangulation, are incorporated into the method to improve its accuracy, efficiency and robustness. The experimental results demonstrate the proposed framework is general, rigorous and practical.

In the future, we will generalize the framework to compute optimal/worst transportation maps on Riemannian manifolds.

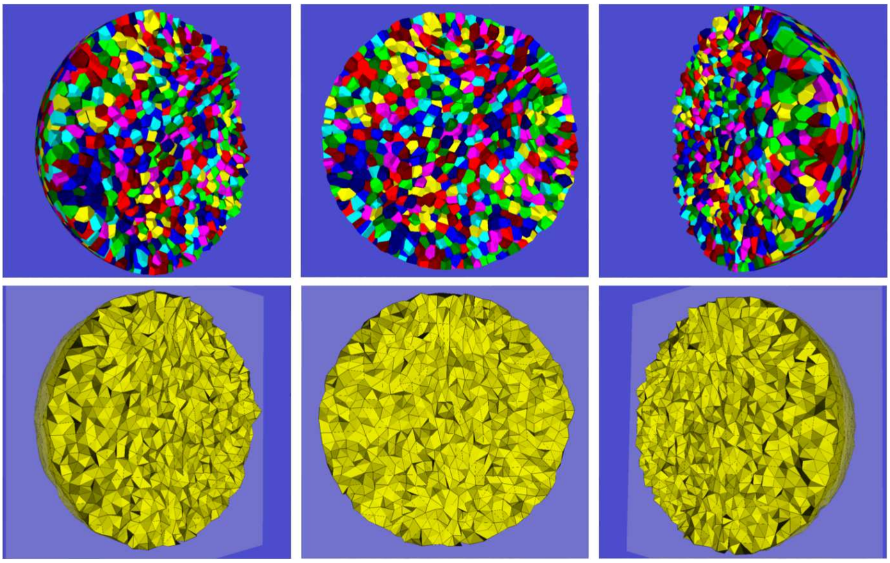


Figure 19: The OT map from the Lebesgue measure to the push-forward measure induced by the volumetric harmonic map, the top row shows the power diagram, the bottom row shows the piece-wise linear map.

References

[1] Alexandroff, A. (1939). Almost everywhere existence of the second differential of a convex function and some properties of convex surfaces connected with it. *Leningrad State Univ. Annals [Uchenye Zapiski] Math. Ser* **6** 3–35. MR0003051

[2] Alexandrov, A. D. (2005). *Convex polyhedra Translated from the 1950 Russian edition by N. S. Dairbekov, S. S. Kutateladze and A. B. Sossinsky*. Springer Monographs in Mathematics. MR2127379

[3] An, D., Xie, J. and Li, P. (2021). Learning Deep Generative Models by Short-run MCMC Inference with Optimal Transport Correction. In *Conference on Computer Vision and Pattern Recognition (CVPR).*

[4] An, D., Guo, Y., Lei, N., Luo, Z., Yau, S.-T. and Gu, X. (2020a). AE-OT: A new Generative Model based on extended semi-discrete optimal transport. In *International Conference on Learning Representations.*

[5] An, D., Guo, Y., Zhang, M., Qi, X., Lei, N. and Gu, X. (2020b). AE-OT-GAN: Training GANs from data specific latent distribution. In *ECCV*.

[6] An, D., Lei, N., Chen, W., Luo, Z., Zhao, T., Si, H. and Gu, X. (2021a). Efficient Approximation of Optimal Transportation Map by Pogorelov Map. In *Proceedings of International Meshing Roundtable*.

[7] An, D., Lei, N., Zhao, T., Si, H. and Gu, X. (2021b). A Moving Mesh Adaptation Method by Optimal Transport. In *Proceedings of International Meshing Roundtable*.

[8] Arjovsky, M., Chintala, S. and Bottou, L. (2017). Wasserstein generative adversarial networks. In *ICML* 214–223.

[9] Bentley, J. L. and Ottmann, T. A. (1979). Algorithms for reporting and counting geometric intersections. *IEEE Transactions on Computers* **C-28** 643 – 647.

[10] Berg, M. d., Cheong, O., Kreveld, M. v. and Overmars, M. (2008). *Computational Geometry: Algorithms and Applications*, 3rd ed. ed. Springer-Verlag TELOS, Santa Clara, CA, USA. MR2723879

[11] Brenier, Y. (1987). Polar decomposition and increasing rearrangement of vector fields. *C. R. Acad. Sci. Paris Sr. I Math.* **305** 805-808. MR0923203

[12] Brenier, Y. (1991). Polar factorization and monotone rearrangement of vector-valued functions. *Comm. Pure Appl. Math.* **44** 375-417. MR1100809

[13] Courty, N., Flamary, R., Tuia, D. and Rakotomamonjy, A. (2017). Optimal Transport for Domain Adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **39** 1853-1865.

[14] Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in Neural Information Processing Systems* 2292–2300.

[15] Delon, J. (2004). Midway image equalization. *J. Math. Imaging Vision*. MR2090128

[16] Edelsbrunner, H. and Shah, N. R. (1996). Incremental topological flipping works for regular triangulations. *Algorithmica* **15** 223–241. MR1368251

[17] Galichon, A. (2016). *Optimal Transport Methods in Economics.* Princeton University Press. MR3586373

[18] Genevay, A., Cuturi, M., Peyré, G. and Bach, F. (2016). Stochastic optimization for large-scale optimal transport. In *Advances in Neural Information Processing Systems* 3440–3448.

[19] Glimm, T. and Oliker, V. (2003). Optical design of single reflector systems and the Monge–Kantorovich mass transfer problem. *Journal of Mathematical Sciences* **117** 4096–4108. MR2027449

[20] Gu, X., Wang, Y., Chan, T. F., Thompson, P. M. and Yau, S.-T. (2004). Genus zero surface conformal mapping and its application to brain surface mapping. *IEEE Transactions on Medical Imaging (TMI)* **23** 949–958.

[21] Gu, D. X., Luo, F., Sun, J. and Yau, S.-T. (2016a). Variational principles for Minkowski type problems, discrete optimal transport, and discrete Monge-Ampère equations. *Asian Journal of Mathematics* **20** 383–398. MR3480024

[22] Gu, X., Luo, F., Sun, J. and Yau, S.-T. (2016b). Variational principles for Minkowski type problems, discrete optimal transport, and discrete Monge–Ampère equations. *Asian Journal of Mathematics* **20** 383–398. MR3480024

[23] Guennebaud, G., Jacob, B. et al. (2010). Eigen v3. http://eigen.tuxfamily.org.

[24] Haber, E., Rehman, T. and Tannenbaum, A. (2010). An efficient numerical method for the solution of the l2 optimal mass transfer problem. *SIAM J. Sci. Comput.* **32** 197–211. MR2599774

[25] Haker, S., Zhu, L., Tannenbaum, A. and Angenent, S. (2004). Optimal Mass Transport for Registration and Warping. *International Journal of Computer Vision* **60** 225–240.

[26] Kusner, M., Sun, Y., Kolkin, N. and Weinberger, K. (2015). From word embeddings to document distances. In *Proceedings of the 32nd International Conference on Machine Learning* 957–966.

[27] Lai, R. and Zhao, H. (2017). Multiscale Nonrigid Point Cloud Registration Using Rotation-Invariant Sliced-Wasserstein Distance via Laplace–Beltrami Eigenmap. *SIAM Journal on Imaging Sciences* **10** 449–483. MR3631386

[28] Lawson, C. L. (1972). Transforming triangulations. *Discrete Mathematics*. MR0311491

[29] Lei, N. and Gu, X. (2021). *Optimal Transportation Theory and Computation*. Higher Education Press.

[30] Lévy, Bruno (2015). A Numerical Algorithm for L2 Semi-Discrete Optimal Transport in 3D. *ESAIM: M2AN* **49** 1693-1715. MR3423272

[31] Lévy, B. and Schwindt, E. L. (2018). Notions of optimal transport theory and how to implement them on a computer. *Computers & Graphics* **72** 135–148.

[32] McRae, A. T. T., Cotter, C. J. and Budd, C. J. (2018). Optimal-Transport–Based Mesh Adaptivity on the Plane and Sphere Using Finite Elements. *SIAM Journal on Scientific Computing* **40** A1121-A1148. MR3790083

[33] Miyato, T., Kataoka, T., Koyama, M. and Yoshida, Y. (2018). Spectral normalization for generative adversarial networks. In *ICLR*.

[34] Nader, G. and Guennebaud, G. (2018). Instant Transport Maps on 2D Grids. *ACM Trans. Graph.* **37**.

[35] Nguyen, X. (2013). Convergence of latent mixing measures in finite and infinite mixture models. *Ann. Statist* **41** 370–400. MR3059422

[36] Peyré, G. and Cuturi, M. (2018). *Computational Optimal Transport*. https://arxiv.org/abs/1803.00567.

[37] Pitie, F., Kokaram, A. C. and Dahyot, R. (2007). Automated colour grading using colour distribution transfer. *Computer Vision and Image Understanding* **107** 123–137.

[38] Rabin, J., Ferradans, S. and Papadakis, N. (2014). Adaptive color transfer with relaxed optimal transport.

[39] Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., Lee, L., Chen, J., Brumbaugh, J., Rigollet, P., Hochedlinger, K., Jaenisch, R., Regev, A. and Lander, E. (2019). Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell* **176** 928–943.

[40] Seguy, V., Damodaran, B. B., Flamary, R., Courty, N., Rolet, A. and Blondel, M. (2018). Large-scale optimal transport and mapping estimation. *Stat* **1050** 26.

[41] Si, H. (2015). TetGen, a Delaunay-Based Quality Tetrahedral Mesh Generator. *ACM Trans. Math. Softw.* **41** 11:1–11:36. MR3318083

[42] Su, Z., Zeng, W., Shi, R., Wang, Y., Sun, J. and Gu, X. (2013). Area preserving brain mapping. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 2235–2242.

[43] Su, Z., Wang, Y., Shi, R., Zeng, W., Sun, J., Luo, F. and Gu, X. (2015). Optimal Mass Transport for Shape Matching and Comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **37** 2246–2259.

[44] Su, K., Cui, L., Qian, K., Lei, N., Zhang, J., Zhang, M. and Gu, X. (2016). Area-preserving mesh parameterization for poly-annulus surfaces based on optimal mass transportation. *Computer Aided Geometric Design* **46** 76–91. MR3534084

[45] Sutherland, I. and Hodgman, G. W. (1974). Reentrant Polygon Clipping. *Communications of the ACM* **17** 32–42.

[46] Tolstikhin, I., Bousquet, O., Gelly, S. and Schoelkopf, B. (2018). Wasserstein Auto-Encoders. In *ICLR*.

[47] Villani, C. (2003). *Topics in Optimal Transportation* **58**. American Mathematical Society. MR1964483

[48] Villani, C. (2008). *Optimal transport: old and new* **338**. Springer Science & Business Media. MR2459454

[49] Wang, X.-J. (1996). On the design of a reflector antenna. *Inverse Problems* **12** 351. MR1391544

[50] Wang, X.-J. (2004). On the design of a reflector antenna II. *Calculus of Variations and Partial Differential Equations* **20** 329–341. MR2062947

[51] Yurochkin, M., Claici, S., Chien, E., Mirzazadeh, F. and Solomon, J. M. (2019). Hierarchical Optimal Transport for Document Representation. In *Advances in Neural Information Processing Systems 32*.

[52] Zeng, W., Samaras, D. and Gu, D. (2010). Ricci Flow for 3D Shape Analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **32** 662–677.

[53] Zhang, M., An, D., Lei, N., Wu, J., Zhao, T., Xu, X., Wang, Y. and Gu, X. (2021). Cortical Morphometry Analysis based on Worst Transportation Theory. In *International Conference on Information Processing in Medical Imaging* 163–176. Springer, Cham.

[54] Zhao, X., Su, Z., Gu, X. D., Kaufman, A., Sun, J., Gao, J. and Luo, F. (2013). Area-preservation mapping using optimal mass transport. *IEEE Transactions on Visualization and Computer Graphics* **19** 2838–2847.

Dongsheng An
State University of New York at Stony Brook
USA
E-mail address: `doan@cs.stonybrook.edu`

Na Lei
Dalian University of Technology
China
E-mail address: `nalei@dlut.edu.cn`

Li Cui
Beijing Normal University
China
E-mail address: `licui@bnu.edu.cn`

Kehua Su
Wuhan University
China
E-mail address: `skh@whu.edu.cn`

Xiaoyin Xu
Brigham and Women's Hospital, Harvard Medical School
USA
E-mail address: `xxu@bwh.harvard.edu`

Feng Luo
Rutgers University
USA
E-mail address: `fluo@math.rutgers.edu`

Xianfeng Gu
State University of New York at Stony Brook
USA
E-mail address: `gu@cs.stonybrook.edu`

Shing-Tung Yau
Tsinghua University
China
Harvard University
USA
E-mail address: yau@math.harvard.edu

Received March 5, 2021