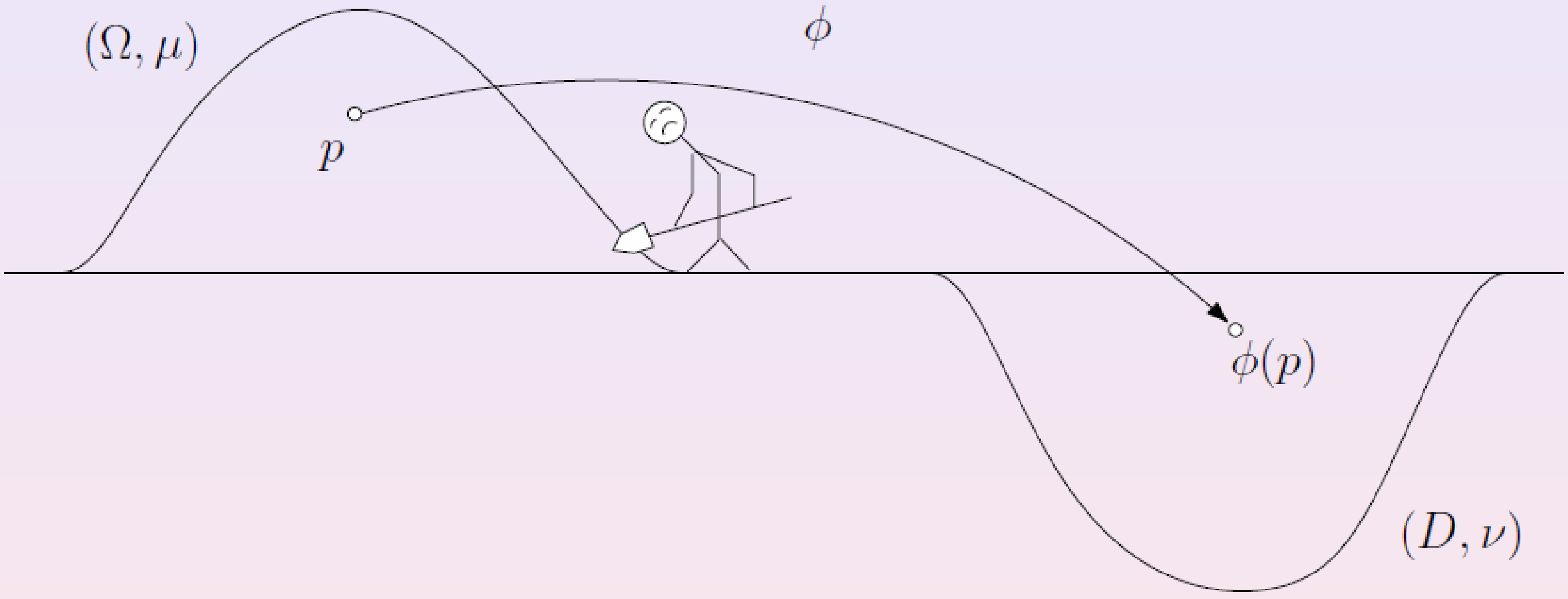
Tutorial on Optimal Mass Transprot for Computer Vision

David Gu
State University of New York at Stony Brook

Graduate Summer School: Computer Vision, IPAM, UCLA July 31, 2013

Joint work with Shing-Tung Yau, Feng Luo and Jian Sun

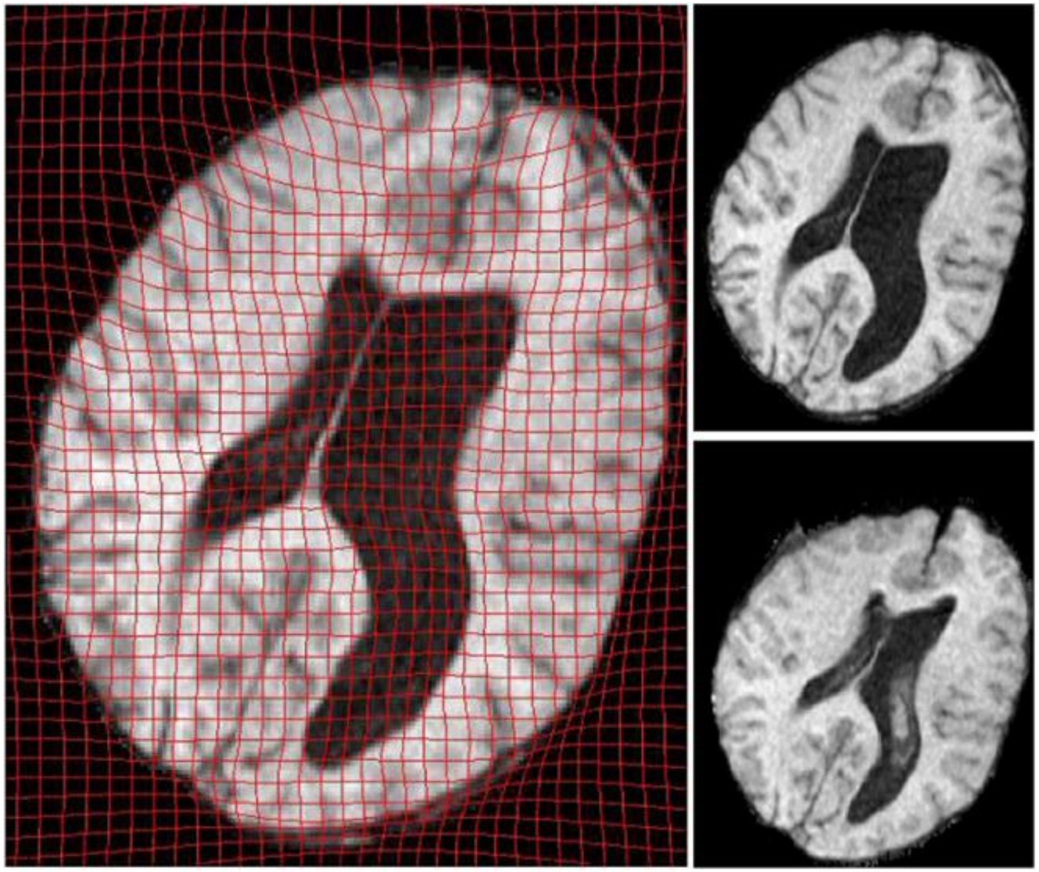
Optimal Mass Transportation Problem



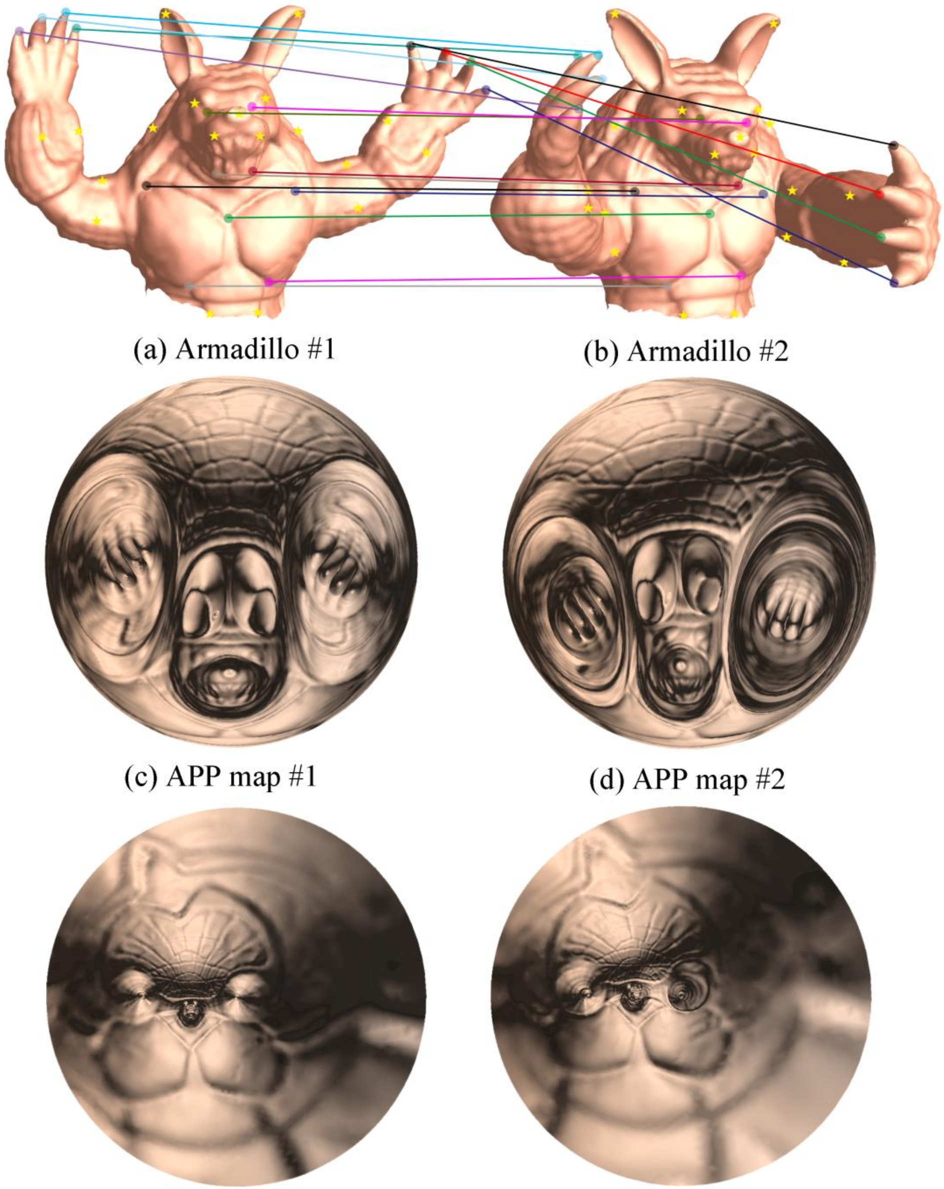
Earth movement cost.

Motivation

- Tannenbaum: Medical image registration



Motivation: Surface registration



(a) Armadillo #1

(b) Armadillo #2

(c) APP map #1

(d) APP map #2

(e) Conformal map #1

(f) Conformal map #2

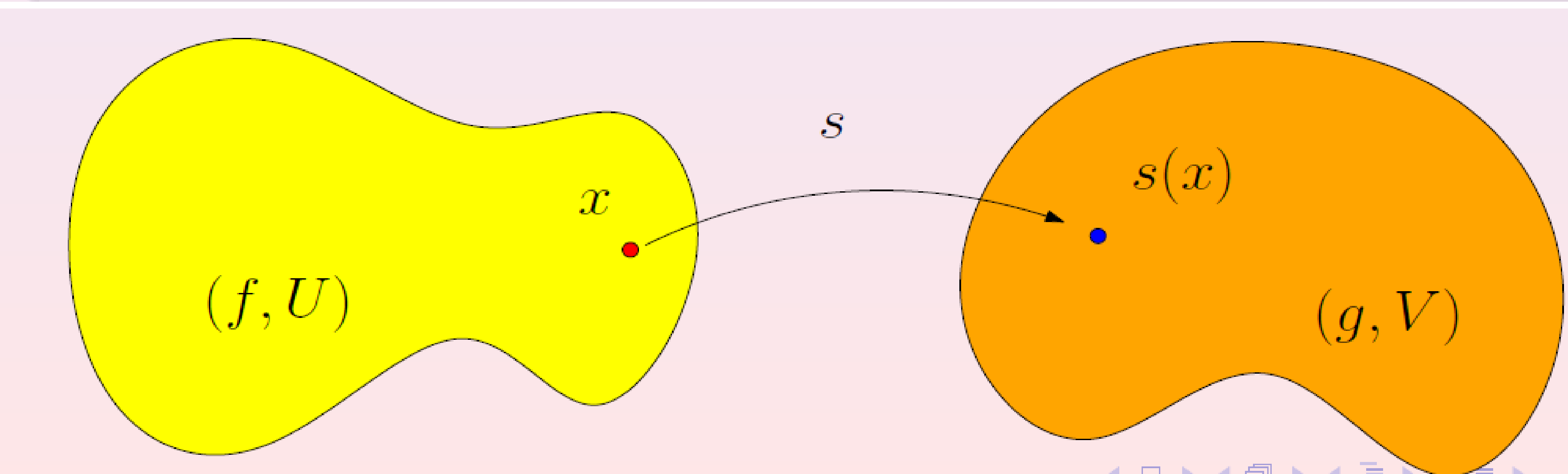
Optimal Mass Transportation Problem

Find a best scheme of transporting one mass distribution (f, U) to another one (g, V) such that the total cost is minimized,

$$U, V: \ \textit{two bounded domains in } \mathbb{R}^n$$

$$\begin{aligned} 0 &\leq f \in L^1(U) \\ 0 &\leq g \in L^1(V) \end{aligned}$$

$$\int_U f = \int_V g.$$



Optimal Mass Transportation Problem

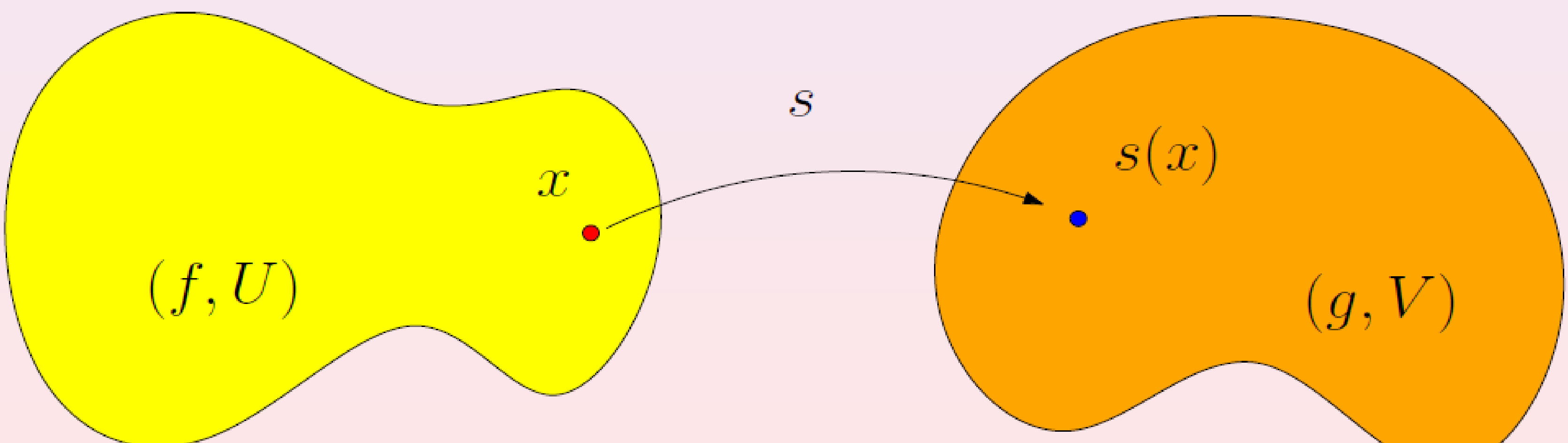
For a transport scheme s (a mapping from U to V)

$$s : x \in U \to y = s(x) \in V$$

The *total cost* is

$$\mathscr{C}(s) = \int_U f(x)c(x,s(x))dx$$

where $c(x,y)$ is the cost function.



Cost Functions

The cost of moving a unit mass from point x to point y.

$$Monge(1781) : c(x,y) = |x-y|.$$

This is the natural cost function. Other cost functions include

$$\begin{aligned} c(x,y) &= |x-y|^p, p \neq 0 \\ c(x,y) &= -\log|x-y| \\ c(x,y) &= \sqrt{\varepsilon + |x-y|^2}, \varepsilon > 0 \end{aligned}$$

Any function can be cost function. It can be negative.

Optimal Mass Transportation Problem

Problem

Is there an optima mapping $T : U \to V$ such that the total cost $\mathscr{C}$ is minimized,

$$\mathscr{C}(T) = \inf\{\mathscr{C}(s) : s \in \mathscr{S}\}$$

where $\mathscr{S}$ is the set of all measure preserving mappings, namely $s : U \to V$ satisfies

$$\int_{s^{-1}(E)} f(x)dx = \int_E g(y)dy, \forall\ Borel\ set\ E \subset V$$

Applications

- Economy: producer-consumer problem, gas station with capacity constraint,
- Probability: Wasserstein distance
- Image processing: image registration
- Digital geometry processing: surface registration

Duality and potential functions

A breakthrough in the study of optimal transportation is the introduction of the duality.

- Kantorovich introduced the dual functional

$$I(\varphi,\psi)=\int_U \varphi(x)f(x)+\int_V \psi(y)g(y), (\varphi,\psi)\in K$$

$$K=\{(\varphi,\psi):\varphi(x)+\psi(y)\le c(x,y)\}$$

- Duality:

$$\inf\{\mathscr{C}(s):s\in\mathscr{S}\}=\sup\{I(\varphi,\psi):(\varphi,\psi)\in K\}.$$

- Kantorovich won Nobel Prize for this work.

Duality and potential functions

- The dual functional I is linear, the set of K is convex.
- $\exists$ a maximizer (u, v):

$$I(u, v) = \sup\{I(\varphi, \psi) : (\varphi, \psi) \in K\}$$

- The maximizer is unique in the sense

$$I(u + a, v - a) = I(u, v)$$

  for any constant a.
- The maximizer (u, v) is called potential functions.

Brenier's Approach

Theorem (Brenier)

If $f, g > 0$ and U is convex, and the cost function is quadratic distance,

$$c(x, y) = |x - y|^2$$

then there exists a convex function $u : U \to \mathbb{R}$ unique upto a constant, such that the unique optimal transportation map is given by the gradient map

$$T : x \to \nabla u(x).$$

Brenier's Approach

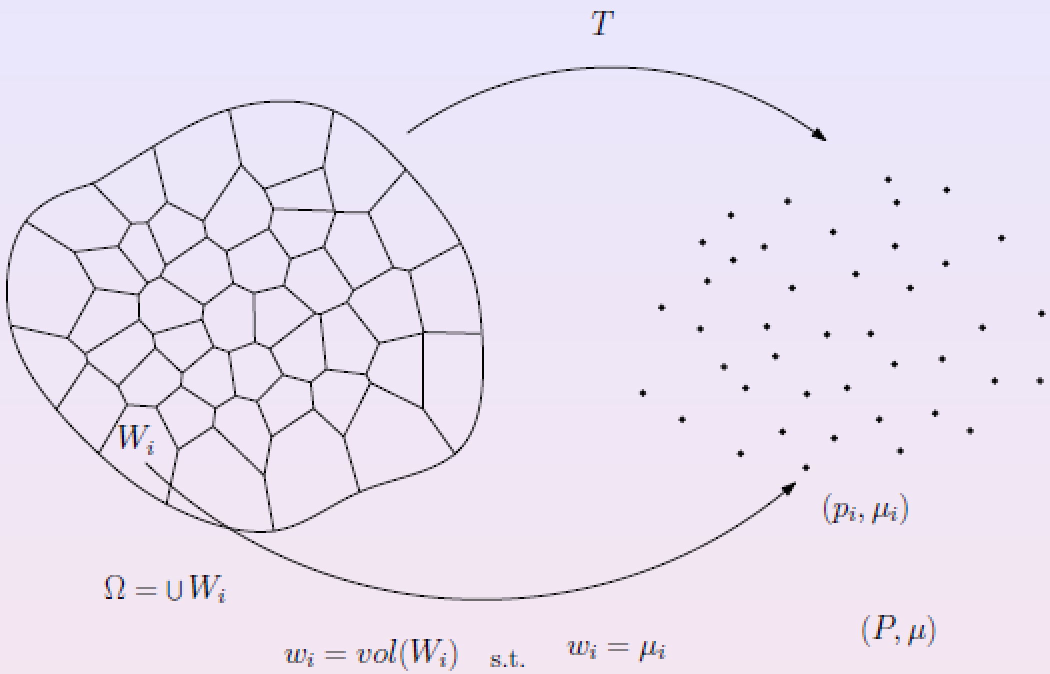
In smooth case, the Brenier potential $u : \Omega \to \mathbb{R}$ statisfies the Monge-Ampere equation

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

and $\nabla u : \Omega \to D$ minimizes the quadratic cost

$$\min_u \int_\Omega |x - \nabla u(x)|^2 dx.$$

Brenier's Approach

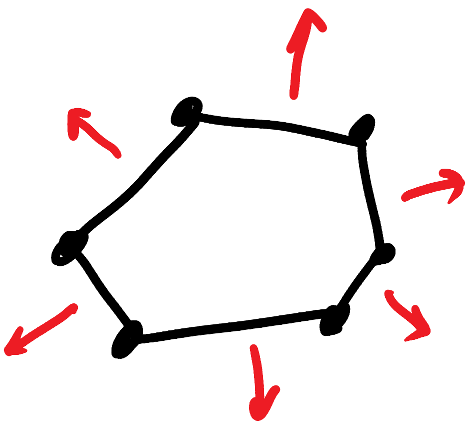


Discretize the target D to $P = \{(p_1, w_1), (p_2, w_2), \cdots, (p_n, w_n)\}$, Decompose Ω to cells $\{C_1, C_2, \cdots, C_n\}$, such that $T(C_k) = p_k$, $vol(C_k) = w_k$, and the mapping minimizes the quadratic cost

$$\sum_k \int_{C_k} |p - p_k|^2 dx.$$

Minkowski problem and several related problems

Eg. A convex polygon P in R^2 is determined by its edge lengths A_i and unit normal vectors n_i.



Take any $u \in R^2$ and project P to u,

$$\sum A_i \, n_i \cdot u = 0,$$

$$\sum A_i n_i = 0.$$

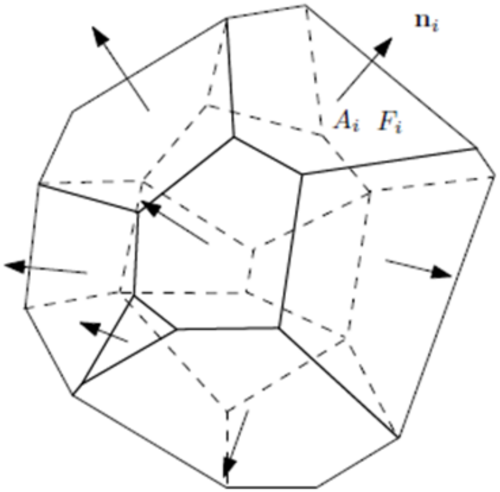
Minkowski Problem. Given k unit vectors $n_1, \ldots n_k$ not contained in a half-space in R^N and $A_1, \ldots, A_k > 0$ s.t.,

$$\sum_i A_i n_i = 0,$$

find a cpt convex polytope P with exactly k codim-1 faces $F_1, \ldots, F_k$ s.t.,

(a) area $(F_i) = A_i$ and

(b) $n_i \perp F_i$.



THM (Minkoswki) P exists and is unique up to translations.

Minkowski's proof is variational and constructs P.

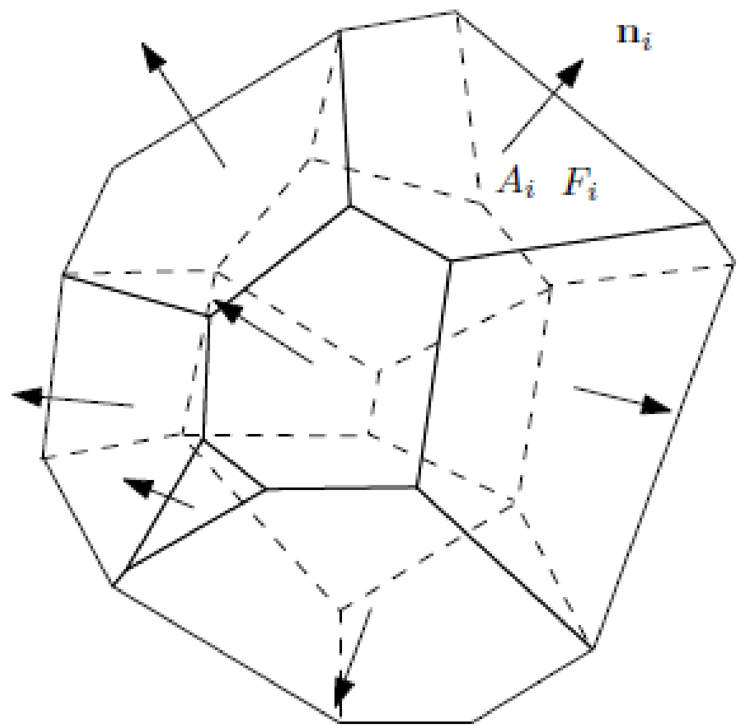
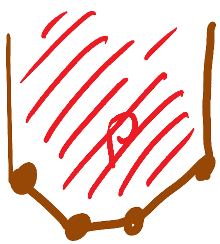


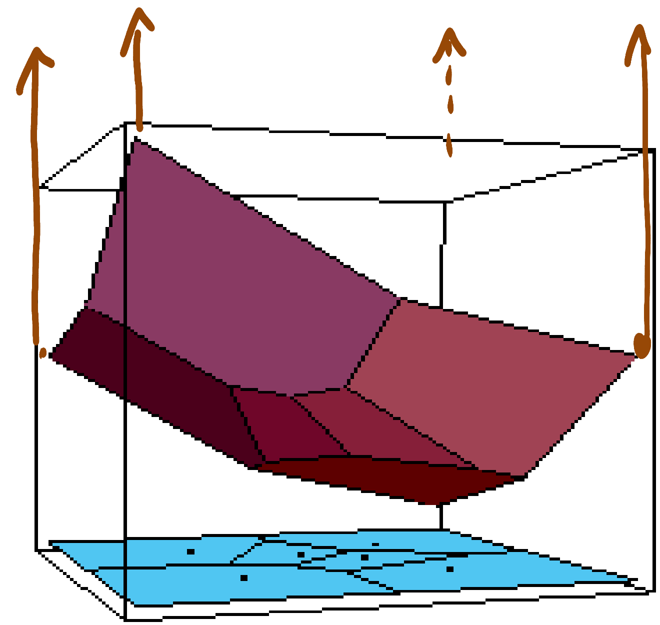
Figure 1: Minkowski problem

Q1. What is Minkowski problem for non-compact polyhedra?

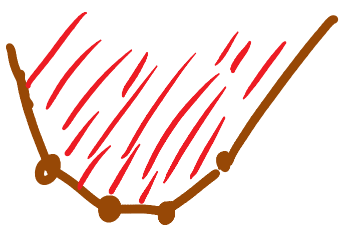
P.S. Alexandrov:



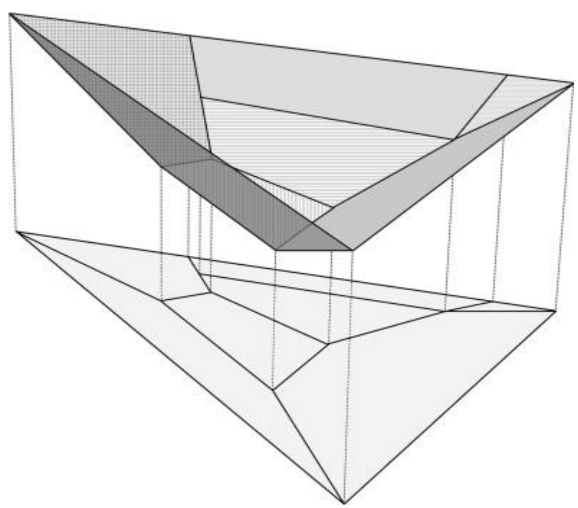
Polyhedron P



Discrete optimal transport

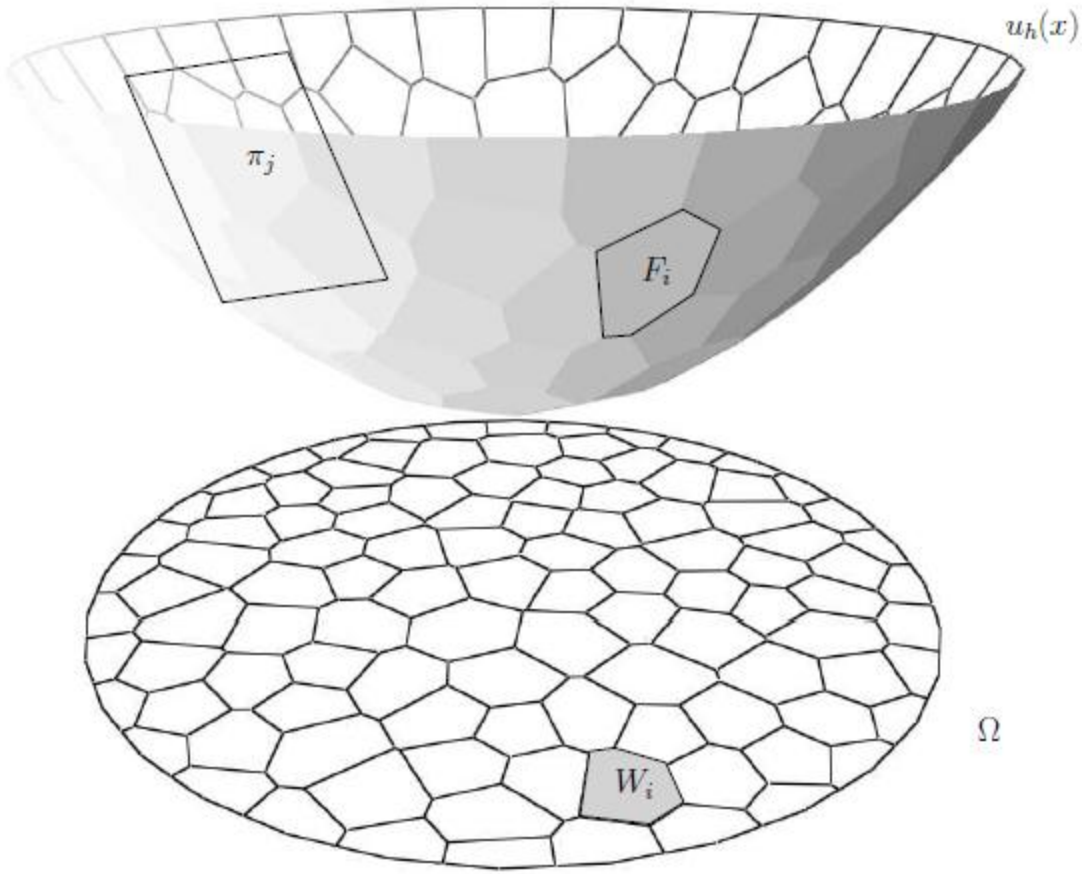


Pogorelov:



Discrete Monge-Ampere equation

Their results: MP solvable for bound faces with unbounded faces fixed.



PL-convex function and its induced convex subdivision.

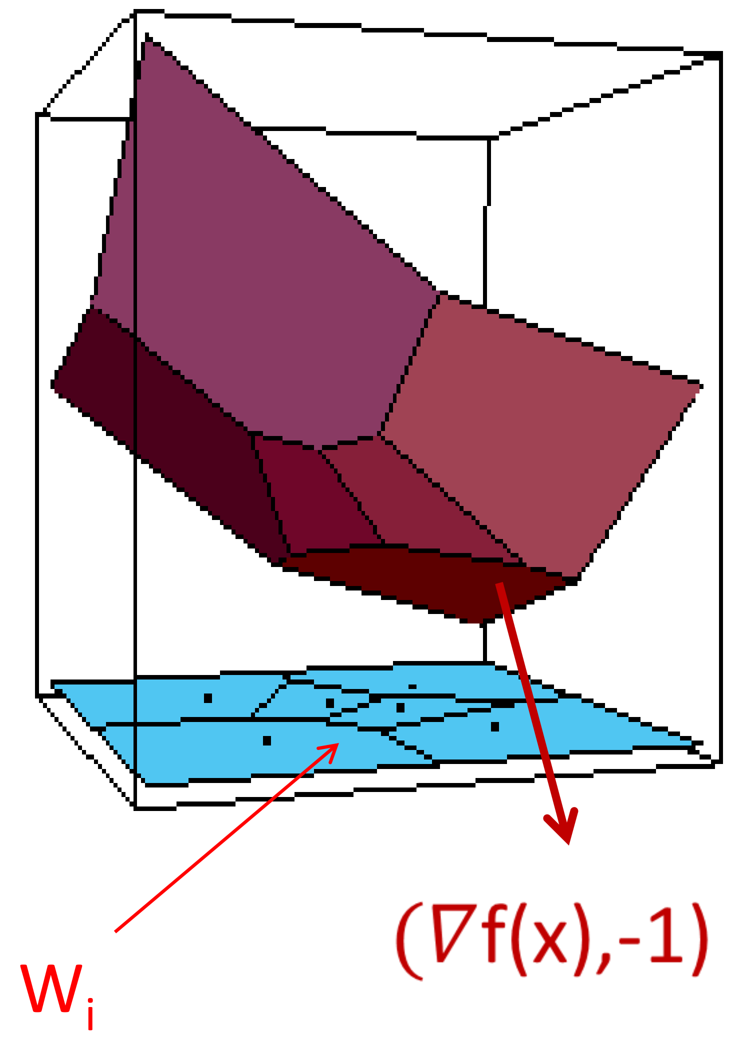
PL convex function

$f(x) = \max\{ x \cdot p_i + h_i \mid i=1,\ldots,k\}$

produces a convex cell decomposition W_i of R^N:

$W_i = \{x \mid x \cdot p_i + h_i \geq x \cdot p_j + h_j, \text{ all } j\}$

$= \{ x \mid \nabla f(x) = p_i\}$



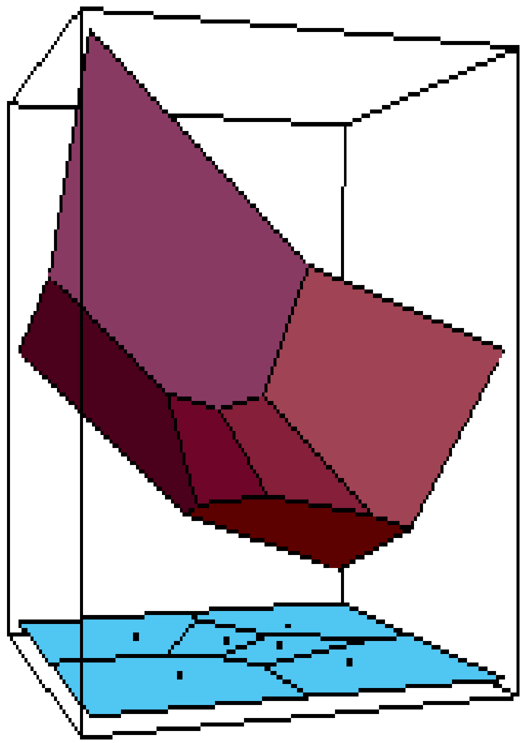
Alexandrov (1950): Given X cpt convex domain in R^N, $p_1, \ldots, p_k$ distinct in R^N, $A_1,\ldots,A_k>0$, s.t. $\sum A_i = \mathrm{vol}(X)$,

$\exists$ PL convex function

$$f(x) = \max_i\{ x \cdot p_i + h_i\},$$

unique up to translation s.t.,

$$\mathrm{Vol}(\{x \in X \mid \nabla f(x) = p_i\}) = A_i.$$



$We\ call\ \nabla f$ the Alexandrov map.

Alexandrov's proof is not variational and is topological.
On page 321 of his book "Convex polyhedra", he asked if there exists a variational proof of his thm. He said such a proof "is of prime importance by itself".

Pogorelov theorem

Suppose $v_1, \ldots, v_m$ in R^N, s.t.

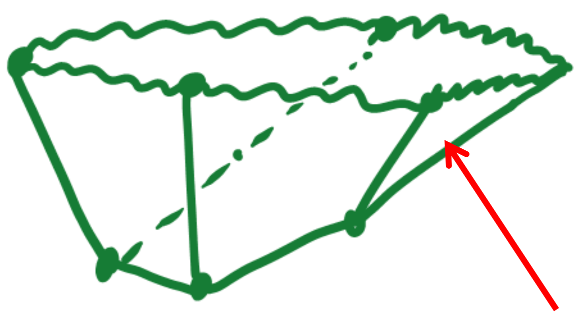
v_i not in $con\{v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_m\}$, and $g_1, \ldots, g_m$ in R.

$\forall \{p_1, \ldots, p_k\} \subset int(conv\{v_1, \ldots, v_m\})$ and $A_1, \ldots, A_K > 0$,

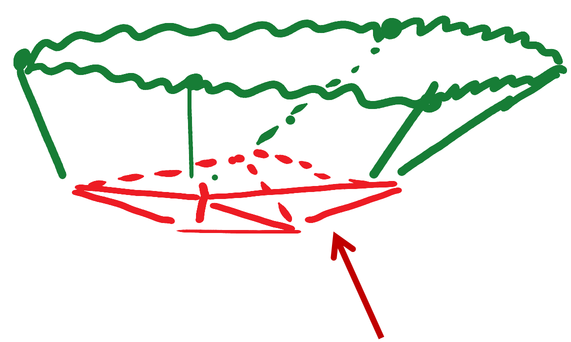
$\exists ! \, h_1, \ldots, h_k$, s.t. the PL convex function

$$f(x) = \max\{\max\{x \cdot p_i + h_i\}, \max\{x \cdot v_j + g_j\}\}$$

satisfies, $\mathrm{vol}\{x \mid \nabla f = p_i\} = A_i$.



$\max\{x \cdot v_j + g_j\}$ unbounded faces



$\max\{x \cdot p_i + h_i\}$, bounded faces

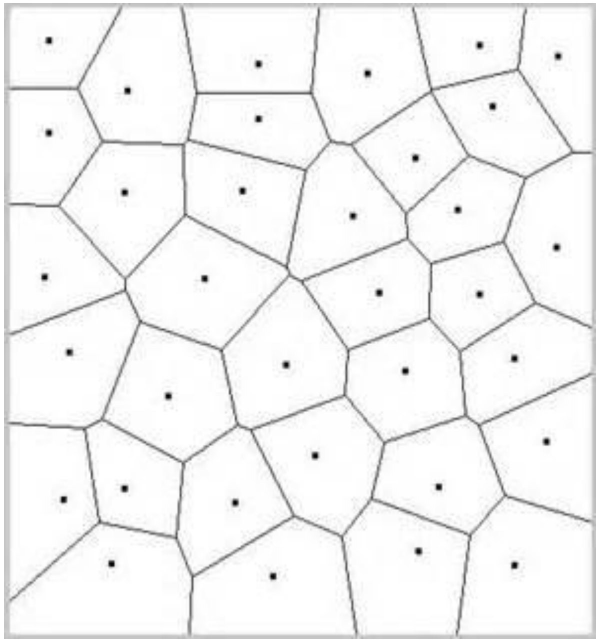
Our main result: there exist variational proofs of Alexandrov's and Pogorelov's theorems.

Basically the same as Minkowski's original proof.

Thus, there exists an algorithm to compute the Alexandrov map ∇f .

We are motivated by computational problems from computer graphics, discrete optimal transportation and discrete Monge-Ampere equation.

Voronoi decomposition and power diagrams

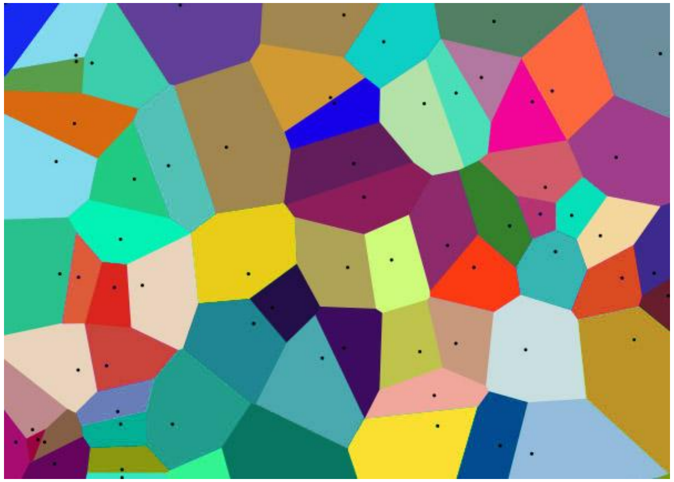


Given $p_1, \ldots, p_k$ in R^N, the Voronoi cell V_i at p_i is:

$$V_i=\{x \mid |x-p_i|^2 \leq |x-p_j|^2, \text{ all } j\}$$

A generalization: power diagram, given $p_1, \ldots, p_k$ in R^N and weights $a_1,\ldots,a_k$ in R, the power diagram at p_i is

$$W_i=\{x \mid |x-p_i|^2+a_i \leq |x-p_j|^2+a_j, \text{ all } j\}$$



PL convex function $f(x)=\max\{x \cdot p_i + h_i\}$ and power diagram

Lemma 1. If $f(x) = max\{x \cdot p_i + h_i\}$, *then* $W_i = \{x \mid \nabla f = p_i\}$ *is a power diagram.*

Proof. By definition $\{x \mid \nabla f = p_i\} = \{x \mid x \cdot p_i + h_i \geq x \cdot p_j + h_j, \text{ all } j\}$

$x \cdot p_i + h_i \geq x \cdot p_j + h_j$ is the same as

$x \cdot x - 2x \cdot p_i + p_i \cdot p_i - 2h_i - p_i \cdot p_i \leq x \cdot x - 2x \cdot p_j + p_j \cdot p_j - 2h_j - p_j \cdot p_j$,

i.e., $|x_i - p_i|^2 - 2h_i - p_i \cdot p_i \leq |x - p_j|^2 - 2h_j - p_j \cdot p_j$ for all j

Lemma 2. If $b_1, \ldots, b_k : X \to [0, \infty)$, let $W_i = \{x \mid b_i(x) \geq b_j(x), \text{ all } j\}$, Then for any partition $\{X_1, \ldots, X_k\}$ of X,

$$\sum_i b_i(x)\, \chi_{X_i}(x) \leq \sum_i b_i(x)\, \chi_{W_i}(x)$$

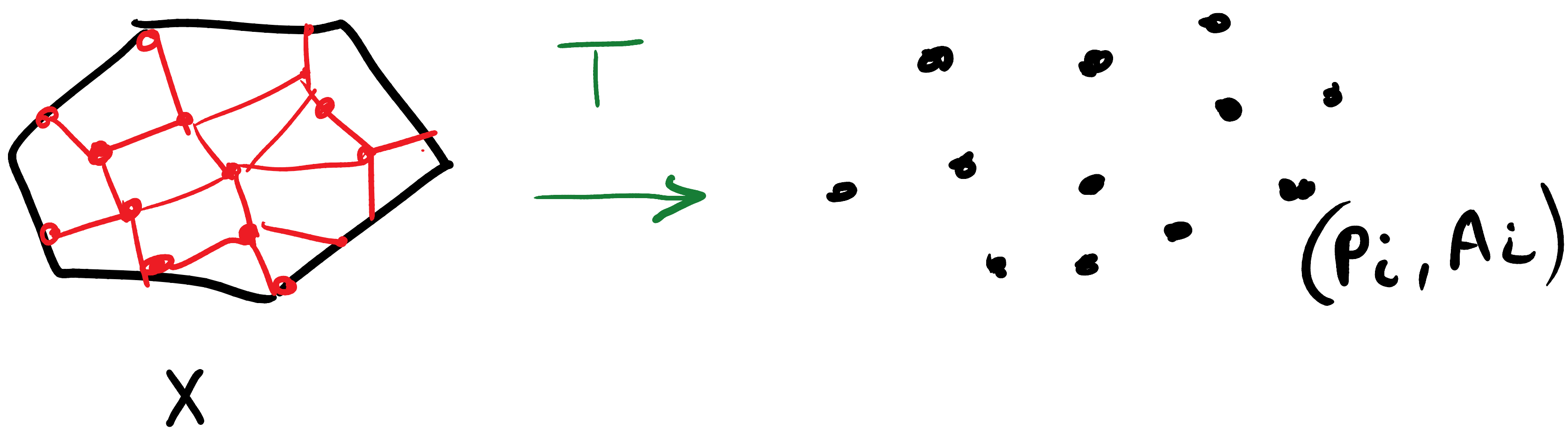
Proof. Take x in X, say x in X_j and also in W_i. Then

LHS= $b_j(x)$

RHS $\geq b_i(x) \geq b_j(x)$ =LHS.

Discrete optimal transport problem (Monge)

Given a compact convex domain X In R^N and
$p_1, \ldots, p_k$ in R^N and $A_1, \ldots, A_k > 0$,
find a transport map $T: X \to \{p_1, \ldots, p_k\}$ with
$vol(T^{-1}(p_i)) = A_i$ so that T minimizes
the cost $\int_X |x - T(x)|^2 \, dx.$ (Y. Brenier)



Recall

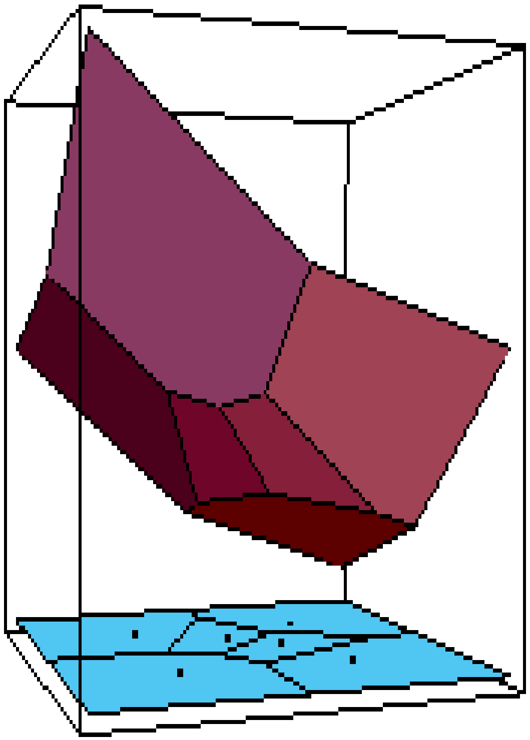
Alexandrov thm: Given X cpt convex domain in R^N, p_1, …, p_k distinct in R^N, $A_1,\ldots,A_k>0$ s.t. $\sum A_i = \mathrm{vol}(X)$.

Then ∃ PL convex function

$$f(x) = \max_i\{ x \cdot p_i + h_i\},$$

unique up to translation s.t.,

$$\mathrm{Vol}(\{x \in X \mid \nabla f(x) = p_i\}) = A_i.$$



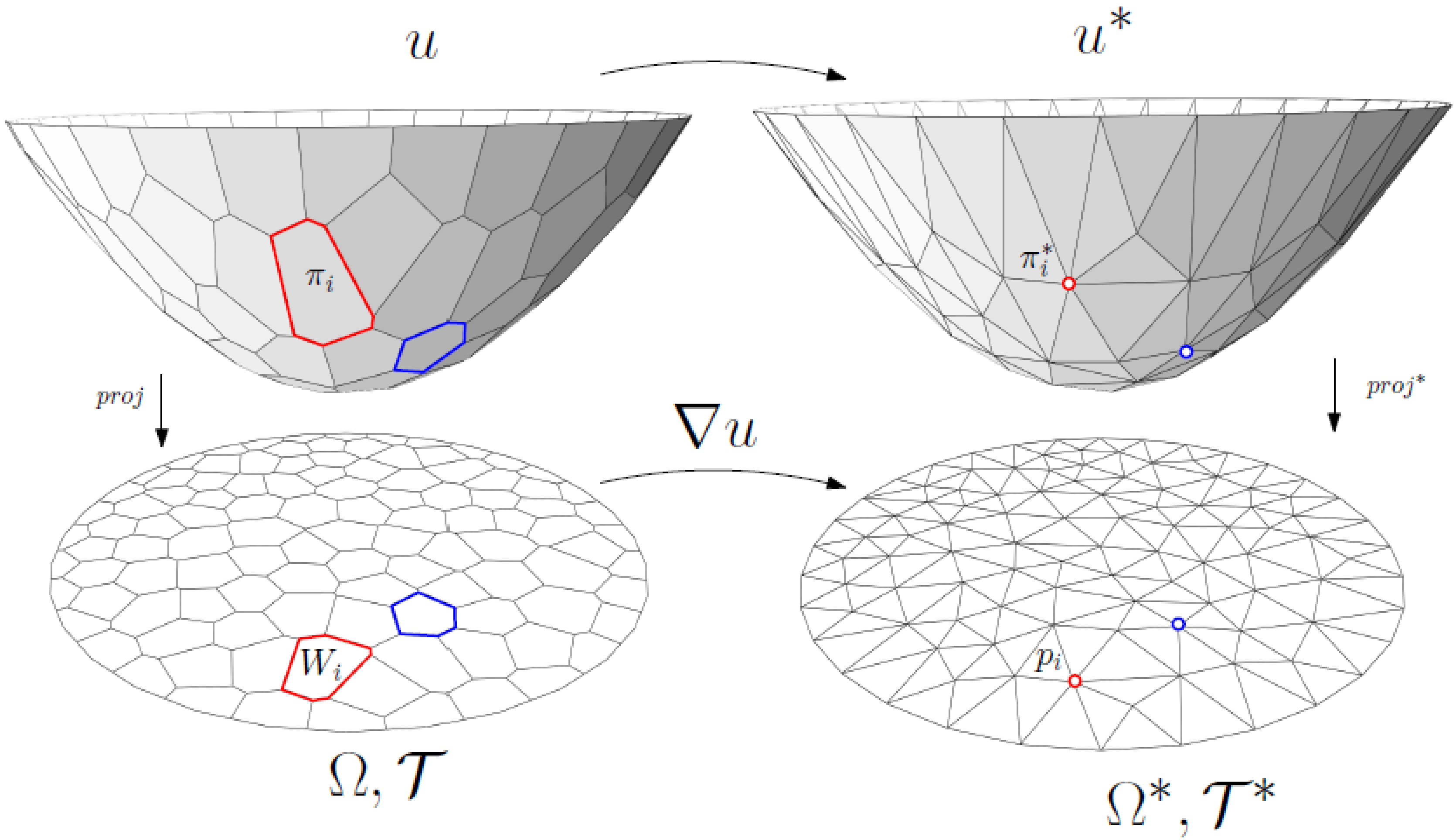
Theorem(Aurenhammer- Hoffmann- Aronov, (1998))

Alexandrov map ∇f is the optimal transport map.

Proof . Let $W_i=\{x \mid x \cdot p_i+h_i \geq x \cdot p_j+h_j, \text{ all } j\}$

Suppose $X_1,\ldots,X_n$ is a partition of X s.t., $vol(X_i)=A_i$ and $T(X_i)=p_i$. Then

$$
\begin{aligned}
cost(T) &= \int_X |x - T(x)|^2\, dx \\
&= \sum \int_{X_i} |x-p_i|^2 dx \\
&= \sum \int_{X_i} (|x-p_i|^2 + w_i) dx - \sum w_i A_i && (vol(X_i)=A_i) \\
&\geq \sum \int_{W_i} (|x-p_i|^2 + w_i) dx - \sum w_i A_i && (\text{lemma2, } vol(W_i)=vol(X_i)=A_i) \\
&= \sum \int_{W_i} |x-p_i|^2 dx \\
&= \int_X |x - \nabla f(x)|^2\, dx \\
&= cost(\nabla f).
\end{aligned}
$$



u
u^*
π_i
π_i^*
$proj$
$proj^*$
∇u
W_i
p_i
$\Omega, \mathcal{T}$
$\Omega^*, \mathcal{T}^*$

Recall

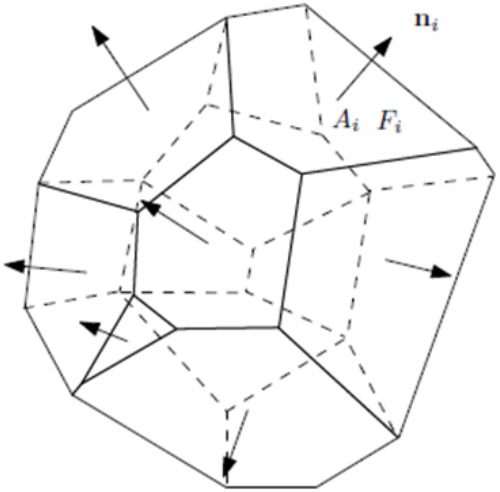
Minkowski thm. Given k unit vectors $n_1, \ldots n_k$ not contained in a half-space in R^N and $A_1, \ldots, A_k > 0$ s.t.,

$$\sum_i A_i n_i = 0,$$

∃, unique up to translation, cpt convex polytope P with exactly k codim-1 faces $F_1, \ldots, F_k$ s.t.,

(a) area $(F_i) = A_i$ and

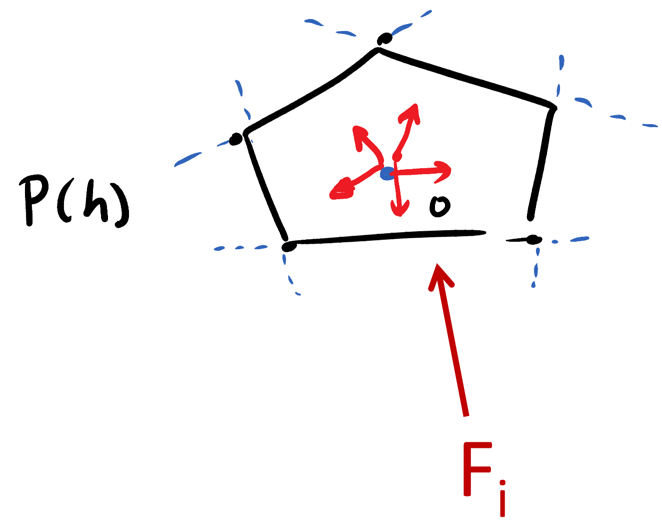
(b) $n_i \perp F_i$.



Mikowski's proof of his thm

Given $h=(h_1, \ldots, h_k)$, $h_i>0$, define cpt convex polytope

$$P(h)=\{x \mid x \cdot n_i \le h_i, \text{ all } i\}.$$



Let Vol: $R_+^k \to R$ be vol(h)=vol(P(h)).

Then, $$\frac{\partial Vol(h)}{\partial h_i} = \text{area}(F_i)$$

The solution h (up to scaling) to MP is the critical point of Vol on $\{ h \mid h_i \ge 0, \ \sum h_i A_i = 1\}$, using Lagrangian multiplier.

Uniqueness part is proved using Brunn-Minkowski inequality which implies $(Vol(h))^{1/N}$ is concave in h.

So far, this is the ONLY proof of uniqueness.

Alexandrov thm: Given X cpt convex domain in R^N, $p_1, \ldots, p_k$ distinct in R^N, $A_1,\ldots,A_k>0$ s.t., $\sum A_i = \mathrm{vol}(X)$, ∃ a PL convex function

$$f(x) = \max_i\{ x \cdot p_i + h_i\},$$

unique up to translation s.t.,

$$\mathrm{Vol}(\{x \in X \mid \nabla f(x) = p_i\}) = A_i.$$

Our Proof. For $h = (h_1, \ldots, h_k)$ in R^k, define f as above and let $W_i(h) = \{ x \mid x \cdot p_i + h_i \geq x \cdot p_j + h_j, \text{ all } j\}$ and $w_i(h) = \mathrm{vol}(W_i(h))$.

Step 1. $H = \{ h \in R^k \mid w_i(h) > 0, \text{ all } i\}$ is non-empty open convex set in R^k.

Step 2. (Key step) $\frac{\partial w_i(h)}{\partial h_j} = \frac{\partial w_j(h)}{\partial h_i} \leq 0$, for $i \neq j$.

Thus the differential 1-form $\sum_i w_i(h)\, dh_i$ is closed in H.

Therefore, ∃ a smooth $F: H \to R$ so that $\frac{\partial F}{\partial h_i} = w_i(h)$

Step 3. $\sum_i \frac{\partial w_i(h)}{\partial h_j} = 0$, due to $\sum_i w_i(h)$=vol(X).

This shows F(h) is convex in H (since the Hessian of F is diagonally dominated)

Step 4. F is strictly convex in $H_0=\{h \in H \mid \sum h_i=0\}$ so that $\nabla F = (w_1, \ldots, w_k)$.

Lemma 3. If F strictly convex on an open convex set Ω in R^m then $\nabla F: \Omega \to R^m$ is $1-1$.

This shows the uniqueness part of Alexandrov's thm.

We show that the concave function

$$G(h) = F(h) - \sum h_i A_i$$

has a minimum point in H_0. The min point h is the solution to Alexandrov's them.

Exactly the same proof works for Pogorelov's thm.

Thm(Gu-L-Sun-Yau). X cpt convex domain in R^N, $p_1, \ldots, p_k$ distinct in R^N, $s: X \to R$ positive continuous.
For any $A_1, \ldots, A_K > 0$ with $\sum A_i = \int_X s(x)\, dx$, $\exists$ a vector $(h_1, \ldots, h_k)$ so that

$$f(x) = \max\{ x \cdot p_i + h_i \}$$

satisfies $\int_{W_i \cap X} s(x)\, dx = A_i$ where $W_i = \{x \mid \nabla f = p_i\}$. Furthermore, h is the minimum point of the convex function

$$E(y) = \int_a^y \sum_i \int_{W_i \cap X} s(x)\, dx\, dy_i - \sum_i A_i y_i.$$

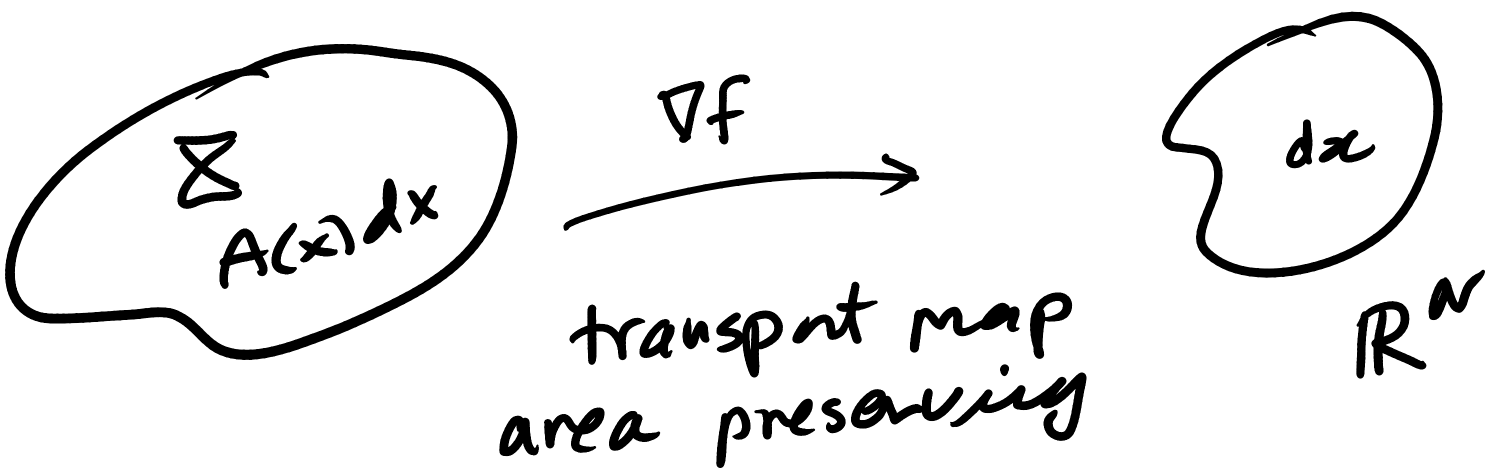
Alexandrov thm corresponds to s(x)=1. Y. Brenier proved a more general form.

Discrete Monge-Ampere Eq (DMAE)

Simplest version: X domain in R^N, A: X $\rightarrow R_{>0}$ find f: X $\rightarrow R$, s.t.,

$$\begin{cases} \det(\mathrm{Hess}(f)) = A, \\ f|_{\partial X} = g \end{cases}$$

This is related to Monge's optimal transport problem:

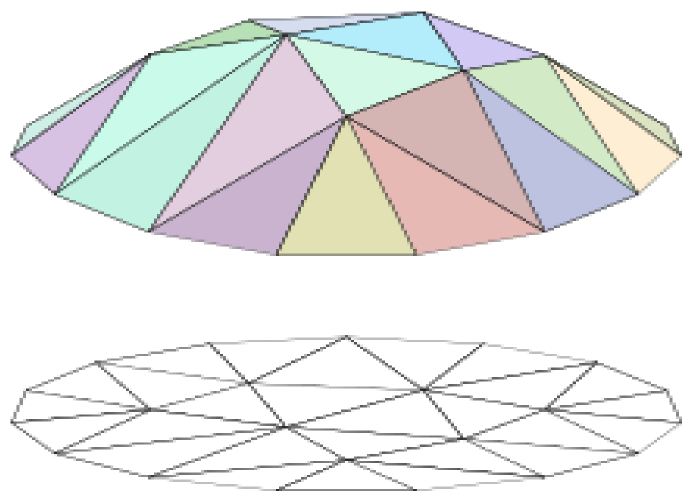


Q2: Given A, g how to compute f?

Q3. What is the discrete det(Hess(f))?

Let X =conv{v_1, …, v_k} a domain in R^N,

u : X → **R** is PL convex function w.r.t a convex cell decomposition $\mathscr{T}$. Then the discrete Hessian det of u sends v $\in \mathscr{T}^{(0)}$ to the volume of the convex hull of the gradients of u at top-dim cells adjacent to v.



Thm(Pogorelov). Suppose X=conv$\{v_1, \ldots, v_m\}$ convex domain in R^N, s.t. v_i not in con$\{v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_m\}$, and $g_1, \ldots, g_m$ in **R**.

$\forall$ $p_1, \ldots, p_k$ in int(X) and $A_1, \ldots, A_K > 0$, then

$\exists\,!$ PL convex function $w: X \to R$ having vertices exactly at p_i, s.t.

(a) the discrete Hessian det of w at p_i is A_i,

(b) $w(v_j)=g_j$ all j.

Indeed, $w(y)=\sup\{x \cdot y - f(x) \mid x\}$ is the Fenchel-Legendre dual of the solution to Pogorelov's thm:

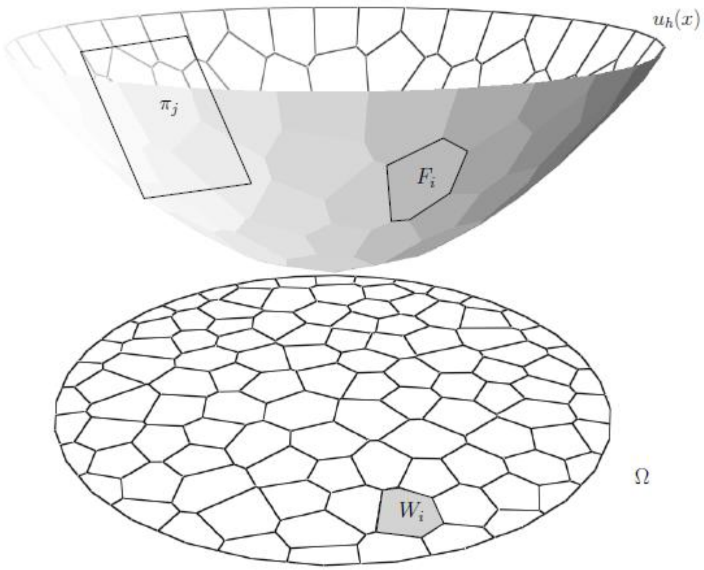
$$f(x)=\max\{\max\{x \cdot p_i + h_i\}, \max\{x \cdot v_j + g_j\}\}.$$

Our result shows that w can be constructed by a finite dim variational principle since dual of PK convex function is computable using linear programming.

Algorithm

- Convex Hull
- Delaunay Triangulation
- Vornoi diagram
- Power Diagram – upper envelope
- Optimal Transportation Map

Computational Algorithm

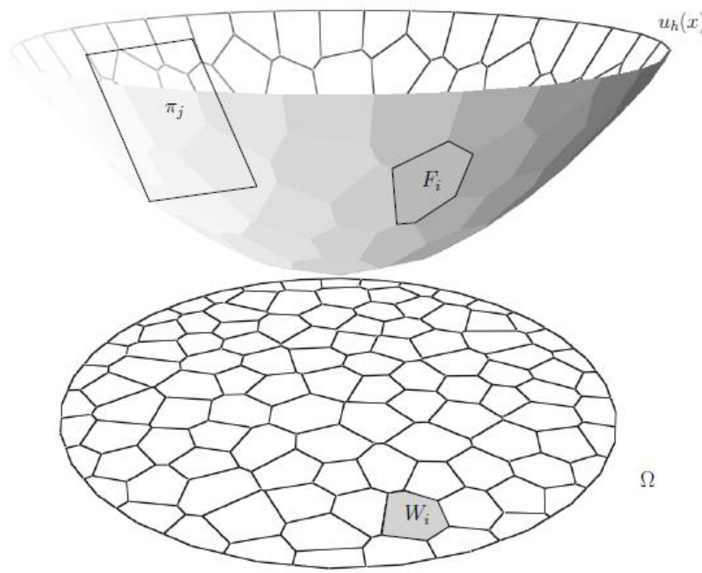


The convex energy is

$$E(h_1, h_2, \cdots, h_k) = \sum_{i=1}^{k} A_i h_i - \int_{\mathbf{0}}^{\mathbf{h}} \sum_{j=1}^{k} W_j dh_j,$$

Geometrically, the energy is the volume beneath the parabola.

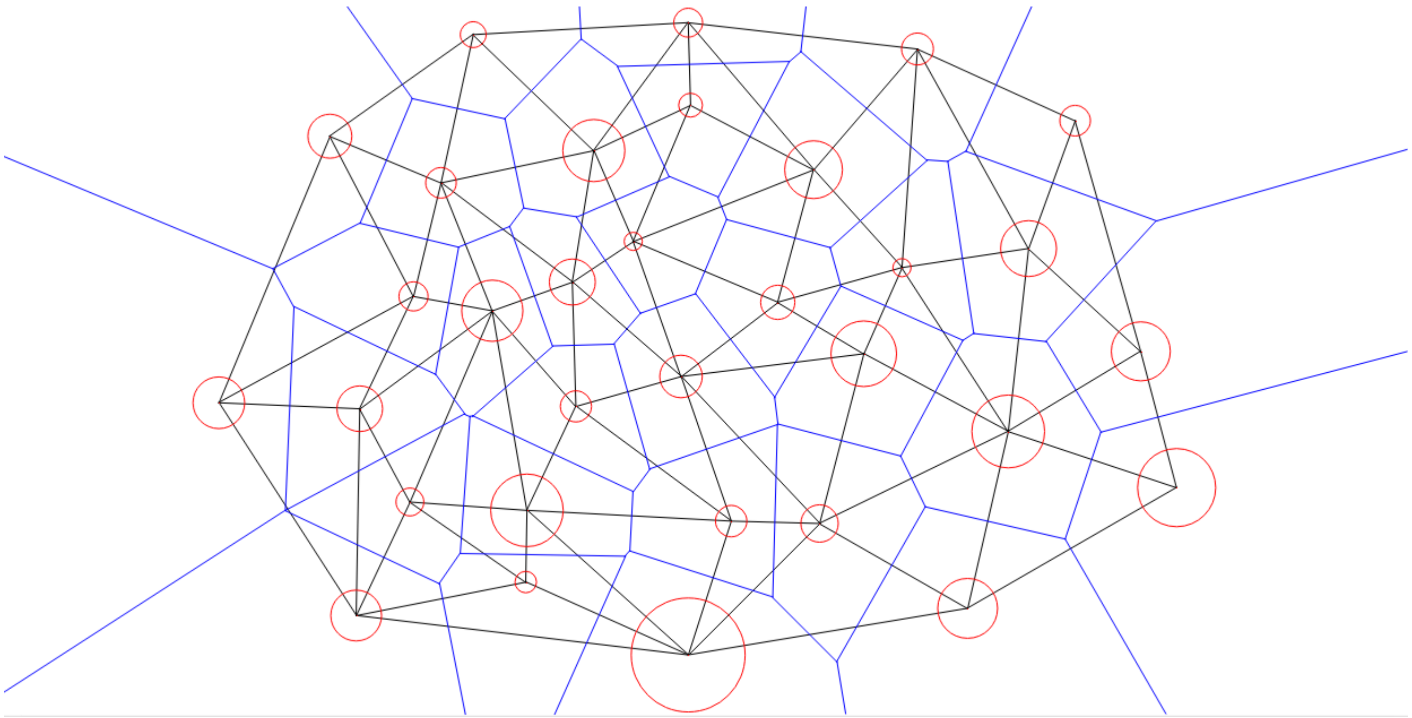
Computational Algorithm



The gradient of the energy is the areas of the cells

$$\nabla E(h_1, h_2, \cdots, h_k) = (A_1 - w_1, A_2 - w_2, \cdots, A_k - w_k)$$

Computational Algorithm



The Hessian of the energy is the length ratios of edge and dual edges,

$$\frac{\partial w_i}{\partial h_j} = \frac{|e_{ij}|}{|\bar{e}_{ij}|}$$

Computational Algorithm

1. Initialize $\mathbf{h} = \mathbf{0}$
2. Compute the Power Voronoi diagram, and the dual Power Delaunay Triangulation
3. Compute the cell areas, which gives the gradient ∇E
4. Compute the edge lengths and the dual edge lengths, which gives the Hessian matrix of E, $Hess(E)$
5. Solve linear system

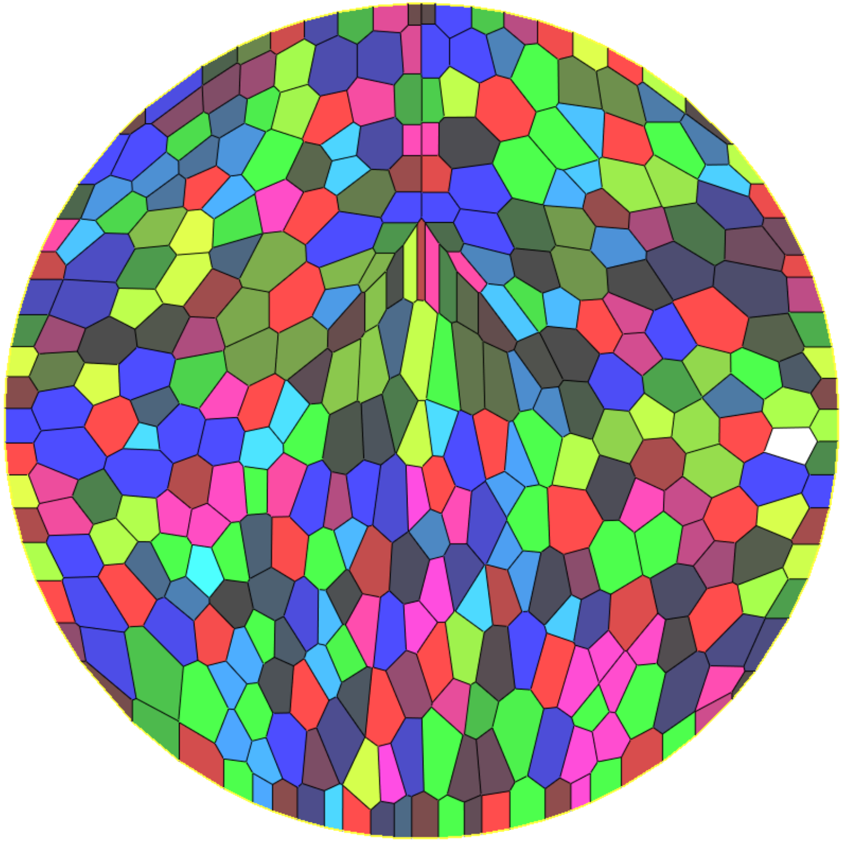
$$\nabla E = Hess(E) d\mathbf{h}$$

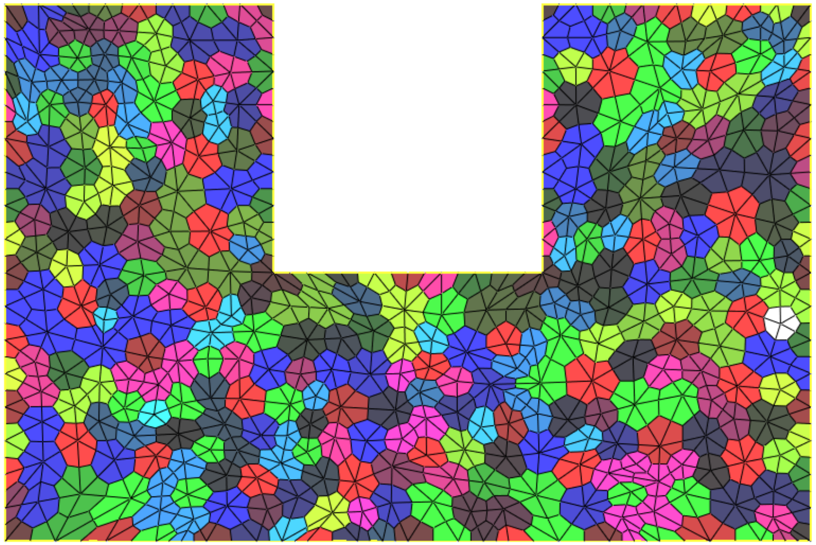
6. Update the height vector

$$(h) \leftarrow \mathbf{h} - \lambda d\mathbf{h},$$

   where λ is a constant to ensure that no cell disappears
7. Repeat step 2 through 6, until $\|d\mathbf{h}\| < \varepsilon$.

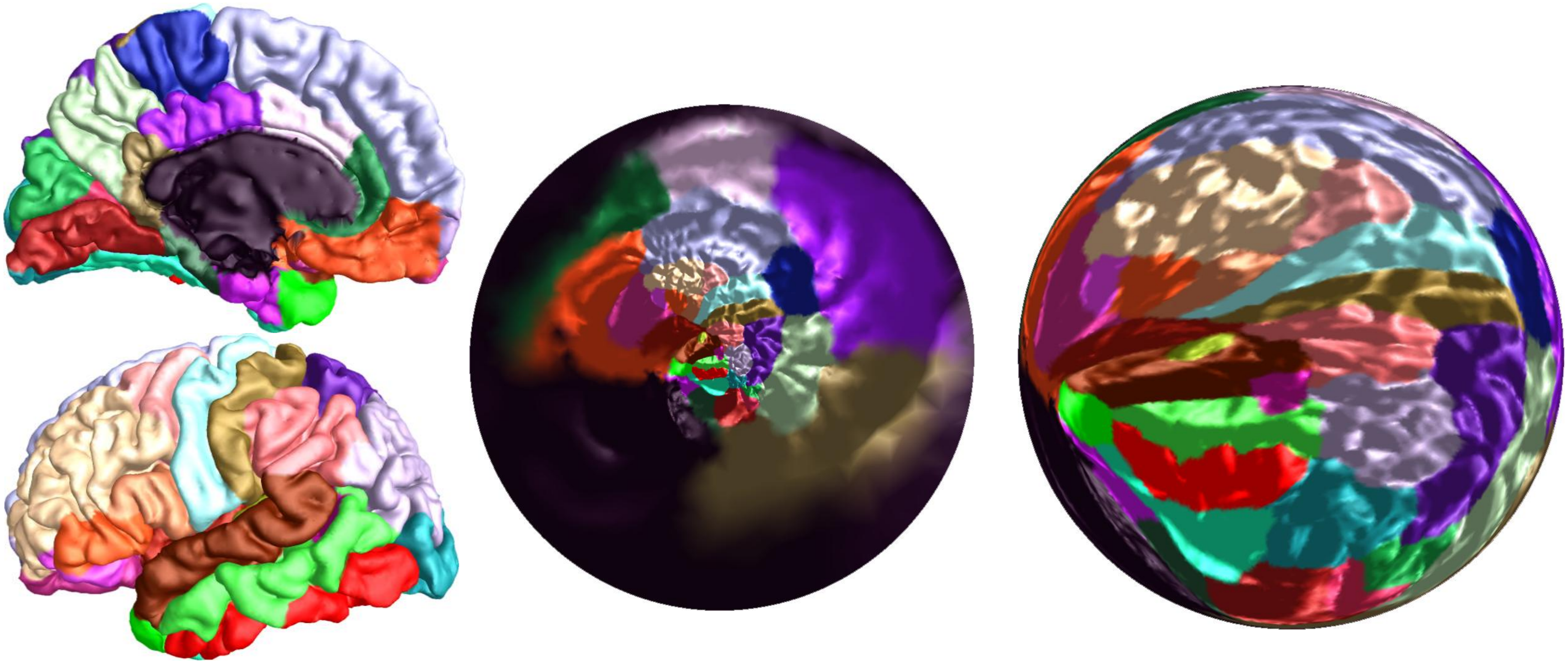
Examples

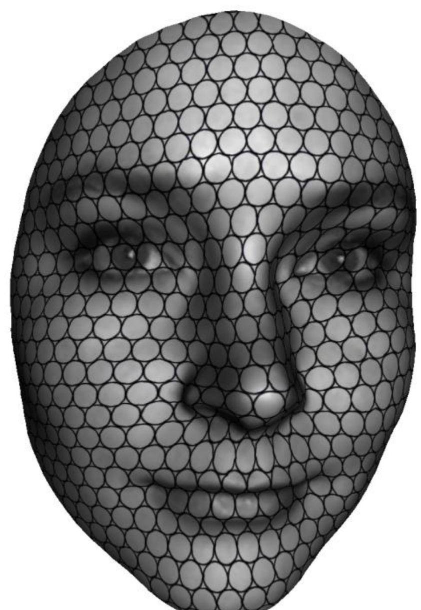


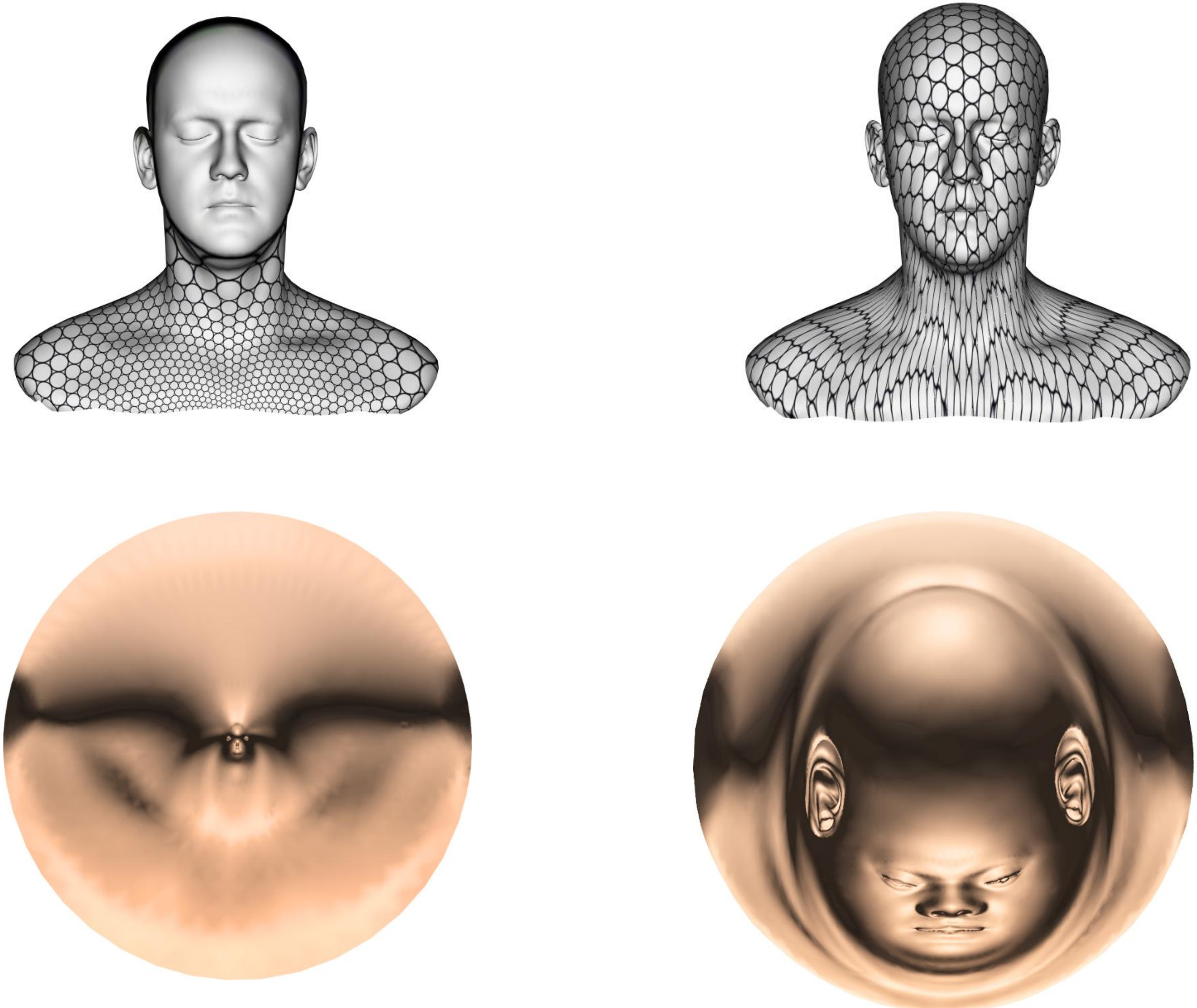


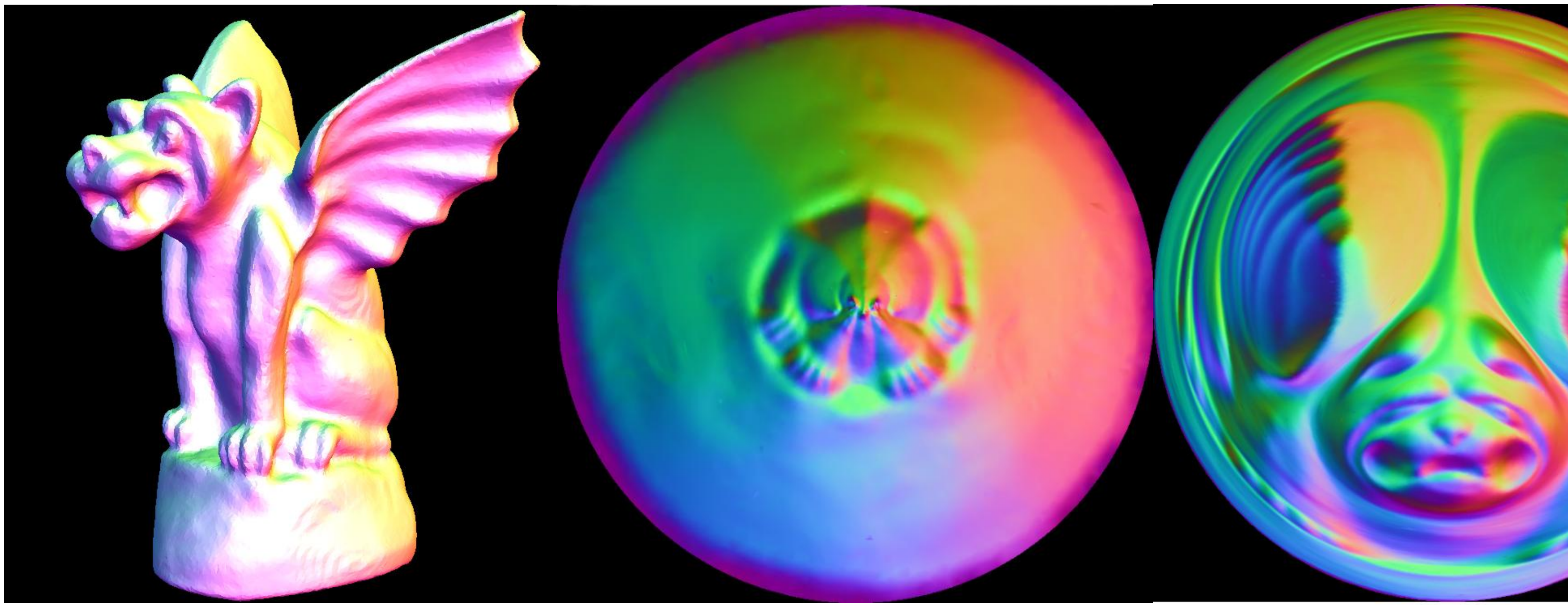
Examples

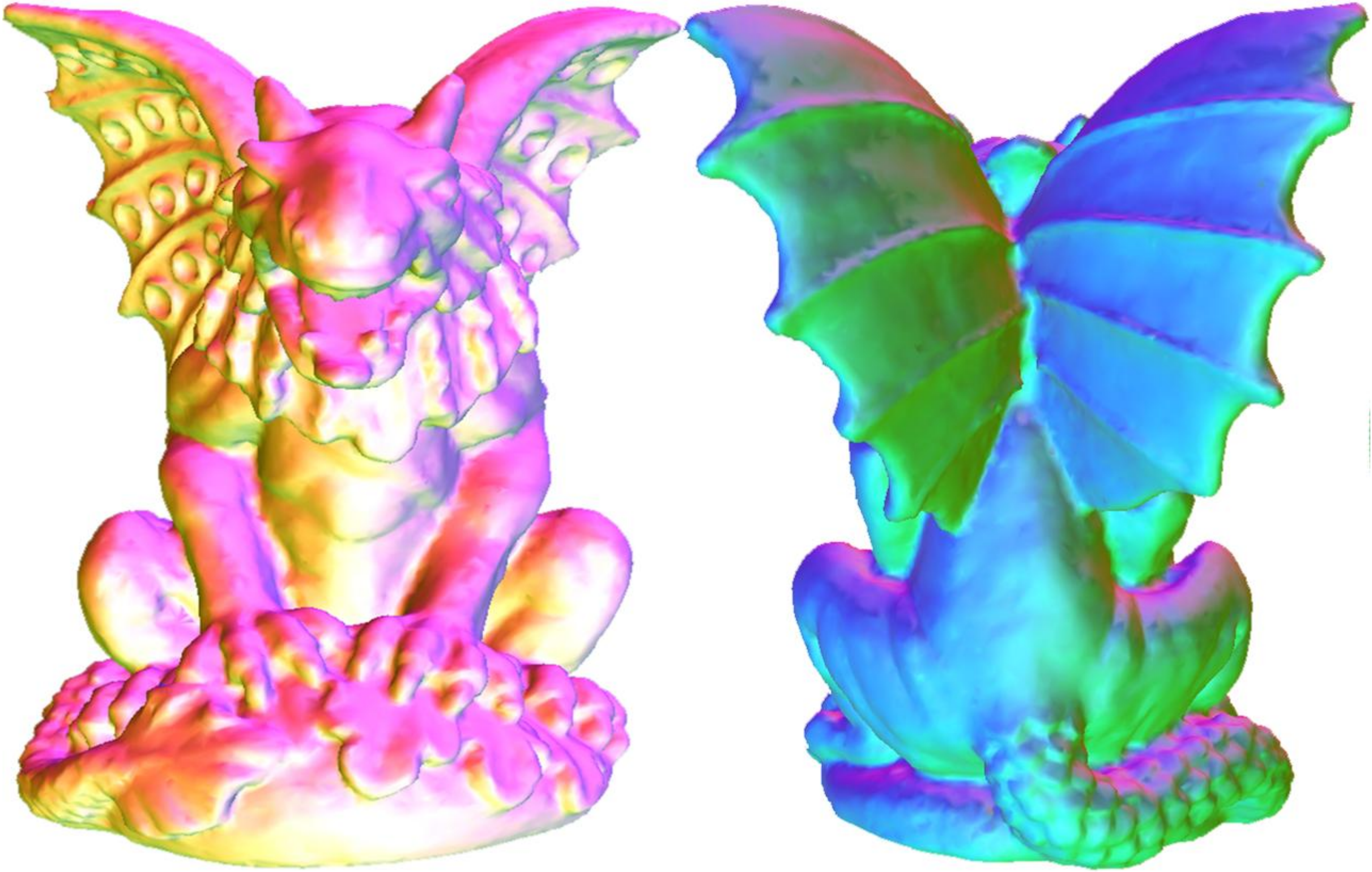
Examples

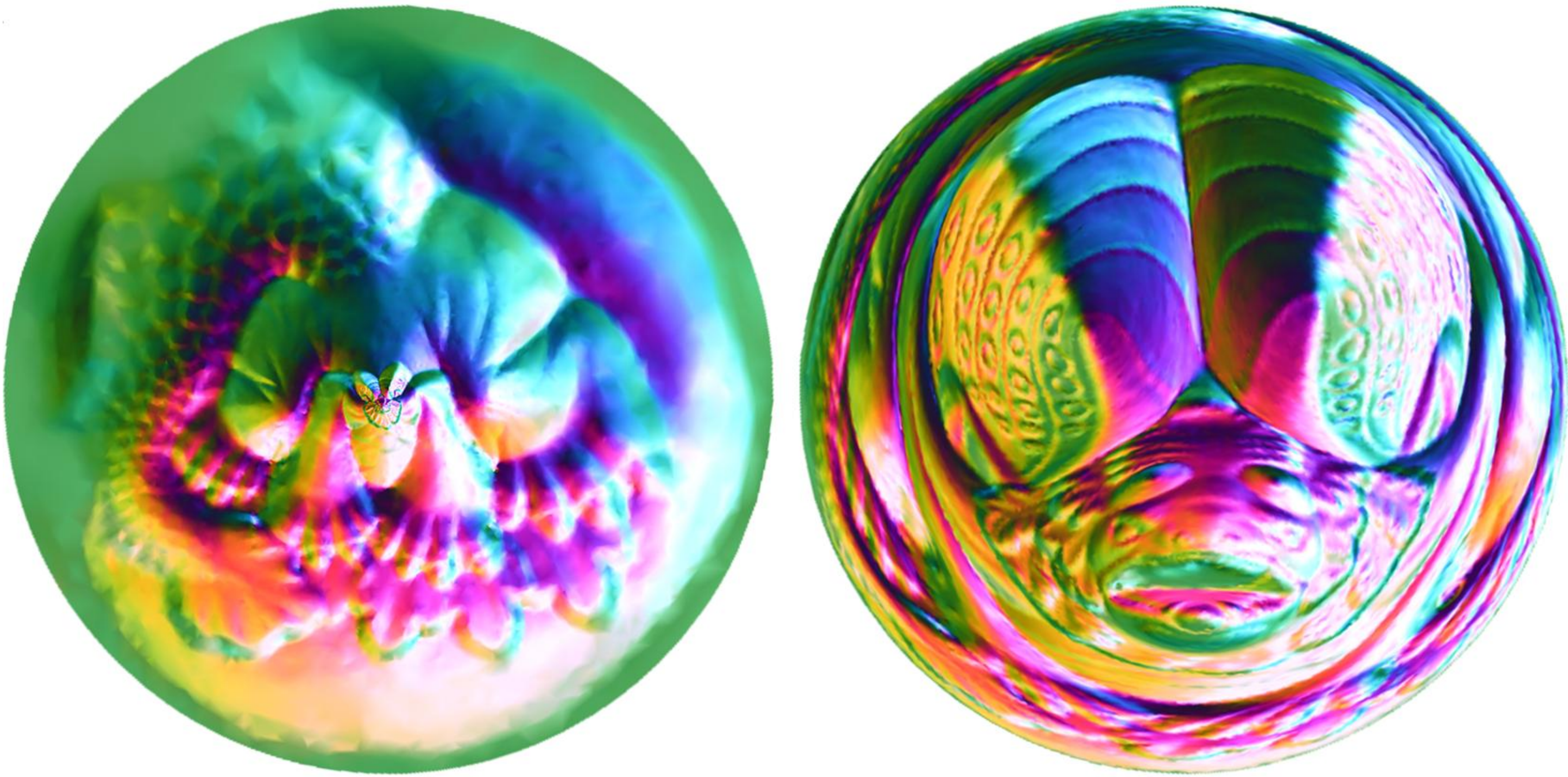




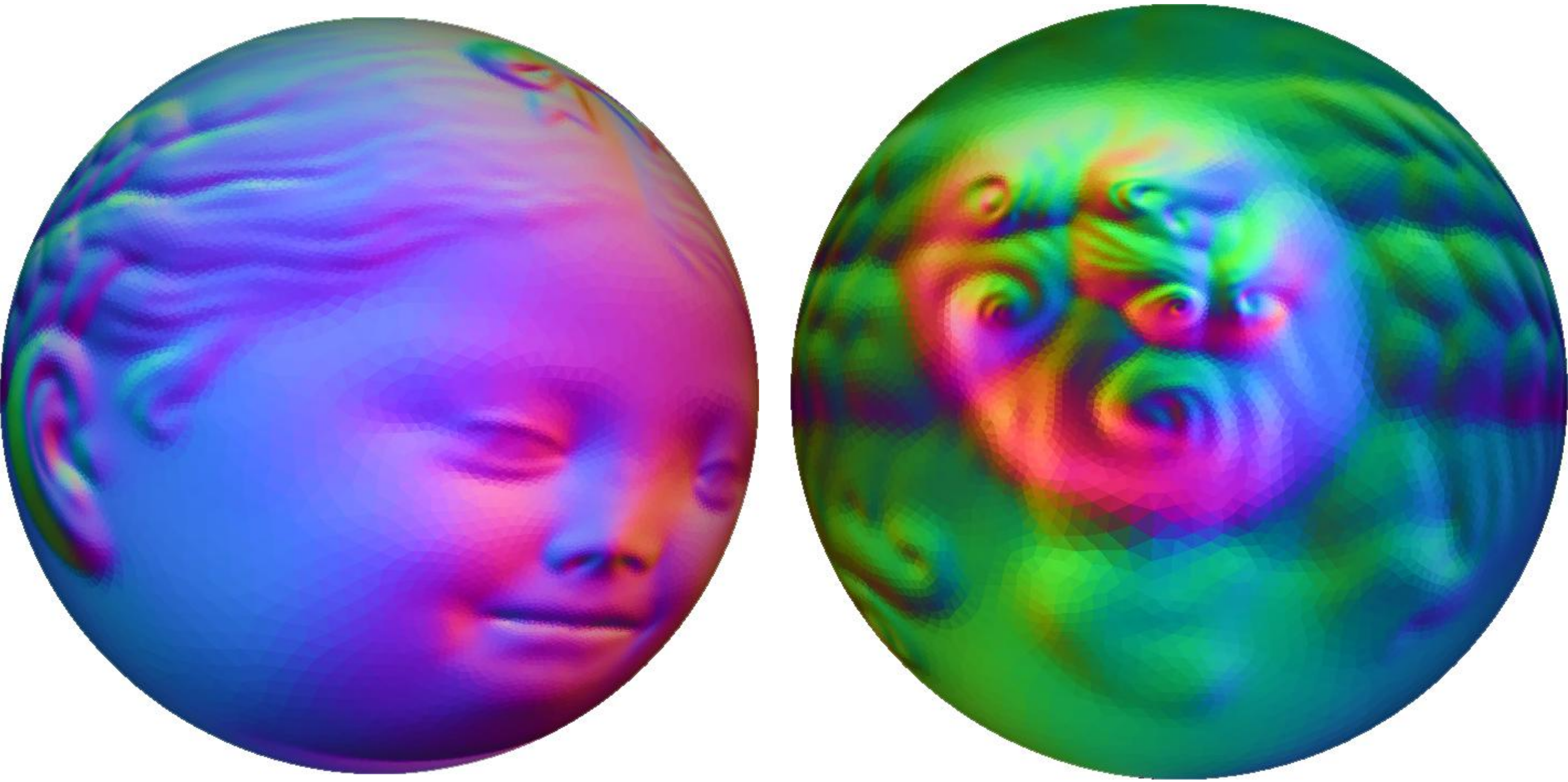


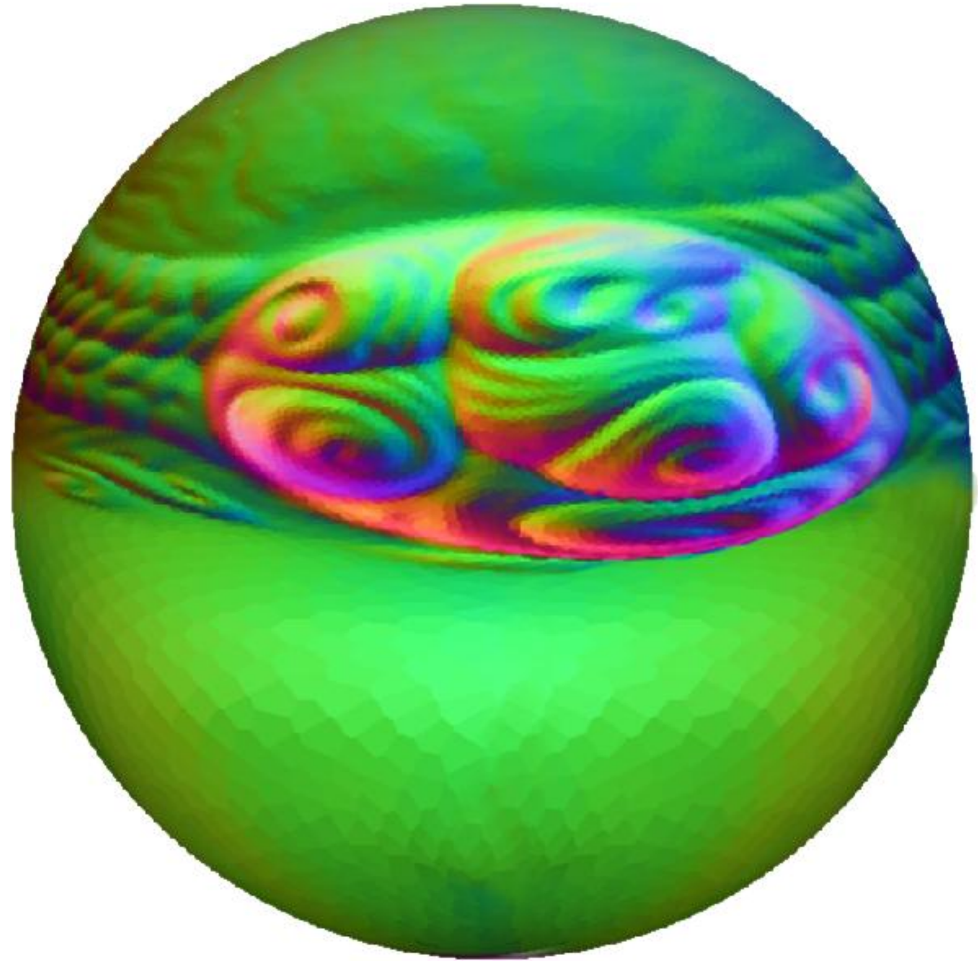



(a)



(b) (c)



Thank you.