## Regression

# Mladen Victor WICKERHAUSER 

Washington University in St. Louis, Missouri<br>victor@wustl.edu<br>http://www.math.wustl.edu/~victor

Dimensionality Reduction and Manifold Estimation
PMF - University of Zagreb
Winter, 2022

## Smoothness Assumptions

Suppose that $X$ and $Y$ are metric spaces.
Desirable properties of a nice function $f: X \rightarrow Y$ are:

- $x^{\prime}$ near $x$ implies that $f\left(x^{\prime}\right)$ is near $f(x)$.
- $f(B(x, \epsilon))$ can be estimated (with error controlled by $\epsilon$ ) from a few values $\left\{f\left(x^{\prime}\right): x^{\prime} \in B(x, \epsilon)\right\}$.
Examples:
- continuously differentiable $f: \mathbf{R} \rightarrow \mathbf{R}$.
- continuous piecewise linear $f$ for linear spaces $X, Y$.
- smooth vector field $f$ on differentiable manifold $X=\mathcal{M}$, taking values in tangent bundle $Y=\mathcal{M} \times \mathbf{E}^{d}$.


## Interpolation Sets

One-dimensional special case: $f: \mathbf{R} \rightarrow \mathbf{R}$, exact fit.
Given:

- abscissas $\left\{x_{0}, x_{1}, \ldots, x_{n}\right\} \subset \mathbf{R}$, all distinct.
- ordinates $\left\{y_{0}, y_{1}, \ldots, y_{n}\right\} \subset \mathbf{R}$, arbitrary.

Find: continuous function $f: \mathbf{R} \rightarrow \mathbf{R}$ such that $f\left(x_{i}\right)=y_{i}$, all $i$.
Idea: fix functions $\left\{f_{0}, f_{1}, \ldots\right\}$, solve for $\left\{a_{0}, a_{1}, \ldots\right\}$ such that

$$
y_{i}=f\left(x_{i}\right) \stackrel{\text { def }}{=} \sum_{j} a_{j} f_{j}\left(x_{i}\right), \quad i=0,1, \ldots, n
$$

Exercise: Unique solution $\left\{a_{j}\right\}$ exists for each $\left\{y_{i}: 0 \leq i \leq n\right\}$ iff matrix $\left\{f_{j}\left(x_{i}\right): 0 \leq i, j \leq n\right\}$ is nonsingular.

## Polynomial Interpolation

Special case: $f_{j}(x)=x^{j}, j=0,1, \ldots, n$.
Seek polynomial $f(x)=p_{n} x^{n}+p_{n-1} x^{n-1}+\cdots+p_{1} x+p_{0}$.
Coefficients $\left\{p_{n}, p_{n-1}, \ldots, p_{1}, p_{0}\right\}$ satisfy

$$
V \mathbf{p}=\mathbf{y} ; \quad\left(\begin{array}{ccccc}
x_{0}^{n} & x_{0}^{n-1} & \cdots & x_{0} & 1 \\
x_{1}^{n} & x_{1}^{n-1} & \cdots & x_{1} & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
x_{n}^{n} & x_{n}^{n-1} & \cdots & x_{n} & 1
\end{array}\right)\left(\begin{array}{c}
p_{n} \\
p_{n-1} \\
\vdots \\
p_{1} \\
p_{0}
\end{array}\right)=\left(\begin{array}{c}
y_{0} \\
y_{1} \\
\vdots \\
y_{n}
\end{array}\right) .
$$

Exercise: Distinct $x_{i}$ 's make Vandermonde matrix $V$ invertible.
Remark. Using Octave/Matlab: Get $V=$ vander $(x)$ and solve for $\mathbf{p}$. Or, compute $\mathrm{p}=$ polyfit ( $\mathrm{x}, \mathrm{y}, \mathrm{n}$ ) directly and evaluate $f(z)=\operatorname{polyval}(\mathrm{p}, \mathrm{z})$ at any point $z$.

## Spline Interpolation

Special case: $x_{0}<x_{1}<\cdots<x_{n}$, and $f(x)$ is a piecewise polynomial equal to cubic $f_{j}$ on $\left[x_{j-1}, x_{j}\right]$ for $j=1, \ldots, n$ Method: find coefficients $a_{j k}$ of

$$
f_{j}(x)=a_{j 3} x^{3}+a_{j 2} x^{2}+a_{j 1} x+a_{j 0}
$$

to match $y$ values, zeroth, first, and second derivatives at $x_{1}, \ldots, x_{n-1}$. Then put

$$
f(x)=f_{j}(x), \quad \text { if } x_{j-1} \leq x \leq x_{j}
$$

Advantage: $C^{2}$ smooth without undesirable high degree for large $n$.
Remark. Get $\left\{a_{j k}\right\}$ (in a returned data structure) with $\mathrm{pp}=\operatorname{spline}(\mathrm{x}, \mathrm{y})$. Evaluate $f(z)=\mathrm{ppval}(\mathrm{pp}, \mathrm{z})$ at any point $x_{0} \leq z \leq x_{n}$ using Octave/Matlab.

## General Linear Regression

Fix $f_{1}, \ldots, f_{m}$ linearly independent functions from $\mathcal{M} \rightarrow \mathbf{R}$. Given $\left\{x_{1}, \ldots, x_{n}\right\} \subset \mathcal{M}$ and $\left\{y_{1}, \ldots, y_{n}\right\} \subset \mathbf{E}^{d}$, find a vector $\mathbf{p}=\left(p_{1}, \ldots, p_{m}\right)$ of expansion coefficients so that

$$
f(x) \stackrel{\text { def }}{=} p_{1} f_{1}(x)+p_{2} f_{2}(x)+\cdots+p_{m} f_{m}(x)
$$

minimizes the squared error

$$
E(\mathbf{p}) \stackrel{\text { def }}{=} \sum_{i=1}^{n}\left\|f\left(x_{i}\right)-y_{i}\right\|^{2}=\sum_{i=1}^{n}\left\|\sum_{j=1}^{m} p_{j} f_{j}\left(x_{i}\right)-y_{i}\right\|^{2}
$$

## Normal Equations

Global minimum exists at the unique critical point $\nabla E(\mathbf{p})=\mathbf{0}$.
This equation gives a system of $m$ linear normal equations for $\mathbf{p}$ :

$$
\mathbf{F}^{T} \mathbf{F} \mathbf{p}=\mathbf{F}^{T} \mathbf{y}, \quad \text { for } \quad \mathbf{F}=\left(\begin{array}{ccc}
f_{1}\left(x_{1}\right) & \cdots & f_{m}\left(x_{1}\right) \\
\vdots & \ddots & \vdots \\
f_{1}\left(x_{n}\right) & \cdots & f_{m}\left(x_{n}\right)
\end{array}\right)
$$

In terms of matrix coordinates $1 \leq i, j \leq m$,

$$
\mathbf{F}^{\top} \mathbf{F}(i, j)=\sum_{k=1}^{n} f_{i}\left(x_{k}\right) f_{j}\left(x_{k}\right) \in \mathbf{R}, \quad \mathbf{F}^{T} \mathbf{y}(i)=\sum_{k=1}^{n} f_{i}\left(x_{k}\right) y_{k} \in \mathbf{R}
$$

Matrix $\mathbf{F}^{\top} \mathbf{F} \in \mathbf{R}^{m \times m}$ is invertible whenever $\mathbf{F}$ has rank $m$, which requires $n \geq m$ and at least $m$ distinct values in $\left\{x_{1}, \ldots, x_{n}\right\}$.

## Geometric Interpretation

If $\mathbf{F}^{T} \mathbf{F} \in \mathbf{R}^{m \times m}$ is invertible, then

$$
\hat{\mathbf{y}} \stackrel{\text { def }}{=} \mathbf{F}\left(\mathbf{F}^{T} \mathbf{F}\right)^{-1} \mathbf{F}^{T} \mathbf{y}
$$

is the orthogonal projection of the vector $\mathbf{y}$ onto the column space of $\mathbf{F}$, which is an m-dimensional subspace of $\mathbf{E}^{n}$. Column vectors

$$
\mathbf{f}_{1} \stackrel{\text { def }}{=}\left(\begin{array}{c}
f_{1}\left(x_{1}\right) \\
\vdots \\
f_{1}\left(x_{n}\right)
\end{array}\right), \ldots, \mathbf{f}_{m} \stackrel{\text { def }}{=}\left(\begin{array}{c}
f_{m}\left(x_{1}\right) \\
\vdots \\
f_{m}\left(x_{n}\right)
\end{array}\right)
$$

are a basis $\mathbf{f}$ for this subspace, and

$$
\hat{\mathbf{p}} \stackrel{\text { def }}{=}\left(\mathbf{F}^{T} \mathbf{F}\right)^{-1} \mathbf{F}^{T} \mathbf{y}
$$

gives the expansion coefficients of $\hat{\mathbf{y}}$ with respect to $\mathbf{f}$.

## Statistical Interpretation

Model: for $k=1, \ldots, n$, make measurements

$$
y_{k}=f\left(x_{k}\right)=\sum_{i=1}^{m} p_{i} f_{i}\left(x_{k}\right)+\epsilon_{k} .
$$

Assumptions:

- Functions $f_{1}, \ldots, f_{m}$ and abscissas $x_{1}, \ldots, x_{n}$ are given.
- Parameters $\mathbf{p}=\left(p_{1}, \ldots, p_{m}\right)$ are fixed but unobservable.
- Each measurement has unobservable random error $\epsilon_{k}$.
- Measurement vector $\mathbf{y} \stackrel{\text { def }}{=}\left(y_{1}, \ldots, y_{n}\right)$ is observable but random (because of the errors in each component).
Estimator $\hat{\mathbf{p}}=\hat{\mathbf{p}}(\mathbf{y}) \in \mathbf{R}^{m}$ is some fixed function of $\mathbf{y}$.
Idea: estimate $\mathbf{p} \approx \hat{\mathbf{p}}=\hat{\mathbf{p}}(\mathbf{y})$ from particular measurements $\mathbf{y}$.


## Estimator Properties

Assume that estimator $\hat{\mathbf{p}}=\hat{\mathbf{p}}(\mathbf{y})$ for $\mathbf{p}$ has finite expectation and finite variance. Say that $\hat{\mathbf{p}}$ :

- is linear if $\hat{\mathbf{p}}(\mathbf{y})$ is a linear function of $\mathbf{y}$,
- is unbiased if $\mathrm{E}(\hat{\mathbf{p}})=\mathbf{p}$.
- has risk defined by $\mathrm{E}\left(\|\hat{\mathbf{p}}-\mathrm{E}(\mathbf{p})\|^{2}\right)$.

Note: for unbiased $\hat{\mathbf{p}}$, the risk is $\mathrm{E}\left(\|\hat{\mathbf{p}}-\mathbf{p}\|^{2}\right)$.
Say that random variables $r, s$ with finite expectation and variance are uncorrelated iff $\mathrm{E}(r s)=\mathrm{E}(r) \mathrm{E}(s)$.
Exercise: If $\hat{\mathbf{p}}$ is linear and $\mathbf{y}$ has finite expectation and variance, then $\hat{\mathbf{p}}(\mathbf{y})$ has finite variance and expectation.

## Optimality of Least Squares

When $\mathbf{F}$ has rank $m$, then

$$
\hat{\mathbf{p}} \stackrel{\text { def }}{=}\left(\mathbf{F}^{T} \mathbf{F}\right)^{-1} \mathbf{F}^{T} \mathbf{y}
$$

defines the ordinary least squares (OLS) estimator for the "true" expansion coefficients $\mathbf{p}$. It is evidently linear.
Suppose $\left\{\epsilon_{1}, \ldots, \epsilon_{n}\right\}$ are uncorrelated samples of a random variable $\epsilon$ with $\mathrm{E}(\epsilon)=0$ and $\operatorname{Var}(\epsilon)=\sigma^{2}<\infty$. Then $\hat{\mathbf{p}}$ will be unbiased with finite variance. In addition:
Theorem (Gauss-Markov)
Estimator $\hat{\mathbf{p}}$ has minimal variance (or risk) among all linear unbiased estimators for $\mathbf{p}$.

## Piecewise Linear Interpolation

Special case, like for splines: $x_{0}<x_{1}<\cdots<x_{n}$, but $f(x)$ is a piecewise linear polynomial.
Method: explicit formula

$$
f(x)=f_{j}(x)=\frac{\left(x_{j}-x\right) y_{j-1}+\left(x-x_{j-1}\right) y_{j}}{x_{j}-x_{j-1}}, \quad x_{j-1} \leq x \leq x_{j}
$$

Advantages:

- continuity: $f\left(x_{j}\right)=f_{j}\left(x_{j}\right)=f_{j+1}\left(x_{j}\right)=y_{j}$ for all $j$.
- positivity: $y_{j}, y_{j-1} \geq 0 \Longrightarrow\left(\forall x \in\left[x_{j-1}, x_{j}\right]\right) f(x) \geq 0$
- piecewise constant derivative almost everywhere;

Remark. Using Octave/Matlab: Get $\left\{f_{j}\right\}$ (in a returned data structure) with pl=interp1(x,y,"linear", "pp"). Evaluate $f(z)=\operatorname{ppval}(\mathrm{pl}, \mathrm{z})$ at any point $x_{0} \leq z \leq x_{n}$.

## Higher Dimensional Piecewise Linear Interpolation

Goal: Find neighborhoods in $\mathbf{E}^{d}$ better suited to linear approximation than open balls.

Idea: use fixed finite point sets, take linear combinations.
Choices:

- cube: $\left[a_{1}, b_{1}\right] \times \cdots \times\left[a_{d}, b_{d}\right] \subset \mathbf{E}^{d}, 2 d$ coordinates.
- grid of $d$-cubes, $m$ per axis, $m^{d}$ in all.
- specify gridpoints in a $d \times(m+1)$ matrix, using $d$ increasing sequences of length $m+1$;
- plot $m^{d}$ values, one for each cube.
- simplex: $d+1$ vertices specifying a $d$-dimensional set.
- tesselation of $d$-simplexes formed from $m \geq d+1$ vertices;
- specify the vertices in a $d \times m$ matrix;
- specify $k \leq\binom{ m}{d+1}$ simplexes in a $k \times(d+1)$ matrix;
- plot $k$ values, one for each simplex.


## Convex Hulls and Simplexes

- $S \subset \mathbf{E}^{d}$ is convex iff

$$
\mathbf{x}, \mathbf{y} \in S \Longrightarrow(\forall t \in[0,1]) t \mathbf{x}+(1-t) \mathbf{y} \in S
$$

- Convex hull of $S \subset \mathbf{E}^{d}$ is $\cap\left\{K \subset \mathbf{E}^{d}: K\right.$ convex, $\left.S \subset K\right\}$;
- Convex hull is convex, hence minimal, hence unique.
- For finite $S=\left\{\mathbf{x}_{1}, \ldots, \mathbf{x}_{m}\right\} \subset \mathbf{E}^{d}$, use convex combinations:

$$
\operatorname{ConvHull}(S)=\left\{p_{1} \mathbf{x}_{1}+\cdots+p_{m} \mathbf{x}_{m}:(\forall i) p_{i} \geq 0 ; \sum_{i=1}^{m} p_{i}=1\right\}
$$

- Simplexes in $\mathbf{E}^{d}$ :
- d-simplex: convex hull of $d+1$ points with nonempty interior.
- $n$-simplex, $1 \leq n<d$ : convex hull of $n+1$ points that contains a nonempty relatively open $n$-ball.

Exercise: $K, K^{\prime}$ convex implies $K \cap K^{\prime}$ is convex.

## QuickHull

Recursive subdivision algorithm to compute the convex hull of a set $S$ of $N$ points in $\mathbf{E}^{d}$ :

- complexity: $O(N \log N)$ expected, $O\left(N^{2}\right)$ worst case.
- conditioning: poor, due to degenerate simplexes.
- rounding error: small diameter $S$ very far from 0 .

Links to articles on these practical issues:

- Original article: QuickHull.pdf
- Software and documentation: http://www.qhull.org/


## Graph Interpolation

Special case: 2 dimensional domain.

- abscissas $X=\left\{\mathbf{x}_{1}, \ldots, \mathbf{x}_{n}\right\} \subset \mathbf{E}^{2}$, distinct, noncollinear.
- ordinates $Y=\left\{y_{1}, \ldots, y_{n}\right\} \subset \mathbf{R}$, arbitrary.

Given $\mathbf{x} \in \operatorname{ConvHull}(X)$, find the piecewise linear graph interpolation value $y$ by:

- finding neighbors $\mathbf{x}_{a}, \mathbf{x}_{b}, \mathbf{x}_{c} \in X$ that form a 2-simplex ("triangle") containing $\mathbf{x}$;
- write $\mathbf{x}=p_{a} \mathbf{x}_{a}+p_{b} \mathbf{x}_{b}+p_{c} \mathbf{x}_{c}$ as a convex combination;
- define $y=p_{a} y_{a}+p_{b} y_{b}+p_{c} y_{c}$ as the same convex combination.

Exercise: If $\mathbf{x}$ is in the 2 -simplex defined by $\mathbf{x}_{a}, \mathbf{x}_{b}, \mathbf{x}_{c}$, then the convex combination $\mathbf{x}=p_{a} \mathbf{x}_{a}+p_{b} \mathbf{x}_{b}+p_{c} \mathbf{x}_{c}$ is unique.

## Delaunay Triangulation

Given: $S=\left\{\mathbf{p}_{1}, \ldots, \mathbf{p}_{m}\right\} \subset \mathbf{E}^{2}$, ConvHull $(S)$ has a nonempty interior (iff there exist three noncollinear points in $S$ ).

Goal: find a triangulation of ConvHull (S).
Delaunay's method: Any three noncollinear points in $S$ determine a 2-simplex and its circumcircle. Call it a Delaunay triangle iff the circumcircle contains no other points of $S$ in its interior.
Nonoverlapping $U, V \in \mathbf{E}^{2}: U \cap V$ has no interior.
$S$ is in general position iff: no 3 points are collinear and, if 4 points share a circumcircle, then there is a 5 th point in its interior.
Theorem
If $S$ is in general position, then ConvHull (S) has a unique decomposition as a nonoverlapping union of Delaunay triangles.

## Delaunay Tesselation

Tesselation of a finite set $S=\left\{\mathbf{p}_{1}, \ldots, \mathbf{p}_{m}\right\} \subset \mathbf{E}^{d}$ is a union of nonoverlapping $d$-simplexes, each determined by $d+1$ points of $S$, that equals the convex hull of $S$.

Any $d$-simplex determines a unique circumsphere.
$S$ is in general position in $\mathbf{E}^{d}$ iff:

- any $d+1$ distinct points of $S$ form a $d$-simplex.
- if $d+2$ distinct points of $S$ lie on the same circumsphere, then there is another point of $S$ inside that circumsphere.
Delaunay simplex: Convex hull of $d+1$ points of $S \subset \mathbf{E}^{d}$ that (1) is a $d$-simplex, and (2) contains no other points of $S$ in the interior of its circumsphere.
Theorem
Any finite $S \subset \mathbf{E}^{d}$ in general position has a unique tesselation of Delaunay simplexes.


## Delaunay Tesselation (proof sketch)

Given finite set $S=\left\{\mathbf{x}_{1}, \ldots, \mathbf{x}_{m}\right\} \subset \mathbf{E}^{d}$ in general position:
$\checkmark$ define injection $i: \mathbf{E}^{d} \rightarrow \mathbf{E}^{d+1}$ by $i(\mathbf{x})=\left(\mathbf{x},\|\mathbf{x}\|^{2}\right)$, so $i(S)$ is a finite subset of a hyper-paraboloid ("lifted" $S$ ).

- let $S^{\prime}=\operatorname{ConvHull}(i(S)) \subset \mathbf{E}^{d+1}$.
- let $S^{\prime \prime} \subset S^{\prime}$ be the $d$-simplexes that:
- are on the boundary of $S^{\prime}$, and
- have no points of $S^{\prime}$ below them ("bottom facets" of $S^{\prime}$, with smallest $d+1$ coordinates).
- define projection $p: \mathbf{E}^{d+1} \rightarrow \mathbf{E}^{d}$ so that $p \circ i(\mathbf{x})=\mathbf{x}$ for all $\mathbf{x} \in \mathbf{E}^{d}: p\left(x_{1}, \ldots, x_{d}, x_{d+1}\right)=\left(x_{1}, \ldots, x_{d}\right)$.
Then $p\left(S^{\prime \prime}\right)$ is the Delaunay tesselation of $S$.
Exercise: Complete the proof. Hint: see K.Q.Brown (1979).


## Riemannian Metrics and Arc Length

Example metrics for a differentiable manifold $\mathcal{M}$ :

- original d() from the abstract definition of $\mathcal{M}$,
- relative metric from Whitney embedding $\mathcal{M} \rightarrow \mathbf{E}^{N}$,
- Riemann metric defined by a local positive definite inner product $\langle\mathbf{u}, \mathbf{v}\rangle_{x}$ for $\mathbf{u}, \mathbf{v} \in T_{x} \mathcal{M}$ at every $x \in \mathcal{M}$.
(Embedding gives a natural $\langle\mathbf{u}, \mathbf{v}\rangle_{x}$ since $T_{x} \mathcal{M}=\mathbf{E}^{d}$ at each $x$.)
Each inner product defines an arc length, along each differentiable $\gamma:[a, b] \rightarrow \mathcal{M}$, between $p=\gamma(a)$ and $q=\gamma(b)$ :

$$
g_{\gamma}(p, q) \stackrel{\text { def }}{=} \int_{a}^{b} \sqrt{\left\langle\gamma^{\prime}(t), \gamma^{\prime}(t)\right\rangle_{\gamma(t)}} d t
$$

Note that $g_{\gamma}(p, q) \geq 0$.

## Geodesic Distance

For each $p, q \in \mathcal{M}$, let $\Gamma(p, q)$ be the set of differentiable curves $\gamma:[a, b] \rightarrow \mathcal{M}$ with $\gamma(a)=p$ and $\gamma(b)=q$.

Fix $\left\{\langle\cdot, \cdot\rangle_{x}: x \in \mathcal{M}\right\}$. Define the corresponding geodesic distance to be

$$
g(p, q) \stackrel{\text { def }}{=} \min \left\{g_{\gamma}(p, q): \gamma \in \Gamma(p, q)\right\} .
$$

Then $g(p, q)$ is a metric on $\mathcal{M}$. It is called the Riemann metric, though $\langle\cdot, \cdot\rangle_{x}$ also has that name and symbol $g$.

A geodesic between $p, q \in \mathcal{M}$ is any $\gamma \in \Gamma$ with $g_{\gamma}(p, q)=g(p, q)$.
Examples: lines in $\mathbf{E}^{d}$, great circles in $S^{2}$.

## Delaunay Triangulation of a 2-Manifold

Let $\mathcal{M}$ be a differentiable 2-manifold. Fix a Riemann metric $g$. Substitute for plane geometric notions as follows:

- the geodesic between $p$ and $q$ replaces the line segment $\overline{p q}$,
- the geodesic circle $\{x \in \mathcal{M}: g(x, c)=r\}$, of radius $r$ centered at $c$, replaces the plane circle,
- the open geodesic disk $\{x \in \mathcal{M}: g(x, c)<r\}$, uniquely defined by its boundary geodesic circle, replaces $B(c, r)$.
Say that a finite set of points in $\mathcal{M}$ is in general position iff
- no 3 lie on a single geodesic curve, and
- if 4 lie on a single geodesic circle, then a 5th lies inside it.

Theorem
A set in general position has a unique Delaunay triangulation consisting of geodesics.

## Voronoi Cells

For $S=\left\{x_{1}, \ldots, x_{m}\right\} \subset \mathcal{M}$, define the Voronoi cell of $x_{i}$ to be

$$
F_{i} \stackrel{\text { def }}{=}\left\{x \in \mathcal{M}:(\forall j) g\left(x, x_{i}\right) \leq g\left(x, x_{j}\right)\right\}
$$

- If $x$ is closest to some unique $x_{i}$, then $x$ is inside $F_{i}$.
- If $x$ is equidistant from exactly two centers $x_{i} \neq x_{j}$, then $x$ lies on a geodesic curve in $F_{i} \cap F_{j}$.
- If $x$ is equidistant from exactly three centers, then $x$ lies on the intersection of geodesics separating three Voronoi cells.

Exercise: If $S$ is in general position, then no $x \in \mathcal{M}$ can be equidistant from 4 or more centers.

## Voronoi Tesselations on 2-Manifolds

For 2-manifold $\mathcal{M}$, suppose that $S=\left\{x_{i}: i=1, \ldots, m\right\} \subset \mathcal{M}$ is in general position, and let $\left\{F_{i}: i=1, \ldots, m\right\}$ be the corresponding Voronoi cells.

The Voronoi tesselation of $\mathcal{M}$ from $S$ is a graph ( $V, E, F$ ) consisting of Voronoi cells (Faces), geodesic curve segments (Edges) separating pairs of faces, and intersection points (Vertices) equidistant from exactly 3 centers.

Compare with the Delaunay tesselation (triangulation) of $\mathcal{M}$ from $S$, the graph ( $V^{\prime}, E^{\prime}, F^{\prime}$ ) with vertices $V^{\prime}=S$, faces $F^{\prime}$ that are the Delaunay triangles, and edges $E^{\prime}$ that are the sides of the Delaunay triangles.

## Isomorphisms and Duals of Graphs

## Definition

Two graphs $(V, E, F)$ and $(\tilde{V}, \tilde{E}, \tilde{F})$ are isomorphic, written
$(\tilde{V}, \tilde{E}, \tilde{F}) \leftrightarrow(V, E, F)$, iff $V \leftrightarrow \tilde{V}, E \leftrightarrow \tilde{E}$, and $F \leftrightarrow \tilde{F}$.
Definition
( $V, E, F$ ) and ( $V^{\prime}, E^{\prime}, F^{\prime}$ ) are dual graphs iff

- each face in $F$ corresponds to a unique vertex in $V^{\prime}: F \leftrightarrow V^{\prime}$,
- each vertex in $V$ corresponds to a unique face in $F^{\prime}: V \leftrightarrow F^{\prime}$,
- each edge in $E$ corresponds to a unique edge in $E^{\prime}: E \leftrightarrow E^{\prime}$, namely, they are duals iff $(V, E, F) \leftrightarrow\left(F^{\prime}, E^{\prime}, V^{\prime}\right)$.

Exercise: The dual of the dual of $(V, E, F)$ is isomorphic to $(V, E, F)$.

## Voronoi and Delaunay are Duals

## Theorem

For 2-manifold $\mathcal{M}$ and finite subset $S \subset \mathcal{M}$ in general position, the Voronoi tesselation of $\mathcal{M}$ from $S$ is isomorphic to the dual graph of the Delaunay triangulation.

Remark. Voronoi cells give nonoverlapping regions, with piecewise geodesic boundaries, that are analogues of open balls.

Atlas of Voronoi cell charts needs no transition functions.
Delaunay triangulations have Delaunay refinements useful in subdivision and approximation.

Duality allows choice and best-adaption to specific problems.

## Voronoi Tesselation of Planet Earth



Figure: Hrvoje Lukatela, Hipparchus atlas, 1987

## Delaunay Tesselation of a Human Head



Figure: Dyer, Zhang, and Möller, "3D Cutaway," Survey..., 2009

## References

- C. Bradford Barber, David P. Dobkin, and Hannu Huhdanpaa. "The Quickhull Algorithm for Convex Hulls," ACM Trans. on Mathematical Software 22:4(1996),pp.469-483
- Kevin Q. Brown, "Voronoi Diagrams from Convex Hulls," Information Processing Letters, 9:5(1979), pp.223-228.
- Mark de Berg, Otfried Cheong (née Schwarzkopf), Marc van Kreveld, and Mark Overmars. Computational Geometry: Algorithms and Applications, 3rd Ed., Springer-Verlag, 2008.
- Ramsay Dyer, Hao Zhang, and Torsten Möller. "A survey of Delaunay structures for surface representation." GrUVi Lab, School of Computing Science, Simon Fraser University, 2009. DelaunayStructuresforSurfaces.pdf
- Hrvoje Lukatela. AUTO-CARTO 8, Baltimore, March 1987. http://www.lukatela.com/hrvoje/geomedia.html

