Voronoi diagrams in information geometry

— Statistical Voronoi diagrams and their applications —

Frank NIELSEN

École Polytechnique
Sony CSL
e-mail: `Frank.Nielsen@acm.org`

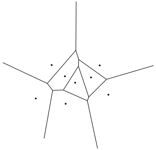
MaxEnt 2014

Outline of the tutorial

1. **Euclidean Voronoi diagrams**
2. Discriminant analysis, **Mahalanobis Voronoi diagrams**, and **anisotropic Voronoi diagrams**
3. **Fisher-Hotelling-Rao Voronoi diagrams** (Riemannian curved geometries)
4. **Kullback-Leibler Voronoi diagrams** (Bregman Voronoi diagrams, dually flat geometries)
5. Bayes' error, Chernoff information and **statistical Voronoi diagrams**

Euclidean (ordinary) Voronoi diagrams

$\mathcal{P} = \{P_1, ..., P_n\}$: n distinct **point generators** in Euclidean space $\mathbb{E}^d$



$$\boxed{V(P_i) = \{X : D_E(P_i, X) \leq D_E(P_j, X),\ \forall j \neq i\} = \cap_{i=1}^{n} \mathrm{Bi}^+(P_i, P_j)}$$

$D_E(P, Q) = \|\theta(P) - \theta(Q)\|_2 = \sqrt{\sum_{i=1}^{d} (\theta_i(P) - \theta_i(Q))^2}$

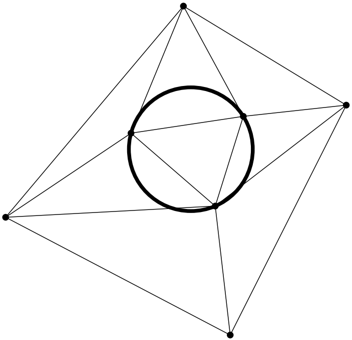
$\theta(P) = p$: Cartesian coordinate system with $\theta_j(P_i) = p_i^{(j)}$.

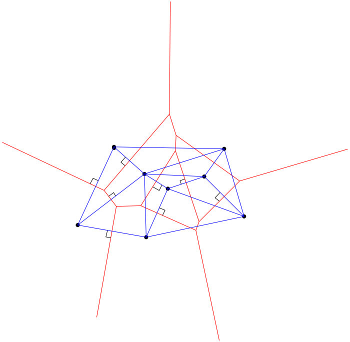
Bisectors $\mathrm{Bi}(P, Q) = \{X : D_E(P, X) \leq D_E(Q, X)\}$: **hyperplanes**

Voronoi diagram = **cell complex** $V(P_i)$**'s with their faces**

$\Rightarrow$ Many applications : crystal growth, codebook/quantization, molecule interfaces/docking, motion planning, etc.

Voronoi diagrams and dual Delaunay simplicial complex

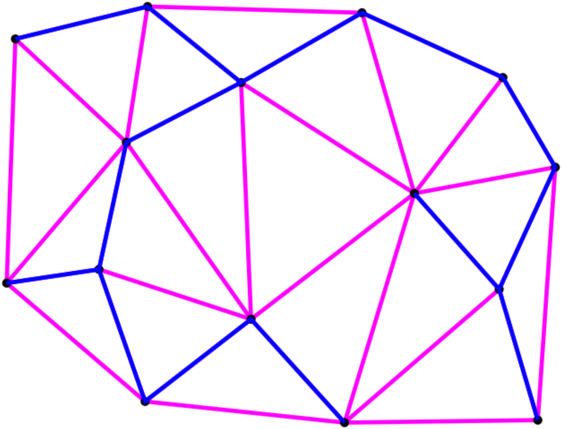




- $\Rightarrow$ **Empty sphere** property, **max min angle** triangulation, etc
- $\Rightarrow$ General position : no $(d+2)$ points cospherical
- $\Rightarrow$ Voronoi & dual **Delaunay triangulation**
- $\Rightarrow$ Bisector $\mathrm{Bi}(P, Q)$ **perpendicular** $\perp$ to segment $[PQ]$

Minimum spanning tree $\subset$ Delaunay triangulation

All edges of the Euclidean MST are Delaunay edges
$\rightarrow$ Prim's greedy algorithm in Delaunay graph in $O(n \log n)$

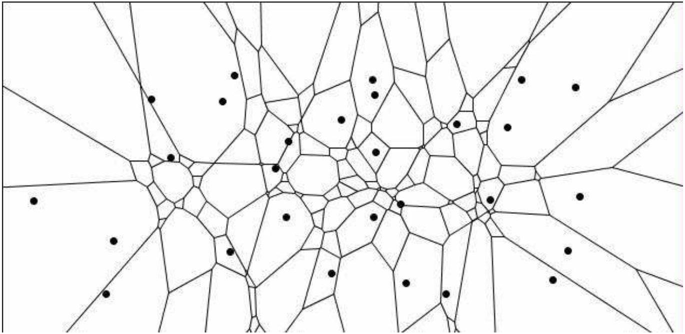


Convex hull : boundary ∂ of the Delaunay triangulation

Order k Voronoi diagrams

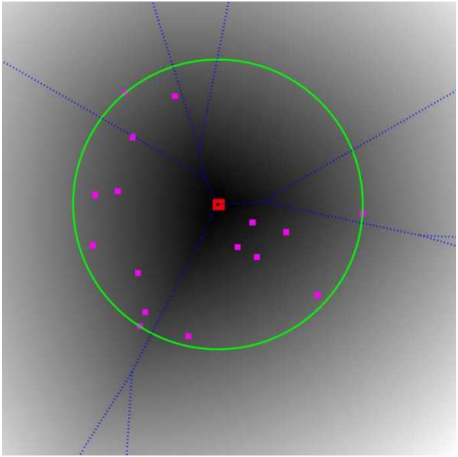
All subsets of size k, $\mathcal{P}_k = \binom{\mathcal{P}}{k} = \{\mathcal{K}_1, ..., \mathcal{K}_N\}$ with $N = \binom{n}{k}$.
partition the space into **non-empty** k**-order Voronoi cells** :

$$\boxed{\mathrm{Vor}_k(\mathcal{K}_i) = \{x : \forall q \in \mathcal{K}_i, \forall r \in \mathcal{P} \backslash \mathcal{K}_i,\ D(x,q) \leq D(x,r)\}}$$



Combinatorial complexity not yet settled ! ! ! (k-sets/levels)
$\equiv$ projection of k-levels of arrangements of hyperplanes in $\mathbb{R}^{d+1}$.

Farthest order Voronoi and **Minimum Enclosing Ball**



Circumcenter/minmax center on the farthest Voronoi diagram.
Non-differentiable optim. at the farthest Voronoi [16]

Voronoi & Delaunay : Complexity and algorithms

- Complexity : $\Theta(n^{\lceil \frac{d}{2} \rceil})$ (quadratic in 3D)
  Moment curve : $t \mapsto (t, t^2, .., t^d)$, etc.
- Construction : $\Theta(n \log n + n^{\lceil \frac{d}{2} \rceil})$, output-sensitive algorithms $\Omega(n \log n + f)$, **not yet optimal output-sensitive algorithms**.
- Voronoi diagram : $\boxed{\mathrm{Vor}_D(\mathcal{P}) = \mathrm{Vor}_{f(D(\cdot,\cdot))}(\mathcal{P})}$ for a strictly monotonically increasing function $f(\cdot)$
  Vor. Euclidean $\equiv$ Vor. squared Euclidean $\subset$ Vor. Bregman
- Geometric toolbox : **space of spheres** (polarity, orthogonality between spheres, etc.)

Multiple Hypothesis Testing [9, 10]

Given a random variable X with n hypothesis

$$\begin{aligned} H_1 &: X \sim P_1 \\ &: \ldots \\ H_n &: X \sim P_n \end{aligned}$$

decide for a IID sample $x_1, \ldots, x_m \sim X$ which hypothesis holds true ?

$$P^m_{\mathrm{correct}} = 1 - P^m_{\mathrm{error}}$$

Asymptotic regime : **Error exponent** α

$$\boxed{\lim_{m\rightarrow\infty} -\frac{1}{m} \log P_e^m = \alpha}$$

Bayesian hypothesis testing

- Prior probabilities : $w_i = \mathbb{P}(X \sim P_i) > 0$ (with $\sum_{i=1}^n w_i = 1$)
- Conditional probabilities : $\mathbb{P}(X = x | X \sim P_i)$.

$$\mathbb{P}(X = x) = \sum_{i=1}^{n} \mathbb{P}(X \sim P_i) \mathbb{P}(X = x | X \sim P_i) = \sum_{i=1}^{n} w_i \mathbb{P}(X | P_i)$$

- Cost *design* matrix $[c_{ij}]$= with $c_{i,j}$ = cost of deciding H_i when in fact H_j is true
- $p_{i,j}(u(x))$ = probability of making this decision using **rule** u.

Bayesian detector : MAP rule

- Minimize the *expected cost* :

$$E_X[c(r(x))], \quad c(r(x)) = \sum_i \left(w_i \sum_{j \neq i} c_{i,j} p_{i,j}(r(x)) \right)$$

- Special case : Probability of error P_e : $c_{i,i} = 0$ and $c_{i,j} = 1$ for $i \neq j$:

$$\boxed{P_e = E_X \left[\sum_i \left(w_i \sum_{j \neq i} p_{i,j}(r(x)) \right) \right]}$$

- Maximum *a posteriori* probability (MAP) rule :

$$\boxed{\mathrm{map}(x) = \mathrm{argmax}_{i \in \{1,\dots,n\}}\ w_i p_i(x)}$$

where $p_i(x) = \mathbb{P}(X = x | X \sim P_i)$ are the conditional probabilities.

$\rightarrow$ MAP Bayesian detector minimizes P_e over all **rules** [7]

MVNs, MAP rule, and Mahalanobis distance (1930)

MultiVariate Normals

(MVNs)

- Say, $X_1 \sim N(\mu_1, \Sigma)$ and $X_2 \sim N(\mu_2, \Sigma)$
- Probability of error (misclassification) for $w_1 = w_2 = \frac{1}{2}$:

$$\boxed{P_e = \frac{1}{2}(1 - \mathrm{TV}(P_1, P_2)) = \Phi\left(-\frac{1}{2}\Delta(P_1, P_2)\right)}$$

$\Phi(\cdot)$: standard normal distribution function

- Mahalanobis metric distance $\Delta(\cdot, \cdot)$:

$$\boxed{\Delta^2(X_1, X_2) = (\mu_1 - \mu_2)^\top \Sigma^{-1} (\mu_1 - \mu_2) = \Delta\mu^\top \Sigma^{-1} \Delta\mu}$$

- Measure the density overlapping (the greater, the lesser)
- generalize Euclidean distance ($\Sigma = I$)
- Δ^2 : only symmetric Bregman divergence.

$\Rightarrow$ easy to approximate $\tilde{P}_e$ experimentally

Mahalanobis Voronoi diagrams

MHT for isotropic MVNs with uniform weight $w_i = \frac{1}{n}$

Cholesky decomposition $\Sigma = LL^\top$

$$
\begin{aligned}
\Delta^2(X_1, X_2) &= (\mu_1 - \mu_2)^\top \Sigma^{-1} (\mu_1 - \mu_2), \\
&= (\mu_1 - \mu_2)^\top (L^{-1})^\top L^{-1} (\mu_1 - \mu_1), \\
&= (L^{-1}\mu_1 - L^{-1}\mu_2)^\top (L^{-1}\mu_1 - L^{-1}\mu_2)
\end{aligned}
$$

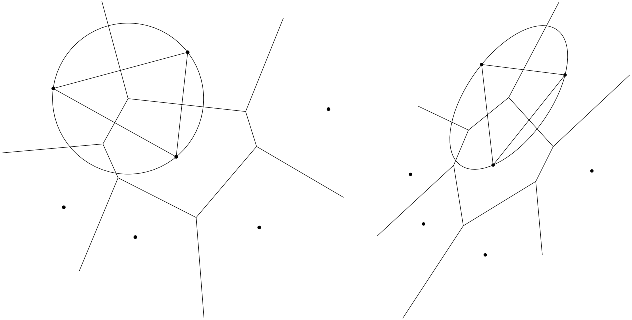
$$\boxed{\Delta^2(X_1, X_2) = D_E(L^{-1}\mu_1, L^{-1}\mu_2)}$$

Mahalanobis Voronoi diagram :

- Cholesky decomposition : $\Sigma = LL^\top$
- Map $\mathcal{P} = \{p_1, ..., p_n\}$ to $\mathcal{P}' = \{p'_1, ..., p'_n\}$ with $p'_i = L^{-1}x$.
- Ordinary Voronoi diagram $\mathrm{Vor}_E(\mathcal{P}')$
- Map $\mathrm{Vor}_E(\mathcal{P}')$ to $\mathrm{Vor}_\Sigma(\mathcal{P}) = L\mathrm{Vor}_E(\mathcal{P}')$.

Mahalanobis Voronoi diagrams

Σ account for both **correlation** and dimension (feature) **scaling**



Dual structure $\equiv$ anisotropic Delaunay triangulation
$\Rightarrow$ "empty circumellipse" property
Recall : MAP rule from Voronoi diagram

Anisotropic Voronoi diagrams [6] (MHT for MVNs)

Classification : Generator $X_i \sim N(\mu_i, \Sigma_i)$, $w_i = \frac{1}{n}$.

Discriminant functions are **quadratic bisectors**
Orphans, islands, dual not a triangulation
(need wedged/visibility conditions)
Applications : Crystal growth in fields (not Riemannian smooth metric)

Statistics : Estimators

Given $x_1, ..., x_n$ IID observations (from a population), estimate the underlying distribution $X \sim p$.

- **empirical CDF** : $\hat{F}_n(x) = \frac{1}{n}\sum_i I_{(-\infty,x)}(x_i)$.
  Glivenko-Cantelli theorem : $\sup_x |\hat{F}_n(x) - F(x)| \rightarrow 0$ a.s.
- Fisher approach : Density p belongs to a **parametric family** $p(x|\theta)$, $D = \#$parameters (order)
  - **Method of moments** : Match distribution moments with sample moments :

$$\mathbb{E}_X[X^l] = \frac{1}{n}\sum_{i=1}^{n} x_i^l,$$

    $\Rightarrow$ any D independent equations yields an estimator
  - Fisher Maximum (log-)Likelihood Estimator **MLE** :

$$\max_\theta l(\theta; x_1, ..., x_n) = \max_\theta \prod_{i=1}^{n} p(x_i;\theta)$$

Estimation : Variance Lower Bound

Which estimator $\hat{\theta}_n$ shall we choose ?

Estimator is also a random variable on a random vector

- Consistent : $\lim_{n\rightarrow\infty} \hat{\theta}_n \rightarrow \theta$
- Unbiased : $\mathbb{E}_\theta[\hat{\theta}_n] - \theta = 0$
- Minimum square error (MSE) : $\mathbb{E}[(\hat{\theta} - \theta)^2] = \mathbb{B}[\hat{\theta}]^2 + \mathbb{V}[\hat{\theta}]$
- **Fisher information matrix** :

$$\boxed{I(\theta) = \left[I_{i,j}(\theta) = \mathbb{E}_\theta \left[\frac{\partial}{\partial\theta_i} \log p(x|\theta) \frac{\partial}{\partial\theta_j} \log p(x|\theta) \right] \right]}$$

- **Fréchet Darmois Cramér-Rao lower bound** for unbiased estimators :

$$\boxed{\mathbb{V}[\hat{\theta}_n] \succeq \frac{1}{n} I^{-1}(\theta)}$$

- Efficient : estimator reaches the Cramèr-Rao lower bound

Sufficient statistics and exponential families

- **Sufficient statistic** $t(x)$:

$$\boxed{\mathbb{P}(x|t(x)=t,\theta)=\mathbb{P}(x|t(x)=t)}$$

  All information for inference contained in $t(x)$. ($\neq$ ancillary)
- For univariate Gaussians, $D=2$: $t_1(x)=x$ and $t_2(x)=x^2$.

$$\hat{\mu}=\bar{x}=\frac{1}{n}\sum_i x_i,\quad \hat{\sigma}^2=\frac{1}{n}\sum_i(x_i-\bar{x})^2=\left(\frac{1}{n}\sum_i x_i^2\right)-(\bar{x})^2$$

- **Exponential families** have finite dimensional sufficient statistics (Koopman) : Reduce n data to D statistics.

$$\boxed{\forall x\in\mathcal{X},\ \mathbb{P}(x|\theta)=\exp(\theta^\top t(x)-F(\theta)+k(x))}$$

  $F(\cdot)$: log-normalizer/cumulant/partition function
  $k(x)$: auxiliary term for carrier measure

Exponential families : Fisher information and MLE

Common distributions are exponential families :
Poisson, Gaussians, Gamma, Beta, Dirichlet, etc.

- Fisher information :

$$I(\theta) = \nabla^2 F(\theta)$$

- MLE for an exponential family :

$$\nabla F(\theta) = \frac{1}{n} \sum_i t(X_i) = \eta$$

MLE exists iff.

$$\boxed{\eta = \nabla F(\theta) \in \mathrm{int}(\mathrm{CH}(\mathcal{X}))}$$

Population space : Hotelling (1930) [5] & Rao (1945) [25]

Birth of **differential-geometric methods in statistics**.

- Fisher information matrix (positive definite) can be used as a (smooth) Riemannian metric tensor.
- Distance between two populations indexed by θ_1 and θ_2 : Riemannian distance (metric length)
- Fisher-Hotelling-Rao (FHR) geodesic distance used in classification : Find the closest population to a given population
- Used in tests of significance (null versus alternative hypothesis), power of a test : $\mathbb{P}(\text{reject } H_0 | H_0 \text{ is false })$ (surfaces in population spaces)

Rao's distance (1945, introduced by Hotelling 1930 [5])

- Infinitesimal length element :

$$\mathrm{d}s^2 = \sum_{ij} g_{ij}(\theta)\mathrm{d}\theta_i \mathrm{d}\theta_j = \mathrm{d}\theta^T I(\theta)\mathrm{d}\theta$$

- Geodesic and distance are hard to explicitly calculate :

$$\boxed{\rho(p(x;\theta_1), p(x;\theta_2)) = \min_{\substack{\theta(s)\\ \theta(0)=\theta_1\\ \theta(1)=\theta_2}} \int_0^1 \sqrt{\left(\frac{\mathrm{d}\theta}{\mathrm{d}s}\right)^T I(\theta)\frac{\mathrm{d}\theta}{\mathrm{d}s}}\mathrm{d}s}$$

- Metric property of ρ, many tools [1] : Riemannian Log/Exp tangent/manifold mapping

Fisher Rao Hotelling Voronoi : Riemannian Voronoi diagrams

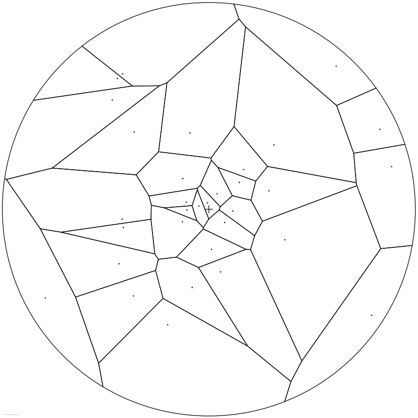
- Location-scale $2D$ families have *non-positive curvature* (Hotelling, 1930) : **FHR Voronoi diagrams** amount to **hyperbolic Voronoi diagrams** or Euclidean diagrams (location families only like isotropic Gaussians)
- Multinomial family has **spherical** geometry on the positive orthant : **Spherical Voronoi diagram** (stereographic projection $\propto$ Euclidean Voronoi diagrams)
- Arbitrary families $p(x|\theta)$: Geodesics not in closed forms $\rightarrow$ limited computational framework in practice...

Hyperbolic Voronoi diagrams [19, 22]

- In Klein disk, the hyperbolic Voronoi diagram amounts to a **clipped affine Voronoi diagram**, or a **clipped power diagram**. Efficient clipping algorithm [2].
- Convert to other models of hyperbolic geometry : Poincaré disk, upper half space, hyperboloid, **Beltrami** hemisphere, etc.
- **Conformal** (good for vizualizing) versus **non-conformal** (good for computing) models. (conformal metric $G(x) = \lambda(x)I$ is scaled identity metric)

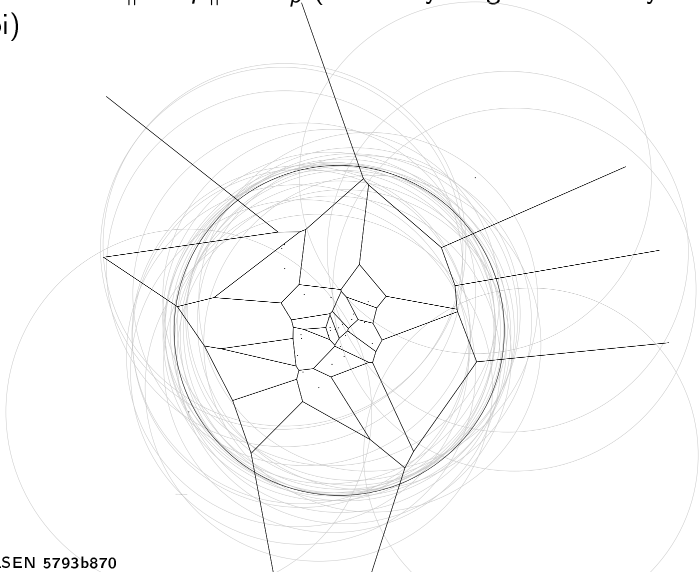
Hyperbolic Voronoi diagrams [19, 22]

Hyperbolic Voronoi diagram in Klein disk :



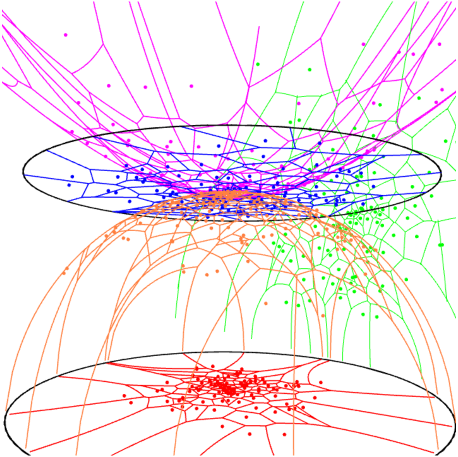
Hyperbolic Voronoi diagrams [19, 22]

Affine Voronoi diagram equivalent to power diagram.
Power distance : $\|x - p\|^2 - w_p$ (additively weighted ordinary Voronoi)



Hyperbolic Voronoi diagrams [19, 22]

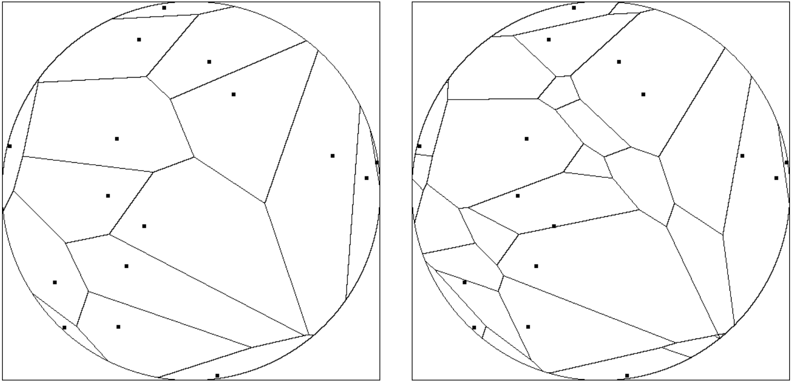
5 common models of the abstract hyperbolic geometry



https://www.youtube.com/watch?v=i9IUzNxeH4o

ACM Symposium on Computational Geometry (SoCG'14)

k-order hyperbolic Voronoi diagram



$k = 1$ $\qquad$ $k = 2$

Affine diagram $\Rightarrow$ equivalent to a power diagram [21]
Watch k-order Voronoi :
`https://www.youtube.com/watch?v=i9IUzNxeH4o`

Constrast function and dually flat space

- Convex and strictly differentiable function $F(\theta)$ admits a Legendre-Fenchel convex conjugate $F^*(\eta)$:

$$\boxed{F^*(\eta) = \sup_{\theta}(\theta^\top \eta - F(\theta)), \quad \nabla F(\theta) = \eta = (\nabla F^*)^{-1}(\theta)}$$

- Young's inequality gives rise to **canonical divergence** [8] :

$$F(\theta) + F^*(\eta') \geq \theta^\top \eta' \Rightarrow \boxed{A_{F,F^*}(\theta, \eta') = F(\theta) + F^*(\eta') - \theta^\top \eta'}$$

- Writing using single coordinate system, get dual **Bregman divergences** :

$$\begin{aligned} B_F(\theta_p : \theta_q) &= F(\theta_p) - F(\theta_q) - (\theta_p - \theta_q)^\top \nabla F(\theta_q) \\ &= B_{F^*}(\eta_q : \eta_p) \\ &= A_{F,F^*}(\theta_p, \eta_q) = A_{F^*,F}(\eta_q : \theta_p) \end{aligned}$$

Discriminant analysis exponential families

- Consider $X_1 \sim E_F(\theta_1)$, $X_2 \sim E_F(\theta_2)$ (or moment parameterization $\eta_i = \nabla F(\theta_i)$)
- Bayes' rule : Classify x from X_1 iff. $\mathbb{P}(x|\theta_1) > \mathbb{P}(x|\theta_2)$
- Use bijection between exponential families and Bregman divergences :

$$\boxed{\log p(x|\theta) = -B_{F^*}(t(x) : \eta) + F^*(t(x)) + k(x)}$$

- MAP rule ($w_1 = w_2 = \frac{1}{2}$) :

$$\boxed{B_{F^*}(t(x) : \eta_1) < B_{F^*}(t(x) : \eta_2)}$$

- Bregman bisector :

$$\mathrm{Bi}_{F^*}(\eta_1, \eta_2) = \{\eta \mid B_{F^*}(\eta : \eta_1) = B_{F^*}(\eta : \eta_1)\}$$

Bregman dual bisectors [3, 17, 20]

Right-sided bisector : $\rightarrow$ Hyperplane (θ-hyperplane)

$$H_F(p,q) = \{x \in \mathcal{X} \mid B_F(x:p) = B_F(x:q)\}.$$

H_F :

$$\boxed{\langle \nabla F(p) - \nabla F(q), x\rangle + (F(p) - F(q) + \langle q, \nabla F(q)\rangle - \langle p, \nabla F(p)\rangle) = 0}$$

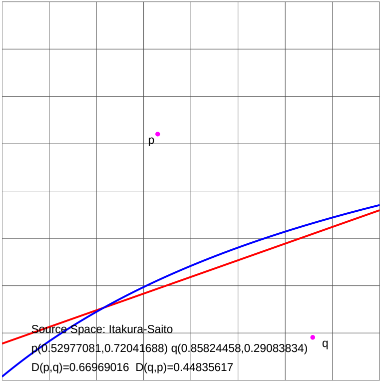
Left-sided bisector : $\rightarrow$ θ-Hypersurface (η-hyperplane)

$$H'_F(p,q) = \{x \in \mathcal{X} \mid B_F(p:x) = B_F(q:x)\}$$

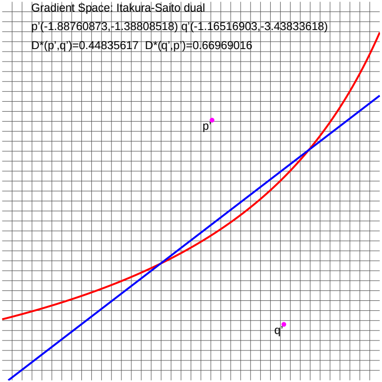
$$\boxed{H'_F : \langle \nabla F(x), q - p\rangle + F(p) - F(q) = 0}$$

Visualizing Bregman bisectors

Primal coordinates θ
natural parameters



Dual coordinates η
expectation parameters



Space of Bregman spheres and Bregman balls [3]

Dual Bregman balls (bounding Bregman spheres) :

$$\begin{aligned} \mathrm{Ball}_F^r(c,r) &= \{x \in \mathcal{X} \mid B_F(x:c) \leq r\} \\ \text{and} \quad \mathrm{Ball}_F^l(c,r) &= \{x \in \mathcal{X} \mid B_F(c:x) \leq r\} \end{aligned}$$

Legendre duality :

$$\boxed{\mathrm{Ball}_F^l(c,r) = (\nabla F)^{-1}(\mathrm{Ball}_{F^*}^r(\nabla F(c),r))}$$

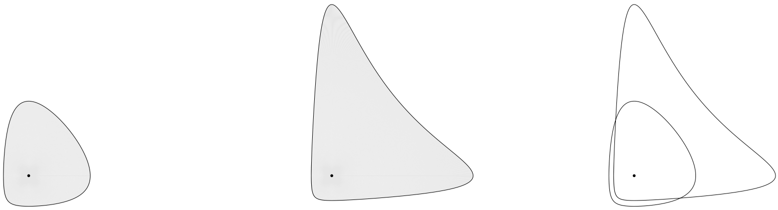
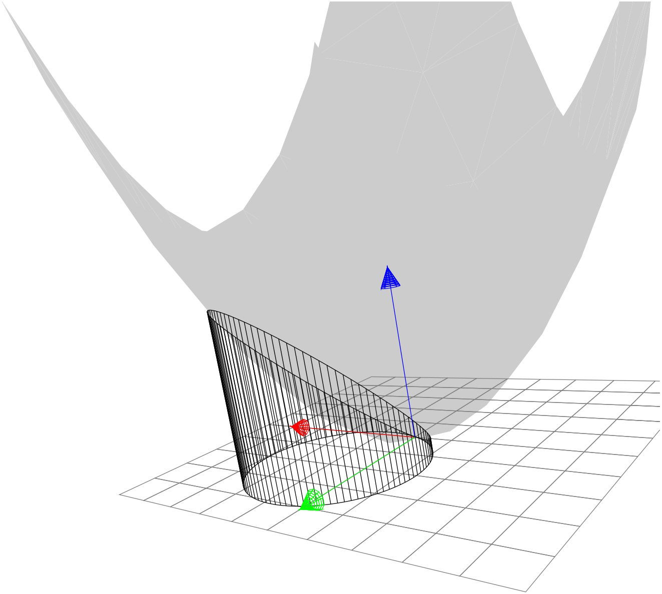


Illustration for Itakura-Saito divergence, $F(x) = -\log x$

Lifting/Polarity : Potential function graph $\mathcal{F}$



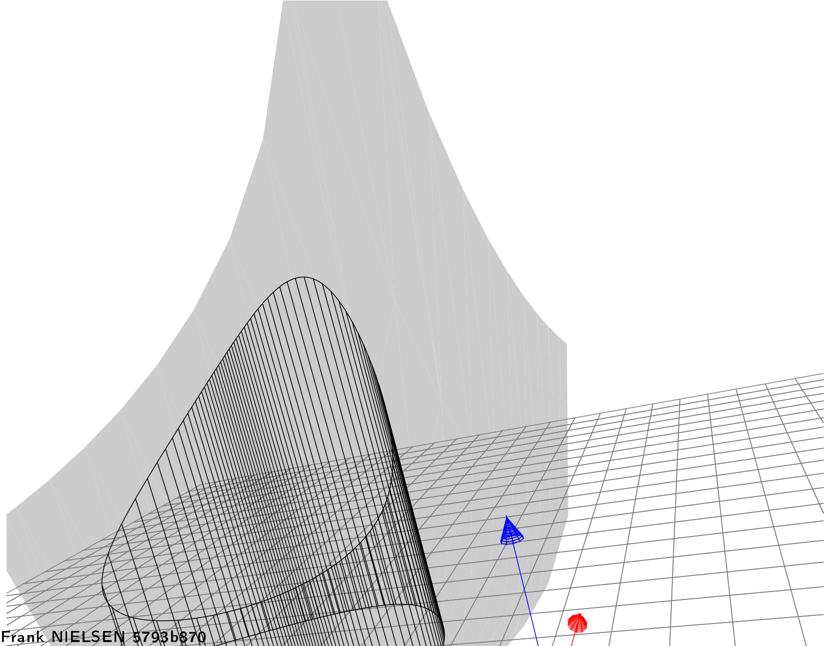
Space of Bregman spheres : Lifting map [3]

$\mathcal{F} : x \mapsto \hat{x} = (x, F(x))$, hypersurface in $\mathbb{R}^{d+1}$.
H_p : Tangent hyperplane at $\hat{p}$, $z = H_p(x) = \langle x - p, \nabla F(p) \rangle + F(p)$

- Bregman sphere $\sigma \longrightarrow \hat{\sigma}$ with supporting hyperplane
  $H_\sigma : z = \langle x - c, \nabla F(c) \rangle + F(c) + r$.
  (// to H_c and shifted vertically by r)
  $\hat{\sigma} = \mathcal{F} \cap H_\sigma$.
- intersection of any hyperplane H with $\mathcal{F}$ projects onto $\mathcal{X}$ as a Bregman sphere :

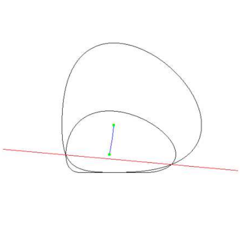
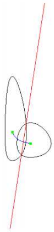
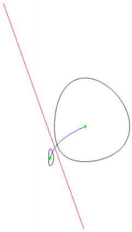
$$H : z = \langle x, a \rangle + b \rightarrow \sigma : \mathrm{Ball}_F(c = (\nabla F)^{-1}(a), r = \langle a, c \rangle - F(c) + b)$$

Lifting with Itakura-Saito potential function



Space of Bregman spheres [3]

- Union/intersection of Bregman spheres from representational polytope [3]
- Radical axis of two Bregman balls is an hyperplane : Applications to Nearest Neighbor search trees like Bregman ball trees or Bregman vantage point trees [23].

Space of spheres : **Minimum enclosing ball** [14, 24]

To a hyperplane $H_\sigma = H(a,b) : z = \langle a, x\rangle + b$ in $\mathbb{R}^{d+1}$, corresponds a ball $\sigma = \mathrm{Ball}(c,r)$ in $\mathbb{R}^d$ with center $c = \nabla F^*(a)$ and radius :

$$r = \langle a, c\rangle - F(c) + b = \langle a, \nabla F^*(a)\rangle - F(\nabla F^*(a)) + b = F^*(a) + b$$

since $F(\nabla F^*(a)) = \langle \nabla F^*(a), a\rangle - F^*(a)$ (Young equality)
SEB : Find halfspace $H(a,b)^- : z \leq \langle a, x\rangle + b$ that contains all lifted points :

$$\min_{a,b} r = F^*(a) + b,$$
$$\forall i \in \{1, ..., n\}, \qquad \langle a, x_i\rangle + b - F(x_i) \geq 0.$$

$\Rightarrow$ **Convex Program (CP) with linear inequality constraints**
$\Rightarrow F(\theta) = F^*(\eta) = \frac{1}{2}x^\top x$: CP $\rightarrow$ **Quadratic Programming** (QP) [4].

InSphere predicates wrt Bregman divergences [3]

Implicit representation of Bregman spheres/balls :

- **Is x inside the Bregman ball defined by $d+1$ support points ?**

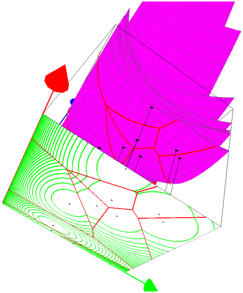
$$\mathrm{InSphere}(x; p_0, ..., p_d) = \begin{vmatrix} 1 & ... & 1 & 1 \\ p_0 & ... & p_d & x \\ F(p_0) & ... & F(p_d) & F(x) \end{vmatrix}$$

- $\mathrm{InSphere}(x; p_0, ..., p_d)$ is negative, null or positive depending on whether x lies inside, on, or outside σ.

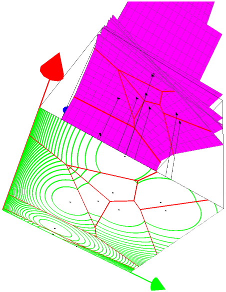
Bregman Voronoi diagrams as minimization diagrams [3]

A subclass of affine diagrams which have all non-empty cells .
Minimization diagram of the n functions
$D_i(x) = B_F(x : p_i) = F(x) - F(p_i) - \langle x - p_i, \nabla F(p_i)\rangle$.
$\equiv$ minimization of n linear functions :

$$\boxed{H_i(x) = (p_i - x)^T \nabla F(q_i) - F(p_i)}$$



$\Longleftrightarrow$



Watch `https://www.youtube.com/watch?v=L7v1wuN_9Wg`

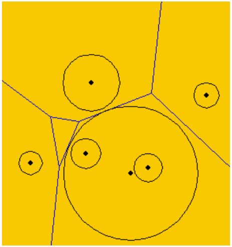
Bregman Voronoi from Power diagrams [15, 20]

Any affine diagram can be built from a **power diagram**.
(power diagrams defined in full space $\mathbb{R}^d$)

- Power distance of x to $\mathrm{Ball}(p, r)$: $||p - x||^2 - r^2$.
- Laguerre diagram : minimization diagram of $D_i(x) = ||p_i - x||^2 - r_i^2$
- Power bisector of $\mathrm{Ball}(p_i, r_i)$ and $\mathrm{Ball}(p_j, r_j)$= radical hyperplane :

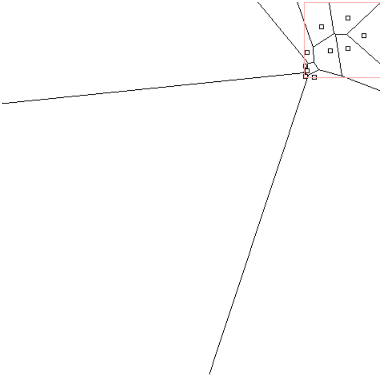
$$2\langle x, p_j - p_i\rangle + ||p_i||^2 - ||p_j||^2 + r_j^2 - r_i^2 = 0.$$

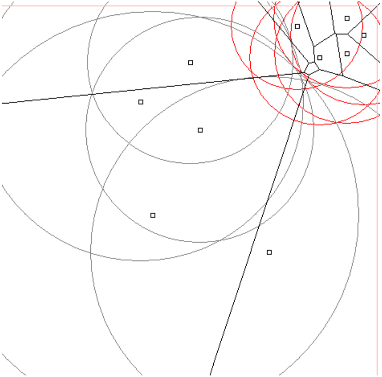
Universality : $\boxed{\text{Affine Bregman Voronoi diagram} \Leftrightarrow \text{Power diagram}}$



Affine Bregman Voronoi diagrams as power diagrams

Equivalence : $B(\nabla F(p_i), r_i)$ with
$r_i^2 = \langle \nabla F(p_i), \nabla F(p_i) \rangle + 2(F(p_i) - \langle p_i, \nabla F(p_i) \rangle)$
(imaginary radii shown in red, WLOG. $r_i \geq 0$ by shifting)

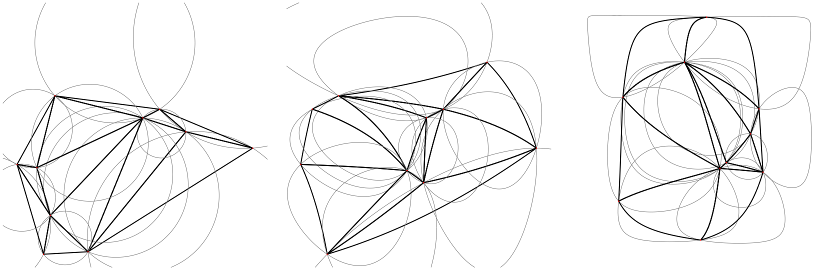




Some cells may be empty in the Laguerre diagram (power diagram) but not in the Bregman diagram

http://www.csl.sony.co.jp/person/nielsen/BVDapplet/

Bregman dual Delaunay triangulations



Delaunay　Exponential　Hellinger-like

- empty Bregman sphere property,
- geodesic triangles.

BVDs extends Euclidean Voronoi diagrams with similar complexity/algorithms.

Riemannian tensor and orthogonality

Dually flat space.

- $F(\theta)$ convex on Θ yields convex conjugate $F^*(\eta)$ on H
- **Same** Riemannian tensor can be expressed in both coordinate systems :

  $$g(\theta) = \nabla^2 F(\theta), \quad g^*(\eta) = \nabla^2 F(\eta), \quad g(\theta(X))g^*(\eta(X)) = I, \forall X$$

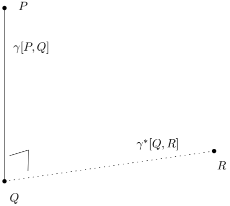
  Same infinitesimal length element $\mathrm{d}s^2 = (\mathrm{d}s^*)^2$.
- Two curves $\gamma_1(t)$ and $\gamma_2(t)$ are orthogonal at $Q = \gamma_1(t_0) = \gamma_2(t_0)$ iff.

  $$\langle \dot\gamma_1(t_0), \dot\gamma_2(t_0) \rangle_{g(\theta(Q))} = \langle \dot\gamma_1(t_0), \dot\gamma_2(t_0) \rangle^*_{g^*(\eta(Q))} = 0$$
- Geodesic $\gamma(P, Q)$ and dual geodesic $\gamma^*(P, Q)$ (and geodesic segments $\gamma[P, Q]$ and $\gamma^*[P, Q]$)

Orthogonality

3-point property (generalized law of cosines) :

$$\boxed{B_F(p:r) = B_F(p:q) + B_F(q:r) - (p-q)^T(\nabla F(r) - \nabla F(q))}$$



$\gamma(P,Q)$ orthogonal to $\gamma^*(Q,R)$ iff.

$$\boxed{B_F(p:r) = B_F(p:q) + B_F(q:r)}$$

Equivalent to $\langle \theta_p - \theta_q, \eta_r - \eta_q \rangle = 0$

Extend Pythagoras' theorem

$$\gamma(P,Q) \perp \gamma^*(Q,R)$$

Dually orthogonal Bregman Voronoi & Triangulations

Ordinary Voronoi diagram is perpendicular to Delaunay triangulation. (Vor k-face $\perp$ Del $d-k$-face)
Dual line segment geodesics :

$$\begin{aligned} \gamma(P,Q) &= \{\theta = \theta_p + (1-\lambda)\theta_q \mid \lambda \in [0,1]\} \\ \gamma^*(P,Q) &= \{\eta = \eta_p + (1-\lambda)\eta_q \mid \lambda \in [0,1]\} \end{aligned}$$

Bisectors :

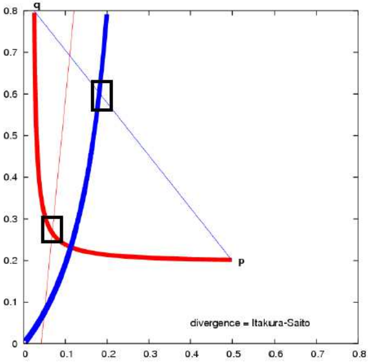
$$\begin{aligned} \mathrm{Bi}_\theta(p,q) &: \langle x, \theta_q - \theta_p \rangle + F(\theta_p) - F(\theta_q) = 0 \\ \mathrm{Bi}_\eta(p,q) &: \langle x, \eta_q - \eta_p \rangle + F^*(\eta_p) - F^*(\eta_q) = 0 \end{aligned}$$

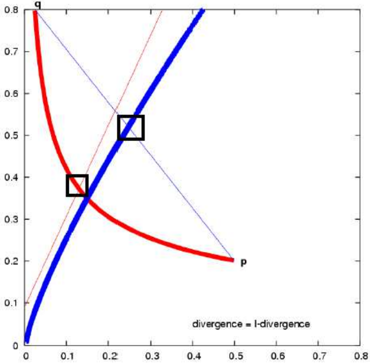
Dual orthogonality :

$$\begin{aligned} \mathrm{Bi}_\eta(p,q) &\perp \gamma^*(P,Q) \\ \gamma(P,Q) &\perp \mathrm{Bi}_\theta(p,q) \end{aligned}$$

Dually orthogonal Bregman Voronoi & Triangulations

$$
\begin{aligned}
\mathrm{Bi}_\eta(p,q) &\perp \gamma^*(P,Q) \\
\gamma(P,Q) &\perp \mathrm{Bi}_\theta(p,q)
\end{aligned}
$$





MAP rule and additive Bregman Voronoi diagrams

Optimal MAP decision rule :

$$
\begin{aligned}
\mathrm{map}(x) &= \mathrm{argmax}_{i \in \{1,\dots,n\}} w_i p_i(x) \\
&= \mathrm{argmax}_{i \in \{1,\dots,n\}} - B^*(t(x) : \eta_i) + \log w_i, \\
&= \mathrm{argmin}_{i \in \{1,\dots,n\}} B^*(t(x) : \eta_i) - \log w_i
\end{aligned}
$$

$\Rightarrow$**dual Bregman Voronoi with additive weights** (affine diagram).
Vertical projection of the intersections of the lifted half-spaces on the potential $\mathcal{F}^*$ shifted by the weights $-\log w_i$.

Relative entropy for exponential families [18]

- Kullback-Leibler divergence (cross-entropy minus entropy) :

$$
\begin{aligned}
\mathrm{KL}(P:Q) &= \int p(x)\log\frac{p(x)}{q(x)}\mathrm{d}x \geq 0 \\
&= \underbrace{\int p(x)\log\frac{1}{q(x)}\mathrm{d}x}_{H^{\times}(P:Q)+c} - \underbrace{\int p(x)\log\frac{1}{p(x)}\mathrm{d}x}_{H(p)=H^{\times}(P:P)-c} \\
&= F(\theta_Q) - F(\theta_P) - \langle \theta_Q - \theta_P, \nabla F(\theta_P)\rangle \\
&= B_F(\theta_Q:\theta_P) = B_{F^*}(\eta_P:\eta_Q)
\end{aligned}
$$

Bregman divergence B_F defined for a strictly convex and differentiable function up to some affine terms.

- Proof $\mathrm{KL}(P:Q) = B_F(\theta_Q:\theta_P)$ follows from

$$X \sim E_F(\theta) \Longrightarrow \boxed{E[t(X)] = \nabla F(\theta)} = \eta$$

Upper bounds P_e using Chernoff Information [11]

- Trick : $\min(a,b) \leq a^\alpha b^{1-\alpha}, \forall \alpha \in (0,1)$ (for $a, b > 0$)

$$E^* = \int \min(\mathbb{P}(C_1|x), \mathbb{P}(C_2|x)) p(x)\mathrm{d}x \leq w_1^\alpha w_2^{1-\alpha} \int p_1^\alpha(x) p_2^{1-\alpha}(x)\mathrm{d}x$$

- Upper bound the minimum error E^*

$$E^* \leq w_1^\alpha w_2^{1-\alpha} c_\alpha(p_1 : p_2),$$

$c_\alpha(p_1 : p_2) = \int p_1^\alpha(x) p_2^{1-\alpha}(x)\mathrm{d}x$: Chernoff α-coefficient.

$$c^*(p_1 : p_2) = c_{\alpha^*}(p_1 : p_2) = \min_{\alpha \in (0,1)} \int p_1^\alpha(x) p_2^{1-\alpha}(x)\mathrm{d}x.$$

- *Chernoff information* (or Chernoff divergence) :

$$C^*(p_1 : p_2) = C_{\alpha^*}(p_1 : p_2) = \max_{\alpha \in (0,1)} -\log \int p_1^\alpha(x) p_2^{1-\alpha}(x)\mathrm{d}x$$

extend methodoloy with **quasi-arithmetic means** [11]

Chernoff coefficient/information : exponential families

$$
\begin{aligned}
C_\alpha(p:q) &= -\log c_\alpha(p,q) = J_F^{(\alpha)}(\theta_p:\theta_q),\\
c_\alpha(p:q) &= e^{-C_\alpha(p:q)} = e^{-J_F^{(\alpha)}(\theta_p:\theta_q)}.
\end{aligned}
$$

Skewed Jensen divergence (on natural parameters) :

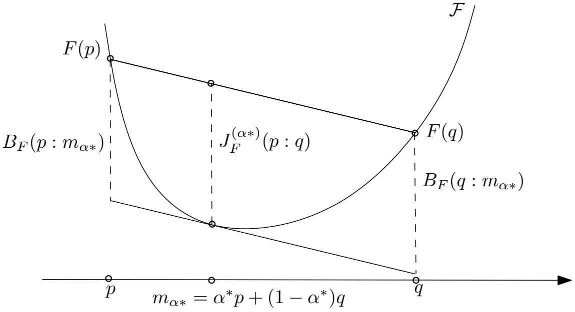
$$
\boxed{J_F^{(\alpha)}(\theta_p:\theta_q) = (\alpha F(\theta_p) + (1-\alpha)F(\theta_q)) - F(\alpha\theta_p + (1-\alpha)\theta_q)}
$$

Maximizing skew Jensen divergence

$$\boxed{\alpha^* = \arg \max_{0<\alpha<1} J_F^{(\alpha)}(p : q)}$$

$$\boxed{J_F^{(\alpha^*)}(p : q) = B_F(p : m_{\alpha^*}) = B_F(q : m_{\alpha^*})}$$

$m_\alpha = \alpha p + (1-\alpha) q$: α-mixing of p and q.



Maximum skew Jensen divergence amounts to Bregman divergences.

Geometry of the best error exponent : binary hypothesis [11]

Chernoff distribution P^* :

$$\boxed{P^* = P_{\theta_{12}^*} = G_e(P_1, P_2) \cap \mathrm{Bi}_m(P_1, P_2)}$$

e-geodesic :

$$G_e(P_1, P_2) = \{E_{12}^{(\lambda)} \mid \theta(E_{12}^{(\lambda)}) = (1-\lambda)\theta_1 + \lambda\theta_2, \lambda \in [0,1]\},$$

m-bisector :

$$\mathrm{Bi}_m(P_1, P_2) : \{P \mid F(\theta_1) - F(\theta_2) + \eta(P)^\top \Delta\theta = 0\},$$

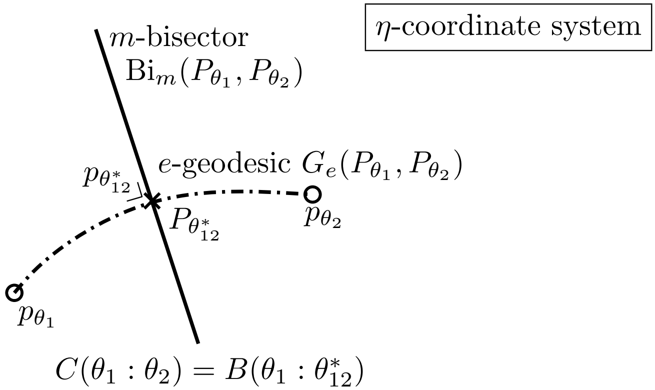
Optimal natural parameter of P^* :

$$\theta^* = \theta_{12}^{(\alpha^*)} = \mathrm{argmin}_{\theta \in \Theta} B(\theta_1 : \theta) = \mathrm{argmin}_{\theta \in \Theta} B(\theta_2 : \theta).$$

$\rightarrow$ closed-form for order-1 family, or efficient bisection search.

Geometry of the best error exponent : binary hypothesis

$$\boxed{P^* = P_{\theta_{12}^*} = G_e(P_1, P_2) \cap \mathrm{Bi}_m(P_1, P_2)}$$



BHT : P_e bounded using Bregman divergence between Chernoff distribution and class-conditional distributions.

Geometry of the best error exponent : multiple hypothesis

n-ary MHT [7] from *minimum pairwise Chernoff distance* :

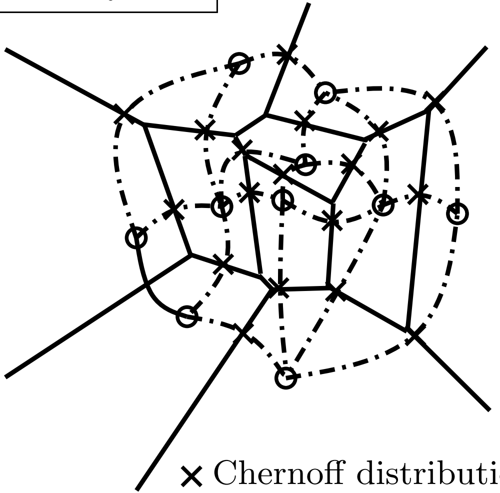
$$\boxed{C(P_1, ..., P_n) = \min_{i, j \neq i} C(P_i, P_j)}$$

$$P_e^m \leq e^{-mC(P_{i^*}, P_{j^*})}, \quad (i^*, j^*) = \mathrm{argmin}_{i, j \neq i} C(P_i, P_j)$$

Compute for each pair of natural neighbors P_{θ_i} and P_{θ_j}, the Chernoff distance $C(P_{\theta_i}, P_{\theta_j})$, and choose the pair with minimal distance. (Proof by contradiction using Bregman Pythagoras theorem.)

Multiple Hypothesis testing & Chernoff information

$\boxed{\eta\text{-coordinate system}}$



× Chernoff distribution between natural neighbours

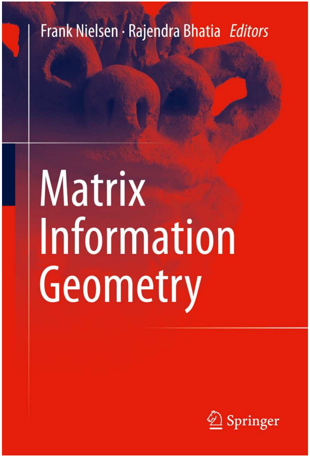
Summary : Statistical Voronoi diagrams

- Mahalanobis Voronoi diagrams, anisotropic Voronoi diagrams (additively-weighted)
- Fisher-Hotelling-Rao Riemannian Voronoi diagrams
- Kullback-Leibler Voronoi diagrams for exponential families = Bregman Voronoi diagrams
  - Extend ordinary Voronoi diagrams ($\equiv$ isotropic Gaussians)
  - Space of spheres : Lifting/polarity
  - bisectors $\perp$ geodesics
  - dual regular triangulation with empty spheres

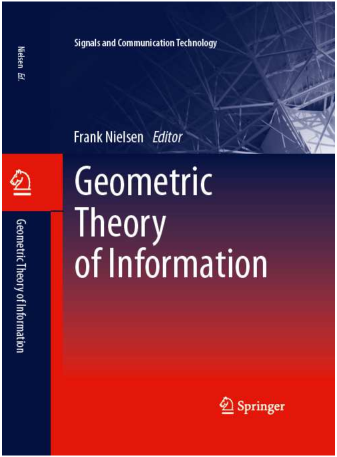
Information geometry : affine differential geometry of "parameter spaces", invariance principles

Many other kinds of Voronoi : Jensen-Bregman, (u, v)-structures , conformal divergences, total Bregman & total Jensen divergences, etc.

Computational Information Geometry : Edited books



[13]



[12]

http://www.springer.com/engineering/signals/book/978-3-642-30231-2
http://www.sonycsl.co.jp/person/nielsen/infogeo/MIG/MIGBOOKWEB/

http://www.springer.com/engineering/signals/book/978-3-319-05316-5
http://www.sonycsl.co.jp/person/nielsen/infogeo/GTI/GeometricTheoryOfInformation.html

Marc Arnaudon and Frank Nielsen.
On approximating the Riemannian 1-center.
Comput. Geom. Theory Appl., 46(1) :93–104, January 2013.

Jean-Daniel Boissonnat and Christophe Delage.
Convex hull and Voronoi diagram of additively weighted points.
In Gerth StÃÿlting Brodal and Stefano Leonardi, editors, *ESA*, volume 3669 of *Lecture Notes in Computer Science*, pages 367–378. Springer, 2005.

Jean-Daniel Boissonnat, Frank Nielsen, and Richard Nock.
Bregman Voronoi diagrams.
Discrete and Computational Geometry, 44(2) :281–307, April 2010.

Bernd Gärtner and Sven Schönherr.
An efficient, exact, and generic quadratic programming solver for geometric optimization.
In *Proceedings of the sixteenth annual symposium on Computational geometry*, pages 110–118. ACM, 2000.

Harold Hotelling.

Francois Labelle and Jonathan Richard Shewchuk.
Anisotropic Voronoi diagrams and guaranteed-quality anisotropic mesh generation.
In *Proceedings of the nineteenth annual symposium on Computational geometry*, pages 191–200. ACM, 2003.

C. C. Leang and D. H. Johnson.
On the asymptotics of M-hypothesis Bayesian detection.
IEEE Transactions on Information Theory, 43(1) :280–282, January 1997.

Frank Nielsen.
Legendre transformation and information geometry.
Technical Report CIG-MEMO2, September 2010.

Frank Nielsen.
Hypothesis testing, information divergence and computational geometry.

In Frank Nielsen and FrÃĺdÃĺric Barbaresco, editors, *Geometric Science of Information*, volume 8085 of *Lecture Notes in Computer Science*, pages 241–248. Springer Berlin Heidelberg, 2013.

Frank Nielsen.
Generalized bhattacharyya and chernoff upper bounds on bayes error using quasi-arithmetic means.
Pattern Recognition Letters, 42 :25–34, 2014.

Frank Nielsen.
Generalized bhattacharyya and chernoff upper bounds on bayes error using quasi-arithmetic means.
Pattern Recognition Letters, 42 :25–34, 2014.

Frank Nielsen.
Geometric Theory of Information.
Springer, 2014.

Frank Nielsen and Rajendra Bhatia, editors.
Matrix Information Geometry (Revised Invited Papers). Springer, 2012.

Frank Nielsen and Richard Nock.
On the smallest enclosing information disk.
Information Processing Letters (IPL), 105(3) :93–97, 2008.

Frank Nielsen and Richard Nock.
Quantum Voronoi diagrams and Holevo channel capacity for 1-qubit quantum states.
In *IEEE International Symposium on Information Theory (ISIT)*, pages 96–100, Toronto, Canada, July 2008. IEEE.

Frank Nielsen and Richard Nock.
Approximating smallest enclosing balls with applications to machine learning.
Int. J. Comput. Geometry Appl., 19(5) :389–414, 2009.

Frank Nielsen and Richard Nock.
The dual Voronoi diagrams with respect to representational Bregman divergences.
In *International Symposium on Voronoi Diagrams (ISVD)*, pages 71–78, 2009.

Frank Nielsen and Richard Nock.
Entropies and cross-entropies of exponential families.
In *Proceedings of the International Conference on Image Processing, ICIP 2010, September 26-29, Hong Kong, China*, pages 3621–3624, 2010.

Frank Nielsen and Richard Nock.
Hyperbolic Voronoi diagrams made easy.
In *2013 13th International Conference on Computational Science and Its Applications*, pages 74–80. IEEE, 2010.

Frank Nielsen and Richard Nock.
Hyperbolic Voronoi diagrams made easy.
In *International Conference on Computational Science and its Applications (ICCSA)*, volume 1, pages 74–80, Los Alamitos, CA, USA, march 2010. IEEE Computer Society.

Frank Nielsen and Richard Nock.
Further results on the hyperbolic Voronoi diagrams.
ISVD, 2014.

Frank Nielsen and Richard Nock.
Visualizing hyperbolic Voronoi diagrams.
In *Symposium on Computational Geometry*, page 90, 2014.

Frank Nielsen, Paolo Piro, and Michel Barlaud.
Bregman vantage point trees for efficient nearest neighbor queries.
In *Proceedings of the 2009 IEEE International Conference on Multimedia and Expo (ICME)*, pages 878–881, 2009.

Richard Nock and Frank Nielsen.
Fitting the smallest enclosing bregman balls.
In *16th European Conference on Machine Learning (ECML)*, pages 649–656, October 2005.

Calyampudi Radhakrishna Rao.
Information and the accuracy attainable in the estimation of statistical parameters.
Bulletin of the Calcutta Mathematical Society, 37 :81–89, 1945.