# Jensen-Bregman Voronoi diagrams and centroidal tessellations

Frank Nielsen

www.informationgeometry.org

École Polytechnique, LIX, France

Sony Computer Science Laboratories, FRL, Japan

(joint work with Richard Nock)

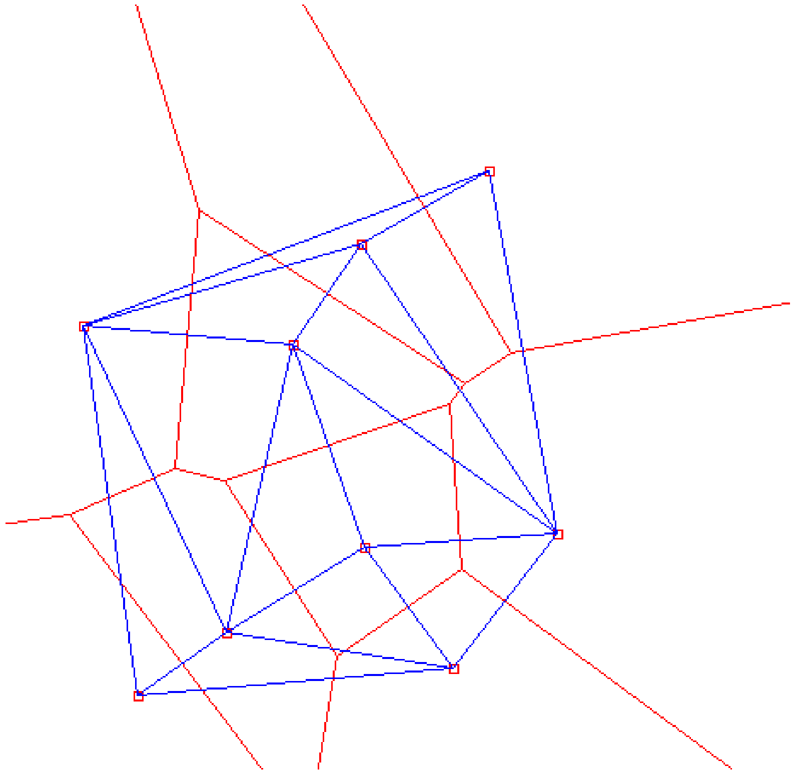
International Workshop on Voronoi Diagrams (ISVD 2010)

June 28, 2010 (Mon.)

Last updated, June 23rd 2010

# Voronoi diagrams

Fundamental combinatorial structure for proximty locations:

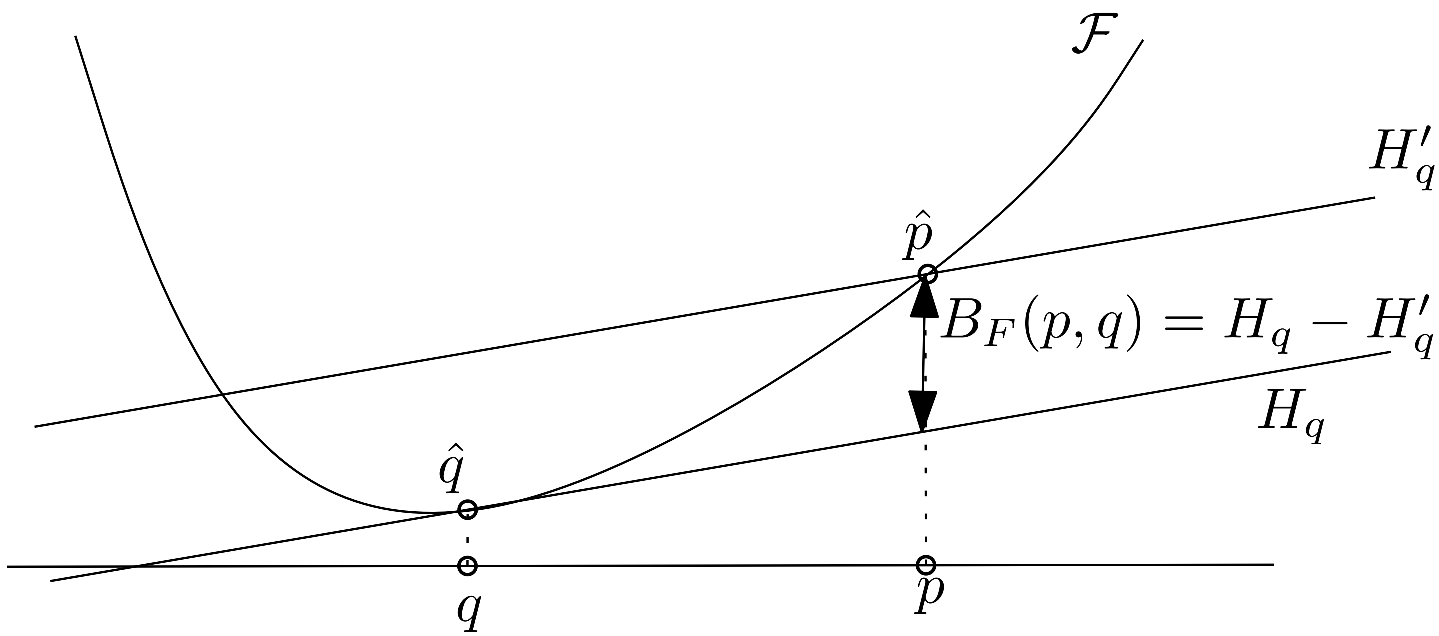


Dual Delaunay triangulation

$\rightarrow$ Extend Voronoi diagrams to Jensen-Bregman divergences.

# Bregman divergences

$$B_F(p,q) = F(p) - F(q) - \langle p-q, \nabla F(q)\rangle,$$




- Kullback-Leibler ($F(x) = x\log x$): $\mathrm{KL}(p,q) = \sum_{i=1}^d p^{(i)} \log \frac{p^{(i)}}{q^{(i)}}$
- Squared Euclidean $L_2^2$ ($F(x) = x^2$): $L_2^2(p,q) = \sum_{i=1}^d (p^{(i)} - q^{(i)})^2 = \|p-q\|^2$

# Symmetrizing Bregman divergences

- **Jeffreys-Bregman divergences.**

$$
\begin{aligned}
S_F(p;q) &= \frac{B_F(p,q)+B_F(q,p)}{2} \\
&= \frac{1}{2}\langle p-q, \nabla F(p) - \nabla F(q)\rangle,
\end{aligned}
$$

- **Jensen-Bregman divergences (diversity index).**

$$
\begin{aligned}
J_F(p;q) &= \frac{B_F(p,\frac{p+q}{2})+B_F(q,\frac{p+q}{2})}{2} \\
&= \frac{F(p)+F(q)}{2} - F\left(\frac{p+q}{2}\right) = \mathrm{BR}_F(p,q)
\end{aligned}
$$

# ensen-Bregman divergences: Burbea-Rao divergences

Based on Jensen's inequality for a convex function $F$:

$$d(x,p) = \frac{F(x)+F(p)}{2} - F\left(\frac{x+p}{2}\right) \stackrel{\text{equal}}{=} \mathrm{BR}_F(x,p) \geq 0.$$

strictly convex function $F(\cdot)$.

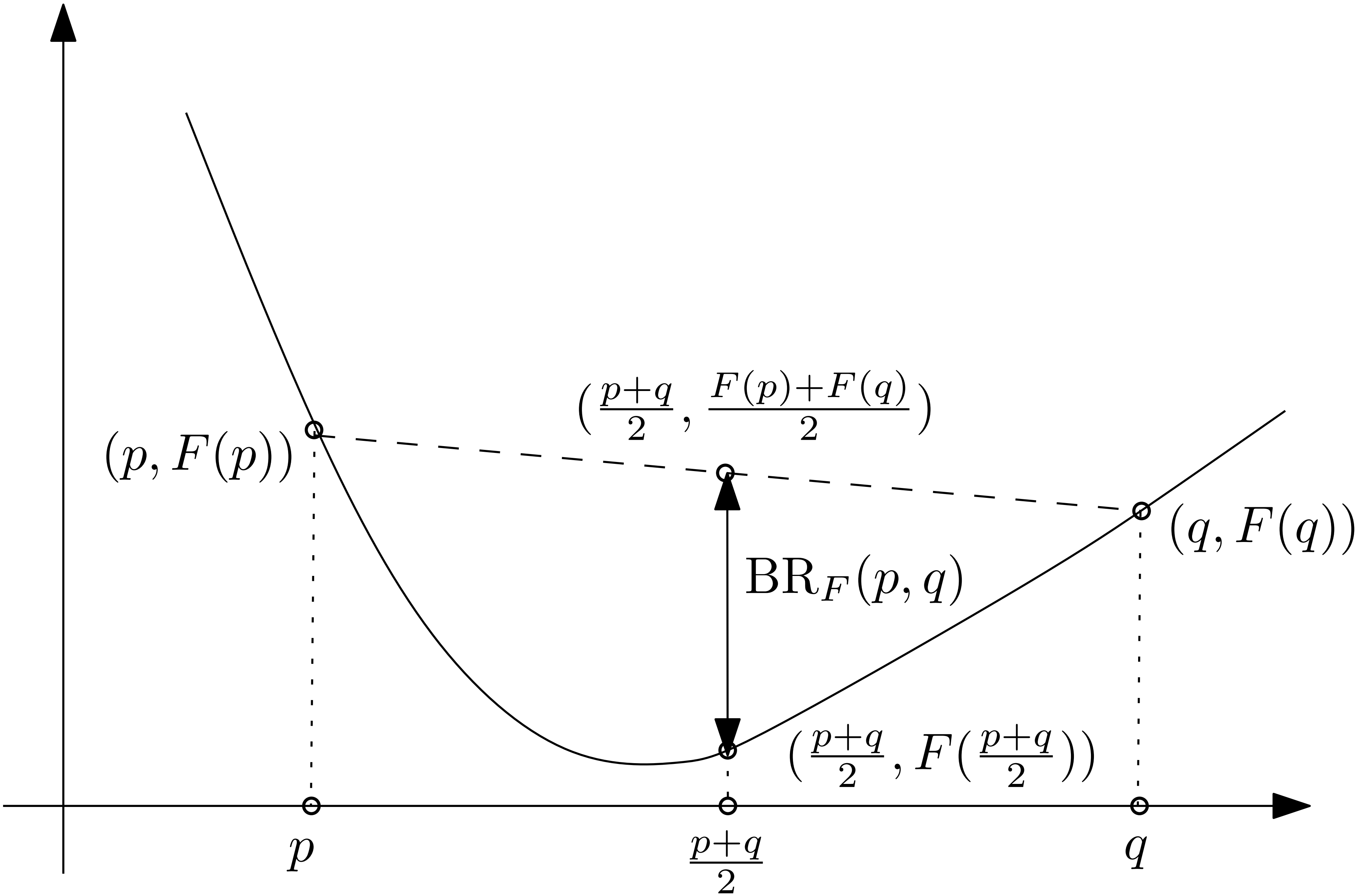
$$\mathrm{BR}_F(p,q) = \sum_{i=1}^{d} \mathrm{BR}_F(p^{(i)}, q^{(i)}),$$

Includes the special case of Jensen-Shannon divergence:

$$\mathrm{JS}(p,q) = H\left(\frac{p+q}{2}\right) - \frac{H(p)+H(q)}{2}$$

$F(x) = -H(x)$, the negative Shannon entropy $H(x) = -x\log x$.
$\rightarrow$ generators are convex and entropies are concave (negative generators)

# Visualizing Burbea-Rao divergences



# Burbea-Rao divergences: Squared Mahalanobis

$$
\begin{aligned}
\mathrm{BR}_F(p,q) &= \frac{F(p)+F(q)}{2} - F\left(\frac{p+q}{2}\right)\\
&= \frac{2\langle Qp,p\rangle + 2\langle Qq,q\rangle - \langle Q(p+q), p+q\rangle}{4}\\
&= \frac{1}{4}(\langle Qp,p\rangle + \langle Qq,q\rangle - 2\langle Qp,q\rangle)\\
&= \frac{1}{4}\langle Q(p-q), p-q\rangle = \frac{1}{4}\|p-q\|_Q^2.
\end{aligned}
$$

(Not a metric. square root of Jensen-Shannon is a metric but not the square roots of all Burbea-Rao divergences.)
For $Q = I$, we get the squared Euclidean distance.

$\rightarrow$ Ordinary Voronoi diagrams are a special case of Jensen-Bregman Voronoi diagrams.

# Skew Burbea-Rao divergences

$$\begin{aligned} \mathrm{BR}_F^{(\alpha)} &: \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+ \\ \mathrm{BR}_F^{(\alpha)}(p,q) &= \alpha F(p) + (1-\alpha)F(q) - F(\alpha p + (1-\alpha)q) \end{aligned}$$

$$\begin{aligned} \mathrm{BR}_F^{(\alpha)}(p,q) &= \alpha F(p) + (1-\alpha)F(q) - F(\alpha p + (1-\alpha)q) \\ &= \mathrm{BR}_F^{(1-\alpha)}(q,p) \end{aligned}$$

Skew symmetrization of Bregman divergences:

$$\alpha B_F(p, \alpha p + (1-\alpha)q) + (1-\alpha) B_F(q, \alpha p + (1-\alpha)q) \stackrel{\text{equal}}{=} \mathrm{BR}_F^{(\alpha)}(p,q)$$

= skew Jensen-Bregman divergences.

# Bregman as asymptotic skewed Burbea-Rao

$$B_F(p,q) = \lim_{\alpha\rightarrow 1} \frac{1}{1-\alpha}\mathrm{BR}_F^{(\alpha)}(p,q)$$

$$B_F(q,p) = \lim_{\alpha\rightarrow 0} \frac{1}{\alpha}\mathrm{BR}_F^{(\alpha)}(p,q)$$

Proof: $F(\alpha p+(1-\alpha)q) = F(p+(1-\alpha)(q-p)) \simeq_{\alpha\simeq 1} F(p)+(1-\alpha)(q-p)\nabla F(p)$

Taylor

$F(\alpha p+(1-\alpha)q)-\alpha F(p)-(1-\alpha)F(q) \simeq_{\alpha\rightarrow 1} (1-\alpha)F(p)+(1-\alpha)(q-p)\nabla F(p)-(1-\alpha)F(q)$

$\simeq_{\alpha\rightarrow 1} (1-\alpha)\left(F(p)-F(q)-(p-q)\nabla F(p)\right)$

$\lim_{\alpha\rightarrow 1} \mathrm{BR}_F^{(\alpha)}(p,q) = (1-\alpha)B_F(p,q)$

For $0<\alpha<1$, swap arguments by setting $\alpha\rightarrow 1-\alpha$:

$$\mathrm{BR}_F^{(\alpha)}(p,q) = \mathrm{BR}_F^{(1-\alpha)}(q,p)$$

$\rightarrow$ Extend to arbitrary $\alpha\in\mathbb{R}$ by dividing by $\alpha(1-\alpha)$:

$\mathrm{BR'}_F^{(\alpha)}(p,q) = \frac{1}{\alpha(1-\alpha)}\mathrm{BR}_F^{(\alpha)}(p,q)$ or $\mathrm{BR'}_F^{(\alpha')}(p,q) = \frac{4}{(1-\alpha'^2)}\mathrm{BR}_F^{(\frac{1-\alpha'}{2})}(p,q)$

(with $\alpha = \frac{1-\alpha'}{2}$)

$\rightarrow$ Bregman Voronoi diagrams are special cases of (scaled) skew Jensen-Bregman Voronoi diagrams.

Bregman Voronoi Diagrams, Discrete and Computational Geometry (Springer), 10.1007/s00454-010-9256-1, 2010.

# Non-homogeneous distances

Jensen-Bregman divergences are usually not homogeneous ($D(kp,kq)=k^\lambda D(p,q)$) except for

- Burg entropy ($\lambda=0$) $F(x)=-\log x$, $J_F(p,q)=\log\frac{p+q}{2\sqrt{pq}}$. (logarithm of the ratio of the arithmetic mean over the geometric mean),
- Shannon entropy ($\lambda=1$) $F(x)=x\log x$, $J_F(p,q)=\frac{1}{2}\left(p\log\frac{2p}{p+q}+q\log\frac{2q}{p+q}\right)$
- Squared entropy ($\lambda=2$) $F(x)=x^2$, $J_F(p,q)=\frac{1}{4}(p-q)^2$

For homogeneous distances,

$$D(p,q)=q^\lambda D\left(\frac{p}{q},1\right).$$

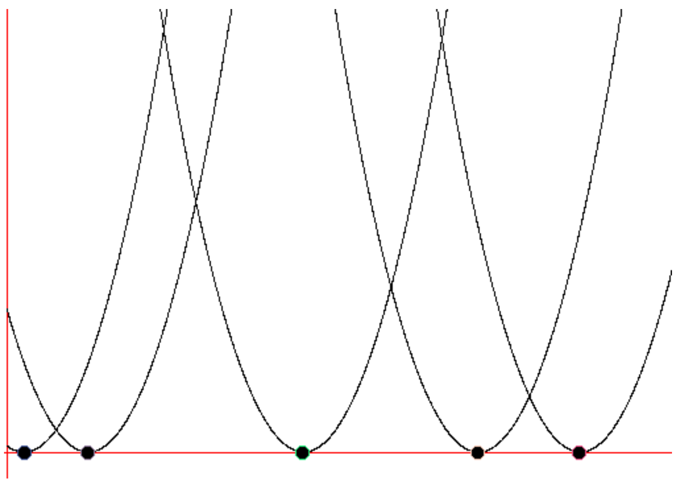
# Voronoi diagram as a minimization diagram

Sites (generators), cells.

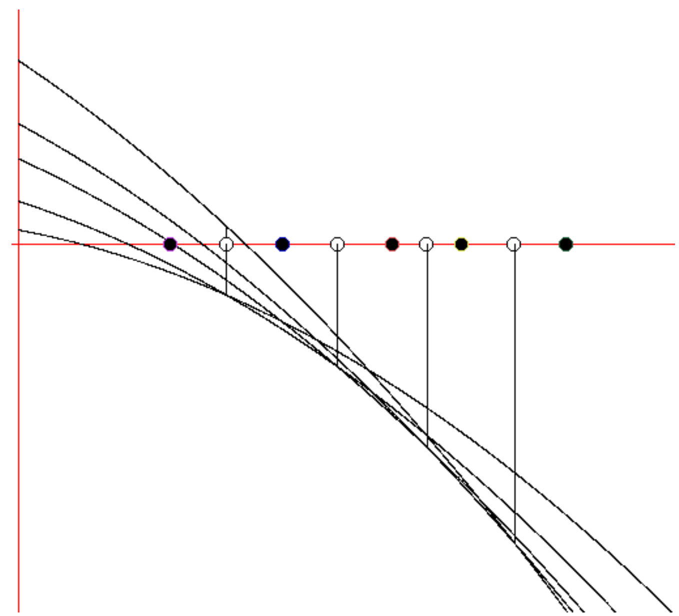
$$V(p_i) = \{p \mid J_F(p, p_i) < J_F(p, p_j)\ \forall j \neq i\}.$$

*Anchored* distance function

$$D_i(x) = J_F(x, p_i) = \frac{F(p_i)+F(x)}{2} - F\left(\frac{p_i+x}{2}\right)$$
$$\equiv D_i'(x) = \frac{1}{2}F(p_i) - F\left(\frac{p_i+x}{2}\right).$$

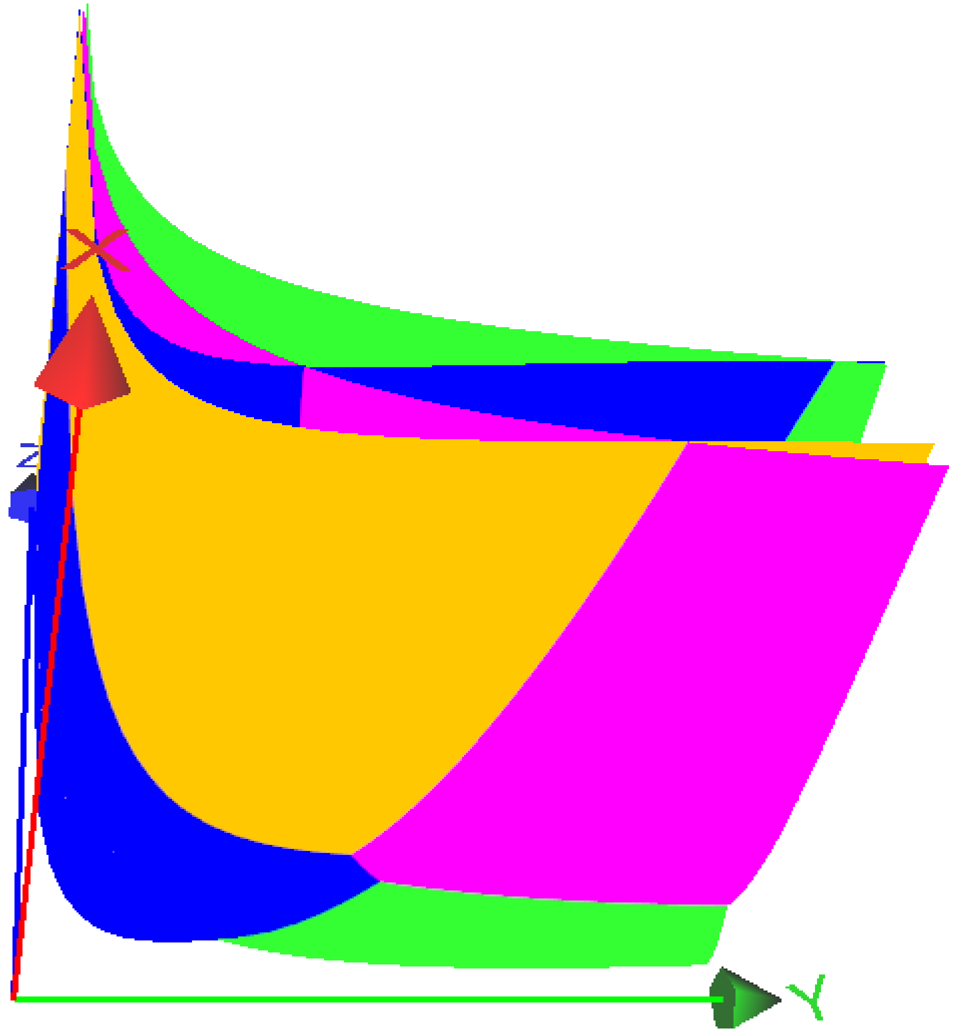


lower enveloppe Shannon $D_i(x)$



lower enveloppe Shannon $D_i'(x)$

# 3D Jensen-Burg envelope




$\rightarrow$ Vertical projection yields 2D Jensen-Burg Voronoi diagram.

# Jensen-Bregman Voronoi

Jensen-Bregman divergences are not necessarily convex.

For example, consider $F(x) = x^3$ on $\mathbb{R}^+$ (with $F''(x) = 6x$). We have $D_p''(x) = 3(x - \frac{p+x}{4})$; This is non-negative for $x \geq \frac{p}{3}$ only.

However, we have a special structure of the lower envelope.
Minimization diagram $\min_i D_i(x)$ is equivalent to the minimization diagram of the functions

$$D_i(x) \equiv D_i'(x) = \frac{1}{2}F(p_i) - F\left(\frac{p_i + x}{2}\right).$$

($D'$ does not mean derivative function but simply $D_i(x) = D_i'(x) + \frac{F(x)}{2}$ )

# Jensen-Bregman bisector

Bisector $(p, q)$ (after removing $\frac{F(x)}{2}$ terms):

$$\frac{F(p)}{2} - F\left(\frac{x+p}{2}\right) = \frac{F(q)}{2} - F\left(\frac{x+q}{2}\right).$$

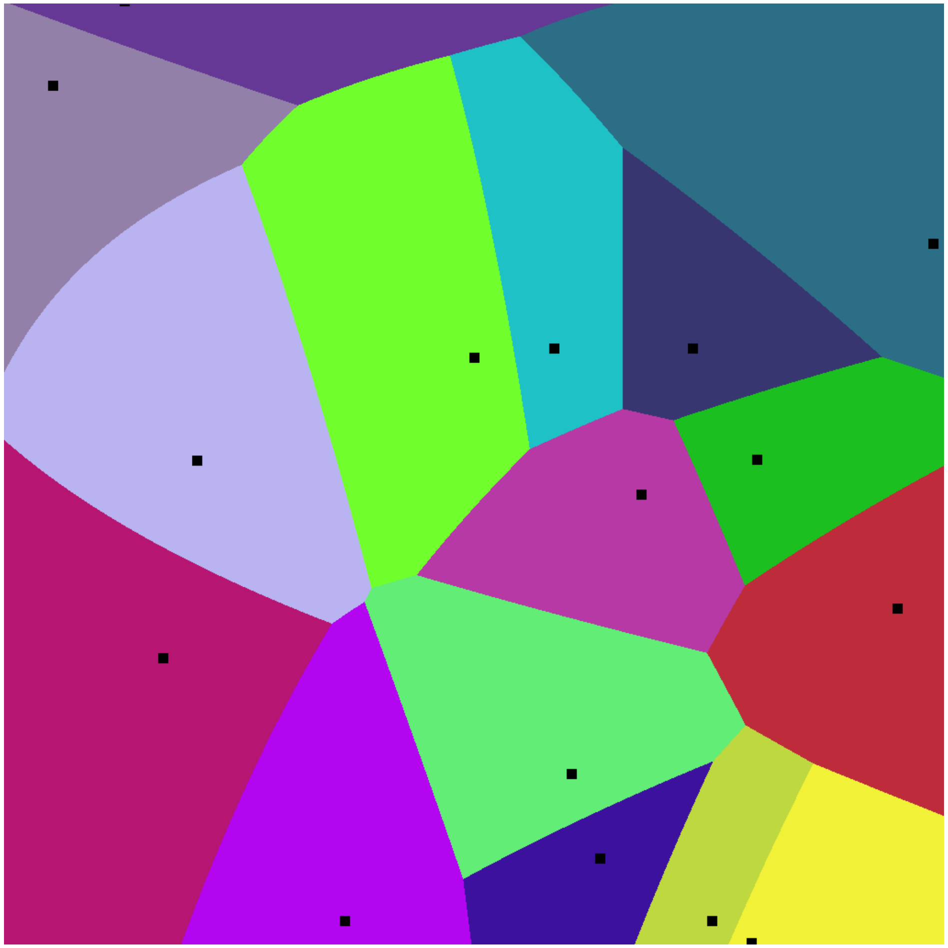
$$\underbrace{F\left(\frac{x+q}{2}\right)}_{\text{convex}} \underbrace{-F\left(\frac{x+p}{2}\right)}_{\text{Concave}} + \underbrace{\frac{F(p)}{2} - \frac{F(q)}{2}}_{\text{Constante}} = 0$$

Interpreted as the sum/intersection of a *convex* function $F\left(\frac{x+q}{2}\right) - \frac{F(p)}{2}$ with a *concave* function $-F\left(\frac{x+p}{2}\right) - \frac{F(q)}{2}$.

In 2D, the iso-distance curves intersect in at most two points.
(proof by contradiction, require strict convexity).

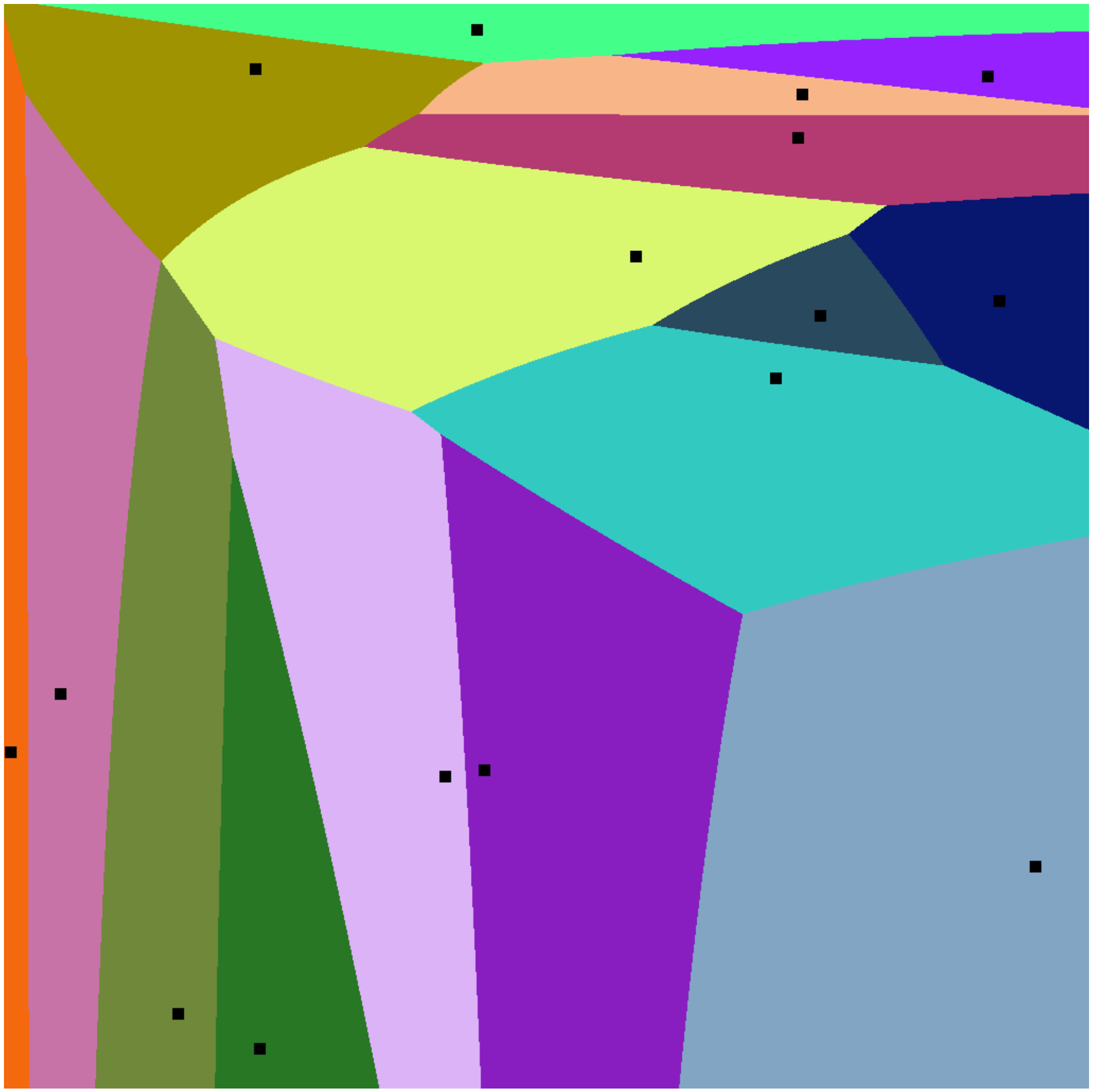
# Jensen-Bregman Voronoi

In 2D, bisectors are pseudo-lines (pseudo-circles iso-distance contours)
Complexity of the Jensen-Bregman Voronoi diagram is linear
(planar graph)



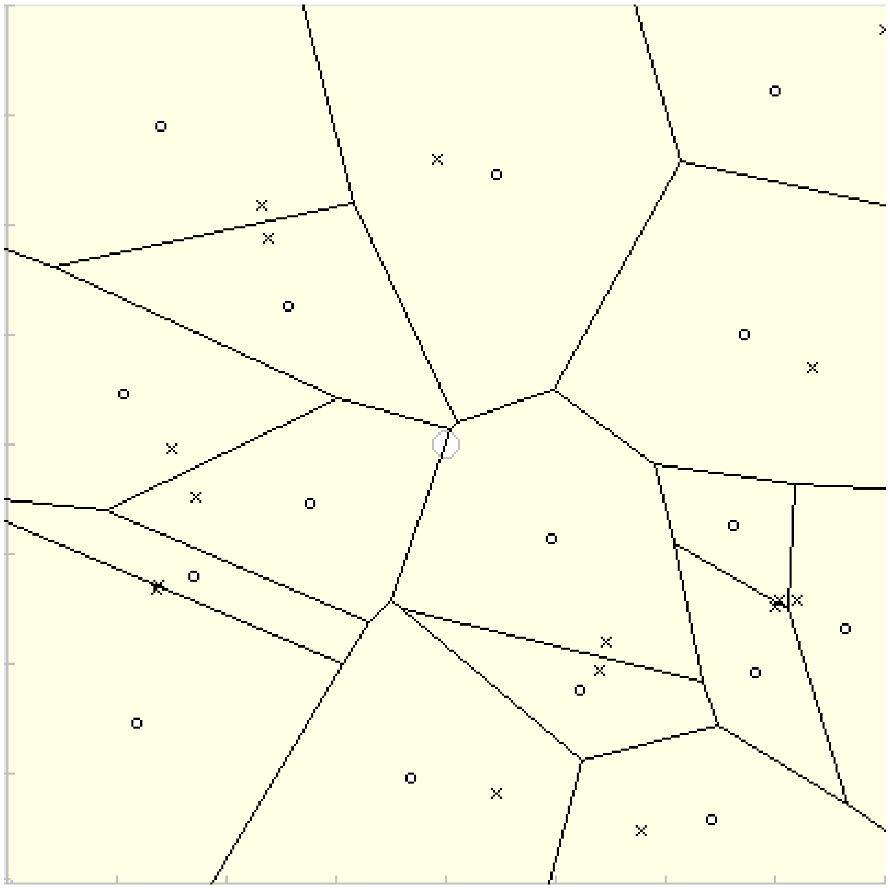
Jensen-Shannon Voronoi diagram

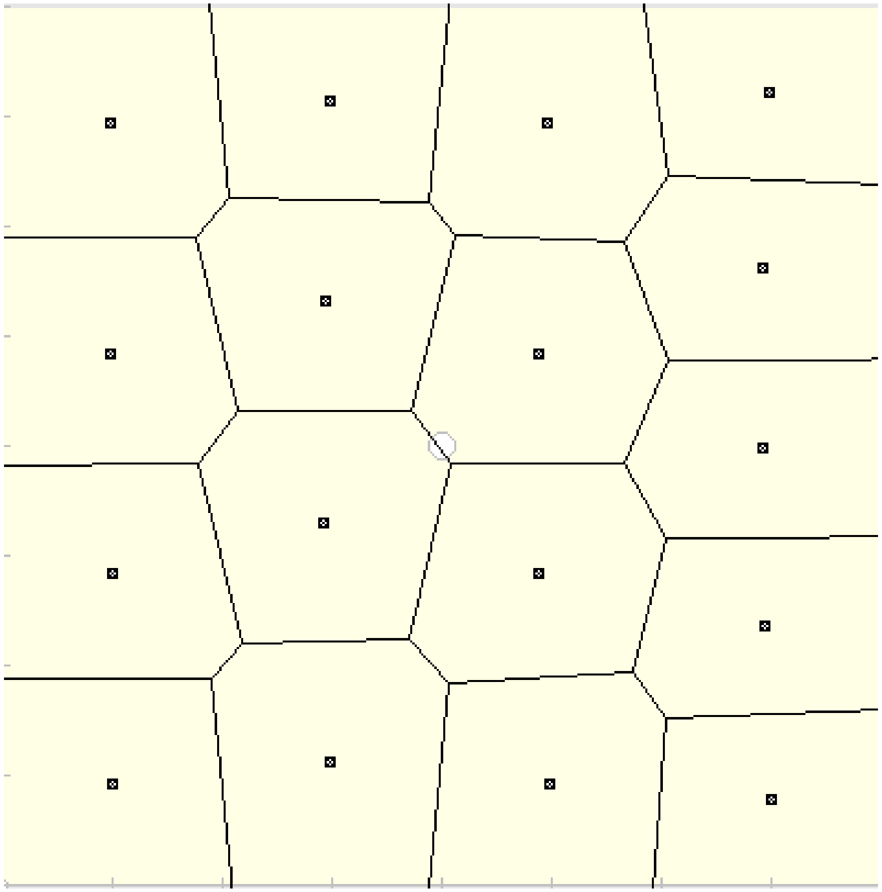
# 2D Jensen-Burg Voronoi



# Centroidal Voronoi Tesselations

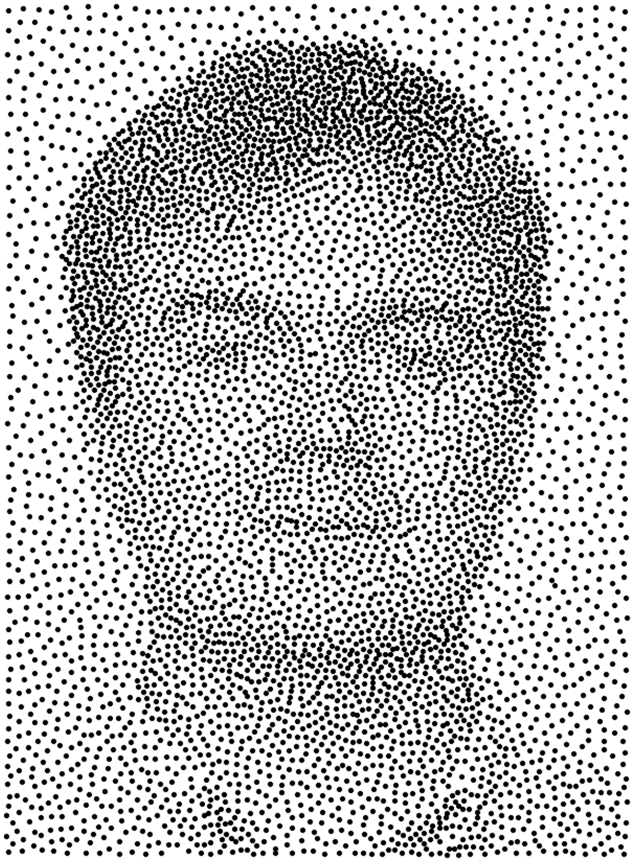
CVT using Lloyd or L-BFGS algorithms (Limited BroydenŰFletcherŰGoldfarbŰShanno).





# Applications of CVTs: Stippling

Also called pointillism in computer graphics.
(non-photorealistic renderings)



# Jensen-Bregman centroids

Consider a finite point set $\{p_1, ..., p_n\}$.

$$c^* = \arg\min_c \sum_{i=1}^n w_i J_F(p_i, c) = \arg\min_c L(c)$$

Minimization can be decomposed as

$$L(c) = L_{\mathrm{convex}}(c) + L_{\mathrm{concave}}(c).$$

(Under mild assumptions any function can be decomposed as a sum of a convex plus concave function)
For the Jensen-Bregman centroid, this decomposition is given *explicitly*:

$$\begin{aligned} L_{\mathrm{convex}}(c) &= \frac{F(c)}{2} \\ L_{\mathrm{concave}}(c) &= -\sum_{i=1}^n F\left(\frac{p_i + c}{2}\right), \end{aligned}$$

# Jensen-Bregman centroids (and barycenters)

Use framework of CCCP (ConCave Convex Procedure):

$$\nabla F(x) = \sum_{i=1}^{n} w_i \nabla F\left(\frac{x+p_i}{2}\right)$$

$$x = \nabla F^{-1}\left(\sum_{i=1}^{n} w_i \nabla F\left(\frac{x+p_i}{2}\right)\right)$$

Start from arbitrary $c_0$ and iterate:

$$x_{t+1} = \nabla F^{-1}\left(\sum_{i=1}^{n} w_i \nabla F\left(\frac{x_t+p_i}{2}\right)\right).$$

Guaranteed convergence to a (local) minimum.
(gradient descent methods require to fix learning rates)

# Jensen-Bregman centroids

Consider a dense region $X$ with an underlying density function $\rho(\cdot)$.

$$\arg\min_{c \in X} \int_{p \in X} \rho(p) \mathrm{BR}_F(p, c) \mathrm{d}p$$

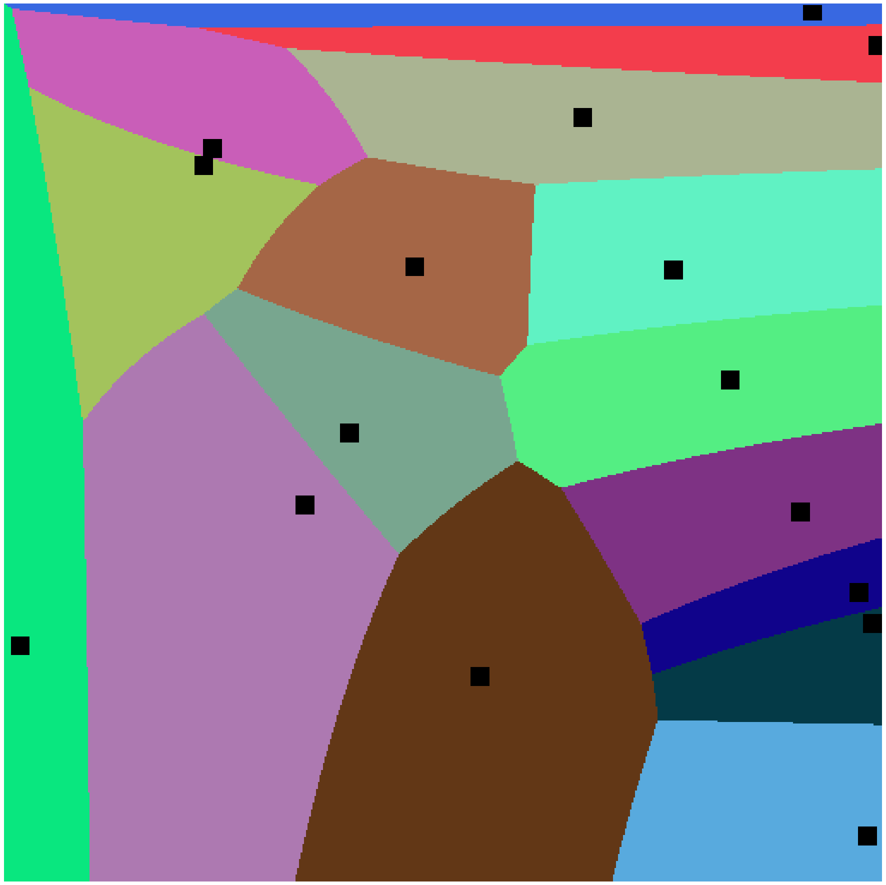
Integral CCCP using $c_0 = \int_{p \in X} \rho(p) \mathrm{d}p$, update as follows:

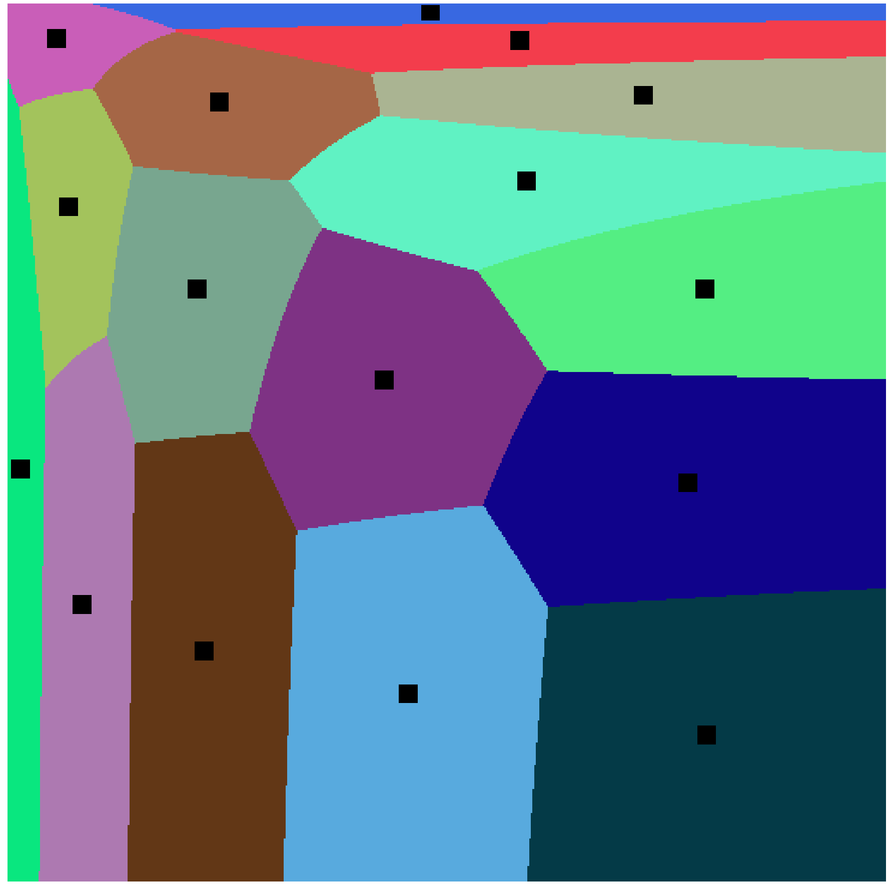
$$c_{t+1} = \nabla F^{-1} \left( \frac{\int_{p \in X} \rho(p) \nabla F \left( \frac{p + c_t}{2} \right) \mathrm{d}p}{\int_{p \in X} \rho(p) \mathrm{d}p} \right)$$

$\rightarrow$ In general, difficult to compute in closed-form (except for squared Mahalanobis). Approximate by discretizing.

# Jensen-Burg centroidal Voronoi tesselation

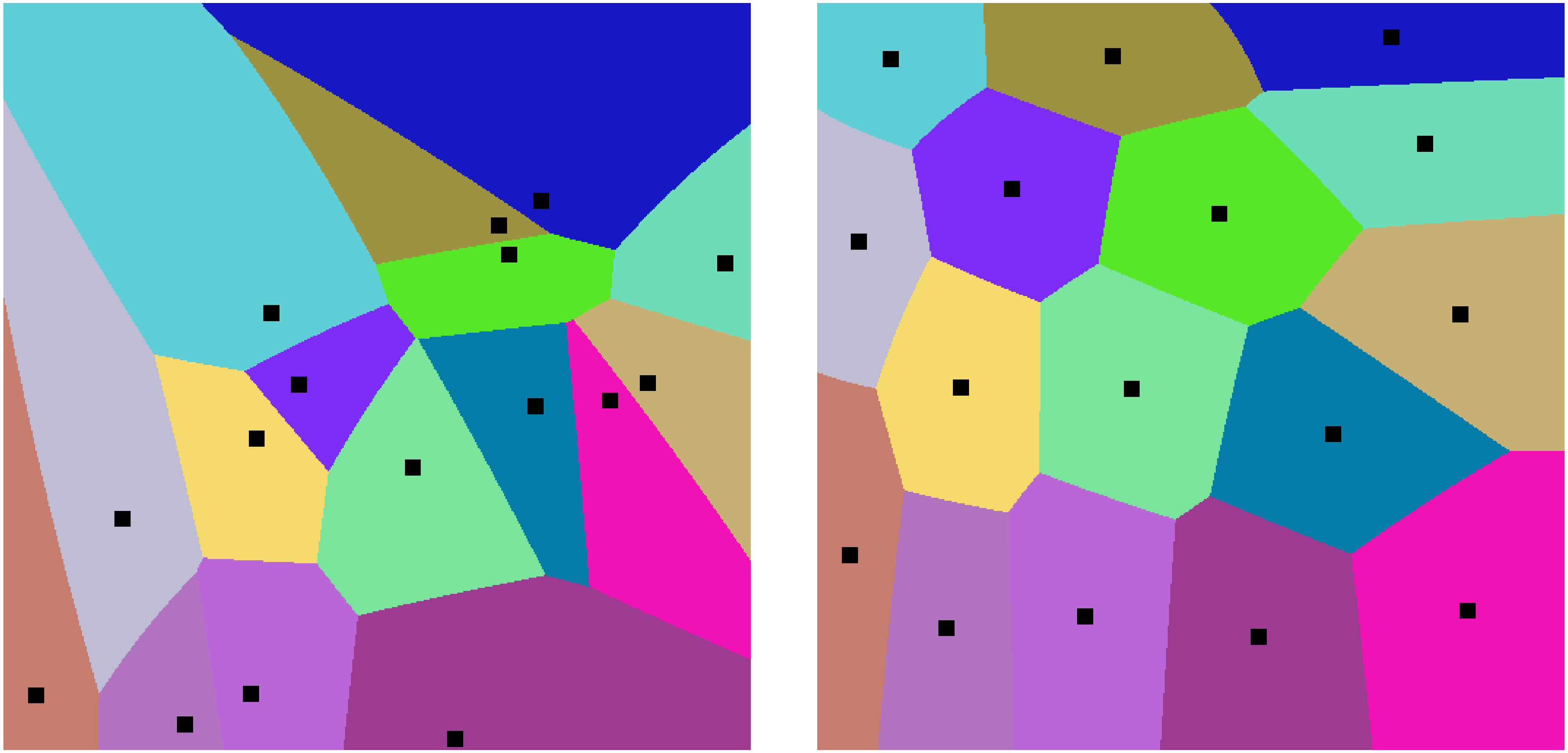
We discretize cells.





# Jensen-Shannon centroidal Voronoi tesselation

We discretize cells



# Application to statistical Voronoi diagrams

For probability densities, Bhattacharyya similarity coefficient (and non-metric symmetric distance):

$$C_(p,q) = \int \sqrt{p(x)q(x)}\mathrm{d}x, \qquad 0 < C(p,q) \leq 1, \qquad B(p,q) = -\ln C(p,q).$$

(coefficient is always *strictly* positive)
Hellinger metric

$$H(p,q) = \sqrt{\frac{1}{2}\int (\sqrt{p(x)} - \sqrt{q(x)})^2 \mathrm{d}x},$$

such that $0 \leq H(p,q) \leq 1$.

$$\begin{aligned} H(p,q) &= \sqrt{\frac{1}{2}\left(\int p(x)\mathrm{d}x + \int q(x)\mathrm{d}x - 2\int \sqrt{p(x)}\sqrt{q(x)}\mathrm{d}x\right)} \\ &= \sqrt{1 - C(p,q)}. \end{aligned}$$

# Chernoff coefficients/$\alpha$-divergences

Skew Bhattacharrya divergences based on Chernoff $\alpha$-coefficients.

$$\begin{aligned} B_\alpha(p,q) &= -\ln \int_x p^\alpha(x) q^{1-\alpha}(x)\mathrm{d}x = -\ln C_\alpha(p,q) \\ &= -\ln \int_x q(x) \left(\frac{p(x)}{q(x)}\right)^\alpha \mathrm{d}x \\ &= -\ln E_q[L^\alpha(x)] \end{aligned}$$

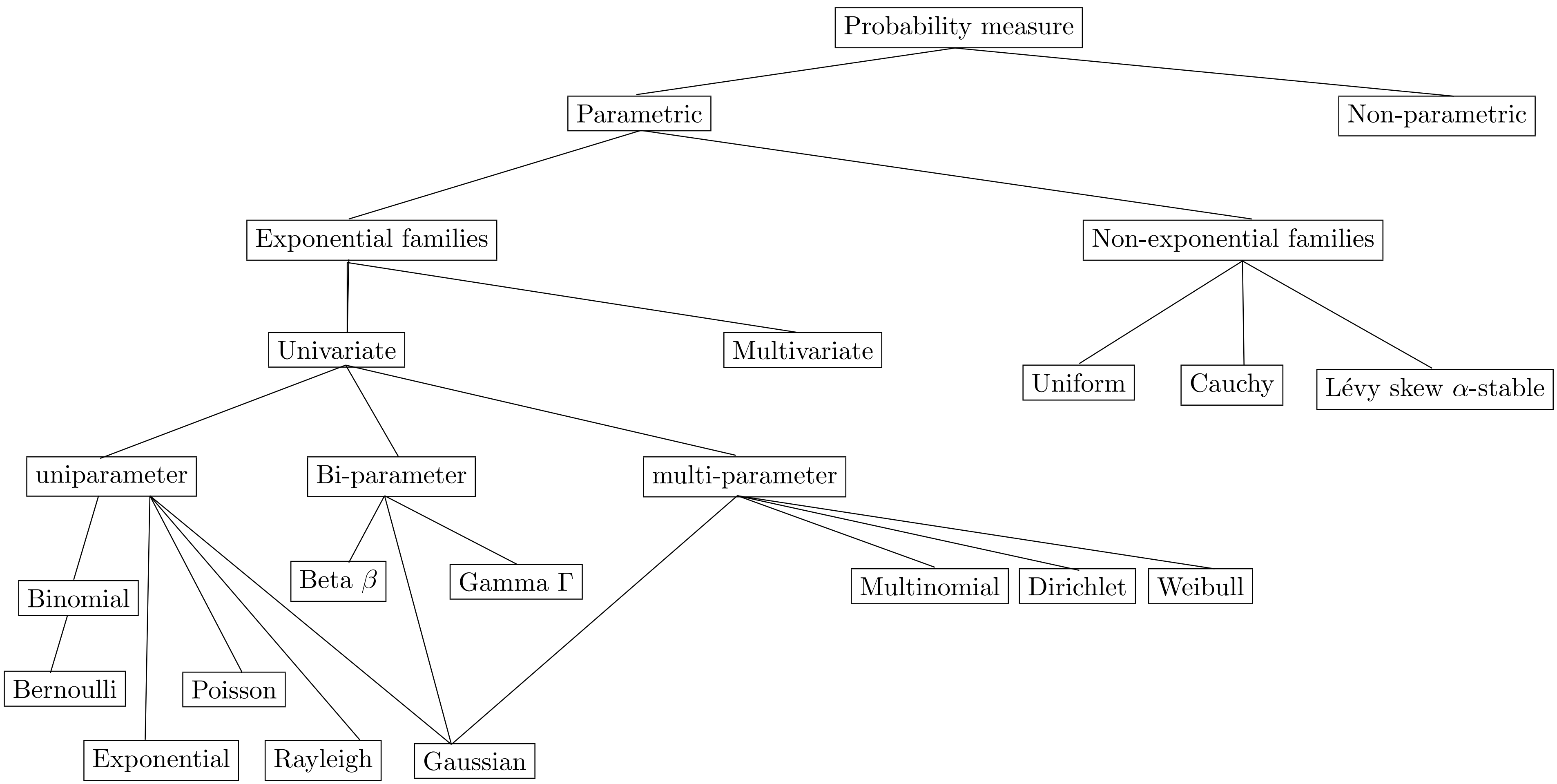
Amari $\alpha$-divergence:

$$D_\alpha(p||q) = \begin{cases} \frac{4}{1-\alpha^2}\left(1 - \int p(x)^{\frac{1-\alpha}{2}} q(x)^{\frac{1+\alpha}{2}} \mathrm{d}x\right), & \alpha \neq \pm 1, \\ \int p(x) \log \frac{p(x)}{q(x)} \mathrm{d}x = \mathrm{KL}(p,q), & \alpha = -1, \\ \int q(x) \log \frac{q(x)}{p(x)} \mathrm{d}x = \mathrm{KL}(q,p), & \alpha = 1, \end{cases}$$

$$D_\alpha(p||q) = D_{-\alpha}(q||p)$$

Remapping $\alpha' = \frac{1-\alpha}{2}$ ($\alpha = 1 - 2\alpha'$) to get Chernoff $\alpha'$-divergences

# Exponential families in statistics




Finite moments of all orders.

# Exponential families in statistics

Gaussian, Poisson, Bernoulli/multinomial, Gamma/Beta, etc.:

$$p(x;\lambda) = p_F(x;\theta) = \exp\left(\langle t(x),\theta\rangle - F(\theta) + k(x)\right).$$

Example: Poisson distribution

$$p(x;\lambda) = \frac{\lambda^x}{x!}\exp(-\lambda),$$

- the sufficient statistic $t(x) = x$,
- $\theta = \log\lambda$, the natural parameter,
- $F(\theta) = \exp\theta$, the log-normalizer,
- and $k(x) = -\log x!$ the carrier measure (with respect to the counting measure).

# Gaussians as an exponential family

$$p(x;\lambda) = p(x;\mu,\Sigma) = \frac{1}{2\pi\sqrt{\det\Sigma}} \exp\left(-\frac{(x-\mu)^T\Sigma^{-1}(x-\mu))}{2}\right)$$

- $\theta = (\Sigma^{-1}\mu, \frac{1}{2}\Sigma^{-1}) \in \Theta = \mathbb{R}^d \times \mathbb{K}_{d\times d}$, with $\mathbb{K}_{d\times d}$ cone of positive definite matrices,
- $F(\theta) = \frac{1}{4}\mathrm{tr}(\theta_2^{-1}\theta_1\theta_1^T) - \frac{1}{2}\log\det\theta_2 + \frac{d}{2}\log\pi$,
- $t(x) = (x, -x^Tx)$,
- $k(x) = 0$.

Inner product : composite, sum of a dot product and a matrix trace :

$$\langle\theta,\theta'\rangle = \theta_1^T\theta_1' + \mathrm{tr}(\theta_2^T\theta_2').$$

The coordinate transformation $\tau : \Lambda \to \Theta$ is given for $\lambda = (\mu, \Sigma)$ by

$$\tau(\lambda) = \left(\lambda_2^{-1}\lambda_1, \frac{1}{2}\lambda_2^{-1}\right), \qquad \tau^{-1}(\theta) = \left(\frac{1}{2}\theta_2^{-1}\theta_1, \frac{1}{2}\theta_2^{-1}\right)$$

# Bhattacharyya/Chernoff of exponential families

Equivalence with skew Burbea-Rao distances:

$$B_\alpha(p_F(x;\theta_p), p_F(x;\theta_q)) = \mathrm{BR}_F^{(\alpha)}(\theta_p,\theta_q) = \alpha F(\theta_p)+(1-\alpha)F(\theta_q)-F(\alpha\theta_p+(1-\alpha)\theta_q)$$

Proof: Chernoff coefficients $C_\alpha(p,q)$ of members $p = p_F(x;\theta_p)$ and $q = p_F(x;\theta_q)$ of the *same* exponential family $\mathcal{E}_F$:

$C_\alpha(p,q) = \int p^\alpha(x)q^{1-\alpha}(x)\mathrm{d}x = \int p_F^\alpha(x;\theta_p)p_F^{1-\alpha}(x;\theta_q)\mathrm{d}x$
$= \int \exp(\alpha(\langle x,\theta_p\rangle - F(\theta_p))) \times \exp((1-\alpha)(\langle x,\theta_q\rangle - F(\theta_q)))\mathrm{d}x$
$= \int \exp\left(\langle x, \alpha\theta_p + (1-\alpha)\theta_q\rangle - (\alpha F(\theta_p) + (1-\alpha)F(\theta_q)\right)\mathrm{d}x$
$= \exp -(\alpha F(\theta_p) + (1-\alpha)F(\theta_q)) \times$
$\int \exp\left(\langle x, \alpha\theta_p + (1-\alpha)\theta_q\rangle - F(\alpha\theta_p + (1-\alpha)\theta_q) + F(\alpha\theta_p + (1-\alpha)\theta_q)\right)\mathrm{d}x$
$= \exp\left(F(\alpha\theta_p + (1-\alpha)\theta_q) - (\alpha F(\theta_p) + (1-\alpha)F(\theta_q)\right) \times \int \exp\langle x, \alpha\theta_p + (1-\alpha)\theta_q\rangle -$
$F(\alpha\theta_p + (1-\alpha)\theta_q)\mathrm{d}x$

$$= \exp\left(F(\alpha\theta_p + (1-\alpha)\theta_q) - (\alpha F(\theta_p) + (1-\alpha)F(\theta_q)\right) \times \underbrace{\int p_F(x;\alpha\theta_p + (1-\alpha)\theta_q)\mathrm{d}x}_{=1}$$

$= \exp(-\mathrm{BR}_F^{(\alpha)}(\theta_p,\theta_q)) > 0.$ Coefficient is always strictly positive. For $\theta_p = \theta_q$, $C_\alpha(\theta_p,\theta_q) = \exp -0 = 1$ and $B_\alpha(\theta_p,\theta_q) = 0$.

# Summary of paper & results

- Extending Jensen-Shannon divergence to arbitrary convex (information) function (=negative entropy)
- Jensen-Bregman divergences (=Jensen or Burbea-Rao divergences)
- 2D Voronoi diagram is in *linear complexity* (bisector). Jensen-Bregman Voronoi diagrams extend Bregman Voronoi diagrams.
- Jensen-Bregman centroids (solved iteratively using CCCP)
- Jensen-Bregman centroidal Voronoi tesselations (CVTs, by discretization)
- Bhattacharyya distance of members of the same exponential family = Jensen-Bregman divergence on the natural parameters
- Statistical Voronoi diagrams, extending Bregman Voronoi diagrams (since asymptotically skewed Jensen-Bregman divergences yields Bregman divergences)

# References


- "Bhattacharyya clustering with applications to mixture simplifications," ICPR 2010. arXiv, 2010. `http://arxiv.org/abs/1004.5049`
- "Sided and symmetrized Bregman centroids," IEEE Transactions on Information Theory, vol. 55, no. 6, pp. 2048-2059, June 2009.
- "Bregman Voronoi diagrams," Discrete & Computational Geometry, 2010.
- "On the convexity of some divergence measures based on entropy functions," IEEE Transactions on Information Theory, vol. 28, no. 3, pp. 489-495, 1982.
- "Statistical exponential families: A digest with flash cards," 2009, arXiv.org:0911.4863
- A. Yuille and A. Rangarajan, "The concave-convex procedure," Neural Computation, vol. 15, no. 4, pp. 915-936, 2003.
- J. Zhang, "Divergence Function, Duality, and Convex Analysis," Neural Computation, vol. 6, 159-195, 2004.

# Thank you

www.informationgeometry.org
www.informationgeometry.org/JensenBregman/

blog.informationgeometry.org

http://www.twitter.com/FrnkNlsn

Acknowledgements:
We gratefully acknowledge financial support from DIGITEO (grant GAS 2008-16D) and French National Research Agency (ANR, grant GAIA 07-BLAN-0328-01).