On Bregman Voronoi Diagrams

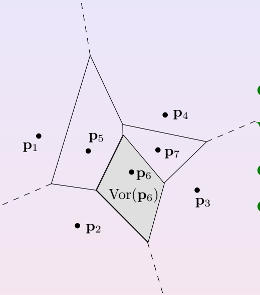
Frank Nielsen[1] Jean-Daniel Boissonnat[2] Richard Nock[3]

[1]Sony Computer Science Laboratories, Inc.
Fundamental Research Laboratory
Frank.Nielsen@acm.org

[2]INRIA Sophia-Antipolis
Geometrica
Jean-Daniel.Boissonnat@sophia.inria.fr

[3]University of Antilles-Guyanne
CEREGMIA
Richard.Nock@martinique.univ-ag.fr

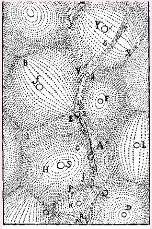
July 2006 — January 2007



- Voronoi diagram $\mathrm{Vor}(\mathcal{S})$ s.t.

$\mathrm{Vor}(\mathbf{p}_i) \stackrel{\mathrm{def}}{=} \{\mathbf{x} \in \mathbb{R}^d \mid ||\mathbf{p}_i\mathbf{x}|| \leq ||\mathbf{p}_j\mathbf{x}|| \ \forall \mathbf{p}_j \in \mathcal{S}\}$

- Voronoi sites (static view).
- Voronoi generators (dynamic view).

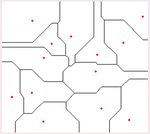


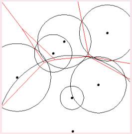
$\rightarrow$ René Descartes, 17th century.
$\rightarrow$ Partition the Euclidean space $\mathbb{E}^d$ wrt
the Euclidean distance $||\mathbf{x}||_2 = \sqrt{\sum_{i=1}^d x_i^2}$.

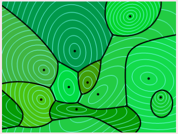
Generalizing Voronoi Diagrams

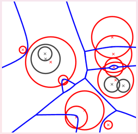
Voronoi diagrams *widely studied* in comp. geometry [AK'00]:

- Manhattan (taxi-cab) diagram (L_1 norm): $||\mathbf{x}||_1 = \sum_{i=1}^{d} |x_i|$,
- Affine diagram (power distance): $||\mathbf{x} - \mathbf{c}_i||^2 - r_i^2$,
- Anisotropic diagram (quad. dist.): $\sqrt{(\mathbf{x} - \mathbf{c}_i)^T \mathbf{Q}_i (\mathbf{x} - \mathbf{c}_i)}$,
- Apollonius diagram (circle distance): $||\mathbf{x} - \mathbf{c}_i|| - r_i$,
- Möbius diagram (weighted distance): $\lambda_i ||\mathbf{x} - \mathbf{c}_i|| - \mu_i$,
- Abstract Voronoi diagrams [Klein'89], etc.


Taxi-cab diagram

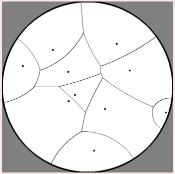

Power diagram


Anisotropic diagram

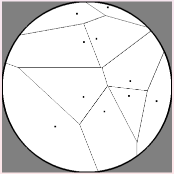

Apollonius diagram

Non-Euclidean Voronoi diagrams

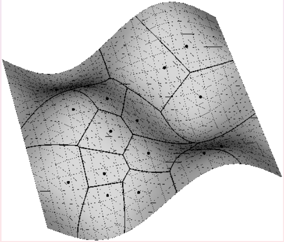
- Hyperbolic Voronoi: Poincaré disk [B+'96], Poincaré half-plane [OT'96], etc.
- Kullback-Leibler divergence (statistical Voronoi diagrams) [OI'96] & [S+'98]
  Divergence between two statistical distributions
  $\mathrm{KL}(p||q) = \int_x p(x) \log \frac{p(x)}{q(x)} \mathrm{d}x$ [relative entropy]
- Riemannian Voronoi diagrams: geodesic length (aka geodesic Voronoi diagrams) [LL'00]



Hyperbolic Voronoi (Poincaré)



Hyperbolic Voronoi (Klein)



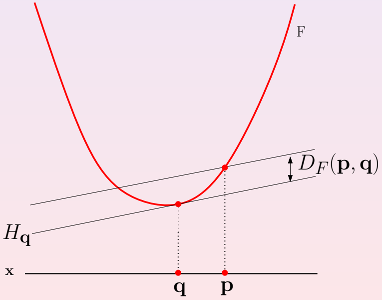
Riemannian Voronoi

F a strictly convex and differentiable function defined over a convex set domain $\mathcal{X}$

$D_F(\mathbf{p}, \mathbf{q}) = F(\mathbf{p}) - F(\mathbf{q}) - \langle \mathbf{p} - \mathbf{q}, \nabla_F(\mathbf{q}) \rangle$

not a distance

(not necessarily symmetric nor does triangle inequality hold)



Example: The squared Euclidean distance

- $F(x) = x^2$: strictly convex and differentiable over $\mathbb{R}^d$

(Multivariate $F(\mathbf{x}) = \sum_{i=1}^d x_i^2$)

$$\begin{aligned} D_F(\mathbf{p},\mathbf{q}) &= F(\mathbf{p}) - F(\mathbf{q}) - \langle \mathbf{p}-\mathbf{q}, \nabla_F(\mathbf{q})\rangle \\ &= \mathbf{p}^2 - \mathbf{q}^2 - \langle \mathbf{p}-\mathbf{q}, 2\mathbf{q}\rangle \\ &= \|\mathbf{p}-\mathbf{q}\|^2 \end{aligned}$$

Voronoi diagram equivalence classes

Since $\mathrm{Vor}(\mathcal{S}; d_2) = \mathrm{Vor}(\mathcal{S}; d_2^2)$, the ordinary Voronoi diagram is interpreted as a Bregman Voronoi diagram.

(Any strictly monotone function f of d_2 yields the same ordinary Voronoi diagram: $\mathrm{Vor}(\mathcal{S}; d_2) = \mathrm{Vor}(\mathcal{S}; f(d_2))$.)

Bregman divergences for probability distributions

- $F(\mathbf{p}) = \int p(x) \log p(x) \, \mathrm{d}x$ (Shannon entropy)

(Discrete distributions $F(\mathbf{p}) = \sum_x p(x) \log p(x) \, \mathrm{d}x$)

$$\begin{aligned} D_F(\mathbf{p}, \mathbf{q}) &= \int (p(x) \log p(x) - q(x) \log q(x) \\ &\qquad - \langle p(x) - q(x), \log q(x) + 1 \rangle)) \, \mathrm{d}x \\ &= \int p(x) \log \frac{p(x)}{q(x)} \, \mathrm{d}x \qquad \text{(KL divergence)} \end{aligned}$$

Kullback-Leiber divergence also known as: relative entropy or I-divergence.

(Defined either on the probability simplex or *extended* on the full positive quadrant.)

Bregman divergences: A versatile measure

Bregman divergences are *versatile*, suited to mixed type data.

(Build multivariate divergences *dimensionwise* using elementary univariate divergences.)

Fact (Linearity)

Bregman divergence is a linear operator:
$\forall F_1 \in \mathcal{C}\ \forall F_2 \in \mathcal{C} \quad D_{F_1+\lambda F_2}(\mathbf{p}||\mathbf{q}) = D_{F_1}(\mathbf{p}||\mathbf{q}) + \lambda D_{F_2}(\mathbf{p}||\mathbf{q})$ *for any* $\lambda \geq 0$.

Fact (Equivalence classes)

Let $G(\mathbf{x}) = F(\mathbf{x}) + \langle \mathbf{a}, \mathbf{x} \rangle + b$ *be another strictly convex and differentiable function, with* $\mathbf{a} \in \mathbb{R}^d$ *and* $b \in \mathbb{R}$. *Then* $D_F(\mathbf{p}||\mathbf{q}) = D_G(\mathbf{p}||\mathbf{q})$.

(For Voronoi diagrams, relax the classes to any monotone function of D_F*: relative vs absolute divergence.)*

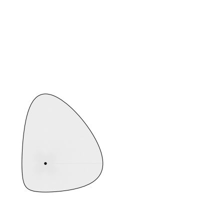
Bregman divergences for sound processing

$F(\mathbf{p}) = -\int_x \log p(x)\,\mathrm{d}x$ (Burg entropy)

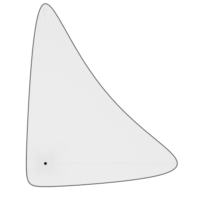
$D_F(\mathbf{p},\mathbf{q}) = \int_x \left(\frac{p(x)}{q(x)} - \log\frac{p(x)}{q(x)} - 1\right)\mathrm{d}x$ (Itakura-Saito)

Convexity & Bregman balls

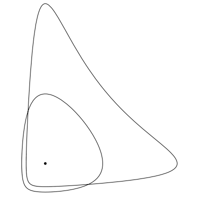
$D_F(\mathbf{p}||\mathbf{q})$ is *convex* in its first argument $\mathbf{p}$ but *not necessarily* in its second argument $\mathbf{q}$.



$\mathrm{ball}(\mathbf{c}, r) = \{\mathbf{x} \mid D_F(\mathbf{x},\mathbf{c}) \leq r\}$



$\mathrm{ball}'(\mathbf{c}, r) = \{\mathbf{x} \mid D_F(\mathbf{c},\mathbf{x}) \leq r\}$



Superposition of I.-S. balls

Dual divergence

Convex conjugate

Unique *convex conjugate* function G $(= F^*)$ obtained by the Legendre transformation: $G(\mathbf{y}) = \sup_{\mathbf{x}\in\mathcal{X}}\{\langle \mathbf{y}, \mathbf{x}\rangle - F(\mathbf{x})\}$.

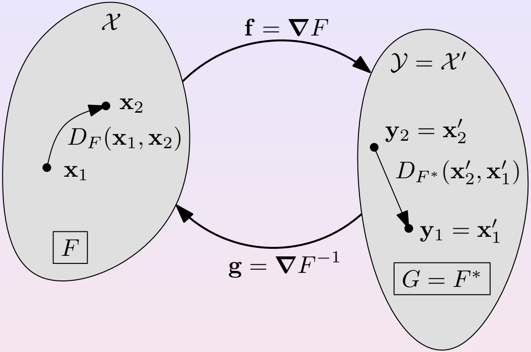
$$\nabla G(\mathbf{y}) = \nabla(\langle \mathbf{y}, \mathbf{x}\rangle - F(\mathbf{x})) = 0 \quad \rightarrow \quad \boxed{\mathbf{y} = \nabla_F(\mathbf{x})}.$$

(thus we have $\mathbf{x} = \nabla_F{}^{-1}(\mathbf{y})$)
$D_F(\mathbf{p}||\mathbf{q}) = F(\mathbf{p}) - F(\mathbf{q}) - \langle \mathbf{p}-\mathbf{q}, \mathbf{q}'\rangle$ with $(\mathbf{q}' = \nabla_F(\mathbf{q}))$. F^* $(= G)$ is a Bregman generator function such that $(F^*)^* = F$.

Dual Bregman divergence

$D_F(\mathbf{p}||\mathbf{q}) = F(\mathbf{p}) + F^*(\mathbf{q}') - \langle \mathbf{p}, \mathbf{q}'\rangle = D_{F^*}(\mathbf{q}'||\mathbf{p}')$

Legendre transformation: $F^*(\mathbf{x}') = -F(\mathbf{x}) + \langle \mathbf{x}, \mathbf{x}' \rangle$.



$$
\begin{aligned}
D_F(\mathbf{x}_1, \mathbf{x}_2) &= F(\mathbf{x}_1) - F(\mathbf{x}_2) - \langle \mathbf{x}_1 - \mathbf{x}_2, \mathbf{x}'_2 \rangle \\
&= -F^*(\mathbf{x}'_1) + \langle \mathbf{x}_1, \mathbf{x}'_1 \rangle + F^*(\mathbf{x}'_2) - \langle \mathbf{x}_1, \mathbf{x}'_2 \rangle \\
&= D_{F^*}(\mathbf{x}'_2, \mathbf{x}'_1)
\end{aligned}
$$

Examples of dual divergences

- **Exponential loss $\longleftrightarrow$ unnormalized Shannon entropy.**
  $F(x) = \exp(x) \longleftrightarrow G(y) = y \log y - y = F^*(x')$.

$F(x) = \exp x$	$D_F(x_1 \| \| x_2) = \exp x_1 - \exp x_2 - (x_1 - x_2) \exp x_2$	$f(x) = \exp x = y$
$G(y) = y \log y - y$	$D_G(y_1 \| \| y_2) = y_1 \log \frac{y_1}{y_2} + y_2 - y_1$	$g(y) = \log y = x$

- **Logistic loss $\longleftrightarrow$ Bernouilli-like entropy.**
  $F(x) = x \log x + (1-x) \log(1-x) \longleftrightarrow G(y) = \log(1 + \exp(y))$

$F(x) = \log(1 + \exp x)$	$D_F(x_1 \| \| x_2) = \log \frac{1+\exp x_1}{1+\exp x_2} - (x_1 - x_2) \frac{\exp x_2}{1+\exp x_2}$	$f(x) = \frac{\exp x}{1+\exp x} = y$
$G(y) = y \log \frac{y}{1-y} + \log(1-y)$	$D_G(y_1 \| \| y_2) = y_1 \log \frac{y_1}{y_2} + (1 - y_1) \log \frac{1-y_1}{1-y_2}$	$g(y) = \log \frac{y}{1-y} = x$

Bregman (Dual) divergences

Dual divergences have gradient entries swapped in the table:

(Because of equivalence classes, it is sufficient to have $f = \Theta(g)$.)

Dom. $\mathcal{X}$	Function F (or dual $G = F^*$)	Gradient $(f = g^{-1})$	Inv. grad. $(g = f^{-1})$	Divergence $D_F(p, q)$
$\mathbb{R}$	Squared function$\star$ x^2	$2x$	$\frac{x}{2}$	Squared loss (norm) $(p-q)^2$
$\mathbb{R}^+$	Unnorm. Shannon entropy $x \log x - x$	$\log x$	$\exp(x)$	Kullback-Leibler div. (I-div.) $p \log \frac{p}{q} - p + q$
$\mathbb{R}$	Exponential $\exp x$	$\exp x$	$\log x$	Exponential loss $\exp(p) - (p - q + 1)\exp(q)$
$\mathbb{R}^{+}*$	Burg entropy$\star$ $-\log x$	$-\frac{1}{x}$	$-\frac{1}{x}$	Itakura-Saito divergence $\frac{p}{q} - \log \frac{p}{q} - 1$
$[0, 1]$	Bit entropy $x \log x + (1-x)\log(1-x)$	$\log \frac{x}{1-x}$	$\frac{\exp x}{1+\exp x}$	Logistic loss $p \log \frac{p}{q} + (1-p)\log \frac{1-p}{1-q}$
$\mathbb{R}$	Dual bit entropy $\log(1 + \exp x)$	$\frac{\exp x}{1+\exp x}$	$\log \frac{x}{1-x}$	Dual logistic loss $\log \frac{1+\exp p}{1+\exp q} - (p-q)\frac{\exp q}{1+\exp q}$
$[-1, 1]$	Hellinger$\star$ $-\sqrt{1-x^2}$	$\frac{x}{\sqrt{1-x^2}}$	$\frac{x}{\sqrt{1+x^2}}$	Hellinger $\frac{1-pq}{\sqrt{1-q^2}} - \sqrt{1-p^2}$

(Self-dual divergences are marked with an asterisk $\star$. Note that $f = \nabla_F$ and $g = \nabla_F^{-1}$.)

Self-dual Bregman divergences: Legendre duals

Legendre duality: Consider functions *and* domains:

$$(F, \mathcal{X}) \leftrightarrow (F^*, \mathcal{X}^*)$$

- Squared Euclidean distance: $F(\mathbf{x}) = \frac{1}{2}\langle \mathbf{x}, \mathbf{x} \rangle$ is self-dual on $\mathcal{X} = \mathcal{X}^* = \mathbb{R}^d$.
- Itakura-Saito divergence. $F(x) = -\sum \log x_i$. Domains are $\mathcal{X} = \mathbb{R}^+*$ and $\mathcal{X}^* = \mathbb{R}^-*$ $(G = F^* = -\log(-x))$ $(D_F(p||q) = \frac{p}{q} - \log\frac{p}{q} - 1 = \frac{q'}{p'} - \log\frac{q'}{p'} - 1 = D_F(q'||p')$ with $q' = -\frac{1}{q}$ and $p' = -\frac{1}{p})$

It can be difficult to compute for a given F its convex conjugate: $\int \nabla_F^{-1}$

(eg, $F(x) = x \log x$; Liouville's non exp-log functions).

Exponential families in Statistics

Canonical representation of the proba. density func. of a r.v. X

$p(x|\theta) \stackrel{\text{def}}{=} \exp\{\langle\theta, \mathbf{f}(x)\rangle - F(\theta) - k(\mathbf{f}(x))\}$,

with $\mathbf{f}(x)$: *sufficient statistics* and θ: *natural parameters*.

F: *cumulant function* (or *log-partition* function).

Example: Univariate Gaussian distributions $\mathcal{N}(\mu, \sigma)$

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\{-\frac{(x-\mu)^2}{2\sigma^2}\} = \exp\left\{\langle [x\ x^2]^T, [\frac{\mu}{\sigma^2}\ \frac{-1}{2\sigma^2}]^T\rangle - (\frac{\mu^2}{\sigma^2} + \log\sigma) - \frac{1}{2}\log 2\pi\right\}$$

Minimal statistics $\mathbf{f}(x) = [x\ x^2]^T$, natural parameters $[\theta_1\ \theta_2]^T = [\frac{\mu}{\sigma^2}\ \frac{-1}{2\sigma^2}]^T$,

cumulant function $F(\theta_1, \theta_2) = -\frac{\theta_1^2}{4\theta_2} - \frac{1}{2}\log(-2\theta_2)$, and $k(\mathbf{f}(x)) = \frac{1}{2}\log 2\pi$.

Duality [B+'05]

Bregman functions $\leftrightarrow$ exponential families.

(Bijection primordial for designing tailored clustering divergences [B+05])

Exponential families include many usual distributions: Gaussian, Poisson, Bernouilli, Multinomial, Raleigh, etc.

Exponential family ($\exp\{\langle \boldsymbol{\theta}, \mathbf{f}(x)\rangle - F(\boldsymbol{\theta}) - k(\mathbf{f}(x))\}$)				
Name	Natural $\boldsymbol{\theta}$	Sufficient stat. $\mathbf{f}(x)$	Cumulant $F(\boldsymbol{\theta})$	Density cond. $k(\mathbf{f}(x))$
Bernouilli	$\log \frac{q}{1-q}$	x	$\log(1+\exp\theta)$	0
Gaussian	$[\frac{\mu}{\sigma^2}\ \frac{-1}{2\sigma^2}]^T$	$[x\ x^2]^T$	$-\frac{\theta_1^2}{4\theta_2} - \frac{1}{2}\log(-2\log\theta_2)$	$\frac{1}{2}\log 2\pi$
Poisson	$\log\lambda$	x	$\exp\theta$	$\log x!$

Cumulant function F fully characterizes the family.
W.l.o.g., simplify the p.d.f. to

$$g(\mathbf{w}|\boldsymbol{\theta}) = \exp(\langle \boldsymbol{\theta}, \mathbf{w}\rangle - F(\boldsymbol{\theta}) - h(\mathbf{w}))$$

(Indeed, $\frac{f(\mathbf{x}|\boldsymbol{\theta})}{g(\mathbf{w}|\boldsymbol{\theta})}$ is independent of $\boldsymbol{\theta}$.)

Multivariate Normal distributions: $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$

Probability density function of d-variate $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ ($\mathcal{X} = \mathbb{R}^d$) is

$$\frac{1}{(2\pi)^{\frac{d}{2}}\sqrt{\det \boldsymbol{\Sigma}}} \exp\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}(\mathbf{x}-\boldsymbol{\mu})\},$$

where $\boldsymbol{\mu}$ is the mean and $\boldsymbol{\Sigma}$ is the covariance matrix. (positive definite)

- Natural parameters:
  $\boldsymbol{\theta} = (\boldsymbol{\theta}_A, \boldsymbol{\theta}_B)$ with $\boldsymbol{\theta}_A = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ and $\boldsymbol{\theta}_B = -\frac{1}{2}\boldsymbol{\Sigma}^{-1}$.
- Sufficient statistics: $\mathbf{f}_A(\mathbf{x}) = \mathbf{x}$ and $\mathbf{f}_B(\mathbf{x}) = \mathbf{x}\mathbf{x}^T$
- Cumulant function F:

$$F(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{2}\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \frac{1}{2}\log\det(2\pi\boldsymbol{\Sigma}).$$

$$F(\boldsymbol{\theta}) = -\frac{1}{4}\boldsymbol{\theta}_A^T \boldsymbol{\theta}_B^{-1}\boldsymbol{\theta}_A + \frac{1}{2}\log\det(-\pi\boldsymbol{\theta}_B^{-1}).$$

- $k(\mathbf{f}(\mathbf{x})) = 0$.

Kullback-Leiber divergence $\mathrm{KL}(p||q)$ of two probability distributions p and q : $\mathrm{KL}(p||q) = \int_x p(x) \log \frac{p(x)}{q(x)} \mathrm{d}x$.

Kullback-Leibler: Bregman divergence for the cumulant function

$\mathrm{KL}(\theta_p||\theta_q) = D_F(\theta_q||\theta_p) = F(\theta_\mathbf{q}) - F(\theta_\mathbf{p}) - \langle(\theta_\mathbf{q} - \theta_\mathbf{p}), \theta_\mathbf{p}[\mathbf{f}]\rangle$,
with $\theta_p[\mathbf{f}] = E_{p(x|\theta)}[X] = \mathrm{d}\eta$ denoting *expectation parameters*.

$\theta_p[\mathbf{f}] = [\int_x \mathbf{f}(x) \exp\{\langle\theta_\mathbf{p}, \mathbf{f}(x)\rangle - F(\theta_\mathbf{p}) - k(\mathbf{f}(x))\}\mathrm{d}x]$

(Beware of argument swapping.)

Kullback-Leibler and Legendre duality

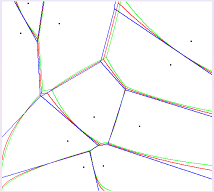
$\mathrm{KL}(\theta_p||\theta_q) = D_F(\theta_q||\theta_p) = D_{F^*}(\mathrm{d}\eta_p||\mathrm{d}\eta_q)$.

Univariate Normal distributions: natural parameters $\theta = [\frac{\mu}{\sigma^2} \ \frac{-1}{2\sigma^2}]^T$, expectation parameters $\mathrm{d}\eta = [\mu \ \mu^2 + \sigma^2]^T$.

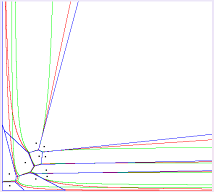
Bregman Voronoi diagrams: Bisectors

Two types of Voronoi diagrams defined by bisectors:

- First-type $H_{\mathbf{pq}}$ $H(\mathbf{p},\mathbf{q}) = \{\mathbf{x} \in \mathcal{X} \mid D_F(\mathbf{x}||\mathbf{p}) = D_F(\mathbf{x}||\mathbf{q})\}$ ()
- Second-type $H'_{\mathbf{pq}}$: $H'(\mathbf{p},\mathbf{q}) = \{\mathbf{x} \in \mathcal{X} \mid D_F(\mathbf{p}||\mathbf{x}) = D_F(\mathbf{q}||\mathbf{x})\}$



(Kullback-Leibler)



(Itakura-Saito)

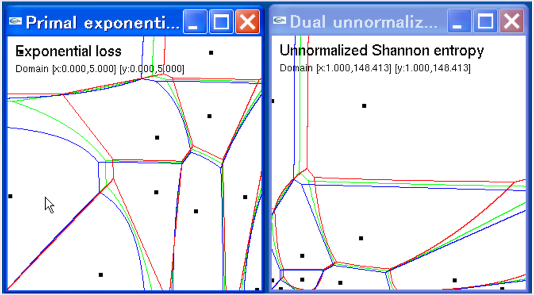
Affine/Curved Voronoi diagrams & Dualities

- Hyperplane $H(\mathbf{p},\mathbf{q})$: $\langle \mathbf{x}, \mathbf{p}' - \mathbf{q}' \rangle + F(\mathbf{p}) - \langle \mathbf{p}, \mathbf{p}' \rangle - F(\mathbf{q}) + \langle \mathbf{q}, \mathbf{q}' \rangle = 0$.
- Hypersurface $H'(\mathbf{p},\mathbf{q})$: $\langle \mathbf{x}', \mathbf{q} - \mathbf{p} \rangle + F(\mathbf{p}) - F(\mathbf{q}) = 0$.
  Curved in $\mathbf{x}$ but linear in $\mathbf{x}'$.

  (Dual of first-type bisector for gradient point set $\mathcal{S}'$ and D_{F*}; $\Rightarrow$ image by ∇_F is a hyperplane in $\mathcal{X}'$.)

Duality: $\mathrm{vor}_F(\mathcal{S}) \overset{\text{dual}}{\equiv} \mathrm{vor}'_{F*}(\mathcal{S}')$ and $\mathrm{vor}'_F(\mathcal{S}) \overset{\text{dual}}{\equiv} \mathrm{vor}_{F*}(\mathcal{S}')$.

Bregman Voronoi diagrams: Videos



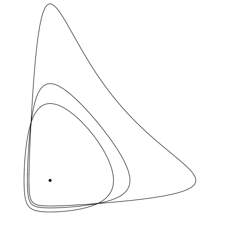
(Rasterized real-time on GPU, or computed exactly using the `qhull` package.)

Visit `http://www.csl.sony.co.jp/person/nielsen/BregmanVoronoi`

Symmetrized Bregman divergence is a Bregman divergence

$S_F(\mathbf{p},\mathbf{q}) = \frac{1}{2}(D_F(\mathbf{p},\mathbf{q}) + D_F(\mathbf{q},\mathbf{p})) = \frac{1}{2}\langle \mathbf{p}-\mathbf{q}, \mathbf{p}'-\mathbf{q}'\rangle.$

$$\begin{aligned} S_F(\mathbf{p},\mathbf{q}) &= \frac{1}{2}\left(D_F(\mathbf{p},\mathbf{q}) + D_F(\mathbf{q},\mathbf{p})\right) \\ &= \frac{1}{2}\left(D_F(\mathbf{p},\mathbf{q}) + D_{F^*}(\mathbf{p}',\mathbf{q}')\right) \\ &= D_{\hat{F}}(\hat{\mathbf{p}},\hat{\mathbf{q}}) \end{aligned}$$



where $\hat{\mathbf{p}} = (\mathbf{p},\mathbf{p}')$ and $\hat{F}(\hat{\mathbf{p}}) = \frac{1}{2}\left(F(\mathbf{p}) + F^*(\mathbf{p}')\right)$

Symmetrized bisector & Symmetrized Voronoi diagram

$H_{S_F}(\mathbf{p},\mathbf{q}) = \mathrm{proj}_{\mathcal{X}}\left(H_{D_{\hat{F}}}(\hat{\mathbf{p}},\hat{\mathbf{q}}) \cap \mathcal{M}\right)$ with $\mathcal{M} = \{(\mathbf{x},\mathbf{x}')\} \subset \mathbb{R}^{2d}$

(Double space dimension: from d-variate F to $2d$-variate $\hat{F}$.)

Space of Bregman spheres

Bregman spheres $\sigma(\mathbf{c}, r) = \{\mathbf{x} \in \mathcal{X} \mid D_F(\mathbf{x}, \mathbf{c}) = r\}$

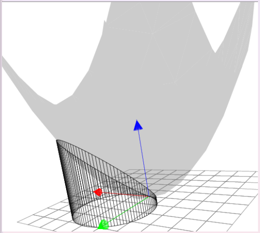
Lemma

The lifted image $\hat{\sigma}$ onto $\mathcal{F}$ of a Bregman sphere σ is contained in a hyperplane H_σ.

$(H_\sigma : z = \langle \mathbf{x} - \mathbf{c}, \mathbf{c}' \rangle + F(\mathbf{c}) + r)$

Conversely, the intersection of any hyperplane H with $\mathcal{F}$ projects vertically onto a Bregman sphere.

$(H : z = \langle \mathbf{x}, \mathbf{a} \rangle + b \longrightarrow \sigma = (\nabla_F{}^{-1}(\mathbf{a}), \langle \nabla_F{}^{-1}(\mathbf{a}), \mathbf{a} \rangle - F(\nabla_F{}^{-1}(\mathbf{a})) + b))$



(eg, usual paraboloid of revolution $\mathcal{F}$)

Algorithmics of Bregman spheres/balls

Union/intersection of Bregman balls

The union/intersection of n Bregman balls of $\mathcal{X}$ has complexity $\Theta(n^{\lfloor \frac{d+1}{2} \rfloor})$ and can be computed in time $\Theta(n \log n + n^{\lfloor \frac{d+1}{2} \rfloor})$

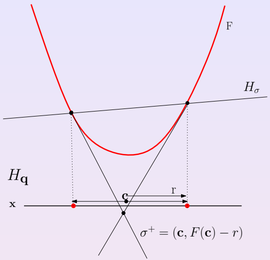
- Boundary of $\bigcap_i \sigma_i$: $\mathrm{Proj}_\perp$ of $\mathcal{F} \cap \left(\cap_{i=1}^n H_{\sigma_i}^\uparrow \right)$.
- Boundary of $\bigcup_i \sigma_i$: $\mathrm{Proj}_\perp$ of complement of $\mathcal{F} \cap \left(\cap_{i=1}^n H_{\sigma_i}^\uparrow \right)$.

Generalize the (Euclidean) space of spheres to Bregman spaces of spheres: radical axes, pencils of spheres, etc.

(Important for manifold reconstruction since every solid is a union of balls; ie., medial axis — power crust)

Polarity for symmetric divergences

The pole of H_σ is the point $\sigma^+ = (\mathbf{c}, F(c) - r)$ common to all the tangent hyperplanes at $H_\sigma \cap \mathcal{F}$



Polarity

Polarity preserves incidences $\sigma_1^+ \in H_{\sigma_2} \Leftrightarrow \sigma_2^+ \in H_{\sigma_1}$

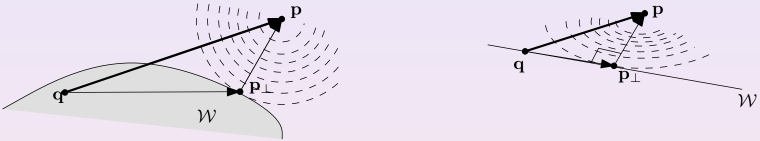
$$\begin{aligned} \sigma_1^+ \in H_{\sigma_2} & \Leftrightarrow F(\mathbf{c}_1) - r_1 = \langle \mathbf{c}_1 - \mathbf{c}_2, \mathbf{c}_2' \rangle + F(\mathbf{c}_2) + r_2 \\ & \Leftrightarrow D_F(\mathbf{c}_1, \mathbf{c}_2) = r_1 + r_2 = D_F(\mathbf{c}_2, \mathbf{c}_1) \\ & \Leftrightarrow \sigma_2^+ \in H_{\sigma_1} \end{aligned}$$

(Require symmetric divergences.)

Generalized Pythagoras theorem

Fact (Three-point)

$D_F(\mathbf{p}||\mathbf{q}) + D_F(\mathbf{q}||\mathbf{r}) = D_F(\mathbf{p}||\mathbf{r}) + \langle \mathbf{p}-\mathbf{q}, \mathbf{r}'-\mathbf{q}'\rangle.$



Fact (Bregman projection)

$\mathbf{p}_\perp$ *Bregman projection of t* $\mathbf{p}$ *onto convex subset* $\mathcal{W} \subseteq \mathcal{X}$*:*

$$\mathbf{p}_\perp = \mathrm{argmin}_{\mathbf{w}\in\mathcal{W}} D_F(\mathbf{w},\mathbf{p}).$$

Equality iff $\mathcal{W}$ is an affine set.

Dual orthogonality of bisectors with geodesics

$\mathrm{l}(\mathbf{p},\mathbf{q}) = \{\mathbf{x} : \mathbf{x} = \lambda\mathbf{p} + (1-\lambda)\mathbf{q}\}$ Straight line segment $[\mathbf{pq}]$

$\mathrm{c}(\mathbf{p},\mathbf{q}) = \{\mathbf{x} : \mathbf{x}' = \lambda\mathbf{p}' + (1-\lambda)\mathbf{q}'\}$ Geodesic $(\mathbf{p},\mathbf{q})$

Orthogonality (Projection)

X is Bregman orthogonal to Y if

$\forall\mathbf{x}\in X, \forall\mathbf{y}\in Y, \forall\mathbf{t}\in X\cap Y$

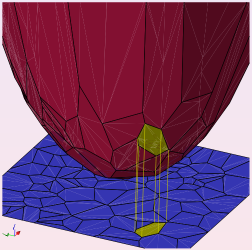
$$D_F(\mathbf{x},\mathbf{t}) + D_F(\mathbf{t},\mathbf{y}) = D_F(\mathbf{x},\mathbf{y}) \quad \Leftrightarrow \quad \langle \mathbf{x}-\mathbf{t}, \mathbf{y}'-\mathbf{t}'\rangle = 0$$

Lemma

$\mathrm{c}(\mathbf{p},\mathbf{q})$ is Bregman orthogonal to $H_{\mathbf{pq}}$

$\mathrm{l}(\mathbf{p},\mathbf{q})$ is Bregman orthogonal to $H'_{\mathbf{pq}}$

Bregman Voronoi diagrams from polytopes



Theorem

The first-type Bregman Voronoi diagram $\mathrm{vor}_F(\mathcal{S})$ is obtained by projecting by $\mathrm{Proj}_\perp$ the faces of the $(d+1)$-dimensional polytope $\mathcal{H} = \cap_i H^{\uparrow}_{\mathbf{p}_i}$ of $\mathcal{X}^+$ onto $\mathcal{X}$.

The Bregman Voronoi diagrams of a set of n d-dimensional points have complexity $\Theta(n^{\lfloor \frac{d+1}{2} \rfloor})$ and can be computed in optimal time $\Theta(n \log n + n^{\lfloor \frac{d+1}{2} \rfloor})$.

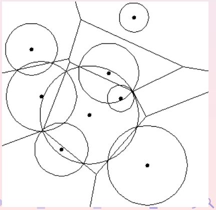
Bregman Voronoi diagrams from power diagrams

Affine Voronoi diagrams

The first-type Bregman Voronoi diagram of n sites of $\mathcal{X}$ is identical to the power diagram of n Euclidean hyperspheres centered at $\nabla_F(\mathcal{S}) = \{\mathbf{p}' \mid \mathbf{p} \in \mathcal{S}\}$

$$
\begin{aligned}
& D_F(\mathbf{x}, \mathbf{p}_i) \leq D_F(\mathbf{x}, \mathbf{p}_j) \\
\iff\quad & -F(\mathbf{p}_i) - \langle \mathbf{x} - \mathbf{p}_i, \mathbf{p}_i' \rangle) \leq -F(\mathbf{p}_j) - \langle \mathbf{x} - \mathbf{p}_j, \mathbf{p}_j' \rangle) \\
\iff\quad & \langle \mathbf{x}, \mathbf{x} \rangle - 2\langle \mathbf{x}, \mathbf{p}_i' \rangle - 2F(\mathbf{p}_i) + 2\langle \mathbf{p}_i, \mathbf{p}_i' \rangle \leq \langle \mathbf{x}, \mathbf{x} \rangle - 2\langle \mathbf{x}, \mathbf{p}_j' \rangle - 2F(\mathbf{p}_j) + 2\langle \mathbf{p}_j, \mathbf{p}_j' \rangle \\
\iff\quad & \langle \mathbf{x} - \mathbf{p}_i', \mathbf{x} - \mathbf{p}_i' \rangle - r_i^2 \leq \langle \mathbf{x} - \mathbf{p}_j', \mathbf{x} - \mathbf{p}_j' \rangle - r_j^2
\end{aligned}
$$

$\mathbf{p}_i \rightarrow \sigma_i = \mathrm{Ball}(\mathbf{p}_i', r_i)$ with $r_i^2 = \langle \mathbf{p}_i', \mathbf{p}_i' \rangle + 2(F(\mathbf{p}_i) - \langle \mathbf{p}_i, \mathbf{p}_i' \rangle)$



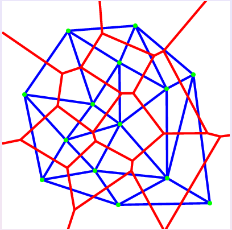
Straight triangulations from polytopes

Several ways to define Bregman triangulations

Definition (Straight triangulation)

$\hat{\mathcal{S}}$: lifted image of $\mathcal{S}$
$\mathcal{T}$: lower convex hull of $\hat{\mathcal{S}}$

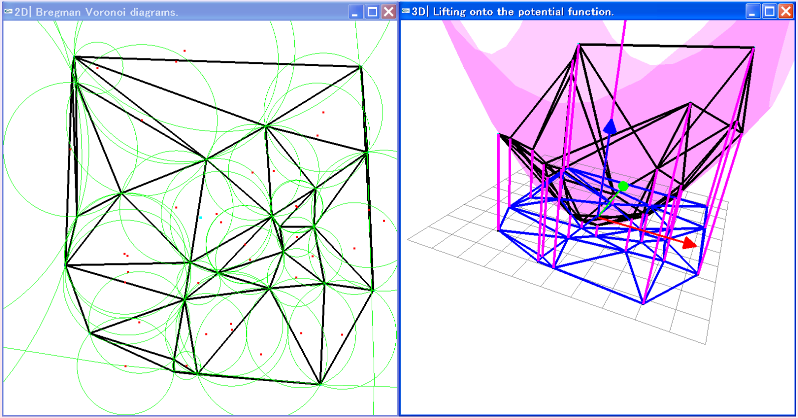
The vertical projection of $\mathcal{T}$ is called the straight Bregman triangulation of $\mathcal{S}$



Properties

- Characteristic property : The Bregman sphere circumscribing any simplex of $BT(\mathcal{S})$ is empty
- Optimality : $BT(\mathcal{S}) = \min_{T \in \mathcal{T}(\mathcal{S})} \max_{\tau \in T} r(\tau)$
  ($r(\tau) =$ radius of the smallest Bregman ball containing τ)
  [Rajan]

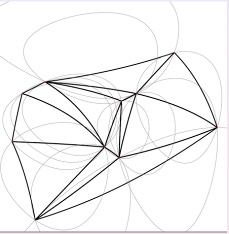
Bregman Voronoi diagram & triangulation from polytopes



(Implemented using OpenGL®.)

Bregman triangulations: Geodesic triangulations

The straight triangulation is not necessarily the dual of the Bregman Voronoi diagram. The dual triangulation is the *geodesic triangulation* (bisector/geodesic Bregman orthogonality).



Geodesic triangulation

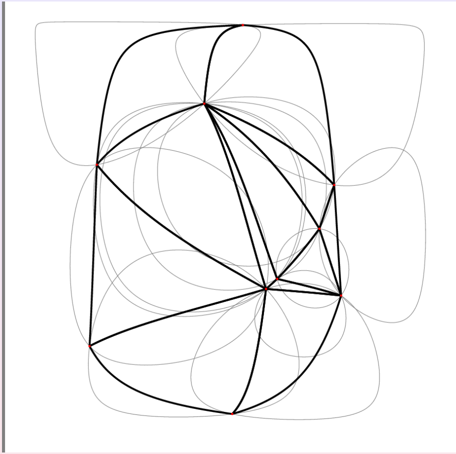
- Image of triangulation by ∇_F^{-1} is a curved triangulation
- Edges: geodesic arcs joining two sites

Symmetric divergences

For symmetric divergences, straight and geodesic triangulations are combinatorially equivalent (polarity).

(For $D_F = L_2^2$, straight and geodesic triangulations are the same ordinary Delaunay triangulations.)

Geodesic triangulation: Hellinger-type



(Hellinger distance $D^{(0)}(p||q) = 2\int(\sqrt{p(x)} - \sqrt{q(x)})\mathrm{d}x$, a particular case of f-divergence)

($D^{(1)}$ is Kullback-Leibler)

Riemannian geometry

Bregman geometry can be tackled from the viewpoint of Riemannian geometry:

- Riemannian geometry & Information geometry [AN'00],

  (mostly Fischer metrics for statistical manifolds)
- Calculus on manifolds: Parallel transports of vector fields and connections,
- geodesic Voronoi diagram $\int_{\mathbf{a}}^{\mathbf{b}} \sqrt{\mathbf{g}_{ij}\gamma^{i'}\gamma_j'}\mathrm{d}s$ (metric $\mathbf{g}_{ij}$)
- Canonical divergences & second- or third-order metric approximations

  ($\mathbf{g}^{(D_F)} = \mathbf{g}^{(D_{F^*})}$ with $\mathbf{g}_{ij}^{(D_F)} = -D_F(\partial_i||\partial_j) = D_F(\partial_i\partial_j||\cdot) = D_F(\cdot||\partial_i\partial_j)$)

Bregman geometry: A special case of Riemannian geometry

- Dual affine coordinate systems ($\mathbf{x}$ and $\mathbf{x}'$) of *non-metric* connections ∇ and ∇^* (D_F and D_{F^*} are non-conformal representations)
- Torsion-free & zero-curvature space (dually flat space).

Extensions

In our paper, we further consider

- Weighted Bregman diagrams:
  $WD_F(\mathbf{p}_i, \mathbf{p}_j) = D_F(\mathbf{p}_i, \mathbf{p}_j) + w_i - w_j$
  (incl. k-order Bregman diagrams)
- k-jet Bregman divergences (tails of Taylor expansions)
- Centroidal Voronoi diagrams (centroid & Bregman information).
- Applications to ϵ-sampling and minmax quantization
- Applications to machine learning

Visit http://www.csl.sony.co.jp/person/nielsen/BregmanVoronoi/

References

B+'96 Boissonnat et al., *Output-sensitive construction of the Delaunay triangulation of points lying in two planes*, Internat. J. Comput. Geom. Appl., 6(1):1-14, 1996.

OT'96 Onishi & Takayama, *Construction of Voronoi diagram on the upper half-plane*, IEICE Trans. 79-A, pp. 533-539, 1996.

S+'98 Sadakane et al., *Voronoi diagrams by divergences with additive weights*, SoCG video, 1998.

AK'00 Aurenhammer & Klein, *Voronoi diagrams*, Handbook of Comp. Geom. , Elsevier , pp. 201-290, 2000.

AN'00 Amari & Nagaoka, *Methods of information geometry*, Oxford University press 2005.

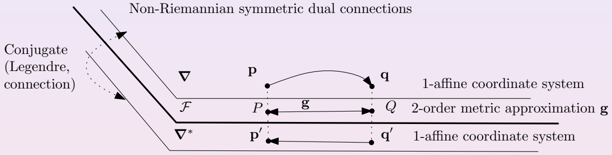
LL'00 Leibon & Letscher, *Delaunay triangulations and Voronoi diagrams for Riemannian manifolds*, SoCG, pp. 341-349, 2000.

B+'05 Banerjee et al., *Clustering with Bregman Divergences*, JMLR 2005.

Thank You!

E-mail: Frank.Nielsen@acm.org

Bregman geometry: Dually flat space



Centroid and Bregman information

Centroid

The (weighted) Bregman centroid of a compact domain $\mathcal{D}$ is $\mathbf{c}^* = \arg\min_{\mathbf{c}\in\mathcal{D}} \int_{\mathbf{x}\in\mathcal{D}} p(\mathbf{x})\, D_F(\mathbf{x},\mathbf{c})\, \mathrm{d}x$

Lemma

$\mathbf{c}^*$ coincides with the (weighted) centroid of $\mathcal{D}$

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{c}} \int_{\mathbf{x}\in\mathcal{D}} p(\mathbf{x})\, D_F(\mathbf{x},\mathbf{c})\, \mathrm{d}x &= \frac{\partial}{\partial \mathbf{c}} \int_{\mathbf{x}\in\mathcal{D}} p(\mathbf{x})\, \left(F(\mathbf{x}) - F(\mathbf{c}) - \langle \mathbf{x}-\mathbf{c}, \nabla_F(\mathbf{c})\rangle\right) \mathrm{d}x \\
&= -\int_{\mathbf{x}\in\mathcal{D}} p(\mathbf{x})\, {\nabla_F}^2(\mathbf{c})(\mathbf{x}-\mathbf{c})\mathrm{d}x \\
&= -{\nabla_F}^2(\mathbf{c}) \left(\int_{\mathbf{x}\in\mathcal{D}} \mathbf{x}p(\mathbf{x})\mathrm{d}x - \mathbf{c}\int_{\mathbf{x}\in\mathcal{D}} p(\mathbf{x})\mathrm{d}x \right)
\end{aligned}$$

vanishes for $\mathbf{c}^* = \frac{\int_{\mathbf{x}\in\mathcal{D}} \mathbf{x}p(\mathbf{x})\mathrm{d}x}{\int_{\mathbf{x}\in\mathcal{D}} p(\mathbf{x})\mathrm{d}x}$

Bregman information

Bregman information

$\mathbf{x}$ random variable, pdf = $p(\mathbf{x})$

Distortion rate
$D_F(\mathbf{c}) = \int_{\mathbf{x}\in\mathcal{X}} p(\mathbf{x})\, D_F(\mathbf{x},\mathbf{c})\,\mathrm{d}x$

Bregman information
$\inf_{\mathbf{c}\in\mathcal{X}} D_F(\mathbf{c})$

Bregman representative
$\mathbf{c}^* = \int_{\mathbf{x}\in\mathcal{X}} \mathbf{x}\, p(\mathbf{x})\,\mathrm{d}x = E(\mathbf{x})$

Centroidal Bregman Voronoi diagrams and least-square quantization

Lloyd relaxation

1. Choose k sites
2. repeat
    1. Compute the Bregman Voronoi diagram of the k sites
    2. Move the sites to the centroids of their cells

   until convergence

ε-sampling and minmax quantization

ε-sample

$\mathrm{error}(P) = \max_{\mathbf{x}\in\mathcal{D}} \min_{\mathbf{p}_i\in P} D_F(\mathbf{x},\mathbf{p}_i)$

A finite set of points P of $\mathcal{D}$ is an ε-sample of $\mathcal{D}$ iff $\mathrm{error}(P) \leq \varepsilon$

Local maxima

$\mathrm{error}(P) = \max_{\mathbf{v}\in V} \min_{\mathbf{p}_i\in P} D_F(\mathbf{x},\mathbf{p}_i)$.

where V consists of the vertices of $BVD(P)$ and intersection points between the edges of $BVD(P)$ and the boundary of $\mathcal{D}$

Ruppert's algorithm

Insert a vertex of the current Bregman Voronoi diagram if its Bregman radius is greater than ε

Implied by the following lemma and a packing argument

Fatness of Bregman balls

If F is of class C^2, there exists two constants γ' and γ'' such that

$$EB(\mathbf{c}, \gamma'\sqrt{r}) \subset B(\mathbf{c}, r) \subset EB(\mathbf{c}, \gamma''\sqrt{r})$$

Number of sample points

$$\frac{\mathrm{area}(\mathcal{D})}{\gamma''\pi\varepsilon^2} \leq |P| \leq \frac{4\mathrm{area}(\mathcal{D}^{+\frac{\varepsilon}{2}})}{\gamma'\pi\varepsilon^2}$$