On the Smallest Enclosing Information Disk

Frank Nielsen[1] Richard Nock[2]

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

[2]University of Antilles-Guyanne
DSI-GRIMAAG
Richard.Nock@martinique.univ-ag.fr

August 2006

Smallest Enclosing Balls

Problem

Given $\mathcal{S} = \{\mathbf{s}_1, ..., \mathbf{s}_n\}$*, compute a simplified description, called the* **center***, that fits well* $\mathcal{S}$ *(i.e., summarizes* $\mathcal{S}$*).*

Two optimization criteria:

- MINAVG: Find a center $\mathbf{c}^*$ which minimizes the *average distortion* w.r.t $\mathcal{S}$: $\mathbf{c}^* = \mathrm{argmin}_{\mathbf{c}} \sum_i d(c, \mathbf{s}_i)$.
- MINMAX: Find a center $\mathbf{c}^*$ which minimizes the *maximal distortion* w.r.t $\mathcal{S}$: $\mathbf{c}^* = \mathrm{argmin}_{\mathbf{c}} \max_i d(c, \mathbf{s}_i)$.

Investigated in Applied Mathematics:

- Computational geometry (1-center problem),
- Computational statistics (1-point estimator),
- Machine learning (1-class classification),

Smallest Enclosing Balls in Computational Geometry

Distortion measure $d(\cdot, \cdot)$ is the geometric distance:

Euclidean distance L_2.

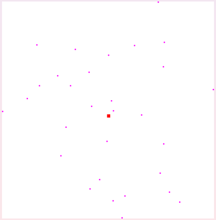
$\mathbf{c}^*$ is the **circumcenter** of $\mathcal{S}$ for MINMAX,

Squared Euclidean distance L_2^2.

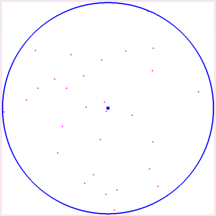
$\mathbf{c}^*$ is the **centroid** of $\mathcal{S}$ for MINAVG ($\rightarrow$ k-means),

Euclidean distance L_2.

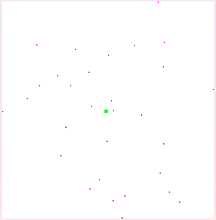
$\mathbf{c}^*$ is the **Fermat-Weber** point for MINAVG.



Centroid
MINAVG L_2^2



Circumcenter
MINMAX L_2



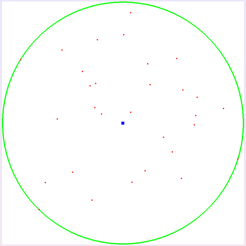
Fermat-Weber
MINAVG L_2

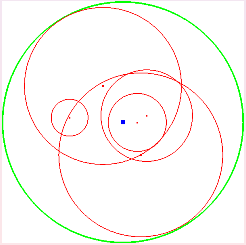
MinMax in Computational Geometry (MINIBALL)

Smallest Enclosing Ball [NN'04]

- Pioneered by Sylvester (1857),
- Unique circumcenter $\mathbf{c}^*$ (radius r^*),
- LP-type, linear-time randomized algorithm (fixed dimension d),
- Weakly polynomial.
- Efficient SOCP numerical solver,
- Fast combinatorial heuristics ($d \geq 1000$).

MinMax point set





MinMax ball set

Distortions: Bregman Divergences

Definition

Bregman divergences are parameterized (F) families of distortions.

Let $F : \mathcal{X} \longrightarrow \mathbb{R}$, such that F is *strictly convex* and *differentiable* on $\mathrm{int}(\mathcal{X})$, for a convex domain $\mathcal{X} \subseteq \mathbb{R}^d$.

Bregman divergence D_F:

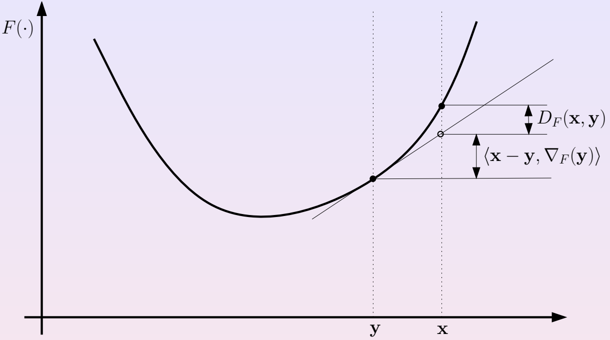
$$D_F(\mathbf{x}, \mathbf{y}) = F(\mathbf{x}) - F(\mathbf{y}) - \langle \mathbf{x} - \mathbf{y}, \nabla_F(\mathbf{y}) \rangle .$$

∇_F : gradient operator of F

$\langle \cdot, \cdot \rangle$: Inner product (dot product)

($\rightarrow D_F$ is the tail of a Taylor expansion of F)

Visualizing F and D_F



$$D_F(\mathbf{x}, \mathbf{y}) = F(\mathbf{x}) - F(\mathbf{y}) - \langle \mathbf{x} - \mathbf{y}, \nabla_F(\mathbf{y}) \rangle .$$

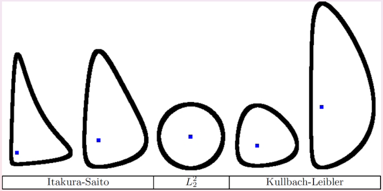
($\rightarrow$ D_F is the a truncated Taylor expansion of F)

Bregman Balls (Information Balls)

- Euclidean Ball: $\mathcal{B}_{\mathbf{c},r} = \{\mathbf{x} \in \mathcal{X} : \|\mathbf{x} - \mathbf{c}\|_2^2 \leq r\}$
  (r: *squared* radius. L_2^2: Bregman divergence $F(\mathbf{x}) = \sum_{i=1}^d x_i^2$)

Theorem [BMDG'04]

The MinAvg Ball for Bregman divergences is the **centroid** .



Two types of Bregman balls

- First-type:
  $\mathcal{B}_{\mathbf{c},r} = \{\mathbf{x} \in \mathcal{X} : D_F(\boxed{\mathbf{c}}, \mathbf{x}) \leq r\}$,
- Second-type:
  $\mathcal{B}'_{\mathbf{c},r} = \{\mathbf{x} \in \mathcal{X} : D_F(\mathbf{x}, \boxed{\mathbf{c}}) \leq r\}$

Lemma

The smallest enclosing Bregman balls $\mathcal{B}_{\mathbf{c}^*,r^*}$ and $\mathcal{B}'_{\mathbf{c}^*,r^*}$ of $\mathcal{S}$ are unique.

$\longrightarrow$ Consider first-type Bregman balls.
(The second-type is obtained as a first-type ball on the dual divergence D_{F^*} using the Legendre-Fenchel transformation.)

Applications of Bregman Balls

Circumcenters of the smallest enclosing Bregman balls encode:

Euclidean squared distance. The <u>closest</u> point to a set of points.

$$D_F(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{d}(q_i - p_i)^2 = ||\mathbf{p}||^2 + ||\mathbf{q}||^2 - 2\langle \mathbf{p},\mathbf{q}\rangle.$$

Itakura-Saito divergence. The <u>closest</u> (sound) signal to a set of signals (speech recognition).

$$D_F(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{d}\left(\frac{p_i}{q_i} - \log\frac{p_i}{q_i} - 1\right), [\leftarrow\ F(\mathbf{x}) = -\sum_{i=1}^{d}\log x_i]$$

Kullback-Leibler. The <u>closest</u> distribution to a set of distributions (density estimation).

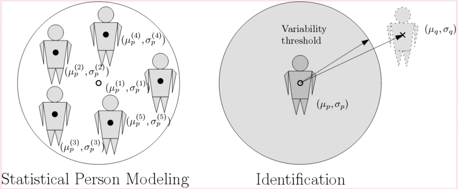
$$D_F(\mathbf{p},\mathbf{q}) = \sum_{i=1}^{d} p_i \log\frac{p_i}{q_i} - p_i + q_i, [F(\mathbf{x}) = -\sum_{i=1}^{d} x_i \log x_i]$$

Information Disks

Problem

Given a set $\mathcal{S} = \{\mathbf{s}_1, ..., \mathbf{s}_n\}$ *of* n *2D vector points, compute the* MINMAX **center**: $\mathbf{c}^* = \mathrm{argmin}_{\mathbf{c}} \max_i d(c, \mathbf{s}_i)$.

- handle geometric points for various distortions,
- handle parametric distributions (*e.g.*, Normal distributions are parameterized by (μ, σ)).



Information Disk is LP-type

Monotonicity. For any $\mathcal{F}$ and $\mathcal{G}$ such that $\mathcal{F} \subseteq \mathcal{G} \subseteq \mathcal{X}$, $r^*(\mathcal{F}) \leq r^*(\mathcal{G})$.

Locality. For any $\mathcal{F}$ and $\mathcal{G}$ such that $\mathcal{F} \subseteq \mathcal{G} \subseteq \mathcal{X}$ with $r^*(\mathcal{F}) = r^*(\mathcal{G})$, and any point $\mathbf{p} \in \mathcal{X}$,

$$r^*(\mathcal{G}) < r^*(G \cup \{\mathbf{p}\}) \rightarrow r^*(\mathcal{F}) < r^*(F \cup \{\mathbf{p}\}).$$

MiniInfoBall($\mathcal{S} = \{\mathbf{p}_1, ..., \mathbf{p}_n\}, \mathcal{B}$):

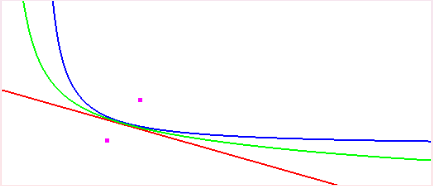
```
◁ Initially B = ∅. Returns B* = (c*, r*) ▷
IF |S ∪ B| ≤ 3
   RETURN B=SolveInfoBasis(S ∪ B)
ELSE
   ◁ Select at random p ∈ S ▷
   B*=MiniInfoBall(S\{p}, B)
   IF p ∉ B*
      ◁ Then add p to the basis ▷
      MiniInfoBall(S\{p}, B ∪ {p})
```

Lemma

The first-type Bregman bisector
$\mathrm{Bisector}(\mathbf{p},\mathbf{q}) = \{\mathbf{c} \in \mathcal{X} \mid D_F(\mathbf{c},\mathbf{p}) = D_F(\mathbf{c},\mathbf{q})\}$ *is linear.*

This is a linear equation in $\mathbf{c}$ (an *hyperplane*). Bisector
$\mathrm{Bisector}(\mathbf{p},\mathbf{q}) = \{\mathbf{x} \mid \langle \mathbf{x}, \mathbf{d_{pq}} \rangle + k_{\mathbf{pq}} = 0\}$ with

- $\mathbf{d_{pq}} = \nabla_F(\mathbf{p}) - \nabla_F(\mathbf{q})$ a vector, and
- $k_{\mathbf{pq}} = F(\mathbf{p}) - F(\mathbf{q}) + \langle \mathbf{q}, \nabla_F(\mathbf{q}) \rangle - \langle \mathbf{p}, \nabla_F(\mathbf{p}) \rangle$ a constant



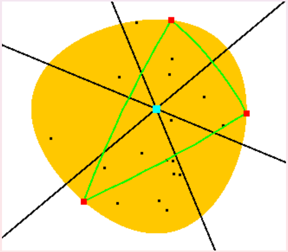
(Itakura-Saito divergence)

Computing basis (SolveInfoBasis)

Basis 3: The circumcenter is the trisector.
(intersection of 3 linear bisectors, enough to consider any two of them).

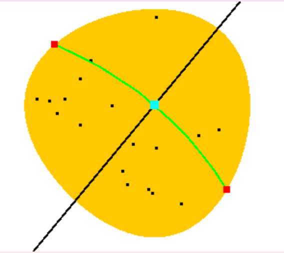
$$\mathbf{c}^* = \mathrm{l}_{12} \times \mathrm{l}_{13} = \mathrm{l}_{12} \times \mathrm{l}_{23} = \mathrm{l}_{13} \times \mathrm{l}_{23},$$

l_{ij}: projective point associated to the linear bisector $\mathrm{Bisector}(\mathbf{p}_i, \mathbf{p_j})$ ($\times$: cross-product)



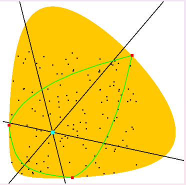
Computing basis (SOLVEINFOBASIS)

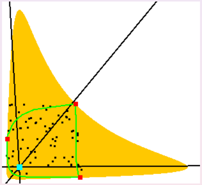
Basis 2: Either minimize $D_F(\mathbf{c}, \mathbf{p})$ s.t. $\mathbf{c}^* \in \mathrm{Bisector}(\mathbf{p}, \mathbf{q})$, or better perform a logarithmic search on $\lambda \in [0, 1]$ s. t. $\mathbf{r}_\lambda = \nabla_F{}^{-1}((1-\lambda)\nabla_F(\mathbf{p}) + \lambda\nabla_F(\mathbf{q}))$ is on the geodesic of $\mathbf{pq}$ ($\nabla_F{}^{-1}$: reciprocal gradient).

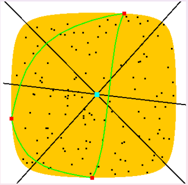


Live Demo

http://www.csl.sony.co.jp/person/nielsen/BregmanBall/MINIBALL/







Statistical application example

Univariate Normal law distribution:
$N(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$.

Consider the Kullback-Leibler divergence of two distributions:
$\mathrm{KL}(f,g) = \int_x f(x) \log \frac{f(x)}{g(x)}$.
Canonical form of an *exponential family*:
$N(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi} Z(\boldsymbol{\theta})} \exp\{\langle \boldsymbol{\theta}, \mathbf{f}(x) \rangle\}$ with:

- $Z(\boldsymbol{\theta}) = \sigma \exp\{\frac{\mu^2}{2\sigma^2}\} = \sqrt{-\frac{1}{2\theta_1}} \exp\{-\frac{\theta_2^2}{4\theta_1}\}$,
- $\mathbf{f}(x) = [x^2 \; x]^T$: *sufficient statistics*,
- $\boldsymbol{\theta} = [-\frac{1}{2\sigma^2} \; \frac{\mu}{\sigma^2}]^T$: *natural parameters*.

Kullback-Leibler of parametric exponential family is a Bregman divergence for $F = \log Z$.

$$\mathrm{KL}(\theta_p || \theta_q) = D_F(\theta_p, \theta_q) = \langle (\theta_p - \theta_q), \theta_p[\mathbf{f}] \rangle + \log \frac{Z(\theta_q)}{Z(\theta_p)}$$

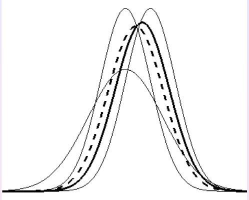
$$\theta_p[\mathbf{f}] = \begin{bmatrix} \int_x \frac{x^2}{Z(\theta_p)} \exp\{\langle \theta_p, \mathbf{f}(x) \rangle\} \\ \int_x \frac{x}{Z(\theta_p)} \exp\{\langle \theta_p, \mathbf{f}(x) \rangle\} \end{bmatrix} = \begin{bmatrix} \mu_p^2 + \sigma_p^2 \\ \mu_p \end{bmatrix}$$

Bisector $\langle (\theta_p - \theta_q), \theta_c[\mathbf{f}] \rangle + \log \frac{Z(\theta_p)}{Z(\theta_q)} = 0$.

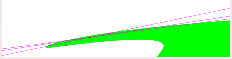
1D Gaussian distribution: change variables

$(\mu, \sigma) \rightarrow (\mu^2 + \sigma^2, \mu) = (x, y)$ (with $x > y > 0$).
It comes $Z(x, y) = \sqrt{x - y^2} \exp\{\frac{y^2}{2(x-y^2)}\}$,
$\log Z(x, y) = \log \sqrt{x - y^2} + \frac{y^2}{2(x-y^2)}$ and
$\nabla_F(x, y) = (\frac{1}{2(x-y^2)} - \frac{y^2}{2(x-y^2)}, \frac{y^3}{(x-y^2)^2})$.

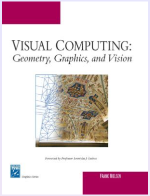
Statistical application example (cont'd)



- MinMax: $(\mu^*, \sigma^*) \simeq (2.67446, 1.08313)$ and

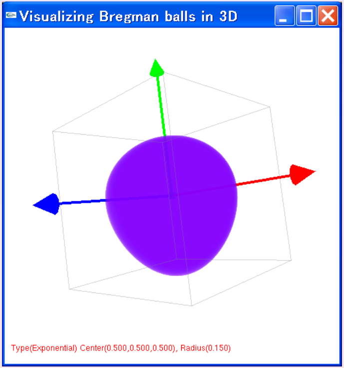
  $r^* \simeq 0.801357$, 
- MinAvg: $(\mu^{*'}, \sigma^{*'}) = (2.40909, 1.10782)$.

Note that $\mathrm{KL}(N_i, N_j) = \frac{1}{2}\left(\frac{\sigma_i^2}{\sigma_j^2} + 2\log\frac{\sigma_j}{\sigma_i} - 1 + \frac{(\mu_j - \mu_i)^2}{\sigma_j^2}\right)$.

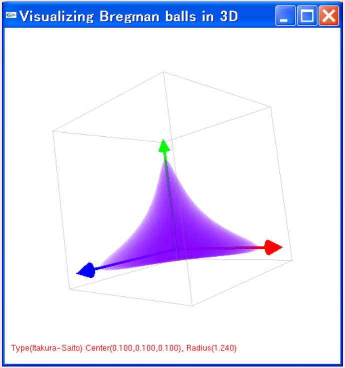


- Java Applet online:
  www.csl.sony.co.jp/person/nielsen/BregmanBall/MINIBALL/
- Source code: Basic MiniBall, Line intersection by projective geometry
  Visual Computing: Geometry, Graphics, and Vision, ISBN 1-58450-427-7, 2005.
- In high dimensions, extend Bădoiu & Clarkson core-set
  See *On approximating the smallest enclosing Bregman Balls* (SoCG'06 video)

3D Bregman balls (video)

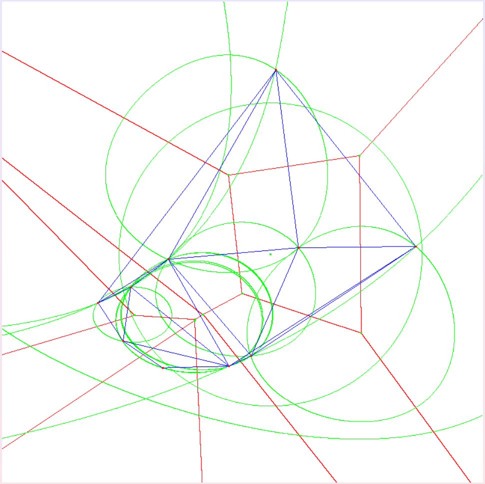


Relative entropy (KL)



Itakura-Saito

Bregman Voronoi/Delaunay



EXP

Bibliography

- Welzl, ”Smallest Enclosing Disks (Balls and Ellipsoids)”, LNCS 555:359-370, 1991.
- Crammer & Singer, ”Learning Algorithms for Enclosing Points in Bregmanian Spheres”, COLT03.
- Nock & Nielsen, ”Fitting the smallest Bregman ball”, ECML05 (SoCG06 video).
- Banerjee et. al, ”Clustering with Bregman divergences” , JMLR05.