CIMPA Research School : Data Science for Engineering and Technology
Tunis 2019

Bregman divergences
a basic tool for pseudo-metrics building for data structured by physics

3- Computational Geometry with Bregman divergences
- 3a- Bregman divergences
- 3b- Bregman gaps

Stéphane ANDRIEUX

ONERA - France

Member of the National Academy of Technologies of France

The essential geometric property of BD

Recalling the centroid definition of a finite set S of N points x_i in IR^N $\quad \mu = \frac{1}{N} \sum_{i=1,N} x_i$

Centroid property for Bregman divergences

For any set S *of* N *points in* IR^n, *and* *any* *Bregman divergence* D_J, *we have*:

$$\mu = \frac{1}{N} \sum_{i=1,N} x_i = \underset{s \in \mathrm{IR}^n}{\arg\min}\, C_J(s) \equiv \frac{1}{N} \sum_{i=1,N} D_J(x_i, s)$$

The property extends to the probabilistic context with the expectation $E_\nu[f(x)] = \frac{1}{N} \sum_{i=1,N} \nu_i f(x_i)$

The mean or centroïd of a finite set S of N points minimizes the sum of Bregman divergences between this point and any other points of the set S, whatever the choice of the Bregman divergence.

Bannerje et al. (2005) proved the reverse proposition : that any function $D(x, y)$ satisfying the property is necessarily a Bergman divergence

The centroid property is then a **characterization of Bregman divergence**

The essential geometric property of BD (*proof*)

$$\mu = \frac{1}{N}\sum_{i=1,N} x_i = \underset{s\,\in\,\mathrm{IR}^n}{\arg\min}\, C_J(s) \equiv \frac{1}{N}\sum_{i=1,N} D_J(x_i, s)$$

$$\begin{aligned} C_J(s) - C_J(\mu) &= \frac{1}{N}\sum_{i=1,N} J(x_i) - J(s) - (\nabla J(s), x_i - s) - \frac{1}{N}\sum_{i=1,N} J(x_i) - J(\mu) - (\nabla J(\mu), x_i - \mu) \\ &= J(\mu) - J(s) - \left(\nabla J(s), \frac{1}{N}\sum_{i=1,N} x_i - s\right) + \left(\nabla J(\mu), \frac{1}{N}\sum_{i=1,N} x_i - \mu\right) \\ &= J(\mu) - J(s) - (\nabla J(s), \mu - s) \end{aligned}$$

Bregman divergence generated by *J* between μ and s, then ≥ *0*

Geometry computational what is it ?

Computational geometry (in its simplest sense) deals with:

- Defining geometrical concepts
  - Ball, sphere or more complicated object
- Defining elementary geometrical operations for a given point
  - Compute the distance to a set, perform a projection, find the nearest neighbor, find the equidistant set of two distinct points

The challenge here :

Most, if not every, concepts or methods rely on a fundamental property of scalar products: the triangle inequality of Pythagorean theorem, a property not satisfied by the Bregman divergences or Bregman gaps

First general geometric properties of BD

Hyperplane separation or bisectors $BBs_J(e_1,e_2)=\{x, D_J(x,e_1)=D_J(x,e_2)\}$

The set of equidistant points x with respect to a Bregman divergence to a pair (e_1,e_2), also called named Bregman bisector of the first type of (e_1, e_2).is an hyperplane. Furthermore the points e_1 and e_2 lie on the different sides of this hyperplane

Affine function of x

$$D_J\left(x,e_1\right)=D_J\left(x,e_2\right) \longrightarrow \cancel{J(x)}-J(e_1)-\left\langle \underline{p}_1, x-e_1\right\rangle = \cancel{J(x)}-J(e_2)-\left\langle \underline{p}_2, x-e_2\right\rangle$$

Bregman projection on a closed convex sets $\forall C\ \forall e\ \exists \tilde{e}\ \ s.t.\ \ \tilde{e}=\arg\min_{h\in C} D_J(h,e)$

For any closed convex subset C of IR^{n}xIR^n and point e, there exists a projection $\tilde{e}$ onto C in the sense of the Bregman divergence, the projection is unique if J is strictly convex

$$\underset{h\in C}{Min}\ D_J(h,e)=\underset{h\in C}{Min}\left[J(h)-\left\langle \underline{p},h\right\rangle\right]-J(e)+\left\langle \underline{p},e\right\rangle$$

J , C are convex , C is closed $\rightarrow$ $\tilde{e}$ exists, belongs to ∂C, unique if J strictly convex

Where we are still interested in triangular inequality

Weak forms of the triangle inequality

For any triplet (e_1, e_2, e_3) :

$$D_J(e_1,e_3) = D_J(e_1,e_2) + D_J(e_2,e_3) - \langle \nabla J(e_1) - \nabla J(e_2), e_1 - e_2 \rangle$$

With $\tilde{e}$ the projection of e on C

If C is a closed convex subset of IR^n

$$\forall e_3 \in IR^n \; and \; \forall e_1 \in C \;,\; D_J(e_1,e_3) \geq D_J(e_1,\tilde{e}_3) + D_J(\tilde{e}_3,e_3)$$

If C is also an affine subspace

$$\forall e_3 \in IR^n \; and \; \forall e_1 \in C \;,\; D_J(e_1,e_3) = D_J(e_1,\tilde{e}_3) + D_J(\tilde{e}_3,e_3)$$

Bregman balls and spheres

The generalization of balls and sphere concepts is straightforward

Bregman balls

The Bregman balls, associated to the generating function J, with center c and pseudo-radius r are the sets:

$$BB_J(c,\rho) = \{e, D_J(e,c) \le \rho\}$$

Convex because of convexity of D_J relatively to its first variable

$$BB'_J(c,\rho) = \{e, D_J(c,e) \le \rho\}$$

Not necessarily convex because of non-convexity of D_J relatively to its second variable

The associated Bregman spheres are respectively

$\partial BB_J(c,\rho)$ and $\partial BB'_J(c,\rho)$

If $0 \in \mathrm{Dom}\, J$

Origin centered Bregman balls If $\nabla J(0)=0$:

$$BB_J(0,\rho) = \{e, J(e) \le \rho\}$$
$$BB'_J(0,\rho) = \{e, \langle \nabla J(e), e\rangle - J(e) \le \rho\}$$

Origin centered Bregman balls for positively homogeneous generating functions

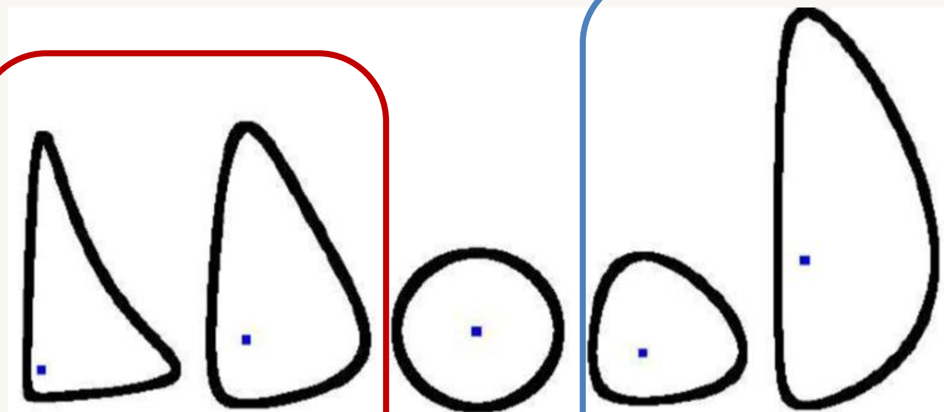
J convex positively homogenous function of degree α (> 1)

$$BB_J(0,\rho) = \{e, J(e) \le \rho\} \quad BB'_J(0,\rho) = \{e, (\alpha-1)J(e) \le \rho\}$$

Convex
Homothetic

Bregman balls, what do they look like?

2D



Itakura-Saito

Kullback-Leibler

centers

3D

Kullback-Leibler   Itakura-Saito   Logistic

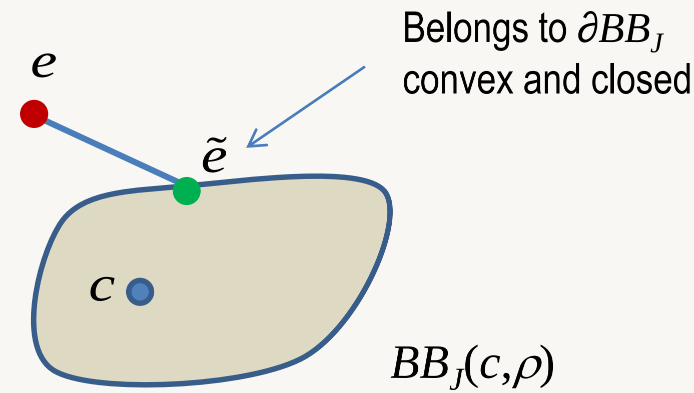
Bregman Projection onto Bregman balls

$$D_J(e_1,e_2) = J(e_1) - J(e_2) - \langle \nabla J(e_2), e_1 - e_2 \rangle$$

Defined for the first type of BB because convex

$$\tilde{e} = \underset{x \in BB_J(c,\rho)}{\arg\min} \; D_J(x,e)$$

$$BB_J(c,\rho) = \{x, D_J(x,c) \le \rho\}$$



$$\begin{aligned} L(x,\mu) &= D_J(x,c) + \mu\left(BB_J(x,c) - \rho\right) \\ &= (1+\mu)J(x) - J(e) - \mu J(c) - \langle \nabla J(e) + \mu \nabla J(c), x \rangle + \langle \nabla J(e), e \rangle + \mu \langle \nabla J(c), c \rangle \end{aligned}$$

$$\begin{aligned} D_x L(\tilde{e},\lambda).\delta x &= (1+\lambda)\langle \nabla J(\tilde{e}), \delta x \rangle - \langle \nabla J(e) + \lambda \nabla J(c), \delta x \rangle \quad \forall \delta x \\ &= \langle (1+\lambda)\nabla J(\tilde{e}) - \nabla J(e) - \lambda \nabla J(c), \delta x \rangle \qquad \forall \delta x \\ &= 0 \end{aligned}$$

Setting $\eta = \lambda / (1+\lambda) \in]0,1[$ $\longrightarrow$ $\boxed{\nabla J(\tilde{e}) = (1-\eta)\nabla J(e) + \eta \nabla J(c)}$

Bregman Projection onto Bregman balls (cont.)

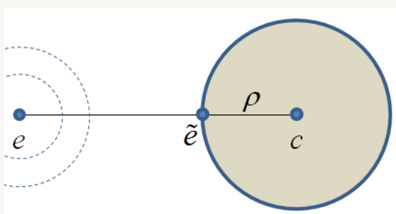
$$D_J(e_1,e_2) = J(e_1) - J(e_2) - \langle \nabla J(e_2), e_1 - e_2 \rangle$$

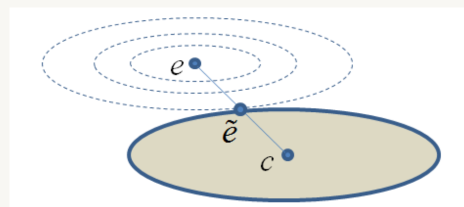
$$\tilde{e} = \underset{x \in BB_J(c,\rho)}{\arg\min} \; D_J(x,e)$$

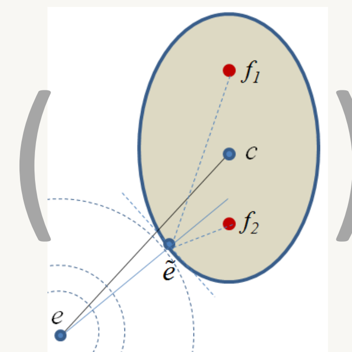
$$BB_J(c,\rho) = \{x, D_J(x,c) \le \rho\}$$

$$\nabla J(\tilde{e}) = (1-\eta)\nabla J(e) + \eta \nabla J(c)$$

Except for the Mahalanobis distances, the projection lies <u>not</u> on the segment [*e*,*c*]







$$D_J\left(\nabla J^*[\lambda_0 \nabla J(e) + (1-\lambda_0)\nabla J(c)], c\right) = \rho \longrightarrow \boxed{\tilde{e} = \nabla J^*[\eta \nabla J(e) + (1-\eta)\nabla J(c)]}$$

Using the conjugate function J* of *J*

$$J^*(p) = \sup_{c} \langle p,c\rangle - J(c) = \langle p,e\rangle - J(e) \;\; \text{with} \;\; p = \nabla J(e)$$

$$J(e) = J^{**}(e) = \sup_{q} \langle q,e\rangle - J^*(q) = \langle p,e\rangle - J^*(e) \;\; \text{with} \;\; e = \nabla J^*(p)$$

The essential geometric property of BG

Recalling the centroid definition of a finite set S of N points x_i in IR^N $\quad \mu = \frac{1}{N}\sum_{i=1,N} x_i$

Centroid property for Symmetrized Bregman Gaps

For any set S of N pairs $[e,p]$ of dual points with respect to any convex function J in IR^{2n}, and for the symmetrized Bregman gap SBD_J generated by J,

$$\begin{bmatrix} e \\ p \end{bmatrix} = \frac{1}{N}\sum_{i=1,N}\begin{bmatrix} e_i \\ p_i \end{bmatrix} = \underset{\begin{bmatrix} x \\ y \end{bmatrix} \in \mathrm{IR}^n}{\arg\min}\; C_J([x,y]) \equiv \frac{1}{N}\sum_{i=1,N} SBD_J([e_i,p_i],[x,y])$$

The property extends to the probabilistic context with the expectation $E_v[f(x)] = \frac{1}{N}\sum_{i=1,N} v_i f(x_i)$

The mean or centroïd pair of a finite set S of N pairs minimizes the sum of Symmetrized Bregman Gap between this point and any other points of the set S, <u>whatever</u> the choice of the Bregman Gap.

First general geometric properties of BG

Hyperplane separation or bisectors

The set of equidistant points x with respect to a Bregman gap to a pair (e_1,e_2), also called named Bregman bisector of the first type of (e_1, e_2).is an hyperplane. Furthermore the points e_1 and e_2 lie on the different sides of this hyperplane

Affine function of x

$$D_J(x,e_1) = D_J(x,e_2) \longrightarrow \cancel{J(x)} - J(e_1) - \langle \underline{p}_1, x - e_1 \rangle = \cancel{J(x)} - J(e_2) - \langle \underline{p}_2, x - e_2 \rangle$$

Bregman projection on a closed convex sets

$$\forall C \ \forall e \ \exists \tilde{e} \ \ s.t. \ \ \tilde{e} = \arg\min_{h \in C} D_J(h,e)$$

For any closed convex subset C of IR^{n}xIR^n and point e, there exists a projection $\tilde{e}$ onto C in the sense of the Bregman divergence, the projection is unique if J is strictly convex

$$\underset{h \in C}{Min} \ D_J(h,e) = \underset{h \in C}{Min} \left[J(h) - \langle \underline{p}, h \rangle \right] - J(e) + \langle \underline{p}, e \rangle$$

J , C are convex , C is closed $\rightarrow$ $\tilde{e}$ exists, belongs to ∂C, unique if J strictly convex

First general geometric properties of BG

Hyperplane separation or bisectors

The set of equidistant points x with respect to a Bregman gap to a pair (e_1,e_2), also called named Bregman bisector of the first type of (e_1, e_2).is an hyperplane. Furthermore the points e_1 and e_2 lie on the different sides of this hyperplane

Affine function of $[e_x, p_x]$

$$SBG_J([x,p_x],[e_1,p_1]) = SBG_J([x,p_x],[e_2,p_2]) \longrightarrow \langle p_1 - p_2, x\rangle + \langle p_x, e_1 - e_2\rangle = \langle p_1, e_1\rangle - \langle p_2, e_2\rangle$$

Bregman Gap projection on a closed convex sets

$$\forall C \quad \exists[\tilde{e},\tilde{p}] \; s.t. \; [\tilde{e},\tilde{p}] = \underset{[e,p]\in C}{\arg\min} \; BG_J([\tilde{e},\tilde{p}],[e,p])$$

For any closed convex subset C of IR^{n}xIR^n and point e, and any a pair of duals quantities [e, p], there exists a projection $[\tilde{e},\tilde{p}]$ on C in the sense of the Bregman gap, the projection is unique if J is strictly convex

Bregman Gap Balls

The Bregman gap ball, associated to the generating function J, with center μ and radius r is the set

$$BGB^J_{\mu,\rho} = \{e, SBG_J(e,\mu) \le \rho\}$$

BG Balls are closed convex sets

Thanks for your attention