Some perspectives on Bregman divergences*

Frank Nielsen

Sony Computer Science Laboratories, Inc.

DO x ML @Kyoto RIMS
May 12-16th, 2025

Sony CSL

Outline of the talk

- Bregman divergences and some usages
- Boolean geometry of Bregman balls
- Bregman divergences derived from comparative convexity

Bregman divergences (1960's)

Lev M. Bregman
(1941 - 2023)
Photo: courtesy of
Alexander Fradkov

- $F: \Theta \subseteq \mathbb{R}^m \rightarrow \mathbb{R}$ a strictly convex and smooth real-valued function on a finite dim. Hilbert space <.,.>

Bregman divergence $B_F: \Theta \times \mathrm{RelInt}(\Theta) \rightarrow \mathbb{R}_{\geq 0}$

$$B_F(\theta_1 : \theta_2) = F(\theta_1) - F(\theta_2) - \langle \theta_1 - \theta_2, \nabla F(\theta_2) \rangle$$

Smooth measure of discrepancy, not a metric distance because it violates the triangle inequality, and is asymmetric when F is not quadratic function. Hence the delimiter notation ":" instead of $B_F(\theta_1, \theta_2)$

BD interpreted as **remainder** of a first order Taylor expression of $F(\theta_1)$ around θ_2:

$$F(\theta_1) = F(\theta_2) + \langle \theta_1 - \theta_2, \nabla F(\theta_2) \rangle + \underbrace{B_F(\theta_1 : \theta_2)}_{\text{Taylor remainder}}$$

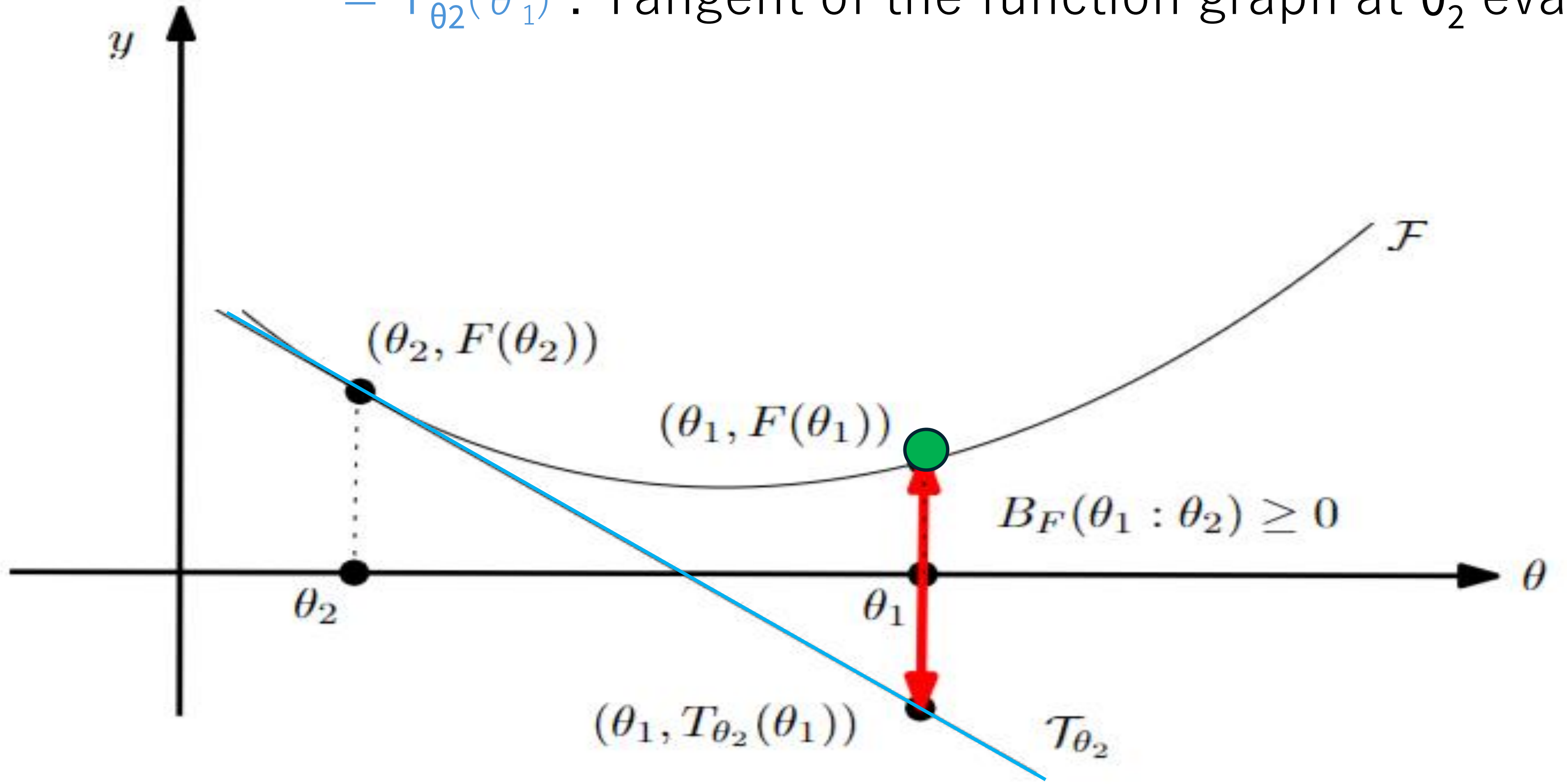
Example of remainder: Lagrange remainder (smooth C^2 generators): $\nabla^2 F$ **SPD** $\Rightarrow B_F(\theta_1 : \theta_2) \geq 0$

$$B_F(\theta_1 : \theta_2) = \tfrac{1}{2} (\theta_2 - \theta_1)^T \nabla^2 F(\theta) (\theta_2 - \theta_1) \geq 0, \ \theta \in [\theta_1 \theta_2]$$

Geometric interpretation as a **vertical gap** using the graph $(\theta, F(\theta))$:

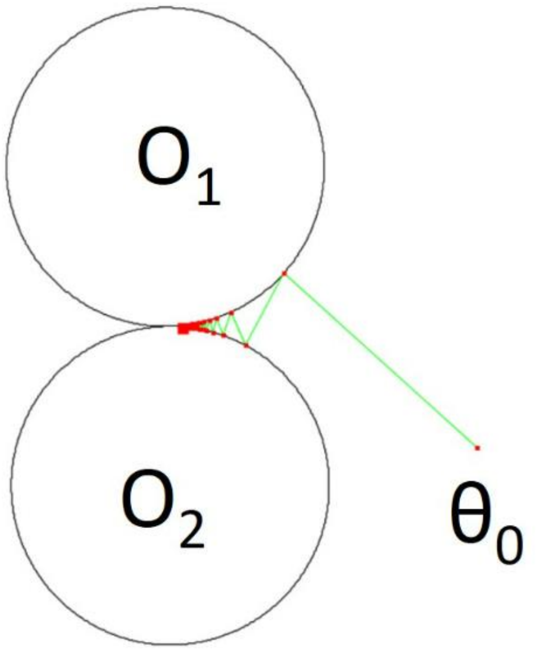
$$B_F(\theta_1 : \theta_2) = F(\theta_1) - \underbrace{(F(\theta_2) + \langle \theta_1 - \theta_2, \nabla F(\theta_2) \rangle)}$$

$= T_{\theta_2}(\theta_1)$: Tangent of the function graph at θ_2 evaluated at θ_1



BDs: Versatile and popular in OR, ML, IT, signal processing

Originally motivated for finding an intersection point in a set of convex objects using **Bregman projections**.
(ex. of convex objects: halfspaces, balls, etc.)



$\theta_0 \in \Theta, t \leftarrow 0$

$$\theta_{t+1} = \arg \min_{\theta \in O_{1+(t \mod n)}} B_F(\theta_t : \theta)$$

BDs unify:

- *squared Euclidean divergence* $F(\theta) = \frac{1}{2} \Sigma_i <\theta, \theta>$
- *Kullback-Leibler divergence* $F(\theta) = \Sigma_i \theta_i \log(\theta_i)$ (relative Shannon entropy)
- *Itakura-Saito divergence* $F(\theta) = \Sigma_i -\log(\theta_i)$ (relative Burg entropy)

$$B_F(\theta_1 : \theta_2) = F(\theta_1) - F(\theta_2) - <\theta_1 - \theta_2, \nabla F(\theta_2)>$$

L22 ($\beta=2$), KLD ($\beta \to 0$), ISD ($\beta=1$), belong to a *family* of **β-divergences**, learn ad hoc $\beta \geq 0$

x,y>0, $\beta \geq 0$

$$d_\beta(x|y) = \begin{cases} \frac{x}{y} - \log(\frac{x}{y}) - 1 & \beta = 0 \\ x(\log x - \log y) + (y - x) & \beta = 1 \\ \frac{x^\beta + (\beta-1)y^\beta - \beta x y^{\beta-1}}{\beta(\beta-1)} & \beta \in \mathbb{R}\backslash\{0,1\} \end{cases}$$

Bregman Generator: $\phi_\beta(x) = \begin{cases} -\log x + x - 1 & \beta = 0 \\ x \log x - x + 1 & \beta = 1 \\ \frac{x^\beta}{\beta(\beta-1)} - \frac{x}{\beta-1} + \frac{1}{\beta} & \text{otherwise.} \end{cases}$

Bregman divergences in machine learning···

- Kullback-Leibler divergence between two probability densities:

$$D_{KL}[p(x):q(x)] = \int p(x) \log (p(x)/q(x))\, d\mu(x)$$

is **difficult to calculate in closed form** because of the integral ∫ ...

- But Kullback-Leibler divergence between two probability densities of a **natural exponential family** with densities $p(x|\theta) \propto \exp(\langle x, \theta \rangle)$

amount to a **reverse Bregman divergence** $B_F^{rev}(\theta_1 : \theta_2) := B_F(\theta_2 : \theta_1)$

$$D_{KL}[p(x|\theta_1) : p(x|\theta_2)] = B_F^{rev}(\theta_1 : \theta_2) = B_F(\theta_2 : \theta_1)$$

Bypass the ∫, ∇F in BD easy to calculate! ⇒ Easy calculations of KLDs

Azoury, Katy S., and Manfred K. Warmuth. "Relative loss bounds for on-line density estimation with the exponential family of distributions." *Machine learning* 43 (2001)

Representational Bregman divergences (2009)

- Use a **representation function** R :

$$B_{F,R}(\lambda_1 : \lambda_2) := B_F(R(\lambda_1):R(\lambda_2))$$
$$= F(R(\lambda_1)) - F(R(\lambda_2)) - \langle R(\lambda_1) - R(\lambda_2), \nabla F(R(\lambda_2)) \rangle$$

Note that F∘R may not be a Bregman generator, i.e., not be strictly convex.

For example, consider the KLD between two densities of a **generic exponential family (natural parameter from representation function)**

$p_\lambda(x) \propto \tilde{p}_\lambda(x) = \exp(\langle \theta(\lambda), t(x) \rangle)\, h(x)$ include normal, Gamma/Beta, Wishart, Poisson, etc.

θ (λ): natural parameter corresponding to λ, representation function R(.)= θ (.)

$$D_{KL}[p(x|\lambda_1) : p(x|\lambda_2)] = B_F^{rev}(\theta(\lambda_1) : \theta(\lambda_2)) = B_F(\theta(\lambda_2) : \theta(\lambda_1))$$

NEF density $p(x|\theta) \propto \exp(\langle x, \theta \rangle)$ $D_{KL}[p(x|\theta_1) : p(x|\theta_2)] = B_F^{rev}(\theta_1 : \theta_2) = B_F(\theta_2 : \theta_1)$

Extended α -divergences are representational BDs

α -divergences extended to m-dimensional positive measures are **representational Bregman divergences**:

$$D_\alpha^+(q_1:q_2) = \begin{cases} \frac{4}{1-\alpha^2}\sum_{i=1}^m \left(\frac{1-\alpha}{2}q_1 + \frac{1+\alpha}{2}q_2 - q_1^{\frac{1-\alpha}{2}} q_2^{\frac{1+\alpha}{2}}\right), \alpha\in\mathbb{R}\backslash\{-1,1\} \\ D_{\mathrm{KL}}^{*}{}^{+}(q_1:q_2) = D_{\mathrm{KL}}^+(q_2:q_1) = \sum_{i=1}^m q_2^i \log\frac{q_2^i}{q_1^i} + q_1^i - q_2^i & \alpha = 1 \\ D_{\mathrm{KL}}^+(q_1:q_2) = \sum_{i=1}^m q_1^i \log\frac{q_1^i}{q_2^i} + q_2^i - q_1^i & \alpha = -1. \end{cases}$$

$$D_\alpha^+(q_1:q_2) = B_{F_\alpha}(R_\alpha(q_1):R_\alpha(q_2))$$

Bregman generator: $F_\alpha(r) = \sum_{i=1}^m f_\alpha(r_i), \quad f_\alpha(x) = \begin{cases} \frac{2}{1+\alpha}\left(\frac{1-\alpha}{2}x\right)^{\frac{2}{1-\alpha}}, & \alpha\neq 1 \\ \log x, & \alpha = 1. \end{cases}$

Representation function: $R_\alpha(q) = (r_\alpha(q_1),\ldots,r_\alpha(q_m)), \quad r_\alpha(x) = \frac{2}{1-\alpha}x^{\frac{1-\alpha}{2}}$

Bregman divergence: $B_F(\theta_1:\theta_2)=F(\theta_1)-F(\theta_2)-<\theta_1-\theta_2, \nabla F(\theta_2)>$

"The dual Voronoi diagrams with respect to representational Bregman divergences." IEEE ISVD *2009*

Convex duality via Legendre-Fenchel transform

- Legendre-Fenchel transform of a convex function F:

  $$F^*(\eta)=\sup_{\theta\in\Theta}\{<\theta,\eta>-F(\theta)\}$$

- Problem: some *tricky functions* with gradient map ∇F domain not convex…

  Example: $h(\xi_1,\xi_2)=[(\xi_1^2/\xi_2)+\xi_1^2+\xi_2^2]/4$ on upper plane domain $\Xi=(\xi_1,\xi_2)$

- Thus, we consider "**nice convex functions**" = **Legendre-type functions** $(\Theta,F(\theta))$

  (i) Θ open, and (ii) $\lim_{\theta\to\partial\Theta}\|\nabla F(\theta)\|=\infty$

Then we get:

❶ **reciprocal gradient maps** $\eta=\nabla F(\theta)$ and $\theta=\nabla F^*(\eta)$, $\nabla F^*=(\nabla F)^{-1}$

❷ conjugation yields $(H,F^*(\eta))$ of Legendre type

❸ biconjugation is an **involution**: $(H,F^*(\eta))^*=(H^*=\Theta,F^{**}=F(\theta))$

- Convex conjugate: $F^*(\eta)=<\nabla F^{-1}(\eta),\eta>-F(\nabla F^{-1}(\eta))$ since $\eta=\nabla F(\theta)$

Fenchel-Young divergences & convex duality

- Young inequality: $F(\theta_1)+F^*(\eta_2) \geq \langle \theta_1, \eta_2 \rangle$ with equality when $\eta_2 = \nabla F(\theta_1)$
- Build the Fenchel-Young divergence from the inequality: lhs-rhs ≥0

$$\mathbf{Y_{F,F^*}(\theta_1, \eta_2) = F(\theta_1)+F^*(\eta_2) - \langle \theta_1, \eta_2 \rangle} \geq 0$$

- **Mixed parameterizations** θ and η : $B_F(\theta_1 : \theta_2) = Y_{F,F^*}(\theta_1, \eta_2)$
- Duality: $\mathbf{B_F(\theta_1 : \theta_2) = Y_{F,F^*}(\theta_1, \eta_2) = Y_{F^*,F}(\eta_2, \theta_1) = B_{F^*}(\eta_2, \eta_1)}$
- Dual BDs + Dual FYs from involution $F^{**}=F$
- Note : $B_F(\theta_1 : \theta_2)=0 \Leftrightarrow \theta_1 = \theta_2 \Leftrightarrow \eta_1 = \eta_2$ i.e., $\nabla F(\theta_1) = \nabla F(\theta_2)$

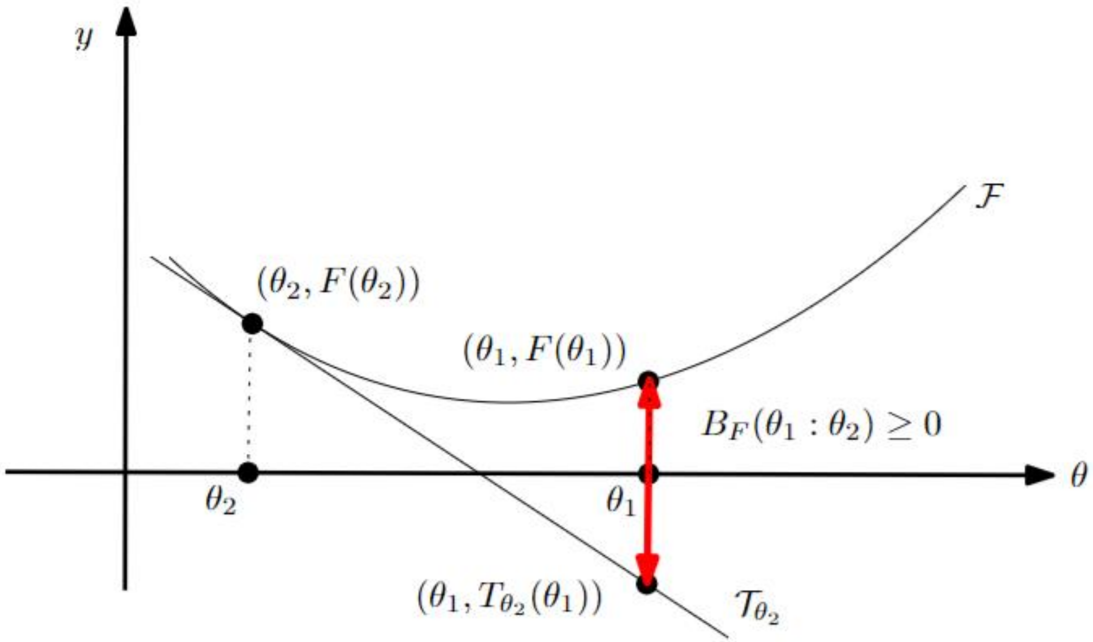
(FY initially called Legendre-Fenchel divergences…)

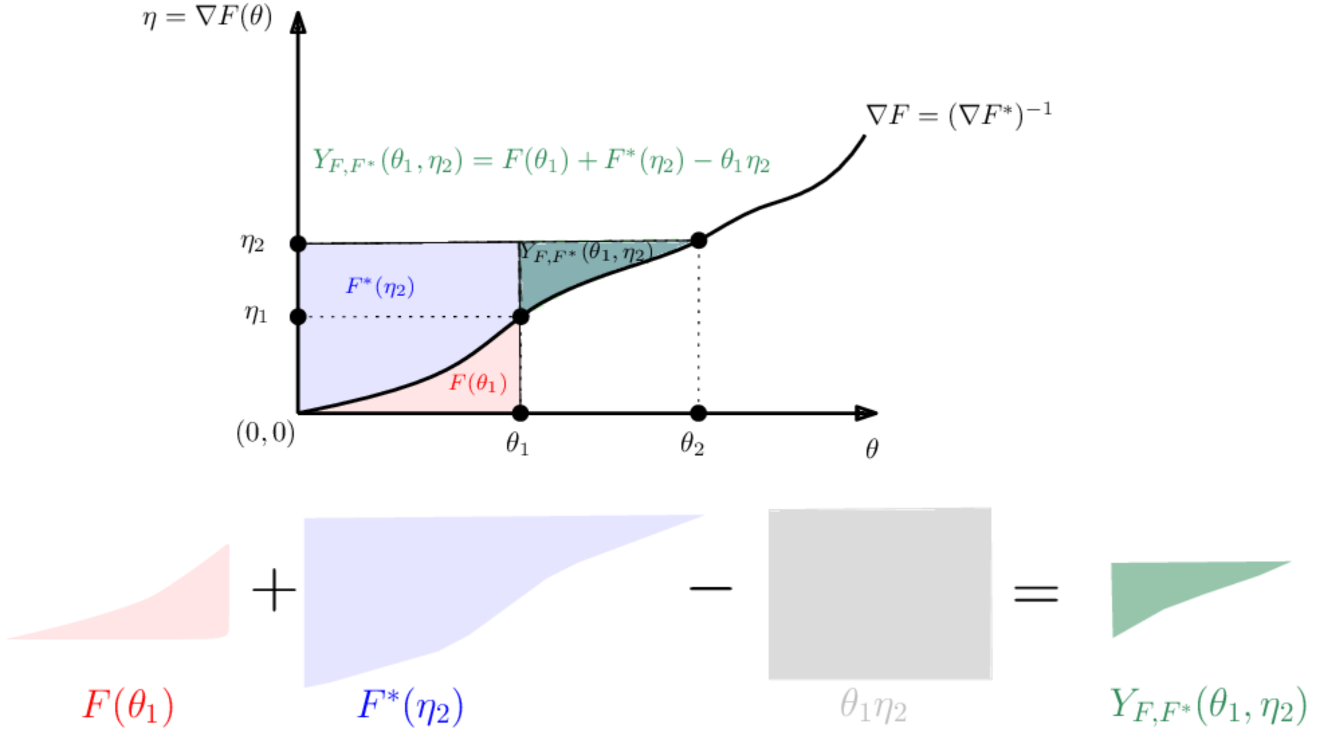
Bregman divergence vs Fenchel-Young divergence

Same parameterization $B_F(\theta_1 : \theta_2) = Y_{F,F^*}(\theta_1, \eta_2)$ **mixed parameterization**

F strictly convex and differentiable

F' ↗ strictly increasing





$B_F(\theta_1 : \theta_2) = F(\theta_1) - F(\theta_2) - \langle \theta_1 - \theta_2, \nabla F(\theta_2) \rangle$

$Y_{F,F^*}(\theta_1, \eta_2) = F(\theta_1) + F^*(\eta_2) - \langle \theta_1, \eta_2 \rangle$

Kullback-Leibler divergence between non-normalized exponential family densities

- Kullback-Leibler divergence between two **positive measures**:

$$D_{KL}^{+}[p_1(x):p_2(x)] = \int \{ p_1(x) \log (p_1(x)/p_2(x)) + p_2(x) - p_1(x) \} \, d\mu(x)$$

- Exponential family density:
  - Normalized: $p(x|\theta) = \exp(<x, \theta> - F(\theta)) \, d\mu(x)$
  - Non-normalized: $q(x|\theta) = \exp(<x, \theta>) \, d\mu(x)$
- Hence, $p(x|\theta) = q(x|\theta)/Z(\theta)$ with **partition function** $Z(\theta)=\exp(F(\theta))$ and **cumulant function** $F(\theta)=\log Z(\theta)$
- When F is convex, Z=exp(F) is log-convex
- log-convex functions are convex functions: So **both F and Z are convex functions**

- KLD between normalized densities = reverse Bregman wrt F:

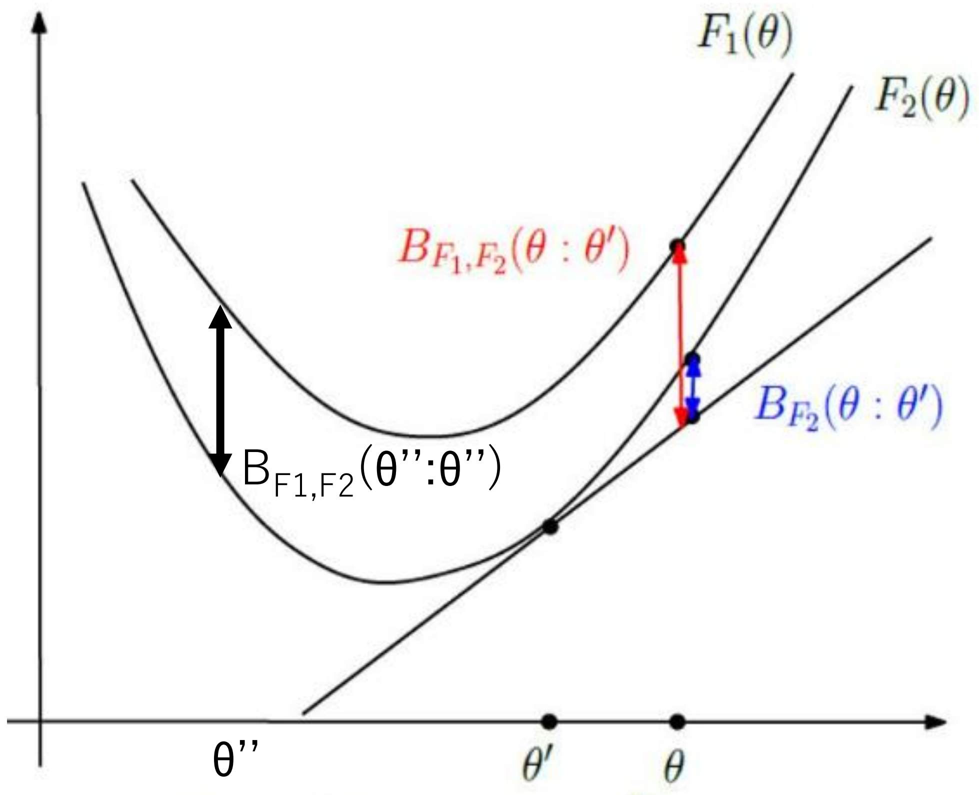
$$D_{KL}[p_{\theta_1}(x):p_{\theta_2}(x)] = B_F^{*}[\theta_1 : \theta_2] = B_F[\theta_2 : \theta_1]$$

- KLD between non-normalized densities = reverse Bregman wrt Z:

$$D_{KL}^{+}[q_{\theta_1}(x):q_{\theta_2}(x)] = B_Z^{*}[\theta_1 : \theta_2] = B_Z[\theta_2 : \theta_1]$$

2312.12849

Duo Bregman divergences: Generalize BDs with a pair of generators



One generator **majorizes** the other one:

$$F_1(\theta) \geq F_2(\theta)$$

Then

$$B_{F_1,F_2}(\theta:\theta') = F_1(\theta) - F_2(\theta') - (\theta-\theta')^\top \nabla F_2(\theta')$$

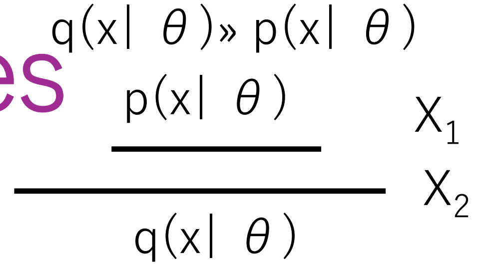
$$\geq B_{F_2}(\theta:\theta')$$

- Recover Bregman divergence when $F_1(\theta) = F_2(\theta) = F(\theta)$

$$B_F(\theta_1:\theta_2) = F(\theta_1) - F(\theta_2) - \langle \theta_1 - \theta_2, \nabla F(\theta_2) \rangle$$

- Only **pseudo-divergence** because $B_{F1,F2}(\theta'':\theta'')$ positive, not zero

KLD between nested exponential families amount to duo Bregman pseudo-divergences



- Consider an exponential family on support X_1:

$D_{KL}[p(x):q(x)]=\int p(x) \log (p(x)/q(x))\, d\mu(x)$

$0 \log(0/0)=0$

$$p(x|\theta) = \exp(\langle x, \theta\rangle - F_1(\theta))\, d\mu(x)$$

with cumulant function $F_1(\theta)=\log \int_{X1} \exp(\langle x, \theta\rangle)\, d\mu(x)$

- Another exponential family with **nested supports: $X_1 \subseteq X_2$**

$$q(x|\theta) = \exp(\langle x, \theta\rangle - F_2(\theta))\, d\mu(x)$$

is an exponential family with $F_2(\theta)=\log \int_{X2} \exp(\langle x, \theta\rangle)\, d\mu(x) \geq F_1(\theta)$

- Then KLD amounts to a **reverse duo Bregman pseudo-divergence**:

$$D_{KL}[p(x|\theta_1) : q(x|\theta_2)]= B_{F2,F1}^{rev}(\theta_1 : \theta_2)=B_{F2,F1}(\theta_2 : \theta_1)$$

"Statistical divergences between densities of truncated exponential families with nested supports: Duo Bregman and duo Jensen divergences." *Entropy* 24.3 (2022)

Curved Bregman divergences

Consider a domain U which maps to a subset of Θ by θ =c(u)
with dim(U)<dim(Θ):

$B_{F,u}(u_1 : u_2) := B_F(c(u_1) : c(u_2))$ is not Bregman when {c(u) | u ∊ U} not convex

usually not a Bregman divergence unless c(.) is affine

Example: Symmetrized Bregman divergences (Jeffreys-Bregman div.)
are curved Bregman divergences: $S_F(\theta_1, \theta_2) = \langle \theta_1 - \theta_2, \eta_1 - \eta_2 \rangle$

$$\begin{aligned} S_F(\theta_1 : \theta_2) &= B_F(\theta_1 : \theta_2) + B_F(\theta_2 : \theta_1), \\ &= B_F(\theta_1 : \theta_2) + B_{F^*}(\nabla F(\theta_1) : \nabla F(\theta_2)) \\ &= \breve{B}_{F_\xi}(\xi(\theta_1) : \xi(\theta_2)), \end{aligned}$$

$$F^*(\eta) = \langle \theta, \eta \rangle - F(\theta) \qquad F_\xi(\theta, \eta) := F(\theta) + F^*(\eta) \qquad \xi(\theta) = (\theta, \nabla F(\theta))$$

$$\mathcal{U} = \{(\theta, \nabla F(\theta)) : \theta \in \Theta\}$$

m-dimensional submanifold in 2m-dimensional space

Curved Bregman centroid is the Bregman projection of the full Bregman centroid

Theorem:

$$\arg\min_{u\in\mathcal{U}} \sum_{i=1}^{n} w_i\, B_F(\theta_i : \theta(u)) = \arg\min_{u\in\mathcal{U}} B_F(\bar\theta : \theta(u))$$

[Bregman projection]

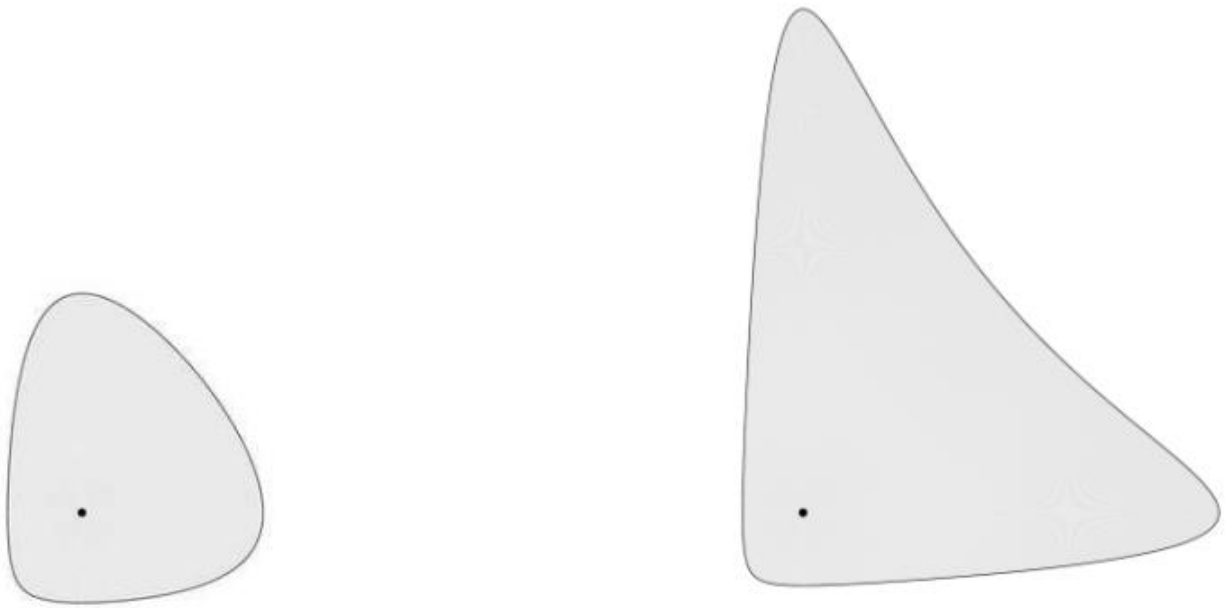
$$\theta_i = \theta(u_i) \qquad \bar\theta = \sum_i w_i\theta_i$$

Proof.

$$\begin{aligned}
\min_{u\in\mathcal{U}} \sum_{i=1}^{n} w_i\, B_F(\theta_i : \theta(u)) &= \sum_{i=1}^{n} w_i(F(\theta_i) - F(\theta(u)) - \langle \theta_i - \theta(u), \nabla F(\theta(u))\rangle),\\
&\equiv -F(\theta(u)) - \langle \bar\theta - \theta(u), \nabla F(\theta(u))\rangle,\\
&\equiv F(\bar\theta) - F(\theta(u)) - \langle \bar\theta - \theta(u), \nabla F(\theta(u))\rangle\\
&= B_F(\bar\theta : \theta(u)).
\end{aligned}$$

"What is... an information projection?" Notices of the AMS 65.3 (2018): 321-324.

Space of Bregman balls



Example:
Itakura-Saito right and left spheres

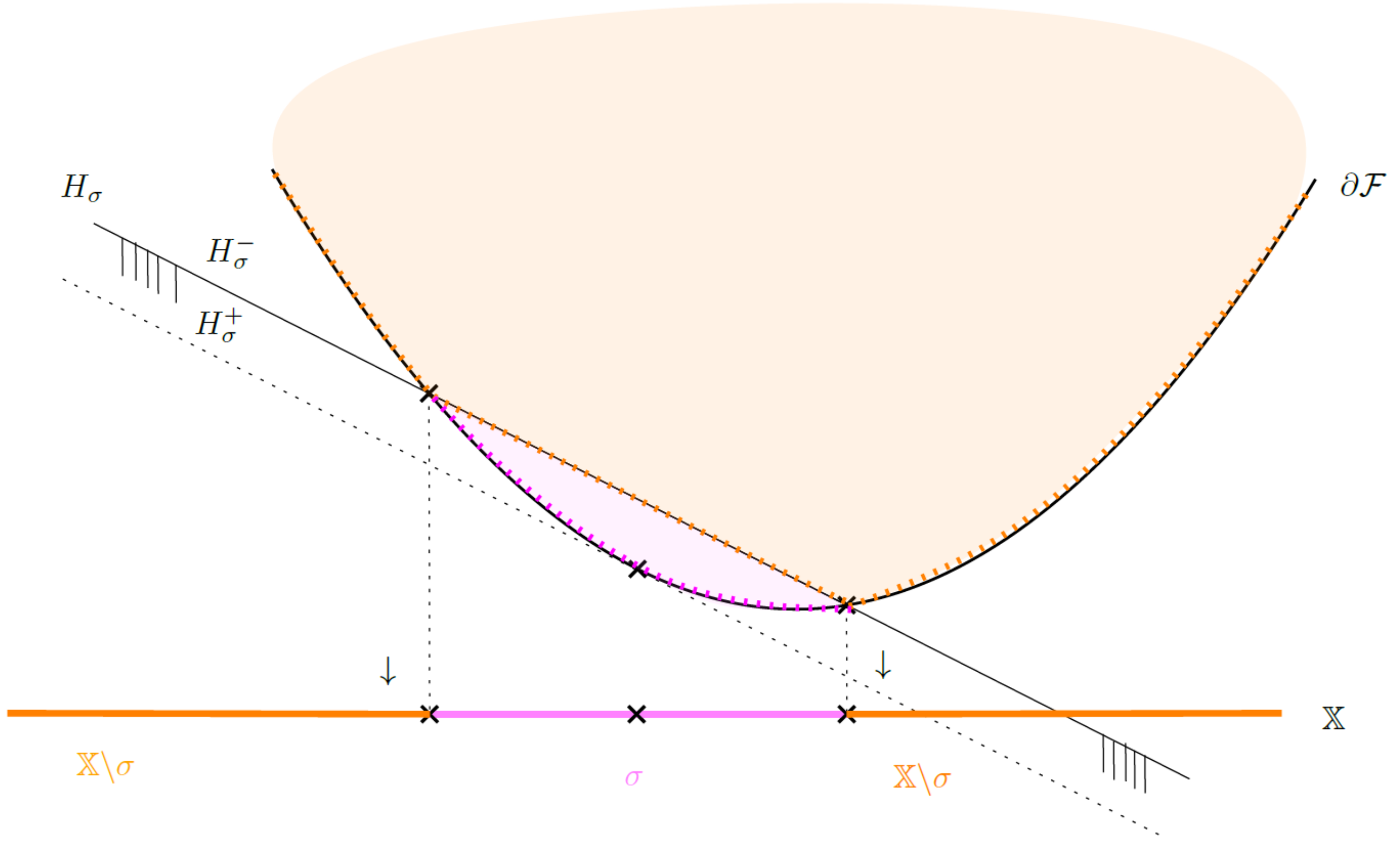
Right-sided Bregman ball: $\sigma_F(\theta, r) = \{\theta' \in \Theta : B_F(\theta' : \theta) \leq r\}$

Left-sided Bregman ball: $\sigma_F^{\star}(\theta, r) = \{\theta' \in \Theta : B_F(\theta : \theta') \leq r\}$

Application: Boolean algebra of unions & intersections of Bregman balls

Right Bregman ball and its complement

$$\mathcal{F} := \{(\theta, y \geq F(\theta)) \ : \ \theta \in \Theta \subset \mathbb{R}^m\} \subset \mathbb{R}^{m+1}$$



$\sigma^c = \mathbb{X}\backslash\sigma = \downarrow (H_\sigma^- \cap \partial\mathcal{F})$ $\sigma = \downarrow (H_\sigma^+ \cap \partial\mathcal{F})$ $\sigma^c = \mathbb{X}\backslash\sigma = \downarrow (H_\sigma^- \cap \partial\mathcal{F})$

$\mathbb{R}$

$\downarrow$ means vertical projection

S^c: complement of set S

To any sphere, associate an hyperplane:

$$H_{\theta,r} : y = \langle \theta' - \theta, \nabla F(\theta) \rangle + F(\theta) + r$$

Reciprocally, to an hyperplane cutting the function graph, associate a sphere

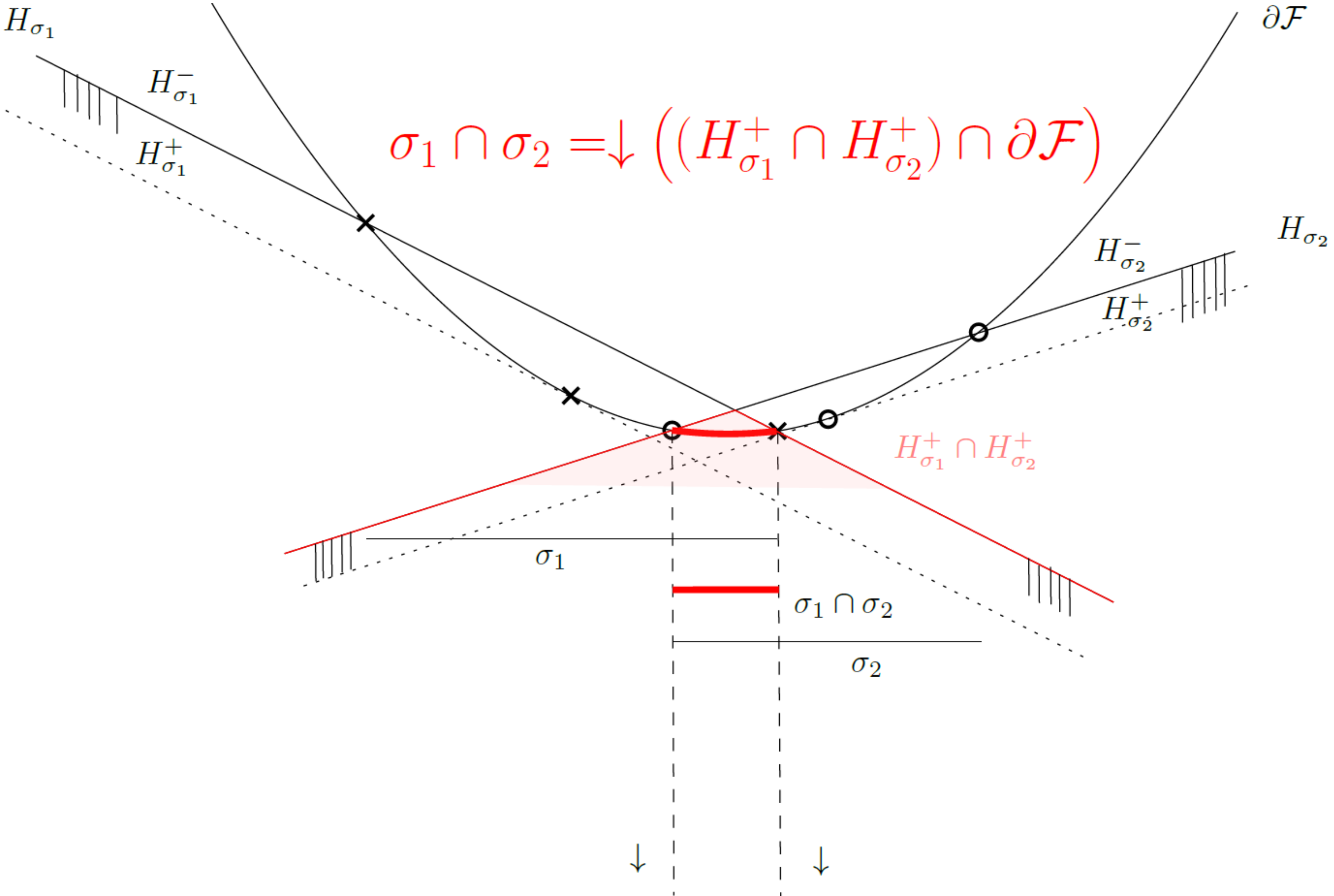
$$z = \langle \mathbf{x}, \mathbf{a} \rangle + b$$

Center: $\mathbf{c} = \boldsymbol{\nabla}^{-1} F(\mathbf{a})$

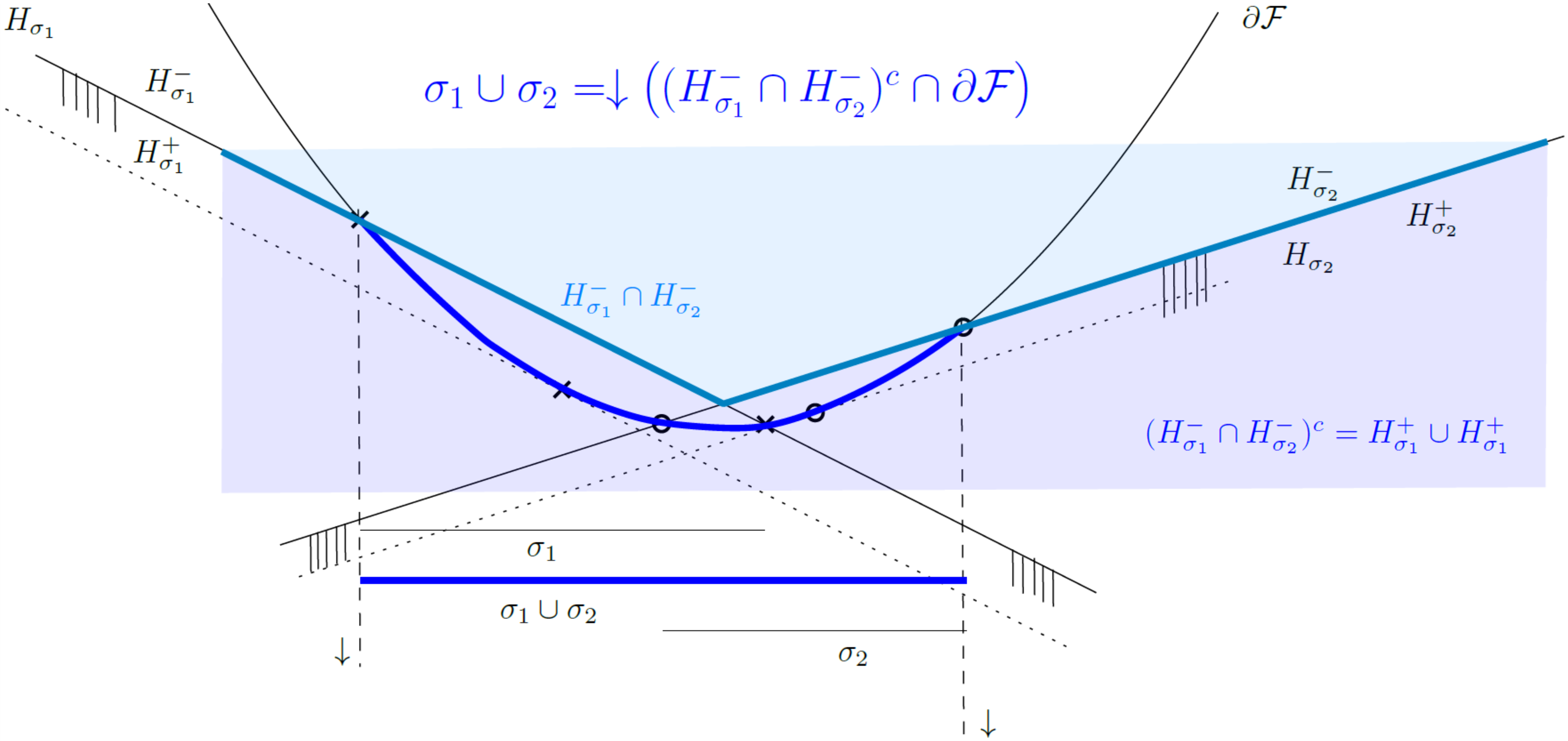
Radius: $\langle \mathbf{a}, \mathbf{c} \rangle - F(\mathbf{c}) + b$

Lifting to potential Bregman generator graph

Intersection of two right Bregman balls



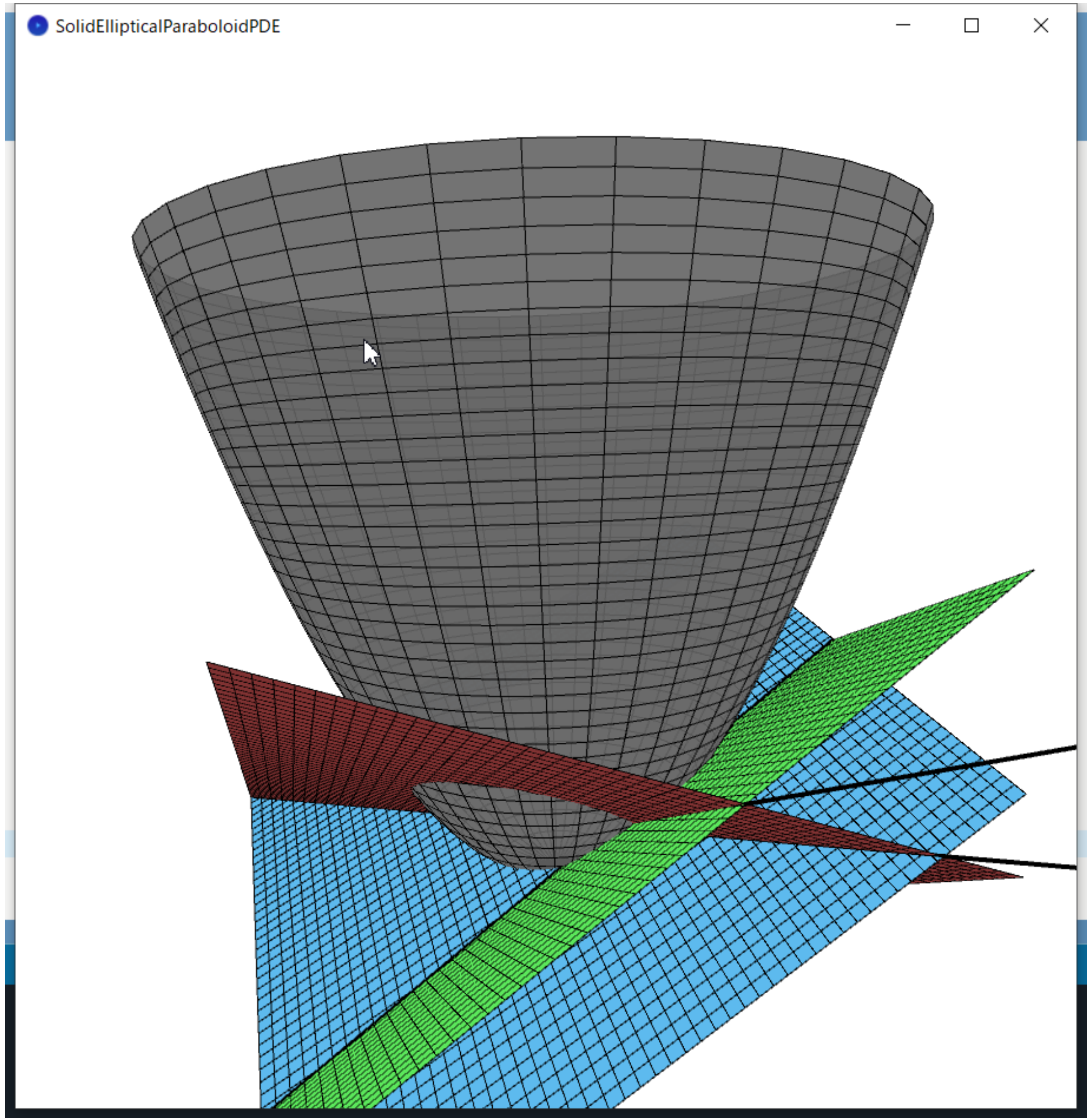
Union of two right Bregman balls



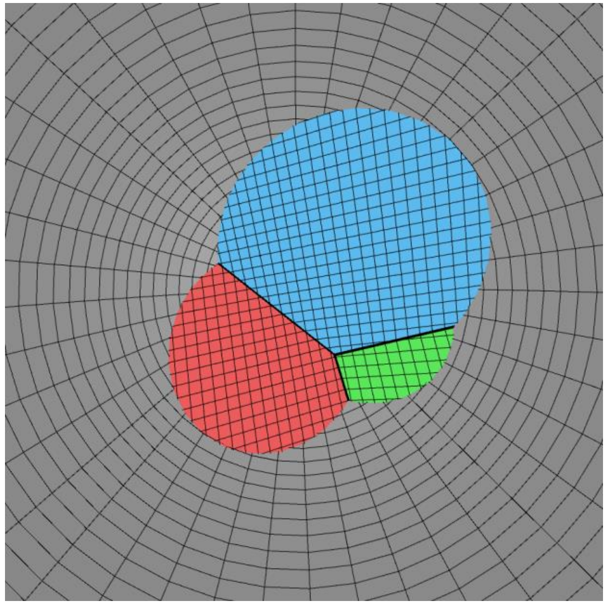
Set Morgan's law: $(A \cup B)^c = A^c \cap B^c$

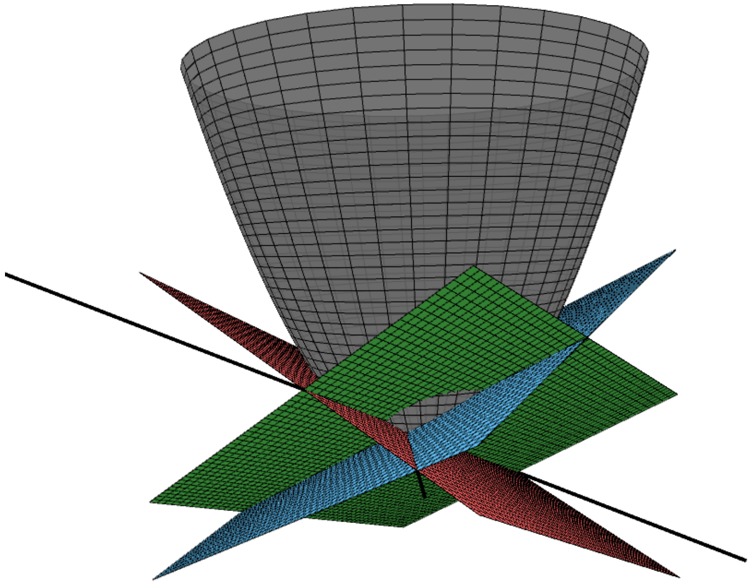
Complement of halfspace $(H^+)^c = H^-$

Example: Euclidean spheres potential function: Paraboloid, L22



Top view displays the union of disks





$$B_F(\theta_1 : \theta_2) = F(\theta_1) - F(\theta_2) - < \theta_1 - \theta_2, \nabla F(\theta_2) >$$

Generalization of Bregman divergences using comparative convexity

Comparative convexity: (M,N)-convexity

Ordinary convexity of a function: $f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$

for all t in [0,1]

- *Definition*: A function Z is **(M,N)-convex** iff for in α in [0,1]:

$$Z(M(x, y; \alpha, 1-\alpha)) \leq N(Z(x), Z(y); \alpha, 1-\alpha)$$

- Ordinary convexity = (A,A)-convexity wrt to arithmetic weighted mean

$A(x, y; \alpha, 1-\alpha) = \alpha x + (1-\alpha)y$ $\quad$ $f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2)$

for all t in [0,1]

- **Log-convexity**: **(A,G)-convexity** wrt to A/Geometric weighted means:

$G(x, y; \alpha, 1-\alpha) = x^{\alpha} y^{1-\alpha}$ $\quad$ $f(tx_1 + (1-t)x_2) \leq f(x_1)^t f(x_2)^{1-t}$

for all t in [0,1]

Since G ≤ A, (A,G)-functions are (A,A)-convex: Log-convex functions are convex

Comparative convexity wrt quasi-arithmetic means

- **quasi-arithmetic mean** for a strictly monotone generator h(u):

$$M_h(x,y;\alpha,1-\alpha) = h^{-1}(\alpha h(x) + (1-\alpha) h(x)).$$

- Includes **power means** which are *homogeneous means*:

$$M_p(x,y;\alpha,1-\alpha) = (\alpha x^p + (1-\alpha) y^p)^{\frac{1}{p}} = M_{h_p}(x,y;\alpha,1-\alpha), \quad p \neq 0$$

$$h_p(u) = \frac{u^p - 1}{p} \qquad h_p^{-1}(u) = (1+up)^{\frac{1}{p}}$$

Include the **geometric mean** in the limit case p→0

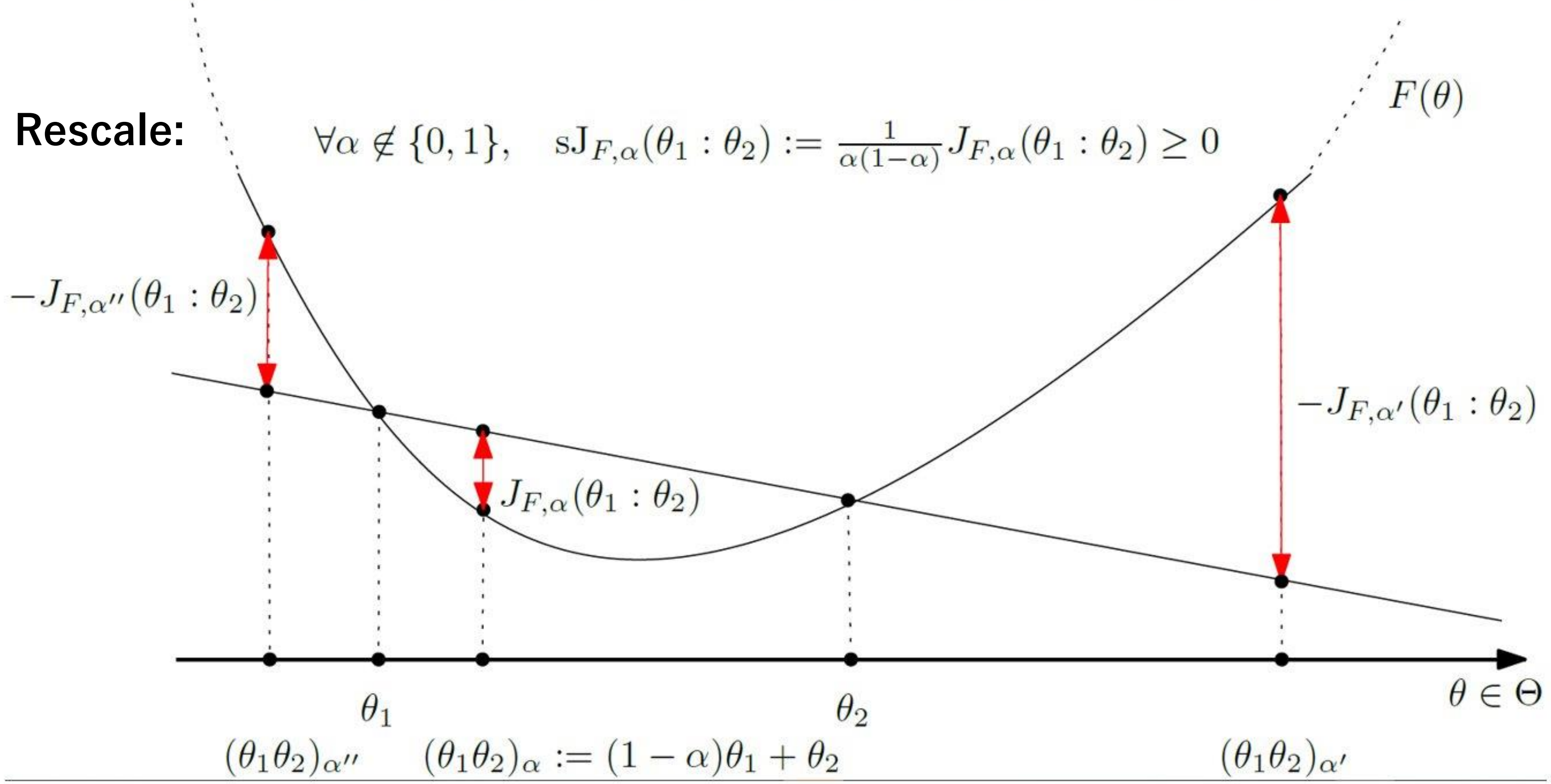
Checking the comparative convexity wrt two quasi-arithmetic means via an ordinary convexity test:

Proposition 6 ([1, 34]). *A function $Z(\theta)$ is strictly (M_ρ, M_τ)-convex with respect to two strictly increasing smooth functions ρ and τ if and only if the function $F = \tau \circ Z \circ \rho^{-1}$ is strictly convex.*

Scaled skewed Jensen divergences & Bregman divergences

$$\forall \alpha \in (0,1), \quad J_{F,\alpha}(\theta_1 : \theta_2) := (1-\alpha)F(\theta_1) + \alpha F(\theta_2) - F((1-\alpha)\theta_1 + \alpha\theta_2)$$

Rescale: $\forall \alpha \notin \{0,1\}, \quad \mathrm{sJ}_{F,\alpha}(\theta_1 : \theta_2) := \frac{1}{\alpha(1-\alpha)} J_{F,\alpha}(\theta_1 : \theta_2) \geq 0$



Jensen divergences

measures the vertical gap induced by a strictly convex function

$\lim_{\alpha \to 0} \mathrm{sJ}_{F,\alpha}(\theta_1 : \theta_2) = B_F(\theta_1 : \theta_2)$ (Bregman divergence) $\lim_{\alpha \to 1} \mathrm{sJ}_{F,\alpha}(\theta_1 : \theta_2) = B_F(\theta_2 : \theta_1)$ (reverse BD)

The Burbea-Rao and Bhattacharyya centroids, IEEE Transactions on Information Theory 57.8 (2011)

Generalizing Bregman divergences with (M,N)-convexity: (M,N)-Bregman divergences

- First, define skew Jensen divergence from (M,N)-comp. convexity:

Definition:

$$J_{F,\alpha}^{M,N}(p:q) = N_\alpha(F(p),F(q)) - F(M_\alpha(p,q)).$$

Non-negative for **(M,N)-convex generators** F, provided **regular means** M and N (e.g. all power means)

Definition 5 (Bregman Comparative Convexity Divergence, BCCD) *The Bregman Comparative Convexity Divergence (BCCD) is defined for a strictly (M,N)-convex function $F : I \to \mathbb{R}$ by*

$$B_F^{M,N}(p:q) = \lim_{\alpha\to 1^-} \frac{1}{\alpha(1-\alpha)} J_{F,\alpha}^{M,N}(p:q) = \lim_{\alpha\to 1^-} \frac{1}{\alpha(1-\alpha)} (N_\alpha(F(p),F(q))) - F(M_\alpha(p,q))) \quad (31)$$

This definition is by analogy to limit of scaled skewed Jensen divergences amount to forward/reverse Bregman divergences.

Generalizing Bregman divergences with quasi-arithmetic mean convexity

Theorem 1 (Quasi-arithmetic Bregman divergences, QABD) *Let $F : I \subset \mathbb{R} \to \mathbb{R}$ be a real-valued (M_ρ, M_τ)-convex function defined on an interval I for two strictly monotone and differentiable functions ρ and τ. The quasi-arithmetic Bregman divergence (QABD) induced by the comparative convexity is:*

$$\boxed{B_F^{\rho,\tau}(p:q) = \frac{\tau(F(p)) - \tau(F(q))}{\tau'(F(q))} - \frac{\rho(p)-\rho(q)}{\rho'(q)} F'(q).} \qquad (45)$$

Amounts to a **conformal representational Bregman divergence** :

$$\boxed{B_F^{\rho,\tau}(p:q) = \frac{1}{\tau'(F(q))} B_G(\rho(p):\rho(q))}$$

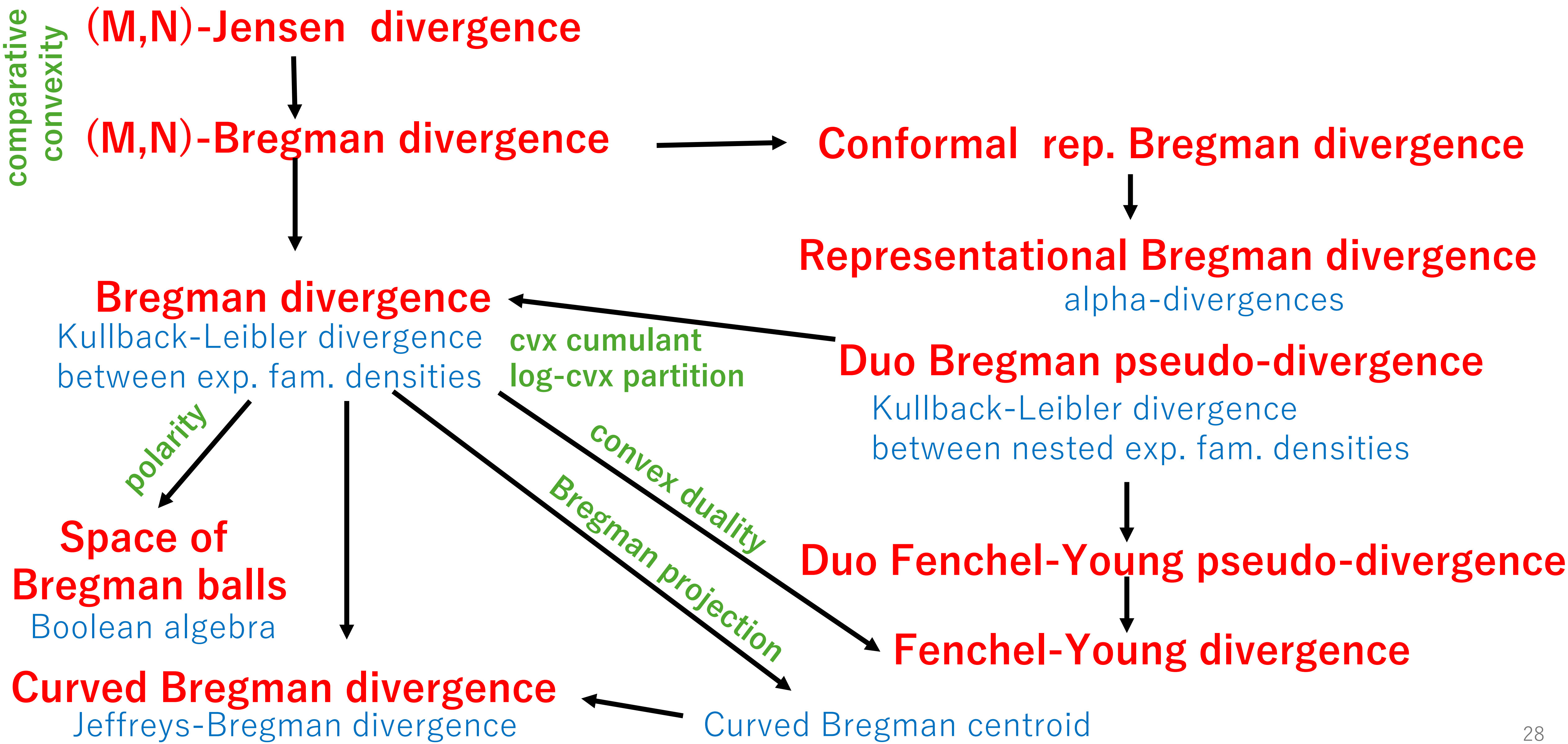
With convex generator:
$G(x) = \tau(F(\rho^{-1}(x)))$

Conformal factor

Remark: Conformal Bregman divergences may yield **robustness** in applications

Shape retrieval using hierarchical total Bregman soft clustering, IEEE Transactions on pattern analysis and machine intelligence (2012)

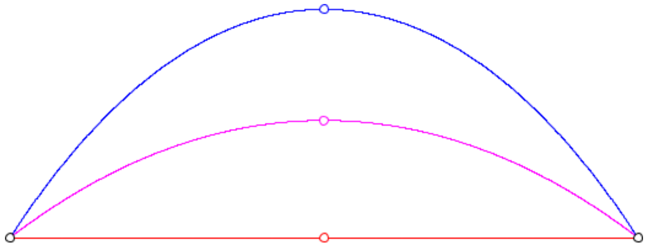
Summary map of this talk

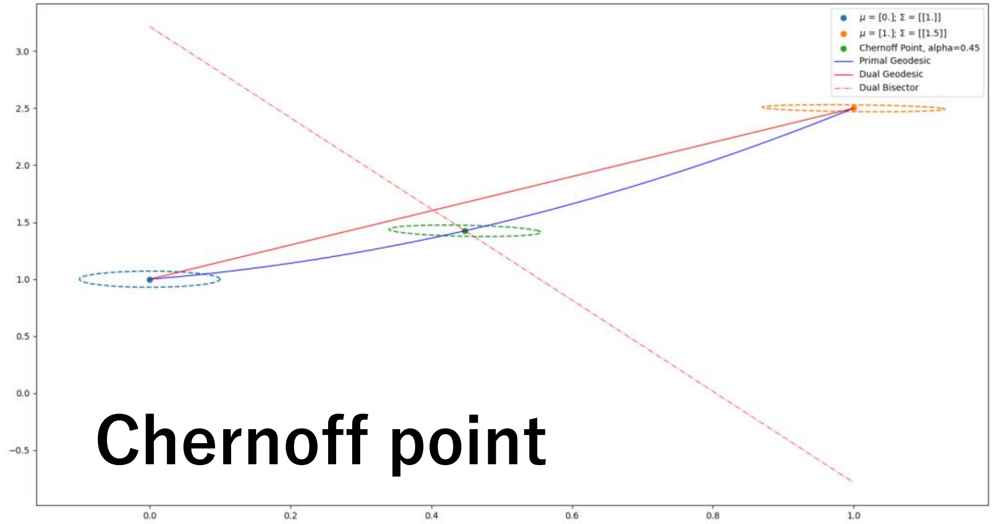


A Python library for geometric computing on Bregman Manifolds

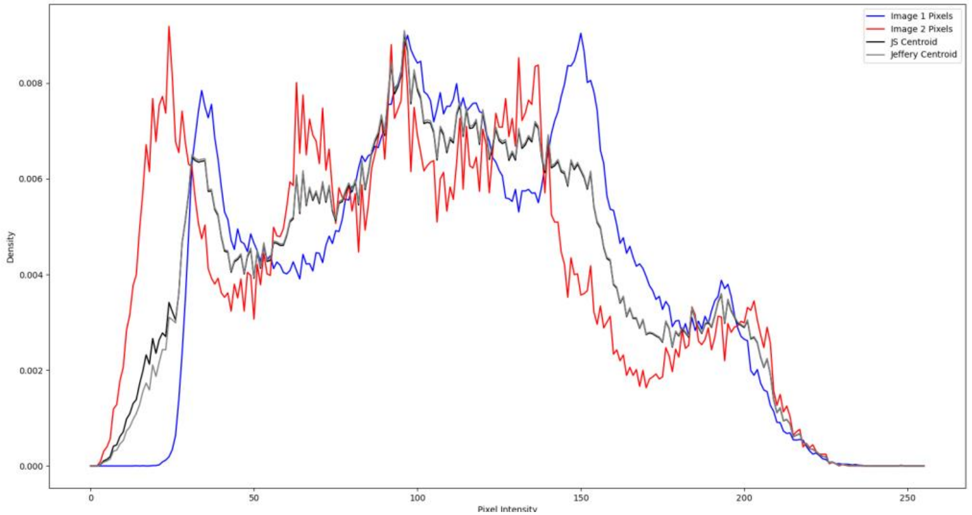
pyBregMan

https://franknielsen.github.io/pyBregMan/

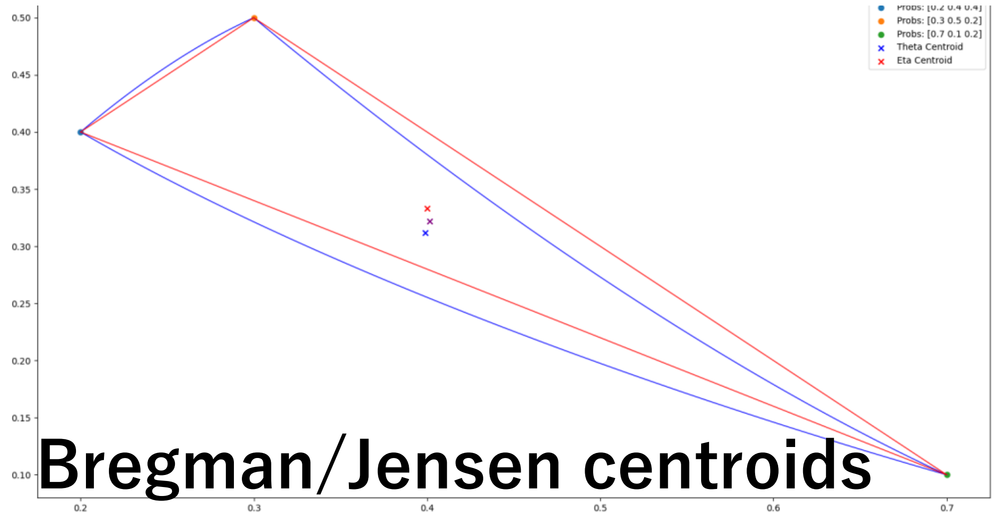




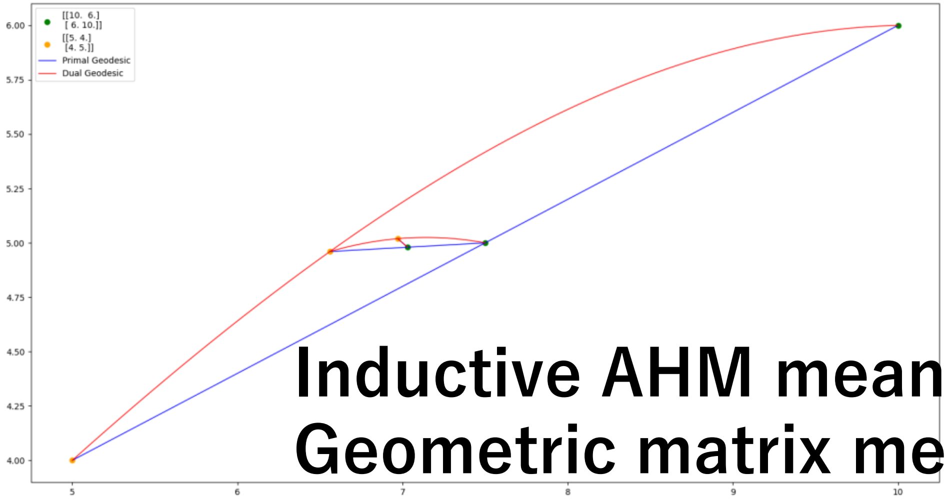
Chernoff point



Jensen-Shannon centroid



Bregman/Jensen centroids



Inductive AHM mean
Geometric matrix mean

Joint work of Frank Nielsen and Alexander Soen

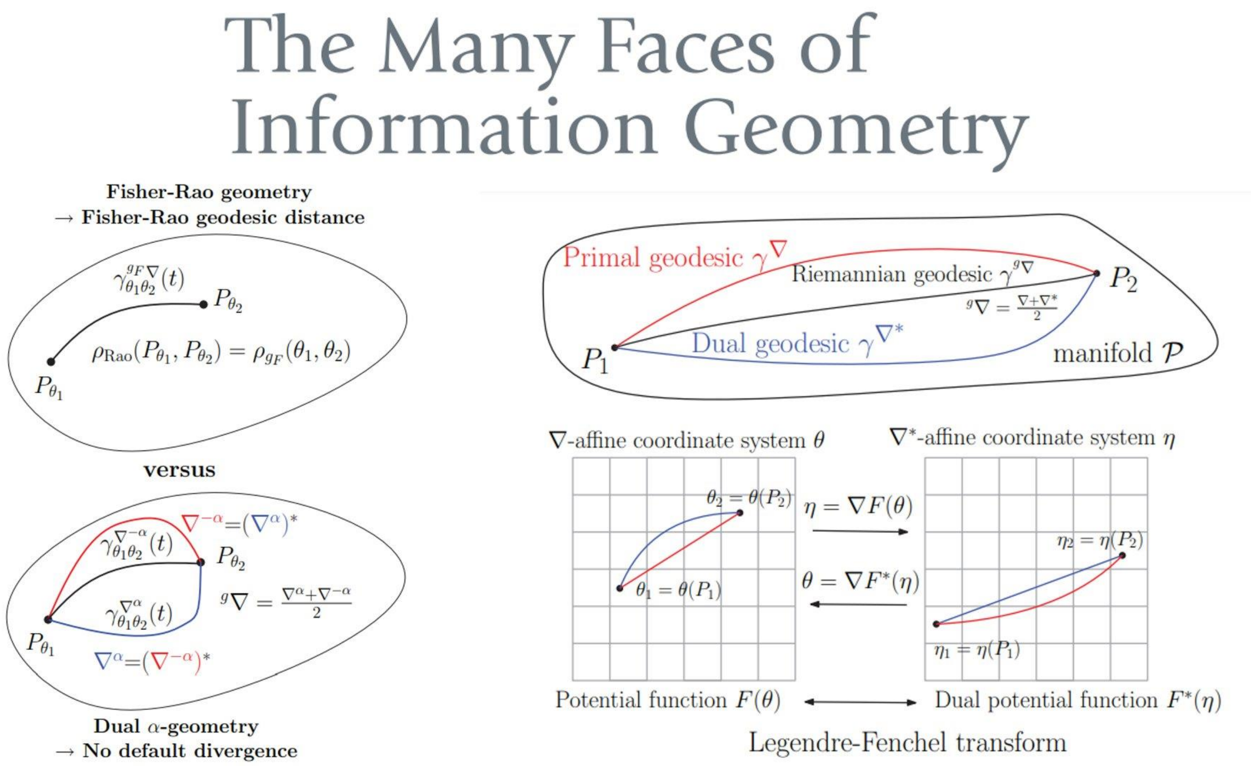
Thank you!

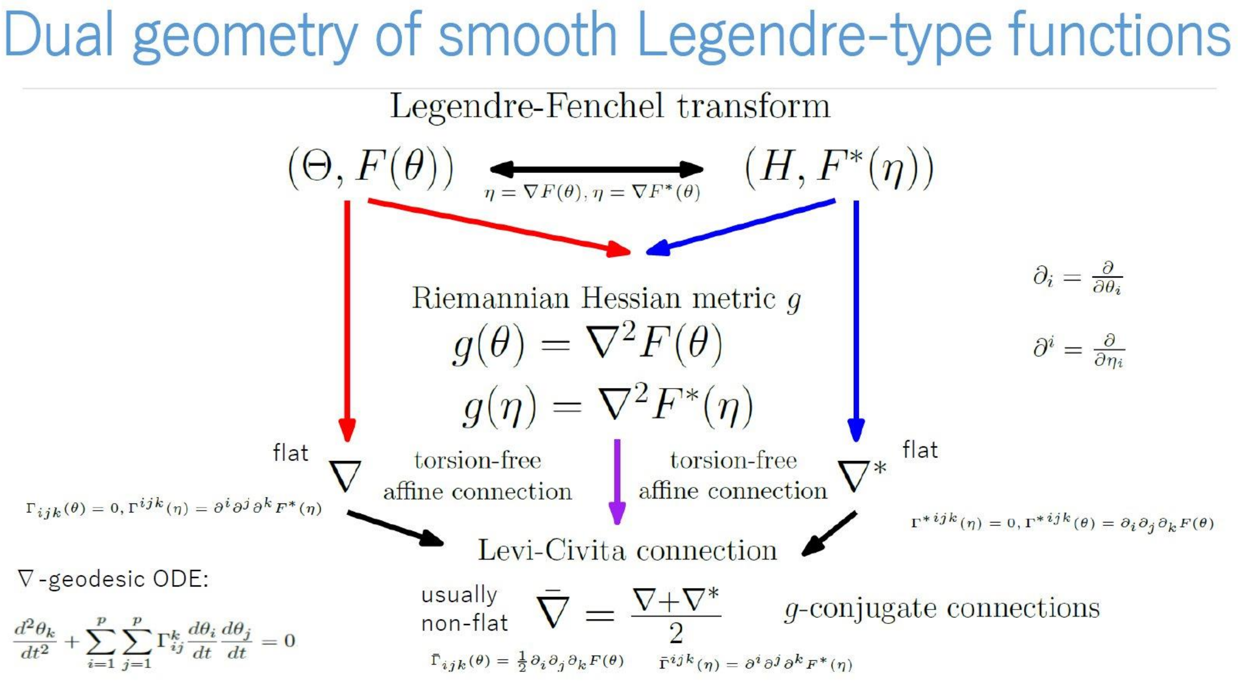
https://franknielsen.github.io/

Many thanks to all my inspiring collaborators.

In particular, special thanks to Richard Nock, Ke Sun, Ehsan Amid, and Alexander Soen

+ invitation to the information geometry of BDs





The many faces of information geometry. Not. Am. Math. Soc, 69(1), pp.36-45.

Some references

- NF and Richard Nock. "**The dual Voronoi diagrams with respect to representational Bregman divergences**." *Sixth International Symposium on Voronoi Diagrams*. IEEE, 2009.

- Boissonnat, Jean-Daniel, FN, and Richard Nock. "**Bregman Voronoi diagrams**." *Discrete & Computational Geometry* 44 (2010): 281-307.

- NF. "**Statistical divergences between densities of truncated exponential families with nested supports: Duo Bregman and duo Jensen divergences**." *Entropy* 24.3 (2022

- NF and Richard Nock. "**Generalizing skew Jensen divergences and Bregman divergences with comparative convexity**." *IEEE Signal Processing Letters* 24.8 (2017)

- NF. "**Curved representational Bregman divergences and their applications**." *arXiv preprint arXiv:2504.05654*