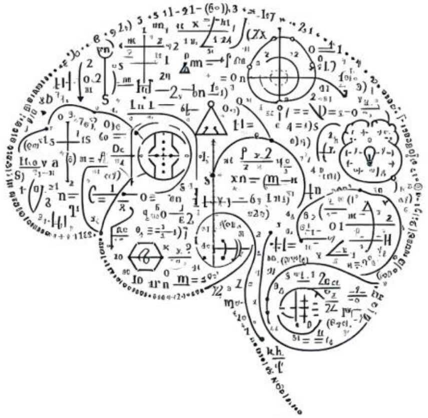
Some generalizations and perspectives on Bregman divergences

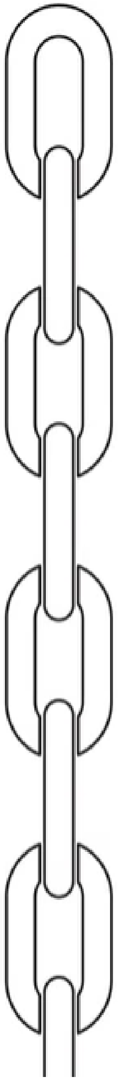
Frank Nielsen

Sony Computer Science Laboratories, Inc.

MML 2025 @ TUHH, Hamburg
September 24th, 2025



Outline: A thread of short stories



1. Quick introduction to **Bregman divergences**
2. **Duo** Bregman pseudo-divergences
3. **Curved** Bregman divergences
4. **Generalized Legendre transforms** and information geometry
5. **Generalized convexity** and Bregman divergences
6. Boolean algebra of **Bregman balls**

Bregman divergences (1960's): 1,2,3

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

① $F: \Theta \subseteq \mathbb{R}^m \rightarrow \mathbb{R}$ **strictly convex and smooth**

Bregman divergence $B_F : \Theta \times \text{Relative Interior}(\Theta) \rightarrow \mathbb{R}_{\geq 0}$

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

Smooth measure of discrepancy, not a metric distance because it violates the triangle inequality and is asymmetric when F is not quadratic function.

② BD = **remainder of a first order Taylor expression** of $F(\theta_1)$ around θ_2:

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

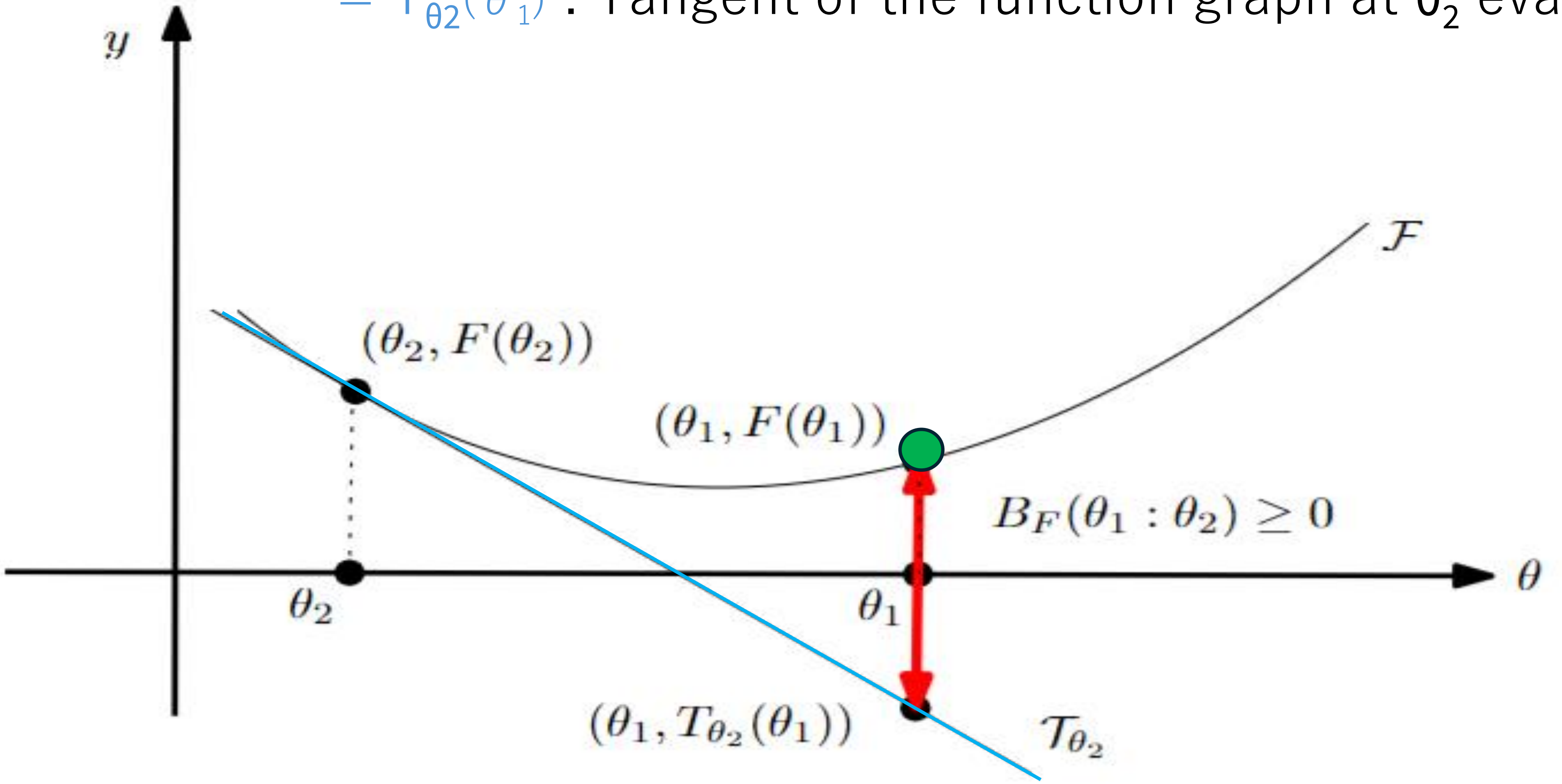
③ Example: **Lagrange remainder** (smooth C^2 generators): $\nabla^2 F$ **SPD** $\Rightarrow B_F(\theta_1 : \theta_2) \geq 0$

$B_F(\theta_1 : \theta_2) = \frac{1}{2} (\theta_2 - \theta_1)^T \nabla^2 F(\theta) (\theta_2 - \theta_1) \geq 0$, $\exists\, \theta \in [\theta_1 \theta_2]$ (squared Mahalanobis/Euclidean)

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

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

$= T_{\theta 2}(\theta_1)$: Tangent of the function graph at θ_2 evaluated at θ_1



Relative entropy between Gaussian distributions?

- Relative entropy = **cross-entropy** minus **entropy**, a dissimilarity measure in information theory, statistics, ML, etc.

  called **Kullback-Leibler divergence** in information theory**:**

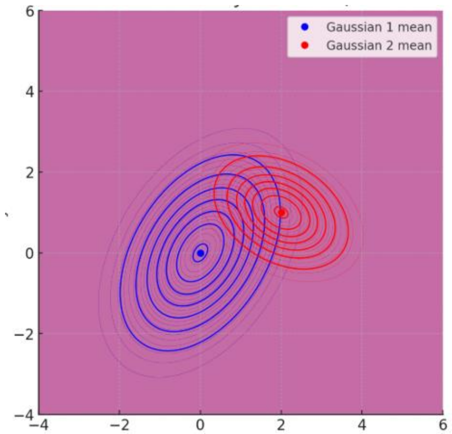
$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

$$D_{KL}(P||Q) = \int P(x) \log \frac{P(x)}{Q(x)} dx$$

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

KLD integrals ∫ is non-trivial but….

- *How to calculate the KLD between multivariate Gaussian distributions?*



$$p(x) = \frac{1}{(2\pi)^{d/2}|\Sigma_p|^{1/2}} \exp\left(-\tfrac{1}{2}(x-\mu_p)^\top \Sigma_p^{-1}(x-\mu_p)\right)$$

$$q(x) = \frac{1}{(2\pi)^{d/2}|\Sigma_q|^{1/2}} \exp\left(-\tfrac{1}{2}(x-\mu_q)^\top \Sigma_q^{-1}(x-\mu_q)\right)$$

Expectation of this term?

$$\log \frac{p(x)}{q(x)} = \frac{1}{2} \log \frac{|\Sigma_q|}{|\Sigma_p|} + \frac{1}{2}(x-\mu_q)^\top \Sigma_q^{-1}(x-\mu_q) - \frac{1}{2}(x-\mu_p)^\top \Sigma_p^{-1}(x-\mu_p)$$

Bregman divergences in ML: Exponential families

- Kullback-Leibler divergence between two probability densities:

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

can be **difficult to calculate in closed form** because of the integral ∫

- But Kullback-Leibler divergence between two probability densities of an **exponential family** with densities $p(x|\theta) = \exp(<t(x), \theta> - F(\theta))$

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

where F is the **cumulant function** (log partition/log Laplace function)

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

BDs wrt CFs between parameters = reverse KLD distributions

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)**

Kullback-Leibler divergence between non-normalized exponential family densities

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

$$D_{KL}^{+}[q_1(x):q_2(x)] = \int \{ q_1(x) \log (q_1(x)/q_2(x)) + q_2(x)-q_1(x) \} \, d\mu(x)$$

- Exponential family density: Normalized $p(x|\theta) = \exp(\langle x, \theta \rangle - F(\theta))\, d\mu(x)$ versus **non-normalized**: $q(x|\theta) = \exp(\langle x, \theta \rangle)\, d\mu(x)$

- 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**, and log-convex functions are also convex functions:

Both F and Z are convex functions, yields Bregman divergences B_F & B_Z

- Well-known: KLD between normalized densities = reverse Bregman wrt F:

$$D_{KL}[p_{\theta 1}(x):p_{\theta 2}(x)] = B_F^{*}[\theta_1 : \theta_2] = B_F[\theta_2 : \theta_1]$$

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

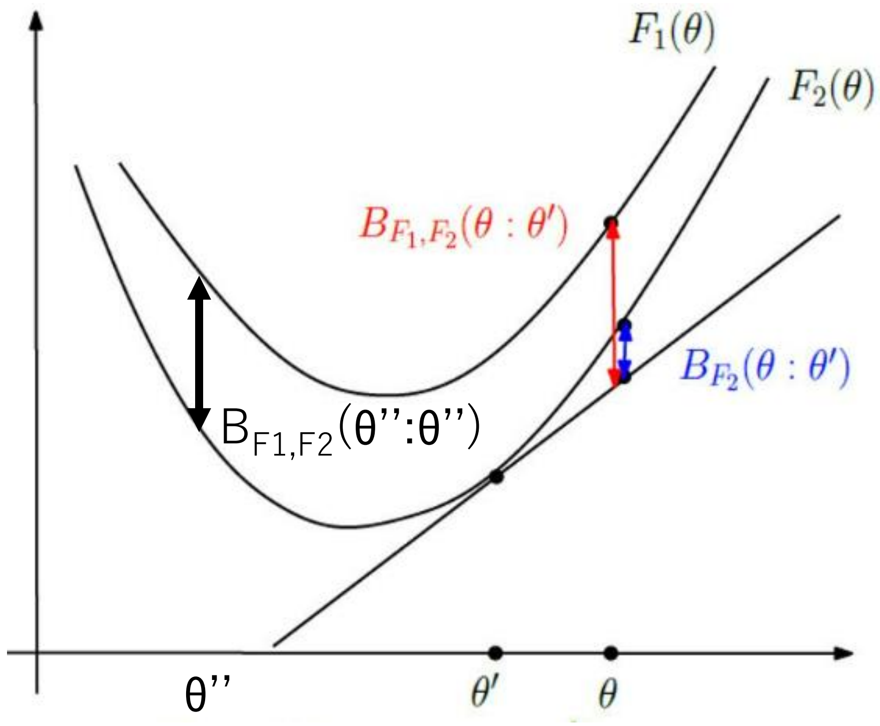
$$D_{KL}^{+}[q_{\theta 1}(x):q_{\theta 2}(x)] = B_Z^{*}[\theta_1 : \theta_2] = B_Z[\theta_2 : \theta_1]$$

2312.12849

Outline

1. A quick introduction to Bregman divergences
2. **Duo Bregman pseudo-divergences**
3. Curved Bregman divergences
4. Generalized Legendre transforms and information geometry
5. Generalized convexity and Bregman divergences
6. Space of Bregman balls

Duo Bregman divergences: Generalize BDs with a pair of generators



One generator **majorizes** the other one:

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

Duo Bregman pseudo-divergence:

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

$$\geq B_{F2}(\theta:\theta')$$

- 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.

But why considering two generators?

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

q(x| θ)» p(x| θ)

p(x| θ) X_1

X_2

q(x| θ)

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

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

- Consider an exponential family on support X_1:

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

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

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

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

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**:

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

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

KL divergence between truncated normal densities

PDF of truncated normal on (a,b): $$p_{m,s}^{a,b}(x) = \frac{1}{\sqrt{2\pi}s\ (\Phi_{m,s}(b) - \Phi_{m,s}(a))} \exp\left(-\frac{(x-m)^2}{2s^2}\right)$$

This is an example, DO NOT READ!
For illustration purpose!

$$\Phi_{m,s}(x) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{x-m}{\sqrt{2}s}\right)\right), \quad \mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\mathrm{d}t.$$

Truncated normal PDFs form an exponential family with log-normalizer :

$$F_{a,b}(m,s) = \frac{m^2}{2s^2} + \frac{1}{2}\log 2\pi s^2 + \log\left(\Phi_{m,s}(b) - \Phi_{m,s}(a)\right)$$

Moment parameters and mean & variance:

$$\mu(m,s;a,b) = m - s\,\frac{\phi(\beta)-\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}, \qquad \phi(x) := \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$$

$$\begin{aligned}\eta_1(m,s;a,b) &= E_{p_{m,s}^{a,b}}[x] = \mu(m,s;a,b),\\ \eta_2(m,s;a,b) &= E_{p_{m,s}^{a,b}}[x^2] = \sigma^2(m,s;a,b) + \mu^2(m,s;a,b).\end{aligned}$$

$$\sigma^2(m,s;a,b) = s^2\left(1 - \frac{\beta\phi(\beta)-\alpha\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)} - \left(\frac{\phi(\beta)-\phi(\alpha)}{\Phi(\beta)-\Phi(\alpha)}\right)^2\right),$$

Kullback-Leibler divergence between nested truncated normal distributions:

$$\begin{aligned}D_{\mathrm{KL}}[p_{m_1,s_1}^{a_1,b_1} : p_{m_2,s_2}^{a_2,b_2}] &= \frac{m_2}{2s_2^2} - \frac{m_1}{2s_1^2} + \log\frac{Z_{a_2,b_2}(m_2,s_2)}{Z_{a_1,b_1}(m_1,s_1)} - \left(\frac{m_2}{s_2^2} - \frac{m_1}{s_1^2}\right)\eta_1(m_1,s_1;a_1,b_1)\\ &\quad - \left(\frac{1}{2s_1^2} - \frac{1}{2s_2^2}\right)\eta_2(m_1,s_1;a_1,b_1)\end{aligned}$$

if nested distributions $(a_1,b_1) \subseteq (a_2,b_2)$

$D_{\mathrm{KL}}[p_{m_1,s_1}^{a_1,b_1} : p_{m_2,s_2}^{a_2,b_2}] = +\infty, (a_1,b_1) \not\subseteq (a_2,b_2)$ otherwise

Convex duality via Legendre-Fenchel transform

- Convex duality: 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 (Θ,F(θ))**

  (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 $\theta = \nabla F^*(\eta)$ and $\nabla F^* = (\nabla F)^{-1}$

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 iff $\eta_2 = \nabla F(\theta_1)$
- Build the **Fenchel-Young divergence** from inequality gap lhs-rhs ≥0

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

- BD with **mixed parameterizations** θ and η : $B_F(\theta_1 : \theta_2) = Y_{F,F^*}(\theta_1, \eta_2)$
- Duality: **Four equivalent expressions** of the "Bregman divergence"

$$B_F(\theta_1 : \theta_2) = Y_{F,F^*}(\theta_1, \eta_2) = Y_{F^*,F}(\eta_2, \theta_1) = B_{F^*}(\eta_2, \eta_1)$$

- Dual BDs + Dual FYs from involution $F^{**}=F$

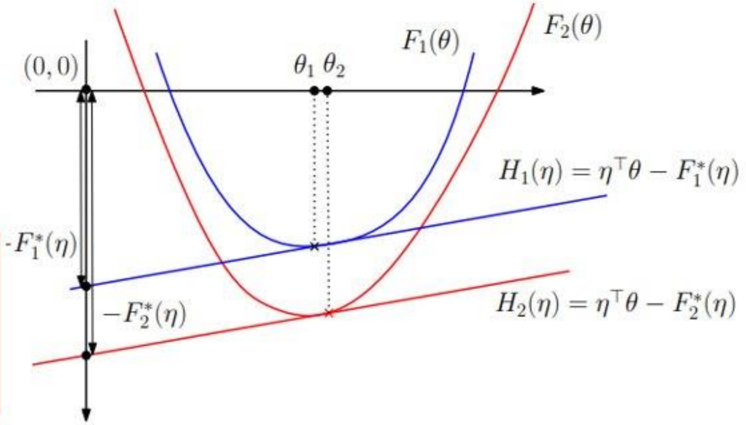
Note : $B_F(\theta_1 : \theta_2)=0 \Leftrightarrow \theta_1 = \theta_2 \Leftrightarrow \eta_1 = \eta_2$ i.e., $\nabla F(\theta_1) = \nabla F(\theta_2)$

Legendre transformation reverses majorization order

Legendre-Fenchel transformation: $F^*(\eta) := \sup_{\theta\in\Theta}\{\eta^\top\theta - F(\theta)\}$

F Legendre-type function, Moreau **biconjugation theorem**: $(F^*)^* = F$

proper+lower semi-continuous+convex

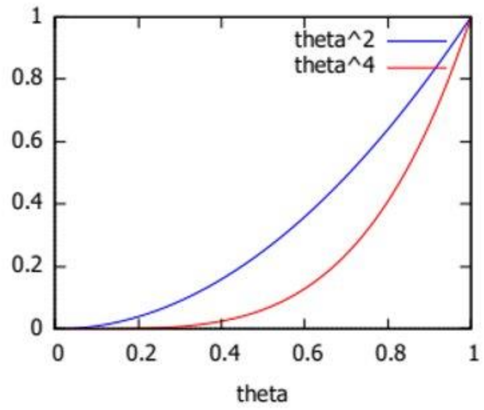


Legendre-Fenchel transform **reverses ordering**:

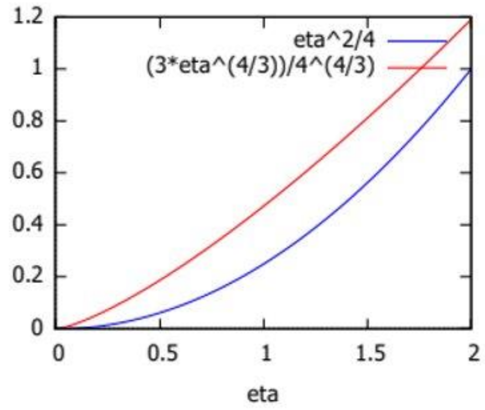
$$\boxed{\forall\theta\in\Theta,\quad F_1(\theta)\geq F_2(\theta) \Leftrightarrow \forall\eta\in H,\quad F_1^*(\eta)\leq F_2^*(\eta)}$$

Proof:

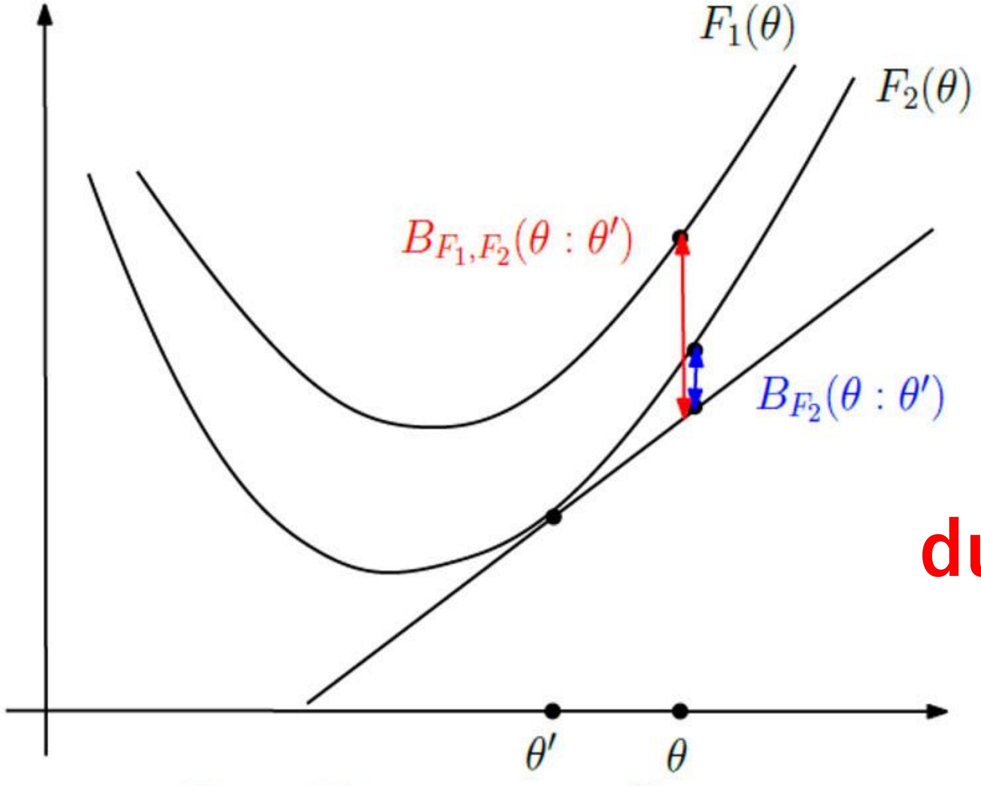
$$\begin{aligned}
F_1^*(\eta) &:= \sup_{\theta\in\Theta}\{\eta^\top\theta - F_1(\theta)\},\\
&= \eta^\top\theta_1 - F_1(\theta_1) \qquad (\text{with } \eta = \nabla F_1(\theta_1))\\
&\leq \eta^\top\theta_1 - F_2(\theta_1),\\
&\leq \sup_{\theta\in\Theta}\{\eta^\top\theta - F_2(\theta)\} =: F_2^*(\eta).
\end{aligned}$$



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



Conjugate functions $F_1^*(\eta)\leq F_2^*(\eta)$



Duality
and
duo Fenchel-Young pseudo-divergences

Duo Bregman divergence

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

Duo Fenchel-Young divergence

$$Y_{F_1,F_2^*}(\theta,\eta') := F_1(\theta) + F_2^*(\eta') - \theta^\top \eta'.$$

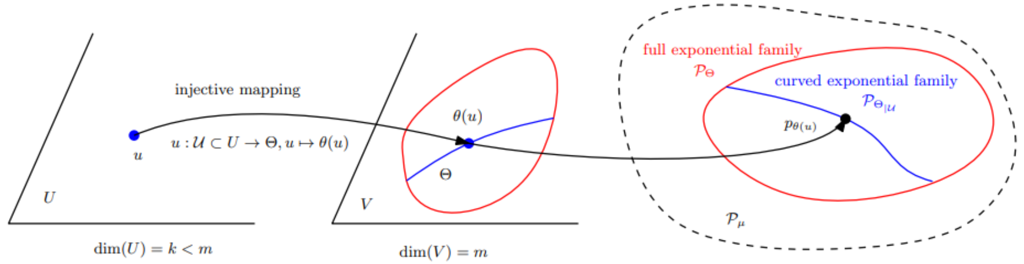
Relationship with truncated exponential families with nested supports:

$$D_{\mathrm{KL}}[p_{\theta_1} : q_{\theta_2}] = Y_{F_2,F_1^*}(\theta_2 : \eta_1) = B_{F_2,F_1}(\theta_2 : \theta_1)$$

Outline

1. A quick introduction to Bregman divergences
2. Duo Bregman pseudo-divergences
3. **Curved Bregman divergences**
4. Generalized Legendre transforms and information geometry
5. Generalized convexity and Bregman divergences
6. Space of Bregman balls

Curved Bregman divergences



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

$\mathbf{B_{F,c}(u_1 : u_2)}$:= $\mathbf{B_F(c(u_1):c(u_2))}$ 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
(usually not cvx affine space, hence not a Bregman divergence)

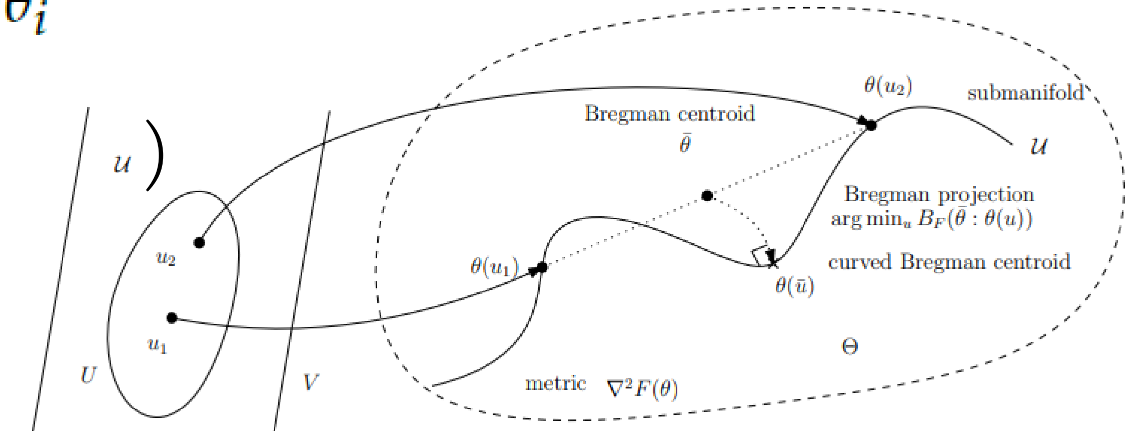
Curved Bregman centroid is the Bregman projection of the full Bregman centroid

Theorem: **Bregman projection**

$\bar{\theta} = \sum_i w_i \theta_i$

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

N-point opt. $\theta_i = \theta(u_i)$ **One-point opt!**



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.

Another example of curved Bregman divergences:

- Consider d-variate **circular complex normal distributio** $\mathcal{CN}_d(\mu_\mathbb{C}, S_\mathbb{C})$
- CNDs handled as **2d real normal distributions** $N_{2d}([\mu_\mathbb{C}]_\mathbb{R}, \frac{1}{2}[S_\mathbb{C}]_\mathbb{R})$

$$[z = a + ib]_\mathbb{R} = (a, b) \qquad [M = A + iB]_\mathbb{R} = \begin{bmatrix} A & -B \\ B & A \end{bmatrix}$$

- Then KLD between CNDs amounts to a **curved Bregman divergence**:

$$D_{\mathrm{KL}}[p_{m_\mathbb{C}, S_\mathbb{C}} : p_{m'_\mathbb{C}, S'_\mathbb{C}}] = D_{\mathrm{KL}}[p_{\mu,\Sigma} : p_{\mu',\Sigma'}] = B_F(\theta' : \theta)$$

Natural parameters

$(\Sigma^{-1}\mu, \frac{1}{2}\Sigma^{-1})$

$\mu = [m_\mathbb{C}]_\mathbb{R}$ $\Sigma = [S_\mathbb{C}]_\mathbb{R}$ and $\mu' = [m'_\mathbb{C}]_\mathbb{R}$ $\Sigma' = [S'_\mathbb{C}]_\mathbb{R}$

- **Curved submanifold** parameter in R^{2d} x Mat(2d)

$$\mathcal{U} = \left\{ (v, M) : v \in \mathbb{R}^{2d}, M = \begin{bmatrix} A & -B \\ B & A \end{bmatrix}, A \in \mathbb{R}^{d\times d} \succ 0, B \in \mathbb{R}^{d\times d} \succ 0 \right\}$$

2504.05654

Outline

1. A quick introduction to Bregman divergences
2. Duo Bregman pseudo-divergences
3. Curved Bregman divergences
4. **Generalized Legendre transforms and information geometry**
5. Generalized convexity and Bregman divergences
6. Space of Bregman balls

Generalized Legendre transforms (2009)

The concept of duality in convex analysis, and the characterization of the Legendre transform

By Shiri Artstein-Avidan and Vitali Milman*

Abstract

In the main theorem of this paper we show that any involution on the class of lower semi-continuous convex functions which is order-reversing, must be, up to linear terms, the well known Legendre transform.

Theorem 1. *Assume a transform* $\mathcal{T} : Cvx(\mathbb{R}^n) \to Cvx(\mathbb{R}^n)$ *(defined on the whole domain* $Cvx(\mathbb{R}^n)$*) satisfies*

1. $\mathcal{T}\mathcal{T}\phi = \phi$,
2. $\phi \le \psi$ *implies* $\mathcal{T}\phi \ge \mathcal{T}\psi$.

Then, $\mathcal{T}$ *is essentially the classical Legendre transform; namely there exists a constant* $C_0 \in \mathbb{R}$, *a vector* $v_0 \in \mathbb{R}^n$, *and an invertible symmetric linear transformation* $B \in \mathrm{GL}_n$ *such that*

$$(\mathcal{T}\phi)(x) = (\mathcal{L}\phi)(Bx + v_0) + \langle x, v_0 \rangle + C_0.$$

Generalized Artstein-Avidan—Milman Legendre transforms

Definition (Generalized Legendre-Fenchel convex conjugates) *Let* $\mathcal{L}_{\lambda,E,f,g,h}$ *denote a generalized Legendre-Fenchel transform:*

$$\mathcal{L}_{\lambda,E,f,g,h}F := \mathcal{L}_P F := \lambda(\mathcal{L}F)(E\eta + f) + \langle \eta, g\rangle + h$$

for the parameter $P = (\lambda, E, f, g, h)$. where $(\mathcal{L}F)(\eta) := \sup_{\theta\in\mathbb{R}^m} \{\langle\theta,\eta\rangle - F(\theta)\}$

Affine-deformed convex functions of both argument and returned value remain convex:

$$F_P(\theta) := \lambda\, F(A\theta + b) + \langle\theta, c\rangle + d \qquad P = (\lambda, A, b, c, d) \in \mathbb{P} := \mathbb{R}_{>0} \times \mathrm{GL}(\mathbb{R}^m) \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$$

Theorem *For any* $F \in \Gamma_0$, *we have* $\mathcal{L}_P(F) := (F^*)_P = \mathcal{L}\left(F_{P^\diamond}\right)$

$$P^\diamond := \left(\lambda, \frac{1}{\lambda}A^{-1}, -\frac{1}{\lambda}A^{-1}c, -A^{-1}b, \langle b, A^{-1}c\rangle - d\right) \in \mathbb{P}$$

$$(P^\diamond)^\diamond = P$$

GLTs are LT on affinely deformed convex functions

arXiv:2507.20577

Generalized Artstein-Avidan—Milman Legendre transforms

Definition **(Generalized Legendre-Fenchel convex conjugates** **)** *Let* $\mathcal{L}_{\lambda,E,f,g,h}$ *denote a generalized Legendre-Fenchel transform:*

$$\mathcal{L}_{\lambda,E,f,g,h}F := \mathcal{L}_P F := \lambda(\mathcal{L}F)(E\eta + f) + \langle \eta, g\rangle + h$$

for the parameter $P = (\lambda, E, f, g, h)$. where $(\mathcal{L}F)(\eta) := \sup_{\theta\in\mathbb{R}^m} \{\langle\theta,\eta\rangle - F(\theta)\}$

Affine-deformed convex functions of both argument and returned value remain convex:

$F_P(\theta) := \lambda\, F(A\theta + b) + \langle\theta, c\rangle + d$ $\quad P = (\lambda, A, b, c, d) \in \mathbb{P} := \mathbb{R}_{>0} \times \mathrm{GL}(\mathbb{R}^m) \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$

Theorem *For any* $F \in \Gamma_0$*, we have* $\mathcal{L}_P(F) := (F^*)_P = \mathcal{L}\left(F_{P^\diamond}\right)$

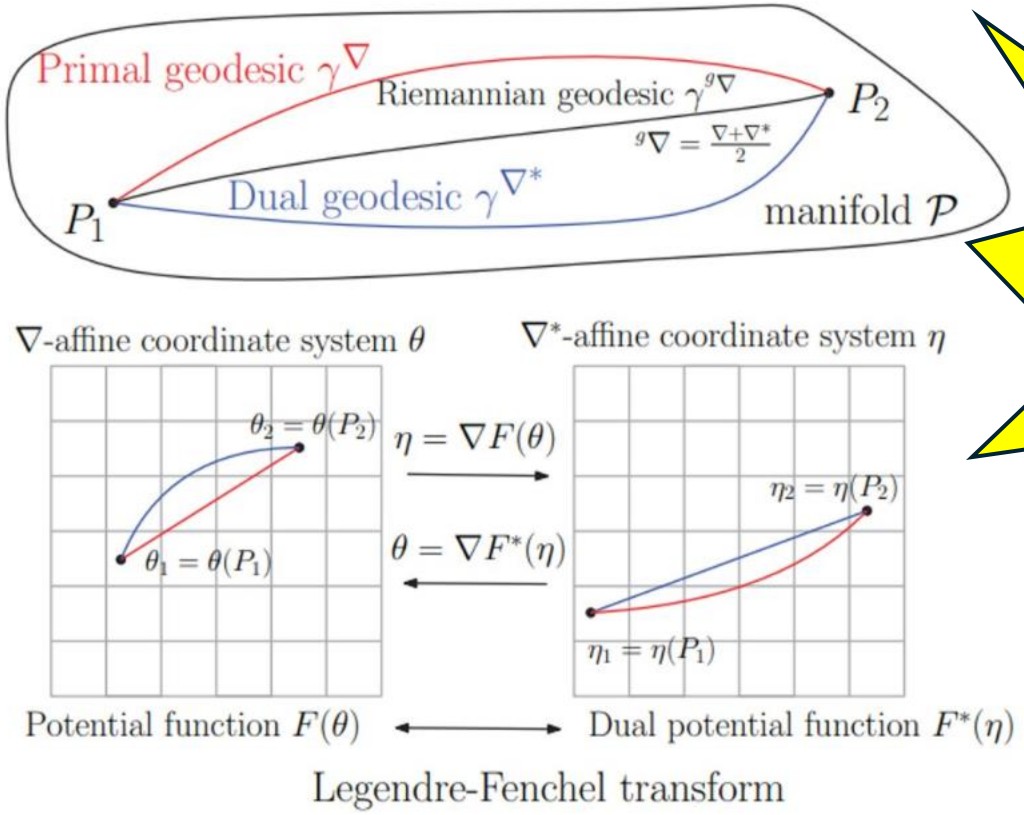
$$P^\diamond := \left(\lambda, \frac{1}{\lambda}A^{-1}, -\frac{1}{\lambda}A^{-1}c, -A^{-1}b, \langle b, A^{-1}c\rangle - d\right) \in \mathbb{P}$$

$$(P^\diamond)^\diamond = P$$

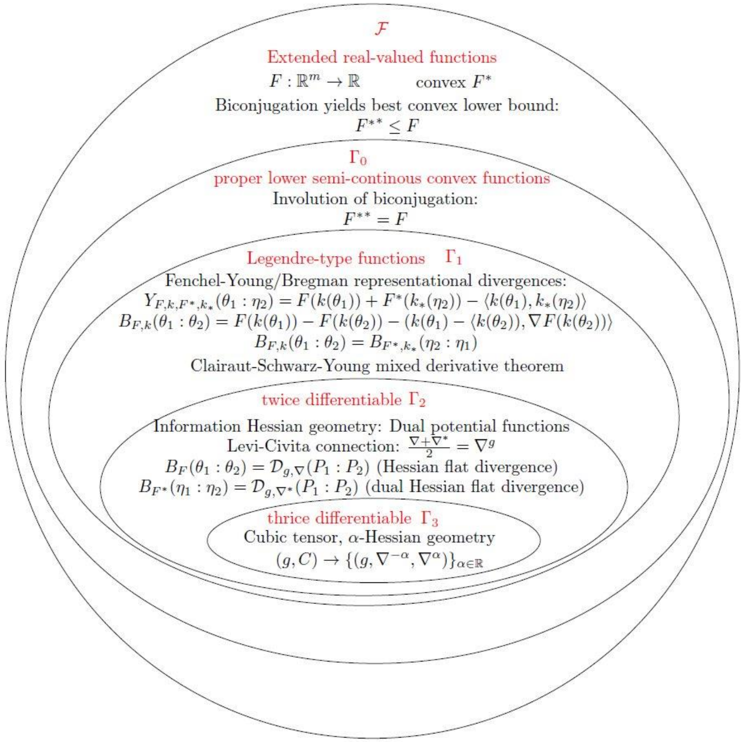
arXiv:2507.20577

GLTs can be explained from information geometry

DO NOT READ!
For illustration purpose!



Dually flat spaces, Hessian manifolds



Degrees of freedom when reconstruction dual potential functions

- Affine coordinate system up to affine transformation

$$\mathcal{L}_{\lambda,E,f,g,h}F := \mathcal{L}_P F := \lambda(\mathcal{L}F)(E\eta + f) + \langle \eta, g\rangle + h$$

- Scaling of potential functions
- Same geometric Legendre transform

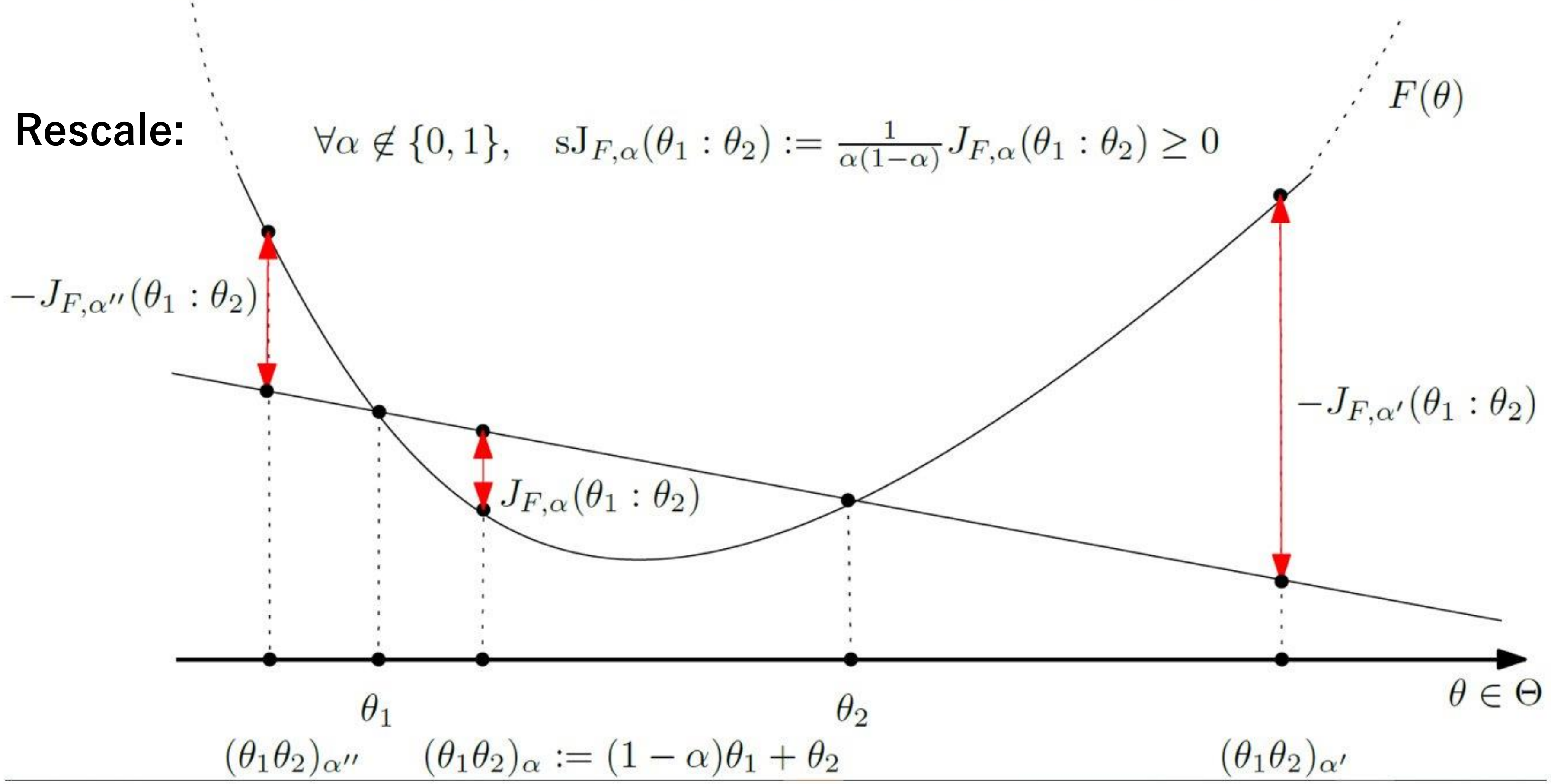
Outline

1. A quick introduction to Bregman divergences
2. Duo Bregman pseudo-divergences
3. Curved Bregman divergences
4. Generalized Legendre transforms and information geometry
5. **Generalized convexity and Bregman divergences**
6. Space of Bregman balls

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)

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

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

for all t in [0,1]

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

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

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 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.*

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

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

Definition:

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

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*

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

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 $F : I \subset \mathbb{R} \rightarrow \mathbb{R}$ 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:*

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

From 1st order Taylor expansion…

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)

Outline

1. A quick introduction to Bregman divergences

2. Duo Bregman pseudo-divergences

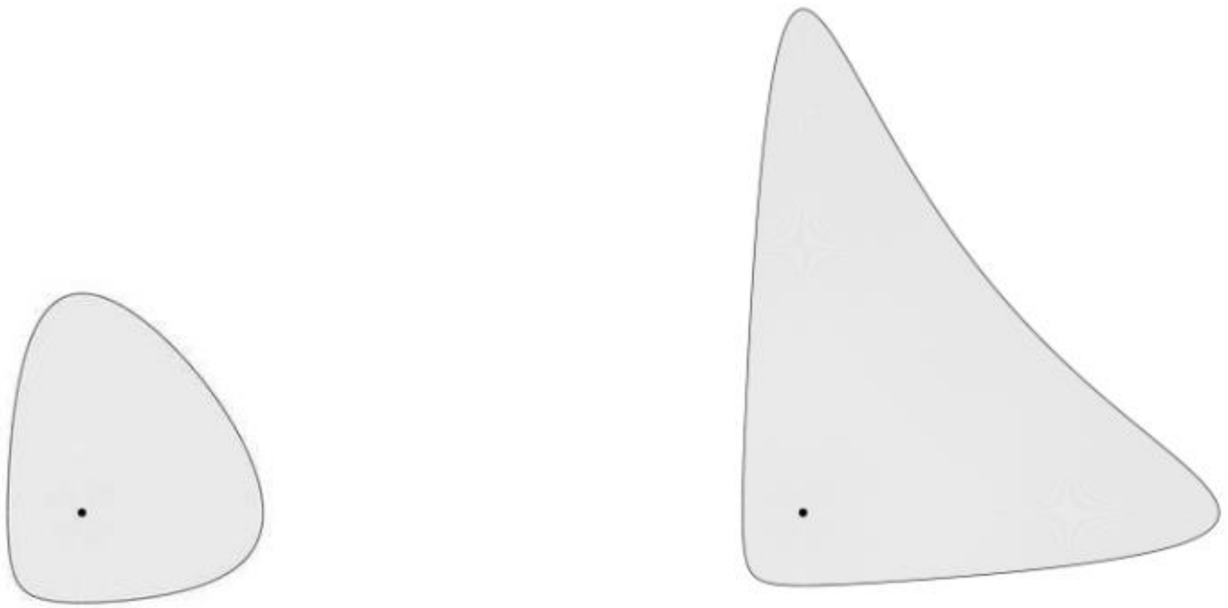
3. Curved Bregman divergences

4. Generalized Legendre transforms and information geometry

5. Generalized convexity and Bregman divergences

6. **Boolean algebra of Bregman balls**

Space of Bregman balls



Example:
Itakura-Saito right and left spheres

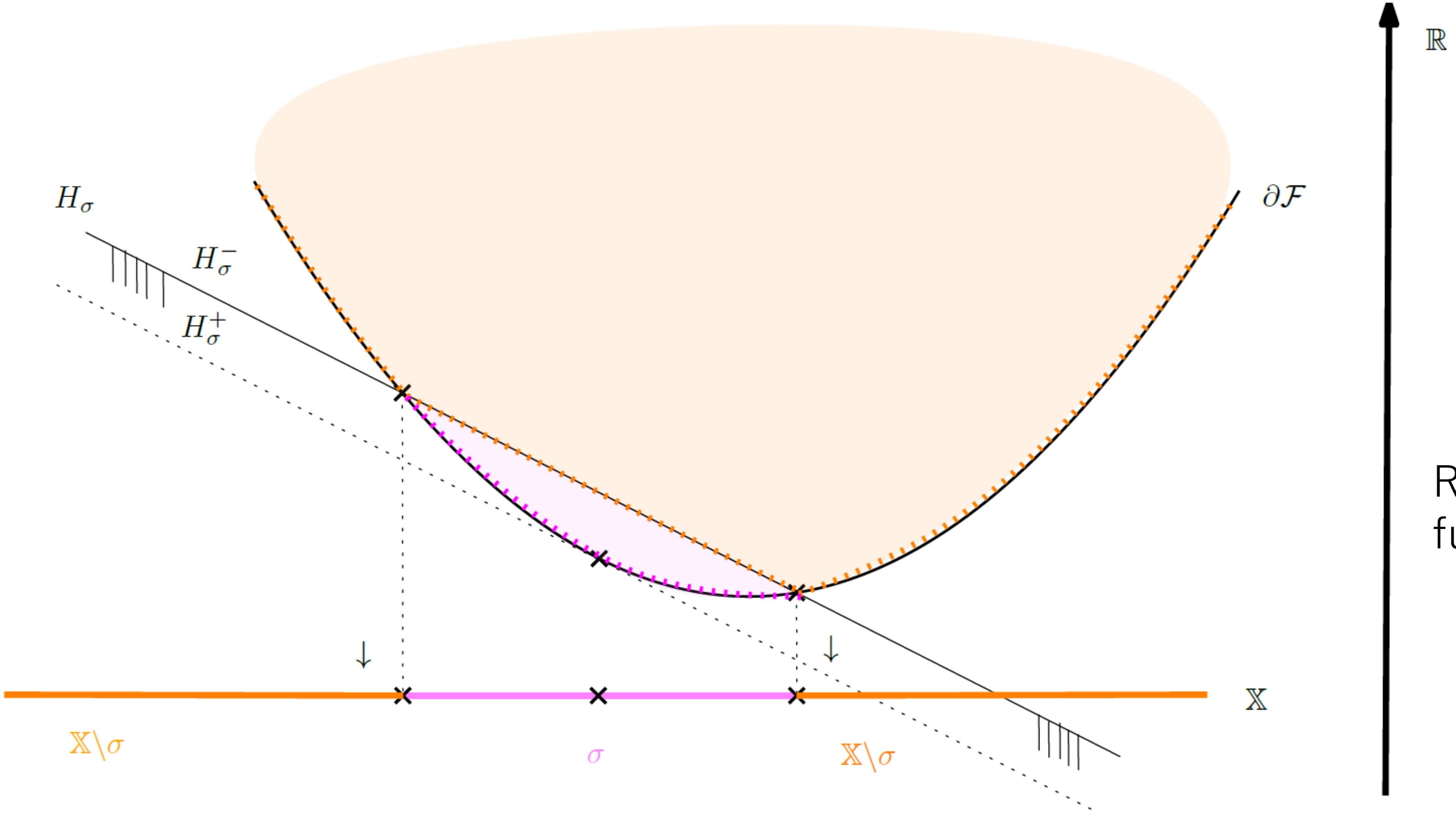
Right-sided Bregman ball: $\sigma_F(\theta, r) = \{\theta' \in \Theta : B_F(\theta' : \theta) \leq r\}$

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

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



$\downarrow$ means vertical projection

S^c: complement of set S

To any sphere, associate a hyperplane:

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

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

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

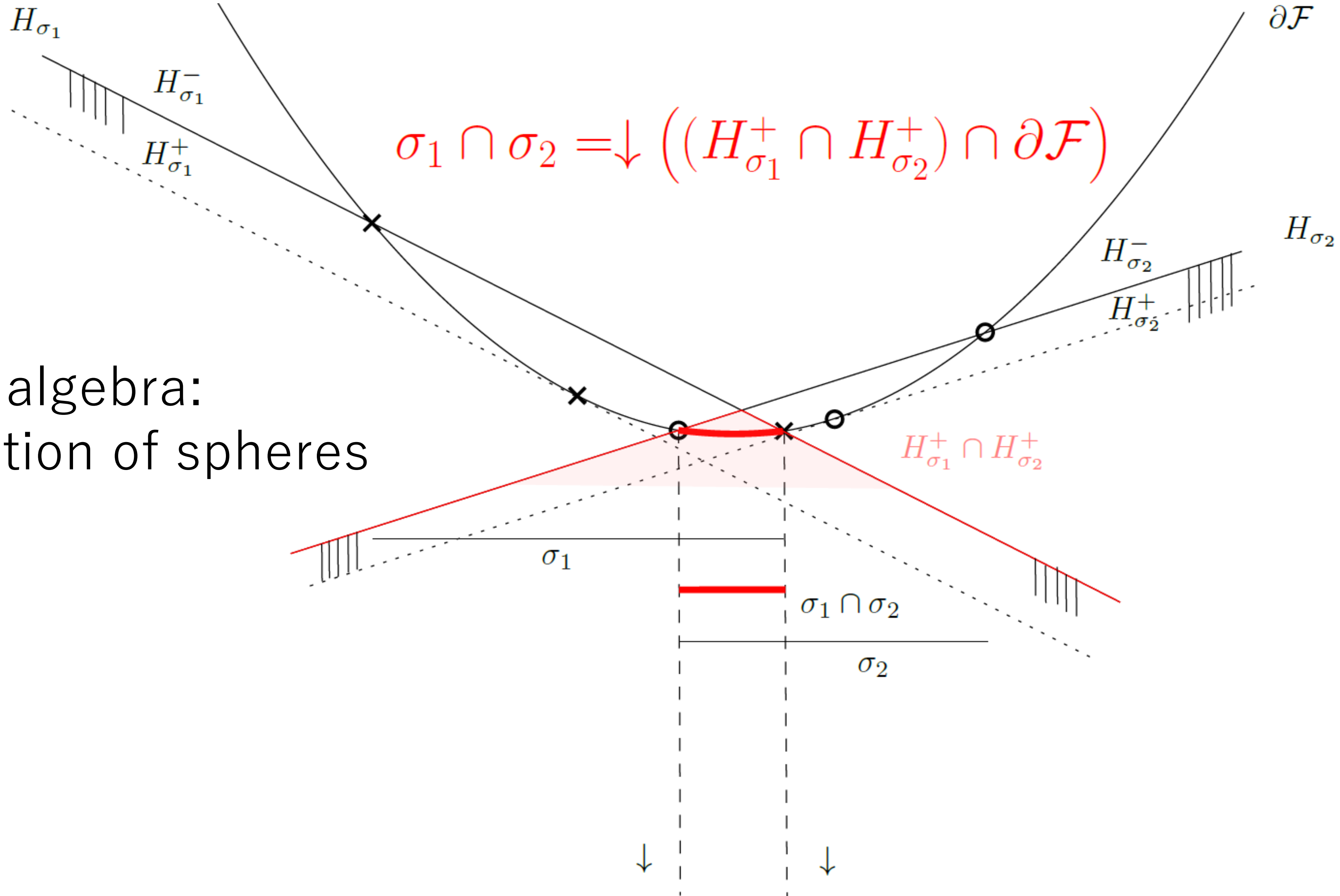
Center: $\mathbf{c} = \boldsymbol{\nabla}^{-1} F(\mathbf{a})$

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

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

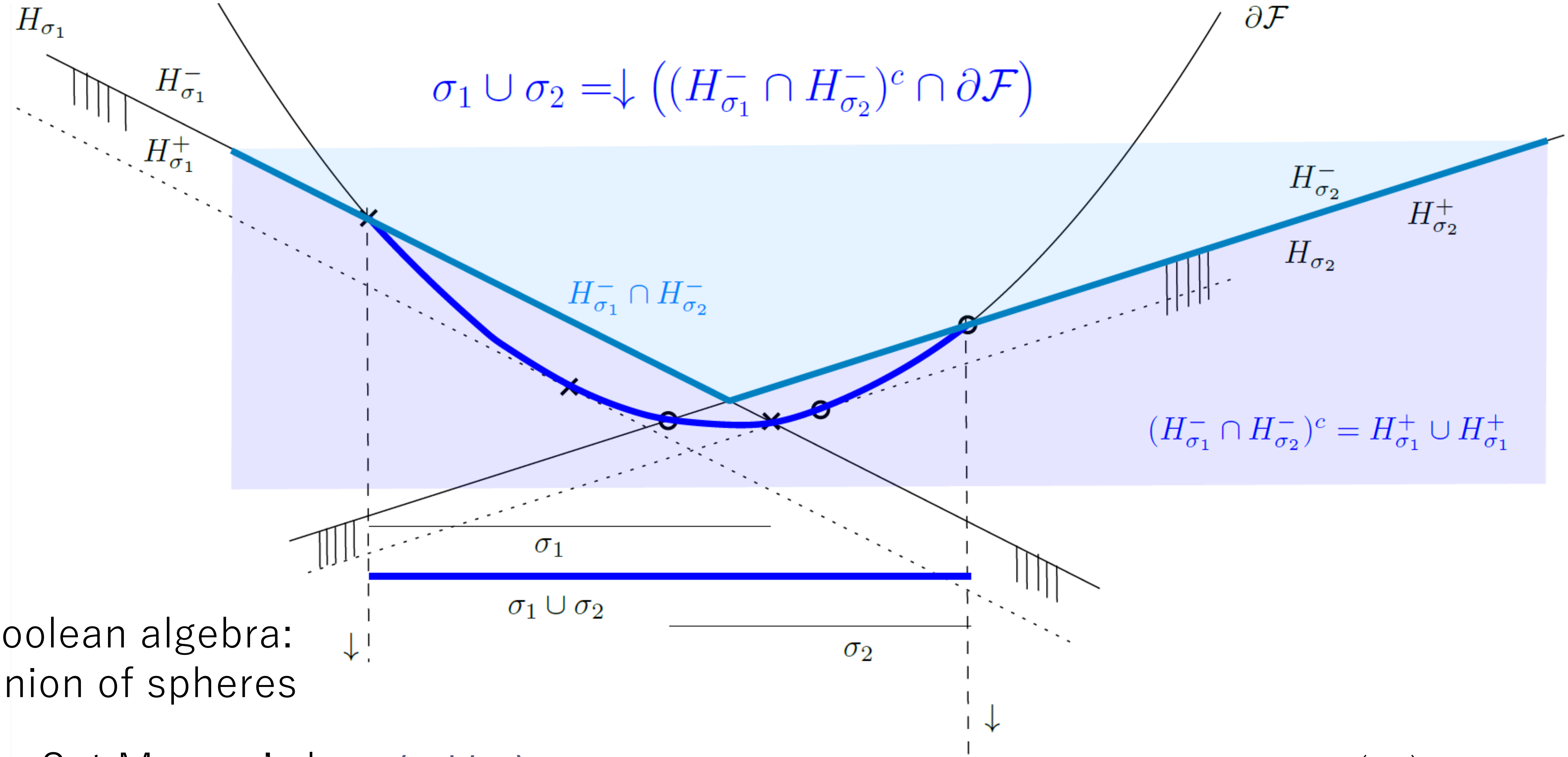
Lifting to potential Bregman generator graph & vertical projection

Intersection of two right Bregman balls



Boolean algebra:
intersection of spheres

Union of two right Bregman balls

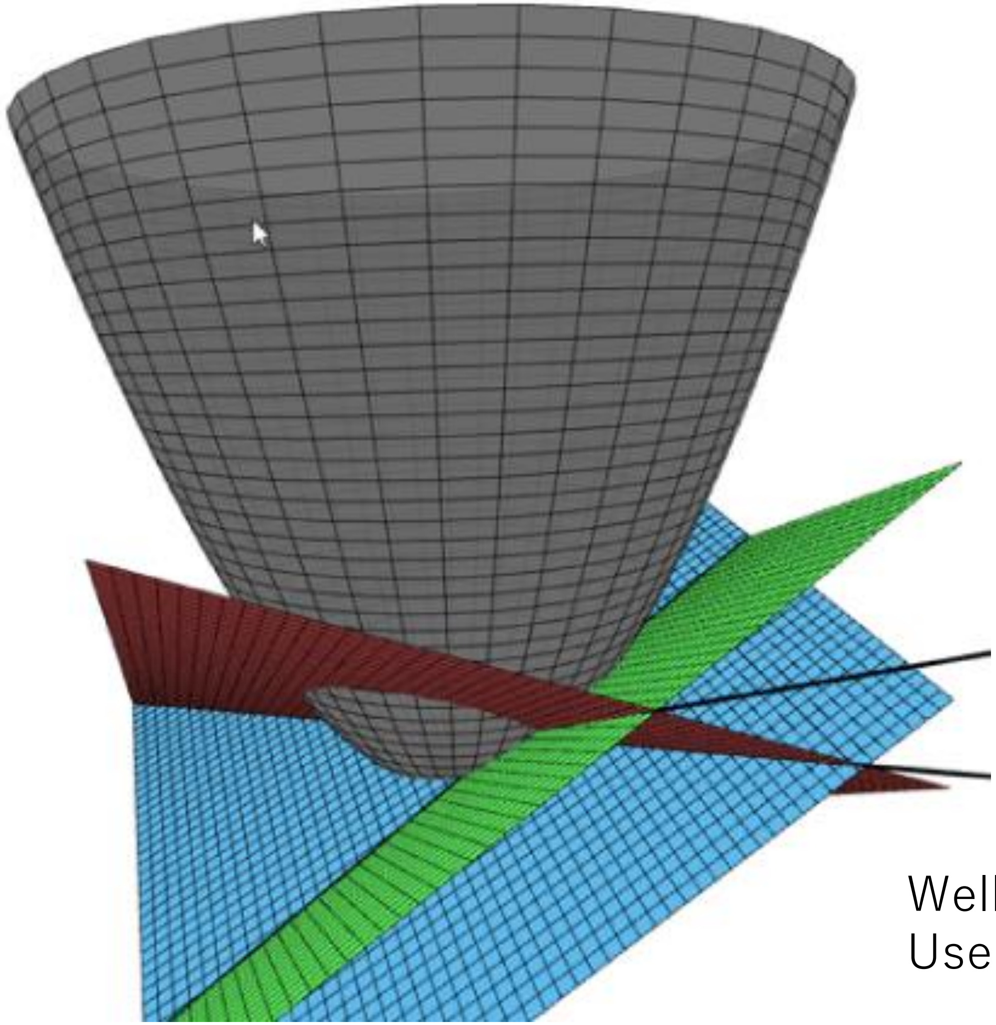


Boolean algebra:
Union of spheres

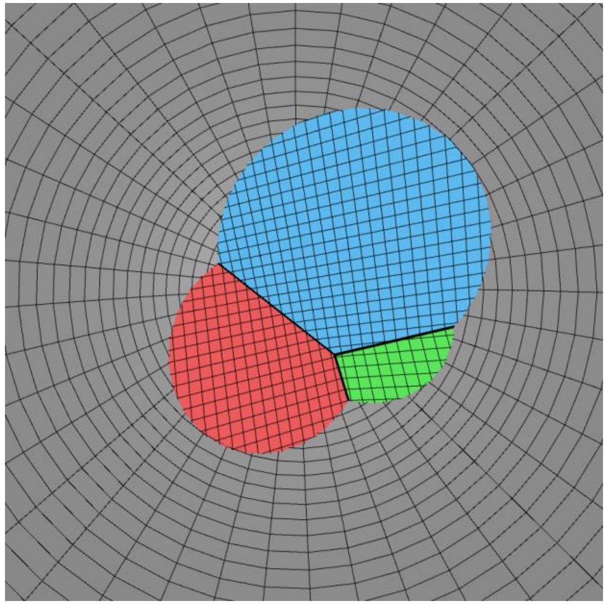
Set Morgan's law: $(A \cup B)^c = A^c \cap B^c$

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

Example: Euclidean spheres potential function: Paraboloid, L22



Top view displays the union of disks



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

Well-known “paraboloid transform” in computational geometry
Used for computing (Bregman) Voronoi and dual Delaunay complex

Wrapping up

- Quick introduction to Bregman divergences

BD reconstructed for partition function Z = reverse extended KLD+

- Duo Bregman pseudo-divergences

BD with 2 majorized generators, applications to KLD between nested exp fams

- Curved Bregman divergences

Symmetrized BD or KLD between circular complex normal, projection theorem

- Generalized Legendre transforms and information geometry

Rev. order involut. transf. = affine-deformed LT = geometric LT of Hessian mfd

- Generalized convexity and Bregman divergences

BDs = limits of scaled skew Jensen div., comparative cvxity, conformal factor

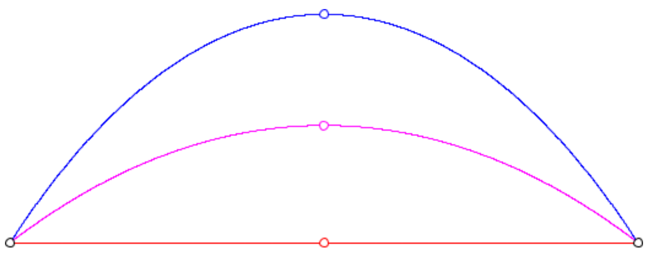
- Space of Bregman balls , Boolean algebra of Bregman balls

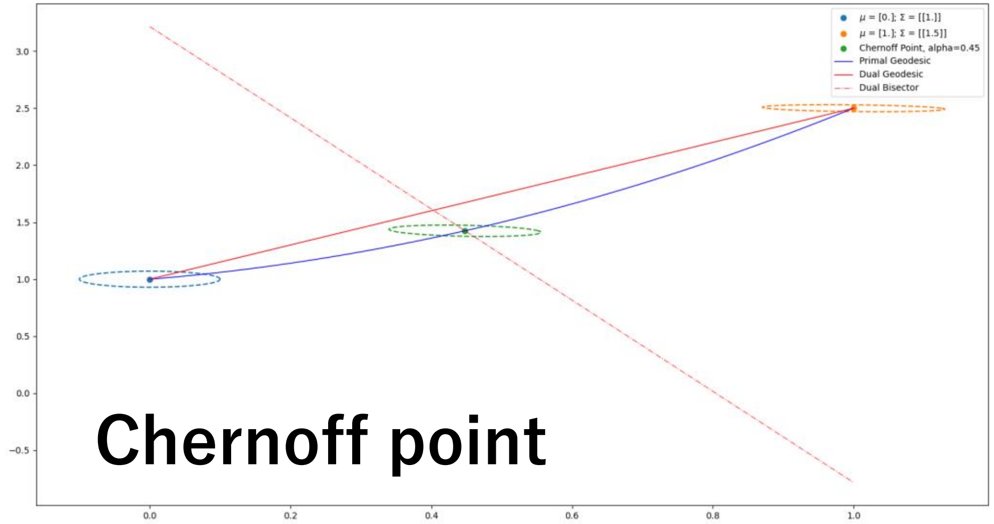
Embedding into higher dim., Boolean operations = intersections of halfspaces

A Python library for geometric computing on Bregman Manifolds

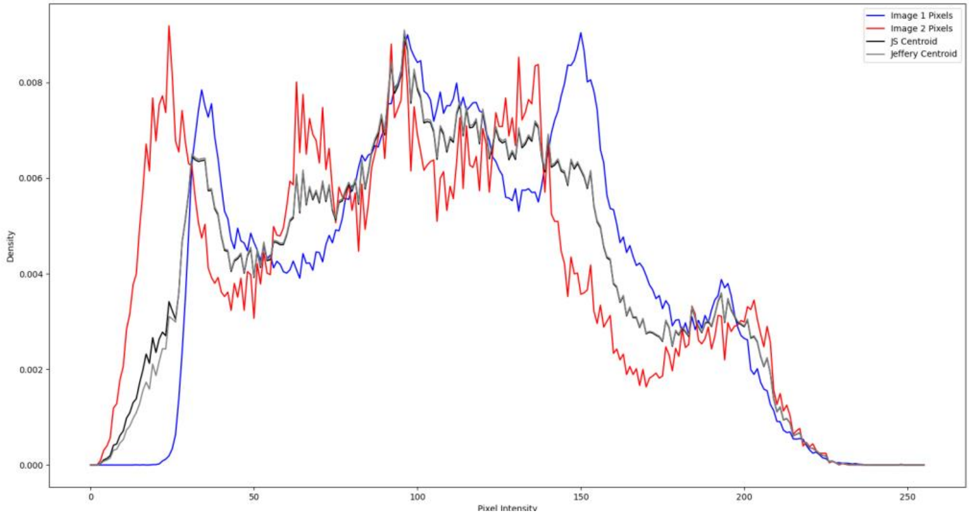
pyBregMan

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

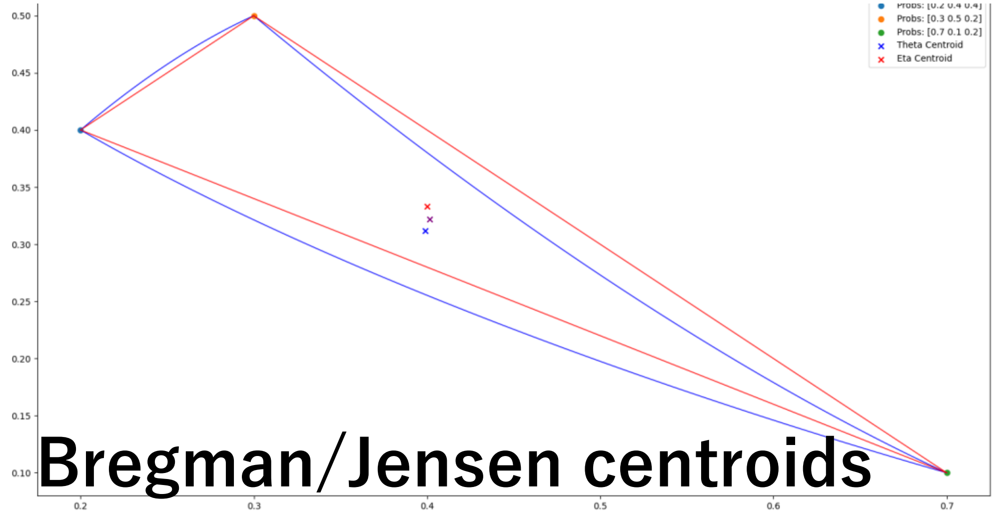




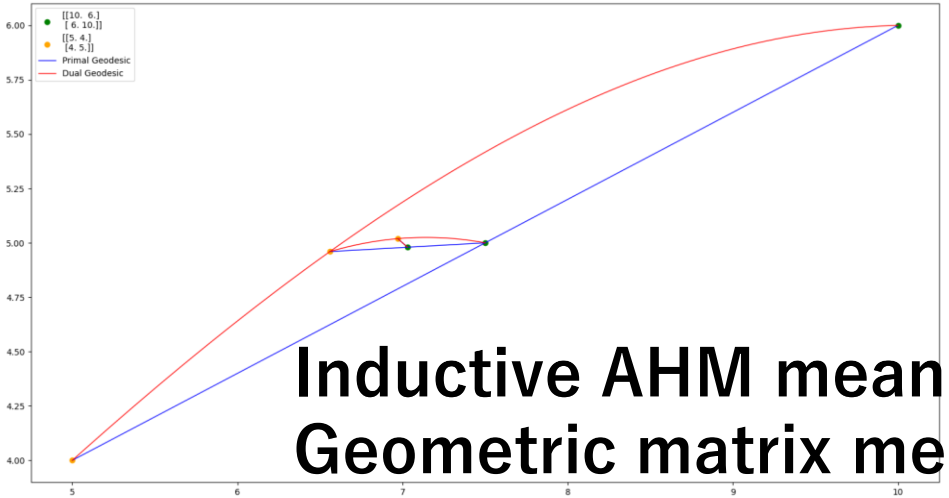
Chernoff point



Jensen-Shannon centroid



Bregman/Jensen centroids



Inductive AHM mean
Geometric matrix mean

Joint work of Frank Nielsen and Alexander Soen

Some references

- A quick introduction to Bregman divergences

Divergences Induced by the Cumulant and Partition Functions of Exponential Families and Their Deformations Induced by Comparative Convexity. Entropy 26(3): 193 (2024)

- Duo Bregman pseudo-divergences

Statistical Divergences between Densities of Truncated Exponential Families with Nested Supports: Duo Bregman and Duo Jensen Divergences. Entropy 24(3): 421 (2022)

- Curved Bregman divergences

Curved representational Bregman divergences and their applications. 2504.05654 (2025)

- Generalized Legendre transforms and information geometry

A note on the Artstein-Avidan-Milman's generalized Legendre transforms. 2507.20577 (2025)

- Generalized convexity and Bregman divergences

Generalizing Skew Jensen Divergences and Bregman Divergences With Comparative Convexity. IEEE Signal Process. Lett. 24(8): 1123-1127 (2017)

- Space of Bregman balls

Bregman Voronoi Diagrams. Discret. Comput. Geom. 44(2): 281-307 (2010)

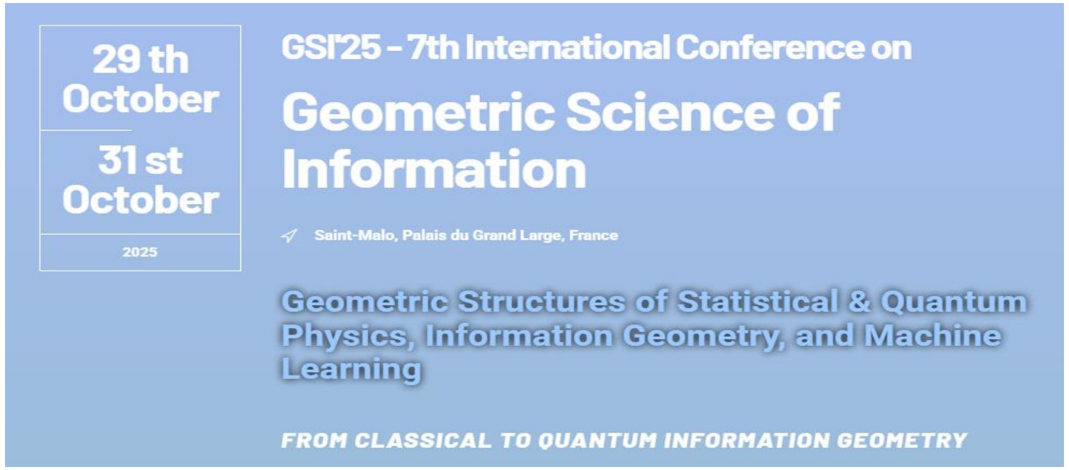
Yet further generalizations…

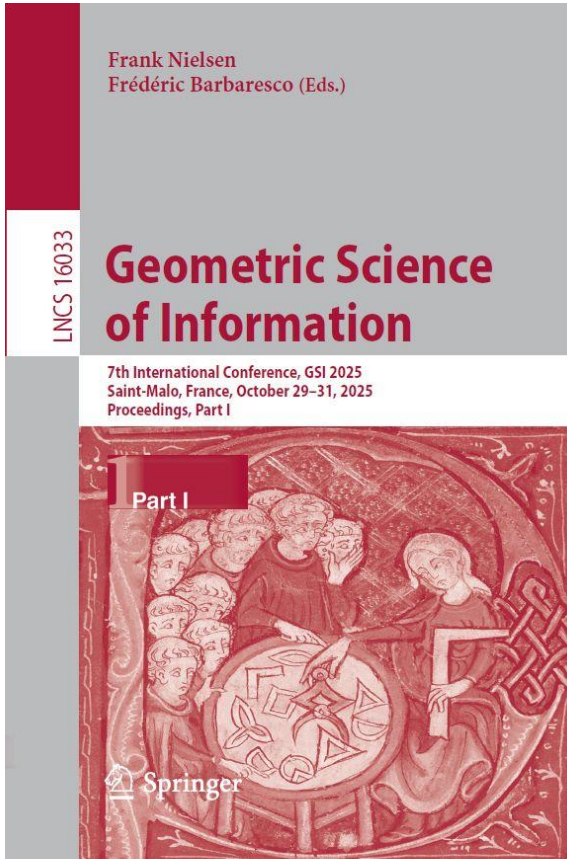
Bregman divergence

- representational Bregman divergence
- curved Bregman divergence
- total Bregman divergence
- conformal Bregman divergence
- duo pseudo-Bregman divergence
- matrix Bregman divergence
- comparative-convexity Bregman divergence
- chord Bregman divergence
- tangent Bregman divergence
- Bregman-Chernoff divergence
- Jensen-Bregman divergence
- quasi-convex Bregman divergence
- Symplectic Bregman divergence
- Symmetrized Bregman divergence

Thank you!

Geometric Science of Information conference:

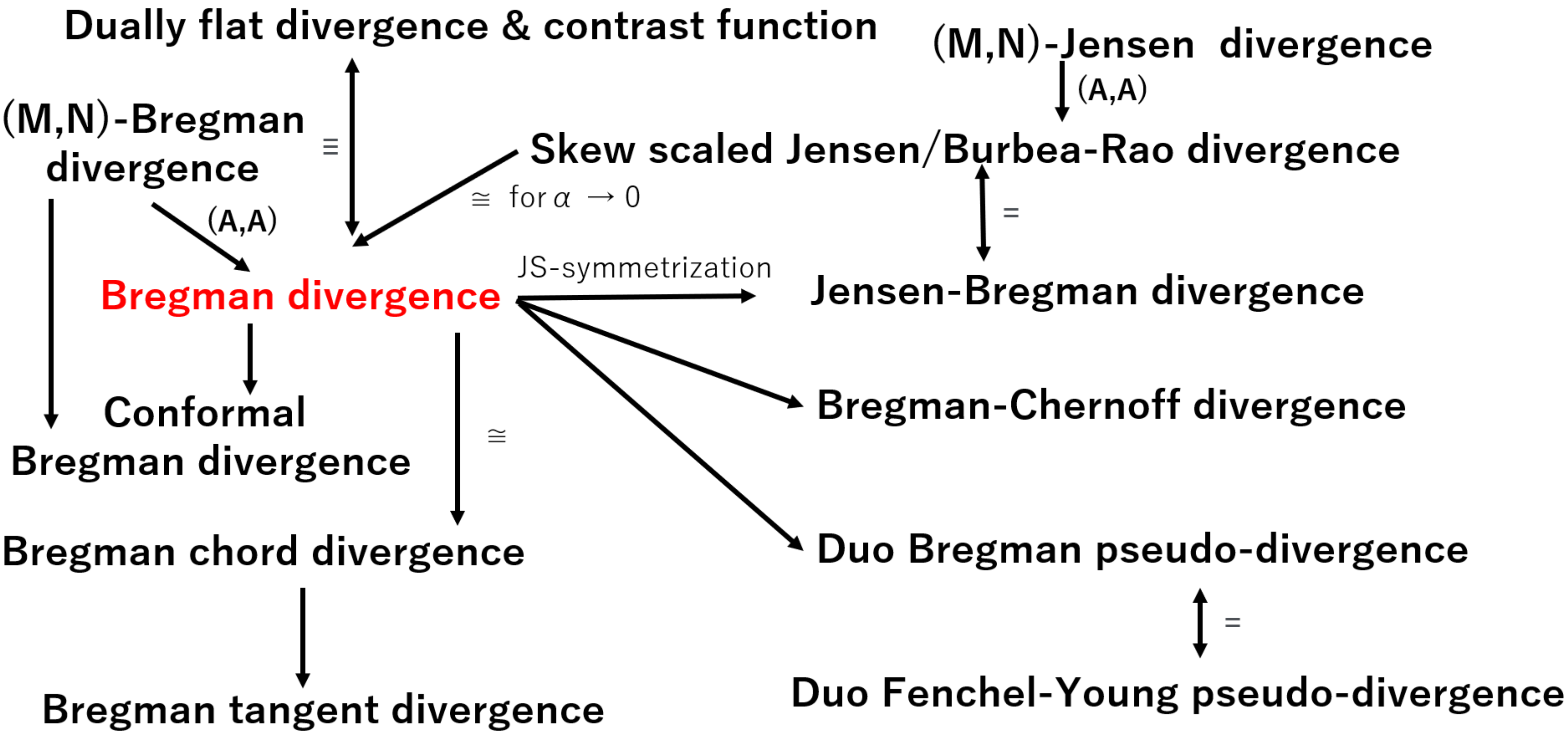




https://franknielsen.github.io/GSI/

Slides: https://franknielsen.github.io/MML25.pdf

Panorama of some generalizations of Bregman divergences



But also matrix Bregman divergence, functional Bregman divergence, etc.