Quasi-arithmetic centers, quasi-arithmetic mixtures, and the Jensen-Shannon ∇-divergences

GSI'23

Frank Nielsen

Sony Computer Science Laboratories Inc

Sony CSL

June 2023

arXiv:2301.10980

Talk outline, and contributions

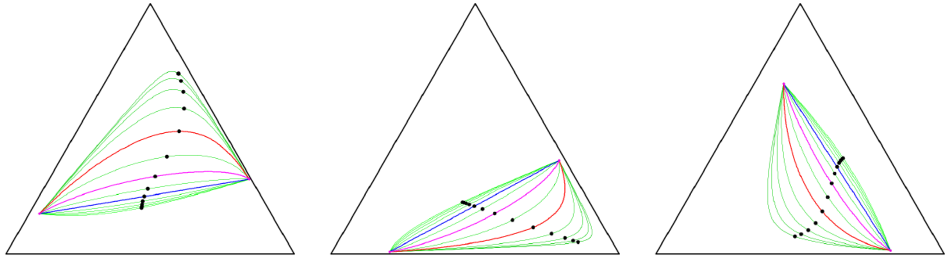
<u>Goals</u>:

I. Generalize scalar quasi-arithmetic means to multivariate cases

II. Show that the dually flat spaces of information geometry yields a natural framework for defining and studying this generalization

<u>Outline of the talk</u>:

1. Weighted quasi-arithmetic means
2. Quasi-arithmetic centers and their invariance and equivariance properties
3. Quasi-arithmetic mixtures
4. Jensen-Shannon ∇-divergences



examples of α-geodesics with midpoints in the probability simplex

Weighted quasi-arithmetic means (QAMs)

Standard (n-1)-dimensional simplex: $\Delta_{n-1} = \{(w_1, \ldots, w_n) : w_i \geq 0, \sum_i w_i = 1\}$

Definition (Weighted quasi-arithmetic mean (1930's)). *Let $f : I \subset \mathbb{R} \to \mathbb{R}$ be a strictly monotone and differentiable real-valued function. The weighted quasi-arithmetic mean (QAM) $M_f(x_1, \ldots, x_n; w)$ between n scalars $x_1, \ldots, x_n \in I \subset \mathbb{R}$ with respect to a normalized weight vector $w \in \Delta_{n-1}$, is defined by*

$$M_f(x_1, \ldots, x_n; w) := f^{-1}\left(\sum_{i=1}^{n} w_i f(x_i)\right).$$

QAMs enjoy the in-betweenness property:

$$\min\{x_1, \ldots, x_n\} \leq M_f(x_1, \ldots, x_n; w) \leq \max\{x_1, \ldots, x_n\}$$

[Kolmogorov 1930] [Nagumo 1930] [De Finetti 1931]

Quasi-arithmetic means (QAMs)

- **Classes of generators** [f]=[g] with f ≡ g yieldings the same QAM:

$$M_g(x,y) = M_f(x,y) \text{ if and only if } \boxed{g(t) = \lambda f(t) + c} \text{ for } \lambda \in \mathbb{R}\backslash\{0\}$$

- So let us fix wlog. strictly increasing and differentiable f since we can always either consider either f or -f (i.e., λ=-1, c=0).

- QAMs include **p-power means** for the smooth family of generators $f_p(t)$:

$$M_p(x,y) := M_{f_p}(x,y) \qquad f_p(t) = \begin{cases} \frac{t^p-1}{p}, & p \in \mathbb{R}\backslash\{0\}, \\ \log(t), & p = 0. \end{cases}, \quad f_p^{-1}(t) = \begin{cases} (1+tp)^{\frac{1}{p}}, & p \in \mathbb{R}\backslash\{0\}, \\ \exp(t), & p = 0. \end{cases}$$

- Pythagoras means: Harmonic (p=-1), Geometric (p=0), Arithmetic (p=1)
- **Homogeneous QAMs** $M_f(\lambda x, \lambda y) = \lambda M_f(x,y) \text{ for all } \lambda > 0$ are exactly p-power means

A generalization of the law of large numbers (LLN) and the central limit theorem (CLT)

- Quasi-arithmetic means for a strictly monotone and smooth function f(u):

$$M_f(x_1, \ldots, x_n) \quad = \quad f^{-1}(\sum_{i=1}^n f(x_i)))$$

- **Quasi-arithmetic expected value** of a random variable X:

$$\mathbb{E}_f[X] = f^{-1}(\mathbb{E}[f(X)])$$

- **Law of large numbers** for an iid random vector with variance V[X]<∞:

$$M_f(X_1, \ldots, X_n) \quad \xrightarrow{a.s.} \quad \mathbb{E}_f[X]$$

- **Central limit theorem**: $\sqrt{n}\ (M_f(X_1, \ldots, X_n) - \mathbb{E}_f[X]) \quad \xrightarrow{d} \quad N\left(0, \frac{\mathbb{V}[f(X)]}{(f'(\mathbb{E}_f[X]))^2}\right)$

Miguel De Carvalho. Mean, what do you mean? The American Statistician, 70(3):270-274, 2016.

Quasi-Arithmetic Centers (QACs) = Multivariate QAMs:

Univariate QAMs: $$M_f(x_1, \ldots, x_n; w) := f^{-1}\left(\sum_{i=1}^{n} w_i f(x_i)\right)$$

Two problems we face when going from univariate to multivariate cases:

1. Define the proper notion of *"multivariate increasing"* function F and its equivalent class of functions
2. In general, the implicit function theorem only proves locally and inverse function F^{-1} of F: $R^d \rightarrow R^d$ provided its Jacobian matrix is not singular

Information geometry provides the right framework to generalize QAMs to quasi-arithmetic centers (QACs) and study their properties.
Consider the dually flat spaces of information geometry

Legendre-type functions

$\Gamma_0(E)$: Cone of lower semi-continuous (lsc) convex functions from E into $\mathbb{R} \cup \{+\infty\}$

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

Problem: Domain H of η may not be convex... $F^* \in \Gamma_0(E)$ $F^{**} = F$

counterexample with $h(\xi_1, \xi_2) = [(\xi_1^2/\xi_2) + \xi_1^2 + \xi_2^2]/4$ [Rockafeller 1967]

To by pass this problem:

Definition Legendre type function . *(Θ, F) is of Legendre type if the function $F : \Theta \subset \mathbb{X} \to \mathbb{R}$ is strictly convex and differentiable with $\Theta \neq \emptyset$ an open convex set and*

$$\lim_{\lambda \to 0} \frac{d}{d\lambda} F(\lambda\theta + (1-\lambda)\bar{\theta}) = -\infty, \quad \forall \theta \in \Theta, \forall \bar{\theta} \in \partial\Theta. \quad (1)$$

Convex conjugate of a Legendre-type function (Θ,F(θ)) is of Legendre-type:

Given by the **Legendre function**: $F^*(\eta) = \langle \nabla F^{-1}(\eta), \eta \rangle - F(\nabla F^{-1}(\eta))$

Gradient map ∇F is globally invertible: ∇F^{-1}

Comonotone functions in inner product spaces

- **Comonotone functions**: $\forall \theta_1, \theta_2 \in \mathbb{X}, \theta_1 \neq \theta_2, \quad \langle \theta_1 - \theta_2, G(\theta_1) - G(\theta_2) \rangle > 0$

(i.e., **co**monotone = monotone with respect to the **identity function**)

Proposition (Gradient co-monotonicity). *The gradient functions $\nabla F(\theta)$ and $\nabla F^*(\eta)$ of the Legendre-type convex conjugates F and F^* in $\mathcal{F}$ are strictly increasing co-monotone functions.*

Proof using symmetrization of Bregman divergences = Jeffreys-Bregman divergence:

$$B_F(\theta_1 : \theta_2) + B_F(\theta_2 : \theta_1) = \langle \theta_2 - \theta_1, \nabla F(\theta_2) - \nabla F(\theta_1) \rangle > 0, \quad \forall \theta_1 \neq \theta_2$$

$$B_{F^*}(\eta_1 : \eta_2) + B_{F^*}(\eta_2 : \eta_1) = \langle \eta_2 - \eta_1, \nabla F^*(\eta_2) - \nabla F^*(\eta_1) \rangle > 0, \quad \forall \eta_1 \neq \eta_2$$

because Bregman divergences(and sums thereof) are always non-negative

$$B_F(\theta_1 : \theta_2) = F(\theta_1) - F(\theta_2) - \langle \theta_1 - \theta_2, \nabla F(\theta_2) \rangle \geq 0,$$

$$B_{F^*}(\eta_1 : \eta_2) = F^*(\eta_1) - F^*(\eta_2) - \langle \eta_1 - \eta_2, \nabla F^*(\eta_2) \rangle \geq 0$$

Remark: **Generalization of monotonicity** because when d=1, f(x) is strictly monotone iff $f(x_1)-f(x_2)$ is of same sign of x_1-x_2 that is, $(f(x_1)-f(x_2))(x_1-x_2)>0$

Quasi-arithmetic centers: Definition generalizing QAMs

Definition (Quasi-arithmetic centers, QACs)). *Let $F : \Theta \to \mathbb{R}$ be a strictly convex and smooth real-valued function of Legendre-type in $\mathcal{F}$. The weighted quasi-arithmetic average of $\theta_1, \ldots, \theta_n$ and $w \in \Delta_{n-1}$ is defined by the gradient map ∇F as follows:*

$$
\begin{aligned}
M_{\nabla F}(\theta_1, \ldots, \theta_n; w) &:= \nabla F^{-1}\left(\sum_i w_i \nabla F(\theta_i)\right), \\
&= \nabla F^*\left(\sum_i w_i \nabla F(\theta_i)\right),
\end{aligned}
$$

where $\nabla F^* = (\nabla F)^{-1}$ is the gradient map of the Legendre transform F^* of F.

This definition generalizes univariate quasi-arithmetic means : $M_f(x_1, \ldots, x_n; w) := f^{-1}\left(\sum_{i=1}^n w_i f(x_i)\right)$

Let $F(t) = \int_a^t f(u)\mathrm{d}u$

Then we have $M_f = M_{F'}$

An illustrating example: The matrix harmonic mean

- Consider the real-value minus **logdet function** $F(\theta) = -\log\det(\theta)$
- Domain F: $\mathrm{Sym}_{++}(d) \to \mathbb{R}$ the cone of symmetric positive-definite matrices
- Inner product: $\langle A, B\rangle := \mathrm{tr}(AB^\top)$
- We have:

$$F(\theta) = -\log\det(\theta),$$ ←Legendre-type function

$$\nabla F(\theta) = -\theta^{-1} =: \eta(\theta),$$

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

$$F^*(\eta) = \langle\theta(\eta), \eta\rangle - F(\theta(\eta)) = -d - \log\det(-\eta)$$ ←Legendre-type function

The quasi-arithmetic center with respect to F: $M_{\nabla F}(\theta_1, \theta_2) = 2(\theta_1^{-1} + \theta_2^{-1})^{-1}$

The quasi-arithmetic center with respect to F*: $M_{\nabla F^*}(\eta_1, \eta_2) = 2\left(\eta_1^{-1} + \eta_2^{-1}\right)^{-1}$

Generalize univariate harmonic mean with F(x)= log x, f(x)=F'(x)=1/x: $H(a,b) = \frac{2ab}{a+b}$ for $a, b > 0$

A Legendre-type function F gives rise to a pair of dual quasi-arithmetic centers

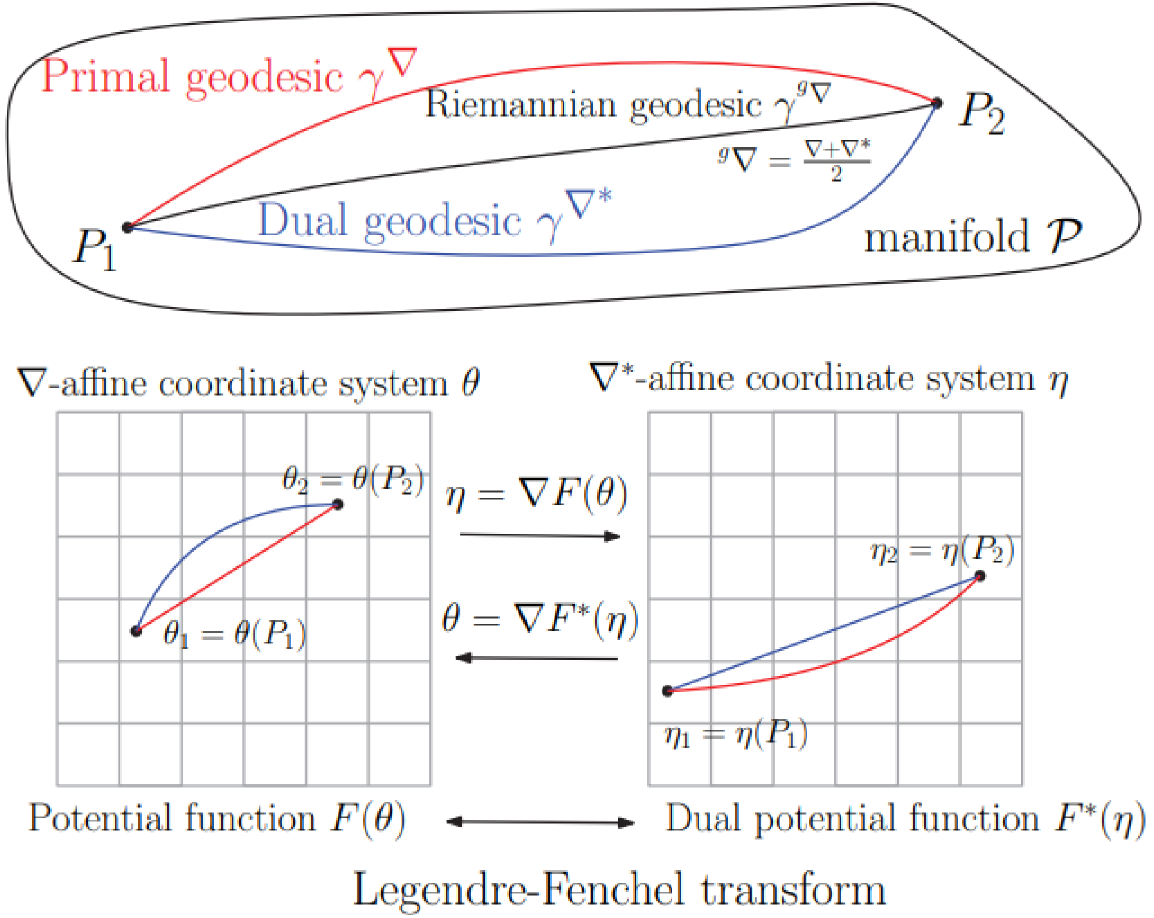
$M_{\nabla F}$ and $M_{\nabla F^*}$: dual operators

Dually flat structures of information geometry

- A Legendre-type Bregman generator F() induces a **dually flat space structure**:

$$(\Theta, g(\theta) = \nabla_\theta^2 F(\theta), \nabla, \nabla^*)$$

- A point P can be either parameterized by θ-coordinate and dual η-coordinate



[AMS 2022]

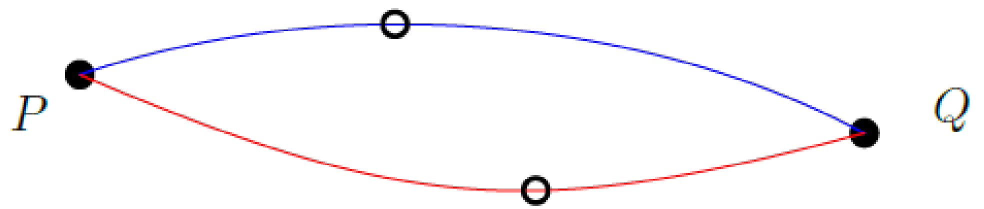
Quasi-arithmetic barycenters and dual geodesics

- The **dual geodesics** induced by the dual flat connections can be expressed using **dual weighted quasi-arithmetic centers**:

∇-geodesic $\gamma_\nabla(P,Q;t) = (PQ)^\nabla(t)$

$$(PQ)^\nabla(t) = \begin{pmatrix} M_{\mathrm{id}}(\theta(P),\theta(Q);1-t,t) \\ M_{\nabla F^*}(\eta(P),\eta(Q);1-t,t) \end{pmatrix}$$

←dual QAC $M_{\nabla F^*}$



∇^*-geodesic $\gamma_{\nabla^*}(P,Q;t) = (PQ)^{\nabla^*}(t)$

$$(PQ)^{\nabla^*}(t) = \begin{pmatrix} M_{\nabla F}(\theta(P),\theta(Q);1-t,t) \\ M_{\mathrm{id}}(\eta(P),\eta(Q);1-t,t) \end{pmatrix}$$

←primal QAC $M_{\nabla F}$

(M, g, ∇, ∇^*)

n-Variable Quasi-arithmetic centers as centroids in dually flat spaces

Consider n points $P_1, \ldots, P_n$ on the DFS (M, g, ∇, ∇^*) (canonical divergence = Bregman divergence)

Right-sided centroid:

$$\bar{C}_R = \arg\min_{P \in M} \sum_{i=1}^{n} \frac{1}{n} D_{\nabla,\nabla^*}(P_i : P)$$

$$\bar{\theta}_R = \arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} B_F(\theta_i : \theta)$$

$$\bar{\theta}_R = \theta(\bar{C}_R) = \frac{1}{n}\sum_{i=1}^{n} \theta_i = M_{\mathrm{id}}(\theta_1, \ldots, \theta_n)$$

$$\bar{\eta}_R = \nabla F(\bar{\theta}_R) = M_{\nabla F^*}(\eta_1, \ldots, \eta_n).$$ ← dual QAC

Reference duality

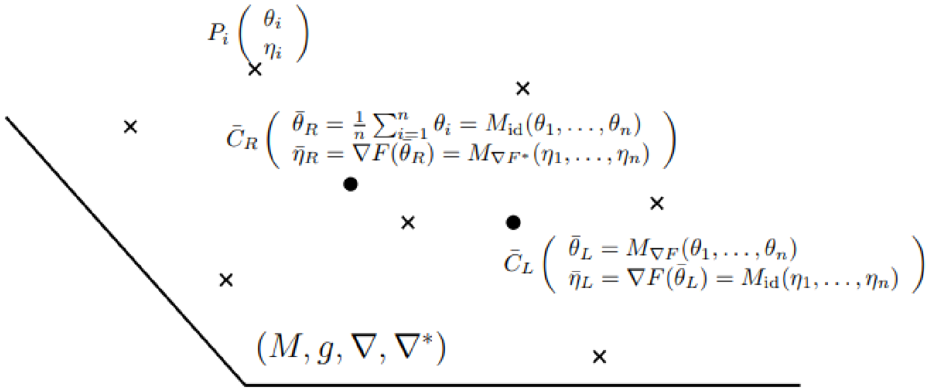
Left-sided centroid:

$$\bar{C}_L = \arg\min_{P \in M} \sum_{i=1}^{n} \frac{1}{n} D_{\nabla,\nabla^*}(P : P_i)$$

$$\bar{\theta}_L = \arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} B_F(\theta : \theta_i)$$

$$\bar{\theta}_L = M_{\nabla F}(\theta_1, \ldots, \theta_n),$$ ← primal QAC

$$\bar{\eta}_L = \nabla F(\bar{\theta}_L) = M_{\mathrm{id}}(\eta_1, \ldots, \eta_n)$$



Notice that when n=2, weighted dual quasi-arithmetic barycenters define the dual geodesics

Invariance/equivariance of quasi-arithmetic centers

Information geometry is well-suited to study the **properties of QACs**:

A dually flat space (DFS) can be **realized** by *a class of Bregman generators*:

$$(M, g, \nabla, \nabla^*) \leftarrow \mathrm{DFS}([\theta, F(\theta); \eta, F^*(\eta)])$$

Affine Legendre invariance of dually flat spaces:

- By adding an affine term...

Same DFS with $\bar{F}(\theta) = F(\theta) + \langle c, \theta \rangle + d.$

Invariance of quasi-arithmetic center:

$$M_{\nabla \bar{F}}(\theta_1, \ldots; \theta_n; w) = M_{\nabla F}(\theta_1, \ldots; \theta_n; w)$$

- By an affine change of coordinate...

Same DFS with $\bar{\theta} = A\theta + b$ such that $\bar{F}(\bar{\theta}) = F(\theta)$

$$\nabla \bar{F}(x) = (A^{-1})^\top \nabla F(A^{-1}(x - b)).$$

Equivariance of quasi-arithmetic center:

$$M_{\nabla \bar{F}}(\bar{\theta}_1, \ldots, \bar{\theta}_n; w) = A\, M_{\nabla F}(\theta_1, \ldots, \theta_n; w) + b$$

$$B_{\bar{F}(\overline{\theta_1} : \overline{\theta_2})} = B_F(\theta_1 : \theta_2)$$

Same canonical divergence of the DFS
(= constrast function on the diagonal of the product manifold)

Canonical divergence versus Legendre-Fenchel/Bregman divergences

- Canonical divergence induced by dual flat connections is between **points**
- dual Bregman divergences B_F and B_{F^*} between **dual coordinates**
- Legendre-Fenchel divergence Y_F between **mixed coordinates**

$$F(\theta) + F^*(\eta) - \langle \theta, \eta \rangle = 0 \qquad \eta = \nabla F(\theta)$$

$$\begin{aligned} B_F(\theta_1 : \theta_2) &:= F(\theta_1) - \underbrace{F(\theta_2)}_{=\langle \theta_2, \eta_2 \rangle - F^*(\eta_2)} - \langle \theta_1 - \theta_2, \nabla F(\eta_2) \rangle \\ &= F(\theta_1) + F^*(\eta_2) - \langle \theta_1, \eta_2 \rangle =: Y_F(\theta_1 : \eta_2) \end{aligned}$$

$$\begin{aligned} (M, g, \nabla, \nabla^*) &\leftarrow \mathrm{DFS}([\Theta, F(\theta), H, F^*(\eta)]) \\ &\leftarrow \mathrm{DFS}([\bar{\Theta}, \bar{F}(\bar{\theta}), \bar{H}, \bar{F}^*(\bar{\eta})]) \end{aligned}$$

$$\begin{aligned} D_{\nabla, \nabla^*}(P_1 : P_2) &= B_F(\theta_1 : \theta_2) = B_{F^*}(\eta_1, \eta_2) = Y_F(\theta_1 : \eta_2) = Y_{F^*}(\eta_2 : \theta_1) \\ &= B_{\bar{F}}(\overline{\theta_1} : \overline{\theta_2}) = B_{\bar{F}^*}(\overline{\eta_1}, \overline{\eta_2}) = Y_F(\overline{\theta_1} : \overline{\eta_2}) = Y_{F^*}(\overline{\eta_2} : \overline{\theta_1}) \end{aligned}$$

Affine Legendre invariance of dually flat spaces plus setting the unit scale of divergences

- Affine Legendre invariance:

$$\bar{F}(\bar{\theta}) = F(A\theta + b) + \langle c, \theta \rangle + d$$
$$\bar{F}^*(\bar{\eta}) = F^*(A^*\eta + b^*) + \langle c^*, \eta \rangle + d^*$$

- Set the unit scale of canonical divergence (DFS differ here, rescaled):

(does not change the quasi-arithmetic center) $D_{\lambda,\nabla,\nabla^*} := \lambda D_{\nabla,\nabla^*}$

amount to scale the potential function λF(θ) vs F(θ)

Proposition (Invariance and equivariance of QACs). *Let $F(\theta)$ be a function of Legendre type. Then $\bar{F}(\bar{\theta}) := \lambda(F(A\theta+b)+\langle c,\theta\rangle+d)$ for $A \in \mathrm{GL}(d)$, $b, c \in \mathbb{R}^d$, $d \in \mathbb{R}^d$ and $\lambda \in \mathbb{R}_{>0}$ is a Legendre-type function, and we have*

$$M_{\nabla\bar{F}} = A\, M_{\nabla F} + b.$$

Illustrating example: Mahalanobis divergence

- **Mahalanobis divergence** = squared Mahalanobis metric distance

$$\Delta^2(\theta_1,\theta_2) = B_{F_Q}(\theta_1 : \theta_2) = \frac{1}{2}(\theta_2-\theta_1)^\top Q\,(\theta_2-\theta_1)$$

fails triangle inequality of metric distances

Primal potential function: $F_Q(\theta) = \frac{1}{2}\theta^\top Q\theta + c\theta + \kappa$

Dual potential function: $F^*(\eta) = \frac{1}{2}\eta^\top Q^{-1}\eta = F_{Q^{-1}}(\eta),$

- The dual QACs induced by the dual Mahalanobis generators F and F* coincide to weighted arithmetic mean M_{id}:

$$M_{\nabla F_Q}(\theta_1,\ldots,\theta_n;w) = Q^{-1}\left(\sum_{i=1}^n w_i Q\theta_i\right) = \sum_{i=1}^n w_i\theta_i = M_{\mathrm{id}}(\theta_1,\ldots,\theta_n;w),$$

$$M_{\nabla F_Q^*}(\eta_1,\ldots,\eta_n;w) = Q\left(\sum_{i=1}^n w_i Q^{-1}\eta_i\right) = M_{\mathrm{id}}(\eta_1,\ldots,\eta_n;w).$$

Quasi-arithmetic mixtures (QAMixs), and α-mixtures

Definition . *The M_f-mixture of n densities $p_1, \ldots, p_n$ weighted by $w \in \Delta_n^\circ$ is defined by*

$$(p_1, \ldots, p_n; w)^{M_f}(x) := \frac{M_f(p_1(x), \ldots, p_n(x); w)}{\int M_f(p_1(x), \ldots, p_n(x); w) \mathrm{d}\mu(x)}.$$

Centroid of n densities with respect to the α-divergences yields a QAMix:

$$(p_1, \ldots, p_n; w)^{M_\alpha} = \arg\min_p \sum_i w_i D_\alpha(p_i, p)$$

D_α denotes the α-divergences:

$$D_\alpha[m(s) : l(s)] = \begin{cases} \int m(s) ds - \int l(s) ds + \int m(s) \log \frac{m(s)}{l(s)} ds & \alpha = -1 \\ \int l(s) ds - \int m(s) ds + \int l(s) \log \frac{l(s)}{m(s)} ds + \int l(s) \log \frac{l(s)}{m(s)} ds & \alpha = 1 \\ \frac{2}{1+\alpha} \int m(s) ds + \frac{2}{1-\alpha} \int l(s) ds - \frac{4}{1-\alpha^2} \int m(s)^{\frac{1-\alpha}{2}} l(s)^{\frac{1+\alpha}{2}} ds, & \alpha \neq \pm 1. \end{cases}$$

α-families of probability distributions [Amari 2007]

[arXiv:2209.07481]

k=2 QAMixs and the ∇-Jensen-Shannon divergence

- **Jensen-Shannon divergence** is bounded symmetrization of KL divergence:

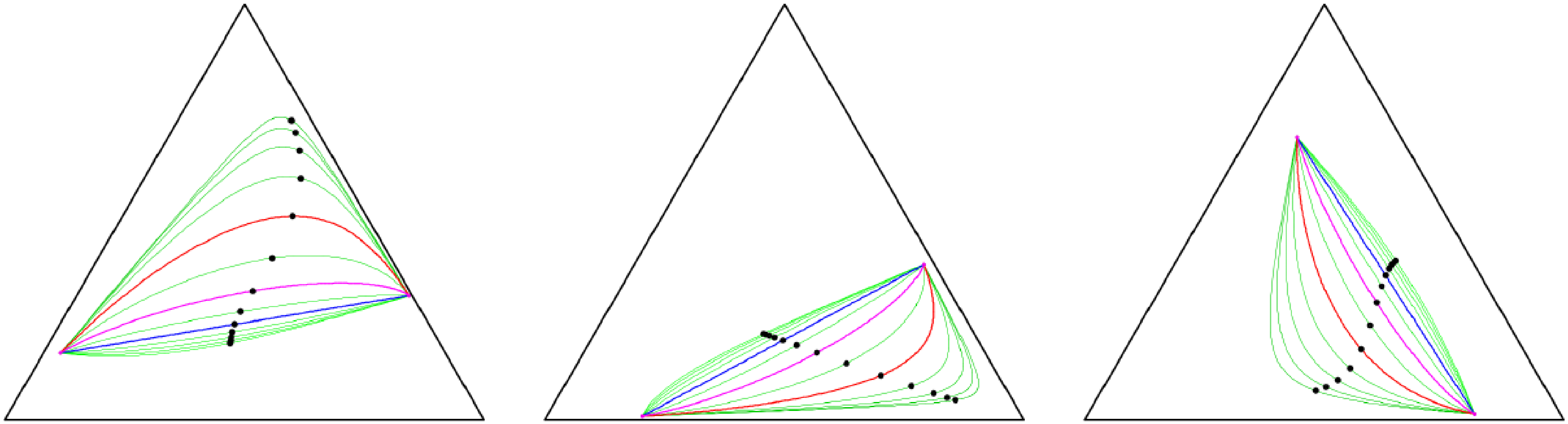
$$D_{\mathrm{JS}}(p,q) = \frac{1}{2}\left(D_{\mathrm{KL}}\left(p : \frac{p+q}{2}\right) + D_{\mathrm{KL}}\left(q : \frac{p+q}{2}\right)\right) \leq \log(2)$$

- Interpret arithmetic mixture as the **midpoint of a mixture geodesic** (wrt to the flat non-parametric mixture connection ∇^m in information geometry).
- Generalize Jensen-Shannon divergence with **arbitrary ∇-connections**:

Definition (Affine connection-based ∇-Jensen-Shannon divergence). *Let ∇ be an affine connection on the space of densities $\mathcal{P}$, and $\gamma_\nabla(p,q;t)$ the geodesic linking density $p = \gamma_\nabla(p,q;0)$ to density $q = \gamma_\nabla(p,q;1)$. Then the ∇-Jensen-Shannon divergence is defined by:*

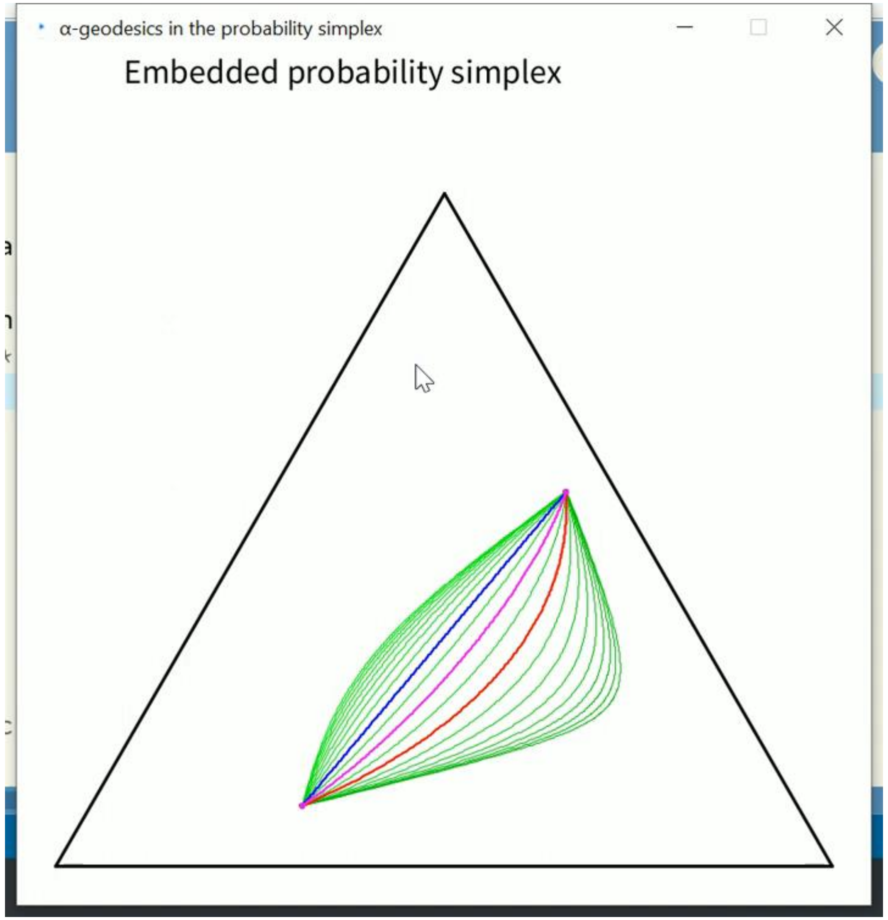
$$D_\nabla^{\mathrm{JS}}(p,q) := \frac{1}{2}\left(D_{\mathrm{KL}}\left(p : \gamma_\nabla\left(p,q;\frac{1}{2}\right)\right) + D_{\mathrm{KL}}\left(q : \gamma_\nabla\left(p,q;\frac{1}{2}\right)\right)\right).$$

∇$^{\alpha}$-connections and geodesics in the probability simplex, ∇$^{\alpha}$-Jensen-Shannon divergence

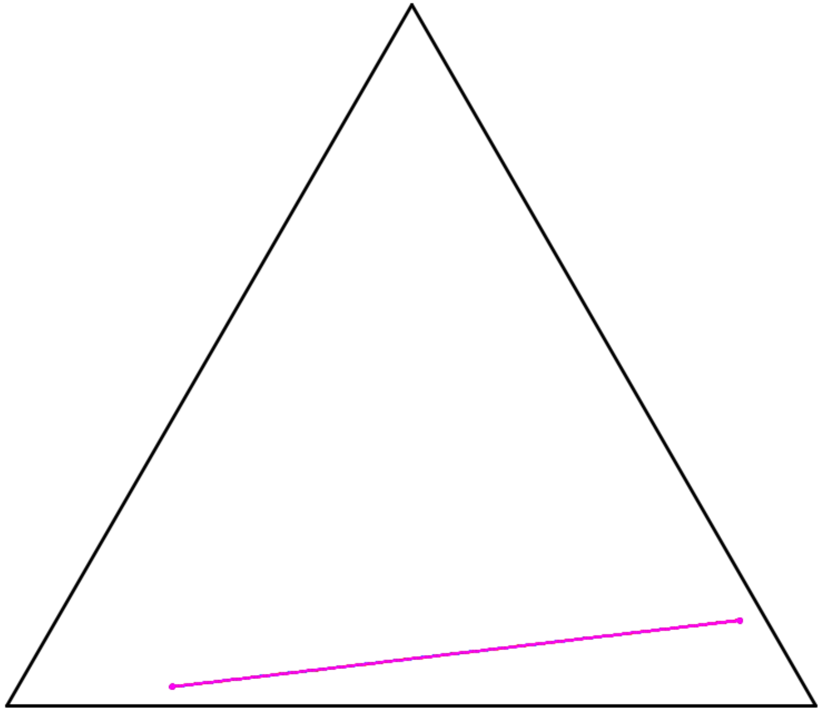


$$D^{\mathrm{JS}}_{\nabla^\alpha}(p,q) = \frac{1}{2}\left(D_{\mathrm{KL}}\left(p : \gamma_{\nabla^\alpha}\left(p,q;\frac{1}{2}\right)\right) + D_{\mathrm{KL}}\left(q : \gamma_{\nabla^\alpha}\left(p,q;\frac{1}{2}\right)\right)\right)$$

α-geodesics coincide when they pass through a standard simplex vertex



non-degenerate



degenerate

grateful for fruitful discussions with Fábio Meneghetti and Sueli Costa

Inductive Means: Geodesics/quasi-arithmetic centers

- Gauss and Lagrange independently studied the following convergence of pairs of iterations:

$$\begin{aligned} a_{t+1} &= \frac{a_t + b_t}{2} \\ b_{t+1} &= \sqrt{a_t b_t} \end{aligned}$$

and proves quadratic convergence to the **arithmetic-geometric mean AGM**

$$\mathrm{AGM}(a_0, b_0) = \frac{\pi}{4} \frac{a_0 + b_0}{K\left(\frac{a_0 - b_0}{a_0 + b_0}\right)}$$

where K is complete elliptic integral of the first kind
AGM also used to approximate ellipse perimeter and π

- In general, choosing two strict means M and M' with interness property will converge but difficult to *analytically express the common limits of iterations*
- When M=Arithmetic and M'=Harmonic, the **arithmetic-harmonic mean AHM** yields the geometric mean:

$$\begin{aligned} a_{t+1} &= A(a_t, h_t) \\ h_{t+1} &= H(a_t, h_t) \end{aligned}$$

$$\mathrm{AHM}(x, y) = \lim_{t \to \infty} a_t = \lim_{t \to \infty} h_t = \sqrt{xy} = G(x, y)$$

Inductive matrix arithmetic-harmonic mean

- Consider the cone of symmetric positive-definite matrices (SPD cone), and extend the AHM to SPD matrices:

$$A_{t+1} = \frac{A_t + H_t}{2} = A(A_t, H_t)$$ ←arithmetic mean

$$H_{t+1} = 2\left(A_t^{-1} + H_t^{-1}\right)^{-1} = H(A_t, H_t)$$ ←harmonic mean

- Then the sequences converge quadratically to the **matrix geometric mean**:

$$\mathrm{AHM}(X,Y) = \lim_{t\to+\infty} A_t = \lim_{t\to+\infty} H_t.$$

$$\boxed{\mathrm{AHM}(X,Y) = X^{\frac{1}{2}}\left(X^{-\frac{1}{2}}\,Y\,X^{-\frac{1}{2}}\right)^{\frac{1}{2}}X^{\frac{1}{2}} = G(X,Y)}$$

which is also the **Riemannian center of mass** with respect to the trace metric:

$$G(X,Y) = \arg\min_{M\in\mathbb{P}(d)} \frac{1}{2}\rho^2(X,M) + \frac{1}{2}\rho^2(Y,M).$$

$$\rho(P_1,P_2) = \sqrt{\sum_{i=1}^{d} \log^2 \lambda_i\left(P_1^{-\frac{1}{2}}\,P_2\,P_1^{-\frac{1}{2}}\right)}$$ Riemannian distance

$$g_P(V_1,V_2) = \mathrm{tr}\left(P^{-1}V_1P^{-1}V_2\right)$$

[Nakamura 2001, Atteia-Raissouli 2001]

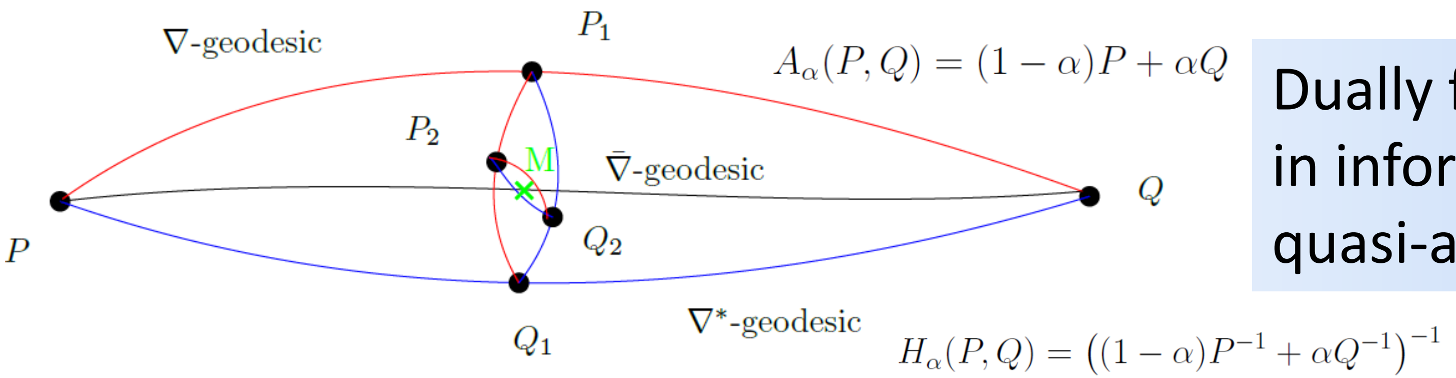
Geometric interpretation of the AHM matrix mean

$$A_{t+1} = \frac{A_t + H_t}{2} = A(A_t, H_t)$$
$$H_{t+1} = 2\left(A_t^{-1} + H_t^{-1}\right)^{-1} = H(A_t, H_t)$$

$$P_{t+1} = \gamma\left(P_t, Q_t : \frac{1}{2}\right)$$
$$Q_{t+1} = \gamma^*\left(P_t, Q_t : \frac{1}{2}\right)$$

(SPD, g^G, ∇^A, ∇^H) is a dually flat space, ∇^G is Levi-Civita connection

$$G_\alpha(P,Q) = P^{\frac{1}{2}}\left(P^{-\frac{1}{2}}\, Q\, P^{-\frac{1}{2}}\right)^\alpha P^{\frac{1}{2}}$$



Dually flat space (SPD, g^G, ∇^A, ∇^H)
in information geometry defines
quasi-arithmetic centers as geodesic midpoints

Primal geodesic midpoint is the arithmetic center wrt Euclidean metric $g_P^A(X,Y) = \mathrm{tr}(X^\top Y)$

Dual geodesic midpoint = harmonic center wrt an isometric Eucl. metric $g_P^H(X,Y) = \mathrm{tr}(P^{-2} X P^{-2} Y)$

Levi-Civita geodesic midpoint is geometric Karcher mean (not QAC) $g_P^G(X,Y) = \mathrm{tr}(P^{-1} X P^{-1} Y)$

[Nakamura 2001]

Summary: Beyond scalar quasi-arithmetic means

Information geometry of dually flat spaces yields a generalization of quasi-arithmetic means:

$$M_f(x_1,\ldots,x_n;w) := f^{-1}\left(\sum_{i=1}^n w_i f(x_i)\right)$$

- 1d monotone function generalize to gradient map of a Legendre-type multivatiate function (comonotone)

$$M_{\nabla F}(\theta_1,\ldots,\theta_n;w) := \nabla F^{-1}\left(\sum_i w_i \nabla F(\theta_i)\right)$$
$$= \nabla F^{*}\left(\sum_i w_i \nabla F(\theta_i)\right)$$

dual quasi-arithmetic centers induced by a Legendre-type function

Applications of QACs:

- dual centers of mass of n≥2 points expressed using weighted quasi-arithmetic centers
- dual geodesics expressed in coordinate systems as weighted quasi-arithmetic centers (n=2)
- invariance/equivariance analyzed from the viewpoint of information geometry

$$\bar{F}(\bar{\theta}) := \lambda(F(A\theta+b)+\langle c,\theta\rangle+d) \Longrightarrow M_{\nabla\bar{F}} = A\,M_{\nabla F}+b.$$

- define quasi-arithmetic mixtures which provides a way to integrate density components
- define ∇-Jensen-Shannon divergences
- Inductive arithmetic-harmonic geometric matrix mean expressed using QACs

Some references

- Amari, Shun-ichi: Information Geometry and Its Applications. Applied Mathematical Sciences, Springer Japan (2016)
- Masrani, V., Brekelmans, R., Bui, T., Nielsen, F., Galstyan, A., Ver Steeg, G., Wood, F.: q-paths: Generalizing the geometric annealing path using power means. In: Uncertainty in Articial Intelligence. pp. 1938-1947. PMLR (2021)
- Nakamura, Y.: Algorithms associated with arithmetic, geometric and harmonic means and integrable systems. Journal of computational and applied mathematics 131(1-2), 161174 (2001)
- Rockafellar, R.T.: Conjugates and Legendre transforms of convex functions. Canadian Journal of Mathematics 19, 200205 (1967)
- Zhang, J.: Nonparametric information geometry: From divergence function to referential-representational biduality on statistical manifolds. Entropy 15(12), 5384-5418 (2013)
- Nielsen, Frank. "The many faces of information geometry." *Not. Am. Math. Soc* 69.1 (2022): 36-45.
- Atteia, Marc, and Mustapha Raïssouli. "Self dual operators on convex functionals: Geometric mean and square root of convex functionals." *Journal of Convex Analysis* 8.1 (2001): 223-240.
- Ben-Tal, A., Charnes, A., Teboulle, M.: Entropic means. Journal of Mathematical Analysis and Applications 139(2), 537551 (1989)
- Amari, Shun-ichi: Integration of stochastic models by minimizing -divergence. Neural computation 19(10), 27802796 (2007)
- Nielsen, F.: On the Jensen-Shannon symmetrization of distances relying on abstract means. Entropy 21(5), 485 (2019)