A divergence center interpretation of general Kubo-Ando means

József Pitrik
Wigner Research Centre for Physics
and
Budapest University of Technology and Economics
Joint work with Dániel Virosztek, Erdős group, IST Austria

Trojan Math Seminar
16 September 2021

Notations

- $\mathcal{H}$ (finite) dimensional complex Hilbert-space
- $\mathcal{B}(\mathcal{H})$ linear operators on $\mathcal{H}$
- $\mathcal{B}(\mathcal{H})^{sa}$ self-adjoint operators on $\mathcal{H}$
- $\mathcal{B}(\mathcal{H})^{+}$ positive semi-definite operators on $\mathcal{H}$
- $\mathcal{B}(\mathcal{H})^{++}$ positive definite (and so invertible) operators on $\mathcal{H}$
- $\langle A \mid B \rangle = \operatorname{Tr} A^* B$ Hilbert-Schmidt inner product of $A, B \in \mathcal{B}(\mathcal{H})$
- $\|A\|_2 = (\operatorname{Tr} A^* A)^{1/2}$ Hilbert-Scmidt (Schatten-2) norm of $A \in \mathcal{B}(\mathcal{H})$
- $\mathbf{D}$ and $\mathbf{D}^2$ denote the first and second Fréchet derivatives, respectively

We consider the Löwner order induced by positivity on $\mathcal{B}(\mathcal{H})^{sa}$, that is, by $A \leq B$ we mean that $B - A$ is positive semi-definite.

Operator (matrix) means in Kubo-Ando sense

A binary operation $\sigma : \mathcal{B}(\mathcal{H})^+ \times \mathcal{B}(\mathcal{H})^+ \to \mathcal{B}(\mathcal{H})^{++}$ is called an **operator connection**, if it satisfies for $A, B, C, D \in \mathcal{B}(\mathcal{H})^+$:

1. $A \leq B$ and $B \leq D$ imply $A\sigma B \leq C\sigma D$ (joint monotonicity)
2. $C(A\sigma B)C \leq (CAC)\sigma(CBC)$ (transformer inequlality)
3. $A_n, B_n \in \mathcal{B}(\mathcal{H})^+$, $A_n \searrow A$, $B_n \searrow B$ imply $A_n\sigma B_n \searrow A\sigma B$ (downward continuity).
   (here $A_n \searrow A$ means that $A_1 \geq A_2 \geq \ldots$ and $A_n \longrightarrow A$ in strong operator topology.)

   An operator connection σ is called an **operator mean (Kubo-Ando mean)** if

4. $I\sigma I = I$, where I is the identity in $\mathcal{B}(\mathcal{H})$.

An operator mean is **symmetric** if $A\sigma B = B\sigma A$.

Kubo-Ando Theorem [1]

For each operator connection σ there exist a unique operator monotone function $f_\sigma : [0, \infty) \to [0, \infty)$, s.t.

$$f_\sigma(t)I = I\sigma(tI), \quad t \geq 0.$$

Furthermore,

- The map $\sigma \mapsto f_\sigma$ is an affine order-isomporphism between the operator connections and the operator monotone functions $f_\sigma : [0, \infty) \to [0, \infty)$.
  (i.e. when $\sigma_i \mapsto f_i$ for $i = 1, 2$, then $A\sigma_1 B \leq A\sigma_2 B$ for all $A, B \in \mathcal{B}(\mathcal{H})^+$ iff $f_1(t) \leq f_2(t)$, for all $t \geq 0$.)

[1] T. Ando, F. Kubo, *Means of positive linear operators*, Math. Ann. **246** (1980), 205–224.

Kubo-Ando Theorem

- If A is invertible, then

$$A\sigma B = A^{1/2} f_\sigma(A^{-1/2} B A^{-1/2}) A^{1/2}.$$

- σ is an operator mean if and only if $f_\sigma(1) = 1$.
  In this case, $A\sigma A = A$, for all A.
- σ is a symmetric operator mean if and only if $f_\sigma(1) = 1$ and $f_\sigma(t) = t f_\sigma(1/t)$, for $t > 0$.

Some well known operator mean

$A, B \in \mathcal{B}(\mathcal{H})^{++}$, $\alpha \in [0, 1]$

- **Weighted arithmetic mean**

$$A \nabla_\alpha B = (1 - \alpha)A + \alpha B$$

Representing function:

$$f_{\nabla_\alpha}(t) = (1 - \alpha) + \alpha t$$

In particular for $\alpha = 1/2$:

$$A \nabla B = (A + B)/2$$

arithmetic mean (symmetric)
Generalization for the positive operators A_j, $j = 1, 2, \ldots, m$:

$$\frac{1}{m} \sum_{j=1}^{m} A_j$$

- **Weighted geometric mean**

$$A\#_\alpha B = A^{1/2}(A^{-1/2}BA^{-1/2})^\alpha A^{1/2}$$

Representing function:

$$f_{\#_\alpha}(t) = t^\alpha, \quad (t > 0)$$

In particular for $\alpha = 1/2$:

$$A\#B = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}$$

geometric mean (symmetric)
If A and B commutes, then $A\#B = (AB)^{1/2}$.
Generalization for $m > 2$ positive operators?

- **Weighted harmonic mean**

$$A!_\alpha B = \left((1-\alpha)A^{-1} + \alpha B^{-1}\right)^{-1}$$

Representing function:

$$f_{!_\alpha}(t) = \frac{t}{(1-\alpha)t + \alpha}$$

In particular for $\alpha = 1/2$:

$$A!B = 2\left(A^{-1} + B^{-1}\right)^{-1}$$

harmonic mean (symmetric)
Generalization for the positive operators A_j, $j = 1, 2, \ldots, m$:

$$m(\sum_{j=1}^{m} A_j^{-1})^{-1}.$$

For $t > 0$

$$\frac{t}{(1-\alpha)t+\alpha} \leq t^{\alpha} \leq (1-\alpha)t + \alpha t$$

holds, which implies thanks to the Kubo-Ando Theorem that

$$A!_{\alpha}B \leq A\#_{\alpha}B \leq A\nabla_{\alpha}B.$$

Furthermore, for an arbitrary operator mean σ with the representing function f_{σ}

$$\frac{t}{(1-\alpha)t+\alpha} \leq f_{\sigma} \leq (1-\alpha)t + \alpha t$$

which implies

$$A!_{\alpha}B \leq A\sigma B \leq A\nabla_{\alpha}B.$$

Barycenters

- motivation from statistics: we perform an uncertain measurement several times with outcomes in a metric space (X, d)
- the most natural estimator of the quantity a we are interested in is the *mean squared error estimator*

$$\hat{a} := \arg\min_{x \in X} \frac{1}{m} \sum_{j=1}^{m} d^2\left(a_j, x\right),$$

where a_j's are the outcomes
- slightly more generally,

$$\hat{a} := \arg\min_{x \in X} \sum_{j=1}^{m} w_j d^2\left(a_j, x\right),$$

where the w_j's are arbitrary weights (not necessarily relative frequencies)

- if $(X, d) = (\mathbb{R}^n, \|\cdot\|)$, then the barycenter is the weighted average,

$$\arg\min_{x \in X} \sum_{j=1}^{m} w_j d^2(a_j, x) = \sum_{j=1}^{m} w_j a_j$$

- sometimes one should consider "squared distance-like" quantities instead of the square of a genuine metric
- a prominent example is the (classical) relative entropy on probability vectors,

$$H(\mathbf{p}, \mathbf{q}) = \sum_{k=1}^{n} p_k (\log p_k - \log q_k),$$

where $0 < p_1, \ldots, p_n, q_1, \ldots, q_n < 1$ and $\sum_{k=1}^{n} p_k = \sum_{k=1}^{n} q_k = 1$

- in this case, we have similar result:

$$\arg\min_{\mathbf{q} \in \mathcal{P}_n} \sum_{j=1}^{m} w_j H(\mathbf{p}_j, \mathbf{q}) = \sum_{j=1}^{m} w_j \mathbf{p}_j$$

- more generally, if $\varphi : (0,1) \to \mathbb{R}$ is a strictly convex C^1 function, and

$$H_\varphi(\mathbf{p}, \mathbf{q}) = \sum_{k=1}^{n} \varphi(p_k) - \varphi(q_k) - \varphi'(q_k)(p_k - q_k)$$

is the associated Bregman divergence, then again,[2]

$$\arg\min_{\mathbf{q} \in \mathcal{P}_n} \sum_{j=1}^{m} w_j H_\varphi(\mathbf{p}_j, \mathbf{q}) = \sum_{j=1}^{m} w_j \mathbf{p}_j,$$

no matter what φ is

- the classical relative entropy corresponds to $\varphi(x) = x \log x - x$

[2] I. S. Dhillon and J. A. Tropp, Matrix nearness problems with Bregman divergences, SIAM J. Matrix Anal. Appl. **29** (2004), 1120-1146, and A. Banerjee, S. Merugu, I. S. Dhillon and J. Ghosh, Clustering with Bregman divergences, J. Mach. Learn. Res. **6** (2005), 1705-1749

The divergence interpretation of the arithmetic mean

- The arithmetic mean $A \nabla B = (A+B)/2$ is the mean squared estimator for the Euclidean metric on positive operators:

$$A \nabla B = \arg\min_{X>0} \frac{1}{2}(\mathrm{Tr}\,(A-X)^2 + \mathrm{Tr}\,(B-X)^2).$$

- Let $\varphi : \mathbb{R}^+ \to \mathbb{R}$ be a differentiable strictly convex function and

$$\Phi(x,y) = \varphi(x) - \varphi(y) - \varphi'(y)(x-y)$$

be the associated **Bregman divergence**. Then[3] for the positive operators A_j

$$\arg\min_{X>0} \sum_{j=1}^{m} \frac{1}{m}\Phi(A_j, X) = \sum_{j=1}^{m} \frac{1}{m} A_j$$

holds, independently of φ.

[3] R. Bhatia, S. Gaubert, T. Jain, *Matrix versions of the Hellinger distance,* Lett. Math. Phys.,vol 109, 1777–1804 (2019)

The Riemannian trace metric (RTM)

- the *Boltzmann entropy* (or H-functional) of a random variable X with probability density f is given by

$$H(X) = -\int_{\mathrm{supp}(X)} f(x) \log f(x) \mathrm{d}x$$

- this is a particularly important functional; for instance, the heat equation

$$\partial_t u = \Delta u$$

can be seen as the gradient flow for the Boltzmann entropy as potential (or "energy") in the differential structure induced by optimal transportation[4]

[4] C. Villani, *Topics in optimal transportation,* Graduate studies in Mathematics vol. 58, American Mathemtical Society, Providence, RI, 2003, Sec. 8.3.

- centered multivariate Gaussians on $\mathbb{R}^n$ are completely described by their positive definite covariance matrix A; the probability density is given by

$$f_{\mathcal{N}(0,A)}(x) = \frac{\exp\left(-\frac{1}{2}x^* A^{-1} x\right)}{\sqrt{(2\pi)^N \mathrm{det} A}}$$

- the Boltzmann entropy of $X \sim \mathcal{N}(0, A)$ is

$$H(X) = \frac{1}{2}\left((\log(2\pi) + 1)N + \mathrm{Tr} \log A\right) = \frac{1}{2}\mathrm{Tr} \log A + C(N)$$

  (Remember, that $\log \mathrm{det} A = \mathrm{Tr} \log A$.)
- so H is a convex functional on non-degenerate centered Gaussians on $\mathbb{R}^n$
- for the sake of simplicity, we will identify these Gaussians with their covariance ($\mathcal{N}(0, A) \longrightarrow A$), and forget the prefactor $1/2$ and the constant $C(n)$

Let $\dim\mathcal{H} = n$. The set $\mathcal{P}_n := \mathcal{B}(\mathcal{H})^{++}$ of positive definite $n \times n$ matrices can be considered as an open subset af the Euclidean space $\mathbb{R}^{n^2}$ and they form a manifold.

- the Boltzmann entropy gives rise to a Riemannian metric by its Hessian
- $H(A) = \mathrm{Tr}\log A$
- $\mathbf{D}H(A)[X] = \mathrm{Tr}A^{-1}X$
- $\mathbf{D}^2H(A)[Y,X] = \mathrm{Tr}A^{-1}YA^{-1}X$
- this is a collection of positive definite bilinear forms on the tangent spaces $T_A\mathcal{P}_n(\mathbb{R}) \simeq M_n^{sa}(\mathbb{R})$ that depends smoothly on the foot point A, and is therefore a Riemannian tensor field
- the metric induced by the Riemannian tensor field

$$g_A(X, Y) := \mathrm{Tr}A^{-1}YA^{-1}X$$

is often called **Riemannian trace metric (RTM)**

- When $\gamma : [0,1] \to \mathcal{P}_n$ is a C^1 curve, the lenght of γ with respect to RTM:

$$L(\gamma) = \int_0^1 \sqrt{g_{\gamma(t)}\left(\gamma'(t), \gamma'(t)\right)} \mathrm{d}t = \int_0^1 \|\gamma(t)^{-1/2}\gamma'(t)\gamma(t)^{-1/2}\|_2 \mathrm{d}t.$$

- the global distance (the RTM) is obtained from the Riemannian structure by a variational formula:

$$d_{RTM}(A, B)$$

$$= \inf \left\{ \int_0^1 \sqrt{g_{\gamma(t)}\left(\gamma'(t), \gamma'(t)\right)} \mathrm{d}t \,\middle|\, \gamma : [0,1] \to \mathcal{P}_n,\ \gamma(0) = A, \gamma(1) = B \right\}$$

- the curve γ with minimal arc length is called *geodesic*

The geometric mean as barycenter in RTM[5]

- For any $A, B \in \mathcal{P}_n$ the geodesic joining A to B in RTM is given by

$$\gamma_{A\to B}(t) = A^{\frac{1}{2}}\left(A^{-\frac{1}{2}}BA^{-\frac{1}{2}}\right)^{t}A^{\frac{1}{2}}$$

that is, the geodesic consists of the weighted geometric means.

- Consequently,

$$\gamma'_{A\to B}(t) = A^{\frac{1}{2}}\left(A^{-\frac{1}{2}}BA^{-\frac{1}{2}}\right)^{t}\log\left(A^{-\frac{1}{2}}BA^{-\frac{1}{2}}\right)A^{\frac{1}{2}},$$

[5] Corach-Porta-Recht, Lawson-Lim, Bhatia-Holbrook, Moakher

- ...and the RTM has a simple closed form:

$$d_{RTM}(A,B) = \int_0^1 \sqrt{g_{\gamma_{A\to B}(t)}\left(\gamma'_{A\to B}(t), \gamma'_{A\to B}(t)\right)}\mathrm{d}t$$

$$= \int_0^1 \sqrt{\mathrm{Tr}\left(\left(\gamma_{A\to B}(t)\right)^{-1}\gamma'_{A\to B}(t)\right)^2}\mathrm{d}t$$

$$= \left\|\log\left(A^{-\frac{1}{2}}BA^{-\frac{1}{2}}\right)\right\|_2$$

- The midpoint of the geodesic curve joining A to B is the geometric mean:

$$\gamma_{A\to B}(\frac{1}{2}) = A\#B = A^{\frac{1}{2}}\left(A^{-\frac{1}{2}}BA^{-\frac{1}{2}}\right)^{\frac{1}{2}}A^{\frac{1}{2}}.$$

- The S-divergence is given by

$$d_S(X,Y) = \sqrt{\mathrm{Tr}\log\left(\frac{X+Y}{2}\right) - \frac{1}{2}\mathrm{Tr}\log X - \frac{1}{2}\mathrm{Tr}\log Y}.$$

The geometric mean is the mean squared error estimator also for the S-divergence[6]

$$A\#B = \arg\min_{X>0}\frac{1}{2}\left(d_S^2(A,X) + d_S^2(B,X)\right)$$

[6] S. Sra, *Positive definite matrices and the S-divergence,* Proc. Amer. Math. Soc. **144** (2016), 2787-2797.

The Karcher mean

- the barycenter of the positive definite matrices $A_1, \dots, A_m$ with weights $w_1, \dots w_m$, which is usually called *Karcher mean* or *multivariate geometric mean* in RTM is

$$\underset{X \in \mathcal{P}_n}{\arg\min} \sum_{j=1}^{m} w_j d^2_{RTM}(A_j, X) = \underset{X \in \mathcal{P}_n}{\arg\min} \sum_{j=1}^{m} w_j \left\| \log \left(A_j^{-\frac{1}{2}} X A_j^{-\frac{1}{2}} \right) \right\|_2^2$$

- the barycenter problem: we aim to find $X_0 \in \mathcal{P}_n$, where the derivative of the mean squared error function vanishes, i.e.,

$$\mathbf{D} \left(\sum_{j=1}^{m} w_j d^2_{RTM}(A_j, \cdot) \right) (X_0)[Y] = 0 \qquad (Y \in M_n^{sa})$$

- keeping in mind that

$$\mathsf{D}\left(\|\log(\cdot)\|_2^2\right)(X)[Y] = 2\mathrm{Tr}\left(X^{-1}\log X\right)Y,$$

we get[7] that the Karcher mean is the solution of the nonlinear matrix equation called *Karcher equation*:

$$\sum_{j=1}^{m} w_j \log\left(X^{\frac{1}{2}}A_j^{-1}X^{\frac{1}{2}}\right) = 0$$

- no explicit formula is known unless all the A_j's commute
- in the commutative case, the Karcher mean coincides with the geometric mean

$$\prod_{j=1}^{m} A_j^{w_j}$$

[7] R. Bhatia, Positive Definite Matrices, Princeton University Press, 2007.

The harmonic mean as barycenter[8]

- We can define a Riemannian metric on $\mathcal{P}_n$ locally at A by the relation

$$\mathrm{d}s = \|A^{-1}\mathrm{d}AA^{-1}\|_2,$$

- Let $\gamma : [0,1] \to \mathcal{P}_n$ be a smooth path. The arc-length along this path is given by

$$L(\gamma) = \int_0^1 \|\gamma(t)^{-1}\gamma'(t)\gamma(t)^{-1}\|_2 \mathrm{d}t.$$

- The corresponding geodesic distance between $A, B \in \mathcal{P}_n$ is defined by

$$\delta(A,B) =$$

$$\inf\left\{\int_0^1 \|\gamma(t)^{-1}\gamma'(t)\gamma(t)^{-1}\|_2 \mathrm{d}t : \gamma(t) \in \mathcal{P}_n, t \in (0,1), \gamma(0) = A, \gamma(1) = B\right\}.$$

[8] P. J., Virosztek D. 2021. A divergence center interpretation of general symmetric Kubo-Ando means, and related weighted multivariate operator means. Linear Algebra and its Applications. 609, 203–217.

- The unique geodesic running from A to B is the weighted harmonic mean:

$$\gamma(t) = A!_t B = \left[(1-t)A^{-1} + tB^{-1}\right]^{-1}, \quad t \in [0,1]$$

- The geodesic distance is:

$$\delta(A,B) = \|B^{-1} - A^{-1}\|_2.$$

A new divergence

Let $\sigma : \mathcal{B}(\mathcal{H})^{++} \times \mathcal{B}(\mathcal{H})^{++} \to \mathcal{B}(\mathcal{H})^{++}$ be a symmetric Kubo-Ando mean with operator monotone representing function $f_\sigma : (0, \infty) \to (0, \infty)$ i.e.

$$A \sigma B = A^{\frac{1}{2}} f_\sigma \left(A^{-\frac{1}{2}} B A^{-\frac{1}{2}} \right) A^{\frac{1}{2}}.$$

Clearly, $f_\sigma(1) = 1$, and the symmetry of σ implies that $f_\sigma(x) = x f_\sigma \left(\frac{1}{x} \right)$ for $x > 0$, and hence $f'_\sigma(1) = 1/2$. We define

$$g_\sigma : (0, \infty) \supseteq \mathrm{ran}\,(f_\sigma) \to [0, \infty)$$

by

$$g_\sigma(x) := \int_1^x \left(1 - \frac{1}{f_\sigma^{-1}(t)} \right) \mathrm{d}t.$$

Obviously, $g_\sigma(1) = 0$, $g'_\sigma(x) = 1 - \frac{1}{f_\sigma^{-1}(x)}$, and $g'_\sigma(1) = 0$ as $f_\sigma(1) = 1$.

Now we define[9] the following quantity for operators $A, B \in \mathcal{B}(\mathcal{H})^{++}$

$$\phi_\sigma(A,B) := \begin{cases} \mathrm{Tr} g_\sigma \left(A^{-1/2} B A^{-1/2}\right), & \text{if } \mathrm{spec}\left(A^{-1/2} B A^{-1/2}\right) \subseteq \mathrm{ran}\left(f_\sigma\right), \\ +\infty, & \text{if } \mathrm{spec}\left(A^{-1/2} B A^{-1/2}\right) \nsubseteq \mathrm{ran}\left(f_\sigma\right). \end{cases}$$

[9] P. J., Virosztek D. 2021. A divergence center interpretation of general symmetric Kubo-Ando means, and related weighted multivariate operator means. Linear Algebra and its Applications. 609, 203–217.

Then ϕ_σ is a **divergence** in the sense of Amari[10], i.e.
for any symmetric Kubo-Ando mean σ, the map

$$\phi_\sigma : \mathcal{B}(\mathcal{H})^{++} \times \mathcal{B}(\mathcal{H})^{++} \to [0, +\infty]; \quad (A, B) \mapsto \phi_\sigma(A, B)$$

satisfies the followings.

- $\phi_\sigma(A, B) \geq 0$ and $\phi_\sigma(A, B) = 0$ if and only if $A = B$.
- The first derivative of ϕ_σ in the second variable vanishes at the diagonal, that is, $\mathbf{D}\left(\phi_\sigma(A, \cdot)\right)[A] = 0 \in \mathrm{Lin}\left(\mathcal{B}(\mathcal{H})^{sa}, \mathbb{R}\right)$ for all $A \in \mathcal{B}(\mathcal{H})^{++}$.
- The second derivative of Φ_σ in the second variable is positive at the diagonal, that is, $\mathbf{D}^2\left(\phi_\sigma(A, \cdot)\right)[A](Y, Y) \geq 0$ for all $Y \in \mathcal{B}(\mathcal{H})^{sa}$.

[10] S. Amari, *Information Geometry and its Applications*, Springer (Tokyo), 2016.

Further properties of the divergence ϕ_σ

For any Kubo-Ando mean σ and for any $A, B \in \mathcal{B}(\mathcal{H})^{++}$ we have

- $\phi_\sigma(A^{-1}, B^{-1}) = \phi_\sigma(B, A)$
- $\phi_\sigma\left(TAT^*, TBT^*\right) = \phi_\sigma(A, B)$
  for an arbitrary invertible operator $T \in \mathcal{B}(\mathcal{H})$
- The divergence ϕ_σ is symmetric in its arguments, that is

$$\phi_\sigma(A, B) = \phi_\sigma(B, A)$$

  holds for all $A, B \in \mathcal{B}(\mathcal{H})^{++}$, if and only if $\sigma = \#$ is the geometric mean.

Kubo-Ando means as divergence centers with respect to ϕ_σ

Theorem

[a] *For any $A, B \in \mathcal{B}(\mathcal{H})^{++}$,*

$$\underset{X \in \mathcal{B}(\mathcal{H})^{++}}{\arg\min} \frac{1}{2}\left(\phi_\sigma(A,X) + \phi_\sigma(B,X)\right) = A\sigma B.$$

That is, $A\sigma B$ is a unique minimizer of the function

$$X \mapsto \frac{1}{2}\left(\phi_\sigma(A,X) + \phi_\sigma(B,X)\right)$$

on $\mathcal{B}(\mathcal{H})^{++}$.

[a] P. J., Virosztek D. 2021. A divergence center interpretation of general symmetric Kubo-Ando means, and related weighted multivariate operator means. Linear Algebra and its Applications. 609, 203–217.

Given a symmetric Kubo-Ando mean σ, a finite set of positive definite operators $\mathbf{A} = \{A_1, \ldots, A_m\} \subset \mathcal{B}(\mathcal{H})^{++}$, and a discrete probability distribution $\mathbf{w} = \{w_1, \ldots, w_m\} \subset (0, 1]$ with $\sum_{j=1}^{m} w_j = 1$ we define the corresponding loss function $Q_{\sigma,\mathbf{A},\mathbf{w}} : \mathcal{B}(\mathcal{H})^{++} \to [0, \infty]$ by

$$Q_{\sigma,\mathbf{A},\mathbf{w}}(X) := \sum_{j=1}^{m} w_j \phi_\sigma \left(A_j, X\right).$$

From now on, we assume that the range of f_σ is maximal, that is, $\operatorname{ran}\left(f_\sigma\right) = (0, \infty)$. Consequently, ϕ_σ is always finite, and hence so is $Q_{\sigma,\mathbf{A},\mathbf{w}}$ on the whole positive definite cone $\mathcal{B}(\mathcal{H})^{++}$.

Let $\sigma : \mathcal{B}(\mathcal{H})^{++} \times \mathcal{B}(\mathcal{H})^{++} \to \mathcal{B}(\mathcal{H})^{++}$ be a symmetric Kubo-Ando operator mean such that the operator monotone representing function $f_\sigma : (0, \infty) \to (0, \infty)$ is surjective. We call the optimizer

$$\mathbf{bc}\,(\sigma, \mathbf{A}, \mathbf{w}) := \underset{X \in \mathcal{B}(\mathcal{H})^{++}}{\arg\min}\; Q_{\sigma,\mathbf{A},\mathbf{w}}$$

the weighted barycenter of the operators $\{A_1, \dots, A_m\}$ with weights $\{w_1, \dots, w_m\}$.

To find the barycenter $\mathbf{bc}(\sigma, \mathbf{A}, \mathbf{w})$, we have to solve the critical point equation

$$\mathbf{D}Q_{\sigma,\mathbf{A},\mathbf{w}}[X](\cdot) = 0$$

for the strictly convex loss function $Q_{\sigma,\mathbf{A},\mathbf{w}}$, where the symbol

$$\mathbf{D}Q_{\sigma,\mathbf{A},\mathbf{w}}[X](\cdot) \in \mathrm{Lin}\left(\mathcal{B}(\mathcal{H})^{sa}, \mathbb{R}\right)$$

stands for the Fréchet derivative of $Q_{\sigma,\mathbf{A},\mathbf{w}}$ at the point $X \in \mathcal{B}(\mathcal{H})^{++}$.

That is, the equation to be solved is

$$\sum_{j=1}^{m} w_j A_j^{-\frac{1}{2}} g'_\sigma \left(A_j^{-\frac{1}{2}} X A_j^{-\frac{1}{2}} \right) A_j^{-\frac{1}{2}} = 0.$$

By the definition of g_σ, $g'_\sigma(t) = 1 - \frac{1}{f_\sigma^{-1}(t)}$ for $t \in (0, \infty)$, and hence the critical point of the loss function $Q_{\sigma,\mathbf{A},\mathbf{w}}$ is described by the equation

$$\sum_{j=1}^{m} w_j A_j^{-\frac{1}{2}} \left(I - \left(f_\sigma^{-1} \left(A_j^{-\frac{1}{2}} X A_j^{-\frac{1}{2}} \right) \right)^{-1} \right) A_j^{-\frac{1}{2}} = 0.$$

The barycenter corresponding to the geometric mean

For $\sigma = \#$ the generating function is $f_{\#}(x) = \sqrt{x}$, and hence the inverse is $f_{\#}^{-1}(t) = t^2$. In this case, the critical point equation describing the barycenter $\mathbf{bc}(\#, \mathbf{A}, \mathbf{w})$ reads as follows:

$$\sum_{j=1}^{m} w_j \left(A_j^{-1} - X^{-1} A_j X^{-1} \right) = 0.$$

It can be rearranged as

$$X \left(\sum_{j=1}^{m} w_j A_j^{-1} \right) X = \sum_{j=1}^{m} w_j A_j.$$

This is the Ricatti equation for the weighted multivariate harmonic mean $\left(\sum_{j=1}^{m} w_j A_j^{-1} \right)^{-1}$ and arithmetic mean $\sum_{j=1}^{m} w_j A_j$.

Hence the barycenter $\mathbf{bc}\,(\#, \mathbf{A}, \mathbf{w})$ coincides with the weighted $\mathcal{A}\#\mathcal{H}$-mean of Kim, Lawson, and Lim[11], that is,

$$\mathbf{bc}\,(\#, \mathbf{A}, \mathbf{w}) = \left(\sum_{j=1}^{m} w_j A_j^{-1} \right)^{-1} \# \left(\sum_{j=1}^{m} w_j A_j \right).$$

[11] S. Kim, J. Lawson, Y. Lim, *The matrix geometric mean of parametrized, weighted arithmetic and harmonic means,* Linear Algebra Appl. **435** (2011), 2114–2131.

Elementary properties of the barycenter

The barycenter $\mathbf{bc}(\sigma, \mathbf{A}, \mathbf{w})$ satisfies the following properties:

- Idempotency: $\mathbf{bc}(\sigma, \{A, \dots, A\}, \mathbf{w}) = A$ for any symmetric Kubo-Ando mean σ, any $A \in \mathcal{B}(\mathcal{H})^{++}$, and any probability vector $\mathbf{w}$.
- Homogeneity: $\mathbf{bc}(\sigma, t\mathbf{A}, \mathbf{w}) = t\mathbf{bc}(\sigma, \mathbf{A}, \mathbf{w})$ where the shorthand $t\mathbf{A}$ denotes $\{tA_1, \dots, tA_m\}$ if $\mathbf{A} = \{A_1, \dots A_m\}$
- Permutation invariance: $\mathbf{bc}(\sigma, \mathbf{A}_\pi, \mathbf{w}_\pi) = \mathbf{bc}(\sigma, \mathbf{A}, \mathbf{w})$ where π is a permutation of $\{1, \dots, m\}$, and $\mathbf{A}_\pi = \left\{A_{\pi(1)}, \dots, A_{\pi(m)}\right\}$, $\mathbf{w}_\pi = \left\{w_{\pi(1)}, \dots, w_{\pi(m)}\right\}$.

- Congruence invariance:

$$\mathbf{bc}\left(\sigma, T\mathbf{A}T^{*}, \mathbf{w}\right) = T\mathbf{bc}\left(\sigma, \mathbf{A}, \mathbf{w}\right) T^{*}$$

  for any invertible $T \in \mathcal{B}(\mathcal{H})$, where $T\mathbf{A}T^{*} = \{TA_1T^{*}, \dots, TA_mT^{*}\}$ if $\mathbf{A} = \{A_1, \dots A_m\}$.

- The weighted multivariate harmonic mean is a lower bound for the barycenter

$$\left(\sum_{j=1}^{m} w_j A_j^{-1}\right)^{-1} \leq \mathbf{bc}\left(\sigma, \mathbf{A}, \mathbf{w}\right).$$

Thanks for Your attention!

Acknowledgement

The speaker was supported by the Hungarian National Research, Development and Innovation Office (NKFIH) via

- the "Frontline" Research Excellence Programme (Grant No. KKP133827)
- grants no. K119442, no. K124152, and no. KH129601.

and by the Hungarian Academy of Sciences Lendület-Momentum Grant for Quantum Information Theory, No. 96 141.