**Acta Scientiarum Mathematicarum**

# Operator means, barycenters, and fixed point equations

Dániel Virosztek

**Abstract.** The seminal work of Kubo and Ando (Math Ann 246:205–224, 1979/80) provided us with an axiomatic approach to means of positive operators. As most of their axioms are algebraic in nature, this approach has a clear algebraic flavour. On the other hand, it is highly natural to take the geomeric viewpoint and consider a distance (understood in a broad sense) on the cone of positive operators, and define the mean of positive operators by an appropriate notion of the center of mass. This strategy often leads to a fixed point equation that characterizes the mean. The aim of this survey is to highlight those cases where the algebraic and the geometric approaches meet each other.

**Mathematics Subject Classification.** Primary 47A64, Secondary 47A60.

**Keywords.** Operator mean, Barycenter, Fixed point equation.

## Contents

## 1. An algebraic approach to operator menas

Although the *paralell sum* $A : B = \left(A^{-1} + B^{-1}\right)^{-1}$ of positive definite matrice was considered already in 1969 by *Anderson* and *Duffin* [2], who

used this operation to study electrical networks, and the *geometric mean* $A \# B = A^{\frac{1}{2}} \left( A^{-\frac{1}{2}} B A^{-\frac{1}{2}} \right)^{\frac{1}{2}} A^{\frac{1}{2}} = B^{\frac{1}{2}} \left( B^{-\frac{1}{2}} A B^{-\frac{1}{2}} \right)^{\frac{1}{2}} B^{\frac{1}{2}}$ was introduced by *Pusz* and *Woronowicz* in 1975 [23], the first systematic study of means of positive operators is due to *Kubo* and *Ando* from 1980 [17], who provided an axiomatic approach to operator means which proved extremely influential in the last decades. The binary operations characterized by their axioms are now called *Kubo-Ando* connections and means, and the study of these means is flourishing research field within operator theory with intimate connections to quantum information theory.

This first section is very brief introduction to the Kubo-Ando theory which, however, aims to cover the most important concepts and phenomena. First, we need to fix some notation. Let $\mathcal{H}$ be a complex Hilbert space, either finite or infinite dimensional, and let $\mathcal{B}(\mathcal{H})$ denote the set of all bounded linear operators on $\mathcal{H}$. Let $\mathcal{B}(\mathcal{H})^{sa}$, $\mathcal{B}(\mathcal{H})^{+}$, and $\mathcal{B}(\mathcal{H})^{++}$ stand for the set of all bounded self-adjoint, positive semidefinite, and positive definite (i.e., positive semidefinite and invertible) operators, respectively. The symbol $I$ stands for the identity of $\mathcal{B}(\mathcal{H})$, and 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 semidefinite, and $A < B$ means that $B - A$ is positive definite. The spectrum of $X \in \mathcal{B}(\mathcal{H})$ is denoted by $\operatorname{spec}(X)$. The symbols $\mathbf{D}$ and $\mathbf{D}^2$ denote the first and second Fréchet derivatives, respectively.

**Definition 1.** (Operator connections and means) A binary operation
$$\sigma : \mathcal{B}(\mathcal{H})^{+} \times \mathcal{B}(\mathcal{H})^{+} \to \mathcal{B}(\mathcal{H})^{+}; (A, B) \mapsto A \sigma B$$
is called an *operator connection* if it satisfies the following three properties:

- (P1) Monotonocity in both variables:
$$\text{if } A \leq A' \text{ and } B \leq B' \text{ then } A \sigma B \leq A' \sigma B'$$

- (P2) Transformer inequality:
$$C (A \sigma B) C \leq (CAC)\sigma(CBC) \text{ for all } A, B, C \in \mathcal{B}(\mathcal{H})^{+}$$

- (P3) Continuity for decreasing sequences: if $A_1 \geq A_2 \geq A_3 \geq \ldots$ and $A_n \to A$ in the strong operator topology, and similarly, $B_1 \geq B_2 \geq B_3 \geq \ldots$ and $B_n \to B$ strongly, then
$$(A_n) \, \sigma \, (B_n) \to A \sigma B \text{ in the strong operator topology}$$

Furthermore, *operator means* are those operator connections that satisfy the

- (P4) Normalization condition: $I \sigma I = I$

One of the most striking results of [17] is that there is one-by-one correspondence between operator connections and operator monotone functions mapping the positive half-line into itself. Namely, for every operator connection $\sigma$ there is a unique operator monotone function $f : (0, \infty) \to (0, \infty)$ such that

$$A \sigma B = A^{\frac{1}{2}} f \left( A^{-\frac{1}{2}} B A^{-\frac{1}{2}} \right) A^{\frac{1}{2}} \qquad (A, B \in \mathcal{B}(\mathcal{H})^{++}). \qquad (1)$$

The $\sigma \leftrightarrow f$ correspondence described in (1) is in fact an affine order-isomorphism. It is instructive to take a look at those particular connections that have already been mentioned, and note that the parallel sum $A : B = \left(A^{-1} + B^{-1}\right)^{-1}$ corresponds to the operator monotone function $f(x) = \frac{x}{x+1}$ while the geometric mean $A \# B = A^{\frac{1}{2}} \left(A^{-\frac{1}{2}} B A^{-\frac{1}{2}}\right)^{\frac{1}{2}} A^{\frac{1}{2}}$ corresponds to $f(x) = \sqrt{x}$.

Operator monotone functions mapping the positive half-line $(0, \infty)$ into itself admit a transparent integral-representation by Löwner's theory. In [17], the following integral representation was considered:

$$f(x) = \int_{[0,\infty]} \frac{x(1+t)}{x+t} \mathrm{d}m(t) \qquad (x > 0), \tag{2}$$

where $m$ is a positive Radon measure on the extended half-line $[0, \infty]$. There is another integral-representation formula for operator monotone functions which may be even more convenient than (2). By a simple push-forward of $m$ by the transformation $T : [0, \infty] \to [0, 1]; t \mapsto \lambda := \frac{t}{t+1}$, we get the following integral-representation of positive operator monotone functions on $(0, \infty)$ :

$$f_\mu(x) = \int_{[0,1]} \frac{x}{(1-\lambda)x + \lambda} \mathrm{d}\mu(\lambda) \qquad (x > 0), \tag{3}$$

where $\mu = T_\# m$, that is, $\mu(A) = m\left(T^{-1}(A)\right)$ for every Borel set $A \subseteq [0, 1]$. This representation is also well-known and appears — among others — in [10]. Note that if $m$ is absolutely continuous with respect to the Lebesgue measure and $\mathrm{d}m(t) = \rho(t)\mathrm{d}t$, then the density of $\mu = T_\# m$ is given by $\mathrm{d}\mu(\lambda) = \frac{1}{(1-\lambda)^2} \rho\left(\frac{\lambda}{1-\lambda}\right) \mathrm{d}\lambda$.

In particular, by (1) and (3), there is an affine isomorphism between operator connections and positive Radon measures on $[0, 1]$. We note that the normalization condition $I\sigma I = I$ is satisfied if and only if $\mu([0, 1]) = 1$.

We denote by $\mathcal{P}([0, 1])$ the set of all Borel probability measures on $[0, 1]$, and by $c(\mu) := \int_{[0,1]} \lambda \mathrm{d}\mu(\lambda)$ the center of mass of $\mu$. There is a natural way to assign a weight parameter to a mean $\sigma$, namely, $W(\sigma) := f'(1) = c(\mu)$, where $f$ generates $\sigma$ in the sense of (1) and $\mu$ generates $f$ in the sense of (3). This weight parameter will play an essential role later when we turn to the discussion of *generalized Hellinger distances*. Here we only mention that for the weighted arithmetic, geometric, and harmonic means generated by the functions

$$a_\lambda(x) = (1 - \lambda) + \lambda x, \; g_\lambda(x) = x^\lambda, \text{ and } h_\lambda(x) = \left((1 - \lambda) + \lambda x^{-1}\right)^{-1},$$

respectively, we have $W(\sigma_{a_\lambda}) = W(\sigma_{g_\lambda}) = W(\sigma_{h_\lambda}) = \lambda$. That is, this weight parameter coincides with the usual one in the most important special cases.

The *convex order* is a well-known relation between probability measures; for $\mu, \nu \in \mathcal{P}([0, 1])$, we say that $\mu \preccurlyeq \nu$ if for all convex functions $u : [0, 1] \to \mathbb{R}$ we have $\int_{[0,1]} u \, \mathrm{d}\mu \leq \int_{[0,1]} u \, \mathrm{d}\nu$. It is clear that for all $\mu \in \mathcal{P}([0, 1])$ with $c(\mu) = \lambda$ we have $\delta_\lambda \preccurlyeq \mu \preccurlyeq (1 - \lambda)\delta_0 + \lambda\delta_1$, where $\delta_x$ denotes the Dirac

mass concentrated on $x$. For any fixed $x > 0$, the map $\lambda \mapsto \frac{x}{(1-\lambda)x+\lambda}$ — which is the core of the integral representation (3) — is convex. Therefore, if $\mu \preccurlyeq \nu$, then $f_\mu(x) \leq f_\nu(x)$ for all $x > 0$, and hence $A\sigma_\mu B \leq A\sigma_\nu B$ for all $A, B \in \mathcal{B}(\mathcal{H})^{++}$, where $f_\mu$ denotes the generating function corresponding to $\mu$ via (3) and $\sigma_\mu$ denotes the operator mean corresponding to $f_\mu$ via (1). Consequently, if $\nu = (1 - c(\mu))\delta_0 + c(\mu)\delta_1$, then $A\sigma_\nu B - A\sigma_\mu B$ is always positive, in particular, $\mathrm{tr}\,(A\sigma_\nu B - A\sigma_\mu B) \geq 0$. This quantity is exactly the generalized quantum Hellinger divergence we will discuss in Sect. 2.3.

Let us turn to the description of two interesting involutive operations on operator connections. The *adjoint* of a conncetion $\sigma$ is denoted by $\sigma^*$ and is defined by

$$A\sigma^* B = \left(A^{-1}\sigma B^{-1}\right)^{-1} \tag{4}$$

for $A, B \in \mathcal{B}(\mathcal{H})^{++}$. If $f$ is the generating function of $\sigma$ in the sense of (1), then $\sigma^*$ is generated by $f^*(x) = 1/f(1/x)$. Note that taking the inverse on positive definite operators is an order-reversing operation, and hence $f^*$ is operator monotone whenever $f$ is so. Important examples of self-adjoint connections are the *weighted geometric means* generated by the power functions $[0, \infty) \ni x \mapsto x^p$ for $0 \leq p \leq 1$. The *transpose* of $\sigma$ is denoted by $\sigma^t$ and is defined by

$$A\sigma^t B = B\sigma A \qquad (A, B \in \mathcal{B}(\mathcal{H})^+). \tag{5}$$

An important fact is that if $\sigma$ is represented by $f$, then its transpose $\sigma^t$ is represented by $xf(1/x)$. Therefore, a connection $\sigma$ is symmetric if and only if its generator $f$ satisfies $f(x) = xf(1/x)$.

All in all, the Kubo-Ando theory is a beautiful and satisfactory theory of two-variable operator means. However, it leaves the problem multivariate operator means untouched. A clear advantage of the geometric approach to be presented in the next section is that it produces natural candidates for means of several positive operators. Apparently, a substantial part of the studies of Riemannian geometries on positive operators was motivated by the problem of finding appropriate multivariate counterparts of well-established bi-variate operator means.

## 2. A geometric approach to operator means

The notion of barycenter—or least squares mean—plays a central role in averaging procedures related to various topics in mathematics and mathematical physics. Given a metric space $(X, d)$ and an $m$-tuple $a_1, \ldots, a_m$ in $X$ with positive weights $w_1, \ldots, w_m$ such that $\sum_{j=1}^m w_j = 1$, the barycenter is defined to be

$$\arg\min_{x \in X} \sum_{j=1}^m w_j d^2(a_j, x). \tag{6}$$

The object defined above by (6) plays an important role in various areas of mathematics, and hence can be found under various names: it is sometimes called *Fréchet mean* or *Karcher mean* or *Cartan mean*. It is also called

*mean squared error estimator* for the following reason: imagine that we want to determine an unknown element of the metric space $(X, d)$ and we can perform imperfect measurements many times. If we measure $a_j$ with relative frequency $w_j$, then an attractive estimation of the unknown object is (6) which minimizes the weighted mean squared error from the measured objects.

In the sequel, we review two distinguished metrics on the cone of positive definite operators on finite dimensional Hilbert spaces, namely the *Riemannian trace metric* and the *Bures-Wasserstein metric*. The focus of these review will be on the barycenters determined by these metrics.

## 2.1. The Riemannian trace metric

The convex *Boltzmann entropy* (or H-functional) of a random variable $X$ with probability density $\varrho$ is given by

$$H(X) = \int_{\text{supp}(X)} \varrho(x) \log \varrho(x) \mathrm{d}x. \tag{7}$$

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

$$\partial_t \varrho = \Delta \varrho$$

can be seen as the gradient flow for the Boltzmann entropy as potential (or "energy") in the differential structure induced by optimal transportation [11–13].

Let us restrict our attention to special random variables. A circularly-symmetric centered complex Gaussian distribution on $\mathbb{C}^N$ is completely described by its covariance matrix $\Sigma$. The probability density of such a zero-mean complex Gaussian $Z \sim \mathcal{N}_{\mathbb{C}}(0, \Sigma)$ is given by

$$f_{\mathcal{N}_{\mathbb{C}}(0,\Sigma)}(z) = \frac{\exp\left(-z^* \Sigma^{-1} z\right)}{\pi^N \det \Sigma}.$$

Consequently, the Boltzmann entropy of $Z \sim \mathcal{N}_{\mathbb{C}}(0, \Sigma)$ can be given in the following simple closed form:

$$H(Z) = -\text{tr} \log \Sigma + C(N) \tag{8}$$

where $C(N)$ is an irrelevant constant depending only on the dimension $N$. So the Boltzmann entropy is a smooth convex functional on the sub-manifold of centered, circularly-symmetric, non-degenerate complex Gaussians. We identify this sub-manifold with the cone of positive definite $N \times N$ matrices by the convenient identification of the random variable with its covariance matrix.

Direct computations shows that the Hessian (that is, the second derivative) of the Boltzmann entropy (8) on these appropriate Gaussians is given by

$$\mathbf{D}^2 H(A)[Y, X] = \text{tr} A^{-1} Y A^{-1} X. \tag{9}$$

This is a collection of positive definite bilinear forms on the tangent spaces

$$T_A\left(M_N^{++}(\mathbb{C})\right) \simeq M_N^{sa}(\mathbb{C})$$

that depends smoothly on the foot point $A$. Therefore, (9) is a Riemannian metric tensor field. The global metric induced by this Riemannian tensor field

$$g_A(X, Y) := \text{tr} A^{-1} Y A^{-1} X$$

is called the *Riemannian trace metric (RTM)*.

A particularly nice feature of the Riemannian trace metric is that the geodesic curve connecting the points $A, B \in \mathcal{B}(\mathcal{H})^{++}$ has the following simple closed form:

$$\gamma_{A \to B}(t) = A^{\frac{1}{2}} \left( A^{-\frac{1}{2}} B A^{-\frac{1}{2}} \right)^t A^{\frac{1}{2}}. \tag{10}$$

That is, the geodesic consists of the *weighted geometric means* where the weight parameter $t$ runs from 0 to 1. Consequently, the derivative of the geodesic reads as

$$\gamma'_{A \to B}(t) = A^{\frac{1}{2}} \left( A^{-\frac{1}{2}} B A^{-\frac{1}{2}} \right)^t \log \left( A^{-\frac{1}{2}} B A^{-\frac{1}{2}} \right) A^{\frac{1}{2}}, \tag{11}$$

and the RTM has the following simple closed form:

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

$$= \int_0^1 \sqrt{\mathrm{tr} \left( \left( (\gamma_{A \to B}(t))^{-1} \gamma'_{A \to B}(t) \right)^2 \right)} \mathrm{d}t = \left\| \log \left( A^{-\frac{1}{2}} B A^{-\frac{1}{2}} \right) \right\|_2. \tag{12}$$

The notation $||\cdot||_2$ stands here and throughout this paper for the Hilbert-Schmidt norm $||X||_2 := \sqrt{\mathrm{tr}\,(X^*X)}$.

Therefore, the barycenter (6) of the positive definite operators $A_1, \ldots, A_m \in \mathcal{B}(\mathcal{H})^{++}$ with probability weights $w_1, \ldots w_m$ is

$$\underset{X \in \mathcal{B}(\mathcal{H})^{++}}{\arg\min} \sum_{j=1}^m w_j d_{RTM}^2 (A_j, X) = \underset{X \in \mathcal{B}(\mathcal{H})^{++}}{\arg\min} \sum_{j=1}^m w_j \left\| \log \left( X^{-\frac{1}{2}} A_j X^{-\frac{1}{2}} \right) \right\|_2^2 \tag{13}$$

One can use Karcher's formula [14, Theorem 2.1] to compute the gradient of the objective function

$$X \mapsto \sum_{j=1}^m w_j \left\| \log \left( X^{-\frac{1}{2}} A_j X^{-\frac{1}{2}} \right) \right\|_2^2 \tag{14}$$

and deduce that the barycenter (13)—which is more often called *Karcher mean* in this context—coincides with the unique positive definite solution of the *Karcher equation*

$$\sum_{j=1}^m w_j \log \left( X^{\frac{1}{2}} A_j^{-1} X^{\frac{1}{2}} \right) = 0. \tag{15}$$

See [18] and [4] for alternative approaches on the derivation of the Karcher Eq. (15).

## 2.2. The Bures-Wasserstein metric

The *classical optimal transport (OT) problem* is to arrange the transportation of goods from producers to consumers in an optimal way, given the distribution of the sources and the needs (described by probability measures $\mu$ and $\nu$), and the cost $c(x, y)$ of transporting a unit of goods from $x$ to $y$. Accordingly, a *transport plan* is modeled by a probability distribution $\pi$ on the product of the initial and the target spaces, where $\mathrm{d}\pi(x, y)$ is the amount of goods

to be transferred from $x$ to $y$, and hence the marginals of $\pi$ are $\mu$ and $\nu$. So the *optimal transport cost* is the minimum of a convex optimization problem with linear loss function:

$$\text{Cost}\,(\mu,\nu,c) = \min\left\{\iint_{X\times Y} c(x,y)\mathrm{d}\pi\,(x,y)\,\middle|\,(\pi)_1 = \mu,\,(\pi)_2 = \nu\right\} \quad (16)$$

where $(\pi)_i$ denotes the $i$th marginal of $\pi$, and $X$ is the initial and $Y$ is the target space.

OT costs (16) give rise to OT distances (Wasserstein distances) on measures for certain cost functions $c(.,.)$. A prominent example is the *quadratic Wasserstein distance* between probabilities on $\mathbb{R}^n$ having finite second moment defined by

$$d^2_{\mathcal{W}_2}(\mu,\nu) = \inf\left\{\iint_{\mathbb{R}^n\times\mathbb{R}^n} |x-y|^2\,\mathrm{d}\pi\,(x,y)\,\middle|\,(\pi)_1 = \mu,\,(\pi)_2 = \nu\right\}. \quad (17)$$

The above definition (17) of the 2-Wasserstein metric is *static in nature*. It refers only to the initial and final distributions of the mass to be transported. The *dynamical theory of mass transport* concerns on the contrary flows of measures connecting the initial and final states. The optimization problem is *minimizing the total kinetic energy* needed to perform the transport. More precisely, the task is to minimize the kinetic energy over flows of measures $(\rho_t)_{t=0}^T$ connecting the initial and final distributions and time-depending velocity fields $(v_t)_{t=0}^T$ governing the flows—we say that the velocity field $(v_t)_{t=0}^T$ governs the flow $(\rho_t)_{t=0}^T$ if they satisfy the linear transport equation (or continuity equation)

$$\frac{\partial \rho_t}{\partial t} + \nabla_x \cdot (\rho_t v_t) = 0. \quad (18)$$

Accordingly, the formula for the minimal kinetic energy (MKE) needed to transform $\mu$ to $\nu$, if the total time allocated for the transport is $T$, reads as follows:

$$\text{MKE}_T\,(\mu,\nu) =$$

$$= \inf\left\{\int_0^T \int_{\mathbb{R}^n} \rho_t(x)\,|v_t(x)|^2\,\mathrm{d}x\,\mathrm{d}t\,\middle|\,\frac{\partial \rho_t}{\partial t} + \nabla_x \cdot (\rho_t v_t) = 0,\,\rho_0 = \mu,\,\rho_T = \nu\right\}. \quad (19)$$

A seminal result of Benamou and Brenier [3] connects the static and the dynamic theory beautifully: the static 2-Wasserstein distance (17) and the minimal kinetic energy needed to perform the dynamics are essentially the same. More precisely, the Benamou-Brenier formula [3] tells us that the minimizing flow of measures in (19) is given by the *displacement interpolation*, and

$$\text{MKE}_T\,(\mu,\nu) = \frac{1}{T}d^2_{\mathcal{W}_2}(\mu,\nu). \quad (20)$$

Now, if we take another look at (19), keeping (20) in mind, we may observe that the 2-Wasserstein distance $d_{\mathcal{W}_2}$ is given by a formula that looks very

much like a Riemannian geodesic formula. And indeed, there is a Riemannian metric tensor field on the space of probabilities that gives rise to the 2-Wasserstein metric. The discussion of this Riemannian metric in general is beyond the scope of this survey—we will be interested only in the special case of centered Gaussian measures. We only mention that this line of research was pioneered by Otto [20] and we refer the interested reader to Subsection 8.1.2. of [25] for a detailed description of the theory. We must note, however, the groundbreaking discovery of Jordan et al. [11–13] who proved that the heat flow is the gradient flow of the Boltzmann entropy with respect to the 2-Wasserstein Riemannian metric on probability densities.

We shall restrict our attention to non-degenerate centered Gaussian measures on $\mathbb{C}^n$ that we identify with their non-singular covariance matrices. If $\mu$ is the law of the random variable $X \sim \mathcal{N}_{\mathbb{C}}(0, A)$ and $\nu$ is the law of $Y \sim \mathcal{N}_{\mathbb{C}}(0, B)$, then the quadratic Wasserstein distance between $\mu$ and $\nu$ admits the following closed form that refers only to the covariance matrices [1, 24]:

$$d_{\mathcal{W}_2}(\mu, \nu) = \left( \mathrm{tr}A + \mathrm{tr}B - 2\mathrm{tr}\left( A^{\frac{1}{2}} B A^{\frac{1}{2}} \right)^{\frac{1}{2}} \right)^{\frac{1}{2}}. \tag{21}$$

The distance between the positive definite operators $A, B \in \mathcal{B}(\mathcal{H})^{++}$ acting on the Hilbert space $\mathcal{H} = \mathbb{C}^n$ that appears on the right-hand side of (21) has a quantum information theoretic interpretation, as well. In that context, the name *Bures distance* is used more frequently. We refer the reader to [8] for a thorough study of this *Bures-Wasserstein* distance given for $A, B \in \mathcal{B}(\mathcal{H})^{++}$ by

$$d_{BW}^2(A, B) = \mathrm{tr}A + \mathrm{tr}B - 2\mathrm{tr}\left( A^{\frac{1}{2}} B A^{\frac{1}{2}} \right)^{\frac{1}{2}}. \tag{22}$$

We note furthermore that the isometries of the density spaces of $C^*$-algebras with respect to (22) have been determined by Molnár in [19]. The geodesic line segment in the Bures-Wasserstein metric connecting $A$ with $B$ has the following simple closed form [7, 8]:

$$A \Diamond_t B = (1-t)^2 A^2 + t^2 B^2 + t(1-t)\left( (AB)^{\frac{1}{2}} + (BA)^{\frac{1}{2}} \right) \tag{23}$$

where $t$ runs from 0 to 1, and the square roots of the non-Hermitian operators are understood as follows: $(AB)^{\frac{1}{2}} = A^{\frac{1}{2}} \left( A^{\frac{1}{2}} B A^{\frac{1}{2}} \right)^{\frac{1}{2}} A^{-\frac{1}{2}}$ and

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

The elements of the geodesic segment (23) are Bures-Wasserstein barycenters with appropriate weights [7], namely,

$$A \Diamond_t B = \underset{X \in \mathcal{B}(\mathcal{H})^{++}}{\arg\min} (1-t)d_{BW}^2(A, X) + td_{BW}^2(B, X) \qquad (A, B \in \mathcal{B}(\mathcal{H})^{++}). \tag{24}$$

An attractive feature of the Bures-Wasserstein barycenter is that it is characterized by a rather simple fixed point equation: the unique minimizer of

the functional

$$X \mapsto \sum_j w_j d_{BW}^2(A_j, X) \tag{25}$$

on the positive definite cone coincides with the unique positive definite solution of the operator equation

$$X = \sum_{j=1}^m w_j \left( X^{\frac{1}{2}} A_j X^{\frac{1}{2}} \right)^{\frac{1}{2}}. \tag{26}$$

This striking result was first proved in [1] and a very transparent presentation of the proof can be found, e.g., in [8, Section 6].

### 2.3. Barycenters for generalized quantum Hellinger distances

As one can see in (22), the Bures-Wasserstein distance is the square root of the distance between the trace of the arithmetic mean of $A$ and $B$ and the trace of a certain geometric mean of the same operators. Therefore it is a non-commutative version of the *Hellinger distance* of probability vectors defined by

$$d_H^2\left((p_1, \ldots, p_n), (q_1, \ldots q_n)\right) = \sum_{j=1}^n \left(\sqrt{p_j} - \sqrt{q_j}\right)^2. \tag{27}$$

A thorough study of a variety of quantum Hellinger distances is presented in [5]. Somewhat later, *generalized* quantum Hellinger distances were introduced and studied with a strong emphasis on the characterization of the barycenter [21]. Very recently, a one-parameter family of distances including the Bures-Wasserstein distance and a certain Hellinger distance was proposed and studied in [16].

Given a Borel probability measure $\mu \in \mathcal{P}([0,1])$, the corresponding *generalized quantum Hellinger divergence* is given by

$$\phi_\mu(A, B) := \operatorname{tr}\left((1 - c(\mu)) A + c(\mu) B - A\sigma_{f_\mu} B\right) \qquad \left(A, B \in \mathcal{B}(\mathcal{H})^{++}\right), \tag{28}$$

where $c(\mu) = \int_{0,1} \lambda \mathrm{d}\mu(\lambda)$ is the center of mass of $\mu$, and $\sigma_{f_\mu}$ is the Kubo-Ando mean generated by the operator monotone function $f_\mu$ in the sense of (1) and $f_\mu$ is determined by $\mu$ in the sense of (3). We believe that the characterization of the barycenter of finitely many positive operators by a fixed point equation is instructive, especially as we did not present the proof of the analogous result for the Bures-Wasserstein metric. The following result and its proof appeared originally in [21].

**Theorem 1.** *Let $\mu \in \mathcal{P}[0,1]$ and let $\phi_\mu$ be the generalized quantum Hellinger divergence generated by $\mu$ given in (28). The barycenter of the positive definite operators $A_1, \ldots, A_m$ with positive weights $w_1, \ldots, w_m$ with respect to $\phi_\mu$, i.e.,*

$$\operatorname*{arg\,min}_{X \in \mathcal{B}(\mathcal{H})^{++}} \sum_{j=1}^m w_j \phi_\mu\left(A_j, X\right) \tag{29}$$

*coincides with the unique positive definite solution of the fixed point equation*

$$X = \frac{1}{c(\mu)} \sum_{j=1}^{m} w_j \int_{[0,1]} \lambda \left| (1-\lambda) A_j^{-1} X^{\frac{1}{2}} + \lambda X^{-\frac{1}{2}} \right|^{-2} \mathrm{d}\mu(\lambda) \qquad (30)$$

*where $|\cdot|$ stands for the absolute value of an operator, that is, $|Z| = (Z^* Z)^{\frac{1}{2}}$.*

*Proof.* Assume that the positive definite operators $A_1, \ldots, A_m$ and the weights $w_1, \ldots w_m$ are given. By the strict concavity of $f_\mu$, the function

$$X \mapsto \phi_\mu(A, X) = \mathrm{tr}\left( (1-c(\mu)) A + c(\mu) X - A^{\frac{1}{2}} f_\mu\left( A^{-\frac{1}{2}} X A^{-\frac{1}{2}} \right) A^{\frac{1}{2}} \right)$$

is strictly convex on $\mathcal{B}(\mathcal{H})^{++}$, see, e.g., [9, 2.10. Thm.]. Therefore, there is a unique solution $X_0$ of (29), and it is necessarily a critical point of the function $X \mapsto \sum_{j=1}^{m} w_j \phi_\mu(A_j, X)$. That is, it satisfies

$$\mathbf{D}\left( \sum_{j=1}^{m} w_j \phi_\mu(A_j, \cdot) \right)(X_0)[Y] = 0 \qquad (Y \in \mathcal{B}(\mathcal{H})^{sa}). \qquad (31)$$

Easy computations give that

$$\mathbf{D}\left( \sum_{j=1}^{m} w_j \phi_\mu(A_j, \cdot) \right)(X)[Y] = c(\mu)\,\mathrm{tr}\,Y - \sum_{j=1}^{m} w_j \mathrm{tr}\,\mathbf{D}F_{\mu, A_j}(X)[Y], \quad (32)$$

where for a positive definite operator $A$, the map $F_{\mu, A} : \mathcal{B}(\mathcal{H})^{++} \to \mathcal{B}(\mathcal{H})^{++}$ is defined by

$$F_{\mu, A}(X) := A \sigma_{f_\mu} X = A^{\frac{1}{2}} f_\mu\left( A^{-\frac{1}{2}} X A^{-\frac{1}{2}} \right) A^{\frac{1}{2}}. \qquad (33)$$

By differentiating (3), we have

$$\mathbf{D}f_\mu(X)[Y] = \int_{[0,1]} \lambda \left( (1-\lambda) X + \lambda I \right)^{-1} Y \left( (1-\lambda) X + \lambda I \right)^{-1} \mathrm{d}\mu(\lambda) \quad (34)$$

for $X \in \mathcal{B}(\mathcal{H})^{++}$, $Y \in \mathcal{B}(\mathcal{H})^{sa}$. Consequently,

$$\mathbf{D}F_{\mu, A_j}(X)[Y]$$

$$= \int_{[0,1]} \lambda A_j^{\frac{1}{2}} \left( (1-\lambda) A_j^{-\frac{1}{2}} X A_j^{-\frac{1}{2}} + \lambda I \right)^{-1} A_j^{-\frac{1}{2}} Y^{\frac{1}{2}} \times$$

$$\times Y^{\frac{1}{2}} A_j^{-\frac{1}{2}} \left( (1-\lambda) A_j^{-\frac{1}{2}} X A_j^{-\frac{1}{2}} + \lambda I \right)^{-1} A_j^{\frac{1}{2}} \mathrm{d}\mu(\lambda) =$$

$$= \int_{[0,1]} \lambda \left( (1-\lambda) X A_j^{-1} + \lambda I \right)^{-1} Y \left( (1-\lambda) A_j^{-1} X + \lambda I \right)^{-1} \mathrm{d}\mu(\lambda). \quad (35)$$

By the linearity and the cyclic property of the trace, we get from (32) and (35) that (31) is equivalent to

$$\mathrm{tr}\left[ Y \left( c(\mu) I - \sum_{j=1}^{m} w_j \int_{[0,1]} \lambda \left| (1-\lambda) A_j^{-1} X + \lambda I \right|^{-2} \mathrm{d}\mu(\lambda) \right) \right] = 0 \, (Y \in \mathcal{B}(\mathcal{H})^{sa}).$$

$$(36)$$

This latter equation amounts to

$$c\left(\mu\right)I = \sum_{j=1}^{m} w_j \int_{[0,1]} \lambda \left|(1-\lambda)A_j^{-1}X + \lambda I\right|^{-2} \mathrm{d}\mu(\lambda). \qquad (37)$$

Multiplying by $\frac{1}{\sqrt{c(\mu)}}X^{1/2}$ from both left and right gives the desired operator Eq. (30). $\hfill\square$

## 3. Connections between the algebraic and the geometric approaches

This section is devoted to the phenomenon when the algebraic and the geometric approach to operator means meet each other, that is, when Kubo-Ando means admit barycentric interpretation.

It is well known that special Kubo-Ando operator means, namely, the arithmetic and the geometric means admit divergence center interpretations. The arithmetic mean $A\nabla B = (A+B)/2$ is clearly the barycenter for the Euclidean metric on positive operators:

$$A\nabla B = \underset{X>0}{\arg\min} \frac{1}{2}\left(\mathrm{tr}(A-X)^2 + \mathrm{tr}(B-X)^2\right).$$

A much more interesting fact is that the geometric mean $A\#B$ is the barycenter for the Riemannian trace metric $d_{RTM}(X,Y) = \left\|\log\left(X^{-\frac{1}{2}}YX^{-\frac{1}{2}}\right)\right\|_2$, that is,

$$A\#B = \underset{X>0}{\arg\min} \frac{1}{2}\left(d_{RTM}^2(A,X) + d_{RTM}^2(B,X)\right). \qquad (38)$$

The barycentric representation (38) of the bivariate geometric mean opened the gate for the *definition* of the multivariate geometric mean as the barycenter with respect to the Riemannian trace metric. This definition was introduced by Moakher [18] and Bhatia-Holbrook [6]. A recent result in [22] tells us that *every* symmetric Kubo-Ando means admits a divergence center interpretation. Now we turn to the detailed explanation of this latter result, and we follow the presentation of [22].

Let $\sigma : \mathcal{B}(\mathcal{H})^{++} \times \mathcal{B}(\mathcal{H})^{++} \to \mathcal{B}(\mathcal{H})^{++}$ be a symmetric Kubo-Ando operator mean, and let $f_\sigma : (0,\infty) \to (0,\infty)$ be the operator monotone function representing $\sigma$ in the sense that

$$A\sigma B = A^{\frac{1}{2}} f_\sigma\left(A^{-\frac{1}{2}}BA^{-\frac{1}{2}}\right)A^{\frac{1}{2}}. \qquad (39)$$

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

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

by

$$g_\sigma(x) := \int_1^x \left(1 - \frac{1}{f_\sigma^{-1}(t)}\right) \mathrm{d}t. \qquad (40)$$

Obviously, $g_\sigma(1) = 0$, $g'_\sigma(x) = 1 - \frac{1}{f_\sigma^{-1}(x)}$, and $g'_\sigma(1) = 0$ as $f_\sigma(1) = 1$. Since $f_\sigma$ is strictly monotone increasing, so is $g'_\sigma$, and hence $g_\sigma$ is strictly convex on its domain. Now we define the quantity

$$\phi_\sigma(A, B) := \mathrm{tr}\left(g_\sigma\left(A^{-1/2}BA^{-1/2}\right)\right), \tag{41}$$

for positive definite operators $A, B \in \mathcal{B}(\mathcal{H})^{++}$ such that the spectrum of $A^{-1/2}BA^{-1/2}$ is contained in $\mathrm{ran}\,(f_\sigma)$. We define $\phi_\sigma(A, B) := +\infty$ if spec $\left(A^{-1/2}BA^{-1/2}\right) \not\subseteq \mathrm{ran}\,(f_\sigma)$. It will be important in the sequel that by [9, 2.10. Thm.] the strict convexity of $g_\sigma$ implies that $X \mapsto \phi_\sigma(A, X)$ is strictly convex (whenever finite) for every $\sigma$ and $A$. Now we are in the position to formalize the divergence center interpretation of symmetric Kubo-Ando means. The precise statement reads as follows.

**Theorem 2.** *For any* $A, B \in \mathcal{B}(\mathcal{H})^{++}$,

$$\underset{X \in \mathcal{B}(\mathcal{H})^{++}}{\arg\min} \frac{1}{2}\left(\phi_\sigma(A, X) + \phi_\sigma(B, X)\right) = A\sigma B. \tag{42}$$

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

*Proof.* By the strict convexity of $X \mapsto \frac{1}{2}\left(\phi_\sigma(A, X) + \phi_\sigma(B, X)\right)$ it is sufficient to show that $A\sigma B$ is a critical point, and therefore a unique minimizer. First we compute the derivative

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \phi_\sigma(A, X + tY) = \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \mathrm{tr}g_\sigma\left(A^{-1/2}XA^{-1/2} + tA^{-1/2}YA^{-1/2}\right) =$$

$$\mathrm{tr}A^{-1/2}g'_\sigma\left(A^{-1/2}XA^{-1/2}\right)A^{-1/2}Y \tag{43}$$

for all $Y \in \mathcal{B}(\mathcal{H})^{sa}$. Since $g'_\sigma(x) = 1 - (f_\sigma^{-1}(x))^{-1}$, we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \frac{1}{2}\phi_\sigma(A, X + tY) = \frac{1}{2}\mathrm{tr}\left(A^{-1/2}\left(I - \left[f_\sigma^{-1}\left(A^{-1/2}XA^{-1/2}\right)\right]^{-1}\right)A^{-1/2}Y\right) \tag{44}$$

for all $Y \in \mathcal{B}(\mathcal{H})^{sa}$. Substituting $X = A\sigma B = A^{1/2}f_\sigma(A^{-1/2}BA^{-1/2})A^{1/2}$ into the derivative above, the right hand side of (44) becomes

$$\frac{1}{2}\mathrm{tr}\left(A^{-1/2}\left(I - \left[f_\sigma^{-1}\left(A^{-1/2}A^{1/2}f_\sigma(A^{-1/2}BA^{-1/2})A^{1/2}A^{-1/2}\right)\right]^{-1}\right)A^{-1/2}Y\right)$$

$$= \frac{1}{2}\mathrm{tr}(A^{-1} - B^{-1})Y. \tag{45}$$

Since the operator mean $\sigma$ is symmetric, that is

$$A\sigma B = B\sigma A = B^{1/2}f_\sigma\left(B^{-1/2}AB^{-1/2}\right)B^{1/2},$$

a similar computation for the derivative

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} \frac{1}{2}\phi_\sigma(B, X + tY)$$

at $X = A\sigma B$ gives

$$\frac{1}{2}\mathrm{tr}(B^{-1} - A^{-1})Y \tag{46}$$

for all $Y \in \mathcal{B}(\mathcal{H})^{sa}$. Using (45) and (46) we get for the derivative

$$\frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0}\left\{\frac{1}{2}\phi_\sigma(A, X+tY) + \frac{1}{2}\phi_\sigma(B, X+tY)\right\}\bigg|_{X=A\sigma B}$$
$$= \frac{1}{2}\mathrm{tr}(A^{-1} - B^{-1})Y + \frac{1}{2}\mathrm{tr}(B^{-1} - A^{-1})Y = 0$$

for all $Y \in \mathcal{B}(\mathcal{H})^{sa}$. So we obtained that $A\sigma B$ is a critical point and hence a unique minimizer of $X \mapsto \frac{1}{2}\left(\phi_\sigma(A, X) + \phi_\sigma(B, X)\right)$. $\qquad\square$

The above characterization of symmetric Kubo-Ando means as barycenters (Theorem 2) naturally leads to the idea of defining weighted and multivariate versions of Kubo-Ando means as minimizers of appropriate loss functions derived from the divergence $\phi_\sigma$. Given a symmetric Kubo-Ando mean $\sigma$, 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

$$Q_{\sigma, \mathbf{A}, \mathbf{w}}(X) := \sum_{j=1}^m w_j \phi_\sigma\left(A_j, X\right) \tag{47}$$

where $\phi_\sigma$ is defined by (41).

However, in the weighted multivariate setting, when $\mathrm{ran}\left(f_\sigma\right)$ is smaller than the whole positive half-line $(0, \infty)$, then some undesirable phenomena occur which are illustrated by the next example.

Consider the arithmetic mean generated by $f_\nabla(x) = (1 + x)/2$ with $\mathrm{ran}\left(f_\nabla\right) = \left(\frac{1}{2}, \infty\right)$. Let $A_1, A_2 \in \mathcal{B}(\mathcal{H})^{++}$ satisfy $A_1 < \frac{1}{3}A_2$. In this case, for any $\alpha \in (0, 1)$, the loss function $Q_{\nabla, \{A_1, A_2\}, \{1-\alpha, \alpha\}}(X)$ is finite only if $X > \frac{1}{2}A_2$. So the barycenter of $A_1$ and $A_2$ with weights $\{1 - \alpha, \alpha\}$ is separated from $A_1$ for every $\alpha \in (0, 1)$, even for values very close to 0.

To exclude such phenomena, from now on, we assume that the range of $f_\sigma$ is maximal, that is, $\mathrm{ran}\left(f_\sigma\right) = (0, \infty)$, and hence $g_\sigma(x) = \int_1^x \left(1 - \frac{1}{f_\sigma^{-1}(t)}\right)\mathrm{d}t$ is defined on the whole positive half-line $(0, \infty)$. Consequently, $\phi_\sigma$ is always finite, and hence so is $Q_{\sigma, \mathbf{A}, \mathbf{w}}$ on the whole positive definite cone $\mathcal{B}(\mathcal{H})^{++}$.

**Definition 2.** 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 $f_\sigma : (0, \infty) \to (0, \infty)$, which is the operator monotone function representing $\sigma$ in the sense of (39), is surjective. Let $g_\sigma$ be defined as in (40), and $\phi_\sigma$ be defined as in (41). We call the optimizer

$$\mathbf{bc}\left(\sigma, \mathbf{A}, \mathbf{w}\right) := \underset{X \in \mathcal{B}(\mathcal{H})^{++}}{\arg\min}\; Q_{\sigma, \mathbf{A}, \mathbf{w}} \tag{48}$$

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

By Theorem 2, this barycenter may be considered as a weighted multivariate version of Kubo-Ando means. To find the barycenter $\mathbf{bc}\,(\sigma, \mathbf{A}, \mathbf{w})$, we have to solve the critical point equation

$$\mathbf{D}Q_{\sigma,\mathbf{A},\mathbf{w}}[X](\cdot) = 0 \tag{49}$$

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

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

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

$$\mathbf{D}Q_{\sigma,\mathbf{A},\mathbf{w}}[X](Y) = \sum_{j=1}^{m} w_j \mathbf{D}\left(\phi_\sigma\left(A_j, \cdot\right)\right)[X](Y)$$

$$= \sum_{j=1}^{m} w_j \mathbf{D}\left(\mathrm{tr}\left(g_\sigma\left(A_j^{-\frac{1}{2}} \cdot A_j^{-\frac{1}{2}}\right)\right)\right)[X](Y)$$

$$= \sum_{j=1}^{m} w_j \mathrm{tr}\left(g_\sigma'\left(A_j^{-\frac{1}{2}} X A_j^{-\frac{1}{2}}\right) A_j^{-\frac{1}{2}} Y A_j^{-\frac{1}{2}}\right)$$

$$= \mathrm{tr}\left(\left(\sum_{j=1}^{m} w_j A_j^{-\frac{1}{2}} g_\sigma'\left(A_j^{-\frac{1}{2}} X A_j^{-\frac{1}{2}}\right) A_j^{-\frac{1}{2}}\right) Y\right)$$

that is, the equation to be solved is

$$\sum_{j=1}^{m} w_j A_j^{-\frac{1}{2}} g_\sigma'\left(A_j^{-\frac{1}{2}} X A_j^{-\frac{1}{2}}\right) A_j^{-\frac{1}{2}} = 0. \tag{50}$$

By the definition of $g_\sigma$, see (40), $g_\sigma'(t) = 1 - \frac{1}{f_\sigma^{-1}(t)}$ 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

$$\sum_{j=1}^{m} w_j A_j^{-\frac{1}{2}} \left(I - \left(f_\sigma^{-1}\left(A_j^{-\frac{1}{2}} X A_j^{-\frac{1}{2}}\right)\right)^{-1}\right) A_j^{-\frac{1}{2}} = 0. \tag{51}$$

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 (51) describing the barycenter $\mathbf{bc}\,(\#, \mathbf{A}, \mathbf{w})$ reads as follows:

$$\sum_{j=1}^{m} w_j \left(A_j^{-1} - X^{-1} A_j X^{-1}\right) = 0. \tag{52}$$

Note that (52) may be considered as a *generalized Riccati equation*, and in the special case $m = 2, w_1 = w_2 = \frac{1}{2}$, the solution of (52) is the symmetric geometric mean $A_1 \# A_2$.

More generally, if $m = 2, w_1 = 1 - \alpha$, and $w_2 = \alpha$, then (52) has the following form:

$$(1 - \alpha)A_1^{-1} + \alpha A_2^{-1} = (1 - \alpha)X^{-1} A_1 X^{-1} + \alpha X^{-1} A_2 X^{-1},$$

or equivalently

$$X\left[(1 - \alpha)A_1^{-1} + \alpha A_2^{-1}\right] X = (1 - \alpha)A_1 + \alpha A_2. \tag{53}$$

Recall that for positive definite $A$ and $B$, the Riccati equation

$$XA^{-1}X = B$$

has a unique positive definite solution, that is the geometric mean

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

We can observe that (53) is the Riccati equation for the weighted harmonic mean

$$A_1!_\alpha A_2 = [(1-\alpha)A_1^{-1} + \alpha A_2^{-1}]^{-1}$$

and the weighted arithmetic mean $A_1 \nabla_\alpha A_2 = (1-\alpha)A_1 + \alpha A_2$, ie

$$X(A_1!_\alpha A_2)^{-1}X = A_1 \nabla_\alpha A_2. \tag{54}$$

Hence the solution of (53) is the geometric mean of the weighted harmonic and the weighted arithmetic mean

$$X = (A_1!_\alpha A_2)\#(A_1\nabla_\alpha A_2). \tag{55}$$

It means that in this case the weighted barycenter with respect to $\phi_\#$ does not coincide with the weighted geometric mean, nevertheless

$$\mathbf{bc}\left(\#,\{A_1,A_2\},\{1-\alpha,\alpha\}\right) = (A_1!_\alpha A_2)\#(A_1\nabla_\alpha A_2)$$

that is, $\mathbf{bc}\left(\#,\{A_1,A_2\},\{1-\alpha,\alpha\}\right)$ is the Kubo-Ando mean of $A_1$ and $A_2$ with representing function

$$f_{(!_\alpha \# \nabla_\alpha)}(x) = \sqrt{\frac{x(1-\alpha+\alpha x)}{(1-\alpha)x + \alpha}}.$$

These means were widely investigated in [15].

We note that the critical point equation (52) can be rearranged as

$$X\left(\sum_{j=1}^{m} w_j A_j^{-1}\right)X = \sum_{j=1}^{m} w_j A_j. \tag{56}$$

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 $\mathbf{bc}\left(\#,\mathbf{A},\mathbf{w}\right)$ coincides with the weighted $\mathcal{A}\#\mathcal{H}$-mean of Kim, Lawson, and Lim [15], that is,

$$\mathbf{bc}\left(\#,\mathbf{A},\mathbf{w}\right) = \left(\sum_{j=1}^{m} w_j A_j^{-1}\right)^{-1} \# \left(\sum_{j=1}^{m} w_j A_j\right). \tag{57}$$

**Declarations**

**Conflict of interest** On behalf of all authors, the corresponding author states that there is no Conflict of interest.

# References

[1] Agueh, M., Carlier, G.: Barycenters in the Wasserstein space. SIAM J. Math. Anal. **43**(2), 904–924 (2011). https://doi.org/10.1137/100805741

[2] Anderson, W.N., Jr., Duffin, R.J.: Series and parallel addition of matrices. J. Math. Anal. Appl. **26**, 576–594 (1969). https://doi.org/10.1016/0022-247X(69)90200-5

[3] Benamou, J.-D., Brenier, Y.: A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. Numer. Math. **84**(3), 375–393 (2000). https://doi.org/10.1007/s002110050002

[4] Bhatia, R.: Positive definite matrices, paperback Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ (2007)

[5] Bhatia, R., Gaubert, S., Jain, T.: Matrix versions of the Hellinger distance. Lett. Math. Phys. **109**(8), 1777–1804 (2019). https://doi.org/10.1007/s11005-019-01156-0

[6] Bhatia, R., Holbrook, J.: Riemannian geometry and matrix geometric means. Linear Algebra Appl. **413**(2–3), 594–618 (2006). https://doi.org/10.1016/j.laa.2005.08.025

[7] Bhatia, R., Jain, T., Lim, Y.: Inequalities for the Wasserstein mean of positive definite matrices. Linear Algebra Appl. **576**, 108–123 (2019). https://doi.org/10.1016/j.laa.2018.03.017

[8] Bhatia, R., Jain, T., Lim, Y.: On the Bures-Wasserstein distance between positive definite matrices. Expo. Math. **37**(2), 165–191 (2019). https://doi.org/10.1016/j.exmath.2018.01.002

[9] Carlen, E.: Trace inequalities and quantum entropy: an introductory course. In: Entropy and the quantum, volume 529 of Contemp. Math., pp. 73–140.

Amer. Math. Soc., Providence, RI (2010). https://doi.org/10.1090/conm/529/10428

[10] Hansen, F.: The fast track to Löwner's theorem. Linear Algebra Appl. **438**(11), 4557–4571 (2013). https://doi.org/10.1016/j.laa.2013.01.022

[11] Jordan, R., Kinderlehrer, D., Otto, F.: Free energy and the Fokker-Planck equation. Phys. D **107**(2–4), 265–271 (1997). https://doi.org/10.1016/S0167-2789(97)00093-6

[12] Jordan, R., Kinderlehrer, D., Otto, F.: The route to stability through Fokker-Planck dynamics. In: Differential equations and applications (Hangzhou, 1996), pp. 108–126. Int. Press, Cambridge, MA, [1997]

[13] Jordan, R., Kinderlehrer, D., Otto, F.: The variational formulation of the Fokker-Planck equation. SIAM J. Math. Anal. **29**(1), 1–17 (1998). https://doi.org/10.1137/S0036141096303359

[14] Karcher, H.: Riemannian center of mass and mollifier smoothing. Comm. Pure Appl. Math. **30**(5), 509–541 (1977). https://doi.org/10.1002/cpa.3160300502

[15] Kim, S., Lawson, J., Lim, Y.: The matrix geometric mean of parameterized, weighted arithmetic and harmonic means. Linear Algebra Appl. **435**(9), 2114–2131 (2011). https://doi.org/10.1016/j.laa.2011.04.010

[16] Komálovics, Á, Molnár, L.: On a parametric family of distance measures that includes the Hellinger and the Bures distances. J. Math. Anal. Appl.**529**(2):Paper No. 127226, 31 (2024). https://doi.org/10.1016/j.jmaa.2023.127226

[17] Kubo, F., Ando, T.: Means of positive linear operators. Math. Ann. **246**(3):205–224, 1979/80. https://doi.org/10.1007/BF01371042

[18] Moakher, M.: A differential geometric approach to the geometric mean of symmetric positive-definite matrices. SIAM J. Matrix Anal. Appl. **26**(3), 735–747 (2005). https://doi.org/10.1137/S0895479803436937

[19] Molnár, L.: Bures isometries between density spaces of $C^*$-algebras. Linear Algebra Appl. **557**, 22–33 (2018). https://doi.org/10.1016/j.laa.2018.07.008

[20] Otto, F.: The geometry of dissipative evolution equations: the porous medium equation. Comm. Part. Differ. Equs. **26**(1–2), 101–174 (2001). https://doi.org/10.1081/PDE-100002243

[21] Pitrik, J., Virosztek, D.: Quantum Hellinger distances revisited. Lett. Math. Phys. **110**(8), 2039–2052 (2020). https://doi.org/10.1007/s11005-020-01282-0

[22] Pitrik, J., Virosztek, D.: A divergence center interpretation of general symmetric Kubo-Ando means, and related weighted multivariate operator means. Linear Algebra Appl. **609**, 203–217 (2021). https://doi.org/10.1016/j.laa.2020.09.007

[23] Pusz, W., Woronowicz, S.L.: Functional calculus for sesquilinear forms and the purification map. Rep. Math. Phys. **8**(2), 159–170 (1975). https://doi.org/10.1016/0034-4877(75)90061-0

[24] Takatsu, A.: Wasserstein geometry of Gaussian measures. Osaka J. Math. **48**(4):1005–1026, (2011). URL: http://projecteuclid.org/euclid.ojm/1326291215

[25] Villani, C.: Topics in optimal transportation, vol. 58 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI (2003). https://doi.org/10.1090/gsm/058

Dániel Virosztek (✉)
HUN-REN Alfréd Rényi Institute of Mathematics
Reáltanoda U. 13-15
Budapest H-1053
Hungary
e-mail: `virosztek.daniel@renyi.hu`