isid/ms/2011/02
March 17, 2011
http://www.isid.ac.in/~statmath/eprints

The matrix geometric mean

Rajendra Bhatia
Rajeeva L. Karandikar

Indian Statistical Institute, Delhi Centre
7, SJSS Marg, New Delhi–110 016, India

The matrix geometric mean

Rajendra Bhatia
Indian Statistical Institute,
New Delhi-110 016, India
rbh@isid.ac.in

and

Rajeeva L. Karandikar
Chennai Mathematical Institute,
Siruseri-603103, India
rlk@cmi.ac.in

Abstract An attractive candidate for the geometric mean of m positive definite matrices $A_1, \ldots, A_m$ is their Riemannian barycentre G. One of its important properties, monotonicity in the m arguments, has been established recently by J. Lawson and Y. Lim. We give a much simpler proof of this result, and prove some other inequalities. One of these says that, for every unitarily invariant norm, $|||G|||$ is not bigger than the geometric mean of $|||A_1|||, \ldots, |||A_m|||$.

Mathematics Subject Classification (2010): 15B48, 47A64, 53C20
Keywords: Positive definite matrix, geometric mean, Riemannian metric, barycentre, variance inequality.

1 Introduction

Operator theorists, physicists, engineers and statisticians have long been interested in various averaging operations (means) on positive definite matrices. When just two matrices are involved the theory is very well developed. See the foundational work of Kubo and Ando [10], and the recent exposition in Chapter 4 of [4].

Particularly intriguing has been the notion of *geometric mean*. For two positive definite matrices A and B this is given by an explicit formula

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

credited to Pusz and Woronowicz [14]. For more than two matrices an appropriate definition of a geometric mean with some natural properties remained elusive for long. Progress was made recently by making a connection with differential geometry.

The space $\mathbb{P}$ consisting of positive definite matrices (of a fixed size d) is endowed with a Riemannian metric δ_2 defined as

$$\delta_2(A, B) = \left[\sum_{j=1}^{d} \log^2 \lambda_j(A^{-1}B)\right]^{\frac{1}{2}}, \tag{2}$$

where $\lambda_j(X)$ denote the eigenvalues of X. Any two points A, B of $\mathbb{P}$ can be joined by a unique geodesic, for which a natural parametrisation is given by

$$A\#_t B = A^{\frac{1}{2}}(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})^t A^{\frac{1}{2}}, \quad 0 \le t \le 1. \tag{3}$$

The geometric mean (1) is evidently the midpoint of this geodesic.

With this understanding it is natural to think that the geometric mean of m positive definite matrices $A_1, \ldots, A_m$ should be defined as the "centre" of the convex set spanned by these m points in the metric space $(\mathbb{P}, \delta_2)$. The right candidate for this would seem to be the *barycentre* $G(A_1, \ldots, A_m)$ defined as

$$G(A_1, \ldots, A_m) = \underset{X}{\operatorname{argmin}} \sum_{j=1}^{m} \frac{1}{m}\delta_2^2(X, A_j), \tag{4}$$

the unique point X_0 in $\mathbb{P}$ at which the sum in (4) is minimised. Geometric properties of the space $\mathbb{P}$ as the Riemannian symmetric space $GL(n)/U(n)$ have been studied for long. It is a classical theorem of E. Cartan (see [2] p.234, [8] p.66, [9]) that the minimiser in (4) exists and is unique. M. Moakher [13] proposed (4) as the right candidate for the geometric mean, and obtained many of its interesting properties, including a useful characterisation of G as the unique solution in $\mathbb{P}$ for the equation

$$\sum_{j=1}^{m} \log(X^{-1}A_j) = 0. \tag{5}$$

The same definition for the geometric mean was also proposed by Bhatia and Holbrook [5]. In this paper and the exposé [6] they further highlighted the geometric aspects of the problem. Among other things, they pointed out that another definition of the geometric mean proposed a little earlier by Ando, Li and Mathias [1] had a nice geometric interpretation. This definition, in terms of the limit of an interactive process, leads to a "centre" of the convex set spanned by $A_1, \ldots, A_m$ which is not always the same as the barycentre.

Important in operator theory, though ignored by geometers, is the order relation on $\mathbb{P}$. We say that $A \leq B$, if for all vectors x we have the inequality $\langle Ax, x\rangle \leq \langle Bx, x\rangle$ between inner products. Among the several conditions that a mean is expected to fulfill is *monotonicity* with respect to this order. For G this means that we must have $G(A_1, \ldots, A_m) \leq G(B_1, \ldots, B_m)$ whenever $A_j \leq B_j$ for $1 \leq j \leq m$. The two-variable mean (1) does satisfy this condition. While several facts about $G(A_1, \ldots, A_m)$ were proved in [4], [5], [13], it is only recently that Lawson and Lim [11] have succeeded in establishing this crucial monotonicity property for the case $m \geq 3$.

To prove their theorem, Lawson and Lim have borrowed tools from an unexpected source–the work of Sturm [16] on probability measures on metric spaces of nonpositive curvature. It is known that $(\mathbb{P}, \delta_2)$ is such a space. The principal goal of this paper is to give a vastly simplified proof. We use no measure theory, just some simple counting arguments and basic inequalities for the metric δ_2. In addition to this we obtain some more mean-like properties of $G(A_1, \ldots, A_m)$.

Ando, Li and Mathias [1] listed ten properties that a geometric mean of m matrices should satisfy, and showed that their mean possesses all of them. For the barycentre mean G, many of these properties had been established earlier. With the work of Lawson and Lim [11] it is now known that G too has all the ten properties. Other notions of geometric mean with all the ten properties have been proposed recently [7]. The barycentre mean has been used in diverse applications such as elasticity, signal processing, medical imaging and computer vision. See [11] for some references.

Chapters 4 and 6 of [4] provide a convenient summary of basic results on matrix means. Other facts about matrix analysis that we use can be found in [3].

2 Some inequalities

For the convenience of the reader, we collect some fundamental inequalities that we need for our proofs.

It is a fundamental fact of Loewner's theory [3, Chapter V] that for $0 \le t \le 1$, the function $A \longmapsto A^t$ is operator monotone. Consequently, $A\#_t B$ as defined in (3) is monotone in B. Since $A\#_t B = B\#_{1-t}A$, it is monotone in A as well. So we have the following well-known result.

Proposition 2.1 *Let $A_1 \le B_1$ and $A_2 \le B_2$. Then for all $0 \le t \le 1$ we have*

$$A_1\#_t A_2 \le B_1\#_t B_2.$$

Let $\mathbb{H}$ be the real linear space consisting of all Hermitian matrices (of size d) with the Euclidean norm

$$||T||_2 = (\operatorname{tr} T^*T)^{\frac{1}{2}} = \left(\sum_{i,j} |t_{ij}|^2\right)^{\frac{1}{2}} \tag{6}$$

A fundamental inequality, from which several facts about the metric δ_2 can be derived, is the *exponential metric increasing property* (EMI in brief). This says that the exponential map exp from $(\mathbb{H}, ||\cdot||_2)$ onto $(\mathbb{P}, \delta_2)$ increases distances in general, and preserves distances along rays through the origin. More precisely:

Proposition 2.2 (EMI) *For all S, T in $\mathbb{H}$ we have*

$$\delta_2(e^S, e^T) \ge ||S - T||_2.$$

The two sides are equal when S and T commute. In particular, this is so if $S = \alpha T$ for some real number α.

See [4, pp.203-204]. As a corollary, we see that if $\{A_n\}$ is a sequence in $\mathbb{P}$ such that $\delta_2(A_n, A) \longrightarrow 0$, then $||A_n - A||_2 \longrightarrow 0$. Using this one can see the following:

Proposition 2.3 *Let $\{A_n\}$ and $\{B_n\}$ be two sequences in $\mathbb{P}$, such that $A_n \le B_n$ for all n. Suppose $\delta_2(A_n, A) \longrightarrow 0$ and $\delta_2(B_n, B) \longrightarrow 0$. Then $A \le B$.*

A consequence of the EMI is the following.

Proposition 2.4 *For all A, B, C in $\mathbb{P}$ and for $0 \le t \le 1$ we have*

$$\delta_2^2(C, A\#_t B) \le (1-t)\delta_2^2(C, A) + t\delta_2^2(C, B) - t(1-t)\delta_2^2(A, B). \tag{7}$$

The special case $t = \frac{1}{2}$ of (7) is called the *semiparallelogram law*. See [4, p.207] for a simple proof. Using the special case we can prove (7) inductively for all dyadic rationals t in $[0, 1]$, and then by continuity extend it to all t.

Let $a_1, \ldots, a_m$ be elements of a Hilbert space $\mathcal{H}$ and let $g = \frac{1}{m}(a_1 + \cdots + a_m)$. Then for all $z \in \mathcal{H}$ we have

$$||z - g||^2 = \sum_{j=1}^{m} \frac{1}{m}[\ ||z - a_j||^2 - ||g - a_j||^2\ ]. \tag{8}$$

This can be easily seen by reducing the general case to the special one with $a_1 + a_2 + \cdots + a_m = 0$.

In the space $(\mathbb{P}, \delta_2)$ the equality (8) is replaced by an inequality:

Theorem 2.5 (Variance Inequality) Let $A_1, A_2, \ldots, A_m$ *be any elements of* $\mathbb{P}$ *and let* $G = G(A_1, \ldots, A_m)$. *Then for all* $Z \in \mathbb{P}$ *we have*

$$\delta_2^2(Z, G) \leq \sum_{j=1}^{m} \frac{1}{m}[\delta_2^2(Z, A_j) - \delta_2^2(G, A_j)]. \tag{9}$$

Proof For every nonsingular matrix X, the map Γ_X, defined on $\mathbb{P}$ as $\Gamma_X(A) = X^*AX$, is an isometry for the metric δ_2. See [4,p.202]. Use this fact with $X = G^{-\frac{1}{2}}$ to see that it suffices to prove (9) in the special situation when $G = I$. In this case the desired inequality (9) becomes

$$\delta_2^2(Z, I) \leq \sum_{j=1}^{m} \frac{1}{m}[\delta_2^2(Z, A_j) - \delta_2^2(I, Aj)]. \tag{10}$$

We have remarked that G is the unique positive definite solution of the equation (5). When $G = I$ this reduces to

$$\sum_{j=1}^{m} \log(A_j) = 0.$$

Now, $\log(A_j)$ are points in the Hilbert space $(\mathbb{H}, ||\cdot||_2)$, and their sum is 0. So from (8) we get

$$||\log Z||_2^2 = \sum_{j=1}^{m} \frac{1}{m}[\ ||\log Z - \log A_j||_2^2 - ||\log A_j||_2^2\ ]. \tag{11}$$

Using Proposition 2.2 (EMI) we see that

$$||\log Z||_2 = \delta_2(Z, I), \quad ||\log A_j||_2 = \delta_2(A_j, I),$$

and

$$||\log Z - \log A_j||_2 \leq \delta_2(Z, A_j).$$

So the inequality (10) follows from (11). $\square$

A more general version of the inequality (9) with integrals in place of sums is Proposition 4.4 of Sturm [16]. The discrete version is adequate for our purpose. The role of EMI is clearly brought out in our proof.

3 Monotonicity of the geometric mean

Let $A_1, \ldots, A_m$ be elements of $\mathbb{P}$, let $G = G(A_1, \ldots, A_m)$, and

$$\alpha = \sum_{j=1}^{m} \frac{1}{m} \delta_2^2(G, A_j). \tag{12}$$

For $n \geq 1$ let $\mathcal{J}_n$ be the set of all ordered n-tuples $(j_1, j_2, \ldots, j_n)$, with $j_k \in \{1, 2, \ldots, m\}$. This is a set with m^n elements. For each element of this set we define a "mean" $\mathcal{M}_n(j_1, \ldots, j_n)$ of the given matrices $A_1, \ldots, A_m$ by the following inductive procedure:

$$\mathcal{M}_1(j) = A_j, \text{ for all } j \in \mathcal{J}_1;$$

$$\mathcal{M}_{n+1}(j_1, \ldots, j_n, k) = \mathcal{M}_n(j_1, \ldots, j_n) \#_{\frac{1}{n+1}} A_k,$$

for all $(j_1, \ldots, j_n)$ in $\mathcal{J}_n$ and k in $\mathcal{J}_1$. The heart of our proof is in the following estimate.

Theorem 3.1 *For every n we have*

$$\frac{1}{m^n} \sum_{(j_1, \ldots, j_n) \in \mathcal{J}_n} \delta_2^2(G, \mathcal{M}_n(j_1, \ldots, j_n)) \leq \frac{1}{n} \alpha. \tag{13}$$

Proof We prove this by induction on n. When $n = 1$, the two sides of (13) are equal. Assuming that (13) is true for n, we will show that it is true for $n+1$. We have

$$\sum_{k=1}^{m} \delta_2^2(G, \mathcal{M}_{n+1}(j_1, \ldots, j_n, k)) = \sum_{k=1}^{m} \delta_2^2(G, \mathcal{M}_n(j_1, \ldots, j_n) \#_{\frac{1}{n+1}} A_k). \tag{14}$$

Using Proposition 2.4 we see that the right hand side of (14) is less than or equal to

$$\sum_{k=1}^{m} \left[\frac{n}{n+1} \delta_2^2(G, \mathcal{M}_n(j_1, \ldots, j_n)) + \frac{1}{n+1} \delta_2^2(G, A_k) - \frac{n}{(n+1)^2} \delta_2^2(\mathcal{M}_n(j_1, \ldots, j_n), A_k) \right]$$

$$= \frac{mn}{n+1} \delta_2^2(G, \mathcal{M}_n(j_1, \ldots, j_n)) + \frac{m}{n+1} \alpha - \frac{n}{(n+1)^2} \sum_{k=1}^{m} \delta_2^2(\mathcal{M}_n(j_1, \ldots, j_n), A_k). \tag{15}$$

By the Variance Inequality (9) we have

$$m\delta_2^2(\mathcal{M}_n(j_1,\dots,j_n),G)+\sum_{k=1}^m \delta_2^2(G,A_k)\le \sum_{k=1}^m \delta_2^2(\mathcal{M}_n(j_1,\dots,j_n),A_k). \tag{16}$$

From (14), (15) and (16) we obtain

$$\begin{aligned}\frac{1}{m}\sum_{k=1}^m \delta_2^2(G,\mathcal{M}_{n+1}(j_1,\dots,j_n,k)) &\le \frac{n}{n+1}\delta_2^2(G,\mathcal{M}_n(j_1,\dots,j_n))+\frac{1}{n+1}\alpha\\ &\quad -\frac{n}{(n+1)^2}[\delta_2^2(G,\mathcal{M}_n(j_1,\dots,j_n))+\alpha]\\ &= \frac{n^2}{(n+1)^2}\delta_2^2(G,\mathcal{M}_n(j_1,\dots,j_n))+\frac{1}{(n+1)^2}\alpha.\end{aligned} \tag{17}$$

Sum up the two sides of (17) over $(j_1,\dots,j_n)\in\mathcal{J}_n$, divide by m^n, and use the induction hypothesis. This shows that

$$\frac{1}{m^{n+1}}\sum_{(j_1,j_2,\dots,j_{n+1})\in\mathcal{J}_{n+1}} \delta_2^2(G,\mathcal{M}_{n+1}(j_1,j_2,\dots,j_{n+1}))\le \frac{n^2}{(n+1)^2}\frac{\alpha}{n}+\frac{1}{(n+1)^2}\alpha=\frac{\alpha}{n+1}.$$

This shows that the inequality (13) is valid for all n. □

Now we are in a position to give our proof of the following theorem first proved by Lawson and Lim.

Theorem 3.2 *Let $A_1,\dots,A_m$ and $A_1',\dots,A_m'$ be positive definite matrices such that $A_j\le A_j'$ for $1\le j\le m$. Then*

$$G(A_1,\dots,A_m)\le G(A_1',\dots,A_m'). \tag{18}$$

Proof Let us denote by G and G' the two sides of (18), by α and α' the sums in (12) corresponding to the m-tuples $(A_1,\dots,A_m)$ and $(A_1',\dots,A_m')$, and by $\mathcal{M}_n$ and $\mathcal{M}_n'$ the "means" defined before Theorem 3.1.

Let ε be any given positive number, and choose a positive integer n such that $\frac{\alpha}{n}<\frac{\varepsilon^2}{4}$ and $\frac{\alpha'}{n}<\frac{\varepsilon^2}{4}$. Let $\mathcal{I}_n$ be the subset of $\mathcal{J}_n$ such that $\delta_2(G,\mathcal{M}_n(j_1,\dots,j_n))>\varepsilon$ if $(j_1,\dots,j_n)\in\mathcal{I}_n$. From the inequality (13) we can conclude that the cardinality of the set $\mathcal{I}_n$ cannot be bigger than one fourth of the cardinality of $\mathcal{J}_n$. By the same reasoning, the cardinality of the subset $\mathcal{I}_n'$ consisting of indices $(j_1,\dots,j_n)$ for which $\delta_2(G',\mathcal{M}_n'(j_1,\dots,j_n))>\varepsilon$ cannot be bigger than one fourth of the cardinality of $\mathcal{J}_n$.

This shows that the intersection of the two sets

$$\mathcal{C}_1=\{(j_1,\dots,j_n)\in\mathcal{J}_n:\delta_2(G,\mathcal{M}_n(j_1,\dots,j_n))<\varepsilon\}$$

and

$$\mathcal{C}_2 = \{(j_1, \ldots, j_n) \in \mathcal{J}_n : \delta_2(G', \mathcal{M}'_n(j_1, \ldots, j_n)) < \varepsilon\}$$

contains at least one element. (In fact it contains at least $\frac{m^n}{2}$ elements.)

So let $(j_1^*, \ldots, j_n^*) \in \mathcal{C}_1 \cap \mathcal{C}_2$. Using Proposition 2.1 we see that

$$\mathcal{M}_n(j_1^*, \ldots, j_n^*) \leq \mathcal{M}'_n(j_1^*, \ldots, j_n^*).$$

For $\varepsilon = \frac{1}{k}$, choose n and $(j_1^*, \ldots, j_n^*)$ as described above and let $D_k = \mathcal{M}_n(j_1^*, \ldots, j_n^*)$ and $D'_k = \mathcal{M}'_n(j_1^*, \ldots, j_n^*)$. Then we have

$$\delta_2(G, D_k) \leq \frac{1}{k}, \; \delta_2(G', D'_k) \leq \frac{1}{k}, \text{ and } D_k \leq D'_k.$$

Using Proposition 2.3 we conclude that $G \leq G'$. $\square$

Remark In essence we have proved and used a weak law of large numbers in this special situation. Sturm [16] proves a weak law and a strong law in a more general set up of probability measures on metric spaces of nonpositive curvature. Lawson and Lim [11] derive Theorem 3.2 from this more difficult strong law. We have avoided the complications of both the strong law and of measures.

Let us note here that our arguments give simplifications of the proofs of two other important properties of the geometric mean. The first is *joint concavity.* This says that for any two m-tuples $(A_1, \ldots, A_m)$ and $(A'_1, \ldots, A'_m)$ of positive definite matrices, we have for all $0 \leq t \leq 1$

$$(1-t)G(A_1, \ldots, A_n) + tG(A'_1, \ldots, A'_n) \leq G((1-t)A_1 + tA'_1, \ldots, (1-t)A_n + tA'_n). \quad (19)$$

This concavity property is known to hold for the function $f(A, B) = A\#_s B$ for each $0 \leq s \leq 1$. From there it is carried over to the means $\mathcal{M}_n(A_1, \ldots, A_m)$ and then to $G(A_1, \ldots, A_m)$. The argument is similar to the one we have used in Theorem 3.2.

The second property is *continuity.* This follows from the interesting inequality given in the next result first proved in [11].

Theorem 3.3 *For all $A_1, \ldots, A_m$ and $A'_1, \ldots, A'_m$ in $\mathbb{P}$*

$$\delta_2(G(A_1, \ldots, A_m), G(A'_1, \ldots, A'_m)) \leq \sum_{j=1}^{m} \frac{1}{m} \delta_2(A_j, A'_j). \quad (20)$$

Proof We will use the following well-known convexity property of the metric: for any four points A_1, A_2, B_1, B_2 in $\mathbb{P}$ we have for all $0 \leq t \leq 1$

$$\delta_2(A_1 \#_t A_2, B_1 \#_t B_2) \leq (1-t)\delta_2(A_1, B_1) + t\delta_2(A_2, B_2). \quad (21)$$

See Corollary 6.1.11 [4]. Let $\mathcal{M}_n$ and $\mathcal{M}'_n$ be the "means" defined at the beginning of the proof of Theorem 3.2. Then for each $(j_1,\ldots,j_n)\in\mathcal{J}_n$ we have, using (21) repeatedly,

$$\begin{aligned}
&\delta_2(\mathcal{M}_n(j_1,\ldots,j_n),\mathcal{M}'_n(j_1,\ldots,j_n))\\
\leq\ & \frac{n-1}{n}\delta_2(\mathcal{M}_{n-1}(j_1,\ldots,j_{n-1}),\mathcal{M}'_{n-1}(j_1,\ldots,j_{n-1}))+\frac{1}{n}\delta_2(A_{j_n},A'_{j_n})\\
\leq\ & \frac{n-2}{n}\delta_2(\mathcal{M}_{n-2}(j_1,\ldots,j_{n-2}),\mathcal{M}'_{n-2}(j_1,\ldots,j_{n-2}))+\frac{1}{n}\delta_2(A_{j_{n-1}},A'_{j_{n-1}})+\frac{1}{n}\delta_2(A_{j_n},A'_{j_n})\\
\leq\ & \ldots\ldots\\
\leq\ & \frac{1}{n}\sum_{k=1}^{n}\delta_2(A_{j_k},A'_{j_k}).
\end{aligned}$$

Now we will show that given an ε, for large n, there are lots of choices of $(j_1,\ldots,j_n)$ for which the last sum is smaller than $\frac{1}{m}\sum_{k=1}^{m}\delta_2(A_k,A'_k)+\varepsilon$. With this aim, let

$$\theta=\sum_{j=1}^{m}\frac{1}{m}\delta_2(A_j,A'_j)$$

and

$$\gamma=\sum_{j=1}^{m}\frac{1}{m}\big(\delta_2(A_j,A'_j)-\theta\big)^2.$$

It can be shown that

$$\frac{1}{m^n}\sum_{(j_1,\ldots,j_n)\in\mathcal{J}_n}\Big(\frac{1}{n}\sum_{k=1}^{n}\delta_2(A_{j_k},A'_{j_k})-\theta\Big)^2\leq\frac{1}{n}\gamma.$$

The steps are similar to (and far simpler than) the proof of (13). Then, given an ε, we can choose n such that $\frac{\gamma}{n}\leq\frac{\varepsilon^2}{4}$, and then for this choice the set

$$\mathcal{C}_3=\Big\{(j_1,\ldots,j_n)\in\mathcal{J}_n:\Big|\frac{1}{n}\sum_{k=1}^{n}\delta_2(A_{j_k},A'_{j_k})-\theta\Big|<\varepsilon\Big\}$$

contains at least $\frac{3}{4}m^n$ elements.

Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be the sets defined in the proof of Theorem 3.3. We have shown that we can choose n such that cardinality of each of $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$ is at least $\frac{3}{4}m^n$. Thus

$$\mathcal{C}_1\cap\mathcal{C}_2\cap\mathcal{C}_3$$

is non-empty (its cardinality is at least $\frac{1}{4}m^n$). Let $(j_1^*, \ldots, j_n^*)$ be a common element in $\mathcal{C}_1$, $\mathcal{C}_2$ and $\mathcal{C}_3$ and let $M_n = \mathcal{M}_n(j_1^*, \ldots, j_n^*)$, $M_n' = \mathcal{M}_n'(j_1^*, \ldots, j_n^*)$ and $\theta_n = \frac{1}{n}\sum_{k=1}^n \delta_2(A_{j_k^*}, A'_{j_k^*})$.

Then for this choice of n and $(j_1^*, \ldots, j_n^*)$, we have seen that

$$\delta_2(G(A_1, \ldots, A_m), M_n) \leq \varepsilon,$$

$$\delta_2(G(A_1', \ldots, A_m'), M_n') \leq \varepsilon,$$

and

$$|\theta_n - \theta| \leq \varepsilon.$$

Thus we conclude

$$\delta_2(G(A_1, \ldots, A_m), G(A_1', \ldots, A_m')) \leq \theta + 3\varepsilon.$$

Since ε is arbitrary, this completes the proof of (20). $\square$

The inequalities (19) and (20) have been proved by Lawson and Lim [11], using the full force of Sturm's theorems. For the proof of (20) they make use of the Wasserstein distance between probability measures. Our proofs are much simpler.

4 More properties of the geometric mean

The geometric mean (4) enjoys several other properties. Some of them are shown in this section.

Theorem 4.1 *Let Φ be a positive unital linear map from the matrix algebra $\mathbb{M}(d)$ to $\mathbb{M}(k)$. Then for all positive definite matrices $A_1, \ldots, A_m$ in $\mathbb{M}(d)$ we have*

$$\Phi(G(A_1, \ldots, A_m)) \leq G(\Phi(A_1), \ldots, \Phi(A_m)). \tag{22}$$

Proof It is well known that the two variable mean (1) has this property. See Theorem 4.1.5 in [4]. From this the property is inherited by the "mean" (3); i.e., for all $0 \leq t \leq 1$ we have $\Phi(A\#_t B) \leq \Phi(A)\#_t\Phi(B)$. Now let $\mathcal{M}_n(j_1, \ldots, j_n)$ be the "means" constructed from $A_1, \ldots, A_m$ in Section 3, and let $\mathcal{M}_n^{\Phi}(j_1, \ldots, j_n)$ be the corresponding objects obtained by replacing $A_1, \ldots, A_m$ with $\Phi(A_1), \ldots, \Phi(A_m)$. Then we have

$$\Phi(\mathcal{M}_n(j_1, \ldots, j_n)) \leq \mathcal{M}_n^{\Phi}(j_1, \ldots, j_n)$$

for all $(j_1, \ldots, j_n) \in \mathcal{J}_n$. From this we obtain the inequality (22) by the argument used in the proof of Theorem 3.2. $\square$

Applying this to the positive linear functional $\phi(A) = \langle Ax, x\rangle$, where x is a unit vector in $\mathbb{C}^d$, we obtain the following result proved by Yamazaki [17].

Corollary 4.2 *For all positive definite $A_1, \ldots, A_m$ and all vectors x*, we have

$$\langle G(A_1, \ldots, A_m)x, x\rangle \leq \left(\prod_{j=1}^{m} \langle A_j x, x\rangle\right)^{\frac{1}{m}}. \tag{23}$$

Let $||A|| = \sup_{||x||=1} ||Ax||$ be the usual operator norm on $\mathbb{M}(d)$. Then from (23) we obtain:

Theorem 4.3 *For all positive definite matrices $A_1, \ldots, A_m$ we have*

$$||G(A_1, \ldots, A_m)|| \leq \prod_{j=1}^{m} ||A_j||^{\frac{1}{m}} \tag{24}$$

Let $1 \leq k \leq d$ and let $\Lambda^k(T)$ denote the kth antisymmetric tensor power of a $d \times d$ matrix T. If A and B are positive definite matrices, then for all $0 \leq t \leq 1$ we have

$$\Lambda^k(A \#_t B) = \Lambda^k(A) \#_t \Lambda^k(B). \tag{25}$$

Once again, using the by now familiar argument, we can deduce the following:

Theorem 4.4 *For all $1 \leq k \leq d$, and for all $d \times d$ positive definite matrices, we have*

$$\Lambda^k G(A_1, \ldots, A_m) = G(\Lambda^k A_1, \ldots, \Lambda^k A_m). \tag{26}$$

As a corollary we have the following result proved by Yamazaki [17].

Corollary 4.5 *We have*

$$\det G(A_1, \ldots, A_m) = \left(\prod_{j=1}^{m} \det A_j\right)^{\frac{1}{m}} \tag{27}$$

Let $\lambda_1(T) \geq \lambda_2(T) \geq \ldots \geq \lambda_N(T)$ be the eigenvalues of an $N \times N$ positive definite matrix T. The inequality (24) says that

$$\lambda_1(G(A_1, \ldots, A_m)) \leq \prod_{j=1}^{m} \lambda_1^{\frac{1}{m}}(A_j). \tag{28}$$

Using the fact

$$\lambda_1(\Lambda^k T) = \prod_{i=1}^{k} \lambda_i(T),$$

we obtain from (26) the inequality

$$\prod_{i=1}^{k} \lambda_i(G(A_1, \ldots, A_m)) \leq \prod_{j=1}^{m} \prod_{i=1}^{k} \lambda_i^{\frac{1}{m}}(A_j), \tag{29}$$

for all $k = 1, 2, \ldots, d$. When $k = d$, this is an equality. Relations (28) and (27) are special cases of this for $k = d$ and $k = 1$, respectively.

Interchange the order of the products in (29), and then use a standard result from the theory of majorisation [3,p.42] to obtain

$$\sum_{i=1}^{k} \lambda_i(G(A_1, \ldots, A_m)) \leq \sum_{i=1}^{k} \prod_{j=1}^{m} \lambda_i^{\frac{1}{m}}(A_j). \tag{30}$$

Now recall Hölder's inequality for any array. If a_{ij}, $1 \leq i \leq k$, $1 \leq j \leq m$, is an array of positive numbers, then

$$\sum_{i=1}^{k} \left(\prod_{j=1}^{m} a_{ij} \right)^{\frac{1}{m}} \leq \prod_{j=1}^{m} \left(\sum_{i=1}^{k} a_{ij} \right)^{\frac{1}{m}}. \tag{31}$$

See [15,p.152]. Using this we obtain from (30)

$$\sum_{i=1}^{k} \lambda_i(G(A_1, \ldots, A_m)) \leq \prod_{j=1}^{m} \left[\sum_{i=1}^{k} \lambda_i(A_j) \right]^{\frac{1}{m}}. \tag{32}$$

Now recall that the Ky Fan k-norm of a matrix T is defined as $||T||_k = s_1(T) + \cdots + s_k(T)$, the sum of the top k singular values of T. Thus (32) can be stated in another way as

$$||G(A_1, \ldots, A_m)||_k \leq \prod_{j=1}^{m} ||A_j||_k^{\frac{1}{m}}, \quad 1 \leq k \leq d. \tag{33}$$

As is well-known, the Ky Fan norms play a very special role in the theory of majorisation [3]. Using a theorem of Li and Mathias [12] we obtain from (33):

Theorem 4.6 *Let $A_1, \ldots, A_m$ be positive definite matrices. Then for every unitarily invariant norm $|||\cdot|||$ we have*

$$|||G(A_1, \ldots, A_m)||| \leq \prod_{j=1}^{m} |||A_j|||^{\frac{1}{m}}. \tag{34}$$

Theorem 4.3 is a special case of this. Among unitarily invariant norms are the Schatten p-norms, much used in operator theory and physics.

5 The weighted geometric mean

A *weight vector* $w = (w_1, \ldots, w_m)$ is an m-tuple of positive numbers w_j, such that $w_1 + \cdots + w_m = 1$. The *weighted geometric mean* of positive definite matrices $A_1, \ldots, A_m$ is defined as the solution of the minimisation problem

$$G(w; A_1, A_2, \ldots, A_m) = \underset{X}{\operatorname{argmin}} \sum_{j=1}^{m} w_j \delta_2^2(X, A_j). \tag{35}$$

It is known that the problem has a unique solution in $\mathbb{P}$, and that is also the unique solution of the matrix equation

$$\sum_{j=1}^{m} w_j \log(X^{-1} A_j) = 0. \tag{36}$$

The mean $G(A_1, \ldots, A_m)$ corresponds to the special choice $w_j = 1/m$, $1 \le j \le m$.

All our results can be proved for the weighted geometric mean. In fact, with continuity of $G(A_1, \ldots, A_m)$ in hand, the more general case can be deduced from the special one. We indicate this briefly.

Suppose all weights w_j are rational. Choosing a common denominator, we can assume that $w_j = p_j/q$, $1 \le j \le m$. Then we consider a q-tuple of matrices in which each A_j occurs with multiplicity p_j. This brings us to the case we have already studied. From rational weights we go to real ones by a continuity argument.

Thus the mean $G(w; A_1, \ldots, A_m)$ is monotone and jointly concave in the variables $A_1, \ldots, A_m$.

Most of our inequalities have a general version in which the term $\frac{1}{m}$ is replaced by w_j at the appropriate places. Thus, for example, instead of (20) we have the inequality

$$\delta_2(G(w; A_1, \ldots, A_m), G(w; A_1', \ldots, A_m')) \le \sum_{j=1}^{k} w_j \delta_2(A_j, A_j'), \tag{37}$$

and instead of (34) we have

$$|||G(w; A_1, \ldots, A_m)||| \le \prod_{j=1}^{m} |||A_j|||^{w_j}, \tag{38}$$

for every unitarily invariant norm.

Lawson-Lim [11], Yamazaki [17] and some other authors prove their results in the weighted setting. Our argument offers an easy transition from the equal-weights case to the general one.

An alternative approach is the following. When $m = 2$ and $w = (1-t, t)$ for some $0 < t < 1$, then $G(w; A, B) = A\#_t B$. Properties such as monotonicity, joint concavity and continuity, for this mean are very well known. In the several variable case with weights $w = (w_1, \ldots, w_m)$ and matrices $A_1, \ldots, A_m$ we can follow the arguments used in Section 3. The modification would be that we now define the "means" $\mathcal{M}_n(j_1, \ldots, j_n)$ inductively as follows:

$$\mathcal{M}_1(j) = A_j, \quad \text{for all} \quad j \in \mathcal{J}_1$$

$$\mathcal{M}_{n+1}(j_1, \ldots, j_n; k) = \mathcal{M}_n(j_1, \ldots, j_n)\#_t A_k,$$

for all $(j_1, \ldots, j_n) \in \mathcal{J}_n$, where

$$t = \frac{w_k}{w_{j_1} + \cdots + w_{j_n} + w_k}.$$

Now with

$$\alpha = \sum_{j=1}^{m} w_j \delta_2^2(G(w; A_1, \ldots, A_m), A_j)$$

we can establish the weighted version of Theorem 3.1, and then use it to prove the corresponding version of Theorems 3.2 and 3.3. Results of Section 4 can also be obtained using this argument.

References

1. T. Ando, C. K. Li and R. Mathias, *Geometric means,* Linear Algebra Appl. **385** (2004), 305-334.

2. M. Berger, *A Panoramic View of Riemannian Geometry,* Springer, 2003.

3. R. Bhatia, *Matrix Analysis,* Springer, 1997.

4. R. Bhatia, *Positive Definite Matrices,* Princeton University Press, 2007.

5. R. Bhatia and J. Holbrook, *Riemannian geometry and matrix geometric means,* Linear Algebra Appl. **413** (2006), 594-618.

6. R. Bhatia and J. Holbrook, *Noncommutative geometric means,* Math. Intelligencer **28** (2006), 32-39.

7. D. Bini, B. Meini and F. Poloni, *An effective matrix geometric mean satisfying the Ando-Li-Mathias properties,* Math. Comp. **79** (2010), 437-452.

8. J. Jost, *Nonpositive Curvature: Geometric and Analytic Aspects*, Lectures in Mathematics ETH Zürich, Birkhäuser, 1997.

9. H. Karcher, *Riemannian center of mass and mollifier smoothing,* Comm. Pure Appl. Math. **30** (1977), 509-541.

10. F. Kubo and T. Ando, *Means of positive linear operators,* Math. Ann., **246** (1980), 205-224.

11. J. Lawson and Y. Lim, *Monotonic properties of the least squares mean,* Math. Ann. to appear.

12. C.-K. Li and R. Mathias, *Generalizations of Ky Fan's dominance theorem,* SIAM J. Matrix Anal. Appl., **19** (1998) 99-106.

13. M. Moakher, *A differential geometric approach to the geometric mean of symmetric positive-definite matrices,* SIAM J. Matrix Anal. Appl. **26** (2005), 735-747.

14. W. Pusz and S. L. Woronowicz, *Functional calculus for sesquilinear forms and the purification map,* Rep. Math. Phys., **8** (1975), 159-170.

15. J. M. Steele, *The Cauchy-Schwarz Master Class,* Cambridge University Press, 2004.

16. K.-T. Sturm, *Probability measures on metric spaces of nonpositive curvature,* in*: Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces,* Eds. P. Auscher et. al., Contemp. Math. 338, Amer. Math. Soc., 2003.

17. T. Yamazaki, *On properties of geometric mean of n-operators via Riemannian metric,* preprint.