Chapter 2
The Riemannian Mean of Positive Matrices

Rajendra Bhatia

2.1 Introduction

Recent work in the study of the geometric mean of positive definite matrices has seen the coming together of several subjects: matrix analysis, operator theory, differential geometry (Riemannian and Finsler), probability and numerical analysis. At the same time the range of its applications has grown from physics and electrical engineering (the two areas in which the subject had its beginnings) to include radar data processing, medical imaging, elasticity, statistics and machine learning.

This article, based on my talk at the Indo-French Seminar on Matrix Information Geometries, is a partial view of the arena from the perspective of matrix analysis. There has been striking progress on one of the problems raised in that talk, and I report on that as well.

A pertinent reference for the theory of matrix means is [8], Chaps. 4 and 6. General facts on matrix analysis used here can be found in [6].

2.2 The Binary Geometric Mean

Let $\mathbb{R}_+$ be the set of positive numbers. A *mean* is a function $m : \mathbb{R}_+ \times \mathbb{R}_+ \longrightarrow \mathbb{R}_+$ that satisfies the following conditions

(i) $m(a,b) = m(b,a)$.
(ii) $\min(a,b) \le m(a,b) \le \max(a,b)$.
(iii) $m(\alpha a, \alpha b) \le \alpha m(a,b)$ for all $\alpha > 0$.
(iv) $a \le a' \Rightarrow m(a,b) \le m(a',b)$.

R. Bhatia (✉)
Theoretical Statistics and Mathematics Unit, Indian Statistical Institute,
7, S. J. S. Sansanwal Marg, New Delhi-110016, India
e-mail: rbh@isid.ac.in

F. Nielsen and R. Bhatia (eds.), *Matrix Information Geometry*,
DOI: 10.1007/978-3-642-30232-9_2, © Springer-Verlag Berlin Heidelberg 2013

(v) m is continuous.

Other requirements may be imposed, if needed, in a particular context. The most familiar examples of means are the arithmetic, geometric and harmonic means, defined as

$$\frac{a+b}{2}, \quad \sqrt{ab}, \quad \left(\frac{a^{-1}+b^{-1}}{2}\right)^{-1}, \tag{2.1}$$

respectively. There are several others such as the *logarithmic mean* defined as

$$L(a,b) = \frac{a-b}{\log a - \log b} = \int_0^1 a^{1-t} b^t \, dt, \tag{2.2}$$

much used in heat flow problems; and the binomial means defined as

$$B_p(a,b) = \left(\frac{a^p+b^p}{2}\right)^{1/p}, -\infty < p < \infty. \tag{2.3}$$

The limits

$$\lim_{p\to 0} \quad B_p(a,b) = \sqrt{ab} \tag{2.4}$$

$$\lim_{p\to \infty} \quad B_p(a,b) = \max(a,b) \tag{2.5}$$

$$\lim_{p\to -\infty} \quad B_p(a,b) = \min(a,b) \tag{2.6}$$

are also means.

In various contexts we wish to have a notion of a mean of two positive definite (positive, for short) matrices. Several interesting problems arise. The first of these is that matrix multiplication is not commutative, and the second that the order relation $A \le B$ on positive matrices has some peculiar features. We say that $A \le B$ if $B - A$ is positive semidefinite. Then $A \le B$ does not necessarily imply $A^2 \le B^2$.

Let $\mathbb{P}(n)$ be the set of $n \times n$ positive matrices. Imitating the five conditions above we could say that a *matrix mean* is a map $M : \mathbb{P}(n) \times \mathbb{P}(n) \to \mathbb{P}(n)$ that satisfies the following conditions:

(i)′ $M(A,B) = M(B,A)$.
(ii)′ If $A \le B$, then $A \le M(A,B) \le B$.
(iii)′ $M(X^*AX, X^*BX) = X^*M(A,B)X$, for all nonsingular matrices X. (Here X^* is the conjugate transpose of X).
(iv)′ $A \le A' \Rightarrow M(A,B) \le M(A',B)$.
(v)′ M is continuous.

The arithmetic and the harmonic means defined, respectively as

$$\frac{A+B}{2}, \quad \left(\frac{A^{-1}+B^{-1}}{2}\right)^{-1}$$

do have the five properties listed above. How about the geometric mean? The matrix $A^{1/2}B^{1/2}$ is not even Hermitian, let alone positive, unless A and B commute. We could imitate the relation (2.4) and consider

$$\lim_{p\to 0}\left(\frac{A^p+B^p}{2}\right)^{1/p}, \tag{2.7}$$

or

$$\exp\left(\frac{\log A+\log B}{2}\right). \tag{2.8}$$

These matrices are positive, but they do not have either of the properties (iii)$'$ and (iv)$'$. (It is well known that the exponential map is not order-preserving, and $A \mapsto A^t$ is order-preserving if and only if $0 \le t \le 1$). It is known that the expressions (2.7) and (2.8) represent the same matrix. In some contexts this matrix is used as a "geometric mean".

The definition that works is

$$A\#B := A^{1/2}\left(A^{-1/2}\,B\,A^{-1/2}\right)^{1/2}A^{1/2}. \tag{2.9}$$

Note that if A and B commute, then this reduces to $A^{1/2}B^{1/2}$. It can be shown that $M(A,B)=A\#B$ has all the properties (i)$'$–(v)$'$. Choosing $X=A^{-1/2}$ in (iii)$'$ one sees that this is the only natural definition of a geometric mean!

Note further that

$$(A\#B)^{-1}=A^{-1}\#B^{-1}, \tag{2.10}$$

which is a desirable property for a geometric mean, and that

$$\det(A\#B)=(\det A\ \det B)^{1/2}. \tag{2.11}$$

The definition (2.9) occurs first in a paper of Pusz and Woronowicz [28] dealing with problems of mathematical physics and operator algebras. It turns out that the matrix (2.9) is the unique positive solution of the Riccati equation

$$XA^{-1}X=B, \tag{2.12}$$

and that can serve as another definition of the geometric mean. In the electrical engineering literature there were other definitions before [28]. The product of two positive matrices has positive eigenvalues. Let $(A^{-1}B)^{1/2}$ be the square root of $A^{-1}B$ that has positive eigenvalues. Then the matrix

$$A(A^{-1}B)^{1/2} \tag{2.13}$$

turns out to be equal to the one in (2.9). This matrix was introduced in [15] as the geometric mean of A and B.

In 1979 T. Ando published a very important paper [1] that brought the geometric mean to the attention of a large community. Among other things, Ando showed that among all Hermitian X for which the 2×2 block matrix $\begin{bmatrix} A & X \\ X & B \end{bmatrix}$ is positive there is a maximum, and this maximum is equal to the geometric mean. In other words,

$$A\#B = \max \left\{ X : \begin{bmatrix} A & X \\ X & B \end{bmatrix} \geq 0 \right\}. \tag{2.14}$$

Ando used this characterisation to prove several striking results about convexity of some matrix functions that are important in matrix analysis and quantum theory. He highlighted the inequality between the harmonic, geometric and arithmetic means:

$$\left(\frac{A^{-1} + B^{-1}}{2} \right)^{-1} \leq A\#B \leq \frac{A+B}{2}, \tag{2.15}$$

and the fact that

$$A\#B \text{ is a jointly concave function of } A \text{ and } B. \tag{2.16}$$

We remark here that the matrix $\begin{bmatrix} A & X \\ X & B \end{bmatrix}$ is positive if and only if there exists a contraction K such that $X = A^{1/2} K B^{1/2}$. The maximal X is characterised by the fact that the K occurring here is unitary. In other words

$$A\#B = A^{1/2} U B^{1/2}, \tag{2.17}$$

where U is unitary, and this condition determines $A\#B$.

The paper of Ando was followed by the foundational work of Kubo and Ando [19] where an axiomatic framework is laid down for a general theory of binary matrix means.

With the success of this work it was natural to look for a good definition of a geometric mean of more than two positive matrices. This turned out to be a tricky problem resisting solution for nearly 25 years. Once again the arithmetic and the harmonic means of m positive matrices can be defined in the obvious way as

$$\frac{1}{m} \sum_{j=1}^{m} A_j \quad \text{and} \quad \left(\frac{1}{m} \sum_{j=1}^{m} A_j^{-1} \right)^{-1},$$

respectively. None of the different ways of defining the geometric mean of two matrices given above can be successfully imitated to yield a good generalisation to the case of m matrices.

This problem has been resolved recently. The approach involves some differential geometry. This is briefly explained in the next section.

2.3 The Differential Geometry Connection

Let $\mathbb{H}(n)$ be the real vector space consisting of $n \times n$ Hermitian matrices equipped with the inner product $\langle A, B\rangle = \operatorname{tr} A^*B$, and the associated norm $\|A\|_2 = (\operatorname{tr} A^*A)^{1/2}$. The exponential map

$$\exp : \mathbb{H}(n) \to \mathbb{P}(n)$$

is a bijection, and induces on $\mathbb{P}(n)$ a Riemannian metric structure. The induced metric on $\mathbb{P}(n)$ is

$$\begin{aligned}\delta_2(A, B) &= \|\log(A^{-1/2}BA^{-1/2})\|_2 \\ &= \left(\sum_{i=1}^{n} \log^2 \lambda_i(A^{-1}B)\right)^{1/2},\end{aligned} \tag{2.18}$$

where $\lambda_i(A^{-1}B)$ are the eigenvalues of $A^{-1}B$.

This metric has several interesting properties:

$$\delta_2(X^*AX,\ X^*BX) = \delta_2(A, B), \tag{2.19}$$

for all $X \in GL(n)$, and

$$\delta_2(A^{-1}, B^{-1}) = \delta_2(A, B). \tag{2.20}$$

A useful consequence of (2.19) is

$$\delta_2(A, B) = \delta_2(I, A^{-1/2}BA^{-1/2}). \tag{2.21}$$

The exponential map $\exp : \mathbb{H}(n) \to \mathbb{P}(n)$ increases distances; i.e.,

$$\delta_2(e^H, e^K) \geq \|H - K\|_2 \tag{2.22}$$

for all $H, K \in \mathbb{H}(n)$. This is called the EMI, *the exponential metric increasing property*. See [7] for a simple proof of it.

Any two points A, B of $\mathbb{P}(n)$ can be joined by a unique geodesic, for which a natural parametrisation is

$$A\#_t B := A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2}, \quad 0 \le t \le 1. \tag{2.23}$$

This shows the geometric mean $A\#B$ defined by (2.9) in a new light. It is the midpoint of the geodesic joining A to B.

A consequence of the parallelogram law for the norm $\|\cdot\|_2$ is the Apollonius theorem: given any A, B, C, let $M = \frac{1}{2}(A+B)$. Then

$$\|A - C\|_2^2 + \|B - C\|_2^2 = 2\left(\|M - C\|_2^2 + \|M - A\|_2^2\right).$$

The EMI can be used to show that for the metric δ_2 there is an analogue in the form of an inequality: given any A, B, C in $\mathbb{P}(n)$, let $M = A\#B$. Then

$$\delta_2^2(M, C) \le \frac{\delta_2^2(A, C) + \delta_2^2(B, C)}{2} - \frac{\delta_2^2(A, B)}{4}. \tag{2.24}$$

This is called the *semiparallelogram law.*

Several authors studied $A\#_t B$ as the "t-geometric mean", considering it a generalisation of the geometric mean but not always making a connection with the Riemannian geometry. Such a connection was noted in the work of Corach and coauthors. See, e.g., [16]. In an excellent expository article [20], Lawson and Lim highlighted this point of view.

This suggests that the geometric mean of m positive definite matrices ought to be the "centre" of the convex set spanned by $A_1, \dots, A_m$. Such a definition was given in two papers, one by Bhatia and Holbrook [11] and the other by Moakher [24]. Before describing this *Riemannian mean* we discuss another object introduced by Ando, Li and Mathias. (We remark here that there has been some very interesting work on means and medians in non-Riemannian (Finsler) geometries as well. See the paper by Arnaudon and Nielsen [3] and references therein).

2.4 The ALM Mean

The paper of Ando, Li and Mathias [2] is very significant as it first clearly articulates ten conditions that a geometric mean $G(A_1, \dots, A_m)$ should satisfy, and then gives a construction of such a mean. The ten conditions (not all independent of each other) are:

1. *Consistency with scalars*. If $A_1, \dots, A_m$ pairwise commute, then

$$G(A_1, \dots, A_m) = (A_1 A_2 \dots A_m)^{1/m}.$$

2. *Joint homogeneity*. For all positive numbers $\alpha_1, \dots, \alpha_m$,

$$G(\alpha_1 A_1, \dots, \alpha_m A_m) = (\alpha_1 \alpha_2 \dots \alpha_m)^{1/m} G(A_1, \dots, A_m).$$

3. *Symmetry*. If σ is any permutation of $\{1, 2, \ldots, m\}$, then

$$G(A_{\sigma(1)}, \ldots, A_{\sigma(m)}) = G(A_1, \ldots, A_m).$$

4. *Monotonicity*. If $A_j \leq B_j$, $1 \leq j \leq m$, then

$$G(A_1, \ldots, A_m) \leq G(B_1, \ldots, B_m).$$

5. *Congruence Invariance*. For all X in $GL(n)$

$$G(X^* A_1 X, \ldots, X^* A_m X) = X^* G(A_1, \ldots, A_m) X.$$

6. *Continuity*. If $\left\{A_j^{(n)}\right\}_n$ is a decreasing sequence of positive matrices converging to A_j, then the sequence $\left\{G(A_1^{(n)}, \ldots, A_m^{(n)})\right\}_n$ converges to $G(A_1, \ldots, A_m)$.
7. *Joint concavity*. If $0 < t < 1$, then

$$\begin{aligned} &G\left((1-t)A_1 + tB_1, \ldots, (1-t)A_m + tB_m\right) \\ &\geq (1-t)G(A_1, \ldots, A_m) + tG(B_1, \ldots, B_m). \end{aligned}$$

8. *Self-duality*. $G(A_1, \ldots, A_m) = G(A_1^{-1}, \ldots, A_m^{-1})^{-1}$.
9. *Determinant identity*. $\det G(A_1, \ldots, A_m) = (\det A_1 \cdot \det A_2 \ldots \det A_m)^{1/m}$.
10. *Arithmetic-geometric-harmonic mean inequality*.

$$\left(\frac{1}{m}\sum_{j=1}^{m} A_j^{-1}\right)^{-1} \leq G(A_1, \ldots, A_m) \leq \frac{1}{m}\sum_{j=1}^{m} A_j.$$

When $m = 2$, the binary mean $G(A_1, A_2) = A_1 \# A_2$ satisfies all these conditions. For $m > 2$ the ALM mean is defined inductively. Suppose a geometric mean $G^{\#}$ has been defined for $(m-1)$ tuples. Then given an m-tuple $A = (A_1, \ldots, A_m)$ of positive matrices define the m-tuple $T(A)$ as

$$\begin{aligned} T(A) = (&G^{\#}(A_2, \ldots, A_m), \ldots, \\ &G^{\#}(A_1, \ldots, \widehat{A}_j, \ldots, A_m), \ldots, G^{\#}(A_1, \ldots, A_{m-1})) \end{aligned}$$

where the circumflex indicates the term under it has been dropped. Then it can be shown that the sequence $\{T^k(A)\}$ converges to an m-tuple of the form $(X, X, \ldots, X)$. We then define

$$G^{\#}(A_1, \ldots, A_n) = X.$$

In the case $m = 3$, this process can be visualised as follows. Given A, B, C let $\triangle_1$ be the "triangle" with vertices A, B, C, and successively construct a sequence of triangles $\triangle_{k+1}$ by joining the "midpoints" of the vertices of $\triangle_k$.

From the semiparallelogram law (2.24) it follows that

$$\delta_2(A\#B,\ A\#C) \leq \frac{\delta_2(B, C)}{2}. \tag{2.25}$$

(If the geometry was Euclidean, the two sides of (2.25) would have been equal). This, in turn, shows that the diameter of $\triangle_{k+1}$ is at most $\frac{1}{2^k}$ times the diameter of $\triangle_1$. The space $(\mathbb{P}(n), \delta_2)$ is a complete metric space. So the intersection of the nested sequence $\{\triangle_k\}$ is single point. This point is $G^{\#}(A, B, C)$.

This interpretation of the ALM mean was given in [11].

2.5 The Riemannian Mean

The *Riemannian barycentre*, or the *centre of mass* of m elements $A_1, \ldots, A_m$ is defined as

$$G(A_1, \ldots, A_m) = \arg\min \sum_{j=1}^{m} \delta_2^2(X, A_j), \tag{2.26}$$

where the notation $\arg\min f(X)$ means the point X_0 at which the function f attains its minimum value. It is a classical theorem of E. Cartan that the minimum in (2.26) is attained at a unique point X_0. It can be shown that this point is the solution of the matrix equation

$$\sum_{j=1}^{m} \log\left(A_j^{-1/2} X A_j^{-1/2}\right) = 0. \tag{2.27}$$

This $G(A_1, \ldots, A_m)$ was proposed as the geometric mean of $A_1, \ldots, A_m$ in [11] and [24].

It is clear from the definition that G is symmetric in the m variables. The invariance properties (2.19) and (2.20) for the metric δ_2 lead to Properties 5 and 8 in the ALM list. Some others like 1, 2 and 6 can be derived without much difficulty. Properties like monotonicity and concavity are not at all obvious. This was left unresolved in [11], and in the expositions of this work in [10] and [8].

Given m points $a_1, \ldots, a_m$ in a Euclidean space, the function

$$\sum_{j=1}^{m} \|x - a_j\|^2$$

has a unique minimum at

$$\bar{a} = \frac{1}{m}(a_1 + \cdots + a_m),$$

the arithmetic mean of $a_1, \ldots, a_m$. This is the "Euclidean barycentre" of these points. When $m = 3$, the point $\bar{a}$ is the point where the three medians of the triangle with vertices a_1, a_2, a_3 intersect. This is also the point that lies in the intersection of the nested sequence of triangles $\{\triangle_k\}$ obtained by the procedure outlined at the end of Sect. 2.4.

It was pointed out in [11] that in the Riemannian space $(\mathbb{P}(n), \delta_2)$ the three medians of a triangle do not always intersect each other, and that the Riemannian barycentre and the ALM mean are not always the same.

The interpretation of the ALM mean as a procedure for reaching the "centre" of a triangle inspired the construction of another mean by Nakamura [26] and by Bini, Meini and Poloni [14]. Given A, B, C define sequences $\{A^{(k)}\}$, $\{B^{(k)}\}$, $\{C^{(k)}\}$ as follows

$$\left(A^{(0)}, B^{(0)}, C^{(0)}\right) = (A, B, C),$$

and for $k \geq 0$

$$\begin{aligned} &\left(A^{(k+1)}, B^{(k+1)}, C^{(k+1)}\right) \\ &= \left(A^{(k)}\#_{2/3}\left(B^{(k)}\#C^{(k)}\right), B^{(k)}\#_{2/3}\left(A^{(k)}\#C^{(k)}\right), C^{(k)}\#_{2/3}\left(A^{(k)}\#B^{(k)}\right)\right). \end{aligned}$$

Then the three sequences $\{A^{(k)}\}$, $\{B^{(k)}\}$, $\{C^{(k)}\}$ converge to a common limit $\widetilde{G}(A, B, C)$.

In the analogous situation in Euclidean geometry $A^{(1)}$ is obtained by going from A two-thirds of the distance towards the midpoint of B and C. Thus the points $A^{(1)}$, $B^{(1)}$ and $C^{(1)}$ all coincide with the centre of the triangle $\triangle(A, B, C)$.

It was shown in [14] that in the case of $\mathbb{P}(n)$, the mean $\widetilde{G}(A, B, C)$ is, in general, different from the ALM mean $G^{\#}(A, B, C)$. It is remarkable that the mean $\widetilde{G}(A, B, C)$ also has the ten properties enjoyed by the ALM mean. With this work it became clear that when $m > 2$, there are infinitely many possible definitions of a geometric mean that satisfy the ten conditions stipulated in [2]. In a recent paper Palfia [27] has proposed a general method for extending the definition of binary matrix means to the multivariable case.

We point out that both $G^{\#}$ and $\widetilde{G}$ are realised as limits of sequences of two-variable geometric means. Since the binary mean $A\#B$ is monotone in A and B, this property is inherited by $G^{\#}$ and $\widetilde{G}$ when more than two variables are involved. Some other properties like the arithmetic-geometric-harmonic mean inequality for $G^{\#}$ and $\widetilde{G}$ too can be derived from the two-variable case. The definition (2.26) for G involves all the m matrices at the same time, and this argument is not readily available.

Though the Riemannian mean has long been of interest to geometers, questions concerning its monotonicity, eigenvalues, norms etc. have not arisen naturally in that context. More recently there has been vigorous interest in this mean because of its use in image and signal processing problems. (See the article by Barbaresco [4] for an excellent account). Thus it becomes more important to know whether it has all the properties listed above. It turns out that it does. This was first proved using probabilistic ideas that we explain next.

2.6 Reaching the Riemannian Barycentre

Let $a_1, \ldots, a_m$ be vectors in a Euclidean space and consider the averages s_j defined as

$$\begin{aligned} s_1 &= a_1, \\ s_2 &= \frac{1}{2}(a_1 + a_2), \\ s_3 &= \frac{2}{3}s_2 + \frac{1}{3}a_3 = \frac{1}{3}(a_1 + a_2 + a_3), \\ &\vdots \\ s_k &= \frac{k-1}{k}\, s_{k-1} + \frac{1}{k}\, a_k. \end{aligned}$$

Clearly $s_m = \frac{1}{m}(a_1 + \cdots + a_m)$. The procedure that we now describe is inspired by this idea.

Let $A_1, \ldots, A_m$ be positive matrices and consider the "asymmetric averages"

$$\begin{aligned} S_1 &= A_1, \\ S_2 &= (A_1 \#_{1/2} A_2), \\ S_3 &= S_2 \#_{1/3} A_3 \\ &\vdots \\ S_k &= S_{k-1} \#_{1/k} A_k. \end{aligned}$$

We cannot quite expect, as in Euclidean geometry, that S_m would be the Riemannian barycentre $G(A_1, \ldots, A_m)$. However there is an adaptation of this idea—a sequence of such averages that converges to G.

The space $(\mathbb{P}(n), \delta_2)$ is a complete metric space of nonpositive curvature. (These are spaces whose metric satisfies the semiparallelogram law). A general theory of probability measures on such spaces has been developed. From the work of Sturm [29] on this topic, Lawson and Lim [21] extracted the following idea pertinent to our discussion.

Carry out a sequence of independent trials in which an integer is chosen from the set $\{1, 2, \ldots, m\}$ with equal probability. Let $\mathcal{I} = \{i_1, i_2, \ldots\}$ be a sequence thus obtained. Let $\{S_k(\mathcal{I}, A)\}$ be the sequence whose terms S_k are defined as

$$S_1 = A_{i_1}, \;\; S_2 = S_1 \#_{1/2} A_{i_2}, \ldots, \;\; S_k = S_{k-1} \#_{1/k} A_{i_k}.$$

It follows from a theorem of Sturm [29] that for almost all $\mathcal{I}$ the sequence $\{S_k(\mathcal{I}, A)\}$ converges to $G(A_1, \ldots, A_m)$.

The first import of this result is that the Riemannian mean $G(A_1, \ldots, A_m)$ is the limit of a sequence constructed from $A_1, \ldots, A_m$ by taking at each step a binary geometric mean. Second, since the convergence takes place for almost all sequences $\mathcal{I}$, given two m-tuples $(A_1, \ldots, A_m)$ and $(B_1, \ldots, B_m)$ we can find a sequence $\mathcal{I}$ such that

$$G(A_1, \ldots, A_m) = \lim_{k\to\infty} S_k(\mathcal{I}, A)$$

and

$$G(B_1, \ldots, B_m) = \lim_{k\to\infty} S_k(\mathcal{I}, B).$$

The monotonicity of G follows from this because of the known properties of binary means: if $A_j \le B_j$, $1 \le j \le m$, then

$$S_k(\mathcal{I}, A) \le S_k(\mathcal{I}, B) \quad \text{for} \quad k = 1, 2, 3, \ldots.$$

This and other properties of G like its joint concavity were obtained by Lawson and Lim [21].

A much simplified argument was presented in [12]. In this paper it is noted that for deriving the property mentioned above (monotonicity) we need only that there is *one* common sequence $\mathcal{I}$ for which both $S_k(\mathcal{I}, A)$ and $S_k(\mathcal{I}, B)$ converge to their respective limits $G(A_1, \ldots, A_m)$ and $G(B_1, \ldots, B_m)$. For this we do not need the "strong law of large numbers" proved by Sturm which says that the convergence takes place for almost all $\mathcal{I}$. It is adequate to have a "weak law of large numbers" that would say that for each m-tuple $A = (A_1, \ldots, A_m)$ the convergence takes place for $\mathcal{I}$ in a set of large measure (large here means anything bigger than $1/2$ of the full measure). Then given two m-tuples A and B, these two sets of large measure intersect each other. So there is a sequence $\mathcal{I}$ for which both $S_k(\mathcal{I}, A)$ and $S_k(\mathcal{I}, B)$ converge. Further such a weak law of large numbers can be proved using rather simple counting arguments and familiar matrix analysis ideas.

A several variables version of the fundamental inequality (2.24) is proved in [12]. This could be useful in other contexts. Let $G = G(A_1, \ldots, A_m)$. Then for any point C of $\mathbb{P}(n)$ we have

$$\delta_2^2(G, C) \le \sum_{j=1}^{m} \frac{1}{m} \left[\delta_2^2(A_j, C) - \delta_2^2(A_j, G)\right]. \tag{2.28}$$

When $m = 2$, this reduces to (2.24).

The main argument in [12] is based on the following inequality. Let $\mathcal{I}_n$ be the set of all ordered n-tuples $(j_1, \ldots, j_n)$ with $j_k \in \{1, 2, \ldots, m\}$. This is a set with m^n elements. For each element of this set we define, as before, averages $S_n(j_1, \ldots, j_n; A)$ inductively as follows: $S_1(j; A) = A_j$ for all $j \in \mathcal{I}_1$,

$$S_n(j_1, \ldots, j_{n-1}, k; A) = S_{n-1}(j_1, \ldots, j_{n-1}; A) \#_{1/n} A_k,$$

for all $(j_1, \ldots, j_{n-1})$ in $\mathcal{I}_{n-1}$ and k in $\mathcal{I}_1$. Let $G = G(A_1, \ldots, A_m)$ and

$$\alpha = \frac{1}{m} \sum_{j=1}^{m} \delta_2^2(G, A_j), \tag{2.29}$$

then

$$\frac{1}{m^n} \sum_{(j_1,\ldots,j_n)\in\mathcal{I}_n} \delta_2^2 \left(G, S_n(j_1, \ldots, j_n; A)\right) \leq \frac{1}{n} \alpha. \tag{2.30}$$

This inequality says that on an average (over $\mathcal{I}_n$) $\delta_2^2 \left(G, S_n(j_1, \ldots, j_n; A)\right)$ is smaller than $\frac{1}{n}\alpha$. So if $\frac{1}{n}\alpha < \frac{\varepsilon}{3}$, then at most one third of the terms in the sum on the left hand side of (2.30) can be bigger than ε. This is the "weak law" that suffices for the argument mentioned earlier. Let $A' = (A'_1, \ldots, A'_m)$ be another m-tuple of positive matrices, $G' = G(A'_1, \ldots, A'_m)$ and α' the corresponding quantity defined by (2.29). Given ε choose n such that $\frac{1}{n}\alpha < \frac{\varepsilon}{3}$ and $\frac{1}{n}\alpha' < \frac{\varepsilon}{3}$. Then for at least $2/3$ of $(j_1, \ldots, j_n)$ in $\mathcal{I}_n$ we have

$$\delta_2^2 \left(G, S_n(j_1, \ldots, j_n; A)\right) < \varepsilon,$$

and also for at least $2/3$ of them

$$\delta_2^2 \left(G', S_n(j_1, \ldots, j_n; A')\right) < \varepsilon.$$

So for at least $1/3$ of elements of $\mathcal{I}_n$ both these inequalities are simultaneously true. If $A_j \leq A'_j$, $1 \leq j \leq m$, then $S_n(j_1, \ldots, j_n; A) \leq S_n(j_1, \ldots, j_n; A')$. From this we can conclude that $G \leq G'$.

At this stage one wonders whether there is a probability-free proof of this. The argument of Lawson and Lim is based on the fact, proved by Sturm, that for almost all sequences $\mathcal{I}$ with their terms coming from the set $\{1, 2, \ldots, m\}$ the averages $S_k(\mathcal{I}, A)$ converge to $G(A_1, \ldots, A_m)$. If this happens for almost all sequences, it should happen for the most natural sequence whose terms are i_k with $i_k = k (\operatorname{mod} m)$. In my talk at the MIG Seminar [9] this question was raised as "*Is G really playing dice?*" This has now been answered by Holbrook in [17]. He has shown that one can reach $G(A_1, \ldots, A_m)$ as a limit of a "deterministic walk". More precisely he proves the following. For any $X \in \mathbb{P}(n)$ let $\varphi_r(X) = X \#_{1/r} A_k$ where $k = r (\operatorname{mod} m)$. Let

$$\varphi_{r,n} = \varphi_{r+n-1} \cdots \varphi_{r+1} \cdot \varphi_r.$$

Then for all X, and for all positive integers r,

$$\lim_{n\to\infty} \varphi_{r,n}(X) = G(A_1, \ldots, A_m). \tag{2.31}$$

Choosing $X = A_1$ and $r = 1$, we see that the sequence

$$((((\,(A_1 \#_{1/2} A_2)\#_{1/3})A_3)\ldots\#_{1/m} A_m)\#_{1/m+1} A_1)\#_{1/m+2} A_2 \ldots$$

converges to $G(A_1, \ldots, A_m)$.

By its definition $G(A_1, \ldots, A_m)$ is the unique minimiser of the strictly convex function

$$f(X) = \sum_{j=1}^{m} \delta_2^2(X, A_j).$$

Therefore, one can prove (2.31) by showing that

$$\lim_{n\to\infty} f(\varphi_{r,n}(X)) = f(G).$$

The gradient of the function f is known [8]. So tools of calculus can be brought in. The essential idea of Holbrook's proof is to show that as n runs through all positive integers the distance of $\varphi_{r,n}(X)$ from G is reduced after every m steps.

Almost simultaneously with Holbrook's work has appeared a very interesting paper by Lim and Palfia [22]. Here the Riemannian mean is realised as the limit of another sequence. To understand the idea behind this it is helpful to start with the case of positive numbers $a_1, \ldots, a_m$. We have

$$\lim_{t\to\infty} \left(\frac{a_1^t + \cdots + a_m^t}{m}\right)^{1/t} = \exp\left(\frac{\log\, a_1 + \cdots + \log\, a_m}{m}\right) = (a_1 \ldots a_m)^{1/m}. \tag{2.32}$$

The quantities

$$x_t := \left(\frac{a_1^t + \cdots + a_m^t}{m}\right)^{1/t}, \quad t \neq 0 \tag{2.33}$$

are the classical power means. Going to positive matrices $A_1, \ldots, A_m$ we do have

$$\lim_{t\to 0} \left(\frac{A_1^t + \cdots + A_m^t}{m}\right)^{1/t} = \exp\left(\frac{\log\, A_1 + \cdots + \log\, A_m}{m}\right), \tag{2.34}$$

as was observed in [5]. However, the positive operator in (2.34) is not the same as $G(A_1, \ldots, A_m)$, except in very special cases. Again, there is an ingenious adaptation of this in [22] that successfully tackles the noncommutativity.

Note that the power mean x_t can be characterised as the unique solution of the equation

$$x = \frac{1}{m} \sum_{j=1}^{m} x^{1-t}\, a_j^t. \tag{2.35}$$

Inspired by this, consider the equation

$$X = \frac{1}{m} \sum_{j=1}^{m} X \#_t A_j, \tag{2.36}$$

for $0 < t < 1$. Lim and Palfia show that this equation has a unique solution. Applying a congruence we see that this solution X_t satisfies the equation

$$\begin{aligned} I &= \frac{1}{m} \sum_{j=1}^{m} I \#_t \left(X_t^{-1/2} A_j X_t^{-1/2} \right) \\ &= \frac{1}{m} \sum_{j=1}^{m} \left(X_t^{-1/2} A_j X_t^{-1/2} \right)^t, \end{aligned}$$

which, in turn, leads to

$$\sum_{j=1}^{m} \frac{\left(X_t^{-1/2} A_j X_t^{-1/2} \right)^t - I}{t} = 0, \quad 0 < t < 1. \tag{2.37}$$

On the other hand $G(A_1, \ldots, A_m)$ is the solution of the equation

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

Now recall that

$$\lim_{t \downarrow 0} \frac{x^t - 1}{t} = \log x.$$

A calculation based on this is then used to show that

$$\lim_{t \downarrow 0} X_t = G(A_1, \ldots, A_m). \tag{2.38}$$

Once again, taking advantage of the fact that G is a limit of objects defined via binary geometric means, several properties of G like monotonicity and concavity can easily be derived.

In [12] it was shown that in addition to the ten properties listed at the beginning of Sect. 2.4, the mean G has other interesting properties important in operator theory. If Φ is a positive unital linear map on the matrix algebra $\mathbb{M}(n)$, then

$$\Phi(G(A_1, \ldots, A_m)) \leq G(\Phi(A_1), \ldots, \Phi(A_m)). \tag{2.39}$$

If $|||\cdot|||$ is any unitarily invariant norm on $\mathbb{M}(n)$, then

$$|||G(A_1, \ldots, A_m)||| \le \prod_{j=1}^{m} |||A_j|||^{1/m}. \tag{2.40}$$

Special cases of (2.39) and (2.40) were proved earlier by Yamazaki [30]. It turns out that both the means $G^{\#}$ and $\widetilde{G}$ considered in Sects. 2.4 and 2.5 also satisfy (2.39) and (2.40).

It is of interest to know what properties characterise the Riemannian mean G among all means. One such property has been found in [31] and [22]. In the first of these papers, Yamazaki showed that

$$\sum_{j=1}^{m} \log A_j \le 0 \quad \text{implies} \quad G(A_1, \ldots, A_m) \le I. \tag{2.41}$$

In [22] Lim and Palfia show that this condition together with congruence invariance and self-duality (conditions 5 and 8 in the ALM list) uniquely determine the mean G. To see this consider any function $g(A_1, \ldots, A_m)$ taking positive matrix values. If it satisfies the condition (2.41) and is self-dual, then $\sum_{j=1}^{m} \log A_j = 0$ implies $g(A_1, \ldots, A_m) = I$. If $X = G(A_1, \ldots, A_m)$, then we have $\sum_{j=1}^{m} \log \left(X^{-1/2} A_j X^{-1/2}\right) = 0$. Hence

$$g\left(X^{-1/2}A_1X^{-1/2}, \ldots, X^{-1/2}A_mX^{-1/2}\right) = I.$$

If g is congruence-invariant, then from this it follows that

$$g(A_1, \ldots, A_m) = X = G(A_1, \ldots, A_m).$$

2.7 Summary

The Riemannian mean, also called the Cartan mean or the Karcher mean, has long been of interest in differential geometry. Recently it has been used in several areas like radar and medical imaging, elasticity, machine learning and statistics. It is also an interesting topic for matrix analysts and operator theorists. Some questions (like operator monotonicity and concavity) that are intrinsically more natural to these subjects have led to a better understanding of this object. In particular several new characterisations of the Riemannian mean have been found in 2010–2011. These show the mean as a limit of (explicitly constructed) sequences. They will be useful for devising numerical algorithms for computation of the mean.

References

1. Ando, T.: Concavity of certain maps on positive definite matrices and applications to Hadamard products. Linear Algebra Appl. **26**, 203–241 (1979)
2. Ando, T., Li, C.K., Mathias, R.: Geometric means. Linear Algebra Appl. **385**, 305–334 (2004)
3. Arnaudon, M., Nielsen, F.: Medians and means in finsler geometry, hal-00540625, June 2011
4. Barbaresco, F.: Interactions between symmetric cone and information geometry: Bruhat-Tits and Siegel space models for higher resolution autoregressive Doppler imagery. In Emerging Trends in Visual Computing: LIX Fall Colloquium, ETVC 2008, Palaiseau, France, 2008. Revised Invited Papers, pp. 124–163. Springer-Verlag (2009)
5. Bhagwat, K.V., Subramanian, R.: Inequalities between means of positive operators. Math. Proc. Camb. Philos. Soc. **83**, 393–401 (1978)
6. Bhatia, R.: Matrix Analysis. Springer-Verlag, New York (1997)
7. Bhatia, R.: On the exponential metric increasing property. Linear Algebra Appl. **375**, 211–220 (2003)
8. Bhatia, R.: Positive definite matrices. In: Princeton Series in Applied Mathematics. Princeton University Press, Princeton (2007)
9. Bhatia, R.: Matrix means. Indo-French Seminar on Matrix Information Geometries, Palaiseau, February (2011)
10. Bhatia, R., Holbrook, J.: Noncommutative geometric means. Math. Intell. **28**, 32–29 (2006)
11. Bhatia, R., Holbrook, J.: Riemannian geometry and matrix geometric means. Linear Algebra Appl. **413**, 594–618 (2006)
12. Bhatia, R., Karandikar, R.: Monotonicity of the matrix geometric mean, to appear in Math. Ann. (2011)
13. Bini, D., Iannazzo, B.: Computing the Karcher mean of symmetric positive definite matrices, to appear in Linear Algebra Appl. (2012)
14. Bini, D., Meini, B., Poloni, F.: An effective matrix geometric mean satisfying the Ando-Li-Mathias properties. Math. Comp. **79**, 437–452 (2010)
15. Carlin, H.J., Noble, G.A.: Circuit properties of coupled dispersive lines with applications to wave guide modelling. In: Skwirzynki, J.K., Scanlan, J.O. (eds.) Proceedings on Network and Signal Theory, pp. 258–269. Peter Pergrinus (1973)
16. Corach, G., Porta, H., Recht, L.: Convexity of the geodesic distance on spaces of positive operators. Ill. J. Math. **38**, 87–94 (1994)
17. Holbrook, J.: No dice: a determine approach to the Cartan centroid, to appear in J. Ramanujan Math. Soc.
18. Karcher, H.: Riemannian center of mass and mollifier smoothing. Comm. Pure Appl. Math. **30**, 509–541 (1977)
19. Kubo, F., Ando, T.: Means of positive linear operators. Math. Ann. **246**, 205–224 (1980)
20. Lawson, J., Lim, Y.: The geometric mean, matrices, metrics and more. Am. Math. Mon. **108**, 797–812 (2001)
21. Lawson, J., Lim, Y.: Monotonic properties of the least squares mean. Math. Ann. **351**, 267–279 (2011)
22. Lim, Y., Pálfia, M.: The matrix power means and the Karcher mean. J. Funct. Anal. **262**, 1498–1514 (2012)
23. Moakher, M.: Means and averaging in the group of rotations. SIAM J. Matrix Anal. Appl. **24**, 1–16 (2002)
24. Moakher, M.: A differential geometric approach to the geometric mean of symmetric positive definite matrices. SIAM J. Matrix Anal. Appl. **26**, 735–747 (2005)
25. Moakher, M.: On the averaging of symmetric positive-definite tensors. J. Elast. **82**, 273–296 (2006)
26. Nakamura, N.: Geometric means of positive operators. Kyungpook Math. J. **49**, 167–181 (2009)
27. Palfia, M.: A multivariable extension of two-variable matrix means. SIAM J. Matrix Anal. Appl. **32**, 385–393 (2011)

28. Pusz, W., Woronowicz, S.L.: Functional calculus for sesquilinear forms and the purification map. Rep. Math. Phys. **8**, 159–170 (1975)
29. Sturm, K.-T.: Probability measures on metric spaces of nonpositive curvature. In: Auscher P. et al. (eds.) Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces (Contemp. Math. vol. 338, Amer. Math. Soc. 2003)
30. Yamazaki, T.: An elementary proof of arithmetic-geometric mean inequality of the weighted Riemannian mean of positive definite matrices, to appear in Linear Algebra Appl.
31. Yamazaki, T.: The Riemannian mean and matrix inequalities related to the Ando-Hiai inequality and chaotic order, to appear in Operators and Matrices

http://www.springer.com/978-3-642-30231-2

Matrix Information Geometry
(Eds.)F. Nielsen; R. Bhatia
2013, XII, 454 p. 104 illus., Hardcover
ISBN: 978-3-642-30231-2