British Journal of Mathematics & Computer Science
5(1): 1-12, 2015, Article no.BJMCS.2015.001
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

A Review of Geometric Mean of Positive Definite Matrices

Wen-Haw Chen[*1]

[1] *Department of Applied Mathematics, Tunghai University, Taichung 40704, Taiwan.*

Article Information
DOI: 10.9734/BJMCS/2015/13026
Editor(s):
(1) Jaime Rangel-Mondragon, Faculty of Informatics, Queretaros Institute of Technology, Mexico and Faculty of Computer Science, Autonomous University of Quertaro, Mexico.
Reviewers:
(1) Anonymous, University of Camerino, Italy.
(2) Anonymous, Sony Computer Science Laboratories, Japan.
(3) Anonymous, Northwest Normal University, P.R. China.
Complete Peer review History:
http://www.sciencedomain.org/review-history.php?iid=707&id=6&aid=6430

Review Article

Received: 29 July 2014
Accepted: 22 September 2014
Published: 09 October 2014

Abstract

In the paper [1], Pusz and Woronowicz first gave the geometric mean of two positive definite matrices. This mean has similar properties to those of the geometric mean of two positive numbers. In [2], Ando, Li and Mathias listed ten properties that a geometric mean of m positive definite matrices should satisfy. Then gave a definition of geometric mean of m matrices by a iteration which satisfies these ten properties. For the geometric mean of two positive definite matrices, there is an interesting relationship between matrix geometric mean and the information metric. Consider the set of all positive matrices as a Riemannian manifold with the information metric. Then the geometric mean of two matrices in the manifold is just the middle point of the geodesic connecting them. In this paper, we review this notion and present two different proofs, the variation method and the exponential map method, for proving the relationship.

Keywords: Geometric means; Positive definite matrices; Geodesics.
2010 Mathematics Subject Classification: 53C21; 83C05; 57N16

**Corresponding author: E-mail: whchen@thu.edu.tw*

1 Introduction

Study of geometric means on positive definite matrices is an important topic to many disciplines of science such as operator theory, physics, engineering and statistics etc (c.f. [3], [4] and [5]). However, the product of two positive matrices is not always positive definite. For example, let $A = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$. Then $AB = \begin{bmatrix} -1 & 3 \\ -3 & 8 \end{bmatrix}$ is not positive definite since AB is not symmetric. Hence we can not define the geometric mean to be $(AB)^{\frac{1}{2}}$, since the geometric mean of two positive matrices should be positive.

The geometric mean of two positive definite matrices was first given by Pusz and Woronowicz in [1]. They defined the geometric mean $A\#B$ of two positive definite matrices A and B by

$$A\#B = A^{\frac{1}{2}}(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})^{\frac{1}{2}}A^{\frac{1}{2}}. \tag{1.1}$$

More detailed study on the mean can be found in Kubo and Ando's paper [3]. The geometric mean $A\#B$ has some similar properties as those of the geometric mean of positive numbers. For example, it satisfies the arithmetic-geometric-harmonic-mean inequality (c.f. [6] and [5] e.t.c.).

On the other hand, the geometric mean of positive definite matrices may be defined geometrically. Information geometry began as the geometric study of statistical estimation, and consider the set of probability distributions which constitute a statistical model as a Riemannian manifold with the Fisher metric. In [7], Rao had already pointed out in his paper that the Fisher information matrix determines a Riemannian metric on a statistical manifold (see also [8]). In addition, there is an interesting relationship between matrix geometric mean and the information metric. It can be shown that the geometric mean of A, B is just the middle point of the geodesic which connecting A and B.

In the rest of this paper, we will review the Fisher information metric on a statistical model in section 2. In section 3, we will verify that the geometric mean is the midpoint of the geodesic segment by the variation method and the exponential method. Moreover, a definition of geometric mean for three or more positive definite matrices by a iteration will be reviewed in section 4. In the final section, we discuss the geometric meaning of the geometric mean for more positive matrices.

2 Fisher Information Metric of a Statistical Model

We first introduce the Fisher information metric of a statistical model (c.f. [8] and [9]). Information geometry has been used in diverse applications such as statistics, control theory, and information theory. It began as the geometric study of statistical estimation. This involved viewing a statistical model as a Riemannian manifold with the Fisher metric.

Definition 2.1. (c.f. [8]) Let S be a family of probability distributions on $\mathcal{X}$. Suppose each element of S, that is a probability distribution, may be parameterized using n real-valued variables $[\xi^1, \cdots, \xi^n$ so that

$$S = \{p_\xi = p(x;\xi) \,|\, \xi = [\xi^1, \cdots, \xi^n \in \Xi\}],$$

where Ξ is a subset of $\mathbb{R}^n$ and the mapping $\xi \mapsto p_\xi$ is injective, then we call such S an n-dimensional statistical model on $\mathcal{X}$.

For example, consider the multivariate normal distribution. Let $\mathcal{X} = \mathbb{R}^k$ be the k-dimensional real space, $n = k + \frac{k(k+1)}{2}$ and $\xi = [\mu, \Sigma]$, where μ is the mean and Σ is a $k \times k$ positive definite matrix. Define $\Xi = \{[\mu, \Sigma] \,|\, \mu \in \mathbb{R}^k,\, \Sigma \in \mathbb{R}^{k\times k}\}$. Then

$$p(x;\xi) = (2\pi)^{-\frac{k}{2}}(det\Sigma)^{-\frac{1}{2}} \exp\big\{-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)\big\}$$

Definition 2.2. Consider an n-dimensional statistical model $S = \{p_\theta \mid \theta \in \Theta\}$ where $p_\theta = p(x;\theta)$ are probability distribution functions and Θ is a subset of $\mathbb{R}^n$. Then the Fisher information matrix of S at θ is the $n \times n$ matrix $G(\theta) = \big[g_{ij}(\theta)\big]$ where $g_{ij}(\theta)$ is defined by

$$\begin{aligned} g_{ij}(\theta) &= E_\theta\big[\frac{\partial}{\partial\theta^i}\log p(x;\xi)\frac{\partial}{\partial\theta^j}\log p(x;\xi)\big] \\ &= \int_{\mathcal{X}}\big[\frac{\partial}{\partial\theta^i}\log p(x;\xi)\frac{\partial}{\partial\xi^j}\log p(x;\theta)\big]p(x;\theta)\,dx. \end{aligned}$$

We assume that Θ is an open subset of $\mathbb{R}^n$ and for each $x \in \mathcal{X}$, the function $\xi \mapsto p(x;\xi)$ $(\Xi \to \mathbb{R})$ is C^∞ so that we can define $\frac{\partial}{\partial\xi^i}p(x;\xi)$ and $\frac{\partial}{\partial\xi^i}\frac{\partial}{\partial\xi^j}p(x;\xi)$.

In addition, we assume that the order of integration and differentiation may be freely rearranged. For example, we shall often use formulas such as

$$\int_{\mathcal{X}}\frac{\partial}{\partial\xi^i}p(x;\xi)\,dx = \frac{\partial}{\partial\xi^i}\int_{\mathcal{X}}p(x;\xi)\,dx = \frac{\partial}{\partial\xi^i}1 = 0.$$

We also assume that $p(x;\xi) > 0$ for all $\xi \in \Xi$ and all $x \in \mathcal{X}$, and in the discussions below, we define $\partial_i = \frac{\partial}{\partial\xi^i}$. By the assumptions, it is easy to see that G is positive semidefinite. We assume that G is positive definite.

Now we can define the Riemannian metric $g_\theta =<,>_\theta$ on the tangent space $T_\theta(S)$ at θ by

$$< (\partial_i)_\theta, (\partial_j)_\theta >_\theta = g_{ij}(\theta) = E_\theta[\partial_i l_\theta \partial_j l_\theta, \quad (\partial_i)_\theta, (\partial_j)_\theta \in T_\theta(S). \tag{2.1}$$

We call this the Fisher metric or the information metric.

An important example is the multivariate normal distributions with 0 expectation. The distributions are given by

$$p_A(x) = \frac{1}{\sqrt{(2\pi)^n det(A)}}\exp\{-\frac{1}{2}x^TA^{-1}x\}, \tag{2.2}$$

where A is positive definite real matrix and $x \in \mathbb{R}^n$. The tangent space at a point p_A can be identified as the set of all symmetric real matrices and the information metric was given by Skovgaard. The formula is

$$\begin{aligned} g_A(H_1,H_2) &= \frac{1}{2}Tr(A^{-1}H_1A^{-1}H_2) \\ &= \frac{1}{2}Tr(A^{-\frac{1}{2}}H_1A^{-\frac{1}{2}}A^{-\frac{1}{2}}H_2A^{-\frac{1}{2}}), \end{aligned}$$

where H_1, H_2 are symmetric real matrices. It coincides with the Hilbert-Schmidt inner product scaled by $\frac{1}{2}$.

Remark 2.1. The distance on the manifold of multivariate normal distributions with zero expectation has been provided by inference method. In [10], the authors presented an analytical computation of the asymptotic temporal behavior of the information geometric complexity (IGC) of finite-dimensional Gaussian statistical manifolds in the presence of microcorrelations (correlations between microvariables), and observed a power law decay of the IGC at a rate determined by the correlation coefficient. They found that microcorrelations lead to the emergence of an asymptotic information geometric compression of the statistical macrostates explored by the system at a faster rate than that observed in absence of microcorrelations. This finding uncovered an important connection between (micro)-correlations and (macro)-complexity in Gaussian statistical dynamical systems.

In the next section, we will find that the geometric mean of two matrices is the middle point of geodesic when we consider the set of real positive definite matrices as a Riemannian manifold with the information metric.

3 Geometric Mean of Two Positive Definite Matrices

We know that the arithmetic mean can be extend to the matrix

$$\frac{A_1 + \cdots + A_n}{n},$$

for n positive definite matrices $A_1, \cdots, A_n$ where $n \geq 2$. Now we want to extend the notion of geometric mean for positive numbers to positive definite matrices.

In fact, the geometric mean of two positive definite matrices A and B is defined by

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

The set of all Hermitian matrices is denote by H_n, and the set of all positive definite matrices is denote by P_n.

Next we show that the geometric mean of two positive definite matrices A, B is the middle point of the geodesic which connects A and B when P_n is viewed as a Riemannian manifold. (c.f. [11] and [12])

Let us review some results of Riemannian geometry. A Riemannian manifold is a differentiable manifold M equipped with a Riemannian metric $g(\cdot,\cdot)$ denoted by the pair (M, g). A proper variation of a piecewise differentiable curve $c : [0, a] \to M$ in a Riemannian manifold (M, g) is a variation $f : (-\epsilon, \epsilon) \times [0, a] \to M$ with the same initial point and endpoint and $f(0, t) = c(t)$. The energy functional $E(s)$ of the curve c is defined by

$$E(s) = \int_0^a |\frac{\partial f}{\partial t}(s,t)|^2 dt, \qquad s \in (-\epsilon, \epsilon).$$

It is known that from the first variation formula, the curve c is a geodesic on (M, g) if and only if c is a critical point of the energy functional $E(c(s))$.

Consider P_n as a manifold. The tangent space of a point A can be identified as H_n. Define the Riemannian metric at A by the differential

$$ds = ||A^{-\frac{1}{2}} dA A^{-\frac{1}{2}}||_2 = [Tr((A^{-1}dA)^2)]^{\frac{1}{2}}. \tag{3.1}$$

If $\gamma : [a, b] \to P_n$ is a differentiable curve in P_n, then we define its length as

$$L(\gamma) = \int_a^b ||\gamma^{-\frac{1}{2}}(t)\gamma'(t)\gamma^{-\frac{1}{2}}(t)||_2 \, dt.$$

For each invertible X, $\Gamma_X(A) = X^*AX$ is a bijection of P_n onto P_n. If γ is a differentiable curve in P_n, then the composition $\Gamma_X \circ \gamma$ is another differentiable curve in P_n. One can prove that

Lemma 3.1. *For each invertible X and for each differentiable curve γ*

$$L(\Gamma_X \circ \gamma) = L(\gamma).$$

The following theorem states that the geometric mean of two positive definite matrices is the middle point of the geodesic which connection them. We will introduce two different proofs, one is by the first variation formula and the other is by the exponential map.

Theorem 3.2. *The geodesic connecting A, B is*

$$\gamma(t) = A^{\frac{1}{2}}(A^{-\frac{1}{2}} B A^{-\frac{1}{2}})^t A^{\frac{1}{2}}$$

where $0 \leq t \leq 1$ and the geometric mean of A, B is the middle point of the geodesic which connecting A and B.

3.1 Proof by variation

By the Lemma 3.1, we may assume that $A = I, \gamma(t) = B^t$. Let $l(t)$ be a curve such that $l(0) = l(1) = 0$. Then the variations of energy function is given by

$$\frac{d}{d\epsilon}[\int_0^1 (\sqrt{g_{\gamma(t)+\epsilon l(t)}(\gamma'(t)+\epsilon l'(t), \gamma'(t)+\epsilon l'(t))})^2\, dt]|_{\epsilon=0}$$

$$= Tr(\int_0^1 -(B^t)^{-1}(\log B)^2 l(t)dt + (B^t)^{-1}(\log B)l(t)|_0^1 + \int_0^1 (B^t)^{-1}(\log B)^2 l(t)\, dt). \quad (3.2)$$

Since $l(0) = l(1) = 0$, the first term vanishes here and the derivative at ϵ is 0. On the other hand

$$g_{\gamma(t)}(\gamma'(t), \gamma'(t)) = Tr((B^t)^{-1}(B^t \log B)(B^t)^{-1}(B^t \log B))$$
$$= Tr((\log B)^2)$$

does not depend on t, we conclude that $\gamma(t) = B^t$ is the geodesic curve between I and B.

3.2 Proof by exponential map

Let X and Y be Banach space and U be an open subset of X. A map $f : U \to Y$ is said to be differential at $u \in U$ if there exist a bounded linear operator T from X to Y such that

$$\lim_{v\to 0} \frac{||f(u+v) - f(u) - T(v)||_Y}{||v||_X} = 0,$$

We call T the derivative of f at u and denote T by $Df(u)$. Note that $Df(u)(w) = \frac{d}{dt}|_{t=0} f(u+tw)$.

Now let I be an open interval and $H_n(I)$ be the collection of all Hermitian matrices whose eigenvalues are in I. Then a function f in $C^1(I)$ induces a map from $H_n(I)$ into H_n, where $C^1(I)$ is the space of continuously differentiable real-valued function on I. If $f \in C^1(I)$ and $A \in H_n(I)$, then we define $f^{[1]}(A)$ as the matrix whose i, j entry is

$$f^{[1]}(\lambda_i, \lambda_j) = \begin{cases} \frac{f(\lambda_i)-f(\lambda_j)}{\lambda_i-\lambda_j} & \lambda_i \neq \lambda_j \\ f'(\lambda_i) & \lambda_i = \lambda_j \end{cases},$$

where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. This is called the Loewner matrix of f at A. (c.f. [1] and [13])

The function f on $H_n(I)$ is differentiable. Its derivative at A, denoted as $Df(A)$, is a linear map on H_n. We have

$$Df(A)(H) = \frac{d}{dt}|_{t=0} f(A+tH).$$

An interesting expression for this derivative in terms of Loewner matrices is given in the following theorem.

Theorem 3.3. *[14] Let $f \in C^1(I)$ and $A \in H_n(I)$. Then*

$$Df(A)(H) = f^{[1]}(A) \bullet H$$
$$= U[f^{[1]}(\Lambda) \bullet (U^* H U)]U^*,$$

where Λ is diagonal and $A = U\Lambda U^*$ and $\bullet$ denotes the Schur product.

We write De^H for the derivative of the exponential map at a point H of H_n. This is a linear map on H_n and the action is given by

$$De^H(K) = \frac{d}{dt}|_{t=0}(e^{H+tK}) = \lim_{t\to 0} \frac{e^{H+tK} - e^H}{t}.$$

Theorem 3.4. *[6] For all H and K in H_n we have*

$$||e^{-\frac{H}{2}}De^{H}(K)e^{-\frac{H}{2}}||_2 \geq ||K||_2.$$

Proof. First we claim that $||X^*KX||_2 = ||K||_2$ if $X^*X = I$. Note that

$$\begin{aligned}||X^*KX||_2 &= (Tr[(X^*KX)(X^*KX)^*)^{\frac{1}{2}}\\ &= (Tr[KK^*)^{\frac{1}{2}}\\ &= ||K||_2.\end{aligned}$$

Now since H is Hermitian, $H = U\Lambda U^*$ where Λ is a diagonal matrix and $UU^* = I$. By Theorem 3.3 we have

$$\begin{aligned}De^{H}(K) &= U[[x_{ij} \bullet U^*KU]U^*]\\ &= U[[x_{ij} \bullet B]U^*]\end{aligned}$$

where $x_{ij} = \begin{cases}\frac{e^{\lambda_i}-e^{\lambda_j}}{\lambda_i-\lambda_j} & \lambda_i \neq \lambda_j\\ e^{\lambda_i} & \lambda_i = \lambda_j\end{cases}$ and $B = U^*KU$.

Note that $e^{-\frac{H}{2}} = U\begin{bmatrix}e^{-\frac{\lambda_1}{2}} & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & e^{-\frac{\lambda_n}{2}}\end{bmatrix}U^*$. Hence we have

$$\begin{aligned}&||e^{-\frac{H}{2}}De^{H}(K)e^{-\frac{H}{2}}||_2\\ &= ||U\begin{bmatrix}e^{-\frac{\lambda_1}{2}} & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & e^{-\frac{\lambda_n}{2}}\end{bmatrix}U^*U[[x_{ij}]\bullet\\ &B]U^*U\begin{bmatrix}e^{-\frac{\lambda_1}{2}} & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & e^{-\frac{\lambda_n}{2}}\end{bmatrix}U^*||_2\\ &= ||\begin{bmatrix}e^{-\frac{\lambda_1}{2}} & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & e^{-\frac{\lambda_n}{2}}\end{bmatrix}[[x_{ij}]\bullet B]\begin{bmatrix}e^{-\frac{\lambda_1}{2}} & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & e^{-\frac{\lambda_n}{2}}\end{bmatrix}||_2\\ &= ||[a_{ij}b_{ij}||_2\end{aligned}$$

where $[b_{ij}] = B$,

$$\begin{aligned}a_{ij} &= e^{-\frac{\lambda_i}{2}}\frac{e^{\lambda_i}-e^{\lambda_j}}{\lambda_i-\lambda_j}e^{-\frac{\lambda_j}{2}}\\ &= \frac{\sinh(\frac{\lambda_i-\lambda_j}{2})}{\frac{\lambda_i-\lambda_j}{2}}\end{aligned}$$

if $\lambda_i \neq \lambda_j$ and $a_{ij} = 1$ if $\lambda_i = \lambda_j$. Since $\frac{\sinh(x)}{x} \geq 1$ for all $x \neq 0$, $a_{ij} \geq 1$ for all i, j. Hence

$$\begin{aligned} ||e^{-\frac{H}{2}} De^H(K)e^{-\frac{H}{2}}||_2 &= ||[a_{ij}b_{ij}||_2 \\ &= (\sum_{ij} a_{ij}^2 b_{ij}\overline{b_{ij}})^{\frac{1}{2}} \\ &\geq (\sum_{ij} b_{ij}\overline{b_{ij}})^{\frac{1}{2}} \\ &= ||B||_2 \\ &= ||U^*KU||_2 \\ &= ||K||_2. \end{aligned}$$

□

Theorem 3.5. *[6] Let $H(t)$, $a \leq t \leq b$ be any curve in H_n and let $\gamma(t) = e^{H(t)}$. Then we have*

$$L(\gamma) \geq \int_a^b ||H'(t)||_2 \, dt.$$

Proof. By the chain rule, we have

$$\gamma'(t) = e^{H(t)}H'(t) = \frac{d}{dh}|_{h=0}[e^{H(t)+hH'(t)} = De^{H(t)}(H'(t)).$$

Theorem 3.4 implies that

$$\begin{aligned} ||\gamma^{-\frac{1}{2}}(t)\gamma'(t)\gamma^{-\frac{1}{2}}(t)||_2 &= ||e^{-\frac{H(t)}{2}} De^{H(t)}(H'(t))e^{-\frac{H(t)}{2}}||_2 \\ &\geq ||H'(t)||_2. \end{aligned}$$

Integrating over t we complete the proof. □

Now we define a metric δ_2 on P_n. For any $A, B \in P_n$, we define $\delta_2(A, B)$ by

$$\delta_2(A, B) = \inf\{L(\gamma) : \gamma \text{ is a curve from } A \text{ to } B\}.$$

According to Lemma 1, each Γ_X is an isometry for the length L. Hence it is also an isometry for the metric δ_2, that is,

$$\delta_2(A, B) = \delta_2(\Gamma_X(A), \Gamma_X(B)),$$

for all A, B in P_n and invertible X.

If $\gamma(t)$ is any curve joining A and B in P_n, then $H(t) = \log \gamma(t)$ is a curve joining $\log A$ and $\log B$ in H_n. Since $\int_a^b ||H'(t)||_2 \, dt$ is the length of $H(t)$ in H_n and H_n is a convex subspace of Euclidean space M_n, $\int_a^b ||H'(t)||_2 \, dt$ is bounded below by the length of the straight line segment $(1-t)\log A + t \log B$ which joining $\log A$ and $\log B$ where $0 \leq t \leq 1$. Hence by Theorem 3.5,

$$L(\gamma) \geq ||\log A - \log B||_2$$

and we have the following theorem.

Theorem 3.6. *[6] For each pair of points A, B in P_n we have*

$$\delta_2(A, B) \geq ||\log A - \log B||_2.$$

In other words for any two matrices H and K in H_n

$$\delta_2(e^H, e^K) \geq ||H - K||_2.$$

Thus the exponential map

$$\exp : (H_n, ||\cdot||_2) \to (P_n, \delta_2)$$

increases distances.

We write $[H, K]$ for the line segment joining H and K and $[A, B]$ for the geodesic from A to B where H, K in H_n and A, B in P_n.

Theorem 3.7. *[6] Let A and B be commuting matrices in P_n. Then the exponential function maps the line segment $[\log A, \log B]$ in H_n to the geodesic $[A, B]$ in P_n. In this case*

$$\delta_2(A, B) = ||\log A - \log B||_2.$$

Proof. We claim that

$$\gamma(t) = \exp((1-t)\log A + t\log B),$$

where $0 \le t \le 1$ is the unique curve of shortest length joining A and B in the space (P_n, δ_2). Since A and B commute, $\gamma(t) = A^{1-t}B^t$ and $\gamma'(t) = (\log B - \log A)\gamma(t)$. Thus

$$\begin{aligned} L(\gamma) &= \int_0^1 ||\gamma^{-\frac{1}{2}}\gamma'(t)\gamma^{-\frac{1}{2}}||_2 \, dt \\ &= \int_0^1 ||\log A - \log B||_2 \, dt \\ &= ||\log A - \log B||_2. \end{aligned}$$

Theorem 3.6 says that no curve can be shorter than this.

Now suppose $\tilde{\gamma}$ is another curve that joins A and B and has the same length as that of γ. Then $\tilde{H}(t) = \log \tilde{\gamma}(t)$ is a curve that joins $\log A$ and $\log B$ in H_n, and by Theorem 3.5, this curve has length $||\log A - \log B||_2$. But in a Euclidean space the straight line segment is the unique shortest curve between two points. So $\tilde{H(t)}$ is a re-parametrization of the line segment $[\log A, \log B]$. ☐

When A and B commute, the natural parametrization of the geodesic $[A, B]$ is given by

$$\gamma(t) = A^{1-t}B^t, \; 0 \le t \le 1,$$

in the sense that $\delta_2(A, \gamma(t)) = t\delta_2(A, B)$ for each t. The general case is obtained from this and the isometries Γ_X.

Theorem 3.8. *[6] Let A and B be any two elements of P_n. Then there exists a unique geodesic $[A, B]$ joining A and B. This geodesic has a parametrization*

$$\gamma(t) = A^{\frac{1}{2}}(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})^t A^{\frac{1}{2}}, \; 0 \le t \le 1,$$

which is natural in the sense that

$$\delta_2(A, \gamma(t)) = t\delta_2(A, B)$$

for each t. Furthermore, we have

$$\delta_2(A, B) = ||\log(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})||_2.$$

Proof. The matrices I and $A^{-\frac{1}{2}}BA^{-\frac{1}{2}}$ commute. By Theorem 3.7, the geodesic $[I, A^{-\frac{1}{2}}BA^{-\frac{1}{2}}$ is naturally parametrized as

$$\gamma_0(t) = (A^{-\frac{1}{2}}BA^{-\frac{1}{2}})^t.$$

Applying the isometry $\Gamma_{A^{\frac{1}{2}}}$ we obtain the curve

$$\gamma(t) = \Gamma_{A^{\frac{1}{2}}}(\gamma_0(t)) = A^{\frac{1}{2}}(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})^t A^{\frac{1}{2}}$$

joining the points $\Gamma_{A^{\frac{1}{2}}}(I) = A$ and $\Gamma_{A^{\frac{1}{2}}}(A^{-\frac{1}{2}}BA^{-\frac{1}{2}}) = B$. Since $\Gamma_{A^{\frac{1}{2}}}$ is an isometry, this curve is the geodesic $[A, B]$. The equality $\delta_2(A, \gamma(t)) = t\delta_2(A, B)$ follows from the similar property for $\gamma_0(t)$ noted earlier. We see that by Lemma 1

$$\begin{aligned}\delta_2(A, B) &= \delta_2(I, A^{-\frac{1}{2}}BA^{-\frac{1}{2}}) \\ &= || \log I - \log(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})||_2 \\ &= || \log(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})||_2.\end{aligned}$$

□

4 Geometric Mean of Three or More Matrices

Ando, Li and Mathias [2] listed ten properties that a geometric mean of m matrices should satisfy, which we call the ALM properties. For simplicity, we report this list in the case $m = 3$.

1. Consistency with scalars. If A, B, C commute then $G(A, B, C) = (ABC)^{\frac{1}{3}}$.
2. Joint homogeneity. $G(\alpha A, \beta B, \gamma C) = (\alpha\beta\gamma)^{\frac{1}{3}}G(A, B, C)$, for α, β, $\gamma > 0$.
3. Permutation invariance. For any permutation $\pi(A, B, C)$ of (A, B, C), it holds that $G(A, B, C) = G(\pi(A, B, C))$.
4. Monotonicity. If $A \geq A_0$, $B \geq B_0$, and $C \geq C_0$, then $G(A, B, C) \geq G(A_0, B_0, C_0)$.
5. Continuity from above. If $\{A_n\}$, $\{B_n\}$, $\{C_n\}$ are monotonic decreasing sequences converging to A, B, C, respectively, then $\{G(A_n, B_n, C_n)\}$ converges to $G(A, B, C)$.
6. Congruence invariance. For any invertible S, it holds that
$$G(S^*AS, S^*BS, S^*CS) = S^*G(A, B, C)S.$$
7. Joint concavity. For $0 < \lambda < 1$,
$$\begin{aligned}&G(\lambda A_1 + (1-\lambda)A_2, \lambda B_1 + (1-\lambda)B_2, \lambda C_1 + (1-\lambda)C_2) \\ &\qquad \geq \lambda G(A_1, B_1, C_1) + (1-\lambda)G(A_2, B_2, C_2).\end{aligned}$$
8. Self-duality. $G(A^{-1}, B^{-1}, C^{-1}) = (G(A, B, C))^{-1}$.
9. Determinant identity.
$$\det G(A, B, C) = (\det A \det B \det C)^{\frac{1}{3}}.$$
10. Harmonic-geometric-arithmetic mean inequality.
$$(\frac{A^{-1} + B^{-1} + C^{-1}}{3})^{-1} \leq G(A, B, C) \leq \frac{A + B + C}{3}.$$

They also give a definition of geometric mean of m matrices by a iteration. Denote $G_2(A_1, A_2) = A_1\#A_2$ and suppose the mean G_{m-1} of $m-1$ matrices is already defined. Given $A_1, \ldots, A_m$, define m sequences by
$$A_i^{j+1} = G_{m-1}(A_1^j, A_2^j, \ldots, A_{i-1}^j, A_{i+1}^j, \ldots, A_m^j)$$
for $j = 1, 2, \ldots$ and $A_i^1 = A_i$. They proved that the sequences $\{A_i^j\}_{j=1}^{\infty}$ converge to a common matrix which satisfy the ALM properties. We denote by G_m^{ALM}.

5 Conclusion

So far, we state the proof of the theorem that the geometric mean is the middle point of the geodesic. This allows us to find a geometric meaning.

Consider the ALM geometric mean G_m^{ALM}, when $m = 3$, another geometric mean is defined in the same way by Bini, Meini and Poloni [15] which is denoted by G_3^{BMP}, but the iteration is replaced by

$$A_1^{j+1} = G_2(A_2^j, A_3^j)\#_{\frac{1}{3}} A_1^j,$$

$$A_2^{j+1} = G_2(A_1^j, A_3^j)\#_{\frac{1}{3}} A_2^j,$$

$$A_3^{j+1} = G_2(A_1^j, A_2^j)\#_{\frac{1}{3}} A_3^j,$$

where $A\#_t B$ is defined by $A^{\frac{1}{2}}(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})^t A^{\frac{1}{2}}$. It has been proved that the three matrix sequences have common limit which is different from the G_3^{ALM}, and satisfies the ALM properties.

The idea which the geometric mean can be viewed as the middle point of a geodesic is a very important result. In fact, for two positive numbers, we know that the arithmetic mean is the middle point of the geodesic connect these two scalar. In addition, for m positive numbers, the arithmetic mean minimizes the sum of the squared distances to the given points x_k

$$\bar{x} = \frac{1}{m}\sum_{k=1}^{m} x_k = \text{argmin}_{x>0} \sum_{k=1}^{m} \text{d}_e^2(x, x_k),$$

where $\text{d}_e(x, y) = |x - y|$ is the Euclidean distance in $\mathbb{R}$, and the geometric mean also minimizes the sum of the squared hyperbolic distances to the given points x_k

$$\tilde{x} = \sqrt[m]{x_1 x_2 \cdots x_m} = \text{argmin}_{x>0} \sum_{k=1}^{m} \text{d}_h^2(x, x_k),$$

where $\text{d}_h(x, y) = |\log x - \log y|$ is the hyperbolic distance between x and y on positive number. So Moakher [16] and Bhatia and Holroo [12] gave a definition of geometric mean of m positive definite matrices $A_1, \ldots, A_m$ which is defined by

$$G(A_1, \ldots, A_m) = \text{argmin}_{X \in P_n} \sum_{j=1}^{m} \delta_2^2(X, A_j)$$

. We call $G(A_1, \ldots, A_m)$ the barycenter or the center of mass. It can be shown that there is a unique X_0 such that $\sum_{j=1}^{m} \delta_2^2(X, A_j)$ is minimized. When $m = 2$, we have $G(A, B) = A\#B$.

In fact, the barycenter also satisfies the ALM properties and is not always the same as G_m^{ALM}. So there are many different geometric mean which satisfy the ALM properties.

The barycenter mean has been used in diverse applications such as elasticity, signal processing, medical imaging and computer vision. Recently, Barycenter means are extensively studied in generalized notions of centroids and barycenter to the broad class of information-theoretic distortion measures called Bregman divergences ([17]). It is also a comprehensive reference to generalized means in the framework of information geometry, already extensively applied in image processing. Another quite recent work shows the utility of geometric mean of KL-divergence as a complexity measure ([18]). Note that KL-divergence is known to asymptotically converge to Fisher information matrix. Therefore, the geometric mean of KL-divergence corresponds to the integral of geometric products of Fisher information matrices along the m-geodesic on the statistical manifold.

Acknowledgment

This paper is partially supported by a Taiwan NSC grant (NSC 102-2511-S-029-002-).

Competing Interests

The author declares that no competing interests exist.

References

[1] Pusz W, Woronowicz SL. Functional calculus for sesquilinear forms and the purification map, Reports on Mathematical Physics. 1975;8(2):159-170.

[2]Ando T, Li CK, Mathias R. Geometric means, Linear algebra and its applications. 2004;385:305-334.

[3] Kubo F, Ando T. Means of positive linear operators, Mathematische Annalen. 1980; 246(3): 205-224.

[4] Anderson WN, Trapp GE. Operator means and electrical networks, Proc. 1980 IEEE International Symposium on Circuits and Systems. 1980;523-527.

[5] Petz D. Matrix Analysis with some Applications; 2011.
Available: http://www.math.bme.hu/ petz

[6] Bhatia R. Positive definite matrices, Princeton University Press; 2009.

[7] Rao CR. Information and accuracy attainable in the estimation of statistical parameters, Bulletin of the Calcutta Mathematical Society. 1945;37(3):81-91.

[8] Amari S, Nagaoka H. Methods of information geometry, American Mathematical Society. 2000.

[9] Skovgaard LT. A Riemannian geometry of the multivariate normal model, Scandinavian Journal of Statistics. 1984;11:211-223.

[10] Ali SA, Cafaro C, Kim DM, Mancini S. The effect of the microscopic correlations on the information geometric complexity of Gaussian statistical models, Physica A 2010;389:3117.

[11] Carmo MP.Do. Riemannian geometry, Birkhäuser Boston; 1992.

[12] Bhatia R, Holbrook J. Riemannian geometry and matrix geometric means, Linear algebra and its applications. 2006;413(2):594-618.

[13] Schechter M. Principles of functional analysis, American Mathematical Society; 2002.

[14] Bhatia R. Matrix analysis, Springer. 1997.

[15] Bini D, Meini B, [Poloni] F. An effective matrix geometric mean satisfying the Ando-Li-Mathias properties, Mathematics of Computation. 2010;79(269):437-452.

[16] Moakher M. A differential geometric approach to the geometric mean of symmetric positive-definite matrices, SIAM Journal on Matrix Analysis and Applications. 2005;26(3):735-747.

[17] Nielsen F, Nock R. Sided and Symmetrized Bregman Centroids, IEEE Transaction on Information Theory; 2009.
Avaliable: http://www.lix.polytechnique.fr/ nielsen/SidedandSymmetrizedBregmanCentroids.pdf

[18] Funabashi M. Network Decomposition and Complexity Measures: An Information Geometrical Approach. Entropy. 2014;16(7):4132-4167.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
www.sciencedomain.org/review-history.php?iid=707&id=6&aid=6430