Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Linear Algebra and its Applications 413 (2006) 594–618
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

Riemannian geometry and matrix geometric means

Rajendra Bhatia [a], John Holbrook [b,*]

[a] *Indian Statistical Institute, 7 S.J.S. Sansanwal Marg, New Delhi 110 016, India*
[b] *Department of Mathematics and Statistics, University of Guelph, Guelph, Ont., Canada N1G 2W1*

Received 21 April 2005; accepted 30 August 2005
Available online 2 November 2005
Submitted by H. Schneider

Dedicated to Richard Brualdi

Abstract

The geometric mean of two positive definite matrices has been defined in several ways and studied by several authors, including Pusz and Woronowicz, and Ando. The characterizations by these authors do not readily extend to three matrices and it has been a long-standing problem to define a natural geometric mean of three positive definite matrices. In some recent papers new understanding of the geometric mean of two positive definite matrices has been achieved by identifying the geometric mean of A and B as the midpoint of the geodesic (with respect to a natural Riemannian metric) joining A and B. This suggests some natural definitions for a geometric mean of three positive definite matrices. We explain the necessary geometric background and explore the properties of some of these candidates.

AMS classification: 15A45; 15A48; 53B21; 53C22

Keywords: Positive definite matrix; Geometric mean; Riemannian manifold; Semi-parallelogram law; Gradient

* Corresponding author.
E-mail addresses: rbh@isid.ac.in (R. Bhatia), jholbroo@uoguelph.ca (J. Holbrook).

0024-3795/$ - see front matter
doi:10.1016/j.laa.2005.08.025

1. Introduction

The goals of this article are partly expository but include an analysis of recent approaches to the definition of a geometric mean for three (or more) positive definite matrices. Effective definitions for this concept have long been elusive although related ideas have appeared as our work was in progress (see [2,9]). We first review some standard constructions from Riemannian geometry with stress on the matrix analytic aspects of these constructions. We then explain how this setting allows a better understanding of the geometric mean of two positive definite matrices. Finally we turn to the problem of extending these ideas to three matrices.

One well-established interpretation of the "geometric mean" $A\#B$ of two positive definite matrices A, B says that

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

There is a natural hyperbolic geometry (i.e. one with nonpositive curvature) on the space $\mathbb{P}_n$ of $n \times n$ positive definite matrices in which this $A\#B$ has a pleasing conceptual meaning: $A\#B$ is the midpoint of the geodesic joining A and B (see for example [5]). A preferred interpretation of the "geometric mean" of *three* matrices $A, B, C \in \mathbb{P}_n$ is not so well-established. There are several competing definitions, helpfully discussed in [2]. In that paper the authors highlight a particular interpretation, which we denote by $\mathrm{alm}(A, B, C)$, that is obtained by a limit procedure successively replacing the vertices of the "triangle" by the geometric means of its sides. More precisely, starting from $\Delta_0 = \{A_0, B_0, C_0\}$ we define by induction

$$\Delta_{m+1} = \{A_m\#B_m, B_m\#C_m, C_m\#A_m\}, \tag{2}$$

and set

$$\mathrm{alm}(\Delta_0) = \lim_{m\to\infty} \Delta_m, \tag{3}$$

or to be exact $\mathrm{alm}(\Delta_0) = M_\infty$ where $\lim_{m\to\infty} \Delta_m = (M_\infty, M_\infty, M_\infty)$.

In these notes we explore the hyperbolic-geometry setting for constructions such as (3), pointing out for example that the convergence result essential to (3) is especially evident in that geometric setting. We examine certain other candidates for the "geometric mean" of a triple $\Delta = \{A, B, C\}$ that appear natural in the geometric setting, with emphasis on the "least squares" point Z minimizing $\delta^2(A, Z) + \delta^2(B, Z) + \delta^2(C, Z)$, where $\delta(A, Z)$ denotes the geodesic distance from A to Z. This point, denoted by $\mathrm{ls}(\Delta)$, is that $Z \in \mathbb{P}_n$ (it turns out to be unique) such that, for all $Y \in \mathbb{P}_n$, $\mathrm{ss}_\Delta(Z) \leqslant \mathrm{ss}_\Delta(Y)$, where

$$\mathrm{ss}_\Delta(Y) = \delta^2(A, Y) + \delta^2(B, Y) + \delta^2(C, Y). \tag{4}$$

This construction is classic, going back to Élie Cartan in the early 20th century (see for example [6, p. 178]), but it is perhaps only recently that $\mathrm{ls}(\Delta)$ has been considered as a matrix geometric mean. We have learned that our own work runs parallel, in several respects, to that of Moakher in [9]. We have tried to make our account self-contained wherever it is reasonable to do so, and have included many geometric details that may be convenient for readers coming from a background in matrix analysis similar to our own.

In this paragraph we provide a guide to some of the literature related to this article; other references appear in later sections. The effective definition of the geometric mean for two positive definite matrices seems to have first appeared in Pusz and Woronowicz [12]. Ando [1] provided the first systematic development of many of its basic properties, giving equivalent characterizations and applications to matrix inequalities that are otherwise difficult to prove. Trapp [13] is a good survey of matrix means, including the geometric, and relates these concepts to the earlier electrical engineering literature. The geometric mean has been linked to differential geometry in Corach–Porta–Recht [7] and Lawson–Lim [8], for example.

2. The natural metric on $\mathbb{P}_n$

Here we review the definition of the hyperbolic geometry for $\mathbb{P}_n$ and obtain some of its properties, notably the "exponential metric increasing property" and the "semi-parallelogram law". In part this is a reworking of (some of) the material in [5]. Our notation includes the following: $M_n(\mathbb{C})$ denotes the space (★-algebra) of all $n \times n$ complex matrices, GL_n denotes the space (group) of all invertible elements in $M_n(\mathbb{C})$, $\mathbb{S}_n$ denotes the space (real-linear subspace) of all self-adjoint elements in $M_n(\mathbb{C})$, and $\mathbb{P}_n$ denotes the space (cone) of all positive definite (pd) elements in $M_n(\mathbb{C})$.

A Riemannian metric on $\mathbb{P}_n$ is determined locally (at A) by the relation

$$ds = \|A^{-\frac{1}{2}} dA A^{-\frac{1}{2}}\|_2, \tag{5}$$

where $\|X\|_2$ denotes the Frobenius, Hilbert–Schmidt, or Schatten-2 norm of a matrix $X \in M_n(\mathbb{C})$, i.e. $\|X\|_2 = (\sum_{i,j} |x_{ij}|^2)^{\frac{1}{2}}$. The mnemonic (5) is interpreted as a recipe for computing the "length" $L(\gamma)$ of a (differentiable) path $\gamma : [a, b] \to \mathbb{P}_n$:

$$L(\gamma) = \int_a^b \|\gamma^{-\frac{1}{2}}(t)\gamma'(t)\gamma^{-\frac{1}{2}}(t)\|_2 \,\mathrm{d}t. \tag{6}$$

A key observation is that we have a large class of bijections $\Gamma : \mathbb{P}_n \to \mathbb{P}_n$ that are isometric with respect to this notion of length. Indeed, given any $X \in GL_n$, let $\Gamma_X : \mathbb{P}_n \to \mathbb{P}_n$ be defined by $\Gamma_X(A) = XAX^*$. Given a path γ as above, the composition $\Gamma_X \circ \gamma : [a, b] \to \mathbb{P}_n$ is another such path and we have

$$(\forall X \in GL_n) \quad L(\Gamma_X \circ \gamma) = L(\gamma), \tag{7}$$

since for each t

$$\begin{aligned} &\|(X\gamma(t)X^*)^{-\frac{1}{2}}(X\gamma(t)X^*)'(X\gamma(t)X^*)^{-\frac{1}{2}}\|_2 \\ &\quad = \|\gamma^{-\frac{1}{2}}(t)\gamma'(t)\gamma^{-\frac{1}{2}}(t)\|_2. \end{aligned} \tag{8}$$

To obtain (8), recall that for any $Y \in M_n(\mathbb{C})$ we have $\|Y\|_2^2 = \sum_1^n s_k^2(Y)$, where $s_k(Y)$ denotes the kth singular value of Y. In the case of positive definite matrices such as $A^{-\frac{1}{2}}BA^{-\frac{1}{2}}$ $(A, B \in \mathbb{P}_n)$ we see that $\|A^{-\frac{1}{2}}BA^{-\frac{1}{2}}\|_2$ depends only on the eigenvalues of $A^{-\frac{1}{2}}BA^{-\frac{1}{2}}$; these, by similarity, are the same as the eigenvalues of BA^{-1}. Thus (8) follows from the observation that $(X\gamma'(t)X^*)(X\gamma(t)X^*)^{-1}$ and $\gamma'(t)\gamma^{-1}(t)$ have the same eigenvalues: indeed, $(X\gamma'(t)X^*)(X\gamma(t)X^*)^{-1} = X(\gamma'(t)\gamma^{-1}(t))X^{-1}$.

Based on the notion of length, introduced above, we define the geodesic distance $\delta(A, B)$ between any two $A, B \in \mathbb{P}_n$:

$$\delta(A, B) = \inf\{L(\gamma) : \gamma \text{ is a (differentiable) path from } A \text{ to } B\}. \tag{9}$$

We shall soon see that this infimum is attained by a path uniquely determined by A and B. This path is called the geodesic joining A to B and it will be denoted by $[A, B]$; this notation should not be confused with the Lie bracket notation (which would also make sense here!). In any case, it is clear that (9) defines a metric on $\mathbb{P}_n$. In particular, the triangle inequality $\delta(A, B) \leqslant \delta(A, C) + \delta(C, B)$ follows from the observation that a path γ_1 from A to C can be adjoined to a path γ_2 from C to B to obtain a path "$\gamma_1 + \gamma_2$" from A to B having length $L(\gamma_1) + L(\gamma_2)$. By definition $\delta(A, B) \leqslant L(\gamma_1 + \gamma_2) = L(\gamma_1) + L(\gamma_2)$, and taking infima on the right we obtain the triangle inequality.

Because of the isometry of the mapping Γ_X with respect to length L, it is clear that each Γ_X is also an isometry with respect to δ:

$$(\forall X \in GL_n, \quad \forall A, B \in \mathbb{P}_n) \quad \delta(\Gamma_X(A), \Gamma_X(B)) = \delta(A, B). \tag{10}$$

This observation, together with the identification of certain special geodesics, will allow us to find $\delta(A, B)$ directly and to compute the geodesic $[A, B]$ explicitly.

The main new ingredient is the infinitesimal exponential metric increasing property (IEMI):

$$(H, K \in \mathbb{S}_n) \quad \|(\mathrm{e}^H)^{-\frac{1}{2}} D\mathrm{e}^H(K)(\mathrm{e}^H)^{-\frac{1}{2}}\|_2 \geqslant \|K\|_2, \tag{11}$$

where $D\mathrm{e}^H$ denotes the Fréchet derivative of the (matrix) exponential function exp at H. This is a linear map on $\mathbb{S}_n$ and its action is given by the formula

$$D\mathrm{e}^H(K) = \lim_{t \to 0} \frac{\mathrm{e}^{H+tK} - \mathrm{e}^H}{t}.$$

Let $f : \mathbb{R} \to \mathbb{R}$ be any differentiable function (here we are concerned with the case $f(t) = \exp(t) = \mathrm{e}^t$). It is well-known that, working with respect to an orthonormal basis of eigenvectors for $H \in \mathbb{S}_n$,

$$(\forall K \in \mathbb{S}_n) \quad [Df(H)](K) = \left[\frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j}\right] \circ K, \tag{12}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of H and $\circ$ denotes here the Schur or entry-wise product of $n \times n$ matrices. See, for example, [4, Theorem V.3.3]. Note that when $\lambda_i = \lambda_j$ (for example, when $i = j$) the corresponding entry in the "Löwner matrix" on the right of (12) is interpreted as $f'(\lambda_i)$.

Thus we may compute $(\mathrm{e}^H)^{-\frac{1}{2}} D\mathrm{e}^H(K)(\mathrm{e}^H)^{-\frac{1}{2}}$ as

$$\operatorname{diag}\{\mathrm{e}^{-\lambda_i/2}\}\left(\left[\frac{\mathrm{e}^{\lambda_i}-\mathrm{e}^{\lambda_j}}{\lambda_i-\lambda_j}\right]\circ K\right)\operatorname{diag}\{\mathrm{e}^{-\lambda_j/2}\}=\left[\frac{\mathrm{e}^{\frac{\lambda_i-\lambda_j}{2}}-\mathrm{e}^{-\frac{\lambda_i-\lambda_j}{2}}}{\lambda_i-\lambda_j}\right]\circ K,$$

so that (11) follows from the elementary fact that $\frac{\mathrm{e}^{t/2}-\mathrm{e}^{-t/2}}{t}\geqslant 1$ for all t.

Proposition 1. *Given any differentiable path* $\gamma:[a,b]\to\mathbb{P}_n$, *parametrized as* $\gamma(t)=\mathrm{e}^{H(t)}$ (*i.e. setting* $H(t)=\log\gamma(t)$),

$$L(\gamma)\geqslant\int_a^b\|H'(t)\|_2\,\mathrm{d}t. \tag{13}$$

For any $A, B\in\mathbb{P}_n$,

$$\delta(A,B)\geqslant\|\log A-\log B\|_2. \tag{14}$$

Proof. By the chain rule $\gamma'(t)=D\mathrm{e}^{H(t)}(H'(t))$ so that (13) follows directly from (6) and (11). Let $\gamma:[a,b]\to\mathbb{P}_n$ be any (differentiable) path from A to B. Then $H(t)=\log\gamma(t)$ defines a path in the Euclidian space $(\mathbb{S}_n,\|\cdot\|_2)$. The RHS of (13) is just the Euclidian length of $H(\cdot)$ so that it is bounded below by $\|H(a)-H(b)\|_2=\|\log A-\log B\|_2$. Thus, for any such γ, we have $L(\gamma)\geqslant\|\log A-\log B\|_2$, and (14) follows. □

Proposition 2. *Let* $A, B\in\mathbb{P}_n$ *be commuting matrices. Then the exponential function exp maps the Euclidian line segment* $[\log A,\log B]_2\subset\mathbb{S}_n$ *isometrically to the geodesic* $[A,B]$ *in* $\mathbb{P}_n$. *In particular*, $\delta(A,B)=\|\log A-\log B\|_2$.

Proof. Consider the path claimed to trace out the geodesic $[A,B]$, namely $\gamma(t)=\exp((1-t)\log A+t\log B)$. We must verify that $\gamma:[0,1]\to\mathbb{P}_n$ traverses the unique path of shortest length joining A to B. Because A and B commute, $\gamma(t)=A^{1-t}B^t$ and $\gamma'(t)=(\log B-\log A)\gamma(t)$. Hence, directly from (6),

$$L(\gamma)=\int_0^1\|\log A-\log B\|_2\,\mathrm{d}t=\|\log A-\log B\|_2. \tag{15}$$

Proposition 1 says that $\|\log B-\log A\|_2$ is the least possible length for a path from A to B so that $\gamma(t)=A^{1-t}B^t$ does attain the minimum. Any other path $\tilde{\gamma}$ with this length would, applying (13) to $\tilde{\gamma}$, be such that $\tilde{H}(t)=\log\tilde{\gamma}(t)$ had Euclidian length equal to $\|\log B-\log A\|_2$, i.e. the distance between its end-points. Thus $\tilde{H}(\cdot)$ would be a reparametrization of $[\log A,\log B]_2$. It is the affine parametrization $H(t)=(1-t)\log A+t\log B$, however, that maps isometrically to $[A,B]$ throughout the whole interval (apply the calculation of (15) to the restriction of $H(\cdot)$ to any subinterval $[a,b]$ of $[0,1]$). □

Note that Proposition 2 tells us that when A and B commute the natural parametrization of $[A, B]$ in $\mathbb{P}_n$ is given by $\gamma(t) = A^{1-t}B^t$ in the sense that $\delta(A, \gamma(t)) = t\delta(A, B)$ for each $t \in [0, 1]$. Combining this information with the isometries Γ_X we can proceed to the general case.

Proposition 3. *For any* $A, B \in \mathbb{P}_n$ *the geodesic* $[A, B]$ *is naturally parametrized by* $\gamma : [0, 1] \to [A, B]$ *with*

$$\gamma(t) = A^{\frac{1}{2}} \left(A^{-\frac{1}{2}} B A^{-\frac{1}{2}}\right)^t A^{\frac{1}{2}}, \tag{16}$$

in the sense that $\delta(A, \gamma(t)) = t\delta(A, B)$ *for each* $t \in [0, 1]$. *Moreover*,

$$\delta(A, B) = \| \log(A^{-\frac{1}{2}} B A^{-\frac{1}{2}})\|_2. \tag{17}$$

Proof. The matrices I and $A^{-\frac{1}{2}} B A^{-\frac{1}{2}}$ commute so that $\gamma_0(t) = (A^{-\frac{1}{2}} B A^{-\frac{1}{2}})^t$ naturally parametrizes $[I, A^{-\frac{1}{2}} B A^{-\frac{1}{2}}]$. Applying the isometry $\Gamma_{A^{\frac{1}{2}}}$ to γ_0 we obtain the natural parametrization of $[A, B] = \left[\Gamma_{A^{\frac{1}{2}}}(I), \Gamma_{A^{\frac{1}{2}}}(A^{-\frac{1}{2}} B A^{-\frac{1}{2}})\right]$, namely $\gamma(t) = \Gamma_{A^{\frac{1}{2}}}(\gamma_0(t)) = A^{\frac{1}{2}}(A^{-\frac{1}{2}} B A^{-\frac{1}{2}})^t A^{\frac{1}{2}}$, as claimed. Moreover, $\delta(A, B) = \delta(I, A^{-\frac{1}{2}} B A^{-\frac{1}{2}}) = \| \log I - \log(A^{-\frac{1}{2}} B A^{-\frac{1}{2}})\|_2 = \| \log(A^{-\frac{1}{2}} B A^{-\frac{1}{2}})\|_2$. $\square$

As a special case we see that $\gamma(1/2)$, the geodesic midpoint of A and B, is given by $A^{\frac{1}{2}}(A^{-\frac{1}{2}} B A^{-\frac{1}{2}})^{\frac{1}{2}} A^{\frac{1}{2}}$, which is the geometric mean $A\#B$ of A and B as defined by (1). The formula (1) appears unlikely at first glance since the symmetry in A and B is obscured. Proposition 3 reveals the symmetry geometrically since, reversing the roles of A and B, the proposition tells us that the midpoint of $[A, B]$ can equally well be expressed as $B^{\frac{1}{2}}(B^{-\frac{1}{2}} A B^{-\frac{1}{2}})^{\frac{1}{2}} B^{\frac{1}{2}}$. Note also that $\delta(A^{-1}, B^{-1}) = \delta(A, B)$ (see (17), for example); this makes it clear geometrically that the geometric mean respects matrix inversion, i.e. $A^{-1}\#B^{-1} = (A\#B)^{-1}$.

Proposition 4. *If, for some* $A, B \in \mathbb{P}_n$, I *lies on the geodesic* $[A, B]$, *then* A *and* B *commute*, $[A, B]$ *is the isometric image via the map exp of a line segment through* 0 *in* $\mathbb{S}_n$, *and*

$$\log B = -\frac{1-t}{t} \log A, \tag{18}$$

where $t = \delta(A, I)/\delta(A, B)$.

Proof. From Proposition 3 we know that, for some t, $I = A^{\frac{1}{2}}(A^{-\frac{1}{2}} B A^{-\frac{1}{2}})^t A^{\frac{1}{2}}$. Thus $B = A^{\frac{1}{2}} A^{-1/t} A^{\frac{1}{2}} = A^{-(1-t)/t}$ so that A and B commute and (18) holds. Proposition 2 ensures that $[A, B]$ is the isometric image of the Euclidian line segment $[\log A, \log B]_2$ in $\mathbb{S}_n$ and, of course, $0 = \log I$ lies in this segment. $\square$

The following proposition establishes the "semi-parallelogram law" for $\mathbb{P}_n$.

Proposition 5. *Given $A, B \in \mathbb{P}_n$, let M $(= A\#B)$ be the midpoint of the geodesic $[A, B]$. Then for any $C \in \mathbb{P}_n$ we have*

$$\delta^2(M, C) \leqslant \frac{\delta^2(A, C) + \delta^2(B, C)}{2} - \frac{1}{4}\delta^2(A, B). \tag{19}$$

Proof. Applying the isometry $\Gamma_{M^{-\frac{1}{2}}}$ to all the matrices involved, we may assume that $M = I$. By Proposition 4 we have $\log B = -\log A$ and $\delta(A, B) = \|\log A - \log B\|_2 = 2\|\log A\|_2$. That proposition also applies to $[M, C] = [I, C]$ so that $\delta(M, C) = \|\log M - \log C\|_2 = \|\log C\|_2$. In the Euclidian space $(\mathbb{S}_n, \|\cdot\|_2)$ we have

$$\|\log C\|_2^2 + \|\log A\|_2^2 = \frac{\|\log C - \log A\|_2^2 + \|\log C + \log A\|_2^2}{2}$$

(a form of the parallelogram law). Recalling the relations above, we may write this as

$$\delta^2(M, C) + (\delta(A, B)/2)^2 = \frac{\|\log C - \log A\|_2^2 + \|\log C - \log B\|_2^2}{2}$$

so that, by Proposition 1,

$$\delta^2(M, C) + \frac{1}{4}\delta^2(A, B) \leqslant \frac{\delta^2(A, C) + \delta^2(B, C)}{2}. \qquad \square$$

The impact of Propositions 1 and 4 may be summarized by saying that the mapping exp: $\mathbb{S}_n \to \mathbb{P}_n$ is isometric on line segments through 0 and is metric nondecreasing in general. These features are conventionally expressed as the "exponential metric increasing" property (EMI). The semi-parallelogram law and EMI reflect the nonpositive curvature of $(\mathbb{P}_n, \delta)$, though we'll not define curvature formally here. Another useful aspect of this nonpositive curvature will be the fact that, for any $A, B, C \in \mathbb{P}_n$ and any $t \in [0, 1]$

$$\delta(A\#_t B, A\#_t C) \leqslant t\delta(B, C), \tag{20}$$

where we use $A\#_t B$ to denote the point T on $[A, B]$ such that $\delta(A, T) = t\delta(A, B)$ (in particular we have $A\#_{\frac{1}{2}} B = A\#B$, the midpoint). In terms of this notation, Proposition 3 yields the relation

$$A\#_t B = A^{\frac{1}{2}}(A^{-\frac{1}{2}} B A^{-\frac{1}{2}})^t A^{\frac{1}{2}}. \tag{21}$$

The relation (20) is known as "convexity of the metric". See for example [5], p. 218. Here we first provide a proof for the case we need later ($t = 1/2$), basing it on the semi-parallelogram law.

Proposition 6. *For any $A, B, C \in \mathbb{P}_n$ we have*

$$\delta(A\#B, A\#C) \leqslant \frac{1}{2}\delta(B, C). \tag{22}$$

Proof. Let $M_1 = A\#B$ and $M_2 = A\#C$. Since M_1 is the midpoint of the side of the geodesic triangle $\{A, B, C\}$ opposite C, Proposition 5 (the semi-parallelogram law) tells us that

$$\delta^2(C, M_1) \leqslant \frac{\delta^2(C, A) + \delta^2(C, B)}{2} - \frac{1}{4}\delta^2(A, B).$$

Since M_2 is the midpoint of the side of the geodesic triangle $\{A, M_1, C\}$ opposite M_1, Proposition 5 also tells us that

$$\delta^2(M_1, M_2) \leqslant \frac{\delta^2(M_1, C) + \delta^2(M_1, A)}{2} - \frac{1}{4}\delta^2(C, A).$$

Applying the first inequality to the second, we obtain

$$\begin{aligned}\delta^2(M_1, M_2) \leqslant{}& \frac{1}{4}\delta^2(C, A) + \frac{1}{4}\delta^2(C, B) - \frac{1}{8}\delta^2(A, B)\\ &+ \frac{1}{2}\delta^2(M_1, A) - \frac{1}{4}\delta^2(C, A).\end{aligned}$$

Since $\delta(M_1, A) = \frac{1}{2}\delta(A, B)$, the RHS simplifies to $\frac{1}{4}\delta^2(C, B)$ and (22) follows. □

As a corollary we obtain (20) in a somewhat more general form: (20) follows from (23) by setting $B' = C' = A$.

Corollary. *Given* $B, C, B', C' \in \mathbb{P}_n$, *the function* $f(t) = \delta(B'\#_t B, C'\#_t C)$ *is convex on* $[0, 1]$, *i.e.*

$$\delta(B'\#_t B, C'\#_t C) \leqslant (1 - t)\delta(B', C') + t\delta(B, C). \tag{23}$$

Proof. Since f is continuous it is sufficient to prove that it is midpoint-convex. Let $M_1 = B'\#B$, $M_2 = C'\#C$, and $M = B'\#C$. Proposition 6 implies that $\delta(M_1, M) \leqslant \frac{1}{2}\delta(B, C)$ and $\delta(M, M_2) \leqslant \frac{1}{2}\delta(B', C')$. Hence

$$\delta(M_1, M_2) \leqslant \delta(M_1, M) + \delta(M, M_2) \leqslant \frac{1}{2}\delta(B, C) + \frac{1}{2}\delta(B', C'). \qquad \square$$

3. Convex hulls

We say a subset $\mathscr{S}$ of $\mathbb{P}_n$ is *convex* if $A, B \in \mathscr{S} \Longrightarrow [A, B] \subseteq \mathscr{S}$. Evidently the intersection of any family of convex subsets is itself convex. It is then natural to define the *convex hull* $\mathrm{conv}(\mathscr{T})$ of a subset $\mathscr{T}$ of $\mathbb{P}_n$ by

$$\mathrm{conv}(\mathscr{T}) = \bigcap\{\mathscr{S} : \mathscr{T} \subseteq \mathscr{S} \text{ and } \mathscr{S} \text{ is convex}\}. \tag{24}$$

Thus conv($\mathscr{T}$) is the smallest convex set containing $\mathscr{T}$. Since $[A, B]$ is itself convex, it is clear that for any *two* points $A, B \in \mathbb{P}_n$ we have $\mathrm{conv}(\{A, B\}) = [A, B]$. The convex hull of *three* points $A, B, C \in \mathbb{P}_n$ is, however, harder to describe in general. Note 6.1.3.1 in [3] comments on this problem. Although it is natural to regard any $A\#_t B$ ($t \in [0, 1]$) as a "geometric" convex combination of A and B, we do not have a consistent notion of geometric convex combination for three or more elements. Nevertheless, there is a more "constructive" approach to computing conv($\mathscr{T}$) than (24) reveals.

Proposition 7. *Given* $\mathscr{T} = \mathscr{T}_0 \subseteq \mathbb{P}_n$, *define inductively the sets* $\mathscr{T}_m$ *via*

$$\mathscr{T}_{m+1} = \bigcup\{[A, B] : A, B \in \mathscr{T}_m\}.$$

Then $\mathrm{conv}(\mathscr{T}) = \bigcup_0^\infty \mathscr{T}_m$.

Proof. It is clear by induction that each $\mathscr{T}_m \subseteq \mathrm{conv}(\mathscr{T})$. Hence $\bigcup_0^\infty \mathscr{T}_m \subseteq \mathrm{conv}(\mathscr{T})$. It only remains to show that $\bigcup_0^\infty \mathscr{T}_m$ is convex. Note that $\mathscr{T}_{m+1} \supseteq \mathscr{T}_m$ since $A \in [A, B]$. Thus given any particular $A, B \in \bigcup_0^\infty \mathscr{T}_m$ there is some fixed m' such that $A, B \in \mathscr{T}_{m'}$. Then $[A, B] \subseteq \mathscr{T}_{m'+1} \subseteq \bigcup_0^\infty \mathscr{T}_m$. $\square$

If, in the construction of Proposition 7, we take a "triangle" $\mathscr{T} = \{A, B, C\}$, it is clear that $\mathscr{T}_1$ is the union of the "edges", i.e. $\mathscr{T}_1 = [A, B] \cup [B, C] \cup [C, A]$. However, $\mathscr{T}_2$ is not in general a "surface", as we might expect by analogy with Euclidian space, but rather a "fatter" object. Fig. 1 attempts to portray a part of $\mathscr{T}_2$.

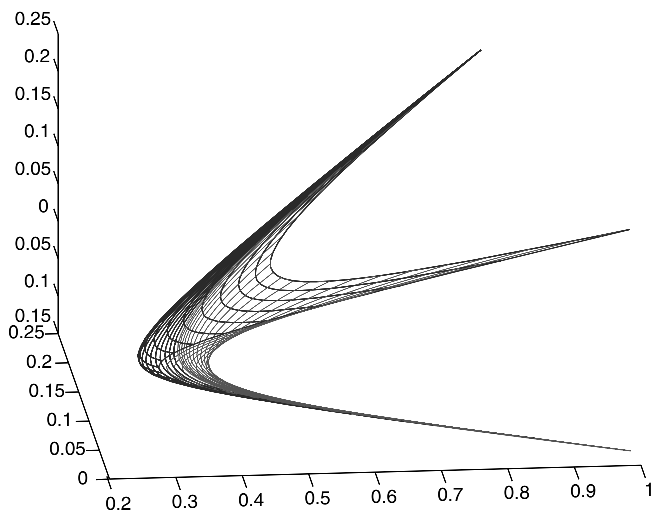

Fig. 1. Two pieces of conv(A, B, C), namely the "surfaces" $[A, [B, C]]$ and $[B, [C, A]]$.

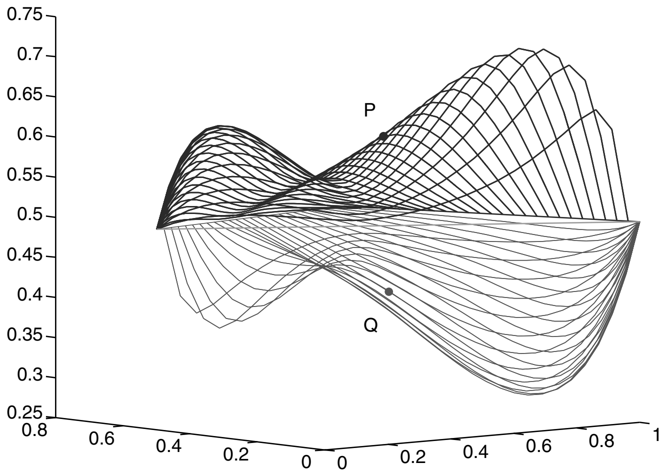

Fig. 2. In general, conv(A, B, C) lacks consistent convex coordinates.

The blue[1] "surface" in Fig. 1 represents the union of geodesics joining A to points of the opposite edge $[B, C]$. That is, the blue curves sketch out $\bigcup\{[A, B\#_t C], t \in [0, 1]\}$, which we may denote by $[A, [B, C]]$; again, we warn the reader that we are *not* using the Lie bracket notation. The red curves, on the other hand, sketch out $[B, [C, A]]$, and it is clear from this example that $[A, [B, C]]$ and $[B, [C, A]]$ do not in general belong to a simple surface bounded by $\mathscr{T}_1$. Indeed, it appears in Fig. 1 that $[A, [B, C]]$ and $[B, [C, A]]$ intersect only along the edges $\mathscr{T}_1$ of the triangle. The points A, B, and C in this demonstration have been chosen "at random" from $\mathbb{P}_2$ and normalized so that $\|A\|_2 = 1$, etc. The coordinates used to plot Fig. 1 correspond to a choice of orthonormal basis in $\mathbb{S}_2$. While the dimension of $\mathbb{S}_2$ is 4, Fig. 1 plots the projection of (some of) $\mathscr{T}_2$ on the subspace spanned by 3 of the 4 orthonormal basis elements. We remark that (except for Fig. 3b) the matrices chosen are "generic", but that the views have been selected carefully to reveal certain features.

In Fig. 2 we focus on the impossibility (in general) of assigning consistent convex coordinates to the points of conv$(\{A, B, C\})$. Indeed, it turns out that it is not appropriate to speak of "consistent convex coordinates" in this setting; see however our discussion of the "Cartan surface" following Proposition 17. Here $\mathscr{T}_1 = [A, B] \cup [B, C] \cup [C, A]$ is artificially rendered by the sides of an affine triangle. The blue geodesics representing $[A, [B, C]]$ are plotted with a positive vertical displacement proportional to their distances from the points in $[B, [C, A]]$ to which they might be expected to correspond (and which they would indeed match perfectly if,

[1] For interpretation of color in figures, the reader is referred to the Web version of this article.

for example, A, B, and C commuted—see Proposition 9). As an illustration, $P = A\#_{\frac{2}{3}}(B\#_{\frac{1}{2}}C)$, which under favorable conditions would be a "centroid" for $\{A, B, C\}$, is plotted a distance $\|P - Q\|_2$ above the affine triangle, where Q is the "matching" point in $[B, [C, A]]$: $Q = B\#_{\frac{2}{3}}(C\#_{\frac{1}{2}}A)$. Likewise the red geodesics representing $[B, [C, A]]$ are plotted *below* the affine triangle by amounts related to their distances from "matching" elements in $[A, [B, C]]$.

We next examine those favorable cases in which conv$(\{A, B, C\})$ *can* be viewed as a surface spanning $[A, B] \cup [B, C] \cup [C, A]$. We say $A, B, C \in \mathbb{P}_n$ are Γ-commuting if there exists $X \in GL_n$ such that $\Gamma_X(A), \Gamma_X(B), \Gamma_X(C)$ commute with one another. Note that A, B, C themselves need not commute, though that situation is a special case of Γ-commutativity. For example, $\{A, B, B\}$ Γ-commute for arbitrary $A, B \in \mathbb{P}_n$ (take $X = A^{-\frac{1}{2}}$). On the other hand, Γ-commutativity is quite a restrictive condition, as (b) of Proposition 8 makes clear.

Proposition 8. *Given $A, B, C \in \mathbb{P}_n$, the following conditions are equivalent*:

(a) A, B, C *Γ-commute, i.e.*

$(\exists X \in GL_n)$ *such that* $\Gamma_X(A), \Gamma_X(B), \Gamma_X(C)$ *commute*;

(b) $AB^{-1}C = CB^{-1}A$;

(c) $A^{-\frac{1}{2}}BA^{-\frac{1}{2}}$ *and* $A^{-\frac{1}{2}}CA^{-\frac{1}{2}}$ *commute*.

Proof. Suppose that (a) holds. Since XAX^*, XBX^*, XCX^* commute, so do XAX^*, $(XBX^*)^{-1}$, XCX^* so that

$$XAX^*(XBX^*)^{-1}XCX^* = XCX^*(XBX^*)^{-1}XAX^*,$$

i.e. $XAB^{-1}CX^* = XCB^{-1}AX^*$. Thus (a) implies (b).

Reversing the steps above we see that if (b) holds we have, for any $Y \in \mathbb{P}_n$, $\Gamma_Y(A)\Gamma_Y(B)^{-1}\Gamma_Y(C) = \Gamma_Y(C)\Gamma_Y(B)^{-1}\Gamma_Y(A)$. Taking $Y = A^{-\frac{1}{2}}$ we have $\Gamma_{A^{-\frac{1}{2}}}(B)^{-1}\Gamma_{A^{-\frac{1}{2}}}(C) = \Gamma_{A^{-\frac{1}{2}}}(C)\Gamma_{A^{-\frac{1}{2}}}(B)^{-1}$. Thus (b) implies (c). Finally, (c) provides a specific X, namely $A^{-\frac{1}{2}}$, for (a). $\square$

Remark. The symmetric condition of Γ-commutativity is equivalent to the easily computable but less-obviously symmetric (b). It is easy to check directly that condition (b) of Proposition 8 *is* in fact symmetric in A, B, C.

Proposition 9. *Let A, B, C be Γ-commuting and choose any $X \in GL_n$ such that $\Gamma(A), \Gamma(B), \Gamma(C)$ commute, where $\Gamma = \Gamma_X$. Then conv$(\{A, B, C\})$ is isometric to the affine triangle conv$(\{\log\Gamma(A), \log\Gamma(B), \log\Gamma(C)\})$ (in the Euclidian space $(\mathbb{S}_n, \|\cdot\|_2)$) via the map $\Gamma^{-1} \circ \exp$. Thus consistent "convex" coordinates may be assigned to points in conv$(\{A, B, C\})$ with (r, s, t) $(r + s + t = 1)$ corresponding to $\Gamma^{-1}((\Gamma(A))^r(\Gamma(B))^s(\Gamma(C))^t)$.*

Proof. Since Γ is a δ-isometry, we need only show that exp maps conv($\{\log \Gamma(A), \log \Gamma(B), \log \Gamma(C)\}) \subset \mathbb{S}_n$ isometrically to conv($\{\Gamma(A), \Gamma(B), \Gamma(C)\}$).

The fact that exp is isometric on conv($\{\log \Gamma(A), \log \Gamma(B), \log \Gamma(C)\}$) is a consequence of Proposition 2, and commutativity ensures that

$$\exp(r \log \Gamma(A) + s \log \Gamma(B) + t \log \Gamma(C)) = (\Gamma(A))^r (\Gamma(B))^s (\Gamma(C))^t.$$

It remains to show that

$$\mathscr{T} = \{(\Gamma(A))^r (\Gamma(B))^s (\Gamma(C))^t : r, s, t \geqslant 0, r + s + t = 1\}$$

is, in fact, conv($\{\Gamma(A), \Gamma(B), \Gamma(C)\}$). Certainly $\mathscr{T}$ contains each of $\Gamma(A)$, $\Gamma(B)$, and $\Gamma(C)$, and $\mathscr{T}$ is convex: if $T(r, s, t) = (\Gamma(A))^r (\Gamma(B))^s (\Gamma(C))^t \in \mathscr{T}$ and $T(r', s', t') \in \mathscr{T}$ then

$$T(r, s, t)\#_u T(r', s', t') = T(r(1 - u) + r'u, s(1 - u) + s'u, t(1 - u) + t'u)$$

is also in $\mathscr{T}$. In fact, we see that if $\{\Gamma(A), \Gamma(B), \Gamma(C)\}$ is denoted by $\mathscr{T}_0$ then, in terms of the notation of Proposition 7, $\mathscr{T} = \mathscr{T}_2$. $\square$

Figs. 3a and 3b illustrate the effect of the Γ-commutativity in Proposition 9. In Fig. 3a we see a simulation of conv($\{A, B, C\}$) when A, B, C are chosen at random in $\mathbb{P}_2$. In Fig. 3b we see the same sort of simulation applied to a triple of matrices chosen so that they Γ-commute. The simulations were computed via finite approximations to the towers of sets occurring in Proposition 7. If $\{A, B, C\}$ is denoted by $\mathscr{T}_0$ then,

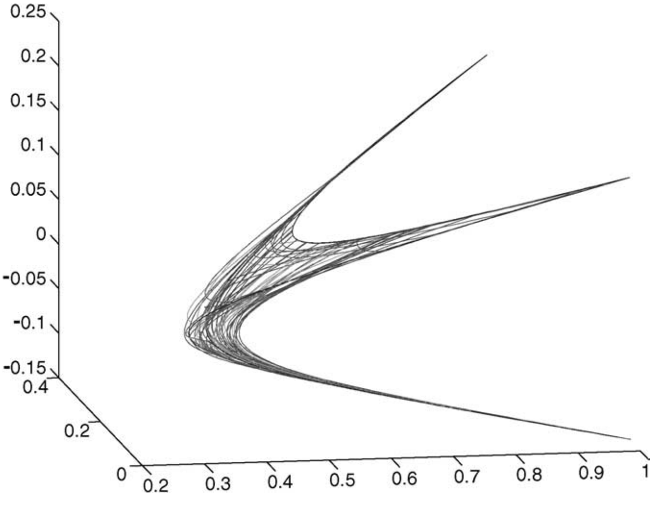

Fig. 3a. Part of conv(A, B, C) where A, B, C are random in P_2.

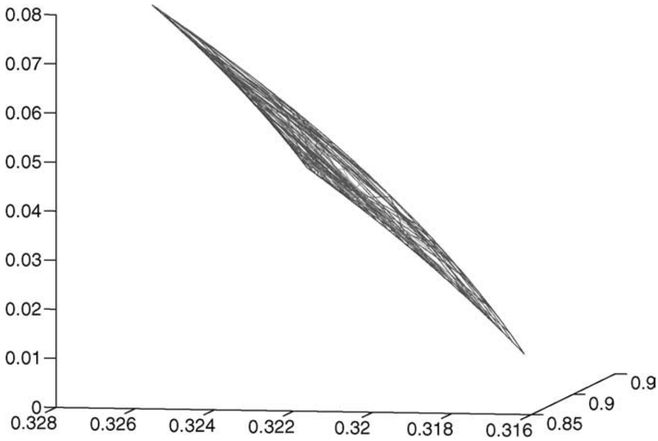

Fig. 3b. Part of conv(A, B, C) where A, B, C in P_2 Gamma-commute.

in terms of the notation of Proposition 7, $\mathscr{T}_1$ (union of the triangle's edges) is shown in blue, part of $\mathscr{T}_2$ is shown in red, and part of $\mathscr{T}_3$ is shown in green.

Remark. In [9] (see Section 3.3) Moakher studies a different situation where, as in Fig. 3b, conv$(\{A, B, C\})$ is a surface. He analyzes $\mathscr{SP}(2)$, the space of *real* pd 2×2 matrices with determinant 1. Since $\mathscr{SP}(2)$ itself has real dimension 2, if A, B, C are chosen from $\mathscr{SP}(2)$ we may expect a picture rather like Fig. 3b, but for different reasons. Moakher notes that $\mathscr{SP}(2)$ is a hyperboloid (of constant negative curvature), but conv$(\{A, B, C\})$ will not admit consistent convex coordinates in the sense of Proposition 9.

4. Completeness and metric projection

In contrast to the matrix norms, the geodesic metric δ makes $\mathbb{P}_n$ into a *complete* metric space.

Proposition 10. *The metric space* $(\mathbb{P}_n, \delta)$ *is complete*.

Proof. Suppose that $\{A_n\}_1^\infty$ is a δ-Cauchy sequence in $\mathbb{P}_n$. By (14) of Proposition 1, $\{\log A_n\}_1^\infty$ is a Cauchy sequence in the Euclidian space $\mathbb{S}_n$ so that it has a limit $L \in \mathbb{S}_n$: $\|\log A_n - L\|_2 \to_n 0$. We can conclude that $\delta(A_n, \exp L) \to 0$ once we are convinced that $\exp : (\mathbb{S}_n, \|\cdot\|_2) \to (\mathbb{P}_n, \delta)$ is continuous. Suppose $L_n \to L$ in $(\mathbb{S}_n, \|\cdot\|_2)$. In view of (17) we have $\delta(\exp L_n, \exp L) = \|\log((\exp L)^{-\frac{1}{2}} \exp L_n (\exp L)^{-\frac{1}{2}})\|_2$, so that to see that $\delta(\exp L_n, \exp L) \to 0$ we need only observe that the eigenvalues of $(\exp L)^{-\frac{1}{2}} \exp L_n (\exp L)^{-\frac{1}{2}}$ tend to 1. This follows by the continuity of exp with respect to any convenient matrix norm (eg the operator norm), combined with spectral continuity. $\square$

Note that $\mathbb{P}_n$ is *not* complete with respect to matrix norms. Instead, it has a boundary consisting of the singular positive *semi*-definite (psd) matrices. In terms of $(\mathbb{P}_n, \delta)$ these "boundary points" are "points at infinity". Proposition 11 shows that they may be approached along appropriate geodesics. It is conventional to extend certain matrix operations from pd to psd matrices by means of the "$+\epsilon I$" device. For example, we may define the geometric mean $S_1 \# S_2$ of psd matrices S_1, S_2 by

$$S_1 \# S_2 = \lim_{\epsilon \downarrow 0} (S_1 + \epsilon I) \# (S_2 + \epsilon I).$$

Proposition 11 makes it less surprising that this operation, while continuous on $\mathbb{P}_n$, is no longer continuous when so extended to the psd matrices.

Proposition 11. *Let S be a singular psd matrix in $\mathbb{S}_n$. Then $S = \lim_{t\to\infty} A\#_t B$ for certain pairs $A, B \in \mathbb{P}_n$. Commuting A, B may be chosen, in fact, and in this case $S = \lim_{t\to\infty} A^{1-t}B^t$. These limits may be computed with respect to any convenient matrix norm. With respect to δ, on the other hand, we have $\lim_{t\to\infty} \delta(A, A\#_t B) = \infty$.*

Remark. We may safely extend our notation $A\#_t B$ from $t \in [0, 1]$ to arbitrary real t using the relation (21).

Proof. Working with an orthonormal basis of eigenvectors for S, we have $S = \operatorname{diag}\{\lambda_k\}$ where $\lambda_k \geqslant 0$ and, for some k, $\lambda_k = 0$. Let $A = \operatorname{diag}\{\alpha_k\}$ and $B = \operatorname{diag}\{\beta_k\}$, where $\alpha_k = \beta_k = \lambda_k$ if $\lambda_k > 0$ and $\alpha_k = 1, \beta_k = 1/2$ if $\lambda_k = 0$. Then it is clear that $S = \lim_{t\to\infty} A^{1-t}B^t$ with respect to any matrix norm, while $\delta(I, A^{1-t}B^t) = \|\log A^{1-t}B^t\|_2 \geqslant |\log 2^{-t}| = t\log 2 \to \infty$ as $t \to \infty$. $\square$

As in a Hilbert space, we can define metric projection onto closed convex subsets of any space, such as $(\mathbb{P}_n, \delta)$, that is complete and satisfies the semi-parallelogram law (19).

Proposition 12. *Let $\mathscr{S}$ be a closed convex set in $(\mathbb{P}_n, \delta)$. For each $A \in \mathbb{P}_n$ there is a unique closest point C to A in $\mathscr{S}$, i.e. $C \in \mathscr{S}$ and for any other $S \in \mathscr{S}$ we have $\delta(A, S) > \delta(A, C)$.*

Proof. Consider a sequence C_n in $\mathscr{S}$ such that $\delta(A, C_n) \to_n \mu$ where

$$\mu = \inf\{\delta(A, S) : S \in \mathscr{S}\}.$$

The semi-parallelogram law (19) implies that

$$\delta^2(C_n, C_m) \leqslant 2(\delta^2(A, C_n) + \delta^2(A, C_m)) - 4\delta^2(A, M),$$

so that

$$\delta^2(C_n, C_m) \leqslant 2(\delta^2(A, C_n) + \delta^2(A, C_m)) - 4\mu^2, \tag{25}$$

since M, the geodesic midpoint of $[C_n, C_m]$, is in $\mathscr{S}$ by the assumed convexity. As $n, m \to \infty$ the RHS of (25) tends to $2(\mu^2 + \mu^2) - 4\mu^2 = 0$, so that $\{C_n\}$ is a Cauchy sequence. By Proposition 10 $\{C_n\}$ has a limit C and continuity of the metric ensures that $\delta(A, C) = \lim \delta(A, C_n) = \mu$. Since $\mathscr{S}$ is closed, $C \in \mathscr{S}$. By the definition of μ, $\delta(A, S) \geqslant \delta(A, C)$ for any $S \in \mathscr{S}$. Finally, if S is any element of $\mathscr{S}$ such that $\delta(A, S) = \mu$ then putting $C_n = C$ and $C_m = S$ in (25) shows that $\delta(C, S) = 0$, i.e. $S = C$. $\square$

The mapping $\pi : \mathbb{P}_n \to \mathscr{S}$ defined by $\pi(A) = C$, where $\mathscr{S}$, A, C are as in Proposition 12, may be called *metric projection* onto $\mathscr{S}$. Metric projection shrinks distances in an appropriate context of nonpositive curvature (see [6, pp. 176–177]). Here we give a proof of a special case needed later, basing our argument directly on the semi-parallelogram law. Thus, the argument given in Proposition 13 for (P_n, δ) applies also to any *Bruhat-Tits* space (compare [6, p. 163]).

Proposition 13. *If π is metric projection onto a closed convex subset $\mathscr{S}$ of $\mathbb{P}_n$, $A \in \mathbb{P}_n$, $C = \pi(A)$, and $D \in \mathscr{S}$, then $\delta^2(A, D) \geqslant \delta^2(C, D) + \delta^2(A, C)$.*

Proof. Let $M_0 = D$ and let M_{n+1} be the geodesic midpoint of $[C, M_n]$. Then $M_n \to_n C = M_\infty$ and in fact $\delta(C, M_n) = 2^{-n}\delta(C, D)$. The semi-parallelogram law (19) implies that

$$2\delta^2(A, M_{n+1}) \leqslant \delta^2(A, M_n) + \delta^2(A, C) - \frac{1}{2}\delta^2(C, M_n),$$

i.e.

$$\delta^2(A, M_n) - \delta^2(A, M_{n+1}) \geqslant \frac{1}{2}\frac{1}{4^n}\delta^2(C, D) + \delta^2(A, M_{n+1}) - \delta^2(A, C).$$

Summing these inequalities,

$$\begin{aligned}&\sum_{n=0}^{\infty}(\delta^2(A, M_n) - \delta^2(A, M_{n+1}))\\ &\quad \geqslant \left(\frac{1}{2}\sum_{n=0}^{\infty}\frac{1}{4^n}\right)\delta^2(C, D) + \sum_{n=0}^{\infty}(\delta^2(A, M_{n+1}) - \delta^2(A, C)).\end{aligned}$$

It is easy to see that these series are absolutely convergent. For example,

$$\begin{aligned}&|\delta^2(A, M_{n+1}) - \delta^2(A, C)|\\ &\quad = (\delta(A, M_{n+1}) + \delta(A, C))|\delta(A, M_{n+1}) - \delta(A, C)|\\ &\quad \leqslant \left(2\delta(A, C) + \frac{1}{2^{n+1}}\delta(C, D)\right)\frac{1}{2^{n+1}}\delta(C, D).\end{aligned}$$

Thus

$$s_0 \geqslant \frac{2}{3}\delta^2(C, D) + \sum_{n=1}^{\infty} s_n,$$

where $s_n = \delta^2(A, M_n) - \delta^2(A, C)$. By the same argument

$$s_n \geqslant \frac{2}{3}\delta^2(C, M_n) + \sum_{k=n+1}^{\infty} s_k,$$

so that

$$\begin{aligned}
\delta^2(A, D) - \delta^2(A, C) &= s_0 \\
&\geqslant \frac{2}{3}\delta^2(C, D) + s_1 + \sum_{k=2}^{\infty} s_k \\
&\geqslant \frac{2}{3}\delta^2(C, D) + \frac{2}{3}\delta^2(C, M_1) + 2\sum_{k=2}^{\infty} s_k \\
&= \frac{2}{3}\left(1 + \frac{1}{4}\right)\delta^2(C, D) + 2\sum_{k=2}^{\infty} s_k \\
&= \frac{2}{3}\left(1 + \frac{1}{4}\right)\delta^2(C, D) + 2s_2 + 2\sum_{k=3}^{\infty} s_k \\
&\geqslant \frac{2}{3}\left(1 + \frac{1}{4}\right)\delta^2(C, D) + 2\left(\frac{2}{3}\delta^2(C, M_2) + \sum_{k=3}^{\infty} s_k\right) + 2\sum_{k=3}^{\infty} s_k \\
&= \frac{2}{3}\left(1 + \frac{1}{4} + \frac{2}{4^2}\right)\delta^2(C, D) + 4\sum_{k=3}^{\infty} s_k,
\end{aligned}$$

etc. Since each $M_k \in \mathscr{S}$ we have $s_k \geqslant 0$. Thus

$$\begin{aligned}
&\delta^2(A, D) - \delta^2(A, C) \\
&\quad \geqslant \frac{2}{3}\left(1 + \frac{1}{4} + \frac{2}{4^2} + \frac{4}{4^3} + \frac{8}{4^4} + \cdots\right)\delta^2(C, D) \\
&\quad = \frac{2}{3}\left(1 + \frac{1}{4}\left(1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots\right)\right)\delta^2(C, D) \\
&\quad = \frac{2}{3}\left(1 + \frac{1}{4}\cdot 2\right)\delta^2(C, D) = \delta^2(C, D). \qquad \square
\end{aligned}$$

5. A comparison of means for $A, B, C \in \mathbb{P}_n$

The authors of [2] establish that $\mathrm{alm}(A, B, C)$, defined as in (3), has many of the properties that might be expected of the "geometric mean" of three pd matrices. In particular, it has properties that we shall describe as *consistency*, *permutation invariance*, and *monotonicity*. The geometric mean $\mathrm{alm}(A, B, C)$ is consistent in the sense that when A, B, C commute the value is as we would expect for simultaneously diagonalizable matrices: $\mathrm{alm}(A, B, C) = (ABC)^{\frac{1}{3}}$. It is permutation invariant in the sense that $\mathrm{alm}(A, B, C)$ is independent of the order in which A, B, C are listed. Finally, it is monotone in the sense that it respects the natural order on $\mathbb{P}_n$, whereby $A \leqslant A'$ means that $A' - A$ is psd. Thus, if $A, A', B, B', C, C' \in \mathbb{P}_n$ and $A \leqslant A'$, $B \leqslant B'$, $C \leqslant C'$, then $\mathrm{alm}(A, B, C) \leqslant \mathrm{alm}(A', B', C')$. We shall see that $\mathrm{ls}(A, B, C)$, defined as in our Introduction, is consistent and permutation invariant and that there are good reasons for believing that it is monotone as well. It is not usually the same as $\mathrm{alm}(A, B, C)$ yet $\mathrm{ls}(A, B, C)$ may be seen as a reasonable alternative interpretation of the geometric mean for three pd matrices.

First we point out that the convergence required in the definition of $\mathrm{alm}(\Delta)$ follows naturally from the geometric features of $\mathbb{P}_n$ discussed in previous sections. Convergence always takes place and the limit $\mathrm{alm}(\Delta)$ lies in the closure of $\mathrm{conv}(\Delta)$. Note that the closure of any convex subset of $\mathbb{P}_n$ is again convex. This follows from the geodesic formula (16), for example. Since it involves sets, rather than points, in $\mathbb{P}_n$, the relation (3) requires some clarification—perhaps in terms of a Hausdorff metric. Instead, we interpret it as in the following proposition, which also makes it clear that $\mathrm{alm}(A, B, C)$ is permutation invariant.

Proposition 14. *Given any $A, B, C \in \mathbb{P}_n$ set $A_0 = A$, $B_0 = B$, $C_0 = C$ and let $A_m = A_{m-1}\#B_{m-1}$, $B_m = B_{m-1}\#C_{m-1}$, and $C_m = C_{m-1}\#A_{m-1}$, for each positive integer m. Then for any choice of $Z_m \in \mathrm{conv}(\{A_m, B_m, C_m\})$ the sequence $\{Z_m\}$ converges to a point $\mathrm{alm}(A, B, C)$ that is independent of the choice of Z_m and lies in the closure of $\mathrm{conv}(\{A, B, C\})$.*

Proof. Let $M_0 = \max\{\delta(A_0, B_0), \delta(B_0, C_0), \delta(C_0, A_0)\}$. It is convenient to note that the diameter $\mathrm{diam}(\mathrm{conv}(\{A_0, B_0, C_0\}))$, i.e. $\max\{\delta(X, Y) : X, Y \in \mathrm{conv}(\{A_0, B_0, C_0\})\}$, is M_0. To see this recall Proposition 7 and observe that if $\mathrm{diam}(\{X_0, X_1, Y_0, Y_1\}) \leqslant M$ then $\delta(X, Y) \leqslant M$ for any $X \in [X_0, X_1]$ and $Y \in [Y_0, Y_1]$. Indeed, let $X = X_0\#_t X_1$ and $Y = Y_0\#_s Y_1$. We may assume that $0 \leqslant s \leqslant t \leqslant 1$. By (20), $d_1 = \delta(X, Y_0\#_t X_1) \leqslant (1-t)\delta(X_0, Y_0) \leqslant (1-t)M$ and $d_3 = \delta(Y_0\#_s X_1, Y) \leqslant s\delta(X_1, Y_1) \leqslant sM$. By Proposition 3, $d_2 = \delta(Y_0\#_t X_1, Y_0\#_s X_1) = (t-s)\delta(Y_0, X_1) \leqslant (t-s)M$. Adding the three inequalities we obtain $d_1 + d_2 + d_3 \leqslant M$; a fortiori, $\delta(X, Y) \leqslant M$.

By Proposition 6, $M_1 = \max\{\delta(A_1, B_1), \delta(B_1, C_1), \delta(C_1, A_1)\} \leqslant \frac{1}{2}M_0$, and it follows as above that $\mathrm{diam}(\mathrm{conv}(\{A_1, B_1, C_1\})) = M_1 \leqslant \frac{1}{2}M_0$. Similarly we obtain $\mathrm{diam}(\mathrm{conv}(\{A_m, B_m, C_m\})) \leqslant 2^{-m}M_0$. Evidently, if $k > m$ we have $Z_k \in$

$\mathrm{conv}(\{A_m, B_m, C_m\})$ so that $\delta(Z_m, Z_k) \leqslant 2^{-m} M_0$. Hence $\{Z_m\}$ is a Cauchy sequence and converges in view of Proposition 10. The uniqueness of the limit follows from the possibility of interlacing such Cauchy sequences and since all the sequences live in $\mathrm{conv}(\{A_0, B_0, C_0\})$, $\mathrm{alm}(A, B, C)$ lies in the closure. □

Remark. Recently Petz and Temesi (see [11, Theorem 2]) have given an elementary proof of the convergence of sequences defining $\mathrm{alm}(\Delta)$.

The following proposition justifies our definition of $\mathrm{ls}(A, B, C)$. We have chosen to base it on the completeness of $(\mathbb{P}_n, \delta)$ (and the semi-parallelogram law) rather than on local compactness.

Proposition 15. *Given any triangle Δ with vertices $A, B, C \in \mathbb{P}_n$, the function* $\mathrm{ss}_\Delta(Y)$ *defined by* (4) *is strictly convex and achieves a unique* (*local and global*) *minimum at the point we denote by* $\mathrm{ls}(\Delta)$.

Proof. By strict convexity of ss_Δ we mean, of course, that for $Y_1 \neq Y_2$ we have $\mathrm{ss}_\Delta(Y_1 \#_t Y_2) < (1-t)\mathrm{ss}_\Delta(Y_1) + t\,\mathrm{ss}_\Delta(Y_2)$ whenever $0 < t < 1$. This is clear because the semi-parallelogram law (19) implies that each term of ss_Δ is strictly (midpoint) convex: e.g.

$$\delta^2(A, Y_1 \# Y_2) \leqslant \frac{\delta^2(A, Y_1) + \delta^2(A, Y_2)}{2} - \frac{1}{4}\delta^2(Y_1, Y_2).$$

Let $m = \inf\{\mathrm{ss}_\Delta(Y) : Y \in \mathbb{P}_n\}$, and consider any $Y_k \in \mathbb{P}_n$ such that $\mathrm{ss}_\Delta(Y_k) \to m$ as $k \to \infty$. Using the semi-parallelogram law again we see that

$$\begin{aligned}\frac{3}{4}\delta^2(Y_k, Y_j) &\leqslant \frac{1}{2}(\mathrm{ss}_\Delta(Y_k) + \mathrm{ss}_\Delta(Y_j)) - \mathrm{ss}_\Delta(Y_k \# Y_j)\\ &\leqslant \frac{1}{2}(\mathrm{ss}_\Delta(Y_k) + \mathrm{ss}_\Delta(Y_j)) - m.\end{aligned}$$

Thus $\delta(Y_k, Y_j) \to 0$ as $k, j \to \infty$ and the Cauchy sequence converges via Proposition 10 to some limit Z. The function ss_Δ is clearly continuous, so that $\mathrm{ss}_\Delta(Z) = m$. By strict convexity Z $(= \mathrm{ls}(\Delta))$ is the *unique* minimal point for ss_Δ. □

Note that $\mathrm{ls}(A^{-1}, B^{-1}, C^{-1}) = (\mathrm{ls}(A, B, C))^{-1}$, since $\delta(X^{-1}, Y^{-1}) = \delta(X, Y)$; the mean $\mathrm{alm}(A, B, C)$ also respects matrix inversion, inheriting this property from the corresponding property for two variables.

The following lemma will be useful in identifying the "gradient" for ss_Δ. Given functions $f, g : (0, \infty) \to \mathbb{R}$, $X \in \mathbb{P}_n$ and $Y \in \mathbb{S}_n$, we write $[D(f, g, X)](Y)$ for the limit

$$\lim_{t \to 0} \frac{\langle f(X+tY), g(X+tY)\rangle_F - \langle f(X), g(X)\rangle_F}{t}, \tag{26}$$

provided this limit exists. Here $\langle A, B\rangle_F$ denotes the Frobenius inner product for matrices A, B, i.e. $\langle A, B\rangle_F = \sum_{i,j} a_{ij}\overline{b_{ij}} = \mathrm{trace}(B^*A)$.

Lemma. *If $f, g : (0, \infty) \to \mathbb{R}$ are continuously differentiable then*

$$[D(f, g, X)](Y) = \langle f'(X)g(X) + f(X)g'(X), Y\rangle_F,$$

for any $X \in \mathbb{P}_n$ and $Y \in \mathbb{S}_n$.

Proof. Writing the difference quotient in (26) as

$$\frac{\langle f(X+tY), g(X+tY)\rangle_F - \langle f(X), g(X+tY)\rangle_F}{t} + \frac{\langle f(X), g(X+tY)\rangle_F - \langle f(X), g(X)\rangle_F}{t},$$

we see that $[D(f, g, X)](Y) = \langle [Df(X)](Y), g(X)\rangle_F + \langle f(X), [Dg(X)](Y)\rangle_F$, where $Df(X)$ and $Dg(X)$ are Fréchet derivatives, as discussed in Section 2. Working with respect to an orthonormal basis of eigenvectors for X we have, as in (12),

$$\langle [Df(X)](Y), g(X)\rangle_F = \left\langle \left[\frac{f(\lambda_i) - f(\lambda_j)}{\lambda_i - \lambda_j}\right] \circ Y, g(X)\right\rangle_F,$$

where $X = \mathrm{diag}\{\lambda_k\}$. Since $g(X) = \mathrm{diag}\{g(\lambda_k)\}$,

$$\langle [Df(X)](Y), g(X)\rangle_F = \langle \mathrm{diag}\{f'(\lambda_k)\} \circ Y, g(X)\rangle_F = \langle f'(X)Y, g(X)\rangle_F.$$

By the commutativity of the trace and the fact that $f'(X)$ and $g(X)$ commute, we also have $\langle [Df(X)](Y), g(X)\rangle_F = \langle f'(X)g(X), Y\rangle_F$. Similarly, $\langle f(X), [Dg(X)](Y)\rangle_F = \langle f(X)g'(X), Y\rangle_F$. $\square$

Condition (28) of the following proposition provides a useful criterion for $Z = \mathrm{ls}(\Delta)$. We first learned about it from [9], but it perhaps dates back to much earlier work by Élie Cartan (see [3, Section 6.1.5]).

Proposition 16. *Given $A, B, C, Z \in \mathbb{P}_n$ and $\Delta = \{A, B, C\}$, the matrix*

$$G(A, B, C, Z) = 2\Big(Z^{-\frac{1}{2}} \log\big(Z^{\frac{1}{2}}A^{-1}Z^{\frac{1}{2}}\big)Z^{-\frac{1}{2}} + Z^{-\frac{1}{2}} \log\big(Z^{\frac{1}{2}}B^{-1}Z^{\frac{1}{2}}\big)Z^{-\frac{1}{2}} + Z^{-\frac{1}{2}} \log\big(Z^{\frac{1}{2}}C^{-1}Z^{\frac{1}{2}}\big)Z^{-\frac{1}{2}}\Big)$$

is the gradient of $\mathrm{ss}_\Delta(Z)$ in the sense that, for any $Y \in \mathbb{S}_n$,

$$\lim_{t\to 0} \frac{\mathrm{ss}_\Delta(Z+tY) - \mathrm{ss}_\Delta(Z)}{t} = \langle G(A, B, C, Z), Y\rangle_F. \tag{27}$$

Hence, if $Z = \mathrm{ls}(\Delta)$, $G(A, B, C, Z) = 0$; equivalently

$$\log(A^{-1}Z) + \log(B^{-1}Z) + \log(C^{-1}Z) = 0. \tag{28}$$

Proof. For convenience, set $A_1 = A$, $A_2 = B$, and $A_3 = C$. We have (in view of Proposition 3)

$$\begin{aligned}\mathrm{ss}_\Delta(Z) &= \sum_{k=1}^{3} \left\| \log\left(A_k^{-\frac{1}{2}} Z A_k^{-\frac{1}{2}} \right) \right\|_2^2 \\ &= \sum_{k=1}^{3} \left\langle \log\left(A_k^{-\frac{1}{2}} Z A_k^{-\frac{1}{2}} \right), \log\left(A_k^{-\frac{1}{2}} Z A_k^{-\frac{1}{2}} \right) \right\rangle_F .\end{aligned}$$

Using the lemma above we see that

$$\begin{aligned}&\lim_{t\to 0} \frac{\mathrm{ss}_\Delta(Z+tY) - \mathrm{ss}_\Delta(Z)}{t} \\ &\quad = \sum_{k=1}^{3} \left[D\left(\log, \log, A_k^{-\frac{1}{2}} Z A_k^{-\frac{1}{2}} \right) \right] \left(A_k^{-\frac{1}{2}} Y A_k^{-\frac{1}{2}} \right) \\ &\quad = \sum_{k=1}^{3} 2 \left\langle \log\left(A_k^{-\frac{1}{2}} Z A_k^{-\frac{1}{2}} \right) \left(A_k^{-\frac{1}{2}} Z A_k^{-\frac{1}{2}} \right)^{-1}, A_k^{-\frac{1}{2}} Y A_k^{-\frac{1}{2}} \right\rangle_F \\ &\quad = \sum_{k=1}^{3} 2 \cdot \operatorname{trace}\left(A_k^{-\frac{1}{2}} Y A_k^{-\frac{1}{2}} \log\left(A_k^{-\frac{1}{2}} Z A_k^{-\frac{1}{2}} \right) A_k^{\frac{1}{2}} Z^{-1} A_k^{\frac{1}{2}} \right).\end{aligned}$$

Using commutativity of the trace, we may rewrite this as

$$\begin{aligned}&\sum_{k=1}^{3} 2 \cdot \operatorname{trace}\left(Y A_k^{-\frac{1}{2}} \log\left(A_k^{-\frac{1}{2}} Z A_k^{-\frac{1}{2}} \right) A_k^{\frac{1}{2}} Z^{-1} \right) \\ &\quad = \left\langle 2 \sum_{k=1}^{3} A_k^{-\frac{1}{2}} \log\left(A_k^{-\frac{1}{2}} Z A_k^{-\frac{1}{2}} \right) A_k^{\frac{1}{2}} Z^{-1}, Y \right\rangle_F .\end{aligned}$$

The matrix functional calculus allows us to extend the logarithm function to matrices with eigenvalues in $\mathbb{C} \setminus (-\infty, 0]$ in such a way that similarity is respected, i.e.

$$\log(SXS^{-1}) = S \log(X) S^{-1}. \tag{29}$$

Thus we may write

$$\lim_{t\to 0} \frac{\mathrm{ss}_\Delta(Z+tY) - \mathrm{ss}_\Delta(Z)}{t} = \left\langle 2 \sum_{k=1}^{3} \log(A_k^{-1} Z) Z^{-1}, Y \right\rangle_F ,$$

so that

$$G(A, B, C, Z) = 2 \sum_{k=1}^{3} \log(A_k^{-1} Z) Z^{-1}. \tag{30}$$

To obtain the form of the gradient set out in the proposition, note that another application of (29) shows that

$$\begin{aligned}\log(A_k^{-1}Z)Z^{-1} &= Z^{-\frac{1}{2}}Z^{\frac{1}{2}}\log(A_k^{-1}Z)Z^{-\frac{1}{2}}Z^{-\frac{1}{2}}\\ &= Z^{-\frac{1}{2}}\log\left(Z^{\frac{1}{2}}A_k^{-1}Z^{\frac{1}{2}}\right)Z^{-\frac{1}{2}}.\end{aligned}$$

Finally, (30) shows that the vanishing of $G(A, B, C, Z)$ is equivalent to (28). □

Numerical experiments reveal that, in general, $\mathrm{alm}(A, B, C) \neq \mathrm{ls}(A, B, C)$. Indeed, $\mathrm{alm}(A, B, C)$ may differ from any of the *weighted* least square points $\mathrm{ls}(A, B, C, w_A, w_B, w_C)$ discussed below. Given weights $w_A, w_B, w_C \in [0, 1]$ with $w_A + w_B + w_C = 1$, we define $\mathrm{ss}_\Delta(Y, w_A, w_B, w_C)$ as

$$w_A\delta^2(A, Y) + w_B\delta^2(B, Y) + w_C\delta^2(C, Y),$$

and $Z = \mathrm{ls}(A, B, C, w_A, w_B, w_C)$ is the point such that $\mathrm{ss}_\Delta(Z, w_A, w_B, w_C) \leqslant \mathrm{ss}_\Delta(Y, w_A, w_B, w_C)$ for all $Y \in \mathbb{P}_n$. It is easy (extending the argument of Proposition 15) to establish the existence of $\mathrm{ls}(A, B, C, w_A, w_B, w_C)$ and its uniqueness. Note that $\mathrm{ls}(A, B, C) = \mathrm{ls}(A, B, C, 1/3, 1/3, 1/3)$.

Proposition 17. *For any $A, B, C \in \mathbb{P}_n$ and weights $w_A, w_B, w_C \in [0, 1]$ with $w_A + w_B + w_C = 1$, the weighted least squares point* $\mathrm{ls}(A, B, C, w_A, w_B, w_C)$ *lies in the closure of* $\mathrm{conv}(\{A, B, C\})$. *In particular*, $\mathrm{ls}(A, B, C)$ *lies in the closure of* $\mathrm{conv}(\{A, B, C\})$.

Proof. Let $\mathscr{S}$ denote the closure of $\mathrm{conv}(\{A, B, C\})$ and let π denote metric projection onto $\mathscr{S}$. Since $A \in \mathscr{S}$, Proposition 13 ensures that $\delta(A, Y) \geqslant \delta(A, \pi(Y))$ for any $Y \in \mathbb{P}_n$. The same argument applies to B and C, so that $\mathrm{ss}_\Delta(Y, w_A, w_B, w_C) \geqslant \mathrm{ss}_\Delta(\pi(Y), w_A, w_B, w_C)$. Thus $\mathrm{ls}(A, B, C, w_A, w_B, w_C)$ cannot lie outside $\mathscr{S}$. □

The set of all weighted least squares points may be viewed as a surface spanning the triangle $[A, B] \cup [B, C] \cup [C, A]$ and lying within the closure of $\mathrm{conv}(\{A, B, C\})$. We may call this the Cartan spanning surface, since the idea appears to go back to Élie Cartan (see [3, Section 6.1.5]). Using the gradient formula of Proposition 16 (and its analogues for $\mathrm{ss}_\Delta(Z, w_A, w_B, w_C)$) we may design effective computer algorithms for approximating the various points $\mathrm{ls}(A, B, C, w_A, w_B, w_C)$ on the Cartan spanning surface. For example, we may modify the fixed point method of Moakher (discussed below) to compute the Cartan surface. Thus Fig. 4 depicts the Cartan surface with a colouring based on the weights w_A, w_B, w_C: at the vertices, where w_A, w_B, or w_C equals 1, we see pure colours (red, green, or blue) whereas $\mathrm{ls}(A, B, C)$ is located where the colours blend evenly ($w_A = w_B = w_C = 1/3$). Fig. 4 also shows, as a small black circle, the location of $\mathrm{alm}(A, B, C)$—lying slightly off the Cartan surface (but close to $\mathrm{ls}(A, B, C)$) in this example.

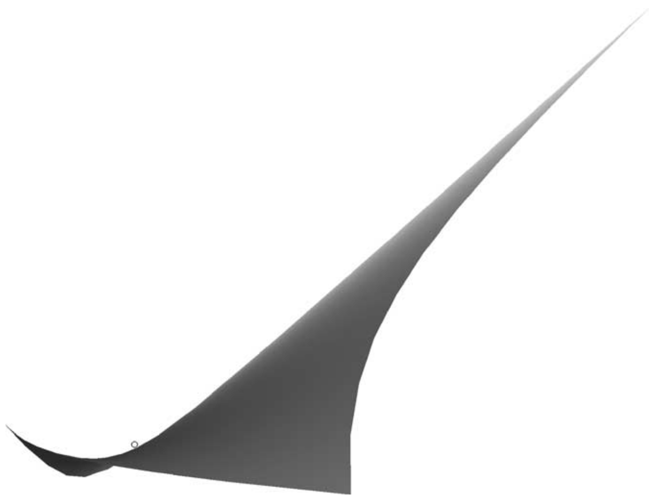

Fig. 4. The Cartan surface vs $\mathrm{alm}(A, B, C)$.

The case of Γ-commuting A, B, C presents a simpler picture, as the following proposition makes clear.

Proposition 18. *If A, B, C are Γ-commuting then the Cartan spanning surface coincides with* $\mathrm{conv}(\{A, B, C\})$ *and* $\mathrm{ls}(A, B, C) = \mathrm{alm}(A, B, C)$. *If $X \in GL_n$ is such that $\Gamma(A), \Gamma(B), \Gamma(C)$ commute, with $\Gamma = \Gamma_X$, then*

$$\mathrm{ls}(A, B, C, w_A, w_B, w_C) = \Gamma^{-1}(\Gamma(A)^{w_A}\Gamma(B)^{w_B}\Gamma(C)^{w_C}). \tag{31}$$

In particular, if A, B, C commute then $\mathrm{alm}(A, B, C) = \mathrm{ls}(A, B, C) = (ABC)^{\frac{1}{3}}$.

Proof. By Proposition 17 each point $\mathrm{ls}(A, B, C, w_A, w_C, w_C)$ lies in the closure of $\mathrm{conv}(\{A, B, C\})$ and, by Proposition 9, $\mathrm{conv}(\{A, B, C\})$ is isometric to the affine triangle $\mathrm{conv}(\{\log \Gamma(A), \log \Gamma(B), \log \Gamma(C)\})$. In such a Euclidian setting we know that, for any $w_k \in [0, 1]$ with $\sum_k w_k = 1$, and vectors x, x_k,

$$\operatorname{arcmin}\left(\sum_k w_k \|x - x_k\|^2\right) = \sum_k w_k x_k,$$

so that (31) follows. The construction converging to $\mathrm{alm}(A, B, C)$ (see Proposition 14) corresponds in $\mathrm{conv}(\{\log \Gamma(A), \log \Gamma(B), \log \Gamma(C)\})$ to the successive selection of affine triangles with vertices obtained at each stage as midpoints of the sides of the previous stage. These triangles converge to $(\log \Gamma(A) + \log \Gamma(B) + \log \Gamma(C))/3$ so that $\mathrm{alm}(A, B, C) = \Gamma^{-1}(\exp((\log \Gamma(A) + \log \Gamma(B) + \log \Gamma(C))/3)) = \Gamma^{-1}(\Gamma(A)^{1/3}\Gamma(B)^{1/3}\Gamma(C)^{1/3}) = \mathrm{ls}(A, B, C)$. $\square$

Note that one consequence of Proposition 18 is that both $\mathrm{alm}(\Delta)$ and $\mathrm{ls}(\Delta)$ are "consistent" in the sense of the opening paragraph of this section. Our final task is to examine evidence for our conjecture that $\mathrm{ls}(\Delta)$ is also "monotone".

In view of the invariance of δ under the congruence $\Gamma_{A^{-\frac{1}{2}}}$ it is sufficient to study $Z = \mathrm{ls}(\Delta)$ when $\Delta = (I, B, C)$. Proposition 16 then tells us that Z is the unique solution to

$$\log(Z) + \log\left(Z^{\frac{1}{2}} B^{-1} Z^{\frac{1}{2}}\right) + \log\left(Z^{\frac{1}{2}} C^{-1} Z^{\frac{1}{2}}\right) = 0.$$

Moakher has observed (see [10, Section 2.4.2]) that this means $S = \log(Z)$ is the unique fixed point of any F_α with $\alpha \in (0, 1)$, where

$$F_\alpha(X) = \alpha X + (\alpha - 1)\left(\log\left(\mathrm{e}^{X/2} B^{-1} \mathrm{e}^{X/2}\right) + \log\left(\mathrm{e}^{X/2} C^{-1} \mathrm{e}^{X/2}\right)\right).$$

Moreover, it appears that, for an appropriate choice of α, F_α acts as if it were a contraction mapping so that the iterates $F_\alpha^m(X)$ converge to $S = \log(\mathrm{ls}(\Delta))$ as $m \to \infty$. This technique allows the reliable computation of $\mathrm{ls}(\Delta)$ in most cases and computer experiments appear to support the conjecture that $\mathrm{ls}(\Delta)$ is monotone. For example, millions of "random" choices of $\Delta = (I, B, C)$ and $\Delta' = (I, B', C)$, with $B' \geqslant B$ have resulted in $Z = \mathrm{ls}(\Delta)$ and $Z' = \mathrm{ls}(\Delta')$ with $Z' \geqslant Z$ in every case.

Note that (because log is matrix monotone) the conjecture $Z' \geqslant Z$ implies also that $S' \geqslant S$ where S' is the fixed point of

$$F'_\alpha(X) = \alpha X + (\alpha - 1)\left(\log\left(\mathrm{e}^{X/2} (B')^{-1} \mathrm{e}^{X/2}\right) + \log\left(\mathrm{e}^{X/2} C^{-1} \mathrm{e}^{X/2}\right)\right).$$

Computer experiments suggest that F_α is monotone at S, i.e. that $T \geqslant S$ implies $F_\alpha(T) \geqslant F_\alpha(S) = S$. If this could be established it would strongly support the (weaker?) conjecture that $S' \geqslant S$. To see this note first that $F'_\alpha(X) \geqslant F_\alpha(X)$ for all X (because $(B')^{-1} \leqslant B^{-1}$ and log is matrix monotone). Thus we would have $T \geqslant S$ implying $F'_\alpha(T) \geqslant F_\alpha(T) \geqslant S$ and, inductively, $(F'_\alpha)^m(T) \geqslant S$ for all m. Since we can probably choose some $T \geqslant S$ such that $(F'_\alpha)^m(T) \to_m S'$ ($T = S$ would be a reasonable choice if B' is close to B), it would follow that $S' \geqslant S$.

We are perhaps rather far from a proof of the conjecture, but we state it formally as worthy of further study.

Conjecture. *The mean* $\mathrm{ls}(\Delta)$ *is monotone with respect to its arguments, i.e. if* $A, A', B, B', C, C' \in \mathbb{P}_n$ *and* $A \leqslant A'$, $B \leqslant B'$, $C \leqslant C'$, *then* $\mathrm{ls}(A, B, C) \leqslant \mathrm{ls}(A', B', C')$.

Further remarks. It is sufficient to show that $Z \leqslant Z'$ where $Z = \mathrm{ls}(A, B, C)$ and $Z' = \mathrm{ls}(A', B, C)$. In view of Proposition 16 the gradients $G(A, B, C, Z)$ and $G(A', B, C, Z')$ vanish and, since $1/x$ is matrix decreasing while log is matrix monotone (see [4, Chapter V, especially pp. 114 and 135]), $G(A', B, C, Z) \leqslant G(A, B, C, Z) = 0$, i.e. $-G(A', B, C, Z) \geqslant 0$. Since we have $\mathrm{ss}_\Delta(A', B, C, Z') < \mathrm{ss}_\Delta(A', B, C, Z)$ (provided $Z' \neq Z$), and $-G(A', B, C, Z) \geqslant 0$ is the direction of steepest descent for ss_Δ, we might expect that $Z' \geqslant Z$.

More precisely, let X be psd and let $L(t)$ denote $\mathrm{ls}(A + tX, B, C)$. We wish to show that $L'(0) \geqslant 0$. Consider the difference quotient

$$DQ(t) = \frac{G(A + tX, B, C, L(t)) - G(A + tX, B, C, L(0))}{t}.$$

Since $G(A + tX, B, C, L(t)) = 0$ (Proposition 16) and $-G(A + tX, B, C, L(0)) \geqslant 0$ (when $t > 0$) by the last paragraph, this difference quotient is psd. On the other hand, subject to appropriate smoothness conditions,

$$\lim_{t\to 0} DQ(t) = \lim_{t\to 0} \frac{G(A, B, C, L(t)) - G(A, B, C, L(0))}{t},$$

and, by the chain rule, this is $[D_4G(A, B, C, L(0))](L'(0))$, where D_4G denotes the Fréchet derivative of G with respect to its fourth variable. Thus $[D_4G(A, B, C, L(0))]$ $(L'(0)) \geqslant 0$ and we may try to evaluate D_4G with a view to concluding that $L'(0) \geqslant 0$. Note that we may write

$$G(A, B, C, Z) = 2\Big(A^{-\frac{1}{2}} \log\big(A^{-\frac{1}{2}} Z A^{-\frac{1}{2}}\big)\big(A^{-\frac{1}{2}} Z A^{-\frac{1}{2}}\big)^{-1} A^{-\frac{1}{2}} + \cdots\Big),$$

so that it may be pertinent to analyze the Fréchet derivative $D\log(Y)Y^{-1}$.

Acnowledgements

We are indebted to Roy Mathias and Maher Moakher for helpful discussions of many of the points treated above. We thank Jean–Pierre Schoch for designing and performing many of the relevant computer experiments. The second author is grateful to NSERC of Canada, the University of Guelph sabbatical program, and the hospitality of the Indian Statistical Institute (Delhi Centre) for their support of this project. Some of the results in this paper were presented by the first author at the ILAS conference held at Coimbra in July 2004.

References

[1] T. Ando, Concavity of certain maps on positive definite matrices and applications to Hadamard products, Linear Algebra Appl. 26 (1979) 203–241.
[2] T. Ando, C.-K. Li, R. Mathias, Geometric means, Linear Algebra Appl. 385 (2004) 305–334.
[3] M. Berger, A Panoramic View of Riemannian Geometry, Springer, 2003.
[4] R. Bhatia, Matrix Analysis, Springer, 1997.
[5] R. Bhatia, On the exponential metric increasing property, Linear Algebra Appl. 375 (2003) 211–220.
[6] M. Bridson, A. Haefliger, Metric Spaces of Non-Positive Curvature, Springer, 1999.
[7] G. Corach, H. Porta, L. Recht, Geodesics and operator means in the space of positive operators, Int. J. Math. 4 (1993) 193–202.
[8] J.D. Lawson, Y. Lim, The geometric mean, matrices, metrics, and more, Amer. Math. Monthly 108 (2001) 797–812.
[9] M. Moakher, A differential geometric approach to the geometric mean of symmetric positive-definite matrices, SIAM J. Matrix Anal. Appl. 26 (2005) 735–747.

[10] M. Moakher, On the averaging of symmetric positive-definite tensors, 2004, preprint.
[11] D. Petz, R. Temesi, Means of positive numbers and matrices, 2005, preprint.
[12] W. Pusz, S.L. Woronowicz, Functional calculus for sesquilinear forms and the purification map, Reports Math. Phys. 8 (1975) 159–170.
[13] G.E. Trapp, Hermitian semidefinite matrix means and related matrix inequalities—an introduction, Linear and Multilinear Algebra 16 (1984) 113–123.