1-1-2013

The least squares mean of positive Hilbert-Schmidt operators

Jimmie Lawson
Louisiana State University

Yongdo Lim
Sungkyunkwan University

J. Math. Anal. Appl. 403 (2013) 365–375

Contents lists available at SciVerse ScienceDirect

Journal of Mathematical Analysis and Applications

journal homepage: www.elsevier.com/locate/jmaa

The least squares mean of positive Hilbert–Schmidt operators

Jimmie Lawson [a,*], Yongdo Lim [b]

[a] Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803, USA

[b] Department of Mathematics, Sungkyunkwan University, Suwon 440-746, Republic of Korea

ARTICLE INFO

Article history:
Received 1 November 2012
Available online 13 February 2013
Submitted by Gustavo Corach

Keywords:
Positive Hilbert–Schmidt operators
Least squares mean
Karcher equation
Riemannian manifold
Nonpositive curvature

ABSTRACT

We show that, the least squares mean on the Riemannian manifold Σ of positive operators in the extended Hilbert–Schmidt algebra of linear operators on a Hilbert space equipped with the canonical trace metric is the unique solution of the corresponding Karcher equation. This allows us to conclude that, the least squares mean is the restriction of the Karcher mean on the open cone of all bounded positive definite operators, and hence inherits the basic properties of that mean. Conversely, the Karcher mean on the positive definite operators is shown to be the unique monotonically strongly continuous extension of the least squares mean on Σ.

1. Introduction

Positive definite matrices have arisen recently in a dazzling variety of applications: image processing, medical imaging, radar systems, learning machines, brain–computer interfacing, subdivision schemes, to cite a few. The Riemannian trace metric on the open convex cone $\mathbb{P} = \mathbb{P}_m$ of $m \times m$ positive definite matrices plays an important role in these applied areas involving matrix interpolation, filtering, estimation, optimization and averaging, where it has been increasingly recognized that the Euclidean distance is often not the most suitable for the set of positive definite matrices and that working with the appropriate metric geometry does matter in computational problems.

A natural and attractive candidate of means and averaging for positive definite matrices is the least squares mean (Riemannian geometric mean, Cartan centroid, Karcher mean) from the Riemannian trace metric. (Recall the trace metric distance between two positive definite matrices is given by $\delta(A, B) = \left(\sum_{i=1}^{k} \log^2 \lambda_i(A^{-1}B)\right)^{\frac{1}{2}}$, where $\lambda_i(X)$ denotes the i-th eigenvalue of X in non-decreasing order.) First M. Moakher [17] and then Bhatia and Holbrook [4] suggested the least squares mean as the unique minimizer of the sum of the squares of the distances:

$$\Lambda(A_1, \ldots, A_n) = \underset{X \in \mathbb{P}}{\arg\min} \sum_{i=1}^{n} \delta^2(X, A_i). \tag{1.1}$$

This idea had been anticipated by Élie Cartan (see, for example, Section 6.1.5 of [1]), who showed among other things that such a unique minimizer exists if the points all lie in a convex ball in a Riemannian manifold, which is enough to deduce the existence of the least-squares mean globally for $\mathbb{P}$. A more detailed study of Riemannian centers of mass in the setting of Riemannian manifolds was carried out by H. Karcher [10]. Using the Karcher's formula for the gradient of the objective

* Corresponding author.
E-mail addresses: lawson@math.lsu.edu (J. Lawson), ylim@skku.edu (Y. Lim).

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2013.02.013

function (Theorem 2.1 of [10]) or computing appropriate derivatives as in [2,17] yields that the Karcher mean coincides with the unique positive definite solution of the *Karcher equation*

$$\sum_{i=1}^{n} \log(X^{1/2}A_i^{-1}X^{1/2}) = 0. \tag{1.2}$$

A currently active research topic in linear algebra is understanding, finding properties of, and computing efficiently, the least squares mean. The monotonicity of the Karcher mean, one of the key axiomatic properties of matrix geometric means, conjectured by Bhatia and Holbrook [4], was recently established by Lawson and Lim [14] via a probabilistic convergence of approximations and by Bhatia and Karandikar [5] via some probabilistic counting arguments, both arguments depending heavily on basic inequalities for the Riemannian metric. Interesting deterministic approaches to the least squares mean have recently been discovered independently by Holbrook via a "no dice" approach [8] and by Lim and Pálfia [16] via an arithmetic power mean approach. We recommend a nice survey paper by Bhatia [3] for more on the least squares mean of positive definite matrices. A key ingredient in the derivation of many of these properties, the monotonicity in particular, is the fact that the trace metric on the positive definite matrices gives them the structure of a Cartan–Hadamard–Riemannian manifold, in particular a manifold of nonpositive curvature. This implies that, the metric is a complete metric that satisfies the *semiparallelogram law*: for each X, Y, there exists a (necessarily unique) P satisfying

$$\delta^2(P, Z) \le \frac{1}{2}\delta^2(X, Z) + \frac{1}{2}\delta^2(Y, Z) - \frac{1}{4}\delta^2(X, Y) \tag{1.3}$$

for all $Z \in \mathbb{P}$. Such complete metrics are sometimes called Bruhat–Tits metrics and the resulting metric spaces are called metric spaces of nonpositive curvature, NPC-spaces for short, or CAT_0-spaces, and there is now an extensive literature treating such spaces (see, for example, [6,19], and, in particular, [11, Chapter 11] for a basic treatment as they arise in the context of Cartan–Hadamard–Riemannian manifolds).

The main purpose of this paper is to extend these results from positive definite matrices to the infinite dimensional setting, specifically to the set Σ of positive operators in the extended Hilbert–Schmidt algebra $\mathrm{HS}(H) + \mathbb{R}I$ on a (separable) Hilbert space H, where

$$\Sigma := \{A + \lambda I > 0 : A^* = A,\ A \in \mathrm{HS}(H),\ \lambda \in \mathbb{R}\}.$$

As has recently been worked out by G. Larotonda [12,13], Σ is an infinite dimensional Cartan–Hadamard–Riemannian manifold equipped with a canonical trace metric, where scalar operators are orthogonal to the Hilbert–Schmidt operators $\mathrm{HS}(H)$. The structure of the Riemannian manifold Σ in many ways closely parallels that of finite dimensional Riemannian manifolds of positive definite matrices equipped with the Riemannian trace metric.

Our main result, an extension from the finite dimensional setting, is that, the weighted least squares mean for a weighting $\omega = (w_1, \ldots, w_n)$ on Σ is the unique solution of the Karcher equation

$$\sum_{i=1}^{n} w_i \log(X^{1/2}A_i^{-1}X^{1/2}) = 0.$$

In Section 2 we recall some of Larotonda's basic results and elaborate on them in what follows. We clarify in Section 3 the relationship between the least-squares mean on the infinite dimensional Cartan–Hadamard–Riemannian manifold Σ and the critical point of the objective function in both the Euclidean and the Riemannian sense. We establish in Section 4 that, in the context of the extended Hilbert–Schmidt operators, the solution of the Karcher equation may be interpreted as the solution of the gradient of the least-squares distance function being 0, hence the equation's close connection to the least squares mean.

In Section 5 we establish basic properties of the least squares mean on Σ, which follow almost immediately from properties of the Karcher mean established in [15], since we show that, the least squares mean on Σ is the restriction of the Karcher mean on $\mathbb{P}$, the cone of all (bounded) positive definite operators on the Hilbert space H. Conversely, we show that, the Karcher mean on the positive definite operators is the unique monotonically strongly continuous extension of the least squares mean on Σ. Thus the Karcher mean on the open cone $\mathbb{P}$ of positive definite operators is in two senses the "ghost of a departed quantity" (to use Bishop Berkeley's famous expression). Firstly it satisfies the Karcher equation, which in the finite dimensional or extended Hilbert–Schmidt setting, means the gradient of the function given in (1.1) vanishes, and secondly it is the (strongly directed) limiting case of the least squares mean on the extended Hilbert–Schmidt algebra.

2. Hilbert–Schmidt operators

Let H be a complex Hilbert space. We assume in all that follows that H is separable, but see Remark 5.6. For a complex Hilbert space H let $\mathcal{B}(H)$ be the Banach space of bounded linear operators on H equipped with the operator norm. We denote the adjoint of $A \in \mathcal{B}(H)$ by A^* and call A Hermitian if $A = A^*$. Let $\mathcal{S}(H)$ denote the closed subspace of bounded Hermitian linear operators, and let $\mathbb{P} = \mathbb{P}(H) \subseteq \mathcal{S}(H)$ be the open convex cone of positive definite operators. The Banach Lie group $\mathrm{GL}(H)$ of bounded invertible linear operators (with operation composition) acts on $\mathbb{P}$ via congruence transformations:

$\Gamma_C(X) = CXC^*$. For $X, Y \in \mathcal{S}(H)$, we write $X \leq Y$ if $Y - X$ is positive semidefinite, and $X < Y$ if $Y - X$ is positive definite. Note that, $X \leq Y$ if and only if $\langle x, Xx\rangle \leq \langle x, Yx\rangle$ for all $x \in H$.

Let $\mathrm{HS}(H)$ denote the bilateral ideal of Hilbert–Schmidt operators of $\mathcal{B}(H)$. Recall that, $\mathrm{HS}(H)$ is a Banach algebra (without unit) when given the norm $\|A\|_2 = \mathrm{Tr}(AA^*)^{1/2}$. In $\mathcal{B}(H)$ we define

$$\mathcal{H}_{\mathbb{C}} = \{A + \lambda I : A \in \mathrm{HS}(H), \lambda \in \mathbb{C}\},$$

a complex linear subalgebra that we call the *extended Hilbert–Schmidt algebra*. There is a natural Hilbert space structure for this subspace (where scalar operators are orthogonal to Hilbert–Schmidt operators) which is given by the inner product

$$\langle A + \lambda I, B + \mu I\rangle_2 = \mathrm{Tr} AB^* + \lambda\overline{\beta},$$

where $\mathrm{Tr}AB^*$ denotes the trace of AB^*. The algebra $\mathcal{H}_{\mathbb{C}}$ is complete with respect to the corresponding Hilbert norm. Our focus is the symmetric or real part of $\mathcal{H}_{\mathbb{C}}$,

$$\mathcal{H}_{\mathbb{R}} = \{A + \lambda I : A^* = A,\ A \in \mathrm{HS}(H),\ \lambda \in \mathbb{R}\},$$

which inherits the structure of a (real) Banach space, and with the same inner product becomes a real Hilbert space, and on its positive part $\Sigma = \mathbb{P} \cap \mathcal{H}_{\mathbb{R}}$, an open subcone. We denote $\mathrm{HS}(H) \cap \mathcal{H}_{\mathbb{R}}$ by $\mathrm{HS}_{\mathbb{R}}$.

We define a Riemannian metric on Σ as follows. We identify $T\Sigma$ with $\Sigma \times \mathcal{H}_{\mathbb{R}}$, and endow the tangent space at A with the Hilbert metric

$$\langle X, Y\rangle_A = \langle A^{-1}X, YA^{-1}\rangle_2.$$

The structure of the Riemannian manifold Σ closely parallels that of finite dimensional Riemannian manifolds of positive definite matrices equipped with the Riemannian trace metric, as has been worked out by Larotonda [12,13].

Lemma 2.1. *The inner product $\langle\cdot, \cdot, \rangle_2$ satisfies the following basic properties:*

(i) $\langle XY, Y^*X^*\rangle_2 = \langle YX, X^*Y^*\rangle_2$ *for any* $X, Y \in \mathcal{H}_{\mathbb{C}}$;
(ii) $\langle ZX, YZ\rangle_2 = \langle XZ, ZY\rangle_2$ *for* $X, Y \in \mathcal{H}_{\mathbb{C}}$ *and* $Z \in \mathcal{H}_{\mathbb{R}}$;
(iii) $\langle AU, V\rangle_2 = \langle U, A^*V\rangle_2$ *for* $A \in \mathcal{B}(H)$ *and* $U, V \in \mathrm{HS}(H)$;
(iv) $\|X\|_A = \|A^{-1/2}XA^{-1/2}\|_2$ *for* $X \in \mathcal{H}_{\mathbb{R}}$ *and* $A \in \Sigma$.

Proof. Items (i) and (ii) follow from direct computation and the commutativity properties of the trace. For (iii) we recall that, for any orthonormal basis $\{e_i : i \in I\}$ of H, the inner product on $\mathrm{HS}(H)$ is given by $\langle U, V\rangle = \mathrm{Tr}(UV) = \sum_{i\in I}\langle Ue_i, Ve_i\rangle$. For $A \in \mathcal{B}(H)$, AU and A^*V are again in $\mathrm{HS}(H)$ since the latter is an ideal in $\mathcal{B}(H)$. We thus have

$$\langle AU, V\rangle_2 = \sum_{i\in I}\langle AUe_i, Ve_i\rangle = \sum_{i\in I}\langle Ue_i, A^*Ve_i\rangle = \langle U, A^*V\rangle_2.$$

For (iv), we first observe from item (ii) that

$$\begin{aligned}\|X\|_A^2 &= \langle X, X\rangle_A = \langle A^{-1}X, XA^{-1}\rangle_2\\ &= \langle A^{-1/2}A^{-1/2}X, XA^{-1/2}A^{-1/2}\rangle_2\\ &= \langle A^{-1/2}XA^{-1/2}, A^{-1/2}XA^{-1/2}\rangle_2 = \|A^{-1/2}XA^{-1/2}\|_2^2.\end{aligned}$$

Taking square roots gives the desired result. □

We recall some of Larotonda's basic results (see [12, Section 3]) and elaborate on them in what follows.

Theorem 2.2 (*[12]*). *The Riemannian manifold Σ is an NPC-space (i.e., a complete metric space satisfying the semiparallelogram law (1.3)) with respect to the Riemannian distance function δ. Moreover, for distinct $A, B \in \Sigma$:*

(i) *the curve* $t \mapsto A\#_tB = A^{1/2}(A^{-1/2}BA^{-1/2})^tA^{1/2}$ *is (up to parametrization) the unique geodesic line passing through A and B*;
(ii) *the exponential and logarithm maps at A are diffeomorphisms given by*

$$\mathrm{Exp}_A(X) = A^{1/2}e^{A^{-1/2}XA^{-1/2}}A^{1/2}, \qquad \mathrm{Log}_A(P) = A^{1/2}\log(A^{-1/2}PA^{-1/2})A^{1/2},$$

where X and P vary over $\mathcal{H}_{\mathbb{R}}$ and Σ, respectively, e^X is the usual exponential map restricted to $\mathcal{H}_{\mathbb{R}}$, and $\log : \Sigma \to \mathcal{H}_{\mathbb{R}}$ *is its inverse;*
(iii) *the Riemannian distance between A and B is*

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

Remark 2.3. Larotonda does not show directly that the metric space is an NPC-space, but this follows readily by standard results from his Theorem 3.5 that the Riemannian exponential function has at all points an expansive differential; see e.g., [11, Chapter 11]. The function $\mathrm{Exp}_A : T_A\Sigma \to \Sigma$ having expansive differential means $\|D\,\mathrm{Exp}_A(U)(W)\|_{\mathrm{Exp}_A(U)} \geq \|W\|_A$.

It should be noted that, the preceding results are generalizations of similar results for the set of positive definite matrices equipped with the Riemannian trace metric, as was worked out by Mostow much earlier [18, Section 2]. In this case, the distance trace metric is given by $\delta(A, B) = \left[\sum_{i=1}^{m} \log^2 \lambda_i(A^{-1}B)\right]^{\frac{1}{2}}$, where the $\lambda_i(X)$ denote the eigenvalues of X. In this fashion the set of positive definite matrices is endowed with the structure of a Cartan–Hadamard–Riemannian manifold, in particular a manifold of nonpositive curvature. This whole theory has been extended to the setting of Finsler geometry by Conde and Larotonda [7] for the case of a manifold with each of the tangent spaces equipped with a p-uniformly convex norm.

The factorization of $\mathcal{H}_{\mathbb{R}}$ as the direct sum of the Hilbert–Schmidt ideal $\mathrm{HS}_{\mathbb{R}}$ and $\mathbb{R}I$ induces a corresponding factorization of Σ.

Lemma 2.4. *If $A \in \mathrm{HS}_{\mathbb{R}}$ and $I + A \in \Sigma$, then $I + A = e^B$ for some (unique) $B \in \mathrm{HS}_{\mathbb{R}}$. Conversely for $B \in \mathrm{HS}_{\mathbb{R}}$, $e^B = I + A$ for some unique $A \in \mathrm{HS}_{\mathbb{R}}$. Thus every $X \in \Sigma$ has a unique factorization in the form $X = e^{\lambda}e^B = e^{\lambda}(I + A)$, where $\lambda \in \mathbb{R}$ and $A, B \in \mathrm{HS}_{\mathbb{R}}$.*

Proof. As a special case of Theorem 2.2(ii) the map $\exp(X) = e^X$ is a diffeomorphism from $\mathcal{H}_{\mathbb{R}}$ to Σ with inverse log. Thus $I + A = \exp(\lambda I + B)$ for some unique $\lambda I + B \in \mathcal{H}_{\mathbb{R}}$ if $I + A \in \Sigma$. It follows that

$$I + A = \exp(\lambda I + B) = e^{\lambda}e^B = e^{\lambda}(I + (e^B - I)).$$

Note that, $e^B - I = \sum_{n=1}^{\infty} B^n/n! \in \mathrm{HS}_{\mathbb{R}}$. Thus $I + A = e^{\lambda}I + e^{\lambda}(e^B - I)$, hence $I = e^{\lambda}I$, so that $\lambda = 0$. Also $A = e^{\lambda}(e^B - I) = e^B - I$ implies $I + A = e^B$.

Conversely let $B \in \mathrm{HS}_{\mathbb{R}}$. Then $e^B = I + \sum_{n=1}^{\infty} B^n/n!$. The second term $A = \sum_{n=1}^{\infty} B^n/n! \in \mathrm{HS}_{\mathbb{R}}$, since the latter is an ideal, in particular a subalgebra, of $\mathcal{H}_{\mathbb{R}}$. That A is unique follows from the fact $\mathcal{H}_{\mathbb{R}}$ is the direct sum of the Hilbert–Schmidt ideal $\mathrm{HS}_{\mathbb{R}}$ and $\mathbb{R}I$.

For the last assertion, the fact that exp is a diffeomorphism yields $X = \exp(\lambda I + B) = e^{\lambda}e^B$ for some unique $\lambda I + B \in \mathcal{H}_{\mathbb{R}}$, where $\lambda \in \mathbb{R}$ and $B \in \mathrm{HS}_{\mathbb{R}}$. The alternative form $e^B = I + A$ follows from the previous part of the proof. $\square$

Let us henceforth denote $(I + \mathrm{HS}_{\mathbb{R}}) \cap \Sigma = \exp(\mathrm{HS}_{\mathbb{R}})$ by Σ_I.

Corollary 2.5. *The map $\exp : \mathrm{HS}_{\mathbb{R}} \to \Sigma_I$ is a diffeomorphism with inverse diffeomorphism $\log : \Sigma_I \to \mathrm{HS}_{\mathbb{R}}$, where $\Sigma_I = \exp(\mathrm{HS}_{\mathbb{R}})$.*

Corollary 2.6. *The distance squared between $A = e^{\lambda}e^C$ and $B = e^{\mu}e^D$, where $C, D \in \mathrm{HS}_{\mathbb{R}}$, is given by $\delta^2(A, B) = (\mu - \lambda)^2 + \mathrm{Tr}(\log^2(e^{C/2}e^{-D}e^{C/2}))$, where $e^{C/2}e^{-D}e^{C/2} \in \Sigma_I$.*

Proof. We first note that, $e^{C/2}e^{-D}e^{C/2}$ is positive definite, hence in Σ. From Lemma 2.4 we may write $e^{C/2} = I + E$ and $e^{-D} = I + F$, where $E, F \in \mathrm{HS}_{\mathbb{R}}$. Then $e^{C/2}e^{-D}e^{C/2} = (I + E)(I + F)(I + E) = I + G$ for some $G \in \mathrm{HS}_{\mathbb{R}}$, since $\mathrm{HS}_{\mathbb{R}}$ is an ideal in $\mathcal{H}_{\mathbb{R}}$. Thus $e^{C/2}e^{-D}e^{C/2} \in \Sigma_I$ and hence $\log(e^{C/2}e^{-D}e^{C/2}) \in \mathrm{HS}_{\mathbb{R}}$.

We compute from Lemma 2.4 and Theorem 2.2(iii) that

$$\begin{aligned}\delta^2(A, B) &= \langle \log(e^{\lambda/2}e^{C/2}e^{-\mu}e^{-D}e^{\lambda/2}e^{C/2}), \log(e^{\lambda/2}e^{C/2}e^{-\mu}e^{-D}e^{\lambda/2}e^{C/2})\rangle \\ &= \langle (\lambda - \mu)I + \log(e^{C/2}e^{-D}e^{C/2}), (\lambda - \mu)I + \log(e^{C/2}e^{-D}e^{C/2})\rangle \\ &= (\lambda - \mu)^2 + \mathrm{Tr}(\log^2(e^{C/2}e^{-D}e^{C/2})). \quad \square\end{aligned}$$

The following properties are straightforward generalizations from the finite-dimensional setting.

Lemma 2.7. *Let $A, B \in \Sigma$. Then*

(i) $M(A\#_tB)M^* = (MAM^*)\#_t(MBM^*)$ *for invertible* $M \in \mathcal{H}_{\mathbb{C}}$;
(ii) $A\#_tB = B\#_{1-t}A$, $(A\#_tB)^{-1} = A^{-1}\#_tB^{-1}$;
(iii) $(A\#_tB)\#_s(A\#_uB) = A\#_{(1-s)t+su}B$ *for any* $s, t, u \in \mathbb{R}$.

Lemma 2.8. *For $A, B \in \Sigma$ and invertible $M \in \mathcal{H}_{\mathbb{C}}$,*

(i) $\delta(A, B) = \delta(A^{-1}, B^{-1}) = \delta(MAM^*, MBM^*)$;
(ii) $\delta(A\#B, A) = \delta(A\#B, B) = \frac{1}{2}\delta(A, B)$;
(iii) $\delta(A\#_tB, A\#_sB) = |s - t|\delta(A, B)$ *for all* $t, s \in [0, 1]$; *and*
(iv) $\delta(A^t, B^t) \leq t\delta(A, B)$ *for all* $t \in [0, 1]$.

Proof. Invariance under inversion follows from

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

Invariance under congruence transformations follows directly from

$$\|MUM^*\|_{MXM^*} = \|U\|_X$$

for invertible $M \in \mathcal{H}_{\mathbb{C}}$; see [12, Lemma 2.5].

Items (ii) and (iii) follow directly from Theorem 2.2(i). The last item is a special case of [12, Corollary 3.10]. □

Remark 2.9. Using polarization and part (i) of the preceding yields

$$\langle MUM^*, MVM^*\rangle_{MXM^*} = \langle U, V\rangle_X. \tag{2.4}$$

The following results hold for any NPC-space [19].

Lemma 2.10. *For all $t \in [0, 1]$ and $A, B, C, D \in \Sigma$,*

$$\delta^2(A\#_t B, C) \le (1-t)\delta^2(A, C) + t\delta^2(B, C) - (1-t)t\delta^2(A, B), \tag{2.5}$$

$$\delta(A\#_t B, C\#_t D) \le (1-t)\delta(A, C) + t\delta(B, D). \tag{2.6}$$

3. The least squares mean on Σ

Let $\omega = (w_1, \ldots, w_n) \in \Delta_n$, the simplex of positive probability vectors in $\mathbb{R}^n$, the convex hull of the set of unit coordinate vectors, and let $A_1, \ldots, A_n \in \Sigma$. The ω-least squares mean of $A_1, \ldots, A_n$ for the metric δ on Σ is defined by

$$\Lambda(\omega; A_1, \ldots, A_n) = \underset{X\in\Sigma}{\arg\min} \sum_{i=1}^{n} w_i \delta^2(X, A_i). \tag{3.7}$$

Since (Σ, δ) is an NPC-space, the least-squares mean exists and is unique [19].

As first order of business, we show that, finding the least squares mean on Σ reduces to finding it for the special cases of $\mathbb{R}I$ and $\Sigma_I = \exp(\mathrm{HS}_{\mathbb{R}})$.

Definition 3.1. A nonempty subset P of Σ is convex if for any $A, B \in P$, we have $A\#_t B \in P$ for $0 \le t \le 1$ (see Theorem 2.2(i)).

Lemma 3.2. *The set Σ_I is a closed convex subset of Σ which contains the ω-least squares mean of any $A_1, \ldots, A_n \in \Sigma_I$.*

Proof. The fact that Σ_I is closed under taking powers $A \mapsto A^r$ for $r \in \mathbb{R}$ follows directly from $\Sigma_I = \exp(\mathrm{HS}_{\mathbb{R}})$. The proof of Corollary 2.6 shows that $e^{C/2}e^D e^{C/2} \in \Sigma_I$ whenever $C, D \in \mathrm{HS}_{\mathbb{R}}$. The convexity of Σ_I then follows from Theorem 2.2(i). That Σ_I is closed in Σ follows from the fact $\mathrm{HS}_{\mathbb{R}}$ is closed in $\mathcal{H}_{\mathbb{R}}$ and $\Sigma_I = \Sigma \cap (I + \mathrm{HS}_R)$.

That any closed convex set contains the least squares mean of any finite set of its elements has been shown by Bhatia and Holbrook [4]. Indeed the result is true for general NPC-spaces; see [9, Lemma 3.3.4]. □

Lemma 3.3. *For $i = 1, \ldots, n$, let $A_i = \lambda_i B_i$, where $\lambda_i > 0$ and $B_i \in \Sigma_I$ (Lemma 2.4). The ω-least squares mean of $A_1, \ldots, A_n$ is given by $\left(\prod_{i=1}^n \lambda_i^{w_i}\right) \Lambda(\omega; B_1, \ldots, B_n)$.*

Proof. For any $X = \lambda Y \in \Sigma$, where $Y \in \Sigma_I$, by Corollary 2.6

$$\sum_{i=1}^{n} w_i \delta^2(X, A_i) = \sum_{i=1}^{n} w_i (\log\lambda - \log\lambda_i)^2 + \sum_{i=1}^{n} w_i \mathrm{Tr}(\log^2(B_i^{1/2} Y^{-1} B_i^{1/2})). \tag{3.8}$$

Thus the ω-least squares mean is obtained by choosing λ so the first summation is minimized and choosing Y so that the second summation is satisfied. Differentiating the first summation with respect to λ, setting the derivative equal to 0, and solving for λ yields $\log(\lambda) = \sum_{i=1}^n w_i \log(\lambda_i)$. By definition and Corollary 2.6 the minimizer of the second summation is $\Lambda(\omega; B_1, \ldots, B_n)$. □

We observe that, the least squares mean is the unique minimizer of the function $f : \Sigma \to \mathbb{R}$ defined by $f(X) = (1/2)\sum_{i=1}^n w_i \delta^2(X, A_i)$. We next seek to clarify the relationship between the least squares mean and the critical point of the objective function f in both the Euclidean and the Riemannian sense.

Let $g : \Sigma \to \mathbb{R}$ be a differentiable function. The (Euclidean) gradient of g is defined at $U \in \mathcal{H}_{\mathbb{R}}$ by

$$Dg(X)(U) = \langle \nabla g(X), U\rangle_2.$$

The Riemannian gradient of g at X is defined by

$$\langle \nabla^{\mathrm{Rie}} g(X), U\rangle_X = (g \circ \gamma)'(0)$$

for all geodesics γ satisfying $\gamma(0) = X$ and $\dot{\gamma}(0) = U$. We note that, for such a geodesic γ,

$$\langle \nabla g(X), U\rangle_2 = Dg(X)(U) = (g \circ \gamma)'(0) = \langle \nabla^{\mathrm{Rie}} g(X), U\rangle_X. \tag{3.9}$$

In particular,

$$\nabla g(X) = 0 \Longleftrightarrow \nabla^{\mathrm{Rie}} g(X) = 0. \tag{3.10}$$

Remark 3.4. In the finite dimensional setting,

$$\nabla g(X) = X^{-1}\left[\nabla^{\mathrm{Rie}} g(X)\right] X^{-1}. \tag{3.11}$$

Indeed,

$$\langle \nabla^{\mathrm{Rie}} g(X), U\rangle_X = \langle X^{-1}U, [\nabla^{\mathrm{Rie}} g(X)]X^{-1}\rangle_2 = \langle U, X^{-1}[\nabla^{\mathrm{Rie}} g(X)]X^{-1}\rangle_2$$

where the second equality follows immediately from the cyclicity of the trace functional. However, this formula fails in our current setting, since the inner product is no longer a true trace inner product, but also contains a summand from scalar operators λI.

In the remainder of this section we work exclusively in the Euclidean inner product setting and thus for convenience drop the subscript 2 for the trace inner product. We seek an explicit expression for the gradient $\nabla f(X)$ for X restricted to Σ_I. We consider the inverse functions $\exp : \mathrm{HS}_{\mathbb{R}} \to \Sigma_I$ and $\log : \Sigma_I \to \mathrm{HS}_{\mathbb{R}}$. Let $f_I : \Sigma_I \to \mathbb{R}$ be defined by

$$f_I(X) = \frac{1}{2}\delta^2(X, I) = \frac{1}{2}\|\log X\|_2^2 = \frac{1}{2}\langle \log X, \log X\rangle$$

and note that

$$Df_I(X)(U) = \langle D\log(X)(U), \log X\rangle.$$

If the operator $D\log(X)$ acting on $\mathrm{HS}_{\mathbb{R}}$ is a symmetric operator w.r.t. the trace inner product, then

$$Df_I(X)(U) = \langle D\log(X)(\log X), U\rangle = \langle X^{-1}\log X, U\rangle, \text{ since}$$
$$\lim_{t\to 0^+} \frac{\log(X + t\log X) - \log X}{t} = \lim_{t\to 0^+} \frac{\log(I + t(\log X)X^{-1})}{t} = X^{-1}\log X,$$

which yields

$$\nabla f_I(X) = X^{-1}\log X = X^{-1/2}(\log X)X^{-1/2}. \tag{3.12}$$

The symmetry of $D\log(\cdot)$ is equivalent to that of $D\exp(\cdot)$. We show that, $D\exp(A)$ is symmetric for each $A \in \mathrm{HS}_{\mathbb{R}}$. For $U \in \mathrm{HS}_{\mathbb{R}}$, set $S_k(A, U) = \sum_{j=0}^{k-1} A^jUA^{k-j-1} = \sum_{j=0}^{k-1} A^{k-j-1}UA^j$. From differentiation of the power series formula for $\exp(\cdot)$, we obtain

$$D\exp(A)(U) = \lim_{t\to 0} \frac{\exp(A + tU) - \exp(A)}{t} = \sum_{k=1}^{\infty} \frac{S_k(A, U)}{k!}.$$

For any $V \in \mathrm{HS}_{\mathbb{R}}$, we then have

$$\begin{aligned}\langle D\exp(A)(U), V\rangle &= \left\langle \sum_{k=1}^{\infty} \frac{S_k(A, U)}{k!}, V\right\rangle = \sum_{k=1}^{\infty} \frac{1}{k!}\langle S_k(A, U), V\rangle \\ &= \sum_{k=1}^{\infty} \frac{1}{k!}\langle U, S_k(A, V)\rangle \\ &= \langle U, D\exp(A)(V)\rangle\end{aligned}$$

where the third equality follows from the fact that

$$\langle A^{k-j-1}UA^j, V\rangle = \mathrm{Tr}(A^{k-j-1}UA^jV) = \mathrm{Tr}(UA^jVA^{k-j-1}) = \langle U, A^jVA^{k-j-1}\rangle.$$

This completes the proof of (3.12).

Remark 3.5. (i) In the finite dimensional setting, Mostow [18] has given a similar computation to ours, for establishing the symmetry property of $D\log(X)$. Alternatively, it follows from the Dalecki–Krein formula in terms of Hadamard product [2, p. 60].

(ii) The technique of the previous proof may also be used to show that the derivative of the map $X \mapsto X^n$ on $\mathrm{HS}_{\mathbb{R}}$ is a symmetric operator. From this it follows that, if $f : \mathrm{HS}_{\mathbb{R}} \to \mathrm{HS}_{\mathbb{R}}$ is an analytic function, then Df is a symmetric operator. The same is true for $g = f + I : \mathrm{HS}_{\mathbb{R}} \to \Sigma_I$.

Next, we find a formula for the gradient of our least squares objective function.

Proposition 3.6. *Let $A \in \Sigma_I$. Define $f_A : \Sigma_I \to \mathbb{R}$ by $f_A(X) = (1/2)\delta^2(X, A)$. Then*

$$\nabla f_A(X) = X^{-1/2} \log(X^{1/2}A^{-1}X^{1/2})X^{-1/2}.$$

More generally, for $A_1, \dots, A_n \in \Sigma_I$,

$$\nabla f(X) = X^{-1/2} \left[\sum_{i=1}^{n} w_i \log(X^{1/2}A_i^{-1}X^{1/2})\right] X^{-1/2}, \tag{3.13}$$

where $f : \Sigma_I \to \mathbb{R}, f(X) = (1/2)\sum_{i=1}^{n} w_i\delta^2(X, A_i)$. Furthermore, $\nabla f(X) = 0$ if and only if X is a solution of the following equation

$$\sum_{i=1}^{n} w_i \log(X^{1/2}A_i^{-1}X^{1/2}) = 0. \tag{3.14}$$

Proof. We note $f_A(X) = (1/2)\delta^2(X, A) = (1/2)\delta^2(A^{-1/2}XA^{-1/2}, I) = (f_I \circ \Gamma_{A^{-1/2}})(X)$, where $\Gamma_{A^{-1/2}}(X) = A^{-1/2}XA^{-1/2}$. Let $X \in \Sigma_I, U \in \mathrm{HS}_{\mathbb{R}}$. Set $B = A^{-1/2}XA^{-1/2} \in \Sigma$. Then

$$\begin{aligned}
\langle \nabla f_A(X), U\rangle &= D(f_I \circ \Gamma_{A^{-1/2}})(X)(U) = Df_I(A^{-1/2}XA^{-1/2})(A^{-1/2}UA^{-1/2}) \\
&= \langle \nabla f_I(B), A^{-1/2}UA^{-1/2}\rangle \overset{(3.12)}{=} \langle B^{-1/2}(\log B)B^{-1/2}, A^{-1/2}UA^{-1/2}\rangle \\
&= \langle A^{-1/2}A^{1/2}B^{-1/2}(\log B)B^{-1/2}, A^{-1/2}UA^{-1/2}\rangle \\
&= \langle A^{1/2}B^{-1/2}(\log B)B^{-1/2}A^{-1/2}, A^{-1}U\rangle \quad (\text{Lemma 2.1(ii)}) \\
&= \langle A^{-1/2}B^{-1/2}(\log B)B^{-1/2}A^{-1/2}, U\rangle. \quad (\text{Lemma 2.1(iii)})
\end{aligned}$$

As a result it suffices to show that

$$X^{-1/2} \log(X^{1/2}A^{-1}X^{1/2})X^{-1/2} = A^{-1/2}B^{-1/2}(\log B)B^{-1/2}A^{-1/2}.$$

Put $Z = X^{1/2}A^{-1/2}(A^{1/2}X^{-1}A^{1/2})^{1/2} = X^{1/2}A^{-1/2}B^{-1/2}$. One can directly see that $Z^* = Z^{-1}$ and $ZBZ^* = X^{1/2}A^{-1}X^{1/2}$. Then

$$Z(\log B)Z^* = \log ZBZ^* = \log(X^{1/2}A^{-1}X^{1/2}),$$

which establishes the asserted formula for ∇f_A. The remaining assertions of the proposition now follow directly. □

Let $\mathbb{P} = \mathbb{P}(H) \subseteq \mathcal{S}(H)$ be the open convex cone of positive definite operators on the Hilbert space H.

Definition 3.7. Let $\mathbb{A} = (A_1, \dots, A_n) \in \mathbb{P}^n, \omega = (w_1, \dots, w_n) \in \Delta_n$. The following nonlinear operator equation on $\mathbb{P}$ is called the *Karcher equation*:

$$\sum_{i=1}^{n} w_i \log(X^{1/2}A_i^{-1}X^{1/2}) = 0. \tag{3.15}$$

We have the following restatement of the last part of Proposition 3.6.

Corollary 3.8. *On the metric manifold (Σ_I, δ) the least squares mean $\Lambda(\omega; A_1, \dots, A_n)$ satisfies the corresponding Karcher equation.*

In the finite-dimensional setting it is known that, the least-squares mean is the unique positive solution of the Karcher equation. We extend this result to the infinite-dimensional setting in the next section.

4. Convex functions on Σ

A map $g : \Sigma \to \mathbb{R}$ is called uniformly convex if there is a strictly increasing function $\phi : [0, \infty) \to [0, \infty)$ such that

$$g(A\#B) \le \frac{1}{2}(g(A) + g(B)) - \phi(\delta(A, B))$$

for all $A, B \in \mathbb{P}$. For a continuous uniformly convex function g, there is a unique minimizer of $g(X)$ (see [19]):

$$\arg\min_{X \in \mathbb{P}} g(X).$$

Now the objective function $f : \Sigma \to \mathbb{R}, f(X) = (1/2)\sum_{i=1}^{n} w_i \delta^2(X, A_i)$ is uniformly convex in the Riemannian sense. Indeed Lemma 2.10 ensures that

$$\begin{aligned} f(X\#Y) &= \frac{1}{2}\sum_{i=1}^{n} w_i \delta^2(X\#Y, A_i) \\ &\le \frac{1}{2}\sum_{i=1}^{n} w_i \left[\frac{1}{2}(\delta^2(X, A_i) + \delta^2(Y, A_i)) - \frac{1}{4}\delta^2(X, Y)\right] \\ &= \frac{1}{2}(f(X) + f(Y)) - \frac{1}{4}\delta^2(X, Y). \end{aligned}$$

Definition 4.1. A map $g : \Sigma \to \mathbb{R}$ is said to be convex if $g(X\#_t Y) \le (1-t)g(X) + g(Y)$ for all $X, Y \in \Sigma$ and $t \in [0, 1]$ and strictly convex if $g(X\#_t Y) < (1-t)g(X) + tg(Y)$ for all $X \neq Y \in \Sigma$ and $t \in (0, 1)$.

Any strictly convex function on a NPC space has a unique minimizer, if the minimizer exists (cf. [19]).

Proposition 4.2. *Let $g : \Sigma \to \mathbb{R}$ be strictly convex (uniformly convex) and differentiable. For $X \in \Sigma$, the following are equivalent;*

(i) *X is the minimizer of g;*
(ii) *for all $Y \in \Sigma$,*

$$0 = \lim_{t\to 0} \frac{g(X\#_t Y) - g(X)}{t}; \tag{4.16}$$

(iii) *$(g \circ \gamma)'(0) = 0$ for all Riemannian geodesics γ satisfying $\gamma(0) = X$;*
(iv) *$\nabla g(X) = 0$;*
(v) *$\nabla^{\mathrm{Rie}} g(X) = 0$.*

Proof. The equivalence between (iv) and (v) follows from (3.10). We note that, any geodesic γ satisfying $\gamma(0) = X$ is of the form $g(t) = X\#_t Y$ for some $Y \in \Sigma$, which proves the equivalence of (ii) and (iii).

(i) implies (ii): Suppose that, X is the global (or even local) minimizer of g. Then one can easily see that

$$\lim_{t\to 0} \frac{g(X\#_t Y) - g(X)}{t} = 0$$

for all $Y \in \Sigma$.

(iii) if and only if (iv): Let $U \in \mathcal{H}_{\mathbb{R}}$. Choose $Y > 0$ such that $\gamma'(0) = U$ for $\gamma(t) = X\#_t Y$. The following equalities establish the equivalence of (iii) and (iv):

$$\langle \nabla g(X), U\rangle_2 = Dg(X)(U) = Dg(\gamma(0))(\gamma'(0)) = (g \circ \gamma)'(0).$$

(iii) implies (i): For $X \neq Y \in \Sigma$, define $\gamma(t) = X\#_t Y$. Then $g \circ \gamma$ is a smooth strictly convex function from $\mathbb{R} \to \mathbb{R}$, which by hypothesis has derivative 0 at 0. It follows that $g \circ \gamma$ must have its unique minimum at 0, and hence $g(X) = g(\gamma(0)) < g(\gamma(1)) = g(Y)$. Since Y was arbitrary, the unique global minimizer occurs at X. □

Now the objective function $f : \Sigma \to \mathbb{R}, f(X) = (1/2)\sum_{i=1}^{n} w_i \delta^2(X, A_i)$ is strictly convex in the Riemannian sense. Indeed let $X, Y \in \Sigma$ and let $t \in (0, 1)$. Then Lemma 2.10 ensures that

$$\begin{aligned} f(X\#_t Y) &= (1/2)\sum_{i=1}^{n} w_i \delta^2(X\#_t Y, A_i) < (1/2)\sum_{i=1}^{n} w_i \left[(1-t)\delta^2(X, A_i) + t\delta^2(Y, A_i)\right] \\ &= (1-t)f(X) + tf(Y). \end{aligned}$$

Combining this together with Lemma 3.2 and Propositions 3.6 and 4.2 yields the following.

Theorem 4.3. *The least-squares mean $\Lambda(\omega; A_1, \ldots, A_n)$ is uniquely determined and is the unique solution of the Karcher equation (3.15) on Σ.*

Proof. We write $A_i = \lambda_i B_i$, where $\lambda_i > 0$ and $B_i \in \Sigma_I$ for each i. By Lemma 3.3 and its proof, particularly Eq. (3.8), the least squares mean $\Lambda(\omega; A_1, \ldots, A_n)$ is given by λY, where $\lambda = \lambda_1^{w_1} \cdots \lambda_n^{w_n}$ and $X = Y$ minimizes $\sum_{i=1}^{n} w_i \mathrm{Tr}(\log^2(X^{1/2} B_i^{-1} X^{1/2}))$. We note that $x = \lambda$ is the unique solution to

$$0 = \log(x(\lambda_1^{w_1} \cdots \lambda_n^{w_n})^{-1}) = \sum_{i=1}^{n} w_i \log(x^{1/2}(\lambda_i^{-1})x^{1/2}) \tag{4.17}$$

and by Propositions 3.6 and 4.2, $X = Y$ is the unique solution to

$$\sum_{i=1}^{n} w_i \log(X^{1/2}B_i^{-1}X^{1/2}) = 0. \tag{4.18}$$

Adding the two previous equations together yields that $xX = \lambda Y$ satisfies

$$\begin{aligned} 0 &= \sum_{i=1}^{n} w_i \log(x^{1/2}(\lambda_i^{-1})x^{1/2})I + \sum_{i=1}^{n} w_i \log(X^{1/2}B_i^{-1}X^{1/2}) \\ &= \sum_{i=1}^{n} w_i \log((xX)^{1/2}A_i^{-1}(xX)^{1/2}). \end{aligned}$$

Since $\mathcal{H}_{\mathbb{R}}$ is the direct sum of $\mathbb{R}I$ and $\mathrm{HS}_{\mathbb{R}}$, it follows that the preceding sum is equal to 0 iff each summand is, and thus the solution is unique as a result of the uniqueness of the solution in each of the summands. □

Remark 4.4. Once one identifies Σ as a Riemannian manifold of nonpositive curvature, it is possible to appeal directly to a general theorem of H. Karcher [10, Theorem 1.2] that gives the gradient of the least-squares distance function in a general form of the left hand side of the Karcher equation and establishes that this gradient vanishes precisely at the minimum value. However, we are of the opinion that, the Riemannian geometry of the positive Hilbert–Schmidt operators is an important new tool for the study of these operators, and that, it is thus of value to put on record concrete calculations for the gradient, exponential map, etc. in this setting. Additionally, our proof is quite elementary in the sense that we employ only basic operator theory in our proof, together with a minimal knowledge of Riemannian geometry.

5. Properties of the least squares means

In a recent paper [15] the authors have studied the cone $\mathbb{P}$ of positive define operators contained in the algebra $\mathcal{B}(H)$ of bounded linear operators on an arbitrary Hilbert space H and shown that its Karcher equation

$$\sum_{i=1}^{n} w_i \log(X^{1/2}A_i^{-1}X^{1/2}) = 0 \tag{5.19}$$

always admits a unique solution, called the ω-weighted Karcher mean of $A_1, \ldots, A_n$, and denoted $\tilde{\Lambda}(\omega; A_1, \ldots, A_n)$. The extended Hilbert–Schmidt subalgebra sits inside of $\mathcal{B}(H)$ and Σ sits inside $\mathbb{P}$, and as we have seen in Theorem 4.3, the least squares mean on Σ satisfies the Karcher equation. Hence the Karcher mean on $\mathbb{P}$ restricted to Σ is precisely the least squares mean of Σ. Via this identification standard properties of the Karcher mean established in [15] carry over to the least squares mean on Σ.

Theorem 5.1. *Let $\omega \in \Delta_n$ and $\mathbb{A} = (A_1, \ldots, A_n) \in \Sigma^n$. Then the least squares mean $\Lambda(\omega; A_1, \ldots, A_n)$ is the unique solution in Σ, and more generally in $\mathbb{P}$, of the Karcher equation*

$$0 = \sum_{i=1}^{n} \omega_i \log(X^{1/2}A_i^{-1}X^{1/2}).$$

Furthermore, $\Lambda : \Delta_n \times \Sigma^n \to \Sigma$ satisfies the following properties:

(P1) *(Consistency with scalars) $\Lambda(\omega; \mathbb{A}) = A_1^{w_1} \cdots A_n^{w_n}$ if the A_i's commute;*
(P2) *(Joint homogeneity) $\Lambda(\omega; a_1A_1, \ldots, a_nA_n) = a_1^{w_1} \cdots a_n^{w_n} \Lambda(\omega; \mathbb{A})$;*
(P3) *(Permutation invariance) $\Lambda(\omega_\sigma; \mathbb{A}_\sigma) = \Lambda(\omega; \mathbb{A})$, where $\omega_\sigma = (w_{\sigma(1)}, \ldots, w_{\sigma(n)})$;*
(P4) *(Monotonicity) If $B_i \leq A_i$ for all $1 \leq i \leq n$, then $\Lambda(\omega; \mathbb{B}) \leq \Lambda(\omega; \mathbb{A})$;*
(P5) *(Continuity) $\delta(\Lambda(\omega; \mathbb{A}), \Lambda(\omega; \mathbb{B})) \leq \sum_{i=1}^{n} w_i\delta(A_i, B_i)$ for the Riemannian trace metric δ; in particular $\Lambda(\omega, \cdot)$ is continuous;*
(P6) *(Congruence invariance) $\Lambda(\omega; M^*\mathbb{A}M) = M^*\Lambda(\omega; \mathbb{A})M$ for any invertible M;*
(P7) *(Joint concavity) $\Lambda(\omega; \lambda\mathbb{A} + (1-\lambda)\mathbb{B}) \geq \lambda\Lambda(\omega; \mathbb{A}) + (1-\lambda)\Lambda(\omega; \mathbb{B})$ for $0 \leq \lambda \leq 1$;*
(P8) *(Self-duality) $\Lambda(\omega; A_1^{-1}, \ldots, A_n^{-1})^{-1} = \Lambda(\omega; A_1, \ldots, A_n)$; and*
(P9) *(AGH weighted mean inequalities) $\left(\sum_{i=1}^{n} w_iA_i^{-1}\right)^{-1} \leq \Lambda(\omega; \mathbb{A}) \leq \sum_{i=1}^{n} w_iA_i$.*

Remark 5.2. All the preceding properties follow directly as restrictions of the more general version for $\mathbb{P}$ given in [15], except for (P5), which was shown in [15] for the Thompson metric. But the version given here holds in general for NPC-metric spaces, in particular for (Σ, δ); see, for example, [14].

We have seen how the least squares mean on Σ can be viewed as the restriction of the Karcher mean on $\mathbb{P}$ to Σ. We now examine a reverse construction: extending the least squares mean on Σ to $\mathbb{P}$.

For an infinite-dimensional Hilbert space H, let $\{\alpha\}_{\alpha\in\Delta}$ be a nonempty collection of non-zero finite-dimensional subspaces of H such that given any finite-dimensional subspace, there is some $\alpha \in \Delta$ containing it. Then the family Δ ordered by inclusion is directed, since for $\alpha, \beta \in \Delta$ there is a finite-dimensional subspace containing both of them and a γ containing this subspace. Let $P_\alpha : H \to H$ denote the orthogonal projection onto the subspace α. We view $\{P_\alpha : \alpha \in \Delta\}$ as a net indexed by Δ that strongly converges to the identity I, since for any $x \in H$, $P_\alpha(x) = x$ eventually (i.e., there exists $\beta \in \Delta$ such that $P_\alpha(x) = x$ for all $\alpha \geq \beta$). We note that, P_α satisfies $\|P_\alpha\| = 1$ for all α, that P_α is Hermitian, positive semidefinite, and idempotent, and that $P_\alpha \leq I$. If $\alpha \leq \beta$, then $P_\alpha = P_\alpha P_\beta = P_\beta P_\alpha = P_\beta P_\alpha P_\beta$, so

$$\langle x, P_\alpha x\rangle = \langle x, P_\beta P_\alpha P_\beta x\rangle = \langle P_\beta x, P_\alpha(P_\beta x)\rangle \leq \langle P_\beta x, P_\beta x\rangle = \langle x, P_\beta x\rangle$$

since $\langle x, P_\alpha x\rangle \leq \langle x, x\rangle$ for all x. We conclude that, $P_\alpha \leq P_\beta$ for $\alpha \leq \beta$ so $\{P_\alpha : \alpha \in \Delta\}$ is a monotonically increasing net strongly converging to its supremum the identity I.

Since $\{P_\alpha : \alpha \in \Delta\}$ is bounded, the net $\{P_\alpha A P_\alpha\}$ strongly converges to A for any $A \in \mathcal{B}(H)$. For A Hermitian, it is a monotonically increasing net with supremum A. (One can show that A is the supremum directly or use the standard fact that any monotonically increasing net of symmetric operators that is bounded above strongly converges to its supremum.)

We consider the extended Hilbert–Schmidt subalgebra $\mathcal{H}_{\mathbb{C}}$ of $\mathcal{B}(H)$ consisting of all operators of the form $\lambda I + B$, where B comes from the closed ideal of Hilbert–Schmidt operators. Any $P_\alpha A P_\alpha$ has finite-dimensional rank, in particular is a Hilbert–Schmidt operator. We have seen that, the open cone Σ of positive elements has a least squares mean with a rich theory, which we wish to extend to the cone $\mathbb{P}$ of positive operators in $\mathcal{B}(H)$.

Proposition 5.3. *Let $A_1, \ldots, A_n \in \mathbb{P}$, and let $\omega = (w_1, \ldots, w_n)$ be a weight. Choose m large enough such that $e^{-m}I < A_i < e^m I$ for $1 \leq i \leq n$. Then*

$$X_\alpha = \Lambda(\omega; e^m I - P_\alpha(e^m I - A_1)P_\alpha, \ldots, e^m I - P_\alpha(e^m I - A_n)P_\alpha)$$

is a monotonically decreasing net in Σ bounded below by $e^{-m}I$ that strongly converges to its infimum, which is equal to the Karcher mean $\tilde{\Lambda}(\omega; A_1, \ldots, A_n)$.

Proof. Since the net $\{P_\alpha(e^m I - A_i)P_\alpha\}_\alpha$ is a monotonically increasing net strongly converging to its supremum $e^m I - A_i$, the net $\{e^m I - P_\alpha(e^m I - A_i)P_\alpha\}_\alpha$ is a decreasing net strongly converging to it infimum $e^m I - (e^m I - A_i) = A_i \geq e^{-m}I$. By the idempotency and monotonicity (Property (P4) of Theorem 5.1) of the least-squares mean, we have that X_α is a decreasing net bounded below by $e^{-m}I$, and hence strongly converges to its infimum, call it Y.

By Theorem 4.3 each X_α satisfies the Karcher equation:

$$\sum_{i=1}^{n} \omega_i \log(X_\alpha^{1/2}[e^m I - P_\alpha(e^m I - A_i)P_\alpha]^{-1} X_\alpha^{1/2}) = 0.$$

Since the function $f(X) = \sum_{i=1}^{n} w_i \log(X^{1/2}A_i^{-1}X^{1/2})$ is strongly continuous on the bounded order interval $[e^{-m}I, e^m I]$ (see [15, Lemma 5.4]), we conclude that

$$0 = f(Y) = \sum_{i=1}^{n} w_i \log(Y^{1/2}A_i^{-1}Y^{1/2}).$$

Hence Y is equal to the Karcher mean $\tilde{\Lambda}(\omega; A_1, \ldots, A_n)$. □

From the proof one extracts the following special case of strong continuity of Λ.

Corollary 5.4. *Let $\mathbb{A}_\alpha = (A_{1,\alpha}, \ldots, A_{n,\alpha})$ be a decreasing resp. increasing net in $\mathbb{P}^n$ that strongly converges to its infimum resp. supremum $\mathbb{A} = (A_1, \ldots, A_n) \in \mathbb{P}^n$. Then $\tilde{\Lambda}(\omega; \mathbb{A}_\alpha)$ is a decreasing resp. increasing net strongly converging to $\tilde{\Lambda}(\omega; \mathbb{A})$.*

Remark 5.5. Not only is Σ strongly dense in $\mathbb{P}$, but, as we have seen, every member of $\mathbb{P}^n$ can be obtained as an infimum resp. supremum of a decreasing resp. increasing net in Σ^n, which implies that the net is strongly convergent to that member. By Corollary 5.4 one has monotonic and strong convergence of the corresponding Karcher means. In this sense the Karcher mean on $\mathbb{P}$ is the unique extension of the least squares mean on Σ that is strongly continuous on monotonic nets.

Remark 5.6. We have assumed throughout this paper the separability of the Hilbert space H, primarily because this is the underlying assumption of [12], upon which our work depends. However, one can obtain the major results of this paper for the Karcher and least squares mean of bounded operators $\{A_1, \ldots, A_n\}$ on a general Hilbert space by taking the separable C^*-subalgebra generated by these operators and using the separable version of the Gelfand–Naimark theorem to reinterpret these operators as acting on a separable Hilbert space.

Acknowledgment

The authors thank the referee for pointing out the content included in Remark 3.5. The work of the second author was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (No. 2012-005191).

References

[1] M. Berger, A Panoramic View of Riemannlan Geometry, Springer-Verlag, 2003.
[2] R. Bhatia, Positive Definite Matrices, in: Princeton Series in Applied Mathematics, Princeton University Press, Princeton, NJ, 2007.
[3] R. Bhatia, The Riemannian mean of positive matrices, in: F. Nielsen, R. Bhatia (Eds.), Matrix Information Geometry, Springer, 2012, pp. 35–51.
[4] R. Bhatia, J. Holbrook, Riemannian geometry and matrix geometric means, Linear Algebra Appl. 413 (2006) 594–618.
[5] R. Bhatia, R. Karandikar, Monotonicity of the matrix geometric mean, Math. Ann. 353 (2012) 1453–1467.
[6] M. Bridson, A. Haefliger, Metric Spaces of Non-Positive Curvature, Springer, Berlin, 1999.
[7] C. Conde, G. Larotonda, Manifolds of semi-negative curvature, Proc. Lond. Math. Soc. 100 (2010) 670–704.
[8] J. Holbrook, No dice: a determinic approach to the Cartan centroid, J. Ramanujan Math. Soc. (in press).
[9] J. Jost, Nonpositive Curvature: Geometric and Analytic Aspects, in: Lectures in Mathematics, ETH Zürich, Birkhäuser, Basel, 1997.
[10] H. Karcher, Riemannian center of mass and mollifier smoothing, Comm. Pure Appl. Math. 30 (1977) 509–541.
[11] S. Lang, Fundamentals of Differential Geometry, in: Graduate Texts in Mathematics, Springer, 1999.
[12] G. Larotonda, Nonpositive curvature: a geometric approach to Hilbert–Schmidt operators, Differential Geom. Appl. 25 (2007) 679–700.
[13] G. Larotonda, Geodesic Convexity, Symmetric Spaces and Hilbert–Schmidt Operators, Ph.D. Thesis, Universidad Nacional de Bueros Aires, 2005.
[14] J. Lawson, Y. Lim, Monotonic properties of the least squares mean, Math. Ann. 351 (2011) 267–279.
[15] J. Lawson, Y. Lim, Karcher means and Karcher equations of positive definite operators, Trans. Amer. Math. Soc. (in press).
[16] Y. Lim, M. Pálfia, The matrix power means and the Karcher mean, J. Funct. Anal. 262 (2012) 1498–1514.
[17] M. Moakher, A differential geometric approach to the geometric mean of symmetric positive-definite matrices, SIAM J. Matrix Anal. Appl. 26 (2005) 735–747.
[18] G. Mostow, Some new decomposition theorems for semi-simple groups, Mem. Amer. Math. Soc. 1955 (1955) 31–54.
[19] K.-T. Sturm, Probability measures on metric spaces of nonpositive curvature, in: Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces, (Paris, 2002), in: Contemp. Math., vol. 338, Amer. Math. Soc., Providence, RI, 2003, pp. 357–390.