9-24-2013

Weighted means and Karcher equations of positive operators

Jimmie Lawson
Louisiana State University

Yongdo Lim
Sungkyunkwan University

Recommended Citation

Lawson, J., & Lim, Y. (2013). Weighted means and Karcher equations of positive operators. *Proceedings of the National Academy of Sciences of the United States of America, 110* (39), 15626-15632.
https://doi.org/10.1073/pnas.1313640110

Weighted means and Karcher equations of positive operators

Jimmie Lawson[a,1] **and Yongdo Lim**[b,1]

[a]Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803; and [b]Department of Mathematics, Sungkyunkwan University, Suwon 440-746, Korea

Edited by Charles L. Epstein, University of Pennsylvania, Philadelphia, PA, and accepted by the Editorial Board August 8, 2013 (received for review July 16, 2013)

The Karcher or least-squares mean has recently become an important tool for the averaging and study of positive definite matrices. In this paper, we show that this mean extends, in its general weighted form, to the infinite-dimensional setting of positive operators on a Hilbert space and retains most of its attractive properties. The primary extension is via its characterization as the unique solution of the corresponding Karcher equation. We also introduce power means P_t in the infinite-dimensional setting and show that the Karcher mean is the strong limit of the monotonically decreasing family of power means as $t \to 0^+$. We show each of these characterizations provide important insights about the Karcher mean.

Hilbert–Schmidt algebra | Riemannian manifold | Thompson metric

1. Introduction

Positive definite matrices have become fundamental computational objects in many areas of engineering, computer science, physics, statistics, and applied mathematics. They appear in a diverse variety of settings: covariance matrices in statistics, elements of the search space in convex and semidefinite programming, kernels in machine learning, density matrices in quantum information, data points in radar imaging, and diffusion tensors in medical imaging, to cite only a few. A variety of computational algorithms have arisen for approximations, interpolation, filtering, estimation, and averaging.

The process of averaging typically involves taking some type of matrix mean for some finite number of positive matrices of fixed dimension. Since the pioneering paper of Kubo and Ando (1), an extensive theory of two-variable means has sprung up for positive matrices and operators, but the n-variable case for $n>2$ has remained problematic. Once one realizes, however, that the matrix geometric mean $g_2(A,B)=A\#B:=A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}$, the appropriate noncommutative analog of the real geometric mean $\sqrt{ab}$, is the metric midpoint of A and B for the trace metric δ on the set $\mathbb{P}$ of positive definite matrices of some fixed dimension—see, e.g., refs. 2 and 3—it is natural to use an averaging technique over this metric to extend this mean to a larger number of variables. First, Moakher (4) and then Bhatia and Holbrook (5) suggested extending the geometric mean to n-points by taking the mean to be the unique minimizer of the sum of the squares of the distances:

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

where $\delta(X,A_i)=\left\|\log X^{-1/2}A_iX^{-1/2}\right\|_F$. This idea had been anticipated by Élie Cartan (ref. 6, section 6.1.5), who showed 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 Karcher (7), whose ideas are important to the present work.

Another approach to generalizing the geometric mean to n-variables, independent of metric notions, was suggested by Ando, Li, and Mathias (8) via a "symmetrization procedure" and induction. The Ando–Li–Mathias paper was also important for listing, and deriving for their mean, 10 desirable properties for multivariable geometric means. Moaker and Bhatia and Holbrook were able to establish a number of these important properties for the least-squares mean, but the important question of the monotonicity of this mean, conjectured by Bhatia and Holbrook (5), was left open. However, the authors were recently able to show (9) that all of the properties, in particular the monotonicity, are satisfied in the more general setting of weighted means for any weight $\omega=(w_1,\ldots w_n)$ of nonnegative entries summing to 1 and positive matrix n-tuple $\mathbb{A}=(A_1,\ldots,A_n)$:

- (P1) (consistency with scalars) $g_n(\omega;\mathbb{A})=A_1^{w_1}\cdots A_n^{w_n}$ if the A_i 's commute;
- (P2) (joint homogeneity) $g_n(\omega;a_1A_1,\ldots,a_nA_n)=a_1^{w_1}\cdots a_n^{w_n}g_n(\omega;\mathbb{A})$;
- (P3) (permutation invariance) $g_n(\omega_\sigma;\mathbb{A}_\sigma)=g_n(\omega;\mathbb{A})$, where $\omega_\sigma=(w_{\sigma(1)},\ldots,w_{\sigma(n)})$;
- (P4) (monotonicity) if $B_i\le A_i$ for all $1\le i\le n$, then $g_n(\omega;\mathbb{B})\le g_n(\omega;\mathbb{A})$;
- (P5) (continuity) the map $g_n(\omega;\cdot)$ is continuous;
- (P6) (congruence invariance) $g_n(\omega;M^*\mathbb{A}M)=M^*g_n(\omega;\mathbb{A})M$ for any invertible M;
- (P7) (joint concavity) $g_n(\omega;\lambda\mathbb{A}+(1-\lambda)\mathbb{B})\ge\lambda g_n(\omega;\mathbb{A})+(1-\lambda)g_n(\omega;\mathbb{B})$;
- (P8) (self-duality) $g_n(\omega;A_1^{-1},\ldots,A_n^{-1})^{-1}=g_n(\omega;A_1,\ldots,A_n)$;
- (P9) (AGH weighted mean inequalities) $\left(\sum_{i=1}^n w_iA_i^{-1}\right)^{-1}\le g_n(\omega;\mathbb{A})\le\sum_{i=1}^n w_iA_i$;
- (P10) (determinental identity) $\mathrm{Det}g_n(\omega;\mathbb{A})=\prod_{i=1}^n(\mathrm{Det}A_i)^{w_i}$.

Significance

As positive matrices and operators have gained increased prominence in theoretical, applied, and computational settings, finding appropriate methods for averaging them has become an important task. In recent years, the minimizer of the (weighted) sum of the distances (in an appropriately chosen metric) to the points to be averaged has been shown to exhibit many attractive features. In this paper, we extend most of these results to the infinite-dimensional setting, where the metric definition needs to be replaced by a solution shown to be unique of a corresponding equation called the Karcher equation. A multivariable weighted operator mean results that in many senses generalizes the geometric mean of a finite number of positive real numbers.

Author contributions: J.L. and Y.L. contributed equally to this work and its writing and are joint first authors.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. C.L.E. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. E-mail: lawson@math.lsu.edu or ylim@skku.edu.

 www.pnas.org/cgi/doi/10.1073/pnas.1313640110

A key ingredient in the derivation of many of these properties, the monotonicity in particular, is the fact that the trace metric on the manifold of positive definite matrices gives them the structure of a Cartan–Hadamard Riemannian manifold, in particular a manifold of nonpositive curvature. This implies that equipped with the Riemannian distance metric the manifold is a complete metric space that satisfies the "semiparallelogram law": for each X, Y, there exists P satisfying

$$\delta^2(P,Z) \le \frac{1}{2}\delta^2(X,Z) + \frac{1}{2}\delta^2(Y,Z) - \frac{1}{4}\delta^2(X,Y) \qquad [1.1]$$

for all Z. The point P turns out to be the unique metric midpoint between X and Y and lies on a metric geodesic between them. Such spaces are called metric spaces of nonpositive curvature, NPC-spaces for short, or CAT_0 -spaces, a widely studied class of metric spaces with a rich structure (see, e.g., refs. 10, 11, and 12, chapter 11).

Because in the statistical, quantum, and other settings as well, one may be interested in the more general case of positive bounded linear operators on an infinite-dimensional Hilbert space, one would like to have a suitable and effective averaging procedure for this context also. However, the significant theory that has developed for the multivariable least-squares mean does not readily carry over to the setting of positive operators on a Hilbert space, because one has no such Riemannian structure nor NPC-metric available. Fortunately, there is an alternative path through which one may approach this mean besides the least-squares path. If $X = g_n(\omega; \mathbb{A})$, then by invariance under congruence transformations $I = g_n(\omega; X^{-1/2}\mathbb{A}X^{-1/2})$. However, in the finite-dimensional setting, the latter equation holds if and only if $0 = \log I$ is the ω-weighted arithmetic mean of $\log(X^{-1/2}A_1X^{-1/2}), \ldots, \log(X^{-1/2}A_nX^{-1/2})$, i.e.,

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

We refer to this equation as the Karcher equation, a version of which appears in Karcher's paper (7). We show that the Karcher equation of positive definite operators has a unique solution in $\mathbb{P}$, call the solution "the Karcher mean," and establish that the preceding properties (P1) to (P9) are satisfied by it.

In *Section 2*, we recall the Thompson metric and list properties of it that will be important for our development. *Section 3* introduces the important tool of power means, which we need to establish existence of the Karcher mean, but the fact that well-behaved power means exist for the Hilbert operator setting is of independent interest. Lim and Pálfia (13) have recently shown in the finite-dimensional setting that the Karcher or least-squares mean is the limit as $t \to 0^+$ of the power means P_t. We show additionally that they are monotonically decreasing, which allows us (*Section 4*) to deduce the existence of their limit in the strong topology in the general Hilbert space setting. We use this fact for our initial provisional definition of the Karcher mean. We show in *Section 5* that the Karcher mean defined in this way does indeed satisfy the Karcher equation, and in *Section 6* we establish that it is the unique solution and present a list of the fundamental properties of the Karcher mean. Although for convenience we carry out our work in $\mathcal{B}(H)$, the C^* -algebra of all bounded linear operators on a Hilbert space H, our constructions only require that we be working in a monotone complete subalgebra of $\mathcal{B}(H)$.

In the finite-dimensional Riemannian setting the Karcher equation is (equivalent to) the condition for the vanishing of the gradient of the least-squares distance function. In *Section 7*, we indicate that this remains true for the Riemannian manifold of extended Hilbert–Schmidt operators. The Karcher mean on $\mathbb{P}$ is then the continuous extension via directed strong limits of the least-squares mean on the Riemannian manifold of extended Hilbert–Schmidt operators. Thus, a ghostly connection is also maintained to the least-squares mean, even in the infinite-dimensional setting.

2. The Operator Geometric Mean and the Thompson Metric

For a Hilbert space H, let $\mathcal{B}(H)$ be the Banach space of bounded linear operators on H equipped with the operator norm, $\mathcal{S}(H)$ the closed subspace of bounded self-adjoint 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: $\Gamma_M(X) = MXM^*$. For $X, Y \in \mathcal{S}(H)$, we write $X \le Y$ if $Y - X$ is positive semidefinite, and $X < Y$ if $Y - X$ is positive definite. Note that $X \le Y$ if and only if $\langle x, Xx\rangle \le \langle x, Yx\rangle$ for all $x \in H$.

It is natural to define the operator geometric mean $I\#A$ of the identity I and $A \in \mathbb{P}$ to be $\sqrt{IA} = A^{1/2}$, and more generally the t-weighted geometric mean by $I\#_tA = I^{1-t}A^t = A^t$ (the geometric mean being the case $t = 1/2$); see property (P1) in *Introduction*. If one extends these definitions to $\mathbb{P} \times \mathbb{P}$ so that the resulting weighted geometric mean is invariant under congruence transformations (property (P6)), then the "t-weighted geometric mean" is uniquely given by the following:

$$A\#_tB = A^{1/2}\left(A^{-1/2}BA^{-1/2}\right)^t A^{1/2}. \qquad [2.2]$$

The following properties for the weighted geometric mean are well known (1, 14, 15).

Lemma 2.1. *Let $A, B, C, D \in \mathbb{P}$ and let $t \in \mathbb{R}$. Then*

(*i*) $A\#_t B = A^{1-t}B^t$ *for* $AB = BA$;

(*ii*) $(aA)\#_t(bB) = a^{1-t}b^t(A\#_t B)$ *for* $a, b > 0$;

(*iii*) (*Loewner–Heinz inequality*) $A\#_t B \le C\#_t D$ *for* $A \le C, B \le D$ *and* $t \in [0,1]$;

(*iv*) $M(A\#_t B)M^* = (MAM^*)\#_t(MBM^*)$ *for* $M \in \mathrm{GL}(H)$;

(*v*) $A\#_t B = B\#_{1-t}A$, $(A\#_t B)^{-1} = A^{-1}\#_t B^{-1}$;

(*vi*) $(\lambda A + (1-\lambda)B)\#_t(\lambda C + (1-\lambda)D) \ge \lambda(A\#_t C) + (1-\lambda)(B\#_t D)$ *for* $\lambda, t \in [0,1]$;

(*vii*) $((1-t)A^{-1} + tB^{-1})^{-1} \le A\#_t B \le (1-t)A + tB$ *for* $t \in [0,1]$;

(*viii*) $(A\#_t B)\#_s(A\#_u B) = A\#_{(1-s)t+su} B$ *for any* $s, t, u \in \mathbb{R}$.

As mentioned in Introduction, we do not have available in the infinite-dimensional setting a metric comparable to the trace metric that endows $\mathbb{P}$ with the structure of a nonpositively curved metric space. However, there is a useful metric on $\mathbb{P}$ that satisfies a weaker property of nonpositivity. The Thompson metric on $\mathbb{P}$ is defined by $d(A,B) = \|\log(A^{-1/2}BA^{-1/2})\|$, where $\|X\|$ denotes the operator norm of X. It is known that d is a complete metric on $\mathbb{P}$, that it induces on $\mathbb{P}$ the operator norm topology, and that $d(A,B) = \max\{\log M(B/A), \log M(A/B)\}$, where $M(B/A) = \inf\{\alpha > 0 : B \le \alpha A\}$ (14, 16, 17). We note that the Thompson metric (in the second form) exists on all normal cones of real Banach spaces. For the following lemma, see refs. 14, 15, and 18.

Lemma 2.2. *Basic properties of the Thompson metric on $\mathbb{P}$ include the following:*

(*i*) $d(A,B) = d(rA, rB) = d(A^{-1}, B^{-1}) = d(MAM^*, MBM^*)$;

(*ii*) $d(A\#B, A) = d(A\#B, B) = \frac{1}{2}d(A,B)$;

(*iii*) $d(A\#_tB, C\#_t D) \le (1-t)d(A,C) + td(B,D),\ t \in [0,1]$;

(*iv*) $d(A\#_tB, A\#_s B) = |s-t|d(A,B), s, t \in [0,1]$.

Property (*iii*) is a weakened version of nonpositivity for a metric and is often referred to as "Busemann nonpositive curvature" of the metric.

MATHEMATICS

Remark 2.3. It follows from $d(A,B)=\|\log(A^{-1/2}BA^{-1/2})\|$ that the exponential function $\exp:\mathcal{S}(H)\to\mathbb{P}$ is an isometry when restricted to any one-dimensional subspace of $\mathcal{S}(H)$ equipped with the operator norm distance to its image in $\mathbb{P}$ equipped with the Thompson metric. The Thompson metric is characterized as the only metric on $\mathbb{P}$ with this property that is invariant under congruence transformations.

The following nonexpansive property of addition for the Thompson metric will be useful for our purpose [see ref. 19, lemma 10.1, (*iv*)].

Lemma 2.4. *Let $A_i,B_i\in\mathbb{P}, i=1,2,\ldots,n$. Then*

$$d\left(\sum_{i=1}^{n}A_i,\sum_{i=1}^{n}B_i\right)\le\max_{1\le i\le n}\{d(A_i,B_i)\}.$$

Proof: For $n=2$, suppose that $d(A_1,B_1)\le d(A_2,B_2)=\log r$. Then $A_2\le rB_2$, $B_2\le rA_2$, $A_1\le rB_1$, $B_1\le rA_1$, and thus $A_1+A_2\le rB_1+rB_2=r(B_1+B_2)$, $B_1+B_2\le rA_1+rA_2=r(A_1+A_2)$. Hence $d(A_1+A_2,B_1+B_2)\le\log r=d(A_2,B_2)$. The general case easily follows by induction. ∎

3. Power Means

For positive real numbers $a_1,\ldots,a_n$, a weighting $\omega=(w_1,\ldots,w_n)$, and $t\neq0$, the "ω-weighted power mean of order t" is given by $P_t(\omega;a_1,\ldots,a_n)=\left(\sum_{i=1}^{n}w_ia_i^t\right)^{1/t}$. By elementary algebra, $x=\left(\sum_{i=1}^{n}w_ia_i^t\right)^{1/t}$ satisfies the equation $x=\sum_{i=1}^{n}w_ix^{1-t}a_i^t$. The formula for the power mean does not readily extend to the case of positive operators, but its equational characterization does, as observed by Lim and Pálfia (13) in the case of positive definite matrices. Their notion and most of their results readily extend to the setting of positive operators on a Hilbert space, as we point out in this section.

In what follows we let $\Delta_n=\{(w_1,\ldots,w_n)\in[0,1]^n:\sum_{i=1}^{n}w_i=1\}$.

Theorem 3.1. *Let $A_1,\ldots,A_n\in\mathbb{P}$ and let $\omega=(w_1,\ldots,w_n)\in\Delta_n$. Then for each $t\in(0,1]$, the following equation has a unique positive definite solution*:

$$X=\sum_{i=1}^{n}w_i(X\#_tA_i). \quad [3.3]$$

Proof: The map $f:\mathbb{P}\to\mathbb{P}$ defined by $f(X)=\sum_{i=1}^{n}w_i(X\#_tA_i)$ is a strict contraction with respect to the Thompson metric, because for $X,Y>0$

$$\begin{aligned}d(f(X),f(Y))&\le\max_{1\le i\le n}\{d(w_i(X\#_tA_i),w_i(Y\#_tA_i))\}\\&\le\max_{1\le i\le n}\{d(X\#_tA_i,Y\#_tA_i)\}\\&\le\max_{1\le i\le n}\{(1-t)d(X,Y)\}=(1-t)d(X,Y).\end{aligned}$$

by *Lemma 2.2*, (*i*) and (*iii*), and *Lemma 2.4*. Thus, f has a unique fixed point. ∎

Definition 3.2. [Power means] Let $\mathbb{A}=(A_1,\ldots,A_n)\in\mathbb{P}^n$ and $\omega\in\Delta_n$. For $t\in(0,1]$, we denote by $P_t(\omega;\mathbb{A})$ the unique solution of Eq. **3.3**. For $t\in[-1,0)$, we define $P_t(\omega;\mathbb{A})=P_{-t}(\omega;\mathbb{A}^{-1})^{-1}$, where $\mathbb{A}^{-1}=(A_1^{-1},\ldots,A_n^{-1})$. We call $P_t(\omega;\mathbb{A})$ the ω-weighted power mean of order t of $A_1,\ldots,A_n$.

The power mean $P_t(\omega;\mathbb{A})$ has an interesting geometrical interpretation for $t>0$: it is the unique point X having the property that when we move along the Thompson metric geodesic curve $t\mapsto X\#_tA_i$ a t th amount of the distance from X toward each A_i and take the weighted arithmetic average of the resulting points, we recover X. More importantly, it satisfies the axiomatic properties listed in *Introduction*, or mild variants thereof.

Proposition 3.3. *For $\mathbb{A}=(A_1,\ldots,A_n),\mathbb{B}=(B_1,\ldots,B_n)\in\mathbb{P}^n,\omega\in\Delta_n,\sigma\in S^n$, a permutation on n-letters, and $t\in[-1,1]\backslash\{0\}$:*

(1) $P_t(\omega;\mathbb{A})=\left(\sum_{i=1}^{n}w_iA_i^t\right)^{\frac{1}{t}}$ *if the A_i 's commute;*

(2) $P_t(\omega;aA_1,\ldots,aA_n)=aP_t(\omega;\mathbb{A})$;

(3) $P_t(\omega_\sigma;\mathbb{A}_\sigma)=P_t(\omega;\mathbb{A})$ *for any permutation σ;*

(4) $P_t(\omega;\mathbb{A})\le P_t(\omega;\mathbb{B})$ *if $A_i\le B_i$ for all $i=1,2,\ldots,n$;*

(5) $d(P_t(\omega;\mathbb{A}),P_t(\omega;\mathbb{B}))\le\max_{1\le i\le n}\{d(A_i,B_i)\}$;

(6) $(1-u)P_{|t|}(\omega;\mathbb{A})+uP_{|t|}(\omega;\mathbb{B})\le P_{|t|}(\omega;(1-u)\mathbb{A}+u\mathbb{B})$ *for any $u\in[0,1]$;*

(7) $P_t(\omega;M\mathbb{A}M^*)=MP_t(\omega;\mathbb{A})M^*$ *for any invertible M;*

(8) $P_t(\omega;\mathbb{A}^{-1})^{-1}=P_{-t}(\omega;\mathbb{A})$;

(9) $\left(\sum_{i=1}^{n}w_iA_i^{-1}\right)^{-1}\le P_t(\omega;\mathbb{A})\le\sum_{i=1}^{n}w_iA_i$.

Proof: Most of the proofs follow from properties of the two-variable weighted geometric mean and the fact that power means are equationally defined from it. We illustrate with the proof of the important monotonicity property, item (4).

Suppose that $A_i\le B_i$ for all $i=1,2,\ldots,n$. Let $t\in(0,1]$. Define $f(X)=\sum_{i=1}^{n}w_i(X\#_tA_i)$ and $g(X)=\sum_{i=1}^{n}w_i(X\#_tB_i)$. Then $P_t(\omega;\mathbb{A})=\lim_{k\to\infty}f^k(X)$ and $P_t(\omega;\mathbb{B})=\lim_{k\to\infty}g^k(X)$ for any $X\in\mathbb{P}$, by the Banach fixed point theorem. By the Loewner–Heinz inequality [*Lemma 2.1*, (*iii*)], $f(X)\le g(X)$ for all $X\in\mathbb{P}$, and $f(X)\le f(Y), g(X)\le g(Y)$ whenever $X\le Y$. Let $X_0>0$. Then $f(X_0)\le g(X_0)$ and $f^2(X_0)=f(f(X_0))\le g(f(X_0))\le g^2(X_0)$. Inductively, we have $f^k(X_0)\le g^k(X_0)$ for all $k\in\mathbb{N}$. Therefore, $P_t(\omega;\mathbb{A})=\lim_{k\to\infty}f^k(X_0)\le\lim_{k\to\infty}g^k(X_0)=P_t(\omega;\mathbb{B})$.

We also derive item (5), because we need it later. Let $X=P_t(\omega;\mathbb{A})$ and $Y=P_t(\omega;\mathbb{B})$. Then from *Lemma 2.2*, (*i*) and (*iii*), and *Lemma 2.4*,

$$\begin{aligned}d(X,Y)&=d\left(\sum_{i=1}^{n}w_i(X\#_tA_i),\sum_{i=1}^{n}w_i(Y\#_tB_i)\right)\\&\le\max_{1\le i\le n}d(X\#_tA_i,Y\#_tB_i)\\&\le\max_{1\le i\le n}[(1-t)d(X,Y)+td(A_i,B_i)]\\&=(1-t)d(X,Y)+t\max_{1\le i\le n}d(A_i,B_i).\end{aligned}$$

Other proofs are similar to those of ref. 13. ∎

4. The Power Mean Limit

In ref. 13, Lim and Pálfia have shown in the finite-dimensional setting that the Karcher or least-squares mean is the limit as $t\to0^+$ of the (monotonically decreasing) family of power means P_t. We take this characterization as the launch point for establishing existence of the infinite-dimensional Karcher mean.

We recall that the strong topology on the space $\mathcal{B}(H)$ of bounded linear operators is the topology of pointwise convergence. If a net of positive semidefinite operators A_α converges strongly to A, then the nonnegative values $\langle x,A_\alpha x\rangle$ must converge to a nonnegative $\langle x,Ax\rangle$, so the cone $\{A:0\le A\}$ is strongly closed. Hence the partial order $\{(A,B)\in\mathcal{S}(H)\times\mathcal{S}(H):A\le B\}$ is strongly closed. We recall also the well-known fact that any monotonically decreasing net of self-adjoint operators that is bounded below possesses an infimum A to which it strongly converges [see, for example, theorem 4.28(b) of ref. 20]. Dually a monotonically increasing net that is bounded above strongly converges to it supremum.

For $G,H:\Delta_n\times\mathbb{P}^n\to\mathbb{P}$, we define $G\le H$ if $G(\omega;\mathbb{A})\le H(\omega;\mathbb{A})$ for all $\omega\in\Delta_n$ and $\mathbb{A}\in\mathbb{P}^n$. We note that $\mathcal{H}\le\mathcal{A}$, the arithmetic-harmonic mean inequality [Proposition 3.3, (9)].

Theorem 4.1. *Let $\omega\in\Delta_n$ and $\mathbb{A}\in\mathbb{P}^n$. Then there exist $X_\pm\in\mathbb{P}$ such that*

$$\lim_{t\to 0^\pm} P_t(\omega;\mathbb{A}) = X_\pm$$

under the strong-operator topology. Define $P_{0^\pm}(\omega;\mathbb{A}) = X_\pm$. *Then for* $0 < t \le s \le 1$,

$$\mathcal{H} = P_{-1} \le P_{-s} \le P_{-t} \le \cdots \le P_{0^-} \le P_{0^+} \le \cdots \le P_t \le P_s \le \mathcal{A}.$$

Proof: Let $\omega = (w_1,\ldots,w_n) \in \Delta_n$ and $\mathbb{A} = (A_1,\ldots,A_n) \in \mathbb{P}^n$. To indicate how the inequalities are established, we show only that $P_t \le P_s \le P_1 = \mathcal{A}$ for $0 < t \le s \le 1$.

Let $f : \mathbb{P} \to \mathbb{P}$ be defined by $f(X) = \sum_{i=1}^n w_i (X \#_t A_i)$. By the Banach fixed point theorem, $P_t(\omega;\mathbb{A}) = \lim_{k\to\infty} f^k(X)$ for any $X \in \mathbb{P}$. We observe from the fact $X = X\#_0 A_i$ and *Lemma 2.1*, (*vii*) and (*viii*), that

$$f(X) = \sum_{i=1}^n w_i (X\#_t A_i) = \sum_{i=1}^n w_i \Big[X \#_{\frac{t}{s}} (X\#_s A_i)\Big]$$
$$\le \sum_{i=1}^n w_i \Big[\Big(1-\frac{t}{s}\Big)X + \frac{t}{s}(X\#_s A_i)\Big] = \Big(1-\frac{t}{s}\Big)X + \frac{t}{s}\sum_{i=1}^n w_i (X\#_s A_i).$$

Applying the preceding to $X_0 = P_s(\omega;\mathbb{A})$ yields

$$f(X_0) \le \Big(1-\frac{t}{s}\Big)X_0 + \frac{t}{s}\sum_{i=1}^n w_i (X_0 \#_s A_i) = \Big(1-\frac{t}{s}\Big)X_0 + \frac{t}{s}X_0 = X_0.$$

Because f is monotonic [*Lemma 2.1*, (*iii*)], $f^{k+1}(X_0) \le f^k(X_0) \le \cdots \le f(X_0) \le X_0$ for all $k \in \mathbb{N}$. Therefore, $P_t(\omega;\mathbb{A}) = \lim_{k\to\infty} f^k(X_0) \le X_0 = P_s(\omega;\mathbb{A})$.

The nets $\{P_t(\omega;\mathbb{A})\}_{t>0}$ and $\{P_{-t}(\omega;\mathbb{A})\}_{t>0}$ are monotonic and bounded between $\mathcal{H}(\omega;\mathbb{A})$ and $\mathcal{A}(\omega;\mathbb{A})$. Therefore, there exist $X_\pm \in \mathbb{P}$ such that

$$\lim_{t\to 0^+} P_t(\omega;\mathbb{A}) = X_+, \quad \lim_{t\to 0^-} P_t(\omega;\mathbb{A}) = X_-$$

under the strong-operator topology. From the basic inequalities $P_t \ge P_{-t}$ for all $t \in (0,1]$ and the strong closedness of the partial order on $\mathcal{S}(H)$, their strong limits satisfy $X_+ \ge X_-$.

Definition 4.2. We set $\Lambda(\omega; A_1,\ldots,A_n) = P_{0^+}(\omega; A_1,\ldots,A_n)$ and call it the ω-weighted Karcher mean of $A_1,\ldots,A_n$. We set $\Lambda^*(\omega; A_1,\ldots,A_n) = P_{0^-}(\omega; A_1,\ldots,A_n)$.

With the help of *Theorem 4.1*, the basic properties of power means in *Proposition 3.3* carry over to limiting case of the Karcher mean and yield most of the axiomatic properties (P1) to (P9) given in *Introduction*.

Theorem 4.3. *The properties* (P1) *through* (P9), *except for* (P2) *and* (P8), *hold for the Karcher mean. Property* (P5) *holds in the strengthened form*:

(P5) *(continuity)* $d(\Lambda(\omega;\mathbb{A}), \Lambda(\omega;\mathbb{B})) \le \max_{1\le i\le n}\{d(A_i,B_i)\}$.

Proof: We provide a proof only for (P5), which we need later. Let $X_t = P_t(\omega;\mathbb{A})$ and $Y_t = P_t(\omega;\mathbb{B})$. Let $\alpha := \max_{1\le i\le n}\{d(A_i,B_i)\}$. By *Proposition 3.3*, $d(X_t,Y_t) \le \alpha$ for the Thompson metric d and thus $X_t \le e^\alpha Y_t$ for $0 < t \le 1$. Because $\Lambda(\omega;\mathbb{A}) \le X_t$ for each t, we have $\Lambda(\omega;\mathbb{A}) \le e^\alpha Y_t$. By the strong closedness of the order and the strong convergence of Y_t to $\Lambda(\omega;\mathbb{B})$ as $t \to 0^+$, $\Lambda(\omega;\mathbb{A}) \le e^\alpha \Lambda(\omega;\mathbb{B})$. Similarly, $\Lambda(\omega;\mathbb{B}) \le e^\alpha\Lambda(\omega;\mathbb{A})$. It follows from definition of the Thompson metric that $d(\Lambda(\omega;\mathbb{A}), \Lambda(\omega;\mathbb{B})) \le \alpha = \max\{d(A_i,B_i) : 1 \le i \le n\}$. ∎

In *Section 6*, we derive the missing two properties, further strengthen (P5), and list them all explicitly.

5. The Karcher Equation

Let $\mathbb{A} = (A_1,\ldots,A_n) \in \mathbb{P}^n$, $\omega = (w_1,\ldots,w_n) \in \Delta_n$. We consider the following nonlinear operator equation on $\mathbb{P}$, called the "Karcher equation":

$$\sum_{i=1}^n w_i \log\Big(X^{-1/2} A_i X^{-1/2}\Big) = 0. \qquad [5.4]$$

Note that multiplying by −1 yields the equivalent equation $\sum_{i=1}^n w_i \log(X^{1/2} A_i^{-1} X^{1/2}) = 0$, and we pass freely between the two.

In the finite-dimensional setting, it is known that the least-squares mean (the Karcher mean) satisfies the Karcher equation, indeed is the unique positive solution of the Karcher equation. Our goal in this section is to show that the Karcher mean we have defined in the preceding section satisfies the Karcher equation (and hence is aptly named). We work mainly in the strong topology and use heavily the following lemma, a special case of theorem 3.6 from Kadison's study of strongly continuous operator functions (21) on self-adjoint operators.

Lemma 5.1. *Let* Q *be an open or closed subset of* $\mathbb{R}$ *and let* $f : Q \to \mathbb{R}$ *be a continuous bounded function. Then the corresponding operator function* $\hat{f}$ *is strong-operator continuous on the set* $\mathcal{S}H_Q$ *of bounded self-adjoint operators on a Hilbert space* H *with spectra in* Q.

Lemma 5.2. *The following functions* (*i*) *and* (*ii*) *are strongly continuous on* $[e^{-m}I, e^m I] = \{A \in \mathcal{S}(H) : e^{-m}I \le A \le e^m I\}$ *and function* (*iii*) *is strongly continuous on* $[e^{-m}I, e^m I]^2$.

(*i*) *The logarithm map* $A \mapsto \log A$, *which is also monotonic.*

(*ii*) *The power map* $A \mapsto A^r$ *for* $-1 \le r \le 1$.

(*iii*) *The binary weighted mean map* $(A,B) \mapsto A\#_t B$ *for* $0 \le t \le 1$.

The last two functions have image contained in $[e^{-m}I, e^m I]$.

The following shows that the previously defined Karcher mean is indeed a solution of the Karcher equation (Eq. **5.4**).

Theorem 5.3. *For each* $\omega \in \Delta_n$ *and* $\mathbb{A} \in \mathbb{P}^n$, $\Lambda(\omega;\mathbb{A})$ *satisfies the Karcher equation*.

Proof: We sketch the proof of this crucial result; for details, see ref. 22. Let $t \in (0,1]$. Let $X_t = P_t(\omega;\mathbb{A})$ and let $X_0 = \Lambda(\omega;\mathbb{A}) = P_{0^+}(\omega;\mathbb{A})$. By our provisional definition of $\Lambda(\omega;\mathbb{A})$, with respect to the strong topology $X_t \to X_0$ monotonically as $t \to 0^+$ and $X_0 \le X_t \le X_1$ for all $t \in [0,1]$. Pick m such that for all i, $A_i, X_1, X_0 \in [e^{-m/2}I, e^{m/2}I]$, the Loewner order interval of operators between $e^{-m/2}I$ and $e^{m/2}I$. It follows that $X_t \in [e^{-m/2}I, e^{m/2}I]$ for all $0 \le t \le 1$.

By the order reversal of inversion $[e^{-m/2}I, e^{m/2}I]$ is closed under inversion so that $X_t^{-1/2} A_i X_t^{-1/2} \in [e^{-m}I, e^m I]$ for all $t \in [0,1]$ and $i = 1,\ldots,n$. By *Lemma 5.2*, (*ii*), $X_t^{-1/2} A_i X_t^{-1/2}$ converges strongly to $X_0^{-1/2} A_i X_0^{-1/2}$. By strong continuity of log on $[e^{-m}I, e^m I]$ [*Lemma 5.2*, (*i*)], $U_i := \log(X_t^{-1/2} A_i X_t^{-1/2}) \to V_i := \log(X_0^{-1/2} A_i X_0^{-1/2})$ for all i. One can then argue, although the argument is a bit delicate, that in any open ball $B_r(0)$, $r > e^m$, in the strong topology

$$\lim_{t\to 0} \frac{\Big(X_t^{-1/2} A_i X_t^{-1/2}\Big)^t - I}{t} = \lim_{t\to 0}\frac{e^{tU_i} - I}{t} = V_i = \log\Big(X_0^{-1/2} A_i X_0^{-1/2}\Big) \qquad [5.5]$$

for all $i = 1,\ldots,n$.

By definition, $X_t = \sum_{i=1}^n w_i (X_t \#_t A_i)$. Premultiplying and postmultiplying this equation by $X_t^{-1/2}$ and substituting from Eq. **2.2** for the weighted mean yields for $t > 0$:

$$I = \sum_{i=1}^n w_i \Big(X_t^{-1/2} A_i X_t^{-1/2}\Big)^t,$$

that is, $0 = \sum_{i=1}^n w_i \Big[\frac{(X_t^{-1/2} A_i X_t^{-1/2})^t - I}{t}\Big]$. By Eq. **5.5**,

MATHEMATICS

$$0 = \lim_{t \to 0^+} \sum_{i=1}^n w_i \left[\frac{\left(X_t^{-1/2} A_i X_t^{-1/2} \right)^t - I}{t} \right]$$

$$= \sum_{i=1}^n w_i \lim_{t \to 0^+} \left[\frac{\left(X_t^{-1/2} A_i X_t^{-1/2} \right)^t - I}{t} \right]$$

$$= \sum_{i=1}^n w_i \log\left(X_0^{-1/2} A_i X_0^{-1/2} \right).$$

This shows that $X_0 = \Lambda(\omega; \mathbb{A})$ is a solution of the Karcher equation. ∎

Corollary 5.4. *Dually, the operator* $\Lambda^*(\omega; A_1, \ldots, A_n) = P_{0^-}(\omega; A_1, \ldots, A_n)$ *also satisfies the Karcher equation.*

6. Uniqueness of the Karcher Mean

In this section, we establish that the Karcher equation (Eq. **5.4**) has unique solution the Karcher mean and summarize its fundamental properties. We begin locally.

Theorem 6.1. *Let* $\omega \in \Delta_n$. *Then there exists* $\epsilon_\omega > 1$ *such that the Karcher equation has unique solution in* $[\epsilon_\omega^{-1} I, \epsilon_\omega I]$ *the Karcher mean. Furthermore, the Karcher mean* $\Lambda(\omega; \cdot)$ *is* C^∞ *on a neighborhood of* $\mathbb{I} = (I, \ldots, I)$ *in* $\mathbb{P}^n$.

Proof: One considers the map $F^\omega : \mathbb{P}^n \times \mathbb{P} \to \mathcal{S}(H)$ defined by

$$F^\omega(A_1, \ldots, A_n, X) = \sum_{i=1}^n w_i \log\left(X^{1/2} A_i^{-1} X^{1/2} \right),$$

checks that the conditions of the implicit function theorem are satisfied at $(\mathbb{I}, I)$, and concludes that there exist open neighborhoods $U_{\mathbb{I}}$ of $\mathbb{I}$ in $\mathbb{P}^n$ and V_I of I in $\mathbb{P}$ and a C^∞ -mapping $g : U_{\mathbb{I}} \to V_I$ such that $g(\mathbb{I}) = I$ and $F^\omega(\mathbb{A}, X) = 0$ if and only if $X = g(\mathbb{A})$ for all $\mathbb{A} \in U_{\mathbb{I}}, X \in V_I$. One chooses $[\epsilon_\omega^{-1} I, \epsilon_\omega I] \subseteq V_I$ so that $[\epsilon_\omega^{-1} I, \epsilon_\omega I]^n \subseteq U_{\mathbb{I}}$ and notes that g and $\Lambda(\omega; \cdot)$ must agree on $[\epsilon_\omega^{-1} I, \epsilon_\omega I]^n$. ∎

Remark 6.2: By invariance of the Karcher mean under congruence transformations, the preceding result may be extended to a neighborhood of the diagonal in $\mathbb{P}^n$.

Theorem 6.3. *For all* $\omega = (w_1, \ldots, w_n) \in \Delta_n, \mathbb{A} = (A_1, \ldots, A_n) \in \mathbb{P}^n$ *and* $t \in (0, 1]$, *the following conditions are satisfied.*

(*i*) *the Karcher mean* $X = \Lambda(\omega; A_1, \ldots, A_n)$ *is the unique solution of the Karcher equation,* $0 = \sum_{i=1}^n \omega_i \log(X^{-1/2} A_i X^{-1/2})$.

(*ii*) $\Lambda : \Delta_n \times \mathbb{P}^n \to \mathbb{P}$ *is jointly homogeneous, that is,*
$\Lambda(\omega; a_1 A_1, \ldots, a_n A_n) = a_1^{w_1} \cdots a_n^{w_n} \Lambda(\omega; A_1, \ldots, A_n)$.

(*iii*) *For the Thompson metric,* $d(\Lambda(\omega; \mathbb{A}), \Lambda(\omega; \mathbb{B})) \leq \sum_{i=1}^n w_i d(A_i, B_i)$.

(*iv*) *The equation* $X = \Lambda(\omega; X \#_t A_1, \ldots, X \#_t A_n)$ *has a unique solution in* $\mathbb{P}$.

Furthermore, $X = \Lambda(\omega; X \#_t A_1, \ldots, X \#_t A_n)$ *if and only if* $X = \Lambda(\omega; A_1, \ldots, A_n)$.

Proof: We first show that Λ satisfies condition (*iv*). Fix $\omega \in \Delta_n$ and $\mathbb{A} = (A_1, \ldots, A_n)$. Define $f_t : \mathbb{P} \to \mathbb{P}$ by $f_t(X) = \Lambda(\omega; X \#_t A_1, \ldots, X \#_t A_n)$. Then

$$\begin{aligned} d(f_t(X), f_t(Y)) &= d(\Lambda(\omega; X \#_t A_1, \ldots, X \#_t A_n), \Lambda(\omega; Y \#_t A_1, \ldots, Y \#_t A_n)) \\ &\leq d\left(X \#_t A_j, Y \#_t A_j \right) := \max\{ d(X \#_t A_i, Y \#_t A_i) : 1 \leq i \leq n \} \\ &\leq (1-t) d(X, Y), \end{aligned}$$

where the first inequality follows from *Theorem 4.3*, (P5), and the second from *Lemma 2.2*, (*iii*). It follows that f_t is a strict contraction for the Thompson metric and hence has a unique fixed point, which is the unique solution for the equation of (*iv*).

(*iv*) implies (*i*): Let X satisfy the Karcher equation for $(\omega; \mathbb{A})$. Pick $\epsilon_\omega > 1$ as in *Theorem 6.1* and $t \in (0, 1)$ such that $(X^{-1/2} A_i X^{-1/2})^t \in [\epsilon_\omega^{-1} I, \epsilon_\omega I]$, $i = 1, \ldots, n$. Then $\sum_{i=1}^n w_i \log(X^{-1/2} A_i X^{-1/2}) = 0 = \sum_{i=1}^n w_i \log((X^{-1/2} A_i X^{-1/2})^t)$. Clearly $Y = I$ is then a solution of $\sum_{i=1}^n w_i \log(Y^{-1/2} (X^{-1/2} A_i X^{-1/2})^t Y^{-1/2}) = 0$. By monotonicity and idempotency of Λ, $\Lambda(\omega; (X^{-1/2} A_1 X^{-1/2})^t, \ldots, (X^{-1/2} A_n X^{-1/2})^t)$ must belong to $[\epsilon_\omega^{-1} I, \epsilon_\omega I]$. By the uniqueness of the Karcher solution on $[\epsilon_\omega^{-1} I, \epsilon_\omega I]$ (*Theorem 6.1*), $I = \Lambda(\omega; (X^{-1/2} A_1 X^{-1/2})^t, \ldots, (X^{-1/2} A_n X^{-1/2})^t)$. By invariancy under congruence transformations [property (P6) of *Theorem 4.3*],

$$\begin{aligned} X &= \Lambda\left(\omega; X^{1/2} \left(X^{-1/2} A_1 X^{-1/2} \right)^t X^{1/2}, \ldots, X^{1/2} \left(X^{-1/2} A_n X^{-1/2} \right)^t X^{1/2} \right) \\ &= \Lambda(\omega; X \#_t A_1, \ldots, X \#_t A_n). \end{aligned}$$

Because $\Lambda(\omega; A_1, \ldots, A_n)$ is one possibility for our original choice of X, property (*iv*) implies $X = \Lambda(\omega; \mathbb{A})$. This shows also that $\Lambda(\omega; \mathbb{A})$ is the unique solution of $X = \Lambda(\omega; X \#_t A_1, \ldots, X \#_t A_n)$.

(*i*) implies (*ii*): Let $Y = a_1^{w_1} \cdots a_n^{w_n} X$. Then

$$\begin{aligned} \sum_{i=1}^n w_i \log\left(Y^{-1/2} (a_i A_i) Y^{-1/2} \right) &= \sum_{i=1}^n w_i \log\left(\frac{a_i}{\prod_{i=1}^n a_i^{w_i}} X^{-1/2} A_i X^{-1/2} \right) \\ &= \sum_{i=1}^n w_i \left[\log\left(\frac{a_i}{\prod_{i=1}^n a_i^{w_i}} I \right) + \log\left(X^{-1/2} A_i X^{-1/2} \right) \right] \\ &= \sum_{i=1}^n w_i \log\left(X^{-1/2} A_i X^{-1/2} \right), \end{aligned}$$

where the second equality follows from the fact that $\log tA = \log tI + \log A$ for any $t > 0$ and $A > 0$. Therefore, the left-hand side equals 0 iff the right-hand side does, which translates to $Y = \Lambda(\omega; a_1 A_1, \ldots, a_n A_n)$ iff $Y = a_1^{w_1} \cdots a_n^{w_n} \Lambda(\omega; \mathbb{A})$.

(*ii*) implies (*iii*): By definition of the Thompson metric, $A_i \leq e^{d(A_i, B_i)} B_i$ and $B_i \leq e^{d(A_i, B_i)} A_i$ for all $i = 1, \ldots, n$. By joint homogeneity and the monotonicity of Λ (*Theorem 4.3*),

$$\begin{aligned} \Lambda(\omega; A_1, \ldots, A_n) &\leq \Lambda\left(\omega; e^{d(A_1, B_1)} B_1, \ldots, e^{d(A_n, B_n)} B_n \right) \\ &= e^{\sum_{i=1}^n w_i d(A_i, B_i)} \Lambda(\omega; B_1, \ldots, B_n) \end{aligned}$$

and similarly $\Lambda(\omega; B_1, \ldots, B_n) \leq e^{\sum_{i=1}^n w_i d(A_i, B_i)} \Lambda(\omega; A_1, \ldots, A_n)$. This implies that $d(\Lambda(\omega; \mathbb{A}), \Lambda(\omega; \mathbb{B})) \leq \sum_{i=1}^n w_i d(A_i, B_i)$.

(*iii*) implies (*iv*): For $t \in (0, 1)$, it follows from *Lemma 2.2*, (*iii*), and the hypothesis that the map $f_t : \mathbb{P} \to \mathbb{P}$ defined by $f_t(X) = \Lambda(\omega; X \#_t A_1, \ldots, X \#_t A_n)$ is a strict contraction for the Thompson metric and hence has a unique fixed point on $\mathbb{P}$, i.e., $X = \Lambda(\omega; X \#_t A_1, \ldots, X \#_t A_n)$ has a unique solution in $\mathbb{P}$. ∎

Remark 6.4. In the light of *Theorem 6.3*, it is more natural to redefine the Karcher mean $\Lambda(\omega; A_1, \ldots, A_n)$ to be the unique solution of the corresponding Karcher equation. That was certainly our motivation in naming it the Karcher mean from the beginning.

By *Corollary 5.4*, $\Lambda^*(\omega; \mathbb{A}) = \lim_{t \to 0^-} P_t(\omega; \mathbb{A})$ also satisfies the same Karcher equation as $\Lambda(\omega; \mathbb{A})$, and hence by the uniqueness of solution, the two are equal. This yields the following corollary.

Corollary 6.5. *For a weight* $\omega = (w_1, \ldots w_n) \in \Delta_n$ *and* $\mathbb{A} = (A_1, \ldots, A_n) \in \mathbb{P}^n$, $\Lambda^*(\omega; \mathbb{A}) = \Lambda(\omega; \mathbb{A})$, *and thus* $\Lambda(\omega; \mathbb{A}) = \lim_{t \to 0} P_t(\omega; \mathbb{A})$.

We gather together our results about the fundamental properties of the Karcher mean.

Theorem 6.6. *For a weight* $\omega=(w_1,\ldots w_n)\in\Delta_n$ *and* $\mathbb{A}=(A_1,\ldots,A_n)\in\mathbb{P}^n$, *the following properties hold*:

(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* $B_i\le A_i$ *for all* $1\le i\le n$, *then* $\Lambda(\omega;\mathbb{B})\le\Lambda(\omega;\mathbb{A})$;

(P5) (*continuity*) $d(\Lambda(\omega;\mathbb{A}),\Lambda(\omega;\mathbb{B}))\le\sum_{i=1}^n w_i d(A_i,B_i)$, *d the Thompson metric*;

(P6) (*congruence invariance*) $\Lambda(\omega;M^*\mathbb{A}M)=M^*\Lambda(\omega;\mathbb{A})M$ *for any invertible* M;

(P7) (*joint concavity*) $\Lambda(\omega;\lambda\mathbb{A}+(1-\lambda)\mathbb{B})\ge\lambda\Lambda(\omega;\mathbb{A})+(1-\lambda)\Lambda(\omega;\mathbb{B})$;

(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*) $\left(\sum_{i=1}^n w_iA_i^{-1}\right)^{-1}\le\Lambda(\omega;\mathbb{A})\le\sum_{i=1}^n w_iA_i$.

Proof: By *Theorem 4.3* and *Theorem 6.3*, Λ is jointly homogeneous and satisfies (P5). Property (P8) follows from *Theorem 4.1*, the preceding corollary and the fact that $P_{-t}(\omega;\mathbb{A})=P_t(\omega;\mathbb{A}^{-1})^{-1}$ for $t>0$. The remaining properties appeared in *Theorem 4.3*. ∎

Remark 6.7. The Karcher mean is uniquely determined by congruence invariancy (P6), self-duality (P8), and the following property:

$$(Y)\ \sum_{i=1}^n w_i\log A_i\le 0 \text{ implies } \Lambda(\omega;A_1,\ldots,A_n)\le I$$

for all $\omega\in\Delta_n$ and $(A_1,\ldots,A_n)\in\mathbb{P}^n$. For a finite-dimensional Hilbert space, property (Y) and the previous characterization for the Karcher mean appear in refs. 23 and 13, respectively. The proofs are similar to those of refs. 23 and 13.

7. Hilbert–Schmidt Operators

In this section, we briefly sketch from ref. 24 a distinctly different approach to the infinite-dimensional Karcher mean that connects it more closely to the least-squares mean of the finite-dimensional setting.

Let HS(H) denote the bilateral ideal of Hilbert–Schmidt operators of $\mathcal{B}(H)$, the algebra of bounded linear operators on a complex Hilbert space H. Recall that 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) given by the inner product

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

Our focus is on 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 with the restricted inner product becomes a real Hilbert space, and on its positive part $\Sigma=\mathbb{P}\cap\mathcal{H}_\mathbb{R}$, the open subcone of positive definite operators in $\mathcal{H}_\mathbb{R}$. We note that $\lambda>0$ is a necessary condition for membership in Σ.

We define a Riemannian metric on Σ by identifying $T\Sigma$ with $\Sigma\times\mathcal{H}_\mathbb{R}$, and endowing the tangent space at $A\in\Sigma$ with the Hilbert metric

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

We note that $\|X\|_A=\langle X,X\rangle_A^{\frac{1}{2}}=\|A^{-1/2}XA^{-1/2}\|_2$.

The structure of the Riemannian manifold Σ closely parallels that of the finite-dimensional Riemannian manifolds of positive definite matrices equipped with the Riemannian trace metric, as has been worked out by Larotonda (25). In particular, Σ is a Riemannian manifold of nonpositive curvature, and its distance metric δ is an NPC metric (Eq. **1.1**). Hence the weighted least-squares minimizer

$$g_n(\omega;A_1,\ldots,A_n)=\underset{X\in\mathbb{P}}{\arg\min}\sum_{i=1}^n w_i\delta^2(X,A_i)$$

uniquely exists. Using methods of Riemannian geometry and Karcher's result (ref. 7, theorem 1.2) or a more operator-theoretic approach (24), one can show the least-squares mean is the unique point where the gradient of the least-squares objective function $f(X)=\sum_{i=1}^n w_i\delta^2(X,A_i)$ vanishes, which leads to the alternative characterization of the least-squares mean as the unique solution to the Karcher equation, which (up to a scalar multiple) arises from setting the gradient equal to 0. We summarize:

Theorem 7.1. The ω-weighted least-squares mean of $(A_1,\ldots,A_n)\in\Sigma^n$ in the Riemannian manifold Σ is the Karcher mean $\Lambda(\omega;A_1,\ldots,A_n)$, the solution of the Karcher equation (Eq. **1.2**).

It follows from Theorem 7.1 that the least-squares mean in Σ is the restriction of the Karcher mean on $\mathbb{P}$ to Σ. We now present a reverse construction: extending the least-squares mean, equivalently the restricted Karcher mean, on Σ to $\mathbb{P}$. Let $\{\alpha\}_{\alpha\in\Delta}$ denote the collection of nonzero finite-dimensional subspaces of H ordered by inclusion, a directed family. Let $P_\alpha:H\to H$ denote the orthogonal projection onto the subspace α. We note that each P_α is hermitian, positive semidefinite, idempotent, and has finite rank, hence is Hilbert–Schmidt. We view $\{P_\alpha:\alpha\in\Delta\}$ as a monotonically increasing net indexed by Δ that strongly converges to its supremum, the identity I, because for any $x\in H$, $P_\alpha(x)=x$ for all large enough α.

Because $\{P_\alpha:\alpha\in\Delta\}$ is bounded, the net $\{P_\alpha AP_\alpha\}$ 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.)

Proposition 7.2. *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* $e^{-m}I<A_i<e^mI$ *for* $1\le i\le n$. *Then*

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

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* $\Lambda(\omega;A_1,\ldots,A_n)$.

Proof: Because the net $\{P_\alpha(e^mI-A_i)P_\alpha\}_\alpha$ is a monotonically increasing net strongly converging to its supremum e^mI-A_i, the net $\{e^mI-P_\alpha(e^mI-A_i)P_\alpha\}_\alpha$ is a decreasing net strongly converging to its infimum $e^mI-(e^mI-A_i)=A_i\ge e^{-m}I$. By the idempotency and monotonicity [property (P4) of *Theorem 6.6*] of the Karcher 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 7.1*, each X_α satisfies the Karcher equation:

$$\sum_{i=1}^n\omega_i\log\left(X_\alpha^{1/2}B_{i,\alpha}^{-1}X_\alpha^{1/2}\right)=0,$$

where $B_{i,\alpha}=e^mI-P_\alpha(e^mI-A_i)P_\alpha$. Because (by the previous paragraph) $B_{i,\alpha}$ converges strongly to A_i and because the function

MATHEMATICS

$f(X) = \sum_{i=1}^{n} w_i \log(X^{1/2}A_i^{-1}X^{1/2})$ is strongly continuous on the bounded order interval $[e^{-m}I, e^m I]$ by *Lemma 5.2*, we conclude that

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

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

Remark 7.3. We note that *Proposition 7.2* provides an algorithm of sorts for approximating the Karcher mean. One picks finite-dimensional subspaces α of increasing dimension, computes on α the restriction of the least-squares mean given in the proposition, and uses these as approximations.

From the proof of *Proposition 7.2*, one extracts the following special case of strong continuity of Λ.

Corollary 7.4. *Let $\mathbb{A}_\alpha = (A_{1,\alpha}, \ldots, A_{n,\alpha})$ be a decreasing respectively (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 $\Lambda(\omega; \mathbb{A}_\alpha)$ is a decreasing resp. increasing net strongly converging to $\Lambda(\omega; \mathbb{A})$.*

Remark 7.5. 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 7.4*, 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.

It remains an open question whether $\Lambda : \mathbb{P}^n \to \mathbb{P}$ is strongly continuous.

8. Subalgebras

For convenience and ease of presentation, we have limited our considerations to the full algebra $\mathcal{B}(H)$ of bounded linear operators. However, we observe that the constructions can be carried out in large classes of subalgebras (which we assume always to contain the identity I). For any norm-closed C^* -subalgebra $\mathcal{A}$ of $\mathcal{B}(H)$, $\mathbb{P}_{\mathcal{A}} = \mathcal{A} \cap \mathbb{P}$ will be its open cone of positive operators, and will be closed under the operation of taking weighted geometric means $A \#_t B$. Hence it will be closed under taking power means $P_t(\omega; \mathbb{A})$, because the power mean is the limit in the Thompson metric of a contractive map defined from the weighted geometric means on $\mathbb{P}_{\mathcal{A}}$, and because the topology of the Thompson metric agrees with the relative operator norm topology. Because we have defined $\Lambda(\omega; \mathbb{A})$ to be the strong limit of the monotonically decreasing family $\{P_t : t > 0\}$, we need that the subalgebra is monotone complete (actually, monotone σ-complete will suffice because one can restrict to $t = 1/n$ and obtain the same infimum). Once one has closure under the Karcher mean for the subalgebra, then one sees readily that its properties that we have derived for the full algebra $\mathcal{B}(H)$ are inherited by the subalgebra, in particular its characterization as the unique solution of the corresponding Karcher equation. Because the von Neumann subalgebras are strongly closed, they in particular have Karcher means defined in the manner of this paper and satisfying the properties derived for it.

ACKNOWLEDGMENTS. We express our gratitude to Prof. Dick Kadison for helpful insights about the strong topology, especially for his pointing us to ref. 21. The work of Y.L. was supported by National Research Foundation of Korea Grant 2012-005191 funded by the Korean government (Ministry of Education, Science and Technology).

1. Kubo F, Ando T (1980) Means of positive linear operators. *Math Ann* 246(3):205–224.
2. Lawson J, Lim Y (2001) The geometric mean, matrices, metrics, and more. *Am Math Mon* 108(9):797–812.
3. Bhatia R (2007) *Positive Definite Matrices*. Princeton Series in Applied Mathematics (Princeton Univ Press, Princeton).
4. Moakher M (2005) A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM J Matrix Anal Appl* 26(3):735–747.
5. Bhatia R, Holbrook J (2006) Riemannian geometry and matrix geometric means. *Linear Algebra Appl* 413(2-3):594–618.
6. Berger M (2003) *A Panoramic View of Riemannlan Geometry* (Springer, Berlin).
7. Karcher H (1977) Riemannian center of mass and mollifier smoothing. *Commun Pure Appl Math* 30(5):509–541.
8. Ando T, Li CK, Mathias R (2004) Geometric means. *Linear Algebra Appl* 385:305–334.
9. Lawson J, Lim Y (2011) Monotonic properties of the least squares mean. *Math Ann* 351(2):267–279.
10. Bridson M, Haefliger A (1999) *Metric Spaces of Non-Positive Curvature* (Springer, Berlin).
11. Sturm KT (2003) Probability measures on metric spaces of nonpositive curvature. *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces*. Contemporary Mathematics, eds Auscher P, Coulhon T, Grigor'yan A (American Mathematical Society, Providence, RI), Vol 338, pp 357–390.
12. Lang S (1999) *Fundamentals of Differential Geometry*. Graduate Texts in Mathematics (Springer, Berlin).
13. Lim Y, Pálfia M (2012) The matrix power means and the Karcher mean. *J Funct Anal* 262(4):1498–1514.
14. Corach G, Porta H, Recht L (1994) Convexity of the geodesic distance on spaces of positive operators. *Illinois J Math* 38(1):87–94.
15. Lawson J, Lim Y (2007) Metric convexity of symmetric cones. *Osaka J Math* 44(4): 795–816.
16. Thompson AC (1963) On certain contraction mappings in a partially ordered vector space. *Proc Am Math Soc* 14:438–443.
17. Nussbaum RD (1994) Finsler structures for the part metric and Hilbert's projective metric and applications to ordinary differential equations. *Differ Integral Equ* 7(5-6):1649–1707.
18. Bhatia R (2003) On the exponential metric increasing property. *Linear Algebra Appl* 375:211–220.
19. Lawson J, Lim Y (2008) A general framework for extending means to higher orders. *Colloq Math* 113(2):191–221.
20. Weidmann J (1980) *Linear Operators in Hilbert Spaces* (Springer, Berlin).
21. Kadison R (1968) Strong continuity of operator functions. *Pac J Math* 26:121–129.
22. Lawson J, Lim Y (2013) Karcher means and Karcher equations of positive definite operators. *Trans Am Math Soc*, in press.
23. Yamazaki T (2012) The Riemannian mean and matrix inequalities related to the Ando-Hiai inequality and chaotic order. *Oper Matrices* 6(3):577–588.
24. Lawson J, Lim Y (2013) The least squares mean of positive Hilbert-Schmidt operators. *J Math Anal Appl* 403(2):365–375.
25. Larotonda G (2007) Nonpositive curvature: A geometric approach to Hilbert-Schmidt operators. *Differ Geom Appl* 25(6):679–700.