Math. Ann. (2011) 351:267–279
DOI 10.1007/s00208-010-0603-6

Mathematische Annalen

Monotonic properties of the least squares mean

Jimmie Lawson · Yongdo Lim

Received: 27 July 2010 / Revised: 8 October 2010 / Published online: 16 November 2010

Abstract We settle an open problem of several years standing by showing that the least squares mean for positive definite matrices is monotone for the usual (Loewner) order. Indeed we show this is a special case of its appropriate generalization to partially ordered complete metric spaces of nonpositive curvature. Our techniques extend to establish other basic properties of the least squares mean such as continuity and joint concavity. Moreover, we introduce a weighted least squares mean and derive our results in this more general setting.

Mathematics Subject Classification (2000) 15A48

1 Introduction

Not only does the study of positive definite matrices remain a flourishing area of mathematical investigation (see e.g., the recent monograph of Bhatia [6] and references therein), but positive definite matrices have become fundamental computational objects in many applied areas. They appear as covariance matrices in statistics, as elements of the search space in convex and semidefinite programming, as kernels in machine learning, as density matrices in quantum information, and as diffusion tensors in medical imaging, to cite a few. A variety of metric-based computational algorithms for positive definite matrices have arisen for approximations, interpolation, filtering,

J. Lawson (✉)
Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803, USA
e-mail: lawson@math.lsu.edu

Y. Lim
Department of Mathematics, Kyungpook National University, Taegu 702-701, Korea
e-mail: ylim@knu.ac.kr

estimation, and averaging, the last being the concern of this paper. In recent years, it has been increasingly recognized that the Euclidean distance is often not the most suitable for the space $\mathbb{P}$ of positive definite matrices and that working with the appropriate geometry does matter in computational problems. It is thus not surprising that there has been increasing interest in the trace metric, the distance metric arising from the natural Riemannian structure on $\mathbb{P}$ making it a Riemannian manifold, indeed a symmetric space, of negative curvature. (Recall the trace metric distance between two positive definite matrices is given by $\delta(A, B) = (\sum_{i=1}^{k} \log^2 \lambda_i(A^{-1}B))^{\frac{1}{2}}$, where $\lambda_i(X)$ denotes the ith eigenvalue of X in non-decreasing order.) Recent contributions that have advocated the use of this metric in applications include [12,24,28] for tensor computation in medical imaging and [4] for radar processing.

Since the pioneering paper of Kubo and Ando [16], 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 $\mathfrak{G}_2(A, B) = A\#B := A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}$ is the metric midpoint of A and B for the trace metric (see, e.g., [6,18]), it is natural to use an averaging technique over this metric to extend this mean to a larger number of variables. First Moakher [23] and then Bhatia and Holbrook [7,8] 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:

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

This idea had been anticipated by Élie Cartan (see, for example, section 6.1.5 of [5]), who showed among other things 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 $\mathbb{P}$.

Another approach, independent of metric notions, was suggested by Ando, Li, and Mathias [2] via a "symmetrization procedure" and induction. The Ando–Li–Mathias paper was also important for listing, and deriving for their mean, ten desirable properties for extended geometric means $g : \mathbb{P}^n \to \mathbb{P}$ that one might anticipate from properties of the two-variable geometric mean, where $\mathbb{P} = \mathbb{P}_m$ denotes the convex cone of $m \times m$ positive definite Hermitian matrices equipped with the Loewner order $\leq$. The Ando–Li–Mathias mean proved to be computationally cumbersome, and Bini, Meini, and Poloni [9] suggested an alternative with more rapid convergence properties, which also satisfies the ten axioms. One notes in particular that while the axioms characterize the two-variable case, this is no longer true in the n-variable case, $n > 2$.

The ten properties may be generalized to the setting of weighted geometric means. We recall that the two-variable weighted geometric mean is given by

$$t \mapsto \mathfrak{G}_2(1-t, t; A, B) = A\#_t B =: A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2},$$

which is a geodesic parametrization of the unique geodesic passing through A and B for $A \neq B$. A weighted geometric mean of n-positive definite matrices should be defined for each weight, where the weights $\mathbf{w} = (w_1, \ldots, w_n)$ vary over Δ_n, the simplex of positive probability vectors convexly spanned by the unit coordinate vectors. We define a *weighted geometric mean* of n positive definite matrices to be a map $g : \Delta_n \times \mathbb{P}^n \to \mathbb{P}$ satisfying the following properties:

(P1) (Consistency with scalars) $g(\mathbf{w}; \mathbb{A}) = A_1^{w_1} \ldots A_n^{w_n}$ if the A_i's commute;
(P2) (Joint homogeneity) $g(\mathbf{w}; a_1 A_1, \ldots, a_n A_n) = a_1^{w_1} \ldots a_n^{w_n} g(\mathbf{w}; \mathbb{A})$;
(P3) (Permutation invariance) $g(\mathbf{w}_\sigma; \mathbb{A}_\sigma) = g(\mathbf{w}; \mathbb{A})$, where $\mathbf{w}_\sigma = (w_{\sigma(1)}, \ldots, w_{\sigma(n)})$;
(P4) (Monotonicity) If $B_i \leq A_i$ for all $1 \leq i \leq n$, then $g(\mathbf{w}; \mathbb{B}) \leq g(\mathbf{w}; \mathbb{A})$;
(P5) (Continuity) The map $g(\mathbf{w}; \cdot)$ is continuous;
(P6) (Congruence invariance) $g(\mathbf{w}; M^*\mathbb{A}M) = M^* g(\mathbf{w}; \mathbb{A}) M$ for any invertible M;
(P7) (Joint concavity) $g(\mathbf{w}; \lambda\mathbb{A} + (1 - \lambda)\mathbb{B}) \geq \lambda g(\mathbf{w}; \mathbb{A}) + (1 - \lambda) g(\mathbf{w}; \mathbb{B})$ for $0 \leq \lambda \leq 1$;
(P8) (Self-duality) $g(\mathbf{w}; A_1^{-1}, \ldots, A_n^{-1})^{-1} = g(\mathbf{w}; A_1, \ldots, A_n)$;
(P9) (Determinant identity) $\mathrm{Det}\, g(\mathbf{w}; \mathbb{A}) = \prod_{i=1}^n (\mathrm{Det} A_i)^{w_i}$; and
(P10) (AGH weighted mean inequalities) $(\sum_{i=1}^n w_i A_i^{-1})^{-1} \leq g(\mathbf{w}; \mathbb{A}) \leq \sum_{i=1}^n w_i A_i$.

We note that the two-variable weighted geometric mean $\mathfrak{G}_2(1 - t, t; A, B) = A \#_t B$, $t \in [0, 1]$, satisfies $(P1)-(P10)$, and that the original ten properties for the unweighted case arise by specializing to the weight $(1/n, \ldots, 1/n)$.

In their study of the unweighted least squares mean, Moakher [23] and Bhatia and Holbrook [7,8] have derived some of the axiomatic properties (P1)–(P10) satisfied by the Ando–Li–Mathias geometric mean: consistency with scalars, joint homogeneity, permutation invariance, congruence invariance, and self-duality (the last two being true since congruence transformations and inversion are isometries). Further, based on computational experimentation, Bhatia and Holbrook conjectured monotonicity for the least squares mean (problem 19 in "Open problems in matrix theory" by Zhan [29]). Providing a positive solution (Corollary 2) to this conjecture was the original motivation for this paper.

In this paper we introduce the *weighted least squares mean* $\mathfrak{G}_n(\mathbf{w}; A_1, \ldots, A_n)$ of $(A_1, \ldots, A_n)$ with the weight $\mathbf{w} = (w_1, \ldots, w_n) \in \Delta_n$, which is defined to be

$$\mathfrak{G}_n(\mathbf{w}; A_1, \ldots, A_n) = \underset{X \in \mathbb{P}}{\arg\min} \sum_{i=1}^n w_i \delta^2(X, A_i). \tag{1}$$

Computing appropriate derivatives as in [6,23] yields that the weighted least squares mean coincides with the unique positive definite solution of the equation

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

It is not difficult to see from (1) and (2) and some elementary facts about matrices and the trace metric that the weighted least squares mean satisfies $(P1)-(P3)$, $(P6)$, $(P8)$ and $(P9)$. In this paper we show that the weighted least squares mean satisfies all the properties $(P1)-(P10)$ by verifying all the additional properties (P4), (P5), (P7), and (P10). As far as we know, this is the first verification of properties (P4) and (P7) in both the weighted and unweighted cases and of (P10) in the weighted case, the unweighted case having been shown by Yamazaki [27]. We thus see that the (weighted) least squares mean provides another important example of a (weighted) geometric mean. We further show that the weighted least squares mean is non-expansive: $\delta(\mathfrak{G}_n(\mathbf{w}; A_1, \ldots, A_n), \mathfrak{G}_n(\mathbf{w}; B_1, \ldots, B_n)) \leq \sum_{i=1}^{n} w_i \delta(A_i, B_i)$.

The main tools of the paper involve the theory of nonpositively curved metric spaces and techniques from probability and random variable theory and the recent combination of the two, particularly by Sturm [26]. Not only are these tools crucial for our developments, but also, we believe, significantly enhance the potential usefulness of the least squares mean.

2 Metric spaces and means

The setting appropriate for our considerations is that of *globally nonpositively curved metric spaces*, which we call *NPC spaces* for short (since we do not consider the locally nonpositively curved spaces). These are complete metric spaces M such that for each $x, y \in M$, there exists an $m \in M$ satisfying

$$d^2(m, z) \leq \frac{1}{2}d^2(x, z) + \frac{1}{2}d^2(y, z) - \frac{1}{4}d^2(x, y) \tag{3}$$

for all $z \in M$. Such spaces are also called (global) CAT(0)-spaces or Hadamard spaces. The theory of such spaces is quite extensive; see, e.g., [3,10,14,26]. In particular the m appearing in (3) is the unique metric midpoint between x and y. By inductively choosing midpoints for dyadic rationals and extending by continuity, one obtains for each $x \neq y$ a unique metric *minimal geodesic* $\gamma : [0, 1] \to M$ satisfying $d(\gamma(t), \gamma(s)) = |t - s|d(x, y)$. We denote $\gamma(t)$ by $x\#_t y$ and call it the *t-weighted mean* of x and y. The midpoint $x\#_{1/2}y$ we denote simply as $x\#y$. We remark that by uniqueness $x\#_t y = y\#_{1-t}x$; in particular, $x\#y = y\#x$.

Remark 1 Equation (3) is sometimes referred to as the *semiparallelogram law*, since it can derived from the parallelogram law in Hilbert spaces by replacing the equality with an inequality (see [18]). It is satisfied by the length metric in any simply connected nonpositively curved Riemannian manifold [17]. Hence the metric definition represents a metric generalization of nonpositive curvature. The trace metric on the Riemannian symmetric space of positive definite matrices is a particular example [17,18].

Equation (3) admits a more general formulation in terms of the weighted mean (see e.g. [26, Proposition 2.3]). For all $0 \leq t \leq 1$ we have

$$d^2(x\#_t y, z) \leq (1 - t)d^2(x, z) + td^2(y, z) - t(1 - t)d^2(x, y). \tag{4}$$

An *n-mean* on a set X is a function $\mu : X^n \to X$ satisfying the idempotency law $\mu(x, x, \ldots, x) = x$. It is *symmetric* if it is invariant under all permutations σ

of $\{1, \ldots, n\}$, i.e., $\mu(x_1, \ldots, x_n) = \mu(x_{\sigma(1)}, \ldots, x_{\sigma(n)})$. For a metric space X with weighted mean, the operation $x\#_t y$ is a 2-mean for each t. A special case is the midpoint mean $x\#y$ for $t = 1/2$, which is symmetric.

The problem of extending the geometric mean of two positive definite matrices to an n-variable mean for $n \geq 3$ generalizes to the setting of metric spaces with unique midpoints. Under appropriate metric hypotheses, all of which are implied by the NPC condition, the symmetrization procedure applies and inductively yields n-means extending $\mu(x, y) = x\#y$ for each $n \geq 3$; see Es-Sahib and Heinich [11] and the authors [19]. (Recall that the symmetrization procedure extends a k-variable mean μ_k to a $k+1$-variable mean μ_{k+1} defined by $\mu_{k+1}(x_1, \ldots, x_{k+1}) = y$ if $y = \lim_{n\to\infty} x_i^n$ for each $1 \leq i \leq k+1$, where inductively $x_i^{n+1} = \mu_k(x_1^n, \ldots, x_{i-1}^n, x_{i+1}^n, \ldots, x_k^n), \quad x_i^0 = x_i$.) Extension methods for weighted 2-means and for the mean of Bini, Meini, and Poloni [9] also generalize to NPC-spaces, and even weaker metric settings [21].

The weighted least squares mean can be immediately formulated in any metric space (M, d). Given $(a_1, \ldots, a_n) \in M^n$, and positive real numbers $w_1, \ldots, w_n$ summing to 1, we define

$$\mathfrak{G}_n(w_1, \ldots, w_n; a_1, \ldots, a_n) := \arg\min_{z \in M} \sum_{i=1}^{n} w_i d^2(z, a_i), \tag{5}$$

provided the minimizer exists and is unique. In general the minimizer may fail to exist or fail to be unique, but existence and uniqueness always holds for NPC spaces as can be readily deduced from the uniform convexity of the metric; see [26, Propositions 1.7, 4.3]. Note that the mean in (5) is permutation invariant in the sense of property (P3) for weighted means given in the Introduction. By taking $w_i = 1/n$ for each $i = 1, \ldots, n$, we see that the unweighted least squares mean is a special case of the weighted one, so we work with the weighted case in what follows. Although this mean is sometimes referred to as the Karcher mean in light of its appearance in his work on Riemannian manifolds [15], we will refer to it as the *weighted least squares mean*, or simply as the *least squares mean*.

One other mean will play an important role in what follows, one that we shall call the *inductive mean* following the terminology of [26], although it appeared earlier in [1,25]. It is defined inductively for NPC spaces (or more generally for metric spaces with weighed means $x\#_t y$) for each $k \geq 2$ by $S_2(x, y) = x\#y$ and for $k \geq 3$, $S_k(x_1, \ldots, x_k) = S_{k-1}(x_1, \ldots, x_{k-1})\#_{\frac{1}{k}} x_k$.

3 Random variables and barycenters

In recent years significant portions of the classical theory of real-valued random variables on a probability space have been successfully generalized to the setting in which the random variables take values in a metric space M. We quickly recall some of this theory as worked out, for example, by Es-Sahib and Heinich [11] and particularly by Sturm [26].

Let $(\Omega, \mathcal{A}, P)$ be a probability space: a set Ω equipped with a σ-algebra $\mathcal{A}$ of subsets, and a σ-additive probability measure P on $\mathcal{A}$. We write the measure or probability of $A \in \mathcal{A}$ by $P(A)$. For a metric space (M, d), an *M-valued random variable* is a function $X : \Omega \to M$ which is measurable in the sense that $X^{-1}(B) \in \mathcal{A}$ for every Borel subset B of M. We further impose the technically useful assumption that the image $X(\Omega)$ is a separable subset of M.

The push-forward of the measure P by X is denoted and defined by $q_X(B) = P(X^{-1}(B))$ for each Borel subset B of M. It is a probability measure on the Borel sets of M and is called the *distribution* of X. A sequence of random variables $\{X_n\}$ is *identically distributed* (i.d.) if all have the same distribution. For any q_X-integrable function $\phi : M \to \mathbb{R}$, one has the basic formula $\int_M \phi\, dq_X = \int_\Omega \phi X\, dP$.

A collection of random variables $\{X_i : i \in I\}$ is *independent* if for every finite $F \subseteq I$, $P(\bigcap_{i \in F} X_i^{-1}(B_i)) = \prod_{i \in F} P(X_i^{-1}(B_i))$, where $\{B_i : i \in I\}$ is any collection of Borel subsets of M. A sequence $\{X_n\}$ is i.i.d. if it is both independent and identically distributed.

Assume henceforth that M is an NPC-space. Let $\mathcal{P}(M)$ denote the set of probability measures with separable support on $(M, \mathcal{B}(M))$, where $\mathcal{B}(M)$ is the collection of Borel sets. We define the collection $\mathcal{P}^1(M)$, resp. $\mathcal{P}^2(M)$, of probability measures $q \in \mathcal{P}(M)$ to be those satisfying $\int_M d(z, x)q(dx) < \infty$, resp. $\int_M d^2(z, x)q(dx) < \infty$, for some (hence all) $z \in M$. Members of $\mathcal{P}^1(M)$ are called *integrable* and those in $\mathcal{P}^2(M)$ are called *square integrable*. We define a random variable $X : \Omega \to M$ to be in L^1, resp. L^2, if its distribution is integrable resp. square integrable. In particular, it is integrable if $\int_\Omega d(z, X(\omega))P(d\omega) = \int_M d(z, x)q_X(dx) < \infty$ for $z \in M$.

Following Sturm [26], we define the *barycenter* $b(q)$ of $q \in \mathcal{P}^1(M)$ by

$$b(q) = \underset{z \in M}{\arg\min} \int_M [d^2(z, x) - d^2(y, x)]q(dx). \tag{6}$$

Sturm uses the uniform convexity of $z \mapsto d^2(z, x)$ to show that independently of y there is a unique $z = b(q)$, the barycenter (by definition), at which this minimum is obtained [26, Proposition 4.3], and that for the case that q is square integrable the barycenter can be alternatively characterized by

$$b(q) = \underset{z \in M}{\arg\min} \int_M d^2(z, x)q(dx). \tag{7}$$

Remark 2 For the case that $q = \sum_{i=1}^n w_i \delta_{x_i}$, where $(w_1, \ldots, w_n)$ is a weight and δ_{x_i} is the point mass at x_i, we have

$$\begin{aligned} b(q) &= \underset{z \in M}{\arg\min} \int_M d^2(z, x)q(dx) \\ &= \underset{z \in M}{\arg\min} \sum_{i=1}^n w_i d^2(z, x_i) = \mathfrak{G}_n(w_1, \ldots, w_n; x_1, \ldots, x_n). \end{aligned}$$

Thus in this case q is square integrable and its barycenter $b(q)$ agrees with the weighted least squares mean of $(x_1, \ldots, x_n)$.

For $X : \Omega \to M$ integrable, we define its *expected value* EX by

$$EX = \underset{z \in M}{\arg\min} \int\limits_{\Omega} \left[d^2(z, X(\omega)) - d^2(y, X(\omega)) \right] P(d\omega)$$
$$= \underset{z \in M}{\arg\min} \int\limits_{M} [d^2(z, x) - d^2(y, x)] q_X(dx) = b(q_X). \tag{8}$$

From this definition it is clear that integrable i.d. random variables have the same expectation.

It is also possible to define and prove notions of a Law of Large Numbers for a sequence of i.i.d. random variables into a metric space M. Let $\{X_k : k \in \mathbb{N}\}$ be a sequence of i.i.d. random variables on some probability space $(\Omega, \mathcal{A}, P)$ into M. Let μ_k be an k-mean on M for each k, for example the inductive mean or one obtained by the symmetrization procedure. We use these means to form the "average" Y_k of the given random variables according to the rule $Y_k(\omega) := \mu_k(X_1(\omega), \ldots, X_k(\omega))$. Now under suitable hypotheses Sturm [26] and Es-Sahib and Heinich [11] showed that a strong law of large numbers is satisfied, that is, the Y_k converge pointwise a.e. to a common point b. The principal result of Sturm [26, Theorem 4.7] is crucial for our purposes.

Theorem 1 *Let $\{X_k\}_{k \in \mathbb{N}}$ be a sequence of bounded i.i.d. random variables from a probability space $(\Omega, \mathcal{A}, P)$ into an NPC space M. Let S_k denote the inductive mean for each $k \geq 2$, and set $Y_k(\omega) = S_k(X_1(\omega), \ldots, X_k(\omega))$. Then $Y_k(\omega) \to EX_1$ as $k \to \infty$ for almost all $\omega \in \Omega$.*

Theorem 1 provides an essential tool for our study of the least squares mean. Many properties of the weighted 2-means of an NPC space can be shown to extend to their finite iterations S_n, and then shown to be preserved in passing to the limit, the expectation. By Remark 2 and the equation following it, the expectation is the weighted least squares mean if the push-forward measure q_{X_1} agrees with a given weighted finitely supported measure. In this way we are able to deduce properties of the weighted least squares mean from properties of the weighted 2-mean.

4 A basic construction

In this section we specialize Theorem 1 to the case of finitely supported probability measures, the case of interest to us. We first recall a standard construction in probability theory. Let $(\Omega, P) = \prod_{k=1}^{\infty} (\Omega_k, P_k)$ be the infinite product probability space, where $\Omega_k = \{\xi_1, \ldots, \xi_n\}$ and $P_k = \sum_{i=1}^{n} w_i \delta_{\xi_i}$ for all k. Define a sequence of random variables $X_k : \Omega \to M$ by $X_k(\omega) = x_i$ if the kth component of $\omega \in \Omega$ is ξ_i. Then $\{X_k\}$ is i.i.d. and the distribution of X_k is $\sum_{i=1}^{n} w_i \delta_{x_i}$.

We define $Y_k : \Omega \to M$ for each k by $Y_k(\omega) = S_k(X_1(\omega), \dots, X_k(\omega))$, where S_k is the inductive mean. By Theorem 1 we have that $\lim_{k\to\infty} Y_k(\omega) = EX_1 = b(q_{X_1})$ a.e. From Remark 2 it follows that that $\lim_{k\to\infty} Y_k(\omega) = \mathfrak{G}_n(\mathbf{w}; x_1, \dots, x_n)$ a.e. We summarize this special case of Theorem 1.

Corollary 1 *Let (M, d) be an NPC space, let $\{x_1, \dots, x_n\} \subseteq M$, and let $\mathbf{w} = (w_1, \dots, w_n)$ be a weight. Then $\lim_{k\to\infty} Y_k(\omega) = \mathfrak{G}_n(\mathbf{w}; x_1, \dots, x_n)$ a.e. for the $\{Y_k\}$ given in the preceding construction.*

We consider a basic example, which will be used in Sect. 6.

Proposition 1 *Let $\mathcal{H}$ be a Hilbert space endowed with the metric induced by the inner product. Then*

(i) *$\mathcal{H}$ is an NPC space.*
(ii) *The binary t-weighted mean of x and y is given by $(1-t)x + ty$.*
(iii) *The inductive mean is given by $S_k(x_1, \dots, x_k) = \sum_{i=1}^{k}(1/k)x_i$.*
(iv) *The weighted least squares mean for weight $\mathbf{w} = (w_1, \dots, w_n)$ is given by $\mathfrak{G}_n(\mathbf{w}; x_1, \dots, x_n) = \sum_{i=1}^{n} w_i x_i$.*
(v) *For $\{X_k\}_{k\in\mathbb{N}}$ and a weight $\mathbf{w} = (w_1, \dots, w_n)$ as given in the preceding construction, we have*

$$\lim_{k\to\infty} \sum_{i=1}^{k}(1/k)X_i(\omega) \to \mathfrak{G}_n(\mathbf{w}; x_1, \dots, x_n) = \sum_{i=1}^{n} w_i x_i \text{ a.e.}$$

Proof (i) It is standard that Hilbert spaces satisfy the parallelogram law, hence the semiparallelogram law (3), and hence are NPC spaces (see e.g. [26, Proposition 3.5]).

(ii) The map on $[0, 1]$ given by $t \mapsto (1-t)x + ty$ is a metric geodesic taking 0 to x and 1 to y. Since such geodesics are unique in NPC spaces, it must give the t-weighted mean.

(iii) By definition and induction

$$S_k(x_1, \dots, x_k) = \frac{k-1}{k} S_{k-1}(x_1, \dots, x_{k-1}) + \frac{1}{k}x_k$$
$$= \frac{k-1}{k}\sum_{i=1}^{k-1}\frac{1}{k-1}x_i + \frac{1}{k}x_k = \sum_{i=1}^{k}\frac{1}{k}x_i.$$

(iv) Consider the measure $q = \sum_{i=1}^{n} w_i \delta_{x_i}$. Then for any $y \in \mathcal{H}$,

$$\langle \mathfrak{G}_n(\mathbf{w}; x_1, \dots, x_n), y\rangle = \langle b(q), y\rangle = \int_{\mathcal{H}} \langle x, y\rangle q(dx)$$
$$= \sum_{i=1}^{n} w_i \langle x_i, y\rangle = \left\langle \sum_{i=1}^{n} w_i x_i, y \right\rangle,$$

where the first equality follows from Remark 2 and the second is the content of [26, Proposition 5.4]. The conclusion of (iv) is now immediate.

(v) In the earlier construction of this section we have $Y_k(\omega) = \sum_{i=1}^{k} \frac{1}{k} X_i(\omega)$ by part (iii). The conclusion of (v) then follows from Corollary 1 and (iv).

5 Monotonicity and Loewner–Heinz NPC spaces

The fundamental Loewner–Heinz inequality for positive definite matrices asserts that $A^{1/2} \le B^{1/2}$ whenever $A \le B$. This can be written alternatively as $A\#I \le B\#I$ whenever $A \le B$ and extends to the equivalent monotonicity property that $A_1\#A_2 \le B_1\#B_2$ whenever $A_1 \le B_1$ and $A_2 \le B_2$. These considerations motivate the next definition.

Definition 1 A *Loewner–Heinz NPC space* is an NPC space equipped with a closed partial order $\le$ satisfying $x_1\#x_2 \le y_1\#y_2$ whenever $x_i \le y_i$ for $i = 1, 2$. (Recall that a partial order on a topological space X is closed if $\{(x, y) : x \le y\}$ is closed in $X \times X$ equipped with the product topology.)

A mean $\mu : M^n \to M$ on a partially ordered metric space is called *order-preserving* or *monotonic* if $x_i \le y_i$ for $i = 1, \dots, n$ implies $\mu(x_1, \dots, x_n) \le \mu(y_1, \dots, y_n)$.

Lemma 1 *The inductive mean S_k on a Loewner–Heinz NPC space is monotonic for every $k \ge 2$.*

Proof We first observe that $x_1\#_t x_2 \le y_1\#_t y_2$ whenever $x_i \le y_i$ for $i = 1, 2$ by the standard argument of extending the inequality to the dyadic weighted means by induction for the case of the dyadic rationals, and then extending to general $t \in [0, 1]$ by continuity in t and the closedness of the relation $\le$. Assuming that the inductive k-mean S_k is monotonic, it follows that $S_{k+1}(x_1, \dots, x_{k+1}) = S_k(x_1, \dots, x_k)\#_{\frac{1}{k+1}} x_{k+1}$ is monotonic since S_k and the t-weighted mean both are.

Theorem 2 *Let $(M, d, \le)$ be a Loewner–Heinz NPC space. Then for a fixed weight $\mathbf{w} = (w_1, \dots, w_n)$ the weighted least squares mean $\mathfrak{G}_n$ is monotonic for $n \ge 2$.*

Proof Assuming $x_i \le y_i$ for $1 \le i \le n$, we show $\mathfrak{G}_n(\mathbf{w}; x_1, \dots, x_n) \le \mathfrak{G}_n(\mathbf{w}; y_1, \dots, y_n)$, where $\mathfrak{G}_n$ is the least squares mean on M^n. Let Ω_k be a copy of the n-element set $\{\xi_1, \dots, \xi_n\}$ equipped with the measure $P_k = \sum_{i=1}^{n} w_i \delta_{\xi_i}$. Let $(\Omega, P) = \prod_{k=1}^{\infty} (\Omega_k, P_k)$ and X_k be as defined in the beginning of Sect. 4. Similarly we define $\tilde{X}_k : \Omega \to M$ by $\tilde{X}_k(\omega) = y_i$ if $\pi_k(\omega) = \xi_i$. As we have seen in the previous section $\{X_k\}$ is i.i.d. with distribution $\sum_{i=1}^{n} w_i \delta_{x_i}$, while $\{\tilde{X}_k\}$ is i.i.d. with distribution $\sum_{i=1}^{n} w_i \delta_{y_i}$. Finally we note that $(X_1(\omega), \dots, X_k(\omega))$ is coordinatewise less than or equal to $(\tilde{X}_1(\omega), \dots, \tilde{X}_k(\omega))$ since $x_i \le y_i$ for each $i = 1, \dots, n$.

We define $Y_k, \tilde{Y}_k : \Omega \to M$ by $Y_k(\omega) = S_k(X_1(\omega), \dots, X_k(\omega))$ and $\tilde{Y}_k(\omega) = S_k(\tilde{X}_1(\omega), \dots, \tilde{X}_k(\omega))$. It follows from Lemma 1 that $Y_k(\omega) \le \tilde{Y}_k(\omega)$ for each $\omega \in \Omega$. By Corollary 1 we have that $\lim_{k\to\infty} Y_k(\omega) = \mathfrak{G}_n(\mathbf{w}; x_1, \dots, x_n)$ a.e. and $\lim_{k\to\infty} \tilde{Y}_k(\omega) = \mathfrak{G}_n(\mathbf{w}; y_1, \dots, y_n)$ a.e. By the closedness of the partial order (and the fact that the intersection of two sets of measure 1 still has measure 1), we conclude that $\mathfrak{G}_n(\mathbf{w}; x_1, \dots, x_n) \le \mathfrak{G}_n(\mathbf{w}; y_1, \dots, y_n)$.

Since the trace metric on the space $\mathbb{P}$ of $m \times m$ positive definite (real or complex) matrices makes it a Loewner–Heinz NPC space with respect to the Loewner order (see e.g. [18]), we have the following corollary.

Corollary 2 *The weighted least squares mean on the set $\mathbb{P}$ of positive definite matrices is monotonic.*

Remark 3 Loewner [22] proved that a function defined on an open interval is operator monotone if and only if it allows an analytic continuation into the complex upper half-plane with nonnegative imaginary part. The function $f(t) = t^\alpha, \alpha \in [0, 1]$ is operator monotone on the positive reals, that is, $X \leq Y$ implies $X^\alpha \leq Y^\alpha$ for positive definite matrices X and Y. The inequality was independently proved by Heinz [13]. It is equivalent to the extended monotonicity property of the weighted geometric mean: $B_1 \#_t B_2 \leq A_1 \#_t A_2, t \in [0, 1]$, whenever $B_1 \leq A_1$ and $B_2 \leq A_2$. It is natural to consider the monotonicity of the least squares mean $\mathfrak{G}_n(\omega; \mathbb{B}) \leq \mathfrak{G}_n(\omega; \mathbb{A})$ whenever $B_i \leq A_i$ for each i as an n-variable Loewner-Heinz inequality for positive definite matrices.

A function $F : \mathbb{P}^n \to \mathbb{P}$ is *jointly concave* if for any $(A_1, \ldots, A_n), (B_1, \ldots, B_n) \in \mathbb{P}^n$ and $0 \leq t \leq 1$, we have

$$tF(A_1, \ldots, A_n) + (1-t)F(B_1, \ldots, B_n) \leq F(tA_1 + (1-t)B_1, \ldots, tA_n + (1-t)B_n).$$

Proposition 2 *The least squares mean $\mathfrak{G}_n : \mathbb{P}^n \to \mathbb{P}$ for the trace metric is jointly concave for each $n \geq 2$.*

Proof It is a standard result that the two-variable weighted geometric mean on $\mathbb{P}$ is jointly concave. It follows directly by induction that the inductive mean S_n of positive definite matrices is jointly concave for $n \geq 2$ (see the proof of Lemma 1).

Fix $(A_1, \ldots, A_k), (B_1, \ldots, B_k) \in \mathbb{P}^k$ and a weight $\mathbf{w} = (w_1, \ldots, w_k)$. Construct random variables $\{X_k\}, \{\tilde{X}_k\}$ as in the proof of Theorem 2 with A_i replacing x_i and B_i replacing y_i for each i. For $Y_k = S_k(X_1, \ldots, X_k)$ and $\tilde{Y}_k = S_k(\tilde{X}_1, \ldots, \tilde{X}_k\}$, we conclude from the concavity of S_k that

$$tY_k + (1-t)\tilde{Y}_k \leq S_k(tX_1 + (1-t)\tilde{X}_1, \ldots, tX_k + (1-t)\tilde{X}_k) = S_k(Z_1, \ldots, Z_k),$$

where $Z_i = tX_i + (1-t)\tilde{X}_i$ for $1 \leq i \leq k$. Note that the Z_k are i.i.d. with each Z_k having distribution that assigns mass w_i to each $tA_i + (1-t)B_i$, $1 \leq i \leq n$. From Corollary 1 the limit of both sides exists a.e. and is given by the appropriate least squares mean, and from the closedness of the order we conclude

$$t\mathfrak{G}_n(\mathbf{w}; A_1, \ldots, A_n) + (1-t)\mathfrak{G}_n(\mathbf{w}; B_1, \ldots, B_m) \leq \mathfrak{G}_n(\mathbf{w}; C_1, \ldots, C_n),$$

where $C_i = tA_i + (1-t)B_i$ for each i.

6 Other properties of the least squares mean

The fact that the unweighted least squares mean is bounded above by the arithmetic mean, and hence below by the harmonic mean has been recently shown by

Yamazaki [27]. We give an alternative approach via probabilistic methods and derive the result for the weighted least squares mean.

Proposition 3 *For* $(A_1, \ldots, A_n) \in \mathbb{P}^n$ *and a weight* $\mathbf{w} = (w_1, \ldots, w_n)$*, we have*

$$\left(\sum_{i=1}^{n} w_i A_i^{-1}\right)^{-1} \leq \mathfrak{G}_n(\mathbf{w}; A_1, \ldots, A_n) \leq \sum_{i=1}^{n} w_i A_i.$$

Proof It is a standard result that the two-variable weighted geometric mean on $\mathbb{P}$ is below the corresponding weighted arithmetic mean: $A\#_t B \leq (1-t)A + tB$ for $0 \leq t \leq 1$. It follows by induction that the inductive mean satisfies for each k

$$\begin{aligned} S_k(B_1, \ldots, B_k) &= S_{k-1}(B_1, \ldots, B_{k-1})\#_{1/k} B_k \\ &\leq \frac{k-1}{k} \sum_{i=1}^{k-1} \frac{1}{k-1} B_i + \frac{1}{k} B_k = \frac{1}{k} \sum_{i=1}^{k} B_i. \end{aligned}$$

Construct a sequence of i.i.d. random variables $\{X_k\}$ as in Sect. 4 such that the distribution is $\sum_{i=1}^{n} w_i \delta_{A_i}$ for each X_k. Set $Y_k = S_k(X_1, \ldots, X_k)$ for each k. From Corollary 1 $\lim_{k\to\infty} Y_k(\omega) = \mathfrak{G}_n(\mathbf{w}; A_1, \ldots, A_n)$ a.e.

Endow the space of Hermitian matrices $\mathbb{H}$ including $\mathbb{P}$ with the Hilbert space structure with inner product $\langle A, B\rangle = \mathrm{tr} A^* B$. Then $\mathbb{P}$ is an open subset of $\mathbb{H}$. Set $Z_k = \sum_{i=1}^{k} (1/k) X_i$, where $\{X_k\}$ are the random variables of the previous paragraph. By Proposition 1(v), $\lim_{k\to\infty} Z_k(\omega) \to \sum_{i=1}^{n} w_i A_i$ a.e. By the first paragraph $Y_k(\omega) \leq Z_k(\omega)$ for all k, ω. From the closure of the order, we conclude that $\mathfrak{G}_n(\mathbf{w}; A_1, \ldots, A_n) \leq \sum_{i=1}^{n} w_i A_i$.

The first inequality in the conclusion of the proposition follows from the second and the fact that inversion is an isometry for the trace metric and hence the least squares mean is self-dual, i.e., $\mathfrak{G}_n(\mathbf{w}; A_1^{-1}, \ldots, A_n^{-1})^{-1} = \mathfrak{G}_n(\mathbf{w}; A_1, \ldots, A_n)$.

Let M be an NPC space. Given probability measures $p, q \in \mathcal{P}(M)$, we say that a probability measure $\mu \in \mathcal{P}(M^2)$ is a *coupling* of p and q if the *marginals* of μ are p and q, that is, if for all Borel sets $B \in \mathcal{B}(M)$

$$\mu(B \times M) = p(B) \text{ and } \mu(M \times B) = q(B). \tag{9}$$

Definition 2 The (L^1)-*Wasserstein distance* ρ on $\mathcal{P}^1(M)$ is given by

$$W(p, q) = \inf\left\{ \int_{M\times M} d(x, y)\mu(dx\,dy) : \mu \text{ is a coupling of } p \text{ and } q \right\}.$$

We adopt the most common name for the metric, the *Wasserstein* distance, although it also appears under a variety of other names such as the *Kantorovich–Rubinstein* distance.

Proposition 4 *For $(x_1, \dots, x_n), (y_1, \dots, y_n) \in M^n$, a weight $\mathbf{w} = (w_1, \dots, w_n)$, and the corresponding finitely supported probability measures $q_1 = \sum_{k=1}^n w_i \delta_{x_i}$ and $q_2 = \sum_{k=1}^n w_i \delta_{y_i}$ on M,*

$$d(\mathfrak{G}_n(\mathbf{w}; x_1, \dots, x_n), \mathfrak{G}_n(\mathbf{w}; y_1, \dots, y_n)) \le W(q_1, q_2) \le \sum_{i=1}^n w_i d(x_i, y_i).$$

Hence, in particular, the least squares mean $\mathfrak{G}_n$ is continuous for each $\mathbf{w}$.

Proof Define μ on $M \times M$ by $\mu = \sum_{i=1}^n w_i \delta_{(x_i, y_i)}$. One sees readily that μ is a coupling of q_1 and q_2, and thus $W(q_1, q_2) \le \int_{M \times M} d(x, y)\mu(dxdy) = \sum_{i=1}^n w_i d(x_i, y_i)$. By Theorem 6.3 of [26], the barycentric map $b : \mathcal{P}^1(M) \to M$ satisfies, for all p, q, the fundamental contraction property $d(b(p), b(q)) \le W(p, q)$. By Remark 2 $b(q_1) = \mathfrak{G}_n(\mathbf{w}; x_1, \dots, x_n)$ and similarly $b(q_2) = \mathfrak{G}_n(\mathbf{w}; y_1, \dots, y_n)$. Thus

$$\begin{aligned} d(\mathfrak{G}_n(\mathbf{w}; x_1, \dots, x_n), \mathfrak{G}_n(\mathbf{w}; y_1, \dots, y_n)) &= d(b(q_1), b(q_2)) \le W(q_1, q_2) \\ &\le \sum_{i=1}^n w_i d(x_i, y_i). \end{aligned}$$

The fact that the right-hand of the preceding is larger than the left-hand directly establishes the continuity of $\mathfrak{G}_n$.

From this result together with Corollary 2 and Propositions 2 and 3 we conclude that the least squares mean of positive definite matrices satisfies the continuity, monotonicity, joint concavity, and AGM inequality properties, and hence all the fundamental properties of the geometric means of positive definite matrices defined for and satisfied by the Ando–Li–Mathias and Bini–Meini–Poloni constructions [2,9]; see [27] for other properties.

We note in closing that Corollary 2 and Propositions 2, 3 and 4 extend to the positive symmetric cones of finite-dimensional Euclidean Jordan algebras; see [20].

Acknowledgments This paper has benefited from the careful reading, editing, and thoughtful suggestions of the referee.

References

1. Ahn, E., Kim, S., Lim, Y.: An extended Lie–Trotter formula and its applications. Linear Algebra Appl. **427**, 190–196 (2007)
2. Ando, T., Li, C.K., Mathias, R.: Geometric means. Linear Algebra Appl. **385**, 305–334 (2004)
3. Ballman, W.: Lectures on Spaces of Nonpositive Curvature. Birkhäuser Verlag, Basel (1995)
4. Barbaresco, F.: Innovative tools for radar signal processing based on Cartan's geometry of symmetric positive definite matrices and information geometry. In: Proceedings of the IEEE International Radar Conference (Rome, Italy 2008)

5. Berger, M.: A Panoramic View of Riemannlan Geometry. Springer, Berlin (2003)
6. Bhatia, R.: Positive definite matrices. In: Princeton Series in Applied Mathematics. Princeton University Press, Princeton, NJ (2007)
7. Bhatia, R., Holbrook, J.: Noncommutative geometric means. Math. Intell. **28**, 32–39 (2006)
8. Bhatia, R., Holbrook, J.: Riemannian geometry and matrix geometric means. Linear Algebra Appl. **413**, 594–618 (2006)
9. Bini, D., Meini, B., Poloni, F.: An effective matrix geometric mean satisfying the Ando–Li–Mathias properties. Math. Comp. **79**, 437–452 (2010)
10. Bridson, M., Haefliger, A.: Metric Spaces of Non-Positive Curvature. Springer, Berlin (1999)
11. Es-Sahib, A., Heinich, H.: Barycentre canonique pour un espace métrique à courbure négative. Séminaire de probabilités (Strasbourg) **33**, 355–370 (1999)
12. Fletcher, P.D., Joshi, S.: Riemannian geometry for the statistical analysis of difusion tensor data. Signal Process. **87**, 250–262 (2007)
13. Heinz, E.: Beitäge zur Störungstheorie der Spectralzerlegung. Math. Ann. **123**, 415–438 (1951)
14. Jost, J.: Nonpositive curvatures: geometric and analytic aspects. In: Lectures in Mathematics ETH Zürich. Birkhäuser, Basel (1997)
15. Karcher, H.: Riemannian center of mass and mollifier smoothing. Commun. Pure Appl. Math. **30**, 509–541 (1977)
16. Kubo, F., Ando, T.: Means of positive linear operators. Math. Ann. **246**, 205–224 (1980)
17. Lang, S.: Fundamentals of differential geometry. In: Graduate Texts in Math Springer, Berlin (1999)
18. Lawson, J.D., Lim, Y.: The geometric mean, matrices, metrics, and more. Am. Math. Monthly **108**, 797–812 (2001)
19. Lawson, J.D., Lim, Y.: A general framework for extending means to higher orders. Colloq. Math. **113**, 191–221 (2008)
20. Lawson, J.D., Lim, Y.: Monotonic properties of the least squared mean (preprint). arXiv:1007.4792v1
21. Lawson, J.D., Lee, H., Lim, Y.: Weighted geometric means. Forum. Math. (to appear)
22. Löwner, K.: Über monotone Matrixfunktionen. Math. Z. **38**, 177–216 (1934)
23. Moakher, M.: A differential geometric approach to the geometric mean of symmetric positive-definite matrices. SIAM J. Matrix Anal. Appl. **26**, 735–747 (2005)
24. Pennec, X., Fillard, P., Ayache, N.: A Riemannian framework for tensor computing. Int. J. Comput. Vis. **66**, 41–66 (2006)
25. Sagae, M., Tanabe, K.: Upper and lower bounds for the arithmetic-geometric-harmonic means of positive definite matrices. Linear Multilinear Algebras **37**, 279–282 (1994)
26. Sturm, K.-T.: Probability measures on metric spaces of nonpositive curvature. In: Auscher, P. et al. (eds.) Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces. Contemp. Math. vol. 338. Amer. Math. Soc. (AMS), Providence (2003)
27. Yamazaki, T.: On properties of geometric mean of n-operators via Riemannian metric (preprint)
28. Zerai, M., Moakher, M.: Riemannian curvature-driven flows for tensor-valued data. In: Sgallari, F., Murli, A., Paragios, N. (eds) Scale Space and Variational Methods in Computer Vision, pp. 592–602. Springer, Berlin (2007)
29. Zhan, X.: Open problems in matrix theory. In: Ji, L., Liu, K., Yang, L., Yau S.-T. (eds.) Proceedings of the 4th International Congress of Chinese Mathematicians, vol. I. Higher Education Press, Beijing, pp. 367–382 (2008)