The Metric Geometry of the Multivariable Matrix Geometric Mean

Jimmie Lawson

Joint Work with Yongdo Lim

Department of Mathematics

Louisiana State University

Baton Rouge, LA 70803, USA

Introduction

Positive definite matrices have become fundamental computational objects in a diverse variety of applied areas. They appear as covariance matrices in statistics, elements of the search space in convex and semidefinite programming, kernels in machine learning, density matrices in quantum information, and diffusion tensors in medical imaging, to cite a few.

Introduction

Positive definite matrices have become fundamental computational objects in a diverse variety of applied areas. They appear as covariance matrices in statistics, elements of the search space in convex and semidefinite programming, kernels in machine learning, density matrices in quantum information, and 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, estimation, and averaging, the last being the concern of this talk.

The Trace Metric

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 δ, the distance metric arising from the natural Riemannian structure on $\mathbb{P}$ making it a Riemannian manifold, indeed a symmetric space, of negative curvature:

$$\delta(A, B) = \left(\sum_{i=1}^{k} \log^2 \lambda_i(A^{-1}B)\right)^{\frac{1}{2}},$$

where $\lambda_i(X)$ denotes the ith eigenvalue of X in non-decreasing order.

Basic Geometric Properties

We list some basic properties of $\mathbb{P}$ endowed with the trace metric.

(1) The matrix geometric mean $A\#B = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}$ is the unique metric midpoint between A and B.

(2) There is a unique metric geodesic line through any two distinct points $A, B \in \mathbb{P}$ given by the weighted means $\gamma(t) = A\#_t B = A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2}$.

(3) (Congruence Invariance) Congruence transformations $A \mapsto CAC^*$ for C invertible are isometries.

(4) Inversion $A \mapsto A^{-1}$ is an isometry.

(5) (Monotonicity) $A \leq B$, $C \leq D \Rightarrow A\#_t B \leq C\#_t D$.

Extending the Geometric Mean

Once one realizes 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 δ, it is natural to use an averaging technique over this metric to extend this mean to n-variables. First M. Moakher (2005) and then Bhatia and Holbrook (2006) suggested the least squares mean, taking the mean to be the unique minimizer of the sum of the squares of the distances:

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

Some Background

This idea had been anticipated by Élie Cartan, 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 $\mathbb{P}$.

Some Background

This idea had been anticipated by Élie Cartan, 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 $\mathbb{P}$.

The mean is frequently called the Karcher mean in light of its appearance in his work on Riemannian manifolds (1977). Indeed, he considered general probabilistic means that included weighted least squares mean:

$$\mathfrak{G}_n(w_1, \ldots, w_n; A_1, \ldots, A_n) = \arg\min_{X \in \mathbb{P}} \sum_{i=1}^{n} w_i \delta^2(X, A_i),$$

where the non-negative w_i satisfy $\sum_{i=1}^{n} w_i = 1$.

A Natural Question

In a 2004 LAA article called “Geometric Means" T. Ando, C.K. Li and R. Mathias gave a construction (frequently called “symmetrization") that extended the two-variable matrix geometric mean to n-variables for each $n \geq 3$ and identified a list of ten properties that this extended mean satisfied. Both contributions–the construction and the axiomatic properties–were important and have been influential in subsequent developments.

A Natural Question

In a 2004 LAA article called “Geometric Means" T. Ando, C.K. Li and R. Mathias gave a construction (frequently called “symmetrization") that extended the two-variable matrix geometric mean to n-variables for each $n \geq 3$ and identified a list of ten properties that this extended mean satisfied. Both contributions–the construction and the axiomatic properties–were important and have been influential in subsequent developments.

Question: Do the Ando-Li-Mathias properties extend to the least squares mean? In particular, Bhatia and Holbrook (2006) asked whether the least squares mean was monotonic in each of its arguments. Computer calculations indicated “Yes."

NPC Spaces

The answer is indeed “yes," but showing it required new tools: the theory of nonpositively curved metric spaces, techniques from probability and random variable theory, and the fairly recent combination of the two, particularly by K.-T. Sturm (2003).

NPC Spaces

The answer is indeed “yes," but showing it required new tools: the theory of nonpositively curved metric spaces, techniques from probability and random variable theory, and the fairly recent combination of the two, particularly by K.-T. Sturm (2003).

The setting appropriate for our considerations is that of globally nonpositively curved metric spaces, or *NPC spaces* for short: These are complete metric spaces M satisfying for each $x, y \in M$, there exists $m \in M$ such that for all $z \in M$

$$d^2(m, z) \leq \frac{1}{2}d^2(x, z) + \frac{1}{2}d^2(y, z) - \frac{1}{4}d^2(x, y). \qquad \text{(NPC)}$$

Such spaces are also called (global) CAT(0)-spaces or Hadamard spaces.

Geodesics and Weighted Means

The theory of such NPC spaces is quite extensive. In particular the m appearing in

$$d^2(m,z) \le \frac{1}{2}d^2(x,z) + \frac{1}{2}d^2(y,z) - \frac{1}{4}d^2(x,y) \qquad \text{(NPC)}$$

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 $d(\gamma(t), \gamma(s)) = |t-s| d(x,y)$, $\gamma(0) = x$, $\gamma(1) = y$.

Geodesics and Weighted Means

The theory of such NPC spaces is quite extensive. In particular the m appearing in

$$d^2(m,z) \le \frac{1}{2}d^2(x,z) + \frac{1}{2}d^2(y,z) - \frac{1}{4}d^2(x,y) \qquad \text{(NPC)}$$

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 $d(\gamma(t),\gamma(s)) = |t-s|d(x,y)$, $\gamma(0) = x$, $\gamma(1) = y$.

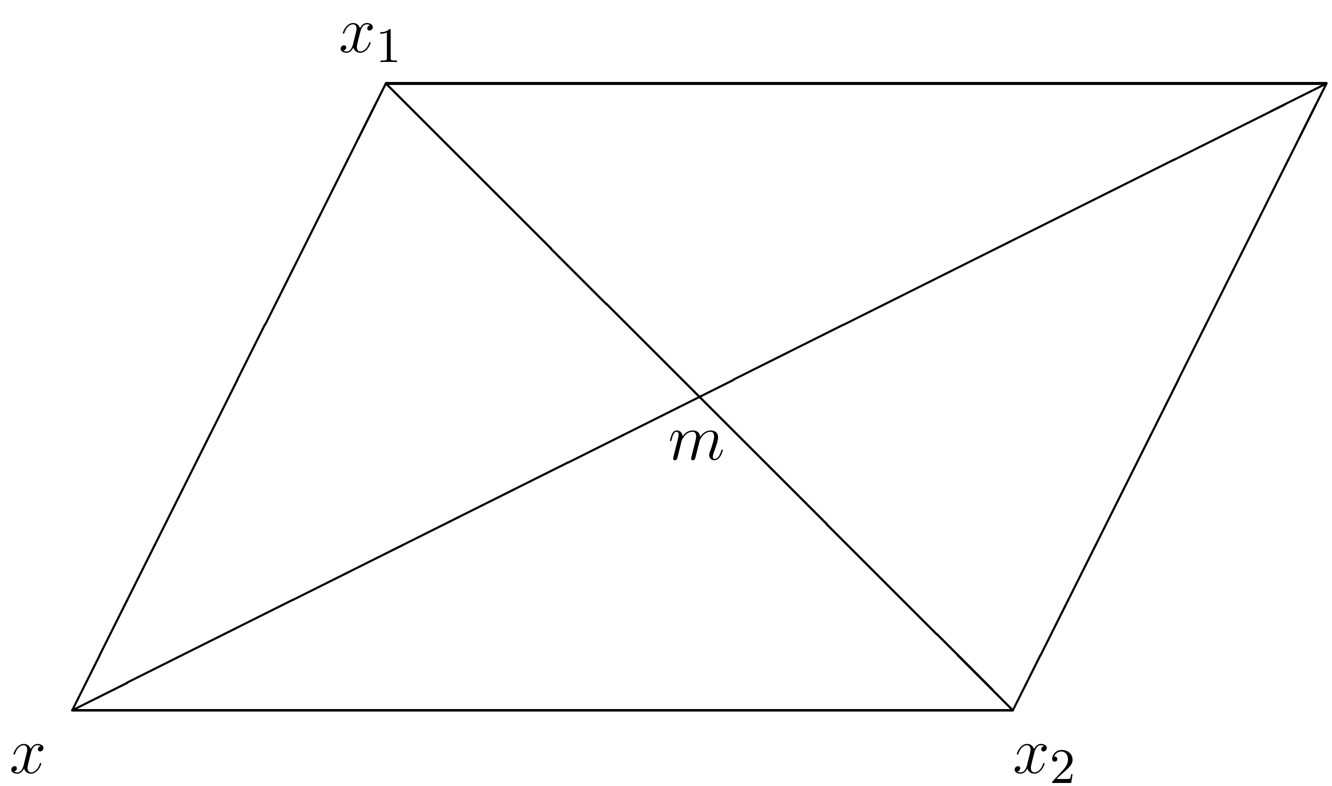
We denote $\gamma(t)$ by $x\#_t y$ and call it the t-*weighted mean* of x and y. The midpoint $x\#_{1/2}y$ we denote simply as $x\#y$. We remark that by uniqueness $x\#_t y = y\#_{1-t}x$; in particular, $x\#y = y\#x$.

The Semiparallelogram Law

Weakening the parallelogram law in Hilbert space to an inequality yields (NPC) or the semiparallelogram law:

$$\text{sum of 2 diagonals squared} \quad \leq \quad \text{sum of 4 sides squared}$$

$$d^2(x_1, x_2) + 4d^2(x, m)(= (2d(x,m))^2) \quad \leq \quad 2d^2(x, x_1) + 2d^2(x, x_2)$$



Metrics and Curvature

Equation (NPC), the *semiparallelogram law*, holds in both euclidean and hyperbolic geometry. More generally, it is satisfied by the length metric in any simply connected nonpositively curved Riemannian manifold. Hence the metric definition represents a metric generalization of nonpositive curvature.

Fact: *The trace metric on the nonpositively curved Riemannian symmetric space of positive definite matrices* $\mathbb{P}$ *is a particular and important example of an NPC space*.

Metrics and Curvature

Equation (NPC), the *semiparallelogram law*, holds in both euclidean and hyperbolic geometry. More generally, it is satisfied by the length metric in any simply connected nonpositively curved Riemannian manifold. Hence the metric definition represents a metric generalization of nonpositive curvature.

Fact: *The trace metric on the nonpositively curved Riemannian symmetric space of positive definite matrices* $\mathbb{P}$ *is a particular and important example of an NPC space*.

Equation (NPC) admits a more general formulation in terms of the weighted mean. For all $0 \leq t \leq 1$ we have

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

Metric Least Squares Means

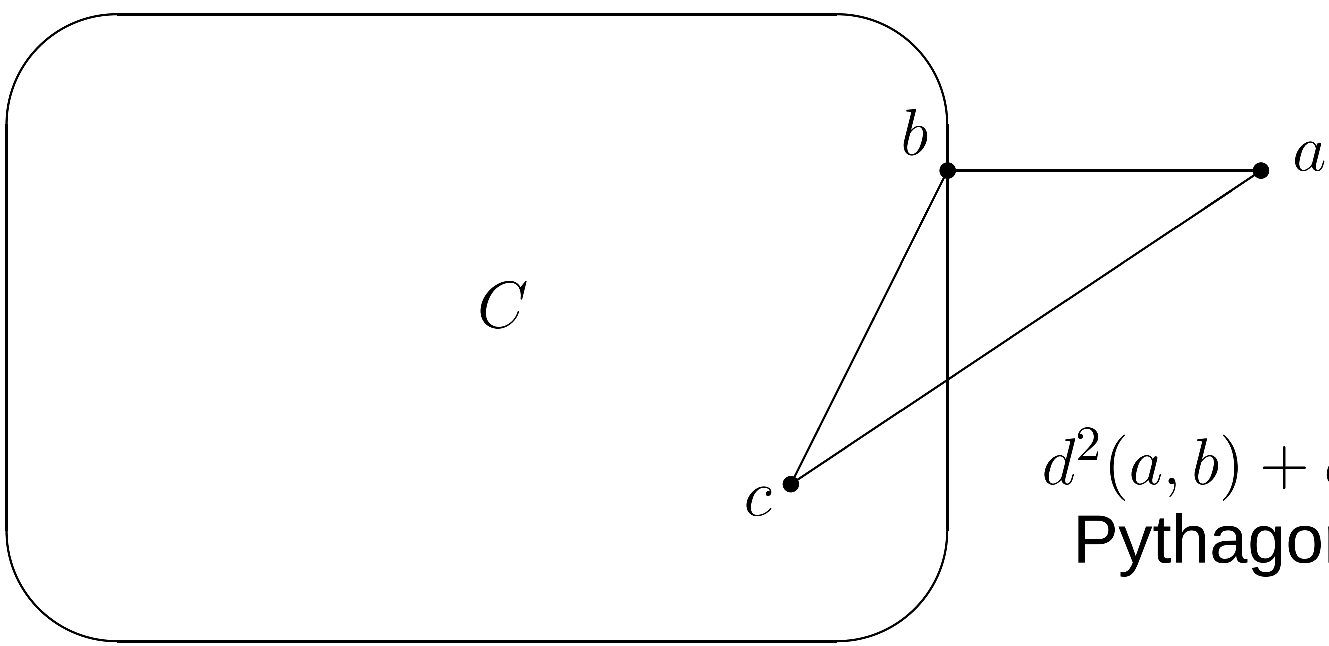
The weighted least squares mean $\mathfrak{G}(\cdot;\cdot)$ can be easily formulated in any metric space (M,d). Given $(a_1,\ldots,a_n)\in M^n$, and positive real numbers $w_1,\ldots,w_n$ summing to 1, we define

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

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.

Pythagorean Inequality

Given a closed geodesically convex set C and a point $a \notin C$ in an NPC-space, there exists a unique point $b \in C$ closest to a, called the projection of a on C. Taking any point $c \in C$, $c \neq b$, it is the case that $d^2(a,b) + d^2(b,c) \leq d^2(a,c)$. Since intuitively $\angle abc \geq 90°$, it makes sense to call this inequality the Pythagorean inequality.



$$d^2(a,b) + d^2(b,c) \leq d^2(a,c)$$

Pythagorean Inequality

A Basic Theorem

Theorem. *The weighted geometric mean* $\mathfrak{G}_n(w_1,\ldots,w_n;a_1,\ldots,a_n)$ *belongs to the convex hull of* $\{a_1,\ldots,a_n\}$.

Proof. If $a=\mathfrak{G}_n(w_1,\ldots,w_n;a_1,\ldots,a_n)$ lies outside the convex hull C of $\{a_1,\ldots,a_n\}$, then by the Pythagorean Inequality for the projection b of a onto C, $d^2(b,a_i)\leq d^2(a,a_i)-d^2(a,b)<d^2(a,a_i)$ for each i, which contradicts the fact that a is the *least* squares mean.

The Inductive Mean

One other mean will play an important role in what follows, one that we shall call the inductive mean, following the terminology of K.-T. Sturm (2003). It appeared elsewhere in the work of M. Sagae and K. Tanabe (1994) and Ahn, Kim, and Lim (2007). It is defined inductively for NPC spaces (or more generally for metric spaces with weighted binary means $x\#_t y$) for each $k \geq 2$ by $S_2(x, y) = x\#y$ and for $k \geq 3$, $S_k(x_1, \ldots, x_k) = S_{k-1}(x_1, \ldots, x_{k-1})\#_{\frac{1}{k}} x_k$. In euclidean space it collapses to $S_n(x_1, \ldots, x_n) = (1/n) \sum_{i=1}^{n} x_i$.

The Inductive Mean

One other mean will play an important role in what follows, one that we shall call the inductive mean, following the terminology of K.-T. Sturm (2003). It appeared elsewhere in the work of M. Sagae and K. Tanabe (1994) and Ahn, Kim, and Lim (2007). It is defined inductively for NPC spaces (or more generally for metric spaces with weighted binary means $x\#_t y$) for each $k \geq 2$ by $S_2(x,y) = x\#y$ and for $k \geq 3$, $S_k(x_1,\ldots,x_k) = S_{k-1}(x_1,\ldots,x_{k-1})\#_{\frac{1}{k}} x_k$. In euclidean space it collapses to $S_n(x_1,\ldots,x_n) = (1/n)\sum_{i=1}^n x_i$.

Note that this mean at each stage is defined from the previous stage by taking the appropriate two-variable weighted mean, which is monotone in each variable for $\mathbb{P}$. Thus the inductive mean is monotone.

Random Walks and Sturm's Theorem

Let (X, d) be an NPC metric space, $\{x_1, \ldots, x_m\} \subseteq X$. Set $\mathbb{N}_m = \{1, 2, \ldots, m\}$ and assign to $k \in \mathbb{N}_m$ the probability w_k, where $0 \leq w_k \leq 1$ and $\sum_{k=1}^{m} w_i = 1$. For each $\omega \in \prod_{n=1}^{\infty} \mathbb{N}_m$, define inductively a sequence $\sigma = \sigma_\omega$ in X by $\sigma(1) = x_{\omega(1)}$, $\sigma(k) = S_k(x_{\omega(1)}, \ldots, x_{\omega(k)})$, where S_k is the inductive mean. (The sequence σ_ω may be viewed as a "walk" starting at $\sigma(1) = x_{\omega(1)}$ and obtaining $\sigma(k)$ by moving from $\sigma(k-1)$ toward $x_{\omega(k)}$ a distance of $(1/k)d(\sigma(k-1), x_{\omega(k)})$.)

Random Walks and Sturm's Theorem

Let (X, d) be an NPC metric space, $\{x_1, \ldots, x_m\} \subseteq X$. Set $\mathbb{N}_m = \{1, 2, \ldots, m\}$ and assign to $k \in \mathbb{N}_m$ the probability w_k, where $0 \leq w_k \leq 1$ and $\sum_{k=1}^{m} w_i = 1$. For each $\omega \in \prod_{n=1}^{\infty} \mathbb{N}_m$, define inductively a sequence $\sigma = \sigma_\omega$ in X by $\sigma(1) = x_{\omega(1)}$, $\sigma(k) = S_k(x_{\omega(1)}, \ldots, x_{\omega(k)})$, where S_k is the inductive mean. (The sequence σ_ω may be viewed as a "walk" starting at $\sigma(1) = x_{\omega(1)}$ and obtaining $\sigma(k)$ by moving from $\sigma(k-1)$ toward $x_{\omega(k)}$ a distance of $(1/k)d(\sigma(k-1), x_{\omega(k)})$.)

Sturm's Theorem. *Giving $\prod_{n=1}^{\infty} \mathbb{N}_m$ the product probability, the set $\{\omega \in \prod_{n=1}^{\infty} \mathbb{N}_m : \lim_n \sigma_\omega(n) = \mathfrak{G}(\mathbf{w}; x_1, \ldots, x_m)\}$ has measure 1, i.e., $\sigma_\omega(n) \to \mathfrak{G}(\mathbf{w}; x_1, \ldots, x_m)$ for almost all ω.*
More generally, Sturm establishes a version of the Strong Law of Large Numbers for random variables into an NPC metric space, with limit the least squares mean.

The Matrix Geometric Mean

Using Sturm's Theorem, Lawson and Lim (2011) were able to show:

(1) The least squares mean $\mathfrak{G}$ on $\mathbb{P}$, the limit a.e. of the inductive mean, is monotone: $A_i \leq B_i$ for $1 \leq i \leq n$ implies $\mathfrak{G}(A_1, \ldots, A_n) \leq \mathfrak{G}(B_1, \ldots B_n)$.

(2) All ten of the Ando-Li-Mathias (ALM) axioms hold for $\mathfrak{G}$.

(3) In a natural way $\mathfrak{G}$ can be extended to a weighted mean, and appropriate weighted versions of the ten properties hold.

The Matrix Geometric Mean

Using Sturm's Theorem, Lawson and Lim (2011) were able to show:

(1) The least squares mean $\mathfrak{G}$ on $\mathbb{P}$, the limit a.e. of the inductive mean, is monotone: $A_i \leq B_i$ for $1 \leq i \leq n$ implies $\mathfrak{G}(A_1, \ldots, A_n) \leq \mathfrak{G}(B_1, \ldots B_n)$.

(2) All ten of the Ando-Li-Mathias (ALM) axioms hold for $\mathfrak{G}$.

(3) In a natural way $\mathfrak{G}$ can be extended to a weighted mean, and appropriate weighted versions of the ten properties hold.

Note: The ALM mean is typically distinct from the least squares mean for $n \geq 3$. Thus the ALM axioms do not characterize a mean. The latter fact had already been noted by Bini, Meini and Poloni (2010), who introduced a much more computationally efficient variant of the ALM mean.

Follow-ups

Bhatia (2012) was able to give a simplified self-contained probabilistic proof of a weaker law of large numbers for the case $\mathbb{P}$ of the positive matrices that sufficed to derive monotonicity and other properties of the multivariable geometric mean. Holbrook (2012) gave a non-probabilistic (“no dice") proof of a direct convergence scheme of the inductive mean to $\mathfrak{G}$.

In some remarkable very recent work M. Bačák has given a nonprobabilistic version of the Proximal Point Algorithm for general NPC-spaces that gives a general convergence scheme for computing the least squares mean.

The Proximal Point Algorithm

Let (X,d) be an NPC-space, $\mathbf{a} = (a_1,\ldots,a_m) \in X^n$, $\mathbf{w} = (w_1,\ldots,w_m)$ a weight. Let $\{\lambda_k > 0\}_{k\geq 0}$ be a sequence satisfying $\sum_k \lambda_k = \infty$, $\sum_k \lambda_k^2 < \infty$.

Given $x_0 \in X$, inductively for $k \geq 0$ set

$$
\begin{aligned}
x_{km+1} &= x_{km}\#_t a_1 \text{ for } t = \frac{\lambda_k w_1}{1+\lambda_k w_1} \\
x_{km+2} &= x_{km+1}\#_t a_2 \text{ for } t = \frac{\lambda_k w_2}{1+\lambda_k w_2} \\
\ldots\ldots &= \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
x_{km+m}(= x_{k(m+1)}) &= x_{km+m-1}\#_t a_n \text{ for } t = \frac{\lambda_k w_n}{1+\lambda_k w_n}
\end{aligned}
$$

Then $x_j \to \mathfrak{G}_m(\mathbf{w};\mathbf{a})$, the least squares mean.

The Karcher Equation

The uniform convexity of the trace metric d on $\mathbb{P}$ yields that the least squares mean is the unique critical point for the function $X \mapsto \sum_{k=1}^{n} d^2(X, A_k)$. The least squares mean is thus characterized by the vanishing of the gradient, which is equivalent to its being a solution of the following Karcher equation:

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

The Karcher Equation

The uniform convexity of the trace metric d on $\mathbb{P}$ yields that the least squares mean is the unique critical point for the function $X \mapsto \sum_{k=1}^{n} d^2(X, A_k)$. The least squares mean is thus characterized by the vanishing of the gradient, which is equivalent to its being a solution of the following Karcher equation:

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

This latter equivalence provides a method for defining a mean on bounded positive definite operators on an infinite-dimensional Hilbert space (where the trace metric is no longer available) as the solution to the Karcher equation.

Power(ful) Means

Power means for positive definite matrices have recently been introduced by Lim and Palfia (2012).

Theorem. *Let* $A_1, \ldots, A_n \in \Omega$ *and let* $\mathbf{w} = (w_1, \ldots, w_n)$ *be a weight. Then for each* $t \in (0, 1]$, *the following equation has a unique positive definite solution* $X = P_t(\mathbf{w}; A_1, \ldots, A_n)$:

$$X = \sum_{i=1}^{n} w_i (X \#_t A_i).$$

Power(ful) Means

Power means for positive definite matrices have recently been introduced by Lim and Palfia (2012).
Theorem. *Let* $A_1, \ldots, A_n \in \Omega$ *and let* $\mathbf{w} = (w_1, \ldots, w_n)$ *be a weight. Then for each* $t \in (0, 1]$, *the following equation has a unique positive definite solution* $X = P_t(\mathbf{w}; A_1, \ldots, A_n)$:

$$X = \sum_{i=1}^{n} w_i (X \#_t A_i).$$

No closed formula is known, but when restricted to the positive reals (or commuting matrices), the power mean reduces to the usual power mean

$$P_t(\mathbf{w}; a_1, \ldots, a_n) = \left(w_1 a_1^t + \cdots + w_n a_n^t\right)^{\frac{1}{t}}.$$

Power Means II

Power means are invariant under permutations, monotonic, and invariant under congruence transformations. Their geometry appears to be more suited to the Thompson metric than the trace metric. In particular, they are non-expansive with respect to the Thompson metric and the right hand side of the defining equation is a strict contraction (properties 3 and 4, next slide).

Power Means II

Power means are invariant under permutations, monotonic, and invariant under congruence transformations. Their geometry appears to be more suited to the Thompson metric than the trace metric. In particular, they are non-expansive with respect to the Thompson metric and the right hand side of the defining equation is a strict contraction (properties 3 and 4, next slide).

The power means are decreasing, $s < t$ implies $P_s(\cdot\,;\,\cdot) \le P_t(\cdot\,;\,\cdot)$, and most importantly

$$\lim_{t \to 0^+} P_t(\cdot\,;\,\cdot) = \mathfrak{G}(\cdot\,;\,\cdot).$$

Thus by passing to limits the power means provide an alternative (probability free) route to the monotonicity of the least squares mean $\mathfrak{G}$.

The Thompson Metric

The Thompson metric on $\mathbb{P}$ is defined by

$$d(A,B) = \log\{M(A/B), M(B/A)\},\ M(A/B) = \inf\{t : A \leq tB\}.$$

For $\mathbb{P}$ finite or infinite dimensional, the following hold.
1. d is a complete metric on $\mathbb{P}$, and its topology agrees with the relative operator topology.
2. The mappings $\gamma(t) = A\#_t B$ form a family of distinguished metric geodesics, but are no longer unique.
3. The Thompson metric satisfies a weaker nonpositive curvature condition than NPC-metrics, namely Busemann nonpositive curvature: $d(A\#B, A\#C) \leq \frac{1}{2}d(A,B)$. (This works out to be a geometric version of the Loewner-Heinz inequality: $A \leq B \Rightarrow A^{1/2} \leq B^{1/2}$.)
4. $d(\sum_{i=1}^n A_i, \sum_{i=1}^n B_i) \leq \max_{1\leq i\leq n}\{d(A_i, B_i)\}$.

The Infinite Dimensional Case

For $\mathbb{P}$ the positive operators on an infinite-dimensional Hilbert space, properties of the Thompson metric yield the existence and basic properties of the power means P_t. The strong limit of the decreasing family $\{P_t : t > 0\}$ gives a solution to the Karcher equation, which retains most of the ALM properties, as shown by Lawson and Lim. The Thompson metric is a key tool in establishing its uniqueness.

The Extended Hilbert-Schmidt Algebra

In recent work G. Larotonda has extended the Riemannian manifold structure of finite-dimensional $\mathbb{P}$ to the setting of the positive Hilbert-Schmidt positive operators augmented with all positive scalar multiplies of the identity, all operating on an infinite-dimensional Hilbert space. This manifold $\mathbb{P}_\omega$ has an analog of the trace metric making $\mathbb{P}_\omega$ into an NPC-space. The least squares mean is again the unique solution of the Karcher equation. The Karcher mean of the manifold $\mathbb{P}$ of all positive operators is then the continuous extension via directed strong limits of the least squares mean on the Riemannian manifold $\mathbb{P}_\omega$ of extended Hilbert-Schmidt positive operators.

References

1. T. Ando, C.-K. Li, and R. Mathias, Geometric means, Linear Algebra and Appl. 385(2004), 305–334.

2. M. Bačák, Computing Medians and Means In Hadamard spaces, preprint.

3. R. Bhatia and J. Holbrook, Riemannian geometry and matrix geometric means, Linear Algebra Appl. **413** (2006), 594-618.

4. J. Lawson and Y. Lim, Monotonic properties of the least squares mean, Math. Ann. **351** (2011), 267-279.

5. J. Lawson and Y. Lim, Karcher means and Karcher equations of positive definite operators, to appear Trans. Amer. Math. Soc.

6. Y. Lim and M. Pálfia, The matrix power means and the Karcher mean, J. Funct. Anal. **262** (2012), 1498-1514.