数理解析研究所講究録
第 1737 巻 2011 年 56-64

MEANS, METRICS, AND MEASURES

JIMMIE LAWSON

ABSTRACT. We survey some recently introduced concepts and techniques that have been applied to settle affirmatively an open problem of several years standing of whether the least-squares mean for positive definite matrices is monotone for the usual (Loewner) order. The approach involves treating the set of positive definite matrices equipped with the trace metric as a nonpositively curved (NPC) metric space and applying the probability and random variable theory of such spaces to the problem at hand. These 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 means and extend our results to this setting. Least squares mean, positive definite matix, monotonicity, metric nonpositive curvature, symmetric cone, Loewner-Heinz space, metric random variables, barycenter [2000]15A48, 53C70, 60B05, 60G50, 52A55

1. INTRODUCTION

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, estimation, and averaging, the last being the principal 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* δ, the distance metric arising from the natural Riemannian structure on $\mathbb{P}$ making it a

Date: December, 2010.

Riemannian manifold, indeed a symmetric space, of negative curvature:

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

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

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 [12] and then Bhatia and Holbrook [5] suggested the *least squares mean*, taking the mean to be the unique minimizer of the sum of the squares of the distances:

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

This idea had been anticipated by Élie Cartan (see, for example, Section 6.1.5 of [3], 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 sometimes called the *Karcher mean* in light of its appearance in his work on Riemannian manifolds [9]. Indeed, he considered a *weighted least squares mean*:

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

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

In [2] 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 properties that this extended mean satisfied. Both contributions were important and have been influential in subsequent developments. In light of this paper it is natural to the the following

Question. *Are the Ando-Li-Mathias properties valid for the least squares mean?*

In particular, Bhatia and Holbrook asked whether the least squares n-mean was monotonic in each of its arguments. Computer calculations indicated "Yes." The answer is indeed "yes," as has recently been shown in [11], but showing it required new tools:

the theory of nonpositively curved metric spaces, techniques from probability and random variable theory, and the recent combination of the two, particularly by K.-T. Sturm [14].

2. NPC SPACES

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 $z \in M$

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

Such spaces are also called (global) CAT(0)-spaces or Hadamard spaces. Equation (NPC) is sometimes referred to as the *semiparallelogram law*, since it is a reformulation of the parallelogram law in Hilbert spaces with the equality replaced by an inequality (see, e.g., [10]). It is satisfied by the length metric in any simply connected nonpositively curved Riemannian manifold. Hence the metric definition yields a metric generalization of nonpositive curvature. We record the important

Fact. *The trace metric on the Riemannian symmetric space of positive definite matrices is a particular and important example of an NPC space.*

The theory of such NPC spaces is quite extensive (see, e.g., [7]). In particular the m appearing in

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

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 the defining property $d(\gamma(t), \gamma(s)) = |t-s|d(x,y)$. 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, $x\#y = y\#x$. Equation (NPR) admits a more general equivalent formulation in terms of the weighted mean. For all $0 \leq t \leq 1$ we have

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

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

$$\mathfrak{G}_n(w_1,\ldots,w_n;a_1,\ldots,a_n):=\underset{z\in M}{\arg\min}\sum_{i=1}^{n}w_id^2(z,a_i), \tag{2.1}$$

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.

One other mean will play an important role in what follows, one that we shall call the *inductive mean* following the terminology of Sturm. It appeared earlier in the work of [13, 1]. It is defined inductively for NPC spaces (or more generally for metric spaces with weighed means $x\#_ty$) for each $k\geq 2$ by $S_2(x,y)=x\#y$ and for $k\geq 3$, $S_k(x_1,\ldots,x_k)=S_{k-1}(x_1,\ldots,x_{k-1})\#_{\frac{1}{k}}x_k$.

The new proof strategy for showing monotonicity and other properties for the least squares mean consists of two major steps.
Step 1: Using induction one shows that a cer tain property (e.g. monotonicity) of the mean $x\#_ty$ carries over to the inductive mean S_n. This is often possible since the inductive mean is dened directly from the weighted 2-mean.
Step 2: One shows that the property in question transfers from the inductive mean to the least squares mean by using Sturms theorem that the least squares mean $\mathfrak{G}$ is the pointwise limit $\lim_{n\to\infty}S_n$ a.e.

3. Metric-valued random variables

In recent years signicant 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 [8] and Sturm [14].

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 separable metric space (M,d), an M-valued random variable is a function $X:\Omega\to M$ that is measurable in the sense that $X^{-1}(B)?\mathcal{A}$ for every

Borel subset B of M. The *push-forward* of the measure P by a random variable $X : \Omega \to M$ is denoted and dened 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 nite $F \subseteq I$, $P(\bigcap_{i\in F} X_i^{-1}(B_i)) = \prod_{i\in F} P(X^{-1}(B_i))$, 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.

We assume henceforth that (M, d) is a separable NPC-space. Let $\mathcal{P}(M)$ denote the set of probability measures 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 ($= L^1$) if

$$\int_\Omega d(z, X(\omega)) \, P(d\omega) = \int_M d(z, x) \, q_X(dx) < \infty \text{ for } z \in M.$$

Following Sturm [14], we dene the barycenter $b(q)$ by

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

for $q \in \mathcal{P}^2(M)$. 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.

Remark 3.1. 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

$$b(q) = \arg\inf_z \int_M d^2(z, x)q(dx) = \arg\inf_z \sum_{i=1}^n w_i d^2(z, x_i) = \mathfrak{G}_n(w_1, \ldots, w_n; x_1, \ldots, x_n).$$

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

$$EX = \arg \inf_{z \in M} \int_\Omega d^2(z, X(\omega))\, \sigma(d\omega) = \arg\min_{z \in M} \int_M d^2(z, x)\, q_X(dx) = b(q_X).$$

From this definition it is clear that 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 $\{X_n : n \in \mathbb{N}\}$ be a sequence of independent, identically distributed random variables on some probability space $(\Omega, \mathcal{A}, \sigma)$ into M. Let μ_n be an n-mean on M for each n, for example one obtained by the symmetrization procedure or least squares. We use these means to form the "average" Y_n of the given random variables according to the rule $Y_n(\omega) := \mu_n(X_1(\omega), \ldots, X_n(\omega))$. Now under suitable hypotheses Es-Sahib and Heinich [8] and Sturm [14] show that a strong law of large numbers is satisfied, that is, the Y_n converge pointwise a.e. to a common point b. The principal result of Sturm [14, Theorem 4.7] is crucial for our purposes.

Theorem 3.2. *Let $\{X_n\}_{n \in \mathbb{N}}$ be a sequence of bounded i.i.d. random variables from a probability space $(\Omega, \mathcal{A}, \sigma)$ into an NPC space M. Let S_n denote the inductive mean for each $n \geq 2$, and set $Y_n(\omega) = S_n(X_1(\omega), \ldots, X_n(\omega))$. Then $Y_n(\omega) \to EX_1$ as $n \to \infty$ for almost all $\omega \in \Omega$.*

4. Loewner-Heinz spaces

The fundamental Loewner-Heinz inequality for positive definite matrices asserts that $A^{1/2} \leq B^{1/2}$ whenever $A \leq B$. This can be written alternatively as $A \# I \leq B \# I$ whenever $A \leq B$ and extends to the equivalent monotonicity property that $A_1 \# A_2 \leq B_1 \# B_2$ whenever $A_1 \leq B_1$ and $A_2 \leq B_2$. These considerations motivate us to define a *Loewner-Heinz NPC space* as an NPC space equipped with a closed partial order $\leq$ satisfying $x_1 \# x_2 \leq y_1 \# y_2$ whenever $x_i \leq y_i$ for $i = 1, 2$. (Recall that a partial order on a topological space X is closed if $\{(x, y) : x \leq y\}$ 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 $x_i \leq y_i$ for $i = 1, \ldots, n$ implies $\mu(x_1, \ldots, x_n) \leq \mu(y_1, \ldots, y_n)$.

From the Loewner-Heinz property and by induction one obtains that the inductive mean S_n is order-preserving. This, together with Sturm's theorem yields the

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

Method of proof: Assume $x_i \leq y_i$ for $1 \leq i \leq n$. Let Ω_k be a copy of the n-element set $\{\xi_1, ..., \xi_n\}$ equipped with the measure $\sum_{i=1}^n w_i \delta_{\xi_i}$ and $\Omega = \prod_{k=1}^\infty \Omega_k$ be the countable product of the Ω_k with the product measure. Let $X_k : \Omega \to M$ be defined by $X_k(\omega) = x_i$ if $\pi_k(\omega) = \xi_i$, where $\pi_k : \Omega \to \Omega_k$ is projection into the kth-coordinate. Similarly we define $\tilde{X}_k : \Omega \to M$ by $\tilde{X}_k(\omega) = y_i$ if $\pi_k(\omega) = \xi_i$. Then $\{X_k\}$ 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 $\sum_{i=1}^n w_i \delta_{y_i}$. We note that $(X_1(\omega), \ldots, X_k(\omega))$ is coordinatewise less than or equal to $(\tilde{X}_1(\omega), \ldots, \tilde{X}_k(\omega))$ since $x_i \leq y_i$ for each $i = 1, \ldots, n$.

We define $Y_k, \tilde{Y}_k : \Omega \to M$ by $Y_k(\omega) = S_k(X_1(\omega), \ldots, X_k(\omega))$ and $\tilde{Y}_k(\omega) = S_k(\tilde{X}_1(\omega), \ldots, \tilde{X}_k(\omega))$. By monotonicity of the inductive mean $Y_k(\omega) \leq \tilde{Y}_k(\omega)$ for each $\omega \in \Omega$. By Sturm's Theorem we have that $\lim_{k\to\infty} Y_k(\omega) = \mathfrak{G}_n(\mathbf{w}; x_1, \ldots, x_n)$ a.e. and $\lim_{k\to\infty} \tilde{Y}_k(\omega) = \mathfrak{G}_n(\mathbf{w}; y_1, \ldots, y_n)$ a.e. By the closedness of the partial order $\mathfrak{G}_n(\mathbf{w}; x_1, \ldots, x_n) \leq \mathfrak{G}_n(\mathbf{w}; y_1, \ldots, y_n)$.

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. [10]), we have the following

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

5. Other properties and open problems

Using the techniques of the previous section, we can establish other basic properties of the least squares mean.

(1) The least squares mean is jointly concave.

(2) The weighted least squares mean is bounded above the the corresponding weighted arithmetic mean and below by the correponding harmonic mean.

(3) The least squares mean satisfies the Busemann-type inequality

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

One loses the NPC property of the metric when one passes to positive definite operators on an infinite dimensional Hilbert space. The question then arises whether some weakened version of Sturm's theorem remains valid. If the inductive means converge a.e., then the limit would be a natural candidate for a generalization of the least squares mean to this more general setting and the techniques under consideration should apply to deduce some of its basic properties.

In the finite-dimensional setting, it is only in the case of averaging with the inductive mean that the limit via the Law of Large Numbers is known. Can one say anything if one averages instead with the Ando-Li-Mathias mean or the Bini-Meini-Poloni mean [6]?

References

[1] E. Ahn, S. Kim, and Y. Lim, An extended Lie-Trotter formula and its applications, Linear Algebra Appl. **427** (2007), 190-196.

[2] T. Ando, C.K. Li and R. Mathias, Geometric means, Linear Algebra Appl. **385** (2004), 305–334.

[3] M. Berger, *A Panoramic View of Riemannlan Geometry*, Springer Verlag, 2003.

[4] R. Bhatia and J. Holbrook, Noncommutative geometric means, Math. Intelligencer **28** (2006), 32-39.

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

[6] D. Bini, B. Meini and F. Poloni, An effective matrix geometric mean satisfying the Ando-Li-Mathias properties, Math. Comp. **79**, (2010), 437-452.

[7] M. Bridson and A. Haefliger, *Metric Spaces of Non-Positive Curvature*, Springer, Berlin, 1999.

[8] A. Es-Sahib and H. Heinich, Barycentre canonique pour un espace métrique à courbure négative, Séminaire de probabilités (Strasbourg) **33** (1999), 355–370.

[9] H. Karcher, Riemannian center of mass and mollifier smoothing, Comm. Pure Appl. Math. **30** (1977), 509–541.

[10] J. D. Lawson and Y. Lim, The geometric mean, matrices, metrics, and more, Amer. Math. Monthly **108** (2001), 797-812.

[11] —, Order monotonicity for the least squares mean, to appear Math. Annalen.

[12] M. Moakher, A differential geometric approach to the geometric mean of symmetric positive-definite matrices, SIAM J. Matrix Anal. Appl. **26** (2005), 735-747.

[13] M. Sagae and K. Tanabe, Upper and lower bounds for the arithmetic-geometric-harmonic means of positive definite matrices, Linear and Multilinear Algebras **37** (1994), 279-282.

[14] K.-T. Sturm, Probability measures on metric spaces of nonpositive curvature, in: Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces, Eds. P. Auscher et. al., Contemp. Math. 338, Amer, Math. Soc. (AMS), Providence, 2003.

DEPARTMENT OF MATHEMATICS, LOUISIANA STATE UNIVERSITY, BATON ROUGE, LA70803, USA

E-mail address: `lawson@math.lsu.edu`