Ando-Hiai Inequality for Probability Measures

Yongdo Lim
joint work with Sejong Kim and Hosoo Lee

September 8, 2016

Overview

The Ando-Hiai inequality of positive definite matrices A, B and $t \in [0, 1]$ is given by

$$A \#_t B \leqslant I \quad \Longrightarrow \quad A^p \#_t B^p \leqslant I, \;\; p \geqslant 1.$$

In this talk we present an extension of the Ando-Hiai inequality for L^2-probability measures and show that it is closely related to a characteristic property (called Yamazaki's inequality) of the Cartan barycenter among other invariant barycenters on the cone of positive definite matrices equipped with the Riemannian trace metric.

In fact, it is known that there are infinitely many distinct such barycenteric maps and that the center of gravity, also called the Cartan barycenter, is a canonical example.

Contractive barycenters

An isometry invariant barycentric map on the space of probability measures on a metric space with finite first or second moment with respect to the Wasserstein distance (Kantorovich-Rubinstein) plays a fundamental role in the fields of metric/ geometric/convex analysis, statistical analysis, probability measure theory, optimal transport theory, to cite only a few.

K.-T. Sturm develops a theory of barycenters of probability measures for metric spaces of nonpositive curvature, particularly that class of metric spaces known as CAT(0)-spaces. The canonical barycenter on a CAT(0)-space (M, d) is the Cartan barycenter; the unique point $\Lambda(\mu)$ that minimizes the variance function

$$z \mapsto \int_M d^2(z, x) d\mu(x).$$

It turns out that the Cartan barycentric map contracts the Wasserstein metric, a property that has been called the fundamental contraction property.

Riemannian Trace Metric

In recent years, it has been increasingly recognized that the Euclidean distance is often not the most suitable for the space $\mathbb{P} = \mathbb{P}_m$ of $m \times m$ 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:

- $T_A(\mathbb{P}) = \mathbb{H}$ with $\langle X|Y\rangle_A = \mathrm{Tr}(A^{-1}XA^{-1}Y)$,
- [Riemannian trace distance]

$$d(A, B) = \|\log A^{-1}B\|_F = \left[\sum_{i=1}^{m} \log^2 \lambda_j (A^{-\frac{1}{2}} B A^{-\frac{1}{2}})\right]^{\frac{1}{2}},$$

where $\lambda_i(X)$ denotes the i-th eigenvalue of X.

Basic Properties

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

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

(3) (Congruence Invariance) Congruence transformations $X \mapsto CXC^*$ for invertible C are isometries of $\mathbb{P}$.

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

(5) $(\mathbb{P}, \delta)$ is complete and is an important example of Hadamard spaces(global NPC, Bruhat-Tits, CAT_0 spaces).

(6) $d^2(A\#_t B, C\#_t D) \leqslant$
$(1-t)d^2(A, C) + td^2(B, D) - (1-t)t[d(A, B) - d(C, D)]^2$.

(7) $X \mapsto d^2(X, A)$ is **uniformal convex** in Riemannian sense.

Wasserstein spaces

Let $\mathcal{B} := \mathcal{B}(\mathbb{P})$ be the algebra of Borel sets, the smallest σ-algebra containing the open sets of $\mathbb{P}$.

Let $\mathcal{P}$ be the set of all probability measures on $(\mathbb{P}, \mathcal{B})$ and $\mathcal{P}^0$ the set of all $\mu \in \mathcal{P}$ of the form $\mu = (1/n)\sum_{j=1}^{n} \delta_{A_j}$, where δ_A is the point measure of mass 1 at $A \in \mathbb{P}$.

For $p \in [1, \infty)$ let $\mathcal{P}^p$ be the set of probability measures with *finite* p*-moment*: for some (and hence all) $Y \in \mathbb{P}$,

$$\int_{\mathbb{P}} d^p(X, Y) d\mu(X) < \infty.$$

Note that

$$\mathcal{P}^0 \subset \mathcal{P}^q \subset \mathcal{P}^p \subset \mathcal{P}^1, \quad 1 \leqslant p \leqslant q < \infty.$$

Wasserstein spaces

We say that $\omega \in \mathcal{P}(\mathbb{P} \times \mathbb{P})$ is a *coupling* for $\mu, \nu \in \mathcal{P}$ and the μ, ν are *marginals* for ω if for all $B \in \mathcal{B}$

$$\omega(B \times \mathbb{P}) = \mu(B) \qquad \text{and} \qquad \omega(\mathbb{P} \times B) = \nu(B).$$

We denote the set of all couplings for $\mu, \nu \in \mathcal{P}$ by $\Pi(\mu, \nu)$.

The Wasserstein distance d_p^W on $\mathcal{P}^p$ is defined by

$$d_p^W(\mu_1, \mu_2) := \left[\inf_{\pi \in \Pi(\mu_1, \mu_2)} \int_{\mathbb{P} \times \mathbb{P}} d^p(X, Y) d\pi(X, Y) \right]^{\frac{1}{p}}.$$

It is known that d_p^W is a complete metric on $\mathcal{P}^p$ and $\mathcal{P}^0$ is dense in $\mathcal{P}^p$ and $d_p^W \leqslant d_q^W, \quad 1 \leqslant p \leqslant q < \infty$.

- For $\mu = \frac{1}{n} \sum_{j=1}^n \delta_{A_j}$, $\nu = \frac{1}{n} \sum_{j=1}^n \delta_{B_j}$, and $1 \leqslant p < \infty$

$$d_p^W(\mu, \nu) = \min_{\sigma \in S^n} \left[\frac{1}{n} \sum_{j=1}^n d^p(A_j, B_{\sigma(j)}) \right]^{1/p}$$

Isometric actions

For $M \in \mathrm{GL}_m$, $A \in \mathbb{P}$, $\mu \in \mathcal{P}^1$, $t \in \mathbb{R} \setminus \{0\}$, and $\mathcal{O} \in \mathcal{B}(\mathbb{P})$, we let

$$M.A := MAM^*, \quad M.\mathcal{O} = \{MAM^* : A \in \mathcal{O}\}, \quad \mathcal{O}^t := \{A^t : A \in \mathcal{O}\}$$

and

$$(M.\mu)(\mathcal{O}) = \mu(M^{-1}.\mathcal{O}), \qquad \mu^t(\mathcal{O}) := \mu(\mathcal{O}^{\frac{1}{t}}).$$

Example

The actions are natural and comparable with discrete measures:

$$\mu^t = \frac{1}{n}\sum_{j=1}^{n} \delta_{A_j^t}, \qquad \mu = \frac{1}{n}\sum_{j=1}^{n} \delta_{A_j} \in \mathcal{P}^0$$

$$M.\mu = \frac{1}{n}\sum_{j=1}^{n} \delta_{M.A_j} = \frac{1}{n}\sum_{j=1}^{n} \delta_{MA_jM^*}, \qquad M \in \mathrm{GL}_m.$$

Invariant and contractive barycenters

Definition

A map $\beta : \mathcal{P}^2 \to \mathbb{P}$ is said to be a *barycenter* if it is idempotent in the sense that $\beta(\delta_X) = X$ for all $X \in \mathbb{P}$. A barycentric map β is said to be contractive if

$$d(\beta(\mu_1), \beta(\mu_2)) \leqslant d_2^W(\mu_1, \mu_2), \quad \forall \mu_1, \mu_2 \in \mathcal{P}^2$$

and is said to be invariant if for all $M \in GL_m$ and $\mu \in \mathcal{P}^2$,

(i) $\beta(M.\mu) = M.\beta(\mu)$; and

(ii) $\beta(\mu^{-1}) = \beta(\mu)^{-1}$.

• A metric space admitting a contractive barycenter, called a barycentric metric space, is an important object to study in a variety of pure and applied areas. In fact, a complete, simply connected Riemannian manifold admits a contractive barycenter if and only if it has nonpositive curvature (Sturm)

Cartan Barycenter

The Cartan barycenter $\Lambda : \mathcal{P}^2 \to \mathbb{P}$ is defined by

$$\Lambda(\mu) = \operatorname*{arg\,min}_{Z \in \mathbb{P}} \int_{\mathbb{P}} d^2(Z, X) d\mu(X).$$

The uniqueness and existence of the minimizer is well known in general setting of Hadamard spaces (CAT(0 spaces).

Theorem (Fundamental Contraction Property, Sturm)

For $\mu, \nu \in \mathcal{P}^2$,

$$d(\Lambda(\mu), \Lambda(\nu)) \leqslant d_1^W(\mu, \nu) \leqslant d_2^W(\mu, \nu).$$

- For $\mu = \frac{1}{n} \sum_{j=1}^{n} \delta_{A_j}$,

$$\Lambda(\mu) = \operatorname*{arg\,min}_{X > 0} \sum_{j=1}^{n} w_j d^2(X, A_j).$$

Cartan Barycenter

The Cartan barycenter $\Lambda : \mathcal{P}^2 \to \mathbb{P}$ is defined by

$$\Lambda(\mu) = \arg\min_{Z \in \mathbb{P}} \int_{\mathbb{P}} d^2(Z, X) d\mu(X).$$

The uniqueness and existence of the minimizer is well known in general setting of Hadamard spaces.

• The Cartan barycenter Λ on $\mathcal{P}^2$ can be extended to $\mathcal{P}^1$: for $\mu \in \mathcal{P}^1$, the unique minimizer of the uniformly convex, continuous function

$$Z \mapsto \int_{\mathbb{P}} [d^2(Z, X) - d^2(Y, X)] d\mu(X).$$

This point is independent of Y and coincides with $\Lambda(\mu)$ for $\mu \in \mathcal{P}^2$.

• It is an invariant barycenter and there are infinitely many invariant and contractive barycenters on $\mathbb{P}$.

Karcher Equation

The Cartan barycenter on $\mathcal{P}^2$

$$\Lambda(\mu) = \underset{Z \in \mathbb{P}}{\arg\min} \int_{\mathbb{P}} d^2(Z, X) d\mu(X)$$

arises as the unique point where the gradient of the variance function

$$Z \mapsto \int_{\mathbb{P}} d^2(Z, X) d\mu(X)$$

vanishes:

$$\int_{\mathbb{P}} \log_Z(X) d\mu(X) = 0,$$

as general setting of Riemannian manifolds with nonpositive curvature (Karcher). The Cartan barycenter $\Lambda(\mu)$ is the unique positive definite solution Z of the Karcher equation

$$\int_{\mathbb{P}} \log(Z^{-1/2} X Z^{-1/2}) d\mu(X) = 0. \tag{1}$$

Monotonicity

Definition

A subset $\mathcal{U} \subset \mathbb{P}$ is called an upper set, if whenever $A \in \mathcal{U}$ and $A \leqslant B$, then $B \in \mathcal{U}$. This gives rise to a partial order on $\mathcal{P}^2$:

$$\mu \leqslant \nu \Longleftrightarrow \mu(\mathcal{U}) \leqslant \nu(\mathcal{U}), \quad \forall \text{upper set} \;\; \mathcal{U}.$$

- For $\mu = \frac{1}{n}\sum_{j=1}^{n} \delta_{A_j}$ and $\nu = \frac{1}{n}\sum_{j=1}^{n} \delta_{B_j}$,

$$A_j \leqslant B_j, \forall j \quad \Longrightarrow \quad \mu \leqslant \nu.$$

Theorem (Kim and Lee)

$$\Lambda(\mu) \leqslant \Lambda(\nu) \;\; \text{for} \;\; \mu \leqslant \nu.$$

This extends that for discrete measures, which is known as the monotonicity conjecture suggested by Bhatia and Holbrook and settled by Lawson and Lim.

Inequality Characterizing Barycenters

Theorem
Let $\beta : \mathcal{P}^2 \to \mathbb{P}$ be an invariant barycenter on $\mathbb{P}$ satisfying

$$\int_{\mathbb{P}} \log X \, d\mu(X) \leqslant 0 \quad \text{implies} \quad \beta(\mu) \leqslant I \tag{2}$$

for all $\mu \in \mathcal{P}^2$. Then $\beta = \Lambda$. Moreover, Λ satisfies (2).

• (Yamazaki Inequality) $\mu = (1/n) \sum_{j=1}^n \delta_{A_j} \in \mathcal{P}^0$,

$$\frac{1}{n} \sum_{j=1}^n \log A_j \leqslant 0 \Longrightarrow \Lambda(A_1, \ldots, A_n) \leqslant I.$$

• Palfia and Lim; Kracher mean characterization on $\mathcal{P}^0$.

• Key tools: Karcher equation, Monotonicity, Fundamental Contraction Property.

Key Steps

Let $\mu \in \mathcal{P}^2$. Assume that $\int_{\mathbb{P}} \log(X) d\mu(X) \leqslant 0$.

Step 1. By Löwner ordering, there exists $A \geqslant I$ such that

$$\int_{\mathbb{P}} \log(X) d\mu_A(X) = \frac{1}{2} \int_{\mathbb{P}} \log(X) d\mu(X) + \frac{1}{2} \log(A) = 0,$$

where $\mu_A := \frac{1}{2}\mu + \frac{1}{2}\delta_A \in \mathcal{P}^2$. By the Karcher equation, $\Lambda(\mu_A) = I$.
(Note $A \leqslant B \Longrightarrow \mu_A \leqslant \mu_B$)

Step 2. The sequence G_k on $\mathbb{P}$ defined inductively by

$$G_0 = \Lambda(\mu_I) \text{ for } \mu_I = \frac{1}{2}\mu + \frac{1}{2}\delta_I \quad \text{and} \quad G_{k+1} = \Lambda(\mu_{G_k}), \quad k = 0, 1, \ldots$$

satisfies $0 < G_k \leqslant \cdots \leqslant G_1 \leqslant G_0 \leqslant I$.

Step 3. $\lim_{k\to\infty} G_k = \Lambda(\mu)$. Then $\Lambda(\mu) = \lim_{k\to\infty} G_k \leqslant I$.

Ando-Hiai Inequality

Recall the Ando-Hiai inequality

$$A\#_t B \leqslant I \quad \Longrightarrow \quad A^p\#_t B^p \leqslant I, \quad t \in [0,1],\ p \geqslant 1.$$

- For $\mu = (1-t)\delta_A + t\delta_B$ with $t \in [0,1]$,

$$\Lambda(\mu) = A\#_t B \quad \text{and} \quad \mu^p = (1-t)\delta_{A^p} + t\delta_{B^p}.$$

Corollary

$$\Lambda(\mu) \leqslant I \Longrightarrow \Lambda(\mu^p) \leqslant I, \forall p \geqslant 1, \quad \mu \in \mathcal{P}^2.$$

- For $\mu \in \mathcal{P}^2$,

$$\int_{\mathbb{P}} \log X d\mu(X) \leqslant 0 \quad \text{implies} \quad \Lambda(\mu^p) \leqslant I, \forall p > 0.$$

The reverse implication holds true for $\mu \in \mathcal{P}^0$ (Yamazaki):

$$\int_{\mathbb{P}} \log X d\mu(X) \leqslant 0 \quad \text{if and only if} \quad \Lambda(\mu^p) \leqslant I, \forall p > 0.$$

Norm Inequality

Theorem

Let $\mu \in \mathcal{P}^2$*, and let* $0 < p \leqslant q$*. Then*

$$\Lambda(\mu^p) \leqslant I \quad \text{implies} \quad \Lambda(\mu^q) \leqslant I.$$

Moreover for the operator norm $\|\cdot\|$,

$$\|\Lambda(\mu^q)^{\frac{1}{q}}\| \leqslant \|\Lambda(\mu^p)^{\frac{1}{p}}\|.$$

- For $\mu = \frac{1}{n}\sum_{j=1}^{n} \delta_{A_j} \in \mathcal{P}^0$ and $0 < p \leqslant q$,

$$\Lambda(\mu^q)^{\frac{1}{q}} \underset{(\log)}{\prec} \Lambda(\mu^p)^{\frac{1}{p}}$$

and therefore for all unitarily invariant norms $|||\cdot|||$,

$$|||\Lambda(\mu^q)^{\frac{1}{q}}||| \leqslant |||\Lambda(\mu^p)^{\frac{1}{p}}|||, \qquad 0 < p \leqslant q.$$

Golden- Thompson Inequality

Using the fact that

$$\lim_{p\to 0^+} \Lambda(\mu^p)^{\frac{1}{p}} = \exp((1/n)\sum_{j=1}^{n} \log A_j),$$

we obtain

$$\lim_{p\to 1^-} |||\Lambda(\omega;\mu^{\frac{1}{p}})^p||| = |||\Lambda(\mu)||| \leqslant \lim_{p\to 0^+} |||\Lambda(\mu^p)|||^{\frac{1}{p}} = |||e^{(1/n)\sum_{j=1}^{n}\log A_j}|||,$$

which is a multivariate version of complementary Golden-Thompson inequality and settles a question of Bhatia and Grover. It remains open for general $\mu \in \mathcal{P}^2$. An appropriate version in the setting of $\mu \in \mathcal{P}^2$ is

$$|||\Lambda(\mu)||| \leqslant \lim_{p\to 0^+} |||\Lambda(\mu^p)|||^{\frac{1}{p}} = |||e^{\int_{\mathbb{P}} \log X \, d\mu(X)}|||.$$

Symmetric cones and Hilbert-Schmidt operators

The derived results are valid for a general class of partially ordered symmetric spaces of non-compact type, namely the class of symmetric cones. Symmetric cones, also called domains of positivity, are open convex self-dual cones in Euclidean space which have a transitive group of symmetries. By the Koecher-Vinberg theorem these correspond to the cone of squares in Euclidean Jordan algebras, originally classified by Jordan, von Neumann and Wigner.

Also for the Riemannian manifold of positive operators in the extended Hilbert-Schmidt algebra of linear operators on a Hilbert space equipped with the canonical trace metric.

Further work

Although the order inequality characterizing invariant barycenters is new and quite attractive, particularly in the theory of matrix analysis, analysis on symmetric cones and probability measures, it depends heavily on the Karcher equation (also, monotonicity and Sturm's fundamental contraction theorem for the Cartan barycenter) and it does not carry over in the $\mathcal{P}^1$-setting. The Cartan barycener on $\mathcal{P}^1$ is also an important object in related research areas and so finding a (order inequality) characterizing property still remains open in our context.

Infinite dimensional setting

The *Karcher mean* $\Lambda = \{\Lambda_n\}$ of positive operators on a Hilbert space is defined as the unique solution in $\mathbb{P}$ of the Karcher equation

$$X = \Lambda_n(A_1, \ldots A_n) \Leftrightarrow \sum_{j=1}^{n} \log(X^{-1/2} A_j X^{-1/2}) = 0.$$

Theorem

There exists a uniquely determined contractive (*for the Thompson metric* $d(A, B) = \| \log A^{-1} B \|$) *barycentric map*

$$\beta : \mathcal{P}^1(\mathbb{P}) \to \mathbb{P}$$

satisfying

$$\beta((1/n) \sum_{i=1}^{n} \delta_{A_i}) = \Lambda_n(A_1, \ldots, A_n).$$

- Ando-Hiai inequality/Yamazaki inequality for β?

Conjecture

For $\mu \in \mathcal{P}^2(\mathbb{P})$, the equation (Karcher equation)

$$\int_{\mathbb{P}} \log(Z^{-1/2} X Z^{-1/2}) d\mu(X) = 0. \tag{3}$$

has a unique positive definite solution.

- (Lawson and Lim) It is for $\mu \in \mathcal{P}^0(\mathbb{P})$

References

(1) T. Yamazaki, The Riemannian mean and matrix inequalities related to the Ando-Hiai inequality and chaotic order, Operators and Matrices (2012).

(2) Y. Lim and M. Miklos, Matrix power means and the Karcher mean, J. Funct. Anal.

(3) S. Kim, H. Lee and Y. Lim, An order inequality characterizing invariant barycenters on symmetric cones, J. Math. Anal. Appl. 2016.

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

(5) —, Weighted means and Karcher equations of positive operators, Proc. Natl. Acad. Sci. USA. **110** (2013), 15626-15632.

(6) —, The least squares mean of positive Hilbert?Schmidt operators, J. Math. Anal. Appl. 2013.

(7) —, Karcher means and Karcher equations of positive definite operators, Trans. Amer. Math. Soc.(2014).