J. OPERATOR THEORY
77:1(2017), 87–107
doi: 10.7900/jot.2015dec24.2111

CONTRACTIVE BARYCENTRIC MAPS

JIMMIE D. LAWSON and YONGDO LIM

Communicated by Marius Dădârlat

ABSTRACT. We first develop in the context of complete metric spaces a one-to-one correspondence between the class of means $G = \{G_n\}_{n\geqslant 2}$ that are symmetric, multiplicative, and contractive and the class of contractive (with respect to the Wasserstein metric) barycentric maps on the space of L^1-probability measures. We apply this equivalence to the recently introduced and studied Karcher mean on the open cone $\mathbb{P}$ of positive invertible operators on a Hilbert space equipped with the Thompson metric to obtain a corresponding contractive barycentric map. In this context we derive a version of earlier results of Sturm and Lim and Palfia about approximating the Karcher mean with the more constructive inductive mean. This leads to the conclusion that the Karcher barycenter lies in the strong closure of the convex hull of the support of a probability measure. This fact is a crucial ingredient in deriving a version of Jensen's inequality, with which we close.

KEYWORDS: *Positive operator, operator mean, barycentric map, Karcher mean, geodesic metric space, Jensen's inequality.*

MSC (2010): Primary 47B65; Secondary 47L07, 15A48.

INTRODUCTION

In [28] 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 $\mathrm{CAT}(0)$-spaces or alternatively Hadamard spaces. For these spaces one has available a method for "averaging" a finite set of points, or more generally finding the barycenter of a probability measure, via an approach stretching back to Cartan by finding the point that minimizes the sum or integral of the distances squared to the point. A particularly important example of a Hadamard space is the open cone of positive definite matrices equipped with the trace metric. Since this cone forms a Hadamard space, one can use the least squares mean for averaging, which can be useful in a variety of applications where positive definite matrices appear as "data points".

Our main goal in this paper is to extend this theory to the integrable measures on the open cone of positive operators on an arbitrary Hilbert space. The appropriate metric to consider in this case is the Thompson metric, but for this metric one no longer has unique minimizers. However, in this setting the Karcher equation, which in the finite-dimensional case captures the fact that the gradient of the least squares mapping vanishes, still has a unique solution that defines a mean retaining most of the properties of the Cartan mean [22], [23]. We develop in Section 3 general methods of extending means to barycentric maps on measures, which apply in the setting of the open cone of positive operators (Section 5). We suggest that the method we introduce of obtaining barycentric maps via means or finitely supported measures is a useful technique in a much broader range of settings.

Sturm [28] uses an approximation of the least squares or Cartan mean via the inductive mean, a mean built up inductively and constructively from the two-variable mean, to drive his development of the theory of the least squares barycenter. We show, using a key result of Lim and Palfia [25], that one can do a similar approximation in the strong topology for the Karcher mean (Section 6). To demonstrate the usefulness of this result we derive some results about convexity and the Karcher mean with the main application being a derivation of a version of Jensen's inequality in this setting.

A key distinctive feature of our considerations is the fact that although the open cone of positive operators equipped with the Thompson metric is a geodesic metric space, the geodesics are no longer unique. Our approach is to identify for each pair of points a distinguished geodesic between the points and work with this family of geodesics. We think that this setting, metric spaces with distinguished geodesics, is important to consider, because it is often the situation one encounters in the infinite-dimensional setting.

1. METRIC SPACES

A *path* in a metric space (X, d) is a continuous map $\alpha : I \to X$, where $I = [a, b]$ is some closed interval in $\mathbb{R}$. Given any partition of $a = t_0 < t_1 < \cdots < t_n = b$ of $[a, b]$, one can form the sum $\sum_{k=1}^{n} d(\alpha(t_{k-1}), \alpha(t_k))$, and the length $L(\alpha)$ is the supremum of such sums over all partitions of $[a, b]$. The path α is called a *geodesic path* if $L(\alpha|_{[c,d]}) = d(\alpha(c), \alpha(d))$ for all subintervals $[c, d]$ of I. From the triangle inequality geodesics are paths of minimal length, namely the distance $d(\alpha(a), \alpha(b))$. Any geodesic path can be reparametrized by arc length, and the resulting parametrization is an isometry. The image of a geodesic path is called a *geodesic segment*. We note that one can define geodesic paths and segments more generally that satisfy locally the preceding conditions, in which case what we have defined are then called minimal geodesic paths and segments, but we will have no need of the more general notion.

The space X is called a *geodesic space* if for any two points $x, y \in X$, there exists a geodesic path α defined on some interval $I = [a, b]$ such that $\alpha(a) = x$ and $\alpha(b) = y$. In general one may have a myriad of geodesic segments between two points. A *geodesic selector* σ chooses for each (x, y) a geodesic path $\sigma_{x,y} : [0, 1] \to X$ such that $\sigma_{x,y}(0) = x$ and $\sigma_{x,y}(1) = y$. We alternatively write $\sigma_{x,y}(t)$ as $x\#_t y$ and $\sigma_{x,y}(1/2)$, a midpoint, simply as $x\#y$. A geodesic metric space equipped with a geodesic selector is called a *distinguished geodesic space*. A selector is said to be *symmetric* if for all (x, y), $\sigma_{x,y}(t) = \sigma_{y,x}(1-t)$, i.e., $x\#_t y = y\#_{1-t}x$. In particular $x\#y = y\#x$ in this case.

Of course the important case in which a geodesic space has a unique geodesic segment between any two points is an example of a distinguished geodesic space. An important specific example is that of *Hadamard spaces*, complete metric spaces satisfying for all x, y, z the semiparallelogram law

$$d^2(x, y) \leqslant \frac{1}{2}d^2(x, z) + \frac{1}{2}d^2(y, z) - \frac{1}{4}d^2(x\#y, z),$$

where $x\#y$ is the midpoint, necessarily unique, between x and y. In the literature these are often referred to as (global) CAT(0)-spaces or spaces of nonpositive curvature (NPC), see K.-T. Sturm [28] for the latter terminology.

For a metric space X, let $\mathcal{B}(X)$ be the algebra of Borel sets, the smallest σ-algebra containing the open sets. Let $\mathcal{P}(X)$ be the set of all probability measures on $(X, \mathcal{B}(X))$ with support that is separable and has measure 1 and $\mathcal{P}_0(X)$ the set of all $\mu \in \mathcal{P}(X)$ of the form $\mu = (1/n) \sum_{j=1}^{n} \delta_{x_j}$ with $n \in \mathbb{N}$, where δ_x is the point measure of mass 1 at x.

REMARK 1.1. It is known, apparently not widely so, that in any metric space X the support of a Borel probability measure, the points for which each neighborhood has positive measure, is separable. Additionally the support has measure 1 if the metric space is separable, but for general metric spaces one needs to require that this be true.

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

$$\int_X d^p(x, y)\mathrm{d}\mu(y) < \infty.$$

For $p = \infty$, $\mathcal{P}^\infty(X)$ denotes the set of probability measures with bounded separable support.

For metric spaces X and Y, a continuous $f : X \to Y$ induces a *push-forward* map $f_* : \mathcal{P}(X) \to \mathcal{P}(Y)$ defined by $f_*(\mu)(B) = \mu(f^{-1}(B))$ for $\mu \in \mathcal{P}(X)$ and $B \in \mathcal{B}(Y)$. Note that $\mathrm{supp}(f_*(\mu)) = f(\mathrm{supp}(\mu))^-$, the closure of the image of the support of μ; in particular $f_*(\mu)$ has separable support.

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

$$\omega(B \times X) = \mu(B) \quad \text{and} \quad \omega(X \times B) = \nu(B).$$

Equivalently μ and ν are the push-forwards of ω under the projection maps π_1 and π_2 respectively. We note that one such coupling is the product measure $\mu \times \nu$, and that for any coupling ω it must be the case that $\operatorname{supp}(\omega) \subseteq \operatorname{supp}(\mu) \times \operatorname{supp}(\nu)$. We denote the set of all couplings by $\Pi(\mu, \nu)$.

For $1 \leqslant p < \infty$, the p-Wasserstein distance $\mathfrak{w}_p$ (alternatively Kantorovich–Rubinstein distance) on $\mathcal{P}^p(X)$ is defined by

$$\mathfrak{w}_p(\mu_1, \mu_2) := \Big(\inf_{\pi \in \Pi(\mu_1, \mu_2)} \int_{X \times X} d^p(x, y) \mathrm{d}\pi(x, y) \Big)^{1/p}.$$

It is known that $\mathfrak{w}_p$ is a complete metric on $\mathcal{P}^p(X)$ whenever X is a complete metric space and $\mathcal{P}_0(X)$ is $\mathfrak{w}_p$-dense in $\mathcal{P}^p(X)$ [5], [28]. Furthermore, it follows from the Hölder inequality that $\mathfrak{w}_p \leqslant \mathfrak{w}_{p'}$ whenever $p \leqslant p'$. The last observation makes possible the definition of $\mathfrak{w}_\infty(\mu_1, \mu_2) = \lim_{p \to \infty} \mathfrak{w}_p(\mu_1, \mu_2)$ on $\mathcal{P}^\infty(X)$. The limit is finite on the bounded measures and yields a complete metric space. Alternatively the ∞-metric is given by

$$\mathfrak{w}_\infty(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \sup\{d(x, y) : (x, y) \in \operatorname{supp}(\pi)\}. \tag{1.1}$$

For the following see the introduction of [30], also [4], [6], [26].

EXAMPLE 1.2. For $\mu = (1/n) \sum_{j=1}^{n} \delta_{x_j}$, $\nu = (1/n) \sum_{j=1}^{n} \delta_{y_j}$, and $1 \leqslant p < \infty$

$$\mathfrak{w}_p(\mu, \nu) = \min_{\sigma \in S^n} \Big(\frac{1}{n} \sum_{j=1}^{n} d^p(x_j, y_{\sigma(j)}) \Big)^{1/p}$$

where S^n denotes the permutation group on n-letters. For the case $p = \infty$,

$$\mathfrak{w}_\infty(\mu, \nu) = \min_{\sigma \in S^n} \max\{d(x_j, y_{\sigma(j)}) : 1 \leqslant j \leqslant n\}.$$

LEMMA 1.3. *Let $f : X \to Y$ be a Lipschitz map with Lipschitz constant C. Then $f_* : \mathcal{P}^p(X) \to \mathcal{P}^p(Y)$ is Lipschitz with Lipschitz constant C for $1 \leqslant p \leqslant \infty$.*

Proof. We note first that for any coupling π of $\mu, \nu \in \mathcal{P}$, it is straightforward to verify that $(f \times f)_*(\pi)$ is a coupling of $f_*(\mu), f_*(\nu) \in \mathcal{P}(Y)$. Using a standard change of variables, we have for $\mu, \nu \in \mathcal{P}^p(X)$, $1 \leqslant p < \infty$,

$$\begin{aligned} \mathfrak{w}_p^p(f_*(\mu), f_*(\nu)) &= \inf_{\pi \in \Pi(f_*(\mu), f_*(\nu))} \int_{Y \times Y} d^p(y_1, y_2) \mathrm{d}\pi \\ &\leqslant \inf_{\pi \in \Pi(\mu, \nu)} \int_{Y \times Y} d^p(y_1, y_2) \mathrm{d}(f \times f)_*(\pi) \end{aligned}$$

$$= \inf_{\pi \in \Pi(\mu,\nu)} \int_{X\times X} d^p(f(x_1), f(x_2))\mathrm{d}\pi$$

$$\leqslant \inf_{\pi \in \Pi(\mu,\nu)} \int_{X\times X} C^p d^p(x_1, x_2)\mathrm{d}\pi = C^p \mathfrak{w}_p^p(\mu,\nu).$$

Taking the p-th root of both sides yields the desired result. For $p = \infty$ the lemma follows directly from equation (1.1). ∎

2. CONTRACTIVE MEANS AND BARYCENTERS

Throughout this section let (X, d) be a complete metric space. An *n-mean* G_n on X for $n \geqslant 2$ is a continuous map $G_n : X^n \to X$ that is idempotent in the sense that $G_n(x, \ldots, x) = x$ for all $x \in X$. An n-mean G_n is *symmetric* or *permutation invariant* if $G_n(\mathbf{x}_\sigma) = G_n(\mathbf{x})$, where $\mathbf{x}_\sigma = (x_{\sigma(1)}, \ldots, x_{\sigma(n)})$ for each permuation σ of $\{1, \ldots, n\}$. A (symmetric) *mean* G on X is a sequence of means $\{G_n\}$, one (symmetric) mean for each $n \geqslant 2$.

For $\mathbf{x} = (x_1, \ldots, x_n) \in X^n$, we let

$$\mathbf{x}^k = (x_1, \ldots, x_n, x_1, \ldots, x_n, \ldots, x_1, \ldots, x_n) \in X^{nk},$$

where the number of blocks is k. We define the *carrier* $S(\mathbf{x})$ of $\mathbf{x}$ to be the set of entries in $\mathbf{x}$, i.e., the smallest finite subset F such that $\mathbf{x} \in F^n$. We set $[\mathbf{x}]$ equal to the equivalence class of all n-tuples obtained by permuting the coordinates of $\mathbf{x} = (x_1, \ldots, x_n)$. Note that the operation $[\mathbf{x}]^k = [\mathbf{x}^k]$ is well-defined and that all members of $[\mathbf{x}]$ all have the same carrier set $S(\mathbf{x})$.

A tuple $\mathbf{x} = (x_1, \ldots, x_n) \in X^n$ *induces* a finitely supported probability measure μ on $S(\mathbf{x})$ by $\mu = \sum_{i=1}^{n} (1/n)\delta_{x_i}$, where δ_{x_i} is the point measure of mass 1 at x_i. Since the tuple may contain repetitions of some of its entries, each singleton set $\{x\}$ for $x \in \{x_1, \ldots, x_n\}$ will have measure k/n, where k is the number of times that it appears in the listing $x_1, \ldots, x_n$. Note that every member of $[\mathbf{x}]$ induces the same finitely supported probability measure.

LEMMA 2.1. *For each probability measure μ on X with finite support F for which $\mu(x) (= \mu(\{x\}))$ is rational for each $x \in F$, there exists a unique $[\mathbf{x}]$ inducing μ such that any $[\mathbf{y}]$ inducing μ is equal to $[\mathbf{x}]^k$ for some $k \geqslant 1$.*

Proof. For each $x \in F$, $\mu(x)$ is a positive rational number, which we may assume is reduced to lowest terms. Let n be the least common multiple of the denominators of $\mu(x)$ for $x \in F$, and let $\mu(x) = k_x/n$ for each $x \in F$. Let $\mathbf{x} \in F^n$ be chosen so that each $x \in F$ appears k_x times in $\mathbf{x}$. Then $[\mathbf{x}]$ induces μ.

Suppose that $\mathbf{y} \in F^m$ induces μ. If j_x is the number of times that $x \in F$, appears in $\mathbf{y}$, then it must be the case the $j_x/m = \mu(x) = k_x/n$. Thus m is a common multiple of the reduced denominators of the $\mu(x)$, $x \in F$, and hence a multiple of the least common multiple n, i.e., $m = qn$. It follows that $j_x = qk_x$ for each $x \in F$, and hence that $[\mathbf{x}]^q = [\mathbf{y}]$. ∎

DEFINITION 2.2. A mean $G = \{G_n\}$ on X is said to be *multiplicative* if for all $n, k \geqslant 2$ and all $\mathbf{x} = (x_1, \ldots, x_n) \in X^n$,

$$G_n(\mathbf{x}) = G_{nk}(\mathbf{x}^k).$$

If G is also symmetric, then G is called *intrinsic*.

We have the following corollary to Lemma 2.1.

COROLLARY 2.3. *Let G be an intrinsic mean. Then for any finitely supported probability measure μ with support F and taking on rational values, we may define $G(\mu) = G_n(\mathbf{x})$, for any $\mathbf{x} \in F^n$ that induces μ.*

Proof. By Lemma 2.1 such $\mathbf{x}$ exist and any such will yield the same result since G is intrinsic. ∎

The following notion of what we call a contractive mean has appeared in other work; see e.g. [17].

DEFINITION 2.4. An n-mean $G_n : X^n \to X$ is said to be *p-contractive* for $p \in [1, \infty)$ if

$$d(G_n(\mathbf{x}), G_n(\mathbf{y})) \leqslant \Big(\frac{1}{n}\sum_{j=1}^n d^p(x_j, y_j)\Big)^{1/p} = \frac{1}{n^{1/p}}\Big(\sum_{j=1}^n d^p(x_j, y_j)\Big)^{1/p} \tag{2.1}$$

for all $\mathbf{x} = (x_1, \ldots, x_n), \mathbf{y} = (y_1, \ldots, y_n) \in X^n$. The special case $p = 1$ we refer to simply as a *contractive* n-mean:

$$d(G_n(\mathbf{x}), G_n(\mathbf{y})) \leqslant \frac{1}{n}\sum_{j=1}^n d(x_j, y_j).$$

The limiting case of p-contractivity for $p = \infty$ is given by

$$d(G_n(\mathbf{x}), G_n(\mathbf{y})) \leqslant \max\{d(x_i, y_i) : 1 \leqslant i \leqslant n\} = \lim_{p\to\infty}\Big(\frac{1}{n}\sum_{j=1}^n d^p(x_j, y_j)\Big)^{1/p}.$$

A mean $G = \{G_n\}$ is *p-contractive* for $p \in [1, \infty]$ if each G_n is p-contractive.

We observe that our notion of a contractive mean implies that it is strictly contractive coordinatewise. Indeed G_n is p-contractive, $1 \leqslant p < \infty$, if and only if it is coordinatewise $(1/n^{1/p})$-contractive: for all $x, y, a_j \in X$ and $j = 1, \ldots, n$:

$$d(G_n(a_1, \ldots, a_{j-1}, x, a_{j+1}, \ldots, a_n), G_n(a_1, \ldots, a_{j-1}, y, a_{j+1}, \ldots, a_n)) \leqslant \frac{1}{n^{1/p}} d(x, y).$$

REMARK 2.5. Recall that for $a_1, \ldots, a_n \geqslant 0$, the *power means* are defined for $1 \leqslant p < \infty$ by

$$\mathcal{M}^p(a_1, \ldots, a_n) = \Big(\frac{1}{n}\sum_{j=1}^n a_i^p\Big)^{1/p}$$

and for $p = \infty$ by $\mathcal{M}^\infty(a_1, \dots a_n) = \max\{a_i : 1 \leqslant i \leqslant n\}$. In terms of these means the definition of a p-contractive mean $G_n : X_n \to X$ can be rewritten as

$$d(G_n(\mathbf{x}), G_n(\mathbf{y})) \leqslant \mathcal{M}^p(d(x_1, y_1), \dots, d(x_n, y_n)).$$

We find this a useful alternative notation.

We generalize the notion of Sturm [28] of a contractive barycentric map on the set of probability measures of finite first moment on a complete metric space.

DEFINITION 2.6. A barycentric map $\beta : \mathcal{P}^p(X) \to X$ is *p-contractive* if
(i) $\beta(\delta_x) = x$ for all x;
(ii) $d(\beta(\mu_1), \beta(\mu_2)) \leqslant \mathfrak{w}_p(\mu_1, \mu_2)$ for all $\mu_1, \mu_2 \in \mathcal{P}^p(X)$.

A complete metric space having a contractive barycentric map must be a geodesic metric space [28]. Indeed it is naturally a distinguished geodesic space with distinguished geodesics for $x, y \in X$ given by $t \mapsto x\#_t y := \beta((1-t)\delta_x + t\delta_y)$, a geodesic from x to y. Moreover, $\beta((1/2)\delta_x + (1/2)\delta_y)$ is a midpoint between x and y and the distance function $t \mapsto d(x\#_t y, a\#_t b)$ is convex. Sturm [28] has established an empirical law of large numbers from a contractive barycentric map on a complete metric space via "barycentric mean values".

A fundamental relationship exists between contractive intrinsic means and barycentric maps, as worked out in the following.

PROPOSITION 2.7. *In a complete metric space (X, d) for $1 \leqslant p < \infty$, there is a bijective correspondence Φ between the set of p-contractive intrinsic means G on X and the set of p-contractive barycentric maps β on $\mathcal{P}^p(X)$. The correspondence Φ is uniquely determined by the requirement that $\Phi(G) = \beta$ precisely when for all $(x_1, \dots, x_n) \in X^n$ $G_n(x_1, \dots, x_n) = \beta\Big((1/n) \sum_{j=1}^{n} \delta_{x_j}\Big)$. The result remains valid for $p = \infty$ if one restricts to the domain of β to $\mathcal{P}_{\mathrm{cpt}}(X)$, the space of Borel measures with compact support.*

Proof. Let $G = \{G_n\}$ be a p-contractive and intrinsic mean for $1 \leqslant p < \infty$. Let $\mu = (1/k) \sum_{i=1}^{k} \delta_{x_i}$, $\nu = (1/m) \sum_{i=1}^{m} \delta_{y_i} \in \mathcal{P}_0(X)$ and correspondingly let $\mathbf{x} = (x_1, \dots, x_k) \in X^k$ and $\mathbf{y} = (y_1, \dots, y_m) \in X^m$. Then $\mathbf{x}^m, \mathbf{y}^k \in X^{mk}$, $G_k(\mathbf{x}) = G_{km}(\mathbf{x}^m)$ and $G_m(\mathbf{y}) = G_{mk}(\mathbf{y}^k)$ since G is intrinsic, and by Lemma 2.1 $\mathbf{x}^m$ induces μ and $\mathbf{y}^k$ induces ν. Let $n = km$. Then by (2.1), Example 1.2, and the symmetric property

$$d(G_k(\mathbf{x}), G_m(\mathbf{y})) = d(G_n(\mathbf{x}^m), G_n(\mathbf{y}^k)) \leqslant \min_{\sigma \in S^n} \Big(\frac{1}{n} \sum_{j=1}^{n} d^p(x_j, y_{\sigma(j)})\Big)^{1/p} = \mathfrak{w}_p(\mu, \nu).$$

By density of $\mathcal{P}_0(X)$ in $\mathcal{P}^p(X)$ and completeness of (X, d), the contractive map $(1/n) \sum_{j=1}^{n} \delta_{x_j} \mapsto G_n(x_1, \dots, x_n)$ extends uniquely to a map $\beta_G : \mathcal{P}^p(X) \to X$ which satisfies (ii) of Definition 2.6.

Conversely, let $\beta : \mathcal{P}^p(X) \to X$ be a p-contractive barycentric map. Then the mean $G = \{G_n\}$ defined by $G_n : X^n \to X$, $G_n(\mathbf{x}) := \beta\Big((1/n)\sum_{j=1}^{n}\delta_{x_j}\Big)$ is a p-contractive intrinsic mean.

Similarly for $p = \infty$

$$d(G_n(\mathbf{x}), G_n(\mathbf{y})) \leqslant \min_{\sigma\in S^n} \max\{d(x_j, y_{\sigma(j)} : 1 \leqslant j \leqslant n\} = \mathfrak{w}_\infty(\mu,\nu).$$

However, $\mathcal{P}_0(X)$ need not, in general, be dense in $\mathcal{P}^\infty(X)$, but is dense in $\mathcal{P}_{\mathrm{cpt}}(X)$, in which case the previous proof goes through for $p = \infty$. ∎

REMARK 2.8. A metric space is sometimes called *proper* if bounded and closed sets are compact. Such metric spaces are complete and $\mathcal{P}_{\mathrm{cpt}}(X) = \mathcal{P}^\infty(X)$, so in this setting the preceding proposition holds for the case $p = \infty$.

EXAMPLE 2.9. We consider the arithmetic mean $A = \{A_n\}$ on a Banach space E defined by $A_n(x_1,\dots,x_n) = (1/n)\sum_{i=1}^{n} x_i$. It is easily checked to be symmetric and multiplicative, i.e., intrinsic. To see that A is contractive we compute with the help of the triangle inequality for $\mathbf{x} = (x_1,\dots,x_n)$ and $\mathbf{y} = (y_1,\dots,y_n)$ in E^n,

$$d(A_n(\mathbf{x}), A_n(\mathbf{y})) = \Big\|\frac{1}{n}\sum_{i=1}^{n} x_i - \frac{1}{n}\sum_{i=1}^{n} y_i\Big\| \leqslant \frac{1}{n}\sum_{i=1}^{n}\|x_i - y_i\| = \frac{1}{n}\sum_{i=1}^{n} d(x_i, y_i).$$

Hence A extends to a contractive barycentric map $\beta_A : \mathcal{P}^1(E) \to E$.

EXAMPLE 2.10. On a Hadamard space X the Cartan barycenter is a well-known and classical notion of a barycenter [28]. The *Cartan barycentric map* carries $\mu \in \mathcal{P}^1(X)$ to the unique point that minimizes (independently of y) the function

$$z \mapsto \int_X [d^2(z,x) - d^2(y,x)]\,\mathrm{d}\mu(x).$$

It is known that the Cartan barycentric map is contractive for the Wasserstein metric $\mathfrak{w}_1$, a property that has been called the *fundamental contraction property* ([28], Theorem 6.3). By Proposition 2.7 the Cartan barycentric map is $\Phi(\Lambda)$, where Λ is the *least squares mean* given by $\Lambda_n(x_1,\dots,x_n) = x^*$ if and only if x^* is the unique minimizer of the function $x \mapsto \sum_{i=1}^{n} d^2(x, x_i)$. In particular, we conclude that the least squares mean on Hadamard spaces is contractive.

PROPOSITION 2.11. *Let X and Y be complete metric spaces equipped with p-contractive intrinsic means G and H respectively. Let $\Phi(G) = \beta : \mathcal{P}^p(X) \to X$ and $\Phi(H) = \theta : \mathcal{P}^p(Y) \to Y$ be the corresponding contractive barycentric maps. Then $X \times Y$ equipped with the product mean $G \times H = \{G_n \times H_n\}$, which is intrinsic, may be equipped with a complete metric generating the product topology such that $G \times H$ is p-contractive and the projection maps $X \times Y \to X$ and $X \times Y \to Y$ are Lipschitz* *continuous. For any metric on $X \times Y$ satisfying these two conditions, the contractive barycentric map $\Phi(G \times H) : \mathcal{P}^p(X \times Y) \to X \times Y$ is given by $\omega \mapsto (\beta(\omega_X), \theta(\omega_Y))$, where ω_X, respectively ω_Y is the X-marginal, respectively Y-marginal of ω.*

Proof. The fact that $G \times H$ operates coordinatewise and that G and H are intrinsic directly implies that $G \times H$ is intrinsic. Define a metric ρ on $X \times Y$ by

$$\rho((x_1, y_1), (x_2, y_2)) = (d^p(x_1, x_2) + d^p(y_1, y_2)\ )^{1/p}.$$

For $(x_1, \ldots, x_n), (u_1, \ldots, u_n) \in X^n$ and $(y_1, \ldots, y_n), (v_1, \ldots, v_n) \in Y^n$

$$\begin{aligned}
&\rho^p(G_n \times H_n((x_1, y_1), \ldots, (x_n, y_n)), G_n \times H_n((u_1, v_1), \ldots, (u_n, v_n))) \\
&\quad = \rho^p((G_n(x_1, \ldots, x_n), H_n(y_1, \ldots, y_n)), (G_n(u_1, \ldots, u_n), H_n(v_1, \ldots, v_n))) \\
&\quad = d^p(G_n(x_1, \ldots, x_n), G_n(u_1, \ldots, u_n)) + d^p(H_n(y_1, \ldots, y_n), H_n(v_1, \ldots, v_n)) \\
&\quad \leqslant \sum_{i=1}^n d^p(x_i, u_i) + \sum_{i=1}^n d^p(y_i, v_i) = \sum_{i=1}^n \rho((x_i, y_i), (u_i, v_i))^p,
\end{aligned}$$

which establishes that $G \times H$ is p-contractive. The projection maps $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ are clearly Lipschitz with Lipschitz constant 1.

For the last assertion, the fact that $G \times H$ is intrinsic and p-contractive implies that a corresponding contractive barycentric map $\Phi(G \times H)$ exists by Proposition 2.7. Since

$$\begin{aligned}
\Phi(G \times H)\Big(\frac{1}{n}\sum_{i=1}^n \delta_{(x_i, y_i)}\Big) &= (G_n \times H_n)((x_1, y_1), \ldots, (x_n, y_n)) \\
&= (G_n(x_1, \ldots, x_n), H_n(y_1, \ldots, y_n)) \\
&= \Big(\beta\Big(\frac{1}{n}\sum_{i=1}^n \delta_{x_i}\Big), \theta\Big(\frac{1}{n}\sum_{i=1}^n \delta_{y_i}\Big)\Big),
\end{aligned}$$

$\Phi(G \times H)$ agrees with $\omega \mapsto (\beta(\omega_X), \theta(\omega_Y))$ on $\mathcal{P}_0(X \times Y)$. Since $\pi_X : X \times Y \to X$ is Lipschitz, by Lemma 1.3 the function $(\pi_X)_* : \mathcal{P}^p(X \times Y) \to \mathcal{P}^p(X)$ is continuous. Since the first coordinate of $\omega \mapsto (\beta(\omega_X), \theta(\omega_Y))$ is equal to the composition $\beta(\pi_X)_*$, it is continuous and similarly the second coordinate is continuous. Since $\mathcal{P}_0(X \times Y)$ is dense in $\mathcal{P}^p(X \times Y)$, there is at most one continuous extension, so the two maps agree on all of $\mathcal{P}^p(X \times Y)$. ∎

3. A CONSTRUCTION

In this section we consider a construction that under appropriate hypotheses can convert a mean G that is not intrinsic into a mean G^* that is intrinsic while preserving the p-contractive property.

LEMMA 3.1. *Let $G = \{G_n\}$ be a mean on a metric space X such that for each $\mathbf{x} \in X^n$, $n \geqslant 2$, the limit $G_{nk}(\mathbf{x}^k)$ as $k \to \infty$ exists and is denoted by $G_n^*(\mathbf{x})$. Then $G^* = \{G_n^*\}$ is multiplicative, respectively symmetric, respectively p-contractive whenever G is.*

Proof. The multiplicative property of G^* is immediate from

$$G^*_{nk}(\mathbf{x}^k) = \lim_{l\to\infty} G_{nkl}(\mathbf{x}^{kl}) = G^*_n(\mathbf{x}).$$

If G is symmetric, then $G^*_n(\mathbf{x}_\sigma) = \lim_{k\to\infty} G_{nk}((\mathbf{x}_\sigma)^k) = \lim_{k\to\infty} G_{nk}(\mathbf{x}^k) = G^*_n(\mathbf{x})$ for all permutations σ. From

$$\begin{aligned} d(G^*_n(\mathbf{x}), G^*_n(\mathbf{y})) &= \lim_{k\to\infty} d(G_{nk}(\mathbf{x}^k), G_{nk}(\mathbf{y}^k)) \leqslant \lim_{k\to\infty} \Big(\frac{1}{nk}\sum_{i=1}^{k}\sum_{j=1}^{n} d^p(x_j, y_j)\Big)^{1/p} \\ &= \Big(\frac{1}{n}\sum_{j=1}^{n} d^p(x_j, y_j)\Big)^{1/p}, \end{aligned}$$

each G^*_n is p-contractive if G is. ∎

There are two different motivations that might lead one to employ the preceding construction. The first is that one wants to use some useful approximation to study some specific barycentric map. The prime example has been the approximation of the Cartan barycentric map of Example 2.10 with what Sturm [28] calls the inductive mean. We recall that in a metric space (X, d) equipped with distinguished geodesics, the *inductive mean* S is defined inductively by $S_1(a_1) = a_1$ and $S_k(a_1, \ldots, a_k) = S_{k-1}(x_1, \ldots, x_{k-1})\#_{1/k} x_k$. In [28] Sturm showed that in a Hadamard space the Cartan barycentric map and related least squares mean could be approximated almost everywhere (in an appropriate sense) as a limit of the inductive mean. Lim and Palfia [25] showed that for the case of n points the least squares mean could be approximated deterministically by the inductive mean. We recall a slightly recast and simplified version of the main result of Lim and Palfia.

THEOREM 3.2. *Let $\mathbf{x} = (x_1, \ldots, x_n) \in X^n$, where X is a Hadamard space. Then $\lim_{k\to\infty} S_{kn}(\mathbf{x}^k) = \Lambda_n(\mathbf{x})$, where S is the inductive mean and Λ the least squares mean. This convergence takes place uniformly over all Hadamard spaces in the sense that for each k*

$$d(\Lambda(\mathbf{x}), S_{kn}(\mathbf{x})) \leqslant \frac{4}{k}\Delta^2(x_1, \ldots, x_n), \tag{3.1}$$

where $\Delta(x_1, \ldots, x_n)$ is the diameter of the set $\{x_1, \ldots, x_n\}$, the maximum distance between any pair of the points in the set.

The theorem shows that $\Lambda = S^*$ and thus, in particular, that S^* exists. What is a bit surprising about this example is that a non-symmetric mean S gives rise to a symmetric $S^* = \Lambda$.

REMARK 3.3. It is interesting to note that one sees easily (by induction) that the inductive mean on a Hadamard space is contractive, hence Λ is contractive as well, and thus by Proposition 2.7 we conclude that the Cartan barycentric map is contractive also. This approach to these results gives an alternative to the probabilistic one of Sturm.

In extending various theories such as ergodic theory to the setting of metric spaces there is sometimes a need to have available a contractive barycentric map that is invariant under isometries. This need can provide a second motivation for employing the machinery of Lemma 3.1. We briefly mention an important example without going into the details.

The construction scheme in Lemma 3.1 for a specific contractive mean on metric spaces of nonpositive curvature in the sense of Busemann (a weaker notion than that of a Hadamard space) has been used by Es-Sahib and Heinich in [8]. Its convergence in the sense of Lemma 3.1 has been established more generally with a more elementary proof by A. Navas [26]. The resulting limiting intrinsic mean plays a key role in establishing a general version of the Birkhoff ergodic theorem on Busemann NPC spaces [26]. The starting contractive mean of Es-Sahib and Heinich is constructed inductively by an approach that has been called the symmetrization procedure. Surprisingly, this symmetrization procedure arose independently in work by Ando, Li, and Mathias [1], who used it to give a solution of the long standing open problem of extending the matrix geometric mean to from two to n variables. It has become one of the standard procedures for providing contractive symmetric means. Quite general conditions for the existence of the mean in metric spaces were given in [20].

4. THE KARCHER MEAN AND ITS BARYCENTRIC MAP

For a Hilbert space H let $\mathcal{B}(H)$ be the Banach space of bounded linear operators on H equipped with the operator norm, $\mathcal{S}(H)$ the closed subspace of bounded self-adjoint linear operators, and let $\mathbb{P} = \mathbb{P}(H) \subseteq \mathcal{S}(H)$ be the open convex cone of positive definite operators. The Banach–Lie group $\mathrm{GL}(H)$ of bounded invertible linear operators (with operation composition) acts on $\mathbb{P}$ via congruence transformations: $\Gamma_C(X) = CXC^*$. For $X, Y \in \mathcal{S}(H)$, we write $X \leqslant Y$ if $Y - X$ is positive semidefinite, and $X < Y$ if $Y - X$ is positive definite. Note that $X \leqslant Y$ if and only if $\langle x, Xx\rangle \leqslant \langle x, Yx\rangle$ for all $x \in H$.

For $A, B \in \mathbb{P}$ and $t \in \mathbb{R}$, the t-weighted geometric mean of A and B is defined by

$$A\#_t B = A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2}. \tag{4.1}$$

Some basic properties of the t-weighted mean are:

(i) (Loewner–Heinz inequality) $A\#_t B \leqslant C\#_t D$ for $A \leqslant C, B \leqslant D$ and $t \in [0,1]$;
(ii) $M(A\#_t B)M^* = (MAM^*)\#_t(MBM^*)$ for $M \in \mathrm{GL}(H)$;
(iii) $A\#_t B \leqslant (1-t)A + tB$ for $t \in [0,1]$.

For $t = 1/2$, $A\#_{1/2}B = B\#_{1/2}A$ is called simply the geometric mean of A and B and denoted $A\#B$.

The Thompson metric on $\mathbb{P}$ is defined by $d(A,B) = \|\log(A^{-1/2}BA^{-1/2})\|$, where $\|X\|$ denotes the operator norm of X. It is known that d is a complete metric on $\mathbb{P}$, that the metric topology agrees with the relative topology induced by the operator norm, and that

$$d(A,B) = \max\{\log M(B/A), \log M(A/B)\},$$

where $M(B/A) = \inf\{\alpha > 0 : B \leqslant \alpha A\}$; see [7], [27], [29]. Furthermore, $A\#B$ is a midpoint of A and B in the Thompson metric and $t \mapsto A\#_t B$, $0 \leqslant t \leqslant 1$, is a metric geodesic from A to B.

We note that the Thompson metric (in the second form) exists on all normal cones of real Banach spaces. For instance,

$$d((s_1,\ldots,s_n),(t_1,\ldots,t_n)) = \max_{1\leqslant i\leqslant n} \left|\log \frac{s_i}{t_i}\right| \tag{4.2}$$

on $\mathbb{R}_+^n$, where $\mathbb{R}_+ = (0,\infty)$.

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

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

It has been shown in [23] that this equation does indeed have a unique solution in $\mathbb{P}$ and that the resulting mean Λ_n for $n \geqslant 2$ has the following properties:

(i) Λ_n is symmetric and idempotent;

(ii) (Monotonicity) If $B_i \leqslant A_i$ for all $1 \leqslant i \leqslant n$, then $\Lambda_n(B_1,\ldots,B_n) \leqslant \Lambda_n(A_1,\ldots,A_n)$;

(iii) (Contractivity) $d(\Lambda_n(A_1,\ldots,A_n),\Lambda_n(B_1,\ldots,B_n)) \leqslant (1/n)\sum_{i=1}^{n} d(A_i,B_i)$ for the Thompson metric d.

We note also that the Karcher mean Λ is intrinsic since the left hand side of the Karcher equation for $(A_1,\ldots,A_n)^k$ is just k times that for $(A_1,\ldots,A_n)$, and hence still equal to 0 for the same X. We thus have from Proposition 2.7 the next proposition.

PROPOSITION 4.1. *The correspondence of Proposition* 2.7 *yields a uniquely determined contractive barycentric map* $\beta_\Lambda : \mathcal{P}^1(\mathbb{P}) \to \mathbb{P}$ *satisfying* $\beta\Big((1/n)\sum_{i=1}^{n}\delta_{A_i}\Big) = \Lambda_n(A_1,\ldots,A_n)$.

5. APPROXIMATION OF THE KARCHER MEAN

We maintain the setting in which $\mathcal{B}(H)$ denotes the C^*-algebra of bounded linear operators on a Hilbert space H, $\mathcal{S}(H)$ is the closed subspace of bounded hermitian operators, and $\mathbb{P}$ is the open cone of positive operators. The next lemma asserts a variant of normality for the closed cone $\mathbb{P}_0 = \{A \in \mathcal{S}(H) : 0 \leqslant A\}$ in $\mathcal{S}(H)$.

LEMMA 5.1. *Assume that $0 \leqslant B_\alpha \leqslant A_\alpha$ for nets $\{B_\alpha\}$ and $\{A_\alpha\}$, that $\{A_\alpha\}$ converges strongly to 0, and that the set $\{B_\alpha\}$ is bounded above. Then the net $\{B_\alpha\}$ converges strongly to 0.*

Proof. If $0 \leqslant B_\alpha \leqslant C$ for each α, then $0 \leqslant B_\alpha^{1/2} \leqslant C^{1/2}$ by the Loewner–Heinz theorem, and thus $\|B_\alpha^{1/2}\| \leqslant \|C^{1/2}\|$ (see, e.g., Lemma 2.3(ii) of [23]). For $x \in H$,

$$0 \leqslant \langle B_\alpha^{1/2}x, B_\alpha^{1/2}x\rangle = \langle x, B_\alpha x\rangle \leqslant \langle x, A_\alpha x\rangle.$$

By hypothesis the right hand side converges to 0, so $\|B_\alpha^{1/2}(x)\|^2 \to 0$. Then

$$\|B_\alpha(x)\| = \|B_\alpha^{1/2}B_\alpha^{1/2}(x)\| \leqslant \|C^{1/2}\|\,\|B_\alpha^{1/2}(x)\| \to 0$$

and thus $B_\alpha \to 0$ strongly. ∎

LEMMA 5.2. *Suppose that $\{A_\alpha\}$, $\{B_\alpha\}$, $\{C_\alpha\}$ are nets in $\mathcal{S}(H)$ satisfying:*
(i) *$C_\alpha \leqslant B_\alpha \leqslant A_\alpha$ for each α;*
(ii) *the set $\{B_\alpha\}$ is bounded above and the net C_α is bounded below in $\mathcal{S}(H)$;*
(iii) *the nets $\{A_\alpha\}$ and $\{C_\alpha\}$ both converge strongly to some $B \in \mathcal{S}(H)$.*
Then the net $\{B_\alpha\}$ converges strongly to B.

Proof. Suppose that $B_\alpha \leqslant D$ and $E \leqslant C_\alpha$ for each α. Set $F_\alpha = A_\alpha - C_\alpha$ and $G_\alpha = B_\alpha - C_\alpha$. Then for each α, $0 \leqslant G_\alpha \leqslant F_\alpha$, $G_\alpha \leqslant D - E$. Furthermore, the net $\{F_\alpha\}$ converges strongly to $B - B = 0$. It then follows from Lemma 5.1 that $G_\alpha \to 0$ strongly. Therefore the net $B_\alpha = (B_\alpha - C_\alpha) + C_\alpha$ converges strongly to $0 + B = B$. ∎

A subset K of $\mathcal{S}(H)$ is said to be *order-convex* if $A, C \in K$ and $C \leqslant B \leqslant A$ implies that $B \in K$. If for an arbitrary set K, one takes $\langle K\rangle_{\mathrm{o}}$ to be the set of all $B \in \mathcal{S}(H)$ such that $C \leqslant B \leqslant A$ for some $A, C \in K$, then $\langle K\rangle_{\mathrm{o}}$ is the smallest order convex set containing K.

LEMMA 5.3. *Let $D \leqslant E$ in $\mathcal{S}(H)$, set $[D, E] = \{A \in \mathcal{S}(H) : D \leqslant A \leqslant E\}$, let $X \in [D, E]$, and let V be a strongly open set containing X. Then there exists a strongly open set U such that*

$$X \in U \cap [D, E] \subseteq \langle U \cap [D, E]\rangle_{\mathrm{o}} \subseteq V.$$

Hence the order interval $[D, E]$ equipped with the relative strong topology has a basis of order-convex neighborhoods at each of its points.

Proof. Let $D \leqslant X \leqslant E$ and let V be a strongly open set containing X. Suppose that $\langle U \cap [D,E]\rangle_{\mathrm{o}}$ is not contained in V for every strongly open set U containing X. Then there exist nets $D \leqslant C_\alpha \leqslant B_\alpha \leqslant A_\alpha \leqslant E$, where the indices α run over the strongly open neighborhoods U of X, such that $C_\alpha, A_\alpha \to X$ in the strong topology, but $B_\alpha \notin V$. But this contradicts Lemma 5.2. ∎

Let $\mathrm{HS}(H)$ denote the bilateral ideal of Hilbert–Schmidt operators of $\mathcal{B}(H)$. Then with respect to the norm $\|A\|_2 = \mathrm{tr}(AA^*)^{1/2}$, $\mathrm{HS}(H)$ is a Banach algebra (without unit). In $\mathcal{B}(H)$ we define

$$\mathcal{H}_{\mathbb{C}} = \{A + \lambda I : A \in \mathrm{HS}(H), \lambda \in \mathbb{C}\},$$

a complex linear subalgebra that we call the *extended Hilbert–Schmidt algebra*. There is a natural Hilbert space structure for this subspace (where scalar operators are orthogonal to Hilbert–Schmidt operators) given by the inner product

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

Our focus is on the symmetric or real part of $\mathcal{H}_{\mathbb{C}}$,

$$\mathcal{H}_{\mathbb{R}} = \{A + \lambda I : A^* = A,\ A \in \mathrm{HS}(H),\ \lambda \in \mathbb{R}\},$$

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

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

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

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

The structure of the Riemannian manifold Σ closely parallels that of the finite dimensional Riemannian manifolds of positive definite matrices equipped with the Riemannian trace metric, as has been worked out by Larotonda [16]; see also the last part of [25]. In particular Σ is a Riemannian manifold of nonpositive curvature, and hence equipped with its distance metric δ is a Hadamard space. Hence the least squares minimizer

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

uniquely exists. Furthermore, the topology induced by the metric δ on Σ agrees with the relative topology of Hilbert space topology of $\mathcal{H}_{\mathbb{R}}$, and is finer than the relative operator norm topology.

One can identify the tangent space $T_I(\Sigma)$ at I with $\mathcal{H}_{\mathbb{R}}$ and the exponential mapping on the tangent space with the usual exponential mapping of operators. The mapping $\exp : \mathcal{H}_{\mathbb{R}} \to \Sigma$ is a diffeormorphism with inverse denoted log. Using methods of Riemannian geometry and Karcher's result ([11], Theorem 1.2) or a more operator-theoretic approach [23], one can show the least squares mean is the unique point where the gradient of the least squares objective function $f(X) = \sum_{i=1}^{n} \delta^2(X, A_i)$ vanishes, which leads to the alternative characterization of the least squares mean as the unique solution to the Karcher equation, which (up to a scalar multiple) arises from setting the gradient equal to 0. We summarize:

THEOREM 5.4. *The least squares mean of $(A_1, \dots, A_n) \in \Sigma^n$ in the Riemannian manifold Σ is the Karcher mean $\Lambda(A_1, \dots, A_n)$, the solution of the Karcher equation*

$$\sum_{i=1}^{n} \log(X^{1/2} A_i^{-1} X^{1/2}) = 0. \tag{5.1}$$

It follows from Theorem 5.4 that the least squares mean in Σ is the restriction of the Karcher mean on $\mathbb{P}$ to Σ. We now present a reverse construction: extending the least squares mean, equivalently the restricted Karcher mean, on Σ to $\mathbb{P}$. Let $\{\alpha\}_{\alpha \in \Delta}$ denote the collection of non-zero finite-dimensional subspaces of H ordered by inclusion, a directed family. Let $P_\alpha : H \to H$ denote the orthogonal projection onto the subspace α. We note that each P_α is hermitian, positive semidefinite, idempotent, and has finite rank, hence is Hilbert–Schmidt. We view $\{P_\alpha : \alpha \in \Delta\}$ as a monotonically increasing net indexed by Δ that strongly converges to its supremum, the identity I, since for any $x \in H$, $P_\alpha(x) = x$ for all large enough α.

Since $\{P_\alpha : \alpha \in \Delta\}$ is bounded, the net $\{P_\alpha A P_\alpha\}$ strongly converges to A for any $A \in \mathcal{B}(H)$. For A hermitian, it is a monotonically increasing net with supremum A. (One can show that A is the supremum directly or use the standard fact that any monotonically increasing net of symmetric operators that is bounded above strongly converges to its supremum.)

PROPOSITION 5.5. *For $A_1, \dots, A_n \in \mathbb{P}$ choose m large enough such that $\mathrm{e}^{-m} I < A_i < \mathrm{e}^{m} I$ for $1 \leqslant i \leqslant n$. Then*

$$X_\alpha = \Lambda(\mathrm{e}^m I - P_\alpha(\mathrm{e}^m I - A_1) P_\alpha, \dots, \mathrm{e}^m I - P_\alpha(\mathrm{e}^m I - A_n) P_\alpha)$$

is a monotonically decreasing net in Σ bounded below by $\mathrm{e}^{-m} I$ that strongly converges to its infimum, which is equal to the Karcher mean $\Lambda(A_1, \dots, A_n)$. Similarly

$$Z_\alpha = \Lambda(\mathrm{e}^{-m} I + P_\alpha(A_1 - \mathrm{e}^{-m} I) P_\alpha, \dots, \mathrm{e}^{-m} I + P_\alpha(A_n - \mathrm{e}^{-m} I) P_\alpha)$$

is a monotonically increasing net in Σ bounded above by $\mathrm{e}^m I$ that strongly converges to its supremum, again the Karcher mean $\Lambda(A_1, \dots, A_n)$.

Proof. Since the net $\{P_\alpha(\mathrm{e}^m I - A_i) P_\alpha\}_\alpha$ is a monotonically increasing net strongly converging to its supremum $\mathrm{e}^m I - A_i$, the net $\{\mathrm{e}^m I - P_\alpha(\mathrm{e}^m I - A_i) P_\alpha\}_\alpha$ is a decreasing net strongly converging to its infimum $\mathrm{e}^m I - (\mathrm{e}^m I - A_i) = A_i \geqslant \mathrm{e}^{-m} I$. By the idempotency and monotonicity ([23], Theorem 6.8) of the Karcher mean, we have that X_α is a decreasing net bounded below by $\mathrm{e}^{-m} I$, and hence strongly converges to its infimum, call it Y.

By Theorem 5.4 each X_α satisfies the Karcher equation

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

where $B_{i,\alpha} = e^m I - P_\alpha(e^m I - A_i)P_\alpha$. Since (by the previous paragraph) $B_{i,\alpha}$ converges strongly to A_i and since the function $f(X) = \sum_{i=1}^{n} \log(X^{1/2} A_i^{-1} X^{1/2})$ is strongly continuous on the bounded order interval $[e^{-m}I, e^m I]$ by Lemma 5.4 of [23], we conclude that

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

Hence Y is equal to the Karcher mean $\Lambda(\omega; A_1, \ldots, A_n)$. The similar assertions for the net Z_α follow in an analogous manner. ∎

From the proof of Proposition 5.5 one extracts the following special case of strong continuity of Λ.

COROLLARY 5.6. *Let $\mathbb{A}_\alpha = (A_{1,\alpha}, \ldots, A_{n,\alpha})$ be a decreasing respectively increasing net in $\mathbb{P}^n$ that strongly converges to its infimum respectively supremum $\mathbb{A} = (A_1, \ldots, A_n) \in \mathbb{P}^n$. Then $\Lambda(\mathbb{A}_\alpha)$ is a decreasing respectively increasing net strongly converging to $\Lambda(\mathbb{A})$.*

We come now to our generalization of the Lim–Palfia theorem, Theorem 3.2.

THEOREM 5.7. *For $\mathbb{A} = (A_1, \ldots, A_n) \in \mathbb{P}^n$, we have*

$$\Lambda(A_1, \ldots, A_n) = \lim_{k\to\infty} S_{kn}(\mathbb{A}^k),$$

where the limit is taken in the strong topology.

Proof. Let V be a strongly open set containing $\widehat{A} = \Lambda(A_1, \ldots, A_n)$. Pick

$$e^{-m}I \leqslant Z_\alpha \leqslant \widehat{A} \leqslant X_\alpha \leqslant e^m I$$

as in Proposition 5.5. By Lemma 5.3 there exists a strongly open set U containing $\widehat{A}$ such that $\langle U \cap [e^{-2m}I, e^{2m}I]\rangle_o \subseteq V$. The strong convergence $Z_\alpha, X_\alpha \to \widehat{A}$ implies there exists β such that $Z_\alpha, X_\alpha \in U \cap [e^{-m}I, e^m I]$ for $\alpha \geqslant \beta$.

It is a standard result that the Hilbert–Schmidt norm $\|\cdot\|_2$ bounds the operator norm $\|\cdot\|$ on $HS(H)$. Thus for $A + \lambda I \in \Sigma$

$$\|A + \lambda I\|_{\mathcal{H}_\mathbb{R}} = \|A\|_2 + \lambda \geqslant \|A\| + \lambda \geqslant \|A + \lambda I\|.$$

Since the set $Q = \{Y \in \mathcal{P} : e^{-2m}I < Y < e^{2m}\}$ is open in the Thompson metric, hence norm topology, and contains $[e^{-m}I, e^m I]$, there exists $\varepsilon > 0$, such that the norm open balls of radius ε around Z_β and X_β respectively are contained in $Q \cap U$. Then the smaller ε-balls $\mathcal{N}_\varepsilon(Z_\beta)$ and $\mathcal{N}_\varepsilon(X_\beta)$ in $\mathcal{H}_\mathbb{R}$ will also lie in $Q \cap U$.

Let $X_{j,\beta} = \mathrm{e}^m I - P_\beta(\mathrm{e}^m I - A_j)P_\beta$ for $1 \leqslant j \leqslant n$. By definition $X_\beta = \Lambda(X_{1,\beta}, \ldots, X_{n,\beta})$. Similarly $Z_\beta = \Lambda(Z_{1,\beta}, \ldots, Z_{n,\beta})$ for $Z_{j,\beta}$ defined in an analogous manner. Then by Theorem 3.2 for the inductive mean sequences

$$\lim_{k\to\infty} S_{kn}((X_{1,\beta}, \ldots, X_{n,\beta})^k) = \Lambda(X_{1,\beta}, \ldots, X_{n,\beta}) = X_\beta,$$

where the limit is taken in the manifold topology of Σ, which is the relative Hilbert space topology of $\mathcal{H}_{\mathbb{R}}$. Similarly $\lim_{k\to\infty} S_{kn}((Z_{1,\beta}, \ldots, Z_{n,\beta})^k) = Z_\beta$. Thus for large enough N,

$$\begin{aligned} S_{Nn}((X_{1,\beta}, \ldots, X_{n,\beta})^N), S_{Nn}((Z_{1,\beta}, \ldots, Z_{n,\beta})^N) &\in \mathcal{N}_\varepsilon(X_\beta) \cup \mathcal{N}_\varepsilon(Z_\beta) \\ &\subseteq [\mathrm{e}^{-2m} I, \mathrm{e}^{2m} I] \cap U. \end{aligned}$$

Since the two-variable weighted geometric means $X\#_t Y$ are monotonic for $0 < t < 1$, it follows easily by induction that the inductive means are monotonic. Hence

$$S_{Nn}((Z_{1,\beta}, \ldots, Z_{n,\beta})^N) \leqslant S_{Nn}((A_1, \ldots, A_n)^N) \leqslant S_{Nn}((X_{1,\beta}, \ldots, X_{n,\beta})^N),$$

and thus by the order convexity of $\langle [\mathrm{e}^{-2m} I, \mathrm{e}^{2m} I] \cap U \rangle_{\mathrm{o}}$, $S_{Nn}((A_1, \ldots, A_n)^N) \in V$ for large enough N. ∎

6. CONVEXITY AND JENSEN'S INEQUALITY

A subset K of $\mathbb{P}$ is said to be *convex* if $A\#_t B \in K$, $0 < t < 1$, whenever $A, B \in K$. Since any $A\#_t B$ can be approximated arbitrarily close in the norm topology (which is also the metric topology for the Thompson metric) by iterated midpoints, a closed subset is convex if and only if it is closed under taking midpoints $A\#B$. Since the intersection of (closed) convex sets is again (closed) convex, every subset has a (closed) convex hull, the smallest such containing the subset.

LEMMA 6.1. *Suppose that a convex set $K \subseteq \mathbb{P}$ is bounded in the Thompson metric, i.e., contained in the order interval $[\mathrm{e}^{-m} I, \mathrm{e}^m I]$ for some $m > 0$. Then the strong closure of K is convex.*

Proof. Since the cone $\mathbb{P}_0$ of positive semidefinite operators is closed in the strong topology, it follows that $[\mathrm{e}^{-m} I, \mathrm{e}^m I] = (\mathrm{e}^{-m} I + \mathbb{P}_0) \cap (\mathrm{e}^m I - \mathbb{P}_0)$ is strongly closed. Let $A_\alpha \to A$, $B_\alpha \to B$ in the strong topology, where $\{A_\alpha\}, \{B_\alpha\} \subseteq K$. Then $A, B \in [\mathrm{e}^{-m} I, \mathrm{e}^m I] \subseteq \mathbb{P}$. By Lemma 5.4(iii) of [23] the map sending X, Y to $X\#_t Y$ for $0 \leqslant t \leqslant 1$ is jointly continuous with respect to the strong topology when restricted to $[\mathrm{e}^{-m} I, \mathrm{e}^m I]$. From this it follows that $A_\alpha\#_t B_\alpha \to A\#_t B$ strongly, and thus the closure is convex. ∎

PROPOSITION 6.2. *For $A_1, \ldots, A_n \in \mathbb{P}$, $\Lambda(A_1, \ldots, A_n)$ is in the strong closure of the convex hull of $\{A_1, \ldots, A_n\}$, in particular in the closed convex hull.*

Proof. This result is a corollary of Theorem 5.7, since each member in the sequence of inductive means belongs to the convex hull of $\{A_1, \dots, A_n\}$. ∎

PROPOSITION 6.3. *Let $\beta_\Lambda : \mathcal{P}^1(\mathbb{P}) \to \mathbb{P}$ be the contractive barycentric map of Proposition* 4.1. *Then for $\mu \in \mathcal{P}^1(\mathbb{P})$, $\beta_\Lambda(\mu)$ belongs to the strong closure of the convex hull of the support of μ.*

Proof. Let Σ be the support of μ. Then $\beta_\Lambda\Big((1/n)\sum_{i=1}^{n}\delta_{A_i}\Big) = \Lambda(A_1, \dots, A_n)$ is in the strong closure of the convex hull of Σ for each $A_1, \dots, A_n \in \Sigma$ by Proposition 6.2. Since $\mathcal{P}_0(\Sigma)$ is dense in $\mathcal{P}^1(\Sigma)$, its image under β_Λ is dense, hence strongly dense, in the image of $\mathcal{P}^\infty(\mathbb{P})$, and hence $\beta_\Lambda(\mu)$ is in the strong closure of the convex hull of Σ. ∎

REMARK 6.4. (i) For $\mathbb{P} \times \mathbb{R}$ equipped with $d((A,a),(B,b)) = d(A,B) + d(a,b)$, the sum of the Thompson metric and the usual metric on $\mathbb{R}$ and hence a complete metric, we define the product mean $G = \{G_n\}$ on $\mathbb{P} \times \mathbb{R}$ by

$$G_n((A_1,a_1), \dots, (A_n,a_n)\,) = \Big(\Lambda_n(A_1, \dots, A_n), \frac{1}{n}\sum_{i=1}^{n} a_i\Big).$$

For G_2 the mean yields the standard midpoint pair $(A_1\#A_2, a_1/2 + a_2/2)$ with corresponding geodesic paths $t \mapsto (A_1\#_t A_2, (1-t)a_1 + ta_2)$ for $0 \leqslant t \leqslant 1$.

(ii) That G is intrinsic and contractive follows from Proposition 2.11, from which we also conclude that there is an induced contractive barycentric map β_G that is defined by taking the barycenter of each of the marginals of any given member of $\mathcal{P}^1(X \times \mathbb{R})$.

(iii) For $\mathbf{A} = (A_1, \dots, A_n) \in \mathbb{P}^n$ and $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$, we observe for the inductive mean $S = \{S_n\}$ that

$$\lim_{k\to\infty} S_{kn}(\mathbf{A},\mathbf{a})^k = \Big(\Lambda_n(\mathbf{A}), \frac{1}{n}\sum_{i=1}^{n} a_i\Big) = G_n(\mathbf{A},\mathbf{a}) = \beta_G\Big(\frac{1}{n}\sum_{i=1}^{n}\delta_{(A_i,a_i)}\Big), \tag{6.1}$$

where in the first coordinate the limit is taken in the strong topology. The first equality follows in the first coordinate from Theorem 5.7 and in the second from the fact, easily established by induction, that $S_m(a_1, \dots, a_m) = (1/m)\sum_{i=1}^{m} a_i$.

PROPOSITION 6.5 (Jensen's inequality). *Let $F : \mathbb{P} \to \mathbb{R}$ be lower semicontinuous with respect to the strong topology and convex, i.e., $F(A\#_t B) \leqslant (1-t)F(A) + tF(B)$ for all $A, B \in \mathbb{P}$. Then for any measure $\mu \in \mathcal{P}^1(\mathbb{P})$, $F(\beta_\Lambda(\mu)) \leqslant \int_{\mathbb{P}} F(X)\,\mathrm{d}\mu(X)$.*

Proof. We equip $\mathbb{P} \times \mathbb{R}$ with the product structure of the preceding remark. We observe that the set $K = \{(A,a) : F(A) \leqslant a\}$ is closed in $\mathbb{P} \times \mathbb{R}$ in the product of the strong and standard topologies by the lower semicontinuity assumption and convex by the convexity of F. For $\mathbf{A} = (A_1, \dots, A_n) \in \mathbb{P}^n$ and $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$ such that $(\mathbf{A},\mathbf{a}) \in K$, we have from the inductive definition of S that $S_{kn}(\mathbf{A}, \mathbf{a})^k \in K$ for each k, since K is convex. It follows from equation (6.1) that $\beta_G\Big((1/n) \sum_{i=1}^{n} \delta_{(A_i,a_i)}\Big) \in K$ from the fact that K is appropriately closed.

For $\mu \in \mathcal{P}^1(\mathbb{P})$ consider the map $j : \mathbb{P} \to \mathbb{P} \times \mathbb{R}$ defined by $j(A) = (A, F(A))$. The image of $\mathbb{P}$ under j is contained in K, so the push-forward measure $j^*(\mu)$ has support contained in K. Since $\mathcal{P}_0(K)$ is dense in $\mathcal{P}^1(K)$, β_G is continuous, and K is closed, we conclude from the previous paragraph that $\beta_G(j^*(\mu)) \subseteq K$.

From part (ii) and equation (6.1) of the preceding remark, it follows that β_G acts on $\mathcal{P}^1(\mathbb{P} \times \mathbb{R})$ by applying $\beta_\Lambda \times \beta_\mathbb{R}$ to the marginals of any μ, where $\beta_\mathbb{R}$ is the contractive barycentric map corresponding to the arithmetic mean on $\mathbb{R}$. Since j followed by the projection into $\mathbb{P}$ is the identity, we conclude that $\beta_G(j^*(\mu)) = (\beta_\Lambda(\mu), \beta_\mathbb{R}(F^*(\mu)))$. Since this ordered pair is in K by the preceding paragraph, we conclude that

$$\beta_\Lambda(\mu) \leqslant \beta_\mathbb{R}(F^*(\mu)) = \int_{\mathbb{P}} F(X)\,\mathrm{d}\mu(X).$$

This concludes the proof. ∎

7. SUBALGEBRAS

For convenience and ease of presentation we have limited our considerations in Sections 5, 6, and 7 to the full algebra $\mathcal{B}(H)$ of bounded linear operators. However, we observe that the constructions can be carried out in large classes of subalgebras (which we assume always to contain the identity I), and hence by the Gelfand–Naimark theorem hold for large classes of C^*-algebras. For any norm-closed C^*-subalgebra $\mathcal{A}$ of $\mathcal{B}(H)$, $\mathbb{P}_\mathcal{A} = \mathcal{A} \cap \mathbb{P}$ will be its open cone of positive operators, and will be closed under the operation of taking weighted geometric means $A\#_t B$. However to obtain the existence of the Karcher mean, one needs to assume further that the subalgebra $\mathcal{A}$ is monotone complete (actually, monotone σ-complete will suffice). Since the von Neumann subalgebras are strongly closed, the previous results hold in that setting.

8. AN OPEN PROBLEM

In [28] Sturm has derived a version of the strong law of large numbers (SLLN) for random variables into a Hadamard space. An interesting question is whether a similar SLLN holds for the space $\mathbb{P}(H)$ of positive operators on a Hilbert space. Theorem 6.7 may be viewed as a deterministic version of this result for measures with finite support.

Acknowledgements. The work of Y. Lim was supported by the National Research Foundation of Korea (NRF) grants Nos. 2015R1A3A2031159 and 2016R1A5A1008055 funded by the Korea government (MEST).

REFERENCES

[1] T. ANDO, C.K. LI, R. MATHIAS, Geometric means, *Linear Algebra Appl.* **385**(2004), 305–334.

[2] T. AUSTIN, A CAT(0)-valued pointwise ergodic theorem, *J. Topol. Anal.* **3**(2011), 145–152.

[3] D BINI, B. MEINI, F. POLONI, An effective matrix geometric mean satisfying the Ando–Li–Mathias properties, *Math. Comp.* **79**(2010), 437–452.

[4] J. BOCHI, A. NAVAS, A geometric path from zero Lyapunov exponents to rotation cocycles, *Ergodic Theory Dynam. Systems* **35**(2013), 374–402.

[5] F. BOLLEY, Separability and completeness for the Wasserstein distance, in *Séminaire de Probabilités* XLI, Lecture Notes in Math., vol. 1934, Springer, Berlin 2008, pp. 371–377.

[6] T. CHAMPION, L. DE PASCALE, R. JUUTINEN, The ∞-Wasserstein distance: local solutions and existence of optimal transport maps, *SIAM J. Math. Anal.* **40**(2008), 1–20.

[7] G. CORACH, H. PORTA, L. RECHT, Convexity of the geodesic distance on spaces of positive operators, *Illinois J. Math.* **38**(1994), 87–94.

[8] A. ES-SAHIB, H. HEINICH, Barycentre canonique pour un espace metrique à courbure négative, in *Séminaire de Probabilités* XXXIII, Lecture Notes in Math., vol. 1709, Springer, Berlin 1999, pp. 355–370.

[9] B. JEURIS, R. VANDEBRIL, B. VANDEREYCKEN, A survey and comparison of contemporary algorithms for computing the matrix geometric mean, *Electron. Trans. Numer. Anal.* **39**(2012), 379–402.

[10] J. JOST, *Nonpositive Curvature: Geometric and Analytic Aspects*, Lectures in Math., ETH Zürich, Birkhäuser, Basel 1997.

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

[12] A. KARLSSON, G.A. MARGULIS, A multiplicative ergodic theorem and nonpositively curved spaces, *Comm. Math. Phys.* **208**(1999), 107–123.

[13] M. KELL, Uniform convex metric spaces, *Anal. Geom. Metr. Spaces* **2**(2014), 359–380.

[14] W. KENDALL, A survey of Riemannian centres of mass for date, in *Proceedings 59th ISI World Statistics Congress*, Hong Kong 2013 (Session STS035), pp. 25–30.

[15] F. KUBO, T. ANDO, Means of positive linear operators, *Math. Ann.* **246**(1980), 205–224.

[16] G. LAROTONDA, Nonpositive curvature: A geometric approach to Hilbert–Schmidt operators, *Differential Geom. Appl.* **25**(2007), 679–700.

[17] J. LAWSON, H. LEE, Y. LIM, Weighted geometric means, *Forum Math.* **24**(2012), 1067–1090.

[18] J. LAWSON, Y. LIM, Metric convexity of symmetric cones, *Osaka J. Math.* **44**(2007), 795–816.

[19] J. LAWSON, Y. LIM, Symmetric spaces with convex metrics, *Forum Math.* **19** (2007), 571–602.

[20] J. LAWSON, Y. LIM, A general framework for extending means to higher orders, *Colloq. Math.* **113**(2008), 191–221.

[21] J. LAWSON, Y. LIM, Monotonic properties of the least squares mean, *Math. Ann.* **351**(2011), 267–279.

[22] J. LAWSON, Y. LIM, Weighted means and Karcher equations of positive operators, *Proc. Natl. Acad. Sci. USA* **110**(2013), 15626–15632.

[23] J. LAWSON, Y. LIM, Karcher means and Karcher equations of positive operators, *Trans. Amer. Math. Soc. Ser. B* **1**(2014), 1–22.

[24] Y. LIM, Geometry of midpoint sets for Thompson's metric, *Linear Algebra Appl.* **439**(2013), 211–227.

[25] Y. LIM, M. PALFIA, Approximations to the Karcher mean on Hadamard spaces via geometric power means, *Forum Math.* **27**(2015), 2609–2635.

[26] A. NAVAS, An L^1 ergodic theorem with values in a non-positively curved space via a canonical barycenter map, *Ergodic Theory Dynam. Systems* **33**(2013), 609–623.

[27] R.D. NUSSBAUM, Hilbert's projective metric and iterated nonlinear maps, *Mem. Amer. Math. Soc.* **75**(1988), no. 391.

[28] K.-T. STURM, Probability measures on metric spaces of nonpositive curvature, in *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces (Paris, 2002)*, Contemp. Math., vol. 338, Amer. Math. Soc., Providence, RI 2003, pp. 357–390.

[29] A.C. THOMPSON, On certain contraction mappings in a partially ordered vector space, *Proc. Amer. Math. Soc.* **14**(1963), 438–443.

[30] C. VILLANI, *Topics in Optimal Transportation*, Grad. Stud. Math., vol. 58, Springer, Berlin 2003.

JIMMIE D. LAWSON, DEPARTMENT OF MATHEMATICS, LOUISIANA STATE UNIVERSITY, BATON ROUGE, LA 70803, U.S.A.
E-mail address: lawson@math.lsu.edu

YONGDO LIM, DEPARTMENT OF MATHEMATICS, SUNGKYUNKWAN UNIVERSITY, SUWON 440-746, KOREA
E-mail address: ylim@skku.edu

Received December 24, 2015; revised July 12, 2016.