Ergodic theorem in Hadamard spaces in terms of inductive means

Jorge Antezana[1,2], Eduardo Ghiglioni[1, 2], and Demetrio Stojanoff[1, 2]

[1]Departamento de de Matemática, FCE-UNLP, Calles 50 y 115, (1900) La Plata, Argentina.
[2] Instituto Argentino de Matemática, 'Alberto P. Calderón', CONICET, Saavedra 15 3er. piso, (1083) Buenos Aires, Argentina.

Abstract

Let $(G, +)$ be a compact, abelian, and metrizable topological group. In this group we take $g \in G$ such that the corresponding automorphism τ_g is ergodic. The main result of this paper is a new ergodic theorem for functions in $L^1(G, M)$, where M is a Hadamard space. The novelty of our result is that we use inductive means to average the elements of the orbit $\{\tau_g^n(h)\}_{n\in\mathbb{N}}$. The advantage of inductive means is that they can be explicitly computed in many important examples. The proof of the ergodic theorem is done firstly for continuous functions, and then it is extended to L^1 functions. The extension is based in a new construction of mollifiers in Hadamard spaces. This construction has the advantage that it only uses the metric structure and the existence of barycenters, and do not require the existence of an underlyng vector space. For this reason, it can be used in any Hadamard space, in contrast with those results that need to use the tangent space or some chart to define the mollifier.[1]

1 Introduction

1.1 Motivation

The initial motivation for this work was the study of ergodic theorems involving the barycenters in the space of $n \times n$ strictly positive matrices $\mathcal{M}_n(\mathbb{C})^+$. More precisely, we pursued an ergodic type theorem in terms of the so called inductive means, which in many important cases can be explicitly computed.

Recall that the set $\mathcal{M}_n(\mathbb{C})^+$ is an open cone in the real vector space of selfadjoint matrices $\mathcal{H}(n)$. In particular, it is a differential manifold and the tangent spaces can

[1]Keywords and phrases: Ergodic theorem, Inductive means, Barycenter, Hadamard space, non-positively curved space.

be indentified for simplicity with $\mathcal{H}(n)$. The manifold $\mathcal{M}_n(\mathbb{C})^+$ can be endowed with a natural Riemannian structure. With respect to this metric structure, if $\alpha : [a,b] \to \mathcal{M}_n(\mathbb{C})^+$ is a piecewise smooth path, its length is defined by

$$L(\alpha) = \int_a^b \|\alpha^{-1/2}(t)\alpha'(t)\alpha^{-1/2}(t)\|_2 \, dt,$$

where $\|\cdot\|_2$ denotes the Frobenius or Hilbert-Schmidt norm. In this way, $\mathcal{M}_n(\mathbb{C})^+$ becomes a Riemannian manifold with non-positive curvature. As usual, a distance δ can be defined by

$$\delta(A,B) = \inf\{L(\alpha) : \ \alpha \text{ is a piecewise smooth path connecting } A \text{ with } B\}.$$

The infimum is actually a minimum, and the geodesic connecting two positive matrices A and B has the following simple expression

$$\gamma_{AB}(t) = A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2}.$$

It is usual in matrix analysis to use the notation $A\#_t B$ instead of $\gamma_{AB}(t)$. The midpoint $A\#_{\frac{1}{2}} B$ is called **geometric mean** or **barycenter** between A and B, and it also admits the following variational characterization

$$A\#_{\frac{1}{2}} B = \arg \min_{C \in \mathcal{M}_n(\mathbb{C})^+} \Big(\delta^2(A,C) + \delta^2(B,C)\Big).$$

There is no reason to restrict our attention to only two matrices. The notion of geometric mean can be generalized for more than two matrices in the obvious way

$$\Gamma(A_1,\ldots,A_n) := \arg \min_{C \in \mathcal{M}_n(\mathbb{C})^+} \Big(\sum_{j=1}^n \delta^2(A_j,C)\Big).$$

The solution of this least square minimization problem exists and is unique because of the convexity properties of the distance $\delta(\cdot,\cdot)$. The geometric means naturally appear in many applied problems. For instance, they appear in the study of radar signals. In these problems, each signal is detected by more than one sensor. The information of each sensor is codified in a covariance kernel, and these kernels have to be averaged to get the final output. It turns out that the best way to average the kernels is not the standard arithmetic mean, but the geometric mean introduced above (see [12] and the references therein for more details). Another typical application of the geometric means is in problems related with the gradient or Newton like optimization methods (see [9],[29]).

The usual problem dealing with geometric means is that the geometric mean of three or more matrices does not have in general a closed formula. For this reason, there has been an intensive research with the aim to find good ways to approximate the geometric mean of more than two matrices ([8], [19], [25]). In [15], Holbrook proved that they can be approximated by the so called **inductive means**. To motivate the definition of inductive means, note that given a sequence $\{a_n\}_{n\in\mathbb{N}}$ of complex numbers

$$\begin{aligned}\frac{a_1+a_2+a_3}{3} =& \frac{2}{3}\left(\frac{a_1+a_2}{2}\right)+\frac{1}{3}\,a_3\\ &\vdots\\ \frac{a_1+\ldots+a_n}{n} =& \frac{n-1}{n}\left(\frac{a_1+\ldots+a_{n-1}}{n-1}\right)+\frac{1}{n}\,a_n.\end{aligned}$$

Let $\gamma_{a,b}(t) = t\,b + (1-t)a$, and for a moment allow us to use the notation $a\,\natural_t\,b = \gamma_{a,b}(t)$. Then

$$\begin{aligned}\frac{a_1+a_2+a_3}{3} =& (a_1\,\natural_{\frac{1}{2}}\,a_2)\,\natural_{\frac{1}{3}}\,a_3\\ \frac{a_1+a_2+a_3+a_4}{4} =& ((a_1\,\natural_{\frac{1}{2}}\,a_2)\,\natural_{\frac{1}{3}}\,a_3)\,\natural_{\frac{1}{4}}\,a_4.\end{aligned}$$

and so on and so forth. The segments are the geodesics in the euclidean space. Thus, in our setting, we can replace the segments by the geodesic associated to the Riemannian structure. This is the idea that leads to the definition of the inductive means. Given a sequence of strictly positive matrices $\{A_n\}_{n\in\mathbb{N}}$, the inductive means are defined as follows:

$$\begin{aligned}&\Gamma_1(A) = A_1\\ &\Gamma_n(A) = \Gamma_{n-1}(A)\#_{\frac{1}{n}}A_n \qquad (n \geq 2).\end{aligned}$$

Now, consider d positive matrices $A_0, \ldots, A_{d-1}$, and define the function $F : \mathbb{Z}_d \to \mathcal{M}_n(\mathbb{C})^+$ by $F(\overline{k}) = A_k$, where $\mathbb{Z}_d$ denotes the abelian group of integers mod d. Then, if we define the periodic sequence $A = \{F(\overline{n})\}_{n\in\mathbb{N}}$, then Holbrook proved

$$\lim_{n\to\infty} \Gamma_n(A) = \Gamma(A_0, \ldots, A_{d-1}).$$

1.2 About our results

Let $(G, +)$ be a compact topological group, endowed with a Haar measure m, and let $\tau : G \to G$ be an ergodic map. The classical Birkhoff ergodic theorem says that, given $f \in L^1(G)$, then for m-almost every $h \in G$

$$\frac{1}{n}\sum_{k=0}^{n-1} f(\tau^k(\omega)) \xrightarrow[n\to\infty]{} \int_G f(\omega)dm(\omega). \tag{1.1}$$

Holbrook's theorem can be seen as an ergodic theorem for functions defined in $\mathbb{Z}^p$ and taking values in $\mathcal{M}_n(\mathbb{C})^+$. Indeed, let the inductive means play the role of the averages at the left hand side, and let the joint geometric mean $\Gamma(A_0, \ldots, A_{d-1})$ play the role of the integral of F with respect to the Haar measure in $\mathbb{Z}_d$.

This interpretation of Holbrook result suggests that we can get computable approximations of the geometric means of several matrices in terms of ergodic averages. Our main goal is to prove such ergodic theorems in terms of the inductive means in a much more general context. More precisely, consider a dynamical system (G, τ), where G is a compact group and τ is ergodic with respect to the Haar measure m. Note that the group has to be abelian because the orbits $\{\tau^n g\}_{n\in\mathbb{N}}$ are dense in G. In this setting recall that given a function $F : G \to \mathcal{M}_n(\mathbb{C})^+$, we say that $F \in L^p(G, \mathcal{M}_n(\mathbb{C})^+)$ if

$$\int_G \delta^p(F(g), B))\, dm(g) < \infty,$$

where B is any positive matrix. By the triangular inequality, this definition does not depend on the choice of B. Following Sturm [36], the barycenter of $F \in L^1(G, \mathcal{M}_n(\mathbb{C})^+)$ is defined as

$$\beta_F := \arg \min_{C\in\mathcal{M}_n(\mathbb{C})^+} \int_G \delta^2(F(g), C)) - \delta^2(F(g), B))\, dm(g).$$

As before, this definition does not depend on B. Note that if $F \in L^2(G, \mathcal{M}_n(\mathbb{C})^+)$

$$\beta_F := \arg \min_{C\in\mathcal{M}_n(\mathbb{C})^+} \int_{\mathbb{T}} \delta^2(F(g), C))dm(g)$$

and we get a natural generalization of the geometric means defined before. In this paper we prove that, given $A \in L^1(G, \mathcal{M}_n(\mathbb{C})^+)$, for almost every $g \in G$

$$\lim_{n\to\infty} \Gamma_n\big(A(g), A(\tau(g)), \ldots, A(\tau^{n-1}(g))\big) = \beta_A. \tag{1.2}$$

Moreover, we prove this result not only for functions taking values in $\mathcal{M}_n(\mathbb{C})^+$, but also in any Hadamard space M(see subsection 2.1 for a precise definition). In [36], Sturm developed a theory of barycenters of probability measures for Hadamard spaces (see Subsection 2.2 for some definitions and basic results). Endowed with this barycenter, Hadamard spaces play an important role in the theory of integrations (random variables, expectations and variances), law of large numbers, ergodic theory, Jensen's inequality (see [10], [14], [22], [30], and [36]), stochastic generalization of Lipschitz retractions and extension problems of Lipschitz and Hölder maps (see [23], [28], and [31]) and optimal transport theory on Riemannian manifolds (see [33], and [34]), etc.

The generalization from $\mathcal{M}_n(\mathbb{C})^+$, or even from more general Riemannian manifolds with non-positive curvature, to general Hadamard spaces is not straightforward. One of the reasons is that a general Hadamard space does not necessarily have an underlying finite dimensional vector space, as in the case of manifolds. This makes some steps of the proof much more involved, and lead us to a new definition of mollifiers that only uses the metric structure (see Subsection 3.3.1).

1.3 Comments on related works

In [3] Austin proved a very general ergodic theorem for Hadamard spaces, which in our setting says that, given $A \in L^2(G, M)$, for almost every $g \in G$ it holds that

$$\lim_{n\to\infty} \Gamma\big(A(g), A(\tau(g)), \ldots, A(\tau^{n-1}(g))\,\big) = \beta_A.$$

Later on, in [30] Navas extended the Austin's result to functions in $A \in L^1(G, M)$, and taking values in more general metric spaces. In these results, the discrete arithmetic means of Birkhoff's theorem are replaced by the (joint) geometric mean of the n-tuple

$$\big(A(g), A(\tau(g)), \ldots, A(\tau^{n-1}(g))\big).$$

On the other hand, the integral in Birkhoff theorem is replaced by the barycenter β_A. Note that in the sequence

$$A(g),\, \Gamma\Big(A(g), A(\tau(g))\Big),\, \Gamma\Big(A(g), A(\tau(g), A(\tau^2(g))\Big),\, \ldots$$

from the second term on, the elements of the sequence do not have a closed formula. In this direction, the inductive means are simpler and provide in many concrete instances a computable approximation sequence to the barycenter. This is an advantage of our result, but there is a price to pay for this. On one hand, we need a good control on the convexity of the metric. For this reason we work on Hadamard spaces, as in the case of Austin's result [3]. On the other hand, the inductive means are simple averages of points. This simplicity is good from the computational point of view, but it confine our result to $\mathbb{Z}$-actions. The aforementioned results also hold for more general actions.

1.4 Organization of the paper

The paper is organized as follows. Section 2 is devoted to collect some preliminaries on Hadamard spaces, as well as, barycenters and inductive means in Hadamard spaces. In section 3 we prove our main result for functions defined in a Kronecker systems and taking values in a Hadamard space. Firstly we will prove the result for continuous function (Theorem 3.1). In order to extend this result to L^1 functions (Theorem 3.7), first of all we prove in Subsection 3.3.1 some results related with approximation by continuous functions in general Hadamard spaces. These results are interesting by themselves, and generalize some results proved by Karcher in [17]. Finally, in the Subsection 3.3.2 we complete the proof of the L^1 version of the ergodic theorem.

2 Preliminaries

2.1 Hadamard spaces

In this section we summarize some basic facts about Hadamard spaces, also called (global) CAT(0) spaces or non-positively curved (NPC) spaces. This subject started with the works by Alexandrov [1] and Reshetnyak [35]. Nowadays there exists a huge bibliography on the subject. The interested reader is referred to the monographs [4], [5], [11], and [16] for more information.

Definition 2.1. *A complete metric space (M, δ) is called a **Hadamard space** if it satisfies the semiparallelogram law, i.e., for each $x, y \in M$ there exists $m \in M$ satisfying*

$$\delta^2(m, z) \leq \frac{1}{2}\delta^2(x, z) + \frac{1}{2}\delta^2(y, z) - \frac{1}{4}\delta^2(x, y) \tag{2.1}$$

for all $z \in M$. The point m is called (metric) ***midpoint*** *between x and y.*

Taking $z = x$ and $z = y$ in the inequality (2.1), it is easy to conclude that $\delta(x, m) = \delta(m, y) = \frac{1}{2}\delta(x, y)$. Moreover, this inequality also implies that the midpoint is unique. The existence and uniqueness of midpoints give rise to a unique (metric) geodesic $\gamma_{a,b} : [0, 1] \to M$ connecting any given two points a and b. Indeed, firstly define $\gamma_{a,b}(1/2)$ to be the midpoint of a and b. Then, using an inductive argument, we define the geodesic for all dyadic rational numbers in $[0, 1]$. Finally by completeness, it can be extended to all $t \in [0, 1]$. Throughout this paper, we will use the notation $a\#_t b$ instead of $\gamma_{a,b}(t)$.

The inequality (2.1) also extends to arbitrary points on geodesics.

Proposition 2.2. *Let (M, δ) be a Hadamard space. Then, for all $t \in [0, 1]$ and $x, y, z \in M$,*

$$\delta^2(x\#_t y, z) \leq (1-t)\delta^2(x, z) + t\delta^2(y, z) - t(1-t)\delta^2(x, y). \tag{2.2}$$

A consequence of this result, that we will use later, is:

Corollary 2.3. *Given four points $a, a^{'}, b, b^{'} \in M$ let*

$$f(t) = \delta(a\#_t a^{'}, b\#_t b^{'}).$$

Then f is convex on $[0, 1]$; i.e.

$$\delta(a\#_t a^{'}, b\#_t b^{'}) \leq (1-t)\delta(a, b) + t\delta(a^{'}, b^{'}). \tag{2.3}$$

We conclude this subsection with the so called Reshetnyak's Quadruple Comparison theorem.

Theorem 2.4. *Let (M, δ) be a Hadamard space. For all $x_1, x_2, x_3, x_4 \in M$,*

$$\delta^2(x_1, x_3) + \delta^2(x_2, x_4) \leq \delta^2(x_2, x_3) + \delta^2(x_1, x_4) + 2\delta(x_1, x_2)\delta(x_3, x_4). \tag{2.4}$$

2.2 Barycenters

Let (M, δ) be a Hadamard space, and let $\mathcal{B}(M)$ the σ-algebra of Borel sets (i.e. the smallest σ-algebra that contains the open sets). Denote by $\mathcal{P}(M)$ the set of all probability measures on $\mathcal{B}(M)$ with separable support, and for $1 \leq \theta < \infty$, let $\mathcal{P}^\theta(M)$ denote the set of $\mu \in \mathcal{P}(M)$ such that

$$\int \delta^\theta(x, y)d\mu(y) < \infty,$$

for some (hence all) $x \in M$. By means of $\mathcal{P}^\infty(M)$ we will denote the set of all measures in $\mathcal{P}(M)$ with bounded support.

Proposition 2.5. *Let (M,δ) be a Hadamard space and fix $y \in M$. For each $\mu \in \mathcal{P}^1(M)$ there exists a unique point $\beta_\mu \in M$ which minimizes the uniformly convex, continuous function*

$$z \mapsto \int_M [\delta^2(z,x) - \delta^2(y,x)] d\mu(x).$$

This point is independent of y.

Following Sturm's paper [36], the point β_μ is called **barycenter** of μ. If $\mu \in \mathcal{P}^2(M)$ then β_μ coincides with the usual Cartan's definition of barycenter:

$$\operatorname*{argmin}_{z\in M} \int_M \delta^2(z,x) d\mu(x).$$

The following inequality satisfied by the baricenter will be very important in the sequel.

Proposition 2.6 (Variance Inequality)**.** *Let (M,δ) be a Hadamard space. For any probability measure $\mu \in \mathcal{P}^1(M)$ and for all $z \in M$:*

$$\int_M [\delta^2(z,x) - \delta^2(\beta_\mu, x)] d\mu(x) \geq \delta^2(z, \beta_\mu). \tag{2.5}$$

Now, let (Ω, P) be an arbitrary probability space and let $F : \Omega \to M$ be a measurable map. This function defines a probability measure $F_*P \in \mathcal{P}(M)$ by

$$F_*P(A) := P(F^{-1}(A)) = P(\{\omega \in \Omega : F(\omega) \in A\}) \qquad (\forall A \in \mathcal{B}(M)),$$

which is called **pushforward measure** of P by the function F. In probabilistic language, the pushforward measure F_*P is called distribution of F.

Given $1 \leq \theta \leq \infty$, we say that $F \in L^\theta(\Omega, M)$ if $F_*P \in \mathcal{P}^\theta(M)$. In other words, for $1 \leq \theta < \infty$, we say that $F \in L^\theta(\Omega, M)$ if for some (and hence for all) $y \in M$ it holds that

$$\int_\Omega \delta^\theta(F(\omega), y)\, dP(\omega) < \infty. \tag{2.6}$$

On the other hand, we say that $F \in L^\infty(\Omega, M)$ if for some (and hence for all) $y \in M$ the function $\omega \mapsto \delta(F(\omega), y)$ is essentially bounded.

2.3 The inductive mean

As in the case of strictly positive matrices considered in the introduction, we define the inductive means in general Hadamard spaces as follows.

Definition 2.7. *(Inductive mean). Let (M,δ) be a Hadamard space. Given $a \in M^{\mathbb{N}}$ set*

$$\begin{aligned} S_1(a) &= a_1 \\ S_n(a) &= S_{n-1}(a) \#_{\frac{1}{n}} a_n \qquad (n \geq 2). \end{aligned}$$

Example. Suppose that M is $\mathbb{C}^n$ with the usual euclidean distance. Since the geodesics in this case are the line segments, if we take a sequence $\{a_n\}_{n\in\mathbb{N}}$ in $\mathbb{C}^n$, then

$$\begin{aligned} S_1(a) &= a_1, \\ S_2(a) &= \frac{a_1 + a_2}{2}, \\ S_3(a) &= \frac{2}{3}\left(\frac{a_1 + a_2}{2}\right) + \frac{1}{3}\, a_3 = \frac{a_1 + a_2 + a_3}{3}, \end{aligned}$$

and so on and so forth. Therefore, in this case the inductive means coincides the the arithmetic means. ▲

From now on, let (M, δ) be a Hadamard space. As a consequence of (2.3), we directly get the following result.

Corollary 2.8. *Given $a, b \in M^{\mathbb{N}}$, then*

$$\delta(S_n(a), S_n(b)) \leq \frac{1}{n} \sum_{i=1}^{n} \delta(a_i, b_i). \tag{2.7}$$

The next lemma follows from (2.2), and it is a special case of a weighted inequality considered by Lim and Pálfia in [26].

Lemma 2.9. *Given $a \in M^{\mathbb{N}}$ and $z \in M$, for every $k, m \in \mathbb{N}$*

$$\begin{aligned} \delta^2(S_{k+m}(a), z) \leq\ & \frac{k}{k+m}\ \delta^2(S_k(a), z) + \frac{1}{k+m} \sum_{j=0}^{m-1} \delta^2(a_{k+j+1}, z) \\ & - \frac{k}{(k+m)^2} \sum_{j=0}^{m-1} \delta^2(S_{k+j}(a), a_{k+j+1}). \end{aligned}$$

Proof. By the inequality (2.2) applied to $S_{n+1}(a) = S_n(a) \,\#_{n+1}\, (a_{n+1})$ we obtain

$$(n+1)\ \delta^2(S_{n+1}(a), z) - n\ \delta^2(S_n(a), z) \leq \delta^2(a_{n+1}, z) - \frac{n}{(n+1)} \delta^2(S_n(a), a_{n+1}).$$

Summing these inequalities from $n = k$ until $n = k + m - 1$ we get that the difference

$$(k+m)\ \delta^2(S_{k+m}(a), z) - k\ \delta^2(S_k(a), z),$$

obtained from the telescopic sum of the left hand side, is less or equal than

$$\sum_{j=0}^{m-1} \left(\delta^2(a_{k+j+1}, z) - \frac{k+j}{(k+j+1)} \delta^2(S_{k+j}(a), a_{k+j+1}) \right).$$

Finally, using that $\frac{k+j}{k+j+1} \geq \frac{k}{k+m}$ for every $j \in \{0, \ldots, m-1\}$, this sum is bounded from the above by

$$\sum_{j=0}^{m-1} \left(\delta^2(a_{k+j+1}, z) - \frac{k}{(k+m)} \delta^2(S_{k+j}(a), a_{k+j+1}) \right),$$

which completes the proof. ■

Given a sequence $a \in M^{\mathbb{N}}$, let $\Delta(a)$ denote the diameter of its image, that is

$$\Delta(a) := \sup_{n,m\in\mathbb{N}} \delta(a_n, a_m).$$

Note that, also by (2.2), $\delta(S_n(a), a_k) \leq \Delta(a)$ for all $n, k \in \mathbb{N}$.

Lemma 2.10. *Given $a \in M^{\mathbb{N}}$ such that $\Delta(a) < \infty$, then for all $k, m \in \mathbb{N}$ it holds that*

$$\frac{1}{m}\sum_{j=0}^{m-1} \delta^2(S_k(a), a_{k+j+1}) \leq \tilde{R}_{m,k} + \frac{1}{m}\sum_{j=0}^{m-1} \delta^2(S_{k+j}(a), a_{k+j+1}).$$

where $\tilde{R}_{m,k} = \left(\frac{m^2}{(k+1)^2} + 2\frac{m}{k+1}\right)\Delta^2(a)$.

Proof. Note that by (2.7) and all k,

$$\delta(S_{k+j}(a), S_{k+j+1}(a)) \leq \frac{1}{k+j+1}\Delta(a).$$

Hence

$$\begin{aligned}
\delta(S_k(a), a_{k+j+1}) &\leq \delta(S_k(a), S_{k+j}(a)) + \delta(S_{k+j}(a), a_{k+j+1}) \\
&\leq \sum_{h=1}^{j} \frac{1}{k+h}\Delta(a) + \delta(S_{k+j}(a), a_{k+j+1}) \\
&\leq \frac{m}{k+1}\Delta(a) + \delta(S_{k+j}(a), a_{k+j+1}).
\end{aligned}$$

Therefore, , for every $j \leq m$

$$\delta^2(S_k(a), a_{k+j+1}) \leq \left(\frac{m^2}{(k+1)^2} + 2\frac{m}{k+1}\right)\Delta^2(a) + \delta^2(S_{k+j}(a), a_{k+j+1}),$$

where we have used that $\delta(S_{k+j}(a), a_{k+j+1}) \leq \Delta(a)$ for every $k, j \in \mathbb{N}$. Summing up these inequalities and dividing by m, we get the desired result. ■

3 Ergodic formulae associated to inductive means

3.1 The framework and basic notation

Let $(G, +)$ be a compact, abelian, and metrizable topological group. In this group we fix a Haar measure m, and we take an ergodic automorphism $\tau(h) = h + g$ for some $g \in G$. A shift invariant metric G is denoted by d_G. Throughout this section we work with the dynamical system (G, τ).

Remark. A topological dynamical system (Ω, τ) is called a **Kronecker system** if it is isomorphic to a group dynamical system (G, τ) as the one described above. Every topological Kronecker system $(\Omega; x \to x + \alpha)$ can be canonically converted into a measure-preserving system which is compact. It is well known that any isometric (or equicontinuous) and minimal dynamical system is a Kronecker system.

On the other hand, we will fix a Hadamard space (M, δ). Given a function $A : G \to M$, we define $a^\tau : G \to M^{\mathbb{N}}$ by

$$a^\tau(x) := \{a_j^\tau(x)\}_{n\in\mathbb{N}} \quad \text{where} \quad a_j^\tau(x) = A(\tau^j(x)). \tag{3.1}$$

3.2 The continuous case

In this section we will prove the ergodic formula for continuous functions.

Theorem 3.1. *Let M be a Hadarmard space and $A : G \to M$ a continuous function. Then*

$$\lim_{n\to\infty} S_n(a^\tau(g)) = \beta_A, \tag{3.2}$$

uniformly in $g \in G$.

With this aim, we firstly prove some technical results, which at the end of this subsection are combined to get a proof of Theorem 3.1.

Lemma 3.2. *Let $A : G \to M$ be a continuous function, and let K be any compact subset of M. For each $n \in \mathbb{N}$, define $F_n : G \times K \to \mathbb{R}$ by*

$$F_n(g, x) = \frac{1}{n}\sum_{j=0}^{n-1} \delta^2(a_j^\tau(g), x).$$

Then, the family $\{F_n\}_{n\in\mathbb{N}}$ is equicontinuous.

Proof. By the triangular inequality, the map $y \mapsto \delta^2(A(\cdot), y)$ is continuous from (K, δ) into the set of real valued continuous functions defined on G endowed with the uniform norm. Since K is compact, the family $\{\delta^2(A(\cdot), x)\}_{x\in K}$ is (uniformly) equicontinuous. Hence, given $\varepsilon > 0$, there exists $\delta > 0$ such that if

$$d_G(g_1, g_2) < \delta \quad \text{then} \quad |\delta^2(A(g_1), x) - \delta^2(A(g_2), x)| < \frac{\varepsilon}{2},$$

for every $x \in K$. Since τ is isometric and $d_G(g_1, g_2) < \delta$, we get that

$$|F_n(g_1, x) - F_n(g_2, x)| = \left|\frac{1}{n}\sum_{j=0}^{n-1} \delta^2(a_j^\tau(g_1), x) - \delta^2(a_j^\tau(g_2), x)\right| < \frac{\varepsilon}{2}.$$

Let Δ be the diameter of the set (Image$(A) \times K$) in M^2. Since both sets are compact, $\Delta < \infty$. So, take (g_1, x_1) and (g_2, x_2) such that $d_G(g_1, g_2) < \delta$ and $\delta(x_1, x_2) < \frac{\varepsilon}{4\Delta}$. Then

$$\begin{aligned}|F_n(g_1, x_1) - F_n(g_2, x_2)| &\le |F_n(g_1, x_1) - F_n(g_1, x_2)| + |F_n(g_1, x_2) - F_n(g_2, x_2)| \\ &\le \frac{2\Delta}{n}\sum_{k=0}^{n-1} \delta(x_1, x_2) \; + \; \frac{\varepsilon}{2} < \varepsilon.\end{aligned}$$

■

Now, as a consequence of Arzelà-Ascoli and Birkhoff theorems we get

Proposition 3.3. *Let $A : G \to M$ be a continuous function, and K a compact subset of M. Then*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} \delta^2(a_j^\tau(g), x) = \int_G \delta^2(A(\gamma), x)dm(\gamma),$$

and the convergence is uniform in $(g, x) \in G \times K$.

From now on we will fix the continuous function $A : G \to M$. Let

$$\alpha := \min_{x\in M} \int_G \delta^2(A(g), x)dm(g),$$

and β_A is the point where this minimum is attained, i.e., β_A is the barycenter of the pushforward by A of the Haar measure in G. Then we obtain the following upper estimate.

Lemma 3.4. *For every $\varepsilon > 0$, there exists $m_0 \in \mathbb{N}$ such that, for all $m \geq m_0$ and for all $k \in \mathbb{N}$,*

$$\delta^2(S_{k+m}(a^\tau(g)), \beta_A) \leq \frac{k}{k+m}\delta^2(S_k(a^\tau(g)), \beta_A) + \frac{m}{k+m}(\alpha+\varepsilon) - \frac{km}{(k+m)^2}\left(\frac{1}{m}\sum_{j=0}^{m-1} \delta^2\big(S_{k+j}(a^\tau(g)), a_{k+j+1}^\tau(g)\big)\right).$$

Proof. For every $\varepsilon > 0$, exists $m_0 \in \mathbb{N}$ such that for all $m \geq m_0$,

$$\left|\frac{1}{m}\sum_{j=0}^{m-1} \delta^2(a_{k+j+1}^\tau(g), \beta_A) - \alpha\right| < \varepsilon.$$

Note that m_0 is independient of k by Proposition 3.3. Now, by Lemma 2.9

$$\begin{aligned}
\delta^2(S_{k+m}(a^\tau(g)), \beta_A) &\leq \frac{k}{k+m}\ \delta^2(S_k(a^\tau(g)), \beta_A) + \frac{1}{k+m}\sum_{j=0}^{m-1} \delta^2(a_{k+j+1}^\tau(g), \beta_A) \\
&\quad - \frac{k}{(k+m)^2}\sum_{j=0}^{m-1} \delta^2(S_{k+j}(a^\tau(g)), a_{k+j+1}^\tau(g)) \\
&= \frac{k}{k+m}\ \delta^2(S_k(a^\tau(g)), \beta_A) + \frac{m}{k+m}\left(\frac{1}{m}\sum_{j=0}^{m-1} \delta^2(a_{k+j+1}^\tau(g), \beta_A)\right) \\
&\quad - \frac{km}{(k+m)^2}\left(\frac{1}{m}\sum_{j=0}^{m-1} \delta^2(S_{k+j}(a^\tau(g)), a_{k+j+1}^\tau(g))\right) \\
&\leq \frac{k}{k+m}\ \delta^2(S_k(a^\tau(g)), \beta_A) + \frac{m}{k+m}(\alpha+\varepsilon) \\
&\quad - \frac{km}{(k+m)^2}\left(\frac{1}{m}\sum_{j=0}^{m-1} \delta^2(S_{k+j}(a^\tau(g)), a_{k+j+1}^\tau(g))\right).
\end{aligned}$$

∎

Recall that, given a sequence $a \in M^{\mathbb{N}}$, then $\Delta(a)$ denotes the diameter of its image, i.e.,

$$\Delta(a) := \sup_{n,m\in\mathbb{N}} \delta(a_n, a_m).$$

Since $A : G \to M$ is continuous, note that

$$C_a := \sup_{g\in G} \Delta(a^\tau(g)) < \infty.$$

Lemma 3.5. *For every $\varepsilon > 0$, there exists $m_0 \in \mathbb{N}$ such that for all $m \geq m_0$ and for all $k \in \mathbb{N}$*

$$\delta^2(S_k(a^\tau(g)), \beta_A) - \varepsilon + \alpha - R_{m,k} \leq \frac{1}{m}\sum_{j=0}^{m-1} \delta^2(S_{k+j}(a^\tau(g)), a^\tau_{k+j+1}(g)),$$

where $R_{m,k} = \left(\frac{m^2}{(k+1)^2} + 2\frac{m}{k+1}\right) C_a^2$.

Proof. Consider the compact set

$$K := \overline{cc\,\{S_k(a^\tau(x)) : k \in \mathbb{N}\}},$$

where the convex hull is in the geodesic sense. For every $\varepsilon > 0$, there exists $m_0 \in \mathbb{N}$ such that for all $m \geq m_0$, by the variance inequality (Proposition 2.6) and Proposition 3.3, it holds that

$$\begin{aligned}\delta^2(S_k(a^\tau(x)), \beta_A) &\leq \int_G \delta^2\big(S_k(a^\tau(g)), A(\gamma)\big)\, dm(\gamma) \;-\; \alpha \\ &\leq \varepsilon + \frac{1}{m}\sum_{j=0}^{m-1} \delta^2(S_k(a^\tau(g)), a^\tau_{k+j+1}(g)) - \alpha.\end{aligned}$$

Finally, by Lemma 2.10

$$\delta^2(S_k(a^\tau(x)), \beta_A) \leq \varepsilon + \frac{1}{m}\sum_{j=0}^{m-1} \delta^2(S_{k+j}(a^\tau(x)), a^\tau_{k+j+1}(x)) + R_{m,k} - \alpha,$$

where $R_{m,k} = \left(\frac{m^2}{(k+1)^2} + 2\frac{m}{k+1}\right) C_a$. ■

Lemma 3.6. *Given $\varepsilon > 0$, there exists $m_0 \geq 1$ such that for every $\ell \in \mathbb{N}$*

$$\delta^2(S_{\ell m_0}(a^\tau(g)), \beta_A) \leq \frac{L}{\ell} + \varepsilon,$$

uniformly in $g \in G$, where $L = \alpha + 3C_a^2$.

Proof. Fix $\varepsilon > 0$. By Lemmas 3.4 and 3.5, there exists $m_0 \geq 1$ such that for all $k \in \mathbb{N}$,

$$\delta^2(S_{k+m_0}(a^\tau(g)), \beta_A) \leq \frac{k}{k+m_0}\delta^2(S_k(a^\tau(g)), \beta_A) + \frac{m_0}{k+m_0}(\alpha+\varepsilon) - \frac{km_0}{(k+m_0)^2}\left(\frac{1}{m_0}\sum_{j=0}^{m_0-1}\delta^2(S_{k+j}(a^\tau(x)), a^\tau_{k+j+1}(x))\right),$$

and

$$\frac{1}{m_0}\sum_{j=0}^{m_0-1}\delta^2(S_{k+j}(a^\tau(g)), a^\tau_{k+j+1}(g)) \geq \delta^2(S_k(a^\tau(g)), \beta_A) - \varepsilon + \alpha - R_{m_0,k}.$$

Therefore, combining these two inequalities we obtain

$$\delta^2(S_{k+m_0}(a^\tau(g)), \beta_A) \leq \frac{k}{k+m_0}\delta^2(S_k(a^\tau(g)), \beta_A) + \frac{m_0}{k+m_0}(\alpha+\varepsilon) - \frac{km_0}{(k+m_0)^2}\left(\delta^2(S_k(a^\tau(g)), \beta_A) - \varepsilon + \alpha - R_{m_0,k}\right).$$

Consider now the particular case where $k = \ell m_0$. Since $R_{m_0,\ell m_0} \leq \frac{3}{\ell}C_a^2$ we get

$$\begin{aligned}\delta^2(S_{(\ell+1)m_0}(a^\tau(g)), \beta_A) &\leq \frac{\ell}{\ell+1}\delta^2(S_{\ell m_0}(a^\tau(g)), \beta_A) + \frac{1}{l+1}(\alpha+\varepsilon) - \\ &\quad \frac{\ell}{(\ell+1)^2}\left(\delta^2(S_{\ell m_0}(a^\tau(g)), \beta_A) - \varepsilon + \alpha - R_{m_0,\ell m_0}\right) \\ &\leq \frac{\ell^2\ \delta^2(S_{\ell m_0}(a^\tau(g)), \beta_A) + (2\ell+1)\varepsilon + \alpha + 3C_a^2}{(\ell+1)^2}. \end{aligned} \tag{3.3}$$

Using this recursive inequality, the result follows by induction on ℓ. Indeed, if $\ell = 1$ then

$$\delta^2(S_{m_0}(a^\tau(g)), \beta_A) \leq C_a^2 \leq L.$$

On the other hand, if we assume that the result holds for some $\ell \geq 1$, i.e.

$$\delta^2(S_{\ell m_0}(a^\tau(x)), g) \leq \frac{L}{\ell} + \varepsilon,$$

then combining this inequality with (3.3) we have that

$$\delta^2(S_{(\ell+1)m_0}(a^\tau(g)), \beta_A) \leq \frac{\ell L + \ell^2\varepsilon + (2\ell+1)\varepsilon + \alpha + 3C_a^2}{(\ell+1)^2} = \frac{L}{\ell+1} + \varepsilon.$$

∎

Now we are ready to prove the ergodic formula for continuous functions.

of Theorem 3.1. Given $\varepsilon > 0$, by Lemma 3.6, there exists $m_0 \in \mathbb{N}$ such that,

$$\delta^2(S_{\ell m_0}(a^\tau(g)), \beta_A) \le \frac{L}{\ell} + \frac{\varepsilon^2}{8},$$

for every $\ell \in \mathbb{N}$. Take $\ell_0 \in \mathbb{N}$ such that for all $\ell \ge \ell_0$,

$$\delta^2(S_{\ell m_0}(a^\tau(g)), \beta_A) \le \frac{\varepsilon^2}{4}. \tag{3.4}$$

Let $n = \ell m_0 + d$ such that $\ell \ge \ell_0$ and $d \in \{1, \ldots, m_0 - 1\}$. Since $x\#_t x = x$ for all $x \in M$, using Corollary 2.8 with the sequences

$$(\ a_1^\tau(g), \ldots, a_{\ell m_0}^\tau(g), \underbrace{S_{\ell m_0}(a^\tau(g)), \ldots, S_{\ell m_0}(a^\tau(g))}_{\text{d times}}\ )$$

and

$$(\ a_1^\tau(g), \ldots, a_{\ell m_0}^\tau(g),\ a_{\ell m_0+1}^\tau(g)\ ,\ \ldots\ ,\ a_{\ell m_0+d}^\tau(g)\ ).$$

Taking into account that $\delta(S_{\ell m_0}(a^\tau(g)), a_{\ell m_0+j}^\tau(g)) \le C_a$ for every $j \in \{1, \ldots, m_0 - 1\}$, we get

$$\begin{aligned} \delta(S_{\ell m_0}(a^\tau(g)), S_{\ell m_0+d}(a^\tau(g))) &\le \frac{1}{\ell m_0 + d} \sum_{j=1}^{d} \delta(S_{\ell m_0}(a^\tau(g)), a_{\ell m_0+j}^\tau(g)) \\ &\le \frac{d}{\ell m_0 + d} C_a \le \frac{1}{\ell} C_a \xrightarrow[k\to\infty]{} 0. \end{aligned}$$

Combining this with (3.4) we obtain that for n big enough $\delta(S_n(a^\tau(g)), \beta_A) < \varepsilon$. ■

3.3 The L^1 case

The natural framework for the ergodic theorem is L^1. In this section we will prove the following ergodic theorem for functions in $L^1(G, M)$ in terms of the inductive means, which is the main result of this paper.

Theorem 3.7. *Given $A \in L^1(G, M)$, for almost every $g \in G$*

$$\lim_{n\to\infty} S_n(a^\tau(g)) = \beta_A. \tag{3.5}$$

Remark. Let M be the field of complex numbers with the usual distance. As we observed in the Example 2.3, the inductive means $S_n(a^\tau(g))$ become the usual arithmetic mean

$$\frac{1}{n} \sum_{k=0}^{n-1} A(\tau^n(g)).$$

On the other hand, the barycenter β_A is just the integral of A with respect to the Haar measure m. So, equation (3.5) takes the form

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}A(\tau^n(g))=\int_G A(g)\,dm(g)$$

which is the usual Birkhoff ergodic theorem. ▲

The strategy of the proof consists in constructing good approximations by continuous functions, and get the result of L^1 functions as a consequence of the theorem for continuous functions (Theorem 3.1 above). So, the first questions that appear are: what does good approximation mean?, and what should we require to the approximation in order to get the L^1 case as a limit of the continuous case? The next two lemmas contain the clue to answer these two questions.

Lemma 3.8. *Let $(\Omega,\mathcal{B},P)$ be a probability space, and $A,B\in L^1(X,M)$. If*

$$\beta_A=\operatorname*{argmin}_{z\in M}\int_\Omega[\delta^2(A(\omega),z)-\delta^2(A(\omega),y)]\;dP(\omega),$$
$$\beta_B=\operatorname*{argmin}_{z\in M}\int_\Omega[\delta^2(B(\omega),z)-\delta^2(B(\omega),y)]\;dP(\omega),$$

then

$$\delta(\beta_A,\beta_B)\le\int_\Omega\delta(A(\omega),B(\omega))dP(\omega). \tag{3.6}$$

Remark. Recall that the definition of β_A (resp. β_B) does not depend on the chosen $y\in M$.

Proof. By the variance inequality (Proposition 2.6) we get

$$\delta^2(\beta_A,\beta_B)\le\int_\Omega\delta^2(\beta_A,B(\omega))-\delta^2(\beta_B,B(\omega))dP(\omega),$$
$$\delta^2(\beta_A,\beta_B)\le\int_\Omega\delta^2(\beta_B,A(\omega))-\delta^2(\beta_A,A(\omega))dP(\omega),$$

and the combination of these two inequalities leads to

$$\begin{aligned}2\delta^2(\beta_A,\beta_B)\le\int_\Omega&\delta^2(\beta_A,B(\omega))+\delta^2(\beta_B,A(\omega))\\&-\delta^2(\beta_B,B(\omega))-\delta^2(\beta_A,A(\omega))dP(\omega).\end{aligned}$$

Finally, using the Reshetnyak's cuadruple comparison (Theorem 2.4) we obtain

$$2\delta^2(\beta_A,\beta_B)\le2\delta(\beta_A,\beta_B)\int_\Omega\delta(A(\omega),B(\omega))\,dP(\omega),$$

which is, after an algebraic simplification, the desired result. ■

Lemma 3.9. *Let $A, B \in L^1(G, M)$. Given $\varepsilon > 0$, for almost every $g \in G$ there exists n_0, which may depends on g, such that*

$$\delta\big(S_n(a^\tau(g)), S_n(b^\tau(g))\big) \leq \varepsilon + \int_G \delta(A(g), B(g))dm(g), \tag{3.7}$$

provided $n \geq n_0$.

Proof. Indeed, by Corollary 2.8

$$\delta\big(S_n(a^\tau(g)), S_n(b^\tau(g))\big) \leq \frac{1}{n}\sum_{k=0}^{n-1} \delta\big(a_k^\tau(g), b_k^\tau(g)\big) = \frac{1}{n}\sum_{k=0}^{n-1} \delta\big(A(\tau^k(g)), B(\tau^k(g))\big),$$

and therefore, the lemma follows by Birkhoff ergodic Theorem. ■

3.3.1 Good approximation by continuous functions

The previous two lemmas indicate that we need a kind of L^1 approximation. More precisely, given $A \in L^1(G, M)$ and $\varepsilon > 0$, we are looking for a continuous function $A_\varepsilon : G \to M$ such that

$$\int_G \delta(A(g), A_\varepsilon(g))\, dm(g) < \varepsilon.$$

In some cases there exists an underlyng finite dimensional vector space. This is the case, for instance, when M is the set of (strictly) positive matrices, or more generally, when M is a Riemannian manifold with non-positive curvature. In these cases, the function A_ε can be constructed by using mollifiers. This idea was used by Karcher in [17]. In the general case, we can use a similar idea.

Given $\eta > 0$, let U_η be a neighborhood of the identity of G so that $m(U_\eta) < \eta$, whose diameter is also less than η. Fix any $y \in M$, and define

$$A_\eta(g_0) = \operatorname*{argmin}_{z\in M} \int_{U_\eta} [\delta^2(z, A(g + g_0)) - \delta^2(y, A(g + g_0))]\ dm(g). \tag{3.8}$$

Equivalently, $A_\eta(g_0)$ is the barycenter of the pushforward by A of the Haar measure restricted to $g_0 + U_\eta$. This definition follows the idea of mollifiers, replacing the arithmetic mean by the average induced by barycenters. We will prove that, as in the case of usual mollifiers, these continuous functions provide good approximation in L^1 (Theorem 3.13 below). With this aim, firstly we will prove the following lemma.

Lemma 3.10. *Let $A \in L^p(G, M)$ where $1 \leq p < \infty$. Define the function $\varphi : G \to [0, +\infty)$ by*

$$\varphi(h) = \int_G \delta^p\big(A(g), A(g + h)\big)dm(g),$$

is a continuous function.

Proof. Fix $z_0 \in M$, and define the measure on the Borel sets of G

$$\nu(B) := \int_B \delta^p(A(g), z_0) dm(g).$$

By definition, ν is absolutely continuous with respect to the Haar measure m. In consequence, given $\varepsilon > 0$, there exists $\eta > 0$, such that, whenever a Borel set B satisfies

$$\int_B dm(g) < \eta,$$

it holds that

$$\nu(B) = \int_B \delta^p(A(g), z_0) dm(g) < \frac{\varepsilon}{2^{p+2}}, \tag{3.9}$$

By Lusin Theorem [13, Thm 7.5.2], there is a compact set $C_\eta \subset G$ such that $m(C_\eta) \geq 1 - \eta/2$ and the restriction of A to C_η is (uniformly) continuous.

Since m is a Haar measure, it is enough to prove the continuity of φ at the identity. With this aim, take a neighborhood of the identity U so that whenever $g_1, g_2 \in C_\eta$ satisfy that $g_1 - g_2 \in U$, it holds that

$$\delta^p(A(g_1), A(g_2)) \leq \frac{\varepsilon}{2}.$$

Given $h \in U$, define $\Omega := C_\eta \cap (C_\eta + h)$, and $\Omega^c := G \setminus \Omega$. Then

$$\begin{aligned}
\int_G \delta^p(A(g), A(g+h)) dm(g) &= \int_\Omega + \int_{\Omega^c} \delta^p(A(g), A(g+h)) dm(g) \\
&\leq \frac{\varepsilon}{2} + \int_{\Omega^c} \delta^p(A(g), A(g+h)) dm(g) \\
&\leq \frac{\varepsilon}{2} + \int_{\Omega^c} [\delta(A(g), z_0) + \delta(A(g+h), z_0)]^p \, dm(\omega) \\
&= \frac{\varepsilon}{2} + 2^{p+1} \int_{\Omega^c} \delta^p(A(g), z_0) dm(g),
\end{aligned}$$

where in the last identity we have used that m is shift invariant. Since $|\Omega^c| < \delta$ we obtain that

$$\int_G \delta^p(A(g), A(g+h)) dm(g) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

■

Corollary 3.11. *For every $\eta > 0$, the functions A_η are continuous.*

Proof. Indeed, by Lemma 3.8

$$\begin{aligned}
\delta(A_\eta(h_1), A_\eta(h_2)) &\leq \frac{1}{m(U_\eta)} \int_{U_\eta} \delta(A(g+h_1), A(g+h_2)) \, dm(g) \\
&\leq \frac{1}{m(U_\eta)} \int_G \delta(A(g+h_1), A(g+h_2)) \, dm(g) \\
&\leq \frac{1}{m(U_\eta)} \int_G \delta[\, A(g), A(g+(h_2-h_1)) \,] \, dm(g).
\end{aligned}$$

So, the continuity of A_η is a consequence of the continuity of φ at the identity. ■

The map $A \mapsto A_\varepsilon$ has the following useful continuity property.

Lemma 3.12. *Let $A, B \in L^1(G, M)$, and $\eta > 0$. For every $\varepsilon > 0$, there exists $\rho > 0$ such that if*

$$\int_G \delta(A(g), B(g)) dm(g) \leq \rho,$$

then the corresponding continuous functions A_η and B_η satisfy that

$$\max_{g \in G} \delta(A_\eta(g), B_\eta(g)) \leq \varepsilon.$$

Proof. Indeed, given $\varepsilon > 0$, take $\rho = m(U_\eta)\varepsilon$. Then, by Lemma 3.8

$$\begin{aligned}\delta(A_\eta(g), B_\eta(g)) &\leq \frac{1}{|U_\eta|} \int_{U_\eta} \delta(A(g+h), B(g+h)) dm(h) \\ &\leq \frac{1}{|U_\eta|} \int_G \delta(A(h), B(h)) dm(h) \leq \varepsilon,\end{aligned}$$

for all $g \in G$. ■

We arrive to the main result of this section.

Proposition 3.13. *Given a function $A \in L^1(G, M)$, if A_η are the continuous functions defined by* (3.8) *then*

$$\lim_{\eta \to 0^+} \int_G \delta(A(g), A_\eta(g))\, dm(g) = 0.$$

Proof. Firstly, assume that $A \in L^2(G, M)$. In this case, by the variance inequality, it holds that

$$\delta^2(A(g), A_\eta(g)) \leq \frac{1}{|U_\eta|} \int_{U_\eta} \delta^2(A(g), A(g+h)) dm(h).$$

So, using Fubini's theorem we get

$$\begin{aligned}\int_G \delta^2(A(g), A_\eta(g))\, dm(g) &\leq \frac{1}{|U_\eta|} \int_{U_\eta} \int_G \delta^2(A(g), A(g+h))\, dm(g)\, dm(h) \\ &= \frac{1}{|U_\eta|} \int_{U_\eta} \varphi(h)\ dm(h).\end{aligned}$$

By Lemma 3.10, the function φ is continuous. In consequence, if e denotes the identity of G

$$\lim_{\eta \to 0^+} \frac{1}{|U_\eta|} \int_{U_\eta} \varphi(h)\ dm(h) = \varphi(e) = 0.$$

This proves the result for functions in $L^2(G, M)$ since by Jensen's inequality

$$\int_G \delta(A(g), A_\eta(g))\, dm(g) \leq \left(\int_G \delta^2(A(g), A_\eta(g))\, dm(g) \right)^{1/2}.$$

Now, consider a general $A \in L^1(G, M)$. Fix $z_0 \in M$, and for each natural number N define the truncations

$$A^{(N)}(g) := \begin{cases} A(g) & \text{if } \delta(A(g), z_0) < N \\ z_0 & \text{if } \delta(A(g), z_0) \geq N \end{cases}.$$

For each N we have that $A^{(N)} \in L^1(G, M) \cap L^\infty(G, M)$, and therefore it also belongs to $L^2(G, M)$. On the other hand, since the function defined on G by $g \mapsto \delta(A(g), z_0)$ is integrable, it holds that

$$\int_G \delta(A(g), A^{(N)}(g))\, dm(g) = \int_{\{g:\, \delta(A(g), z_0) \geq N\}} \delta(A(g), z_0)\, dm(g) \xrightarrow[N \to \infty]{} 0 \tag{3.10}$$

So, if A_η and $A_\eta^{(N)}$ are the continuous functions associated to A and $A^{(N)}$ respectively, then

$$\begin{aligned} \int_G \delta(A(g), A_\eta(g))\, dm(g) \leq & \int_G \delta(A(g), A^{(N)}(g))\, dm(g) \\ & + \int_G \delta(A^{(N)}(g), A_\eta^{(N)}(g))\, dm(g) \\ & + \int_G \delta(A_\eta^{(N)}(g), A_\eta(g))\, dm(g). \end{aligned}$$

Note that each term of the right hand side tends to zero: the first one by (3.10), the second one by the L^2 case done in the first part, and the last one by Lemma 3.12. ■

3.3.2 Proof of Theorem 3.7

Let $\varepsilon > 0$. For each $k \in \mathbb{N}$ let A_k be a continuous function such that

$$\int_G \delta(A(g), A_k(g))\, dm(g) \leq \frac{1}{k}.$$

By Lemma 3.9, we can take a set of measure zero $N \subseteq G$ such that if we take $g \in G \setminus N$ and $k \in \mathbb{N}$, there exists n_0, which may depend on g and k, so that

$$\delta\big(S_n(a^\tau(g)), S_n(a^\tau_{(k)}(g))\big) \leq \frac{\varepsilon}{4} + \int_G \delta(A(g), A_k(g)) dm(g),$$

provided $n \geq n_0$. In this expression, $a^\tau_{(k)}$ is the sequence defined in terms of A_k and τ as in (3.1). Fix $g \in G \setminus N$. Taking k so that $1/k < \varepsilon/4$, we get that

$$\delta\big(S_n(a^\tau(g)), S_n(a^\tau_{(k)}(g))\big) \leq \frac{\varepsilon}{2},$$

for every $n \geq n_0$. By Lemma 3.8, it also holds that $\delta(\beta_A, \beta_{A_k}) \leq \frac{\varepsilon}{4}$ where

$$\beta_A = \operatorname*{argmin}_{z \in M} \int_G [\delta^2(A(g), z) - \delta^2(A(g), y)]\, dm(g),$$
$$\beta_{A_k} = \operatorname*{argmin}_{z \in M} \int_G [\delta^2(A_k(g), z) - \delta^2(A_k(g), y)]\, dm(g).$$

Finally, by Theorem 3.1, there exists $n_1 \geq 1$ such that for every $n \geq n_1$

$$\delta(S_n(a^\tau_{(k)}(g), \beta_{A_k}) \leq \frac{\varepsilon}{4}.$$

Combining all these inequalities we obtain that

$$\begin{aligned}\delta\big(S_n(a^\tau(g)), \beta_A\big) &\leq \delta\big(S_n(a^\tau(g)), S_n(a^\tau_{(k)}(g))\big) \\ &\quad + \delta\big(S_n(a^\tau_{(k)}(g)), \beta_{A_k}\big) + \delta(\beta_{A_k}, \beta_A) \leq \varepsilon,\end{aligned}$$

which concludes the proof.

Acknowledgements:

This work was supported by Consejo Nacional de Investigaciones Científicas y Técnicas-Argentina (PIP-152), Agencia Nacional de Promoción de Ciencia y Tecnología-Argentina (PICT 2015-1505), Universidad Nacional de La Plata-Argentina (UNLP-11X585) and Ministerio de Economía y Competitividad-España (MTM2016-75196-P).

References

[1] Alexandrov, A. D.: A theorem on triangles in a metric space and some applications, Trudy Math. Inst. Steklov 38 (1951), 5-23.

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

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

[4] Ballmann W.: Lectures on spaces of nonpositive curvature. DMV Seminar Band 25, Birkhäuser Verlag, Basel.

[5] Ballmann W., Gromov M., Schroeder V.: Manifolds of nonpositive curvature. Progress in Mathematics 61. Birkhäuser Boston Inc., Boston, MA

[6] Berger M.: A Panoramic View of Riemannian Geometry, Springer, 2003.

[7] Bhatia R.: Positive Definite Matrices, Princeton Series in Applied Mathematics, Princeton University Press, 2007.

[8] Bhatia R., Karandikar R.: Monotonicity of the matrix geometric mean. Math. Ann. 353(4)(2012) 1453-1467.

[9] Bini D., Iannazzo B.: Computing the Karcher mean of symmetric positive definite matrices, Linear Algebra Appl. 438 (2013) 1700-1710.

[10] Bochi J., Navas A.: A geometric path from zero Lyapunov exponents to rotation cocycles, Ergodic Theory Dynam. Systems 35 (2015) 374-402.

[11] Bridson M.R. , Haefliger A.: Metric spaces of non-positive curvature, Grundlehren der Mathematischen Wissenschaften, 319. Springer-Verlag, Berlin.

[12] Barbaresco F.: Interactions between symmetric cone and information geometries: Bruhat-Tits and Siegel spaces models for higher resolution autoregressive Doppler imagery, Emerging Trends in Visual Computing, Lecture Notes in Computer Science 5416 (2009) 124-163.

[13] Dudley R. M.: Real analysis and probability, Cambridge Studies in Advanced Mathematics, 74. Cambridge University Press, Cambridge.

[14] Es-Sahib A., Heinich H.: Barycentre canonique pour un espace metrique ?a courbure negative, in: Seminaire de Probabilites, in: Lecture Notes in Math., vol. 1709, Springer, Berlin, 1999.

[15] Holbrook J.: No dice: a determinic approach to the Cartan centroid, J. Ramanujan Math. Soc. 27 (2012) 509-521.

[16] Jost J.: Nonpositive Curvature: Geometric and Analytic Aspects, Lectures in Mathematics ETH Zurich, Birkhauser, 1997.

[17] Karcher H.: Riemannian center of mass and mollifier smoothing, Comm. Pure Appl. Math. 30 (1977) 509-541.

[18] Lawson J., Lim Y.: A general framework for extending means to higher orders, Colloq. Math. 113 (2008) 191-221.

[19] Lawson J., Lim Y.: Monotonic properties of the least squares mean, Math. Ann. 351 (2011) 267-279.

[20] Lawson J., Lim Y.: Weighted means and Karcher equations of positive operators, Proc. Natl. Acad. Sci. USA 110 (2013) 15626-15632.

[21] Lawson J., Lim Y.: Karcher means and Karcher equations of positive definite operators, Trans. Amer. Math. Soc. Ser. B1 (2014) 1-22.

[22] Lawson J., Lim Y.: Contractive barycentric maps, J. Operator Theory 77 (2017) 87-107.

[23] Lee J., Naor A.: Extending Lipschitz functions via random metric partitions, Invent. Math. 160 (2005) 59-95.

[24] Lim Y.: Riemannian distances between Geometric means. SIAM J. Matrix Anal. Appl. 34 (2013), 932-945.

[25] Lim Y., Pálfia M.: Matrix power mean and the Karcher mean, J. Funct. Anal. 262 (2012) 1498-1514.

[26] Lim Y., Pálfia M.: Weighted deterministic walks and no dice approach for the least squares mean on Hadamard spaces, Bull. Lond. Math. Soc. 46 (2014) 561-570.

[27] Lim Y., Pálfia M.: Approximations to the Karcher mean on Hadamard spaces via geometric power means, Forum Math. 27 (2015) 2609-2635.

[28] Mendel M., Naor A.: Spectral calculus and Lipschitz extension for barycentric metric spaces, Anal. Geom. Metr. Spaces 1 (2013) 163-199.

[29] Moakher M., Zerai M.: The Riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data, J. Math. Imaging Vision 40 (2011) 171-187.

[30] Navas A.: 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.

[31] Ohta S.: Extending Lipschitz and Hölder maps between metric spaces, Positivity 13 (2009) 407-425.

[32] Pálfia M.: Means in metric spaces and the center of mass, J. Math. Anal. Appl. 381 (2011) 383-391.

[33] Pass B.: Uniqueness and Monge solutions in the multimarginal optimal transportation problem, SIAM J. Math. Anal. 43 (2011) 2758-2775.

[34] Pass B.: Optimal transportation with infinitely many marginals, J. Funct. Anal. 264 (2013) 947-963.

[35] Reshetnyak Y. G.: (1968): Inextensible mappings in a space of curvature no greater than K. Sib. Math. Jour. 9, 683-689

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