

# Ergodic theorem in CAT(0) spaces in terms of inductive means

JORGE ANTEZANA , EDUARDO GHIGLIONI and DEMETRIO STOJANOFF

*Instituto Argentino de Matemática ‘Alberto P. Calderón’ (IAM-CONICET),  
CABA, Argentina*

*and*

*Departamento de Matemática, Facultad de Ciencias Exactas,  
Universidad Nacional de La Plata, Buenos Aires, Argentina  
(e-mail: antezana@mate.unlp.edu.ar, eghiglioni@mate.unlp.edu.ar,  
demetrio@mate.unlp.edu.ar)*

*(Received 3 February 2021 and accepted in revised form 14 January 2022)*

*Abstract.* Let  $(G, +)$  be a compact, abelian, and metrizable topological group. In this group we take  $g \in G$  such that the corresponding automorphism  $\tau_g$  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 on 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 does not require the existence of an underlying vector space. For this reason, it can be used in any Hadamard space, in contrast to those results that need to use the tangent space or some chart to define the mollifier.

**Key words:** ergodic theorem, inductive means, barycenter, Hadamard space, non-positively curved space

**2020 Mathematics Subject Classification:** 37A30, 51F99 (Primary)

## *Contents*

1	Introduction	1434
1.1	Framework and related results	1434
1.2	Main results	1436
1.3	Organization of the paper	1437
2	Preliminaries	1438

2.1	CAT(0) spaces	1438
2.2	Barycenters in CAT(0) spaces	1438
2.3	The inductive means	1440
3	Proof of the main results	1442
3.1	Continuous case	1442
3.2	Preparation for the $L^1$ case	1446
3.3	Good approximation by continuous functions	1447
3.4	The $L^1$ case and almost everywhere convergence	1451
3.5	The $L^p$ results	1451
	Acknowledgements	1453
	References	1453

## 1. Introduction

One of the classical results in ergodic theory is the Birkhoff ergodic theorem.

**THEOREM 1.1. (Birkhoff)** *Let  $(X, \mathfrak{X}, m)$  be a probability space,  $\tau : X \rightarrow X$  an ergodic map, and  $f \in L^1(X, \mathbb{C})$ . Then, for  $m$ -almost every  $x \in X$ ,*

$$\frac{1}{n} \sum_{k=0}^{n-1} f(\tau^k(x)) \xrightarrow{n \rightarrow \infty} \int_X f(t) dm(t). \quad (1.1)$$

The aim of this work is to study this result in the context of CAT(0) spaces. The traditional extensions of the Birkhoff theorem to this setting replace the arithmetic means by the so-called barycenters. In particular, the barycenters are used to average the function along the ergodic orbit. Our motivation to consider variants of these traditional extensions is that usually the barycenters cannot be computed explicitly. Moreover, in many important cases, the usual ways to approximate the barycenters using convex optimization methods are not useful for applications (see Example 2.3). A similar situation can be found in the extensions of other well-known theorems to CAT(0) spaces. This is the case for the law of large numbers. In [30] Sturm introduced the so-called inductive means, which can be computed easily in those spaces where the geodesics are known. Using these means to average the independent copies of the random variable, he obtained in [30] a new version of the law of large numbers in CAT(0) spaces. Motivated by this result, we studied a version of the classical Birkhoff ergodic theorem using the inductive means defined by Sturm to average the functions along the ergodic orbit. As a consequence, we get new ways to approximate a barycenter of integrable function with values in a CAT(0) space.

**1.1. Framework and related results.** Recall that a CAT(0) space, also called a Hadamard space, is a complete metric space  $(M, \delta)$  whose metric satisfies the following semiparallelogram law: given  $x, y \in M$ , there exists  $m \in M$  satisfying

$$\delta^2(m, z) \leq \frac{1}{2}\delta^2(x, z) + \frac{1}{2}\delta^2(y, z) - \frac{1}{4}\delta^2(x, y), \quad (1.2)$$

for all  $z \in M$ . The point  $m$  is unique, and it is called the midpoint between  $x$  and  $y$ , because, taking  $z = x$  and  $z = y$  the following identities hold:

$$\delta(x, m) = \delta(m, y) = \frac{1}{2}\delta(x, y).$$

The existence and uniqueness of midpoints give rise to a unique (continuous) geodesic which we will denote by  $\gamma : [0, 1] \rightarrow M$  (see §2 for more details). We will denote this curve by  $x\#_t y$  instead of  $\gamma(t)$ . Typical examples of CAT(0) spaces are the Riemannian manifolds with non-positive sectional curvature, and certain types of graphs such as trees or spiders. The systematic study of these spaces started with work by Alexandrov [1] and Reshetnyak [28], and the subject was strongly influenced by the works of Gromov [13, 14]. Nowadays there exists a huge bibliography on the subject. The interested reader is referred to the monographs [4, 5, 10, 15].

The convexity properties of the metric allow us to define a notion of barycenter in CAT(0) spaces. 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 [9, 12, 18, 24, 30]), stochastic generalization of Lipschitz retractions and extension problems of Lipschitz and Hölder maps (see [19, 22, 25]), optimal transport theory on Riemannian manifolds (see [26, 27]), and so on.

Roughly speaking, the barycenter constitutes a way to average points in  $M$ , taking into account the metric properties of the space. More precisely, the barycenter is defined for some measures with separable support (see §2 for the formal definition). Given  $n$  points in the space  $M$ , let  $\beta(x_1, \dots, x_n)$  denote the barycenter of the points  $x_1, \dots, x_n$  (more precisely, the barycenter of the point measure  $\mu = \delta_{x_1} + \dots + \delta_{x_n}$ ). On the other hand, if  $(X, \mathfrak{X}, \mu)$  is a measure space and  $f : X \rightarrow M$  is a measurable function such that for some  $y \in M$  (and therefore for any  $y \in M$ )

$$\int_X \delta(f(x), y) d\mu(x) < \infty, \quad (1.3)$$

then  $\beta_f$  denotes the barycenter of the pushforward measure  $f_*(\mu)$ .

Any Hilbert space  $\mathcal{H}$ , and in particular  $\mathbb{C}$ , is a CAT(0) space with the metric induced by the norm, and the barycenter in  $\mathcal{H}$  is precisely the arithmetic mean. Therefore, the natural extension of the Birkhoff theorem for a function  $f$  satisfying (1.3) is obtained by replacing the arithmetic means by the barycenters

$$\beta(f(x), \dots, f(\tau^{n-1}(x))) \xrightarrow{n \rightarrow \infty} \beta_f. \quad (1.4)$$

This result was proved by Austin in [2] for functions satisfying the integrability condition

$$\int_X \delta(f(x), y)^2 d\mu(x) < \infty, \quad (1.5)$$

instead of (1.3). Later on, in [24] Navas proved it for functions satisfying (1.3). In both cases, the authors considered not only  $\mathbb{Z}$ -actions but also much more general actions given by amenable groups. Moreover, Navas's theorem holds not only in CAT(0) spaces but also in metric spaces of non-positive curvature in the sense of Busemann.

However, the barycenters  $\beta(f(x), \dots, f(\tau^{n-1}(x)))$  in (1.4) may be very difficult to compute for  $n \geq 4$ . Hence, it is natural to look for an alternative way to average the points  $f(x), \dots, f(\tau^{n-1}(x))$ . This leads to the definition of inductive means. To motivate their definition, note that, given a sequence  $\{a_n\}_{n \in \mathbb{N}}$  of complex numbers,

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

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

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

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 Hadamard space. This is the idea that leads to the definition of the inductive means. Given a sequence  $\{a_n\}_{n \in \mathbb{N}}$  whose elements belong to a CAT(0) space  $M$ , the inductive means are defined as follows:

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

These means were introduced by Sturm in [30], where he proved the following version of the law of the large numbers.

**THEOREM 1.2.** (Sturm) *Let  $(X, \mathfrak{X}, \mu)$  be a probability space, and let  $A = \{A_j\}_{j \in \mathbb{N}}$  be a sequences of independent and identically distributed bounded random variables satisfying (1.5). Then, almost surely,*

$$S_n(A) \xrightarrow[n \rightarrow \infty]{} \beta_{A_1}.$$

This result suggests the possibility of finding extensions of the Birkhoff ergodic theorem using inductive means to average the values of the function on the ergodic orbit. Note that, if we want to use inductive means, then we are compelled to consider only  $\mathbb{Z}$ -actions.

**1.2. Main results.** Let  $(G, +)$  be a compact and metrizable topological group. In this group we fix a Haar measure  $m$ , a shift-invariant metric  $d_G$ , and we take an ergodic automorphism  $\tau(h) = h + g$  for some  $g \in G$ . Note that the existence of such an ergodic automorphism implies that the group must be abelian (see [32, Theorem 1.9])

On the other hand, let  $(M, \delta)$  be a fixed CAT(0) space. Given a function  $A : G \rightarrow M$ , we define  $a^\tau : G \rightarrow M^{\mathbb{N}}$  by

$$a^\tau(x) := \{a_j^\tau(x)\}_{j \in \mathbb{N}} \quad \text{where } a_j^\tau(x) = A(\tau^j(x)). \quad (1.6)$$

Our first main theorem is the following continuous version of the ergodic theorem.

**THEOREM 1.3.** *Let  $M$  be a Hadamard space and  $A : G \rightarrow M$  a continuous function. Then*

$$\lim_{n \rightarrow \infty} S_n(a^\tau(g)) = \beta_A, \quad (1.7)$$

uniformly in  $g \in G$ .

To extend this result to  $L^1(G, M)$  functions, we need to find ‘good  $L^1$ -approximations by continuous functions’. These approximations are obtained in §3.3, where we study mollifiers in general Hadamard spaces. The results on mollifiers obtained in this subsection are of interest in their own right, since they generalize some results proved by Karcher in [16] for Riemannian manifolds. Using this  $L^1$ -approximation we get the following  $L^1$  version of the ergodic theorem.

**THEOREM 1.4.** *Given  $A \in L^1(G, M)$ , for almost every  $g \in G$ ,*

$$\lim_{n \rightarrow \infty} S_n(a^\tau(g)) = \beta_A. \quad (1.8)$$

From this result, using standard techniques we get the following  $L^p$  versions.

**THEOREM 1.5.** *Let  $1 \leq p < \infty$  and  $A \in L^p(G, M)$ . Then*

$$\lim_{n \rightarrow \infty} \int_G \delta^p(S_n(a^\tau(g)), \beta_A) dm(g) = 0. \quad (1.9)$$

Recall that a topological dynamical system  $(\Omega, \tau)$  is called a *Kronecker system* if it is isomorphic to a group dynamical system  $(G, \tau)$  like the one described above. Also recall that any equicontinuous dynamical system becomes an isometric system by changing the metric, and any minimal isometric dynamical system is a Kronecker system (see, for example, [31, §2.6]). Therefore, using standard arguments, all the main results of this work can be extended to equicontinuous systems. In order to go further and consider more general dynamical systems we think that a different approach is required.

**1.3. Organization of the paper.** The paper is organized as follows. Section 2 is devoted to gathering together some preliminaries on Hadamard spaces, barycenters and inductive means that will be used throughout the paper. Section 3 is devoted to the proofs of our main results. In this section we also prove those results related to approximation by continuous functions in general Hadamard spaces.

2. Preliminaries

In this section we recall some results on CAT(0) spaces, barycenters, as well as proving some results on inductive means that we will need later. The interested reader is referred to the monographs [3–5, 10, 15] for more information.

2.1. *CAT(0) spaces.* Recall that a CAT(0) space, also known as Hadamard space, is a complete metric space  $(M, \delta)$  that satisfies the following semiparallelogram law: given  $x, y \in M$ , 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 unique, and it is called midpoint between  $x$  and  $y$ , since  $\delta(x, m) = \delta(m, y) = \frac{1}{2}\delta(x, y)$ . Recall that, given a continuous curve  $\gamma : [a, b] \rightarrow M$ , its length is computed as

$$\inf \sum_{n=1}^N \delta(\gamma(t_{n+1}), \gamma(t_n))$$

where the infimum is taken over all the partitions  $\{t_1, \dots, t_N\}$  of the interval  $[a, b]$ . The existence and uniqueness of midpoints give rise to a unique (continuous) geodesic  $\gamma_{x,y} : [0, 1] \rightarrow M$  connecting any given two points  $x$  and  $y$ . Indeed, we first define  $\gamma_{x,y}(1/2)$  to be the midpoint of  $x$  and  $y$ . 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  $t \in [0, 1]$ . It can be proved that this curve is the shortest path connecting  $x$  and  $y$ . As we mentioned in the introduction, we will use the notation  $x\#_t y$  instead of  $\gamma_{x,y}(t)$ . It is not difficult to see that the points of this geodesic satisfy the following generalized semiparallelogram inequality:

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

As consequence of this inequality the next result on the convexity of the metric is obtained (see, for example, [30, Corollary 2.5]).

PROPOSITION 2.1. *Given four points  $a, a', b, b' \in M$ , let*

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

*Then  $f$  is convex on  $[0, 1]$ ; that is,*

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

Another very important result in CAT(0) spaces is the so-called Reshetnyak quadruple comparison theorem (see, for example, [30, Proposition 2.4]).

THEOREM 2.2. *Let  $(M, \delta)$  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 in CAT(0) spaces.* Let  $\mathcal{B}(M)$  be the  $\sigma$ -algebra of Borel sets (that is, the smallest  $\sigma$ -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  $\mathcal{P}^\theta(M)$  be the set of those measures  $\mu \in \mathcal{P}(M)$  such that

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

for some (and hence for 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. Finally, given a measure space  $(X, \mathfrak{X}, \mu)$  and a measurable function  $f : X \rightarrow M$ , we say that  $f$  belongs to  $L^p(X, M)$  if the pushforward of  $\mu$  by  $f$  belongs to  $\mathcal{P}^p(M)$  ( $1 \leq p \leq \infty$ ).

If  $\mu \in \mathcal{P}^2(M)$ , then the usual Cartan definition of barycenter  $\beta_\mu$  can be extrapolated to this setting:

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

The existence of a unique minimizer is guaranteed by the convexity properties of the metric. This definition can be extended to measures in  $\mathcal{P}^1(M)$ . Following the ideas of Sturm in [30], given any point  $y \in M$ , the barycenter of a measure  $\mu \in \mathcal{P}^1(M)$  is defined as the unique minimizer of the functional

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

Although the functional depends on the point  $y$ , it is easy to see that the minimizer is independent of it. Hence, the barycenter is well defined. Moreover, if  $\mu \in \mathcal{P}^2(M)$  this definition coincides with Cartan's definition. Note that in this case the quantity

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

can be thought as a variance. Moreover, the barycenters in this case also satisfy the following inequality known as the *variance inequality*:

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

Therefore, sometimes the barycenter is considered as a nonlinear version of the expectations. For instance, this idea was used by Sturm to extend different result from probability theory to this nonlinear setting (see [30, 29] and the references therein).

*Special notation.* As we mentioned in the introduction, we will use a special notation in the following two cases. On the one hand, let  $(X, \mathfrak{X}, \mu)$  be a measure space and let  $f : X \rightarrow M$  be a measurable function in  $L^1(X, M)$ . By means of  $\beta_f$  we will denote the barycenter of the pushforward measure  $f_*(\mu)$ . On the other hand, given  $n$  points  $x_1, \dots, x_n \in M$ , by means of  $\beta(x_1, \dots, x_n)$  we will denote the barycenter of the point measure  $\mu = \delta_{x_1} + \dots + \delta_{x_n}$ .

The main issue dealing with barycenters is that they are difficult to compute. The computation of the barycenter of three or more points may be difficult. Although there exists a very rich convex theory in Hadamard space (see, for instance, [3]), sometimes

the approximation of the barycenter using convex optimization is not satisfactory. A good example of this situation is as follows.

*Example 2.3.* (Positive matrices) Recall that the set of positive invertible matrices  $\mathcal{M}_n(\mathbb{C})^+$  is an open cone in the real vector space of self-adjoint matrices  $\mathcal{H}(n)$ . In particular, it is a differentiable manifold and the tangent spaces can be identified 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  $\alpha : [a, b] \rightarrow \mathcal{M}_n(\mathbb{C})^+$  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, and in particular a CAT(0) space. The geodesic connecting two positive matrices  $A$  and  $B$  has the following simple expression:

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

So, the barycenter of the measure  $\mu = \frac{1}{2}(\delta_A + \delta_B)$  is given by

$$A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2} A^{1/2}.$$

However, if we add an atom to  $\mu$ , there is no longer a closed formula for the barycenter (also called the geometric mean in this setting). As a consequence, simple questions such as the monotonicity of the barycenter with respect to the usual order of matrices become difficult. Using convex optimization, it is possible to construct a sequence that approximates the barycenter of a measure. However, that sequence does not contain enough information in order to prove that the barycenter is monotone. This issue, for instance, motivated intensive research with the aim of finding good ways to approximate the barycenters of more than two matrices [7, 17, 20]. The barycenters in this setting have attracted much attention in recent years because of their interesting applications in signal processing (see [6] and the references therein), and gradient or Newton-like optimization methods (see [8, 23]).

2.3. *The inductive means.* Recall that, given  $a \in M^{\mathbb{N}}$ , the inductive means are define as:

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

As a consequence of (2.3), we directly get the following result.

COROLLARY 2.4. *Given  $a, b \in M^{\mathbb{N}}$ , then*

$$\delta(S_n(a), S_n(b)) \leq \frac{1}{n} \sum_{i=1}^n \delta(a_i, b_i). \tag{2.6}$$

The next lemma follows from (2.2), and it is a special case of a weighted inequality considered by Lim and Pálfi in [21].



LEMMA 2.5. Given  $a \in M^{\mathbb{N}}$  and  $z \in M$ , for every  $k, m \in \mathbb{N}$ ,

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

*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  $n = k$  to  $n = k + m - 1$ , 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 than or equal to

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

Finally, using that  $(k+j)/(k+j+1) \geq k/(k+m)$  for every  $j \in \{0, \dots, m-1\}$ , this sum is bounded from 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 2.6. Given  $a \in M^{\mathbb{N}}$  such that  $\Delta(a) < \infty$ , we have for all  $k, m \in \mathbb{N}$  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  $\tilde{R}_{m,k} = ((m^2/(k+1)^2) + 2(m/(k+1)))\Delta^2(a)$ .

*Proof.* Note that by (2.6) and for all  $k$ ,

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

Hence

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

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. Proof of the main results

3.1. *Continuous case.* In this section we will prove Theorem 1.3. Recall that, given a function  $A : G \rightarrow M$ , we define  $a^\tau : G \rightarrow M^{\mathbb{N}}$  by

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

The proof is rather long and technical, so we split it into several lemmas and a technical result, which will be combined at the end of the section to provide the proof of Theorem 1.3.

LEMMA 3.1. *Let  $A : G \rightarrow M$  be a continuous function, and let  $K$  be any compact subset of  $M$ . For each  $n \in \mathbb{N}$ , define  $F_n : G \times K \rightarrow \mathbb{R}$  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  $(K, \delta)$  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  $d_G(g_1, g_2) < \delta$  then

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

for every  $x \in K$ . Since  $\tau$  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  $\Delta$  be the diameter of the set  $(\text{Image}(A) \times K)$  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  $\delta(x_1, x_2) < \varepsilon/4\Delta$ .

Then

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

Now, as a consequence of the Arzelà–Ascoli and Birkhoff theorems, we get the following proposition.

**PROPOSITION 3.2.** *Let  $A : G \rightarrow M$  be a continuous function, and  $K$  a compact subset of  $M$ . Then*

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

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

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

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

and  $\beta_A$  is the point where this minimum is attained, that is,  $\beta_A$  is the barycenter of the pushforward by  $A$  of the Haar measure in  $G$ . Then we obtain the following upper estimate.

**LEMMA 3.3.** *For every  $\varepsilon > 0$ , there exists  $m_0 \in \mathbb{N}$  such that, for all  $m \geq m_0$  and for all  $k \in \mathbb{N}$ ,*

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

*Proof.* For every  $\varepsilon > 0$ , there 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 independent of  $k$  by Proposition 3.2. Now, by Lemma 2.5,

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

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

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

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

where, as we have defined before Lemma 2.6,  $\Delta(a^\tau(g))$  denotes the diameter of the image of the sequence  $a^\tau(g)$ .

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  $k \in \mathbb{N}$ ,*

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

where  $R_{m,k} = (m^2/(k+1)^2) + 2(m/(k+1))C_a^2$ .

*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 (2.5) and Proposition 3.2,

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

Finally, by Lemma 2.6,

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

where  $R_{m,k} = ((m^2/(k+1)^2) + 2(m/(k+1)))C_a$ . □

LEMMA 3.5. 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 3.3 and 3.4, there exists  $m_0 \geq 1$  such that for all  $k \in \mathbb{N}$ ,

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

and

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

Therefore, combining these two inequalities, we obtain

$$\begin{aligned} \delta^2(S_{k+m_0}(a^\tau(g)), \beta_A) &\leq \frac{k}{k+m_0} \delta^2(S_k(a^\tau(g)), \beta_A) + \frac{m_0}{k+m_0}(\alpha + \varepsilon) \\ &\quad - \frac{km_0}{(k+m_0)^2} (\delta^2(S_k(a^\tau(g)), \beta_A) - \varepsilon + \alpha - R_{m_0,k}). \end{aligned}$$

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

$$\begin{aligned} \delta^2(S_{(\ell+1)m_0}(a^\tau(g)), \beta_A) &\leq \frac{\ell}{\ell+1} \delta^2(S_{\ell m_0}(a^\tau(g)), \beta_A) + \frac{1}{\ell+1}(\alpha + \varepsilon) \\ &\quad - \frac{\ell}{(\ell+1)^2} (\delta^2(S_{\ell m_0}(a^\tau(g)), \beta_A) - \varepsilon + \alpha - R_{m_0, \ell m_0}) \\ &\leq \frac{\ell^2 \delta^2(S_{\ell m_0}(a^\tau(g)), \beta_A) + (2\ell+1)\varepsilon + \alpha + 3C_a^2}{(\ell+1)^2}. \quad (3.1) \end{aligned}$$

Using this recursive inequality, the result follows by induction on  $\ell$ . 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  $\ell \geq 1$ , that is,

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

then, combining this inequality with (3.1), we have that

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

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

*Proof of Theorem 1.3.* Given  $\varepsilon > 0$ , by Lemma 3.5, there exists  $m_0 \in \mathbb{N}$  such that

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

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

$$\delta^2(S_{\ell m_0}(a^\tau(g)), \beta_A) \leq \frac{\varepsilon^2}{4}. \tag{3.2}$$

Let  $n = \ell m_0 + d$  such that  $\ell \geq \ell_0$  and  $d \in \{1, \dots, m_0 - 1\}$ . Since  $x\#_t x = x$  for all  $x \in M$ , using Corollary 2.4 with the sequences

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

and

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

we get

$$\delta(S_{\ell m_0}(a^\tau(g)), S_{\ell m_0+d}(a^\tau(g))) \leq \frac{1}{\ell m_0 + d} \sum_{j=1}^d \delta(S_{\ell m_0}(a^\tau(g)), a_{\ell m_0+j}^\tau(g)).$$

Now, taking into account that  $\delta(S_{\ell m_0}(a^\tau(g)), a_{\ell m_0+j}^\tau(g)) \leq C_a$  for every  $j \in \{1, \dots, m_0 - 1\}$ , we obtain that

$$\delta(S_{\ell m_0}(a^\tau(g)), S_{\ell m_0+d}(a^\tau(g))) \leq \frac{d}{\ell m_0 + d} C_a \leq \frac{1}{\ell} C_a \xrightarrow{k \rightarrow \infty} 0.$$

Combining this with (3.2) we conclude that, for  $n$  big enough,  $\delta(S_n(a^\tau(g)), \beta_A) < \varepsilon$ .  $\square$

**3.2. Preparation for the  $L^1$  case.** The natural framework for the ergodic theorem is  $L^1$ . In this section we will firstly prove Theorem 1.4, which is the main result of this paper.

The strategy of the proof involves constructing good approximations by continuous functions, and obtaining the result for  $L^1$  functions as a consequence of the theorem for continuous functions (Theorem 1.3 above). So, the first questions that arise are: what does good approximation mean and what should we require of the approximation in order to get the  $L^1$  case as a limit of the continuous case? The next two lemmas contain the clues to answer these two questions. The first lemma can be found as [30, Theorem 6.3], and it is often called *fundamental contraction property*. For the sake of completeness we include a simple proof of this fact.

**LEMMA 3.6.** *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

$$\delta(\beta_A, \beta_B) \leq \int_{\Omega} \delta(A(\omega), B(\omega)) dP(\omega). \quad (3.3)$$

*Remark 3.7.* Recall that the definition of  $\beta_A$  (respectively,  $\beta_B$ ) does not depend on the chosen  $y \in M$ .

*Proof.* By the variance inequality (2.5) we get

$$\delta^2(\beta_A, \beta_B) \leq \int_{\Omega} \delta^2(\beta_A, B(\omega)) - \delta^2(\beta_B, B(\omega)) dP(\omega),$$

$$\delta^2(\beta_A, \beta_B) \leq \int_{\Omega} \delta^2(\beta_B, A(\omega)) - \delta^2(\beta_A, A(\omega)) dP(\omega),$$

and the combination of these two inequalities leads to

$$2\delta^2(\beta_A, \beta_B) \leq \int_{\Omega} \delta^2(\beta_A, B(\omega)) + \delta^2(\beta_B, A(\omega)) - \delta^2(\beta_B, B(\omega)) - \delta^2(\beta_A, A(\omega)) dP(\omega).$$

Finally, using the Reshetnyak quadruple comparison theorem (Theorem 2.2), we obtain

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

which is, after some algebraic simplification, the desired result.  $\square$

**LEMMA 3.8.** *Let  $A, B \in L^1(G, M)$ . Given  $\varepsilon > 0$ , for almost every  $g \in G$  there exists  $n_0$ , which may depend on  $g$ , such that*

$$\delta(S_n(a^\tau(g)), S_n(b^\tau(g))) \leq \varepsilon + \int_G \delta(A(g), B(g)) dm(g), \quad (3.4)$$

provided  $n \geq n_0$ .

*Proof.* Indeed, by Corollary 2.4,

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

and therefore, the lemma follows from the Birkhoff ergodic theorem.  $\square$

**3.3. 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  $A_\varepsilon : G \rightarrow M$  such that

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

In some cases there exists an underlying 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  $A_\varepsilon$  can

be constructed by using mollifiers. This idea was used by Karcher in [16]. In the general case, we can use a similar idea.

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

$$A_\eta(g_0) = \operatorname{argmin}_{z \in M} \int_{U_\eta} [\delta^2(z, A(g + g_0)) - \delta^2(y, A(g + g_0))] dm(g). \tag{3.5}$$

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 3.12 below). With this aim, we first prove the following lemma.

LEMMA 3.9. *Let  $A \in L^p(G, M)$  where  $1 \leq p < \infty$ . The function defined by  $\varphi : G \rightarrow [0, +\infty)$  by*

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

*is a continuous function.*

*Proof.* Fix  $z_0 \in M$ , and define the measure

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

on the Borel sets of  $G$ . By definition,  $\nu$  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

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

the inequality

$$\nu(B) = \int_B \delta^p(A(g), z_0) dm(g) < \frac{\varepsilon}{2^{p+1}} \tag{3.6}$$

holds. By the Lusin theorem [11, Theorem 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  $C_\eta$  is (uniformly) continuous.

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

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

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

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



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

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

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

COROLLARY 3.10. For every  $\eta > 0$ , the functions  $A_\eta$  are continuous.

*Proof.* Indeed, by Lemma 3.6

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

So the continuity of  $A_\eta$  is a consequence of the continuity of  $\varphi$  at the identity. □

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

LEMMA 3.11. 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  $A_\eta$  and  $B_\eta$  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 3.6,

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

for all  $g \in G$ . □

We arrive at the main result on approximation.

PROPOSITION 3.12. Given a function  $A \in L^1(G, M)$ , if  $A_\eta$  are the continuous functions defined by (3.5) then

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

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

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

holds. So, using Fubini’s theorem, we obtain

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

By Lemma 3.9, the function  $\varphi$  is continuous. In consequence, if  $e$  denotes the identity of  $G$ , then

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

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, we have that

$$\int_G \delta(A(g), A^{(N)}(g)) \, dm(g) = \int_{\{g: \delta(A(g), z_0) \geq N\}} \delta(A(g), z_0) \, dm(g) \xrightarrow{N \rightarrow \infty} 0. \tag{3.7}$$

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

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

Note that each term of the right-hand side tends to zero: the first by (3.7), the second by the  $L^2$  case done in the first part, and the last by Lemma 3.11. □

3.4. *The  $L^1$  case and almost everywhere convergence.* 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 3.8, 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  $k$ , such that

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

provided  $n \geq n_0$ . In this expression,  $a_{(k)}^\tau$  is the sequence defined in terms of  $A_k$  and  $\tau$  as in (1.6). Fix  $g \in G \setminus N$ . Taking  $k$  so that  $1/k < \varepsilon/4$ , we get that

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

for every  $n \geq n_0$ . By Lemma 3.6, we also have that  $\delta(\beta_A, \beta_{A_k}) \leq \varepsilon/4$ , 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 1.3, there exists  $n_1 \geq 1$  such that, for every  $n \geq n_1$ ,

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

Combining all these inequalities, we obtain that

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

which concludes the proof.  $\square$

3.5. *The  $L^p$  results.* We conclude this section by proving the  $L^p$  ergodic theorems.

**THEOREM 3.13.** *Let  $1 \leq p < \infty$  and  $A \in L^p(G, M)$ . Then*

$$\lim_{n \rightarrow \infty} \int_G \delta^p(S_n(a^\tau(g)), \beta_A) \, dm(g) = 0. \quad (3.8)$$

*Proof.* Let us define the following measure on the Borel sets of  $G$ :

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

By definition,  $\nu$  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

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

we have that

$$v(B) = \int_B \delta^P(A(g), \beta_A) dm(g) < \frac{\varepsilon}{2}. \tag{3.9}$$

By Egoroff’s theorem, as

$$\delta^P(S_n(a^\tau(g)), \beta_A) \xrightarrow{n \rightarrow \infty} 0$$

converge almost everywhere on a finite measure space, there exists a set  $C_\eta \subset G$  with  $m(C_\eta) < \eta$  such that

$$\delta^P(S_n(a^\tau(g)), \beta_A) \xrightarrow{n \rightarrow \infty} 0 \tag{3.10}$$

uniformly on  $G \setminus C_\eta$ .

On the other hand,

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

Therefore, by Jensen’s inequality,

$$\delta^P(S_n(a^\tau(g)), \beta_A) \leq \frac{1}{n} \sum_{k=0}^{n-1} \delta^P(A(\tau^k(g)), \beta_A). \tag{3.11}$$

Now, there exists  $n_0 \in \mathbb{N}$  such that, for all  $n \geq n_0$ ,

$$\int_{G \setminus C_\eta} \delta^P(S_n(a^\tau(g)), \beta_A) dm(g) < \frac{\varepsilon}{2},$$

as a consequence of (3.10). Therefore

$$\begin{aligned} \int_G \delta^P(S_n(a^\tau(g)), \beta_A) dm(g) &= \int_{G \setminus C_\eta} \delta^P(S_n(a^\tau(g)), \beta_A) dm(g) \\ &\quad + \int_{C_\eta} \delta^P(S_n(a^\tau(g)), \beta_A) dm(g) \\ &= \frac{\varepsilon}{2} + \int_{C_\eta} \delta^P(S_n(a^\tau(g)), \beta_A) dm(g). \end{aligned}$$

On the other hand, taking integral over  $C_\eta$  in (3.11), we obtain

$$\begin{aligned} \int_{C_\eta} \delta^P(S_n(a^\tau(g)), \beta_A) dm(g) &\leq \int_{C_\eta} \frac{1}{n} \sum_{k=0}^{n-1} \delta^P(A(\tau^k(g)), \beta_A) dm(g) \\ &= \sum_{k=0}^{n-1} \frac{1}{n} \int_{C_\eta} \delta^P(A(\tau^k(g)), \beta_A) dm(g). \end{aligned}$$

Since  $\int_{C_\eta} dm(g) < \eta$ , by (3.9),

$$v(C_\eta) = \int_{C_\eta} \delta^p(A(g), \beta_A) dm(g) < \frac{\varepsilon}{2}.$$

So, for all  $n \in \mathbb{N}$ ,

$$\sum_{k=0}^{n-1} \frac{1}{n} \int_{C_\eta} \delta^p(A(\tau^k(g)), \beta_A) dm(g) < \frac{\varepsilon}{2}.$$

Finally, combining this two bounds, given  $\varepsilon$ , there exists  $n_0 \in \mathbb{N}$  such that, for all  $n \geq n_0$ ,

$$\int_G \delta^p(S_n(a^\tau(g)), \beta_A) dm(g) < \varepsilon,$$

which concludes the proof.  $\square$

*Acknowledgements.* This work was supported by the Consejo Nacional de Investigaciones Científicas y Técnicas, Argentina (PIP-152), the Agencia Nacional de Promoción de Ciencia y Tecnología, Argentina (PICT 2015-1505), the Universidad Nacional de La Plata, Argentina (UNLP-11X585), and the Ministerio de Economía y Competitividad, Spain (MTM2016-75196-P). The authors would like to thanks to Enrique Pujals for fruitful conversations on an earlier version of this paper. Also we would like to thank the referee for his/her careful reading of the manuscript and his/her suggestions and corrections which have helped us to improve the paper.

## REFERENCES

- [1] A. D. Alexandrov. A theorem on triangles in a metric space and some applications. *Tr. Mat. Inst. Steklova* **38** (1951), 5–23.
- [2] T. Austin. A CAT(0) valued pointwise ergodic theorem. *J. Topol. Anal.* **3** (2011), 145–152.
- [3] M. Bačák. *Convex Analysis and Optimization in Hadamard Spaces (De Gruyter Series in Nonlinear Analysis and Applications, 22)*. De Gruyter, Berlin, 2014.
- [4] W. Ballmann. *Lectures on Spaces of Nonpositive Curvature (DMV Seminar, 25)*. Birkhäuser Verlag, Basel, 1995.
- [5] W. Ballmann, M. Gromov and V. Schroeder. *Manifolds of Nonpositive Curvature (Progress in Mathematics, 61)*. Birkhäuser, Boston, 1985.
- [6] F. Barbaresco. 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)*. Ed. F. Nielsen. Springer, Berlin, 2009, pp. 124–163.
- [7] R. Bhatia and R. Karandikar. Monotonicity of the matrix geometric mean. *Math. Ann.* **353**(4) (2012), 1453–1467.
- [8] D. Bini and B. Iannazzo. Computing the Karcher mean of symmetric positive definite matrices. *Linear Algebra Appl.* **438** (2013), 1700–1710.
- [9] J. Bochi and A. Navas. A geometric path from zero Lyapunov exponents to rotation cocycles. *Ergod. Th. & Dynam. Sys.* **35** (2015), 374–402.
- [10] M. R. Bridson and A. Haefliger. *Metric Spaces of Non-positive Curvature (Grundlehren der Mathematischen Wissenschaften, 319)*. Springer-Verlag, Berlin, 1999.
- [11] R. M. Dudley. *Real Analysis and Probability (Cambridge Studies in Advanced Mathematics, 74)*. Cambridge University Press, Cambridge, 2002.

- [12] A. Es-Sahib and H. Heinich. Barycentre canonique pour un espace metrique a courbure negative. *Seminaire de Probabilites (Lecture Notes in Mathematics, 1709)*. Eds. J. Azéma, M. Émery, M. Ledoux and M. Yor. Springer, Berlin, 1999.
- [13] M. Gromov. *Structures métriques pour les variétés Riemanniennes (Rédigé par J. Lafontaine et P. Pansu, Textes Math., 1)*. CEDIC/Fernand Nathan, Paris, 1981.
- [14] M. Gromov. Hyperbolic groups. *Essays in Group Theory (Mathematical Sciences Research Institute Publications, 8)*. Ed. S. M. Gerten. Springer-Verlag, New York, 1987, pp. 75–264.
- [15] J. Jost. *Nonpositive Curvature: Geometric and Analytic Aspects (Lectures in Mathematics ETH Zurich)*. Birkhäuser, Basel, 1997.
- [16] H. Karcher. Riemannian center of mass and mollifier smoothing. *Comm. Pure Appl. Math.* **30** (1977), 509–541.
- [17] J. Lawson and Y. Lim. Monotonic properties of the least squares mean. *Math. Ann.* **351** (2011), 267–279.
- [18] J. Lawson and Y. Lim. Contractive barycentric maps. *J. Operator Theory* **77** (2017), 87–107.
- [19] J. Lee and A. Naor. Extending Lipschitz functions via random metric partitions. *Invent. Math.* **160** (2005), 59–95.
- [20] Y. Lim and M. Pálfia. Matrix power mean and the Karcher mean. *J. Funct. Anal.* **262** (2012), 1498–1514.
- [21] Y. Lim and M. Pálfia. Weighted deterministic walks and no dice approach for the least squares mean on Hadamard spaces. *Bull. Lond. Math. Soc.* **46** (2014), 561–570.
- [22] M. Mendel and A. Naor. Spectral calculus and Lipschitz extension for barycentric metric spaces. *Anal. Geom. Metr. Spaces* **1** (2013), 163–199.
- [23] M. Moakher and M. Zerai. 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.
- [24] A. Navas. An  $L^1$  ergodic theorem with values in a non-positively curved space via a canonical barycenter map. *Ergod. Th. & Dynam. Sys.* **33** (2013), 609–623.
- [25] S. Ohta. Extending Lipschitz and Hölder maps between metric spaces. *Positivity* **13** (2009), 407–425.
- [26] B. Pass. Uniqueness and Monge solutions in the multimarginal optimal transportation problem. *SIAM J. Math. Anal.* **43** (2011), 2758–2775.
- [27] B. Pass. Optimal transportation with infinitely many marginals. *J. Funct. Anal.* **264** (2013), 947–963.
- [28] Y. G. Reshetnyak. Inextensible mappings in a space of curvature no greater than  $K$ . *Sib. Math. J.* **9** (1968), 683–689.
- [29] K.-T. Sturm. Nonlinear martingale theory for processes with values in metric spaces of nonpositive curvature. *Ann. Probab.* **30**(3) (2002), 1195–1222.
- [30] K.-T. Sturm. Probability measures on metric spaces of nonpositive curvature. *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces (Contemporary Mathematics, 338)*. Eds. P. Auscher, T. Coulhon and A. Grigor’yan. American Mathematical Society, Providence, RI, 2003.
- [31] T. Tao. *Poincare’s Legacies: Pages from Year Two of a Mathematical Blog. Part I*. American Mathematical Society, Providence, RI, 2009.
- [32] P. Walters. *An Introduction to Ergodic Theory (Graduate Texts in Mathematics, 79)*. Springer-Verlag, New York, 1982.