Ergodic theorems for the L^1-Karcher mean

Jorge Antezana[1,2,3*], Eduardo Ghiglioni[2,3†], Yongdo Lim[4†], Miklós Pálfia[5,6†]

[1*]Departamento de Matemática de la Universidad Autónoma de Madrid.
[2*]Departamento de Matemática, FCE-UNLP, La Plata, Argentina.
[3*]Instituto Argentino de Matemática, 'Alberto P. Calderón', CONICET, CABA, Argentina.
[4]Department of Mathematics, Sungkyunkwan University, Suwon 440-746, Korea.
[5]Department of Mathematics, Corvinus University of Budapest, H-1093 Budapest, Fövám tér 8., Hungary.
[6]Bolyai Institute, Interdisciplinary Excellence Centre, University of Szeged, H-6720 Szeged, Hungary.

*Corresponding author(s). E-mail(s): jorge.antezana@uam.es;
Contributing authors: eghiglioni@mate.unlp.edu.ar; ylim@skku.edu; miklos.palfia@uni-corvinus.hu;
†These authors contributed equally to this work.

Abstract

Recently the Karcher mean has been extended to the case of probability measures of positive operators on infinite-dimensional Hilbert spaces as the unique solution of a nonlinear operator equation on the convex Banach-Finsler manifold of positive operators. Let (Ω, μ) be a probability space, and let $\tau : \Omega \to \Omega$ be a totally ergodic map. The main result of this paper is a new ergodic theorem for functions $F \in L^1(\Omega, \mathbb{P})$, where $\mathbb{P}$ is the open cone of the strictly positive operators acting on a (separable) Hilbert space. In our result, we use inductive means to average the elements of the orbit, and we prove that almost surely these averages converge to the Karcher mean of the push-forward measure $F_*(\mu)$. From our result we recover the strong law of large numbers and the "no dice" results proved by the third and fourth authors in the article *Strong law of large numbers for the* L^1*-Karcher mean*, Journal of Func. Anal. 279 (2020). From our main result, we also deduce an ergodic theorem for Markov chains with state space included in $\mathbb{P}$.

Keywords: Karcher mean; Inductive means; Ergodic theorem; Law of large numbers.

MSC Classification: 47A56; 47A64; 37A30; 47A35; 60F15

1 Introduction

Let $\mathcal{H}$ be a complex and separable Hilbert space. In this note $B(\mathcal{H})$, $\mathbb{S}$ and $\mathbb{P}$ stand for the sets of (bounded) linear operators, selfadjoint operators, and strictly positive operators acting in $\mathcal{H}$ respectively. The unitary group of $B(\mathcal{H})$ is indicated by $\mathcal{U}(\mathcal{H})$. If $X \in B(\mathcal{H})$, then $\|X\|$ stands for the usual operator norm, and we will use $|\cdot|$ to indicate the modulus of an operator, i.e. $|X| = \sqrt{X^*X}$. On $\mathbb{S}$ the closure $\overline{\mathbb{P}}$ generates a partial order $\leq$, often known as Löwner order.

When $\mathcal{H}$ is finite-dimensional, then $\mathbb{P}$ can be equipped with the a trace metric

$$ds = \left(\operatorname{tr}(A^{-1}dA)^2\right)^{1/2} = \|A^{-1/2}dA\,A^{-1/2}\|_2,$$

where $\|\cdot\|_2$ denotes the Frobenius, also known as Hilbert-Schmidt, norm. In other words, if $\alpha : [a,b] \to \mathbb{P}$ is a piecewise smooth curve, its length is defined by

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

This induces a Riemannian structure in $\mathbb{P}$, in which the distance between $A, B \in \mathbb{P}$ defined by

$$d_2(A,B) = \inf\{\mathrm{L}(\alpha) : \ \alpha \text{ is a piecewise smooth curve connecting } A \text{ with } B\}$$

can be computed by the formula

$$d_2(A,B) = \|\log(A^{-1/2}BA^{-1/2})\|_2.$$

The geodesic joining A and B also has a closed form expression

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

for $0 \leq t \leq 1$. In operator theory, $\gamma_{AB}(t)$ is usually denoted by $A\#_t B$, and it is called the weighted geometric mean of A and B. From the metric point of view, $\mathbb{P}$ endowed with the metric d_2 is a CAT(0) or Hadamard space. In this setting, the aforementioned weighted geometric means are just a particular case of the notion of a barycenter. We say that a measure μ defined in the Borel sets of $\mathbb{P}$ belongs to the space $\mathcal{P}^1(\mathbb{P})$ if for some (thus all) $A \in \mathbb{P}$

$$\int d_2(X,A)d\mu(X) < \infty.$$

For such a measure we define the barycenter of μ, also known as the Karcher mean of μ, by

$$\Lambda(\mu) = \arg\min_{X\in\mathbb{P}} \int_{\mathbb{T}} \delta^2(A,X)) - \delta^2(A,B))\, d\mu(A), \tag{1}$$

where B is any element of $\mathbb{P}$. It is not difficult to check that $\Lambda(\mu)$ does not depend on the choice of $B \in \mathbb{P}$. It is also easy to check that the weighted geometric means mentioned before are the barycenter of the atomic measures $(1-t)\delta_A + t\delta_B$, where δ_C denotes the unit mass probability measure at the point $C \in \mathbb{P}$. For these particular cases, we have a closed formula for the barycenter, given by

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

This is no longer true in general for measures with three or more points in their support. This leads to investigating different ways to approximate the Karcher's mean.

In [26], Sturm defined the inductive means. To motivate their definition, take a sequence $\{a_n\}_{n\in\mathbb{N}}$ of complex numbers. Then, trivially we can rewrite the arithmetic means as follows:

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

Let $s_{a,b}(t) = t\,b + (1-t)a$, and for a moment allow us to use the notation $a \natural_t b = s_{a,b}(t)$. 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. 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 in $\mathbb{P}$[1]. Given a sequence $A = \{A_n\}_{n\in\mathbb{N}}$ whose elements belong to $\mathbb{P}$, then the inductive means are defined as follows:

$$\begin{aligned} &S_1(A) = A_1 \\ &S_n(A) = S_{n-1}(A) \#_{\frac{1}{n}} A_n \qquad (n \geq 2). \end{aligned}$$

[1]This generalization can be done in any CAT(0) space in the same way, because in these spaces there exists a notion of (metric) geodesics.

Note that these means can be computed explicitly, and they inherit from the weighted geometric means the monotonicity with respect to the Löwner order. Using these inductive means, Sturm in [26] proved the following version of the law of large numbers:

Theorem A (Sturm)**.** *Let Y_i be a sequence of i.i.d. bounded random variables taking values in $\mathbb{P}$. If μ denotes the (common) law of the variables Y_i, then*

$$S_n(\{Y_i\}_i) \xrightarrow[n\to\infty]{a.s.} \Lambda(\mu).$$

Sturm's law of large numbers was used by Lawson-Lim [14] and Bhatia-Karandikar [4] to prove the monotonicity of

$$\Lambda\Big(\frac{1}{k}\sum_{i=1}^{k}\delta_{A_i}\Big)$$

with respect to the variables $A_i \in \mathbb{P}$ and the Löwner order, an important conjecture in matrix analysis. Later a deterministic, also called "no dice", version of Sturm's law that periodically recycles all the points A_i was proved by Holbrook [9]. This result can be thought of as an ergodic theorem in a finite cyclic group. This reinterpretation of Holbrook's result was the motivation of the following Birkhoff-type ergodic theorem proved in [2]. 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 consider an ergodic automorphism $\tau(h) = h + g$ for some $g \in G$. In this framework, the following result holds.

Theorem B. *Let $A : G \to \mathbb{P}$ such that the push forward measure $\mu_A = A_*(m)$ belongs to $\mathcal{P}^1(\mathbb{P})$. Then, for almost every $a \in G$*

$$\lim_{n\to\infty} S_n(\, A^\tau(a)\,) = \Lambda(\mu_A). \tag{2}$$

where $A^\tau(a)$ is the forward ergodic orbit $\{A(\tau^n(a))\}_{n\in\mathbb{N}}$.

1.1 The infinite dimensional setting

It turns out that this metric formulation of Λ is no longer possible when $\mathcal{H}$ is infinite dimensional. In the infinite dimensional setting the available metrics on $\mathbb{P}$ are no longer 2-convex, thus it is far from being a CAT(0) space. Recall that a function $h : \mathcal{H} \to (-\infty, \infty]$ is strongly convex with parameter $\kappa > 0$ if,

$$h((1-t)x + ty) \le (1-t)h(x) + th(y) - \kappa t(1-t)d(x,y)^2,$$

for every $x, y \in \mathcal{H}$ and $t \in [0, 1]$. Having established this terminology, it can be proved that a geodesic metric space (X, d) is CAT(0) if and only if the function $x \mapsto d(x, z)^2$ is strongly convex with parameter $\kappa = 1$, for each $z \in X$ (see [3] for further information).

As a matter of fact, the natural metric structure on $\mathbb{P}$, is a Finsler structure where the length of a piecewise smooth curve $\alpha : [a,b] \to \mathbb{P}$, and $A, B \in \mathbb{P}$ we define

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

Recall that $\|\cdot\|$ is the usual operator norm. Using this definition of length of a curve we can define the following distance

$$d_\infty(A,B) = \inf\{\mathrm{L}(\alpha) : \ \alpha \text{ is a piecewise smooth curve connecting } A \text{ with } B\}.$$

This metric structure makes the natural actions of $GL(\mathcal{H})$ by conjugation isometries. In [5], this metric is characterized in the following way

$$d_\infty(A,B) = \|\log(A^{-1/2}BA^{-1/2})\|.$$

This proved that the Finsler metric coincides with the so called Thompson metric, which can be alternatively characterized as

$$\min\{r \geq 0 : \ e^{-r}\, A \leq B \leq e^r\, A\ \}.$$

Let $\mathcal{P}(\mathbb{P})$ denote the set of τ-additive Borel probability measures. Recall that finite Borel measures on separable metric spaces are known to be τ-additive (see the beginning of Section 3.1 for the definition). For $1 \leq \theta < \infty$, let $\mathcal{P}^\theta(\mathbb{P})$ be the set of measures $\mu \in \mathcal{P}(\mathbb{P})$ such that

$$\int d_\infty^\theta(X,A)d\mu(X) < \infty,$$

for some (and therefore for all) $A \in \mathbb{P}$. Note that the minimization problem (1) has more than one solution if we replace d_2 by d_∞. Therefore, the barycenter can not be defined through the minimization problem. A breakthrough idea was to define the Karcher mean of a measure $\mu \in \mathcal{P}^1(\mathbb{P})$ by using the Karcher equation

$$\int_{\mathbb{P}} \log_X A d\mu(A) = 0, \tag{3}$$

where $\log_X A := X^{1/2}\log(X^{-1/2}AX^{-1/2})X^{1/2}$. This approach requires proving that the Karcher equation has a unique solution, which was a very challenging problem. The following theorem is the final result of a deep work done in [13], [15], and [17].

Theorem C. *Let $\mu \in \mathcal{P}^1(\mathbb{P})$. Then, the Karcher equation*

$$\int_{\mathbb{P}} \log_X A d\mu(A) = 0, \tag{4}$$

has a unique positive definite solution $\Lambda(\mu)$.

This theorem allowed the following definition.

Definition 1 (Karcher mean). *For a* $\mu \in \mathcal{P}^1(\mathbb{P})$ *the* Karcher mean *is defined as the unique solution* $\Lambda(\mu)$ *of* (6).

Defined in this way, the Karcher mean is an extension of the mean defined for matrices. Using this definition, Lim and Pálfia proved in [18] the following generalization of Sturm's Law of Large Numbers.

Theorem D. *Let* Y_i *be a sequence of i.i.d. random variables taken values in* $\mathbb{P}$*. If* $\mu \in \mathcal{P}^1(\mathbb{P})$ *denotes the (common) law of the variables* Y_i*, then*

$$S_n(A) \xrightarrow[n\to\infty]{a.s.} \Lambda(\mu).$$

2 Main results in this work

Let (Ω, μ) be a probability space, and let $\tau : \Omega \to \Omega$ be a totally ergodic map, that is, a map τ such that

$$\tau^n = \underbrace{\tau \circ \ldots \circ \tau}_{\text{n times}}$$

is ergodic for every $n \in \mathbb{N}$. Recall that a characterization of ergodicity is the following: given measurable sets A and B then

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \mu(\tau^{-k}(A) \cap B) = \mu(A)\mu(B).$$

The map τ is called **weak mixing** if

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \left|\mu(\tau^{-k}(A) \cap B) - \mu(A)\mu(B)\right| = 0$$

and **strong mixing** if

$$\lim_{n\to\infty} \mu(\tau^{-n}(A) \cap B) = \mu(A)\mu(B).$$

It is well known that strong mixing $\implies$ weak mixing $\implies$ total ergodicity. Now, using this terminology, our main result can be stated in the following way

Theorem 1. *Let* (Ω, π) *be a probability space, and* $\tau : \Omega \to \Omega$ *be a totally ergodic map. If* $A : \Omega \to \mathbb{P}$ *is a measurable function such that* $\mu = A_*\pi$ *belongs to* $\mathcal{P}^1(\mathbb{P})$*, then for almost every* $\omega \in \Omega$

$$\lim_{n\to\infty} S_n(A^\tau(\omega)) = \Lambda(\mu). \tag{5}$$

where $A^\tau(\omega)$ *is the forward ergodic orbit* $\{A(\tau^n(\omega))\}_{n\in\mathbb{N}}$*.*

This theorem not only generalizes to the infinite-dimensional setting the Theorem B, but it also generalizes that theorem in the sense that it applies to a wider class of ergodic maps. Indeed, using standard arguments in ergodic theory, Theorem B can be applied to equicontinuous dynamical systems. However, equicontinuity is necessary in the proof. In Theorem 1, the topological restrictions disappears. This allows us to apply the theorem, for instance, to the often-called shift maps. Using these maps and a well-known trick, we can deduce Lim-Pálfia's Law of Large Numbers as a consequence of Theorem 1. More precisely, let $X_n : \Omega \to \mathbb{P}$ be a sequence of random vectors defined in a probability space (Ω, μ). Define the map $\varphi : \Omega \to \mathbb{P}^{\mathbb{N}}$ by

$$\varphi(\omega) = (X_1(\omega), X_2(\omega), \ldots),$$

and the measure $\nu = \varphi_*(\mu)$. In $\mathbb{P}^{\mathbb{N}}$ we consider the shift operator

$$\tau(A_1, A_2, A_3, \ldots) = (A_2, A_3, \ldots).$$

It is well known that this map is strong mixing. On the other hand, if $p_1 : \mathbb{P}^{\mathbb{N}} \to \mathbb{P}$ denotes the projection onto the first coordinate, then

$$S_n(\{X_i(\omega)\}) = S_n(\{p_1(\tau^k(x)\}_k) \quad \text{and} \quad (p_1)_*(\nu) = \mu.$$

In this way, we can get Theorem D as a consequence of Theorem 1. The same trick can be applied to get an ergodic theorem in the setting of operator valued Markov chains (see Section 5 for more details)

2.1 Organization of the paper

The paper is organized as follows. In Section 3 we list some preliminaries that we will need throughout the paper. Section 4 is devoted to the proof of Theorem 1. Firstly we will prove the case where the function $A : \Omega \to \mathbb{P}$ is bounded, and later on we will deduce from this particular case the general L^1 case. Finally, Section 5 is devoted to some application of our main result to the study of operator valued Markov chains.

3 Preliminaries

3.1 The Karcher mean

Recall that $\mathcal{P}(\mathbb{P})$ denotes the set of τ-additive Borel probability measures, that is, $\mu \in \mathcal{P}(\mathbb{P})$ satisfies

$$\mu\Big(\bigcup_{\alpha} U_\alpha\Big) = \sup_{\alpha} \mu(U_\alpha),$$

for any directed family $\{U_\alpha\}$ of open sets. Equivalently, a σ-finite probability measure is τ-additive if and only if it is fully supported, that is $\mu(\mathrm{supp}(\mu)) = 1$ (See [8, 12]). For $1 \leq \theta < \infty$, $\mathcal{P}^\theta(\mathbb{P})$ is the set of measures $\mu \in \mathcal{P}(\mathbb{P})$ such that

$$\int d_\infty^\theta(X, A) d\mu(X) < \infty,$$

for some (and therefore for all) $A \in \mathbb{P}$. For a measure $\mu \in \mathcal{P}^1(\mathbb{P})$. the Karcher equation

$$\int_{\mathbb{P}} \log_X A d\mu(A) = 0, \tag{6}$$

has a unique positive definite solution $\Lambda(\mu)$, and by definition this solution is the *Karcher mean*. As in the finite dimensional setting, the Karcher mean in $\mathbb{P}$ is contractive with respect to the Wasserstein distance. Recall that, the L^1-Wasserstein distance between $\mu, \nu \in \mathcal{P}^1(\mathbb{P})$ is defined as

$$W_1(\mu, \nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\mathbb{P}\times\mathbb{P}} d_\infty(A, B)\, d\gamma(A, B) \tag{7}$$

where $\Pi(\mu, \nu)$ denotes the set of all τ-additive Borel probability measures on the product space $\mathbb{P} \times \mathbb{P}$ with marginals μ and ν.

Theorem 2 (see Proposition 2.5. [13])**.** *Let $A_i, B_i \in \mathbb{P}$ for $1 \leq i \leq n$. Then Λ for $\mu = \frac{1}{n}\sum_{i=1}^n \delta_{A_i}$ and $\nu = \frac{1}{n}\sum_{i=1}^n \delta_{B_i}$ satisfies*

$$d_\infty(\Lambda(\mu), \Lambda(\nu)) \leq \frac{1}{n}\sum_{i=1}^n d_\infty(A_i, B_i). \tag{8}$$

In particular, by permutation invariance of Λ in the variables $(A_1, \ldots, A_n)$, we have

$$d_\infty(\Lambda(\mu), \Lambda(\nu)) \leq \min_{\sigma \in S_n} \frac{1}{n}\sum_{i=1}^n d_\infty(A_i, B_{\sigma(i)}) = W_1(\mu, \nu). \tag{9}$$

As a consequence, the following result is obtained in [18] for general measures.

Theorem 3. *For all $\mu, \nu \in \mathcal{P}^1(\mathbb{P})$ it holds that*

$$d_\infty(\Lambda(\mu), \Lambda(\nu)) \leq W_1(\mu, \nu). \tag{10}$$

3.2 Inductive means

Let $A, B \in \mathbb{P}$ and $t \in [0, 1]$, recall that *weighted geometric mean $A\#_t B$* is

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

If $\mu = (1-t)\delta_A + t\delta_B$, then $A\#_t B$ is the unique solution of the corresponding Karcher equation

$$\int_{\mathbb{P}} \log_X A d\mu(A) = (1-t)\log_X A + t\log_X B = 0$$

As we mentioned in the introduction, given a sequence $A \in \mathbb{P}^{\mathbb{N}}$, the *inductive means* are define as:

$$
\begin{aligned}
S_1(A) &= A_1 \\
S_n(A) &= S_{n-1}(A) \#_{\frac{1}{n}} A_n \qquad (n \geq 2).
\end{aligned}
$$

As a consequence of Theorem 2, applied to the weighted geometric mean, we directly get by induction the following result.

Lemma 4. *Given $A, B \in \mathbb{P}^{\mathbb{N}}$, then*

$$
d_\infty(S_n(A), S_n(B)) \leq \frac{1}{n} \sum_{i=1}^{n} d_\infty(A_i, B_i). \tag{11}
$$

Remark 1. *In the finite dimensional setting, Theorem 2, Theorem 3, and Lemma 4 also hold replacing d_∞ by d_2.* ▲

3.3 Evolution systems related to Λ

The fundamental W_1-contraction property (10) leads to the development of an ODE flow theory for Λ, which resembles the gradient flow theory for its potential function in the finite dimensional case or in CAT(0)-spaces (see [16, 23] and the monograph [3] for the CAT(0) gradient flow). The development of this theory for Λ and the Thompson metric is done in [18]. In this last part of the preliminaries, we collect some results from that theory that we will need in this work.

Definition 2 (Resolvent operator)**.** *Given $\mu \in \mathcal{P}^1(\mathbb{P})$ we define the resolvent operator for $\lambda > 0$ and $X \in \mathbb{P}$ as*

$$
J_\lambda^\mu(X) := \Lambda\left(\frac{\lambda}{\lambda+1}\mu + \frac{1}{\lambda+1}\delta_X\right), \tag{12}
$$

that is, the unique solution of the Karcher equation

$$
\frac{\lambda}{\lambda+1}\int_{\mathbb{P}} \log_Z A \, d\mu(A) + \frac{1}{\lambda+1}\log_Z(X) = 0.
$$

Many results involving the resolvent operator that hold in CAT(0) spaces have been also proved in this setting, for instance, the following contraction property of the resolvent operator.

Proposition 5 (Resolvent contraction)**.** *[See [18], Proposition 4.1, for a proof] Given $\mu \in \mathcal{P}^1(\mathbb{P})$, for $\lambda > 0$ and $X, Y \in \mathbb{P}$ we have*

$$
d_\infty(J_\lambda^\mu(X), J_\lambda^\mu(Y)) \leq \frac{1}{1+\lambda} d_\infty(X, Y). \tag{13}
$$

From the definition it is not difficult to see that $J_\lambda^\mu(\Lambda(\mu)) = \Lambda(\mu)$. Moreover, we have the following convergence result.

Proposition 6. *[See [18], Proposition 5.1, for a proof] Let $\mu \in \mathcal{P}^1(\mathbb{P})$, $d \geq 0$ an integer and $X \in \mathbb{P}$. Let*

$$\begin{cases} X_0 := X \\ X_k := J^{\mu}_{1/(k+d)}(X_{k-1}) \quad \text{if } k > 0. \end{cases}$$

Then $d_\infty(X_k, \Lambda(\mu)) \to 0$.

The next result will be very important to relate the ergodic theorems with respect to inductive means and the classical Birkhoff theorem.

Lemma 7. *Let $\mu, \nu_i \in \mathcal{P}^1(\mathbb{P})$ for $i \in \mathbb{N}$, $l \geq 0$ an integer and $X_0, Y_0 \in \mathbb{P}$. Let*

$$X_{k+1} := J^{\mu}_{1/(l+k+1)}(X_k) \quad \text{and} \quad Y_{k+1} := J^{\nu_{k+1}}_{1/(l+k+1)}(Y_k).$$

Then

$$d_\infty(X_{k+1}, Y_{k+1}) \leq \frac{l+1}{k+l+1} d_\infty(X_0, Y_0) + \frac{1}{k+l+1} \sum_{i=l+1}^{k+1} W_1(\mu, \nu_i).$$

Proof. As $d_\infty(\Lambda(\mu), \Lambda(\nu)) \leq W_1(\mu, \nu)$ and as the W_1 distance is convex we have that for any $\lambda > 0, x, y \in \mathbb{P}$ and $\mu, \eta \in \mathcal{P}^1(\mathbb{P})$,

$$\begin{aligned} d_\infty(J^{\mu}_{\lambda}(x), J^{\eta}_{\lambda}(x)) &= d_\infty\left(\Lambda\left(\frac{1}{1+\lambda}\delta_x + \frac{\lambda}{1+\lambda}\mu\right), \Lambda\left(\frac{1}{1+\lambda}\delta_y + \frac{\lambda}{1+\lambda}\eta\right)\right) \\ &\leq W_1\left(\frac{1}{1+\lambda}\delta_x + \frac{\lambda}{1+\lambda}\mu, \frac{1}{1+\lambda}\delta_y + \frac{\lambda}{1+\lambda}\eta\right) \\ &\leq \frac{1}{1+\lambda} W_1(\delta_x, \delta_y) + \frac{\lambda}{1+\lambda} W_1(\mu, \eta) \\ &= \frac{1}{1+\lambda} d_\infty(x, y) + \frac{\lambda}{1+\lambda} W_1(\mu, \eta). \end{aligned}$$

Thus when $\lambda = 1/k$ we get that

$$d_\infty(J^{\mu}_{1/k}(x), J^{\eta}_{1/k}(x)) \leq \frac{k}{1+k} d_\infty(x, y) + \frac{1}{1+k} W_1(\mu, \eta),$$

so applying this inequality to $d_\infty(X_{k+1}, Y_{k+1})$ and induction we get

$$d_\infty(X_{k+1}, Y_{k+1}) \leq \frac{l+1}{k+l+1} d_\infty(X_0, Y_0) + \frac{1}{k+l+1} \sum_{i=l+1}^{k+1} W_1(\mu, \nu_i).$$

proving the assertion. ∎

4 Proofs of the main results

This section is devoted to the proof of Theorem 1. Throughout this section, (Ω, π) is a probability space, and $\tau : \Omega \to \Omega$ is a totally ergodic map. Given a measurable function $A : \Omega \to \mathbb{P}$, we say that $F \in L^1(\Omega, \mathbb{P})$ if $A_*\pi \in \mathcal{P}^1(\mathbb{P})$.

Also recall that, for the sake of simplicity, we use the following notation. Given a function $A : \Omega \to \mathbb{P}$, let $A^\tau : \Omega \to \mathbb{P}^\mathbb{N}$ be defined by

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

In other words, for each $\omega \in \Omega$, the sequence $A^\tau(\omega)$ is the forward ergodic orbit.

4.1 The bounded case

Our first main result is the following bounded version of the Theorem 1.

Theorem 8. *Let (Ω, π) be a probability space, and $\tau : \Omega \to \Omega$ be a totally ergodic map. If $A : \Omega \to \mathbb{P}$ is a bounded measurable function and $\mu = A_*\pi$, then for almost every $\omega \in \Omega$,*

$$\lim_{n\to\infty} S_n(A^\tau(\omega)) = \Lambda(\mu). \tag{14}$$

To begin with, we will consider an arbitrary sequence $A \in \mathbb{P}^\mathbb{N}$, and only at the end we will consider the sequence associated to the ergodic orbit. The proof of Theorem 8 is a combination of different techniques developed in [18], among them the following two results. The proofs of these results can be found in [18] Lemma 5.5. and Theorem 5.6 equation (47) respectively.

Lemma 9. *Let $A \in \mathbb{P}^\mathbb{N}$. Then, for $n \geq 1$ big enough and every $k \in \mathbb{N}$ there exists $E \in \mathbb{S}$ such that*

$$S_{kn}(A) - S_{k(n+1)}(A) + E + \frac{1}{n+1}\,\frac{1}{k}\sum_{j=1}^{k} \log_{S_{k(n+1)}(A)} A_{kn+j} = 0 \tag{15}$$

and for some constant $C > 0$

$$\|E\| \leq C\,\frac{\operatorname{diam}\operatorname{supp}(\mu)}{n^2}.$$

Lemma 10. *Let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathbb{P}^\mathbb{N}$. Given $k \in \mathbb{N}$, define*

$$\mu_{k,n} = \frac{1}{k}\sum_{j=1}^{k} \delta_{A_{kn+j}}.$$

Then, there exists $N \geq 1$ big enough such that for every $n \geq N$ there exists $\overline{S}_{kn}$ such that

$$d_\infty(S_{kn}(A), \overline{S}_{kn}) = O(n^{-2}), \tag{16}$$

and

$$S_{k(n+1)}(A) = J^{\mu_{k,n}}_{1/(n+1)} \overline{S}_{kn}. \tag{17}$$

We will also need the following number theoretical lemma (see for instance [20]).
Lemma 11. *Let $a_k \geq 0$ be a sequence such that*

$$a_{k+1} \leq \left(1 - \frac{1}{k+1}\right) a_k + \frac{\beta}{(k+1)^2},$$

where $\alpha, \beta > 0$. Then

$$a_k \leq \frac{\beta(1 + \log(k+1))}{k+1}.$$

Now we are ready to proceed to the proof of Theorem 8.

Proof of Theorem 8. Assume that N is big enough. Using the notation of Lemma 10, define

$$\begin{cases} \widetilde{S}_{kn} = \overline{S}_{kN} & \text{if } 1 \leq n \leq N \\ \widetilde{S}_{k(n+1)} = J^{\mu_{k,n}}_{1/(n+1)} \widetilde{S}_{kn} & \text{if } n > N \end{cases}.$$

Then, by the resolvent contraction (Proposition 5) and (16) we get that

$$d_\infty(S_{k(n+1)}(A), \widetilde{S}_{k(n+1)}) \leq \left(1 - \frac{1}{n+2}\right) d_\infty(S_{kn}, \widetilde{S}_{kn}) + O(n^{-2}).$$

Therefore, by Lemma 11

$$d_\infty(S_{kn}(A), \widetilde{S}_{kn}) \leq C \frac{\log n}{n}.$$

Now we define

$$\begin{cases} \widehat{S}_n = \widetilde{S}_{nk} & \text{if } 1 \leq n \leq N \\ \widehat{S}_{(n+1)} = J^{\mu}_{1/(n+1)} \widehat{S}_n & \text{if } n > N \end{cases}$$

By Lemma 7 we get that

$$d_\infty(\widetilde{S}_{kn}, \widehat{S}_n) \leq \frac{1}{n} \sum_{j=N}^{n} W_1(\mu, \mu_{k,j}).$$

By Proposition 6 , we know that $\widehat{S}_n \xrightarrow[n\to\infty]{} \Lambda(\mu)$. Recall that

$$\mu_{k,n} = \frac{1}{k}\sum_{j=1}^{k} \delta_{A_{kn+j}}.$$

Till now, we have considered an arbitrary sequence $\{A_n\}_{n\in\mathbb{N}}$. Now, consider the ergodic orbit $\{A(\tau^n(\omega))\}_{n\in\mathbb{N}}$. Then the measures μ_{nk} depend on the point $\omega \in \Omega$. In particular

$$\mu_{k,n+1}(\omega) = \mu_{k,n}(\tau^k(\omega)).$$

Since τ is totally ergodic,

$$\begin{aligned}\frac{1}{n+1}\sum_{j=0}^{n} W_1(\mu, \mu_{k,j}(\omega)) &= \frac{1}{n+1}\sum_{j=1}^{n} W_1(\mu, \mu_{k,0}((\tau^k)^j(\omega))) \\ &\xrightarrow[n\to\infty]{} \int_\Omega W_1(\mu, \mu_{k,0}(\omega))\, d\pi.\end{aligned}$$

Note that the supports of the measure μ and the measures $\mu_{k,0}$ are contained in a bounded set X_0. Hence the functions $\omega \mapsto W_1(\mu, \mu_{k,0}(\omega))$ are uniformly bounded. In consequence, if we can prove that for almost every $\omega \in \Omega$

$$\lim_{k\to\infty} W_1(\mu, \mu_{k,0}(\omega)) = 0, \tag{18}$$

then by the dominated convergence theorem we get that

$$\lim_{k\to\infty} \int_\Omega W_1(\mu, \mu_{k,0}(\omega))\, d\pi = 0.$$

This would conclude the proof. In order to prove (18), first of all note that X_0 can be assumed not only bounded but also separable. So, on the one hand, the boundedness of X_0 implies that the measures $\{\mu_{k,0}\}$ are uniformly integrable in the sense that

$$\lim_{R\to\infty} \limsup_{k\to\infty} \int_{d_\infty(x,A)\geq R} d_\infty(x, A) d\mu_{k,0}(A) = 0.$$

Theferofe, the topology generated by W_1 coincides with the weak-$*$ (also called weak) topology of X_0 (see [18, Prop. 2.2] for a proof of this fact). On the other hand, since any separable metric space admits a totally bounded metrization that preserves the topology (see [6, Thm 2.8.2]), the weak convergence can be tested using only bounded Lipschitz functions BL. Since this space is separable with respect to the sup-norm, if $\{f_n\}_{n\in\mathbb{N}}$ is a dense set in BL, the convergence

$$\int_{X_0} f_n\ d\mu_{k,0} \xrightarrow[k\to\infty]{} \int_{X_0} f_n\ d\mu$$

for each function f_n, follows by the standard Birkhoff ergodic theorem for every ω in a complement of a zero measure set N_n in Ω. So, in the complement of the union N of the sets N_n, we get the desired weak convergence of the measures $\mu_{k,0}$ to μ. ■

4.2 The L^1 case

This section is devoted to extend Theorem 8 to any integrable function. The strategy is to use an approximation argument based on the following two results.

Lemma 12. *Let $A, B \in L^1(\Omega, \mathbb{P})$. If $\beta_A = \Lambda(A_*\pi)$ and $\beta_B = \Lambda(B_*\pi)$, then*

$$d_\infty(\beta_A, \beta_B) \leq \int_\Omega d_\infty(A(\omega), B(\omega))d\pi(\omega). \tag{19}$$

Proof. By Theorem 3 we know that

$$d_\infty(\beta_A, \beta_B) \leq W_1(A_*\pi, B_*\pi).$$

Let $F : \Omega \to \mathbb{P}^2$ be defined by $F(\omega) = (A(\omega), B(\omega))$. Then taking in (7) $\gamma = F_*(\pi)$ we get the desired inequality. ■

Lemma 13. *Let $A, B \in L^1(\Omega, \mathbb{P})$. Given $\varepsilon > 0$, for almost every $\omega \in \Omega$ there exists n_0, which may depend on ω, such that*

$$d_\infty\big(S_n(A^\tau(\omega)), S_n(B^\tau(\omega))\big) \leq \varepsilon + \int_\Omega d_\infty(A(\omega), B(\omega))d\pi(\omega), \tag{20}$$

provided $n \geq n_0$.

Proof. Indeed, by Corollary 4

$$\begin{aligned} d_\infty\big(S_n(A^\tau(g)), S_n(B^\tau(g))\big) &\leq \frac{1}{n}\sum_{k=0}^{n-1} d_\infty\big(A_k^\tau(g), B_k^\tau(g)\big) \\ &= \frac{1}{n}\sum_{k=0}^{n-1} d_\infty\big(A(\tau^k(g)), B(\tau^k(g))\big), \end{aligned}$$

and therefore, the lemma follows from the classical Birkhoff ergodic theorem. ■

Proof of Theorem 1. Let $A \in L^1(\Omega, \mathbb{P})$. As in Lemma 12, let $\beta_A = \Lambda(A_*\pi)$. Define the sequence of bounded functions

$$A_n(\omega) = \begin{cases} A(\omega) & \text{if } d_\infty(A(\omega), \beta_A) \leq n \\ \beta_A & \text{if } d_\infty(A(\omega), \beta_A) > n \end{cases}.$$

Then (see for instance the beginning of the proof of Lemma 6.2 in [18])

$$\int_\Omega d_\infty(A(\omega), A_n(\omega))d\pi(\omega) \xrightarrow[n\to\infty]{} 0.$$

Now, the result follows combining Lemmas 12 and 13. ■

5 Some consequences of the ergodic theorem

In this section we will deduce from Theorem 1, an ergodic theorem with respect to the inductive means for Markov chains with values in $\mathbb{P}$. The trick is the same as the one used in the introduction to deduce Lim-Pálfia's law of large numbers. Let $X_n : \Omega \to \mathbb{P}$ be a sequence random vectors defined in a probability space (Ω, μ). Define the map $\varphi : \Omega \to \mathbb{P}^{\mathbb{N}}$ by

$$\varphi(\omega) = (X_1(\omega), X_2(\omega), \ldots),$$

and the measure $\nu = \varphi_*(\mu)$. In $\mathbb{P}^{\mathbb{N}}$ we consider the shift operator

$$\tau(A_1, A_2, A_3, \ldots) = (A_2, A_3, \ldots).$$

Note that, if $p_1 : \mathbb{P}^{\mathbb{N}} \to \mathbb{P}$ denotes the projection onto the first coordinate, then

$$S_n(\{X_i(\omega)\}) = S_n\big(\{p_1(\tau^k(x)\}_k\big).$$

5.1 Ergodic theorem for Markov chains

Assume that $\{X_n\}_{n\in\mathbb{N}}$ is a Markov chain whose state space is a finite subset Σ of $\mathbb{P}$. Recall that this means that the sequence satisfies the Markov property

$$\mu(X_{n+1} = \sigma_{n+1}|X_n = \sigma_n, \ldots, X_0 = \sigma_0) = \mu(X_{n+1} = \sigma_{n+1}|X_n = \sigma_n), \tag{21}$$

and for any $\sigma_i, \sigma_j \in \Sigma$ the quantity

$$p(i,j) = \mu(X_{n+1} = \sigma_j|X_n = \sigma_i),$$

is independent of n. Let P be the transition matrix, that is, the matrix such that $P_{j,i} = p(i,j)$. Assume that the Markov chain is regular, that is, there exists $n > 0$ such that P^n has all its entries positive[2]. This implies the existence and uniqueness of a stationary distribution π[3].

Now, assume that $\mu(X_1 = \sigma_j) = \pi_j$. Since the chain is regular, the shift map is not only ergodic, but also it is strong mixing with respect to the measure $\nu = \varphi_*(\mu)$.

---

[2]These matrices are also known as primitive matrices in the setting of Perron-Frobenius theory.

[3]There exist weaker conditions under which the stationary measure also exists and it is unique. See for instance chapter 8 of [10] for a matrix analysis approach of this result, or the books [21] and [25] for a probabilistic approach.

This implies that τ is totally ergodic with respect to ν. Since

$$S_n(\{X_i(\omega)\}) = S_n\big(\{p_1(\tau^k(x)\}_k\big) \quad \text{and} \quad (p_1)_*(\nu) = \sum_j \pi_j \delta_{\sigma_j},$$

by Theorem 1 we get that

$$S_n(\{X_i(\omega)\}) \xrightarrow[n\to\infty]{} \Lambda\left(\sum_j \pi_j \delta_{\sigma_j}\right) \quad a.s.$$

in d_∞. In other words, we have proved the following result.

Theorem 14. *Let (Ω, w) be a finite probability space, and let $\{X_n\}_{n\in\mathbb{N}}$ be an ergodic Markov chain. If π denotes the stationary measure associated to the chain, then*

$$S_n(\{X_i(\omega)\}) \xrightarrow[n\to\infty]{} \Lambda\left(\sum_j \pi_j \delta_{\sigma_j}\right) \quad a.s.$$

in d_∞.

6 Concluding remarks

In this paper, we proved a new ergodic theorem for functions $F \in L^1(\Omega, \mathbb{P})$, where $\mathbb{P}$ is the open cone of the strictly positive operators acting on a (separable) Hilbert space. Firstly, we proved the case where the function $F : \Omega \to \mathbb{P}$ is bounded. Later on, we deduce from this particular case the general L^1 case. In our approach, we use inductive means to average the elements of the orbit, and we proved that almost surely these averages converge to the Karcher mean of the push-forward measure $F_*(\mu)$. Finally, we use this result to get some application to the study of operator valued Markov chains. Possible extension of this ergodic theorem could be interesting. For example, extend this result to finite-dimensional symmetric cones or to the positive cone of a JB-algebra.

Acknowledgments. The authors wish to express their gratitude to the referee for valuable comments and suggestions, which helped us in improving the quality of the manuscript. The work of J. Antezana was partially supported by the National Scientific and Technical Research Council of Argentina (Grant: PIP11220200103209CO), the National University of La Plata (Grant: 11X974), and the Spanish State Investigation Agency (PID2022-142202NB-I00). The work of E. Ghiglioni was partially supported by by the National Scientific and Technical Research Council of Argentina (Grant: PIP11220200103209CO), and the National University of La Plata (Grant: 11X974). The work of M. Pálfia was supported by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund and financed under the TKP2021-NVA funding scheme, Project no. TKP2021-NVA-09; the Hungarian National Research, Development and Innovation Office NKFIH, Grant

No. FK128972, Grant No. ÚNKP-23-5 and by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences, Grant No. BO/00998/23/3; and the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) No.2015R1A3A2031159, No.2016R1C1B1011972 and No.2019R1C1C1006405.

Declarations

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

[1] M. Alfsen, W. Schultz, and E. Stormer, A Gelfand-Neumark theorem for Jordan algebras, Advances in Math. 28, (1978), pp. 11-56.

[2] J. Antezana, E. Ghiglioni and D. Stojanoff, Ergodic theorem in CAT(0) spaces in terms of inductive means, Ergodic Theory and Dynamical Systems, 43:5 (2023), pp. 1433-1454.

[3] M. Bačák, Convex analysis and optimization in Hadamard spaces, De Gruyter Series in Nonlinear Analysis and Applications 22 (2014).

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

[5] G. Corach, H. Porta and L. Recht, Geodesics and operator means in the space of positive operators, Int. J. Math., 4:2 (1993) pp. 193–202.

[6] R.M. Dudley, Real Analysis and Probability, Cambridge Studies in Advanced Mathematics, 74. Cambridge University Press, Cambridge (2002).

[7] J. Faraut, A. Korányi, Analysis on symmetric cones, Clarendon Press, Oxford (1994).

[8] F. Hiai, J. Lawson and Y. Lim, The stochastic order of probability measures on ordered metric spaces, J. Math. Anal. Appl., 464:1 (2018), pp. 707–724,

[9] J. Holbrook, No dice: a determinic approach to the Cartan centroid, J. Ramanujan Math. Soc. 27:4 (2012), pp. 509–521.

[10] R. A. Horn, C. R. Johnson. Matrix analysis. Second. Cambridge University Press, Cambridge, 2013.

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

[12] J. Lawson, Ordered Probability Spaces, J. Math. Anal. Appl. 455:1 (2017), pp. 167–179.

[13] J. Lawson and Y. Lim, Karcher means and Karcher equations of positive definite operators, Tran. of the AMS, 1 (2014), pp. 1–22.

[14] J. Lawson and Y. Lim, Monotonic properties of the least squares mean, Math. Ann. 351:2 (2011), pp 267–279.

[15] J. Lawson and Y. Lim, Weighted means and Karcher equations of positive operators, Proceedings of the National Academy of Sciences, 110:39 (2013), pp. 15626–15632.

[16] Y. Lim and M. Pálfia, Weighted deterministic walks for the least squares mean on Hadamard spaces, Bull. of the London Math. Soc. 46:3 (2014), pp. 561–570.

[17] Y. Lim and M. Pálfia, Existence, uniqueness and an ODE approach to the L^1-Karcher mean, Adv. Math. Volume 376, 6 January 2021, 107435.

[18] Y. Lim and M. Pálfia, Strong law of large numbers for the L^1-Karcher mean, Journal of Func. Anal. 279 (2020) 108672.

[19] A. Navas, An L^1 ergodic theorem with values in a nonpositively curved space via a canonical barycenter map, Ergodic Theory and Dynamical Systems 33 (2013), pp. 609–623.

[20] A. Nedic and D. P. Bertsekas, Convergence Rate of Incremental Subgradient Algorithms, Stochastic optimization: algorithms and applications (Gainesville, FL, 2000), 223–264, Appl. Optim., 54, Kluwer Acad. Publ., Dordrecht, 2001.

[21] J. R. Norris, Markov chains. Cambridge Series in Statistical and Probabilistic Mathematics, 2. Cambridge University Press, Cambridge, 1997.

[22] S. Ohta and M. Pálfia, Discrete-time gradient flows and law of large numbers in Alexandrov spaces, Calc. Var. PDE 54 (2015), pp. 1597-1610.

[23] S. Ohta and M. Pálfia, Gradient flows and a Trotter-Kato formula of semi-convex functions on CAT(1)-spaces, Amer. J. Math. 139:4 (2017), 937-965.

[24] M. Pálfia, Operator means of probability measures and generalized Karcher equations, Adv. Math. 289 (2016), pp. 951–1007.

[25] S. M. Ross, Stochastic processes, Second edition. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, 1996.

[26] K.-T. Sturm. 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.

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

[28] C. Villani, Topics in Optimal Transportation (Graduate Studies in Mathematics, 58), Springer, Berlin, 2003.