An *L* [1] ergodic theorem with values in a non-positively curved space via a canonical barycenter map

ANDRÉS NAVAS

Link to this article: http://journals.cambridge.org/abstract_S0143385711001015

How to cite this article:
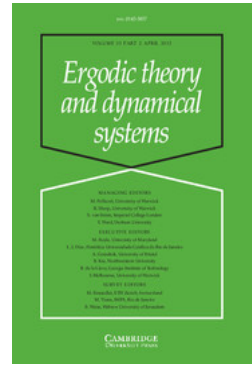ANDRÉS NAVAS (2013). An *L* [1] ergodic theorem with values in a non-positively curved space via a canonical barycenter map. Ergodic Theory and Dynamical Systems, 33, pp 609-623 doi:10.1017/S0143385711001015

Request Permissions : Click here

An L^1 ergodic theorem with values in a non-positively curved space via a canonical barycenter map

ANDRÉS NAVAS

Dep. de Matemáticas, Fac. de Ciencia, Univ. de Santiago, Alameda 3363,
Estación Central, Santiago, Chile
(e-mail: andres.navas@usach.cl)

Abstract. We give a general version of the Birkhoff ergodic theorem for functions taking values in non-positively curved spaces. In this setting, the notion of a Birkhoff sum is replaced by that of a barycenter along the orbits. The construction of an appropriate barycenter map is the core of this note. As a byproduct of our construction, we prove a fixed point theorem for actions by isometries on a Buseman space.

1. *Introduction*

The extension of classical ergodic theorems to a geometric—non-positively curved—setting has been one of the most fascinating developments in ergodic theory in recent years; see [**6**] for a nice survey containing most of the relevant results for functions (cocycles) taking values in isometry groups.

In a different though related direction, Es-Sahib and Heinich proved in [**4**] an ergodic type theorem for L^1 independent and identically distributed (i.i.d.) random variables taking values in a non-positively curved space. An analogous result for L^2 i.i.d. random variables was given by Sturm in [**7**]. Recently, Austin proved a nice extension of Sturm's result to arbitrary measure-preserving actions of amenable groups (see [**1**]). Unfortunately, Austin's L^2 setting is not the most appropriate one in view of the fact that the most powerful framework of the ergodic theorem is that of L^1 spaces. In this work, we prove a general ergodic theorem for L^1 functions taking values in non-positively curved spaces, where the notion of Birkhoff sums is replaced by that of barycenters along the orbits.

Let us begin by recalling a classical construction. Given a complete CAT(0)-space (X, d), we consider the space $P^2(X)$ of probability measures with finite second moment, that is,

$$\int_X d(x, y)^2 \, d\mu(y) < \infty$$

(this condition does not depend on the point $x \in X$). Following Cartan (see, for instance, [5]), to each $\mu \in P^2(X)$ one may associate a *barycenter* bar(μ), namely the unique point that minimizes the function

$$x \to \int_X d(x, y)^2 \, d\mu(y).$$

A crucial property of bar $: P^2(X) \to X$ is that it is 1-Lipschitz for the 2-Wasserstein metric [7]:

$$d(\mathrm{bar}(\mu_1), \mathrm{bar}(\mu_2)) \le W_2(\mu_1, \mu_2) := \inf_{\nu \in (\mu_1 | \mu_2)} \sqrt{\int_{X \times X} d(x, y)^2 \, d\nu(x, y)},$$

where $(\mu_1 | \mu_2)$ denotes the set of all probability measures ν on $X \times X$ that project into μ_1 and μ_2 on the first and the second factor, respectively (see [8] for more details on this metric).

The first task of this work was to introduce an analogous notion for the space $P^1(X)$ of probability measures with finite first moment:

$$\int_X d(x, y) \, d\mu(y) < \infty.$$

It was after we developed a notion of barycenter adapted to our needs that we discovered the equivalent construction of [4]. We decided to include our approach here because it is more elementary in that, unlike [4], it does not rely on deep probabilistic results. Although this makes our computations a little more involved, it has the advantage of allowing us to avoid the (finite) local compactness hypothesis of [4] for the underlying space, thus solving a problem formulated in [7, Example 6.5]. Summarizing, let (X, d) be a complete metric space with non-positive curvature in the sense of Buseman (which, for convenience, we call a *Buseman space*). Assuming that X is separable, in §2 we construct a map bar$^\star : P^1(X) \to X$ that is 1-Lipschitz for the 1-Wasserstein metric:

$$d(\mathrm{bar}^\star(\mu_1), \mathrm{bar}^\star(\mu_2)) \le W_1(\mu_1, \mu_2) := \inf_{\nu \in (\mu_1 | \mu_2)} \int_{X \times X} d(x, y) \, d\nu(x, y).$$

By elementary reasoning, this also applies to any separable Banach space, where geodesics are understood as being segments of lines.

The map constructed above is equivariant with respect to the natural action of isometries. At the end of §2, we give an application of this fact, namely, we prove that every compact group of isometries of a Buseman space has a fixed point. The novelty here is that we do not assume any hypothesis of strict convexity (with such a hypothesis, the result is elementary and well known).

We next describe the goal of this work. Given an amenable group G with a measure-preserving action T on a probability space $(\Omega, \mathcal{P})$, let (F_n) be a *tempered* Følner sequence in G, that is, a Følner sequence for which there exists $C > 0$ such that for all $n \in \mathbb{N}$,

$$m_G \left(\bigcup_{k < n} F_k^{-1} F_n \right) \le C m_G(F_n),$$

where m_G denotes the left Haar measure on X. Let $\varphi : \Omega \to X$ be a measurable function lying in $L^1(\mathcal{P}, X)$, that is, such that for some (equivalently, all) $x \in X$,

$$\int_\Omega d(\varphi(\omega), x)\, d\mathcal{P}(\omega) < \infty.$$

Notice that $L^1(\mathcal{P}, X)$ becomes a metric space when endowed with the distance

$$d_1(\varphi, \psi) := \sqrt{\int_\Omega d(\varphi(\omega), \psi(\omega))\, d\mathcal{P}(\omega)}.$$

MAIN THEOREM. *With the notation above, assume that X is either a separable Banach space or a separable Buseman space. Then*

$$\omega \mapsto \mathrm{bar}^\star\left(\frac{1}{m_G(F_n)} \int_{F_n} \delta_{\varphi(T^g \omega)}\, dm_G(g)\right)$$

is a sequence of maps that converges pointwise and in $L^1(\mathcal{P}, X)$ to a T-invariant function from Ω to X.

For Banach spaces, the barycenter of a measure $(1/m)(\delta_{x_1} + \cdots + \delta_{x_m})$ is just the Dirac measure concentrated at the point $(1/m)(x_1 + \cdots + x_m)$. In particular, when $G \sim \mathbb{Z}$, $X = \mathbb{R}$ and $F_n = \{0, \ldots, n-1\}$, the theorem reduces to the classical (invertible) Birkhoff ergodic theorem for $\varphi \in L^1(\mathcal{P}, \mathbb{R})$.

The proof of our main theorem uses the general strategy of [1], that is, the contractivity properties of the barycenter maps transform the desired convergence into that of suitable sequences of real-valued functions to which Lindenstrauss's pointwise ergodic theorem [3] applies. Recall that in the setting of [1], the probability measure lies in $P^2(X)$ and one considers functions $\varphi : \Omega \to X$ lying in the space $L^2(\mathcal{P}, X)$, that is, such that for some (equivalently, all) $x \in X$,

$$\int_\Omega d(\varphi(\omega), x)^2\, d\mathcal{P}(\omega) < \infty.$$

This space may be naturally endowed with the distance

$$d_2(\varphi, \psi) := \int_\Omega d(\varphi(\omega), \psi(\omega))^2\, d\mathcal{P}(\omega).$$

Austin's theorem then asserts that for every $\varphi \in L^2(\mathcal{P}, X)$, the sequence of maps

$$\omega \mapsto \mathrm{bar}\left(\frac{1}{m_G(F_n)} \int_{F_n} \delta_{\varphi(T^g \omega)}\, dm_G(g)\right) \tag{1}$$

converges pointwise and in $L^2(\mathcal{P}, X)$ to a T-invariant function from Ω to X.

Quite interestingly, Austin's theorem is not a consequence of our main theorem. Indeed, although—as in the classical case—our theorem extends to an L^p version by a straightforward and well-known argument, the barycenters bar and bar* may differ, even for very nice spaces; see Remark 2.3. Despite this, the map bar is also 1-Lipschitz for the 1-Wasserstein metric; see [7, Proposition 4.3]. Using the methods of §3, this allows us to show that the convergence of the sequence of maps (1) actually holds in $L^1(\mathcal{P}, X)$. We point out that this still holds for probability measures in $P^1(X)$ for a clever modification of Cartan's barycenter (see [7, Proposition 4.3]).

2. *The barycenter map*

For a Banach space X, a natural definition of the barycenter of a measure $\mu \in P^1(X)$ is

$$\mathrm{bar}^\star(\mu) := \int_X x \, d\mu(x).$$

Notice that given μ_1, μ_2 in $P^1(X)$, for each $\nu \in (\mu_1|\mu_2)$ we have

$$\int_{X \times X} \|x - y\| \, d\nu(x, y) \geq \left\| \int_{X \times X} (x - y) \, d\nu(x, y) \right\|$$

$$= \left\| \int_{X \times X} x \, d\nu(x, y) - \int_{X \times X} y \, d\nu(x, y) \right\|$$

$$= \left\| \int_X x \, d(\pi_1 \nu)(x) - \int_X y \, d(\pi_2 \nu)(y) \right\|$$

$$= \|\mathrm{bar}^\star(\mu_1) - \mathrm{bar}^\star(\mu_2)\|.$$

As a consequence,

$$\|\mathrm{bar}^\star(\mu_1) - \mathrm{bar}^\star(\mu_2)\| \leq W_1(\mu_1, \mu_2).$$

A definition with an analogous property for non-positively curved spaces is much more subtle. In what follows, X will denote a Buseman space (separability will be needed later). Recall that this means that X is geodesic and the distance function along geodesics is convex. Equivalently, given any two pairs of points x, y and x', y', their corresponding (unique) midpoints m, m' satisfy

$$d(m, m') \leq \frac{d(x, x')}{2} + \frac{d(y, y')}{2}. \tag{2}$$

This property allows us to define a barycenter $\mathrm{bar}_n(x_1, \ldots, x_n)$ of any finite family $(x_1, \ldots, x_n)$ of (not necessarily distinct) points as follows. For $n = 1$, we let $\mathrm{bar}_1(x) := x$. For $n = 2$, we let $\mathrm{bar}_2(x, y)$ be the midpoint between x and y. Now, assuming that the barycenters $\mathrm{bar}_n(\cdot, \ldots, \cdot)$ of all families of n points have been defined, we define $\mathrm{bar}_{n+1}(x_1, \ldots, x_n, x_{n+1})$ as follows: starting with $(x_1, \ldots, x_{n+1}) =: (x_1^{(0)}, \ldots, x_{n+1}^{(0)})$, we replace each x_i by the (already defined) barycenter of $(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{n+1})$. Then we do the same with the resulting set $(x_1^{(1)}, \ldots, x_{n+1}^{(1)})$, thus yielding a new set $(x_1^{(2)}, \ldots, x_{n+1}^{(2)})$. Repeating this procedure and passing to the limit along the Cauchy sequences $(x_i^{(k)})_{k \in \mathbb{N}}$, the corresponding set will collapse to a single point, which we call the barycenter of $(x_1, \ldots, x_{n+1})$ and we denote it by $\mathrm{bar}(x_1, \ldots, x_n) = \mathrm{bar}(x_i; i = 1, \ldots, n)$. The proof of this convergence will be accomplished inductively together with the following crucial relation:

$$d(\mathrm{bar}_n(x_1, \ldots, x_n), \mathrm{bar}_n(x_1, \ldots, x_n)) \leq \frac{1}{n} \sum_{i=1}^n d(x_i, y_i). \tag{3}$$

First, for $n = 2$, the barycenter is already defined, and (3) reduces to (2). Now, assuming that we have shown the existence of the barycenter as well as inequality (3) for families of n points, let us consider a family $(x_1, \ldots, x_{n+1})$. For each $i \neq j$ in $\{1, \ldots, n + 1\}$, we

have

$$d(x_i^{(1)}, x_j^{(1)}) = d(\text{bar}_n(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{n+1}),$$

$$\text{bar}_n(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_{n+1})) \leq \frac{d(x_i, x_j)}{n}.$$

Therefore,

$$\text{diam}\{x_1^{(1)}, \ldots, x_{n+1}^{(1)}\} \leq \frac{1}{n} \text{diam}\{x_1, \ldots, x_{n+1}\},$$

and more generally, for all $k \geq 1$,

$$\text{diam}\{x_1^{(k)}, \ldots, x_{n+1}^{(k)}\} \leq \frac{1}{n^k} \text{diam}\{x_1, \ldots, x_{n+1}\}.$$

By this inequality and Lemma 2.1 below, the diameter of the convex closure of $\{x_1^{(k)}, \ldots, x_{n+1}^{(k)}\}$ converges to zero as k goes to infinity. Since $x_i^{(l)}$ belongs to this convex closure for all $l \geq k$, this shows that $\text{bar}_{n+1}(x_1, \ldots, x_{n+1})$ is well defined.

Next, take two families $(x_1, \ldots, x_{n+1})$ and $(y_1, \ldots, y_{n+1})$. By the inductive hypothesis, for each index $i \in \{1, \ldots, n+1\}$,

$$d(x_i^{(1)}, y_i^{(1)}) = d(\text{bar}_n(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{n+1}),$$

$$\text{bar}_n(y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_{n+1})) \leq \frac{1}{n} \sum_{j \neq i} d(x_j, y_j).$$

Summing over all $i = 1, \ldots, n+1$, this yields

$$\sum_{i=1}^{n+1} d(x_i^{(1)}, y_i^{(1)}) \leq \sum_{i=1}^{n+1} d(x_i, y_i).$$

More generally, for all $k \geq 1$,

$$\sum_{i=1}^{n+1} d(x_i^{(k)}, y_i^{(k)}) \leq \sum_{i=1}^{n+1} d(x_i^{(k-1)}, y_i^{(k-1)}) \leq \cdots \leq \sum_{i=1}^{n+1} d(x_i, y_i).$$

Letting k go to infinity, all the points $x_i^{(k)}$ (respectively $y_i^{(k)}$) converge to $\text{bar}_{n+1}(x_1, \ldots, x_{n+1})$ (respectively $\text{bar}_{n+1}(y_1 \ldots, y_{n+1})$). Hence, passing to the limit in the previous inequality, we obtain

$$(n+1) \, d(\text{bar}_{n+1}(x_1, \ldots, x_{n+1}), \text{bar}_{n+1}(y_1 \ldots, y_{n+1})) \leq \sum_{i=1}^{n+1} d(x_i, y_i),$$

as we wanted to show.

LEMMA 2.1. *The diameter of the convex closure of every bounded subset of X equals its own diameter.*

Proof. An explicit inductive description of the convex closure of a bounded subset B of X (i.e. the smallest convex subset of X containing B) proceeds as follows. Letting $B_0 := B$ and having defined $B_1, \ldots, B_n$, we let B_{n+1} be the union of all geodesics with endpoints in B_n. Then $B_n \subset B_{n+1}$, and the closure of the union $B_\infty := \bigcup_n B_n$ is the convex

closure of B. Since B_∞ contains B, we have $\mathrm{diam}(B_\infty) \geq \mathrm{diam}(B)$. To show the converse inequality, it suffices to show that for all $n \geq 0$,

$$\mathrm{diam}(B_{n+1}) \leq \mathrm{diam}(B_n). \tag{4}$$

To check this, given arbitrary points x, y in B_{n+1}, we may find x_0, x_1 and y_0, y_1 in B_n such that x (respectively y) lies in the geodesic joining x_0 and x_1 (respectively y_0 and y_1). The convexity of the distance along geodesics shows that

$$d(x, y_0) \leq \max\{d(x_0, y_0), d(x_1, y_0)\} \leq \mathrm{diam}(B_n),$$

$$d(x, y_1) \leq \max\{d(x_0, y_1), d(x_1, y_1)\} \leq \mathrm{diam}(B_n).$$

Another application of this convexity then shows that

$$d(x, y) \leq \max\{d(x, y_0), d(x, y_1)\} \leq \mathrm{diam}(B_n).$$

Since x, y were arbitrary points of B_{n+1}, this shows (4). $\square$

By the symmetry of the construction, for every permutation σ of $\{1, \ldots, n\}$,

$$\mathrm{bar}_n(x_1, \ldots, x_n) = \mathrm{bar}_n(x_{\sigma(1)}, \ldots, x_{\sigma(n)}).$$

With this in mind, (3) implies that

$$d(\mathrm{bar}_n(x_1, \ldots, x_n), \mathrm{bar}_n(y_1, \ldots, y_n)) \leq \frac{1}{n} \min_{\sigma \in S_n} \sum_{i=1}^{n} d(x_i, y_{\sigma(i)}).$$

The important observation here is that (by a theorem of Garrett Birkhoff; see [**8**, Introduction]) the right-hand-side expression above corresponds to the 1-Wasserstein distance between certain probability measures. More precisely,

$$\frac{1}{n} \min_{\sigma \in S_n} \sum_{i=1}^{n} d(x_i, y_{\sigma(i)}) = W_1(\mu_1, \mu_2),$$

where $\mu_1 := (1/n)(\delta_{x_1} + \cdots + \delta_{x_n})$ and $\mu_2 := (1/n)(\delta_{y_1} + \cdots + \delta_{y_n})$. In order to obtain a barycenter map that is 1-Lipschitz for the 1-Wasserstein metric, we would need to define the barycenter of $(1/n)(\delta_{x_1} + \cdots + \delta_{x_n})$ as $\mathrm{bar}_n(x_1, \ldots, x_n)$. However, such a definition is not intrinsic. For instance, though the n-set $(x_1, x_2, \ldots, x_n)$ and the $2n$-set $(x_1, x_1, x_2, x_2, \ldots, x_n, x_n)$ should be identified as measures, the points $\mathrm{bar}_n(x_1, x_2, \ldots, x_n)$ and $\mathrm{bar}_{2n}(x_1, x_1, x_2, x_2, \ldots, x_n, x_n)$ do not necessarily coincide. For example, the reader may easily check that for X a tripod of endpoints x, y, z and edges of the same length ℓ, the points $\mathrm{bar}_4(x, x, y, x)$ and $b_8(x, x, x, x, y, y, z, z)$ are different. (The former is at distance $7\ell/9$ from x, while the latter is at distance $2533\ell/3150$ from the same vertex; see Figure 1.)

To solve the problem above, we will slightly modify the definition of the barycenter of finite families of points so that it becomes invariant under the procedure—at the level of measures—of 'subdivision of mass along the atoms'. Given an arbitrary family $Q = (x_1, \ldots, x_n)$ of points in X, we let

$$Q^k := (x_1, \ldots, x_n, x_1, \ldots, x_n, \ldots, x_1, \ldots, x_n),$$
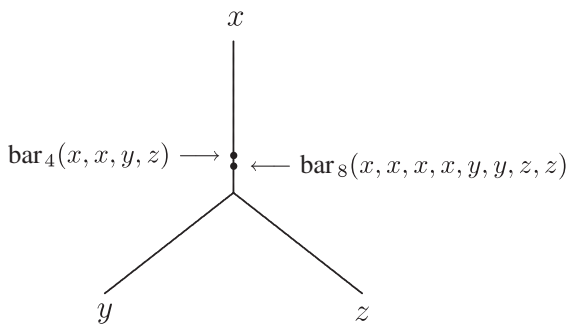
where the number of blocks is k.

FIGURE 1.

PROPOSITION 2.2. *The sequence of barycenters $\mathrm{bar}_{nk}(Q^k)$ is a Cauchy sequence.*

Assuming that this proposition holds, and since X is supposed to be complete, we may define the (canonical) barycenter

$$\mathrm{bar}^\star\left(\frac{1}{n}(\delta_{x_1} + \cdots + \delta_{x_n})\right)$$

as the limit point of the sequence $\mathrm{bar}_{nk}(Q^k)$. Indeed, one can easily check that this limit point depends only on the corresponding measure and not on any particular way of writing it as an equally weighted mean of Dirac measures (with not necessarily different atoms). Moreover, we still have the crucial relation

$$d\left(\mathrm{bar}^\star\left(\frac{1}{n}(\delta_{x_1} + \cdots + \delta_{x_n})\right), \mathrm{bar}^\star\left(\frac{1}{n}(\delta_{x_1} + \cdots + \delta_{x_n})\right)\right) \le \frac{1}{n} \min_{\sigma \in S_n} \sum_{i=1}^{n} d(x_i, y_{\sigma(i)}).$$

Thus, denoting by $P_\mathbb{Q}(X)$ the set of atomic probability measures on X all of whose atoms have rational mass, we have a well-defined map $\mathrm{bar}^\star : P_\mathbb{Q}(X) \to X$, and the previous inequality means that this map is 1-Lipschitz for the 1-Wasserstein metric: for all μ_1, μ_2 in $P_\mathbb{Q}(X)$,

$$d(\mathrm{bar}^\star(\mu_1), \mathrm{bar}^\star(\mu_2)) \le W_1(\mu_1, \mu_2). \tag{5}$$

If X is separable, then it is known that $P_\mathbb{Q}(X)$ is W_1-dense in $P^1(X)$. We may thus extend the map $\mathrm{bar}^\star$ to all $P^1(X)$ so that (5) holds for all μ_1, μ_2 in $P^1(X)$. This concludes our construction.

Remark 2.3. It is worth pointing out that for CAT(0)-spaces, $\mathrm{bar}^\star$ does not necessarily coincide with the Cartan barycenter. Indeed, for the example illustrated in Figure 1, the Cartan barycenter of the measure $\delta_x/2 + \delta_y/4 + \delta_z/4$ is the origin, though the barycenter $\mathrm{bar}^\star$ of this measure lies on the axis joining the origin to x (see the final remark of [**4**, Section I.2]).

To close this section, we next give a proof of Proposition 2.2. We observe that this proposition is also proved in [**4**] by means of a quite indirect argument that uses a deep martingale theorem and requires X to satisfy a weak local-compactness property.

Although this very elegant approach does not seem to be the most appropriate one in view of the purely geometric nature of the statement, the reader will still recognize a certain probabilistic flavor in our computations below. The key estimate for the distance between the barycenters of Q^k and Q^{k+l} is provided by the following lemma.

LEMMA 2.4. *For every $1/2 < \alpha < 2/3$, there exists a constant $C = C(\alpha) > 0$ and $L \gg 1$ such that for all positive integers l, k satisfying $L \leq l \leq \sqrt{k}$,*

$$d(\mathrm{bar}_{nk}(Q^k), \mathrm{bar}_{n(k+l)}(Q^{k+l})) \leq CD\frac{l^{3\alpha-1}}{k}, \tag{6}$$

where D denotes the diameter of the set $\{x_1, \ldots, x_n\}$. Moreover, for $0 \leq l \leq L$, there is the weaker estimate

$$d(\mathrm{bar}_{nk}(Q^k), \mathrm{bar}_{n(k+l)}(Q^{k+l})) \leq D\frac{l}{k}. \tag{7}$$

Assuming that this lemma holds, let us prove Proposition 2.2. Given $\varepsilon > 0$, fix an integer $k_\varepsilon \geq \max\{L, 10\}$ such that

$$\frac{D}{k_\varepsilon} + \frac{3^{3-3\alpha}CD}{(2-3\alpha)(k_\varepsilon - 1)^{2-3\alpha}} < \varepsilon,$$

where C is the constant provided by Lemma 2.4. For any $k_1 < k_2$ larger than k_ε, define the sequence (ℓ_j) by $\ell_1 := k_\varepsilon^2$ and $\ell_{j+1} := \ell_j + [\sqrt{\ell_j}]$. One can easily check by induction that $\ell_j \geq (k_\varepsilon + j)^2/9$ holds for all $j \geq 1$. Choose $m \geq 1$ such that $\ell_m < k_2 \leq \ell_{m+1}$. By Lemma 2.4,

$$d(\mathrm{bar}_{n\ell_j}(Q^{\ell_j}), \mathrm{bar}_{n\ell_{j+1}}(Q^{\ell_{j+1}})) \leq CD\frac{[\ell_{j+1} - \ell_j]^{3\alpha-1}}{\ell_j}$$

$$\leq \frac{CD}{\ell_j^{(3-3\alpha)/2}}, \quad j = 1, 2, \ldots, m-1.$$

Moreover,

$$d(\mathrm{bar}_{n\ell_m}(Q^{\ell_m}), \mathrm{bar}_{nk_2}(Q^{k_2})) \leq D\frac{[\ell_{m+1} - \ell_m]}{\ell_m} \leq \frac{D}{\ell_m^{1/2}}.$$

Using the triangle inequality, this yields

$$d(\mathrm{bar}_{nk_1}(Q^{k_1}), \mathrm{bar}_{nk_2}(Q^{k_2})) \leq \frac{D}{\ell_m^{1/2}} + \sum_{j=1}^{m-1} \frac{CD}{\ell_j^{(3-3\alpha)/2}}$$

$$\leq \frac{D}{k_\varepsilon} + \sum_{j=1}^{\infty} \frac{3^{3-3\alpha}CD}{(k_\varepsilon + j)^{3-3\alpha}}$$

$$\leq \frac{D}{k_\varepsilon} + 3^{3-3\alpha}CD \int_{k_\varepsilon-1}^{\infty} \frac{dx}{x^{3-3\alpha}}$$

$$\leq \frac{D}{k_\varepsilon} + \frac{3^{3-3\alpha}CD}{(2-3\alpha)(k_\varepsilon - 1)^{2-3\alpha}} < \varepsilon,$$

thus showing the Cauchy property.

It remains to prove Lemma 2.4. For this purpose we require the following lemma.

LEMMA 2.5. *Given integers $1 \le l < m$ and points $x, y_1, \ldots, y_m$ in X, the distance between x and $\mathrm{bar}_m(y_1, \ldots, y_m)$ is smaller than or equal to the mean distance between x and the points of the form $\mathrm{bar}_{m-l}(y_1, \ldots, \hat{y}_{i_1}, \ldots, \hat{y}_{i_l}, \ldots, y_m)$, where $i_1, \ldots, i_l$ range over all possible choices of different values in $\{1, \ldots, m\}$ (and each weight equals $m(m-1)\cdots(m-l+1) = m!/(m-l)!$).*

Proof. For $l = 1$, this follows as an application of (3) to

$$\mathrm{bar}_m(y_1, \ldots, y_m) = \mathrm{bar}_m(\mathrm{bar}_{m-1}(y_1, \ldots, \hat{y}_i, \ldots, y_m)); \quad i = 1, \ldots, m).$$

The general case easily follows by an inductive argument using (3) again. □

The idea of the proof of Lemma 2.4 consists in viewing the process of 'reduction of coordinates' for passing from Q^{k+l} to Q^k as a random process, which should imitate a Bernoulli trial for large values of $k \gg l$ (this process has a hypergeometric multivariate distribution). For each index j, the final associated error (i.e. the difference between l and the number of deleted entries x_j) should be—in mean—much smaller than ln. This allows us to pass from the elementary though useless upper bound $\sim Dl/k$ for the distance between the barycenters to the much better upper bound $\sim CDl^{3\alpha-1}/k$.

Proof of Lemma 2.4. As explained above, estimate (7) follows as a direct application of Lemma 2.5, so let us concentrate on (6). Lemma 2.5 again implies that the distance from $\mathrm{bar}_{nk}(Q^k)$ to $\mathrm{bar}_{n(k+l)}(Q^{k+l})$ is smaller than or equal to the mean of the distance between $\mathrm{bar}_{nk}(Q^k)$ and the points $\mathrm{bar}_{nk}(y_1, \ldots, y_{kn})$, where $(y_1, \ldots, y_{kn})$ ranges over all families that coincide with Q^{k+l} except for the deletion of ln entries. Among these families, the number of those for which the deleted entries correspond to an x_j-position a number of times equal to i_j (with $i_1 + \cdots + i_n = nl$) is

$$\binom{k+l}{i_1}\binom{k+l}{i_2}\cdots\binom{k+l}{i_n}.$$

Moreover, the distance from the barycenter of such a family to $\mathrm{bar}(Q^k)$ is smaller than or equal to

$$\frac{D}{kn}(|i_1 - l| + |i_2 - l| + \cdots + |i_n - l|).$$

By Lemma 2.5, this implies that $d(\mathrm{bar}_{nk}(Q^k), \mathrm{bar}_{n(k+l)}(Q^{k+l}))$ is smaller than or equal to

$$\frac{D}{kn} \sum_{i_1+\cdots+i_n=nl} \frac{\binom{k+l}{i_1}\binom{k+l}{i_2}\cdots\binom{k+l}{i_n}}{\binom{n(k+l)}{nl}}(|i_1 - l| + |i_2 - l| + \cdots + |i_n - l|)$$

$$= \frac{D}{k} \sum_{i=0}^{nl} \frac{\binom{k+l}{i}\binom{(n-1)(k+l)}{nl-i}}{\binom{n(k+l)}{nl}}|i - l|$$

$$= \frac{D}{k} \sum_{i=0}^{nl-l} \frac{\binom{k+l}{l+i}\binom{(n-1)(k+l)}{nl-l-i}}{\binom{n(k+l)}{nl}}i + \frac{D}{k} \sum_{i=0}^{l} \frac{\binom{k+l}{l-i}\binom{(n-1)(k+l)}{nl-l+i}}{\binom{n(k+l)}{nl}}i.$$

We will estimate the first of the two sums above, leaving to the reader the task of carrying out analogous computations for the second sum. First, notice that

$$
\frac{D}{k}\sum_{i=0}^{nl-l}\frac{\binom{k+l}{l+i}\binom{(n-1)(k+l)}{nl-l-i}}{\binom{n(k+l)}{nl}}i = \frac{D}{k}\sum_{i=0}^{l^\alpha}\frac{\binom{k+l}{l+i}\binom{(n-1)(k+l)}{nl-l-i}}{\binom{n(k+l)}{nl}}i + \frac{D}{k}\sum_{i=l^\alpha}^{nl-l}\frac{\binom{k+l}{l+i}\binom{(n-1)(k+l)}{nl-l-i}}{\binom{n(k+l)}{nl}}i
$$

$$
\leq \frac{Dl^\alpha}{k}\sum_{i=0}^{l^\alpha}\frac{\binom{k+l}{l+i}\binom{(n-1)(k+l)}{nl-l-i}}{\binom{n(k+l)}{nl}}
$$

$$
+ \frac{D(nl-l)}{k}\sum_{i=l^\alpha}^{nl-l}\frac{\binom{k+l}{l+i}\binom{(n-1)(k+l)}{nl-l-i}}{\binom{n(k+l)}{nl}}
$$

$$
\leq \frac{Dl^\alpha}{k} + \frac{D(n-1)l}{k}\left(1 - \sum_{i=0}^{l^\alpha}\frac{\binom{k+l}{l+i}\binom{(n-1)(k+l)}{nl-l-i}}{\binom{n(k+l)}{nl}}\right).
$$

The proof will then follow from an estimate of the form

$$
1 - \sum_{i=0}^{l^\alpha}\frac{\binom{k+l}{l+i}\binom{(n-1)(k+l)}{nl-l-i}}{\binom{n(k+l)}{nl}} \leq \frac{C}{l^{2-3\alpha}}. \tag{8}
$$

To show this, first rewrite

$$
\frac{\binom{k+l}{l+i}\binom{(n-1)(k+l)}{nl-l-i}}{\binom{n(k+l)}{nl}}
$$

$$
= \frac{\binom{k+l}{l}\binom{(n-1)(k+l)}{(n-1)l}}{\binom{n(k+l)}{nl}}\frac{k(k-1)\cdots(k-i+1)}{(l+1)(l+2)\cdots(l+i)}\frac{((n-1)l)((n-1)l-1)\cdots((n-1)l-i+1)}{((n-1)k)((n-1)k+1)\cdots((n-1)k+i)}.
$$

Now, using the improved version of Stirling's inequality (see [**2**, Ch. II.9])

$$
\sqrt{2\pi m}\left(\frac{m}{e}\right)^m e^{1/(12m+1)} \leq m! \leq \sqrt{2\pi m}\left(\frac{m}{e}\right)^m e^{1/12m},
$$

one can easily check that for a certain $e^{9/(12(l+1))} \leq \lambda \leq e^{9/12l}$,

$$
\frac{\binom{k+l}{l+i}\binom{(n-1)(k+l)}{(n-1)l-i}}{\binom{n(k+l)}{nl}} = \lambda\sqrt{\frac{(k+l)n}{2\pi kl(n-1)}} \geq \lambda\sqrt{\frac{n}{2\pi l(n-1)}}. \tag{9}
$$

On the other hand, choosing $L \gg 1$ and $c > 0$ such that $|\log(1+x) - x| \leq cx^2$ holds for all $|x| \leq 1/L^{2-2\alpha}$, for all $l \geq L$,

$$
\log\left(\frac{k(k-1)\cdots(k-i+1)}{(l+1)(l+2)\cdots(l+i)}\frac{((n-1)l)((n-1)l-1)\cdots((n-1)l-i+1)}{((n-1)k)((n-1)k+1)\cdots((n-1)k+i)}\right)
$$

$$
= \log\left(\frac{(1-(1/k))\cdots(1-((i-1)/k))}{(1+(1/l))(1+(2/l))\cdots(1+(i/l))}\right.
$$

$$
\left.\times\frac{(1-(1/(n-1)l))\cdots(1-(i-1)/(n-1)l)}{(1+(1/(n-1)k))\cdots(1+(i/(n-1)k))}\right)
$$

$$
\geq -\sum_{m=1}^{i-1}\frac{m}{(n-1)l} - \sum_{m=1}^{i}\frac{m}{l} - \sum_{m=1}^{i-1}\frac{m}{k} - \sum_{m=1}^{i}\frac{m}{(n-1)k} - 2c\left(\frac{i^3}{l^2}\right) - 2c\left(\frac{i^3}{k^2}\right)
$$

$$
\geq -\frac{i^2 n}{2l(n-1)} - \frac{i(n-2)}{2l(n-1)} - 4cl^{3\alpha-2}.
$$

Putting this together with (9) and using the inequality $1 - x \le e^{-x}$, one easily concludes that

$$\sum_{i=0}^{l^\alpha} \frac{\binom{k+l}{l+i}\binom{(n-1)(k+l)}{nl-l-i}}{\binom{n(k+l)}{nl}} \ge \left(1 - \frac{\bar{C}}{l^{2-3\alpha}}\right) \sum_{i=0}^{l^\alpha} \sqrt{\frac{n}{2\pi l(n-1)}} e^{-(i^2 n/2(n-1)l)}.$$

The series involved can obviously be compared with an integral:

$$\sum_{i=0}^{l^\alpha} \sqrt{\frac{n}{2\pi l(n-1)}} e^{-(i^2 n)/(2(n-1)l)}$$

$$= \sqrt{\frac{n}{2\pi(n-1)}} \sum_{i=0}^{l^\alpha} \frac{e^{-(i^2 n)/(2(n-1)l)}}{\sqrt{l}}$$

$$\ge \sqrt{\frac{n}{2\pi(n-1)}} \int_0^{l^\alpha} e^{-(x^2 n)/(2(n-1)l)} \, dx$$

$$\ge \frac{1}{\sqrt{2\pi}} \int_0^{l^\alpha \sqrt{n/((n-1)l)}} e^{-x^2/2} \, dx = 1 - \int_{l^\alpha \sqrt{n/((n-1)l)}}^{\infty} e^{-x^2/2} \, dx$$

$$\ge 1 - 2e^{-\ell^\alpha/2\sqrt{n/((n-1)l)}}.$$

Putting all of this together, one easily obtains (8), which concludes the proof. $\quad\square$

An application: a fixed point theorem. By construction, the map bar* is equivariant under the action of isometries. As a consequence, every action of a compact group by isometries of a Buseman space has a fixed point. Indeed, the push-forward of the Haar measure along an orbit is an invariant probability measure for the action. By equivariance, the barycenter bar* of this measure must remain fixed.

Despite the simple argument above, it is worth pointing out that a much stronger result holds: if a group action by isometries of a Buseman space has a (non-empty) compact invariant set, then it has a fixed point. (In particular, actions on a proper such space with bounded orbits must have fixed points.) Although the author was convinced that this was quite well known, according to the specialists it is apparently new, so we sketch the argument of the proof below (the details are left to the reader).

We will use the following construction. Given a compact subset B of X, we let B^* be the set of all midpoints between points of B whose distance realizes the diameter. By Lemma 2.1,

$$\text{diam}(B^*) \le \text{diam}(B) =: D.$$

Moreover, if equality holds, then there are points x_1, x_2, x_3, x_4 in B such that the distance between any of them equals D. Indeed, let y, z in B^* be such that $d(y, z) = D$. Let x_1, x_2 (respectively x_3, x_4) be points in B such that y (respectively z) is the midpoint between x_1 and x_2 (respectively x_3 and x_4) and $d(x_1, x_2) = d(x_3, x_4) = D$. Using

$$D = d(y, z) \le \frac{d(x_1, x_3)}{2} + \frac{d(x_2, x_4)}{2} \le D,$$

we conclude that $d(x_1, x_3) = d(x_2, x_4) = D$. Similarly, using

$$D = d(y, z) \leq \frac{d(x_1, x_4)}{2} + \frac{d(x_2, x_3)}{2} \leq D,$$

we conclude that $d(x_1, x_4) = d(x_2, x_3) = D$.

The preceding argument easily allows us to show the following generalization: starting with $B_1 := B$ of diameter D, define inductively $B_n := (B_{n-1})^*$. If $\mathrm{diam}(B_N) = D$, then there exist 2^N points $x_1, \ldots, x_{2^N}$ in B such that the distance between any of them equals D.

Assume now that Γ acts on X preserving a compact set $\hat{B}$. Compactness type arguments easily yield a compact invariant subset B of $\hat{B}$ of minimal diameter D. We claim that B is a single point (hence a fixed point for the action). Indeed, assume otherwise and cover B by finitely many (say, M) open balls of radius $D/2$. Since all the B_n are also compact and invariant, the minimality of D yields $\mathrm{diam}(B_n) = D$ for all $n \geq 1$. Fix N such that $2^N > M$. By the discussion above, there exists a sequence of points $x_1, \ldots, x_{2^N}$ in B such that the distance between any of them equals $D > 0$. However, this is impossible by the choice of N.

3. The L^1 ergodic theorem

To simplify, given $\varphi : \Omega \to X$, let us denote by

$$\mu_{n,\varphi}(\omega) := \frac{1}{m_G(F_n)} \int_{F_n} \delta_{\varphi(T^g \omega)} \, dm_G(g)$$

the nth *empirical measure* associated to φ. Notice that for all φ, ψ in $L^1(\mathcal{P}, X)$ and all $n \geq 1$,

$$\int_\Omega d(\mathrm{bar}^\star(\mu_{n,\varphi}(\omega)), \mathrm{bar}^\star(\mu_{n,\psi}(\omega))) \, d\mathcal{P}(\omega)$$

$$= \int_\Omega d\left(\mathrm{bar}^\star\left(\frac{1}{m_G(F_n)} \int_{F_n} \delta_{\varphi(T^g \omega)} \, dm_G(g) \right), \right.$$

$$\left. \mathrm{bar}^\star\left(\frac{1}{m_G(F_n)} \int_{F_n} \delta_{\psi(T^g \omega)} \, dm_G(g) \right) \right) d\mathcal{P}(\omega)$$

$$\leq \int_\Omega \frac{1}{m_G(F_n)} \int_{F_n} d(\varphi(T^g \omega), \psi(T^g \omega)) \, dm_G(g) \, d\mathcal{P}(\omega)$$

$$= \int_\Omega d(\varphi(\omega), \psi(\omega)) \, d\mathcal{P}(\omega),$$

hence

$$\int_\Omega d(\mathrm{bar}^\star(\mu_{n,\varphi}(\omega)), \mathrm{bar}^\star(\mu_{n,\varphi}(\omega))) \, d\mathcal{P}(\omega) \leq d_1(\varphi, \psi). \tag{10}$$

To prove the main theorem, let us first assume that φ takes values in a finite set, say $\{x_1, \ldots, x_k\}$, and let Ω_i be the preimage of $\{x_i\}$ under φ. A direct application of Lindenstrauss's ergodic theorem [3] to the characteristic function of Ω_i yields the existence almost everywhere of the following limit:

$$\lambda_i(\omega) := \lim_{n \to \infty} \frac{m_G(\{g \in F_n : T^g \omega \in \Omega_i\})}{m_G(F_n)}.$$

We claim that almost surely we have the convergence

$$\text{bar}^\star(\mu_{n,\varphi}) \longrightarrow \text{bar}^\star\left(\sum_{i=1}^{k} \lambda_i(\omega)\delta_{x_i}\right). \tag{11}$$

Indeed, since $\text{bar}^\star$ is 1-Lipschitz for W_1, given $\varepsilon > 0$ we have that for almost every $\omega \in \Omega$ there exists $n(\omega, \varepsilon) \geq 1$ such that for all $n \geq n(\omega, \varepsilon)$,

$$d\left(\text{bar}^\star(\mu_{n,\varphi}), \text{bar}^\star\left(\sum_{i=1}^{k} \lambda_i(\omega)\delta_{x_i}\right)\right)$$

$$\leq W_1\left(\sum_{i=1}^{k} \frac{m_G(\{g \in F_n : T^g\omega \in \Omega_i\})}{m_G(F_n)}\delta_{x_i}, \sum_{i=1}^{k} \lambda_i(\omega)\delta_{x_i}\right)$$

$$\leq \sum_{i=1}^{k}\left|\frac{m_G(\{g \in F_n : T^g\omega \in \Omega_i\})}{m_G(F_n)} - \lambda_i(\omega)\right|\text{diam}\{x_1, \ldots, x_k\}$$

$$\leq \varepsilon.$$

This shows the convergence (11). Now notice that by construction, both $\text{bar}^\star(\mu_{n,\varphi})$ and $\text{bar}^\star(\sum_{i=1}^{k} \lambda_i(\omega)\delta_{x_i})$ belong to the convex closure of $\{x_1, \ldots, x_k\}$. By Lemma 2.1, this implies that for all $n \geq 1$, the distance between these two points is less than or equal to $\text{diam}\{x_1, \ldots, x_k\}$. A direct application of the dominated convergence theorem then shows that the convergence (11) also holds in $L^1(\mathcal{P}, X)$.

In order to deal with the general case we will need the next lemma.

LEMMA 3.1. *There exists a constant $C > 0$ (depending only on the sequence (F_n)) such that for all φ, ψ in $L^1(X, \mu)$ and all $\lambda > 0$,*

$$\mathcal{P}[\omega \in \Omega : \sup_{n \geq 1} d(\text{bar}^\star(\mu_{n,\varphi}(\omega)), \text{bar}^\star(\mu_{n,\psi}(\omega))) \geq \lambda] \leq \frac{C}{\lambda}d_1(\varphi, \psi). \tag{12}$$

Proof. Since $\text{bar}^\star$ is 1-Lipschitz for W_1, the set involved in the inequality above is contained in $\{\omega \in \Omega : \sup_{n \geq 1} W_1(\mu_{n,\varphi}, \mu_{n,\psi}) \geq \lambda\}$. Now, noticing that the measure

$$\nu_n := \frac{1}{m_G(F_n)} \int_{F_n} \delta_{(\varphi(T^g\omega), \psi(T^g\omega))} \, dm_G(g)$$

lies in $(\mu_{n,\varphi}|\mu_{n,\psi})$, we obtain

$$W_1(\mu_{n,\varphi}, \mu_{n,\psi}) \leq \frac{1}{m_G(F_n)} \int_{F_n} d(\varphi(T^g\omega), \psi(T^g\omega)) \, dm_G(g).$$

Thus, the left-hand-side expression in (12) is smaller than or equal to

$$\mathcal{P}\left[\omega \in \Omega : \sup_{n \geq 1} \frac{1}{m_G(F_n)} \int_{F_n} d(\varphi(T^g\omega), \psi(T^g\omega)) \, dm_G(g) \geq \lambda\right].$$

Now, a direct application of Lindenstrauss's maximal ergodic theorem (see [3, Theorem 3.2]) yields the existence of a constant $C > 0$ (depending only on (F_n)) such that this last probability is smaller than or equal to

$$\frac{C}{\lambda} \int_{\Omega} d(\varphi(\omega), \psi(\omega)) \, d\mathcal{P}(\omega),$$

as desired. $\qquad\square$

We may now complete the proof of the main theorem. Since X is assumed to be separable, for each $\varphi \in L^1(\mathcal{P}, X)$ there exists a sequence of finite-valued functions $\varphi_k : \Omega \to X$ that converges to φ in the L^1 sense. Thus, given $\varepsilon > 0$, we may fix $\psi := \varphi_{k_\varepsilon}$ such that $d_1(\varphi, \psi) \le \varepsilon^2$. By (12),

$$\mathcal{P}\left[\omega \in \Omega : \sup_{n \ge 1} d(\text{bar}^\star(\mu_{n,\varphi}(\omega)), \text{bar}^\star(\mu_{n,\psi}(\omega))) \ge \varepsilon\right] \le \frac{C}{\varepsilon} d_1(\varphi, \psi) \le C\varepsilon.$$

Since $\text{bar}^\star(\mu_{n,\psi})$ is known to converge almost everywhere, this inequality implies that on a set of measure at least $1 - C\varepsilon$, the sequence $(\text{bar}^\star(\mu_{n,\varphi}(\omega)))$ asymptotically oscillates by at most 2ε. Since this is true for all $\varepsilon > 0$, this shows that $\text{bar}^\star(\mu_{n,\varphi}(\omega))$ converges almost surely.

Finally, to show the convergence in $L^1(\Omega, X)$, just notice that by (10),

$$\int_\Omega d(\text{bar}^\star(\mu_{n,\varphi}(\omega)), \text{bar}^\star(\mu_{m,\varphi}(\omega)))\, d\mathcal{P}(\omega)$$
$$\le \int_\Omega [d(\text{bar}^\star(\mu_{n,\varphi}), \text{bar}^\star(\mu_{n,\varphi_k})) + d(\text{bar}^\star(\mu_{n,\varphi_k}), \text{bar}^\star(\mu_{m,\varphi_k}))$$
$$+ d(\text{bar}^\star(\mu_{m,\varphi_k}), \text{bar}^\star(\mu_{m,\varphi}))]\, d\mathcal{P}(\omega)$$
$$\le 2d_1(\varphi, \varphi_k) + \int_\Omega d(\text{bar}^\star(\mu_{n,\varphi_k}), \text{bar}^\star(\mu_{m,\varphi_k}))\, d\mathcal{P}(\omega).$$

For a given $\varepsilon > 0$, we may fix k large enough so that $d_1(\varphi, \varphi_k) \le \varepsilon/3$. Since $\text{bar}^\star(\mu_{n,\varphi_k})$ converges in $L^1(\mathcal{P}, X)$ as n goes to infinity, we may fix n_ε so that for all n, m larger than n_ε,

$$\int_\Omega d(\text{bar}^\star(\mu_{n,\varphi_k}), \text{bar}^\star(\mu_{m,\varphi_k}))\, d\mathcal{P}(\omega) \le \frac{\varepsilon}{3}.$$

Putting all of this together, we obtain that for all n, m larger than n_ε,

$$\int_\Omega d(\text{bar}^\star(\mu_{n,\varphi}(\omega)), \text{bar}^\star(\mu_{m,\varphi}(\omega)))\, d\mathcal{P}(\omega) \le \varepsilon.$$

Hence, $\text{bar}^\star(\varphi_{n,\varphi})$ is a Cauchy sequence in $L^1(\mathcal{P}, X)$, as we wanted to show.

REFERENCES

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

[2] W. Feller. *An Introduction to Probability Theory and its Applications, Vol. 1 (Wiley Series in Probability and Mathematical Statistics)*. Wiley, New York, 1950.

[3] E. Lindenstrauss. Pointwise theorems for amenable groups. *Invent. Math.* **146** (2001), 259–295.

[4]   A. Es-Sahib and H. Heinich. Barycentre canonique pour un espace métrique à courbure négative. *Séminaire de Probabilités XXXIII (Lecture Notes in Mathematics, 1709)*. Springer, Berlin, 1999, pp. 355–370.

[5]   J. Jost. *Nonpositive curvature: geometric and analytic aspects (Lectures in Mathematics)*. ETH Zürich, 1997.

[6]   A. Karlsson and F. Ledrappier. Noncommutative ergodic theorems. *Geometry, Rigidity, and Group Actions (Chicago Lectures in Mathematics)*. Eds. B. Farb and D. Fisher. University of Chicago Press, Chicago, IL, 2011, pp. 396–418.

[7]   K.-T. Sturm. Probability measures on metric spaces of nonpositive curvature. *Heat Kernels and Analysis on Manifolds, Graphs, and Metric Spaces (Contemporary Mathematics, 338)*. American Mathematical Society, Providence, RI, 2003, pp. 357–390.

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