

ERGODIC THEOREM FOR NONSTATIONARY RANDOM WALKS ON COMPACT ABELIAN GROUPS

GRIGORII MONAKOV

ABSTRACT. We consider a nonstationary random walk on a compact metrizable abelian group. Under a classical strict aperiodicity assumption we establish a weak-* convergence to the Haar measure, Ergodic Theorem and Large Deviation Type Estimate.

1. INTRODUCTION

1.1. Main result. Consider a topological compact metrizable abelian group $(\mathbb{A}, 0, +)$ with shift-invariant metric¹ d :

$$d(x + \alpha, y + \alpha) = d(x, y) \quad \forall \alpha, x, y \in \mathbb{A}.$$

Let $\mathcal{B}$ denote the Borel σ -algebra. It is well-known [Ha, Chapter XI] that there exists a unique shift invariant Borel probability measure on $\mathbb{A}$ that is usually referred to as Haar measure. Let us denote this measure by h .

Denote the space of all probability measures on $(\mathbb{A}, \mathcal{B})$ by $\mathcal{M}$. We define convolution of two measures $\mu, \nu \in \mathcal{M}$ by the following formula:

$$(1.1) \quad (\mu * \nu)(B) = \int_{\mathbb{A}} \mathbf{1}_B(\alpha + \beta) d\mu(\beta) d\nu(\alpha) \quad \forall B \in \mathcal{B}.$$

One can view the measure $\mu * \nu$ as a distribution of the image of a random point $\alpha \in \mathbb{A}$, distributed with respect to ν , shifted by an independent random element $\beta \in \mathbb{A}$, distributed with respect to μ .

We will need the following **strict aperiodicity** condition (see, for instance, [Bol]):

Definition 1.1. We say that a random shift defined by a measure μ is **strictly aperiodic** if the smallest semigroup containing $\{\alpha - \beta \mid \alpha, \beta \in \text{supp}(\mu)\}$ is dense in $\mathbb{A}$ (we denote by $\text{supp}(\mu)$ the topological support of the measure μ , i. e. the smallest closed subset of a full measure).

It is not hard to see that the condition above is equivalent to $\text{supp}(\mu)$ not being contained in a coset of any proper closed (normal) subgroup of $\mathbb{A}$.

The main result of this paper can be stated as follows:

Theorem 1.2. *Consider a compact (in weak-* topology) $K \subset \mathcal{M}$, such that any $\mu \in K$ is **strictly aperiodic**. Let $\mu_1, \mu_2, \mu_3, \dots$ be a sequence of measures in K .*

2010 *Mathematics Subject Classification.* Primary: 37H05, 37B05, 60J05.

¹It is worth noticing that the statement that all compact separable abelian groups of real power are metrizable can be neither proved nor disproved in ZFC [H]. However, it is well-known (see [Bi] and [K]) that any topological group is metrizable if and only if it is first countable, and, moreover, the metric can always be taken to be left-invariant.

Then for any $\varphi \in C(\mathbb{A}, \mathbb{R})$ and any $x \in \mathbb{A}$ almost surely

$$(1.2) \quad \frac{1}{n} \sum_{k=0}^{n-1} \varphi(x + \alpha_1 + \alpha_2 + \dots + \alpha_k) \rightarrow \int_{\mathbb{A}} \varphi dh \quad \text{as } n \rightarrow \infty,$$

where the elements $\alpha_1, \alpha_2, \dots$ are independent random elements in $\mathbb{A}$ and α_j is distributed with respect to μ_j .

Moreover, an analogue of the Large Deviations Theorem holds. Namely, for any $\varepsilon > 0$ there exist $C, \delta > 0$ such that for any $x \in \mathbb{A}$ and any $n \geq 1$

$$(1.3) \quad \mathbb{P} \left(\left| \frac{1}{n} \sum_{k=0}^{n-1} \varphi(x + \alpha_1 + \alpha_2 + \dots + \alpha_k) - \int_{\mathbb{A}} \varphi dh \right| > \varepsilon \right) < C \exp(-\delta n),$$

where

$$\mathbb{P} = \prod_{n=1}^{\infty} \mu_n.$$

Some classical examples of abelian groups to which our results can be applied include flat tori, p -adic integers, and Cantor dyadic group. Additional motivation for studying stationary and nonstationary random shifts on tori comes from the recent interest to dynamics of mixed random-quasiperiodic cocycles, see [CDK1, CDK2].

Remark 1.3. The proof presented in this paper relies heavily on the commutativity of $\mathbb{A}$. Nevertheless, we expect a similar result to hold in a noncommutative case as well and plan to study it in future publications.

1.2. Random and nonstationary ergodic theorems. Ergodic Theorems and Large Deviation Type Estimates have been intensively studied for both stationary and nonstationary stochastic dynamical systems. Let us mention some classical and recent results in that area.

Let X be a compact metric space and μ a Borel probability measure on the space of continuous functions $C(X, X)$. For any Borel probability measure ν on X we can define its convolution with μ by

$$(1.4) \quad (\mu * \nu)(A) = \int_{C(X, X)} \nu(f^{-1}(A)) d\mu(f)$$

for any Borel set $A \subset X$. The resulting measure $\mu * \nu$ can be understood in the following way: it is the distribution of $f(x)$, where $x \in X$ is distributed with respect to measure ν , the map $f \in C(X, X)$ is distributed with respect to the measure μ and f and x are chosen independently from one another. Notice that formula (1.1) is a special case of (1.4).

A probability measure ν on X is called stationary if $\mu * \nu = \nu$. A stationary measure always exists. For instance, one can observe that any limit point of the sequence

$$\frac{1}{n} \sum_{j=0}^{n-1} \mu^{*j} * \nu$$

will be a stationary measure (for details see [BQ, Lemma 2.10]). Now the Birkhoff Ergodic Theorem for stationary stochastic dynamical systems can be formulated as follows [DMPS, Theorem 5.2.9]:

Theorem 1.4. *Let ν be any μ -stationary ergodic (extremal point of the set of μ -stationary measures) measure on X and $\varphi \in C(X, \mathbb{R})$ be a continuous function. Then for ν -almost every point $x \in X$ and $\mu^{\mathbb{N}}$ -almost every sequence $f_1, f_2, \dots \in C(X, X)$ one has*

$$\frac{1}{n} \sum_{k=0}^{n-1} \varphi(f_k \circ f_{k-1} \circ \dots \circ f_1(x)) \rightarrow_{n \rightarrow \infty} \int_X \varphi(x) d\nu(x).$$

Large Deviation Type Estimates are also well studied for the stationary case. For that setting it turns out to be useful to operate in terms of Markov processes instead of stochastic dynamical systems. Some of the results can be found in [V, A, DK] and references therein. In their recent paper Cai, Duarte and Klein obtained an abstract Large Deviations Theorem (see Theorem 1.1 and Remark 2.3 in [CDK3]). One application of this result is a Large Deviation Theorem for random translations on the torus (see Theorem 1.2). Although this result is similar to our Theorem 1.2, it provides a more accurate estimate for the probability of large deviations in a more restrictive setting and uses completely different methods. In classical Probability Theory Large Deviation Type Estimates are also known for nonstationary sequences of independent random variables. This fact is often referred to as Gärtner-Ellis theorem (see, for example [DZ]).

For the nonstationary deterministic setting some results are also known (sometimes such systems are referred to as *sequential*). For instance, in [BB] authors deal with the situation when all transformations have a common invariant measure, in [CR] authors consider nonstationary sequences of piecewise expanding transformations of $[0, 1]$, and in [GBK] authors generalize Krylov-Bogolyubov theorem and other classical results from ergodic theory of smooth dynamical systems for a non autonomous (non stationary) dynamical system, defined by a sequence of uniformly convergent maps.

Ergodic Theorem for random nonstationary systems has also been studied from the point of view of Markov processes (and is referred to as strong law of large numbers). For various results on that topic see [G, LL1, LL2, LY, Y1, Y2, ZYW]. Large Deviation Type Estimates for inhomogeneous Markov chains can be found in [DS]. A recent monograph by Dolgopyat and Sarig [DSa] investigates various limit theorems including Large Deviations Theorem for nonstationary Markov processes (see also references therein). Finally, an abstract version of ergodic theorem that we use as a black box to establish our result (see Theorem 2.4 below) was recently obtained by Gorodetski and Kleptsyn in [GK].

1.3. Random walks on groups. The literature concerning random walks on groups is vast and goes far beyond the scope of our paper, for example, one can see surveys [BQ, BL, Fu] and references therein. In this section we will concentrate on random walks on compact groups only. Limit theorems for this case have been studied intensively since 1940s. For a detailed survey of results about random walks on flat tori see [DF]. Generalizing results of Lévy [L] for a random walk on $\mathbb{R}/\mathbb{Z}$ several convergence results for the sequence μ^{*n} were obtained in pioneering works of Kawada and Itô [KI], Kawada [Ka], Prékopa, Rényi and Urbanik [PRU] and Urbanik [U]. In the work [KI] by Kloss one will find a uniform stationary version of Ergodic Theorem and first results on the weak-* convergence of a non-stationary sequence $\mu_n * \mu_{n-1} * \dots * \mu_1$. In [S] Stromberg proves the most general result about weak-* convergence in the stationary case (see Remark 2.6 below). In

[BE] it is proved (in noncommutative case) that Ergodic Theorem for a stationary random walk is equivalent to the smallest closed subgroup containing $\text{supp}(\mu)$ being equal to the whole group (such measures μ are usually referred to as *adapted*). Results about convergence of μ^{*n} in the sense of total variation can be found in [B, RX, AG]. Finally, Borda in the recent works [Bo1, Bo2] established the Central Limit Theorem and Law of Iterated Logarithm for the stationary random walks on groups. The works mentioned above also include several nonstationary forms of the ergodic theorem. However, all those theorems impose some additional Doeblin-type conditions on measures $\mu_1, \mu_2, \dots, \mu_n$, which is not required for our approach.

Our proof of Theorem 1.2 consists of two independent parts: we prove that under suitable conditions the sequence of measures $\mu_n * \mu_{n-1} * \dots * \mu_1$ converges (uniformly with respect to the choice of $\mu_1, \mu_2, \dots, \mu_n$) to h in weak-* topology (see Theorem 2.5), and then apply a recent result by Gorodetski and Kleptsyn (see Theorem 2.4) to deduce the ergodic theorem. Even though the weak-* convergence in Theorem 2.5 is similar in spirit to the classical theorem by Kawada and Itô (see [KI, Theorem 8]), our proof uses a completely different technique. In their work Kawada and Itô used the machinery of representation theory to prove weak-* convergence in the stationary case. However, we don't know if these methods can be pushed further to establish a uniform nonstationary result like the one in Theorem 2.5.

2. PROOF OF THE MAIN RESULT

Recently, A. Gorodetski and V. Kleptsyn proved a nonstationary analog of the stochastic version of Birkhoff Ergodic Theorem [GK]. In order to state it we would first need to recall the definition of the Wasserstein metric. Let X be a compact metric space and $\mathcal{M}$ be the space of Borel probability measures on X . Then the Wasserstein metric can be defined as follows:

Definition 2.1. Let ν_1, ν_2 be two probability measures in $\mathcal{M}$. Then the *Wasserstein distance* between them is defined as

$$W(\nu_1, \nu_2) = \inf_{\gamma} \iint_{X \times X} d(x, y) d\gamma(x, y),$$

where the infimum is taken over all probability measures γ on $(X \times X, \sigma(\mathcal{B} \times \mathcal{B}))$ with the marginals (projections on the x and y coordinates) $P_x(\gamma) = \nu_1$ and $P_y(\gamma) = \nu_2$.

Remark 2.2. It is well known (see, for example [Vi, Theorem 6.9]) that convergence in Wasserstein metric is equivalent to the one in weak-* topology.

Remark 2.3. Another important property of Wasserstein metric that we will use is the following convexity:

$$W(t\nu_1 + (1-t)\nu_2, \nu_3) \leq tW(\nu_1, \nu_3) + (1-t)W(\nu_2, \nu_3) \quad \forall \nu_1, \nu_2, \nu_3 \in \mathcal{M}, \quad \forall t \in (0, 1).$$

This statement follows directly from Definition 2.1.

Let us also introduce the following

Standing Assumption: Let X be a compact metric space. We will say that a sequence of distributions $\mu_1, \mu_2, \mu_3, \dots$ on $C(X, X)$ satisfies the *Standing Assumption* if for any $\delta > 0$ there exists $m \in \mathbb{N}$ such that the images of any two initial

measures after averaging over m random steps after any initial moment n become δ -close to each other in terms of Wasserstein distance:

$$W(\mu_{n+m} * \dots * \mu_{n+1} * \nu, \mu_{n+m} * \dots * \mu_{n+1} * \nu') < \delta$$

for any probability measures ν and ν' on X and any $n \in \mathbb{N}$.

Now the nonstationary stochastic version of Birkhoff Ergodic Theorem can be formulated as follows :

Theorem 2.4 ([GK, Theorem 3.1]). *Let X be a compact metric space. Suppose the sequence of distributions $\mu_1, \mu_2, \mu_3, \dots$ satisfies the **Standing Assumption** above. Given any Borel probability measure ν_0 on X , define*

$$\nu_n := \mu_n * \nu_{n-1}, \quad n = 1, 2, \dots$$

Then for any $\varphi \in C(X, \mathbb{R})$ and any $x \in X$, almost surely

$$(2.1) \quad \frac{1}{n} \left| \sum_{k=0}^{n-1} \varphi(f_k \circ \dots \circ f_1(x)) - \sum_{k=0}^{n-1} \int_X \varphi d\nu_k \right| \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Moreover, an analogue of the Large Deviations Theorem holds. Namely, for any $\varepsilon > 0$ there exist $C, \delta > 0$ such that for any $x \in X$ and any $n \geq 1$

$$(2.2) \quad \mathbb{P} \left(\frac{1}{n} \left| \sum_{k=0}^{n-1} \varphi(f_k \circ \dots \circ f_1(x)) - \sum_{k=0}^{n-1} \int_X \varphi d\nu_k \right| > \varepsilon \right) < C \exp(-\delta n),$$

where

$$\mathbb{P} = \prod_{n=1}^{\infty} \mu_n.$$

Choosing $X = \mathbb{A}$ we see that to prove Theorem 1.2 it is enough to verify the **Standing Assumption** for the sequence $\mu_1, \mu_2, \mu_3, \dots$, which brings us to the next

Theorem 2.5. *Consider a compact (in weak-* topology) $K \subset \mathcal{M}$, such that any $\mu \in K$ is **strictly aperiodic**. Then for any $\varepsilon > 0$ there exists $m \in \mathbb{N}$, such that for any sequence of measures $\mu_1, \mu_2, \dots, \mu_m \in K$ and any starting measure $\nu \in \mathcal{M}$ the following inequality holds:*

$$(2.3) \quad W(\mu_m * \mu_{m-1} * \dots * \mu_1 * \nu, h) < \varepsilon.$$

Remark 2.6. A classical theorem by Stromberg [S] states that in a stationary case a weak-* limit $\lim_{n \rightarrow \infty} \mu^{*n}$ exists if and only if μ is **strictly aperiodic**. Moreover, said limit is always equal to the Haar measure h . It can be easily seen from Stromberg's result that Theorem 2.5 will not hold if there is a not **strictly aperiodic** $\mu \in K$. In fact, a simple counterexample can be constructed for $\mathbb{A} = \mathbb{R}/\mathbb{Z}$: consider $\mu = \delta_\alpha$ for some irrational α . Despite the fact that $\bigcup_{n \in \mathbb{N}} \text{supp}(\mu^{*n})$ is dense in $\mathbb{R}/\mathbb{Z}$, there is still no weak-* limit for μ^{*n} .

The compactness of K is also essential. For the same $\mathbb{A} = \mathbb{R}/\mathbb{Z}$ fix some irrational $\alpha \in (0, 1)$ and define

$$\mu_n = \frac{1}{2} \delta_0 + \frac{1}{2} \delta_{\alpha/2^n}.$$

Even though every μ_n is **strictly aperiodic** we have

$$\text{supp}(\mu_n * \mu_{n-1} * \dots * \mu_1) \subset [0, \alpha],$$

hence the weak-* limit of $\mu_n * \mu_{n-1} * \dots * \mu_1$ cannot be equal to $h = \text{Leb}|_{[0,1]}$.

Postponing the proof of Theorem 2.5 until the next section we would like to deduce Theorem 1.2:

Proof of Theorem 1.2. It follows from inequality (2.3) that for any sequence $\mu_1, \mu_2, \mu_3 \dots \in K$ the **Standing Assumption** holds. Applying Theorem 2.4 we obtain that statements (2.1) and (2.2) hold. Since convergence in Wasserstein metric is equivalent to weak-* convergence inequality (2.3) also implies that

$$\frac{1}{n} \sum_{k=0}^{n-1} \int_{\mathbb{A}} \varphi d\nu_k \rightarrow \int_{\mathbb{A}} \varphi dh.$$

After that formulae (1.2) and (1.3) follow directly from (2.1) and (2.2). $\square$

3. PROOF OF THEOREM 2.5: WEAK-* CONVERGENCE TO THE HAAR MEASURE

We will denote by $B(x, r)$ an open ball in metric d of radius r centered at x . We would like to start with a following

Definition 3.1. We say that a set $Q \in \mathcal{B}$ is ε -wide if it is contained in a ball of radius ε and contains a ball of radius $\varepsilon/3$.

Remark 3.2. Note that it follows directly from Vitali covering lemma [Vit] that any compact metric space can be partitioned into finitely many ε -wide sets for any fixed $\varepsilon > 0$.

Let us postpone the proof of the following technical statement until Section 4:

Proposition 3.3. *Let K be a compact set in $\mathcal{M}$, such that any $\mu \in K$ is **strictly aperiodic**. For any $\varepsilon > 0$ there exists $m \in \mathbb{N}$ and $\delta > 0$, such that for any $\mu_1, \mu_2, \dots, \mu_m \in K$, any $\nu \in \mathcal{M}$ and any ε -wide set Q we have*

$$(3.1) \quad [\mu_m * \mu_{m-1} * \dots * \mu_1 * \nu](Q) \geq \delta.$$

Before proving Theorem 2.5 let us establish the following

Lemma 3.4. *For any $\mu, \nu \in \mathcal{M}$ one has*

$$(3.2) \quad W(\mu * \nu, h) \leq W(\nu, h).$$

Proof of Lemma 3.4. Let γ be a probability measure on $\mathbb{A} \times \mathbb{A}$, such that $P_x(\gamma) = \nu$ and $P_y(\gamma) = h$. Define a measure

$$d\tilde{\gamma}(x, y) = \int_{\mathbb{A}} d\gamma(x - \alpha, y - \alpha) d\mu(\alpha).$$

It is clear that $P_x(\tilde{\gamma}) = \mu * \nu$ and $P_y(\tilde{\gamma}) = h$. Now

$$\begin{aligned} \iint_{\mathbb{A} \times \mathbb{A}} d(x, y) d\tilde{\gamma}(x, y) &= \\ &= \iiint_{\mathbb{A} \times \mathbb{A} \times \mathbb{A}} d(x, y) d\gamma(x - \alpha, y - \alpha) d\mu(\alpha) = \\ &= \iiint_{\mathbb{A} \times \mathbb{A} \times \mathbb{A}} d(x + \alpha, y + \alpha) d\gamma(x, y) d\mu(\alpha) = \\ &= \iiint_{\mathbb{A} \times \mathbb{A} \times \mathbb{A}} d(x, y) d\gamma(x, y) d\mu(\alpha) = \\ &= \iint_{\mathbb{A} \times \mathbb{A}} d(x, y) d\gamma(x, y) \end{aligned}$$

and inequality (3.2) follows. $\square$

Proof of Theorem 2.5. Let us fix $\varepsilon > 0$ and $\nu \in \mathcal{M}$. We will first show that for $m \in \mathbb{N}$ and δ satisfying (3.1) the measure $\nu^{(m)} := \mu_m * \mu_{m-1} * \dots * \mu_1 * \nu$ can be decomposed into a sum

$$(3.3) \quad \nu^{(m)} = \nu_0 + \nu_1,$$

where $\nu_1(\mathbb{A}) = \delta$ and

$$(3.4) \quad W\left(\frac{\nu_1}{\nu_1(\mathbb{A})}, h\right) \leq \varepsilon.$$

In order to construct that decomposition we fix a partition of $\mathbb{A}$ into ε -wide sets $\{Q_j\}_{1 \leq j \leq k}$. Next, define a measure

$$d\nu_1(x) = \sum_{j=1}^k \frac{\delta h(Q_j)}{\nu^{(m)}(Q_j)} \mathbf{1}_{Q_j}(x) d\nu^{(m)}(x).$$

Notice that according to (3.1) we have $\frac{\delta h(Q_j)}{\nu^{(m)}(Q_j)} < \frac{\delta}{\nu^{(m)}(Q_j)} \leq 1$, so $\nu_0 = \nu^{(m)} - \nu_1$ is a well-defined positive measure and $\nu_1(Q_j) = \delta h(Q_j)$. Clearly, $\nu_1(\mathbb{A}) = \delta$. Now let us show that (3.4) holds. We would like to construct a probability measure γ on $\mathbb{A} \times \mathbb{A}$, such that

$$(3.5) \quad \text{supp}(\gamma) \subset \cup_{j=1}^k Q_j \times Q_j,$$

$P_x(\gamma) = \frac{\nu_1}{\delta}$, and $P_y(\gamma) = h$. It immediately follows from (3.5) that

$$\begin{aligned} \iint_{\mathbb{A} \times \mathbb{A}} d(x, y) d\gamma(x, y) &= \sum_{j=1}^k \iint_{Q_j \times Q_j} d(x, y) d\gamma(x, y) \leq \\ &\leq \varepsilon \sum_{j=1}^k \gamma(Q_j \times Q_j) = \varepsilon. \end{aligned}$$

The construction of required γ can be done in the following way: for each Q_j define

$$\gamma|_{Q_j \times Q_j} = \frac{\nu_1|_{Q_j} \otimes h|_{Q_j}}{h(Q_j)}.$$

It's easy to verify that γ satisfies required properties and hence (3.4) holds.

Now we can apply the same reasoning to the measure ν_0 . Namely, after applying another m iterations of our random walk we will be able to decompose $\mu_{2m} * \mu_{2m-1} * \dots * \mu_{m+1} * \nu_0$ the same way as in (3.3) (up to scaling). That will give the following decomposition for $\nu^{(2m)} := \mu_{2m} * \mu_{2m-1} * \dots * \mu_1 * \nu$:

$$\nu^{(2m)} = \nu_0^{(2)} + \nu_1^{(2)} + \nu_2^{(2)},$$

where $\nu_1^{(2)} = \mu_{2m} * \mu_{2m-1} * \dots * \mu_{m+1} * \nu_1$, $\nu_2^{(2)}(\mathbb{A}) = \delta(1 - \delta)$ and

$$W\left(\frac{\nu_2^{(2)}}{\nu_2^{(2)}(\mathbb{A})}, h\right) \leq \varepsilon.$$

In addition, combining Lemma 3.4 and inequality (3.4) we get

$$W\left(\frac{\nu_1^{(2)}}{\nu_1^{(2)}(\mathbb{A})}, h\right) \leq \varepsilon.$$

Repeating this procedure $r > \log_{1-\delta}(\varepsilon)$ times we will obtain a decomposition

$$\nu^{(mr)} = \nu_0^{(r)} + \nu_1^{(r)} + \dots + \nu_r^{(r)},$$

such that

$$(3.6) \quad W\left(\frac{\nu_j^{(r)}}{\nu_j^{(r)}(\mathbb{A})}, h\right) = W\left(\frac{\nu_j^{(r)}}{(1-\delta)^{j-1}\delta}, h\right) \leq \varepsilon \quad \text{for any } 1 \leq j \leq r$$

and $\nu_0^{(r)}(\mathbb{A}) = (1-\delta)^r < \varepsilon$. Finally, applying convexity of W stated in Remark 2.3, we observe that

$$\begin{aligned} W\left(\sum_{j=0}^r \nu_j^{(r)}, h\right) &\leq \nu_0^{(r)}(\mathbb{A})W\left(\frac{\nu_0^{(r)}}{\nu_0^{(r)}(\mathbb{A})}, h\right) + \sum_{j=1}^r \nu_j^{(r)}(\mathbb{A})W\left(\frac{\nu_j^{(r)}}{\nu_j^{(r)}(\mathbb{A})}, h\right) \leq \\ &\leq \text{Diam}(\mathbb{A})\varepsilon + \varepsilon = (\text{Diam}(\mathbb{A}) + 1)\varepsilon \end{aligned}$$

and since ε was chosen an arbitrary positive constant Theorem 2.5 is proven. $\square$

4. PROOF OF PROPOSITION 3.3: ε -DENSITY OF THE SUPPORT

We would like to start by proving the following

Lemma 4.1. *Consider $\mu \in \mathcal{M}$ that satisfies the **strict aperiodicity** condition. Then for any $\varepsilon > 0$ there exists $m \in \mathbb{N}$, such that $\text{supp}(\mu^{*m})$ is ε -dense in $\mathbb{A}$.*

Proof. It follows from **strict aperiodicity** of μ that for a fixed $\varepsilon > 0$ there exists $m_0 \in \mathbb{N}$, such that the set of points of the form $\sum_{j=1}^n \alpha_j - \beta_j$ with $\alpha_j, \beta_j \in \text{supp}(\mu)$ and $n \leq m_0$ is ε -dense in $\mathbb{A}$. Let us choose a finite number of these sums $\sum_{j=1}^{n_k} \alpha_j^{(k)} - \beta_j^{(k)}$ for $1 \leq k \leq N$ that forms a finite ε -net in $\mathbb{A}$. First of all, notice that we can make $n_k = m_0$ for all $1 \leq k \leq N$ by adding some bogus terms of the form $\alpha - \alpha$ for some $\alpha \in \text{supp}(\mu)$. Define an angle

$$\beta = \sum_{k=1}^N \sum_{j=1}^{m_0} \beta_j^{(k)}.$$

Then all the sums of the form $\sum_{j=1}^{m_0} (\alpha_j^{(k)} - \beta_j^{(k)}) + \beta$ are elements of $\text{supp}(\mu^{*m_0N})$. Hence $\text{supp}(\mu^{*m_0N}) + \beta$ is ε -dense in $\mathbb{A}$ and so is $\text{supp}(\mu^{*m_0N})$, hence $m = m_0N$ satisfies the required statement. $\square$

Remark 4.2. Notice that if $\text{supp}(\mu)$ is ε -dense then so is $\text{supp}(\tilde{\mu} * \mu) = \text{supp}(\mu * \tilde{\mu})$ for any $\mu, \tilde{\mu} \in \mathcal{M}$.

Now, consider a compact set $K \subset \mathcal{M}$ such that any measure $\mu \in K$ has the **strict aperiodicity** property and let us prove the key property of such compact set:

Proposition 4.3. *For any $\varepsilon > 0$ there exists m , such that for any sequence $\mu_1, \mu_2, \dots, \mu_m \in K$ the support $\text{supp}(\mu_m * \mu_{m-1} * \dots * \mu_1)$ is ε -dense in $\mathbb{A}$.*

We will start by proving the following

Lemma 4.4. *For any $\varepsilon > 0$ and any measure $\mu \in K$ there exist $\delta > 0$ and $m \in \mathbb{N}$, such that for any sequence of measures $\mu_1, \mu_2, \dots, \mu_m \in K$, such that $W(\mu, \mu_j) < \delta$ the support $\text{supp}(\mu_m * \mu_{m-1} * \dots * \mu_1)$ is ε -dense in $\mathbb{A}$.*

Proof of Lemma 4.4. Since $\mu \in K$ it is **strictly aperiodic**, we can apply Lemma 4.1 and conclude that for a fixed ε there exists m , such that $\text{supp}(\mu^{*m})$ is $\varepsilon/2$ -dense in $\mathbb{A}$. We would like to denote $\tilde{\varepsilon} = \varepsilon/(2m)$ and prove the following technical

Lemma 4.5. *There exists $\delta > 0$, such that for any $\tilde{\mu}$, such that $W(\mu, \tilde{\mu}) < \delta$ we have $\text{supp}(\mu) \subset \text{supp}(\tilde{\mu}) + B(0, \tilde{\varepsilon})$.*

Proof of Lemma 4.5. Assume such δ does not exist. Then for any positive $\delta > 0$ there is a measure $\tilde{\mu}_\delta$ and a point x_δ , such that $x_\delta \in \text{supp}(\mu)$ but $x_\delta \notin \text{supp}(\tilde{\mu}_\delta) + B(0, \tilde{\varepsilon})$. Then $B(x_\delta, \tilde{\varepsilon}) \cap \text{supp}(\tilde{\mu}_\delta) = \emptyset$ and hence $\mu(B(x_\delta, \tilde{\varepsilon}/2)) \cdot \tilde{\varepsilon}/2 \leq W(\mu, \tilde{\mu})$. The last inequality is true because for any measure γ on $\mathbb{A} \times \mathbb{A}$, such that $P_x(\gamma) = \mu$ and $P_y(\gamma) = \tilde{\mu}_\delta$ we will have

$$\begin{aligned} \iint_{\mathbb{A} \times \mathbb{A}} d(x, y) d\gamma(x, y) &\geq \iint_{B(x_\delta, \tilde{\varepsilon}/2) \times (\mathbb{A} \setminus B(x_\delta, \tilde{\varepsilon}))} d(x, y) d\gamma(x, y) \geq \\ &\geq \frac{\tilde{\varepsilon}}{2} \cdot \mu(B(x_\delta, \tilde{\varepsilon}/2)) \cdot \tilde{\mu}_\delta(\mathbb{A} \setminus B(x_\delta, \tilde{\varepsilon}/2)) = \frac{\tilde{\varepsilon}}{2} \cdot \mu(B(x_\delta, \tilde{\varepsilon}/2)). \end{aligned}$$

Recalling that $W(\mu, \tilde{\mu}_\delta) < \delta$ we arrive to

$$\mu(B(x_\delta, \tilde{\varepsilon}/2)) < \frac{2\delta}{\tilde{\varepsilon}}.$$

Taking $\delta \rightarrow 0$ and choosing a convergent subsequence of x_δ by compactness we find a point $x_0 \in \text{supp}(\mu)$, such that $\mu(B(x_0, \tilde{\varepsilon}/2)) = 0$, which contradicts the definition of $\text{supp}(\mu)$. $\square$

Applying Lemma 4.5 to every individual measure μ_j for $1 \leq j \leq m$ we see that for the same δ we have

$$\text{supp}(\mu^{*m}) \subset \text{supp}(\mu_m * \mu_{m-1} * \dots * \mu_1) + B(0, \varepsilon/2),$$

which fact guarantees that $\text{supp}(\mu_m * \mu_{m-1} * \dots * \mu_1)$ is ε -dense in $\mathbb{A}$. $\square$

Proof of Proposition 4.3. Let us fix some $\varepsilon > 0$. First, since K is compact we can choose such $m_0 \in \mathbb{N}$ and $\delta > 0$ that for any sequence of measures $\mu_1, \mu_2, \dots, \mu_{m_0} \in K$ with $W(\mu_j, \mu_k) < \delta$ the support $\text{supp}(\mu_{m_0} * \mu_{m_0-1} * \dots * \mu_1)$ will be ε -dense in $\mathbb{A}$. We can also choose a finite covering of K by balls of radius $\delta/2$. Assuming that covering has n balls let us take $m = nm_0$. Then any sequence of m measures will have at least m_0 terms $\mu_{n_1}, \mu_{n_2}, \dots, \mu_{n_{m_0}}$ from the same ball of radius $\delta/2$ and hence $W(\mu_{n_j}, \mu_{n_k}) < \delta$ will hold. Due to commutativity we can rearrange the measures in a following way:

$$\mu_m * \mu_{m-1} * \dots * \mu_1 = \dots * \mu_{n_{m_0}} * \mu_{n_{m_0}-1} * \dots * \mu_{n_1}.$$

The support $\text{supp}(\mu_{n_{m_0}} * \mu_{n_{m_0}-1} * \dots * \mu_{n_1})$ is ε -dense because of our choice of δ . Applying Remark 4.2 we conclude that the support $\text{supp}(\mu_m * \mu_{m-1} * \dots * \mu_1)$ is also ε -dense. $\square$

Using Proposition 4.3 we can finally finish the proof of Proposition 3.3. Assume that it is not true and fix $\varepsilon > 0$ for which the statement fails. Then by Proposition 4.3 there exists $m \in \mathbb{N}$, such that for any sequence $\mu_1, \mu_2, \dots, \mu_m \in K$ the support $\text{supp}(\mu_m * \mu_{m-1} * \dots * \mu_1)$ is $\varepsilon/8$ -dense in $\mathbb{A}$. Let us fix that m . By our assumption

for any $n \in \mathbb{N}$ there exists a sequence $\mu_1^{(n)}, \mu_2^{(n)}, \dots, \mu_m^{(n)} \in K$, $\nu^{(n)} \in \mathcal{M}$ and an ε -wide set Q_n , such that

$$[\mu_m^{(n)} * \mu_{m-1}^{(n)} * \dots * \mu_1^{(n)} * \nu^{(n)}](Q_n) < \frac{1}{n}.$$

Due to compactness of $\mathbb{A}$ we can choose an infinite subsequence Q_{n_k} , such that all Q_{n_k} cover the same ball B of radius $\varepsilon/4$ (see Definition 3.1). After that we can choose a convergent subsequence $m+1$ more times, so that $\mu_j^{(n)} \rightarrow_{n \rightarrow \infty} \mu_j$ for $1 \leq j \leq m$ and $\nu^{(n)} \rightarrow_{n \rightarrow \infty} \nu$. Now we see that for some $\mu_1, \mu_2, \dots, \mu_m \in K$ and $\nu \in \mathcal{M}$ we have

$$[\mu_m * \mu_{m-1} * \dots * \mu_1 * \nu](B) = 0,$$

which contradicts the fact that $\text{supp}(\mu_m * \mu_{m-1} * \dots * \mu_1)$ (and hence, by Remark 4.2 also $\text{supp}(\mu_m * \mu_{m-1} * \dots * \mu_1 * \nu)$) is $\varepsilon/8$ -dense. That concludes the proof of Proposition 3.3.

ACKNOWLEDGEMENTS

The author is grateful to A. Gorodetski and V. Kleptsyn for permanent attention to the work and fruitful discussions, to S. Klein and D. Dolgopyat for providing helpful references, and both anonymous reviewers for their detailed feedback and valuable comments. The author was supported in part by NSF grant DMS-2247966 (PI: A. Gorodetski).

REFERENCES

- A. A. de Acosta, Large deviations for Markov chains. Cambridge Tracts in Math., **229** Cambridge University Press, Cambridge, 2022. xii+249 pp.
- AG. M. Anoussis, D. Gatzouras, A spectral radius formula for the Fourier transform on compact groups and applications to random walks, *Adv. Math.* **188** (2004), no.2, pp. 425–443.
- BQ. Y. Benoist, J.-F. Quint, Random walks on reductive groups. *Ergeb. Math. Grenzgeb.* (3), 62 [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics], Springer, Cham, 2016. xi+323 pp.
- BB. D. Berend, V. Bergelson, Ergodic and mixing sequences of transformations. *Ergodic Theory Dynam. Systems* **4** (1984), no.3, 353–366.
- BE. A. Berger, S. Evans, A limit theorem for occupation measures of Lévy processes in compact groups, *Stoch. Dyn.* **13** (2013), no.1, 1250008, 16 pp.
- B. R. Bhattacharya, Speed of convergence of the n -fold convolution of a probability measure on a compact group, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **25** (1972/1973), pp. 1–10.
- Bi. G. Birkhoff, A note on topological groups. *Compositio Math.* **3** (1936), 427–430.
- Bo1. B. Borda, Equidistribution of random walks on compact groups, *Ann. Inst. Henri Poincaré Probab. Stat.* **57** (2021), no.1, pp. 54–72.
- Bo2. B. Borda, Equidistribution of random walks on compact groups II. The Wasserstein metric, *Bernoulli* **27** (2021), no.4, pp. 2598–2623.
- BL. P. Bougerol and J. Lacroix, Products of Random Matrices with Applications to Schrödinger Operators, Birkhauser, Boston, 1985.
- CDK1. A. Cai, P. Duarte, S. Klein, Mixed random-quasiperiodic cocycles. *Bull. Braz. Math. Soc. (N.S.)* **53** (2022), no.4, 1469–1497.
- CDK2. A. Cai, P. Duarte, S. Klein, Furstenberg Theory of Mixed Random-Quasiperiodic Cycles. *Commun. Math. Phys.* (2023). <https://doi.org/10.1007/s00220-023-04726-5>
- CDK3. A. Cai, P. Duarte, S. Klein, Statistical properties for mixing Markov chains with applications to dynamical systems, *arXiv:2210.16908* (2022).
- CR. J.-P., Conze, A. Raugi, Limit theorems for sequential expanding dynamical systems on $[0, 1]$. *Ergodic theory and related fields*, 89–121. *Contemp. Math.*, **430** American Mathematical Society, Providence, RI, 2007.

- DZ. A. Dembo, O. Zeitouni, Large deviations techniques and applications. Stoch. Model. Appl. Probab., **38** Springer-Verlag, Berlin, 2010. xvi+396 pp.
- DS. Z. Dietz, S. Sethuraman, Large deviations for a class of nonhomogeneous Markov chains. Ann. Appl. Probab. **15** (2005), no.1A, 421–486.
- DF. D. Dolgopyat, B. Fayad, Limit theorems for toral translations. Hyperbolic dynamics, fluctuations and large deviations, 227–277. Proc. Sympos. Pure Math., **89**, American Mathematical Society, Providence, RI, 2015.
- DSa. D. Dolgopyat, O. Sarig, Local Limit Theorems for Inhomogeneous Markov Chains (to appear).
- DMPs. R. Douc, E. Moulines, P. Priouret, P. Soulier, Markov chains. Springer Ser. Oper. Res. Financ. Eng. Springer, Cham, 2018. xviii+757 pp.
- DK. P. Duarte, S. Klein, Lyapunov exponents of linear cocycles. Continuity via large deviations. Atlantis Stud. Dyn. Syst., **3** Atlantis Press, Paris, 2016. xiii+263 pp.
- Fu. A. Furman, Random walks on groups and random transformations, *Handbook of dynamical systems*, vol. 1A, 931–1014, North-Holland, Amsterdam, 2002.
- G. B. Gidas, Nonstationary Markov chains and convergence of the annealing algorithm. J. Statist. Phys. **39** (1985), no.1-2, 73–131.
- GBK. P. Góra, A. Boyarsky, C. Keefe, Absolutely continuous invariant measures for non-autonomous dynamical systems. J. Math. Anal. Appl. **470** (2019), no.1, 159–168.
- GK. A. Gorodetski, V. Kleptsyn, Non-stationary version of Ergodic Theorem for random dynamical systems. Mosc. Math. J. **23** (2023), no.4, 515–532.
- H. D. Hackler, Compact groups of real power need not be metrizable. Proc. Amer. Math. Soc. **63** (1977), no.1, 187.
- Ha. P. Halmos, Measure theory, Springer New York, 1950.
- K. S. Kakutani, Über die Metrisation der topologischen Gruppen. (German) Proc. Imp. Acad. Tokyo **12** (1936), no.4, 82–84.
- Ka. Y. Kawada, On the probability distribution on a compact group. II. Proc. Phys.-Math. Soc. Japan (3) **23** (1941), 669–686.
- KI. Y. Kawada, K. Itô, On the probability distribution on a compact group. I., Proc. Phys.-Math. Soc. Japan (3) **22** (1940), pp. 977–998.
- Kl. B. Kloss, О вероятностных распределениях на бикомпактных топологических группах (Russian. English summary), Teor. Veroyatnost. i Primenen **4** (1959), pp. 255–290, (Title translation) Probability Distributions on Bicomact Topological Groups.
- L. P. Lévy, L'addition des variables aléatoires définies sur un circonférence (French), Bull. Soc. Math. France **67** (1939), pp. 1–41.
- LL1. G. Liu, W. Liu, On the strong law of large numbers for functionals of countable nonhomogeneous Markov chains. Stochastic Process. Appl. **50** (1994), no.2, 375–391.
- LL2. W. Liu, G. Liu, A class of strong laws for functionals of countable nonhomogeneous Markov chains. Statist. Probab. Lett. **22** (1995), no.2, 87–96.
- LY. W. Liu, W. Yang, An extension of Shannon-McMillan theorem and some limit properties for nonhomogeneous Markov chains. Stochastic Process. Appl. **61** (1996), no.1, 129–145.
- PRU. A. Prékopa, A. Rényi, K. Urbanik, О предельном распределении для сумм независимых случайных величин на бикомпактных коммутативных топологических группах (Russian. English summary), Acta Math. Acad. Sci. Hungar. **7** (1956), pp. 11–16, (Title translation) On the limiting distribution of sums of independent random variables in bicomact commutative topological groups.
- RX. K. Ross, D. Xu, Norm convergence of random walks on compact hypergroups, Math. Z. **214** (1993), no.3, pp. 415–423.
- S. K. Stromberg, Probabilities on a compact group, Trans. Amer. Math. Soc. **94** (1960), pp. 295–309.
- U. K. Urbanik, On the limiting probability distribution on a compact topological group, Fund. Math. **44** (1957), pp. 253–261.
- V. S. Varadhan, Large deviations and applications. CBMS-NSF Regional Conf. Ser. in Appl. Math., **46** Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1984. v+75 pp.
- Vi. C. Villani, Optimal transport. Old and new, Grundlehren Math. Wiss., **338** [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 2009. xxii+973 pp.

- Vit. G. Vitali, Sui gruppi di punti e sulle funzioni di variabili reali, Atti dell'Accademia delle Scienze di Torino (Italian), **43** (1908), pp. 75–92, (Title translation) On groups of points and functions of real variables.
- Y1. W. Yang, Convergence in the Cesàro sense and strong law of large numbers for nonhomogeneous Markov chains. Ninth special issue on linear algebra and statistics, Linear Algebra Appl. **354** (2002), 275–288.
- Y2. W. Yang, Strong law of large numbers for countable nonhomogeneous Markov chains. Linear Algebra Appl. **430** (2009), no.11-12, 3008–3018.
- ZYW. P. Zhang, W. Yang, B. Wang, The strong law of large numbers for multivariate functions of continuous-state nonhomogeneous Markov chains, Dynamic Systems and Applications **27** (2018), no. 2, pp. 257-269.

GRIGORII MONAKOV, DEPARTMENT OF MATHEMATICS, UNIVERSITY OF CALIFORNIA, IRVINE, CA 92697, USA

Email address: gmonakov@uci.edu