

# NON-STATIONARY VERSION OF ERGODIC THEOREM FOR RANDOM DYNAMICAL SYSTEMS

ANTON GORODETSKI AND VICTOR KLEPTSYN

*To our teacher, Professor Yulij Sergeevich Ilyashenko*

ABSTRACT. We prove a version of pointwise Ergodic Theorem for non-stationary random dynamical systems. Also, we discuss two specific examples where the result is applicable: non-stationary iterated function systems and non-stationary random matrix products.

## 1. INTRODUCTION

Birkhoff Ergodic Theorem is one of the key tools of the theory of dynamical systems. In the context of topological dynamics it claims that for a continuous dynamical system  $f : X \rightarrow X$  on a metric compact  $X$  and an  $f$ -invariant measure  $\nu$  on  $X$ , given a “test function”  $\varphi \in C(X)$ , the time averages

$$\frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x))$$

converge for  $\nu$ -a.e.  $x \in X$ , and if the measure  $\nu$  is ergodic, the limit almost everywhere equals the space average  $\int_X \varphi d\nu$ .

It admits a natural generalization for the case of random dynamical systems. Namely, let  $X$  be a compact metric space, and  $\mu$  be a probability measure on the space of continuous maps  $C(X, X)$ . The iterations of the corresponding random dynamical system are the sequences of compositions

$$f_1, f_2 \circ f_1, \dots, f_n \circ \dots \circ f_1, \dots,$$

where  $f_i : X \rightarrow X$  are chosen randomly and independently w.r.t. the measure  $\mu$ .

An analogue of the notion of an invariant measure is the one of a stationary measure. Namely, a  $\mu$ -stationary measure  $\nu$  is a probability measure on  $X$  such that  $\mu * \nu = \nu$ , where the convolution  $\mu * \nu$  is the law of  $f(x)$  for independent  $f$  chosen w.r.t.  $\mu$  and  $x$  w.r.t.  $\nu$ . Such a measure always exists (this can be shown via an analogue of the Krylov-Bogolyubov averaging procedure), though is not unique in general (in the same way as an invariant measure is not unique for a classical deterministic dynamical system in general).

---

*Date:* June 27, 2023.

A. G. was supported in part by NSF grants DMS-1855541 and DMS-2247966. Part of the work was done while A. G. was in residence at Institute Mittag-Leffler in Djursholm, Sweden, during the Spring semester of 2023, and was supported by the Swedish Research Council under grant no. 2016-06596.

V.K. was supported in part by ANR Gromev (ANR-19-CE40-0007) and by Centre Henri Lebesgue (ANR-11-LABX-0020-01).

The following theorem corresponds to the Birkhoff Ergodic Theorem for the stationary random dynamics (e.g. see [DMPS, Theorem 5.2.9]):

**Theorem 1.1** (Random Birkhoff Ergodic Theorem). *For any ergodic stationary measure  $\nu$  on  $X$ , for any  $\varphi \in C(X, \mathbb{R})$ , for  $\nu$ -a.e.  $x \in X$ ,  $\mu^{\mathbb{N}}$ -almost surely one has*

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

where  $f_1, f_2, \dots$  are chosen randomly and independently, with respect to the distribution  $\mu$ .

However, this generalisation heavily relies on the fact that  $f_n$ 's are identically distributed. What can be said if  $f_n$ 's are still chosen independently, but each  $f_n$  is distributed w.r.t. a different measure  $\mu_n$ ? This can be represented symbolically by the following diagram:

$$X \xrightarrow[f_1]{\mu_1} X \xrightarrow[f_2]{\mu_2} X \rightarrow \dots \rightarrow X \xrightarrow[f_n]{\mu_n} X \rightarrow \dots$$

At first glance, many notions and tools disappear, starting with the one of the stationary measure: the measures  $\mu_n$  on different steps may differ, and the system of equations  $\nu = \mu_n * \nu$  for all  $n$  is usually incompatible. Thus even finding a proper statement, not to mention proving it, is non-evident.

Meanwhile, such a setting naturally arises in some situations. For example, the study of one-dimensional Anderson Localization problem [GK3] in presence of a background potential leads to the study of products of independent non-identically distributed matrices, the non-stationary Furstenberg Theorem [GK2], that is within the frames of non-stationary ergodic theory. It also motivated the study of Hölder regularity of averaged images of a given initial measure in [GKM, Theorem 2.8].

The goal of the present paper is to discuss a paradigm for ergodic theorems in a non-stationary setting, presenting both (counter)-examples and a restricted version of an ergodic theorem, see Theorem 3.1 below. The latter is only a “proof-of-concept”; it definitely can be vastly generalized, and we hope it soon will be: our goal is to motivate the farther research in this direction.

## 2. PRELIMINARIES

Throughout this paper, we set  $X$  to be a compact metric space, and assume that a “test function”  $\varphi \in C(X)$  is given. Also, for every  $n$  we assume that a measure  $\mu_n$  on  $C(X, X)$  is given (if the dynamics is assumed to be invertible, one can ask instead for the measure on the set of homeomorphisms of  $X$ ). We then denote

$$(1) \quad \mathbb{P} := \prod_{n=1}^{\infty} \mu_n,$$

and our goal is to describe time averages along  $\mathbb{P}$ -almost every sequence of iterations.

Let us denote by  $\mathcal{M} = \mathcal{M}(X)$  the space of Borel probability measures on the compact metric space  $X$ . We consider it to be equipped with the Wasserstein distance (that is one of the ways to metrize the weak-\* topology in the space of probability measures):

**Definition 2.1.** Let  $\nu_1, \nu_2 \in \mathcal{M}$  be two probability measures on  $X$ . The *Wasserstein distance* between  $\nu_1$  and  $\nu_2$  is defined as

$$(2) \quad \text{dist}_{\mathcal{M}}(\nu_1, \nu_2) = \inf_{\gamma} \iint_{X \times X} d_X(x, y) d\gamma(x, y),$$

where the infimum is taken over all probability measures  $\gamma$  on  $X \times X$  with the marginals  $\nu_1$  and  $\nu_2$ , that is,  $(\pi_j)_* \gamma = \nu_j$ ,  $j = 1, 2$ , where  $\pi_{1,2} : X \times X \rightarrow X$  is the projection on the first and second factor respectively.

Next, note that for a random orbit

$$(x_n)_{n \in \mathbb{N}}, \quad x_n = f_n(x_{n-1})$$

the non-stationarity of the dynamics implies that the law of  $x_n$  can vary quite strongly, thus the same applies to the law of  $\varphi(x_n)$  and to its expectation. Thus, we should not expect the time averages  $\frac{1}{n} \sum_{k=0}^{n-1} \varphi(x_k)$  to converge. Instead, we should expect these to have a deterministic behaviour, similarly to the non-stationary Law of Large Numbers. Namely, under reasonable assumptions for independent, but not identically distributed random variables  $\xi_n$  one has

$$\frac{1}{n} \left| \sum_{k=1}^n \xi_k - \sum_{k=1}^n \mathbb{E} \xi_k \right| \rightarrow 0, \quad n \rightarrow \infty.$$

As the Birkhoff Ergodic Theorem generalises the Law of Large Numbers, we should expect

$$(3) \quad \frac{1}{n} \left| \sum_{k=1}^n \varphi(x_k) - \sum_{k=1}^n \int_X \varphi(x) d\nu_k(x) \right| \rightarrow 0, \quad n \rightarrow \infty$$

for some reasonable non-random measures  $\nu_k$ . Indeed, this turns out to be the case: see Theorem 3.1 below.

### 3. STATEMENT OF THE MAIN RESULT

Our main result will be stated under the following assumption. This assumption, in a sense, generalizes (and slightly strengthens) the *unique* ergodicity. The latter requires the stationary measure to be unique, so this setting is more restrictive than “simple” ergodicity; however, in this setting such an assumption seems to be appropriate, see the discussion in Section 6.2 below, and Example 6.3 therein.

**Standing Assumption:** *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  $\delta$ -close to each other:*

$$\forall \nu, \nu' \in \mathcal{M}, \quad \forall n \in \mathbb{N}, \quad \text{dist}_{\mathcal{M}}(\mu_{n+m} * \dots * \mu_{n+1} * \nu, \mu_{n+m} * \dots * \mu_{n+1} * \nu') < \delta.$$

Note that if after a given number  $m$  of steps the averaged images of any two initial measures  $\nu, \nu'$  are close to each other, one can pick any of these images and state that all the others are close to it. So one can pick any measure  $\nu_0$  (for instance, a Dirac one at a given initial point  $x_0$ ) and define

$$(4) \quad \nu_n := \mu_n * \nu_{n-1}, \quad n = 1, 2, \dots$$

Then, for any given measure  $\nu$ , the convolution  $\mu_{n+m} * \dots * \mu_{n+1} * \nu$  is close to  $\nu_{n+m}$ , and this motivates the appearance of measures  $\nu_n$  in (3).

Here is the main result of this paper:

**Theorem 3.1.** *Suppose the sequence of distributions  $\mu_1, \mu_2, \mu_3, \dots$  satisfies the Standing Assumption above. Given any Borel probability measure  $\nu_0$  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*

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

*where the measures  $\nu_n$  are defined by (4).*

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

$$(5) \quad \forall n \in \mathbb{N}, \quad \mathbb{P} \left( \frac{1}{n} \left| \sum_{k=1}^n \varphi(f_k \circ \dots \circ f_1(x)) - \sum_{k=1}^n \int_X \varphi d\nu_k \right| > \varepsilon \right) < C \exp(-\delta n).$$

**Remark 3.2.** While the Standing Assumption above might seem quite restrictive, it does hold in a natural way in some interesting cases. In the next section we prove that it holds in two settings, namely, for random dynamical systems defined by contractions, also known as random iterated function systems, and for random dynamical systems on a projective space generated by random matrix products.

**Remark 3.3.** In the stationary case, when  $\mu_i = \mu$  for all  $i \in \mathbb{N}$ , Standing Assumption implies that the random dynamical system defined by the distribution  $\mu$  is uniquely ergodic, i.e. has a unique stationary measure. Notice, however, that the Standing Assumption is actually stronger (i.e. more restrictive) than the unique ergodicity. Namely, unique ergodicity is equivalent to the statement that *time averages* of convoluted images of any initial measure  $\nu$  converge to the unique stationary measure  $\bar{\nu}$  uniformly in  $\nu$ :

$$\frac{1}{m} \sum_{j=0}^m \mu^{(*j)} * \nu \rightarrow \bar{\nu}, \quad m \rightarrow \infty,$$

where  $\mu^{(*j)}$  is  $j$ -th convolution power. Example 6.3 explains why the Standing Assumption in the non-stationary setting should indeed be stronger.

**Remark 3.4.** Large Deviation Estimates for stationary Markov chains were heavily studied, and numerous stationary versions of (5) are known, e.g. see [DV, Ga] for classical texts, [Ac] for a recent monograph on the subject, as well as [DK1, Theorem 5.2], [DK2, Theorem 4.4].

**Remark 3.5.** We would like to point out that there are different settings where non-stationary versions of ergodic theorems could be and were established. For example, the setting where a sequence of maps that preserve the same fixed measure was considered in [BB]. The case when all maps are non-singular with respect to the Lebesgue measure, so the transfer operator technics are applicable, was studied in [CR, HNTV]. Sometimes a term *sequential dynamical system* is used in these contexts. We emphasize that both the setting and the underlying mechanisms there are different from ours.

## 4. EXAMPLES OF APPLICABILITY

We will give two examples of non-stationary random dynamical systems for which Standing Assumption is satisfied and, hence, Theorem 3.1 is applicable.

**4.1. Random Iterated Function Systems.** An iterated function system is defined by a finite collection of contractions of a complete metric space, usually  $\mathbb{R}^d$  or a compact subset of  $\mathbb{R}^d$ . In many cases it leads to a fractal attractor. For example, the standard Cantor set can be defined via two contractions of the unite interval,  $x \mapsto \frac{x}{3}$  and  $x \mapsto \frac{x}{3} + \frac{2}{3}$ . Many self-similar fractals can be generated this way. If probabilities are assigned to each of the contractions, one can consider a corresponding random dynamical system. It is known that it must have unique stationary measure [Hu, Sz]. Properties of this stationary measure were a subject of intense studies. There is a vast amount of literature on exact-dimensionality and fractal dimension of this measure; we will only mention a recent survey [FS]. Hölder regularity of the this measure was studied in [FL, Proposition 2.2] and [GKM, Section 1.2]. For a survey of the theory of Bernoulli convolutions that also fall into this setting see [PSS] or a more recent [Va].

One can consider a non-stationary random dynamical system generated by contractions, where a different distribution on the space of contraction can be chosen on each step. This setting was considered, for example, in [GKM]; it was shown there that under some mild and natural conditions the averaged iterates of any given initial measure must converge to the space of Hölder regular measures exponentially fast.

One can easily see that in this setting the Standing Assumption holds, and hence, Theorem 3.1 is applicable:

**Proposition 4.1.** *Let  $X$  be a compact metric space, and  $\lambda \in (0, 1)$  be a constant. Suppose  $\{\mu_i\}_{i \in \mathbb{N}}$  be a sequence of probability distributions in the space of contractions  $X \rightarrow X$  with Lipschitz constant at most  $\lambda$ . Then Standing Assumption holds, i.e. for any Borel probability measures  $\nu, \nu'$  on  $X$  we have*

$$\text{dist}_{\mathcal{M}}(\mu_{n+m} * \dots * \mu_{n+1} * \nu, \mu_{n+m} * \dots * \mu_{n+1} * \nu') \rightarrow 0 \quad \text{as } m \rightarrow \infty,$$

*uniformly in  $(\nu, \nu')$  and  $n \in \mathbb{N}$ .*

*Proof.* Indeed, for any  $\mu_i$ , any map  $f \in \text{supp } \mu_i$  is a contraction with Lipschitz constant at most  $\lambda$ . Hence for any  $\nu_1, \nu_2 \in \mathcal{M}$  we have

$$\text{dist}_{\mathcal{M}}(\mu * \nu_1, \mu * \nu_2) \leq \lambda \text{dist}_{\mathcal{M}}(\nu_1, \nu_2),$$

and Proposition 4.1 follows.  $\square$

**4.2. Random Matrix Products.** Let us now show that under some non-degeneracy assumptions a non-stationary random dynamical system defined by projective maps must satisfy Standing Assumption.

Suppose  $\mathbf{K}$  is a compact set in the space of probability distributions on  $SL(2, \mathbb{R})$  such that for any  $\mu \in \mathbf{K}$  the *measures condition* is satisfied (we use terminology from [GKM]), i.e. there are no probability measures  $\nu_1, \nu_2$  on  $\mathbb{RP}^1$  such that  $(f_A)_*(\nu_1) = \nu_2$  for all  $A \in \text{supp } \mu$ . Slightly abusing the notation, we will treat  $\mu$  also as a measure on the space of projective maps  $f_A : \mathbb{RP}^1 \rightarrow \mathbb{RP}^1$ .

**Proposition 4.2.** *Suppose the measure condition holds for each  $\mu \in \mathbf{K}$ . Then for any sequence  $\{\mu_i\} \in \mathbf{K}^{\mathbb{N}}$  and any probability measures  $\nu, \nu' \in \mathcal{M}$  we have:*

$$\text{dist}_{\mathcal{M}}(\mu_{n+m} * \dots * \mu_{n+1} * \nu, \mu_{n+m} * \dots * \mu_{n+1} * \nu') \rightarrow 0 \quad \text{as } m \rightarrow \infty,$$

*uniformly in  $(\nu, \nu')$  and  $n \in \mathbb{N}$ .*

We will deduce this proposition from the following statement, that is also of independent interest:

**Proposition 4.3.** *Suppose the measure condition holds for each  $\mu \in \mathbf{K}$ . Then for any  $\varepsilon > 0$  there exists  $m$  such that for any measures  $\mu_1, \dots, \mu_m \in \mathbf{K}$  and any points  $x, y \in \mathbb{R}P^1$  one has*

$$(\mu_m * \dots * \mu_1)\{F \mid d(F(x), F(y)) < \varepsilon\} > 1 - \varepsilon,$$

*where the composition  $F = f_m \circ \dots \circ f_1$  is formed by independent random maps  $f_i$  distributed w.r.t.  $\mu_i$ . In other words, the random composition  $F$  brings  $x$  and  $y$  to the distance less than  $\varepsilon$  with probability at least  $1 - \varepsilon$ .*

This deduction is indeed immediate:

*Proof of Proposition 4.2.* Note first that

$$\begin{aligned} \text{dist}_{\mathcal{M}}(\mu_{n+m} * \dots * \mu_{n+1} * \nu, \mu_{n+m} * \dots * \mu_{n+1} * \nu') &\leq \\ \iint_{\mathbb{R}P^1} \text{dist}_{\mathcal{M}}(\mu_{n+m} * \dots * \mu_{n+1} * \delta_x, \mu_{n+m} * \dots * \mu_{n+1} * \delta_y) d\nu(x) d\nu'(y); \end{aligned}$$

this follows from the convexity of the transport distance: mixing the couplings provides a coupling for the averaged measures.

At the same time, we have

$$\begin{aligned} (6) \quad \text{dist}_{\mathcal{M}}(\mu_{n+m} * \dots * \mu_{n+1} * \delta_x, \mu_{n+m} * \dots * \mu_{n+1} * \delta_y) &\leq \\ &\leq \int d(F(x), F(y)) d(\mu_{n+m} * \dots * \mu_{n+1})(F). \end{aligned}$$

Take an arbitrary  $\varepsilon > 0$ ; once  $m$  is taken sufficiently large so that the conclusion of Proposition 4.3 holds, the right hand side of (6) does not exceed

$$\begin{aligned} \varepsilon \cdot (\mu_{n+m} * \dots * \mu_{n+1})\{F \mid d(F(x), F(y)) < \varepsilon\} + \\ + 1 \cdot (\mu_{n+m} * \dots * \mu_{n+1})\{F \mid d(F(x), F(y)) \geq \varepsilon\} \leq \varepsilon \cdot 1 + 1 \cdot \varepsilon = 2\varepsilon. \end{aligned}$$

As  $\varepsilon > 0$  is arbitrary, we get the desired convergence that is uniform in  $n$ ,  $\nu$ , and  $\nu'$ .  $\square$

Let us now return to Proposition 4.3. Before proceeding to the proof, note that if the dynamics was stationary and if the expectation of  $\log \|A\|$  was bounded, the conclusion of the proposition would be implied by the famous Furstenberg Theorem. This motivates to approach it via a non-stationary version of Furstenberg Theorem. Such versions were studied recently in [GK2] and [G].

*Proof of Proposition 4.3.* Assume that the expectation of  $\|A\|$  is bounded uniformly in  $\mu \in \mathbf{K}$ :

$$(7) \quad \exists C_{\mathbf{K}} : \quad \forall \mu \in \mathbf{K} \quad \int \|A\| d\mu(A) < C_{\mathbf{K}}.$$

Then we are in a situation when Theorems 1.1 and 1.4 from [GK2] are applicable. Namely, denote

$$L_n := \mathbb{E} \log \|A_n \dots A_1\|.$$

Theorem 1.1 from [GK2] then implies a linear lower bound for  $L_n$ : there exists  $\lambda_{\mathbf{K}} > 0$  such that

$$L_n \geq n\lambda_{\mathbf{K}};$$

moreover,  $\lambda_{\mathbf{K}}$  does not depend on the individual choices of  $\mu_i$ , but only on the compact  $\mathbf{K}$  itself.

Now, Theorem 1.4 from [GK2] states that a Large Deviations type estimate holds: for any fixed  $\delta > 0$ , there exists  $\delta'$  such that for any  $v_0 \in \mathbb{R}^2$ ,  $|v_0| = 1$ , and any sufficiently large  $n$  (that is, for all  $n$  greater than some  $n_0$ ) one has

$$\mathbb{P}(|\log |A_n \dots A_1 v_0| - L_n| > n\delta) < e^{-n\delta'}.$$

Moreover (see [GK2, Remark 1.4]), the constants  $\delta'$  and  $n_0$  can be chosen uniformly in all possible sequences of  $\mu_i \in \mathbf{K}$ .

To establish the conclusion of the proposition in this particular case, i.e. when (7) holds, take  $\delta := \frac{1}{2}\lambda_{\mathbf{K}} > 0$  and consider the corresponding  $n_0, \delta'$ . Then, once the inequality

$$(8) \quad |\log |A_n \dots A_1 v_0| - L_n| < n\delta$$

holds, it implies

$$\log |A_n \dots A_1 v_0| > L_n - n\delta \geq \frac{1}{2}\lambda_{\mathbf{K}}n.$$

In particular, for any given  $\varepsilon > 0$ , taking  $n_1$  to be the integer part of  $\frac{\log 2\varepsilon^{-1}}{\frac{1}{2}\lambda_{\mathbf{K}}}$ , we have

$$|A_n \dots A_1 v_0| > 2\varepsilon^{-1}$$

once  $n > n_1$  and (8) holds. In turn, this inequality implies that the angle between the image  $A_n \dots A_1 v_0$  and the image of the most expanded unit vector does not exceed  $\frac{\varepsilon}{2}$ . Hence, having it for two vectors  $v_0, v'_0$  corresponding to two points  $x, y \in \mathbb{RP}^1$  implies that

$$(9) \quad \text{dist}(F(x), F(y)) < \varepsilon,$$

where  $F = f_{A_n \dots A_1}$ . On the other hand, for any  $n > n_2 := \lceil \frac{1}{\delta'} \log 2\varepsilon^{-1} \rceil$  one has

$$e^{-n\delta'} < \frac{\varepsilon}{2}.$$

Hence, for  $n > \max(n_0, n_1, n_2)$  we get the desired (9) with the probability at least

$$1 - 2e^{-n\delta'} > 1 - 2\frac{\varepsilon}{2} = 1 - \varepsilon$$

for arbitrary  $\mu_1, \dots, \mu_n \in \mathbf{K}$  and  $x, y \in \mathbb{RP}^1$ .

Now, let us pass to the general case. The following tool is useful to study random dynamical systems on the circle. Consider an inverse stationary measure, that is, a measure  $\nu^-$  such that

$$\nu^- = \mathbb{E} [(f^{-1})_* \nu^-].$$

Then the measures of forward random iterations of an interval form a martingale:

$$\mathbb{E} \nu^-(f(J)) = \mathbb{E} [((f^{-1})_* \nu^-)](J) = \nu^-(J).$$

In the non-stationary setting the notion of *one* stationary measure is no longer applicable. Thus, similarly to the statement of Theorem 3.1, we will replace this

notion by a sequence of measures related by convolution. Namely, for every given  $n$  and measures  $\mu_1, \dots, \mu_n \in \mathbf{K}$ , we define

$$\nu_n^- := \text{Leb}, \quad \forall i = 1, \dots, n \quad \nu_{i-1}^- := \mathbb{E}_{\mu_i} [(f^{-1})_* \nu_i^-] = \mu_i^- * \nu_i^-,$$

where the measure  $\mu_i^-$  is the push-forward image of  $\mu_i$  under the inversion map  $f \mapsto f^{-1}$ . Then we also have the martingale property for the measures of the iterations: for any  $i$  and any interval  $J$  one has

$$(10) \quad \mathbb{E}_{\mu_i} [\nu_i^-(f(J))] = \nu_{i-1}^-(J),$$

and thus

$$(11) \quad \mathbb{E}_{\mu_n * \dots * \mu_i} [\nu_n^-(F(J))] = \nu_{i-1}^-(J).$$

For any given  $x, y \in \mathbb{RP}^1$  consider the sequence of their random iterations: let  $f_i$  be chosen independently w.r.t.  $\mu_i$ , and denote

$$x_0 := x, y_0 := y, \quad x_i = f_i(x_{i-1}), y_i = f_i(y_{i-1}).$$

We then have the following lemma.

**Lemma 4.4.** *Assume that for some  $i$  after  $i$  iterations we have  $\nu_i^-([x_i, y_i]) < \frac{1}{2}\varepsilon^2$  or  $\nu_i^-([y_i, x_i]) < \frac{1}{2}\varepsilon^2$ . Then with the probability at least  $1 - \frac{\varepsilon}{2}$  we have*

$$\text{dist}(x_n, y_n) < \varepsilon.$$

*Proof.* Assume that we are in the first case. Then as

$$\mathbb{E}(\nu_n^-([x_n, y_n]) \mid f_1, \dots, f_i) = \nu_i^-([x_i, y_i]) < \varepsilon \cdot \frac{\varepsilon}{2},$$

due to the Markov inequality with the probability at least  $1 - \frac{\varepsilon}{2}$  we have

$$\text{dist}(x_n, y_n) = \text{Leb}([x_n, y_n]) = \nu_n^-([x_n, y_n]) < \varepsilon.$$

The case  $\nu_i^-([y_i, x_i]) < \frac{1}{2}\varepsilon^2$  can be treated in the same way.  $\square$

Roughly speaking, Lemma 4.4 says that if the random iterations of  $x$  and  $y$  approach each other in the sense of the measure  $\nu_i^-$ , they most probably will stay sufficiently close to each other in the usual sense at time  $n$ , too. The next observation is that measures  $\nu_i^-$  are non-atomic, uniformly in both  $i$  and possible choices of measures  $\mu_j$ . Namely, we have the following lemma.

**Lemma 4.5.** *There exists  $\varepsilon_1 > 0$  such that for any  $n, i$  and  $\mu_1, \dots, \mu_n \in \mathbf{K}$  and any interval  $J$  of length  $|J| < \varepsilon_1$  we have  $\nu_i^-(J) < \frac{1}{2}\varepsilon^2$ .*

*Proof.* Note first that the measures  $\mu^-$  corresponding to the inverse maps also satisfy the measures condition. Indeed, if we had  $f_*^{-1}\nu = \nu'$  for  $\mu$ -a.e.  $f$  for some  $\mu \in \mathbf{K}$ , then we would also have  $f_*\nu' = \nu$ , and that would be a contradiction.

Now, the Atoms Dissolving Theorem 1.13 from [GK2] is applicable, and it states that there exists some  $k_0$  such that for any  $\mu'_1, \dots, \mu'_{k_0} \in \mathbf{K}$  and any  $x', x'' \in \mathbb{RP}^1$

$$(12) \quad ((\mu'_1)^- * \dots * (\mu'_{k_0})^- * \delta_{x'}) (\{x''\}) < \frac{1}{2}\varepsilon^2.$$

In other words, any (reverse) length  $k_0$  convolution makes weights of atoms dissolve so they do not exceed  $\frac{1}{2}\varepsilon^2$ . This implies that for some  $\varepsilon'$  and any  $\mu_1, \dots, \mu_{k_0} \in \mathbf{K}$  one has

$$(13) \quad (\mu_1^- * \dots * \mu_{k_0}^- * \delta_{x'}) (J) < \frac{1}{2}\varepsilon^2 \quad \forall J, |J| \leq \varepsilon'.$$



Indeed, if this statement would be violated for any  $\varepsilon'_m = \frac{1}{m}$ , then for some  $J_m$ ,  $\mu_{1,m}, \dots, \mu_{k_0,m}$ , we would find a convergent subsequence of these data,

$$J_{m_j} \rightarrow x'', \quad \mu_{i,m_j} \rightarrow \mu'_i \in \mathbf{K},$$

and would get a contradiction with (12). Averaging (12) over  $x'$  w.r.t.  $\nu_{i+k_0}^-$ , we obtain the conclusion of the lemma for all  $i \leq n - k_0$ .

Finally, all the measures  $\nu_{n-k}^-$  for  $k = 1, \dots, k_0$  are non-atomic, as by definition  $\nu_n^- = \text{Leb}$ . Again, this implies that there exists  $\varepsilon''$  such that for any  $\mu_{n-k+1}, \dots, \mu_n \in \mathbf{K}$

$$(14) \quad (\mu_{n-k+1}^- * \dots * \mu_n^- * \nu_n^-)(J) < \frac{1}{2}\varepsilon^2 \quad \forall J, |J| \leq \varepsilon''.$$

Indeed, otherwise we find a convergent subsequence of intervals  $J_m$  and measures  $\mu_{n-j}^- \in \mathbf{K}$ , and obtain a contradiction with the non-atomicity mentioned above.

Taking  $\varepsilon_1 := \min(\varepsilon', \varepsilon'')$  concludes the proof of Lemma 4.5.  $\square$

Combining the statements of Lemma 4.5 and Lemma 4.4, we get the following corollary:

**Corollary 4.6.** *For any  $\varepsilon > 0$  there exists  $\varepsilon_1 > 0$  such that the following holds. Assume that for some  $i$  after  $i$  iterations we have  $\text{dist}(x_i, y_i) < \varepsilon_1$ . Then with the probability at least  $1 - \frac{\varepsilon}{2}$  we have*

$$\text{dist}(x_n, y_n) < \varepsilon.$$

We are now ready to conclude the proof of Proposition 4.3 by considering two different cases. Assume first that there exists  $\varepsilon_2 > 0$  and  $k_2$  such that for any  $x', y' \in \mathbb{R}\mathbb{P}^1$  and any  $\mu_1, \dots, \mu_{k_2} \in \mathbf{K}$  with the probability at least  $\varepsilon_2$  the following event holds:

$$\exists i \leq k_2 : \quad \text{dist}(x'_i, y'_i) \leq \varepsilon_1,$$

where

$$x'_0 := x', y'_0 := y', \quad x'_i = f_i(x'_{i-1}), y'_i = f_i(y'_{i-1}),$$

and  $f_i$  are random maps chosen independently w.r.t.  $\mu_i$ .

In this case, for any  $k_3$  the probability that the iterations of two initial points  $x$  and  $y$  do not approach each other closer than  $\varepsilon_1$  during  $k_2 \cdot k_3$  iterations does not exceed  $(1 - \varepsilon_2)^{k_3}$ : we have  $k_3$  attempts with at least  $\varepsilon_2$  chance of success at each of these. Taking  $k_3$  sufficiently large so that  $(1 - \varepsilon_2)^{k_3} < \frac{\varepsilon}{2}$  and defining  $n := k_2 \cdot k_3$ , we get that with the probability at least  $1 - \frac{\varepsilon}{2}$  there exists  $i \leq n$  such that

$$\text{dist}(x_i, y_i) < \varepsilon_1,$$

and hence  $\text{dist}(x_n, y_n) < \varepsilon$  with the probability at least  $(1 - \frac{\varepsilon}{2})^2 > 1 - \varepsilon$ .

Finally, if such  $\varepsilon_2$  and  $k_2$  do not exist, we take a sequence of candidates  $\varepsilon_{2,(m)} = \frac{1}{m}$ ,  $k_{2,(m)} = m$  and consider the points  $x'_{(m)}, y'_{(m)}$  and measures  $\mu_{1,(m)}, \dots, \mu_{m,(m)} \in \mathbf{K}$ , for which the desired statement fails. Extracting a convergent subsequence, we find  $x', y'$  and  $\mu_1, \mu_2, \dots$ , for which almost surely the images  $x'_i, y'_i$  always stay at the distance at least  $\varepsilon'$ .

Now, Atoms Dissolving Theorem 1.13 from [GK2] together with the compactness argument that we have already applied imply that for some  $k_4$ ,  $\varepsilon_3 > 0$  for any interval  $J$  of length  $|J| \leq \varepsilon_3$  we have

$$\mathbb{P}(x_{k_4} \in J), \mathbb{P}(y_{k_4} \in J) \leq \frac{1}{3}.$$

In particular, for any  $i > k_4$  the product of matrices  $T_{[k_4, i]} := A_i \dots A_{k_4+1}$  cannot have norm higher than  $\frac{4}{\varepsilon_3 \varepsilon_2}$ , as otherwise with the probability at least  $\frac{1}{3}$  both points  $x_{k_4}, y_{k_4}$  will be at the distance at least  $\frac{\varepsilon_3}{2}$  from its most contracted direction, and thus their images

$$x_i = f_{T_{[k_4, i]}}(x_{k_4}), \quad y_i = f_{T_{[k_4, i]}}(y_{k_4})$$

will be at the distance at most  $\frac{1}{\|T_{[k_4, i]}\| \cdot \frac{\varepsilon_3}{2}} < \frac{\varepsilon_2}{2}$  from the image of the most expanded direction, and hence closer than  $\varepsilon_2$  to each other.

Thus, the products  $A_i \dots A_{k_4+1}$  almost surely satisfy a uniform upper bound by  $\frac{4}{\varepsilon_3 \varepsilon_2}$ . Hence, the norms of the matrices  $A_i$  are also uniformly bounded.

Consider now the compact set  $\mathbf{K}'$  that is the closure of  $\{\mu_{k_4+1}, \mu_{k_4+2}, \dots\}$ . This is a closed subset of  $\mathbf{K}$ , and hence a compact set of measures, on which the norms of matrices are uniformly bounded and for which the measure condition is satisfied. Hence, we are again in the assumptions of the Nonstationary Furstenberg Theorem ([GK2, Theorem 1.4]). And thus the uniform bound on the norms of products  $T_{[k_4, i]}$  provides us a contradiction, making this second case impossible.

This concludes the proof of Proposition 4.3.  $\square$

**Remark 4.7.** In case of a higher dimension, without additional assumptions (an analogue of absence of a finite invariant set of planes) the measure condition does not suffice to ensure the nonstationary ergodicity, as [GK2, Example A.1] shows. Nevertheless, with some extra assumptions (that would be sufficient to guarantee a non-stationary version of “simplicity of Lyapunov spectrum”, or at least “simplicity of the first Lyapunov exponent”) the conclusion of the proposition could be generalised to the case of  $SL(d, \mathbb{R})$ ,  $d > 2$ .

**Remark 4.8.** Contraction of random orbits for general stationary dynamics on the circle was established and studied in detail in [A], [KN], [Mal]. Even though in some parts of the proof of Proposition 4.3 we have used the fact that we are composing projective maps, we expect that under some suitable assumptions the statement and the proof can be adapted to a more general case of non-stationary dynamics of circle homeomorphisms.

## 5. PROOF OF THEOREM 3.1

We will establish the second part of the theorem, the Large Deviations estimates, first; it implies the first part by an easy application of the Borel–Cantelli Lemma.

For a closed interval  $J \subset \mathbb{R}$ , denote by  $\mathcal{M}(J)$  the space of Borel probability measures on  $J$  equipped with the Wasserstein metric.

Fix any  $\varphi \in C(X, \mathbb{R})$  and set  $M = \max(1, \max_{x \in X} |\varphi(x)|)$ . The continuous map  $\varphi : X \rightarrow \mathbb{R}$  induces a map on the space of Borel probability measures  $\varphi_* : \mathcal{M}(X) \rightarrow \mathcal{M}([-M, M])$ .

**Lemma 5.1.** *The map  $\varphi_* : \mathcal{M}(X) \rightarrow \mathcal{M}([-M, M])$  is continuous.*

*Proof.* Indeed, as  $X$  is compact and thus  $\varphi$  is uniformly continuous on  $X$ , there exists  $\delta > 0$  such that  $d_X(x, y) < \delta$  implies  $|\varphi(x) - \varphi(y)| < \frac{\varepsilon'}{2}$ . On the other hand, take  $\varepsilon'' := \frac{\varepsilon' \delta}{4M}$ . If for some coupling  $\gamma$  between two measures  $\nu, \nu'$  the integral (2) takes value less than  $\varepsilon''$ , then by Markov inequality

$$\gamma(\{(x, y) \mid d_X(x, y) > \delta\}) < \frac{\varepsilon''}{\delta} = \frac{\varepsilon'}{4M},$$

and hence, using the pushforward  $\gamma' := (\varphi, \varphi)_* \gamma$  as a coupling between  $\varphi_* \nu$  and  $\varphi_* \nu'$ , we have

$$\begin{aligned} \iint |u - v| d\gamma'(u, v) &= \iint_{X \times X} |\varphi(x) - \varphi(y)| d\gamma(x, y) \leq \\ &\leq \frac{\varepsilon'}{2} \cdot \gamma(\{(x, y) \mid d_X(x, y) \leq \delta\}) + 2M \cdot \gamma(\{(x, y) \mid d_X(x, y) > \delta\}) \\ &< \frac{\varepsilon'}{2} \cdot 1 + 2M \cdot \frac{\varepsilon'}{4M} = \frac{\varepsilon'}{2} + \frac{\varepsilon'}{2} = \varepsilon'. \end{aligned}$$

Thus, the inequality  $\text{dist}_{\mathcal{M}}(\nu, \nu') < \varepsilon''$  implies  $\text{dist}_{\mathcal{M}([-M, M])}(\varphi_* \nu, \varphi_* \nu') < \varepsilon'$ , which proves Lemma 5.1.  $\square$

Since  $\mathcal{M}(X)$  is compact, Lemma 5.1 implies that  $\varphi_*$  is uniformly continuous. Combined with the Standing Assumption this implies that for any  $\varepsilon' > 0$  there exists  $m$  such that for any  $n$  and for any initial measure  $\nu \in \mathcal{M}$ ,

$$(15) \quad \text{dist}_{\mathcal{M}([-M, M])}(\varphi_*(\mu_{n+m} * \dots * \mu_{n+1} * \nu), \varphi_* \nu_{n+m}) < \varepsilon',$$

where the sequence of measures  $\{\nu_i\}_{i \geq 0}$  is given by (4).

Now, fix  $\varepsilon > 0$ , and let us obtain the Large Deviations estimates for this  $\varepsilon$ . Fix  $m$  sufficiently large to make sure that (15) holds with  $\varepsilon' = \frac{\varepsilon^2}{30M}$ .

Since  $\varphi$  is bounded, it suffices to establish (5) for all  $n \geq 1$  that are divisible by  $m$ . For any  $n = mq$ ,  $q \in \mathbb{N}$ , decompose the iterations  $k = 1, \dots, mq$  into  $m$  arithmetic sequences  $k = r + jm$ ,  $j = 0, \dots, q-1$  with the difference  $m$ , indexed by the residue  $r = 1, \dots, m$ .

It suffices to check that Large Deviations estimate holds for each of these subsequences and  $\varepsilon > 0$  that we fixed above. Indeed,

$$\begin{aligned} \frac{1}{n} \left| \sum_{k=1}^n \varphi(f_k \circ \dots \circ f_1(x)) - \sum_{k=1}^n \int_X \varphi d\nu_k \right| &\leq \\ &\leq \frac{1}{m} \sum_{r=1}^m \left| \frac{1}{q} \left( \sum_{j=0}^{q-1} \varphi(f_{jm+r} \circ \dots \circ f_1(x)) - \sum_{j=0}^{q-1} \int_X \varphi d\nu_{jm+r} \right) \right|. \end{aligned}$$

Therefore, it suffices to prove that for each given  $r = 1, 2, \dots, m$  we have for all  $q \geq 1$  that

$$(16) \quad \mathbb{P} \left( \frac{1}{q} \left| \sum_{j=0}^{q-1} \varphi(f_{jm+r} \circ \dots \circ f_1(x)) - \sum_{j=0}^{q-1} \int_X \varphi d\nu_{jm+r} \right| > \varepsilon \right) < \frac{C}{m} e^{-\delta n},$$

since that will imply (5).

Fix  $r \in \{1, 2, \dots, m\}$ , and denote

$$y_j := x_{r+mj} = f_{r+mj} \circ \dots \circ f_1(x);$$

then, the first sum in (16) is given by  $\sum_{j=0}^{q-1} \varphi(y_j)$ .

To establish the estimate (16), we are going to compare this sum to a sum of *independent* random variables  $\xi_j$  that are distributed w.r.t.  $\varphi_* \nu_{jm+r}$  respectively. To do so, for the technical reasons of the construction, we will add two more random variables per iteration that are distributed uniformly on  $[0, 1]$  and that are

independent from the iterations and altogether. Namely, we consider

$$\Omega := \{(f_n, z_n, z'_n)_{n \in \mathbb{N}}\} = C(X, X)^{\mathbb{N}} \times [0, 1]^{\mathbb{N}} \times [0, 1]^{\mathbb{N}},$$

equipping it with the probability measure

$$\tilde{\mathbb{P}} := \mathbb{P} \times \text{Leb}^{\mathbb{N}} \times \text{Leb}^{\mathbb{N}},$$

where  $\mathbb{P}$  is defined by (1).

**Lemma 5.2.** *There exist random variables  $\xi_j, \xi'_j$  on the space  $(\Omega, \tilde{\mathbb{P}})$  such that:*

- $\xi_j$  are independent for all  $j$ , and the law of  $\xi_j$  is  $\varphi_* \nu_{r+jm}$ ;
- $\xi'_j$  are i.i.d Bernoulli variables, taking value 1 with the probability  $\frac{\varepsilon}{8M}$ .
- For every  $j$ , if  $\xi'_j$  takes value 0, then  $|\varphi(y_j) - \xi_j| \leq \frac{\varepsilon}{3}$ .

We postpone the (slightly technical) proof of this lemma until the end of this section. Note that it allows to complete the proof of Theorem 3.1. Indeed, note first that the last conclusion of Lemma 5.2 implies the inequality

$$(17) \quad |\varphi(y_j) - \xi_j| \leq \frac{\varepsilon}{3} + 2M\xi'_j;$$

notice that here we are using the fact that  $|\varphi| \leq M$ , and hence  $|\varphi(y_j) - \xi_j| \leq 2M$ .

The sum in (16) can be estimated as

$$(18) \quad \begin{aligned} & \frac{1}{q} \left| \sum_{j=0}^{q-1} \varphi(f_{jm+r} \circ \dots \circ f_1(x)) - \sum_{j=0}^{q-1} \int_X \varphi d\nu_{jm+r} \right| = \\ & = \frac{1}{q} \left| \sum_{j=0}^{q-1} \varphi(y_j) - \sum_{j=0}^{q-1} \mathbb{E}\xi_j \right| \leq \frac{1}{q} \sum_{j=0}^{q-1} |\varphi(y_j) - \xi_j| + \frac{1}{q} \left| \sum_{j=0}^{q-1} (\xi_j - \mathbb{E}\xi_j) \right| \leq \\ & \leq \frac{\varepsilon}{3} + 2M \cdot \frac{1}{q} \sum_{j=0}^{q-1} \xi'_j + \frac{1}{q} \left| \sum_{j=0}^{q-1} (\xi_j - \mathbb{E}\xi_j) \right| \end{aligned}$$

Now, note, that as  $\mathbb{E}\xi'_j = \frac{\varepsilon}{8M} < \frac{\varepsilon}{6M}$ , one has

$$\tilde{\mathbb{P}} \left( \frac{1}{q} \sum_{j=0}^{q-1} \xi'_j > \frac{\varepsilon}{6M} \right) < C_1 e^{-\delta_1 q}$$

for some  $\delta_1 > 0$  due to the Large Deviations estimate for the independent variables  $\xi'_j$ . In the same way, as  $\xi_j$  are independent and uniformly bounded by  $M$ , from the Large Deviations estimate one has

$$\tilde{\mathbb{P}} \left( \left| \frac{1}{q} \sum_{j=0}^{q-1} (\xi_j - \mathbb{E}\xi_j) \right| > \frac{\varepsilon}{3} \right) < C_2 e^{-\delta_2 q}$$

Now, if both  $\frac{1}{q} \sum_{j=0}^{q-1} \xi'_j \leq \frac{\varepsilon}{6M}$  and  $\frac{1}{q} \left| \sum_{j=0}^{q-1} (\xi_j - \mathbb{E}\xi_j) \right| < \frac{\varepsilon}{3}$  hold, (18) implies

$$\frac{1}{q} \left| \sum_{j=0}^{q-1} \varphi(y_j) - \sum_{j=0}^{q-1} \mathbb{E}\xi_j \right| \leq \frac{\varepsilon}{3} + 2M \cdot \frac{\varepsilon}{6M} + \frac{\varepsilon}{3} = \varepsilon.$$

Hence, the probability in (16) is bounded from above by the

$$\begin{aligned} \mathbb{P} \left( \frac{1}{q} \left| \sum_{j=0}^{q-1} \varphi(y_j) - \sum_{j=0}^{q-1} \mathbb{E} \xi_j \right| > \varepsilon \right) &\leq \\ &\leq \tilde{\mathbb{P}} \left( \frac{1}{q} \sum_{j=0}^{q-1} \xi'_j > \frac{\varepsilon}{6M} \right) + \tilde{\mathbb{P}} \left( \left| \frac{1}{q} \sum_{j=0}^{q-1} (\xi_j - \mathbb{E} \xi_j) \right| > \frac{\varepsilon}{3} \right) \\ &< C_1 e^{-\delta_1 q} + C_2 e^{-\delta_2 q}, \end{aligned}$$

and we obtain the desired exponentially small bound with  $\delta = \min(\delta_1, \delta_2)$ .

Since we have obtained the desired bound (16) that holds for each  $r = 1, \dots, m$ , this completes the proof of Theorem 3.1 modulo Lemma 5.2.

*Proof of Lemma 5.2.* On the space  $(\Omega, \tilde{\mathbb{P}})$ , define a sequence of  $\sigma$ -algebras

$$\mathcal{F}_j := \sigma(\{(f_k, z_k, z'_k) \mid k \leq r + jm\}).$$

Note that the random values  $\varphi(y_i) = \varphi(x_{r+mi})$ ,  $i = 0, \dots, j-1$  are measurable w.r.t.  $\mathcal{F}_{j-1}$ ; meanwhile, the conditional distribution of  $\varphi(y_j)$  with respect to this  $\sigma$ -algebra is given by

$$D_{j, y_{j-1}} := \varphi_*(\mu_{r+mj} * \dots * \mu_{r+1+m(j-1)} * \delta_{y_{j-1}}).$$

Due to the choice of  $m$ , this conditional distribution is  $\varepsilon'$ -close to  $\varphi_* \nu_{r+jm}$ , that is the law of the desired random variable  $\xi_j$ . This suggests to construct  $\xi_j$  and  $\xi'_j$  in the following way. Assume that

- (i) for every  $j$  the random variables  $\xi_j, \xi'_j$  are measurable w.r.t.  $\mathcal{F}_j$ ;
- (ii) the conditional distribution of  $\xi_j$  w.r.t.  $\{(f_k, z_k, z'_k) \mid k \leq r + (j-1)m\}$  is  $\varphi_* \nu_{r+jm}$ ;
- (iii) the conditional distribution of  $\xi'_j$  w.r.t.  $\{(f_k, z_k, z'_k) \mid k \leq r + (j-1)m\}$  is Bernoulli with the success probability  $\frac{\varepsilon}{8M}$ ;
- (iv) If  $|\varphi(y_j) - \xi_j| > \frac{\varepsilon}{3}$ , then  $\xi'_j$  takes value 1.

Then these random variables satisfy the conditions of Lemma 5.2. Indeed, property (ii) implies that  $\xi_j$  is independent from  $\sigma$ -algebra  $\mathcal{F}_{j-1}$ . As  $\xi_0, \dots, \xi_{j-1}$  are measurable w.r.t.  $\mathcal{F}_{j-1}$  due to the property (i), the random variable  $\xi_j$  is independent from  $\xi_0, \dots, \xi_{j-1}$ .

In the same way,  $\xi'_j$  is independent from  $\xi'_0, \dots, \xi'_{j-1}$  due to (i) and (iii). The last conclusion of the lemma will follow from (iv).

Let us now construct random variables  $\xi_j, \xi'_j$ , satisfying (i)–(iv).

Recall that between two probability measures on the real line, the optimal transport can be constructed explicitly. Namely, assume that we are given two probability measures  $\rho, \rho'$  on the real line. Take two independent random variables  $\eta, \zeta$ , where  $\eta$  is distributed w.r.t.  $\rho$ , and  $\zeta$  is uniformly distributed on  $[0, 1]$ . Let

$$\Phi_\rho(y, z) := \rho((-\infty, y)) + z \cdot \rho(\{y\});$$

then  $\Phi(\eta, \zeta)$  is a uniformly distributed random variable (this is the standard principle of substituting a random variable to its own distribution function, where the independent random variable is used to “dissolve the atoms”). On the other hand, an application of the function

$$\Psi_{\rho'}(s) = \sup\{y : \rho'((-\infty, y)) \leq s\},$$

that is an inverse to the distribution function of  $\rho'$ , sends the Lebesgue measure to  $\rho'$ . Hence, their composition,  $F_{(\rho, \rho')} := \Psi_{\rho'} \circ \Phi_{\rho}$ , provides a coupling between  $\rho$  and  $\rho'$ . Namely, the random variable  $\xi := F_{(\rho, \rho')}(\eta, \zeta)$  is distributed w.r.t.  $\rho'$ , and the desired coupling is given by the joint distribution  $(\eta, \xi)$ . This coupling minimizes the Wasserstein distance [V].

Now, for each  $j$  consider random variables

$$\xi_j := F_{(D_j, y_{j-1}, \varphi_* \nu_{jm+r})}(\varphi(y_j), z_{r+jm}).$$

Properties (i) and (ii) then are satisfied by construction.

Due to the optimality of this transport and the assumption on  $m$ , conditionally to any  $y_{j-1}$  we have

$$\mathbb{E}(|\xi_j - \varphi(y_j)| \mid y_{j-1}) = \text{dist}_{\mathcal{M}([-M, M])}(D_j, y_{j-1}, \varphi_* \nu_{jm+r}) \leq \varepsilon'.$$

Now, consider the event

$$A_j := \left\{ |\xi_j - \varphi(y_j)| > \frac{\varepsilon}{3} \right\}.$$

Due to Markov inequality, conditionally to any  $y_{j-1}$ , its probability does not exceed

$$\tilde{\mathbb{P}}(A_j \mid y_{j-1}) \leq \frac{3}{\varepsilon} \cdot \mathbb{E}(|\xi_j - \varphi(y_j)| \mid y_{j-1}) \leq \frac{3\varepsilon'}{\varepsilon} = \frac{3}{\varepsilon} \cdot \frac{\varepsilon^2}{30M} < \frac{\varepsilon}{8M}.$$

Now, define  $\xi'_j$  by

$$(19) \quad \xi'_j = \mathbf{1}_{A_j} + \mathbf{1}_{|\xi_j - \varphi(y_j)| \leq \frac{\varepsilon}{3}} \cdot \mathbf{1}_{z_{r+jm} \leq \psi(j, y_{j-1})},$$

where the value

$$\psi(j, y_{j-1}) := \frac{\frac{\varepsilon}{8M} - \tilde{\mathbb{P}}(A_j \mid y_{j-1})}{1 - \tilde{\mathbb{P}}(A_j \mid y_{j-1})} \in (0, 1)$$

is chosen in such a way that

$$\tilde{\mathbb{P}}(\xi'_j \mid y_{j-1}) = \frac{\varepsilon}{8M}$$

for any value of  $y_{j-1}$ . The choice (19) ensures both property (iii) (measurability follows from the explicit formulae) and (iv) (that follows directly from the construction).

This completes the construction of random variables  $\xi_j, \xi'_j$  and the verification of their properties, thus the proof of Lemma 5.2.  $\square$

## 6. CONCLUDING REMARKS

**6.1. Different phase spaces at different times.** A general remark is that in the non-stationary setting the phase spaces at different iterations are no longer canonically identified with each other. Namely, consider a random orbit

$$x_n = f_n(x_{n-1}).$$

An application of a time-dependent change of variables  $y_n = h_n(x_n)$ , where each  $h_n$  is a homeomorphism of  $X$ , sends an orbit of the initial system to

$$y_n = g_n(y_{n-1}), \quad g_n = h_n \circ f_n \circ h_{n-1}^{-1}.$$

This corresponds to the random iterations of another system from the same class, with measures  $\mu'_n = (T_n)_* \mu_n$  that are push-forward images of  $\mu_n$  under the “change of variable” maps

$$T_n : C(X, X) \rightarrow C(X, X), \quad T_n : f \mapsto h_n \circ f_n \circ h_{n-1}^{-1}.$$

However, the equality between some  $x_n$  and  $x_m$  does not imply the equality between their images  $y_n = h_n(x_n)$  and  $y_m = h_m(x_m)$ . This is why one can (and should) think of the spaces  $X$  at different moments of time as of *different* fibers.

Moreover, in the absence of stationarity we do not need to assume that the phase space stays the same all the time. Rather, we can consider a sequence of compact metric spaces  $X_0, X_1, X_2, \dots$ , and a sequence of probability measures  $\mu_n, n \geq 1$ , on  $C(X_{n-1}, X_n)$ :

$$X_0 \xrightarrow[f_1]{\mu_1} X_1 \xrightarrow[f_2]{\mu_2} X_2 \rightarrow \dots \rightarrow X_{n-1} \xrightarrow[f_n]{\mu_n} X_n \rightarrow \dots$$

Then, given a sequence of “observable” functions  $\varphi_n \in C(X_n, \mathbb{R})$ , we can ask the same questions as before.

It is reasonable to require a uniform bound on the diameters of  $X_n$ , and a uniform modulus of continuity for the functions  $\varphi_n$ ; Theorem 3.1 then can be generalized to such a setting:

**Theorem 6.1.** *Assume that the diameters of  $X_n$  are uniformly bounded, and that the Standing Assumption (with the replacement of  $X$  by corresponding  $X_n$ ’s), holds. Then for any sequence  $\varphi_n \in C(X_n, \mathbb{R})$  of functions admitting a uniform modulus of continuity and any  $x \in X_0$  we have that*

$$\frac{1}{n} \left| \sum_{k=1}^n \varphi_k(f_k \circ \dots \circ f_1(x)) - \sum_{k=1}^n \int_{X_k} \varphi_k d\nu_k \right| \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

where  $\nu_n = \mu_n * \nu_{n-1}$  for all  $n \geq 1$ , and  $\nu_0$  is any initial Borel probability measure on  $X_0$ .

Moreover, an analogue of the Large Deviations Theorem holds: for any  $\varepsilon > 0$  there exists  $C, \delta > 0$  such that for any  $x \in X_0$

$$\forall n \in \mathbb{N}, \quad \mathbb{P} \left( \frac{1}{n} \left| \sum_{k=1}^n \varphi_k(f_k \circ \dots \circ f_1(x)) - \sum_{k=1}^n \int_{X_k} \varphi_k d\nu_k \right| > \varepsilon \right) < C \exp(-\delta n).$$

The proof of Theorem 6.1 is almost verbatim repetition of the proof of Theorem 3.1, and as such will be omitted.

**6.2. Counter-examples: assumptions that cannot be avoided.** This section is devoted to presenting the examples where Nonstationary Random Ergodic Theorem does not hold. The first of these shows why one has to require at least some upper bound on the number of iterations needed to “diffuse” the initial measure in the Standing Assumption:

**Example 6.2.** *Let the set  $X = \{a, b\}$  consist of two points only. Fix a (sufficiently quickly growing) sequence  $n_k = 10^{k^2}$ , and let measure  $\mu_n$  be defined as  $\mu_n = \delta_{\text{id}}$  for  $n \neq n_k$  and*

$$\mu_{n_k} = \frac{1}{2} \delta_{\text{id}} + \frac{1}{2} \delta_{\sigma},$$

where  $\sigma$  is a transposition that interchanges  $a$  and  $b$ . Then, on one hand, for any initial  $n$ , any two measures  $\nu, \nu'$  on  $X$ , and any  $n_k > n$ , one has

$$\mu_{n_k} * \dots * \mu_{n+1} * \nu' = \mu_{n_k} * \dots * \mu_{n+1} * \nu = \frac{1}{2} \delta_a + \frac{1}{2} \delta_b =: \bar{\nu}.$$

On the other hand, any individual orbit  $(x_n)$  spends each interval of time between  $n_k$  and  $n_{k+1}$  either fully at  $a$ , or fully at  $b$ , with each of these probabilities occurring

equiprobably and independently for different  $k$ . As the quotient  $\frac{n_{k+1}-n_k}{n_k}$  tends to infinity, this easily implies that the time averages of any function  $\varphi$  almost surely have both  $\varphi(a)$  and  $\varphi(b)$  as its accumulation points. In particular, these time averages (for non-constant  $\varphi$ ) do not converge.

The next example shows why in the Standing Assumption we avoid averaging in time. Actually, the absence of the natural identification between the phase spaces at different times (discussed in the previous section) already suggests that time-average of images of a measure is not a good object to be considered: that would require an addition of measures on the different spaces. However, even if we are dealing with a uniquely ergodic deterministic dynamical system, so that such an addition can be considered, the Nonstationary Ergodic Theorem for  $\varphi_n$  depending on  $n$  without the Standing Assumption may not hold:

**Example 6.3.** Let  $X = S^1 = \mathbb{R}/\mathbb{Z}$ , and  $f(x) = x + \alpha \bmod 1$  be an irrational rotation. Then, the (classical) dynamical system  $(X, f)$  is uniquely ergodic (its unique invariant measure is Lebesgue measure); however, the Standing Assumption does not hold (there is no randomness, and the distances between the orbits do not decrease). At the same time, Cesaro averages of the images of any two initial measures on the circle converge to Lebesgue measure with some uniform rate.

Now, take any function  $\varphi \in C(X)$  and consider the family  $\varphi_n = \varphi \circ f^{-n}$ . This family of functions is equicontinuous on  $X$ . At the same, the conclusion of Theorem 6.1 does not hold for these functions: for any initial point  $x \in X$  one has

$$\frac{1}{n} \sum_{k=1}^n \varphi_k(f_k \circ \dots \circ f_1(x)) = \frac{1}{n} \sum_{k=1}^n \varphi \circ f^{-k}(f^k(x)) = \varphi(x),$$

so there is no constant-like behaviour. Also, replacing the choice of functions by

$$\varphi_n = \varphi \circ f^{-n+r(n)},$$

where  $r(n) := \max\{k \mid n > n_k\}$  for a fast-growing sequence  $n_k = 10^{k^2}$ , similar to Example 6.2, one gets the absence of a limit of “non-stationary” time averages for every initial point  $x$ .

#### ACKNOWLEDGMENTS

We are grateful to Andrew Török for his remarks on the preliminary draft of this paper.

#### REFERENCES

- [Ac] A. de Acosta, Large deviations for Markov chains, *Cambridge Tracts in Math.*, 229 Cambridge University Press, Cambridge, 2022, xii+249 pp.
- [A] V. A. Antonov, Modeling of processes of cyclic evolution type. Synchronization by a random signal. *Vestnik Leningrad. Univ. Mat. Mekh. Astronom.* 1984, no. 2, pp. 67–76.
- [BB] D. Berend, V. Bergelson, Ergodic and mixing sequences of transformations, *Ergodic Theory and Dynamical Systems*, **4** (1984), pp. 353–366.
- [CR] J.-P. Conze, A. Raugi, Limit theorems for sequential expanding dynamical systems on  $[0, 1]$ , *Ergodic theory and related fields*, *Contemp. Math.*, **430**, Amer. Math. Soc., Providence, RI, 2007, pp. 89–121.
- [DK] B. Deroin, V. Kleptsyn, Random conformal dynamical systems, *Geom. Funct. Anal.* **17** (2007), no. 4, pp. 1043–1105.
- [DKN] B. Deroin, V. Kleptsyn, A. Navas, On the question of ergodicity for minimal group actions on the circle, *Moscow Mathematical Journal*, **9** (2009), pp. 263–303.



- [DV] M.D. Donsker, S.R.S. Varadhan, Asymptotic evaluation of certain Markov process expectations for large time. I. II. *Comm. Pure Appl. Math.* **28** (1975), 1–47; **28** (1975), pp. 279–301.
- [DMPS] R. Douc, E. Moulines, P. Priouret, P. Soulier, Markov chains, *Springer Ser. Oper. Res. Financ. Eng.*, Springer, Cham, 2018, xviii+757 pp.
- [DK1] P. Durte, S. Klein, Lyapunov exponents of linear cocycles, *Atlantis Stud. Dyn. Syst.*, 3 Atlantis Press, Paris, 2016, xiii+263 pp.
- [DK2] P. Durte, S. Klein, Continuity of the Lyapunov exponents of linear cocycles *Publ. Mat. IMPA 31 o Coloq. Bras. Mat. Instituto Nacional de Matematica Pura e Aplicada (IMPA)*, Rio de Janeiro, 2017, ii+142 pp.
- [FL] D. Feng, K. Lau, Multifractal formalism for self-similar measures with weak separation condition, *J. Math. Pures Appl.* **9** (2009), pp. 407–428.
- [FS] D. Feng, K. Simon, Dimension estimates for  $C^1$  iterated function systems and  $C^1$  repellers, a survey, *Thermodynamic formalism*, pp. 421–467, Lecture Notes in Math., 2290, CIRM Jean-Morlet Ser., Springer, Cham, 2021.
- [Hu] J. Hutchinson, Fractals and self-similarity, *Indiana Univ. Math. J.* **30** (1981), pp. 713–747.
- [FurK] H. Furstenberg, H. Kesten, Products of random matrices, *Ann. Math. Statist.* **31** (1960), pp. 457–469.
- [FurKif] H. Furstenberg, Y. Kifer, Random matrix products and measures on projective spaces, *Israel Journal of Mathematics* **46** (1983), pp. 12–32.
- [Ga] J. Gartner, On Large Deviations from the Invariant Measure, *Theory of Probability and Its Applications* **22** (1977), pp. 24–39.
- [G] I. Goldsheid, Exponential growth of products of non-stationary Markov-dependent matrices, *International Mathematics Research Notices*, volume 2022, issue 8, April 2022, pp. 6310–6346.
- [GK1] A. Gorodetski, V. Kleptsyn, Parametric Furstenberg Theorem on Random Products of  $SL(2, \mathbb{R})$  matrices, *Adv. Math.* **378** (2021), Paper No. 107522, 81 pp.
- [GK2] A. Gorodetski, V. Kleptsyn, Non-stationary version of Furstenberg Theorem on random matrix products, preprint, arXiv:2210.03805
- [GK3] A. Gorodetski, V. Kleptsyn, Non-stationary versions of Anderson Localization and Parametric Furstenberg Theorem, work in progress.
- [GKM] A. Gorodetski, V. Kleptsyn, G. Monakov, Hölder regularity of stationary measures, arXiv:2209.12342
- [KN] V. Kleptsyn, M. Nalskii, Convergence of orbits in random dynamical systems on a circle, *Funct. Anal. Appl.* vol. 38 (2004), no. 4, pp. 267–282.
- [HNTV] N. Haydn, M. Nicol, A. Török, S. Vaienti, Almost sure invariance principle for sequential and non-stationary dynamical systems, *Trans. Amer. Math. Soc.* **369** (2017), pp. 5293–5316.
- [KS] L. Korolov, Ya. Sinai, Theory of Probability and Random Processes, *Springer-Verlag Berlin Heidelberg*, Universitext, 2007, xii+353 pp.
- [La] Thierry de La Rue, Joinings in Ergodic Theory, *Encyclopedia of Complexity and Systems Science*, Springer, Berlin, Heidelberg 2020.
- [Mal] D. Malicet, Random walks on  $\text{Homeo}(S^1)$ , *Comm. Math. Phys.* **356** (2017), pp. 1083–1116.
- [PSS] Y. Peres, W. Schlag, B. Solomyak, Sixty years of Bernoulli convolutions, *Fractal geometry and stochastics, II* (Greifswald/Koserow, 1998), pp. 39–65, Progr. Probab., 46, Birkhäuser, Basel, 2000.
- [Sz] T. Szarek, Invariant measures for iterated function systems, *Ann. Polon. Math.* **75** (2000), pp. 87–98.
- [V] S. Vallander, Calculation of the Wasserstein distance between probability distributions on the line, *Theory of Probability and its Applications*, **18** (1974), pp. 784–786.
- [Va] P. Varju, Recent progress on Bernoulli convolutions, *European Congress of Mathematics*, pp. 847–867, *Eur. Math. Soc.*, Zürich, 2018

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

*Email address:* `asgor@uci.edu`

CNRS, INSTITUTE OF MATHEMATICAL RESEARCH OF RENNES, IRMAR, UMR 6625 DU CNRS

*Email address:* `victor.kleptsyn@univ-rennes1.fr`