TRANSACTIONS OF THE
AMERICAN MATHEMATICAL SOCIETY, SERIES B
Volume 8, Pages 105–129 (February 17, 2021)
https://doi.org/10.1090/btran/60

ENTROPY AND DIMENSION OF DISINTEGRATIONS OF STATIONARY MEASURES

PABLO LESSA

Abstract. We extend a result of Ledrappier, Hochman, and Solomyak on exact dimensionality of stationary measures for $\mathrm{SL}_2(\mathbb{R})$ to disintegrations of stationary measures for $\mathrm{GL}(\mathbb{R}^d)$ onto the one dimensional foliations of the space of flags obtained by forgetting a single subspace.

The dimensions of these conditional measures are expressed in terms of the gap between consecutive Lyapunov exponents, and a certain entropy associated to the group action on the one dimensional foliation they are defined on. It is shown that the entropies thus defined are also related to simplicity of the Lyapunov spectrum for the given measure on $\mathrm{GL}(\mathbb{R}^d)$.

1. Introduction

It was shown by Ledrappier [Led84], Hochman and Solomyak [HS17], that if ν is a probability on the projective space of $\mathbb{R}^2$ which is stationary with respect to a probability μ on $\mathrm{SL}_2(\mathbb{R})$ with finite Lyapunov exponents, then ν is exact dimensional and its dimension is $\frac{\kappa}{2\chi}$ where κ is the Furstenberg entropy and χ is the largest Lyapunov exponent (hence 2χ is the gap between the two Lyapunov exponents).

Suppose now that μ is a probability on $\mathrm{SL}_3(\mathbb{R})$ and ν is a μ-stationary probability on the space of flags in $\mathbb{R}^3$ (i.e. pairs (L, P) where $L \subset P$, L is a one dimensional subspace, and P is a two dimensional subspace), which is a three-dimensional manifold.

We consider here the two foliations of the space of flags obtained by partitioning into sets of flags sharing the same one dimensional subspace on the one hand, and flags sharing the same two dimensional subspace on the other. These are foliations by circles, and furthermore the action of any invertible linear self mapping of $\mathbb{R}^3$ preserves both foliations.

In this context we show that the conditional measures obtained by disintegrating ν with respect to these two foliations, are exact dimensional. Furthermore we express the dimension of these disintegrations in terms of the gap between consecutive Lyapunov exponents as well as two entropies κ_1, κ_2. Before establishing the dimension formula we show that the entropies κ_i bound the gaps between exponents from below and therefore, in principle, yield a criteria for simplicity of the Lyapunov spectrum.

We prove our results in a slightly more general context, that of actions of $\mathrm{GL}(\mathbb{R}^d)$ on the space of complete flags in $\mathbb{R}^d$. In this context there are $d-1$ associated one

Received by the editors August 5, 2019, and, in revised form, July 14, 2020, and November 13, 2020.
2020 *Mathematics Subject Classification.* Primary 37F35.

©2021 by the author under Creative Commons Attribution-Noncommercial 3.0 License (CC BY NC 3.0)

dimensional foliations which correspond to "forgetting" the i-dimensional subspace of all flags for some $i \in \{1, \ldots, d-1\}$.

1.1. **Preliminaries.** Let $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_d(A) > 0$ denote the singular values of an element $A \in \mathrm{GL}(\mathbb{R}^d)$ with respect to the standard inner product.

We denote by $\mathrm{Flags}(\mathbb{R}^d)$ the space of complete flags in $\mathbb{R}^d$, an element $F \in \mathrm{Flags}(\mathbb{R}^d)$ is of the form $F = (S_0, S_1, \ldots, S_d)$ where S_i is an i-dimensional subspace of $\mathbb{R}^d$ for each $i = 0, \ldots, d$ and $S_i \subset S_{i+1}$ for $i = 0, \ldots, d-1$.

Let $\mathrm{Flags}_i(\mathbb{R}^d)$ denote the space of flags missing their i-dimensional subspace. For a given complete flag $F = (S_0, \ldots, S_d)$ we denote by F_i its projection to $\mathrm{Flags}_i(\mathbb{R}^d)$ (i.e. the sequence obtained by removing S_i from F).

We use the notation $X \overset{(d)}{=} Y$ for equality in distribution between random elements X and Y. And $\nu_1 \ll \nu_2$ to mean that the probability ν_1 is absolutely continuous with respect to ν_2.

If X and Y are random elements taking values in complete separable metric spaces (a version of) the conditional distribution of X given Y is a $\sigma(Y)$-measurable random probability ν_Y on the range of X such that

$$\int f(x) d\nu_Y(x) = \mathbb{E}\left(f(X)|Y\right)$$

for all continuous bounded real functions (here the right-hand side is the conditional expectation of $f(X)$ with respect to the σ-algebra generated by Y). Such a conditional distribution is well defined up to sets of zero measure but we will abuse notation slightly referring to 'the conditional distribution'.

It is always the case that there exists a Borel mapping $y \mapsto \nu(y)$ from the range of Y to the space of probabilities on the range of X such that $\nu(Y)$ is a version of the conditional distribution of X given Y. Fixing such a mapping one may speak of ν_y for y non-random in the range of Y.

The lower local dimension of a probability measure ν on a metric space at a point x is defined by

$$\underline{\dim}_x(\nu) = \liminf_{r \to 0} \frac{\log\left(\nu(B_r(x)\right)}{\log(r)},$$

while the upper local dimension is defined by

$$\overline{\dim}_x(\nu) = \limsup_{r \to 0} \frac{\log\left(\nu(B_r(x)\right)}{\log(r)},$$

where $B_r(x)$ is the ball of radius r centered at x.

If the lower and upper dimensions of ν are equal to the same constant ν-almost everywhere then we say that ν is exact dimensional and define its global dimension $\dim(\nu)$ as the given constant.

1.2. **Statement of main results.** Suppose that A is a random element of $\mathrm{GL}(\mathbb{R}^d)$ with distribution μ such that

$$\mathbb{E}\left(\left|\log\left(\sigma_i(A)\right)\right|\right) < +\infty \text{ for } i = 1, \ldots, d,$$

and let $F = (S_0, \ldots, S_d)$ be a random element of $\mathrm{Flags}(\mathbb{R}^d)$ with distribution ν which is independent from A and such that

$$F \overset{(d)}{=} AF.$$

The existence of such a pair (A, F) is equivalent to the fact that ν is a μ-stationary probability, as first defined in [Fur63].

The Lyapunov exponents $\chi_1, \ldots, \chi_d$ of μ relative to ν are defined by the equations

$$\chi_1 + \cdots + \chi_i = \mathbb{E}\left(\log\left(|\det_{S_i}(A)|\right)\right) \text{ for } i = 1, \ldots, d,$$

where $|\det_S(A)|$ is the Jacobian of the restriction of A to the subspace S (where the volume measure induced by standard inner product is used on S and its image). In the degenerate case where $S = \{0\}$ one has $|\det_S(A)| = 1$, and if S is one dimensional one has $|\det_S(A)| = \|A_{|S}\|$.

The Lyapunov exponents given by the multiplicative ergodic theorem of [Ose68] for a product of i.i.d. random matrices of distribution μ are obtained by maximizing the sums $\chi_1 + \cdots + \chi_i$ over all stationary probabilities ν as shown in [FK83].

Fix $i \in \{1, \ldots, d-1\}$, let ν_i be the projection of ν to $\mathrm{Flags}_i(\mathbb{R}^d)$, and let ν_{F_i} be the conditional distribution of F given F_i.

Theorem 1 (Inequality between entropy and gap between exponents)**.** *If ν is the unique stationary probability on $\mathrm{Flags}(\mathbb{R}^d)$ which projects to ν_i then $A\nu_{F_i} \ll \nu_{AF_i}$ almost surely,*

$$0 \le \kappa_i = \mathbb{E}\left(\log\left(\frac{dA\nu_{F_i}}{d\nu_{AF_i}}(AF)\right)\right) \le \chi_i - \chi_{i+1},$$

and $\kappa_i = 0$ if and only if $A\nu_{F_i} = \nu_{AF_i}$ almost surely.

Theorem 2 (Dimension of conditional measures)**.** *If ν is ergodic, is the unique stationary probability on $\mathrm{Flags}(\mathbb{R}^d)$ which projects to ν_i, and $\kappa_i > 0$, then almost surely ν_{F_i} is exact dimensional and*

$$\dim(\nu_{F_i}) = \frac{\kappa_i}{\chi_i - \chi_{i+1}}.$$

In the case $d = 2$ both theorems above are known. A proof of Theorem 1 in this case was first given in [Led84]. In the same work the formula for dimension in Theorem 2 is shown to hold for a slightly different notion of dimension. The exact dimensionality of stationary measures when $d = 2$ was first proved in [HS17] and this implies the formula above for the same notion of dimension we use here.

Theorem 1 implies that the Lyapunov spectrum is simple (i.e. all exponents are different) if there does not exist a family of conditional probabilities $F_i \mapsto \nu_{F_i}$ satisfying $A\nu_{F_i} = \nu_{AF_i}$ for μ almost every A. This suggests a connection to criteria for simplicity dating back to [GdM89] and [GR89] though we do not explore this issue further here.

Part 1. Entropy, mutual information, and Lyapunov exponent gaps

2. Entropy and mutual information

We will define below $I(A, AF|AF_i)$ the conditional mutual information between A and AF given AF_i. This is a non-negative $\sigma(AF_i)$-measurable random variable which may take the value $+\infty$.

The purpose of this section is to prove that:

Lemma 1 (Entropy and mutual information)**.** *If $I(A, AF|AF_i) < +\infty$ almost surely then $A\nu_{F_i} \ll \nu_{AF_i}$ almost surely and $\kappa_i = \mathbb{E}\left(I(A, AF|AF_i)\right)$.*

Conversely, if $A\nu_{F_i} \ll \nu_{AF_i}$ almost surely then $\kappa_i = \mathbb{E}\left(I(A, AF|AF_i)\right)$ whether κ_i is finite or not.

This result reduces the problem of showing that $A\nu_{F_i} \ll \nu_{AF_i}$ almost surely and that $0 \le \kappa_i < +\infty$ to that of bounding the conditional mutual information between A and AF given AF_i.

A general reference covering mutual information including Dobrushin's theorem and the Gelfand-Yaglom-Perez theorem is [Pin64].

2.1. **Conditional mutual information.**

2.1.1. *Mutual information.* Let X and Y be random elements of two Polish spaces $\mathcal{X}$ and $\mathcal{Y}$, and denote $\mu_X, \mu_Y, \mu_{(X,Y)}$ the distribution of X, Y, and (X,Y) respectively.

The mutual information between X and Y is defined by

$$I(X,Y) = \sup \sum_{A \in P} \log\left(\frac{\mu_{(X,Y)}(A)}{(\mu_X \times \mu_Y)(A)}\right)\mu_{(X,Y)}(A)$$

where the supremum is over all finite partitions P of $\mathcal{X} \times \mathcal{Y}$ into Borel sets.

Directly from the definition one sees that $I(X,Y) = I(Y,X)$.

By Jensen's inequality $0 \le I(X,Y) \le +\infty$ with equality to 0 if and only if X and Y are independent. If X takes countably many values and has finite entropy $H(X)$ in the sense of [Sha48] one has $I(X,Y) \le H(X)$.

It was shown in [Dob59] that $I(X,Y)$ is the supremum over any sequence of partitions which generate the Borel σ-algebra in $\mathcal{X} \times \mathcal{Y}$ (see also [Gra11, Lemma 7.3]). This has the following important corollary:

Proposition 1 (Semi-continuity of mutual information). *If* $\lim_{n\to+\infty}(X_n, Y_n) = (X,Y)$ *in the sense of distributions then* $I(X,Y) \le \liminf_{n\to+\infty} I(X_n, Y_n)$.

It was shown in [GfY59] and [Per59] that if $I(X,Y) < +\infty$ then $\mu_{(X,Y)} \ll \mu_X \times \mu_Y$ and

$$I(X,Y) = \mathbb{E}\left(\log\left(\frac{d\mu_{(X,Y)}}{d(\mu_X \times \mu_Y)}(X,Y)\right)\right).$$

Conversely, if $\mu_{(X,Y)} \ll \mu_X \times \mu_Y$ then

$$I(X,Y) = \mathbb{E}\left(\log\left(\frac{d\mu_{(X,Y)}}{d(\mu_X \times \mu_Y)}(X,Y)\right)\right),$$

whether the right hand side is finite or not.

These results are usually called the Gelfand-Yaglom-Perez Theorem.

In our context, when $d = 2$, this yields the following result:

Proposition 2. *If $d = 2$ and $I(A, AF) < \infty$ then $A\nu \ll \nu$ almost surely and* $0 \le \kappa = \mathbb{E}\left(\log\left(\frac{dA\nu}{d\nu}(AF)\right)\right) = I(A, AF) < +\infty$.

Conversely, if $A\nu \ll \nu$ almost surely then $\kappa = I(A, AF)$ whether $I(A, AF)$ is finite or not.

Proof. The marginal distributions of (A, AF) are μ and ν respectively. However the conditional distribution of AF given A is $A\nu$.

Therefore letting m be the joint distribution of (A, AF) one has

$$\int f(a,x)dm(a,x) = \int\int f(a,x)da\nu(x)d\mu(a),$$

for all measurable functions f.

If $A\nu \ll \nu$ almost surely then

$$\int f(a,x)dm(a,x) = \int\int f(a,x)\frac{da\nu}{d\nu}(x)d\nu(x)d\mu(a)$$
$$= \int f(a,x)\frac{da\nu}{d\nu}(x)d(\mu\times\nu)(a,x),$$

so that $\frac{dm}{d(\mu\times\nu)}(a,x) = \frac{da\nu}{d\nu}(x)$ at $(\mu\times\nu)$-almost every point (a,x).

On the other hand if $m \ll (\mu\times\nu)$ then setting $g(a,x) = \frac{dm}{d(\mu\times\nu)}(a,x)$ one has

$$\int f(a,x)dm(a,x) = \int\int f(a,x)da\nu(x)d\mu(a) = \int\int f(a,x)g(a,x)d\nu(x)d\mu(a),$$

for all measurable funtions f.

Letting $f(a,x) = \mathbb{1}_A(a)h(x)$ where $\mathbb{1}_A$ is the indicator of an arbitrary subset A of $\mathrm{GL}(\mathbb{R}^d)$, and h is continuous on the compact space $\mathrm{Flags}(\mathbb{R}^2)$, one obtains that

$$\int h(x)da\nu(x) = \int h(x)g(a,x)d\nu(x),$$

for μ-almost every a. Intersecting the μ-full measure sets where this holds over a countable dense set of functions h, one obtains a full measure set for μ where $\frac{da\nu}{d\nu}(x) = g(a,x)$.

Hence, the distribution of (A, AF) is absolutely continuous with respect to $\mu\times\nu$ if and only if $A\nu \ll \nu$ almost surely and in this case the Radon-Nikodym derivative between the two at (A, AF) is given by $\frac{dA\nu}{d\nu}(AF)$. □

2.1.2. *Conditional mutual information.* Let $\mathcal{F}$ be a σ-algebra of measurable sets in the probability space on which the random elements X and Y are defined.

The mutual information between X and Y conditioned on $\mathcal{F}$ is the unique up to modifications on null sets random variable $I(X,Y|\mathcal{F})$ obtained as above but using the conditional distribution of (X,Y) conditioned on $\mathcal{F}$. In the case $\mathcal{F} = \sigma(Z_1, Z_2, \ldots, Z_k)$ we use the notation $I(X,Y|Z_1,Z_2,\ldots,Z_k) = I(X,Y|\mathcal{F})$.

One still has $0 \le I(X,Y|\mathcal{F}) = I(Y,X|\mathcal{F}) \le +\infty$ almost surely. Almost sure equality to zero occurs if and only if X and Y are conditionally independent given $\mathcal{F}$.

In general there is no relation between $I(X,Y)$ and $I(X,Y|\mathcal{F})$ or even $\mathbb{E}\left(I(X,Y|F)\right)$.

To see this suppose for example that X, Y are i.i.d. taking the values ±1 with probability $1/2$ and $Z = XY$, then one has $I(X,Y) = 0$ while $I(X,Y|Z) = \log(2)$ almost surely.

On the other hand for any Markov chain X_1, X_2, X_3 one has $I(X_1, X_3|X_2) = 0$ almost surely, and one may construct examples with $I(X_1,X_3) > 0$. For example, setting $X_1 = Y_1, X_2 = Y_1 + Y_2$ and $X_3 = Y_1 + Y_2 + Y_3$ where the Y_i are i.i.d. with $\mathbb{P}(Y_i = \pm1) = 1/2$ suffices.

The following semi-continuity property holds:

Proposition 3 (Semi-continuity of conditional mutual information)**.** *If the conditional distribution of (X_n, Y_n) given $\mathcal{F}$ converges almost surely to the conditional distribution of (X,Y) given $\mathcal{F}$ then $I(X,Y|\mathcal{F}) \le \liminf_{n\to+\infty} I(X_n,Y_n|\mathcal{F})$ almost surely.*

Proof. This is a direct consequence of Proposition 1. □

The following monotonicity property follows immediately from the definition of mutual information

$$I(X,Y|\mathcal{F}) \leq I(X,(Y,Z)|\mathcal{F}).$$

A more precise version of monotonicity is the following:

Proposition 4 (Chain rule for conditional mutual information)**.** *If X,Y,Z are random elements and $\mathcal{F}$ a σ-algebra of events of the probability space on which they are defined, then*

$$I(X,(Y,Z)|\mathcal{F}) = \mathbb{E}\left(I(X,Y|Z,\mathcal{F})|\mathcal{F}\right) + I(X,Z|\mathcal{F}).$$

Proof. When $\mathcal{F}$ is trivial this is [Gra11, Corollary 7.14] (notice that what said reference denotes by $I(X,Y|Z)$ is $\mathbb{E}\left(I(X,Y|Z)\right)$ in our notation). The general case follows by applying this to the conditional distributions given $\mathcal{F}$. □

2.2. **Proof of Lemma 1.** We will calculate the marginal distributions and the joint distribution of (A,AF) conditioned on AF_i and apply the Gelfand-Yaglom-Perez Theorem as in Proposition 2.

To begin we simply let μ_{AF_i} be the conditional distribution of A given AF_i.

By stationarity of ν the conditional distribution of AF given AF_i is ν_{AF_i}.

For the joint distribution notice that the distribution of AF conditioned on $\sigma(A,AF_i)$ is the same as conditioned on $\sigma(A,F_i)$ and therefore it is $A\nu_{F_i}$.

Hence the joint conditional distribution of (A,AF) given AF_i satisfies (and is determined by the equation)

$$\mathbb{E}\left(f(A,AF)|AF_i\right) = \int\int f(a,x)da\nu_{F_i}(x)d\mu_{AF_i}(a)$$

for all continuous bounded f.

By the Gelfand-Yaglom-Perez Theorem if $I(A,AF|AF_i) < +\infty$ almost surely then $A\nu_{F_i} \ll \nu_{AF_i}$ almost surely and

$$\mathbb{E}\left(\log\left(\frac{dA\nu_{F_i}}{d\nu_{AF_i}}(AF)\right)|AF_i\right) < +\infty$$

almost surely.

And conversely, if $A\nu_{F_i} \ll \nu_{AF_i}$ almost surely one has

$$I(A,AF|AF_i) = \mathbb{E}\left(\log\left(\frac{dA\nu_{F_i}}{d\nu_{AF_i}}(AF)\right)|AF_i\right).$$

The result now follows by taking expectation.

3. Proof of Theorem 1

In this section we will prove Theorem 1.

The strategy is to approximate (A,F) by pairs with the property that the conditional distributions ν_{F_i} are absolutely continuous with respect to the natural geometric measure on their domain of definition.

For the approximating pairs there is a direct relation between the distortion of the conditional measures by a linear mapping A and its determinants on certain subspaces. This argument establishes equality between the entropy κ_i and the Lyapunov exponent gap $\chi_i - \chi_{i+1}$ for the approximating pairs.

The result is then obtained by passing to the limit using the properties of conditional mutual information discussed in the previous section. At this step equality is lost, and one obtains only an inequality between entropy and the Lyapunov exponent gap.

An important technical issue is that one must maintain the same conditioning σ-algebra for the approximating pairs and the limit pair (A, F) in order to apply Proposition 3.

The idea of approximating a probability μ by one whose stationary probability is absolutely continuous with respect to the natural geometric measure is already present in [Fur63, Theorem 8.6].

3.1. **Jacobians of linear actions on flags.** We will now briefly, for the duration of this subsection, abandon the context where A and F are random satisfying $AF \overset{(d)}{=} F$ in order to discuss a result for a deterministic transformation A and flag F.

Denote the mapping $F \mapsto F_i$ which removes from each flag in $\mathrm{Flags}(\mathbb{R}^d)$ its i-dimensional subspace by π_i, and notice that the fibers $\mathrm{Flags}_{F_i}(\mathbb{R}^d) = \pi_i^{-1}(F_i)$ are 1-dimensional. We consider on each $\mathrm{Flags}_{F_i}(\mathbb{R}^d)$ the the unique probability measure η_{F_i} which is invariant under the action of orthogonal transformations which fix F_i.

Notice that any element $A \in \mathrm{GL}(\mathbb{R}^d)$ leaves the family of measures η_{F_i} quasi-invariant. We will need the explicit Jacobian of the action of A on this family of measures.

Lemma 2. *If* $A \in \mathrm{GL}(\mathbb{R}^d)$*,* $F = (S_0, S_1, \ldots, S_d) \in \mathrm{Flags}(\mathbb{R}^d)$*, and* $i \in \{1, \ldots, d-1\}$*, then*

$$\frac{dA\eta_{F_i}}{d\eta_{AF_i}}(AF) = \frac{|\det_{S_i}(A)|^2}{|\det_{S_{i-1}}(A)||\det_{S_{i+1}}(A)|}.$$

Proof. We begin by proving the case $d = 2$ (this case is included in the statement of [Fur63, Lemma 8.8] though the proof is omitted there).

In this case $F = (S_0, S_1, S_2)$ and the only non-trivial subspace is S_1 which has dimension 1 in $\mathbb{R}^2$. Therefore, we are looking to calculate the Jacobian of the action of A on the projective space of lines in $\mathbb{R}^2$ at the line S_1 with respect to the unique rotationally invariant probability η.

For this purpose consider a unit length vector $v \in S_1$ and an orthogonal vector w of length δ. Let R be the rectangle $\{sv + tw : s, t \in [0, 1]\}$.

Since we are considering the action of A on projective space, it is equivalent to consider the transformation $B = A/|\det_{S_1}(A)| = A/|Av|$ so that Bv has length one.

Notice that BR is a paralelogram with a side in AS_1 of length 1, and area ϵ which is the length of the orthogonal projection of BR onto the subspace orthogonal to AS_1. Calculating the determinant of B one obtains explicitly

$$\epsilon = |\det(B)|\delta = \frac{|\det(A)|}{|\det_{S_1}(A)|^2}\delta.$$

Taking the limit as $\epsilon \to 0$ we obtain that the derivative of the action of A on projective space at the point S_1 is $\frac{|\det(A)|}{|\det_{S_1}(A)|^2}$ from which it follows that

$$\frac{dA\eta}{d\eta}(AS_1) = \frac{|\det_{S_1}(A)|^2}{|\det(A)|}$$

as claimed.

We will now show that the general case may be reduced to the two dimensional case.

For this purpose suppose now that $d > 2$, $F = (S_0, \ldots, S_d)$, and $i \in \{1, \ldots, d-1\}$.

Notice that the quotient space S_{i+1}/S_{i-1} is two dimensional and inherits an inner product from $\mathbb{R}^d$ which makes it isometric to the orthogonal complement of S_{i-1} within S_{i+1}. The same is true for AS_{i+1}/AS_{i-1}.

Therefore, letting $B : S_{i+1}/S_{i-1} \to AS_{i+1}/AS_{i-1}$ be the linear map induced by A one has

$$\frac{dA\eta_{F_i}}{d\eta_{AF_i}}(AF) = \frac{|\det_{S_i}(B)|^2}{|\det(B)|},$$

where on the right hand side the space S_i is considered as a one-dimensional subspace of S_{i+1}/S_{i-1}.

The result follows from the observation that $|\det(B)| = |\det_{S_{i+1}}(A)|/|\det_{S_{i-1}}(A)|$ and $|\det_{S_i}(B)| = |\det_{S_i}(A)|/|\det_{S_{i-1}}(A)|$. □

3.2. **Proof of Theorem 1.** We return now to the notation and context of the statement of Theorem 1. In particular A and $F = (S_0, \ldots, S_d)$ are independent random elements with distribution μ and ν respectively and such that $AF \overset{(d)}{=} F$. Recall that ν_i is the projection of ν onto $\mathrm{Flags}_i(\mathbb{R}^d)$ and ν_{F_i} is the conditional distribution of F given F_i.

3.2.1. *Representation.* Since the statement of the theorem only depends on the joint distribution of (A, F) we are at liberty to change (A, F) to any other pair with the same distribution.

For this purpose fix a Borel mapping $(u, m) \mapsto \rho(u, m)$ where $u \in [0, 1]$, m is a Borel probability on $\mathrm{GL}(\mathbb{R}^d)$, and $\rho(u, m) \in \mathrm{GL}(\mathbb{R}^d)$, such that if U is a uniformly distributed random variable on $[0, 1]$ then $\rho(U, m)$ has distribution m.

Assume furthermore for any convergent sequence of probabilities $m_n \to m$ one has $\rho(U, m_n) \to \rho(U, m)$ almost surely. Such a representation ρ exists by the main result of [BD83].

In the same way fix a representation $(u, m) \mapsto \rho_{\mathrm{Flags}}(u, m)$ into $\mathrm{Flags}(\mathbb{R}^d)$, and representation $(u, m) \mapsto \rho_{\mathrm{Flags}_i}(u, m)$ into $\mathrm{Flags}_i(\mathbb{R}^d)$.

Let ν_i be the distribution of the incomplete flag F_i, and ν_{F_i} the conditional distribution of F given F_i.

Setting $(A', F_i', F') = (\rho(u_1, \mu), \rho_{\mathrm{Flags}_i}(u_2, \nu_i), \rho_{\mathrm{Flags}}(u_3, \nu_{F_i'}))$ where u_1, u_2, u_3 are i.i.d. uniform in $[0, 1]$, one has that (A', F') has distribution $\mu \times \nu$ which is the joint distribution of (A, F).

To simplify notation we assume from now on $(A, F) = (A', F')$.

3.2.2. *Perturbation.* Let $\{R_t, t \geq 0\}$ be defined so that conditioned on AF_i it is a Brownian motion starting at the identity on the group of orthogonal transformations which fix AF_i. To clarify dependence on the other random elements we assume $\{R_t, t \geq 0\}$ is $\sigma(AF_i, u_4)$-measurable where u_4 is uniform on $[0, 1]$ and independent from (u_1, u_2, u_3).

Now for each $t \geq 0$ let $A_t = R_t A$ and notice that $A_t F_i = AF_i$ almost surely and $A_t \to A$ when $t \to 0$ almost surely.

We denote by $C(\mathrm{Flags}(\mathbb{R}^d), \mathbb{R})$ the space of real valued continuous functions on $\mathrm{Flags}(\mathbb{R}^d)$ with the topology of uniform convergence, and consider for each $t \geq 0$

the operator $P_t : C(\text{Flags}(\mathbb{R}^d), \mathbb{R}) \to C(\text{Flags}(\mathbb{R}^d), \mathbb{R})$ defined by

$$(P_t f)(x) = \mathbb{E}\left(f(A_t x)\right).$$

Notice that $P_t 1 = 1$ and if $f \geq 0$ then $P_t f \geq 0$. Therefore there is an associated action of P_t on the space of probability measures on $\text{Flags}(\mathbb{R}^d)$ defined by

$$\int f(x) d(P_t)^* m(x) = \int (P_t f)(x) dm(x).$$

Lemma 3. *For each $t > 0$ there is a P_t^*-invariant probability measure ν_t on $\text{Flags}(\mathbb{R}^d)$ whose projection onto $\text{Flags}_i(\mathbb{R}^d)$ is ν_i.*

Furthermore picking for each $t > 0$ a measure ν_t as above one has $\lim_{t\to 0} \nu_t = \nu$, and letting $x_i \mapsto \nu_{t,x_i}$ be the disintegration of ν_t with respect to the projection to $\text{Flags}_i(\mathbb{R}^d)$ the following properties hold:

(1) *Almost surely ν_{t,F_i} is absolutely continuous with respect to η_{F_i}.*
(2) *There is a compact subinterval $I_t \subset (0, +\infty)$ such that $\frac{d\nu_{t,F_i}}{d\eta_{F_i}}$ takes values in I_t almost surely.*

Proof. Let $\pi : \text{Flags}(\mathbb{R}^d) \to \text{Flags}_i(\mathbb{R}^d)$ be the canonical projection.

Let $Q_t f(x) = \int f(rx) d\lambda_{t,x_i}(r)$, where $x_i = \pi(x)$, and λ_{t,x_i} is the distribution of the time of t of Brownian motion starting at the identity on the group of orthogonal transformations fixing x_i.

Notice that

$$P(Q_t f)(x) = \int (Q_t f)(ax) d\mu(a) = \int\int f(rax) d\lambda_{t,ax_i}(r) d\mu(a) = P_t f(x).$$

Since Q_t preserves the set of functions of the form $f(x) = g(\pi(x))$ one obtains that $\pi_* Q_t^* m = \pi_* m$ for all probabilities m.

In particular $P_t^* = (Q_t)^* P^*$ preserves the space of probabilities which project onto ν_i. By the Markov-Kakutani fixed point theorem, this implies that there is at least one fixed point for P_t^* in this space.

Because λ_{t,x_i} has a continuous positive density with respect to the invariant measure on group of orthogonal transformations stabilizing x_i it follows that, for any probability m on $\text{Flags}_i(\mathbb{R}^d)$ the measure $Q_t^* m$ satisfies properties 1 and 2 in the statement above.

In particular for any P_t^*-invariant probability ν_t with $\pi_* \nu_t = \nu_i$ one has $\nu_t = P_t^* \nu_t = Q_t^*(P^* \nu_t)$, and therefore ν_t satisfies properties 1 and 2.

Finally, let f be any continuous function and, supose $m = \lim_{n\to+\infty} \nu_{t_n}$ where $\lim_{n\to+\infty} t_n = 0$. Using the notation $\lambda(f)$ for the integral of f with respect to the measure λ, we have

$$\begin{aligned}
|m(f) - m(Pf)| &= \lim_{n\to+\infty} |\nu_{t_n}(f) - \nu_{t_n}(Pf)| \\
&= \lim_{n\to+\infty} |\nu_{t_n}(PQ_t f) - \nu_{t_n}(Pf)| \\
&\leq \lim_{n\to+\infty} \|P(Q_{t_n} f - f)\|_\infty \\
&\leq \lim_{n\to+\infty} \|Q_{t_n} f - f\|_\infty,
\end{aligned}$$

where we have used that $|Pf(x)| \leq \int |f(ax)| d\mu(a) \leq \|f\|_\infty$ so P decreases the L^∞ norm.

Notice that λ_{t,x_i} converges to the point mass at the identity when $t \to 0$. The convergence is uniform in the sense that given $r > 0$ and letting B_r be the ball of radius r centered at the identity in the full orthogonal group, for each $\epsilon > 0$ there exists $T > 0$ such that $\lambda_{t,x_i}(B_r) > 1 - \epsilon$ for all $t < T$ and all x_i. It follows that

$$\lim_{t\to 0} Q_t f(x) = \lim_{t\to 0} \int f(rx) d\lambda_{t,\pi(x)}(r) = f(x),$$

for all x and the convergence is uniform.

Since $\|Q_{t_n} f - f\|_\infty$ goes to zero we conclude that $m(f) = m(Pf)$. Since this holds for all f one has that $P^* m = m$. By hypothesis ν is the unique measure with this property with projection ν_i, therefore $m = \nu$.

We have shown that ν is the only limit point of ν_t when $t \to 0$. The space of probabilities on $\text{Flags}(\mathbb{R}^d)$ is compact and metrizable, and therefore this implies $\lim_{t\to 0} \nu_t = \nu$ as claimed. □

3.2.3. *Conclusion of the proof.* We will fix from now on a sequence t_n given by the following claim (c.f. [CLP19, section 6.1.6]):

Claim 1. There exists a sequence of positive numbers with $\lim_{n\to+\infty} t_n = 0$ such that, letting ν_{t_n} and ν_{t_n,x_i} be given by lemma 3 one has

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=1}^{n} \nu_{t_j,F_i} = \nu_{F_i}$$

almost surely.

Proof. To begin fix any sequence of positive numbers with $\lim_{m\to+\infty} s_m = 0$.

Let $\{f_k : k = 1, 2, \ldots\}$ be a dense sequence of continuous functions on $\text{Flags}(\mathbb{R}^d)$.

Notice that $x_i \mapsto \nu_{s_m,x_i}(f_j), m = 1, 2, \ldots$ is a bounded sequence in $L^1(\text{Flags}_i(\mathbb{R}^d), \nu_i)$.

By Komlos' theorem (see [Kom67]) there exists a subsequence $\{m_{1,j} : j = 1, 2, \ldots\}$ such that

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=1}^{n} \nu_{s_{m_{1,j}},x_i}(f_1)$$

exists for ν_i-almost every x_i, and any further subsequence has the same property.

For each $k = 1, 2, \ldots$, using Komlos' theorem as above, we may define $m_{k+1,1}, m_{k+1,2}, \ldots$ a subsequence of $m_{k,1}, m_{k,2}, \ldots$ such that

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=1}^{n} \nu_{s_{m_{k+1,j}},x_i}(f_{k+1})$$

exists for ν_i-almost every x_i, and any further subsequence has the same property.

Letting $t_n = s_{m_{n,n}}$ we have that

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{j=1}^{n} \nu_{t_j,x_i}(f_k)$$

exists for ν_i-almost every x_i and all k.

For each x_i the the restriction of $\{f_k\}$ to $\pi^{-1}(x_i)$ is dense. Since the space of probabilities on $\pi^{-1}(x_i)$ is compact, this implies that there exist probabilities m_{x_i}

such that

$$\lim_{n\to+\infty}\frac{1}{n}\sum_{j=1}^{n}\nu_{t_j,x_i}=m_{x_i}$$

for ν_i-almost every x_i.

By lemma 3 one has $\lim_{n\to+\infty}\nu_{t_n}=\nu$. Therefore, for any continuous f by dominated convergence one has

$$\int m_{x_i}(f)d\nu_i(x_i)=\int\lim_{n\to+\infty}\frac{1}{n}\sum_{j=1}^{n}\nu_{t_j,x_i}(f)d\nu_{x_i}=\lim_{n\to+\infty}\frac{1}{n}\sum_{j=1}^{n}\nu_{t_j}(f)=\nu(f),$$

from where $\int m_{x_i}d\nu_i(x_i)=\nu$ and $m_{x_i}=\nu_{x_i}$ for ν_i-almost every x_i. □

For each n let T_n be uniform $\{t_1,\ldots,t_n\}$ and independent from (u_1,u_2,u_3,u_4), and let $X_n=\rho_{\mathrm{Flags}}(u_3,\nu_{T_n,F_i})$ and $A_n=R_{T_n}A$.

Claim 2. One has that $X_n\overset{(d)}{=}A_nX_n$.

Proof. For any continuous function $f:\mathrm{Flags}(\mathbb{R}^d)\to\mathbb{R}$ one has

$$\begin{aligned}\mathbb{E}\left(f(X_n)\right)&=\mathbb{E}\left(\mathbb{E}\left(f(X_n)|F_i\right)\right)\\&=\mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{n}\nu_{t_k,F_i}(f)\right)\\&=\frac{1}{n}\sum_{k=1}^{n}\int\int f(x)d\nu_{t_k,x_i}(x)d\nu_i(x_i)\\&=\frac{1}{n}\sum_{k=1}^{n}\int\int P_{t_k}f(x)d\nu_{t_k,x_i}(x)d\nu_i(x_i)\\&=\frac{1}{n}\sum_{k=1}^{n}\int\int\int\int f(rax)d\lambda_{t_k,ax_i}(r)d\nu_{t_k,x_i}(x)d\nu_i(x_i)d\mu(a)\\&=\mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{n}\int\int f(rAx)d\lambda_{t_k,AF_i}(r)d\nu_{t_k,F_i}(x)\right)\\&=\mathbb{E}\left(\frac{1}{n}\sum_{k=1}^{n}\int f(R_{t_k}Ax)d\nu_{t_k,F_i}(x)\right)\\&=\mathbb{E}\left(f(A_nX_n)|A,F_i\right)\\&=\mathbb{E}\left(f(A_nX_n)\right).\end{aligned}$$

□

Since T_n converges in distribution to 0 there exists a subsequence such that $\lim_{k\to+\infty}T_{n_k}=0$ almost surely. We fix such a subsequence from now on.

Claim 3. The conditional distribution of $(A,A_{n_k}X_{n_k})$ given AF_i converges almost surely to the conditional distribution of (A,F) given AF_i.

Proof. It suffices to show that for all bounded and uniformly continuous f one has

$$\lim_{k\to+\infty}\mathbb{E}\left(f(A,A_{n_k}X_{n_k})|AF_i\right)=\mathbb{E}\left(f(A,AF)|AF_i\right),$$

almost surely.

The distance between $A_{n_k}X_{n_k} = R_{T_{n_k}}AX_{n_k}$ and AX_{n_k} goes to 0 almost surely. Therefore, since f is uniformly continuous, one has

$$\lim_{k\to+\infty} |f(A, A_{n_k}X_{n_k}) - f(A, AX_{n_k})| = 0,$$

almost surely.

Because f is bounded, by dominated convergence, the limit above also holds in the L^1 sense, and therefore

$$\lim_{k\to+\infty} \mathbb{E}\left(f(A, A_{n_k}X_{n_k}) - f(A, AX_{n_k})|AF_i\right) = 0,$$

almost surely.

Noticing that $\sigma(AF_i) \subset \sigma(A, F_i) = \sigma(A, AF_i)$, we now calculate

$$\begin{aligned}
\lim_{k\to+\infty} \mathbb{E}\left(f(A, A_{n_k}X_{n_k})|AF_i\right) &= \lim_{k\to+\infty} \mathbb{E}\left(f(A, AX_{n_k})|AF_i\right) \\
&= \lim_{k\to+\infty} \mathbb{E}\left(\mathbb{E}\left(f(A, AX_{n_k})|A, F_i\right)|AF_i\right) \\
&= \lim_{k\to+\infty} \mathbb{E}\left(\frac{1}{n_k}\sum_{j=1}^{n_k}\int f(A, Ax)d\nu_{t_j,F_i}(x)|AF_i\right) \\
&= \mathbb{E}\left(\int f(A, Ax)d\nu_{F_i}(x)|AF_i\right) \\
&= \mathbb{E}\left(\mathbb{E}\left(f(A, AF)|A, F_i\right)|AF_i\right) \\
&= \mathbb{E}\left(f(A, AF)|AF_i\right)
\end{aligned}$$

where we have used the almost sure convergence of $\frac{1}{n_k}\sum_{j=1}^{n_k}\nu_{t_j,F_i}$ to ν_{F_i} and boundedness of f to move the limit inside the expected value in the third to last step. $\square$

In view of the above claim by, lemma 1, proposition 1, and Fatou's lemma we have

$$\begin{aligned}
\kappa_i &= \mathbb{E}\left(I(A, AF|AF_i)\right) \\
&\le \mathbb{E}\left(\liminf_{k\to+\infty} I(A, A_{n_k}X_{n_k}|AF_i)\right) \\
&\le \liminf_{k\to+\infty} \mathbb{E}\left(I(A, A_{n_k}X_{n_k}|AF_i)\right).
\end{aligned}$$

By monotonocity and the chain rule (Proposition 4) we continue

$$\begin{aligned}
&\liminf_{k\to+\infty} \mathbb{E}\left(I(A, A_{n_k}X_{n_k}|AF_i)\right) \\
&\quad\le \liminf_{k\to+\infty} \mathbb{E}\left(I(A, (T_{n_k}, A_{n_k}X_{n_k})|AF_i)\right) \\
&\quad= \liminf_{k\to+\infty} \mathbb{E}\left(I(A, T_{n_k}|AF_i) + I(A, A_{n_k}X_{n_k}|T_{n_k}, AF_i)\right),
\end{aligned}$$

Because T_{n_k} is independent from A and AF_i we have

$$\begin{aligned}
&\liminf_{k\to+\infty} \mathbb{E}\left(I(A, T_{n_k}|AF_i) + I(A, A_{n_k}X_{n_k}|T_{n_k}, AF_i)\right) \\
&\quad= \liminf_{k\to+\infty} \mathbb{E}\left(I(A, A_{n_k}X_{n_k}|T_{n_k}, AF_i)\right).
\end{aligned}$$

And, finally, by monotonicity of mutual information

$$\liminf_{k\to+\infty} \mathbb{E}\left(I(A, A_{n_k}X_{n_k}|T_{n_k}, AF_i)\right) \le \liminf_{k\to+\infty} \mathbb{E}\left(I((R_{n_k}, A), A_{n_k}X_{n_k}|T_{n_k}, AF_i)\right).$$

We conclude the proof by establishing the following:

Claim 4. In the above context one has:

$$\liminf_{k\to+\infty} \mathbb{E}\left(I((R_{n_k}, A), A_{n_k}X_{n_k}|T_{n_k}, AF_i)\right) = \chi_i - \chi_{i+1}.$$

Proof. We first claim that the the conditional distribution of $A_{n_k}X_{n_k}$ given $\mathcal{F}_1 = \sigma(T_{n_k}, A_{n_k}F_i) = \sigma(T_{n_k}, AF_i)$ is $\nu_{T_{n_k},AF_i}$.

Since AF_i has distribution ν_i and is independent from T_{n_k} it suffices to prove that the conditional distribution of $(A_{n_k}X_{n_k}, AF_i)$ given T_{n_k} coincides with that of (X_{n_k}, F_i) given T_{n_k}.

Since $F_i = \pi(X_{n_k})$ and $AF_i = \pi(A_{n_k}X_{n_k})$, we only need to verify that the conditional distribution of X_{n_k} given T_{n_k} (which is $\nu_{T_{n_k}}$ coincides with that of $A_{n_k}X_{n_k}$ given T_{n_k}. This follows immediately since $\nu_{T_{n_k}}$ is $P^*_{T_{n_k}}$-invariant.

Now notice that the conditional distribution of X_{n_k} given $\mathcal{F}_2 = \sigma((R_{T_{n_k}}, A), \mathcal{F}_1)$ $= \sigma(R_{T_{n_k}}, A, T_{n_k}, AF_i) = \sigma(R_{T_{n_k}}, A, T_{n_k}, F_i)$ is also $\nu_{T_{n_k},F_i}$. This implies that the conditional distribution of $A_{n_k}X_{n_k}$ given $\mathcal{F}_2$ is $R_{n_k}A\nu_{T_{n_k},F_i}$.

Let m denote the conditional distribution of (R_{n_k}, A) given $\mathcal{F}_1$.

We have shown that the joint distribution of $(R_{n_k}, A), A_{n_k}X_{n_k}$ given $\mathcal{F}_1$ has projections m and $\nu_{T_{n_k},AF_i}$, while its disintegration onto the factor m has conditional measures $R_{n_k}A\nu_{T_{n_k},F_i}$.

Applying the Gelfand-Yaglom-Perez theorem this yields

$$\mathbb{E}\left(I((R_{n_k}, A), A_{n_k}X_{n_k}|T_{n_k}, AF_i)\right) = \mathbb{E}\left(\frac{dR_{n_k}A\nu_{T_{n_k},AF_i}}{d\nu_{T_{n_k},AF_i}}(R_{n_k}AX_{n_k})\right).$$

Let $S_{n,i}$ be the i-dimensional subspace of X_n and $\varphi_{t_n,F_i} = \frac{d\nu_{t_n,F_i}}{d\eta_{F_i}}$. Using lemma 2 we obtain

$$\begin{aligned}\mathbb{E}\left(\frac{dR_{n_k}A\nu_{T_{n_k},AF_i}}{d\nu_{T_{n_k},AF_i}}(R_{n_k}AX_{n_k})\right) &= \mathbb{E}\left(\log\left(\frac{|\det_{S_{n_k,i}}(R_{n_k}A)|^2}{|\det_{S_{i-1}}(R_{n_k}A)||\det_{S_{i+1}}(R_{n_k}A)|}\right)\right)\\ &\quad + \mathbb{E}\left(\log\left(\frac{\varphi_{T_{n_k},F_i}(X_{n_k})}{\varphi_{T_{n_k},AF_i}(A_{n_k}X_{n_k})}\right)\right)\\ &= \mathbb{E}\left(\log\left(\frac{|\det_{S_{n_k,i}}(R_{n_k}A)|^2}{|\det_{S_{i-1}}(R_{n_k}A)||\det_{S_{i+1}}(R_{n_k}A)|}\right)\right)\end{aligned}$$

where for the last equality we have used that $(T_{n_k}, X_{n_k}, \pi(X_{n_k})) = (T_{n_k}, X_{n_k}, F_i)$ has the same distribution as $(T_{n_k}, R_{n_k}AX_{n_k}, \pi(R_{n_k}A)) = (T_{n_k}, R_{n_k}AX_{n_k}, AF_i)$.

Since R_{n_k} is an orthogonal transformation the determinants of $R_{n_k}A$ and A coincide on all subspaces. Therefore the right hand side above is equal to

$$\mathbb{E}\left(\log\left(\frac{|\det_{S_{n_k,i}}(A)|^2}{|\det_{S_{i-1}}(A)||\det_{S_{i+1}}(A)|}\right)\right).$$

Since X_{n_k} converges in distribution to F we have that $S_{n_k,i}$ converges in distribution to S_i the i-dimensional subspace of F. Because the logarithm of the determinant of A on any subspace is bounded between constant multiples of $\log(\sigma_1(A))$ and $\log(\sigma_d(A))$ both of which are integrable, we can pass to the limit (e.g. using dominated convergence after replacing $S_{n_k,i}$ by a sequence with the same individual distributions but which converges almost surely, see [Bil99, Theorem 6.7]) obtaining

$$\lim_{k\to+\infty}\mathbb{E}\left(\log\left(\frac{|\det_{S_{n_k,i}}(A)|^2}{|\det_{S_{i-1}}(A)||\det_{S_{i+1}}(A)|}\right)\right)=\mathbb{E}\left(\log\left(\frac{|\det_{S_i}(A)|^2}{|\det_{S_{i-1}}(A)||\det_{S_{i+1}}(A)|}\right)\right).$$

Finally since $\mathbb{E}\left(|\det_{S_j}(A)|\right)=\chi_1+\cdots+\chi_j$ for $j=1,\ldots,d$, one obtains

$$\mathbb{E}\left(\log\left(\frac{|\det_{S_i}(A)|^2}{|\det_{S_{i-1}}(A)||\det_{S_{i+1}}(A)|}\right)\right)=\chi_i-\chi_{i+1},$$

which concludes the proof. □

Part 2. Exact dimensionality and dimension of conditional probabilities

In this part of the article we will prove Theorem 2. We now specify notation and context that will be used throughout.

Recall that μ is a probability on $\mathrm{GL}(\mathbb{R}^d)$ with respect to which the logarithm of all singular values are integrable and ν is a μ-stationary probability on $\mathrm{Flags}(\mathbb{R}^d)$.

A dimension $i\in\{1,\ldots,d-1\}$ is fixed throughout, ν_i is the projection of ν on the space $\mathrm{Flags}_i(\mathbb{R}^d)$ of incomplete flags missing their i-dimensional subspace. It is assumed that ν is the unique stationary probability with projection ν_i.

A disintegration $F_i\mapsto\nu_{F_i}$ of ν with respect to ν_i is fixed (so $\nu=\int\nu_{F_i}d\nu_i(F_i)$).

We consider an i.i.d. sequence $(A(n))_{n\in\mathbb{Z}}$ with common distribution μ and a stationary sequence of random random flags $(F(n))_{n\in\mathbb{Z}}$ with common distribution ν such that

$$A(n+k)\cdots A(n)F(n)=F(n+k)$$

for all $n\in\mathbb{Z}$ and $k\geq 0$. We will use $S_j(n)$ for the j-dimensional subspace of the flag $F(n)$ and $F_i(n)$ as before for the incomplete flag obtained by removing the subspace $S_i(n)$.

By hypothesis ν is ergodic (i.e. extremal among stationary probabilities) this implies that the stationary sequence $((F(n),A(n)))_{n\in\mathbb{Z}}$ is ergodic.

As before, Lyapunov exponents $\chi_1,\ldots,\chi_d$ are defined by the equations

$$\chi_1+\cdots+\chi_j=\mathbb{E}\left(\log\left(\left|\det_{S_i(n)}(A(n))\right|\right)\right).$$

By Theorem 1 one has $A\nu_{F_i(n)}\ll\nu_{F_i(n+1)}$ almost surely and

$$0\leq\kappa_i=\mathbb{E}\left(\log\left(\frac{dA\nu_{F_i(n)}}{d\nu_{F_i(n+1)}}(F(n+1))\right)\right)\leq\chi_i-\chi_{i+1}.$$

We assume from now on that $\kappa_i>0$.

4. Non-atomicity of conditional measures

Our first step in the proof of Theorem 2 is that $\nu_{F_i(n)}$ is almost surely non-atomic (i.e. all points have measure zero).

Lemma 4. *Almost surely $\nu_{F_i(n)}$ is non-atomic for all n.*

Proof. By ergodicity and one has

$$\kappa_i=\lim_{n\to+\infty}\frac{1}{n}\log\left(\frac{dA(n-1)\cdots A(0)\nu_{F_i(0)}}{d\nu_{F_i(n)}}(F(n))\right),$$

almost surely.

Suppose for the sake of contradiction that $\mathbb{P}\left(\nu_{F_i(0)}(F(0)) > 0\right) > 0$. Conditioning on this event the equation above becomes

$$\kappa_i = \lim_{n\to+\infty} \frac{1}{n} \log\left(\frac{\nu_{F_i(0)}(F(0))}{\nu_{F_i(n)}(F(n))}\right).$$

However, by Poincaré recurrence $\nu_{F_i(n)}(F(n))$ is recurrent almost surely (i.e. almost surely there exists a subsequence such that $\lim_k \nu_{F_i(n_k)}(F(n_k)) = \nu_{F_i(0)}(F(0))$). This implies that $\kappa_i = 0$ which contradicts the hypothesis that $\kappa_i > 0$. Hence, $\nu_{F_i(0)}(F(0)) = 0$ almost surely, as claimed. □

5. The multiplicative ergodic theorem

From Theorem 1 and the hypothesis that $\kappa_i > 0$ one obtains that $\chi_i > \chi_{i+1}$. We will now apply the multiplicative ergodic theorem of [Ose68] to the mappings induced by the sequence $A(n)$ between the quotient spaces $S_{i+1}(n)/S_{i-1}(n)$ to obtain the following result:

Lemma 5. *Almost surely for each n one has*

$$\lim_{k\to+\infty} \frac{1}{k} \log\left(|\det_{S_i(n)}(A(n+k-1)\cdots A(n))|\right) = \chi_1 + \cdots + \chi_{i-1} + \chi_i,$$

and there exists a unique i-dimensional subspace $S_i'(n)$ containing $S_{i-1}(n)$ and contained in $S_{i+1}(n)$ such that

$$\lim_{k\to+\infty} \frac{1}{k} \log\left(|\det_{S_i'(n)}(A(n+k-1)\cdots A(n))|\right) = \chi_1 + \cdots + \chi_{i-1} + \chi_{i+1}.$$

Furthermore, $S_i(n)$ and $S_i'(n)$ are conditionally independent given $F_i(n)$, and $S_i(n) \neq S_i'(n)$ almost surely.

Finally, the logarithm of the angle between the projections of $S_i(n)$ and $S_i'(n)$ to $S_{i+1}(n)/S_{i-1}(n)$ is $o(|n|)$ when $n \to \pm\infty$.

Proof. For each n consider the quotient space $V(n) = S_{i+1}(n)/S_{i-1}(n)$ with the induced inner product coming from $\mathbb{R}^d$, let $E^u(n)$ be the one-dimensional subspace in $V(n)$ which is the projection of $S_i(n)$, and let $T(n) : V(n) \to V(n+1)$ be mapping induced by $A(n)$.

Notice that almost surely each $V(n)$ is isometric to $\mathbb{R}^2$ with the usual inner product. Furthermore the random sequence

$$\cdots \overset{T(n-1)}{\mapsto} (V(n), E^u(n)) \overset{T(n)}{\mapsto} (V(n+1), E^u(n+1)) \overset{T(n+1)}{\mapsto} \cdots$$

is stationary and ergodic.

One has

$$\mathbb{E}\left(\log\left(\left|\det_{E^u(n)}(T_n)\right|\right)\right) = \chi_i$$

which implies by Birkhoff's theorem that almost surely

$$\lim_{k\to+\infty} \frac{1}{k} \log\left(\|T(n-k)^{-1}\cdots T(n-1)^{-1}v\|\right) = -\chi_i$$

and

$$\lim_{k\to+\infty} \frac{1}{k} \log\left(\|T(n+k-1)\cdots T(n)v\|\right) = \chi_i$$

for all $v \in E^u(n) \setminus \{0\}$.

On the other hand

$$\mathbb{E}\left(\log\left(|\det(T_n)|\right)\right) = \chi_i + \chi_{i+1}.$$

which implies that almost surely

$$\lim_{k\to+\infty} \frac{1}{k}\log\left(|\det(T(n+k-1)\cdots T(n))|\right) = \chi_i + \chi_{i+1}.$$

By hypothesis $\kappa_i > 0$ which implies by Theorem 1 that $\chi_i > \chi_{i+1}$. Hence, one obtains from the multiplicative ergodic theorem of [Ose68] that almost surely

$$E^u(n) = \{0\} \cup \{v \in V(n) : \lim_{k\to+\infty}\frac{1}{k}\log\left(\|T(n-k)^{-1}\cdots T(n-1)^{-1}v\|\right) = -\chi_i\}$$

and

$$E^s(n) = \{0\} \cup \{v \in V(n) : \lim_{k\to+\infty}\frac{1}{k}\log\left(\|T(n+k-1)\cdots T(n)v\|\right) = \chi_{i+1}\},$$

are complementary one-dimensional subspaces, and the angle between them is $e^{o(n)}$.

From the equations above it follows that $E^u(n)$ is $\sigma(F_i(n), A(n-1), A(n-2),\ldots)$-measurable, while $E^s(n)$ is $\sigma(F_i(n), A(n), A(n+1),\ldots)$-measurable. Since $F_i(n)$ is $\sigma(A(n-1), A(n-2),\ldots)$-measurable one has that $(A(n-1), A(n-2),\ldots)$ and $(A(n), A(n+1),\ldots)$ are conditionally independent given $F_i(n)$. In particular, conditioned on $F_i(n)$ one has that $E^u(n)$ and $E^s(n)$ are independent.

Setting $S_i'(n)$ to be the subspace in $S_{i+1}(n)$ which projects to $E^s(n)$ in $S_{i+1}(n)/S_{i-1}(n)$ one obtains the desired result. □

6. Proof of Theorem 2

6.1. **Random circle diffeomorphisms.** We fix from now on a Borel measurable projection from $\text{Flags}_i(\mathbb{R}^d)$ to $\mathbb{R}^2$ which consists of mapping S_{i+1}/S_{i-1} to $\mathbb{R}^2$ isometrically (where S_j denotes the j-dimensional subspace of the flag). Furthermore we fix an isometry between the unit circle S^1 with the usual arc-length distance scaled by one half dist, and the space of one-dimensional subspaces of $\mathbb{R}^2$ with the distance given by the angle. The composition of these mappings will be used to identify each fiber of the projection from $\text{Flags}(\mathbb{R}^d)$ to $\text{Flags}_i(\mathbb{R}^d)$ with the unit circle. Equivalently, given an incomplete flag $F_i = (S_0,\ldots,S_d)$ we have chosen an isometry from the projective space of S_{i+1}/S_{i-1} to the unit circle, and therefore each i-dimensional subspace between S_{i-1} and S_{i+1} corresponds to a point on the unit circle.

With these identifications let $\mathcal{F}_n = \sigma(F_i(n))$, ν_n be be the projection of $\nu_{F_i(n)}$ to S^1, x_n be the projection of $S_i(n)$ to S^1, y_n be the projection of $S_i'(n)$ (given by Lemma 5) to S^1, T_n the diffeomorphism of S^1 obtained by projecting the action of $A(n)$ between $S_{i+1}(n)/S_{i-1}(n)$ and $S_{i+1}(n+1)/S_{i-1}(n+1)$, and for convenience let $\kappa = \kappa_i$ and $\chi = \chi_i - \chi_{i+1}$. Finally, we let η be the rotationally invariant probability on the unit circle.

The proof of Theorem 2 will proceed as follows (see Figure 1): We will construct a sequence of random intervals I_n containing x_n and such that $T_{-1}\circ\cdots\circ T_{-n}(I_{-n})$ is roughly of size $e^{-\chi n}$. We will then show that $\nu_0(T_{-1}\circ\cdots\circ T_{-n}(I_{-n}))$ is roughly $e^{-\kappa n}$. These two facts will yield that the local dimension of ν_0 at x_0 is almost surely κ/χ so that in particular that ν_0 is exact dimensional.

A few technical issues arise which we have concealed with the word ‘roughly’ in the previous paragraph. For example, the estimates for the measure of the intervals will hold only for some values of n, but these values are sufficiently dense to imply the needed dimension estimates.

We begin with a simple consequence of lemma 5.

Proposition 5. *Let $k \in \mathbb{Z}$ and $\epsilon \in (0,1)$ be fixed and let $I = S^1 \setminus B_{\epsilon\,\mathrm{dist}(x_k,y_k)}(x_k)$.*

Then the length of $T_{k-n}^{-1} \circ \cdots \circ T_{k-1}^{-1}(I)$ converges to 0 exponentially quickly when $n \to +\infty$.

Proof. The interval I corresponds to a cone C of one dimensional subspaces in $S_{i+1}(k)/S_{i-1}(k)$ whose angle (with respect to the standard inner product inherited from $\mathbb{R}^d$) with the projection E^u of $S_i(k)$ is larger than ϵ times the angle between E^u and the projection E^s of $S_i'(k)$.

By lemma 5, under the action the linear mapping L_n corresponding to $T_{k-n}^{-1} \circ \cdots \circ T_{k-1}^{-1}$ the norm of vectors in E^u are multiplied by a factor of $e^{-\chi_i n + o(n)}$ while those in E^s are multiplied by a factor of $e^{-\chi_{i+1} n + o(n)}$.

We fix on the domain of L_n the inner product for which the norm on E^u, E^s coincides with the standard one, but for which these subspaces are orthogonal.

Similarly on the range of L_n we pick the inner product where L_nE^u, L_nE^s are orthogonal and the restriction of the norm on both subspaces coincides with the usual one.

With respect to these inner products the angle between any two subspaces in C decreases by a factor of $e^{-(\chi_i - \chi_{i+1})n + o(n)}$ under L_n.

However, once again by lemma 5, the angle between L_nE^u, L_nE^s is $e^{o(n)}$ for the standard inner product. This implies that, measured with the standard inner product the angle between any two subspaces of C decreases by the same factor up to a multiplicative $e^{o(n)}$. □

6.2. **Stationary intervals.** We now construct the sequence of intervals that will be used in our argument. The key points for what follows are that: the construction is stationary, the intervals contain x_n but not y_n, their size is controlled by $\mathrm{dist}(x_n, y_n)$, and frequently $\nu_n(I_n)$ is not close to zero.

Lemma 6 (Stationary intervals)**.** *Setting*

$$I_n = S^1 \setminus B_{\frac{1}{2}\mathrm{dist}(x_n,y_n)}(y_n),$$

one has $\mathbb{P}\left(\nu_n(I_n) \geq 1/2\right) \geq 1/2$ for all n.

Proof. Since almost surely ν_n is non-atomic there is a smallest positive radius r_n such that $\nu_n(B_{r_n}(y_n)) = \nu_n(S^1 \setminus B_{r_n}(y_n)) = 1/2$.

By lemma 5, conditioned on $\mathcal{F}_n$ one has that x_n has distribution ν_n and is independent from r_n and y_n. Therefore $\mathbb{P}(x_n \in B_{r_n}(y_n)|\mathcal{F}_n) = \nu_n(B_{r_n}(y_n)) = 1/2$ and taking expected value $\mathbb{P}(x_n \in B_{r_n}(y_n)) = 1/2$.

In the event that $x_n \in B_{r_n}(y_n)$ one has that $S^1 \setminus B_{r_n}(y_n) \subset I_n$ and therefore that $\nu_n(I_n) \geq 1/2$. This proves the claim. □

What remains is to estimate the size and ν_0 probability of the sequence $T_{-1} \circ \cdots \circ T_{-n}(I_{-n})$.

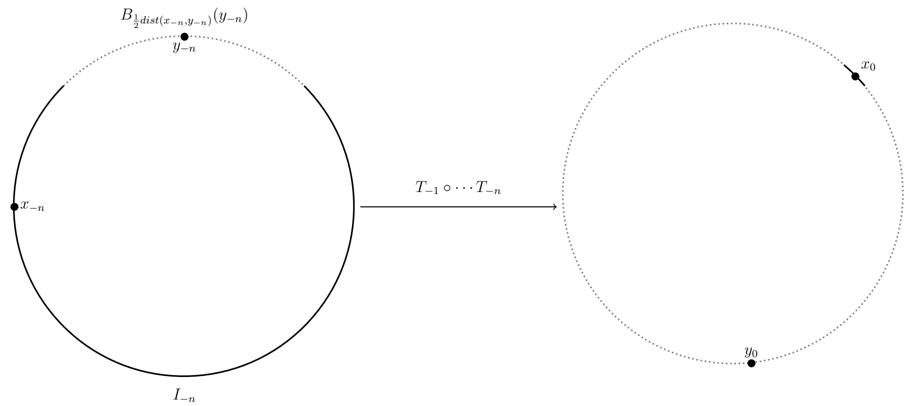

FIGURE 1. For large n the transformation $T_{-1} \circ \cdots \circ T_{-n}$ contracts the large interval I_{-n} to an interval of size roughly $e^{-\chi n}$ (see Lemma 7). With frequency at least 1/2 the ν_0-measure of the image interval is roughly $e^{-\kappa n}$ (see lemmas 6 and 11).

6.3. **Length of distinguished intervals.** The point of what follows is that the intervals $T_{-1} \circ \cdots \circ T_{-n}(I_{-n})$ contain x_0 and are roughly of size $e^{-\chi n}$.

We will use the following result which is essentially Maker's theorem [Mak40, Theorem 1] or [Bre57, Theorem 1].

Theorem (Maker's theorem)**.** *Let $(X_{n,k})_{k,n\in\mathbb{Z}}$ be a family of random variables which is stationary in the sense that its distribution equals that of $(Y_{k,n})_{k,n\in\mathbb{Z}}$ where $Y_{k,n} = X_{k+1,n}$.*

Suppose that the limit $X_k = \lim_{n\to+\infty} X_{k,n}$ exists almost surely and that $\mathbb{E}(\sup_n |X_{k,n}|) < +\infty$ for all (or equivalently due to stationary, for some) k.

Then $\lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} X_{-k,n-k} = \lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} X_{-k}$ almost surely.

Proof. By Birkhoff's ergodic theorem

$$X = \lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} X_{-k},$$

exists almost surely and $\mathbb{E}(X) = \mathbb{E}(X_0)$ is finite.

Following [Bre57, Theorem 1] we write

$$\frac{1}{n}\sum_{k=0}^{n-1} X_{-k,n-k} = \frac{1}{n}\sum_{k=0}^{n-1} X_{-k} + \frac{1}{n}\sum_{k=0}^{n-1}(X_{-k,n-k} - X_{-k}).$$

The first term converges to X almost surely. Letting Y_n be the second term notice that for any fixed N we have

$$\begin{aligned}\limsup_{n\to+\infty} |Y_n| &= \limsup_{n\to+\infty} \left|\frac{1}{n}\sum_{k=0}^{n-1}(X_{-k,n-k} - X_{-k})\right| \\ &\le \lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} \sup_{n\ge N} |X_{-k,n} - X_{-k}| = Z_N,\end{aligned}$$

where the limit defining Z_N exists almost surely and satisfies

$$\mathbb{E}(Z_N) = \mathbb{E}\left(\sup_{n\geq N} |X_{0,n} - X_0|\right) < +\infty$$

by Birkhoff's ergodic theorem.

Since $\sup_{n\geq N} |X_{0,n} - X_0|$ decreases monotonely to 0 we obtain

$$\mathbb{E}\left(\limsup_{n\to+\infty} |Y_n|\right) \leq \lim_{N\to+\infty} \mathbb{E}(Z_N) = 0,$$

so that $\limsup_{n\to+\infty} |Y_n| = 0$ almost surely. □

Lemma 7 (Length of distinguished intervals)**.** *For all $\epsilon > 0$ almost surely one has*

$$B_{r_n}(x_0) \subset T_{-1}\cdots T_{-n} I_{-n} \subset B_{R_n}(x_0)$$

for all n large enough, where $r_n = \exp(-(\chi+\epsilon)n)$ and $R_n = \exp(-(\chi-\epsilon)n)$.

Proof. Recall that η denotes the rotationally invariant probability on the unit circle S^1.

By Lemma 2 one has

$$\chi = \mathbb{E}\left(\log\left(\frac{dT_{k-1}\eta}{d\eta}(x_k)\right)\right),$$

for all k.

For each n let J_n be the connected component of $I_n \setminus \{x_n\}$ which is counter-clockwise from x_n and define

$$X_{k,n} = \log\left(\frac{\eta(T_{k-2}\circ\cdots\circ T_{k-n}(J_{k-n}))}{\eta(T_{k-1}\circ\cdots\circ T_{k-n}(J_{k-n}))}\right),$$

and

$$X_k = \log\left(\frac{dT_{k-1}\eta}{d\eta}(x_k)\right).$$

By lemma 5 one has that $S^1 \setminus J_{n-k}$ contains a ball of radius $e^{o(n)}$ centered at y_{n-k}. In view of proposition 5 this implies that for all $\epsilon \in (0,1)$ almost surely eventually $T_{k-1}\circ\cdots\circ T_{k-n}(J_{k-n}) \subset B_{\epsilon\,\mathrm{dist}(x_k,y_k)}(x_k)$. This shows that almost surely the length of $T_{k-1}\circ\cdots\circ T_{k-n}(J_{k-n})$ goes to zero when $n\to+\infty$ and therefore one one has $\lim_{n\to+\infty} X_{k,n} = X_k$ almost surely for all k.

Notice that for each $x \in S^1$ one has, again by lemma 2, that

$$\frac{dT_{k-1}\eta}{d\eta}(x) = \frac{|\det_S(A(k-1))|^2}{|\det_{S_{i-1}(k-1)}(A(k-1))||\det_{S_{i+1}(k-1)}(A(k-1))|}$$

for some i-dimensional subspace S between $S_{i-1}(k-1)$ and $S_{i+1}(k-1)$.

In particular this implies that $\min_x \log\left(\frac{dT_{k-1}\eta}{d\eta}(x)\right)$ and $\max_x \log\left(\frac{dT_{k-1}\eta}{d\eta}(x)\right)$ have finite expectation since they are controlled by the logarithms of singular values of $A(k-1)$.

This yields that $\sup_{n\to+\infty} |X_{k,n}|$ has finite expectation for all k.

Applying Maker's theorem we obtain

$$\begin{aligned}\chi &= \lim_{n\to+\infty}\frac{1}{n}\sum_{k=0}^{n-1}\log\left(\frac{dT_{-k-1}\eta}{d\eta}(x_{-k})\right)\\ &= \lim_{n\to+\infty}\frac{1}{n}\sum_{k=0}^{n-1}\log\left(\frac{\eta(T_{-k-2}\circ\cdots\circ T_{-n}(J_{-n}))}{\eta(T_{-k-1}\circ\cdots\circ T_{-n}(J_{-n}))}\right)\\ &= \lim_{n\to+\infty}\frac{1}{n}\log\left(\frac{\eta(J_{-n})}{\eta(T_{-1}\circ\cdots\circ T_{-n}(J_{-n}))}\right).\end{aligned}$$

Finally, since $\eta(J_{-n}) = e^{o(n)}$ when $n\to+\infty$ by Lemma 5 one obtains:

$$\lim_{n\to+\infty}\frac{1}{n}\log\left(\eta(T_{-1}\circ\cdots\circ T_{-n}(J_{-n})\right) = -\chi.$$

The same argument shows that $\eta(T_{-1}\circ\cdots T_{-n}(I_{-n}\setminus J_{-n})) = e^{-\chi n}$ which establishes the claims. □

6.4. **Probability of distinguished intervals.** We will now essentially repeat the argument of the previous subsection replacing the rotationally invariant probability measure (which is equivalent to length up to a factor) with the random probabilities ν_n.

In this case one wishes to replace (in the ergodic averages) the terms of the form $\frac{dT_{k-1}\nu_{k-1}}{d\nu_k}(x_k)$ with approximating terms calculated using the intervales I_n. Almost sure convergence of the approximating terms boils down to the theorem on differentiation of measures. However, the integrability of the supremum of the approximating terms is more subtle.

The issue is that the singular values of $A(k-1)$ do not directly control the maximum and minimum of $\frac{dT_{k-1}\nu_{k-1}}{d\nu_k}(x)$ on the circle. In fact, this density may be unbounded with positive probability. Instead, control of the approximation comes from the $x\log(x)$-integrability of the density with respect to ν_k which follows from the fact that $\kappa < +\infty$ (that is Theorem 1).

6.4.1. *Orlicz regularity and a maximal inequality.* For each k let $f_k(x) = \frac{dT_{k-1}\nu_{k-1}}{d\nu_k}(x)$ and notice that it is $\sigma(\mathcal{F}_{k-1},\mathcal{F}_k,T_{k-1})$-measurable.

Notice that x_{k-1} and $A(k-1)$ are independent conditioned on $\mathcal{F}_{k-1}$. Since the conditional distribution of x_{k-1} given $\mathcal{F}_{k-1}$ is ν_{k-1} one obtains that the distribution of $x_k = T_{k-1}(x_{k-1})$ conditioned on $\sigma(\mathcal{F}_{k-1},A(k-1))$ has density f_k with respect to ν_k. Since this conditional distribution is $\sigma(\mathcal{F}_{k-1},\mathcal{F}_k,T_{k-1})$-measurable and $\sigma(\mathcal{F}_{k-1},\mathcal{F}_k,T_{k-1})\subset\sigma(\mathcal{F}_{k-1},A(k-1))$ one obtains that the conditional distribution of x_k given $\sigma(\mathcal{F}_{k-1},\mathcal{F}_k,T_{k-1})$ has density f_k with respect to ν_k. Therefore,

$$\begin{aligned}\kappa &= \mathbb{E}\left(\log\left(f_k(x_k)\right)\right) = \mathbb{E}\left(\mathbb{E}\left(\log\left(f_k(x_k)\right)|\mathcal{F}_{k-1},\mathcal{F}_k,T_{k-1}\right)\right)\\ &= \mathbb{E}\left(\int f_k(x)\log(f_k(x))d\nu_k(x)\right).\end{aligned}$$

In particular $f_k\log(f_k)$ is almost surely integrable with respect to ν_k. In other words, f_k almost surely belongs to an Orlicz space which is slightly smaller than $L^1(\nu_k)$ and the expected value of the corresponding Orlicz norm is finite. This fact, which follows from the finiteness of κ given by Theorem 1, will allow us to control the maximal function of f_k.

We define the maximal function of a function $f : S^1 \to \mathbb{R}$ with respect to a probability λ as

$$M_\lambda f(x) = \sup_{x\in I} \frac{1}{\nu(I)} \int_I |f(y)| d\lambda(y)$$

where the supremum is over all intervals containing x.

We will need the following maximal inequality the proof of which is adapted from the proof of [Ste70, Theorem 1].

Lemma 8 (Maximal inequality)**.** *There exists a constant $C > 0$ such that for any probability λ on S^1 and any λ-integrable function f one has*

$$t\lambda\left(\{x : M_\lambda f(x) > t\}\right) \le C \int |f| \mathbb{1}_{\{|f|>t/2\}} d\lambda$$

for all $t > 0$.

Proof. Given λ,f, and t consider a compact set $K \subset \{M_\lambda f > t\}$ such that

$$\lambda(\{M_\lambda f > t\}) \le 2\lambda(K).$$

By definition, each point in K belongs to an interval I such that

$$t\lambda(I) < \int |f| \mathbb{1}_I d\lambda.$$

Since K is compact one may cover it with finitely many such intervals.

Applying the Besicovitch covering lemma (e.g. see [dG75, Theorem 1.1]) there exists a constant c (which does not depend on λ nor f) such that a subcover may be found so that no more than c intervals intersect simultaneously.

Summing over such a subcover one has

$$t\lambda(\{M_\lambda f > t\}) \le 2t\lambda(K) \le 2c \int |f| d\lambda.$$

This inequality has been established for all λ-integrable f and all $t > 0$. Applying it to $g = f\mathbb{1}_{\{|f|>t/2\}}$ one obtains (observing that $M_\lambda f \le t/2 + M_\lambda g$) that

$$t\lambda(\{M_\lambda f > t\}) \le t\lambda(\{M_\lambda g > t/2\}) \le 4c \int |f| \mathbb{1}_{\{|f|>t/2\}} d\lambda$$

which establishes the claim. □

We now use Lemma 8 to control the typical maximal function of f_k. The argument is adapted from [Nev75, Proposition IV-2-10], see the appendix of said work for discussion of this type of results in the context of general Orlicz spaces.

Lemma 9 (Average maximal function)**.** *In the context above one has*

$$\mathbb{E}\left(\log\left(M_{\nu_k} f_k(x_k)\right)\right) < +\infty.$$

Proof. As observed at the beginning of section 6.4.1 the conditional distribution of x_k given $\sigma(\mathcal{F}_{k-1}, T_{k-1}, \mathcal{F}_k)$ has density f_k with respect to ν_k. Therefore,

$$\begin{aligned}\mathbb{E}\left(\log\left(M_{\nu_k} f_k(x_k)\right)\right) &= \mathbb{E}\left(\mathbb{E}\left(\log\left(M_{\nu_k} f_k(x_k)\right) | \mathcal{F}_{k-1}, T_{k-1}, \mathcal{F}_k\right)\right) \\ &= \mathbb{E}\left(\int f_k(x) \log\left(M_{\nu_k} f_k(x)\right) d\nu_k(x)\right).\end{aligned}$$

The lower bound $f_k \log(f_k) \le f_k \log(M_{\nu_k} f_k)$ which holds ν_k-almost everywhere reduces the problem to showing that the expected value on the right is not $+\infty$.

Applying the inequality $a\log(b) \le a\log(a) + b/e$ (valid for $a, b \ge 0$) one obtains

$$\mathbb{E}\left(\int f_k(x)\log\left(M_{\nu_k} f_k(x)\right) d\nu_k(x)\right) \le \kappa + \frac{1}{e}\mathbb{E}\left(\int M_{\nu_k} f_k(x) d\nu_k(x)\right).$$

We now conclude by using Lemma 8 as follows

$$\begin{aligned}\mathbb{E}\left(\int M_{\nu_k} f_k(x) d\nu_k(x)\right) &\le 1 + \mathbb{E}\left(\int_1^{+\infty} \nu_k\left(\{M_{\nu_k} f_k \ge t\}\right) dt\right) \\ &\le 1 + C\mathbb{E}\left(\int_1^{+\infty}\int f_k(x)\frac{1}{t}\mathbb{1}_{\{f_k \ge t/2\}} d\nu_k(x) dt\right) \\ &\le 1 + C\mathbb{E}\left(\int f_k(x)\log(2f_k(x)) d\nu_k(x)\right) \\ &= 1 + C\log(2) + C\kappa.\end{aligned}$$

□

6.4.2. *Domination of approximating terms.* We will now establish the main estimate needed to apply Maker's theorem as in Lemma 7. For the needed upper bound Lemma 9 suffices. For the lower bound we mimic the argument of [Chu61].

Lemma 10. *For each $n = 1, 2, \ldots$ let $J_n = T_{-1} \circ \cdots \circ T_{-n}(I_{-n})$ and*

$$X_n = \log\left(\frac{T_{-1}\nu_{-1}(J_n)}{\nu_0(J_n)}\right).$$

Then $\mathbb{E}\left(\sup_n |X_n|\right) < +\infty$.

Proof. Notice first that

$$X_n = \log\left(\frac{1}{\nu_0(J_n)}\int_{J_n} f_0(x) d\nu_0(x)\right) \le \log\left(M_{\nu_0} f_0(x_0)\right).$$

In view of Lemma 9 this bounds $\sup_n X_n$ from above by an integrable random varaible.

For the lower bound consider the event that $X_n \le -t$ and notice that this implies

$$\int_{J_n} f_0(x) d\nu_0(x) \le e^{-t}\nu_0(J_n).$$

Given f_0 and ν_0 define the bad set B_t as the set of points in the circle belonging to an interval I such that

$$\int_I f_0 d\nu_0 \le e^{-t}\nu_0(I). \tag{1}$$

Following the proof of Lemma 8 consider a compact set $K \subset B_t$ with

$$\int_{B_t} f_0 d\nu_0 \le 2\int_K f_0 d\nu_0.$$

By considering a finite covering of K by intervals satisfying equation 1 and summing over a Besicovitch subcover where no more than c intervals overlap (here the constant c does not depend on f_0 nor ν_0) we obtain:

$$\int_{B_t} f_0\nu_0 \le 2\int_K f_0 d\nu_0 \le 2ce^{-t}.$$

Using that the conditional distribution of x_0 given f_0 and ν_0 is $f_0\nu_0$ we obtain

$$\mathbb{P}\left(\inf_n X_n \le -t\right) \le \mathbb{P}\left(x_0 \in B_t\right) = \mathbb{E}\left(\mathbb{P}\left(x_0 \in B_t | f_0, \nu_0\right)\right)$$
$$= \mathbb{E}\left(\int_{B_t} f_0\nu_0\right) \le 2ce^{-t}$$

which shows that $\inf_n X_n$ is integrable as claimed. □

6.4.3. *Probability estimates.* Having solved the main technical issues we now repeat the argument of Lemma 7 replacing the uniform measure η with the random measure ν_0 to obtains the desired estimate on the ν_0-measure of a sequence of intervals shrinking to x_0.

Lemma 11 (Probability of distinguished intervals)**.** *Almost surely one has*

$$\lim_{n\to+\infty} \frac{1}{n} \log\left(\frac{\nu_{-n}(I_{-n})}{\nu_0(T_{-1}\circ\cdots\circ T_{-n}(I_{-n}))}\right) = \kappa.$$

Proof. For each $k \in \mathbb{Z}$ and $n = 1, 2, \ldots$ let $J_{k,n} = T_{k-1}\circ\cdots\circ T_{k-n}(I_{k-n})$,

$$X_{k,n} = \log\left(\frac{T_{k-1}\nu_{k-1}(J_{k,n})}{\nu_k(J_{k,n})}\right),$$

and

$$X_k = \log\left(\frac{dT_{k-1}\nu_{k-1}}{d\nu_k}(x_k)\right).$$

Notice that for each n the sequence $X_{k,n}$ is stationary and almost surely $\lim_{n\to+\infty} X_{k,n} = X_k$.

Furthermore $\sup_n |X_{k,n}|$ is integrable by Lemma 10.

Applying Maker's theorem as in lemma 7, almost surely one has

$$\begin{aligned}
\kappa &= \lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} \log\left(\frac{dT_{-k-1}\nu_{-k-1}}{d\nu_{-k}}(x_{-k})\right) \\
&= \lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} X_{-k} \\
&= \lim_{n\to+\infty} \frac{1}{n}\sum_{k=0}^{n-1} X_{-k,n-k} \\
&= \lim_{n\to+\infty} \frac{1}{n} \log\left(\frac{\nu_{-n}(I_{-n})}{\nu_0(T_{-1}\circ\cdots\circ T_{-n}(I_{-n}))}\right),
\end{aligned}$$

as claimed. □

A technical issue in what follows is that the asymptotic lower bound for $\nu_0(T_{-1}\circ\cdots\circ T_{-n}(I_{-n}))$ just obtained, is bad when $\nu_{-n}(I_{-n})$ is small. However, in view of Lemma 6, $\nu_{-n}(I_{-n}) \ge 1/2$ 'half of the time', and this suffices for our needs.

6.5. **Proof of Theorem 2.** Let $n_1 < n_2 < \cdots$ be the (random) sequence of values of n for which $\nu_{-n}(I_{-n}) \ge 1/2$. By Lemma 6 this occurs with probability at least $1/2$ for each fixed n. Hence, by the ergodic theorem, taking a subsequence we may assume that $n_k = 2k + o(k)$ almost surely.

For each k let $J_k = (T_{-1}\circ\cdots\circ T_{-n_k})(I_{-n_k})$.

Fix $\epsilon > 0$ and let $r_n = \exp(-(\chi+\epsilon)n)$ and $R_n = \exp(-(\chi-\epsilon)n)$.

Choose two integer valued functions $\ell(r) \le k(r)$ such that

$$\ell(r) = \frac{-(1-\epsilon)\log(r)}{2(\chi+\epsilon)} + o(\log(r))$$

and

$$k(r) = \frac{-(1+\epsilon)\log(r)}{2(\chi-\epsilon)} + o(\log(r))$$

as $r \to 0$.

Notice that eventually one has $R_{n_{k(r)}} \le r \le r_{n_{\ell(r)}}$ and therefore by Lemma 7 almost surely

$$J_{k(r)} \subset B_{R_{n_{k(r)}}}(x_0) \subset B_r \subset B_{r_{n_{\ell(r)}}} \subset J_{\ell(r)},$$

for all r small enough.

Combining these facts one obtains the bounds

$$\frac{-\log(\nu_0(J_{\ell(r)}))}{-\log(R_{n_{k(r)}})} \le \frac{-\log(\nu_0(B_r(x_0)))}{-\log(r)} \le \frac{-\log(\nu_0(J_{k(r)}))}{-\log(r_{n_{\ell(r)}})}$$

By Lemma 11 almost surely

$$-\log(\nu_0(J_k)) = \kappa n_k + o(k),$$

when $k \to +\infty$.

This implies that almost surely

$$\frac{(1-\epsilon)\kappa}{\chi+\epsilon} \le \underline{\dim}_{x_0}(\nu) \le \overline{\dim}_{x_0}(\nu) \le \frac{(1+\epsilon)\kappa}{\chi-\epsilon}.$$

By intersecting over the corresponding full measure sets for a countable sequence $\epsilon_n \to 0$ one obtains that almost surely ν_0 is exact dimensional with dimension κ/χ as claimed.

Acknowledgments

The author is grateful to François Ledrappier for many helpful discussions. The author would also like to thank an anonymous referee for pointing out an error in a previous version of the proof of theorem 1, and for helping to improve the general quality of the article.

References

[BD83] David Blackwell and Lester E. Dubins, *An extension of Skorohod's almost sure representation theorem*, Proc. Amer. Math. Soc. **89** (1983), no. 4, 691–692, DOI 10.2307/2044607. MR718998

[Bil99] Patrick Billingsley, *Convergence of probability measures*, 2nd ed., A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York, 1999. MR1700749

[Bre57] Leo Breiman, *The individual ergodic theorem of information theory*, Ann. Math. Statist. **28** (1957), 809–811, DOI 10.1214/aoms/1177706899. MR92710

[Chu61] K. L. Chung, *A note on the ergodic theorem of information theory*, Ann. Math. Statist. **32** (1961), 612–614, DOI 10.1214/aoms/1177705069. MR131782

[CLP19] Matias Carrasco, Pablo Lessa, and Elliot Paquette, *On the speed of distance stationary sequences*, arXiv e-prints, page `arXiv:1912.12523`, December 2019.

[dG75] Miguel de Guzmán, *Differentiation of integrals in R^n*, Lecture Notes in Mathematics, Vol. 481, Springer-Verlag, Berlin-New York, 1975. With appendices by Antonio Córdoba, and Robert Fefferman, and two by Roberto Moriyón. MR0457661

[Dob59] R. L. Dobrušin, *A general formulation of the fundamental theorem of Shannon in the theory of information* (Russian), Uspehi Mat. Nauk **14** (1959), no. 6 (90), 3–104. MR0107574

[FK83] H. Furstenberg and Y. Kifer, *Random matrix products and measures on projective spaces*, Israel J. Math. **46** (1983), no. 1-2, 12–32, DOI 10.1007/BF02760620. MR727020

[Fur63] Harry Furstenberg, *Noncommuting random products*, Trans. Amer. Math. Soc. **108** (1963), 377–428, DOI 10.2307/1993589. MR163345

[GdM89] I. Ya. Gol′dsheĭd and G. A. Margulis, *Lyapunov exponents of a product of random matrices* (Russian), Uspekhi Mat. Nauk **44** (1989), no. 5(269), 13–60, DOI 10.1070/RM1989v044n05ABEH002214; English transl., Russian Math. Surveys **44** (1989), no. 5, 11–71. MR1040268

[GfY59] I. M. Gel′fand and A. M. Yaglom, *Calculation of the amount of information about a random function contained in another such function*, Amer. Math. Soc. Transl. (2) **12** (1959), 199–246. MR0113741

[GR89] Yves Guivarc'h and Albert Raugi, *Propriétés de contraction d'un semi-groupe de matrices inversibles. Coefficients de Liapunoff d'un produit de matrices aléatoires indépendantes* (French, with English summary), Israel J. Math. **65** (1989), no. 2, 165–196, DOI 10.1007/BF02764859. MR998669

[Gra11] Robert M. Gray, *Entropy and information theory*, 2nd ed., Springer, New York, 2011. MR3134681

[HS17] Michael Hochman and Boris Solomyak, *On the dimension of Furstenberg measure for $SL_2(\mathbb{R})$ random matrix products*, Invent. Math. **210** (2017), no. 3, 815–875, DOI 10.1007/s00222-017-0740-6. MR3735630

[Kom67] J. Komlós, *A generalization of a problem of Steinhaus*, Acta Math. Acad. Sci. Hungar. **18** (1967), 217–229, DOI 10.1007/BF02020976. MR210177

[Led84] F. Ledrappier, *Quelques propriétés des exposants caractéristiques* (French), École d'été de probabilités de Saint-Flour, XII—1982, Lecture Notes in Math., vol. 1097, Springer, Berlin, 1984, pp. 305–396, DOI 10.1007/BFb0099434. MR876081

[Mak40] Philip T. Maker, *The ergodic theorem for a sequence of functions*, Duke Math. J. **6** (1940), 27–30. MR2028

[Nev75] J. Neveu, *Discrete-parameter martingales*, Revised edition, North-Holland Publishing Co., Amsterdam-Oxford; American Elsevier Publishing Co., Inc., New York, 1975. Translated from the French by T. P. Speed; North-Holland Mathematical Library, Vol. 10. MR0402915

[Ose68] V. I. Oseledec, *A multiplicative ergodic theorem. Characteristic Ljapunov, exponents of dynamical systems* (Russian), Trudy Moskov. Mat. Obšč. **19** (1968), 179–210. MR0240280

[Per59] Albert Perez, *Information theory with an abstract alphabet. Generalized forms of McMillan's limit theorem for the case of discrete and continuous times*, Theor. Probability Appl. **4** (1959), 99–102, DOI 10.1137/1104007. MR122613

[Pin64] M. S. Pinsker, *Information and information stability of random variables and processes*, Translated and edited by Amiel Feinstein, Holden-Day, Inc., San Francisco, Calif.-London-Amsterdam, 1964. MR0213190

[Sha48] C. E. Shannon, *A mathematical theory of communication*, Bell System Tech. J. **27** (1948), 379–423, 623–656, DOI 10.1002/j.1538-7305.1948.tb01338.x. MR26286

[Ste70] Elias M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Mathematical Series, No. 30, Princeton University Press, Princeton, N.J., 1970. MR0290095

Facultad de Ingeniería, IMERL, Julio Herrera y Reissig 565, 11300 Montevideo, Uruguay

Email address: `plessa@fing.edu.uy`