Journal of Statistical Physics, Vol. 113, Nos. 1/2, October 2003 (© 2003)

Random Versus Deterministic Exponents in a Rich Family of Diffeomorphisms

François Ledrappier,[1] Michael Shub,[2] Carles Simó,[3] and Amie Wilkinson[4]

Received October 7, 2002; accepted April 2, 2003

We study, both numerically and theoretically, the relationship between the random Lyapunov exponent of a family of area preserving diffeomorphisms of the 2-sphere and the mean of the Lyapunov exponents of the individual members. The motivation for this study is the hope that a rich enough family of diffeomorphisms will always have members with positive Lyapunov exponents, that is to say, positive entropy. At question is what sort of notion of richness would make such a conclusion valid. One type of richness of a family—invariance under the left action of $SO(n+1)$—occurs naturally in the context of volume preserving diffeomorphisms of the n-sphere. Based on some positive results for families linear maps obtained by Dedieu and Shub, we investigate the exponents of such a family on the 2-sphere. Again motivated by the linear case, we investigate whether there is in fact a lower bound for the mean of the Lyapunov exponents in terms of the random exponents (with respect to the push-forward of Haar measure on $SO(3)$) in such a family. The family $\mathscr{F}_\varepsilon$ that we study contains a twist map with stretching parameter ε. In the family $\mathscr{F}_\varepsilon$, we find strong numerical evidence for the existence of such a lower bound on mean Lyapunov exponents, when the values of the stretching parameter ε are not too small. Even moderate values of ε like $\varepsilon \geqslant 10$ are enough to have an average of the metric entropy larger than that of the random map. For small ε the estimated average entropy seems positive but is definitely much less than the one of the random map. The numerical evidence is in favor of the existence of exponentially small lower and upper bounds (in the present example, with an

Dedicated to Yakov Sinai on his 65th birthday.

[1] Centre de Mathématiques, École Polytechnique, 91128 Palaiseau Cedex, France; e-mail: ledrappi@math.polytechnique.fr

[2] Department of Mathematics, University of Toronto, 100 St. George Street, Toronto, Ontario M5S 3G3 Canada and IBM T. J. Watson Research Center, P.O. Box 218, Yorktown Heights, New York 10598; e-mail: mshub@math.toronto.edu

[3] Departamento de Matemàtica, Aplicada i Anàlisi, Universitat de Barcelona, 08071 Barcelona, Spain; e-mail: carles@maia.ub.es

[4] Mathematics Department, Northwestern University, Evanston, Illinois 60208-273; e-mail: wilkinso@math.northwestern.edu

0022-4715/03/1000-0085/0

analytic family). Finally, the effect of a small randomization of fixed size δ of the individual elements of the family $\mathscr{F}_\varepsilon$ is considered. Now the mean of the local random exponents of the family is indeed asymptotic to the random exponent of the entire family as ε tends to infinity.

KEY WORDS: Lyapunov estimates; random diffeomorphism; twist maps; rich families.

1. INTRODUCTION

Numerical experiments with area-preserving surface diffeomorphisms often produce the following dynamical picture: elliptical islands floating in ergodic seas. A reasonable guess is that these ergodic seas typically have positive measure, and further, that the Lyapunov exponents on these seas are on average nonzero. An example of tiny elliptical islands in the context of differential equations can be found in ref. 1, where all rough numeric tests are in favor of ergodic behavior.

In this paper, we add to the pile of experimental evidence in favor of this conjecture. We also discuss a *possible* theoretical approach to finding positive Lyapunov exponents in certain families of area-preserving diffeomorphisms of the sphere S^2. The possibility of such an approach was discussed in ref. 2. The families we consider are not obtained from a specific set of equations, but from the following construction. Let $f: S^2 \to S^2$ be an area-preserving diffeomorphism of the round sphere, and let $SO(3)$ be the isometry group of S^2. Let

$$\mathscr{F} = \{g \circ f \mid g \in SO(3)\}$$

be the left $SO(3)$-coset of f in $\mathrm{Diff}(S^2)$, and let ν be the push-forward of Haar measure on $SO(3)$ to $\mathscr{F}$. Provided that f is not itself an isometry, the family $\mathscr{F}$ has nonzero *random* Lyapunov exponents with respect to ν (see Proposition 2.2 later). The question this paper addresses is whether these random exponents can somehow be connected to the Lyapunov exponents of individual members of $\mathscr{F}$, at least *on ν-average*.

To test whether there might be such a connection, we chose f to be a twist map, all of whose Lyapunov exponents are zero. The resulting family $\mathscr{F}$ has similarities to the standard family on the 2-torus. The dynamics of the individual elements of $\mathscr{F}$ and how they depend on parameters is an interesting topic, but we only study here some key properties in the case of small ε. We mainly focus on two quantities, the *random exponent* $R(\nu)$ and the *average exponent* $\Lambda(\nu)$, which we now define.

Let μ be Lebesgue measure on S^2 normalized to be a probability measure. Suppose for now that ν is an arbitrary Borel probability measure supported on a subset $\mathscr{F}$ of $\mathrm{Diff}_\mu(S^2)$, the space of μ-preserving diffeomorphisms of S^2. For $f \in \mathscr{F}$, the largest Lyapunov of f at $x \in S^2$ is found by computing the limit:

$$\lim_{n \to \infty} \frac{1}{n} \log \|T_x f^n\| =: \lambda_1(x, f), \tag{1}$$

which exists for μ-almost every x by the subadditive ergodic theorem. We define the *average exponent of f* to be

$$\lambda(f) = \int_{S^2} \lambda_1(x, f)\, d\mu(x) \tag{2}$$

and the *average exponent of ν* to be

$$\Lambda(\nu) = \int_{\mathrm{Diff}^r_\mu(S^2)} \lambda(f)\, d\nu(f). \tag{3}$$

Rather than iterate a single diffeomorphism $f \in \mathscr{F}$, we might choose instead a sequence of diffeomorphisms $\{f_1, f_2, \ldots\} \subset \mathscr{F}$ and form their composition:

$$f^{(n)} := f_n \circ f_{n-1} \circ \cdots \circ f_1.$$

If the sequence is chosen to be independent and identically distributed with respect to ν, then almost surely the limit

$$\lim_{n \to \infty} \frac{1}{n} \log \|T_x f^{(n)}\| =: R(x, (f_i)_1^\infty, \nu) \tag{4}$$

will exist, for μ-almost every x. (This too follows from the subadditive ergodic theorem, applied in the appropriate context.) Further, the integral of $R(x, (f_i)_1^\infty, \nu)$ with respect to μ is almost surely independent of the sequence $(f_i)_1^\infty$. We define the *random exponent* of ν to be this integral:

$$R(\nu) = \int_{S^2} R(x, (f_i)_1^\infty, \nu)\, d\mu(x), \tag{5}$$

(see Kifer for an introduction to the subject of random diffeomorphisms and their exponents. We also give a self-contained introduction in Section 2). The random exponent $R(\nu)$ is usually positive, unless ν is fairly degenerate.[3]

The quantity $\Lambda(\nu)$ is mysterious from a computational perspective, but useful from a dynamical one. The quantity $R(\nu)$ is relatively easy to estimate and is often positive.

Our goal is to understand in general if there is a notion of richness for a probability measure ν on the volume preserving diffeomorphisms of a closed manifold M such that the positivity of $R(\nu)$ implies the positivity of $\Lambda(\nu)$. Here we are investigating whether the $SO(3)$ invariance of the measure ν on $\mathrm{Diff}_\mu(S^2)$ might provide such a notion of richness. In ref. 2 we asked if even more might hold, that we might bound $\Lambda(\nu)$ from below in terms of $R(\nu)$.

Question 1.1. Is there a positive constant C—perhaps 1—such that $\Lambda(\nu) \geqslant CR(\nu)$?

Some motivation for Question 1.1 can be found in similar question for the iterates of linear maps (see ref. 4 where an affirmative answer to the analogue of Question 1.1 is proven with $C = 1$). In Section 3, we describe a theoretical framework in which to address Question 1.1 and related questions. We discuss the linear case in Section 4.

Returning to the specific family of diffeomorphisms mentioned earlier, we now describe the experiment in more detail.

For $\varepsilon > 0$, we define a one-parameter family of twist maps f_ε as follows. Express S^2 as the sphere of radius $1/2$ centered at $(0, 0)$ in $\mathbf{R} \times \mathbf{C}$, so that the coordinates $(r, z) \in S^2$ satisfy the equation

$$|r|^2 + |z|^2 = 1/4.$$

In these coordinates define a twist map $f_\varepsilon\colon S^2 \to S^2$, for $\varepsilon > 0$, by

$$f_\varepsilon(r, z) = (r, \exp(2\pi i(r + 1/2)\, \varepsilon)\, z).$$

Let $\mathscr{F}_\varepsilon$ be the orbit $SO(3)\, f_\varepsilon$. Let ν be the push-forward of Haar measure on $SO(3)$. We denote the resulting random and average Lyapunov exponents by $R(\varepsilon)$ and $\Lambda(\varepsilon)$, respectively.

The numerical results are described in Section 6. To summarize these results, it appears that the inequality $\Lambda(\varepsilon) \geqslant R(\varepsilon)$ is satisfied for large ε, and it is definitely not satisfied for small ε. We now have rigorous results to confirm some of these observations. The strongest of these results is proved in Section 7: for ε close to 0, there is no $C > 0$ satisfying the inequality in Question 1.1: in fact, we show in Corollary 7.7 that for small ε, $\Lambda(\varepsilon)$ is less than $\varepsilon^3 \sim R(\varepsilon)^{3/2}$. The numerics support an upper bound on $\Lambda(\varepsilon)$ that is exponentially small, and we show in Theorem 7.6 that on most of $\mathscr{F}_\varepsilon$ this is indeed the case. On the other hand, $\Lambda(\varepsilon)$ does appear to be positive for positive ε in all of the range where we can meaningfully compute. Section 7 also contains a study of the bifurcation structure of fixed points inside the family $\mathscr{F}_\varepsilon$. We include this analysis because it sheds light on where, in both $\mathscr{F}_\varepsilon$ and S^2, new elliptic periodic points and their surrounding islands are produced. Homoclinic bifurcations also give rise to horseshoes, which are associated (at least heuristically) with positive measure sets with nonzero exponents.

For the case of large ε, we show in Section 8, that an inequality like that in Question 1.1 is satisfied when a small amount of noise is introduced. More precisely, we first prove in Section 3 some results about the quantities $R(\varepsilon)$ and $\Lambda(\varepsilon)$ and a third quantity $R(\varepsilon, \delta)$, which measures the exponents of the "in-between" process in which each element of $\mathscr{F}_\varepsilon$ has added noise in a δ-ball inside $\mathscr{F}_\varepsilon$ (see Section 3 for details). In particular, we prove that any element h of a family $\mathscr{F} = SO(3)\, f$ described above will have average δ-diffused exponents $R(h, \nu, \delta)$ that are positive, unless f is an isometry. In addition, there exists a stationary measure $m_{h,\delta}$ for such a process on the projective bundle PS^2 that is absolutely continuous with respect to Liouville measure and projects to Lebesgue measure μ on S^2; this measure is unique among stationary measures with these properties and is the unique fixed point of a "simple" linear operator. The integrated measure

$$m_\delta = \int_{\mathscr{F}} m_{h,\delta}\, d\nu(h)$$

determines $R(\nu, \delta)$, which is the average of the δ-diffused exponents $R(h, \nu, \delta)$ over $h \in \mathscr{F}$. Whenever m_δ is equal to Lebesgue measure m, we have the equality: $R(\nu, \delta) = \Lambda(\nu)$. For the family $\mathscr{F}_\varepsilon$, denote by $m_{\varepsilon,\delta}$ this integrated measure. In Section 8, we prove (Theorem 8.1) that for any $\delta > 0$, $\lim_{\varepsilon \to \infty} m_{\varepsilon,\delta} = m$. Using this result, we prove that if enough noise is introduced, then the inequality in Question 1.1 is satisfied as $\varepsilon \to \infty$; in particular, we show in Corollary 8.2 that $R(\varepsilon, \delta) - R(\varepsilon)$ tends to 0 as $\delta \to 0$ and $\varepsilon \to \infty$ sufficiently quickly, for instance $\varepsilon > \delta^{-25}$. Results of a similar nature for the standard family were obtained by Carleson and Spencer[5] and are described in Section 8 later.

We suspect that a further study of these measures m_δ would be interesting. Even for an $SO(n)$- or $SU(n)$-invariant family of matrices, the properties of the analogous "in-between" measures m_δ are, for the most part, unknown. In Section 4, we discuss what is known about these measures. For $SO(2)$, we prove that m_δ is Lebesgue measure for all $\delta > 0$, and for $SU(n)$, m_δ is *not* Lebesgue measure if δ is sufficiently small.

2. BACKGROUND ON RANDOM TRANSFORMATIONS AND EXPONENTS

In this section we introduce some notation and gather together some facts and propositions. What we have to say in this section and the next is standard and can be found for example in refs. 3, 6–10 in most cases in greater generality. We have outlined proofs here in order to be reasonably self contained.

If $\mathscr{H} \subset \mathrm{Diff}(M)$, and ν is a probability measure on $\mathscr{H}$, then $(\mathscr{H}, \nu)$ generates a random process given by selecting an independent, ν-distributed sequence $(h_i)_1^\infty \subset \mathscr{H}$ and forming the compositions:

$$h^{(n)} = h_n \circ h_{n-1} \circ \cdots \circ h_1.$$

To study all possible outcomes of this experiment, we introduce the following auxiliary spaces and transformations: the shift space, $\mathscr{H}^\infty := \Pi_{j=1}^\infty \mathscr{H}$, the one-sided shift $\sigma\colon \mathscr{H}^\infty \hookleftarrow$ given by:

$$\sigma(h_1, h_2, \ldots) = (h_2, h_3, \ldots),$$

and the skew product $\tau\colon \mathscr{H}^\infty \times M \hookleftarrow$ given by:

$$\tau((h_i)_1^\infty, x) = (\sigma((h_i)_1^\infty), h_1(x)).$$

Then σ has a natural invariant measure ν^∞, the product measure induced by ν, but *a priori* τ has no preferred invariant measure.

Definition 2.1. Let ν be a probability measure on $\mathscr{H} \subset \mathrm{Diff}(M)$. A measure μ on M is *stationary* for the random process given by $(\mathscr{H}, \nu)$ if any of the following equivalent conditions is satisfied:

1. $\tau_*(\nu^\infty \times \mu) = \nu^\infty \times \mu$
2. $ev_*(\nu \times \mu) = \mu$, where $ev\colon \mathscr{H} \times M \to M$ is the evaluation map:

$$ev(h, x) = h(x)$$

3. $\mu \star \nu = \mu$, where $\star$ is the convolution operator defined by:

$$\mu \star \nu(A) = \int_{\mathscr{H}} \mu(h^{-1}(A))\, d\nu(h),$$

for every μ-measurable subset $A \subset M$.

Stationary measures always exist[6] and are the random analogue of invariant measures in the nonrandom setting. Part of the focus of this paper is to find natural stationary measures in the case where $M = T_1S^2$ and $\mathscr{H}$ and ν are derived from Haar measure on $SO(3)$.

Given an injective linear map $A: V \to W$ between normed vector spaces we denote by $A_\#$ the induced map from the unit sphere in V to the unit sphere in W, which is defined by $v \to \frac{A(v)}{\|A(v)\|}$. We use the same notation for the induced map on the projective space PV. We denote the tangent bundle of S^2 by TS^2, the unit tangent bundle by T_1S^2, and the projective bundle by PS^2. We let m denote the normalized Liouville measure on PS^2, so m is a probability measure which pushes forward under projection to S^2 to μ. The fibers of TS^2 and PS^2 over a point $z \in S^2$ are denoted by T_zS^2 and P_zS^2. For any manifolds M, N and differentiable map $F: M \to N$ the derivative of F at $x \in M$ is denoted by T_xF; for $v \in T_xM$ we will usually write "TFv" instead of $T_xF(v)$. Finally, we denote by $F_\#: T_1M \to T_1N$ the map that covers F and is $(T_xF)_\#$ on the fiber over $x \in M$. Since the tangent map to $g \in SO(3)$ preserves unit tangent vectors, we will write "g" for $g_\#$.

Now let $f \in \mathrm{Diff}_\mu(S^2)$, let $\mathscr{F} = \{g \circ f \mid g \in SO(3)\}$, and let ν be the push-forward to $\mathscr{F}$ of Haar measure on $SO(3)$. Let m be normalized Liouville measure on PS^2. Associated to $\mathscr{F}$ we then have the set

$$\mathscr{F}_\# = \{h_\# \mid h \in \mathscr{F}\} = \{g \circ f_\# \mid g \in SO(3)\},$$

and the measure $\nu_\#$, the push-forward to $\mathrm{Diff}(PS^2)$ of Haar measure on $SO(3)$. Let $\sigma: \mathscr{F}^\infty \hookleftarrow$, $\tau: \mathscr{F}^\infty \times S^2 \hookleftarrow$, $\sigma_\#: \mathscr{F}_\#^\infty \hookleftarrow$, and $\tau_\#: \mathscr{F}_\#^\infty \times PS^2 \hookleftarrow$ be the associated auxiliary transformations to the random processes generated by $(\mathscr{F}, \nu)$ and $(\mathscr{F}_\#, \nu_\#)$ respectively.

Lemma 2.1. The measures μ and m are stationary for ν and $\nu_\#$ respectively.

The transformations τ, σ, and $\sigma_\#$ are ergodic with respect to $\nu^\infty \times \mu$, ν^∞, and $\nu_\#^\infty$, respectively.

Proof. It is straightforward to check that they are stationary. Ergodicity is not much harder to check. ∎

Now we can compute $R(\nu)$ more explicitly:

Proposition 2.2. Let $f \in \mathrm{Diff}_\mu(S^2)$, let $\mathscr{F} = \{g \circ f \mid g \in SO(3)\}$, and let ν be the push-forward to $\mathscr{F}$ of Haar measure on $SO(3)$. Let m be normalized Liouville measure on PS^2, the projective bundle of S^2. Then

$$R(\nu) = \int_{PS^2} \log \|Tfv\| \, dm(v).$$

Moreover, $R(\nu) > 0$, unless f is an isometry.

Proof. We first apply Birkhoff's Ergodic Theorem to the measure-preserving transformation $\tau_{\#}\colon \mathscr{F}_{\#}^{\infty} \times PS^2 \hookleftarrow$ and the function

$$\psi((h_{i\#})_1^{\infty}, v) = \log \|Tfv\|$$

to obtain that

$$\begin{aligned}
\lim_{p\to\infty} \frac{1}{p} \log \|Th_p \cdots Th_1(v)\| &= \lim_{p\to\infty} \frac{1}{p} \sum_{j=1}^{p} \log \|Tf(h_{j\#} \cdots h_{1\#}(v))\| \\
&= \lim_{p\to\infty} \frac{1}{p} \sum_{j=0}^{p-1} \psi(\tau_{\#}^{j}((h_{i\#})_1^{\infty}, v)) \\
&=: L((h_{i\#})_1^{\infty}, v)
\end{aligned}$$

exists a.e. in $\mathscr{F}_{\#}^{\infty} \times PS^2$. The integral of this limit L is

$$\begin{aligned}
\int L((h_{i\#})_1^{\infty}, v)\, d(\nu_{\#}^{\infty} \times m) &= \int_{\mathscr{F}_{\#}^{\infty} \times PS^2} \psi\, d(\nu_{\#}^{\infty} \times m) \\
&= \int_{PS^2} \log \|Tfv\|\, dm(v).
\end{aligned}$$

Next, we apply Oseledec's theorem to the map $\tau\colon \mathscr{F}^{\infty} \times S^2 \hookleftarrow$ and the cocycle $((h_i)_1^{\infty}, x) \mapsto T_x h_1$. We obtain that for almost all $(h_i)_1^{\infty}$, almost all $x \in S^2$, and for almost all $v \in T_{1,x}S^2$, the limit

$$K((h_i)_1^{\infty}, x) := \lim_{p\to\infty} \frac{1}{p} \log \|Th_p \cdots Th_1(v)\|$$

exists, is independent of v, and has $\nu^{\infty} \times \mu$-integral equal to $R(\nu)$. The function:

$$\begin{aligned}
K((h_i)_1^{\infty}) &:= \int_{S^2} K((h_i)_1^{\infty}, x)\, d\mu(x) \\
&= \int_{x\in S^2} \int_{u \in T_{1,x}S^2} \lim_{p\to\infty} \frac{1}{p} \log \|Th_p \cdots T_x h_1(u)\|\, du\, d\mu(x) \\
&= \int_{PS^2} \lim_{p\to\infty} \frac{1}{p} \log \|Th_p \cdots Th_1(v)\|\, dm(v)
\end{aligned}$$

is σ-invariant and has ν^∞-integral equal to $R(\nu)$; ergodicity of σ implies that it is a.e. constant and therefore equal to $R(\nu)$. We conclude that

$$R(\nu) = \int_{PS^2} \log \|Tf(v)\| \, dm(v).$$

It remains to see that $\int_{PS^2} \log \|Tfv\| \, dm(v) > 0$ if f is not an isometry. For this we use the following elementary lemma.

Lemma 2.3. Let $A \in SL(2, \mathbf{R})$. Then the Jacobian of $A_\sharp$ with respect to Lebesgue measure on S^1 is given by:

$$\mathrm{Jac}(A_\sharp)(v) = \|Av\|^{-2},$$

for $v \in S^1$.

Since f preserves μ, it follows from this lemma that the Jacobian of $f_\sharp$ with respect to m at $v \in T_1 S^2$ is $\|Tf(v)\|^{-2}$. Since $Tf_\sharp$ is a diffeomorphism,

$$\begin{aligned}\int_{PS^2} \|Tf(v)\|^{-2} \, dm(v) &= \int_{PS^2} \mathrm{Jac}(f_\sharp)(v) \\ &= 1.\end{aligned}$$

By Jensen's inequality,

$$\begin{aligned}\int_{PS^2} \log \|Tf(v)\|^{-2} \, dm(v) &\leqslant \log\left(\int_{PS^2} \|Tf(v)\|^{-2} \, dm(v)\right) \\ &= 0\end{aligned}$$

with inequality holding unless $\log \|Tfv\|$ is constant and equal to 0. Rearranging the inequality, we see that, unless f is an isometry, we must have $\int_{PS^2} \log \|Tf(v)\| \, dm(v) > 0$. ∎

3. A THEORETICAL FRAMEWORK

3.1. Connecting $R(\varepsilon)$ to $\Lambda(\varepsilon)$

In this section, we attempt to interpolate between $R(\varepsilon)$ and $\Lambda(\varepsilon)$ via a third quantity, $R(\varepsilon, \delta)$, which we call the *random δ-diffused exponent*. When δ is greater than or equal to the radius of $SO(3)$, $R(\varepsilon, \delta)$ is equal to $R(\varepsilon)$; as δ approaches 0, $R(\varepsilon, \delta)$ approaches (in lim sup) a lower bound for $\Lambda(\varepsilon)$. Roughly speaking, $R(\varepsilon, \delta)$ is the exponent (averaged over $\mathcal{F}_\varepsilon$) obtained by introducing random perturbations (viewed as noise) of order δ to each element of $\mathscr{F}_\varepsilon$, staying within the family $\mathscr{F}_\varepsilon$. In Lemma 3.2 we show that $\limsup_{\delta\to 0} R(\varepsilon, \delta) \leqslant \Lambda(\varepsilon)$. On the other hand, we derive in Proposition 3.3 a formula for $R(\varepsilon, \delta)$:

$$R(\varepsilon, \delta) = \int_{PS^2} \log \|Tfv\| \, dm_{\varepsilon,\delta}(v).$$

The probability measure $m_{\varepsilon,\delta}$ in this formula has nice properties: it projects to Lebesgue measure μ on S^2, and is absolutely continuous with respect to m, with smooth density.

Now, recall (Proposition 2.2) that

$$R(\varepsilon) = \int_{PS^2} \log \|Tfv\| \, dm(v).$$

If it were the case that $m_{\varepsilon,\delta} \to m$ as $\delta \to 0$, then it would follow that:

$$\begin{aligned}
\Lambda(\varepsilon) &\geqslant \limsup_{\delta\to 0} R(\varepsilon, \delta) \\
&= \limsup_{\delta\to 0} \int_{PS^2} \log \|Tfv\| \, dm_{\varepsilon,\delta}(v) \\
&= \int_{PS^2} \log \|Tfv\| \, dm(v) \\
&= R(\varepsilon).
\end{aligned}$$

Hence the properties of this measure $m_{\varepsilon,\delta}$ are potentially quite interesting with regard to Question 1.1.

Here we collect some properties of $m_{\varepsilon,\delta}$ and $R(\varepsilon, \delta)$. First of all, $R(\varepsilon, \delta)$ is always positive for $\delta > 0$ (in fact, we prove in Corollary 3.4 that this is true not just on average, but for individual elements of $\mathscr{F}_\varepsilon$). In other words, introducing noise (no matter how small) to an element $h \in \mathscr{F}_\varepsilon$ invariably produces positive exponents.

The measure $m_{\varepsilon,\delta}$ has additional properties as well. We prove that we can write:

$$m_{\varepsilon,\delta} = \int_{SO(3)} m_{g,\varepsilon,\delta} \, dg,$$

where, for each $g \in SO(3)$, $m_{g,\varepsilon,\delta}$ is the unique probability measure on PS^2 with the properties:

1. $m_{g,\varepsilon,\delta}$ is stationary for the δ-diffused process about gf_ε;

2. $m_{g,\varepsilon,\delta}$ is absolutely continuous with respect to m, with smooth density;

3. $m_{g,\varepsilon,\delta}$ projects to Lebesgue measure μ on S^2.

As part of the proof, we show that each measure $m_{g,\varepsilon,\delta}$ is the unique fixed point of a "simple" linear operator.

Finally, from the way $m_{\varepsilon,\delta}$ is constructed, it follows that $m_{\varepsilon,\delta}$ shares all of the symmetries of f_ε. In particular, the density $\varphi_{\varepsilon,\delta}$ is invariant under all rotations that fix the North pole. The further study of these measures $m_{g,\varepsilon,\delta}$ and $m_{\varepsilon,\delta}$ might be of independent interest. We discuss the linear version of $m_{\varepsilon,\delta}$ in Section 4. In Section 8, we examine the behavior of $m_{\varepsilon,\delta}$ as $\varepsilon \to \infty$.

We now turn to the proofs of assertions 1–3. We first prove a standard semicontinuity result for random exponents.

Lemma 3.1. Let $\{\gamma_i\}$ be a sequence of probability measures on $\mathscr{H} \subset \mathrm{Diff}(M)$ that converges weakly to a probability measure γ. Suppose that μ is stationary for the random process generated by $(\mathscr{H}, \gamma_i)$, for every i, (for example, if $\mathscr{H} \subset \mathrm{Diff}_\mu(M)$). Then

$$\limsup_{\gamma_i \to \gamma} R(\gamma_i) \leqslant R(\gamma).$$

Proof. Let $a_n((h_i)_1^\infty, x) = \log \|T_x h^{(n)}\|$. Then $a_n\colon \mathscr{H}^\infty \times M \to \mathbf{R}$ is subadditive with respect to τ. By the subadditive ergodic theorem it then follows that

$$\begin{aligned} R(\gamma) &= \int_{\mathscr{H}^\infty \times M} \lim_{n\to\infty} \frac{1}{n} a_n \, d(\gamma^\infty \times \mu) \\ &= \lim_{n\to\infty} \frac{1}{n} \int_{\mathscr{H}^\infty \times M} a_n \, d(\gamma^\infty \times \mu) \\ &= \inf_n \frac{1}{n} \int_{\mathscr{H}^\infty \times M} a_n \, d(\gamma^\infty \times \mu). \end{aligned}$$

Now for any fixed n we have

$$\begin{aligned} \limsup_{\gamma_i \to \gamma} R(\gamma_i) &\leqslant \limsup_{\gamma_i \to \gamma} \frac{1}{n} \int_{\mathscr{H}^\infty \times M} a_n \, d(\mu \times \gamma_i^\infty) \\ &= \frac{1}{n} \int_{\mathscr{H}^\infty \times M} a_n \, d(\mu \times \gamma^\infty). \end{aligned}$$

So

$$\limsup_{\gamma_i \to \gamma} R(\gamma_i) \leqslant \inf_n \frac{1}{n} \int_{\mathscr{H}^\infty \times M} a_n \, d(\mu \times \gamma^\infty)$$
$$= R(\gamma). \quad \blacksquare$$

We will apply this lemma to the situation where γ_i is supported on a small ball in $\mathscr{F}_\varepsilon$ converging, as $i \to \infty$, to a Dirac measure supported on an element of $\mathscr{F}_\varepsilon$.

Let $\delta > 0$, and let U_δ be a symmetric δ ball around the identity in $SO(3)$. Give U_δ the restriction of Haar measure, normalized to be a probability measure and similarly for $\mathscr{F}_{g,\varepsilon,\delta} := U_\delta g f_\varepsilon$, for every $g \in SO(3)$ and $\varepsilon > 0$. We denote this last measure by $\nu_{g,\varepsilon,\delta}$. Let $R(g, \varepsilon, \delta) = R(\nu_{g,\varepsilon,\delta})$.

Definition 3.1. The (δ-)*diffused random exponent* is the average over $SO(3)$ of $R(g, \varepsilon, \delta)$:

$$R(\varepsilon, \delta) = \int_{g \in SO(3)} R(g, \varepsilon, \delta) \, d\nu(g).$$

Lemma 3.2.

$$\limsup_{\delta \to 0} \int_{g \in SO(3)} R(g, \varepsilon, \delta) \, dg \leqslant \Lambda(\varepsilon).$$

Proof. Note that $\lim_{\delta \to 0} \nu_{g,\varepsilon,\delta} = \delta_{gf_\varepsilon}$, Dirac measure supported on gf_ε. By the previous lemma,

$$\begin{aligned} \limsup_{\delta \to 0} R(g, \varepsilon, \delta) &= \limsup_{\delta \to 0} R(\nu_{g,\varepsilon,\delta}) \\ &\leqslant R(\delta_{gf_\varepsilon}) \\ &= \lambda(gf_\varepsilon) \end{aligned}$$

for each $g \in SO(3)$, so the same is true for the integral. $\blacksquare$

Now let $h_0 \colon S^2 \to S^2$ be any μ-preserving diffeomorphism, and let $\mathscr{H} = U_\delta h_0$. As in the previous section, define the space

$$\mathscr{H}_\# = \{h_\# \mid h \in \mathscr{H}\},$$

and evaluation maps

$$ev\colon \mathscr{H} \times S^2 \to S^2, \qquad ev_\#\colon \mathscr{H} \times PS^2 \to PS^2.$$

Let ν_δ and $\nu_{\delta\#}$ be the push-forwards of normalized Haar measure to $\mathscr{H}$ and $\mathscr{H}_\#$, respectively. Note that μ is stationary for the process generated by $(\mathscr{H}, \nu_\delta)$.

Proposition 3.3. If h_0 is not an isometry, then for fixed $\delta > 0$ the random process on PS^2 generated by $(\mathscr{H}_\#, \nu_{\delta\#})$ has a stationary measure m_δ that is absolutely continuous with smooth density, covers μ, and is the unique such stationary measure.

Moreover,

$$R(\nu_\delta) = \int_{PS^2} \log \|Th_0 v\| \, dm_\delta(v) > 0.$$

Corollary 3.4. For fixed $\delta > 0$, $\varepsilon \neq 0$, and $g \in SO(3)$, the random process on PS^2 generated by $(\mathscr{F}_{g,\varepsilon,\delta}, \nu_{g,\varepsilon,\delta})$ has a stationary measure $m_{g,\varepsilon,\delta}$ that is absolutely continuous with smooth density, covers μ, and is the unique such stationary measure.

Moreover,

$$R(g, \varepsilon, \delta) = \int_{PS^2} \log \|Tf_\varepsilon v\| \, dm_{g,\varepsilon,\delta} > 0.$$

Proof of Corollary 3.4. Apply Proposition 3.3 to the case where $h_0 = gf_\varepsilon$.

Proof of Proposition 3.3. Let h_0 and $\delta > 0$ be given. We break the proof into steps.

Step 1. Construction of m_δ. Recall that the convolution of a probability measure π on $\mathscr{H}_\#$ and a probability measure m on PS^2 is a probability measure $\pi \star m$ on PS^2 defined by

$$\pi \star m(E) = \int_{\mathscr{H}_\#} m(h_\#^{-1} E) \, d\pi(h_\#)$$

for every m-measurable $E \subseteq PS^2$. That a measure m is stationary for the measure π is equivalent to the fact that $\pi_\# \star m = m$. For $k > 1$ we let $\pi^k \star m = \pi \star (\pi^{k-1} \star m)$. For any probability measure m on PS^2 any weak limit of the Cesàro sums $\frac{1}{n}\sum_1^n \pi^k \star m$ is a stationary measure for π. Beginning with a measure m which pushes forward under projection to μ produces an invariant measure by this process with the same property. If we start with m as Liouville measure on PS^2 and $\pi = \nu_{\delta\#}$ we call this limiting measure m_δ.

Step 2. m_δ Is Absolutely Continuous, with Smooth Density. For any measurable set $A \subseteq PS^2$, we have:

$$\begin{aligned} m_\delta(A) &= (\nu_{\delta\#} \times m_\delta)\, ev_\#^{-1}(A) \\ &= (\nu_{\delta\#} \times m_\delta)\{(h_\#, v) \mid h_\#(v) \in A\} \\ &= \int_{v \in PS^2} \nu_{\delta\#}\{h_\# \mid h_\#(v) \in A\}\, dm_\delta(v). \end{aligned}$$

Now if the Liouville measure $m(A)$ equals zero, then $\nu_\delta\{h_\# \mid h_\#(v) \in A\}$ must also be zero, for every $v \in PS^2$. Thus $m_\delta(A)$ is zero. It follows that $m_\delta(A)$ is absolutely continuous with respect to m. So there is a non-negative integrable function φ_δ defined on PS^2 so that for any measurable $A \subseteq PS^2$,

$$m_\delta(A) = \int_A \varphi_\delta(x)\, dm(x).$$

In Lemma 3.6 we will prove that φ_δ satisfies the following formula:

$$\begin{aligned} \varphi_\delta(x) &= \frac{1}{m(B(x,\delta))} \int_{h_{0\#}^{-1} B(x,\delta)} \varphi_\delta(y)\, dm(y) \\ &= \frac{1}{m(B(x,\delta))} \int_{B(x,\delta)} \varphi_\delta(h_{0\#}^{-1} z)\, \mathrm{Jac}(h_{0\#}^{-1})(z)\, dm(z). \end{aligned}$$

It follows now fairly directly that φ_δ is as smooth as h_0, since the average over a δ-ball of an L^1 function is continuous, and of a C^k function, is C^{k+1}.

Step 3. $R(\nu_\delta)$ Satisfies the Integral Formula, and the Exponents of m_δ Are Nonzero. The argument that $R(\nu_\delta) = \int_{PS^2} \log \|Th_0 v\|\, dm_\delta$ is now the same as in the proof of Proposition 2.2, using Birkhoff's and Oseledec's theorems, where $\mathscr{F}_\#$ is replaced by $\mathscr{H}_\#$, ν by ν_δ and m by m_δ.

Next we will prove that the largest exponent is positive and from that we will deduce uniqueness. As in Proposition 2.2, for any $h = gh_0 \in \mathscr{H}$ the Jacobian of $h_\#$ with respect to m_δ at the vector v in T_1S^2 is

$$\rho(h, v) = \frac{\varphi_\delta(h_\# v)}{\|Thv\|^2\, \varphi_\delta(v)} = \frac{\varphi_\delta(gh_{0\#} v)}{\|Th_0 v\|^2\, \varphi_\delta(v)} \tag{6}$$

provided that $\varphi_\delta(v) \neq 0$.

We claim that the function $\rho(h, v)$ cannot be $\nu_\delta \times m_\delta$—almost everywhere equal to 1. Suppose for the sake of contradiction that $\rho(h, v) = 1$ a.e.

Since φ_δ is continuous, if we fix m_δ-a.e. v, then the function $\rho(\cdot, v)$ is continuous. Thus, for m_δ-a.e. v, we must have $\rho(h, v) = 1$ for *every* $h \in U_\delta$.

Next, notice in the expression (6) for $\rho(h, v)$, that the only term that depends on $g \in U_\delta$ is the numerator $\varphi_\delta(gh_{0\sharp}v)$. Rewriting this expression, we have, for almost every v in the support of φ_δ,

$$\varphi_\delta(gh_{0\sharp}v) = \|Th_0 v\|^2\, \varphi_\delta(v), \tag{7}$$

for every $g \in U_\delta$. Since m_δ projects to μ, and φ_δ is continuous, we have that for μ-a.e. $x \in S^2$, the set $O = \{v \,|\, \varphi_\delta(v) > 0\}$ is an open, $h_{0\sharp}$-invariant set in PS^2 that intersects almost every fiber. Equation (7) implies φ_δ must be constant on connected components of O, since varying g inside of U_δ, the vector $gh_{0\sharp}v$ covers an open neighborhood of $h_{0\sharp}v$ in PS^2.

But, again by Eq. (7), on each such component of O, we must have that $\|Th_0\|$ is constant. Since O intersects almost every fiber of PS^2, we obtain that for almost every $x \in S^2$, there exists a connected open set I_x in the fiber of PS^2 on which $T_x h_0$ has constant norm. Since h_0 preserves area, we must have $\|T_x h_0\| = 1$ for μ-a.e. x, contradicting the assumption that h_0 is not an isometry.

So $\rho(h, v)$ is not a.e. equal to 1, and by Jensen's inequality, we have:

$$\begin{aligned}\int_{\mathcal{H}\times PS^2} \log \rho(h, v)\, d\nu_\delta(h)\, dm_\delta(v) &< \log \int_{\mathcal{H}\times PS^2} \rho(h, v)\, d\nu_\delta(h)\, dm_\delta(v)\\ &= 0.\end{aligned}$$

But

$$\begin{aligned}R(\nu_\delta) &= \int_{PS^2} \log \|Th_0 v\|\, dm_\delta(v)\\
&= -\frac{1}{2}\int_{\mathcal{H}\times PS^2} \log \|Thv\|^{-2}\, d\nu_\delta(h)\, dm_\delta(v)\\
&= -\frac{1}{2}\int_{\mathcal{H}\times PS^2} \log\left(\|Thv\|^{-2}\frac{\varphi_\delta(h_\sharp(v))}{\varphi_\delta(v)}\right) d\nu_\delta(h)\, dm_\delta(v)\\
&= -\frac{1}{2}\int_{\mathcal{H}\times PS^2} \log \rho(h, v)\, d\nu_\delta(h)\, dm_\delta(v)\\
&> 0.\end{aligned}$$

(Here we used the stationarity of m_δ to conclude that the integral of $\log(\varphi_\delta(h_\sharp(v))/\varphi_\delta(v))$ is 0.)

Step 4. $m_\delta(A)$ Is Unique. It remains to prove that the measure m_δ is unique among absolutely continuous stationary measures which cover μ, now that we know that the random Lyapunov exponents are not zero.

Let γ be any such measure. Then, as for m_δ, there is a nonnegative smooth function ψ such that for any measurable $A \subseteq PS^2$,

$$\gamma(A) = \int_A \psi(v)\, dm(v).$$

Let γ_x be the disintegration of γ on the fiber $T_{1,x}S^2$. The density of γ_x with respect to Lebesgue measure on the fiber $T_{1,x}S^2$ is the restriction of ψ to the fiber.

We need the notion of a natural extension of a non-invertible transformation. Let $(\Omega, \mathscr{A}, p)$ be a probability space with p-preserving transformation $T: \Omega \to \Omega$. Let

$$\hat{\Omega} = \{(\ldots, \omega_{-1}, \omega_0) \in \Pi^1_{-\infty}\Omega \mid T(\omega_{-i}) = \omega_{-i+1}, \forall i \geqslant 1\}.$$

Let $\hat{T}: \hat{\Omega} \hookleftarrow$ be the map:

$$\hat{T}(\ldots, \omega_{-1}, \omega_0) = (\ldots, \omega_{-1}, \omega_0, T(\omega_0)),$$

and let $\hat{\pi}_0: \hat{\Omega} \to \Omega$ be the projection onto the first factor:

$$\hat{\pi}_0(\ldots, \omega_{-1}, \omega_0) = \omega_0.$$

Let $\hat{\mathscr{A}}$ be the smallest σ-algebra on $\hat{\Omega}$ so that $\hat{\pi}_0$ and $\hat{T}$ are both measurable. On $(\hat{\Omega}, \hat{\mathscr{A}})$, there is a unique probability measure $\hat{p}$, invariant under $\hat{T}$, that pushes forward under $\hat{\pi}$ to p. The measure-preserving system $(\hat{\Omega}, \hat{\mathscr{A}}, \hat{p}) \hookleftarrow \hat{T}$ is called the *natural extension* of $(\Omega, \mathscr{A}, p) \hookleftarrow T$. The natural extension $\hat{T}$ is invertible, and ergodic if T is ergodic.

Let τ and $\tau_\#$ be the associated auxiliary maps to the processes generated by $(\mathscr{H}, \nu_\delta)$ and $(\mathscr{H}_\#, \nu_{\delta\#})$, respectively. The natural extension of $\tau_\#$ with respect to the measure $\nu_{\#\delta}^\infty \times \gamma$ fibers over the natural extension of τ with respect to $\nu_\delta^\infty \times \mu$; the fiber over $((h_i)_{-\infty}^\infty, x)$ is $T_{1,x}S^2$. For $\hat{\nu}^\infty$-almost every $\mathbf{h} = (h_i)_{-\infty}^\infty$ and almost every fiber $T_{1,x}S^2$ there is a measure $\hat{\gamma}_{\mathbf{h},x}$ which is the disintegration of $\widehat{\nu^\infty \times \gamma}$ along the fiber. Note that the extension $\widehat{\nu^\infty \times \gamma}$ is determined by this system of measures, and therefore so is γ.

By (ref. 9, Proposition 1.1, p. 131), if $(\mathbf{h} = (h_i)_{-\infty}^\infty, x) \in \mathscr{H}_{-\infty}^\infty \times S^2 = \widehat{\mathscr{H}^\infty \times S^2}$, then $\hat{\gamma}_x$ is the limit of the push-forwards:

$$\hat{\gamma}_{\mathbf{h},x} = \lim_{n \to \infty} (h_{-n\#} \cdots h_{1\#})_* (\gamma_{h_{-n}^{-1} \cdots h_1^{-1}(x)}). \tag{8}$$

Lemma 3.5. The limit (8) does not depend on the initial choice of absolutely continuous stationary measure γ covering μ.

Proof. The average exponents of τ are nonzero, so the average exponents of $\hat{\tau}$ are also nonzero. It is easy to see that τ is ergodic, and thus, so is $\hat{\tau}$ and the exponents of $\hat{\tau}$ are in fact nonzero μ-a.e.

Let $u(\mathbf{h}, x) \in PS^2$ be the unstable Lyapunov direction for $\hat{\tau}$ over $(\mathbf{h}, x)$. For every $\epsilon > 0$, there exists an $n > 0$, and a set $G \subset \mathscr{H}_{-\infty}^{\infty} \times S^2$ such that

- $\nu_{-\infty}^{\infty}\mu(G) > 1 - \epsilon$
- for every $(\mathbf{h}, x) \in G$, the $\gamma_{\mathbf{h}, x}$-measure of an ϵ-neighborhood of $u(\mathbf{h}, x)$ in $T_{1,x}S^2$ is at least $1 - \epsilon$.

It follows that the limit in (8) is concentrated on the point $u(\mathbf{h}, x)$. ▮

Thus the natural extension of τ and γ is the same as the natural extension of τ and m_δ, so τ and m_δ are themselves equal. ▮

The next lemma completes the proof of Proposition 3.3.

Lemma 3.6. Let γ be an absolutely continuous measure with respect to m on PS^2 that is stationary for ν_δ. Then the density function ψ defining γ satisfies

$$\psi(x) = \frac{1}{m(B(x, \delta))} \int_{h_{0\#}^{-1} B(x, \delta)} \psi(y)\, dm(y).$$

Proof. For any measurable set $A \subseteq PS^2$, we have:

$$\begin{aligned}
\int_A \psi(y)\, dm(y) &= \gamma(A) \\
&= \nu_\delta \times \gamma(ev_{\#}^{-1}(A)) \\
&= \nu_\delta \times \gamma\{(h, y) \in \mathscr{H}_{\#} \times PS^2 \mid h(y) \in A\} \\
&= \int_{g \in U_\delta} \gamma\{h_{0\#}^{-1} g^{-1} A\}\, d\nu_\delta(g) \\
&= \int_{g \in U_\delta} \int_{h_{0\#}^{-1} g^{-1} A} \psi(y)\, dm(y)\, d\nu_\delta(g) \\
&= \int_{g \in U_\delta} \int_A \psi(h_{0\#}^{-1} g^{-1}(x))\, J(h_{0\#}^{-1} g^{-1}(x))^{-1}\, dm(x)\, d\nu_\delta(g) \\
&= \int_A \int_{g \in U_\delta} \psi(h_{0\#}^{-1} g^{-1}(x))\, J(h_{0\#}^{-1} g^{-1}(x))^{-1}\, d\nu_\delta(g)\, dm(x).
\end{aligned}$$

So,

$$\psi(x) = \int_{g \in U_\delta} \psi(h_{0\#}^{-1} g^{-1}(x)\, J(h_{0\#}^{-1} g^{-1}(x))^{-1}\, d\nu_\delta(g).$$

Since the integrand only depends on the point $x \in PS^2$ we push the measure ν_δ forward to PS^2, use that $J(g)^{-1} = 1$ and that Haar measure on $SO(3)$ pushes forward to Liouville measure on PS^2 to obtain

$$\psi(x) = \frac{1}{m(B(x, \delta))} \int_{y \in B(x, \delta)} \psi(h_{0\#}^{-1}(y))\, J(h_{0\#}^{-1}(y))\, dm(y).$$

Finally, changing variables one more time, gives:

$$\psi(x) = \frac{1}{m(B(x, \delta))} \int_{h_{0\#}^{-1} B(x, \delta)} \psi(y)\, dm(y). \quad \blacksquare$$

This completes the proof of Proposition 3.3.

Returning to discussion of the family $\mathcal{F}_\varepsilon$, we have verified that properties 1–3 hold for the measures $m_{g, \varepsilon, \delta}$. Now let

$$m_{\varepsilon, \delta} = \int_{g \in SO(3)} m_{g, \varepsilon, \delta}\, dg.$$

Since

$$\int_{SO(3)} R(g, \varepsilon, \delta)\, dg = \int_{PS^2} \log \|Tfv\|\, dm_{g, \varepsilon, \delta}(v)\, dg,$$

it follows from Corollary 3.4 and Lemma 3.2 that:

Proposition 3.7.

$$\Lambda(\varepsilon) \geqslant \limsup_{\delta \to 0} \int_{PS^2} \log \|Tf_\varepsilon(x)\, v\|\, dm_{\varepsilon, \delta}.$$

3.2. What Is this Measure $m_{\varepsilon, \delta}$?

For the family of twist maps $\mathcal{F}_\varepsilon$ under consideration, we prove in Section 7 that the answer to Question 1.1 is "no," and for small ε, we have $\Lambda(\varepsilon) < R(\varepsilon)$. It then follows from Proposition 3.7 that for small ε, Lebesgue measure m is not a weak limit of $m_{\varepsilon, \delta}$ as $\delta \to 0$. At the opposite extreme, we show in Section 8 that as ε tends to infinity, the measures $m_{\varepsilon,\delta}$ do approach Lebesgue measure, for $\delta > 0$ fixed. We hope that a future experiment will reveal more precisely how these measures behave in δ, for moderate values of ε. As will be seen in Section 6, the tiny differences between $\Lambda(\varepsilon)$ and $R(\varepsilon)$ for moderate and large values of ε, are expected to give rise to numerical difficulties in estimating the behavior with respect to δ.

In Section 4, we show that in $SO(2)$-invariant families of 2×2 matrices, if m_δ is the analogous "in-between" measure, averaged over the family, then m_δ is Lebesgue measure on S^1, for all $\delta > 0$. On the other hand, we also show that for unitarily-invariant families, as $\delta \to 0$, the m_δ do *not* approach the natural unitarily invariant measure on the appropriate Grassmannian manifold, but instead they limit on an even "better" measure, in the sense that this measure forces the (strict) inequality $\Lambda > R$. We describe the construction of m_δ for matrices in Section 4.

4. THE LINEAR CASE

Question 1.1 was originally motivated by a result of Dedieu–Shub about random and deterministic exponents for families of matrices. In this section, we describe these results and apply the framework of the previous section to the matrix setting.

Let L_i be a sequence of linear maps mapping finite dimensional normed vector spaces V_i to V_{i+1} for $i \in \mathbf{N}$. Let $v \in V_0 \setminus \{0\}$. If the limit $\lim \frac{1}{k} \log \|L_{k-1} \cdots L_0(v)\|$ exists it is called a Lyapunov exponent of the sequence. It is easy to see that if two vectors have the same exponent then so does every vector in the space spanned by them. It follows that there are at most $\dim(V_0)$ exponents. We denote them λ_j where $j \leqslant k \leqslant \dim(V_0)$. We order the λ_i so that $\lambda_i \geqslant \lambda_{i+1}$.

Given a probability measure μ on $GL(n, \mathbf{C})$, the space of invertible $n \times n$ complex matrices, we may form infinite sequences of elements chosen at random from μ by taking the product measure on $GL(n, \mathbf{C})^{\mathbf{N}}$. Thus we may also talk about the Lyapunov exponents of sequences or almost all sequences in $GL(n, \mathbf{C})^{\mathbf{N}}$.

For measures μ on $GL(n, \mathbf{C})$ satisfying a mild integrability condition, we have, by Oseledec's Theorem, n Lyapunov exponents $r_1 \geqslant r_2 \geqslant \cdots \geqslant r_n \geqslant -\infty$ such that, for almost every sequence $\cdots g_k \cdots g_1 \in GL(n, \mathbf{C})$, the limit

$$\lim \frac{1}{k} \log \|g_k \cdots g_1 v\|$$

exists for every $v \in \mathbf{C}^n \setminus \{0\}$ and equals one of the r_i, $i = 1 \cdots n$, see Gol'dsheid and Margulis[10] or Ruelle[11] or Oseledec.[12] We may call the numbers $r_1, \dots, r_n$ random Lyapunov exponents or even just random exponents. If the measure is concentrated on a point A, these numbers:

$$\lambda_i(A) = \lim \frac{1}{n} \log \|A^n v\|, \qquad i = 1 \dots n,$$

are $\log |e_1|, \dots, \log |e_n|$, where $e_i(A) = e_i$, $i = 1 \dots n$, are the eigenvalues of A written with multiplicity and $|e_1| \geqslant |e_2| \geqslant \cdots \geqslant |e_n|$.

The integrability condition for Oseledec's Theorem is

$$g \in GL(n, \mathbf{C}) \to \log^+(\|g\|) \qquad \text{is } \mu\text{-integrable}$$

where for a real valued function f, $f^+ = \max[0, f]$. Here we will assume more so that all our integrals are defined and finite, namely:

$$g \in GL(n, \mathbf{C}) \to \log^+(\|g\|) \quad \text{and} \quad \log^+(\|g^{-1}\|) \quad \text{are } \mu\text{-integrable.} \tag{$*$}$$

In this matrix setting, there are rigorous lower bounds for the average exponents (=logarithms of moduli of eigenvalues) of unitarily-invariant families in $GL(n, \mathbf{C})$. In ref. 4, the following bound is proved:

Theorem 4.1 (ref. 4). If μ is a unitarily invariant measure on $GL(n, \mathbf{C})$ satisfying ($*$) then, for $k = 1, \dots, n$,

$$\int_{A \in GL(n, \mathbf{C})} \sum_{i=1}^{k} \log |\lambda_i(A)| \, d\mu(A) \geqslant \sum_{i=1}^{k} r_i.$$

By unitary invariance we mean $\mu(U(X)) = \mu(X)$ for all unitary transformations $U \in U(n, \mathbf{C})$ and all μ-measurable $X \subset GL(n, \mathbf{C})$.

We can rephrase a special case of this theorem in a form similar to Question 1.1. Fix $A \in GL(n, \mathbf{C})$. As above, let v be normalized Haar measure on $U(n, \mathbf{C})$, and also denote by v the push-forward of v to the coset $U(n, \mathbf{C})\, A \subset GL(n, \mathbf{C})$. Let $R(A) = r_1(A)$ be the largest random exponent of v, and let

$$\Lambda(A) = \int_{B \in U(n, \mathbf{C})\, A} \log |e_1(B)| \, dv(B).$$

Then we have:

Corollary 4.2 (ref. 4). For $n \geqslant 2$, and for any $A \in GL(n, \mathbf{C})$,

$$\Lambda(A) \geqslant R(A).$$

Equality holds if and only if $A \in U(n, \mathbf{C})$.

Thus non-zero Lyapunov exponents for the family, i.e., non-zero random exponents, implies that at least some of the individual linear maps have non-zero exponents, i.e., eigenvalues of modulus not equal to 1. Hence the question that we posed for diffeomorphisms has a positive answer for sufficiently rich (i.e., unitarily-invariant) families of matrices.

Remark 4.3. Theorem 4.1 is not true for general measures on $GL(n, \mathbf{C})$ or $GL(n, \mathbf{R})$ even for $n = 2$. Consider

$$A_1 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \qquad A_2 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

and give probability $1/2$ to each. The left hand integral is zero but as is easily seen the right hand sum is positive. So, in this case the inequality goes the other way. We do not know a characterization of measures which make Theorem 4.1 valid.

We expect similar results for orthogonally invariant probability measures on $GL(n, \mathbf{R})$ but the only case in which such a result has been proved is in dimension 2, where we have:

Theorem 4.4 (ref. 4). Let μ be a probability measure on $GL(2, \mathbf{R})$ satisfying

$$g \in GL(2, \mathbf{R}) \to \log^+(\|g\|) \quad \text{and} \quad \log^+(\|g^{-1}\|) \quad \text{are } \mu\text{-integrable.}$$

(a) If μ is a $SO(2, \mathbf{R})$ invariant measure on $GL^+(2, \mathbf{R})$ then,

$$\int_{A \in GL^+(2, \mathbf{R})} \log |\lambda_1(A)| \, d\mu(A) = \int_{A \in GL^+(2, \mathbf{R})} \int_{x \in S^1} \log \|Ax\| \, dx \, d\mu(A).$$

(b) If μ is a $SO(2, \mathbf{R})$ invariant measure on $GL^-(2, \mathbf{R})$, whose support is not contained in $\mathbb{R}O(2, \mathbf{R})$, i.e., in the set of scalar multiples of orthogonal matrices, then

$$\int_{A \in GL^-(2, \mathbf{R})} \log |\lambda_1(A)| \, d\mu(A) > \int_{A \in GL^-(2, \mathbf{R})} \int_{x \in S^1} \log \|Ax\| \, dx \, d\mu(A).$$

Here $GL^+(2, \mathbf{R})$ (resp. $GL^-(2, \mathbf{R})$) is the set of invertible matrices with positive (resp. negative) determinant. To rephrase Theorem 4.4(a), fix $A \in GL^+(2, \mathbf{R})$, and let ν be the push forward of Haar measure on $SO(2, \mathbf{R})$ to the coset $SO(2, \mathbf{R})\, A$. Let $R(A)$ be the largest random exponent for the process induced by ν, and let $\Lambda(A) = \int_{B \in SO(2,\mathbf{R})A} \log |e_1(A)| \, d\nu(A)$. The we have:

Corollary 4.5 (refs. 4 and 13). For any $A \in GL^+(2, \mathbf{R})$,

$$\Lambda(A) = R(A).$$

We give an alternate proof of Corollary 4.5 in the following subsection.

4.1. m_δ for Matrices

Let $A \in GL(n, \mathbf{C})$ or $GL(n, \mathbf{R})$ and μ be the Haar measure on $U(n, \mathbf{C})$ or $SO(n, \mathbf{R})$, respectively, normalized to be a probability measure. Let G denote $GL(n, \mathbf{C})$ or $GL(n, \mathbf{R})$. As we did for families of diffeomorphisms in Section 3, we now interpolate between random products and deterministic powers of matrices by changing μ. Let $\delta > 0$ and G_δ be the δ-neighborhood of the identity in G. For $g \in G$, $G_\delta gA$ is a neighborhood of gA in GA. We normalize Haar measure restricted to G_δ and push it forward to $G_\delta gA$. Let us call this measure $\mu_{\delta,g}$. Let $r_1(\delta, g)$ be the largest random exponent for this measure. At the end of this subsection, we prove:

Proposition 4.6. $\lim_{\delta \to 0} r_1(\delta, g) = \log |e_1(gA)|$.

Let $G_{n,k}(\mathbf{C})$ denote the Grassmannian manifold of k dimensional vector subspaces in $\mathbf{C}^n$, and let m be the natural unitarily invariant probability measure on $G_{n,k}(\mathbf{C})$. Any $n \times n$ complex matrix acts on the homogeneous space $G_{n,k}$ by left-multiplication. For $A \in GL(n, \mathbf{C})$ and $P \in G_{n,k}(\mathbf{C})$, denote by $A \mid P$ the restriction of A to the subspace P. Now let $m_{\delta,g}$ be the stationary measure on the Grassmannian $G_{n,1}(= \mathbf{C}P^{n-1})$ induced by $\mu_{\delta,g}$.

Proposition 4.7. $r_1(\delta, g) = \int_{P \in G_{n,1}} \log \|(A \mid P)\| \, dm_{\delta,g}$.

Let $m_\delta = \int_{g \in G} \nu_{\delta,g} \, d\mu$. It follows that:

Proposition 4.8. $\int_{g \in G} \lambda_1(gA) \, d\mu = \lim_{\delta \to 0} \int_{P \in G_{n,1}} \log \|(A \mid P)\| \, dm_\delta$.

Now

$$r_1 = \int_{P \in G_{n,1}} \log \|(A \mid P)\| \, dm(P).$$

So a comparison of $\int_{g \in G} \lambda_1(gA) \, d\mu$ and r_1 can be achieved via an understanding of the relationship between m and m_δ. We have two results in this direction.

First, recall that inequality in Corollary 4.2 is strict when $n \geqslant 2$ unless A is an isometry. So for $G = SU(n)$ the measures m_δ favor the expanding directions of A as $\delta \to 0$. By Proposition 4.8, we obtain the immediate corollary:

Corollary 4.9. For $G = SU(n)$, $n \geqslant 2$,

$$\lim_{\delta \to 0} m_\delta \neq m,$$

unless A is an isometry.

Experimentally the same seems to hold for $SO(n)$ when $n > 2$, but we have not checked this very carefully.

By contrast, the equality in Corollary 4.5 is consistent with $\lim_{\delta \to 0} m_\delta = m$. In fact, more is true:

Theorem 4.10. For $G = SO(2)$,

$$m_\delta = m,$$

for all $\delta > 0$.

Combined with Proposition 4.8, Theorem 4.10 gives another proof of Corollary 4.5.

Proof of Proposition 4.6. Following the proof of Lemma 3.2, one obtains that

$$\log |e_1(gA)| \geqslant \lim_{\delta \to 0} r_1(\delta, g).$$

If A is replaced by cA, for $c \in \mathbf{C} \setminus \{0\}$, then both sides of the equality change by $\log |c|$. So we may assume that $|\det A| = 1$, and it will be enough to prove that $\log |e_1(gA)| \leqslant \lim_{\delta \to 0} r_1(\delta, g)$ under the hypothesis that $|e_1(gA)| > 1$.

Let $E \subset \mathbf{C}^n$ be the generalized eigenspace of the eigenvalues of gA whose modulus equals $|e_1(gA)|$. Then given $\varepsilon > 0$, there is a metric on $\mathbf{C}^n$, a closed cone $K \subset \mathbf{C}^n$ containing E in its interior, and a $\delta > 0$ such that, for any $B \in GL(n, \mathbf{C})$ in the δ-neighborhood of gA, we have:

1. $B(K) \subset K$, and
2. $\|Bv\| \geqslant (|e_1(gA)| - \varepsilon)\,\|v\|$, for all $v \in K$.

It follows that, for any $\varepsilon > 0$, there is a $\delta > 0$ such that,

$$\|B_n \cdots B_1\| \geqslant (|e_1(gA)| - \varepsilon)^n,$$

for all sequences $B_1, \ldots, B_n$ in $U_\delta gA$. Hence

$$r_1(\delta, g) \geqslant \log(|e_1(gA)| - \varepsilon),$$

and so

$$\lim_{\delta \to 0} r_1(\delta, g) \geqslant \log |e_1(ga)|. \quad \blacksquare$$

We now turn to the proof of Theorem 4.10.

Proof of Theorem 4.10. By an argument presented in the proof of Proposition 4.6, we may assume that $\det A = 1$. The projective action of $SL(2, \mathbf{R})$ on $\mathbf{R}P^1$ is conjugate to the standard action on the circle $S^1 = \{z \in \mathbf{C} \mid |z| = 1\}$ by linear fractional transformations. The conjugacy sends the rotation by θ to rotation by 2θ. Let $f\colon S^1 \to S^1$ be the linear fractional transformation induced by A, and let $\mathscr{F} = \{\alpha f \mid \alpha \in S^1\}$. Let

$$U_{\alpha,\delta} = \{\beta\alpha f \mid \arg(\beta) \in (-\delta, \delta)\},$$

and let $\nu_{\alpha,\delta}$ be normalized Lebesgue measure on $U_{\alpha,\delta}$, pushed forward from $(-\delta, \delta)$. Denote by $m_{\alpha,\delta}$ the stationary measure on S^1 induced by $\nu_{\alpha,\delta}$. We will show that:

$$m_\delta = \int_{S^1} m_{\alpha,\delta}\, d\alpha$$

is Lebesgue measure on S^1.

The same argument as in the proof of Lemma 3.6 shows that for every $\alpha \in S^1$,

$$dm_{\alpha,\delta}(z) = \varphi_{\alpha,\delta}(z)\, dz,$$

where

$$\varphi_{\alpha,\delta}(z) = \frac{1}{2\delta} \int_{y \in (\alpha f)^{-1} B(z,\delta)} \varphi_{\alpha,\delta}(y)\, dy,$$

and $B(z, \delta) = \{\beta z \mid \arg(\beta) \in (-\delta, \delta)\}$. Setting

$$k_\delta(z) = \frac{1}{2\delta} 1_{B(1,\delta)}(z),$$

we have that

$$\varphi_{\alpha,\delta}(z) = \int_{y \in S^1} k_\delta(\alpha \bar{z} f(y))\, \varphi_{\alpha,\delta}(y)\, dy,$$

where we use $\bar{z}$ to denote the multiplicative inverse of $z \in S^1$. Note that $\int_{S^1} k_\delta(z)\, dz = 1$.

Consider the following, more general setting. Let k be a non-negative function on S^1 such that $\int k = 1$ and $\int k^2 < \infty$ (all integrals are with respect to the normalized Lebesgue measure on S^1). For example, $k = k_\delta$. Let $f \colon S^1 \mapsto S^1$ be a fractional linear transformation, so that we can write, for $|z| = 1$:

$$f(z) = \sum_{n \geqslant 0} c_n z^n.$$

For $\alpha \in S^1$, define the operator $L_{\alpha,f}$ on real functions on S^1 by:

$$L_{\alpha,f}\varphi(z) = \int k(\alpha \bar{z} f(y))\, \varphi(y)\, dy.$$

The operator $L_{\alpha,f}$ is a positive operator, $\int L_{\alpha,f}\varphi = \int \varphi$. There exists a unique function φ_α satisfying $L_{\alpha,f}\varphi_\alpha = \varphi_\alpha$ and $\int \varphi_\alpha = 1$. The function φ_α is positive and continuous, upper and positive lower bounds for φ_α can be chosen uniformly in α. In the case $k = k_\delta$, we have $\varphi_\alpha = \varphi_{\alpha,\delta}$.

Proposition 4.11. We have, for all $z \in S^1$, $\int \varphi_\alpha(z)\, d\alpha = 1$.

The proposition follows directly from the following two claims:

Claim 1. For all $m \geqslant 0$, all $z \in S^1$, $\int L^m_{\alpha,f} 1(z)\, d\alpha = 1$.

Claim 2. The sequence $\frac{1}{n}\sum_{m=1}^{n} L^m_{\alpha,f} 1(z)$ converges to $\varphi_\alpha(z)$ in $L^1(d\alpha, dz)$ as $n \to \infty$.

Claim 2 follows from the ergodic theorem for Markov (i.e., $L1 = 1$) operators applied to the operator $\psi(\alpha, z) \to \frac{1}{\varphi_\alpha(z)} L_{\alpha,f}(\varphi_\alpha \psi(\alpha, .))(z)$ and the initial function $\frac{1}{\varphi_\alpha(z)}$.

In order to prove Claim 1, we compute, for a function $\varphi \in L^2$, $\varphi(z) = \sum_n \gamma_n z^n$, the Fourier coefficients γ'_n of the function $L_{\alpha,f}\varphi$. We find, for $n \geqslant 0$:

$$\begin{aligned}\gamma'_n &= \int \bar{z}^n L_{\alpha, f}\varphi(z)\, dz \\ &= \int \bar{z}^n k(\alpha \bar{z} f(y))\, \varphi(y)\, dy\, dz \\ &= \alpha^n \hat{k}(-n) \int (f(y))^n\, \varphi(y)\, dy \\ &= \alpha^n \hat{k}(-n) \int \sum_{k \geqslant 0} c_k^{(n)} y^k \varphi(y)\, dy \\ &= \alpha^n \hat{k}(-n) \sum_{k \geqslant 0} c_k^{(n)} \gamma_k\end{aligned}$$

and, analogously:

$$\gamma'_{-n} = \alpha^{-n} \hat{k}(n) \sum_{k \geqslant 0} c_k^{(n)} \gamma_{-k},$$

where we wrote $(f(z))^n = \sum_{k \geqslant 0} c_k^{(n)} z^k$, $(f(z))^0 = 1$, and $\hat{k}$ is the Fourier transform of k.

Iterating these formulas, we obtain for a function $\varphi \in L^2$, $\varphi(z) = \sum_n \gamma_n z^n$, the Fourier coefficients $\gamma_n^{(m)}$ of the function $L^m_{\alpha, f}\varphi$:

$$\gamma_n^{(m)} = \alpha^n \hat{k}(-n) \sum_{n_1, \ldots, n_m \geqslant 0} (\Pi_{s=1}^{m-1} \alpha^{n_s} \hat{k}(-n_s))\, c_{n_1}^{(n)} c_{n_2}^{(n_1)} \cdots c_{n_m}^{(n_{m-1})} \gamma_{n_m}$$

for $n \geqslant 0$, and

$$\gamma_{-n}^{(m)} = \alpha^{-n} \hat{k}(n) \sum_{n_1, \ldots, n_m \geqslant 0} (\Pi_{s=1}^{m-1} \alpha^{-n_s} \hat{k}(n_s))\, c_{n_1}^{(n)} c_{n_2}^{(n_1)} \cdots c_{n_m}^{(n_{m-1})} \gamma_{-n_m}$$

for $n \leqslant 0$.

To get the Fourier coefficients of the bounded continuous function $\int L^m_{\alpha, f} 1(z)\, d\alpha$, we integrate in α the Fourier coefficients of the bounded continuous functions $L^m_{\alpha, f} 1(z)$. In the above sum, all terms vanish, except the ones with $n + n_1 + \cdots + n_{m-1} = 0$. Since all n_i have the same sign, the one nonzero integral corresponds to $n = n_1 = \cdots = n_{m-1} = 0$. Claim 1 follows. ∎

Remark 4.12. Theorem 4.10 holds even without randomization. Suppose that A has determinant equal to 1 and let O vary over $SO(2, \mathbf{R})$.

Then for almost every O the eigenvalues of OA are either complex with irrational argument or real and there is one eigenvalue of modulus bigger than one. In the first case Cesàro sums of the push forward of Lebesgue measure by $OA_\sharp$ converge to the unique invariant measure of $OA_\sharp$. In the second case to the Dirac measure supported on the expanding eigenspace. Call these measures m_{OA}, then $\int m_{OA}\, dO$ is Lebesgue measure. The proof is the same, but easier.

5. EXPERIMENTAL METHOD

5.1. Haar Measure on $SO(3)$

It is clear that an element of $SO(3)$ is determined by its axis and angle of rotation. Here we describe how to pick axis and angle uniformly with respect to Haar measure on $SO(3)$.

Let S^2 be the usual two sphere with measure μ. Let S^1 be the usual unit circle of angles from 0 to 2π given the probability measure with density function $(1-\cos\theta)/(2\pi)$. Let $S^2 \times S^1$ be the product space with the product measure which we denote by γ. There is a natural map $P: S^2 \times S^1 \to SO(3)$ which maps a vector x and an angle θ to the orthogonal transformation which fixes x and rotates by angle θ around x according to the right hand rule. The map P sends $(x, 0)$ to the identity in $SO(3)$ for all $x \in S^2$ and is two to one when $\theta \neq 0$, (x, θ), and $(-x, -\theta)$ map to the same point. P maps γ to the Haar measure on $SO(3)$. If we identify $(x, \theta) \sim (-x, -\theta)$ we obtain $S^2 \times S^1/\sim$ which is a circle bundle over real projective 2-space. P induces a map $P^\sim: S^2 \times S^1/\sim \to SO(3)$.

Proposition 5.1. $P^\sim$ is one-one off of the zero section of $S^2 \times S^1/\sim$ and gives a measurable isomorphism between $(S^2 \times S^1/\sim, \gamma)$ and $(SO(3), \text{Haar})$.

Proof. That $P^\sim$ is one-one off the zero section of $S^2 \times S^1/\sim$ is easily verified.

Fix the standard product metric on $S^2 \times S^1$, normalized so that each factor has total volume 1. Normalized Haar measure on $SO(3)$ is Riemannian volume with respect to the bi-invariant metric we now describe. The Lie algebra of $SO(3)$ is the space of anti-symmetric matrices $so(3)$; on this algebra, we put the inner product:

$$\langle A, B\rangle = \frac{1}{2c^2}\, \mathrm{tr}(AB^t),$$

where $c=2\pi^{2/3}$. An orthonormal basis for $so(3)$ is $\{X, Y, Z\}$, where

$$X=\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -c \\ 0 & c & 0 \end{pmatrix}, \qquad Y=\begin{pmatrix} 0 & 0 & -c \\ 0 & 0 & 0 \\ c & 0 & 0 \end{pmatrix},$$

and

$$Z=\begin{pmatrix} 0 & c & 0 \\ -c & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

(Note $[X, Y]=cZ$, etc.). In the bi-invariant metric induced by this inner product, $SO(3)$ has constant sectional curvatures, see ref. 14:

$$\lambda=\frac{c^2}{4}=\frac{\|[X, Y]\|^2}{4}=\frac{\|[X, Z]\|^2}{4}=\frac{\|[Y, Z]\|^2}{4}.$$

The diameter of $SO(3)$ in this metric is $r=\pi/c$ and the total volume is

$$\frac{\pi}{\lambda}\left(2r-\frac{\sin(2\sqrt{\lambda}\, r)}{\sqrt{\lambda}}\right)=1.$$

Let $\rho(x, \theta)\, d\mu(x)\, d\theta$ be the pullback of the volume form on $SO(3)$ to $S^2\times S^1$ under P. To prove that $P^{\sim}$ is an isomorphism, it suffices to show that ρ is the density of γ, that is, to show that:

$$\rho(x, \theta)=(1-\cos\theta)/\pi,$$

for all $(x, \theta)\in S^2\times S^1$.

For any $x, y\in S^2$, if $N=A\times\Theta\subset S^2\times S^1$ is a product neighborhood of (x, θ) of Lebesgue measure δ, then there exists a $g\in SO(3)$ such that $\hat{N}=gA\times\Theta$ is a neighborhood of (y, θ) of Lebesgue measure δ. From the definition of P, it follows that $P(\hat{N})=gP(N)\, g^{-1}$. Since $\rho(x, \theta)\, d\mu(x)\, d\theta$ is the pullback of an $SO(3)$-invariant form, we obtain that for any $x, y\in S^2$, $\theta\in S^1$,

$$\rho(x, \theta)=\rho(y, \theta)=:\rho(\theta).$$

Finally, we compute $\rho(\theta)$. Since the geodesics of $SO(3)$ through I are precisely the one-parameter subgroups, the image under $P^{\sim}$ of the curve $t\mapsto(p, t)$ is a geodesic through I of speed $1/c$. It follows that for any

$\theta \in (0, \pi)$, $P^{\sim}$ sends $S^2 \times (0, \theta)/\sim$ diffeomorphically onto $B_{c^{-1}\theta}(I) \setminus \{0\}$, the punctured ball of radius $c^{-1}\theta$ about the identity in $SO(3)$. The volume of such a ball is

$$\begin{aligned}\int_0^\theta \rho(t)\, dt &= \mathrm{vol}(B_{c^{-1}\theta}(I)) \\ &= \frac{\pi}{\lambda}\left(2c^{-1}\theta - \frac{\sin(2\sqrt{\lambda}\, c^{-1}\theta)}{\sqrt{\lambda}}\right) \\ &= \frac{1}{\pi}(\theta - \sin\theta).\end{aligned}$$

By the Fundamental Theorem of Calculus,

$$\rho(\theta) = \frac{1}{\pi}(1 - \cos(\theta)), \tag{9}$$

which completes the proof. ∎

We call P or more appropriately $P^{\sim}$ polar coordinates on $SO(3)$.

5.2. Computing Random Exponents

Recall that $R(\varepsilon) = \int_{PS^2} \log \|Tf_\varepsilon v\|\, dm(v)$. Using the results of ref. 13 we can reduce the right hand integral to a one variable integral which we can then evaluate numerically very accurately. This is how the random Lyapunov exponents are computed.

Proposition 5.2.

$$R(\varepsilon) = \int_0^{\frac{1}{2}} \log(1 + (2\pi\varepsilon x(1-x))^2)\, dx.$$

Proof. Consider the "inverse Archimedean projection"

$$\Psi\colon (\theta, x) \mapsto (g(x)\cos\theta, g(x)\sin\theta, x - \tfrac{1}{2}),$$

where $g(x) = \sqrt{x(1-x)}$. This map sends the cylinder $C = S^1 \times [0, 1]$ onto the sphere S^2 and is area-preserving: the pullback $\Psi^*\, d\mu$ is a multiple of the Lebesgue volume form $d\theta\, dx$ on C. The Riemannian metric on S^2 pulls back to the metric:

$$\langle v, w\rangle_{(\theta, x)} = v^t B(x)^2\, w,$$

where

$$B(x) = [D\Psi(x,\theta)^t\, D\Psi(x,\theta)]^{1/2} = \begin{pmatrix} g(x) & 0 \\ 0 & (2g(x))^{-1} \end{pmatrix}.$$

Setting $\tilde{f}_\varepsilon = \Psi^{-1} \circ f_\varepsilon \circ \Psi$, we have that:

$$\tilde{f}_\varepsilon(\theta, x) = (\theta + 2\pi\varepsilon x, x).$$

We next compute:

$$R(\varepsilon) = \int_{PS^2} \log \|Tf_\varepsilon v\|\, dm(v).$$

$$= \int_{T_1C} \log \|T\tilde{f}_\varepsilon v\|\, \Psi^*_\sharp\, dm(v).$$

In the second equation, the unit tangent bundle T_1C and the quantity $\|T\tilde{f}_\varepsilon v\|$ are defined with respect to the Ψ-pullback Riemannian metric on C.

If $v \in TC$ is a unit vector with respect to the pullback metric, then $u = Bv$ is a unit vector with respect to the Euclidean metric, and $\|T\tilde{f}_\varepsilon v\| = \|BT\tilde{f}_\varepsilon B^{-1}u\|_{\text{Eucl.}}$. Hence we can write:

$$R(\varepsilon) = \frac{1}{2\pi} \int_{(x,\theta)\in C} \int_{u \in S^1} \log \|BT_{(x,\theta)}\tilde{f}_\varepsilon B^{-1}u\|_{\text{Eucl.}}\, dx\, d\theta\, du$$

$$= \frac{1}{2\pi} \int_{(x,\theta)\in C} \int_{u \in S^1} \log \left\| \begin{pmatrix} 1 & 4\pi\varepsilon x(1-x) \\ 0 & 1 \end{pmatrix} u \right\|_{\text{Eucl.}} dx\, d\theta\, du.$$

For $A \in SL(2, \mathbf{R})$, ref. 13 show that

$$\int_{u \in S^1} \log \|Au\|_{\text{Eucl.}}\, du = \log((\|A\|_{\text{Eucl.}} + \|A\|^{-1}_{\text{Eucl.}})/2).$$

Applying this to $A = \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix}$, we obtain that

$$\int_{u \in S^1} \log \|Au\|_{\text{Eucl.}}\, du = \tfrac{1}{2} \log(1 + \alpha^2/4).$$

It follows that:

$$R(\varepsilon) = \int_0^1 \int_{u \in S^1} \log \left\| \begin{pmatrix} 1 & 4\pi\varepsilon x(1-x) \\ 0 & 1 \end{pmatrix} u \right\|_{\text{Eucl.}} du\, dx$$

$$= \frac{1}{2} \int_0^1 \log(1 + (4\pi\varepsilon x(1-x))^2/4)\, dx$$

$$= \int_0^{1/2} \log(1 + (2\pi\varepsilon x(1-x))^2)\, dx.$$

6. EXPERIMENTAL RESULTS

In this section we describe several numerical experiments and their results. First we obtain experimental values of $R(\varepsilon)$ for random maps as introduced in Section 2. We use a method similar to those we use later to compute estimates for $\Lambda(\varepsilon)$. Proposition 5.2 allows to us to check the accuracy of the computed estimate of $R(\varepsilon)$ against the precise value of $R(\varepsilon)$ given there. Then we pass to the computation of $\Lambda(\varepsilon)$. Three different approaches to the computation of $\Lambda(\varepsilon)$ are presented and discussed which will allow us to obtain accurate enough values to draw conclusions. Finally, a sample of numerical estimates of $\Lambda(\varepsilon)$ is shown.

6.1. The Case of Random Maps

To obtain experimental values of $R(\varepsilon)$ for different ε we proceed as in Proposition 2.2. A random point x in S^2 and a random vector ξ in $T_{1,x}S^2$ is chosen. A random sequence g_i in $SO(3)$ is selected and the derivative of the maps $g_i f_\varepsilon$ are applied to the tangent vector ξ. The rate of increase of the logarithm of $\|T_x f^{(n)}(\xi)\|$ is described, where $f^{(n)} = g_n f_\varepsilon \cdots g_1 f_\varepsilon$. The results are the same with probability 1. For brevity, we refer loosely to the use of formula (4).

Let us describe the selection of random elements:

- An initial point can be described in polar coordinates by a longitude λ_x and a latitude β_x. The value of λ_x is chosen at random in $[0, 2\pi]$ with uniform probability. Concerning β_x, a random value $z \in [-1, 1]$ is selected with uniform probability and then we let $\beta_x = \sin^{-1} z$. This gives the uniform probability for $x \in S^2$.

- A tangent vector $\xi \in T_{1,x}S^2$ is generated by choosing an angle $\psi \in [0, 2\pi]$ with uniform probability and letting ξ make an angle ψ with the unit tangent vector to the latitude through x taken in the positive sense.

We call this last vector the horizontal vector at x. It is not defined at the poles, but as the poles have measure zero this is irrelevant in the current context.

- A random rotation $g \in SO(3)$ is determined by an axis and an angle of rotation. As described in Section 5.1 one can take these as elements in $S^2 \times S^1$. It is not necessary to carry out the identification described there. The axis is selected just as the point $x \in S^2$ was above. Let θ be the rotation angle. To select it according to formula (9) pick a random value $z \in [0, 2\pi]$ with uniform probability and solve the equation $z = \theta - \sin(\theta)$ for θ. The equation is nothing other than the well known Kepler equation with eccentricity equal to 1. There are efficient solvers for it.

Then, given initial values of (x, ξ) one can apply formula (4) to approximate $R(x, \varepsilon)$ by using a finite number of iterates, N. In turn, to approximate $R(\varepsilon)$ the integral in formula (5) can be computed using a sample of size M in PS^2. Let $R_{M,N}(\varepsilon)$ be the value obtained.

This value is compared to the one given by Proposition 5.2 which has been computed using a Simpson method with iterative mesh refinement (the values of $R(\varepsilon)$ are shown in Fig. 7). The following limit approximations are straightforward to derive:

$$\begin{aligned} R(\varepsilon) &= \frac{\pi^2}{15}\varepsilon^2 - \frac{2\pi^4}{315}\varepsilon^4 + O(\varepsilon^6) && \text{for} \quad \varepsilon \to 0, \\ R(\varepsilon) &= \log(2\pi\varepsilon) - 2 + \frac{1}{2\varepsilon} + O\left(\frac{\log(\varepsilon)}{\varepsilon^2}\right) && \text{for} \quad \varepsilon \to \infty. \end{aligned} \tag{10}$$

Skipping the O terms one has relative errors less than 0.01 if $\varepsilon < 0.30$ and $\varepsilon > 3.19$, respectively.

Let

$$\Delta_{M,N}(\varepsilon) = R_{M,N}(\varepsilon) - R(\varepsilon). \tag{11}$$

We have verified experimentally that $\Delta_{M,N}(\varepsilon)$ has essentially zero average and a standard deviation like

$$\sigma_{M,N}(\varepsilon) \approx \frac{\kappa(\varepsilon)}{\sqrt{MN}},$$

provided N is large enough.

Tests have been done for several choices of N, M, and ε. Using $M = 10^3$ and $N = 10^k$, $k = 2, \ldots, 6$, the values of $\kappa(\varepsilon)$ have been estimated for ε ranging from 10^{-1} to 10^3. There are no significant differences from

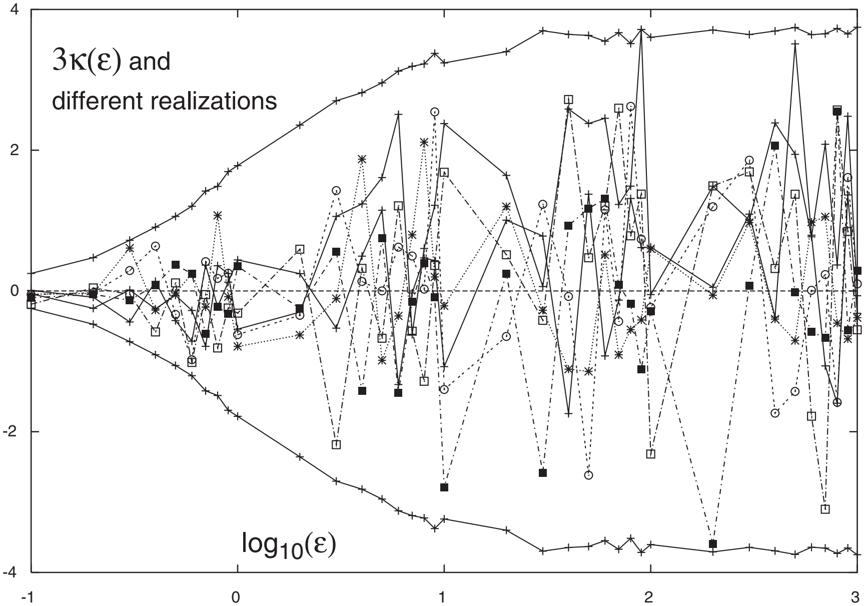

Fig. 1. Scaled intervals which are estimated to contain 99.8% of the results in the case of random maps and some realizations, as functions of ε.

$k = 4$ on. Figure 1 displays, for different values of ε, the interval $[-3\kappa(\varepsilon), 3\kappa(\varepsilon)]$ and the results of single runs (i.e., taking $M = 1$) for $N = 10^k$, $k = 4, \dots, 9$. More concretely, the plotted values are the deviations $\Delta_{1,N}(\varepsilon)$ given by formula (11) multiplied by $\sqrt{N}$.

These results indicate that $R_{M,N}(\varepsilon)$ and $R(\varepsilon)$ agree to order 10^{-5} taking $MN = 10^{10}$. Further checks have been done for larger values of ε (up to $\varepsilon = 10^6$) which show no significant variation of $\kappa(\varepsilon)$ between $\varepsilon = 10^3$ and $\varepsilon = 10^6$.

6.2. Computing the Lyapunov Exponent in the Deterministic Case

In principle one can follow a similar scheme to compute $\Lambda(\varepsilon)$. That is, using (1) with a finite number of iterates, N, an approximation of $\lambda_1(x, g \circ f_\varepsilon)$ is obtained. Then, (2) is computed using a Montecarlo method sampling $x \in S^2$ as described with samples of size M_p. Finally an estimate of $\Lambda(\varepsilon)$ is computed by applying again a Montecarlo method to (3), sampling $g \in SO(3)$ as explained above and using samples of size M_r. In any case the samples are taken according to the appropriate measures. The total number of iterates of the maps and their differential is, hence, $M_r M_p N$. Let us denote by $\Lambda_{M_r, M_p, N}(\varepsilon)$ a value obtained in this way.

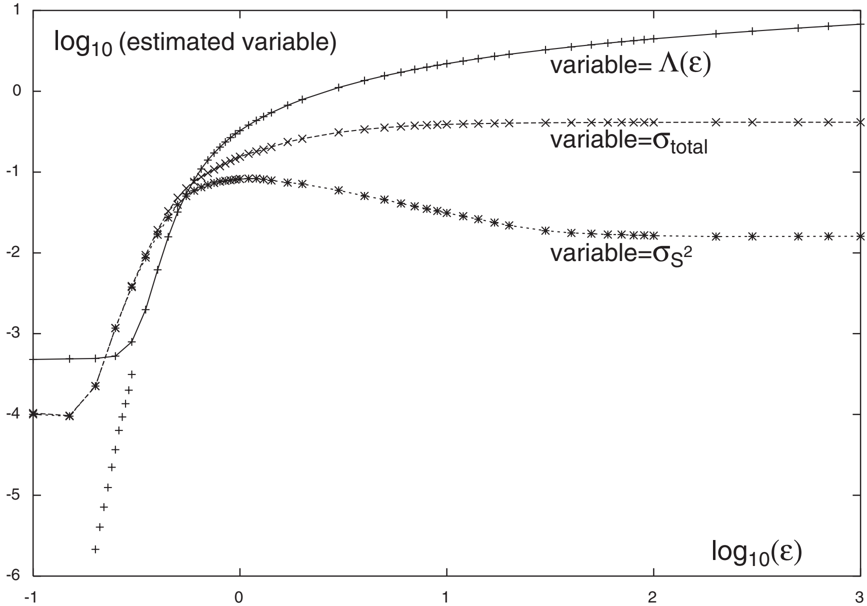

Fig. 2. Values of $\Lambda(\varepsilon)$, estimated by using mainly probabilistic methods, as a function of ε. Log_{10} scales used in both axes. The estimated values of the standard deviations are also shown. The results displayed here are obtained using $N = 8192$, $M_p = 5000$, $M_r = 5000$. The dots which appear in the lower left part correspond to the values computed using a different estimator, given by formula (12) with $(m, n) = (2, 0)$ and are closer to the real values of $\Lambda(\varepsilon)$. See the text for additional explanations.

Results of this approach are shown in Fig. 2 for different values of ε. They require some explanation. For a fixed $g \in SO(3)$ the values of $\lambda_1(x, g \circ f_\varepsilon)$ are estimated for M_p random values of x. The standard deviation of the values of $\lambda_1(x, g \circ f_\varepsilon)$ is then computed. This value, σ_g, depends of the choice of g. Let σ_{S^2} be the average value of σ_g when a full sample of $g \in SO(3)$ is considered. On the other hand all the $M_r M_p$ determinations of $\lambda_1(x, g \circ f_\varepsilon)$ can be used to estimate a global standard deviation, σ_{total}.

It is clear that σ_{S^2} measures the average dispersion of the maximal Lyapunov exponent when different points are taken in the phase space. The dispersion depends on the concrete rotation g taken. Typically, for the g such that relatively small values of the average $\lambda(g \circ f_\varepsilon)$ of $\lambda_1(x, g \circ f_\varepsilon)$ on S^2 are obtained, it is seen that σ_g is larger. It should also be mentioned that the errors in the determination of $\lambda_1(x, g \circ f_\varepsilon)$, due to the finiteness of the number of iterates N, also contribute to this dispersion.

On the other hand there is also a dispersion in $\lambda(g \circ f_\varepsilon)$ when different g are taken. The standard deviation σ_{total} measures the cumulative effect of both dispersions.

Now let us make several comments on the observed behavior, based on computations carried out with quite different values of M_r, M_p, and N.

- The estimates of $\Lambda(\varepsilon)$ are close to $R(\varepsilon)$ for ε large. For instance, for $\varepsilon > 3$ one has that $|\Lambda_{5000,\,5000,\,8192}(\varepsilon) - R(\varepsilon)|$, in the runs done, is below 0.00325. The situation is worse for small ε because the difference reaches the value -0.105 for $\varepsilon = 0.55$. If we proceed to compute the relative error (r.e.) $= (\Lambda_{5000,\,5000,\,8192}(\varepsilon) - R(\varepsilon))/R(\varepsilon)$ the agreement is even worse. For $\varepsilon = 1$ one has r.e. ≈ -0.14 and r.e. < -0.9 if $\varepsilon < 0.42$. In these comparisons one should take into account that the discrepancies also include the errors done in the estimates of $\Lambda(\varepsilon)$ (see Sections 6.3 and 6.4). But the present results already indicate that the differences for ε small are not only due to statistical errors. To make this more evident some additional computations have used a total number of iterates (for some selected ε) with $M_r M_p N$ largely exceeding 10^{12}.

Furthermore it seems also clear that for $\varepsilon < 0.3$ there is a "saturation" in the behavior of the estimates of $\Lambda_{M_r,\,M_p,\,N}(\varepsilon)$ and of the standard deviations. Indeed, it can be seen that the trend on the figure changes completely (this is also the purpose to use logarithmic scales). Systematic errors occur which completely invalidate the statistical results. To make this more evident some values, for ε in the range $[0.2, 0.3]$, computed also in a probabilistic way but with a different estimator (see Section 6.3) are also shown as dots in the lower left part. For these computations $M_r = 14400$, $M_p = 16384$, and $N = 16384$ have been used.

The maximal value $\lambda_{1,\max}(\varepsilon)$ of $\lambda_1(x, g \circ f_\varepsilon)$ for $x \in S^2$ and $g \in SO(3)$ is larger than $R(\varepsilon)$. This requires samples with M_r, M_p large to be detected if ε is small. It will be clear from Section 7 and the upper formula in (10) that the quotient $q(\varepsilon) = \lambda_{1,\max}(\varepsilon)/R(\varepsilon)$ is unbounded if $\varepsilon \to 0$. If large values of ε are considered, it is observed that $q(\varepsilon)$ tends slowly to 1 when $\varepsilon \to \infty$. In fact it follows from the lower formula in (10) and the analysis in Section 7 (which is partly valid for any ε) that the difference $\lambda_{1,\max}(\varepsilon) - R(\varepsilon)$ is bounded by $2 - \log(2) + O(\varepsilon^{-1})$. To see differences close to the bound one has to use very large values of M_r, M_p.

- The estimates of σ_{S^2} are mildly sensitive to the concrete values of M_r, M_p, and N, provided these values are not too small, and assuming $0.3 < \varepsilon < 3$. For $\varepsilon > 30$ a clear dependence with respect to N, of the form $N^{-1/2}$, is seen. From $\varepsilon = 3$ to $\varepsilon = 30$ there is a gradual increase in the dependence with respect to N. For $\varepsilon < 0.3$ a tendency towards a behavior of the form N^{-1}, which increases when ε decreases, is clear.

If extremal values of σ_{S^2} are considered when a sample of g is taken, it is clear that the minimum must be zero. But there is a significant difference for $\varepsilon < 2$, because the minimum is already close to zero for samples of moderate size, while for larger ε the minimum, which is almost insensitive to ε goes to zero slowly when M_r increases. On the other hand, the maximum of the observed values of σ_{S^2} increases until $\varepsilon \approx 10$. It only stabilizes to a value with small dependence on ε for $\varepsilon \approx 100$.

• The values of σ_{total} are much larger than σ_{S^2} for $\varepsilon > 1$, while for $\varepsilon < 0.3$ they are essentially equal. In particular they have a small dependence with respect to M_r, M_p, and N, if these are not too small and assuming $\varepsilon > 0.3$.

An analysis of the reasons of the observed behavior is useful because it helps in three different aspects: (a) to understand the different contributions to the errors in the estimates of $\Lambda(\varepsilon)$; (b) to see the main differences between the cases of random and deterministic maps; and (c) to suggest alternative methods to obtain better estimates.

• In the deterministic case the dynamics on S^2 for a given rotation g is relevant. This dynamics is "destroyed" (or "smoothed," "averaged") in the case of random maps. Hence, the initial point is irrelevant and the estimates improve in a probabilistic way, depending on the total number of iterates MN, in the random case.

• Changing g in the deterministic case produces dramatic changes in the dynamics and, hence, on $\lambda(g \circ f_\varepsilon)$. This is specially clear for g close to the identity (axis close to the pole or small rotation) or for rotations of angle very close to π around an axis of small latitude. The variability of $\lambda(g \circ f_\varepsilon)$ with respect to g is a major source of dispersion in the results, specially for large ε. A standard deviation σ_{total} around 0.4, mainly due to the variation of $\lambda(g \circ f_\varepsilon)$, would require sampling with M_r of the order of 10^8, at least, to have accurate results.

• For a fixed g, changing $x \in S^2$ has a very mild effect for large ε. Despite the possible existence of tiny islands (see Section 7) the dynamics "looks" ergodic. A similar behavior has been observed for standard–like maps in ref. 15 and in the case of volume preserving flows it is seen in ref. 1, where an analysis of the places where the islands should be expected is carried out before finding them explicitly.

Figure 3 displays a sample of orbits in S^2 for fixed g and two different values of ε. For $\varepsilon = 0.3$ the dynamics is mainly dominated by an integrable behavior, with many invariant curves and small chaotic seas, the largest one seen in the front part. This is persistent with respect to changes in g.

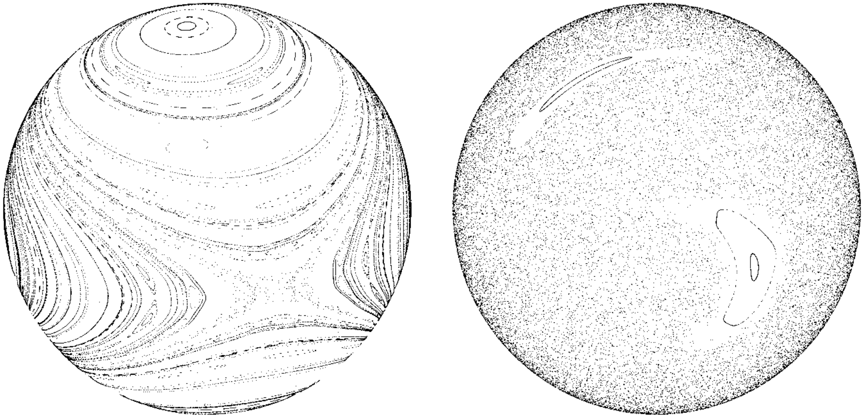

Fig. 3. Left: 3D view of orbits in S^2 for $\varepsilon = 0.3$. In $SO(3)$ a rotation of angle θ and axis of latitude β with $(\beta, \theta) = (\frac{18}{25} 2\pi\gamma, \frac{3}{25} 2\pi\gamma, \gamma = (\sqrt{5} - 1)/2)$ as in Fig. 4. A total of 128 random initial points has been taken and for each point 1024 iterates have computed. Points on the back have been plotted only partially and with smaller dots. Right: Vertical projection on the horizontal (x, y)-plane of points with $z < 0$ for $\varepsilon = 2$ and same θ and β as before.

The system is even more integrable (that is, invariant curves fill up a larger area) for most of the rotations $g \in SO(3)$. For $\varepsilon = 2$ (which is not so large!) only minor islands subsist, and they can even disappear for different g. For values like $\varepsilon = 10$ it is hard to see any island unless g is selected on a set of small measure. In the random case one observes a uniform distribution of iterates in S^2 and the same is essentially true for large ε in the deterministic case.

• On the other hand, for fixed g and small ε the value of $\lambda_1(x, g \circ f_\varepsilon)$ depends strongly on x. But the behavior is typically rather sharp. Either one obtains a moderate value of the order of $\varepsilon^{1.5}$ or it is zero. The smaller the value of ε, the larger the measure of the x with exponent zero, of course. The average value can be very small and despite the standard deviation σ_{S^2} also being small, large samples with respect to x have to be taken if small relative errors are desired.

• The worst point concerning accuracy, especially for small ε, are the errors in the computation of $\lambda_1(x, g \circ f_\varepsilon)$. Indeed, for an integrable motion (e.g., x in an invariant curve) λ_1 is zero, but the estimates $\frac{1}{n} \log \|T_x f^n\|$ are, generically, of the order of $\frac{\log(n)}{n}$. This implies that the convergence to zero is slow. For large ε this effect is relatively not so dramatic, but oscillations in the behavior of the quotients and different trends can be expected.

6.3. Improved Procedures

Several alternative procedures have been used in previous computations to determine the maximal Lyapunov exponent of a given map, averaged on the phase space:

(1) If the dynamics has a uniform hyperbolicity but with superimposed strong periodic or quasiperiodic oscillations, the following strategy has been used in ref. 16. It consists in detecting, by an iterative procedure, an upper envelope of the plot of the quotients $\frac{1}{n}\log\|T_x f^n\|$ as a function of n. Then, and after skipping a transient regime, one fits a function of the form $\alpha+\beta/n$ to the envelope. The value of α is a good estimator for $\lambda_1(x, f)$ and, as the system in ref. 16 is a skew product with linear action on the fibers, the value of x is irrelevant.

(2) If the values of $\lambda_1(x, f)$ depend strongly on x, it is possible to divide the phase space in pixels of a given size (in general, d-dimensional pixels) and start the computations at a point in each pixel. However, if the number of pixels is large and the system depends on several additional parameters, the method can be prohibitive from a computational point of view. Then, together with each initial pixel one considers all the pixels visited by the orbit. The estimated Lyapunov exponent is assigned to all of them. One requires each pixel to be visited a minimal number of times (in case of need one takes several initial points in the pixel) and an averaged Lyapunov exponent is assigned to the pixel. Later on this is averaged over the full phase space.

This method has been used in ref. 17 to study the classical Hill's problem and how the degree of chaos behaves with respect to the energy.

One should also take into account the stickiness of invariant curves. An initial point in a chaotic sea can remain close to an island for a large number of iterates. Hence, it is a good strategy to take a larger number of initial points even if one has to decrease the number of iterates for each one, provided this number is not too small. Furthermore, the local slope of $\log\|T_x f^n\|$ can have quite different trends if the number of iterates is large. Statistically this is not a problem because the interesting magnitude is the average behavior.

(3) The previous method still suffers from slow convergence of the quotients to $\lambda_1(x, f)$. An alternative method has been used in the context of flows with applications to galactic potentials in ref. 18 and later on extended to discrete transformations in ref. 19, where references to other applications can be found. It is mainly intended to discriminate between regular and chaotic motion (that is, to decide if one can accept $\lambda_1(x, f)=0$ or not), but it also supplies an estimate of the Lyapunov exponent.

Given a map f, an initial point x on a manifold $\mathscr{M}$ and a random vector $\xi \in T_{1,x}\mathscr{M}$, let $\xi_0 = \xi$ and define $\xi_k = (T_{f^{k-1}(x)} f)(\xi_{k-1})$. For fixed integers m and n and $N > 0$ let

$$Y_{m,n}(N) = N^n \sum_{k=1}^{N} \log\left(\frac{\|\xi_k\|}{\|\xi_{k-1}\|}\right) k^m, \qquad \bar{Y}_{m,n}(N) = \sum_{k=1}^{N} Y_{m,n}(k).$$

Then, for a chaotic orbit the estimator

$$\hat{Y}_{m,n}(N) = (m+1)(m+n+2)\,\frac{\bar{Y}_{m,n}(N)}{N^{n+m+2}} \tag{12}$$

tends to $\lambda_1(x, f)$, while for a generic regular orbit it behaves like $\frac{(m+1)(m+n+2)}{m(m+n+1)}\frac{1}{N}$. The basic idea is to average the exponential rate of increase of the length of ξ_k so that the transience has small relevance and to smooth out the irregularities of the quotients. Hence, it is a measure of the mean exponential growth of nearby orbits (MEGNO) and depends on the couple (m, n). Suitable values (according to numerical experience) are $m = 2$, $n = 0$, and then it is denoted as MEGNO20.

For $(m, n) = (2, 0)$ a slightly better estimator for $\lambda_1(x, f)$ is obtained by using $12\,\frac{\bar{Y}_{2,0}(N)}{N^4+4N^3+5N^2}$. Furthermore, when this method is used with these (m, n), one can check for a behavior of the form $\frac{2}{N}$ to decide $\lambda_1(x, f) \approx 0$. Typically $\hat{Y}_{2,0}(N) - \frac{2}{N} = O(N^{-2})$ for regular orbits.

Given a maximal number of iterates $N_{\max}$ to be used in the estimates, an additional question is whether it can be better to use another value $N < N_{\max}$ as a better choice to estimate $\lambda_1(x, f)$. In ref. 19 a "right stop" criterion is introduced. It is specially relevant if the orbit is close to be regular, to prevent an overestimate of $\lambda_1(x, f)$, but it has not been used in the present computations.

Figure 4 illustrates the different behavior of MEGNO20 and the quotients in formula (1) in several cases. Details on the parameters used for the plots are as follows.

On the upper row, left plot, for $\varepsilon = 0.3$ a rotation of angle θ and axis of latitude β with $\frac{\beta}{2\pi} = \frac{3}{25}\,g$, $\frac{\theta}{2\pi} = \frac{18}{25}\,g$ where $g = (\sqrt{5}-1)/2$ has been selected. Random initial conditions are chosen. After a transient of 512 iterates, estimates of $\lambda_1(x, f)$ are produced and plotted for the next 2048 iterates. Solid lines (the lower ones) correspond to the estimates using MEGNO20, while discontinuous lines are produced by the classical formula. The middle part displays a similar plot for different initial conditions and the right part is similar to the middle one but for $\varepsilon = 10$.

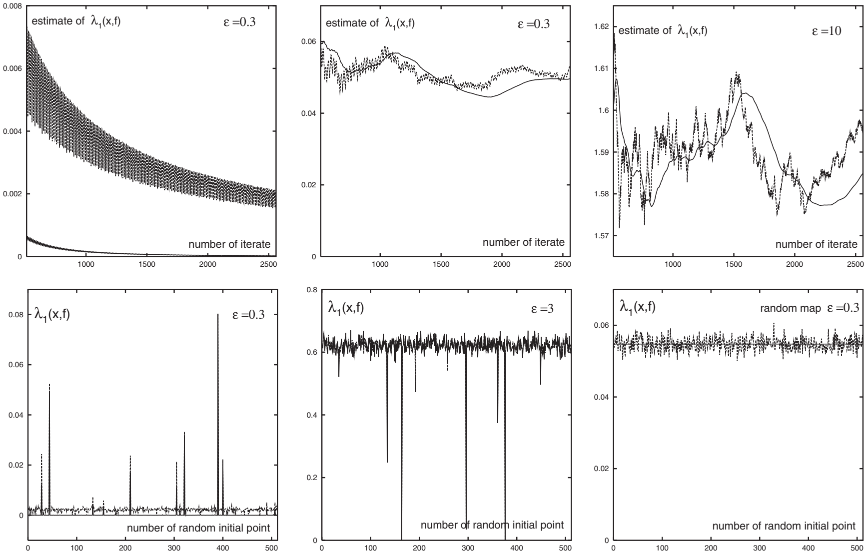

Fig. 4. Upper row: An illustration of different computations of the largest Lyapunov exponent. Left: A regular orbit for $\varepsilon = 0.3$. Middle: An irregular orbit also for $\varepsilon = 0.3$. Right: An irregular orbit for $\varepsilon = 10$. Lower row: Estimates of the largest Lyapunov exponent for random initial conditions. From left to right deterministic cases with $\varepsilon = 0.3$ and $\varepsilon = 3$ and a random case with $\varepsilon = 0.3$. See the text for details.

On the lower row estimates of the largest Lyapunov exponent for random initial conditions are shown. On the left part $\varepsilon = 0.3$ and θ and β as before have been used, and 512 random initial points are plotted. Solid (resp. discontinuous) lines correspond to MEGNO20 (resp. classical) estimators. Transient and number of iterates are as before. The left (middle) plot in the upper row corresponds to number 4 (43) of these points. The middle part is similar but for $\varepsilon = 3$. For completeness estimates in the random case using finite n in (4) are also shown on the right for $\varepsilon = 0.3$. Each iterate uses a random element in $SO(3)$ with density $\hat{\rho}(\beta, \theta) = \cos(\beta)(1 - \cos(\theta))/(2\pi)$. The solid line shows the value of $R(0.3) \approx 0.0547518$.

See also ref. 20 for additional methods and applications.

Due to the good properties of the procedure, estimates of $\lambda_1(x, g \circ f_\varepsilon)$ have been computed using procedure (3) earlier. For the integrations in S^2 a Montecarlo method has been used. This is good enough for large ε, because of the mild dependence of $\lambda_1(x, g \circ f_\varepsilon)$ with respect to x (for, say, $\varepsilon > 1$) for most of the $x \in S^2$ and most of the $g \in SO(3)$. For ε small and especially if $\varepsilon < 0.3$, a method such as the one presented in (2) would be suitable, but there are additional problems, due to the smallness of the Lyapunov exponents, to be discussed later. Furthermore it will turn out that it is relevant to compute $\Lambda(\varepsilon)$ with a small relative error for large ε, to allow for a careful comparison with $R(\varepsilon)$. But for ε small it will be clear from the results, even those obtained with a moderate accuracy, that $\Lambda(\varepsilon)$ is far below $R(\varepsilon)$. In any case, it seems that numerical estimates of $\Lambda(\varepsilon)$ for $\varepsilon < 0.3$ with small relative error require an enormous computational effort.

Finally, for the integrations in $SO(3)$ and taking into account the large standard deviation found for moderate and large values of ε, it has been found more convenient to use numerical quadrature formulas based on a grid of points. More concretely, a product Simpson method has been used with respect to the latitude β_g of the axis of rotation and the rotated angle θ_g. The longitude of the axis λ_g is irrelevant: changes in this longitude are equivalent to changes in the longitude λ_x of $x \in S^2$. Using a grid with $\theta_g \in [0, 2\pi]$, $\beta_g \in [0, \pi/2]$ requires that the estimate of $\lambda(g \circ f_\varepsilon)$ be multiplied by the factor $\cos(\beta_g)(1-\cos(\theta_g))$.

Initial estimates for the results shown in the next Section use, for the elements in $SO(3)$ the following data: $\lambda_g = 0$, (θ_g, β_g) on a grid of $M_g \times M_g$ with $M_g = 200$. Then a sample of $M_p = 1600$ random initial points and tangent vectors in PS^1 and $N = 8000$ iterates are used. The programs have been done in such a way that beyond the estimates for these values, also estimates using grids with $M_g = 100$, 50 or using samples with $M_p = 800$, 400, 200, 100 and doing a number of iterates equal to $N = 4000$, 2000, 1000, 500 are computed. This allows for a check on the internal consistency of the results.

It turns out that the use of different grids in $SO(3)$ stabilizes quickly. Concerning the dependence with respect to M_p and N, it is clearly seen that there is no need for very large values of M_p except in the case of small ε and one is interested in having small relative error. The dependence with respect to N is clearly of the form ctant/N. Hence, extrapolations with respect to N have been used. The initial estimates allow for a fine tuning of the most suitable values for the grid, M_p and N. For instance, assuming that one can accept a total of 2^{41} iterates (for every value of ε), for large ε a typical choice is $M_g = 2^8$, $M_p = 2^9$, $N = 2^{16}$, while for small ε it is $M_g = 2^7$, $M_p = 2^{10}$, $N = 2^{17}$. Even with this large N the results start to be not very good if $\varepsilon < 0.2$.

Figure 5 shows 3D views of $h(\theta_g, \beta_g, \varepsilon) = \lambda(g \circ f_\varepsilon)\cos(\beta_g)(1-\cos(\theta_g)) \times \frac{\pi}{2}$ as a function of (θ_g, β_g) for different values of ε. Level lines of these surfaces are displayed in Fig. 6. The plots give a good evidence of the smooth behavior of $h(\theta_g, \beta_g, \varepsilon)$ for moderate and large values of ε, and how the behavior becomes wilder, with sharp changes for small ε. It is clear that the results are the same if (θ, β) is replaced by $(2\pi-\theta, -\beta)$.

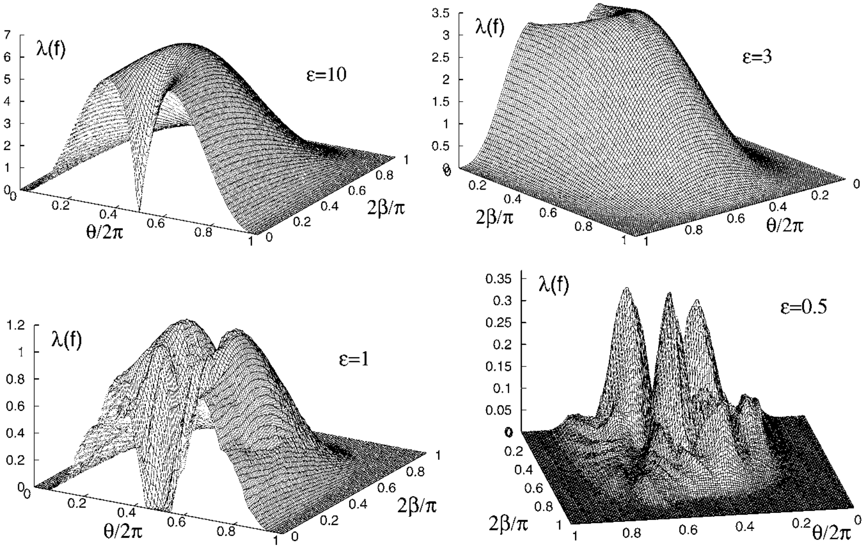

Fig. 5. Plots of the average exponent $\lambda(f)$ as a function of (β, θ) for the elements in $SO(3)$. The displayed values have been already corrected by the factor $\cos(\beta)(1-\cos(\theta))\,\pi/2$. The four plots correspond to different values of ε (10, 3, 1, 0.5) as shown.

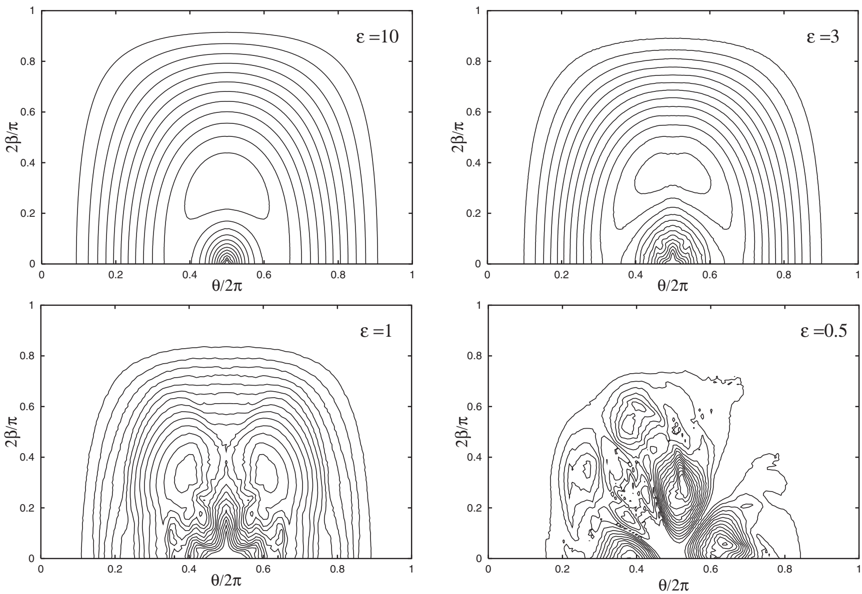

Fig. 6. Level lines of the surfaces shown in Fig. 5. The distances between consecutive levels are 0.5, 0.25, 0.08, and 0.025, respectively, for the decreasing values of ε.

Furthermore, note that for integer ε the symmetries of $g \circ f_\varepsilon$ imply that the results should be the same if θ is changed to $2\pi-\theta$, as clearly seen in the first three plots. For $\varepsilon = 0.5$ the lack of symmetry $\theta \leftrightarrow 2\pi-\theta$ is clear, but for large non-integer ε this lack of symmetry is harder to detect.

6.4. A Sample of Results

The results of applying the methodology just described are shown in Fig. 7. Typical values for the number of iterates, initial data and grid have been given before.

On the top plots general views of $R(\varepsilon)$ and estimates of $\Lambda(\varepsilon)$, to be denoted as $\Lambda(\varepsilon)_{\text{num}}$, can be seen. In particular, for large and small ε it is easy to check the limit behavior of $R(\varepsilon)$ predicted by (10). On top left only values $\varepsilon \leqslant 10$ are shown. On this scale no differences can be seen between $R(\varepsilon)$ and $\Lambda(\varepsilon)_{\text{num}}$ for $\varepsilon > 3$. For small ε the differences are clear and they are quite dramatic for $\varepsilon < 0.4$.

On the bottom left plot the tiny differences $\Lambda(\varepsilon)_{\text{num}} - R(\varepsilon)$ are displayed. It seems that they tend to 0 as $\varepsilon \to \infty$, in agreement with the

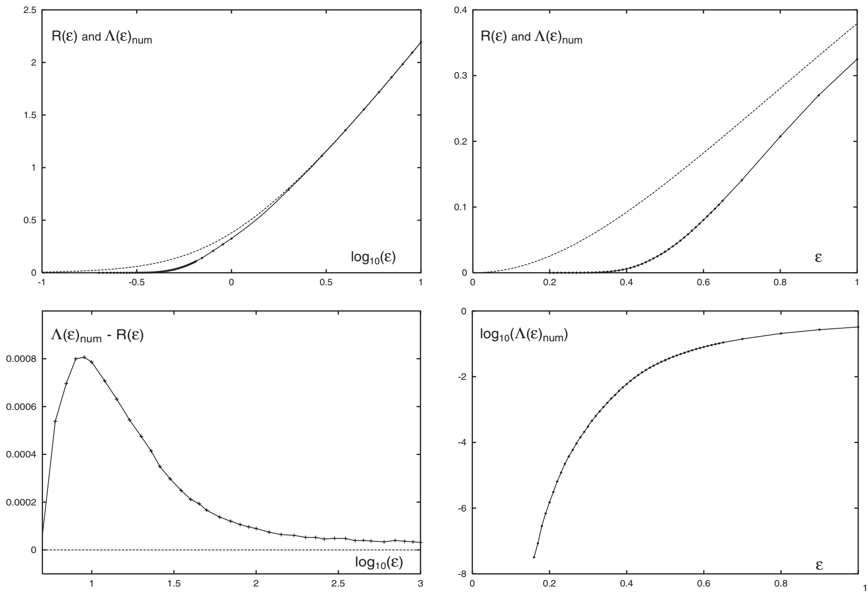

Fig. 7. Results of the numerical computation of $\Lambda(\varepsilon)$. Upper row: Left: A general view where the curve for $R(\varepsilon)$ is seen as the upper one. For larger values of ε one cannot see the difference with $R(\varepsilon)$. Right: A magnification for small ε (in $[0, 1]$). Lower row: Left: The small difference between $\Lambda(\varepsilon)_{\text{num}}$ and $R(\varepsilon)$ for ε large (in $[5, 1000]$). Right: The values of $\Lambda(\varepsilon)_{\text{num}}$ in logarithmic scale for ε small.

results in Section 8. Finally the behavior of $\Lambda(\varepsilon)_{\mathrm{num}}$ for ε small is seen in detail on the bottom right part. A logarithmic scale has been used to reveal that $\log(\Lambda(\varepsilon)_{\mathrm{num}})$ is dominated by a function of the form $-c/\varepsilon$ for some $c>0$. Due to the smallness of the estimates and to the fact that it would require an enormous effort to estimate $\Lambda(\varepsilon)$ for ε close to 0.1 (unless other methods, from deterministic analysis, are used), it is not completely clear what the correct behavior is. Even with a reduced set of data a fit of the values obtained for $0.16 \leqslant \varepsilon \leqslant 0.3$ gives a result of the form

$$\log(\Lambda(\varepsilon)_{\mathrm{num}}) \approx 2.45 - \frac{3.16}{\varepsilon} \tag{13}$$

which must be taken with caution, but seems to give the correct trend. This suggests that the inequality in Question 1.1 is not satisfied for small ε, which will be confirmed theoretically in Section 7. This fact is not a surprise, because similar facts occur in generic analytic families of area-preserving diffeomorphisms. The smallness of the Lyapunov exponent is related to the area of the chaotic seas which in turn is related to the splitting of the separatrices of fixed and periodic point for maps close to the identity. See ref. 21, where general upper estimates can be found. In fact, as with many a priori exponentially small upper bounds, this result can also be obtained as a corollary of averaging theory for analytic systems, see ref. 22 and also Section 7.2.

Finally it should be mentioned that some computations have been done for large values of ε (up to 10^6). Due to the strong chaotic properties it is enough to take small values (say $N_p = 256$) of the number of initial points in S^2. But the grid in the parameters (θ_g, β_g) has to contain more points. Typical values of N_g to have a good determination of $\Lambda(\varepsilon)_{\mathrm{num}}$ are 2^9 and 2^{10}. The results confirm what is seen in the left lower part of Fig. 7, that is, $\Lambda(\varepsilon)_{\mathrm{num}} > R(\varepsilon)$ and the difference goes to zero slowly.

7. THE CASE OF SMALL ε

As it is clear that the greatest problems occur for small ε, it is worth to carry out a preliminary analysis of the dynamics in this case. The first item to be studied is the location and stability of fixed points. This can be carried out, with the same effort, for any ε. Furthermore this allows us to see how bifurcations give rise to new elliptic fixed points with the corresponding creation of islands. Later on the global behavior of $g \circ f_\varepsilon$ on S^2 is discussed. In what follows it is assumed that $\varepsilon > 0$.

7.1. Fixed Points and Their Stability

To look for fixed points of $g \circ f_\varepsilon$ it is enough to consider axes of rotation with zero longitude. Let (β, θ) be the latitude of the axis and the angle of the rotation. As follows from Section 5.1, it is not restrictive to assume $\beta \in [0, \pi/2]$. Then a fixed point A is mapped by f_ε to a point A' which by g returns to A. Let b be the latitude of A. It is clear that A and A' must have symmetric longitudes, $-\delta(b)$ and $\delta(b)$, respectively, where $\delta(b) = \frac{\pi}{2}\varepsilon(1+\sin(b))$. It is easy to derive the condition for the fixed points

$$\sin\left(\frac{\theta}{2}\right)\sin(\beta)\cos(b)\cos(\delta(b)) - \sin\left(\frac{\theta}{2}\right)\cos(\beta)\sin(b) + \cos\left(\frac{\theta}{2}\right)\cos(b)\sin(\delta(b)) = 0. \tag{14}$$

For ε small one has $\sin(\delta(b)) = O(\varepsilon)$ and $\cos(\delta(b)) = 1 - O(\varepsilon^2)$. For the analysis of this case it is more convenient to write (14) in the form

$$\sin\left(\frac{\theta}{2}\right)\sin(\beta - b) - \varepsilon^2\frac{\pi^2}{4}\sin\left(\frac{\theta}{2}\right)\sin(\beta)\cos(b)(1+\sin(b))^2 + \varepsilon\frac{\pi}{2}\cos\left(\frac{\theta}{2}\right)\cos(b)(1+\sin(b)) + O(\varepsilon^3) = 0. \tag{15}$$

From (15) it is clear that if $\sin(\frac{\theta}{2}) = O(1)$, that is, θ is not too close to 0 from the positive or the negative side, then one has solutions for b of the form

$$b = \beta + \gamma\varepsilon + O(\varepsilon^2) \qquad \text{or} \qquad b = \pi + \beta + \gamma\varepsilon + O(\varepsilon^2),$$

where γ is independent of ε. The value for γ is given by

$$\gamma = \frac{\pi}{2}\cos(\beta)(1 \pm \sin(\beta)) \Big/ \tan\left(\frac{\theta}{2}\right),$$

where the + sign is used in the first case and the − sign in the second. In both cases b is close to either β or $\beta + \pi$ and an important thing is that there are exactly two fixed points for $g \circ f_\varepsilon$.

Otherwise one can write $\theta = m\pi\varepsilon$, $m > 0$ or $\theta = 2\pi + m\pi\varepsilon$, $m < 0$. The dominant terms in (15) become in both cases

$$m\sin(b - \beta) - \cos(b)(1+\sin(b)) = 0. \tag{16}$$

Equation (16) has to be seen as an equation for b depending on β and m (which accounts for θ). As it has zero average it should have at least two different zeros. To look for more solutions it is relevant to compute the lines (in (β, m)) where double zeros occur. The angle b can be used to parameterize these lines. They are represented as

$$\begin{aligned} m^2 &= 2+\sin^3(b)-\tfrac{3}{4}\sin^2(2b), \\ \beta &= b-\arg[\sin(b)-\cos(2b)-\sqrt{-1}(\cos(b)+\tfrac{1}{2}\sin(2b))]. \end{aligned} \tag{17}$$

As $\beta\in[0, \pi/2]$, inspection of (16) shows that no double zeros can occur in the case $m>0$. Hence the value of m is confined to $[-2, 0]$. The bounds on β also imply that the parameter b in (17) has to be in $[\pi/2, 3\pi/2]$. It is elementary to discuss the behavior of (m, β) as a function of b. It is better seen by looking at Fig. 8. It can just be said that two curves of double zeros appear with $\beta\in[\pi/4, \pi/2]$. They meet at $m=-\sqrt{2}$, $\beta=\pi/4$, where a triple zero appears. Between both curves there are exactly four zeros.

When additional powers of ε are included, a routine application of the Implicit Function Theorem permits us to conclude the same behavior for the full equation (14). It should be noted that $\beta=\pi/2$ has to be excluded from the previous analysis: in that case the axis of rotation is also the axis of the twist.

The preceding analysis can be summarized as

Proposition 7.1. For ε small enough and any $g\in SO(3)$ there are always at least two fixed points of $g\circ f_\varepsilon$. Bifurcations to exactly four fixed point appear for any longitude of the rotation axis and for latitude of the axis and angle of rotation $(\beta, \theta=\pi(2-m\varepsilon+O(\varepsilon^2)))$ along two lines described by formulae (17) when the parameter b ranges in $(\pi/2, \pi)$ and $(\pi, 3\pi/2)$, respectively.

To discuss bifurcations of the fixed points for general values of ε is an elementary but cumbersome task. As an illustration the case of bifurcations appearing on $\beta=0$ is presented. Then (14) reduces to

$$\cos\left(\frac{\theta}{2}\right)\cos(b)\sin(\delta(b))-\sin\left(\frac{\theta}{2}\right)\sin(b)=0 \tag{18}$$

and the condition for a double root becomes

$$\frac{2}{\pi\varepsilon}\tan\left(\frac{\pi\varepsilon}{2}(1+\sin(b))\right)=\sin(b)\cos^2(b). \tag{19}$$

The degenerate cases $|b| = \pi/2$ must be excluded in (19). It is immediate that new double fixed points appear on S^2 with $\beta = 0$ if and only if ε is a positive integer. The number of double fixed points with $\beta = 0$ (and some θ) increases with ε. Also from (18) it follows that new zeros appear near $\theta = 0$, one for $\theta > 0$ and the other for $\theta < 0$. These zeros move towards $\theta = \pi$ without ever reaching it. So, it is a simple matter to state how many fixed points exist for $\beta = 0$ (except at the bifurcation values of ε): there are at most $2(1 + E(\varepsilon))$, where E denotes the integer part of ε. For a given non-integer ε there are always values of θ such that this number $2(1 + E(\varepsilon))$ is the exact number of fixed points. This has an elementary dynamical interpretation: new fixed points emanate from the north pole of S^2 when the rotation number of f_ε at the north pole (defined by continuity) passes through 0 (mod 1).

To study the stability of the fixed points we recall that they are generically elliptic (eigenvalues μ in $S^1 \setminus \{\pm 1\}$), hyperbolic (real positive eigenvalues) and hyperbolic with reflection (real negative eigenvalues). Let E, H, R denote the number of fixed points of each type. Euler–Poincaré formula gives $E - H + R = 2$ (for simple fixed points). At the creation of new fixed points E and H increase by 1. When double eigenvalues are equal to -1 then, generically, E decreases by 1 and R increases by 1.

An analytic discussion on the stability of the fixed points is elementary (at least for small ε) but cumbersome. It is worth mentioning that, for any ε the maximum eigenvalue at a fixed point is achieved on $b = 0$ and has the expression

$$\mu_{\max} = \frac{\pi\varepsilon}{2} + \left(1 + \left(\frac{\pi\varepsilon}{2}\right)^2\right)^{1/2}. \tag{20}$$

A sample of illustrations is shown in Fig. 8 having $\theta/2\pi$ as horizontal variable and β/π as vertical one. A region containing i (resp. j, k) fixed points of type E (resp. H, R) is denoted as $R^k H^j E^i$. On the top left plot and for $\varepsilon = 0.1$ the two curves on the upper part of the plot are the curves of double zeros given by (17). Only in the region bounded by them there are 4 fixed points; the code is $H^1 E^3$. The codes for the black, dark grey, and light grey regions are, respectively, R^2, E^2, and $R^1 E^1$. On the top right plot, the value of ε is ≈ 3.456789. The region containing the point (0.5, 0) has exactly 2 fixed points while the regions which contact with this one through arcs have 4. The darker region has 6 and the small region near the upper right corner has 10. The regions around $\theta = 0$ have 8 fixed points. The solid lines give the location of all bifurcations and changes of stability.

In the bottom left plot, computed for $\varepsilon \approx 9.876543$, all the lines of bifurcation or change of stability are plotted. The number of fixed points,

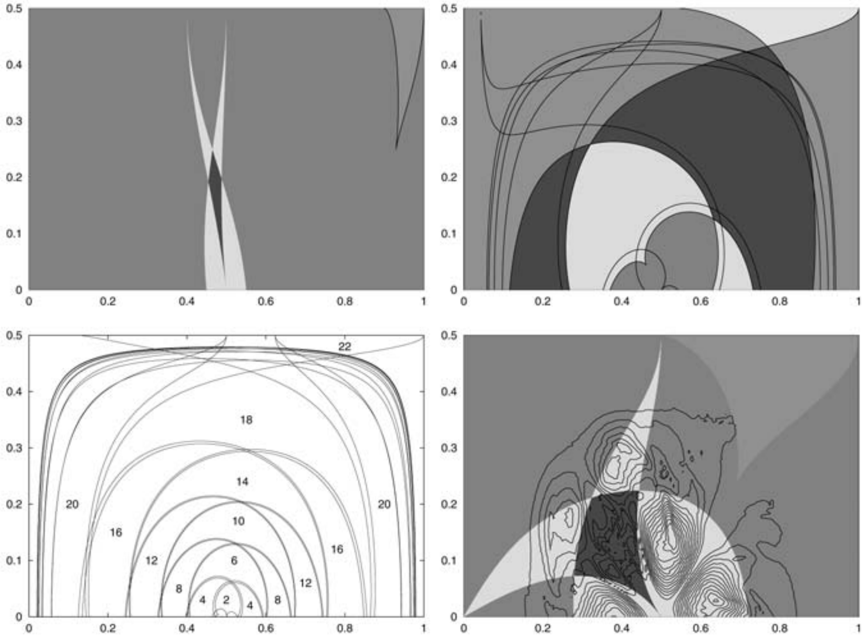

Fig. 8. Illustrations on the number of fixed points and their stability for different values of ε. On the horizontal (resp. the vertical) axis the variable $\theta/2\pi$ (resp. β/π) is displayed. Top left: The different regions correspond to different number of fixed points or to stability changes for $\varepsilon = 0.1$. Top right: Plot for $\varepsilon \approx 3.456789$. Different grey levels correspond to different number of fixed points. The lines are drawn where stability changes occur. Bottom left: For $\varepsilon \approx 9.876543$ the lines of stability changes are show. The numbers refer to how many fixed points exist in each region. Bottom right: Plot for $\varepsilon = 0.5$ where the level lines from Fig. 6 are also shown. See the text for more details.

NFP, in the major regions is shown. The typical transitions are as follow: Consider, for instance, a passage from NFP $= 16$ to NFP $= 18$ near $\beta = 0$ with increasing θ. First a line of creation of an elliptic and hyperbolic points is found. One passes from a code R^9H^7 to $R^9H^8E^1$. This is followed by a change of stability by passing to $R^{10}H^8$. Later on, inside the region with NFP $= 20$, the points of R type become again of type E. So the code passes, in the different changes, from $R^{11}H^9$ to H^9E^{11}.

Finally, in the bottom right plot, for $\varepsilon = 0.5$, regions similar to the case $\varepsilon = 0.1$ can be seen, with a different configuration. The level lines of Fig. 6 are also displayed. It is checked that the highest levels correspond to domains where the map has exactly one elliptic and one hyperbolic with reflection fixed points. This fact is also present for smaller values of ε.

7.2. The Maps as Perturbed Twists

The object of interest is the global dynamics of $g \circ f_\varepsilon$ in S^2. To this end it is convenient to write these maps in a slightly different, but equivalent, way. For this study S^2 will be taken as the sphere of radius 1 centered at the origin. Instead of considering f_ε as a twist around the z-axis of angle $\pi\varepsilon(1+z)$, it will be taken as a twist around the axis of zero longitude and latitude β with angle of rotation around this axis of a point of coordinates (x, y, z) equal to $\pi\varepsilon(1+x\cos(\beta)+z\sin(\beta))$. Then, the rotation g is simply a rotation of angle θ around the z-axis, to be denoted by $R_\theta^{(z)}$. Up to the substitution of $\pi/2-\beta$ for β, the relative positions of the axes in this formulation and in the previous one are equivalent.

It is instructive to first consider the case $\theta = 2\pi\frac{p}{q}$, where p, q are coprime integers. Let us introduce $\delta = \pi(1+x\cos(\beta)+z\sin(\beta))$. Then

$$f_\varepsilon\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \varepsilon\delta\begin{pmatrix} -y\sin(\beta) \\ x\sin(\beta)-z\cos(\beta) \\ y\cos(\beta) \end{pmatrix} + O(\varepsilon^2). \tag{21}$$

The next step is the computation of the map $M_{q,\theta,\beta,\varepsilon} := (R_\theta^{(z)} \circ f_\varepsilon)^q$, the parameter β being the latitude of the axis of f_ε. It is clear that at order zero in ε one has $M_{q,\theta,\beta,0} = \mathrm{Id}$. An elementary computation using formula (21) for f_ε and the expression of δ as a function of x, z, and β, gives

$$M_{q,\theta,\beta,\varepsilon} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} R_\gamma^{(z)} \begin{pmatrix} x \\ y \\ z \end{pmatrix} + O(\varepsilon^2), \tag{22}$$

where $R_\gamma^{(z)}$ is now a rotation around the z-axis in each one of the horizontal planes with angle of rotation depending on z as follows

$$\gamma = \pi\varepsilon q(\sin(\beta)+zP_2(\sin(\beta))), \tag{23}$$

where P_2 denotes the second Legendre polynomial ($P_2(w) = \frac{3}{2}w^2 - \frac{1}{2}$).

This result tells us that the rotation $R_\theta^{(z)}$ averages the effect of the map f_ε in a good way. Let us remark that in (22) it is assumed p/q fixed and ε sufficiently small. From (23) it follows that the angle γ is still small provided that εq is small. In the trivial case $p=0$, $q=1$ the rotation is the identity and then the twist can be be taken with $\beta=\pi/2$, recovering in formula (23) the angle rotated in the twist.

To pass to the general case for θ one needs a preliminary lemma.

Lemma 7.2. Let $\rho \in (0, 1)$ and $N \in \mathbf{N}$. Let

$$S_{\rho, N} = \bigcup_{1 \leqslant q \leqslant N, 0 \leqslant p \leqslant q, (p, q) = 1} \left[\frac{p-\rho}{q}, \frac{p+\rho}{q} \right].$$

Then, if $N+1 \geqslant \rho^{-1}$ one has $[0, 1] \subset S_{\rho, N}$.

Proof. Let $\alpha \in [0, 1]$. If $\alpha = r/s \in \mathbf{Q}$ with $(r, s) = 1$ and $s \leqslant N$, there is nothing to prove. Otherwise consider the approximants to α given by the continued fraction algorithm. Assume that $\frac{p_1}{q_1}$ and $\frac{p_2}{q_2}$ are consecutive approximants with $q_1 \leqslant N$ and $q_2 > N$. Then

$$\left| \alpha - \frac{p_1}{q_1} \right| \leqslant \frac{1}{q_1 q_2} \leqslant \frac{1}{q_1(N+1)} \leqslant \frac{\rho}{q_1}. \quad \blacksquare$$

Hence $\frac{\theta}{2\pi}$ can be written as $\frac{p}{q} + \frac{\mu}{q}$, where $|\mu| \leqslant \rho$, $q+1 \leqslant \rho^{-1}$, where ρ is not specified for the moment. Hence one can represent the map $g \circ f_\varepsilon$ as something similar to the previous case, that is, a rotation whose angle is a rational multiple of 2π, composed with a map close to the identity, by writing

$$g \circ f_\varepsilon = R^{(z)}_{2\pi p/q} \circ R^{(z)}_{2\pi\mu/q} f_\varepsilon.$$

By a direct computation one obtains expression (22) again with the following modifications:

- If the rational which approximates $\frac{\theta}{2\pi}$ is 0, then $q = 1$ and there is no average, so that we keep the map $g \circ f_\varepsilon$,
- The rotation is now $\gamma = \pi\varepsilon q(\sin(\beta) + zP_2(\sin(\beta))) + 2\pi\mu$,
- The error terms are, uniformly in θ, of the form $O(\varepsilon^2 q + \frac{\rho^2}{q})$.

If one takes $\rho = \varepsilon^{2/3}$ then the maps $(g \circ f_\varepsilon)^q$ are, in all cases, $\varepsilon^{1/3}$-close to the identity and the error terms are at most $\varepsilon^{4/3}$. Note that besides the choice $\rho = \varepsilon^{2/3}$ there are other possibilities, but $\rho = \varepsilon^{2/3}$ is good enough to prove Corollary 7.7. Finally, it is clear that (23), or the modification just mentioned adding $2\pi\mu$, is a twist except for $\beta = \beta_{\text{crit}}$ such that $P_2(\sin(\beta_{\text{crit}})) = 0$ ($\beta_{\text{crit}} = \sin^{-1}(1/\sqrt{3})$).

To summarize we state the following

Proposition 7.3. If θ is not $\varepsilon^{2/3}$-close to zero and $\beta \neq \beta_{\text{crit}}$ the maps $g \circ f_\varepsilon$ for ε small enough, have a power which is $\varepsilon^{1/3}$-close to the identity. This power satisfies a twist condition of order at least $\varepsilon P_2(\sin(\beta))$.

Remark 7.4. Rotations g with small θ are irrelevant, for the present purpose, due to the Haar measure in $SO(3)$, and a single exceptional case (non-twist) is also unimportant. This will be seen later in detail.

Remark 7.5. In the case of small θ it is still possible to show that $g \circ f_\varepsilon$ produces a twist effect on each meridian in S^2. A problem which appears though is that the angle rotated by the different points can pass through an extremum, losing in this way the twist property. In fact this is not so important because the existence of invariant curves when the twist condition is lost at some point has been established in ref. 23. But this refinement is not necessary in the present context.

Theorem 7.6. With the possible exclusion of an open set $\mathscr{B}$ in $SO(3)$ of small measure, there exist ε_0 such that for $\varepsilon < \varepsilon_0$ the maps $g \circ f_\varepsilon$ have a dynamics exponentially close to an integrable flow in S^2.

Proof. The proof is divided into steps.

1. The maps $M_{q,\theta,\beta,\varepsilon}$, being a power of $g \circ f_\varepsilon$, have the same dynamics as $g \circ f_\varepsilon$. In all cases (including $q = 1$ and the exceptional value of β) they are $\varepsilon^{1/3}$-close to the identity. Hence there exists a suspension given by the flow of a 1-periodic vector field in S^2 such that the time-1 map associated to this flow coincides with $M_{q,\theta,\beta,\varepsilon}$. The vector field is "slow" (of the order of $\varepsilon^{1/3}$) and the dominant terms do not depend on time. See ref. 24 for details and an explicit construction.

It is relevant to note that the vector field is analytic with respect to the phase space variables (that is, the points in S^2) while the dependence in ε is discontinuous in $SO(3)$ (moving $\theta \in [0, 2\pi]$ changes the value of q), but the relevant thing is that it is *bounded* in ε. Furthermore the dependence with respect to t can be made of class $\mathscr{C}^r$ for any $r > 0$, but continuity in t is sufficient for what follows. Furthermore the vector field is Hamiltonian.

2. The next step is to "average" the vector field with respect to t. This is the content of Neishtadt's theorem.[22] See ref. 25 for a detailed proof. As a consequence the vector field can be written as an autonomous part and a remainder which is exponentially small in the current small parameter; that is, the remainder is bounded by $\exp(-c\varepsilon^{-1/3})$ for some $c > 0$. Furthermore, the averaged vector field is still Hamiltonian (see ref. 26 for a sketch of the proof).

3. As the averaged system is a Hamiltonian in S^2, it is integrable and, hence, foliated by invariant curves except on the separatrices, which are a set of zero measure. Most of the invariant curves subsist as a consequence

of Moser's twist theorem. To this end one should have that the perturbation is small compared with the twist condition. Hence, it is enough to exclude a neighborhood of the critical latitude β_{crit} which can be taken also exponentially small. Furthermore the set of points in S^2 not covered by invariant curves of the full system has a measure bounded by the square root of the perturbation, again exponentially small in ε.

Summarizing, when arbitrary $g \in SO(3) \setminus \mathscr{B}$ are considered the dynamics in S^2 is ordered (the points lie on invariant curves) except for points in a subset of S^2 of exponentially small measure. Furthermore $\mathscr{B}$ consists of a neighborhood of the identity of size $O(\varepsilon^{2/3})$ and a neighborhood of β_{crit} which is exponentially small in ε. ∎

Corollary 7.7. For ε small enough $\Lambda(\varepsilon) < A\varepsilon^3$ for some fixed $A > 0$.

Proof. It is sufficient to make Remark 7.4 more explicit. The differentials of the maps f_ε increase the length of the vectors in TS^2 by a factor of the form $1 + O(\varepsilon)$ and composing with $g \in SO(3)$ produces no essential changes in the factor. Hence the values of $\lambda_1(x, g \circ f_\varepsilon)$ are bounded by $C\varepsilon$, where C is a positive constant. The contribution to $\Lambda(\varepsilon)$ of the g to which Theorem 7.6 applies is bounded by $C\varepsilon$ times an exponentially small amount. On the other hand the contribution of the excluded neighborhood of β_{crit} is also exponentially small.

Therefore the main contribution to $\Lambda(\varepsilon)$ can only come from the neighborhood of the identity excluded in Theorem 7.6. But the Haar measure of this set is of the order of

$$\int_0^{\varepsilon^{2/3}} (1 - \cos(\theta))\, d\theta = O(\varepsilon^2).$$

This bound and the previous one on $\lambda_1(x, g \circ f_\varepsilon)$ give the result. ∎

If we want to consider more "realistic" upper bounds it is possible to proceed along the ideas in Remark 7.5. A further consideration is that the largest stochastic zones are typically associated to the splitting of separatrices of the hyperbolic fixed points. From ref. 21 it follows that the splitting can be bounded by

$$\exp\left(-\frac{c}{\log(\mu_{\max})}\right),$$

where $\mu_{\max}$ is the maximal eigenvalue at the fixed points and $c > 0$. From (20) one has that for ε small $\log(\mu_{\max}) = \frac{\pi\varepsilon}{2} + O(\varepsilon^2)$. This "heuristic" prediction is in good agreement with the observed behavior for ε small.

8. THE CASE OF LARGE ε

Let us call the subbundle of PS^2 tangent to the invariant circles of f_ε the *horizontal bundle* and denote it by H. As $\varepsilon \to \infty$, a large portion of PS^2 is sucked into a small neighborhood of H under $f_{\varepsilon\#}$. The measure $\delta_{g(H)}$ on PS^2 that is atomic in each tangent space and supported on $g(H)$ looks more and more like an invariant measure for $gf_{\varepsilon,\#}$. These measures integrate to give Lebesgue:

$$m = \int_{g \in SO(3)} \delta_{g(H)}\, d\nu(g).$$

This yields a heuristic argument for why the inequality in Question 1.1 should hold when $\varepsilon = \infty$. In this section, we make this argument rigorous by adding some δ-noise in $\mathscr{F}_\varepsilon$, and replacing invariant measures with stationary measures. We prove:

Theorem 8.1. Let $m_{\varepsilon,\delta}$ be defined as in Section 3, and let $\varphi_{\varepsilon,\delta}$ be the density of $m_{\varepsilon,\delta}$:

$$dm_{\varepsilon,\delta} = \varphi_{\varepsilon,\delta}\, dm.$$

There exists $C > 0$ such that, for all $\varepsilon, \delta > 0$,

$$\|\varphi_{\varepsilon,\delta} - 1\|_1 < C\delta^{-11}\varepsilon^{-1/2}$$

where $\|\cdot\|_1$ is the L^1-norm with respect to Lebesgue measure m on PS^2.

This has the corollary:

Corollary 8.2. There exists a $C > 0$ such that for all $\varepsilon, \delta > 0$,

$$|R(\varepsilon,\delta) - R(\varepsilon)| \leqslant C\delta^{-11}\varepsilon^{-1/2}\log \varepsilon.$$

In particular, for all $\delta > 0$,

$$\lim_{\varepsilon \to \infty} |R(\varepsilon,\delta) - R(\varepsilon)| = 0,$$

where $R(\varepsilon,\delta)$ is the random diffused exponent defined in Section 3.

As we were finishing this paper, unpublished work of L. Carleson and T. Spencer came to our attention.[5] For the standard map on the 2-torus:

$$g_\varepsilon \colon (x, y) \mapsto (2x + \varepsilon \sin(2\pi x) - y, x)$$

where ε measures the strength of the nonlinearity, they prove that by adding a noise of strength $\exp(-\varepsilon^2)$ to the element g_ε, a Lyapunov exponent of order $\log \varepsilon$ can be established.

Proof of Corollary 8.2. From the definitions,

$$R(\varepsilon, \delta) = \int_{PS^2} \log \|Tf_\varepsilon v\| \, dm_{\varepsilon,\delta}(v)$$

$$= \int_{PS^2} \log \|Tf_\varepsilon v\| \, \varphi_{\varepsilon,\delta}(v) \, dm(v),$$

whereas $R(\varepsilon)$ is the integral of $\log \|Tf_\varepsilon v\|$ with respect to m. Hence,

$$|R(\varepsilon, \delta) - R(\varepsilon)| = \left| \int_{PS^2} \log \|Tf_\varepsilon v\| \, (\varphi_{\varepsilon,\delta}(v) - 1) \, dm(v) \right|$$

$$\leqslant \|\log Tf_\varepsilon\|_\infty \, \|\varphi_{\varepsilon,\delta} - 1\|_1$$

$$\leqslant C\delta^{-11}\varepsilon^{-1/2} \log \varepsilon,$$

by Theorem 8.1. ∎

Proof of Theorem 8.1. $SO(3)$ acts transitively on T_1S^2 by isometries and with trivial stabilizer. From now on we identify points in T_1S^2 with elements of $SO(3)$, and use the group structure in writing our formulas. We will use x, y, z to denote elements of PS^2, p, q for elements of S^2, and (p, v), (q, w) for elements of PS^2 (or T_1S^2).

Recall from Section 3 that

$$\varphi_{\varepsilon,\delta} = \int_{PS^2} \varphi_{\varepsilon,\delta,g} \, d\nu(g),$$

where $\varphi_{\varepsilon,\delta,g}$ is the fixed point of the operator $L_{\varepsilon,\delta,g}$ defined by:

$$L_{\varepsilon,\delta,g}\psi(x) = \frac{1}{m(B(x,\delta))} \int_{(gf_{\varepsilon\#})^{-1} B(x,\delta)} \psi(y) \, dm(y).$$

Setting

$$k_\delta(x) = \frac{1}{m(B(e,\delta))} 1_{B(e,\delta)}(x),$$

where e is the identity element of $SO(3)$, we rewrite $L_{\varepsilon,\delta,g}$ as:

$$L_{\varepsilon,\delta,g}\psi(x) = \int_{PS^2} k_\delta(x^{-1}gf_{\varepsilon\#}(y))\,\psi(y)\,dm(y). \tag{24}$$

Let $L_{\varepsilon,\delta} = L_{\varepsilon,\delta,e}$. It is clear from (24) that $L_{\varepsilon,\delta,g}\psi(x) = L_{\varepsilon,\delta}\psi(g^{-1}x)$.

Denote by $K_{\varepsilon,\delta}\colon PS^2 \times PS^2 \to \mathbf{R}_+$ the kernel of the operator $L_{\varepsilon,\delta}$, so that

$$K_{\varepsilon,\delta}(x,y) = k_\delta(x^{-1}f_{\varepsilon\#}(y)).$$

Let $\pi\colon PS^2 \to S^2$ be the projection along tangent fibers. By averaging along fibers, we shall approximate $K_{\varepsilon,\delta}$ by a new kernel $\hat{K}_{\varepsilon,\delta}$ that is constant along fibers of the second PS^2-factor. Define $\hat{K}_{\varepsilon,\delta}\colon PS^2 \times PS^2 \to \mathbf{R}_+$ by

$$\hat{K}_{\varepsilon,\delta}(x,y) = \int_{\pi^{-1}\pi y} K_{\varepsilon,\delta}(x,z)\,dm_{\pi y}(z),$$

where, for $p \in S^2$, m_p denotes the disintegration of m along the fiber $\pi^{-1}p$. For $g \in SO(3)$, we obtain a new operator $\hat{L}_{\varepsilon,\delta,g}$ on $L^\infty(PS^2)$, given by:

$$\hat{L}_{\varepsilon,\delta,g}\phi(x) = \int_{PS^2} \hat{K}_{\varepsilon,\delta}(g^{-1}x,y)\,\phi(y)\,dm(y).$$

Let $\hat{L}_{\varepsilon,\delta} = \hat{L}_{\varepsilon,\delta,e}$.

The next lemma shows that the operators $\hat{L}_{\varepsilon,\delta,g}$ have a good averaging property when applied to densities of measures that project to Lebesgue measure μ on S^2.

Lemma 8.3. Let $\hat{K}\colon PS^2 \times PS^2 \to \mathbf{R}_+$ be any L^1 function such that:

1. for all $p \in S^2$,

$$\int_{\pi^{-1}p}\int_{PS^2} \hat{K}(x,y)\,dm_p(x)\,dm(y) = 1;$$

2. if $\pi(y) = \pi(z)$, then for all x, $\hat{K}(x,y) = \hat{K}(x,z)$.

For $\phi \in L^\infty(PS^2)$, and $g \in SO(3)$, define $\hat{L}_g\phi \in L^\infty(PS^2)$ by

$$\hat{L}_g\phi(x) = \int_{PS^2} \hat{K}(g^{-1}x,y)\,\phi(y)\,dm(y).$$

Then:

(a) $\hat{\varphi}_g = \hat{L}_g \varphi$ is the unique fixed point of $\hat{L}_g$, for any $\varphi \in L^\infty(PS^2)$ that satisfies:

$$\int_{\pi^{-1}p} \varphi(z)\, dm_p(z) = 1,$$

for all $p \in S^2$.

(b) for all $x \in PS^2$, we have:

$$\int_{SO(3)} \hat{\varphi}_g(x)\, dv(g) = 1.$$

Remark 8.4. Lemma 8.3 applies to the operators $\hat{L}_{\varepsilon, \delta, g}$. An example of a function φ that satisfies the hypotheses of Lemma 8.3 is the density $\varphi_{\varepsilon, \delta, g}$, for any ε, δ, g. We do not use that $\hat{\varphi}_g = \hat{L}_g \hat{\varphi}_g$ below.

Proof of Lemma 8.3. Since it is constant along fibers of the second factor, $\hat{K}$ projects to a function on $PS^2 \times S^2$, which we shall also call $\hat{K}$.

For $g \in SO(3)$, define $\hat{\varphi}_g$ by:

$$\hat{\varphi}_g(x) = \hat{L}_g 1(x) = \int_{p \in S^2} \hat{K}(g^{-1}x, p)\, d\mu(p).$$

We compute directly that, for any φ satisfying the hypotheses of (a),

$$\begin{aligned}
\hat{L}_g(\varphi)(x) &= \int_{y \in PS^2} \hat{K}(g^{-1}x, y)\, \varphi(y)\, dm(y) \\
&= \int_{p \in S^2} \int_{z \in \pi^{-1}p} \hat{K}(g^{-1}x, p)\, \varphi(z)\, dm_p(z)\, d\mu(p) \\
&= \int_{p \in S^2} K(g^{-1}x, p)\, d\mu(p) \\
&= \hat{\varphi}_g(x).
\end{aligned}$$

To see that $\hat{L}_g \hat{\varphi}_g = \hat{\varphi}_g$ and finish the proof of (a) it is now sufficient to verify that

$$\int_{\pi^{-1}p} \hat{\varphi}_g(z)\, dm_p(z) = 1.$$

But $\hat{\varphi}_g(z) = \hat{L}_g 1(z) = L_g 1(z)$ since the function 1 is constant and hence constant on fibers. Now $L_g(1)$ is the density function of a measure on PS^2 which covers Lebesgue measure on S^2 and hence its integral on fibers equals 1. This proves (a).

Integrating $\hat{\varphi}_g(x)$ with respect to g we obtain:

$$\begin{aligned}\int_{g \in SO(3)} \hat{\varphi}_g(x)\, d\nu(g) &= \int_{g \in SO(3)} \int_{p \in S^2} \hat{K}(g^{-1}x, p)\, d\mu(p)\, d\nu(g)\\ &= \int_{y \in PS^2} \int_{p \in S^2} \hat{K}(y, p)\, d\mu(p)\, dm(y)\\ &= 1,\end{aligned}$$

completing the proof of (b). ▮

Returning to the proof of Theorem 8.1, let $\hat{\varphi}_{\varepsilon,\delta,g} = \hat{L}_{\varepsilon,\delta,g} 1 = \hat{L}_{\varepsilon,\delta,g}\varphi_{\varepsilon,\delta,g}$ be the unique fixed point of $\hat{L}_{\varepsilon,\delta,g}$ given by Lemma 8.3. We now have:

$$\begin{aligned}\|\varphi_{\varepsilon,\delta} - 1\|_1 &= \left\| \int_{g \in SO(3)} (\varphi_{\varepsilon,\delta,g} - 1)\, d\nu(g) \right\|_1\\ &= \left\| \int_{g \in SO(3)} (\varphi_{\varepsilon,\delta,g} - \hat{\varphi}_{\varepsilon,\delta,g})\, d\nu(g) \right\|_1 + \left\| \int_{g \in SO(3)} (\hat{\varphi}_{\varepsilon,\delta,g} - 1)\, d\nu(g) \right\|_1\\ &\leqslant \left\| \int_{g \in SO(3)} (L_{\varepsilon,\delta,g}\varphi_{\varepsilon,\delta,g} - \hat{L}_{\varepsilon,\delta,g}\varphi_{\varepsilon,\delta,g})\, d\nu(g) \right\|_1\\ &\quad + \left\| \int_{g \in SO(3)} (\hat{\varphi}_{\varepsilon,\delta,g} - 1)\, d\nu(g) \right\|_1\\ &= \left\| \int_{g \in SO(3)} (L_{\varepsilon,\delta}\varphi_{\varepsilon,\delta,g} \circ g^{-1} - \hat{L}_{\varepsilon,\delta}\varphi_{\varepsilon,\delta,g} \circ g^{-1})\, d\nu(g) \right\|_1\\ &\leqslant \int_{g \in SO(3)} \|L_{\varepsilon,\delta}\varphi_{\varepsilon,\delta,g} \circ g^{-1} - \hat{L}_{\varepsilon,\delta}\varphi_{\varepsilon,\delta,g} \circ g^{-1}\|_1\, d\nu(g)\end{aligned}$$

where we used Lemma 8.3 to obtain the second to last inequality.

Propositions 8.5 and 8.6, which we state and prove below, now imply that

$$\begin{aligned}
&\int_{g \in SO(3)} \|L_{\varepsilon,\delta}\varphi_{\varepsilon,\delta,g} \circ g^{-1} - \hat{L}_{\varepsilon,\delta}\varphi_{\varepsilon,\delta,g} \circ g^{-1}\|_1 \, d\nu(g) \\
&\qquad \leqslant C_1\delta^{-3}\varepsilon^{-1/2} \int_{g \in SO(3)} \|\varphi_{\varepsilon,\delta,g}\|_\infty \, d\nu(g) \\
&\qquad \leqslant C_2\delta^{-3}\varepsilon^{-1/2} \int_{g \in SO(3)} \delta^{-8} \, d\nu(g) \\
&\qquad = C\delta^{-11}\varepsilon^{-1/2},
\end{aligned}$$

completing the proof of the theorem. It remains to state and prove Propositions 8.5 and 8.6.

Proposition 8.5. There is a $C > 0$ such that, if $\varepsilon \geqslant \delta^{-4}$, then for any $\phi \in L^\infty(PS^2)$,

$$\|L_{\varepsilon,\delta}(\phi) - \hat{L}_{\varepsilon,\delta}(\phi)\|_1 \leqslant C \, \|\phi\|_\infty \, \delta^{-3}\varepsilon^{-1/2}.$$

Proposition 8.6. There exists $C > 0$ such that, for all $\varepsilon > 0$ and $\delta > 0$,

$$\|\varphi_{\epsilon,\delta,g}\|_\infty \leqslant C\delta^{-8}.$$

Proof of Proposition 8.5.

$$\begin{aligned}
&\|L_{\varepsilon,\delta}(\phi) - \hat{L}_{\varepsilon,\delta}(\phi)\|_1 \\
&\qquad \leqslant \|\phi\|_\infty \, \|K_{\varepsilon,\delta} - \hat{K}_{\varepsilon,\delta}\|_1 \\
&\qquad = \|\phi\|_\infty \int |K_{\varepsilon,\delta}(x,y) - \hat{K}_{\varepsilon,\delta}(x,y)| \, dm(x) \, dm(y) \\
&\qquad = \|\phi\|_\infty \int \left| K_{\varepsilon,\delta}(x,y) - \int_{\pi^{-1}\pi y} K_{\varepsilon,\delta}(x,z) \, dm_{\pi y}(z) \right| dm(x) \, dm(y).
\end{aligned}$$

Let $x = (p, v)$, $y = (q, w)$ be elements of PS^2. For a fixed p, v, q, the map $w \mapsto K_{\varepsilon,\delta}((p, v), (q, w))$ is a constant multiple $c\delta^{-3}$ of the characteristic function for $\pi^{-1}q \cap f_{\varepsilon\#}^{-1}B((p, v), \delta)$. Note that, for any measurable set B in a probability space with measure m, the average value of the function $|m(B) - 1_B|$ is $2m(B)(1 - m(B))$. It follows that:

$$\begin{aligned}
&\|K_{\varepsilon,\delta} - \hat{K}_{\varepsilon,\delta}\|_1 \\
&\quad = 2c\delta^{-3} \int_{(p,v) \in PS^2} \int_{q \in f_\varepsilon^{-1}B(p,\delta)} |\beta(p, v, q)(1 - \beta(p, v, q))| \, dm((p, v)) \, d\mu(q),
\end{aligned}$$

where

$$\beta(p, v, q) = m_q(f_{\varepsilon\#}^{-1} B((p, v), \delta)).$$

We next show that for some $k > 0$, and for $\varepsilon > \delta^{-4}$, there is a set $G \subset PS^2$, with $m(G) \geqslant 1 - \varepsilon^{-1/2}$ such that, for all $(p, v) \in G$, there is a set $G' = G'(p, v) \subset f_\varepsilon^{-1} B(p, \delta)$ with $\mu(G') \geqslant \mu(B(p, \delta)) - \varepsilon^{-1/2}$, such that, for $q \in G'$,

$$\beta(p, v, q) \leqslant k\varepsilon^{-1/2} \qquad \text{or} \qquad \beta(p, v, q) \geqslant 1 - k\varepsilon^{-1/2}. \tag{25}$$

This implies that

$$\begin{aligned} \|K_{\varepsilon,\delta} - \hat{K}_{\varepsilon,\delta}\|_1 &\leqslant 2c\delta^{-3}(k\varepsilon^{-1/2}\mu(B(p, \delta)) + 2\varepsilon^{-1/2}) \\ &\leqslant C\delta^{-3}\varepsilon^{-1/2}, \end{aligned}$$

which implies the result.

Fix $\delta < 1/2$, and assume that $\varepsilon > \delta^{-4}$. For $\alpha > 0$, denote by C_α the α-neighborhood of the horizontal bundle $H \subset PS^2$. In other words, C_α is the set of lines in $PS^2 \backslash \pi^{-1}\{NP, SP\}$ at angle $\leqslant \alpha$ with the latitudinal circles.

It is not difficult to see that if the distance from p to the poles is greater than $\varepsilon^{-1/4}$, then

$$m_{f_\varepsilon^{-1}p}(f_{\varepsilon\#}^{-1}(C_{\varepsilon^{-1/2}} \cap \pi^{-1}p)) \geqslant 1 - \varepsilon^{-1/2}. \tag{26}$$

Let G_0 be the set of p for which (26) holds, and let $G = \pi^{-1}G_0$. Then $m(G) \geqslant 1 - \varepsilon^{-1/2}$.

Fix $(p, v) \in G$, and consider a point $f_\varepsilon q \in B(p, \delta) \cap G_0$. The intersection of $B((p, v), \delta)$ with the fiber $\pi^{-1} f_\varepsilon q$ is an interval I. If the endpoints of I are disjoint from the interval $J = C_{\varepsilon^{-1/2}} \cap \pi^{-1} f_\varepsilon q$, then either $I \supset J$ or $I \cap J = \varnothing$. In the former case, the length of $f_{\varepsilon\#}^{-1} I$ is greater than the length of $f_{\varepsilon\#}^{-1} J$, which by (26) is greater than $1 - \varepsilon^{-1/2}$. In the latter case, the length of $f_{\varepsilon\#}^{-1} I$ is less than $\varepsilon^{-1/2}$. Hence if we let $G' = G'(p, v)$ be the set of $q \in f_\varepsilon^{-1} B(p, \delta) \cap G_0$ satisfying:

$$\partial B((p, v), \delta) \cap C_{\varepsilon^{-1/2}} \cap \pi^{-1} f_\varepsilon q = \varnothing, \tag{27}$$

then (25) holds for all $q \in G'$. It remains to show that $\mu(G') \geqslant \mu(B(p, \delta)) - \varepsilon^{-1/2}$.

For $(p, v) \in PS^2$, denote by $S((p, v), \delta)$ the geodesic sphere of radius δ centered at (p, v), so $S((p, v), \delta) = \partial B((p, v), \delta)$. We will use the following lemma here and later in the proof of Proposition 8.6.

Lemma 8.7. There exists $C>0$ such that, for all $(p,v)\in\pi^{-1}p$, and all $\alpha>0$,

$$\mu(\pi(S((p,v),\delta)\cap C_\alpha))\leqslant C\delta\alpha.$$

Proof. The claim follows from the following facts:

1. On $T_1S^2\setminus\pi^{-1}\{NP,SP\}$, the subbundle H (regarded as a submanifold) is uniformly transverse to the fibers of PS^2.

2. There exists a $C>0$ such that for all δ sufficiently small, and for all $(p,v)\in PS^2$, the intersection $S((p,v),\delta)\cap H$ is contained in a smooth curve of length $\leqslant C\delta$.

The verification of these facts is left as an exercise. ∎

From Lemma 8.7 it follows that:

$$\begin{aligned}\mu(G')&\geqslant\mu(B(p,\delta))-\mu(G_0)-\mu(f_\varepsilon^{-1}\pi(S((p,v),\delta)\cap C_{\varepsilon^{-1/2}}))\\&\geqslant\mu(B(p,\delta))-\varepsilon^{-1/2}.\end{aligned}$$

This completes the proof of Proposition 8.5. ∎

Proof of Proposition 8.6. We know that $\varphi_{\epsilon,\delta,g}$ is a function in L^1 which satisfies, for all $(p,v)\in PS^2$,

$$\varphi_{\epsilon,\delta,g}(p,v)=c\delta^{-3}\int_{f_{\varepsilon\#}^{-1}B(g^{-1}(p,v),\delta)}d\varphi_{\epsilon,\delta,g}(q,w)\,d\mu(q)\,dm_q(w)\tag{28}$$

and, for all $p\in S^2$,

$$\int_{\pi^{-1}p}\varphi_{\epsilon,\delta,g}(p,v)\,dm_p(v)=1.\tag{29}$$

Then, by (28), the function $\varphi_{\epsilon,\delta,g}$ is continuous, and therefore has a maximum $M_{\epsilon,\delta,g}$ that we denote by M. The idea is that a Hölder constant for $\varphi_{\epsilon,\delta,g}$ can be estimated in terms of M. Reporting in (29) gives a bound for M which is independent of ε, g. Since we want to use (29) at the end, it suffices to consider the Hölder constant along the fibers. So, let $(p,v),(p,v')\in\pi^{-1}p$. We have:

$$\begin{aligned}&|\varphi_{\epsilon,\delta,g}(p,v)-\varphi_{\epsilon,\delta,g}(p,v')|\\&\qquad\leqslant cM\delta^{-3}m(f_{\varepsilon\#}^{-1}B(g^{-1}(p,v),\delta)\,\Delta f_{\varepsilon\#}^{-1}B(g^{-1}(p,v'),\delta)),\end{aligned}$$

where m is Lebesgue measure on $T_1 S^2$, and $A \Delta B$ stands for the set of points which belong to only one of the subsets A or B.

Lemma 8.8. There exists $C > 0$ such that, for all $v, v' \in T_{1,p}S^2$,

$$m(f_{\varepsilon\#}^{-1} B(g^{-1}(p, v), \delta) \,\Delta\, f_{\varepsilon\#}^{-1} B(g^{-1}(p, v'), \delta)) \leqslant C\delta d(v, v')^{1/4}.$$

Remark 8.9. In Proposition 3.3, we prove that $\varphi_{\epsilon, \delta, g}$ is as smooth as f_ε. The point of the arguments here is to get a Hölder constant independent of ε.

From this lemma, it then follows that $M \leqslant 5(Cc)^4 \delta^{-8}/2$, since:

$$\begin{aligned} 1 &= \int_{T_p S^2} \varphi_{\epsilon, \delta, g}(p, v)\, dv \\ &\geqslant M \int_{-\infty}^{\infty} (1 - Cc\delta^{-2} |t|^{1/4})^+ \, dt \\ &= \frac{2M}{5} (Cc)^{-4} \delta^8. \end{aligned}$$

We now prove Lemma 8.8.

We have the two balls $B((p, v), \delta))$ and $B((p, v'), \delta)$. Let $\alpha = \sqrt{d(v, v')}$. We may assume that $\alpha \ll \delta$. The set $B((p, v), \delta)) \,\Delta B((p, v'), \delta)$ meets the fiber $\pi^{-1} q$ in a pair of intervals, each of length $\leqslant \alpha^2 \ll \sqrt{\alpha}$. If the endpoints of these intervals do not lie in $C_{\sqrt{\alpha}}$, then the entire intervals must be disjoint from $C_{\sqrt{\alpha}}$. In other words, if

$$\pi^{-1} q \cap (S((p, v), \delta) \cup S((p, v'), \delta)) \cap C_{\sqrt{\alpha}} = \varnothing, \tag{30}$$

then

$$\pi^{-1} q \cap B((p, v), \delta)) \,\Delta B((p, v'), \delta) \cap C_{\sqrt{\alpha}} = \varnothing. \tag{31}$$

Let $G \subset S^2$ be the set of q satisfying (30). By Lemma 8.7, $\mu(G) \geqslant 1 - 2C\delta\sqrt{\alpha}$.

Claim 8.10. There exists a $C > 0$ such that, for all $\alpha \leqslant 1$, $p \in S^2$, and $\varepsilon \geqslant 0$, if $(p, v) \notin C_{\sqrt{\alpha}}$, then

$$\|T_{(p, v)} f_{\varepsilon,\#}^{-1}|_{T_{1,p}S^2}\| \leqslant C\alpha^{-1}.$$

Proof. Recall that $\|T_{(p,v)} f_{\varepsilon,\#}^{-1}|_{\pi^{-1}p}\| = \|T_p f_\varepsilon^{-1} v\|^{-2}$. With respect to the orthonormal basis of $T_p S^2$ of the form $\{e_1(p), e_2(p)\}$, where $e_1(p) \in H$ points in the direction of f_ε-twist and $e_2(p)$ points toward the north pole NP, $T_p f_\varepsilon$ takes the form:

$$T_p f_\varepsilon^{-1} = \begin{pmatrix} 1 & -\beta \\ 0 & 1 \end{pmatrix},$$

for some $\beta \geqslant 0$. A direct computation shows that there exists a constant $C > 0$ such that, for all $\alpha, \beta \geqslant 0$, if the angle between a unit vector $v \in \mathbf{R}^2$ and the x-axis is greater than $\sqrt{\alpha}$, then:

$$\left\| \begin{pmatrix} 1 & -\beta \\ 0 & 1 \end{pmatrix} v \right\|^{-2} \leqslant C\alpha^{-1}$$

From this the claim follows. ▮

Claim 8.10 and (31) imply that for $q \in G$, the derivative of $f_{\varepsilon\#}$ on $\pi^{-1}q \cap B((p,v),\delta)\, \Delta B((p,v'),\delta)$ is bounded:

$$\|T_{(q,w)} f_{\varepsilon\#}|_{T_{1,q}S^2}\| \leqslant \alpha^{-1},$$

for all w such that $(q,w) \in B((p,v),\delta))\, \Delta B((p,v'),\delta)$. But for $q \in G$,

$$\begin{aligned} m_{f_\varepsilon^{-1}q}(f_{\varepsilon\#}^{-1}(B((p,v),\delta))\, \Delta B((p,v'),\delta) &\leqslant \alpha^{-1} m_q(B((p,v),\delta))\, \Delta B((p,v'),\delta)) \\ &\leqslant \alpha^{-1} d(v,v') \\ &= d(v,v')^{1/2}. \end{aligned}$$

But then

$$\begin{aligned} m(B((p,v),\delta))\, \Delta B((p,v'),\delta) &\leqslant 2C_1\delta\alpha^{1/2} + C_2\delta^2 d(v,v')^{1/2} \\ &\leqslant C\delta d(v,v')^{1/4}, \end{aligned}$$

completing the proof of Proposition 8.6 and of Theorem 8.1. ▮

9. DISCUSSION

We have wondered[(2)] about the relationship of the random Lyapunov exponent of a measure on the space of volume preserving diffeomorphisms of a manifold to the mean of the Lyapunov exponents of the individual members. The point of the question we raised was to be able to conclude that in a rich enough family of diffeomorphisms there must be some with positive Lyapunov exponents, that is to say positive entropy. At question is what sort of notion of richness would make such a conclusion valid. We even proposed that much more might conceivably be true, a lower bound for the mean of the Lyapunov exponents in terms of the random exponents for orthogonally invariant measures on volume preserving diffeomorphisms of the sphere. The orthogonal invariance of the measure was to provide the necessary "richness."

In the studied family strong numerical evidence has been found about the existence of such a lower bound when the values of the stretching parameter ε are not too small. In some sense strong stretching has an effect similar to randomization, but it depends in a clear way on the concrete map. More concretely

- Even moderate values of ε like $\varepsilon \geqslant 10$ are enough to have an average of the metric entropy larger than the one corresponding to the random map.
- There exist unbounded parameters ε for which islands are born. The range of existence of these islands is small, but only the islands associated to fixed points have been considered.
- For small ε the estimated average entropy seems positive and definitely to be much less than the one of the random map. The numerical evidence is in favor of the existence of exponentially small lower and upper bounds (in the present example, with an analytic family).

The problems in numerically estimating exponents and how to overcome them have been discussed. A partial analysis of the family of maps has been done for ε small. Even a rough estimate of an upper bound of the averaged entropy is enough to show that the this averaged entropy falls below any constant multiple of the entropy of the randomized system, if ε is sufficiently small.

Finally, the effect of a small randomization of fixed size δ of the individual elements of the family $\mathscr{F}_\varepsilon$ is considered. Now the mean of the local random exponents of the family is indeed asymptotic to the random exponent of the entire family as ε tends to infinity; that is, $R(\varepsilon, \delta)$ and $R(\varepsilon)$ are asymptotic.

ACKNOWLEDGMENTS

We thank Roy Adler, Victor Klepstyn, Yuri Kifer, Marco Martens, Lai-Sang Young, and especially Tolya Katok for conversations about this work. The computing facilities of the Dynamical Systems Group of the University of Barcelona have been widely used. We thank the supporting institutions. Michael Shub was partly supported by NSF Grant #DMS-9988809. Carles Simó was partly supported by Grants DGICYT BFM2000-805, CIRIT 2001SGR-70, and INTAS00-221. Amie Wilkinson was partly supported by NSF Grant #DMS-0100314.

REFERENCES

1. A. Neishtadt, C. Simó, and A. Vasiliev, Geometric and statistical properties induced by separatrix crossings in volume-preserving systems, *Nonlinearity* **16**:521–557 (2003).
2. K. Burns, C. Pugh, M. Shub, and A. Wilkinson, Recent results about stable ergodicity, *Proc. Symposia A.M.S.* **69**:321–366 (2001).
3. A. Carverhill, Furstenberg's theorem for non-linear stochastic systems, *Probab. Theory Related Fields* **74**:529–534 (1987).
4. J. P. Dedieu and M. Shub, On random and mean exponents for unitarily invariant probability measures on $GL(n, \mathbf{C})$, to appear in *Astérisque*.
5. L. Carleson and T. Spencer, personal communication; and T. Spencer, *Standard Map Conjectures*, Einstein Chair Lecture at CUNY, videotape #329.
6. Y. Kifer, Ergodic theory of random transformations, in *Progress in Probability and Statistics*, Vol. 10 (Birkhäuser Boston, Boston, MA, 1986).
7. Y. Kifer, Random perturbations of dynamical systems, in *Progress in Probability and Statistics*, Vol. 16 (Birkhäuser Boston, Boston, MA, 1988).
8. Y. Kifer, Random dynamics and its applications, in *Proc. of Int. Congress of Math.*, Vol. II (Berlin, 1998). Doc. Math 1998, Extra Vol. II, 809–818 (electronic).
9. Pei-Dong Liu and Min Qian, Smooth ergodic theory of random dynamical systems, *Lecture Notes in Math.*, No. 1606 (Springer, 1995).
10. I. Ya. Gol'shied and G. A. Margulis, Lyapunov indices of a product of random matrices, *Russian Math. Surveys* **44**:11–71 (1989).
11. D. Ruelle, *Ergodic Theory of Differentiable Dynamical Systems*, Vol. 50 (Publications Mathématiques de l'IHES, 1979), pp. 27–58.
12. V. I. Oseledec, A multiplicative ergodic theorem. Lyapunov characteristic numbers for dynamical systems, *Trans. Moscow Math. Soc.* **19**:197–231 (1968).
13. A. Avila and J. Bochi, A formula with applications to the theory of Lyapunov exponents, *Israel J. Math*, to appear.
14. M. Do Carmo, *Riemannian Geometry* (Birkhauser, Boston, 1992).
15. A. Giorgilli, V. F. Lazutkin, and C. Simó, Visualization of a hyperbolic structure in area-preserving maps, *Regular & Chaotic Dynamics* **2**:47–61 (1997).
16. H. Broer and C. Simó, Hill's equation with quasi–periodic forcing: Resonance tongues, instability pockets, and global phenomena, *Bull. Soc. Bras. Mat.* **29**:253–293 (1998).
17. C. Simó and T. Stuchi, Central stable/unstable manifolds and the destruction of KAM tori in the planar Hill problem, *Physica D* **140**:1–32 (2000).
18. P. M. Cincotta and C. Simó, Simple tools to study global dynamics in non–axisymmetric galactic potentials, I, *Astronom. & Astrophys. Supp.* **147**:205–228 (2000).
19. P. M. Cincotta, C. M. Giordano, and C. Simó, Phase space structure of multidimensional systems by means of the mean exponential growth factor of nearby orbits (MEGNO), *Physica D*, in press.
20. C. Simó, Global dynamics and fast indicators, in *Global Analysis of Dynamical Systems*, H. W. Broer *et al.*, eds. (IOP Publishing, Bristol, 2001), pp. 373–390.

21. E. Fontich and C. Simó, The splitting of separatrices for analytic diffeomorphisms, *Ergodic Theory Dynam. Systems* **10**:295–318 (1990).
22. A. Neishtadt, The separation of motions in systems with rapidly rotating phase, *J. Appl. Math. Mech.* **48**:133–139 (1984).
23. C. Simó, Invariant curves of perturbations of non twist integrable area preserving maps, *Regular & Chaotic Dynamics* **3**:180–195 (1998).
24. H. Broer, R. Roussarie, and C. Simó, Invariant circles in the Bogdanov–Takens bifurcation for diffeomorphisms, *Ergodic Theory Dynam. Systems* **16**:1147–1172 (1996).
25. C. Simó, Averaging under fast quasi-periodic forcing, in *Integrable and Chaotic Behaviour in Hamiltonian Systems*, I. Seimenis, ed. (Plenum, New York, 1994), pp. 13–34.
26. C. Simó and C. Valls, A formal approximation of the splitting of separatrices in the classical Arnold's example of diffusion with two equal parameters, *Nonlinearity* **14**:1707–1760 (2001).
27. R. Abraham and J. Robbin, *Transversal Mappings and Flows* (Benjamin, New York/ Amsterdam, 1967).
28. R. Abraham and S. Smale, Nongenericity of Ω-stability, in *1970 Global Analysis, Proc. Sympos. Pure Math.*, Vol. 14 (Berkeley, CA, 1968), pp. 5–8.
29. D. V. Anosov, Geodesic flows on closed Riemannian manifolds of negative curvature, *Proc. Steklov. Inst. Math.* **90** (1967).