University of **Reading**

Asymptotic stability of the optimal filter for random chaotic maps

Article

Accepted Version

It is advisable to refer to the publisher's version if you intend to cite from the work.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

ASYMPTOTIC STABILITY OF THE OPTIMAL FILTER FOR RANDOM CHAOTIC MAPS

JOCHEN BRÖCKER AND GIANLUIGI DEL MAGNO

ABSTRACT. The asymptotic stability of the optimal filtering process in discrete time is revisited. The filtering process is the conditional probability of the state of a Markov process, called the signal process, given a series of observations. Asymptotic stability means that the distance between the true filtering process and a wrongly initialised filter converges to zero as time progresses. In the present setting, the signal process arises through iterating an i.i.d. sequence of uniformly expanding random maps. It is showed that for such a signal, the asymptotic stability is exponential provided that its initial conditions are sufficiently smooth. Similar to previous work on this problem, Hilbert's projective metric on cones is employed as well as certain mixing properties of the signal, albeit with important differences. Mixing and ultimately filter stability in the present situation are due to the expanding dynamics rather than the stochasticity of the signal process. In fact, the conditions even permit iterations of a fixed (nonrandom) expanding map.

1. INTRODUCTION

This paper concerns the long term properties of the filtering process, a subject that has attracted considerable interest for some time. Suppose that the state of a system is modeled by a discrete-time homogenous Markov chain $\{X_n\}_{n\geq 0}$ with values in some measurable space E. Suppose also that the state X_n is not directly observed, and that it can only be observed through observations that are realisations of some stochastic process $\{Y_n\}_{n>0}$ with state space $\mathbb{R}^d$. Concerning the precise connection between the processes $\{X_n\}$ and $\{Y_n\}$, we assume, as most authors do, a memoryless channel (although more general setups have been considered), which means that conditionally on $\{X_n\}_{n\geq 0}$, the observations $\{Y_n\}_{n>0}$ are independent and identically distributed with some nonsingular distribution (precise definitions will be given later). The stochastic filtering problem amounts to computing the conditional probability π_n of X_n given the observations $Y_0, \ldots, Y_n$. The process $\{\pi_n\}_{n\geq 0}$ takes values in the space of probabilities of E, and is called the filter process.

An important problem in stochastic filtering is to establish the asymptotic stability of π_n depending on its initial state π_0. The value

π_n of the filter process at time n can be obtained from π_{n-1} and Y_n by an essentially Bayesian calculation. There exists a family of linear operators $\{\mathcal{L}_y\}_{y\in\mathbb{R}^d}$ on the space of signed measures over E, leaving invariant the set of probability measures such that the filtering process satisfies the iterative relation

$$(1) \qquad \pi_n = \frac{\mathcal{L}_{Y_n}\pi_{n-1}}{\mathcal{L}_{Y_n}\pi_{n-1}(E)} =: \hat{\mathcal{L}}_{Y_n}\pi_{n-1}$$

with initial condition $\pi_0 = \mathbb{P}(X_0 \in \cdot)$. The operators $\mathcal{L}_{Y_n}$ and $\hat{\mathcal{L}}_{Y_n}$ are called the unnormalised and the normalised filtering operators, respectively. Note that $\hat{\mathcal{L}}_{Y_n}$ is a nonlinear operator due to the normalisation.

One of the difficulties with applying the nonlinear filter in practice is that the initial condition π_0 is required to be known. An experimentor who sets up an optimal filter for a series of incoming observations though will hardly know the correct initial distribution of X_0. Using the wrong initial condition in (1) will result in a filter with suboptimal performance. This would not be so much of a problem though if the filter were insensitive to misspecification of the initial condition, in the sense that it "forgets" the initial condition as time progresses. In other words, considering the wrongly initialised filtering process

$$(2) \qquad \pi_n^{(\rho)} = \hat{\mathcal{L}}_{Y_n}\pi_{n-1}^{(\rho)}$$

with initial condition $\pi_0^{(\rho)} = \rho$, where ρ is some probability measure over E, we are interested in the difference between $\pi_n^{(\rho)}$ and π_n as n becomes large.

Insensitivity of the optimal filter with respect to the initial condition has been investigated by many authors and from different angles. In [14], the authors distinguish "intrinsic" approaches from approaches which consider (1) as the sequential application of positive random operators acting on signed measures. According to the authors, intrinsic approaches exploit the probabilistic interpretation of the filtering process as a conditional probability, as opposed to non–intrinsic ones, where filter stability is investigated using for example the Hilbert projective metric or the Dobrushin ergodic coefficient (see e.g. [2, 3, 10, 11, 12, 21, 27, 26]); our results will develop the latter approach further. (Although the authors of [14] identify strong mixing properties of the signal—regardless of the observation structure—as one "intrinsic" mechanism ensuring filter stability, there is a strong overlap with approaches employing positive random operator tools, as the latter will necessarily have to exploit the properties of the signal process.)

In this paper, we assume $\{X_n\}$ to be generated by compositions of random maps. More precisely, we consider a sequence of independent identically distributed transformations $\phi_n \colon E \to E$ of a compact connected Riemannian manifold E with canonical volume m. If X_0 is

independent from all ϕ_n, then the relation

$$X_n = \phi_{n-1}(X_{n-1})$$

defines a homogenous Markov process. Our main conditions (to be made precise later) will be that

(1) ϕ is a random local diffeomorphism of a manifold E of class $C^{1+\alpha}$ such that $\|D_x\phi\, v\| \geq \sigma_0\|v\|$ for every $v \in T_x E$, and for some nonrandom constant $\sigma_0 > 1$.
(2) Conditionally on $\{X_n\}_{n \geq 0}$, the observations are independent and identically distributed, and the (conditional) distribution has a density (the likelihood) with respect to some σ-finite measure.
(3) The logarithm of the likelihood function is of class C^α on E.
(4) Certain integrability conditions apply to the Hölder coefficients of $\log|\det D_x\phi|$ and the logarithm of the likelihood function.

Condition 1 above implies that each realisation of ϕ is uniformly expanding. Expanding maps form a nontrivial but still relatively simple class of "chaotic" maps with strong ergodic properties such as mixing and exponential decay of correlations.

The main results of this paper are the following. First, we will argue that under the conditions above, we can assume the observation process to be ergodic. We may then assume that there exists an ergodic automorphism $T \colon \Omega \to \Omega$ preserving the probability $\mathbb{P}$ so that $Y_n(\omega) = Y_1(T^{n-1}\omega)$. Our main Theorem 3.1 shows that there exist a unique stationary filtering density, that is a positive function $h \colon \Omega \times E \to \mathbb{R}$ so that $h(\omega, \cdot) \in C^\alpha$ and $\int_E h(\omega, x)\mathrm{d}m(x) = 1$ so that if $\rho_\omega = h_\omega m$, then we have

(1) $\{\rho_{T^n\omega}\}_{n \geq 0}$ is a stationary orbit of the filter equation (1).
(2) ρ_ω emerges as the limit of "going backwards" in Equation (2), with exponential speed of convergence in $\|\cdot\|_{\mathrm{TV}}$.
(3) Initial conditions q with C^α densities are forgotten exponentially fast, that is

$$\varlimsup_{n \to +\infty} \frac{1}{n} \log \|\hat{\mathcal{L}}_n(\omega)q - \rho_{T^n\omega}\|_{\mathrm{TV}} \leq \xi.$$

(We have used the short hand $\hat{\mathcal{L}}_n(\omega)q = \hat{\mathcal{L}}_{Y_1(T^{n-1}\omega)} \circ \cdots \circ \hat{\mathcal{L}}_{Y_1(\omega)}q$.) The norm $\|\cdot\|_{\mathrm{TV}}$ denotes the total variation norm, which for a signed measure μ on E is defined as

$$\|\mu\|_{\mathrm{TV}} = \sup_A \mu(A) - \inf_A \mu(A),$$

where sup and inf run over all Borel sets of E. Note also that $\|\cdot\|_{\mathrm{TV}}$ is $\|\cdot\|_1$ on the level of densities with respect to the Riemannian volume m on E.

The proof of Theorem 3.1 consists of two major steps. Firstly, we show that under our assumptions on the system ϕ, the operators $\mathcal{L}_\omega$

preserve certain cones consisting of positive Hölder continuous functions. By equipping these cones with Hilbert's projective metric, we show that the operators $\mathcal{L}_\omega$ are contractions using classical results from [5, 6]. As a second step, the existence of the function h is proved employing a Fixed Point Theorem for random Lipschiz maps contracting on average (generalising a theorem of Bougerol [8]).

Our paper is similar to previous work on the stability of the nonlinear filter (e.g. [2, 10, 27]) in that we employ Hilbert's projective metric on cones, albeit with important differences. In those papers, the authors work with a fixed cone $\mathcal{C}$, namely the cone of positive measures in the Banach space of signed measures with the total variation norm, and require that there is a finite measure λ on the state space E of the Markov process so that the transition kernel $K(x, B)$ satisfies the strong mixing condition

$$(3) \qquad c\lambda \leq K(x,.) \leq \frac{1}{c}\lambda \qquad \text{for all } x \in E \text{ and some } c > 0.$$

This condition ensures that the image of the cone $\mathcal{C}$ under the Markov kernel has finite diameter in the Hilbert metric. The contracting effect of the filter is due to the stochasticity of the signal process, while the observation process is largely unimportant under these circumstances.

In our setting, the condition (3) is not satisfied. In fact, our conditions permit situations in which the ϕ_n are not even random but equal to a fixed (expanding) map ϕ, in which case $K(x, B) = \delta_{\phi(x)}(B)$ and hence condition (3) cannot hold. Mixing and ultimately filter stability in our situation is due to the expanding dynamics, which has a tendency to render smooth densities even smoother and thus acts similar to the stochasticity in the situation of mixing Markov kernels. In order to exploit this though, we cannot work with a single cone but have to construct a random family of cones. Another important aspect of the present setting is that the filter will *not* be stable with respect to arbitrary initial conditions; the initial conditions need to exhibit a certain smoothness already, as explained in item 3 above.

Nonrandom expanding maps have been extensively investigated, see for instance [4, 31] and references therein. Random expanding maps have been considered too, and many techniques presented in the present paper have been used in previous publications to find random invariant measures of random dynamical systems; this problem bears strong similarity to ours. For instance [22, 23, 24, 34] consider the case of random maps expanding on average. In [34], Hilbert's projective metric is used to find random invariant measures of random maps expanding on average, including measures that are absolutely continuous with respect to Lebesgue measure. The strongly related problem of random subshifts was considered by [7, 19]; see also [4] for many interesting applications.

In [33], invariant measures of random Lasota-Yorke maps were investigated, again using Hilbert's projective metric. In that work, the

author proves a general result for what he calls *good transfer operators*, and then shows that transfer operators for random Lasota-Yorke maps emerge as a special case. The result of Buzzi however requires that the operators preserve the integral of nonnegative functions; hence it cannot directly be applied to filter operators (at least not without modification). Further, good transfer operators (in the sense of Buzzi) work on function spaces with a notion of variation, while in the present paper we prefer to use functions that are log–Hölder continuous which seem to be more amenable to our setting. Building on the work of Buzzi, Froyland et al established the stability of the invariant measure of random Lasota- Yorke maps under various types of perturbations [32]. The result is formulated for abstract random linear operators as in the work of Buzzi, and for the same reasons the result does not apply, at least not directly, to our filter operators.

The paper is organized as follows. In Section 2, we briefly review the theory of nonlinear filtering and prove two lemmas concerning the ergodic properties of the observations. Further, we present the particular signal processes discussed in this paper. In Section 3, we formulate our main result in Theorem 3.1. As already explained, this theorem is proved by using Hilbert's projective metric on cones and a Fixed Point Theorem for random maps. The latter is proved in Section 4. In Section 5, we introduce Hilbert's projective metric on cones and recall relevant results for this metric obtained by Garrett Birkhoff. Finally, Section 6 contains the proof of Theorem 3.1.

2. Nonlinear filtering – general theory

In this section we will explain the notion of nonlinear filtering in an abstract setting.

2.1. **Basic Definitions and assumptions.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The *signal process* is a homogeneous Markov process $\{X_n \colon n \in \mathbb{N}_0\}$ on a polish space E endowed with the Borel σ-algebra $\mathcal{B}$. By K we will denote the transition kernel of $\{X_n\}$ (i.e. K is regular and $K(x, B) = \mathbb{P}(X_1 \in B | X_0 = x)$ a.s.); further, π_0 denotes the distribution of X_0. Throughout the paper, we will use the abbreviations $K(x, \varphi) = \int_E \varphi(z) K(x, \mathrm{d}z)$ and $K\mu(B) = \int_E K(x, B) \mathrm{d}\mu(x)$.

For any $I \subset \mathbb{Z}$, we write $\mathcal{X}_I$ for the σ-algebra generated by $\{X_k \colon k \in I \cap \mathbb{N}_0\}$. The *observation process* $\{Y_n \colon n \in \mathbb{N}\}$ is a process on $\mathbb{R}^d$. By $\mathcal{A}$, we denote the Borel algebra of $\mathbb{R}^d$. Let $\mathcal{Y}_I$ be the σ-algebra generated by $\{Y_k \colon k \in I \cap \mathbb{N}\}$. We assume that $\mathcal{Y}_I$ (resp. $\mathcal{X}_I$) are trivial if $I \cap \mathbb{N}$ (resp. $I \cap \mathbb{N}_0$) are empty.

Definition 2.1. The *filtering process* $\{\pi_n \colon n \in \mathbb{N}_0\}$ is a sequence of regular conditional probabilities on E so that for $\mathbb{P}$–a.a. ω,

(4) $$\pi_n(B) = \mathbb{P}\left(X_n \in B | \mathcal{Y}_{[1,n]}\right) \qquad \forall B \in \mathcal{B}.$$

Note that π_0 is the distribution of X_0 in agreement with our previous definition of π_0.

The problem of calculating π_n is called *nonlinear filtering*. Provided that further assumptions apply (to be specified later), the filtering process can be calculated in an iterative fashion.

We make the assumption of a homogeneous memoryless channel throughout the paper. This means that given $A_k \in \mathcal{A}$ for $k = 1, \ldots, n$, we have

$$(5) \qquad \mathbb{P}\left(Y_1 \in A_1, \ldots, Y_n \in A_n | \mathcal{X}_{[1,n]}\right) = \prod_{k=1}^{n} \mathbb{P}(Y_k \in A_k | X_k).$$

Since $\mathbb{R}^d$ is separable, there exist regular conditional probabilities

$$\Gamma_n \colon \mathcal{A} \times E \to [0, 1]$$

so that for any $A \in \mathcal{A}$ we have $\mathbb{P}(Y_n \in A | X_n) = \Gamma_n(A, X_n)$ a.s. We further assume that Γ_n does not depend on n, and that $\Gamma(\cdot, x)$ is absolutely continuous with respect to some σ-finite Borel measure λ on $\mathbb{R}^d$ for all x. Define the *likelihood function*

$$g(y, x) := \frac{\mathrm{d}\Gamma(\cdot, x)}{\mathrm{d}\lambda}(y).$$

Again due to the separability of $\mathbb{R}^d$, we can assume that g is measurable on $(\mathbb{R}^d \times E, \mathcal{A} \otimes \mathcal{B})$ (see Part (3) of Lemma A.1). Further, due to Tonelli's theorem we also have that for any probability measure ν on $(E, \mathcal{B})$,

$$\int_{\mathbb{R}^d \times E} g \, \mathrm{d}(\lambda \otimes \nu) = \int_E \int_{\mathbb{R}^d} g(y, x) \mathrm{d}\lambda(y) \mathrm{d}\nu(x) = 1,$$

hence g is integrable with respect to $\lambda \otimes \nu$ and $g(y, \cdot)$ is integrable with respect to ν except for y in some λ–null set.

2.2. Filtering equations.

Proposition 2.2. *Under the assumptions on $\{X_n, Y_n\}$ imposed in Section 2.1, the filtering process $\{\pi_n\}$ satisfies the following recursion equations:*

$$(6) \qquad \pi_n(\varphi) = \frac{\int_E \varphi(x) g(Y_n, x) \mathrm{d}\pi_{n-1}^+(x)}{\int_E g(Y_n, x) \mathrm{d}\pi_{n-1}^+(x)},$$

$$(7) \qquad \pi_n^+(\varphi) = \int_E K(z, \varphi) \pi_n(\mathrm{d}z).$$

The denominator in (6) *is* $\mathbb{P}$*–a.s. positive.*

Equations (6) and (7) are referred to as prediction and update step, respectively. For a proof, see [15, Section 3.2, Theorem 3.9]. In fact, in [15], different conditions are imposed which however contain an error, as a closer inspection of the proof will reveal. The first starred condition

in Section 3.2 of [15] should state that $\{X_n\}$ is Markov with respect to the filtration $\mathcal{X}_{[1,n]} \vee \mathcal{Y}_{[1,n]}$, rather than just with respect to its own filtration [16]. Then it can be shown that the conditions in [15] are equivalent to ours.

The proof of the following lemma is straightforward and is left to the reader.

Lemma 2.3. *Suppose that π_n^+ is absolutely continuous with respect some σ-finite measure m on E. Then π_{n+1} is absolutely continuous with respect to m as well. Moreover,*

$$(8) \qquad \frac{\mathrm{d}\pi_{n+1}}{\mathrm{d}m}(x) = \frac{1}{c}g(Y_n, x)\frac{\mathrm{d}\pi_n^+}{\mathrm{d}m}(x),$$

where

$$(9) \qquad c = \int_E g(Y_n, x)\frac{\mathrm{d}\pi_n^+}{\mathrm{d}m}(x)\mathrm{d}m(x).$$

The following two results will be relevant later for the type of signal processes considered mainly in this paper (Sec.2.3).

Lemma 2.4. *Suppose that the signal process $\{X_n\}$ is stationary, and the memoryless channel assumption holds. Then the process $\{X_n, Y_n\}$ is a stationary Markov process.*

Again, the proof is simple and left to the reader. The following proposition will be important later in our work. It relates the ergodicity of the observation process, which we will need, to the ergodicity of the signal process, resulting from the assumptions we will impose.

Proposition 2.5. *Assume the memoryless channel assumption holds, and suppose that $\{X_n\}$ is ergodic. Then $\{Y_n\}$ is also ergodic.*

Proof. For sets $A_1 \in \mathcal{Y}_{[1,k]}$ and $A_2 \in \mathcal{Y}_{[k+1,k+m]}$, the relation

$$(10) \qquad \mathbb{P}(A_1 \cap A_2|\mathcal{X}_{[1,n]}) = \mathbb{P}(A_1|\mathcal{X}_{[1,n]}) \cdot \mathbb{P}(A_2|\mathcal{X}_{[1,n]}) \qquad \text{a.s.}$$

follows from the memoryless channel assumption as long as $k+m \leq n$. We also have that if B is an m–dimensional cylinder set in $\mathcal{A}^{\mathbb{N}}$, then

$$(11) \qquad \begin{aligned} &\mathbb{P}(\{(Y_{k+1}, \ldots, Y_{k+m}) \in B\}|\mathcal{X}_{[1,n]} = (x_1, \ldots, x_n)) \\ &= \mathbb{P}(\{(Y_1, \ldots, Y_m) \in B\}|\mathcal{X}_{[1,n-k]} = (x_{k+1}, \ldots, x_n)) \qquad \text{a.s.} \end{aligned}$$

again from the memoryless channel assumption as long as $k+m \leq n$ (here we use the fact that Γ_n does not depend on n). Letting $n \to \infty$ in these relations, we get by martingale convergence

$$(12) \qquad \mathbb{P}(A_1 \cap A_2|\mathcal{X}_{[1,\infty]}) = \mathbb{P}(A_1|\mathcal{X}_{[1,\infty]}) \cdot \mathbb{P}(A_2|\mathcal{X}_{[1,\infty]}) \qquad \text{a.s.}$$

from Equation (10) and

$$(13) \qquad \begin{aligned} &\mathbb{P}(\{(Y_{k+1}, \ldots, Y_{k+m}) \in B\}|\mathcal{X}_{[1,\infty]} = (x_1, x_2, \ldots)) \\ &= \mathbb{P}(\{(Y_1, \ldots, Y_m) \in B\}|\mathcal{X}_{[1,\infty]} = (x_{k+1}, x_{k+2}, \ldots)) \qquad \text{a.s.} \end{aligned}$$

from Equation (11), respectively. Since $\mathbb{P}$ is regular, (12) holds for $A_2 \in \mathcal{Y}_{[k+1,\infty]}$, and (13) holds for $B \in \mathcal{A}^{\mathbb{N}}$, in the sense that

$$(14) \quad \begin{aligned} &\mathbb{P}(\{(Y_{k+1}, Y_{k+2}, \ldots) \in B\}|\mathcal{X}_{[1,\infty]} = (x_1, x_2, \ldots)) \\ &= \mathbb{P}(\{(Y_1, Y_2, \ldots) \in B\}|\mathcal{X}_{[1,\infty]}) = (x_{k+1}, x_{k+2}, \ldots)) \qquad \text{a.s.} \end{aligned}$$

If $A \in \mathcal{Y}_{[1,\infty]}$ is invariant, i.e., if there is $B \in \mathcal{A}^{\mathbb{N}}$ such that

$$A = \{(Y_{k+1}, Y_{k+2}, \ldots) \in B\} \qquad \text{for every } k,$$

then (14) shows that $\mathbb{P}(A|\mathcal{X}_{[1,\infty]})$ is an invariant random variable and thus almost sure equal to $\mathbb{P}(A)$ by ergodicity of $\{X_n\}$. On the other hand, an invariant event is in $\mathcal{Y}_{[k,\infty]}$ for every k, so (12) applies with A in place of A_2, and we get

$$\mathbb{P}(A_1 \cap A|\mathcal{X}_{[1,\infty]}) = \mathbb{P}(A_1|\mathcal{X}_{[1,\infty]})\mathbb{P}(A),$$

or after taking expectations

$$(15) \qquad\qquad \mathbb{P}(A_1 \cap A) = \mathbb{P}(A_1)\mathbb{P}(A)$$

for invariant A and $A_1 \in \mathcal{Y}_{[1,k]}$. We now approximate A with sets $A^k \in \mathcal{Y}_{[1,k]}$ so that $A^k \to A$ in symmetric difference; (15) holds for any such A^k in place of A_1. Taking limits in (15) then gives $\mathbb{P}(A) = \mathbb{P}(A)^2$ for any invariant A, establishing ergodicity of $\{Y_n\}$. $\qquad\square$

2.3. **Random mappings.** The results of this paper apply to signal processes of a particular kind. We assume that $\{X_n\}$ is generated by iterating a random sequence of transformations of a manifold. The transformations are assumed to be independent and identically distributed. More specifically, let now E be a compact connected smooth Riemannian manifold. Denote by $\mathcal{B}$ the Borel σ-algebra and by m the Riemannian volume of E. Further, let

$$\phi \colon \Omega \times E \to E, \quad (\omega, x) \mapsto \phi(\omega, x)$$

be a measurable mapping so that $\phi(\omega, .)$ is a local C^1–diffeomorphism for all ω. We will refer to such mappings as *random local diffeomorphisms*. Since $\phi(\omega, \cdot)$ is continuous for all ω, it is sufficient that ϕ is measurable only in ω for all x to obtain joint measurability in (ω, x) (see Part (4) of Lemma A.1). A random local diffeomorphism ϕ gives rise to a random variable $\omega \to \phi(\omega, \cdot)$ with values in E^E with the product σ-algebra $\mathcal{B}^{\otimes E}$; by slight abuse of notation, this random variable will again denoted by ϕ.

Assume that we are given a sequence ϕ_n of random local diffeomorphisms that are independent and identically distributed (as random variables in the sense defined above), and let $X_0 \colon \Omega \to E$ be a random variable independent from all ϕ_n. Then the relation

$$X_n = \phi_n(\omega, X_{n-1}), \quad n \geq 1$$

defines a homogenous Markov process on E with transition kernel given by
$$K(x, B) = \mathbb{P}(\{\omega \colon \phi_n(\omega, x) \in B\}) = \mathbb{E}(\mathbb{1}_B \circ \phi_n(\omega, x)).$$
and initial distribution $\pi_0(B) = \mathbb{P}(X_0 \in B)$, where $B \in \mathcal{B}$ and $x \in E$. For a proof of these assertions, see [1, Theorem 2.1.7.].

For the remainder of this section, we will establish a representation of the filtering process in terms of densities where the Frobenius–Perron operator plays an essential role. (A very fruitful approach to studying the statistical properties of deterministic dynamical system goes via investigating the spectrum of the associated Frobenius–Perron operator and more general transfer operators, see [4, 31].)

Definition 2.6. The *Frobenius–Perron operator* associated with a local diffeomorphism ϕ is defined as
$$\mathcal{P}\psi(x) = \sum_{y \in E; \phi(y) = x} \psi(y) |\det D\phi(y)|^{-1}$$
for $\psi \colon E \to \mathbb{R}$.

The number of $y \in E$ with $\phi(y) = x$ (i.e. the number of terms in the sum representing the Frobenius–Perron operator) is called the *degree of x*. Due to our assumptions on ϕ, the degree is a finite constant [25, Proposition 4.46]. Note that by substitution, we have the identity
$$\int \varphi \circ \phi(x)\, \psi(x)\, \mathrm{d}m(x) = \int \varphi(x)\, \mathcal{P}\psi(x)\, \mathrm{d}m(x)$$
for φ measurable and bounded and $\psi \in L^1(m)$, the space of all Borel measurable functions $\psi \colon E \to \mathbb{R}$ with $\int_E |\psi(y)| \mathrm{d}m(y) < \infty$.

Lemma 2.7. *Suppose that $h \colon \Omega \times E \to \mathbb{R}$ is nonnegative, measurable, and $h(\omega, \cdot) \in L^1(m)$ for every $\omega \in \Omega$. Then the same is true for the mapping $(\omega, x) \mapsto \mathcal{P}(\omega)h(\omega, x)$.*

Proof. The function $(\omega, x) \mapsto \mathbb{1}_B \circ \phi(\omega, x)h(\omega, x)$ is nonnegative, measurable, and in $L^1(m)$ for every ω, so
$$\omega \mapsto \int \mathbb{1}_B \circ \phi(\omega, x)h(\omega, x)\mathrm{d}m(x) = \int_B \mathcal{P}(\omega)h_\omega(x)\mathrm{d}m(x)$$
is finite and measurable by Tonelli's theorem. Hence
$$(\omega, x) \mapsto \mathcal{P}(\omega)h_\omega(x)$$
is nonnegative and in $L^1(m)$ for every ω, and joint measurability follows from Part (3) of Lemma A.1. $\qquad\square$

Lemma 2.8. *Suppose that π is a probability measure on E with density $h(x)$, then*
$$\frac{\mathrm{d}K\pi}{\mathrm{d}m}(x) = \mathbb{E}\left(\mathcal{P}(\omega)h(x)\right) = \int \mathcal{P}(\omega)h(x)\mathbb{P}(\mathrm{d}\omega).$$

Proof. Since $(\omega, x) \mapsto \mathcal{P}(\omega)h(x)$ is measurable by Lemma 2.7, the statement is an easy application of Tonelli's theorem. □

If we now combine Lemmas 2.3 and 2.8 with Proposition 2.2, we obtain the following representation of the filtering process in terms of densities.

Proposition 2.9. *Suppose the signal process is generated as stated in the beginning of this section, and the observations satisfy the memoryless channel assumption. Let $\{\pi_n \colon n \in \mathbb{N}_0\}$ be the filtering process, and assume that π_n for some n has a density $p_n(x)$ with respect to the Riemannian volume m. Then also π_{n+1} has a density p_{n+1} with respect to m given by*

$$p_{n+1}(x) = \frac{1}{w_n} g(Y_{n+1}, x) \overline{\mathcal{P}} p_n(x)$$

with

$$\overline{\mathcal{P}} \colon h \mapsto \int \mathcal{P}(\omega) h(x) \mathrm{d}\mathbb{P}(\omega)$$

and the normalisation factor

$$w_n = \int g(Y_{n+1}, x) \overline{\mathcal{P}} p_n(x) m(\mathrm{d}x)$$

In view of Proposition 2.9 and to facilitate the subsequent discussion, we make the following definitions.

Definition 2.10. With a given likelihood function g and a given random local diffeomorphism ϕ as defined at the beginning of this section, we associate the family of operators

$$\mathcal{L}(y) \colon h \mapsto g(y, x) \overline{\mathcal{P}} h(x), \quad \hat{\mathcal{L}}(y) \colon h \mapsto \frac{\mathcal{L}(y)h(x)}{\|\mathcal{L}(y)h\|_1},$$

which are called the unnormalised and the L^1–normalised filter operators, respectively. The L^1–norm is taken with respect to m, and the operator $\overline{\mathcal{P}}$ is as defined in Proposition 2.9.

Remark 2.11. Our conditions imposed later on the likelihood function will ensure that the denominator $\|\mathcal{L}(y)h(x)\|_1$ is not zero for every nonnegative and measurable h with $\|h\|_1 = 1$. Note that in the filter recursion, positivity of the denominator is ensured by Proposition 2.2.

3. MAIN RESULT

We will first formulate our main result in a somewhat more abstract setting and later establish the connection with filtering. Suppose that $\{Y_n \colon n \in \mathbb{N}\}$ is a stationary and ergodic process with state space $\mathbb{R}^d$. By stationarity (and possibly after extending the probability space), we can construct random variables $Y_0, Y_{-1}, Y_{-2}, \ldots$ so that $\{Y_n, n \in \mathbb{Z}\}$ becomes a stationary and ergodic process over $\mathbb{Z}$. We may then assume that there exists an ergodic automorphism $T \colon \Omega \to \Omega$ preserving the

probability $\mathbb{P}$ so that $Y_n(\omega) = Y_1(T^{n-1}\omega)$. If at the same time $\{Y_n\}$ were the observations of some signal process in the sense of Proposition 2.9, then (according to that proposition) the densities $p_n(x)$ of the filtering process for any n could be generated by applying an ergodic sequence of filter operators to the initial density p_0. Namely, we would have

$$p_n(x) = \hat{\mathcal{L}}_{Y_1(T^{n-1}\omega)} \circ \cdots \circ \hat{\mathcal{L}}_{Y_1(T^k\omega)} p_k(x).$$

The subject of our main theorem is the asymptotic behaviour of stationary operator sequences of this form, where the operators are as in Definition 2.10.

To formulate the theorem, let $0 < \alpha \leq 1$, and ϕ be a random local diffeomorphism of a manifold E of class $C^{1+\alpha}$. Define C_+^α as the set of strictly positive functions on E whose logarithm is Hölder continuous with exponent α. Let g be a likelihood function on E with $g(y, \cdot) \in C_+^\alpha$ for all y. The least expansion coefficient of $\phi(\omega, \cdot)$ is given by

$$\sigma(\omega) = \min_{x \in E} \|(D_x\phi(\omega, \cdot))^{-1}\|^{-1}.$$

The function $\sigma(\omega)$ is strictly positive, because $\phi(\omega, \cdot)$ is a local diffeomorphism. Next, define

$$b(\omega) = |\log|\det D_x\phi(\omega, \cdot)||_\alpha \quad \text{and} \quad c(\omega) = \max\{|\log g(Y_1(\omega), \cdot)|_\alpha, 1\},$$

where $|\cdot|_\alpha$ denotes the Hölder constant with respect to Hölder exponent α. The separability of E implies that the functions σ, b, c are measurable (see Lemma 6.6). We further impose on σ, b, c the following conditions:

C1 (Uniform Expansion and Regularity): there exist $b_0 > 0$ and $\sigma_0 > 1$ such that $\sigma_0 \leq \sigma(\omega)$ and $b(\omega) \leq b_0$ for a.e. $\omega \in \Omega$;

C2 (Temperedness): $\overline{\lim}_{n \to \pm\infty} |n|^{-1} \log c(T^n\omega) = 0$ for a.e. $\omega \in \Omega$.

At the end of this section, we discuss an example satisfying these assumptions.

For the purpose of the following theorem, we use the shorthand $\mathcal{L}(\omega)$ instead of $\mathcal{L}(Y_1(\omega))$ and put $\mathcal{L}_n(\omega) = \mathcal{L}(T^{n-1}\omega) \circ \cdots \circ \mathcal{L}(\omega)$. Similar notation is used for $\hat{\mathcal{L}}$.

Theorem 3.1. *Suppose that ϕ and g satisfy conditions C1 and C2, and let $\mathcal{L}(\omega)$ and $\hat{\mathcal{L}}(\omega)$ be the associated filter operators as in Definition 2.10. Let $\xi = \int r(\omega) d\mathbb{P}(\omega) < 0$, where r is as in the proof of Lemma 6.5. Then there exist a T-invariant set $\Omega_0 \subset \Omega$ of full measure, a unique[1] pair of measurable functions $h: \Omega \times E \to \mathbb{R}$ and $\Lambda: \Omega \to (0, +\infty)$ such that for $\omega \in \Omega_0$ and $\varphi \in C_+^\alpha$, we have*

(1) $h(\omega, \cdot) \in C_+^\alpha(E)$ and $\|h(\omega, \cdot)\|_1 = 1$,

(2) $\overline{\lim}_{n \to +\infty} \dfrac{1}{n} \log \|h(\omega, \cdot) - \hat{\mathcal{L}}_n(T^{-n}\omega)\mathbb{1}\|_\infty \leq \xi$,

---

[1] The uniqueness has to be understood in the sense explained in the footnote of Theorem 4.3.

(3) $\mathcal{L}(\omega)h(\omega,\cdot) = \Lambda(\omega)h(T\omega,\cdot),$

(4) $\overline{\lim}_{n\to+\infty} \dfrac{1}{n} \log \|h(T^n\omega,\cdot) - \hat{\mathcal{L}}_n(\omega)\varphi\|_\infty \leq \xi,$

(5) $\overline{\lim}_{n\to\pm\infty} \dfrac{1}{|n|} \log \|h(T^n\omega,\cdot)\|_\infty \leq 0.$

The interpretation of Theorem 3.1 in terms of filtering is straight forward if indeed the observations $\{Y_n\}$ are ergodic. In this situation, $\mathcal{L}(T^{n-1}\omega)$ is the filtering operator associated with the observation Y_n, and the eigenfunction h is to be interpreted as a stationary solution of the filtering equations, in the following sense: If we combine conclusions (1) and (3) for the L^1–norm, we obtain

$$\hat{\mathcal{L}}(\omega)h(\omega,\cdot) = h(T\omega,\cdot).$$

The conclusion (2) of Theorem 3.1 provides information as to how the stationary filtering process h is constructed, namely as the limit of the sequence $\{\hat{\mathcal{L}}_n(T^{-n}\omega)\mathbb{1}\}_{n\in\mathbb{N}}$. This means that the stationary filtering process is essentially a filtering process that started in the infinitely remote past, and also provides a way to approximate h. Conclusion (4) of Theorem 3.1 gives a related but slightly different result, namely that a forward orbit of the form $\{\hat{\mathcal{L}}_n(\omega)\varphi\}_{n\in\mathbb{N}}$ with $\varphi \in C^\alpha_+(E)$, that is, a filtering process initialised with some sufficiently smooth but otherwise arbitrary density φ (when driven by the same observations) is asymptotically attracted by the stationary filtering process $\{h(T^n\omega,\cdot)\}_{n\in\mathbb{N}}$ with exponential speed of convergence.

The ergodicity of $\{Y_n\}$ required in Theorem 3.1 would follow if $\{X_n\}$ were known to be ergodic, thanks to Proposition 2.5. But this can in fact be obtained as a corollary of Theorem 3.1 which requires only the condition C1 on the signal process $\{X_n\}$ (this result might be of independent interest):

Corollary 3.2. *Suppose that ϕ satisfies condition C1. Then there exist a constant $\xi < 0$ and a function $\bar{h} \in C^\alpha_+(E)$ with $\|\bar{h}\|_1 = 1$ such that for any $\varphi \in C^\alpha_+(E)$ with $\|\varphi\|_1 = 1$, we have*

(1) $\overline{\mathcal{P}}\bar{h} = \bar{h},$

(2) $\overline{\lim}_{n\to+\infty} \dfrac{1}{n} \log \|\overline{\mathcal{P}}^n\varphi - \bar{h}\|_\infty \leq \xi.$

Furthermore, $\bar{h}m$ is the unique invariant measure for the Markov kernel K which is equivalent to the Riemannian volume, and letting $\bar{h}m$ be the distribution of X_0 renders the signal process ergodic.

Proof. Statements 1 and 2 follow directly from Theorem 3.1 by considering the case $g = 1$. Further, statement 1 shows that $\bar{h}m$ is an invariant measure for the Markov kernel which is clearly equivalent to the Riemannian volume m because $\bar{h}$ is positive. The uniqueness

statement follows from ergodicity, which we will prove now. The proof is fairly standard. Let C be an invariant event for $\{X_n\}$ and put $\psi(x) = \mathbb{P}(C|X_0 = x)$. It can be shown (see [9], proof of Theorem 7.16) that ψ can assume the values 0 or 1 only $\bar{h}m$-a.s., and further that

$$(16) \qquad \psi(x) = \int \psi(y) K(x, \mathrm{d}y).$$

Thus we have to show that any ψ bounded and in $L^1(m)$ satisfying Equation (16) is $\bar{h}m$-a.s. equal to a constant. With $\varphi \in C_+^\alpha(E)$, we have (by Eq. (16), Lemma 2.8, and item 2, respectively)

$$
\begin{aligned}
\int \varphi(x)\psi(x)\,\mathrm{d}m(x) &= \|\varphi\|_1 \int \frac{\varphi(x)}{\|\varphi\|_1} K^n(\psi, x)\,\mathrm{d}m(x) \\
(17) \qquad &= \|\varphi\|_1 \int \overline{\mathcal{P}}^n\left(\frac{\varphi}{\|\varphi\|_1}\right)(x)\,\psi(x)\,\mathrm{d}m(x) \\
&\to \|\varphi\|_1 \int \overline{h}(x)\psi(x)\,\mathrm{d}m(x),
\end{aligned}
$$

which implies that $\psi(x) = \int \overline{h}(z)\psi(z)\mathrm{d}m(z)$, completing the proof. $\qquad\square$

Finally, the following theorem covers some situations in which the signal process is not ergodic.

Theorem 3.3. *Suppose that ϕ_n is an i.i.d. series of random local diffeomorphism of class $C^{1+\alpha}$ satisfying C1, that $x \to \mathbb{E}(\log_+(c)|X_1 = x)$ is in $L^1(m)$, and that $\frac{\mathrm{d}\mathbb{P}(X_0 \in \cdot)}{\mathrm{d}m} = \pi_0$ exists. Then there exist a constant $\xi < 0$ and a set $\Omega_0 \subset \Omega$ of full measure so that for any $\omega \in \Omega_0$ and $\varphi, \psi \in C_+^\alpha(E)$ we have*

$$(18) \qquad \varlimsup_{n \to +\infty} \frac{1}{n} \log \|\hat{\mathcal{L}}_n(\omega)\varphi - \hat{\mathcal{L}}_n(\omega)\psi\|_\infty \le \xi$$

where we write $\hat{\mathcal{L}}_n(\omega) = \hat{\mathcal{L}}(Y_n(\omega)) \circ \ldots \circ \hat{\mathcal{L}}(Y_1(\omega))$. In particular, if $\pi_0 \in C_+^\alpha(E)$ we have for the filtering process π_n and any $\varphi \in C_+^\alpha(E)$

$$\varlimsup_{n \to +\infty} \frac{1}{n} \log \|\hat{\mathcal{L}}_n(\omega)\varphi - \pi_n(\omega)\|_\infty \le \xi$$

Proof. We can introduce a new probability measure $\tilde{\mathbb{P}}$ on $\mathcal{Y}_\mathbb{N} \vee \mathcal{X}_{\mathbb{N}_0}$ without changing the distribution of $\{\phi_n\}$ by putting $\frac{\mathrm{d}\tilde{\mathbb{P}}(X_0 \in \cdot)}{\mathrm{d}m} = \bar{h}$ where $\bar{h}$ is as in Corollary 3.2. This renders $\{X_n\}$ ergodic and thus also $\{Y_n\}$ by Proposition 2.5. We can thus extend the probability space $(\Omega, \tilde{\mathbb{P}})$ so that $\{Y_n\}$ extends to negative times, and we are in the setting of Theorem 3.1. Condition C1 is satisfied by assumption. Condition C2 is satisfied if we can show that c is integrable with respect to $\tilde{\mathbb{P}}$ (see

Remark 4.2). But this is the case because

$$\tilde{\mathbb{E}}(c) = \int_E \mathbb{E}(c|X_1 = x)\bar{h}(x)\,\mathrm{d}m$$

$$\leq \|\bar{h}\|_\infty \int_E \mathbb{E}(c|X_1 = x)\,\mathrm{d}m < \infty$$

by assumption on c. Statement (18) now follows from item 4 in Theorem 3.1 and the triangle inequality. Since the set Ω_0 on which (18) is valid is in $\mathcal{Y}_\mathbb{N}$, and since $\frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\tilde{\mathbb{P}}}|_{\mathcal{Y}_\mathbb{N} \vee \mathcal{X}_{\mathbb{N}_0}} = \frac{\pi_0}{h}$, we have $\mathbb{P}(\Omega_0) = 0$, as claimed. The last statement follows because $\pi_n = \hat{\mathcal{L}}_n(\omega)\pi_0$. $\qquad\square$

Example 3.4. Let $E = \mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$ be the d–dimensional flat torus, and $A_k \in \mathbb{Z}^{d \times d}$ for $k = 1 \ldots l$ with $\|A_k^{-1}\|^{-1} \geq \sigma_0 > 1$ for some σ_0. Let $\{Z_n : n \in \mathbb{N}\}$ be a series of independent and identically distributed random variables on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in $\{1 \ldots l\}$. Then $\phi_n : (\omega, x) \mapsto A_{Z_n(\omega)}x \mod 1$ is an i.i.d. series of random local diffeomorphisms of $\mathbb{T}^d$ which satisfies assumption C1. Further, let $\{R_n : n \in \mathbb{N}\}$ be a series of iid standard normal random variables, independent of the $\{Z_n : n \in \mathbb{N}\}$. We assume that X_0 is independent of $\{R_n\}$ and $\{Z_n : n \in \mathbb{N}\}$, and that the signal process $\{X_n : n \in \mathbb{N}_0\}$ is generated as in Subsection 2.3. Now for some constant $s > 0$ and $w : \mathbb{T}^d \to \mathbb{R}$ a α–Hölder continuous function, let the observation process be $Y_n = w(X_n) + s \cdot R_n$ for $n \in \mathbb{N}$. With λ being the Lebesgue measure on $\mathbb{R}$, we get the likelihood $g(y, x) = \exp(-(y - w(x))^2/(2s^2))/(\sqrt{2\pi}s)$. We are now in the setting of Proposition 2.9.

Further, $|\log g(y, \cdot)|_\alpha = \frac{1}{s^2}|yw - \frac{1}{2}w^2|_\alpha \leq \frac{1}{s^2}(|w|_\alpha|y| + \frac{1}{2}|w|_\alpha^2)$ so that $|\log g(Y_1(\omega), \cdot)|_\alpha \leq \frac{1}{s^2}(|w|_\alpha|R_1| + |w|_\alpha\|w\|_\infty + \frac{1}{2}|w|_\alpha^2)$ which gives that $\mathbb{E}(c|X_1 = x)$ is bounded by a finite constant independent of x. Hence if $\frac{\mathrm{d}\mathbb{P}(X_0 \in \cdot)}{\mathrm{d}m} = \pi_0$ exists, the conditions of Theorem 3.3 are satisfied. Further, the conditions of Corollary 3.2 are also satisfied and if we let $\pi_0 = \bar{h}$, then the conditions of Theorem 3.1 are satisfied, too.

4. RANDOM FIXED POINT THEOREM

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space together with an automorphism $T : \Omega \to \Omega$ preserving the probability $\mathbb{P}$. The abstract dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, T)$ is assumed to be ergodic. In the rest of the paper, any almost everywhere true statement on Ω has to be understood with respect to the measure $\mathbb{P}$. We assume that there exist a collection $\{X_\omega\}_{\omega \in \Omega}$ of subsets of some space X and a complete metric d_ω on each X_ω. Let $U = \bigcup_{\omega \in \Omega} (\{\omega\} \times X_\omega)$. We also assume that there exists a transformation

$$\phi : U \to X$$

such that $\phi(\omega, X_\omega) \subseteq X_{T\omega}$. The iterations $\phi_n \colon U \to X$ of ϕ are given by

$$\phi_n(\omega, x) = \begin{cases} x & \text{if} \quad n = 0, \\ \phi(T^{n-1}\omega, \cdot) \circ \cdots \circ \phi(\omega, x) & \text{if} \quad n > 0. \end{cases}$$

For every $\omega \in \Omega$ and every $x \in X_\omega$, the sequence $\{\phi_n\}_{n \geq 0}$ satisfies the *cocycle property*:

$$\phi_{n+m}(\omega, x) = \phi_n(T^m\omega, \phi_m(\omega, x)), \qquad m, n = 0, 1, 2, \ldots,$$

and is called a random dynamical system on X over $(\Omega, \mathcal{F}, \mathbb{P}, T)$ generated by ϕ, or simply the *random dynamical system ϕ*.

Remark 4.1. The notion of random dynamical system presented here coincides essentially with the one of a discrete time random dynamical system generated by the random map ϕ given in Arnold's book [1]. The only difference between the two definitions is that our definition does not require the map ϕ to be measurable.

4.1. **Random fixed point theorem.** We now prove a random fixed point theorem (Theorem 4.3), which will be instrumental in proving our main result. The main hypotheses of the theorem are Conditions A1 and A2 below. The extra Condition A3 is required for the measurability and uniqueness of the random fixed point.

A1 (Contraction on average): The map $\phi(\omega, \cdot)$ is Lipschitz for all $\omega \in \Omega$. The Lipschitz constant of $\phi_n(\omega, \cdot) \colon (X_\omega, d_\omega) \to (X_{T^n\omega}, d_{T^n\omega})$ is denoted by $\rho_n(\omega)$ for all $\omega \in \Omega$ and $n \geq 0$. There exists an integrable function $r \colon \Omega \to \mathbb{R}$ such that $\log \rho_1(\omega) \leq r(\omega)$ for a.e. $\omega \in \Omega$ and $\beta := \int_\Omega r(\omega) \mathrm{d}\mathbb{P}(\omega) < 0$.

A2 (Temperedness): There exist a map $x_0 \colon \Omega \to X$ and a measurable function $g \colon \Omega \to (0, +\infty)$ with $\overline{\lim}_{n \to \pm\infty} |n|^{-1} \log g(T^n\omega) = 0$ for a.e. $\omega \in \Omega$ such that $x_0(\omega) \in X_\omega$ for every $\omega \in \Omega$, and $l(\omega) := d_\omega(x_0(\omega), \phi(T^{-1}\omega, x_0(T^{-1}\omega))) \leq g(\omega)$ for a.e. $\omega \in \Omega$.

A3 (Measurability): The set X is equipped with a metric d such that $d|_{X_\omega} \leq d_\omega$ for every $\omega \in \Omega$. Denote by $\mathcal{X}$ and $\mathcal{U}$ the Borel σ-algebra of X generated by d and the restriction of $\mathcal{F} \otimes \mathcal{X}$ to U, respectively. The transformations $\phi \colon (U, \mathcal{U}) \to (X, \mathcal{X})$ and $x_0 \colon (\Omega, \mathcal{F}) \to (X, \mathcal{X})$ are both measurable.

Remark 4.2. We recall that if $f \colon \Omega \to \mathbb{R}$ is measurable, then

$$\overline{\lim_{n \to \infty}} \frac{1}{n} f(T^n\omega) = \overline{\lim_{n \to \infty}} \frac{1}{n} f(T^{-n}\omega) \in \{0, +\infty\} \qquad \text{for a.e. } \omega \in \Omega.$$

and that the limit is equal to 0 if $f^+ := \max\{0, f\} \in L^1(\mathbb{P})$ [1, Proposition 4.1.3].

Theorem 4.3. *Let ϕ be a random dynamical system satisfying conditions A1 and A2. Let β be the constant in A1, and let x_0 be the map in A2. Then there exist a T-invariant set $\Omega_1 \subset \Omega$ of full measure and a map $Z : \Omega \to X$ such that for every $\omega \in \Omega_1$, we have*

(1) $Z(\omega) \in X_\omega$,

(2) $\overline{\lim}_{n \to +\infty} \frac{1}{n} \log d_\omega \left(Z(\omega), \phi_n(T^{-n}\omega, x_0(T^{-n}\omega)) \right) \le \beta$,

(3) $Z(T^{n+1}\omega) = \phi(T^n\omega, Z(T^n\omega))$ *for $n \in \mathbb{Z}$,*

(4) $\overline{\lim}_{n \to +\infty} \frac{1}{n} \log d_{T^n\omega}(Z(T^n\omega), \phi_n(\omega, x)) \le \beta$ *for $x \in X_\omega$,*

(5) $\overline{\lim}_{n \to \pm\infty} \frac{1}{|n|} \log d_{T^n\omega} \left(Z(T^n\omega), x_0(T^n\omega) \right) \le 0$.

Moreover, if condition A3 holds as well, then Z is measurable wrt to the Borel σ-algebra of (X, d), is the unique[2] measurable map satisfying conclusions (1) and (3), and conclusions (2),(4) and (5) hold with respect to the metric d as well.

Proof. By the Birkhoff Ergodic Theorem, there exists a measurable set $\Omega_* \subset \Omega$ of full measure such that

$$\lim_{n \to \pm\infty} \frac{1}{|n|} \sum_{k=0}^{n-1} r(T^k\omega) = \beta \qquad \text{for } \omega \in \Omega_*.$$

One can easily construct a T-invariant set $\Omega_1 \subset \Omega_*$ of full measure such that both conditions A1 and A2 are satisfied for every $\omega \in \Omega_1$.

For every $\omega \in \Omega_1$ and every $\epsilon > 0$, there exists $\bar{k}(\omega, \epsilon) > 0$ such that for every $k > \bar{k}(\omega, \epsilon)$,

$$(19) \qquad \left| \sum_{i=0}^{k-1} \left(r(T^{\pm i}\omega) - \beta \right) \right| \le \epsilon k \qquad \text{and} \qquad g(T^{-k}\omega) \le e^{\epsilon k}.$$

Hence,

$$(20) \qquad \rho_k(\omega) \le \prod_{i=0}^{k-1} \rho_1(T^i\omega) \le e^{\sum_{i=0}^{k-1} r(T^i\omega)} \le e^{(\beta+\epsilon)k},$$

$$(21) \qquad \rho_k(T^{-k}\omega) \le \prod_{i=0}^{k-1} \rho_1(T^{-i}\omega) \le e^{\sum_{i=0}^{k-1} r(T^{-i}\omega)} \le e^{(\beta+\epsilon)k},$$

and

$$(22) \qquad l(T^{-k}\omega) \le g(T^{-k}\omega) \le e^{\epsilon k}.$$

---

[2]This means that if $Y : \Omega \to X$ is a measurable map satisfying conclusions (1) and (3) on an invariant set $\Omega_2 \subset \Omega$ of positive measure, then $Y = Z$ a.e. on Ω_2.

For each $n \geq 0$, define

$$Z_n(\omega) = \begin{cases} \phi_n(T^{-n}\omega, x_0(T^{-n}\omega)) & \text{if } \omega \in \Omega_1, \\ x_0(\omega) & \text{if } \omega \in \Omega \setminus \Omega_1. \end{cases}$$

By A2, we have $Z_n(\omega) \in X_\omega$ for every $\omega \in \Omega$.

Let $\omega \in \Omega_1$ and $0 < \epsilon < -\beta/2$. From (21) and (22), it follows that if $n > m > \bar{k}(\omega, \epsilon)$, then

$$
\begin{aligned}
d_\omega\left(Z_m(\omega), Z_n(\omega)\right) &\leq \sum_{k=m}^{n-1} d_\omega\left(Z_k(\omega), Z_{k+1}(\omega)\right) \\
&\leq \sum_{k=m}^{n-1} \rho_k(T^{-k}\omega) l(T^{-k}\omega) \\
&\leq \sum_{k=m}^{n-1} e^{(\beta+2\epsilon)k} \leq \frac{e^{(\beta+2\epsilon)m}}{1 - e^{\beta+2\epsilon}}.
\end{aligned}
$$
(23)

To obtain the second inequality, we used the fact that $Z_{k+1}(\omega)$ is equal to $\phi_k(T^{-k}\omega, \phi(T^{-k-1}\omega, x_0(T^{-k-1}\omega)))$. It follows that $\{Z_n(\omega)\}$ is a Cauchy sequence. Since (X_ω, d_ω) is complete, there exists $Z(\omega) \in X_\omega$ such that

$$\lim_{n \to +\infty} d_\omega(Z_n(\omega), Z(\omega)) = 0,$$
(24)

proving the existence of Z and conclusion (1) of the theorem. By taking the limit in (23) as $n \to +\infty$, we obtain

$$d_\omega(Z_m(\omega), Z(\omega)) \leq \frac{e^{(\beta+2\epsilon)m}}{1 - e^{\beta+2\epsilon}} \qquad \text{for } m > \bar{k}(\omega, \epsilon).$$

Since ϵ is arbitrarily small, the previous inequality gives conclusion (2) of the theorem.

Let $\omega \in \Omega_1$. The set Ω_1 is T-invariant, and so $T^n\omega \in \Omega_1$ for every $n \in \mathbb{Z}$. From (24), it follows that

$$Z(T^{n+1}\omega) = \lim_{k \to +\infty} \phi_k(T^{-k+n+1}\omega, x_0(T^{-k+n+1}\omega)).$$

Using the cocycle property of ϕ and the continuity of $\phi(\omega, \cdot)$, we obtain conclusion (3) of the theorem:

$$
\begin{aligned}
Z(T^{n+1}\omega) &= \lim_{k \to +\infty} \phi_k(T^{-k+n+1}\omega, x_0(T^{-k+n+1}\omega)) \\
&= \lim_{k \to +\infty} \phi(T^n\omega, \phi_{k-1}(T^{-k+n+1}\omega, x_0(T^{-k+n+1}\omega))) \\
&= \phi(T^n\omega, Z(T^n\omega)).
\end{aligned}
$$

In particular, we have $Z(T^n\omega) = \phi_n(\omega, Z(\omega))$. This together (20) implies that if $x \in X_\omega$, then

$$
\begin{aligned}
d_{T^n\omega}\left(\phi_n(\omega, x), Z(T^n\omega)\right) &= d_{T^n\omega}\left(\phi_n(\omega, x), \phi_n(\omega, Z(\omega))\right) \\
&\leq \rho_n(\omega)d_\omega\left(x, Z(\omega)\right) \\
&\leq e^{(\beta+\epsilon)n}d_\omega(x, Z(\omega))
\end{aligned}
$$

for $\epsilon > 0$ and $n > \bar{k}(\omega, \epsilon)$, giving conclusion (4) of the theorem.

By the definition of Z and using the first two inequalities in (23), we obtain

$$
\begin{aligned}
d_\omega\left(Z(\omega), x_0(\omega)\right) &= \lim_{n\to+\infty} d_\omega\left(Z_n(\omega), Z_0(\omega)\right) \\
&\leq \lim_{n\to+\infty} \sum_{k=0}^{n-1} \rho_k(T^{-k}\omega)l(T^{-k}\omega) \\
&\leq \sum_{k=0}^{\infty} e^{\sum_{i=0}^{k-1} r(T^{-i}\omega)} g(T^{-k}\omega) =: f(\omega).
\end{aligned}
$$

The function f is well defined, because $\sum_{i=0}^{k-1} r(T^{-i}\omega) \sim k\beta < 0$ as $k \to \infty$, and $g \circ T^{-i}$ has sub-exponential growth. Note also that f is measurable, since it is the limit of measurable functions with values in $\mathbb{R}$. To prove conclusion (5), it is enough to show that

$$
\overline{\lim_{m\to+\infty}} \frac{1}{m} \log f(T^{-m}\omega) \leq 0 \qquad \text{for } \omega \in \Omega_1.
$$

Indeed, by Remark 4.2, we then have

$$
\overline{\lim_{m\to+\infty}} \frac{1}{m} \log f(T^{-m}\omega) = \overline{\lim_{m\to+\infty}} \frac{1}{m} \log f(T^m\omega) = 0 \quad \text{for a.e. } \omega \in \Omega.
$$

Fix $m > \bar{k}(\omega, \epsilon)$. From (19), it follows that

$$
\sum_{i=0}^{k-1} r(T^{-i-m}\omega) = \sum_{i=0}^{k+m-1} r(T^{-i}\omega) - \sum_{i=0}^{m-1} r(T^{-i}\omega) \leq \beta k + 2\epsilon m,
$$

and

$$
g(T^{-k-m}\omega) \leq e^{(k+m)\epsilon}.
$$

Now, fix $0 < \epsilon < \beta$. Combining all together, we obtain

$$
\begin{aligned}
f(T^{-m}\omega) &= \sum_{k=0}^{\infty} e^{\sum_{i=0}^{k-1} r(T^{-i-m}\omega)} g(T^{-k-m}\omega) \\
&\leq e^{3m\epsilon} \sum_{k=0}^{\infty} e^{(\beta+\epsilon)k} = \frac{e^{3m\epsilon}}{1 - e^{\beta+\epsilon}},
\end{aligned}
$$

and so

$$
\overline{\lim_{m\to+\infty}} \frac{1}{m} \log f(T^{-m}\omega) \leq 3\epsilon.
$$

Since ϵ is arbitrary, we obtain the wanted conclusion.

Condition A3 implies that each map $Z_n \colon \Omega \to X$ is measurable, and that the sequence $\{Z_n\}$ converges pointwise to Z with respect to the metric d, since it does with respect to the metric d_ω. We can therefore conclude that $Z \colon \Omega \to X$ is measurable (for example, see [18, Theorem 4.2.2]).

To prove the uniqueness of Z, we argue by contradiction. Suppose that there exist a T-invariant subset $\Omega_2 \subset \Omega$ of positive measure and a measurable $Y \colon \Omega_2 \to X$ satisfying conclusions (1) and (3) on Ω_2. Also, suppose that $X \neq Y$ on a measurable subset of $\Omega_1 \cap \Omega_2$ of positive measure. It follows that there exists $\eta > 0$ such that the measurable set $A = \{\omega \in \Omega_1 \cap \Omega_2 \colon d(Z(\omega), Y(\omega)) > \eta\}$ has positive measure. By the Poincaré Recurrence Theorem, there exist $\omega \in A$ and a diverging sequence of positive integers such that $T^{n_k}\omega \in A$ for every $k > 0$. Then, by applying conclusions (1) and (3) to Y and using conclusion (4), we obtain the contradiction,

$$\eta < d_{T^{n_k}\omega}\left(Z(T^{n_k}\omega), Y(T^{n_k}\omega)\right)$$
$$= d_{T^{n_k}\omega}\left(Z(T^{n_k}\omega), \phi_{n_k}(\omega, Y(\omega))\right) \xrightarrow[k\to+\infty]{} 0.$$

Thus, $X = Y$ a.e. on Ω_2, and conclusion (6) holds. $\qquad\square$

4.2. **Affine systems.** When ϕ is affine, then the map Z of Theorem 4.3 has an explicit representation. Indeed, suppose that $X_\omega = \mathbb{R}^k$ with $d_\omega = d$ being the distance generated by the Euclidean norm $\|\cdot\|$. Let E_k be the space of $k \times k$ matrices. Also, suppose that $\omega \mapsto A(\omega) \in E_k$ and $\omega \mapsto B(\omega) \in \mathbb{R}^k$ are measurable. The random dynamical system generated by $\phi(\omega, x) = A(\omega)x + B(\omega)$, $\omega \in \Omega, x \in \mathbb{R}^d$ is called *affine*. Systems of this type will appear in Section 6.1.

Proposition 4.4. *Suppose that ϕ is an affine random dynamical system satisfying*

(1) $\int \log \|A(\omega)\| \mathrm{d}\mathbb{P}(\omega) < 0$,
(2) $\overline{\lim}_{n\to\pm\infty} |n|^{-1} \log \max\{1, \|B(T^n\omega)\|\} = 0$.

Then all the conclusion of Theorem 4.3 hold for ϕ. Moreover,

$$Z(\omega) = \sum_{k=1}^{\infty} \left(\prod_{i=1}^{k-1} A(T^{-i}\omega)\right) B(T^{-k}\omega) \qquad \text{for a.e. } \omega \in \Omega.$$

Proof. By choosing $x_0(\omega) \equiv 0$, we see that Condition A3 is immediately satisfied. Since $\rho_1(\omega) = \|A(\omega)\|$, the first hypothesis of the proposition implies Condition A1. Finally, since $l(\omega) = \|B(T^{-1}(\omega))\| \leq \max\{1, \|B(T^{-1}(\omega))\|\}$, the second hypothesis of the proposition implies Condition A2. Now, the first conclusion of the proposition is just Theorem 4.3; Part (2) of Theorem 4.3 and a straighforward computation yield the second conclusion of the proposition. $\qquad\square$

5. Hilbert's projective metric

In this section, we recall the essential definitions and properties concerning Hilbert's projective metric on cones. This metric plays a key role in the proof of Theorem 3.1. For a more complete account on the subject, the reader is referred to [4, 5, 6, 28, 29, 31].

5.1. **Hilbert's metric on cones.**

Let $(\mathcal{B}, |\cdot|)$ be a real normed space with the topology induced by the norm $|\cdot|$. A subset $\mathcal{C}$ of $\mathcal{B} \setminus \{0\}$ is called a *convex cone* if $\lambda\varphi + \mu\psi \in \mathcal{C}$ for every $\varphi, \psi \in \mathcal{C}$ and every $\lambda, \mu > 0$. We say that $\mathcal{C}$ is *closed* if $\mathcal{C} \cup \{0\}$ is closed. Any cone $\mathcal{C}$ defines a partial ordering on $\mathcal{B}$ by the rule $\varphi \preceq \psi$ with $\varphi, \psi \in \mathcal{B}$ if and only if $\psi - \varphi \in \mathcal{C} \cup \{0\}$. If $\mathcal{C}$ is also closed, then $\preceq$ is continuous, i.e., if $\varphi_n, \psi \in \mathcal{B}$ such that $\psi \preceq \varphi_n$ for every $n > 0$ and $\lim_{n \to +\infty} \varphi_n = \varphi$, then $\psi \preceq \varphi$. Henceforth, a cone is always understood to be closed and convex, unless said otherwise.

We say that two elements φ, ψ of a cone $\mathcal{C}$ are *comparable* and write $\varphi \sim \psi$ if and only if $\lambda\varphi \preceq \psi \preceq \mu\varphi$ for some $\lambda, \mu > 0$. The relation $\sim$ is an equivalence relation, and the equivalence classes of $\mathcal{C}$ are called *components* of $\mathcal{C}$. We denote by $\mathcal{C}_\psi$ the component of $\mathcal{C}$ containing the element $\psi \in \mathcal{C}$. A component of $\mathcal{C}$ has all the property of the cone $\mathcal{C}$ except possibly the closedness.

The *Hilbert metric* θ on a cone $\mathcal{C}$ is defined as follows. Let

$$a(\varphi, \psi) = \sup\{\lambda > 0 : \lambda\varphi \preceq \psi\},$$
$$b(\varphi, \psi) = \inf\{\mu > 0 : \psi \preceq \mu\varphi\}.$$

Then for any pair $\varphi, \psi \in \mathcal{C}$, define

$$\theta(\varphi, \psi) = \begin{cases} \log \frac{b(\varphi,\psi)}{a(\varphi,\psi)} & \text{if } \varphi \sim \psi, \\ +\infty & \text{otherwise.} \end{cases} \tag{25}$$

It is easy to check that the restriction of θ to each component $\mathcal{C}_\psi$ of $\mathcal{C}$ with $\psi \in \mathcal{C}$ is a pseudo-metric, and the restriction of θ to the set $\{\varphi \in \mathcal{C}_\psi : |\varphi| = 1\}$ is a metric.

5.2. **Birkhoff's contraction coefficient.**

Let $(\mathcal{C}_1, \theta_1)$ and $(\mathcal{C}_2, \theta_2)$ be cones of the real normed spaces $(\mathcal{B}_1, |\cdot|_1)$ and $(\mathcal{B}_2, |\cdot|_2)$, respectively, with their respective Hilbert metrics. Suppose that $L : \mathcal{B}_1 \to \mathcal{B}_2$ is a linear transformation such that $L\mathcal{C}_1 \subset \mathcal{C}_2$. Then it can be shown that the restriction of L to $\mathcal{C}_1$ is a contraction with respect to θ_1 and θ_2 [31, Section 2.1], i.e.

$$\theta_2(L\varphi, L\psi) \leq \theta_1(\varphi, \psi) \qquad \text{for } \varphi, \psi \in \mathcal{C}_1.$$

The next result due to Birkhoff shows that L is a strict contraction if the diameter of $L\mathcal{C}_1$ is finite.

Proposition 5.1. *Suppose that $D = \sup\{\theta_2(L\varphi, L\psi) : \varphi, \psi \in \mathcal{C}_1\}$ is finite, then*

$$(26) \qquad \theta_2(L\varphi, L\psi) \leq \tanh\left(\frac{D}{4}\right) \theta_1(\varphi, \psi) \qquad \text{for } \varphi, \psi \in \mathcal{C}_1.$$

For the proof of this proposition, see [5]. The factor $\tanh(D/4)$ is called the *Birkhoff coefficient of L*.

5.3. **Cones of Hölder continuous functions.** We now introduce the family of cones used in the proof of Theorem 3.1. Let E be a connected compact smooth Riemannian manifold, and consider the Banach space $(C(E), \|\cdot\|_\infty)$. We denote by C_+ the set of strictly positive continuous functions on E, and (as in Section 3) we put C_+^α the set of strictly positive functions on E whose logarithm is Hölder continuous with exponent $\alpha > 0$.

For every $t \geq 0$, let

$$\mathcal{C}(t) = \{\varphi \in C_+(E) : \varphi(x) \leq e^{td(x,y)^\alpha}\varphi(y) \text{ for } x, y \in E\}$$

One can easily check that each $\mathcal{C}(t)$ is a cone of $(C(E), \|\cdot\|_\infty)$ and $C_+^\alpha = \bigcup_{t \geq 0} \mathcal{C}(t)$. The Hilbert projective metric θ_t on $\mathcal{C}(t)$ is given by

$$\theta_t(\varphi, \psi) = \log\left(\sup_{\substack{x \neq y \\ u \neq v}} \frac{e^{td(x,y)^\alpha}\varphi(x) - \varphi(y)}{e^{td(x,y)^\alpha}\psi(x) - \psi(y)} \cdot \frac{e^{td(u,v)^\alpha}\psi(u) - \psi(v)}{e^{td(u,v)^\alpha}\varphi(u) - \varphi(v)}\right)$$

for $\varphi, \psi \in \mathcal{C}(t)$. For a proof, see [4, Lemma 2.3].

Let $\preceq_t$ be the partial ordering on $C(E)$ generated by $\mathcal{C}(t)$. We will simply write $\preceq$ when there is no danger of ambiguity. Let $\psi \in \mathcal{C}(t)$, and define $\Sigma_\psi(t) = \{\varphi \in \mathcal{C}_\psi(t) : \|\varphi\|_1 = 1\}$. For the proof of the next lemma, see [28, Lemma 1.3].

Lemma 5.2. *For every $\psi_1, \psi_2 \in \Sigma_\psi(t)$, we have*

$$\|\varphi_1 - \varphi_2\|_1 \leq \left(e^{\theta_t(\varphi_1, \varphi_2)} - 1\right),$$

and

$$\|\varphi_1 - \varphi_2\|_\infty \leq \left(e^{\theta_t(\varphi_1, \varphi_2)} - 1\right) \min\{\|\varphi_1\|_\infty, \|\varphi_2\|_\infty\}.$$

The following proposition follows immediately from [29, Theorem 1.2].

Proposition 5.3. *$(\Sigma_\psi(t), \theta_t|_{\Sigma_\psi(t)})$ is a complete metric space for $k = 1, \infty$.*

Denote by $\mathbb{1}$ the characteristic function on E. Clearly, $\mathbb{1} \in \mathcal{C}(t)$ for every $t \geq 0$. The next lemma shows that $\mathcal{C}(s) \subset \mathcal{C}_{\mathbb{1}}(t)$ (the component of $\mathcal{C}(t)$ containing $\mathbb{1}$) for every $0 \leq s < t$.

Lemma 5.4. *Let $\varphi \in \mathcal{C}(\delta t)$ for some $t \geq 0$ and $0 < \delta < 1$. Then*

$$\theta_t(\mathbb{1}, \varphi) \leq \log \frac{1 + \delta}{1 - \delta} + (\operatorname{diam} E)^\alpha t\delta.$$

Proof. It is not difficult to check that for $z > 1$ and $0 < \delta < 1$, we have

$$1 - \delta < \frac{z - z^\delta}{z - 1} < \frac{z - z^{-\delta}}{z - 1} < 1 + \delta.$$

Using the previous inequality and the fact that $\varphi \in \mathcal{C}(\delta t)$, we obtain

$$\frac{e^{td(x,y)^\alpha}\varphi(x) - \varphi(y)}{e^{td(x,y)^\alpha} - 1} = \varphi(x)\frac{e^{td(x,y)^\alpha} - \frac{\varphi(y)}{\varphi(x)}}{e^{td(x,y)^\alpha} - 1}$$

$$= \varphi(x)\frac{e^{td(x,y)^\alpha} - e^{-t\delta d(x,y)^\alpha}}{e^{td(x,y)^\alpha} - 1}$$

$$\leq \varphi(x) \cdot \sup_{z>1}\left\{\frac{z - z^{-\delta}}{z - 1}\right\}$$

$$\leq \varphi(x)(1 + \delta) \qquad \text{for } x \neq y.$$

Similarly, we obtain

$$\frac{e^{td(u,v)^\alpha}\varphi(u) - \varphi(v)}{e^{td(u,v)^\alpha} - 1} \geq (1 - \delta)\varphi(u) \qquad \text{for } u \neq v.$$

We can now conclude that

$$\theta_t(\mathbb{1}, \varphi) = \log \sup_{\substack{x \neq y \\ u \neq v}}\left\{\frac{e^{td(x,y)^\alpha}\varphi(x) - \varphi(y)}{e^{td(x,y)^\alpha} - 1} \cdot \frac{e^{td(u,v)^\alpha} - 1}{e^{td(u,v)^\alpha}\varphi(u) - \varphi(v)}\right\}$$

$$\leq \log\left(\frac{1 + \delta}{1 - \delta} \cdot \sup_{x,u\in E}\left\{\frac{\varphi(x)}{\varphi(u)}\right\}\right)$$

$$\leq \log\frac{1 + \delta}{1 - \delta} + \log \sup_{x,u\in E} e^{t\delta d(x,u)^\alpha}$$

$$\leq \log\frac{1 + \delta}{1 - \delta} + t\delta(\text{diam } E)^\alpha.$$

$$\square$$

6. Proof of Theorem 3.1

In this section, we prove Theorem 3.1. Recall that E is a connected compact smooth Riemannian manifold (without boundaries), ϕ is a random local diffeomorphism on E, and γ is a likelihood function on E of class C^α with $0 < \alpha \leq 1$. We assume that ϕ and γ satisfy Conditions C1 and C2. The main idea of the proof is to apply Theorem 4.3 to the random dynamical system generated by the random operator $\mathcal{L}$.

6.1. **Dynamics in the space of cones.** Let $\{\mathcal{C}(t)\}_{t\geq 0}$ be the family of cones defined in Subsection 5.3.

Lemma 6.1. *Suppose that $\varphi \in \mathcal{C}(t)$ with $t \geq 0$. Then*

$$(27) \qquad \mathcal{L}(\omega)\varphi \in \mathcal{C}\left(\frac{t + b_0}{\sigma_0^\alpha} + c(\omega)\right) \qquad \text{for } \omega \in \Omega.$$

Proof. Since E is compact, the map $\phi(\omega, \cdot)$ is proper for every ω. But $\phi(\omega, \cdot)$ is also a local diffeomorphism for every ω, and so $\phi(\omega, \cdot)$ is a $N(\omega)$-covering map of E with $N(\omega) \in \mathbb{N}$ being the degree of $\phi(\omega, \cdot)$ (see [25, Proposition 4.46]). Covering maps have the *path lifting property* [25, Proposition A.77]: given a curve $\xi : [0,1] \to E$ and a point $\tilde{x} \in \phi(\omega, \cdot)^{-1}(\xi(0))$, there exists a unique curve $\tilde{\xi} : [0,1] \to E$ such that $\tilde{\xi}(0) = \tilde{x}$ and $\phi(\omega, \tilde{\xi}) = \xi$. From this property, it follows that given any two points $x, y \in E$, we can order the preimages $\{\tilde{x}_1, \ldots, \tilde{x}_{N(\omega)}\} = \phi(\omega, \cdot)^{-1}(x)$ and $\{\tilde{y}_1, \ldots, \tilde{y}_{N(\omega)}\} = \phi(\omega, \cdot)^{-1}(y)$ so that

$$(28) \qquad \sigma(\omega) d(\tilde{x}_i, \tilde{y}_i) \le d(x, y) \qquad \text{for } i = 1, \ldots, N(\omega).$$

Define $\eta(\omega, x) = \log |\det D_x \phi(\omega, x)|$ for $\omega \in \Omega$ and $x \in E$, and let $\varphi \in \mathcal{C}(t)$ for $t \ge 0$. Let $x, y \in E$. From (28) and the properties of φ, we obtain

$$
\begin{aligned}
(\mathcal{P}_\omega \varphi)(x) &= \sum_{i=1}^{N(\omega)} \exp\left[\eta(\omega, \tilde{x}_i)\right] \varphi(\tilde{x}_i) \\
&\le \sum_{i=1}^{N(\omega)} \exp\left[\eta(\omega, \tilde{y}_i) + b(\omega) d(\tilde{x}_i, \tilde{y}_i)^\alpha\right] \varphi(\tilde{y}_i) \exp\left[t d(\tilde{x}_i, \tilde{y}_i)^\alpha\right] \\
&\le \exp\left(\frac{t + b(\omega)}{\sigma(\omega)^\alpha} d(x, y)^\alpha\right) \sum_{i=1}^{N(\omega)} \exp\left[\eta(\omega, \tilde{y}_i)\right] \varphi(\tilde{y}_i) \\
&= \exp\left(\frac{t + b(\omega)}{\sigma(\omega)^\alpha} d(x, y)^\alpha\right) (\mathcal{P}_\omega \varphi)(y).
\end{aligned}
$$

The previous inequality combined with Condition C1 gives

$$\overline{\mathcal{P}}\varphi(x) \le \exp\left(\frac{t + b_0}{\sigma_0^\alpha}\right) \overline{\mathcal{P}}\varphi(y).$$

Now, the wanted conclusion follows easily from the definitions of the operator $\mathcal{L}$ and the function c. $\qquad \square$

Inclusion (27) defines an action of $\mathcal{L}$ on the cone-parameter t. We associate to this action an affine random dynamical system as follows. Recall that $\sigma_0 > 1$ by Condition C1. Pick $\sigma_0^{-\alpha} < \delta < 1$, and let $\beta = -\log(\delta \sigma_0^\alpha) < 0$. Let $q \colon \Omega \times \mathbb{R} \to \mathbb{R}$ be the map given by

$$q(\omega, t) = e^\beta t + b_0 e^\beta + c(\omega) \qquad \text{for } \omega \in \Omega \text{ and } t \in \mathbb{R}.$$

This map generates an affine random dynamical system on $\mathbb{R}$. Denote by q_n the nth iterate of q.

Proposition 6.2. *There exists a full measure T-invariant set $\Omega_1 \subset \Omega$ such that*

$$Z(\omega) := \frac{b_0 e^\beta}{1 - e^\beta} + \sum_{k=1}^\infty e^{\beta(k-1)} c(T^{-k}\omega) > 0 \qquad \text{for } \omega \in \Omega_1,$$

is the only measurable function on Ω_1 with the following properties:

(1) $Z(T^{n+1}\omega) = q(T^n\omega, Z(T^n\omega))$ *for* $\omega \in \Omega_1$ *and* $n \in \mathbb{Z}$,

(2) $\overline{\lim}_{n\to+\infty} \frac{1}{n} \log|Z(T^n\omega) - q_n(\omega, t)| \leq \beta$ *for* $\omega \in \Omega_1$ *and* $t \in \mathbb{R}$,

(3) $\overline{\lim}_{n\to\pm\infty} \frac{1}{|n|} \log Z(T^n\omega) = 0$ *for* $\omega \in \Omega_1$.

Proof. The proposition is now a corollary of Proposition 4.4. Indeed, in the notation of Proposition 4.4, we have $A(\omega) = e^\beta$ and $B(\omega) = b_0 e^\beta$. The first hypothesis of Proposition 4.4 is trivially satisfied, whereas the second one follows directly from Condition C2. $\square$

Lemma 6.3. *Let Ω_1 and Z be as in Proposition 6.2. Then for every $n > 0$, we have*

(1) $\mathcal{L}_n(\omega)\mathcal{C}(t) \subset \mathcal{C}(\delta q_n(\omega, t)) \subset \mathcal{C}_{\mathbb{1}}(q_n(\omega, t))$ *for* $\omega \in \Omega$ *and* $t \geq 0$,

(2) $\mathcal{L}_n(\omega)\mathcal{C}(Z(\omega)) \subset \mathcal{C}(\delta Z(T^n\omega)) \subset \mathcal{C}_{\mathbb{1}}(Z(T^n\omega))$ *for* $\omega \in \Omega_1$.

Proof. The lemma is an easy consequence of Lemmas 5.4 and 6.1 and Proposition 6.2. $\square$

Theorem holds for functions $\varphi \in C^\alpha_+(E)$. The next proposition allows us to reduce the proof of Theorem 3.1 to the special case when $\varphi \in \mathcal{C}(Z(\omega))$.

Proposition 6.4. *Let Ω_1 and Z be as in Proposition 6.2. Then*

(1) *for every $t \in \mathbb{R}$ and a.e. $\omega \in \Omega_1$, there exists $n > 0$ such that $\delta q_n(\omega, t) < Z(T^n\omega)$;*

(2) *for every $\varphi \in C^\alpha_+(E)$ and a.e. $\omega \in \Omega_1$, there exists an integer $n > 0$ such that $\mathcal{L}_n(\omega)\varphi \in \mathcal{C}(Z(T^n\omega))$.*

Proof. By conclusion (1) of Theorem 6.2, we can find $\epsilon > 0$ such that the set $\tilde{\Omega} = \{\omega \in \Omega_1 : Z(\omega) > \epsilon\delta/(1-\delta)\}$ has positive $\mathbb{P}$-measure. Since $(\Omega, \mathcal{F}, \mathbb{P}, T)$ is ergodic, there exists a set $\Omega_2 \subset \Omega_1$ of full measure such that for every $\omega \in \Omega_2$, we can find an increasing sequence of positive integers $\{n_k(\omega)\}_{k>0}$ with the property that $T^{n_k(\omega)}\omega \in \tilde{\Omega}$ for every $k > 0$.

Fix $\omega \in \Omega_2$ and $t \in \mathbb{R}$, and write n_k for $n_k(\omega)$. From conclusion (3) of Theorem 6.2, it follows that $q_{n_k}(\omega, t) \leq Z(T^{n_k}\omega) + \epsilon$ for k sufficiently large. Since $T^{n_k}\omega \in \tilde{\Omega}$, we obtain

$$q_{n_k}(\omega, t) \leq Z(T^{n_k}\omega) + \left(\frac{1}{\delta} - 1\right) Z(T^{n_k}\omega) \leq \frac{1}{\delta} Z(T^{n_k}\omega),$$

which proves part (1) of the proposition.

Given $\varphi \in C^\alpha_+(E)$, there exists $t > 0$ such that $\varphi \in \mathcal{C}(t)$. By Lemma 6.3, we know that $\mathcal{L}_n(\omega)\varphi \subset \mathcal{C}(\delta q_n(\omega, t))$ for $\omega \in \Omega$ and $n > 0$. Part (2) of the proposition now follows from Part (1). $\square$

6.2. **The random dynamical system generated by $\hat{\mathcal{L}}$.** Let X be the Banach space $(C(E), \|\cdot\|_\infty)$. Let Ω_1 be the set as in Proposition 6.2. Without loss of generality, we can assume that $\Omega_1 = \Omega$ and that the function Z is strictly positive everywhere on Ω. Recall that $\mathbb{1}$ is the characteristic function on E. As before, let $\mathcal{C}_{\mathbb{1}}(t)$ be the component of $\mathcal{C}(t)$ containing $\mathbb{1}$. Define

$$X_\omega = \{\varphi \in \mathcal{C}_{\mathbb{1}}(Z(\omega)) \colon \|\varphi\|_1 = 1\} = \Sigma_{\mathbb{1}}(Z(\omega)) \qquad \text{for } \omega \in \Omega.$$

Denote by θ_ω the Hilbert metric of $\mathcal{C}(Z(\omega))$. By Proposition 5.3, the metric space $(X_\omega, \theta_\omega|_{X_\omega})$ is complete. By Lemma 6.3, we have $\mathcal{L}(\omega)\mathcal{C}_{\mathbb{1}}(Z(\omega)) \subseteq \mathcal{C}_{\mathbb{1}}(Z(T\omega))$, and so $\hat{\mathcal{L}}(\omega)X_\omega \subseteq X_{T\omega}$.

Lemma 6.5. *The random dynamical system generated by $\hat{\mathcal{L}}$ satisfies Conditions A1 and A2.*

Proof. Fix $\omega \in \Omega$. By Lemma 5.4, the diameter of $\mathcal{C}(\delta Z(T\omega))$ computed with respect to the metric $\theta_{T\omega}$ is bounded above by the measurable function

$$D(\omega) := 2\log\frac{1+\delta}{1-\delta} + 2\delta\,(\operatorname{diam} E)^\alpha\, Z(T\omega).$$

Proposition 5.1 implies that $\hat{\mathcal{L}}(\omega) : (X_\omega, \theta_\omega) \to (X_{T\omega}, \theta_{T\omega})$ is a strict contraction with Lipschitz constant $\rho(\omega) \leq \tanh(D(\omega)/4)$. Since $0 < Z < +\infty$ a.e. on Ω, it is easy to check that the function $r(\omega) := \log\tanh(D(\omega)/4)$ is negative and bounded a.e. on Ω. This implies that $r \in L^1(\mathbb{P})$ and $\int r(\omega)d\mathbb{P}(\omega) < 0$. Hence, Condition A1 is satisfied.

We now prove that A2 is satisfied with $x_0 \equiv \mathbb{1}$. Note that $\mathbb{1} \in X_\omega$ for every $\omega \in \Omega$. By Lemma 5.4, we have $\theta_\omega(\mathbb{1}, \hat{\mathcal{L}}(T^{-1}\omega)\mathbb{1}) \leq D(\omega)$. So it is enough to observe that $\overline{\lim}_{n\to\pm\infty}|n|^{-1}\log D(T^n\omega) = 0$ for a.e. $\omega \in \Omega$, which follows easily from conclusion (3) of Proposition 6.2. $\qquad\square$

Lemma 6.6. *Suppose that $\varphi\colon \Omega\times E \to \mathbb{R}$ is a measurable function such that $\varphi(\omega, \cdot) \in C(E)$ for every $\omega \in \Omega$. Then the function $\omega \mapsto \|\varphi(\omega, \cdot)\|_1$ is measurable.*

Proof. The lemma follows directly from Tonelli's Theorem. $\qquad\square$

We will now conclude the proof of Theorem 3.1. By Lemma 6.5, conclusions (1)-(5) of Theorem 4.3 hold for the random dynamical system $\hat{\mathcal{L}}$. This means precisely that there exist a T-invariant set $\Omega_0 \subset \Omega$ of full measure and a function $h\colon \Omega \times E \to \mathbb{R}$ such that for $\omega \in \Omega_0$ and $\varphi \in X_\omega$, we have

(i) $h(\omega, \cdot) \in X_\omega$,
(ii) $\overline{\lim}_{n\to+\infty} n^{-1}\log\theta_\omega(h(\omega, \cdot), \hat{\mathcal{L}}_n(T^{-n}\omega)\mathbb{1}) \leq \xi$,
(iii) $\hat{\mathcal{L}}_n(\omega)h(\omega, \cdot) = h(T^n\omega, \cdot)$ for every $n \in \mathbb{Z}$,
(iv) $\overline{\lim}_{n\to+\infty} n^{-1}\log\theta_{T^n\omega}(h(T^n\omega, \cdot), \hat{\mathcal{L}}_n(\omega)\varphi) \leq \xi$,
(v) $\overline{\lim}_{n\to\pm\infty} |n|^{-1}\log\theta_{T^n\omega}(h(T^n\omega), \mathbb{1}) \leq 0$.

Conclusions (1)-(4) of the theorem are straightforward consequences of statements (i)-(iv) above. Indeed, conclusion (1) is just statement (i). By Lemma 5.2, we have

$$\|h(\omega, \cdot) - \hat{\mathcal{L}}_n(T^{-n}\omega)\mathbb{1}\|_\infty \leq (e^{\Theta_\omega(h(\omega, \cdot), \hat{\mathcal{L}}_n(T^{-n}\omega))} - 1)\|h(\omega, \cdot)\|_\infty.$$

This together with (ii) implies conclusion (2). Conclusion (5) is derived similarly from (v). To prove conclusion (3), we define $\Lambda(\omega) = \|\mathcal{L}(\omega)h(\omega, \cdot)\|_1$, and then use (iii). The proof of Conclusion (4) is similar to that of (2), and follows directly from (iv), (v), Lemma 5.2 and the second part of Proposition 6.4. Note that in (iv), it is assumed that $\varphi \in X_\omega$ with $\omega \in \Omega_0$. We can assume that this is the case even when $\varphi \in C_+^\alpha(E)$ in virtue of the second part of Proposition 6.4 and the T-invariance of Ω_0.

Let $g_n(\omega, x) = (\mathcal{L}(T^{-n}\omega)\mathbb{1})(x)$ for $\omega \in \Omega$, $x \in E$ and $n > 0$. By Lemma 2.7, each g_n is measurable, and by Lemma 6.6, $\omega \mapsto \|g_n(\omega, \cdot)\|_1$ is also measurable. Hence, $h_n(\omega, x) := g_n(\omega, x)/\|g_n(\omega, \cdot)\|_1$ is measurable. Now, by conclusion (2), the sequence $\{h_n(\omega, \cdot)\}_{n>0}$ converges uniformly to $h(\omega, \cdot)$ for every $\omega \in \Omega_0$. Hence, the restriction $h|_{\Omega_0 \times E}$ is measurable. By construction of h (see the proof of Theorem 4.3) and our choice $x_0(\omega, \cdot) = \mathbb{1}$, we also have $h(\omega, \cdot) = \mathbb{1}$ for $\Omega \setminus \Omega_0$. We conclude that h is measurable on the entire set $\Omega \times E$.

We now show that Λ is measurable. Since h is measurable, Lemma 2.7 implies that $(\omega, x) \mapsto (\mathcal{L}(\omega)h(\omega, \cdot))(x)$ is measurable. Moreover, since $h(\omega, \cdot) \in C(E)$ is continuous, $\mathcal{L}(\omega)h(\omega, \cdot)$ is continuous as well. Hence, $(\omega, x) \mapsto (\mathcal{L}(\omega)h(\omega, \cdot))(x)$ satisfies the hypotheses of Lemma 6.6, and so Λ is measurable.

It remains to show that h and Λ are unique. The proof is the same as the one of the uniqueness of Z in Theorem 4.3, and so we omit it. Finally, we remark that if Z is bounded in ω (for instance if c is bounded), then the converge in Theorem 3.1 is uniform in ω.

Appendix A. Disintegrations and their Radon–Nikodym derivatives

In this paper, we consider mappings and measures on product spaces $\Omega \times E$, where E is metric. The following facts prove useful in this context.

Lemma A.1. *Suppose E is polish space, and let $\mathcal{B}$ be the Borel algebra on E.*

(1) *Let $\mathbb{Q}$ be a probability measure on $(\Omega, \mathcal{A}) \times (E, \mathcal{B})$ with marginal $\mathbb{P}$ on Ω. There exists a regular conditional probability μ_ω which disintegrates $\mathbb{Q}$. That is, μ_ω is a probability measure on $(E, \mathcal{B})$ for every ω, the function $\omega \to \mu_\omega(B)$ is measurable for every*

$B \in \mathcal{B}$, *and*

$$\mathbb{Q}(A, B) = \int_A \mu_\omega(B) \, \mathrm{d}\mathbb{P}.$$

(2) If μ_ω disintegrates $\mathbb{Q}$ and $\varphi \in L^1(\mathbb{Q})$, then the mapping

$$\mu_\omega(\varphi_\omega) \colon \omega \to \int \varphi(\omega, x) \, \mathrm{d}\mu_\omega(x)$$

is in $L^1(\mathbb{P})$, and

$$\int \varphi \mathrm{d}\mathbb{Q} = \int \mu_\omega(\varphi_\omega) \, \mathrm{d}\mathbb{P}.$$

(3) *If μ_ω is regular and $\mu_\omega \ll \lambda$ for every ω with some σ–finite measure λ, then there is a $\mathcal{A} \times \mathcal{B}$–measurable version of $q(\omega, x) = \frac{\mathrm{d}\mu_\omega}{\mathrm{d}\lambda}(x)$.*
(4) *If F is metric and $\varphi \colon \Omega \times E \to F$ is so that $\varphi(., x)$ is measurable for every $x \in E$ and $\varphi(\omega, .)$ is continuous for every ω, then there is a version of $\varphi \colon \Omega \times E \to F$ jointly measurable in (ω, x).*

For a proof of (1), see [20], p.196. Item (2) is an easy consequence of the monotone convergence theorem. Item (3) is a known fact; for convenience of the reader, a proof is given below. Item (4) can be found in [13], Lemma 3.14. Note that for every $A \in \mathcal{A}$,

$$\mathbb{P}(A) = \mathbb{Q}(A, E) = \int_A \mu_\omega(E) \, \mathrm{d}\mathbb{P}.$$

whence $\mu_\omega(E) = 1$ a.s.

Proof. (of Item 3) A very similar statement is demonstrated in the proof of theorem 3.1 in [30] for λ a probability measure, see also [17], Ch. XI, Sec. 17. At first, let λ be a probability measure, and consider the probability space $(E, \mathcal{B}, \lambda)$ with expectation $\mathbb{M}$. Clearly, $q(\omega, \cdot)$ is in L^1 for every ω and $\omega \to \mathbb{M}(\mathbb{1}_B(\cdot)q(\omega, \cdot)) = \mu_\omega(B)$ is measurable for every $B \in \mathcal{B}$. Let $B_1, B_2, \ldots$ be the countable basis for $\mathcal{B}$ and put $\mathcal{B}_n = \sigma(B_1, \ldots, B_n)$. The function $(\omega, x) \to q_n(\omega, x) = \mathbb{M}(q(\omega, \cdot)|\mathcal{B}_n)(x)$ is easily seen to be measurable as it can be explicitely constructed using a finite disjoint $\mathcal{B}_n$–measurable partition of E. But $q_n(\omega, \cdot)$ is a L^1–martingale for every ω, so by martingale convergence $\lim_n q_n(\omega, x) = q(\omega, x)$ for any x not in some set N_ω with $\lambda(N_\omega) = 0$. This means that the measurable set N of points (ω, x) where q_n does not converge has ω–sections of zero measure, and Fubini implies $(\lambda \times \mathbb{P})(N) = 0$, that is, q_n converges $\lambda \times \mathbb{P}$–almost everywhere. Hence q is measurable. The extension to σ–finite λ is analogous to the corresponding statement for Radon–Nikodym derivatives. $\square$

Acknowledgements

References

[1] L. Arnold, *Random dynamical systems*, Springer Monographs in Mathematics, Springer, Berlin, 1998.

[2] R. Atar and O. Zeitouni, Exponential stability for nonlinear filtering, Ann. Inst. H. Poincaré Probab. Statist. **33** (1997), no. 6, 697–725.

[3] R. Atar and O. Zeitouni, Lyapunov exponents for finite state nonlinear filtering, SIAM J. Control Optim. **35** (1997), no. 1, 36–55.

[4] V. Baladi, *Positive transfer operators and decay of correlations*, Advanced Series in Nonlinear Dynamics, 16, World Sci. Publishing, River Edge, NJ, 2000.

[5] G. Birkhoff, Extensions of Jentzsch's theorem, Trans. Amer. Math. Soc. **85** (1957), 219–227.

[6] G. Birkhoff, *Lattice theory*, Third edition. American Mathematical Society Colloquium Publications, Vol. XXV, Amer. Math. Soc., Providence, RI, 1967.

[7] T. Bogenschütz and V. M. Gundlach, Ruelle's transfer operator for random subshifts of finite type, Ergodic Theory Dynam. Systems **15** (1995), no. 3, 413–447.

[8] P. Bougerol, Kalman filtering with random coefficients and contractions, SIAM J. Control Optim. **31** (1993), no. 4, 942–959.

[9] Leo Breiman. *Probability*. Addison-Wesley, Reading, Mass, 1973.

[10] A. Budhiraja and H. J. Kushner, Robustness of nonlinear filters over the infinite time interval, SIAM J. Control Optim. **36** (1998), no. 5, 1618–1637.

[11] A. Budhiraja and D. Ocone, Exponential stability of discrete-time filters for bounded observation noise, Systems Control Lett. **30** (1997), no. 4, 185–193.

[12] A. Budhiraja and D. Ocone, Exponential stability in discrete-time filtering for non-ergodic signals, Stochastic Process. Appl. **82** (1999), no. 2, 245–257.

[13] C. Castaing and M. Valadier, *Convex analysis and measurable multifunctions*, Lecture Notes in Mathematics, Vol. 580, Springer, Berlin, 1977.

[14] P. Chigansky, R. Liptser and R. Van Handel, Intrinsic methods in filter stability, in *The Oxford handbook of nonlinear filtering*, 319–351, Oxford Univ. Press, Oxford.

[15] P. Chigansky, *Introduction to nonlinear filtering*, Lecture Notes, 2005, Unpublished, `http://pluto.huji.ac.il/~pchiga/teaching/Filtering/nlf.html`

[16] P. Chigansky, private communication.

[17] J. L. Doob, *Measure theory*, Graduate Texts in Mathematics, 143, Springer, New York, 1994.

[18] R. M. Dudley, *Real analysis and probability*, revised reprint of the 1989 original, Cambridge Studies in Advanced Mathematics, 74, Cambridge Univ. Press, Cambridge, 2002.

[19] P. Ferrero and B. Schmitt, Produits aléatoires d'opérateurs matrices de transfert, Probab. Theory Related Fields **79** (1988), no. 2, 227–248.

[20] P. Gänssler and W. Stute, *Wahrscheinlichkeitstheorie*, Springer, Berlin, 1977.

[21] K. Heine and D. Crisan, Uniform approximations of discrete-time filters, Adv. in Appl. Probab. **40** (2008), no. 4, 979–1001.

[22] K. Khanin and Y. Kifer, Thermodynamic formalism for random transformations and statistical mechanics, in *Sinai's Moscow Seminar on Dynamical Systems*, 107–140, Amer. Math. Soc. Transl. Ser. 2, 171, Amer. Math. Soc., Providence, RI.

[23] Y. Kifer, Equilibrium states for random expanding transformations, Random Comput. Dynam. **1** (1992/93), no. 1, 1–31.

[24] Y. Kifer, Thermodynamic formalism for random transformations revisited, Stoch. Dyn. **8** (2008), no. 1, 77–102.

[25] J. M. Lee, *Introduction to smooth manifolds*, Graduate Texts in Mathematics, 218, Springer, New York, 2003.

[26] F. LeGland and N. Oudjane, A robustification approach to stability and to uniform particle approximation of nonlinear filters: the example of pseudo-mixing signals, Stochastic Process. Appl. **106** (2003), no. 2, 279–316.

[27] F. Le Gland and N. Oudjane, Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters, Ann. Appl. Probab. **14** (2004), no. 1, 144–187.

[28] C. Liverani, Decay of correlations, Ann. of Math. (2) **142** (1995), no. 2, 239–301.

[29] R. D. Nussbaum, Iterated nonlinear maps and Hilbert's projective metric. II, Mem. Amer. Math. Soc. **79** (1989), no. 401, iv+118 pp.

[30] D. Orstein and L. Sucheston, An operator theorem on L_1 convergence to zero with applications to Markov kernels, Ann. Math. Statist. **41** (1970), 1631–1639.

[31] M. Viana, *Stochastic dynamics of deterministic systems*, Colóquio Brasileiro de Matemática, 21, IMPA, Rio de Janeiro, 1997.

[32] Gary Froyland, Cecilia González-Tokman, and Anthony Quas, Stability and approximation of random invariant densities for Lasota–Yorke map cocycles, Nonlinearity, **27** (2014), no. 4.

[33] Jérôme Buzzi, Exponential decay of correlations for random Lasota–Yorke maps, Communications in Mathematical Physics, **208** (1999), no. 1, 25–54.

[34] Jochen Bröcker and Gianluigi Del Magno, Invariant measures for random transformations expanding on average, Publ. Math. Uruguay, **14** (2013), 78–104.

DEPARTMENT OF MATHEMATICS AND STATISTICS, SCHOOL OF MATHEMATICAL AND PHYSICAL SCIENCES, WHITEKNIGHTS, PO BOX 220, READING RG6 6AX, UNITED KINGDOM
  E-mail address: `j.broecker@reading.ac.uk`

UNIVERSIDADE FEDERAL DA BAHIA, INSTITUTO DE MATEMÁTICA, AVENIDA ADHEMAR DE BARROS, ONDINA, 40170110 - SALVADOR, BA - BRASIL
  E-mail address: `gdelmagno@ufba.br`