AN ABSTRACT CONTINUITY THEOREM FOR
LYAPUNOV EXPONENTS OF LINEAR COCYCLES
AND AN APPLICATION TO RANDOM COCYCLES

PEDRO DUARTE AND SILVIUS KLEIN

Abstract. We devise an abstract scheme to prove continuity of
the Lyapunov exponents for a general class of linear cocycles. The
main assumption is a large deviation type (LDT) estimate, uniform
in the data. We provide a modulus of continuity that depends
explicitly on the sharpness of the LDT estimate. Moreover, we
derive such estimates for a class of irreducible random cocycles of
Markovian type, thus proving Hölder continuity of their Lyapunov
exponents.
This is a preliminary version.

1. Introduction

A linear cocycle is the dynamical system underlying a skew-product
map acting on a vector bundle. The base dynamics is given by an
ergodic transformation, while the action on the fiber is given by a ma-
trix valued function with a certain regularity. Lyapunov exponents are
quantities that measure the average exponential growth of the iterates
of the cocycle along invariant subspaces of the fibers. These invariant
subspaces define a filtration of each fiber, which depends measurably on
the base point. The existence of both Lyapunov exponents and the fil-
tration, which is called the Oseledets filtration, are guaranteed to hold
under mild integrability conditions on the cocycle by the celebrated
Oseledets multiplicative ergodic theorem (see [1, 26]).

Two classes of general linear cocycles have been extensively studied
so far: the class of random cocycles, where the base dynamics is a
Bernoulli shift, and the class of quasiperiodic cocycles, where the base
dynamics is a torus translation.

In this paper we study continuity properties of the Lyapunov expo-
nents for a general class of linear cocycles. The base dynamics will be
fixed. Hence we identify the cocycle with the matrix valued function
that determines its fiber action. Continuity is meant as a function of
the cocycle. Our method applies to both random and quasiperiodic
cocycles giving quantitative results, i.e. a modulus of continuity under appropriate further conditions.

The main assumptions required by the method described in this paper are the availability of large deviation type (LDT) estimates on the base dynamics for a rich enough class of observables, and the existence of uniform LDT estimates on the fiber dynamics. Large deviations in classical probabilities or for multiplicative systems associated to a dynamical system describe the asymptotic behavior of tail events in terms of a rate function. We require a somewhat different type of large deviations. Instead of a precise rate, only a good estimate on the decay of the tail event is needed. However, we require that such estimates hold for all iterates of the system, after a certain threshold, and that in the case of the fiber dynamics, this threshold of applicability as well as various other parameters describing the LDT estimates are stable under small perturbations of the cocycle, a property we refer to as uniform fiber LDT estimates.

Kingman’s subadditive ergodic theorem allows us to describe Lyapunov exponents as limits, when the number \( n \) of iterates grows, of finite scale Lyapunov exponents, which are defined as the phase space average of quantities related to the singular values of the \( n \)th iterate of the cocycle.

The mechanism for obtaining quantitative continuity properties of the limit objects (i.e. the Lyapunov exponents) is a deterministic result called the Avalanche Principle. Roughly speaking, the avalanche principle (AP) allows us to relate singular values of a long block (i.e. product) of matrices to certain averages of singular values of individual components of the block. This holds provided certain geometric conditions (which we call "gaps" and "angles" conditions) on the individual components are satisfied. The gap condition means that a pattern on the relative sizes of consecutive singular values holds uniformly for all elements of the block, while the angle condition ensures that most expanding singular directions of consecutive elements of the block are not almost orthogonal, hence they are not canceling each other out.

In order to effectively apply the AP to long blocks made up of iterates of a cocycle, the geometric conditions need to be satisfied for a large enough set of phases. This is where the LDT estimates on the fiber action are used, as they turn estimates on finite scale Lyapunov exponents, which are phase space averages, into pointwise estimates which hold for a large number of phases, and correspondingly they imply the geometric conditions of the AP for that large set of phases.

The sharpness of the LDT determines how long a block of matrices in the AP can be, before running out of phases satisfying the geometric
conditions. This argument is then used repeatedly, in an inductive procedure, where the previous long block becomes a typical component of the next much larger block, and the LDT estimate is used again to guarantee the geometric conditions for sufficiently many phases, and hence the applicability of the AP in the next stage of the induction.

This method of proving continuity of Lyapunov exponents was first introduced by M. Goldstein and W. Schlag in [13] in the context of quasiperiodic, analytic Schrödinger cocycles. These types of cocycles form a one-parameter family of SL(2, R)-valued cocycles associated to a discrete, one-dimensional quasiperiodic Schrödinger operator with analytic potential. The base dynamics is a torus translation by a Diophantine frequency number (or vector, for a higher dimensional torus) and the family is indexed by the energy parameter. The authors also prove a (sharp, for one dimensional torus translations) LDT estimate for these cocycles and obtain Hölder (or log-Hölder, for higher dimensional torus translations) continuity of the Lyapunov exponent as a function of the energy parameter, under the assumption of a uniform positive lower bound on the Lyapunov exponents.

A higher dimensional version of the AP, along with a higher dimensional version of the result in [13], were obtained in [27] for Schrödinger-like cocycles, under the restrictive assumption that all Lyapunov exponents are simple. It was also indicated in this paper that this method is in some sense modular, a statement that motivated in part our present work.

The AP in [13] along with various versions of LDT estimates have been widely used for problems related to Lyapunov exponents of quasiperiodic cocycles or to problems on the spectral theory of discrete, quasiperiodic Schrödinger operators (see for instance [6, 14, 16, 20, 10] and references therein).

We should also mention a related problem, that of joint continuity, in frequency and cocycle, of the Lyapunov exponent, which is treated in [8, 7, 17, 2].

Our work in [10] presents a geometric, conceptual approach to the Avalanche Principle, which allows us to generalize it to higher dimensions, namely to blocks of GL(m, R) matrices. We use this general AP in [10] to prove Hölder (or log-Hölder for multifrequency translations) continuity of the Lyapunov exponents of GL(m,R)-valued quasiperiodic, analytic cocycles in a neighborhood of a cocycle with simple Lyapunov exponents, and simple continuity everywhere else.
At the other end of the type of ergodic behavior of the base dynamics - the random case, continuity results for linear cocycles over Bernoulli shifts in the generic case go back to H. Furstenberg and Kifer, see [12].

E. Le Page proved in [24] Hölder continuity of the top Lyapunov exponent for a one-parameter family of cocycles over the Bernoulli shift, under irreducibility and contraction assumptions, which are assumed to hold uniformly throughout this family.

C. Bocker-Neto and M. Viana [3] proved continuity of the Lyapunov exponents for two-dimensional cocycles over Bernoulli shifts without any irreducibility assumptions (the result does not provide a modulus of continuity though). A higher dimensional version of this result was announced by A. Ávila, A. Eskin and M. Viana (see the monograph [28]). An extension of results from [3] to a particular type of cocycles over Markov systems (particular in the sense that the cocycle still depends on one coordinate, as in the Bernoulli case) was obtained in [23].

We note, for the interested reader, that a general one-stop reference for continuity results for random cocycles is M. Viana’s monograph [28].

The first goal of this paper is to present an abstract, modular approach to proving continuity of the Lyapunov exponents, which uses an inductive procedure based on the deterministic, general Avalanche Principle. The main advantage of this approach, besides the fact that it provides quantitative estimates, is its versatility.

This approach applies to both quasiperiodic and random cocycles (or to any other types of base dynamics) as long as appropriate LDT estimates are satisfied. This is the context of theorem 2.1.

Most existing results regarding quasiperiodic, analytic cocycles already fit this approach, so they can be obtained as a consequence of theorem 2.1. Moreover, in forthcoming papers we will also consider within this scheme, quasiperiodic cocycles with singularities (i.e. not $\text{GL}(m, \mathbb{R})$ valued). It should be noted that while we do require Diophantine translations, and the frequency is fixed, this method applies equally to translations on the one or the higher dimensional torus, which is not possible with the methods in [2].

The second goal of this paper is to establish LDT estimates for a general class of random cocycles under a certain irreducibility assumption. Apart from this technical hypothesis, the approach is very general and works for cocycles with finite memory over mixing Markov systems of finite order. A Laplace-Markov operator can be associated to each
cocycle over a Markov system. If the system is mixing, and the cocycle irreducible, then the Laplace-Markov operator acts quasi-compactly on some appropriate Banach algebra of measurable functions, having a single eigenvalue of maximum absolute value. The large deviation estimates are a by-product of a control on the top eigenvalue of the Laplace-Markov operator. The basic ideas of the proof are due to Le Page [24], [22], who proved central limit theorems, as well as the continuity of the top Lyapunov exponent for random i.i.d. cocycles, and to Bougerol [4] who extended Le Page’s work to cocycles over mixing Markov systems. See also the books [5] and [15]. All large deviation theorems in the previous references are asymptotic statements, somewhat stronger than the large deviation estimates we require, but also not providing the kind of uniformity in the cocycle that is essential to apply the abstract theorem 2.1. This requirement forced us to dwell on the subject, following closely a general framework introduced in [15] to address this type of problem. The necessity of the irreducibility assumption is explained in more detail at the end of section 2. The continuity of the Lyapunov exponents is a corollary of the abstract theorem 2.1, where the given modulus of continuity corresponds in this case to Hölder continuity.

Comparing with the continuity theorem of Le Page [24], our results do not require any contraction assumption and provide the continuity at all exponents, regardless of the gaps in the Lyapunov spectrum. The statement is about continuity in the space of irreducible cocycles and not just for one-parameter families. It is also more general since we address cocycles over mixing Markov shifts, and not just over the Bernoulli shifts.

This paper is organized as follows.

Section 2 contains a description of the abstract setting of our continuity result and its formulation. It then describes the class of irreducible cocycles over a Markov system, the LDT estimate for them and its consequence, via the abstract result, to the continuity of the Lyapunov exponents.

Section 3 presents an extension of the Avalanche Principle in [10] to noninvertible matrices.

Section 4 contains a technical estimate that will ensure that the gap condition in the AP holds for a large set of phases. It is a crucial ingredient in our argument, as it allows us to handle systems (e.g. Markov shifts) where large deviation type estimates are available only in the presence of a gap in the Lyapunov spectrum (which can be interpreted...
as a projective contraction property). This estimate follows from a type of uniform upper semicontinuity of the top Lyapunov exponent of a linear cocycle, a result that can be of independent interest.

Section 5 contains the proof of the main abstract statement regarding Lyapunov exponents and of its extensions. It also contains finite scale uniform estimates that are of independent interest.

Section 6 contains a LDT estimate theorem in some abstract functional analytic context, which specializes a more general framework in [15].

Section 7 proves base and fibre LDT estimate theorems for measurable spaces of irreducible cocycles over a Markov system, reducing them to the abstract theorem of the previous section.

2. Definitions and statements

An ergodic dynamical system $(X, \mathcal{F}, \mu, T)$ consists of a set $X$, a $\sigma$-algebra $\mathcal{F}$, a probability measure $\mu$ on $(X, \mathcal{F})$ and a transformation $T: X \to X$ which is ergodic and measure preserving.

Two important classes of ergodic dynamical systems are the shift over a stochastic process (i.e. a Bernoulli shift or a Markov shift) and the torus translation by an incommensurable frequency vector.

A linear cocycle over an ergodic system $(X, \mathcal{F}, \mu, T)$ is a skew-product map on the vector bundle $X \times \mathbb{R}^m$ given by

$$X \times \mathbb{R}^m \ni (x, v) \mapsto (Tx, A(x)v) \in X \times \mathbb{R}^m$$

hence $T$ is the base dynamics while $A$ defines the fiber action. Since the base dynamics will be fixed throughout this paper, we identify the cocycle with just its fiber action $A$.

The iterates of the cocycle are $(T^n x, A^{(n)}(x) v)$, where

$$A^{(n)}(x) = A(T^{n-1} x) \cdot \ldots \cdot A(T x) \cdot A(x)$$

Given an ergodic system $(X, \mathcal{F}, \mu, T)$, we introduce the main actors - a space of cocycles and a set of observables - and we describe the main assumptions on them - certain uniform large deviation type (LDT) estimates and a uniform $L^p$-boundedness. Then we formulate an abstract criterion for the continuity of the corresponding Lyapunov exponents.

Finally, we describe the Markov shift, the corresponding space of cocycles and some general assumptions which ensure that this model falls within the scope of the general criterion - namely that it satisfies appropriate uniform LDT estimates - which allows us to conclude that the corresponding Lyapunov exponents are continuous functions of the cocycle.
Cocycles and observables.

**Definition 2.1.** A space of measurable cocycles $C$ is any class of matrix valued functions $A : X \to \text{Mat}_m(\mathbb{R})$, where $m \in \mathbb{N}$ is not fixed, such that every $A : X \to \text{Mat}_m(\mathbb{R})$ in $C$ has the following properties:

1. $A$ is $\mathcal{F}$-measurable
2. $\|A\| \in L^\infty(\mu)$
3. The forward iterates $A^{(n)}$, $n \geq 1$ are in $C$
4. The exterior powers $\wedge_k A : X \to \text{Mat}_{(m)}(\mathbb{R})$ are in $C$, for $k \leq m$.

Each subspace $C_m := \{ A \in C \mid A : X \to \text{Mat}_m(\mathbb{R}) \}$ is a-priori endowed with a distance $\text{dist} : C_m \times C_m \to [0, +\infty)$ which is at least as fine as the $L^\infty$ distance, i.e. for all $A, B \in C_m$ we have

$$\text{dist}(B, A) \geq \|B - A\|_{L^\infty}$$

We assume a correlation between the distances on each of these subspaces, in the sense that the map

$$C_m \ni A \mapsto \wedge_k A \in C_{(m)}$$

is locally Lipschitz.

Let $A \in C$ be a measurable cocycle. Since $\|A\| \in L^\infty$, we have $\log^+\|A\| \in L^1$, hence Furstenberg-Kesten’s theorem (the non-invertible, one-sided case, see chapter 3 in [1] applies. In particular, if we denote

$$L_1^{(n)}(A) := \int_X \frac{1}{n} \log \|A^{(n)}(x)\| \mu(dx)$$

then as $n \to \infty$, $L_1^{(n)}(A) \to L_1(A)$ (the top Lyapunov exponent).

We call $L_1^{(n)}(A)$ finite scale (top) Lyapunov exponents.

We will need stronger integrability assumptions on the measurable functions $\frac{1}{n} \log \|A^{(n)}(x)\|$. Let $1 \leq p \leq \infty$.

**Definition 2.2.** A cocycle $A \in C$ is called $L^p$ - bounded if there is $C < \infty$, which we call its $L^p$-bound, such that for all $n \geq 1$ we have:

$$\left\| \frac{1}{n} \log \|A^{(n)}(\cdot)\| \right\|_{L^p} < C$$

**Definition 2.3.** A cocycle $A \in C_m$ is called uniformly $L^p$ - bounded if there are $\delta = \delta(A) > 0$ and $C = C(A) < \infty$ such that for all $B \in C_m$ with $\text{dist}(B, A) < \delta$ and for all $n \geq 1$ we have:

$$\left\| \frac{1}{n} \log \|B^{(n)}(\cdot)\| \right\|_{L^p} < C$$
It is not difficult to show that if a cocycle $A \in \mathcal{C}$ satisfies the bounds

$$\|\log \|A^\pm\|\|_{L^p} \leq C < \infty$$

then for all $n \in \mathbb{Z}$ we have

$$\|\log \|A^{(n)}\|\|_{L^p} \leq C|n| \quad (2.4)$$

Hence if we assume that

$$\log \|A^\pm\| \in L^p \quad (2.5)$$

for all cocycles $A \in \mathcal{C}$, and if we endow $\mathcal{C}_m$ with the distance given by

$$\text{dist}_p(A, B) := \|A - B\|_{L^\infty} + \|\log \|A^{-1}\| - \log \|B^{-1}\|\|_{L^p}$$

when $1 \leq p < \infty$, and by

$$\text{dist}_\infty(A, B) := \|A - B\|_{L^\infty}$$

when $p = \infty$, then every cocycle $A \in \mathcal{C}$ is uniformly $L^p$-bounded.

The application we have in this paper to random cocycles assumes from the beginning the integrability condition (2.5), so uniform $L^p$-boundedness is automatic. However, we want that our scheme be applicable also to cocycles that are very singular (i.e. non-invertible everywhere), which is the case of a forthcoming paper on quasiperiodic cocycles, and that is why we make the weaker, uniform $L^p$-boundedness assumption.

Given a cocycle $A \in \mathcal{C}$ and an integer $N \in \mathbb{N}$, denote by $\mathcal{F}_N(A)$ the algebra generated by the sets $\{x \in X : \|A^{(n)}(x)\| \leq c\}$ or $\{x \in X : \|A^{(n)}(x)\| \geq c\}$ where $c \geq 0$ and $0 \leq n \leq N$.

Let $\Xi$ be a set of measurable functions $\xi : X \to \mathbb{R}$, which we call observables. Let $A \in \mathcal{C}$.

**Definition 2.4.** We say that $\Xi$ and $A$ are compatible if for every integer $N \in \mathbb{N}$, for every set $F \in \mathcal{F}_N(A)$ and for every $\epsilon > 0$, there is an observable $\xi \in \Xi$ such that:

$$1_F \leq \xi \quad \text{and} \quad \int_X \xi \, d\mu \leq \mu(F) + \epsilon \quad (2.6)$$

**Large deviation type estimates.** As mentioned in the introduction, the main tools in our results are some appropriate large deviations type (LDT) estimates for the given dynamical systems (meaning the base and the fiber dynamics). A LDT estimate for the base dynamics says that given an observable $\xi : X \to \mathbb{R}$, we have

$$\mu \{x \in X : \left| \frac{1}{n} \sum_{j=0}^{n-1} \xi(T^j x) - \int_X \xi \, d\mu \right| > \epsilon \} < \iota(n, \epsilon)$$
where $\epsilon = o(1)$ and $i(n, \epsilon) \to 0$ (as $n \to \infty$) represent, respectively, the size of the deviation from the mean and the measure of the deviation set. The above inequality should hold for all integers $n \geq n_0(\xi, \epsilon)$.

In classical probabilities, when dealing with i.i.d. random variables, large deviations are precise asymptotic statements, and the measure of the deviation set decays exponentially. For our purposes, and for the given dynamical systems, we need slightly different types of estimates (not as precise, but for all iterates of the system and satisfying some uniformity properties). Moreover, in some of our applications (e.g. to certain types of quasi-periodic cocycles), the available decay of the measure of the deviation set is not exponential in the number of iterates, but slower than exponential. This is the motivation behind the following formalism.

From now on, $\xi, \psi: (0, \infty) \to (0, \infty)$ will represent functions that describe respectively, the size of the deviation from the mean and the measure of the deviation set. We assume that the deviation size functions $\xi(t)$ are non-increasing. We assume that the deviation set measure functions $\psi(t)$ are continuous and strictly decreasing to 0, as $t \to \infty$, at least like a power and at most like an exponential, in other words we assume that:

$$\log t \lesssim \log \frac{1}{\psi(t)} \lesssim t \quad \text{as } t \to \infty$$

Denote by $\phi_{\psi}(t)$ the inverse of the map $t \mapsto \psi_{\psi}(t) := t [\psi(t)]^{1/2}$. We then also assume that the increasing function $\phi_{\psi}(t)$ does not grow too fast, or more precisely that:

$$\lim_{t \to \infty} \frac{\phi_{\psi}(2t)}{\phi_{\psi}(t)} < 2$$

We will use the notation $\epsilon_n := \epsilon(n)$ and $\psi_n := \psi(n)$ for integers $n$.

In the applications we have thus far, the deviation size functions are either constant functions $\xi(t) \equiv \epsilon$ for some $0 < \epsilon \ll 1$ or powers $\xi(t) \equiv t^{-a}$ for some $a > 0$, while deviation set measure functions are exponentials $\psi(t) \equiv e^{-ct}$ for some $c > 0$ or sub-exponentials $\psi(t) \equiv e^{-ct/b}$ or $\psi(t) \equiv e^{-ct/(\log t)^b}$.

Let $E$ and $I$ be some spaces of functions, with $E$ containing deviation size functions $\xi(t)$ and $I$ containing deviation set measure functions $\psi(t)$. We assume that $I$ is a convex cone, i.e., the functions $a \cdot \xi(t)$ and $\xi_1(t) + \xi_2(t)$ belong to $I$ for any $a > 0$ and $\xi_1, \xi_2 \in I$. Let $P$ be a set of triplets $p = (n_0, \xi, \psi)$, where $n_0 \in \mathbb{N}$, $\xi \in E$ and $\psi \in I$. An element $p \in P$ is referred to as a LDT parameter. Our set of LDT parameters $P$ should satisfy the condition: for all $\epsilon > 0$ there is $p = p(\epsilon) = (n_0, \xi, \psi) \in P$
such that $\epsilon_{n_0} \leq \epsilon$. This says that $\mathcal{P}$ contains LDT parameters with arbitrarily small deviation size functions.

We now define the base and fiber LDT estimates, which are relative to given spaces of deviation functions $\mathcal{E}, \mathcal{I}$ and set of parameters $\mathcal{P}$.

**Definition 2.5.** An observable $\xi: X \to \mathbb{R}$ satisfies a base-LDT estimate if for every $\epsilon > 0$ there is $p = p(\xi, \epsilon) \in \mathcal{P}$, $p = (n_0, \epsilon, \iota)$, such that for all $n \geq n_0$ we have

$$\mu \{ x \in X : \frac{1}{n} \sum_{j=0}^{n-1} \xi(T^j x) - \int_X \xi \, d\mu \leq \epsilon \} < \iota_n \quad (2.7)$$

We note that in the applications we have so far, when deriving a base-LDT for an observable $\xi$, the same will hold for $-\xi$, hence in fact $2.7$ becomes

$$\mu \{ x \in X : \frac{1}{n} \sum_{j=0}^{n-1} \xi(T^j x) - \int_X \xi \, d\mu \geq \epsilon \} < \iota_n$$

**Definition 2.6.** A measurable cocycle $A \in \mathcal{C}$ satisfies a fiber-LDT estimate if for every $\epsilon > 0$ there is $p = p(A, \epsilon) \in \mathcal{P}$, $p = (n_0, \epsilon, \iota)$, such that for all $n \geq n_0$ we have $\epsilon_n \leq \epsilon$ and

$$\mu \{ x \in X : \frac{1}{n} \log \|A^{(n)}(x)\| - L_1^{(n)}(A) \geq \epsilon \} < \iota_n \quad (2.8)$$

We will need a stronger form of the fiber-LDT, one that is uniform in a neighborhood of the cocycle, in the sense that estimate (2.8) holds with the same LDT parameter for all nearby cocycles.

**Definition 2.7.** A measurable cocycle $A \in \mathcal{C}_m$ satisfies a uniform fiber-LDT if for all $\epsilon > 0$ there are $\delta = \delta(A, \epsilon) > 0$ and $p = p(A, \epsilon) \in \mathcal{P}$, $p = (n_0, \epsilon, \iota)$, such that if $B \in \mathcal{C}_m$ with $\text{dist}(B, A) < \delta$ and if $n \geq n_0$ then $\epsilon_n \leq \epsilon$ and

$$\mu \{ x \in X : \frac{1}{n} \log \|B^{(n)}(x)\| - L_1^{(n)}(B) \geq \epsilon \} < \iota_n$$

**Abstract continuity statements.**

**Theorem 2.1.** Consider an ergodic system $(X, \mathcal{F}, \mu, T)$, a space of measurable cocycles $\mathcal{C}$, a set of observables $\Xi$, a set of LDT parameters $\mathcal{P}$ with corresponding spaces of deviation functions $\mathcal{E}, \mathcal{I}$, let $1 < p \leq \infty$ and assume the following:

1. $\Xi$ is compatible with every cocycle $A \in \mathcal{C}$.
2. Every observable $\xi \in \Xi$ satisfies a base-LDT.
3. Every $A \in \mathcal{C}$ with $L_1(A) > -\infty$ is uniformly $L^p$-bounded.
(4) Every cocycle \( A \in \mathcal{C} \) for which \( L_1(A) > L_2(A) \) satisfies a uniform fiber-LDT.

Then all Lyapunov exponents \( L_k : \mathcal{C} \to [-\infty, \infty), \; 1 \leq k \leq m, \; m \in \mathbb{N} \) are continuous functions of the cocycle.

Moreover, given \( A \in \mathcal{C} \) and \( 1 \leq k \leq m \), if the Lyapunov exponent \( L_k(A) \) is simple, then locally near \( A \) the map \( L_k \) has a modulus of continuity \( \omega(h) := \frac{\xi(c \log \frac{1}{h})}{1 - \frac{1}{p}} \) for some \( \xi = \xi(A) \in \mathcal{I} \) and \( c = c(A) > 0 \).

In section 5, where we present the proof of this theorem, we give a more general formulation of this quantitative continuity result, one that applies to any multiplicity patterns in the Lyapunov spectrum of a cocycle.

**Applications to random cocycles.** Let \( \Sigma \) be a compact metric space, \( \mathcal{F} \) be its Borel \( \sigma \)-field, and \( K \) a Markov kernel on \((\Sigma, \mathcal{F})\), according to the following definition.

**Definition 2.8.** A function \( K : \Sigma \times \mathcal{F} \to [0, 1] \) such that

1. for every \( x \in \Sigma, \; A \mapsto K(x, A) \) is a probability measure in \( \Sigma \), also denoted by \( K_x \),
2. for every \( A \in \mathcal{F}, \) the function \( x \mapsto K(x, A) \) is \( \mathcal{F} \)-measurable,

is called a Markov kernel on \((\Sigma, \mathcal{F})\).

The iterated Markov kernels are defined recursively, setting \( K^1 = K \) and for \( n \geq 1, \; K^{n+1}(x, A) = \int_{\Sigma} K^n(y, A) K(x, dy) \). Each power \( K^n \) is itself a Markov kernel on \((\Sigma, \mathcal{F})\). A probability measure \( \mu \) on \((\Sigma, \mathcal{F})\) is called \( K \)-stationary if

\[ \mu(A) = \int K(x, A) \mu(dx), \; \forall A \in \mathcal{F}. \]

A pair \((K, \mu)\), where \( \mu \) is a \( K \)-stationary probability measure will be referred as a Markov system. From now on, we consider \((K, \mu)\) to be a fixed Markov system. It follows from Kolmogorov’s consistency theorem that there exists a unique probability measure \( \mathbb{P}_\mu \) on the sequence space \( \Omega = \Sigma^\mathbb{Z} \) such that the following holds. Let \( e_n : \Omega \to \Sigma, \; e_n(\omega) = x_n, \) where \( \omega = (x_n)_n \), and let \( \mathcal{F}^\mathbb{Z} \) be the Borel \( \sigma \)-field on \( \Omega \). Then \( \{ e_n : \Omega \to \Sigma \}_n \) is a stationary \( \Sigma \)-valued Markov process with common distribution \( \mu \) and transition kernel \( K \). The shift map \( T : \Omega \to \Omega \), defined by \( T(\omega) = (x_{n+1})_n \), preserves the measure \( \mathbb{P}_\mu \). The measure preserving dynamical system \((T, \Omega, \mathcal{F}^\mathbb{Z}, \mathbb{P}_\mu)\) is called a Markov shift. We say that \((K, \mu)\) is strongly mixing if there are constants \( C > 0 \)
and $0 < \rho < 1$ such that for all $f \in L^\infty(\Sigma)$ and $x \in \Sigma$,
\[
\left| \int_{\Sigma} f(y) K^n(x, dy) - \int_{\Sigma} f \, d\mu \right| \leq C \rho^n \| f \|_\infty \quad \forall n \in \mathbb{N} .
\]
This in particular implies that the deterministic counterpart of $(K, \mu)$, the Markov shift $(T, \Omega, \mathcal{F}, \mathbb{P}_\mu)$, is an ergodic and mixing dynamical system. From now we shall assume that $(K, \mu)$ is strongly mixing.

Let $\mathcal{B}(K, \infty)$ be the space of functions $A : \Sigma \times \Sigma \rightarrow \text{GL}(d, \mathbb{R})$ such that
\begin{enumerate}
\item $A$ is $\mathcal{F} \otimes \mathcal{F}$-measurable,
\item $\| A(x, y) \|$ and $\| A(x, y)^{-1} \|$ are uniformly bounded.
\end{enumerate}

The set $\mathcal{B}(K, \infty)$ is a metric space with the following distance
\[
d_\infty(A, B) = \| A - B \|_\infty := \sup_{x, y \in \Sigma} | A(x, y) - B(x, y) | .
\]

More generally, for each $a > 0$ we introduce the space $\mathcal{B}(K, a)$ of functions $A : \Sigma \times \Sigma \rightarrow \text{GL}(d, \mathbb{R})$ such that
\begin{enumerate}
\item $A$ is $\mathcal{F} \otimes \mathcal{F}$-measurable,
\item $\| A(x, y) \|$ is uniformly bounded,
\item for some constant $C > 0$ and all $x \in \Sigma$,
\[
\eta^a_A(x) := \int_{\Sigma} \| A(x, y)^{-1} \|^a K(x, dy) \leq C .
\]
\end{enumerate}

The set $\mathcal{B}(K, a)$ is a metric space with the distance
\[
d_a(A, B) := \| A - B \|_\infty + \| \eta^a_A - \eta^a_B \|_\infty .
\]

Each function $A \in \mathcal{B}(K, a)$, $0 < a \leq \infty$, determines the following linear cocycle $F_A : \Omega \times \mathbb{R}^d \rightarrow \Omega \times \mathbb{R}^d$, defined by
\[
F_A(\omega, v) := (T \omega, \hat{A}(\omega) v) ,
\]
where $\hat{A} : \Omega \rightarrow \text{GL}(d, \mathbb{R})$ is the function $\hat{A}(\omega) := A(x_0, x_1)$, for all $\omega = (x_n)_{n \in \mathbb{Z}}$. This cocycle is determined by, and sometimes identified with, the data $(K, \mu, A)$. The iterates of $F_A$ are the maps $F^n_A : \Omega \times \mathbb{R}^d \rightarrow \Omega \times \mathbb{R}^d$ given by $F^n_A(\omega, v) = (T^n \omega, \hat{A}^{(n)}(\omega) v)$, with $\hat{A}^{(n)} : \Omega \rightarrow \text{GL}(d, \mathbb{R})$ is defined for all $\omega = (x_n)_{n \in \mathbb{Z}}$ by
\[
\hat{A}^{(n)}(\omega) := A(x_{n-1}, x_n) \ldots A(x_1, x_2) A(x_0, x_1) .
\]

Let $\text{Gr}(\mathbb{R}^d)$ denote the Grassmann manifold of the Euclidean space $\mathbb{R}^d$. A $\mathcal{F}$-measurable function $V : \Sigma \rightarrow \text{Gr}(\mathbb{R}^d)$ is said to be $A$-invariant when
\[
A(x_{n-1}, x_n) V(x_{n-1}) = V(x_n) \quad \text{for } \mathbb{P}_\mu\text{-a.e. } \omega = (x_n)_{n \in \mathbb{Z}} .
\]
The ergodicity of \((T, \mu)\) implies that the subspaces \(V(x)\) have constant dimension \(\mu\)-a.e., denoted by \(\dim(V)\). We say that this family is proper if \(0 < \dim(V) < d\). Next we introduce the concept of irreducible cocycle. See definition 2.7 in [4]. A cocycle \(A \in \mathcal{B}(K, a)\) is called irreducible w.r.t. \((K, \mu)\) if it admits no measurable proper \(A\)-invariant family of linear subspaces. We denote by \(\mathcal{B}_{\text{irred}}(K, a)\) the subclass of irreducible cocycles in \(\mathcal{B}(K, a)\).

**Theorem 2.2.** If the Markov system \((K, \mu)\) is strongly mixing then for \(a \geq 2\), the space \(\mathcal{B}_{\text{irred}}(K, a)\) of measurable cocycles over \((T, \Omega, \mathcal{B}, \mathcal{P}_\mu)\) satisfies uniform fiber and base LDT estimates.

**Proof.** See Theorems 7.1 and 7.2.

From this we obtain the following.

**Corollary 2.1.** Under the same assumptions, for every \(m \in \mathbb{N}\) the Lyapunov exponents \(L_i : \mathcal{B}_{\text{irred}}^m(K, a) \to [-\infty, \infty], 1 \leq i \leq m\), are continuous functions. Moreover, given \(A \in \mathcal{B}_{\text{irred}}^m(K, a)\), and \(1 \leq k \leq m\), if the Lyapunov exponent \(L_k(A)\) is simple, then locally near \(A\), the map \(L_k\) is a H"older continuous functions of the cocycle.

**Proof.** Apply Theorem 2.1 to the space \(\mathcal{B}_{\text{irred}}(K, a)\) of measurable cocycles over the base dynamical system \((T, \Omega, \mathcal{B}, \mathcal{P}_\mu)\), with the spaces of LDT parameters \(\mathcal{P}\), and deviation function sets \(\mathcal{E}\) and \(\mathcal{I}\), introduced in the last part of section 6. The chosen set of observables is the Banach algebra \(\Xi = \mathcal{H}_\alpha(\Omega^-)\), for some \(\alpha > 0\) small enough. The compatibility condition (1) in Theorem 2.1 is automatic because \(\Xi\) contains all functions \(1_F\) with \(F \in \mathcal{F}_N(A), N \in \mathbb{N}\) and \(A \in \mathcal{B}(K, a)\). A simple computation shows that the modulus of continuity associated to the choice of deviation function sets \(\mathcal{E}\) and \(\mathcal{I}\) is the modulus of H"older continuity.

**Remark 2.1.** The concept of irreducibility extends to Markov systems of finite order. In this more general context, Corollary 2.1 also applies cocycles \(A : \Omega \to \text{GL}(m, \mathbb{R})\) which depend on finitely many symbols. The reduction to the previous framework is standard and corresponds to redefine the space of symbols to be some product \(\Sigma^k\), with large \(k\).

We need the irreducibility assumption to prove uniform fibre LDT estimates in Theorem 2.2. The proof exploits the fact that for irreducible cocycles there is some Banach algebra of measurable functions, independent of the cocycle, where the associated Laplace-Markov operators act as quasi-compact and simple operators (see section 7). For reducible cocycles this fact is also true, and leads to fibre LDT estimates, but the Banach algebra must be tailored to the cocycle, and
hence the scheme of proof presented here does not provide the required uniformity.

Le Page has proven in [22] the Hölder continuity of the top Lyapunov exponent of strongly irreducible contracting cocycles over a Bernoulli shift. We are not aware of any generalization of Le Page’s theorem, on the continuity of the top Lyapunov exponent, for irreducible cocycles over strongly mixing Markov shifts. The main novelty in our paper is that we address the Hölder continuity of all Lyapunov exponents.

3. A general deterministic estimate

In a previous work [10] we have generalized the classical Avalanche Principle of M. Goldstein and W. Schlag (see [13]) to invertible matrices in $GL(m, \mathbb{R})$. The aim of this section is to extend this result to general matrices in $Mat(m, \mathbb{R})$. To state this Avalanche Principle we need to introduce several concepts like gap patterns, flags, most expanding singular flags and angles between flags.

**Gap Patterns.** Given a matrix $g \in Mat(m, \mathbb{R})$ and a signature $\tau = (\tau_1, \ldots, \tau_k)$, we say that the singular spectrum of $g$ has a $\tau$-gap pattern if and only if

$$s_{\tau_1} > s_{\tau_1+1} \geq s_{\tau_2} > s_{\tau_2+1} \geq \ldots \geq s_{\tau_k} \geq 0.$$

To measure the size of a $\tau$-gap pattern we introduce the quantity

$$\sigma_\tau(g) := \max_{1 \leq j \leq k} \sigma_{\tau_j}(g), \quad \text{where} \quad \sigma_{\tau_j}(g) := \frac{s_{\tau_j+1}(g)}{s_{\tau_j}(g)}.$$

The singular spectrum of $g$ has a $\tau$-gap pattern if and only if $\sigma_\tau(g) < 1$.

**Flags.** Given a signature $\tau = (\tau_1, \ldots, \tau_k)$, any sequence $F = (F_1, \ldots, F_k)$ of vector subspaces $F_1 \subset F_2 \subset \ldots \subset F_k \subset \mathbb{R}^m$ such that $\dim F_j = \tau_j$, for each $j = 1, \ldots, k$, will be called a $\tau$-flag. Given a vector subspace $V \subset \mathbb{R}^m$, we say that the flag $F$ is transversal to $V$ when $F_k \cap V = \{0\}$. We denote by $\mathcal{F}_\tau$ the manifold of all $\tau$-flags, and by $\mathcal{F}_\tau^m[V]$ the submanifold of all $\tau$-flags transversal to $V$. Two special cases of flag manifolds are the projective space $\mathbb{P}^{m-1}$, when $\tau = (1)$, and the Grassmannian manifold $Gr_k^m$, when $\tau = (k)$.

There is a partial action of the algebra $Mat(m, \mathbb{R})$ on flags, where each $g \in Mat(m, \mathbb{R})$ determines the map $\varphi_g : \mathcal{F}_\tau^m[\text{Ker}(g)] \to \mathcal{F}_\tau^m[\text{Ker}(g^*)]$ defined by $\varphi_gF = (gF_1, \ldots, gF_k)$.

The Grassmannian $Gr_k^m$ can be identified with a submanifold of the projective space $\mathbb{P}(\wedge_k \mathbb{R}^m)$, by identifying each vector subspace $V \subset \mathbb{R}^m$ of dimension $k$ with the exterior product $v_1 \wedge v_2 \wedge \ldots \wedge v_k$, where
\{v_1, v_2, \ldots, v_k\} is any orthonormal basis of \(V\). With the canonical Riemannian metric, projective spaces have diameter \(\pi/2\). We normalize the distance so that projective spaces, and in particular Grassmanian manifolds \(\text{Gr}_k\), have diameter one. Consider \(\mathcal{F}_\tau^m\) as a submanifold of \(\text{Gr}_{\tau_1} \times \text{Gr}_{\tau_2} \times \ldots \times \text{Gr}_{\tau_k}\) which with the induced (max) product distance, has also diameter 1. With this metric, each flag manifold \(\mathcal{F}_\tau^m\) is a compact metric space with diameter 1.

**Singular Flags.** Assuming \(\|g\| = s_1(g) > s_2(g)\), denote by \(\hat{v}_\pm(g) \in \mathbb{P}^{m-1}\) the most expanding singular directions of \(g\). Letting \(v_\pm(g) \in \mathbb{R}^m\) be a unit vector in the line \(\hat{v}_\pm(g)\), we have

\[
\begin{align*}
g v_-(g) &= \|g\| v_+(g) \\
g^*v_+(g) &= \|g\| v_-(g)
\end{align*}
\]

Let \(v_1^\pm(g), \ldots, v_m^\pm(g)\) be orthonormal singular eigen-basis of \(g\). These basis are characterized by the relations

\[
g v_j^\pm(g) = s_j(g) v_j^\pm(g), \quad 1 \leq j \leq m
\]

Assume now that \(g\) has a \(\tau\)-gap pattern. We define the most expanding \(\tau\)-flags of \(g\) as

\[
\hat{v}_{\tau,\pm}(g) := (\hat{V}_{\tau,1}^\pm(g), \ldots, \hat{V}_{\tau,k}^\pm(g)) \in \mathcal{F}_\tau^m
\]

where \(\hat{V}_{\tau,j}^\pm(g)\) is the linear span \(\langle v_1^\pm(g), \ldots, v_j^\pm(g) \rangle\) for \(1 \leq j \leq k\). The subspaces \(\hat{V}_{\tau,j}^\pm(g)\) do not depend on the choice of the singular eigen-basis, precisely because \(g\) has a \(\tau\)-gap pattern. Identifying each vector subspace \(\hat{V}_{\tau,j}^\pm(g) \subset \mathbb{R}^m\) with a simple \(\tau_j\)-vector \(V_{\tau,j}^\pm(g) \in \bigwedge_{\tau_j} \mathbb{R}^m\), the following relation holds

\[
(\bigwedge_{\tau_j} g) V_{\tau,j}^\pm(g) = p_{\tau_j}(g) V_{\tau,j}^\pm(g), \quad \text{where } p_{\tau_j}(g) = \|\bigwedge_{\tau_j} g\|.
\]

The flag \(\hat{v}_{\tau,-}(g)\) is transversal to \(\text{Ker}(g)\), while \(\hat{v}_{\tau,+}(g)\) is transversal to \(\text{Ker}(g^*)\). Furthermore,

\[
\varphi_g \hat{v}_{\tau,-}(g) = \hat{v}_{\tau,+}(g)
\]

**Angles between Flags.** We call the correlation of \(u\) and \(v \in \mathbb{P}^{m-1}\) to the number

\[
\alpha(u, v) := |\langle u, v \rangle|.
\]

The correlation \(\alpha(u, v)\) is the sine of the angle between \(u\) and the orthogonal complement \(v^\perp\). Similarly, given subspaces \(U, V \in \text{Gr}_k\) we define the correlation of \(U\) and \(V\) to be

\[
\alpha(U, V) := |\langle u_1 \wedge \ldots \wedge u_k, v_1 \wedge \ldots \wedge v_k \rangle|,
\]
where \{u_1, \ldots, u_k\} and \{v_1, \ldots, v_k\} are any orthonormal basis of \(U\) and \(V\), respectively. Finally, given flags \(F, G \in \mathcal{F}_\tau^m\), we set
\[
\alpha(F, G) := \min_{1 \leq j \leq k} \alpha(F_j, G_j).
\]

Clearly, two directions \(u, v \in \mathbb{P}^{m-1}\) are orthogonal if and only if \(\alpha(u, v) = 0\). We say that two subspaces \(U, V \in \text{Gr}_k^m\) are orthogonal, and write \(U \perp V\), if and only if there are unit vectors \((u, v)\) such that \(u \in U \cap V^\perp\) and \(v \in V \cap U^\perp\). It follows easily that
\[
\alpha(U, V) = 0 \iff U \perp V.
\]

Finally we say that two flags \(F, G \in \mathcal{F}_\tau^m\) are orthogonal, and write \(F \perp G\), if and only if for some \(1 \leq j \leq k\), \(F_j \perp G_j\). Again we have
\[
\alpha(F, G) = 0 \iff F \perp G.
\]

The angle between two lines in \(\mathbb{P}^{m-1}\), two subspaces in \(\text{Gr}_k^m\), or two flags in \(\mathcal{F}_\tau^m\), is defined by \(\angle(u, v) := \arccos(\alpha(u, v))\). This angle relates with the normalized distance of these flag manifolds as follows
\[
\angle(F, G) = \frac{\pi}{2} d(F, G).
\]

For each flag \(F \in \mathcal{F}_\tau^m\) we define the orthogonal flag hyperplane:
\[
\Sigma(F) = \{ G \in \mathcal{F}_\tau^m : \alpha(G, F) = 0 \} = \{ G \in \mathcal{F}_\tau^m : G \perp F \}
\]

Lemma 3.1. For any flags \(F, G \in \mathcal{F}_\tau^m\),
\[
(a) \ \alpha(F, G) = \sin \left( \frac{\pi}{2} d(G, \Sigma(F)) \right),
\]
\[
(b) \ d(G, \Sigma(F)) \leq \alpha(F, G) \leq \frac{\pi}{2} d(G, \Sigma(F)).
\]

Proof. See lemma 3.2 in [10].

Assuming that \(g\) and \(g'\) have \(\tau\)-gap patterns, we define
\[
\alpha_\tau(g, g') := \alpha(\hat{v}_{\tau,+}(g), \hat{v}_{\tau,-}(g')).
\] (3.1)

The statement. Next theorem says that given a chain of matrices \(g_0, g_1, \ldots, g_{n-1}\) in \(\text{Mat}(m, \mathbb{R})\), with some quantified \(\tau\)-gap pattern, and with minimum ‘angles’ between the most expanding singular flags for pairs of consecutive matrices, then their product keeps the same pattern. Given \(0 \leq j < i < n\), let us write \(g^{(i)} := g_{i-1} \ldots g_1 g_0\).

Theorem 3.1. There exists a constant \(c > 0\) such that given \(0 < \varepsilon < 1, 0 < \kappa \leq c \varepsilon^2\) and \(g_0, g_1, \ldots, g_{n-1} \in \text{Mat}(m, \mathbb{R})\), if
\[
(a) \ \sigma_\tau(g_i) \leq \kappa, \text{ for } 0 \leq i \leq n-1, \text{ and }
\]
\[
(b) \ \alpha_\tau(g_{i-1}, g_i) \geq \varepsilon, \text{ for } 1 \leq i \leq n-1,
\]
then
(1) \(d(\hat{\nu}_{\tau,+}(g^{(n)}), \hat{\nu}_{\tau,+}(g_{n-1})) \lesssim \kappa \varepsilon^{-1}\)
(2) \(d(\hat{\nu}_{\tau,-}(g^{(n)}), \hat{\nu}_{\tau,-}(g_0)) \lesssim \kappa \varepsilon^{-1}\)
(3) \(\sigma_{\tau}(g^{(n)}) \leq \left(\frac{\kappa (1+\varepsilon)}{\varepsilon^2}\right)^n\)
(4) for any \(\tau\)-s.v.p. function \(\pi,\)
\[
|\log \pi(g^{(n)}) + \sum_{j=1}^{n-2} \log \pi(g_i) - \sum_{\ell=1}^{n-1} \log \pi(g_{\ell \ell-1})| \lesssim n \frac{K}{\varepsilon^2} \tag{3.2}
\]

**Remark 3.1.** All constants in this theorem can be made explicit. For instance, the non optimal value \(c = \frac{2-\sqrt{2}}{2\pi + 2\pi}\) leads to the explicit upper bound \(2\pi \kappa \varepsilon^{-1}\) in (1) and (2). Of course the conclusion of the theorem is only interesting when \(\kappa \ll \varepsilon^2\).

**On the proof.** The strategy to prove conclusions (1), (2) and (3) is to look at the contracting action of matrices \(g_j \in \text{Mat}(m, \mathbb{R})\) on the flag manifold \(\mathcal{F}_m^n\). Because of assumption (a), each \(g_i \in \text{Mat}(m, \mathbb{R})\) acts as a contraction on \(\mathcal{F}^m_\tau\) away from the critical set
\[
\Sigma^{-\varepsilon}_\tau(g_i) := \{ F \in \mathcal{F}^m_\tau : \alpha(F, \hat{\nu}_{\tau,-}(g_i)) < \varepsilon \}
\]
Assumption (b) ensures that \(\varphi_{g_i}(\mathcal{F}^m_\tau \setminus \Sigma^{-\varepsilon}_\tau(g_i))\) does not intersect the next critical set \(\Sigma^{-\varepsilon}_\tau(g_{i+1})\), at least for some possibly smaller \(\varepsilon > 0\). Hence the composition \(\varphi_{g^{(n)}}\) acts as a contraction on some appropriate domain of the flag manifold \(\mathcal{F}^m_\tau\).

More precisely we use an abstract lemma on chains of continuous contracting maps \(g_i : X \setminus \Sigma_i \to X \ (0 \leq i \leq n)\) on a compact metric space \(X\). Denote by \(B_\varepsilon(\Sigma)\) the \(\varepsilon\)-neighbourhood of a subset \(\Sigma \subset X\),
\[
B_\varepsilon(\Sigma) = \{ x \in X : d(x, \Sigma) < \varepsilon \}
\]

**Lemma 3.2.** Let \(\varepsilon, \delta > 0\) and \(0 < \kappa < 1\) such that \(\delta/(1-\kappa) < \varepsilon < 1/2\) and \(\delta < \kappa\). Given a compact metric space \(X\) with diameter 1, closed subsets \(\Sigma_0, \ldots, \Sigma_{n-1} \subset X\), a chain of continuous mappings \(g_j : X \setminus \Sigma_j \to X\), for \(j = 0, 1, \ldots, n-1\), and pairs of points \((x_0, y_0), (x_1, y_1), \ldots, (x_{n-1}, y_{n-1})\) in \(X \times X\) such that for every \(0 \leq i \leq n-1:\)
- (a) \(g_i(x_i) = y_i\),
- (b) \(d(x_i, \Sigma_i) = 1\) and \(d(y_i, \Sigma_{i+1}) \geq 2\varepsilon\),
- (c) \(g_i(X \setminus B_\varepsilon(\Sigma_i)) : X \setminus B_\varepsilon(\Sigma_i) \to X\) has Lipschitz constant \(\leq \kappa\),
- (d) \(g_i(X \setminus B_\varepsilon(\Sigma_i)) \subset B_\delta(y_i)\).

Then setting \(g^{(n)} := g_{n-1} \circ \ldots \circ g_1 \circ g_0\), the following hold:
- (1) \(d\left(y_{n-1}, g^{(n)}(x_0)\right) \leq \frac{\delta}{1-\kappa}\),
- (2) if \(x_0 = y_{n-1}\) then there is a unique point \(x^* \in X\) such that \(g^{(n)}(x^*) = x^*\) and \(d\left(x_0, x^*\right) \leq \frac{\delta}{(1-\kappa)(1-\kappa^2)}\).
Proof. See Lemma 3.1 of [10]. □

This lemma is applied with \( X = \mathcal{F}_\tau^m \) to the chains of maps \( \varphi_{g_0}, \varphi_{g_1}, \ldots, \varphi_{g_i}: \varphi_{g_i}, \ldots, \varphi_{g_0} \) and \( \varphi_{g_i}', \ldots, \varphi_{g_0}', \varphi_{g_i}, \ldots, \varphi_{g_i} \) to infer conclusions (1) and (2) on the most expanding singular flags \( \hat{v}_{\tau,\pm}(g^{(i)}) \).

The main difference with the previous (non-singular) Avalanche Principle is that any \( g \in \text{GL}(m, \mathbb{R}) \) induces a diffeomorphism \( \varphi_{g}: \mathcal{F}_\tau^m \to \mathcal{F}_\tau^m \), while a singular matrix \( g \in \text{Mat}(m, \mathbb{R}) \) will only induce a partial map \( \varphi_{g}: \mathcal{F}_\tau^m[\text{Ker}(g)] \to \mathcal{F}_\tau^m[\text{Ker}(g^*)] \). However, this will not matter because the partial maps \( \varphi_{g} \) will always be defined over the region in the flag manifold \( \mathcal{F}_\tau^m \) where their action is analysed. Given \( g \in \text{Mat}(m, \mathbb{R}) \), we define the singular critical sets

\[
\Sigma_{\tau}^\pm(g) := \{ F \in \mathcal{F}_\tau^m : \alpha(F, \hat{v}_{\tau,-}(g)) = 0 \},
\]

as well as their \( \varepsilon \)-neighbourhoods (\( \varepsilon > 0 \))

\[
\Sigma_{\tau}^{\pm,\varepsilon}(g) := \{ F \in \mathcal{F}_\tau^m : \alpha(F, \hat{v}_{\tau,\pm}(g)) < \varepsilon \}.
\]

By definition,

\[
\mathcal{F}_\tau^m \setminus \Sigma_{\tau}^-(g) \subseteq \mathcal{F}_\tau^m[\text{Ker}(g)] \quad \text{and} \quad \mathcal{F}_\tau^m \setminus \Sigma_{\tau}^+(g) \subseteq \mathcal{F}_\tau^m[\text{Ker}(g^*)].
\]

This shows that \( \varphi_{g} \) is defined on \( \mathcal{F}_\tau^m \setminus \Sigma_{\tau}^{-,\varepsilon}(g) \) and \( \varphi_{g^*} \) is defined on \( \mathcal{F}_\tau^m \setminus \Sigma_{\tau}^{+,\varepsilon}(g^*) \), for every \( g \in \text{Mat}(m, \mathbb{R}) \). Next lemma (Lemma 3.3 in [10]) shows that assumption (a), \( \sigma_\tau(g_i) \leq \kappa \), implies \( \varphi_{g_i}: \mathcal{F}_\tau^m \setminus \Sigma_{\tau}^{-,\varepsilon}(g_i) \to \mathcal{F}_\tau^m \) is a Lipschitz contraction. The proof is exactly the same and shall be omitted.

**Proposition 3.3.** Given \( \varepsilon > 0 \) and \( \kappa > 0 \), for any \( g \in \text{Mat}(m, \mathbb{R}) \) with \( \sigma_\tau(g) \leq \kappa \), the restriction mapping \( \varphi_{g}: \mathcal{F}_\tau^m \setminus \Sigma_{\tau}^{-,\varepsilon}(g) \to \mathcal{F}_\tau^m \) has Lipschitz constant \( \kappa (1 + \varepsilon)/\varepsilon^2 \) and \( \varphi_{g}(\mathcal{F}_\tau^m \setminus \Sigma_{\tau}^{-,\varepsilon}(g)) \subseteq B_{\kappa/\varepsilon}(\hat{v}_{\tau,+}(g)) \).

It follows that for every \( i = 1, \ldots, n \),

\[
\sigma_\tau(g^{(i)}) = \max_{1 \leq j \leq k} \frac{s_{\tau_j+1}(g)}{s_{\tau_j}(g)} \leq \max_{1 \leq j \leq k} \text{Lip}(\varphi_{\Lambda_{\tau_j}g^{(i)}}) \leq \left( \frac{\kappa (1 + \varepsilon)}{\varepsilon^2} \right)^i,
\]

where \( \text{Lip}(\varphi_{\Lambda_{\tau_j}g^{(i)}}) \) denotes the the Lipschitz constant of the map \( \varphi_{\Lambda_{\tau_j}g^{(i)}} \) in some appropriate domain of the projective space \( \mathcal{F}_\tau^m = \mathbb{P}(\Lambda_{\tau_j} \mathbb{R}^m) \). This proves conclusion (3) of the Avalanche Principle. It also shows that \( g^{(i)} \) has a \( \tau \)-gap pattern for every \( i = 1, \ldots, n \).

Conclusion (4) of the Avalanche Principle is based on a couple of facts that we transcribe here from [10]. First a definition. Given \( g, g' \in \text{Mat}(m, \mathbb{R}) \) with \( \tau \)-gap patterns let

\[
\beta_\tau(g, g') := \sqrt{\sigma_\tau(g)^2 + \alpha_\tau(g, g')^2 + \sigma_\tau(g')^2},
\]

(3.3)
where the operation
\[ a \oplus b := a + b - a b \]
induces a commutative semigroup structure on the interval \([0, 1]\). It follows that
\[ 1 \leq \frac{\beta_{\tau}(g, g')}{\alpha_{\tau}(g, g')} \leq \sqrt{1 + \frac{\sigma_{\tau}(g)^2 + \sigma_{\tau}(g')^2}{\alpha_{\tau}(g, g')^2}}. \quad (3.4) \]

We respectively write \( \alpha(g, g') \) and \( \beta(g, g') \) instead of \( \alpha_{\tau}(g, g') \) and \( \beta_{\tau}(g, g') \) when \( \tau = (1) \). We also write \( \alpha_k(g, g') \) and \( \beta_k(g, g') \) instead of \( \alpha_{\tau}(g, g') \) and \( \beta_{\tau}(g, g') \) when \( \tau = (k) \).

**Proposition 3.4.** Given \( g, g' \in \text{Mat}(m, \mathbb{R}) \) with a \((1)\)-gap pattern,
\[ \alpha(g, g') \leq \frac{\|g'g\|}{\|g'\| \|g\|} \leq \beta(g, g'). \]

The proof of this fact does not depend on the invertibility of matrices \( g, g' \) (see Proposition 3.8 in [10]). These inequalities then imply the following more general fact (see Corollary 3.10 in [10]).

**Proposition 3.5.** Given \( g_0, g_1, \ldots, g_{n-1} \in \text{Mat}(m, \mathbb{R}) \), and the \( \tau \)-s.v.p. function \( \pi = \pi_{\tau,j} \), if every matrix \( g_i \) and \( g(i) = g_{i-1} \ldots g_0 \) has a singular spectrum with a \( \tau \)-gap pattern then
\[ \prod_{i=1}^{n-1} \frac{\alpha_{\tau_j}(g(i), g_i)}{\beta_{\tau_{j-1}}(g(i), g_i)} \leq \frac{\pi(g_{n-1} \ldots g_1 g_0)}{\pi(g_{n-1}) \ldots \pi(g_1) \pi(g_0)} \leq \prod_{i=1}^{n-1} \frac{\beta_{\tau_j}(g(i), g_i)}{\alpha_{\tau_{j-1}}(g(i), g_i)} \]

Finally to prove (3.2), we consider a \( \tau \)-s.v.p. function \( \pi \) and apply Proposition 3.5 to the factors of the following ratio
\[ \frac{\pi(g_{n-1} \ldots g_1 g_0) \pi(g_{n-2}) \pi(g_{n-2}) \ldots \pi(g_1)}{\pi(g_{n-1} g_{n-2}) \ldots \pi(g_1 g_0)} = \frac{\pi(g_{n-1} \ldots g_1 g_0)}{\pi(g_{n-1}) \ldots \pi(g_1) \pi(g_0)} \pi(g_{n-2}) \ldots \pi(g_1 \pi(g_0)) \frac{\pi(g_{n-1}) \pi(g_{n-2})}{\pi(g_{n-1} g_{n-2})} \ldots \frac{\pi(g_1) \pi(g_0)}{\pi(g_{n-1} g_{n-2})}. \]

Using (3.4) we can get upper and lower bounds for this expression written as products of \( 2(n - 1) \) factors close to 1. The proof is exactly the same as in [10].

Finally we rephrase the Avalanche Principle in the form that will be used throughout the rest of the paper. See the proof in [10].
Proposition 3.6. There exists \( c > 0 \) such that given \( 0 < \epsilon < 1 \), \( 0 < \kappa \leq c\epsilon^2 \) and \( g_0, g_1, \ldots, g_{n-1} \in \text{Mat}(m, \mathbb{R}) \), if

\[
\begin{align*}
(g\text{aps}) & \quad \rho(g_i) > \frac{1}{\kappa} \quad \text{for all } \tau \text{- s.v.r. } \rho, \quad 0 \leq i \leq n - 1 \\
(\text{angles}) & \quad \frac{\pi(g_i \cdot g_{i-1})}{\pi(g_i) \cdot \pi(g_{i-1})} > \epsilon \quad \text{for all } \tau \text{- s.v.p. } \pi, \quad 1 \leq i \leq n - 1
\end{align*}
\]

then we have:

\[
\rho(g^{(n)}) > \left(\frac{\epsilon^2}{\kappa}\right)^n \quad \text{for all } \tau \text{- s.v.r. } \rho
\]

\[
\left| \log \pi(g^{(n)}) + \sum_{i=1}^{n-2} \log \pi(g_i) - \sum_{i=1}^{n-1} \log \pi(g_i \cdot g_{i-1}) \right| \lesssim n \cdot \frac{\kappa}{\epsilon^2}
\]

for all \( \tau \text{- s.v.p. } \pi \)

4. Upper semicontinuity of the top Lyapunov exponent

Given are an ergodic system \((X, \mathcal{F}, \mu, T)\), a space of measurable cocycles \( \mathcal{C} \), a set of observables \( \Xi \) and a set of LDT parameters \( \mathcal{P} \) with corresponding spaces of deviation functions \( \mathcal{E} \) and \( \mathcal{I} \).

It is well known that the top Lyapunov exponent is upper semicontinuous as a function of the cocycle. Our argument requires a much more precise version of the upper semicontinuity, one that is uniform in the number \( n \) of iterates and in the phase \( x \). Such results are available, see [18], [11], and they are based on a stopping time argument used by Katznelson and Weiss [19] in their proofs of the Birkhoff’s and Kingman’s ergodic theorems. However, the results in [18], [11] require unique ergodicity of the system, a property that Bernoulli and Markov shifts do not satisfy. By replacing unique ergodicity with a weaker property - namely that a base-LDT holds for a large enough set of observables, which we show later to hold for Markov shifts - we obtain a (weaker) version of the uniform upper semicontinuity in [18], one which holds for a large enough set of phases.

Proposition 4.1 (nearly uniform upper semicontinuity). Let \( A \in C_m \) be a measurable cocycle such that \( \Xi \) and \( A \) are compatible and every observable \( \xi \in \Xi \) satisfies a base-LDT.

(i) Assume that \( L_1(A) > -\infty \) and that \( A \) is \( L^1 \)-bounded.
For every \( \epsilon > 0 \), there are \( \delta = \delta(A, \epsilon) > 0 \), \( n_0 = n_0(A, \epsilon) \in \mathbb{N} \) and \( \iota = \iota(A, \epsilon) \in \mathbb{I} \), such that if \( B \in C_m \) with \( d(B, A) < \delta \), and if \( n \geq n_0 \), then the upper bound

\[
\frac{1}{n} \log \| B^{(n)}(x) \| \leq L_1(A) + \epsilon
\]  

holds for all \( x \) outside of a set of measure \( < \iota_n \).

Up to a zero measure set, the exceptional set depends only on \( A, \epsilon \).

(ii) Assume that \( L_1(A) = -\infty \).

For every \( t < \infty \), there are \( \delta = \delta(A, t) > 0 \), \( n_0 = n_0(A, t) \in \mathbb{N} \) and \( \iota = \iota(A, t) \in \mathbb{I} \), such that if \( B \in C_m \) with \( d(B, A) < \delta \), and if \( n \geq n_0 \), then the upper bound

\[
\frac{1}{n} \log \| B^{(n)}(x) \| \leq -t
\]  

holds for all \( x \) outside of a set of measure \( < \iota_n \).

Up to a zero measure set, the exceptional set depends only on \( A, t \).

Proof. Throughout this proof, \( C \) will stand for a positive, finite, large enough constant that depends only on the cocycle \( A \), and which may change slightly from one estimate to another.

If \( B \in C \) is at some small distance from \( A \), then it will be close enough to \( A \) in the \( L^\infty \) distance as well, hence we will assume that for \( \mu \) a.e. \( x \in X \) we have \( \| B(x) \| < C \).

Moreover, in the case (i), when \( L_1(A) \) is finite, since we also assume \( A \) to be \( L^1 \) bounded, we may choose the constant \( C \) such that for all \( n \geq 1 \) we have \( \left\| \frac{1}{n} \log \| A^{(n)}(\cdot) \| \right\|_{L^1} < C \) and hence also \( |L_1(A)| < C \).

The proofs for each of the two cases are similar, but the argument will differ in some parts. We first present the case \( L_1(A) > -\infty \) in detail, then indicate how to modify the argument for the case \( L_1(A) = -\infty \).

(i) Fix \( \epsilon > 0 \). By Kingman’s subadditive ergodic theorem,

\[
\lim_{n \to \infty} \frac{1}{n} \log \| A^{(n)}(x) \| = L_1(A) \quad \text{for } \mu \text{ a.e. } x
\]

hence the number

\[
\iota_n(x) := \min\{n \geq 1: \frac{1}{n} \log \| A^{(n)}(x) \| < L_1(A) + \epsilon\}
\]  

is defined for \( \mu \) a.e. \( x \in X \).

For every integer \( N \), let

\[
\mathcal{U}_N := \{x: \iota_n(x) \leq N\} = \bigcup_{n=1}^{N} \{x: \frac{1}{n} \log \| A^{(n)}(x) \| < L_1(A) + \epsilon\}
\]
Then $U_N^c \in \mathcal{F}_N(A)$, $U_N \subset U_{N+1}$ and $\cup_N U_N$ has full measure. Therefore, there is $N = N(\epsilon, A)$ such that $\mu(U_N^c) < \epsilon$.

We fix this integer $N$ for the rest of the proof and denote the set $\mathcal{U} = \mathcal{U}(\epsilon, A) := U_N$. Therefore, $U_N^c \in \mathcal{F}_N(A)$, $\mu(U_N^c) < \epsilon$ and we have: if $x \in \mathcal{U}$ then $1 \leq n(x) \leq N$ and

$$\log \| A^{(n(x))}(x) \| \leq n(x)L_1(A) + n(x)\epsilon$$  \hspace{1cm} (4.4)

Next we will bound from above $\log \| B^{(n)}(x) \|$ by $\log \| A^{(n)}(x) \| + o(1)$ for all cocycles $B$ with $\text{dist}(B, A) < \delta$ where $\delta$ will be chosen later, for all $1 \leq n \leq N$ and for a large set of phases $x \in X$.

Since $A$ is $L^1$-bounded, $\log \| A^{(n)} \| \in L^1(X, \mu)$, so $A^{(n)}(x) \neq 0$ for $\mu$-a.e. $x \in X$.

Moreover, if $B \in \mathcal{C}_m$ with $\text{dist}(B, A) < \delta$ (where $\delta \ll 1$ is chosen below), we have $\| B(x) - A(x) \| < \delta$ and $\| B(x) \| < C$ for $\mu$-a.e. $x \in X$.

Then for $x$ outside a null set and for $1 \leq n \leq N$, we have:

$$\log \| B^{(n)}(x) \| - \log \| A^{(n)}(x) \| = \log \frac{\| B^{(n)}(x) \|}{\| A^{(n)}(x) \|}$$

$$\leq \log \left[ \frac{\| B^{(n)}(x) - A^{(n)}(x) \| + 1}{\| A^{(n)}(x) \|} \right] \leq \frac{\| B^{(n)}(x) - A^{(n)}(x) \|}{\| A^{(n)}(x) \|}$$

$$\leq nC^{-1}\delta \cdot \frac{1}{\| A^{(n)}(x) \|} \leq NC^{-1}\delta \cdot \frac{1}{\| A^{(n)}(x) \|}.$$  \hspace{1cm} Hence

$$\log \| B^{(n)}(x) \| \leq \log \| A^{(n)}(x) \| + \delta NC^{-1} \cdot \frac{1}{\| A^{(n)}(x) \|}$$  \hspace{1cm} (4.5)

for all $x$ outside a zero measure set and for all $1 \leq n \leq N$.

Let $t := e^{-N^2/C/\epsilon}$. Consider the set $\mathcal{V} := \bigcap_{n=1}^N \{ x : \| A^{(n)}(x) \| > t \}$. Clearly $\mathcal{V}^c \in \mathcal{F}_N(A)$, and we will show that $\mathcal{V}^c$ has measure at most $\epsilon$.

If for some $1 \leq n \leq N$ and $x \in X$ we have $\| A^{(n)}(x) \| \leq t$ ($< 1$), then

$$\left| \frac{1}{n} \log \| A^{(n)}(x) \| \right| > \frac{\log 1/t}{n}$$

hence

$$\mathcal{V}^c \subset \bigcup_{n=1}^N \{ x : \left| \frac{1}{n} \log \| A^{(n)}(x) \| \right| > \frac{\log 1/t}{n} \}$$

Since $A$ is $L^1$-bounded, there is $C = C(A) < \infty$ such that for all $n \geq 1$

$$\left\| \frac{1}{n} \log \| A^{(n)}(\cdot) \| \right\|_{L^1} < C$$
Then by Chebyshev's inequality,

$$\mu \left\{ x : \frac{1}{n} \log \| A^{(n)}(x) \| > \frac{\log 1/t}{n} \right\} < \frac{Cn}{\log 1/t} \leq \frac{CN}{\log 1/t} = \frac{\epsilon}{N}$$

Therefore

$$\mu(V^c) < N \epsilon/N = \epsilon$$

and if $1 \leq n \leq N$ then for $\mu$ a.e. $x \in V$ we have

$$\log \| B^{(n)}(x) \| \leq \log \| A^{(n)}(x) \| + \delta N C^{N-1} e^{N^2C/\epsilon} < \log \| A^{(n)}(x) \| + \epsilon$$

provided we choose $\delta < \delta(\epsilon, C, N) = \delta(\epsilon, A)$ small enough.

Let $O := U \cap V$. Then $O^c \in F_N(A)$ and $\mu(O^c) < 2\epsilon$. We conclude that for $\mu$ almost every $x \in O$, we have:

$$\log \| B^{(n(x))}(x) \| \leq n(x) L_1(A) + n(x) 2\epsilon \quad (4.6)$$

Let $n_0 = n_0(\epsilon, A) := \frac{CN}{\epsilon}$.

Fix $x \in X$ and define inductively for all $k \geq 1$ the sequence of phases $x_k = x_k(x) \in X$ and the sequence of integers $n_k = n_k(x) \in \mathbb{N}$ as follows:

- $x_1 = x$
- $x_2 = T^{n_1}x_1$
- $\ldots$
- $x_{k+1} = T^{n_k}x_k$
- $n_1 = \begin{cases} n(x_1) & \text{if } x_1 \in O \\ 1 & \text{if } x_1 \notin O \end{cases}$
- $n_2 = \begin{cases} n(x_2) & \text{if } x_2 \in O \\ 1 & \text{if } x_2 \notin O \end{cases}$
- $\ldots$
- $n_{k+1} = \begin{cases} n(x_{k+1}) & \text{if } x_{k+1} \in O \\ 1 & \text{if } x_{k+1} \notin O \end{cases}$

Note that for all $k \geq 1$, $x_{k+1} = T^{n_{k} + \ldots + n_1} x$ and $1 \leq n_k \leq N$.

For any $n \geq n_0(> N \geq n_1)$, there is $p \geq 1$ such that

$$n_1 + \ldots + n_p \leq n < n_1 + \ldots + n_p + n_{p+1}$$

so $n = n_1 + \ldots + n_p + m$, where $0 \leq m < n_{p+1} \leq N$.

For any cocycle $B$ such that $d(B, A) < \delta$, let $b_n(x) := \log \| B^{(n)}(x) \|$. Then clearly for $\mu$ - a.e. $x \in X$ we have $b_n(x) \leq n C$, where $C$ is a constant that depends on $A$ and $b_n(x)$ is a sub-additive process, meaning:

$$b_{n+m}(x) \leq b_n(x) + b_m(T^nx)$$

for all $n, m \geq 1$ and for $\mu$ almost every $x \in X$. 

Using this sub-additivity and the definition of \( x_k(x), n_k(x) \), we have:

\[
\log \| B^{(n)}(x) \| = b_n(x) = b_{n_1+\ldots+n_p+m}(x) \leq \sum_{k=1}^{p} b_{n_k}(x_k) + b_m(x_{p+1})
\]

(4.7)

We will estimate each term separately. Each estimate is valid for \( x \) outside a null set.

For the last term we use the trivial bound:

\[
b_m(x_{p+1}) \leq mC < NC
\]

(4.8)

For every \( 1 \leq k \leq p \) we have:

- Either \( x_k \in \emptyset \), so \( n_k = n(x_k) \), in which case, using (4.6) we get:

\[
b_{n_k}(x_k) = \log \| B^{(n(x_k))}(x_k) \| \leq n(x_k) L_1(A) + 2\epsilon n(x_k)
\]

= \( (L_1(A) + 2\epsilon) n_k \)

- Or \( x_k \notin \emptyset \), so \( n_k = 1 \), in which case

\[
b_{n_k}(x_k) = \log \| B(x_k) \| \leq C
\]

Therefore,

\[
b_{n_k}(x_k) = b_{n_k}(x_k) 1_\emptyset(x_k) + b_{n_k}(x_k) 1_{\neg \emptyset}(x_k)
\]

\[
\leq (L_1(A) + 2\epsilon) n_k 1_\emptyset(x_k) + C 1_{\neg \emptyset}(x_k)
\]

\[
= (L_1(A) + 2\epsilon) n_k - (L_1(A) + 2\epsilon) n_k 1_{\emptyset}(x_k) + C 1_{\neg \emptyset}(x_k)
\]

\[
= (L_1(A) + 2\epsilon) n_k - (L_1(A) + 2\epsilon) 1_{\emptyset}(x_k) + C 1_{\neg \emptyset}(x_k)
\]

where in the last equality we used the fact that \( n_k = 1 \) when \( x_k \in \emptyset \).

We conclude:

\[
b_{n_k}(x_k) \leq (L_1(A) + 2\epsilon) n_k + (C - L_1(A) - 2\epsilon) 1_{\neg \emptyset}(x_k)
\]

\[
< (L_1(A) + 2\epsilon) n_k + 3C 1_{\neg \emptyset}(x_k)
\]

(4.9)

Adding up the estimates (4.8) and (4.9) for all \( 1 \leq k \leq p \), and combining with (4.7), we have:

\[
\log \| B^{(n)}(x) \| \leq (n_1 + \ldots + n_p)(L_1(A) + 2\epsilon) + 3C \sum_{k=1}^{p} 1_{\neg \emptyset}(x_k) + CN
\]

\[
\leq n (L_1(A) + 2\epsilon) + 3C \sum_{j=0}^{n-1} 1_{\neg \emptyset}(T^j x) + CN
\]
Divide both sides by \( n \) to conclude that for \( \mu \) a.e. \( x \in X \) and for all \( n \geq n_0 \) we have:

\[
\frac{1}{n} \log \|B^{(n)}(x)\| \leq L_1(A) + 2\epsilon + 3C \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{1}_{\mathcal{O}^c(T^j x)} + \frac{CN}{n} \quad (4.10)
\]

By the choice of \( n \) we have \( \frac{CN}{n} < \epsilon \), so all is left is to estimate the Birkhoff average above. We use the compatibility condition. Since \( \mathcal{O}^c \in \mathcal{F}_N(A) \), there is an observable \( \xi = \xi(A,\epsilon) \in \Xi \) such that \( \|\mathbb{1}_{\mathcal{O}^c(x)}\| \leq \xi \) and \( \int_X \xi \, d\mu < \mu(\mathcal{O}^c) + \epsilon < 3\epsilon \). Then, applying the base-LDT to \( \xi \), there is \( p = p(\epsilon,\xi) = p(A,\epsilon) \in \mathcal{P} \), \( p = (n_0,\epsilon,\xi) \), such that for \( n \geq n_0 \) we have \( \epsilon_n \leq \epsilon \) and

\[
\frac{1}{n} \sum_{j=0}^{n-1} \mathbb{1}_{\mathcal{O}^c(T^j x)} \leq \frac{1}{n} \sum_{j=0}^{n-1} \xi(T^j x) < \int_X \xi \, d\mu + \epsilon_n < 4\epsilon
\]

provided we choose \( x \) outside a set of measure \( \iota_n \). This ends the proof in the case \( L_1(A) > -\infty \).

(ii) The case \( L_1(A) = -\infty \). Let \( t \) be large enough, say \( t > C + 1 \). We apply again Kingman’s subadditive theorem, and define the integers

\[
n(x) := \min\{n \geq 1: \frac{1}{n} \log \|A^{(n)}(x)\| < -2t\} \quad (4.11)
\]

for \( \mu \) a.e. \( x \in X \).

Define the sets \( \mathcal{U}_N \) similarly, and fix \( N = N(A,t) \), and then \( \mathcal{U} = \mathcal{U}(A,t) = \mathcal{U}_N \) so that \( \mu(\mathcal{U}^c) < 1/t \). Furthermore, as before, \( \mathcal{U}^c \in \mathcal{F}_N(A) \) and if \( x \in \mathcal{U} \) then \( 1 \leq n(x) \leq N \) and

\[
\log\|A^{(n(x))}(x)\| \leq -2tn(x) \quad (4.12)
\]

We will show that (4.12) holds also for cocycles \( B \) in a small neighborhood of \( A \). This is the part where the argument differs from the case \( L_1(A) > -\infty \).

Let \( 0 < \delta < \frac{e^{-2nt}}{e\int N C^N - 1} \), and let \( B \in \mathcal{C}_m \) with \( \text{dist}(B,A) < \delta \), so \( \|B(x) - A(x)\| < \delta \) for \( \mu \) a.e. \( x \in X \).

Then clearly, for any \( 1 \leq m \leq N \) and for \( \mu \) a.e. \( x \in X \) we have:

\[
\|B^{(m)}(x) - A^{(m)}(x)\| < m C^{m-1} \delta \leq N C^{N-1} \delta < \frac{e^{-2nt}}{t} \leq \frac{e^{-2mt}}{t} \quad (4.13)
\]

For \( \mu \) a.e. \( x \in X \) and for every \( 1 \leq m \leq N \), there are two cases.

Case 1. \( \|A^{(m)}(x)\| \) is extremely small, i.e.

\[
\|A^{(m)}(x)\| < e^{-2tm}
\]
In this case, using (4.13) we get $\|B^{(m)}(x)\| < e^{-2tm} (1 + 1/t)$, so
\[
\log \|B^{(m)}(x)\| < -2tm + 1/t
\] (4.14)

**Case 2.** $\|A^{(m)}(x)\|$ has a lower bound:
\[
\|A^{(m)}(x)\| \geq e^{-2tm}
\]
Then using again (4.13)
\[
\log \|B^{(m)}(x)\| - \log \|A^{(m)}(x)\| \leq \frac{\|B^{(m)}(x) - A^{(m)}(x)\|}{\|A^{(m)}(x)\|}
\]
\[
\leq \frac{e^{-2tm}}{t} e^{2tm} = 1/t
\]

hence
\[
\log \|B^{(m)}(x)\| < \log \|A^{(m)}(x)\| + 1/t
\] (4.15)

Then if $m = n(x)$, using (4.15) in the second case and using directly (4.14) in the first case, we conclude that for $\mu$ a.e. $x \in \mathcal{U}$ we have
\[
\log \|B^{(n(x))}(x)\| \leq -2tn(x) + 1/t \leq (-2t + 1/t) n(x)
\] (4.16)

which is the analogue of (4.6) when $L_1(A) > -\infty$.

The rest of the proof then follows exactly the same pattern as when $L_1(A) > -\infty$, the role of $L_1(A) + 2\epsilon$ being now played by $-2t + 1/t$, while the small set $\emptyset$ is simply $\mathcal{U}$, since there was no extra small set excluded when deriving (4.16). □

**Remark 4.1.** Note that since our cocycles are in $L^\infty$, proposition 4.1 above also implies the upper semicontinuity of the top Lyapunov exponent as a function of the cocycle. In particular, this gives continuity at cocycles $A$ with $L_1(A) = -\infty$, and since $L_1(A) \geq L_2(A) \geq \ldots \geq L_m(A)$, this implies that every Lyapunov exponent is continuous at $A$.

Therefore, from now on, we may assume that $L_1(A) > -\infty$.

The main application of proposition 4.1 is the following lemma, which we use repeatedly throughout the inductive argument. It gives us a lower bound on the gap between the first two singular values of the iterates of a cocycle, thus ensuring the gap condition in the Avalanche Principle.

Throughout this paper, if $A \in \mathcal{C}$ is such that $L_1(A) > L_2(A) \geq -\infty$, then $\kappa(A)$ will denote the gap between the first two Lyapunov exponents, i.e. $\kappa(A) := L_1(A) - L_2(A) > 0$ if $L_2(A) > -\infty$, while if $L_2(A) = -\infty$ then $\kappa(A)$ will be a fixed, large enough finite constant.

**Lemma 4.2.** Let $A \in \mathcal{C}_m$ be a cocycle for which $L_1(A) > L_2(A)$ and let $\epsilon > 0$. There are $\delta_0 = \delta_0(A, \epsilon) > 0$, $n_0 = n_0(A, \epsilon) \in \mathbb{N}$ and
\[ \tau = \tau(A, \epsilon) \in \mathcal{I} \text{ such that for all } B \in \mathcal{C}_m \text{ with } \text{dist}(B, A) < \delta_0 \text{ and for all } n \geq n_0, \text{ if} \]

\[ \left| L_1^{(n)}(B) - L_1^{(n)}(A) \right| < \theta \]  

(4.17)

then for all phases \( x \) outside a set of measure \( < \iota_n \) we have:

\[ \frac{1}{n} \log \rho(B^{(n)}(x)) > \kappa(A) - 2\theta - 3\epsilon \]  

(4.18)

Moreover,

\[ L_1^{(n)}(B) - L_2^{(n)}(B) > (\kappa(A) - 2\theta - 3\epsilon)(1 - \iota_n) \]  

(4.19)

**Proof.** Fix \( \epsilon > 0 \). If \( L_2(A) = -\infty \), let \( t = t(A) := -2L_1(A) + \kappa(A) \).

Since \( L_1(A) > L_2(A) \), the cocycle \( A \) satisfies a uniform fiber-LDT with a parameter \( p = p(A, \epsilon) \in P \) and in a neighborhood around \( A \) of size \( \delta(A, \epsilon) > 0 \).

The compatibility condition holds for all cocycles in \( \mathcal{C} \), hence also for \( \wedge_2 A \). Note that \( L_1(\wedge_2 A) = L_1(A) + L_2(A) \), hence \( L_1(\wedge_2 A) > -\infty \) iff \( L_2(A) > -\infty \).

The nearly uniform upper semicontinuity proposition \([4.1]\) can then be applied to \( \wedge_2 A \), and it gives parameters \( \delta > 0, \tau \in \mathcal{I}, n_0 \in \mathbb{N} \) that define the range of validity of \([4.1]\) and \([4.2]\) respectively. These parameters depend on \( A \) and \( \epsilon \) when \( L_2(A) > -\infty \) and only on \( A \) when \( L_2(A) = -\infty \).

Pick \( \delta = \delta(A, \epsilon) > 0, n_0 = n_0(A, \epsilon) \in \mathbb{N}, t = t(A, \epsilon) \in \mathcal{I} \) such that both the uniform fiber-LDT and proposition \([4.1]\) apply for all cocycles \( B \in \mathcal{C}_m \) with \( \text{dist}(B, A) < \delta \), for all \( n \geq n_0 \) and for all \( x \) outside a set of measure \( < \iota_n \). Fix such \( B, n, x \).

For any matrix \( g \in \text{Mat}(m, \mathbb{R}) \) we have

\[ \rho(g) = \frac{s_1(g)}{s_2(g)} = \frac{\|g\|^2}{\|\wedge_2 g\|} \]  

(4.20)

Note that \( \rho(g) \) could be \(+\infty\), in which case the corresponding lower bounds are trivial.

From \([4.20]\) we get

\[ \frac{1}{n} \log \rho(B^{(n)}(x)) = 2 \frac{1}{n} \log \|B^{(n)}(x)\| - \frac{1}{n} \log \|\wedge_2 B^{(n)}(x)\| \]  

(4.21)

Use the uniform fiber-LDT to obtain a lower bound on the first term on the right hand side of \([4.21]\):

\[ \frac{1}{n} \log \|B^{(n)}(x)\| > L_1^{(n)}(B) - \epsilon_n > L_1^{(n)}(B) - \epsilon \]

Moreover, from assumption \([4.17]\) we have

\[ L_1^{(n)}(B) > L_1^{(n)}(A) - \theta \geq L_1(A) - \theta \]
hence
\[ \frac{1}{n} \log \| B^{(n)}(x) \| > L_1(A) - \theta - \epsilon \quad (4.22) \]

We use proposition 4.1 for the cocycle \( \land_2 A \) to obtain an upper bound on \( \frac{1}{n} \log \| \land_2 B^{(n)}(x) \| \).

If \( L_2(A) > -\infty \), so \( L_1(\land_2 A) > -\infty \), from part (i) of proposition 4.1 we get
\[ \frac{1}{n} \log \| \land_2 B^{(n)}(x) \| < L_1(B) + \epsilon = L_1(A) + L_2(A) + 2\epsilon \quad (4.23) \]

Combine (4.21), (4.22), (4.23) to conclude that for all chosen \( B, n, x \) we have:
\[ \frac{1}{n} \log \rho(B^{(n)}(x)) > \kappa(A) - 2\theta - 3\epsilon \quad (4.24) \]
which proves (4.18). Since this estimate holds for all \( x \) outside a set of measure \( \eta_n \), by integrating in \( x \) we derive (4.19) as well, with a factor \( (1 - 2\eta_n) \) and provided \( \kappa(A) - 2\theta - 3\epsilon \geq \kappa(A)/2 \).

Now if \( L_2(A) = -\infty \), so \( L_1(\land_2 A) = -\infty \), use part (ii) of proposition 4.1 to get
\[ \frac{1}{n} \log \| \land_2 B^{(n)}(x) \| < -t = 2L_1(A) - \kappa(A) \quad (4.25) \]

Combine (4.21), (4.22), (4.25) and get (4.18) in this case as well. Then (4.19) follows as above. \( \square \)

5. The proof of continuity of the Lyapunov exponents

We are in the context of the main theorem 2.1: a given ergodic system \((X, \mathcal{F}, \mu, T)\), a space of measurable cocycles \( \mathcal{C} \), a set of observables \( \Xi \) and a set of LDT parameters \( \mathcal{P} \) with corresponding spaces of deviation functions \( \mathcal{E} \) and \( \mathcal{I} \). We assume the compatibility condition 2.4 between \( \Xi \) and any cocycle \( A \in \mathcal{C} \), the base-LDT for any observable \( \xi \in \Xi \), the uniform \( L^p \)-boundedness condition (put \( p = 2 \) for simplicity of notation) on any cocycle \( A \in \mathcal{C} \) with \( L_1(A) > -\infty \) and the uniform fiber-LDT for any cocycle \( A \in \mathcal{C} \) with \( L_1(A) > L_2(A) \). These LDT estimates hold for parameters \( p \in \mathcal{P} \).

Finite scale uniform continuity.

Proposition 5.1 (finite scale uniform continuity). Let \( A \in \mathcal{C}_m \) be a cocycle for which \( L_1(A) > L_2(A) \). There are \( \delta_0 = \delta_0(A) > 0 \), \( n_{01} = n_{01}(A) \), \( C_1 = C_1(A) > 0 \) and \( \iota = \iota(A) \in \mathcal{I} \) such that for any two
cocycles $B_1, B_2 \in \mathcal{C}_n$ with $\text{dist}(B_i, A) \leq \delta_0$ where $i = 1, 2$, if $n \geq n_0$, and $\text{dist}(B_1, B_2) < e^{-C_1 n}$, then

$$\left| L_1^{(n)}(B_1) - L_1^{(n)}(B_2) \right| < \epsilon_n^{1/2} \quad (5.1)$$

**Proof.** Let $\epsilon_0 := \kappa(A)/10 > 0$. Since $L_1(A) > L_2(A)$, the uniform fiber-LDT and lemma 4.2 hold for $A$, $\epsilon_0$. Choose parameters $p = p(A) \in \mathcal{P}$, $p = (n_0, \epsilon \leq \underline{\epsilon})$ and $\delta_0 = \delta_0(A) > 0$ such that $\epsilon_{n_0} \leq \epsilon_0$ and lemma 4.2 and the fiber-LDT hold for all cocycles $B \in \mathcal{C}_n$ with $\text{dist}(B, A) \leq \delta_0$ and for all $n \geq n_0$.

Let $C_0 = C_0(A) > 0$ such that for all such cocycles $B$ we have $\|B\|_{L\infty} \leq e^{C_0}$ and for all $n \geq 1$,

$$\left| L_1^{(n)}(B) \right| \leq \left\| \frac{1}{n} \log \|B^{(n)}(x)\| \right\|_{L^2} \leq C_0$$

Pick $C_1 > 2C_0 + \epsilon_0$ and $n_0 \geq n_0$ such that $e^{-C_1 n_0} < \delta_0$.

Let $n \geq n_0$ and $B_i \in \mathcal{C}_n$ with $\text{dist}(B_i, A) \leq \delta_0$ ($i = 1, 2$) be arbitrary but fixed. Assume that $\text{dist}(B_1, B_2) < e^{-C_1 n}$.

Apply the fiber-LDT to each $B_i$ and conclude that for all $x$ outside a set $\mathcal{B}_n$, with $\mu(\mathcal{B}_n) < \epsilon_n$ we have:

$$\frac{1}{n} \log \|B_i^{(n)}(x)\| > L_1^{(n)}(B_i) - \epsilon_n \geq -C_0 - \epsilon_0 \quad (5.2)$$

Let $\mathcal{B}_n = \mathcal{B}_n^1 \cup \mathcal{B}_n^2$, so $\mu(\mathcal{B}_n) < 2 \epsilon_n$ and if $x \in \mathcal{B}_n^c$ then we have:

$$\|B_1^{(n)}(x)\|, \|B_2^{(n)}(x)\| > e^{-(C_0 + \epsilon_0) n} \quad (5.3)$$

Moreover, for $\mu$ - a.e. $x \in \mathcal{B}_n^c$ and all $0 \leq j \leq n - 1$ we also have:

$$\|B_1(T^j x) - B_2(T^j x)\| \leq \text{dist}(B_1, B_2) < e^{-C_1 n}$$

Therefore, for $\mu$ - a.e. $x \in \mathcal{B}_n^c$ we get:

$$\left| \frac{1}{n} \log \|B_1^{(n)}(x)\| - \frac{1}{n} \log \|B_2^{(n)}(x)\| \right| = \frac{1}{n} \left| \log \frac{\|B_1^{(n)}(x)\|}{\|B_2^{(n)}(x)\|} \right| \leq \frac{1}{n} \frac{\|B_1^{(n)}(x) - B_2^{(n)}(x)\|}{\min\{\|B_1^{(n)}(x)\|, \|B_2^{(n)}(x)\|\}} \leq \frac{1}{n} \sum_{j=0}^{n-1} e^{(C_0 + \epsilon_0) n} \|B_2^{(n-j-1)}(T^{j+1} x)\| \|B_1(T^j x) - B_2(T^j x)\| \|B_1^{(j)}(x)\| \leq \frac{1}{n} \sum_{j=0}^{n-1} e^{(C_0 + \epsilon_0) n} e^{C_0 (n-j-1)} e^{-C_1 n} e^{C_0 j} \leq e^{-n(C_1 - 2C_0 - \epsilon_0)}$$
Integrating in $x$ we conclude:
\[
\int_{\mathbb{R}^n} \left| \frac{1}{n} \log \| B_1^{(n)}(x) \| - \frac{1}{n} \log \| B_2^{(n)}(x) \| \right| \mu(dx) < e^{-n(C_1 - 2C_0 - \epsilon_0)} \tag{5.4}
\]

By Cauchy-Schwarz we have
\[
\int_{\mathbb{R}^n} \left| \frac{1}{n} \log \| B_1^{(n)}(x) \| - \frac{1}{n} \log \| B_2^{(n)}(x) \| \right| \mu(dx) \leq \left\| \frac{1}{n} \log \| B_1^{(n)}(x) \| \right\|_{L^2} \cdot \mu(B_n)^{1/2} + \left\| \frac{1}{n} \log \| B_2^{(n)}(x) \| \right\|_{L^2} \cdot \mu(B_n)^{1/2}
\]

hence
\[
\int_{\mathbb{R}^n} \left| \frac{1}{n} \log \| B_1^{(n)}(x) \| - \frac{1}{n} \log \| B_2^{(n)}(x) \| \right| \mu(dx) \lesssim C_0 \frac{1}{n} \tag{5.5}
\]

Since $1 \in I$ decays at most exponentially, we may of course assume that $e^{-n(C_1 - 2C_0 - \epsilon_0)} < \frac{1}{n^{1/2}}$, so (5.4) and (5.5) imply
\[
|L_1^{(n)}(B_1) - L_1^{(n)}(B_2)| \leq \int_X \left| \frac{1}{n} \log \| B_1^{(n)}(x) \| - \frac{1}{n} \log \| B_2^{(n)}(x) \| \right| \mu(dx) < \frac{1}{n} \tag{5.6}
\]

which proves (5.1).

\[\square\]

**The inductive step procedure.** In this subsection we derive the main technical result used to prove our continuity theorem, an inductive tool based on the avalanche principle 3.6, the uniform fiber-LDT 2.7 and the nearly uniform upper semicontinuity 4.1. All estimates involving two consecutive scales $n_0, n_1$ of the inductive procedure will carry errors of order at most $\frac{n_0}{n_1}$. We begin with a simple lemma which shows that we may always assume that $n_1$ is a multiple of $n_0$, otherwise an extra error term of the same order is accrued.

**Lemma 5.2.** Let $A \in C$ be an $L^1$-bounded cocycle, and let $C$ be its $L^1$-bound. If $n_0, n_1, n, r \in \mathbb{N}$ are such that $n_1 = n \cdot n_0 + r$ and $0 \leq r \leq n_0$, then
\[
-2C \frac{n_0}{n_1} + L_1^{(n_1)n_0}(A) \leq L_1^{(n_1)}(A) \leq L_1^{(n_0)}(A) + 2C \frac{n_0}{n_1} \tag{5.6}
\]

**Proof.** The cocycle $A$ is $L^1$-bounded by $C$ hence for all $m \geq 1$ we have
\[
|L_1^{(m)}(A)| \leq C
\]

Since $n_1 = n \cdot n_0 + r$ and $r \geq 0$, we can write
\[
A^{(n_1)}(x) = A^{(r)}(T^{n_0}x) \cdot A^{(n_0)}(x)
\]

hence
\[
\|A^{(n_1)}(x)\| \leq \|A^{(r)}(T^{n_0}x)\| \cdot \|A^{(n_0)}(x)\|
\]
Taking logarithms, dividing by $n_1$ then integrating in $x$ we get:

$$L_1^{(n_1)}(A) \leq \frac{n \eta}{n_1} L_1^{(n \eta)}(A) + \frac{r}{n_1} L_1^{(r)}(A)$$

which then implies

$$L_1^{(n_1)}(A) - L_1^{(n \eta)}(A) \leq \frac{r}{n_1} [L_1^{(r)}(A) - L_1^{(n \eta)}(A)] \leq 2C \frac{r}{n_1}$$

which proves the right hand side of (5.6).

Now write $(n + 1) n_0 = n_1 + q$, where $q = n_0 - r$, so $0 \leq q \leq n_0$. Then

$$A^{((n+1)n_0)}(x) = A^q(T^{m_1}x) \cdot A^{(n_1)}(x)$$

hence

$$\|A^{((n+1)n_0)}(x)\| \leq \|A^q(T^{m_1}x)\| \|A^{(n_1)}(x)\|$$

Taking logarithms, dividing by $(n + 1) n_0$ then integrating in $x$ we get:

$$L_1^{((n+1)n_0)}(A) \leq \frac{n_1}{(n + 1) n_0} L_1^{(n_1)}(A) + \frac{q}{(n + 1) n_0} L_1^{(q)}(A)$$

which then implies

$$L_1^{((n+1)n_0)}(A) - L_1^{(n_1)}(A) \leq \frac{q}{(n + 1) n_0} [L_1^{(q)}(A) - L_1^{(n_1)}(A)] \leq 2C \frac{1}{(n + 1)}$$

which proves the left hand side of (5.6).

\[\square\]

Lemma 5.3. Let $B \in C$ satisfying a fiber-LDT with parameter $p = (n_0, \varepsilon, L) \in \mathcal{P}$. Let $m_1, m_2, n \in \mathbb{N}$ and $\eta > 0$ be such that $m_i \geq n \geq n_0$ for $i = 1, 2$ and

$$|L_1^{(m_2+m_1)}(B) - L_1^{(m_i)}(B)| < \eta$$

Then

$$\|B^{(m_2+m_1)}(x)\| \|B^{(m_2)}(T^{m_1}x)\| \|B^{(m_1)}(x)\| > e^{-(m_1+m_2)(\eta+2\epsilon_n)} \quad (5.7)$$

for all $x$ outside a set of measure $< 3 \iota_n$.

Proof. Applying (one side inequality in) the fiber-LDT to the cocycle $B$ at scale $m_2 + m_1$, for all $x$ outside a set of measure $< \iota_{m_2+m_1} < \iota_n$, we have:

$$\frac{1}{m_2 + m_1} \log \|B^{(m_2+m_1)}(x)\| > L_1^{(m_2+m_1)}(B) - \epsilon_{m_2+m_1} \geq L_1^{(m_2+m_1)}(B) - \epsilon_n$$

hence

$$\|B^{(m_2+m_1)}(x)\| > e^{(m_1+m_2)} L_1^{(m_1+m_2)}(B) - (m_2+m_1) \epsilon_n \quad (5.8)$$
Applying (the other side inequality in) the fiber-LDT to the cocycle $B$ at scales $m_2, m_1$, for all $x$ outside a set of measure $< \iota_{m_2} + \iota_{m_1} < 2\iota_n$, we have:

\[
\frac{1}{m_1} \log \|B^{(m_1)}(x)\| < L_1^{(m_1)}(B) + \epsilon_{m_1} < L_1^{(m_1)}(B) + \epsilon_n
\]

and

\[
\frac{1}{m_2} \log \|B^{(m_2)}(T^{m_1}x)\| < L_1^{(m_2)}(B) + \epsilon_{m_2} < L_1^{(m_2)}(B) + \epsilon_n
\]

hence

\[
\|B^{(m_1)}(x)\| < e^{m_1L_1^{(m_1)}(B)+m_1\epsilon_n} \quad (5.9)
\]

\[
\|B^{(m_2)}(T^{m_1}x)\| < e^{m_2L_1^{(m_2)}(B)+m_2\epsilon_n} \quad (5.10)
\]

Combining (5.8), (5.9) and (5.10), for all $x$ outside a set of measure $< 3\iota_n$ we have:

\[
\frac{\|B^{(m_2+m_1)}(x)\|}{\|B^{(m_2)}(T^{m_1}x)\| \cdot \|B^{(m_1)}(x)\|} > e^{m_1(L^{(m_2+m_1)}_1(B)-L^{(m_1)}_1(B))} + m_2(L^{(m_2+m_1)}_1(B)-L^{(m_1)}_1(B))-2(m_2+m_1)\epsilon_n
\]

\[
> e^{-(m_1+m_2)(\eta+2\epsilon_n)}
\]

which proves the lemma.

\[\square\]

**Proposition 5.4** (inductive step procedure). Let $A \in C_m$ be a measurable cocycle such that $\kappa(A) := L_1(A) - L_2(A) > 0$. Fix $0 < \epsilon < \kappa(A)/20$.

There are $C = C(A) > 0$, $\delta = \delta(A, \epsilon) > 0$, $n_{00} = n_{00}(A, \epsilon) \in \mathbb{N}$, $\iota = \iota(A, \epsilon) \in \mathbb{I}$ such that for any $n_0 \geq n_{00}$, if the inequalities

(a) $L_1^{(n_0)}(B) - L_1^{(2n_0)}(B) < \eta_0$ 

(b) $|L_1^{(n_0)}(B) - L_1^{(n_0)}(A)| < \theta_0$ 

hold for a cocycle $B \in C_m$ with $\text{dist}(B, A) < \delta$, and if the positive numbers $\eta_0, \theta_0$ satisfy

\[
4\eta_0 + 2\theta_0 < \kappa(A) - 12\epsilon 
\]

then for any integer $n_1$ such that

\[
n_0^{1+} \leq n_1 \leq n_0 \cdot \iota_{n_0}^{-1/2} 
\]

we have

\[
|L_1^{(n_1)}(B) + L_1^{(n_0)}(B) - 2L_1^{(2n_0)}(B)| < C_{n_0} n_1 
\]




Furthermore,

(a++) \( L_1^{(n_1)}(B) - L_1^{(2n_1)}(B) < \eta_1 \) \hspace{1cm} (5.16)

(b++) \( |L_1^{(n_1)}(B) - L_1^{(n_1)}(A)| < \theta_1 \) \hspace{1cm} (5.17)

where

\[ \theta_1 = \theta_0 + 4\eta_0 + C\frac{n_0}{n_1} \] \hspace{1cm} (5.18)

\[ \eta_1 = C\frac{n_0}{n_1} \] \hspace{1cm} (5.19)

Proof. Since \( L_1(A) > L_2(A) \), the cocycle \( A \) satisfies a uniform fiber-LDT. Moreover, lemma \ref{lemma4.2} also applies.

Pick \( \delta = \delta(A,\epsilon) > 0 \), \( n_0 = n_0(A,\epsilon) \in \mathbb{N} \), \( \iota = \iota(A,\epsilon) \in \mathcal{I} \) such that for any \( n \geq n_0 \), both the uniform fiber-LDT and lemma \ref{lemma4.2} apply for all cocycles \( B \in C_m \) with \( \text{dist}(B, A) < \delta \), for all \( n \geq n_0 \) and for all \( x \) outside a set of measure \( < \iota_n \).

Let \( n_0 \geq n_00 \) and assume (5.11) and (5.12) hold for all cocycles \( B \in C_m \) with \( \text{dist}(B, A) < \delta \).

From (5.11), applying lemma \ref{lemma5.3}, we have:

\[ \frac{\|B^{(n_0)}(x)\|}{\|B^{(n_0)}(T^{n_0}x)\| \cdot \|B^{(n_0)}(x)\|} > e^{-n_0(2\eta_0+4\epsilon n_0)} \geq e^{-n_0(2\eta_0+4\epsilon)} =: \varepsilon \hspace{1cm} (5.20) \]

for all \( x \) outside a set of measure \( < 3\iota_n \).

(The reader should mind the difference between \( \epsilon \) and \( \varepsilon \).)

This estimate will ensure that the angles condition in the avalanche principle (proposition \ref{proposition3.6}) holds. Moreover, due to the assumption (5.12), applying lemma \ref{lemma4.2} for \( x \) outside a set of measure \( < \iota_n \), we have:

\[ \rho(B^{(n_0)}(x)) > e^{n_0(\kappa(A)-2\theta_0-3\epsilon)} =: \frac{1}{\kappa} \hspace{1cm} (5.21) \]

which will ensure that the gaps condition in the avalanche principle also holds.

Let \( \mathcal{B}_{n_0} \) be the union of the exceptional sets in (5.20) and (5.21). To simplify notations, replace the deviation set measure function \( \iota \) by \( 2\iota \), so we may assume \( \mu(\mathcal{B}_{n_0}) < \iota_n \) (we will tacitly do this throughout the paper).

Let \( n_1 \) be an integer such that \( n_0^{1+} \leq n_1 \leq n_0 \cdot \iota_n^{-1/2} \). Since \( \iota(t) \) decreases at least like \( t^{-c} \) (as \( t \to \infty \)) for some \( c > 0 \), and since \( \iota \) depends on \( \epsilon \) and \( A \), \( n_00 \) might need to be chosen larger, depending on \( \epsilon \) and \( A \) so that if \( n_0 \geq n_00 \) then the integer interval \([n_0^{1+}, n_0 \cdot \iota_n^{-1/2}] \) is large enough.

Moreover, due to lemma \ref{lemma5.2} we may assume that \( n_1 = n \cdot n_0 \) for some \( n \in \mathbb{N} \). To see this, note that once (5.15) is proven for scales
that are multiples of \( n_0 \), in particular for the scales \( n_1 = n n_0 \) and \( n_1'' = (n + 1) n_0 \), then using (5.6) we derive (5.15) for any scale \( n_1 \) such that \( n n_0 \leq n_1 \leq (n + 1) n_0 \). Furthermore, (5.16) and (5.17) will be derived directly from (5.15).

For every \( 0 \leq i \leq n - 1 \) define

\[
g_i = g_i(x) := B^{(n_0)}(T^{i n_0} x)
\]

Then clearly \( g_i^{(n)} = B^{(n_1)}(x) \) and \( g_i \cdot g_{i-1} = B^{(2 n_0)}(T^{(i-1)n_0} x) \) for all \( 1 \leq i \leq n - 1 \).

Let \( \mathcal{B}_{n_0} := \bigcup_{i=0}^{n-1} T^{-i n_0} \mathcal{B}_{n_0} \), so \( \mu(\mathcal{B}_{n_0}) < n \tau_{n_0} \) and if \( x \notin \mathcal{B}_{n_0} \) then

\[\rho(g_i) > \frac{1}{\kappa} \quad \text{for all} \quad 0 \leq i \leq n - 1\]

\[\frac{\|g_i \cdot g_{i-1}\|}{\|g_i\| \cdot \|g_{i-1}\|} > \varepsilon \quad \text{for all} \quad 1 \leq i \leq n - 1\]

Note also that condition (5.13) implies \( \kappa \ll \varepsilon^2 \).

Therefore, we can apply the avalanche principle (proposition 3.6) and obtain:

\[
\left| \log \|g^{(n)}\| + \sum_{i=1}^{n-2} \log \|g_i\| - \sum_{i=1}^{n-1} \log \|g_i \cdot g_{i-1}\| \right| \lesssim n \cdot \frac{\kappa}{\varepsilon^2}
\]

But

\[
\frac{\kappa}{\varepsilon^2} = e^{-n_0} (\kappa(A)-4n_0-2\theta_0-11\epsilon) < e^{-\epsilon n_0}
\]

Since the deviation set measure functions \( \iota \in J \) decay at most exponentially fast, we may assume that our \( \iota \) was chosen already so that \( e^{-\epsilon t} \leq \iota(t) \) for \( t \geq n_0 \). Hence we have \( \frac{\kappa}{\varepsilon^2} < \tau_{n_0} \).

The avalanche principle applied to our data then becomes:

\[
\left| \log \|B^{(n_1)}(x)\| + \sum_{i=1}^{n-2} \log \|B^{(n_0)}(T^{i n_0} x)\| - \sum_{i=1}^{n-1} \log \|B^{(2 n_0)}(T^{(i-1)n_0} x)\| \right| \lesssim n \tau_{n_0}
\]

(5.22)

for all \( x \notin \mathcal{B}_{n_0} \), where \( \mu(\mathcal{B}_{n_0}) < n \tau_{n_0} \).

Now divide both sides of (5.22) by \( n_1 = n \cdot n_0 \) and get:

\[
\left| \frac{1}{n_1} \log \|B^{(n_1)}(x)\| + \frac{1}{n} \sum_{i=1}^{n-2} \frac{1}{n_0} \log \|B^{(n_0)}(T^{i n_0} x)\| - \frac{2}{n} \sum_{i=1}^{n-1} \log \frac{1}{2n_0} \|B^{(2 n_0)}(T^{(i-1)n_0} x)\| \right| \lesssim \tau_{n_0}
\]
Now integrate in $x$ and conclude:

$$\left| L_1^{(n_1)}(B) + \frac{n-2}{n} L_1^{(n_0)}(B) - \frac{2(n-1)}{n} L_1^{(2n_0)}(B) \right| \lesssim \epsilon_n + \frac{C(A) \epsilon_n}{n} < C \epsilon_n$$

The term on the left hand side of the above inequality can be rewritten in the form

$$\left| L_1^{(n_1)}(B) + L_1^{(n_0)}(B) - 2L_1^{(2n_0)}(B) - \frac{2}{n} [L_1^{(n_0)}(B) - L_1^{(2n_0)}(B)] \right|$$

hence we conclude:

$$\left| L_1^{(n_1)}(B) + L_1^{(n_0)}(B) - 2L_1^{(2n_0)}(B) \right| < C \epsilon_n + \frac{2}{n} [L_1^{(n_0)}(B) - L_1^{(2n_0)}(B)] < C \frac{n_0}{n_1}$$

Clearly the same argument leading to (5.22) will hold for $2n_1$ instead of $n_1$, which via the triangle inequality proves (5.16).

We can rewrite (5.23) in the form

$$\left| L_1^{(n_1)}(B) - L_1^{(n_0)}(B) + 2[L_1^{(n_0)}(B) - L_1^{(2n_0)}(B)] \right| < C \frac{n_0}{n_1}$$

Using (5.24) for $B$ and $A$ we get:

$$\left| L_1^{(n_1)}(B) - L_1^{(n_1)}(A) \right| < \left| L_1^{(n_1)}(B) - L_1^{(n_0)}(B) + 2[L_1^{(n_0)}(B) - L_1^{(2n_0)}(B)] \right|$$

$$+ \left| L_1^{(n_1)}(A) - L_1^{(n_0)}(A) + 2[L_1^{(n_0)}(A) - L_1^{(2n_0)}(A)] \right|$$

$$+ \left| L_1^{(n_0)}(B) - L_1^{(n_0)}(A) \right|$$

$$+ 2 \left| L_1^{(n_0)}(B) - L_1^{(2n_0)}(B) \right| + 2 \left| L_1^{(n_0)}(A) - L_1^{(2n_0)}(A) \right|$$

$$< \theta_0 + 4\eta_0 + C \frac{n_0}{n_1}$$

\[
\square
\]

**Theorem 5.1.** Let $A \in \mathcal{C}_m$ be a measurable cocycle for which $L_1(A) > L_2(A)$. Then the map $\mathcal{C}_m \ni B \mapsto L_1(B)$ is continuous at $A$ and the map $\mathcal{C}_m \ni B \mapsto L_1(B) - L_2(B)$ is lower semicontinuous at $A$.

**Proof.** Let $0 < \epsilon < \kappa(A)/100$ be arbitrary but fixed.

Since $L_1^{(n)}(A) \to L_1(A)$ as $n \to \infty$, there is $n_{02} = n_{02}(A, \epsilon) \in \mathbb{N}$ such that for all $n \geq n_{02}$ we have

$$L_1^{(n)}(A) - L_1^{(2n)}(A) < \epsilon$$

(5.25)

**General continuity theorem.**

**Theorem 5.1.** Let $A \in \mathcal{C}_m$ be a measurable cocycle for which $L_1(A) > L_2(A)$. Then the map $\mathcal{C}_m \ni B \mapsto L_1(B)$ is continuous at $A$ and the map $\mathcal{C}_m \ni B \mapsto L_1(B) - L_2(B)$ is lower semicontinuous at $A$.

**Proof.** Let $0 < \epsilon < \kappa(A)/100$ be arbitrary but fixed.

Since $L_1^{(n)}(A) \to L_1(A)$ as $n \to \infty$, there is $n_{02} = n_{02}(A, \epsilon) \in \mathbb{N}$ such that for all $n \geq n_{02}$ we have

$$L_1^{(n)}(A) - L_1^{(2n)}(A) < \epsilon$$

(5.25)
We will apply the inductive step proposition 5.4 repeatedly. We first choose the relevant parameters (which will depend on $A$ and $\epsilon$) so that both the inductive step proposition 5.4 and the finite scale continuity proposition 5.1 apply. The latter will ensure that the assumptions (5.11), (5.12) and (5.13) of the inductive step proposition 5.4 are satisfied for a large enough scale $n_0 = n_0(A, \epsilon)$, so we can start running the inductive argument with that scale.

Let $\iota \in I$ be the sum of the corresponding deviation measure functions in the inductive step proposition 5.4 and the finite scale continuity proposition 5.1.

Let $\delta_0$ be less than the size of the neighborhood of $A \in C_m$ from the inductive step proposition 5.4 and from the finite scale continuity proposition 5.1 respectively. Let $C_1$ be the constant in the finite scale continuity proposition 5.1 and let $C$ be the constant in the inductive step proposition 5.4.

Finally, let the scale $n_0$ be greater than the thresholds $n_{00}$ from the inductive step proposition 5.4, $n_{01}$ from the finite scale continuity proposition 5.1 and $n_{02}$ from (5.25) above. Moreover, assume $n_0$ to be large enough so that $e^{-C_1 2n_0} < \delta_0$, $\frac{1}{2} \tau_{n_0} < \epsilon$ and $n_0^{0+} \ll \tau_{n_0}^{-1/2}$ for $n \geq n_0$ and $C n_0^{-0} \lesssim \epsilon$.

Let $\delta := e^{-C_1 2n_0}$ and let $B \in C_m$ with dist($B, A$) $< \delta$.

Since $\delta = e^{-C_1 2n_0} < e^{-C_1 n_0}$, we can apply the finite scale continuity proposition 5.1 (with $B_2 = B$ and $B_1 = A$) at scales $2n_0$ and $n_0$ and get:

\[
\left| L^{(n_0)}_1(B) - L^{(2n_0)}_0(A) \right| < \frac{1}{2} \tau_{n_0}^{-1/2} =: \theta_0 < \epsilon \quad (5.26)
\]

\[
\left| L^{(2n_0)}_1(B) - L^{(2n_0)}_0(A) \right| < \frac{1}{2} \tau_{2n_0}^{1/2} = \theta_0 \quad (5.27)
\]

Then (5.25), (5.26), (5.27) imply

\[
L^{(n_0)}_1(B) - L^{(2n_0)}_1(A) < 2 \frac{1}{2} \tau_{n_0}^{-1/2} + \epsilon =: \eta_0 < 3 \epsilon \quad (5.28)
\]

The inequalities (5.28) and (5.27) imply the assumptions (5.11) and (5.12) in the inductive step proposition 5.4 and since

\[
2\theta_0 + 4\eta_0 < 2\epsilon + 12\epsilon = 14\epsilon < \kappa(A) - 12\epsilon
\]

the condition (5.13) between parameters is also satisfied.

We can apply the inductive step proposition 5.4 and conclude that

\[
L^{(n_0)}_1(B) - L^{(2n_0)}_1(B) > (\kappa(A) - \theta_0 - 2\epsilon) \cdot (1 - \tau_{n_0}) \quad (5.29)
\]
and say for $n_1 \preceq n_0^{1+}$ we have:

\[
L_1^{(n_1)}(B) - L_1^{(2n_1)}(B) < \eta_1 \quad (5.30)
\]

where

\[
\theta_1 = \theta_0 + 4\eta_0 + C \frac{n_0}{n_1} \quad (5.32)
\]

\[
\eta_1 = C \frac{n_0}{n_1} \quad (5.33)
\]

But

\[
2\theta_1 + 4\eta_1 = (2\theta_0 + 8\eta_0) + 4C \frac{n_0}{n_1} < (2\varepsilon + 24\varepsilon) + \varepsilon < \kappa(A) - 12\varepsilon
\]

hence the inductive step proposition 5.4 applies again and we get:

\[
L_1^{(n_1)}(B) - L_2^{(n_1)}(B) > (\kappa(A) - \theta_1 - 2\varepsilon) \cdot (1 - \iota_{n_1}) \quad (5.34)
\]

and say for $n_2 \preceq n_1^{1+}$ we have:

\[
L_1^{(n_2)}(B) - L_1^{(2n_2)}(B) < \eta_2 \quad (5.35)
\]

where

\[
\theta_2 = \theta_1 + 4\eta_1 + C \frac{n_1}{n_2} \quad (5.37)
\]

\[
\eta_2 = C \frac{n_1}{n_2} \quad (5.38)
\]

Moreover,

\[
2\theta_2 + 4\eta_2 = (2\theta_0 + 8\eta_0) + [10C \frac{n_0}{n_1} + 6C \frac{n_1}{n_2}] < (2\varepsilon + 24\varepsilon) + \varepsilon < \kappa(A) - 12\varepsilon
\]

It is now clear how we continue this procedure. Going from step $k$ to step $k + 1$, we choose a scale $n_{k+1} \preceq n_k^{1+}$ and we have:

\[
L_1^{(n_k)}(B) - L_2^{(n_k)}(B) > (\kappa(A) - \theta_k - 2\varepsilon) \cdot (1 - \iota_{n_k}) \quad (5.39)
\]

and

\[
L_1^{(n_{k+1})}(B) - L_1^{(2n_{k+1})}(B) < \eta_{k+1} \quad (5.40)
\]

\[
|L_1^{(n_{k+1})}(B) - L_1^{(n_{k+1})}(A)| < \theta_{k+1} \quad (5.41)
\]

where

\[
\eta_{k+1} = C \frac{n_k}{n_{k+1}} \quad (5.42)
\]
and

\[
\theta_{k+1} = \theta_k + 4\eta_k + C \frac{n_1}{n_2}
\]

\[
= (\theta_0 + 4\eta_0) + 5C \sum_{i=0}^{k-1} \frac{n_i}{n_{i+1}} + C \frac{n_k}{n_{k+1}}
\]

\[
< (\theta_0 + 4\eta_0) + 5C \sum_{i=0}^{\infty} \frac{n_i}{n_{i+1}}
\]

\[
< (\theta_0 + 4\eta_0) + 10C n_0^{-0} < (\epsilon + 12\epsilon) + 10\epsilon = 23\epsilon
\]

hence

\[
\theta_{k+1} < 23\epsilon \tag{5.43}
\]

Moreover

\[
2\theta_{k+1} + 8\eta_{k+1} = (2\theta_0 + 8\eta_0) + 10C \sum_{i=0}^{k-1} \frac{n_i}{n_{i+1}} + 6C \frac{n_k}{n_{k+1}}
\]

\[
< (2\theta_0 + 8\eta_0) + 10C \sum_{i=0}^{\infty} \frac{n_i}{n_{i+1}}
\]

\[
< (2\theta_0 + 8\eta_0) + 20C n_0^{-0} < (2\epsilon + 24\epsilon) + 20\epsilon = 46\epsilon
\]

hence

\[
2\theta_{k+1} + 8\eta_{k+1} < \kappa(A) - 12\epsilon
\]

which ensures that the inductive process runs indefinitely.

Now take the limit as \( k \to \infty \) in (5.41), and using (5.43) we have

\[
|L_1(B) - L_1(A)| \leq 23\epsilon
\]

which proves the continuity at \( A \) of the top Lyapunov exponent \( L_1 \).

Moreover, taking the limit as \( k \to \infty \) in (5.39), and using again (5.43) we have

\[
L_1(B) - L_2(B) \geq \kappa(A) - 23\epsilon - 2\epsilon = L_1(A) - L_2(A) - 25\epsilon
\]

which proves the lower semicontinuity at \( A \) of the gap between the first and second Lyapunov exponents.

\( \square \)

Note that estimate (5.41) in the proof of theorem 5.1 says that if the cocycle \( B \) is close enough to \( A \), then

\[
|L_1^{(n)}(B) - L_1^{(n)}(A)| \lesssim \epsilon \tag{5.44}
\]

holds for an increasing sequence of scales \( n = n_{k+1}, k \geq 0 \).

In fact, a slight modification of the argument shows that (5.44) holds in fact for all large enough scales \( n \).
Indeed, it is enough to first ensure that the base step of the inductive procedure, i.e. that the estimates
\[
\left| L_1^{(n_0)}(B) - L_0^{(n_0)}(A) \right| < \epsilon_0^{1/2} =: \theta_0 < \epsilon \\
\left| L_1^{(2n_0)}(B) - L_0^{(2n_0)}(A) \right| < \epsilon_0^{1/2} < \epsilon_0^{1/2} =: \theta_0
\]
hold not just for a single scale \( n_0 \), but for a whole (finite) interval of scales \( N_0 = [n_{00}, e^{2n_{00}}] =: [n_0^-, n_0^+] \), where \( n_{00} \) is greater than the applicability threshold of various estimates (e.g. uniform fiber-LDT, finite scale continuity etc).

Let \( \psi(t) := t^{1+} \) and define inductively the intervals of scales \( N_1 = [\psi(n_0^-), \psi(n_0^+)] =: [n_1^-, n_1^+] \), \( N_{k+1} = [\psi(n_k^-), \psi(n_k^+)] =: [n_{k+1}^-, n_{k+1}^+] \) for all \( k \geq 0 \).

It follows that if \( n = n_1 \in N_1 \), then \( n = n_1 \propto \psi(n_0) = n_0^{1+} \) for some \( n_0 \in N_0 \), and so (5.30) and (5.31) hold for all \( n_1 \in N_1 \).

Continuing inductively, for every \( k \geq 1 \), if \( n = n_{k+1} \in N_{k+1} \), then there is \( n_k \in N_k \) such that \( n = n_{k+1} \propto \psi(n_k) = n_k^{1+} \) and then (5.40) and (5.41) hold as well.

The intervals \( N_0 \) and \( N_1 \) overlap because
\[
n_1^- = \psi(n_0^-) = n_0^{1+} < e^{n_{00}} = n_0^+
\]
Then since \( \psi \) is increasing and \( N_{k+1} \propto \psi(N_k) \), the intervals \( N_k \) and \( N_{k+1} \) will overlap for all \( k \geq 0 \).

Therefore, if \( n \geq n_1^- \) then \( n \in N_{k+1} \) for some \( k \geq 0 \) and so (5.44) holds. This means, moreover, that we may apply lemma 4.2 at all such scales and conclude that for all \( x \) outside a set of measure \( < \iota_n \),
\[
\frac{1}{n} \log \rho(B^{(n)}(x)) > \kappa(A) - 5\epsilon
\]
We conclude that the following uniform, finite scale statement holds.

**Lemma 5.5.** Given a cocycle \( A \in \mathcal{C}_m \) with \( L_1(A) > L_2(A) \) and \( 0 < \epsilon < \kappa(A)/100 \), there are \( \delta = \delta(A, \epsilon) > 0 \) and \( n_0 = n_0(A, \epsilon) \in \mathbb{N} \), such that for all \( n \geq n_0 \) and for all \( B \in \mathcal{C}_m \) with \( \text{dist}(B, A) < \delta \) we have:
\[
\left| L_1^{(n)}(B) - L_1^{(n)}(A) \right| < \epsilon \tag{5.45}
\]
\[
\frac{1}{n} \log \rho(B^{(n)}(x)) > \kappa(A) - 5\epsilon \tag{5.46}
\]
for all \( x \) outside a set of measure \( < \iota_n \).

**Corollary 5.6.** For all \( m \geq 1 \), and for all \( 1 \leq k \leq m \), the Lyapunov exponents \( L_k : \mathcal{C}_m \to (-\infty, \infty) \) are continuous functions.
Proof. Let $A \in \mathcal{C}$ be a measurable cocycle. If $L_1(A) > L_2(A)$, then we can conclude, from theorem 5.1 above that $L_1$ is continuous at $A$ and that $L_1 - L_2$ is lower semicontinuous at $A$. 

If $A$ has a different gap pattern, by taking appropriate exterior powers, we can always reduce the problem to one where there is a gap between the first two Lyapunov exponents. 

For instance, if $L_1(A) = L_2(A) > L_3(A) \geq \ldots$, consider instead the cocycle $\wedge_2 A$. Clearly $L_1(\wedge_2 A) = L_1(A) + L_2(A)$ and $L_2(\wedge_2 A) = L_1(A) + L_3(A)$, hence $L_1(\wedge_2 A) - L_2(\wedge_2 A) = L_2(A) - L_3(A) > 0$, hence there is a gap between the first two Lyapunov exponents of $\wedge_2 A$. This will imply, using theorem 5.1 for $\wedge_2 A$, that the Lyapunov block $L_1 + L_2$ is continuous and the gap $L_2 - L_3$ is lower semi-continuous at $A$. 

This argument shows that given a cocycle $A \in \mathcal{C}$ with any gap pattern, the corresponding Lyapunov blocks are all continuous at $A$, while the corresponding gaps are lower semicontinuous at $A$. 

Moreover, the general assumptions made on the space of cocycles ensure that the map 

$$
\mathcal{C}_m \ni B \mapsto L_1(B) + \ldots + L_m(B) = \int_X \log |\det[B(x)]| \mu(dx)
$$

is continuous everywhere. 

It is then a simple exercise (see lemma 6.1 and theorem 6.2 in [10] for its solution) to see that this is all that is needed to conclude continuity of each individual Lyapunov exponent, irrespective of any gap pattern. \hfill \Box

Modulus of continuity. The following proposition, which is also interesting in itself, will be the main ingredient in obtaining the modulus of continuity of the top Lyapunov exponent. It gives the rate of convergence of the finite scale exponents $L_1^{(n)}(B)$ to the top Lyapunov exponent $L_1(B)$ and it gives an estimate on the proximity of these finite scale exponents at different scales. 

These estimates are uniform in a neighborhood of a cocycle $A \in \mathcal{C}_m$ for which $L_1(A) > L_2(A)$, and they depend on a deviation measure function $\xi = \xi(A) \in J$, which will be fixed in the beginning. Define the map $\psi(t) = \psi_\xi(t) := t \cdot [\xi(t)]^{-1/2}$, and let $\phi = \phi_\xi$ be its inverse. Moreover, for every integer $n \in \mathbb{N}$, denote $n++ := \lfloor \psi(n) \rfloor = \lfloor n \xi^{-1/2} \rfloor$ and $n-- := \lfloor \phi(n) \rfloor$, so $(n++) - - \approx n$. 

These estimates will be obtained by applying repeatedly the inductive step proposition 5.4. In order to obtain the sharpest possible estimate, when going from one scale to the next, we will make the greatest
possible jump, which is why we have defined the 'next scale' \( n++ \) above as \( \lfloor n^{-1/2} \rfloor \).

**Proposition 5.7** (uniform speed of convergence). Let \( A \in C_m \) be a measurable cocycle for which \( L_1(A) > L_2(A) \). There are \( \delta = \delta(A) > 0 \), \( C = C(A) \), \( n_{00} = n_{00}(A) \in \mathbb{N} \), \( \iota = \iota(A) \in I \) such that, with the above notations, for all \( n \geq n_{00} \) and for all \( B \in C_m \) with \( \text{dist}(B, A) < \delta \) we have:

\[
L_1^{(n)}(B) - L_1(B) < C \frac{\phi(n)}{n} \leq n^{-1/2} \quad (5.47)
\]

\[
| L_1^{(n++)}(B) + L_1^{(n)}(B) - 2L_1^{(2n)}(B) | < C \frac{n}{n++} \leq n^{1/2} \quad (5.48)
\]

**Proof.** To prove (5.47) it is enough to show, under similar constraints on \( B \) and \( n \), and for some function \( \iota = \iota(A) \in I \), that

\[
L_1^{(n)}(B) - L_1^{(2n)}(B) < C \frac{\phi(n)}{n} \quad (5.49)
\]

This would imply, for all \( k \geq 0 \),

\[
L_1^{(2^k n)}(B) - L_1^{(2^{k+1} n)}(B) < \frac{\phi(2^k n)}{2^k n} \quad (5.50)
\]

Since we assume that for all deviation measure functions \( \iota \in I \), the corresponding map \( \phi = \phi_\iota \) satisfies

\[
\lim_{t \to \infty} \frac{\phi_\iota(2t)}{\phi_\iota(t)} < 2
\]

there is \( 0 < r < 1 \) such that for large enough \( n \) we have \( \phi(2n) \leq 2r \phi(n) \), hence for all \( k \) we have:

\[
\frac{\phi(2^k n)}{2^k n} \leq r^k \frac{\phi(n)}{n}
\]

We can then sum up for \( k \) from 0 to \( \infty \) in (5.50) and derive (5.47).

To prove (5.49), (5.48) we follow the same procedure, based on the inductive step proposition 5.4, used in the proof the general continuity theorem 5.1 and of its extension lemma 5.5, but with some modifications. We will work again with intervals of scales instead of individual scales. We fix \( \epsilon := \kappa(A)/100 \) so all subsequent parameters, including the deviation measure function \( \iota \), will be fixed and dependent only upon the cocycle \( A \). Let \( \iota \in I \), \( \delta_0 > 0 \), \( C_1 > 0 \) and \( C > 0 \) be as in the beginning of the proof of theorem 5.1.

Choose \( n^-_{00} \in \mathbb{N} \) large enough that for \( n \geq n^-_{00} \) the inductive step proposition 5.4 and the finite scale continuity proposition 5.1 apply, and that \( L_1^{(n)}(A) - L_1^{(2n)}(A) < \epsilon_0 \).
Moreover, assume \( n_0^- \) to be large enough that \( e^{-C_1 e^{n_0^-}} < \epsilon_0, \nu_n^{1/2} < \epsilon_0 \), \( C(n_0^-)^{-0-} < \epsilon_0, n^{0+} \ll \nu_n^{-1/2} \) for \( n \geq n_0^- \) and since \( \nu_n \) decays at most exponentially, we may also assume that \( n_0^- \nu_n^{-1/2} < e^{n_0} \).

Now set \( n_0^+ := \lfloor e^{n_0^-} \rfloor, N_0 := [n_0^-, n_0^+] \) and \( \delta := e^{-C_1 2n_0^+} \).

The assumptions above ensure that for all cocycles \( B \in \mathcal{C}_m \) with \( \text{dist}(B, A) < \delta \), and for all \( n_0 \in N_0 \), since \( \delta = e^{-C_1 2n_0^+} \leq e^{-C_1 2n_0} < e^{-C_1 n_0} \), the finite scale continuity proposition \( 5.1 \) applies at scales \( 2n_0, n_0 \). This implies, as in the proof of theorem \( 5.1 \), the assumptions in the inductive step proposition \( 5.4 \) for every \( n_0 \in N_0 \).

Let \( n_1 := \lfloor \psi(n_0^-) \rfloor, n_1^+ := \lfloor \psi(n_0^+) \rfloor \) and \( N_1 := [n_1^-, n_1^+] \supseteq \psi(N_0) \).

We may assume that for every \( n_1 \in N_1 \) there is \( n_0 \in N_0 \) such that \( n_1 = n_0 \lfloor \nu_n^{-1/2} \rfloor \) \( \approx n_0 \nu_n^{-1} \) (this is because by lemma \( 5.2 \), the estimates involving scales \( n_1 \in N_1 \) which are not divisible by \( n_0 \) will only carry an additional error of order \( \frac{n_0}{n_1} \)).

We apply the inductive step proposition \( 5.4 \) and obtain:

\[
L_1^{(n_1)}(B) - L_1^{(2n_1)}(B) < C \frac{n_0}{n_1} \approx C \frac{\phi(n_1)}{n_1} \\
\left| L_1^{(n_1)}(B) + L_1^{(n_0)}(B) - 2L_1^{(2n_0)}(B) \right| < C \frac{n_0}{n_1} \approx C \nu_n^{1/2} \]

(Since \( n_1 \approx \psi(n_0) \), we have \( \phi(n_1) \approx n_0 \), as \( \phi \) is the inverse of \( \psi \).)

The procedure continues in the same way, with intervals of scales defined inductively by \( N_{k+1} = [n_k^-, n_k^+] \supseteq \psi(N_k) \) for all \( k \geq 0 \). Again, each two consecutive intervals of scales \( N_k \) and \( N_{k+1} \) overlap. Therefore, if \( n \geq n_1 \), then \( n = n_{k+1} \in N_{k+1} \) for some \( k \geq 0 \), so there is \( n_k \in N_k \) such that \( n_{k+1} = n_k \lfloor \nu_n^{-1/2} \rfloor \approx n_k + + \). We then have:

\[
L_1^{(n_{k+1})}(B) - L_1^{(2n_{k+1})}(B) < C \frac{n_k}{n_{k+1}} \approx C \frac{\phi(n_{k+1})}{n_{k+1}} \\
\left| L_1^{(n_{k+1})}(B) + L_1^{(n_k)}(B) - 2L_1^{(2n_k)}(B) \right| < C \frac{n_k}{n_{k+1}} \approx C \nu_n^{1/2} \]

which completes the proof.

\[\Box\]

The following theorem shows that locally near any cocycle \( A \in \mathcal{C}_m \) for which \( L_1(A) > L_2(A) \), the top Lyapunov exponent has a modulus of continuity given by a map that depends explicitly on a deviation measure function \( \lambda \), hence on the strength of the large deviation type estimates satisfied by the dynamical system.

**Theorem 5.2** (modulus of continuity). Let \( A \in \mathcal{C}_m \) be a measurable cocycle for which \( L_1(A) > L_2(A) \). There are \( \delta = \delta(A) > 0, \lambda = \)
for all cocycles $B_i \in \mathcal{C}_m$ with $\text{dist}(B_i, A) < \delta$, where $i = 1, 2$, we have

$$\left| L_1(B_1) - L_1(B_2) \right| \leq \left[ \varepsilon (c \log(1/\text{dist}(B_1, B_2))) \right]^{1/2} \quad (5.51)$$

**Proof.** Choose parameters $\delta_0 = \delta_0(A) > 0$, $n_{00} = n_{00}(A) \in \mathbb{N}$ and $\varepsilon = \varepsilon(A) \in \mathcal{J}$ such that both the finite scale uniform continuity proposition [5.1] and the uniform speed of convergence proposition [5.7] apply with deviation measure function $\varepsilon$ for all cocycles $B \in \mathcal{C}_m$ with $\text{dist}(B, A) < \delta_0$ and for all $n \geq n_{00}$.

Let $C_1 = C_1(A) > 0$ be the constant from proposition [5.1] and set $\delta := \min\{\delta_0, \varepsilon^{1/4} e^{-C_1 4n_{00}}\}$.

Let $B_i \in \mathcal{C}_m$ be measurable cocycles with $\text{dist}(B_i, A) < \delta$ ($i = 1, 2$) and put $\text{dist}(B_1, B_2) =: h$ ($< 2\delta \leq e^{-C_1 4n_{00}}$).

Set $n := \left[ \frac{1}{2c_2} \log(1/h) \right] \in \mathbb{N}$. Then clearly $e^{-C_1 4n} \leq h \leq e^{-C_1 2n}$ hence $\text{dist}(B_1, B_2) = h \leq e^{-C_1 2n}$ and $n \geq n_{00}$.

All of this preparation shows that we can apply the finite scale uniform continuity proposition [5.1] to $B_1, B_2$ at scales $n$ and $2n$ and get:

$$\left| L_1^{(n)}(B_1) - L_1^{(n)}(B_2) \right| < \varepsilon_n^{1/2} \quad (5.52)$$

$$\left| L_1^{(2n)}(B_1) - L_1^{(2n)}(B_2) \right| < \varepsilon_n^{1/2} \quad (5.53)$$

Since $\text{dist}(B_i, A) < \delta \leq \delta_0$ and $n \geq n_{00}$, we can also apply the uniform speed of convergence proposition [5.7] to $B_i$ ($i = 1, 2$) at scale $n$ and have:

$$L_1^{(n++)}(B_i) - L_1(B_i) < \varepsilon_n^{1/2} \quad (5.54)$$

$$\left| L_1^{(n++)}(B_1) + L_1^{(n)}(B_1) - 2L_1^{(2n)}(B_1) \right| < \varepsilon_n^{1/2} \quad (5.55)$$

Combining (5.52), (5.53), (5.54), (5.55) we conclude:

$$\left| L_1(B_1) - L_1(B_2) \right| \lesssim \varepsilon_n^{1/2} \leq \left[ \varepsilon (1/(2C_1) \log(1/h)) \right]^{1/2}$$

\[ \square \]

A more general statement than theorem [5.2] holds. We need to introduce the appropriate terminology first (see also [10] for more details).

If for some $1 \leq i < m$, $L_i(A) > L_{i+1}(A)$, we say that the Lyapunov spectrum of $A$ has a gap at dimension $i$. If the Lyapunov spectrum of $A$ has gaps at dimensions $1 \leq \tau_1 < \tau_2 < \ldots < \tau_k < m$, and if we denote by $\tau = (\tau_1, \tau_2, \ldots, \tau_k)$, then we say that the Lyapunov spectrum of $A$ has a gap pattern encoded by the signature $\tau$, or in short, a $\tau$-gap pattern. In particular, the case of simple top Lyapunov spectrum is encoded by the signature $\tau = (1)$. 
Let $A \in C$ and let $\tau = (\tau_1, \tau_2, \ldots, \tau_k)$ be a signature. We define the Lyapunov spectrum $\tau$-blocks of $A$ as:

$$L_{\pi,1}(A) := L_1(A) + \ldots + L_{\tau_1}(A)$$
$$L_{\pi,2}(A) := L_{\tau_1+1}(A) + \ldots + L_{\tau_2}(A)$$

and in general, with the convention that $\tau_0 = 0$,

$$L_{\pi,j}(A) := L_{\tau_j+1}(A) + \ldots + L_{\tau_{j+1}}(A)$$

for all $0 \leq j \leq k - 1$.

Since the cocycle $A \in C$ has a $\tau$-gap pattern, hence a gap at dimension $\tau_1$, and since

$$L_1(\wedge_{\tau_1} A) - L_2(\wedge_{\tau_1} A) = L_{\tau_1}(A) - L_{\tau_1+1}(A) > 0$$

then the cocycle $\wedge_{\tau_1} A \in C_{(\tau_1)}$ has a simple top Lyapunov exponent.

Therefore, theorem 5.2 applies to $\wedge_{\tau_1} A$ and it implies that the map

$$C_{(\tau_1)} \ni \hat{B} \mapsto L_1(\hat{B}) \in [-\infty, \infty)$$

has a modulus of continuity given by (5.51), in a neighborhood of $\wedge_{\tau_1} A$.

Since the map $C_m \ni B \mapsto \wedge_{\tau_1} B$ is locally Lipschitz, and since clearly

$$L_1(\wedge_{\tau_1} B) = L_1(B) + \ldots + L_{\tau_1}(B) = L_{\tau_1}(B),$$

we conclude that the map

$$C_m \ni B \mapsto L_{\tau_1}(B)$$

has a modulus of continuity similar to (5.51).

We can do the same with the next dimension $\tau_2$, where there is another gap in the spectrum of $A$. We will conclude that the map

$$C \ni B \mapsto L_1(B) + \ldots + L_{\tau_2}(B) \in [-\infty, \infty)$$

has a similar modulus of continuity to (5.51).

Then by subtraction, the same can be said about the second Lyapunov spectrum $\tau$-block $L_{\pi,2}$ and so on.

We have obtained the following general continuity statement, under precisely the same assumptions as in theorem 2.1.

**Theorem 5.3.** For every $m$, the maps $L_k: (C_m, \text{dist}) \to [-\infty, \infty)$, $1 \leq k \leq m$ are continuous everywhere. Moreover, if $A \in C_m$ has a $\tau$-gap pattern, then locally near $A$, the corresponding Lyapunov spectrum $\tau$-blocks have a modulus of continuity given by $\omega(h) := \left[ \frac{1}{h} (c \log \frac{1}{h}) \right]^{1-1/p}$ for some $\ell = \ell(A) \in I$ and $c = c(A) > 0$. This applies, in particular, to any individual Lyapunov exponent $L_k$ whenever $L_k(A)$ is simple.
6. LDT FOR RANDOM COCYCLES

In this section we prove an abstract LDT theorem that will be used in the next section to derive uniform base and fibre LDT estimates for random cocycles.

Let $\Sigma$ be a compact metric space, $\mathcal{F}$ its Borel $\sigma$-field and $K$ a Markov kernel on $(\Sigma, \mathcal{F})$. Recall that $K$ is a map that to each point $x \in \Sigma$ associates a probability measure $K_x$ on $\Sigma$. See Definition 2.8. We use the following notations for the integral of a $\mathcal{F}$-measurable function $h : \Sigma \to \mathbb{R}$ w.r.t. $K_x$

$$\int_{\Sigma} h dK_x = \int_{\Sigma} h(y) K(x, dy).$$

**Laplace-Markov Operators.** We review here some concepts and definitions of a theory developed by Le Page, Bougerol and Lacroix, following closely the book [15] of Hennion and Hervé.

Given a Banach space $(B, \|\cdot\|)$, we denote by $L(B)$ the Banach algebra of bounded operators on $B$. Given $Q \in L(B)$, the operator norm of $Q$ is defined by $\|Q\| := \sup_{\|f\|=1} \|Qf\|$, while its spectral radius, denoted by $r(Q)$, is the limit

$$r(Q) := \lim_{n \to +\infty} \|Q^n\|^{1/n} = \inf_{n \geq 1} \|Q^n\|^{1/n}.$$

Let $(L^\infty(\Sigma), \|\cdot\|_\infty)$ denote the Banach algebra of complex bounded $\mathcal{F}$-measurable functions with the sup norm $\|f\|_\infty = \sup_{x \in \Sigma} |f(x)|$.

**Definition 6.1.** The linear operator $Q_K : L^\infty(\Sigma) \to L^\infty(\Sigma)$

$$(Q_K f)(x) := \int_{\Sigma} f(y) K(x, dy),$$

determined by a Markov kernel $K$, is called a Markov operator.

Denoting by $\mathbf{1}$ the constant function $\mathbf{1}(x) \equiv 1$, we have $Q_K \mathbf{1} = \mathbf{1}$ and $\|Q_K\| = 1$, because $\int_{\Sigma} K(x, dy) = 1$. In particular the operator $Q_K$ has spectral radius $r(Q_K) = 1$. The adjoint of a Markov operator $Q_K$ is the linear map $Q_K^* : M(\Sigma) \to M(\Sigma)$ defined by $Q_K^* \nu = \int_{\Sigma} K_x d\nu(x)$, acting on the space $M(\Sigma)$ of complex finite measures. By definition, for every $f \in L^\infty(\Sigma)$ and $\nu \in M(\Sigma)$,

$$\langle f, Q_K^* \nu \rangle := \int_{\Sigma} f d(Q_K^* \nu) = \int_{\Sigma} (Q_K f) d\nu =: \langle Q_K f, \nu \rangle.$$

Moreover, since $K$ is a Markov kernel, the adjoint operator $Q_K^*$ maps probability measures to probability measures. A probability measure $\mu$ on $(\Sigma, \mathcal{F})$ is called $K$-stationary if and only if $Q_K^* \mu = \mu$. 

From now on, we assume that $\mu$ is a $K$-stationary probability on $(\Sigma, \mathcal{F})$. We extend the ‘strong mixing’ concept introduced in section 2.

Given a normed space $((\mathcal{B}, \|\cdot\|)$ with a bounded inclusion $\mathcal{B} \subset L^\infty(\Sigma)$, we say that $K$ acts continuously on $((\mathcal{B}, \|\cdot\|)$ if $Q_K(\mathcal{B}) \subseteq \mathcal{B}$, and there is $M > 0$ such that for all $f \in \mathcal{B}$, $\|Q_K(f)\| \leq M \|f\|$.

**Definition 6.2.** Given $C > 0$ and $0 < \rho < 1$, we say that $(K, \mu)$ is strongly mixing on $\mathcal{B}$, with constants $(C, \rho)$, if

(1) $K$ acts continuously on $\mathcal{B}$, and

(2) for every $f \in \mathcal{B}$ and $n \in \mathbb{N}$,

$$\|(Q_K)^n f - \langle f, \mu \rangle 1 \| \leq C \rho^n \|f\|.$$

Examples of strongly mixing systems arise naturally from Markov kernels satisfying the so called Doeblin condition. See [9]. We say that $K$ satisfies the Doeblin condition if there is a positive finite measure $\rho$ on $(\Sigma, \mathcal{F})$ and some $\varepsilon > 0$ such that for all $x \in \Sigma$ and $A \in \mathcal{F}$,

$$K(x, A) \geq 1 - \varepsilon \quad \Rightarrow \quad \rho(A) \geq \varepsilon.$$

Given $A \in \mathcal{F}$, define

$$L^\infty(A) := \{ f \in L^\infty(\Sigma, \mathcal{F}) : f|_{\Sigma \setminus A} \equiv 0 \},$$

which is a closed Banach subspace of $(L^\infty(\Sigma), \|\cdot\|_\infty)$.

**Theorem 6.1.** Let $K$ be a Markov kernel on $(\Sigma, \mathcal{F})$. If $K$ satisfies the Doeblin condition then there is a ‘mixing period’ $p \in \mathbb{N}$, there are sets $\Sigma_1, \ldots, \Sigma_m$ in $\mathcal{F}$, and there are probability measures $\nu_1, \ldots, \nu_m$ in $(\Sigma, \mathcal{F})$ such that for all $i, j = 1, \ldots, m$,

(1) $\Sigma_i \cap \Sigma_j = \emptyset$ if $i \neq j$,

(2) $\Sigma_i$ is $K^p$-forward invariant,

(3) $\nu_i$ is $K^p$-stationary and ergodic with $\nu_i(\Sigma_j) = \delta_{ij}$,

(4) $\lim_{n \to +\infty} K^{pn}(x, \Sigma_1 \cup \ldots \cup \Sigma_m) = 1$, with geometric uniform speed of convergence, for all $x \in \Sigma$,

(5) $\nu(\Sigma_1 \cup \ldots \cup \Sigma_m) = 1$, for every $K^p$-stationary probability $\nu$,

(6) $(K^p, \nu_i)$ is strongly mixing on $L^\infty(\Sigma_i)$.

**Proof.** See section V-5 of the book [9].

An operator $Q : \mathcal{B} \to \mathcal{B}$ is called quasi-compact if there is a $Q$-invariant decomposition in two closed subspaces $\mathcal{B} = F \oplus H$, with $\dim F < +\infty$ and $r(Q|_H) < r(Q)$. The spectral radius $r(Q|_H)$ will be referred as the inner spectral radius of $Q$. A quasi-compact operator $Q$ is called simple if it has a unique eigenvalue, denoted by $\lambda(Q)$, of modulus equal to $r(Q)$.

A complex Banach space, of complex valued functions, $\mathcal{B}$, is called a lattice if for all $f \in \mathcal{B}$, $\overline{f}, |f| \in \mathcal{B}$.
Proposition 6.1. Let $K$ be a Markov kernel on $(\Sigma, \mathcal{F})$ which acts continuously on a Banach lattice $(\mathcal{B}, \|\cdot\|)$ with a bounded embedding in $L^\infty(\Sigma)$. Then the following statements are equivalent:

1. $K$ is strongly mixing on $\mathcal{B}$,
2. $Q_K|_\mathcal{B} : \mathcal{B} \to \mathcal{B}$ is quasi-compact and simple.

Proof. The proof is left as an exercise. \qed

Definition 6.3. Given a measurable function $\xi : \Sigma \to \mathbb{R}$, the operator $Q_{K, \xi, z} : L^\infty(\Sigma) \to L^\infty(\Sigma)$ defined by

$$(Q_{K, \xi, z} f)(x) := \int_{\Sigma} f(y) e^{z \xi(y)} K(x, dy),$$

is called the Laplace-Markov operator of the pair $(K, \xi)$.

Of course some assumptions must be imposed on $\xi$ for this operator to be well-defined.

Basic Assumptions. Because the setting in [15] is too general for our purposes, we make here stronger assumptions to derive an abstract LDT theorem, from which both base and fiber LDT theorems will be deduced.

Consider a scale of complex Banach algebras $(\mathcal{B}_\alpha, \|\cdot\|_\alpha)$ of bounded measurable functions indexed in $\alpha \in [0, 1]$. More precisely, we assume there are seminorms $v_\alpha : \mathcal{B}_\alpha \to [0, +\infty[$ such that for all $0 \leq \alpha \leq 1$,

1. $\|f\|_\alpha = v_\alpha(f) + \|f\|_\infty$, for all $f \in \mathcal{B}_\alpha$,
2. $\mathcal{B}_0 = L^\infty(\Sigma)$, and $\|\cdot\|_0$ is equivalent to $\|\cdot\|_\infty$,
3. $\mathcal{B}_\alpha$ is a lattice, i.e., if $f \in \mathcal{B}_\alpha$ then $|f|, |f| \in \mathcal{B}_\alpha$,
4. $\mathcal{B}_\alpha$ is a Banach algebra with unity $1 \in \mathcal{B}_\alpha$ and $v_\alpha(1) = 0$.

Moreover we assume that this family is a scale in the sense that for all $0 \leq \alpha_0 < \alpha_1 < \alpha_2 \leq 1$,

1. $\mathcal{B}_{\alpha_2} \subset \mathcal{B}_{\alpha_1} \subset \mathcal{B}_{\alpha_0}$,
2. $v_{\alpha_0}(f) \leq v_{\alpha_1}(f) \leq v_{\alpha_2}(f)$, for all $f \in \mathcal{B}_{\alpha_2}$,
3. $v_{\alpha_1}(f) \leq v_{\alpha_0}(f)^{\alpha_2-\alpha_0} v_{\alpha_2}(f)^{\alpha_1-\alpha_0}$, for all $f \in \mathcal{B}_{\alpha_2}$.

An example of such a scale is formed by the Banach algebras $\mathcal{K}_\alpha(\Sigma)$, of Hölder continuous functions on $\Sigma$, with exponent $\alpha$, w.r.t. some normalized distance $d : \Sigma \times \Sigma \to [0, 1]$. The norms on these spaces are defined as follows: for all $\alpha \in [0, 1]$ and $f \in L^\infty(\Sigma)$,

$$\|f\|_\alpha := v_\alpha(f) + \|f\|_\infty, \quad \text{with} \quad v_\alpha(f) := \sup_{x, y \in \Sigma, x \neq y} \frac{|f(x) - f(y)|}{d(x, y)^\alpha}.$$
The spaces $\mathcal{K}_\alpha(\Sigma) := \{ f \in L^\infty(\Sigma) : v_\alpha(f) < +\infty \}$ satisfy the assumptions (B1)-(B7). See for instance [21].

Let us call observed Markov system to any triple $(K, \mu, \xi)$, where $(K, \mu)$ is a Markov system on $\Sigma$, and $\xi : \Sigma \to \mathbb{R}$ is a measurable observable. Consider a family of observed Markov systems $\{ (K_\gamma, \mu_\gamma, \xi_\gamma) \}_{\gamma \in \mathcal{X}}$, and write $Q_{\gamma,z} := Q_{K_\gamma, \mu_\gamma, \xi_\gamma}$, $\lambda_\gamma(z) := \lambda(Q_{\gamma,z})$ (the maximum eigenvalue of $Q_{\gamma,z}$), and $c_\gamma(t) := \log \lambda_\gamma(t)$, for every index $\gamma \in \mathcal{X}$.

Assume $C, M, \theta, a, \beta, \alpha_0, \alpha_1, \sigma$ are positive constants with

$$0 < \alpha_1 < \frac{\alpha_0}{2}, \quad 0 < \beta < a, \quad 0 < \sigma < 1 \quad (6.1)$$

such that the following properties hold for $\gamma, \gamma_1, \gamma_2 \in \mathcal{X}$, $\alpha_1 \leq \alpha \leq \alpha_0$ and $f \in \mathcal{B}_\alpha$:

(A1) (simple quasi-compactness) $\|Q_{\gamma,0}^n f - (f, \mu_\gamma) 1\|_\alpha \leq C \sigma^n \|f\|_\alpha$.

(A2) (continuous action) $\|Q_{\gamma,0}(e^{a|\xi|})\|_\infty \leq M$.

(A3) (analytic action) for all $z \in \mathbb{D}_\beta := \{ w \in \mathbb{C} : |w| < \beta \}$, $f \in \mathcal{B}_\alpha$ and $i = 0, 1, 2$,

(a) $Q_{\gamma,z}(f \xi_i^\gamma) \in \mathcal{B}_\alpha$,

(b) $v_\alpha(Q_{\gamma,z}(f \xi_i^\gamma)) \leq M \|f\|_\alpha$.

(A4) (Hölder continuity) $\|Q_{\gamma_1,z}f - Q_{\gamma_2,z}f\|_\infty \leq M \|f\|_\alpha d(\gamma_1, \gamma_2)^\theta$.

By (A2), $Q_{\gamma,z}$ acts continuously on $L^\infty(\Sigma)$, for $z \in \mathbb{D}_\alpha$. From (A3) it follows that $Q_{\gamma,z}$ acts continuously on $\mathcal{B}_\alpha$, for all $z \in \mathbb{D}_\beta$. Moreover, these conditions imply that $Q_\gamma : \mathbb{D}_\beta \to \mathcal{L}(\mathcal{B}_\alpha)$, $z \mapsto Q_{\gamma,z}$, is an analytic function. (A1) says that $(K_\gamma, \mu_\gamma)$ is strongly mixing on $\mathcal{B}_\alpha$, with constants $\sigma(C, \sigma)$, i.e., the operator $Q_{\gamma,0}|_{\mathcal{B}_\alpha} : \mathcal{B}_\alpha \to \mathcal{B}_\alpha$ is quasi-compact and simple. Finally from (A4), the function $\gamma \mapsto \lambda_\gamma(z)$ is uniformly Hölder continuous. In the sequel we prove all these facts uniformly in $\gamma \in \mathcal{X}$.

**Proposition 6.2.** Assume (A1) for all $\alpha_1 \leq \alpha \leq \alpha_0$ and suppose that

1. $K_\gamma = K$ and $\mu_\gamma = \mu$ are fixed, for all $\gamma \in \mathcal{X}$,
2. the function $\mathcal{X} \to \mathcal{B}_{\alpha_0}, \gamma \mapsto \xi_\gamma$, is bounded and Lipschitz.

Then for any $a > 0$ and $\beta > 0$ there is $M > 0$ such that the conditions (A2)-(A4) hold for any $\alpha_1 \leq \alpha \leq \alpha_0$.

**Proof.** Fix $a, \beta > 0$, let $L = \sup\{ \|\xi_\gamma\|_{\alpha_0} : \gamma \in \mathcal{X} \}$ and take

$M := \max \{ C \sigma, e^{aL}, C \sigma e^{BL}, C \sigma L e^{BL}, C \sigma L^2 e^{BL} \}$.

To prove (A2) note that

$$\|Q_K(e^{a|\xi|})\|_\infty \leq \|e^{a|\xi|}\|_\infty \leq e^{a\|\xi\|_\infty} \leq M.$$
For (A3) we use (A1) and the fact that \((B, ||·||_α)\) is a Banach algebra. Given \(z \in D_β\), and \(α_1 \leq α \leq α_0\),

\[
v_α(Q_γ,z(f_γ^i)) = v_α(Q_K(f_γ^i e^{z_γ})) ≤ C σ \|f_γ^i e^{z_γ}\|_α \\
≤ C σ \|f\|_α · ||ξ_γ||_α · e^{z_γ} \|_α \\
≤ C σ \|f\|_α · ||ξ_γ||_α · _α · e^β \|_α ≤ M \|f\|_α .
\]

Let us now prove (A4). By (A1),

\[
\|Q_K f\|_α ≤ C σ \|f\|_α ≤ M \|f\|_α .
\]

By (2), \(\|ξ_γ − ξ_2\|_α ≤ d(γ_1, γ_2)\), for all \(γ_1, γ_2 \in X\). Thus for all \(z \in D_β\),

\[
\|Q_γ,z f − Q_γ,z f\|_∞ ≤ \|Q_K(1) e^{z_1} - e^{z_2}\|_α \\
≤ M \|f\|_α \|e^{z_1} - e^{z_2}\|_α \\
≤ M e^β \|f\|_α \|e^{z_1} - z_2 - 1\|_α \\
≤ M e^β \|f\|_α (e^β \|ξ_1 - ξ_2\|_α - 1) \\
≤ M e^β \|f\|_α \|ξ_1 - ξ_2\|_α \\
≤ M e^β \|f\|_α d(γ_1, γ_2) .
\]

Hence, enlarging \(M\) if necessary, (A4) holds. □

**Proposition 6.3.** For all \(γ \in X\), \(z \in D_β\) and \(f \in L^∞(Σ)\),

\[
\|Q_γ,z(f)\|_∞ ≤ M \|f\|_∞ \quad (6.2) \\
\|Q_γ(z f_γ)\|_∞ ≤ \frac{M}{(a - β)} e \|f\|_∞ \quad (6.3) \\
\|Q_γ(z f_γ^2)\|_∞ ≤ \frac{4 M}{(a - β)^2 e^2} \|f\|_∞ \quad (6.4)
\]

**Lemma 6.4.** Given \(0 < β < a\), for all \(x \in ]0, +∞[\),

\[
x e^{βx} ≤ \frac{e^{αx}}{(a - β) e} \quad (6.5) \\
x^2 e^{βx} ≤ \frac{e^{αx}}{(a - β)^2 e^2} . \quad (6.6)
\]

**Proof.** It is enough to check that the absolute minima of the proper functions \(f(x) = x^{-1} e^{(a - β)x}\) and \(g(x) = x^{-2} e^{(a - β)x}\) on \(]0, +∞[\) are respectively equal to \(\frac{1}{(a - β)e}\) and \(\frac{1}{(a - β)^2 e^2}\). □
Proof of Proposition 6.3. Inequality (6.2) follows from (A2). To prove (6.4) we use (A2) and (6.6)

\[ |Q_{\gamma,z}(f \xi^2)| \leq \int_{\Sigma} |f(t)| |\xi_{\gamma}(y)|^2 e^{(\text{Re}z)\xi(y)} K_{\gamma}(x,dy) \]

\[ \leq \|f\|_{\infty} \int_{\Sigma} |\xi_{\gamma}(y)|^2 e^{\beta |\xi_{\gamma}(y)|} K_{\gamma}(x,dy) \]

\[ \leq \frac{4 \|f\|_{\infty}}{(a-\beta)^2 e^2} \int_{\Sigma} e^a |\xi_{\gamma}(y)| K_{\gamma}(x,dy) \leq \frac{4 M}{(a-\beta)^2 e^2} \|f\|_{\infty} . \]

Analogously, (6.3) follows from (A2) and (6.5). □

Let

\[ M' := \max \left\{ M, \frac{M}{(a-\beta) e}, \frac{4 M}{(a-\beta)^2 e^2} \right\} \]

Proposition 6.5. For all \( \gamma \in \mathcal{X} \), \( \alpha_1 \leq \alpha \leq \alpha_0 \), \( z \in \mathbb{D}_\beta \), \( f \in \mathcal{B}_\alpha \) and \( i = 0, 1, 2 \),

\[ \|Q_{\gamma,z}(f \xi_i)| \|_{\alpha} \leq M' \|f\|_{\alpha} . \]

Proof. Combine Proposition 6.3 with (A1). □

It follows from the previous proposition that for \( i = 0, 1, 2 \) and \( z \in \mathbb{D}_\beta \), \( f \mapsto Q_{\gamma,z}f \) is a bounded operator in \( \mathcal{L}(\mathcal{B}_\alpha) \). In particular the function \( Q_{\gamma} : \mathbb{D}_\beta \to \mathcal{L}(\mathcal{B}_\alpha) \) is well-defined.

Proposition 6.6. The function \( Q_{\gamma} : \mathbb{D}_\beta \to \mathcal{L}(\mathcal{B}_\alpha) \) is analytic with

\[ \left[ \frac{d^i}{dz^i} Q_{\gamma,z} \right](f) = Q_{\gamma,z}(f \xi^i) \text{ for } f \in \mathcal{B}_\alpha , \]

for all \( \gamma \in \mathcal{X} \), and \( \alpha_1 \leq \alpha \leq \alpha_0 \).

Proof. Given \( b \in \mathbb{R} \), for all \( z, z_0 \in \mathbb{C} \),

\[ \frac{e^{zb} - e^{z_0 b}}{z - z_0} - b e^{z_0 b} = \int_{z_0}^z b^2 e^{b} \frac{z - \zeta}{z - z_0} d\zeta . \]

This is the first order Taylor remainder formula for \( f(z) = e^{bz} \) at \( z = z_0 \). To shorten the notation let us write \( Q_z \) for \( Q_{\gamma,z} \), and \( \xi(z) \) for \( \xi_{\gamma}(z) \). Replacing \( b \) by \( \xi(y) \), multiplying by \( f(y) K_{\gamma}(x,dy) \) and integrating over \( \Sigma \) we get

\[ \frac{Q_z f - Q_{z_0} f}{z - z_0} - Q_{z_0} (f \xi) = \int_{z_0}^z Q_\zeta (f \xi^2) \frac{z - \zeta}{z - z_0} d\zeta . \]
Hence, using Proposition 6.5, for all \( z \in \mathbb{D}_\beta \),
\[
v_\alpha \left( \frac{Q_z f - Q_{z_0} f}{z - z_0} - Q_{z_0} (f \xi) \right) \leq \int_{z_0}^z v_\alpha(Q_\zeta (f \xi^2)) \left| \frac{z - \zeta}{z - z_0} \right| d\zeta \leq M' \|f\|_\alpha \left| z - z_0 \right|.
\]
Analogously, using Proposition 6.3, for all \( z \in \mathbb{D}_\beta \),
\[
\| \frac{Q_z f - Q_{z_0} f}{z - z_0} - Q_{z_0} (f \xi) \|_\infty \leq \int_{z_0}^z \|Q_\zeta (f \xi^2)\|_\infty \left| \frac{z - \zeta}{z - z_0} \right| d\zeta \leq M' \|f\|_\alpha \left| z - z_0 \right|,
\]
which proves that the following limit exists in \( \mathcal{L}(\mathcal{B}_\alpha) \),
\[
\lim_{z \to z_0} \frac{Q_z f - Q_{z_0} f}{z - z_0} = Q_{z_0} (f \cdot).
\]

When the Markov system \((K, \mu)\) is fixed, and the observable \( \xi \) varies in \( \mathbb{B}_\alpha \), we can say a bit more.

**Proposition 6.7.** Consider a family \( \{(K, \mu, \xi)\}_{\xi \in \mathbb{B}_\alpha} \) of observed Markov systems and assume it satisfies (A1). Then for all \( \alpha_1 \leq \alpha \leq \alpha_0 \),

1. \( Q_{\xi, z} \) acts as a bounded operator on \( \mathcal{B}_\alpha \), for all \( z \in \mathbb{C} \),
2. The map \( \mathcal{B}_\alpha \times \mathbb{C} \ni (\xi, z) \mapsto Q_{\xi, z} \in \mathcal{L}(\mathcal{B}_\alpha) \) is analytic.

**Proof.** For all \( z \in \mathbb{C} \), \( e^{z \xi} \in \mathbb{B}_\alpha \) and hence the multiplication by this function is a bounded operator on the Banach algebra \( \mathcal{B}_\alpha \). The lemma follows because the Laplace-Markov operator \( Q_{\xi, z} \) is the composition of this multiplication operator with the Markov operator \( Q_K \), i.e., \( Q_{\xi, z}(f) = Q_K(e^{z \xi} f) \). Item (2) is a simple exercise. \( \square \)

Let us now focus on the proof that \( Q_z = Q_{\gamma, z} \) is quasi-compact and simple for \( z \in \mathbb{D}_\beta \).

**Proposition 6.8.** Consider a family \( \{(K_{\gamma}, \mu_{\gamma}, \xi_{\gamma})\}_{\gamma \in \mathcal{X}} \) of observed Markov systems satisfying (A1)-(A4). Given \( \varepsilon > 0 \) there exist \( M'', C' > 0 \) and \( 0 < \beta_0 < \beta \), depending only on \( \varepsilon, \sigma, \beta, C, M \), such that for all \( \gamma \in \mathcal{X} \), \( z \in \mathbb{D}_{\beta_0} \) and \( \alpha_1 \leq \alpha \leq \alpha_0 \) there is a one dimensional subspace \( E_{\gamma, z} \subset \mathcal{B}_\alpha \) a hyperplane \( H_{\gamma, z} \subset \mathcal{B}_\alpha \), \( \lambda_{\gamma}(z) \in \mathbb{C} \), and a linear map \( P_{\gamma, z} : \mathcal{B}_\alpha \to E_{\gamma, z} \) satisfying for all \( \gamma \in \mathcal{X} \), \( z \in \mathbb{D}_{\beta_0} \) and \( \alpha_1 \leq \alpha \leq \alpha_0 \)

1. \( \mathcal{B}_\alpha = E_{\gamma, z} \oplus H_{\gamma, z} \) is a \( Q_{\gamma, z} \)-invariant decomposition,
2. \( P_{\gamma, z} \) is a projection onto \( E_{\gamma, z} \), parallel to \( H_{\gamma, z} \),
3. \( Q_{\gamma, z} \circ P_{\gamma, z} = P_{\gamma, z} \circ Q_{\gamma, z} = \lambda_{\gamma}(z) P_{\gamma, z} \),
4. \( Q_{\gamma, z} f = \lambda_{\gamma}(z) f \) for all \( f \in E_{\gamma, z} \),
\( \lambda : \mathbb{D}_{\beta_0} \to \mathbb{C} \) is analytic in a neighbourhood of \( \mathbb{D}_{\beta_0} \).

Furthermore, for all \( \gamma, \gamma_1, \gamma_2 \in X \), \( z \in \mathbb{D}_{\beta_0} \) and \( 0 < \alpha \leq \alpha_0 \)

(a) \( |\lambda_{\gamma}(z)| \geq 1 - \varepsilon \),
(b) \( \|Q_{\gamma,z}^{n} - \lambda_{\gamma}(z)^n P_{\gamma,z} f\|_\alpha \leq C' (\sigma + \varepsilon)^n \|f\|_\alpha \),
(c) \( |\lambda_{\gamma_1}(z) - \lambda_{\gamma_2}(z)| \leq M'' d(\gamma_1, \gamma_2)^{\frac{\varepsilon}{2}} \).

Define the operators
\[
P_{\gamma,z} := \frac{1}{2\pi i} \int_{\Gamma_1} R_{\gamma,z}(w) \, dw \quad \text{(6.8)}
\]
\[
L_{\gamma,z} := \frac{1}{2\pi i} \int_{\Gamma_1} w R_{\gamma,z}(w) \, dw \quad \text{(6.9)}
\]
\[
N_{\gamma,z} := \frac{1}{2\pi i} \int_{\Gamma_0} w R_{\gamma,z}(w) \, dw \quad \text{(6.10)}
\]

where \( \Gamma_0 \) and \( \Gamma_1 \) are the positively oriented circles
\[
\Gamma_0 = \{ \, w \in \mathbb{C} : \, |w| = \frac{1 + 2\sigma}{3} \, \},
\]
\[
\Gamma_1 = \{ \, w \in \mathbb{C} : \, |w - 1| = \frac{1 - \sigma}{3} \, \},
\]
and \( R_{\gamma,z}(w) \) stands for the resolvent of \( Q_{\gamma,z} \), i.e., \( R_{\gamma,z}(w) := (w I - Q_{\gamma,z})^{-1} \).

We prove below, see Lemma \( \text{[6.10]} \) that for all \( w \notin \text{int}(\Gamma_0) \cup \text{int}(\Gamma_1) \), the norm \( \|R_{\gamma,z}(w)\| \) is uniformly bounded. This implies that the spectrum \( \Sigma_{\gamma,z} \) of \( Q_{\gamma,z} \) is contained in \( \text{int}(\Gamma_0) \cup \text{int}(\Gamma_1) \), and hence we can write \( \Sigma_{\gamma,z} = \Sigma_{\gamma,z}^0 \cup \Sigma_{\gamma,z}^1 \) with \( \Sigma_{\gamma,z}^i \subset \text{int}(\Gamma_i) \), for \( i = 0, 1 \). By the spectral theory of bounded operators on Banach spaces, see for instance chapter IX in [25], if we denote by \( H_{\gamma,z} \) and \( E_{\gamma,z} \) the subspaces of \( \mathcal{B}_\alpha \), respectively associated to the spectrum components \( \Sigma_{\gamma,z}^0 \) and \( \Sigma_{\gamma,z}^1 \), then for all \( z \in \mathbb{D}_{\beta_0} \), with \( \beta_0 > 0 \) small enough,

(1) the operators \( Q_{\gamma,z}, P_{\gamma,z}, L_{\gamma,z} \) and \( N_{\gamma,z} \) commute,
(2) \( L_{\gamma,z} f = Q_{\gamma,z} f \in E_{\gamma,z} \), for all \( f \in E_{\gamma,z} \),
(3) \( N_{\gamma,z} f = Q_{\gamma,z} f \in H_{\gamma,z} \), for all \( f \in H_{\gamma,z} \),
(4) \( Q_{\gamma,z} = L_{\gamma,z} + N_{\gamma,z} \),
(5) \( \mathcal{B}_\alpha = E_{\gamma,z} \oplus H_{\gamma,z} \),
(6) \( P_{\gamma,z} \) is the projection to \( E_{\gamma,z} \) parallel to \( H_{\gamma,z} \).

For \( z = 0 \), the condition (A1) implies that the operator \( Q_{\gamma,0}|_{E_{\alpha}} \) is quasi-compact and simple, with spectrum \( \Sigma_{\gamma,0}^0 \subset \mathbb{D}_\sigma \) and \( \Sigma_{\gamma,0}^1 = \{1\} \). Since 1 is a simple eigenvalue, \( E_{\gamma,0} = \{1\} \) is the space of constant functions. We must have \( H_{\gamma,0} = \{ f \in \mathcal{B}_\alpha : \int f \, d\mu_{\gamma} = 0 \} \) because the operator \( Q_{\gamma,0} \) acts invariantly on this space, as a contraction with
skeletal radius \( \leq \sigma \). Thus for all \( f \in \mathcal{B}_\alpha \), \( P_{\gamma,0}f = (\int f \, d\mu_\gamma) \mathbf{1} \) and \( N_{\gamma,0}f = Q_{\gamma,0}f - (\int f \, d\mu_\gamma) \mathbf{1} \). Since 1 is a simple eigenvalue of \( Q_{\gamma,0} \), a continuity argument implies that \( \Sigma_{\gamma,z}^1 \) is a singleton, i.e., \( \Sigma_{\gamma,z}^1 = \{ \lambda_z(z) \} \), for all \( z \in \mathbb{D}_\beta \). It follows easily that \( \dim(E_{\gamma,z}) = 1 \), and \( \lambda_\gamma(z) = \langle L_{\gamma,z} \mathbf{1}, \mu \rangle / \langle P_{\gamma,z} \mathbf{1}, \mu \rangle \). By perturbation theory, and Proposition 6.6, the function \( \lambda_\gamma : \mathbb{D}_\beta \to \mathbb{C} \) is analytic. Hence, to finish the proof of Proposition 6.8, it is enough to establish items (a), (b) and (c) of this proposition. First we need a couple of lemmas.

**Lemma 6.9.** Given \( T, T_0 \in \mathcal{L}(\mathcal{B}) \), if \( T_0 \) is invertible with \( \|T_0^{-1}\| \leq C \), and \( \|T - T_0\| \leq \varepsilon \) then

1. \( T \) is invertible, with \( \|T^{-1}\| \leq \frac{C}{1 - C\varepsilon} \);
2. \( \|T^{-1} - T_0^{-1}\| \leq \frac{C^2}{1 - C\varepsilon} \|T - T_0\| \).

**Lemma 6.10.** There exist constants \( C_0 > 0 \) and \( 0 < \beta < \beta_0 \), depending only on \( C, M, \sigma \) and \( \beta \), such that for \( \gamma \in \mathcal{X} \), \( z \in \mathbb{D}_{\beta_0} \), and any of the operators \( T_{\gamma,z} \in \{ Q_{\gamma,z}, R_{\gamma,z}(w), L_{\gamma,z}, N_{\gamma,z}, P_{\gamma,z} \} \), with \( w \notin \text{int}(\Gamma_0) \cup \text{int}(\Gamma_1) \),

1. \( \|T_{\gamma,z}\| \leq C_0 \),
2. \( \|T_{\gamma,z} - T_{\gamma,0}\| \leq C_0 \|z\| \).

**Proof.** First note that \( \|L_{\gamma,0}\| = \|P_{\gamma,0}\| = 1 \) and \( \|N_{\gamma,0}\| \leq C\sigma \), so that \( \|Q_{\gamma,0}\| = \|L_{\gamma,0} + N_{\gamma,0}\| \leq 1 + C\sigma \). Let us go through the given operators, one at a time. Assume \( 0 < \beta_0 < \beta \) is small and take \( z \in \mathbb{D}_{\beta_0} \). For \( Q_{\gamma,z} \), item (1) follows from Proposition 6.5 taking \( C_0 := M' \), while (2) follows from Propositions 6.5 and 6.6 with the same constant. For the operator \( R_{\gamma,z}(w) \), we have

\[
R_{\gamma,0}(w) = w^{-1} (I - w^{-1} Q_{\gamma,0})^{-1} = w^{-1} \sum_{n=0}^{\infty} \frac{Q_{\gamma,0}^n}{w^n} = w^{-1} \sum_{n=0}^{\infty} \frac{P_{\gamma,0}^n}{w^n} + w^{-1} \sum_{n=0}^{\infty} \frac{N_{\gamma,0}^n}{w^n} = \frac{P_{\gamma,0}}{w - 1} + \sum_{n=0}^{\infty} \frac{N_{\gamma,0}^n}{w^{n+1}}.
\]

Notice also that \( w \notin \text{int}(\Gamma_0) \cup \text{int}(\Gamma_1) \) implies \( |w - 1| \geq \frac{1 - \sigma}{3} \) and \( |w| \geq \frac{1 + 2\sigma}{3} \), and hence

\[
\|R_{\gamma,0}(w)\| \leq \frac{\|P_{\gamma,0}\|}{|w - 1|} + C \frac{\sum_{n=0}^{\infty} \left( \frac{\sigma}{|w|} \right)^n}{|w|} \leq \frac{3}{1 - \sigma} + \frac{3C}{1 + 2\sigma} \sum_{n=0}^{\infty} \left( \frac{3\sigma}{1 + 2\sigma} \right)^n = \frac{3 + 3C}{1 - \sigma} =: C_1.
\]
Therefore, applying Lemma \[6.9\] to \(w I - Q_{\gamma,z}\) and \(w I - Q_{\gamma,0}\), item (1) holds with \(C_2 := \frac{c_1}{1 - c_1 c_0 \beta_0}\), while (2) holds with \(C_3 := \frac{c_2^2 c_0}{1 - c_1 c_0 \beta_0}\). Of course we have to pick \(0 < \beta_0 < \beta\) small enough to make the denominators in the constants \(C_2\) and \(C_3\) are positive. For the remaining operators \(P_{\gamma,z}\), \(L_{\gamma,z}\) and \(N_{\gamma,z}\) we use the integral formulas \(6.8\), \(6.9\) and \(6.10\) to reduce to the previous case, using the same constants \(C_2\) and \(C_3\) as before.

**Proof of Proposition 6.8** Take \(0 < \beta_0 < \beta\) according to Lemma \(6.10\).

Fixing a probability measure \(\mu\) on \(\Sigma\), we can write, for all \(z \in \mathbb{D}_\beta\),

\[
\lambda_\gamma(z) = \frac{\langle L_{\gamma,z} 1, \mu \rangle}{\langle P_{\gamma,z} 1, \mu \rangle}.
\]  

(6.11)

Note that by Lemma \(6.10\) for all \(\gamma \in \mathcal{X}\),

\[
\langle P_{\gamma,z} 1, \mu \rangle \geq 1 - \|P_{\gamma,z} 1 - P_{\gamma,0} 1\|\alpha \geq 1 - C_0 \beta_0.
\]

Hence, for all \(z \in \mathbb{D}_\beta\),

\[
\left|\lambda_\gamma(z) - 1\right| \leq \frac{\left|\langle L_{\gamma,z} 1, \mu \rangle - \langle L_{\gamma,0} 1, \mu \rangle\right|}{\langle P_{\gamma,z} 1, \mu \rangle}.
\]

\[
\leq \frac{\left|\langle L_{\gamma,z} 1 - L_{\gamma,0} 1, \mu \rangle\right|}{1 - C_0 \beta_0} + \frac{C_0\left|\langle P_{\gamma,z} 1 - P_{\gamma,0} 1, \mu \rangle\right|}{1 - C_0 \beta_0}
\]

\[
\leq \frac{C_0 \beta_0}{1 - C_0 \beta_0} + \frac{C_0^2 \beta_0}{(1 - C_0 \beta_0)^2} = O(\beta_0).
\]

Thus, given \(\varepsilon > 0\) we can make \(\beta_0 > 0\) small enough so that for all \(\gamma \in \mathcal{X}\), and all \(z \in \mathbb{D}_\beta\), \(\left|\lambda_\gamma(z) - 1\right| < \varepsilon\). This implies (a). To prove (b), choose \(p \in \mathbb{N}\) such that \(C \sigma^p \leq (\sigma + \varepsilon)^p\), and make \(\beta_0 > 0\) small enough so that

\[
p C_0^{p-1} \delta_0 < (\sigma + \varepsilon)^p - (\sigma + \varepsilon)^2 = O(\varepsilon).
\]

We have then

\[
\|N_{\gamma,z}^p\| \leq \|N_{\gamma,0}^p\| + \|N_{\gamma,z}^p - N_{\gamma,0}^p\|
\]

\[
\leq C \sigma^p + p C_0^{p-1} \|N_{\gamma,z} - N_{\gamma,0}\| \leq C \sigma^p + p C_0^{p-1} \delta_0
\]

\[
\leq C \sigma^p + (\sigma + \varepsilon)^p - (\sigma + \varepsilon)^2 < (\sigma + \varepsilon)^p.
\]

It follows that for all \(n \in \mathbb{N}\), \(\|N_{\gamma,z}^n\| \leq C_0^{p-1} (\sigma + \varepsilon)^n\). This proves (b) with \(C' = C_0^{p-1}\). Finally, to prove (c), we claim that for all \(\gamma_1, \gamma_2 \in \mathcal{X}\), \(z \in \mathbb{D}_\beta\), \(2 \alpha_1 \leq \alpha \leq \alpha_0\), and \(f \in \mathcal{B}_{\alpha}\),

\[
v'_2(Q_{\gamma_1,z} f - Q_{\gamma_2,z} f) \lesssim \|f\|_\alpha d(\gamma_1, \gamma_2)^\frac{\theta}{2}.
\]  

(6.12)
In fact by (B7), (B2) and (A4), we have
\[
v_2\left(Q_{\gamma_1,z} f - Q_{\gamma_2,z} f\right) \leq v_0\left(Q_{\gamma_1,z} f - Q_{\gamma_2,z} f\right) \frac{1}{2} v_\alpha\left(Q_{\gamma_1,z} f - Q_{\gamma_2,z} f\right) \frac{1}{2}
\]
\[
\lesssim \left\|Q_{\gamma_1,z} f - Q_{\gamma_2,z} f\right\|_2 v_\alpha\left(Q_{\gamma_1,z} f - Q_{\gamma_2,z} f\right) \frac{1}{2}
\]
\[
\lesssim \|f\|_\alpha d(\gamma_1, \gamma_2)^{\frac{\theta}{2}}.
\]
Equation (6.12) implies, for all $\gamma_1, \gamma_2, z, \alpha$ and $f$ as above, and all $w \notin \text{int}(\Gamma_0) \cup \text{int}(\Gamma_1)$,
\[
v_2\left(R_{\gamma_1,z}(w) f - R_{\gamma_2,z}(w) f\right) \lesssim \|f\|_\alpha d(\gamma_1, \gamma_2)^{\frac{\theta}{2}}. \tag{6.13}
\]
This follows from (6.12), Lemma 6.10, and the algebraic relation
\[
R_{\gamma_1,z}(w) - R_{\gamma_2,z}(w) = -R_{\gamma_1,z}(w) \circ (Q_{\gamma_1,z} - Q_{\gamma_2,z}) \circ R_{\gamma_2,z}(w).
\]
Thus, integrating (6.8) and (6.9), we obtain
\[
\|P_{\gamma_1,z} f - P_{\gamma_2,z} f\|_2 \lesssim \|f\|_\alpha d(\gamma_1, \gamma_2)^{\frac{\theta}{2}},
\]
\[
\|L_{\gamma_1,z} f - L_{\gamma_2,z} f\|_2 \lesssim \|f\|_\alpha d(\gamma_1, \gamma_2)^{\frac{\theta}{2}}.
\]
Finally, from (6.11) and the previous inequalities, we prove (c), arguing as above. \qed

**Abstract Large Deviation Estimates.** Consider constant functions to measure the deviation sizes, $\mathcal{E} := \{\xi(t) \equiv \varepsilon : \varepsilon > 0\}$, and decaying exponential functions to bound the corresponding deviation set measures, $\mathcal{J} := \{\xi(t) = Ae^{-ct} : c > 0, A > 0\}$. Let $\mathcal{P}$ be the set of LDT parameters $(\mathcal{P}_0, \mathcal{P}_1, \mathcal{P}_2)$ in $\mathbb{N} \times \mathcal{E} \times \mathcal{J}$ such that for some constants $c > 0$ and $A > 0$, if $\xi(t) \equiv \varepsilon$ then $\xi(t) = Ae^{-c\varepsilon^2 t}$. The LDT estimates obtained in the sequel refer to this set $\mathcal{P}$ of LDT parameters.

Let $\{(K_{\gamma, \mu_\gamma, \xi_\gamma})_{\gamma \in \mathcal{X}}\}$ be a family of observed Markov systems, and $\Omega = \Sigma^\mathbb{Z}$ be the space of full sequences in the compact metric space $\Sigma$. Denote by $\mathcal{F}$ the Borel $\sigma$-field of $\Omega$, and consider the $\mathcal{F}$-measurable process $\{e_n : \Omega \rightarrow \Sigma\}_{n \geq 0}$ defined by $e_n(\omega) = x_n$, where $\omega = (x_n)_{n \in \mathbb{Z}}$. Given $x \in \Sigma$, we define $\mathbb{P}_x = \mathbb{P}_{\gamma,x}$ as the unique probability measure on $(\Omega, \mathcal{F})$ such that $\{e_n : \Omega \rightarrow \Sigma\}_{n \geq 0}$ is a Markov process with initial distribution $\delta_x$ and transition kernel $K_{\gamma}$. In general, given a probability measure $\mu$ on $\Sigma$, we set $\mathbb{P}_{\mu} := \int_\Sigma \mathbb{P}_x \mu(dx)$. The process $\{e_n\}_{n \geq 0}$ is then a Markov process on $(\Omega, \mathcal{F}, \mathbb{P}_{\mu})$ with initial distribution $\mu$ and transition kernel $K_{\gamma}$. Moreover, if the probability $\mu$ is $K_{\gamma}$-stationary then $\{e_n\}_{n \geq 0}$ is a stationary process. Define $S_n(\xi_\gamma)$ as the sum process on $(\Omega, \mathcal{F})$,
\[
S_n(\xi_\gamma) := \xi(e_0) + \xi(e_1) + \ldots + \xi(e_{n-1}).
\]
As before let $\lambda_\gamma(t) := \lambda(Q_{\gamma,t})$ denote the maximum modulus eigenvalue of $Q_{\gamma,t}$, while $c_\gamma(t) := \log \lambda_\gamma(t)$.

**Theorem 6.2.** Let $\{(K_{\gamma}, \mu_\gamma, \xi_\gamma)\}_{\gamma \in X}$ be a family of observed Markov system. Assume this family satisfies (A1)-(A4) for some positive constants $C, M, \theta, a, \beta, \alpha_0, \alpha_1, \sigma$ for which relations (6.1) hold. Then given $\gamma_0 \in X$ and $h > (c_{\gamma_0})''(0)$ there exist a neighbourhood $\mathcal{V}$ of $\gamma_0$ in $\mathcal{X}$, $A > 0$ and $\varepsilon_0 > 0$ such that for all $\gamma \in \mathcal{V}$, $n \in \mathbb{N}$, $0 < \varepsilon < \varepsilon_0$, and $x \in \Sigma$,

$$
P_x \left[ \left| \frac{1}{n} S_n(\xi_\gamma) - \mathbb{E}_\mu(\xi_\gamma) \right| \geq \varepsilon \right] \leq A e^{-\frac{\varepsilon^2}{2h} n}.
$$

**Corollary 6.11.** Let $(K, \mu)$ be a Markov system. Consider a family $\{(K, \mu, \xi)\}_{\xi \in \mathbb{B}_{a_0}}$ of observed Markov systems satisfying (A1) for some positive constants $C, \alpha_0, \alpha_1, \sigma$ fulfilling the relations (6.1). Then given $h > (c_{\alpha_0})''(0)$ there exist a neighbourhood $\mathcal{V}$ of $\xi$ in $\mathbb{B}$, and constants $A > 0$ and $\varepsilon_0 > 0$ such that for all $\xi \in \mathcal{V}$, $n \in \mathbb{N}$, $0 < \varepsilon < \varepsilon_0$ and $x \in \Sigma$,

$$
P_x \left[ \left| \frac{1}{n} S_n(\xi) - \mathbb{E}_\mu(\xi) \right| \geq \varepsilon \right] \leq A e^{-\frac{\varepsilon^2}{2h} n}.
$$

**Proof.** Given $\xi \in \mathbb{B}_{a_0}$ take a bounded neighbourhood $\mathcal{V}$ of $\xi$ in $\mathbb{B}_{a_0}$. By Proposition 6.2, the family $\{(K, \mu, \xi)\}_{\xi \in \mathcal{V}}$ also satisfies (A2)-(A4). The conclusion follows by applying Theorem 6.2 to this family of observed Markov systems. ■

**Remark 6.1.** Averaging in $x$, w.r.t. $\mu_\gamma$, the probabilities in theorem 6.2 we get for all $\gamma \in \mathcal{V}$, $n \in \mathbb{N}$ and $0 < \varepsilon < \varepsilon_0$,

$$
P_{\mu_\gamma} \left[ \left| \frac{1}{n} S_n(\xi_\gamma) - \mathbb{E}_\mu(\xi_\gamma) \right| \geq \varepsilon \right] \leq A e^{-\frac{\varepsilon^2}{2h} n}.
$$

In the rest of this section assume (A1)-(A4) hold for the family $\{(K_{\gamma}, \mu_\gamma, \xi_\gamma)\}_{\gamma \in X}$, and take $0 < \beta_0 < \beta$ according to Proposition 6.8.

**Lemma 6.12.** For all $\gamma \in \mathcal{X}$, $n \in \mathbb{N}$, $z \in \mathbb{D}_{\beta_0}$ and $x \in \Sigma$,

$$
((Q_{\gamma,z})^n 1)(x) = \mathbb{E}_x \left[ e^z S_n(\xi_\gamma) \right] = \int_{\Omega} e^z S_n(\xi_\gamma) d\mathbb{P}_x.
$$

In particular, for all $z \in \mathbb{D}_{\beta_0}$,

$$
\mathbb{E}_{\mu_\gamma}((Q_{\gamma,z})^n 1) = \mathbb{E}_{\mu_\gamma} \left[ e^z S_n(\xi_\gamma) \right].
$$
There exist $B > 0$ and a sequence $\delta_n$ converging to 0 geometrically such that for $\gamma \in \mathcal{X}$, $t \in [-\beta_0, \beta_0]$, $n \in \mathbb{N}$ and $x \in \Sigma$,

$$\left| E_x \left[ e^{tS_n(x_\gamma)} \right] - n \log \lambda(t) \right| \leq B |t| + \delta_n .$$

Proof. We shall use the notation of Proposition 6.8, choosing $\varepsilon > 0$ small enough so that $\sigma + \varepsilon < 1 - \varepsilon$. By Lemma 6.12, $(Q^n_{\gamma, t}1)(x) = E_x \left[ e^{tS_n(x_\gamma)} \right]$. By Lemma 6.10 there is $B > 0$ such that for all $t \in [-\beta_0, \beta_0]$, $\|P_{\gamma, t} - I\|_\alpha \leq B |t|$. Hence

$$\left| E_x \left[ e^{tS_n(x_\gamma)} \right] - \lambda(t)^n \right| \leq \left| (Q^n_{\gamma, t}1)(x) - \lambda(t)^n \right| \leq \|Q^n_{\gamma, t}1 - \lambda(t)^n P_{\gamma, t}1\|_\alpha + \lambda(t)^n \|1 - P_{\gamma, t}1\|_\alpha \leq N^n_{\gamma, t}1\|_\alpha + \lambda(t)^n \|1 - P_{\gamma, t}1\|_\alpha \leq C (\sigma + \varepsilon)^n + B |t| \lambda(t)^n .$$

Thus

$$\left| \log E_x \left[ e^{tS_n(x_\gamma)} \right] - n \log \lambda(t) \right| = \left| \log E_x \left[ e^{tS_n(x_\gamma)} \right] - \log \lambda(t)^n \right| \leq \left| \frac{E_x \left[ e^{tS_n(x_\gamma)} \right] - \lambda(t)^n}{E_x \left[ e^{tS_n(x_\gamma)} \right]} \right| \min\{\lambda(t)^n, E_x \left[ e^{tS_n(x_\gamma)} \right]\} \leq \frac{B |t| \lambda(t)^n + C (\sigma + \varepsilon)^n}{(1 - B |t|) \lambda(t)^n - C (\sigma + \varepsilon)^n} \leq \frac{B |t| + \delta_n}{1 - B |t|} \leq 2 (B |t| + \delta_n) ,$$

where $\delta_n := C (\sigma + \varepsilon)^n \leq C \left( \frac{\sigma + \varepsilon}{1 - \varepsilon} \right)^n$ converges geometrically to zero. 

Proof of Theorem 6.2. Denote by $\mathcal{H}(\mathcal{D}_{\beta_0})$ the Banach space of analytic functions on $\mathcal{D}_{\beta_0}$ with a continuous extension to $\mathcal{D}_{\beta_0}$. By Proposition 6.8(5), $\lambda_\gamma \in \mathcal{H}(\mathcal{D}_{\beta_0})$ for all $\gamma \in \mathcal{X}$. Since $h > (c_{\gamma_0}''(0)$, by Proposition 6.8(c) there is a neighbourhood $\mathcal{V}$ of $\gamma_0$ in $\mathcal{X}$ such that $h > (c_{\gamma})''(0)$ for all $\gamma \in \mathcal{V}$. Assume $E_\mu(x_\gamma) = 0$. Otherwise work with
\( \xi' = \xi - \mathbb{E}_\mu(\xi) \mathbf{1} \), for which \( \mathbb{E}_\mu(\xi') = 0 \). By Proposition 6.13:

\[
\frac{1}{n} \log \mathbb{E}_x [e^{tS_n(\xi')} ] \leq c_n(t) + \frac{B|t| + \delta_n}{n},
\]

with \( \delta_n \) decreasing to 0 geometrically. Because \( h > (c_\gamma)'(0) \), using again the equi-continuity in 6.8(c) there is a small neighbourhood \( ] - t_0, t_0[ \), of \( t = 0 \) such that \( c_\gamma(t) \leq \frac{ht^2}{2} \), for all \( \gamma \in \mathcal{V} \). Applying Chebyshev’s inequality for \( |t| < t_0 \)

\[
\mathbb{P}_x[S_n(\xi_\gamma) > n\varepsilon ] \leq e^{-t\varepsilon} \mathbb{E}_x[e^{tS_n(\xi)} ]
\]

\[
\leq e^{-(t\varepsilon - c_\gamma(t)) n + B|t| + \delta_n}
\]

\[
\leq e^{-(t\varepsilon - \frac{ht^2}{2}) n + B|t| + \delta_n}.
\]

Thus, given \( 0 < \varepsilon < \varepsilon_0 := \beta t_0 \), picking \( t = \frac{\varepsilon}{h} \in ]0, t_0[ \) we get

\[
\mathbb{P}_x[S_n(\xi_\gamma) > n\varepsilon ] \leq e^{-\frac{2}{\pi} n + B|\frac{\varepsilon}{h}| + \delta_n}.
\]

The assumptions (A1)-(A4) are symmetric in \( \xi \), in the sense that if the family \( \{(K_\gamma, \mu_\gamma, \xi_\gamma)\}_{\gamma \in \mathcal{X}} \) satisfies them then so does \( \{(K_\gamma, \mu_\gamma, -\xi_\gamma)\}_{\gamma \in \mathcal{X}} \). Therefore we can derive the same conclusion for \( -\xi_\gamma \),

\[
\mathbb{P}_x[S_n(\xi_\gamma) < -n\varepsilon ] = \mathbb{P}_\mu[S_n(-\xi_\gamma) > n\varepsilon ] \leq e^{-\frac{2}{\pi} n + B|\frac{\varepsilon}{h}| + \delta_n}.
\]

Thus, for all \( \gamma \in \mathcal{V} \), \( n \in \mathbb{N} \) and \( 0 < \varepsilon < \varepsilon_0 \),

\[
\mathbb{P}_x[|S_n(\xi_\gamma)| > n\varepsilon ] \leq e^{-\frac{2}{\pi} n + B|\frac{\varepsilon}{h}| + \delta_n} + e^{-\frac{2}{\pi} n + B|\frac{\varepsilon}{h}| + \delta_n} \leq e^{-\frac{2}{\pi} n}.
\]

\[
\square
\]

7. Applications to random cocycles

In this section we prove uniform base and fibre LDT estimates for the class of irreducible cocycles over strongly mixing Markov shifts.

Base Large Deviation Estimates. Consider a strongly mixing Markov system \((K, \mu)\) on the compact metric space \( \Sigma \). Let \( \Omega^- = \Sigma_0^\mathbb{Z} \) and \( \Omega = \Sigma^\mathbb{Z} \) denote the compact spaces of sequences of states in \( \Sigma \), where \( \Sigma_0^- := \{ \ldots, -2, -1, 0 \} \). We endow these spaces with the Borel \( \sigma \)-fields generated by cylinders, respectively denoted by \( \mathcal{F}^- \) and \( \mathcal{F} \). Let \( T : \Omega \to \Omega \) denote the full shift map \( T(x_n)_{n \in \mathbb{Z}} = (x_{n+1})_{n \in \mathbb{Z}} \). Consider also the measure \( \mathbb{P}_\mu \) in \( \Omega \), introduced just before Theorem 6.2. This probability measure is \( T \)-invariant, i.e., \( \mathbb{P}_\mu = T_* \mathbb{P}_\mu \), and hence determines the mixing dynamical system \((\Omega, \mathcal{F}, \mathbb{P}_\mu, T)\). We shall denote by \( \mathbb{P}^- \) the push-forward of \( \mathbb{P}_\mu \) by the canonical projection \( \Pi : \Omega \to \Omega^- \),
Consider the process \( \{ e_n : \Omega \to \Omega^\sim \}_{n \geq 0} \) in \((\Omega, \mathbb{P}_\mu)\) defined by \( e_n(\omega) := (\ldots, x_{n-1}, x_n) \), where \( \omega = (x_n)_{n \in \mathbb{Z}} \). This process has initial distribution \( \mathbb{P}_\mu^\sim \) and transition kernel \( \tilde{K} \), and because \( \{ e_n : \Omega \to \Sigma \}_{n \in \mathbb{Z}} \) is a stationary process it follows that \( \{ e_n \}_{n \in \mathbb{Z}} \) is also a stationary process. Thus, \( \mathbb{P}_\mu \) is a \( \tilde{K} \)-stationary measure. The Markov operator of \( \tilde{K} \) is the linear transformation \( Q_{\tilde{K}} : L^\infty(\Omega^\sim) \to L^\infty(\Omega^\sim) \),

\[
(Q_{\tilde{K}} f)(\ldots, x_{-1}, x_0) := \int_\Sigma f(\ldots, x_{-1}, x_0, x_1) K(x_0, dx_1).
\]

Next we introduce a scale of Banach algebras satisfying (B1)-(B7), where the operators \( Q_{\tilde{K}} \) shall act. Consider the metric \( \tilde{d} : \Omega \times \Omega \to [0, 1] \), where for all \( \omega = (x_k)_{k \in \mathbb{Z}} \) and \( \omega' = (x'_k)_{k \in \mathbb{Z}} \) in \( \Omega \),

\[
\tilde{d}(\omega, \omega') := 2^{-\inf\{|k| : k \in \mathbb{Z}, x_k \neq x'_k\}}.
\]

Note that \( \Omega \) is not compact for the topology induced by \( \tilde{d} \), unless \( \Sigma \) is finite. Given \( k \in \mathbb{N} \), \( \alpha \geq 0 \) and \( f \in L^\infty(\Omega) \) define

\[
\begin{align*}
v_k(f) &:= \sup\{|f(x) - f(y)| : \tilde{d}(x, y) \leq 2^{-k}\}, \\
V_\alpha(f) &:= \sup\{2^{\alpha k} v_k(f) : k \in \mathbb{N}\}, \\
\|f\|_\alpha &:= \|f\|_\infty + V_\alpha(f), \\
\mathcal{H}_\alpha(\Omega) &:= \{ f \in L^\infty(\Omega) : V_\alpha(f) < +\infty \}.
\end{align*}
\]

The last set, \( \mathcal{H}_\alpha(\Omega) \), is the space of Hölder continuous functions with exponent \( \alpha \) w.r.t. the distance \( \tilde{d} \) on \( \Omega \). In fact it follows easily from the definition that

\[
V_\alpha(f) = \sup_{\omega \neq \omega'} \frac{|f(\omega) - f(\omega')|}{\tilde{d}(\omega, \omega')^\alpha},
\]

which shows that \( \mathcal{H}_\alpha(\Omega) \) is the space of functions \( f : \Omega \to \mathbb{C} \) where the variation \( v_k(f) \) decays exponentially with \( k \) at rate \( 2^{-\alpha} \), i.e., for some constant \( C > 0 \), and all \( k \in \mathbb{N} \), \( v_k(f) \leq C 2^{-\alpha k} \).

**Proposition 7.1.** For all \( k \in \mathbb{N} \) and \( \alpha \geq 0 \), the functions \( v_k \) and \( V_\alpha \) are seminorms on \( \mathcal{H}_\alpha(\Omega) \) such that for all \( f, g \in \mathcal{H}_\alpha(\Omega) \),
Exercise. another scale of Banach algebras satisfying the assumptions (B1)-(B7).

Proposition 7.2. \((\mathcal{H}_\alpha(\Omega), \| \cdot \|_\alpha)\) is a unital Banach algebra, and also a lattice, for all \(0 \leq \alpha \leq 1\).

For \(\alpha = 0\), the seminorm \(V_0\) measures the variation of \(f\) because

\[ V_0(f) = \sup \{ |f(\omega) - f(\omega')| : \omega, \omega' \in \Omega \} . \]

Hence \(\mathcal{H}_0(\Omega) = L^\infty(\Omega)\), while the norm \(\| \cdot \|_0\) is equivalent to \(\| \cdot \|_\infty\). These considerations show that \(\{\mathcal{H}_\alpha(\Omega)\}_{\alpha \in [0,1]}\) satisfies the assumptions (B1)-(B4), while the remaining (B5)-(B7) follow from the observation made after the statement of these conditions. See also [21].

Any function \(\xi \in L^\infty(\Omega^-)\) can, and will, be regarded as a function \(\xi(\omega)\) of \(\omega = (x_n)_{n \in \mathbb{Z}} \in \Omega\), that does not depend on the ‘future’ coordinates \(x_n\) with \(n \geq 0\). Hence \(L^\infty(\Omega^-)\) is a subspace of \(L^\infty(\Omega)\). Making these identifications, the spaces \(\mathcal{H}_\alpha(\Omega^-) := \mathcal{H}_\alpha(\Omega) \cap L^\infty(\Omega^-)\), form another scale of Banach algebras satisfying the assumptions (B1)-(B7).

We are now going to prove that the Markov operator \(Q_K\) acts continuously on \(\mathcal{H}_\alpha(\Omega^-)\).

Proposition 7.3. For all \(f \in \mathcal{H}_\alpha(\Omega^-)\) and \(n \in \mathbb{N}\),

1. \(\|(Q_K)^n f\|_\infty \leq \|f\|_\infty\).
2. \(V_\alpha((Q_K)^n f) \leq \max \{2\|(Q_K)^n f\|_\infty, 2^{-n} V_\alpha(f)\}\).

Proof. We shall write \(Q = Q_K\). The first inequality follows because \(\int_{\Omega} K(x_0, dx_1) = 1\). For the second, note that if \(k \geq 1\) then \(v_k(Q^n f) \leq v_{k+n}(f)\). Indeed, for \(\omega = (x_n)_{n \leq 0}\) and \(\omega' = (x'_n)_{n \leq 0}\) in \(\Omega^-\) such that \(\Delta(\omega, \omega') \leq 2^{-k}\) with \(k \geq 1\), we have \(x_0 = x'_0\). Thus

\[
\|(Q^n f)(\ldots, x_{-1}, x_0) - (Q^n f)(\ldots, x'_{-1}, x'_0)\| \leq \int_{\Sigma^n} |f(\ldots, x_0, x_1, \ldots, x_n) - f(\ldots, x'_0, x_1, \ldots, x_n)| \prod_{j=0}^{n-1} K(x_j, dx_{j+1})
\]

\[
\leq v_{k+n}(f) \int_{\Sigma^n} \prod_{j=0}^{n-1} K(x_j, dx_{j+1}) = v_{k+n}(f),
\]

and taking the sup in \(\omega, \omega' \in \Omega^-\) such that \(\Delta(\omega, \omega') \leq 2^{-k}\), the inequality \(v_k(Q^n f) \leq v_{k+n}(f)\) follows. Hence, for \(k \geq 1\),

\[
2^{nk}v_k(Q^n f) = 2^{-n\alpha}(2^{\alpha(k+n)}v_{k+n}(f)) \leq 2^{-n\alpha}V_\alpha(f).
\]
For $k = 0$ note that $\nu_0(Q^n f)$ is the variation of $Q^n f$. Thus $\nu_0(Q^n f) \leq 2 \|Q^n f\|_{\infty}$. Taking the sup in $k \in \mathbb{N}$, item (2) follows. □

Let $\mathcal{V}$ be any bounded set $\mathcal{H}_1(\Omega^-)$ and consider the family of observed Markov systems $\{(K, \mu, \xi)\}_{\xi \in \mathcal{V}}$. We claim that given $0 < \beta < a$, there are constants $C, M > 0$ and $0 < \sigma < 1$ for which this family satisfies the assumptions (A1)-(A4) with $\theta = \alpha_0 = 1$. By Proposition 6.2, it is enough to check condition (A1), which is done in the next proposition.

**Proposition 7.4.** If $(K, \mu)$ is strongly mixing, then given $0 < \alpha_1 < \frac{1}{2}$ there are constants $C > 0$ and $0 < \sigma < 1$ such that for all $\alpha_1 \leq \alpha \leq 1$, $(\tilde{K}, \mathbb{P}_\mu)$ is strongly mixing on $\mathcal{H}_\alpha(\Omega^-)$ with constants $(C, \sigma)$.

**Proof.** Given a function $f \in \mathcal{H}_\alpha(\Omega^-)$, denote by $f_k : \Omega^- \to \mathbb{C}$ the following function

$$f_k(\ldots, x_0) := \int_{\Omega^-} f(\ldots, x_k, \ldots, x_0) d\mathbb{P}_\mu(\ldots, x_k).$$

Note that if $\mathcal{F}_k^-$ is the sub $\sigma$-field of $\mathcal{F}^-$ generated by the cylinders in the coordinates $x_{-k+1}, \ldots, x_{-1}, x_0$, we have $f_k = \mathbb{E}_\mu(f|\mathcal{F}_k^-)$, and in particular $\mathbb{E}_\mu(f_k) = \mathbb{E}_\mu(f)$, for all $k \in \mathbb{N}$. By definition of $f_k$,

$$\|Q^n(f - f_k)\|_{\infty} \leq \|f - f_k\|_{\infty} \leq v_k(f) \leq 2^{-\alpha k} V_\alpha(f). \tag{7.1}$$

Because $(K, \mu)$ is strongly mixing, there are constants $C > 0$ and $0 < \rho < 1$ such that for any function $h \in L^\infty(\Sigma)$ with $\int_\Sigma h \, d\mu = 0$,

$$\left| \int_\Sigma h(y) K^n(x, dy) \right| \leq C \rho^n \|h\|_{\infty}.$$

Now, if $h \in L^\infty(\Omega^-)$ is a function with zero average, i.e., $\mathbb{E}_\mu(h) = 0$, which depends only on the first coordinate $x_0$, then $Q^n h$ also depends only on the first coordinate, and is given by

$$(Q^n h)(\ldots, x_0) := \int_\Sigma h(y) K^n(x_0, dy).$$

Hence

$$\|Q^n h\|_{\infty} \leq C \rho^n \|h\|_{\infty}. \tag{7.2}$$

We claim that $h = Q^k(f_k - \mathbb{E}_\mu(f)) 1$ is a function with zero average that depends only on the first coordinate. The first part of claim follows because $Q$ preserves averages and, as remarked above, $\mathbb{E}_\mu(f_k) = \mathbb{E}_\mu(f)$. For the second part note two things: first $Q$ ‘preserves’ functions that depend only on the first coordinate $x_0$; second, $Q$ maps a function $f$ that depends only on the coordinates $x_{-k}, \ldots, x_{-1}, x_0$ to a function that
depends only on the coordinates \(x_{-k+1}, \ldots, x_{-1}, x_0\), in other words \(Qf\) looses dependence in \(x_{-k}\). Therefore, from (7.2)

\[
\|Q^n(f_k - \mathbb{E}_\mu^- (f) 1)\|_\infty = \|Q^n f_h\|_\infty \leq C \rho^{-k} \|h\|_\infty \\
\leq C \rho^{-k} \|Q^k (f_k - \mathbb{E}_\mu^- (f) 1)\|_\infty \\
\leq C \rho^{-k} \|f_k - \mathbb{E}_\mu^- (f) 1\|_\infty \leq 2C \rho^{-k} \|f\|_\infty
\]  

(7.3)

Setting \(\sigma = \max\{2^{-\frac{n}{2}}, \sqrt{q}\}\) we have \(0 < \sigma < 1\). From the inequalities (7.1) and (7.3), with \(k = n/2\), we have

\[
\|Q^n f - \mathbb{E}_\mu^- (f) 1\|_\infty \leq \|Q^n (f - f_k)\|_\infty + \|Q^n (f_k - \mathbb{E}_\mu^- (f) 1)\|_\infty \\
\leq 2^{-\alpha_2} V_\alpha (f) + 2C \rho^2 \|f\|_\infty \\
\leq \sigma^n V_\alpha (f) + 2C \sigma^n \|f\|_\infty .
\]

On the other hand, by item (2) of Proposition 7.3

\[
V_\alpha (Q^n f - \mathbb{E}_\mu^- (f) 1) = V_\alpha (Q^n (f - \mathbb{E}_\mu^- (f) 1)) \\
\leq \max\{\|Q^n f - \mathbb{E}_\mu^- (f) 1\|_\infty, 2^{-\alpha_1} V_\alpha (f)\} \\
\leq \max\{\sigma^n V_\alpha (f) + 2C \sigma^n \|f\|_\infty, \sigma^{2n} V_\alpha (f)\} \\
\leq \sigma^n V_\alpha (f) + 2C \sigma^n \|f\|_\infty .
\]

Thus, for all \(f \in \mathcal{H}_\alpha (\Omega^-)\),

\[
\|Q^n f - \mathbb{E}_\mu^- (f) 1\|_\alpha \leq 4C \sigma^n \|f\|_\alpha ,
\]

which proves that \((Q^K, \mathbb{P}_\mu^-)\) is strongly mixing in \(\mathcal{H}_\alpha (\Omega^-)\) with constants \((4C, \sigma)\).

\[\square\]

**Theorem 7.1.** If \((K, \mu)\) is strongly mixing then for all \(0 < \alpha \leq 1\), any observable \(\xi \in \mathcal{H}_\alpha (\Omega^-)\) satisfies a uniform base-LDT estimate, with parameter space \(\mathcal{P}\), for the measure preserving dynamical system \((\Omega, \mathcal{F}, \mu, T)\).

**Proof.** Given \(0 < \alpha \leq 1\) and \(\xi \in \mathcal{H}_\alpha (\Omega^-)\), take \(\alpha_0 = \alpha, \alpha_1 = \frac{\alpha}{2}\) and a neighbourhood \(V\) of \(\xi\) in \(\mathcal{H}_\alpha (\Omega)\). As explained above the family \(\{(K, \mu, \xi)\}_{\xi \in V}\) satisfies (A1), and hence, by Proposition 6.2 it also satisfies (A2)-(A4). Thus by Theorem 6.2, given \(\xi \in V\) there exist a neighbourhood \(U \subset V\) of \(\xi\), and constants \(\varepsilon_0 > 0, A > 0\) and \(h > 0\) such that for all \(\xi \in V\), all \(0 < \varepsilon < \varepsilon_0\) and all \(n \in \mathbb{N}\),

\[
\mathbb{P}_\mu^- \left[ \left| \frac{1}{n} S_n (\xi) - \mathbb{E}_\mu (\xi) \right| \geq \varepsilon \right] \leq A e^{-\frac{\varepsilon^2}{2n} n} .
\]
Since we have for the measure preserving projection $\Pi : \Omega \to \Omega$,
\[
\Pi^{-1}\left[\left|\frac{1}{n}S_n(\zeta) - \mathbb{E}_\mu(\zeta)\right| \geq \varepsilon\right] = \left\{\omega \in \Omega : \left|\frac{1}{n}\sum_{j=0}^{n-1}\xi(T^j(\omega)) - \int_\Omega \zeta\,d\mathbb{P}_\mu\right| \geq \varepsilon\right\}
\]
the observable $\xi$ satisfies a uniform base-LDT estimate. 

**Fiber Large Deviation Estimates.** Consider a strongly mixing Markov system $(K, \mu)$ on the compact metric space $\Sigma$, and the spaces of measurable cocycles $\mathcal{B}(K,a)$, with $1 < a \leq \infty$, introduced in section 2. Elements in $\mathcal{B}(K,a)$ are measurable functions $A : \Sigma \times \Sigma \to \text{GL}(d, \mathbb{R})$. Let $\mathbb{P}(\mathbb{R}^m)$ denote the real projective space, a compact metric space. Identifying each $p \in \mathbb{P}(\mathbb{R}^m)$ with any of its representative unit vectors $p \in \mathbb{R}^m$, we consider the distance in $\mathbb{P}(\mathbb{R}^m)$ defined by
\[
\delta(p, q) := \frac{\|p \wedge q\|}{\|p\|\|q\|}.
\]
A cocycle $A \in \mathcal{B}(K,a)$ induces a map $F_A : \Omega \times \mathbb{P}(\mathbb{R}^m) \to \Omega \times \mathbb{P}(\mathbb{R}^m)$ defined by $F_A(\omega, v) = (T\omega, A(x_0, x_1)v)$, where $\omega = (x_n)_{n \in \mathbb{N}}$. The linear transformation $Q_A : L^\infty(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m)) \to L^\infty(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m))$,
\[
(Q_A f)(x, y, p) := \int_\Sigma f(y, z, A(y, z)p) K(y, dz),
\]
is the Markov operator of the following kernel on $\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m)$
\[
K_A(x, y, p) := \int_\Sigma \delta_{(y, z, A(y, z)p)} K(y, dz).
\]
Identifying $L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ with the subspace of functions $f(x, y, p)$ in $L^\infty(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m))$ that do not depend on the first coordinate $x$,
\[
Q_A(L^\infty(\Sigma \times \Sigma \times \mathbb{P}(\mathbb{R}^m))) \subseteq L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^m)) .
\]
In particular the subspace $L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ is $Q_A$-invariant, and the restriction of $Q_A$ to this subspace is given by
\[
(Q_A f)(x, p) := \int_\Sigma f(y, A(x, y)p) K(x, dy).
\]
This operator is associated to the following Markov kernel on $\Sigma \times \mathbb{P}(\mathbb{R}^m)$,
\[
K_A(x, p) := \int_\Sigma \delta_{(y, A(x, y)p)} K(x, dy).
\]
Next we introduce a scale of Banach algebras satisfying (B1)-(B7). Given $0 \leq \alpha \leq 1$ and $f \in L^\infty(\Sigma \times \Sigma \times \mathbb{P}^m)$ define
\[
\|f\|_\alpha := V_\alpha(f) + \|f\|_\infty, \tag{7.8}
\]
\[
V_\alpha(f) := \sup_{x,y, \in \Sigma \atop p \neq q} \left| \frac{f(x,y,p) - f(x,y,q)}{\delta(p,q)^\alpha} \right|, \tag{7.9}
\]
\[\mathcal{H}_\alpha(\Sigma \times \Sigma \times \mathbb{P}^m) := \{ f \in L^\infty(\Sigma \times \Sigma \times \mathbb{P}^m) : V_\alpha(f) < +\infty \}.
\]
\[\mathcal{H}_\alpha(\Sigma \times \Sigma \times \mathbb{P}^m)\] is the space of measurable functions which are Hölder continuous, with exponent $\alpha$, in the last variable. The fact that $\mathcal{H}_\alpha(\Sigma \times \Sigma \times \mathbb{P}^m)$ is a Banach algebra follows from the following proposition.

**Proposition 7.5.** For $0 \leq \alpha \leq 1$, the function $V_\alpha$ is a seminorm on $\mathcal{H}_\alpha(\Sigma \times \Sigma \times \mathbb{P}^m)$ such that for $f, g \in \mathcal{H}_\alpha(\Sigma \times \Sigma \times \mathbb{P}^m)$,
\[
V_\alpha(fg) \leq \|f\|_\infty V_\alpha(g) + \|g\|_\infty V_\alpha(f).
\]
Moreover, $(\mathcal{H}_\alpha(\Sigma \times \Sigma \times \mathbb{P}^m), \|\cdot\|_\alpha)$ is a unital Banach algebra, as well as a lattice.

Consider the family of Banach spaces $\{ (\mathcal{H}_\alpha(\Sigma \times \Sigma \times \mathbb{P}^m)), \|\cdot\|_\alpha \}_{\alpha \in [0,1]}$ for which conditions (B1), (B3) and (B4) are automatic. For $\alpha = 0$, the seminorm $V_0$ measures the variation of $f$ on the projective coordinate
\[
V_0(f) = \sup \{|f(x,y,p) - f(x,y,p')| : x, y \in \Sigma, p, p' \in \mathbb{P}^m \}.
\]
Hence $\mathcal{H}_0(\Sigma \times \Sigma \times \mathbb{P}^m) = L^\infty(\Sigma \times \Sigma \times \mathbb{P}^m)$, while the norm $\|\cdot\|_0$ is equivalent to $\|\cdot\|_\infty$. This proves (B2). Conditions (B5) and (B6) hold because the projective metric $\delta$ takes values in $[0,1]$. Finally (B7) follows from the equality
\[
\frac{\Delta}{d^{\alpha_1}} = \left( \frac{\Delta}{d^{\alpha_0}} \right)^{\frac{\alpha_2 - \alpha_0}{\alpha_2 - \alpha_0}} \left( \frac{\Delta}{d^{\alpha_2}} \right)^{\frac{\alpha_1 - \alpha_0}{\alpha_2 - \alpha_0}},
\]
which holds for all $\Delta \geq 0$ and $d > 0$.

We shall consider families of observed Markov systems of the form $\{ (K_A, \mu_A, \xi_A) \}_{A \in \mathcal{V}}$, where $\mathcal{V}$ is an open set in $\mathcal{B}(K,a)$, and for every cocycle $A \in \mathcal{V}$, $K_A$ is the kernel \cite{11}, $\mu_A$ is the unique $K_A$-stationary measure in $\Sigma \times \Sigma \mathbb{P}^m$, and $\xi_A : \Sigma \times \Sigma \times \mathbb{P}^m \to \mathbb{R}$ is the measurable function defined by
\[
\xi_A(x,y,p) := \log \|A(x,y) p\|.
\]
Note that the scale of Banach sub-algebras $\{(H_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m)), \|\cdot\|_\alpha)\}_{\alpha \in [0,1]}$, where the Markov operators act invariantly, also satisfies the conditions (B1)-(B7).

Throughout the rest of this section we assume that

(C1) $(K, \mu)$ is strongly mixing on $\Sigma$,
(C2) $A \in \mathcal{B}^{\text{irred}}(K, a)$ with $a \geq 2$,
(C3) $L_1(A, \mu) > L_2(A, \mu)$.

Our goal is to show that if $\mathcal{V}$ is small enough, then conditions (A1)-(A4) hold for the family of observed Markov systems $\{(K_B, \mu_B, \xi_B)\}_{B \in \mathcal{V}}$, this in order to apply Theorem 6.2.

Next proposition implies (A1).

**Proposition 7.6.** Given $(K, \mu, A)$ satisfying (C1)-(C3), there exists a neighbourhood $\mathcal{V}$ of $A$ in $\mathcal{B}(K, a)$, and constants $\alpha_0, \alpha_1, \sigma, C > 0$ such that $0 < \sigma < 1$, $0 < \alpha_1 < \frac{\alpha_0}{2} \leq 1$, and for all $B \in \mathcal{V}$ and $f \in H_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ with $\alpha_1 \leq \alpha \leq \alpha_0$,

$$\|(Q_{K_B})^n f - \langle f, \mu_B \rangle \|_\alpha \leq C \sigma^n \|f\|_\alpha,$$

where $\mu_B$ denotes a $K_B$-stationary measure on $\Sigma \times \mathbb{P}(\mathbb{R}^m)$.

This proposition follows from the proof of Theorem 3.7 in [4], using the conclusion of Proposition 7.20.

**Proposition 7.7.** If $(K, \mu, A)$ satisfies (C1)-(C3) then there is a unique $K_A$-stationary probability measure $\mu_A$ on $\Sigma \times \mathbb{P}(\mathbb{R}^m)$.

**Proof.** See Lemma 3.5 in [4] for the existence proof. The uniqueness is not really needed here, but it follows from Proposition 7.6 above. □

The proof of Proposition 7.6 relies on a long chain of lemmas. We begin with five lemmas ruling the Lipschitz action of an invertible matrices on the projective space. Given non collinear vectors $p, q \in \mathbb{R}^d$ with $\|p\| = \|q\| = 1$, let

$$v_p(q) := \|q - \langle p, q \rangle p\|^{-1} (q - \langle p, q \rangle p)$$

be the versor of the orthogonal projection of $q$ onto $p^\perp$.

**Lemma 7.8.** Given $A \in \text{GL}(d, \mathbb{R})$, and points $p \neq q$ in $\mathbb{P}(\mathbb{R}^m)$,

$$\frac{\delta(Ap, Aq)}{\delta(p, q)} = \|Ap \wedge Av_p(q)\|.$$
Proof. We can write \( q = (\cos \theta) p + (\sin \theta) v_p(q) \), where \( \theta = \angle(p, q) \).

Hence
\[
\delta(p, q) = \|p \wedge q\| = (\sin \theta) \|p \wedge v_p(q)\| = \sin \theta ,
\]
and
\[
\delta(Ap, Aq) = (\sin \theta) \|Ap \wedge Av_p(q)\| ,
\]
which proves the claimed formula.

Lemma 7.9. Given \( A, B \in \text{GL}(d, \mathbb{R}) \), and \( p \neq q \) in \( \mathbb{P}(\mathbb{R}^m) \),
\[
\| \frac{\delta(Ap, Aq)}{\delta(p, q)} - \frac{\delta(Bp, Bq)}{\delta(p, q)} \| \leq C \|A - B\| ,
\]
where \( C = \max\{\|A\|, \|B\|\} \).

Proof.
\[
\left| \frac{\delta(Ap, Aq)}{\delta(p, q)} - \frac{\delta(Bp, Bq)}{\delta(p, q)} \right| = \left| \|Ap \wedge Av_p(q)\| - \|Bp \wedge Bv_p(q)\| \right|
\]
\[
\leq \|Ap \wedge Av_p(q)\| - \|Bp \wedge Bv_p(q)\|
\]
\[
\leq \|Ap \wedge (Av_p(q) - Bv_p(q))\| + \|(Ap - Bp) \wedge Bv_p(q)\|
\]
\[
\leq \|A\| \|A - B\| + \|A - B\| \|B\|
\]
\[
\leq \max\{\|A\|, \|B\|\} \|A - B\| .
\]

Lemma 7.10. Given \( A, B \in \text{GL}(d, \mathbb{R}) \),
\[
\|A^{-1}\|^{-2} \leq \frac{\delta(Ap, Aq)}{\delta(p, q)} \leq \|A\|^2 .
\]

Proof. Given unit vectors \( p, v \in \mathbb{R}^m \) with \( \|p\| = \|v\| = 1 \) and \( \langle p, v \rangle = 0 \),
\[
\|Ap \wedge Av\| = \|\Lambda_2 A(p \wedge v)\| \leq \|\Lambda_2 A\| \leq \|A\|^2 .
\]
Similarly
\[
\|Ap \wedge Av\| = \|\Lambda_2 A(p \wedge v)\| \geq \|\Lambda_2 A^{-1}\|^{-1} \geq \|A^{-1}\|^{-2} .
\]

Given \( M \in \text{GL}(d, \mathbb{R}) \) define
\[
\ell(M) := \max\{\log\|M\|, \log\|M^{-1}\|\} .
\]

(7.10)

Lemma 7.11. For every \( M \in \text{GL}(d, \mathbb{R}) \) and \( p, q \in \mathbb{P}(\mathbb{R}^m) \),
\[
-2\ell(M) \leq \log \left[ \frac{\delta(M p, M q)}{\delta(p, q)} \right] \leq 2\ell(M) .
\]

Proof. Follows from Lemma 7.10. See also Theorem 1 in [22].

\[\hfill\]
Lemma 7.12. Given $A, B \in \text{GL}(d, \mathbb{R})$, $0 < \alpha \leq 1$ and $p \neq q$ in $\mathbb{P}(\mathbb{R}^m)$, 
\[
\left| \left( \frac{\delta(Ap, Aq)}{\delta(p, q)} \right)^{\alpha} - \left( \frac{\delta(Bp, Bq)}{\delta(p, q)} \right)^{\alpha} \right| \leq C \|A - B\| ,
\]
where $C = \alpha \max\{\|A^{-1}\|, \|B^{-1}\|\}^{2(1-\alpha)} \cdot \max\{\|A\|, \|B\|\}$.

Proof. Setting $\Delta_A := \frac{\delta(Ap, Aq)}{\delta(p, q)}$ and $\Delta_B := \frac{\delta(Bp, Bq)}{\delta(p, q)}$, from lemmas 7.9 and 7.10 we get 
\[
|\Delta_A^\alpha - \Delta_B^\alpha| \leq \alpha \max\{\Delta_A^{-1}, \Delta_B^{-1}\} |\Delta_A - \Delta_B|
\]
\[
\leq \alpha \max\{\|A^{-1}\|, \|B^{-1}\|\}^{2(1-\alpha)} \|\Delta_A - \Delta_B\|
\]
\[
\leq \alpha \max\{\|A^{-1}\|, \|B^{-1}\|\}^{2(1-\alpha)} \max\{\|A\|, \|B\|\} \|A - B\| .
\]

Recall that by definition of $\mathcal{B}(K, a)$, given $A \in \mathcal{B}(K, a)$, the following function is uniformly bounded
\[
\eta^a_A(x) := \int_{\Sigma} \|A(x, y)^{-1}\|^a K(x, dy) = \mathbb{E}_x \left( \|\hat{A}^{-1}\|^a \right) .
\]

Consider the norm $\|\cdot\|_\infty$ in the space $L^\infty(\Omega, \text{GL}(d, \mathbb{C}))$ of bounded measurable functions $\hat{A} : \Omega \rightarrow \text{GL}(d, \mathbb{C})$,
\[
\|\hat{A}\|_\infty := \sup_{\omega \in \Omega} \|\hat{A}(\omega)\| .
\]

We shall denote by $d_\infty(\cdot, \cdot)$ the associated distance
\[
d_\infty(\hat{A}, \hat{B}) := \|\hat{A} - \hat{B}\|_\infty .
\]

For all $A, B \in \mathcal{B}(K, a)$, $d_\infty(\hat{A}, \hat{B}) = d_\infty(A, B)$, where the right-hand-side stands for the distance in $\mathcal{B}(K, a)$.

Next lemma describes semigroup properties of the cocycles in the spaces $\mathcal{B}(K, a)$ with $0 < a \leq \infty$.

Lemma 7.13. Let $A \in \mathcal{B}(K, a)$ with $0 < a \leq \infty$.

1. $\mathbb{E}_x(\|\hat{A}^{(n)}\|^{-1}) \leq (\|\eta^a_A\|_\infty)^n$, for all $x \in \Sigma$.

2. if $a = \infty$ then $\|\hat{A}^{(n)}\|_\infty \leq (\|A^{-1}\|_\infty)^n$.

3. $\|\hat{A}^{(n)}\|_\infty \leq (\|\hat{A}\|_\infty)^n$.

4. $d_\infty(\hat{A}^{(n)}, \hat{B}^{(n)}) \leq n \max\{\|\hat{A}\|_\infty, \|\hat{B}\|_\infty\}^{n-1} d_\infty(\hat{A}, \hat{B})$. 
Proof. Assume that \( A \in \mathcal{B}(K,a) \), and \( \mathbb{E}_x(\|\hat{A}^{-1}\|^a) \leq C \) for all \( x \in \Sigma \).

\[
\mathbb{E}_{x_0}(\|\hat{A}^{(n)}(\omega)^{-1}\|^a) = \int_{\Omega} \|\hat{A}^{(n)}(\omega)^{-1}\|^a \, d\mathbb{P}_{x_0}(\omega)
\leq \int_{\Omega} \|\hat{A}^{(n-1)}(\omega)^{-1}\|^a \|A(x_{n-1}, x_n)^{-1}\|^a \, d\mathbb{P}_{x_0}(\omega)
\leq \int_{\Sigma} \|\hat{A}^{(n-1)}(\omega)^{-1}\|^a \|A(x_{n-1}, x_n)^{-1}\|^a \prod_{j=0}^{n-1} K(x_j, dx_{j+1})
\leq C \int_{\Sigma} \|\hat{A}^{(n-1)}(\omega)^{-1}\|^a \prod_{j=0}^{n-2} K(x_j, dx_{j+1})
= C \int_{\Sigma} \|\hat{A}^{(n-1)}(\omega)^{-1}\|^a \, d\mathbb{P}_{x_0}(\omega) = C \mathbb{E}_{x_0}(\|\hat{A}^{(n-1)}(\omega)^{-1}\|^a).
\]

By induction we get
\[
\mathbb{E}_x(\|\hat{A}^{(n)}(\omega)^{-1}\|^a) \leq C^n,
\]
which proves item (a). Items (b) and (c) are straightforward. To prove (d) we use the formula
\[
\hat{A}^{(n)} - \hat{B}^{(n)} = \sum_{j=0}^{n-1} (\hat{A}^{(j)} \circ T^{n-j}) (\hat{A} \circ T^{n-1-j} - \hat{B} \circ T^{n-1-j}) \hat{B}^{(n-1-j)}.
\]

Assumption (C2), on the irreducibility of the cocycle \( A \), is used in the following lemma.

**Lemma 7.14.** If \((K, \mu, A)\) satisfies (C1)-(C3) then
\[
\lim_{n \to +\infty} \frac{1}{n} \mathbb{E}_x(\log \|\hat{A}^{(n)}(p)\|) = L_1(A, \mu),
\]
with uniform convergence in \((x, p) \in \Sigma \times \mathbb{P}(\mathbb{R}^m)\).

**Proof.** See Lemma 3.1 in [4].

Given \( A \in \mathcal{B}(K,a) \) and \( 0 < \alpha \leq 1 \), define for all \( n \in \mathbb{N} \),
\[
\xi_n^\alpha(A) := \sup_{x \in \Sigma, p \neq q} \mathbb{E}_x \left[ \left( \frac{\delta(\hat{A}^{(n)}(p), \hat{A}^{(n)}(q))}{\delta(p, q)} \right)^\alpha \right] \in [0, +\infty] \quad (7.11)
\]

The following lemma highlights the importance of this quantity.
Lemma 7.15. Given $A \in \mathcal{B}(K, a)$, $f \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ and $n \in \mathbb{N}$, 

$$V_\alpha((Q_K A)^n f) \leq \xi_n^\alpha(A) V_\alpha(f).$$

Proof. Note that for any $f \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$, and $(x_0, p) \in \Sigma \times \mathbb{P}(\mathbb{R}^m)$,

$$(Q_{\hat{K}} A)^n f)(x_0, p) =$$

$$= \int_{\Sigma^n} f(x_n, A(x_{n-1}, x_n) \ldots A(x_0, x_1) p) \prod_{j=0}^{n-1} K(x_j, dx_{j+1})$$

$$= \mathbb{E}_{x_0} \left[ f(e_n, \hat{A}(n)p) \right].$$

Hence

$$V_\alpha((Q_{\hat{K}} A)^n f) = \sup_{x \in \Sigma, p \neq q} \frac{|(Q_{\hat{K}} A)^n f)(x, p) - (Q_{\hat{K}} A)^n f)(x, q)|}{\delta(p, q)\alpha}$$

$$= \sup_{x \in \Sigma, p \neq q} \frac{\mathbb{E}_x \left[ |f(e_n, \hat{A}(n)p) - f(e_n, \hat{A}(n)q)|\right]}{\delta(p, q)\alpha}$$

$$\leq \sup_{x \in \Sigma, p \neq q} \mathbb{E}_x \left[ \left(\frac{\delta(\hat{A}(n)p, \hat{A}(n)q)}{\delta(p, q)}\right)^\alpha\right]$$

$$= V_\alpha(f) \sup_{x \in \Sigma, p \neq q} \mathbb{E}_x \left[ \left(\frac{\delta(\hat{A}(n)p, \hat{A}(n)q)}{\delta(p, q)}\right)^\alpha\right]$$

$$= V_\alpha(f) \xi_n^\alpha(A).$$

\[\Box\]

Lemma 7.16. Given $A \in \mathcal{B}(K, a)$ and $n \in \mathbb{N}$, there exist constants $\alpha_0 > 0$ and $C > 0$, depending only on $\|A\|_\infty$ and $\|\eta_A\|_\infty$, such that $\xi_n^\alpha(A) \leq C$ for all $0 < \alpha \leq \alpha_0$.

Proof. To simplify the notation we write $M_n$ instead of $\hat{A}(n)$. By Lemma 7.11, given $x \in \Sigma$, and $p \neq q$ in $\mathbb{P}(\mathbb{R}^m)$,

$$\mathbb{E}_x \left[ \left(\frac{\delta(M_n p, M_n q)}{\delta(p, q)}\right)^\alpha\right] = \mathbb{E}_x \left[ \exp \left( \alpha \log \frac{\delta(M_n p, M_n q)}{\delta(p, q)}\right)\right]$$

$$\leq \mathbb{E}_x \left[ e^{2\alpha \ell(M_n)}\right] \leq \mathbb{E}_x \left[ \prod_{j=0}^{n-1} e^{2\alpha \ell(M_{n-j+1})}\right]$$

Since $A \in \mathcal{B}(K, a)$, we have $\mathbb{E}_x \left[ e^{\alpha \ell(M)}\right] \leq C$, for all $x \in \Sigma$. Thus $\mathbb{E}_\nu \left[ e^{\alpha \ell(M)}\right] \leq C$ for any initial distribution $\nu$ on $\Sigma$, and in particular,
$\mathbb{E}_x \left[ e^{\alpha \ell(M_1 \circ T^j)} \right] \leq C$ for all $j \geq 0$. Hence if $0 < \alpha \leq \frac{a}{2n}$, by the Hölder inequality we get
\[
\mathbb{E}_x \left[ \prod_{j=0}^{n-1} e^{2\alpha \ell(M_1 \circ T^j)} \right] \leq \prod_{j=0}^{n-1} \left| e^{2\alpha \ell(M_1 \circ T^j)} \right|_{L^p(x)}^{1/n}
\]
\[
= \left( \prod_{j=0}^{n-1} \mathbb{E}_x(e^{2\alpha \ell(M_1 \circ T^j)}) \right)^{1/n}
\]
\[
\leq \left( \prod_{j=0}^{n-1} \mathbb{E}_x(e^{\alpha \ell(M_1 \circ T^j)}) \right)^{1/n} \leq C.
\]
Taking the sup in $x$, and $p \neq q$, we obtain $\xi_n^\alpha \leq C$. \hfill \Box

**Corollary 7.17.** Given $A \in \mathcal{B}(K,a)$, for all $0 < \alpha \leq 1$ small enough the subspace $\mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ is invariant under $Q_{KA}$.

**Lemma 7.18.** The sequence $\{\xi_n^\alpha(A)\}_{n \geq 0}$ is sub-multiplicative, i.e.,
\[
\xi_n^\alpha(A) \leq \xi_n^\alpha(A) \xi_m^\alpha(A) \quad \text{for} \quad n, m \in \mathbb{N}.
\]
In particular,
\[
\lim_{n \to +\infty} \xi_n^\alpha(A)^{1/n} = \inf\{ \xi_n^\alpha(A)^{1/n} : n \in \mathbb{N} \}.
\]
**Proof.** We shall write $M_n = \hat{A}^{(n)}$. Given $x \in \Sigma$, $p \neq q$ in $\mathbb{P}(\mathbb{R}^m)$,
\[
\mathbb{E}_x \left[ \left( \frac{\delta(M_{n+m}p, M_{n+m}q)}{\delta(p,q)} \right)^\alpha \right] \leq
\leq \mathbb{E}_x \left[ \left( \frac{\delta((M_n \circ T^m)M_{mp}, (M_n \circ T^m)M_{mq})}{\delta(M_{mp}, M_{mq})} \right)^\alpha \left( \frac{\delta(M_{mp}, M_{mq})}{\delta(p,q)} \right)^\alpha \right]
\]
\[
\leq \xi_m^\alpha \mathbb{E}_x \left[ \left( \frac{\delta((M_n \circ T^m)M_{mp}, (M_n \circ T^m)M_{mq})}{\delta(M_{mp}, M_{mq})} \right)^\alpha \right]
\]
\[
\leq \xi_m^\alpha \sup_{p \neq q} \mathbb{E}_{K^m(x, \cdot)} \left[ \left( \frac{\delta(M_{mp}, M_{mq})}{\delta(p,q)} \right)^\alpha \right] \leq \xi_m^\alpha \xi_n^\alpha,
\]
and taking the sup we get $\xi_{n+m}^\alpha \leq \xi_n^\alpha \xi_m^\alpha$. \hfill \Box

**Lemma 7.19.** For some $n \in \mathbb{N}$, all $x \in \Sigma$ and $p \neq q$ in $\mathbb{P}(\mathbb{R}^m)$,
\[
\mathbb{E}_x \left[ \log \frac{\delta(\hat{A}^{(n)}p, \hat{A}^{(n)}q)}{\delta(p,q)} \right] \leq -1.
\]
Proof. We shall write \( M_n = \hat{A}^{(n)} \). Given \( x \in \Sigma \) and \( p \neq q \) in \( \mathbb{P}(\mathbb{R}^m) \),
\[
\frac{1}{n} \mathbb{E}_x \left[ \log \frac{\delta(M_n p, M_n q)}{\delta(p, q)} \right] \leq 0
\]
\[
\leq \frac{1}{n} \mathbb{E}_x \left[ \log \frac{\| (M_n p) \wedge (M_n q) \|}{\| M_n p \| \| M_n q \|} \frac{\| p \| \| q \|}{\| p \wedge q \|} \right] \leq \frac{1}{n} \mathbb{E}_x \left[ \log \frac{\| (M_n p) \wedge (M_n q) \|}{\| M_n p \| \| M_n q \|} \| p \| \| q \| \right] \]
\[
\leq \frac{1}{n} \mathbb{E}_x \left[ \log \| \Lambda_2 \hat{A}^{(n)}(\| \|) \right] - \frac{1}{n} \mathbb{E}_x \left[ \log \| \hat{A}^{(n)}(p) \| \right] - \frac{1}{n} \mathbb{E}_x \left[ \log \| \hat{A}^{(n)}(q) \| \right] ,
\]
and the right hand side converges to \( L^{(1)} + L^{(2)} - 2 L^{(1)} = L^{(2)} - L^{(1)} < 0 \). By Lemma 7.14 we have
\[
\lim_{n \to +\infty} \sup_{x \in \Sigma, p \neq q} \frac{1}{n} \mathbb{E}_x \left[ \log \frac{\delta(M_n p, M_n q)}{\delta(p, q)} \right] \leq L^{(2)} - L^{(1)} < 0 .
\]
Hence taking \( n \) large enough such that \( m (L^{(2)} - L^{(1)}) < -1 \) the Lemma follows.

\[\square\]

**Proposition 7.20.** If \((K, \mu, A)\) satisfies (C1)-(C3), there is a neighbourhood \( V \) of \( A \) in \( B(K, a) \) and there are positive constants \( \alpha_0, \alpha_1 \) and \( C \) such that \( 0 < \alpha_1 < \frac{\alpha_0}{2}, 0 < \sigma < 1 \) and
\[
V_\sigma((Q_B)^n f) \leq C \sigma^n V_\alpha(f) ,
\]
for all \( B \in \mathcal{V}, \alpha_1 \leq \alpha \leq \alpha_0, f \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m)) \) and \( n \in \mathbb{N} \).

**Proof.** We begin deriving a modulus of continuity for the functions \( B \mapsto \xi_n^\alpha(B) \). Fix a neighbourhood \( V \) of \( A \) in \( B(K, a) \) such that for \( B \in \mathcal{V}, \)
\[1 \quad \| \hat{B} \|_\infty \leq C, \]and
\[2 \quad \text{for all } B \in \mathcal{V} \text{ and } x \in \Sigma, \quad \mathbb{E}_x(\| \hat{B}^{-1} \|^2) \leq C. \]
Then, by Lemma 7.13, for \( B \in \mathcal{V}, x \in \Sigma \) and \( n \in \mathbb{N} \),
\[
\| \hat{B}^{(n)} \|_\infty \leq C^n , \quad \text{and} \quad \mathbb{E}_x(\| (\hat{B}^{(n)})^{-1} \|^{2(1-\alpha)}) \leq \mathbb{E}_x(\| (\hat{B}^{(n)})^{-1} \|)^2 \leq C^n .
\]
Therefore, using lemmas 7.12 and 7.13
\[
|\xi_n^\alpha(A) - \xi_n^\alpha(B)| \leq \sup_{x \in \Sigma, p \neq q} \mathbb{E}_x \left[ \left( \frac{\delta(\hat{A}^{(n)} p, \hat{A}^{(n)} q)}{\delta(p, q)} \right)^\alpha - \left( \frac{\delta(\hat{B}^{(n)} p, \hat{B}^{(n)} q)}{\delta(p, q)} \right)^\alpha \right] \]
\[
\leq \alpha \left( \sup_{B \in \mathcal{V}} \mathbb{E}_x(\| (\hat{B}^{(n)})^{-1} \|^{2(1-\alpha)}) \right) \left( \sup_{B \in \mathcal{V}} \| \hat{B}^{(n)} \|_\infty \right) \| \hat{A}^{(n)} - \hat{B}^{(n)} \|_\infty \]
\[
\leq \alpha C^{2n} \| \hat{A}^{(n)} - \hat{B}^{(n)} \|_\infty \leq \alpha n C^{2n} C^{-n} \| \hat{A} - \hat{B} \|_\infty = \alpha n C^{3n-1} d_\alpha(A, B). \]
Write $M_\alpha = \tilde{A}^{(\alpha)}$. We claim that for some $n_0 \in \mathbb{N}$ and $0 < \alpha_0 \leq 1$ small enough, $\xi_{n_0}^{\alpha}(A) < 1$. For that we use the following inequality

$$e^x \leq 1 + x + \frac{x^2}{2} |e^x|.$$ 

Choose $n_0 \in \mathbb{N}$ given by Lemma 7.19. Given $x \in \Sigma$, $p \neq q$ in $\mathbb{P}(\mathbb{R}^m)$,

$$\mathbb{E}_x \left[ \left( \frac{\delta(M_{n_0}p, M_{n_0}q)}{\delta(p, q)} \right)^{\alpha} \right] = \mathbb{E}_x \left[ \exp \left( \alpha \log \frac{\delta(M_{n_0}p, M_{n_0}q)}{\delta(p, q)} \right) \right]$$

$$\leq \mathbb{E}_x \left[ 1 + \alpha \log \frac{\delta(M_{n_0}p, M_{n_0}q)}{\delta(p, q)} + \frac{\alpha^2}{2} \log^2 \frac{\delta(M_{n_0}p, M_{n_0}q)}{\delta(p, q)} \left( \frac{\delta(M_{n_0}p, M_{n_0}q)}{\delta(p, q)} \right)^\alpha \right]$$

$$\leq 1 - \alpha + \frac{\alpha^2}{2} \mathbb{E}_x \left[ 16 \ell(M_{n_0})^2 \exp(\alpha \ell(M_{n_0})) \right] \leq 1 - \alpha + O(\alpha^2).$$

Arguing as in Lemma 7.16 we see that that $\mathbb{E}_x \left[ 16 \ell(M_{n_0})^2 \exp(\alpha \ell(M_{n_0})) \right]$ is finite and uniformly bounded in $x$, $p$ and $q$, for all $\alpha > 0$ small enough. This explains the last inequality. Taking $\alpha > 0$ sufficiently small the right-hand-side above is less than 1, which proves that $\xi_{n_0}^{\alpha}(A) < 1$.

In fact we can choose $0 < \alpha_1 < \frac{\alpha_0}{7}$ and $0 < \rho < 1$ such that for all $\alpha_1 \leq \alpha \leq \alpha_0$, $\xi_{n_0}^{\alpha}(A) \leq \rho$. To extend this inequality to all cocycles $B \in \mathcal{V}$, pick any $\rho' \in ]\rho, 1[\)$ and choose $\delta > 0$ such that $\alpha_0 n_0 C^{3\alpha_0 - 1} \delta < \rho' - \rho$. Make the neighbourhood $\mathcal{V}$ small enough so that $d_{\Sigma}(A, B) < \delta$ for all $B \in \mathcal{V}$. Then for every $B \in \mathcal{V}$ and $\alpha_1 \leq \alpha \leq \alpha_0$,

$$|\xi_{n_0}^{\alpha}(A) - \xi_{n_0}^{\alpha}(B)| < \rho' - \rho$$,

which implies

$$\xi_{n_0}^{\alpha}(B) \leq \xi_{n_0}^{\alpha}(A) + |\xi_{n_0}^{\alpha}(A) - \xi_{n_0}^{\alpha}(B)| < \rho'.$$

By Lemma 7.16 there are constants $\alpha_0' > 0$ and $C' > 0$, uniform in the neighbourhood $\mathcal{V}$ of $A$, such that $\xi_{\alpha_0}(B) \leq C'$ for all $B \in \mathcal{V}$, $0 < \alpha \leq \alpha_0'$ and $0 \leq j \leq n_0$. Shrinking if necessary the constants $\alpha_1$ and $\alpha_0$ above, we may assume that $\alpha_0' = \alpha_0$. Thus, because the sequence $\{\xi_{n}^{\alpha}(B)\}_{n \geq 0}$ is sub-multiplicative, letting $\sigma = (\rho')^{1/n_0}$ we have $\xi_{n}^{\alpha}(B) \leq C' \sigma^n$ for all $B \in \mathcal{V}$, $n \in \mathbb{N}$ and $\alpha_1 \leq \alpha \leq \alpha_0$. By Lemma 7.15 this concludes the proof.

\textit{Proof of Proposition 7.6} We begin with some preliminary considerations necessary to the proof. Given $(K, \mu, A)$ satisfying (C1)-(C3) take the neighbourhood $\mathcal{V}$ of $A$ provided by Proposition 7.20. For each $B \in \mathcal{V}$ let $\mu_B$ denote the $K_B$-stationary measure on $\Sigma \times \mathbb{P}(\mathbb{R}^m)$, and for each $x \in \Sigma$, denote by $\mu_{B,x} := \mu_B(\cdot | x)$ the conditional measure on $\mathbb{P}(\mathbb{R}^m)$. The canonical projection $\pi : \Sigma \times \mathbb{P}(\mathbb{R}^m) \to \Sigma$, $\pi(x, p) = x$,
binds the probability measures $\mu_B$ on $\Sigma \times \mathbb{P}(\mathbb{R}^m)$ with $\mu = \pi_* \mu_B$ on $\Sigma$, and the conditional measures $\mu_{B,x}$ on $\mathbb{P}(\mathbb{R}^m)$ through the following relation
\[
\int \int f(x,p) \mu_B(dx,dp) = \int \int_{\mathbb{P}(\mathbb{R}^m)} f(x,p) \mu_{B,x}(dp) \mu(dx),
\]
which holds for all $f \in L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^m))$. We claim that for $\mu$-a.e. $x \in \Sigma$ the following projective measures coincide
\[
\int \int_{\mathbb{P}(\mathbb{R}^m)} f(x,p) \nu_x(dp) \mu(dx) = \int \int_{\mathbb{P}(\mathbb{R}^m)} f(x,p) \nu'_x(dp) \mu(dx),
\]
for every $f \in L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^m))$. A straightforward computation shows that integrating $f(x,p)$ w.r.t. the left-hand-side in (7.12) we get $\int f \mu_B$, while integrating the same function w.r.t. the right-hand-side we obtain $\int Q_B f d\mu_B$. Thus, (7.12) holds because $\mu_B$ is $K_B$-stationary. Consider now the projection operator $\Pi_B : L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^m)) \to L^\infty(\Sigma)$ defined by
\[
(\Pi_B f)(x) := \int_{\mathbb{P}(\mathbb{R}^m)} f(x,p) \mu_{B,x}(dp).
\]
Because of (7.12), the following diagram commutes
\[
\begin{array}{ccc}
L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^m)) & \xrightarrow{Q_B} & L^\infty(\Sigma \times \mathbb{P}(\mathbb{R}^m)) \\
\Pi_B \downarrow & & \downarrow \Pi_B \\
L^\infty(\Sigma) & \xrightarrow{Q_K} & L^\infty(\Sigma)
\end{array}
\]
Define the three linear subspaces of $\mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$:
\[
E_B := \{ \text{constant functions} \} \\
G_B := \{ f \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m)) : \Pi_B f = 0 \} \\
H_B := \{ f \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m)) : V_\alpha(f) = 0 \}
\]
All these spaces are $Q_B$-invariant. The first since $Q_B 1 = 1$. $G_B$ is invariant because of the commutative diagram above. Finally, the subspace $H_B$ is invariant by Proposition 7.20. Hence we have the following
$Q_B$-invariant direct sum decomposition

$$\mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m)) = E_B \oplus G_B \oplus H_B.$$  

In fact, each function $f \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$ can be written as

$$f = \langle f, \mu_B \rangle 1 + (h - \Pi_B h) + \Pi_B h,$$

(7.13)  

with $h = f - \langle f, \mu_B \rangle 1$. The first term is clearly in $E_B$, the second lies in $G_B$ since $\Pi_B$ is a projection, and the third belongs to $H_B$ because the range of the projection $\Pi_B$ is contained in $H_B$.

To prove the proposition consider the following seminorms on the space $\mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m))$

$$\|f\|^\ast := \sup_{x \in \Sigma} \left| \int_{\mathbb{P}(\mathbb{R}^m)} f(x, p) \mu_{B,x}(dp) \right|,$$

$$\|f\|^\ast_\alpha := V_\alpha(f) + \|f\|^\ast.$$  

The second, $\|\cdot\|^\ast_\alpha$, is a norm equivalent to $\|\cdot\|_\alpha$, a fact that follows from the inequalities

$$\|f\|^\ast \leq \|f\|_\infty \leq V_0(f) + \|f\|^\ast \leq V_\alpha(f) + \|f\|^\ast.$$  

Using the decomposition (7.13) and Proposition 7.20, we have

$$V_\alpha((Q_B)^n f - \langle f, \mu_B \rangle 1) = V_\alpha((Q_B)^n[f - \langle f, \mu_B \rangle 1]) = V_\alpha((Q_B)^n(h - \Pi_B h + \Pi_B h)) = V_\alpha((Q_B)^n(h - \Pi_B h)) \leq C \sigma^n V_\alpha(h - \Pi_B h) = C \sigma^n V_\alpha(h) = C \sigma^n V_\alpha(f).$$

On the other hand, because of (C1) we may assume that the constants $C > 0$ and $0 < \sigma < 1$ are such that for all $n \in \mathbb{N}$ and $g \in L^\infty(\Sigma)$,

$$\|(Q_K)^n g - \langle g, \mu \rangle 1\|_\infty \leq C \sigma^n \|g\|_\infty.$$  

Therefore, considering $g = \Pi_B h$,

$$\|(Q_B)^n f - \langle f, \mu_B \rangle 1\|^\ast = \|(Q_B)^n[f - \langle f, \mu_B \rangle 1]\|^\ast = \|(Q_B)^n(h)\|^\ast = \|(Q_B)^n(\Pi_B h)\|^\ast = \|(Q_K)^n(\Pi_B h)\|_\infty \leq C \sigma^n \|\Pi_B h\|_\infty = C \sigma^n \|h\|^\ast \leq C \sigma^n \|f\|^\ast.$$  

The two driven inequalities imply

$$\|(Q_B)^n f - \langle f, \mu_B \rangle 1\|^\ast_\alpha \leq C \sigma^n \|f\|^\ast_\alpha,$$

which finishes the proof. □
Next lemma implies (A2).

**Lemma 7.21.** Given $A \in \mathcal{B}(K,a)$, for all $f \in \mathcal{H}_a(\Sigma \times \mathbb{P}^1(\mathbb{R}^m))$, and $z \in \mathbb{C}$ such that Re $z \leq b$,

$$∥Q_{A,z}f∥_∞ \leq ∥A∥^b_∞ ∥f∥_∞ .$$

**Proof.** Given $z \in \mathbb{C}$ such that Re $z \leq b$,

$$|(Q_{A,z}f)(x,p)| = \left| \mathbb{E}_x \left[ \|\tilde{A}p\|^z f(e_1,\tilde{A}p) \right] \right| ≤ ∥f∥_∞ \mathbb{E}_x \left[ ∥\tilde{A}∥^{\text{Re}z} \right] ≤ ∥f∥_∞ ∥A∥^b_∞ .$$

□

**Lemma 7.22.** Given $M \in \text{GL}(d,\mathbb{R})$ and $b > 0$, there are positive constants $κ_i$, $1 ≤ i ≤ 3$, depending only on $b$ and $∥M∥$, such that for any $0 < α ≤ 1$, $p, q \in \mathbb{P}(\mathbb{R}^m)$ and $z \in \mathbb{C}$ with $|z| ≤ b$,

$$(1) \sup_{p \neq q} \frac{∥Mp∥^z - ∥Mq∥^z}{δ(p,q)^α} ≤ κ_1,$$

$$(2) \sup_{p \neq q} \frac{∥\log∥Mp∥∥^i - ∥\log∥Mq∥∥^i}{δ(p,q)^α} ≤ κ_2 \text{ for } i = 1, 2,$$

$$(3) \sup_{p \neq q} \frac{∥\log∥Mp∥∥^i ∥Mp∥^z - ∥\log∥Mq∥∥^i ∥Mq∥^z}{δ(p,q)^α} ≤ κ_3 \text{ for } i = 1, 2 .$$

**Proof.** See Lemma 4.2 in chapter V of [5] for items (1)-(2). For the sake of completeness we include the proofs here. The following inequalities are needed: for any $p, q \in \mathbb{P}(\mathbb{R}^m)$, there are unit vector representatives of $p$ and $q$ such that

$$∥p - q∥ ≤ √2 \text{ and } ∥p - q∥ ≤ √2 δ(p,q) .$$

Defining $F_z : \mathbb{P}(\mathbb{R}^m) \rightarrow \mathbb{C}$, $F_z(p) := ∥Mp∥^z$, we have

$$∥(DF_z)_p∥ ≤ |z| ∥M∥ ∥Mp∥^{\text{Re}z - 1} ≤ b ∥M∥^b .$$

Hence

$$\frac{∥Mp∥^z - ∥Mq∥^z}{δ(p,q)^α} ≤ b ∥M∥^b \left( \frac{∥p - q∥}{δ(p,q)} \right)^α ∥p - q∥^{1-α} ≤ b ∥M∥^b (√2)^α (√2)^{1-α} = √2 b ∥M∥^b .$$

Analogously, defining $G : \mathbb{P}(\mathbb{R}^m) \rightarrow \mathbb{R}$, $G(p) := ∥\log∥Mp∥$, we have

$$∥DG_p∥ = \frac{∥(DF_1)_p∥}{∥Mp∥} ≤ \frac{∥M∥}{∥Mp∥} = \frac{∥M∥ ∥M^{-1}(Mp)∥}{∥Mp∥} ≤ ∥M∥ ∥M^{-1}∥ .$$
and

\[ \frac{|\log \| M p \| - \log \| M q \| |}{\delta(p, q)^\alpha} \leq \| M \| \| M^{-1} \| \left( \left\| p - q \right\| \frac{\delta(p, q)}{\delta(p, q)^\alpha} \right) \| p - q \|^{1-\alpha} \]

\[ \leq \| M \| \| M^{-1} \| (\sqrt{2})^\alpha (\sqrt{2})^{1-\alpha} = \sqrt{2} \| M \| \| M^{-1} \| , \]

and

\[ \frac{(\log \| M p \|)^2 - (\log \| M q \|)^2}{\delta(p, q)^\alpha} \leq 2 \log \| M \| \frac{|\log \| M p \| - \log \| M q \| |}{\delta(p, q)^\alpha} \]

\[ \leq 2 \sqrt{2} (\log \| M \|) \| M \| \| M^{-1} \| . \]

Finally (3) follows from (1) and (2) using the following inequality

\[ V_\alpha(F G) \leq V_\alpha(F) \| G \|_\infty + \| F \|_\infty V_\alpha(G). \]

Notice that \( \| F \|_\infty \leq \| M \| b \) and \( \| G \|_\infty \leq \max\{\log \| M \|, \log^2 \| M \|\} \).

Choose \( 0 < \alpha_1 < \alpha_0 \leq 1 \) according to Proposition 7.20. Assumption (A3) follows from:

**Lemma 7.23.** Given \( A \in \mathcal{B}(K, a) \) and \( b > 0 \), there is a constant \( C_1 > 0 \), depending only on \( b \), \( \| A \|_\infty \) and \( \xi_2^i(A) \) such that for all \( f \in \mathcal{H}_\alpha(\Sigma \times \mathbb{P}(\mathbb{R}^m)) \) with \( \alpha_1 \leq \alpha \leq \alpha_0 \), and all \( z \in \mathbb{C} \) such that \( |z| \leq b \),

\[ V_\alpha(Q_{A,z}(f \xi_A^i)) \leq C_1(b, A) \| f \|_\alpha \quad i = 0, 1, 2. \]

**Proof.** Applying Lemma 7.22 (1), we get

\[ \frac{|Q_{A,z}(f \xi_A^i)(x, p) - Q_{A,z}(f \xi_A^i)(x, q)|}{\delta(p, q)^\alpha} = \]

\[ \leq \frac{\mathbb{E}_x \left[ \left| \| \hat{A} p \| \cdot |\log \| \hat{A} p \| |^i f(e_1, \hat{A} p) - \| \hat{A} q \| \cdot |\log \| \hat{A} q \| |^i f(e_1, \hat{A} q) | \right] }{\delta(p, q)^\alpha} \]

\[ \leq \frac{\mathbb{E}_x \left[ \left| \| \hat{A} p \| \cdot |\log \| \hat{A} p \| |^i \| \hat{A} p \|^z - \| \log \| \hat{A} q \| |^i \| \hat{A} q \|^z | \right] }{\delta(p, q)^\alpha} + V_\alpha(f) \mathbb{E}_x \left[ |\log \| \hat{A} \| |^i \| \hat{A} \|^{\Re z} \left( \frac{\delta(\hat{A} p, \hat{A} q)}{\delta(p, q)} \right)^\alpha \right] \]

\[ \leq \kappa_3 \| f \|_\infty + V_\alpha(f) \| \log \| A \|_\infty |^i \| A \|_\infty \sup_{x, y \neq q} \mathbb{E}_x \left[ \left( \frac{\delta(\hat{A} p, \hat{A} q)}{\delta(p, q)} \right)^\alpha \right] \]

\[ \leq C_1 \| f \|_\alpha , \]

where \( C_1 \) is the maximum between \( \kappa_3 \) and \( \| \log \| A \|_\infty |^i \| A \|_\infty \|^{b_0} \xi_1^i(A) \), for \( i = 0, 1, 2. \)
Finally the next lemma proves condition (A4).

**Lemma 7.24.** Given $A, B \in \mathcal{B}(K, a)$ and $b > 0$, there is a constant $C_2 > 0$, depending only on $b$, $\|A\|_\infty$ and $\|B\|_\infty$ such that for all $f \in \mathcal{H}_\alpha (\Sigma \times \mathbb{P}(\mathbb{R}^m))$, and all $z \in \mathbb{C}$ such that $\Re z \leq b$,

$$\|Q_{A,z}f - Q_{B,z}f\|_\infty \leq C_2 \|A - B\|_\infty^\alpha \|f\|_\alpha .$$

**Proof.** A simple computation shows that for all $z \in \mathbb{C}$ with $\Re z \leq b$, and all $A, B \in \text{GL}(d, \mathbb{R})$,

$$\|A p\|^z - \|B p\|^z \leq b \max\{\|A\|^{b-1}, \|B\|^{b-1}\} \|A - B\| .$$

Hence

$$| (Q_{A,z}f - Q_{B,z}f)(x, p) | \leq \mathbb{E}_x \left[ | \|\hat{A} p\|^z f(e_1, \hat{A} p) - \|\hat{B} p\|^z f(e_1, \hat{B} p) | \right]$$

$$\leq \|f\|_\infty \mathbb{E}_x \left[ | \|\hat{A} p\|^z - \|\hat{B} p\|^z | \right]$$

$$+ \|\hat{B}\|_\infty^b \mathbb{E}_x \left[ | f(e_1, \hat{A} p) - f(e_1, \hat{B} p) | \right]$$

$$\leq b \max\{\|A\|^{b-1}, \|B\|^{b-1}\} \|A - B\|_\infty \|f\|_\infty$$

$$+ \|\hat{B}\|_\infty^b V_\alpha(f) \mathbb{E}_x \left[ \delta(\hat{A} p, \hat{B} p)^\alpha \right]$$

$$\leq b \max\{\|A\|^{b-1}, \|B\|^{b-1}\} \|A - B\|_\infty \|f\|_\infty$$

$$+ \|\hat{B}\|_\infty^b V_\alpha(f) \|A - B\|_\infty^\alpha \leq C_2 \|f\|_\alpha \|A - B\|_\infty ,$$

where $C_2 = \max\{\|B\|_\infty^b, b \|B\|_\infty^{b-1}, b \|A\|_\infty^{b-1}\}$. \qed

**Theorem 7.2.** If $(K, \mu)$ is strongly mixing then any cocycle $A \in \mathcal{B}(K, a)$ for which (C1)-(C3) hold satisfies a uniform fiber-LDT estimate, with parameter space $\mathcal{F}$, for the measure preserving dynamical system $(\Omega, \mathcal{F}, \mathbb{P}_\mu, T)$.

**Proof.** As we have seen, the family of observed Markov systems $\{(K_A, \mu_A, \xi_A)\}_{A \in \mathcal{V}}$, where $\mathcal{V}$ is the neighbourhood of $A$ in $\mathcal{B}(K, a)$ provided by Proposition 7.6, satisfies last section’s assumptions (A1)-(A4). Hence, by Theorem 6.2, there is a smaller neighbourhood $\mathcal{V}_0$ of $A$ in $\mathcal{B}(K, a)$, there exist constants $\varepsilon_0 > 0$, $C > 0$ and $h > 0$ such that for all $B \in \mathcal{V}_0$, $0 < \varepsilon < \varepsilon_0$, $n \in \mathbb{N}$, $x \in \Sigma$ and $p \in \mathbb{P}(\mathbb{R}^m)$,

$$\mathbb{P}_x \left[ \left| \frac{1}{n} \log \|B^{(n)} p\| - L_1(B, \mu) \right| \geq \varepsilon \right] \leq C e^{-\frac{\varepsilon^2}{2h} n} ,$$

and integrating w.r.t. $\mu$ we get for all $p \in \mathbb{P}(\mathbb{R}^m)$,

$$\mathbb{P}_\mu \left[ \left| \frac{1}{n} \log \|B^{(n)} p\| - L_1(B, \mu) \right| \geq \varepsilon \right] \leq C e^{-\frac{\varepsilon^2}{2h} n} .$$
Choose the canonical basis \( \{ e_1, \ldots, e_d \} \) of \( \mathbb{R}^m \) and consider the following norm \( \| \cdot \|' \) on Mat\(_d(\mathbb{R})\), \( \| M \|' := \max_{1 \leq j \leq d} \| M e_j \| \). Since this norm is equivalent to the operator norm, for all \( B \in \mathcal{V}_0 \), \( n \in \mathbb{N} \) and \( p \in \mathbb{P}(\mathbb{R}^m) \),

\[
\| B^{(n)} p \| \leq \| B^{(n)} \| \leq C \| B^{(n)} \|' = C \max_{1 \leq j \leq d} \| B^{(n)} e_j \| .
\]

Thus a simple comparison of the deviation sets gives

\[
\mathbb{P}_\mu \left[ \frac{1}{n} \log \| B^{(n)} \| - L_1(B, \mu) \geq \varepsilon \right] \lesssim e^{-\varepsilon^2 n}
\]

for all \( B \in \mathcal{V}_0 \), \( 0 < \varepsilon < \varepsilon_0 \) and \( n \in \mathbb{N} \).

\[ \square \]

Acknowledgments. Both authors would like to acknowledge fruitful conversations they had with A. Baraviera, G. Del Magno and J. P. Gaivão about large deviations estimates for random cocycles.

The first author was partially supported by Fundação para a Ciência e a Tecnologia through the Program POCI 2010.

The second author was supported by the Norwegian Research Council project no. 213638, ”Discrete Models in Mathematical Analysis”. He is also grateful to the University of Lisbon and to the Institut Mittag-Leffler (through its ”Research in Peace” program) for their hospitality during the summer of 2013, when this project began.

References

[1] Ludwig Arnold, Random dynamical systems, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 1998. MR 1723992 (2000m:37087)
[2] A. Ávila, S. Jitomirskaya, and C. Sadel, Complex one-frequency cocycles, preprint (2013), 1–15.
[3] C. Bocker-Neto and M. Viana, Continuity of Lyapunov exponents for random 2d matrices, preprint (2010), 1–38.
[4] Philippe Bougerol, Théorèmes limite pour les systèmes linéaires à coefficients markoviens, Probab. Theory Related Fields 78 (1988), no. 2, 193–221. MR 945109 (89i:60122)
[5] Philippe Bougerol and Jean Lacroix, Products of random matrices with applications to Schrödinger operators, Progress in Probability and Statistics, vol. 8, Birkhäuser Boston, Inc., Boston, MA, 1985. MR 886674 (88f:60013)
[6] J. Bourgain, Green’s function estimates for lattice Schrödinger operators and applications, Annals of Mathematics Studies, vol. 158, Princeton University Press, Princeton, NJ, 2005. MR 2100420 (2005j:35184)
[7] ________, Positivity and continuity of the Lyapounov exponent for shifts on \( \mathbb{T}^d \) with arbitrary frequency vector and real analytic potential, J. Anal. Math. 96 (2005), 313–355. MR 2177191 (2006i:47064)
ABSTRACT CONTINUITY OF LYAPUNOV EXponents

[8] J. Bourgain and S. Jitomirskaya, Continuity of the Lyapunov exponent for quasiperiodic operators with analytic potential, J. Statist. Phys. 108 (2002), no. 5-6, 1203–1218, Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.

[9] J. L. Doob, Stochastic processes, Wiley Classics Library, John Wiley & Sons, Inc., New York, 1990, Reprint of the 1953 original, A Wiley-Interscience Publication. MR 1038526 (91d:60002)

[10] Pedro Duarte and Silvius Klein, Continuity of the Lyapunov Exponents for Quasiperiodic Cocycles, Comm. Math. Phys. 332 (2014), no. 3, 1113–1166. MR 3262622

[11] Alex Furman, On the multiplicative ergodic theorem for uniquely ergodic systems, Ann. Inst. H. Poincaré Probab. Statist. 33 (1997), no. 6, 797–815. MR 1484541 (98i:28018)

[12] H. Furstenberg and Y. Kifer, Random matrix products and measures on projective spaces, Israel J. Math. 46 (1983), no. 1-2, 12–32. MR 727020 (85i:22010)

[13] Michael Goldstein and Wilhelm Schlag, Hölder continuity of the integrated density of states for quasi-periodic Schrödinger equations and averages of shifts of subharmonic functions, Ann. of Math. (2) 154 (2001), no. 1, 155–203. MR 1847592 (2002h:82055)

[14] ———, Fine properties of the integrated density of states and a quantitative separation property of the Dirichlet eigenvalues, Geom. Funct. Anal. 18 (2008), no. 3, 755–869. MR 2438997 (2010h:47063)

[15] Hubert Hennion and Loïc Hervé, Limit theorems for Markov chains and stochastic properties of dynamical systems by quasi-compactness, Lecture Notes in Mathematics, vol. 1766, Springer-Verlag, Berlin, 2001. MR 1862393 (2002h:60146)

[16] S. Jitomirskaya and C. A. Marx, Continuity of the Lyapunov exponent for analytic quasi-periodic cocycles with singularities, J. Fixed Point Theory Appl. 10 (2011), no. 1, 129–146. MR 2825743 (2012h:37095)

[17] ———, Analytic quasi-periodic cocycles with singularities and the Lyapunov exponent of extended Harper’s model, Comm. Math. Phys. 316 (2012), no. 1, 237–267. MR 2989459

[18] Svetlana Jitomirskaya and Rajinder Mavi, Continuity of the measure of the spectrum for quasiperiodic Schrödinger operators with rough potentials, Comm. Math. Phys. 325 (2014), no. 2, 585–601. MR 3148097

[19] Yitzhak Katznelson and Benjamin Weiss, A simple proof of some ergodic theorems, Israel J. Math. 42 (1982), no. 4, 291–296. MR 682312 (84i:28020)

[20] Silvius Klein, Localization for quasiperiodic Schrödinger operators with multivariable Gevrey potential functions, To appear, Journal of Spectral Theory (2013), 1–24.

[21] S G Krein and Yu I Petunin, Scales of banach spaces, Russian Mathematical Surveys 21 (1966), no. 2, 85.

[22] Émilie Le Page, Théorèmes limites pour les produits de matrices aléatoires, Probability measures on groups (Oberwolfach, 1981), Lecture Notes in Math., vol. 928, Springer, Berlin-New York, 1982, pp. 258–303. MR 69072 (84d:60012)

[23] E. Malheiro and M. Viana, Lyapunov exponents of linear cocycles over markov shifts, preprint (2014), 1–25.
[24] Émile le Page, \textit{Régularité du plus grand exposant caractéristique des produits de matrices aléatoires indépendantes et applications}, Annales de l’institut Henri Poincaré (B) Probabilités et Statistiques \textbf{25} (1989), no. 2, 109–142 (fre).

[25] F. Riesz and B. Szőkefalvi-Nagy, \textit{Functional analysis}, Ungar, 1955.

[26] David Ruelle, \textit{Ergodic theory of differentiable dynamical systems}, Inst. Hautes Études Sci. Publ. Math. (1979), no. 50, 27–58. MR 556581 (81f:58031)

[27] Wilhelm Schlag, \textit{Regularity and convergence rates for the Lyapunov exponents of linear cocycles}, J. Mod. Dyn. \textbf{7} (2013), no. 4, 619–637. MR 3177775

[28] M. Viana, \textit{Lectures on Lyapunov exponents}, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2014.

Departamento de Matemática and CMAF, Faculdade de Ciências, Universidade de Lisboa, Portugal

\textit{E-mail address: pduarte@ptmat.fc.ul.pt}

Department of Mathematical Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway, and IMAR, Bucharest, Romania

\textit{E-mail address: silvius.klein@math.ntnu.no}