The Annals of Applied Probability
2014, Vol. 24, No. 1, 422–446
DOI: 10.1214/13-AAP929

LOSS OF MEMORY OF HIDDEN MARKOV MODELS AND LYAPUNOV EXPONENTS[1]

BY PIERRE COLLET[2] AND FLORENCIA LEONARDI[3]

Ecole Polytechnique and Universidade de São Paulo

In this paper we prove that the asymptotic rate of exponential loss of memory of a finite state hidden Markov model is bounded above by the difference of the first two Lyapunov exponents of a certain product of matrices. We also show that this bound is in fact realized, namely for almost all realizations of the observed process we can find symbols where the asymptotic exponential rate of loss of memory attains the difference of the first two Lyapunov exponents. These results are derived in particular for the observed process and for the filter; that is, for the distribution of the hidden state conditioned on the observed sequence. We also prove similar results in total variation.

1. Introduction. Let $(X_t)_{t\in\mathbb{Z}}$ be a Markov chain over a finite alphabet $\mathscr{A}$. We consider a probabilistic function $(Z_t)_{t\in\mathbb{Z}}$ of this chain, a model introduced by Petrie (1969). More precisely, there is another finite alphabet $\mathscr{B}$ and for any X_t we choose at random a Z_t in $\mathscr{B}$. The random choice of Z_t depends only on the value X_t of the original process at time t. The process (Z_t) is the observed process and (X_t) is the hidden process. This model is called a hidden Markov process.

We are interested in the asymptotic loss of memory of the processes $(X_t)_{t\in\mathbb{Z}}$ and $(Z_t)_{t\in\mathbb{Z}}$ conditioned on the observed sequence. For example, if the conditional probability of Z_t given X_t does not depend on X_t, the process $(Z_t)_{t\in\mathbb{Z}}$ is an independent process. Another trivial example is when there is no random choice, namely $Z_t = X_t$, in this case the process $(Z_t)_{t\in\mathbb{Z}}$ is Markovian. However, as we will see, under natural assumptions, the process $(Z_t)_{t\in\mathbb{Z}}$ has infinite memory. On the other hand, a particularly interesting question from the point of view of applications is to consider the loss of memory of the filter; that is, the distribution of X_0 conditioned on the past observed sequence $Z_{-1}, \dots, Z_{-n+1}$ and for different initial conditions on X_{-n}; see, for example, Cappé, Moulines and Rydén (2005).

Received November 2011.

[1]Supported in part by FAPESPs project *Consistent estimation of stochastic processes with variable length memory* (2009/09411-8), CNPqs projects *Stochastic systems with variable length interactions* (476501/2009-1) and USP project *Mathematics*, *computation*, *language and the brain* (11.1.9367.1.5).

[2]Supported in part by the DynEurBraz European project.

[3]Supported in part by a CNPq-Brazil fellowship.

MSC2010 subject classifications. Primary 60J2; secondary 62M09.

Key words and phrases. Random functions of Markov chains, Oseledec's theorem, perturbed processes.

Our goal is to investigate how fast these processes loose memory.

Exponential upper bounds for this asymptotic loss of memory have been obtained in various papers; see, for example, Douc, Moulines and Ritov (2009), Douc et al. (2009) and references therein. For the case of projections of Markov chains and the relation with Gibbs measures, see Chazottes and Ugalde (2011) and references therein.

In the present paper, under generic assumptions, we prove that the asymptotic rate of exponential loss of memory is bounded above by the difference of the first two Lyapunov exponents of a certain product of matrices. We also show that this bound is in fact realized, namely for almost all realizations of the process $(Z_t)_{t\in\mathbb{Z}}$, we can find symbols where the asymptotic exponential rate of loss of memory attains the difference of the first two Lyapunov exponents. As far as we know our results provide the first lower bounds for the loss of memory of these processes. Similar results (in particular lower bounds) are also obtained in the total variation distance.

Conditioned on the observed sequence $Z_{-1}, \dots, Z_{-n+1}$, we have considered different possibilities for the initial distribution at time $-n$, namely one can either give the initial distribution of X_{-n} or the initial distribution of Z_{-n}. Similarly one can ask for the distribution of X_0 (the hidden present state) or of Z_0 (the observable present state).

As an application, we consider the case of a randomly perturbed Markov chain with two symbols. We show that the asymptotic rate of loss of memory can be expanded in powers of the perturbation with a logarithmic singularity. This was our original motivation coming from our previous work with Galves [Collet, Galves and Leonardi (2008)].

The relation between product of random matrices and hidden Markov models was previously described in Jacquet, Seroussi and Szpankowski (2008). In this paper it was proved in particular that the first Lyapunov exponent is the opposite of the entropy of the process.

The content of the paper is as follows. In Section 2 we give a precise definition of the asymptotic exponential rate of loss of memory and state the main results about the relation of this rate with the first two Lyapunov exponents.

Proofs are given in Section 3. They rely on more general propositions which allow to treat at once the different situations of initial distributions and present distributions. In Section 4 we give the application to the random perturbation of a two states Markov chain.

2. Definitions and main results. Let $(X_t)_{t\in\mathbb{Z}}$ be an irreducible aperiodic Markov chain over a finite alphabet $\mathscr{A}$ with transition probability matrix $p(\cdot|\cdot)$ and unique invariant measure π. Without loss of generality we will assume $\mathscr{A} = \{1, 2, \dots, k\}$. In the sequel, we will use the shorthand notation x_r^s for a sequence of symbols $(x_r, \dots, x_s)$ $(r \le s)$. Consider another finite alphabet $\mathscr{B} = \{1, 2, \dots, \ell\}$, and a process $(Z_t)_{t\in\mathbb{Z}}$, a probabilistic function of the Markov chain $(X_t)_{t\in\mathbb{Z}}$

over $\mathscr{B}$. That is, there exists a matrix $q(\cdot|\cdot) \in \mathbb{R}^{k\times\ell}$ such that for any $n \geq 0$, any $z_0^n \in \mathscr{B}^{n+1}$ and any $x_0^n \in \mathscr{A}^{n+1}$, we have

$$\mathbb{P}\big(Z_0^n = z_0^n | X_0^n = x_0^n\big) = \prod_{i=0}^{n} \mathbb{P}(Z_i = z_i | X_i = x_i) = \prod_{i=0}^{n} q(z_i|x_i). \tag{2.1}$$

From now on, the symbol $\underline{z}$ will represent an element in $\mathscr{B}^{\mathbb{Z}}$. Define the shift-operator $\mathscr{S} : \mathscr{B}^{\mathbb{Z}} \to \mathscr{B}^{\mathbb{Z}}$ by

$$(\mathscr{S}\underline{z})_n = z_{n+1}.$$

The shift is invertible, and its inverse is given by

$$\big(\mathscr{S}^{-1}\underline{z}\big)_n = z_{n-1}.$$

To state our results we will need the following hypotheses:

(H1) $\min_{i,j} p(j|i) > 0$, $\min_{i,m} q(m|i) > 0$.
(H2) $\det(p) \neq 0$.
(H3) $\operatorname{rank}(q) = k$.

Note that hypothesis (H3) implies $\ell \geq k$.

For the convenience of the reader we recall Oseledec's theorem in finite dimension; see, for example, Katok and Hasselblatt (1995), Ledrappier (1984). As usual, we denote by $\log^+(x) = \max(\log(x), 0)$.

OSELEDEC'S THEOREM. *Let (Ω, μ) be a probability space and let T be a measurable transformation of Ω such that μ is T-ergodic. Let L_ω be a measurable function from Ω to $\mathscr{L}(\mathbb{R}^k)$ (the space of linear operators of $\mathbb{R}^k$ into itself). Assume the function L_ω satisfies*

$$\int \log^+ \|L_\omega\| \, d\mu(\omega) < +\infty.$$

Then, there exist $\lambda_1 > \lambda_2 > \cdots > \lambda_s$, with $s \leq k$ and there exists an invariant set $\tilde{\Omega} \subset \Omega$ of full measure ($\mu(\Omega \setminus \tilde{\Omega}) = 0$) such that for all $\omega \in \tilde{\Omega}$ there exist $s+1$ sub-vector spaces

$$\mathbb{R}^k = V_\omega^{(1)} \supsetneq V_\omega^{(2)} \supsetneq \cdots \supsetneq V_\omega^{(s+1)} = \{\vec{0}\}$$

such that for any $\vec{v} \in V_\omega^{(j)} \setminus V_\omega^{(j+1)}$ $(1 \leq j \leq s)$ we have

$$\lim_{n\to+\infty} \frac{1}{n} \log\big\|L_\omega^{[n]}\vec{v}\big\| = \lambda_j,$$

where $L_\omega^{[n]} = L_{T^{n-1}(\omega)} \cdots L_\omega$. Moreover, the subspaces satisfy the relation

$$L_\omega V_\omega^{(j)} \subseteq V_{T\omega}^{(j)}.$$

The numbers $\lambda_1, \lambda_2, \ldots, \lambda_s$ are called the Lyapunov exponents.

In the sequel we will use this theorem with $\Omega = \mathscr{B}^{\mathbb{Z}}$, μ the stationary ergodic measure of the process $(Z_t)_{t\in\mathbb{Z}}$ [Cappé, Moulines and Rydén (2005)], $T = \mathscr{S}^{-1}$ and $L_{\underline{z}}$ the linear operator in $\mathbb{R}^k$ with matrix given by

$$(L_{\underline{z}})_{i,j} = q(z_0|i)p(j|i).$$

With this notation, we have, for example,

$$\mathbb{P}(X_0 = a, Z_{-n+1}^{-1} = z_{-n+1}^{-1}, X_{-n} = b) = \langle \vec{\theta}_b, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{1}_a \rangle,$$

where $(\vec{\theta}_b)_i = p(i|b)$ and $\vec{1}_a$ is the basis vector with component number a equal to one.

From now on we will use the ℓ^2 norm $\|\cdot\|$ and the corresponding scalar product on $\mathbb{R}^k$. Note that from our definition of $L_{\underline{z}}$ we have

$$\sup_{\underline{z}} \|L_{\underline{z}}\| < +\infty.$$

Therefore we can apply Oseledec's theorem to get the existence of the Lyapunov exponents.

For any $\underline{z} \in \mathscr{B}^{\mathbb{Z}}$, for probabilities ρ on $\mathscr{A}$, η on $\mathscr{B}$ and any integer n, we define two probabilities on $\mathscr{A}$ by

$$\nu^{[n]}_{\underline{z},\rho}(a) = \frac{\sum_{b\in\mathscr{A}} \mathbb{P}(X_0 = a, Z_{-n+1}^{-1} = z_{-n+1}^{-1}, X_{-n} = b)\rho(b)}{\sum_{b\in\mathscr{A}} \mathbb{P}(Z_{-n+1}^{-1} = z_{-n+1}^{-1}, X_{-n} = b)\rho(b)}, \qquad a \in \mathscr{A},$$

and

$$\sigma^{[n]}_{\underline{z},\eta}(a) = \frac{\sum_{c\in\mathscr{B}} \mathbb{P}(X_0 = a, Z_{-n+1}^{-1} = z_{-n+1}^{-1}, Z_{-n} = c)\eta(c)}{\sum_{c\in\mathscr{B}} \mathbb{P}(Z_{-n+1}^{-1} = z_{-n+1}^{-1}, Z_{-n} = c)\eta(c)}, \qquad a \in \mathscr{A}.$$

These are the probabilities of X_0 conditioned on the observed string z_{-n+1}^{-1} when the distribution of X_{-n} is ρ (resp., the distribution of Z_{-n} is η).

When ρ is a Dirac measure concentrated on b we will simply denote the measure $\nu^{[n]}_{\underline{z},\rho}$ by $\nu^{[n]}_{\underline{z},b}$, and similarly for $\sigma^{[n]}_{\underline{z},\eta}$.

We can state now our main results.

THEOREM 2.1. *Under the hypothesis* (H1), *for each $a \in \mathscr{A}$, for any probabilities ρ and ρ' on $\mathscr{A}$*,

$$\limsup_{n\to+\infty} \frac{1}{n} \log\left|\nu^{[n]}_{\underline{z},\rho}(a) - \nu^{[n]}_{\underline{z},\rho'}(a)\right| \le \lambda_2 - \lambda_1,$$

μ-almost surely. Similarly, under the hypothesis (H1), *for each $a \in \mathscr{A}$, for any probabilities η and η' on $\mathscr{B}$*,

$$\limsup_{n\to+\infty} \frac{1}{n} \log\left|\sigma^{[n]}_{\underline{z},\eta}(a) - \sigma^{[n]}_{\underline{z},\eta'}(a)\right| \le \lambda_2 - \lambda_1,$$

μ-almost surely.

REMARK. When $\mathscr{A} = \mathscr{B}$ and q is the identity matrix, $(Z_t)_{t\in\mathbb{Z}} = (X_t)_{t\in\mathbb{Z}}$ is a Markov chain. The second part of hypothesis (H1) does not hold, but it is easy to adapt the proof of Theorem 2.1 for this particular case. It is easy to verify recursively that the matrices $L_{\underline{z}}^{[n]}$ are of rank one. The Lyapunov exponents can be computed explicitly. One gets $\lambda_1 = -H(p)$ (the entropy of the Markov chain with transition probability p) from the ergodic theorem, and $\lambda_2 = -\infty$ with multiplicity $k-1$.

THEOREM 2.2. *Under hypotheses* (H1)–(H2), *for μ-almost all $\underline{z}$ there exists $a, b, c \in \mathscr{A}$* (*which may depend on $\underline{z}$*) *such that*

$$\limsup_{n\to+\infty} \frac{1}{n} \log\left|\nu_{\underline{z},b}^{[n]}(a) - \nu_{\underline{z},c}^{[n]}(a)\right| = \lambda_2 - \lambda_1.$$

Under hypotheses (H1)–(H3), *for μ-almost all $\underline{z}$ there exists $a \in \mathscr{A}$, $b, c \in \mathscr{B}$* (*which may depend on $\underline{z}$*) *such that*

$$\limsup_{n\to+\infty} \frac{1}{n} \log\left|\sigma_{\underline{z},b}^{[n]}(a) - \sigma_{\underline{z},c}^{[n]}(a)\right| = \lambda_2 - \lambda_1.$$

As a corollary, we derive equivalent results for the loss of memory of the process $(Z_t)_{t\in\mathbb{Z}}$. For any $\underline{z} \in \mathscr{B}^{\mathbb{Z}}$, for probabilities ρ on $\mathscr{A}$, η on $\mathscr{B}$ and any integer n, we define two probabilities on $\mathscr{B}$ by

$$\tilde{\nu}_{\underline{z},\rho}^{[n]}(e) = \frac{\sum_{b\in\mathscr{A}} \mathbb{P}(Z_0 = e, Z_{-n+1}^{-1} = z_{-n+1}^{-1}, X_{-n} = b)\rho(b)}{\sum_{b\in\mathscr{A}} \mathbb{P}(Z_{-n+1}^{-1} = z_{-n+1}^{-1}, X_{-n} = b)\rho(b)}, \qquad e \in \mathscr{B},$$

and

$$\tilde{\sigma}_{\underline{z},\eta}^{[n]}(e) = \frac{\sum_{c\in\mathscr{B}} \mathbb{P}(Z_0 = e, Z_{-n+1}^{-1} = z_{-n+1}^{-1}, Z_{-n} = c)\eta(c)}{\sum_{c\in\mathscr{B}} \mathbb{P}(Z_{-n+1}^{-1} = z_{-n+1}^{-1}, Z_{-n} = c)\eta(c)}, \qquad e \in \mathscr{B}.$$

COROLLARY 2.3. *Under the hypothesis* (H1), *for each $e \in \mathscr{B}$, for any probabilities ρ and ρ' on $\mathscr{A}$*,

$$\limsup_{n\to+\infty} \frac{1}{n} \log\left|\tilde{\nu}_{\underline{z},\rho}^{[n]}(e) - \tilde{\nu}_{\underline{z},\rho'}^{[n]}(e)\right| \le \lambda_2 - \lambda_1,$$

μ-almost surely. Similarly, under the hypothesis (H1), *for each $e \in \mathscr{B}$, for any probabilities η and η' on $\mathscr{B}$*,

$$\limsup_{n\to+\infty} \frac{1}{n} \log\left|\tilde{\sigma}_{\underline{z},\eta}^{[n]}(e) - \tilde{\sigma}_{\underline{z},\eta'}^{[n]}(e)\right| \le \lambda_2 - \lambda_1,$$

μ-almost surely.

Moreover, under hypotheses (H1)–(H3), *for μ-almost all $\underline{z}$ there exists $e \in \mathscr{B}$, $b, c \in \mathscr{A}$ (which may depend on $\underline{z}$) such that*

$$\limsup_{n\to+\infty} \frac{1}{n} \log\bigl|\tilde{\nu}_{\underline{z},b}^{[n]}(e) - \tilde{\nu}_{\underline{z},c}^{[n]}(e)\bigr| = \lambda_2 - \lambda_1.$$

Under hypotheses (H1)–(H3), *for μ-almost all $\underline{z}$ there exists $e, b, c \in \mathscr{B}$ (which may depend on $\underline{z}$) such that*

$$\limsup_{n\to+\infty} \frac{1}{n} \log\bigl|\tilde{\sigma}_{\underline{z},b}^{[n]}(e) - \tilde{\sigma}_{\underline{z},c}^{[n]}(e)\bigr| = \lambda_2 - \lambda_1.$$

From a practical point of view, one can prove various lower bounds for the quantity $\lambda_2 - \lambda_1$. As an example we give the following result.

Let

$$\Gamma = \frac{1}{\min_{m,i}\{q(m|i)\}}.$$

PROPOSITION 2.4. *Under hypotheses* (H1)–(H2) *we have*

$$\lambda_2 - \lambda_1 \geq \frac{1}{k-1} \log\bigl|\det(p)\bigr| - \frac{k}{k-1} \log \Gamma.$$

We now state a related result using the total variation distance between the distributions. This result will only use a weaker version of hypothesis (H3), namely (H3′), there exists $b, c \in \mathscr{B}$ in and $i \in \mathscr{A}$ such that

$$q(b|i) \neq q(c|i). \tag{2.2}$$

Note that under this hypothesis, we do not assume any relation between the cardinality of $\mathscr{A}$ and the cardinality of $\mathscr{B}$ (we require of course the cardinality of $\mathscr{B}$ being at least two).

We recall that the total variation distance $\mathrm{TV}(\nu_1, \nu_2)$ between two measures ν_1 and ν_2 on $\mathscr{A}$ is defined by

$$\mathrm{TV}(\nu_1, \nu_2) = \frac{1}{2} \sum_{a\in\mathscr{A}} \bigl|\nu_1(a) - \nu_2(a)\bigr|.$$

Similar definitions are given for two measures on $\mathscr{B}$.

It follows at once that under the hypothesis (H1) we have μ-almost surely

$$\limsup_{n\to+\infty} \frac{1}{n} \log \mathrm{TV}\bigl(\nu_{\underline{z},\rho}^{[n]}(a) - \nu_{\underline{z},\rho'}^{[n]}(a)\bigr) \leq \lambda_2 - \lambda_1,$$

and similarly for the measures σ, $\tilde{\nu}$ and $\tilde{\sigma}$. In order to state a lower bound for these quantities, we need to recall a result about Lyapunov dimensions.

We denote by s the number of different Lyapunov exponents, and by m_i ($1 \leq i \leq s$) the multiplicity of the exponent λ_i, namely

$$m_i = \dim(V_\omega^{(i)}) - \dim(V_\omega^{(i+1)}).$$

It follows from Oseledec's theorem that these numbers are μ-almost surely constant.

THEOREM 2.5. *Assume hypotheses* (H1)–(H2). *Then for μ-almost every $\underline{z}$ and for any pair (b, c) of elements in $\mathscr{B}$ satisfying* (H3′) *we have*

$$\begin{aligned}\liminf_{n\to\infty} \frac{1}{n} \log \mathrm{TV}(\sigma_{\underline{z},b}^{[n]}, \sigma_{\underline{z},c}^{[n]}) &\geq \lambda_s - \lambda_1 + \sum_{i=2}^{s} m_i(\lambda_i - \lambda_2)\\ &\geq 2(\log|\det p| - k \log \Gamma).\end{aligned}$$

REMARK. Similar lower bounds for the total variation distance between the measures $\nu_{\underline{z},b}^{[n]}$ and $\nu_{\underline{z},c}^{[n]}$ can be proven under hypotheses (H1)–(H2). In the case of the total variation distance between $\tilde{\sigma}_{\underline{z},b}^{[n]}$ and $\tilde{\sigma}_{\underline{z},c}^{[n]}$ (resp., between $\tilde{\nu}_{\underline{z},b}^{[n]}$ and $\tilde{\nu}_{\underline{z},c}^{[n]}$) we can prove also the same lower bounds, but this requires the full set of hypotheses (H1)–(H3).

3. Proofs. We begin by proving some lemmas which will be useful later. We introduce the order $(\mathbb{R}^k, \leq)$ given by $\vec{v} \leq \vec{w}$ if and only if $v_i \leq w_i$ for all $i = 1, \dots, k$. When needed, we will also make use of the symbols $<$, $>$ and $\geq$, defined in an analogous way. Note that since the matrices $L_{\underline{z}}$ have strictly positive entries, if $\vec{v} \leq \vec{w}$, then $L_{\underline{z}}\vec{v} \leq L_{\underline{z}}\vec{w}$. We will use the notation $\vec{1} \in \mathbb{R}^k$ for the vector with components $(\vec{1})_i = 1$ for each $i = 1, \dots, k$ and the notation $\vec{1}_a \in \mathbb{R}^k$ for the vector with components $(\vec{1}_a)_a = 1$ and $(\vec{1}_a)_i = 0$ for $i \neq a$.

LEMMA 3.1. *Under hypothesis* (H1), *if $\vec{\xi} \in V_{\underline{z}}^{(2)} \setminus \{\vec{0}\}$, then $\vec{\xi}$ has two nonzero components of opposite signs, μ-almost surely.*

PROOF. Assume there exits $\vec{\xi} \in V_{\underline{z}}^{(2)} \setminus \{\vec{0}\}$ with $\vec{\xi}_i \geq 0$ for all $i = 1, \dots, k$. Then, from hypothesis (H1) it follows that there exists $\alpha > 0$ such that, for all $\underline{z}$,

$$L_{\underline{z}}\vec{\xi} \geq \alpha \|\vec{\xi}\| \vec{1}.$$

One may take, for example,

$$\alpha = \frac{1}{\sqrt{k}} \inf_{z_0,i,j} q(z_0|i) p(j|i) = \frac{1}{\sqrt{k}} \inf_{\underline{z},i,j} (L_{\underline{z}})_{i,j}.$$

We can apply $L_{\mathscr{S}^{-1}\underline{z}}^{[n-1]}$ to both sides, use monotonicity and take norms, to obtain

$$\|L_{\underline{z}}^{[n]}\vec{\xi}\| \geq \alpha \|\vec{\xi}\| \|L_{\mathscr{S}^{-1}\underline{z}}^{[n-1]}\vec{1}\|.$$

Let $\vec{w} \in V^{(1)}_{\mathscr{S}^{-1}\underline{z}} \setminus V^{(2)}_{\mathscr{S}^{-1}\underline{z}}$. Then

$$\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{w}\| \le \|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}|\vec{w}|\| \le \|\vec{w}\|\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{1}\| \le \frac{\|\vec{w}\|}{\alpha\|\vec{\xi}\|}\|L^{[n]}_{\underline{z}}\vec{\xi}\|.$$

Therefore

$$\|L^{[n]}_{\underline{z}}\vec{\xi}\| \ge \frac{\alpha\|\vec{\xi}\|}{\|\vec{w}\|}\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{w}\|,$$

and using Oseledec's theorem we have μ-almost surely that

$$\lim_{n\to+\infty}\frac{1}{n}\log\|L^{[n]}_{\underline{z}}\vec{\xi}\| \ge \lim_{n\to+\infty}\frac{1}{n}\log\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{w}\| = \lambda_1,$$

which contradicts the fact that $\vec{\xi} \in V^{(2)}_{\underline{z}} \setminus \{\vec{0}\}$. $\square$

LEMMA 3.2. *Under hypothesis* (H1) *we have* $\operatorname{Codim}(V^{(2)}_{\underline{z}}) = 1$, *$\mu$-almost surely.*

PROOF. Assume $\operatorname{Codim}(V^{(2)}_{\underline{z}}) \ge 2$. Since any vector $\vec{w}_1$ of norm one in the cone $\mathscr{C}_k = \{\vec{w} : \vec{w} > 0\}$ does not belong to $V^{(2)}_{\underline{z}}$ (by Lemma 3.1), the vector space $V^{(2)}_{\underline{z}} \oplus \mathbb{R}\vec{w}_1$ is of codimension at least one, μ-almost surely. Therefore we can find a vector $\vec{w}_2$ of norm one in $\mathscr{C}_k \setminus (V^{(2)}_{\underline{z}} \oplus \mathbb{R}\vec{w}_1)$. Note that

$$\inf_{\vec{y}\in V^{(2)}_{\underline{z}},\gamma} \|\vec{w}_1 - \gamma\vec{w}_2 - \vec{y}\| > 0 \tag{3.1}$$

since otherwise, the minimum is reached at a finite nonzero pair $(\gamma, \vec{y})$ which would contradict $\vec{w}_2 \in \mathscr{C}_k \setminus (V^{(2)}_{\underline{z}} \oplus \mathbb{R}\vec{w}_1)$. Let $\underline{z}$ be a fixed element in $\mathscr{B}^{\mathbb{Z}}$.

Define

$$\gamma_n = \max_i \frac{(L^{[n]}_{\underline{z}}\vec{w}_1)_i}{(L^{[n]}_{\underline{z}}\vec{w}_2)_i} \quad \text{and} \quad \delta_n = \min_i \frac{(L^{[n]}_{\underline{z}}\vec{w}_1)_i}{(L^{[n]}_{\underline{z}}\vec{w}_2)_i}.$$

Let

$$\phi = \inf_{\underline{z}} \min_{i,j,r,s} \frac{(L_{\underline{z}})_{r,j}(L_{\underline{z}})_{s,i}}{(L_{\underline{z}})_{s,j}(L_{\underline{z}})_{r,i}}.$$

It follows from hypothesis (H1) that $\phi > 0$. Let

$$\alpha = \frac{1-\sqrt{\phi}}{1+\sqrt{\phi}} < 1.$$

From the Birkhoff–Hopf theorem [see, e.g., Cavazos-Cadena (2003)], there exists a constant $\beta > 0$ such that for all $\underline{z} \in \mathscr{B}^{\mathbb{Z}}$ and all n,

$$1 \le \frac{\gamma_n}{\delta_n} \le 1 + \beta\alpha^n. \tag{3.2}$$

We now prove that

$$\frac{\gamma_n}{1+\beta\alpha^n} \le \delta_{n+1} \le \gamma_{n+1} \le \gamma_n.$$

To see this observe that $\delta_{n+1} \le \gamma_{n+1}$ by definition. We also have by monotonicity of $L_{\underline{z}}$

$$\begin{aligned}\gamma_{n+1} &= \max_i \frac{(L_{\underline{z}}^{[n+1]}\vec{w}_1)_i}{(L_{\underline{z}}^{[n+1]}\vec{w}_2)_i} = \max_i \frac{(L_{\mathscr{S}^{-n}\underline{z}} L_{\underline{z}}^{[n]}\vec{w}_1)_i}{(L_{\underline{z}}^{[n+1]}\vec{w}_2)_i} \\ &\le \max_i \frac{(L_{\mathscr{S}^{-n}\underline{z}}\gamma_n L_{\underline{z}}^{[n]}\vec{w}_2)_i}{(L_{\underline{z}}^{[n+1]}\vec{w}_2)_i} = \gamma_n\end{aligned}$$

and also

$$\delta_{n+1} \ge \delta_n = \gamma_n \frac{\delta_n}{\gamma_n} \ge \frac{\gamma_n}{1+\beta\alpha^n}.$$

Since the sequence (γ_n) is decreasing, there exists γ^* and $\beta' > 0$ such that

$$|\gamma_n - \gamma^*| \le \beta'\alpha^n.$$

On the other hand, it follows immediately from (3.2) that for any $i = 1, \dots, k$, we have

$$-\frac{\gamma_n \beta\alpha^n (L_{\underline{z}}^{[n]}\vec{w}_2)_i}{1+\beta\alpha^n} \le (L_{\underline{z}}^{[n]}\vec{w}_1)_i - \gamma_n (L_{\underline{z}}^{[n]}\vec{w}_2)_i \le 0.$$

Then there exists $\beta'' > 0$ such that

$$\frac{\|L_{\underline{z}}^{[n]}\vec{w}_1 - \gamma_n L_{\underline{z}}^{[n]}\vec{w}_2\|}{\|L_{\underline{z}}^{[n]}\vec{w}_2\|} \le \beta''\alpha^n.$$

This implies

$$\begin{aligned}\frac{\|L_{\underline{z}}^{[n]}\vec{w}_1 - \gamma^* L_{\underline{z}}^{[n]}\vec{w}_2\|}{\|L_{\underline{z}}^{[n]}\vec{w}_2\|} &\le \frac{|\gamma_n - \gamma^*|\,\|L_{\underline{z}}^{[n]}\vec{w}_2\|}{\|L_{\underline{z}}^{[n]}\vec{w}_2\|} + \frac{\|L_{\underline{z}}^{[n]}\vec{w}_1 - \gamma_n L_{\underline{z}}^{[n]}\vec{w}_2\|}{\|L_{\underline{z}}^{[n]}\vec{w}_2\|} \\ &\le (\beta' + \beta'')\alpha^n.\end{aligned}$$

Since $\vec{w}_1$ and $\vec{w}_2$ are linearly independent, we have $\vec{w}_1 - \gamma^*\vec{w}_2 \ne \vec{0}$. This and the previous inequality imply that

$$\lim_{n\to+\infty} \frac{1}{n} \log\|L_{\underline{z}}^{[n]}(\vec{w}_1 - \gamma^*\vec{w}_2)\| \le \lambda_1 + \log\alpha < \lambda_1,$$

then $\vec{w}_1 - \gamma^*\vec{w}_2 \in V_{\underline{z}}^{(2)} \setminus \{0\}$, and this contradicts (3.1). □

The proof of Theorem 2.1 will follow from the next proposition.

PROPOSITION 3.3. *Let $(\vec{\psi}_a)_{a\in\mathscr{A}}$ be a basis of $\mathbb{R}^k$ satisfying $\vec{\psi}_a \geq 0$ for any a. Let $\vec{\theta}_1 > 0$ and $\vec{\theta}_2 > 0$ be two vectors in $\mathbb{R}^k$. Then*

$$\limsup_{n\to\infty} \frac{1}{n} \log\left| \frac{\langle\vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}_a\rangle}{\langle\vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\sum_a\vec{\psi}_a\rangle} - \frac{\langle\vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}_a\rangle}{\langle\vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\sum_a\vec{\psi}_a\rangle}\right| \leq \lambda_2 - \lambda_1.$$

PROOF. By Lemma 3.1, since $(\vec{\psi}_a)_i \geq 0$ for any $i = 1,\dots,k$ then $\vec{\psi}_a \notin V^{(2)}_{\underline{z}}$, μ-almost surely. In the same way, from Lemma 3.1 we have

$$\vec{\psi} = \sum_{a\in\mathscr{A}} \vec{\psi}_a \in V^{(1)}_{\underline{z}} \setminus V^{(2)}_{\underline{z}}.$$

Note also that since $(\vec{\psi}_a)_{a\in\mathscr{A}}$ form a basis of nonnegative vectors, we must have $\vec{\psi}_i > 0$ for all $i = 1,\dots,k$. Therefore, by Lemma 3.2 we have that for any $a \in \mathscr{A}$,

$$\vec{\psi}_a = u_a\vec{\psi} + \vec{\xi}_a, \tag{3.3}$$

where $\vec{\xi}_a \in V^{(2)}_{\underline{z}}$, $u_a \neq 0$, and this decomposition is unique. Then

$$\frac{\langle\vec{\theta}_j, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}_a\rangle}{\langle\vec{\theta}_j, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\sum_a\vec{\psi}_a\rangle} = u_a + \frac{\langle\vec{\theta}_j, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\xi}_a\rangle}{\langle\vec{\theta}_j, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle}, \qquad j = 1, 2.$$

Define for any n and $\underline{z}$

$$\gamma(n,\underline{z}) = \frac{\langle\vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle}{\langle\vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle} \tag{3.4}$$

and let

$$R = \max\left\{\sup_i \frac{(\vec{\theta}_1)_i}{(\vec{\theta}_2)_i}, \sup_i \frac{(\vec{\theta}_2)_i}{(\vec{\theta}_1)_i}\right\}.$$

Then we have

$$\begin{aligned}\langle\vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle &= \sum_{i=1}^k (\vec{\theta}_1)_i \big(L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\big)_i \leq R\sum_{i=1}^k (\vec{\theta}_2)_i\big(L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\big)_i \\ &= R\langle\vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle\end{aligned}$$

and similarly

$$\langle\vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle \leq R\langle\vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle.$$

In other words for any n and $\underline{z}$,

$$R^{-1} \leq \gamma(n,\underline{z}) \leq R. \tag{3.5}$$

Then

$$\frac{\langle \vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\psi}_a \rangle}{\langle \vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \sum_a \vec{\psi}_a \rangle} - \frac{\langle \vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\psi}_a \rangle}{\langle \vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \sum_a \vec{\psi}_a \rangle}$$
$$= \big(\langle \vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\psi} \rangle\big)^{-1} \langle \vec{\theta}_1 - \gamma(n, \underline{z}) \vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\xi}_a \rangle.$$

Note that

$$\big|\langle \vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\psi} \rangle\big| \geq \frac{1}{\sqrt{k}} \| L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\psi} \| \inf_i \{(\vec{\theta}_1)_i\}$$

and

$$\big|\langle \vec{\theta}_1 - \gamma(n, \underline{z}) \vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\xi}_a \rangle\big| \leq \|\vec{\theta}_1 - \gamma(n, \underline{z}) \vec{\theta}_2\| \cdot \| L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\xi}_a \|$$
$$\leq \big(\|\vec{\theta}_1\| + R\|\vec{\theta}_2\|\big) \| L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\xi}_a \|.$$

Then, using Oseledec's theorem the result follows. □

PROOF OF THEOREM 2.1. We observe that

$$\sum_{b \in \mathscr{A}} \mathbb{P}\big(X_0 = a, Z_{-n+1}^{-1} = z_{-n+1}^{-1}, X_{-n} = b\big) \rho(b)$$
$$= \sum_{b \in \mathscr{A}} \sum_{x_{-n+1}^{-1} \in \mathscr{A}^{n-1}} p(x_{-n+1}|b) \rho(b) q(z_{-1}|x_{-1}) p(a|x_{-1})$$
$$\times \prod_{l=1}^{n-2} q(z_{-l-1}|x_{-l-1}) p(x_{-l}|x_{-l-1})$$
$$= \langle \vec{\theta}_\rho, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\psi}_a \rangle,$$

where $\vec{\theta}_\rho, \vec{\psi}_a \in \mathbb{R}^k$ are given by

$$(\vec{\theta}_\rho)_i = \sum_{b \in \mathscr{A}} \rho(b) p(i|b) \quad \text{and} \quad \vec{\psi}_a = \vec{1}_a. \tag{3.6}$$

Therefore,

$$\nu^{[n]}_{\underline{z},\rho}(a) = \frac{\langle \vec{\theta}_\rho, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\psi}_a \rangle}{\langle \vec{\theta}_\rho, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \sum_a \vec{\psi}_a \rangle},$$

and the first statement of Theorem 2.1 follows from Proposition 3.3 since the conditions on $(\psi_a)_{a \in \mathscr{A}}$, $\vec{\theta}_\rho$ and $\vec{\theta}_{\rho'}$ can be immediately verified using hypothesis (H1).

The second part follows similarly by noting that

$$\sum_{c \in \mathscr{B}} \mathbb{P}\big(X_0 = a, Z_{-n+1}^{-1} = z_{-n+1}^{-1}, Z_{-n} = c\big) \eta(c) = \langle \vec{\theta}_\eta, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\psi}_a \rangle,$$

where

$$(\vec{\theta}_\eta)_i = \sum_{c\in\mathscr{B}} \sum_{x\in\mathscr{A}} p(i|x)q(c|x)\eta(c) \quad \text{and} \quad \vec{\psi}_a = \vec{1}_a. \tag{3.7}$$

This finishes the proof of Theorem 2.1. □

Before proceeding with the proof of Theorem 2.2 we will prove a useful lemma.

LEMMA 3.4. *Let $(\vec{\psi}_a)_{a\in\mathscr{A}}$ be a basis of $\mathbb{R}^k$ such that $\vec{\psi}_a \geq 0$ for all a. Let $\tilde{\Omega}$ be a set of full μ-measure where the Oseledec's theorem holds. Then for any $\underline{z} \in \bar{\Omega}$, there exists a symbol $a = a(\underline{z}) \in \mathscr{A}$ such that $\vec{\xi}_a \in V_{\underline{z}}^{(2)} \setminus V_{\underline{z}}^{(3)}$, where $\vec{\xi}_a$ is the unique vector in $V_{\underline{z}}^{(2)}$ satisfying*

$$\vec{\psi}_a = u_a \sum_{b\in\mathscr{A}} \vec{\psi}_b + \vec{\xi}_a$$

for some real number u_a.

PROOF. Assume $\vec{\xi}_a \in V_{\underline{z}}^{(3)}$ for all a. Then, as $\operatorname{Codim}(V_{\underline{z}}^{(3)}) \geq 2$, the set $\{\vec{\psi}_a\}$ generates a sub-space of co-dimension 1. This contradicts the fact that the set of vectors $\{\vec{\psi}_a : a \in \mathscr{A}\}$ forms a basis of $\mathbb{R}^k$. □

The proof of Theorem 2.2 will follow from the next proposition.

PROPOSITION 3.5. *Let $(\vec{\psi}_a)_{a\in\mathscr{A}}$ be a basis of $\mathbb{R}^k$ satisfying $\vec{\psi}_a \geq 0$ for any a. Let $(\vec{\theta}_j)_{j\in\mathscr{A}}$ be another basis of $\mathbb{R}^k$ such that $\vec{\theta}_j > 0$. Then for μ-almost every $\underline{z}$ there exist $a \in \mathscr{A}$ and two indices $r, s \in \{1, \dots, k\}$ such that*

$$\limsup_{n\to\infty} \frac{1}{n} \log\left| \frac{\langle \vec{\theta}_r, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\psi}_a\rangle}{\langle \vec{\theta}_r, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \sum_a \vec{\psi}_a\rangle} - \frac{\langle \vec{\theta}_s, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\psi}_a\rangle}{\langle \vec{\theta}_s, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \sum_a \vec{\psi}_a\rangle} \right| \geq \lambda_2 - \lambda_1. \tag{3.8}$$

PROOF. Let $\tilde{\Omega}$ be a set of full μ-measure where the Oseledec's theorem holds. Applying Lemma 3.4, for any $\underline{z} \in \tilde{\Omega}$ we find a symbol $a = a(\underline{z}) \in \mathscr{A}$ such that

$$\vec{\psi}_a = u_a \sum_{b\in\mathscr{A}} \vec{\psi}_b + \vec{\xi}_a$$

with $\vec{\xi}_a \in V_{\underline{z}}^{(2)} \setminus V_{\underline{z}}^{(3)}$. Let

$$\tilde{\xi}_a(n, \underline{z}) = \frac{L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\xi}_a}{\| L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\xi}_a \|} \in V^{(2)}_{\mathscr{S}^{-n}\underline{z}}. \tag{3.9}$$

We now show that there exist r and s such that

$$\limsup_{n\to\infty} \left|\langle\tilde\theta_r(n,\underline{z}) - \tilde\theta_s(n,\underline{z}), \tilde\xi_a(n,\underline{z})\rangle\right| > 0,$$

where the vectors $\tilde\theta_j(n,\underline{z})$ are defined by

$$\tilde\theta_j(n,\underline{z}) = \frac{\langle\vec\theta_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \sum_a \vec\psi_a\rangle}{\langle\vec\theta_j, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \sum_a \vec\psi_a\rangle}\vec\theta_j.$$

Assume this is not the case, namely that for any r and s,

$$\lim_{n\to\infty} \left|\langle\tilde\theta_r(n,\underline{z}) - \tilde\theta_s(n,\underline{z}), \tilde\xi_a(n,\underline{z})\rangle\right| = 0. \tag{3.10}$$

Choose for any n (and fixed $\underline{z}$) a normalized vector $\vec f(n,\underline{z})$ orthogonal to $V^{(2)}_{\mathscr{S}^{-n}\underline{z}}$. Such a vector exists by Lemma 3.2. Note that for any j, n and $\underline{z}$, we have

$$0 < R^{-1}\min_m \|\vec\theta_m\| \le \|\tilde\theta_j(n,\underline{z})\| \le R\max_m\|\vec\theta_m\|,$$

where

$$R = \sup_{j,m}\sup_i \frac{(\vec\theta_j)_i}{(\vec\theta_m)_i}.$$

This implies that the vectors $(\vec f(n,\underline{z}), \tilde\xi_a(n,\underline{z}), \tilde\theta_1(n,\underline{z}), \dots, \tilde\theta_k(n,\underline{z}))$ belong to a compact subset of $\mathbb{R}^{k+2}$. Therefore, we can find a subsequence (n_j) of integers such that

$$\lim_{j\to\infty}\big(\vec f(n_j,\underline{z}), \tilde\xi_a(n_j,\underline{z}), \tilde\theta_1(n_j,\underline{z}), \dots, \tilde\theta_k(n_j,\underline{z})\big)$$
$$= \big(\bar f(\underline{z}), \bar\xi_a(\underline{z}), \bar\theta_1(\underline{z}), \dots, \bar\theta_k(\underline{z})\big).$$

The vectors $\bar f(\underline{z})$ and $\bar\xi_a(\underline{z})$ have norm one, and the vectors $\bar\theta_j(\underline{z})$ have nonnegative components and satisfy

$$0 < R^{-1}\min_m\|\vec\theta_m\| \le \|\bar\theta_j(\underline{z})\| \le R\max_m\|\vec\theta_m\|.$$

We have also for any r and s

$$\langle\bar\theta_r(\underline{z}) - \bar\theta_s(\underline{z}), \bar\xi_a(\underline{z})\rangle = 0.$$

We now show that the set of vectors $\{\bar\theta_m(\underline{z})\}$ is a basis of $\mathbb{R}^k$. This follows from

$$\left|\det\big(\bar\theta_1(\underline{z}), \dots, \bar\theta_k(\underline{z})\big)\right| = \left|\det(\vec\theta_1, \dots, \vec\theta_k)\right| \lim_{j\to\infty}\prod_{m=1}^k\left|\frac{\langle\vec\theta_1, L^{[n_j-1]}_{\mathscr{S}^{-1}\underline{z}}\sum_a\vec\psi_a\rangle}{\langle\vec\theta_m, L^{[n_j-1]}_{\mathscr{S}^{-1}\underline{z}}\sum_a\vec\psi_a\rangle}\right|$$
$$\ge R^{-k}\left|\det(\vec\theta_1, \dots, \vec\theta_k)\right| > 0.$$

Let

$$\zeta(n,\underline{z})=\frac{1}{k}\sum_{m=1}^{k}\tilde{\theta}_m(n,\underline{z})$$

and

$$\bar{\zeta}(\underline{z})=\lim_{j\to\infty}\zeta(n_j,\underline{z})=\frac{1}{k}\sum_{m=1}^{k}\bar{\theta}_m(\underline{z}).$$

We now observe that since all the components of the vector $\zeta(n,\underline{z})$ are strictly positive, and since by Lemma 3.1 any vector in $V^{(2)}_{\mathscr{S}^{-n}\underline{z}}$ has two components of opposite sign, we get

$$\begin{aligned}\frac{|\langle\vec{f}(n,\underline{z}),\zeta(n,\underline{z})\rangle|}{\|\zeta(n,\underline{z})\|}&=\inf_{\vec{y}\in V^{(2)}_{\mathscr{S}^{-n}\underline{z}}}\left\|\frac{\zeta(n,\underline{z})}{\|\zeta(n,\underline{z})\|}-\vec{y}\right\|\\&\geq\min_i\left\{\frac{(\zeta(n,\underline{z}))_i}{\|\zeta(n,\underline{z})\|}\right\}\geq\frac{1}{kR^2}\frac{\min_{m,i}(\vec{\theta}_m)_i}{\max_{m,i}(\vec{\theta}_m)_i}>0.\end{aligned}$$

Taking the limit we get

$$\frac{|\langle\bar{f}(\underline{z}),\bar{\zeta}(\underline{z})\rangle|}{\|\bar{\zeta}(\underline{z})\|}\geq\frac{1}{kR^2}\frac{\min_{m,i}(\vec{\theta}_m)_i}{\max_{m,i}(\vec{\theta}_m)_i}>0.$$

We now define the orthogonal projection $\mathcal{P}$ on the orthogonal $\bar{f}^{\perp}$ of $\bar{f}$ parallel to $\bar{\zeta}$, namely for any vector v

$$\mathcal{P}v=v-\bar{\zeta}\frac{\langle\bar{f},v\rangle}{\langle\bar{f},\bar{\zeta}\rangle}.$$

We claim that the vectors $(\mathcal{P}(\bar{\theta}_m(\underline{z})-\bar{\theta}_{m+1}(\underline{z})))_{m=1,\dots,k-1}$ form a basis of $\bar{f}^{\perp}$. Indeed, if this is not true, there exist real numbers $\alpha_1,\dots,\alpha_{k-1}$, with at least one nonzero, such that

$$\sum_{m=1}^{k-1}\alpha_m\mathcal{P}\big(\bar{\theta}_m(\underline{z})-\bar{\theta}_{m+1}(\underline{z})\big)=0.$$

In other words, there exists a number α such that

$$\sum_{m=1}^{k-1}\alpha_m\big(\bar{\theta}_m(\underline{z})-\bar{\theta}_{m+1}(\underline{z})\big)=\alpha\bar{\zeta}.$$

But this is impossible since the vectors $(\bar{\theta}_m(\underline{z})-\bar{\theta}_{m+1}(\underline{z}))_{m=1,\dots,k-1}$ and $\bar{\zeta}(\underline{z})$ form a basis of $\mathbb{R}^k$. Since

$$\langle\bar{f}(\underline{z}),\bar{\xi}_a(\underline{z})\rangle=\lim_{j\to\infty}\langle f(n_j,\underline{z}),\tilde{\xi}_a(n_j,\underline{z})\rangle=0,$$

we obtain that the normalized vector $\bar{\xi}_a(\underline{z})$ would be orthogonal to the basis $(\mathcal{P}(\bar{\theta}_m(\underline{z}) - \bar{\theta}_{m+1}(\underline{z})))_{m=1,\dots,k-1}$ of $\bar{f}^{\perp}$ which is a contradiction with (3.10). In other words, there exists $a = a(\underline{z})$, $r = r(\underline{z})$ and $s = s(\underline{z})$ such that

$$\limsup_{n\to\infty} \left|\langle \tilde{\theta}_r(n,\underline{z}) - \tilde{\theta}_s(n,\underline{z}), \tilde{\xi}_a(n,\underline{z})\rangle\right| > 0.$$

By Schwarz's inequality we have

$$\begin{aligned}
&\left|\frac{\langle \vec{\theta}_r, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\psi}_a\rangle}{\langle \vec{\theta}_r, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \sum_a \vec{\psi}_a\rangle} - \frac{\langle \vec{\theta}_s, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\psi}_a\rangle}{\langle \vec{\theta}_s, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \sum_a \vec{\psi}_a\rangle}\right| \\
&\quad = \frac{\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\xi}_a\|}{|\langle \vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \sum_a \vec{\psi}_a\rangle|} \left|\langle \tilde{\theta}_r(n,\underline{z}) - \tilde{\theta}_s(n,\underline{z}), \tilde{\xi}_a(n,\underline{z})\rangle\right| \\
&\quad \geq \frac{\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\xi}_a\|}{\|\vec{\theta}_1\| \|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \sum_a \vec{\psi}_a\|} \left|\langle \tilde{\theta}_r(n,\underline{z}) - \tilde{\theta}_s(n,\underline{z}), \tilde{\xi}_a(n,\underline{z})\rangle\right|.
\end{aligned}$$

Therefore, for this choice of $a(\underline{z}) \in \mathscr{A}$, $r(\underline{z})$ and $s(\underline{z})$, we have

$$\limsup_{n\to\infty} \frac{1}{n} \log\left|\frac{\langle \vec{\theta}_r, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\psi}_a\rangle}{\langle \vec{\theta}_r, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \sum_a \vec{\psi}_a\rangle} - \frac{\langle \vec{\theta}_s, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{\psi}_a\rangle}{\langle \vec{\theta}_s, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \sum_a \vec{\psi}_a\rangle}\right| \geq \lambda_2 - \lambda_1. \qquad \square$$

PROOF OF THEOREM 2.2. As in the proof of Theorem 2.1 we take for any $a \in \mathscr{A}$ the vector $\vec{\psi}_a = \vec{1}_a$. We also take for any $b \in \mathscr{B}$ the vector $\vec{\theta}_b$ in $\mathbb{R}^k$ as the vector $\vec{\theta}_\rho$ in (3.6) with ρ the Dirac measure concentrated on b, that is, $(\vec{\theta}_b)_i = p(i|b)$. Under (H2), these definitions verify the hypotheses of Proposition 3.5 and the first part of Theorem 2.2 follows.

For the second part, we define for any $c \in \mathscr{B}$ the vector $(\vec{\theta}_c)$ as the vector $(\vec{\theta}_\eta)$ in (3.7) with η the Dirac measure concentrated on c, that is,

$$(\vec{\theta}_c)_i = \sum_{x\in\mathscr{A}} p(i|x) q(c|x).$$

It follows from (H2) and (H3) that we can choose $c_1, \dots, c_k$ such that $(\vec{\theta}_{c_j})_{1\leq j\leq k}$ is a basis of $\mathbb{R}^k$. The result follows again from Proposition 3.5. $\square$

PROOF OF COROLLARY 2.3. The upper bound follows by noting that for all $\underline{z} \in \mathscr{B}^{\mathbb{Z}}$, for any $e \in \mathscr{B}$ and for any measures ρ and ρ' on $\mathscr{A}$, we have

$$\tilde{\nu}^{[n]}_{\underline{z},\rho}(e) - \tilde{\nu}^{[n]}_{\underline{z},\rho'}(e) = \sum_{x_0\in\mathscr{A}} q(e|x_0)\big(\nu^{[n]}_{\underline{z},\rho}(x_0) - \nu^{[n]}_{\underline{z},\rho'}(x_0)\big), \tag{3.11}$$

and applying Theorem 2.1, the second upper bound follows similarly.

We now prove that the upper bound is reached for almost all $\underline{z} \in \mathscr{B}^{\mathbb{Z}}$.

By (H3), as $\mathrm{rank}(q) = k$ there exists symbols $e_1, \dots, e_k \in \mathscr{B}$ such that the matrix $M \in \mathbb{R}^{k\times k}$ with elements $M_{i,j} = q(e_i|j)$ is invertible. For $b, c \in \mathscr{A}$, denote by $U^{[n]}_{b,c,\underline{z}}$ and $V^{[n]}_{b,c,\underline{z}}$ the vectors in $\mathbb{R}^k$ with elements $(U^{[n]}_{b,c,\underline{z}})_i = \tilde{\nu}^{[n]}_{\underline{z},b}(e_i) - \tilde{\nu}^{[n]}_{\underline{z},c}(e_i)$ and $(V^{[n]}_{b,c,\underline{z}})_i = \nu^{[n]}_{\underline{z},b}(i) - \nu^{[n]}_{\underline{z},c}(i)$. By (3.11) we have

$$U^{[n]}_{b,c,\underline{z}} = M V^{[n]}_{b,c,\underline{z}}$$

and as M is invertible

$$V^{[n]}_{b,c,\underline{z}} = M^{-1} U^{[n]}_{b,c,\underline{z}}.$$

Then, for all $a, b, c \in \mathscr{A}$

$$\begin{aligned}\left|\nu^{[n]}_{\underline{z},b}(a) - \nu^{[n]}_{\underline{z},c}(a)\right| &\le \|V^{[n]}_{b,c,\underline{z}}\| \le \|M^{-1}\|\|U^{[n]}_{b,c,\underline{z}}\| \\ &\le \sqrt{k}\|M^{-1}\| \max_i\{\left|\tilde{\nu}^{[n]}_{\underline{z},b}(e_i) - \tilde{\nu}^{[n]}_{\underline{z},c}(e_i)\right|\}.\end{aligned}$$

Applying the logarithm on both sides, dividing by n an taking limits, we have that for all $\underline{z}$ on a set of positive measure, for all $a, b, c \in \mathscr{A}$, and for all $e \in \mathscr{B}$

$$\limsup_{n\to\infty} \frac{1}{n} \log\left|\nu^{[n]}_{\underline{z},b}(a) - \nu^{[n]}_{\underline{z},c}(a)\right| \le \max_{e\in\mathscr{B}} \limsup_{n\to\infty} \frac{1}{n} \log\left|\tilde{\nu}^{[n]}_{\underline{z},b}(e) - \tilde{\nu}^{[n]}_{\underline{z},c}(e)\right|,$$

and the third part of Corollary 2.3 follows from Theorem 2.2. The last part follows by the same arguments. □

PROOF OF PROPOSITION 2.4. It is well known that the sequence of Lyapunov exponents satisfy

$$\lambda_1 + m_2\lambda_2 + \cdots + m_s\lambda_s = \mathbb{E}_\mu\left[\log|\det L_{\underline{z}}|\right],$$

where the numbers m_i denote the multiplicity of λ_i, namely $\dim(V^{(j)}_{\underline{z}}) = m_j + \cdots + m_s$; see Katok and Hasselblatt (1995), Ledrappier (1984). In particular, $1 + m_2 + \cdots + m_s = k$. Let $E = \mathbb{E}_\mu[\log|\det L_{\underline{z}}|]$. Then we have

$$E \le \lambda_1 + (k-1)\lambda_2$$

and

$$\lambda_2 - \lambda_1 \ge \frac{E}{k-1} - \frac{k}{k-1}\lambda_1.$$

Note that by Lemma 3.1, for almost all $\underline{z}$ we have

$$\lambda_1 = \lim_{n\to\infty} \frac{1}{n} \log\left\|L^{[n]}_{\underline{z}}\vec{1}\right\| \le \lim_{n\to\infty} \frac{1}{n} \log\|\vec{1}\| = 0.$$

Moreover,

$$\det L_{\underline{z}} = \left(\prod_{i=1}^{k} q(z_0|i)\right)\det(p).$$

Therefore,

$$\begin{aligned}\lambda_2 - \lambda_1 &\geq \frac{1}{k-1}\log\bigl|\det(p)\bigr| + \frac{1}{k-1}\sum_{i=1}^{k}\mathbb{E}_\mu\bigl[\log q(\cdot|i)\bigr]\\ &\geq \frac{1}{k-1}\log\bigl|\det(p)\bigr| - \frac{k}{k-1}\log\Gamma.\end{aligned}$$

□

Before proving Theorem 2.5, we prove a lemma in linear algebra which will be useful for the proof. This lemma is probably well known but we could not find a reference. We give the proof here for the convenience of the reader.

LEMMA 3.6. *Let $e_1, \dots, e_k$ be a basis of $\mathbb{R}^k$ and assume that all the vectors $\|e_j\|$ have norm one. Then, for any vector $v \in \mathbb{R}^k$ of norm one, we have*

$$\sup_{1\le j\le k}\bigl|\langle v, e_j\rangle\bigr| \geq \frac{1}{k^{3/2}(k-1)!\det A},$$

where A is a matrix mapping the basis (e_i) to an orthonormal basis.

PROOF. Let

$$\delta = \sup_{1\le j\le k}\bigl|\langle v, e_j\rangle\bigr|.$$

Let (f_j) be an orthonormal basis of $\mathbb{R}^k$. Let A be the matrix mapping the basis (e_j) to the basis (f_j), namely $Ae_j = f_j$ for $1 \le j \le k$. We have

$$\langle v, e_j\rangle = \langle v, A^{-1} f_j\rangle = \langle A^{-1^{\mathrm{t}}} v, f_j\rangle.$$

Therefore,

$$\|A^{-1^{\mathrm{t}}} v\| \le \delta\sqrt{k}$$

and

$$1 = \|v\| = \|A^{\mathrm{t}} A^{-1^{\mathrm{t}}} v\| \le \|A^{\mathrm{t}}\|\delta\sqrt{k}.$$

On the other hand, since $(e_j)_\ell = A^{-1}_{j,\ell}$ we have $|A^{-1}_{j,\ell}| \le 1$ for $1 \le j \le k$ and $1 \le \ell \le k$. This implies by the well-known formula expressing the elements of an inverse matrix in terms of minors and determinant that for any $1 \le j \le k$ and $1 \le \ell \le k$

$$|A_{j,\ell}| \le \frac{(k-1)!}{|\det A^{-1}|} = (k-1)!\det A.$$

Therefore $\|A^{\mathrm{t}}\| = \|A\| \le k(k-1)!\det A$ (the Hilbert–Schmidt norm), and we finally get

$$\delta \ge \frac{1}{k^{3/2}(k-1)!|\det A|}. \qquad \square$$

Theorem 2.5 will be a consequence of the following proposition.

PROPOSITION 3.7. *Assume hypotheses* (H1)–(H2) *hold. Let* $(\vec{\psi}_a)_{a\in\mathscr{A}}$ *be a basis of* $\mathbb{R}^k$ *satisfying* $\vec{\psi}_a \ge 0$ *for any* a. *Let* $\vec{\theta}_1 > 0$ *and* $\vec{\theta}_2 > 0$ *be two vectors in* $\mathbb{R}^k$ *with* $\vec{\theta}_1$ *independent of* $\vec{\theta}_2$. *Then*

$$\liminf_{n\to\infty} \frac{1}{n} \log \sum_{a\in\mathscr{A}} \left| \frac{\langle\vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}_a\rangle}{\langle\vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\sum_a \vec{\psi}_a\rangle} - \frac{\langle\vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}_a\rangle}{\langle\vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\sum_a \vec{\psi}_a\rangle} \right|$$

$$\ge \lambda_s - \lambda_1 + \sum_{i=2}^{s} m_i(\lambda_i - \lambda_2) \ge 2\bigl(\log|\det p| - k\log\Gamma\bigr).$$

PROOF. As in the proof of Proposition 3.3 let

$$\vec{\psi} = \sum_{a\in\mathscr{A}} \vec{\psi}_a \quad \text{and} \quad \gamma(n,\underline{z}) = \frac{\langle\vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle}{\langle\vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle}.$$

We also define the vector $\vec{\eta}(n,\underline{z}) = \vec{\theta}_1 - \gamma(n,\underline{z})\vec{\theta}_2$ that satisfies

$$\langle\vec{\eta}(n,\underline{z}), L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle = 0.$$

For any $a \in \mathscr{A}$, we denote as before by $\vec{\xi}_a$ the unique vector in $V^{(2)}_{\underline{z}}$ such that $\vec{\psi}_a = u_a\vec{\psi} + \vec{\xi}_a$, with u_a a real number.

Let

$$\tilde{\xi}_a(n,\underline{z}) = \frac{L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\xi}_a}{\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\xi}_a\|} \in V^{(2)}_{\mathscr{S}^{-n}\underline{z}}. \tag{3.12}$$

Let $a_2, \dots, a_k$ be any given collection of $k-1$ different elements of $\mathscr{A}$. Then, for any $\underline{z} \in \Omega$, the set of vectors $\vec{\psi}, \vec{\xi}_{a_2}, \dots, \vec{\xi}_{a_k}$ form a basis of $\mathbb{R}^k$ (since $\vec{\psi} \notin V^{(2)}_{\underline{z}}$ by Lemma 3.1).

By the hypotheses (H1)–(H2) we have that for any $\underline{z}$, $\det(L_{\underline{z}}) \ne 0$, and therefore for any integer n, $\det(L^{[n]}_{\underline{z}}) \ne 0$. This implies that the collection of vectors

$$\left\{ \frac{L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}}{\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\|}, \tilde{\xi}_{a_2}(n,\underline{z}), \dots, \tilde{\xi}_{a_k}(n,\underline{z}) \right\} \tag{3.13}$$

is also a basis of $\mathbb{R}^k$ and

$$\begin{aligned}
&\sum_{j=2}^{k}\left|\frac{\langle\vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}_{a_j}\rangle}{\langle\vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle} - \frac{\langle\vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}_{a_j}\rangle}{\langle\vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle}\right| \\
&\qquad \geq \sum_{j=2}^{k}\frac{\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\xi}_{a_j}\|}{\|\vec{\theta}_1\|\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\|}\left|\langle\vec{\theta}_1 - \gamma(n,\underline{z})\vec{\theta}_2, \tilde{\xi}_{a_j}(n,\underline{z})\rangle\right| \\
&\qquad = \sum_{j=2}^{k}\frac{\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\xi}_{a_j}\|}{\|\vec{\theta}_1\|\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\|}\left|\langle\vec{\eta}(n,\underline{z}), \tilde{\xi}_{a_j}(n,\underline{z})\rangle\right|.
\end{aligned}$$

We now apply Lemma 3.6 and obtain

$$\begin{aligned}
&\sum_{j=2}^{k}\left|\frac{\langle\vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}_{a_j}\rangle}{\langle\vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle} - \frac{\langle\vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}_{a_j}\rangle}{\langle\vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle}\right| \\
&\qquad \geq \frac{\inf_{a\in\mathscr{A}}\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\xi}_a\|}{\|\vec{\theta}_1\|\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\|}\frac{\|\vec{\eta}(n,\underline{z})\|}{k^{3/2}(k-1)!|\det M|},
\end{aligned}$$

where M is the matrix formed by the vectors in (3.13). We now observe that

$$\|\vec{\eta}(n,\underline{z})\| = \|\vec{\theta}_1 - \gamma(n,\underline{z})\vec{\theta}_2\| \geq \inf_{\alpha}\|\vec{\theta}_1 - \alpha\vec{\theta}_2\| = \sqrt{\|\vec{\theta}_1\|^2 - \frac{\langle\vec{\theta}_1,\vec{\theta}_2\rangle^2}{\|\vec{\theta}_2\|^2}} > 0,$$

since $\vec{\theta}_1$ is independent of $\vec{\theta}_2$. We also have

$$\begin{aligned}
\det M &= \frac{\det(L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\xi}_{a_2}, \ldots, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\xi}_{a_k})}{\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\|\prod_{j=2}^{k}\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\xi}_{a_j}\|} \\
&= \frac{\det(L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}})}{\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\|\prod_{j=2}^{k}\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\xi}_{a_j}\|}\det(\vec{\psi}, \vec{\xi}_{a_2}, \ldots, \vec{\xi}_{a_k}).
\end{aligned}$$

We observe that, for any $a \in A$, by Oseledec's theorem we have μ-almost surely

$$\liminf_{n\to\infty}\frac{1}{n}\log\|L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\xi}_a\| \geq \lambda_s.$$

Therefore [see, e.g., Katok and Hasselblatt (1995), Ledrappier (1984)], since

$$\lim_{n\to\infty}\frac{1}{n}\log\left|\det\left(L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\right)\right| = \sum_{j=1}^{s} m_j\lambda_j,$$

we get

$$\begin{aligned}
&\liminf_{n\to\infty}\frac{1}{n}\log\sum_{a\in\mathscr{A}}\left|\frac{\langle\vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}_a\rangle}{\langle\vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle} - \frac{\langle\vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}_a\rangle}{\langle\vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle}\right| \\
&\quad\ge \liminf_{n\to\infty}\frac{1}{n}\log\sum_{j=2}^{k}\left|\frac{\langle\vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}_{a_j}\rangle}{\langle\vec{\theta}_1, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle} - \frac{\langle\vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}_{a_j}\rangle}{\langle\vec{\theta}_2, L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\vec{\psi}\rangle}\right| \\
&\quad\ge \lambda_s - \lambda_1 + \sum_{j=1}^{s} m_j\lambda_j - (k-1)\lambda_2 - \lambda_1 \\
&\quad= \lambda_s - \lambda_1 + \sum_{j=2}^{s} m_j(\lambda_j - \lambda_2),
\end{aligned}$$

which is the first part of the lower bound. We also have

$$\left|\det\left(L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}}\right)\right| = |\det p|^{n-1}\prod_{j=1}^{n}\left(\prod_{\ell=1}^{k} q(z_{-j}, l)\right) \ge |\det p|^{n-1}\Gamma^{-nk}.$$

Therefore

$$\sum_{j=1}^{s} m_j\lambda_j \ge \log|\det p| - k\log\Gamma.$$

Since all the Lyapunov exponents are nonpositive, we get

$$\begin{aligned}
\lambda_s - \lambda_1 + \sum_{j=2}^{s} m_j(\lambda_j - \lambda_2) &\ge \lambda_s + \sum_{j=2}^{s} m_j\lambda_j \\
&\ge 2\sum_{j=1}^{s} m_j\lambda_j \\
&\ge 2\log|\det p| - 2k\log\Gamma. \qquad \square
\end{aligned}$$

PROOF OF THEOREM 2.5. The result follows immediately from Proposition 3.7 using the same choices for the vectors $(\vec{\psi}_a)_{a\in\mathscr{A}}$ and $(\vec{\theta}_b)_{b\in\mathscr{B}}$ as in the proof of Theorem 2.2. $\square$

4. Perturbed processes over a binary alphabet. Consider the Markov chain $(X_t)_{t\in\mathbb{Z}}$ over the alphabet $\mathscr{A} = \{0, 1\}$ with matrix of transition probabilities given by

$$P = \begin{pmatrix} p_0 & 1-p_0 \\ p_1 & 1-p_1 \end{pmatrix},$$

where we assume $p_0 \neq p_1$ and

$$0 < \beta = \min\{p_0, p_1, 1 - p_0, 1 - p_1\}.$$

The quantities $p(j|i)$ are given by $p(j|i) = P_{i,j}$.

Consider also the process $(Z_t)_{t\in\mathbb{Z}}$ over the alphabet $\mathscr{B} = \{0, 1\}$ with output matrix $q_\varepsilon(j|i) = \mathbb{P}(Z_0 = j|X_0 = i) = (1-\varepsilon)\mathbb{1}_{\{i=j\}} + \varepsilon\mathbb{1}_{\{i\neq j\}}$. From now on we will assume $\varepsilon \in (0, 1) \setminus \{1/2\}$. Then, as $z_0 \in \{0, 1\}$,

$$L_{\underline{z},\varepsilon} = \begin{pmatrix} [z_0\varepsilon + (1-z_0)(1-\varepsilon)]p_0 & [z_0\varepsilon + (1-z_0)(1-\varepsilon)](1-p_0) \\ [z_0(1-\varepsilon) + (1-z_0)\varepsilon]p_1 & [z_0(1-\varepsilon) + (1-z_0)\varepsilon](1-p_1) \end{pmatrix}.$$

We have the following equality:

$$\lambda_1 + \lambda_2 = \mathbb{E}_\mu\bigl[\log|\det L_{\cdot,\varepsilon}|\bigr];$$

see, for example, Ledrappier (1984) or Katok and Hasselblatt (1995) for a proof. Therefore

$$\begin{aligned} \lambda_1 + \lambda_2 &= \mathbb{P}(Z_0 = 0)\log\bigl((1-\varepsilon)\varepsilon\bigl|p_0(1-p_1) - p_1(1-p_0)\bigr|\bigr) \\ &\quad + \mathbb{P}(Z_0 = 1)\log\bigl(\varepsilon(1-\varepsilon)\bigl|p_0(1-p_1) - p_1(1-p_0)\bigr|\bigr) \\ &= \log\varepsilon + \log(1-\varepsilon) + \log\bigl|\det(P)\bigr|. \end{aligned} \tag{4.1}$$

From the above expression for $L_{\underline{z},\varepsilon}$ we have

$$L_{\underline{z},\varepsilon} = M_{z_0} + \varepsilon A_{z_0},$$

where

$$M_{z_0} = \begin{pmatrix} (1-z_0)p_0 & (1-z_0)(1-p_0) \\ z_0 p_1 & z_0(1-p_1) \end{pmatrix}$$

and

$$A_{z_0} = (2z_0 - 1)\begin{pmatrix} p_0 & (1-p_0) \\ -p_1 & -(1-p_1) \end{pmatrix}.$$

For $z_0 \in \{0, 1\}$ define the vectors

$$\vec{e}_{z_0} = \begin{pmatrix} 1 - z_0 \\ z_0 \end{pmatrix} \quad \text{and} \quad \vec{f}_{z_0} = \begin{pmatrix} z_0 \\ 1 - z_0 \end{pmatrix}.$$

These vectors have norm 1 and satisfy

$$M_{z_0}\vec{e}_{z_1} = \rho_0(z_0, z_1)\vec{e}_{z_0} \quad \text{and} \quad M^t_{z_0}\vec{f}_{z_0} = \vec{0},$$

where

$$\rho_0(z_0, z_1) = (1-z_1)\bigl(p_0(1-z_0) + p_1 z_0\bigr) + z_1\bigl((1-p_0)(1-z_0) + (1-p_1)z_0\bigr),$$

since z_0 and z_1 equal zero or one.

We recall that a distance d can be defined on Ω as follows. For $\underline{z}$ and $\underline{z}'$ in Ω, let

$$\tilde{d}(\underline{z}, \underline{z}') = \inf\{|i|, z_i \neq z_i'\}.$$

Then

$$d(\underline{z}, \underline{z}') = e^{-\tilde{d}(\underline{z}, \underline{z}')}.$$

We refer to Bowen (2008) for details, in particular Ω equipped with this distance is a compact metric space. We now prove the following result.

LEMMA 4.1. *There exist two constants $\varepsilon_0 > 0$ and $D > 0$ and two continuous functions $\rho(\varepsilon, \underline{z})$ and $h(\varepsilon, \underline{z})$ such that for any $\varepsilon \in [0, \varepsilon_0]$, the vectors*

$$\vec{g}(\varepsilon, \underline{z}) = \vec{e}_{z_1} + \varepsilon h(\varepsilon, \underline{z}) \vec{f}_{z_1}$$

satisfy

$$L_{\underline{z},\varepsilon} \vec{g}(\varepsilon, \underline{z}) = \rho(\varepsilon, \underline{z}) \vec{g}(\varepsilon, \mathscr{S}^{-1}\underline{z}).$$

Moreover, there is a constant $U > 1$ such that for any $\varepsilon \in [0, \varepsilon_0]$, any n and any $\underline{z} \in \Omega$,

$$\|\vec{g}(\varepsilon, \underline{z}) - \vec{e}_{z_1}\| \leq U\varepsilon, \qquad |\rho(\varepsilon, \underline{z}) - \langle M_{z_1} \vec{e}_{z_2}, \vec{e}_{z_1} \rangle| \leq U\varepsilon$$

and

$$U^{-1} \varepsilon \vec{1} \leq \vec{g}(\varepsilon, \underline{z}) \leq U \vec{1}.$$

PROOF. The equation for $\vec{g}$ is equivalent to

$$L_{\mathscr{S}\underline{z},\varepsilon} \vec{g}(\varepsilon, \mathscr{S}\underline{z}) = \rho(\varepsilon, \mathscr{S}\underline{z}) \vec{g}(\varepsilon, \underline{z}). \tag{4.2}$$

Note that

$$\vec{g}(\varepsilon, \mathscr{S}\underline{z}) = \vec{e}_{z_2} + \varepsilon h(\varepsilon, \mathscr{S}\underline{z}) \vec{f}_{z_2} \quad \text{and} \quad L_{\mathscr{S}\underline{z},\varepsilon} = M_{z_1} + \varepsilon A_{z_1}.$$

Taking the scalar product of both terms in equation (4.2) with $\vec{e}_{z_1}$ and $\vec{f}_{z_1}$ we get

$$\begin{aligned} \rho(\varepsilon, \mathscr{S}\underline{z}) &= \langle M_{z_1} \vec{e}_{z_2}, \vec{e}_{z_1} \rangle + \varepsilon h(\varepsilon, \mathscr{S}\underline{z}) \langle M_{z_1} \vec{f}_{z_2}, \vec{e}_{z_1} \rangle \\ &\quad + \varepsilon \langle A_{z_1} \vec{e}_{z_2}, \vec{e}_{z_1} \rangle + \varepsilon^2 h(\varepsilon, \mathscr{S}\underline{z}) \langle A_{z_1} \vec{f}_{z_2}, \vec{e}_{z_1} \rangle \end{aligned} \tag{4.3}$$

and since $M_{z_1}^t \vec{f}_{z_1} = 0$ and $\langle \vec{f}_{z_1}, \vec{e}_{z_1} \rangle = 0$,

$$\rho(\varepsilon, \mathscr{S}\underline{z}) h(\varepsilon, \underline{z}) = \langle A_{z_1} \vec{e}_{z_2}, \vec{f}_{z_1} \rangle + \varepsilon h(\varepsilon, \mathscr{S}\underline{z}) \langle A_{z_1} \vec{f}_{z_2}, \vec{f}_{z_1} \rangle.$$

We denote by $\mathcal{D}$ the Banach space of continuous functions on $[0, \varepsilon_0] \times \Omega$ equipped with the sup norm. On the ball B_D of radius $D = 4\beta^{-1}$ centered at the origin in $\mathcal{D}$ we define a transformation $\mathscr{T}$ given by

$$\mathscr{T}(h)(\varepsilon, \underline{z}) = \frac{u_1(\varepsilon, \underline{z}) + \varepsilon u_2(\varepsilon, \underline{z}) h(\varepsilon, \mathscr{S}\underline{z})}{u_3(\varepsilon, \underline{z}) + \varepsilon u_4(\varepsilon, \underline{z}) h(\varepsilon, \mathscr{S}\underline{z})}, \tag{4.4}$$

where

$$u_1(\varepsilon,\underline{z}) = \langle A_{z_1}\vec{e}_{z_2}, \vec{f}_{z_1}\rangle, \qquad u_2(\varepsilon,\underline{z}) = \langle A_{z_1}\vec{f}_{z_2}, \vec{f}_{z_1}\rangle,$$
$$u_3(\varepsilon,\underline{z}) = \langle M_{z_1}\vec{e}_{z_2}, \vec{e}_{z_1}\rangle + \varepsilon\langle A_{z_1}\vec{e}_{z_2}, \vec{e}_{z_1}\rangle$$

and

$$u_4(\varepsilon,\underline{z}) = \langle M_{z_1}\vec{f}_{z_2}, \vec{e}_{z_1}\rangle + \varepsilon\langle A_{z_1}\vec{f}_{z_2}, \vec{e}_{z_1}\rangle.$$

Direct computation shows that for all $(\varepsilon,\underline{z}) \in [0,\varepsilon_0]\times\Omega$ we have

$$\beta \le |u_1(\varepsilon,\underline{z})| \le 1, \qquad \beta \le |u_2(\varepsilon,\underline{z})| \le 1, \qquad \beta \le \frac{u_1(\varepsilon,\underline{z})}{u_3(\varepsilon,\underline{z})} \le \beta^{-1},$$
$$\beta-\varepsilon \le |u_3(\varepsilon,\underline{z})| \le 1+\varepsilon, \qquad \beta-\varepsilon \le |u_4(\varepsilon,\underline{z})| \le 1+\varepsilon.$$

We first prove that $\mathscr{T}$ maps B_D into itself. Indeed for $h \in B_D$, since $D = 4\beta^{-1}$ there exists $\varepsilon_0' > 0$ small enough such that for any $\varepsilon \in [0,\varepsilon_0']$,

$$|\mathscr{T}(h)(\varepsilon,\underline{z})| \le \frac{1+\varepsilon D}{\beta-\varepsilon-\varepsilon D(1+\varepsilon)} \le D.$$

We leave to the reader the proof that $\mathscr{T}(h)$ is a continuous function of ε and $\underline{z}$. We now prove that $\mathscr{T}$ is a contraction on B_D. For h and h' in B_D, since $D = 4\beta^{-1}$ there exists $\varepsilon_0 > 0$ small enough, and smaller than ε_0', such that for any $\varepsilon \in [0,\varepsilon_0]$ we have

$$\begin{aligned}
&|\mathscr{T}(h)(\varepsilon,\underline{z}) - \mathscr{T}(h')(\varepsilon,\underline{z})| \\
&\quad = \varepsilon\left|\frac{u_1(\varepsilon,\underline{z})u_4(\varepsilon,\underline{z}) - u_2(\varepsilon,\underline{z})u_3(\varepsilon,\underline{z})}{(u_3(\varepsilon,\underline{z})+\varepsilon u_4(\varepsilon,\underline{z})h(\varepsilon,\mathscr{S}\underline{z}))(u_3(\varepsilon,\underline{z})+\varepsilon u_4(\varepsilon,\underline{z})h'(\varepsilon,\mathscr{S}\underline{z}))}\right| \\
&\qquad \times |h(\varepsilon,\underline{z}) - h'(\varepsilon,\underline{z})| \\
&\quad \le \varepsilon\frac{4}{(\beta-\varepsilon-\varepsilon D(1+\varepsilon))^2}|h(\varepsilon,\underline{z}) - h'(\varepsilon,\underline{z})| \le \frac{1}{2}|h(\varepsilon,\underline{z}) - h'(\varepsilon,\underline{z})|.
\end{aligned}$$

By the contraction mapping principle [see, e.g., Dieudonné (1969)], the map $\mathscr{T}$ has a unique fixed point h in B_D. It follows at once that the vectors

$$\vec{g}(\varepsilon,\underline{z}) = \vec{e}_{z_1} + \varepsilon h(\varepsilon,\underline{z})\vec{f}_{z_1}$$

satisfy equation (4.2). The estimate on $\vec{g}(\varepsilon,\underline{z})$ follows immediately from the fact that $h \in B_D$, and from (4.4),

$$h(\varepsilon,\underline{z}) = \frac{u_1(\varepsilon,\underline{z})}{u_3(\varepsilon,\underline{z})} + \mathcal{O}(\varepsilon).$$

The estimate on $\rho(\varepsilon,\underline{z})$ follows from (4.3). □

REMARK. An easy improvement of the above proof allows to show that ρ and h depend analytically on ε in a small (complex) neighborhood of 0.

By the estimate on $\vec{g}(\varepsilon, \underline{z})$ of the previous lemma and Lemma 3.1 applied to the vector $\vec{1}$, we have μ-almost surely

$$\lim_{n\to\infty} \frac{1}{n} \log \| L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{g}(\varepsilon, \underline{z}) \| = \lim_{n\to\infty} \frac{1}{n} \log \| L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{1} \| = \lambda_1.$$

On the other hand from Lemma 4.1 it follows that

$$\log \| L^{[n-1]}_{\mathscr{S}^{-1}\underline{z}} \vec{g}(\varepsilon, \underline{z}) \| = \sum_{j=0}^{n} \log \rho(\varepsilon, \mathscr{S}^{-j}\underline{z}) + \| \vec{g}(\varepsilon, \mathscr{S}^{-n}\underline{z}) \|.$$

Using again the estimate on $\vec{g}(\varepsilon, \underline{z})$ from Lemma 4.1, the Birkhoff ergodic theorem [Krengel (1985)] and the ergodicity of μ, we have

$$\lambda_1 = \int \log \rho(\varepsilon, \underline{z}) \, d\mu(\underline{z}).$$

The first Lyapunov exponent λ_1 is equal to H the entropy of the process $(Z_t)_{t\in\mathbb{Z}}$ and this entropy has an expansion in terms of ε; see Jacquet, Seroussi and Szpankowski (2008). Therefore

$$H = H_0 + \mathcal{O}(\varepsilon),$$

where H_0 is the entropy of the Markov chain $(X_t)_{t\in\mathbb{Z}}$.

The following theorem is an immediate consequence of the above estimates.

THEOREM 4.2. *If $p_0 \neq p_1$, $\min\{p_0, p_1 1 - p_0, 1 - p_1\} > 0$ and $\varepsilon > 0$ is small enough, we have μ-almost surely*

$$\limsup_{n\to+\infty} \frac{1}{n} \log \left| \nu^{[n]}_{\underline{z},b}(a) - \nu^{[n]}_{\underline{z},c}(a) \right| \leq \log \varepsilon + \log \left| \det(P) \right| - 2H_0 + \mathcal{O}(\varepsilon).$$

Moreover, for μ-almost all $\underline{z}$ there is a triplet (a, b, c) (which may depend on $\underline{z}$) where the equality holds.

PROOF. It is easy to verify that hypotheses (H1)–(H2) are satisfied. We therefore apply Theorems 2.1 and 2.2. The result follows from (4.1) and the above estimate on λ_1. $\square$

As λ_1 and λ_2 are fixed, the above estimate also applies to the asymptotic rate of exponential loss of memory of the measures $\sigma^{[n]}_{\underline{z},\eta}$, $\tilde{\nu}^{[n]}_{\underline{z},\rho}$ and $\tilde{\sigma}^{[n]}_{\underline{z},\eta}$.

Acknowledgments. The authors thank A. Galves for interesting discussions about the subject, and an anonymous referee for useful suggestions. This work is part of FAPESP's project *NeuroMat* (2011/51350-6) and CNPq's project *Probabilistic modeling of brain activity* (480108/2012-9). F. Leonardi is also thankful to the Brazilian–French agreement for cooperation in mathematics and USP-COFECUB project (2009.1.820.45.8).

REFERENCES

BOWEN, R. (2008). *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, revised ed. *Lecture Notes in Math*. **470**. Springer, Berlin. MR2423393

CAPPÉ, O., MOULINES, E. and RYDÉN, T. (2005). *Inference in Hidden Markov Models*. Springer, New York. MR2159833

CAVAZOS-CADENA, R. (2003). An alternative derivation of Birkhoff's formula for the contraction coefficient of a positive matrix. *Linear Algebra Appl*. **375** 291–297. MR2013472

CHAZOTTES, J.-R. and UGALDE, E. (2011). On the preservation of Gibbsianness under symbol amalgamation. In *Entropy of Hidden Markov Processes and Connections to Dynamical Systems*. *London Mathematical Society Lecture Note Series* **385** 72–97. Cambridge Univ. Press, Cambridge. MR2866665

COLLET, P., GALVES, A. and LEONARDI, F. (2008). Random perturbations of stochastic processes with unbounded variable length memory. *Electron*. *J*. *Probab*. **13** 1345–1361. MR2438809

DIEUDONNÉ, J. (1969). *Foundations of Modern Analysis*. *Pure and Applied Mathematics* **10-I**. Academic Press, New York. Enlarged and corrected printing. MR0349288

DOUC, R., MOULINES, E. and RITOV, Y. (2009). Forgetting of the initial condition for the filter in general state-space hidden Markov chain: A coupling approach. *Electron*. *J*. *Probab*. **14** 27–49. MR2471658

DOUC, R., FORT, G., MOULINES, E. and PRIOURET, P. (2009). Forgetting the initial distribution for hidden Markov models. *Stochastic Process*. *Appl*. **119** 1235–1256. MR2508572

JACQUET, P., SEROUSSI, G. and SZPANKOWSKI, W. (2008). On the entropy of a hidden Markov process. *Theoret*. *Comput*. *Sci*. **395** 203–219. MR2424508

KATOK, A. and HASSELBLATT, B. (1995). *Introduction to the Modern Theory of Dynamical Systems*. *Encyclopedia of Mathematics and Its Applications* **54**. Cambridge Univ. Press, Cambridge. MR1326374

KRENGEL, U. (1985). *Ergodic Theorems*. *de Gruyter Studies in Mathematics* **6**. de Gruyter, Berlin. MR0797411

LEDRAPPIER, F. (1984). Quelques propriétés des exposants caractéristiques. In *École D'été de Probabilités de Saint-Flour, XII*—1982. *Lecture Notes in Math*. **1097** 305–396. Springer, Berlin. MR0876081

PETRIE, T. (1969). Probabilistic functions of finite state Markov chains. *Ann*. *Math*. *Statist*. **40** 97–115. MR0239662

CENTRE DE PHYSIQUE THÉORIQUE
CNRS UMR 7644
ECOLE POLYTECHNIQUE
91128 PALAISEAU CEDEX
FRANCE
E-MAIL: pierre.collet@cpht.polytechnique.fr

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
UNIVERSIDADE DE SÃO PAULO
CEP 05508-090, CIDADE UNIVERSITÁRIA
SÃO PAULO, SP
BRASIL
E-MAIL: florencia@usp.br