Available online at www.sciencedirect.com
SCIENCE DIRECT®

Theoretical Computer Science 332 (2005) 543–557

Theoretical Computer Science

www.elsevier.com/locate/tcs

An upper bound for the largest Lyapunov exponent of a Markovian product of nonnegative matrices☆

Reza Gharavi[a], V. Anantharam[b,*]

[a] *School of Electrical Engineering, Cornell University, Ithaca, NY 14843, USA*
[b] *EECS Department, University of California, Berkeley, CA 94720, USA*

Received 9 February 2004; accepted 25 December 2004
Communicated by W. Szpankowski

Abstract

We derive an upper bound for the largest Lyapunov exponent of a Markovian product of nonnegative matrices using Markovian type counting arguments. The bound is expressed as the maximum of a nonlinear concave function over a finite-dimensional convex polytope of probability distributions.

Keywords: Entropy; Lyapunov exponent; Markov chains

1. Introduction

In this paper we derive an upper bound for the largest Lyapunov exponent of a Markovian product of nonnegative matrices. The bound, given in Section 4, is expressed as the maximum of a nonlinear concave function over a finite-dimensional convex set of probability distributions. The bound is derived using Markovian type counting arguments [12], a technique familiar in information theory [11].

In this section, we define the problem, and then give a brief review of some of the related literature. In Section 2 we develop the basic notions underlying the Markovian type counting

☆ Research supported by ONR MURI N00014-1-0637, DARPA grant No. N66001-00-C-8062, and by NSF contract ECS 0123512.

* Corresponding author.
E-mail address: ananth@EECS.Berkeley.EDU (V. Anantharam).

0304-3975/$ - see front matter
doi:10.1016/j.tcs.2004.12.025

technique that we will use to derive our upper bound. The bound itself is derived in Section 3, and in Section 4 it is expressed as a concave optimization problem over a convex polytope of probability distributions. In Section 5 we discuss the quality of the bound.

Let $\mathbf{R}_+ \stackrel{\text{def}}{=} [0, \infty)$ denote the set of nonnegative real numbers. Let $(X_n, n \geqslant 0)$ be an irreducible Markov chain on $\mathcal{X} \stackrel{\text{def}}{=} \{A_1, \ldots, A_K\}$, where $A_i \in \mathbf{R}_+^{p\times p}$, $1 \leqslant i \leqslant K$ are fixed (deterministic) matrices with nonnegative entries. Let P denote the transition probability matrix and π the (unique) stationary distribution of the Markov chain $(X_n, n \geqslant 0)$. We assume that the chain is initialized with its stationary distribution. Note that $\pi(i) > 0$ for all $1 \leqslant i \leqslant K$. For the basic results on finite state Markov chains that we mention without proof, see e.g. [31].

Let λ denote the largest Lyapunov exponent of $(X_n, n \geqslant 0)$.

$$\lambda \stackrel{\text{def}}{=} \lim_{n\to\infty} n^{-1} \mathrm{E} \log \|X_{n-1} \cdots X_0\|. \tag{1}$$

Here we take $\|A\| = \sum_{ij} |A_{ij}|$. The existence of the limit in Eq. (1) is well known [14,24,28,30]; further, it is easy to see that the limit does not depend on the choice of matrix norm. The exact determination of λ is well known to be a difficult problem [1,2,6–8,10,13,21,25,27,29]. The purpose of this paper is to derive an upper bound for λ.

In this paper we assume that $\lambda > -\infty$. It is easy to see that this is equivalent to requiring that there be no sequence $(i_0, \ldots, i_n)$ with

$$\pi(i_0)P(i_0, i_1) \ldots P(i_{n-1}, i_n) > 0 \tag{2}$$

such that the matrix product $A_{i_{n-1}} \cdots A_{i_0}$ is the zero matrix. Since the entries of $A_1, \ldots, A_K$ are nonnegative and finite, it is easily seen by path counting that $\lambda > -\infty$ implies that there are finite constants $-\infty < a_*, a^* < \infty$ such that we have the pointwise bound

$$a_* \leqslant n^{-1} \log \|X_{n-1} \cdots X_0\| \leqslant a^* \tag{3}$$

almost surely.

We will therefore assume the existence of bounds of this type.

To close this section, we briefly discuss the related literature. Excellent surveys of the basic theory of Lyapunov exponents, including historical remarks, are available in [3] and [33]; the latter pays particular attention to products of random matrices. The existence of the limit in Eq. (1), for a general stationary ergodic sequence of matrices $(X_n, n \geqslant 0)$, which defines the largest Lyapunov exponent, is best seen as a small part of Oseledec's multiplicative ergodic theorem, [28], proofs of which are also available in [30] and [9]. It can also be seen as a simple consequence of Kingman's subadditive ergodic theorem [23,24,32], when one has submultiplicativity of the matrix norm, i.e., $\|AB\| \leqslant \|A\|\|B\|$. Little is known in general about how to compute the largest Lyapunov exponent, even for the case of i.i.d. matrices. Several papers consider matrices with nonnegative entries in the i.i.d. case, and sometimes also in the Markov case. A notable paper of Key [22] provides a technique for proving lower bounds to the largest Lyapunov exponent of an i.i.d. product of nonnegative matrices, under some additional hypotheses. Hennion [17] and Peres [29] both study the largest Lyapunov exponent of a product of i.i.d. nonnegative matrices as a function of certain parameters; the latter paper also considers the Markov case. The convergence in distribution of the product of i.i.d. nonnegative matrices with spectral radius 1 is studied by Kesten and Spitzer [20] and in several papers of Mukherjea, see [26] for a survey.

To the best of our knowledge, the approach used in this paper to upper bound the largest Lyapunov exponent of a Markovian product of nonnegative matrices has not appeared earlier in the literature. We were led to study the problem considered here by a desire to understand the convergence behavior of asynchronous computation [4] with a probabilistic model for the delays between the processors [15]. Another reason for the interest in this problem is its relevance to the computation of the entropy of hidden Markov models [5], which has recently begun to be discussed in some depth in the information theory literature [18,19].

2. Preliminaries

In this section we will first define the *Markov type* of a sequence in $\mathcal{X}^n$. We will then give upper and lower bounds for the cardinality and probability of sequences having the same type [1] in terms of the entropy and the information discrimination function. Most of the definitions and results of this section are from [12]. From now on, with some abuse of notation, we identify the set $\mathcal{X}$ with $\{1, \ldots, K\}$, and write elements of $\mathcal{X}^n$ as $x = (i_0 i_1 \ldots i_{n-1})$.

Definition. Let M denote the space of probability measures on $\mathcal{X} \times \mathcal{X}$. We think of M as the unit simplex in $\mathbf{R}^{K \times K}$. Define $v^n: \mathcal{X}^{n+1} \to M$ by $v^n(x) = v$ where $v(i, j) = n^{-1} N(i, j|x)$, $1 \leqslant i, j \leqslant K$. Here $N(i, j|x)$ denotes the number of transitions from i to j in x. We call $v^n(x)$ the *Markov type* of x.

Definition. For $v \in M$ and $1 \leqslant i \leqslant K$, let $v(i, *)$ denote $\sum_j v(i, j)$ and $v(*, i)$ denote $\sum_j v(j, i)$. Define $M^n \stackrel{\text{def}}{=\!=} \text{image}(v^n)$. M^n is called the set of *n-types*.

Remark. Clearly, for all $v \in M^n$ we have

$$|v(i, *) - v(*, i)| \leqslant \frac{1}{n}, \quad 1 \leqslant i \leqslant K. \tag{4}$$

Definition. For $v \in M^n$, define

$$\mathcal{C}^n(v) \stackrel{\text{def}}{=\!=} \{x \in \mathcal{X}^{n+1}: v^n(x) = v\}.$$

$\mathcal{C}^n(v)$ are the sequences of length $n + 1$ of Markov type v.

[1] Since we only consider Markov types in this paper, we will abbreviate this to “type".

Upper and lower bounds for the cardinality and the stationary probability of $\mathcal{C}^n(v)$ in terms of the entropy and the information discrimination function are provided in [12]. To state these results, we first recall the following standard definitions.

Definition. For $v \in M$, the *entropy* of v is defined as

$$H(v) \stackrel{\text{def}}{=\!=} -\sum_{i,j} v(i,j) \log \frac{v(i,j)}{v(i,*)},$$

and the *information discrimination* of v with respect to P is defined as

$$D(v,P) \stackrel{\text{def}}{=\!=} \sum_{i,j} v(i,j) \log \frac{v(i,j)}{v(i,*)P(i,j)},$$

with the convention that $0/0 = 1$ and $0 \log 0 = 0$. All logarithms are to base 2.

Remark. Ignoring the all-zero rows, $H(v)$ is just the entropy of the normalized rows of v averaged over the distribution $\{v(i,*), 1 \leqslant i \leqslant K\}$ and $D(v,P)$ is just the information discrimination between the normalized rows of v and rows of P averaged over the distribution $\{v(i,*), 1 \leqslant i \leqslant K\}$. With a slight abuse of notation, we write $v \ll P$ if and only if $P(i,j) = 0$ implies $v(i,j) = 0$. It is well known that $0 \leqslant H(\cdot) \leqslant \log K$, $H(\cdot)$ is a continuous concave function on M, and $D(\cdot, P)$ is a nonnegative continuous convex function on M that is finite on $\{v \in M : v \ll P\}$; see e.g. [11].

Now let $x = (i_0 \cdots i_n) \in \mathcal{X}^{n+1}$ and $v = v^n(x)$. Then, writing exp for $\exp_2$, we have

$$\begin{aligned} \Pr(x) &= \pi(i_0) \prod_{i,j} P(i,j)^{n v(i,j)} \\ &= \pi(i_0) \exp\left[n \sum_{i,j} v(i,j) \log P(i,j) \right] \\ &= \pi(i_0) \exp[-n\,(D(v,P) + H(v))]. \end{aligned} \tag{5}$$

Therefore, if $v \in M^n$, then for all $x \in \mathcal{C}^n(v)$ we have

$$\pi_* \exp[-n\,(D(v,P) + H(v))] \leqslant \Pr(x) \leqslant \exp[-n\,(D(v,P) + H(v))], \tag{6}$$

where $\pi_* = \min_i \pi(i) > 0$.

The following bounds on the cardinality of $\mathcal{C}^n(v)$, $v \in M^n$, can be derived in a straightforward way from the results in [12]: There is a polynomial $r(n)$ such that

$$[r(n)]^{-1} \exp[nH(v)] \leqslant |\mathcal{C}^n(v)| \leqslant K \exp[nH(v)]. \tag{7}$$

In conjunction with (6), this yields the bounds on $\Pr(\mathcal{C}^n(v))$, $v \in M^n$:

$$\pi_* [r(n)]^{-1} \exp[-nD(v,P)] \leqslant \Pr(\mathcal{C}^n(v)) \leqslant K \exp[-nD(v,P)]. \tag{8}$$

3. An upper bound for λ

In this section we find an upper bound for λ, by focusing on the ε-typical sequences in $\mathcal{X}^{n+1}$. In the next section, this bound will be expressed as a concave optimization problem over a convex polytope of probability distributions.

Definition. Given $\varepsilon > 0$, define the set of ε-*typical n-types* by

$$D_\varepsilon^n \stackrel{\text{def}}{=\!=} \{v \in M^n : D(v, P) \leqslant \varepsilon\},$$

and define the set of ε-*typical sequences of length* $n+1$ by

$$\mathcal{D}_\varepsilon^n \stackrel{\text{def}}{=\!=} \mathcal{C}^n(D_\varepsilon^n) = \{x \in \mathcal{X}^{n+1} : D(v^n(x), P) \leqslant \varepsilon\}.$$

We now estimate $\Pr(\mathcal{D}_\varepsilon^n)$ by estimating $\Pr\left(\mathcal{X}^{n+1} - \mathcal{D}_\varepsilon^n\right) = \Pr(\mathcal{C}^n(M^n - D_\varepsilon^n))$. We write

$$\begin{aligned}\Pr(\mathcal{X}^{n+1} - \mathcal{D}_\varepsilon^n) &= \sum_{v \in M^n - D_\varepsilon^n} \Pr(\mathcal{C}^n(v)) \\ &\leqslant (n+1)^{K^2} \max_{v \in M^n - D_\varepsilon^n} \Pr(\mathcal{C}^n(v)) \\ &\leqslant K(n+1)^{K^2} \max_{v \in M^n - D_\varepsilon^n} \exp[-nD(v, P)] \\ &\leqslant K(n+1)^{K^2} \exp[-n\varepsilon],\end{aligned}$$

where $|M^n| \leqslant (n+1)^{K^2}$ was used in the first inequality and (8) was used in the second. Note that, for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \Pr(\mathcal{X}^{n+1} - \mathcal{D}_\varepsilon^n) = 0 \quad \text{and} \quad \lim_{n\to\infty} \Pr(\mathcal{D}_\varepsilon^n) = 1. \tag{9}$$

Therefore, for all $\varepsilon > 0$, we can write

$$\begin{aligned}\lambda &= \lim_{n\to\infty} \Big[n^{-1} \mathrm{E} 1\{\mathcal{D}_\varepsilon^n\} \log \|X_{n-1} \cdots X_0\| \\ &\qquad + n^{-1} \mathrm{E} 1\{\mathcal{X}^{n+1} - \mathcal{D}_\varepsilon^n\} \log \|X_{n-1} \cdots X_0\| \Big] \\ &= \lim_{n\to\infty} n^{-1} \mathrm{E} 1\{\mathcal{D}_\varepsilon^n\} \log \|X_{n-1} \cdots X_0\| \\ &= \lim_{n\to\infty} n^{-1} [\Pr(\mathcal{D}_\varepsilon^n)]^{-1} \mathrm{E} 1\{\mathcal{D}_\varepsilon^n\} \log \|X_{n-1} \cdots X_0\|,\end{aligned} \tag{10}$$

where (3) and (9) were used in the second equality and (9) was used in the third. Observing that $[\Pr(\mathcal{D}_\varepsilon^n)]^{-1} \mathrm{E} 1\{\mathcal{D}_\varepsilon^n\}(\cdot)$ is an expectation operator, we use Jensen's inequality to write

$$\begin{aligned}\lambda &\leqslant \varliminf_{n\to\infty} n^{-1} \log [\Pr(\mathcal{D}_\varepsilon^n)]^{-1} \mathrm{E} 1\{\mathcal{D}_\varepsilon^n\} \|X_{n-1} \cdots X_0\| \\ &= \varliminf_{n\to\infty} n^{-1} \log \mathrm{E} 1\{\mathcal{D}_\varepsilon^n\} \|X_{n-1} \cdots X_0\|,\end{aligned}$$

where (9) was again used in that last equality. Since the last quantity is nonincreasing as $\varepsilon \downarrow 0$, we can upper bound λ by

$$\hat{\lambda} \stackrel{\text{def}}{=\!=} \lim_{\varepsilon\downarrow 0} \varliminf_{n\to\infty} n^{-1} \log \mathrm{E} 1\{\mathcal{D}_\varepsilon^n\} \|X_{n-1} \cdots X_0\|.$$

4. Calculation of $\hat{\lambda}$

In this section, we will reformulate $\hat{\lambda}$ in terms of the solution to a concave optimization problem. Starting from its definition, we write

$$\begin{aligned}
\hat{\lambda} &= \lim_{\varepsilon \downarrow 0} \varliminf_{n\to\infty} n^{-1} \log \mathrm{E}1\{\mathcal{D}^n_\varepsilon\}\|X_{n-1}\cdots X_0\| \\
&= \lim_{\varepsilon \downarrow 0} \varliminf_{n\to\infty} n^{-1} \log \mathrm{E} \sum_{v\in D^n_\varepsilon} 1\{\mathcal{C}^n(v)\}\|X_{n-1}\cdots X_0\| \\
&= \lim_{\varepsilon \downarrow 0} \varliminf_{n\to\infty} n^{-1} \log \max_{v\in D^n_\varepsilon} \mathrm{E}1\{\mathcal{C}^n(v)\}\|X_{n-1}\cdots X_0\| \\
&= \lim_{\varepsilon \downarrow 0} \varliminf_{n\to\infty} n^{-1} \log \max_{v\in D^n_\varepsilon} \sum_{x\in\mathcal{C}^n(v)} \Pr(x)\|A_{i_{n-1}}\cdots A_{i_0}\| \\
&= \lim_{\varepsilon \downarrow 0} \varliminf_{n\to\infty} n^{-1} \log \max_{v\in D^n_\varepsilon} \sum_{x\in\mathcal{C}^n(v)} \sum_{u_0,\ldots,u_n} \Pr(x) \prod_{l=0}^{n-1} A^{\mathrm{T}}_{i_l}(u_l, u_{l+1}), \qquad (11)
\end{aligned}$$

where $|D^n_\varepsilon| \leqslant |M^n| \leqslant (n+1)^{K^2}$ was used in the third equality.

We now compute this double sum combinatorially. We do this by introducing an extended alphabet $\mathcal{Y}$ and replacing the double sum in (11) by a single sum over sequences of this extended alphabet.

Definition. Let $\mathcal{Y} \stackrel{\text{def}}{=} \{1, \ldots, K\} \times \{1, \ldots, p\}$. Let $\mathcal{M}$ denote the space of probability measures on $\mathcal{Y} \times \mathcal{Y}$. A sequence $y \in \mathcal{Y}^{n+1}$ will be written $(i_0, u_0; \ldots; i_n, u_n)$. Define $\eta^n : \mathcal{Y}^{n+1} \to \mathcal{M}$ by $\eta^n(y) = \eta$ where $\eta(i, u; j, v) = n^{-1} N^n(i, u; j, v|y)$. Here $N^n(i, u; j, v|y)$ denotes the number of transitions from (i, u) to (j, v) in y. We call $\eta^n(y)$ the *extended Markov type* of y. Let $\mathcal{M}^n$ denote the image of η^n. $\mathcal{M}^n$ is called the set of *extended n-types*. For $\eta \in \mathcal{M}^n$, we write $\mathcal{C}^n(\eta)$ for $\{y \in \mathcal{Y}^{n+1} : \eta^n(y) = \eta\}$.

Remark. For all $\eta \in \mathcal{M}^n$ we have

$$|\ \eta(i, u; *, *) - \eta(*, *; i, u)\ | \leqslant 1, \quad (i, u) \in \mathcal{Y}, \qquad (12)$$

where $\eta(i, u; *, *)$ denotes $\sum_{j,v} \eta(i, u; j, v)$ and $\eta(*, *; i, u)$ denotes $\sum_{j,v} \eta(j, v; i, u)$.

Remark. We have a map $m : \mathcal{M} \to M$ given by $m(\eta) = v$ where

$$v(i, j) = \sum_{u,v} \eta(i, u; j, v).$$

Note that m maps $\mathcal{M}$ onto M. For $v \in M^n$, we write $\mathcal{M}^n(v) = \{\eta \in \mathcal{M}^n : m(\eta) = v\}$.

Definition. The set of *ε-typical extended n-types* is defined as

$$T^n_\varepsilon = \{\eta \in \mathcal{M}^n : m(\eta) \in D^n_\varepsilon\}.$$

Notation. For $x \in \mathcal{X}^{n+1}$, we write $\mathcal{Y}^{n+1}(x)$ for the set of $y = (i_0, u_0; \ldots; i_n, u_n)$ such that $x = (i_0, \ldots, i_n)$.

We may now continue from Eq. (11) to write

$$\hat{\lambda} = \lim_{\varepsilon\downarrow 0} \varliminf_{n\to\infty} n^{-1} \log \max_{v\in D_\varepsilon^n} \sum_{x\in\mathcal{C}^n(v)} \sum_{y\in\mathcal{Y}^{n+1}(x)} \pi(i_0) \prod_{l=0}^{n-1} P(i_l, i_{l+1}) A_{i_l}^{\mathrm{T}}(u_l, u_{l+1}).$$

Noting that $0 < \pi_* \leqslant \pi(i_0) \leqslant 1$, this can be rewritten as

$$\hat{\lambda} = \lim_{\varepsilon\downarrow 0} \varliminf_{n\to\infty} n^{-1} \log \max_{v\in D_\varepsilon^n} \sum_{\eta\in\mathcal{M}^n(v)} |\mathcal{C}^n(\eta)| \prod_{i,u;\,j,v} \left(P(i,j) A_i^{\mathrm{T}}(u,v)\right)^{n\eta(i,u;\,j,v)}.$$

Let

$$H(\eta) \stackrel{\text{def}}{=\!=} -\sum_{i,u;\,j,v} \eta(i,u;\,j,v) \log \frac{\eta(i,u;\,j,v)}{\eta(i,u;\,*,*)}.$$

Then, as in (7), there is a polynomial $r(n)$ such that

$$[r(n)]^{-1} \exp[nH(\eta)] \leqslant |\mathcal{C}^n(\eta)| \leqslant K \exp[nH(\eta)].$$

Also we have $|\mathcal{M}^n(\eta)| \leqslant |\mathcal{M}^n| \leqslant (n+1)^{(Kp)^2}$. Hence we get

$$\hat{\lambda} = \lim_{\varepsilon\downarrow 0} \varliminf_{n\to\infty} \max_{\eta\in T_\varepsilon^n} n^{-1} \log \exp[n(H(\eta) + F_0(\eta) + F(\eta))],$$

$$\hat{\lambda} = \lim_{\varepsilon\downarrow 0} \varliminf_{n\to\infty} \max_{\eta\in T_\varepsilon^n} [H(\eta) + F_0(\eta) + F(\eta)],$$

where

$$F_0(\eta) \stackrel{\text{def}}{=\!=} \sum_{i,u;\,j,v} \eta(i,u;\,j,v) \log P(i,j),$$

$$F(\eta) \stackrel{\text{def}}{=\!=} \sum_{i,u;\,j,v} \eta(i,u;\,j,v) \log A_i^{\mathrm{T}}(u,v).$$

Now $H(\cdot)$ is a concave continuous function on $\mathcal{M}$ and $F_0(\cdot)$ and $F(\cdot)$ are upper semi-continuous functions that are finite on the closed convex subsets given by the intersection of $\mathcal{M}$ with the linear subspaces defined by

$$\eta(i,u;\,j,v) = 0 \qquad \text{for all } (i,u;\,j,v) \text{ such that } P(i,j) = 0, \text{ and}$$
$$\eta(i,u;\,j,v) = 0 \qquad \text{for all } (i,u;\,j,v) \text{ such that } A_i^{\mathrm{T}}(u,v) = 0,$$

respectively.

From these observations we see that the $\varliminf$ is actually a limit and we have

$$\begin{aligned} \hat{\lambda} &= \lim_{\varepsilon\downarrow 0} \max_{\eta: D(m(\eta),P)\leqslant\varepsilon} \left[H(\eta) + F_0(\eta) + F(\eta)\right] \\ &= \max_{\eta: m(\eta)=\pi P} \left[H(\eta) + F_0(\eta) + F(\eta)\right], \end{aligned} \tag{13}$$

where $\pi P \in M$ is defined by $\pi P(i,j) = \pi(i) P(i,j)$.

We next note that the term corresponding to $F_0(\cdot)$ is superfluous in the optimization problem to determine $\hat{\lambda}$. This is because, for any $\eta \in \mathcal{M}$ with $m(\eta) = \pi P$, we have

$$\begin{aligned} F_0(\eta) &= \sum_{i,u;j,v} \eta(i,u;\ j,v) \log P(i,\ j) \\ &= \sum_{i,j} \pi(i) P(i,\ j) \log P(i,\ j) \\ &= -H(\pi P). \end{aligned}$$

So we may write

$$\hat{\lambda} = \max_{\eta \in \mathcal{M}} \left[H(\eta) + F(\eta) \right] - H(\pi P), \tag{14}$$

subject to the constraints

$$m(\eta) = \pi P, \tag{15}$$

$$\eta(i,u;\ j,v) = 0 \qquad \text{for all } (i,u;\ j,v) \text{ such that } A_i^{\mathrm{T}}(u,v) = 0, \tag{16}$$

$$\eta(i,u;*,*) = \eta(*,*;i,u), \qquad 1 \leqslant i \leqslant K,\ 1 \leqslant u \leqslant p. \tag{17}$$

This is our upper bound for the largest Lyapunov exponent λ. Here the constraint (15) already appeared in (13), the constraint (16) is imposed because $F(\eta) = -\infty$ for all $\eta \in \mathcal{M}$ that do not satisfy this constraint, and (17) is a consequence of (12), which every extended n-type must satisfy.

5. Discussion

We first verify that the domain of the constrained optimization problem defined by Eqs. (14–17) is nonempty whenever $\lambda > -\infty$. This is already clear from the result that the solution of the optimization problem, namely $\hat{\lambda}$, is an upper bound to λ, but can also easily be directly verified. The condition for $\lambda > -\infty$ is that for every sequence of states $(i_0, \ldots, i_n)$ satisfying (2) there is a sequence of coordinates $(u_0, \ldots, u_n)$ with

$$A_{i_0}^{\mathrm{T}}(u_0,u_1) A_{i_1}^{\mathrm{T}}(u_1,u_2) \ldots A_{i_{n-1}}^{\mathrm{T}}(u_{n-1},u_n) > 0. \tag{18}$$

Pick one such sequence of coordinates for each such sequence of states, in an arbitrary way, and associate to such a sequence of states $(i_0, \ldots, i_n)$ the extended Markov type of the chosen sequence $(i_0, u_0; \ldots, i_n, u_n)$. Let η be a limit of a sequence of such extended types for which the type of the underlying sequence of states converges to πP. The compactness of $\mathcal{M}$ and the ergodic theorem for the underlying irreducible Markov chain ensure the existence of at least one such η. It is straightforward to verify that any such η lies in the domain of the optimization problem defined by Eqs. (14–17). We thus have $\hat{\lambda} > -\infty$ whenever $\lambda > -\infty$.

Next suppose one of the matrices A_i, $1 \leqslant i \leqslant K$ is the zero matrix. Then $\lambda = -\infty$. Also, the domain of the constrained optimization problem defined by Eqs. (14–17) is empty, because it is impossible to find $\eta \in \mathcal{M}$ satisfying both conditions (16) and (15). Thus we also have $\hat{\lambda} = -\infty$.

On the other hand, it is possible to have $\hat{\lambda} > -\infty$ even when $\lambda = -\infty$, as shown by the following simple example.

Example 1. Let $K = 3$ and $p = 2$. The transition probability matrix of the underlying Markov chain is given by

$$P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

This Markov chain is irreducible with stationary distribution

$$\pi = \left[\frac{1}{2}, \frac{1}{4}, \frac{1}{4}\right].$$

Let the nonnegative matrices corresponding to the individual states of the Markov chain be given by

$$A_1^{\mathrm{T}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A_2^{\mathrm{T}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \text{and } A_3^{\mathrm{T}} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}.$$

Consider the sequence of states $(i_0, i_1, i_2, i_3) = (2, 1, 2, 1)$. This has strictly positive probability, equal to $\frac{1}{8}$, in the underlying Markov chain. However

$$A_2^{\mathrm{T}} A_1^{\mathrm{T}} A_2^{\mathrm{T}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

This verifies that $\lambda = -\infty$.

In this example we have

$$\pi P = \begin{bmatrix} 0 & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & 0 & 0 \\ \frac{1}{4} & 0 & 0 \end{bmatrix}.$$

It is straightforward to verify that the probability distribution η on $\{(i, u) : 1 \leqslant i \leqslant 3, 1 \leqslant p \leqslant 2\}$, with the rows and columns indexed in lexicographic order, given by

$$\eta = \begin{bmatrix} 0 & 0 & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{4} \\ 0 & \frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{19}$$

lies in the domain of the optimization problem defined by Eqs. (14–17). It follows that $\hat{\lambda} > -\infty$.[2]

The basic feature of Example 1 is the existence of an infinite path through the states of the underlying Markov chain the types of whose finite initial segments converge to πP, and for which the product of the matrices corresponding to any finite initial segment of the path is not identically zero. It is of course possible to construct many examples of this kind, including ones where a transition is possible in the underlying Markov chain between every pair of states. One sees immediately from this that one can construct examples with $\lambda > -\infty$ where $\hat{\lambda}$ is a rather poor upper bound for λ. Indeed, one can start with an example where $\lambda = -\infty$ and having the feature identified in Example 1 as giving $\hat{\lambda} > -\infty$ and then modify the underlying matrices so as to make $0 \geqslant \lambda > -\infty$ without significantly affecting $\hat{\lambda}$. For instance, the following example is constructed from Example 1 by following this approach.

Example 2. Let $K = 3$, $p = 2$, and let the underlying Markov chain have transition probability matrix P as in Example 1, so that π and πP are as defined there. Let the nonnegative matrices corresponding to the individual states of the Markov chain be given by

$$A_1^{\mathrm{T}} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad A_2^{\mathrm{T}} = \begin{bmatrix} \delta & 1 \\ 0 & 0 \end{bmatrix}, \quad \text{and } A_3^{\mathrm{T}} = \begin{bmatrix} 0 & 0 \\ 1 & \delta \end{bmatrix},$$

where $1 \geqslant \delta > 0$.

In this example one can check that $\lambda = \frac{1}{4} \log \delta$. However $\hat{\lambda}$ must be at least as big as the one in Example 1, since the choice of η in Eq. (19) continues to satisfy the constraints (15–17) for this problem and the objective of the optimization problem (14) evaluates to the same number at this η in both examples.

Let us return to the situation where $\lambda = -\infty$. Thus, there exists a sequence of states $(i_0, \ldots, i_n)$ satisfying (2) for which $A_{i_0}^{\mathrm{T}} A_{i_1}^{\mathrm{T}} \ldots A_{i_{n-1}}^{\mathrm{T}}$ is the zero matrix. This tells us that if, for the underlying Markov chain, instead of the originally given one we take the one whose states are comprised of blocks of states of the original chain, then, if the block size is sufficiently large, one cannot have a phenomenon like that in Example 1: indeed for the upper bound defined as the solution to the optimization problem (14) subject to the constraints (15–17) with the new underlying chain, the domain of the problem becomes empty, so the upper bound once again becomes $-\infty$. We call the process of working with blocks of the underlying Markov chain *amortization*, since time has to be normalized to be on the same scale for all blocks. We now formally define, for each $L \geqslant 1$, the *L-amortized upper bound*.

---

[2] One can check that η given in Eq. (19) is the unique point in the domain of the optimization problem, so that here $\hat{\lambda} = -\frac{1}{2}$.

Definition. Write $i^{(L)}$ for $(i_0, \dots, i_{L-1}) \in \mathcal{X}^L$, and similarly write $j^{(L)}$. Let $\mathcal{X}^{(L)}$ be given by

$$\mathcal{X}^{(L)} \stackrel{\text{def}}{=\!=} \{i^{(L)} \in \mathcal{X}^L \ : \ P(i_0, i_1) \dots P(i_{L-2}, i_{L-1}) > 0\}.$$

Let $M^{(L)}$ denote the space of probability measures on $\mathcal{X}^{(L)} \times \mathcal{X}^{(L)}$. Let $(\pi P)^{(L)} \in M^{(L)}$ be defined by

$$(\pi P)^{(L)}(i^{(L)}, j^{(L)}) \stackrel{\text{def}}{=\!=} \pi(i_0) P(i_0, i_1) \dots P(i_{L-2}, i_{L-1}) \\ P(i_{L-1}, j_0) P(j_0, j_1) \dots P(j_{L-2}, j_{L-1}).$$

Let $\mathcal{Y}^{(L)}$ denote $\mathcal{X}^{(L)} \times \{1, \dots, p\}$, and let $\mathcal{M}^{(L)}$ denote the space of probability measures on $\mathcal{Y}^{(L)} \times \mathcal{Y}^{(L)}$. Let $m^{(L)} : \mathcal{M}^{(L)} \mapsto M^{(L)}$ be given by $m^{(L)}(\eta^{(L)}) = v^{(L)}$, where

$$v^{(L)}(i^{(L)}, j^{(L)}) = \sum_{u,v} \eta^{(L)}(i^{(L)}, u;\ j^{(L)}, v).$$

For $\eta^{(L)} \in \mathcal{M}^{(L)}$, the entropy $H(\eta^{(L)})$ is defined, in the usual way, as

$$H(\eta^{(L)}) \stackrel{\text{def}}{=\!=} - \sum_{i^{(L)}, u;\, j^{(L)}, v} \eta^{(L)}(i^{(L)}, u;\ j^{(L)}, v) \log \frac{\eta^{(L)}(i^{(L)}, u;\ j^{(L)}, v)}{\eta^{(L)}(i^{(L)}, u; *, *)},$$

where

$$\eta^{(L)}(i^{(L)}, u; *, *) \stackrel{\text{def}}{=\!=} \sum_{j^{(L)}, v} \eta^{(L)}(i^{(L)}, u;\ j^{(L)}, v).$$

Define the function $F^{(L)}$ on $\mathcal{M}^{(L)}$ by

$$F^{(L)}(\eta^{(L)}) \stackrel{\text{def}}{=\!=} \sum_{i^{(L)}, u;\, j^{(L)}, v} \eta^{(L)}(i^{(L)}, u;\ j^{(L)}, v) \log A^{\mathrm{T}}_{i^{(L)}}(u, v),$$

where

$$A^{\mathrm{T}}_{i^{(L)}} \stackrel{\text{def}}{=\!=} A^{\mathrm{T}}_{i_0} A^{\mathrm{T}}_{i_1} \dots A^{\mathrm{T}}_{i_{L-1}}.$$

Then the L-amortized upper bound is defined as

$$\hat{\lambda}^{(L)} = \frac{1}{L} \left(\max_{\eta^{(L)} \in \mathcal{M}^{(L)}} \left[H(\eta^{(L)}) + F^{(L)}(\eta^{(L)}) \right] \right) - H(\pi P), \tag{20}$$

subject to the constraints

$$m^{(L)}(\eta^{(L)}) = (\pi P)^{(L)}, \tag{21}$$

$$\eta^{(L)}(i^{(L)}, u;\ j^{(L)}, v) = 0 \qquad \text{for all } (i^{(L)}, u;\ j^{(L)}, v) \text{ with } A^{\mathrm{T}}_{i^{(L)}}(u, v) = 0, \tag{22}$$

$$\eta(i^{(L)}, u; *, *) = \eta(*, *; i^{(L)}, u), \qquad \text{for all } (i^{(L)}, u). \tag{23}$$

That $\hat{\lambda}^{(L)}$ is an upper bound for the largest Lyapunov exponent is an immediate consequence of the earlier development, once one recognizes that $H((\pi P)^{(L)}) = LH(\pi P)$, where the entropy, $H((\pi P)^{(L)})$ of $(\pi P)^{(L)}$ is defined in the usual way.[3] We now have the following result:

Theorem.

$$\lim_{L\to\infty} \hat{\lambda}^{(L)} = \lambda.$$

Proof. While we are only interested in the case $\lambda > -\infty$, observe that we have already argued that if $\lambda = -\infty$ then $\lim_{L\to\infty} \hat{\lambda}^{(L)} = -\infty$, so the theorem holds in this case. Now suppose that $\lambda > -\infty$. Then we have $\hat{\lambda}^{(L)} > -\infty$ for all $L \geqslant 1$, so that there is at least one $\eta^{(L)}$ in the domain of the optimization problem (20) with constraints (21–23).

It is straightforward to check that for any such $\eta^{(L)}$ we have

$$H(\eta^{(L)}) \leqslant H((\pi P)^{(L)}) + \log(p^2).$$

This is a direct consequence of standard entropy inequalities [11] using (21) once one recognizes that the conditional distribution of $\eta^{(L)}(i^{(L)}, u; j^{(L)}, v)$ given $(i^{(L)}, j^{(L)})$ lives on a set of cardinality at most p^2.

We also observe that for every $i^{(L)} \in \mathcal{X}^{(L)}$ and every $(u, v) \in \{1, \dots, p\} \times \{1, \dots, p\}$, we have $A^{\mathrm{T}}_{i^{(L)}}(u, v) \leqslant \|A_{i^{(L)}}\|$. Hence

$$\begin{aligned}
F^{(L)}(\eta^{(L)}) &\stackrel{\text{def}}{=} \sum_{i^{(L)},u;j^{(L)},v} \eta^{(L)}(i^{(L)}, u; j^{(L)}, v) \log A^{\mathrm{T}}_{i^{(L)}}(u, v) \\
&\leqslant \sum_{i^{(L)},u;j^{(L)},v} \eta^{(L)}(i^{(L)}, u; j^{(L)}, v) \log \|A_{i^{(L)}}\| \\
&= \sum_{i^{(L)}} \pi(i_0) P(i_0, i_1) \dots P(i_{L-2}, i_{L-1}) \log \|A_{i^{(L)}}\| \\
&= E \log \|X_{L-1} \dots X_0\|,
\end{aligned}$$

where the notation in the last equation is as in Eq. (1).

Putting these observations together, we get

$$\hat{\lambda}^{(L)} \leqslant \frac{1}{L} E \log \|X_{L-1} \dots X_0\| + \frac{1}{L} \log p^2,$$

where we have used the fact that $H((\pi P)^{(L)}) = LH(\pi P)$. Since we already know that $\hat{\lambda}^{(L)}$ is an upper bound for λ, taking the limit as $L \to \infty$ and appealing to (1) proves the theorem. □

We now describe the results of some numerical experiments we carried out, which suggest that the upper bound can sometimes be quite good, even without the need for amortization.

[3] Strictly speaking, if the underlying Markov chain is periodic with period d, then unless L is coprime with d the new underlying Markov chain at the level of blocks is no longer irreducible, as was assumed in the earlier development. It is not hard to show that $\hat{\lambda}^{(L)}$ is still an upper bound for the largest Lyapunov exponent of the original problem even in this case.

To explain how we arrived at the numbers reported here, we first say a few words about the simulation methodology. Since an analytic expression for λ is not known, we estimated λ using (1) as follows. Define

$$\lambda_n \stackrel{\text{def}}{=} n^{-1}\mathrm{E}\log\|X_{n-1}\cdots X_0\|.$$

Two issues need to be addressed. First, how fast does λ_n converge to λ; and second, how should λ_n for a given n be estimated. To address the second issue, we assumed $L_n \stackrel{\text{def}}{=} n^{-1}\log\|X_{n-1}\cdots X_0\|$ to have a normal distribution. To estimate the mean of L_n, i.e., λ_n, the mean and variance of samples obtained (using a random number generator) were used to construct confidence intervals. All of the confidence intervals used have a confidence coefficient of at least 0.999.

To address the first issue, we estimated $\lambda_{5\times10^7}$ and compared its value to an estimate for $\lambda_{5\times10^6}$. In all cases below, the estimate for $\lambda_{5\times10^6}$ was within the confidence interval of $\lambda_{5\times10^7}$ (and its confidence interval was the same as that of $\lambda_{5\times10^7}$). So, in the following example, we assumed λ to lie within the confidence interval of $\lambda_{5\times10^7}$.

Example 4. Consider the asynchronous computation of the equation $x_{n+1} = Ax_n$, where A is a 2×2 matrix, each component is handled by a separate processor. We assume that the computation proceeds as

$$x_n^i = \sum_j a_{ij}x^j_{n-d_{ij}(n)}.$$

Here the matrix $d(n) = [d_{ij}(n)]$ is a matrix of delays, which is assumed to evolve in a Markovian way: it can be one of two values d_1 or d_2, with transition matrix P. Consider the numerical values

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}, \quad d_1 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad d_2 = \begin{bmatrix} 1 & 2 \\ 2 & 2 \end{bmatrix}, \quad \text{with } P = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}.$$

It is straightforward to see that the evolution of the computation can be described through a Markovian product of fixed nonnegative matrices. The underlying Markov chain is a two state chain with transition probability matrix P. The matrices are 4×4 matrices: the matrix applied to determine $[x_n^1, x_n^2]$ can be thought of as determining $[x_n^1, x_n^2, x_{n-1}^1, x_{n-1}^2]$ in terms of $[x_{n-1}^1, x_{n-1}^2, x_{n-2}^1, x_{n-2}^2]$; the matrix that is applied depends on the state of the underlying Markov chain.

The following results were obtained using our technique to bound the largest Lyapunov exponent of the computation, without any amortization

	$\lambda_{5\times10^7}$	$\hat{\lambda}$	Error (%)
$p = 0.1$	1.0950 ± 10^{-4}	1.0969	≈ 0.2
$p = 0.3$	1.2086 ± 10^{-4}	1.2164	≈ 0.7
$p = 0.5$	1.2926 ± 10^{-4}	1.3012	≈ 0.7
$p = 0.7$	$1.3566 \pm 2\times10^{-4}$	1.3618	≈ 0.4
$p = 0.9$	$1.4053 \pm 3\times10^{-4}$	1.4064	≈ 0.08

Observe that λ is an increasing function of p. It appears that our estimate performed very well for a wide range of p.

Acknowledgements

This major part of this work was done in 1994, and formed part of the first author's doctoral dissertation at Cornell University [15], which concerned asynchronous computation [4] with random delays. The major contribution of this thesis, excluding these results, was eventually published in [16]. We would like to thank Wojciech Szpankowski for encouraging us to also submit these results for publication.

References

[1] L. Arnold, H. Crauel, J.-P. Eckmann (Eds.), Lyapunov Exponents: Proceedings, Oberwolfach 1990, Lecture Notes in Mathematics, Vol. 1486, Springer, Berlin, 1991.
[2] L. Arnold, V. Wihstutz (Eds.), Lyapunov Exponents: Proceedings, Bremen 1984, Lecture Notes in Mathematics, Vol. 1186, Springer, Berlin, 1986.
[3] L. Arnold, V. Wihstutz, Lyapunov exponents: a survey, in: L. Arnold, V. Wihstutz (Eds.), Lyapunov Exponents: Proceedings, Bremen 1984, Lecture Notes in Mathematics, Vol. 1186, Springer, Berlin, 1986, pp. 1–26.
[4] D.P. Bertsekas, J.N. Tsitsiklis, Parallel and Distributed Computation, Prentice Hall, Englewood Cliffs, NJ, 1989.
[5] D. Blackwell, The entropy of functions of finite-state Markov chains, Transactions of the First Prague Conference on Information Theory, Statistical Decision Functions, and Random Processes, Liblice, Czechoslovakia, November 28–30, 1956, Publishing House of the Czechoslovak Academy of Sciences, Prague, 1957.
[6] V.D. Blondel, J.N. Tsitsiklis, The Lyapunov exponent and joint spectral radius of pairs of matrices are hard—when not impossible—to compute and to approximate, Mathematics of Control, Signals, and Systems, Vol. 10, no. 1, pp. 31–40, 1997. Correction published in Vol. 10, no. 4, p. 381, 1997.
[7] P. Bougerol, J. Lacroix, Products of Random Matrices with Applications to Schrödinger Operators, Progress in Probability and Statistics, Vol. 8, Birkhäuser, Boston, 1985.
[8] J.E. Cohen, H. Kesten, C.M. Newman (Eds.), Random Matrices and their Applications, Vol. 50, American Mathematical Society, Providence, 1986.
[9] J.E. Cohen, H. Kesten, C.M. Newman, Oseledec's multiplicative ergodic theorem: a proof, in: J.E. Cohen, H. Kesten, C. M. Newman (Eds.), Random Matrices and Their Applications, Vol. 50, American Mathematical Society, Providence, 1986, pp. 23–30.
[10] A. Crisanti, G. Paladin, A. Vulpiani, Products of Random Matrices in Statistical Physics, Springer Series in Solid State Sciences, Vol. 104, Springer, Berlin, 1993.
[11] I. Csiszár, J. Körner, Information Theory: Coding Theorems for Discrete Memoryless Systems, Academic Press, New York, 1981.
[12] L.E. Davisson, G. Longo, A. Sgarro, The error exponent for the noiseless encoding of finite ergodic Markov sources, IEEE Trans. Inform. Theory (1981) 431–438.
[13] J.M. Deutsch, G. Paladin, Products of random matrices in a microcanonical ensemble, Physical Rev. Lett. 62 (7) (1989) 695–699.
[14] H. Furstenberg, H. Kesten, Products of random matrices, Ann. Math. Statist. 31 (1960) 457–469.
[15] R. Gharavi, A Structure Theorem for Partially Asynchronous Relaxations with Random Delays, Doctoral Dissertation, Cornell University, Ithaca, New York, 1995.
[16] R. Gharavi, V. Anantharam, Structure theorems for partially asynchronous iterations of a nonnegative matrix with random delays, SADHANA: Proceedings of the Indian Academy of Sciences, Vol. 24, Parts 4 and 5, 1999, pp. 369–424.

[17] H. Hennion, Dérivabilité du plus grand exposant caractéristique des produits de matrices aléatoires indépendantes à coefficients positifs, Ann. Inst. H. Poincaré Probab. Statist. 27 (1) (1991) 27–59.
[18] T. Holliday, P. Glynn, A. Goldsmith, On entropy and Lyapunov exponents for finite-state channels, IEEE Trans. Information Theory, submitted for publication. Preprint available online at http://wsl.stanford.edu/Publications/THolliday/Lyapunov.pdf.
[19] P. Jacquet, G. Seroussi, W. Szpankowski, On the entropy of a hidden Markov process, Proceedings of the Data Compression Conference, Snowbird, Utah, March 23–25, 2004, to appear. Preprint available online at http://www.cs.purdue.edu/homes/spa/publications.html.
[20] H. Kesten, F. Spitzer, Convergence in distribution of products of random matrices, Zeitschrift für Wahrscheinlichkeit und verwandete Gebiete, Vol. 67, 1984, pp. 363–386.
[21] E.S. Key, Computable examples of the maximal Lyapunov exponent, Probab. Theory Related Fields 75 (1987) 97–107.
[22] E.S. Key, Lower bounds for the maximal Lyapunov exponent, J. Theoret. Probab. 3 (3) (1990) 477–488.
[23] J.F.C. Kingman, The ergodic theory of subadditive stochastic processes, J. Roy. Statist. Soc. Series B 30 (1968) 499–510.
[24] J.F.C. Kingman, Subadditive processes, in: École d'Été de Probabilités de Saint-Flour, V-1975, Lecture Notes in Mathematics, Vol. 539, Springer, Berlin, 1976, pp. 167–223.
[25] R. Mainieri, Zeta function for the Lyapunov exponent of a product of random matrices, Phys. Rev. Lett. 68 (13) (1992) 1965–1968.
[26] A. Mukherjea, Topics in products of random matrices, Tata Institute of Fundamental Research Lectures on Mathematics, Vol. 87, Narosa Publishing House, New Delhi, 2000.
[27] C.M. Newman, Lyapunov exponents for some products of random matrices: exact expressions and asymptotic distributions, in: J.E. Cohen, H. Kesten, C.M. Newman (Eds.), Random Matrices and their Applications, Vol. 50, American Mathematical Society, Providence, 1986, pp. 121–141.
[28] V.I. Oseledec, A multiplicative ergodic theorem: Liapunov characteristic numbers for dynamical systems, Trans. Moscow Math. Soc. 19 (1968) 197–231.
[29] Y. Peres, Domains of analytic continuation for the top Lyapunov exponent, Ann. Inst. H. Poincaré Probab. Statist. 28 (1) (1992) 131–148.
[30] M.S. Raghunathan, A proof of Oseledec's multiplicative ergodic theorem, Israel J. Math. 32 (1979) 356–362.
[31] E. Seneta, Nonnegative Matrices and Markov Chains, Springer, Berlin, 1981.
[32] J.M. Steele, Kingman's subadditive ergodic theorem, Ann. Inst. H. Poincaré Probab. Statist. 25 (1) (1989) 93–98.
[33] J.C. Watkins, Limit theorems for products of random matrices, in: J.E. Cohen, H. Kesten, C.M. Newman (Eds.), Random Matrices and their Applications, Vol. 50, American Mathematical Society, Providence, 1986, pp. 5–22.