Computing Integrals, Lyapunov exponents and entropy

Mark Pollicott

mpollic@maths.warwick.ac.uk

Warwick University, England

The aim of the talk

We want to describe an approach to studying Lyapunov exponents for random matrix products.

The aim of the talk

We want to describe an approach to studying Lyapunov exponents for random matrix products.

This not intended to be a method for computing numerical approximations. However, we will give examples when they are more amusing than the proofs.

The aim of the talk

We want to describe an approach to studying Lyapunov exponents for random matrix products.

This not intended to be a method for computing numerical approximations. However, we will give examples when they are more amusing than the proofs.

To motivate some of the ideas we will first consider a somewhat simpler setting (integrals), then progress to the more subtle case (Lyapunov exponents) and eventually dabble a little with entropy.

The aim of the talk

We want to describe an approach to studying Lyapunov exponents for random matrix products.

This not intended to be a method for computing numerical approximations. However, we will give examples when they are more amusing than the proofs.

To motivate some of the ideas we will first consider a somewhat simpler setting (integrals), then progress to the more subtle case (Lyapunov exponents) and eventually dabble a little with entropy.

The underlying method is actually based on some ideas from Thermodynamic Formalism (basically Dynamical Systems, with a dash of Statistical Physics).

The scheme of the talk

- 1. Computing integrals

The scheme of the talk

- 1. Computing integrals
- 2. Computing Lyapunov exponents

The scheme of the talk

- 1. Computing integrals
- 2. Computing Lyapunov exponents
- 3. Some examples

The scheme of the talk

- 1. Computing integrals
- 2. Computing Lyapunov exponents
- 3. Some examples
- 4. The ingredients in the proof

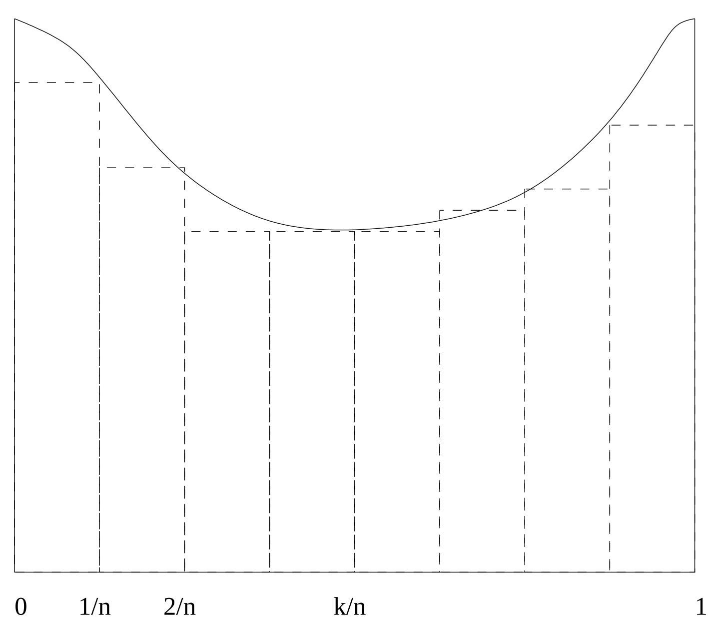
The scheme of the talk

- 1. Computing integrals
- 2. Computing Lyapunov exponents
- 3. Some examples
- 4. The ingredients in the proof
- 5. Computing entropy (rates)

1. Computing integrals

We could try to approximate the integral of an analytic function $f : [0,1] \to \mathbb{R}$ by a discrete summation over a finite set of points. For $n \geq 1$ we can naively write

$$\int_0^1 f(x)\, dx \approx \frac{1}{n} \sum_{k=1}^{n} f\left(\frac{k}{n}\right) .$$



Classical approximations to integrals

There are various classical methods which use different combinations of the same values $f(k/n)$ to give better approximations to the integral (e.g., Newton-Cotes).

For example, the best known is Simpson's rule

$$\frac{1}{3n}\left(f(0) + 2\sum_{k=1}^{n/2-1} f\left(\frac{2k}{n}\right) + 4\sum_{k=1}^{n/2} f\left(\frac{2k-1}{n}\right) + f(1)\right).$$

Classical approximations to integrals

There are various classical methods which use different combinations of the same values $f(k/n)$to give better approximations to the integral (e.g., Newton-Cotes).

For example, the best known is Simpson's rule

$$\frac{1}{3n}\left(f(0)+2\sum_{k=1}^{n/2-1} f\left(\frac{2k}{n}\right)+4\sum_{k=1}^{n/2} f\left(\frac{2k-1}{n}\right)+f(1)\right).$$

For each of these classical methods the error is $O(n^{-k})$ as $n \to \infty$ for some $k > 0$.

Other approximations to integrals

Let us replace $\{k/n : 1 \leq j \leq n\}$ by the following finite sets:

Notation. For each $N \geq 1$ we define a family of points

$$P_N = \left\{ \frac{j}{(2^m - 1)} : 0 \leq j \leq 2^m - 1, 1 \leq m \leq N \right\}.$$

This has cardinality $|P_N| \approx 2^{N+1}$.

Other approximations to integrals

Let us replace $\{k/n : 1 \leq j \leq n\}$ by the following finite sets:

Notation. For each $N \geq 1$ we define a family of points

$$P_N = \left\{ \frac{j}{(2^m - 1)} : 0 \leq j \leq 2^m - 1, 1 \leq m \leq N \right\}.$$

This has cardinality $|P_N| \approx 2^{N+1}$.

Let us consider the following question:

Is there a method for combining the values $\{f(x) : x \in P_N\}$ so as to get a "good" estimate for the integral ?

Some notation

For $n \geq 0$, we define (for no obvious reason)

$$e_n = (-1)^n \frac{2^{-n(n+1)/2}}{\prod_{i=1}^{n}(1-2^{-i})} \, .$$

Some notation

For $n \geq 0$, we define (for no obvious reason)

$$e_n = (-1)^n \frac{2^{-n(n+1)/2}}{\prod_{i=1}^{n}(1-2^{-i})} \, .$$

For each $1 \leq m \leq N$, define

$$\beta_m(f) = \frac{1}{2^m} \sum_{j=0}^{2^m-1} f\left(\frac{j}{2^m-1}\right)$$

Some notation

For $n \geq 0$, we define (for no obvious reason)

$$e_n = (-1)^n \frac{2^{-n(n+1)/2}}{\prod_{i=1}^{n}(1-2^{-i})} .$$

For each $1 \leq m \leq N$, define

$$\beta_m(f) = \frac{1}{2^m} \sum_{j=0}^{2^m-1} f\left(\frac{j}{2^m-1}\right)$$

and

$$\mu_N(f) = \frac{\sum_{m=1}^{N} \frac{e_{N-m}}{1-2^{-m}} \beta_m(f)}{\sum_{m=1}^{N} \frac{e_{N-m}}{1-2^{-m}}} .$$

Some notation

For $n \geq 0$, we define (for no obvious reason)

$$e_n = (-1)^n \frac{2^{-n(n+1)/2}}{\prod_{i=1}^{n}(1-2^{-i})}.$$

For each $1 \leq m \leq N$, define

$$\beta_m(f) = \frac{1}{2^m} \sum_{j=0}^{2^m-1} f\left(\frac{j}{2^m-1}\right)$$

and

$$\mu_N(f) = \frac{\sum_{m=1}^{N} \frac{e_{N-m}}{1-2^{-m}} \beta_m(f)}{\sum_{m=1}^{N} \frac{e_{N-m}}{1-2^{-m}}}.$$

In particular, $\mu_N(f)$ depends only on the values f takes on the points in P_N.

A good approximation

Theorem 0: For any analytic function $f : [0,1] \to \mathbf{R}$,

$$|\int_0^1 f(x)\,dx - \mu_N(f)| = O\left(\exp(-cN^2)\right) \quad \text{as } N \to \infty\,,$$

for some $c > 0$.

A good approximation

Theorem 0: For any analytic function $f : [0,1] \to \mathbf{R}$,

$$|\int_0^1 f(x)\,dx - \mu_N(f)| = O\left(\exp(-cN^2)\right) \quad \text{as } N \to \infty\,,$$

for some $c > 0$.

Since $P_N = 2^{N+1}$ the Theorem tells us that error is super-polynomial in $|P_N|$, i.e., for some $C > 0$,

$$|\int_0^1 f(x)\,dx - \mu_N(f)| = O\left(\exp(-C\log^2|P_N|)\right).$$

A good approximation

Theorem 0: For any analytic function $f : [0,1] \to \mathbf{R}$,

$$|\int_0^1 f(x)\,dx - \mu_N(f)| = O\left(\exp(-cN^2)\right) \quad \text{as } N \to \infty\,,$$

for some $c > 0$.

Since $P_N = 2^{N+1}$ the Theorem tells us that error is super-polynomial in $|P_N|$, i.e., for some $C > 0$,

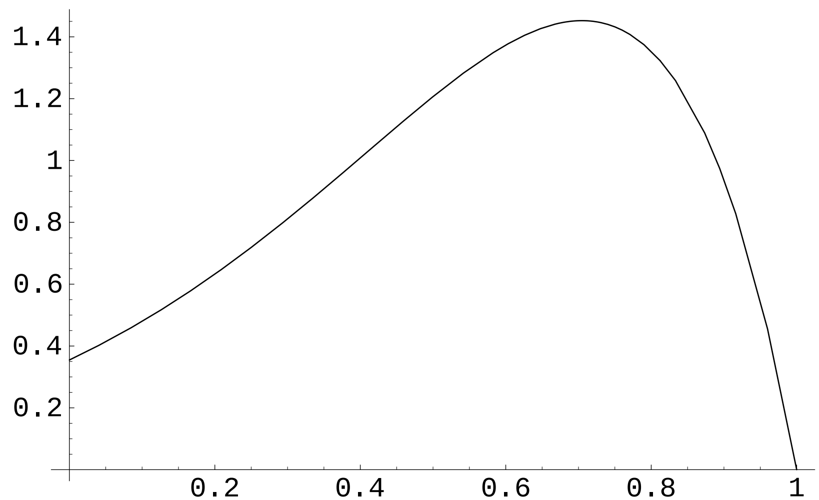
$$|\int_0^1 f(x)\,dx - \mu_N(f)| = O\left(\exp(-C\log^2|P_N|)\right).$$

Aside: The value $c > 0$ can be chosen arbitrarily close to $\log 2$.

An example (in lieu of a proof)

Example: We can apply the theorem to $\int_0^1 f(x)\,dx$ where

$$f(x) = \frac{5\pi}{2}(e^{\pi} - 2)^{-1} e^{\pi x} \cos(\pi x/2)).$$



The integral is actually equal to 1.

Approximations $\mu_N(f)$ to the integral

$N = 1$	$0.185755068918523803464237806443$
$N = 2$	$-0.841124284383205603881801616792$
$N = 3$	$0.406087753333242837879890606782$
$N = 4$	$1.09233276774560006235284301088$
$N = 5$	$0.996815103527958716565339733814$
$N = 6$	$1.00004673478255155995818417282$
$N = 7$	$0.999999773986333380177009909343$
$N = 8$	$0.999999999827806434982440128231$
$N = 9$	$1.00000000000326837809455213367$
$N = 10$	$0.99999999999999360196233983786$
$N = 11$	$1.0000000000000000043615982678651$
$N = 12$	$0.99999999999999999999890263764894006$
$N = 13$	$1.0000000000000000000000000742107229$
$N = 14$	$1.00000000000000000000000000000513617960$
$N = 15$	$0.9999999999999999999999999999999993061845$

3. Lyapunov exponents

Let $\{A_1, A_2\}$ be a fixed pair of 2×2 matrices.

3. Lyapunov exponents

Let $\{A_1, A_2\}$ be a fixed pair of 2×2 matrices.

The Lyapunov exponent λ is given by the limit

$$\lambda := \lim_{n \to +\infty} \frac{1}{n} \sum_{i_1, \cdots, i_n} \frac{1}{2^n} \log ||A_{i_1} \cdots A_{i_n}||$$

3. Lyapunov exponents

Let $\{A_1, A_2\}$ be a fixed pair of 2×2 matrices.

The Lyapunov exponent λ is given by the limit

$$\lambda := \lim_{n \to +\infty} \frac{1}{n} \sum_{i_1, \cdots, i_n} \frac{1}{2^n} \log ||A_{i_1} \cdots A_{i_n}||$$

We could consider more matrices and other measures, but this setting illustrates the results while keeping the notation (relatively) simple.

3. Lyapunov exponents

Let $\{A_1, A_2\}$ be a fixed pair of 2×2 matrices.

The Lyapunov exponent λ is given by the limit

$$\lambda := \lim_{n \to +\infty} \frac{1}{n} \sum_{i_1, \cdots, i_n} \frac{1}{2^n} \log ||A_{i_1} \cdots A_{i_n}||$$

We could consider more matrices and other measures, but this setting illustrates the results while keeping the notation (relatively) simple.

For simplicity, in the examples we shall also assume that both the matrices A_1 and A_2 have determinant 1.

A classical result for typical points

By a famous result of Kesten and Furstenberg one has the following pointwise estimate with repesct to the Bernoulli measure $\mu = (\frac{1}{2}, \frac{1}{2})^{\mathbb{Z}^+}$:

Theorem (Furstenberg-Kesten) For a.e. (μ) $\underline{i} \in \Sigma$ one has

$$\lambda = \lim_{n \to +\infty} \frac{1}{n} \log ||A_{i_1} \cdots A_{i_n}||.$$

A classical result for typical points

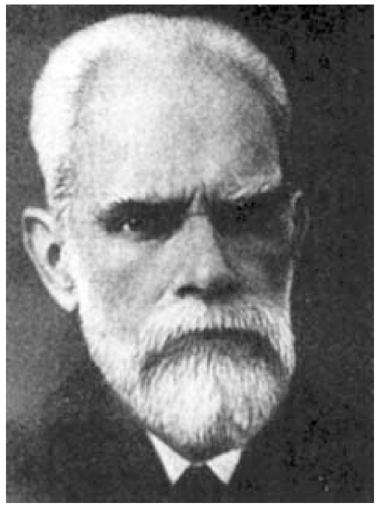
By a famous result of Kesten and Furstenberg one has the following pointwise estimate with repesct to the Bernoulli measure $\mu = (\frac{1}{2}, \frac{1}{2})^{\mathbb{Z}^+}$:

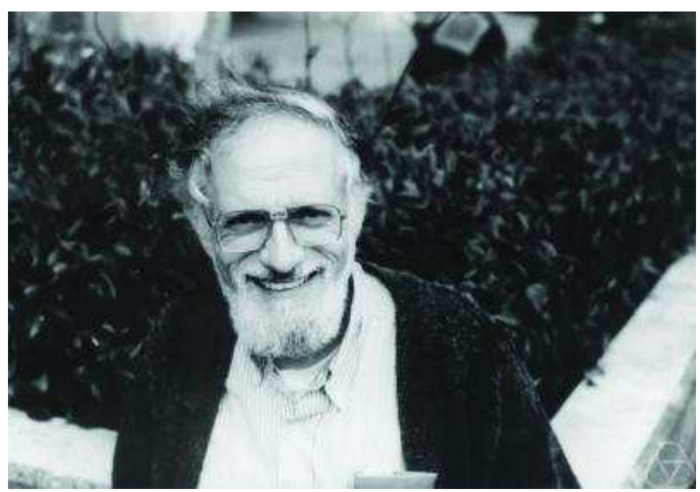
Theorem (Furstenberg-Kesten) For a.e. (μ) $\underline{i} \in \Sigma$ one has

$$\lambda = \lim_{n \to +\infty} \frac{1}{n} \log ||A_{i_1} \cdots A_{i_n}||.$$

Standing assumption: Henceforth, we shall assume that the matrices are all positive (i.e., each of the entries of each of the matrices A_1, A_2 are strictly positive).

Lyapunov, Kesten and Furstenberg





Some notation

Given a finite string $\underline{i} = (i_1, i_2, \cdots, i_n) \in \{1,2\}^n$ let $|\underline{i}| = n$ denote its length and let $\sigma \underline{i} = (i_2, \cdots, i_n, i_1)$ denote a cyclic permutation.

Some notation

Given a finite string $\underline{i} = (i_1, i_2, \cdots, i_n) \in \{1, 2\}^n$ let $|\underline{i}| = n$ denote its length and let $\sigma\underline{i} = (i_2, \cdots, i_n, i_1)$ denote a cyclic permutation.

Let $A_{\underline{i}} = A_{i_1} \cdots A_{i_n}$ denote the product matrix and let $A_{\underline{i}} x_{\underline{i}} = \lambda_{\underline{i}} x_{\underline{i}}$, where $\lambda_{\underline{i}}$ and $x_{\underline{i}}$ are the maximal positive eigenvalue and eigenvector (by the Perron-Frobenius Theorem).

Some notation

Given a finite string $\underline{i} = (i_1, i_2, \cdots, i_n) \in \{1,2\}^n$ let $|\underline{i}| = n$ denote its length and let $\sigma\underline{i} = (i_2, \cdots, i_n, i_1)$ denote a cyclic permutation.

Let $A_{\underline{i}} = A_{i_1} \cdots A_{i_n}$ denote the product matrix and let $A_{\underline{i}} x_{\underline{i}} = \lambda_{\underline{i}} x_{\underline{i}}$, where $\lambda_{\underline{i}}$ and $x_{\underline{i}}$ are the maximal positive eigenvalue and eigenvector (by the Perron-Frobenius Theorem).

For $j = 1, 2$ and $t \in \mathbb{R}$ we define a complex function

$$d^{(j)}(z,t) = \exp\left(-\sum_{n=1}^{\infty} \frac{z^n}{2^n n} \sum_{|\underline{i}|=n} \frac{1}{(1-\lambda_{\underline{i}}^{-2})} \left(\prod_{l=0}^{n-1} \frac{||A_j x_{\sigma^l \underline{i}}||}{||x_{\sigma^l \underline{i}}||} \right)^t \right).$$

which converges to an analytic function for $|z|$ small.

The "explicit" expression

We can deduce the following.

Theorem 1: For each $j = 1, 2$, the function $d^{(j)}(z, t)$ is analytic for $z \in \mathbb{C}$ and $t \in \mathbb{R}$ and we can write

$$\lambda = \frac{\frac{\partial d^{(1)}}{\partial t}(1,0)}{\frac{\partial d^{(1)}}{\partial z}(1,0)} + \frac{\frac{\partial d^{(2)}}{\partial t}(1,0)}{\frac{\partial d^{(2)}}{\partial z}(1,0)}.$$

The "explicit" expression

We can deduce the following.

Theorem 1: For each $j = 1, 2$, the function $d^{(j)}(z,t)$ is analytic for $z \in \mathbb{C}$ and $t \in \mathbb{R}$ and we can write

$$\lambda = \frac{\frac{\partial d^{(1)}}{\partial t}(1,0)}{\frac{\partial d^{(1)}}{\partial z}(1,0)} + \frac{\frac{\partial d^{(2)}}{\partial t}(1,0)}{\frac{\partial d^{(2)}}{\partial z}(1,0)}.$$

Moreover, expanding $d^{(j)}(z,t) = 1 + \sum_{n=1}^{\infty} a_n^{(j)}(t) z^n$

The "explicit" expression

We can deduce the following.

Theorem 1: For each $j = 1, 2$, the function $d^{(j)}(z, t)$ is analytic for $z \in \mathbb{C}$ and $t \in \mathbb{R}$ and we can write

$$\lambda = \frac{\frac{\partial d^{(1)}}{\partial t}(1,0)}{\frac{\partial d^{(1)}}{\partial z}(1,0)} + \frac{\frac{\partial d^{(2)}}{\partial t}(1,0)}{\frac{\partial d^{(2)}}{\partial z}(1,0)}.$$

Moreover, expanding $d^{(j)}(z,t) = 1 + \sum_{n=1}^{\infty} a_n^{(j)}(t) z^n$ we have:
(i) the $a_n^{(j)}$ depend only on the maximal eigenvectors and the eigenvalues of the matrices $A_{\underline{i}}$, with $|\underline{i}| \leq n$; and

The "explicit" expression

We can deduce the following.

Theorem 1: For each $j = 1, 2$, the function $d^{(j)}(z,t)$ is analytic for $z \in \mathbb{C}$ and $t \in \mathbb{R}$ and we can write

$$\lambda = \frac{\frac{\partial d^{(1)}}{\partial t}(1,0)}{\frac{\partial d^{(1)}}{\partial z}(1,0)} + \frac{\frac{\partial d^{(2)}}{\partial t}(1,0)}{\frac{\partial d^{(2)}}{\partial z}(1,0)}.$$

Moreover, expanding $d^{(j)}(z,t) = 1 + \sum_{n=1}^{\infty} a_n^{(j)}(t) z^n$ we have:

(i) the $a_n^{(j)}$ depend only on the maximal eigenvectors and the eigenvalues of the matrices $A_{\underline{i}}$, with $|\underline{i}| \leq n$; and

(ii) we can bound $|a_n^{(j)}| = O\left(\exp\left(-Cn^2\right)\right)$.

Approximating λ

Theorem 2: There are approximations λ_n to λ defined in terms of the maximal eigenvectors and the eigenvalues of the 2^{n+1} matrices $A_{\underline{i}}$, with $|\underline{i}| \leq n$. Moreover, for some $C > 0$

$$|\lambda - \lambda_n| = O\left(\exp\left(-Cn^2\right)\right).$$

Approximating λ

Theorem 2: There are approximations λ_n to λ defined in terms of the maximal eigenvectors and the eigenvalues of the 2^{n+1} matrices $A_{\underline{i}}$, with $|\underline{i}| \leq n$. Moreover, for some $C > 0$

$$|\lambda - \lambda_n| = O\left(\exp\left(-Cn^2\right)\right).$$

Theorem 2 shows that $\lambda_n \to \lambda$ faster then any exponential.

Approximating λ

Theorem 2: There are approximations λ_n to λ defined in terms of the maximal eigenvectors and the eigenvalues of the 2^{n+1} matrices $A_{\underline{i}}$, with $|\underline{i}| \leq n$. Moreover, for some $C > 0$

$$|\lambda - \lambda_n| = O\left(\exp\left(-Cn^2\right)\right).$$

Theorem 2 shows that $\lambda_n \to \lambda$ faster then any exponential.
We can deduce Theorem 2 from Theorem 1, by setting

$$\lambda_n = \frac{\sum_{k=1}^{n} \frac{\partial a_k^{(1)}}{\partial t}(0)}{\sum_{k=1}^{n}(k+1)a_k^{(1)}(0)} + \frac{\sum_{k=1}^{n} \frac{\partial a_k^{(2)}}{\partial t}(0)}{\sum_{k=1}^{n}(k+1)a_n^{(2)}(0)}$$

3. Example 1

If we consider the matrices

$$A_1 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \text{ and } A_2 = \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix}$$

then $\lambda_{n=9}$ this gives an approximation

$$\lambda = 1.1433110351029492458432518536555882994025 \cdots$$

to the Lyapunov exponent λ.

This is presented to the number of decimal places to which it empirically appears to be accurate (i.e., by comparison with the other terms $\{\lambda_n\}$ in the approximation).

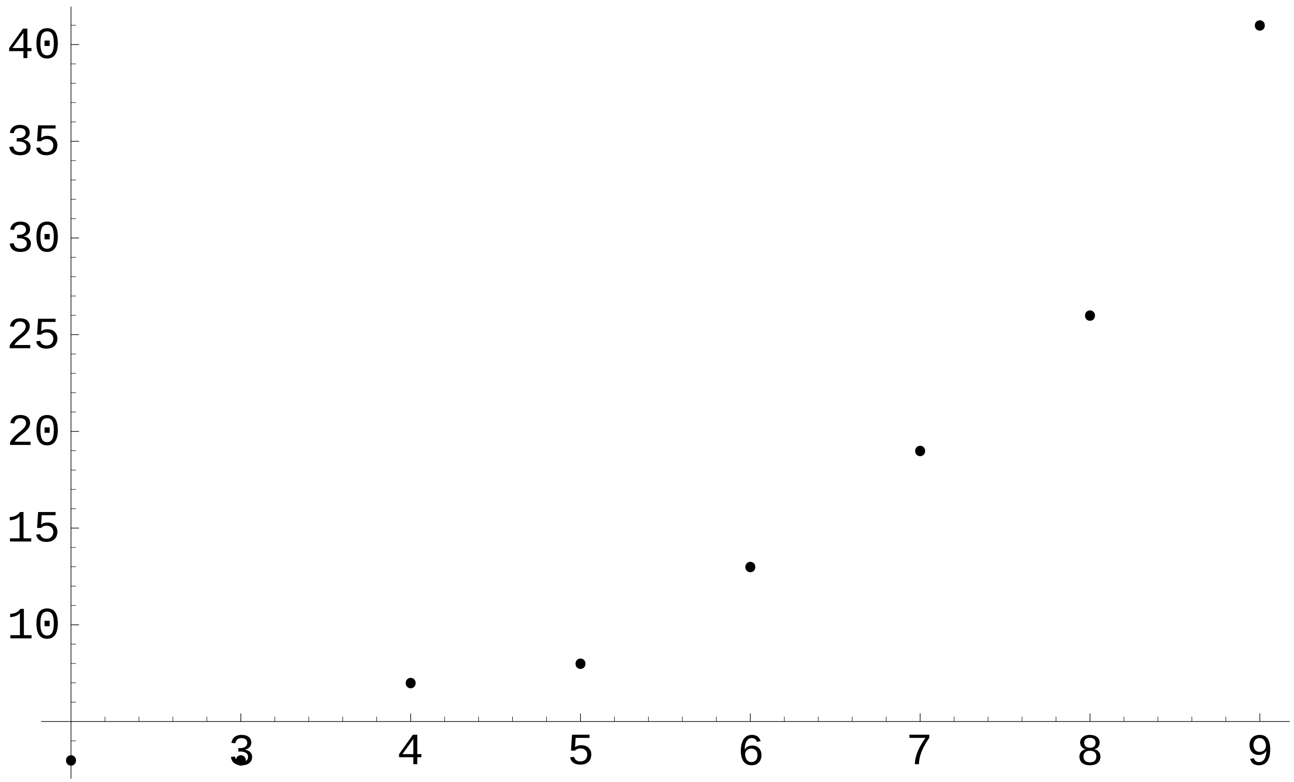
The convergent sequence

Using the complex functions to estimate λ we get :

n	λ_n
1	1.1435949601546489930611282560219921476826
2	1.1432978985074534413937646485571388968329
3	1.1433110787994660471763348416564186089168
4	1.1433110350856466164727559958382071786676
5	1.1433110351029501308232209336496360457362
6	1.1433110351029492458371384694231808633421
7	1.1433110351029492458432518595030145277475
8	1.1433110351029492458432518536555875134112
9	1.1433110351029492458432518536555882994025

For $n = 9$ this is empirically accurate to the 40 decimal places presented.

Super exponential convergence



This plots the number of decimal places in which λ_n and λ_{n+1} agree, against n.

Example 2

Consider the matrices

$$A_1 = \begin{pmatrix} 10 & 1 \\ 9 & 1 \end{pmatrix} \text{ and } A_2 = \begin{pmatrix} 9 & 1 \\ 8 & 1 \end{pmatrix}$$

then the approximations to λ are:

n	λ_n
1	2.341004439933296719772529760910725271776314889929526146521359936127823993879 9437581862818170
2	2.341001098971159429528556454970844691040896647105606075784302295458470606105 2709917152303839
3	2.341001099849930351103674905118261282575383610954371455389408463437561113313 5031699228554668
4	2.341001099849928912869691647574932766880686312348274316277170054896209376480 2328020799059556
5	2.341001099849928912869710477313030753335154640635330757905600135584517519416 8891870156381798
6	2.341001099849928912869710477313028599757800886036333444280974287802196538005 8195559306184339
7	2.341001099849928912869710477313028599757800888289781051642495952865120788671 3598850637666836
8	2.341001099849928912869710477313028599757800888289781051642473759039904012072 4650671884850113
9	2.3410010998499289128697104773130285997578 00888289781051642473759039904012074 5618922420869645

For $n = 9$ this is empirically accurate to the 91 decimal places presented.

Example 3

Consider the matrices

$$A_1 = \begin{pmatrix} 7 & 3 \\ 2 & 1 \end{pmatrix} \text{ and } A_2 = \begin{pmatrix} 10 & 3 \\ 3 & 1 \end{pmatrix}.$$

The approximations to λ are:

n	λ_n
1	2.226794680565867691678454463893275332781235903017721818110653556780474152705 849221790
2	2.226797724235394025007002616723657576800065563173304513358098585220170446096 613205087
3	2.226797722849062871074016189569685470507762455874642623505891442348129450228 371844908
4	2.226797722849068825806764639909071500115833975908048142996879783203444994079 979986570
5	2.226797722849068825806431036394280152446363204985417176642180397999067283381 567867328
6	2.226797722849068825806431036394550829772989705424769806111599542942516596093 789602569
7	2.226797722849068825806431036394550829772986383248925907363709940039787927644 325901024
8	2.226797722849068825806431036394550829772986383248925907993812202447480671254 123242280
9	2.226797722849068825806431036394550829772986383248925907993812202447478802262 891529130

For $n = 9$ this is empirically accurate to the 84 decimal places presented.

4. Overview of the proof

The proof draws on old ideas from the work of Ruelle [1976] (based on even earlier work of Grothendieck [1955]).

Ultimately we need to connect λ to the complex functions $d^{(j)}(z,t)$, which play a rôle roughly akin to characteristic polynomials for matrices.

The role of matrices is replaced by (Ruelle) transfer operators defined on spaces of (analytic) functions.

Strategy of the proof (of Theorem 1)

- Step 1: Consider action of A_1, A_2 on projective space

Strategy of the proof (of Theorem 1)

- Step 1: Consider action of A_1, A_2 on projective space
- Step 2: Associate (transfer) operators and show they are trace class

Strategy of the proof (of Theorem 1)

- Step 1: Consider action of A_1, A_2 on projective space
- Step 2: Associate (transfer) operators and show they are trace class
- Step 3: Relate Lyapunov exponents to transfer operators

Strategy of the proof (of Theorem 1)

- Step 1: Consider action of A_1, A_2 on projective space
- Step 2: Associate (transfer) operators and show they are trace class
- Step 3: Relate Lyapunov exponents to transfer operators
- Step 4: Relate the transfer operators to the complex functions $d^{(j)}(z,t)$ and show that the terms in the expansion of $z \mapsto d^{(j)}(z,t)$ converge quickly.

Strategy of the proof (of Theorem 1)

- Step 1: Consider action of A_1, A_2 on projective space
- Step 2: Associate (transfer) operators and show they are trace class
- Step 3: Relate Lyapunov exponents to transfer operators
- Step 4: Relate the transfer operators to the complex functions $d^{(j)}(z,t)$ and show that the terms in the expansion of $z \mapsto d^{(j)}(z,t)$ converge quickly.

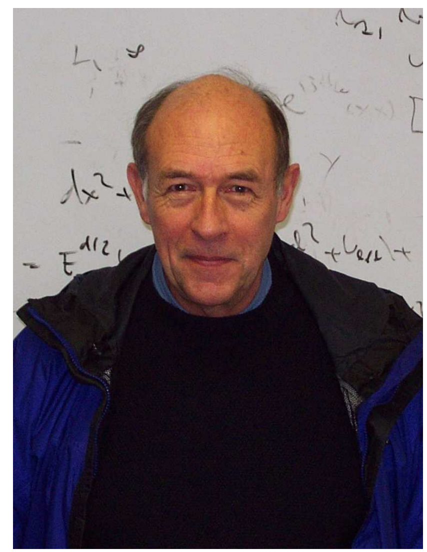
Strategy of the proof (of Theorem 1)

- Step 1: Consider action of A_1, A_2 on projective space
- Step 2: Associate (transfer) operators and show they are trace class
- Step 3: Relate Lyapunov exponents to transfer operators
- Step 4: Relate the transfer operators to the complex functions $d^{(j)}(z,t)$ and show that the terms in the expansion of $z \mapsto d^{(j)}(z,t)$ converge quickly.

Steps 1 and 3 are standard.

Grothendieck and Ruelle

Alexander Grothendieck http://www-history.mcs.st-andrews.ac.uk/PictDisplay/Grothendiec...



1 of 1 3/10/07 06:41

Step 1: Projective actions

Consider the real projective space

$$\mathbb{R}P^2 = (\mathbb{R}^2 - \{(0,0)\})/\sim$$

where $v \sim w$ if there exists $\beta \neq 0$ such that $\beta v = w$.

Step 1: Projective actions

Consider the real projective space

$$\mathbb{R}P^2 = (\mathbb{R}^2 - \{(0,0)\})/\sim$$

where $v \sim w$ if there exists $\beta \neq 0$ such that $\beta v = w$.

It is often convenient to think of $\mathbb{R}P^2$ *in terms of representative vectors* $x \in \mathbb{R}^2$ *with* $||x|| = 1$

Step 1: Projective actions

Consider the real projective space

$$\mathbb{R}P^2 = (\mathbb{R}^2 - \{(0,0)\})/\sim$$

where $v \sim w$ if there exists $\beta \neq 0$ such that $\beta v = w$.

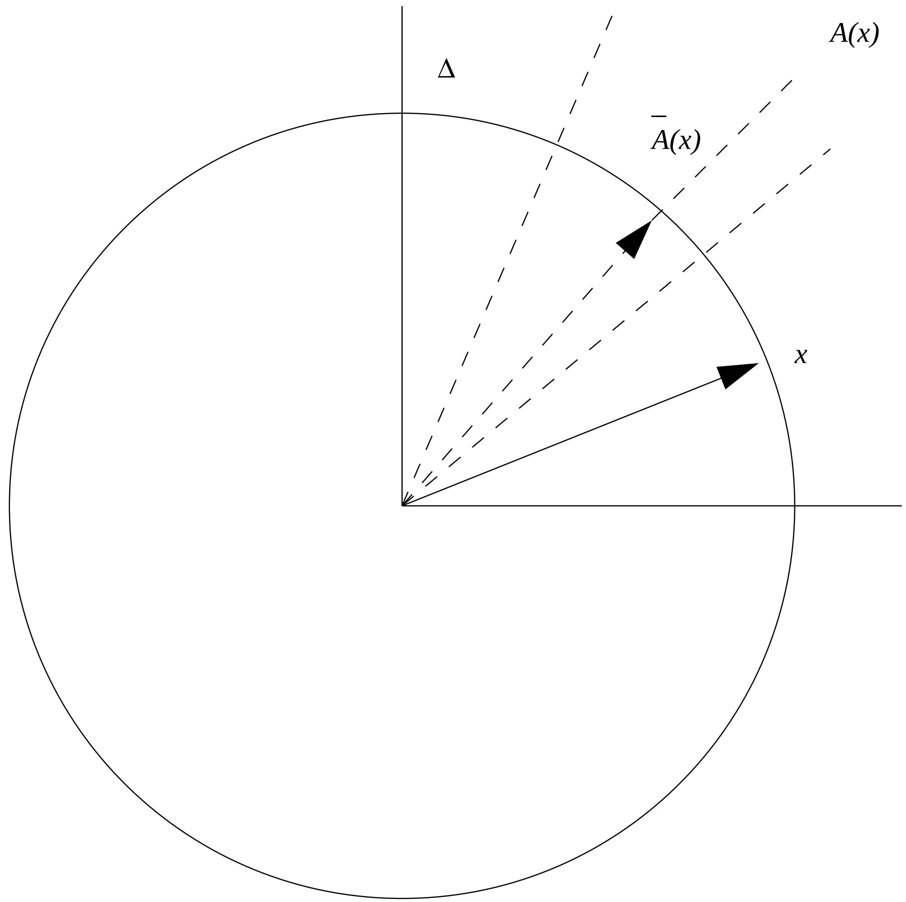
It is often convenient to think of $\mathbb{R}P^2$ *in terms of representative vectors* $x \in \mathbb{R}^2$ *with* $||x|| = 1$

The projective action $\overline{A}_j : \mathbb{R}P^2 \to \mathbb{R}P^2$ takes the form

$$\overline{A}_j(x) = \frac{A_j x}{||A_j x||},$$

for $j = 1, 2$.

Linear action and projective action



Let $\Delta \subset \mathbb{R}P^2$ correspond to the positive quadrant.

The projective actions as contractions

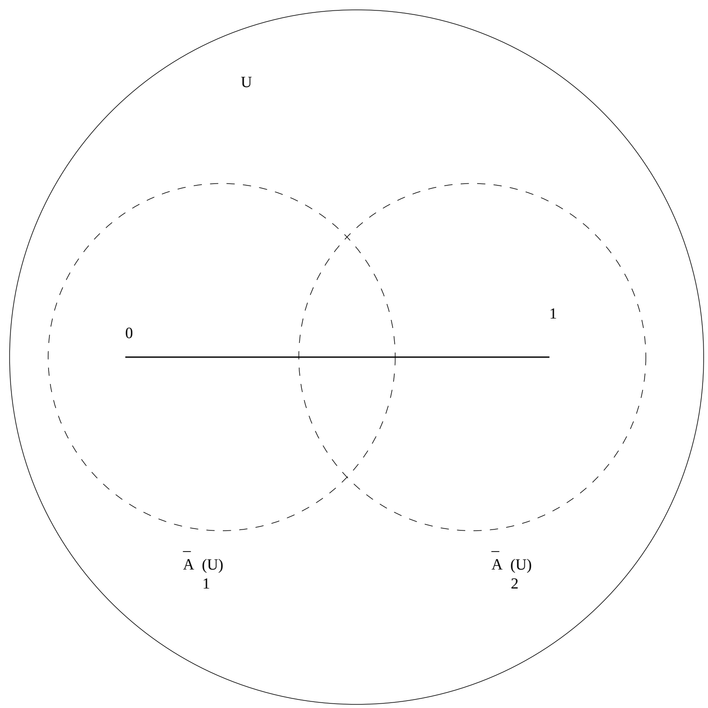
The one-dimensional set Δ can be easily identified with the interval $[0, 1]$, say.

The projective actions as contractions

The one-dimensional set Δ can be easily identified with the interval $[0,1]$, say.
In these coordinates we can choose a neighbourhood $[0,1] \subset U \subset \mathbb{C}$ (the complexification) such that we have $\overline{A}_1(U), \overline{A}_2(U) \subset$ closure(U).

The projective actions as contractions

The one-dimensional set Δ can be easily identified with the interval $[0, 1]$, say.
In these coordinates we can choose a neighbourhood $[0, 1] \subset U \subset \mathbb{C}$ (the complexification) such that we have $\overline{A}_1(U), \overline{A}_2(U) \subset$ closure(U).



Step 2: Ruelle transfer operator

Let $C^\omega(U)$ denote the Hilbert space of analytic functions on U which are square integrable.

Step 2: Ruelle transfer operator

Let $C^\omega(U)$ denote the Hilbert space of analytic functions on U which are square integrable.

We can define a bounded linear operator $\mathcal{L} : C^\omega(U) \to C^\omega(U)$ by

$$\mathcal{L}w(x) = \frac{1}{2}w(\overline{A}_1 x) + \frac{1}{2}w(\overline{A}_2 x).$$

This is a (Ruelle) transfer operator.

Step 2: Ruelle transfer operator

Let $C^\omega(U)$ denote the Hilbert space of analytic functions on U which are square integrable.

We can define a bounded linear operator $\mathcal{L} : C^\omega(U) \to C^\omega(U)$ by

$$\mathcal{L}w(x) = \frac{1}{2}w(\overline{A}_1 x) + \frac{1}{2}w(\overline{A}_2 x).$$

This is a (Ruelle) transfer operator.

Crucially, this operator is trace class (on the analytic functions $C^\omega(U)$). Moreover, each tr$(\mathcal{L}^n)$ can be explicitly expressed in terms of the matrices $A_{\underline{i}}$ with $|\underline{i}| = n$.

Step 3: transfer operators and λ

The maximal eigenvalue 1 for $\mathcal{L}$ has an eigenprojection μ i.e., $\mathcal{L}^*\mu = \mu$.

Step 3: transfer operators and λ

The maximal eigenvalue 1 for $\mathcal{L}$ has an eigenprojection μ i.e., $\mathcal{L}^*\mu = \mu$.
Define functions $f_j : \mathbb{R}P^2 \to \mathbb{R}$ by

$$f_j(x) := \log\left(\frac{||A_j x||}{||x||}\right), \text{ for } j = 1, 2.$$

Step 3: transfer operators and λ

The maximal eigenvalue 1 for $\mathcal{L}$ has an eigenprojection μ i.e., $\mathcal{L}^*\mu = \mu$.
Define functions $f_j : \mathbb{R}P^2 \to \mathbb{R}$ by

$$f_j(x) := \log\left(\frac{||A_j x||}{||x||}\right), \text{ for } j = 1, 2.$$

Lemma (Lyapunov exponent from transfer operator).

$$\lambda = \frac{1}{2}\mu(f_1) + \frac{1}{2}\mu(f_2).$$

("Furstenberg's formula").

Step 4: Determinants

We can define *determinants* by

$$d(z) := \exp\left(-\sum_{n=1}^{\infty}\frac{z^n}{n}\mathrm{tr}(\mathcal{L}^n)\right)$$

Step 4: Determinants

We can define *determinants* by

$$d(z) := \exp\left(-\sum_{n=1}^{\infty} \frac{z^n}{n} \mathrm{tr}(\mathcal{L}^n)\right)$$

Theorem (Grothendieck, Ruelle). We can expand

$$d(z) = 1 + \sum_{n=1}^{\infty} a_n z^n, \text{ where } |a_n| = O(e^{-cn^2}).$$

Step 4: Determinants

We can define *determinants* by

$$d(z) := \exp\left(-\sum_{n=1}^{\infty} \frac{z^n}{n} \mathrm{tr}(\mathcal{L}^n)\right)$$

Theorem (Grothendieck, Ruelle). We can expand

$$d(z) = 1 + \sum_{n=1}^{\infty} a_n z^n, \text{ where } |a_n| = O(e^{-cn^2}).$$

If we replaced $\mathcal{L}$ by a matrix then $d(z)$ would just be a polynomial.

The complex functions $d^{(j)}(z,t)$ have similar properties. The rest of the proof is then routine stuff.

5. Speculation on entropy (rates)

We might like to try to adapt the previous approach to describe the entropy rate of some Hidden Markov Processes.

5. Speculation on entropy (rates)

We might like to try to adapt the previous approach to describe the entropy rate of some Hidden Markov Processes.

Part of the motivation comes from existing connection between entropy rates and Lyapunov exponents.

5. Speculation on entropy (rates)

We might like to try to adapt the previous approach to describe the entropy rate of some Hidden Markov Processes.

Part of the motivation comes from existing connection between entropy rates and Lyapunov exponents.

However, it seems like a better idea to try to approach estimating integrals (with respect to the Blackwell measure)

Two transformations

Fix $\epsilon > 0$ and $0 < p < 1$.

Two transformations

Fix $\epsilon > 0$ and $0 < p < 1$.

Consider the transformations $f_1 : \mathbb{R}^+ \to \mathbb{R}^+$ to $f_2 : \mathbb{R}^+ \to \mathbb{R}^+$ defined by

$$f_1(x) = \frac{(1-\epsilon)}{\epsilon} \frac{px + (1-p)}{(1-p)x + p}$$

and

$$f_2(x) = \frac{\epsilon}{(1-\epsilon)} \frac{(1-p)x + p}{px + (1-p)}.$$

A measure

We can also associate weight functions $r_1 : \mathbb{R}^+ \to \mathbb{R}$ and $r_2 : \mathbb{R}^+ \to \mathbb{R}$ defined by

$$r_1(x) = \frac{((1-\epsilon)p + \epsilon(1-p))x + (1-\epsilon)(1-p) + \epsilon p)}{x+1}$$

$$r_2(x) = \frac{\epsilon p + (1-\epsilon)(1-p))x + \epsilon)(1-p) + (1-\epsilon)p)}{x+1}.$$

A measure

We can also associate weight functions $r_1 : \mathbb{R}^+ \to \mathbb{R}$ and $r_2 : \mathbb{R}^+ \to \mathbb{R}$ defined by

$$r_1(x) = \frac{((1-\epsilon)p + \epsilon(1-p))x + (1-\epsilon)(1-p) + \epsilon p)}{x+1}$$

$$r_2(x) = \frac{\epsilon p + (1-\epsilon)(1-p))x + \epsilon)(1-p) + (1-\epsilon)p)}{x+1}.$$

Definition. We can define the (Blackwell) probability measure μ by the identity

$$\mu(A) = \int_{f_1^{-1}A} r_1(x) d\mu(x) + \int_{f_2^{-1}A} r_2(x) d\mu(x)$$

HMP and the entropy rate

Consider a binary symmetric channel with crossover probability $\epsilon > 0$ and input Markov chain with transition matrix

$$\begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}.$$

Theorem (Han-Marcus). The entropy rate is given by the following:

$$H = -\int (r_1(x) \log r_1(x) + r_2(x) \log r_2(x)) \, d\mu(x)$$

HMP and the entropy rate

Consider a binary symmetric channel with crossover probability $\epsilon > 0$ and input Markov chain with transition matrix

$$\begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}.$$

Theorem (Han-Marcus). The entropy rate is given by the following:

$$H = -\int (r_1(x) \log r_1(x) + r_2(x) \log r_2(x)) \, d\mu(x)$$

This identity has an explicit integrand and a mysterious measure. But one can hope to try estimate this integral using the same method as before.

A transfer operator

We can consider the associate transfer operator $\mathcal{L}$ defined on analytic functions on $\mathbb{R}^+$ by

$$\mathcal{L}w(x) = w(f_1 x) r_1(x) + w(f_2 x) r_2(x).$$

A transfer operator

We can consider the associate transfer operator $\mathcal{L}$ defined on analytic functions on $\mathbb{R}^+$ by

$$\mathcal{L}w(x) = w(f_1 x) r_1(x) + w(f_2 x) r_2(x).$$

By analogy with the approach for Lyapunov exponents, we might try to estimate the integral for the entropy rate using complex functions defined in terms of the fixed points

$$f_{i_1} \cdots f_{i_n} x = x$$

of the maps $f_{i_1} \cdots f_{i_n} : \mathbb{R}^+ \to \mathbb{R}^+$, where $i_1, \cdots, i_n \in \{1, 2\}$ and $n \geq 1$, and weights

$$r_{i_1}(x) r_{i_2}(f_{i_1} x) \cdots r_{i_n}(f_{i_{n-1}} \cdots f_{i_1} x).$$

Example (in lieu of a theorem or proof)

Consider the specific case of $\epsilon = 0.01$ and $p = 0.3$.
We get the following approximations to the entropy rate H:

n	nth approximation to H
1	0.6193020462238054585978760034752462 4771
2	0.6188019801176990362221728543989839 4068
3	0.6188021214839795457364547350024867 8392
4	0.6188021162904104147644028303257558 1772
5	0.6188021162904437719741107127221680 4499
6	0.6188021162904437719770974068510844 2687
7	0.6188021162904437719770974068578338 9657