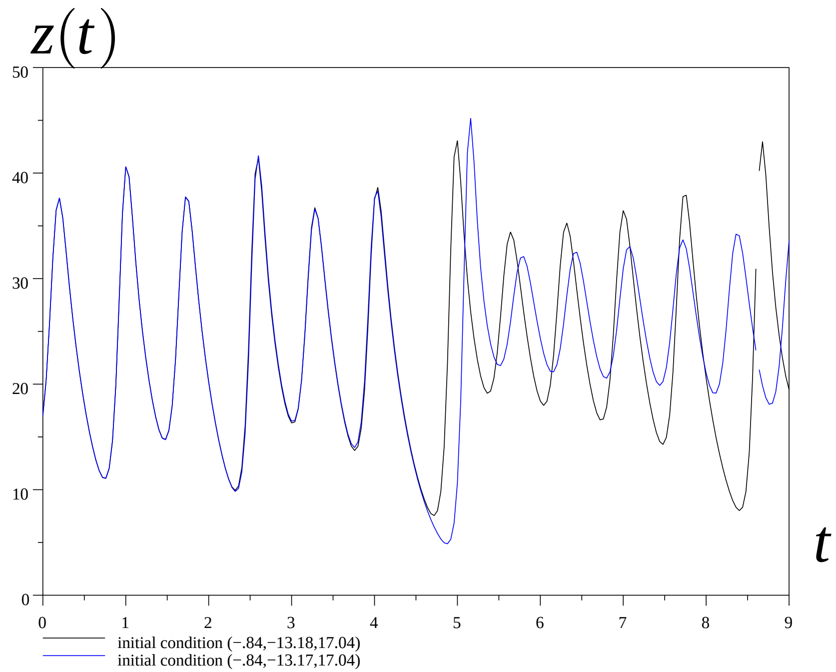
SDIC and Lyapunov exponents.

One the two ingredients for chaotic behaviour is the sensitive dependence to initial conditions (SDIC).
Two nearby initial conditions in generic position give rise to trajectories which separate with time. In other words, small (generic) errors are amplified. This can be illustrated on the Lorenz dynamical system.



This is the well known butterfly effect of E.Lorenz.

The second ingredient for chaos is the boundedness of phase space. If two trajectories confined in a bounded domain separate in time, their distance should reach the diameter of the domain. By that time the nonlinearities are not negligible anymore and the trajectories are folded. Eventually they may come back near to one another at a later time. We will come back to this global effect later on.

This argument does not work for linear time evolution that does not lead to chaotic behaviour and will not be considered anymore.

How to quantify SDIC (the amplification of small errors?)

Let x_0 and y_0 be two nearby initial conditions: $y_0 = x_0 + \vec{h}$, $\vec{h}$ small.
By Taylor's formula we have

$$y_1 = T\left(y_0\right) = T\left(x_0 + \vec{h}\right) = T\left(x_0\right) + DT_{x_0}\vec{h} + \mathcal{O}\left(\vec{h}^2\right)$$

where DT_x is the differential of T at x, namely the matrix of partial derivatives.

More generally, as long as the two trajectories stay nearby, *i.e.* up to a time n not too large, one can write

$$y_n = T^n(y_0) = T^n(x_0 + \vec{h}) = x_n + DT^n_{x_0}\vec{h} + \mathscr{O}\left(\vec{h}^2\right) .$$

The initial small error $\vec{h}$ is amplified by the “factor” $DT^n_{x_0}$ and we will study its growth in n for n large.

One should be careful here about the order of limits.
One first let $\vec{h}$ tend to 0 then n to infinity.
The two trajectories must stay close at least up to time n, otherwise Taylor’s formula may not apply anymore (and the result may be very different).

We will now discuss the behaviour of $DT^n_{x_0}$ in several cases of increasing generality (and complexity).

First (easy) case: dimension one, the phase space is an interval ([-1,1]).

$$f^n(x+h) = f^n(x) + f^{n\prime}(x)h + \mathscr{O}(h^2) \; .$$

By the chain rule ($f^n = f \circ f \circ \cdots \circ f$, n times), we have

$$f^{n\prime}(x) = \prod_{j=0}^{n-1} f'\big(f^j(x)\big) \; .$$

This naturally suggests an exponential growth in n, and we look for the exponential growth rate per unit of time:

$$\frac{1}{n} \log\big|f^{n\prime}(x)\big| = \frac{1}{n} \sum_{j=0}^{n-1} \log\big|f'\big(f^j(x)\big)\big| \; .$$

We see appearing a temporal average and we can apply Birkhoff's ergodic Theorem.

Let μ be an ergodic invariant measure such that the function $\log|f'|$ has an integrable modulus. Then except on a set of μ measure zero, we have convergence of the temporal average, and moreover

$$\lim_{n\to\infty} \frac{1}{n} \log\left|f^{n\prime}(x)\right| = \int \log\left|f'(\cdot)\right| d\mu \ .$$

This number is called the Lyapunov exponent of the measure μ for the transformation f.

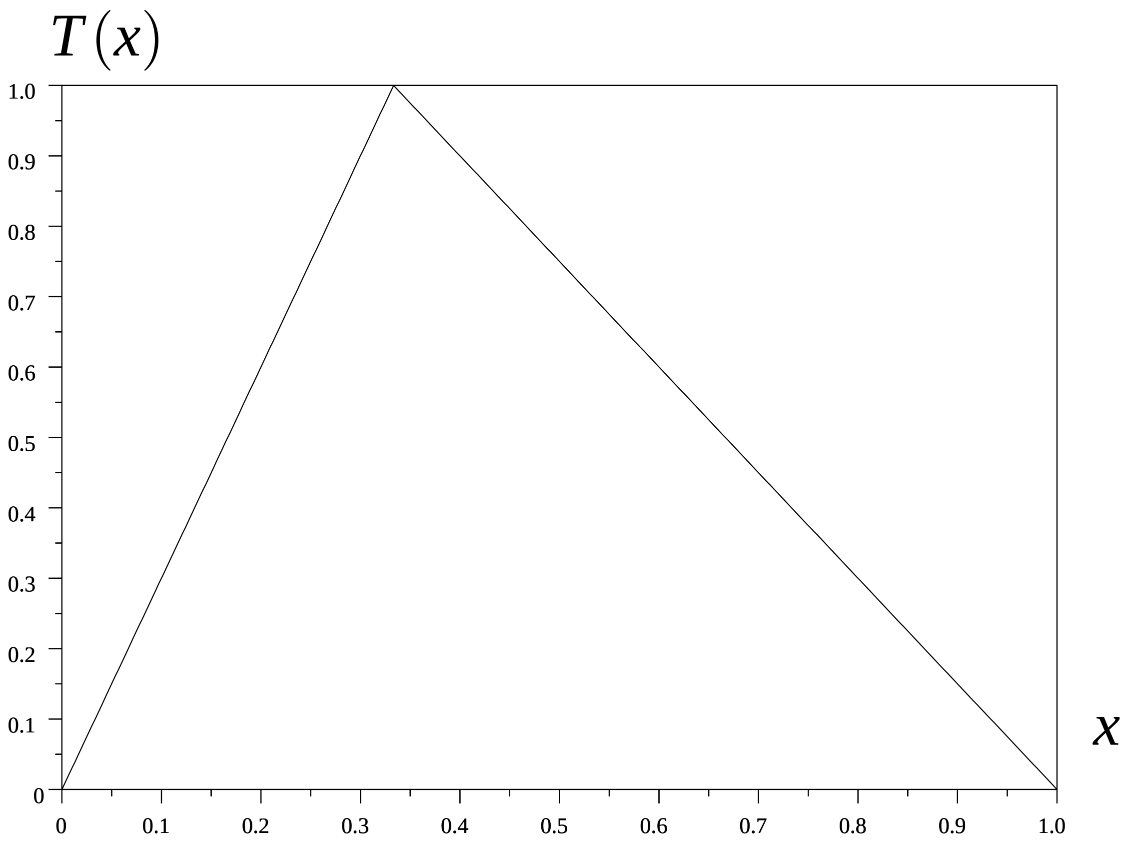
Here one should stress again the importance of the initial condition.

There are many initial conditions for which the limit does not exist.

For many other initial conditions, the limit exists, but take another value. For example for the initial conditions typical of a different ergodic measure.

Example : $\Omega = [0,1]$

$$T(x) = \begin{cases} 3x & \text{if } 0 \le x < 1/3 \\ 3(1-x)/2 & \text{if } 1/3 \le x \le 1. \end{cases}$$



$\Omega = [0, 1]$

$$T(x) = \begin{cases} 3x & \text{if } 0 \le x < 1/3 \\ 3(1-x)/2 & \text{if } 1/3 \le x \le 1. \end{cases}$$

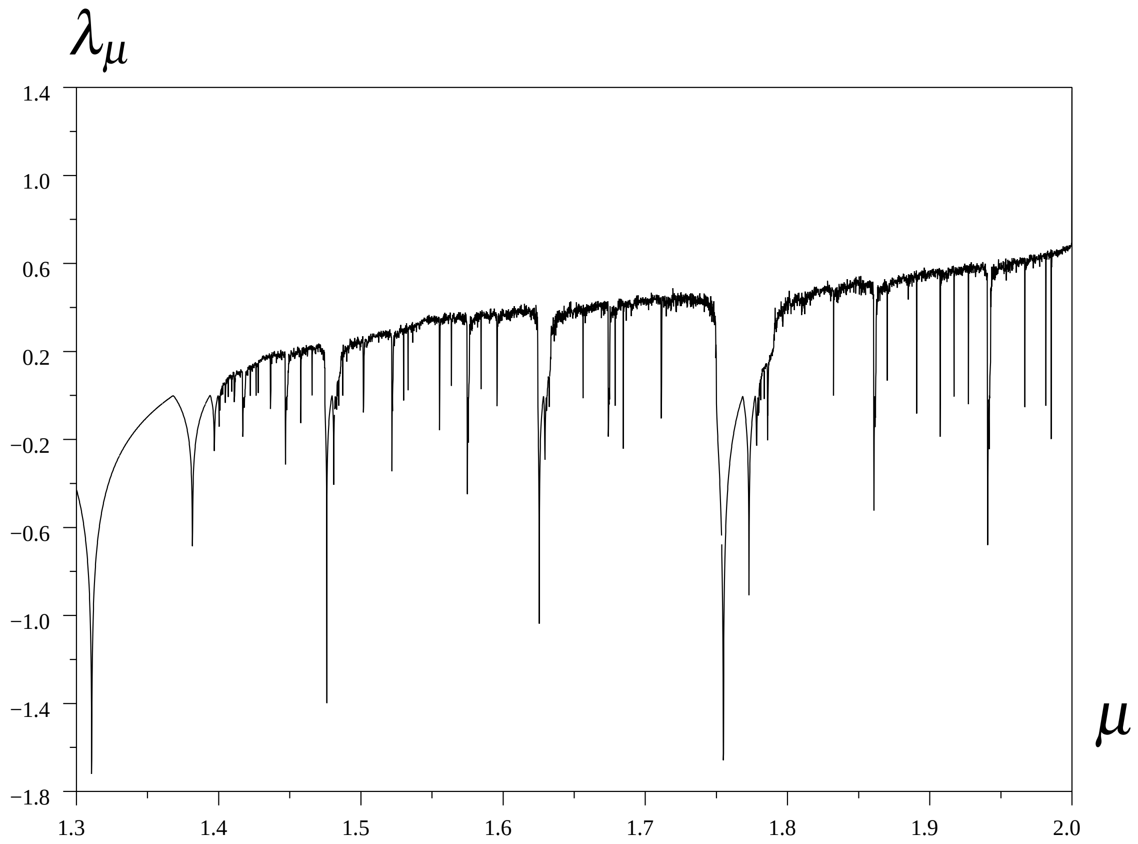
The Lebesgue measure is T invariant and ergodic. Its Lyapunov exponent is given by

$$\frac{1}{3} \log 3 + \frac{2}{3} \log(3/2) \ .$$

The transformation T has a fixed point $x = 3/5$. The Dirac measure in this point is also invariant and ergodic. Its Lyapunov exponent is $\log(3/2)$.

There is an uncountable set of invariant ergodic measures, all with different Lyapunov exponents (taking all the values between $\log 3/2$ and $\log 3$).

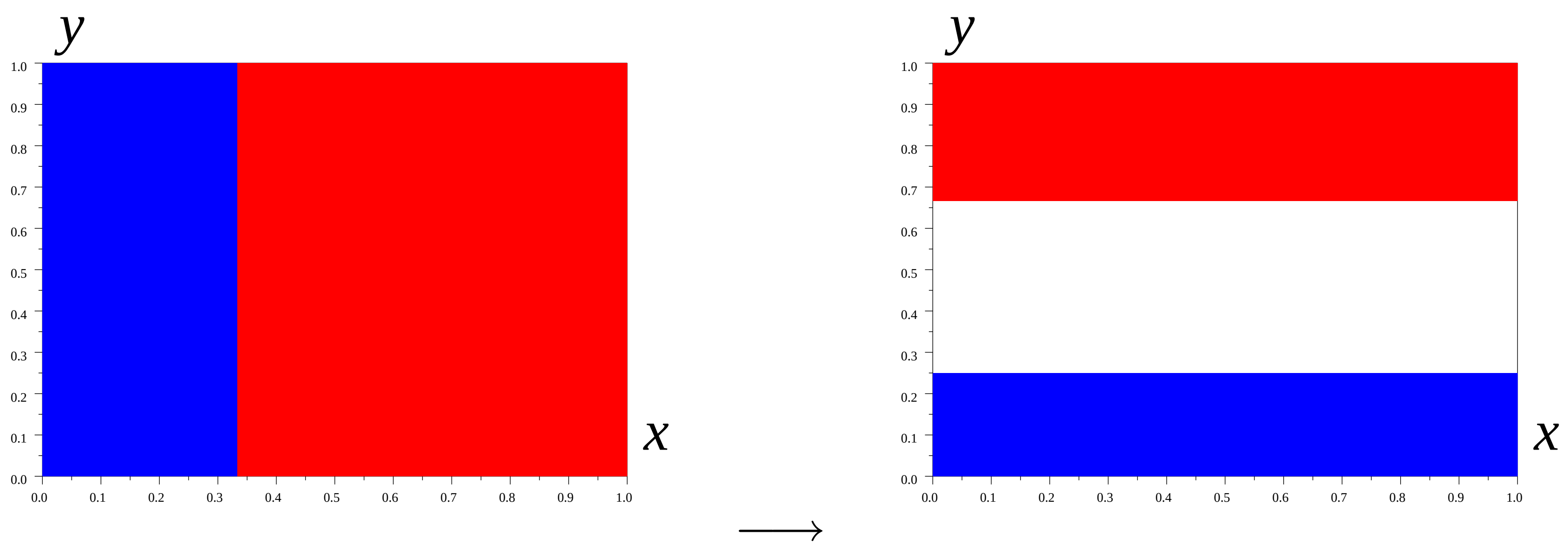
Lyapunov exponents depend in general of the parameters. For example for the one parameter family of maps $x \to 1 - \mu x^2$ (with phase space $[-1,1]$ and $0 \leq \mu \leq 2$)



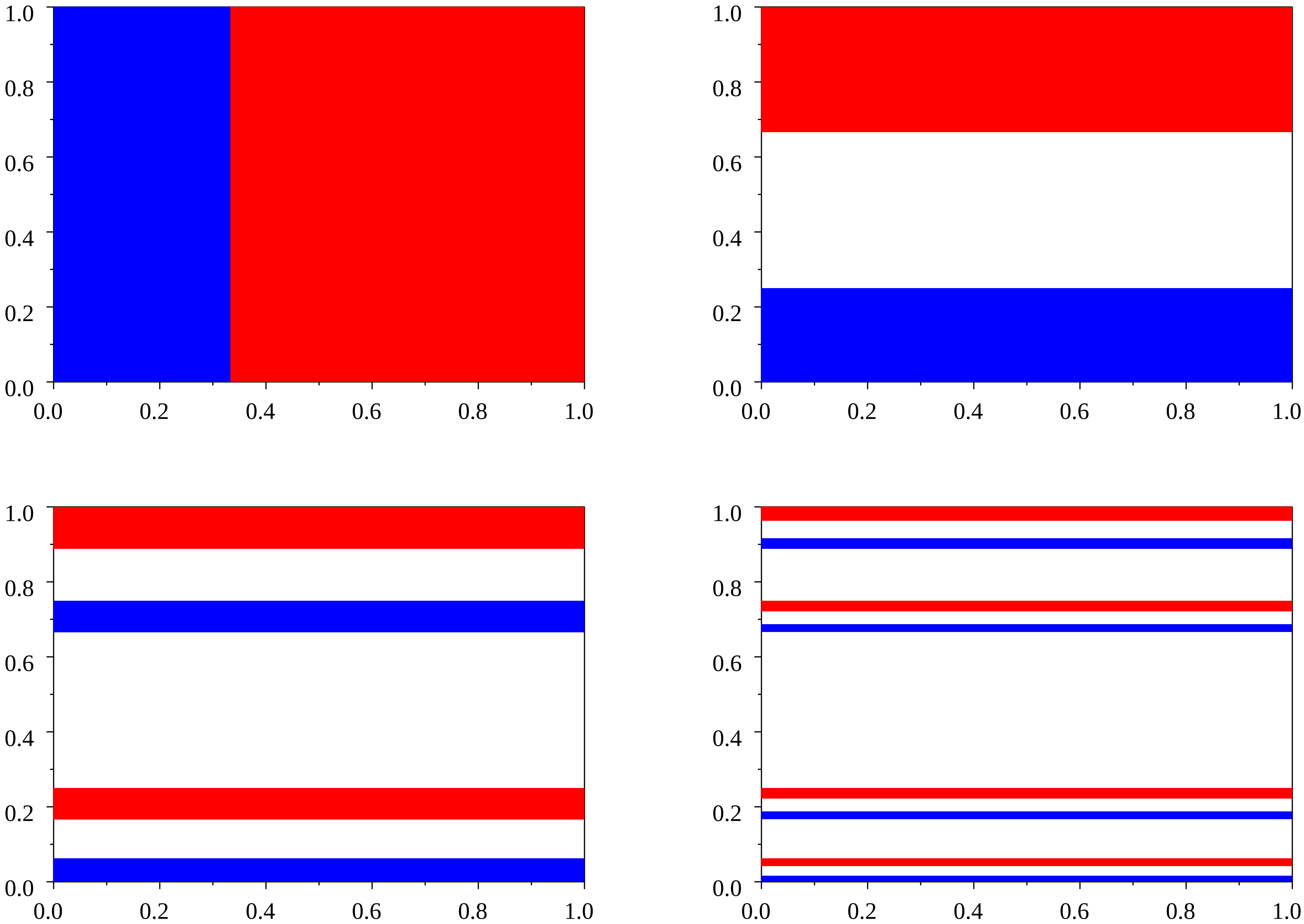
Next level of difficulty: dimension larger than one.

A simple example: the dissipative baker's map. The phase space is the unit square $[0,1]\times[0,1]$. The map is given by

$$T(x,y) = \begin{cases} (3x, y/4) & \text{if } 0 \le x < 1/3 \\ (3(x-1/3)/2, (2+y)/3) & \text{if } 1/3 \le x \le 1. \end{cases}$$



The attractor is a strange set: product of a segment by a Cantor set. The SRB measure is the product of the Lebesgue measure along the segment by a Cantor measure.


1.0
0.8
0.6
0.4
0.2
0.0
0.0 0.2 0.4 0.6 0.8 1.0
1.0
0.8
0.6
0.4
0.2
0.0
0.0 0.2 0.4 0.6 0.8 1.0
1.0
0.8
0.6
0.4
0.2
0.0
0.0 0.2 0.4 0.6 0.8 1.0
1.0
0.8
0.6
0.4
0.2
0.0
0.0 0.2 0.4 0.6 0.8 1.0

The differential is given by

$$DT_{(x,y)} = \begin{cases} \begin{pmatrix} 3 & 0 \\ 0 & 1/4 \end{pmatrix} & \text{if } 0 \le x < 1/3 \\ \begin{pmatrix} 3/2 & 0 \\ 0 & -1/3 \end{pmatrix} & \text{if } 1/3 \le x \le 1. \end{cases}$$

To estimate $DT^n_{(x,y)}$ we now have to perform a product of matrices. If we start from the initial point (x,y), with an initial error $\vec{h}$, we have after n iteration steps (using the chain rule) an error given by

$$DT^n_{(x,y)}\vec{h} = DT_{T^{n-1}(x,y)} DT_{T^{n-2}(x,y)} \cdots DT_{(x,y)}\vec{h} \,.$$

In general matrices do not commute (hence one should be careful with the order in the product). However here they commute (and they are even diagonal).

Therefore, to obtain the product matrix, it is enough to take the product of the diagonal elements:

$$DT_{T^{n-1}(x,y)} DT_{T^{n-2}(x,y)} \cdots DT_{(x,y)} = \begin{pmatrix} \prod_{j=0}^{n-1} u\big(T^j(x,y)\big) & 0 \\ 0 & \prod_{j=0}^{n-1} v\big(T^j(x,y)\big) \end{pmatrix}$$

where

$$u(x,y) = 3\chi_{[0,1/3]}(x) + (3/2)\chi_{[1/3,1]}(x)$$

and

$$v(x,y) = (1/4)\chi_{[0,1/3]}(x) - (1/3)\chi_{[1/3,1]}(x)\ .$$

We can now take the $\log$ of the absolute value of each diagonal entry of the product and apply Birkhoff's ergodic theorem as in the one dimensional case.

Since the functions u and v do not depend on y, the integral with respect to the SRB measure reduces to the integration with respect to the one dimensional Lebesgue measure. Therefore, for Lebesgue almost any (x,y) (two dimensional Lebesgue measure, recall the definition of SRB)

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \log \left| u\left(T^j(x,y)\right|\right. = (1/3)\log 3 + (2/3)\log(3/2)$$

$$= \log\left((27/4)^{1/3}\right).$$

Similarly

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \log \left| v\left(T^j(x,y)\right|\right. = -(1/3)\log 4 - (2/3)\log(3)$$

$$= \log\left((36)^{-1/3}\right).$$

Conclusion :

$$DT_{T^{n-1}(x,y)} DT_{T^{n-2}(x,y)} \cdots DT_{(x,y)} \approx \begin{pmatrix} (27/4)^{n/3} & 0 \\ 0 & (36)^{-n/3} \end{pmatrix}.$$

We now have to interpret this result.

If we take a vector $\vec{h} = \left(h_1, h_2\right)$, such that $h_1 \neq 0$, we see that the first component of $DT^n_{(x,y)}\vec{h}$ grows exponentially fast with growth rate $\log\left((27/4)^{1/3}\right)$, and this component dominates the other one which decreases exponentially fast at the rate $\log\left(36^{-1/3}\right)$ (< 0).

In other words, almost any error grows exponentially fast with rate $\log\left((27/4)^{1/3}\right)$, this is the maximal Lyapunov exponent.

But there is another Lyapunov exponent $\log\left(36^{-1/3}\right)$ (< 0) corresponding to special initial errors satisfying $h_1 = 0$. These errors do not grow but decay exponentially fast.

This is similar to the diagonalization of matrices but the interpretation is slightly more involved.

We have to distinguish two subspaces. The first one is the entire space $E_0 = \mathbb{R}^2$. The second one is the subspace of codimension one $E_1 = \{(0, h_2)\} \subset E_0$.
If $\vec{h} \in E_0 \backslash E_1$ the initial error grows exponentially fast with rate $\log\left(\left(27/4\right)^{1/3}\right)$.

If $\vec{h} \in E_1$ the initial error decreases exponentially fast with rate $\log\left(36^{-1/3}\right)$.

Positive Lyapunov exponents are obviously responsible for the Sensitive Dependence on Initial Conditions. Their corresponding "eigen" directions are tangent to the attractor.

Transversally to the attractor one gets contracting directions, namely negative Lyapunov exponents. In some cases the attractor can also contain directions of negative Lyapunov exponents.

This is a local picture, we will discuss later on a more global picture.

The general case combines the two preceding ideas (product of matrices and ergodic theorem) together with a new fact: the subspaces E_0, E_1, etc. depend on the initial condition: they vary from point to point.

Moreover the matrices appearing in the product do not commute (be careful with the order).

One has to use a more sophisticated version of the ergodic theorem called the sub-additive ergodic theorem.

We are only interested in the size of the vector $DT^n_{x_0}\vec{h}$. It is convenient to look at the size of its square norm

$$\left\langle DT^n_{x_0}\vec{h}\middle|DT^n_{x_0}\vec{h}\right\rangle = \left\langle\vec{h}\middle|\left(DT^n_{x_0}\right)^t DT^n_{x_0}\vec{h}\right\rangle .$$

The result can be formulated in terms of the asymptotic behaviour of the matrix

$$\left(DT^n_{x_0}\right)^t DT^n_{x_0} .$$

The Oseledec theorem.

Let μ be an ergodic invariant measure. For μ almost any initial condition x_0, the sequence of (positive) matrices

$$\left(\left(DT^n_{x_0}\right)^t DT^n_{x_0}\right)^{1/2n}$$

converges to a matrix Λ symmetrical and positive.

The logarithms of the (different) eigenvalues of Λ denoted by $\lambda_0 > \lambda_1 > \cdots > \lambda_k$ are called the Lyapunov exponents (Lyapunov spectrum). Of course some eigenvalues can occur with multiplicity larger than one.

The Lyapunov exponents are invariant by (invertible and regular) changes of variables.

Second part of The Oseledec theorem.
Except on a set of points x in the phase space of μ measure zero, there exists a decreasing sequence of sub vector spaces

$$E_0(x) \supsetneq E_1(x) \supsetneq \cdots \supsetneq E_k(x)$$

with the following properties.

$E_0(x)$ is the entire space.

$$DT_x E_m(x) = E_m\big(T(x)\big)$$

If $\vec{h} \in E_m(x) \backslash E_{m+1}(x)$ then

$$\lim_{n\to\infty} \frac{1}{n} \log \left\| DT_x^n \vec{h} \right\| = \lambda_m \; .$$

This looks somewhat like the decomposition in Jordan blocks.

Remarks

The theorem says

$$\left\| DT_x^n \vec{h} \right\| \approx e^{n\lambda_m}$$

only in a logarithmic sense. There can be a very large (or very small) sub-exponential prefactor which depends on n, x and $\vec{h}$. These prefactors play an important role in some finer questions.

The rate of convergence is in general unknown.

How to determine the Lyapunov exponents?

Three difficulties.

- Compute $\left(M_n^t M_n\right)^{1/2n}$ with

$$M_n = A_n A_{n-1} \cdots A_1$$

where the matrices A_m do not commute.

- If the phase space is known but not the transformation, one has to estimate $A_n = DT_{x_{n-1}}$.

- Reconstruct the phase space if it is unknown.

We will deal with these problems one after the other, the last one in the next lecture.

There are also several numerical difficulties. One would like to take n as large as possible to get a better statistics. However, when performing the multiplication of the matrices $A_n A_{n-1} \cdots A_1$ one gets exponentially growing quantities. This problem is solved by summing logarithms, not by computing products.

In general the data consists of only one (hopefully typical) trajectory of given fixed length.

Some experimental or numerical noise is often present.

A simpler problem is to determine the largest exponent (test of chaos: SDIC).

Two generic, close, initial conditions separate exponentially fast with this exponential rate.

A concrete algorithm consists in fixing a small number $\varepsilon > 0$ and to compute for τ integer the quantity

$$L_N(\tau, \varepsilon, x) =$$

$$\frac{1}{N-\tau-1}\sum_{j=0}^{N-\tau}\log\left(\frac{1}{\left|\mathscr{U}_\varepsilon\left(T^j(x)\right)\right|}\sum_{y\in\mathscr{U}_\varepsilon\left(T^j(x)\right)}\|T^{j+\tau}(x)-T^\tau(y)\|\right)$$

where $\mathscr{U}_\varepsilon\left(T^j(x)\right)$ is the set of points of the orbit of x at a distance less than ε of $T^j(x)$.

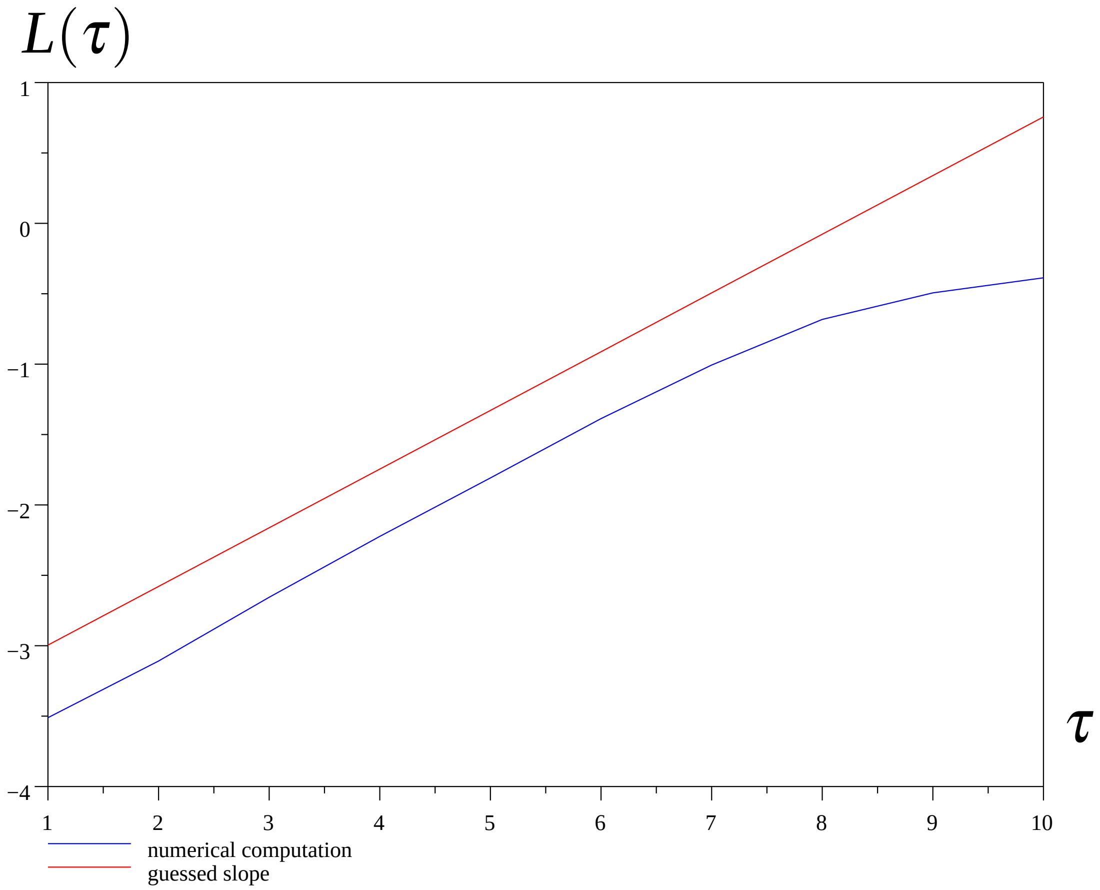
The idea is that each term is of order $\exp(\lambda_{\max}\tau)$. For τ large enough, one expects a linear growth.

But this linear growth should saturate when $\varepsilon \exp(\lambda_{\max}\tau)$ is of the order of the size of the attractor.

The summation over the neighboring points should eliminate the fluctuations.

One plots the curve $L(\tau)$ for various values of ε to see if there is a zone with constant slope.

Maximum exponent for the Hénon map, original values of the parameters.



How to compute the logarithm of the eigenvalues of $\left(M_n^t M_n\right)^{1/2n}$.

The first problem is not to loose too much precision and computer time for large n. The problem gets even more difficult if there are many exponents.

There are several efficient algorithms to compute $\left(M_n^t M_n\right)$.

One is based on the **QR** method.

Any (real) matrix can be written as a product QR with Q orthogonal (and real) and R upper triangular (and real).

If $M_n = Q_n R_n$, there are two real orthogonal matrices U_n et V_n such that $R_n = U_n D_n V_n$ with D_n a real diagonal matrix with the same diagonal elements as R_n. (This gives the singular value decomposition of M_n).

Then $M_n^t M_n = V_n^t D_n^2 V_n$, hence

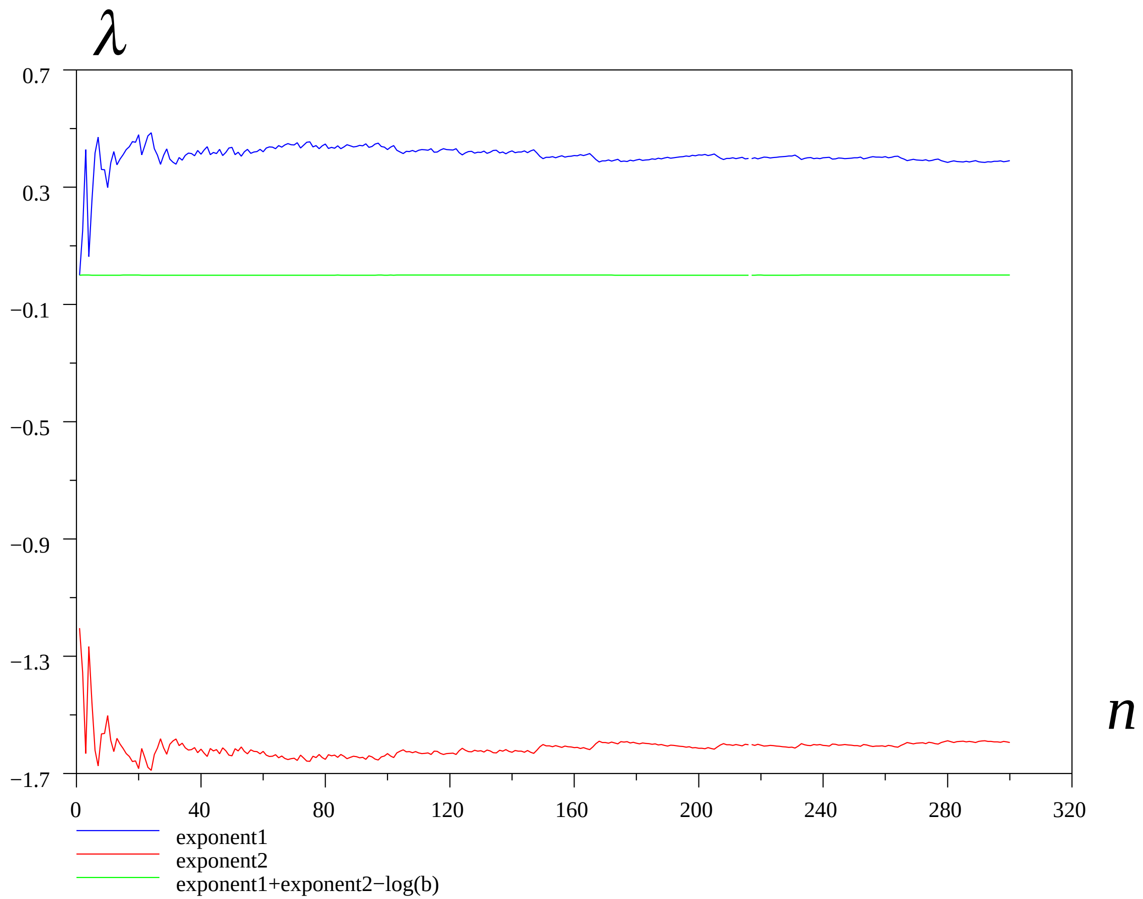
$$\left(M_n^t M_n\right)^{1/2n} = V_n^t D_n^{2/2n} V_n = V_n^t D_n^{1/n} V_n \ .$$

The Lyapunov exponents are (approximated by) the logarithms of the diagonal elements of D_n divided by n.

From a computational point of view, $M_{n+1} = A_{n+1} M_n = A_{n+1} Q_n R_n$. One performs the **QR** decomposition of the matrix $A_{n+1} Q_n = Q_{n+1} W_n$ and then $R_{n+1} = W_n R_n$.

In practice, it is enough to keep only the orthogonal matrix Q_n and to accumulate the logarithms of the diagonal elements of W_n.

Example of the Hénon map (original values).



Estimation of DT_x.

Quite often one does not know DT_x explicitly. A notable exception is in numerical computations. Therefore one has to determine this matrix form the data.

An experimental problem (and a little bit numerical also) is that the data are noisy. In other words, instead of having the sequence of points $x_1, x_2, \cdots, x_n$ from an orbit in phase space, (namely $T(x_j) = x_{j+1}$), one has measured

$$y_j = x_j + \varepsilon_j \ ,$$

ε_j small (hopefully), random (not always).

In general one assumes that the noises (ε_j) form a stationary sequence of independent Gaussian random variables. It is always a good idea to try to verify these hypothesis using some statistical tests.

To get an approximation of DT_x one uses Taylor's formula. Consider a small neighborhood $\mathscr{U}_x$ of x (often a ball).

If a point x_l of an orbit falls in $\mathscr{U}_x$, we have by the Taylor formula

$$T(x_l) = T(x) + DT_x\big(x_l - x\big) + \mathscr{O}\big((x_l - x)^2\big)\ .$$

If we neglect the quadratic correction, the transformation T is affine in $\mathscr{U}_x$ (a translation plus a linear transformation). In dimension d it is therefore given by $d^2 + d$ parameters ($T(x)$ and the matrix DT_x). It suffices to have enough points in the neighborhood $\mathscr{U}_x$ to determine these parameters.

In the presence of noise

$$T(x_l) = T(x) + DT_x\big(x_l - x\big) - \varepsilon_l + \mathscr{O}\big((x_l - x)^2\big) \ .$$

If we neglect the quadratic term in Taylor's formula, we have to determine an affine transformation in the presence of noise. This is a classical problem which can be solved for example by a least square method.

In other words, one computes

$$\mathrm{argmin}_A \sum_{x_j \in \mathscr{U}_x} \left\| T(x_j) - T(x) - A(x_j - x) \right\|^2 \ .$$

A gives (the approximation of) DT_x. In general, x is a point on the orbit.

Some remarks.

The choice of the neighborhood $\mathscr{U}_x$ is delicate. There are two opposite conditions.

- The neighborhood should be small enough so that the affine approximation is good. Some authors proposed to use quadratic corrections but the computation becomes heavier.

If the neighborhood is too big, the non linearities are not negligible anymore. Fake exponents appear which are multiple of the true ones. Hence in the presence of exponents which are multiple of others one should change the size of the neighborhood to check if the relation still holds.

- One wants to take the neighborhood large enough to have many points inside and get a good statistics in the noise elimination (least square).

If the map is invertible, the inverse transform has the opposite exponents (for the same invariant measure).

If the map T is volume preserving, that is to say $\det\left(DT_x\right) = 1$, the sum of the exponents is zero. This generalizes immediately to the case of constant determinant of DT_x as in the Hénon map.

For the symplectic transformations (as in mechanical systems), if λ is a Lyapunov exponent, $-\lambda$ is also a Lyapunov exponent with the same multiplicity.

For the continuous time one can perform a discrete time sampling (often delivered by the experimental observation) and use the discrete time algorithms.
There is also a continuous time versions of the **QR** method.