Chapter 3

Nonlinear time series analysis

3.1 Deterministic dynamical systems

While statisticians, when trying to explain the real world, are starting from a "random world" by introducing correlations or dependencies, respectively, physicians often think about the world as a deterministic one[1] and stochasticity (noise) is introduced as an approximation of effects which are either too high—dimensional or fluctuate too fastly to take them explicitly into account. So our starting point is a deterministic dynamic system living in a state space X which should be at the moment finite dimensional. The dynamics is either defined for discrete times

$$\boldsymbol{x}_{n+1} = F(\boldsymbol{x}_n) \tag{3.1}$$

thus defining a map or for continuous times

$$\dot{\boldsymbol{x}}(t) = f(\boldsymbol{x}) \ . \tag{3.2}$$

by a system of coupled ordinary differential equations which defines a flow $\boldsymbol{x}(t+t_0) = \phi^t(\boldsymbol{x}(t_0))$. There are several possibilites to relate the two descriptions to each other. Very often one considers the stroboscopic map of (3.2) for a given time T with $\boldsymbol{x}_n = \phi^T(\boldsymbol{x}_{n-1})$, e.g. in the case of periodically driven systems, or the Poincare surface of section (Poincare map) - the section of the flow with a hyperplane transversal to the flow. Formally it is defined

[1] With the exception of quantum mechanics, but even there the evolution of the wave function is deterministic

in the neighbourhood of a periodic orbit, but often it can be extented to the whole phase space. A simple way to generate the hyperplane, is to set one coordinate of the dynamical system to a fixed value (TISEAN program: *poincare*). In the following we only consider maps F, be it generically maps, maps generated by sampling flow data with a fixed sampling interval or Poincare maps.

3.1.1 Characterization — Dynamical invariants

One of the objectives of time series analysis is the characterisation of the system which generated the time series in question. In the case of deterministic dynamic systems there are quantities available which are better suited for this task than simply taking the model parameters. Deterministic dynamical systems can be charaterised by quantities which are invariant with respect to coordinate transformations and therefore independent of the "channel" by which we observe the system. We will came back to that in 3.1.2.

Attractor dimension

The first invariant is the *attractor dimension.* There are several definitions of attractors of dynamical systems around. Intuitively an attractor is the set of points in the phase space which are visited by the system asymptotically if the transient is discarded. A little bit more mathematically one could say that an attractor is an invariant set, which is attracting - in contrast to a repellor or a saddle point. To be attracting the set A must be a subset of an open set U, its neighbourhood, with

$$\lim_{n\to\infty} \inf_{y\in A} ||F^n(x) - y|| \to 0 \qquad \forall x \in U \ .$$

Sometimes it is only required that A attracts a set of positive measure, which leads to the different concept of Milnor attractors.
With respect to the dimension we can distinguish between dimensions of a set or dimensions of a measure. The first simply considers all points of a set, the latter also takes into account how often this points are visited by the system.
Let us first consider the box-counting dimension, which is an example of the first, but can be considered also in the more general framework of the latter.

Box-counting dimension of a set A: There are several equivalent definitions of this dimension. One possibility is to partition the phase space of our system by hypercubes with a side length ϵ. Then we call $N_\epsilon(A)$ the number of cells, which are intersected by the attractor A. The box-counting dimension D_0 is then defined as

$$D_0 = \lim_{\epsilon \to 0} -\frac{\log N_\epsilon(A)}{\log \epsilon} \; .$$

This is a property of the set only. It is invariant with respect to smooth invertible transformations of the phace space.
Examples: Fixed point attractors have dimension zero, limit circles 1, quasiperiodic motion on a torus 2. Middle thirs cantor set: Repellor of

$$x_{n+1} = \begin{array}{cl} 3x & \text{if} \quad x <= 1/2 \\ 3 - 3x & \text{if} \quad x > 1/2; . \end{array}$$

is a Cantor set. With $\epsilon = 3^{-n}$ and $N_\epsilon = 2^n$ we get $D_0 = \log 2 / \log 3 = log_3 2$. Before considering dimensions that take the measure into account, let us first discuss the entropy.

KS-entropy

While the dimension gives us information about the number of active degrees of freedom of the dynamical system, there is a second, complementary quantitiy, the metric or Kolmogrov-Sinai entropy which tells us about the randomness or irregularity of the dynamics. Basically it measures the uncertainty of the next observation given all the observations from the past.
To describe this we use the notion of an invariant measure. Remeber the probability space $(\Omega, \mathcal{B}, P)$ containing of a set of possible events Ω, a σ-algebra of subsets $\mathcal{B}$ (Set of subsets of Ω) and the probability measure P. Each set of events $A \subseteq \mathcal{B}$ has a probability $0 \leq P(A) \leq 1$, $P(\Omega) = 1$. Now our set of events is the phase space X of our dynamical system. We say a measure μ is invariant under a transformation $F : X \to X$, or F is a measure preserving transformation wrt to μ if

$$\mu(F^{-1}A) = \mu(A) \quad \forall \, A \in \mathcal{B} \; . \tag{3.3}$$

Let us consider some probability space $(X, \mathcal{B}, \mu)$ and a finite or countable index set I. A collection of measurable subsets, $\xi = \{C_\alpha \in \mathcal{B} | \alpha \in \mathcal{I}\}$ is called a **measurable partition** of X if

1. $\mu(X \setminus \cup_{\alpha \in I} C_\alpha) = 0$, i.e. the partition “contains” the whole measure.

2. $\mu(C_{\alpha_1} \cap C_{\alpha_2}) = 0$ if $\alpha_1 \neq \alpha_2$, i.e. the cells C_α of the partition are disjoint.

The entropy of μ with resepect to the partition ξ is then

$$H(\xi) := H_\mu(\xi) = -\sum_{\alpha\in I} \mu(C_\alpha) \log \mu(C_\alpha) \geq 0 \ . \tag{3.4}$$

Example: Logistic map

$$x_{n+1} = 1 - 2x^2 \tag{3.5}$$

with the partition $C_1 = [-1, 0)$,$C_2 = [0, 1]$. $\mu(C_1) = \mu(C_2) = 1/2$. Therefore $H(\xi) = \log 2$.
Now let us consider two partitions $\xi = \{C_\alpha|\alpha \in I\}$ and $\eta = \{D_\beta|\beta \in J\}$. Then the joint partition $\xi \vee \eta$ is defined as

$$\xi \vee \eta := \{C \cap D|C \in \xi, D \in \eta, \mu(C \cap D) > 0\}$$

It is also possible to define the conditional entropy of ξ given η using the notation $\mu(A|B) = \mu(A \cap B)/\mu(B)$ as

$$H(\xi|\eta) := -\sum_{\beta\in J} \mu(D_\beta) \sum_{\alpha\in I} \mu(C_\alpha|D_\beta) \log \mu(C_\alpha|D_\beta) \tag{3.6}$$

which can be written alternatively

$$H(\xi|\eta) = H(\xi \vee \eta) - H(\eta) \ .$$

Now we are able to define the entropy of the transformation F with respect to the partition ξ. First we introduce the joint partition of ξ and its preimages under F

$$\xi^F_{-n} := \xi \vee F^{-1}(\xi) \vee \ldots \vee F^{-n+1}(\xi) \ .$$

Example: ξ^F_{-2} for the logistic map (3.5) consists of the intervals between the points $-1, -\sqrt(1/2), 0, \sqrt(1/2), 1$, with $H(\xi^F_{-2}) = \log 4$ and $H(\xi^F_{-2}|\xi^F_{-1}) = \log 2$.

A this point we can also employ a complementary way to introduce these entropies, namely as entropies of a symbol sequence. Think of using the partition ξ to encode the phase space of the dynamical system. The trajectory $\{x_1, \ldots, x_n\}$ is encoded by a symbol sequence $\{\alpha_1, \ldots, \alpha_n\}$, if $x_1 \in C_{\alpha_1}$,$x_2 \in C_{\alpha_2}$ and so on. If we denote the probability to observe a certain symbol by $p(\alpha) = \mu(C_\alpha)$ we get for the entropy

$$H(\xi) = H(\alpha) = -\sum_{\alpha\in I} p(\alpha) \log p(\alpha) \ .$$

with α denoting the random variable which can have the value α with probability $p(\alpha)$. What corresponds then to ξ^F_{-n}? Being in a cell of this partition means that the trajectory was at time n in C_{α_n}, at $n-1$ in $C_{\alpha_{n-1}}$ and so on. Thus the measure of one cell of this partition corresponds to the joint probability $p(\alpha_n, \alpha_{n-1}, \ldots, \alpha_1)$, i.e. the probability of a certain subsequence of the string. Consequently the conditional entropy

$$H(\xi^F_{-2}|\xi^F_{-1}) = H(\alpha_2|\alpha_1) = -\sum_{\alpha_1,\alpha_2\in I} p(\alpha_2, \alpha_1) \log p(\alpha_2|\alpha_1)$$

is denoting the uncertainty of observing the symbol α_2 after α_1 was seen. The **metric entropy** of the transformation F relative to the partition ξ (sometimes also called the entropy rate of the process generated by F) is defined as

$$h(F,\xi) := h_\mu(F,\xi) := \lim_{n\to\infty} \frac{1}{n} H(\xi^F_{-n}) \tag{3.7}$$

which is equivalent to

$$h(F,\xi) = \lim_{n\to\infty} H(\xi|F^{-1}(\xi^F_{-n}) \ . \tag{3.8}$$

$H(\xi|F^{-1}(\xi^F_{-n})$ is monotonically decreasing. This can be shown using the representation via the symbol sequences:

$$H(\xi|F^{-1}(\xi^F_{-n}) = H(\alpha_0|\alpha_{-1}, \ldots, \alpha_{-n+1}) := h_n$$

Then

$$\begin{aligned} h_n - h_{n+1} &= H(\alpha_0|\alpha_{-1}, \ldots, \alpha_{-n+1}) - H(\alpha_0|\alpha_{-1}, \ldots, \alpha_{-n}) \\ &= MI(\alpha_0 : \alpha_n|\alpha_{-1}, \ldots, \alpha_{-n+1}) \geq 0 \end{aligned}$$

is a conditional mutual information.

The *KS-entropy* of F with respect to μ is then defined as the supremum over all partitions:

$$h_{KS}(F) := h_\mu(F) := \sup_{\xi, h(\xi)<\infty} h_\mu(F,\xi) \ . \tag{3.9}$$

A generating partition ξ_g is a partition for which the metric entropy is maximal, i.e.

$$h(F,\xi_g) = h_{KS}(F) \ .$$

There is however, in general no algorithm to find generating partitions for arbitrary dynamical systems. For 1-dimensional maps it is known how to

find them and for 2-d also an algorithm exists, which allowed to determine the generating partitions for well known systems, such as the henon map Grassberger and Kantz (1985) or the standard map **?**.
But if we cannot find a generating partition, is it possible to estimate the KS-entropy? Yes, because in most cases (nonatomic Borel measure on a compact metric space) a finer and finer refinement of the partition allows to get better and better estimates. Or more formally: If I consider a sequence of partitions ξ_i with $\mathrm{diam}(\xi_i) \to 0$ ($\mathrm{diam}(\xi_i) := \sup_{C \in \xi} \mathrm{diam}(C)$), then $h(F, \xi_i) \to h_{KS}(F)$. An important property of the KS-entropy is that

$$h_{KS}(F^k) = k h_{KS}(F);. \tag{3.10}$$

This should be taken into account, when estimating entropies from flow data using a delay embedding.

Lyapunov exponents

The central property of chaotic dynamics is its sensitive dependence on the initial conditions, i.e. the exponential divergence of initially neighbouring trajectories. In order to keep the dynamics bounded, however, this “stretching” of the attractor has to be complemented by a folding mechanism, which brings points together which were far away from each other. If we look only locally at the dynamics wo only see the stretching. So, if we denote the distance between two trajectories at time n by $\Delta_n = ||\boldsymbol{x}_n - \boldsymbol{x}'_n|| = ||F^n(\boldsymbol{x}_0) - F^n(vx'_0)||$ then we expect

$$\Delta_n \propto e^{\lambda n}$$

with the Lyapunov exponent

$$\lambda = \lim_{n \to \infty} \lim_{\Delta_0 \to 0} \frac{1}{n} \log \Delta_n \ . \tag{3.11}$$

As we will see in a moment, one can define a whole spectrum of exponents and (3.11) ist the largest one. This Lyapunov exponents allow already a classification of deterministic dynamical systems:

- stable fixed point: $\lambda < 0$
- stable limit cycle: $\lambda = 0$
- chaotic behaviour: $\lambda > 0$

Note that, however, for a diffusion process (random walk) $\Delta_n \propto \sqrt(n)$, i.e. $\lambda \propto \frac{log(n)}{n} \to 0$ for $n \to \infty$.
Now let us analyze the dynamics of the difference between two trajectories $\boldsymbol{x}_n$ and $\boldsymbol{y}_n = \boldsymbol{x}_n + \boldsymbol{\Delta}_n$ in more detail. Because we consider in the end infinitesimal differences, their dynamics is governed by the linearization of the map $\boldsymbol{F}$ (3.1),i.e. its Jacobian

$$\boldsymbol{J}(\boldsymbol{x}_n) = \left(\frac{\partial \boldsymbol{F}}{\partial \boldsymbol{x}}\right)_{\boldsymbol{x}=\boldsymbol{x}_n} \qquad J_{ij}(\boldsymbol{x}_n) = \left(\frac{\partial F_i}{\partial x_j}\right)_{\boldsymbol{x}=\boldsymbol{x}_n} .$$

This leads to a linear dynamical system with time dependent coefficients for the perturbations $\boldsymbol{\Delta}$

$$\boldsymbol{\Delta}_{n+1} = \boldsymbol{J}(\boldsymbol{x}_n)\boldsymbol{\Delta}_n .$$

The long term dynamics is the governed by the eigenvalues Λ_i of the product of th Jacobians

$$\left(\prod_{n=1}^{N} \boldsymbol{J}(\boldsymbol{x}_n)\right) \boldsymbol{u}_i^{(N)} = \Lambda_i^{(N)} \boldsymbol{u}_i^{(N)} . \tag{3.12}$$

with $\boldsymbol{u}_i^{(N)}$ denoting the eigenvectors of the product of the N Jacobians.
The Lyapunov exponent λ_i is then defined as the normalized logarithm of the modulus of the ith eigenvalue Λ_i of the product of all Jacobians along the trajectory (in time order) in the limit of an infinitely long trajectory:

$$\lambda_i = \lim_{N\to\infty} \frac{1}{N} \log |\Lambda_i^{(N)}| \tag{3.13}$$

Usually the eigenvalues are ordered according their magnitude, starting with the largest. The fact that the limit (3.13) exists and is unique was schown by **?** and is known as multiplicative ergodic theorem. This is a highly non-trivial result because the multiplication of matrices is non-commutative and the logarithm cannot be exchanged with the formation of the eigenvalues. In the case of one-dimensional maps, however, the definition reduces to

$$\lambda = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \log |F'(x_n)|$$

and the existence and uniqueness is established by the usual (Birkhoff) ergodic theorem.
Some properties:

- The Lyapunov exponents are invariant under smooth transformations of the phase space.
$$\tilde{\boldsymbol{F}}(\tilde{\boldsymbol{x}}) = \boldsymbol{g} \circ \boldsymbol{F} \circ \boldsymbol{g}^{-1}(\tilde{\boldsymbol{x}})$$
Then
$$\prod_{n=1}^{N} \tilde{\boldsymbol{J}}_n(\boldsymbol{x}_n) = \tilde{\boldsymbol{J}}_N^{(\boldsymbol{g})} \prod \boldsymbol{J}_n \tilde{\boldsymbol{J}}_1^{(\boldsymbol{g}^{-1})}$$
yields in the limit $N \to \infty$ the same eigenvalues and thus the same Lyapunov spectrum as the original dynamics. This ensures that the Lyapunov exponents are indeed invariants of a dynamical system.
- If μ is invariant under $\boldsymbol{F}$ then it is also under $\boldsymbol{F}^{-1}$. The absolute values of the Lyapunov exponents of $\boldsymbol{F}^{-1}$ remain the same but the sign of the exponents becomes reversed.
- Flow data have always at least one $\lambda_j = 0$.
- The Lyapunov spectra of Hamiltonian systems are symmetric wrt to zero, because the dynamics remains invariant wrt to time reversal.

Relation between the invariants

In many cases the Lyapunov spetrum contains all informations about the invariants of a dynamical system: The entropy is equal to the sum of the positiv Lyapunov exponents, the so called PESIN identity
$$h_{KS} = \sum_{\lambda_k > 0} \lambda_k \ . \tag{3.14}$$
The KAPLAN-YORKE formula makes the connection between the Lyapunov exponents and the fractal dimension of the attractor. If there is only one fractal diraction, one has
$$D_{KY} = n + \frac{\sum_{i=1}^{n} \lambda_i}{|\lambda_{n+1}|} \ , \sum_{i=1}^{n} \lambda_i \geq 0 > \sum_{i=1}^{n+1} \lambda_i$$
D_{KY} is called the KAPLAN-YORKE or also the LYAPUNOV dimension. A more general theorem was proven by Ledrappier and Young (1985):
$$\sum_i D_i \lambda_i = 0.$$
with D_i being partial dimensions, i.e. dimensions in a certain direction, with values between 0 and 1.

Some Examples

- Linear Systems: What about the deterministic parts of the linear systems considered by the statisticians? If they are stable, they have fixed point attractors, i.e. $D = 0$, only negative Lyapunov exponents and thus zero entropy.

- One dimensional maps: They have only one Lyapunov exponent. If $\lambda > 0$, $h_{KS} = \lambda$, $D_{KS} = 1$.

- Two dimensional maps: If $\lambda_1 > 0 > \lambda_2$:

$$D_{KS} = \begin{cases} 2 & \text{if} \quad |\lambda_1| > |\lambda_2| \\ 1 + \frac{\lambda_1}{|\lambda_2|} & else \end{cases}$$

3.1.2 Phase space recontruction — embedding theorems

Very often one can only observe one or a few variables of a higher dimensional dynamics. The question then is: Can we reconstruct the phase space of the underlying dynamical systems in order to estimate dimension, entropy and the Lyapunov exponents? The answer is yes and is founded on the embedding theorems by Whitney (1936), Takens (1980) and its extension by Sauer et al. (1991). The basic idea is the following: If we oberve a dynamical system

$$\boldsymbol{x}_{n+1} = F(\boldsymbol{x}_n)$$

via an observation function $y = h(\boldsymbol{x})$, the dynamical system (3.1) gives rise to a dynamic of y. Takens proposed to reconstruct the original phase space using the so called *delay coordinates* $\boldsymbol{y}_n = (y_n, y_{n-1}, \ldots, y_{n-m+1})$. The question is now under which conditions there exists a deterministic dynamical system G for the dynamics of $\boldsymbol{y}$ and how it is related to F? Obviously, $\boldsymbol{F}$ induces a dynamics for $\boldsymbol{y}$ because

$$y_{n+k} = h(\boldsymbol{x}_{n+k}) = h(F^k(\boldsymbol{x}_n)) \ .$$

But exists there a also a dynamical system

$$y_{n+1} = G(\boldsymbol{y}_n)$$

and is the map between the two state spaces invertible, i.e. will we have a one to one relationship between $\boldsymbol{x}$ and $\boldsymbol{y}$? The answer is that under generic conditions m has to be large enough to ensure this one to one relationship.

Whitney proved that every D-dimensional smooth manifold can be embedded in the $\mathbb{R}^{2D+1}$, and that the set of maps forming an embedding is a dense and open set in the space of C^1 (continously differentiable) maps. Thus for an arbitrary map $\in C^1$ there exists an embedding in its neighborhood. Takens applied this to attractor reconstruction using delay coordinates. Sauer et al. improved the result of Takens and extented it to more general situations. Their central result is, that the recontructed state space has to be at least of dimension $m > 2D_0$, with D_0 the box counting dimension of the attractor, in order to have almost every embedding of the original phase space being one to one for the states and the Jacobian (Immersion).
Sauer et al. also considered the question, whether filtering the data could affect to possibility of a proper embedding. The result was, that the application of finite impuls response (FIR) filters to the delay coordinates would still allow an embedding, as long as enough independent observables will be considered. On the other hand, IIR filters might change the properties of the dynamical system (they are a dynamical system by themselves) and therefore affect the dimension and entropies of the whole system. Consider for instance the following extented Henon map:

$$x_{n+1} = 1 - Ax_n^2 + By_n \tag{3.15}$$

$$y_{n+1} = x_n \tag{3.16}$$

$$z_{n+1} = \alpha z_n + x_n \tag{3.17}$$

Even for $|\alpha| < 1$ this additional degree of freedon can increase the attractor dimension.
Up to this point we only discussed to which extend the properties of a given dynamical system can be recovered in the reconstructed phase space, e.g. by using a delay embedding. Here two remarks are in order:

1. For practical applications there might be better or worse phase space reconstruction. For instance, in the case of the delay embedding the delay time τ has to be selected, which we set to 1 so far, but which can be set arbitrarily in principle — with some exceptions for periodic processes, remember the discussion of the aliasing problem. Also, if more then one observable is available, one can ask, which coordinates should be used, delay coordinates from only one, or some mixed delay vector of the two, but which one? There is up to now no general method to find optimal state space reconstructions, but there are some pragmatical approaches available, which we will discuss later.

2. Up to now we started with a dynamical system and a given "true" state space. This is, however, not the situation, which we will find

in practice. There we want to characterize the system, which has produced the data, but there is nothing like a "true" state space - there are only equivalent representations of one physical system and one of them is our state space reconstruction. There might be, however, some of them easier to interpret than others.

False nearest neighbors

TISEAN program: *false_nearest*
How can we detect a sufficiently large embedding dimension m? One Possibility is to look for so called *false nearest neighbors* (Kennel et al. (1992)). The idea is to use the geometrical structure induced by the deterministic character of the dynamics, i.e. the fact that the attractor lies in a low-dimensional manifold. As long as the embedding dimension is too low, there is no one to one embedding of the attractor and neighbouring points in the embedding space might not be neighbours in the phase space. Thus if a point $\boldsymbol{x}_i$ is a nearest neighbour to $\boldsymbol{x}_j$ in m dimensions, but not in $m+1$ dimensions, it is called a false neighbor [2] .
Then with increasing m the fraction of false nearest neighbours is estimated. If this fraction drops for some m^* this is a good candidate for a minimal embedding dimension. Usually, it drops already for $m > D_0$, which might not be sufficient as an embedding dimension, depending what one wants to analyze. There are, however, some pitfalls of this algorithm, which one has to aware of:

- In chaotic systems also true neighbours become more seperated when increasing the embedding dimension due to the effect of the chaotic dynamics.

- If the data are noisy the signature of the determinism becomes weakened.

- If the attractor is strongly folded in the reconstructed phase space the neighborhood size has to be small enough to separate several sheets of the folded attractor.

[2] In *false_nearest* a slighty different criterion is used: if the distance in $m+1$ is larger than *factor* times the distance in m dimensions it is considered a false nearest neighbor.

3.1.3 Dimension and entropy estimation

Box-counting dimensions and — entropies

TISEAN implementation: *boxcount*.
Univariate data: Let us start with a time series of N data points $\{x_1, \dots, x_N\}$. If we have data from an interval $[x_{min}, x_{max}]$ encoding the data with k-symbols corresponds to a partition of the reconstructed phase space with hypercubes of side length $\epsilon = \frac{x_{max}-x_{min}}{k}$. In the m-dimensional reconstructed phase space spanned by the points $\boldsymbol{x}_n = (x_n, x_{n-1}, \dots, x_{n-m+1})$ we can count how often each of the hypercubes is visited. The relative frequencies defines a probability distribution on the cells of this partition and we can estimate its Shannon entropy

$$H(m, \epsilon) = -\sum p_j \log p_j \quad \text{with} \quad p_j = \frac{n_j}{N} \tag{3.18}$$

The information dimension can then be estimated by looking at the slope of $H(m, \epsilon)$ with respect to $-\log \epsilon$, because

$$H(m, \epsilon) = \text{const} - D_1 \log \epsilon + \mathcal{O}(\epsilon) \ .$$

Clearly, for $k = 1$ and therefore $\epsilon = x_{max} - x_{min}$ only one box is filled, with $p = 1$ and $H(\epsilon) = 0$. On the other hand, for sufficciently small ϵ, each cell of the partition contains only one point therefore $p_j = 1/N$ and $H(\epsilon) = \log N$. This is clearly a finite sample effect. The entropy (3.18) is only a good estimate of the entropy of the invariant measure if ϵ is not too small, or N is large enough, respectively. For a more detailed discussion of finite sample effects and its correction see Grassberger (2003).
The dimension

$$D_1 = \lim_{\epsilon \to 0} -\frac{H_\epsilon}{\log \epsilon} \tag{3.19}$$

is called the *information dimension*. It is possible to introduce a whole family of D_q, the so called Renyi dimensions, using the Renyi entropies

$$H^{(q)}(\epsilon) = \frac{1}{1-q} \log \sum_j p_j^q \tag{3.20}$$

and corresponding dimensions

$$D^{(q)}(m, \epsilon) = \lim_{\epsilon \to 0} -\frac{H_\epsilon^{(q)}}{\log \epsilon} \ . \tag{3.21}$$

Exercise: Show using the rule of l'Hospital that

$$\lim_{q\to 1} H^{(q)}(\epsilon) = -\sum p_j \log p_j \ .$$

If the D_q are different the system is called *multifractal*. Several methods, like multifractal analysis and the thermodynamic formalism builds upon the Renyi entropies and dimensions, respectively.

Estimating the KS-entropie

Estimating the KS-entropy using the box-counting entropy estimates $H(m,\epsilon)$ is straightforward:

1. Find an embedding dimension m_0, which is large enough, at least $m_0 > D_0$.

2. Estimate $H(m,\epsilon)$ for some values of $m \geq m_0$. Estimate the conditional entropies
$$h(m,\epsilon) = H(m+1,\epsilon) - H(m,\epsilon) \tag{3.22}$$
plot $h(m,\epsilon)$ as a function of $\log\epsilon$ and look for a plateau $h(m,\epsilon) \approx$ const at some ϵ range. The ϵ has to be large enough to minimize finite sample effects, but also small enough to resolve the deterministic structure.

3. If the $h(m,\epsilon)$ remains also constant for increasing m the value might be used as an estimate for h_{KS}.

There are, however, severe problems with this procedure. Although the $h(m,\epsilon)$ are monotonically decreasing, i.e. $h(m,\epsilon) \geq h(m+1,\epsilon)$ we cannot expect that $h(m,\epsilon)$ estimated from the data gives an upper bound for the $h(\infty,\epsilon)$, because the finite sample effects lead to an underestimation of the conditional entropies. Thus also alternative methods should be used to estimate the KS-entropy from a time series, such as the correlation entropy and the Lyapunov exponents.

The correlation dimension

TISEAN implementation *d2*
The most popular quantity from nonlinear time series analysis is the correlation dimension. For low dimensional data it can be reliably estimated already from relatively short data sets with a relatively simple algorithm.

Mathematically the correlation dimension of a measure μ is defined as follows:

$$D_2 = -\lim_{\epsilon \to 0} \frac{\log \int_X \mu(B(\boldsymbol{x}, \epsilon)) d\mu(\boldsymbol{x})}{\log \epsilon} \tag{3.23}$$

with $B(\boldsymbol{x}, \epsilon)$ denoting the Ball of radius ϵ centered at point $\boldsymbol{x}$, i.e. the set of points $\boldsymbol{y}$ with $||\boldsymbol{x} - \boldsymbol{y}|| < \epsilon$. It is estimated from N data points via the correlation sum

$$C(\epsilon) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \Theta(\epsilon - ||\boldsymbol{x}_i - \boldsymbol{x}_j||)$$

with Θ being the Heaviside step function $\Theta(x) = 0$ if $x \leq 0$ and $\Theta(x) = 1$ if $x > 1$. That means we count the fraction of distances between data points in the phase space, which is smaller than ϵ. In the limit $N \to \infty$ we expect C to scale like a power law, $C(\epsilon) \propto \epsilon^D$, and we can define the correlation dimension by

$$\begin{aligned} D_2(N, \epsilon) &= \frac{\partial C(\epsilon, N)}{\partial \log \epsilon} && (3.24) \\ D_2 &= \lim_{\epsilon \to 0} \lim_{N \to \infty} D_2(N, \epsilon) && (3.25) \end{aligned}$$

In practice, however, we have only a finite amount of data, so we cannot perform the limits and so he have to estimate the dimension at finite resolution ϵ. Therefore one usually plots $D_2(m, \epsilon)$ via $\log \epsilon$ (see Fig. **??**) for the example of the henon map. Then one has to identify a region, where it is approximately constant and can then estimate it by fitting a stright line in the log-log plot of $C(\epsilon)$. This plateau or scaling range is limited on the large scales, because if ϵ is too large, the structure of the attractor cannot be resolved and usually the dimension is overestimated[3], while the lower end might be determined by the accuracy of the measurement (how many digits), the number of data points and/or the amount of noise.

If the embedding dimension m is too small and the amount of data is sufficiently large the plateau should appear at the value of the embedding dimension $D_2 = m$. Only if the embedding dimension is larger than D_2 we can expect to find a plaeau et the value of the attractor dimension. This happens usually alread for $m > D_0$ and not only for $m > 2D_0$, the correct embedding dimension. The explanation is that the self-intersections of the attractor have zero measure and therefore do not affect our dimension

[3] Note that this might be totally different for strongly correlated data, such as highly sampled flow data.

estimates. However, for prediction or modelling this self-intersections are important and that m might be too small.

Temporal correlations and the Theiler correction

There is an important practical problem, which lead to many spurious dimension estimates in the past, the problem of temporal correlations. We use the number of neighbours of $\boldsymbol{x}$ with a distance smaller than ϵ to estimate the measure of $\mu(B(\boldsymbol{x}, \epsilon))$, i.e. the probability to find a point in the ϵ neighbourhood of $\boldsymbol{x}$. If now the actual neighbours of these points are not only neighbours in the phase space but also neighbours in time, we get obviously a biased estimate. There might be even contributions to this bias from the other points in the neighbourhood and their temporal correlated neighbours if they are also neighbours of the first point. Theiler (1986) proposed therefore to exclude all points in a teporal window around the refence point from the calculation. This is sometimes called the "Theiler window" n_{TW}. The formula for the correlation sum then reads

$$C(\epsilon) = \frac{2}{(N - n_{TW}(N - 1 - n_{TW})} \sum_{i=1}^{N} \sum_{j=i+n_{TW}+1}^{N} \Theta(\epsilon - ||\boldsymbol{x}_i - \boldsymbol{x}_j||) \ .$$

To determine a good value of this window Provenzale et al. (1992) introduced the so called space time separation plot (in TISEAN *stp*).

3.1.4 Estimating the Lyapunov exponents

The largest Lyapunov exponent

TISEAN: *lyap_k, lyap_r* For the restimation of the largest Lyapunov exponent the expansion rate has to be estimated. This is done by calculating the logarithm of the mean difference between points which were initially in the neighbourhood of a reference point and finally also averaging over these reference points:

$$S(\Delta n) = \frac{1}{N - \Delta n} \sum_{n=1}^{N-\Delta n} \log \left(\frac{1}{|\mathcal{U}(\boldsymbol{x}_\prime)} | \sum_{\boldsymbol{y}_n \in \mathcal{U}(\boldsymbol{x}_\prime)} |v y_{n+\Delta n} - \boldsymbol{x}_{n+\Delta n}| \right) \quad (3.26)$$

Then the linear slope of $S(\Delta n)$ should be an estimate of the largest Lyapunov exponent, becaue the difference will be dominated by the largest exponent. Beside the usual embedding parameters one has also to specify the neighbourhood, either by its diameter ϵ or by the number of neighbours. The program *lyap_k* estimates the stretching factor for a set of neighbourhood sizes and provides some statistics about the numbers of neighbours found.

Lyapunov spectrum

A reliable estimation of the Lyapunov spectrum is in most cases only possible if the system equations are known or a global model is available for a given data set. In this case we can estimate the Jacobian from the equations. Nevertheless, the product of the Jacobians will become singular, so it cannot be evaluated. Therefore usually a procedure introduced by Benettin et al. (1980a,b) is used: A set of orthogonal vectors, spanning the phase space is iterated using the linearized dynamics. After a few steps the vectors become more and more aligned in the direction of the largest Lyapunov exponent. Therefore the vectors are iteratively othonormalized, beginning with the largest one and the scaling factors are stored. The Lyapunov exponents are then estimated by the averages of the logarithms of the scaling factors.
One possibility to estimate the Lyapunov spectra from data would be to estimate the Jacobians directly from the data (TISEAN: *lyap_spec*). This corresponds to fitting loacally linear models, which will be discussed later. At this point we will only mentions some problems of this approach:

- The local neighbourhoods used for the linear fit has to be large enough to avoid fitting the pecularities of the noise.
- On the other hand side the local neighbourhood has to be small enough not to smear out the nonlinear structures of the attractor.
- Very often this method does not provide robust estimates of the exponents. Thus, although conceptionally appealing because it contains all information about the invariants, only estimating the Lyapunov spectrum from data might be actually a bad idea.

Bibliography

Benettin, G., Galgani, L., Giorgilli, A., Strelcyn, J.-M., 1980a. Lyapunov characteristic exponents for smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 1: Theory. Meccanica 15 (1), 9–20.
URL http://dx.doi.org/10.1007/BF02128236

Benettin, G., Galgani, L., Giorgilli, A., Strelcyn, J.-M., 1980b. Lyapunov characteristic exponents for smooth dynamical systems and for hamiltonian systems; a method for computing all of them. part 2: Numerical application. Meccanica 15 (1), 21–30.
URL http://dx.doi.org/10.1007/BF02128237

Grassberger, P., 2003. Entropy estimates from insufficient samplings. arXiv:physics, 0307138v2.

Grassberger, P., Kantz, H., 1985. Generating partitions for the dissipative Hénon map. Phys. Lett. A 113 (5), 235–238.

Kennel, M. B., Brown, R., Abarbanel, H. D. I., 1992. Determining embedding dimension for phase-space reconstruction using a geometrical construction. Phys. Rev. A 45, 3403–3411.

Ledrappier, F., Young, L.-S., 1985. The metric entropy of diffeomorphisms part II : Relations between entropy, exponents, and dimension. Ann. Math. 122, 540.

Provenzale, A., Smith, L., Vio, R., Murante, G., 1992. Distinguishing between low-dimensional dynamics and randomness in measured time series. Physica D 58, 31–49.

Sauer, T., Yorke, J. A., Casdagli, M., 1991. Embedology. J. Stat. Phys. 65, 579–616.

Takens, F., 1980. Detecting strange attractors in turbulence. In: Rand, D. A., Young, L.-S. (Eds.), Dynamical Systems and Turbulence (Warwick 1980). Vol. 898 of Lecture Notes in Mathematics. Springer-Verlag, Berlin, pp. 366–381.

Theiler, J., 1986. Spurious dimensions from correlation algorithms applied to limited time-series data. Phys. Rev. A 34, 2427–2432.

Whitney, H., 1936. Differentiable manifolds. Ann. Math. 37, 645.