Contents

Chapter 1

Introduction

This diplom thesis treats the analysis of spatio-temporal data. Such data occur in many fields of physics both in experimental investigations and in theoretical physics, mainly from computer simulations. Various methods of data analysis are compared and the validity of their results and possible pitfalls are pointed out. Finally these methods are applied to the data from computer simulations of nonlinear deterministic models of two different reaction-diffusion systems.

Spatio-temporal data can be regarded as the solutions of partial differential equations. These show a much wider range of solutions compared to ordinary differential equations. This is especially the case for nonlinear partial differential equations, as an example consider the Navier-Stokes equation. It seems impossible to find an analytic classification of the solutions similar to the case of ordinary linear differential equations. For numerical analysis one always has to introduce discretizations in both space and time[1], which have to be chosen in a way that no spurious results emerge.

The typical problems that arise in the analysis of spatio-temporal data are different from those in time series analysis: Long transients are common for such systems, more probable if more 'typical wavelengths' fit into the spatial extent. For some systems the transient time may be so long that one cannot expect ever to observe a stationary dynamics. If the transient time is unknown it may be difficult or impossible to decide wether the examined data is part of the transient or not. Moreover, the size of the data samples can be very big when compared to simple time series data, especially if the system has two or more spatial dimensions. Therefore data reduction methods like the pattern recognition algorithm using the singular value decomposition elaborated in this thesis are highly desirable. In many practical cases only data reduction makes the samples accessible for analysis on computers with their limited resources of memory and computing speed.

From all the above it can be seen that the investigation of spatio-temporal data has experimental character. Familiarity with the methods of analysis and the possible artefacts produced by their unsuitable application is essential.

[1]Moreover there is also a discretization due to the representation of values as floating point numbers in digital computer systems. Its effects are not treated in this thesis.

This thesis is organized as follows: In the first part the investigated systems, the heterostructure hot-electron diode in an electric circuit and the heterogeneous catalytic oxidation of carbon monoxide on a monocrystalline platinum surface, and the models used in the simulations are briefly introduced. Then several methods for data analysis are explained. These methods are applied to the data samples from computer simulations of the models at different values of their dynamical parameters. The results obtained are discussed in the last part of this thesis.

Chapter 2

The Systems Examined

2.1 Heterostructure Hot-Electron Diode

The heterostructure hot-electron diode (HHED) is a layered semiconductor structure that
consists ... ers are s-
lightly ... nsport of
electron ... conduc-
tivity th ... bias, one
with a l ... e GaAs-
layer. F ... As-layer)
the mea ... nneling.
This res ... a greater
fraction ... diode is
caused ...

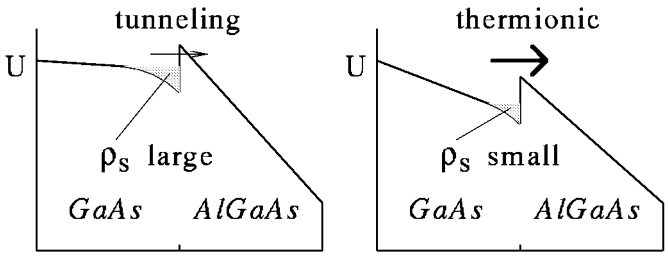

Figure 2.1: **The two possible cases for the conduction in the HHED.** *Conduction band diagram for the low current state (left) and the high current state (right).*

These are both stationary states and bistability can actually be found in a certain range of voltage (cf. [16] p.431). But these different conduction types can also be observed in one system on different locations in the junction. This leads to 'current filaments' which may be stationary or have a time dependence. Thus many types of interesting dynamics are possible, from the periodic case over (temporal) chaos to turbulence.

The system examined here is a heterostructure hot-electron diode driven under current-controlled conditions in the circuit shown in figure 2.2. The current is determined by the (large) resistor R. The parameters are the total current I_0 through the HHED and the external capacita

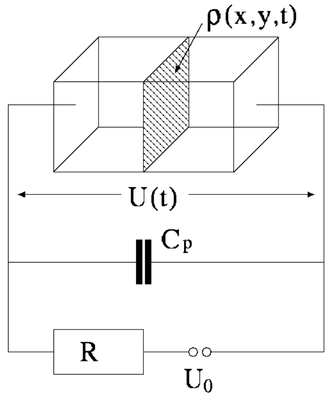

Figure 2.2: **The Circuit that is used to operate the HHED.**

The dynamic variables for the model are the interface charge density, $\rho_s(x,y,t)$, and $U(t)$, the voltage across the HHED. Here x and y are the spatial coordinates parallel to the interface. For the derivation of the dynamic model see [15] chapter 2, p.9.

The model equations are:

$$\frac{\partial}{\partial t}\rho_s(x,y,t) = j_1 - j_2 + \frac{\mu}{\varepsilon_1/L_1+\varepsilon_2/L_2}\nabla\cdot(|\rho_s|\,\nabla\rho_s) \qquad (2.1)$$

$$\frac{d}{dt}U(t) = \frac{1}{C+C_p}\left(I_0 - \int_A dx\,dy\left(\frac{C}{C_1}j_1+\frac{C}{C_2}j_2\right)\right) \qquad (2.2)$$

where

L_1, L_2 are the layer widths,
F_1, F_2 are the electric fields in the layers,
$\varepsilon_2, \varepsilon_1$ are the permittivities of the layers,
A is area of the interface,
C_p is the capacity of a capacitor parallel to the HHED,
$C_1 = 4\,A\,\varepsilon_1/3\,L_1$ is the effective capacity of layer 1,
$C_2 = A\,\varepsilon_2/L_2$ is the effective capacity of layer 2,
$C = (1/C_1+1/C_2)^{-1}$ is the capacity of the HHED,
$j_1(\vec{x},t)$ is the current density in layer 1,
$j_2(\vec{x},t)$ is the current density in layer 2,
μ is the mobility of the 2-dimensional electron gas at the interface,
I_0 is the total current through the structure.

(Cf. [13] chapter 4.1, p.42).

After introducing the dimensionless variables

$$u := U \frac{C}{\varepsilon_1 F_0 A} \tag{2.3}$$
$$v := \rho_s \frac{C}{\varepsilon_1 F_0 C_2} \tag{2.4}$$
$$j_0 := I_0 \frac{1}{A \sigma F_0} \tag{2.5}$$

and normalized time and space coordinates

$$\tau_s := \varepsilon_1 C_2/(\sigma C) \tag{2.6}$$
$$l_s := \sqrt{\frac{D \varepsilon_1 C_2}{\sigma C}} \tag{2.7}$$

the equations can with certain simplifications be rewritten as:

$$\frac{\partial}{\partial t} v = \frac{u - v}{(u-v)^2 + 1} - T v + \Delta v \tag{2.8}$$
$$\frac{d}{dt} u = \alpha (j_0 - (u - \langle v \rangle)) \tag{2.9}$$

where

$\alpha = C_2/(C + C_p)$
$T = T_0 \tau_s$ the normalized tunneling rate and
$\langle v \rangle = \int dx\, dy\, v(x, y)/A$ is the spatial average of $v(x, y)$.
(Cf. [13] chapter 5.1, p.61).
The control parameters for this model are now α and j_0.

This model has solutions that show 'spiking', as does the more complicated one above. The expression 'spiking' is used for the short time occurrence of current filaments that result in typical spikes in the voltage, which can be regular or irregular (cf. [13] p.53). It may very well be one of simplest possible models showing spiking and may therefore be called a 'generic' model for spiking ([13] p.62).

For $T = 0.05$ and $du/dj_0 < 0$ this is an activator-inhibitor system with v acting as activator and u as inhibitor.

Equation 2.8 is a reaction-diffusion equation of the type

$$\frac{\partial a(\vec{x}, t)}{\partial t} = f(a(\vec{x}, t), U) + D \Delta a(\vec{x}, t) \tag{2.10}$$

(cf. [13] p.77).

The data samples that were examined are outputs of computer simulations of the dynamics of the generic model (without noise) with one spatial coordinate.

2.2 Heterogeneous Catalysis on a Monocrystalline Surface

The second system examined is that of the catalytic oxidation of CO on a Pt-110 surface in a stirred reactor at low pressures ($p < 10^{-3}$Torr).

This system is experimentally well examined. Many types of interesting behavior have been observed for different parameter values. Depending on the reaction conditions one observes e.g. standing or moving waves, spiral-waves, concentric elliptic waves and turbulence (cf. [18]).

The oxidation of CO with O_2 happens entirely on the Pt(110) surface between adsorbed particles (Langmuir-Hinshelwood mechanism, cf. [14] p.75). Oxidation of CO to CO_2 in the gas phase or between one adsorbed particle and one arriving from the gas phase does not occur. The catalytic effect may be explained as follows: The O_2 dissociates, when adsorbed, into two O-atoms. The recombination would require about 55 kcal/mol, whereas the formation of CO_2 out of an adsorbed O-atom and an adsorbed CO-molecule needs only about 25 kcal/mol of activation energy. Therefore the latter process has a greater probability. The CO_2 leaves the Pt-surface immediately after its formation.
(Cf. [18] p.340 and [14] p.75).

The reaction equations are

$$
\begin{aligned}
CO + * &\rightleftharpoons CO_{ad} && (2.11)\\
O_2 + 2* \rightleftharpoons O_{2,ad} &\rightarrow 2\,O_{ad} && (2.12)\\
O_{ad} + CO_{ad} &\rightarrow CO_2 + 2* && (2.13)
\end{aligned}
$$

Where the $*$ denotes a free adsorption site on the Pt-surface.

These add to the net reaction:

$$2\,CO + O_2 \rightarrow 2\,CO_2 \tag{2.14}$$

The simplest mesoscopic model that correctly describes the experimentally observed behavior in all examined parameter regions needs three dynamic variables (cf. [14] p.94).

In the following let $c(\vec{x},t)$ denote the covering with carbon monoxide, $o(\vec{x},t)$ denote the covering with oxygen and $a(\vec{x},t)$ denote the portion of the Pt-surface that has an 1x1 structure. Sketches of the two types of surface structures are shown in figure 2.4. If the Pt-surface is covered mainly with CO (and has a 1x1 structure) the adsorption of O_2 is inhibited and so the reaction rate is low. At low coverings with CO (and 1x2 'missing row' structure, cf. [14] p.72) the oxygen can be adsorbed better and there is a high reaction rate. These two phases with its different adsorption probabilities for oxygen have to be included in the model ([14] p.94). In the equilibrium a approaches a value that is determined by a function $f(c)$ shown in figure 2.3. (cf. [14] p.97)

There exists a global coupling mechanism for the system. The adsorption or desorption of gas molecules changes the corresponding partial pressures in the gas phase. But spatial differences in the partial pressures are decreased through the motion of the gas molecules.

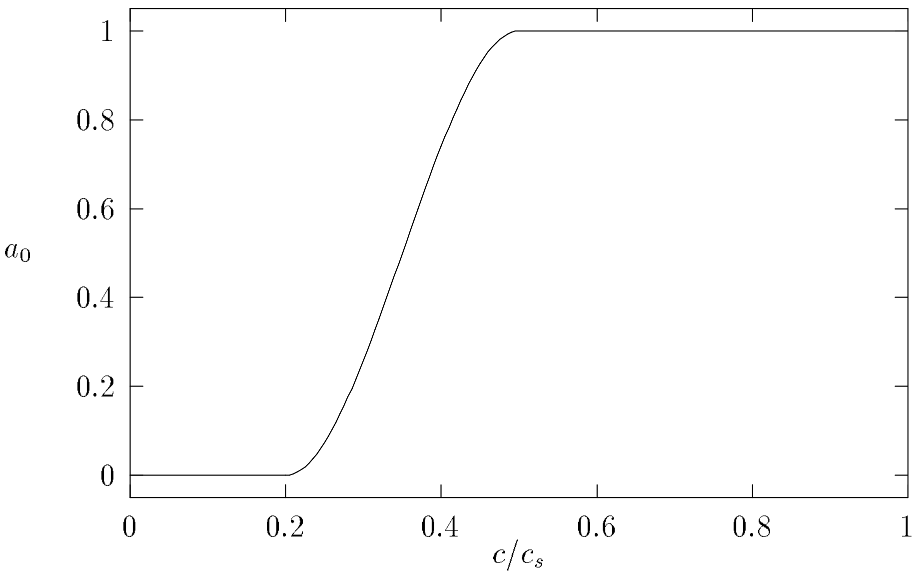

Figure 2.3: **Equilibrium function $a_0 = f(c/c_s)$ for the portion of 1x1 structure of the Pt-surface versus CO covering.**
Between 0.2 and 0.5 the function is a third order polynomial in c/c_s which is completely determined by the assumptions that f is a continuous and differentiable function being 0 for values $c/c_s < 0.2$ and 1 for $c/c_s > 0.5$ (cf. [14] p.95).

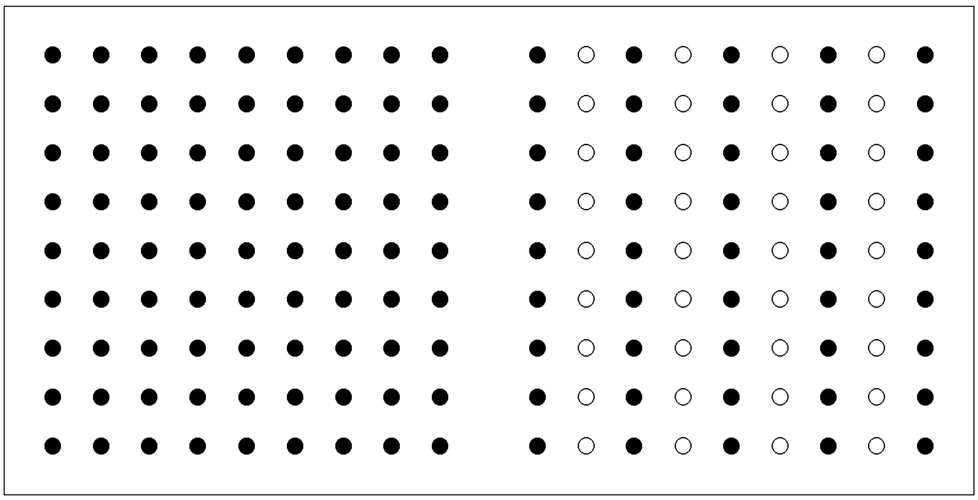

Figure 2.4: **Sketch of the surface structures of the Pt-110 surface.**
Left side: 1x1-type, right side: 1x2-(missing row) type. The disks symbolise Pt-atoms.

In the range of temperatures and pressures used in the experiments the velocities of the particles are in the order of $1000\,m/sec$. This velocity is much greater than that of the typical spatial processes in the system. Therefore changes in the partial pressures at one location immediately affect the dynamics of the whole system (cf. [18] p.343).
(See [14] p.94, p.72).

The model equations are (cf. [19])

$$\begin{aligned}
\frac{\partial c}{\partial t} &= \left(D_1 \frac{\partial^2 c}{\partial r_1^2} + \frac{\partial}{\partial r_2}\left(D_2(a)\frac{\partial c}{\partial r_2}\right)\right) + \\
&\quad + k_1 \cdot p_{CO} \cdot s_c \left(1 - (c/c_s)^3\right) - k_2 \cdot c - k_3 \cdot c \cdot o && (2.15)\\
\frac{\partial o}{\partial t} &= k_4 \cdot p_{o2} \cdot (s_{o1} \cdot a + s_{o2} \cdot (1-a))\left(1 - (o/o_s) - (c/c_s)\right)^2 - k_3 \cdot c \cdot o && (2.16)\\
\frac{\partial a}{\partial t} &= k_5 \cdot \begin{cases} -a & 0.0 \le \frac{c}{c_s} \le 0.2 \\ -a - \frac{\left(\frac{c}{c_s}\right)^3 - 1.05\left(\frac{c}{c_s}\right)^2 + 0.3\left(\frac{c}{c_s}\right) - 0.026}{0.0135} & 0.2 < \frac{c}{c_s} < 0.5 \\ -a + 1 & 0.5 \le \frac{c}{c_s} \le 1.0 \end{cases} && (2.17)\\
&= k_5 \left(f(\frac{c}{c_s}) - a\right)
\end{aligned}$$

where

$$D_2(a) := D_2 \left(1 + (\frac{D_1}{D_2} - 1)\, a^3\right) \tag{2.18}$$

$$\frac{dp_{co}}{dt} = \frac{J_{io}}{V}\left\{p_{coe} - p_{co}\left[1 + \frac{V_{ML}}{J_{io}\,A}\int \left(k_1\, p_{co}\, s_c\, \left[1 - (c/c_s)^3\right] - k_2\, c\right)\, dr^2\right]\right\} \tag{2.19}$$

where
r_i $(i = 1, 2)$ denote the two inequivalent spatial directions on the surface,
J_{io} is the gas flow through the reactor,
A is the size of the Pt-surface,
p_{coe} is the partial pressure of CO in the gas inlet,
p_{CO} and p_{O_2} are the partial pressures of CO and O, respectively,
s_c and s_o are the adhesion coefficients of CO and O, respectively,
c_s and o_s are the saturation coverings of CO and O, respectively, (the quantities that correspond to one monolayer),
k_1 is the hit rate of the CO-molecules on the Pt-surface: $k_1 = (2\pi m_{CO}\, k\, T)^{-1/2}$,
k_2 is the kinetic constant of the desorption of CO: $k_2 := k_2' exp(-E\, A_2/R\, T)$
k_3 is the reaction constant for the formation of CO_2: $k_3 := k_3' exp(-E\, A_3/R\, T)$
k_4 is the hit rate of the O_2 molecules on the Pt-surface: $k_4 = (2\pi m_O\, k\, T)^{-1/2}$,
k_5 is the reaction constant for the relaxation of a towards the equilibrium $a_0 = f(c/c_s)$.

The variables $k_1 ... k_4$ are responsible for the temperature dependence of the system. Equation 2.19 describes the global coupling over the gas phase. Equation 2.15 is a reaction-diffusion equation of the type

$$\frac{\partial a(\vec{x}, t)}{\partial t} = f(a(\vec{x}, t), U) + D\, \Delta a(\vec{x}, t) \tag{2.20}$$

The first two terms on the right hand side of that equation 2.15 are the Laplacian operator for the two inequivalent directions r_1, r_2. These terms describe the diffusion of the adsorbed CO-molecules. The terms with the factor k_3 in equations 2.15 and 2.16 are due to the formation of CO_2 from one CO-molecule with one O-atom (equation 2.13) according to standard kinetics. The term with k_2 in equation 2.15 corresponds to thermic desorption of CO. Desorption of O_2 needs not being taken into account. The terms with k_1 and k_4 in equations 2.15 and 2.16, respectively, describe the adsorption process of CO and O_2. The CO is adsorbed with precursor kinetics: the adsorption happens over a transient state where the molecule is not yet in its final place. It may move to its final location, therefore the adsorption probability does not decrease linearly with the covering, it is nearly one up to a certain threshold value and falls rapidly for greater coverings (cf. [14] p.80). The exponent 2 in the term multiplied with k_4 in equation 2.16 is due to the fact that two empty sites are needed for the adsorption of each O_2 molecule. The third equation (2.17) describes the relaxation of a towards its equilibrium state. The temperature dependence of the model comes from that of the reaction constants.

Introducing the dimensionless variables

$$u := \frac{c}{c_s} \tag{2.21}$$

$$v := \frac{o}{o_s} \tag{2.22}$$

$$\tau := k_5\, t \tag{2.23}$$

$$z_i := \frac{k_5\, r_i}{\sqrt{D_2\, k_1\, p_{CO}}} \tag{2.24}$$

and defining

$$\varepsilon = \frac{k_5}{k_1\, p_{coe}} \tag{2.25}$$

$$\beta = \frac{k_2}{k_1\, p_{coe}} \tag{2.26}$$

$$\gamma = \frac{k_3\, o_s}{k_1\, p_{coe}} \tag{2.27}$$

$$\delta = \frac{k_4\, p_{o2}\, (s_{o1} - s_{o2})}{k_1\, p_{coe}} \tag{2.28}$$

$$\sigma = \frac{s_{o2}}{s_{o1} - s_{o2}} \tag{2.29}$$

$$\nu = \frac{p_{co}}{p_{coe}} \tag{2.30}$$

(where s_{o1} and s_{o2} are the adhesion coefficients for O for the 1x1 and the 1x2 surface structure, respectively).

the equation system becomes (for the one-dimensional case):

$$\varepsilon \frac{\partial u}{\partial \tau} = \varepsilon^2 \left(\frac{\partial^2 u}{\partial x^2}\right) + \nu\,(1 - u^3) - \beta\, u - \gamma\, u\, v \tag{2.31}$$

$$o_s\,\varepsilon\,\frac{\partial v}{\partial \tau} = \delta\,(a+\sigma)\,(1-u-v)^2 - \gamma\,u\,v \tag{2.32}$$

$$\frac{\partial a}{\partial \tau} = f(u) - a \tag{2.33}$$

where x is the remaining spatial coordinate.

The data sets that were examined are the output of computer simulations of the dynamics of the one-dimensional case using the equations 2.31, 2.32 and 2.33.

Chapter 3

Methods Of Data Analysis

3.1 Data Preparation

3.1.1 Filtering and Trend Elimination

In order to prepare the data for further analysis it is often necessary to subtract the mean value, eliminate 'slow' processes, trends or annual cycles. All these manipulations are special cases of filtering. Trend elimination corresponds to high pass filtering and is used if only 'oscillatory' parts of the signal are of interest. If one examines climatic temperature data and is interested only in the monthly or daily changes the annual cycles had to be removed which corresponds to applying a notch filter. The smoothing of a time series in order to reduce noise corresponds to low pass filtering. Applying a band pass filter to the data in order to select the interesting part of the dynamics is the most critical filtering operation because one decides in advance which frequencies contribute to the 'real' dynamics and spurious results may occur very likely. Of course all the filtering can be done in Fourier space but this is very often not the best idea.

In the following useful techniques are described briefly. Details can be found in Appendix A and in the text books listed in the bibliography, especially in [1].

Convolution Filters

In many cases it is more convenient to use the technique of the finite impulse response (FIR) filter (also called nonrecursive filter) which works in the time domain. This means applying a convolution with a set of coefficients b_k, $(k = 0, ..., N-1)$ and computing

$$x_{t,new} \;:=\; \sum_{k=0}^{N-1} b_k \, x_{t-k} \qquad (3.1)$$

From the transfer function[1]

$$H(\omega) \;=\; \sum_{k=0}^{N-1} b_k \, e^{-i\,k\,\omega} \tag{3.2}$$

one can see what the manipulation does in the frequency domain. Applying several coefficient sets in succession yields a transfer function which is the product of the individual transfer functions, this is due to the property of the Fourier transform that convolution products in the time domain correspond to ordinary products in frequency domain. The best known example is that of data smoothing with the coefficient set $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$ which is the convolution product $(\frac{1}{2}, \frac{1}{2}) * (\frac{1}{2}, \frac{1}{2}) = \left(\frac{1}{2}, \frac{1}{2}\right)^2$. Here the symmetric form is considered, i.e. the coefficients have indices from -1 to +1:

$$H(\omega) \;=\; \sum_{k=-1}^{k=+1} b_k \, e^{-i\,k\,\omega} = \frac{e^{+i\,\omega}}{4} + \frac{1}{2} + \frac{e^{-i\,\omega}}{4} = \frac{1}{2} + \frac{\cos(\omega)}{2} = cos\left(\frac{\omega}{2}\right)^2 \tag{3.3}$$

thus

$$H(\omega) \;=\; \left(\frac{e^{+i\,\omega/2}}{2} + \frac{e^{-i\,\omega/2}}{2}\right)^2 \tag{3.4}$$

applying $\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$ r times gives

$$H_r(\omega) \;=\; cos\left(\frac{\omega}{2}\right)^r = \left(\frac{e^{+i\,\omega/2}}{2} + \frac{e^{-i\,\omega/2}}{2}\right)^{2r} \tag{3.5}$$

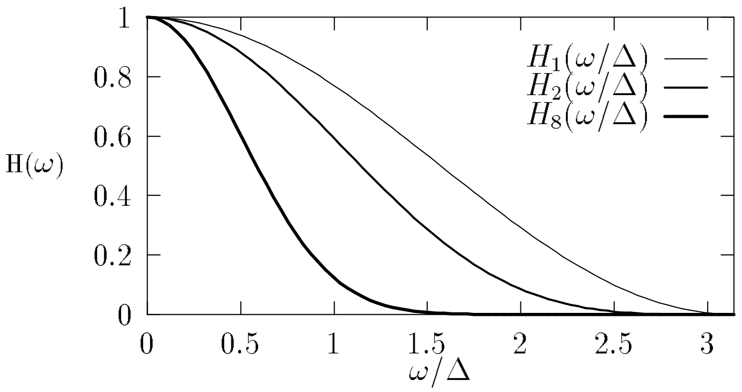

Figure 3.1: **Transfer functions for the $(\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$-filter** *when applied once, twice and 8 times. Δ denotes the sampling interval of the time series. The Nyquist frequency (see chapter 3.2.1) corresponds to $\omega/\Delta = \pi$.*

[1]The transfer function of a filter describes its effect in the frequency domain (cf. chapter 3.2.3).

The transfer functions H_r are real and hence no phase shift occurs. The extension of this method to several dimensions is straight forward, e.g. for two-dimensional data:

$$x_{new}(t_1, t_2) \;:=\; \sum_{k_1=0}^{N_1-1} \sum_{k_2=0}^{N_2-1} b(k_1, k_2)\, x(t_1 - k_1, t_2 - k_2) \tag{3.6}$$

The theory of FIR filtering is treated in many text books, some with emphasis on time series ('digital filtering in the time domain', e.g. [3] chapter 13.5, p.558) others with emphasis on two dimensional data ('digital picture processing'). By working in the time domain one saves the effort of Fourier transforming two times, which means a speedup (and programming simplification) especially for very long time series or two-dimensional data. The apparent limitation that reasonably small coefficient sets only give certain possible filter characteristics does not seem to be a real restriction in practical work. When a desired filter characteristic leads to a very 'big' coefficient set one would use filtering in the frequency domain. In that very case one might probably manipulate the data in a way that leads to spurious results in further analysis.

Methods of Trend Elimination

Trends can be eliminated by calculating the mean value of a subset of the data inside a window that is moved through the data and subtracting that 'local' mean from each value. The width of the window is an important parameter which determines the highest frequency that is removed and has to be chosen individually for each data set.

Many refinements and derived procedures exist, e.g. one can use a polynomial or a spline fit through a set of local means and subtract from each data point the value of the fitted curve.

Another possibility for trend elimination is to examine the series of first differences ($x_{n+1} - x_n, \quad n = 1...N$) instead of the time series itself. It has the advantage of the great computational simplicity, the corresponding transfer function is

$$H(\omega) \;=\; (1 - e^{-i\,\omega}) \tag{3.7}$$

The term $z\, f_0$ from the z-transform is neglected here (cf. Appendix A, formula A.6). Apart from the phase shift the effect of this operation in the frequency domain is

$$|H(\omega)| \;=\; \sqrt{(1-\cos\omega)^2 + (\sin\omega)^2} = \sqrt{2}\cdot\sqrt{1-\cos\omega} = 2\,\sin(\omega/2) \tag{3.8}$$

3.1.2 Nonlinear Transformations

The most frequently used nonlinear transformation is the logarithm. When taking the logarithm of the data set one uses the functional equation of the logarithm, ratios (orders of magnitude) are changed into differences. It is very common to plot power spectra in logarithmic scale in order to be able to see at the same time the dominating components and those with little amplitude. An exponential decay or a relaxation process is easily identified in a logarithmic plot. Before taking the logarithm the data set has to be mapped linearly into an interval with both limits greater than zero.

One can determine the nonlinearity of the operation by choosing this interval. If limits close to $+1$ are used the distortion will be small. If many orders of magnitude should become visible one would choose an interval of one of the forms $[small, 1]$, $[1, big]$ or $[small, big]$. An output symmetric to zero is achieved by choosing $[value, 1/value]$.

Signals with peak behavior can often be transformed into more symmetric oscillations. When embedded in a pseudo phase space these give a more uniform attractor reconstruction than the unmodified data.

Several properties of time series data, for example the dimension of the attractor reconstruction are conserved under nonlinear transformations. These properties may generally be determined more easily after some transformation that changes a rather nonuniform distribution of the data to a more uniform distribution. The distribution can be checked with the histogram (cf. chapter 3.2.5).

Another possible nonlinear transformation is the arctan function. It has the advantage that there are no poles for real arguments. It is also null-symmetric but the functional equation is not as simple as that for the logarithm.

3.1.3 Surrogate Data

Surrogates are 'pseudo time series' derived from a time series which share certain properties with the original while corresponding to a nondeterministic (random) process. With the classical methods of data analysis a surrogate cannot be distinguished from the data it was derived of. Surrogates are used to counter check the results of the nonlinear part of the analysis (e.g. the attractor reconstruction). For that purpose the analysis procedures are also performed with the surrogates. If similar results are found for the surrogate and the original data they have to be considered as meaningless.

A *random phase surrogate* of a given time series is a surrogate with the same power spectral density but completely random phases. It can be produced in the following way:

- take the FFT of the original time series
- change from real-imaginary to modulus-phase representation
- randomise the phases (and set phase(N–k)=–phase(k) for k=0...N/2–1, so that it is still the Fourier transform of a real time series)
- change back to real-imaginary representation
- apply the inverse Fourier transform

The resulting data corresponds to a set of N/2 (N be the length of the time series) independent harmonic oscillators of different amplitudes, i.e. to a completely random process.

Another form of surrogate is generated by a realization of an autoregressive model (cf. chapter 3.2.3) that was fitted to the time series. Again the preserved property is the power spectral density. This case corresponds to the spectrum estimation with the maximum entropy method (cf. chapter 3.2.4).

3.2 Classical (Linear) Methods

3.2.1 Fourier Transform

Definitions

The *Fourier transform* is a unitary transformation of square integrable functions (more precisely: functions of the Schwarz class).
For the function f

$$f : R^n \to C \quad \vec{x} \mapsto f(\vec{x}) \tag{3.9}$$

the Fourier transform is defined by the integral:

$$F(\vec{\omega}) := \frac{1}{\sqrt{(2\pi)^n}} \int_{R^n} f(\vec{x})\, e^{-i\,\vec{x}\,\vec{\omega}} d^n x \tag{3.10}$$

The back-transformation is:

$$f(\vec{x}) = \frac{1}{\sqrt{(2\pi)^n}} \int_{R^n} F(\vec{\omega})\, e^{+i\,\vec{x}\,\vec{\omega}} d^n \omega \tag{3.11}$$

i.e. the complex conjugate transform. For periodic functions one considers the periodicity interval L and uses the semi-continuous form of the Fourier transform:

$$c_k := \frac{1}{\sqrt{L}} \int_L f(x)\, e^{-i\,k\,x} dx \tag{3.12}$$

$$f(x) = \sum_{k=-\infty}^{k=+\infty} c_k\, e^{i\,k\,x} \tag{3.13}$$

The last expression holds for values of x where f(x) is continuous, at discontinuities the Fourier sum converges to $\frac{f(x+0)+f(x-0)}{2}$.
The following equivalent form is frequently given for the semi-continuous form:

$$a_k := \frac{1}{\sqrt{L}} \int_L f(x)\, \cos \frac{2\pi\, k\, x}{L} dx, \qquad k = 0, 1, 2, ... \tag{3.14}$$

$$b_k := \frac{1}{\sqrt{L}} \int_L f(x)\, \sin \frac{2\pi\, k\, x}{L} dx, \qquad k = 1, 2, ... \tag{3.15}$$

then back-transformation is

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos \frac{2\pi\, k\, x}{L} + b_k \sin \frac{2\pi\, k\, x}{L} \right) \tag{3.16}$$

The expansion coefficients of these two forms are related by

$$c_k = \begin{cases} \frac{a_0}{2} & (k = 0) \\ \frac{1}{2}(a_k - i b_k) & (k > 0) \\ \frac{1}{2}(a_k + i b_k) & (k < 0) \end{cases} \tag{3.17}$$

In numerical calculations one always has a discrete set of data values f_j instead of a continuous function f(x). In that case the Fourier sums have to be used:

$$c_k := \frac{1}{\sqrt{N}} \sum_{j=0}^{N-1} f_j\, e^{-i\, j\, k} \tag{3.18}$$

The backtransformation again is a 'symmetric' expression:

$$f_j = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} c_k\, e^{+i\, j\, k} \tag{3.19}$$

See [6] chapter 4.4.2.1., there are numerous textbooks that treat the Fourier transform.

Implementation: The Fast Fourier Transform

The computational effort of the direct implementation (i.e. simply calculating all N Fourier sums of N summands each) is of the order N^2. An efficient way of computing the Fourier transform is given by the so called *fast Fourier transform* (FFT) algorithm. The common implementation of the FFT is that for data sets of length $N = 2^n$ which is especially effective and simple, such a FFT implementation reduces the work to proportional $N \log_2(N)$. This is done by a recursive procedure that uses the fact that the N-th roots of unity $(\cos \frac{2\pi k}{N} + i \sin(\frac{2\pi k}{N}), k = 0 \ldots N-1)$ form a multiplicative group of order N. A reduction of the computational work is achieved whenever N has nontrivial factorization, i.e. N is not prime (cf. [6] chapter 4.4.1.3, p.616). Special implementations exist for several special cases of factorizations of N into small primes. The implementations for N not of the form 2^n are quite complicated and should be used only in very special cases. Additional computational speedups can achieved for real data and/or if only the spectral power is of interest.

For a data sets not of the size $N = 2^n$ one can either truncate to the next smaller power of two or pad zeros to the next greater number 2^n. Zero padding will in most cases produce artefacts in the spectrum. Another method is sampling up (by a noninteger 'stretch factor', cf. Appendix B.3) to the next greater number of the form 2^n and possibly reduce the number of the frequency values afterwards by the 'inverse' procedure (cf. Appendix B.4). If it is important to use all the data at hand and one wants to avoid every artefact possibly involved by the above procedures, the direct implementation of the Fourier transform is the method of choice (and a bit of waiting for the result might be necessary).

For an implementation of the FFT for $N = 2^n$ see [7] p.83, the book gives two different implementations (bit reversal before or after the main procedure, called Cooley Tukey and Sande Tukey method respectively) in a pseudo algol language that can be easily translated to any computer language. The number of the sine and cosine calculations can be reduced to one (one sine and one cosine) for arbitrary N by using the relation

$$e^{-2\pi i (k+1)/N} = e^{-2\pi i k/N} e^{-2\pi i/N} \tag{3.20}$$

and recursively computing all the sin-cos values with two multiplications and one addition for each of them.

When interpreting FFT spectra it is important to know some properties and the possible artefacts produced by the discrete Fourier transform:

The maximal frequency (*Nyquist frequency*[2]) is located in the middle of the frequency scale, the values k with $N/2 < k \leq N$ must be considered as negative frequency values, this is called 'wrap-around order' of the frequencies.

The cutting of the sampled function at the beginning and the end of the interval causes clip-effects. These can be effectively reduced by the method of 'windowing' of the data, i.e. multiplying the data with a window-function before transforming. The windowing functions are positive functions that decrease to zero at the interval border and are multiplied with the data before transforming. A sensible windowing function will not affect the power spectrum to much. The exact effect of the windowing is that the Fourier transform of the data after windowing is the convolution product of the Fourier transforms of the unmodified data and the windowing function[3]. Therefore the Fourier transform of the window function should be similar to a δ-peak. Most commonly used today is the Welch windowing:

$$Welch(n) \quad := \quad \sin \frac{n\,\pi}{N} \tag{3.21}$$

'Aliasing' means that frequencies higher than the Nyquist frequency are wrapped into the Nyquist interval. The corresponding peaks appear at the beat frequency with the Nyquist frequency. The only way to avoid misinterpretation because of aliasing is (prior) knowledge of the signal properties. Without heuristic arguments it is impossible to distinguish alias peaks from regular ones. Therefore, if possible, the signal should be low pass filtered when sampled.

Due to the form of the spectrum estimate that the Fourier transform has one must not expect to get a 'delta peak' if a finite sample of a perfect sine wave is transformed, cf. chapter 3.2.4. Instead the peaks will look like that in figure 3.2.
(Cf. [5], [3] chapter 12 or [7] chapter 2.3).

[2]In the Fourier transform of a time series with sampling rate Δ the maximal frequency is $f_{Nyquist} = 1/(2\,\Delta)$ (or $\omega_{Nyquist} = \pi/\Delta$).

[3]Cf. relation A.3.

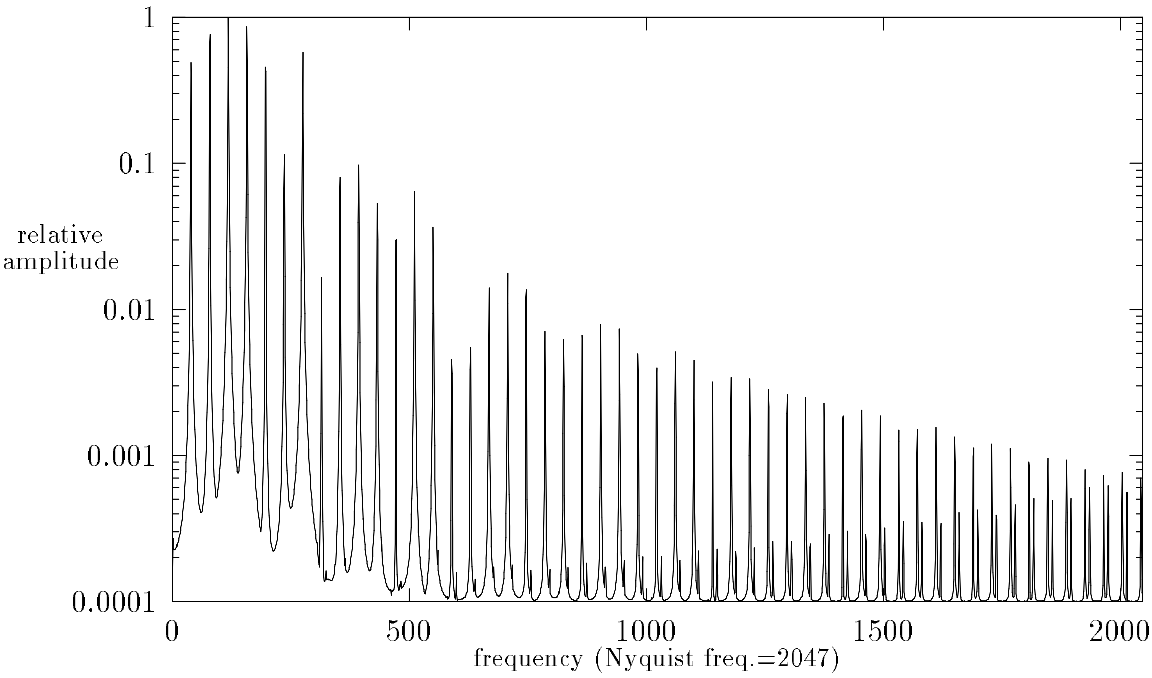

Figure 3.2: **Aliasing:** *Peaks that would lie above the Nyquist frequency are wrapped inside the spectrum in reverse order. This is the Fourier transform of the global data for the periodic case of the circuit with the heterostructure hot-electron diode (data set E, cf. chapters 2.1 and 4.1).*

3.2.2 The Autocorrelation Function

Definitions

The *cross correlation* for two random variables is defined by

$$r(x,y) := \frac{C(x,y)}{\sqrt{C(x,x)\,C(y,y)}} \tag{3.22}$$

where

$$C(x,y) := \langle x\, y\rangle - \langle x\rangle\,\langle y\rangle \tag{3.23}$$

The estimate for time average is the mean value of the data

$$\langle x\rangle := \text{time average, here: } \langle x\rangle := \frac{1}{N}\sum_{i=1}^{N} x_i \tag{3.24}$$

C(x,x) is the variance of x.

For two statistically independent random variables x,y r(x,y) is zero. The reverse statement does not hold, the correlation function measures only linear relations of the random variables, i.e. those where the distribution is concentrated around a straight line[4]. The *autocorrelation function*

$$r(\tau) = r_\tau(x,x) := \frac{C_\tau(x,x)}{C_0(x,x)}, \text{ where } C_\tau(x,x) := \langle x_i x_{i+\tau}\rangle - \langle x_i\rangle^2 \tag{3.25}$$

is a measure for the 'linear memory' of a function, $C_\tau(x,x)$ is called autocovariance of x. For a stationary process the autocorrelation function is symmetrical with respect to zero, r(0)=1, $-1 \le r(x) \le +1$ for $x \neq 0$.

Estimation Methods

Estimates for autocorrelation coefficients can be calculated directly by formula 3.25: First the autocovariance $C_\tau(x,x)$ is computed for each τ. Then all values are divided by $C_0(x,x)$ to get $r_\tau(x,x)$. It is very important that possible trends in the time series have been removed before otherwise spurious autocorrelation coefficients will occur. One must also eliminate the errors that emerge from the different number of points available for the calculation for each lag τ. For a data set of N values the following formula gives the direct estimate for the autocorrelation function after the mean value and trends have been removed:

$$r_\tau(x,x) \approx \frac{\sum_{k=1}^{N-\tau} x_k\, x_{k+\tau}}{(N-\tau)\, C_0(x,x)} \tag{3.26}$$

As the fluctuations of the estimated values are proportional to $1/\sqrt{(N-\tau)}$ and thus become big for τ close to N it is sensible to use this method of estimation only for $\tau < N/4$ (some authors say $\tau < N/10$).

The computational effort involved is proportional N^2, which is the same asymptotics as for the direct implementation of the Fourier transform. The way out is again the use of the Fast Fourier transform which reduces the computational work to proportional $N\log_2(N)$. In the following let $\mathcal{F}f$ denote the Fourier transform of f(x). The Fourier transform of the cross-correlation of two functions $f(x)$ and $g(x)$ is

$$\mathcal{F}C_\tau(f(x),g(x)) = \int\int f(x)\, g(x+\tau) dx\, e^{-i\,\omega\,\tau} d\tau = \tag{3.27}$$

(substituting $z := \tau + x, dz = d\tau$)

$$\int\int f(x)\, g(z) dx e^{-i\,\omega\,(z-x)} dz = \int f(x)\, e^{+i\,\omega\,x} dx \int g(z)\, e^{-i\,\omega\,z} dz = \tag{3.28}$$

$$= (\mathcal{F}^{-1}f)\,(\mathcal{F}g) \tag{3.29}$$

[4] An example of a quantity that vanishes if and only if two random variables are independent is the so called *transinformation* (also called *mutual information*, cf. [2] p.125): $H(x,y) := \int\int P(x,y) \log \frac{P(x,y)}{P(x)\,P(y)} dx dy$

therefore

$$C_\tau(f(x), g(x)) \;=\; \mathcal{F}^{-1}\left((\mathcal{F}^{-1} f)\,(\mathcal{F} g)\right) \tag{3.30}$$

(Cf. [3] chapter 13.2, p.545).

If $g := f$ (noting that for $f(x)$ real $\mathcal{F}^{-1}f$ is the complex conjugate of $\mathcal{F}f$) this relation becomes:

$$C_\tau(f(x), f(x)) \;=\; \mathcal{F}^{-1}\left(|\mathcal{F}f|^2\right) \tag{3.31}$$

The autocovariance function is the Fourier transform of the power spectral density. For time series this relation is often formulated as (e.g. [1] p.136 or [2] p.126 formula 6.10).

$$|\mathcal{F}f|^2 = \sum_{k=1}^{N} C_\tau(x, x)\, \cos\frac{2\pi\,\tau\, k}{N} \tag{3.32}$$

This formulation takes advantage of the fact that the imaginary part always vanishes because of the symmetry of the autocovariance function.

When implementing the estimation of the autocorrelation function via the FFT one first has to append as many zeros to the data as the biggest lag τ one is interested in. This must be done in order to avoid a wrap around effect that leads to spurious autocorrelation coefficients (without this zero padding the first part of the time series is 'correlated' backwards with the last part, cf. [3] chapter 13.1, p.541).

3.2.3 Fitting Linear Models in the Time Domain, Linear Prediction

Fitting linear models to a given time series is equivalent to an estimation of the power spectrum. As a stochastic system ('filtered noise') is completely determined by its power spectrum, all information describing the stochastic properties is contained in a linear model if we allow for enough degrees of freedom.

On the other hand linear models are blind for nonlinear interdependencies in the data: the output of the logistic map $x \mapsto r\,x\,(1-x)$ for r=4 has the autocorrelation zero for all time lags τ (except for $\tau = 0$ where it is one) and therefore appears to be white noise to every linear model.

Nonetheless it makes sense to fit linear models to data possibly having nonlinear deterministic properties. As much as (and mathematically equivalent to) looking at the spectrum one gets first information on system properties.

Moving Average Model

A *moving average model of order p* (abbreviated $MA(p)$) is defined by:

$$x_t \;=\; \sum_{n=0}^{p} b_n\, e_{t-n} \tag{3.33}$$

where e_{t-n} is the input white noise. Equation 3.33 expresses each value of a time series as the linear combination of the noise terms at p successive times. It is also called *finite impulse response* (FIR) or *nonrecursive* filter. The transfer function describing the filter characteristics is found by the z-transformation (see Appendix A):

$$X(z) = B(z)\,E(z) \quad \text{here } X(z) \text{ denotes } Z\{x_n\}$$

$$H(\omega) = \sum_{k=0}^{p} b_k e^{-i\,k\,\omega\,\Delta}$$

where Δ denotes the sampling interval, so that $\omega\,\Delta \in [-\pi, +\pi]$ or $f\,\Delta \in [-1/2, +1/2]$. This is a power series like form of an approximation for the spectral density.

Autoregressive Model

An *autoregressive model of order p* (abbreviated AR(p)) is defined by:

$$x_t \;=\; \sum_{m=1}^{p} a_m\,x_{t-m} + e_t \tag{3.34}$$

where e_t is the (white) noise added at each time, $a_m, m = 1...p$ are the coefficients that determine the properties of the model. Equation 3.34 expresses each value of a time series as a linear combination of its p predecessors plus a noise term. It also called *AR model of operator length p*, *infinite impulse response* (IIR) or *recursive* filter. Again the transfer function can be found using the z-transform:

$$X(z) = A(z)\,X(z) + E(z), \text{ so } X(z) = \frac{1}{1 - A(z)}\,E(z)$$

$$H(\omega) = \frac{1}{1 - \sum_{k=1}^{p} a_k e^{-i\,k\,\omega\,\Delta}} \tag{3.35}$$

The *characteristic polynomial* for the model is defined by

$$P(z) \;:=\; 1 - \sum_{i=1}^{p} a_i\,z^i \tag{3.36}$$

The model is called *stable* if all the roots of the *characteristic equation*

$$P(z) \;=\; 0 \tag{3.37}$$

lie outside[5] the unit circle: $|z_i| \geq 1, i = 1...p$

[5]Some authors define the characteristic polynomial as $P(z) := z^p - \sum_{i=1}^{p} d_i\,z^{p-i}$, the reversed polynomial. As the roots a the reversed polynomial are the inverse roots of the original polynomial the stability condition is then that all roots have to lie inside the unit circle.

By continued substitution a finite order AR model can always be expressed as an infinite order MA model:

$$\frac{1}{1-\sum_{k=1}^{p} a_k\, z^k} = \sum_{k=0}^{\infty} b_k\, z^k \tag{3.38}$$

Multiplication with the denominator gives

$$\begin{aligned} -1 &= (-1\, z^0 + a_1\, z^1 + a_2\, z^2 + a_3\, z^3 + a_4\, z^4 + ... + a_p\, z^p) \\ &\quad (b_0\, z^0 + b_1\, z^1 + b_2\, z^2 + b_3\, z^3 + b_4\, z^4 + ...) \end{aligned} \tag{3.39}$$

Comparing coefficients leads to a recursion for the MA-coefficients b_n:

$$b_0 = 1 \quad \text{and} \quad b_n = \sum_{k=1}^{min(n,p)} b_{n-k}\, a_k \quad \forall n > 0 \tag{3.40}$$

In the upper limit of summation[6] it was taken into account that the coefficients a_k of the AR model are zero for $k > p$.

From the relation 3.38 it can be seen that the histogram of a realization of an AR model is always a Gaussian distribution. Each term on the right hand side of the equation 3.38 contributes at each instance a random variable that has a Gaussian distribution like the input white noise. The sum of two or more random variables that have a Gaussian distribution again has a Gaussian distribution.

It is common to all linear models (i.e. filters) that without input (white noise) they produce no (stationary) output. In the case of a stable autoregressive model the initial conditions produce an exponentially decaying signal.

If an AR model shall be found for some time series one has to determine the appropriate coefficients a_m. By multiplying the equation 3.34 by $x_{t-\tau}$, taking expectation values and normalizing one gets a set of the equations called the *Yule-Walker equations*:

$$r(\tau) \;=\; \sum_{m=1}^{p} a_m\, r(\tau - m), \quad \tau > 0 \tag{3.41}$$

These equations allows the calculation of the autocorrelation coefficients $r(.)$ from the model coefficients a_m. But they also allow to fit a AR model of desired order to a time series if the autocorrelation function is known. One just has to solve the system of linear equations for the a_m.

For the right choice of the order p of the AR model one has to look at the autocorrelation function: the number of coefficients must be chosen big enough to 'cover' the characteristic part of the autocorrelation function, e.g. if it quickly decays after a few oscillations one would use as many coefficients as there are autocorrelation coefficients in these first oscillations (cf. figures 4.37 and 4.39). If the time series is oversampled, i.e. one has many autocorrelation coefficients in each oscillation, it is a good idea to use a

[6] It would normally be equal to n.

downsampled version of the data to fit the AR model in order to avoid dealing with too many AR-coefficients.

Statements about quality of the fitted model can be drawn from the spectra of realizations of the model. With white noise as input one should get a realization with the same power spectral density as the original time series. Equivalently, the residual signal, i.e. the difference signal of a realization of the model with the original time series should be white noise. If not, one must possibly try a model of greater order p.

The AR model derived by the above procedure needs not be stable, which is easily recognized by the 'exploding' realizations unstable models produce (cf. figure 4.38). For rather periodical signals, where the AR model is mathematically appropriate, because it can reproduce peaks in the power spectrum very well, the roots of the characteristic equation tend to be located near the unit circle (and may lie outside). One might 'correct' an unstable model by moving the roots z_i of the characteristic equation with $|z_i| > 1$ onto the unit circle (by the mapping $z_i \rightarrow z_i/|z_i|$) or reflecting them across the unit circle ($z_i \rightarrow 1/z_i^*$). The disadvantage of these algorithms is that one has to operate with the roots of the characteristic equation which have to be determined numerically. Especially for large models ($p > 50$) the values of the roots are sensitive to roundoff errors. (cf. [3] p.567)

For those processes with sharp peaks in the power spectral density where the AR model is good in a mathematical sense the fitted AR models tend to be unstable and one often has to move the roots of the characteristic equation.

The use of autoregressive models for signal prediction is called *linear prediction*. For that purpose one simply takes p values of the time series as initial conditions for a fitted AR(p)-model and produces the subsequent data values using the model equations. The amount of white noise[7] that has to be added at each step of the prediction[8] can be determined by a procedure derived from equation 3.38: The standard deviation σ_{AR} of an autoregressive model is equal to that of the corresponding MA-model[9] (σ_{MA}). Let σ_{noise} denote the standard deviation of the input white noise for the autoregressive model and σ_{data} that of the data. Then

$$\sigma_{data} = \sigma_{AR}\,\sigma_{noise} = \sigma_{MA}\,\sigma_{noise} \tag{3.42}$$

Thus

$$\sigma_{noise} = \frac{\sigma_{data}}{\sigma_{MA}} \tag{3.43}$$

The amplitude of a realization of a moving average model is proportional to the square root of the sum of the squares of its coefficients.

I suggest using relation 3.40 to compute this sum recursively up to a finite number of summands. A C-code implementation of that procedure is given in Appendix B.2.

[7]On a computer white noise may be produced from pairs of evenly distributed random numbers R_1 and R_2: Two independent random numbers z_1 and z_2 with Gaussian distribution are given by $z_1 = \sqrt{-2\,\log(R_1)}\,cos(2\,\pi\,R_2)$ and $z_1 = \sqrt{-2\,\log(R_1)}\,sin(2\,\pi\,R_2)$.

[8]the term e_k on the right hand side of equation 3.34

[9]Both models with white noise with standard deviation one as input.

This procedure[10] gives more precise results[11] than the direct estimation procedure that measures the standard deviation of a realization of the AR model with white noise with standard deviation one as input. It can also be used to determine whether a given AR model is stable. A necessary condition for the stability of the AR model is that the sum of the squares of the MA coefficients converges. Recursive calculation of the MA coefficients and observing whether they decrease towards zero gives a strong hint on the stability of AR model without having to solve its characteristic equation.

The accuracy of a prediction using an autoregressive model is determined by the autocorrelation function of the data. Less (linear) correlated regions can only be predicted less accurate by a linear model. Individual realizations often appear more like a surrogate, less like a prediction. In practice linear prediction is used for voice transmission and speech synthesizer chips, the voice data is reduced to a set of a few AR coefficients that have to be corrected several times each second, this method is called *linear predictive coding* (cf. [3] p.571). In this way the data transfer rate necessary for voice transmission can be reduced drastically.

Autoregressive Moving Average Model

An *autoregressive moving average model* of order p,q (ARMA(p,q)) is defined by

$$x_t = \sum_{m=0}^{p} a_m \, x_{t-m} + \sum_{n=0}^{q} b_n \, e_{t-n} \tag{3.44}$$

This is a combination of AR and MA models and specializes to these for q=0, p=0 respectively. The z-transform of the output

$$X(z) \quad = \quad A(z)\,X(z) + B(z)\,E(z) \tag{3.45}$$

solved for $X(z)$

$$X(z) \quad = \quad \frac{B(z)}{1 - A(z)} E(z) \tag{3.46}$$

gives the transfer function

$$H(f) \quad = \quad \frac{\sum_{k=0}^{q} b_k e^{-i\,k\,\omega\,\Delta}}{1 - \sum_{k=1}^{p} a_k e^{-i\,k\,\omega\,\Delta}} \tag{3.47}$$

(Δ is again the sampling interval). The Nyquist interval is $-1/2 < f\,\Delta < +1/2$ (or equivalently $-\pi < \omega\,\Delta < +\pi$). The transfer function will vanish at zeros of the AR term (B(z)=0) and diverges at poles (A(z)=1) of the AR term, unless these are cancelled by a vanishing numerator.

[10] Note that this procedure is equivalent to making a prediction with the AR model without noise for the initial values (0,0,...,0,1) and then computing the standard deviation of the prediction.

[11] It is also faster.

The underlying form of a power spectrum estimation is that of a Padé approximation. A *Padé approximant* for some function $f(x)$ is a rational function $R(x)$ of the form

$$R(x) = \frac{\sum_{k=0}^{q} b_k\, x^k}{1 - \sum_{k=1}^{p} a_k x^k} \tag{3.48}$$

if its power series expansion agrees with that of the function to the highest possible order.

Generally Padé approximants, compared to power series, tend to give superior approximations to a function. The determination of the a_k and the b_k for a given time series is equivalent to making a Padé approximation to the Fourier transform. (For Padé approximants see [3] chapter 5.12, p.200).

A variant of the ARMA model is the *autoregressive integration moving average* (ARIMA) model, here an ARMA model is fitted to the first differences of the time series. A realization of the ARIMA model is given by the sum of the values of the ARMA realization.

3.2.4 Spectrum Estimation Techniques

Fourier Transform

Consider the mapping $z := e^{i\,\omega\,\Delta}$ which maps the Nyquist interval $[-\pi/\Delta, +\pi/\Delta]$ to the unit circle. The true spectrum of an infinite signal may be written as a Laurent series:

$$P(\omega) \quad = \quad \left| \sum_{k=-\infty}^{+\infty} c_k\, z^k \right|^2 \qquad \text{where} \quad z := e^{i\,\omega\,\Delta} \tag{3.49}$$

The Fourier Transform spectrum estimate derived from a finite data sample of the length N has the form of a truncated Laurent series:

$$P(\omega) \quad \approx \quad \left| \sum_{k=-N/2}^{N/2-1} c_k\, z^k \right|^2 \tag{3.50}$$

It is inherent to this form that it can have zeros but no poles in the frequency range. Therefore this method of spectrum estimation is called *direct method, all-zero model* or *moving average model* [12]. Due to the existence of the fast Fourier transformation (FFT) algorithm this the most widespread technique used for estimation of the power spectrum.

Maximum Entropy Method

Another form of a power spectrum estimation is that of a rational function with constant numerator, the *maximum entropy spectral estimation* (MES).

$$P(\omega) \approx \frac{1}{\left|1 - \sum_{k=1}^{M} a_k\, z^k\right|^2} \qquad \text{where} \quad z := e^{i\,\omega\,\Delta} \tag{3.51}$$

This approximation can have poles (i.e. infinite power spectral density) for at most M frequencies. It is called *all-poles model, maximum entropy method* (MEM), *maximum entropy spectral estimation* (MES), *autoregressive model* [13] (AR). The name maximum entropy method is due to a condition satisfied by the coefficients a_k:

$$\frac{\alpha}{\left|1 - \sum_{k=1}^{M} a_k\, z^k\right|^2} \quad \approx \quad \sum_{k=-M}^{+M} r(k)\, z^k \qquad \alpha \in \mathbb{R} \tag{3.52}$$

The real parameter α is needed make the expression on the left hand side equal to one for $z = 0$. In principle one could determine the a_k by that equation, practically one would rather use the Yule-Walker equations (3.41). The approximation sign has to be understood that way that the power series for the left hand side of the equation must agree

[12]The theory of linear models is equivalent to that of the estimation of power spectra, cf. 3.2.3 .

[13]Dito, cf. 3.2.3 .

to the right hand side for the powers of z from -M to M. Among all possible extrapolations this one has the maximum entropy in a 'definable information-theoretic sense' (cited from [3] p.574), i.e. the least possible constraints are put on the continuation.

The MEM is mathematically especially good if the signal consists mainly of periodical parts corresponding to sharp peaks in the spectrum. Unfortunately, in this very case the AR model fitted tends to be unstable (cf. chapter 3.2.3). Therefore the MEM is a good technique to find the locations of sharp spectral peaks. It is no good procedure to determine the energy within the oscillations corresponding to these peaks. Practically, this technique seems to be of great importance if the data consists of only few samples (cf. [42]).

ARMA Related Spectrum Estimation

For time series with both peaks and continuous parts in the spectrum one might well suppose that a spectrum estimation corresponding to the ARMA-model would be the best choice. Analogous to superiority of function approximation using a Padé series the results should be better compared to a purely AR or MA based spectrum estimation with the same number of coefficients. The fact that the ARMA form is not commonly used is due to the greater difficulties in fitting a Padé series and the many possible choices for a special Padé series.

Wavelets

The *wavelet transform* is a linear orthogonal transformation (a rotation in function space) that shares some properties with the Fourier transform. Like the FFT there is a 'fast' algorithm for the transformation for data of the length 2^n. The back-transform is also very similar to the original transform. If written as a matrix it is just the transposed matrix (for the Fourier transform it is the complex conjugate matrix). The basis functions (called *wavelets*) have the property of being localized in space. It is an especially good choice for signals with structures on many different scales (a 'fractal' structure).
(See [3] chapter 13.10 p.591).

3.2.5 Invariant Density (Histogram)

The *histogram* (equivalent to the invariant density in the case of an ergodic system) of a time series $x(t), \quad t = 0...N$ may be defined by

$$h(y) := \sum_{\{t:x(t)\in[y-\epsilon,y+\epsilon]\}} 1 \qquad \forall y \in Y \tag{3.53}$$

where Y is a set of equidistant[14] values (the 'boxes' in the histogram-plots) ranging from the global minimum of the data to the global maximum. This is the 'distribution statistics'

[14] Their distance is $2\,\varepsilon$.

of the data sample. A histogram of a purely random process is the Gaussian distribution (the sum of many distributions of arbitrary type is the Gaussian distribution). Histograms allow first statements about the data, e.g. an exponential falling slope may indicate that it is useful to take the logarithm of the data.

3.3 Attractor Reconstruction and Geometrical Methods

3.3.1 Delay Coordinates

Stochastic and chaotic systems can have the same power spectral density (and hence autocorrelation function), so they are indistinguishable for the classical methods. A property that distinguishes deterministic systems obeying a dissipative dynamics is that they possess an attractor in phase space. The trajectories of the system approach the attractor and after a typical transient time their trajectories lie arbitrary close to it. If one could find the attractor of a certain system then it would be successfully distinguished from a stochastic system. The method allowing to reconstruct the attractor is that of the *delay coordinates*. Consider the set of points $\vec{X}$:

$$\vec{X} = \{x(t), x(t+\tau), x(t+2\tau), ..., x(t+(m-1)\,\tau)\} \qquad \forall t \tag{3.54}$$

The plots produced with delay coordinates are often called *phase portraits* or *time-delay plots*. For a continuous signal this defines a trajectory in an abstract space. The so called *pseudo phase space* which is of dimension m. By a theorem of Takens ([32]) this trajectory is topological equivalent to the real attractor for allmost all values of the delay τ and of the embedding dimension m with

$$m \;\geq\; 2\,D_H + 1 \tag{3.55}$$

(where D_H denotes the Hausdorff dimension[15] of the real attractor). One only needs one dynamical variable to reconstruct the dynamics of a system with several degrees of freedom (cf. [8] chapter 4.4.1).

As a trivial example consider the harmonic oscillator

$$x(t) \;=\; sin(\omega\, t) \tag{3.56}$$

If one plots the curve

$$\vec{X}(t) \;=\; \{\sin(\omega\, t), \sin(\omega\,(t+\tau))\} \tag{3.57}$$

in the 2-dimensional pseudo phase space one generally obtains ellipses which are only in those special cases degenerated to straight lines where the delaytime is a integer multiple of the half period.

[15]Cf. chapter 3.4.2 .

The choice of the delaytime τ is essential, generally it has to be a fraction (1/10...1/3) of the slowest frequency of the system. For discretely sampled data it is of course limited to multiples of the sampling interval and one only gets a cloud of points in the pseudo phase space. Connecting these points to get a phase portrait may be difficult if the sampling interval is not much smaller than the typical oscillation time of the system. If there are not many points at all or the data is noisy this may even be entirely impossible.

Typical reconstructions are distorted but the interesting invariant properties like the attractor dimension, Lyapunov exponent and Kolmogorov entropy are invariant to the distorting transformation and can in principle be estimated by certain algorithms.

However, one must be very careful in interpreting attractor reconstructions and especially the depending quantities like the attractor dimension.

(See [2] chapter 6.3, [11] chapter 5.3).

3.3.2 Poincaré Section, Maximum Map, Amplitude- and Period Contours

Poincaré sections are sets of intersection points of the attractor with a plane in phase space. The continuous dynamics in phase space is mapped onto a discrete dynamics. Let $\vec{P}_n$ be a sequence of points in the order that is induced by following a certain trajectory on the attractor. For a deterministic system one can write

$$\vec{P}_{n+1} = f(\vec{P}_n) \tag{3.58}$$

where f is defined by the nonlinear mapping

$$\vec{X}(t+T) = \vec{X}(t) + \int_t^{t+T} \vec{F}(\vec{X}(\tau))\, d\tau \tag{3.59}$$

if the differential equations for the system are

$$\dot{\vec{X}}(t) = \vec{F}(\vec{X}(t)) \tag{3.60}$$

The properties of discrete nonlinear mappings (examples are the Henon map, the logistic map, the tent map, the circle map and the Bernoulli map) are well established[16], and comparing the easy depictable (2-dimensional) plots with the wellknown discrete mappings can give valueable information about the attractor. Iterated nonlinear mappings that satisfy quite general conditions have some remarkable universal properties that were discovered by Feigenbaum (see [33] or [34]).

In many practical cases the Poincaré sections are helpful to estimate the dimension of the attractor. This is due to the fact that the cloud of points generated by the Poincaré section is an object with the dimension of the attractor minus one. Theoretically a limit cycle should lead to a finite set of single points[17]. A torus should give a closed one

[16] Cf. [9] p.212.
[17] Typically only a few.

dimensional curve and a chaotic attractor corresponds to a fractal set of points. Because of the finite sampling interval it may be difficult to distinguish these cases in practice. Of course attractor dimensions greater than three cannot be identified in two dimensional Poincaré sections.

Among the Poincaré sections those variants that can be produced directly from the data without having to do any operations in phase space are of greatest practical interest.

The first and most important of them is the *maximum map* (or *first return map*). The set of the local maxima (or minima) of the time series is determined and one plots extremum versus next extremum for all local extrema in the time series. Correction of the extremum values by interpolation improves the results in many cases.

Maximum maps for periodic signals often give artefacts looking similar to to the one shown in figure 3.3. Instead of the single point that one expects a beat effect with the sampling frequency leads to a meaningless crescent like shape. The separation into two parts of maximum map using interpolated extrema in figure 3.3 is due to the unsymmetrical form of the peaks. Figure 3.4 shows the same type of artefact for a perfect sine wave. The deviations from the theoretical value[18] can be decreased with quadratic interpolation. As the terms in the Taylor expansion around the extrema which are of higher than second order become more important at a greater distance from the maxima it is also clear that lower sampling rates lead to greater errors in the maximum map.

As an example for the maximum map consider figure 4.10. The dimension of the point set is apparently close to one. In fact the underlying chaotic dynamics has an attractor of dimension slightly greater two. In this case the shape of the point cloud suggests to be considered as the plot of the function (defined by equation 3.59) that is iterated by the dynamics. The degree of unpredictability is determined by the slope of the function. Derivatives greater than one are equivalent to error amplification.

In the case of a circular shaped point set one would identify the dynamics with that of a circle map (cf. [11] chapter 6.2).

[18]The value of the maximum should be equal to one, leading to a maximum map consisting of the single point (1,1).

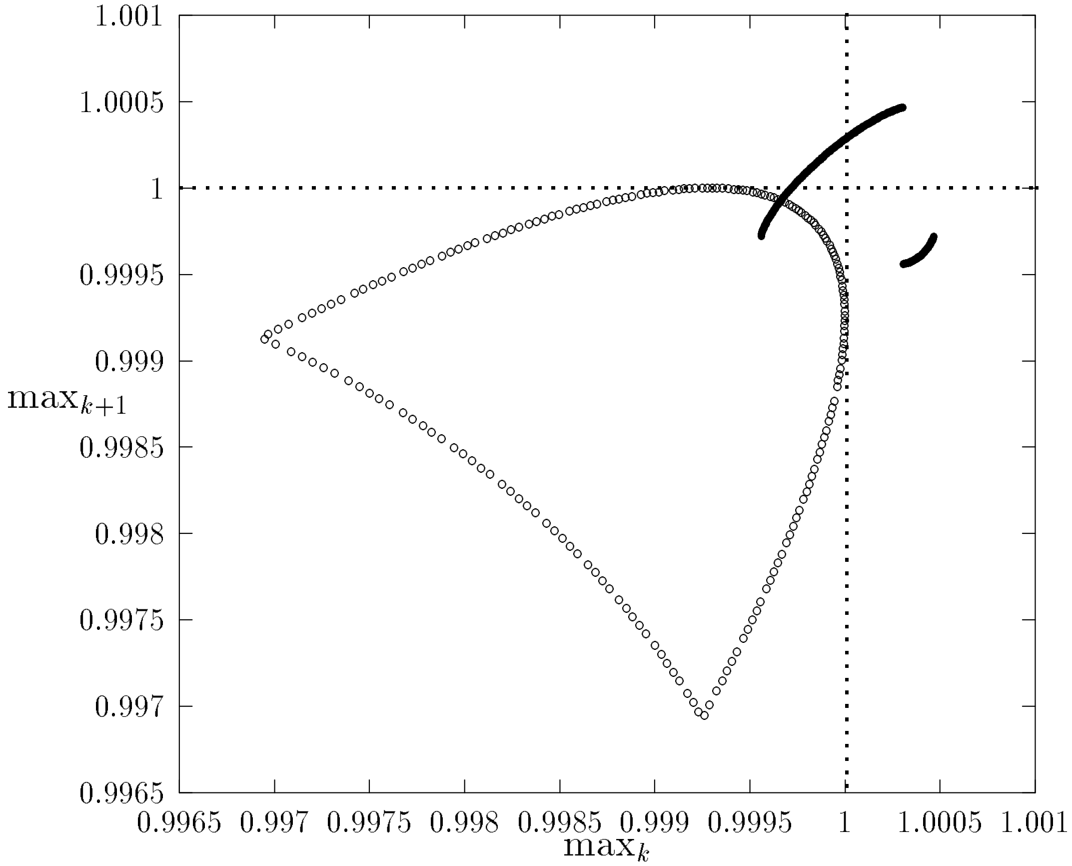

Figure 3.3: **Spurious maximum maps for the spikes in the global data set E.** *Local maxima were used for this plot. Quadratic interpolation was used for the two parts close to (1,1) where the individual points are not resolved. Their separation is due to the non-symmetric form of the peaks.*

A *maximum contour* is the graph of the amplitudes of the successive extrema of the time series, a *period contour* is the graph of the duration of the successive periods (between extrema or between zero crossings) of the time series. The degree of frequency modulation m_f of the signal can be read off the period contour, it is defined as

$$m_f \;:=\; \frac{p_{max} - p_{min}}{p_{max}} \tag{3.61}$$

where p_{min} and p_{max} denote the minimum- and maximum value of the period. The amplitude modulation m_a is defined analogously

$$m_a \;:=\; \frac{a_{max} - a_{min}}{a_{max}} \tag{3.62}$$

where a_{min} and a_{max} are the minimum- and maximum value of the amplitude. These definitions only hold for nullsymmetric signals, then the modulations as defined above measure the ratio of the biggest with the smallest amplitude (period). For unsymmetrical signals the contours of the minima and the contours of the maxima may lead to different values for the modulations.

Correction of the positions used for the period contour is often essential, without one can only obtains integer period length. The formulae for the quadratic interpolation of an extremum x_i are
for the value of a extremum

$$x_{int} \;=\; x_i + \frac{(x_{i-1} - x_{i+1})^2}{8\,(x_{i-1} + x_{i+1} - 2\,x_i)} \tag{3.63}$$

and its position i

$$i_{int} \;=\; i + \frac{(x_{i-1} - x_{i+1})}{2\,(x_{i-1} + x_{i+1} - 2\,x_i)} \tag{3.64}$$

Other types of interpolation such as spline interpolation can of course also be used.

The maximum contour or the period contour may also be used to eliminate the main periodicity of data. E.g. fitting an autoregressive process often leads to divergent models if the time series is highly periodical, fitting an accurate model to the maximum contour is a good idea in this case.

For the determination of local extrema many procedures are possible, noisy data may need low-pass filtering before. A sensible condition for considering a value as local maximum is that it has to be the maximum among a certain number of neighboring points. A threshold to select only extrema beyond a certain value, or analogous for the periods a minimum length to select the slowest oscillation may be appropriate in some cases.

For certain data sets it can be advisable to apply a two-step procedure: First the local extrema are determined for the data set after strong smoothing. Then the extrema are searched in the original data, but only in the regions of the extrema in the smoothed data.

A *stroboscopic map* is produced by taking the value of one dynamic variable at periodic time intervals. It is especially useful for periodically driven systems. In this case the time interval would be chosen equal to the period of the driving force.

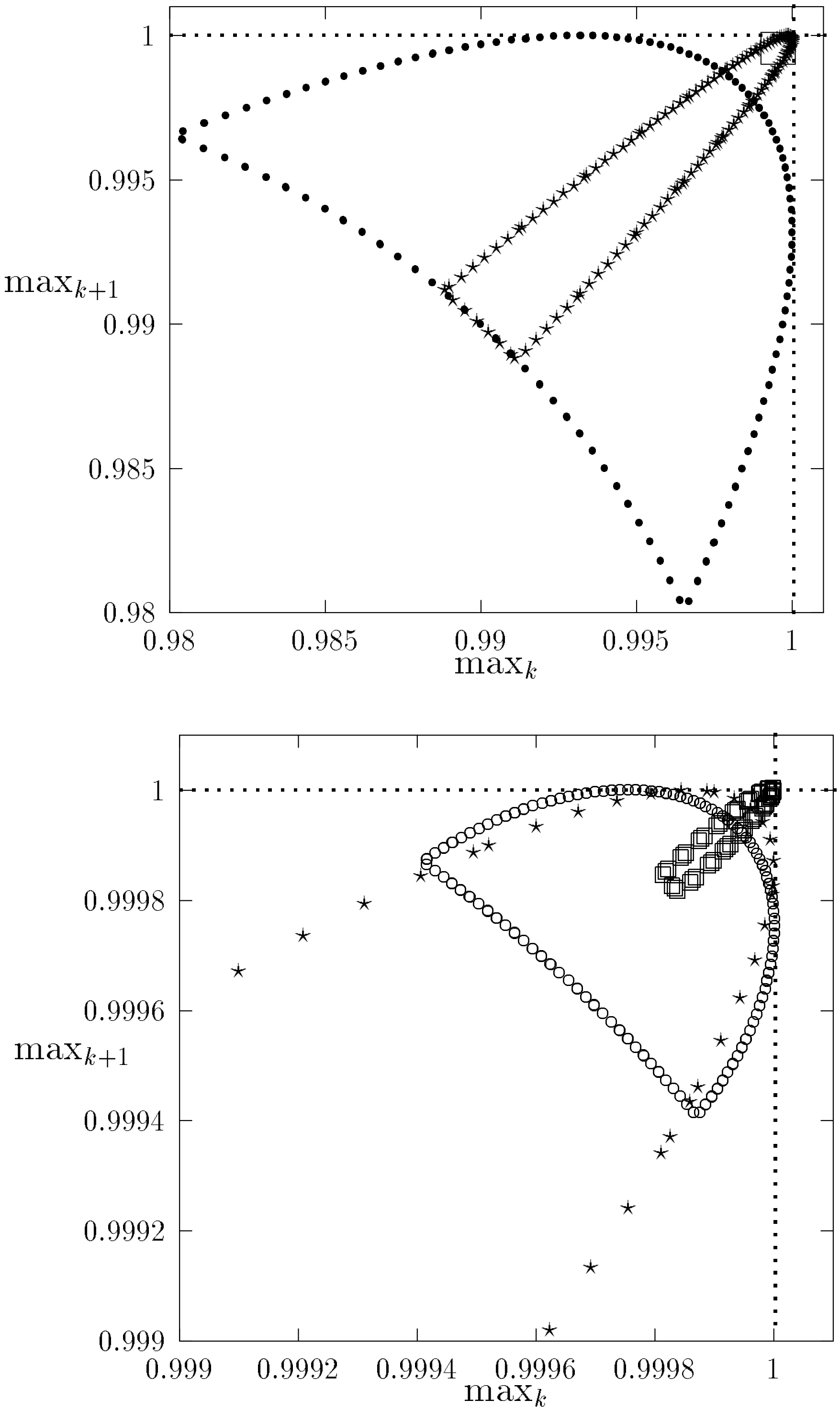

Figure 3.4: **Spurious maximum maps for a sine wave for different sampling rates. Top: Non-interpolated extrema. Bottom: Detail (box in the upper plot) with interpolated extrema.**
asterisks: non-interpolated maxima, sampling rate 0.3,
disks (●): non-interpolated maxima, sampling rate 0.4,
boxes (□): interpolated maxima, sampling rate 0.3,
circles (○): interpolated maxima, sampling rate 0.4 .

3.4 Analysis of Invariant Properties of the Attractor

3.4.1 Estimation of Lyapunov Exponents

Definition

Consider a point in phase space obeying the dynamics of a certain system. A small ball B around that point will be deformed and translated by the dynamics in a way that can for infinitesimal times be approximated linearly as the action of a matrix plus an offset.

$$\vec{x}_{t+dt} \approx \mathbf{A}\,\vec{x}_t + \vec{b} \qquad \forall \vec{x} \in B \tag{3.65}$$

Evolution for short times is approximated by repeated application

$$\vec{x}_{t+k\,dt} \approx \mathbf{A}^k\,\vec{x}_t + \vec{c} \qquad \text{where } \vec{c} = \sum_{i=0}^{k-1} \mathbf{A}^i\,\vec{b} \tag{3.66}$$

This is the viewpoint of linear stability analysis. In the general case the length of some axis will increase exponentially while others decrease exponentially in time. At the same time the ellipsoid is rotated in phase space. The exponents that correspond to the rate of growth of the principal axes are called *Lyapunov exponents* of that system. For a conservative system the sum of the Lyapunov exponents is zero, the volume of the ball is conserved (this is Liouville's theorem). For dissipative systems this sum is smaller than zero.

One definition of a chaotic system is that one Lyapunov exponent must be greater zero, which implies the exponential growth of little differences in the initial conditions for the system. Because of the finite accuracy of the values for the initial conditions this leads to the well known unpredictability of any trajectory of a chaotic system after some time (cf. weather forecasting).

Estimation Procedures

For a system where the law of motion (e.g. the differential equations) is known the algorithm to find the Lyapunov exponents is straight forward: One just has to follow the motion of the epsilon sphere (in a computer simulation) and observe the growth of the principal axis. For a system of differential equations

$$\dot{\vec{X}}(t) = \vec{F}(\vec{X}(t)) \tag{3.67}$$

the set of Lyapunov exponents could in principle be extracted from the Jacobian matrices of the flux vector field $\vec{F}(\vec{X})$.

For time series data the situation is more complicated because the law of motion and the attractor are unknown. The deformation of the ε-sphere has to be simulated by following the trajectories of some points on the attractor that lie close to each other and form approximately a spherical region. From the short time evolution of this point set the Jacobian can be extracted with a least squares fit. After some way on the trajectory a new point set has to be chosen because the distance between them is no longer small enough and the region is not spherical anymore. The difficulties of the implementation arise mainly from the restriction in the choice of the points.

The set of Jacobian matrices and offset vectors can also be used to make predictions for the system, so called *local linear predictions* (cf. [30], [49], [31]) or noise reduction (cf. [40]).

If one is only interested in the largest Lyapunov exponent the procedure is easier: One has to choose only two points and simulate their evolution, after a certain time replacing them by two new points that lie apart in the same direction but with smaller distance. Due to the automatic selection of the most unstable direction this procedure always measures the corresponding (greatest) Lyapunov exponent.

The paper [24] gives FORTRAN programs for the determination of the greatest Lyapunov exponent and for the Lyapunov spectrum. A comparison of different algorithms for the determination of Lyapunov exponents from experimental data can be found in [25]. See [36] for the determination of the Lyapunov spectrum for various model systems. A program for the determination of Lyapunov exponents was written by E.J.Kostelich, for the documentation see [50]. For the limitations for determining Lyapunov exponents see [29].
(See also [2] chapter 6.5, [8] chapter 4.4.2, [36], [45], [46] and [28].)

3.4.2 Estimation of Attractor Dimensions

Definitions

The *capacity* $D_K(A)$ (also called *fractal dimension* denoted by D_F) of a set $A \subset R^n$ is defined as:

$$D_K(A) \quad := \quad -\lim_{\delta \to 0} \frac{\log N(\delta, A)}{\log \delta} \tag{3.68}$$

where $N(\delta, A)$ is the smallest number of $\delta-$cubes that are needed for a complete covering of the set A.

Let $\beta(\delta) := \{B_i\}$ be a countable family of subsets $B_i \subset R^n$ such that $\delta_i := \sup\{|x - y||x, y \in B_i\} \leq \delta$ holds for all B_i in $\beta(\delta)$ and $\bigcup_i B_i \supseteq A$. Then the Hausdorff measure (Hausdorff, 1919) of A for the dimension d is defined as

$$m_d \quad := \quad \lim_{\delta \to 0} \inf_{\beta(\delta)} \sum_i \delta_i^d \tag{3.69}$$

The *Hausdorff dimension* (or *Hausdorff-Besicovitch dimension*) is defined as the real number $D_H(A)$ for which

$$m_d(a) \;=\; \begin{cases} 0, & if \quad d > D_H(A) \\ \infty, & if \quad d < D_H(A) \end{cases} \tag{3.70}$$

The greater number of possible sets allowed for the covering in the definition of the Hausdorff measure has the effect that

$$D_H \;\leq\; D_K \tag{3.71}$$

But for 'usual' attractors one can expect both dimensions to coincide ([35]).

Consider a partition of the phase space into boxes of diameter ε, let $N(\varepsilon)$ be the number of boxes that contain trajectories of the system (i.e. those with relative frequency $p_i > 0$).

Then the *Renyi entropies* (or *generalized entropies*) are defined by

$$I_q(\varepsilon) \;:=\; \frac{1}{1-q}\log(\sum_{i=1}^{N(\varepsilon)} p_i^q) \tag{3.72}$$

These allow the definition of the *Renyi dimensions* (or *generalized dimensions*):

$$D_q \;:=\; \lim_{\varepsilon\to 0}\frac{I_q(\varepsilon)}{|\log\varepsilon|} \tag{3.73}$$

The D_q are generally monotonically decreasing with q, the dependence on q measures the inhomogeneity of the attractor.

The case $q = 1$ must be excluded, but the limit $q \to 1$ exists

$$\lim_{q\to 1} I_q(\varepsilon) \;=\; I(\varepsilon) := -\sum_{i=1}^{N(\varepsilon)} p_i \log(p_i) \tag{3.74}$$

This quantity gives the amount of information one gains by measuring with accuracy ε. The corresponding dimension is

$$D_1 \;=\; \lim_{\varepsilon\to 0}\frac{I(\varepsilon)}{|\log\varepsilon|} \tag{3.75}$$

D_1 counts the number of bits by which $I(\varepsilon)$ increases if the measuring accuracy is doubled.

The Renyi dimension for $q = 0$ is the fractal dimension, for $q = 2$ it is the so called *correlation dimension*.

Let the Lyapunov exponents of a system be sorted in decreasing order, i.e.

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_f \tag{3.76}$$

Then the *Lyapunov dimension* (or *Kaplan-Yorke dimension*) is defined as

$$D_L \;:=\; j + \frac{\sum_{i=1}^{j}\lambda_i}{|\lambda_{j+1}|} \qquad \text{where } \sum_{i=1}^{j}\lambda_i \geq 0 > \sum_{i=1}^{j+1}\lambda_i \tag{3.77}$$

The Lyapunov dimension is an upper bound for the fractal dimension:

$$D_L \quad \geq \quad D_1 \tag{3.78}$$

For a chaotic attractor it is always greater than 2, because $j \geq 2$: there is always one $\lambda_i = 0$ corresponding to perturbations along the trajectories, and one with $\lambda_i > 0$ as the system is chaotic.
(Cf. [12] chapter 3.2, p.71, [10], chapter VI.4).

Dimension Estimation

There are two types of algorithms for the estimation of dimensions for the attractor underlying some time series data: The first type uses the *local densities*, one makes a statistics about the probability to find a point in an ε-sphere for different ε around some fixed point. The alternative is the *nearest neighbor statistics*, a statistics about the typical distance to the k-th nearest neighbor of a point in phase space. In both methods the computation is repeated for several points of the attractor. The final result is the average of the values obtained for the set of points. These methods are sometimes referred to as the method of 'fixed size' and 'fixed mass', respectively.

local density

For a point set, part of an attractor of dimension D, the relation

$$\langle n(\varepsilon) \rangle \quad \propto \quad \varepsilon^D \tag{3.79}$$

should hold over 'many orders of magnitude'. $n(\varepsilon)$ denotes the number of points in an ε-sphere around some point of the attractor. For finite point sets there exists a lower limit for ε where the relation must fail: the typical distance to the nearest neighbor. In the other direction there will be a deviation if ε is in the order of the diameter of the attractor.

For typical data where the dimension can successfully be estimated one would try to find a segment of constant slope D in a plot $n(\varepsilon)$ versus $\log(\varepsilon)$ for some small value for ε up to 1..15% of the attractor diameter. The range of the parameter ε where 3.79 approximately holds is called the 'scaling region'. The computation has to be done for different embedding dimensions. There should be a range of embedding dimensions, beginning from the minimum number of dimensions necessary for a correct attractor reconstruction (cf. formula 3.55) up to some maximum value where the estimated value of D does not change. If there is a significant dependence between embedding dimension and D it is likely that D is a spurious value. The maximal accessible attractor dimension is mainly determined by the number of data points.

The distance in relation 3.79 may be calculated using any norm, for numerical calculations the maximum norm is most convenient.

Different averaging procedures used for the calculation of left hand side of relation 3.79 lead to different types of dimensions:
The *generalized mean* defined by

$$M_q(\varepsilon) \;:=\; \left(\frac{1}{N}\sum_{i=1}^{N} n_i^{q-1}(\varepsilon)\right)^{\frac{1}{1-q}} \tag{3.80}$$

allows a generalization of 3.79 for the Renyi dimensions D_q

$$M_q(\varepsilon) \;\propto\; \varepsilon^{D_q} \tag{3.81}$$

The harmonic mean leads to the fractal dimension or capacity D_0:

$$\langle n(\varepsilon)\rangle_{harm} \;\propto\; \varepsilon^{D_0} \tag{3.82}$$

The geometric mean leads to D_1:

$$\langle n(\varepsilon)\rangle_{geom} \;\propto\; \varepsilon^{D_1} \tag{3.83}$$

The problem of using the harmonic and the geometric mean is that any vanishing density leads to the mean zero.
Dimensions computed with the arithmetic mean are correlation dimensions:

$$\langle n(\varepsilon)\rangle_{arith} \;\propto\; \varepsilon^{D_2} \tag{3.84}$$

These give a lower limit for D_1.

For inhomogeneous attractors there might be no scaling regions because each typical part of the attractor has its own scaling region. The process of averaging then leaves no scaling region at all.

(See [2] chapter 6.4).

nearest neighbor statistics

Let $\langle r_k\rangle$ be the average distance of a given point on the attractor to its k-th nearest neighbor. Then for a set of N points lying on an attractor of dimension D the relation

$$\langle r_k\rangle \;\propto\; G(k,D)\,k^{1/D}\,N^{-1/D} \tag{3.85}$$

should hold in some scaling region of k and N. The function $G(k,D)$ is a correction for small values of k which is very close to 1 for $k>5$:

$$G(k,D) \;:=\; k^{-1/D}\,\frac{\Gamma(k+\frac{1}{D})}{\Gamma(k)} \tag{3.86}$$

If in a log-log plot of $\langle r_k \rangle / G(k, D)$ versus k, N or k/N there is a region of constant slope one can read off this slope as estimation value for the dimension. Again possible dependencies of D on the embedding dimension indicate spurious results. The dimension of inhomogeneous attractors can in contrast to the method of local densities be successfully estimated. The relation 3.85 can be generalized for the moments of r_k (cf. [48]):

$$\langle r_k^{\gamma} \rangle \quad \propto \quad G(k, D(\gamma))\, k^{\gamma/D(\gamma)}\, N^{-\gamma/D(\gamma)} \tag{3.87}$$

$D(\gamma)$ is a dimension function that specializes to D for $\gamma = 1$. For $D(\gamma)$ and D_q the following statements hold

$$D(\gamma) = D_q \quad \text{if} \quad q = 1 - \frac{\gamma}{D(\gamma)} \tag{3.88}$$

and

$$D(\gamma) = D_q \quad \text{if} \quad \gamma = (1 - q)\, D_q \tag{3.89}$$

From 3.88 it can be seen that D is D_q for some value of q between 0 and 1, depending on the value of D.
(See [22], [44] and [47]. Cf. also the documentation for the software package for the dimension estimation using a nearest neighbor algorithm by E.J.Kostelich, [50]).

When extracting estimations for the attractor dimension from time series data one always has two avoid to use two points from the same trajectory in each step of one of the above procedures. Using them would result in estimates that are too low. This restricts the maximum distance one can take for the calculations if there is no special procedure that explicitely removes the points of the trajectory where the given point was taken from.

It is always necessary to be 'familiar' with the data and look at Poincaré sections and attractor reconstructions before one tries to make dimension estimates. Figure 4.1.2 shows an example of the output of a dimension estimation procedure using nearest neighbor statistics. The system is known to be chaotic and the dimension is close to two. All values values obtained are smaller than two which cannot be true for a chaotic system. These completely wrong dimension estimates for a system without noise where it is known that the attractor dimension is very low again should show that the output of such procedures must be interpreted with greatest care.

For experimental data where noise is always present it is common that the numerical value from such procedures depends on both the delay time and the embedding dimension. Smoothing procedures applied before the dimension estimation introduce one more unknown quantity: By smoothing the data strongly enough one can obtain arbitrary dimension estimates between one and the 'real' value.
(For the limitations for estimating attractor dimensions see [29]).

3.5 Methods for the Analysis of Spatio-Temporal Data

3.5.1 Classifying Spatio-Temporal Behavior

For spatio-temporal data there are of course more different types of dynamics possible than in the purely temporal case, a first coarse classification might be:

spatial coherence	coherence in time	
yes	yes	periodical case
yes	no	chaos in time
no	yes	'frozen' chaos
no	no	turbulence

Table 3.1: Types of spatio-temporal dynamics

3.5.2 Empirical Orthogonal Functions

Let a 'movie' (space-time data) be written as X×T-matrix, either as column of T 'photos'[19] $\vec{p}_i$ or as row of X columns $\vec{a}_k$, each describing the dynamics at one location:

$$\mathbf{A} := \left(\vec{a}_1, \vec{a}_2, ..., \vec{a}_X\right) = \begin{pmatrix} photo(1) \\ photo(2) \\ \vdots \\ \vdots \\ \vdots \\ photo(T) \end{pmatrix} = \begin{pmatrix} \vec{p}_1 \\ \vec{p}_2 \\ \vdots \\ \vdots \\ \vdots \\ \vec{p}_T \end{pmatrix} \tag{3.90}$$

For data sets with two or more spatial dimensions the notation needs not be changed, more-dimensional photos can be represented as row vectors by simply identifying the vector components with the locations of the photos in any arbitrary order. The *method of empirical orthogonal functions*, also called the *method of empirical eigenfunctions*, *proper orthogonal decomposition* (POD), *principal component analysis*, *singular spectrum analysis*, *Karhunen-Loéve expansion* or *bi-orthogonal decomposition* is a pattern recognition algorithm that uses linear correlations in space to find coherent structures, the *empirical eigenfunctions*, also called *empirical orthogonal functions* (EOFs), *topos* or *empirical modes* in spatio-temporal data. For the definition of the empirical modes one first has to define the *covariance matrix*, which is a measure for the linear interdependence of

[19] A vector holding the spatial data samples of a certain time.

the dynamics of the different locations. Each entry i,j of the covariance matrix (two point correlation matrix) R is the covariance of location i with location j. Subject to the condition that the time average at every location is zero, it may be written as cf. 3.2.2):

$$\mathbf{R} := \begin{pmatrix} \langle \vec{a}_1, \vec{a}_1 \rangle & \langle \vec{a}_1, \vec{a}_2 \rangle & \dots & \langle \vec{a}_1, \vec{a}_X \rangle \\ \langle \vec{a}_2, \vec{a}_1 \rangle & \langle \vec{a}_2, \vec{a}_2 \rangle & \dots & \langle \vec{a}_2, \vec{a}_X \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \vec{a}_X, \vec{a}_1 \rangle & \langle \vec{a}_X, \vec{a}_2 \rangle & \dots & \langle \vec{a}_X, \vec{a}_X \rangle \end{pmatrix} = \mathbf{A}^T\,\mathbf{A} \tag{3.91}$$

The empirical orthogonal functions of the movie $\mathbf{A}$ are defined as the eigenvectors $\vec{e}_i$ of $\mathbf{R}$, sorted with respect to the eigenvalues e_i (largest first).

The sequence of scalar products of the photos with one particular mode is called the *mode amplitude* (or *chronos*) of that mode. This is time series like data set. For the spatio-temporal data obtained from a linear system (e.g. coupled linear oscillators) this method finds the normal modes of the system as empirical eigenfunctions.

An alternative definition of the empirical eigenfunctions may be given with the following extremality condition: Let the first empirical mode be the one that has the greatest variance of its corresponding mode amplitude among all possible modes. The second mode be the one that has the greatest variance of its mode amplitude among all possible modes remaining and so on. As $\mathbf{R}$ is symmetric the eigenvalues e_i are positive and the eigenvectors $\vec{e}_i$ form a complete orthogonal system (i.e. a basis) for the X-dimensional vector space.

The problem of finding the eigensystem of $\mathbf{R}$ is (numerically) solved by computing the *singular value decomposition* (SVD) of $\mathbf{A}$, i.e. finding the unique representation of the matrix $\mathbf{A}$ of the form

$$\begin{aligned} \mathbf{A} &= \mathbf{U} \cdot \omega \cdot \mathbf{V}^T && (3.92) \\ &= \mathbf{U} \cdot diag(\omega_1, \omega_2, \dots, \omega_Y) \cdot \mathbf{V}^T && (3.93) \\ \mathbf{A}(T \times X) &= \mathbf{U}(T \times X) \cdot \mathbf{\Omega}(X \times X) \cdot \mathbf{V}(X \times X)^T && (3.94) \end{aligned}$$

where $\mathbf{U}$ and $\mathbf{V}$ are column orthogonal matrices (i.e. their columns are pairwise orthogonal vectors)

$$\mathbf{U}^T \cdot \mathbf{U} = \mathbf{V}^T \cdot \mathbf{V} = \mathbf{1} \quad (= \mathbf{V} \cdot \mathbf{V}^T \text{ since } \mathbf{V} \text{ square}) \tag{3.95}$$

(For the singular value decomposition see [3] chapter 2.6, p.59.)

The row vectors of $\mathbf{V}^T$ are the empirical modes defined above which can be seen by writing the covariance matrix as

$$\mathbf{R} = \mathbf{A}^T\,\mathbf{A} = \left(\mathbf{U}\,\omega\,\mathbf{V}^T\right)^T \left(\mathbf{U}\,\omega\,\mathbf{V}^T\right) = \left(\mathbf{V}\,\omega\,\mathbf{U}^T\right) \left(\mathbf{U}\,\omega\,\mathbf{V}^T\right) \tag{3.96}$$

yielding

$$\mathbf{R} = \mathbf{V}\,\omega^2\,\mathbf{V}^T \tag{3.97}$$

Let $\mathbf{E}$ denote the matrix of the eigenvectors of the symmetric matrix $\mathbf{R}$,

$$\mathbf{E} \;:=\; \begin{pmatrix} \vec{e}_1 \\ \vec{e}_2 \\ \vdots \\ \vec{e}_X \end{pmatrix} \tag{3.98}$$

then $\mathbf{E}$ diagonalizes $\mathbf{R}$:

$$\mathrm{diag}(e_1, e_2, ..., e_X) \;=\; \mathbf{E}^T\,\mathbf{R}\,\mathbf{E} = \mathbf{E}^T\,\mathbf{V}\,\omega^2\,\mathbf{V}^T\,\mathbf{E} \tag{3.99}$$

Multiplying equation 3.97 from the left hand side with $\mathbf{V}^T$ and from the right hand side with $\mathbf{V}$ gives

$$\mathbf{V}^T\,\mathbf{R}\,\mathbf{V} \;=\; \mathbf{V}^T\,\mathbf{V}\,\omega^2\,\mathbf{V}^T\,\mathbf{V} = \mathbf{1}\,\omega^2\,\mathbf{1} = \omega^2 \tag{3.100}$$

Comparing equations 3.99 and 3.100 shows that the eigenvectors of the covariance matrix are given by the row vectors of the matrix $\mathbf{V}^T$ in the SVD and the eigenvalues are the squares of the singular values which are the elements of the diagonal matrix ω. The SVD supplies even more information: the columns of the matrix $\mathbf{U}$ are the normalized mode amplitudes for the empirical modes and singular values ω_k, $(k = 1...X)$ are the standard deviation (giving the variance) of the mode amplitudes. The SVD is a compact form of writing the 'photos' in $\mathbf{A}$ as an expansion with the empirical modes. To see this one first writes down the SVD for one element of A_{tx} of the matrix $\mathbf{A}$

$$A_{tx} \;=\; \sum_{k=1}^{X} \omega_k\,U_{tk}\,V^T_{kx} \tag{3.101}$$

which for a complete photo becomes

$$\vec{p}_t \;=\; \sum_{k=1}^{X} \omega_k\,U_{tk}\,\vec{V}^T_k \tag{3.102}$$

If the empirical modes $\vec{\Phi}_k(x) \quad k = 1...X$ are identified with the the row vectors of $\mathbf{V}^T$ the expansion coefficient $c_k(t)$ for photo number t with empirical mode number k is then $\omega_k \cdot U_{tk}$ in the above equation and one can write the expansion in its usual form

$$\vec{p}_t \;=\; \sum_{k=1}^{X} c_k(t)\cdot\vec{\Phi}_k(x) \tag{3.103}$$

$$\mathbf{A} \;=\; \begin{pmatrix} \vec{p}_1 \\ \vec{p}_2 \\ \vdots \\ \vdots \\ \vdots \\ \vec{p}_T \end{pmatrix} = \begin{pmatrix} \vec{u}_1 \\ \vec{u}_2 \\ \vdots \\ (\text{expansion coeff for photo}_i) \\ \vdots \\ \vec{u}_T \end{pmatrix} \cdot \begin{pmatrix} \omega_1 & & & \\ & \omega_2 & & \\ & & \ddots & \\ & & & \omega_X \end{pmatrix} \cdot \begin{pmatrix} \Phi_1 \\ \Phi_2 \\ \vdots \\ \Phi_X \end{pmatrix} \tag{3.104}$$

$$= \begin{pmatrix} m & m & & m \\ o & o & & o \\ d & d & & d \\ e & e & & e \\ a & a & & a \\ m & m & & m \\ p & p & & p \\ l & l & & l \\ i & i & \cdots & i \\ t & t & & t \\ u & u & & u \\ d & d & & d \\ e & e & & e \\ \# & \# & & \# \\ 1 & 2 & & X \end{pmatrix} \cdot \begin{pmatrix} \omega_1 & & & \\ & \omega_2 & & \\ & & \ddots & \\ & & & \omega_X \end{pmatrix} \cdot \begin{pmatrix} normalmode(1) \\ normalmode(2) \\ \vdots \\ normalmode(X) \end{pmatrix} \tag{3.105}$$

Reliable computer subroutines for the calculation of the singular value decomposition may be found in many linear algebra packages or in [3]. Still another advantage of the SVD is the drastic reduction of the computational work involved in comparison with the direct approach of first computing the covariance matrix, then its eigensystem and finally the mode amplitudes using scalar products.

As the empirical modes defined as the eigenvectors of the autocovariance matrix, some properties and limitations of the POD follow directly from statements about the autocovariance matrix.

The amount of data needed for a determination of the modes exactly the amount necessary for the computation of the spatial autocovariance. For periodic cases this is just a few oscillations (one should be enough), for other types of behavior the input data should show all the typical behavior of the system.

For the sampling rate, both in time and space, there is an optimum: The lower limit can be read from a spectrum, all oscillations with amplitudes that are essential for the dynamics should be covered. But too high sampling rates typically result in many modes with great corresponding singular values, the method 'sees' many modes that differ only in unimportant details.

The POD only finds structures that do not move: even in a one-dimensional system with periodic boundary conditions a simple peak of constant shape that moves with constant speed will produce as many (useless) empirical modes (with singular values all equal) as there are spatial sampling locations. This is the reason that the POD fails for turbulent data.

The POD is often used for the reduction of data: the output of a simulation of partial differential equations with many spatial sampling points quickly leads to huge amounts of data. The empirical modes may in such cases be used to characterise the 'patterns' of motion, each mode amplitude can then be examined separately with the methods of time series analysis to characterise the type of the dynamics. (See [20] for an application of the POD for the reduction of the output of computer simulations of vocal fold oscillations).

See [21] for an alternative way for the computation of the POD for very large data sets (the so called *method of snapshots*).

(Cf. [26] for an example of using the method of EOFs to build a low-dimensional model for a hydrodynamic flow of complex geometry. An example of data reduction with the POD is given in [27] p.92. For a complex version of the POD see [37]. Cf. also [38], [41], [23] , [21] and [39]).