DEUTSCHES ELEKTRONEN-SYNCHROTRON

DESY 94-122
July 1994

Thermodynamic and Multifractal Formalism and the Bowen-Series Map

O. Rudolph
II. Institut für Theoretische Physik, Universität Hamburg

ISSN 0418-9833

NOTKESTRASSE 85 - 22603 HAMBURG

DESY 94-122
July 1994

ISSN 0418-9833

Thermodynamic and Multifractal Formalism and the Bowen-Series Map

Oliver Rudolph*

II. Institut für Theoretische Physik, Universität Hamburg
Luruper Chaussee 149
22761 Hamburg, Germany

* Work supported in part by Deutsche Forschungsgemeinschaft

Contents

0 Overview

In the theory of quantum chaos one studies the semiclassical behaviour of quantum mechanical systems whose corresponding classical counterparts exhibit chaos.
The geodesic motion of a free classical particle on closed Riemann surfaces with constant negative curvature is strongly chaotic, cf. Anosov [1]. Selberg's theory relates the classical and the quantum mechanical systems [2, 3, 4]. These systems are sometimes considered as model systems in the theory of quantum chaos since they are well understood from a mathematical point of view.
In this work we study the multifractal formalism for the geodesic flow on surfaces with constant negative curvature. The multifractal analysis of measures has been developed in order to characterize the scaling behaviour of measures on attractors of *classical* chaotic dynamical systems globally. In order to relate the multifractal formalism with quantities usually considered in the study of the geodesic flow on Riemann surfaces with constant negative curvature, it is necessary to establish the assertions of the multifractal formalism in a mathematically rigorous way. This will be achieved with the help of the thermodynamic formalism for hyperbolic dynamical systems developed by Ruelle, Bowen and others.
The treatment is intended to be both introductory and archival and I make no apologies for giving detailed and sometimes elementary proofs. However, this work is by no means a complete treatise and many important topics and results have been omitted. The interested reader should consult the References.
We now give a short overview over the contents of the present work. It should be remarked that the Sections 1, 2 and 3 are to a great extent self-contained and can be read separately.

In **Section 1** we discuss first of all fractal sets. We introduce several notions of fractal dimension of a fractal (or non-fractal) set and study their properties. Moreover, we introduce generalized dimensions which take into account a density distribution spread on the considered set. Our definitions are motivated by the physical literature; however, the concepts are modified in such a way that they also make sense from a mathematical point of view, which is usually not the case in the physical literature. Similar results have been independently obtained by Pesin [5].
In Section 1.3 we discuss the heuristic multifractal formalism developed by Halsey et al. [6].
The term 'multifractal' reflects the fact that the considered set (provided with the density distribution) is divided into uncountably many fractal subsets.
The generalized dimensions of the set (provided with the density distribution) and the spectrum of the dimensions of the uncountably many subsets satisfy relations which are analogous to well-known relations from statistical mechanics. This leads to the thermodynamic formalism for multifractals developed in Section 1.4. The Subsection 1.4 is based on Kohmoto [7].

In **Section 2** we give a brief introduction to the thermodynamic formalism for hyperbolic dynamical systems, which has been developed by Bowen, Ruelle and others. However, we restrict ourselves mainly to symbolic dynamical systems. We consider the concepts of the metric entropy, of the topological pressure, of equilibrium states and of the transfer operator. Proofs are only given when they are necessary to show that the considered quantities are well defined.
Relevant References are [8] - [21].

In **Section 3** we discuss in Section 3.1 basic concepts of hyperbolic geometry, of the geodesic flow *on Riemann surfaces of constant negative curvature and of Fuchsian groups.* Furthermore, we discuss in Section 3 the geometrical Markov coding of geodesics on compact surfaces with con-

stant negative curvature. The geodesic flow on these surfaces is mapped to a symbolic dynamical Markov system. We define and study the Bowen-Series map.
All assertions in this section are supplied with full proofs (with the exceptions of Theorem 3.10 and of the well-known facts from hyperbolic geometry in Section 3.1). The geometrical Markov coding of geodesics is due to C. Series and coauthors [22] - [28]. However, many proofs have been modified or are entirely or at least in part new.
In Section 3.6 we suggest two methods to compute the eigenvalues of the Laplace-Beltrami-operator on the considered Riemann surface with constant negative curvature using the Markov coding of the geodesics.

In **Section 4** we develop the mathematically rigorous multifractal formalism for the Bowen-Series map. *We imbed the multifractal formalism in the thermodynamic formalism studied in* Section 2. This program is for the first time carried out for the Bowen-Series map. For a special dynamical system, called *cookie-cutters*, this program has been already carried out by Rand [29]. Our treatment parallels that in Rand [29]. Related Results can also be found in Collet et al. [30].

In the **Appendices A** and **B** we state basic facts from measure and ergodic theory. Before proceeding with Section 1 the reader is invited to go rapidly through the Appendices A and B.

In **Appendix C** we discuss briefly some in the thermodynamical formalism often considered zeta functions. We consider only zeta functions for symbolic dynamical systems. We further discuss the relation between Selberg's zeta function for the Riemann surfaces with constant negative curvature (cf. Section 3.6) and the Ruelle zeta function for the symbolic dynamical Markov system constructed in Section 3.

1 Fractals and multifractals

1.1 Fractal sets

Scale invariance is an important concept in theoretical physics. A classical example for scale invariance occuring in physics is the Brownian motion. Scale invariance is also an important concept in quantum field theory: for example, scale invariance is closely related to the Bjorken scaling in deep inelastic lepton-hadron scattering.
Scale invariance is often accompanied with certain irregular fragmented sets which are called 'fractals' and which can not or only in part be described with the help of standard geometry.

Definition 1.1 *Let X be a metric space with metric d. A function $h : X \to X$ is called a* SIMILARITY TRANSFORMATION *or a* SIMILITUDE *with (fixed) quotient $s > 0$ if: $d(h(x), h(y)) = sd(x, y)$ for all $x, y \in X$. If $s < 1$, then h is said to be a* CONTRACTIVE SIMILITUDE.
Furthermore, h is called a CONTRACTION MAPPING *with* CONTRACTIVITY FACTOR *r for h if $d(h(x), h(y)) \leq rd(x, y)$ for all $x, y \in X$..*
A set S is called (strictly) SELF-SIMILAR *if S can be devided into disjoint subsets $S = \cup_i S^i$ and if for every i a contractive similitude h^i with quotient s (independent of i) can be found such that $S^i = h^i S$.*
A set S is called (strictly) SELF-SIMILAR WITH QUOTIENT $(s^1, s^2, ..., s^n)$ *if S can be devided into n disjoint subsets $S^k, 1 \leq k \leq n$, and if for each k a contractive similitude h^k with quotient s^k can be found such that $S^k = h^k S$.*

We will consider in this section exclusively the case $X = I\!R^n$ endowed with the usual Euclidean metric.
The notion *fractal* has been introduced and popularisized by Mandelbrot, but there is no generally accepted formal mathematical definition of fractals which includes all sets which are commonly considered to be fractals and excludes all those sets which are commonly considered to be non-fractal. The term 'fractal' is not defined by a legalistic statement, but is used as a broad concept with dubious mathematical status. We will not try to give a formal definition of fractals here.
According to Mandelbrot a fractal set F is an - in the terms of standard geometry - irregular and fragmented set which, however, is not totally chaotic, but in a certain way regular (e.g. scale invariant); a fractal set F has usually the following properties: F possesses a fine structure at arbitrary small scales; F is composed of subsets which are in some way (e.g. statistical, strictly or approximately) similar to F; F possesses no characteristic length; F can not (neither locally nor globally) completely described in terms of standard geometry.
Many examples for fractal sets can be found in the books of Barnsley [31] and Mandelbrot [32]. In [32] Mandelbrot states a provisional definition:

> *A fractal is a set $F \subset I\!R^n$ whose Hausdorff dimension* [1] *is greater than the topological dimension.*

This definition is unsatisfactory since on the one hand it excludes some sets with 'fractal character' (examples can be found in an appendix in [32]) and on the other hand includes totally irregular sets without any structure.
However, as a rule we can state that the several 'fractal dimensions' associated with F are typically greater than the topological dimension of F.
B.B. Mandelbrot [32] distinguishes between those fractal sets occuring in the 'real physical' world and those occuring in mathematical theories. Sets of the second type are called "abstract" fractal sets by Mandelbrot. In the following we will only be concerned with "abstract" fractal sets. It is an empirical fact that fractal sets of the first type have always an upper and a lower characteristic length. Scale invariance is realized only between these characteristic lengths. We will call this type of fractals *quasi-fractals*. For example, coastlines, the surfaces of solids, etc. are quasi-fractal at certain scales. Many other examples can be found in [32].
Abstract fractal sets are typically strictly fractal. An important example where abstract fractals occur are nonlinear dynamical systems. A dynamical system is often given by an equation of the following form

$$\frac{d\mathbf{x}(t)}{dt} = F(\mathbf{x}(t)) \text{ with } \mathbf{x} \in D \subset I\!R^m. \tag{1}$$

Here the vector $\mathbf{x}$ is a set of observables which describe the state of the system in phase space; the real valued function F is defined on D and describes the (nonlinear) time evolution of the system. (The dimension m should be choosen minimal of course; i.e. m is the minimal number of real quantities that are sufficient to characterize the state of the system completely.) We write a solution of Equation 1 with initial condition $\mathbf{x}(0)$ (if existing) as follows

$$\mathbf{x}(t) = \mathbf{f}^t(\mathbf{x}(0)). \tag{2}$$

where $\mathbf{f}$ is a real valued vector of functions and where t is choosen such that $\mathbf{f}^t$ is well defined.
The numerical analysis of nonlinear dynamical systems has shown that the points $\mathbf{f}^t x$ accumulate for large $t \to \infty$ typically on certain subsets of phase space independent of x. These subsets are

[1] The definition of the Hausdorff dimension is given below in Section 1.2.2.

called *attractors*. We will speak of *strange attractors* when the time evolution of the dynamical system depends sensitively on the initial conditions. The concept of 'strange attractor' has been first introduced by Ruelle and Takens [33, 34]. For nonlinear systems these strange attractors are typically - at least in some sections - strictly fractal. It is often useful to consider instead of the continuous-time system in Equation 1 a discrete-time system, especially when one is not interested in the detailed structure of the trajectories.

In a *stroboscopic study* one looks at the trajectory in phase space $\{\mathbf{x}(t)\}_{t\in I\!R^+}$ only at certain times $\{t_i\}_{i\in I\!N}$ with $t_{i+1} > t_i$. In this way for every trajectory $\{\mathbf{x}(t)\}_{t\in I\!R^+}$ an orbit of the form $\{\mathbf{x}(t_i)\}_{i\in I\!N} = \{\mathbf{f}^{t_i - t_{i-1}}\mathbf{x}(t_{i-1})\}_{i\in I\!N}$ is obtained.
A *Poincaré section* is a subset H of an $(n-1)$-dimensional hypersurface in phase space $I\!R^n$. Instead of the trajectory $\{\mathbf{x}(t)\}_{t\in I\!R^+}$ one considers the set $\{\mathbf{x}(i)\}_{i\in I\!N}$ of the successive intersection points of the trajectory $\mathbf{x}(t)$ with H and the POINCARÉ-MAP or the FIRST RETURN MAP is defined by $P(\mathbf{x}(i)) = \mathbf{x}(i+1)$.
In the following we will talk mainly about discrete-time dynamical systems $T: X \to X$ (here X is an arbitrary set, called *phase space*).
Some important mathematical notions and results of measure theory and ergodic theory, which we will use in this work, are collected in the Appendices A and B.

Given the attractor of a dynamical system, e.g. of the invertible measurable map $T: X \to X$, then in order to completely characterize the properties of the dynamical system not only the geometrical structure of the attractor, but also the relative frequency with which a 'typical' orbit visits different regions of the attractor has to be considered. We will consider a certain ergodic measure μ (which is called the NATURAL or the PHYSICAL MEASURE) given by the following property:

$$\mu(A) = \lim_{m\to\infty} \#\{k : 1 \le k \le m \text{ and } T^k(x) \in A\}, \tag{3}$$

where $A \subset X$ and where x is a 'typical' initial value; in order for this operational definition to make sense, $\mu(A)$ has to be given by Equation 3 for almost all in a numerical study possible (= 'typical') initial values $x \in X$ and for almost all in a numerical study possible subsets $A \subset X$.
That the resulting measure exists and is ergodic has to be checked on a case by case basis. However, in the physical literature this is usually only numerically checked. That μ is ergodic means that for every observable g (i.e. a real valued function of the phase space coordinates) the time-average $\lim_{t\to\infty} \frac{1}{t}\int_0^t g(\mathbf{x}(t'))dt'$ equals the μ-integral (= phase space average) of g for μ-almost all $\mathbf{x} = \mathbf{x}(0)$ (cf. Theorem B.1).
The formal limit in Equation 3 can, of course, not be performed in a numerical experiment. A suitable large m has to be choosen.

Of course the physicist is interested in characterizing quantitatively the fractal sets occuring in physics and the dynamics on them. Often it is unnecessary to know exactly the complicated topological structure of single trajectories; instead it is enough to study global ergodic properties of the considered system.

Often there is some distribution or density spread on the fractal (or even non-fractal) set under consideration (e.g. charge, mass, the measure in Equation 3 (in case of dynamical systems), ...) which has to be taken into account in order to yield a physically complete description. However, fractal sets have a complex topological structure and the measures given on fractal sets have often a rather complicated scaling behaviour.

Early Ansätze attempted to characterize fractal attractors of dynamical systems by only a few universal numbers (e.g. Lyapunov exponents, fractal dimensions,...). The basic idea [2] is to characterize the scaling properties of measures on fractals by fractal dimensions (especially dimensions of Hausdorff-type). These fractal dimensions are usually heuristically defined as global scaling exponents of the 'mass' moments $\langle p_i^q \rangle := \sum_i p_i^{q+1}$ in analogy to the critical indices in critical phenomena. Here the sum is taken over the elements of a covering of the fractal set and the p_i are the local probabilities corresponding to the covering (details will be given below). Historically first only fractal dimensions corresponding to the q values $q = 0, 1, -1$ were studied. Only seldom other values $q \in \mathbb{Z}$ have been studied.
However, the scaling exponents to integer q reflect only partly the complicated structure of the fractal sets.
A new phenomenological approach to characterize fractal sets has been developed in Halsey et al. [6]. In this approach the basic idea is to associate with every fractal a continuously parametrized spectrum of generalized dimensions with the help of a measure . This is accomplished by considering the scaling behaviour of $\langle p_i^q \rangle$ not only for integer q, but instead allow q to take arbitrary real values $q \in \mathbb{R}$. This approach is known as MULTIFRACTAL analysis of measures since the fractal (or non-fractal) set is devided into many fractal subsets with the help of the measure, one subset for each possible value of the scaling exponent of $\langle p_i \rangle$.
In the next section we will define the so-called box-counting dimensions and state their most important properties. We will further consider the Hausdorff dimensions, which are more *fundamental from a mathematical point of view.*

1.2 Fractal dimensions

Since the dimension of a set should intuitively be equal to the number of independent variables needed to specify single points of a set, we expect that fractal sets in general have a non-integer dimension. This idea is formulated in a quantitative manner by the concept of fractal dimension. The idea to use fractal dimensions to characterize a physical system has been put forward especially by Mandelbrot [32]. Certainly the dimension concept introduced by Hausdorff [35] is one of the most useful concepts to characterize fractal sets. Hausdorff has associated with every subset of a metric space a dimension, which characterizes the topological structure globally.
We are not only interested in the fractal set, but also in the distribution or density spread on it (given e.g. by an ergodic measure). Therefore we have to use further quantities characterizing the set besides the box resp. Hausdorff dimensions. We will see that it is possible to introduce generalized dimensions which take into account the distribution spread on the fractal set.

In this section we define the most often used dimension type characteristics for subsets of metric spaces. For sake of simplicity, we only consider subsets of $\mathbb{R}^n$ endowed with the usual Euclidean metric. *Further informations can be found in [36] - [44] and in Pesin [45].*

The concept of *topological dimension* of metric spaces will not be considered here. However, we state the following facts (proofs can be found for instance in Hurewicz & Wallmann [46]):

[2] Because of the lack of a satisfactory definition of fractals this idea has conversely been used by Mandelbrot to define the notion of fractals as remarked above.

Remark: The topological dimension is an integer by definition.

Theorem 1.1 *$\mathbb{R}^n$ has topological dimension n.*

Theorem 1.2 *A subset $N \subset \mathbb{R}^n$ has topological dimension n if and only if N contains a non-empty open subset of $\mathbb{R}^n$.*

Theorem 1.3 *Homeomorphic metric spaces have the same topological dimension.*

1.2.1 Box dimensions

The box dimensions are often also called: box-counting dimensions, Rényi dimensions, capacities, fractal dimensions or entropy dimensions. The box dimension is the most often considered dimension type in physics. The box dimensions can only be defined for *bounded* subsets of $\mathbb{R}^n$.

We begin with a simple argument to motivate the definition of the box dimensions.
Let Z be a non-empty bounded subset of $\mathbb{R}^n$ and let $N_\delta(Z)$ be the *smallest* number of n-dimensional cubes of side δ needed to cover Z. It is intuitively appealing to say that a set has dimension D if:

$$N_\delta(Z) \sim \delta^{-D}$$

in the limit $\delta \to 0$. The covering gives us an approximation of the d-dimensional volume of Z as $N_\delta(Z)\delta^d$. For d smaller than D, the d-dimensional volume diverges in the limit $\delta \to 0$ for d larger than D, the d-dimensional volume vanishes in the limit $\delta \to 0$. Solving for D gives the following definition:

Definition 1.2 *Let Z a bounded subset of $\mathbb{R}^n$. The quantities:*

$$\dim_B^L(Z) := \liminf_{\delta \to 0} \frac{\log N_\delta(Z)}{-\log \delta}$$

$$\dim_B^U(Z) := \limsup_{\delta \to 0} \frac{\log N_\delta(Z)}{-\log \delta}$$

are called the LOWER *resp.* UPPER BOX DIMENSION *of Z.*
If both limits are equal, then we call the common value the BOX DIMENSION *of Z and write:*

$$\dim_B(Z) := \lim_{\delta \to 0} \frac{\log N_\delta(Z)}{-\log \delta}. \qquad (4)$$

Here and in the following 'log' denotes the natural logarithm. However, we have the freedom to choose any other logarithm instead; the actual choice of the logarithm is irrelevant for what follows.
The so defined box dimension has been first introduced by Kolmogorov under the name of capacity. *There exist several equivalent formulations of Definition 1.2, cf. Falconer [39]. For example,* instead of considering coverings of Z consisting of n-dimensional cubes of side δ, it is also possible to consider coverings consisting of n-dimensional circles of diameter δ.
As mentioned at the end of the last section, the box dimension corresponds to the scaling exponent of $\langle p_i^q \rangle$ for $q = -1$.

Theorem 1.4 *Let Z and Z' be bounded subsets of $\mathbb{R}^m$ and let $(Z_i)_{i \in \mathbb{N}}$ be a family of bounded subsets of $\mathbb{R}^m$ such that also $\cup_{i \in \mathbb{N}} Z_i$ is bounded. Then we have* *i)* $\dim_B^U(\emptyset) = \dim_B^L(\emptyset) = 0$;

i)' $\dim_B^U(\{x\}) = \dim_B^L(\{x\}) = 0$ *for all* $x \in \mathbb{R}^m$;

ii) $\dim_B^{L,U}(Z') \leq \dim_B^{L,U}(Z)$ *for* $Z' \subseteq Z$;

iii) $\dim_H(Z) \leq \dim_B^L(Z) \leq \dim_B^U(Z)$;

iv) $\dim_B^{L,U}(\cup_{i\in\mathbb{N}} Z_i) \geq \sup_{i\in\mathbb{N}}\{\dim_B^{L,U}(Z_i)\}$;

v) $\dim_B^U(\cup_{i=1}^n Z_i) = \max_{1\leq i\leq n}\{\dim_B^U(Z_i)\}$;

vi) $\dim_B^{L,U}(Z) = \dim_B^{L,U}(\overline{Z})$.

Here $\dim_H$ denotes the Hausdorff dimension defined below and $\overline{Z}$ denotes the closure of Z. Proofs of the assertions in the theorem can be found in Falconer [39] or Pesin [45].
i) and *ii)* and the second inequality in *iii)* follow immediately by Definition 1.2. The first inequality in *iii)* follows by Theorem 1.5 and by the definition of the Hausdorff dimension below. *iv)* follows by *ii)*. A proof for *v)* can be found in [45]. A statement analogous to *v)* can not be proven for $\dim_B^L$. Property *vi)* is clear since it is possible without loss of generality to cover Z with *closed* cubes of side δ (cf. Falconer [39]).

The 'pathological' property *vi)* limits the usefulness of the box dimensions. For example, given a fractal subset F of an open set $O \subset \mathbb{R}^m$ which lies dense in O, then it is easy to check that the box dimension of F equals m.

The box dimension is a pure geometrical concept: in order to take into account a measure given on Z we have to generalize the box dimensions, cf. Grassberger [40], Hentschel & Procaccia [47]. We choose a σ-finite Borel measure μ on $\mathbb{R}^n$. The q-box dimensions defined below depend on the chosen measure μ. As above, we consider $Z \subset \mathbb{R}^n$ to be *bounded*.
We consider finite or countably infinite coverings of the bounded set $Z \subset \mathbb{R}^n$ consisting of cubes W_i of side δ. We define $p_i := \mu(W_i)$. To take into account the measure μ on Z, we weight the several cubes with p_i-dependent factors. For arbitrary $q \in \mathbb{R}, q \neq 1$, we define:

$$\tilde{I}_q(\delta) := \frac{1}{1-q}\log\sum_i p_i^q.$$

This quantity is often called the RÉNYI INFORMATION of order q [48] or ENTROPY OF THE COVERING of order q with respect to μ. For $q = 1$ the Rényi information coincides with the usual information

$$\tilde{I}_1(\delta) = \lim_{q\to 1}\frac{1}{1-q}\log\sum_i p_i^q = -\frac{\sum_i p_i\log p_i}{\sum_i p_i}.$$

For $q = 1$ the cubes are therefore weighted with the corresponding portion of the Shannon information $-p_i\log p_i$.
The dimensions to be defined should be independent of the covering and therefore we define:

$$I_q(\delta) := \frac{1}{1-q}\inf_\delta\left\{\log\sum_i p_i^q\right\} \quad \text{for } q \neq 1. \tag{5}$$

$$\begin{aligned} I_1(\delta) &:= \inf_\delta\left\{-\frac{\sum_i p_i\log p_i}{\sum_i p_i}\right\} \\ &= \inf_\delta\left\{\log\exp\left(-\frac{\sum_i p_i\log p_i}{\sum_i p_i}\right)\right\}. \end{aligned} \tag{6}$$

The infimum is taken over all possible finite or countably infinite coverings of Z consisting of cubes of side δ. The elements of the covering of Z are not necessarily disjoint and the measure μ is not assumed to be normalized, so therefore in general $\sum_i p_i \neq 1$.

Definition 1.3 *Let Z be a bounded subset of $\mathbb{R}^n$. The quantities*

$$\dim_B^{q,L}(Z) := \liminf_{\delta\to 0}\frac{I_q(\delta)}{-\log\delta} \tag{7}$$

$$\dim_B^{q,U}(Z) := \limsup_{\delta\to 0}\frac{I_q(\delta)}{-\log\delta} \tag{8}$$

are called the LOWER *resp.* UPPER *q-*RÉNYI DIMENSION *or also the* LOWER *resp.* UPPER *q-*BOX DIMENSION *with respect to μ and in case of equality of both limits we set*

$$\dim_B^q(Z) := \dim_B^{q,U}(Z) \tag{9}$$

*and call $\dim_B^q(Z)$ the q-*RÉNYI DIMENSION *or q-*BOX DIMENSION *with respect to μ.*

For $q = 0$ the q-box dimensions reduce to the ordinary box dimensions, cf. Definition 1.2. The quantity $\dim_B^1(Z)$ is also called the INFORMATION DIMENSION of Z. Notice that $\dim_B^{q,L,U}(Z)$ could be $\pm\infty$.

It is possible to define the box dimensions in a somewhat different way. We state this definition here in order to motivate the definition of the Hausdorff dimensions below. In Theorem 1.5 $\mathcal{B}(\mathbb{R}^n)$ denotes the Borel σ-algebra on $\mathbb{R}^n$ (cf. Appendix A).

Theorem 1.5 *Let I be a countable set of indices and let $\beta(\delta) := (W_i)_{i\in I}$ be a covering of Z with cubes $W_i \in \mathcal{B}(\mathbb{R}^n)$ of side $\delta_i = \delta$ for all $i \in I$. Let $q \in \mathbb{R}$ and $\kappa \in \mathbb{R}$ and $p_i := \mu(W_i)$. We set*

$$K_\kappa^q(\delta, Z) := \begin{cases} \inf_{\beta(\delta)}\left\{\left(\sum_{i\in I} p_i^q\delta_i^{(1-q)\kappa}\right)\right\} & : \quad q \neq 1 \\ \inf_{\beta(\delta)}\left\{\exp\left(-\frac{\sum_{i\in I} p_i\log\frac{p_i}{\delta_i^\kappa}}{\sum_{i\in I} p_i}\right)\right\} & : \quad q = 1 \end{cases}$$

The equations in Definition 1.3 are for $q \neq 1$ equivalent to:

$$\liminf_{\delta\to 0} K_\kappa^q(\delta, Z) = \begin{cases} \infty & : \quad (1-q)\kappa < (1-q)\dim_B^{q,L}(Z) \\ 0 & : \quad (1-q)\kappa > (1-q)\dim_B^{q,L}(Z) \end{cases}$$

and

$$\limsup_{\delta\to 0} K_\kappa^q(\delta, Z) = \begin{cases} \infty & : \quad (1-q)\kappa < (1-q)\dim_B^{q,U}(Z) \\ 0 & : \quad (1-q)\kappa > (1-q)\dim_B^{q,U}(Z) \end{cases}$$

Definition 1.3 is for $q = 1$ equivalent to:

$$\liminf_{\delta\to 0} K_\kappa^1(\delta, Z) = \begin{cases} \infty & : \quad \kappa < \dim_B^{1,L}(Z) \\ 0 & : \quad \kappa > \dim_B^{1,L}(Z) \end{cases}$$

and

$$\limsup_{\delta\to 0} K_\kappa^1(\delta, Z) = \begin{cases} \infty & : \quad \kappa < \dim_B^{1,U}(Z) \\ 0 & : \quad \kappa > \dim_B^{1,U}(Z) \end{cases}$$

Proof: Let $\delta < 1$ and $\beta(\delta)$ be a covering of Z as above. Then we have for $q < 1$ with $\kappa_1 > \kappa_2$:

$$\sum_{i\in I} p_i^q \delta_i^{(1-q)\kappa_1} = \delta^{(1-q)(\kappa_1-\kappa_2)} \sum_{i\in I} p_i^q \delta_i^{(1-q)\kappa_2}.$$

Taking the infimum yields $\mathcal{K}^q_{\kappa_1}(\delta, Z) = \delta^{(1-q)(\kappa_1-\kappa_2)}\mathcal{K}^q_{\kappa_2}(\delta, Z)$. *We take on both sides the limit* $\liminf_{\delta\to 0}$ and we see that from $\liminf_{\delta\to 0}\mathcal{K}^q_{\kappa_2}(\delta, Z) < \infty$ it follows that $\liminf_{\delta\to 0}\mathcal{K}^q_{\kappa_1}(\delta, Z) = 0$. We have therefore shown that there exists a critical value κ_c such that $\liminf_{\delta\to 0}\mathcal{K}^q_{\kappa}(\delta, Z) = \infty$ for $\kappa < \kappa_c$ and $\liminf_{\delta\to 0}\mathcal{K}^q_{\kappa}(\delta, Z) = 0$ for $\kappa > \kappa_c$.
Similarly, there exist a change-over value κ_c' such that $\limsup_{\delta\to 0}\mathcal{K}^q_{\kappa}(\delta, Z) = \infty$ for $\kappa < \kappa_c'$ and $\limsup_{\delta\to 0}\mathcal{K}^q_{\kappa}(\delta, Z) = 0$ for $\kappa > \kappa_c'$.
For $q > 1$ we can show analogously that $\mathcal{K}^q_{\kappa_1}(\delta, Z) = \delta^{(1-q)(\kappa_1-\kappa_2)}\mathcal{K}^q_{\kappa_2}(\delta, Z)$ *and it follows again* that there exists a change-over value $(1-q)\kappa_c$ (resp. $(1-q)\kappa_c'$) such that $\liminf_{\delta\to 0}\mathcal{K}^q_{\kappa}(\delta, Z)$ (resp. $\limsup_{\delta\to 0}\mathcal{K}^q_{\kappa}(\delta, Z)$) is zero when $(1-q)\kappa < (1-q)\kappa_c$ (resp. $(1-q)\kappa < (1-q)\kappa_c'$) and is infinite when $(1-q)\kappa > (1-q)\kappa_c$ (resp. $(1-q)\kappa > (1-q)\kappa_c'$).
For $q = 1$ we can show similarly that $\mathcal{K}^1_{\kappa_1}(\delta, Z) = \delta^{(\kappa_1-\kappa_2)}\mathcal{K}^1_{\kappa_2}(\delta, Z)$ and it follows again that there exists a change-over value κ_c (resp. κ_c') such that $\liminf_{\delta\to 0}\mathcal{K}^q_{\kappa}(\delta, Z)$ (resp. $\limsup_{\delta\to 0}\mathcal{K}^q_{\kappa}(\delta, Z)$) *is zero when $\kappa < \kappa_c$ (resp. $\kappa < \kappa_c'$) and is infinite when $\kappa > \kappa_c$ (resp. $\kappa > \kappa_c'$).*
By Equations 5 and 6, it follows immediately that

$$\mathcal{K}^q_{\kappa}(\delta, Z) = \begin{cases} \exp\left((1-q)I_q(\delta) + (1-q)\kappa\ln\delta\right) & : \quad q \neq 1 \\ \exp\left(I_1(\delta) + \kappa\ln\delta\right) & : \quad q = 1 \end{cases}$$

This equation and Definition 1.3 imply the assertion. □

The next theorem collects some important properties of the box dimensions.

Theorem 1.6 *Let Z and Z' be bounded subsets of $\mathbb{R}^m$ and let $(Z_i)_{i\in\mathbb{N}}$ be a family of bounded subsets of $\mathbb{R}^m$ such that also $\cup_{i\in\mathbb{N}} Z_i$ is a bounded set. Then we have*

i) $(1-q)\dim_B^{q,L,U}(Z') \leq (1-q)\dim_B^{q,L,U}(Z)$ *and* $\dim_B^{1,L,U}(Z') \leq \dim_B^{1,L,U}(Z)$ *for* $Z' \subseteq Z$;

ii) $(1-q)\dim_H^q(Z) \leq (1-q)\dim_B^{q,L}(Z) \leq (1-q)\dim_B^{q,U}(Z)$

ii)' $\dim_H^1(Z) \leq \dim_B^{1,L}(Z) \leq \dim_B^{1,U}(Z)$.

iii) $(1-q)\dim_B^{q,L,U}\left(\cup_{i\in\mathbb{N}} Z_i\right) \geq \sup_{i\in\mathbb{N}}\left\{(1-q)\dim_B^{q,L,U}(Z_i)\right\}$

iii)' $\dim_B^{1,L,U}\left(\cup_{i\in\mathbb{N}} Z_i\right) \geq \sup_{i\in\mathbb{N}}\left\{\dim_B^{1,L,U}(Z_i)\right\}$.

Here $\dim_H^q$ denotes the q-Hausdorff dimension defined below.
Proof: *i)* follows by the same argument as in Theorem 1.4 *ii)*; *ii)* follows by definition; *iii)* is trivial by *i)*; The item *i)* in Theorem 1.4 can not be generalized to the case $q \neq 0$; instead we have in general that $\dim_B^{q,L,U}(\{x\}) \neq 0$ (not even necessarily > 0) for $x \in Z$, cf. Equation 17. □

We see that actually the quantities $(1-q)\dim_B^{q,L,U}$ behave like proper dimensions (apart from possible non-positivity). Unfortunately, however, the terminology from Definition 1.3 is in common use in the physical literature and so we stick to it.

The following inequality holds:

$$\dim_B^{q_1}(Z) \geq \dim_B^{q_2}(Z) \quad \text{if } 1 < q_1 \leq q_2 \text{ or } q_1 \leq q_2 < 1. \tag{10}$$

This inequality is also true for $q_1 < 1 < q_2$ provided $0 < \mu(Z) < \infty$. If $\mu(Z) = 0$, then inequality 10 is, however, in general false for $q_1 < 1 < q_2$, see Equation 17. To prove inequality 10, one only has to differentiate $\tilde{I}_q(\delta)$ with respect to q, cf. Grassberger [40]. The inequality follows then immediately. In Hentschel & Procaccia [47] a heuristic argument for the inequality 10 has been given for *strictly self-similar fractal sets.*

If we consider a discrete-time dynamical system $T : X \to X$, where $X \subset \mathbb{R}^n$, then it is possible to give heuristic physical reasons for the introduction of the Rényi dimensions $\dim_B^q$ provided q is a positive integer. (cf. Hentschel & Procaccia [47], Gershenfeld [49]). [3] The following argument is only applicable to subsets $Z \subset X \subset \mathbb{R}^m$ which satisfy $0 < \mu(Z) < \infty$.
We consider the return map $\tilde{T}$ on Z induced by the transformation $T : X \to X$. Then we choose a 'typical' initial value x (in the sense of Eq. 3) and consider for $M \in \mathbb{N}$ the set $Z_M := \{\tilde{T}^i(x) \mid 0 \leq i \leq M\}$. We define the CORRELATION INTEGRAL OF ORDER q by

$$C_q(\delta) := \lim_{M\to\infty} \frac{1}{M^q} \#\left\{(x_1, x_2, ..., x_q) \mid x_i \in Z_M \text{ for } 0 \leq i \leq q \text{ and } |x_k - x_l| < \delta \text{ for } 0 \leq k, l \leq q\right\}.$$

We now make an important approximation; this approximation expresses the correlation sums $C_q(\delta)$ by probabilities. It is useful to define for $0 \leq j \leq M$:

$$C_q^j(\delta) := M^{1-q}\#\{(i_1, ..., i_{q-1}) \mid 0 \leq i_k \leq M \text{ and } 0 \neq |\tilde{T}^{i_k}(x) - \tilde{T}^j(x)| < \delta \text{ and } |\tilde{T}^{i_k}(x) - \tilde{T}^{i_r}(x)| < \delta \text{ for all } 0 < k, r < q\}.$$

For the sake of simplicity we consider in the following only the case $q = 2$: However, the general case can be dealed with analogously [49].
We see that $C_2(\delta)$ is the average of the $C_2^j(\delta)$:

$$C_2(\delta) = \lim_{M\to\infty} \frac{1}{M} \sum_{i=0}^{M} C_2^i(\delta).$$

For the sake of simplicity we assume that $\mu(Z) = 1$; here the assumption $0 < \mu(Z) < \infty$ is used. Let a covering of Z with cubes W_n of side δ be given. We denote with W_j the cube which contains $\tilde{T}^j(x)$ and set $p_j = \mu(W_j)$ and denote with M_j the number of points in $Z_M \cap W_j$. $C_2^i(\delta)$ is the number of points in Z_M whose distance from $\tilde{T}^j(x)$ is smaller than δ. The approximation is

$$C_2^j(\delta) \approx \frac{M_j}{M} = p_j.$$

The idea hereby is that the distance of most points in $W_j \cap Z_M$ from $\tilde{T}^j(x)$ is smaller than δ. On the one hand, we include points in $W_j \cap Z_M$ whose distance to $\tilde{T}^j(x)$ is larger than δ, but on the other hand, we exclude some points whose distance to $\tilde{T}^j(x)$ is smaller than δ since they are not contained in $W_j \cap Z_M$. The basic hypothesis in our approximation is that the error made hereby is negligible in the limit $M \to \infty$.
With this approximation it follows for $C_2(\delta)$:

$$C_2(\delta) = \lim_{M\to\infty} \frac{1}{M} \sum_{i=0}^{M} C_2^i(\delta)$$

[3] These give also in principle an algorithm for the numerical computation of $\dim_B^q$.

$$\begin{aligned} &\approx \lim_{M\to\infty} \frac{1}{M} \sum_{\{W_j\}} \sum_{k\in W_j} p_k \\ &= \lim_{M\to\infty} \frac{1}{M} \sum_{\{W_j\}} M_j p_j \\ &= \sum_{\{W_j\}} p_j^2 \end{aligned}$$

Therefore we conclude that $C_2(\delta) = \exp((-1)\tilde{I}_2(\delta))$. In the general case $q \neq 2$ it is possible to show $C_q(\delta) = \exp((1-q)\tilde{I}_q(\delta))$ using the same type of arguments. Therefore we see that the Rényi informations of order q are a measure for the correlation of order q.
If $\dim_B^q(Z)$ exists, then we write for $C_q(\delta)$ in the limit $\delta \to 0$:

$$C_q(\delta) \sim \delta^{\tau(q)},$$

where the CORRELATION EXPONENT $\tau(q)$ OF ORDER q is related to $\dim_B^q(Z)$ by

$$\tau(q) = (q-1)\dim_B^q(Z). \tag{11}$$

1.2.2 Hausdorff dimensions

We introduce in this section Hausdorff type dimensions for subsets of $\mathbb{R}^n$. We begin with the most general form. The notion of outer measure is defined in Appendix A.

The most general form of the Hausdorff-Carathéodory dimensions:
Let $(\nu_\kappa)_{\kappa\in\mathbb{R}}$ be a family of outer measures on $\mathbb{R}^n$. We suppose that for every subset $Z \subset \mathbb{R}^n$ there exists a critical value κ_0 such that $\nu_\kappa(Z) = 0$ for $\kappa > \kappa_0$ and $\nu_\kappa(Z) = \infty$ for $\kappa < \kappa_0$, whereas $\nu_{\kappa_0}(Z) \in [0,\infty]$. The number κ_0 is then called the dimension of Z with respect to the family $(\nu_\kappa)_{\kappa\in\mathbb{R}}$.

In analogy to the definition of the q-box dimensions we will define *q-Hausdorff dimensions* of fractal sets on which a measure is given. The q-Hausdorff dimensions are generalizations of the classical concept of the Hausdorff dimension [35] to fractal sets provided with a measure. In principle, we can associate with every subset of $\mathbb{R}^n$ (provided with a σ-finite measure) its q-Hausdorff dimension for every $q \in \mathbb{R}$. In the following we choose a fixed σ-finite Borel measure μ on $\mathbb{R}^n$. The q-Hausdorff dimensions depend on the chosen σ-finite Borel measure μ for $q \neq 0$.

The classical concept of Hausdorff dimension can be obtained by setting $q = 0$ in all following equations.

Let I be a countable set of indices. Let Z be any subset of $\mathbb{R}^n$ (not necessarily compact or bounded). Let further $\mathfrak{Z}(\delta) := (B_i)_{i\in I}$ be a covering of Z with $B_i \in \mathcal{B}(\mathbb{R}^n)$ and $\delta_i := \sup\{|\mathbf{x}-\mathbf{y}| : \mathbf{x},\mathbf{y} \in B_i\} \leq \delta$ for all $i \in I$. Notice that here the diameters δ_i may be explicitely i-dependent, whereas in the definition of the box dimensions only coverings with fixed cube side δ are allowed. $\mathcal{B}(\mathbb{R}^n)$ denotes the Borel σ-algebra on $\mathbb{R}^n$.
With $q \in \mathbb{R}$ and $\kappa \in \mathbb{R}$ and the abbreviation $p_i := \mu(B_i)$ we define

$$\mathcal{H}_\kappa^q(\delta, Z) := \begin{cases} \inf_{\mathfrak{Z}(\delta)} \left\{ \left(\sum_{i\in I} p_i^q \delta_i^{(1-q)\kappa} \right) \right\} & : \ q \neq 1 \\ \inf_{\mathfrak{Z}(\delta)} \left\{ \exp\left(- \frac{\sum_{i\in I} p_i \log \frac{p_i}{\delta_i^\kappa}}{\sum_{i\in I} p_i} \right) \right\} & : \ q = 1 \end{cases} \tag{12}$$

$$\mathcal{H}_\kappa^q(Z) := \lim_{\delta\to 0} \mathcal{H}_\kappa^q(\delta, Z). \tag{13}$$

When δ tends to 0, then the number of admissible coverings over which the infimum in Equation 12 is taken decreases. Therefore $\mathcal{H}_\kappa^q(\delta, Z)$ is a monoton increasing function of δ and the limit exists (pointwise!) in $[0,\infty]$.

Theorem 1.7 *$\mathcal{H}_\kappa^q$ defines for all $q \in \mathbb{R}$ and $\kappa \in \mathbb{R}$ an outer measure.*

The definition of the notion of outer measure is given in Appendix A.
Proof: [36, 50]: Let $q \neq 1$: Let Z_1 be a subset of Z, then every covering of Z is also a covering of Z_1. Therefore we have $\mathcal{H}_\kappa^q(Z_1) \leq \mathcal{H}_\kappa^q(Z)$. Let $(Z_n)_{n\in\mathbb{N}}$ be a sequence of subsets of Z. Choose for Z_n a δ-covering $(B_{n,i})_{i\in I}$ consisting of Borel measurable sets $B_{n,i}$ such that: $\sum_i p_{n,i}^q \delta_{n,i}^{(1-q)\kappa} \leq \mathcal{H}_\kappa^q(\delta, Z_n) + \frac{\epsilon}{2^n} \leq \mathcal{H}_\kappa^q(Z_n) + \frac{\epsilon}{2^n}$. All $B_{n,i}$ together build a covering of $\cup_n Z_n$ with $\sum_{n,i} p_{n,i}^q \delta_{n,i}^{(1-q)\kappa} \leq \sum_n \mathcal{H}_\kappa^q(Z_n) + \epsilon$. Therefore we have $\mathcal{H}_\kappa^q(\cup_n Z_n) \leq \sum_n \mathcal{H}_\kappa^q(Z_n)$. For $q = 1$ this can be shown using the same argument. Therefore we have that $\mathcal{H}_\kappa^q$ is an outer measure on Z for all q and κ. □

We call $\mathcal{H}_\kappa^q$ the OUTER κ-DIMENSIONAL q-HAUSDORFF MEASURE OF Z with respect to μ.

Theorem 1.8 *For $q \leq 1$ there exists to every $Z \subset \mathbb{R}^n$ a κ_0^q such that $\mathcal{H}_\kappa^q(Z) = \infty$ for $\kappa < \kappa_0^q$ and $\mathcal{H}_\kappa^q(Z) = 0$ for $\kappa > \kappa_0^q$; for $q > 1$ there exists to every $Z \subset \mathbb{R}^n$ a κ_0^q such that $\mathcal{H}_\kappa^q(Z) = \infty$ for $\kappa > \kappa_0^q$ and $\mathcal{H}_\kappa^q(Z) = 0$ for $\kappa < \kappa_0^q$*

Proof: Let $\delta < 1$ and $\beta(\delta)$ be a covering of Z as above. Then we have for $q < 1$ with $\kappa_1 > \kappa_2$:

$$\sum_{i\in I} p_i^q \delta_i^{(1-q)\kappa_1} \leq \delta^{(1-q)(\kappa_1-\kappa_2)} \sum_{i\in I} p_i^q \delta_i^{(1-q)\kappa_2}$$

and taking the infimum yields $\mathcal{H}_{\kappa_1}^q(\delta, Z) \leq \delta^{(1-q)(\kappa_1-\kappa_2)} \mathcal{H}_{\kappa_2}^q(\delta, Z)$. Taking the limit $\delta \to 0$ we see that $\mathcal{H}_{\kappa_2}^q(Z) < \infty$ implies $\mathcal{H}_{\kappa_1}^q(Z) = 0$. We have therefore seen that the graph of $\mathcal{H}_\kappa^q(Z)$ 'jumps' at a critical value of κ from ∞ to 0. For $q > 1$ it is possible to show $\mathcal{H}_{\kappa_1}^q(\delta, Z) \geq \delta^{(1-q)(\kappa_1-\kappa_2)} \mathcal{H}_{\kappa_2}^q(\delta, Z)$ using the same argument. For $q = 1$ it is possible to show $\mathcal{H}_{\kappa_1}^1(\delta, Z) \leq \delta^{(\kappa_1-\kappa_2)} \mathcal{H}_{\kappa_2}^1(\delta, Z)$ using the same argument. The assertion of the theorem follows. □

Definition 1.4 *The quantity*

$$\dim_H^q(Z) := \kappa_0^q \tag{14}$$

will be called the q-HAUSDORFF DIMENSION of Z with respect to μ.

It can be shown that the topological dimension $\dim_T(Z)$ of a separable metric space Z is given by

$$\dim_T(Z) = \inf\{\dim_H^0(\tilde{Z}) \mid \tilde{Z} \text{ is homeomorphic to } Z\}.$$

The following theorem shows that the q-Hausdorff-dimensions behave like proper dimensions.

Theorem 1.9

i) $\dim_H^0(\emptyset) = \dim_H^0(\{x\}) = 0$ *for all* $x \in \mathbb{R}^n$;

ii) $(1-q)\dim_H^q(Z_1) \leq (1-q)\dim_H^q(Z_2)$ *for all* $Z_1 \subset Z_2 \subset \mathbb{R}^n$;

ii)' $\dim_H^1(Z_1) \leq \dim_H^1(Z_2)$ *for all* $Z_1 \subset Z_2 \subset I\!R^n$;

iii) $(1-q)\dim_H^q(\cup_{i\in I\!N} Z_i) = \sup_{i\in I\!N}(1-q)\dim_H^q(Z_i)$ *for all* $Z_i \subset I\!R^n$;

iii)' $\dim_H^1(\cup_{i\in I\!N} Z_i) = \sup_{i\in I\!N} \dim_H^1(Z_i)$ *for all* $Z_i \subset I\!R^n$;

iv) If $F \subset I\!R^n$ *is open, then* $\dim_H^0(F) = n$;

Proof: *i)* is valid by definition; for $q \neq 0$ the assertion in *i)* is in general not true [this follows by Theorem 1.6 *ii)* and Equation 17]; *ii)* follows immediately from $\mathcal{H}_\kappa^q(Z_1) \leq \mathcal{H}_\kappa^q(Z_2)$; *iii)* Let $q < 1$: If $\dim_H^q(Z_i) < \kappa$ for all i, then we have $\mathcal{H}_\kappa^q(Z_i) = 0$ and since $\mathcal{H}_\kappa^q$ is subadditive, it follows that $\mathcal{H}_\kappa^q(\cup_i Z_i) = 0$ and therefore $\dim_H^q(\cup_i Z_i) \leq \kappa$. Thus $\dim_H^q(\cup_i Z_i) \leq \sup_i \dim_H^q(Z_i)$. The converse inequality follows by *ii)*, the case $q \geq 1$ can be proven using the same argument; *iv)* cf. Falconer [39]; □

We see that actually the quantities $(1-q)\dim_H^q$ behave like proper dimensions.

We see that the box dimensions can be obtained in much the same way as the Hausdorff dimensions provided one allows only those coverings $\beta(\delta)$ which satisfy $\delta_i = \delta$ for all i, i.e. provided one replaces $\mathcal{H}_\kappa^q$ by $\mathcal{K}_\kappa^q$. Since $\mathcal{K}_\kappa^q$ is in general not monoton with δ, it is in general also not true that $\mathcal{K}_\kappa^q(\delta, Z) \leq \mathcal{K}_\kappa^{q,L,U}(Z)$ and therefore in the limit $\delta \to 0$ we obtain no outer measure. This is the reason for the unwanted properties of the box dimensions (Theorem 1.9 *iii* is in general not valid for box dimensions since $\mathcal{K}_\kappa^{q,L,U}$ is not subadditive).

The importance of the box dimensions lies in the fact that they are easier to compute (as well numerically as analytically) than the Hausdorff dimensions. In order to compute the Hausdorff dimensions, it is necessary to consider all possible coverings of a set with diameters smaller than every arbitrarily small number. In numerical computations this is impossible; therefore we have to use approximate methods. The box dimensions are often assumed to be approximately equal to (resp. to be at least of the same order of magnitude as) the Hausdorff dimensions. In many cases this yields good approximations. Nonevertheless, we see that the Hausdorff dimension is the more fundamental and more natural quantity.

It is often assumed in the physical literature that $\dim_B^q(Z)$ exists. Furthermore, many authors conjecture (for example [6], [37], [41], [43], [44]) that for typical attractors arising in physics the *box dimension equals the Hausdorff dimension* $\dim_B^q(Z) = \dim_H^q(Z)$ (at least as an order of magnitude equation). This conjecture is based on numerical results.

Mathematically the existence of the box dimensions and the equality of box and Hausdorff dimensions can be proven only under certain additional simplifying assumptions or on a case by case basis. Mathematical rigorous results in this direction can be found in Pesin [45, 5] resp. Sinai [51]. *Recently some physical systems (e.g. DLA (Diffusion limited aggregration)) have been found* for which the box dimensions are not well defined. For a detailed discussion of the numerical and conceptional problems, the reader is referred to the literature, especially to Mandelbrot [52], [53] and [54].

At the end of this section a few remarks are in order. Our definition of the q-Hausdorff dimension differs from the definition usually stated in the physical literature [6, 41, 43, 44, 55, 56, 57]. For $q > 1$ the infimum over all admissible coverings in the definition of the q-Hausdorff measure (cf. Equation 12) is in the literature usually replaced by the supremum over all admissible coverings. *We do not think that this is useful because of the following reasons:*

1. We want to define the Hausdorff dimensions in analogy to the box dimensions. In view of Theorem 1.5, it is natural to define the Hausdorff dimensions by Equation 12.

Another possibility is to change the definition of the box dimensions. However, we want here to stick to the usual definition of the box dimension used in almost all physical texts [5, 29, 38, 39, 45].

2. If we choose in Equation 12 the supremum instead of the infimum for $q > 1$, then we have to impose additional conditions on the admissible coverings in order to retain $\mathcal{H}_\kappa^q(\delta, Z)$ finite. Therefore the dimension which is obtained for $q = 0$ differs in general from the classical concept of the Hausdorff dimension (where the infimum is taken over *all* finite and countably infinite coverings without these additional conditions).
   The construction given by Hausdorff [35] is much more general than Equation 12. Our q-Hausdorff dimension defined above is only a special case of the original concept invented by Hausdorff.
   For example, Halsey et al. [6] consider only *finite* coverings consisting of *disjoint* subsets of Z. Therefore their so called 'Hausdorff dimension' can in general only be defined for compact $Z \subset I\!R^n$.

Therefore we conclude that our choice for $\mathcal{H}_\kappa^q$ in Equation 12 yields the most natural generalization of the usual Hausdorff dimension.

Finally, we have to add a few remarks about the choice of the σ-finite measure μ on $I\!R^m$.
Our measure μ *is not assumed to be a probability measure on* Z. *In the literature, however, the* quantities $\dim_B^{q,L,U}(Z)$ and $\dim_H^q(Z)$ are usually defined with respect to a probability measure $\tilde{\mu}$ on Z, i.e. $\tilde{\mu}(Z) = 1$. The measures considered in the literature are often generated by a computer study according to Equation 3 and normalized to 1. Compared with our natural definition this procedure has the following drawbacks

- The condition $\tilde{\mu}(Z) = 1$ makes only sense for Borel measurable sets Z. Therefore, if we impose the condition $\tilde{\mu}(Z) = 1$, then the q-Hausdorff dimensions and q-box dimensions can only be defined for Borel measurable subsets of $I\!R^n$.

- In order to compare the q-dimensions of different sets, $\tilde{\mu}$ has to be normalized to 1 on each set separately.

- q-Hausdorff and q-box dimensions are not defined for sets of zero $\tilde{\mu}$-measure. It is impossible to normalize $\tilde{\mu}$ to 1 on sets of zero $\tilde{\mu}$-measure.

- The q-Hausdorff dimensions are not defined for unbounded sets which satisfy $\tilde{\mu}(Z) = \infty$. It is impossible to normalize $\tilde{\mu}$ to 1 on sets of infinite $\tilde{\mu}$-measure.

- It is not possible to apply the Hausdorff-Carathéodory construction from page 13 in order to introduce the notion of dimension. The quantity corresponding to $\mathcal{H}_\kappa^q$ is not defined on *sets of zero* $\tilde{\mu}$*-measure and, furthermore, not even countably subadditive and therefore not* an outer measure.

Notice that the values of the q-box and q-Hausdorff dimensions for a set $Z \subset I\!R^m$ with $0 < \mu(Z) < \infty$ do not change when μ is normalized to 1. This follows immediately from the definitions of the q-dimensions.

1.3 Multifractal analysis of measures

In this section we describe the phenomenological theory of multifractals, which was first introduced by Mandelbrot in his work about turbulent systems and later generalized by Halsey et al. [6] and by Frisch and Parisi [58]. A summary of Mandelbrot's work can be found in [53, 59]. However, since Mandelbrot's method is unnecessarily restricted to measures which are generated by a multiplicative process (so called 'self-similar' measures), we will not discuss it here.
Our approach is heuristic and we set no value upon mathematical rigour. A summary of our heuristic assumptions is given at the end of the section. Further literature for this section is [55, 56, 57, 60, 61].

We start with a (fractal or non-fractal) set $X \subset I\!R^n$ and a σ-finite measure μ on $I\!R^n$. The purpose of the multifractal formalism is to achieve a quantitative characterization of the global and the local scaling behaviour of μ on X.
First of all, we again consider coverings $\mathcal{B}(\delta)$ of X consisting of circles with diameter δ. As mentioned above on page 7, the global scaling behaviour of μ is best characterized by the the scaling exponents of the moments $\langle p_i^q \rangle = C_{q+1}(\delta) = \sum_i p_i^{q+1}$, $q \in I\!R$, where the sum is taken over all elements of the covering $\mathcal{B}(\delta)$ and where the limit $\delta \to 0$ is considered, i.e. we write

$$\langle p_i^q \rangle \sim \delta^{\tau(q+1)}. \tag{15}$$

$\tau(q)$ and the Rényi dimensions are related by:

$$\tau(q) = (q-1)\dim_B^q(X).$$

Having in this way obtained a global characterization of the scaling behaviour of μ by generalized box dimensions, the next step is of course to ask for the local scaling behaviour of μ and for its relation to the Rényi dimensions. We consider a point $x \in X$ and a circle K_i in $\mathcal{B}(\delta)$ which contains x. We define the local scaling exponent $\alpha(x)$ of μ at the point x by

$$\mu(K_i) = p_i \sim \delta^{\alpha(x)}, \text{ for } \delta \to 0. \tag{16}$$

We call $\alpha(x)$ the HÖLDER-EXPONENT of μ at x, or the POINTWISE DIMENSION of μ at x, or the STRENGHT OF THE SINGULARITY of μ at x.

Since $p_i \sim \delta^{\alpha(x)}$, the Hölder-exponent $\alpha(x)$ is a measure for the local probability p_i. A heuristic argument clarifies the term 'pointwise dimension' for $\alpha(x)$. It relates $\alpha(x)$ with the q-box dimension of the point x. We consider the q-box dimension of the set $\{x\}$, where $x \in X$ and $q \neq 1$:

$$\dim_B^q(\{x\}) = \frac{1}{q-1} \lim_{\delta \to 0} \frac{\inf_\delta \log \sum_i p_i^q}{\log \delta}.$$

To cover a single point, we need, of course, only one cube of side δ; we choose those cube i_0 for which $\log p_{i_0}^q = \inf_i \log p_i^q$ holds and write:

$$\begin{aligned} \dim_B^q(\{x\}) &= \frac{q}{q-1} \lim_{\delta \to 0} \frac{\log p_{i_0}}{\log \delta} \\ &\simeq \frac{q}{q-1} \lim_{\delta \to 0} \frac{\log \delta^{\alpha(x)}}{\log \delta} \\ &= \frac{q\alpha(x)}{q-1}. \end{aligned} \tag{17}$$

where we have used Equation 16. For $q = 1$ the same argument yields by Equation 6 and by Definition 1.3: $\dim_B^1(\{x\}) = 0$.
By Theorem 1.6 i) and by $\alpha(x) \geq 0$, it follows from this heuristic argument that $\dim_B^q$ is positive for $q < 0$ and for $q > 1$. For $q \in [0,1]$, however, $\dim_B^q$ is in general not positive definite.

Furthermore, we notice that in general $\alpha(x)$ changes rapidly with x. We are now able to devide the set X with the help of the measure μ in many fractal subsets of X, namely

$$S(\alpha) := \{x \in X \mid \alpha(x) = \alpha\}.$$

We denote with $f(\alpha)$ the box dimension of $S(\alpha)$. We call $f(\alpha)$ also the SPECTRUM OF SINGULARITIES or simply $f(\alpha)$-SPECTRUM.
We now seek a relation between the global and the local scaling behaviour of μ, i.e. we seek a relation between the Rényi dimensions $\dim_B^q(X)$ and the spectrum of singularities $f(\alpha)$: Once again we proceed heuristically: we write the number of times $\alpha(x)$ takes on a value in the interval $[\alpha', \alpha' + d\alpha']$ as

$$d\alpha' \rho(\alpha') \delta^{-f(\alpha')}, \tag{18}$$

with an appropriately chosen function ρ. The reader should notice that the number of cubes of side δ which are needed to cover $S(\alpha)$ tends as $\delta^{-f(\alpha)}$ to ∞ in the limit $\delta \to 0$. Therefore the ansatz in Equation 18 is plausible.
In the limit $\delta \to 0$ we write the moment $\langle p_i^{q-1} \rangle$ as

$$C_q(\delta) = \langle p_i^{q-1} \rangle \sim \int d\alpha' \rho(\alpha') \delta^{-f(\alpha')} \delta^{q\alpha'}. \tag{19}$$

Since δ tends to zero the integral is dominated by the value of α' which minimizes $q\alpha' - f(\alpha')$, provided $\rho(\alpha')$ is not zero. Therefore we define $\alpha(q)$ by

$$\frac{d}{d\alpha'} \left[q\alpha' - f(\alpha')\right]\Big|_{\alpha'=\alpha(q)} = 0, \tag{20}$$

$$\frac{d^2}{d\alpha'^2} \left[q\alpha' - f(\alpha')\right]\Big|_{\alpha'=\alpha(q)} > 0 \tag{21}$$

and we further *assume* that $\alpha(q)$ is uniquely determined by this two equations. By this two equations, it follows at once:

$$f'(\alpha(q)) = q, \tag{22}$$

$$f''(\alpha(q)) < 0. \tag{23}$$

In the limit $\delta \to 0$ the correlation integral $C_q(\delta)$ behaves as:

$$C_q(\delta) \sim \delta^{-f(\alpha(q)) + q\alpha(q)}.$$

We now use the relation

$$C_q(\delta) = \exp((1-q)I_q(\delta)),$$

stated in the last section and the definition of the Rényi dimension:

$$\dim_B^q(X) = \lim_{\delta \to 0} \frac{1}{q-1} \frac{\log(C_q(\delta))}{\log(\delta)}.$$

By this equations, it follows

$$\tau(q) = (q-1)\dim_B^q(X) = q\alpha(q) - f(\alpha(q)). \tag{24}$$

This equation is the desired relation between the Rényi dimensions and the $f(\alpha)$-spectrum: $f(\alpha)$ and $\tau(q)$ are related by a Legendre transformation.
If one knows the function $f(\alpha)$ (and especially the spectrum of possible α-values), then one can use Equation 24 to compute $\dim_B^q(X)$. On the other hand, if one knows the Rényi dimensions $\dim_B^q(X)$, then differentiation of Equation 24 with the respect to q yields:

$$\alpha(q) = \frac{d}{dq}[(q-1)\dim_B^q(X)]. \tag{25}$$

and therefore we can determine $f(q)$ with the help of Equation 24.

Notice that the $f(\alpha)$-spectrum is convex and that its slope equals q at every point. Furthermore, we obtain for the box dimension ($q=0$): $\dim_B^0(X) = f(\alpha(0))$. By $f'(\alpha(0)) = 0$, it follows that the box dimension equals the maximum of the $f(\alpha)$ curve. For the information dimension ($q=1$) a similar geometric construction can be given: By $1 = f'(\alpha(1))$ and $\alpha(1) = f(\alpha(1))$, it follows that the information dimension $\dim_B^1(X) = \alpha(1)$ is determined by the conditions $\alpha = f(\alpha)$ and $f'(\alpha) = 1$. In the limit $q \to \infty$ it is possible to make the following approximation:

$$C_q(\delta) \sim p_{max}^q, \text{ for } q \to \infty,$$

and by $(q-1)\dim_B^q(X) \approx \log C_q(\delta)/\log\delta$ we have in the limit $\delta \to 0$:

$$p_{max} \sim \delta^{\dim_B^\infty(X)}.$$

Since on the other hand $p_{max} \sim \delta^{\alpha_{min}}$, we find

$$\alpha_{min} = \alpha(\infty) = \dim_B^\infty(X). \tag{26}$$

(Notice that α_{min} could be zero.) Similarly we find in the limit $q \to -\infty$:

$$\alpha_{max} = \alpha(-\infty) = \dim_B^{-\infty}(X) \tag{27}$$

(Notice that α_{max} could be ∞.)
It follows that the subset with the largest probability is a fractal with box dimension $f(\alpha_{min})$ and that the subset with the smallest probability is a fractal with box dimension $f(\alpha_{max})$.

The Equations 15, 16, 20 - 23 and 24 summarize what is usually called the MULTIFRACTAL ANALYSIS of measures. A measure μ is called a MULTIFRACTAL MEASURE if the multifractal analysis can be carried out for μ and gives a non-trivial result. [An example for a trivial result is: $\dim_B^q(X)$ is independent of q, i.e. $\alpha_{min} = \alpha_{max} = \dim_B(X)$; that is, only one Hölder-exponent is allowed. The domain of $f(\alpha)$ is only one point.]
The term 'multifractal' reflects the fact that the set X is devided into uncountably many fractal sets $S(\alpha)$ with the help of the Hölder-exponent.

Side remark: B.B. Mandelbrot restricts in his work [52, 53, 54, 59] the term 'multifractal' to measures which are generated by a multiplicative process (by analogy with the fractal self-similar

set). For this class of measures he gives a rigorous justification of the multifractal formalism. However, we will not use his constricted terminology in this work.

The multifractal analysis of measures makes it possible to characterize the complexity of a system quantitatively. The multifractal spectrum contains information about the geometric and dynamical properties of the considered system. Furthermore, the multifractal formalism is useful for numerical studies and is applicable (and indeed has been applied) to many systems (for examples, cf. Paladin & Vulpiani [60], McCauley [55, 56]). In a numerical multifractal analysis usually the natural measure given by Equation 3 is studied.

The multifractal formalism has led to a widespread use of concepts from statistical physics in the theory of dynamical systems. Equation 24 corresponds to a well-known thermodynamical relation: $\tau(q)$ corresponds to the free energy and $\langle p_i^q \rangle$ to the partition function. Even the sums in the definition of the quantities $\mathcal{H}_\chi^q(\delta, Z)$ (Equation 12) can be viewed as generalized partition functions.
The consequent development of this idea has led to a 'thermodynamic formalism' for multifractals, which has to be considered as heuristic generalization for arbitrary dynamical systems of the 'thermodynamic formalism' for hyperbolic dynamical systems developed by Ruelle, Bowen and others [9, 17]. We will give an outline of the heuristic thermodynamic formalism for multifractals due to Kohmoto [7]. The thermodynamic formalism developed by Bowen, Ruelle and others will be discussed in the following section.

The multifractal formalism described in this section contains several heuristic assumptions, which need further justification in a rigorous multifractal formalism. However, this justification has to be made on a case by case basis since for arbitrary dynamical systems these assumptions are in general false.
We collect here all heuristic assumptions which have been made during the course of this section.

- We have supposed that the box dimensions $\dim_B^q(X)$ exist. In a rigorous approach to the multifractal analysis we expect that the q-box dimensions can be replaced by the more fundamental q-Hausdorff dimensions and that Equation 24 can be derived for $\dim_H^q(X)$.
- We have assumed that the Hölder-exponent - defined by Equation 16 - is a well-defined quantity. However, the Hölder-exponent may not exist at every point x.
  One can circumvent this difficulty by setting $\alpha(x) = \limsup_{\delta\to 0}\frac{\log p_x}{\log\delta}$; another possibility to circumvent this difficulty is to consider only those $x \in X$ for which the Hölder-exponent exists.
- We have defined: $f(\alpha) = \dim_B(S(\alpha))$. However, there exist measures [59, 62] for which the set $S(\alpha)$ lies dense in X for all allowed $\alpha \in [\alpha_{min}, \alpha_{max}]$; by Theorem 1.4 *vi*) it follows that $f(\alpha) = \dim_B(X)$ for all $\alpha \in [\alpha_{min}, \alpha_{max}]$. We see again that the concept of box dimension is of limited applicability. These difficulty can be bypassed by considering the Hausdorff dimension instead of the box dimension and defining: $f(\alpha) := \dim_H(S(\alpha))$.
- We have further supposed that $f(\alpha)$ is continuous and twice differentiable.
- *$\alpha(q)$ may not be uniquely determined by the Equations 20 and 21.*
- We have assumed that the mass moments $\langle p_i^q \rangle$ scale as in Equation 15 for all $q \in \mathbb{R}$. However, for an arbitrary measure this may be false.

Until now only in some special cases it has been possible to give a mathematically rigorous justification for the above assumptions, see [29, 30, 59, 62, 63, 64, 65].
In the following we will study Markov maps on the interval $I = [0, 1]$ and we will see that the multifractal formalism for these maps can be imbedded mathematically rigorous in the thermodynamic formalism of Bowen, Ruelle and others. We will prove below that $f(\alpha) = \dim_H(S(\alpha))$ and $\tau(q)$ are real analytic functions of their arguments and are Legendre transforms of each other (cf. Theorems 4.2 and 4.3). Below we will not use the above definition of the Hölder-exponent (see Equation 16), but we will alter the definition of α (cf. Definition 4.1).

1.4 Thermodynamic formalism for multifractals

As announced in the last section, we discuss in this section the thermodynamic formalism for arbitrary dynamical systems. We set no value on mathematical rigor. The description of a mathematical rigorous thermodynamic formalism will be given in the next sections.
In this section we consider a fixed ergodic dynamical system $(X, \mathcal{A}, \mu, T)$, where $X \subset I\!R^n$, cf. Appendix B. For example, μ could be the measure given by Equation 3.
We have already mentioned that the quantities $\mathcal{H}^q_\kappa(\delta, X)$ in Equation 12 can also be viewed as partition functions:

$$\mathcal{H}^q_\kappa(\delta, X) = \inf_{\mathfrak{Z}(\delta)} \sum_i \delta_i^{(1-q)\kappa} p_i^q.$$

The analogy to the usual partition functions of statistical mechanics can easiest be seen in the case $q = 0$ and using the abbreviation $\epsilon_i := -\ln \delta_i$:

$$\mathcal{H}^0_\kappa(\delta, X) = \inf_{\mathfrak{Z}(\delta)} \sum_i \exp(-\kappa \epsilon_i).$$

The parameter κ takes over the role of the inverse temperature. ($(1-q)\kappa$ takes over the role of the inverse temperature for $q \neq 0$.) The partition function of statistical mechanics is a defined as a sum over all possible microstates of the considered system. A 'microstate' of a dynamical system is given by a single point in phase space X so it is natural to consider the limit $\delta \to 0$. This limit corresponds formally to the thermodynamic limit of statistical mechanics.
We expect that the quantities $\delta_i^{(1-q)\kappa}$ and p_i^q (remember Equation 16: $p_i \sim \delta_i^{\alpha(x)}$ for $\delta \to 0$) vanish like some power of δ in the limit $\delta \to 0$ and we write in the limit $\delta \to 0$:

$$\delta_i^{(1-q)\kappa} \sim \delta^{(1-q)\epsilon_i\kappa} \qquad (28)$$

$$p_i^q \sim \delta^{q\epsilon_i\alpha(x)}. \qquad (29)$$

Here we have introduced the new variable $\epsilon_i := \frac{\ln \delta_i}{\ln \delta}$.
Furthermore, we expect that the minimal number of sets needed to cover X (with diameter $\leq \delta$) diverges like some power of δ^{-1} in the limit $\delta \to 0$.
Finally, we expect therefore that the following bivariate Gibbs potential

$$P_G(q, \tau) := \lim_{\delta \to 0} -\frac{\log \mathcal{H}^q_\kappa(\delta, X)}{\log \delta} \qquad (30)$$

exists. Here we have introduced the new independent variable $\tau := (q-1)\kappa$.

Remark: In the theory of chaotic dynamical systems often a Gibbs potential of one real variable is considered [7, 66, 67, 68, 69, 70]. This Gibbs potential can be obtained from $P_G(q, \tau)$ by

$$P(\kappa) := P_G(0, \kappa).$$

Sometimes the quantity $P(\kappa)$ is also called 'pressure' or 'free energy'. Compare with the remark on page 88. Actually, the quantity $-P_G(0, \tau)$ is analogous to the free energy of statistical mechanics. Some authors denote therefore the quantitiy which in our notation is $-P(\kappa)$ with $P(\kappa)$.

From

$$\lim_{\delta \to 0} \mathcal{H}^q_\kappa(\delta, X) = \begin{cases} \infty & : \ (1-q)\kappa > (1-q)\kappa_0^q \\ 0 & : \ (1-q)\kappa < (1-q)\kappa_0^q \end{cases}$$

it follows that

$$P_G(q, (q-1)\kappa_0^q) = 0 \qquad (31)$$

holds; i.e. the q-Hausdorff dimension $\kappa_0^q = \dim_H^q(X)$ can be obtained from the zero of $P_G(q, \cdot)$. The Equation 31 shows that the Gibbs potential $P_G(q, \tau)$ contains more information about the considered dynamical system than the spectrum of the q-Hausdorff dimensions.

As above (in Eq. 19) we write in the limit $\delta \to 0$

$$C_q(\delta) \sim \int d\alpha \rho(\alpha) \delta^{-f(\alpha)+q\alpha}.$$

In the Equations 28 and 29 we have two local scaling exponents: first, the Hölder-exponent $\alpha(x)$ and second, the ratio $\epsilon_i = \frac{\log \delta_i}{\log \delta}$.
Generalizing Equation 18 we write the number of times $\alpha(x)$ takes on a value in the interval $[\alpha', \alpha' + d\alpha']$ *and* ϵ takes on a value in the interval $[\epsilon', \epsilon' + d\epsilon']$ as

$$d\epsilon' d\alpha' \rho(\alpha', \epsilon') \delta^{-Q_G(\epsilon', \alpha')}, \qquad (32)$$

with an appropriately chosen function ρ. Now we obtain the following generalization of Equation 19

$$\mathcal{H}^q_\kappa(\delta, X) \sim \int d\epsilon d\alpha \rho(\alpha, \epsilon) \delta^{-Q_G(\epsilon, \alpha) + q\epsilon\alpha + (1-q)\epsilon\kappa}, \qquad (33)$$

in the limit $\delta \to 0$.
We duplicate now the argumentation which has led to the Equations 20 - 23. The integral in Equation 33 is dominated by the minimum of $-Q_G(\epsilon, \alpha) + q\epsilon\alpha + (1-q)\epsilon\kappa$ in the limit $\delta \to 0$. We denote with $\alpha(q, \tau)$ and $\epsilon(q, \tau)$ those values of α and ϵ for which $-Q_G(\epsilon, \alpha) + q\epsilon\alpha - \epsilon\tau$ is minimal. We implicitly *assume* that this condition determines $\alpha(q, \tau)$ and $\epsilon(q, \tau)$ uniquely. We have:

$$\left.\frac{\partial Q_G(\epsilon, \alpha)}{\partial \epsilon}\right|_{\epsilon = \epsilon(q,\tau), \alpha = \alpha(q,\tau)} = \alpha(q, \tau) q - \tau \qquad (34)$$

$$\left.\frac{\partial Q_G(\epsilon, \alpha)}{\partial \alpha}\right|_{\epsilon = \epsilon(q,\tau), \alpha = \alpha(q,\tau)} = \epsilon(q, \tau) q \qquad (35)$$

Here and in the following we assume that both P_G and Q_G are differentiable. Therefore we obtain for the behaviour of $\mathcal{H}^q_\kappa(\delta, X)$ in the limit $\delta \to 0$:

$$\mathcal{H}^q_\kappa(\delta, X) \sim \delta^{-Q_G(\epsilon(q,\tau), \alpha(q,\tau)) + q\epsilon(q,\tau)\alpha(q,\tau) + (1-q)\kappa\epsilon(q,\tau)}$$

and therefore we have by Equation 30

$$P_G(q, \tau) = Q_G(\epsilon(q, \tau), \alpha(q, \tau)) - q\epsilon(q, \tau)\alpha(q, \tau) + \tau\epsilon(q, \tau). \qquad (36)$$

Using Equations 34 and 35 we obtain by Equation 36

$$\epsilon(q,\tau) = +\frac{\partial P_G(q,\tau)}{\partial \tau} \tag{37}$$

$$\epsilon(q,\tau)\alpha(q,\tau) = -\frac{\partial P_G(q,\tau)}{\partial q}. \tag{38}$$

By these two equations, the following *Maxwell relation* follows immediately:

$$-\frac{\partial}{\partial \tau}\left(\epsilon(q,\tau)\alpha(q,\tau)\right) = \frac{\partial}{\partial q}\epsilon(q,\tau).$$

Finally, we obtain

$$Q_G\left(\epsilon(q,\tau),\alpha(q,\tau)\right) = P_G(q,\tau) - \tau\frac{\partial P_G(q,\tau)}{\partial \tau} - q\frac{\partial P_G(q,\tau)}{\partial q}.$$

I.e. $P_G(q,\tau)$ and $Q_G(\epsilon,\alpha)$ are Legendre transforms of each other. $Q_G(\epsilon,\alpha)$ plays the role of a bivariate entropy function in our thermodynamic formalism. The Equation 32 corresponds to Boltzmann's principle $S = k\ln\Omega$.

Next, we show how to reobtain the multifractal spectrum from $P_G(q,\tau)$.
We insert Equation 36 into Equation 31 and find by Equation 34

$$Q_G\left(\epsilon(q,(q-1)\kappa_0^q),\alpha(q,(q-1)\kappa_0^q)\right) = \epsilon\left(q,(q-1)\kappa_0^q\right)\frac{\partial Q_G\left(\epsilon,\alpha(q,(q-1)\kappa_0^q)\right)}{\partial \epsilon}\Bigg|_{\epsilon=\epsilon(q,(q-1)\kappa_0^q)} \tag{39}$$

We define a function $\mathbf{f}$ by

$$\mathbf{f}\left(\alpha(q,(q-1)\kappa_0^q)\right) := q\alpha(q,(q-1)\kappa_0^q) + (1-q)\kappa_0^q \tag{40}$$

Here we have made the additional assumption that the map $q \mapsto \alpha(q,(q-1)\kappa_0^q)$ is invertible. Comparision with Equations 34 and 35 yields

$$\mathbf{f}\left(\alpha(q,(q-1)\kappa_0^q)\right) = \frac{\partial Q_G(\epsilon,\alpha)}{\partial \epsilon}\Bigg|_{\epsilon=\epsilon(q,(q-1)\kappa_0^q),\alpha=\alpha(q,(q-1)\kappa_0^q)} \tag{41}$$

Differentiation of Equation 31 yields:

$$\frac{d(1-q)\kappa_0^q}{dq} = -\alpha(q,(q-1)\kappa_0^q), \tag{42}$$

where we have used Equations 37 and 38. Differentiation of Equation 40 yields immediately

$$\frac{d\mathbf{f}(\alpha)}{d\alpha}\Bigg|_{\alpha=\alpha(q,(q-1)\kappa_0^q)} = q. \tag{43}$$

By Equations 40, 42 and 43, it follows that $\mathbf{f}$ is the Legendre transform of the spectrum $(1-q)\kappa_0^q$ of the generalized dimensions. However, by Equations 22, 24 and 25 also f is the Legendre transform of the spectrum of the generalized dimensions. If we identify $\alpha(q) = \alpha(q,(q-1)\kappa_0^q)$, then we see that the Equations 40, 42 and 43 for $\mathbf{f}$ agree with the Equations 22, 24 and 25 for f. Therefore we identify $\mathbf{f} = f$. Finally, we conclude that we have reobtained (cf. Equation 41) the spectrum of singularities $f(\alpha)$ in our thermodynamic formalism.

2 Thermodynamic formalism

2.1 Symbolic dynamics

Definition 2.1 *Let Σ be a set consisting of bi-infinite sequences of symbols $\{\omega_i\}_{i\in\mathbb{Z}}$ from a finite set $\mathcal{S}$ of N different symbols, called* ALPHABET. *Furthermore, define $\sigma : \Sigma \to \Sigma, \sigma(\{\omega_i\}) := \{\omega_i'\}$, where $\omega_i' := \omega_{i+1}$. We call σ a* SHIFT TRANSFORMATION *or a* SHIFT OPERATOR *on Σ. If Σ is invariant under the action of σ, i.e. $\sigma(\Sigma) = \Sigma$, then the pair (Σ,σ) is called a* (TWO-SIDED) SYMBOLIC DYNAMICAL SYSTEM.
Let Σ^+ be a set consisting of one-sided infinite sequences of symbols $\{\omega_i\}_{i\in\mathbb{N}}$ from a finite set $\mathcal{S}$ of N different symbols, called ALPHABET. *Furthermore, define $\sigma^+ : \Sigma^+ \to \Sigma^+, \sigma^+(\{\omega_i^+\}) := \{\tilde\omega_i^+\}$, where $\tilde\omega_i^+ := \omega_{i+1}^+$. We call σ^+ a* SHIFT-TRANSFORMATION *or a* SHIFT OPERATOR *on Σ^+. If Σ^+ is invariant under the action of σ^+, i.e. $\sigma^+(\Sigma^+) = \Sigma^+$, then the pair (Σ^+,σ^+) is called a* (ONE-SIDED) SYMBOLIC DYNAMICAL SYSTEM.

Remarks:

- Sometimes it is useful to allow the alphabet $\mathcal{S}$ of a symbolic dynamical system to contain an infinite number of different symbols. An example for such a symbolic dynamical system arises in the study of symbolic dynamics for the geodesic flow on the modular surface [71], [72] and [73].
- Everywhere in this work (Σ,σ) is used to denote a symbolic dynamical system consisting of bi-infinite symbol sequences and (Σ^+,σ^+) is used to denote a symbolic dynamical system consisting of one-sided infinite symbol sequences.

Example: We consider the set

$$\Sigma_k := \prod_{-\infty}^{\infty} Z_k = \{0,1,\dots,k-1\}^{\mathbb{Z}} = \{(x_i)_{-\infty}^{\infty} \mid x_i \in Z_k\}$$

of *all* bi-infinite sequences of symbols from a finite alphabet $Z_k := \{0,1,\dots,k-1\}$. Let σ_k denote the corresponding shift operator on Σ_k. The system (Σ_k,σ_k) is a symbolic dynamical system. We discuss some special properties of the system (Σ_k,σ_k) in Appendix B. Especially it is shown there that σ_k is a homeomorphism on Σ_k if Z_k is endowed with the discrete topology and Σ_k with the product topology. (For a general symbolic dynamical system (Σ,σ), it is shown there that σ is a homeomorphism on Σ if the alphabet $\mathcal{S}$ is endowed with the discrete topology and Σ with the product topology.)
The pair (Σ_k,σ_k) is also called the FULL k-SHIFT. A symbolic dynamical system (Σ',σ') with $\Sigma' \subset \Sigma_k$ is called a SUBSHIFT. Alternatively the set Σ_k can be thought of as all bi-infinite walks on a complete directed graph with k vertices. Complete means here that every ordered pair of vertices is connected by a directed edge in the graph; especially to every vertex there is an edge which connects the vertex with itself.
Removing some edges from the graph of a full shift yields the graph of a subshift. However, not every subshift can be represented by such an incomplete subgraph.

Definition 2.2 *(a) Let (Σ,σ) be a symbolic dynamical system and denote with $\mathcal{S}$ its alphabet containing k symbols. If there exists an irreducible $k\times k$ matrix $A = (a_{m,n})$ with matrix elements $a_{m,n} \in \{0,1\}$ such that Σ is the set consisting of all those symbol sequences $\{\omega_i\}$ which satisfy: $A_{\omega_i,\omega_{i+1}} = 1$ for all i, then (Σ,σ) is called a* TOPOLOGICAL MARKOV SHIFT *or a* TOPOLOGICAL MARKOV CHAIN. *We write in this case (Σ_A,σ_A).*

(b) *A symbolic dynamical system* (Σ, σ) *given by a finite list of forbidden blocks of the form* ${}_h[a_h, a_{h+1}, ..., a_l]_l := \{(x_i)_{-\infty}^{\infty} \mid x_j = a_j \text{ for } h \leq j \leq l\}$ *will be called a* SUBSHIFT OF FINITE TYPE. [4] Σ *contains precisely those sequences which contain none of the forbidden blocks as subsequences.*

This definition makes also sense for one-sided symbolic dynamical systems.
Obviously the topological Markov shifts are precisely those subshifts of a full k-shift which can be represented by an incomplete subgraph of the full graph of a full k-shift. The condition that A is irreducible means that the incomplete subgraph is connected and is not the union of several disconnected graphs.
Given the graph of a topological Markov shift, the matrix A can easily be constructed by setting $a_{i,j} = 1$ resp. 0 if the ith vertex is (resp. is not) connected with the jth vertex by an edge.
We have the following lemma:

Lemma 2.1 *Let* (Σ_A, σ_A) *be a topological Markov shift.* σ_A *is topologically mixing if and only if there exists an* $M > 0$ *such that* $A^M > 0$. *(The condition* $A^M > 0$ *means that:* $A^M_{i,j} > 0$ *for all* i, j.*)*

A proof can be found in Bowen [9]. The notion 'topologically mixing' is defined in Appendix B. A matrix A for which there exists an $M > 0$ with $A^M > 0$ is also called APERIODICAL. The condition that A is aperiodical means according to Lemma 2.1 that every two vertices in the graph of Σ_A can be connected by an edge path of length M.

We are now interested in the problem in what cases a general dynamical system (X, T) can be coded by a symbolic dynamical system (Σ, σ). This means that there exists a bijection ϕ which maps every point $x \in X$ to its unique symbol sequence $\phi(x) = \{\omega_n(x)\} \in \Sigma$ and which furthermore satisfies: $\phi \circ T = \sigma \circ \phi$.
In practise the construction of ϕ goes as follows: first of all, we choose a partition $X = \cup_i X_i$ of X with disjoint sets X_i. For every $x \in X$, we define $\phi(x) := \{\omega_n(x)\}$ if $T^n(x) \in X_{\omega_n}$. If the so defined ϕ is one-to-one, then we have coded (X, T) by (Σ', σ'), where Σ' is the ϕ-image of X. If T is invertible, then $n \in \mathbb{Z}$ and (Σ', σ') is a two-sided symbolic dynamical system. If T is non-invertible, then $n \in \mathbb{N}$ and (Σ', σ') is a one-sided symbolic dynamical system.
The partition $\{X_i\}$ of X is called a MARKOV PARTITION OF X and the system (X, T) a MARKOV SYSTEM if (X, T) is mapped to a topological Markov shift with the help of this partition. Finally, we can construct the matrix A as follows:

$$a_{i,j} = \begin{cases} 1 & : \quad T(X_i) \cap X_j \neq \emptyset \\ 0 & : \quad otherwise \end{cases} .$$

It is easy to see that for a Markov partition of a Markov system (X, T)

$$T(X_i) \cap X_j \neq \emptyset \Longrightarrow X_j \subset T(X_i)$$

holds. Coding dynamical systems by symbolic systems is useful because the often complicated action of T on X is replaced by the simpler shift action. E.g. the existence of an everywhere dense T-orbit is equivalent to the existence of a sequence in Σ' which contains every finite allowed

[4] In the literature the notions 'subshift of finite type' and 'topological Markov shift' are often used synonymously. The reason for this is that it can be shown that every subshift of finite type can be transformed into a topological Markov shift by replacing some finite words by new symbols. Details can be found e.g. in Adler and Flatto [71]

block an infinite number of times.
Not every dynamical system can be mapped to a symbolic dynamical system (and therefore also not to a topological Markov shift). The main problem is always to find a suitable partition.
There remains the question why a deterministic system (X, T) can be mapped to an indeterministic random process, especially to a Markov chain. To understand this better, let us consider the members of the partition $X = \cup_i X_i$ as possible outcomes of measurements. To simplify matters, we will assume that the X_i are disjoint. We consider X as the set of all possible states of the dynamical system (X, T). As in Appendix B we will call X also the phase space of (X, T). We are considering an idealized experiment to be performed on the system (X, T). The system may be in the initial state $x_0 \in X$. If $x_0 \in X_i$, then the outcome of the measurement will be i. However, since in general $T^n X_i$ will intersect more than one element of the partition of X, the outcome of the measurement at 'time' 0 does not determine the outcome of a measurement at 'time' n uniquely. Only the precise knowledge of the initial point x_0 in phase space at 'time' 0 determines the state $T^n x_0$ at all future 'times' n completely. If T is invertible, the initial point x_0 determines even all 'past' states $T^{-n} x_0$. The random element in the description of general dynamical systems by symbolic dynamical systems is therefore not a property of the dynamical system itself but comes in through the assumption that it is impossible to determine the state of the system in phase space X precisely.

2.2 The metric entropy

In this section we state some basic facts of the metric entropy. The treatment is by no means exhaustive; we only discuss those properties which we need in the sequel. The treatment in this and the following two sections is close to that in Bowen [9] and Walters [21], where much additional information can be found.
Let $(X, \mathcal{A}, \mu)$ be a probability space (cf. Definition A.2) and $\xi = \{X_1, ..., X_k\}$ a partition of X. We remind the reader that a partition of X consists of disjoint and measurable sets which cover X (cf. Definition A.4). We consider in this section only finite partitions of X.
As explained at the end of the last section, it is possible to consider the elements of ξ as possible outcomes of measurements. The probability for the outcome i equals $\mu(X_i)$. We now want to introduce the notion of information gained (resp. uncertainty lost) by performing the measurement ξ on X. We write $H_\mu(\xi)$ for the information gain to be defined below. We make a priori the natural assumption that $H_\mu(\xi)$ depends on ξ only through $\mu(X_1), ..., \mu(X_k)$ and write $H_\mu(\xi) = H_\mu(\mu(X_1), \mu(X_2), ..., \mu(X_k))$. The function $H_\mu(\mu(X_1), \mu(X_2), ..., \mu(X_k))$ is already uniquely determined through some simple reasonable properties up to a multiplicative positive constant as Theorem 2.1 below shows. Before we can state our theorem, we have to introduce the following notations:

- Let $\eta := \{Y_1, ..., Y_l\}$ be a second partition of X. We define the JOIN of ξ and η by: [5]

$$\xi \vee \eta := \{X_i \cap Y_j \mid X_i \in \xi, Y_j \in \eta\}. \tag{44}$$

  ξ and η correspond to two experiments that we denote with the same symbols ξ and η respectively.

[5] The partitions $\xi_1 \vee \xi_2 \vee ... \vee \xi_n$ where $\xi_1, ..., \xi_n$ are finite partitions of X are defined inductively. We define analogously a covering $\alpha_1 \vee ... \vee \alpha_n$ for arbitrary coverings $\alpha_1, ..., \alpha_n$ of X.

- We define the abbreviation:

$$H_\mu(\xi \mid \eta) := \sum_{j=1}^{l} \mu(Y_j) H_\mu\left(\frac{\mu(X_1 \cap Y_j)}{\mu(Y_j)}, ..., \frac{\mu(X_k \cap Y_j)}{\mu(Y_j)}\right). \tag{45}$$

  $H_\mu(\xi \mid \eta)$ can be interpretated as the average information gained by performing the ξ-measurement when the outcome of the η-measurement is known.

Theorem 2.1 *Let $D_k := \{(p_1, p_2, ..., p_k) \mid p_i \in [0,1], \sum_i p_i = 1\}$ and $H : \cup_{k=1}^{\infty} D_k \to I\!R^+$ be a function satisfying the following properties:*

1. *$H(p_1, ..., p_k) = 0$ if and only if some $p_i = 1$;*
2. *$H|_{D_k}$ is continuous for all $k > 0$, symmetric and maximal at $(1/k, ..., 1/k)$;*
3. *$H(\xi \vee \eta) = H(\xi) + H(\eta \mid \xi)$;*
4. *$H(p_1, ..., p_k, 0) = H(p_1, ..., p_k)$.*

Then there exists a number $\rho > 0$ such that $H(p_1, ..., p_k) = -\rho \sum_i p_i \log p_i$.

This theorem has been taken from Walters [21]; a proof can be found in Khinchin [75]. The expression $0 \log 0$ is regarded as 0. The following definition makes sense because of Theorem 2.1:

Definition 2.3 *We define*

$$H_\mu(\xi) := -\sum_{i=1}^{k} \mu(X_i) \log \mu(X_i)$$

and call $H_\mu(\xi)$ the ENTROPY OF THE PARTITION *ξ with respect to μ.*

Lemma 2.2 *$H_\mu(\xi \vee \eta) \le H_\mu(\xi) + H_\mu(\eta)$.*

Proof: By Theorem 2.1: $H_\mu(\xi \vee \eta) - H_\mu(\xi) = H_\mu(\eta \mid \xi)$. The function $\phi : x \mapsto -x \log x$ is concave on $]0,1[$ since the second derivative of ϕ is negative. For all $\{a_i\}_{i=1}^r$ and $\{x_i\}_{i=1}^r$ with $a_i \ge 0$, $x_i \in [0,1]$ and $\sum_{i=1}^r a_i = 1$, we have $\phi(\sum_{i=1}^r a_i x_i) \ge \sum_{i=1}^r a_i \phi(x_i)$. By Equation 45, it follows: $H_\mu(\eta \mid \xi) \le H_\mu(\eta)$. □

Lemma 2.3 *Let $\{a_m\}_{m=1}^{\infty}$ a sequence of real numbers with $a_{n+m} \le a_n + a_m$ for all $n, m \in I\!N$. Then the limit $\lim_{m\to\infty} \frac{a_m}{m}$ exists and equals $\inf_m \frac{a_m}{m}$.*

Proof: Fix an arbitrary $m \in I\!N$. For every $j > 0$, we write $j = km + n$, where $0 \le n < m$. It follows:

$$\limsup_{j\to\infty} \frac{a_j}{j} \le \limsup_{j\to\infty} \left(\frac{a_{km}}{km} + \frac{a_n}{j}\right) \le \limsup_{j\to\infty} \left(\frac{k a_m}{km} + \frac{a_n}{j}\right) = \frac{a_m}{m}.$$

Thus $\limsup_{j\to\infty} \frac{a_j}{j} \le \inf_m \frac{a_m}{m} \le \liminf_{j\to\infty} \frac{a_j}{j}$ and that proves the assertion. Notice that the limit $\lim_{m\to\infty} \frac{a_m}{m}$ could be $-\infty$. □

Lemma 2.4 *Let η be a finite partition of the probability space $(X, \mathcal{A}, \mu)$ and let T be a measure preserving map on $(X, \mathcal{A}, \mu)$, then the limit*

$$h_\mu(T, \eta) := \lim_{m\to\infty} \frac{1}{m} H_\mu(\eta \vee T^{-1}\eta \vee ... \vee T^{-m+1}\eta)$$

exists.

Proof: Define $a_m := H_\mu(\eta \vee T^{-1}\eta \vee ... \vee T^{-m+1}\eta)$. The assertion of the lemma follows immediately by Lemma 2.3, Lemma 2.2 and by the T-invariance of μ and of H_μ. □

$H_\mu(\eta \vee T^{-1}\eta \vee ... \vee T^{-m+1}\eta)$ is the information gained by measuring a word of length m. $h_\mu(T, \eta)$ is the average information when the experiment η is performed an infinite number of times.

Definition 2.4 *Let $(X, \mathcal{A}, \mu)$ be a probability space and T a measure preserving map on X, then we call*

$$h_\mu(T) := \sup_\eta h_\mu(T, \eta),$$

where the supremum is taken over all finite partitions of X the METRIC ENTROPY *or* KOLMOGOROV-SINAI-ENTROPY *of T with respect to μ.*

$h_\mu(T)$ is therefore the maximum average information gain which can be obtained by measurements on the dynamical system. By Definition B.2, it follows immediately for two isomorphic dynamical systems

Theorem 2.2 *Two isomorphic dynamical systems have the same metric entropy.*

There are examples that the converse of Theorem 2.2 is not valid, cf. Walters [21].
We want to state a theorem which is useful for computing the metric entropy in case of expansive homeomorphisms.

Definition 2.5 *Let X be a compact metric space. A homeomorphism $T : X \to X$ is called* EXPANSIVE *if there exists $\epsilon > 0$ such that*

$$d(T^k x, T^k y) \le \epsilon \text{ for all } k \in Z\!\!Z \Longrightarrow x = y.$$

ϵ is called the EXPANSIVE CONSTANT *of T.*

In Walters [21] it is shown that expansiveness is independent of the metric d provided d generates the topology of X. However, the expansive constant ϵ depends on the metric d. Furthermore, expansiveness is invariant under topological conjugacy (cf. Definition B.5).

Theorem 2.3 *Let X be a compact metric space and $T : X \to X$ an expansive homeomorphism with expansive constant ϵ. Let μ be a T-invariant Borel probability measure on X and let $\xi = (X_1, ..., X_k)$ be a finite partition of X with $\operatorname{diam}(X_i) \le \epsilon$ for all i. Then $h_\mu(T) = h_\mu(T, \xi)$.*

A proof can be found e.g. in [9, 21].
We now consider a symbolic dynamical system (Σ, σ). In Appendix B we endow Σ with a topology and show that Σ endowed with this topology is a compact metric space and that σ is a homeomorphism on Σ. We show here that σ with the metric (cf. Equation 106) given in Appendix B is expansive.
Let $\{\omega_n\} \ne \{\omega'_n\}$, then:

$$d_\rho(\sigma^m\{\omega_n\}, \sigma^m\{\omega'_n\}) = \rho^{t(\sigma^m\{\omega_n\}, \sigma^m\{\omega'_n\})} \le 1$$

since by definition $t(\sigma^m\{\omega_n\},\sigma^m\{\omega_n'\}) \geq 0$ for all m.
If $d_\rho(\sigma^m\{\omega_n\},\sigma^m\{\omega_n'\}) \leq \rho$ for all m, then we have: $\omega_m = \omega_m'$ for all $m \in \mathbb{Z}$. Therefore σ is expansive with expansive constant ρ.
Let μ be a σ-invariant Borel probability measure on Σ and $\mathcal{U} := \{U_1,\dots,U_r\}$ with $U_i = \{\{\omega_n\} \in \Sigma \mid \omega_0 = i\}$, then $\operatorname{diam}(U_i) \leq \rho < 1$ and therefore we can apply Theorem 2.3 and get:

Theorem 2.4 *Let (Σ,σ) be a symbolic dynamical system and $\mathcal{U}$ the partition of Σ given above, then for every σ-invariant Borel probability measure on Σ:*

$$h_\mu(\sigma) = h_\mu(\sigma,\mathcal{U}). \tag{46}$$

2.3 The topological pressure

The concept of topological pressure has originally been introduced for expansive dynamical systems by Ruelle [77] and generalized by Walters [21]. The topological pressure is (up to sign) analogous to the free energy of thermodynamics. We will state a theorem (cf. Theorem 2.7) which clearly reveals the analogy between the free energy of thermodynamics and the topological pressure. Again our treatment is by no means exhaustive; more informations can be found e.g. in [17, 21]. We remark that especially the work of Ruelle [17, 77] contains deeper insights into the relation between thermodynamics and dynamical systems. However, we will restrict ourselves here to some introductory remarks.

Let X be a compact metric space and $T : X \to X$ a homeomorphism on X. We denote with $C(X)$ the set of all continuous and real-valued functions on X. Let α be a finite open covering of X, then we define for $\varphi \in C(X)$ a generalized partition function:

$$Z_n(\varphi,\alpha) := \inf_\beta \left\{ \sum_{B\in\beta} e^{(S_n\varphi)(B)} \,\middle|\, \beta \text{ is a subcovering of } \bigvee_{i=0}^{n-1} T^{-i}\alpha \right\}. \tag{47}$$

where we have introduced the following abbreviation:

$$(S_n\varphi)(B) := \sup_{x\in B} \left\{ \sum_{i=0}^{n-1} \varphi(T^i x) \right\}. \tag{48}$$

Lemma 2.5 *The limit*

$$P(\varphi,\alpha) := \lim_{n\to\infty} \frac{1}{n} \log Z_n(\varphi,\alpha)$$

exists and equals $\inf_n \frac{1}{n}\log Z_n(\varphi,\alpha) > -\infty$.

Proof: By definition of Z_n, it follows immediately: $Z_n(\varphi,\alpha) > \exp(-n\,\|\varphi\|)$, where $\|\varphi\| := \sup_{x\in X}|\varphi(x)|$ is the usual norm on $C(X)$. We have $\inf_n \frac{1}{n}\log Z_n(\varphi,\alpha) > -\infty$. By Lemma 2.3, we only have to show that $\log Z_{n+m}(\varphi,\alpha) \leq \log Z_n(\varphi,\alpha) + \log Z_m(\varphi,\alpha)$. Let β_1 be a subcovering of $\bigvee_{i=0}^{n-1} T^{-i}\alpha$ and β_2 a subcovering of $\bigvee_{i=0}^{m-1} T^{-i}\alpha$, then $\beta_1 \vee T^{-n}\beta_2$ is a subcovering of $\bigvee_{i=0}^{n+m-1} T^{-i}\alpha$ and we have

$$\sum_{D\in\beta_1\vee T^{-n}\beta_2} e^{(S_{n+m}\varphi)(D)} \leq \left(\sum_{B_1\in\beta_1} e^{(S_n\varphi)(B_1)}\right)\left(\sum_{B_2\in\beta_2} e^{(S_m\varphi)(B_2)}\right).$$

Thus $\log Z_{n+m}(\varphi,\alpha) \leq \log Z_n(\varphi,\alpha) + \log Z_m(\varphi,\alpha)$. □

Theorem 2.5 *The limit*

$$P(\varphi) := \lim_{\operatorname{diam}(\alpha)\to 0} P(\varphi,\alpha)$$

exists.

Proof: Let $\tilde\alpha$ be a refinement of the covering α of X. Then $\bigvee_{i=0}^{n-1} T^{-i}\tilde\alpha$ is a refinement of $\bigvee_{i=0}^{n-1} T^{-i}\alpha$, i.e. every $\tilde D \in \bigvee_{i=0}^{n-1} T^{-i}\tilde\alpha$ is completely contained in some $D \in \bigvee_{i=0}^{n-1} T^{-i}\alpha$. Furthermore, let $\tau_\alpha := \sup_{x,y\in X}\{|\varphi(x)-\varphi(y)| : d(x,y) < \operatorname{diam}(\alpha)\}$. If $\tilde D \subset D$, then we have: $(S_m\varphi)(D) \leq (S_m\varphi)(\tilde D) + m\tau_\alpha$ and therefore $Z_m(\varphi,\alpha) \leq Z_m(\varphi,\tilde\alpha)e^{m\tau_\alpha}$. It follows immediately: $P(\varphi,\alpha) \leq P(\varphi,\tilde\alpha) + \tau_\alpha$. We choose for α the covering $\tilde\alpha$ in such a way that $\operatorname{diam}(\tilde\alpha)$ is smaller than the Lebesgue number of α. Then $\tilde\alpha$ is a refinement of α and therefore $P(\varphi,\alpha) - \tau_\alpha \leq \liminf_{\operatorname{diam}(\tilde\alpha)\to 0} P(\varphi,\tilde\alpha)$. Thus $\limsup_{\operatorname{diam}(\alpha)\to 0} P(\varphi,\alpha) \leq \liminf_{\operatorname{diam}(\tilde\alpha)\to 0} P(\varphi,\tilde\alpha)$. The assertion follows.
The value of $P(\varphi)$ could be ∞. □

Definition 2.6 *We call $P(\varphi)$ the* TOPOLOGICAL PRESSURE *for T with respect to φ. Occasionally we will write $P_T(\varphi)$ or $P(T,\varphi)$ instead of $P(\varphi)$.*
Furthermore, we call $h_{top}(T) := P_T(0)$ the TOPOLOGICAL ENTROPY *of T.*

In the following theorem we collect some properties of the topological pressure.

Theorem 2.6 *Let X be a compact metric space and $T : X \to X$ a homeomorphism on X, then we have*

1. *$P_T(\cdot)$ is either constantly ∞ or finite valued;*
2. *$P_T(\cdot)$ is convex: $P_T(\lambda\varphi + (1-\lambda)\psi) \leq \lambda P_T(\varphi) + (1-\lambda)P_T(\psi)$ for all $\lambda \in [0,1]$ and all $\varphi,\psi \in C(X)$;*
3. *If $P_T(\cdot) < \infty$, then $|P_T(\varphi) - P_T(\psi)| \leq \|\varphi - \psi\|$ for all $\varphi,\psi \in C(X)$.*
4. *Let Y be a compact metric space and $S : Y \to Y$ a homeomorphism on Y such that (X,T) and (S,Y) are two topologically conjugate dynamical systems with conjugacy $\phi : Y \to X$, then we have for all $\varphi \in C(X)$: $P_T(\varphi) = P_S(\varphi\circ\phi)$.*
5. *Let $\varphi,\psi \in C(X)$, then $\varphi \leq \psi$ implies $P_T(\varphi) \leq P_T(\psi)$.*

A proof can be found in Walters [21], for expansive homeomorphisms also in Ruelle [17, 77]. The notion of topological conjugacy is given in the appendix in Definition B.5.
We now state the property of the topological pressure which unreveals the analogy to the free energy of statistical mechanics (up to the sign).
The following theorem is also known as the *variational principle*.

Theorem 2.7 *Let $T : X \to X$ be a homeomorphism on a compact metric space X and let $\varphi \in C(X)$, then*

$$P_T(\varphi) = \sup_\mu \left\{ h_\mu(T) + \int \varphi\, d\mu \right\}$$

where the supremum is over all T-invariant Borel probability measures $\mu \in M_T(X)$.

A proof can be found e.g. in [21] or for expansive homeomorphisms T in [17, 77].
Especially it follows

$$h_{top}(T) = \sup_{\mu \in M_T(X)} h_\mu(T). \quad (49)$$

According to a well-known result of statistical mechanics, the free energy $F = E - T'S$ of a system with constant temperature and constant volume does not increase and is minimal in the equilibrium state; i.e. nature minimizes (for such systems) the quantity $E - T'S$ and in the equilibrium state

$$F_{eq} = \inf\{E - T'S\}$$

holds. Of course S denotes the entropy, T' the temperature and E the mean energy of the system.
Setting the (constant) temperature T' equal to one, it follows that the variational principle of the statistical mechanics is completely analogous to Theorem 2.7 provided we interpret the T-invariant Borel probability measures on X as 'states' (cf. Definition 2.9) and the quantity $-\mu(\varphi) = -\int \varphi(x) d\mu(x)$ as energy of the dynamical system (X, T). In our thermodynamical formalism the function φ takes over the role of an interaction energy. A point $x \in X$ corresponds to a microstate in phase space and $-\varphi(x)$ corresponds to the energy of this microstate. In the following sections (cf. Definition 2.9) we will use the term 'state' exclusively for Borel probability *measures on X and the term 'microstate' exclusively for single points in phase space X.* If μ is ergodic, the following expression for the mean energy follows by Birkhoff's ergodic Theorem B.1

$$-\mu(\varphi) = \lim_{n\to\infty} -\frac{1}{n}\sum_{i=0}^{n-1} \varphi(T^i x) \quad (50)$$

for μ-almost all $x \in X$.

Definition 2.7 *Let (Σ_A, σ_A) be a topological Markov shift with matrix A. Then we define for $\varphi \in C(\Sigma_A)$:*

$$\mathrm{var}_k\varphi := \sup\{|\varphi(\omega) - \varphi(\omega')|: \omega_i = \omega_i' \text{ for all } |i| \le k\}.$$

Choose δ with $0 < \delta < 1$ and denote with $\mathcal{F}_A^\delta$ the subset of $C(\Sigma_A)$ which consists of all φ for which there exists a $b > 0$ such that: $\mathrm{var}_k\varphi \le b\delta^k$ for all $k \ge 0$. Furthermore, we define $\mathcal{F}_A := \cup_{0<\delta<1}\mathcal{F}_A^\delta$.

Definition 2.8 *Let (X, d) be a metric space. A function $\varphi : X \to I\!R$ is called* HÖLDER CONTINUOUS OF EXPONENT α *if for all $x, y \in X$*

$$|\varphi(x) - \varphi(y)| \le C d(x, y)^\alpha$$

holds with an appropriately choosen $C > 0$. The set of all continuous functions on X which are Hölder continuous of exponent α will be denoted with $C^\alpha(X, d)$ or with $C^\alpha(X)$, when it is clear which metric on X is used. We denote with $C^{r,\alpha}(X, d)$, where $r \in I\!N$, the set of all C^r-functions whose rth derivative is Hölder continuous of exponent α.

Lemma 2.6 *If $0 < \rho < 1$ and $\alpha > 0$, then: $\mathcal{F}_A^{\rho^\alpha} = C^\alpha(\Sigma_A, d_\rho)$. Especially, $\mathcal{F}_A$ equals the set of all functions in $C(\Sigma_A)$ which have a positive Hölder exponent with respect to the metric d_ρ for every $0 < \rho < 1$. (The metric d_ρ is defined by Equation 106.)*

Proof: We choose ρ, α with $\alpha > 0$ and $0 < \rho < 1$. Let $\varphi \in C^\alpha(\Sigma_A, d_\rho)$, then:

$$\mathrm{var}_k\varphi = \sup\{|\varphi(\omega) - \varphi(\omega')|: \omega_i = \omega_i' \text{ for all } |i| \le k\} \le \sup C d_\rho(\omega, \omega')^\alpha \le C\rho^\alpha\rho^{k\alpha}.$$

Therefore $\varphi \in \mathcal{F}_A^{\rho^\alpha} \subset \mathcal{F}_A$.
Conversely, let $\varphi \in \mathcal{F}_A^{\rho^\alpha}$, then there exists a $C > 0$ such that $\mathrm{var}_k\varphi \le C\rho^{k\alpha}$. Therefore we have for all $\omega, \omega' \in \Sigma_A$ which satisfy $t(\omega, \omega') = k + 1$: $|\varphi(\omega) - \varphi(\omega')| \le \mathrm{var}_k\varphi \le C\rho^{k\alpha} = \frac{C}{\rho} d_\rho(\omega, \omega')^\alpha$. Therefore we have: $\varphi \in C^\alpha(\Sigma_A, d_\rho)$. This proves the first assertion. Furthermore, we have seen that a function $\varphi \in C(\Sigma_A)$ which has for all $0 < \rho < 1$ a positive Hölder exponent with respect to d_ρ is also contained in $\mathcal{F}_A$.
It remains to show: $\varphi \in \mathcal{F}_A \Rightarrow \varphi$ has a positive Hölder exponent with respect to d_ρ for all $\rho \in]0, 1[$. Let $\varphi \in \mathcal{F}_A$, then there exists a $\Theta \in]0, 1[$ and a $C > 0$ such that $\mathrm{var}_k\varphi \le C\Theta^k$. Choose α such that $\rho^\alpha = \Theta$, then we have for all $\omega, \omega' \in \Sigma_A$ which satisfy $t(\omega, \omega') = k + 1$: $|\varphi(\omega) - \varphi(\omega')| \le \mathrm{var}_k\varphi \le C\rho^{k\alpha} = \frac{C}{\rho} d_\rho(\omega, \omega')^\alpha$. Therefore we have for all $\rho \in]0, 1[$: $\varphi \in C^\alpha(\Sigma_A, d_\rho)$. This proves the lemma. □

The function $t(\omega, \omega') = t(\{\omega_i\}, \{\omega_i'\})$ used in the proof is defined in Equation 106.
In the following sections we will need the following lemma that sometimes is called the *principle of bounded variation.*

Lemma 2.7 *Let (Σ_A, σ_A) be a topological Markov shift and let $\varphi \in \mathcal{F}_A$ be a Hölder continuous function. Furthermore choose $\omega, \omega' \in \Sigma_A$ such that there exists an $n > 0$ with $\omega_i = \omega_i'$ for all $|i| \le n$. Then there exists a $d > 0$ such that $|\mathsf{S}_n\varphi(\omega) - \mathsf{S}_n\varphi(\omega')| < d$, where we have introduced the abbreviation $\mathsf{S}_n\varphi(\omega) := \left\{\sum_{i=0}^{n-1} \varphi(\sigma_A^i\omega)\right\}$.*

Proof: For $j < n$, we have: $(\sigma_A^j\omega)_i = (\sigma_A^j\omega')_i$ for all $|i| \le n - j$. By Definition 2.7, it follows immediately that there exist a $b > 0$ and a $\delta \in [0, 1]$ such that $|\varphi(\sigma_A^j\omega) - \varphi(\sigma_A^j\omega')| \le b\delta^{n-j}$ and therefore

$$\left|\mathsf{S}_n\varphi(\omega) - \mathsf{S}_n\varphi(\omega')\right| \le \sum_{j=0}^{n-1} |\varphi(\sigma_A^j\omega) - \varphi(\sigma_A^j\omega')| \le \sum_{j=0}^{n-1} b\delta^{n-j} \le \frac{b}{1-\delta}.$$

The assertion follows by setting $d := \frac{b}{1-\delta}$. □

Theorem 2.8 *Let (Σ_A, σ_A) be a topologically mixing topological Markov shift and consider any fixed $\Theta \in]0, 1[$. Pick n functions $\varphi_1, \ldots, \varphi_n \in \mathcal{F}_A^\Theta$ and define a real-valued function by*

$$\mathsf{P} : I\!R^n \to I\!R : \mathsf{P}(s_1, s_2, \ldots, s_n) := P\left(\sum_{i=1}^n s_i\varphi_i\right).$$

The function P is real analytic and therefore infinitely partially differentiable for every choice of the functions $\varphi_1, \ldots, \varphi_n \in \mathcal{F}_A^\Theta$.

The assertion of this theorem is weaker than the statement proven by Ruelle in his book [17]. Compare with Theorem 5.26 and the Corollaries 5.27, 7.10 and 7.12 in [17]. A proof can also be found in Parry & Pollicott [15]. (The notion 'topologically mixing' is defined in Appendix B.) Notice that given a finite number of functions $\varphi_1 \in \mathcal{F}_A^{\Theta_1}, \ldots, \varphi_n \in \mathcal{F}_A^{\Theta_n}$ with $0 < \Theta_i < 1$ for all $0 < i \le n$, *it is always possible to find a Θ (e.g. $\Theta := \max_i \Theta_i$) such that $\varphi_i \in \mathcal{F}_A^\Theta$ for all $0 < i \le n$.*

2.4 Equilibrium states

Definition 2.9 *Let X be a compact metric space. A measure $\mu \in M(X)$ is also called a* STATE *on X. $M(X)$ has been defined in Appendix B as the set of all Borel probability measures on X. Let $T : X \to X$ be a homeomorphism on X, then we denote the set of all T-invariant states with $M_T(X)$.*

With the help of the variational principle it is possible to select - in analogy to statistical mechanics - in a natural way some special states.

Definition 2.10 *Let $T : X \to X$ be a homeomorphism on a compact metric space X and let $\varphi \in C(X)$. A state $\mu \in M_T(X)$ is called an* EQUILIBRIUM STATE WITH RESPECT TO φ *if*

$$P_T(\varphi) = h_\mu(T) + \int \varphi d\mu.$$

Let the set of all equilibrium states with respect to φ be denoted with $M_T(X, \varphi)$.

There exists examples with $M_T(X, \varphi) = \emptyset$ (cf. Walters [21] § 8.3.).
The concept of *Gibbs state* will not be discussed here. Details can be found in the literature, e.g. in [9, 17, 19, 78, 79, 80, 81, 82]. We only want to mention here that for topologically mixing topological Markov shifts it has been proven that $\mu \in M_{\sigma_A}(\Sigma_A)$ is an equilibrium state with respect to $\varphi \in \mathcal{F}_A$ if and only if μ is a σ_A-invariant Gibbs state with respect to φ.

Lemma 2.8 *Let (Σ_A, σ_A) be a topologically mixing topological Markov shift, then the following holds:*

- *To every $\varphi \in C(\Sigma_A)$ there exists at least one equilibrium state;*
- *To every $\varphi \in \mathcal{F}_A$ there exists a unique equilibrium state μ_φ;*
- *The equilibrium state μ_φ with respect to $\varphi \in \mathcal{F}_A$ is ergodic;*
- *Let $\mu_1, \mu_2, \ldots, \mu_n$ be ergodic measures, then there exists a $\psi \in C(\Sigma_A)$ such that $\mu_1, \mu_2, \ldots, \mu_n \in M_{\sigma_A}(\Sigma_A, \psi)$.*

A proof of the first two points can be found e.g. in Bowen [9], a proof of the third point can be found in Walters [21] and a proof for the last point can be found in Ruelle [17] (Remark 6.15.). (It is worthwhile to compare the last point in the lemma with Corollary 3.17 in Ruelle [17] or Theorem V.2.2. in Israel [83].) All assertions of Lemma 2.8 are in general not true for general dynamical systems (X, T).
We now state Bowen's important characterization of equilibrium states.

Theorem 2.9 (Bowen) *Let (Σ_A, σ_A) be a topologically mixing topological Markov shift. A state $\mu \in M_{\sigma_A}(\Sigma_A)$ is the unique equilibrium state with respect to $\varphi \in \mathcal{F}_A$ if and only if there exist constants $C_2 \geq C_1 > 0$ such that*

$$\mu\{\omega : \omega_i = \omega'_i \text{ for all } 0 \leq i < m\} \in [C_1, C_2] \exp\left(-mP(\varphi) + \sum_{k=0}^{m-1} \varphi(\sigma^k \omega')\right)$$

for all $\omega' \in \Sigma_A$ and $m \geq 0$.

A proof is given in Bowen [9].
If $\varphi \in \mathcal{F}_A$, then the unique equilibrium state $\mu = \mu_\varphi$ with respect to φ is according to Lemma 2.8 ergodic and therefore we have $\mu(\varphi) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \varphi(\sigma_A^i \omega')$ for μ-almost all $\omega' \in \Sigma_A$ by Equation 50. By Theorem 2.9, it follows that the equilibrium state μ depends on the 'mean energy' in very much the same way as the probabilities in the canonical ensemble of statistical mechanics do. The inequality in Theorem 2.9 differs from the correponding expressions in statistical mechanics in that on the left hand side not the probability for a 'microstate' in phase space $\omega' \in \Sigma_A$, but the probability for a hole neighbourhood of this 'microstate' is considered. Correspondingly, only the first $m-1$ terms in the expression for the 'energy' $\mu(\varphi)$ are considered. Theorem 2.9 states that it is possible to obtain the state μ up to a multiplicative constant (independent of m) bounded from above and from below using this approximate value for the 'energy'.

Lemma 2.9 *Let (Σ_A, σ_A) be a topologically mixing topological Markov shift and $\phi, \psi \in \mathcal{F}_A$. Then the following statements are equivalent:*

- *$\mu_\phi = \mu_\psi$.*
- *There exists a constant c and a $u \in \mathcal{F}_A$ such that $\phi(\omega) = \psi(\omega) + u(\sigma\omega) - u(\omega) + c$ for all $\omega \in \Sigma_A$.*

If one of the above conditions is satisfied, then $c = P(\phi) - P(\psi)$.

Proof: Suppose the second statement holds. Then it follows for all $n \in I\!N$

$$\left|\sum_{i=0}^{n-1} (\phi(\sigma_A^i \omega) - \psi(\sigma_A^i \omega) - c)\right| = |u(\sigma_A^n \omega) - u(\omega)| \leq 2\,\|u\|\ .$$

Therefore the inequality in Theorem 2.9 for μ_ϕ remains valid if on the right hand side ϕ is replaced by ψ and $P(\phi)$ is replaced by $P(\phi) - c$ and the constants C_2, C_1 are changed accordingly. Further, the inequality for μ_ψ remains valid if on the right hand side ψ is replaced by ϕ and $P(\psi)$ is replaced by $P(\psi) + c$ and the constants C_1, C_2 are changed accordingly. By the second point of Lemma 2.8, it follows that $\mu_\phi = \mu_\psi$ and $c = P(\phi) - P(\psi)$. A proof for the other direction can be found e.g. in Bowen [9]. □

Let $\varphi \in \mathcal{F}_A$ and μ_φ be the unique equilibrium state with respect to φ. We consider the Hölder continuous function $\psi := \varphi - P(\varphi)$ and find that $P(\psi) = 0$ and that $\mu_\psi = \mu_\varphi$ by Lemma 2.9. Therefore we see that by adding a suitable constant to φ, it is always possible to achieve that the term $-mP(\varphi)$ in Theorem 2.9 vanishes.

Lemma 2.10 *Let $\varphi \in \mathcal{F}_A$ and let μ_φ be the corresponding (unique) equilibrium state. Then there exists a $\psi \in \mathcal{F}_A$ such that $\mu_\psi = \mu_\varphi$ and $\psi(\omega) = \psi(\omega')$ for all $\omega, \omega' \in \Sigma_A$, with $\omega_i = \omega'_i$ for $i \geq 0$.*

A proof can be found in Bowen [9]. By the Lemmata 2.9 and 2.10, we see that it is enough to consider only all equilibrium states with respect to functions $\phi \in \mathcal{F}_A^+ := \mathcal{F}_A \cap C(\Sigma_A^+)$.
By Theorem 2.2 and Theorem 2.6, it follows immediately

Theorem 2.10 *Let $X_i, i \in \{1, 2\}$ be two compact metric spaces and let $T_i : X_i \to X_i, i \in \{1, 2\}$ be two homeomorphisms on these spaces such that $(X_i, T_i), i \in \{1, 2\}$ are two topologically conjugate dynamical systems with the conjugacy $\phi : X_1 \to X_2$. Then the following statement is true: A Borel probability measure μ_2 on X_2 is an equilibrium state with respect to $\varphi_2 \in C(X_2)$ if and only if $\mu_1 = \mu_2 \circ \phi$ is an equilibrium state with respect to $\varphi_1 = \varphi_2 \circ \phi$.*

I.e. *given any dynamical system* (X,T) which can be coded in the sense of Section 2.1 by a symbolic dynamical system (Σ,σ), then by the Theorems 2.2, 2.6 and 2.10, it follows that the equilibrium states for (X,T), the topological pressure and the metric entropies of the different *states of the system* (X,T) can be obtained from the corresponding quantities of the symbolic system (Σ,σ) with the help of the conjugacy $\phi : X \to \Sigma$.

The natural measure

We continue here our discussion from Section 1.1.
We have seen in this section that an attractor X (with the map $T : X \to X$) of a dynamical system can be furnished with several (typically infinitely many) ergodic measures.
In Section 1.1 we have on the other hand recognized that the signals obtained from an experiment or a computer study generate at best only one measure, namely the measure which is obtained *from the experimental signals according to Equation 3. We have called this measure already the* natural or the physical measure in Section 1.1. Now how can we distinguish the physical measure - generated by experimental data - from the many other ergodic measures on X. This problem *is in general unsolved.*
For some dynamical systems, however, it is possible to solve this problem. For example, consider axiom A diffeomorphism $f : M \to M$ on a compact manifold M [9, 10]. For axiom A diffeomorphisms, it has been shown that for all x in a set M_0 of positive Lebesgue-measure which satisfies $\mu_\phi(M_0) = 1$ the following equation holds (see [11], [19] and [76]):

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=0}^{n-1} \delta_{f^i x} = \mu_\phi,$$

where $\phi(x) = -\log\lambda(x)$ is a Hölder continuous function and $\lambda(x)$ is the Jacobi determinant of the derivative Df of f. The limit is understood with respect to the weak*-topology. μ_ϕ is also called the SINAI-RUELLE-BOWEN MEASURE or shortly SRB MEASURE. The results of an experimental measurement whose starting point lies in M_0 generate therefore the SRB measure. Even for some other systems similar statements have been proved. For a survey and for a detailled discussion of the properties of the SRB measure, see [85, 86] and references therein.
Also for the Markov maps $f : I \to I$ on the interval $I = [0,1]$ considered in this work, we will see below that the function $\phi(x) = -\log|f'(x)|$ and the equilibrium state with respect to ϕ play a special role.

2.5 The transfer operator

The transfer operator has been introduced by Ruelle [87, 88] and Araki [89] in their work on the statistical mechanics of lattice gases with long range interaction. This operator is a generalization of the well-known transfer matrix for lattice gases with finite range interaction. A detailed discussion of the transfer operator for lattice gases can be found in the book of Mayer [14]. Many further information about lattice gases can be found in the recent book of Simon [90] or in the book of Israel [83].
The application of the transfer operator method to symbolic dynamical systems (or equivalently systems which can be coded by a symbolic dynamical system) is due to Ruelle [84] and has been further developed by Pollicott [91]. We will in this section summarize the main concepts and properties of the transfer operator for symbolic dynamical systems.

To this end we consider in this section one-sided symbolic dynamical systems (Σ^+,σ^+). The shift operator σ maps the space Σ (consisting of bi-infinite symbol sequences) bijectively onto itself. On the other hand, a word $\omega^+ \in \Sigma^+$ has in general more than one (but only a finite number of) inverse images under σ^+. Therefore the following definition makes sense. [6]

Definition 2.11 *Let* (Σ^+,σ^+) *be a one-sided symbolic dynamical system and* $\varphi \in C(\Sigma^+)$. *Then define*

$$\mathcal{L}_\varphi : C(\Sigma^+) \to C(\Sigma^+), (\mathcal{L}_\varphi g)(\omega^+) := \sum_{\tilde\omega^+ \in (\sigma^+)^{-1}\omega^+} \exp(\varphi(\tilde\omega^+))g(\tilde\omega^+).$$

We will call $\mathcal{L}_\varphi$ *the (real)* TRANSFER OPERATOR *of* (Σ,σ) *with respect to* φ.
Let $\phi \in C(\Sigma^+,\mathbb{C})$ *be a complex valued function. Then we call the operator defined by*

$$\mathcal{L}_\phi : C(\Sigma^+,\mathbb{C}) \to C(\Sigma^+,\mathbb{C}), (\mathcal{L}_\phi g)(\omega^+) := \sum_{\tilde\omega^+ \in (\sigma^+)^{-1}\omega^+} \exp(\phi(\tilde\omega^+))g(\tilde\omega^+)$$

the COMPLEX TRANSFER OPERATOR.
If $\varphi \in C(\Sigma^+)$ *and* $\mathcal{L}_\varphi 1 = 1$, *then we will also say that* φ *resp.* $\mathcal{L}_\varphi$ *is* NORMALIZED.

Sometimes $\mathcal{L}_\varphi$ is also called the RUELLE-OPERATOR and if a *one dimensional* system is coded by the considered symbolic dynamical system, then $\mathcal{L}_\varphi$ is also called the PERRON-FROBENIUS OPERATOR.
It is easy to see that $\mathcal{L}_\phi$ is linear and from

$$\|\mathcal{L}_\phi g\| \le \sup_{\omega^+}\left(\sum_{\tilde\omega^+ \in (\sigma^+)^{-1}\omega^+} \exp(\phi(\tilde\omega^+))\,\|g\|\right)$$

it follows that $\mathcal{L}_\phi$ is bounded, where the usual norm $\|\varphi\| := \sup_{\omega^+\in\Sigma^+}|\varphi(\omega^+)|$ on $C(\Sigma^+)$ has been used.
The following theorem is due to Ruelle and is also called the RUELLE-PERRON-FROBENIUS THEOREM.

Theorem 2.11 *Let* (Σ_A^+,σ_A^+) *be a one-sided topological Markov shift with aperiodical, irreducible matrix* A *and let* $\varphi \in \mathcal{F}_A^+ := \mathcal{F}_A \cap C(\Sigma_A^+)$ *(especially* φ *is real valued), then the transfer operator* $\mathcal{L}_\varphi$ *has the following properties:*

1. *There exists a simple, maximal and positive eigenvalue* λ_{max} *of* $\mathcal{L}_\varphi$ *and the corresponding eigenfunction* $g_\varphi \in C(\Sigma^+)$ *satisfies:* $g_\varphi > 0$;
2. $P(\varphi) = \ln\lambda_{max}$;
3. *The remainder of the spectrum of* $\mathcal{L}_\varphi$ *is contained in a disc with radius strictly smaller than* $e^{P(\varphi)}$;
4. *There exists a Borel probability measure* ν_φ *on* Σ_A^+ *such that* $\nu_\varphi(g_\varphi) = \int g_\varphi d\nu_\varphi = 1$ *and* $\mathcal{L}_\varphi^*\nu_\varphi = \lambda_{max}\nu_\varphi$;
5. *The Borel probability measure* $\mu_\varphi := g_\varphi\nu_\varphi$ *is* σ_A^+*-invariant.*

[6] Let X be a topological space, then we denote with $C(X,\mathbb{C})$ the space of all continuous complex valued functions on X.

A proof can be found in [9, 15].
The operator $\mathcal{L}_\varphi^*$ in the theorem is on $\mathcal{C}(\Sigma_A^+)^*$ defined by $(\mathcal{L}_\varphi^*\nu)(g) := \nu(\mathcal{L}_\varphi g)$. With $\mu = g\nu$ it is meant that for all $f \in \mathcal{C}(\Sigma_A^+)$ the following equality holds: $\mu(f) = \nu(gf) = \int g(\omega^+)f(\omega^+)d\nu(\omega^+)$.
The notion 'aperiodical matrix' has been introduced above after Lemma 2.1. Pollicott has generalized the Ruelle-Perron-Frobenius theorem for complex valued Hölder continuous functions ϕ [15, 91].
It is possible to construct from μ_φ with the help of the Ruelle-Perron-Frobenius theorem the equilibrium state with respect to $\varphi \in \mathcal{F}_A^+$. Details and proofs can be found in Bowen's book [9] Chapter 1.C.

3 The Bowen-Series map

3.1 Hyperbolic geometry

3.1.1 The Poincaré upper half-plane

In this section we discuss some basic concepts and results of hyperbolic geometry. Details and the omitted proofs can be found e.g. in Balazs & Voros [92], Adler & Flatto [71, 74], Beardon [93], Bedford et. al. [22], Berndt & Steiner [94], Fischer & Lieb [95], O'Neill [96], Terras [97] and Venkov [98].

In the following we denote with $I\!H$ always the POINCARÉ UPPER HALF-PLANE or LOBACHEVSKY HALF-PLANE which is defined as the set of all complex numbers with positive imaginary part

$$I\!H := \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\},$$

endowed with the non-Euclidean HYPERBOLIC METRIC

$$ds^2 = \frac{dx^2 + dy^2}{y^2}.$$

We write $ds^2 = \sum_{i,j} g_{ij}dx^i dx^j$, where $g_{ij} = \frac{\delta_{ij}}{y^2}$. Now we can apply well-known formulas [96] to compute the *Gaussian curvature* $K(z)$ at every point $z \in I\!H$ with the result

$$K(z) = -1$$

for all $z \in I\!H$. The Poincaré upper half-plane is therefore an example for a Riemannian manifold with constant negative curvature. By multiplying ds^2 with an arbitrary constant one can achieve every negative value for the Gaussian curvature: this reflects the fact that there is no characteristic length.
We can now use the well-known formula for the Laplace-Beltrami-operator

$$\Delta = \frac{1}{\sqrt{g}} \sum_{k,l} \frac{\partial}{\partial x^k}\left(g^{kl}\sqrt{g}\frac{\partial}{\partial x^l}\right),$$

where $g = \det g_{ij}$. We obtain the following expression for the Laplace-Beltrami-operator on $I\!H$:

$$\begin{aligned}\Delta_{I\!H} &= \frac{1}{\lambda^2}\Delta_{Eucl} \\ &= y^2\Delta_{Eucl},\end{aligned}$$

where $\Delta_{Eucl} = 4\frac{\partial^2}{\partial z \partial \bar{z}}$ denotes the Euclidean Laplace-operator. A curve γ that joins z_1 to z_2 is called a GEODESIC ARC or simply a GEODESIC BETWEEN z_1 and z_2 if γ is the shortest path between z_1 and z_2 with respect to the hyperbolic metric ds. A curve γ is called a GEODESIC if γ is a geodesic arc between any pair of points on γ.

Theorem 3.1 *The geodesics on $I\!H$ are exactly the (Euclidean) half-circles orthogonal to the real axis and the straight lines orthogonal to the real axis $I\!R$. Two different geodesics intersect at most once (intersection points on $I\!R$ included).*

We write γ_∞ and $\gamma_{-\infty}$ for the positive and negative endpoints of an oriented geodesic, respectively. $I\!H$ is geodesically complete, i.e. two arbitrary point in $I\!H$ can be joined by a (unique) geodesic arc. The DISTANCE $\rho(z_1, z_2)$ of two point $z_1, z_2 \in I\!H$ is defined as the hyperbolic length of the geodesic arc connecting the points. One finds:

$$\rho(z_1, z_2) = \mathrm{arcosh}\left(1 + \frac{|z_1 - z_2|^2}{2y_1y_2}\right).$$

The pair $(I\!H, \rho)$ is a metric space. The topology induced by ρ equals the usual Euclidean topology. We consider now the automorphism group $\mathrm{Aut}(I\!H)$ of $I\!H$, that is the group (with respect to composition of maps) of biholomorphic maps $g : I\!H \to I\!H$. The term 'biholomorphic' means that g is bijective and that both g and g^{-1} are holomorphic.

Theorem 3.2 $\mathrm{Aut}(I\!H) = PSL(2, I\!R)$.

Here $SL(2, I\!R)$ denotes the special linear group of 2×2 matrices with real entries and determinant 1. We set $PSL(2, I\!R) := SL(2, I\!R)/\{\pm E\}$. ($E$ denotes the 2×2 unit matrix).

Remarks:

- The action of a 2×2 matrix $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ on $z \in I\!H$ is of course given by $gz = g(z) := \frac{az+b}{cz+d}$. The value of gz does not change if one multiplies all entries of g with -1.
- $\mathrm{Aut}(I\!H)$ acts transitively on $I\!H$ because $g_0 = \begin{pmatrix} \sqrt{y} & \frac{x}{\sqrt{y}} \\ 0 & \frac{1}{\sqrt{y}} \end{pmatrix}$ maps i to $z = x + iy \in I\!H$.
- Every transformation of the form

  $$z \mapsto \frac{az + b}{cz + d}, \text{ with } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL^+(2, I\!R),$$

  where $GL^+(2, I\!R)$ denotes the linear group of 2×2 matrices with real entries and positive determinant, can be considered as a representative of an element of $\mathrm{Aut}(I\!H)$. Of course any two matrices which are multiples of each other represent the same element in $\mathrm{Aut}(I\!H)$.
- The metric ds (and therefore also the distance ρ) is invariant under $\mathrm{Aut}(I\!H)$. The transformations $z \mapsto g(z)$, with $g \in \mathrm{Aut}(I\!H)$ are therefore orientation preserving isometries. Furthermore, the transformations $z \mapsto g(-\bar{z})$, with $g \in \mathrm{Aut}(I\!H)$, are orientation reversing isometries of $I\!H$ [22].

- $\mathrm{Aut}(\mathbb{H})$ consists exactly of the biholomorphic transformations $g : \mathbb{H} \to \mathbb{H}$. These transformations map Euclidean circles and straight lines to Euclidean circles and straight lines (however, not necessarily straight lines to straight lines and circles to circles). Especially, geodesics in $\mathbb{H}$ are mapped to geodesics by the isometries in $\mathrm{Aut}\,\mathbb{H}$ [95].
- The Laplace-Beltrami-operator $\Delta_{\mathbb{H}}$ given above is invariant under the isometries of $\mathbb{H}$. This invariance determines $\Delta_{\mathbb{H}}$ uniquely, in the sense that every invariant operator is a polynomial in $\Delta_{\mathbb{H}}$.

Definition 3.1 *Let $SL(2,\mathbb{C})$ be the special linear group of 2×2 matrices with complex entries and determinant 1. Then $SL(2,\mathbb{C})$ operates on $\hat{\mathbb{C}}$ by* LINEAR FRACTIONAL TRANSFORMATIONS *or* MÖBIUS TRANSFORMATIONS,

$$Az \mapsto \frac{az+b}{cz+d}, \text{ for } A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2,\mathbb{C}).$$

This defines a bijection between the Möbius transformations on $\hat{\mathbb{C}}$ and $\mathrm{Aut}(\hat{\mathbb{C}}) := PSL(2,\mathbb{C}) := SL(2,\mathbb{C})/\{\pm E\}$.

Definition 3.2 *Two elements $g_1, g_2 \in \mathrm{Aut}(\hat{\mathbb{C}})$ are called* CONJUGATE *to each other if there exists an isometry $h \in \mathrm{Aut}(\hat{\mathbb{C}})$ such that*

$$g_1 = hg_2h^{-1}.$$

The equivalence relation defined hereby on $\mathrm{Aut}(\hat{\mathbb{C}})$ induces a partition of $\mathrm{Aut}(\hat{\mathbb{C}})$ into CONJUGACY CLASSES.

Definition 3.3 *Let g be an isometry in $\mathrm{Aut}(\hat{\mathbb{C}})$. Then we call g*

- PARABOLIC *if g is conjugate to the translation $T : z \mapsto z+1$,*
- ELLIPTIC *if g is conjugate to a non-trivial Euclidean rotation $\Phi : z \mapsto e^{i\varphi}z$ about the origin,*
- HYPERBOLIC *if g is conjugate to a transformation $z \mapsto \lambda z$, where $\lambda \in \mathbb{R}^+$ is positive real number with $\lambda \neq 1$.*
- LOXODROMIC, *otherwise.*

Theorem 3.3 *Let $g \in \mathrm{Aut}(\hat{\mathbb{C}})$, $g \neq \pm id$, then:*

(1) *g is parabolic $\Leftrightarrow tr(g) \in \mathbb{R}$ and $|tr(g)| = 2$.*

(2) *g is elliptic $\Leftrightarrow tr(g) \in \mathbb{R}$ and $|tr(g)| < 2$.*

(3) *g is hyperbolic $\Leftrightarrow tr(g) \in \mathbb{R}$ and $|tr(g)| > 2$.*

(4) *g is loxodromic $\Leftrightarrow tr(g) \notin \mathbb{R}$.*

By Theorem 3.3, it follows that there are no loxodromic elements in $\mathrm{Aut}(\mathbb{H})$.

Definition 3.4 *Let $g \in \mathrm{Aut}(\mathbb{H})$, $g \neq id$. Then we call the solutions in $\hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ of the equation $g(z) = z$* FIXED POINTS *of g.*
If g is elliptic, hyperbolic or parabolic, then we call the corresponding fixed points ELLIPTIC, HYPERBOLIC *or* PARABOLIC *fixed points, respectively.*

Parabolic, elliptic and hyperbolic isometries can be characterized by the locations of their fixed points:

Theorem 3.4 *Let $g \in \mathrm{Aut}(\mathbb{H})$, $g \neq \pm id$, then we have:*

(1) *g is parabolic $\Leftrightarrow$ g has a unique fixed point in $\hat{\mathbb{C}}$, which lies in $\mathbb{R} \cup \{\infty\}$,*

(2) *g is elliptic $\Leftrightarrow$ g has exactly two fixed points in $\hat{\mathbb{C}}$, one (and only one) of which lies in $\mathbb{H}$ and none of which lies on $\mathbb{R} \cup \{\infty\}$,*

(3) *g is hyperbolic $\Leftrightarrow$ g has exactly two fixed points on $\mathbb{R} \cup \{\infty\}$.*

There is a second realization of the hyperbolic geometry:
It is well-known that the CAYLEY TRANSFORM C

$$C : z \mapsto w = \frac{z-i}{z+i} \tag{51}$$

maps the upper half-plane biholomorphically to the unit disc $\mathbb{D}$. In order to obtain a hyperbolic structure on $\mathbb{D}$, we have to require that the Cayley transform is an isometry between $(\mathbb{H}, \rho)$ and $(\mathbb{D}, \rho')$, with some distance function ρ' determined by this condition:

$$\begin{aligned} ds &= \lambda(z)\,|dz| \\ &= \lambda(C^{-1}w)\,|(C^{-1})'(w)||dw| \\ &= \lambda_{\mathbb{D}}(w)\,|dw| \end{aligned}$$

Thus for $\lambda_{\mathbb{D}}$:

$$\lambda_{\mathbb{D}}(w) = \frac{2}{1-|w|^2}$$

Therefore we have determined the hyperbolic metric on $\mathbb{D}$:

$$ds^2 = 4\frac{dx^2+dy^2}{(1-x^2-y^2)^2}$$

The automorphism group of $\mathbb{D}$ can easily be determined with the help of the Cayley transform:

Theorem 3.5

$$\mathrm{Aut}(\mathbb{D}) = SU(1,1)/\{\pm E\},$$

where $SU(1,1)$ denotes the special unitary group of 2×2 matrices of the form $\begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{pmatrix}$ and determinant 1.

For the Laplace-Beltrami-operator we obtain:

$$\Delta_{\mathbb{D}} = \frac{1}{\lambda_{\mathbb{D}}^2}\Delta_{Eucl} = \frac{1}{4}(1-|z|^2)^2\Delta_{Eucl}.$$

For the distance ρ' we find:

$$\rho'(z_1, z_2) = \log\frac{|1-\bar{z}_1 z_2| + |z_2 - z_1|}{|1-\bar{z}_1 z_2| - |z_2 - z_1|}, \text{ for } z_1, z_2 \in \mathbb{D}. \tag{52}$$

It is clear that all results obtained in the following sections for $\mathbb{D}$ (resp. $\mathbb{H}$) are also valid for $\mathbb{H}$ (resp. $\mathbb{D}$).

3.1.2 The geodesic flow on $\mathbb{H}$

In this section we describe the mathematical concepts needed to study the motion of a free classical particle on $\mathbb{H}$. The Lagrangean of a free classical particle moving on $\mathbb{H}$ is $\mathcal{L} = \frac{1}{2}m(\frac{ds}{dt})^2$, where m is the mass of the particle, t denotes time and ds is the hyperbolic line element. It can be shown that the orbits of a classical free particle with the above Lagrangean are exactly the geodesics of $\mathbb{H}$, see e.g. Balasz & Voros [92].
Let $z \in \mathbb{H}$. We denote with $\mathbb{H}_z$ the set of all *unit* vectors tangent to a geodesic through z. Furthermore, we denote with $T_1\mathbb{H}$ the disjoint union of all $\mathbb{H}_z$, i.e.

$$T_1\mathbb{H} := \bigcup_{z\in\mathbb{H}} \mathbb{H}_z.$$

Every element u in $T_1\mathbb{H}$ can be coordinatized $u = u(z,\theta)$, where z is the base point of u and θ the angle between u and the positive x-axis measured in counter-clockwise direction. In the following we simply write (z,θ) instead of $u(z,\theta)$. Every element $u(z,\theta) \in T_1\mathbb{H}$ determines an oriented geodesic which we denote by $\gamma_u = \gamma(z,\theta)$.
In order to completely determine the state of a particle [7] moving on $\mathbb{H}$ along a geodesic, the value of its energy (resp. momentum or velocity) has to be fixed. The restriction to *unit* vectors corresponds to fixing the value of the velocity of the particle to 1.

Definition 3.5 *The* GEODESIC FLOW ON $\mathbb{H}$ *is a group* $\{T^r\}_{r\in\mathbb{R}}$ *of homeomorphisms on* $T_1\mathbb{H}$ *such that a specific transformation* T^r *consists in moving an element* $u \in T_1\mathbb{H}$ *along the geodesic line which it determines, by the non-Euclidean distance* r.

This definition is equivalent to the following requirements
(i) $\gamma_{T^r u} = \gamma_u$; (ii) T^r maps $(z,\theta) \in T_1\mathbb{H}$ to a unit vector $T^r(z,\theta)$ tangent to $\gamma(z,\theta)$ whose base point satisfies $\rho(z, T^r z) = r$.
Since we have normalized the velocity of the particle to 1, we can also interpret r as the time needed by the particle to cover the hyperbolic distance r. There is another parametrization for elements in $T_1\mathbb{H}$. An element in $T_1\mathbb{H}$ is uniquely determined by the two endpoints $\gamma_{-\infty}$ and γ_∞ of the associated geodesic and the hyperbolic length along the geodesic measured from an arbitratry (but fixed) point p_0 on the geodesic. For $u \in T_1\mathbb{H}$ we write $u = u(\gamma_\infty, \gamma_{-\infty}, s)$ or simply $u = (\gamma_\infty, \gamma_{-\infty}, s)$.
To prove this, we choose the point p_0 for half-circles as the point for which $\theta = 0$, and for straight lines as the point on the geodesic which satisfies $\mathrm{Im}(p_0) = i$.
The action of the geodesic flow can now simply expressed as:

$$T^r(\gamma_\infty, \gamma_{-\infty}, s) = (\gamma_\infty, \gamma_{-\infty}, s + r).$$

We see at once that the measure given by $dm = dAd\theta = \frac{dxdyd\theta}{y^2} = \frac{2d\gamma_\infty d\gamma_{-\infty} ds}{(\gamma_\infty - \gamma_{-\infty})^2}$ on $T_1\mathbb{H}$ is invariant under the action of any transformation T^r. A proof for the relation $dm = \frac{2d\gamma_\infty d\gamma_{-\infty} ds}{(\gamma_\infty - \gamma_{-\infty})^2}$ can be found in Adler & Flatto [71].
In the last section we have studied the action of the group of isometries $PSL(2,\mathbb{R})$ on $\mathbb{H}$. Every transformation $g : \mathbb{H} \to \mathbb{H}$ in Aut$\mathbb{H}$ induces a transformation $\dot{g} : T_1\mathbb{H} \to T_1\mathbb{H}$.

$$\dot{g}(z,\theta) = (gz, \theta + \arg g'(z)). \tag{53}$$

g transforms the geodesic γ_u which is uniquely determined by the tangent $u(z,\theta)$ into the geodesic $g\gamma_u$ which is determined by $u(gz, \theta + \arg g'(z))$. To prove this assertion, we choose some parametrization of the geodesic $t \mapsto \gamma_u(t)$, e.g. by the arclength. Then the tangent vector to γ_u at the parametervalue $t = 0$ is given by $s \mapsto \gamma_u(0) + s\gamma_u'(0)$. The tangent vector to the geodesic $g\gamma_u$ at the point $g\gamma_u(0)$ is given by $s \mapsto g\gamma_u(0) + s(g\gamma_u)'(0) = g\gamma_u(0) + sg'(\gamma_u(0))\gamma_u'(0)$. The assertion Equation 53 follows.
We define implicitely a function $s_0(\gamma_\infty, \gamma_{-\infty}, g)$ by

$$g(\gamma_\infty, \gamma_{-\infty}, 0) = (g\gamma_\infty, g\gamma_{-\infty}, s_0(\gamma_\infty, \gamma_{-\infty}, g)).$$

Since g is an isometry, it follows that

$$\dot{g}(\gamma_\infty, \gamma_{-\infty}, s) = (g\gamma_\infty, g\gamma_{-\infty}, s + s_0(\gamma_\infty, \gamma_{-\infty}, g)).$$

The geodesic flow defined above commutes with the action of g

$$gT^r = T^r g$$

for every $g \in \mathrm{Aut}(\mathbb{H})$. We have already seen that $PSL(2,\mathbb{R})$ acts transitively on $\mathbb{H}$ and therefore also $PSL(2,\mathbb{R})$ acts transitively on $T_1\mathbb{H}$. [8]
The classical motion along the geodesics on $\mathbb{H}$ is *completely integrable*. A system is called (completely) integrable if there exist n integrals of motion in involution, where n is the number of degrees of freedom of the system. (For the Poincaré half-space we have, of course, $n = 2$.) The phase space is foliated by n-dimensional invariant tori and every trajectory lies completely on some torus [99]. In the article of Balazs and Voros [92] it has been shown that the geodesic motion on $\mathbb{H}$ is indeed completely integrable by constructing explicitely the two constants of motion in involution.
However, it is possible to construct non-integrable systems from the Poincaré half plane by introducing periodic boundary conditions. This is achieved with the help of a discrete group of transformations on $\mathbb{H}$ and is the topic of the next section.

[7] In the following we choose our units such that $m = 1$.

[8] We only have to show that there exist isometries in $PSL(2,\mathbb{R})$ which map a unit vector (at an arbitrary base point) to an arbitrarily rotated unit vector (without changing the base point). However, the transformations $\begin{pmatrix} a & \sqrt{1-a^2} \\ -\sqrt{1-a^2} & a \end{pmatrix}$, with $a \in [-1,1]$ maps the vertical upward directed unit vector i_i at the point i to every other rotated unit vector at i.

3.1.3 Fuchsian groups and fundamental domains

We denote with $\text{Aut}\mathbb{H} = PSL(2,\mathbb{R})$ again the automorphism group of $\mathbb{H}$.

Definition 3.6 *A subgroup Γ of* $\text{Aut}\mathbb{H}$ *acts* (PROPERLY) DISCONTINUOUSLY *on* $\mathbb{H}$ *if for every compact subset K of* $\mathbb{H}$ *the set* $g(K) \cap K$ *is non-empty only for finitely many elements* $g \in \Gamma$.

Definition 3.7 *We say that a group* $\Gamma \subset \text{Aut}\mathbb{H}$ *of Möbius transformations is a* FUCHSIAN GROUP *on the Poincaré half plane if Γ acts discontinuously on* $\mathbb{H}$.

Definition 3.8 *A point* $\zeta \in \mathbb{R} \cup \{\infty\}$ *is called a* LIMIT POINT *of a Fuchsian group Γ if there exists a* $z \in \mathbb{H}$ *such that ζ is an accumulation point of* $\{gz : g \in \Gamma\}$. *The* LIMIT SET $\mathcal{L}$ *of Γ is the union of all limit points of Γ.*

Definition 3.9 *A Fuchsian group is called a* FUCHSIAN GROUP OF THE FIRST KIND *if* $\mathcal{L} = \partial\mathbb{H} = \mathbb{R} \cup \{\infty\}$.

Definition 3.10 *A* FUNDAMENTAL DOMAIN *or* FUNDAMENTAL REGION *of a Fuchsian group Γ on* $\mathbb{H}$ *is a non-empty connected open subset* $\mathcal{F} \subset \mathbb{H}$ *with the property*

$$\bigcup_{g\in\Gamma} g(\bar{\mathcal{F}}) = \mathbb{H}, \; g(\mathcal{F}) \cap h(\mathcal{F}) = \emptyset, \; \text{if } g \neq h, \text{ for all } g,h \in \Gamma,$$

where $\bar{\mathcal{F}}$ *denotes the closure of* $\mathcal{F}$ *with respect to* $\mathbb{H}$.

These definitions can equally well be formulated for $\mathbb{D}$ instead of $\mathbb{H}$.
$\mathcal{F}$ is therefore a fundamental domain for Γ if and only if every point in $\mathbb{H}$ lies in the closure of some image $g(\mathcal{F})$ and two different images are disjoint. No two points in $\mathcal{F}$ are Γ-equivalent. We also talk about a tessellation of the Poincaré half plane.
Geodesically convex [9] and locally finite [10] fundamental domains whose boundary consists of a finite and even number of geodesic segments (possibly intervals on $\mathbb{R}$ resp. $\mathbb{D}$) which are paired by elements of Γ are of special interest. We say that such fundamental domains satisfy the condition (P^*) and call them also PROPER fundamental domains.
Given a proper fundamental domain for Γ there are two possibilities to build a quotient space, namely: 1. Identify (i.e. glue together) pairwise corresponding sides of $\mathcal{F}_0$. 2. Form the quotient space $\mathbb{H}/\Gamma$ in the usual way. It is well-known that the two resulting spaces are the same and that $\mathbb{H}/\Gamma$ is a Riemannian manifold with constant negative curvature. A proof for this assertion can be found in the literature cited above.
In the following we will be interested solely in Fuchsian groups of the first kind which contain no elliptic elements. It is possible to show that for finitely generated Fuchsian groups of the first kind which contain no elliptic elements the fundamental domain can always be choosen as proper fundamental domain, cf. [93], Theorems 9.4.2 and 10.1.2.

Theorem 3.6 *Let Γ be a Fuchsian group of the first kind which contains no elliptic elements. Let* $\mathcal{F}$ *be a proper fundamental domain for Γ. The set* $\{g_i\}$ *of (uniquely determined) elements* $g_i \in \Gamma$ *which pair the sides of the fundamental domain* $\mathcal{F}$ *generates Γ, i.e. every* $g \in \Gamma$ *is a finite product of the* g_i *and* g_i^{-1}.

[9] Every geodesic arc connecting two points in $\mathcal{F}$ lies completely in $\mathcal{F}$.
[10] Every compact subset of $\mathcal{F}$ meets only finitely many Γ-images of $\mathcal{F}$.

This theorem follows easily from a theorem stated in [74], see also [22, 97, 100]. A much more general statement can be found in the book of Beardon [93], Theorem 9.3.3.

A Fuchsian group of the first kind has at most finitely many primitive elliptic conjugacy classes, at most finitely many primitive parabolic conjugacy classes and always infinitely many primitive hyperbolic conjugacy classes. Here a conjugacy class is called PRIMITIVE if it is not a power of any other conjugacy class with exponent greater than 1.
Elliptic elements in Γ correspond to ramification points on the Riemann surface $\mathbb{H}/\Gamma$. To every primitive parabolic conjugacy class there corresponds a (parabolic) cusp in the fundamental domain, i.e. a vertex of the fundamental domain lying on $\mathbb{R}$ with interior angle 0. The fundamental domain $\mathbb{H}/\Gamma$ is compact if and only if Γ contains no parabolic elements; in this case Γ is also called COCOMPACT.
For proofs of all this assertions, see the literature cited above and references therein.
Below we will need a technical condition on the fundamental domains:

Definition 3.11 *Let* $\mathcal{F}$ *be a proper fundamental domain of a Fuchsian group Γ and let N be the union of all images of* $\partial\mathcal{F}$ *under Γ. Then we say* $\mathcal{F}$ *has* EVEN CORNERS *if N is the union of (complete) geodesics in* $\mathbb{H}$.

These definition states that the geodesic completion of every side of $\mathcal{F}$ lies in the tessellation $\mathcal{T}$. This condition is much less restrictive than it may seem at first sight. E.g. the standard fundamental domain of $SL(2,\mathbb{Z})$ has even corners. Adler & Flatto [74] give a proof that for cocompact Fuchsian groups it is always possible to choose a fundamental domain that has even corners.

The geodesic flow on $\mathbb{H}/\Gamma$

Let Γ be a Fuchsian group, then we have seen above that every $g \in \Gamma$ induces a map $g : T_1\mathbb{H} \to T_1\mathbb{H}$ (cf. Equation 53). Therefore Γ induces a group $\tilde{\Gamma}$, which acts on $T_1\mathbb{H}$.
We introduce some terminology, namely $\mathbb{M} := \{\Gamma z \mid z \in \mathbb{H}\}$, $\mathcal{M} := \{\tilde{\Gamma} u \mid u \in T_1\mathbb{H}\}$, where $\Gamma z := \{gz \mid g \in \Gamma\}$ and $\tilde{\Gamma} u := \{\tilde{g}u \mid g \in \Gamma\}$.
The projections of $\mathbb{H}$ on $\mathbb{M}$ resp. of $T_1\mathbb{H}$ on $\mathcal{M}$ are denoted by π resp. $\tilde{\pi}$

$$\pi(z) = \Gamma z \text{ and } \tilde{\pi}(u) = \tilde{\Gamma} u.$$

The topology on $\mathbb{M}$ and $\mathcal{M}$ is determined by the requirement that π and $\tilde{\pi}$ are continuous, i.e. a subset $\mathbb{M}_1 \subset \mathbb{M}$ is called open if the inverse image $\pi^{-1}\mathbb{M}_1$ is open in $\mathbb{H}$ and similarly $\mathcal{M}_1 \subset \mathcal{M}$ is called open if $\tilde{\pi}^{-1}\mathcal{M}_1$ is open in $T_1\mathbb{H}$. We identify $\mathbb{M}$ with some fundamental domain $\mathcal{F}$ of Γ and $\mathcal{M}$ with the set of all unit vectors with base point in $\mathcal{F}$.
The geodesic flow $\{T^s\}_{s\in\mathbb{R}}$ on $\mathbb{H}$ induces a geodesic flow $\{\tilde{T}^s\}_{s\in\mathbb{R}}$ on $\mathbb{M}$ and we have

$$\tilde{T}^s = \pi T^s.$$

Two geodesics γ_1, γ_2 in $\mathbb{H}$ induce the same geodesic in $\mathbb{M}$ if and only if there exist a $g \in \Gamma$ such that $\gamma_2 = g\gamma_1$.
The geodesic flow on $\mathbb{M}$ is ergodic with respect to usual Liouville measure on $\mathcal{M}$, that is the measure on $\mathcal{M}$ induced by the measure $dm = dAd\theta = \frac{dxdyd\theta}{y^2} = \frac{2d\gamma_\infty d\gamma_{-\infty}ds}{(\gamma_\infty - \gamma_{-\infty})^2}$ on $\mathbb{H}$ given above.
The proof can be traced back to Artin [72], *Hedlund and Hopf and is given in e.g.* [22].

Periodic geodesics in $I\!H/\Gamma$

Let $\Gamma \subset \mathrm{Aut}I\!H$ be a Fuchsian group, then it is possible to introduce - as in the Definitions 3.2 and 3.3 for $PSL(2,\mathbb{C})$ - in a simple manner conjugacy classes in Γ.

Definition 3.12 *Two elements $g_1, g_2 \in \Gamma$ are called* CONJUGATE IN Γ *if there exists some $h \in \Gamma$ such that*

$$g_1 = hg_2h^{-1}.$$

The equivalence relation defined hereby allows a natural division of Γ in CONJUGACY CLASSES IN Γ.

Let γ_p be a geodesic in $I\!H$ such that $\pi\gamma_p$ is a periodic geodesic in $I\!M$ with primitive period t_0, i.e. $\pi\gamma_p(t + t_0) = \pi\gamma_p(t)$, for all t, then we have for all t:

$$T^{t_0}\pi(\gamma_p(t), \theta(t)) = \tilde{\pi}(\gamma_p(t + t_0), \theta(t + t_0)) = \tilde{\pi}(\gamma_p(t), \theta(t)).$$

Here $(\gamma_p(t), \theta(t))$ denotes the unit vector tangent to γ_p at the point $\gamma_p(t)$. *There exists therefore a canonical one-to-one correspondence between periodic geodesics in $I\!M$ and periodic points of the geodesic flow on $\mathcal{M}$.*
We have seen in Theorem 3.4 that there exists to every hyperbolic $g \in \mathrm{Aut}I\!H, g \neq id$, exactly one geodesic γ_g in $I\!H$ which is invariant under the action of g. Conversely, given a geodesic γ in $I\!H$, it is possible to construct a hyperbolic isometry g_γ in $\mathrm{Aut}I\!H$ which leaves γ invariant. *There exists therefore a one-to-one correspondence between geodesics in $I\!H$ and hyperbolic isometries of $I\!H$.*
A geodesic γ in $I\!H$ is projected to a *periodic* geodesic in $I\!M$ if and only if $g_\gamma \in \Gamma$. Furthermore, we see that two geodesics γ_1 and γ_2 are projected to the same periodic geodesic in $I\!M$ if and only if g_{γ_1} is conjugate in Γ to g_{γ_2}. As above, we call a hyperbolic isometry $g \in \Gamma$ PRIMITIVE if it can not written as a power of any other isometry $g' \in \Gamma$. Primitive hyperbolic isometries in Γ correspond to primitive periodic geodesics in $I\!M$, whereas non-primitive hyperbolic isometries in γ correspond to multiple traversals of the periodic geodesics in $I\!M$. In summary, we have

Theorem 3.7 *There exist a canonical bijection between primitive hyperbolic conjugacy classes in Γ and periodic geodesics in $I\!M$.*

Every hyperbolic isometry $g \in \Gamma$ in $\mathrm{Aut}I\!H$ is conjugate to some matrix of the form $\begin{pmatrix} e^{\frac{\ell}{2}} & 0 \\ 0 & e^{-\frac{\ell}{2}} \end{pmatrix}$.
Therefore the length of the periodic geodesic in $I\!M$ corresponding to g coincides with the distance between z and hz, where $h = \begin{pmatrix} e^{\frac{\ell}{2}} & 0 \\ 0 & e^{-\frac{\ell}{2}} \end{pmatrix}$ and $z \in I\!H$ arbitrary, e.g. $-iz = y_0 \in I\!R^+$:

$$\ell = \int_{iy_0}^{ie^{\ell}y_0} ds = \int_{y_0}^{e^{\ell}y_0} \frac{dy}{y} = 2\mathrm{arcosh}\frac{\mathrm{tr}g}{2}. \qquad (54)$$

On the other hand, it is easy to check that the maximal eigenvalue λ_{max} of g satisfies

$$\ln \lambda_{max} = \mathrm{arcosh}\frac{\mathrm{tr}g}{2} = \frac{\ell}{2}.$$

Finally, we claim

$$\ell = \pm \ln |g'(x_{fix})| > 0. \qquad (55)$$

In this equation x_{fix} denotes one of the two fixed points of g on $\partial I\!D$. The sign in Equation 55 is choosen such that $\ell > 0$.
For the matrix $h := \begin{pmatrix} e^{\frac{\ell}{2}} & 0 \\ 0 & e^{-\frac{\ell}{2}} \end{pmatrix}$ one fixed point is $x_{fix} = 0$ and it is easy to check that $\ell = +\ln|h'(0)|$. The right hand side of this equation is conjugacy invariant; therefore the assertion $\ell = +\ln|g'(x_{fix,1})|$ is true for one fixed point $x_{fix,1}$ of g. For the other fixed point $x_{fix,2}$ of g, we notice that $\ln|g'(x_{fix,2})| = -\ln|(g^{-1})'(x_{fix,2})|$ and prove completely analogously $\ell = -\ln|g'(x_{fix,2})|$ by considering $h^{-1} = \begin{pmatrix} e^{-\frac{\ell}{2}} & 0 \\ 0 & e^{\frac{\ell}{2}} \end{pmatrix}$ instead of h.
We will in the following always choose the orientation of the geodesic γ in $I\!H$ corresponding to g such that: $\ell = +\ln|g'(\gamma_\infty)|$ and $\ell = -\ln|g'(\gamma_{-\infty})|$. The inverse oriented geodesic γ^{-1} is associated with the isometry g^{-1} and we have: $\ell = +\ln|(g^{-1})'((\gamma^{-1})_\infty)| = -\ln|g'(\gamma_{-\infty})|$ and $\ell = -\ln|(g^{-1})'((\gamma^{-1})_{-\infty})| = +\ln|g'(\gamma_\infty)|$.

3.2 General remarks

In this section we study symbolic dynamics for the geodesic flow on $I\!H/\Gamma$, where Γ is a Fuchsian group of the first kind. However, in order to avoid certain technical subleties, we have to restrict ourselves to a special class of Fuchsian groups, specified below. We follow the work of C. Series et al. [22, 23, 25, 26, 27, 28]. Related results can also be found in Adler und Flatto [71, 74].

From now on, let Γ denote a given Fuchsian group of the first kind.
In this section we will use two different methods to code *oriented* geodesics in $I\!D$ and study their interrelation. [11] First, in the KOEBE-MORSE METHOD a geodesic in $I\!D$ is coded by the so called CUTTING SEQUENCE, i.e. the sequence in which the geodesic cuts a fixed set of curves in $I\!D$. This fixed set is subsequently always chosen to be the Γ-image of the sides of a fixed fundamental domains $\mathcal{F}$ of the considered Fuchsian group of the first kind Γ. As remarked on page 43, it is always possible to find a proper fundamental domain of a Fuchsian group of the first kind Γ, i.e. a geodesically convex polygon whose boundary consists of a finite and even number of geodesic segments which are paired by elements of Γ. By Theorem 3.6, the set of those elements in Γ which pair the sides of the fundamental domain generate Γ; every element in Γ can be written as finite product of the generators and their inverses.
We restrict our considerations to fundamental domains $\mathcal{F}$ of Γ which satisfy the following condition

0. $\mathcal{F}$ is a proper fundamental domain.

We are only interested in those geodesics in $I\!D$ which intersect $\mathcal{F}$. The Γ-images of $\mathcal{F}$ tessellate the whole disc. The Koebe-Morse method associates with every such geodesic a bi-infinite (possibly terminating) sequence of generators of Γ. Our fixed fundamental domain $\mathcal{F}$ singles out the zero position in the cutting sequence.
On the other hand, there is the ARTIN METHOD. Here both endpoints of the geodesics are mapped suitably (i.e. by the Bowen-Series map defined below) to some one-sided infinite sequence of generators of Γ. The BOUNDARY EXPANSIONS for both endpoints ξ and η can be combined to a single bi-infinite sequence $\xi^{-1} * \eta$. ξ denotes the boundary expansion of $\gamma_{-\infty}$ and η denotes the boundary expansion of γ_∞. For the Koebe-Morse method we have to study the set of geodesics

[11] We will consider the Poincaré disc $I\!D$ instead of $I\!H$ in the following.

which intersect the fixed fundamental domain $\mathcal{F}$ [12]

$$\mathcal{R} = \{\gamma : \gamma \cap \mathcal{F} \neq \emptyset\}. \qquad (56)$$

and for the Artin method the set $\mathcal{A}$ of those geodesics whose composited boundary expansions *satisfy certain admissibility conditions* (i.e. the composited boundary expansions of the geodesics in $\mathcal{A}$ should form a topological Markov shift Σ_A; details are given below).
We will see below that the sets $\mathcal{A}$ and $\mathcal{R}$ are not equal for the Fuchsian groups considered in this section (cf. Theorem 3.11).
However, there exists a bijection T between $\mathcal{A}$ and $\mathcal{R}$ such that for 'most' geodesics: $T = id$. Furthermore, T is a conjugacy; cf. Theorem 3.14. Geodesics which pass near to some vertex in the tessellation cause sometimes trouble: it may happen that the boundary expansion of some geodesic corresponds to the cutting sequence of another geodesic which passes the considered vertex on the other 'wrong' side.
Furthermore, we will see that the map which determines the boundary expansions (the Bowen-Series map) is a Markov map. The Bowen-Series map for cocompact Fuchsian groups is expansive by Theorem 3.10.
The shift operator on the space of cutting sequences is, however, in general not Markov!

In summary, the purpose of this section is to map the geodesic flow on $\mathbb{D}/\Gamma$ by a natural conjugacy T to some topological Markov shift for a broad class of Fuchsian groups Γ, .

However, for an arbitrary Fuchsian group of the first kind it is not known how to code the geodesic flow on $\mathbb{D}/\Gamma$ by a symbolic dynamical system. Therefore, we will restrict ourselves to a smaller class of Fuchsian groups and to fundamental domains satisfying respectively:

1. The fundamental domain $\mathcal{F}$ has even corners (cf. Definition 3.11);

2. Γ contains no elliptic elements.

The first requirement simplifies the geometric considerations below. Without this requirement it would be not possible to unambigiously associate with every geodesic a cutting sequence.
Therefore we will choose $\mathcal{F}$ always as proper fundamental domain with even corners
We will prove some central theorems only for Fuchsian groups of the first kind satisfying condition 2. Condition 2 can be replaced by two weaker requirements on $\mathcal{F}$, cf. [23]:

2a. $\mathcal{F}$ has at least four sides.

2b. If $\mathcal{F}$ has four sides and no vertex of $\mathcal{F}$ lies on $\partial\mathbb{D}$, then at least three geodesics in N cross at each vertex of $\mathcal{F}$.

Here N denotes the net of all Γ-images of $\partial\mathcal{F}$. For sake of simplicity we will restrict ourselves to Fuchsian groups satisfying condition 2. By Consequence 1 (see below), it follows that condition 2 implies the conditions 2a and 2b.
We are mainly interested in the case when the Markov map which determines the boundary expansions is expansive: without proof we state that this is the case if and only if Γ contains no parabolic elements (cf. Theorem 3.10 and [25]). Therefore, we have to require the following additional restriction:

[12] In Equation 57 below we will give a slightly different definition of $\mathcal{R}$.

3. Γ contains no parabolic elements.

In summary, we only consider in this work Fuchsian groups Γ and fundamental domains $\mathcal{F}$ which satisfy the conditions 0, 1, 2 *and* 3. However, we will state without proof some results for Fuchsian groups which contain parabolic elements. All results obtained and discussed below remain valid *if the condition 2 is replaced by the two conditions 2a and 2b; we refer the interested reader to* the literature cited above.
Now we have immediately the following consequence:

Consequence 1 *If Γ and $\mathcal{F}$ satisfy the conditions 0, 1, 2 and 3, then $\mathcal{F}$ has more than four sides.*

Supposed $\mathcal{F}$ is a triangle, then there would exist a ramification point on the surface. This is a contradiction since Γ contains no elliptic elements. If $\mathcal{F}$ has four sides, then pairwise glueing of different sides results in a torus or in a surface with ramification points. Since the genus g of a compact surface with constant negative curvature is at least 2, cf. [92], the torus has to be excluded.

3.3 The Koebe-Morse method

3.3.1 The Cayley graph of Γ

At the beginning of this section we introduce some terminology which we will need in the sequel. Let $\mathcal{F}$ be a fundamental domain of a Fuchsian group of the first kind Γ such that the conditions 0, 1, 2 and 3 are satisfied. Every side s of the fundamental domain $\mathcal{F}$ is identified by some $g_s \in \Gamma$ with another side $s' = g_s^{-1}s \in \partial\mathcal{F}$. By Theorem 3.6, the set

$$\mathcal{G} := \{g_s \mid s \text{ is side of } \mathcal{F}\}$$

generates Γ. Every side s of the fundamental domain carries two labels: one inside and one outside $\mathcal{F}$. In the interior of $\mathcal{F}$ we label every side s of $\mathcal{F}$ by the associated generator g_s^{-1} belonging to *s and we write this label on the interior of the side s. In the other fundamental domain adjacent* to the side s the side s is labelled by the inverse generator g_s; i.e. the label of s outside $\mathcal{F}$ is the generator g_s (cf. Figure 3.1). We denote with $C(g_s)$ the geodesic completion of s. $C(g_s)$ satisfies the relation $g_s^{-1}C(g_s) = C(g_s^{-1})$.
We denote with N the net of Γ-images of $\partial\mathcal{F}$. Since $\mathcal{F}$ has even corners, N is the union of complete geodesics in $\mathbb{D}$. Consider the fundamental domain $g_1\mathcal{F}$ and an adjacent fundamental *domain $g_2\mathcal{F}$ with the common side s, then we label the side s in the interior of $g_1\mathcal{F}$ with the* generator $g_2^{-1}g_1 \in \mathcal{G}$ and in the interior of $g_2\mathcal{F}$ with the generator $g_1^{-1}g_2 \in \mathcal{G}$. We suppose that $\mathcal{F}$ is so choosen that $0 \in \mathcal{F}$ holds. Since Γ contains no elliptic elements, 0 is not a fixed point of some element in Γ. (Instead of 0 we could equally well consider every other point in $\mathcal{F}$.)
We introduce now the notion of the CAYLEY GRAPH $K(\Gamma, \mathcal{G})$ of Γ relatively to $\mathcal{G}$: the vertices of the Cayley graph are the elements of the group Γ and two vertices g_1, g_2 are related by an oriented edge from g_1 to g_2 if and only if $g_1^{-1}g_2 \in \mathcal{G}$.
Since $\mathcal{G}$ generates Γ, the Cayley graph $K(\Gamma, \mathcal{G})$ is connected. We can think of the Cayley graph as being imbedded in the Poincaré disc, as the dual net N^* to N. The vertices of dual net N^* to N are exactly the Γ-images of 0 and two points g_10, g_20 are joined by an edge if and only if $g_1\mathcal{F}$ and $g_2\mathcal{F}$ have a common side, i.e. $g_1^{-1}g_2 \in \mathcal{G}$ holds. An oriented edge in N^* joining g_10 to g_20 is denoted by $g_1^{-1}g_2 \in \mathcal{G}$.

A path in N^*, also called an EDGE PATH, which joins successively the vertices $g0, g_0g0, g_0g_1g0, \dots, g_0g_1\dots g_ng0$, where $g_i \in \mathcal{G}$, corresponds to the word $g_0g_1\dots g_n$. On the other hand, every word of generators $g_0g_1\dots g_n$ determines an element $\tilde{g} \in \Gamma$ by $\tilde{g} = g_0g_1\dots g_n$. To every edge path we can associate a unique POLYGONAL PATH, consisting of the sequence of adjacent fundamental domains $g\mathcal{F}, g_0g\mathcal{F}, g_0g_1g\mathcal{F}, \dots, g_0g_1\dots g_ng\mathcal{F}$ and conversely, to every polygonal path there corresponds a unique edge path. Sometimes we will consider one-sided infinite and also bi-infinite edge or polygonal paths.

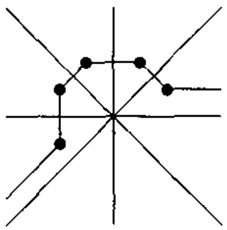

Figure 3.2: half cycle

We are solely interested in paths which never reverse their direction at a vertex; i.e. the corresponding word does not contain a subsequence of the form gg^{-1}. Such paths and words are called REDUCED. The length of a path in N^* is the number of edges contained in it.
A finite path in N^* is called a SHORTEST PATH if the corresponding word is the shortest representation by generators in $\mathcal{G}$ of the element in Γ which is defined by the word. An infinitely long edge resp. polygon path is called SHORTEST PATH if every finite subpath is shortest.

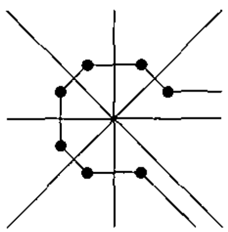

Figure 3.3: long cycle

Pick any vertex v in N. We denote with $n(v)$ the number of geodesics in N which pass through the vertex v. Therefore the number of copies of $\mathcal{F}$ meeting at the vertex v is $2n(v)$. A closed, oriented edge path $K(v)$ in N^* which passes successively through all copies of $\mathcal{F}$ which meet at v consists therefore of $2n(v)$ edges. $K(v)$ goes round v exactly once. Every subpath of $K(v)$ will be called a CYCLE around v. A cycle around the vertex v which contains $n(v)$ edges is also called a HALF CYCLE (cf. Figure 3.2).
A LONG CYCLE contains more than $n(v)$ edges (cf. Figure 3.3). If an edge path contains a long cycle, then the edge path intersects at least one of the geodesics in N which pass through v twice.

We say that a cycle is CLOCKWISE or LEFT-HANDED if the cycle goes clockwise round the corresponding vertex; we say that a cycle is ANTI-CLOCKWISE or RIGHT-HANDED if the cycle goes counter-clockwise round the corresponding vertex. We also write simply L-cycle resp. R-cycle. [13]
Now let $K(v_1), K(v_2)$ be two closed, oriented edge paths in N^* round the vertices v_1 and v_2 with a common edge $\bar{AB}$ and let Z_1, Z_2 be two cycles contained in $K(v_1), K(v_2)$ respectively. If Z_1 ends at B and Z_2 begins at B and if both Z_1 and Z_2 do not pass through A, then the cycles Z_1 and Z_2 are called CONSECUTIVE (cf. Figure 3.4).

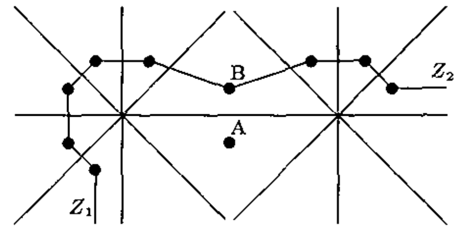

Figure 3.4: consecutive cycles

A sequence of consecutive cycles $Z_1 \subset K(v_1), Z_2 \subset K(v_2), \dots, Z_n \subset K(v_n)$ is called a CHAIN if the corresponding vertices $v_1, v_2, \dots, v_n$ in N are consecutive vertices lying on a geodesic $\tilde{\gamma} \in N$. We also say in this case that a chain RUNS ALONG $\tilde{\gamma}$. A chain of cycles of lengths $n(v_1), n(v_2)-1, n(v_3)-1, \dots, n(v_{n-1})-1, n(v_n)$, where v_i denotes the ith vertex belonging to the ith cycle in the chain is called a LONG CHAIN. A long chain intersects at least one geodesic in N twice, namely the geodesic which passes through the vertices $v_1, v_2, \dots, v_n$ and along which the long chain runs. An infinitely long chain of cycles of lengths $n(v_1)-1, n(v_2)-1, n(v_3)-1, \dots$, where v_i denotes the ith vertex belonging to the ith cycle in the chain, is called a PSEUDO HALF CYCLE. Pseudo half cycles behave similarly as half cycle, as we will see below.

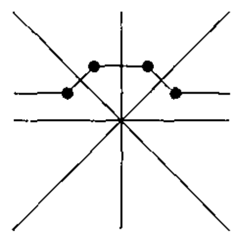

Figure 3.5a

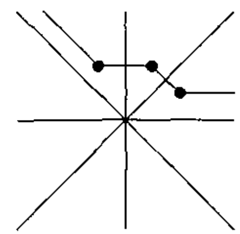

Figure 3.5b

Pick an arbitrary cycle or chain and denote it by $\mathcal{C}$. Then we can associate a COMPLEMENTARY CYCLE or a COMPLEMENTARY CHAIN $\tilde{\mathcal{C}}$ with $\mathcal{C}$ in an obvious way; $\mathcal{C}$ and $\tilde{\mathcal{C}}$ begin and end at the same vertices in N^*; $\mathcal{C}$ and $\tilde{\mathcal{C}}$ cross no common fundamental domain besides the initial and the

[13] This convention differs from that in Series [26].

final fundamental domain; no copy $g\mathcal{F}$ of the fundamental domain $\mathcal{F}$ lies strictly inside the region bounded by $\mathcal{C}$ and $\tilde{\mathcal{C}}$. We also say that $\mathcal{C}$ and $\tilde{\mathcal{C}}$ are *neighbouring* edge paths, see Corollary 3.1.
Let K be an edge path containing a cycle round the vertex v. We say that the VERTEX ANGLE of K at v is π^-, π, π^+ if the number fundamental domains in K adjacent to v equals $n(v)-1, n(v), n(v)+1$ respectively.
Suppose an edge path K has at v the angle π, then K runs before and after passing the vertex v along the same geodesic in N (cf. Figure 3.5a). Suppose the edge path has not the angle π at v, then K runs before and after passing the vertex v along different geodesics in N. If the angle at v is π^+, then K contains a half cycle round v and intersects the geodesic in N along which it has run before passing v (cf. Figure 3.2). If the angle is π^- at v, then K intersects $n(v)-1$ fundamental domains meeting in v and does not intersect the geodesic in N along which it has run before passing v (Figure 3.5b).

3.3.2 Geodesic edge paths

Every geodesic γ in $\mathbb{D}$ defines a geodesic edge path as follows: Suppose that $\gamma \notin N$ and further that γ does not pass through any vertex of N. Then every geodesic arc of γ crosses a sequence of copies of the fundamental domain $h\mathcal{F}, h_1h\mathcal{F}, h_1h_2h\mathcal{F}, ..., h_1h_2...h_nh\mathcal{F}$, where $h \in \Gamma$ and $h_j \in \mathcal{G}$. This defines the GEODESIC POLYGONAL PATH associated with γ. The corresponding edge path $h0, h_1h0, h_1h_2h0, ..., h_1h_2...h_nh0$, where $h_j \in \mathcal{G}$, is called a GEODESIC EDGE PATH. The associated word $h_1h_2...h_n$, with $h_j \in \mathcal{G}$ is also called a GEODESIC WORD.

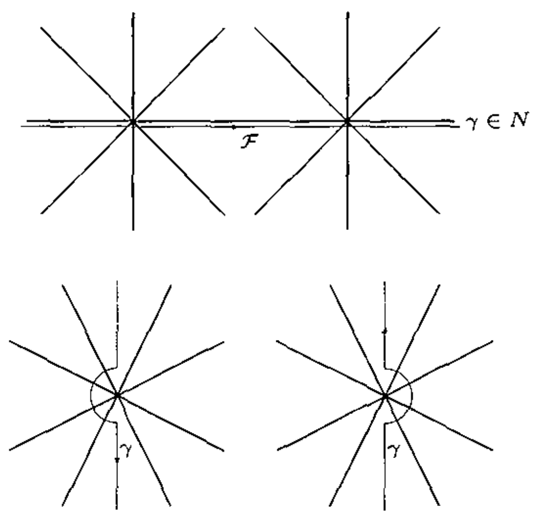

Figure 3.6

If $\gamma \in N$, then we consider instead of γ an infinitesimally deformed curve as shown in Figure

3.6 and determine the edge path associated with this curve. If γ crosses a vertex of N, then we deform γ slightly in the neighbourhood of v such that the deformed curve does not meet the vertex and instead goes around the vertex; we determine then the edge path of the deformed curve. Notice that in both cases there are two possibilities to deform γ. We make the convention that we deform γ always to the right, see Figure 3.6. *In the following it is understood that all geodesics, if necessary, have been deformed in this way.*

Every geodesic defines therefore in a canonical way a doubly infinite geodesic edge path $E(\gamma) = ...h_{-1}\, h_{-2}\, h0, h_{-1}\, h0, h0, h_1\, h0, h_1\, h_2\, h0, ...$, where $h_j \in \mathcal{G}$, resp. the associated infinite geodesic polygonal path $P(\gamma)$. The associated bi-infinite word $...h_{-2}^{-1}h_{-1}^{-1}h_1h_2...$, with $h_j \in \mathcal{G}$, for $j \neq 0$, will be called the CUTTING SEQUENCE of γ. The space of all cutting sequences furnished with the shift operator σ is a symbolic dynamical system in the sense of Definition 2.1; it is, however, in general not a topological Markov shift.
For a geodesic $\gamma \notin N$ which intersects the fundamental domain $\mathcal{F}$ we have: $0 \in E(\gamma)$. It is useful to consider instead of the geodesics in N which intersect $\mathcal{F}$ only those geodesics in N whose deformation - as described above - still intersects $\mathcal{F}$. To be precise, we will not define $\mathcal{R}$ by Equation 56, but we set

$$\mathcal{R} := \{\gamma \mid 0 \in E(\gamma)\}. \tag{57}$$

From now on we denote with $\mathcal{R}$ always the set defined by Equation 57.
By Definition 3.9, it follows that two different, oriented geodesics in $\mathcal{R}$ have different cutting sequences. On the other hand, different cutting sequences belong to different geodesics in $\mathcal{R}$. Since the correspondence between cutting sequences and geodesics in $\mathcal{R}$ is bijective, we can identify $\mathcal{R}$ with the space of all cutting sequences.
Next, we want to state an important theorem characterizing shortest edge paths and, furthermore, we will see that geodesic edge paths are shortest edge paths. First, we need:

Lemma 3.1 *Let s, s' be non-adjacent sides of the fundamental domain $\mathcal{F}$, then the geodesic completions $C(s), C(s') \in N$ of s, s' do not intersect.*

Proof: [25] Suppose that $C(s)$ and $C(s')$ meet at some point $P \notin \partial\mathbb{D}$. We want to show first that without loss of generality we can assume that s and s' are separated by only one side of $\mathcal{F}$. We label the sides of $\mathcal{F}$ between s and s' successively by $s = s_0, s_1, ..., s_p = s'$. Let A denote the intersection point of s and s_1 and let B denote the intersection point of s_{p-1} and s'. Furthermore, let γ denote a geodesic relating A and B. The fundamental domain $\mathcal{F}$ is a proper fundamental domain, hence geodesically convex. The geodesic arc joining A and B is therefore completely contained in $\mathcal{F}$. The side s_{p-1} intersects both γ and $C(s')$ at B. This implies that $C(s_{p-1})$ intersects also $C(s)$. Proceeding inductively, we can assume that s and s' are separated by exact one side s_1 of $\mathcal{F}$.
s_1 is a side of $\mathcal{F}$. The fundamental domain on the other side of s_1 is denoted by $\phi(\mathcal{F})$. s_1 is also a side of $\phi(\mathcal{F})$. The fundamental domain $\phi(\mathcal{F})$ is contained in the triangle APB, but, however, $\phi(\mathcal{F}) \neq APB$ since $\mathcal{F}$ and therefore also $\phi(\mathcal{F})$ contains at least five sides. We denote the two sides of $\phi(\mathcal{F})$ adjacent to s_1 with t and t', respectively. The geodesic completions $C(t), C(t')$ of the sides t, t' meet at some point P_1 in APB or on the boundary of APB since otherwise $C(t)$ or $C(t')$ would intersect some side of APB twice. We now can proceed inductively: we apply the above argument for s, s' and $\mathcal{F}$ to t, t' and $\phi(\mathcal{F})$ and obtain a copy $\phi_2(\mathcal{F})$ of the fundamental domain in ABP_1. $\phi_2(\mathcal{F})$ has two non-adjacent sides u and u' which meet at some point P_2. Proceeding inductively we obtain infinitely many disjoint copies of the fundamental domain $\mathcal{F}$ in the interior

of the triangle ABP. This, however, is impossible since ABP has finite non-Euclidean area.
Now suppose that $C(s)$ and $C(s')$ meet at some point $P \in \partial ID$. Since $\mathcal{F}$ has at least five sides, it is easy to see that in every case $P_2 \notin \partial ID$. Suppose that is not the case. Then $P = P_1 = P_2 \in \partial ID$. Consider the side t_m of $o(\mathcal{F})$ adjacent to t' and different from s_1. The geodesic completion of t_m cuts $C(t')$ between P_1 and B and again at $P_1 = P_2$. This is a contradiction and therefore $P_2 \notin \partial ID$. Now the above argument applies and we are done. □

Theorem 3.8

- *A finite edge path is a shortest edge path if and only if it intersects every geodesic in N at most once.*
- *A finite edge path is reduced and contains no long cycles and no long chains if and only if it intersects every geodesic in N at most once.*
- *A geodesic edge path is a shortest edge path.*

Proof: A proof of the last item and partly of the first two items can be found in Birman & Series [23]. We give another proof.
Given an edge path K_{if} relating the two copies of the fundamental domain $\mathcal{F}_i = h_i\mathcal{F}$ and $\mathcal{F}_f = h_f\mathcal{F}$. These two fundamental domains are separated by a certain subset $N_{if} \subset N$ of geodesics in N. Every path K'_{if} relating both fundamental domains intersects every geodesic in N_{if} at least once, however, always an odd number of times. K'_{if} intersects every geodesic in $N \setminus N_{if}$ an even number of times. Obviously, K'_{if} is a shortest edge path if and only if K'_{if} intersects the geodesics in N_{if} exactly once *and* if K'_{if} intersects no geodesics in $N \setminus N_{if}$. The first assertion follows.
Let $K_{if} \subset E(\gamma)$ be the edge subpath contained in the geodesic edge path $E(\gamma)$ associated with the geodesic γ which joins the vertices $h_i0 = h_10, h_20, \ldots, h_n0 = h_f0$. Supposed that it is possible to find a geodesic $\gamma_N \in N$ such that K_{if} intersects γ_N twice: first between h_10 and h_20 and second between $h_{n-1}0$ and h_n0. Then γ_N *contains* the sides $h_1\mathcal{F} \cap h_2\mathcal{F}$ and $h_{n-1}\mathcal{F} \cap h_n\mathcal{F}$. But also γ intersects these two sides; and therefore γ_N and γ meet twice; this is a contradiction and we have proved that every edge subpath of geodesic edge path $E(\gamma)$ intersects a geodesic in N at most once. Thus the third assertion follows.
A non-reduced edge path intersects some geodesic in N twice in succession. An edge path which contains a long cycle or a long chain intersects a geodesic in N twice. Thus it follows that every edge path which intersects every geodesic in N at most once is reduced and contains no long cycles and no long chains.
It remains to prove the converse direction. Let K_{if} be a reduced edge path which contains no long cycles and no long chains. We suppose that K_{if} intersects some geodesic in N more than once. We prove first the following statement:

* Let $\gamma_1 \in N$ be the first geodesic which is intersected a second time by K_{if}. K_{if} intersects a sequence of fundamental domains between the two intersection points with γ_1 and every fundamental domain in this sequence has at least one of its vertices lying on γ_1. K_{if} intersects no fundamental domain $\mathcal{F}_0$ which satisfies $\mathcal{F}_0 \cap \gamma_1 = \emptyset$.

Supposed this assertion is false. Then K_{if} goes around some fundamental domain $\mathcal{F}_v$ (without intersecting it) which has one vertex v on γ_1. Let the completion of the two sides of $\mathcal{F}_v$ which meet in v to half geodesics be denoted with λ_1 and λ_2. K_{if} intersects both λ_1 and λ_2. Since $\mathcal{F}_v$ has at least five sides (Consequence 1), there exists some side s of $\mathcal{F}_v$ such that the geodesic completion $C(s) \in N$ of s does intersect neither λ_1 nor λ_2 by Lemma 3.1. Thus K_{if} intersects the geodesic $C(s)$ twice. This is a contradiction to the construction of γ_1. Therefore the assertion (*) is proved.
However, (*) states that K_{if} is either not reduced or contains a long cycle or a long chain. This proves the theorem. □

Corollary 3.1

* *Let K_1 and K_2 be two shortest edge paths with coincident initial and final points. Then no copy $g\mathcal{F}$ of the fundamental domain $\mathcal{F}$ lies completely within the region bounded by K_1 and K_2.*

** *Let K_1 and K_2 be two shortest edge paths with coincident initial and final points on ∂ID. If K_1 and K_2 contain no right-handed half cycles and no right-handed pseudo half cycles, then K_1 and K_2 coincide.*

In the first case we also say that K_1 and K_2 are NEIGHBOURING EDGE PATHS.

Proof: (*): Suppose $\mathcal{F}'$ is a fundamental domain lying inside the region bounded by K_1 and K_2. Denote with s_1 resp. s'_1 the two sides of $\mathcal{F}'$ whose geodesic completions $C(s_1)$ resp. $C(s_2)$ intersect K_1 closest to the initial resp. final vertex. Similarly, denote with s_2 resp. s'_2 the two sides of $\mathcal{F}'$ whose geodesic completions $C(s'_1)$ resp. $C(s'_2)$ intersect K_2 closest to the initial resp. final vertex. Since $\mathcal{F}'$ has at least five sides either s_1 and s'_1 or s_2 and s'_2 are non-adjacent sides of $\mathcal{F}'$. We assume without loss of generality that s_1 and s'_1 are non-adjacent sides of $\mathcal{F}'$. By Lemma 3.1, there exists a side s of $\mathcal{F}'$ whose geodesic completion $C(s) \in N$ does not intersect the geodesic completions $C(s_2), C(s'_2)$ of s_2 and s'_2; and therefore K_2 intersects $C(s)$ twice.
Thus K_2 is not a shortest edge path by Theorem 3.8. Notice that the corollary remains true if the edge paths are infinitely long and the common initial and/or final points lie on ∂ID.
(**): The proof given in Series [28] is incomplete. We give a completed and modified proof. By (*), it is clear that K_1 and K_2 are neighbouring edge paths. We suppose that K_1 and K_2 are different. K_1 and K_2 run through a sequence of fundamental domains $\{\mathcal{F}_s\}_{s=-\infty}^{\infty}$ and $\{\mathcal{F}'_s\}_{s=-\infty}^{\infty}$ respectively.
Supposed the angle of both K_1 and K_2 at all vertices is π, then neither K_1 nor K_2 contain a right-handed pseudo half cycle. K_1 and K_2 run completely along some geodesic in N on opposite sides of the geodesic.
Supposed the vertex angles of K_1 and K_2 are not constantly equal to π at every vertex. Then there exists a vertex v_0 such that (i) the edge path K_1 passes through some fundamental domains adjacent to the vertex v_0 which is not crossed by the other edge path K_2 and vice versa and that (ii) one of the two edge paths (without loss of generality K_1) has the vertex angle π^+ at v_0. There exists vertices satisfying (i) since K_1 and K_2 are supposed to be different. Among the vertices satisfying (i) there exists a vertex also satisfying (ii). This can be deduced from the following two facts: (a) the following pairs of vertex angles are forbidden for neighbouring edge paths K_1/K_2 at vertices satisfying (i): π^-/π, π/π^- and π^-/π^- (K_1 and K_2 would not be neighbouring after passing such a vertex); (b) the assumption that at every vertex satisfying (i) the pair of vertex angles of K_1/K_2 equals π/π implies that at least one of the two edge paths contains a right-handed pseudo half cycle.
Since K_1 and K_2 run through non-common fundamental domains while passing the vertex v_0,

it follows that K_1 and K_2 run through a sequence of common fundamental domains either (1) before or (2) after passing the vertex v_0 or (3) neither before nor after passing v_0. We assume that K_1 and K_2 pass through a sequence of common fundamental domains at most *before* passing the vertex v_0. The other case can be treated similarly.
We denote with v_1 the next vertex after v_0 at which K_1 or K_2 has a vertex angle different from π. The angle of K_1 at v_1 is not π^+ since in this case K_1 would contain a long chain which begins at v_0 and ends at v_1.
Suppose the angle of K_1 at v_1 equals π^-, then the angle of K_2 at v_1 equals π^+ since otherwise K_1 and K_2 would not be neighbouring after passing the vertex v_1. If K_1 goes clockwise around v_0 then K_2 goes counter-clockwise around v_1 and vice versa; however, this is impossible by assumption since either K_1 or K_2 would contain a right-handed half cycle.
If the angle of K_1 at v_1 equals π, then the angle of K_2 at v_1 equals π^+ or π^-. We suppose first that the angle of K_2 at v_1 equals π^+. If K_1 goes around the vertex v_0 clockwise, then K_2 goes around the vertex v_1 counter-clockwise and vice versa; however, this is impossible by assumption since either K_1 or K_2 would contain a right-handed half cycle. On the other hand, the angle of K_2 at v_1 can also not be equal to π^- since otherwise K_1 and K_2 would not be neighbouring after passing the vertex v_1 with the anglepair π/π^-.
Finally, there is the possibility that the angles of both K_1 and K_2 are equal to π at every vertex after v_0. This implies that one of the two edge paths K_1 and K_2 contains a right-handed pseudo half cycle. □

3.4 The boundary expansion

Definition and properties of the Bowen-Series map

In this section we discuss the Bowen-Series map. With the help of the Bowen-Series map we will define the boundary expansion of points $\xi \in \partial\mathbb{D}$. We associate with every geodesic in $\mathbb{D}$ the boundary expansions of its endpoints; this method to code oriented geodesics is called the Artin method. Furthermore, we will discuss the most important properties of the Bowen-Series map. Especially, we will see that the Bowen-Series map is a Markov map.
We suppose as above that the considered Fuchsian group Γ and the considered fundamental domain $\mathcal{F}$ satisfy the conditions 0, 1, 2 and 3 from Section 3.2. Further, we label the oriented sides of $\mathcal{F}$ by generators in $\mathcal{G}$ in the way described above. The k generators on the k exterior sides of $\mathcal{F}$ are denoted in counter-clockwise direction as follows: $g_1, g_2, g_3, \ldots, g_k$. On the interior sides the labels are correspondingly: $g_1^{-1}, g_2^{-1}, \ldots, g_k^{-1}$.
As above, the complete geodesic in N which contains the side of $\mathcal{F}$ with the exterior label g_i is denoted by $C(g_i)$. Further, we denote with $A(g_i)$ the interval on $\partial\mathbb{D}$ whose endpoints are simultaneously the endpoints of $C(g_i)$ on $\partial\mathbb{D}$. Hereby we make the convention that the half plane bounded by $C(g_i)$ and $A(g_i)$ does not contain the fundamental domain $\mathcal{F}$. We write $A(g_i) = [P_i, Q_{i+1}]$. Thereby we make the convention that P_i precedes Q_i on $\partial\mathbb{D}$ in counter-clockwise direction. Furthermore, we set $Q_{k+1} = Q_1$ and $P_{k+1} = P_1$. Since Γ is cocompact (condition 3), the set $\{A(g_i) : i \in \{1, 2, \ldots, k\}\}$ is a finite covering of $\partial\mathbb{D}$ with $\cup_{g \in \mathcal{G}} A(g) = \partial\mathbb{D}$.

Remark: Without condition 3 this covering would be countably infinite. The following definition makes sense also when condition 3 does not hold.

Definition 3.13 (Bowen,Series) *The map*

$$f : \partial\mathbb{D} \to \partial\mathbb{D}, f(\xi) := g_i^{-1}(\xi), \text{ if } \xi \in [P_i, P_{i+1}[\tag{58}$$

is called Bowen-Series map *of* Γ *with respect to* $\mathcal{G}$.
The boundary expansion *of* $\xi \in \partial\mathbb{D}$ *is the (infinite) sequence of generators* $\xi_f = g_{i_0} g_{i_1} g_{i_2} \ldots$, *where* $g_{i_n} \in \mathcal{G}$ *and* $f^n(\xi) \in [P_{i_n}, P_{i_n+1}[$ *for all* $n \in \mathbb{N}$.
The backward Bowen-Series map $\bar{f}$ *is defined similarly by:*

$$\bar{f} : \partial\mathbb{D} \to \partial\mathbb{D}, \bar{f}(\xi) := g_i^{-1}(\xi) \text{ if } \xi \in]Q_i, Q_{i+1}]. \tag{59}$$

The backward boundary expansion *of* $\xi \in \partial\mathbb{D}$ *is the (infinite) sequence of generators* $\xi_{\bar{f}} = g_{i_0} g_{i_1} g_{i_2} \ldots$, *where* $g_{i_n} \in \mathcal{G}$ *and* $\bar{f}^n(\xi) \in]Q_{i_n}, Q_{i_n+1}]$ *for all* $n \in \mathbb{N}$.

The Bowen-Series map is defined asymmetrically with respect to the P_i and the Q_i. We want to code the endpoints of oriented geodesics by symbol sequences with the help of the Bowen-Series map. Under time-reversal P_i and Q_{i+1} interchange their roles; if one uses the Bowen-Series map f to code the positive endpoint γ_∞ of a geodesic γ, then it is natural to use the backward Bowen-Series map to code the negative endpoint $\gamma_{-\infty}$ since in the definition of the backward Bowen-Series map the asymmetry of the Bowen-Series map is reversed.
This is pure convention; however, this convention will result in a relatively simple structure of the space of composited symbol sequences; cf. Equation 64.
It is possible to consider the Bowen-Series map as a map on the interval $I = [0, 1]$; we identify the points 0 and 1 in I and map I to $\partial\mathbb{D}$ in such a way that relative lengths are preserved.
Next we state some important properties of the Bowen-Series map: [14]

Lemma 3.2 *There exists a finite set* $W \subset \partial\mathbb{D}$ *such that* $f(W) \subseteq W$.

Proof: [25] Let v_i be those vertex of $\mathcal{F}$ at which the two sides with exterior labels g_{i-1} and g_i meet. We denote with $N(v_i)$ the set of those geodesics in N which pass through the vertex v_i and with $W(v_i)$ the union of the set $\{P_i, Q_i\}$ with the set of all initial and final points on $\partial\mathbb{D}$ of all those geodesics γ in $N(v_i)$ which are not the geodesic completion of some side of $\mathcal{F}$.
We set now $W := \cup_{i=1}^k W(v_i)$. To $w \in W$ there exists an i such that $w \in [P_i, P_{i+1}[$. Thus: $w \in W(v_i) \cup W(v_{i+1})$.
Supposed $w \in W(v_i)$: Then application of f to w yields: $f(w) = g_i^{-1} w$. Since $g_i^{-1} v_i$ is also vertex of $\mathcal{F}$ and since N is Γ-invariant (and therefore especially f-invariant), $g_i^{-1}(\widehat{v_i w})$ is a geodesic arc of a geodesic in N which passes through the vertex $g_i^{-1}(v_i)$. Especially, we have therefore $g_i^{-1}(w) \in W(g_i^{-1} v_i) \subseteq W$. If $w \in W(v_{i+1})$, then similarly $f(w) \in W(g_i^{-1} v_{i+1}) \subseteq W$. Hence, we have seen that $f(W) \subseteq W$. □

Remark: The assertion of the lemma remains also true for Fuchsian groups containing parabolic elements provided one allows W to be also countable. The proof is similar.

The finite set W in Lemma 3.2 induces a partition $\mathcal{P} = \{I_i\}_{i=1}^w$ of $\partial\mathbb{D}$. By definition of W, the covering $\mathcal{P}$ is finer than the covering $\{[P_i, P_{i+1}] : i \in \{1, 2, \ldots, k\}\}$, i.e. for all $j \in \{1, \ldots, w\}$ there exists an $i \in \{1, \ldots, k\}$ such that $I_j \subset [P_i, P_{i+1}]$.

[14] All properties discussed in the following are also valid for the backward Bowen-Series map. In Lemma 3.5 all right-handed half cycles and right-handed pseudo half cycles have to be replaced by left-handed half cycles and left-handed pseudo half cycles respectively.

We now label all points in the set $W(v_i)$ defined in the proof of Lemma 3.2. The labels are illustrated schematically in Figure 3.1. The set $W(v_i)$ consists by construction of $2n(v_i)-2$ points, where $n(v_i)$ denotes the number of the geodesics in N which pass through the vertex v_i. One of the points in $W(v_i)$ is labelled by P_i and one by Q_i; one half of the other points precedes the point P_i (in the counter-clockwise direction). P_i is followed by Q_i and Q_i is followed by the second half of the other points. We denote the $n(v_i)-2$ points before P_i (in the counter-clockwise direction) successively by $S_{i,1}; S_{i,2}; \dots; S_{i,n(v_i)-2}$. We denote the $n(v_i)-2$ points following Q_i (in counter-clockwise direction) successively by $T_{i,n(v_i)-2}; T_{i,n(v_i)-3}; \dots; T_{i,1}$.
Starting at P_1 and and proceeding in counter-clockwise direction along $\partial I\!D$ we obtain the following sequence of points in W (here Lemma 3.1 is used):
$P_1; Q_1; T_{1,n(v_1)-2}; \dots; T_{1,1}; S_{2,1}; \dots; S_{2,n(v_2)-2}; P_2; Q_2; T_{2,n(v_2)-2}; \dots; \dots; \dots; T_{k,1}; S_{k+1,1} = S_{1,1}; \dots; S_{1,n(v_1)-2}; P_{k+1} = P_1$.
The Bowen-Series map f acts on P_i as g_i^{-1} and maps the P_i's bijectively to the Q_j's. This defines implicitely a function $j(i)$. The function $j(i)$ has the property that $g_i = \eta_0$ implies $g_{j(i)-1} = \eta_0^{-1} = g_i^{-1}$. We state the action of g_i^{-1} on several other points in W:

$$\begin{aligned}
P_i &\to Q_{j(i)},\\
T_{i,1} &\to Q_{j(i)+1},\\
T_{i,r} &\to T_{j(i),r-1}, \text{ if } 2 \le r \le n(v_i)-2,\\
S_{i+1,1} &\to P_{j(i)-2},\\
S_{i+1,r} &\to S_{j(i)-1,r-1}, \text{ if } 2 \le r \le n(v_i)-2,\\
T_{i+1,n(v_{i+1})-2} &\to Q_{j(i)-1},\\
Q_{i+1} &\to P_{j(i)-1},\\
P_{i+1} &\to S_{j(i)-1,n(v_i)-2},\\
Q_i &\to T_{j(i),n(v_i)-2}.
\end{aligned} \tag{60}$$

Lemma 3.2 implies immediately the first part of the following theorem:

Theorem 3.9 *There exists a finite partition $\mathcal{P} = \{I_i\}_{i=1}^{n}$ of $S^1 = \partial I\!D$ in intervals such that:*

(M1) *The Bowen-Series map f is strictly monotonic on every $I_i \in \mathcal{P}$ and the restriction f_{I_i} has a holomorphic continuation in a whole neighbourhood $V_i \subset I\!\!C$ of I_i for every i;*

(M2) *If $f(I_i) \cap I_j \neq \emptyset$, then $I_j \subset f(I_i)$;*

(M3) *f satisfies a transitivity condition:*
There exists an $R \in I\!N$ such that $I_j \subset \cup_{r=0}^{R} f^r(I_i)$ for all i,j;

(M4) *f satisfies a finiteness condition:*
if $\mathrm{int}\, I_i =]a_i, b_i[$, then $\cup_{i=1}^{\infty}\{\lim_{h\to 0+} f(a_i+h), \lim_{h\to 0-} f(b_i+h)\}$ is finite.

(M2) is also called the MARKOV PROPERTY of f. We also call f an analytical Markov map because of property (M1).

Proof: [25] We consider the partition $\mathcal{P}$ of $\partial I\!D$ induced by the set W from Lemma 3.2.
(M1) is trivial by virtue of the Definition 3.13 of f.
(M2) follows immediately by Lemma 3.2.

(M3) By Equations 60, we see that if $I_s \in \mathcal{P}$, then there exists a $p > 0$ such that $f^p(I_s)$ contains an interval of the form $A_i := [T_{i,1}, P_{i+1}[$ or $B_i := [Q_i, S_{i+1,1}[$. The restriction of f to the sets $A_i, B_i \subset [P_i, P_{i+1}[$ equals by definition g_i^{-1}.
From now on, let i be fixed. We see that $f([T_{i,1}, P_{i+1}[)$ covers all B_j with the exception of $B_{r(i)} = [Q_{j(i)-1}, S_{j(i),1}[$ and $B_{r'(i)} = [Q_{j(i)}, S_{j(i)+1,1}[$. Similarly, $f([Q_i, S_{i+1,1}[)$ covers all A_l with the exception of $A_{s(i)} = [T_{j(i)-1,1}, P_{j(i)}[$ and $A_{s'(i)} = [T_{j(i)-2,1}, P_{j(i)-1}[$.
A second application of f yields therefore: $f^2([T_{i,1}, P_{i+1}[)$ covers all A_l with at most one exception, namely $\{A_{j(j(i)-1)-2}, A_{j(j(i)-1)-1}\} \cap \{A_{j(j(i))-1}, A_{j(j(i))-2}\}$. This set contains at most one element denoted by $A_{n(i)}$. This follows from the facts (1) that $f(B_{r(i)})$ contains all A_l with exception of $A_{s(j(i)-1)}$ and $A_{s'(j(i)-1)}$ and (2) that $f(B_{r'(i)})$ contains all A_l with exception of $A_{s(j(i))}$ and $A_{s'(j(i))}$ and (3) that $j(i)$ is bijective (a permutation). Choose some A_p, $p \notin \{i, n(i)\}$, then $f^2([T_{p,1}, P_{p+1}[) \subset f^4([T_{i,1}, P_{i+1}[)$ covers all A_l with at most one exception $A_{n(p)} \neq A_{n(i)}$. This inequality follows immediately: The equality $A_{n(p)} = A_{n(i)}$ would imply $A_{n(p)} = A_{n(i)} \in \{A_{j(j(i)-1)-2}, A_{j(j(i)-1)-1}\} \cap \{A_{j(j(i))-1}, A_{j(j(i))-2}\} \cap \{A_{j(j(p)-1)-2}, A_{j(j(p)-1)-1}\} \cap \{A_{j(j(p))-1}, A_{j(j(p))-2}\}$. Since $j(i)$ is a permutation, the intersection on the right hand side is, however, empty for $i \neq p$.
Therefore we conclude that $f^2(A_i) \cup f^4(A_i)$ covers all A_l.
Furthermore, we see that $f(A_i) = [Q_{j(i)+1}, S_{j(i)-1,n(v_i)-2}[$. $\partial I\!D$ is covered by the f-images of four different A_l. Since $\mathcal{F}$ has at least four sides, the assertion (M3) of the theorem follows.
(M4) This is clear since $\mathcal{P}$ is finite. □

Remark: There is an analogous theorem for the Bowen-Series map of a Fuchsian group Γ containing parabolic elements with an in general countable covering $\mathcal{P}$. The proof of (M4) is then not so trivial. In (M3) we have to set $R = \infty$.

Theorem 3.10 *There exists an $N \in I\!N$ and a $\lambda > 1$ such that $|(f^N)'(x)| \geq \lambda > 1$ for all $x \in \partial I\!D$.*

Usually, one sets $\lambda := \inf_{x\in\partial I\!D} |(f^N)'(x)|$. We say that f^N is UNIFORMLY EXPANSIVE and f is (EVENTUALLY) EXPANSIVE in accordance with Definition 2.5. A proof of Theorem 3.10 can be found in Bowen & Series [25] or in Series [27] Theorem 5.1 or in Bowen [108] Lemma 3. Furthermore, for a special choice of the fundamental domain a simple proof of a somewhat weaker statement can be found in Adler & Flatto [74].

Remark: Theorem 3.10 is not valid for the Bowen-Series map of a Fuchsian group containing parabolic elements, i.e. not satisfying condition 3.

Properties of the boundary expansion

We consider now the symbolic dynamical system obtained with the help of the Bowen-Series map f. We denote with Σ_f^+ the space of all (one-sided infinite) boundary expansion of points in $\partial I\!D$:

$$\Sigma_f^+ := \{\xi_f : \xi \in \partial I\!D\} \subset \prod_{i=0}^{\infty} \mathcal{G} = \mathcal{G}^{I\!N}.$$

Definition 3.14 *The set $I(g_{i_0} g_{i_1} \dots g_{i_{n-1}}) := \bigcap_{j=0}^{n} f^{-j} I(g_{i_j}) = \{\eta \in \partial I\!D \mid \eta_f$ begins with the sequence $g_{i_0} g_{i_1} \dots g_{i_{n-1}}\}$ is called an n-CYLINDER.*

$I(g_{i_0} \dots g_{i_n})$ is a possibly empty interval in $\partial I\!D$ for every given word in the generators $g_{i_0} \dots g_{i_n}$.

Lemma 3.3 *There exist constants $\alpha \in]0,1[$ and $C > 0$ such that $\ell(I(g_{i_0}g_{i_1}...g_{i_n})) < C\alpha^n$ for all $n > 0$, where ℓ denotes the Euclidean length on the unit circle $\partial\mathbb{D}$.*

Proof: Choose $g_{i_0}...g_{i_n}$ such that the set

$$I(g_{i_0}g_{i_1}...g_{i_n}) = \bigcap_{j=0}^{n} f^{-j} I(g_{i_j})$$

is a non-empty interval in $\partial\mathbb{D}$. The nth power of the Bowen-Series map acts on $I(g_{i_0}...g_{i_n})$ as $f^n = g_{i_{n-1}}^{-1} g_{i_{n-2}}^{-1} ... g_{i_0}^{-1}$ and therefore $f^n I(g_{i_0}g_{i_1}...g_{i_n}) = I(g_{i_n})$. By Theorem 3.10, it follows that

$$|(f^{nN})'(x)| = \prod_{j=1}^{n} |(f^N)'(f^{(n-j)N}(x))| \geq \lambda^n,$$

for all $x \in \partial\mathbb{D}$. Setting $\lambda_n := \inf_{x\in\partial\mathbb{D}} |(f^n)'(x)|$ yields (*): $\lambda_{n+m} \geq \lambda_n\lambda_m$. On the other hand,

$$|(f^{nN})(x)| = \prod_{j=1}^{N} |(f^n)'(f^{(N-j)n}(x))| \geq \inf_{y\in\partial\mathbb{D}} |(f^n)'(y)|^N = \lambda_n^N,$$

for all $x \in \partial\mathbb{D}$. Now there are two possibilities for every $n \in \mathbb{N}$, namely either

$$\inf_{y\in\partial\mathbb{D}} |(f^n)'(y)| \geq \lambda^{n/N},$$

or

$$\inf_{y\in\partial\mathbb{D}} |(f^n)'(y)| < \lambda^{n/N}.$$

We set $c_n := \lambda^{-n/N} \inf_{y\in\partial\mathbb{D}} |(f^n)'(y)|$, for $0 \leq n < N$ and $c := \min_{0\leq n<N} c_n$. Then we have for all n with $0 \leq n < N$:

$$\lambda_n = \inf_{y\in\partial\mathbb{D}} |(f^n)'(y)| \geq c\lambda^{n/N},$$

and by the inequality (*), it follows

$$\inf_{y\in\partial\mathbb{D}} |(f^{n+kN})'(y)| = \lambda_{n+kN} \geq \lambda_n\lambda_{kN} \geq c\lambda^{n/N}\lambda_N^k = c\lambda^{(n+kN)/N}.$$

Therefore there exists a $c > 0$ such that for all $n \geq 0$:

$$\inf_{y\in\partial\mathbb{D}} |(f^n)'(y)| \geq c\lambda^{n/N}. \tag{61}$$

Hence,

$$\ell(I(g_{i_n})) = \ell(f^n I(g_{i_0}g_{i_1}...g_{i_n})) = \int_{I(g_{i_0}g_{i_1}...g_{i_n})} |(f^n)'(y)|\,dy \geq c\lambda^{n/N}\ell(I(g_{i_0}g_{i_1}...g_{i_n})),$$

and the assertion follows by $\max_{g\in G} \ell(I(g)) \geq \ell(I(g_{i_n}))$. □

Lemma 3.4 *The map $\partial\mathbb{D} \to \Sigma_f^+, \xi \mapsto \xi_f$ is bijective.*

Proof: Let $I(g_{i_1}g_{i_2}...g_{i_n})$ be defined as in Definition 3.14: Obviously, a finite word $w = g_{i_1}...g_{i_n}$ is contained in Σ_f^+ if and only if $I(w)$ is a non-empty interval in $\partial\mathbb{D}$.
The set $I(g_{i_1}g_{i_2}...g_{i_n})$ is by definition an interval in $\partial\mathbb{D}$ (possible empty) and by Lemma 3.3, it follows that $\ell(I(g_{i_1}g_{i_2}...g_{i_n})) \to 0$, for $n \to \infty$, where ℓ denotes the Euclidean length on the unit circle $\partial\mathbb{D}$. Thus at most one $\eta \in \partial\mathbb{D}$ has a given generator sequence as boundary expansion. To every given generator sequence in Σ_f^+ there exists on the other hand at least one $\eta \in \partial\mathbb{D}$ which has this generator sequence as boundary expansion. □

It is possible to characterize the sequences in Σ_f^+:

Lemma 3.5 *A finite word w occurs in the sequences in Σ_f^+ if and only if w is a shortest word and contains no right-handed half cycles.*
No sequence in Σ_f^+ terminates in an infinite chain of anticlockwise cycles.

Proof: A proof of the first assertion can be found in Series [28]. We give for it another proof which is based in part on ideas from Series [22]. We prove first the second assertion. Let $\xi \in \partial\mathbb{D}$ be a point for which ξ_f terminates with an infinite chain of anticlockwise cycles. (We have called such a chain also a right-handed pseudo half cycle.) Then there exists an m such that for $f^m(\xi) =: \eta$ the sequence η_f is an infinite chain of anticlockwise cycles. This boundary expansion corresponds to some P_{i+1}, coded as if P_{i+1} belongs to $[P_i, P_{i+1}[$. We have defined f such that P_{i+1} belongs to $[P_{i+1}, P_{i+2}[$ and therefore P_{i+1} has a boundary expansion which begins with g_{i+1} and which is an infinite chain of clockwise cycles. The second statement follows by Lemma 3.4.
Let $w = g_{i_1}g_{i_2}...g_{i_n}$ be a word. Let $I(g_{i_1}g_{i_2}...g_{i_n})$ as in Definition 3.14: Obviously, a word $w = g_{i_1}...g_{i_n}$ is contained in the sequences of Σ_f^+ if and only if $I(w)$ is a non-empty interval in $\partial\mathbb{D}$.
Let now $w = g_{i_1}...g_{i_n}$ be a word with $I(w) \neq \emptyset$.
Suppose w contains a right-handed half cycle. Without loss of generality we can assume that w begins with a right-handed half cycle. For $\xi \in I(w)$ we have then on the other hand: $\xi \in A(g_{i_1+1})$ and ξ_f begins with g_{i_1+1}. This is a contradiction and we conclude that w contains no right-handed half cycles.
We suppose that w is not a shortest word. By Theorem 3.8, the edge path K_w belonging to w intersects a geodesic in N twice. Without loss of generality we assume that K_w intersects the geodesic $C(g_{i_1})$ first between 0 and $g_{i_1}0$ and second between $g_{i_1}...g_{i_{r-1}}0$ and $g_{i_1}...g_{i_r}0$ for a *suitable choosen* $1 < r \leq n$. *Then, however,* $I(g_{i_1}) \subseteq A(g_{i_1})$ and $g_{i_1}...g_{i_{r-1}}I(g_{i_r}) \subseteq \partial\mathbb{D}\setminus A(g_{i_1})$. Furthermore,

$$\begin{aligned} I(g_{i_1}...g_{i_r}) &= \cap_{t=1}^{r} f^{-t+1} I(g_{i_t}) \\ &= I(g_{i_1}...g_{i_{r-1}}) \cap f^{-r+1} I(g_{i_r}) \\ &= I(g_{i_1}...g_{i_{r-1}}) \cap g_{i_1}...g_{i_{r-1}} I(g_{i_r}), \end{aligned}$$

where in the last step we have used that $f^{r-1}|_{I(g_{i_1}...g_{i_{r-1}})} = (g_{i_1}...g_{i_{r-1}})^{-1}$. This equation implies $I(g_{i_1}...g_{i_r}) = \emptyset$. This is a contradiction; thus w is a shortest word.
Now let w be a shortest word without right-handed half cycles. w defines as usual an element $g \in \Gamma$. We show first that w is uniquely determined by the requirement that w is a shortest word without right-handed half cycles. To this end choose a second shortest word w' without right-handed half cycles defining the same g as w. By Corollary 3.1, the two polygonal paths K_w and $K_{w'}$ defined by w and w' are neighbouring polygonal paths. We show similarly as in the

proof of Corollary 3.1 (**) that K_w and $K_{w'}$ are equal:
There exists a first vertex v_1 which K_w and $K_{w'}$ pass on different sides since otherwise they would be equal. One of the two edge paths, without loss of generality K_w, has the vertex angle π^+ at this vertex v_1 and goes clockwise around the vertex. $K_{w'}$ goes counter-clockwise around v_1 with the vertex angle π. Denote with v the next vertex at which the vertex angle pair is unequal to π/π. It is impossible that K_w goes around the vertex v with the angle π^+ since otherwise K_w would contain a long cycle or a long chain. $K_{w'}$ goes around v with an angle different from π^+ since otherwise $K_{w'}$ would contain a right-handed half cycle. At least one of the two paths goes around v with vertex angle π^-; however, this is impossible since K_w and $K_{w'}$ are neighbouring edge paths. This implies that K_w equals $K_{w'}$ and therefore $w = w'$.
It remains to show that $I(w) \neq \emptyset$ holds. We denote with $H(g_j)$ the closed half plane bounded by $C(g_j)$ and $A(g_j)$ for every $j \in \{1, ..., k\}$, cf. page 55. We now define the action of the Bowen-Series map f on the sets $H(g_j)$. First, on the sets $H(g_j) \cap H(g_{j+1})$ we define $f = g_{j+1}^{-1}$ for all $j \in \{1, ..., k\}$. For all $j \in \{1, ..., k\}$ we define f on the subset of $H(g_j)$ on which f is not defined yet (i.e. on $H(g_j)\backslash[H(g_j) \cap H(g_{j+1}) \cap H(g_j) \cap H(g_{j-1})]$) by $f = g_j^{-1}$. Now f is defined on $\mathbb{D}\backslash\mathcal{F}$. If we restrict the so defined map f to $\partial\mathbb{D}$, we reobtain the definition 3.13 of the Bowen-Series map.
For all $j \in \{1, ..., k\}$ let $B(g_j) := H(g_j)\backslash(H(g_j) \cap H(g_{j+1}))$ and for all $j_0, ..., j_m \in \{1, ..., k\}$ let $B(g_{j_0} ... g_{j_m}) := \cap_{r=0}^{m} f^{-r} B(g_{j_r})$. By definition $I(g_{j_0} ... g_{j_m}) = \partial\mathbb{D} \cap B(g_{j_0} ... g_{j_m})$.
We denote the wordlength of the word w by $|w| = n + 1$. Since the sets $B(g_{j_0} ... g_{j_m})$ are bounded by geodesic arcs of geodesics in N, the open kernels of the sets $B(g_{j_0} ... g_{j_m})$ are the union of some images of the fundamental domain $\mathcal{F}$. Since every edge path relating 0 and $g0$ contains at least $n + 1$ edges, there exists $e_0, e_1, ..., e_n \in \mathcal{G}$ such that $g\mathcal{F} \subset B(e_0 e_1 ... e_n)$. Now we have $\emptyset \neq I(e_0 ... e_n) = \partial\mathbb{D} \cap B(e_0 ... e_n)$ since $B(e_0 ... e_n)$ is non-empty. Similarly to the above reasoning it follows that $e_0 ... e_n$ is a shortest word without right-handed half cycles.
The geodesic in $\mathbb{D}$ relating the points 0 and $g0$ passes through $e_0\mathcal{F}$ and defines a geodesic edge path relating 0 and $g0$, since otherwise we would obtain a contradiction to Theorem 3.8. Therefore there exists a shortest edge path relating 0 and $f(g0) = e_0^{-1} g0$ of length n. Inductively one sees that there exists a shortest edge path relating 0 and $f^{n+1}(g0) = e_m^{-1} ... e_0^{-1} g0$ of length $n - n = 0$. Hence, $g = e_0 ... e_n$; since $e_0 ... e_n$ is a shortest word without right-handed half cycles, it follows $w = e_0 ... e_n$ and therefore $I(w) \neq 0$. This proves the theorem. □

We now prove the following main lemma:

Lemma 3.6 *There exists a topologically mixing topological Markov shift (Σ_A^+, σ_A^+) with matrix A and with finite alphabet $\mathcal{S}$ and a surjective map $\beta : \mathcal{S} \to \mathcal{G}$ such that the induced map $\bar{\beta} : \Sigma_A^+ \to \partial\mathbb{D} \simeq \Sigma_f^+$ is continuous, surjective and injective with the exception of a countable set of points, where $\bar{\beta}$ is two-to-one. Furthermore: $f \circ \bar{\beta} = \bar{\beta} \circ \sigma_A^+$.*

Proof: We consider the partition $\mathcal{P}$ from Theorem 3.9 and set $\mathcal{S} := \cup_{i=1}^{w} \{I_i\}$. Using this ingredients we want to construct a topological Markov shift (Σ_A^+, σ_A^+) (cf. Definition 2.2). We achieve this using the Markov property (M2) of f, cf. Theorem 3.9: the f-image of I_i is an exact union of certain I_j. From this information we obtain immediately the matrix $A = (a_{i,j})$. The matrix together with the alphabet $\mathcal{S}$ and the shift operator σ_A^+ constitutes a topological Markov shift (Σ_A^+, σ_A^+). It follows by the property (M3) from Theorem 3.9 and by Lemma 2.1 that (Σ_A^+, σ_A^+) is topologically mixing: with R given by Theorem 3.9 we have $A^R > 0$.
The map β is defined by $\beta : \mathcal{S} \to \mathcal{G}, \beta(I_j) := g_i$ if $I_j \subset [P_i, P_{i+1}]$. The induced map $\bar{\beta} : \Sigma_A^+ \to \Sigma_f^+$ is obviously surjective.

If $\xi \notin \overline{W} := \cup_{m=0}^{\infty} f^{-m}(W)$, then we can associate with ξ a sequence $s(\xi)$ in Σ_A^+ according to: $s(\xi) = I_{i_0} I_{i_1} I_{i_2} ...$, with $I_{i_j} \in \mathcal{S}$ and $f^n(\xi) \in I_n$, for all $n \in \mathbb{N}$. The sequence $s(\xi)$ does not depend on the way the boundary points of the intervals in $\mathcal{P}$ are assigned to the intervals in $\mathcal{P}$.
If $\xi \in \cup_{m=0}^{\infty} f^{-m}(W)$, then there exists a smallest $m_0 \geq 0$ such that $f^{m_0}(\xi) \in W$, i.e. $f^{m_0}(\xi)$ is a boundary point of some interval in $\mathcal{P}$. There are finitely many possibilities to assign the endpoints of the intervals in $\mathcal{P}$ to the intervals in $\mathcal{P}$. The symbol in $\mathcal{S}$ corresponding to $f^{m_0}(\xi)$ can be arbitrarily choosen as one of the two labels of the intervals adjacent to $f^{m_0}(\xi)$. The assignation of labels for the remaining boundary points of the intervals in $\mathcal{P}$ has to be compatible with the matrix $A = (a_{i,j})$, i.e. if we assign a boundary point ξ_b of the interval I_h to the interval I_h then we have to assign $g_h^{-1}\xi_b$ to $f(I_h) = g_h^{-1} I_h$, where we have set $g_h = g_i$ if $I_h \subset [P_i, P_{i+1}]$.
From this condition and Lemma 3.2 it follows that the assignation of the boundary point $f^{m_0}(\xi)$ to one of the two possible intervals in $\mathcal{P}$ already determines the whole symbol sequence in Σ_A^+ associated with ξ. Thus there are only two different symbol sequences in Σ_A^+ which can be unambigiously assigned to ξ, which we will denote with $s(\xi)$ and $\tilde{s}(\xi)$.
Summarizing, we obtain for every point $\xi \in \partial\mathbb{D}\backslash\overline{W}$ the sequence $s(\xi)$ in Σ_A^+ and for $\xi \in \overline{W}$ we obtain the two sequences $s(\xi), \tilde{s}(\xi)$ in Σ_A^+.
On the other hand, it is clear by construction of Σ_A^+ that every sequence in Σ_A^+ is also contained in $\{s(\xi) \mid \xi \in \partial\mathbb{D}\} \cup \{\tilde{s}(\xi) \mid \xi \in \overline{W}\}$.
Similarly to Lemma 3.4 it can be proven: $\xi_1 \neq \xi_2 \Rightarrow s(\xi_1) \neq s(\xi_2)$, and whenever well-defined $\tilde{s}(\xi_1) \neq \tilde{s}(\xi_2); \tilde{s}(\xi_1) \neq s(\xi_2); s(\xi_1) \neq \tilde{s}(\xi_2)$. Analogously to Lemma 3.4 one defines $J(I_{i_1} I_{i_2} ... I_{i_n}) := \{\eta \in \partial\mathbb{D}\backslash\overline{W} \mid s(\eta) \text{ starts with } I_{i_1} ... I_{i_n}\} \cup \{\eta \in \overline{W} \mid s(\eta) \text{ or } \tilde{s}(\eta) \text{ starts with } I_{i_1} ... I_{i_n}\}$. Notice that every $J(I_{i_1} I_{i_2} ... I_{i_n})$ is contained completely in the closure of some $I(g_1 g_2 ... g_n)$ (cf. Definition 3.14). The Euclidean length of J tends to 0 for $n \to \infty$ by Lemma 3.3. For every given sequence s_0 in Σ_A^+ there exists at most one $\xi \in \partial\mathbb{D}$ with $s(\xi) = s_0$.
On the other hand by Lemma 3.4: $\xi_1 \neq \xi_2 \Leftarrow s(\xi_1) \neq s(\xi_2)$, for $\xi_1, \xi_2 \notin \overline{W}$.
Therefore we have shown that $\bar{\beta} : \Sigma_A^+ \to \partial\mathbb{D}$ is a bijective map (with the exception of only countably many points) which maps every sequence $\omega^+ \in \Sigma_A^+$ to the point $\xi \in \partial\mathbb{D}$ which satisfies $s(\xi) = \omega^+$ or $\tilde{s}(\xi) = \omega^+$. That $\bar{\beta}$ conjugates the shift operator σ_A^+ of Σ_A^+ with f is clear by the definition of f and $s(\xi)$.
It remains to show that $\bar{\beta}$ is continuous. We remark first that we endow $\partial\mathbb{D} \simeq \Sigma_f^+$ with the topology induced by the standard topology of $\mathbb{C}$ and that we endow Σ_A^+ with the topology defined by the metrics $d_\rho, 0 < \rho < 1$ in Equation 108. Σ_A^+ is a compact metric space with this topology. Let $\{\omega_j^+\}_{j \in \mathbb{N}}$ be a sequence with $\omega_j^+ \in \Sigma_A^+$ for all $j \in \mathbb{N}$, which converges to an element $\omega^+ \in \Sigma_A^+$. For all $n \in \mathbb{N}$ there exists a J_n such that for all $j \geq J_n$ the first n symbols of ω_j^+ coincide with the first n symbols of ω^+, i.e. $\bar{\beta}(\omega_j^+) \in I(\omega_0 ... \omega_{n-1})$ for $j > J_n$. In the proof of Lemma 3.4 we have seen that $\ell(I(\omega_1 \omega_2 ... \omega_n)) \to 0$ for $n \to \infty$. Hence it follows that also the sequence $\{\bar{\beta}(\omega_j^+)\}_{j \in \mathbb{N}}$ converges in $\partial\mathbb{D}$. Thus $\bar{\beta}$ is continuous. □

Remark: Lemmata 3.4 and 3.5 are also valid for Fuchsian groups containing parabolic elements. However, Lemma 3.6 is not valid for Fuchsian groups containing parabolic elements. If we allow the alphabet in Lemma 3.6 to be countably infinite then the modified Lemma 3.6 is valid also for Fuchsian groups containing parabolic elements.

The points at which $\bar{\beta}$ is not one-to-one are exactly the points in $\overline{W}$, i.e. exactly the points whose boundary expansion terminates eventually in an infinite chain.
Since $\bar{\beta}$ is bijective on $\partial\mathbb{D}\backslash\overline{W}$, it follows by Lemma 3.6 that the characterization of sequences in Σ_f^+, as stated in Lemma 3.5, is complete.

The Bowen-Series map depends on the considered Fuchsian group Γ and the fixed fundamental domain $\mathcal{F}$. Therefore we do not expect (and this is indeed not the case) that all geodesics in $\mathbb{D}$ can be mapped one-to-one to a symbolic dynamical system with the help of boundary expansions. For instance, geodesics whose initial and final points lie in the interval $[Q_i, P_{i+1}[$ cause trouble; composition of the single boundary expansions for initial and final points yields in general no reduced sequence. We now want to study which geodesics can advantageously be described by boundary expansions.

We have above made the convention that we use the Bowen-Series map f to determine the boundary expansion of the endpoint $\eta := \gamma_\infty$ of a geodesic γ in $\mathbb{D}$ and that we use the backward Bowen-Series map $\bar{f}$ to determine the backward boundary expansion of the initial point $\xi := \gamma_{-\infty}$. We associate to the endpoint γ_∞ the sequence $(\gamma_\infty)_f = \eta_0\eta_1\eta_2...$ and to the initial point $\gamma_{-\infty}$ the sequence $(\gamma_{-\infty})_{\bar{f}} = \xi_0\xi_{-1}\xi_{-2}....$

Above we have already noted that the map $\bar{f}$ has essentially the same properties as f and that all theorems for f in this section are also true for $\bar{f}$, provided one replaces everywhere 'right-handed' by 'left-handed' and vice versa. For instance, in Lemma 3.5 all right-handed cycles and right-handed chains have to be replaced by left-handed cycles and left-handed chains to obtain the correct statement for $\bar{f}$. Especially, it is possible to define the space $\Sigma_{\bar{f}}^+$ which satisfies a lemma analogous to Lemma 3.6 and we have: A finite sequence $e_i...e_{i+r}$ resp. an infinite sequence $e_ie_{i+1}...$ is contained in some element in Σ_f^+ if and only if the 'inverse' sequence $e_{i+r}^{-1}...e_i^{-1}$ resp. $...e_{i+1}^{-1}e_i^{-1}$ is contained in some element in $\Sigma_{\bar{f}}^+$.

We now put together the two sequences $(\gamma_\infty)_f, (\gamma_{-\infty})_{\bar{f}}$ associated with the final and initial points of our geodesic and consider the composited sequence defined by:

$$\gamma_{-\infty} * \gamma_\infty := \xi * \eta := ...\xi_{-2}^{-1}\xi_{-1}^{-1}\xi_0^{-1}\eta_0\eta_1\eta_2.... \tag{62}$$

We denote the space of all sequences constructed in this way of some sequence in $\Sigma_{\bar{f}}^+$ and some sequence in Σ_f^+ by $\mathcal{E}$. The quantity $\gamma_{-\infty} * \gamma_\infty$ is also called the BOUNDARY EXPANSION of γ.

Furthermore, we denote the geodesic γ which joins $\xi = \gamma_{-\infty}$ and $\eta = \gamma_\infty$ with $\gamma(\xi, \eta)$. Obviously, we have defined in this way a by Lemma 3.4 bijective transformation which maps the set of all geodesics in $\mathbb{D}$ to $\mathcal{E}$ by $\gamma(\xi, \eta) \mapsto \xi * \eta$.

The edge path of the boundary expansion relating the vertices $..., \xi_0\xi_{-1}0, \xi_00, 0, \eta_00, \eta_0\eta_10, ...$ will be denoted by $E(\xi * \eta)$. We continue to denote the geodesic edge path of γ by $E(\gamma)$.

We consider the one-sided topological Markov shift (Σ_A^+, σ_A^+) with matrix (A_{ij}). It is possible to construct in a canonical way a two-sided extension (Σ_A, σ_A) of (Σ_A^+, σ_A^+). (Σ_A, σ_A) is the two-sided topological Markov shift with the same alphabet $\mathcal{S}$ as Σ_A^+ and the same matrix (A_{ij}).

As in Lemma 3.6, the map $\beta : \mathcal{S} \to \mathcal{G}$ induces a map $\tilde{\beta} : \Sigma_A \to \tilde{\beta}(\Sigma_A) \subset \mathcal{E}$. By Lemma 3.4, we can *bijectively* associate to every element in $\tilde{\beta}(\Sigma_A) \subset \mathcal{E}$ two points in $\partial\mathbb{D}$ and therefore a geodesic γ which joins these two points. The space of all geodesics obtained in this way is denoted by $\mathcal{A}$. We now have an analogous statement to Lemma 3.6:

Corollary 3.2 *There exists a topologically mixing topological Markov shift (Σ_A, σ_A) with finite alphabet $\mathcal{S}$ and a surjective map $\beta : \mathcal{S} \to \mathcal{G}$ such that the induced map $\tilde{\beta} : \Sigma_A \to \mathcal{A}$ is continuous, surjective and injective with the exception of a countable set of points, where $\tilde{\beta}$ is finite-to-one (at most four-to-one). Furthermore, $f \circ \tilde{\beta} = \tilde{\beta} \circ \sigma_A$.*

Here the Bowen-Series map f on $\mathcal{A}$ is defined by:

$$f(\gamma) = g_i^{-1}\gamma \text{ if } \gamma_\infty \in [P_i, P_{i+1}[. \tag{63}$$

Proof: We prove only the last equation in Corollary 3.2. All other statements in the corollary follow immediately from Lemma 3.6 and the remarks above. Especially, the continuity of $\tilde{\beta}$ can be proven similarly as the continuity of $\bar{\beta}$. That (Σ_A, σ_A) is topologically mixing follows by condition (M3) in Theorem 3.9 together with Lemma 2.1: the matrix A satisfies $A^R > 0$, where R is the constant from Theorem 3.9 .

It is clear that the map $\gamma \mapsto \tilde{\beta}(\sigma_A(\tilde{\beta}^{-1}(\gamma)))$ is the shift operator on the space $\tilde{\beta}(\Sigma_A) \subset \mathcal{E}$. This map maps the sequence $...\xi_{-2}^{-1}\xi_{-1}^{-1}\xi_0^{-1} \mid \eta_0\eta_1\eta_2...$ associated to some geodesic $\gamma \in \mathcal{A}$ to the sequence $...\xi_{-2}^{-1}\xi_{-1}^{-1}\xi_0^{-1}\eta_0 \mid \eta_1\eta_2...$ associated to the geodesic $\tilde{\beta}(\sigma_A(\tilde{\beta}^{-1}(\gamma)))$. Next we want to show that our generalized Bowen-Series map f on $\mathcal{A}$ coincides with $\tilde{\beta} \circ \sigma_A \circ \tilde{\beta}^{-1}$. We only need to show that the boundary expansion of the endpoint of $f(\gamma)$ equals $\eta_1\eta_2...$ and that the backward boundary expansion of the initial point of $f(\gamma)$ equals $\eta_0^{-1}\xi_0\xi_1....$ The first assertion follows immediately from the definition of the boundary expansion.

Now the following statement is true: An element $\tilde{\beta}(\phi) \in \tilde{\beta}(\Sigma_A) \subset \mathcal{E}$ contains no long cycles and no long chains and no right-handed (possibly pseudo) half cycles and furthermore $\xi_0 \neq \eta_0$ holds. An appropiate shift $\tilde{\beta}(\sigma_A^m\phi), m \in \mathbb{Z}$, would otherwise correspond to some geodesic whose endpoint (resp. initial point) has a boundary expansion (resp. backward boundary expansion) which contains a long cycle, a long chain or a right-handed (possibly pseudo) half cycle (resp. left-handed half cycle) or which is not reduced. However, such boundary expansions are excluded, as we have seen above.

Choose i such that $\eta_0 = g_i$ holds. With the notations from Figure 3.1 and Equation 60 we have then $\xi \notin]P_i, T_{i+1,n(v_{i+1})-2}]$; notice that $\xi \in]Q_i, Q_{i+1}]$ would imply $\xi_0 = \eta_0$, and $\xi \in]P_i, Q_i]$ would imply that $\xi * \eta$ contains a long cycle or a long chain (containing $\xi_0^{-1}\eta_0$), and finally $\xi \in]Q_{i+1}, T_{i+1,n(v_{i+1})-2}]$ would imply that $\xi * \eta$ contains a right-handed (if $\xi = T_{i+1,n(v_{i+1})-2}$ pseudo) half cycle. By $\xi \notin]P_i, T_{i+1,n(v_{i+1})-2}]$, it follows (recall the action of $g_i^{-1} = \eta_0^{-1}$ on some points in W given by the Equations 60): $\eta_0^{-1}\xi \in]Q_{j(i)-1}, Q_{j(i)}]$ and $g_{j(i)-1} = \eta_0^{-1}$.

Therefore the backward boundary expansion of $\eta_0^{-1}\xi$ begins with η_0^{-1}. In summary, we conclude that the backward boundary expansion of the initial point of $f(\gamma)$ equals $(\eta_0^{-1}\xi)_{\bar{f}} = \eta_0^{-1}\xi_0\xi_1....$

This proves the corollary. □

We asked above which geodesics can advantageously be described by the boundary expansions of their initial and endpoints. From Corollary 3.2 we see now that e.g. the set $\mathcal{A}$ is such a set of geodesics which can advantageously be described by the boundary expansions of their initial and final points on $\partial\mathbb{D}$. Namely, the action of the Bowen-Series map on $\mathcal{A}$ is topologically conjugate *to the action of a shift operator on a topological Markov shift (with the exception of countably* many points).

Therefore we now want to study whether $\mathcal{A}$ can be characterized by conditions like those in Lemma 3.5. In the proof of Corollary 3.2 we have already found some necessary conditions on geodesics in $\mathcal{A}$. We will immediately see that these conditions in the proof of Corollary 3.2 are essentially already sufficient.

The space Σ_A is σ_A-invariant by Definition 2.1 : $\Sigma_A = \sigma_A\Sigma_A$. Hence, it follows that also $\tilde{\beta}(\Sigma_A) = \tilde{\beta}(\sigma_A\Sigma_A)$.

Since $f \circ \tilde{\beta} = \tilde{\beta} \circ \sigma_A$, we have:

$$f(\tilde{\beta}(\Sigma_A)) = \tilde{\beta}(\Sigma_A).$$

The conditions on $\mathcal{A}$ have to be invariant under f; given a geodesic $\gamma = \gamma(\xi, \eta) \in \mathcal{A}$, the conditions on $\xi * \eta$ have to be independent from the choice of the zero-element in $\xi * \eta$. η_l is a shortest sequence which contains no right-handed half cycles and no right-handed pseudo half cycles. The same is true for $(\xi_f)^{-1} = ...\xi_{-1}^{-1}\xi_0^{-1}$. Therefore it is clear what the conditions on $\mathcal{A}$ are:

$$\mathcal{A} = \{\gamma = \gamma(\xi, \eta) \mid \xi * \eta \text{ is shortest word without right-handed half cycles not beginning or ending with a right-handed pseudo half cycle.}\} \quad (64)$$

$\mathcal{A}$ is completely characterized by the conditions in Equation 64 since the characterization in Lemma 3.5 was complete.

3.5 The relation between the Artin and the Koebe-Morse codings

In this section we study the relation between $\mathcal{A}$ provided with the Bowen-Series map f and the set of geodesics intersecting the fixed fundamental domain $\mathcal{F}$

$$\mathcal{R} := \{\gamma \mid 0 \in E(\gamma)\}$$

provided with the return map τ. Recall that $E(\gamma)$ denotes the geodesic edge path of γ and that we have made the convention to replace every $\gamma \in N$ by a curve deformed infinitesimally to the right of γ and to denote the cutting sequence of this curve by $E(\gamma)$, cf. Figure 3.6. The definition of $\mathcal{R}$ implies that the geodesic with initial point Q_{i+1} and final point P_i (where $0 < i \leq k$ arbitrary) is contained in $\mathcal{R}$; on the other hand, the geodesic with opposite orientation relating P_i and Q_{i+1} is not contained in $\mathcal{R}$.
At the end of the last section we have seen (cf. Corollary 3.2) that the Bowen-Series map on $\mathcal{A}$ is conjugate to a shift operator on some topological Markov shift. The return map $\tau : \mathcal{R} \to \mathcal{R}$ is defined as follows: τ maps a geodesic $\gamma \in \mathcal{R}$ to those equivalent geodesic $\tau(\gamma)$ which enters $\mathcal{F}$ at the point which is equivalent to the point where γ leaves $\mathcal{F}$. We denote with $g(\gamma)$ the interior label of the side through which γ enters $\mathcal{F}$; then the sequence $...g(\gamma)g(\tau(\gamma))g(\tau^2(\gamma))...$ is exactly the cutting sequence of γ.
The simplest relation between $\mathcal{A}$ and $\mathcal{R}$ would of course be $\mathcal{A} = \mathcal{R}$. Indeed, it is possible to show that $\mathcal{A} = \mathcal{R}$ if Γ is, for instance, the freely generated group of two generators [22, 28]. The freely generated group of two generators does not satisfy the conditions 0-3 stated in Section 3.2. In our case, however:

Theorem 3.11 $\mathcal{A} \neq \mathcal{R}$.

Proof: We construct a geodesic contained in $\mathcal{R}$ but not contained in $\mathcal{A}$. In the course of the proof we will make use of Theorem 3.12 which we will state below. The proof of Theorem 3.12 is independent of Theorem 3.11.
We consider again the set W introduced in Lemma 3.2 and label the points in W as in Equations 60. By Lemma 3.1, it follows that the intervals $]T_{i,1}, S_{i+1,1}[$ and $]Q_{i+1}, T_{i+1,n(i,i+1)-2}[$ are non-empty. We choose $\gamma_\infty \in]T_{i,1}, S_{i+1,1}[$ and $\gamma_{-\infty} = T_{i+1,n(i,i+1)-2}$ and consider the oriented geodesic γ relating $\gamma_{-\infty}$ and γ_∞. This geodesic γ intersects $\mathcal{F}$ and goes counter-clockwise around the vertex v_{i+1}. Now we use Theorem 3.12 and obtain a contradiction to the assumption $\gamma \in \mathcal{A}$. The proof remains obviously true if one considers a geodesic γ with $\gamma_\infty \in]T_{i,1}, S_{i+1,1}[$ and $\gamma_{-\infty} = T_{i+1,n(i,i+1)-2} - \epsilon$ with sufficiently small $\epsilon > 0$. □

The proof implies that the geodesics which pass near enough to some vertex $\mathcal{F}$ have a boundary expansion which corresponds to a cutting sequence and an edge path which goes round the considered vertex on the other 'wrong' side than the geodesic itself. We will discuss this point in more detail below. First of all, we collect some definitions and lemmata used in the sequel:

Definition 3.15 *Let $N(v) \subset N$ be the set of all geodesics in N which pass through the vertex v. We say that a geodesic $\gamma(\xi, \eta)$* PASSES NEAR SOME VERTEX *$v \in N$ if the two endpoints ξ, η lie in opposite sectors of the net defined by the geodesics in $N(v)$. (We consider these sectors to be closed.)*
If $\gamma \in \mathcal{R}$, if no endpoint of γ lies in the sector which contains $\mathcal{F}$, and if γ passes near some vertex $v \in \mathcal{F}$, then we say that γ CUTS OFF THE VERTEX *v* ON *$\mathcal{F}$.*

We will use in the following the abbreviation

$$\mathcal{R}\Delta\mathcal{A} := \mathcal{R} - (\mathcal{A} \cap \mathcal{R}) \bigcup \mathcal{A} - (\mathcal{R} \cap \mathcal{A}).$$

Lemma 3.7 *If $\gamma(\xi, \eta) \in \mathcal{R}$ and if $\xi * \eta$ is not a shortest word, then $\gamma \in N$.*

Proof: [28] Suppose first that $\xi_0 = \eta_0 = g_i \in \Gamma$; i.e. that $\xi * \eta$ is not reduced. Then by the definition of the boundary expansion, $\eta \in [P_i, P_{i+1}[$ and $\xi \in]Q_i, Q_{i+1}]$. Especially, γ lies in the half plane bounded by $C(g_i)$ and not containing $\mathcal{F}$. Since γ intersects the fundamental domain $\mathcal{F}$, γ coincides with $C(g_i) \in N$.
Suppose that $\xi * \eta$ is reduced and contains a long cycle or a long chain. This long cycle or this long chain contains $\xi_0^{-1}\eta_0$ since the boundary expansions of η and ξ contain no long cycles or long chains by Lemma 3.5. Therefore the sides $C(\xi_0)$ and $C(\eta_0)$ of $\mathcal{F}$ either intersect in some vertex $v(\xi, \eta)$ or are separated by exactly one side $s(\xi, \eta)$ of $\mathcal{F}$. The edge path $E(\xi * \eta)$ belonging to $\xi * \eta$ intersects some geodesic $C \in N$ twice since $\xi * \eta$ contains a long cycle or long chain. Furthermore, C contains the vertex $v(\xi, \eta)$ or the side $s(\xi, \eta)$. Since the boundary expansions of η and ξ contain no long cycle or long chains, $E(\xi * \eta)$ intersects C exactly twice. Hence γ lies in the half plane bounded by C, which does not contain $\mathcal{F}$. Therefore γ coincides with C. □

The following lemma is in a sense the converse of the previous lemma:

Lemma 3.8 *Let $N(\mathcal{F})$ denote the set of all geodesics in N which have non-empty intersection with $\mathcal{F}$. Let $\gamma \in N(\mathcal{F})$, then:*

$$0 \in E(\gamma) \Longleftrightarrow \gamma \in \mathcal{R} \Longleftrightarrow \gamma \notin \mathcal{A}.$$

Proof: $N(\mathcal{F})$ is the set of all geodesics in N which either meet $\mathcal{F}$ exactly at some vertex $v(\gamma)$ or are the geodesic completion of some side $s(\gamma)$ of $\mathcal{F}$. Recall that for $\gamma \in N(\mathcal{F})$ the statement $0 \in E(\gamma)$ is true if and only if $\mathcal{F}$ lies on the right of γ. We assume first that γ is the completion of one side $s(\gamma)$ of $\mathcal{F}$. In this case there exists an i $(0 < i \leq k)$ such that either $\gamma_{-\infty} = P_i$ and $\gamma_\infty = Q_{i+1}$ or $\gamma_\infty = P_i$ and $\gamma_{-\infty} = Q_{i+1}$ holds.
In the first case $0 \notin E(\gamma)$. If we compute the boundary expansion of Q_{i+1} with the Bowen-Series map f and the backward boundary expansion of P_i with the backward Bowen-Series map $\bar{f}$ and build $P_i * Q_{i+1}$, then we see that $E(P_i * Q_{i+1})$ is a (two-sided) infinite chain of consecutive left-handed cycles. $E(P_i * Q_{i+1})$ passes through all fundamental domains adjacent to γ on the left. Thus: $\gamma \in \mathcal{A}$.
In the second case the same reasoning yields: $0 \in E(\gamma)$. On the other hand, $\gamma \notin \mathcal{A}$. If we composite the boundary expansion of P_i with the backward boundary expansion of Q_{i+1}, then we obtain $E(Q_{i+1} * P_i)$ and we see that $E(Q_{i+1} * P_i)$ begins and ends with a right-handed pseudo half cycle and furthermore that $(P_i)_f$ and $(Q_{i+1})_f$ begin with the same symbol, i.e. $E(Q_{i+1} * P_i)$ is not reduced. Thus: $\gamma \notin \mathcal{A}$.
If γ is not the geodesic completion of some side of $\mathcal{F}$, then there are suitable i, j, with $0 < i \leq$

$k, 0 < j \leq n(v_i) - 2$, such that either $\gamma_\infty = S_{i,j}$ and $\gamma_{-\infty} = T_{i,n(v_i)-1-j}$ or vice versa. The notations have been defined just before Equations 60. The above argument can again applied and we find in the first case that $0 \in E(\gamma)$ and $\gamma \notin \mathcal{A}$ and in the second case that $0 \notin E(\gamma)$ and $\gamma \in \mathcal{A}$. (Remark: In the first case the sequence $E(T_{i,n(v_i)-1-j} * S_{i,j})$ begins and ends with a right-handed pseudo half cycle and contains a long cycle round the vertex $v(\gamma)$.) □

For $\gamma \in N - N(\mathcal{F})$ it is easy to construct examples for which the assertion of Lemma 3.8 is not valid.

Lemma 3.9 *Let $\gamma(\xi,\eta) \in \mathcal{R}\triangle\mathcal{A}$ and if $\gamma(\xi,\eta) \in \mathcal{R}$, then let $\xi * \eta$ be a shortest word. Then the subsequence $\xi_0^{-1}\eta_0$ lies in a cycle or in a chain and γ passes near the vertex $v = v(\xi,\eta)$.*

Here $v(\xi,\eta) \in \mathcal{F}$ denotes as in the proof of Lemma 3.7 the vertex at which the geodesics $C(\xi_0)$ and $C(\eta_0)$ meet.

Proof: [28] Let first $\gamma(\xi,\eta) \in \mathcal{A} - (\mathcal{R} \cap \mathcal{A})$. Then $E(\xi * \eta)$ is a shortest edge path and by Corollary 3.1 (*) $E(\xi * \eta)$ and $E(\gamma)$ are neighbouring. Since $0 \in E(\xi * \eta)$ and since $\gamma \notin \mathcal{R}$, there exists a vertex $v \in \mathcal{F}$ such that $\xi_0^{-1}\eta_0$ lies in a cycle or in a chain round v. Since $E(\gamma)$ is a geodesic edge path, $E(\gamma)$ intersects every geodesic in $N(v)$ at most once. Similarly also $E(\xi * \eta)$ intersects every geodesic in $N(v)$ at most once, since $E(\xi * \eta)$ is a shortest edge path (cf. Theorem 3.8). Since the two edge paths have coincident endpoints on $\partial\mathbb{D}$, they both intersect actually every geodesic in $N(v)$ exactly once (possibly on $\partial\mathbb{D}$), i.e. γ passes near v.
Let now $\gamma(\xi,\eta) \in \mathcal{R} - (\mathcal{A} \cap \mathcal{R})$ and let $\xi * \eta$ be a shortest word. By Lemma 3.5, $\xi * \eta$ contains a right-handed (possibly pseudo) half cycle round some vertex $v \in \mathcal{F}$ which contains $\xi_0^{-1}\eta_0$. The edge path $E(\xi * \eta)$ intersects all sides in $N(v)$ at most once (Theorem 3.8) and since $E(\xi * \eta)$ contains a half cycle round v, also exactly once. I.e. γ passes near the vertex v. □

From the proof of Lemma 3.8 it follows that the conditions of Lemma 3.9 are exactly fulfilled for $\gamma \in [\mathcal{A} - (\mathcal{R} \cap \mathcal{A})] \cup [\mathcal{R} - (\mathcal{A} \cap \mathcal{R}) - (N(\mathcal{F}) \cap \mathcal{R})]$. For $\gamma \in N(\mathcal{F}) \cap \mathcal{R}$ Lemma 3.9 can not be applied since $E(\gamma_{-\infty} * \gamma_\infty)$ is not a shortest edge path.

Theorem 3.12 *Suppose that $\gamma(\xi,\eta)$ passes near the vertex $v = v(\xi,\eta)$ and that $\xi_0^{-1}\eta_0$ lies in a cycle or a chain. Then:*

$$\gamma(\xi,\eta) \in \mathcal{A} \implies (\gamma(\xi,\eta) \in \mathcal{R} \Leftrightarrow \gamma(\xi,\eta) \text{ goes clockwise around } v),$$
$$\gamma(\xi,\eta) \notin \mathcal{A} \implies (\gamma(\xi,\eta) \in \mathcal{R} \Leftrightarrow \gamma(\xi,\eta) \text{ goes counter-clockwise around } v).$$

Proof: [28] We assume at first that $E(\xi * \eta)$ is a shortest edge path. Those sectors in the net on $\partial\mathbb{D}$ defined by $N(v)$ which contain ξ and η are bounded by two geodesics in $N(v)$, say $C_1, C_2 \in N(v)$. $E(\xi * \eta)$ intersects C_1 and C_2 at the points $S_1, S_2 \in \mathbb{D} \cup \partial\mathbb{D}$ respectively. If S_1 or S_2 lie in $\partial\mathbb{D}$, then $E(\xi * \eta)$ begins or ends with a pseudo half cycle, which has the same orientation as the cycle which contains $\xi_0^{-1}\eta_0$. If $S_1, S_2 \notin \partial\mathbb{D}$, then we denote by v_1 resp. v_2 those two vertices in N which (i) lie between $v(\xi,\eta)$ and S_1 resp. S_2 on C_1 resp. C_2 and which (ii) are nearest to the points S_1 resp. S_2. If $v_1 = v_2 = v(\xi,\eta)$, then $E(\xi * \eta)$ contains a half cycle which contains $\xi_0^{-1}\eta_0$. If v_1 resp. $v_2 \neq v(\xi,\eta)$, then $E(\xi * \eta)$ contains a half cycle at v_1 resp. v_2, which has the same orientation as the cycle which contains $\xi_0^{-1}\eta_0$.
Thus we have shown that $E(\xi * \eta)$ contains some (possibly pseudo) half cycle, which has the same orientation as the cycle which contains $\xi_0^{-1}\eta_0$.
Now, $\gamma(\xi,\eta) \in \mathcal{A}$ holds if and only if this cycle is left-handed. On the other hand, $\gamma(\xi,\eta) \in \mathcal{R}$ holds if and only if $E(\xi * \eta)$ and $E(\gamma(\xi,\eta))$ go around the vertex $v(\xi,\eta)$ with the same orientation.

(This follows from $0 \in E(\xi * \eta)$ and from the fact that $E(\gamma(\xi,\eta))$ and $E(\xi * \eta)$ are neighbouring edge paths in the sense of Corollary 3.1). The assertion follows immediately.
Let $E(\xi * \eta)$ be an edge path which is not shortest, then $\gamma(\xi,\eta) \notin \mathcal{A}$. By assumption, $\xi_0 \neq \eta_0$ and therefore $\xi * \eta$ is reduced and contains a long cycle or a long chain.
If $\gamma(\xi,\eta) \in \mathcal{R}$, then by Lemma 3.7 it follows that $\gamma(\xi,\eta) \in N$. In this case $\gamma(\xi,\eta)$ goes by definition counter-clockwise round the vertex $v(\xi,\eta)$.
If $\gamma(\xi,\eta) \notin \mathcal{R}$, then we can as in the proof of Lemma 3.7 deduce that $\gamma(\xi,\eta)$ lies in a half plane which is bounded by some geodesic completion $C \in N(\mathcal{F})$ through $v(\xi,\eta)$ and does not contain $\mathcal{F}$. Since $\gamma(\xi,\eta)$ passes near the vertex $v(\xi,\eta)$, at least one of the two endpoints of $\gamma(\xi,\eta)$ coincides with one of the endpoints of C. In the following we assume that $\gamma_\infty = \eta$ coincides with one of the endpoints of C. The other case that $\xi = \gamma_{-\infty}$ coincides with one of the endpoints of C can be treated similarly.
By Lemma 3.8, $\gamma(\xi,\eta)$ does not equal C; $\gamma(\xi,\eta) = C$ implies $C \notin \mathcal{A}$ and $C \notin \mathcal{R}$.
$E(\xi * \eta)$ and $E(\gamma(\xi,\eta))$ go round $v(\xi,\eta)$ on different sides. Since $E(\xi * \eta)$ contains a long cycle or a long chain, $E(\xi * \eta)$ intersects the geodesic C twice. We denote with S the point at which $E(\xi * \eta)$ intersects the geodesic C after going round the vertex $v(\xi,\eta)$ and with w the next vertex *on N lying between $\eta = \gamma_\infty$ and S on C. Then $E(\xi * \eta)$ contains a half cycle round w which* has the same orientation as the cycle in $\gamma(\xi,\eta)$ round $v(\xi,\eta)$. The cycle round w is completely contained in the one-sided infinite edge path $E(\xi)$ and is therefore a left-handed cycle. Therefore $\gamma(\xi,\eta)$ goes clockwise round the vertex $v(\xi,\eta)$. □

Corollary 3.3 *Suppose that $\gamma(\xi,\eta)$ passes near the vertex $v = v(\xi,\eta)$ and that $\xi_0^{-1}\eta_0$ lies in a cycle or a chain. Then:*

$$\gamma(\xi,\eta) \in \mathcal{R} \implies (\gamma(\xi,\eta) \in \mathcal{A} \Leftrightarrow \gamma(\xi,\eta) \text{ goes clockwise around } v),$$
$$\gamma(\xi,\eta) \notin \mathcal{R} \implies (\gamma(\xi,\eta) \in \mathcal{A} \Leftrightarrow \gamma(\xi,\eta) \text{ goes counter-clockwise around } v).$$

The assertion of the corollary follows at once from Theorem 3.12.

Lemma 3.10 *Let $\gamma(\xi,\eta) \in \mathcal{R}$ and $\eta_f = \eta_0\eta_1\ldots$ and let the cutting sequence of $\gamma(\xi,\eta)$ be $e_0e_1e_2\ldots$ beginning at the side where $\gamma(\xi,\eta)$ leaves the fundamental domain $\mathcal{F}$. Then either $\eta_0 = e_0$ or the sequence $e_0e_1\ldots$ begins with a right-handed chain which ends with a right-handed (possibly pseudo) half cycle, in which case $\gamma(\xi,\eta)$ passes near some vertex of $\mathcal{F}$ and η_0 is the first symbol in the half cycle complementary to $e_0e_1\ldots$.*

In the simplest case the right-handed chain in the theorem consists only of the half cycle with which it ends.
Proof: (see Figure 3.1) Choose i such that $e_0 = g_i$, then $\eta \in [P_i, Q_{i+1}]$ and $\eta_0 = e_0$ unless $\eta \in [P_{i+1}, Q_{i+1}]$. In the second case (i.e. $\eta_0 \neq e_0$): $\eta_0 = g_{i+1}$ and $\gamma(\xi,\eta)$ passes near the vertex v_{i+1} with right-handed orientation (i.e. counter-clockwise). Therefore it is clear that $e_0e_1\ldots$ starts with a right-handed chain. Since $E(\gamma)$ and $E(\eta)$ are neighbouring edge paths by Corollary 3.1, the right-handed chain runs along $C(\eta_0) = C(g_{i+1})$. ($E(\eta)$ denotes the one-sided infinite edge path belonging to η_f and is a shortest edge path by Lemma 3.5.)
The right-handed chain ends at the vertex at which $\gamma(\xi,\eta)$ intersects the geodesic $C(\eta_0) \in N$. Since $E(\gamma(\xi,\eta))$ is a shortest edge path, the right-handed chain ends with a half cycle. Only if $\eta = P_{i+1}$ holds, $E(\gamma(\xi,\eta))$ does not intersect the geodesic $C(\eta_0)$ and the right-handed chain ends with a right-handed pseudo half cycle. The remaining assertions of the lemma are clear. □

The map $\longrightarrow$

Before we define the conjugacy T between $(\mathcal{R},\tau)$ and $(\mathcal{A},f)$, we have to construct a map which maps every pair of shortest oriented polygonal paths $P=\{\mathcal{F}_i\}_{i=-\infty}^{\infty}, P'=\{\mathcal{F}'_i\}_{i=-\infty}^{\infty}$ with coincident endpoints on $\partial\mathbb{D}$ to one another such that the fundamental domains in P are mapped one-to-one to fundamental domains in P' and such that consecutive fundamental domains in P are mapped to consecutive fundamental domains in P'. We will use the symbol $\longrightarrow$ for this map. Furthermore, we require that every fundamental domain which belongs both to P and P' is mapped to itself by $\longrightarrow$.
First of all, we see by Corollary 3.1 that P and P' are neighbouring edge paths. We consider four different cases:

1. P and P' contain no common fundamental domain.
   In this case there are two possibilities.

   - P and P' may run on different sides along some geodesic C in N. The angles of P and P' are then π at all vertices. We pick two adjacent fundamental domains $\mathcal{F}_s, \mathcal{F}'_t$ with a common side on C. We map $\mathcal{F}_s$ and $\mathcal{F}'_t$ to one another:

     $$\mathcal{F}_s \longrightarrow \mathcal{F}'_t.$$

     We map inductively succeeding and preceding fundamental domains in P and P' to one another as follows

     $$\mathcal{F}_{s+r} \longrightarrow \mathcal{F}'_{t+r} \text{ for all } r\in\mathbb{Z}.$$

     This map is obviously independent of the choice of the fundamental domains $\mathcal{F}_s$ and $\mathcal{F}'_t$.
   - P and P' may also run along a sequence of vertices $\dots, v_{-2}, v_{-1}, v_0, v_1, v_2, \dots$ which lie not on a single geodesic in N. A vertex v_i is either flat (i.e. both P and P' have the vertex angle π at v_i) or one of the polygonal paths has at v_i the angle π^+ and the other has the angle π^-. For all other angle combinations either P and P' would not be neighbouring or P and P' would contain a common fundamental domain. Furthermore, notice that if one of the polygonal paths has the angle π^+ (resp. π^-) at v_i, then this polygonal path has at the preceding non-flat vertex and at the next non-flat vertex the angle π^- (resp. π^+).
     We pick now a non-flat vertex v_0 and denote the last fundamental domain in P which is adjacent to v_0 with $\mathcal{F}_s$ and the last fundamental domain in P' which is adjacent to v_0 with $\mathcal{F}'_t$.
     We now map the fundamental domains of P to the fundamental domains of P' by

     $$\mathcal{F}_{s+r\mp1} \longrightarrow \mathcal{F}'_{t+r} \text{ for all } r\in\mathbb{Z}. \tag{65}$$

     Here we make the convention to pick the upper sign if P has at v_0 the vertex angle π^+ and the lower sign if P has at v_0 the vertex angle π^-. We will adopt this convention throughout.
     The so defined map is obviously independent of the non-flat vertex v_0. Let v_{k_1} be the next non-flat vertex. We denote the last fundamental domain in P which is adjacent to v_{k_1} with $\mathcal{F}_p$ and the last fundamental domain in P' which is adjacent to v_{k_1} with $\mathcal{F}'_q$. If P has at v_0 the angle $\pi^\pm$, then P has at v_{k_1} the angle $\pi^\mp$. (Similarly for P'.) We set $\kappa := p-s$. Since P passes at v_{k_1} through $n(v_{k_1})\mp1$ fundamental domains and since P' passes at v_{k_1} through $n(v_{k_1})\pm1$ fundamental domains, we find that $\kappa\pm2=q-t$. If we choose instead of v_0 another vertex, e.g. v_{k_1}, then we are led to: $\mathcal{F}_{p+r\pm1}\longrightarrow\mathcal{F}'_{q+r}$. However, this is equivalent to: $\mathcal{F}_{s+\kappa+r\pm1}\longrightarrow\mathcal{F}'_{t+\kappa+r\pm2}$. And this is equivalent to Equation 65. This argument implies that the map in Equation 65 is independent of the choice of the non-flat vertex v_0.

2. P and P' contain common fundamental domains.
   Every of these common fundamental domains is mapped to itself. Let $P_{i\to p}:=\mathcal{F}_i,\dots,\mathcal{F}_p$ be a sequence of consecutive fundamental domains in P and let $P'_{j\to q}:=\mathcal{F}'_j,\dots,\mathcal{F}'_q$ be a sequence of consecutive fundamental domains in P' such that $\mathcal{F}_i=\mathcal{F}'_j$, $\mathcal{F}_p=\mathcal{F}'_q$ and $\mathcal{F}_r\neq\mathcal{F}_s$ for all $i<r<p$ and $j<s<q$.
   We show now that $P_{i\to p}$ and $P'_{j\to q}$ pass through the same number of fundamental domains. The subsequences $P_{i\to p}$ and $P'_{j\to q}$ are neighbouring and therefore run along a sequence of vertices $v_0, v_1, \dots, v_n$. One of the two polygonal paths (without loss of generality P) has at v_0 the vertex angle π^+, the other (P') has the angle π. P passes therefore through $n(v_0)+1$ fundamental domains at v_0 and P' passes through $n(v_0)$ fundamental domains at v_0. After the vertex v_0 follows a sequence of vertices $v_1,\dots,v_{k_1-1}$ at which both P and P' have the angle π, i.e. both pass through $n(v_j)$ fundamental domains at the vertex v_j for $0<j<k_1<n$. As above, we call such a vertex also a FLAT VERTEX. P has the angle π^- at v_{k_1} and P' has the angle π^+ at v_{k_1}, i.e. P passes through $n(v_{k_1})-1$ fundamental domains and P' passes through $n(v_{k_1})+1$ fundamental domains at v_{k_1}. (All other combinations of vertex angles for P and P' are not allowed since: (i) if the vertex angle of P at v_{k_1} were π^+, then P would contain a long chain or a long cycle, and (ii) if the vertex angles of P/P' at v_{k_1} were $\pi^-/\pi, \pi^-/\pi^-$ resp. π/π^-, then P and P' would not be neighbouring after going round v_{k_1} and (iii) if the vertex angles of P/P' at v_{k_1} were π/π^+, then there would exists an r_1, with $i<r_1<p$, and an s_1, with $j<s_1<q$, such that $\mathcal{F}_{r_1}=\mathcal{F}_{s_1}$. This contradicts the supposition.)
   After the vertex v_{k_1} follows a sequence of vertices $v_{k_1+1},\dots v_{k_2-1}$ at which both P and P' have the angle π. P has the angle π^+ and P' has the angle π^- at the vertex v_{k_2}.
   Proceeding in this way, we arrive finally at the vertex v_n. One polygonal path has at v_n the angle π^+ (namely those polygonal path whose vertex angle at the preceding non-flat vertex is not π^+) and the other has the angle π. Simple counting now yields: $\kappa:=p-i=q-j$ and we obtain the map:

   $$\mathcal{F}_{i+r}\longrightarrow\mathcal{F}'_{j+r} \text{ for } 0\le r\le\kappa.$$

3. Let $P_{i\to\infty}:=\mathcal{F}_i,\mathcal{F}_{i+1},\dots$ and $P'_{j\to\infty}:=\mathcal{F}'_j,\mathcal{F}'_{j+1},\dots$ be two subsequences of fundamental domains of P and of P' respectively such that $\mathcal{F}_i=\mathcal{F}'_j$ and $\mathcal{F}_{i+s}\neq\mathcal{F}'_{j+t}$ for all $s,t>0$. Then $P_{i\to\infty}$ and $P'_{j\to\infty}$ are mapped to one another in the following way:

   $$\mathcal{F}_{i+r}\longrightarrow\mathcal{F}'_{j+r} \text{ for all } r\in\mathbb{N}.$$

4. Let now $P_{-\infty\to i}:=\dots,\mathcal{F}_{i-1},\mathcal{F}_i$ and $P'_{-\infty\to j}:=\dots,\mathcal{F}'_{j-1},\mathcal{F}'_j$ be two one-sided infinite subsequences of fundamental domains of P and of P' respectively such that $\mathcal{F}_i=\mathcal{F}'_j$ and $\mathcal{F}_{i+s}\neq\mathcal{F}'_{j+t}$ for all $s,t<0$. Then $P_{-\infty\to i}$ and $P'_{-\infty\to j}$ are mapped to one another as follows:

   $$\mathcal{F}_{i-r}\longrightarrow\mathcal{F}'_{j-r} \text{ for all } r\in\mathbb{N}.$$

Notice that if a fundamental domain $\mathcal{F}_i$ in P lies in a cycle or a chain round some vertex, then also the fundamental domain in P' to which $\mathcal{F}_i$ is mapped by the map $\longleftrightarrow$ lies in a cycle or in a chain round the same vertex.

We finally remark that the procedure in point 2 above can of course also be used to map two finite shortest complementary cycles or chains to one another.

The conjugacy $T : \mathcal{R} \to \mathcal{A}$

After all this preliminary remarks, we are now enabled to state the definition of the bijective conjugacy T between $(\mathcal{A}, f)$ and $(\mathcal{R}, \tau)$ and to write down the main theorem of this section, namely Theorem 3.14, which states that T conjugates the action of f on $\mathcal{A}$ with the action of τ on $\mathcal{R}$. Because of this theorem we are justified in considering the Bowen-Series map instead of the geodesic flow on $\mathbb{D}$.

We define at first a map $T : \mathcal{R} \to \mathcal{A}$ and a map $S : \mathcal{A} \to \mathcal{R}$.

Definition 3.16

- *Let $\gamma \in \mathcal{R} \cap \mathcal{A}$. Then we set*

$$T(\gamma) := \gamma \tag{66}$$
$$S(\gamma) := \gamma. \tag{67}$$

- *Let $\gamma = \gamma(\xi, \eta) \in \mathcal{A} - (\mathcal{R} \cap \mathcal{A})$. The edge paths $E(\gamma)$ and $E(\xi * \eta)$ have coincident endpoints on $\partial\mathbb{D}$. Furthermore, $0 \in E(\xi * \eta)$ and $0 \notin E(\gamma)$. Since $\gamma \in \mathcal{A}$, $E(\xi * \eta)$ is a shortest edge path by Equation 64. And by Theorem 3.8, $E(\gamma)$ is a shortest edge path. By Corollary 3.1 (*), $E(\xi * \eta)$ and $E(\gamma)$ are neighbouring and therefore they can be mapped to one another with the help of the map $\longleftrightarrow$ defined above: $E(\gamma) \longleftrightarrow E(\xi * \eta)$. Here let $h\mathcal{F}$ denote the fundamental domain in $E(\gamma)$ which is mapped to $\mathcal{F}$ in $E(\xi * \eta)$. Then we set*

$$S(\gamma) := h^{-1}\gamma. \tag{68}$$

- *Let $\gamma \in N(\mathcal{F}) \cap \mathcal{R}$. By definition $E(\gamma)$ lies on the right of γ. We consider now a curve infinitesimally shifted to the left of γ and determine the edge path $E'(\gamma)$ of this curve. $E'(\gamma)$ passes successively through all fundamental domains on the left of γ which have at least one vertex on γ. $E(\gamma)$ and $E'(\gamma)$ have coincident endpoints on $\partial\mathbb{D}$; they are neighbouring by Corollary 3.1 (*) and can be mapped to one another by the map $\longleftrightarrow$. Let thereby $g\mathcal{F}$ denote the fundamental domain in $E'(\gamma)$ which is mapped to $\mathcal{F}$ in $E(\gamma)$. We set*

$$T(\gamma) := g^{-1}\gamma. \tag{69}$$

- *Let $\gamma = \gamma(\xi, \eta) \in \mathcal{R} - (\mathcal{A} \cap \mathcal{R}) - (N(\mathcal{F}) \cap \mathcal{R})$. Then $E(\xi * \eta)$ is a shortest edge path (cf. Lemma 3.7). By Lemma 3.9, γ passes near the vertex $v = v(\xi, \eta)$. Therefore also $E(\xi * \eta)$ passes near v. The sectors in the net defined by $N(v)$ which contain ξ and η are bounded by two geodesics in $N(v)$; we denote this two geodesics by $C_1, C_2 \in N(v)$. $E(\xi * \eta)$ intersects both C_1 and C_2 once at the points $s_1, s_2 \in \mathbb{D} \cup \partial\mathbb{D}$ respectively. We suppose that both s_1 and s_2 lie not on $\partial\mathbb{D}$. We denote the vertices in N which (i) lie on C_1 between v and s_1 resp. on C_2 between v and s_2 and which (ii) are closest to the points s_1 resp. s_2 by v_1 resp. v_2. (Note that it is possible that one (or both) vertices v_1, v_2 coincide with v.) Let $\mathcal{F}_1$ be the first fundamental domain in $E(\xi * \eta)$ which has v_1 as vertex and let $\mathcal{F}_2$ be the last fundamental domain in $E(\xi * \eta)$ which has v_2 as vertex. The edge subpath $E_{1\to 2} := \mathcal{F}_1, \ldots, \mathcal{F}_2$ of $E(\xi * \eta)$ begins with a half cycle at v_1 (angle $= \pi^+$), runs then along C_1 thereby passing between the vertices v_1 and v a sequence of flat vertices on C_1 (i.e. the vertex angle is π), has at v the vertex angle π^- and runs then along C_2 thereby passing between the vertices v and v_2 a sequence of flat vertices on C_2 such that $E_{1\to 2}$ ends with a half cycle at v_2 (angle $= \pi^+$).*
  We consider now the edge path $E^*_{1\to 2}$ which is complementary to $E_{1\to 2}$ and relates $\mathcal{F}_1$ and $\mathcal{F}_2$, i.e. $E^*_{1\to 2}$ and $E(\xi * \eta)$ cross no common fundamental domain besides $\mathcal{F}_1$ and $\mathcal{F}_2$. $E^*_{1\to 2}$ passes between $\mathcal{F}_1$ and $\mathcal{F}_2$ through the same number of fundamental domains as $E_{1\to 2}$ (at each of the vertices v_1 and v_2 one less and at v two more than $E_{1\to 2}$ since $E^*_{1\to 2}$ has at v the angle π^+ and at v_1 and v_2 the angle π).
  Therefore $E_{1\to 2}$ and $E^*_{1\to 2}$ are two shortest complementary edge paths, i.e. $E_{1\to 2}$ and $E^*_{1\to 2}$ are neighbouring edge paths; if one (or both) of the intersection points s_1, s_2 lies on $\partial\mathbb{D}$, then we replace $E_{1\to 2}$ and $E^*_{1\to 2}$ by their one-sided (or two-sided) infinite counterparts. If we remove the subpath $E_{1\to 2}$ from $E(\xi * \eta)$ and insert instead the subpath $E^*_{1\to 2}$, then we obtain an edge path which we denote by $E^*(\xi * \eta)$. $E^*(\xi * \eta)$ is a shortest edge path; $E(\gamma)$ and $E^*(\xi * \eta)$ have coincident endpoints on $\partial\mathbb{D}$; they are therefore neighbouring by Corollary 3.1 (*) and can be mapped to one another with the help of the map $\longleftrightarrow$. Let $g\mathcal{F}$ denote the fundamental domain in $E^*(\xi * \eta)$ which is mapped to $\mathcal{F}$ in $E(\gamma)$; then we set

$$T(\gamma) := g^{-1}\gamma. \tag{70}$$

For the so defined maps T and S the following theorem holds:

Theorem 3.13 *The maps $T : \mathcal{R} \to \mathcal{A}$ and $S : \mathcal{A} \to \mathcal{R}$ are bijective and satisfy:*

$$T \circ S = id_{\mathcal{A}} \text{ and } S \circ T = id_{\mathcal{R}}.$$

Proof: Obviously, we only have to prove that T maps $\mathcal{R} - (\mathcal{R} \cap \mathcal{A})$ bijective to $\mathcal{A} - (\mathcal{R} \cap \mathcal{A})$ and that $T \circ S = id_{\mathcal{A}}$ and $S \circ T = id_{\mathcal{R}}$ holds.

- Let first of all $\gamma = \gamma(\xi, \eta) \in N(\mathcal{F}) \cap (\mathcal{A} - (\mathcal{R} \cap \mathcal{A}))$.
  By construction, $E(\gamma)$ passes through all fundamental domains which are adjacent to γ on the right side of γ and $E(\xi * \eta)$ passes through all fundamental domains which are adjacent to γ on the left side of γ. The map $\longleftrightarrow$ maps $g\mathcal{F} \in E(\gamma)$ to $\mathcal{F} \in E(\xi * \eta)$ and we have: $S(\gamma) = g^{-1}\gamma$.
  This implies that $E(g^{-1}\gamma)$ passes successively through exactly all fundamental domains adjacent to $g^{-1}\gamma$ on the right of $g^{-1}\gamma$ and therefore we have $0 \in E(g^{-1}\gamma)$; it follows by Lemma 3.8 that $g^{-1}\gamma \in N(\mathcal{F}) \cap (\mathcal{R} - (\mathcal{R} \cap \mathcal{A}))$ and we can apply T to $g^{-1}\gamma$. The edge path $E'(g^{-1}\gamma)$ which runs on the left side of $g^{-1}\gamma$ satisfies: $E'(g^{-1}\gamma) = g^{-1}E(\xi * \eta)$ and therefore the map $\longleftrightarrow$ maps $g^{-1}\mathcal{F} \in E'(g^{-1}\gamma)$ to $\mathcal{F} \in E(g^{-1}\gamma)$. Thus: $T(S(\gamma)) = T(g^{-1}\gamma) = gg^{-1}\gamma = \gamma$.

- Let $\gamma \in N(\mathcal{F}) \cap (\mathcal{R} - (\mathcal{R} \cap \mathcal{A}))$.
  Then we can reverse the argument applied in the previous point. $E'(\gamma)$ runs on the left of γ and $E(\gamma)$ runs on the right of γ. The map $\longleftrightarrow$ maps $g\mathcal{F} \in E'(\gamma)$ and $\mathcal{F} \in E(\gamma)$ to one another. Hence: $T(\gamma) = g^{-1}\gamma$.
  On the other hand, we have that $E(g^{-1}\gamma) = g^{-1}E(\gamma)$ runs on the right of $g^{-1}\gamma$. By Lemma

3.8, it follows that $T(\gamma) \in N(\mathcal{F}) \cap (\mathcal{A} - (\mathcal{R} \cap \mathcal{A}))$. Therefore we can apply S to $T(\gamma)$.
$E(g^{-1}\xi * g^{-1}\eta)$ passes through exactly all fundamental domains adjacent to $g^{-1}\gamma$ on the left of $g^{-1}\gamma$. Therefore: $E(g^{-1}\xi * g^{-1}\eta) = g^{-1}E'(\gamma)$ and the map $\longleftrightarrow$ maps $g^{-1}\mathcal{F} \in E(g^{-1}\gamma)$ to $\mathcal{F} \in E(g^{-1}\xi * g^{-1}\eta)$. Thus: $S(T(\gamma)) = S(g^{-1}\gamma) = gg^{-1}\gamma = \gamma$.

- Let now $\gamma = \gamma(\xi, \eta) \in \mathcal{A} - (N(\mathcal{F}) \cap \mathcal{A}) - (\mathcal{A} \cap \mathcal{R})$.
In the definition of S we have seen that in this case both shortest edge paths $E(\gamma)$ and $E(\xi * \eta)$ are mapped according to $E(\xi * \eta) \longleftrightarrow E(\gamma)$ to one another. Let hereby $h\mathcal{F}$ denote the fundamental domain in $E(\gamma)$ which is mapped to $\mathcal{F} \in E(\xi * \eta)$, then we have $S(\gamma) = h^{-1}\gamma$. From $h0 \in E(\gamma)$ it follows at once that $0 \in E(h^{-1}\gamma)$. Thus: $h^{-1}\gamma \in \mathcal{R} - (N(\mathcal{F}) \cap \mathcal{R})$; for if $h^{-1}\gamma \in N(\mathcal{F})$, then also $\gamma \in N(\mathcal{F})$.
By Lemma 3.9, γ passes near the vertex $v = v(\xi, \eta)$ and the sequence $\xi_0^{-1}\eta_0$ lies in a cycle or in a chain. Therefore we can apply Theorem 3.12 and find that γ goes counter-clockwise round the vertex v. Furthermore, that $\xi_0^{-1}\eta_0$ lies in a cycle or in a chain implies that the fundamental domains in $E(\gamma)$ which are mapped by $\longleftrightarrow$ to the fundamental domains in this cycle or in this chain also lie in a cycle or in a chain. Especially it follows that $h\mathcal{F} \in E(\gamma)$ lies in a cycle or in a chain of $E(\gamma)$. This implies that none of the endpoints of γ lies in those sector (of the net defined by the geodesics in $N(v)$ of $\mathbb{D}$) which contains the fundamental domain $h\mathcal{F} \notin E(\xi * \eta)$. Thus, γ cuts off the vertex v on $h\mathcal{F}$ (cf. Definition 3.15).
This implies that $h^{-1}\gamma$ passes near the vertex $h^{-1}v$, cuts off $h^{-1}v$ on $\mathcal{F}$ and goes counter-clockwise round $h^{-1}v$. The assumption $h^{-1}\gamma \in \mathcal{A}$ leads now to a contradiction (recall Theorem 3.12). Thus: $S(\gamma) = h^{-1}\gamma \in \mathcal{R} - (N(\mathcal{F}) \cap \mathcal{R}) - (\mathcal{A} \cap \mathcal{R})$ and we can apply T to $S(\gamma)$.
$E(h^{-1}\xi * h^{-1}\eta)$ is a shortest edge path: otherwise we would have by Lemma 3.7 that $h^{-1}\gamma \in N(\mathcal{F})$ and $\gamma \in N(\mathcal{F})$, contrary to the supposition. Since $S(\gamma) \notin \mathcal{A}$, $E(h^{-1}\xi * h^{-1}\eta)$ contains a right-handed (possibly pseudo) half cycle which contains $\xi_0^{-1}\eta_0$. In the edge path $E^*(h^{-1}\xi * h^{-1}\eta)$ introduced in Definition 3.16 this right-handed (possibly pseudo) half cycle is replaced by the corresponding complementary left-handed (possibly pseudo) half cycle. Since $E(h^{-1}\xi * h^{-1}\eta)$ contains by Lemma 3.5 no further right-handed (pseudo) half cycle, all half cycles and pseudo half cycles in $E^*(h^{-1}\xi * h^{-1}\eta)$ are left-handed.
Similarly $E(\xi * \eta)$ and also $h^{-1}E(\xi * \eta)$ contain only left-handed half cycles and pseudo half cycles. By Corollary 3.1 (**), we have: $E^*(h^{-1}\xi * h^{-1}\eta) = h^{-1}E(\xi * \eta)$ since both edge paths have coincident endpoints on $\partial\mathbb{D}$.
Since the map $E(\gamma) \longleftrightarrow E(\xi * \eta)$ maps the domains $\mathcal{F} \in E(\xi * \eta)$ and $h\mathcal{F} \in E(\gamma)$ to one another, the map $h^{-1}E(\gamma) = E(h^{-1}\gamma) \longleftrightarrow E^*(h^{-1}\xi * h^{-1}\eta) = h^{-1}E(\xi * \eta)$ maps the domains $\mathcal{F} \in E(h^{-1}\gamma)$ and $h^{-1}\mathcal{F} \in E^*(h^{-1}\xi * h^{-1}\eta)$ to one another. Thus: $T(S(\gamma)) = T(h^{-1}\gamma) = hh^{-1}\gamma = \gamma$.

- Let finally $\gamma = \gamma(\xi, \eta) \in \mathcal{R} - (\mathcal{R} \cap N(\mathcal{F})) - (\mathcal{A} \cap \mathcal{R})$.
By Definition 3.16, we have: $E^*(\xi * \eta) \longleftrightarrow E(\gamma)$ and thereby especially: $g\mathcal{F} \longleftrightarrow \mathcal{F}$. Thus $T(\gamma) = g^{-1}\gamma$.
By Lemma 3.7, $E(\xi * \eta)$ is a shortest edge path. By Lemma 3.9, the sequence $\xi_0^{-1}\eta_0$ lies in a cycle or in a chain and $E(\gamma)$ passes near the vertex $v = v(\xi, \eta)$. By Corollary 3.3, $E(\gamma)$ goes counter-clockwise round the vertex v.
Hence, $g^{-1}\gamma$ passes near $g^{-1}v$ and goes counter-clockwise round $g^{-1}v$. Remember that $g0 \notin E(\gamma)$ and $0 \notin E(g^{-1}\gamma)$ and thus $g^{-1}\gamma \notin \mathcal{R}$. Then it follows by Corollary 3.3 that $g^{-1}\gamma \in \mathcal{A} - (N(\mathcal{F}) \cap \mathcal{A}) - (\mathcal{A} \cap \mathcal{R})$. Therefore we can apply S to $T(\gamma)$.
By Equation 64, $E(g^{-1}\xi * g^{-1}\eta)$ is a shortest edge path. Furthermore $E(\xi * \eta)$ is a shortest

edge path: otherwise Lemma 3.7 would imply that $\gamma \in N(\mathcal{F})$, contrary to the supposition. Since $\gamma \notin \mathcal{A}$, $E(\xi * \eta)$ contains a right-handed (possibly pseudo) half cycle which contains $\xi_0^{-1}\eta_0$. In the edge path $E^*(\xi * \eta)$ introduced in Definition 3.16 this right-handed (possibly pseudo) half cycle is replaced by the corresponding complementary left-handed (possibly pseudo) half cycle. Since $E(\xi * \eta)$ contains by Lemma 3.5 no further right-handed (pseudo) half cycles, all half cycles and pseudo half cycles in $E^*(\xi * \eta)$ are left-handed.
Similarly, since $g^{-1}\gamma \in \mathcal{A}$, $E(g^{-1}\xi * g^{-1}\eta)$ contains only left-handed half cycles and pseudo half cycles. By Corollary 3.1 (**), we have: $E(g^{-1}\xi * g^{-1}\eta) = g^{-1}E^*(\xi * \eta)$ since the two edge paths have coincident endpoints on $\partial\mathbb{D}$.
Since the map $E(\gamma) \longleftrightarrow E^*(\xi * \eta)$ maps the domains $\mathcal{F} \in E(\gamma)$ and $g\mathcal{F} \in E^*(\xi * \eta)$ to one another, the map $g^{-1}E(\gamma) = E(g^{-1}\gamma) \longleftrightarrow E(g^{-1}\xi * g^{-1}\eta) = g^{-1}E^*(\xi * \eta)$ maps the domains $\mathcal{F} \in E(g^{-1}\xi * g^{-1}\eta)$ and $g^{-1}\mathcal{F} \in E(g^{-1}\gamma)$ to one another. Hence: $S(T(\gamma) = S(g^{-1}\gamma) = gg^{-1}\gamma = \gamma$. □

Theorem 3.14 *The map $T : \mathcal{R} \to \mathcal{A}$ is a conjugacy between $(\mathcal{R}, \tau)$ and $(\mathcal{A}, f)$, i.e. T satisfies: $T \circ \tau = f \circ T$.*

Proof: The proof given in Series [28] is incomplete. We give a completed and corrected version. However, some parts and the idea of the proof is taken from [28]. In the following proof we always use the notations: $\gamma = \gamma(\xi, \eta), \eta_f = \eta_0\eta_1...$ and $E(\gamma) = e_0e_1....$ With this notation $\tau(\gamma) = e_0^{-1}\gamma$. We distinguish five different cases:

- First of all, let $\gamma = \gamma(\xi, \eta) \in N(\mathcal{F}) \cap \mathcal{R}$.
Then either γ is the geodesic completion of some side of $\mathcal{F}$ or γ passes through some vertex of $\mathcal{F}$. In both cases $\tau(\gamma) \in N(\mathcal{F}) \cap \mathcal{R}$ holds. According to Definition 3.16, we map the edge path $E(\gamma)$ lying to the right of γ to the edge path $E'(\gamma)$ lying to the left of γ: $E(\gamma) \longleftrightarrow E'(\gamma)$. Especially $\mathcal{F} \longleftrightarrow g\mathcal{F}$ and therefore $T(\gamma) = g^{-1}\gamma$.
The fundamental domain in $E(\gamma)$ following $\mathcal{F}$ is $e_0\mathcal{F}$ and the fundamental domain in $E'(\gamma)$ following $g\mathcal{F}$ is denoted by $gh_0\mathcal{F}$. Thus: $e_0\mathcal{F} \longleftrightarrow gh_0\mathcal{F}$.
In order to determine the action of T on $\tau(\gamma)$, we have to consider the mapping $E(\tau(\gamma)) \longleftrightarrow E'(\tau(\gamma))$. Notice that $e_0^{-1}E(\gamma) = E(\tau(\gamma))$ and $e_0^{-1}E'(\gamma) = E'(\tau(\gamma))$. Then we find that especially $\mathcal{F} \longleftrightarrow e_0^{-1}gh_0\mathcal{F}$. Thus, $T(\tau(\gamma)) = h_0^{-1}g^{-1}\gamma$.
It remains to show that $f(T(\gamma)) = h_0^{-1}T(\gamma)$. Obviously, it is sufficient to show that the boundary expansion of $g^{-1}\eta$ begins with h_0.
If γ is a side of $\mathcal{F}$, then choose i such that $g = g_i$ and $\eta = P_i$. The boundary expansion of P_i is a left-handed pseudo half cycle which begins with the sequence $\mathcal{F}, g\mathcal{F}, gh_0\mathcal{F}, ...$; therefore the boundary expansion of η starts with gh_0 and the boundary expansion of $g^{-1}\eta$ begins with h_0.
Consider now the case that γ is a geodesic which passes through only one vertex v of $\mathcal{F}$. Since $\gamma \in N(\mathcal{F})$, γ does not enter the fundamental domain $\mathcal{F}$. The fundamental domain $g\mathcal{F}$ contains also the vertex v. In order to determine the boundary expansion of $g^{-1}\eta$, we consider the geodesic $g^{-1}\gamma$.
$g^{-1}\gamma$ passes through the vertex $g^{-1}v$ of $g^{-1}\mathcal{F}$ and of $\mathcal{F}$. Since $\gamma \in N(\mathcal{F})$, $g^{-1}\gamma$ does enter neither the fundamental domain $\mathcal{F}$ nor $g^{-1}\mathcal{F}$. Hence, $E'(g^{-1}\gamma) = g^{-1}E(\gamma)$ and $E(g^{-1}\gamma) = g^{-1}E'(\gamma)$. This implies that the fundamental domain in $E(g^{-1}\gamma)$ following $\mathcal{F}$ equals $h_0\mathcal{F}$.
We choose i such that $h_0 = g_i$. This choice implies $g^{-1}v = v_{i+1}$. The point $g^{-1}\eta$ coincides then with one of the points $S_{i+1,k}$, which implies that $g^{-1}\eta \in [P_i, P_{i+1}[$. Therefore the boundary expansion of $g^{-1}\eta$ starts with $g_i = h_0$.

- Let $\gamma, \tau(\gamma) \in \mathcal{R} \cap \mathcal{A}$.
This implies that $e_0 = \eta_0$. Suppose that this is not the case. Then γ passes near a vertex v of $\mathcal{F}$ by Lemma 3.10. Further, by Lemma 3.10, it follows that $E(\gamma)$ contains a right-handed chain beginning at the vertex v. (In the proof of Lemma 3.10 we have denoted this vertex by $v = v(e_0e_1 \ldots, \eta)$.) v is also a vertex of $e_0\mathcal{F}$ and therefore it follows that γ passes near some vertex of $e_0\mathcal{F}$ and goes counter-clockwise round this vertex (i.e. in a right-handed cycle). This implies that $e_0^{-1}\gamma$ passes near some vertex of $\mathcal{F}$ and goes counter-clockwise round this vertex (in a right-handed cycle). By Theorem 3.12, it follows that $e_0^{-1}\gamma \notin \mathcal{A}$, contrary to the supposition. Thus $e_0 = \eta_0$.
The proof of Lemma 3.10 implies that $\eta \in [P_i, P_{i+1}[$, where i is such that $e_0 = g_i$. Hence, $T(\tau(\gamma)) = \tau(\gamma) = e_0^{-1}\gamma = f(\gamma) = f(T(\gamma))$. In this equation we have used the definition of f on $\mathcal{A}$ given by Equation 63.

- Let $\gamma = \gamma(\xi, \eta) \in \mathcal{R} \cap \mathcal{A}, \tau(\gamma) \in \mathcal{R} - (\mathcal{R} \cap \mathcal{A})$ and $\gamma \notin N(\mathcal{F})$.
Since $\gamma = \gamma(\xi, \eta) \in \mathcal{R} \cap \mathcal{A}$, the two edge paths $E(\gamma)$ and $E(\xi * \eta)$ are shortest edge paths with coincident endpoints on $\partial \mathbb{D}$. By Corollary 3.1 (*), $E(\gamma)$ and $E(\xi * \eta)$ are neighbouring and therefore can be mapped to one another with the help of the map $\longleftrightarrow$:

$$E(\gamma) \longleftrightarrow E(\xi * \eta).$$

Both edge paths contain the fundamental domain $\mathcal{F}$. The fundamental domain following $\mathcal{F}$ in $E(\gamma)$ is $e_0\mathcal{F}$ and the fundamental domain following $\mathcal{F}$ in $E(\xi * \eta)$ is $\eta_0\mathcal{F}$. Therefore:

$$e_0\mathcal{F} \longleftrightarrow \eta_0\mathcal{F}.$$

$E(e_0^{-1}\xi * e_0^{-1}\eta) = E(\tau(\xi) * \tau(\eta))$ is a shortest edge path. Otherwise, Lemma 3.7 would imply that $\tau(\gamma) \in N$ and therefore $\gamma \in N$, contrary to the supposition. By Equation 64, $E(\tau(\xi) * \tau(\eta))$ contains a right-handed half cycle or right-handed pseudo half cycle which contains $\mathcal{F}$. As above in Definition 3.16, we introduce $E^*(\tau(\xi) * \tau(\eta))$. This right-handed (possibly pseudo) half cycle in $E(\tau(\xi) * \tau(\eta))$ is replaced in $E^*(\tau(\xi) * \tau(\eta))$ by the complementary left-handed (possibly pseudo) half cycle. Thus $E^*(\tau(\xi) * \tau(\eta))$ is a shortest edge path without right-handed half cycles and without right-handed pseudo half cycles.
$e_0^{-1}E(\xi * \eta)$ is also a shortest edge path without right-handed (pseudo) half cycles with the same endpoints as $E^*(\tau(\xi) * \tau(\eta))$. By Corollary 3.1 (**), we have:

$$E^*(\tau(\xi) * \tau(\eta)) = e_0^{-1}E(\xi * \eta).$$

Notice that $e_0^{-1}E(\gamma) = E(\tau(\gamma))$. We find that the neighbouring edge paths $E(\tau(\gamma))$ and $E^*(\tau(\xi) * \tau(\eta))$ can be mapped to one another with the map $\longleftrightarrow$:

$$\begin{array}{ccc} E(\tau(\gamma)) & \longleftrightarrow & E^*(\tau(\xi) * \tau(\eta)) \\ \mathcal{F} & \longleftrightarrow & e_0^{-1}\eta_0\mathcal{F}. \end{array}$$

By Definition 3.16, it follows that: $T(\tau(\gamma)) = \eta_0^{-1}e_0\tau(\gamma) = \eta_0^{-1}\gamma = f(\gamma) = f(T(\gamma))$.

- Let $\gamma = \gamma(\xi, \eta) \in \mathcal{R} - (\mathcal{A} \cap \mathcal{R}), \tau(\gamma) \in \mathcal{R} \cap \mathcal{A}$ and $\gamma \notin N(\mathcal{F})$.
The two edge paths $E(\tau(\gamma))$ and $E(\tau(\xi) * \tau(\eta))$ are shortest edge paths and can therefore be mapped to one another by the map $\longleftrightarrow$:

$$E(\tau(\gamma)) \longleftrightarrow E(\tau(\xi) * \tau(\eta)).$$

Both edge paths contain the fundamental domain $\mathcal{F}$. The fundamental domain in $E(\tau(\gamma))$ preceding $\mathcal{F}$ is $e_0^{-1}\mathcal{F}$ and the fundamental domain in $E(\tau(\xi) * \tau(\eta))$ preceding $\mathcal{F}$ will be denoted by $h_0^{-1}\mathcal{F}$. Thus: $e_0^{-1}\mathcal{F} \longleftrightarrow h_0^{-1}\mathcal{F}$.
As in the previous item we find $E^*(\xi * \eta) = e_0E(\tau(\xi) * \tau(\eta))$ since both edge paths are shortest edge paths without right-handed (pseudo) half cycles with coincident endpoints on $\partial \mathbb{D}$. Notice that $E(\gamma) = e_0E(\tau(\gamma))$; then

$$\begin{array}{ccc} E(\gamma) & \longleftrightarrow & E^*(\xi * \eta) \\ \mathcal{F} & \longleftrightarrow & e_0h_0^{-1}\mathcal{F}. \end{array}$$

Thus: $T(\gamma) = h_0e_0^{-1}\gamma$.
It remains to show that $f(T(\gamma)) = h_0^{-1}T(\gamma)$.
$E(\xi * \eta)$ is a shortest edge path and therefore we can apply Lemma 3.9 and find that $E(\gamma)$ passes near the vertex $v = v(\xi, \eta)$. On the other hand, $E(\gamma)$ and $E^*(\xi * \eta)$ contain both $e_0\mathcal{F}$ since $0 \in E(\tau(\gamma))$ and since $0 \in E(\tau(\xi) * \tau(\eta))$. Therefore the coincident endpoints of $E(\gamma)$ and $E^*(\xi * \eta)$ lie in the sectors bounded by the geodesics $C(\eta_0)$ and $C(h_0)$. Especially we see that $E(\gamma) = e_0e_1 \ldots$ does not begin with a right-handed chain. (This would namely imply $e_1 = h_0^{-1}$.) By Lemma 3.10, it follows that $\eta_0 = e_0$.
The point $f(\eta) = (e_0^{-1}\gamma)_\infty = e_0^{-1}\eta$ has therefore the boundary expansion $(f(\eta))_f = \eta_1\eta_2 \ldots$.
By construction, the edge paths $E^*(\xi * \eta)$ and $E(\xi * \eta)$ coincide in all fundamental domains succeeding the fundamental domain $e_0\mathcal{F}$ (cf. Definition 3.16); the edge sequence is given by $(e_0^{-1}\eta)_f = \eta_1\eta_2 \ldots$. The edge sequence in $E^*(\xi * \eta)$ beginning at the fundamental domain $e_0h_0^{-1}\mathcal{F}$ (which is preceding $e_0\mathcal{F}$) is $h_0\eta_1\eta_2 \ldots$. Since $E^*(\xi * \eta)$ is a shortest edge path without right-handed half cycles and without right-handed pseudo half cycles, also $h_0\eta_1\eta_2 \ldots$ is a shortest edge path without right-handed half cycles and without right-handed pseudo half cycles.
By Lemma 3.4, there exists exactly one $\phi \in \partial \mathbb{D}$ such that $\phi_f = h_0\eta_1\eta_2 \ldots$. By definition of the boundary expansion, the point $f(\phi) = h_0^{-1}\phi$ has the boundary expansion $(f(\phi))_f = \eta_1\eta_2 \ldots$. This implies (cf. Lemma 3.4) that $f(\phi) = e_0^{-1}\eta$. Thus: $\phi = h_0e_0^{-1}\eta$ and we have shown that the boundary expansion of $h_0e_0^{-1}\eta$ begins with h_0; i.e. $f(T(\gamma)) = h_0^{-1}T(\gamma) = h_0^{-1}h_0e_0^{-1}\gamma = \tau(\gamma)$.

- Finally, let $\gamma = \gamma(\xi, \eta) \in \mathcal{R} - (\mathcal{A} \cap \mathcal{R}), \tau(\gamma) \in \mathcal{R} - (\mathcal{A} \cap \mathcal{R})$ and $\gamma \notin N(\mathcal{F})$.
According to Definition 3.16 the map $\longleftrightarrow$ maps the shortest edge paths $E(\gamma)$ and $E^*(\xi * \eta)$ to one another:

$$\begin{array}{ccc} E(\gamma) & \longleftrightarrow & E^*(\xi * \eta) \\ \mathcal{F} & \longleftrightarrow & g\mathcal{F} \\ e_0\mathcal{F} & \longleftrightarrow & gh_0\mathcal{F}. \end{array}$$

Here $e_0\mathcal{F}$ denotes the fundamental domain in $E(\gamma)$ following $\mathcal{F}$ and $gh_0\mathcal{F}$ denotes the fundamental domain in $E^*(\xi * \eta)$ following $g\mathcal{F}$. Thus $T(\gamma) = g^{-1}\gamma$.
Since both $e_0^{-1}E^*(\xi * \eta)$ and $E^*(\tau(\xi) * \tau(\eta))$ are shortest edge paths without right-handed (pseudo) half cycles and since they have coincident endpoints on $\partial \mathbb{D}$, it follows that $e_0^{-1}E^*(\xi * \eta) = E^*(\tau(\xi) * \tau(\eta))$. Hence $E(\tau(\gamma)) \longleftrightarrow E^*(\tau(\xi) * \tau(\eta))$ and especially $\mathcal{F} \longleftrightarrow e_0^{-1}gh_0\mathcal{F}$. Thus we have $T(\tau(\gamma)) = h_0^{-1}g^{-1}e_0\tau(\gamma) = h_0^{-1}g^{-1}\gamma$.
It remains to show that $f(T(\gamma)) = h_0^{-1}T(\gamma)$ holds. Obviously, it is enough to show that the boundary expansion of $g^{-1}\eta$ starts with h_0.

We suppose first that $g = e_0 h_0^{-1}$; this is equivalent to the assertion that $e_0\mathcal{F} \in E(\gamma)$ and $e_0\mathcal{F} \in E^*(\xi * \eta)$; compare with Figure 3.7. Proceeding as in the previous item yields that $E(\gamma) = e_0 e_1 \ldots$ does not begin with a right-handed chain. By Lemma 3.10, it follows that $\eta_0 = e_0$. With literally the same proof as in the previous item one can show that the boundary expansion of $g^{-1}\eta = h_0 e_0^{-1}\eta$ begins with h_0.
In the sequel we suppose therefore $g \neq e_0 h_0^{-1}$. By Lemma 3.7, $\gamma \in \mathcal{R} - (\mathcal{R} \cap \mathcal{A})$ and $\gamma \notin \mathcal{N}$ imply that $E(\xi * \eta)$ is a shortest edge path. By Lemma 3.9, it follows moreover that γ passes near some vertex v of $\mathcal{F}$. Theorem 3.12 implies that γ goes counter-clockwise round this vertex v. Using exactly the same reasoning, we deduce from $\tau(\gamma) \in \mathcal{R} - (\mathcal{R} \cap \mathcal{A})$ that $\tau(\gamma)$ passes near some vertex v' of $\mathcal{F}$ and goes counter-clockwise round this vertex (this is equivalent to the statement that γ passes near a vertex v'' of $e_0\mathcal{F}$ and goes counter-clockwise round v'').

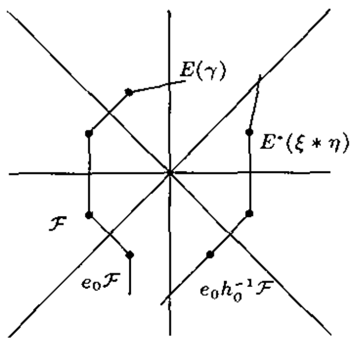

Figure 3.7

In order to determine the boundary expansion of $g^{-1}\eta$, we consider now the geodesic $g^{-1}\gamma$. $g^{-1}\gamma$ passes successively through the fundamental domains $g^{-1}\mathcal{F}, g^{-1}e_0\mathcal{F}, g^{-1}e_0 e_1\mathcal{F}, \ldots$. It follows immediately that $g^{-1}\gamma$ passes near the vertex $g^{-1}v$ of $g^{-1}\mathcal{F}$ and also near the vertex $g^{-1}v''$ of $g^{-1}e_0\mathcal{F}$ and that $g^{-1}\gamma$ goes counter-clockwise round these two vertices.
The fundamental domain $g^{-1}gh_0\mathcal{F} \in g^{-1}E^*(\xi * \eta)$ is not contained in $E(g^{-1}\gamma)$. Otherwise the fundamental domain $g^{-1}gh_0\mathcal{F}$ would be mapped to itself by $\longrightarrow$. Moreover, also the two fundamental domains preceding $g^{-1}gh_0\mathcal{F} = h_0\mathcal{F}$ in $E(g^{-1}\gamma)$ and in $g^{-1}E^*(\xi * \eta)$ would be mapped to one another. The fundamental domain in $g^{-1}E^*(\xi * \eta)$ preceding $h_0\mathcal{F}$ is, however, $\mathcal{F}$. Therefore the fundamental domain preceding $h_0\mathcal{F}$ in $E(g^{-1}\gamma)$ is $g^{-1}\mathcal{F}$. The fundamental domain in $E(g^{-1}\gamma)$ following $g^{-1}\mathcal{F}$ is $g^{-1}e_0\mathcal{F} = h_0\mathcal{F}$. This would imply $gh_0 = e_0$, contrary to the supposition.
We now choose i such that $h_0 = g_i$. (We adopt the convention which we made before Definition 3.13, see also Figure 3.1.) This means especially $g^{-1}v = v_i$. Since $g^{-1}E^*(\xi * \eta)$ is a shortest edge path without right-handed (pseudo) half cycles which passes successively through $\mathcal{F}$ and $h_0\mathcal{F}$, it is clear that $g^{-1}\eta \in [P_i, Q_{i+1}[$. We now show that moreover $g^{-1}\eta \notin]P_{i+1}, Q_{i+1}]$.

Suppose that $g^{-1}\eta \in]P_{i+1}, Q_{i+1}]$. Then $E(g^{-1}\gamma)$ intersects the geodesic completion $C(g_{i+1})$ of those side of $\mathcal{F}$ which is adjacent to $C(h_0)$ but does not contain the vertex $g^{-1}v$; notice that P_{i+1} is an endpoint of $C(g_{i+1})$. On the other hand, as already remarked above, $E(g^{-1}\gamma)$ does not intersect the domain $h_0\mathcal{F}$, i.e. $E(g^{-1}\gamma)$ goes around $h_0\mathcal{F}$. Since $h_0\mathcal{F}$ has at least five sides (cf. Consequence 1), it follows by Lemma 3.1 that $E(g^{-1}\gamma)$ intersects the geodesic completion of at least one side of $h_0\mathcal{F}$ twice. This contradicts the fact that $E(g^{-1}\gamma)$ is a shortest edge path.
Thus $g^{-1}\eta \in [P_i, P_{i+1}[$. And by definition of the boundary expansion, it follows that $(g^{-1}\eta)_f$ starts with h_0. □

In the work of Adler and Flatto [74] the conjugacy T (there denoted by Φ) is explicitely and geometrically constructed for special fundamental domains. [15] The Adler-Flatto construction immediately implies that T is in general discontinuous. The topology on $\mathcal{R}$ and $\mathcal{A}$ is choosen as the topology induced by the standard topology of $\partial\mathbb{D} \times \partial\mathbb{D}$ (every geodesic is identified with the pair of its endpoints). [16] T is therefore no topological conjugacy in the sense of Definition B.5. The map T is nonevertheless useful. With the help of T it is for instance possible to construct at once τ-invariant (resp. ergodic) measures on $\mathcal{R}$ from f-invariant (resp. ergodic) measures on $\mathcal{A}$ and vice versa (cf. Definition B.2 and Theorem B.2).

3.6 Quantum chaos and the Bowen-Series map

In the theory of quantum chaos one is interested in the semiclassical ($\hbar \to 0$) behaviour of quantum mechanical systems whose corresponding classical counterparts ($\hbar = 0$) exhibit chaos, i.e. neighbouring trajectories in phase space diverge at an exponential rate and so the time evolution of the classical system depends sensitively on the initial conditions. The geodesic flow on Riemann surfaces with negative curvature (and especially on Riemann surfaces with constant negative curvature) is strongly chaotic, which has been proven e.g. by Anosov [1]. These systems are well understood from a mathematical point of view and are therefore often used as model systems in the study of quantum chaos, cf. [92].
For a quantum mechanical description of a free particle on a Riemann surface S with constant negative curvature $-K$ one has to study the Schrödinger equation

$$\hat{H}v = -\frac{\hbar^2 K}{2m}\Delta_S v = Ev,$$

where m denotes the mass of the particle. We see that the knowledge of the spectrum of the energy eigenvalues is equivalent to the knowledge of the spectrum of the Laplace-Beltrami-operator Δ_S on the surface S. Explicite expressions for the Laplace-Beltrami-operator on $\mathbb{H}$ resp. $\mathbb{D}$ have been stated already in Section 3.1 above.
In this section we discuss two methods to compute (at least in principle) the eigenvalues of the Laplace-Beltrami-operator on the Riemann surfaces with constant negative curvature, which we have discussed in this section.

[15] The maps T_C, T_R and Φ studied in [74] are essentially our maps τ, f and T. Precisely: τ resp. f are *factors* of T_C resp. T_R.
[16] Of course one can construct topologies on $\mathcal{R}$ and $\mathcal{A}$ such that T is a topological conjugacy in the sense of Definition B.5.

Selberg theory

Let Γ be a Fuchsian group of the first kind which satisfies the conditions 0, 1, 2 and 3 and let $S \simeq \mathbb{H}/\Gamma$ denote the corresponding Riemann surface.

Theorem 3.15 (Selberg) *Let h be a complex-valued function defined on a subset of $\mathbb{C}$ with the following properties:*

- $h(z) = h(-z)$;
- *h is holomorphic in the strip $|\mathrm{Im} z| \leq \frac{1}{2} + \varepsilon$ for some $\varepsilon > 0$;*
- $|h(z)| \leq a(1 + |z|^2)^{-1-\epsilon}$, *in this strip, $a > 0$;*

then (Selberg's trace formula)

$$\sum_{n=0}^{\infty} h(r_n) = \frac{|S|}{2\pi} \int_{-\infty}^{\infty} r h(r) \tanh(\pi r) dr + \sum_{\{\ell_n\}} \sum_{k=1}^{\infty} \frac{g_n \ell_n}{\sinh(\frac{1}{2} k \ell_n)} g(k \ell_n), \tag{71}$$

where $|S|$ denotes the non-Euclidean area of S and $\sum_{\{\ell_n\}}$ denotes the summation over all primitive lengths ℓ_n of periodic geodesics in S with multiplicity g_n; the summation on the left hand side is taken over all solutions of the equations $\frac{1}{4} + r_n^2 = -\lambda_n$, where the λ_n's run through the set of all eigenvalues of the Laplace-Beltrami-operator on S. The function g is defined by $g(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} h(r) e^{-iru} dr$. The integral and the series in Equation 71 converge absolutely.

The energy eigenvalues are determined by $E_n = -\lambda_n = \frac{1}{4} + r_n^2$. There exists a generalization of Selberg's trace formula for Fuchsian groups which do not satisfy the conditions 0, 1, 2, 3.
We see that with the help of Selberg's trace formula it is in principle possible to determine the quantum mechanical energy values by an appropriate choice of the test functions h, provided the classical length spectrum is known. In order to compute the right hand side of Selberg's trace formula, it is necessary to know an algorithm which enables us to compute arbitrarily many lengths of the classical periodic orbits with arbitrarily high accuracy. Such an algorithm has been found for example for the regular octagon, therefore in this example it was possible to compute quantum mechanical energy eigenvalues using Selberg's trace formula, cf. Aurich & Steiner [94, 101, 102, 103] and Aurich, Bogomolny & Steiner [104].

Definition 3.17 *For $s \in \mathbb{C}$, $\mathrm{Re}s > 1$ we define*

$$Z_\Gamma(s) := \prod_{\{\ell_n\}} \prod_{k=0}^{\infty} (1 - e^{-(s+k)\ell_n})^{g_n}. \tag{72}$$

and call Z_Γ SELBERG'S ZETA FUNCTION.

With the help of Theorem 3.15 it is possible to show that the bi-infinite product in Equation 72 converges absolutely for $\mathrm{Re}s > 1$.

Theorem 3.16 • *Z_Γ admits an analytic continuation on the entire complex plane;*

- *Z_Γ has 'trivial' zeros at the points $s = -k$, where $k \in \mathbb{Z}, k \geq 1$. These zeros have multiplicity $(2g-2)(2k+1)$, where g denotes the genus of S;*
- *$s = 1$ is a simple zero and $s = 0$ is a zero with multiplicity $2g - 1$;*

- *The 'non-trivial' zeros of Z_Γ lie at $s_n = \frac{1}{2} \pm i r_n$, where the r_n are defined in Theorem 3.15.*

A proof for the Theorems 3.15 and 3.16 can be found e.g. in Hejhal [105].
The energy eigenvalues are related to the s_n as follows: $E_n = s_n(1 - s_n)$.
Above we have mapped the geodesic flow bijectively to the space $\mathcal{R}$ of cutting sequences. The periodic geodesics in $\mathbb{M} = \mathbb{H}/\Gamma$ have thereby been bijectively mapped to the periodic cutting sequences. The space of the cutting sequences has been mapped in Theorem 3.14 and Corollary 3.2 to a topological Markov shift (Σ_A, σ_A) with the help of the conjugacy $T : \mathcal{R} \to \mathcal{A}$. The equations in Theorem 3.14 and Corollary 3.2 state that the periodic cutting sequences are mapped thereby *bijectively onto the periodic sequences in Σ_A.*
Furthermore, Corollary 3.2 states that we can associate to at most countably many cutting sequences more than one element in Σ_A. The situation is simpler when one considers only periodic cutting sequences. From the proof of Lemma 3.6 it follows that $\tilde{\beta}$ is one-to-one provided none of the endpoints of $\gamma \in \mathcal{A}$ lies in $\overline{W}$. It follows immediately by Lemma 3.2 that for periodic geodesics $\in \mathcal{A}$ the statement $\gamma_\infty \in \overline{W}$ is equivalent to $\gamma_\infty \in W$ and correspondingly that the statement $\gamma_{-\infty} \in \overline{W}$ is equivalent to $\gamma_{-\infty} \in W$. Since W contains only finitely many elements (at least for groups satisfying the conditions 0, 1, 2 and 3), it follows that the restriction of $\tilde{\beta}$ to periodic sequences is only for *finitely* many points not one-to-one. $\tilde{\beta}$ maps these points to geodesics $\in \mathcal{A}$ for which at least one endpoint lies in W.
We consider now an arbitrary geodesic $\gamma \in N(\mathcal{F}) \cap \mathcal{A}$. By Lemma 3.8, there exists an i, with $0 < i \leq k$ and a j, with $0 < j \leq n(v_i) - 2$ such that either $\gamma_{-\infty} = S_{i,j}$ and $\gamma_\infty = T_{i,n(v_i)-2}$ or $\gamma_{-\infty} = P_i$ and $\gamma_\infty = Q_{i+1}$ holds. With the help of Figure 3.1 and Equation 60 we see that the Bowen-Series map on $\mathcal{A}$ (defined by Equation 63) maps the geodesic γ to another geodesic $f(\gamma) \in N(\mathcal{F}) \cap \mathcal{A}$. Inductively, we see: $f^r(\gamma) \in N(\mathcal{F}) \cap \mathcal{A}$ for all $r \geq 0$.
By Corollary 3.2, it follows that the Bowen-Series map f maps the set $\mathcal{A}$ bijectively to $\mathcal{A}$ since the action of f on $\mathcal{A}$ is conjugate to the action of a shift operator on the space of boundary expansions of geodesics in $\mathcal{A}$. Since $N(\mathcal{F}) \cap \mathcal{A}$ is finite, it follows that *the boundary expansion of every geodesic in $N(\mathcal{F}) \cap \mathcal{A}$ is periodic and by Theorem 3.14 that the cutting sequence of every geodesic in $N(\mathcal{F}) \cap \mathcal{R}$ is periodic.*
Suppose that the following assertion holds: given a geodesic $\gamma \in N(\mathcal{F}) \cap \mathcal{A}$, then there exists a further geodesic $\gamma_{period} \in \mathcal{A}$ with endpoint $(\gamma_{period})_\infty = \gamma_\infty \in W$ for which there exists an m_0 such that $f^{m_0}(\gamma_{period}) = h\gamma_{period} = \gamma_{period}$. Especially, $f^{m_0}\gamma_\infty = h\gamma_\infty = \gamma_\infty$ holds. This implies that $f^{m_0}\gamma = h\gamma$ is a geodesic in $N(\mathcal{F}) \cap \mathcal{A}$ with endpoint γ_∞, i.e. $f^{m_0}\gamma = \gamma = \gamma_{period}$.
Therefore we have the following result:
The restriction of $\tilde{\beta}$ to periodic sequences is only for finitely many points not one-to-one. $\tilde{\beta}$ maps these points exactly to the geodesics in $N(\mathcal{F}) \cap \mathcal{A}$ and $T^{-1} \circ \tilde{\beta}$ maps these points exactly to the geodesics in $N(\mathcal{F}) \cap \mathcal{R}$.
Here we have used that T maps the set $N(\mathcal{F}) \cap \mathcal{R}$ bijectively to $N(\mathcal{F}) \cap \mathcal{A}$.
Let now γ be a geodesic in $\mathcal{R}$ whose projection γ_S on $S \simeq \mathbb{M}$ is a periodic geodesic in S. The conjugacy T maps γ to a geodesic $T\gamma$. This geodesic $T\gamma$ has a periodic boundary expansion with the primitive period $m \in \mathbb{N}$. Thus $f^m(T\gamma)_\infty = (T\gamma)_\infty$. f^m is therefore equal to those hyperbolic isometry which leaves $T\gamma$ invariant. Let g_γ be those hyperbolic isometry which leaves γ invariant (cf. Theorem 3.7). Remembering the discussion following Equation 55, we see easily that the length $\ell(\gamma)$ of the projection γ_S of γ on S is given by

$$\ell(\gamma) = +\ln |g_\gamma'(\gamma_\infty)| .$$

Without loss of generality with can always assume that $\ell(\gamma) > 0$: if necessary, we replace g by g^{-1}. The right hand side of this equation is conjugacy invariant. Considering especially the conjugacy T yields $g'(\gamma_\infty) = (T \circ g \circ T^{-1})'((T\gamma)_\infty)$. Recall the discussion following Equation 55 and Theorem 3.7

$$\ell(\gamma) = +\ln |(f^m)'((T\gamma)_\infty)| \,. \tag{73}$$

Since the trace is also conjugacy invariant, it follows at once by Equation 73

$$\ell(\gamma) = 2\text{arcosh}\frac{\text{tr}f^m}{2}. \tag{74}$$

We are now able to give an algorithm which enables us (at least in principle) to compute the lengths of all periodic geodesics:
For all $n \in I\!N$ we construct inductively all finite admissible (i.e. compatible with the matrix A) words of primitive length n in the topological Markov shift Σ_A. That a length n is primitive means here that the word is not an iteration of some shorter word. We associate with every symbol in the alphabet S of Σ_A the corresponding generator; in this way we can construct for every word $(i_0,\dots,i_{n-1})$ the corresponding word in the generators $g_{(i_0,\dots,i_{n-1})} = g_{i_0}\cdots g_{i_{n-1}}$. The isometry $g_{(i_0,\dots,i_{n-1})}$ leaves some geodesic $\gamma_{(i_0,\dots,i_{n-1})} \in \mathcal{A}$ with periodic boundary expansion invariant. The nth power of the Bowen-Series map applied to $(\gamma_{(i_0,\dots,i_{n-1})})_\infty$ equals

$$f^n = g_{i_{n-1}}^{-1}\cdots g_{i_0}^{-1}. \tag{75}$$

Insertion of this expression into one of the Equations 73 or 74 enables us to compute the length of the projection of $T^{-1}\gamma_{(i_0,\dots,i_{n-1})} \in \mathcal{R}$ on $S = I\!M$. Repeating this procedure for all finite admissible words in Σ_A yields successively the lengths of all periodic geodesic in $S = I\!M$. However, as explained above, to get the correct multiplicities of the lengths, we have to be careful with geodesics in $N(\mathcal{F}) \cap \mathcal{A}$. Therefore we take only one representative of the finite words in Σ_A into account which correspond to the same geodesic in $N(\mathcal{F}) \cap \mathcal{A}$. [17] Since σ_A is by Corollary 3.2 conjugate to the action of f on $\mathcal{A}$, it is true that all finite words in Σ_A which correspond to the same geodesic in $N(\mathcal{F}) \cap \mathcal{A}$ have the same primitive wordlength.
The first sum on the right hand side of Equation 71 can now be rewritten as a sum over all finite words in Σ_A:

$$\sum_{n=0}^{\infty} h(r_n) = \frac{|S|}{2\pi}\int_{-\infty}^{\infty} r h(r)\tanh(\pi r)dr + \sum_{\substack{\text{finite prim}\\ \text{words in } \Sigma_A}} \sum_{k=1}^{\infty} \frac{2}{\#}\frac{\text{arcosh}\frac{\text{tr}f^\#}{2}}{\sinh(k\,\text{arcosh}\frac{\text{tr}f^\#}{2})} g\left(2k\,\text{arcosh}\frac{\text{tr}f^\#}{2}\right), \tag{76}$$

where the sum is taken over all finite admissible primitive words in Σ_A and from those finite words in Σ_A which correspond to the same geodesic in $N(\mathcal{F}) \cap \mathcal{A}$ only one representative is included in the sum. The function $\#$ maps every finite word to its length, i.e. $\#(i_0,\dots,i_{n-1}) = n$. The $\#$th power of the Bowen-Series map $f^\#$ corresponding to some word can be obtained as in Equation 75.
The double product in Selberg's zeta function can be rewritten for $\text{Re}\, s > 1$ as follows

$$\begin{aligned} Z_\Gamma(s) &= \prod_{\{\ell_n\}}\prod_{k=0}^{\infty}(1 - e^{-(s+k)\ell_n})^{g_n} \\ &= \prod_{\{\ell_n\}}\prod_{k=0}^{\infty}\exp\left(-\sum_{m=1}^{\infty}\frac{g_n}{m}e^{-(s+k)\ell_n m}\right) \end{aligned}$$

[17] Here we make the convention to identify a finite word with the bi-infinite iteration of this word. Every finite word corresponds to a doubly infinite periodic word in Σ_A. Therefore it makes sense to say that a finite word is contained in Σ_A.

$$\begin{aligned} &= \prod_{k=0}^{\infty}\prod_{\substack{\text{finite prim}\\ \text{words in } \Sigma_A}} \exp\left(\sum_{m=1}^{\infty}\frac{-1}{m\#}|(f^\#)'(\gamma_\infty)|^{-(s+k)m}\right) \\ &= \prod_{k=0}^{\infty}\prod_{\substack{\text{finite prim}\\ \text{words in } \Sigma_A}} \exp\sum_{m=1}^{\infty}\frac{-1}{m\#}\left(\frac{\text{tr}f^\#}{2} + \sqrt{\frac{(\text{tr}f^\#)^2}{4} - 1}\right)^{-2m(s+k)} \\ &= \prod_{k=0}^{\infty}\prod_{\substack{\text{finite words}\\ \text{in } \Sigma_A}} \exp -\frac{1}{\#}\left(\frac{\text{tr}f^\#}{2} + \sqrt{\frac{(\text{tr}f^\#)^2}{4} - 1}\right)^{-2(s+k)} \end{aligned}$$

As above, the function $\#$ maps every finite word to its length. The $\#$th power of the Bowen-Series map $f^\#$ corresponding to some word can be obtained as in Equation 75. The point γ_∞ is the same as in Equation 73.
The product in the last line is taken over all finite, admissible (but not necessarily primitive) words in Σ_A and from those finite words with fixed wordlength in Σ_A corresponding to the same geodesic $N(\mathcal{F}) \cap \mathcal{A}$ only one is included in the product, i.e. multiple traversals of a periodic geodesic are included in the product. For instance, given a geodesic $\gamma \in N(\mathcal{F}) \cap \mathcal{A}$ whose associated primitive word w_γ has length m_γ, then a possible choice is to include the words w_γ, $w_\gamma w_\gamma$, $w_\gamma w_\gamma w_\gamma$, ... etc. of wordlengths $m_\gamma, 2m_\gamma, 3m_\gamma, \dots$ etc. in the product. The product in the third and fourth lines is taken over all finite admissible primitive words in Σ_A and from those finite words in Σ_A corresponding to the same geodesic in $N(\mathcal{F}) \cap \mathcal{A}$ only one is included in the sum.

Finally, notice that for a given Fuchsian group of the first kind Γ it is possible that in general there are several fundamental domains $\mathcal{F}_1, \mathcal{F}_2, \dots$ such that Γ with every of these fundamental domains satisfies the conditions 0, 1, 2 and 3 stated in Section 3.2. The Markov coding of geodesics in $S = I\!D/\Gamma$ described in this section can be achieved for every allowed choice of the fundamental domain. However, coding with respect to different fundamental domains results in general in different topological Markov shifts (cf. Corollary 3.2). Especially, the number of symbols in the alphabets of the topological Markov shifts may not be equal.
For example, Markov coding of geodesics relative to the standard fundamental domain of the octagon group G (i.e. the regular octagon, cf. [94, 102]) yields a topological Markov shift with an alphabet consisting of 48 symbols. However, in Adler & Flatto [74], Theorem 3.1, there is a proof that for a Fuchsian group of the first kind satisfying conditions 2 and 3 it is always possible to find a fundamental domain $\mathcal{F}_0$ with $8g-4$ sides such that also the conditions 0 and 1 are satisfied. The topological Markov shift in Corollary 3.2 has then an alphabet consisting of $16g-8$ elements. This is equivalent to the statement that only two geodesics in N pass through every vertex of $\mathcal{F}_0$. Here g denotes the genus of the Riemann surface $S = I\!D/\Gamma$. For the regular octagon is $g = 2$ and therefore it is possible to find a Markov coding of the geodesics in $I\!D/G$ with 24 symbols.
We see that it may be advantageous in numerical studies to consider fundamental domains with as few sides as possible.

The transfer matrix method

We discuss our second method to compute the eigenvalues of the Laplace-Beltrami-operator on S only qualitatively. The issue of convergence and mathematical subleties are not considered here. The method is based on a lemma of Pollicott [70]. In the following lemma $\mathcal{L}_s$ denotes the transfer operator with respect to the function $\varphi_s = -s \ln |f'|$.

Lemma 3.11 *Let $P = P(-\mathrm{Re}(s) \ln |f'|)$ and let k be so large that $\lambda^k > e^P$ holds, where λ denotes the expansive constant from Theorem 3.10, then $\alpha = s(1-s)$ is an eigenvalue of the Laplace-Beltrami-operator $-\Delta_S$ on S if and only if 1 is an eigenvalue of the transfer operator $\mathcal{L}_s : C^k(\partial I\!D) \to C^k(\partial I\!D)$.*

For the definition of the transfer operator, see Section 2.5. A proof of Lemma 3.11 can be found in Pollicott [70] (This work contains, however, awfully many misprints).
We have changed the domain of the transfer operator somewhat in this lemma: notice, however, that by Corollary 3.2 $\Sigma_A^+ \simeq \partial I\!D$ holds. We once again see the exceptional role of the Hölder function $z \mapsto -s \ln |f'(z)|$.
A method to compute the eigenvalues of the Laplace-Beltrami operator based on Lemma 3.11 has already successfully been used by Bogomolny and Carioli [106].
We rewrite the eigenvalue equation $\mathcal{L}_s h = h$:

$$h(z) = \sum_{y \in f^{-1} z} \frac{h(y)}{|f'(y)|^s}.$$

An iterative procedure suggests itself. We write the equation for h iterative:

$$h_{n+1}(z) = \sum_{y \in f^{-1} z} \frac{h_n(y)}{|f'(y)|^s}. \tag{77}$$

In practice one starts with some appropriate start function h_0 and determines those values of s for which the iteration 77 converges.
For simple functions $f : [0,1] \to [0,1]$ this method has already been used for real s to compute the Gibbs potential $P(q,\tau)$ from Equation 30 with the help of Theorem 2.11, see Feigenbaum, Procaccia & Tél [67] and Kovács & Tél [107]. In case of the Bowen-Series map, the Gibbs potential $P(q,\tau)$ coincides with the topological pressure for a special choice of the Hölder continuous function φ as we will see below. Hopefully, this method is also useful for the numerical computation of energy eigenvalues.

4 Multifractal formalism for the Bowen-Series map

4.1 Preliminaries

As above, let Γ be a Fuchsian group of the first kind which satisfies the conditions 0, 1, 2 und 3 from Section 3.2. The Bowen-Series map associated with Γ is denoted by f. The set $\mathcal{A}$ was defined by *Equation 64 and the action of f on $\mathcal{A}$ is given by Equation 63.*
We have seen in Corollary 3.2 that the conjugacy $\tilde{\beta} : \Sigma_A \to \mathcal{A}$ maps the dynamical system $(\mathcal{A}, f)$ to a topologically mixing topological Markov chain (Σ_A, σ_A). The conjugacy $\tilde{\beta}$ is one-to-one, except for a countable set of points, and onto. If we endow Σ_A with the topology induced by one of the metrics d_ρ, with $0 < \rho < 1$, given by Eq. 106 and identify a geodesic $\gamma \in \mathcal{A}$ with the pair of its endpoints $(\gamma_\infty, \gamma_{-\infty}) \in [0, 2\pi]^2$ and endow $\mathcal{A}$ with the topology induced by $I\!R^2$, then it follows from Corollary 3.2 that $\tilde{\beta}$ is continuous. $\tilde{\beta}$ is not a topological conjugacy in the sense of Definition B.5. In the following we will always identify $\mathcal{A}$ with $[0, 2\pi]^2$.
The aim of this section is to apply the heuristical thermodynamical formalism for multifractals studied in the Sections 1.3 and 1.4 to the Bowen-Series map on $\mathcal{A}$ and to investigate its relation

to the mathematical rigorous thermodynamical formalism described in Section 2.
In Section 1 many objects have been defined in terms of infima over certain coverings (see Equations 5, 6 and 12). For the dynamical system $(\mathcal{A}, f)$ considered here, we have specified the symbolic dynamics in the last section. This symbolic dynamics enables us to define in a canonical way a hierarchical sequence of coverings: First of all, for all $m, n \in I\!N$ we denote by $K(g_{i_{-n}} g_{i_{-n+1}} \cdots g_{i_m})$ the set of all $\gamma \in \mathcal{A}$ whose boundary expansion $(\gamma_\infty)_f$ begins with the $m+1$ generators $g_{i_0} \cdots g_{i_m}$ *and* whose backward boundary expansion $(\gamma_{-\infty})_f$ begins with the n generators $g_{i_{-1}}^{-1} \cdots g_{i_{-n}}^{-1}$:

$$K(g_{i_{-n}} g_{i_{-n+1}} \cdots g_{i_m}) := \{\gamma \in \mathcal{A} \mid (\gamma_\infty)_f \text{ begins with } g_{i_0} g_{i_1} \cdots g_{i_m} \text{ and } (\gamma_{-\infty})_f \text{ begins with } g_{i_{-1}}^{-1} \cdots g_{i_{-n}}^{-1}\}.$$

We will call the sets $K(g_{i_{-n}} g_{i_{-n+1}} \cdots g_{i_{n-1}})$ also n-CYLINDERS.
In this way a hierarchical sequence $(\mathcal{U}_n)_{n \in I\!N}$ of coverings of $\mathcal{A}$ consisting of n-cylinders is obtained. We will see that those objects which have been defined in Section 1 with the help of arbitrary coverings corresponds to objects in this section which we will define below with the help of the covering $\mathcal{U}_n$. The details are given below (cf. Lemma 4.4, Theorem 4.1 and Definition 4.1).
We will not carry out the multifractal analysis for arbitrary measures on $\mathcal{A}$; we rather will restrict ourselves to equilibrium states with respect to certain functions. Since $\tilde{\beta}$ is not a topological conjugacy in the sense of Definition B.5, it is a priori not clear whether Theorem 2.10 is valid for the dynamical systems (Σ_A, σ_A) and $(\mathcal{A}, f)$.
We define in analogy to Definition 2.7 for $\psi \in \mathcal{C}(\mathcal{A})$

$$\mathrm{var}_k \psi := \sup \{|\psi(\gamma_1) - \psi(\gamma_2)|\},$$

where the supremum is over all $\gamma_1, \gamma_2 \in \mathcal{A}$ which satisfy: 1. The boundary expansions of $(\gamma_1)_\infty$ and of $(\gamma_2)_\infty$ coincide in the first $k+1$ generators. 2. The backward boundary expansions of $(\gamma_1)_{-\infty}$ and $(\gamma_2)_{-\infty}$ coincide in the first k generators.

We denote with F_A those $\psi \in \mathcal{C}(\mathcal{A})$ for which constants $c > 0$ and $\epsilon \in]0,1[$ exist such that $\mathrm{var}_k \psi \le c\epsilon^k$ for all $k \ge 0$.
Pick $\psi \in \mathsf{F}_A$, then $\psi \circ \tilde{\beta} \in \mathcal{C}(\Sigma_A)$ and obviously $\psi \circ \tilde{\beta} \in \mathcal{F}_A$. The set $\mathcal{F}_A$ was defined in Section 2, *cf. Definition 2.7. Furthermore, we define κ as follows*

$$\kappa : \mathcal{A} \to \Sigma_A, \kappa(\gamma) := \gamma_{-\infty} * \gamma_\infty. \tag{78}$$

κ maps every geodesic in $\mathcal{A}$ to its boundary expansion (cf. Equation 62). κ is in general not continuous.

We define $\tilde{\Sigma}_A := \kappa(\mathcal{A})$ and denote with $\tilde{\sigma}_A$ the restriction of the shift map to $\tilde{\Sigma}_A$ and with $\tilde{\varphi}$ the restriction of $\varphi \in \mathcal{F}_A$ to $\tilde{\Sigma}_A$. κ is the inverse function of the restriction of $\tilde{\beta}$ to $\tilde{\Sigma}_A$, that is

$$\tilde{\beta} \circ \kappa = id_{\mathcal{A}} \text{ and } \kappa \circ \tilde{\beta}|_{\tilde{\Sigma}_A} = id_{\tilde{\Sigma}_A}.$$

The map $\tilde{\beta}$ is a continuous bijection $\tilde{\Sigma}_A \to \mathcal{A}$ and it follows from Corollary 3.2 that $(\tilde{\Sigma}_A, \tilde{\sigma}_A)$ is a symbolic dynamical system.
For the following lemma it is useful to consider instead of F_A the smaller set $\tilde{\mathsf{F}}_A$, where $\tilde{\mathsf{F}}_A$ is defined as the set of all those $\psi \in \mathsf{F}_A$ for which the - according to Lemma 2.8 uniquely determined - *equilibrium state $\mu_{\psi \circ \tilde{\beta}}$ with respect to $\psi \circ \tilde{\beta}$ on Σ_A satisfies the condition $\mu_{\psi \circ \tilde{\beta}}(\tilde{\Sigma}_A) = 1$.* By Lemma 2.8, we know that the equilibrium state $\mu_{\psi \circ \tilde{\beta}}$ with respect to $\psi \circ \tilde{\beta}$ on Σ_A is an ergodic

measure on Σ_A for all $\psi \in \mathsf{F}_A$. Since $\tilde{\Sigma}_A$ is invariant under σ_A, we see that the restriction to $\tilde{\mathsf{F}}_A$ is equivalent to considering only those equilibrium states μ_{eq} on Σ_A which satisfy $\mu_{eq}(\Sigma_A \backslash \tilde{\Sigma}_A) = 0$.
If $\psi \in \mathsf{F}_A \backslash \tilde{\mathsf{F}}_A$, then the equilibrium state $\mu_{\psi\circ\tilde{\beta}}$ is an ergodic measure on Σ_A and it exists a σ_A-invariant subset $\Sigma^* \subset \Sigma_A \backslash \tilde{\Sigma}_A$ such that $\mu_{\psi\circ\tilde{\beta}}(\Sigma^*) = 1$. We choose a local inverse $\tilde{\beta}_*^{-1}$ of $\tilde{\beta}$ such that $\Sigma^* \subset \tilde{\beta}_*^{-1}\mathcal{A}$. Then $\mu_{\psi\circ\tilde{\beta}} \circ \tilde{\beta}_*^{-1} \circ \tilde{\beta}|_{\tilde{\Sigma}_A}$ is an ergodic measure on $\tilde{\Sigma}_A$. By Lemma 2.8, there exists a $\tilde{\psi} \in \tilde{\mathsf{F}}_A$ such that $\mu_{\psi\circ\tilde{\beta}} \circ \tilde{\beta}_*^{-1} = \mu_{\tilde{\psi}\circ\tilde{\beta}} \circ \kappa$. We are solely interested in ergodic measures on $\mathcal{A}$, so by Lemma 4.1, we are allowed to consider instead of F_A the smaller set $\tilde{\mathsf{F}}_A$.

Lemma 4.1 *Let $\psi \in \tilde{\mathsf{F}}_A$ and $\mu_{\psi\circ\tilde{\beta}}$ be the - by Lemma 2.8 uniquely determined - equilibrium state with respect to $\psi\circ\tilde{\beta} \in \mathcal{F}_A$ on Σ_A. Then $P(\psi) = P(\psi\circ\tilde{\beta})$ and $h_{(\mu_{\psi\circ\tilde{\beta}})\circ\kappa}(f) = h_{\mu_{\psi\circ\tilde{\beta}}}(\sigma_A)$. Especially it follows that $\mu_{\psi\circ\tilde{\beta}} \circ \kappa$ is the uniquely determined equilibrium state with respect to ψ on $\mathcal{A}$.*

Proof: We identify $\mathcal{A}$ with $[0, 2\pi]^2$ and denote with $\mathcal{B}(\mathcal{A})$ the σ-algebra of subsets of $[0, 2\pi]^2$ generated by the open subsets of $[0, 2\pi]^2$ (cf. Appendix A). From an easy computation it follows that also $\Upsilon := \kappa(\mathcal{B}(\mathcal{A}))$ is a σ-algebra of subsets of $\tilde{\Sigma}_A$. Since κ is open, it follows that Υ is contained in the Borel σ-algebra $\mathcal{B}(\tilde{\Sigma}_A)$ of $\tilde{\Sigma}_A$. It is straightforward to check that $\mu_{\psi\circ\tilde{\beta}} \circ \kappa$ is a Borel probability measure on $\mathcal{A}$.
The countable set $\Sigma_A \backslash \tilde{\Sigma}_A$ satisfies $\mu_{\psi\circ\tilde{\beta}}\left(\Sigma_A \backslash \tilde{\Sigma}_A\right) = 0$ since $\psi \in \tilde{\mathsf{F}}_A$. Therefore the dynamical systems $(\Sigma_A, \sigma_A), (\tilde{\Sigma}_A, \tilde{\sigma}_A)$ and $(\mathcal{A}, f)$ are pairwise isomorphic in the sense of Definition B.2. By Theorem 2.2, it follows that $h_{\mu_{\psi\circ\tilde{\beta}}}(\sigma_A) = h_{(\mu_{\psi\circ\tilde{\beta}})\circ\kappa}(f)$.
Since furthermore,

$$\int_{\Sigma_A} \psi \circ \tilde{\beta} d\mu_{\psi\circ\tilde{\beta}} = \int_{\tilde{\Sigma}_A} \psi \circ \tilde{\beta} d\mu_{\psi\circ\tilde{\beta}} = \int_{\tilde{\Sigma}_A} \widetilde{(\psi \circ \tilde{\beta})} d\mu_{\psi\circ\tilde{\beta}} = \int_{\mathcal{A}} \psi d(\mu_{\psi\circ\tilde{\beta}} \circ \kappa),$$

the variational principle Theorem 2.7 implies that $P_f(\psi) = P_{\sigma_A}(\psi \circ \tilde{\beta})$.
Thus

$$\begin{aligned} P_f(\psi) &= P_{\sigma_A}(\psi \circ \tilde{\beta}) \\ &= h_{\mu_{\psi\circ\tilde{\beta}}}(\sigma_A) + \int_{\tilde{\Sigma}_A} \psi \circ \tilde{\beta} d\mu_{\psi\circ\tilde{\beta}} \\ &= h_{(\mu_{\psi\circ\tilde{\beta}})\circ\kappa}(f) + \int_{\mathcal{A}} \psi d(\mu_{\psi\circ\tilde{\beta}} \circ \kappa). \end{aligned}$$

So we have shown that $(\mu_{\psi\circ\tilde{\beta}}) \circ \kappa$ is an equilibrium state with respect to ψ on $\mathcal{A}$.
It remains to show that this state is the *unique* equilibrium state with respect to ψ. Let $\nu \in C(\mathcal{A})$ be a further equilibrium state with respect to ψ, then

$$\begin{aligned} P_{\sigma_A}(\psi \circ \tilde{\beta}) &= P_f(\psi) \\ &= h_\nu(f) + \int_{\mathcal{A}} \psi d\nu \\ &= h_{\nu\circ\tilde{\beta}}(\sigma_A) + \int_{\Sigma_A} \psi \circ \tilde{\beta} d(\nu \circ \tilde{\beta}) \end{aligned}$$

The last equation is valid since (Σ_A, σ_A) is isomorphic to $(\mathcal{A}, f)$ with isomorphism $\tilde{\beta}$. Therefore $\nu \circ \tilde{\beta}$ is a equilibrium state with respect to $\psi \circ \tilde{\beta}$, i.e. $\nu \circ \tilde{\beta} = \mu_{\psi\circ\tilde{\beta}}$ since the equilibrium state with respect to $\psi \circ \tilde{\beta}$ is uniquely determined by Lemma 2.8. □

A lemma analogous to Lemma 4.1 can equally well be formulated if we replace κ by any other local inverse $\tilde{\beta}_*^{-1} : \mathcal{A} \to \Sigma_A$ of $\tilde{\beta}$.
As already indicated above, we will study in this section the relation between the multifractal formalism for the dynamical system $(\mathcal{A}, f)$ and the thermodynamical formalism developed in Section 2. However, we can do this only for equilibrium states with respect to functions in $\tilde{\mathsf{F}}_A$.
In the following let ν be the unique equilibrium state on $\mathcal{A}$ with respect to the function $\chi \in \tilde{\mathsf{F}}_A$. Choose χ such that $P(\chi) = 0$ (this choice is possible by Lemma 2.9 and Lemma 4.1).

The function $z \mapsto -\ln |f'(z)|$

We define two functions

$$\phi_1 : \mathcal{A} \to I\!R \quad , \quad \phi_1(\gamma) := -\ln |f'(\gamma_\infty)|, \tag{79}$$
$$\varphi_1 : \Sigma_A \to I\!R \quad , \quad \varphi_1(\omega) := -\ln |f'(\tilde{\beta}(\omega))|. \tag{80}$$

From the Definition 3.13 of the Bowen-Series map it follows immediately that the function ϕ_1 is not continuous at the points P_i defined in Section 3.4. Furthermore, it is immediate to see that $\phi_1(\gamma)$ depends only on the positive endpoint γ_∞ of the geodesic γ. Correspondingly, $\varphi_1(\omega)$ depends only on $\{\omega_i\}_{i\geq 0}$.

Lemma 4.2 *There exists an $\alpha \in]0, 1[$ and a $d > 0$ such that $|\phi_1(\gamma_1) - \phi_1(\gamma_2)| \leq d\alpha^n$ if $\gamma_1, \gamma_2 \in I(g_{i_0} g_{i_1} \cdots g_{i_n})$.*

Proof: In Lemma 3.3 we have already seen that there exist constants $\alpha \in]0, 1[$ and $C > 0$ such that

$$\ell(I(g_{i_0} g_{i_1} \cdots g_{i_n})) < C\alpha^n,$$

where ℓ denotes the Euclidean length on the unit circle. By Theorem 3.9 (M1), the function ϕ_1 is continuously differentiable on every $I_j \in \mathcal{P}$, where $\mathcal{P}$ denotes the finite covering of $\partial I\!D$ from Theorem 3.9, and therefore ϕ_1 is Lipschitz on every I_j. From this the assertion of the lemma follows immediately. □

Corollary 4.1 *There exists an $\alpha \in]0, 1[$ and a $d > 0$ such that $|\varphi_1(\omega) - \varphi_1(\omega')| \leq d\alpha^n$ if $\omega_i = \omega'_i$ for all $0 \leq i \leq n$.*

It follows from Corollary 4.1 that $\varphi_1 \in \mathcal{F}_A$ is a continuous function provided we endow Σ_A with the usual topology (cf. Equation 105). Nevertheless $\phi_1 = \varphi_1 \circ \kappa \notin \mathsf{F}_A$. This reflects the fact that κ is not continuous.
Let $\psi \in \tilde{\mathsf{F}}_A$, then we consider the functions $\tau\phi_1 + \psi$ and $\tau\varphi_1 + \psi \circ \tilde{\beta}$, where $\tau \in I\!R$. The statements analogous to Lemma 4.2 and Corollary 4.1 are also valid. Thus $\tau\varphi_1 + \psi \circ \tilde{\beta} \in \mathcal{F}_A$. We will denote the unique equilibrium state on Σ_A with respect to $\tau\varphi_1 + \psi \circ \tilde{\beta}$ with $\mu_{\tau\varphi_1+\psi\circ\tilde{\beta}}$. In the following we assume that $\mu_{\varphi_1}(\tilde{\Sigma}_A) = 1$: Otherwise κ has to be replaced by an appropriate $\tilde{\beta}_*^{-1}$ as explained above. Then the same reasoning as in the proof of Lemma 4.1 gives:

1. $\mu_{\tau\varphi_1+\psi\circ\tilde{\beta}} \circ \kappa$ is a Borel probability measure on $\mathcal{A}$;

2. $h_{\mu_{\tau\varphi_1+\psi\circ\tilde{\beta}}}(\sigma_A) = h_{\mu_{\tau\varphi_1+\psi\circ\tilde{\beta}}\circ\kappa}(f)$;

3. $\int_{\Sigma_A}(\tau\varphi_1 + \psi \circ \tilde{\beta})d\mu_{\tau\varphi_1+\psi\circ\tilde{\beta}} = \int_{\mathcal{A}}(\tau\phi_1 + \psi)d\left(\mu_{\tau\varphi_1+\psi\circ\tilde{\beta}} \circ \kappa\right)$.

The topological pressure of $\tau\phi_1 + \psi$ is not defined since $\tau\phi_1 + \psi$ is not continuous. Therefore we define

$$P_f(\tau\phi_1 + \psi) := P_{\sigma_A}\left(\tau\varphi_1 + \psi \circ \tilde{\beta}\right). \quad (81)$$

It is easy to see that the so defined topological pressure $P_f(\tau\phi_1 + \psi)$ can equally well be defined through a limit as in Theorem 2.5 in Section 2.3.
From this it follows that $\mu_{\tau\varphi_1+\psi\circ\tilde{\beta}} \circ \kappa$ is the unique equilibrium state with respect to $\tau\phi_1 + \psi$ on $\mathcal{A}$ (uniqueness can be shown as in the proof of Lemma 4.1).

Lemma 4.3 *There exists a constant $D > 0$ such that*

$$\ell(I(\omega_0\omega_1...\omega_n)) \in [D^{-1}, D]\exp\left(\sum_{j=0}^{n-1}\varphi_1(\sigma_A^j\omega)\right)$$

for all $\omega \in \Sigma_A$.

Proof: [108] Choose $\omega \in \Sigma_A$ and $\gamma_\omega := \tilde{\beta}(\omega)$. For every geodesic ϱ with endpoint $\varrho_\infty \in I(\omega_0\omega_1...\omega_n)$, there exists a $\rho \in \Sigma_A$ with $\tilde{\beta}(\rho) = \varrho$ and $\rho_i = \omega_i$ for all $i \le n$. Thus by Corollary 3.2 and Equation 80,

$$\begin{aligned} |(f^n)'(\varrho_\infty)| &= \prod_{j=0}^{n-1} |f'(\tilde{\beta}(\sigma_A^j\rho))| \\ &= \exp\left(-\sum_{j=0}^{n-1}\varphi_1(\sigma_A^j\rho)\right). \end{aligned}$$

By Lemma 2.7, it follows:

$$\left|\sum_{j=0}^{n-1}\varphi_1(\sigma_A^j\omega) - \sum_{j=0}^{n-1}\varphi_1(\sigma_A^j\rho)\right| \le C$$

since $\varphi_1 \in \mathcal{F}_A$. Therefore $|(f^n)'(\varrho_\infty)|$ differs from $\exp\left(-\sum_{j=0}^{n-1}\varphi_1(\sigma_A^j\omega)\right)$ at most by a bounded multiplicative factor. Therefore there exists a constant $A > 0$ such that

$$\begin{aligned} \ell(I(\omega_n)) = \ell(f^n I(\omega_0...\omega_n)) &= \int_{I(\omega_0...\omega_n)} |(f^n)'(\varrho_\infty)|\, d\varrho_\infty \\ &\in [A^{-1}, A]\exp\left(-\sum_{j=0}^{n-1}\varphi_1(\sigma_A^j\omega)\right)\ell(I(\omega_0...\omega_n)) \end{aligned}$$

and the assertions follows. □

4.2 The multifractal spectrum of the Bowen-Series map

We have ν defined to be the - by Lemma 4.1 unique - equilibrium state with respect to $\chi \in \tilde{\mathsf{F}}_A$. χ has been choosen in such a way that $\chi(\gamma)$ depends only on the positive endpoint γ_∞ of the geodesic $\gamma \in \mathcal{A}$ and that furthermore $P(\chi) = 0$ is satisfied (this is possible by Lemma 2.9, Lemma

2.10 and Lemma 4.1).
To perform the multifractal analysis for the measure ν, we consider the functions

$$\begin{aligned} \phi_{q,\tau} &:= -\tau\phi_1 + q\chi; & (82)\\ \varphi_{q,\tau} &:= -\tau\varphi_1 + q\left(\chi \circ \tilde{\beta}\right), & (83) \end{aligned}$$

where $q, \tau \in \mathbb{R}$. We have seen in the previous section that $\phi_{q,\tau} \notin \mathsf{F}_A$ since ϕ_1 is not continuous. Nonevertheless, we have been able to show that $\mu_{q,\tau} := \mu_{\varphi_{q,\tau}} \circ \kappa$ is the unique equilibrium state with respect to $\phi_{q,\tau}$. The topological pressure of $\phi_{q,\tau}$ has been defined by $P(q,\tau) := P_f(\phi_{q,\tau}) := P_{\sigma_A}(\varphi_{q,\tau})$.

Lemma 4.4 *Let $\mathcal{Z}_n$ denote the set of all n-cylinders of the form $I(g_{i_0}...g_{i_n})$. The topological pressure of $\phi_{q,\tau}$ is given by the limit*

$$P_f(\phi_{q,\tau}) = \lim_{n\to\infty}\frac{1}{n}\ln\sum_{C\in\mathcal{Z}_n}\nu(C)^q\ell(C)^{-\tau}.$$

Proof: By Lemma 4.1 and Theorem 2.9, there exists a $B > 0$ such that

$$\mu_{q,\tau}(I(\omega_0...\omega_n)) \in [B^{-1}, B]\exp\left(-nP(q,\tau) + \sum_{k=0}^{n-1}\varphi_{q,\tau}(\sigma_A^k\omega)\right) \quad (84)$$

for all $\omega \in \Sigma_A$ and $n \ge 0$. Summing over all n-cylinders gives

$$1 = \sum_{C\in\mathcal{Z}_n}\mu_{q,\tau}(C) \in [B^{-1}, B]\sum_{C\in\mathcal{Z}_n}\exp\left(-nP(q,\tau) + \mathsf{S}_n\varphi_{q,\tau}(C)\right), \quad (85)$$

where the abbreviation $\mathsf{S}_n\varphi_{q,\tau}(C) := \sup_{\omega\in C}\left\{\sum_{k=0}^{m-1}\varphi_{q,\tau}(\sigma_A^k\omega)\right\}$ was used. From this it follows immediately

$$P(q,\tau) = \lim_{n\to\infty}\frac{1}{n}\ln\sum_{C\in\mathcal{Z}_n}\exp\left(\mathsf{S}_n\varphi_{q,\tau}(C)\right). \quad (86)$$

Since ν is also an equilibrium state with respect to χ, by Theorem 2.9, Lemma 2.7, Lemma 4.3 and Lemma 4.1, there exists a constant $A > 0$ such that for every n-cylinder $C \in \mathcal{Z}_n$

$$\exp\left(\mathsf{S}_n\varphi_{q,\tau}(C)\right) \in [A^{-1}, A]\nu(C)^q\ell(C)^{-\tau} \quad (87)$$

holds. Insertion in Equation 85 gives

$$1 = \sum_{C\in\mathcal{Z}_n}\mu_{q,\tau}(C) \in [B^{-1}, B]\sum_{C\in\mathcal{Z}_n}\exp\left(-nP(q,\tau)\right)\nu(C)^q\ell(C)^{-\tau}. \quad (88)$$

The assertion follows immediately. □

Remark: In the theory of chaotic dynamical systems sometimes a topological pressure depending on one real variable is considered. This topological pressure can be obtained from $P(q,\tau)$ as follows

$$P(\tau) := P(0,\tau)$$

(cf. [7, 66, 67, 68, 70]). Compare with the remark on page 21.

The topological pressure $P(q,\tau) = P(-\tau\varphi_1 + q(\chi \circ \tilde{\beta}))$ is infinitely differentiable with respect to

q and τ by Theorem 2.8.
It follows from the variational principle Theorem 2.7 and Lemmata 2.8 and 4.1 that $\mu_{q,\tau}$ is the unique f-invariant Borel probability measure on $\mathcal{A}$ which satisfies

$$P(q,\tau) = \sup_{\mu \in M_f(\mathcal{A})} \left(h_\mu(f) + \int \phi_{q,\tau} d\mu \right) = h_{\mu_{q,\tau}}(f) + \int \phi_{q,\tau} d\mu_{q,\tau}.$$

The first partial derivative of $P(q,\tau)$ with respect to τ is therefore

$$\begin{aligned} \frac{\partial P(q,\tau)}{\partial \tau} &= -\int \phi_1 d\mu_{q,\tau} + \frac{\partial}{\partial \tau'} \left(h_{\mu_{q,\tau'}}(f) + q \int \chi d\mu_{q,\tau'} \right)\Big|_{\tau'=\tau} - \tau \frac{\partial}{\partial \tau'} \left(\int \phi_1 d\mu_{q,\tau'} \right)\Big|_{\tau'=\tau} \\ &= -\int \phi_1 d\mu_{q,\tau} + \frac{\partial}{\partial \tau'} \left(h_{\mu_{q,\tau'}}(f) + \int \phi_{q,\tau} d\mu_{q,\tau'} \right)\Big|_{\tau'=\tau}. \end{aligned} \tag{89}$$

The second term in the second equation corresponds to a variation (in τ) of $\mu_{q,\tau}$ with $\phi_{q,\tau}$ held *fixed*. Since $\mu_{q,\tau}$ is an equilibrium state with respect to $\phi_{q,\tau}$, this term vanishes and we obtain

$$\frac{\partial P(q,\tau)}{\partial \tau} = -\int_{\mathcal{A}} \phi_1 d\mu_{q,\tau}. \tag{90}$$

In the same way it can be shown that

$$\frac{\partial P(q,\tau)}{\partial q} = \int_{\mathcal{A}} \chi d\mu_{q,\tau}. \tag{91}$$

$\mu_{q,\tau}$ is an ergodic measure on $\mathcal{A}$ by virtue of Theorem B.2 and Lemmata 4.1 and 2.8 and therefore it follows from Birkhoffs ergodic Theorem B.1 that for $\mu_{q,\tau}$-almost all $\gamma \in \mathcal{A}$:

$$\frac{\partial P(q,\tau)}{\partial \tau} = \lim_{m\to\infty} \frac{1}{m} \sum_{k=0}^{m-1} \ln |f'(f^k\gamma_\infty)| = \lim_{m\to\infty} \frac{1}{m} \ln |(f^m)'(\gamma_\infty)|. \tag{92}$$

If m equals N from Theorem 3.10 or is an integer multiple of N, i.e. $m = nN, n \in I\!N$, then by Theorem 3.10. $|(f^m)'(\gamma_\infty)| \geq \lambda^n > 1$. Thus

$$\frac{\partial P(q,\tau)}{\partial \tau} > 0. \tag{93}$$

In the same way it follows for $\mu_{q,\tau}$-almost all $\gamma \in \mathcal{A}$:

$$\frac{\partial P(q,\tau)}{\partial q} = \lim_{m\to\infty} \frac{1}{m+1} \sum_{k=0}^{m} \chi(f^k\gamma_\infty). \tag{94}$$

We insert $\tau = 0$ and $q = 1$ in Equation 84 and take the natural logarithm of both sides. It follows for all $\omega \in \Sigma_A$ and $n \in I\!N$

$$\begin{aligned} P(\chi) &\geq -\frac{\ln B}{n} - \frac{\ln \mu_{1,0}(I(\omega_0...\omega_n))}{n} + \frac{1}{n}\sum_{k=0}^{n-1} \chi(\beta(\sigma_A^k\omega)) \\ &\geq -\frac{\ln B}{n} + \frac{1}{n}\sum_{k=0}^{n-1} \chi(\beta(\sigma_A^k\omega)). \end{aligned}$$

where we have noticed that $\ln \mu_{1,0}(I(\omega_0...\omega_n)) < 0$. Thus: $0 = P(\chi) \geq \lim_{m\to\infty} \frac{1}{m}\sum_{k=0}^{m-1} \chi(f^k\gamma_\infty)$ for all $\gamma \in \mathcal{A}$. From this it follows immediately

$$\frac{\partial P(q,\tau)}{\partial q} \leq 0. \tag{95}$$

By the Implicite Function Theorem, there exists a real-analytic (infinitely differentiable) function $\tau(q)$ which is defined by $P(q,\tau(q)) = 0$.

Theorem 4.1 *$\tau(q)$ is given by*

$$\sum_{C\in Z_n} \nu(C)^q \ell(C)^{-\tau} \xrightarrow{n\to\infty} \begin{cases} \infty & : \ \tau > \tau(q) \\ 0 & : \ \tau < \tau(q) \end{cases}.$$

Proof: This follows immediately from Equation 88 by noticing that $P(q,\tau)$ is monotonically increasing in τ for fixed q (cf. Equation 93). □

Theorem 4.2 *$\tau(q) \leq (q-1)\dim_H^q(\partial I\!D)$ and $\tau(0) = -\dim_H(\partial I\!D) = -1$.*

Proof: It follows immediately from Theorem 4.1 compared with the definition of the q-Hausdorff dimensions that $(q-1)\dim_H^q(\partial I\!D) \geq \tau(q)$. The relation $\tau(0) = -\dim_H(\partial I\!D)$ has been proven for the first time by Bowen [108]. Bowen's proof has also been given by Ruelle [109]. Bowen's proof is devided into two parts in which the relations $\tau(0) \leq -\dim_H(\partial I\!D)$ and $\tau(0) \geq -\dim_H(\partial I\!D)$ are proved respectively. The proof of the first inequality can easily be generalized to the case $q \neq 0$ yielding in this way a further proof of the relation $\tau(q) \leq (q-1)\dim_H^q(\partial I\!D)$. However, the proof of the second inequality can not be generalized to the case $q \in I\!R$ and therefore it remains open whether the second inequality is valid for arbitrary $q \in I\!R$. □
Now we define the Legendre-transform of $\tau(q)$:

$$\mathfrak{f}(\alpha(q)) := q\alpha(q) - \tau(q), \tag{96}$$

where

$$\alpha(q) := \tau'(q).$$

We call $\mathfrak{f}$ the MULTIFRACTAL SPECTRUM of the Bowen-Series map. Differentiation of the equation $P(q,\tau(q)) = 0$ with respect to q yields:

$$\tau'(q) = -\frac{\frac{\partial P(q,\tau)}{\partial q}}{\frac{\partial P(q,\tau)}{\partial \tau}} = \alpha(q) \geq 0.$$

A second differentiation yields

$$\frac{d^2\tau(q)}{dq^2} = \frac{2\frac{\partial P}{\partial \tau}\frac{\partial P}{\partial q}\frac{\partial^2 P}{\partial q \partial \tau} - \left(\frac{\partial P}{\partial \tau}\right)^2 \frac{\partial^2 P}{\partial q^2} - \left(\frac{\partial P}{\partial q}\right)^2 \frac{\partial^2 P}{\partial \tau^2}}{\left(\frac{\partial P}{\partial \tau}\right)^3},$$

where all derivatives have to be evaluated at $(q,\tau(q))$. The numerator of this expression can be written as $-\langle \vec{v}, (\mathrm{Hess}P)(q,\tau)\vec{v}\rangle$, where $\vec{v} := \left(\frac{\partial P(q,\tau)}{\partial \tau}, -\frac{\partial P(q,\tau)}{\partial q}\right)$ and $(\mathrm{Hess}P)(q,\tau)$ denotes the HESSIAN OF P at the point (q,τ) defined by

$$(\mathrm{Hess}P)(q,\tau) := \begin{pmatrix} \frac{\partial^2 P(q,\tau)}{\partial q^2} & \frac{\partial^2 P(q,\tau)}{\partial q \partial \tau} \\ \frac{\partial^2 P(q,\tau)}{\partial \tau \partial q} & \frac{\partial^2 P(q,\tau)}{\partial \tau^2} \end{pmatrix}.$$

A well-known theorem (cf. e.g. Theorem 3.6 in Fleming [110]) states that $P(q,\tau)$ is convex if and only if $(\mathrm{Hess}P)(q,\tau)$ is a positive semidefinite matrix at every point (q,τ). Furthermore, $P(q,\tau)$ is strictly convex if and only if $(\mathrm{Hess}P)(q,\tau)$ is at every point (q,τ) positive definite. By Theorem 2.6, it follows that $(\mathrm{Hess}P)(q,\tau)$ is a positive semidefinite matrix at every point (q,τ). Using Equation 93 we see that $\tau(q)$ is concave.

$$\frac{d^2\tau(q)}{dq^2} \leq 0. \tag{97}$$

$\alpha(q) = \tau'(q)$ is therefore monotonically decreasing. We notice that together with $\tau(q)$ also $\alpha(q)$ is a real analytic function and that a real analytic functions vanish only at isolated points and therefore it follows for all $q \in I\!R$: $\alpha(q) > 0$ since $\alpha(q)$ is a monotonic function. The inverse $q(\alpha)$ of $\alpha(q)$ is according to known theorems defined on $\alpha(I\!R)$, infinitely differentiable (with the possible exception of some isolated points) and strictly monotonically decreasing. Thus

$$f'(\alpha(q)) = q \tag{98}$$

$$f''(\alpha(q)) = \frac{1}{\tau''(q)} < 0 \tag{99}$$

The second equation is of course not valid at the isolated points where $\tau''(q)$ vanishes.

Definition 4.1 *Let $\omega \in \tilde{\Sigma}_A$ and $I(\omega_0...\omega_n) \in \mathcal{Z}_n$, then we define*

$$\alpha_H(I(\omega_0...\omega_n)) := \frac{\ln \nu(I(\omega_0...\omega_n))}{\ln \ell(I(\omega_0...\omega_n))}$$

and if the limit

$$\alpha_H(\omega) := \alpha_H(\tilde{\beta}(\omega)) := \lim_{n\to\infty} \alpha_H(I(\omega_0...\omega_n))$$

exists, then $\alpha_H(\omega)$ is called the HÖLDER-COEFFICIENT *of ω resp. of $\tilde{\beta}(\omega)$. Furthermore, we define*

$$\mathsf{S}(\alpha) := \{x \in \partial I\!D \mid \alpha_H(x) = \alpha\}. \tag{100}$$

If the Hölder-exponent - defined in Equation 16 - exists, then the Hölder-coefficient just defined in Definition 4.1 exists and equals the Hölder-exponent in Equation 16.

Theorem 4.3 $\mathsf{f}(\alpha) = \dim_H(\mathsf{S}(\alpha))$.

Proof: It follows from the Equations 92 and 94 that for $\mu_{q,\tau(q)}$-almost all $\gamma \in \mathcal{A}$

$$\alpha(q) = -\lim_{m\to\infty} \frac{\frac{1}{m+1}\sum_{k=0}^m \chi(f^k\gamma_\infty)}{\frac{1}{m+1}\ln |(f^{m+1})'(\gamma_\infty)|} \tag{101}$$

is valid. *For every $\gamma \in \mathcal{A}$, there exists a $\omega_\gamma \in \tilde{\Sigma}_A$ such that $\tilde{\beta}(\omega_\gamma) = \gamma$.* Because of $P(\chi) = 0$ we are able with the help of Theorem 2.9 to express the numerator through $\nu(I(\omega_0...\omega_{m+1}))$ and with the help of Lemma 4.3 to express the denominator through $\ell(I(\omega_0...\omega_{m+1}))$:

$$\alpha(q) = \lim_{m\to\infty} \frac{\ln(\nu(I(\omega_0...\omega_{m+1})))}{\ln(\ell(I(\omega_0...\omega_{m+1})))} = \alpha_H(\omega) = \alpha_H(\gamma).$$

I.e. for $\mu_{q,\tau(q)}$-*almost all $\gamma \in \mathcal{A}$ the Hölder-coefficient $\alpha_H(\gamma)$ exists and equals $\alpha(q)$. Equivalently,* the set $\mathsf{S}(\alpha(q))$ has $\mu_{q,\tau(q)}$-measure equal to one:

$$\mu_{q,\tau(q)}(\mathsf{S}(\alpha(q))) = 1.$$

Let $\gamma \in \mathsf{S}(\alpha(q))$ and $\omega = \kappa\gamma$, then by Theorem 2.9 and by Lemma 4.1, there exists a $D > 0$ such that

$$\mu_{q,\tau(q)}(I(\omega_0...\omega_n)) \in [D^{-1}, D] \exp\left(\sum_{k=0}^m \phi_{q,\tau(q)}(f^k\gamma_\infty)\right). \tag{102}$$

Furthermore,

$$\begin{aligned} q\alpha(q) - \tau(q) &= \lim_{m\to\infty} \frac{(q\sum_{k=0}^m \chi(f^k\gamma_\infty) - \tau(q)\sum_{k=0}^m \phi_1(f^k\gamma_\infty))}{\ln \ell(I(\omega_0...\omega_m))} \\ &= \lim_{m\to\infty} \frac{\sum_{k=0}^m \phi_{q,\tau(q)}(f^k\gamma_\infty)}{\ln \ell(I(\omega_0...\omega_n))} \\ &= \lim_{m\to\infty} \frac{\ln \mu_{q,\tau(q)}(I(\omega_0...\omega_n))}{\ln \ell(I(\omega_0...\omega_n))}, \end{aligned}$$

where we have used Lemma 4.3 and Equation 101 in the first equality, the definition of $\phi_{q,\tau(q)}$ in the second equality and Equation 102 in the last equality. We have therefore proved that for all $\gamma \in \mathsf{S}(\alpha(q))$ the limit $\lim_{m\to\infty} \frac{\ln \mu_{q,\tau(q)}(I(\omega_0...\omega_n))}{\ln \ell(I(\omega_0...\omega_n))}$ exists and equals $q\alpha(q) - \tau(q)$.
Application of Lemma 4.5 to $\Lambda \equiv \mathsf{S}(\alpha(q))$ yields

$$\dim_H(\mathsf{S}(\alpha(q))) = q\alpha(q) - \tau(q).$$

□

Lemma 4.5 *Let $\Lambda \subset \partial I\!D$ be a $\mu_{q,\tau(q)}$-measurable set with $\mu_{q,\tau(q)}(\Lambda) > 0$. If for all $x \in \Lambda$*

$$\delta_u \leq \liminf_{n\to\infty} \frac{\ln \mu_{q,\tau(q)}(I(\omega_0...\omega_n))}{\ln \ell(I(\omega_0...\omega_n))} \leq \limsup_{n\to\infty} \frac{\ln \mu_{q,\tau(q)}(I(\omega_0...\omega_n))}{\ln \ell(I(\omega_0...\omega_n))} \leq \delta_o$$

is valid, where we defined $\omega := \kappa(x)$, then it follows

$$\delta_u \leq \dim_H(\Lambda) \leq \delta_o.$$

A proof can be found for instance in Young [111]. The assertion in [111] is slightly different from that in Lemma 4.5; however, one can prove Lemma 4.5 using the same reasoning as in [111]. A direct proof can be found in Billingsley [50] Theorem 14.1.

In the thermodynamical formalism for multifractals we introduced in Section 1.4 a bivariate entropy-function $Q_G(\epsilon,\alpha)$. If the topological pressure $P(q,\tau)$ is strictly convex, we define two new independent variables unambigously through

$$\begin{aligned} \epsilon(q,\tau) &= \frac{\partial P(q,\tau)}{\partial \tau}, \\ \epsilon(q,\tau)\alpha(q,\tau) &= -\frac{\partial P(q,\tau)}{\partial q} \end{aligned}$$

(these equations correspond to the Equations 37 and 38). To this end it is only necessary to show that the functional determinant F of the above 'coordinate-transformation' does not vanish. It is easy to check that

$$\mathsf{F} = \frac{\left(\frac{\partial^2 P(q,\tau)}{\partial\tau\partial q}\right)^2 - \frac{\partial^2 P(q,\tau)}{\partial q^2}\frac{\partial^2 P(q,\tau)}{\partial \tau^2}}{\frac{\partial P(q,\tau)}{\partial \tau}}.$$

If $P(q,\tau)$ is *strictly convex*, then the Hessian of P is positive definite at every point. By a criterion of Hurwitz, it follows that the determinant of the Hessian is positive and non-vanishing. From Equation 93 it follows that the functional determinant F satisfies

$$\mathsf{F} < 0.$$

It is straightforward to show that in this case it is possible to define in analogy to Equation 36 an entropy-function $Q(\epsilon,\alpha)$ which fulfills relations analogous to Equations 34, 35 and 41.

Acknowledgments

I would like to thank Prof. Frank Steiner for his engaged support and for fruitful discussions. This work has been partly supported by Deutsche Forschungsgemeinschaft.

A Measure theory

In this appendix we briefly review some basic elements of measure theory, cf. [36], [112] or [113].

Definition A.1 *Let X be an arbitrary (abstract) set. A family $\mathcal{A}$ of subsets of X is called an* ALGEBRA *if*

$$\begin{aligned} X &\in \mathcal{A}, \\ A \in \mathcal{A} &\Rightarrow A^c \in \mathcal{A}, \\ A, B \in \mathcal{A} &\Rightarrow A \cup B \in \mathcal{A}. \end{aligned}$$

*We say that $\mathcal{A}$ is a σ-*ALGEBRA *if furthermore:*

$$A_i \in \mathcal{A}, i \in I\!N \Longrightarrow \cup_i A_i \in \mathcal{A}.$$

A pair $(X,\mathcal{A})$ consisting of a set X together with a σ-algebra $\mathcal{A}$ is called a MEASURABLE SPACE *and the elements in $\mathcal{A}$ are called* MEASURABLE SETS.

Here $A^c := X \backslash A$ denotes the complement of A with respect to X.
Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two families of subsets of X: we say that $\mathcal{A}_1$ is contained in $\mathcal{A}_2$ if every set in $\mathcal{A}_1$ is also contained in $\mathcal{A}_2$: we write $\mathcal{A}_1 \subset \mathcal{A}_2$. We say that a σ-algebra $\mathcal{A}$ is GENERATED by a family $\mathcal{A}_0$ of subsets of X if $\mathcal{A}_0 \subset \mathcal{A}$ and every σ-algebra $\mathcal{A}'$ which contains $\mathcal{A}_0$ contains also $\mathcal{A}$. $\mathcal{A}$ is therefore the smallest σ-algebra which can be constructed from $\mathcal{A}_0$ by adding further subsets of X to $\mathcal{A}_0$. This definition makes sense since the intersection of a family of σ-algebras of X is again a σ-algebra of subsets of X and since the family of all subsets of X is a σ-algebra.

If X is a topological space, then we call the σ-algebra generated by the open subsets of X BOREL σ-ALGEBRA $\mathcal{B}(X)$ of X. The elements of $\mathcal{B}(X)$ are called BOREL SETS.

From now on let $\mathcal{A}$ denote a σ-algebra of X.

Definition A.2 *A map $\mu : \mathcal{A} \to [0,\infty]$ with the properties*

$$\begin{aligned} \mu(\emptyset) &= 0, \\ \mu\left(\bigcup_{i=1}^{\infty} A_i\right) &= \sum_{i=1}^{\infty} \mu(A_i) \text{ for } A_i \cap A_j = \emptyset, i \neq j. \end{aligned} \qquad (103)$$

is called a MEASURE *on X. μ is also called* COUNTABLY SUBADDITIVE *(because of Equation 103). If X is a topological space and μ is defined on the Borel σ-algebra $\mathcal{B}(X)$ of X, then μ is called a* BOREL MEASURE.
The triple $(X, \mathcal{A}, \mu)$ is called a MEASURE SPACE. *If $\mu(X) < \infty$, then we say that μ is a* FINITE *measure. If furthermore $\mu(X) = 1$, then μ is called* NORMED *and $(X, \mathcal{A}, \mu)$ is called a* PROBABILITY SPACE.
If it is possible to write X as countable union $X = \cup_{k=1}^{\infty} X_k$ with $X_k \in \mathcal{A}$ and if $\mu(X_k) < \infty$ for all $k \in I\!N$, then μ is called a SIGMA-FINITE *(resp. σ-*FINITE*) measure.*

Definition A.3 *Given two measures μ_1 and μ_2 on the measurable space $(X,\mathcal{A})$, then μ_1 is called* ABSOLUTLY CONTINUOUS *with respect to μ_2 if $\mu_2(N) = 0$ implies $\mu_1(N) = 0$ for all $N \in \mathcal{A}$. We write $\mu_1 \ll \mu_2$. Two measures are said to be* EQUIVALENT *if $\mu_1 \ll \mu_2$ and $\mu_2 \ll \mu_1$. We write in this case $\mu_1 \sim \mu_2$. Two measures which are not equivalent are also called* SINGULAR *with respect to each other.*

Definition A.4 *A* PARTITION *of a measure space $(X, \mathcal{A}, \mu)$ is a covering of X consisting of disjoint and measurable sets.*

Definition A.5 *Let C be the class of all subsets of X. A map $\nu : C \to [0,\infty]$ is called an* OUTER MEASURE *if:*

$$\begin{aligned} \nu(\emptyset) &= 0 \\ \nu(Z_1) &\leq \nu(Z_2) \qquad \text{for all } Z_1 \subseteq Z_2 \\ \nu(\cup_i Z_i) &\leq \sum_i \nu(Z_i) \qquad Z_i \in C. \end{aligned}$$

Because of the third property, ν is also called COUNTABLY SUBADDITIVE.

Definition A.6 *A subset Z of X is called ν-*MEASURABLE *or* CARATHÉODORY-MEASURABLE *with respect to an outer measure ν if:*

$$\nu(A) = \nu(A \cap Z) + \nu(A \backslash Z) \qquad (104)$$

for all sets $A \subset Z$.

Theorem A.1 (Carathéodory) *Let ν be an outer measure on X. The family $\mathcal{M}$ of ν-measurable subsets of X is a σ-algebra and the restriction of ν to $\mathcal{M}$ is a measure.*

The proof can be found e.g. in Falconer [38] or Edgar [36] Theorem 5.2.3.

For completeness we sketch now the definition of the integral with respect to μ:
Let $(X, \mathcal{A}, \mu)$ be a measure space. For $A \subset X$ the CHARACTERISTIC FUNCTION f_A OF A is defined by:

$$f_A(x) = \begin{cases} 1 & : \ x \in A \\ 0 & : \ x \notin A \end{cases}$$

We say that a function $f : X \to \mathbb{R}$ is SIMPLE if f can be written as:

$$f = \sum_{i=1}^{N} \lambda_i f_{A_i},$$

where $\lambda_i \in \mathbb{R}, A_i \in \mathcal{A}$ for $1 \le i \le N$. A simple function is called INTEGRABLE *if*

$$\sum_{i=1}^{N} \lambda_i \mu(A_i) < \infty$$

and we define the integral of a simple function by

$$\int_X f d\mu := \sum_{i=1}^{N} \lambda_i \mu(A_i).$$

This value does not depend not on the decomposition of f. We say that a function $f : X \to \mathbb{R}$ is *INTEGRABLE if there exists a family $(f_n)_{n\in\mathbb{N}}$ of simple functions on X such that*

$$\lim_{n\to\infty} f_n(x) = f(x) \text{ for } \mu\text{-almost all } x \in X$$

$$\lim_{n,m\to\infty} \int_X |f_n - f_m|\, d\mu = 0.$$

In this case we define the INTEGRAL of f by

$$\int_X f d\mu := \lim_{n\to\infty} \int_X f_n d\mu.$$

It can be shown that the limit exists and is independent of choice of the family (f_n).
We say that a function $f : X \to \mathbb{C}$ is INTEGRABLE if $\mathrm{Re} f$ and $\mathrm{Im} f$ are integrable and we set

$$\int_X f d\mu := \int_X \mathrm{Re} f d\mu + i \int_X \mathrm{Im} f d\mu.$$

As usual, we define $\mathcal{L}^p(X, \mathcal{A}, \mu)$ to be the set of all functions $f : X \to \mathbb{C}$ such that $|f|^p$ is integrable. We denote with $L^p(X, \mathcal{A}, \mu)$ the set of equivalence classes of functions in $\mathcal{L}^p(X, \mathcal{A}, \mu)$ modulo the equivalence relation 'almost everywhere (a.e.) equal'. Two functions $g : X \to \mathbb{C}$ and $f : X \to \mathbb{C}$ are called μ-ALMOST EVERYWHERE EQUAL if $\mu(\{x \in X \mid g(x) \ne f(x)\}) = 0$. The space $L^p(X, \mathcal{A}, \mu)$ is a Banach space with norm $\|f\|_p := \sqrt[p]{\int |f|^p\, d\mu}$.
A function $f : X \to \mathbb{C}$ is integrable over $A \subset X$ if $f f_A$ is integrable and we set:

$$\int_A f d\mu = \int_X f f_A d\mu.$$

For compact metric spaces X every continuous complex valued function on X is μ-integrable if μ is a finite measure on X.
A more detailed treatment of integration theory can be found e.g. in Berberian [112].

B Ergodic theory

Ergodic theory deals with the action of groups on spaces which have an additional structure: for example, measure spaces or topological spaces. The elements of the group act as structure preserving transformations on the given space. Often the notion 'ergodic theory' is restricted to denote the study of the action of groups on measure spaces.

The word *ergodic* has been introduced by L. Boltzmann. It is composed of the greek words *ergos* and *odos*. Boltzmann's *ergodic hypothesis* states that the orbit of a closed system in phase space during a sufficient long time interval lies dense in phase space (resp. in the submanifold of phase space to constant energy). Boltzmann's target was to deduce the equality of time and phase space averages of physical observables from his hypothesis. A physical observable is a real valued function of the phase space coordinates and the time. However, Boltzmann's ergodic hypothesis is in general false. In analogy to Boltzmann's ideas a transformation T on a measurable space $(X, \mathcal{A}, \mu)$ is called ergodic if for all μ-integrable functions f the integral of f equals the time mean of f for μ-almost all $x \in X$. This is also the content of Birkhoff's ergodic theorem (Theorem B.1).

Measure theoretical ergodic theory

Let $(X, \mathcal{A}, \mu)$ be a measure space. A map $T : X \to X$ is called MEASURABLE with respect to $\mathcal{A}$ if the inverse image satisfies $T^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{A}$. (If X is a topological space and $\mathcal{A} = \mathcal{B}(X)$, then for instance every continuous map is measurable).
The quadruple $(X, \mathcal{A}, \mu, T)$ is in this case also called a (measure theoretical) DYNAMICAL SYSTEM and X is also called the PHASE SPACE of the dynamical system.
For $x \in X$ we denote the ORBIT of x with respect to T by $O_T(x) := \{T^n(x)\}_{n=0}^{\infty}$. We say that a measurable map $T : X \to X$ is (μ-)MEASURE-PRESERVING and that μ is T-INVARIANT if $\mu(A) = \mu(T^{-1}(A))$ for all $A \in \mathcal{A}$.
If X is a topological space, then the set of all Borel probability measures on X is denoted by $M(X)$ and the set of all T-invariant Borel probability measures is denoted by $M_T(X)$.
Further, we say that $T : X \to X$ is an INVERTIBLE MEASURE-PRESERVING TRANSFORMATION (resp. that T is an AUTOMORPHISM on $(X, \mathcal{A}, \mu)$) if T is measure-preserving, bijective and if T^{-1} is also measure-preserving.

Definition B.1 *A measure-preserving transformation T on the measure space $(X, \mathcal{A}, \mu)$ is called* ERGODIC *if for every $A \in \mathcal{A}$ with $T^{-1}(A) = A$ either $\mu(A) = 0$ or $\mu(A^c) = 0$ holds.*
A measure-preserving transformation T on $(X, \mathcal{A}, \mu)$ is called (STRONG) MIXING *if for all $A, B \in \mathcal{A}$*

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B)$$

holds. Given a pair (X, T), we will also say that μ is ergodic (resp. mixing) if T is ergodic (resp. mixing) on $(X, \mathcal{A}, \mu)$.

As above, $A^c = X \setminus A$ denotes the complement of A with respect to X.
That a transformation T is mixing means that for every set A the sequence $T^{-n}(A)$ becomes asymptotically independent of B; ergodicity means that $T^{-n}(A)$ becomes in the mean asymptotically independent of B for all $A, B \in \mathcal{A}$. Of central importance is the following

Theorem B.1 (Birkhoff) *Let $(X, \mathcal{A}, \mu)$ be a probability space and $T : X \to X$ a measure-preserving ergodic transformation, then*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i x) = \int f d\mu$$

for all $f \in L^1(X, \mathcal{A}, \mu)$ and μ-almost all $x \in X$.

A proof can be found for instance in Walters [21].

Definition B.2 *Let $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ be two probability spaces and $T_1 : X_1 \to X_1$ and $T_2 : X_2 \to X_2$ two measurable maps. We say that the two dynamical systems $(X_1, \mathcal{A}_1, \mu_1, T_1)$ and $(X_2, \mathcal{A}_2, \mu_2, T_2)$ are* ISOMORPHIC *if there exist invariant measurable sets $X_i^0 \subset X_i$* [18]*, with $\mu_i(X_i^0) = 1$ and if there exists a measurable, measure-preserving and bijective transformation*

$$\phi : X_1^0 \to X_2^0$$

such that

$$\phi T_1 = T_2 \phi.$$

Theorem B.2 *Let $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ be two probability spaces and $T_1 : X_1 \to X_1$ and $T_2 : X_2 \to X_2$ be two measurable transformations such that $(X_i, \mathcal{A}_i, \mu_i, T_i), i = 1, 2$, are two isomorphic dynamical systems, then we have that T_1 is ergodic (resp. mixing) if and only if T_2 is ergodic (resp. mixing).*

A proof can be found in Walters [21] Theorem 2.13.

Example for a mixing dynamical system:
We consider $(Z, 2^Z, \mu)$, where $Z = \{0, 1, \dots, k-1\}$ and where 2^Z denotes the set of all subsets of Z. The measure μ is given by a probability vector $(p_0, p_1, \dots, p_{k-1})$ by $\mu(\{i\}) = p_i$.
We consider the direct product

$$\Sigma_k := \prod_{-\infty}^{\infty} Z = \{0, 1, \dots, k-1\}^{\mathbb{Z}} = \{(x_i)_{-\infty}^{\infty} \mid x_i \in Z\}.$$

In order to define a product σ-algebra, we consider sets of the form (also called ELEMENTARY RECTANGLES)

$${}_{-n}[a_{-n}, a_{-n+1}, \dots, a_n]_n := \{(x_i)_{-\infty}^{\infty} \mid x_j = a_j \text{ for } -n \le j \le n\}$$

The family of all elementary rectangles is a semi-algebra [19] $\mathcal{S}$, which generates a product σ-algebra $\mathcal{P}$. We write

$$(\Sigma_k, \mathcal{P}) = \prod_{-\infty}^{\infty} (Z, 2^Z).$$

We consider now blocks of the form ${}_h[a_h, a_{h+1}, \dots, a_l]_l := \{(x_i)_{-\infty}^{\infty} \mid x_j = a_j \text{ for } h \le j \le l\}$. We define a measure on these asymmetrical blocks by

$$m({}_h[a_h, a_{h+1}, \dots, a_l]_l) := \prod_{j=h}^{l} p_{a_j}.$$

[18] The spaces X_i^0 are furnished with the σ-algebra $X_i^0 \cap \mathcal{A}_i := \{X_i^0 \cap B \,,\, B \in \mathcal{A}_i\}$ and the measures μ_i are restricted to these σ-algebras.

[19] A family $\mathcal{S}$ of subsets of X is a semi-algebra if

$$\emptyset \in \mathcal{S},$$
$$A, B \in \mathcal{S} \Rightarrow A \cap B \in \mathcal{S},$$
$$A \in \mathcal{S} \Rightarrow A^c = \cup_{i=1}^{n} B_i \text{ with } B_i \in \mathcal{S} \text{ and } B_i \cap B_j = \emptyset, \text{ if } i \neq j$$

holds.

One can prove (cf. Walters [21]) that m can be extended to entire $\mathcal{P}$. m is therefore called the $(p_0, p_1, \dots, p_{k-1})$-product measure on Σ_k.
We define the following shift transformation

$$\sigma_k : \Sigma_k \to \Sigma_k, \sigma_k(\{x_i\}) := \{y_i\},$$

where $y_i = x_{i+1}$.
The pair (Σ_k, σ_k) is a symbolic dynamical system in the sense of Definition 2.1. It is clear that σ_k is measure-preserving with respect to m. (A proof of this fact can e.g. be found in Walters [21].)

Theorem B.3 *The shift operator σ_k on $(\Sigma_k, \mathcal{P}, m)$ is (strong) mixing and ergodic.*

A proof can be found in Walters [21] Theorem 1.12.

Topological dynamics

The measure theoretical ergodic theory deals with measure-preserving transformations. In topological dynamics the properties of continuous maps between topological spaces are studied. We restrict ourselves here to continuous transformations on *compact* metric spaces.
Let X be a compact metric space and let $T : X \to X$ be a homeomorphism on X. The pair (X, T) is called a TOPOLOGICAL DYNAMICAL SYSTEM.
We introduce now a concept which is analogous to ergodicity.

Definition B.3 *A homeomorphism $T : X \to X$ is called* MINIMAL *if the T-orbit $O_T(x) := \{T^n(x) : n \in \mathbb{Z}\}$ lies dense in X for all $x \in X$.*

Given a minimal homeomorphism T, it is possible to show [21] that the only closed subsets $X_1 \subset X$ which satisfy $TX_1 = X_1$ are $\emptyset$ and X. Therefore minimality is analogous to ergodicity.

Definition B.4 *Let $T : X \to X$ be a homeomorphism on a compact metric space.*

- *T is called* TOPOLOGICALLY TRANSITIVE *if there exists to every pair of non-empty, open sets U, V of X an $n \in \mathbb{Z}$ such that*

$$T^n(U) \cap V \neq \emptyset.$$

- *T is called* TOPOLOGICALLY MIXING *if there exists to every pair of non-empty, open sets U, V of X an $N \in \mathbb{N}$ such that*

$$T^n(U) \cap V \neq \emptyset \text{ for all } n > N.$$

A homeomorphism T is topologically transitive if and only if there exists at least *one* $x \in X$ such that $O_T(x)$ lies dense in X. Topological transitivity is therefore a weaker condition than minimality.
A topologically mixing homeomorphism is also topologically transitive. A homeomorphism T is topologically mixing if and only if T^{-1} is.
The following definition replaces Definition B.2

Definition B.5 *Let $T_1 : X_1 \to X_1$ and $T_2 : X_2 \to X_2$ be two homeomorphism on compact metric spaces X_1 resp. X_2. We say that T_1 is* TOPOLOGICALLY CONJUGATE *to T_2 if there exists a homeomorphism $\phi : X_1 \to X_2$ such that $\phi \circ T_1 = T_2 \circ \phi$. The homeomorphism ϕ is called a* (TOPOLOGICAL) CONJUGACY.

We again consider again our above **example**:
Endow $Z = \{0, 1, ..., k-1\}$ with the discrete topology, i.e. every subset of Z is open. Σ_k endowed with the product topology is a topological space. A basis of neighbourhoods at the point $\omega = \{\omega_n\}$ is in this topology given by the sets

$$U_N(\omega) := \{\{\omega'_n\} \mid \omega'_n = \omega_n \text{ for all } |n| \leq N\}, \quad N \geq 1. \tag{105}$$

The topology obtained in this way is metrizable; however, there exists no preferred metric. A possible choice of a metric on Σ_k is

$$d_\rho(\{\omega_n\}, \{\omega'_n\}) := \rho^{t(\{\omega_n\}, \{\omega'_n\})}, \tag{106}$$

where t is defined by $t(\{\omega_n\}, \{\omega'_n\}) := \min\{|n|: n \in \mathbb{Z}, \omega_n \neq \omega'_n\} \geq 0$ and where ρ is an arbitrary number $0 < \rho < 1$. All these metrics generate the topology given above, i.e. two metrics d_{ρ_1} and d_{ρ_2} are pairwise equivalent. This can been seen by the fact that in the metric d_ρ the set U_N is a circle with radius ρ^{N+1}. A further example for a possible metric is given in Walters [21].

Lemma B.1 *Σ_k endowed with any of the metrics d_ρ, $0 < \rho < 1$, is a compact metric space.*

Proof: That d_ρ is a metric for every $\rho \in]0, 1[$ is clear. To show that Σ_k is compact, it is enough to show that every infinite subset $A \subset \Sigma_k$ possesses an accumulation point in Σ_k. Let therefore A infinite. Then at least one of the sets $A_{i_0} := \{\omega \in A \mid \omega_0 = i_0\}, i_0 \in Z$, is infinite. Inductively *it follows that for every $n \in \mathbb{N}$ at least one of the sets*

$$A_{i_{-n} \ldots i_{-1} i_0 i_1 \ldots i_n} := \{\omega \in A \mid \omega_{-n} = i_{-n}, \ldots, \omega_0 = i_0, \ldots, \omega_n = i_n\}, i_j \in Z,$$

is infinite.
We conclude that there exists an inductively defined $\tau \in \Sigma_k$ given by $\tau := (\ldots i_{-r} \ldots i_{-1} i_0 i_1 \ldots i_r \ldots)$. We choose for every n an element τ_n in $A_{i_{-n} \ldots i_{-1} i_0 i_1 \ldots i_n}$ and define in this way a sequence, which converges in the d_ρ-topology to τ. □

The shift operator σ_k on Σ_k is a homeomorphism. Let $U \subset \Sigma_k$ be open, then there exists to every $\omega \in U$ an $N(\omega)$ such that $\omega \in U_{N(\omega)}(\omega) \subset U$. Thus $\sigma_k^{-1}\omega \in U_{N(\omega)+1}(\sigma_k^{-1}\omega) \subset \sigma_k^{-1} U_{N(\omega)}(\omega) \subset \sigma_k^{-1} U$. Therefore σ_k is continuous. The proof that also σ_k^{-1} is continuous uses the same argument. Let $X' \subset \Sigma_k$ be endowed with the topology induced by Σ_k and let X' be invariant under the restriction σ' of σ_k to X', i.e. $\sigma' X' = X'$. Then σ' is a homeomorphism on X'. In conclusion, we see that for every symbolic dynamical system (Σ, σ) (cf. Definition 2.1) consisting of bi-infinite symbol sequences the space Σ can be endowed with a topology such that Σ is compact and such that σ is a homeomorphism on Σ. The compactness of Σ can be shown using the same argument as in Lemma B.1.

We now consider a one-sided symbolic dynamical system (Σ^+, σ^+). A basis of neighbourhoods at every $\omega^+ \in \Sigma^+$ is given by the sets

$$U_N^+(\omega^+) := \{\{\tilde{\omega}_n^+\} \mid \tilde{\omega}_n^+ = \omega_n^+, \text{ for all } n \leq N\}, \quad N \geq 1. \tag{107}$$

As above, we can introduce for every $\rho \in]0, 1[$ a metric d_ρ^+ on Σ^+ by

$$d_\rho^+(\{\omega_n^+\}, \{\tilde{\omega}_n^+\}) := \rho^{t^+(\{\omega_n^+\}, \{\tilde{\omega}_n^+\})}, \tag{108}$$

where t^+ is defined by $t^+(\{\omega_n^+\}, \{\tilde{\omega}_n^+\}) := \min\{n \mid n \in \mathbb{N}, \omega_n \neq \tilde{\omega}_n^+\} \geq 0$. All these metrics generate the topology given above, i.e. two metrics $d_{\rho_1}^+$ and $d_{\rho_2}^+$ are pairwise equivalent. This can been seen by the fact that in the metric d_ρ^+ the set U_N^+ is a circle with radius ρ^{N+1}.

Corollary B.1 *Σ^+ endowed with any of the metrics d_ρ^+, $0 < \rho < 1$ is a compact metric space.*

The proof is similar to the proof of Lemma B.1.

C Zeta functions for symbolic dynamical systems

In this appendix we introduce several zeta functions and discuss some of their properties. The study of zeta functions has become important in several branches of mathematics and physics, see e.g. [15], [84], [114], [115], [116] and [117]. However, here we limit ourselves to the discussion of zeta functions for symbolic dynamical systems.

Artin-Mazur zeta function

Let (Σ, σ) be a symbolic dynamical system. We are interested in the periodic points of (Σ, σ) (cf. Section 3.6) and define

$$N_n(\sigma) := \#\{\omega \in \Sigma \mid \sigma^n \omega = \omega\}.$$

The information contained in the numbers N_n can also be expressed through a single function. To this end, we consider the formal power series

$$\zeta(z) := \exp \sum_{m=1}^{\infty} \frac{z^m}{m} N_m(\sigma), \quad z \in \mathbb{C}. \tag{109}$$

Artin und Mazur [118] have shown that the power series in Equation 109 has non-vanishing radius of convergence and therefore Equation 109 defines a function, which is holomorphic in a neighborhood of $z = 0$.
Example: We consider the full k-shift (Σ_k, σ_k) from Section 2.1 and Appendix B. It is easy to check that $N_n(\sigma_k) = k^n$ and therefore it follows

$$\zeta(z) = \exp \sum_{m=1}^{\infty} \frac{z^m}{m} k^m = \exp(-\ln(1 - kz)) = \frac{1}{1 - kz}.$$

We see that the zeta function for the full k-shift is a meromorphic function on $\mathbb{C}$ with a simple pole at $z = \frac{1}{k}$.
Bowen & Lanford [119] have derived $\zeta(z)$ for a topological Markov shift (Σ_A, σ_A):

$$N_n(\sigma_A) = trA^n$$

and therefore

$$\begin{aligned}\zeta(z) &= \exp\sum_{m=1}^{\infty}\frac{z^m}{m}\operatorname{tr}A^m\\ &= \exp\operatorname{tr}\sum_{m=1}^{\infty}\frac{z^m A^m}{m}\\ &= \exp\operatorname{tr}(-\ln(1-zA))\\ &= \frac{1}{\det(1-zA)}.\end{aligned}$$

Bowen & Lanford [119] have also shown that for an arbitrary symbolic dynamical system it is *not* true that the zeta function in Equation 109 is rational.

Ruelle's zeta function

Ruelle [17, 84, 116] has generalized the Artin-Mazur zeta function by introducing certain weights into the sum 109. We consider for $z \in \mathbb{C}$ and $\varphi \in C(\Sigma, \mathbb{C})$ the formal series

$$\zeta(z,\varphi) := \exp\sum_{m=1}^{\infty}\frac{z^m}{m}\sum_{\omega\in N_m(\sigma)}\exp(S_m\varphi(\omega)). \tag{110}$$

where we have set (cf. Lemma 2.7):

$$S_m\varphi(\omega) := \sum_{i=0}^{m-1}\varphi(\sigma^i\omega).$$

Ruelle's generalization is a very natural one from the point of view of statistical mechanics since - as we have already seen above in Theorem 2.9 - the quantity $-S_m\varphi(\omega)$ can be considered as an approximate value for the quantity $\mu(\varphi)$, which takes over the role of energy of the state μ in the thermodynamical formalism of Section 2. Remember Definition 2.9 and the discussion following Theorem 2.7. The following lemma gives us information about the radius of convergence of the formal power series in Equation 110.

Lemma C.1 *Let (Σ_A, σ_A) be a topologically mixing topological Markov shift and $\varphi \in \mathcal{F}_A$. The radius of convergence of the power series $\sum_{m=1}^{\infty}\frac{z^m}{m}\sum_{\omega\in N_m(\sigma_A)}\exp(S_m\varphi(\omega))$ equals $e^{-P(\mathrm{Re}\varphi)}$, where $P(\mathrm{Re}\varphi)$ denotes the topological pressure with respect to $\mathrm{Re}\varphi$ introduced in Section 2.3. Especially, the radius of convergence of the series $\sum_{m=1}^{\infty}\frac{z^m}{m}N_m(\sigma_A)$ is given by $e^{-h_{top}}$, where h_{top} denotes the topological entropy of (Σ_A, σ_A).*

The notions used in Lemma C.1 are defined in the Definitions 2.1, 2.2 and B.4. A proof of Lemma C.1 can be found for example in Parry & Pollicott [15] or Ruelle [17].
Equation 110 defines a holomorphic function in a neighborhood of $z = 0$ for topologically mixing topological Markov shifts. Often it is of interest whether there exists a meromorphic continuation of this function. Results in this direction can e.g. be found in Mayer [14] chapter IV, Ruelle [17] or Pollicott [115]. We only state the following theorem in this connection:

Theorem C.1 *Let (Σ_A, σ_A) be a topologically mixing topological Markov shift and $\varphi \in \mathcal{F}_A$, then there exists $R(\varphi) > \exp(-P_{\sigma_A}(\varphi))$ such that*

$$d_A(z) := \frac{1}{\zeta_A(z,\varphi)} := \exp\left[-\sum_{m=1}^{\infty}\frac{z^m}{m}\sum_{\omega\in N_m(\sigma_A)}\exp(S_m\varphi(\omega))\right]$$

has an analytic continuation in $\{z : |z| < R(\varphi)\}$ with exactly one zero z_0 in $\{z : |z| < R(\varphi)\}$. This zero z_0 is simple and we have: $z_0 = \exp(-P(\varphi))$.

A proof can be found in Ruelle [17] Theorem 5.29.
For topologically mixing topological Markov shifts (Σ_A, σ_A) and for functions $\varphi(\omega)$ which depend only on a finite number of symbols of the bi-infinite word ω for arbitrary $\omega \in \Sigma_A$, it is possible to prove a generalization of the result of Bowen & Lanford [119], namely that $\zeta(z,\varphi)$ is a rational function on the entire complex plane $\mathbb{C}$ (cf. Parry & Pollicott [15] or Mayer [14]).

We now show how to reobtain Selberg's zeta function (cf. Definition 3.17) from Ruelle's zeta function. We use the notations from Sections 3 and 4.
Let (Σ_A, σ_A) be the symbolic dynamical system from Corollary 3.2. With the help of (Σ_A, σ_A) we have coded the geodesics in $\mathcal{A}$, or equivalently the geodesics in $S = \mathbb{D}/\Gamma$, where Γ denotes a Fuchsian group of the first kind which satisfies the conditions 0, 1, 2 and 3 stated in Section 3.2. The set $\mathcal{A}$ has been defined in Equation 64. The system (Σ_A, σ_A) is a topologically mixing topological Markov shift by Corollary 3.2. We replace the function φ in Equation 110 by the function $s\varphi_1$, where $s \in \mathbb{C}$ and φ_1 is defined in Equation 80. We further denote the positive endpoint of the geodesic $\gamma \in \mathcal{A}$ corresponding to $\tilde{\beta}(\omega)$ with γ_∞. We obtain for the value of the Ruelle zeta function ζ_A for (Σ_A, σ_A) at the argument $z = 1$:

$$\begin{aligned}\zeta_A(1, s\varphi_1) &= \exp\sum_{m=1}^{\infty}\frac{1}{m}\sum_{\omega\in N_m(\sigma_A)} e^{sS_m\varphi_1(\omega)}\\ &= \exp\sum_{m=1}^{\infty}\sum_{\substack{\omega=\sigma_A^m\omega\\ m=\text{primitive}}}\sum_{k=1}^{\infty}\frac{1}{mk}\exp\left(-sk\ln|(f^m)'(\tilde{\beta}\omega)|\right)\\ &= \prod_{\substack{\text{finite prim}\\ \text{words in }\Sigma_A}}\exp\sum_{k=1}^{\infty}\frac{1}{k\#}\exp\left(-sk\ln|(f^\#)'(\tilde{\beta}\omega)|\right)\\ &= \prod_{\substack{\text{finite prim}\\ \text{words in }\Sigma_A}}\exp\left(-\frac{1}{\#}\ln(1-\exp(-s\ln|(f^\#)'(\tilde{\beta}\omega)|))\right)\\ &= \prod_{\substack{\text{finite prim}\\ \text{words in }\Sigma_A}}\left(1-\exp(-s\ln|(f^\#)'(\tilde{\beta}\omega)|)\right)^{-\frac{1}{\#}}.\end{aligned}$$

The function $\#$ maps every word to its length. The infinite products and sums converge by Lemma C.1 if $1 < \exp(-P(\mathrm{Re}(s)\varphi_1))$. It follows by Theorem 4.2 that the infinite sums and products converge for $\mathrm{Re}(s) > -\tau(0) = \dim_H(\partial\mathbb{D}) = 1$.
We want to rewrite this product into a product over the periodic geodesics in $\mathcal{R}$ (cf. Equation 57). We use Corollary 3.2 and Equation 73 resp. Equation 55 and find

$$\zeta_A(1, s\varphi_1) = \left(\prod_{\{l_n\}}(1-\exp(-sl_n))^{-g_n}\right)\left(\prod_{\substack{\text{finite prim}\\ \text{words in }\Sigma_A\setminus\tilde{\Sigma}_A}}\left(1-\exp(-s\ln|(f^\#)'(\tilde{\beta}\omega)|)\right)^{-\frac{1}{\#}}\right)$$

$$= \left(\prod_{\{\ell_n\}} (1-\exp(-s\ell_n))^{-g_n}\right)\left(\prod_{\gamma\in N(\mathcal{F})\cap\mathcal{R}} (1-\exp(-s\ell(\gamma)))^{-g(\gamma)+1}\right).$$

Here the product over $\{\ell_n\}$ is taken over all primitive lengths ℓ_n of periodic geodesics in $S = \mathbb{D}/\Gamma$ and g_n denotes the multiplicity of the length ℓ_n. The space $\tilde{\Sigma}_A$ has been defined in Section 4 by $\tilde{\Sigma}_A := \kappa(\mathcal{A})$, where κ has been defined in Equation 78. The set $N(\mathcal{F})$ has been defined in Lemma 3.8. The number $\ell(\gamma)$ denotes the primitive length of the periodic geodesic $\gamma \in N(\mathcal{F})\cap\mathcal{R}$ and $g(\gamma)$ is equal to the number of periodic words in Σ_A which are mapped to $\gamma \in N(\mathcal{F})\cap\mathcal{R}$ by $\tilde{\beta}$. Therefore we have by Corollary 3.2: $g(\gamma) \le 4$. We use the abbreviation

$$R(s) := \prod_{\gamma\in N(\mathcal{F})\cap\mathcal{R}} (1-\exp(-s\ell(\gamma)))^{g(\gamma)-1}.$$

Finally, we find the connection between Selberg's zeta function $Z_\Gamma(s)$ and Ruelle's zeta function for $s\varphi_1$

$$\zeta_A(1, s\varphi_1) = \frac{Z_\Gamma(s+1)}{Z_\Gamma(s)} R(s), \quad \mathrm{Re}(s) > 1. \tag{111}$$

Above we have shown by Lemma C.1 and Theorem 4.2 that $\zeta_A(1, s\varphi_1)$ is well defined for $\mathrm{Re}(s) > 1$. However, this assertion can also be deduced from Equation 111.
Since $R(s)$ is defined by a finite product, it follows that $R(s)$ is holomorphic on the entire complex plane $\mathbb{C}$. It follows from Theorem 3.16 that $\zeta_A(1, s\varphi_1)$ has a meromorphic continuation in s to the entire complex plane $\mathbb{C}$. The locations of the poles follow from Theorem 3.16.
The reason for the appearance of the factor $R(s)$ in Equation 111 is that the correspondence of periodic words in Σ_A to periodic geodesics in $S = \mathbb{D}/\Gamma$ is not bijective. We consider therefore instead of (Σ_A, σ_A) the symbolic dynamical system $(\tilde{\Sigma}_A, \tilde{\sigma}_A)$ introduced in Section 4.1, where we have set $\tilde{\Sigma}_A := \kappa(\mathcal{A})$. ($\kappa$ is defined in Equation 78 and $\tilde{\sigma}_A$ denotes the shift operator restricted to $\tilde{\Sigma}_A$.) Further, we denote with $\tilde{\varphi}_1$ the restriction of φ_1 to $\tilde{\Sigma}_A$. Using the same manipulations (with reverse order) as above yields

$$\begin{aligned}\frac{Z_\Gamma(s+1)}{Z_\Gamma(s)} &= \left(\prod_{\{\ell_n\}} (1-\exp(-s\ell_n))^{-g_n}\right)\\ &= \prod_{\substack{\text{finite prim}\\ \text{words in } \tilde{\Sigma}_A}} \left(1-\exp(-s\ln|(f^{\#})'(\tilde{\beta}\omega)|)\right)^{-1}\\ &= \tilde{\zeta}_A(1, s\tilde{\varphi}_1).\end{aligned}$$

Here $\tilde{\zeta}_A$ denotes the Ruelle zeta function of the symbolic dynamical system $(\tilde{\Sigma}_A, \tilde{\sigma}_A)$. By Theorem 3.16 and this equation, it follows that $\tilde{\zeta}_A(1, s\tilde{\varphi}_1)$ is well defined for $\mathrm{Re}(s) > 1$ and has a meromorphic continuation to the entire complex plane $\mathbb{C}$. It should be stressed that this statements can not be deduced from Lemma C.1 since $(\tilde{\Sigma}_A, \tilde{\sigma}_A)$ is not a topological Markov shift.
Furthermore, it follows that for $\mathrm{Re}(s) > 1$:

$$Z_\Gamma(s) = \prod_{k=0}^{\infty} (\tilde{\zeta}_A(1, (s+k)\tilde{\varphi}_1))^{-1} \tag{112}$$

holds.

References

[1] Anosov, D., *Geodesic flows on closed Riemann manifolds with negative curvature*, Proc. Steklov Inst. Math. **90** (1967).

[2] Selberg, A., J. Indian Math. Soc. **20** (1956), 47-87.

[3] Gutzwiller, M., Contemp. Math. **53** (1986), 215-251.

[4] Steiner, F., Phys. Lett. B **188** (1987), 447-454.

[5] Pesin, Ya.B., J. Stat. Phys. **71** (1993) 529-547.

[6] Halsey, T.C., M.H. Jensen, L.P. Kadanoff, I. Procaccia, and B. Shraiman, Phys. Rev. A **33** (1986), 1141-1151.

[7] Kohmoto, M., Phys. Rev. A **37** (1988), 1345-1350.

[8] Alekseev, V.M. and M.V. Yakobson, Phys. Rep. **75** (1981), 287-325.

[9] Bowen, R.E., *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture Notes in Mathematics **470**, (Springer, New York, 1975).

[10] Bowen, R.E., *On axiom A diffeomorphisms*, CBMS Regional Conf. Series vol. **35**, (Am. Math. Soc., Providence, R.I., 1978).

[11] Bowen, R.E. and D. Ruelle, Invent. Math. **29** (1975), 181-202.

[12] Coornaert, M. and A. Papadopoulos, *Symbolic Dynamics and Hyperbolic Groups*, Lecture Notes in Mathematics **1539**, (Springer, Berlin, 1993).

[13] Lanford, O.E., *Entropy and equilibrium states in classical statistical mechanics*, in: Statistical Mechanics and mathematical problems, A. Lenard, Ed., Lecture Notes in Physics **20** (Springer, New York, 1971).

[14] Mayer, D.H., *The Ruelle-Araki Transfer Operator in Classical Statistical Mechanics*, Lecture Notes in Physics **123**, (Springer, Berlin, 1980).

[15] Parry, W. and M. Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, Astérisque **187-188** (1990).

[16] Pollicott, M., Am. J. Math. **109** (1987), 183-200.

[17] Ruelle, D., *Thermodynamic formalism*, Encyclopedia of Math. and its Appl. **5**, (Addison-Wesley, Reading, Mass., 1978).

[18] Ruelle, D., Commun. Math. Phys. **125** (1989), 239-262.

[19] Sinai, Ya. G., Russ. Math. Surv. **27** (1972), 21-69.

[20] Smale, S., Bull. Am. Math. Soc. **73** (1967), 747-817.

[21] Walters, P., *An Introduction to Ergodic Theory*, (Springer, New York, 1982).

[22] Bedford, T., M. Keane and C. Series (eds.), *Ergodic Theory, Symbolic Dynamics and Hyperbolic Spaces*, (Oxford University Press, 1991).

[23] Birman, J.S. and C.M. Series, Combinatorial Group Theory and Topology, Annals of Mathematics Studies **111** (Princeton Univ. Press, 1987), 451-478.

[24] Bowen, R.E., Commun. Math. Phys. **69** (1979), 1-17.

[25] Bowen, R.E. and C.M. Series, Publ. Math. IHES **50** (1979), 153-170.

[26] Series, C.M., Acta Math. **146** (1981), 103-128.

[27] Series, C.M., Ergod. Th. & Dynam. Sys. **1** (1981), 337-360.

[28] Series, C.M., *Ergod. Th. & Dynam. Sys.* **6** *(1986), 601-625.*

[29] Rand, D.A., Ergod. Th. & Dynam. Sys. **9** (1989), 527-541.

[30] Collet, P., J.L. Lebowitz and A. Porzio, J. Stat. Phys. **47** (1987), 609-644.

[31] Barnsley, M.F., *Fractals everywhere*, (Academic Press, London, 1988).

[32] Mandelbrot, B.B., *The fractal geometry of nature* (Freeman, New York, 1982).

[33] Ruelle, D. and F. Takens, *Commun. Math. Phys.* **20** *(1971), 167-192.* Commun. Math. Phys. **23** (1971), 343-344.

[34] Ruelle, D., Math. Intelligencer **2** (1982), 126-137.

[35] Hausdorff, F., Math. Ann. **79** (1919), 157-179.

[36] Edgar, G.A., *Measure, topology, and fractal geometry*, (Springer, New York, 1990).

[37] Farmer, J.D., E. Ott, and J.A. Yorke, Physica *D* **7** *(1983), 153-180.*

[38] Falconer, K., *The geometry of fractal sets*, (Cambridge Univ. Press, 1985).

[39] Falconer, K., *Fractal Geometry* (Wiley, Chichester, 1990).

[40] Grassberger, P., Phys. Lett. **97A** (1983), 227-230.

[41] Grassberger, P., Phys. Lett. **107A** (1985), 101-105.

[42] Grebogi, C., E. Ott, and J.A. Yorke, *Phys. Rev. A* **36** *(1987), 3522-3524.*

[43] Grebogi, C., E. Ott, and J.A. Yorke, Phys. Rev. A **37** (1988), 1711-1724.

[44] Leven, R.W., B.P. Koch, and B. Pompe, *Chaos in dissipativen Systemen* (Akademie-Verlag, Berlin, 1989).

[45] Pesin, Ya.B., Russ. Math. Surv. **43**:4 (1988), 111-151.

[46] Hurewicz, W. and H. Wallman, *Dimension Theory*, *(Princeton, 1948).*

[47] Hentschel, H.G.E. and I. Procaccia, Physica D **8** (1983), 435-444.

[48] Rényi, A., *Introduction to information theory*, Appendix in: Probability Theory, (North-Holland, Amsterdam, 1970).

[49] Gershenfeld, N., in: Directions in Chaos Vol.2, Hao Bai-lin (ed.) (World Scientific, Singapur, 1988).

[50] Billingsley, P., *Ergodic Theory and Information*, (Wiley, New York, 1965).

[51] Sinai, Ya. G. (ed.), *Dynamical Systems II*, (Springer, New York, 1989).

[52] Evertsz, C.J.G., Y. Hayakawa and B.B. Mandelbrot, Phys. Rev. A **42** (1990), 4528-4536.

[53] Mandelbrot, B.B., in: 'Random Fluctuations and Pattern Growth', Vol. 157 of NATO Advanced Study Institute, Series E: Applied Science, edited by H.E. Stanley and D. Ostrowsky (Kluwer Academic, Dortrecht, 1988).

[54] Mandelbrot, B.B., Physica A **168** (1990), 95-111.

[55] McCauley, J.L., Phys. Rep. **189** (1990), 225-266.

[56] McCauley, J.L., *Chaos, dynamics and fractals: an algorithmic approach to deterministic chaos*, Cambridge Nonlinear Science Series **2**, (Cambridge University Press, 1993).

[57] Beck, C. and F. Schlögl, *Thermodynamics of chaotic systems: an introduction*, Cambridge Nonlinear Science Series **4**, (Cambridge University Press, 1993).

[58] Frisch, U. and G. Parisi, *in:* 'Turbulence and Predictability of Geophysical Fluid Dynamics and Climate Dynamics', Proc. of the International School of Physics "Enrico Fermi", Course LXXXVIII, Varenna 1983, edited by M. Ghil, R. Benzi and G. Parisi, (North-Holland, New York, 1985).

[59] Evertsz, C.J.G. and B.B. Mandelbrot, *Multifractal measures*, Appendix B in: H.O. Peitgen, H. Jürgens and D. Saupe: Chaos and Fractals (Springer, New York, 1992).

[60] Paladin, G. and A. Vulpiani, Phys. Rep. **156** (1987), 147-225.

[61] Tél, T., Z. Naturforsch. **43a** (1988), 1154-1174.

[62] Edgar, G.A. and R.D. Mauldin, Proc. London Math. Soc. **65** (1992), 604-628.

[63] Brown, G., G. Michon and J. Peyrière, J. Stat. Phys. **66** (1992), 775-790.

[64] Cawley, R. and R.D. Mauldin, Adv. in Math. **92** (1992), 196-236.

[65] Vaienti, S., J. Phys. A **21** (1988), 2313-2320.

[66] Bohr, T. and D. Rand, Physica D **25** (1987), 387-398.

[67] Feigenbaum, M.J., I. Procaccia and T. Tél, Phys. Rev. **A 39** (1989), 5359-5372.

[68] Mayer, D.H., Commun. Math. Phys. **130** (1990), 311-333.

[69] Mayer, D.H., *Thermodynamic Formalism and Quantum Mechanics on the Modular Surface*, preprint (1991).

[70] *Pollicott, M., Adv. in Math.* **85** *(1991), 161-192.*

[71] Adler, R.L. and L. Flatto, in: Ergodic Theory and dynamical systems II, (ed. A. Katok, Birkhäuser, Boston, 1982).

[72] Artin, E., Abh. Math. Sem. Univ. Hamburg **3** (1924), 170-175.

[73] Series, C.M., J. London Math. Soc. **31** (1985), 69-80.

[74] *Adler, R.L. and L. Flatto, Bull. Am. Math. Soc.* **25** *(1991), 229-334.*

[75] Khinchin, A.I., *Mathematical Foundations of Information Theory*, (Dover Publ., New York, 1957).

[76] Ruelle, D., Am. J. Math. **98** (1976), 619-653.

[77] Ruelle, D., Trans. Am. Math. Soc. **185** (1973), 237-251.

[78] Capocaccia, D., Commun. Math. Phys. **48** (1976), 85-88.

[79] Haydn, N.T.A., *Ergod. Th. & Dynam. Sys.* **7** *(1987), 119-132.*

[80] Haydn, N.T.A., Commun. Math. Phys. **134** (1990), 217-236.

[81] Haydn, N.T.A. and D. Ruelle, Commun. Math. Phys. **148** (1992), 155-167.

[82] Ruelle, D., Nonlinearity **5** (1992), 1223-1236.

[83] Israel, R.B., *Convexity in the Theory of Lattice Gases*, (Princeton Univ. Press, 1979).

[84] *Ruelle, D., Invent. Math.* **34** *(1976), 231-242.*

[85] Eckmann, J.P. and D. Ruelle, Rev. mod. Phys. **57** (1985), 617-656.

[86] Ruelle, D., *Chaotic evolution and strange attractors*, Notes prepared by S. Isola, (Cambridge Univ. Press, 1989).

[87] Ruelle, D., Commun. Math. Phys. **9** (1968), 267-278.

[88] Ruelle, D., *Statistical Mechanics: Rigorous Results*, *(Benjamin, New York, 1969).*

[89] Araki, H., Commun. Math. Phys. **14** (1969), 120-157.

[90] Simon, B., *The Statistical Mechanics of Lattice Gases I*, (Princeton University Press, 1993).

[91] Pollicott, M., Ergod. Th. & Dynam. Sys. **4** (1984), 135-146.

[92] Balazs, N.L. and A. Voros, Phys. Rep. **143** (1986), 109-240.

[93] Beardon, A.F., *The Geometry of Discrete Groups*, (Springer, New York, 1983).

[94] Berndt, R. and F. Steiner, *Hyperbolische Geometrie und Anwendungen in der Physik*, (Hamburger Beiträge zur Mathematik, Heft 8, Mathematisches Seminar der Universität Hamburg, 1989).

[95] Fischer, W. and I. Lieb, *Funktionentheorie*, (Vieweg, Braunschweig, 1980).

[96] O'Neill, B., *Semi-Riemannian Geometry, with Applications to Relativity*, (Academic Press, San Diego, 1983).

[97] Terras, A., *Harmonic Analysis on Symmetric Spaces and Applications I*, (Springer, New York, 1985).

[98] Venkov, A.B., Russ. Math. Surv. **34**:3 (1979), 79-153.

[99] Gutzwiller, M.C., *Chaos in Classical and Quantum Mechanics* (Springer, New York, 1990).

[100] Maskit, B., Adv. in Math. **7** (1971), 219-230.

[101] Aurich, R. and F. Steiner, Physica D **32** (1988), 451-460.

[102] Aurich, R. and F. Steiner, Physica D **39** (1989), 169-193.

[103] Aurich, R. and F. Steiner, Proc. R. Soc. Lond. A **437** (1992), 693-714.

[104] Aurich, R., E.B. Bogomolny and F. Steiner, Physica D **48** (1991), 91-101.

[105] Hejhal, D.A., *The Selberg Trace Formula for $PSL(2,\mathbb{Z})$ I*, (Springer, Berlin, 1976).

[106] Bogomolny, E.B. and M. Carioli, Physica D **67** (1993), 88-112.

[107] Kovács, Z. and T. Tél, Phys. Rev. A **45** (1992), 2270-2284.

[108] Bowen, R.E., Publ. Math. IHES **50** (1979), 11-25.

[109] Ruelle, D., Ergod. Th. & Dynam. Sys. **2** (1982), 99-107.

[110] Fleming, W.H., *Functions of Several Variables*, 2nd. Ed. (Springer, New York, 1977).

[111] Young, L.-S., Ergod. Th. & Dynam. Sys. **2** (1982), 109-124.

[112] Berberian, S.K., *Measure and Integration*, (Macmillan, New York, 1965).

[113] Mañé, R., *Ergodic Theory and Differentiable Dynamics*, (Springer, New York, 1987).

[114] Hurt, N.E., Results in Mathematics **23** (1993), 55-120.

[115] Pollicott, M., Invent. math. **85** (1986) 147-164.

[116] Ruelle, D., Bull. Am. Math. Soc. **82** (1976), 153-156.

[117] Steiner, F., *Quantum Chaos*, in: 'Universität Hamburg 1994: Schlaglichter der Forschung zum 75. Jahrestag', R. Ansorge (ed.), published on the occasion of the 75th anniversary of the University of Hamburg, (Dietrich Reimer Verlag, Hamburg, 1994); DESY preprint 94-013.

[118] Artin, M. and B. Mazur, Ann. Math. **81** (1965), 82-99.

[119] Bowen, R.E. and O.E. Lanford III, Proc. Symp. pure math. **14** (1970), 43-49.

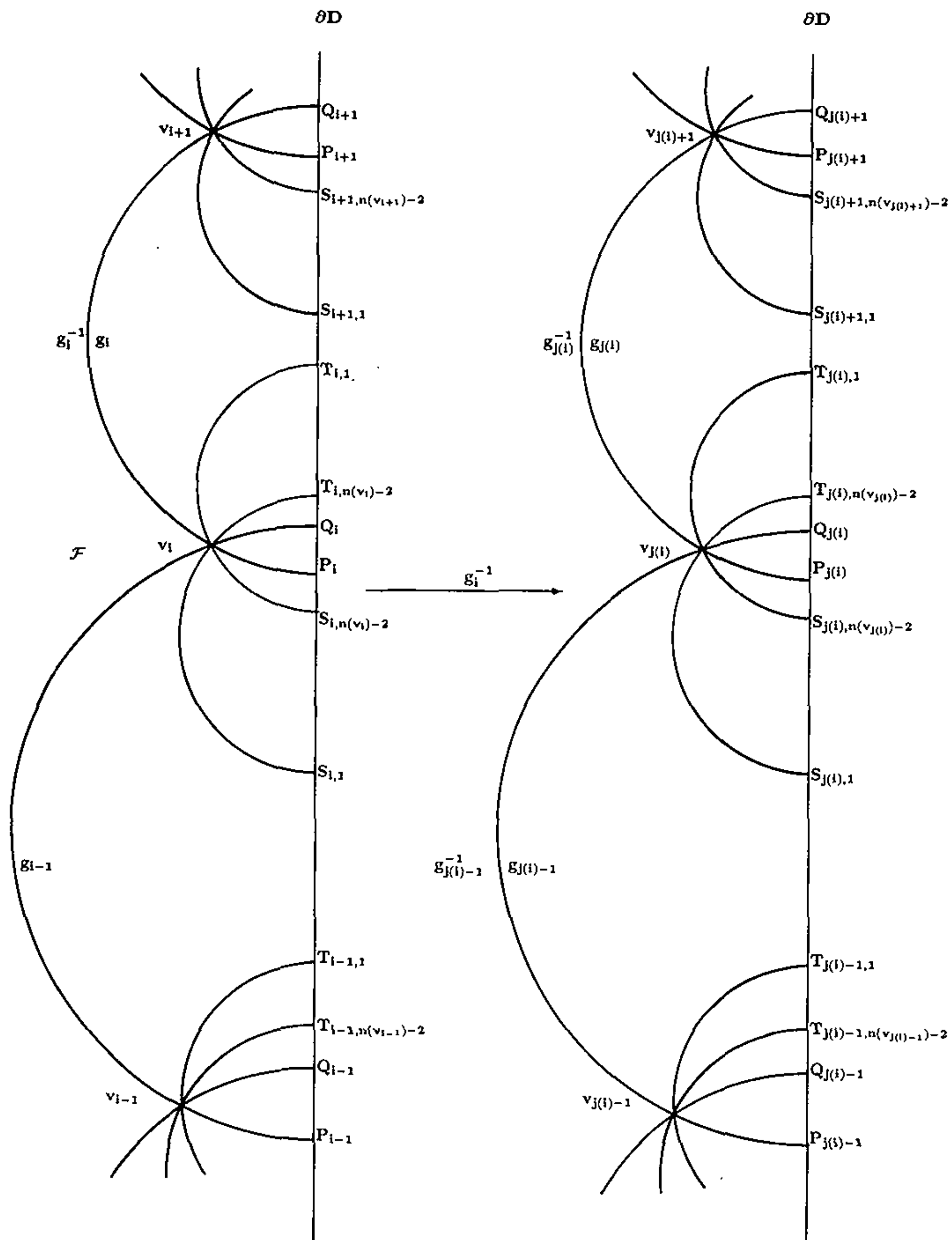

Figure 3.1