universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Rotation, entropy and equilibrium states"

verfasst von / submitted by

Thomas Reitsam, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc.)

New York, 2016

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066821
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Mathematik UG2002
Betreut von / Supervisor:	Professor C. Wolf, City College of New York

Contents

Abstract

For a continuous dynamical system (X, f) on a compact metric space and a continuous potential $\Phi : X \to \mathbb{R}^m$ the generalized rotation set is the subset of $\mathbb{R}^m$ consisting of all integrals of Φ with respect to all f-invariant probability measures. We provide an introduction to notions and results of rotation theory and entropy. For $\alpha \in \mathbb{R}^m$ and the potential $\alpha.\Phi$, following [6], we establish a connection between the rotation vectors of the equilibrium measures and the directional derivative of the pressure. The localized entropy at a point in the rotation set is defined as the supremum of the measure theoretic entropies over all f-invariant probability measures whose integral produce that point. We consider a subshift of finite type and a potential constant on cylinders of length K. Inspired by [19] we show that in this case the rotation set is a polyhedron and give formulas for the localized entropy at extreme points and faces of the rotation set.

1. Introduction

1.1. **Motivation.** A dynamical system is a rule for the evolution of time on a state space. Dynamical systems can be motivated by the real world like for example the motion of a particle in a liquid. It may happen that we have incomplete knowledge about the system. For instance we do not know the exact initial conditions. If a system is chaotic a small change in the initial conditions may impact the outcome. We equip the dynamical system with an invariant probability measure. An important goal in dynamical systems is to understand various typical dynamical behaviors of a given system.

1.2. **History.** The origin of dynamical systems comes from Newtonian physics. H. Poincarè is regarded as the "founder" of dynamical systems. He published two path-breaking works on celestial mechanics at the end of the 19th century and opened the way to modern nonlinear dynamics and chaos. Before 1960 strange and complicated behavior in deterministic systems was regarded as an anomaly but it was not considered to be actually regular or relevant. But with the availability of fast computers, chaotic behavior was recognized to be present in many systems of the real world. The analysis of chaotic phenomena requires different tools, namely rather analytic and especially measure theoretic than geometrical. This changed the perspective to consider the probabilities of outcomes, [2]. This sub-field of dynamics is called ergodic theory and has its actual origin earlier than 1960. It goes back to the famous ergodic hypothesis of Boltzmann who claimed equality of time average and space average for systems in statistical mechanics. Poincarè observed that the existence of a finite invariant measure leads to strong conclusions about recurrence. The development of the terminology of ergodic theory started around 1930. Important contributers are von Neumann, G. D. Birkhoff, E. Hopf, S. Katukani and A. Kolmogorov. Kolmogorov introduced the notion of entropy and brought ergodic theory in a more probabilistic background rather than a functional-analytic one, [7]. Additional to dynamical systems a central concept in this thesis is called rotation theory. It is developed from the idea of Poincarès rotation numbers for circles and later it is brought in the context of general dynamical systems.

1.3. **Set-up and main results.** We work in discrete time, deterministic dynamical systems equipped with at least one invariant probability measure, which means we need an sufficiently nice phase space. To make this more precise, we always consider a compact metrizable space X as the phase space and a continuous map $f : X \to X$ describing the evolution of time. By a dynamical system we mean a tuple (X, f) of such a phase space and a transformation. Invariance of the measure means that the distribution of a probability measure μ does not change under f, i.e. $\mu \circ f^{-1} = \mu$. Sometimes we use one invariant probability measure and sometimes we rather work with a set of invariant probability measures, denoted by $\mathcal{M}(f)$. As the set $\mathcal{M}(f)$ may be rather large for many systems, the question arises as to which f-invariant measure is the natural choice. One idea that might make sense is to consider a measure that maximizes a certain topological complexity, i.e. entropy, among all invariant measures. If such a measure exists, it is called a measure of maximal entropy. The problem may occur that the restriction to one invariant measure implies the loss of other relevant dynamical information. To get on top of this issue we need to get deeper into the set of invariant measures and one way is to work with smaller partitions, [10]. This idea brings us to the generalized rotation set. For a better understanding we explain the development from Poincarè's rotation number to the generalized rotation set, following Misiurewicz [13]. Therefore we consider the circle $\mathbb{T} = \mathbb{R} \backslash \mathbb{Z}$ with the natural projection $\pi : \mathbb{R} \to \mathbb{T}$. Suppose $f : \mathbb{T} \to \mathbb{T}$ is continuous. Then there exists a continuous map $F : \mathbb{R} \to \mathbb{R}$, called lifting of f, such that the diagram

$$\begin{tikzcd} \mathbb{R} \arrow[r, "F"] \arrow[d, "\pi"'] & \mathbb{R} \arrow[d, "\pi"] \\ \mathbb{T} \arrow[r, "f"'] & \mathbb{T} \end{tikzcd}$$

commutes. F is unique up to translation by an integer, i.e. $\tilde{F}(x) = F(x) + k$ for some $k \in \mathbb{Z}$. For now it is sufficient to look only at maps f with degree one. That is, $F(x+1) = F(x) + 1$, for all $x \in \mathbb{R}$. Note that the degree is independent of the choice of lifting. We denote the family of all liftings of degree one of f by $\mathcal{L}_1$. This family is closed under iterates of F and for $k \in \mathbb{Z}$ we have $F^n(x+k) = F^n(x) + k$. We define the upper and lower rotation number of $x \in \mathbb{R}$ for $F \in \mathcal{L}_1$ as

$$\overline{\rho}_F(x) = \varlimsup_{n\to\infty} \frac{F^n(x) - x}{n}, \quad \underline{\rho}_F(x) = \varliminf_{n\to\infty} \frac{F^n(x) - x}{n}. \tag{1}$$

If they coincide we define the rotation number of x for F as $\rho_F(x) = \overline{\rho}_F(x) = \underline{\rho}_F(x)$. The rotation number measures the average movement of any point under iteration of the map. For the first step of generalization we let $\mathbb{T}$ be a m-dimensional real torus. Poincarè proved that if f is an orientation preserving homeomorphism on a circle ($m = 1$) then all limits in (1) coincide and do not depend on x.
For the general theory let $F \in \mathcal{L}_1$ be a lifting of a circle map f. The so called displacement function $\varphi : \mathbb{T} \to \mathbb{R}$ is defined by $\varphi(x) = F(y) - y$, where $\pi(y) = x$. Note that φ is independent of the choice of $y \in \pi^{-1}(x)$. Then we have the telescopic sum

$$F^n(y) - y = \sum_{j=0}^{n-1} \varphi(f^j(x)) =: S_n(\varphi(x)).$$

Thus we obtain the following identity for the rotation number,

$$\rho_F(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} \varphi(f^j(x)). \tag{2}$$

This can again be generalized, namely, let X be a compact metric space, $f : X \to X$ a continuous map and $\Phi : X \to \mathbb{R}^m$ a Borel bounded (in most cases continuous) function, which is called the potential. Also we do not require x to be a fixed point but allow also sequences. In this set-up $\rho_{F,\Phi}(x)$ is called rotation vector instead of rotation number, i.e. if the limit,

$$\rho_{F,\Phi}(x) = \lim_{n_l\to\infty} \frac{1}{n_l} \sum_{j=0}^{n_l-1} \Phi(f^j(x_l)) \tag{3}$$

exists, it is called rotation vector. The set of all rotation vectors of this kind is called pointwise rotation set and denoted by $\mathrm{Rot}_{Pt}(\Phi)$, [10]. To get more information about this statistical average we recall Birkhoff's Ergodic Theorem which establishes Boltzmann's hypothesis. For the theorem let $\mathcal{B}$ be the Borel σ-algebra on X and μ be an f-invariant probability measure.

Definition 1.1. *Let (X, f, μ) be a dynamical system. We call μ (resp. (f, μ)) ergodic if all $B \in \mathcal{B}$ with $f^{-1}(B) = B$ satisfy $\mu(B) \in \{0, 1\}$. By $\mathcal{M}_E \subset \mathcal{M}(f)$ we denote the set of all ergodic measures.*

Theorem 1.2. *Let $f : X \to X$ be a continuous transformation and X be a compact metrizable space. Suppose $f \in L^1(\mu)$ for $\mu \in \mathcal{M}(f)$. Then $(1/n)S_n(\varphi(x))$ converges a.e. to a function $\varphi^* \in L^1(\mu)$. Also $\varphi^* \circ f = \varphi^*$ and $\int \varphi^*\, d\mu = \int \varphi\, d\mu$.*
If (X, f, μ) is ergodic φ^* is constant a.e., in particular $\varphi^* = \int \varphi\, d\mu$ a.e.

For the proof and further information we refer to [16, §1.6]. For $\Phi = (\phi_1, \ldots, \phi_m) : X \to \mathbb{R}^m$ in equation (3) we obtain by the Birkhoff Ergodic

Theorem

$$\frac{1}{n}S_n(\Phi(x)) \to \int \Phi \, d\mu \quad \mu - a.e.,$$

where both sides are m-dimensional vectors. The integral on the right hand side is called rotation vector and denoted by $\mathrm{rv}(\mu)$. The set $\mathrm{Rot}_E(\Phi) = \{\mathrm{rv}(\mu) : \mu \in \mathcal{M}_E\}$ is called the ergodic rotation set of Φ with respect to f. In the last step we do not ask for μ to be ergodic anymore but only f-invariant and define the general rotation set to be $\mathrm{Rot}(\Phi) = \{\mathrm{rv}(\mu) : \mu \in \mathcal{M}(f)\}$. From these three definitions of rotation sets we get the following inclusions:

$$\mathrm{Rot}_E(\Phi) \subset \mathrm{Rot}_{Pt}(\Phi) \subset \mathrm{Rot}(\Phi). \tag{4}$$

First note that both inclusions can be strict. The left-hand side inclusion in equation (4) is a consequence of the Birkhoff Ergodic Theorem. The right-hand side inclusion in equation (4) follows from the Banach-Alaoglu theorem and the sequentially compactness of the set of Borel probability measures. Consider a sequence of measures supported on the corresponding averages in equation (3). In general these measures are not f-invariant. However, any accumulation point is, and the integral over this measure coincides with the statistical limit. These relations are studied by T. Kucherenko and C. Wolf in [9].
But it can be proven that the convex hull of the ergodic and the pointwise rotation sets coincide with the generalized rotation set.

Proposition 1.3. *Let (X, f) be a dynamical system and $\Phi : X \to \mathbb{R}^m$ be a continuous map (X compact). Then*

$$\mathrm{convRot}_E(\Phi) = \mathrm{convRot}_{Pt}(\Phi) = \mathrm{Rot}(\Phi).$$

The proof uses the Ergodic Decomposition Theorem (Section 2.5) to show that for all extreme points $w \in \mathrm{Rot}(\Phi)$ there exists an ergodic measure $\mu \in \mathcal{M}_E(f)$ with $\mathrm{rv}(\mu) = w$.
For various dynamical systems the pointwise rotation set and the generalized rotation set coincide. One example for which they coincide are topologically mixing subshifts of finite type (Section 2.6). It is natural to ask how the generalized rotation set looks like. There is only limited knowledge about the shape of rotation sets. However, K. Ziemian proved in [19] that for subshifts of finite type and potentials which are constant on cylinders of length two, the rotation set is a polyhedron. We generalize this result to cylinders of length K. This result is crucial for the research part in this thesis. Also T. Kucherenko and C. Wolf, [10] gave an example of a rotation set for which they were able to determine the exact shape.
Next, we will measure the complexity of a system. Topological entropy is an invariant to quantify the complexity of a system. Roughly speaking, topological entropy measures the exponential growth rate of the number of different orbits as we increase the length of the orbits, see also Section 2.2. Furthermore it is possible to define a measure theoretic entropy $h_\mu(f)$

for an f-invariant probability measure μ. Roughly speaking, it measures the complexity of the system by ignoring μ-null sets. These two entropy functions are connected by the Variational Principle 2.19. Coming back to our original approach - a better understanding of the set $\mathcal{M}(f)$. We define the localized entropy for $w \in \mathrm{Rot}(\Phi)$ by $h(w) = \sup\{h_\mu(f) : \mu \in \mathcal{M}(f),\ \mathrm{rv}(\mu) = w\}$, following [6]. This notion is strongly related to the idea of a measure of maximal entropy with the difference that we here only consider invariant measures giving a certain rotation vector w. Again one can ask if such a measure exists and if it is unique.
For a dynamical system (X, f) and a continuous function $\varphi : X \to \mathbb{R}$ we define the pressure of φ by

$$P(\varphi) = \sup_{\mu \in \mathcal{M}(f)} \Big(h_\mu(f) + \int \varphi \, d\mu\Big). \tag{5}$$

If there exists a measure $\nu \in \mathcal{M}(f)$ with $h_\nu(f) + \int \varphi d\nu = P(\varphi)$ we call it equilibrium state. The set of all equilibrium states is denoted by ES_φ. For a $\alpha \in \mathbb{R}^m$ and $\Phi : X \to \mathbb{R}^m$ we define $p : \mathbb{R}^m \to \mathbb{R}$ by $p(\alpha) = P(\alpha.\Phi)$. The set of all subdifferentials of p with respect to α is denoted by $\partial p(\alpha)$, see Section 3 for more information. One of the main goals is to work out the proof of the following theorem from Jenkinson's [6].

Theorem 1.4. *Let (X, f) be a dynamical system for which entropy map $\mu \to h_\mu(f)$ is upper semi-continuous and suppose $\Phi : X \to \mathbb{R}^m$ is continuous. Then for $\alpha \in \mathbb{R}^m$ we have $\partial p(\alpha) = \mathrm{rv}(ES_{\alpha.\Phi})$.*

In Section 5 we discuss other main results. This section is also the basis for a paper we will write to extend these results. First we show that for potentials constant on cylinders of length K the corresponding rotation set is a polyhedron. This proof is inspired by Ziemians idea from [19]. In Theorem 5.3 and Theorem 5.4 we give formulas for the localized entropy of extreme points respectively of rotation vectors on faces.

2. Preliminaries

2.1. **Fundamentals.** After giving a broad overview of the objects we will now give precise definitions and discuss the essential background material. The fundamental object in this thesis is the notion of a dynamical system. Recall that throughout the thesis by a dynamical system we mean a tuple (X, f) of a continuous map $f : X \to X$ on a metrizable compact space X, where X is called the phase space of the dynamical system. There are other definitions of a dynamical system but we always assume that f is continuous and X is compact without mentioning it explicitly. The map f is deterministic and we consider systems discrete in time. As we will frequently use the notion of "chaotic" systems we clarify what we mean by that and give a definition. We follow [3].

Definition 2.1. *A point $x \in X$ is called periodic if there exists a $p \geq 1$ such that $f^p(x) = x$. We say x is of period p if p is the smallest such number. A point of period one is called a fixed point of f.*
The set $O_f(x) = \{f^n(x) : n \in \mathbb{Z}\}$ is called orbit of $x \in X$ with respect to f.

Definition 2.2. *A dynamical system (X, f) is called topologically mixing if for any two open non-empty sets $U, V \subset X$ there exists a positive integer $N = N(U, V)$ such that for every $n \geq N$ the intersection $f^n(U) \cap V$ is non-empty.*

Definition 2.3. *A dynamical system (X, f) is called topologically transitive if there exists a point $x \in X$ such that its orbit $O_f(x)$ is dense in X.*

This definition can be also characterized by the following lemma. We do not provide a proof but refer to [7, Lemma 1.4.2].

Lemma 2.4. *The dynamical system (X, f) is topologically transitive if and only if for any two nonempty open sets $U, V \subset X$ there exists an integer $N = N(U, V)$ such that $f^N(U) \cap V$ is nonempty.*

We previously mentioned that chaos is strongly related to sensitive dependence on initial conditions. The next definition clarifies the exact meaning of this.

Definition 2.5. *A dynamical system (X, f) is said to have sensitive dependence on initial conditions if there exists a $\delta > 0$, such that for any $x \in X$ and any $\varepsilon > 0$, there exists $y \in X$ with $d(x, y) < \varepsilon$ and $n \in \mathbb{N}$ such that $d(f^n(x), f^n(y)) > \delta$.*

The following definition is related.

Definition 2.6. *A dynamical system (X, f) is called expansive if there exists $\delta > 0$ such that, for any $x, y \in X$, with $x \neq y$, there exists $n \in \mathbb{N}$ such that $d(f^n(x), f^n(y)) > \delta$. If f is a homeomorphism, (X, f) is called expansive if there exits $\delta > 0$ such that, for any $x, y \in X$, with $x \neq y$, there exists $n \in \mathbb{Z}$ such that $d(f^n(x), f^n(y)) > \delta$.*

There are various ways to define what chaos means, however we go with the following definition of [3]. In Section 2.2 we remark on an alternative definition of chaos.

Definition 2.7. *We say a dynamical system (X, f) is chaotic on X if*

1. *f has sensitive dependence on initial conditions.*
2. *f is topologically transitive.*
3. *Periodic points are dense in X.*

This definition is motivated by saying that it is reasonable for a chaotic map to satisfy the following conditions: unpredictability, indecomposability, and an element of regularity. The system is unpredictable as it sensitively depends on initial conditions. It cannot be decomposed in two subsystems because of the topologically transitivity. But in contrast to a random process we have the regular condition of periodic points being dense.

2.2. **Entropy.** As the level of chaos may differ regarding different systems we want to measure how chaotic a system is. The key notion here is entropy. We will define topological entropy and measure theoretic entropy, respectively Kolmogorov-Sinai Entropy. Although we rather work with measure theoretic entropy we also will introduce topological entropy properly. For more details we refer to [7].
The goal is to obtain an invariant which represents the exponential growth rate of orbit segments distinguishable with arbitrary fine precision. We consider the dynamical system (X, f) and define the so called Bowen metric $d_n^f = d_n$ with respect to the metric d on X by

$$d_n(x,y) = \max_{0 \le i \le n-1} d(f^i(x), f^i(y)). \tag{6}$$

Therefore the new metric d_n measures the distance between the orbit segments $I_n^x = \{x, \dots, f^{n-1}(x)\}$ and I_n^y. Note that d_n induces the same topology on X as d. We denote the open ball around x with respect to d_n by $B_f(x, \varepsilon, n)$, and call $E \subset X$ (n, ε)-spanning if $X \subset \bigcup_{x \in E} B_f(x, \varepsilon, n)$. Let $S_d(\varepsilon, n)$ denote the minimal cardinality of a (n, ε)-spanning set. To give some intuition, this quantity gives the minimal number of initial conditions we need to approximate the behavior of any initial condition up to time n with precision ε. In this context we define the exponential growth rate by

$$h(f, \varepsilon) = \overline{\lim_{n \to \infty}} \frac{1}{n} \log S_d(\varepsilon, n). \tag{7}$$

Observe that $h(f, \varepsilon)$ is non-decreasing as a function of ε and so we can define the topological entropy as followed.

Definition 2.8. *The topological entropy $h(f)$ is defined by*

$$h(f) = \lim_{\varepsilon \to 0} h(f, \varepsilon). \tag{8}$$

Remark 2.9. *At this point it is not clear that the topological entropy does not depend on the metric d on X. However one can show [7, Proposition 3.1.2] that for another metric d' on X which defines the same topology as d, the topological entropy coincides.*

We will now discuss an alternative way to define topological entropy since the concept it uses is important. The definition is via the numbers $N_d(\varepsilon, n)$, the maximal number of points in X with pairwise d_n-distances at least ε. Such sets are called (n, ε)-separated. Those points generate the maximal number of orbit segments of length n that are distinguishable with precision ε. To connect $N_d(\varepsilon, n)$ to $S_d(\varepsilon, n)$ we observe that a maximal (n, ε)-separated set is an (n, ε)-spanning set, because otherwise it would be possible to increase the set by adding any point not covered. Thus

$$N_d(\varepsilon, n) \ge S_d(\varepsilon, n).$$

On the other hand, no ε-ball can contain two points 2ε apart. Thus

$$S_d(\varepsilon, n) \ge N_d(2\varepsilon, n).$$

Skipping some details we obtain

$$h(f) = \lim_{\varepsilon \to 0} \overline{\lim_{n \to \infty}} \frac{1}{n} \log S_d(\varepsilon, n) = \lim_{\varepsilon \to 0} \underline{\lim_{n \to \infty}} \frac{1}{n} \log N_d(\varepsilon, n). \tag{9}$$

For the measure theoretic entropy we need a measure on the dynamical system. In the context of measures we also need a σ-algebra $\mathcal{B}$ which we always choose to be the Borel σ-algebra on X. Note that the map $f : X \to X$ is $\mathcal{B}$-measurable since f is continuous.

Definition 2.10. *Let $\mathcal{M}$ denote the set of all (Borel)-probability measures on X. A measure $\mu \in \mathcal{M}$ is called f-invariant if for all $B \in \mathcal{B}$, $\mu(f^{-1}B) = \mu(B)$ holds true. The set of all f-invariant measures is denoted by $\mathcal{M}(f)$. Sometimes we say the system is measure preserving with respect to μ.*

Lemma 2.11. *For $\mu, \nu \in \mathcal{M}(f)$ we can find $\mu', \nu' \in \mathcal{M}(f)$ and $\lambda \in [0,1]$ such that $\nu = \lambda\nu' + (1-\lambda)\mu'$ and $\mu' \ll \mu \perp \nu'$.*

Proof. Let $\nu, \mu \in \mathcal{M}(f)$. By Lebesgue's Decomposition we can find two measures $\nu_1, \nu_2 \in \mathcal{M}$ such that $\nu = \nu_1 + \nu_2$ and $\nu_1 \ll \mu \perp \nu_2$. It might happen that one of the latter measures is simply the 0-measure, w.l.o.g. let ν' be 0. Then set $\lambda = 1$ and take any measure $\tilde{\nu} \in \mathcal{M}(f)$ which is singular to μ and we are done. Analogously, we proceed if μ' is the 0-measure. From now on we assume neither of the two measures is the 0-measure. We have to show that ν_1, ν_2 are f-invariant. For that, let $A \in \mathcal{B}$ and let u be the density function of ν_1 with respect to μ, note that u exists by Radon-Nikodym. Then we have

$$\nu_1(A) = \int\limits_A u \, d\mu = \int\limits_A u \, d(\mu \circ f^{-1}) = \int\limits_{f^{-1}A} u \, d\mu = \nu_1(f^{-1}),$$

where we use that $\mu \in \mathcal{M}(f)$ and hence ν_1 is f-invariant. For ν_2 we observe

$$\nu_2(A) = \nu(A) - \nu_1(A) = \nu(f^{-1}A) - \nu_1(f^{-1}A) = \nu_2(f^{-1}A).$$

As we now know that ν_1 and ν_2 are both f-invariant and since $\nu_i(X) \leq 1$ for $i = 1, 2$ we can find $\lambda \in (0,1)$ such that $\frac{1}{\lambda}\nu_1(X) = 1$ and $\frac{1}{1-\lambda}\nu_2(X) = 1$. Indeed, one can easily check, that $\lambda = \nu_1(X)$. To complete the proof we define

$$\mu' = \frac{1}{\lambda}\nu_1,$$
$$\nu' = \frac{1}{1-\lambda}\nu_2.$$

We go back to the Lebesgue Decomposition and obtain

$$\nu = \nu_1 + \nu_2 = \frac{\lambda}{\lambda}\nu_1 + \frac{1-\lambda}{1-\lambda}\nu_2 = \lambda\mu' + (1-\lambda)\nu_2.$$

Note that by construction $\mu', \nu' \in \mathcal{M}(f)$ and $\lambda \in (0,1)$. $\square$

The f-invariant measures are essential for describing the statistical properties of the dynamical system. They also play a fundamental role in the definition of rotation sets. Later we prove that in our set-up there is always a f-invariant measure. In order to define the measure theoretic entropy we start with a probability space $(X, \mathcal{B}, \mu)$ and a finite partition ξ of measurable sets. In this case entropy quantifies the amount of knowledge gained when learning about $\xi(x)$. The higher the entropy, the higher the uncertainty.

Definition 2.12. *Let $\xi = \{Z_1, \ldots, Z_n\}$ be a finite partition of X. Then*

$$H_\mu(\xi) = -\sum_{i=1}^{n} \mu(Z_i) \log \mu(Z_i), \tag{10}$$

where $\mu(Z_i) \log \mu(Z_i) = 0$ if $\mu(Z_i) = 0$, is called entropy of ξ.

For the origin of the idea we may consider the amount of yes-or-no questions you have to ask to know in which partition the element x lies. For the next step we extend the probability space to a dynamical system, i.e. we add a function $f : X \to X$ such that the system is measure preserving, and define the average knowledge gained per iteration for ξ as

$$\frac{1}{n} H_\mu(\xi_n) = \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} f^{-i}\xi \right), \tag{11}$$

where

$$\xi_1 \vee \xi_2 = \{Z_1 \cap Z_2 : Z_1 \in \xi_1,\ Z_2 \in \xi_2\}.$$

Definition 2.13. *The entropy for a dynamical system (X, f) with f-invariant probability measure μ with respect to ξ is defined by*

$$h_\mu(f, \xi) = \lim_{n\to\infty} \frac{1}{n} H_\mu(\xi_n). \tag{12}$$

To make this definition consistent one has to prove that the limit in (12) exists. At this point we refer to [16]. Finally we can define the measure theoretic respectively Kolmogorov-Sinai entropy.

Definition 2.14. *The measure theoretic entropy of a measure preserving dynamical system (X, f, μ) is defined by*

$$h_\mu(f) = \sup_{\xi \text{ finite partition of } X} h_\mu(f, \xi). \tag{13}$$

Note that all of the foregoing definitions make also sense if ξ is a countable partition and $H(\xi) < \infty$. In the following we frequently refer to a dynamical system (X, f, μ) where (X, f) is a dynamical system and μ is a f-invariant measure. We will provide propositions that will allow us to work with entropy more easily. To make it easier to deal with the measure theoretic entropy we provide some properties.

Definition 2.15. *Let $\xi = \{C_\alpha : \alpha \in I\}$, $\eta = \{D_\alpha : \alpha \in J\}$ be two measurable partitions of the probability space $(X, \mathcal{B}, \mu)$. The conditional entropy of ξ with respect to η is*

$$H(\xi \mid \eta) = -\sum_{\beta \in J} \mu(D_\beta) \sum_{\alpha \in I} \mu(C_\alpha \mid D_\beta) \log \mu(C_\alpha \mid D_\beta).$$

From this definition we can define a metric on the measurable partitions of finite entropy. The metric is called Rokhlin metric. For two measurable partitions ξ, η with $H(\xi) < \infty$ and $H(\eta) < \infty$ we define the metric by

$$d_R(\xi, \eta) = H(\xi \mid \eta) + H(\eta \mid \xi). \tag{14}$$

Definition 2.16. *Let (X, f) be a dynamical system and ξ be a countable partition.*

1. *For a non-invertible f, ξ is called a generator if the partitions of the form $\bigvee_{i=0}^{k} f^{-i}(\xi)$ ($k \in \mathbb{N}$) form a dense subset in the space of all partitions with finite entropy with respect to the Rokhlin metric* (14).
2. *For invertible f, ξ is called a generator if for the partitions of the form $\bigvee_{i=-k}^{k} f^{i}(\xi)$ ($k \in \mathbb{N}$) the same holds.*

Corollary 2.17. *If ξ is a generator for f then $h_\mu(f) = h_\mu(f, \xi)$.*

Furthermore we have a relation for the entropy of iterations of f, namely:

Lemma 2.18. *For $\mu \in \mathcal{M}(f)$ and $k \in \mathbb{N}$, we have $h_\mu(f^k) = k h_\mu(f)$. If f is invertible then $h_\mu(f^{-1}) = h_\mu(f)$ and hence $h_\mu(f^k) = |k| h_\mu(f)$ for any $k \in \mathbb{Z}$.*

There is also a connection between the topological and the measure theoretic entropy, namely the variational principle. It gives equality of the topological entropy and the supremum of all measure theoretic entropies. In 1968, L. W. Goodwyn proved topological entropy is greater or equal. Then in 1970 E. L. Dinaburg proved equality when X has finite covering dimension. T. N. T. Goodman proved the general case later in 1970. We refer to [16, Theorem 8.6] for the proof.

Theorem 2.19. *Let (X, f) be a dynamical system. Then*

$$h(f) = \sup\{h_\mu(f) : \mu \in \mathcal{M}(f)\}. \tag{15}$$

A measure μ satisfying $h_\mu(f) = h(f)$ is called measure of maximal entropy. As the measure theoretic entropy map $\mu \to h_\mu(f)$ is not necessarily upper semi-continuous there is not always a measure of maximal entropy. Suppose $Y = \{0\} \cup \{(1/n) : n \geq 1\}$ with topology as a subset of $\mathbb{R}$. Let $X = \prod Y$ be the product space and let $f : X \to X$ be the shift homeomorphism (Section 2.6). Let μ_j be the product measure obtained from the measure on Y that gives measure $\frac{1}{2}$ to each of the points $1/(j-1)$ and $1/j$. Then the measure preserving transformation f on $(X, \mathcal{B}, \mu_j)$ is conjugate (Definition 2.36) to the two-sided $(\frac{1}{2}, \frac{1}{2})$-shift and hence $\mathrm{h}_{\mu_j}(f) = \log 2$.

However, $\mu_j \to \mu$ where μ is the atomic measure on X that gives measure 1 to the point $(\dots, 0, 0, 0, \dots)$. Clearly $h_\mu(f) = 0$ so that the entropy map of f is not upper semi-continuous, [16]. To ensure the existence of such a measure we sometimes consider dynamical systems for which the measure theoretic entropy is upper semi-continuous, i.e. $\overline{\lim}_{\mu \to \mu_0} h_\mu(f) = h_{\mu_0}(f)$. It is well known that upper semi-continuous functions attain their maxima on compact spaces. This seems to be a strong restriction, however the following theorem, [16, Theorem 8.2] provides a list of transformations for which the entropy map is always upper semi-continuous.

Theorem 2.20. *Let (X, f) be a dynamical system. If f is an expansive homeomorphism of a compact metric space the measure theoretic entropy map $\mu \to h_\mu(f)$ of f is upper semi-continuous.*

An alternative way of saying a dynamical system is chaotic is if it has positive topological entropy or equivalently there exists at least one f-invariant measure and there is $\mu \in \mathcal{M}(f)$ with positive measure theoretic entropy. This makes perfectly sense since we defined entropy as measure of chaos and positive entropy clearly represents the existence of chaos.

Lemma 2.21. *Let $\mu \in \mathcal{M}$, then the following are equivalent:*

1. *(f, μ) is ergodic (Definition 1.1).*
2. *If $\varphi \in \mathcal{L}^p(X, \mu)$ is f-invariant, $p \geq 1$, then φ is constant μ-a.e.*

Proof. "1⇒2" Let $\varphi \in \mathcal{L}^p$ be f-invariant and define the set $A_c = \{x \in X : \varphi(x) \leq c\}$. Since φ is f-invariant, A_c is f-invariant as well. As we assumed (f, μ) is ergodic that means $\mu(A_c) \in \{0, 1\}$. Hence φ is constant μ a.e.
"2⇒1" Let $A \in \mathcal{B}$ be f-invariant. Therefore $\mathbb{1}_A \in \mathcal{L}^p$ is f-invariant. Since by assumption $\mathbb{1}_A$ is constant μ a.e. we know $\mu(A) \in \{0, 1\}$ and thus (f, μ) is ergodic. □

2.3. **Rotation sets.** After discussing the motivation of rotation sets in the introduction we recall the definition here.

Definition 2.22. *Let $\Phi = (\phi_1, \dots, \phi_m) : X \to \mathbb{R}^m$ be a continuous potential. We define the map $\mathrm{rv} : \mathcal{M}(f) \to \mathbb{R}^m$ by*

$$\mathrm{rv}(\mu) = \left(\int \phi_1 \, d\mu, \dots, \int \phi_m \, d\mu \right), \tag{16}$$

and call $\mathrm{rv}(\mu)$ the rotation vector of μ. The (generalized) rotation set $\mathrm{Rot}(\Phi)$ is defined as the image of all invariant probability measures under $\mathrm{rv}(\cdot)$, i.e.

$$\mathrm{Rot}(\Phi) = \big\{ \mathrm{rv}(\mu) : \mu \in \mathcal{M}(f) \big\}. \tag{17}$$

For $w \in \mathrm{Rot}(\Phi)$ we say that $\mathrm{rv}^{-1}(w)$ is the rotation class of w.

We divide the set of invariant measures into a filtration of partitions. This works as follows. Consider a sequence $(\phi_k)_{k \in \mathbb{N}}$ of continuous potentials that is dense in the Banach space $C(X)$ endowed with the supremum norm. Let

$\mathrm{Rot}(n)$ denote the rotation set of the potential $\Phi_n = (\phi_1, \ldots, \phi_n)$. It follows that

$$\mathrm{Rot}(n+1) = \bigcup_{w \in \mathrm{Rot}(n)} w \times I_w,$$

where I_w is a compact interval defined by

$$I_w = \left\{ \int \phi_{n+1} \, d\mu : \mu \text{ is } f\text{-invariant and } \mathrm{rv}_{\Phi_n} = w \right\}.$$

We say a sequence $(w_n)_{w_n \in \mathrm{Rot}(n)}$ is decreasing if $w_{n+1} = w_n \times \{\alpha_n\}$ for some $\alpha_n \in I_{w_n}$ and all $n \in \mathbb{N}$. It follows that every decreasing sequence $(w_n)_n$ is associated with a decreasing sequence of rotation classes. Moreover, by the Riesz Representation theorem the intersection of the rotation classes contains precisely one invariant measure μ_∞. We say $(\mathrm{Rot}(n))_n$ is a filtration of the space of invariant probability measures. Thus, if n is large we can consider $\mathrm{Rot}(n)$ as a fairly good approximation of the space of invariant probability measures, [10]. As we fix n we obtain $\mathcal{M}(f) = \bigcup_{w \in \mathrm{Rot}(\Phi)} \mathrm{rv}^{-1}(w)$. Related to the measure theoretic entropy we define the localized entropy following [6] on each element of the partition.

Definition 2.23. *The localized entropy for $w \in \mathrm{Rot}(\Phi)$ is defined as*

$$h(w) = \sup\{h_\mu(f) : \mu \in \mathcal{M}(f) \text{ and } \mathrm{rv}(\mu) = w\}. \tag{18}$$

One of the main goals in this thesis is to find a formula to calculate the localized entropy explicitly. Therefore in Section 5 in this thesis we will work with localized entropy.

2.4. **Set structures.** After providing these basic definitions we will now work out some more details about the set structures. We do not provide proofs for all lemmata and theorems, but refer to [16] for more details.

Lemma 2.24. *Let $\mu, \nu \in \mathcal{M}$ be two probability measures on X. If $\int \varphi \, d\mu = \int \varphi \, d\nu$ for all $\varphi \in C(X)$ then $\mu = \nu$.*

Theorem 2.25 (Riesz). *Let X be a compact metric space and $J : C(X) \to \mathbb{R}$ a continuous linear map, such that J is a positive operator (i.e. $\varphi \geq 0 \Rightarrow J(\varphi) \geq 0$) and $J(1) = 1$. Then there exists a unique $\mu \in \mathcal{M}$ such that $J(\varphi) = \int \varphi \, d\mu$, for all $\varphi \in C(X)$.*

Remark 2.26. *So $\mu \to J$ is a bijection between $\mathcal{M}$ and the collection of all normalized positive linear functionals on $C(X)$. As this bijection is affine, $\mathcal{M}$ can be identified with a convex subset of the unit ball of $C^*(X)$, the dual space of $C(X)$. Therefore we get a topology on $\mathcal{M}$ from the weak*-topology on $C^*(X)$.*

Definition 2.27. *The weak*-topology on $\mathcal{M}$ is the smallest topology, such that $\mu \to \int \varphi \, d\mu$, $\varphi \in C(X)$ is continuous.*

Furthermore we can also define a metric which is given by the following lemma and so $\mathcal{M}$ is metrizable.

Lemma 2.28. *If X is a compact metrizable space, the set $\mathcal{M}$ is metrizable in the weak*-topology. Let $\{\varphi_n\}_{n\geq 1} \subset C(X)$ be a dense countable subset, and define the metric*

$$D(\mu,\nu) = \sum_{n\geq 1} \frac{|\int \varphi_n \, d\mu - \int \varphi_n \, d\nu|}{2^n\|\varphi_n\|}.$$

Remark 2.29.

- *$\mu_n \to \mu$ in the weak*-topology in $\mathcal{M}$ if and only if $\int \varphi \, d\mu_n \to \int \varphi \, d\mu$, for all $\varphi \in C(X)$.*
- *For all $A \in \mathcal{A}$, with $\mu(\partial A) = 0$, $\mu_n(A) \to \mu(A)$ if and only if $\mu_n \to \mu$ in the weak*-topology.*

Theorem 2.30 (Theorem 6.5)**.** *If X is a compact metrizable space, then $\mathcal{M}$ is compact in the weak*-topology.*

Proof. Let us denote $\mu(\varphi) = \int \varphi \, d\mu$. Let $\{\mu_n\}_{n\geq 1}$ be a sequence in $\mathcal{M}$. We need to show that there exists a convergent subsequence. Choose $\{\varphi_n\}_{n\geq 1} \subset C(X)$ countable and dense. Here we used that $C(X)$ is separable if X is compact. We apply Cantor's diagonal trick. The sequence $\{\mu_n(\varphi_1)\}_n$ is bounded by $\|\varphi_1\|_\infty$ and hence there exists a convergent subsequence $\{\mu_n^1(\varphi)\}_n$. We do the same for $\{\mu_n^1(\varphi_2)\}_n$, which is again bounded by $\|\varphi_2\|_\infty$ and hence has also a convergent subsequence. We continue inductively and obtain a series of subsequences $\{\mu_n^i\}_i$ such that $\{\mu_n^i(\varphi)\}_n \subset \{\mu_n^{i-1}(\varphi)\}_n \subset \cdots \subset \{\mu_n(\varphi)\}_n$ is converging for $\varphi \in \{\varphi_1, \ldots, \varphi_i\}$. Therefore the diagonal subsequence $\{\mu_n^n(\varphi_i)\}_n$ converges for all $i \in \mathbb{N}$. Since $\{\varphi_i\}_i$ is dense, $\{\mu_n^n(\varphi)\}_n$ converges for all $\varphi \in C(X)$. Therefore we can define $J(\varphi) = \lim_{n\to\infty} \mu_n^n(\varphi)$. One can easily show that J is in $C^*(X)$, normalized and hence there exists $\mu \in \mathcal{M}$, such that $\mu(\varphi) = J(\varphi)$, for all $\varphi \in C(X)$. □

We rather consider $\mathcal{M}(f)$ so it is good to know also more about this space. For the next theorem we define the operator $\tilde{f} : \mathcal{M} \to \mathcal{M}$ by $\tilde{f}(\mu) = \mu \circ f^{-1}$.

Theorem 2.31. *Let X be a compact metrizable space, and $\{\sigma_n\}_{n\geq 1}$ be a sequence in $\mathcal{M}$ and define the sequence $\{\mu_n\}_{n\geq 1}$ by*

$$\mu_n = \frac{1}{n}\sum_{k=0}^{n-1} \tilde{f}^k \sigma_n.$$

Then every accumulation point μ of $\{\mu_n\}_n$ is in $\mathcal{M}(f)$.

Proof. Since $\mathcal{M}$ is compact there exists at least one limit point for every sequence. We now fix one accumulation point $\mu = \lim_{k\to\infty} \mu_{n_k} \in \mathcal{M}$ and show that μ is f-invariant. Let $\varphi \in C(X)$. Then

$$\begin{aligned}\Big|\int \varphi\circ f\ d\mu - \int \varphi\ d\mu\Big| &= \lim_{k\to\infty}\Big|\int \varphi\circ f\ d\mu_{n_k} - \int \varphi\ d\mu_{n_k}\Big| \\ &= \lim_{k\to\infty}\Big|\frac{1}{n_k}\int \sum_{i=0}^{n_k}(\varphi\circ f^{i+1} - \varphi\circ f^i)\ d\sigma_{n_k}\Big| \\ &= \lim_{k\to\infty}\Big|\frac{1}{n_k}\int (\varphi\circ f^{n_k} - \varphi)\ d\sigma_{n_k}\Big| \\ &\le \lim_{k\to\infty}\frac{2\|\varphi\|}{n_k} = 0.\end{aligned}$$

And hence $\mu \in \mathcal{M}(f)$. $\square$

One consequence of this theorem is that $\mathcal{M}(f)$ is non-empty, as $\mathcal{M}$ is compact. This is one of the reasons that we assume X to be compact. Although everything we need from convex analysis can be found in the appendix, we state the definition of extreme points here for convenience. A more detailed list from convex analysis can be found in the appendix.

Definition 2.32. *Let C be a convex set. A point $x \in C$ is an extreme point if for all $y, z \in C$ and all $\lambda \in (0,1)$, $\lambda y + (1-\lambda)z = x$ implies that $x = y = z$.*

The next theorem provides information about the structure of $\mathcal{M}(f)$.

Theorem 2.33. *Let (X, f) be a dynamical system, then*

1. *$\mathcal{M}(f)$ is compact.*
2. *$\mathcal{M}(f)$ is convex.*
3. *μ is an extreme point of $\mathcal{M}(f)$ if and only if μ is ergodic with respect to f.*
4. *If $\mu, \nu \in \mathcal{M}(f)$ are both ergodic and $\mu \neq \nu$ then they are mutually singular.*

Proof. 1. For compactness take a sequence $\{\mu_n\}_n \subset \mathcal{M}(f)$. As the sequence is also in the compact set $\mathcal{M}$, there exists a subsequence converging to $\mu \in \mathcal{M}$. We need to show that μ is f-invariant. Take $\varphi \in C(X)$,

$$\begin{aligned}\int \varphi\ d\tilde{f}\mu = \int \varphi\circ f\ d\mu &= \lim_{n\to\infty}\int \varphi\circ f\ d\mu_n \\ &= \lim_{n\to\infty}\int \varphi\ d\mu_n = \int \varphi\ d\mu.\end{aligned}$$

Thus μ is f-invariant and therefore $\mu \in \mathcal{M}(f)$.
2. For convexity let $\mu, \nu \in \mathcal{M}(f)$, $\lambda \in (0,1)$ and $B \in \mathcal{B}$. Then

$$\lambda\mu(B) + (1-\lambda)\nu(B) = \lambda\mu(f^{-1}(B)) + (1-\lambda)\nu(f^{-1}(B)).$$

Obviously the convex combination is still a probability measure and hence $\mathcal{M}(f)$ convex.

3. We first assume $\mu \in \mathcal{M}(f)$ is an extreme point. Let $f^{-1}(A) = A$ and $\mu(A) > 0$. We will show that $\mu(A) = 1$. Define

$$\nu(B) = \frac{\mu(A \cap B)}{\mu(A)}.$$

Note that also $f^{-1}(A^c) = A^c$ and assume $\mu(A) < 1$. We define

$$\sigma(B) = \frac{\mu(A^c \cap B)}{\mu(A^c)}.$$

By using the f-invariance of μ we know that ν and σ are f-invariant and thus we obtain $\mu = \mu(A)\nu + (1-\mu(A))\sigma$, as we assumed $\mu(A) \in (0,1)$. But this is a contradiction to the extreme point property and therefore $\mu(A) \in \{0,1\}$, which implies that μ is ergodic.
For the reverse direction we assume $\mu \in \mathcal{M}(f)$ is ergodic and $\mu = \lambda\nu + (1-\lambda)\sigma$ for $\nu, \sigma \in \mathcal{M}(f)$ and $\lambda \in (0,1)$. We must show $\mu = \nu = \sigma$. First we observe that ν is absolute continuous with respect to μ ($\nu \ll \mu$) such that there exists the Radon-Nikodym derivative $\frac{d\nu}{d\mu} \geq 0$, i.e. for all $B \in \mathcal{B}$

$$\nu(B) = \int\limits_B \frac{d\nu}{d\mu}\, d\mu.$$

Let

$$B = \left\{ x : \frac{d\nu}{d\mu}(x) < 1 \right\}. \tag{19}$$

We obtain

$$\int\limits_{B \cap f^{-1}B} \frac{d\nu}{d\mu}\, d\mu + \int\limits_{B \setminus f^{-1}B} \frac{d\nu}{d\mu}\, d\mu = \nu(B) = \nu(f^{-1}B)$$
$$\int\limits_{B \cap f^{-1}B} \frac{d\nu}{d\mu}\, d\mu + \int\limits_{f^{-1}B \setminus B} \frac{d\nu}{d\mu}\, d\mu,$$

and thus

$$\int\limits_{B \setminus f^{-1}B} \frac{d\nu}{d\mu}\, d\mu = \int\limits_{f^{-1}B \setminus B} \frac{d\nu}{d\mu}\, d\mu.$$

The Radon-Nikodym derivative is strictly smaller 1 on the left hand side and greater or equal 1 on the right hand side. A short calculation gives

$$\mu(f^{-1}B \setminus B) = \mu(f^{-1}B) - \mu(f^{-1}B \cap B) =$$
$$\mu(B) - \mu(f^{-1}B \cap B) = \mu(B \setminus f^{-1}B),$$

and hence $\mu(f^{-1}B \setminus B) = \mu(B \setminus f^{-1}B) = 0$. This implies $\mu((f^{-1}B \setminus B) \cup (B \setminus f^{-1}B)) = 0$ what means $\mu(B) \in \{0,1\}$. Suppose $\mu(B) = 1$, then

$$\nu(X) = \int\limits_B \frac{d\nu}{d\mu}\, d\mu < \mu(B) = 1$$

contradicting $\nu \in \mathcal{M}(f)$ and so $\mu(B) = 0$. Similarly for

$$\tilde{B} = \left\{ x : \frac{d\nu}{d\mu}(x) > 1 \right\},$$

we obtain $\mu(\tilde{B}) = 0$ such that $\frac{d\nu}{d\mu} = 1$ a.e. Hence we obtain $\nu = \mu = \sigma$ and so μ is an extreme point of $\mathcal{M}(f)$.
4. By the Lebesgue decomposition theorem there exist unique probability measures μ_1, μ_2 and a unique $\lambda \in [0,1]$ such that $\mu = \lambda\mu + (1-\lambda)\mu_2$ where $\mu_1 \ll \nu$ and $\mu_2 \perp \nu$. But since $\mu = \tilde{f}\mu = \lambda\tilde{f}\mu_1 + (1-\lambda)\tilde{f}\mu_2$ and $\tilde{f}\mu_1 \ll \tilde{f}\nu = \nu$ and $\tilde{f}\mu_2$ singular with respect to $\tilde{f}\nu = \nu$ the uniqueness of the decomposition implies $\mu_1, \mu_2 \in \mathcal{M}(f)$. Since μ is an extreme point either $\lambda = 0$ or $\lambda = 1$. If $\lambda = 1$ we have $\mu = \mu_1 \ll \nu$ and as in iii) we get $\mu = \nu$ which contradicts to our assumption. If $\lambda = 0$ we get $\mu = \mu_2 \perp \nu$ which completes the proof. □

2.5. **Ergodic Decomposition.** The Ergodic Decomposition Theorem is a classical theorem which gives probability measures on the set of f-invariant probability measures supported on ergodic measures. The theorem is an application of the more general Choquet's theorem. This tool can be applied to represent measures by ergodic measures and also to see that there are ergodic measures contained in specific sets, for instance in the set of equilibrium states. We do not provide a proof but refer to [14]for further information.

Theorem 2.34. *Suppose that X is a metrizable compact convex subset of a locally convex space M, and that x_0 is an element of X. Then there is a probability measure ν on X which represents x_0 and is supported by the extreme points of X, denoted by $E(X)$, i.e. $x_0 = \int_X x \, d\mu(x) = \int_{E(X)} x \, d\mu(x)$.*

It is easy to apply this theorem to our special case to directly obtain the Ergodic Decomposition theorem. It is well known that $\mathcal{M}$ is locally convex and we proved that $\mathcal{M}(f)$ is convex, compact and ergodic measures are exactly the extreme points, Theorem 2.33. In general the representation is not necessarily unique, however, the last part of Theorem 2.33 ensures that the representation in this case is indeed unique. A space where the decomposition is unique for every element is called Choquet simplex. Thus for every $\mu \in \mathcal{M}(f)$ we can find a unique probability measure ν on $\mathcal{M}(f)$, such that $\mu = \int_{\mathcal{M}(f)} m \, d\nu(m) = \int_{\mathcal{M}_E} m \, d\nu(m)$.

2.6. **Subshift of finite type.** A classical and important class of dynamical systems is the subshift of finite type. In various cases we will work in the set-up of a subshift of finite type. It seems to be rather restrictive to consider only subshifts of finite type but we will see that it actually is not. Later in this section we will explain why it is not such a big restriction and why it is so important. This subsection follows the exposition in [8]. Consider a finite set $\mathcal{A} = \{1, \ldots, d\}$. We call $\mathcal{A}$ alphabet. We equip $\mathcal{A}$ with the discrete metric, that is two distinct points have distance one and a point is distance zero from itself, i.e. $d(j,k) = 1 - \delta_{jk}$ where $\delta_{jk} = 0$ if $j \neq k$ and 1 otherwise. The topology induced by this metric is compact and has discrete topology. There are two different sequence spaces. The first one is the one-sided sequence space $X = \{x = (x_n)_{n\in\mathbb{N}} : x_n \in \mathcal{A}\ \forall n \in \mathbb{N}\} = \{1,\dots,d\}^{\mathbb{N}}$. Please note that by $\mathbb{N}$ we understand the set $\{1,2,3,\dots\}$, i.e. zero is not included unless explicitly indicated. The second one is the two-sided sequence space $\Sigma = \{x = (x_n)_{n\in\mathbb{Z}} : x_n \in \mathcal{A}\ \forall n \in \mathbb{Z}\} = \{1,\dots,d\}^{\mathbb{Z}}$. Both spaces are equipped with the product topology and by Tychonoffs Theorem for products of compact spaces both spaces are compact. The product metric naturally obtained is defined by

$$d(x,y) = \sum_{i=0}^{\infty} \frac{1-\delta_{x_i y_i}}{2^i}$$

on X and analogously on Σ

$$d(x,y) = \sum_{i=-\infty}^{\infty} \frac{1-\delta_{x_i y_i}}{2^{|i|}}.$$

Two equivalent metrics we will use are defined by

$$d(x,y) = \left(\frac{1}{2}\right)^{\min\{i\,:\,x_i \neq y_i\}} \tag{20}$$

for X, respectively for Σ by

$$d(x,y) = \left(\frac{1}{2}\right)^{\min\{|i|:x_i \neq y_i\}}. \tag{21}$$

A cylinder is a set of the form $[x_1,\dots,x_l] = \{y \in X : x_i = y_i\ \forall i \leq l\}$ respectively $[x_{-l},\dots,x_0,\dots,x_l] = \{y \in \Sigma : x_i = y_i\ \forall |i| \leq l\}$. These cylinders build a countable basis for the topology on each space. Therefore every open set can be expressed by a countable union of cylinder sets. In order to get a dynamical system we introduce the shift map $\sigma : X \to X$ respectively $\sigma : \Sigma \to \Sigma$ defined by $(\sigma(x))_n = x_{n+1}$. The tuple (X,σ) is called one-sided (full) shift. In this case σ is a continuous and onto, n-to-1 transformation. Similarly the tuple (Σ,σ) is called two-sided (full) shift. Here the shift map is a homeomorphism.

Sometimes we would like to consider only sequences of a specific type and not all sequences. One class of subshifts are the so-called subshifts of finite type. To determine valid sequences we use a matrix A consisting of zeros and ones, i.e. $A \in \{0,1\}^{d\times d}$. We say a step x_n to x_{n+1} is valid if and only if $A(x_n, x_{n+1}) = 1$. The new one-sided subshift of finite type (X_A,σ) is determined by the space $X_A = \{x \in X : A(x_n,x_{n+1}) = 1\ \forall n \in \mathbb{N}\}$. Analogously $\Sigma_A = \{x \in \Sigma : A(x_n,x_{n+1}) = 1\ \forall n \in \mathbb{Z}\}$ defines the two-sided subshift of finite type (Σ_A,σ). The matrix A is said to be irreducible if for every pair of indices i and j there is an $l > 0$ with $(A^l)_{ij} > 0$. Intuitively this means for a subshift of finite type that we can get from an arbitrary symbol in the

alphabet to any other symbol in finitely many steps. Note that the number of steps needed may differ depending on the pair of indices.
Periodic points play an important role for subshifts of finite type. There are only countably many periodic points and they are dense in the shift space and hence also in the subshift space [8],[16, §5]. This makes the shift space separable. For the shift a periodic point of period p is of the form $(\dots, x_p, x_1, x_2, \dots, x_p, x_1, \dots, x_p, x_1, \dots) = (x_1, \dots, x_p)$, such that $\sigma^p(x) = x$. The right hand side will be the usual notation for periodic points. Furthermore we can construct σ-invariant measures from periodic points, namely for a periodic point $x = (x_1, \dots, x_p)$

$$\mu_x = \frac{1}{p} \sum_{i=0}^{p-1} \delta_{\sigma^i(x)}, \tag{22}$$

where δ is the point measure. Measures of this type are called periodic point measures. For the relation between x being periodic and μ_x being σ-invariant we have the following theorem. As it holds in a more general set-up than shifts we state the general version of it.

Theorem 2.35. *Let (X, f) be a dynamical system. Let $N \geq 1$ and $x \in X$. Then $f^N(x) = x$ if and only if*

$$\mu_x = \frac{1}{N} \sum_{i=0}^{N-1} \delta_{f^i(x)} \in \mathcal{M}(f).$$

From the theorem we get an embedding of X in $\mathcal{M}(f)$. For transitive shifts we can also conclude that the periodic point measures are dense in $\mathcal{M}(f)$.
Subshifts have a lot of convenient properties but seem to be very restrictive as it is one particular dynamical system. An important tool for dynamical systems is topological conjugacy. That basically means to find another (easier) dynamical system which is in some sense "equivalent" to the original one. At this point it is not clear what "equivalent" means. But there are some things that immediately come to our mind which we ask for to call another system equivalent. For instance we would like to have the same periodic points, the same invariant probability measures, and same entropy. Also there should exist a map connecting both systems. This idea leads to the following definition which is not only important for subshifts but for all dynamical systems.

Definition 2.36. *Two continuous maps $f : X \to X$ and $g : Y \to Y$ are called (topologically) conjugate if there is a homeomorphism ψ such that*

$f = \psi^{-1} \circ g \circ \psi$, i.e. the following diagram

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "\psi"'] & X \arrow[d, "\psi"] \\ Y \arrow[r, "g"'] & Y \end{tikzcd}$$

commutes. The map ψ is called conjugacy.

A little bit weaker but still useful concept is called semi-conjugacy respectively factor.

Definition 2.37. *A map $g : Y \to Y$ is a (topological) factor of $f : X \to X$ if there exists a surjective continuous map $\psi : X \to Y$ such that $\psi \circ f = g \circ \psi$. The map ψ is called semiconjugacy.*

In Section 5 we will use the idea of conjugates for two subshifts of finite type. For now we just briefly explain the connection between general dynamical systems and subshifts of finite type. For the idea of coding in a dynamical system (X, f, μ) we take a countable μ-partition ξ of X. Instead of looking at a certain x, we look at the partition element $\xi(x)$ it is contained. Let $\xi = \{Z_i : i \in I\}$, then we define

$$d(x) = i \text{ if } \xi(x) = Z_i.$$

We define the map $D : X \to I^{\mathbb{N}_0}$ by

$$Dx = (d(x), d(f(x)), \dots)$$

and is called coding map. We obtain the commuting diagram

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "D"'] & X \arrow[d, "D"] \\ I^{\mathbb{N}_0} \arrow[r, "\sigma"'] & I^{\mathbb{N}_0} \end{tikzcd}$$

and so D is a semi-conjugacy. This concept is particularly useful if D is invertible. In this case the first n coding numbers tell us in which set of

$$\xi_n = \bigvee_{j=0}^{n-1} f^{-j}\xi$$

x lies. If the fineness of ξ_n increases, $\xi_n(x) \to \{x\}$ holds and so D is invertible. This is for example the case if f has enough "expansion". For more details we refer to [7, §2].
Besides the fact that subshifts look simple they have convenient properties such as the formula for the measure theoretical entropy.

Definition 2.38. *Let (X,σ) be a full shift. For a probability distribution $p=(p_1,\dots,p_d)$, where $0\le p_i\le 1$ for $i=1,\dots,d$ and $\sum_{i=1}^d p_i=1$, the product measure μ_p is given by the probabilities on cylinders*

$$\mu(C_k[a_1,\dots,a_k])=\prod_{i=1}^k p_{a_i},$$

and is called Bernoulli measure.

For a d-full-shift the measure theoretical entropy of a Bernoulli measure μ given by the probability vector $(p_1,\dots,p_d)$ is simply given by

$$h_\mu(\sigma)=-\sum_{i=1}^d p_i\log p_i. \tag{23}$$

There is even more that holds true for full shifts. We state a lemma observed by Rams, [15], which will play a crucial role in the main theorems on localized entropy. Here C_i denotes the i-th cylinder of length 1.

Lemma 2.39. *Let (X,σ) be a full shift. The measure theoretic entropy of $\mu\in\mathcal{M}_E(\sigma)$ is not greater than the metric entropy of the Bernoulli measure defined by the probabilistic vector $(\mu(C_1),\dots,\mu(C_d))$, with equality if and only if μ is Bernoulli.*

Definition 2.40. *Let (X,σ) be a subshift of finite type. A Markov measure is defined by a tuple (π,P), where $\pi=(\pi_i)_i$ with $\pi_i\ge 0$ and $\sum_{i=1}^d\pi_i=1$ is the initial distribution and $P=(p_{ij})_{ij}$ is a $d\times d$ transition matrix such that $\sum_{j=1}^d p_{ij}=1$. The Markov measure ν is given by the probabilities on cylinders,*

$$\mu(C_k[a_1,\dots,a_k])=\pi_{a_1}p_{a_1a_2}\cdots p_{a_{k-1}a_k}.$$

For a Markov measure ν given by (π,P) the measure theoretical entropy is given by

$$h_\nu(\sigma)=-\sum_{i,j=1}^d\pi_i p_{ij}\log p_{ij}. \tag{24}$$

This formula is more general and can also be applied for subshifts of finite type. We refer to [7], [16] for more details. Furthermore topological entropy coincides with the logarithm of the spectral radius of the transition matrix A, which can also be found in [7]. As the concept of subshifts is a well-known standard tool we skipped details and proofs.

2.7. **Cohomology.** Coming back to rotation sets we consider a potential $\Phi: X\to\mathbb{R}^m$. To make not only vacuous statements we want $\mathrm{Rot}(\Phi)$ to be a non-empty subset of $\mathbb{R}^m$ with interior. By that we mean $\mathrm{Rot}(\Phi)$ has interior in $\mathbb{R}^m$, i.e. $\mathrm{int}(\mathrm{Rot}(\Phi))\ne\varnothing$. For the definition of interior and relative interior please see Appendix, section A.2. To ensure this we introduce the terminology of cohomology in the set-up of a subshift of finite type (X_A,f).

We say two real-valued functions $\varphi, \psi : X \to \mathbb{R}$ are essentially cohomologous if there exists a bounded Borel measurable function $u : X \to \mathbb{R}$ and some constant $c \in \mathbb{R}$ such that

$$\varphi - \psi = u \circ f - u + c. \tag{25}$$

The equation in (25) is called cohomological equation. If $c = 0$ we say φ and ψ are cohomologous. We say that φ is an essential coboundary if it is cohomologous to a constant. A vector valued function $\Phi : X \to \mathbb{R}^m$ is called cohomologically full if its coordinate functions $\phi_1, \dots, \phi_m$ are cohomologically independent, i.e. if all non-trivial linear combinations of $\phi_1, \dots, \phi_m$ are not essential coboundaries. We will often assume that the function Φ is cohomologically full to ensure that $\mathrm{Rot}(\Phi)$ has interior as a subset of $\mathbb{R}^m$. This statement is neither trivial nor obvious and therefore we have to do some work in order to establish it. But first of all we mention that the assumption of Φ being cohomologically full is not a restriction, since we can always take a maximal cohomologically independent subset and consider the potential $\tilde{\Phi} : X \to \mathbb{R}^{m'}$ for some $m' \leq m$.
We use Proposition 4.5 in Bowen's book [1] for the case of Hölder-continuous functions. For the following proposition μ_φ denotes the Gibbs measure with respect to φ, see [1] for more information. Although we only need part of the proposition we state all points of the equivalence for completeness.

Proposition 2.41. *Let $\varphi, \psi : X_A \to \mathbb{R}$ be two Hölder-continuous functions. Then the following are equivalent:*

1. $\mu_\varphi = \mu_\psi$.
2. *There is a Hölder function $u : X_A \to \mathbb{R}$ and a constant C so that $\varphi - \psi = C + u \circ f - u$.*
3. *There are constants C and L so that $|S_n\varphi(x) - S_n\psi(x) - Cn| \leq L$ for all $x \in X_A$ and all $n \geq 0$.*
4. *There is a constant C so that $S_n\varphi(x) - S_n\psi(x) = Cn$ when $x \in X_A$ with $f^n(x) = x$.*

To explain the connection of being cohomologically full and having interior as a subset of $\mathbb{R}^m$ for $\Phi : X_A \to \mathbb{R}^m$ we assume for simplicity that $m = 2$, i.e. $\Phi = (\phi_1, \phi_2)$. Higher dimensions follow immediately by using similar arguments. If $\varphi = \lambda_1\phi_1$ and $\psi = \lambda_2\phi_2$ for $\lambda_i \in \mathbb{R}$, condition 2. is equivalent to cohomological dependence of ϕ_1 and ϕ_2, i.e. Φ is not cohomologically full. By the foregoing proposition this is equivalent to condition 4., namely there is a constant such that $S_n\varphi(x) - S_n\psi(x) = Cn$ for all points x of period n and all $n \in \mathbb{N}$. For the following corollary we refer to [8].

Corollary 2.42. *For a transitive subshift of finite type periodic points measures are dense and all periodic point measures are in particular ergodic.*

Dividing the equation from 4. by n and gives

$$\lambda_1 \int \phi_1 \, d\mu - \lambda_2 \int \phi_2 \, d\mu = C, \tag{26}$$

for all ergodic measures μ. This shows us that the two integrals lie on one line segment for all ergodic measures. Since the ergodic measures are dense this means $\dim(\mathrm{Rot}(\Phi)) = 1$ and has therefore no interior in $\mathbb{R}^2$.
For the converse we assume Φ is cohomologically full which is equivalent to condition 2. does not hold and by the proposition it is also equivalent to condition 4. does not hold. So no matter which $\lambda's$ we choose, we can never find just one constant C. This implies that there are ergodic measures lying on different lines and hence $\mathrm{Rot}(\Phi)$ has an interior as subset of $\mathbb{R}^2$.
So far Proposition 2.41 holds only for Hölder-continuous functions but in fact it holds in a little more general set-up, see [6].

2.8. **Example of maximal entropy.** We will construct a subshift with two measures of maximal entropy. We follow Haydn, [5] Therefore let $n \geq 2$ be an integer and define the alphabet $\mathcal{A} = \{0, 1, \dots, 2n\}$. Furthermore we define the sets of symbols $S_1 = \{1, \dots, n\}$ and $S_2 = \{n+1, \dots, 2n\}$, such that we have $\mathcal{A} = \{0\} \cup S_1 \cup S_2$. For simplicity we refer to S_1 as green symbols and to S_2 as yellow symbols. We denote by X the whole sequence space, i.e. $X = \mathcal{A}^{\mathbb{Z}}$ and $X_1 = S_1^{\mathbb{Z}}$, respectively $X_2 = S_2^{\mathbb{Z}}$ the subshift with alphabet S_1, respectively S_2. We will construct our subshift $X(\tau) \subset X$. We let $X(\tau)$ be the union of the two spaces S_1 and S_2, i.e. all monochromatic sequences. Furthermore we add colored sequences in a specific way, namely a word α in one color is separated by a string of zeros γ to a word β of the other color. The length $|\gamma|$ of that string has to be at least $\tau(|\alpha|+|\beta|)$, where $\tau > 0$. If α or β is infinitely long then there must be a infinite sequence of zeros and so there is only one color. To make this more precise we state a formal definition of $X(\tau)$.

Definition 2.43. *Let $\tau > 0$. An element x is in the shift space $X(\tau) \subset X$ if it satisfies one of the following conditions*

1. *$x \in X_1 \cup X_2$, i.e. x is a monochromatic point.*
2. *If bi-colored blocks of x are of the form:*

$$\dots 0 g_1 g_2 \dots g_a 0^{\lambda} y_1 y_2 \dots y_b 0 \dots$$

or

$$0 y_1 y_2 \dots y_a 0^{\lambda} g_1 g_2 \dots g_b 0 \dots$$

where $\lambda \geq \tau(a+b)$, $g_i \in S_1$, $y_i \in S_2$, and 0^{λ} is a string of zeros of length λ.

For an allowed word $\omega = \omega_m \omega_{m+1} \dots \omega_{m+k-1}$ of some length k in $X(\tau)$ we denote by $C(\omega)$ the cylinder set $\{x \in X(\tau) : x_i = \omega_i,\ m \leq i < m+k\}$. The shift transformation $f = \sigma_{|X(\tau)}$ on $X(\tau)$ is defined by $(f(x))_i = x_{i+1},\ i \in \mathbb{Z}$ and $x \in X(\tau)$. Having defined the set-up we go step by step to obtain two ergodic measures of maximal entropy.

Lemma 2.44. *The shift $(X(\tau), f)$ is topologically mixing for every $\tau > 0$.*

Proof. We have to show that for any two words ω and η there exists a number N_0 such that $C(\omega) \cap f^N(C(\eta))$ is non-empty for all $N \geq N_0$. It is enough to show this property on cylinders as they are a generator on the shift space. Tautologically this is true if ω and η are both monochromatic of the same color. If the last symbol of η and the first symbol of ω are of different color then consider the word $\pi = \eta 0^\lambda \omega$. Obviously π is allowed for every $\lambda > \tau(|\eta| + |\omega|)$, and the cylinder set $C(\pi) \subset C(\omega) \cap f^N(C(\eta))$ is non-empty for all $N \geq N_0$, where we choose N_0 to be $(1+\tau)(|\eta| + |\omega|)$.
If the last symbol of η and the first symbol of ω are of the same color, then we consider the word $\pi = \eta 0^\kappa \varepsilon 0^\lambda \omega$, where ε is a symbol of the other color. Here we have $\kappa \geq \tau(|\eta| + |\varepsilon|)$ and $\lambda \geq \tau(|\omega| + |\varepsilon|)$. Clearly $C(\pi)$ is non-empty for these choices of λ and κ. Thus $C(\omega) \cap f^N(C(\eta)) \neq \varnothing$ if $N \geq N_0 = (\tau+1)(|\eta| + 2|\varepsilon| + |\omega|)$. □

Lemma 2.45. *If $\tau > \frac{\log 3}{\log n}$ then the topological entropy $h(f)$ of the shift $(X(\tau), f)$ is equal to $\log n$.*

Proof. First we find a lower bound of the topological entropy $h(f)$ of $X(\tau)$. For that we observe that since $X(\tau)$ contains two full n-shifts, X_1 and X_2, the topological entropy of $X(\tau)$ must be at least $\log n$.
We now estimate the number of the remaining words of length N from above by combinatorical methods. We have to consider two cases:

1. Monochromatic words that might also contain zeros. According to our rules such words begin or end with strings of zeros. Thus we obtain purely yellow or green words of lengths $k = 0, \dots, N$ which at least one side are framed by strings of zeros. By combinatorical arguments their number turn out to be $2\sum_{k=0}^{N}(N-k)n^k \leq 2Nn^N \sum_{k=0}^{N} n^{-k}$ which has exponential growth rate $\log n$.
2. To estimate the number of words of length N that genuinely contain symbols of both colors, we observe that since any such word has at least one transition from yellow to green or vice versa. Therefore it must also contain at least $N\frac{\tau}{\tau+1}$ zeros, that is at most $N' = \lfloor \frac{N}{\tau+1} \rfloor$, where $\lfloor \cdot \rfloor : \mathbb{R} \to \mathbb{N}$ gives the closest lower integer of a real number, colored symbols, i.e. symbols from S_1 and S_2. The colored symbols come in monochromatic blocks of alternating color. Denote by $P_{k,l}$ the number of possibilities in which we can arrange l symbols in k blocks (separated by the appropriate number of zeros), where $k = 1, \dots, l$. One can show that
$$P_{k,l} = \binom{l-1}{k-1}, \tag{27}$$
which is the number of possibilities of picking the first element of every block but the very first one. The number of N-words in $X(\tau)$ which contain l colored symbols arranged in $k \leq l$ monochromatic blocks of alternating colors is $2P_{k,l}n^l$.
Distributing l colored symbols out of N' in k blocks can be done in

$P_{k,l}$ many ways, where $1 \le k \le l \le N'$. This leaves $m = N - (\tau+1)l$ zeros to be distributed on $k+1$ intervals, namely the $k-1$ gaps between the blocks of colored symbols plus the two ends of the entire words. There are

$$Q_{k,m} = \binom{m+k}{k}$$

many possibilities. By Stirling's formula we have

$$\begin{aligned} Q_{k,m} &= \binom{N-(\tau+1)l+k}{k} \le \binom{N-(\tau+1)l+k}{\frac{1}{2}(N-(\tau+1)l+k)} \\ &\le c_1 2^{N-(\tau+1)l+k}\sqrt{N-(\tau+1)l+k} \le c_1 2^{N-(\tau+1)l+k}\sqrt{N}. \end{aligned}$$

We can thus estimate the total number of bi-colored strings of length N by

$$\begin{aligned} Q(N) &\le 2\sum_{l=2}^{N'} n^l \sum_{k=2}^{l} \binom{l-1}{k-1}\binom{N-(\tau+1)l+k}{k} \\ &\le 2c_1\sqrt{N}\sum_{l=2}^{N'} n^l 2^{N-(\tau+1)l}\sum_{k=2}^{l}\binom{l-1}{k-1}2^k \\ &\le 2c_1\sqrt{N}\sum_{l=2}^{N'} n^l 2^{N-(\tau+1)l}3^l \\ &\le c_2\sqrt{N}\begin{cases} 2^N & \text{if } 3n2^{-1-\tau} < 1 \\ 2^N(3n2^{-1-\tau})^{\frac{N}{\tau+1}} & \text{if } 3n2^{-1-\tau} \ge 1. \end{cases} \end{aligned}$$

Thus

$$\lim_{N\to\infty} \frac{1}{N}\log Q(N) \le \max\left(\log 2, \frac{\log 3n}{\tau+1}\right),$$

which is smaller or equal $\log n$ if $\tau \ge \frac{\log 3}{\log n}$.

$\square$

Lemma 2.46. *If $\tau \ge \frac{\log 3}{\log n}$, then there are two mutually singular measures of maximal entropy on $X(\tau)$.*

Proof. For a full n-shift $\{1,\dots,n\}$ the Bernoulli measure with probability vector $(\frac{1}{n},\dots,\frac{1}{n})$ is an ergodic measure of maximal entropy, [16]. Its metric entropy is $\log n$. On $X(\tau)$ we define the Bernoulli measures μ_1 with probabilities $(0,\frac{1}{n},\dots,\frac{1}{n},0,\dots,0)$ and the μ_2 with probabilities $(0,0,\dots,0,\frac{1}{n},\dots,\frac{1}{n})$. Obviously both measures are shift invariant and have metric entropies $\log n$ which by Lemma 2.45 is the topological entropy of $X(\tau)$. Hence μ_1 and μ_2 are distinct ergodic measures of maximal entropy for the subshift $X(\tau)$. $\square$

We will now construct an example of a potential $\Phi : X(\tau) \to \mathbb{R}^m$ which is continuous and such that the two mutually singular measures of maximal entropy from Lemma 2.46 have the same rotation vector but $\mathrm{Rot}(\Phi)$ is non-empty. The construction is inspired by [10].

Example 1. *We give the example for $m = 1$ but it can naturally be extended to higher dimensions. Let $Y(k) + \{x \in X(\tau) : x_l \in S_i \ \forall |l| \leq k, \ i = 1, 2\}$, for $k \geq 3$ and $Y_0 = X(\tau) \setminus \bigcup_{k \geq 3} Y(k)$. Then $\Phi : X(\tau) \to \mathbb{R}$ defined by*

$$\Phi(x) = \begin{cases} 0 & \text{if } x \in Y_0, \\ 1 - \frac{1}{2}^k & \text{if } x \in Y(k), \\ 1 & \text{if } x \in Y(k), \ \forall k \in \mathbb{N}. \end{cases}$$

It is obvious by symmetry that both measures from Lemma 2.46 have the same rotation vector. We can also observe that the rotation set is not just one point. Also continuity can be easily observed.

3. Equilibrium states and subgradients

The goal in this section is to work out details of a proof in Jenkinsons' paper [6]. We work again in the set-up of a dynamical system (X, f), not necessarily a subshift of finite type. First we have to introduce some terminology, starting with pressure. Similar to entropy there is a topological and measure theoretic way to define pressure and analogously the variational principle holds true. For $\varphi : X \to \mathbb{R}$ we define

$$S_n\varphi(x) = \sum_{i=0}^{n-1} \varphi(f^i(x)). \tag{28}$$

The topological pressure (with respect to f) is a map $P_{\text{top}}(\varphi) : C(X) \to \mathbb{R} \cup \{\infty\}$ defined by

$$P_{\text{top}}(\varphi) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log N_\varphi(n, \varepsilon), \tag{29}$$

where

$$N_\varphi(n, \varepsilon) = \sup \left\{ \sum_{x \in F} e^{S_n\varphi(x)} : F \text{ is } (n, \varepsilon)\text{-separated set} \right\}.$$

As already mentioned the pressure satisfies the variational principle, namely,

$$P_{\text{top}}(\varphi) = \sup_{\mu \in \mathcal{M}(f)} \left(h_\mu(f) + \int_X \varphi \, d\mu \right), \tag{30}$$

which brings us to the definition of pressure we will use.

Definition 3.1. *For a continuous function $\varphi : X \to \mathbb{R}$ we define its pressure $P(\varphi)$ with respect to f to be*

$$P(\varphi) = \sup_{\mu \in \mathcal{M}(f)} \left(h_\mu(f) + \int_X \varphi \, d\mu \right). \tag{31}$$

If $\mu \in \mathcal{M}(f)$ attains the supremum it is called an equilibrium state for φ. We denote the set of all equilibrium states with respect to φ by ES_φ.

Note that in general the set of equilibrium states of φ may be empty. To avoid this we frequently restrict our attention to systems with an upper semi-continuous entropy map. Since the entropy map is affine, the set of all equilibrium states is a convex and compact subset of $\mathcal{M}(f)$.

Definition 3.2. *A signed measure μ is called tangent functional of pressure at $\varphi \in C(X)$ if the following inequality holds for all $\psi \in C(X)$*

$$P(\varphi + \psi) - P(\varphi) \geq \int \psi \; d\mu. \tag{32}$$

The set of all tangent functionals of pressure at φ is denoted by t_φ

Similar to standard calculus we define directional derivatives for the pressure.

Definition 3.3. *For $\varphi, \psi \in C(X)$ we define the (one-sided) directional derivative of P at φ in direction ψ by*

$$d^+P(\varphi; \psi) = \lim_{t \downarrow 0^+} \frac{P(\varphi + t\psi) - P(\varphi)}{t}.$$

By convexity of P the map

$$t \to t^{-1}(P(\varphi + t\psi) - P(\varphi)) \tag{33}$$

is increasing in t for all $\varphi, \psi \in C(X)$, see Theorem A.5. This ensures that the directional derivative of P exists at every point. Let us now fix a potential $\Phi : X \to \mathbb{R}^m$ and consider the function $p : \mathbb{R}^m \to \mathbb{R}$ defined by $p(\alpha) = P(\alpha.\Phi)$, where $\alpha.\Phi$ denotes the Euclidean inner product. It is understood in a pointwise way for every $x \in X$. For p we can define the (one-sided) directional derivative in direction $\beta \in \mathbb{R}^m$ in the classical sense, namely,

$$d^+(\alpha; \beta) = \lim_{t \downarrow 0^+} \frac{p(\alpha + t\beta) - p(\alpha)}{t}. \tag{34}$$

Moreover $\gamma \in \mathbb{R}^m$ is called subgradient of p at the point α if

$$p(\alpha + \beta) - p(\alpha) \geq \gamma.\beta, \tag{35}$$

holds for all $\beta \in \mathbb{R}^m$. The set of all subgradients of p at α is called subdifferential of p at α and is denoted by $\partial p(\alpha)$. It is also well-known that p is differentiable at α if and only if the set of subdifferentials at α is a singleton, see Theorem A.6.

Jenkinson [6] proved that in the case of an upper semi-continuous entropy function the subdifferential of p at α coincides with $\mathrm{rv}(ES_{\alpha.\Phi})$. His proof was published in 2001. Recently T. Kucherenko and C. Wolf [9] used this result to prove that under certain conditions the map $w \to h(w)$ is real analytic in the interior of $\mathrm{Rot}(\Phi)$. Jenkinson bases his proof on three lemmata, which we will provide and prove here. The first lemma is pure convex analysis, giving a characterization for subdifferentials and can be found in [18].

Lemma 3.4. *Let $p : \mathbb{R}^m \to \mathbb{R}$ be a convex function which is finite at $\alpha \in \mathbb{R}^m$. Then $\beta \in \mathbb{R}^m$ is a subgradient at α if and only if $d^+p(\alpha;\beta) \geq \gamma.\beta$, for all $\beta \in \mathbb{R}^m$.*

Proof. For the proof we recall that by convexity of p the mapping $t \to t^{-1}(p(\alpha+\beta) - p(\alpha))$ is increasing for all $\alpha, \beta \in \mathbb{R}^m$.
For the first direction we assume $\gamma \in \mathbb{R}^m$ is a subgradient of p at α. Define $\tilde{\beta} = t\beta$ for $t > 0$ and $\beta \in \mathbb{R}^m$. Since γ is a subgradient by assumption the following inequality holds for all $\tilde{\beta} \in \mathbb{R}^m$,

$$p(\alpha + \tilde{\beta}) - p(\alpha) \geq \gamma.\tilde{\beta}.$$

Dividing by t leads to the equivalent inequality

$$\frac{p(\alpha + t\beta) - p(\alpha)}{t} \geq \gamma.\beta, \quad \forall \beta \in \mathbb{R}^m,\ \forall t > 0.$$

Since this inequality holds for all $t > 0$ and the left hand side is monotone in t we let t go to zero from above, i.e. $t \downarrow 0^+$, we obtain

$$d^+p(\alpha;\beta) \geq \gamma.\beta \quad \forall \beta \in \mathbb{R}^m. \tag{36}$$

For the converse we assume inequality (36) holds true. Again we use monotonicity in t, and thus we obtain for all $\beta \in \mathbb{R}^m$,

$$\gamma.\beta \leq d^+p(\alpha;\beta) = \lim_{t\downarrow 0^+} \frac{p(\alpha + t\beta) - p(\alpha)}{t} \leq \frac{p(\alpha + t\beta) - p(\alpha)}{t}, \quad \forall t > 0,$$

where the first inequality holds by assumption, the second one by definition and the last one by monotonicity in t. If we now define $\tilde{\beta} = \beta/t$ we get the result for all $\tilde{\beta} \in \mathbb{R}^m$. $\square$

The next lemma can be found in [17] and gives a relation between subdifferentials and tangent functionals of pressure.

Lemma 3.5. *For $\varphi, \psi \in C(X)$ we have $d^+P(\varphi,\psi) = \sup\{\int \psi \, d\mu : \mu \in t_\varphi\}$.*

Proof. "$\geq$": Let $\mu \in t_\varphi$, and define $\tilde{\psi} = t\psi$, for $t > 0$. Thus we have for all $\psi \in C(X)$

$$\int \psi \, d\mu \leq t^{-1}(P(\varphi + t\psi) - P(\varphi)), \quad \forall t > 0.$$

Again we use the fact that (33) is monotone in t, so we let $t \downarrow 0^+$, and obtain

$$\int \psi \, d\mu \leq d^+P(\varphi;\psi), \quad \forall \psi \in C(X).$$

As $\mu \in t_\varphi$ was arbitrary we can take the supremum and get the first direction.
"$\leq$": Let $a = d^+P(\varphi;\psi)$ and define the linear functional L on $\{t\psi : t \in \mathbb{R}\}$

by $L(t\psi) = ta$. By convexity of P and monotonicity of the function in (33) we obtain

$$L(t\psi) \leq P(\varphi + t\psi) - P(\varphi),$$

so that L is bounded on $\{t\psi : t \in \mathbb{R}\}$ by the convex function $y \to P(\varphi+y) - P(\varphi)$, in particular L is sublinear. Therefore we can apply the Hahn-Banach Theorem and hence L can be extended to a linear functional on $C(X)$, which is also bounded from above by $y \to P(\varphi + y) - P(\varphi)$. Applying the Riesz representation provides the existence of some signed measure μ with $\int \psi \, d\mu = L(\psi) = a$. Since L is bounded we obtain that $\mu \in t_\varphi$. $\square$

This lemma gives a characterization of d^+ with respect to the tangent functionals of pressure but actually we would like to have a connection between ES_φ and d^+. For that we need the next lemma from [16].

Lemma 3.6. *Let (X, f) be a dynamical system, i.e. f is a continuous map on a compact metrizable space X. Let $h_\mu(f) < \infty$, let $\mu_0 \in \mathcal{M}(f)$ and $\varphi \in C(X)$. Then the following holds true:*

1. *$ES_\varphi \subset t_\varphi \subset \mathcal{M}(f)$.*
2. *$h_{\mu_0}(f) = \inf\{P(\psi) - \int \psi \, d\mu_0 : \psi \in C(X)\}$ if and only if $\mu \to h_\mu(f)$ is upper semi-continuous.*
3. *If $\mu \to h_\mu(f)$ is upper semi-continuous, then $t_\varphi = ES_\varphi$.*

Proof. 1. Let $\mu \in ES_\varphi$ and $\psi \in C(X)$. It is easy to see that

$$P(\varphi + \psi) - P(\varphi) \geq h_\mu(f) + \int (\varphi + \psi) \, d\mu - h_\mu(f) - \int \varphi \, d\mu = \int \psi \, d\mu,$$

where we only used $\mu \in ES_\varphi$. And hence $ES_\varphi \subset t_\varphi$.
For the second inclusion let $\mu \in t_\varphi$ and suppose $\psi \geq 0$. We need to show that μ is a f-invariant probability measure. For $\varepsilon > 0$ we obtain

$$\begin{aligned}\int (\psi + \varepsilon) \, d\mu = - \int -(\psi + \varepsilon) \, d\mu &\geq -P(\varphi - (\psi + \varepsilon)) + P(\varphi) \\ &\geq -(P(\varphi) - \inf_{x \in X} (\psi + \varepsilon)) + P(\varphi) = \inf_{x \in X} (\psi + \varepsilon) > 0.\end{aligned}$$

This implies $\int \psi \, d\mu \geq 0$ so that μ is a positive functional. In the next step we show that it is a probability measure, i.e. $\mu(X) = 1$. Let $k \in \mathbb{Z}$, then

$$\int k \, d\mu \leq P(\varphi + k) + P(\varphi) = k,$$

where the first inequality holds since $\mu \in t_\varphi$ by assumption and the right equality since we are taking the supremum only over $\mathcal{M}(f)$ to get P. Taking $k = \pm 1$ gives $\mu(X) = 1$. It remains to show that μ is f-invariant. Let $k \in \mathbb{Z}$ and $\psi \in C(X)$, then

$$k \int (\psi \circ f - \psi) \, d\mu \leq P(\varphi + k(\psi \circ f - \psi)) - P(\varphi) = 0.$$

Taking again $k = \pm 1$ proves that μ is f-invariant and so $t_\varphi \subset \mathcal{M}(f)$.
2. We only prove " $\Leftarrow$ ", for the other direction we refer to [16, Theorem

9.12]. Let $\mu \to h_\mu(f)$ be upper semi-continuous at $\mu_0 \in \mathcal{M}(f)$. By the variational principle we have

$$h_{\mu_0}(f) \leq \inf \left\{ P(\varphi) - \int \varphi \, d\mu_0 : \varphi \in C(X) \right\},$$

which gives the first inequality. For equality we will prove the opposite inequality. Let $b > h_{\mu_0}$ and $C = \{(\mu, t) \in \mathcal{M}(f) \times \mathbb{R} : 0 \leq t \leq h_\mu(f)\}$. By [16, Theorem 8.1] the entropy map is affine, i.e. for $\mu, \nu \in \mathcal{M}(f)$ and $\lambda \in [0,1]$ we have $h_{\lambda\mu+(1-\lambda)\nu}(f) = \lambda h_\mu(f) + (1-\lambda)h_\nu(f)$. Using this it is easy to see that C is convex. Let us consider C as a subset of $C^*(X) \times \mathbb{R}$, where $C^*(X)$ denotes the dual space of $C(X)$. As usual $C^*(X)$ is equipped with the weak*-topology and so by upper semi-continuity of $h_\mu(f)$ at μ_0, $(\mu_0, b) \notin \overline{C}$. The crucial step in this proof is based on Theorem B.1 in the appendix. By this theorem there exists a continuous linear functional $L : C(X)^* \times \mathbb{R} \to \mathbb{R}$, such that $L(\mu, t) < L(\mu_0, b)$ holds for all $(\mu, t) \in \overline{C}$. Since we use the weak*-topology on $C^*(X)$ we know that L is of the form

$$L(\mu, t) = \int \varphi \, d\mu + kt,$$

for some $\varphi \in C(X)$ and some $k \in \mathbb{R}$. Therefore we have

$$\int \varphi \, d\mu + kt < \int \varphi \, d\mu_0 + kb, \quad \forall (\mu, t) \in \overline{C}.$$

If we set $t = h_\mu(f)$ the inequality still holds true for all $\mu \in \mathcal{M}(f)$, so we put $\mu = \mu_0$, which implies $kh_{\mu_0}(f) < kb$, and hence $k > 0$. Thus we can divide by k to obtain,

$$h_\mu(f) + \int \frac{\varphi}{k} \, d\mu < b + \int \frac{\varphi}{k} \, d\mu_0, \quad \forall \mu \in \mathcal{M}(f).$$

By the variational principle we get

$$b \geq P\Big(\frac{\varphi}{k}\Big) - \int \frac{\varphi}{k} \, d\mu_0 \geq \inf \left\{ P(\psi) - \int \psi \, d\mu_0 : \psi \in C(X) \right\}.$$

And so finally

$$h_{\mu_0}(f) \geq \inf \left\{ P(\psi) - \int \psi \, d\mu_0 : \psi \in C(X) \right\}.$$

3. Referring to 1. it only remains to show $t_\varphi \subset ES_\varphi$ if $\mu \to \mathrm{h}_\mu(f)$ is upper semi-continuous. Let $\mu \in t_\varphi$. We have for all $\psi \in C(X)$

$$P(\varphi + \psi) - \int (\varphi + \psi) \, d\mu \geq P(\varphi) - \int \varphi \, d\mu.$$

As latter inequality holds for all $\psi \in C(X)$ we can substitute $\varphi + \psi$ by an arbitrary $\rho \in C(X)$. Since $h_\mu(f)$ is upper semi-continuous, we can apply 2 and take the infimum on the left hand side. Thus we obtain

$$\inf_{\rho \in C(X)} \left(P(\rho) - \int \rho \, d\mu \right) = h_\mu(f) \geq P(\varphi) - \int \varphi \, d\mu.$$

Rearranging this inequality yields $P(\varphi) \leq h_\mu(f) + \int \varphi \, d\mu$, and hence $\mu \in ES_\varphi$. $\square$

Finally we can prove Theorem 1.4 and establish a connection between the subdifferential of p and the set of rotation vectors with respect to the equilibrium states.

Proof of Theorem 1.4. We apply Lemma 3.5 to $\varphi = \alpha.\Phi$ and $\psi = \beta.\Phi$ to obtain for all $\beta \in \mathbb{R}^m$

$$d^+p(\alpha;\beta) = \sup\left\{ \int \beta.\Phi \, d\mu : \mu \in ES_{\alpha.\Phi} \right\}, \tag{37}$$

where we also used 3 of Lemma 3.6 and the upper semi-continuity of $h_\mu(f)$ such that $t_{\alpha.\Phi} = ES_{\alpha.\Phi}$.
We first show that the right hand side is included in the left hand side. By Lemma 3.4, (37), for every $\mu \in ES_{\alpha.\Phi}$ $\mathrm{rv}(\mu)$ is a subgradient of p at α.
For the converse inclusion note that $ES_{\alpha.\Phi}$ is compact combined with (37) implies that for each point γ_0 at the boundary of the convex set $\partial p(\alpha)$ there is a $\mu_0 \in ES_{\alpha.\Phi}$ such that $\mathrm{rv}(\mu_0) = \gamma_0$. But convexity of the two sets $\partial p(\alpha)$ and $ES_{\alpha.\Phi}$ means that in fact for every point γ in $\partial p(\alpha)$ there is a $\mu \in ES_{\alpha.\Phi}$ with $\mathrm{rv}(\mu) = \gamma$. $\square$

Despite Theorem 2.20, the assumption that $h_\mu(f)$ is upper semi-continuous is still a restriction. Therefore it would be nice to have the same result without this assumption. At the end of this section we try to generalize the proposition and it turns out that at least for one inclusion we do not need upper semi-continuity. It is an interesting open problem whether the converse inclusion holds. But this leads beyond the scope of this thesis. First we discuss consequences from this theorem to get more information about rotation sets which will shed some light on the role of equilibrium measures. To simplify notation let us consider the set of equilibrium measures of the form $ES_{\alpha.\Phi}$ for a fixed m-dimensional potential Φ as a function, namely $ES_\Phi : \mathbb{R}^m \to \mathcal{M}(f)$, defined by $ES_\Phi(\alpha) = ES_{\alpha.\Phi}$.

Corollary 3.7. *Let (X, f) be a dynamical system for which the entropy map $\mu \to h_\mu(f)$ is upper semi-continuous, and suppose $\Phi : X \to \mathbb{R}^m$ is continuous. Then p is differentiable at the point $\alpha \in \mathbb{R}^m$ if and only if $\mathrm{rv}(ES_{\alpha.\Phi})$ being a singleton.*

Proof. We recall the fact that p is differentiable at the point $\alpha \in \mathbb{R}^m$ if and only if the set of subdifferentials at the point α is a singleton, Theorem A.6. By Theorem 1.4 this is equivalent to $\mathrm{rv}(ES_{\alpha.\Phi})$ is a singleton. $\square$

We note $\mathrm{rv}(ES_{\alpha.\Phi})$ being a singleton does not imply that $ES_{\alpha.\Phi}$ is a singleton. Obviously, $ES_{\alpha.\Phi}$ being a singleton implies that $\mathrm{rv}(ES_{\alpha.\Phi})$ is also a singleton and thus is p differentiable at α. The converse may be wrong in particular cases. We refer to Section 2.8 where we followed Haydn's idea in [5] and constructed an example of a subshift $f : X(\tau) \to X(\tau)$ with two mutually singular ergodic measures of maximal entropy. We denote these measures here by μ and ν. Since $ES_{0.\Phi} = \sup\{h_\mu(f) : \mu \in \mathcal{M}(f)\} = \{\mu, \nu\}$, we only need to recall Example 1, where the measures μ and ν have the same rotation vector because of symmetry, i.e. $\mathrm{rv}(\mu) = \mathrm{rv}(\nu)$. Thus $\mathrm{rv}(ES_{0.\Phi})$ is a singleton and so p is differentiable at 0 but $ES_{0.\Phi}$ is no singleton.

Proposition 3.8. *Let (X, f) be a dynamical system for which the entropy map $\mu \to h_\mu(f)$ is upper semi-continuous, and suppose $\Phi : X \to \mathbb{R}^m$ is continuous. Then $\mathrm{rv}(\mathcal{M}(f)) \subset \overline{\partial p(\mathbb{R}^m)}$.*

Proof. Let $\mu \in \mathcal{M}(f)$, and define the affine map $F_\mu(\alpha.\Phi) : \mathbb{R}^m \to \mathbb{R}$ by,

$$F_\mu(\alpha.\Phi) := h_\mu(f) + \int \alpha.\Phi \, d\mu. \tag{38}$$

$h\mu(f) + \int \varphi \, d\mu$ is called the free energy of μ with respect to the potential φ. We can easily observe that $F_\mu(\alpha) \leq P(\alpha.\Phi) = p(\alpha)$ holds. The $Graph(F_\mu)$ is a hyperplane in $\mathbb{R}^{m+1}$, and the latter inequality between F_μ and p implies that the $Graph(p)$ lies above this hyperplane (possibly touching it tangentially if p is differentiable). But this implies that the gradient of the $Graph(F_\mu)$ is contained in the closure (regarding asymptotic behavior of p and F_μ) of all subdifferentials of p, since otherwise this graph would intersect transversely. We showed that $\mathrm{rv}(\mu)$ belongs to the closure of $\partial p(\mathbb{R}^m)$ and since μ was arbitrary we conclude $\mathrm{rv}(\mathcal{M}(f)) \subset \overline{\partial p(\mathbb{R}^m)}$. $\square$

Theorem 3.9. *Let (X, f) be a dynamical system for which the entropy map $\mu \to h_\mu(f)$ is upper semi-continuous, and suppose $\Phi : X \to \mathbb{R}^m$ is continuous. Then*

$$\mathrm{rv}(\mathcal{M}(f)) = \overline{\mathrm{rv}(ES_\Phi(\mathbb{R}^m))} = \overline{\partial p(\mathbb{R}^m)}.$$

Proof. Obviously we have $\mathrm{rv}(ES_\Phi(\mathbb{R}^m)) \subset \mathrm{rv}(\mathcal{M}(f))$, since $ES_\Phi(\mathbb{R}^m) \subset \mathcal{M}(f)$. Compactness of $\mathrm{rv}(\mathcal{M}(f))$ implies $\overline{\mathrm{rv}(ES_\Phi(\mathbb{R}^m))} \subset \mathrm{rv}(\mathcal{M}(f))$. By Theorem 1.4 $\partial p(\mathbb{R}^m) = \mathrm{rv}(ES_\Phi(\mathbb{R}^m))$ and therefore Proposition 3.8 completes the proof. $\square$

Definition 3.10. *For a convex set C we define its relative interior by*

$$ri(C) = \{C \in aff(C) : \exists \varepsilon > 0 : B(x, \varepsilon) \cap aff(C) \subset C\},$$

where $aff(C)$ denote the affine hull of C.

For further information see Section A.2 in the appendix.

Corollary 3.11. *Let (X, f) be a dynamical system for which the entropy map $\mu \to h_\mu(f)$ is upper semi-continuous, and suppose $\Phi : X \to \mathbb{R}^m$ is continuous. Then $ri(\mathrm{rv}(\mathcal{M}(f))) \subset \mathrm{rv}(ES_\Phi(\mathbb{R}^m)) = \partial p(\mathbb{R}^m)$.*

Proof. By Theorem 1.4 we know $\mathrm{rv}(ES_\Phi(\mathbb{R}^m)) = \partial p(\mathbb{R}^m)$. Observe that the set $\partial p(\mathbb{R}^m)$ is in general not convex. Let $\hat{p} : \mathbb{R}^m \to [-\infty, \infty]$ denote the convex conjugate of p, defined by $\hat{p}(\hat{\alpha}) = \sup\{\beta.\hat{\alpha} - p(\beta) : \beta \in \mathbb{R}^m\}$, and $dom(\hat{p}) = \{\hat{\alpha} \in \mathbb{R}^m : \hat{p}(\hat{\alpha}) < \infty\}$ is its effective domain. Further information on $\hat{p}$ can be found in [18, §12] or in the Section A.3 in the appendix. By Theorem A.7 we have almost convexity in the sense that

$$ri(dom(\hat{p})) \subset \partial p(\mathbb{R}^m) \subset dom(\hat{p}). \tag{39}$$

Taking the closure followed by the relative interior of all sets in (39) gives the equality $ri(\overline{\partial p(\mathbb{R}^m)} = ri(dom(\hat{p}))$, note that here we used again Section A.2. Combining this equality with (39) gives

$$ri(\overline{\partial p(\mathbb{R}^m)}) \subset \partial p(\mathbb{R}^m).$$

To end this proof we apply Theorem 3.9 to this inclusion and obtain

$$ri(\mathrm{rv}(\mathcal{M}(f))) \subset \mathrm{rv}(ES_\Phi(\mathbb{R}^m)) = \partial p(\mathbb{R}^m).$$

□

Corollary 3.12. *Let (X, f) be a dynamical system for which the entropy map $\mu \to h_\mu(f)$ is upper semi-continuous, and suppose $\Phi : X \to \mathbb{R}^m$ is continuous. If p is strictly convex then $int(\mathrm{rv}(\mathcal{M}(f))) = \mathrm{rv}(ES_\Phi(\mathbb{R}^m)) = \partial p(\mathbb{R}^m)$.*

Proof. Strict convexity of p implies $\partial p(\mathbb{R}^m)$ is both open in $\mathbb{R}^m$ and convex [18, p. 227], so by Theorem 1.4, $\mathrm{rv}(ES_\Phi(\mathbb{R}^m))$ is open and convex. Thus we have $int(\overline{\mathrm{rv}(ES_\Phi(\mathbb{R}^m))}) = \mathrm{rv}(ES_\Phi(\mathbb{R}^m))$, and the result follows by Theorem 3.9. □

As mentioned before we would like to have the same result as in Theorem 1.4 but without asking for upper semi-continuity of $h_\mu(f)$. In this case the set ES_φ might be empty but still we might be able to approximate it and see if this suffices.

We fix a continuous potential $\Phi : X \to \mathbb{R}^m$ and define the set of asymptotic rotation vectors ($ARV_{\alpha.\Phi}$) with respect to $\alpha \in \mathbb{R}^m$ and Φ as

$$\left\{ w \in \mathbb{R}^m : \exists (\mu_n)_{n\in\mathbb{N}} \subset \mathcal{M}(f) : \ h_{\mu_n}(f) + \int \alpha.\Phi \ d\mu_n \to p(\alpha), \ rv(\mu_n) \to w \right\}.$$

If $h_\mu(f)$ is upper semi-continuous this set coincides with $\mathrm{rv}(ES_{\alpha.\Phi})$, hence we have the right candidate to generalize the proposition. The question arises if we still have $ARV_\alpha = \partial p(\alpha)$ without upper semi-continuity.

We will prove $\mathrm{ARV}_{\alpha.\Phi} \subset \partial p(\alpha)$. Note that for $w \in ARV_{\alpha.\Phi}$, $\beta \in \mathbb{R}^m$ and given $\varepsilon > 0$ there is $N \in \mathbb{N}$ such that

$$|\beta.\mathrm{rv}(\mu_n) - \beta.w| < \varepsilon, \text{ and } \left(p(\alpha) - h_{\mu_n}(f) - \int \alpha.\Phi \ d\mu_n \right) < \varepsilon, \ \forall n \geq N. \tag{40}$$

We approximate w and the pressure p and take $N \in \mathbb{N}$ such that both approximations have at least ε-precision. Note also that for the second approximation we do not need the absolute value since we know that the pressure is always greater or equal.

Let $w \in ARV_{\alpha.\Phi}$, we will show that for an arbitrary $\beta \in \mathbb{R}^m$ the inequality

$\beta.w \leq p(\alpha+\beta) - p(\alpha)$ holds. By the approximation in (40) we can directly conclude that

$$\beta.w \leq \beta.\mathrm{rv}(\mu_n) + \varepsilon = \int \beta.\Phi \; d\mu_n \pm h_{\mu_n}(f) \pm \int \alpha.\Phi \; d\mu_n + \varepsilon$$

$$\leq p(\alpha+\beta) - \Big(h_{\mu_n}(f) + \int \alpha.\Phi \; d\mu_n\Big) + \varepsilon \leq p(\alpha+\beta) - p(\alpha) + 2\varepsilon,$$

which means $ARV_{\alpha.\Phi} \subset \partial p(\alpha)$.

4. Hölder continuous potentials

In the introduction we raised the question under which assumptions a measure of the maximal entropy exists and is unique. You can ask the same question for equilibrium states. Bowen answered this question for a family of potentials, namely Hölder continuous potentials and Axiom A systems. But also the dynamical system must satisfy particular properties. We follow [7], where a similar statement for a more specific class of dynamical systems is given.

Definition 4.1. *Let (X, f) be a dynamical system. A specification $S = (\tau, P)$ consists of a finite collection $\tau = \{I_1, \ldots, I_m\}$ of finite intervals $I_i = [a_i, b_i] \subset \mathbb{Z}$ and a map $P : T(\tau) := \bigcup_{i=1}^m I_i \to X$ such that for $t_1, t_2 \in I \in \tau$ we have $f^{t_2 - t_1}(P(t_1)) = P(t_2)$. S is said to be n-spaced if $a_{i+1} > b_i + n$ for all $i \in \{1, \ldots, m\}$ and the minimal such n is called the spacing of S. We say that S parameterizes the collection $\{P_I : I \in \tau\}$ of orbit segments of f. We let $T(S) := T(\tau)$ and $L(S) := L(\tau) := b_m - a_1$. Let d be a metric on X, we say that S is ε-shadowed by $x \in X$ if $d(f^n(x), P(n)) < \varepsilon$ for all $n \in T(S)$. Thus a specification is a parametrized union of orbit segments $P_{|I_i}$ of f.*
If f is a homeomorphism then f is said to have the specification property if for any $\varepsilon > 0$ there exists $M = M_\varepsilon \in \mathbb{N}$ such that any M-spaced specification S is ε-shadowed by a point of X and such that moreover for any $q \geq M + L(S)$ there is a period-q orbit ε-shadowing S.

Roughly speaking this means that (X, f) satisfies the specification property if one can approximate distinct pieces of orbits by single periodic orbits with an certain uniformity.

Definition 4.2. *By $C^f(X)$ we denote the set of functions*

$$\{\varphi \in C(X) : \exists K, \varepsilon > 0 \text{ s.t. } d_n(x,y) \leq \varepsilon \Rightarrow |S_n\varphi(x) - S_n\varphi(y)| < K, \forall n \in \mathbb{N}\},$$

where $S_n\varphi(x)$ is defined as in equation (28) *in Section 3.*

By this definition we have good control of the $S_n\varphi$ uniformly for $n \in \mathbb{N}$ for functions $\varphi \in C^f(X)$. It is well-known that in the case of a hyperbolic set all Hölder-continuous functions are in $C^f(X)$, [7, Proposition 20.2.6.]. We provide a proof for shifts.

Proposition 4.3. *Let (X, f) be a shift. Then every Hölder-continuous function on X is in $C^f(X)$.*

Proof. Let (X, f) be the one-sided shift. The proof works similarly for the two-sided shift. We recall the definition of the metric d on X,

$$d(x, y) = \left(\frac{1}{2}\right)^{\min\{i: x_i \neq y_i\}}. \tag{41}$$

Let $\varphi : X \to \mathbb{R}$ be Hölder-continuous and define $\mathrm{Var}_n(\varphi, f, \varepsilon) = \sup\{|\varphi(x) - \varphi(y)| : d(f^i(x), f^i(y)) < \varepsilon \text{ for } i = 0, \ldots, n\}$. For the shift space with the metric d from (41) we can choose $\varepsilon = \frac{1}{2}$. If for $x, y \in X$, $d(f^i(x), f^i(y)) < \frac{1}{2}$ for $i = 0, \ldots, n$ we conclude that $d(x, y) < (\frac{1}{2})^n$. Since φ is Hölder-continuous we have $|\varphi(x) - \varphi(y)| \leq cd(x, y)^\alpha$ and thus $\mathrm{Var}_n(\varphi, f, \frac{1}{2}) \leq c(\frac{1}{2})^{\alpha n}$. If $d(f^i(x), f^i(y)) < \frac{1}{2}$ for $i = 0, \ldots, k$ we obtain

$$|\varphi(f^k(x)) - \varphi(f^k(y))| \leq \mathrm{Var}_k\Big(\varphi, f, \frac{1}{2}\Big). \tag{42}$$

Thus we have

$$\begin{aligned} |S_n\varphi(x) - S_n\varphi(y)| &\leq \sum_{k=0}^{n} \mathrm{Var}_k\Big(\varphi, f, \frac{1}{2}\Big) \leq \sum_{k=0}^{n} c\left(\frac{1}{2}\right)^{\alpha k} \\ &\leq \sum_{k\geq 0} c\left(\frac{1}{2}\right)^{\alpha k} = \frac{c}{1 - (\frac{1}{2})^\alpha} =: K. \end{aligned} \tag{43}$$

Thus $\varphi \in C^f(X)$. ☐

For the following results we do not provide proofs, but we indicate where they are.

Lemma 4.4 (7, Lemma 4.5.1)**.** *Let X be a compact metric space, $\mu \in \mathcal{M}_X$.*

1. *For all $x \in X$, $\delta > 0$ there exists $\delta' \in (0, \delta)$ such that $\mu(\partial B(x, \delta')) = 0$.*
2. *For all $\delta > 0$ there exists a finite measurable partition $\xi = \{C_1, \ldots, C_k\}$ with diam$(C_i) < \delta$ for all i and $\mu(\partial\xi) = 0$.*

Lemma 4.5 (7, Lemma 20.3.4.)**.** *Let (X, d) be a compact metric space, $f : X \to X$ an expansive homeomorphism with the specification property and $\varphi \in C^f(X)$. Let μ be a accumulation point of the sequence μ_n, defined by*

$$\mu_n = \frac{1}{P_X(\varphi, n)} \sum_{x \in \mathrm{Fix}(f^n)} e^{S_n\varphi(x)}\delta_x \in \mathcal{M},$$

where $P_X(\varphi, n)$ is the normalization factor and $\varepsilon > 0$ as in Definition 4.2 of $C^f(X)$. Then there exists A_ε, $B_\varepsilon > 0$ such that for $x \in X$ and $n \in \mathbb{N}$ we have

$$A_\varepsilon e^{S_n\varphi(x) - nP(\varphi)} \leq \mu(\overline{B_f(x, \varepsilon, n)}) \leq B_\varepsilon e^{S_n\varphi(x) - nP(\varphi)}.$$

Lemma 4.6 (7, Equation (20.3.5))**.** *If $x_i, a_i \geq 0$, $g(x) = x \log x$, then for $b_i = \log a_i$*

$$\sum_{i=1}^{n} x_i(b_i - \log(x_i)) \leq \sum_{i=1}^{n} x_i \log\left(\sum_{j=1}^{n} e^{b_j}\right) + \frac{1}{e},$$

where we used $g(x) \geq -1/e$.

Definition 4.7. *A measure preserving transformation (X, f, μ) is called mixing if for any two measurable sets A, B*

$$\mu(f^{-n}(A) \cap B) \to \mu(A)\mu(B) \text{ as } n \to \infty.$$

Lemma 4.8 (7, Proposition 20.3.6.)**.** *Let f be a homeomorphism of a compact metric space X and μ an f-invariant Borel probability measure such that for any Borel Sets P, Q the inequality*

$$c\mu(P)\mu(Q) \leq \varliminf_{n\to\infty} \mu(P \cap f^{-n}(Q)) \leq \varlimsup_{n\to\infty} \mu(P \cap f^{-n}(Q)) \leq C\mu(P)\mu(Q)$$

holds for some $c, C > 0$. Then is μ mixing.

Theorem 4.9. *Let (X, d) be a compact metric space, $f : X \to X$ an expansive homeomorphism satisfying the specification property, and $\varphi \in C^f(X)$. Then there is exactly one $\mu_\varphi = \mu \in \mathcal{M}(f)$ with $P_\mu(\varphi) := h_\mu(f) + \int \varphi \, d\mu = P(\varphi)$. It is mixing and*

$$\mu_\varphi = \lim_{n\to\infty} \frac{1}{P_X(\varphi, n)} \sum_{x\in \mathrm{Fix}(f^n)} e^{S_n\varphi(x)}\delta_x,$$

where $P_X(\varphi, n)$ is the normalizing factor.

Proof. Let us fix a weak*-accumulation point

$$\mu = \lim_{k\to\infty} \mu_{n_k}$$

of the sequence

$$\mu_n = \frac{1}{P_X(\varphi, n)} \sum_{x\in \mathrm{Fix}(f^n)} e^{S_n\varphi(x)}\delta_x \in \mathcal{M}(f).$$

We will show that if $P_\nu(\varphi) = P(\varphi)$ then $\nu = \mu$, so $P_\mu(\varphi) = P(\varphi)$ and there is only one accumulation point. Since $\nu \in \mathcal{M}(f)$ we can find a convex combination $\nu = \lambda\nu' + (1-\lambda)\mu'$ for $\lambda \in [0, 1]$, and $\nu', \mu' \in \mathcal{M}(f)$ such that $\mu' \ll \mu \perp \nu'$. As μ is ergodic the density function of μ' w.r.t μ is constant μ a.e. and so $\mu' = \mu$. Since we assumed $P_\nu(\varphi) = P(\varphi)$ we have two cases such that $P_\nu(\varphi) = \lambda P_{\nu'}(\varphi) + (1-\lambda)P_\mu(\varphi)$ holds.

1. $\lambda = 0$, and so $\nu = \mu$.
2. $P_{\nu'}(\varphi) = P(\varphi)$.

In the first case we are done and therefore we prove that $P_{\nu'}(\varphi) < P(\varphi)$ as $\nu' \perp \mu$.

For $n \in \mathbb{N}$ and a maximal with respect to the inclusion $(n, 2\varepsilon)$-separated set $E_n = \{x_1, \dots, x_k\}$ we take Borel sets β_x such that $B(x, \varepsilon, n) \subset \beta_x \subset B(x, 2\varepsilon, n)$, and $\mathcal{B}_n := \{\beta_x : x \in E_n\}$ is a partition. Where $B(x, \varepsilon, n)$ is the ε ball around x with respect to the metric d_n. For instance one could define β_x by

$$\beta_{x_1} = B(x_1, 2\varepsilon, n) \Big\backslash \bigcup_{i=2}^{k} B(x_i, \varepsilon, n)$$

$$\beta_{x_{j+1}} = B(x_{j+1}, 2\varepsilon, n) \Big\backslash \bigcup_{i=j+2}^{k} B(x_i, \varepsilon, n) \Big\backslash \bigcup_{i=1}^{j} \beta_{x_j}.$$

Note that by Lemma 4.4 we can choose $\varepsilon > 0$ such that $(\mu + \nu)(\partial \mathcal{B}_n) = 0$. Since f is expansive $\operatorname{diam} f^{-(n/2)}(\mathcal{B}_n) \to 0$ as $n \to \infty$ so if $f(B) = B \subset X$ such that $\mu(B) = 0$ and $\nu(B) = 1$ then there exist finite unions C_n of elements of $\mathcal{B}_n$ such that

$$(\mu + \nu)(C_n \Delta B) = (\mu + \nu)(f^{-(n/2)}(C_n) \Delta B) \xrightarrow{n \to \infty} 0.$$

Furthermore if $\varepsilon < \delta_0/2$ then $\mathcal{B}_n$ is generating for f^n, and so by Lemma 2.18 and Corollary 2.17 $n h_\nu(f) = h_\nu(f^n) = h_\nu(f^n, \mathcal{B}_n) \leq H_\nu(\mathcal{B}_n)$, where $H_\nu(\mathcal{B}_n)$ is defined in equation (11) in Section 2.2. Applying the latter inequality for the entropy and rearrange the integral part of the pressure by using the invariance leads to

$$\begin{aligned} nP_\nu(\varphi) = nh_\nu(f) + n \int \varphi \, d\nu &\leq H_\nu(\mathcal{B}_n) + \sum_{\beta_x \in \mathcal{B}_n} \Bigg(\int_{\beta_x} \sum_{i=0}^{n-1} \varphi \, d\nu \Bigg) \\ &= \sum_{\beta_x \in \mathcal{B}_n} \Bigg(-\nu(\beta_x) \log \nu(\beta_x) + \int_{\beta_x} S_n \varphi \, d\nu \Bigg). \end{aligned} \tag{44}$$

Note that, since $\varphi \in C^f(X)$ we get $S_n \varphi \leq K + S_n \varphi(x)$ on all $\beta_x \in \mathcal{B}_n$. So we get the upper bound for (44)

$$\begin{aligned} &\sum_{x \in E_n; \beta_x \in \mathcal{B}_n} \Big(-\nu(\beta_x) \log \nu(\beta_x) + \nu(\beta_x)(K + S_n \varphi(x)) \Big) \\ =&K + \sum_{x \in E_n; \beta_x \in \mathcal{B}_n} \nu(\beta_x) \Big(S_n \varphi(x) - \log \nu(\beta_x) \Big). \end{aligned} \tag{45}$$

In (45) we can now apply Lemma 4.6 and separate the sum into two parts depending on C_n, and obtain the upper bound for (45)

$$\begin{aligned} &K + \nu(C_n) \log \sum_{x \in E_x; \beta_x \subset C_n} e^{S_n \varphi(x)} \\ &+ \nu(X \backslash C_n) \log \sum_{x \in E_x; \beta_x \cap C_n = \varnothing} e^{S_n \varphi(x)} + \frac{2}{e}. \end{aligned} \tag{46}$$

Finally we apply Lemma 4.5 to (46) to obtain

$$\begin{aligned} n(P_\nu(\varphi) - P(\varphi)) - K - \frac{2}{e} &\le \nu(C_n) \log \sum_{x \in E_x; \beta_x \cap C_n = \varnothing} e^{S_n\varphi(x) - nP(\varphi)} \\ &\quad + \nu(X \backslash C_n) \log \sum_{x \in E_x; \beta_x \cap C_n = \varnothing} e^{S_n\varphi(x) - nP(\varphi)} \\ &\le \nu(C_n) \log(A_\varepsilon^{-1} \mu(C_n)) + \nu(X \backslash C_n) \log(A_\varepsilon^{-1} \mu(X \backslash C_n)). \end{aligned}$$

For $n \to \infty$ the right-hand side goes to $-\infty$, since $\nu(C_n) \to 1$ and $\mu(C_n) \to 0$. And therefore $P_\nu(\varphi) < P(\varphi)$. $\square$

5. Localized entropy

In the Introduction we mentioned that rotation sets can be used to get a partition of the set of invariant measures. This approach is reasonable as the set of invariant measures might be rather big and therefore the question arises which measure reflects most appropriately the relevant information. This may depend on the information one is looking for. Frequently when complexity is measured one considers maximal entropy as criteria. But it turns out that it is hard to tell which measure is the best. Therefore we want to learn more about the behavior of different classes of measures. For instance we want to know how the entropy varies on these classes of measures. Depending on the potential $\Phi = (\phi_1, \dots, \phi_m) : X \to \mathbb{R}^m$ we can obtain different information from the rotation set. As in Section 2.3 we obtain a filtration of the set of invariant probability measures. We fix m and obtain the partition $\bigcup_{w \in \mathrm{Rot}(\Phi)} \mathrm{rv}^{-1}(w) = \mathcal{M}(f)$. On each element of the partition we consider the localized entropy (Definition 2.23). Our goal is to find an explicit formula to calculate the localized entropy. In the first step we consider a subshift of finite type and a potential which is constant on cylinders of lenght K and show that in this case the rotation set is a polyhedron, i.e. the convex hull of finitely many points. The proof is inspired by Ziemian [19]. For this polyhedron we calculate the localized entropy using elementary loops. The set of elementary loops is finite. The question arises if the set of elementary loops is enough to calculate the actual localized entropy rather than computing upper and lower bounds. It turns out that for extreme points we can calculate the localized entropy by calculating the topological entropy for a certain subshift and apply the variational principle for entropy. For w on a face of $\mathrm{Rot}(\Phi)$ we only get a lower bound. To calculate the localized entropy we will need another the result from Lemma 2.39.

Let us define the one-sided subshift of finite type. It turns out everything works exactly the same way for the two-sided shift. The subshift is given by an alphabet $\mathcal{A} = \{1, \dots, d\}$ and the sequence space X_A which is determined by a $d \times d$ matrix composed of 1's and 0's where the step from x_n to x_{n+1} is allowed if $A(x_n, x_{n+1})$ is 1 . Therefore we get the space $X_A = \{x =$

$(x_n)_{n\in\mathbb{N}} : x_n \in \mathcal{A},\ A(x_n, x_{n+1}) = 1\}$. Furthermore we have the shift map $f : X_A \to X_A$ given by $(f(x))_n = x_{n+1}$ and a potential $\Phi : X_A \to \mathbb{R}^m$ constant on cylinders of length K, i.e. for all $y \in C_K(x) = \{z \in X_A : z_1 = x_1, \dots, z_K = x_K\}$, $\Phi(y) = \Phi(x)$. We assume that the dynamical system (X_A, f) is transitive. By combinatorical arguments this implies d^K different cylinders in case of a full shift. We order these cylinders in a lexicographical way, and define associated symbols for the new alphabet. Here we denote a cylinder of length K by $[x_1, \dots, x_K]_K$, where $x_i \in \mathcal{A}$. This gives

$$\begin{aligned} 1' &:= [1, 1, \dots, 1]_K \\ 2' &:= [1, \dots, 1, 2]_K \\ &\vdots \\ d^K &:= [d, d, \dots, d]_K \end{aligned} \tag{47}$$

Thus we obtain the alphabet $\mathcal{A}_C = \{1, \dots, d^K\}$ where each symbol is uniquely identified with one cylinder, and also equipped with the shift map σ in such a way that each sequence in X_A can be uniquely identified with a sequence of cylinders respectively sequences build of symbols of $\mathcal{A}_C$ so that both systems are topologically conjugate (Definition 2.36) to each other. Therefore we have to construct the sequence space Σ_{A_C} on the alphabet $\mathcal{A}_C$ appropriately. To make this precise define the map $\psi : X_A \to \mathcal{A}_C^{\mathbb{N}}$ and define $\Sigma_{A_C} = \psi(X_A)$. Let $x = (x_1, x_2, \dots) \in X_A$, $\xi = (\xi_1, \xi_2, \dots) \in \mathcal{A}_C^{\mathbb{N}}$ and denote by $[x_1, x_2, \dots, x_K]_K$ a cylinder of length K.

$$\begin{aligned} &\psi((x_1, x_2, x_3, x_4, \dots, x_K, x_{K+1}, x_{K+2}, \dots)) = \\ &(\underbrace{[x_1, x_2, \dots, x_K]_K}_{\xi_1}, \underbrace{[x_2, x_3, \dots, x_{K+1}]_K}_{\xi_2}, \underbrace{[x_3, x_4, \dots, x_{K+2}]_K}_{\xi_3}, \dots). \end{aligned} \tag{48}$$

With this map and the lexicographical order from equation (47) we can simply define the associated subshift of finite type, namely

$$\Sigma_{A_C} = \{\xi = (\xi_n)_{n\in\mathbb{N}} : \exists x \in X_A : \psi(x) = \xi,\ \xi_n \in \mathcal{A}_C\}. \tag{49}$$

From this definition we could also define the subshift in the usual way, namely defining a $d^K \times d^K$ matrix A_C which defines the valid sequences. Some symbols may be redundant as we consider the general case that (X_A, f) is a subshift and not a full shift. However, in the case where (X_A, f) is a full shift we can observe that from any symbol ξ_i we only have d possibilities for ξ_{i+1}, since the cylinder ξ_{i+1} is determined by ξ_i up to the last symbol. If (X_A, f) is also a subshift we have even more restrictions. In the next step we will show that ψ is a homeomorphism and the maps f and σ are topological conjugates, i.e. $f = \psi^{-1} \circ \sigma \circ \psi$.

Proposition 5.1. *The dynamical systems (X_A, f) and (Σ_{A_C}, σ) are conjugates with conjugacy $\psi : X_A \to \Sigma_{A_C}$. In particular, for $\tilde{\Phi} = \Phi(\psi^{-1}(\xi))$ the rotation sets and the localized entropy coincide.*

Proof. We have to show that ψ^{-1} exists, i.e. ψ is bijective. Surjectivity follows easily by the definition of Σ_{A_C}. For injectivity let $x, y \in X_A$ and assume $\psi(x) = \psi(y)$. That is

$$\begin{aligned}([x_1,\dots,x_K],[x_2,\dots,x_{K+1}],\dots) &= ([y_1,\dots,y_K],[y_2,\dots,y_{K+1}],\dots)\\ \Leftrightarrow (\xi_1,\xi_2,\dots) &= (\xi_1',\xi',\dots).\end{aligned}$$

Since the association of symbols from $\mathcal{A}_C$ and $\mathcal{A}$ is unique we get $x = y$. So we know that ψ^{-1} exists. Continuity of both ψ and ψ^{-1} is trivial. It remains to show $f = \psi^{-1} \circ \sigma \circ \psi$. For $x \in X_A$ we obtain $(f(x))_n = x_{n+1}$ and

$$(\psi^{-1}\sigma\psi(x))_n = (\psi^{-1}\psi(x))_{n+1} = x_{n+1}. \tag{50}$$

So we have that f and σ are topological conjugate with respect to the homeomorphism ψ. Furthermore we note that for each periodic point $x \in X_A$ we have that $\psi(x)$ is also a periodic point in Σ_C and vice versa. Let $x \in X_A$ be of period p, i.e. $f^p(x) = x$. Then

$$\psi(x) = \psi(f^p(x)) = \psi(\psi^{-1}\sigma^p\psi(x)) = \sigma^p(\psi(x)). \tag{51}$$

The map ψ also preserves measures, namely let μ be a f-invariant probability measure on X_A and define $\tilde{\mu} = \mu \circ \psi^{-1}$. Then for $B \in \mathcal{B}_{X_A}$ and $\tilde{B} = \psi(B)$ we obtain

$$\tilde{\mu}(\sigma^{-1}\tilde{B}) = \mu(\psi^{-1}\sigma^{-1}\psi(B)) = \mu(f^{-1}B) = \mu(B). \tag{52}$$

Next we discuss how to adjust the potential for the new system. We define $\tilde{\Phi} : \Sigma_C \to \mathbb{R}^m$ by

$$\tilde{\Phi}(\xi) = \Phi(\psi^{-1}(\xi)). \tag{53}$$

Since Φ is constant on cylinders of length K, $\tilde{\Phi}$ is constant on cylinders of length 1. Note also that $\tilde{\Phi}$ is well defined since ψ is a homeomorphism. In the last step we will show that not only the systems are equivalent but also their rotation sets. For that we consider the statistical averages from both systems and show that they coincide. Let $x \in X_A$ and $\psi(x) = \xi \in \Sigma_C$. Then

$$\frac{1}{n}\sum_{i=0}^{n-1}\tilde{\Phi}(\sigma^i(\xi)) = \frac{1}{n}\sum_{i=0}^{n-1}\Phi(\psi^{-1}\sigma^i\psi(x)) = \frac{1}{n}\sum_{i=0}^{n-1}\Phi(f^i(x)). \tag{54}$$

All of these pointwise limits give $\mathrm{Rot}_{Pt}(\Phi)$ (resp. $\mathrm{Rot}_{Pt}(\tilde{\Phi})$). As the convex hull of the pointwise rotation set is exactly the general rotation set, the rotation sets of Φ and $\tilde{\Phi}$ coincide. In the last step we prove that also the localized entropy coincides. Let $w \in \mathrm{Rot}(\Phi) = \mathrm{Rot}(\tilde{\Phi})$. By definition we have to prove $h(w) = \sup\{h_\mu(f) : \mu \in \mathcal{M}(f),\ \mathrm{rv}(\mu) = w\} = \sup\{h_{\tilde{\mu}}(\sigma) : \tilde{\mu} \in \mathcal{M}(\sigma),\ \mathrm{rv}(\tilde{\mu}) = w\} = \tilde{h}(w)$. Let $\Theta = \{Z_1, \dots, Z_N\}$ be a finite partition of X_A. The entropy of (X_A, f) of $\mu \in \mathcal{M}(f)$ with respect to Θ is defined by

$$h_\mu(f,\Theta) = \lim_{n\to\infty}\frac{1}{n}H_\mu\bigg(\bigvee_{i=0}^{n-1} f^{-i}\Theta\bigg),$$

see also Definition 2.13. So we get

$$
\begin{aligned}
&H_\mu\Big(\bigvee_{i=0}^{n-1} f^{-i}\Theta\Big) = H_\mu\Big(\bigvee_{i=0}^{n-1}((\psi^{-1}\sigma\psi)^{-i}(\Theta)\Big)\\
&=\sum_{\imath=0}^{n-1}\sum_{j=1}^{N}\mu(\psi^{-1}\sigma^{-i}\psi(Z_j))\log\mu(\psi^{-1}\sigma^{-i}\psi(Z_j)) = H_{\tilde\mu}\Big(\bigvee_{i=0}^{n-1}\sigma^{-i}\tilde\Theta\Big),
\end{aligned}
$$

where $\tilde\mu = \mu\circ\psi^{-1}$ and $\tilde\Theta = \psi(\Theta)$. By equation (52) we know that

$$
H_\mu\Big(\bigvee_{i=1}^{n-1} f^{-i}\Theta\Big) = H_{\tilde\mu}\Big(\bigvee_{i=0}^{n-1}\sigma^{-i}\tilde\Theta\Big).
$$

Since ψ is a homeomorphism we know for $\tilde\mu = \mu\circ\psi^{-1}$ that $h_\mu(f) = h_{\tilde\mu}(\sigma)$. It remains to show that $\mathrm{rv}(\mu) = \mathrm{rv}(\tilde\mu)$.

$$
\mathrm{rv}(\mu) = \int \Phi\ d\mu = \int \Phi\circ(\psi^{-1}\psi)\ d\mu = \int \tilde\Phi\circ\psi\ d\mu = \int \tilde\Psi\ d\tilde\mu = \mathrm{rv}(\tilde\mu).
$$

And therefore we have $h(w) = \tilde h(w)$. $\square$

We obtain the commutative diagram

$$
\begin{tikzcd}
X_A \arrow[r, "f"] \arrow[d, "\psi"'] & X_A \arrow[d, "\psi"] \\
\Sigma_{A_C} \arrow[r, "\sigma"'] & \Sigma_{A_C}\ .
\end{tikzcd}
\tag{55}
$$

Theorem 5.2. *Let (X,f) be a subshift of finite type and suppose $\Phi : X\to\mathbb{R}^m$ is a potential, which is constant on cylinders of length K. Then* $\mathrm{Rot}(\Phi)$ *is a polyhedron, i.e. the convex hull of finitely many points.*

Proof. By Proposition 5.1 (X,f) has the same rotation set as subshift of finite type (Σ_{A_C},σ) with the potential $\tilde\Phi$. Since potentials which are constant on cylinders of length 1 are in particular constant on cylinders of length 2 we know by Ziemian [19] that $\mathrm{Rot}(\Phi)$ is a polyhedron. $\square$

Ziemian proved by induction that the rotation set is a polyhedron if the potential is constant on cylinders of length 2. However, for our approach it is reasonable to work out another proof which gives us more information about the representation of the rotation set by the convex hull. From now on we consider the subshift of finite type (Σ_{A_C},σ) with potential $\tilde\Phi$. We follow Ziemian's idea and work with so called loops. We call a word $\xi_1,\dots,\xi_l,\xi_{l+1}$ a loop of length l if $\xi_1=\xi_{l+1}$ and identify it with the periodic point generated by $(\xi_1,\dots,\xi_l)$. We call a loop of length l elementary if for all $|i-j|<l$, $\xi_i\neq\xi_j$. This means the longest elementary loop is of length d^K. Also note that we basically consider loops independent from their initial point, i.e. $\xi_1,\xi_2,\dots,\xi_l$ and $\xi_2,\xi_3,\dots,\xi_l,\xi_1$ and so on are considered as the same loop. Due to this construction there are only finitely many elementary loops.

By combinatorics the number of elementary loops, disregarding the initial point, in case of a full shift is

$$\sum_{i=1}^{d^K} \binom{d^K}{i} (i-1)!.$$

Hence for a subshift like in our case there are even less elementary loops. The set of all elementary loops is denoted by Ξ. We will cover an arbitrary loop $\xi_1, \ldots, \xi_p$ by elementary loops. This may happen on different levels. The idea is to go to the first symbol which is used more than once and consider the part from the first to the last position with the same symbol. The part in between is considered as a new loop on the next level. Let us bring this in a proper form. The first elementary loop which builds the highest level will look like

$$\xi_1, \ldots, \xi_{i_1-1}, \xi_{l_1}, \ldots, \xi_{i_N-1}, \xi_{l_N}, \ldots, \xi_p, \tag{56}$$

where we inductively define

$$i_n = \min\{l_{n-1} \leq i < p : \exists j > i : \ \xi_i = \xi_j\}, \tag{57}$$

$$l_n = \min\{i_n < l \leq p : \xi_l = \xi_{i_n}\}, \tag{58}$$

and $l_0 = 0$. For the next level we consider the new loops

$$(\xi_{i_1}, \ldots, \xi_{l_1-1}), (\xi_{i_2} \ldots, \xi_{l_2-1}), \ldots, (\xi_{i_N}, \ldots, \xi_{l_N-1})$$

and repeat the procedure iteratively on these loops until we only obtain elementary loops. As in the end it does not make a difference on which level an elementary loop is we keep things easier do not indicate the level. To illustrate this idea please see Figure 1 and Figure 2 on the next page.

Having covered an arbitrary periodic point/loop ξ of length p by (finitely many) elementary loops $\{\chi_1, \ldots, \chi_L\} \subset \Xi$ we can express the rotation vector of the periodic point measure by a convex combination of the rotation vectors of periodic point measures of the elementary loops as follows:

$$\mathrm{rv}(\mu_\xi) = \sum_{i=1}^{L} \frac{p_i}{p} \mathrm{rv}(\mu_i).$$

Note that this representation does not give the same measure, however, for a potential of length 1 the rotation vectors coincide. In the following we discuss this construction. We recall the definition of a periodic point measure,

$$\mu_\xi = \frac{1}{p} \sum_{i=0}^{p-1} \delta_{\sigma^i(\xi)}, \tag{59}$$

where the point measures of the elementary loops are defined analogously. Let $\mu_1, \ldots, \mu_L$ denote these measures. We denote the set of all periodic point measures with respect to an elementary loop by $\mathcal{M}(\Xi)$. Let p_n be the length of the corresponding elementary loop of μ_n. Then we weight $\mathrm{rv}(\mu_n)$

FIGURE 1. Loop on first level

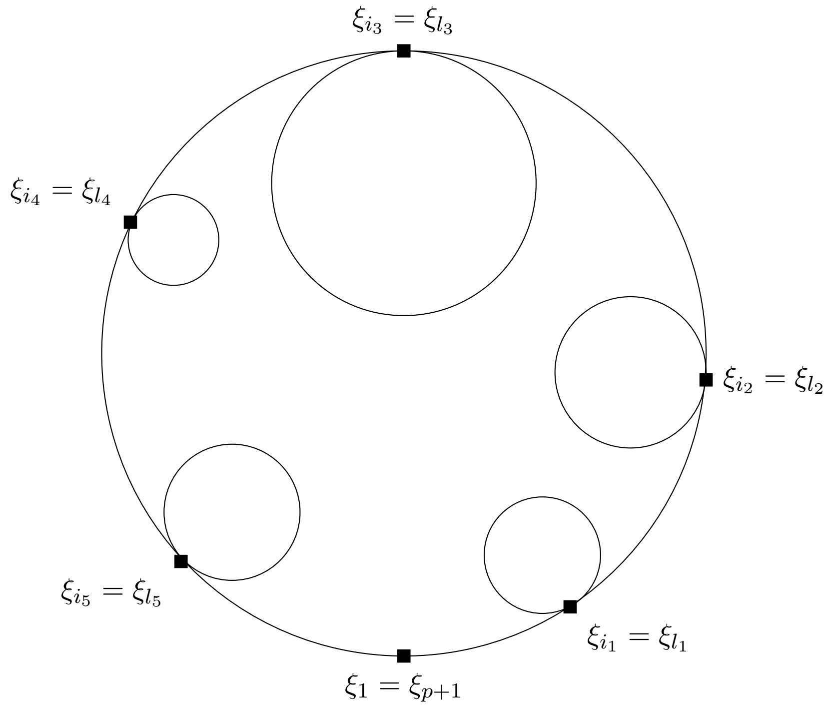
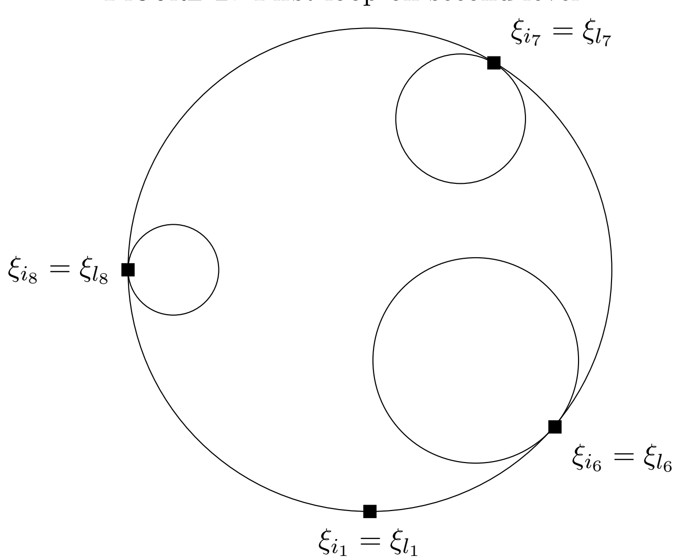


FIGURE 2. First loop on second level



with $\frac{p_n}{p}$. As by construction $\sum_{i=1}^{L} p_i = p$ this is a convex combination. To make this more precise let ξ be a periodic point generated by $(\xi_1, \dots, \xi_p)$ and $\chi_1, \dots, \chi_L$, $L \le p$ elementary loops such that they cover ξ, i.e. $(\xi_1, \dots, \xi_p) = (\chi_1, \dots, \chi_L)$. Note that on the left hand side of the latter equation we have symbols from $\mathcal{A}_C$ and on the right hand side we have elementary loops respectively periodic points. The rotation vector for μ_ξ is given by

$$\mathrm{rv}(\mu_\xi) = \int \tilde{\Phi}\ d\mu_\xi = \sum_{i=1}^{d^K} \mu_\xi(C_i)\tilde{\Phi}(C_i) = \sum_{i=1}^{d^K} \tilde{\Phi}(C_i)\frac{1}{p}\sum_{j=1}^{p} \delta_{\sigma^j(\xi)}(C_i). \tag{60}$$

But on cylinders of length 1 we also have

$$\sum_{j=1}^{p} \delta_{\sigma^j(\xi)}(C_i) = \sum_{k=1}^{L}\sum_{j=1}^{p_k} \delta_{\sigma^j(\chi_k)}(C_i).$$

Thus for (60) we obtain

$$\begin{aligned}\mathrm{rv}(\mu_\xi) &= \sum_{i=1}^{d^k} \tilde{\Phi}(C_i)\frac{1}{p}\sum_{k=1}^{L}\sum_{j=1}^{p_k} \delta_{\sigma^j(\chi_k)}(C_i) = \sum_{i=1}^{d^k} \tilde{\Phi}(C_i)\sum_{k=1}^{L}\frac{p_k}{p}\mu_k(C_i)\\ &= \sum_{k=1}^{L}\frac{p_k}{p}\int \tilde{\Phi}\ d\mu_k = \sum_{k=1}^{L}\frac{p_k}{p}\mathrm{rv}(\mu_k).\end{aligned}$$

Therefore we can represent any rotation vector of a periodic point measure by a finite convex combination of elementary loops, as the periodic point measures are dense we have $\mathrm{Rot}(\tilde{\Phi}) \subset \mathrm{conv}\{\mathrm{rv}(\mu_\xi) : \xi \in \Xi\}$. The converse inclusion is also satisfied as each elementary loop generates a periodic point measure. Since the periodic point measures are dense for transitive subshifts of finite type this is enough to conclude that under our assumptions the rotation set of the new system is a polyhedron. As the two systems have the same rotation set this holds true for the original system. From this construction it may happen that two or more elementary loops have the same rotation vector or that the rotation vector of an elementary loop is not an extreme point. For the shape of the rotation set these loops are redundant but for the localized entropy they do in general matter. For example, if there exists $\pi \in \mathcal{S}_l$, where $\mathcal{S}_l$ denotes the symmetric group with respect to l elements, such that $(\xi_{\pi(1)}, \dots, \xi_{\pi(l)}) = (\tilde{\xi}_1, \dots, \tilde{\xi}_l)$, then they have the same rotation vector. Note that there might be other situations where two elementary loops have the same rotation vector.

The first theorem gives a formula for the localized entropy of an extreme point of the rotation set. The second theorem provides the lower bound for a point on a face and the last one shows how we can calculate the localized entropy in the interior of the rotation set. For all of these theorems we assume the foregoing set-up, namely, (Σ, σ) is a subshift of finite type with alphabet $\mathcal{A}_C$, defined by the map $\psi : (X, f) \to (\Sigma, \sigma)$, and $\tilde{\Phi} : \Sigma \to \mathbb{R}^m$ is a potential constant on cylinders of length 1. Also let $|\Xi| = L$. Let $\chi_1, \dots, \chi_l$

be elementary loops such that $\mathrm{rv}(\chi_1) = \cdots = \mathrm{rv}(\chi_l)$. These elementary loops define a transition matrix Q of a subshift of finite type in the following way: For two symbols $a_i, a_j \in \mathcal{A}_c$ we have $Q(a_i, a_j) = 1$ if and only if there exists $k \in \{1, \ldots, l\}$ such that χ_k has the form $a_1, \ldots, a_i, a_j, \ldots, a_{p_k}$.

Theorem 5.3. *Let $w \in \mathrm{Rot}(\tilde{\Phi})$ be an extreme point, and suppose $\{\mu_1, \ldots, \mu_l\} = \mathcal{M}(\Xi) \cap \mathrm{rv}^{-1}(w)$ with corresponding elementary loops $\chi_1, \ldots, \chi_l$. Then $h(w)$ is determined by the topological entropy of the subshift defined by $\chi_1, \ldots, \chi_l$, namely the logarithm of the spectral radius of the transition matrix.*

Proof. Let $\{\mu_1, \ldots, \mu_l\} = \mathcal{M}(\Xi) \cap \mathrm{rv}^{-1}(w)$ with corresponding elementary loops $\chi_1, \ldots, \chi_l$. These elementary loops define a subshift of finite type. We denote the subshift by $(\tilde{\Sigma}, \tilde{\sigma})$. For this subshift the topological entropy is given by the logarithm of the spectral radius of its transition matrix. Assume there is another measure $\mu \in \mathcal{M}(\sigma)$ which is not supported on the subshift but $\mathrm{rv}(\mu) = w$ with $h_\mu(\sigma) > h(\tilde{\sigma})$. By the ergodic decomposition theorem we have a measure τ on $\mathcal{M}_E(\sigma)$ such that

$$\int_E \int \Phi \; d\nu d\tau(\nu) = \int \Phi \; d\mu, \text{ and } \int_E h_\nu(\sigma) \; d\tau = h_\mu(\sigma). \tag{61}$$

Thus there is at least one ergodic measure, let us say ν' with the same entropy as μ. Note that $\tau\{\nu \in \mathcal{M}_E : \mathrm{rv}(\nu) \neq w\} = 0$ since w is an extreme point. Therefore $\mathrm{rv}(\nu') = w$ and hence ν' is supported on the subshift which completes the proof. $\square$

Theorem 5.4. *Let $w \in \mathrm{Rot}(\tilde{\Phi})$ be a point on a face $F \subset \mathrm{Rot}(\tilde{\Phi})$ and suppose there is no $\mu \in \mathcal{M}(\Xi)$ with $\mathrm{rv}(\mu) = w$. Then $h(w) \geq \sup\{\sum_{i=1}^b \lambda_i h_{\mu_i}(\sigma) : \mu_i \in \mathcal{M}(\sigma),\ \tilde{w}_i$ is extreme point of $F,\ \mathrm{rv}(\mu_i) = \tilde{w}_i,\ \sum_{i=1}^b \lambda_i \tilde{w}_i = w\}$.*

Proof. Let $w \in F$ be as in the Theorem 5.4. Then there exists $(\lambda_1, \ldots, \lambda_b)$ non-negative, where $\dim F = b$, with $\sum_{i=1}^b \lambda_i = 1$ and extreme points of F $\tilde{w}_1, \ldots, \tilde{w}_b$ such that $\sum_{i=1}^b \lambda_i \tilde{w}_i = w$. Obviously we have

$$\Big\{\mu = \sum_{i=1}^b \lambda_i \mu_i : \mu_i \in \mathcal{M}(\sigma),\ \tilde{w}_i \text{ is extr. pt. of } F,\ \mathrm{rv}(\mu_i) = \tilde{w}_i,\ \sum_{i=1}^b \lambda_i \tilde{w}_i = w\Big\}$$
$$\subset \{\mu : \mu \in \mathcal{M}(\sigma),\ \mathrm{rv}(\mu) = w\}.$$

As the measure theoretic entropy is affine we have $h(w) = \sup\{h_\mu(\sigma) : \mu \in \mathcal{M}(\sigma),\ \mathrm{rv}(\mu) = w\} \geq \sup\{\sum_{i=1}^b \lambda_i h_{\mu_i}(f) : \mu_i \in \mathcal{M}(\sigma),\ \tilde{w}_i$ is extr. pt. of F, $\mathrm{rv}(\mu_i) = \tilde{w}_i,\ \sum_{i=1}^b \lambda_i \tilde{w}_i = w\}$. $\square$

We will write a paper where we give a formula to calculate the localized entropy explicitly. There will be a similar theorem for the interior of the rotation set. The goal is to approximate general potentials and their rotation sets by potentials which are constant on cylinders of length K. We are optimistic that in this way we can calculate the localized entropy for general rotation sets.

Appendix A. Convex Analysis

We will give an overview of convex analysis. For most of the theorems we will not provide proofs but refer to [18].

A.1. **Convex sets and functions.** First we define affine sets. A set $A \subset \mathbb{R}^m$ is said to be affine if for all $x, y \in A$ and for all $\lambda \in \mathbb{R}$ also $\lambda x + (1-\lambda)y$ is in A. Further the affine hull of $A \subset \mathbb{R}^m$ is the smallest affine set which contains A and is denoted by $\mathrm{aff}(A)$.
We call $C \subset \mathbb{R}^m$ convex if for any $x, y \in C$ and for all $\lambda \in [0,1]$ $\lambda x + (1-\lambda)y \in C$. The set C is said to be strictly convex if it is convex and every boundary point is an extreme point (Section A.4). We define the convex hull analogously to the affine hull, i.e. the convex hull of C is the smallest convex set which contains C and is denoted by $\mathrm{conv}C$. For any non-zero $b \in \mathbb{R}^m$ and $\beta \in \mathbb{R}$

$$\{x \in \mathbb{R}^m : x.b \leq \beta\}, \qquad \{x \in \mathbb{R}^m : x.b \geq \beta\}$$

are called closed half-spaces. For a function $p : S \to [-\infty, \infty]$, where S is a subset of $\mathbb{R}^m$ we define the epigraph of p by the set

$$\{(x, \mu) : x \in S,\ \mu \in \mathbb{R},\ \mu \geq p(x)\},$$

and denote it by $\mathrm{epi}(p)$. The function p is convex if its epigraph is convex as subset of $\mathbb{R}^{m+1}$. The effective domain of p is defined by

$$\mathrm{dom}(p) = \{x \in \mathbb{R}^m : \exists \mu : (x, \mu) \in \mathrm{epi}(p)\} = \{x \in \mathbb{R}^m : p(x) < \infty\}.$$

We note that this set is again convex, since it is the image of the convex set $\mathrm{epi}(p)$ under a linear transformation. We call p to be proper if its epigraph is non-empty and contains no vertical lines, i.e. $p(x) < \infty$ for at least one x, and $p(x) > -\infty$ for all $x \in S$.

Theorem A.1. *Let $p : S \to (-\infty, \infty]$. Then p is convex if and only if*

$$p(\lambda_1 x_1 + \cdots + \lambda_n x_n) \leq \lambda_1 p(x_1) + \cdots + \lambda_n p(x_n),$$

whenever $\lambda_i \geq 0$ for all i with $\sum_{i=1}^n \lambda_i = 1$.

Some books use this characterization to define convex function.

A.2. **Relative interior.** We equip $\mathbb{R}^m$ with the Euclidean metric and define $B = \{x \in \mathbb{R}^m : d(x, 0) \leq 1\}$ to be the closed unit ball. The closure and interior of $C \subset \mathbb{R}^m$ are defined by

$$\mathrm{cl}(C) = \bigcap \{C + \varepsilon B : \varepsilon > 0\},$$
$$\mathrm{int}(C) = \{x \in \mathbb{R}^m : \exists \varepsilon > 0 : x + \varepsilon B \subset C\}.$$

If we consider a triangle or line segment as subset of $\mathbb{R}^3$, its mathematical interior is empty, although there is a natural interior. In order to fix this issue we define the relative interior, which is the classical interior of C regarded as a subset of its affine hull, i.e.

$$\mathrm{ri}(C) = \{x \in \mathrm{aff}(C) : \exists \varepsilon > 0 : \ (x + \varepsilon B) \cap \mathrm{aff}(C) \subset C\}.$$

We have $\mathrm{ri}(C) \subset C \subset \mathrm{cl}(C)$, $\mathrm{cl}(\mathrm{ri}(C)) = \mathrm{cl}(C)$, and $\mathrm{ri}(\mathrm{cl}(C)) = \mathrm{ri}(C)$.

A.3. **Conjugates.** A proper closed convex set can be written as an intersection of half-spaces, and in particular the epigraph of a proper closed convex function p can be written by such an intersection. Transferring this idea to functions we get the following theorem.

Theorem A.2. *A proper closed convex function p is the pointwise supremum of the collection of all affine functions h that are majorized by p, i.e. $h \leq p$.*

Motivated by this theorem we define the conjugate p^* of p by

$$p^*(x^*) = \sup\{x.x^* - p(x) : x \in \mathbb{R}^m\}.$$

As p^* is the pointwise supremum of affine functions, it is a convex function. Furthermore p^* is a proper closed convex function if and only if p is a proper convex function.

A.4. **Extreme points and faces.** For a convex set $C \subset \mathbb{R}^m$ a convex subset C' of C is called face if every (closed) line segment in C with a relative interior point in C' has both end points in C'. Zero dimensional faces are called extreme points, i.e. x is an extreme point if for all $y, z \in C$ and all $\lambda \in (0, 1)$, $\lambda y + (1 - \lambda)z = x$ implies that $x = y = z$. A point in C is called exposed point if there exists a supporting hyperplane through it containing no other point in C. We note that all exposed points are extreme points. The converse is in general not true. For a counterexample consider the set $\{(x, y) \in \mathbb{R}^2 : y \geq 0, \ x^2 + y^2 \leq 1\} \cup \{(x, y) \in \mathbb{R}^2 : y \leq 0, \ \max(x^2, y^2) \leq 1\}$. The points $(1, 0)$ and $(-1, 0)$ are extreme points but not exposed points.

Theorem A.3. *For a closed convex set C the exposed points are dense in the extreme points, i.e. every extreme point can be approximated by a sequence of exposed points.*

In the context of extreme points we also introduce polyhedral convex sets. A polyhedral convex set or polyhedron in $\mathbb{R}^m$ is defined as a set which can be expressed as intersection of a finite collection of closed half-spaces, i.e. the set of solutions to a finite system of inequalities of the form

$$x.b_i \leq \beta_i, \quad i = 1, \dots, n, \ \beta_i \in \mathbb{R}, \ x, b_i \in \mathbb{R}^m.$$

To give a characterization of polyhedral sets, we state the following theorem.

Theorem A.4. *For a convex set C the following are equivalent:*

1. *C is a polyhedron.*
2. *C is closed and has only finitely many faces.*
3. *C is finitely generated, i.e. it is the convex hull of a finite set of points and directions.*

A.5. **Directional derivatives and subgradients.** We recall the definition of (one-sided) directional derivatives for $p : \mathbb{R}^m \to \mathbb{R}$ in direction $y \in \mathbb{R}^m$,

$$d^+p(x;y) = \lim_{t \downarrow 0^+} \frac{p(x+ty)-p(x)}{t},$$
$$d^-p(x,-y) = \lim_{t \uparrow 0^-} \frac{p(x+ty)-p(x)}{t}.$$

The two sided directional derivative, denoted by $dp(x;y)$, exists if and only if

$$d^+p(x;-y) = d^-p(x;y).$$

Theorem A.5. *Let $p : \mathbb{R}^m \to \mathbb{R}$ be a convex function finite at the point x. For each $y \in \mathbb{R}^m$ the difference quotient in the definition of d^+p is a non-decreasing function of $t > 0$, in particular d^+p exists and*

$$d^+p(x;y) = \inf_{t>0} \frac{p(x+ty)-p(x)}{t}.$$

Proof. We define the function $g(t) = \frac{1}{t}p(x+ty) - p(x)$ and let $\lambda \in (0,1)$. By convexity of p we obtain

$$p(x+\lambda ty) = p((1-\lambda)x + \lambda(x+ty)) \leq (1-\lambda)p(x) + \lambda p(x+ty).$$

Applying this inequality to g brings

$$g(\lambda t) = \frac{p(x+ty)-p(x)}{\lambda t} \leq \frac{\lambda(p(x+ty)-p(x))}{\lambda t} = g(t).$$

Therefore g is non-decreasing in t, and thus the limit $t \to 0$ exists and coincides with the infimum. $\square$

We will now establish a connection between subdifferentials and the differentiability of a function. We recall the definitions of subdifferntials and subgradients. A vector $x^* \in \mathbb{R}^m$ is called subgradient for a convex function p at x, if

$$x^*.(z-x) + p(x) \leq p(z), \quad \forall z \in \mathbb{R}^m.$$

The set of all subgradients of p at x is called subdifferential of p at x and denoted by $\partial p(x)$. Furthermore, a function p from $\mathbb{R}^m$ to $[-\infty, \infty]$, finite at x, is said to be differentiable at x if and only if there exists a vector x^* (unique) such that

$$\lim_{z \to x} \frac{p(z) - p(x) - \langle x^*, z-x \rangle}{|z-x|} = 0.$$

If such an x^* exists, it is called gradient of p at x and is denoted by $\bigtriangledown p(x)$.

Theorem A.6. *Let $p : \mathbb{R}^m \to \mathbb{R}$ be a convex function, and let x be a point where p is finite. If p is differentiable at x, then $\bigtriangledown p$ is the unique subgradient of p at x, so that in particular*

$$p(z) \geq p(x) + \langle \bigtriangledown p(x), z - x \rangle, \quad \forall z \in \mathbb{R}^m.$$

Conversely, if p has a unique subgradient at x, then p is differentiable at x.

Proof. First we assume that p is differentiable at x. Then $dp(x; \cdot)$ is the linear function $\langle \bigtriangledown p(x), \cdot \rangle$. By Lemma 3.4, the subgradients at x are the vectors x^* such that

$$\langle \bigtriangledown p(x), y \rangle \geq \langle x^*, y \rangle, \quad \forall y \in \mathbb{R}^m.$$

This inequality is only satisfied if and only if $x^* = \bigtriangledown p(x)$. Thus $\bigtriangledown p(x)$ is the unique subgradient of p at x.
For the other direction we assume that p has a unique subgradient x^* at x. We define the convex function g by

$$g(y) = p(x + y) - p(x) - \langle x^*, y \rangle.$$

Therefore g has 0 as its unique subgradient in the origin. For differentiablity we must show that this implies

$$\lim_{y \to 0} \frac{g(y)}{|y|} = 0. \tag{62}$$

By Lemma 3.4 the closure of $d^+g(0; \cdot)$ is the support function of $\partial g(0)$, which in this particular case is the constant function 0. Furthermore $d^+g(0; \cdot)$ cannot differ from its closure other than at boundary points of its effective domain. Thus $d^+g(0; \cdot) = 0$, and we have

$$0 = d^+g(0; \alpha) = \lim_{\lambda \downarrow 0} \frac{g(\lambda\alpha) - g(0)}{\lambda}, \quad \forall \alpha \in \mathbb{R}^m.$$

But we know $g(0) = 0$ and by Theorem A.5 the difference quotient is a non-decreasing function of λ. We define the convex functions h_λ by

$$h_\lambda(\alpha) = \frac{g(\lambda\alpha)}{\lambda}, \quad \lambda > 0.$$

Thus h_λ decrease pointwise to the constant function 0 as λ decreases to 0. Let B denote the Euclidean unit ball, and let $\{a_1 \ldots, a_n\}$ be any finite collection of points whose convex hull includes B. Therefore each $\alpha \in B$ can be expressed as a convex combination

$$\alpha = \theta_1 a_1 + \cdots + \theta_n a_n,$$

and so we obtain

$$\begin{aligned} 0 \leq h_\lambda(\alpha) &\leq \sum_{i=1}^{n} \theta_i h_\lambda(a_i) \\ &\leq \max\{h_\lambda(a_i) : i = 1, \ldots, n\}. \end{aligned}$$

Since we know that $h_\lambda(a_i)$ decreases to 0 pointwise for each a_i as $\lambda \downarrow 0$, we can conclude that $h_\lambda(\alpha)$ decreases to 0 uniformly in $\alpha \in B$ as $\lambda \downarrow 0$. That is, for given $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$\frac{g(\lambda\alpha)}{\lambda} \leq \varepsilon, \quad \forall \lambda \in (0, \delta], \quad \forall \alpha \in B.$$

Going back to our actual goal from equation (62), each vector y with $0 < |y| \leq \delta$ can be expressed as $\lambda\alpha$ with $\lambda = |y|$ and $\alpha \in B$. So we have

$$\frac{g(y)}{|y|} < \varepsilon,$$

whenever $0 < |y| \leq \delta$. This proves that the limit we asked for is 0, and thus the function p is differentiable. □

Theorem A.7. *For any proper convex function p and any vector $x \in \mathbb{R}^m$, the following are equivalent:*

1. *$x^* \in \partial p(x)$.*
2. *$z.x^* - p(x)$ achieves its supremum in z at $z = x$.*
3. *$p(x) + p^*(x^*) \leq x.x^*$.*
4. *$p(x) + p^*(x^*) = x.x^*$.*

Corollary A.8. *For a closed proper convex function $p : \mathbb{R}^m \to \mathbb{R}$, ∂p^* is the inverse of ∂p in the sense of multivalued mappings, i.e. $x \in \partial p^*(x^*)$ if and only if $x^* \in \partial p(x)$.*

Appendix B. Further theorems

Theorem B.1 ([4], p. 417, Theorem 10)**.** *If K_1, K_2 are two disjoint closed convex subsets of a locally convex linear topological space $\mathcal{X}$ and if K_1 is compact, then there exists c and $\varepsilon > 0$, and a continuous linear functional $F : \mathcal{X} \to \mathbb{C}$ such that,*

$$\mathcal{R}e(F(K_2)) \leq c - \varepsilon < c < \mathcal{R}e(F(K_1)),$$

where $\mathcal{R}e(F(K))$ denotes the real part of $F(K)$.

Abstract German

Für ein stetiges dynamisches System (X, f) auf einem kompakten metrischen Raum und ein stetiges Potential $\Phi : X \to \mathbb{R}^m$ ist die verallgemeinerte Rotationsmenge definiert als die Teilmenge von $\mathbb{R}^m$, die aus allen Integralen von Φ bezüglich aller f-invarianten Wahrscheinlichkeitsmaße besteht. Wir geben eine Einleitung zu Konzepten und Ergebnisse aus der Rotationstheorie. Für $\alpha \in \mathbb{R}^m$ und dem Potential $\alpha.\Phi$, stellen wir, in Anlehnung an [6], eine Verbindung zwischen den Rotationsvektoren von Gleichgewichtsmaßen und den Richtungsableitungen des Drucks her. Die localized Entropie an einem Punkt in der Rotationsmenge is definiert als das Supremum der Maßtheoretischen Entropie über alle f-invarianten Wahrscheinlichkeitsmaßdessen Integrale diesen Punkt ergeben. Wir betrachten einen Subshift of finite type

und ein Potential, welches konstant auf Zylindern der Länge K ist. Inspiriert durch [19] zeigen wir, dass in diesem Fall die Rotationsmenge ein Polyeder ist und geben eine Formel für die localized entropy an Extrempunkten und Oberflächen der Rotationsmenge.

References

[1] R. Bowen, *Equilibrium states and the ergodic* theory of Anosov diffeomorphisms, Lecture Notes in Math. **470**, Springer-Verlag, Berlin, 1975.

[2] A. Boyarsky and P. Góra, *Laws of Chaos - Invariant measures and Dynamical Systems in One Dimension*, Birkbäuser Boston, 1997

[3] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, Westview Press, ISBN 0-8133-4085-3.

[4] N. Dunford, J. T. Schwartz, *Linear operators Part I*, Wiley-Interscience, 1988, New York.

[5] N. Haydn, *Phase transition in one-dimensional subshifts* Discrete and Continuous Dynamical Systems, (5) **33**, (2013), 1965–1973.

[6] O. Jenkinson, *Rotation, entropy, and equilibrium states*, Trans. Amer. Math. Soc. **353** (2001), 3713–3739.

[7] A. Katok and B. Hasselblatt, *Introduction to the modern theory of dynamical systems. With a supplementary chapter by Katok and Leonardo Mendoza*, Encyclopedia of Mathematics and its Applications, **54**, Cambridge University Press, Cambridge, 1995.

[8] B. P. Kitchens, *Symbolic dynamics - One-sided, Two-sided and Countable State Markov Shifts*, Springer.

[9] T. Kucherenko and C. Wolf, *Geometry and entropy of generalized rotation sets*, Israel Journal of Mathematics **1999** (2014), 791-829.

[10] T. Kucherenko and C. Wolf, *Entropy and rotation sets: A toymodel approach*, Commun. Contemp. Math. DOI: 10.1142/S0219199715500832.

[11] D. Lind and B. Marcus, *An introduction to symbolic dynamics and coding*, Cambridge University Press, 1995.

[12] D. Kwietniak, M. Lacka and P. Oprocha, *A panorama of specification-like properties and their consequences*, eprint arXiv:1503.07355, 03/2015.

[13] M. Misiurewicz, *Rotation Theory*, Misiurewicz's webpage.

[14] R. Phelps, *Lectures on Choquet's theorem*, Lecture Notes in Mathematics, Springer, 2001.

[15] M. Rams, *Measures of maximal dimension for linear horsehoes*, Real Analysis Exchange, Vol 31(1), 2005/2006, pp. 55-62.

[16] P. Walters, *An introduction to ergodic theory*, Graduate Texts in Mathematics 79, Springer, 1981.

[17] P. Walters, *Differentiability properties of the pressure of a continuous transformation on a compact metric space*, J. London Math. Soc. (2), **46** (1992), 471-481.

[18] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.

[19] K. Ziemian, *Rotation sets for subshifts of finite type*, Fund. Math. **146** (1995), 189–201.