Pressure and its Differentiability in Non Compact Spaces

João Felipe Rodrigues
Kelvyn Welsch

March 2021

Contents

1 Introduction and Historical Motivation

The history of the thermodynamic formalism can be traced back to late 1960s, with the seminal work of David Ruelle. He, among others like Yakov Sinai and Rufus Bowen, introduced techniques borrowed from physics to treat difficult questions of the theory of dynamical systems, mainly in compact spaces, although Bowen himself made contributions to the theory of entropy in non-compact spaces in 1973 [Bow73]. Pesin and Pitskel extended this theory to treat pressure in non-compact spaces [PP84].

Nonetheless, yet in 1960s, Boris Markovich Gurevich already considered topological entropy for specific non-compact spaces, namely Markov shifts with countable elements in the alphabet. Zargaryan [Zar86] extend this study to topological pressure in 1986, culminating in the paper of 1998 by Gurevich entitled “Thermodynamic formalism for countable symbolic Markov chains" [GS98]. Both Zargaryan and Gurevich papers follow a process in which a compactification is considered.

Those papers were very prolific, and inspired some advances in the theory. We can cite as example a series of papers by Sarig starting in 1999 [Sar99] as well as this very theory presented here. A remarkable difference is that, while Gurevich consider pressure defined in functions that depend only on finitely many coordinates and Sarig consider only continuous functions with some regularity (like summable variation), we actually consider the whole class of continuous functions.

Further developments on the subject had been made, although most of them assume local compactness and extra assumptions on the regularity of functions.

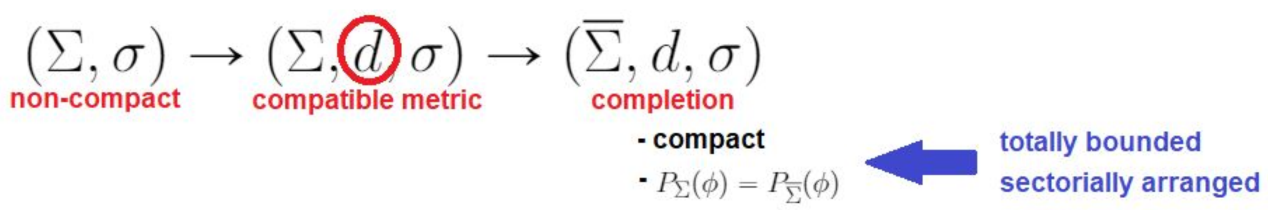
Now, we are going to briefly present the general ideia of the theory that will be developed in this text, originally formulated by Godofredo Iommi and Mike Todd [IT20]. We will be mainly concerned with the topological pressure, which is the principal object of the theory of thermodynamic formalism. For those who are not familiar with this concept, we may consider the pressure as a slightly variation of the entropy, depending upon the choice of a real function defined on the space (particularly continuous functions). The precise definition in compact spaces is the following.

$$P(\phi) = \sup_{\mu \in \mathcal{M}_T} \left(h_\mu(T) + \int \phi d\mu \right)$$

Note that $P(0) = h_{top}(T)$.

Our aim is, therefore, to come up with a suitable notion of pressure in the non-compact case, as well as investigate its differentiability. More precisely, we are going to see that, in the compact case, the pressure is differentiable at large sets (in some sense), and we are going to require that our notion of pressure in the non-compact case be differentiable in large sets as well.

In what follows, we describe our plan to reach the intended aim. We start with a Markov shift (Σ, σ) endowed with the topology generated by cylinders. Then, we define a metric d compatible with the topology and consider the completion of this metric space. We require, then, that the resulting space, $(\overline{\Sigma}, \sigma)$ be (i) compact and that (ii) the pressure coincide with the pressure in the original space. In order to accomplish so, we must impose that (Σ, σ) be (i) totally bounded and (ii) sectorially arranged. Each condition assures the respective requirement. The definition of sectorially arranged will be exposed later. The previous strategy can be summarized in the following scheme:



2 Summary of Thermodynamic Formalism

2.1 Motivation - Helmholtz Free Energy and the Variational Principle

The ergodic theory provides us some results about a dynamical system endowed with an invariant measure. However, different measures usually offer distinct information about the system. A natural question is, therefore, which measure to pick. For some reasons, there is a class of measures inspired by physical problems that gives us good results. The more relevant physical situation is described below.

Suppose we have a system B with temperature T in contact with a system A such that A and B can exchange energy but not particles, the system $A + B$ is isolated and such that B is much larger than A so that A can't affect B temperature. In other words, B is a heat bath for A. Statistical mechanics tell us that the probability of finding the system A in a microstate $\{q_j\}$ with energy $E(\{q_j\})$ is:

$$P(\{q_j\}) = \frac{e^{-\beta E(\{q_j\})}}{Z(\beta)}$$

where $\beta = 1/k_B T$ and:

$$Z(\beta) = \sum_{\{q_j\}} e^{-\beta E(\{q_j\})}$$

is called the partition function. This "probability measure" was first deduced by Boltzmann and Gibbs independently, and named after them. In physics, this is also called the "canonical ensemble".

As expected, $\{q_j\}$ is typically infinity, so that the expressions above are often not well-defined mathematically (although the physicists usually come up with some tricks to "calculate" them). Hence, we cannot use them directly. Instead, we must make some adaptations. This leads us to some different variations of it. The most famous are the DLR measures (based on insights by Dobrushin, Lanford and Ruelle) and the measures due to Bowen.

However, a remarkable fact is that the Gibbs measure is usually obtained in physics by minimizing a function called "Helmholtz free energy", defined by $F = U - TS$ in thermodynamics, where U is the average of E. Supposing we have only a finite number of states, the free energy is given by:

$$F = \sum_{i=1}^{N} p_i U_i - \frac{1}{\beta}\left(-\sum_{i=1}^{N} p_i \log p_i\right)$$

Where the expression for the entropy is motivated by information theory. It's possible to show that, in thermodynamic limit, the first term tends to $\int U d\mu$ and, using Sinai's Generator Theorem, the second tends to something proportional to $-\frac{1}{\beta} h_\mu(T)$, where $h_\mu(T)$ is the metric entropy. Thus, finding a measure that minimizes F is equivalent to finding a measure that maximizes the following:

$$h_\mu(T) + \int \phi d\mu$$

Where ϕ is proportional to $-U$. A measure that maximizes this expression is called an equilibrium measure. Thinking of the pressure as βF, we can formally state the variational principle:

Theorem 2.1 (Variational Principle)**.** *Let X be a compact metric space, suppose $\varphi : X \to \mathbb{R}$ is continuous and the topological entropy is finite, then:*

$$P(\varphi) = \sup\left\{h_\mu(T) + \int \varphi d\mu : \mu \in \mathcal{M}(T)\right\}$$

Where $\mathcal{M}(T)$ is the set of all T-invariant Borel probabilities measures on X.

♣

Although this property can be used to define the pressure, a more direct definition of it is given in the next section.

2.2 Topological Pressure

As Helmholtz free energy is crucial in statistical mechanics, the pressure will be the most important concept in the theory. Furthermore, as seen, we can define it by means of a variational principle, so that this concept seems related to the topological entropy. Indeed, just like the topological entropy, the similarity goes on and the (topological) pressure can also be defined in terms of span and separated sets. Now, we will present the definition using separated sets.

Definition 2.1 (Topological Pressure)**.** Let (X, d) be a compact metric space, $T : X \to X$ be continuous and $\varphi \in C(X)$. For each $n \geq 1$, define:

$$Q_n(T, \varphi, \epsilon) = \sup\left\{\sum_{x\in E} e^{\sum\limits_{i=0}^{n-1} \varphi(T^i x)} : E \subset X \text{ is } (n, \epsilon) \text{ - separated.}\right\}$$

Define $Q(T, \varphi, \epsilon) = \limsup\limits_{n\to+\infty} \frac{1}{n} \log Q_n(T, \varphi, \epsilon)$. The pressure of T is then the functional $P : C(X) \to \mathbb{R} \cup \{\infty\}$ defined by $P(\varphi) = \lim\limits_{\epsilon\to 0} Q(T, \varphi, \epsilon)$.

♣

Remember that every $(n, \epsilon)-$separated set E has finite cardinality, since the space is compact.

Moreover, although this definition may indicate that the pressure depends on the metric, the variational principle assures that it depends only upon the topology (hence the name "topological" pressure). Still with the aid of the variational principle, we can prove the following:

Proposition 2.2. *Let X be a compact metric space such that $T : X \to X$ is continuous, $h_{top}(T) < +\infty$ and $\varphi \in C(X)$. Then:*

1. *If $c \in \mathbb{R}$, then $P(\varphi + c) = P(\varphi) + c$;*
2. *If $\varphi, \psi \in C(X)$ are such that $\varphi \leq \psi$, then $P(\varphi) \leq P(\psi)$;*
3. *The pressure P is Lipschitz continuous and convex.*

Proof. Since all the measures μ in $\mathcal{M}(T)$ are probability measures, we have that $P(\varphi + c) = \sup(X + c)$, where $X = \left\{h_\mu(T) + \int \varphi d\mu : \mu \in \mathcal{M}(T)\right\}$ and $X + c := \{x + c : x \in X\}$. As we know from elementary analysis, $\sup(X + c) = \sup(X) + c$, from where the first proposition holds.

As for the second proposition, define:

$$A := \left\{h_\mu(T) + \int \varphi d\mu : \mu \in \mathcal{M}(T)\right\}$$

$$B := \left\{ h_\mu(T) + \int \psi d\mu : \mu \in \mathcal{M}(T) \right\}$$

Directly from these definitions, we see that for every $x \in A$ there is a $y \in B$ such that $x \leq y \leq \sup B$. Since $\sup B$ is an upper bound for A, we have $\sup A \leq \sup B$. Then, this proves that $P(\varphi) \leq P(\psi)$, as we wanted.

The first two properties can be combined together to prove that P is Lipschitz. In fact:

$$P(\varphi) - |\varphi - \psi| \leq P(\psi) \leq P(\varphi) + |\varphi - \psi| \implies |P(\varphi) - P(\psi)| \leq |\varphi - \psi|$$

Where we used the fact that $\varphi - |\varphi - \psi| \leq \psi \leq \varphi + |\varphi - \psi|$, which in turn implies that $P(\varphi) - |\varphi - \psi| \leq P(\psi) \leq P(\varphi) + |\varphi + \psi|$ (for example, note that for each $x \in X$ we have $-(\varphi(x) - \psi(x)) \leq |\varphi(x) - \psi(x)| \leq |\varphi - \psi|$).

For the final property, note that if $f_\mu : C(X) \to \mathbb{R}$ is defined as $f_\mu(\varphi) = h_\mu(T) + \int \varphi d\mu$ for a measure $\mu \in \mathcal{M}(T)$, then:

$$P(\varphi) = \sup\{f_\mu(\varphi) : \mu \in \mathcal{M}(T)\}$$

Moreover, given such a measure μ, each function f_μ is convex, since given $t \in [0,1]$ and $\varphi, \psi \in C(X)$ we have:

$$f_\mu(t\varphi + (1-t)\psi) = h_\mu(T) + \int (t\varphi + (1-t)\psi) d\mu$$

$$= th_\mu(T) + t\int \varphi d\mu + (1-t)h_\mu(T) + (1-t)\int \psi d\mu$$

$$= t\left(h_\mu(T) + \int \varphi d\mu\right) + (1-t)\left(h_\mu(T) + \int \psi d\mu\right) = tf_\mu(\varphi) + (1-t)f_\mu(\psi)$$

Thus, we only need to show that the pointwise supremum of convex functions is again convex, i.e, the function $f(x) := \sup\limits_{i \in I} f_i(x)$ is convex, if each f_i is. To see why this is true, just note that for each $t \in [0,1]$ and $x, y \in X$ we have:

$$f(tx + (1-t)y) = \sup_{i \in I} f_i(tx + (1-t)y) \leq \sup_{i \in I}(tf_i(x) + (1-t)f_i(y))$$

Since if $g_i(x,y) \leq h_i(x,y)$ for all $i \in I$, then $\sup\limits_{i \in I} g_i(x,y) \leq \sup\limits_{i \in I} h_i(x,y)$. Now, note that for each $i \in I$, we have $tf_i(x) \leq \sup\limits_{i \in I} tf_i(x)$ and $(1-t)f_i(y) \leq \sup\limits_{i \in I}(1-t)f_i(y)$, so that $tf_i(x) + (1-t)f_i(y) \leq \sup\limits_{i \in I} tf_i(x) + \sup\limits_{i \in I}(1-t)f_i(y)$ and then $\sup\limits_{i \in I}(tf_i(x) + (1-t)f_i(x)) \leq \sup\limits_{i \in I} tf_i(x) + \sup\limits_{i \in I}(1-t)f_i(y)$ and thus:

$$f(tx + (1-t)y) \leq \sup_{t \in I} tf_i(x) + \sup_{i \in I}(1-t)f_i(y)$$

Now, remember from elementary real analysis that the supremum of the product by a non-negative constant is the product of the supremum by that constant. Thus:

$$f(tx + (1-t)y) \leq t\sup_{i \in I} f_i(x) + (1-t)\sup_{i \in I} f_i(y) = tf(x) + (1-t)f(y)$$

And thus the proposition is proved.

∎

2.3 Markov Shifts

The most used metric in this setting is:

$$d_m(x, y) = 2^{-\min\{n \in \mathbb{N}; x_n \neq y_n\}} \tag{1}$$

The m subscript stands for "minimum". It will be important when we deal with another metric. If the metric is implicit, though, we will use just d.

The number 2^{-1} is often substituted by an arbitrary number $p \in (0, 1)$. It is easy to see that the open balls in this metric are actually clopens and coincide with the cylinders. Because of that, the topology is often called the topology generated by the cylinders. Another important fact is that the balls are either coincident or disjoint. It means that we can assign a "natural" partition to Σ for each radius ϵ. By the compactness of Σ, it is clear that each partition like those will be finite (it can also be seen by noting that there are finitely many possible cylinders with n letters).

Definition 2.2. Let (Σ, σ) be a Markov shift. We say that Σ is topologically mixing if, for every pair of non-empty open sets, U, V, there exists a natural number N such that, for every $n \geq N$

$$U \cap \sigma^{-n}(V) \neq \emptyset$$

Actually, we are going to use more often the following equivalent condition:

Proposition 2.3. *Σ is topologically mixing if, and only if, for every pair of letters i, j, there exists a number N such that, for every $n \geq N$, there exists an admissible word connecting i and j with length n.*

Proof. Suppose that Σ is topologically mixing. Given i and j, take $[i]$ and $[j]$, that are obviously open sets. By hypothesis, there exists N such that, for any $n > N$, we have a x with $x \in [i]$ and $x \in \sigma^{-n}[j]$. Noticing that $\sigma^{-n}[j]$ is the set of all admissible words such that j is the n-th letter, we have that the first letter of x is i while the n-th is j, whence there is an admissible word with the property required.

Reciprocally, take any non-empty open sets U and V. Then, there are $x \in U$ and $y \in V$. Suppose that j is the first letter of y. By definition of open set, there is a cylinder $[w]$ with $x \in [w] \subset U$, with w a word of length N_o and last letter i. There is a number N_1 such that, for every $n > N_1$, there is an admissible word that connects i and j. But then, there is point $z \in [w]$ such that $z \in U \cap \sigma^{-(N_0+n)}V$. Just make a concatenation of w with the admissible word found and y. This is true for every $n > N_0 + N_1$, as we wanted. ■

Definition 2.3. Let (Σ, σ) be a Markov shift and $\phi : \Sigma \to \mathbb{R}$. The $n-$th variation of ϕ is defined by $V_n(\phi) = \sup\{|\phi(x) - \phi(y)|, x_i = y_i, 0 \leq i \leq n-1\}$ ♣

Obviously, $(V_n(\phi))_n$ is a non-increasing sequence for every ϕ. Eventually, however, we may wish to impose some certain rate of decrease. That is the case of the notion of local Hölder continuity, that will be presented soon. Before, we will see some basic properties of the variation.

Firstly, follows directly from the definition that $V_0(\phi) = diam(\phi(\Sigma))$, hence ϕ is bounded if, and only if $V_0(\phi) < \infty$. More generally, $V_n(\phi) = \sup\{diam(\phi([w])), w \in W_n\}$, where the supremum is taken over W_n, the set of all admissible words of length n. As mentioned, the set of cylinders $[w]$ with $w \in W_n$ always form a partition and is always finite if the alphabet is finite as well. In this setting, we can extend our former consideration and claim that ϕ is bounded if, and only if $V_n(\phi) < \infty$ for *some* n. Indeed, it is well known that the union of a finite number of sets with finite diameter also has finite diameter. The result follows from the fact that the union of the image of a family of sets is the image of the union, which in this case is Σ, once we are talking about a partition. Thus, if the alphabet is finite, $(V_n(\phi))$ is always finite or always infinite. As expected, this is not the case when the alphabet is infinite.

Definition 2.4. A function $\phi : \Sigma \to \mathbb{R}$ is called locally Hölder continuous if there is a constant $A > 0$ such that:

$$V_n(\phi) \leq Ap^n$$

for every $n \geq 1$, where p is the coefficient of the metric. We will denote by $\mathcal{C}^H$ the set of all locally Hölder continuous functions for some p.

It is not difficult to show that ϕ is locally Hölder continuous if, and only if, ϕ is lipschitz in every cylinder, *as long as the Lipschitz constant may be the same for every cylinder.*

Actually, weaker conditions will often do. One of the weakest conditions that lead us to good results is the Walters conditions. In this text, we will be satisfied with the summable variation condition. Like the name suggests, a function with summable variation is a function such the summation of all variations converges.

Now, we are going to see an equivalent condition for local compactness.

Proposition 2.4. *Let (Σ, σ) be a transitive CMS. Then, Σ is locally compact if, and only if, the row sum of the transition matrix is always finite.*

Proof. We will prove this using the metric of the minimum. But this makes no difference, once we already know the topologies are the same.

First, suppose that the row sum is always finite and take a $x \in \Sigma$. We claim that $B(x, 1)$ is compact. Notice that this ball is nothing but the cylinder $[x_o]$. Now, we are going to define a sequence of sets inductively. Take $N_1 = \{n \in \mathbb{N}; A(x_o, n) = 1\}$ and $N_k = \{n \in \mathbb{N}; A(x_{k-1}, n), x_{k-1} \in N_{k-1}\}$, where $A(n, m)$ is the nm element of the transition matrix. Each N_k is finte. Indeed, by hypothesis, $\sum_n A(x_o, n) < \infty$, and therefore, there are finitely many letters that are able do succeed x_o, whence N_1 is finite. Moreover, still by the hypothesis, each $y \in N_{k-1}$ contributes with finitely many letters in N_k. The conclusion follows from the inductive hypothesis. As is known, a subset of the naturals (with the discrete topology) is compact if, and only if, it is finite. We conclude that each N_k is compact. Lastly, it is obvious that $[x_o] = \bigtimes_{k \geq 1} N_k$. Tychonoff theorem then tells us that $[x_o]$ is compact in the product topology. The statement follows from the fact that the cylinder topology is the product topology.

We are now going to prove the contrapositive. Let's assume that there is a row whose sum is infinite. Denote this row by x_o. As the shift is transitive, there is a admissible word beginning and ending with x_o. Take a infinite concatenation of it, that is $x = x_o ... x_o ... x_o ....$ We claim that there is no compact neighborhood of x. Actually, we are proving that there is no compact ball centered in x, but it is enough. It follows trivally from the fact that the balls are clopen and that closed subsets of a compact set are also compact. Using the same notation as in the previous paragraph, given $\epsilon > 0$, $B(x, \epsilon)$ is a cylinder of the form $N = \bigtimes_{k > k_o} N_k$, for some k_o. If N was compact, then N_k would be compact for every $k > k_o$ (look the next paragraph). However, by the construction of x, for every k_o, there is a $i > k_o$ such that N_{i-1} contains x_o. As the x_o row has infinite sum, there are infinitely many letter that are able to succeed x_o. Hence, N_i is certainly infinite, since it contains at least all those letters. N_i is then non-compact and we are left with no hope that $N = B(x, \epsilon)$ is compact.

In fact, let suppose that there is a N_i non-compact but, even so, $N = \bigtimes_k N^k$ is compact. By hypothesis, there is an open cover $(U_i)^j$ of N_i that doesn't admit a finite subcover. Just take $U^j = \bigtimes_k U_k^j$ with $U_k^j = N_k$ for every $k \neq i$ and every n. You can already see that this is an open cover of N (thanks to the definition of product topology but, as we hope N is compact, there must exist a finite subcover. Just check that it implies the existence of a finite subcover of N_i. Contradiction!! ∎

Corollary 2.4.1. *If Σ is a topologically mixing CMS, then local compacteness implies that, for every N, there are $n, m > N$ with $A(n, m) = 1$.*

Proof. We will prove this by a contrapositive. Thus, it is enough to show that, given the existence of a $N > 0$ such that $A(n, m) = 0$ for all $n, m > N$, then there is a $i \leq N$ such that $A(i, j) = 1$, for all $j > N$ (and this will be enough, since sum of the elements of the i'th row will be infinite).

Suppose then that such a $N \geq 1$ exists and take any $k \geq 1$. By the topological mixing property, there is a finite allowed word $x = (x_1, x_2, ..., x_n) \in \Sigma$ such that $(1, x_1, x_2, ..., x_n, N+k, ...)$ is allowed. Since $A(x_n, N+k) = 1$ and $N+k > N$, we know for sure that $x_n \leq N$. Thus, we've proved that for every $j > N$ there exists an element m of $\{1, 2, ..., N\}$ such that $A(m, j) = 1$ (i.e, there are infinitely many arrows going directly from $\{1, 2, ..., N\}$ to $\{N+1, N+2, ...\}$). Now, suppose that for every $m \leq N$ there are finitely many elements $j \geq N$ such that $A(m, j) = 1$. If this were the case, then there would we finitely many $j \geq N$ such that there exists some $m \leq N$ satisfying $A(m, j) = 1$. We can then take a sufficiently big $j \geq N$ such that there are no elements $m \leq N$ such that $A(m, j) = 1$, contradicting what we've just proved. Thus, there is at least one element $i \leq N$ such that $A(i, j) = 1$, for all $j > N$, as we wanted.

■

2.3.1 Ruelle Operator and RPF theorem

A key definition in the study of thermodynamic formalism is the Ruelle operator. In this text, we do not intend to explore this object in all its power, just the necessary for future proofs.

Definition 2.5. Let (Σ, σ) be a Markov shift and $\phi \in \mathcal{C}^H$. We define the operator $\mathcal{L}_\phi : \mathcal{C}(\Sigma) \to \mathcal{C}(\Sigma)$ by:

$$\mathcal{L}_\phi(f)(x) = \sum_{y;\sigma(y)=x} e^{\phi(y)} f(y)$$

The function ϕ may be suppressed if there is no chance of confusion. One can easily check that the the expression below is the same:

$$\mathcal{L}(f)(x) = \sum e^{\phi(ax)} f(ax)$$

Where the summation is taken over all possible values of a. If the alphabet is finite, it is clear that the summation is always defined. If it is not the case, the convergence is more delicate.

By a simple induction we can see that:

$$\mathcal{L}(\mathcal{L}^n f)(x) = \sum_a e^{\phi(ax)} (\mathcal{L}^n f(ax)) = \sum_a e^{\phi(ax)} \left(\sum_{\underline{u}} e^{\phi_n(\underline{u}ax)} f(\underline{u}ax) \right) = \sum_{a,\underline{u}} e^{\phi(ax)+\phi_n(\underline{u}ax)} f(\underline{u}ax)$$

Where $\underline{u}$ is a word of length n. Now, note that the summation above can be summarized in one summation over all possible words of length $n+1$. Then, we have:

$$\sum_w e^{\phi(\sigma^n(wx))+\phi_n(wx)} f(wx) = \sum_w e^{\phi_{n+1}(wx)} f(wx)$$

Hence:

$$\mathcal{L}^n f(x) = \sum_{\underline{u}} e^{\phi_n(\underline{u}x)} f(\underline{u}x)$$

It is easy to see that $f \geq 0 \implies \mathcal{L}f \geq 0$, whence $\mathcal{L}$ is a positive operator. Moreover, if $f \geq 0$, then $\mathcal{L}f = 0 \iff f = 0$.

Recall that, given an operator $\mathcal{L} : V \to W$, its dual $\mathcal{L}^* : W^* \to V^*$ is defined by $\mathcal{L}^*(T) = (T \circ \mathcal{L})$. Thus, given a positive functional $\mu \in \mathcal{C}(\Sigma)$, $\mathcal{L}^*(\mu)$ is also a positive functional. Indeed, given $f \geq 0$

em $\mathcal{C}(\Sigma)$, $\mathcal{L}^*(\mu)(f) = \mu(\mathcal{L}(f))$. But, since the operator is positive, $\mathcal{L}(f) \geq 0$, whence $\mu(\mathcal{L}(f)) \geq 0$, once μ is positive. By the Riesz-Markov theorem, μ is a finite Borel measure and, as we noted, $\mathcal{L}^*(\mu)$ is also a finite Borel measure, since it is a positive functional.

However, the operator is not defined in the set of probability borel measures. In fact, if μ is a borel probability measure, then:

$$(\mathcal{L}^*\mu)(1) = \mu\left(\sum_a e^{\phi(ax)}\right) = \sum_a \int e^{\phi(ax)} d\mu$$

where 1 denotes the constant function, and it is clear that $\mathcal{L}^*(\mu)$ may not be a probability measure, as the last expression may not be 1. Nevertheless, we can normalize. Check that the operator $(\mathcal{L}^*\mu(1))^{-1}\mathcal{L}^*(\mu) : (M)_1(\Sigma) \to (M)_1(\Sigma)$ is well defined, where $(M)_1(\Sigma)$ is the set of all borel probability measures in Σ). By simplicity, we are going go to denote this operator by G and $\mathcal{L}^*\mu(1)$ by λ.

Lemma 2.5. *There is an eigenmeasure for $\mathcal{L}^*$ with eigenvalue λ.*

Proof. By definition of G, we just need to prove that G has a fixed point. This will be done by means of the Tychonoff-Schauder theorem. We already know that $\mathcal{M}_1(\Sigma)$ is a compact convex set. It remains to prove that G is continuous. ■

This lemma is part of one of the most important theorems of this area, the so called Ruelle-Perron-Frobenius theorem. We will stated it below, but we will not prove it completely, although we will prove an statement that is used to prove it.

Theorem 2.6 (Ruelle-Perron-Frobenius)**.** *Let Σ be a topologically mixing finite Markov shift and ϕ locally Hölder continuous. Then, there exists $\lambda > 0$, $h \in \pm$ positive and $\mu \in \mathcal{M}_1(\Sigma)$ such that h and μ are respectively an eigenfunction and an eingenmeasure, both with eigenvalue λ. Also, it holds:*

$$\lim_{n\to\infty} ||\lambda^{-n}\mathcal{L}^n g - \mu(g)h|| + 0$$

for all g continuous

2.3.2 Some technical lemmas

We are going to see some results about the concepts introduced previously, which we are going to use a lot, mainly in the study about the Gurevich Pressure. For that, we should present some definitions: given a set A, $A \cap \alpha_0^{k-1}$ is the set of all cylinders $[w]$ contained in A such that w have length k; if $[w]$ and $[z]$ share respectively the last and first letter, $[w] \cdot [z]$ is the concatenation. For example, if $w = w_o ... w_{n-1}c$ and $z = cz_1..z_m$, $[w] \cdot [z] = [w_o ... w_{n-1}cz_1 ... z_m]$. Given sets of cylinders, $A \cdot B$, when it makes sense, is the union of $[a] \cdot [b]$. Yet, $B_n(\phi) = \exp \sum_{k>n} V_k(\phi)$ and the Birkhoff sum: $\phi_n = \sum_{k=0}^{n-1}(\phi \circ \sigma)^k$ and $\phi_n[w] = \sup\{\phi_n(x); x \in [w]\}$.

Lemma 2.7. *Let ϕ be a summable variation function. The following statements hold:*

1. $\forall n \leq m, V_m(\phi_n) \leq \log B_{m-n}$
2. $\forall n \leq m, |\phi_n(x) - \phi_n[x_o ... x_{m-1}]| \leq \log B_{m-n}$
3. *For some letter $c \in \mathbb{N}$, and n, if $A \subset (\sigma^{-n}([c])) \cap \alpha_0^n$ and $B \subset [c] \cap \alpha_0^m$, then:*

$$\frac{1}{B_1^2} \sum_{[z]\in(A\cdot B)\cap\alpha_o^{n+m}} \exp(\phi_{n+m}[z]) \leq \left(\sum_{[x]\in A\cap\alpha_0^n} \exp(\phi_n[x])\right)\left(\sum_{[y]\in B\cap\alpha_0^m} \exp(\phi_m[y])\right)$$

$$\leq B_1^2 \sum_{[z]\in(A\cdot B)\cap\alpha_o^{n+m}} \exp(\phi_{n+m}[z])$$

Proof. In first place, $B_n(\phi)$ is always well-defined for the function having summable variation. Now:

1. $V_m(\phi_n)$ is the supremum over the set of x, y that share the first m letters of:

$$\left|\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(x)-\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(y)\right|=\left|\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(x)-(\phi\circ\sigma^k)(y)\right|\leq\sum_{k=0}^{n-1}\left|(\phi\circ\sigma^k)(x)-(\phi\circ\sigma^k)(y)\right|$$

   But notice that, If x, y share the first m letters, $\sigma^k(x)$ and $\sigma^k(y)$ share the first $m-k$ letters. Then, $(\phi\circ\sigma^k)(x)-(\phi\circ\sigma^k)(y)\leq V_{m-k}$, therefore:

$$\left|\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(x)-\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(y)\right|\leq\sum_{k=0}^{n-1}V_{m-k}(\phi)=\sum_{k=m-n+1}^{m}V_k(\phi)\leq\sum_{k>m-n}V_k(\phi)$$

   Once this inequality holds for every suitable x, y, the supremum satisfies this too, whence $V_m(\phi_n)\leq\sum_{k=m-n+1}^{m}V_k(\phi)\leq\sum_{k>m-n}V_k(\phi)$

2. It is enough to prove that $|\phi_n(x)-\phi_n[x_o...x_{m-1}]|\leq V_m(\phi_n)$ and use the result of item 1. In fact, let X_m be the subset of Σ^2 such that $(x,y)\in X_m$ if, and only if, x, y have the same m first letters. Then, given x, $\phi_n[x_o...x_{m-1}]$ is just $\sup\phi_n(y)$, where the supremum is taken over $y\in\Sigma$ such that $(x,y)\in X_m$. But then we have that:

$$|\phi_n(x)-\phi_n[x_o...x_{m-1}]|=|\phi_n(x)-\sup\phi_n(y)|\leq|\sup\{\phi_n(x)-\phi_n(y)\}|\leq\sup|\phi_n(x)-\phi_n(y)|$$

   Where every supremum is taken over $y\in\Sigma$ such that $(x,y)\in X_m$. But note that $V_m(\phi_n)$ is almost the same as the last supremum, with the difference that it is taken over the the whole set X_m, that is, all its ordained pairs. Hence, $V_m(\phi_n)$ is greater and this item is done

3. Do not be afraid about the expressions under the summation and the set in which A and B are contained. All they are telling is that A and B are sets of cylinders such that the last letter of the word defining each cylinder in A is the same as the first letter of the words defining each cylinder of B, namely, c, so that the concatenation always make sense.

   Notice that A and B are countable, but there is no reason in principle to believe they converge.

   We will prove first that, given $[x]\in A$ and $[y]\in B$ (and therefore x has length $n+1$ and y length $m+1$), we have that:

$$|\phi_n([x])+\phi_m([y])-\phi_{n+m}([z])|\leq 2\sum_{k>1}V_k(\phi)$$

Actually, we are going to start proving that the expression is positive, whence the modulus is not necessary. Given $x=x_o...x_n$, $y=y_o...y_m$ with $x_n=y_o=c$, $b=y_m$ and $z=x_o...c...y_m$, notice that:

$$\phi_n[x]=\sup_{p\in[x]}\{\phi_n(p)\}=\sup_{r\in[c]}\{\phi_n(x_o^{n-1}r)\}=\sup_{r\in[c]}\left\{\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(x_o^{n-1}r)\right\}$$

Where x_i^j stands for the word formed by the letters $x_i...x_j$. And similarly:

$$\phi_m[y]=\sup_{q\in[b]}\left\{\sum_{k=0}^{m-1}(\phi\circ\sigma^k)(y_o^{m-1}q)\right\}$$

$$\phi_{n+m}[z] = \sup_{q\in[b]}\left\{\sum_{k=0}^{n+m-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q)\right\}$$

Thus:

$$\begin{aligned}\phi_{n+m}[z] &= \sup_{q\in[b]}\left\{\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q) + \sum_{k=n}^{n+m-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q)\right\}\\ &\le \sup_{q\in[b]}\left\{\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q)\right\} + \sup_{q\in[b]}\left\{\sum_{k=n}^{n+m-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q)\right\}\end{aligned}$$

It is easy to see that the second supremum is actually $\phi_m[y]$. Just notice that, by construction of z, $y_o = z_n$ and use the above expression for $\phi_m[y]$. On the other hand, the first supremum is less or equal $\phi_n[x]$. In fact, we could rewrite the above supremum:

$$\sup_{q\in[b]}\left\{\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(x_o^{n-1}y_o^{m-1}q)\right\}$$

And see that the supremum is taken over a subset of the set over which $\phi_n[x]$ is taken. Indeed, given $q\in[b]$, we can always find a $r\in[c]$ such that $r = y_o^{m-1q}$.

To prove the inequality itself, we are going to start by proving that:

$$\sup_{q\in[b]}\left\{\sum_{k=n}^{n+m-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q)\right\} + \sum_{k>1}V_k(\phi) \ge \sup_{r\in[c]}\left\{\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(x_o^{n-1}r)\right\} = \phi_n[x]$$

Take $q\in[b]$ and $r\in[c]$ arbitrarily. We know that $(z_o^{n+m-1}q)_o^n = (x_o^{n-1}r)_o^n$ because $z_o^{n-1} = x_o^{n-1}$ by construction and $z_n = r_o = c$. But, at each iteraction of σ in the sequences, the number of coincident letters decreases. If we take off k elements of the beginning of each sequence, we can only assure that the sequence is equal from 0 to $n-k$. But then, we have that:

$$(\phi\circ\sigma^k)(x_o^{n-1}r) - (\phi\circ\sigma^k)(z_o^{n+m-1}q) \le V_{n-1+k}(\phi)$$

Whence:

$$\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(x_o^{n-1}r) - \sum_{k=0}^{n-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q) \le \sum_{k=0}^{n-1}V_{n+1-k}(\phi) = \sum_{k=2}^{n+1}V_k(\phi)$$

Where the last equality is just a change of variable in the summation. This gives us:

$$\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q) + \sum_{k>1}V_k(\phi) \ge \sum_{k=0}^{n-1}(\phi\circ\sigma^k)(x_o^{n-1}r)$$

Once the equality holds for every $r\in[c]$, given $q\in[b]$:

$$\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q) + \sum_{k>1}V_k(\phi) \ge \sup_{r\in[c]}\left\{\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(x_o^{n-1}r)\right\}$$

Similarly, the equality holds for every $q\in[b]$, so it is obvious that:

$$\sup_{q\in[b]}\left\{\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q) + \sum_{k>1}V_k(\phi)\right\} = \sup_{q\in[b]}\left\{\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q)\right\} + \sum_{k>1}V_k(\phi)$$

$$\geq \sup_{r\in[c]}\left\{\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(x_o^{n-1}r)\right\}$$

Now, let's prove that:

$$\phi_{n+m}[z] = \sup_{q\in[b]}\left\{\sum_{k=0}^{n+m-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q)\right\} + \sum_{k>1}V_k(\phi)$$

$$\geq \sup_{q\in[b]}\left\{\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q)\right\} + \sup_{q\in[b]}\left\{\sum_{k=n}^{n+m-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q)\right\}$$

Given $q, r \in [b]$, as $\sigma^k(z_o^{n+m-1}r) = z_k^{n+m-1}r$ and $\sigma^k(z_o^{n+m-1}q) = z_k^{n+m-1}q$ share the first $n+m+1-k$ letters, we have:

$$\sum_{k=2}^{n+m+1}V_k(\phi) = \sum_{k=0}^{n+m-1}V_{n+m+1-k}(\phi) \geq \sum_{k=0}^{n+m-1}\sup_{q\in[b]}\left\{(\phi\circ\sigma^k)(z_o^{n+m-1}q) - (\phi\circ\sigma^k)(z_o^{n+m-1}q)\right\}$$

$$\Longrightarrow \sum_{k=0}^{n+m-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q) + \sum_{k>1}V_k(\phi) \geq \sum_{k=0}^{n+m-1}\sup_{q\in[b]}\left\{(\phi\circ\sigma^k)(z_o^{n+m-1}q)\right\}$$

$$\geq \sup_{q\in[b]}\left\{\sum_{k=0}^{n-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q)\right\} + \sup_{q\in[b]}\left\{\sum_{k=n}^{n+m-1}(\phi\circ\sigma^k)(z_o^{n+m-1}q)\right\}$$

As the inequality does not depend upon q, we can take the supremum of the first member, and we will be done.

Now, we can join all the work so far to obtain:

$$|\phi_n([x]) + \phi_m([y]) - \phi_{n+m}([z])| \leq 2\sum_{k>1}V_k(\phi)$$

$$\Longrightarrow -2\sum_{k>1}V_k(\phi) + \phi_{n+m}([z]) \leq \phi_n([x]) + \phi_m([y]) \leq 2\sum_{k>1}V_k(\phi) + \phi_{n+m}([z])$$

Taking the exponential:

$$\frac{1}{B_1^2(\phi)}\exp(\phi_{n+m}[z]) \leq \exp(\phi_n[x])\exp(\phi_m[y]) \leq B_1^2(\phi)\exp(\phi_{n+m}[z])$$

Now, notice that this inequality holds for every triplet $[x], [y], [z]$, $[x] \in A$, $[y] \in B$, $[z] \in A\cdot B$. Thus, we can sum up every possible triplets and we finally get the so desired inequality. ■

Proposition 2.8. *Let Σ be a finite topologically mixing shift, $f \in \mathcal{C}(\Sigma)$ be a positive function not identically zero such that $f(x) \leq B_m f(y)$ for every x, y that share their $m+1$ first letters. Then, $\mathcal{L}^n f$ is exponentially bounded from below.*

Proof. As the shift is finite and topologically mixing, there exists N such that, if $n > N$, there exists a word w with length $n-1$ such that z_owx is admissible, for every $x, y \in \Sigma$. Usually, N depends upon the letters, but since the shift is finite, we can take the maximum over all pairs of letters. By the definition of the Ruelle operator, we get $\mathcal{L}^n f(x) \geq \exp(\phi_n(z_owx))f(z_owx)$, because this is just one term of a summation of non-negative terms (recall that f is positive). But realize that $\phi_n(y) = \sum_{k=0}^{n-1}\phi(\sigma k(y)) \geq n(-||\phi||) = -n||\phi||$. Joining the results we get $\mathcal{L}^n f(x) \geq \exp(-n||\phi||)f(z_owx)$. But, by hypothesis, $B_0 f(z_owx) \geq f(z)$, whence $\mathcal{L}^n f(x) \geq \lambda^{-n}\exp(-n||\phi||)f(z_owx) \geq B_0^{-1}K^{-n}f(z)$, where $K = \exp(-||\phi||)$. Thus, for $n > N$, we have:

$$\mathcal{L}^n f(x) \geq B_0^{-1} K^{-n} f(z), \forall x$$

As $f(z) > 0$ for some z, let $A > 0$ be $f(z)$. Then, we get:

$$\mathcal{L}^n f(x) \geq A B_0^{-1} K^{-n}, \forall x$$

■

Lemma 2.9. *Let Σ be a topologically mixing countable Markov shift. Then, for every $n > 0$, there is a topologically mixing subshift $X \subset \Sigma$ with at least n letters*

Proof. In first place, we will show the existence of a topologically mixing subshift with at least two letters. With no loss of generality, take these letters to be 1 and 2.

Since Σ is originally topologically mixing, there are words of length k e $k+1$ connecting 1 to 2 and a word of length j connecting 2 to 1. Let X be the subshift composed by all the letters that form these words. Then, for any $n \geq N$, with $N = (j+k)(j+k-1)+k$, there is a word of length n in X connecting 1 to 2. In fact, notice that, by means of concatenation, we can take words of length $j+k$ and $j+k+1$ that connects 1 to itself. Thus, we can take $j+k$ times the cycle of $j+k-1$ letters (the $j+k$-th letter is 1 again) and then concatenate with the word with k letters, so that the whole cycle connects 1 to 2. Now, let's see that exists a word with length $N+i$, where $i < j+k$ playing the same role. For such, we will take a word similar to the former, with the difference that we will substitute i cycles of $j+k-1$ for cycles of $j+k$. Thereat, the length will be $(j+k-i)(j+k-1)+i(j+k)+k = (j+k)(j+k-1)+k+i = N+i$. More generally, let be $n \geq N$. Therefore, there are q, r such that $n-(j+k)(j+k-1)-k = q(j+k)+r$, with $r < j+k$ and $q \geq 0$. Take the word formed by $q+r$ cycles of $j+k$ letters and $j+k-r$ cycles of $j+k-1$ letters (and the word of k letters). This is an admissible word with length n connecting 1 to 2. Indeed, the length of the word is $q(j+k)+(j+k-r)(j+k-1)+r(j+k)+k = q(j+k)+(j+k)(j+k-1)+r(j+k)-r(j+k)+r+k = [q(j+k)+r]+(j+k)(j+k-1)+k = n$.

With that, it is proven that is always possible to get arbitrarily big words connecting 1 to 2. To show that X is topologically mixing, it remains to prove that it can be done with any other letters from X. Take, therefore, A e B arbitrarily. By construction, there are words w e z connecting A to 1 and 2 to B respectively. Suppose the length of these words are respectively a e b. Then, there is a word with length n connecting A to B for any $m > N+c+d-2$. In fact, it is enough to concatenate a word connecting 1 to 2 with length $n = m-c-d+2 > N$ with w e z.

Now, it just remains to show that we can find X arbitrarily big. For this, just notice that we can find a word connecting 1 to 2 with any subset of letters contained in this word. For example, suppose we want to include 3 and 4. Then, we can take a word connecting 1 to 3, concatenate with a word connecting 3 to 4 and finally concatenate with a word connecting 4 to 1. This is accomplished by the fact tat Σ is topologically mixing.

■

3 Generalizations of Differentiability

In this section we aim to extend the concept of differentiability from $\mathbb{R}^n$ to more general settings. For example, we will extend to real-valued functions defined on arbitrary vector spaces or normed ones. We will also state some very useful facts in the case that the function is convex, which is the case of interest, once the pressure is convex.

There are two relevant notions of differentiability: Gateaux and Fréchet differentibiality. We will restrict our attention to the cases where the codomain is $\mathbb{R}$, but the defitions can be naturally extended to general normed spaces

3.1 Gateaux Differentiability

Definition 3.1 (Gateaux Differenciability)**.** Let V be a vector space and $f : V \to \mathbb{R}$ a function. Then, we say the Gateaux derivative (or directional derivative) of f at $a \in V$ in the direction of $v \in V$ exists when the following limit exists:

$$\lim_{t\to 0} \frac{f(a+tv)-f(a)}{t}$$

In this case, let $f'(a,\cdot) : V \to \mathbb{R}$ be the function defined by the limit above. If $f'(a,.)$ exists for every $v \in V$ and is linear, then f is said to be Gateaux differentiable at a and f' is called the Gateaux differential of f at a. In some occasions, we may also require $f'(a,\cdot)$ to be bounded.

♣

As seen, $f'(a,.)$ is not always linear, although it is straightforward to show that it is always homogeneous. Some authors do not require this function to be linear and bounded to call f (Gateaux) differentiable, but we will always use the more restrictive definition. Indeed, it is really awkward to call a "differential" something that is not linear.

Note that the limit above is nothig besides the natural generalization of the concept of directional derivative. The next two sections will be devoted to prove very important preparatory results for some of the most important theorems in this text.

3.1.1 Banach Spaces

Lemma 3.1. *Let $U \subset \mathbb{R}^n$ be an open set and consider a Lipschitz map $f : U \to \mathbb{R}$. For a given direction v, the derivative of f in direction v exist almost everywhere. Furthermore, the set of point where it exists is a Borel set.*

Proof. Without loss of generality, assume that $U = \mathbb{R}^n$ and for now assume that the result is valid for $U = \mathbb{R}$ (this is a known result: every absolutely continuous function $f : \mathbb{R} \to \mathbb{R}$ is differentiable almost everywhere). We'll first show that the set A_u of all the points where the derivative in direction u exist is a Borel set. In fact, if it holds for $x \in U$, using the Cauchy criterion for functions, we get that for every $n \geq 1$ there exists $\delta > 0$ such that $\left|\frac{f(x+tu)-f(x)}{t} - \frac{f(x+su)-f(x)}{s}\right| \leq \frac{1}{n}$ whenever $0 < |t|, |s| < \delta$. By choosing $m > 0$ such that $\frac{1}{m} < \delta$ and re choosing u and t to be rationals between 0 and $\frac{1}{m}$, we get that $\left|\frac{f(x+tu)-f(x)}{t} - \frac{f(x+su)-f(x)}{s}\right| \leq \frac{1}{n}$ whenever t, s are non-zero rationals such that $0 < |s|, |t| < \frac{1}{m}$. With this, we see that:

$$A_u = \bigcap_{n=1}^{+\infty} \bigcup_{m=1}^{+\infty} \bigcap_{\substack{q_1 \in \mathbb{Q} \\ 0<|q_1|<\frac{1}{m}}} \bigcap_{\substack{q_2 \in \mathbb{Q} \\ 0<|q_2|<\frac{1}{m}}} Gf_u^{-1}(.,q_1,q_2)\left[0,\frac{1}{n}\right]$$

Where $Gf_u(.,q_1,q_2) : U \to \mathbb{R}$ is defined by $Gf_u(x,q_1,q_2) = \left|\frac{f(x+q_1u)-f(x)}{q_1} - \frac{f(x+q_2u)-f(x)}{q_2}\right|$ and we used the result that if the directional derivatives exist in a dense set, then they exist in every direction (to be proved in the next lemma). Since we decomposed A into a countable unions and intersections of closed sets, if follows that A is a Borel set. Now, we are allowed to talk about the measure of A.

For fixed $x \in U$ and $v \in \mathbb{R}^n$, define the function $f_{(x,v)} : \mathbb{R} \to \mathbb{R}$ by $f_{(x,v)}(t) = f(x+tv)$. Note that, if f is Lipschitz with Lipschitz constant $C > 0$, then $|f_{(x,v)}(t) - f_{(x,v)}(s)| = |f(x+tv) - f(x+sv)| \leq C|v||t-s| = M|t-s|$, where $M = C|v|$ is the new Lipschitz constant. Thus, $f_{(x,v)}$ is a Lipschitz function of one real variable, and since we are assuming the result holds for this case, we have that

$f_{(x,v)}$ is differentiable almost everywhere.

Now, fix $v \in \mathbb{R}^n$ and define the sets $B = \{(x,t) : f_{(x,v)}$ is not differentiable at $t \in \mathbb{R}\} \subset \mathbb{R}^{n+1}$, $B_x = \{t \in \mathbb{R} : (x,t) \in B\} \subset \mathbb{R}$ and $B^t = \{x \in \mathbb{R}^n : (x,t) \in B\} \subset \mathbb{R}^n$. By what we've just proved, we have $\lambda^1(B_x) = 0$, for all $x \in U$ and since B is measurable by the first part of the proof aplied to the case of $\mathbb{R}$, the characteristic function $\chi_B : \mathbb{R}^{n+1} \to \mathbb{R}$ is measurable, so by Tonelli's Theorem (for non-negative measurable functions) we conclude the following:

$$\lambda^{n+1}(B) = \int_{\mathbb{R}^{n+1}} \chi_B d\lambda^{n+1} = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}} \chi_B(x,t)dt \right) dx = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}} \chi_{B_x}(t)dt \right) dx$$
$$= \int_{\mathbb{R}^n} \lambda^1(B_x)dx = 0$$

So that:

$$0 = \lambda^{n+1}(B) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}^n} \chi_B(x,t)dx \right) dt = \int_{\mathbb{R}} \left(\int_{\mathbb{R}^n} \chi_{B^t}(x)dx \right) dt = \int_{\mathbb{R}} \lambda^n(B^t)dt$$

And in special $\lambda^n(B^t) = 0$ for almost all $t \in \mathbb{R}$. Thus, there must be some $t_0 \in \mathbb{R}$ such that $\lambda^n(B^{t_0}) = 0$. Now, note that:

$$B^{t_0} = \left\{ x \in \mathbb{R}^n : \nexists \lim_{t \to 0} \frac{f(x + (t+t_0)v) - f(x+tv)}{t} \right\} = \left\{ x \in \mathbb{R}^n : \nexists \lim_{t \to 0} \frac{f(\gamma(x) + tv) - f(\gamma(x))}{t} \right\}$$

Where $\gamma : \mathbb{R}^n \to \mathbb{R}^n$ is given by $\gamma(x) = x + t_0 v$. Finally, we have:

$$\left\{ x \in \mathbb{R}^n : \nexists \lim_{t \to 0} \frac{f(x+tv) - f(x)}{t} \right\} \subset \gamma(B^{t_0})$$

In fact, if x is in the first set, then as γ is surjective, there is $y \in \mathbb{R}^n$ such that $\gamma(y) = x$ and thus $\nexists \lim_{t \to 0} \frac{f(\gamma(y)+tv) - f(\gamma(y))}{t}$ so that the point (y, t_0) is in B^{t_0} as the equality up above shows. Since γ is an isometry, then $\lambda^n(\gamma(B^{t_0})) = \lambda^n(B^{t_0}) = 0$, so that:

$$\lambda^n \left(\left\{ x \in \mathbb{R}^n : \nexists \lim_{t \to 0} \frac{f(x+tv) - f(x)}{t} \right\} \right) \leq \lambda^n(\gamma(B^{t_0})) = 0$$

And this finishes the proof. ∎

An alternate proof:

Proof. Given $v \in \mathbb{R}^{n+1}$, let A denote the set of points such that the derivative of f in the direction of v does not exist. We want to prove that:

$$\mu(A) = \int_{\mathbb{R}^{n+1}} \chi_A d\mu = 0$$

We all know there is an isometry ψ that carries v to $(0, .., |v|)$. We will make a change of variables in the integral using this isometry. Let h and g_A be respectively the composition of the isometry with f and χ_A. Then, g_A becomes the characteristic function of the set of point such that h does not have the directional derivative. By Fubini, we have:

$$\mu(A) = \int_{\mathbb{R}^{n+1}} g_A(\mathbf{x}, y) d(\nu \otimes \lambda)(\mathbf{x}, y) = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}} g_A(\mathbf{x}, y) d\lambda(y) \right) d\nu(\mathbf{x})$$

But, for a given $\mathbf{x} \in \mathbb{R}^n$, it is easy to see that the directional derivative of h exists in $(\mathbf{x}, y)$ if, and only if, $y \mapsto h(\mathbf{x}, y)$ is differentiable in y. As an isometry is Lipschitz, the composition of Lipschitz functions is also Lipschitz and Lipschitz functions defined on intervals are differentiable almost everywhere, we get that $g_A(\mathbf{x}, y)$ is almost everywhere 0, for a given $\mathbf{x}$. Thus, $\mathbf{x} \mapsto \int g_A(\mathbf{x}, y) d\lambda(y)$ is always zero and we get $\mu(A) = 0$, as desired. ■

Proposition 3.2. *Let $f : E \to F$ be a Lipschitz function between two normed spaces such that the Gateaux derivative exists at x_o for every direction v. Then, $f'(x_o, \cdot)$ is continuous.*

Proof. Let $\epsilon > 0$ and $u \in E$. Take δ to be ϵ/K, where K is the Lipschitz constant of f. Then, for every v with $||u - v|| < \delta$:

$$||f'(x_o, u) - f'(x_o, v)|| = \left|\left|\lim_{t\to 0} \frac{f(x_o + tu) - f(x_o + tv)}{t}\right|\right| = \lim_{t\to 0} \frac{||f(x_o + tu) - f(x_o + tv)||}{|t|}$$

Since f is Lipschitz, $||f(x_o + tu) - f(x_o + tv)|| < K||(x_o + tu) - (x_o + tv)|| = K|t|.||u - v|| < |t|\epsilon$, thus:

$$||f'(x_o, u) - f'(x_o, v)|| = \lim_{t\to 0} \frac{|t|\epsilon}{|t|} = \epsilon$$

■

Lemma 3.3. *Let E and F be Banach spaces, and let f be a Lipschitz function from an open set in E into F. Let $G \subset E$ be a dense additive subgroup and assume that for some $x_o \in E$ and for all $u \in G$ the directional derivatives at x_o:*

$$\lim_{t\to 0} h_t(u) = \lim_{t\to 0} \frac{f(x_o + tu) - f(x_o)}{t}$$

exist. Then it exists for every direction. Furthermore, if they are additive as a function of u, then f is Gateaux differentiable at x_o.

Proof. Given $u, v \in E$:

$$||h_t(u) - h_t(v)|| = \frac{1}{|t|}||f(x_0 + tu) - f(x_o - tv)|| < \frac{K}{|t|}||t(u - v)|| = K||u - v||$$

Whence each h_t is lipschitz with the same Lipschitz constant as f. This easily implies that $\{h_t\}_t$ is (uniformly) equicontinuous. In fact, given $\epsilon > 0$, take $\delta = \epsilon/K$. Then, for every $u, v \in E$ and $t \in \mathbb{R}$, if $||u - v|| < \delta$, then

$$||h_t(u) - h_t(v)|| < K||u - v|| < \epsilon$$

Now, take $u \in E = \overline{G}$. We will show that $\lim_{t\to 0} h_t(u)$ exists. It suffices to prove that $h_{t_n}(u)$ is a Cauchy sequence for every sequence $t_n \to 0$. Given such a sequence and $\epsilon > 0$, as G is dense, there is $u_o \in G$ such that $||u - u_o|| < \epsilon/3K$, whence $||h_t(u) - h_t(u_o)|| < \epsilon/3$, for every t, by the equicontinuity. Moreover, by hypothesis, $\lim_{t\to 0} h_t(u_o)$ is a Cauchy sequence, so there is N such that $n, m > N \implies ||h_{t_n}(u_o) - h_{t_m}(u_o)|| < \epsilon/3$.

Then, for $n, m > N$, we have:

$$||h_{t_n}(u) - h_{t_m}(u)|| < ||h_{t_n}(u) - h_{t_n}(u_o)|| + ||h_{t_n}(u_o) - h_{t_m}(u_o)|| + ||h_{t_m}(u_o) - h_{t_m}(u)|| < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$$

Furthermore, by the density of G, given $u, v \in E$, we can find sequences $u_n \to u$ and $v_n \to v$. Then:

$$||f'(x_o, u + v) - (f'(x_o, u) + f'(x_o, v))||$$

$< ||f'(u+v) - f'(u_n+v_n)|| + ||f'(u_n+v_n) - f'(u_n) - f'(v_n))|| + ||f'(u_n) + f'(v_n) - f'(u) - f'(v)||$

Where the dependence on x_o was suppressed by having no risk of confusion. The second term vanishes as f' is additive on G. The other two vanishes by the continuity of f', since it is Lipschitz (see the previous section).

■

Lemma 3.4. *Let $f : U \subset \mathbb{R}^n \to \mathbb{R}$ be a Lipschitz function, with U open. Then f is Gateaux differentiable almost everywhere in U.*

Proof. Given one direction, we already know that the set of point where the directional derivative does not exists is null. However, once G is enumerable, we can join all the sets where the derivative fails to exist to one of the directions and we also get a null set.

Thus, the derivatives in the directions of the vectors of G exist almost everywhere. By the previous lemma is enough to show that these derivatives are additive in the points where they exist.

Firstly, let ϕ be a smooth function with compact support and integral 1. You may take, for example, $\phi(x) = C_n \exp(-1/1 - ||x||^2)$ for $||x|| < 1$ and $\phi(x)$ elsewhere, with C_n a suitable constant. Define $g : \mathbb{R}^n \to \mathbb{R}$ putting $g = f * \phi$, i.e. $g(x) = \int_{\mathbb{R}} f(y)\phi(x-y)dy$. Fixing every variable but one and using the fundamental theorem of calculus, we see there every partial derivative of g exists. Furthermore, as the integrand is continuous, g is actually C^1, and, therefore, differentiable. Thus, $D_g(x)$ is linear for every $x \in \mathbb{R}^n$.

On the other hand, we have:

$$D_g(x)u = \lim_{t\to 0} \frac{g(x+tu) - g(x)}{t} = \lim_{t\to 0} \int_{\mathbb{R}^n} \phi(y) \frac{f(x-y+tu) - f(x-y)}{t} dy$$

Our aim now is to use Lebesgue's dominated convergence theorem to put the limit into the integral. Let's check the hypothesis. We have the application $\mathfrak{F} : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ defined by (when $t \neq 0$):

$$\mathfrak{F}(y,t) = \phi(y) \frac{f(x-y+tu) - f(x-y)}{t}$$

When y is such that the directional derivatives exist and 0 elsewhere. Recall that the first case happens almost everywhere.

Moreover, for a fixed y, take $\mathfrak{F}(y,0)$ such that $\mathfrak{F}(y,\cdot)$ is continuous. This can always happen once ϕ is continuous and the limit of the fraction is the directional derivative almost everywhere and 0 elsewhere. It remains only to find an integrable function such that $|\mathfrak{F}(y,t)| \leq \mathfrak{G}(y)$, for every t. For those y such that $\mathfrak{F}(y,\cdot)$ is always zero, just define $\mathfrak{G}(y) = 0$. Now, suppose y is not one of those. Then, since ϕ is bounded and f is Lipschitz, we have:

$$|\mathfrak{F}(y,t)| \leq M_\phi \frac{K||tu||}{|t|} = M_\phi K||u||$$

Where M_ϕ bounds ϕ and K is the Lipschitz constant. Therefore:

$$D_g(x)u = \int_{\mathbb{R}^n} \phi(y) \lim_{t\to 0} \frac{f(x-y+tu) - f(x-y)}{t} dy = \int_{\mathbb{R}^n} \phi(y) h_u(x-y)dy = (\phi * h_u)(x)$$

Where h_u is any function almost everywhere equals to the derivative in the direction of u. Without loss of generality, h_u can be taken to be bounded. In fact, in the points where the derivative exists, it is bounded by $K||u||$, as said above. Consequently, by the linearity of the convolution and of the differential, $\phi * (h_u + h_v) = \phi * h_u + \phi * h_u = D_g(u) + D_g(v) = D_g(u+v) = \phi * h_{u+v}$, whence:

$$\phi * (h_{u+v} - h_u - h_v) = 0$$

The remaining work is only devoted to show that this implies $h_{u+v} - h_u - h_v = 0$. For such, let $\phi_r(x)$ be $r^{-n}\phi(rx)$. It is not difficulty to see that each one of these functions satisfy all important properties of ϕ, whence $\phi_r * (h_{u+v} - h_u - h_v) = 0$ for every r.

It is enough to prove that $\lim_{r\to 0} \phi_r * f = f$ almost everywhere, for every bounded measurable function f. By Lebesgue's differentiation theorem, since f is integrable, almost every point is a Lebesgue point. Therefore, we are proving that the limit is satisfied when x is a Lebesgue point.

In fact, directly by definition, we have:

$$\lim_{r\to 0^+} \frac{1}{\lambda(B[x,r])} \int_{B[x,r]} |f(y) - f(x)| d\lambda(y) = \lim_{r\to 0^+} \frac{1}{r^n} \int_{B[x,r]} |f(y) - f(x)| d\lambda(y) = 0$$

Once $\lambda(B[x,r])$ is proportional to r^n. Now, we are going to apply a change of variables. Let $\psi : \mathbb{R}^n \to \mathbb{R}^n$ given by $\psi(z) = x - rz$. This is clearly a $\mathcal{C}^1$ diffeomorphism, $A = B[0,1]$ is clearly a measurable set and $|f - f(x)|$ is clearly a measurable function. Given those facts, as $\psi(A) = B[x,r]$, we have:

$$\int_{B[x,r]} |f(y) - f(x)| d\lambda(y) = \int_{B[0,1]} |f(x - rz) - f(x)| r^n d\lambda(z)$$

Where r^n is the determinant of the jacobian of ψ. Thus, we have:

$$\lim_{r\to 0^+} \int_{B[0,1]} |f(x - rz) - f(x)| d\lambda(z) = 0$$

However, as ϕ is bounded, we have that:

$$\left| \int_{B[0,1]} [f(x - rz) - f(x)]\phi(z) d\lambda(z) \right| \leq M_\phi \int_{B[0,1]} |f(x - rz) - f(x)| d\lambda(z)$$

$$\implies \lim_{r\to 0^+} \left| \int_{B[0,1]} [f(x - rz) - f(x)]\phi(z) d\lambda(z) \right| = \left| \lim_{r\to 0^+} \int_{B[0,1]} [f(x - rz) - f(x)]\phi(z) d\lambda(z) \right| = 0$$

$$\implies \lim_{r\to 0^+} \int_{B[0,1]} [f(x - rz) - f(x)]\phi(z) d\lambda(z) = \lim_{r\to 0^+} \int [f(x - rz) - f(x)]\phi(z) d\lambda(z) = 0$$

Where we used that ϕ is zero outside $B[0,1]$. The previous equality gives us:

$$\lim_{r\to 0^+} \int f(x - rz)\phi(z) d\lambda(z) = \int f(x)\phi(z) d\lambda(z) = f(x)$$

Once ϕ has integral 1. Applying the same change of variable but in the reverse direction $(\psi^{-1}(y) = x - y/r)$

$$\int f(x - rz)\phi(z) d\lambda(z) = \int f(y) r^{-n} \phi\left(\frac{x - y}{r}\right) d\lambda(y) = f * \phi_r(x)$$

$$\implies f(x) = \lim_{r\to 0^+} (f * \phi_r)(x)$$

Now, we know that h_u and h_v is almost everywhere additive. Since G is countable, we can find an unique set such the additivity holds for every pair of vectors and such that the complement has measure zero, ending the proof. ■

3.1.2 Convex Functions

Now, we'll need a few definitions. Let the functions below be defined as:

$$f'_+(a,v) := \lim_{t\to 0^+} \frac{f(a+tv)-f(a)}{t}$$
$$f'_-(a,v) := \lim_{t\to 0^-} \frac{f(a+tv)-f(a)}{t}$$

It is not hard to see that $f'_-(a,v) = -f'_+(a,-v)$, so that $f'(a,v)$ exist if, and only if $f'_+(a,\pm v)$ exist and $-f'_+(a,-v) = f'_+(a,v)$ for every $v \in V$. Since we also have $f'_+(a,\lambda v) = \lambda f'_+(a,v)$ for every $v \in V$ and $\lambda \geq 0$, it also follows that $f(a,.)$ is Gateaux differentiable if, and only if $f'(a,.) \in V^*$, where V^* is the algebraic dual of V (i.e, the set of all linear functionals from V to $\mathbb{R}$).

It not entirely clear if these functions are actually well-defined. This property, which follows from the know results of calculus of one real variable, will be extremely important to prove results about Gateaux differentiability. Note that the lateral limits are, a priori, not enough to show that a function is Gateaux differentiable, since we need the additional property of linearity. All these initial problems are solved in the case of convex functions in the next proposition:

Proposition 3.5. *Let V be a normed vector space, $a \in V$ and let $f : V \to \mathbb{R}$ be convex. Then:*

1. *The function $f'_+(a,.) : V \to \mathbb{R}$ is well-defined and is sublinear (i.e, subadditive and positive homogeneous);*
2. *For every $v \in V$ the property $-f'_+(a,-v) \leq f'_+(a,v)$ holds;*
3. *The set:*

$$\tilde{V} = \{v \in V : f'(a,v) \text{ exists}\} = \{v \in V : -f'_+(a,-v) = f'_+(a,v)\}$$

Is a vector subspace of V and $f'|_{\tilde{V}}$ is linear (in special, $f'|_{\tilde{V}} \in V^*$);

4. *If f is continuous, then $f'_+(a,.)$ is Lipschitz continuous.*

Proof. Throughout this proof, we will consider the auxiliary function $\varphi : \mathbb{R} \to \mathbb{R}$ given by (for fixed $a, v \in V$) $\varphi(t) = f(a+tv)$. Then, as seen in any good enough calculus or analysis course, the lateral derivatives of φ exist everywhere, since it is convex. Thus, we have that $f'_+(a,v) = \varphi'_+(0)$ exists and is finite. To see why it is sublinear, just apply the definition and use the fact already stated that $f'_+(a,.)$ is homogeneous:

$$f'_+(a,v+s) = 2f'_+\left(a, \frac{v+s}{2}\right) = 2\lim_{t\to 0^+} \frac{f(a+t\frac{v+s}{2})-f(a)}{t} = 2\lim_{t\to 0^+} \frac{f(\frac{1}{2}(a+tv)+\frac{1}{2}(a+ts))-f(a)}{t}$$
$$\leq 2\lim_{t\to 0^+} \frac{\frac{1}{2}f(a+tv)+\frac{1}{2}f(a+sv)-f(a)}{t} = f'_+(a,v)+f'_+(a,s)$$

Where we used the fact that f is convex in the fourth passage. The second property is so easy we might as well do it in this paragraph: just observe that, since $f'_+(a,.)$ is homogeneous, we have $0 = f'_+(a,0) = f'_+(a,v-v) \leq f'_+(a,v) + f'_+(a,-v) \implies -f'_+(a,-v) \leq f'_+(a,v)$.

Now, let us do the third item. Let $\lambda \geq 0$ be any. We already have $\lambda v \in \tilde{V}$ for every $v \in \tilde{V}$, by positive homogeneity of $f'_+(a,.)$. If $\lambda < 0$ and $v \in \tilde{V}$, then $-\lambda > 0$ and by definition of $\tilde{V}$ we have $-f'_+(a,-\lambda v) = -(-\lambda)f'_+(a,v) = \lambda f'_+(a,v) = -\lambda f'_+(a,-v) = f'_+(a,\lambda v)$. This shows simultaneously that $\lambda v \in \tilde{V}$ and that $f'_+(a, \lambda v) = \lambda f'_+(a, v)$ for all $\lambda \in \mathbb{R}$ and $v \in \tilde{V}$, since we've checked every case. Now, note that, for $v, s \in \tilde{V}$:

$$f'_+(a, v+s) \leq f'_+(a, v) + f'_+(a, s) = -[f'_+(a, -v) + f'_+(a, -s)] \leq -f'_+(a, -v-s) \leq f'_+(a, v+s)$$

Where we used subadditivity and item (2) in the last passages. This shows that $-f'_+(a, -s-v) = f'_+(a, s+v)$, i.e, $s+v \in \tilde{V}$. This also shows that $f'_+(a, s) + f'_+(a, v) = f'_+(a, s+v)$, so that $f'_+(a, .)$ is linear.

As for the last item, since f is continuous, it is locally bounded at every point. We will now prove that f is locally Lipschitz. Without loss of generality, we will prove this result only for the origin, but the proof remains the same if this were not the case. Since the function is locally bounded, there is a $\epsilon > 0$ and a $M > 0$ such that $|f(x)| \leq M$, for all $x \in B(0, 2\epsilon)$. Take any $x_1, x_2 \in B(0, \epsilon)$ and consider the point $x_3 = x_2 + \frac{\epsilon}{\alpha}(x_2 - x_1)$, where $\alpha = |x_2 - x_1|$ (we are just translating x_2 by ϵ on the line connecting x_1 to x_2) and note that:

$$|x_3| \leq |x_2| + \epsilon \leq 2\epsilon$$

So that $x_3 \in B(0, 2\epsilon)$. By noticing that $x_2 = \frac{\alpha}{\epsilon+\alpha}x_3 + \frac{\epsilon}{\epsilon+\alpha}x_1$ and $\frac{\alpha}{\alpha+\epsilon} + \frac{\epsilon}{\alpha+\epsilon} = 1$, we can use the convexity of f to infer that:

$$f(x_2) \leq \frac{\alpha}{\epsilon+\alpha}f(x_3) + \frac{\epsilon}{\alpha+\epsilon}f(x_1) \implies f(x_2) - f(x_1) \leq \frac{\alpha}{\alpha+\epsilon}(f(x_3) - f(x_1))$$

$$\implies f(x_2) - f(x_1) \leq \frac{\alpha}{\epsilon}(f(x_3) - f(x_1)) \leq \frac{2M}{\epsilon}|x_2 - x_1|$$

By interchanging the role of x_1 and x_2, we get $|f(x_2) - f(x_1)| \leq \frac{2M}{\epsilon}|x_2 - x_1|$, so that f is locally Lipschitz. Using this fact, we can easily show that $|f'_+(a, v)| \leq L|v|$ everywhere, for some $L > 0$. In fact, take any neighborhood around a such that f is locally lipschitz in this neighborhood. As we just saw, we can take this to be an spherical neighborhood of radius say $\epsilon > 0$. Now, note that $a + tv \in B(a, \epsilon) \iff |tv| < \epsilon \iff |t| < \frac{\epsilon}{|v|}$, where we can suppose without loss of generality that $v \neq 0$. Then:

$$|f'_+(a, v)| = \lim_{\substack{t \to 0^+ \\ t < \frac{\epsilon}{|v|}}} \frac{|f(a+tv) - f(a)|}{t} \leq L \lim_{\substack{t \to 0^+ \\ t < \frac{\epsilon}{|v|}}} \frac{|tv|}{t} = L|v|$$

Where $t = |t|$ since $t > 0$ and we used the fact that the norm is continuous to pass the absolute value inside the limit. We can now use the subadditivity of $f'_+(a, .)$ to show it is Lipschitz everywhere as follows:

$$f'_+(a, u) - f'_+(a, v) = f'_+(a, v + (u - v)) - f'_+(a, v) \leq f'_+(a, u - v) \leq L|u - v|$$

And this proves the proposition.

■

As an easy corollary of item (3), a function is Gateaux differentiable at $a \in V \iff f'_+(a, .)$ exists at a.

3.2 Fréchet differentiability

Just like Gateaux differentiability extends the notion of directional derivability, Fréchet differentiability extends the notion of differentiability *de facto*.

Definition 3.2 (Fréchet Differentiability)**.** Let V be a normed vector space. A function $P : V \to \mathbb{R}$ is called Fréchet differentiable at ϕ if there is a linear transformation $T : V \to \mathbb{R}$, that is, a linear functional $T \in V^*$ such that:

$$\lim_{||\psi||\to 0} \frac{|P(\phi+\psi) - P(\phi) - T(\psi)|}{||\psi||} = 0$$

♣

Fréchet differentiability is stronger than Gateaux:

Proposition 3.6. *If P is Fréchet differentiable, then it is Gateaux differentiable and there is a differential associated with P. Namely, the linear transformation of the definition of Fréchet differentiability.*

Proof. This is literally the same as in the case of $\mathbb{R}^n$. Given a fixed ψ, take $t\psi$ in the definition of Fréchet differentiability. Then:

$$\lim_{t\to 0} \frac{|P(\phi+t\psi) - P(\phi) - T(t\psi)|}{t||\psi||} = 0 \implies \lim_{t\to 0} \frac{|P(\phi+t\psi) - P(\phi)|}{t} - T(\psi) = 0$$

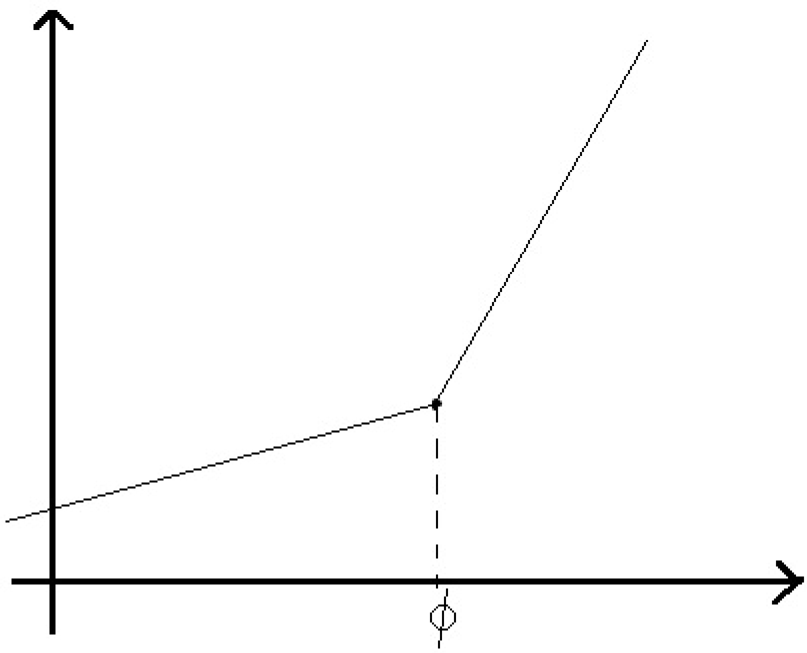
$$\lim_{t\to 0} \frac{|P(\phi+t\psi) - P(\phi)|}{t} = T(\psi)$$

Where we used the linearity of T. We see that $\alpha = T$. ■

It turns out that this definition is too strong for our purposes. [Wal92], corollary 9 states that, a topologically mixing subshift of finite type with finite alphabet, for example, is nowhere Frechet differentiable.

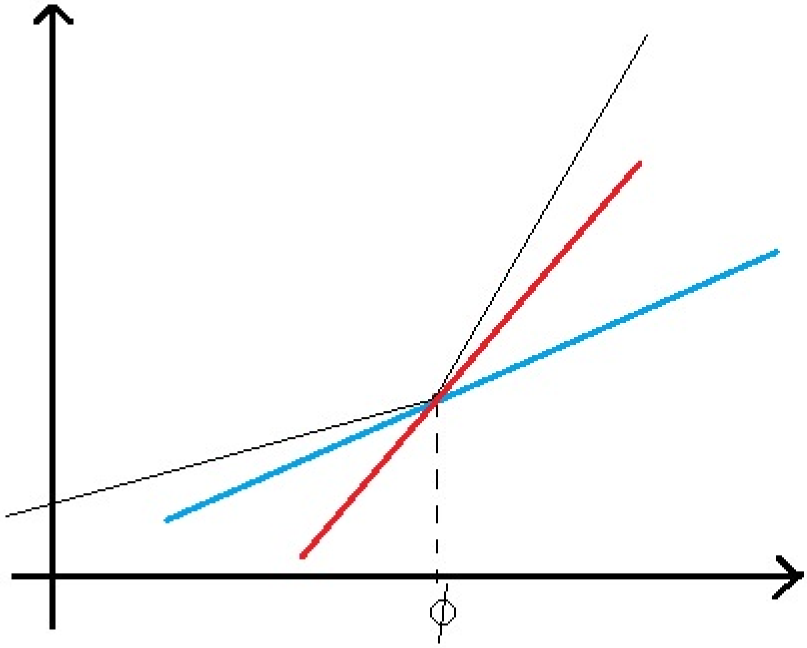
3.3 Subdifferential

The next concept is specific to convex functions, on which we will focus for a moment. In this setting, we can always define a generalization of derivative, even in points where the function is not differentiable. Due to didactic reasons, we will begin exposing the theory for functions defined in $\mathbb{R}$. Take a look at the following convex function:



Note that this function is not differentiable at ϕ. Recall that one can think of the derivative at a point as the slope of the tangent line at this point. Sometimes, a tangent line is defined to be the line that touches the graph at only one point. This definition is not fine for we can have functions that are locally constant. However, if we redefine a tangent line to a convex function to be a line that is contained in the region below the graph (including the graph), then it is true that there is at least one tangent line at each point of the graph of any real convex function. Furthermore, one can shows that the tangent line at a point is unique if, and only if, the function is differentiable at that point.

In the next figure, we draw two tangent lines at ϕ, showing both the existence and the non-uniqueness.



Thus, we can define the subderivative of a convex function at a point ϕ as a number c such that the line with slope c passing through ϕ is tangent to the graph, in the sense above. Equivalently, c is a subderivative if:

$$P(\psi + \phi) - P(\phi) \geq c\psi$$

This equivalence is easily seen, and the intuition for the formula above is that the graph of the function stays above the graph of one such tangent line. One can also show that, for convex functions in the real line, the set of subderivatives is a nonempty closed interval whose extremes are the lateral derivatives at the point (recall from real analysis that the lateral derivatives of such functions always exist). Moreover, the set is unitary if, and only if the point is differentiable.

For convex functions from $\mathbb{R}^2$ to $\mathbb{R}$, the situation is analogue. Every point ϕ has a tangent plane contained in the region below the graph and the plane is unique if, and only if, the function is differentiable at that point. In this case, the plane is parallel to the graph of the differential $dP(\phi)$. If P is not differentiable at ϕ, for each plane satisfying the cited property, we can find a functional whose graph is parallel to it. Those functionals are called subdifferentials.

Generally, for $P : \mathbb{R}^n \to \mathbb{R}$ convex, a subdifferential of P in ϕ is a linear functional $\alpha : \mathbb{R}^n \to \mathbb{R}$ such that the hyperplane passing through $(\phi, P(\phi))$ parallel to the graph of α is entirely contained below the graph of P. The condition for a linear functional be a subdifferential, algebrically, is:

$$P(\phi+\psi)-P(\phi)\geq\alpha(\psi),\forall\psi$$

Now, remembering that we can assign to each measure $\mu\in\mathcal{M}(T)$ a linear functional given by $\varphi\mapsto\int\varphi d\mu$, the next definition will be natural:

Definition 3.3 (Tangent Functional)**.** Let X be a compact metric space, $T:X\to X$ be continuous with finite topological entropy and consider $\varphi\in C(X)$. A finite signed measure μ is called a *tangent functional to P at φ* if, for every $\psi\in C(X)$ we following inequality holds:

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

The set of all tangent functionals to P at φ is denoted by $t_\varphi(T)$ or $\partial P(\varphi)$.

□

We shall now prove three important results about tangent functionals:

Proposition 3.7. *Let $\mu\in t_\varphi(T)$. Then μ is a non-negative $T-$invariant probability measure.*

Proof. For the first property, let $\psi\in C(X)$ be any continuous function satisfying $\psi\geq 0$ and take any $\epsilon>0$. Thus:

$$\int(\psi+\epsilon)d\mu=-\int-(\psi+\epsilon)d\mu\geq-P(\varphi-(\psi+\epsilon))+P(\varphi)\geq-P(\varphi-\inf_{x\in X}(\psi(x)+\epsilon))+P(\varphi)$$
$$=\inf_{x\in X}(\psi(x)+\epsilon)\geq 0$$

Thus, for any $\psi\geq 0$ in $C(X)$ we have $\int\psi d\mu\geq 0$. Now, one can check without difficulties that the class of Boreleans with non-negative measure is a monotonic class (in fact, it is enough to use the upper and lower continuity of μ) and we showed that this class contains all the open sets. By the monotonic class lemma, it also contains all the Boreleans, which prove our claim.

To see why it must be a probability measure, take any $n\in\mathbb{Z}$ and note that:

$$\int nd\mu\leq P(\varphi+n)-P(\varphi)=n\implies n\mu(X)\leq n$$

By taking $n=1$, we see that $\mu(X)\leq 1$. By taking $n=-1$, we see that $\mu(X)\geq 1$, so that $\mu(X)=1$.

Finally, to see why it is T-invariant, first note that $P(\varphi+\psi\circ T)=P(\varphi+\psi)$, since $P=\sup\left\{h_\mu+\int\varphi d\mu+\int\psi\circ Td\mu:\mu\in\mathcal{M}(T)\right\}=\sup\left\{h_\mu+\int\varphi d\mu+\int\psi d\mu:\mu\in\mathcal{M}(T)\right\}=P(\varphi+\psi)$. With this in mind, take any $n\in\mathbb{Z}$ and note that:

$$n\int(\psi\circ T-\psi)d\mu\leq P(n(\psi\circ T-\psi)+\varphi)-P(\varphi)=P(n(\psi-\psi)+\varphi)-P(\varphi)=0$$

By taking $n=1$, we get $\int(\psi\circ T-\psi)d\mu\leq 0$, and by taking $n=-1$ we get the other inequality. Thus, $\int\psi\circ Td\mu=\int\psi d\mu$, so that μ is $T-$invariant.

■

The next theorems will help us characterize even more the set of the tangent functionals:

Theorem 3.8 (Hahn-Banach)**.** *Let X be a vector space, $T:X\to X$ a map and let p be positive homogeneous subadditive function on X. If Y is a vector subspace of X such that $l:Y\to\mathbb{R}$ is a continuous linear functional such that $l(y)\leq p(y)$, for all $y\in Y$, then there exists a linear continuous extension of l, say, $\tilde{l}$ such that $\tilde{l}\leq p$ in all X.*

Proof. Check any good convex analysis book. ■

There is another version of this theorem, which we'll use later on:

Theorem 3.9 (Geometrical Hahn-Banach Theorem). *Let X be a normed vector space, f be convex, $H \subset X$ be an affine set such that a function $h : H \to \mathbb{R}$ is a continuous affine function satisfying $h \leq f$ on H. Assume that:*

$$int(dom(f)) \cap H \neq \varnothing$$

And f is continuous in $int(dom(f))$. Then there exists a continuous affine function $\tilde{h} : X \to \mathbb{R}$ extending h and such that $\tilde{h} \leq f$.

Proof. Check any good convex analysis book. ■

By affine set, we simply mean a set in which the whole line passing through any two points is also in the set. By affine function, we mean any function of the form $h(v) = a_0 + x^*(v)$, where $x^* \in X^*$.

We first have to generalize this result to allow the case where p is only convex. To do this, suppose that $p : X \to \mathbb{R}$ is a convex function and define the function $P(x) = \inf\limits_{t>0} \frac{p(tx)}{t}$, so that $P \leq p$ (take $t = 1$ in the definition of the infimum) and one can easily check that this function is subadditive and positive homogeneous, since p is convex. In fact, it is positive homogeneous trivially, and as for subadditivity, take any $s, v > 0$ and note that:

$$P(x+y) \leq \frac{s+v}{sv} p(\frac{sv}{s+v}(x+y)) = \frac{s+v}{sv} p(\frac{s}{s+v}(vx) + \frac{v}{s+v}(sy)) \leq \frac{1}{v} p(vx) + \frac{1}{s} p(sy)$$

Where in the first inequality we used $P \leq p$ and homogeneity of P (more specifically, we used $P(x+y) = P(\frac{s+v}{sv}\frac{sv}{s+v}(x+y)) = \frac{s+v}{sv} P(\frac{sv}{s+v}(x+y)))$ and in the third inequality we used the fact that p is convex. Since this is true for all $s > 0$, then $P(x+y)$ is a lower bound for the set $\{\frac{1}{v}p(vx) + \frac{1}{s}p(sy)\}$, thus $P(y) + \frac{1}{v}p(vx) \geq P(x+y)$. The same argument for v then results that $P(x+y) \leq P(x) + P(y)$, as we wanted.

Now, remembering that l is linear and $l \leq p$, we have $l(x) = \inf\limits_{t>0} \frac{l(tx)}{t} \leq P(x)$, so that there is an extension $\tilde{l}$ such that $\tilde{l} \leq P$ everywhere. Since $P \leq p$, this proves the generalization.

A few noteworthy facts about $t_\varphi(T)$ should be mentioned:

Proposition 3.10. *In the notation as above, the following remarks are true:*

1. *For every $\varphi \in C(X)$, we have $t_\varphi(T) \neq \varnothing$;*
2. *The set $t_\varphi(T)$ is compact and convex;*
3. *Every equilibrium measure μ for φ belongs to $t_\varphi(T)$.*

Proof. For the first claim, consider the affine set $H = \{\varphi\}$ together with the convex function $P : C(X) \to \mathbb{R}$. Define the continuous affine function $h : H \to \mathbb{R}$ defined by $h(\varphi) = P(\varphi)$. By the Geometrical Hahn-Banach Theorem, there is a continuous affine extension $\tilde{h} : C(X) \to \mathbb{R}$ satisfying $\tilde{h} \leq P$ everywhere. Since it is affine, there exists $x^* \in C(X)^*$ and $a_0 \in \mathbb{R}$ such that $\tilde{h}(\psi) = a_0 + x^*(\psi)$, and since $h(\varphi) = P(\varphi)$, this implies that $a_0 = P(\varphi) - x^*(\varphi)$, so that in fact we have $\tilde{h}(\psi) = P(\varphi) + x^*(\psi - \varphi)$. Using $\tilde{h} \leq P$, we get $P(\psi) - P(\varphi) \geq x^*(\psi - \varphi) \implies x^*(\psi) \leq P(\psi + \varphi) - P(\varphi)$. By Riesz Representation Theorem for signed measures, there is a finite Borel signed measure μ satisfying $x^*(\psi) = \int \psi d\mu$, so that:

$$\int \psi d\mu \leq P(\varphi + \psi) - P(\varphi)$$

For the second claim, note that:

$$t_\varphi(T) = \bigcap_{\psi \in C(X)} \left\{ \mu \in \mathcal{M}(T) : P(\varphi + \psi) - P(\varphi) \geq \int \psi d\mu \right\}$$

Since the intersection of convex sets is again convex, we only need to show that each of the sets in the intersection above in convex. Thus, fix $\psi \in C(X)$, $t \in [0,1]$ and measures $\mu_1, \mu_2 \in \mathcal{M}(T)$. By hypothesis, we have:

$$P(\varphi + \psi) - P(\varphi) \geq \int \psi d\mu_1$$

$$P(\varphi + \psi) - P(\varphi) \geq \int \psi d\mu_2$$

$$\implies tP(\varphi + \psi) - tP(\varphi) \geq \int \psi d(t\mu_1) \text{ and } (1-t)P(\varphi + \psi) - (1-t)P(\varphi) \geq \int \psi d((1-t)\mu_2)$$

Summing these two inequalities, we get:

$$P(\varphi + \psi) - P(\psi) \geq \int \psi d(t\mu_1 + (1-t)\mu_2)$$

So that $t\mu_1 + (1-t)\mu_2 \in t_\varphi(T)$. Now, we'll show it is compact when $C(X)^*$ is endowed with the Weak* Topology. Suppose first that t_φ is norm bounded and weak* - closed. Then, it is contained in a closed ball in the norm topology, which is weak* - compact by Banach Alaoglu's Theorem. Since the weak topology is Hausdorff, then every closed subset of a compact set is compact, so that $t_\varphi(T)$ is compact. Now, to see why $t_\varphi(T)$ is norm bounded, remember that the pressure is Lipschitz with Lipschitz constant of 1. Then, remembering that we can identify $t_\varphi(T)$ with the set of all linear continuous functionals $x^* \in C(X)^*$ such that $P(\varphi + \psi) - P(\varphi) \geq x^*(\psi)$, just note that $||x^*|| = \sup_{|\psi|=1} x^*(\psi) \leq \sup_{|\psi|=1} [P(\varphi + \psi) - P(\varphi)] \leq \sup_{|\psi|=1} |\psi| = 1$, so that $t_\varphi(T)$ is indeed contained inside the closed unit ball in the norm topology. Finally, to see why the set is closed, note that, for fixed $\varphi \in C(X)$, we have:

$$A(\psi) \coloneqq \{x^* \in C(X)^* : x^*(\psi) \leq P(\varphi + \psi) - P(\varphi)\} = w_\psi^{-1}(-\infty, P(\varphi + \psi) - P(\varphi)]$$

Where $w_\psi : C(X)^* \to \mathbb{R}$ is the evaluation functional defined by $w_\psi(x^*) = x^*(\psi)$, which is continuous by definition of the Weak* Topology. With this, we see that $A(\psi)$ is closed, so that $t_\varphi(T) = \bigcap_{\psi \in C(X)} A(\psi)$ is also closed. This finishes the proof that $t_\varphi(T)$ is compact and convex.

As for the last item, if μ is an equilibrium measure, then:

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

$$\implies P(\varphi + \psi) - P(\varphi) = h_\mu(T) + \int \varphi d\mu + \int \psi d\mu - h_\mu(T) - \int \varphi d\mu$$

$$= \int \psi d\mu$$

So that indeed $\mu \in t_\varphi(T)$. This completes the proof.

■

The next results will link subdifferentials to the concept of Gateaux differentiability, and we'll prove them with a bit more of generality, as we'll consider an arbitrary $f : C(X) \to \mathbb{R}$ which is supposed to be convex:

Proposition 3.11. *The next assertions are equivalent:*

1. $\mu \in t_\varphi(T)$;
2. $\int \psi d\mu \leq f'_+(\varphi, \psi)$;
3. $-f'_+(\varphi, -\psi) \leq \int \psi d\mu \leq f'_+(\varphi, \psi)$.

Proof. Assertions (2) and (3) are equivalent, since $-\int \psi d\mu = \int -\psi d\mu \leq f'_+(\varphi, -\psi) \implies -f'_+(\varphi, -\psi) \leq \int \psi d\mu$.

Now, suppose (1) is true. We then have for every $t > 0$:

$$\int \psi d\mu = \frac{1}{t} \int t\psi d\mu \leq \frac{f(\varphi + t\psi) - f(\varphi)}{t}$$

The limit as $t \to 0^+$ gives the result. As for the converse, first note that if we take the function identically equal to 1, we find out that the measure μ is finite. Now, we have:

$$\int (\psi - \varphi) d\mu \leq f'_+(\varphi, \psi + \varphi) = \lim_{t \to 0^+} \frac{f(\varphi + t(\psi - \varphi)) - f(\varphi)}{t} = \inf_{t \in (0,1]} \frac{f(\varphi + t(\psi - \varphi)) - f(\varphi)}{t}$$
$$\leq f(\psi) - f(\varphi)$$

And by taking $\psi \to \psi + \varphi$ we get the desired result.

■

It is interesting to note that, by the Riesz Representation Theorem for signed measures, we can identify every Borel signed finite measure with a continuous linear functional $x^* \in C(X)^*$ and vice-versa. Thus, what we are doing can be done in any other normed vector space.

Lemma 3.12. *For every $\psi \in C(X)$ we have:*

$$f'_+(\varphi, \psi) = \max_{\mu \in t_\varphi(T)} \int \psi d\mu$$

Proof. By the last proposition, we've already established that $f'_+(\varphi, \psi)$ is an upper bound for the set of the $\int \psi d\mu$ with $\mu \in t_\varphi(T)$ and fixed ψ. Thus, it is sufficient to prove that there exists a $\mu \in t_\varphi(T)$ such that $\int \psi d\mu = f'_+(\varphi, \psi)$.

For this, consider the affine subset $H = \varphi + t\psi$ of $C(X)$ and consider the function $h(\varphi + t\psi) = f(\varphi) + tf'_+(\varphi, \psi)$. Of course, we have $f(\varphi + t\psi) \geq f(\varphi) + tf'_+(\varphi, \psi)$, since for $t_0 > 0$, we have $f'_+(\varphi, \psi) = \inf\limits_{t>0} \frac{f(\varphi+t\psi)-f(\varphi)}{t} \leq \frac{f(\varphi+t_0\psi)-f(\varphi)}{t_0}$ and we get the result by rearranging the inequality. The case for $t_0 < 0$ is analogous, just remember that $f'_+(\varphi, \psi) \geq f'_-(\varphi, \psi)$, use that the lateral derivative from the left is the supremum of the fractions as before and remember to change the sign of the inequality when multiplying for $t_0 < 0$ on both sides. The equality is achieved when $t_0 = 0$. This implies that $h \leq f$ on H, so by Hahn-Banach Theorem there is a continuous affine extension $\tilde{h}$ on all $C(X)$ such that $\tilde{h} \leq f$. Since $\tilde{h}(\varphi) = h(\varphi) = f(\varphi)$, this function is of the form $\tilde{h}(\psi) = f(\varphi) + x^*(\varphi - \psi)$, where $x^* \in C(X)^*$. Then:

$$x^*(\psi) = \tilde{h}(\varphi + \psi) - f(\varphi) = h(\varphi + \psi) - f(\varphi) = f'_+(\varphi, \psi)$$

By the Riesz Representation Theorem, there is a finite signed measure μ such that $x^*(\psi) = \int \psi d\mu$, and this proves the Lemma.

$\blacksquare$

Corollary 3.12.1. *Let $f : C(X) \to \mathbb{R}$ be convex. Then f is Gateaux differentiable at φ if, and only if $t_\varphi(T)$ is a singleton. In this case, we have $t_\varphi(T) = \{f'(\varphi, .)\}$ (where this functional should be seen as a signed measure by Riesz).*

Proof. If f is Gateaux differentiable at φ, then $-f'_+(\varphi, -\psi) = f'_+(\varphi, \psi)$, so for any $\mu \in t_\varphi(T)$, item 3 of proposition 3.4 says that $\int \psi d\mu = f'_+(\varphi, \psi)$. By our last result, there is a signed measure ν such that $f'_+(\varphi, \psi) = \int \psi d\nu$, so that $\mu = \nu$. Thus, the set is a singleton with the only signed measure being ν.

Now, assume that f is not Gateaux differentiable at φ, so that there exists $\psi \in C(X)$ such that $f'_+(\varphi, \psi) \neq -f'_+(\varphi, -\psi)$. By the last Lemma, where are measures $\mu, \nu \in t_\varphi(T)$ such that $f'_+(\varphi, \psi) = \int \psi d\mu$ and $-f'_+(\varphi, -\psi) = \int \psi d\nu$. Thus:

$$\int \psi d\nu - \int \psi d\mu = f'_+(\varphi, \psi) + f'_+(\varphi, -\psi) \neq 0$$

So that $\nu \neq \mu$. Thus, $t_\varphi(T)$ is not a singleton.

$\blacksquare$

For the next result, we'll need the following lemmas:

Lemma 3.13 (Separation Theorem)**.** *If K_1, K_2 are disjoint closed convex subsets of a locally convex linear topological space V and if K_1 is compact, then there exists constants $c \in \mathbb{R}$, $\epsilon > 0$ and a continuous linear functional $f : V \to \mathbb{R}$ such that:*

$$f(K_2) \leq c - \epsilon < c \leq f(K_1)$$

$\square$

Proof.

$\blacksquare$

Theorem 3.14. *Let $f \in C(X)$ be any and define:*

$$A := \left\{ \mu \in \mathcal{M}(T) : \exists (\mu_n) \subset \mathcal{M}(T) \text{ with } \mu_n \to \mu \text{ and } h_{\mu_n}(T) + \int f d\mu_n \to P(f) \right\}$$

Then, $t_f(T) = A$.

Proof. The first inclusion is simple. Let $\mu_0 \in A$, so that there exists a sequence of measures $(\mu_n) \subset \mathcal{M}(T)$ satisfying $\mu_n \to \mu$ and $h_{\mu_n} + \int f d\mu_n \to P(f)$. Thus, remembering that $P(f+g) = \sup \left\{ h_\mu + \int (f+g) d\mu : \mu \in \mathcal{M}(T) \right\} \geq h_{\mu_n} + \int (f+g) d\mu_n$, we have:

$$P(f+g) - P(f) \geq h_{\mu_n} + \int (f+g) d\mu_n - P(f) \to \int g d\mu_0$$

So that $A \subset t_f(T)$. As for the second inclusion, suppose that there exists $\mu_0 \in t_f(T) \setminus A$. Take $K_2 = A$ and suppose for now that it is convex, and take $K_1 = \{\mu_0\}$, which is compact and convex trivially and note by assumption that these sets are disjoint. The separation theorem implies that there exists a continuous linear functional $h : C(X)^* \to \mathbb{R}$ such that $h(\mu) \leq c - \epsilon < c \leq h(\mu_0)$, for all $\mu \in A$. In special, we have $\sup_{\mu \in A} h(\mu) \leq c - \epsilon < c \leq h(\mu_0)$, so that $h(\mu_0) > \sup \{h(\mu) : \mu \in A\}$.

Now, since X is a compact metric space, the set of all finite, real valued signed Borel measures has cardinality $2^{\aleph_0}$, so that Riesz Representation Theorem holds for the bidual $C(X)^{**}$, so there is $g \in C(X)$ such that $h(\mu) = \int g d\mu$. Thus, we achieved the inequality:

$$\int g d\mu > \sup\left\{\int g d\mu : \mu \in A\right\}$$

We will now show the opposite inequality to finish the contradiction. For this, by the variational principle for each $n \geq 1$ we can choose a measure $\mu_n \in \mathcal{M}(T)$ such that:

$$h_{\mu_n}(T) + \int \left(f + \frac{g}{n}\right) d\mu_n > P\left(f + \frac{g}{n}\right) - \frac{1}{n^2}$$

For these measures, we have $h_{\mu_n}(T) + \int f d\mu_n \leq \sup\left\{h_\nu(T) + \int f d\nu : \nu \in \mathcal{M}(T)\right\} = P(f) \implies -P(f) \leq -h_{\mu_n} - \int f d\mu_n$. Then:

$$\int g d\mu_0 = n \int \frac{g}{n} d\mu_0 \leq n\left[P\left(f + \frac{g}{n}\right) - P(f)\right]$$

$$\leq n\left[h_{\mu_n} + \int\left(f + \frac{g}{n}\right) + \frac{1}{n^2} - h_{\mu_n} - \int f d\mu_n\right] = \int g d\mu_n + \frac{1}{n}$$

Now, since $M_1(X)$ is weak* - compact, and passing through a subsequence if necessary, we can assume that there is a limit point in which (μ_n) converges weakly, say, to μ^*. In this way, we have $\int g d\mu_0 \leq \int g d\mu^*$. Next, we will show that $\mu^* \in A$. In fact, note that $P(f + \frac{g}{n}) \geq P(f - \frac{||g||}{n}) = P(f) - \frac{||g||}{n}$, since $|g| \leq ||g||$, so that we have $g \geq -||g||$. With this in mind, note that:

$$h_{\mu_n} + \int g d\mu_n > P\left(f + \frac{g}{n}\right) - \int \frac{g}{n} d\mu_n - \frac{1}{n^2} \geq P(f) - 2\frac{||g||}{n} - \frac{1}{n^2} \to P(f)$$

So that in fact $\mu^* \in A$. Thus, we have $\int g d\mu_0 \leq \int g d\mu^* \leq \sup\left\{\int g d\mu : \mu \in A\right\}$, contradicting the other inequality.

Thus, to fully complete the proof, we only need to show that A is convex. To see why this is true, take any $\mu_1, \mu_2 \in A$, so that there exists sequences (μ_n^1) and (μ_n^2) both in $\mathcal{M}(T)$ with $\mu_n^1 \to \mu_1$, $\mu_n^2 \to \mu_2$ and $h_{\mu_n^1} + \int f d\mu_n^1 \to P(f)$ and $h_{\mu_n^2} + \int f d\mu_n^2 \to P(f)$. Note that $(t\mu_n^1 + (1-t)\mu_n^2) \to t\mu_1 + (1-t)\mu_2$. Now, by Theorem 8.1 page 184 [Wal82], we have $h_{t\mu_n^1 + (1-t)\mu_n^2} = th_{\mu_n^1} + (1-t)h_{\mu_n^2}$ for $t \in [0,1]$, so that:

$$h_{t\mu_n^1+(1-t)\mu_n^2} + \int f d(t\mu_n^1 + (1-t)\mu_n^2) = t\left(h_{\mu_n^1} + \int f d\mu_n^1\right) + (1-t)\left(h_{\mu_n^2} + \int f d\mu_n^2\right)$$

$$\to tP(f) + (1-t)P(f) = P(f)$$

So that A is convex.

■

Corollary 3.14.1. *$P : C(X) \to \mathbb{R}$ has a unique tangent functional at $f \in C(X)$ if, and only if there is a unique measure μ_f with the property that whenever $(\mu_n) \subset \mathcal{M}(T)$ is convergent and satisfies $h_{\mu_n} + \int f d\mu_n \to P(f)$, then $\mu_n \to \mu_f$. If this is the case, then the unique tangent functional is μ_f.*

Proof. Assume that $t_f(T) = \{\mu_f\}$ and take any sequence $(\mu_n) \subset \mathcal{M}(T)$ satisfying $h_{\mu_n} + \int f d\mu_n \to P(f)$ and we must prove that $\mu_n \to \mu_f$. Since $A = \{\mu_f\}$ in this case, and the limit $\mu_n \to \mu$ exists, it must be $\mu = \mu_f$, which completes this part of the proof.

As for the converse, take any to $\mu_1, \mu_2 \in t_f(T)$. Then, in the notation of the last theorem, we have $\mu_1, \mu_2 \in A$. Then, there exists sequences (μ_n^1), (μ_n^2) satisfying $h_{\mu_n^i} + \int f d\mu_n^i \to P(f)$ and $(\mu_n^i) \to \mu_i$, for $i \in \{1, 2\}$. By assumption, we have $\mu_n^1 \to \mu_f$ and $\mu_n^2 \to \mu_f$ and also $\mu_n^1 \to \mu_1$ and $\mu_n^2 \to \mu_2$. Since the Weak* Topology is Hausdorff, the limit is unique, and so $\mu_1 = \mu_f$ and $\mu_2 = \mu_f$, so that $\mu_1 = \mu_2$ and thus $t_f(T) = \{\mu_f\}$.

■

As a Corollary, we can get a characterization of when a tangent functional is not an equilibrium measure. First, we'll need a few results about the entropy map, which we'll describe now:

Proposition 3.15. *Let X be a compact metrisable space with $h(T) < +\infty$ and consider a continuous map $T : X \to X$ and a measure $\mu_0 \in \mathcal{M}(T)$. Then $h_{\mu_0}(T) = \inf\left\{P(f) - \int f d\mu_0 : f \in C(X)\right\}$ if, and only if the entropy map $\mu \mapsto h_\mu$ is upper semi-continuous at μ_0.*

Proof. Suppose first that the entropy map is upper semicontinuous at μ_0. By the variational principle given $f \in C(X)$, we have $P(f) \geq h_{\mu_0}(T) + \int f d\mu_0 \implies h_{\mu_0}(T) \leq P(f) - \int f d\mu_0$. Then, $h_{\mu_0}(T)$ is a lower bound for the set $\left\{P(f) - \int f d\mu_0 : f \in C(X)\right\}$, so that $h_{\mu_0}(T) \leq \inf\left\{P(f) - \int f d\mu_0 : f \in C(X)\right\}$.

To see why we can't have the strict inequality, suppose that $h_{\mu_0}(T) < \inf\left\{P(f) - \int f d\mu_0 : f \in C(X)\right\}$ so that there exists a $b > 0$ such that $h_{\mu_0} < b < \inf\left\{P(f) - \int f d\mu_0 : f \in C(X)\right\}$. Define the set $C = \{(\mu, t) \in \mathcal{M}(T) \times \mathbb{R} : 0 \leq t \leq h_\mu(T)\}$. Since the entropy map in concave (and convex), we have that C is convex (since it is the subgraph of a concave function). By using the upper semicontinuity of the entropy map, we can show that $(\mu_0, b) \notin \overline{C}$. In fact, first note that $(\mu_0, b) \notin C$ and that $U = \{\mu \in \mathcal{M}(T) : h_\mu(T) < b\}$ is open (as a matter of fact, given $\mu_1 \in U$, take $0 < \epsilon < b - h_{\mu_1}(T)$ so that there exists an open neighborhood V of $\mathcal{M}(T)$ such that $\mu_1 \in V$ and $h_\mu(T) \leq h_{\mu_1}(T) + \epsilon < b$, so that $V \subset U$) and $\mu_0 \in U$. Then, the set $U \times (\frac{b + h_{\mu_0}(T)}{2}, +\infty)$ is an open neighborhood of (μ_0, b) in $\mathcal{M} \times \mathbb{R}$, i.e, $(\mu_0, b) \notin \overline{C}$. Note actually that we've shown that $C = \overline{C}$, which is a closed convex set. Since $\{(\mu_0, b)\}$ is compact, the separation Lemma (we are now taking the space of all finite Borel measures $B(X)$) applies and we get a continuous linear functional $F : B(X) \times \mathbb{R} \to \mathbb{R}$ such that $F(\mu, t) < F(\mu_0, b)$ for all $(\mu, t) \in \overline{C}$. Now, define $X = \{(\mu, 0) : \mu \in B(X)\}$ and $Y = \{(0, t) : t \in \mathbb{R}\}$ where 0 in Y is the zero measure. Note that $F|_X(\mu) := F(\mu, 0)$ and $F|_Y(t) := F(0, t)$ are continuous and linear, so that by Riesz Representation Theorem for the bidual $C(X)^{**}$ there is a continuous function $f \in C(X)$ such that $F|_X(\mu) = \int f d\mu$ and simmilary $F|_Y(t) = td$, for some $d \in \mathbb{R}$. Then, we have $F(\mu, t) = F(\mu, 0) + F(0, t) = \int f d\mu + td$.

Applying the condition $F(\mu, t) < F(\mu_0, b)$, we get that $\int f d\mu + td < \int f d\mu_0 + db$, for all $(\mu, t) \in \overline{C}$. By taking $(\mu, t) = (\mu_0, h_{\mu_0}(T)) \in C$ and $\mu = \mu_0$, we end up with $h_{\mu_0} d < db$. Since $h_{\mu_0} < b$, the only option is $d > 0$. Then, taking $t = h_\mu(T)$, for all $\mu \in \mathcal{M}(T)$ we have:

$$\int \frac{f}{d} d\mu + h_\mu < \int \frac{f}{d} d\mu_0 + b \implies P\left(\frac{f}{d}\right) \leq b + \int \frac{f}{d} d\mu_0$$

$$\implies b \geq P\left(\frac{f}{d}\right) - \int \frac{f}{d} d\mu_0 \geq \inf\left\{P(g) - \int g d\mu_0 : g \in C(X)\right\}$$

Contradicting the fact that $b < \inf\left\{P(g) - \int g d\mu_0 : g \in C(X)\right\}$. Then, this part of the proof is proved.

Now, for the easy part. Remember that the sets $V_\mu(g, \epsilon) = \left\{\mu \in \mathcal{M}(T) : \left|\int g d\mu - \int g d\mu_0\right| < \epsilon\right\}$ are a base for the Weak* Topology, see [Wal82]. Now, if we have $h_{\mu_0}(T) = \inf\left\{P(f) - \int f d\mu_0 : f \in C(X)\right\}$, then for any $\epsilon > 0$, there is a continuous function $g \in C(X)$ such that $h_{\mu_0}(T) + \epsilon > P(g) - \int g d\mu_0$. Take the neighborhood $V_{\mu_0}(g, \epsilon)$, so that for all μ in this set we have:

$$h_\mu(T) \le P(g) - \int g d\mu < P(g) - \int g d\mu_0 + \epsilon < h_{\mu_0}(T) + 2\epsilon$$

Where in the first inequality we used the Variational Principle. This finishes the proof.

■

Corollary 3.15.1. *In the same hypothesis as before, let the entropy map $\mu \mapsto h_\mu(T)$ be upper semicontinuous at a measure $\mu_0 \in \mathcal{M}(T)$ such that $\mu_0 \in t_f(T)$. Then μ is an equilibrium measure at $f \in C(X)$.*

Proof. Since $\mu \in t_f(T)$, for all $g \in C(X)$ we have $P(f+g) - \int (f+g)d\mu \ge P(f) - \int f d\mu$, so for all $h \in C(X)$ we have $P(h) - \int h d\mu \ge P(f) - \int f d\mu$. Thus, $P(f) - \int f d\mu$ is a lower bound for the set $\left\{P(h) - \int h d\mu : g \in C(X)\right\}$. The last result then says that $h_\mu \ge P(f) - \int f d\mu$. By the Variational Principle, we also have $P(f) \ge h_\mu + \int d\mu \implies h_\mu \le P(f) - \int f \mu$, so that $P(f) = h_\mu + \int f \mu$, i.e, μ is an equilibrium measure for $f \in C(X)$.

■

For CMS of finite entropy, the entropy map is upper semi-continuous at ergodic measures, as proved in [IT18].

Corollary 3.15.2. *Let $f \in C(X)$, $\mu \in \mathcal{M}(T)$ and denote by $M_f(T)$ the set of all equilibrium measures at f. Then, the following are equivalent:*

1. $\mu \in t_f(T) \setminus M_f(T)$;
2. $h_\mu + \int f d\mu < P(f)$ *and there exists* $(\mu)_n \subset \mathcal{M}(T)$ *satisfying* $\mu_n \to \mu$ *and* $h_{\mu_n} + \int f d\mu_n \to P(f)$;
3. $\mu \in t_f(T)$ *and the entropy map* $\mu \to h_\mu$ *is not upper semicontinuous at* μ.

Proof. Suppose (i) first. Observe first that $h_\mu + \int f d\mu \le P(f)$ by the Variational Principle. Since μ it is not an equilibrium measure, the equality before cannot be achieved, so that $h_\mu + \int f d\mu < P(f)$. The second affirmation follows from the observation that μ is an element of A, since it is an element of $t_f(T)$, so that (ii) holds.

Now, assume (ii) holds. Remember that if the entropy map it upper semicontinuous at μ and μ is a tangent functional, then μ is an equilibrium measure, so that it would hold that $h_\mu + \int f d\mu = P(f)$, which contradicts the first hypothesis, so we get that the entropy map is not upper semicontinuous at μ. The fact that $\mu \in t_f(T)$ follows from the observation $\mu \in A$, so (iii) holds.

Finally, assume that (iii) holds. Remember that a function f is not upper semicontinuous at x_0 if there exists an $\epsilon > 0$ such for all neighborhoods U containing x_0, we can find a point $x' \in U$, such that $f(x') > f(x_0) + \epsilon$. In our case, we can consider $\mathcal{M}(T)$ with a metric that induces the weak topology and which metric convergence is equivalent to weak convergence. For each $n \ge 1$ we can choose the ball $B(\mu, \frac{1}{n})$ and a measure $\mu_n \in B(\mu, \frac{1}{n})$ such that $h_{\mu_n} > h_\mu + \epsilon$. Thus, we have $d(\mu, \mu_n) < \frac{1}{n}$ (i.e $\mu_n \to \mu$) and (since $h_{\mu_n} \le P(0)$, the sequence has a convergent subsequence, which we can reorder if necessary to pick a subsequence μ_{n_k} which converges weakly to μ, since the whole sequence converges) $\lim h_{\mu_n} \ge h_\mu + \epsilon > h_\mu$. Then:

$$P(f) \ge \lim(h_{\mu_n} + \int f d\mu_n) > h_\mu + \int d\mu$$

So that $\mu \notin M_f(T)$. This finishes the proof.

■

This result says that a tangent functional μ is not an equilibrium measure precisely when the entropy map $\mu \mapsto h_\mu$ is not upper semicontinuous.

4 Differentiability in the Compact Setting

4.1 Differentiability from a Topological Perspective

We will now prove that the pressure is Gateaux differentiable in a big set in the Topological sense. In fact, this a consequence of a much more general Theorem, which we'll prove now:

Theorem 4.1 (Mazur). *Let E be a separable Banach space and suppose $f : D \to \mathbb{R}$ is a convex continuous function defined on an open convex subset $D \subset E$. Then the set of all points where f is Gateaux differentiable is a dense G_δ set.*

Proof. We will first prove that the set of the points where f is not Gateaux differentiable is a countable union of closed sets. First, since E is separable we can choose a sequence (x_n) in the unit ball of E which is dense in the unit ball. For every $m, n \geq 1$, define:

$$\mathcal{A}_{m,n} := \left\{ x \in D : \exists x^*, y^* \in \partial f(x) \text{ such that } (x^* - y^*)(x_n) \geq \frac{1}{m} \right\}$$

First, remember that f it not Gateaux differentiable at $x \in D$ if, and only if there are distinct $x^*, y^* \in \partial f(x)$. Now, if f is not Gateaux differentiable at x, then there are $x^*, y^* \in \partial f(x)$ such that $x^* \neq y^*$. Suppose then that for all $n \geq 1$ we have $x^*(x_n) = y^*(x_n)$. Since each x^* is continuous in the sup topology, this would imply that $x^* = y^*$ in the unit ball, so that if $y \in E$ then $x^*(y) = x^*\left(\frac{y}{|y|}\right)|y| = y^*\left(\frac{y}{|y|}\right)|y| = y^*(y)$, so $x^* = y^*$, a contradiction. Thus, there exists $n \geq 1$ such that $x^* \neq y^*$, so we can suppose without loss of generality that for this n we have $(x^* - y^*)(x_n) > 0$. There is also $m \geq 1$ such that $(x^* - y^*)(x_n) > \frac{1}{m}$, so that $x \in \mathcal{A}_{m,n}$. Also, given $x \in \mathcal{A}_{m,n}$, there exists x^* and y^* both in $\partial f(x)$ such that $(x^* - y^*)(x_n) \geq \frac{1}{m} > 0$. In special, $x^* \neq y^*$, so that f is not differentiable at x. Thus, we have just proven that $\bigcup_{m=1}^{+\infty} \bigcup_{n=1}^{+\infty} \mathcal{A}_{m,n}$ equals the set of all $x \in D$ such that f is not differentiable.

The next step is to show that each $\mathcal{A}_{m,n}$ is a closed set. For this, take a sequence (z_n) in $\mathcal{A}_{m,n}$ such that $z_n \to z$ and we must show that $z \in \mathcal{A}_{m,n}$. For each $k \geq 1$, we have $z_k \in \mathcal{A}_{m,n}$ so that there are x_k^* and y_k^* both in $\partial f(z_k)$ such that $(x_k^* - y_k^*)(x_n) \geq \frac{1}{m}$. Now, since f is continuous, there is $\epsilon > 0$ and $M > 0$ such that $|f(x)| \leq M$, for all $x \in B(z, \epsilon)$. Now, note that $||x_k^*|| = \frac{2}{\epsilon} \sup_{|x|=\frac{\epsilon}{2}} |x_k^*(x)| \leq \frac{2}{\epsilon} \sup_{|x|=\frac{\epsilon}{2}} |f(z_k + x) - f(z_k)|$. Now, for big enough k, we have $z_k \in B(z, \frac{\epsilon}{2})$ by definition of convergence in metric spaces, so that $|z_k + x - z| \leq |z_k - z| + |x| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$ if $|x| = \frac{\epsilon}{2}$, so that $z_k + x \in B(z, \epsilon)$ for big enough k. Finally, this implies that for big enough k we have $||x_k^*|| \leq \frac{4M}{\epsilon}$. Now, since E is separable, it is a known result that the unit ball in E^* is metrizable in the Weak* Topology, and this result extends to arbitrary balls around the origin. By Banach-Alaoglu's Theorem, these balls are weakly compact. Since the ball $B(0, \frac{4M}{\epsilon})$ is weakly compact and metrizable in the Weak Topology, it is sequentially compact, so that we can extract a converging subsequence of x_k^*. By simplicity, we will assume from now on that the whole sequence converges, without loss of generality. The same process for y_k^* says then that there exists $x^*, y^* \in E^*$ such that $x_k^* \to x^*$ and $y_k^* \to y^*$ weakly. These are our candidates to show that $z \in \mathcal{A}_{m,n}$.

To proceed, remember that in the Weak* Topology if $x_n \to x$ in norm and $f_n \to f$ weakly, then $f_n(x_n) \to f(x)$. Thus, using again the fact that $x_k^* \in \partial f(z_k)$, we have $x_k^*(y - z_k) \leq f(y) - f(z_k)$ for all $y \in E$, so that:

$$x^*(y - z) = \lim_{k \to +\infty} x_k^*(y - z_k) \leq \lim_{k \to +\infty} (f(y) - f(z_k)) = f(y) - f(z)$$

Where we used the continuity of f in the last step. Thus, $x^* \in \partial f(z)$ and the same holds for y^*. Also, since for every such $k \geq 1$ we also have $(x_k^* - y_k^*)(x_n) \geq \frac{1}{m}$, we get $(x^* - y^*)(x_n) =$

$\lim\limits_{k\to+\infty}(x_k^* - y_k^*)(x_n) \geq \frac{1}{m}$, so that in fact $z \in \mathcal{A}_{m,n}$. This finishes the proof that each $\mathcal{A}_{m,n}$ is closed. Thus, so far we've proved that the set of all points $x \in D$ such that f is differentiable equals $\bigcap\limits_{m=1}^{+\infty}\bigcap\limits_{n=1}^{+\infty} D \setminus \mathcal{A}_{m,n}$, which is a G_δ set.

The next step is to show that each set $D \setminus \mathcal{A}_{m,n}$ is dense in D. Take any $x_0 \in D$, $\epsilon > 0$ and define $I := \{r \in \mathbb{R} : x_0 + r(x_n - x_0) \in D\}$, the function $f_1 : I \to \mathbb{R}$ defined by $f_1(r) := f(x_0 + r(x_n - x))$ is convex, and hence differentiable with exception of at most countably many points. Then there is $r_0 \in \mathbb{R}$ such that $|x_0 - x'| < \epsilon$ and such that $f_1'(r_0) = (f|_{[x_0,x_n]})'(x')$ exists, where $x' := x_0 + r_0(x_n - x_0)$.

We only need to show $x' \in D \setminus \mathcal{A}_{m,n}$. Take x^*, y^* in $\partial f(x')$, so that x^* and y^* are subdifferentials to $f|_{[x_0,x_n]}$ at x'. Since $f|_{[x_0,x_n]}$ is differentiable at x', this implies that $x^* = y^*$ in $[x_0, x_n]$, and in special we have $x^*(x_n) = y^*(x_n)$. Thus, $(x^* - y^*)(x_n) = 0 < \frac{1}{m}$, for all $m \geq 1$, so that $x' \in D \setminus \mathcal{A}_{m,n}$ as we wanted.

Finally, by Baire Category Theorem, the intersection of countably many open dense sets in a complete metric space is dense, so that $\bigcap\limits_{m=1}^{+\infty}\bigcap\limits_{n=1}^{+\infty} D \setminus \mathcal{A}_{m,n}$ is dense. This finishes the proof.

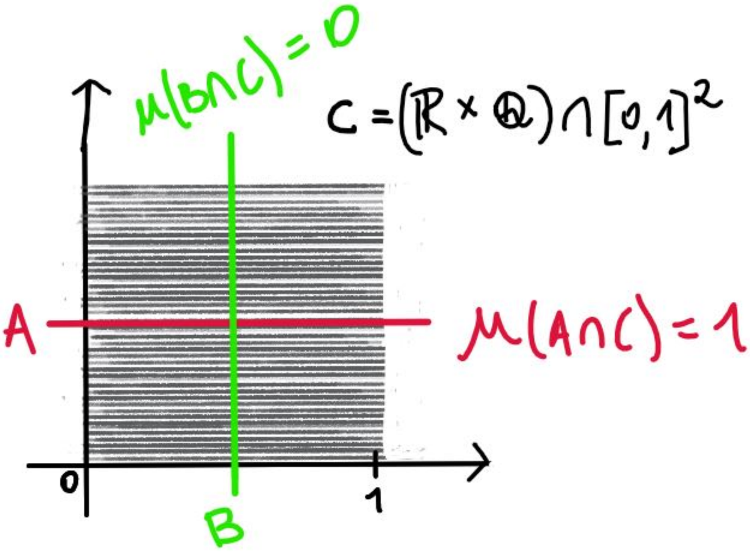
■

By the way, a dense G_δ set is known as a residual set, and this is the "big" set from the Topological point of view we were seeking. Note that Mazur's theorem is applicable since X is compact, so that $C(X)$ is separable.

4.2 Aronszajn null sets

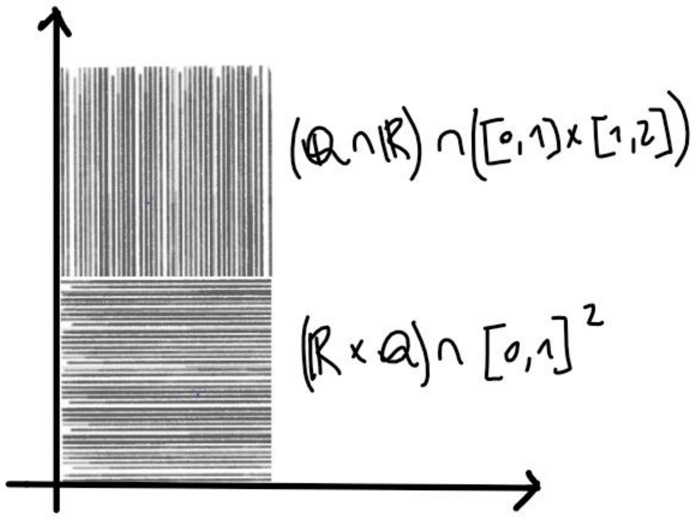
Our aim in this section is to generalize the notion of null set, for example, to separable Banach spaces. We will take the notion of Lebesgue null sets in the real line as primitive, and extend to higher dimensions from this viewpoint.

In first place, let's think about the notion of null set in two dimensions. Take for example the set $\mathbb{Q} \times \mathbb{R}$. We know that the measure of such set is zero by the very definition of product measure. More generally, it is easy to see by means of the Fubini's theorem that, if we have a set in a $n-$dimensional vector space such that, given a basis, its intersection with all the lines parallel to one of the basis vectors have measure zero, then this set is null. With this in mind, we could define a set to be null if, given a basis, its intersection with all the lines parallel to some basis vector has one-dimensional measure equals to zero, as the figure below illustrates



This definition, however, is not quite suitable. It leaves behind one of the most important properties of null sets: the countable union of null sets must also be null.

In fact, consider the set $\mathbb{Q}\times[0,1]\cup\mathbb{R}\times(\mathbb{Q}\cap(1,2])$. Neither the horizontal nor the vertical direction has null intersections with this set. We can, however, split this set in two, such that each part satisfies our previous criterion.



In finite dimension, we only need to check this property to one basis. In the general setting, however, we will explicitly demand that this holds for every sequence spanning a dense subspace (this will be the substitute for the notion of basis). The formal definition is as follows.

Let B be a separable Banach space endowed with Borel $\sigma-$algebra and $a\in B$ different of 0. The set $\mathcal{A}(a)$ is defined as the set of all Borel subsets which intersect each parallel line to a in a Lebesgue measure zero set. More precisely, $\mathcal{A}(a)$ is the set of all Borel sets A such that $\lambda\{t\in\mathbb{R};x+ta\in A\}=0$, $\forall x\in B$. Given a possibly finite sequence of vectors $\{a_k\}$, $\mathcal{A}(\{a_k\})$ is the collection of all sets A that can be split in a family $\{A_k\}$ such that $A=\bigcup_k A_k$. Finally, we say that a set is Aronszajn null if it is in $\bigcap\mathcal{A}(\{a_k\})$, where the intersection is taken over all the possible sequences whose span is dense.

This definition is surely cumbersome. We will prove a result that offers us a shortcut to prove that some sets are Aronszajn null. We present a technical proposition before it.

Proposition 4.2. *Let E be a separable Banach space, $F \subset E$ a $n-$dimensional subspace and λ_n the Lebesgue measure on F induced by the isomorphism of F and $\mathbb{R}^n$. Then, for a given Borel subset B of E, the function $f_B : F \to \mathbb{R}$ defined by $f_B(x) = \lambda_n(F \cap (B + x))$ is measurable.*

Proof. We will prove the assertion first for B bounded and closed. For such, we will prove that $f_B^{-1}([\alpha, +\infty))$ is closed. In fact, take a convergent sequence $(x_k) \subset f_B^{-1}([\alpha, +\infty))$, and we wish to show that $x = \lim_k x_k \in f_B^{-1}([\alpha, +\infty))$.

Notice that:

$$\bigcap_{m \geq 0} \bigcup_{k \geq m} (B + x_k) \subset B + \lim_k x_k$$

In fact, if y is in the first set, it means that $y \in B + x_k$ for infinitely many k. Therefore, there is a sequence $(b_j) \subset B$ such that $y = b_j + x_{k_j}$ for every j. With that, it is clear that (b_j) is convergent, since $b_j = y - x_{k_j}$. Using the fact that B is closed, we get that $b = \lim b_j \in B$. Then, since $y = b + \lim x_k$, we have $y \in B + \lim x_k$.

Now:

$$f_B(x) = \lambda_n \left(F \cap (B + \lim x_k)\right) \geq \lambda_n \left(F \cap \bigcap_{m \geq 0} \bigcup_{k \geq m} (B + x_k) \right) = \lambda_n \left(\bigcap_{m \geq 0} \bigcup_{k \geq m} F \cap (B + x_k) \right)$$

It is obvious that $\bigcup_{k \geq m} F \cap (B + x_k)$ is descrent. Let us see that $\lambda_n(\bigcup_{k \geq 0} F \cap (B + x_k)) < +\infty$, and we will conclude, by the semicontinuity:

$$f_B(x) \geq \lim_m \lambda_n \left(\bigcup_{k \geq m} F \cap (B + x_k) \right) \geq \lim_m \lambda_n \left(F \cap (B + x_m)\right) = \alpha$$

It suffices to prove that $\bigcup_{k \geq 0} F \cap (B + x_k)$ is bounded. Indeed, this set is a subset of $\bigcup_{k \geq 0}(B + x_k)$ and, since (x_k) converges, there is a k_0 with maximum length: $||x_k|| \leq ||x_{k_0}||$. The conclusion follows from the fact that B is bounded.

Clearly we can split $\mathbb{R}^n$ in an enumerable family of disjoint sets, such as spherical shells whose outer surface is open and inner surface is closed. Let E denote one of this shells.

We will prove the above result can be extended to all Borel sets of E. Let $\mathcal{E}$ be the collection of all subsets $B \subset E$ such that f_B is mensurable. We will prove that $\mathcal{E}$ is a $\sigma-$additive class:

(1) It is closed by union of two disjoint sets. Let $A, B \in \mathcal{E}$ disjoint. Then, $f_{A \cup B} = \lambda(E \cap ((A \cup B) + x)) = \lambda(E \cap ((A + x) \cup (B + x))) = \lambda((E \cap (A + x)) \cup (E \cap (B + x))) = \lambda(E \cap (A + x)) + \lambda(E \cap (B + x)) = f_A + f_B$. Where we used the additivity of the measure together with the fact that A, B are disjoint. The result follows from the fact that the sum of measurable functions is mensurable.

(2) If $A, B \in \mathcal{E}$, with $A \subset B$, then $B \backslash A \in \mathcal{E}$. Indeed, let $C := B \backslash A$. $f_C = \lambda(E \cap (C + x)) = \lambda(E \cap ((B + x) \backslash (A + x))) = \lambda((E \cap (B + x)) \backslash (E \cap (A + x))) = \lambda(E \cap (B + x)) - \lambda(E \cap (A + x))$, once E is bounded, therefore has finite measure. Then, $f_C = f_B - f_A$, and f_C is mensurable.

(3) Let (A_k) be a countable family of crescent sets in $\mathcal{E}$ whose union is A. Then $A \in \mathcal{E}$. Just notice that $f_A = \lambda(E \cap (A + x)) = \lambda(\bigcup_k (E \cap (A_k + x))) = \lim_k \lambda(E \cap (A_k + x))$, that is mensurable, once the limit of a sequence of mensurable functions is mensurable.

Using the $\sigma-$additive class lemma, we get that $\mathcal{E}$ is a $\sigma-$ring. It is not difficult to see that E can be write as the limit of compact sets. Thus, $E \in \mathcal{E}$ and $\mathcal{E}$ is the Borel $\sigma-$algebra of E

Now, take any Borel set B of $\mathbb{R}^n$. We can split this set as the union of an enumerable family B_k, such that each element of the family is in one of the spherical shells. Then it is straightforward to notice that f_B is mensurable, by the same argument use to prove (1). ∎

Lemma 4.3. *Let F be an $n-$dimensional subspace of a Banach space E, let λ_n denote the Lebesgue measure on F and $\{y_k\}$ be a basis for F. If A is a Borel subset of E such that $\lambda_n(F \cap (A + x)) = 0$ for every $x \in E$, then $A \in \mathcal{A}(\{y_k\})$.*

Proof. The proof will be by induction. For the initial case, it is straightforward to notice that $F \cap (A + x) = (F + x) \cap A = (\mathbb{R}y_1 + x) \cap A$, whence $\lambda((\mathbb{R}y_1 + x) \cap A) = 0$, for every x, what is the same as saying that $A \in \mathcal{A}(y_1)$.

Now, we are going to show that A can be split into two sets, A_n and A' such that $A_n \in \mathcal{A}(y_n)$ and $A' \in \mathcal{A}(\{y_1^{n-1}\})$, implying $A \in \mathcal{A}(\{y_k\})$, as we want.

Given A, define $A_n := \{x \in A; \lambda_1(A \cap (x + \mathbb{R}y_n)) = 0\}$. By the previous lemma, A_n defined this way is a Borel set. For showing that $A_n \in \mathcal{A}(y_n)$, take $x \in E$. If the line $L = x + \mathbb{R}y_n$ intercepts A_n, then there is $x' \in A_n$ such that the line $L' = x' + \mathbb{R}y_n$ coincides with L. By definition of A_n, $\lambda_1(A \cap L) = \lambda_1(A \cap L') = \implies \lambda(A_n \cap L) = 0$. Otherwise, if L does not intercept A_n and the measure is trivially zero.

Lastly, define $A' := A \backslash A_n$. By the induction hypothesis, we must only prove that $\lambda_{n-1}(G \cap (A' + x)) = 0$, where G is the subspace spanned by $\{y_1, ..., y_{n-1}\}$. Denote $B = G \cap (A' + x)$. By hypothesis, $0 = \lambda_n((A + x) \cap F) \geq \lambda_n((A' + x) \cap F)$. Using Fubini:

$$0 \geq \lambda_n((A' + x) \cap F) = \int_{G \times \mathbb{R}y_n} \chi_{(A'+x)}(u, v) d(\lambda_{n-1} \otimes \lambda_1)(u, v)$$

$$= \int_G \left(\int_{\mathbb{R}y_n} \chi_{(A'+x)}(u, v) d\lambda_1(v) \right) d\lambda_{n-1}(u) = \int_G \lambda_1((A' + x) \cap (u + \mathbb{R}y_n)) d\lambda_{n-1}(u)$$

$$\implies \int_B \lambda_1((A' + x) \cap (u + \mathbb{R}y_n)) d\lambda_{n-1}(u) = 0$$

We will know prove that the integrand is always greater than 0, and we will be forced to admit that $\lambda_{n-1}(B) = 0$.

Indeed, if $u \in B$, then $u \in A' + x$ and $u - x \in A'$. Since the Lebesgue measure is invariant by translation, and $(A \cap B) + x = (A + x) \cap (B + x)$, we have that $\lambda_1((A' + x) \cap (u + \mathbb{R}y_n)) = \lambda_1(A' \cap ((u - x) + \mathbb{R}y_n))$ and the assertion follows by the definition of A'. ■

4.3 Differentiability from a Measure Theoretic Perspective

Now, we present the main dish:

Theorem 4.4. *Let E be a separable Banach space and consider a Lipschitz mapping $f : U \to \mathbb{R}$, where $U \subset E$ is open. Then f fails to be Gateaux differentiable in a Aronszajn null set.*

Proof. First of all, we already proved that the set of points $x \in E$ where f is Gateaux differentiable is Borel. In a similar fashion, one can prove that every other set of interest in this proof will be Borealian.

So, let (x_k) be a sequence of vectors whose span is dense. Denote by V_n the subspace generated by $(x_1, ..., x_n)$ and D_n the subset of points of E such that the derivative of f does not exist in some direction of V_n or is not linear. Then, for every $y \in E$, $(D_n + y) \cap V_n$ is the set where the restriction of the function $f_y(x) = f(x - y)$ fails to be Gateaux differentiable. We already know that this set has measure zero in V_n. It follows from the previous lemmas that $D_n \in \mathcal{A}(x_1, ..., x_n)$. Then, f is not Gateaux differentiable in $\bigcup D_n \in \mathcal{A}(\{x_k\})$.

■

5 Generalizations of Pressure

5.1 Overview

In this section we will discuss briefly about some attempts to generalize the definition of pressure to the non-compact case and the problems addressed to this.

As already discussed, the variational principle is utterly important to a healthy definition of pressure, one of the reasons being the physical motivation. Thus, we expect the definition of pressure in non-compact spaces to satisfy this principle. And, therefore, we could define:

Definition 5.1. Let $\phi : X \to \mathbb{R}$ be a continuous function. The variational pressure of ϕ is defined by:

$$P_{var}(\phi) = \sup\left\{h(\mu) + \int \psi d\mu; \mu \in \mathcal{M}_1(T), \int \psi d\mu > -\infty\right\}$$

The condition of $\int \psi d\mu > -\infty$ assures the lack of undefined expressions $\infty - \infty$, when the entropy is not finite.

However, we want to find a more direct definition, as well as done in the compact setting, with the requirement that this alternative definition must also satisfy the variational principle. The following sections will be devoted to briefly study this definitions, remarking the advantages and drawbacks of each.

Firstly, we claim that the definition with (n, ϵ)-separated sets is not suitable anymore, once this may depend on the metric and not on the topology only, as the variational principle does not depend on the metric.

We will now offer a counter-example. Let $X = (0, \infty)$ endowed with the usual metric and $T(x) = 2x$. Let's estimate an lower bound for the entropy (which is just the pressure on $\phi = 0$) using $(n, \epsilon)-$spanning sets. Take $K = [1, 2]$ as the compact set. Taking ϵ of the form $1/m$ it's not difficult to see that:

$$r_n\left(\frac{1}{m}, K\right) = \begin{cases} \frac{m}{2}, & \text{if } m \text{ is even}, n = 1 \\ \frac{m+1}{2}, & \text{if } m \text{ is odd}, n = 1 \\ m2^{n-2}, & n \geq 2 \end{cases}$$

Thus,

$$\limsup_n \frac{1}{n} \log r_n(1/m, K) = \limsup_n \frac{1}{n} \log(m) + \limsup_n \frac{n-2}{n} \log 2 = \log 2$$

And, finally:

$$h(T) \geq \lim_{\epsilon \to 0} \limsup_n \frac{1}{n} \log r_n(\epsilon, K) = \lim_{m \to \infty} \log 2 = \log 2$$

Whence $h(T) > 0$. We will see now that we can find a metric where $h(T) = 0$. For such, consider the following statement.

Lemma 5.1. *If d is a isometry for T, then $h(T) = 0$.*

Proof. Just notice that, if d is an isometry, then $r_n(\epsilon, K) = r_1(\epsilon, K)$, for every n. Thus, the limsup is zero and hence the entropy. ∎

Now, we only have to find an isometry to T. Indeed, take the metric d' defined by $d'(x, y) = d(x, y)/2^n$ for $x, y \in [2^n, 2^{n+1}]$, where d is the usual one. It is not hard to see that it is in fact a metric and an isometry for T.

5.2 Interior and Exterior Pressure

Two other concepts of the pressure, related to each other, are the interior and exterior pressure. The motivation of these concepts is to be a kind of limit of the pressures of compact sets approximating the original set. The interior pressure takes compact subsets of the original set and the exterior pressure tries to define compact sets containing the original set. The precise definitions are the following.

Definition 5.2. Let ϕ be a continuous function. The interior pressure is:

$$P_{int}(\phi) = \sup\{P_\Lambda(\phi) : \emptyset \neq \Lambda \subset X\}$$

Where Λ must be compact and $T-$invariant. The exterior pressure is:

$$P_{ext}(\phi) = \inf\{P_{\overline{X}}(\phi)\}$$

where the infimum is taken over all possible compact metric spaces $\overline{X}$ such that X can be continuously embedded and ϕ can be continuously extended.

The most important drawback of this definition is that it not always coincides with the variational pressure. Namely, when we have few compact invariant subsets. In reality, we could have *none*. A dynamical system which has no non-trivial closed invariant subset is called minimal. More than stating the existence of such systems, we will prove that *every* compact system has a minimal subset. The proof, unexpectedly, uses the Zorn's lemma.

Proposition 5.2. *Let (X, T) be a compact dynamical system. Then, there is a minimal subset $K \subset X$.*

Proof. Let $\mathcal{S}$ be the set of all closed invariant subsets of X. This is non-empty, once $X \in \mathcal{S}$. Also, the inclusion defines a partial order. Now, we are going to prove that every totally ordered subset of $\mathcal{S}$ has minimal element. In fact, let (A_i) be a family of elements of $\mathcal{S}$ that is totally ordered. Let $A = \bigcap_i A_i$. Since the arbitrary intersection of compact sets is compact, A is compact. Further, it is also non-empty, by the finite intersection property and the total order of the family. Lastly, A is invariant, Indeed, $T^{-1}(A) = \bigcap_i T^{-1}(A_i) = \bigcap_i A_i = A$.

By the Zorn's lemma, $\mathcal{S}$ has a minimal element, which is a minimal set. ■

5.3 Pressure of Pesin-Pitskel

In 1973, Bowen came up with a definition of entropy for subsets of compact sets. His definition was inspired by the definition of Hausdorff dimension and can generalize some results concerning it. Later, Pesin and Pitskel generalized the notion to an analogue of pressure.

There are some similarities among Lebesgue measure, Hausdorff measure and Bowen entropy. Although we cannot draw an exact parallel, I think that may be didactic to follow a presentation emphasizing this similarities. More specifically, all the three concepts follows loosely the same mold, that has to do with defining outer measure via gauge. The following section makes it with more details.

5.3.1 Gauges and Hausdorff Measures

One standard way to construct measures is first considering outer measures. An outer measure is a function $\mu : \mathcal{P}(X) \to [0, +\infty]$ such that $\mu(\emptyset) = 0$ and:

$$\mu(A) \leq \sum_{j=1}^{\infty} \mu(A_j)$$

Provided $A \subset \bigcup_{j=1}^{\infty} A_j$.

Yet, there is a standard way to construct outer measures. We call $(\mathcal{G}, \phi)$ a gauge for $\mathcal{G} \subset \mathcal{P}(X)$ and $\phi : \mathcal{G} \to [0, +\infty]$ if $\inf\{\phi(G), G \in \mathcal{G}\} = 0$ and $\bigcup_j G_j = X$, for some $\{G_j\} \subset \mathcal{G}$. In this case, one can prove that the following function is an outer measure:

$$\mu(A) = \inf\left\{\sum_{j=1}^{\infty} \phi(G_j); A \subset \bigcup_{j=1}^{\infty} G_j, G_j \in \mathcal{G}\right\}$$

Indeed, Lebesgue measure is obtained taking the gauge to be the set of all cubes and ϕ to be their natural volume.

With all that in mind, it is easier to introduce Hausdorff measure. We can think of (s,δ)-Hausdorff measure as a kind of gauge such that $\mathcal{G}$ is the collection of sets with diameter less than δ and ϕ gives the diameter of the set powered by s (apparently, it is not the same to take just balls or cubes). One difference, however, is that we may take arbitrary summations and unions, not only countable ones. The $s-$Hausdorff measure is an outer measure not directly constructed by means of gauge, however. We define it to be $\mathcal{H}^s(A) = \lim_{\delta\to 0}\mathcal{H}^s_\delta(A)$.

The Hausdorff dimension of a set will be the infimum of the exponents such that the $s-$Hausdorff measure is 0.

5.3.2 Entropy of Bowen

To define the entropy, we are going to define an analogue to $(s,\delta)-$Hausdorff measure, that we will name by $(\mathscr{P}, s)-$Bowen measure, where $\mathscr{P}$ is a finite open cover of X. The first drastic difference with the former concept is that here it depends upon an open cover.

The difference that really matters, however, is that ϕ will not account anymore for the pure size of the set, as it was in case of Lebesgue and Hausdorff measure. The “size" now will be influenced by the action of the dynamic f. The function ϕ will be the same as in Hausdorff measure (i.e., something powered by s), but the diameter will be substituted by $\exp(-n(A))$, where n is the biggest integer such that $f^k(A)$ is contained in some member of $\mathscr{P}$ for all previous nonnegative integer.

Let's attempt to construct some intuition on this analogue to the diameter that we will call “dynamical diameter". In first place, the maximum it may take is 1, when A is not contained in any member of $\mathscr{P}$. The dynamical diameter will tend to zero if A still be contained in some member for many iterations of f. Thus, this dynamical diameter, in some sense, still talks about some kind of size — if A is small, it has greater chances to be contained in some member of $\mathscr{P}$, and the smaller is A, greater will be the chances the A will be in some member after many iterations. But this notion also accounts for the dynamic. In regions where the dynamic carries every point near to itself, it will be easier to have a smaller “dynamical diameter". This will be way more difficult in regions where the dynamic is more unstable.

Another difference is that $(\mathscr{P}, s)-$Bowen measure will be obtained by the lim inf as the dynamical diameter tends to zero, not the infimum.

And lastly, the union must always be countable, as it was in the definition of Lebesgue measure.

In summary, the $(\mathscr{P}, s)-$Bowen measure is defined as:

$$m(A) = \liminf_{\delta\to 0}\left\{\sum_{j=1}^{\infty}\exp(-s.n(A_j)); A\subset\bigcup_{j=1}^{\infty}A_j, \exp(-n(A_j)) < \delta\right\}$$

(We could define a $(s,\delta,\mathscr{P})-$Bowen measure such that, for each δ, one would have as $\mathcal{G}$ the collection of sets with dynamical diameter less than δ. The $(\mathscr{P}, s)$ would be the limit as δ tends to zero. This would be more identical to our definition of Hausdorff measure. It may be the case that this alternative definition coincides with the actually presented, but I am not sure.)

The entropy with respect to some open cover $\mathscr{P}$ will be the infimum of the exponents such that the measure is 0, and makes an analogue to Hausdorff dimension. But this still depends upon the cover. The entropy will be the supremum taken over the collection of open covers.

This definition does not satisfies the variational principle always either, as it can be seem in the following example.

Take an alphabet $V = \{a, b\}$ of two symbols and consider the Full Shift $\sigma : \Sigma \to \Sigma$. First of all, we will define a Hölder continuous map $g : \Sigma \to \mathbb{R}$ such that the Birkhoff average around a point $x \in \Sigma$ does not exist. For every $x \in \Sigma$, let $g(x) = 1$ if the first letter of x is a and

$g(x) = -1$ otherwise. If g_n denotes the n'th Birkhoff average of g, then, for every $n \geq 1$ and for $x_0 = abaabbaaaabbbbaaaaaaaabbbbbbbb...$ we have:

$$g_{2^{n+1}-2} = \frac{1}{2^{n+1}-2} \sum_{i=1}^{2^{n+1}-2} g(\sigma^i(x_0)) = 0$$

Since the sequence $(2^{n+1} - 2)_{n \geq 1}$ translates to $(2, 6, 14, ...)$, which one can easily check that $\sigma^{2^{n+1}-2}(x_0)$ always ends on the last b of a sequence of consecutive b's, and since before this sequence there is the same number of $a's$ and $b's$, the sum of $g(\sigma^i(x_0))$ from the first letter up to the $2^{n+1}-2$'th one must be equal to zero. Likewise, we have:

$$g_{3.2^n-2} = \frac{1}{3.2^n - 2} \sum_{i=1}^{3.2^n-2} g(\sigma^i(x_0)) = \frac{2^n}{3.2^n - 2}$$

Since the sequence $(3.2^n - 2)_{n \geq 1}$ translates to $(4, 10, 22, ...)$. One can easily check that $\sigma^{3.2^n-2}(x_0)$ always ends in the end of a sequence of consecutive a's. Since for each $n \geq 1$ up to $3.2^n - 2$ there are 3.2^n-2 elements and $1+2+4+...+2n = 2^n-1$ letters b, the summation is just $3.2^n-2-2(2^n-1) = 2^n$. As n tends to infinity, the second sequence tends to $\frac{1}{3}$ and the first to zero, so the sequence g_n has two converging subsequences which tend to different limits, and so the whole sequence cannot converge. Now, note that g is Hölder continuous, since:

$$V_n(g) = \sup_{\substack{x,y \in \Sigma \\ x_i = y_i \forall i=1,2,...,n-1}} |g(x) - g(y)| \leq 2d(x,y)^n$$

To check this, suppose that $x \neq y$, so that for every $n \geq 1$ we have $d(x,y)^n < 1$ and in this case the first letters of g differ, so that the supremum is zero. Thus, if $x \neq y$ we have $V_n(g) = 0 \leq 2d(x,y)^n$. If $x = y$, then $d(x,y)^n = 1$ and the supremum is always bounded by 2, so that $V_n(g) \leq 2 = 2d(x,y)^n$, so g is Hölder continuous. Now, define:

$$\mathbb{B}(g) = \left\{ x \in \Sigma : \lim_{n \to +\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(\sigma^i(x)) \text{ does not exist} \right\}$$

Now, by Birkhoff's Ergodic Theorem, $\mathbb{B}(g)$ is measurable and for every $\sigma-$invariant measure μ, the measure of $\mathbb{B}(g)$ is zero. Now, to show that $\sigma^{-1}(\mathbb{B}(g)) = \mathbb{B}(g)$, it is enough to show that $\sigma^{-1}(\mathbb{B}(g)^{-1}) = \mathbb{B}(g)^{-1}$, so take $x \in \Sigma$ such that the Birkhoff average converges and note that:

$$\lim_{n \to +\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(\sigma^i(\sigma(x))) = \lim_{n \to +\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(\sigma^{i+1}(x)) = \lim_{n \to +\infty} \frac{1}{n} \sum_{i=1}^{n} g(\sigma^i(x))$$

$$\lim_{n \to +\infty} \frac{1}{n-1} \sum_{i=1}^{n-1} g(\sigma^i(x)) = \lim_{n \to +\infty} \left(\frac{1}{n-1} \sum_{i=0}^{n-1} g(\sigma^i(x)) + \frac{g(x)}{n-1} \right)$$

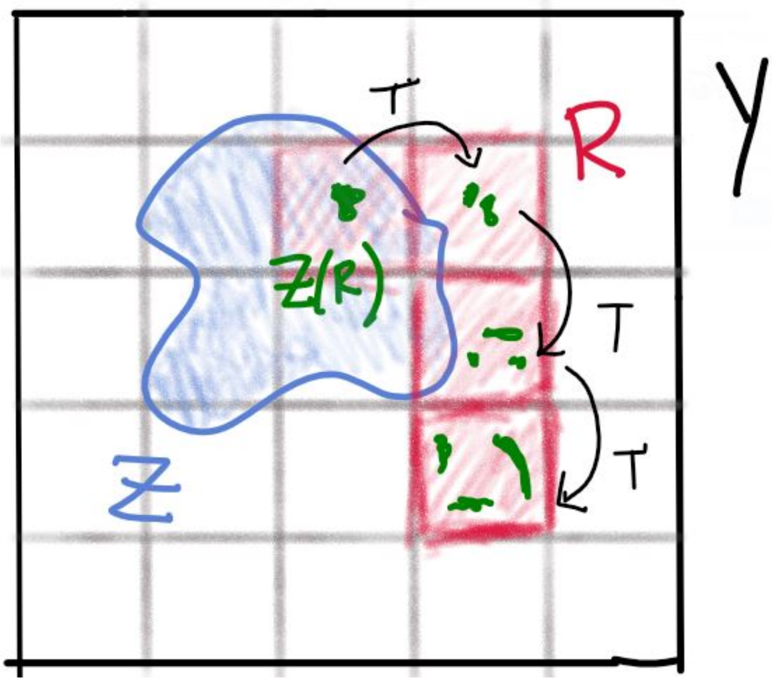
$$= \lim_{n \to +\infty} \frac{1}{n} \frac{n}{n-1} \sum_{i=0}^{n-1} g(\sigma^i(x)) = \lim_{n \to +\infty} \frac{1}{n} \sum_{i=0}^{n-1} g(\sigma^i(x))$$

So that $\mathbb{B}(g)^{-1} \subset \sigma^{-1}(\mathbb{B}(g)^{-1})$. The other inclusion is similar, so that $\mathbb{B}(g)$ is $\sigma-$invariant. Remembering that $f(f^{-1}(X)) \subset X$ for an arbitrary function, it follows that in special $\sigma(\mathbb{B}(g)) = \sigma(\sigma^{-1}(\mathbb{B}(g))) \subset \mathbb{B}(g)$, so that the dynamical system $\sigma : \mathbb{B}(g) \to \mathbb{B}(g)$ is well-defined. Now, we say that g is cohomologous to f if there exists a function $\psi \in C(\Sigma)$ such that $g = f + \psi - \psi \circ \sigma$. In [BS00], Theorem 2.1, it is shown that if a function f is not cohomologous to a constant, then $h(\sigma|_{\mathbb{B}(g)}) = h_{\text{top}}(\sigma) = \log(2)$, where the first entropy is Bowen's entropy. A result in [BS00] says that for Hölder continuous functions f, g, we have that f and g are cohomologous to each other if, and only if $\mathbb{B}(f) = \mathbb{B}(g)$. Now, it is trivial to see that the Birkhoff averages of any constant always exists, so in special $\mathbb{B}(c) = \varnothing$ for every constant c. Since in our construction $\mathbb{B}(g) \neq \varnothing$, it follows that g is not cohomologous to any constant, so that it must hold that $h(\sigma|_{\mathbb{B}(g)}) = \log(2)$. Now, the variational entropy $h_{\text{var}}(\sigma|_{\mathbb{B}(g)}) = \sup\{h_\mu : \mu \in \mathcal{M}(\sigma|_{\mathbb{B}(g)})\}$ is zero, since any measure supported in $\mathbb{B}(g)$ vanishes. Thus, the variational entropy and Bowen's entropy differ in this case.

5.3.3 Pressure of Pesin-Pitskel'

Let X be a compact metric space, $Y \subset X$ and T continuous. Let $\mathscr{P}$ be a finite open cover of X and $Z \subset Y$.

Let R be a finite collection of elements of $\mathscr{P}$. We will name such a collection R by route and by $m-$route a route with m elements, for reasons that will become clear now. Denote by $Z(R)$ the points of Z that are in the first element of R, their images by f are in the second, the images of their images are in the third, and so on. Considering that f is the dynamic, $Z(R)$ is the points that follows the route R. Denote as $S_m\phi(R)$ the supremum of the m-th Birhkoff sum taken over all points that follow a certain $m-$route R, where ϕ is continuous. If $Z(R) = \emptyset$, $S_m\phi(R) = -\infty$.



We will name by circuit a collection of routes (possibly with different lengths). We say that a circuit Γ covers Z every point of Z follows one of the routes of Γ. In other words, $Z = \bigcup_{R\in\Gamma} Z(R)$. A $m-$circuit is a circuit made of routes with length greater than m. Denoting $m(R)$ the number of elements of the route R, we define:

$$\mathfrak{M}(\mathscr{P}, \lambda, Z, \phi, N) = \inf \left\{ \sum_{R\in\Gamma} \exp(-\lambda m(R) + S_m\phi(R)) \right\}$$

Where the infimum is taken over all $N-$circuits that covers Z. Then:

$$\mathfrak{m}(\mathscr{P}, \lambda, Z, \phi) = \lim_{N\to\infty} \mathfrak{M}(\mathscr{P}, \lambda, Z, \phi, N)$$

The function $\mathfrak{m}$ defined above is always a regular outer measure defined on $\mathcal{P}(Y)$.

Just like the Hausdorff measure and the Bowen entropy, this measure gives 0 or ∞ for every exponent λ, except by possibly one. We define then $P_Z(\mathscr{P}, \phi) = \inf\{\lambda : \mathfrak{m}(\mathscr{P}, \lambda, Z, \phi) = 0\}$.

To vanish the dependence on the partition, we define the pressure of Z as:

$$P_Z(\phi) = \lim_{diam(\mathscr{P}) \to 0} P_Z(\mathscr{P}, \phi)$$

Because of this limit, we must ask to Y be totally bounded, in order to always have some finite partition with diameter arbitrarily small.

5.4 Variational Pressure

Due to the lack of a better definition, that always satisfies the variational principle, we will take the variational pressure, already defined, as the official definition.

In the case the non-compact set is a countable Markov shift, there is a more concrete definition that coincides with the variational pressure for all summable variation function. This is the Gurevich Pressure, more detailed in the next section.

5.4.1 Gurevich Pressure

The first step to define pressure is define something that can take the place of the partition function. Given a letter a and a summable variation ϕ, we define:

$$Z_n(\phi, a) = \sum_{\sigma^n x = x; x_0 = a} e^{\phi_n(x)} = \sum_{\sigma^n x = x} e^{\phi_n(x)} \chi_{[a]}(x)$$

Where the summation may be infinite.

The Gurevich pressure will be defined by $\lim_{n \to \infty} \frac{1}{n} \ln Z_n(\phi, a)$. However, we need to prove that this is well defined. In other words, that the limit exists and that it is independent of a. To accomplish so, we must first prove the following lemma:

Lemma 5.3. *Let Σ be a topologically mixing countable markov shift and ϕ of summable variation. Then*

1. $\forall a, b, \exists C, c; Z_n(\phi, a) < C Z_{n+c}(\phi, b)$
2. $\forall a, m, n, Z_n(\phi, a) Z_m(\phi, a) < B_1^2 Z_{n+m}(\phi, a)$

Proof. 1. For $x \in [a]$ such that $\sigma^n x = x$ let $\psi(x)$ be an infinite repetition of $buwx_ov$, denoted by $(buwx_ov)^\infty$, where $w = x_0^{n-1}$ and u, v are words connecting b an a. They exist for the shift is transitive. Let's denote the length of u and v respectively as μ and ν. Let $k = i + j + 2$. Then, we have:

$$|\phi_{n+k}(\psi(x)) - \phi_n(x)| = \left| \sum_{i=0}^{n+k-1} (\phi \circ \sigma^i)((buwx_ov)^\infty) - \sum_{i=o}^{n-1} (\phi \circ \sigma^i)(x) \right| \leq$$

$$\left| \sum_{i=o}^{\mu} (\phi \circ \sigma^i)((buwx_ov)^\infty) \right| + \left| \sum_{\mu+1}^{n+\mu} (\phi \circ \sigma^i)((buwx_ov)^\infty) - \phi_n(x) \right| + \left| \sum_{n+\mu+1}^{n+\mu+\nu+1} (\phi \circ \sigma^i)((buwx_ov)^\infty) \right|$$

$$= \left| \sum_{i=o}^{\mu} (\phi \circ \sigma^i)((buwx_ov)^\infty) \right| + \left| \sum_{0}^{n-1} (\phi \circ \sigma^i)((wx_ov)^\infty) - \phi_n(x) \right| + \left| \sum_{0}^{\nu} (\phi \circ \sigma^i)((x_ov)^\infty) \right|$$

$$\leq |\phi_{\mu+1}[bua]| + |\phi_n((wx_ov)^\infty) - \phi_n(x)| + |\phi_{\nu+1}[avb]| \leq |\phi_{\mu+1}[bua]| + V_{n+1}(\phi_n) + |\phi_{\nu+1}[avb]|$$

$$\implies |\phi_{n+k}(\psi(x)) - \phi_n(x)| \leq |\phi_{\mu+1}[bua]||\phi_{\nu+1}[avb] + \log B_1$$

As we can see, the last expression is finite and depends only upon a, b, v, u, but not on n nor x. Now, define:

$$\log M = |\phi_{n+k}(\psi(x)) - \phi_n(x)| \leq |\phi_{\mu+1}[bua]||\phi_{\nu+1}[avb] + \log B_1$$

As $\phi_n(x) - \log M \leq \phi_{n+k}(\psi(x))$ and as ψ is a bijection, we get:

$$Z_n(\phi, a) = M \sum_{\sigma^n(x)=x;x_0=a} e^{\phi_n(x)-\log M} \leq M \sum_{\sigma^n(x)=x;x_0=a} e^{\phi_{n+k}(\psi(x))}$$

$$\leq M \sum_{\sigma^{n+k}(x)=x;x_0=b} e^{\phi_{n+k}(x)} = MZ_{n+k}(\phi, b)$$

2. We have:

$$Z_n(\phi, a)Z_n(\phi, a) = \sum_{\sigma^n x = x} \sum_{\sigma^m y = y} e^{\phi_n(x)+\phi_m(y)}$$

Obviously, x and y are periodic words of period n and m respectively. Take z to be the $(n+m)-$periodic constructed by repeating the constituent blocks of x and y alternatively, denoted by u and v. Then:

$$Z_n(\phi, a)Z_n(\phi, a) \leq \sum_{\sigma^n x = x} \sum_{\sigma^m y = y} e^{|\phi_n(x)+\phi_m(y)-\phi_{n+m}(z)|} e^{\phi_{n+m}(z)}$$

Now, note that $\phi_{n+m}(z) = \phi_n((uv)^\infty) + \phi_m((vu)^\infty)$, whence $|\phi_n(x) + \phi_m(y) - \phi_{n+m(z)}| \leq |\phi_n(x) - \phi_n((uv)^\infty)| + |\phi_m(y) - \phi_m((vu)^\infty)| \leq 2\log B_1$

$$\implies Z_n(\phi, a)Z_n(\phi, a) \leq B_1^2 \sum_z e^{\phi_{n+m}(z)} \leq B_1^2 \sum_{z\in[a];\sigma^{n+m}z=z} e^{\phi_{n+m}(z)} = B_1^2 Z_{n+m}(\phi, a)$$

Where the first summation is taken over all $z \in [a]$ such that $z = (uv)^\infty$, $u, v \in [a]$ and u, v have a length of respectively n and m.

■

Theorem 5.4. *Let Σ be a topologically mixing countable Markov shift. Let ϕ be a summable variation function. Then, $\lim \frac{1}{n} \log Z_n(\phi, a)$ exists and it is independent of a. It is never $-\infty$*

Proof. In first place, let see that the sequence itself is well-defined. Just for every $n > n_0$ (for somw n_0), there is an admissible word of length n connecting a and a, and then $Z_n(\phi, a) \neq 0$ for $n > n_0$, whence $\zeta_n := \log Z_n(\phi, a)$ is a real number for $n > n_0$.

Now, applying the logarithm in the second item of the lemma above, note that, there exists c such that $\zeta_n + \zeta_m \leq \zeta_{n+m} + c$, for every n, m. This property is known as "almost super-additivity". If $\log Z_n(\phi, a) = +\infty$, this condition implies trivially that the limit is $+\infty$.

Otherwise, fix $m > n$ and write $n = q_n m + r_n$ for $n > m$, with q_n natural and $r_n \in \{0, ..., m-1\}$. Applying the almost super-additivity recursively, we get:

$$\frac{\zeta_n}{n} \geq \frac{q_n\zeta_m + \zeta_r - q_n c}{q_n n m + r_n} \leq \frac{q_n\zeta_m + \zeta_r - q_n c}{p_n m}$$

Where p_n can be $q_n \pm 1$, depending if the numerator is positive or negative. So:

$$\frac{\zeta_n}{n} \geq \left(\frac{q_n}{p_n}\right)\frac{\zeta_m}{m} + \frac{\zeta_r}{p_n m} - \left(\frac{q_n}{p_n}\right)\frac{c}{m}$$

$$\implies \liminf \frac{\zeta_n}{n} \geq \frac{\zeta_m}{m} - \frac{c}{m}$$

Where we applied the liminf on n. Now, taking the sup with respect to m:

$$\liminf \frac{\zeta_n}{n} \geq \sup_{m \geq n_o} \frac{\zeta_m}{m} - \frac{c}{m} \geq \limsup \frac{\zeta_m}{m} - \frac{c}{m} = \limsup \frac{\zeta_m}{m}$$

And this implies that the limit exists.

Let's prove that it does not depends on a. By the item 1 of the previous lemma, given b, there are constants c, C and k, K such that:

$$cZ_{n-k}(\phi, a) \leq Z_n(\phi, b) \leq CZ_{n+K}(\phi, a)$$

$$\implies \log c + \log Z_{n-k}(\phi, a) \leq \log Z_n(\phi, b) \leq \log C + \log Z_{n+K}(\phi, a)$$

For $n > k$. The first inequality gives:

$$\frac{\log c}{n-k} + \frac{\log Z_{n-k}(\phi, a)}{n-k} \leq \frac{\log C}{n-k} + \frac{\log Z_{n+K}(\phi, a)}{n-k}$$

$$\implies \lim_n \frac{\log Z_n(\phi, a)}{n} \leq \lim_n \frac{\log Z_n(\phi, a)}{n}$$

The second inequality gives us the inequality in the other way.

Finally, given $x \in [a]$ such that $\sigma^n x = x$, we have that $\sigma^{kn} x = x$, for k natural. Thus:

$$Z_{kn}(\phi, a) \geq e^{\phi_{kn}(x)} = e^{k\phi_n(x)}$$

$$\implies \lim_k \frac{\log Z_{kn}(\phi, a)}{kn} \geq \frac{\phi_n(x)}{n} > -\infty$$

■

In what follows, we show that the Gurevich pressure is nothing besides the interior pressure, defined earlier.

Lemma 5.5. *Let X be topologically mixing and ϕ of summable variation. Then:*

$$P_G(\phi) = \sup\{P_Y(\phi) : Y \subset X\}$$

Where Y is topologically mixing finite Markov shift and P_Y is the topological pressure.

As always, we have that $h_{top}(T) = P(0)$. This allows us to prove directly from the definition of Gurevich pressure the following lemma:

Lemma 5.6. *Let (Σ, σ) be a topologically mixing Markov shift. Then, if for some n, there is infinitely many periodic points of period n, $h_{top}(T) = \infty$*

From the definition of Gurevich pressure, the entropy is the limit of the logarithm of the $n-$th root of the number of periodic points starting from some a. So, if we have infinitely many periodic points starting from the same letter, the entropy is obviously infinite. However, in the countable case, it is not obvious that the pure existence of infinitely many n-periodic points implies in the existence of infinitely many $n-$periodic points starting from the same letter. So, we will not prove this here completely.

Now, we will prove that the interior pressure equals the Variational Pressure in the case of a Countable Markov Shift (CMS) for uniformly continuous potentials:

Lemma 5.7. *Let $\varphi \in UC_{d_{\theta,\rho}}(\Sigma)$ be bounded, where Σ is a CMS. Then $P^{int}(\varphi) = P_{var}(\varphi)$.*

Proof. First of all, we'll prove that the interior Pressure is Lipschitz with Lipschitz constant equal to 1. The proof of this is analogous to that of Proposition 2.3, and is shown as follows: remembering that $P_{\text{int}}(f) = \sup\{P_A(f) : \varnothing \neq A \subset \Sigma\}$ where each A is compact, each $P_A(f)$ can be seen as the variational pressure on A, so that for each constant $c \in \mathbb{R}$ we have $P_{\text{int}}(f+c) = \sup\{P_A(f+c) : \varnothing \neq A \subset \Sigma\} = \sup\{P_A(f) + c : \varnothing \neq A \subset \Sigma\} = \sup\{P_A(f) : \varnothing \neq A \subset \Sigma\} + c = P_{\text{int}}(f) + c$, where we've utilized Proposition 2.3. Now, if $f \leq g$, then $P_A(f) \leq P_A(g)$ for every non-empty compact set $A \subset X$, so that $\sup\{P_A(f) : \varnothing \neq A \subset \Sigma\} \leq \sup\{P_A(g) : \varnothing \neq A \subset \Sigma\} \implies P_{\text{int}}(f) \leq P_{\text{int}}(g)$ and thus, repeating the proof of the third item in Proposition 2.3 we conclude that $|P_{\text{int}}(\varphi) - P_{\text{int}}(\psi)| \leq |\varphi - \psi|$.

Now, let $\epsilon > 0$ be given and note that the thesis is true for potentials of bounded variation by [Sar99], Theorem 2. Assuming for now that potentials of summable variation are dense in $UC_{d_{\theta,\rho}}$, we can find a potential f of summable variation such that $|f - \varphi| < \epsilon$, so that:

$$\begin{aligned}|P_{\text{int}}(\varphi) - P_{\text{var}}(\varphi)| &< |P_{\text{int}}(\varphi) - P_{\text{int}}(f)| + |P_{\text{int}}(f) - P_{\text{var}}(f)| + |P_{\text{var}}(f) - P_{\text{var}}(\varphi)| \\ &\leq 2|\varphi - f| < 2\epsilon\end{aligned}$$

Where we used that the interior Pressure and variational Pressure are equal for potentials of bounded variation. Thus, we only need to prove that potentials of bounded variation are dense in $UC_{d_{\theta,\rho}}(\Sigma)$.

To see why this is true, take a potential $\varphi \in UC_{d_{\theta,\rho}}(\Sigma)$ and any $\epsilon > 0$ and choose a $\delta > 0$ such that, by uniform continuity, $d(x, y) < \delta \implies |\varphi(x) - \varphi(y)| < \epsilon$. Now, choose $k \geq 1$ such that $\frac{\theta^k}{1-\theta} < \delta$. Now, take the partition of Σ in cylinders of length k as follows:

$$\Sigma = \dot{\bigcup_{\substack{x_i \in \mathbb{N} \\ 1 \leq i \leq k}}} [x_1, x_2, ..., x_k]$$

Consider an arbitrary cylinder of this partition, say Z, and consider any two points $x, y \in Z$. We have $\rho(x_i, y_i) = 0$ for all $1 \leq i \leq k$, so that:

$$d(x, y) = \sum_{n=1}^{+\infty} \theta^n \rho(x_n, y_n) = \sum_{n=k}^{+\infty} \theta^n \rho(x_n, y_n) \leq \sum_{n=k}^{+\infty} \theta^n = \frac{\theta^k}{1-\theta} < \delta$$

In special, for all $x, y \in Z$ we have $|\varphi(x) - \varphi(y)| < \epsilon$. Take any $x_Z \in Z$ and we shall now define $\tilde{\varphi} : \Sigma \to \mathbb{R}$ by $\tilde{\varphi}(x) = \varphi(x_Z)$ for $x \in Z$ (since we have a partition of Σ by cylinder sets with length k, this alone is enough to define $\tilde{\varphi}$). With this, for any $x \in \Sigma$ we can take a cylinder set Z containing x so that $|\varphi(x) - \tilde{\varphi}(x)| < \epsilon$, by what we've said before. All we need to show now is that $\tilde{\varphi}$ is of bounded variation. To see why, take two points $x, y \in \Sigma$ such that $x_i = y_i$, for all $1 \leq i \leq k$, so that $x, y \in [y_1, y_2, ..., y_k]$. Since $\tilde{\varphi}|_{[y_1, y_2, ..., y_k]}$ is constant, we have $|\tilde{\varphi}(x) - \tilde{\varphi}(y)| = 0$, and since x, y were arbitrary this argument shows that $\{|\tilde{\varphi}(x) - \tilde{\varphi}(y)| : x_i = y_i, \forall 1 \leq i \leq k\} = \{0\}$, so that $V_k(\tilde{\varphi}) = 0$. Since the sequence of the $V_n(\tilde{\varphi})$ are decreasing, this shows that $V_n(\tilde{\varphi}) = 0$, for all $n \geq k$. Finally, since φ is bounded, so is $\tilde{\varphi}$ and thus $V_0(\tilde{\varphi}) < +\infty$, so that $V_n(\tilde{\varphi}) < +\infty$, for all $n \geq 1$. In special, this proves that:

$$\sum_{n=0}^{+\infty} V_n(\tilde{\varphi}) = \sum_{n=0}^{k-1} V_n(\tilde{\varphi}) < +\infty$$

Which is what we've wanted to prove and this ends the proof.

■

Lemma 5.8. *Let (Σ_A, σ) be a topologically mixing CMS and $a \in \Sigma$. Let*

$$B_k := \#\{w \textit{ word}; |w| = k, w_0 = a, wa \textit{ is admissible and } w_i \neq a, \textit{ for } i \neq 0\}$$

If $B_k < \infty$ for every k, consider the function f defined by the following power series in the points where it converges.

$$f(x) = \sum_{k=1}^{+\infty} B_k x^k$$

If f is well-defined in some neighborhood of 0, then the following relation holds:

$$h_{top}(\sigma) = \lim_n \frac{1}{n} \log \frac{1}{n!} \frac{d^n}{dx^n} \frac{f(x)}{1 - f(x)}\bigg|_{x=0}$$

provided this limit exists.

Proof. In first place, we must point out that the number of words of length n with exactly $m \leq n$ times the letter a such that wa is admissible can be given by:

$$\frac{1}{n!} \frac{d^n}{dx^n} f^m(x)\bigg|_{x=0}$$

Where f^m, in this lemma, will always stand for power, not composition.

Indeed, first note that the sum of all terms of the form $B_{n_1} B_{n_2} ... B_{n_i}$, where there are exactly m terms in the product and $n_1 + n_2 + ... + n_i = n$ corresponds to the number of words of length n with $m \leq n$ times the letter a in the beginning we were seeking. To see why, just note that any such word can be broken into m pieces, each containing a exactly one time at it's end and each product $B_{n_1} B_{n_2} ... B_{n_i}$ accounts for the number of words w of length $n_1 + n_2 + ... + n_i = n$ (by concatenation) passing through a exactly a times, since each B_{n_k} passes through a exactly one time. Since we need to also take into account all permutations of the indices, our number is just:

$$\sum_{n_1 + n_2 + ... + n_i = n} \frac{n!}{n_1! n_2! ... n_i!} B_{n_1} B_{n_2} ... B_{n_i}$$

And this is known to be the term multiplying x^n in $\left(\sum_k B_k x^k \right)^m = f^m(x)$. Of course, this is just the n'th term in the Taylor series expansion of f^m, which is $\frac{1}{n!} \frac{d^n}{dx^n} f^m(x)\big|_{x=0}$.

Thus, the total number of words of length n that can form a periodic point starting with a is:

$$\sum_{m=1}^{n} \frac{1}{n!} \frac{d^n}{dx^n} f^m(x)\bigg|_{x=0} = \frac{1}{n!} \frac{d^n}{dx^n} \sum_{m=1}^{n} f^m(x)\bigg|_{x=0} = \frac{1}{n!} \frac{d^n}{dx^n} \frac{f(x) - f^{n+1}(x)}{1 - f(x)}\bigg|_{x=0}$$

Since we can have at most n times the letter a in the word. Our next task is to prove that:

$$\frac{d^n}{dx^n}\frac{f^{n+1}(x)}{1-f(x)}\bigg|_{x=0} = 0$$

To see why this must hold, note that $f(x) = x\sum_{k=1}^{+\infty} B_k x^{k-1} = xg(x)$, where $g(x) = \sum_{k=1}^{+\infty} B_k x^{k-1}$ is analytic. Then, we get $f^{n+1}(x) = x^{n+1}g^{n+1}(x)$ and by using the formula:

$$\frac{d^n}{dx^n}\Big|_{x=0}(f.g) = \sum_{k=0}^{n}\binom{n}{k}\frac{d^k}{dx^k}\Big|_{x=0}f(x)\frac{d^{n-k}}{dx^{n-k}}\Big|_{x=0}g(x)$$

In our case, this gives:

$$\frac{d^n}{dx^n}\frac{f^{n+1}(x)}{1-f(x)}\bigg|_{x=0} = \sum_{k=0}^{n}\binom{n}{k}\frac{n!}{k!}x^{n-k+1}\Big|_{x=0}\frac{d^{n-k}}{dx^{n-k}}\Big|_{x=0}h(x) = 0$$

Where $h(x) = \frac{g^{n+1}(x)}{1-f(x)}$, and this proves what we wanted. At the end of the day, we have that the number of periodic points of period n starting with a is:

$$\frac{1}{n!}\frac{d^n}{dx^n}\frac{f(x)}{1-f(x)}\bigg|_{x=0}$$

The lemma follows from the definition of Gurevich pressure for the potencial $\phi = 0$, which corresponds to the topological entropy.

■

6 Structure of $\overline{\Sigma}$

In this section, we are going to start the theory itself. We must mention that we will focus on topologically mixing CMS. So, let (Σ, σ) be a such a shift. Just like the finite case, the metric defined on it is usually given by (1).

Nevertheless, we will be concerned in this text with another class of metrics. Given a metric $\rho : \mathbb{N} \times \mathbb{N} \to [0, 1]$ and $\theta \in (0, 1)$ we define:

$$d_s(x, y) = \sum_{n\geq 0}\theta^n \rho(x_n, y_n)$$

Firstly, note that the metric is well defined in the full shift if, and only if ρ is bounded. Thus, there is not so much loss of generality to take 1 as an upper bound. We could, obviously, take any K as an upper bound. In fact, if ρ is not bounded, we could find sequences (x_n) and (y_n) such that $\rho(x_n, y_n) > \theta^{-n}$, and the series would diverge.

As already pointed out, we aim to consider the completion of Σ and compare the pressure defined in the two cases. Thus, this comparison will only make sense to functions that can be extended to the boundary. Due to a well-known analysis theorem, this is the case for uniformly continuous bounded functions. We know that the set of continuous functions depends only upon the topology. However, the set of uniformly continuous functions depends on a uniform structure (and therefore, on the metric). With this in mind, the greater the set of uniformly continuous functions a metric provides, the better. This can be regarded as one of the reasons to consider a class of metrics, instead of just one. We are left with more freedom to “choose" the set of uniformly continuous functions, by changing ρ and θ. In the next subsections we are going to briefly discuss the topology and the possible uniform structures that this class of metrics can give us.

6.1 Topology

A very natural question, once we are not using the usual metric, is whether the two metrics generate the same topology. In other words, does our metric generate the cylinder topology? Generally, the answer is no. We provide an example in the next paragraph. However, if ρ is discrete, then the answer is yes, and, from now on, we will always assume that ρ is discrete.

Let $\rho : \mathbb{N} \times \mathbb{N} \to [0,1]$ defined by $\rho(a,b) = \left|\frac{1}{a} - \frac{1}{b}\right|$ if $a, b \neq 1$, $\rho(1,b) = \frac{1}{b}$ if $b \neq 1$ and $\rho(1,1) = 0$. Consider also, the bijection $\mathbb{N} \to \{1/n, n > 1\} \cup \{0\}$ given by $n \mapsto 1/n$ if $n \neq 1$ and $1 \mapsto 0$. Then, this bijection is, in fact, an isometry. Note that $\mathbb{N}$ endowed with this metric is compact and 1 is a limit point.

Now, consider the full shift (Σ, σ) with the metric defined by:

$$d(x,y) = \sum_{n \geq 0} \theta^n \rho(x_n, y_n)$$

with $\theta \in (0,1)$ (take, for example, $\theta = 1/2$, but it is not important for what follows).

Take a ball in the metric of the minimum centered in $x = (1,1,...)$ with radius $\epsilon > 0$. This set, $C(x,\epsilon)$, will be the cylinder $[1,...,1]$, such that the size of the cylinder depends on ϵ. (take, for example, $\epsilon = 1/4$ and the ball will be $[1,1]$).

We will now prove that, for every $\delta > 0$, the ball centered in x with radius δ with the metric d will not be such that $B(x,\delta) \subset C(x,\epsilon)$. In other words, for every $\delta > 0$, we will find a $y \in B(x,\delta)$ such that $y \notin C(x,\epsilon)$.

Indeed, take $\delta > 0$. Let $n \in \mathbb{N}$ such that $1/n < \delta$. The word $y = (n,1,1,...)$ satisfies the required properties. Note that:

$$d(x,y) = \rho(1,n) + \sum_{n \geq 1} \theta^n \rho(1,1) = \frac{1}{n} + 0 < \delta$$

Hence, $y \in B(x,\delta)$. But it is very clear that $y \notin C(x,\epsilon) = [1,...1]$, once the first letter of y is not 1.

Indeed, the precise form of the metric is irrelevant. The relevant property is that 1 is a limit point and we can always find a letter arbitrarily close to it.

Now, we will look more carefully into $\mathbb{N}$ endowed with a metric ρ as before, but, as already said, with the additional property that this metric generates the discrete topology of $\mathbb{N}$ (of course, we must need that $\mathbb{N}$ as a subspace of the shift space has the discrete topology, which is the standard topology in the alphabet). If this is true, then we can prove that the metric d above generates the cylinder topology:

Lemma 6.1. *Let $\rho : \mathbb{N}^2 \to [0,1]$ be a metric in $\mathbb{N}$ such that the generated topology is the discrete topology. Then, the metric $d : \Sigma^2 \to \mathbb{R}$ defined by:*

$$d(x,y) = \sum_{n \geq 0} \theta^n \rho(x_n, y_n)$$

With $\theta \in (0,1)$ generates the cylinder topology.

Proof. We will denote the balls in the minimum metric (cylinders) with the letter C and the balls in the other metric with B. Then, it suffices to prove that, for every point x and $\epsilon > 0$ there exists δ_1 such that $C(x,\delta_1) \subset B(x,\epsilon)$ and δ_2 such that $B(x,\delta_2) \subset C(x,\epsilon)$.

We will prove the first statement. Given $\epsilon > 0$, it is clear that there exists n_0 such that $\sum_{n \geq n_o} \theta^n < \epsilon$. Take $\delta \leq 2^{-n_o}$. Thus, $C(x,\delta) \subset B(x,\epsilon)$. Indeed, $y \in C(x,\delta) \implies \min\{n, x_n \neq y_n\} > -\log_2 \delta$. But $\delta \leq 2^{-n_o} \implies \log_2 \delta \leq -n_o \implies -\log_2 \delta \geq n_o$. So, $\min\{n, x_n \neq y_n\} > n_o$, hence $x_n = y_n$, $\forall n \leq n_o$, from where $\rho(x_n, y_n) = 0, n \leq n_o$, and:

$$d(x,y) = \sum_{n\geq 0} \theta^n \rho(x_n, y_n) = \sum_{n\geq n_o} \theta^n \rho(x_n, y_n) \leq \sum_{n\geq n_o} \theta^n < \epsilon$$

Because ρ is always less than 1. Thus $y \in B(x,\epsilon)$.

Furthermore, since the topology of $\mathbb{N}$ is discrete, for every x_n, there exists δ_n such that $\rho(x_n, y_n) < \delta_n \implies x_n = y_n$. Given $\epsilon > 0$, take $n_o = \lceil -\log \epsilon \rceil$ and:

$$\delta = \min_{n\leq n_o} \{\delta_n / \theta_n\}$$

Then $B(x,\delta) \subset C(x,\epsilon)$. Indeed, $y \in B(x,\delta) \implies d(x,y) < \delta \implies \theta^n \rho(x_n, y_n) < \delta$, $\forall n$. In particular:

$$\rho(x_n, y_n) < \delta_n$$

$\forall n \leq n_o$. Thus, $y_n = x_n$, $\forall n \leq n_o$. It tells us that $\min\{n, x_n \neq y_n\} > n_o > -\log \epsilon$, therefore $y \in C(x,\epsilon)$. ■

Of course, at least one of such metric ρ exists (i.e, the discrete metric in $\mathbb{N}$). We will use this metric d in Σ to prove that the completion of Σ is compact provided ρ is totally bounded. First, we'll give some definitions and state a few results.

Definition 6.1. Let $\rho : \mathbb{N}^2 \to [0,1]$ be a metric. Using the terminology of [GS98], we say that ρ is of type 1 if:

$$\inf_{n,m;n\neq m} \rho(n,m) > 0$$

Also, we say that ρ if of type 2, or vanishing (we will use the later more often, maintaining the terminology due to Iommi) if:

$$\lim_{k\to+\infty} \sup_{n,m\geq k} \rho(n,m) = 0$$

♣

Notice that we are identifying the set of vertices with the set of natural numbers. The general case is identical.

Lemma 6.2. *The following assertions hold:*

1. *If ρ is totally bounded, then $d : \Sigma^2 \to \mathbb{R}$ is also totally bounded;*

2. *If ρ is vanishing, then ρ is totally bounded;*

3. *The shift map $\sigma : \Sigma \to \Sigma$ defined by $\sigma(x_1, x_2, ...) = (x_2, x_3, ...)$ is uniformly continuous in the metric d as before;*

4. *If ρ is totally bounded, then $\varphi : \Sigma \to \mathbb{R}$ is uniformly continuous if, and only if it extends to a function $\overline{\varphi} : \overline{\Sigma} \to \mathbb{R}$ which is continuous with respect to $\overline{d}$.*

Proof. 1. First, suppose that ρ is totally bounded and take some $\epsilon > 0$. Now, take $n > 0$ such that $\frac{\theta^n}{1-\theta} < \frac{\epsilon}{2}$ and any $\delta > 0$ such that $\delta < \frac{\epsilon}{2n}$. By hypothesis, there is a finite cover of $\mathbb{N}$ by open balls where each diameter is less than δ, i.e, we have $\mathbb{N} = V_1 \cup V_2 \cup ... \cup V_k$ and each of it's diameter is less than δ. Consider the collection:

$$[V_{i_1}, V_{i_2}, ..., V_{i_n}] := \left\{[x_1, x_2, ..., x_n] : x_j \in V_{i_j}, \forall 1 \leq j \leq n\right\}$$

Where $V_{i_j} \in \{V_1, V_2, ..., V_k\}$. We affirm that this collection covers Σ. To see why, just take any $x \in \Sigma$ and note that each x_m belongs to at least one V_{i_j} for every $1 \leq m \leq n$, since these sets cover $\mathbb{N}$. We'll now show that there is a ball of radius $< \epsilon$ of Σ that covers each set $[V_{i_1}, ..., V_{i_n}]$, and this will be enough, since there are k^n (a finite number) of those sets and they cover Σ. Fix any collection $[V_{i_1}, ..., V_{i_n}]$ and a point y belonging to it. To simplify the proof even more, assume without loss of generality that the collection is just $[V_1, V_2, ..., V_n]$. Since we have $x_i, y_i \in V_i$ for every $1 \leq i \leq n$, we have $\rho(x_i, y_i) < \delta$ for every $1 \leq i \leq n$, so that:

$$d(x,y) = \sum_{i \geq 0} \theta^i \rho(x_i, y_i) \leq \sum_{i=0}^{n} \theta^i \rho(x_i, y_i) + \sum_{i=n}^{+\infty} \theta^i \rho(x_i, y_i)$$

$$\leq \sum_{i=0}^{n} \theta^i \delta + \sum_{i=n}^{+\infty} \theta^i \leq n\delta + \frac{\theta^n}{1-\theta} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

So that $[V_1, ..., V_n] \subset B(y, \epsilon)$. This finishes the proof.

2. Let $\epsilon > 0$ be any. By definition of limit, there is a natural number $N > 0$ such that for all $n \geq N$ we get $\sup_{i,j \geq n} \rho(x_i, x_j) < \epsilon$. Of course, this implies that for all $i, j \geq N$ we also have $\rho(x_i, x_j) < \epsilon$. Take the finite sequence of natural numbers $x_1, x_2, ..., x_N$ and we'll show that the balls $B(x_i, \epsilon)$ cover $\mathbb{N}$. Without loss of generality, we can suppose that $x_i = i$, as the alphabet is countable. Taking any $y \in \mathbb{N}$, if $x_y = y \leq N$ then y trivially belongs to one of the balls before and, if $y > N$, then as $N, y \geq N$ we have $\rho(x_N, y) < \epsilon$, so that $y \in B(x_N, \epsilon)$. This proves that the collection before contains $\mathbb{N}$, as we wanted.

3. Take any $\epsilon > 0$ and choose $\delta = \theta\epsilon > 0$. Then, if $x, y \in \Sigma$ and $d(x,y) < \delta$, we have:

$$d(\sigma(x), \sigma(y)) = \sum_{n \geq 0} \theta^n \rho(x_{n+1}, y_{n+1}) = \frac{1}{\theta} \sum_{n \geq 0} \theta^{n+1} \rho(x_{n+1}, y_{n+1})$$

$$= \frac{1}{\theta} \left(\sum_{n \geq 0} \theta^n \rho(x_n, y_n) - \rho(x_0, y_0) \right) = \frac{d(x,y)}{\theta} - \frac{\rho(x_0, y_0)}{\theta} \leq \frac{1}{\theta} d(x,y) < \epsilon$$

4. The $\Longrightarrow$ part is already known. Suppose now that ρ is totally bounded and that a function $\varphi : \Sigma \to \mathbb{R}$ has a continuous extension $\overline{\varphi} : \overline{\Sigma} \to \mathbb{R}$, and we need to show that φ is uniformly continuous. Now, since ρ is totally bounded, we already proved that d is totally bounded. As we'll prove right after this proposition, this implies that $\overline{d}$ is totally bounded. Since $\overline{\Sigma}$ is complete, then $\overline{\Sigma}$ is compact. Thus, $\overline{\varphi}$ begin a continuous function defined on a compact metric space, it is a uniformly continuous function, and thus $\varphi = \overline{\varphi}|_{\Sigma}$ is a uniformly continuous function.

■

Now, remember that, by definition of completion, there is an isometric immersion $i : \Sigma \to \overline{\Sigma}$ such that the image $i(\Sigma)$ is dense in $\overline{\Sigma}$. With this, we can prove that $\overline{\Sigma}$ is totally bounded if Σ is. To see why, take any $\epsilon > 0$ and take a $\frac{\epsilon}{2}$ finite covering of Σ by open balls and suppose their centers are $x_1, x_2, ..., x_n$. We affirm that the $\epsilon-$balls of center $i(x_1), i(x_2), ..., i(x_n)$ cover $\overline{\Sigma}$ and this will be sufficient. Now, take any $y \in \overline{\Sigma}$ and since $i(\Sigma)$ is dense in $\overline{\Sigma}$, there is $x \in \Sigma$ such that $d_c(i(x), y) < \frac{\epsilon}{2}$, where d_c is the metric in $\overline{\Sigma}$. For this $x \in \Sigma$, there exists a x_i such that $d(x, x_i) < \frac{\epsilon}{2}$, so that:

$$d_c(y, i(x_i)) \leq d_c(y, i(x)) + d_c(i(x), i(x_i)) < \frac{\epsilon}{2} + d(x, x_i) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

And this proves that $y \in B(i(x_i), \epsilon)$, as we wanted. Since any complete and totally bounded space is compact, this shows that $\overline{\Sigma}$ is compact if $\rho : V^2 \to [0,1]$ generates the discrete topology and is totally bounded.

Yet with regard to the types of ρ, we have, defining $\alpha_n = \inf\{\rho(n,m); m \in \mathbb{N}\}$:

Proposition 6.3. *A metric ρ in $\mathbb{N}$ is of type 1 if, and only if,* $\inf \alpha_n > 0$. *More precisely:*

$$\beta := \inf_{n,m;n\neq m} \rho(n,m) = \inf_n \alpha_n$$

Proof. By the definition of α_n, it is clear that $\beta < \alpha_n$, for every n, once the set whose infimum defines β contains the set defining α_n. From this, it follows that β is a lower bound for $\{\alpha_n\}$, remaining to prove that it is the bigger one.

So, take $\epsilon > 0$. By the definition of β, there is n, m such that $\rho(n,m) < \beta + \epsilon$. However, by the definition of α_n, $\alpha_n < \rho(n,m)$. Thus, given $\epsilon > 0$, we have found a n such that $\alpha_n < \beta + \epsilon$, what concludes our proof. ■

Proposition 6.4. *If ρ is of type 1, then ρ is not totally bounded.*

Proof. Indeed, take $0 < \epsilon < \inf\{\rho(n,m), n \neq m\}$. If ρ was totally bounded, we could cover $\mathbb{N}$ with finitely many balls. However, as the set is infinite, there would be a ball centered in n, for example, containing infinitely many letters. But, then, there would be a letter m with $\rho(n,m) < \epsilon$, contradicting the way ϵ was constructed. ■

Corollary 6.4.1. *If ρ is totally bounded, then* $\inf \alpha_n = 0$.

The previous corollary tells us that the set of totally bounded metrics is a subset of the set of metrics that are not of type 1. This inclusion is proper. Take the metric that turns the bijection $n \mapsto n$ if n is even and $n \mapsto 1/n$ if n is odd an isometry. This metric is not totally bounded but is not of type 1 either. Furthermore, the set of metrics which are vanishing is a subset of the set of totally bounded metrics. This inclusion is also proper, and we will see examples later. The next image summarizes this discussion.

6.2 Uniform Equivalences

We recommend this section to be skipped in a first lecture, once the results exposed here are rather a matter of curiosity than usefulness itself.

Proposition 6.5. *Let (Σ, σ) be the full shift. Then, the metrics d_m, of the minimum and $d_s = d_{s,\rho,\theta}$ are not uniformly equivalent if $(\mathbb{N}, \rho)$ is discrete and $\theta \neq 1/2$.*

Proof. We will first prove that it holds when $\theta > 1/2$ It suffices to prove that, for every $k \in \mathbb{N}$, there are $x, y \in \Sigma$ such that:

$$kd_m(x,y) < d_s(x,y)$$

Since ρ is discrete, given $n \in \mathbb{N}$, we will have $\alpha_n > 0$ for every n.

Then, take x formed by $x_n = 0$, for each $n \in \mathbb{N}$ and y formed by $y_n = 0$ if $n < n_o$ and $y_n = 1$ if $n \geq n_o$, where:

$$n_o > \frac{\ln(k/\alpha)}{\ln 2\theta}$$

And $\alpha := \alpha_0$. We will show that this works. In fact, the previous inequality, together with the fact that $\theta > 1/2$ tells us that:

$$\ln(2\theta)^{n_o} > \ln(k/\alpha) \implies (2\theta)^{n_o} > \frac{k}{\alpha} > \frac{k(1-\theta)}{\alpha}$$

$$\implies \alpha\frac{\theta^{n_o}}{(1-\theta)} > \frac{k}{2^{n_o}} = kd_m(x,y)$$

But notice that:

$$d_s(x,y) = \sum_{n=n_o}^{\infty} \theta^n \rho(x_n, y_n) = \rho(0,1)\sum_{n=n_o}^{\infty} \theta^n = \rho(0,1)\frac{\theta^{n_o}}{1-\theta} > \alpha\frac{\theta^{n_o}}{1-\theta}$$

And we are done. In fact, we could have used $\rho(0,1)$ instead of using α.
For $\theta < 1/2$, we are proving that, for every k, there are x, y such that:

$$d_s(x,y) < \frac{1}{k}d_m(x,y)$$

Given k, take x with $x_n = 0$ for every n and y with $y_n = 0$ for every $n \neq n_o$ and $y_{n_o} = 1$, with:

$$n_o > -\frac{\log k\alpha}{\log 2\theta}$$

Note that, since $\theta < 1/2$, the denominator is negative. We have that:

$$\frac{1}{k}d_m(x,y) = \frac{1}{k2^{n_o}}$$

And $d_s(x,y) = \alpha\theta^{n_o}$. But, by the choice of n_o, and once $\log 2\theta < 0$:

$$\log(2\theta)^{n_o} < -\log(k\alpha) \implies (2\theta)^{n_o} < \frac{1}{k\alpha} \implies d_s(x,y) = \alpha\theta^{n_o} < \frac{1}{k2^{n_o}} = \frac{1}{k}d_m(x,y)$$

Where, in this case α is necessarily $\rho(0,1)$. ■

Corollary 6.5.1. *The minimum metric can be slightly modified by replacing $1/2$ by ϕ, so that the distance would be $\phi^{\min n, x_n \neq y_n}$. In this scenario, the metrics would not be uniformly equivalent provided $\theta \neq \phi$*

Proof. Just repeat the proof of the proposition, replacing $1/2$ by ϕ. The modification would yield $\log(\theta/\phi)$ instead and the conclusion would follow. ■

We can easily adapt the proof for transitive shifts with at least one fixed point. The proof for general transitive shifts may be more delicate.

Proposition 6.6. *Let (Σ, σ) be a transitive shift with at least one fixed point. Then, the metrics d_m, of the minimum and $d_s = d_{s,\rho,\theta}$ are not uniformly equivalent if $(\mathbb{N}, \rho)$ is discrete and $\theta \neq 1/2$.*

Proof. Without loss of generality, we can assume that the fixed point is $x = 0...0....$ Start assuming $\theta > 1/2$. Now, since the shift space is transitive, there are letters connecting 0 with 1 and 1 with 0. We can also assume that there are no 0 between them, otherwise we could take a shorter path. We are going to represent the letters connecting 0 and 1 with $\sim$. Then, there is a allowed word, $0 \sim 1 \sim 0$, with no other 0 than those in the ends. Let P be the number of letters between the zeros such that the word is admissible. Notice that we are not requiring P to be as small as possible.

Now, let y be a sequence such that $y_n = 0$ for $n \leq n_o - 1$ and, for $n \geq n_o$, y_n repeats the pattern $0 \sim 1 \sim 0 \sim 1 \sim 0....$ It is clear that y is in the shift space. As an example, think of $y = 0000432156704321567 0....$ Notice that $n_0 = 4$, the repeating pattern is 043215670, or $0 - 432 - 1 - 567 - 0$ (for better visualizing) and $P = 7$.

The rest of the proof will be focused in showing that there is always a n_0 that make the distance in one metric greater than in the other. It is clear that:

$$d_m(x,y) = \frac{1}{2^{n_0+1}}$$

Furthermore, it is straightforward to show that:

$$d_s(x,y) = \sum_{n=n_0+1}^{n_0+P} \theta^n \rho(x_n,y_n) + \sum_{n=n_0+P+2}^{n_0+2P+1} \theta^n \rho(x_n,y_n) + \sum_{n=n_0+2P+3}^{n_0+3P+2} \theta^n \rho(x_n,y_n) + \cdots$$

And the distance takes the following form:

$$d_s(x,y) = \sum_{j=1}^{\infty} \left[\sum_{n=n_o+(j-1)P+j}^{n_o+jP+(j-1)} \theta^n \rho(x_n,y_n) \right]$$

Although the series above seem a bit cumbersome, we will be able to estimate a lower bound for it by means of α, using the same notation of the previous propositions. We have always that:

$$\sum_{n=n_o+(j-1)P+j}^{n_o+jP+(j-1)} \theta^n \rho(x_n,y_n) \geq \alpha \sum_{n=n_o+(j-1)P+j}^{n_o+jP+(j-1)} \theta^n = \alpha \theta^{n_o+(j-1)P+j} \frac{1-\theta^{n_o+jP+(j-1)}}{1-\theta}$$

.

Hence:

$$d_s(x,y) \geq \frac{\alpha\theta^{n_0}}{(1-\theta)\theta^P} \sum_j \theta^{(P+1)j} - \theta^{n_o-1+2jP+2j}$$

The series above can be broken in two, and we get:

$$d_s(x,y) \geq \frac{\alpha\theta^{n_0}}{(1-\theta)\theta^P} \left(\frac{\theta^{P+1}}{1-\theta^{P+1}} - \theta^{n_o-1} \frac{\theta^{2P+2}}{1-\theta^{2P+2}} \right)$$

$$\implies d_s(x,y) \geq \frac{\alpha\theta}{1-\theta} \left(\frac{\theta^{n_o}}{1-\theta^{P+1}} - \frac{\theta^{2n_o+P}}{1-\theta^{2P+2}} \right)$$

Now, we have to show that, for every k, there is a n_o such that

$$\frac{\alpha\theta}{(1-\theta)} \left(\frac{\theta^{n_o}}{1-\theta^{P+1}} - \frac{\theta^{2n_o+P}}{1-\theta^{2P+2}} \right) \geq \frac{k}{2^{n_o+1}}$$

Or, what is equivalent:

$$2^{n_o} \left(\frac{\theta^{n_o}}{1-\theta^{P+1}} - \frac{\theta^{2n_o+P}}{1-\theta^{2P+2}} \right) \geq \frac{k(1-\theta)}{2\alpha\theta}$$

$$\iff 2^{n_o}(B\theta^{n_o} - A\theta^{2n_o}) \geq \frac{k(1-\theta)}{2\alpha\theta} AB$$

where $A = 1-\theta^{P+1}$ and $B = (1-\theta^{2P+2})/\theta^P$

But this is true, because, for $\frac{1}{2} < \theta < 1$, we have that $\lim_{x\to\infty} f(x) = \infty$ for the function:

$$f(x) = 2^x(B\theta^x - A\theta^{2x})$$

for every $A, B > 0$.

Now, assume that $\theta < 1/2$.Take $x = 000...$ and $y = 000... \sim 1 \sim 000...$, where the first letter different of 0 is in the n_o position. Take $M = \max\{\rho(x_n,y_n), n_o \leq n \leq n_o + P\}$. Then:

$$d_s(x,y) = \sum_{n=n_o}^{n_o+P} \theta^n \rho(x_n, 0) \leq M \sum_{n=n_o}^{n_o+P} \theta^n = K\theta^{n_o} \frac{1-\theta^{n_o+P}}{1-\theta}$$

.

We must prove, that, for every k, there is n_o such that:

$$d_s(x,y) \leq K\theta^{n_o} \frac{1-\theta^{n_o+P}}{1-\theta} \leq \frac{1}{k2^{n_o}} = \frac{1}{k} d_m(x,y)$$

It is enough to prove that:

$$(2\theta)^{n_o}(1-\theta^{n_o+P}) < \frac{1-\theta}{Mk}$$

But clearly the limit of the right member is zero, and the proof is complete. ■

Proposition 6.7. *Let (Σ, σ) be a (countable) Markov shift. Then, the minimum metric and the summation metric are uniformly equivalent if (i) $\theta = \phi$ (denoting the coefficients of the metrics) and (ii) ρ is of type 1. Moreover, if Σ is the full shift, the reciprocal is true.*

Proof. Take $m \leq (1-\theta)/K$, where K is an upper bound for ρ, that exists for hypothesis. Given $x, y \in \Sigma$, denote by n_o the smallest n such that $x_n \neq y_n$. Then, we have that:

$$md_s(x,y) = m \sum_{n \geq n_o} \theta^n \rho(x_n, y_n) \leq mK \sum_{n \geq n_o} \theta^n = mK \frac{\theta^{n_o}}{1-\theta} \leq \theta^{n_o} = d_m(x,y)$$

Now, take $M \geq 1/\beta$, where β is the infimum of $\rho(m,n)$ with $m \neq n$. By hypothesis, $\beta > 0$, and M is well defined. Then:

$$Md_s(x,y) = M \sum_{n \geq n_o} \theta^n \rho(x_n, y_n) > M\beta\theta^{n_o} \geq \theta^{n_o} = d_m(x,y)$$

Reciprocally, we already know that, in the full shift, $\theta \neq \phi$ implies that the metrics are not uniformly equivalent, then uniformly equivalence implies $\theta = \phi$. Now, suppose that ρ is unbounded. Than, there are sequences a_k and b_k such that $\rho(a_k, b_k) > k$, for every k. Take $x_k = a_k000...$ and $y_k = b_k000...$. Then $d_m(x_k, y_k)$ is always 1, but $d_s(x_k, y_k)$ is unbounded. Thus, there is no constant m such that $md_s(x,y) < d_m(x,y)$ for every x, y.

Similarly, if d is not of type 1, than there are sequences a_k and b_k such that $\rho(a_k, b_k) < 1/k$, for every k. Taking the same x_k and y_k yields the same distance with d_m, but $d_s(x_k, y_k)$ goes to zero. Thus, there is no constant M such that $d_m(x,y) < Md_s(x,y)$ for every x, y. ■

6.3 Completion

Let $\overline{V}$ be the completion of the set of vertices V and $\partial V := \overline{V} \backslash V$. We will also consider the completion of Σ, $\overline{\Sigma}$ and $\partial\Sigma$ defined analogously. Notice that the boundary of both sets is never empty. The following proposition assures that one can write the elements of $\overline{\Sigma}$ as words composed by symbols of $\overline{V}$, in a similar fashion as Σ.

Proposition 6.8. $\overline{\Sigma} \subset (\overline{V})^{\mathbb{N}}$, *i. e., we can also write the points of the completion $\overline{\Sigma}$ as sequence of letters, provided we can use letters of $\overline{V}$ either.*

Proof. Once $\overline{V}$ is complete, and by the form of the metric defined in $(\overline{V})^{\mathbb{N}}$, it is possible to use standard tools of analysis to show that the later space is also complete. But is obvious that $\Sigma \subset (\overline{V})^{\mathbb{N}}$, and therefore the closures of Σ as well. Once closed sets of complete spaces are complete, the closure of Σ is complete and Σ is dense on it. Then we can identify the completion of Σ with this set, which finishes the proof ■

Lemma 6.9. *If $\rho : V \times V \to [0,1]$ is discrete and of vanishing type for a countable alphabet V, then ∂V is unitary.*

Proof. Using the vanishing property, we will prove that the sequence $x_i = i$ is a Cauchy sequence. In fact, let $\epsilon > 0$ be given. By hypothesis, there exists $N > 0$ such that $k \geq N \implies \sup_{n,m\geq k} \rho(x_n, x_m) < \epsilon$, so that for all $n, m \geq N$ we have (by taking $k = N$ as before) $\rho(x_n, x_m) \leq \sup_{n,m\geq k} \rho(x_n, x_m) < \epsilon$, as we wanted. Since $\overline{V}$ is complete, the sequence (x_i) must converge in $\overline{V}$, so call this limit point $\infty \in \overline{V}$. Of course, if it were the case that $\infty \in V$, then the sequence x_i would need to be eventually constant, since given the $\epsilon > 0$ that isolates ∞, there is $N > 0$ such that $n \geq N \implies \rho(x_n, \infty) < \epsilon$, so that for all $n \geq N$ we would have $n = x_n = \infty$, and absurd. Thus, we have necessarily $\infty \in \partial V$.

Now, take any element $x \in \partial V$ and remembering that the elements of $\overline{V}$ are equivalent classes of Cauchy sequences under the equivalence relation $(x_n) \sim (y_n) \iff \lim_{n\to+\infty} \rho(x_n, y_n) = 0$, then it is enough to show that for any other Cauchy sequence (x_n) of elements of V which is equivalent to an element $a \notin V$ in the completion $\overline{V}$, we have $\lim_{n\to+\infty} \rho(x_n, y_n) = 0$, where $y_n = n$ for all $n \geq 1$, since this would imply that $(1, 2, ...)$ is precisely the equivalence class of the sequence (x_n), and thus $x = \infty$.

Now, if the sequence (x_n) had finitely many alphabet elements in it, it necessarily would be eventually constant. In fact, suppose for the sake or argument that $x_n = 1$ or $x_n = 2$ for all $n \geq 1$. Since it is a Cauchy sequence, there is $N > 0$ such that $m, n \geq N \implies \rho(x_m, x_n) < \rho(1, 2)$, and this is only possible if $\rho(x_m, x_n) = 0$, since either $\rho(x_m, x_n)$ equals 0 or $\rho(1, 2)$, so that $x_m = x_n$ and the sequence is eventually constant. Without loss of generality, suppose that the sequence is eventually constant to an element $a \in V$. Thus, the sequence (x_n) is equivalent to the point $a \in V$ in the completion $\overline{V}$, which is an absurd since we are assuming that $x \notin V$. Thus, the conclusion is that (x_n) has infinitely many alphabet points in itself, so we can extract a subsequence of (x_n), say x_{n_k} which is also a subsquence of $(n)_{n\geq 1}$. Since $n \to +\infty$ (the same infinity as before), then $x_{n_k} \to +\infty$. Since (x_n) is a Cauchy sequence, this means that the whole Cauchy sequence (x_n) must converge to ∞, so that in special $\lim_{n\to+\infty} \rho(x_n, n) = d((x_n), (n)) \leq d((x_n), (\infty)) + d((n), (\infty)) = 0$, as we wanted, where d is the metric in $\overline{V}$. ∎

Proposition 6.10. *A sequence $(x^k)_k$ in Σ is a Cauchy sequence if, and only if, $(x^k_n)_k$ is a Cauchy sequence for every n.*

Proof. This uses only some standard tools of analysis. ∎

With respect to σ, we already know that it is uniformly continuous. Hence, as Σ is dense in $\overline{\Sigma}$, there exists a continuous extension, that we will also denote by σ, from $\overline{\Sigma}$ to itself. A priori, we do not know anything about the extension, but the following result shows that the extension continues doing the same thing.

Proposition 6.11. *If $x \in \overline{\Sigma}$, $x = (x_0, x_1, ...)$, with $x_n \in \overline{V}$, then $\sigma(x) = (x_1, x_2, ...)$. In particular, $\sigma : \overline{\Sigma} \to \overline{\Sigma}$ is well-defined.*

Proof. As $x \in \overline{\Sigma}$, there is a Cauchy sequence (x^k), $x^k = (x^k_0, x^k_1, ...)$ converging to x. As the extension of σ is continuous, $\sigma(x) = \sigma(\lim_k x^k) = \lim_k \sigma(x_k) = \lim_k (x^k_1, x^k_2, ...)$. Since x^k is Cauchy, $(x^k_n)_k$ is Cauchy for every n, each converging to x_n. Then, $(x^k_1, x^k_2, ...)$ is Cauchy and converges to $(x_1, x_2, ...)$. We conclude that $\sigma(x) = (x_1, x_2, ...)$ ∎

Now, let's study some particular subsets of the boundary. It will be much more efficient to name which one of them. We will use the following notation: $\partial\overline{\Sigma}(a, b, c)$ with $a, b \in \mathbb{N} \cup \{\infty\}$ and

$c \in \{0, \infty\}$ to represent the set of all $x = (x_1, x_2, ...) \in \overline{\Sigma}$ such that the first a letters are in $\partial\mathbb{N}$, the next b letters are in $\mathbb{N}$ and the rest is filled with letters of $\partial\mathbb{N}$ if $c = \infty$ and, if $c = 0$, the rest of the word is filled with letters of $\mathbb{N}$. For example, we have that $(\infty, \infty, 1, 2, 3, \infty, \infty, \infty, ...) \in \partial\overline{\Sigma}(2, 3, \infty)$.

We will also use the notation $a \to b$ for $a, b \in \overline{\mathbb{N}}$ to denote the existence of a word $x = (x_1, x_2, x_3, ...)$ such that $x_1 = a$ and $x_2 = b$. We'll also say that $A \to B$ for $A, B \subset \overline{\mathbb{N}}$ if there are $a \in A$ and $b \in B$ such that $a \to b$. Although it is the same notation as that used to denote "a tends to b", we hope that it will be clear what is the meaning of the arrow in each situation.

The next lemma is one of the most important ones in this paper:

Lemma 6.12. *Let (Σ, σ) be a finite entropy countable Markov shift. Then:*

1. *There is no element $x = (x_0, x_1, ...) \in \overline{\Sigma}$ with $x_n = i \in V$, $x_n, ..., x_{n+m} \in \partial V$ and $x_{n+m+1} = j \in V$;*

2. *$\partial\overline{\Sigma}(N+1, \infty, 0) = \sigma^{-N}(\partial\overline{\Sigma}(1, \infty, 0))$;*

3. *For every $N, M < +\infty$, the sets $\partial\overline{\Sigma}(N, \infty, 0)$ are measurable. Moreover, if $\partial\overline{\Sigma}(0, N, \infty)$ and $\partial\overline{\Sigma}(N, M, \infty)$ are also measurable, then $\mu\left(\bigcup_{N=1}^{\infty} \partial\overline{\Sigma}(N, \infty, 0)\right) = \mu\left(\bigcup_{N=0, M=1}^{\infty} \partial\overline{\Sigma}(N, M, \infty)\right) = 0$, for every $\sigma-$invariant measure μ*

4. *$\partial V \to \partial V$ is always allowed;*

5. $$\partial\Sigma = \left[\bigsqcup_{N=1}^{\infty} \partial\Sigma(N, \infty, 0)\right] \sqcup \left[\bigsqcup_{N=0, M=1}^{\infty} \partial\Sigma(N, M, \infty)\right] \sqcup \partial\Sigma(0, 0, \infty)$$

6. *If μ is an ergodic $\sigma-$invariant measure, then $\mu(\partial\Sigma) = 0$ or 1*

Proof.

1. Suppose such a point x exists. Since this point is in the completion, there must exists a Cauchy sequence (y^k) in Σ converging to x. We will show that we have $y^k_n = i$ and $y^k_{n+m+1} = j$ for every $n > N$. Indeed, $(y^k_n)_k$ must be a Cauchy sequence converging to i. However, once V is discrete, the unique way to this sequence converge to $i \in V$ is $(y^k_n) = i$, for every $n > N_1$. The same happens with y^k_{n+m+1}, giving us a N_2 and we can take the maximum between N_1 and N_2. Therefore, there are infinite words of fixed length $m+1$ beginning in i and ending in j. Denote this words by w_k. Once the shift is topologically mixing, there is a word of length ℓ connecting j to i. By concatenation, we have infinite distinct words of length $m + \ell + 1$ connecting i to i, hence the entropy is infinite. Contradiction.

2. This is a simple corollary of the previous item. By proposition 6.11, $\sigma^{-N}(\partial\Sigma(1, \infty, 0))$ is simple the set of points such that $x_N \in \partial V$ and $x_n \in V$ for $n > N$. However, by the previous item, we cannot have $x_n \in V$ for $n < N$. Thus, if $x \in \sigma^{-N}(\partial\Sigma(1, \infty, 0))$, $x_n \in \partial V$, for all $n \leq N$ and $x_n \in V$ for $n > N$. But this is the same as saying that $x \in \partial\Sigma(N+1, \infty, 0)$.

3. We will first prove that $\partial\Sigma(1, \infty, 0)$ is measurable. Just notice that $\partial\Sigma(1, \infty, 0) = \sigma^{-1}(\Sigma)\backslash\Sigma$, purely by definition, that Σ is measurable and σ is measurable, once is continuous. The measurability of $\partial\Sigma^{\infty}_N$ follows from item (2) and again the fact that σ is measurable. Now, to prove that the measure of the sets are zero, we will use Poincaré recurrence theorem. Just notice that no point of $\partial\Sigma(N, \infty, 0)$ comes back to this set. Indeed, such a point has N letters from V, but $\sigma^n(x)$ has a number strictly lesser. By Poincaré recurrence theorem, we must have necessarily $\mu(\partial\Sigma(N, \infty, 0)) = 0$. Once N is arbitrary, the measure of the union is also zero. The argument is completely similar with the others equalities.

4. Suppose that $\partial V \to \partial V$ is not allowed. We'll show that there exists $N > 0$ such that $n, m > N \implies n \not\to m$. To show this, suppose that for all $N > 0$ there are $n, m > N$ such that $n \to m$ and we'll show that $\partial V \to \partial V$ is allowed. For $N = 1$, take $n_1, m_1 > 1$ such that $n_1 \to m_1$ and for $n \leq N_1$

   Now, by Corollary 2.1.1, there is $i \leq N$ such that $i \to j$, for infinitely many $j > N$.

5. One inclusion is obvious. So, take $x \in \partial\Sigma$. If $x_n \in \partial V$ for all n, x is in the third member. Then, let's assume that it is not the case. Denote by N the lesser index such that $x_N \in V$. Suppose that $N = 0$. We already know that $x_n \in \partial V$ for some n, otherwise $x \notin \partial\Sigma$. Let M be the lesser index such that $x_M \in \partial V$. We know that $x_n \in V$ for $0 \leq n \leq M - 1$ and $x_M \in \partial V$. However, by item (1), must be the case that $x_n \in \partial V$ for every $n \geq M$. Then $x \in \partial\Sigma(0, M, \infty)$. Otherwise, suppose that $N > 0$. Then $x_n \in \partial V$, for every $n < N$. If those are the unique letters in ∂V, then $x \in \partial\Sigma(N, \infty, 0)$, so let's assume that it is not. Let M be the lesser index greater than N such that $x_M \in \partial V$. Again, by item (1), we must have $x_n \in \partial V$, for every $n \geq M$. Then, $x \in \partial\Sigma(N, M, \infty)$. In any case, x is in some union.

6. By the very definition of ergodic measure, it is enough to show that $\partial\Sigma$ is invariant with respect to σ. However, this is not the case. Take, for example, $x \in \partial\Sigma(1, \infty, 0)$. More than that, we will prove that $(\partial\Sigma)\Delta(\sigma^{-1}(\partial\Sigma)) = \partial\Sigma(1, \infty, 0)$. First, notice that $\sigma^{-1}(\partial\Sigma) \subset \partial\Sigma$. Indeed, if x is in the first set, $\sigma(x) = (x_1, x_2, ...) \in \partial\Sigma$. Hence, $x_n \in \partial V$ for some $n > 0$. In particular, x has some letter in ∂V and therefore $x \in \partial\Sigma$. With that, it is clear that $(\partial\Sigma)\Delta(\sigma^{-1}(\partial\Sigma)) = (\partial\Sigma)\backslash(\sigma^{-1}(\partial\Sigma))$. One inclusion is direct, once $x \in \partial\Sigma(1, \infty, 0) \implies \sigma(x) \notin \partial\Sigma$. Now, take $x \in (\partial\Sigma)\backslash(\sigma^{-1}(\partial\Sigma))$. We will use item (5) to list the possibilities. Immediately we see that x cannot be in the second nor third member of the union. If this was the case, $\sigma(x)$ would be in $\partial\Sigma$, because we have infinitely many letters of ∂V. On the other hand, if $x \in \partial\Sigma(N, \infty, 0)$ with $N > 1$, then one clearly sees that $\sigma(x)$ would have at least one letter of ∂V in the beginning. The unique conclusion is that $x \in \partial\Sigma(1, \infty, 0)$. This fact, together with item (3), allow us conclude that $\mu((\partial\Sigma)\Delta(\sigma^{-1}(\partial\Sigma))) = 0$. As μ is ergodic, we get that $\mu(\partial\Sigma)$ is 0 or 1. ■

6.4 Basic Examples

Now, we will provide some concrete cases, exemplifying the results seen and the structure presented. In all the examples, we choose the metric to be vanishing. We will have then $\partial V = \{\infty\}$, as already proved. The entropy of the examples are always $\log 2$, as can be found in the literature.

Example 6.1 (Renewal Shift). *The Renewal Shift is defined by the transition matrix such that $a_{1,j} = 1$ for all $j \in \mathbb{N}$, $a_{i,i-1} = 1$ for all $i \geq 2$ and $a_{i,j} = 0$ otherwise. Consider the sequence of admissible words $x^k = (k, k-1, ..., 2, 1, 1, ...)$. This sequence converges to $(\infty, \infty, ...)$ and it follows that $\partial\Sigma(0, 0, \infty) = \{(\infty, \infty, ...)\}$, since we only have one infinity. On the word hand, words like $(x_0, ..., \infty, n, ...)$ are always prohibited for every $n \in \mathbb{N}$. In fact, if such a word was allowed, there would be a Cauchy sequence converging to it. By proposition 6.10, it would be necessary to have a sequence $x^k = (x^k_0, ..., x^k_i, n, ...)$, with $x^k_i \to \infty$. However, the unique allowed letters to be before n are 1 and $n+1$ and these letters can never approach ∞. Thus, $\partial\Sigma(N, \infty, 0) = \partial\Sigma(N, M, \infty) = \emptyset$, for $N, M > 0$. In a similar fashion, one can prove that $(x_0, .., n, \infty, ...)$ is allowed if, and only if $n = 1$, once 1 is the unique letter that can transfer to arbitrarily big letters. We conclude that $\partial\Sigma = \{(w, \infty, \infty, ...)\}$, where w is any admissible word finishing in 1 or nothing.*

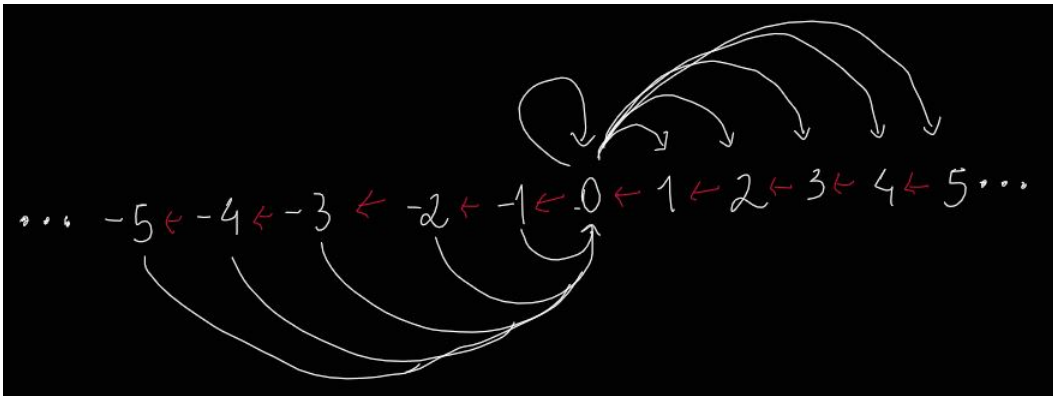
The next example does exactly the opposite of the last example. As we'll see, we shall have $n \not\to \infty$ for every $n \in \mathbb{N}$ and $\infty \to 1$.

Example 6.2 (Backwards Renewal Shift). *The Backwards Renewal Shift is defined by the transition matrix $A = (a_{i,j})$ such that $a_{i,1} = 1$, for all $i \geq 1$, $a_{i,i+1} = 1$ for all $i \geq 1$ and $a_{i,j} = 0$ if otherwise. Take the Cauchy sequence $x^k = (k, 1, 1, ...)$, which is allowed since any k connects with 1 and 1 is allowed to connect with itself. Obviously we have $x^k \to (\infty, 1, 1, ...)$, so $\infty \to 1$ is allowed. Now, suppose that there exists an allowed word of the form $(n, \infty, x_1, x_2, ...)$, with $x_n \in \overline{V}$. By item (1) of the last Lemma, we must have $x_n = \infty$, for every $n \geq 1$, so that there would be a Cauchy sequence of admissible words $x^k \to (n, \infty, \infty, ...)$, with entries in V. Now, the only way this is possible is if $x_i^k \to \infty$, for every $i \geq 2$. Since the only allowed words to be after n are $n+1$ or 1, we would have $x_2^k = n+1$ or $x_2^k = 1$, for every $k \geq 1$, but naturally this sequence can never reach ∞, so we conclude $n \not\to \infty$ as we wanted. Of course, the proof remains the same if we put a finite amount of letters of $\overline{V}$ before n in a coherent way.*

Now, we'll find an identification of $\partial\Sigma$. To do this, note that the Cauchy sequence $(k, k+1, k+2, ...)$, which converges to $(\infty, \infty, ...)$ and, since $\partial V = \{\infty\}$ (there is only one infinity), then $\partial\Sigma(0, 0, \infty) = \{(\infty, \infty, ...)\}$ as before. Following our proof in the last paragraph, we concluded that there cannot be any word of the form $(\infty, \infty, ..., \infty, x_1, x_2, ..., x_n, \infty, \infty, ...)$ with $x_i \in V$ for all $1 \leq i \leq n$, so that we have $\partial\Sigma(N, M, \infty) = \varnothing$, for all $N \geq 0$ and $M \geq 1$. By item (5) of our last Lemma, we then see that $\partial\Sigma = \{(\infty, \infty, ..., \infty, x_n, x_{n+1}, x_{n_2}, ...) : n \geq 2$ and $x_n \in V, \forall n \geq 2\} \sqcup (\infty, \infty, ...)$.

One natural question is wondering if it is really possible to have a word such as $(\infty, 1, \infty, ...) \in \partial\Sigma$. The item (1) of the lemma maybe induces us to think that it is awkward to happen something like this. However, this is really possible, as it will be seen in the next example.

Example 6.3 (Synthetic Renewal Shift). *This example can be seen as a Renewal Shift with "two arms", one being a backwards renewal. It will be more convenient to take $V = \mathbb{Z}$. The transition matrix is such that $A(0, i) = A(-i, 0) = 1$, for every $i \geq 0$, $A(i-1, i) = 1$ and $A(i, j) = 0$ otherwise. While is not such difficult to see that this shift is topologically mixing, it is not so easy to show that its entropy is finite. More specifically, we have $h_{top}(\sigma) = \log(\sqrt{2} + 1)$. Take the sequence $x^k = (-n, 0, n, n-1, ..., 1, 0, 0, ...)$. The limit of this sequence is exactly $(\infty, 0, \infty, \infty, ...)$.. This shift can be seen using the following image.*



Now, we are going to prove, using lemma 5.8 that the entropy of this example is finite. Using $a = 0$, we have that $B_1 = 1$ and $B_k = 2$ for $k > 1$, so:

$$f(x) = x + 2x^2 + 2x^3 + 2x^4 + ...$$

It is straightforward to see that $\limsup \sqrt[n]{B_n} = 1$, so f is well-defined at least in $(-1, 1)$. Remembering that $x + x^2 + x^3 + ... = x/(1-x)$, we have:

$$f(x) = x + 2x\frac{x}{1-x} = \frac{x + x^2}{1-x}$$

$$\implies \frac{f(x)}{1-f(x)} = \frac{x+x^2}{1-2x-x^2}$$

And the job from now on will be to calculate:

$$\frac{d^n}{dx^n}\left(\frac{x+x^2}{1-2x-x^2}\right)\bigg|_{x=0} = \frac{d^n}{dx^n}\left(\frac{x}{1-2x-x^2}\right)\bigg|_{x=0} + \frac{d^n}{dx^n}\left(\frac{x^2}{1-2x-x^2}\right)\bigg|_{x=0}$$

This will be accomplished by means of the expression for the n-th derivative of a product of functions, presented in 5.8 as well. We get, then:

$$\frac{d^n}{dx^n}\left(\frac{x}{1-2x-x^2}\right)\bigg|_{x=0} = n\frac{d^{n-1}}{dx^{n-1}}\left(\frac{1}{1-2x-x^2}\right)\bigg|_{x=0}$$

$$\frac{d^n}{dx^n}\left(\frac{x^2}{1-2x-x^2}\right)\bigg|_{x=0} = n(n-1)\frac{d^{n-2}}{dx^{n-2}}\left(\frac{1}{1-2x-x^2}\right)\bigg|_{x=0}$$

There derivatives are readily evaluate once we take knowledge that:

$$\frac{1}{1-2x-x^2} = \frac{1}{2\sqrt{2}}\left[\frac{1}{\sqrt{2}-1-x} - \frac{1}{-1-\sqrt{2}-x}\right]$$

And:

$$\frac{d^k}{dx^k}\left(\frac{1}{p-x}\right)\bigg|_{x=0} = k!p^{-(k+1)}$$

We get:

$$\frac{d^k}{dx^k}\left(\frac{1}{1-2x-x^2}\right)\bigg|_{x=0} = \frac{k!}{2\sqrt{2}}\left[(\sqrt{2}-1)^{-(k+1)} - (-\sqrt{2}-1)^{-(k+1)}\right]$$

So the number of points of period n starting in 0 is:

$$\frac{1}{2\sqrt{2}}\left[(\sqrt{2}-1)^{-n} + (\sqrt{2}-1)^{-n+1} - (-\sqrt{2}-1)^{-n} - (-\sqrt{2}-1)^{-n+1}\right]$$

Although awkward, this expression gives only integers provided n is integer. Notice that the dominant term is the second one, so the following is an upper bound for the entropy:

$$\lim_n \frac{1}{n}\log\frac{(\sqrt{2}-1)^{-n+1}}{2\sqrt{2}} = \lim_n \frac{n-1}{n}\log\left(\frac{1}{\sqrt{2}-1}\right) = \log\left(\sqrt{2}+1\right)$$

Moreover, it can be shown that, rather than an upper bound, this is the actual value of the entropy.

6.5 Interior Rich

To finish this section, we are going to present a sufficient condition for the pressure being equal in the original and completed spaces.

Definition 6.2. Let (Σ, σ) be a finite entropy, topologically mixing CMS. We say that it is *interior rich* for a totally bounded ρ if, for any $\mu \in \mathcal{M}_{\partial\Sigma}$, $\phi \in C(\overline{\Sigma})$ and $\epsilon > 0$ there is $\mu' \in \mathcal{M}_\Sigma$ such that $h(\mu') > h(\mu) - \epsilon$ and:

$$\int \phi d\mu' > \int \phi d\mu - \epsilon$$

Proposition 6.13. *A topologically mixing CMS* (Σ, σ) *of finite entropy is interior rich if, and only if, the two conditions hold:*

$$\sup_{\mu' \in \mathcal{M}_\Sigma} h(\mu') \geq \sup_{\mu \in \mathcal{M}_{\partial\Sigma}} h(\mu)$$

$$\sup_{\mu' \in \mathcal{M}_\Sigma} \int \phi d\mu' \geq \sup_{\mu \in \mathcal{M}_{\partial\Sigma}} \int \phi d\mu, \forall \phi \in \mathcal{C}(\overline{\Sigma})$$

Proof. It is clear. ■

Lemma 6.14. *If* (Σ, σ) *is interior rich, then* $P_\Sigma(\phi) = P_{\overline{\Sigma}}(\phi)$ *for every uniformly continuous* ϕ.

Proof. First of all, we will need the equality $\mathcal{M}(\overline{\Sigma}) = \mathcal{M}(\Sigma) \sqcup \mathcal{M}(\partial\Sigma)$, where we assume further that the invariant measures in those sets are ergodic. This can be proved as follows: if μ is an ergodic measure in $\mathcal{B}(\overline{\Sigma})$, then by Lemma 6.12 we either have $\mu(\partial\Sigma) = 1$ or $\mu(\partial\Sigma) = 0$. If it is the first case, then necessarily $\mu(\Sigma) = 0$, so this measure is in $\mathcal{M}(\partial\Sigma)$. If not, then $\mu(\partial\Sigma) = 0$ so that the measure is in $\mathcal{M}(\Sigma)$. Of course, the union is disjoint, since if for instance $\mu \in \mathcal{M}(\Sigma)$, then as μ is a probability measure, we have $\mu(\Sigma) = 1$, so in special $\mu(\partial\Sigma) = 0$, so it cannot be in the former set.

Now, by [Wal82], Corollary 9.10.1 we can calculate the pressure in $\overline{\Sigma}$ as follows:

$$P_{\overline{\Sigma}}(\varphi) = \sup \left\{ h(\mu) + \int \varphi d\mu : \mu \in \mathcal{M}(\overline{\Sigma}) \text{ and } \mu \text{ is ergodic} \right\}$$

Now, if A, B are subsets of the extended real numbers (not necessarily bounded), then we have $\sup(A \cup B) = \max\{\sup A, \sup B\}$ (and this is not necessarily finite). In special, by setting:

$$A = \left\{ h(\mu) + \int \varphi d\mu : \mu \in \mathcal{M}(\Sigma) \text{ and } \mu \text{ is ergodic} \right\}$$

$$B = \left\{ h(\mu) + \int \varphi d\mu : \mu \in \mathcal{M}(\partial\Sigma) \text{ and } \mu \text{ is ergodic} \right\}$$

$$\implies \left\{ h(\mu) + \int \varphi d\mu : \mu \in \mathcal{M}(\overline{\Sigma}) \text{ and } \mu \text{ is ergodic} \right\} = A \cup B$$

We thus have $P_{\overline{\Sigma}}(\varphi) = \sup(A \cup B) = \max\{\sup A, \sup B\}$. Now, by the interior rich property we obviously have $\sup_{\mu' \in \mathcal{M}_\Sigma} \left(h(\mu') + \int \phi d\mu' \right) \geq \sup_{\mu \in \mathcal{M}_{\partial\Sigma}} \left(h(\mu) + \int \phi d\mu \right)$, so that $\sup A \geq \sup B$ and in special $\max\{\sup A, \sup B\} = \sup A$, so finally we have $P_{\overline{\Sigma}}(\varphi) = \sup A = P_\Sigma(\varphi)$.

■

7 Sectorially Arranged

In this section we'll define the kind of structure on the set of vertices which makes the main theorem true. Before stating it, we will need some terminology. More specifically, clarify some aspects about connectedness of graphs.

First, we will say that a set $A \subset \overline{V}$ of vertices is connected if, given two vertices $a_1, a_2 \in A$, there is a path between them contained inside A. In this case, the shift must be seen as a non-oriented graph, i.e. the path may not respect the direction of some arrows. This is a standard definition in the theory of non-oriented graphs. On the other hand, we say that a subset A is directionally connected if there is a *directed* path between them contained inside A. In this case, the shift must be seen as oriented and this definition is not standard. Finally, we will say that a subset is strongly connected if it satisfies the standard definition in the theory of oriented graphs, namely, for each pair of vertices, we have two path between them, each in one direction. The illustration below makes this three conceptions clearer. Notice that the first and the third are equivalence relations, whereas the second is not. This fact makes easier to treat the first and the third.

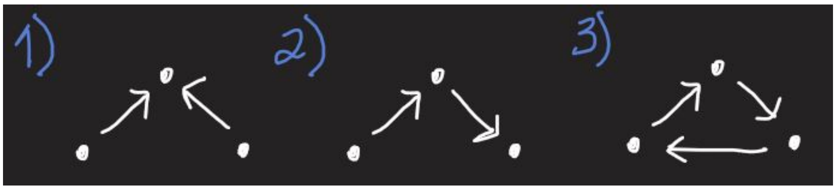


Figure 1: 1) Connected but not directionally nor strongly connected. 2) Directionally connected but not strongly 3) Strongly Connected

Notice that strongly connected implies directionally connected, which implies connected.

We will also say that two set of vertices A and B are disconnected if their union is not connected. This means that there are $a \in A$ and $b \in B$ such that, if there is a path between them, then this path goes necessarily out of $A \cup B$. We define analogously directionally disconnected and strongly disconnected. The direction of the implications is then reversed: disconnected implies directionally disconnected, which implies strongly disconnected.

We can distinguish this from the stricter definition of two totally disconnected sets, which means that there is no path (oriented or not, depending on the underlying meaning of connected) contained in the union between *any* pair of points. In two of the three cases, this definition does not add anything, as we will observe.

Proposition 7.1. *Two connected subsets A, B are disconnected if, and only if, they are totally directionally disconnected.*

Proof. By definition, they are disconnected if there are $a \in A$ and $b \in B$ such that a and b cannot be connected by a non-oriented path in $A \cup B$. However, this is true if, and only if, there is no path in $A \cup B$ connecting any pair of points, each in one set. This happens because connectedness is transitive — if $x \in A$ and $y \in B$ are connected, we can concatenate path and go from a to x to y to b, since each subset is connected. Thus, we have so far that two connected subsets are disconnected if, and only, if, there are no path connecting some pair of points. But this means, by definition, that the two sets are totally directionally disconnected. ■

Furthermore, again only by transitivity, disconnected is equivalent to totally disconnected and strongly disconnected is equivalent to totally strongly disconnected. Hence, from now on, we will just use the terminology "totally disconnected" to represent "totally directionally disconnected".

Now, onto the structure:

Definition 7.1. We say that the vertex set $V = \mathbb{N}$ is sectorially arranged, with respect to the metric ρ in V, if there are sequences $N_k \to \infty$, $\delta_k \to 0$ and

$$V = \{1, ..., N_k\} \cup \bigsqcup_{i=1}^{p_k} V_k^i$$

for $p_k \in \mathbb{N} \cup \{\infty\}$, where each sector V_k^i is infinite and connected with diam $V_k^i < \delta_k$; for a given k, the V_k^i are not connected to each other; and for $k \geq 2$, $V_k^i \subset V_{k-1}^{i'}$.

Note that it is not forbidden for some of the $1, 2, ..., N_k$ to be inside some sector V_k^i. Nonetheless, if there is two sectors or more, some element must be outside every sector. Indeed, by topologically mixing property, one can connect two elements from different sector, but this must be with some element outside every sector, otherwise the sectors would not be disconnected.

A good way to think about the definition is that we can "almost cover" (almost meaning there can be a finite set outside the cover) the set of vertices with sets V_k^i called sectors (here, k represents the "sectorization", which is just a set of sectors, and the indices i represent each sector of a given sectorization k), which get smaller (in the sense that their diameter uniformly tends to zero and the N_k tends to cover the whole space) after each iteration, are disjoint and not connected with each other, each set however being connected and each sectorization refining the last one as the iteration of k gets bigger.

One of the most important features of a sectorially arranged alphabet, that will be used exaustively, is the following:

Lemma 7.2. *Given a sector V_k^i and n natural, we can find an admissible word of length n with elements of V_k^i, provided Σ has finite entropy. If Σ is allowed to have infinite entropy, we can find a concrete case where this not happens.*

Proof. The proof will be made by induction. It is obvious that we always can find words of length 1 and 2 in any sector. But we will now prove that, for every finite set F, one can find an admissible word of length 2 with vertices in $V_k^i \backslash F$. Indeed, suppose it was not the case. Then, the incoming and outgoing arrows of every point of $V \backslash F$, connect them to points in $F \cup \{1, ..., N_k\}$, using the fact that the sectors are disconnected. As $F \cup \{1, ..., N_k\}$ is finite and $V \backslash F$ infinite, by the pigeonhole principle, there must exist a sequence $v_n \in V \backslash F$ such that av_nb is admissible and $a, b \in F$, giving rise to infinite periodic points (of the same period).

Now, let's suppose that, for every finite set F, we can find an admissible word of length ℓ in $V_k^i \backslash F$ and we want to prove that the same holds for words of length $\ell + 1$. This proof will be by contradiction, that is, we will suppose there is a finite set F_0 such that is impossible to find a word of length $\ell + 1$ with letters in $V \backslash F_0$. If it was the case, then the word of length ℓ we can find by hypothesis, w_0, can only connect its endings to letters of $F_0 \cup \{1, ..., N_k\}$. More precisely, there must exist letters u_0 and v_0 in that union such that $u_0w_0v_0$ is allowed. Now, take F_1 to be F_0 joined with the letters of w_0. Again, we can find a word w_1 of length ℓ in $V \backslash F_1$. However, if we were not able to find a word of length $\ell + 1$ in $V \backslash F_0$, we have no hope to find in $V \backslash F_1$. Then, there are necessarily, letters u_1 and v_1 in $F_1 \cup \{1, ..., N_k\}$. However, we cannot have some of these letters in w_0, otherwise we would be able to find a word of length $\ell + 1$ in $V \backslash F_0$. We conclude that these letters are in $F_0 \cup \{1, ..., N_k\}$. Defining recursively w_n, F_n and u_n, v_n, in the end of the day, we can find a sequence of words of the form $u_nw_nv_n$, $u_n, v_n \in F_0 \cup \{1, ..., N_k\}$. As the set of possible u_n and v_n are finite but we have infinitely many distinct w_n, there must be infinite points of the form uw_nv, with u, v fix, giving rise to infinitely many periodic points of the same period, which is a contradiction.

For the concrete example, look at example 7.1.

■

Also, we have:

Proposition 7.3. *Given a sector V_k^i, and letters $v_1, v_2 \in V_k^i$ there are letters $u_1, u_2 \in \{1, ..., N_k\}$ such that one can connect u_1 to v_1 and v_2 to u_2 using just letters from the sector.*

Proof. This follows directly from the fact that the shift is topologically mixing and the sectors are disconnected. Indeed, we already know that some element of $\{1, ..., N_k\}$ must not be in any sector. Without loss of generality, let's suppose this element is 1. There must exist some word connecting 1 to v_1. Let u_1 be the letter in this word before the first letter contained in the sector. Using the fact that the sectors must be disconnected, we must have $u_1 \in \{1, ..., N_k\}$ and every letter after that must be in the sector for the same reason. The second part is done analogously. ■

Corollary 7.3.1. *If V admits some sectorization k with infinite sectors (and Σ is topologically mixing), then there is some element of $\{1, ..., N_k\}$ with infinitely many incoming arrows and other element (possibly the same) with infinitely many outgoing arrows.*

Proof. From the proposition, it follows that, for every sector V_k^i of the sectorization, there must exist a letter v_i such that $v_i \to n_i$ to some letter $n_i \in \{1, ..., N_k\}$. As the set is finite, by the pigeonhole principle, some of these letters have infinitely many incoming arrows. Similarly one can prove that some letter has infinitely many outgoing arrows. ∎

Corollary 7.3.2. *V can only admit some sectorization with infinite sectors if Σ is not locally compact.*

Now, we will explore some properties of this structure. In what follows, we attempt to answer the question: How is the boundary of each sector? The following proposition shows that the boundary is never empty and, more than that, tends to be unitary as we refine the sectorization.

Proposition 7.4. *Let $V_1^{i_1} \subset ... \subset V_k^{i_k}$ be a nested sequence of sectors. Then, $\bigcap_k \overline{V_k^{i_k}}$ is a single point in ∂V.*

Proof. Since the metric ρ is totally bounded, we have that $\overline{V}$ is complete and since each set $\overline{V}_k^{i_k}$ is closed, the sequence is nested and $\mathrm{diam}(V_k^{i_k}) \to 0$, the nested box theorem for complete metric spaces makes us conclude that there is a unique point $x \in \cap_k \overline{V_k^{i_k}} \subset \overline{V}$.

Now, we are going to prove that this point cannot be in V. By contradiction, suppose there is $x \in V$ such that $x \in \overline{V_k^{i_k}}$ for all k. Once the sectors have infinitely many points, for each k we have $y_k \neq x$ with $y_k \in V_k^{i_k}$. As the set is discrete, there is a $\epsilon > 0$ such that $d(x, y_k) > \epsilon$ for all k. But this implies that $diam(V_k^{i_k}) > \epsilon$ for all k, and this implies the contradiction. ∎

Lemma 7.5. *Suppose that V is sectorially arranged and take any $v_\infty, v'_\infty \in \partial V$. Then $v_\infty \to v'_\infty$ if, and only if, $v_\infty = v'_\infty$. More than that, $(v_\infty, v_\infty, ...)$ is always allowed.*

Proof. We are going to separate in two cases. In the first case, we suppose that $v_\infty \in \partial V_k^{i(k)}$ for infinitely many k. That happens, for example, whenever the number of sector is finite, but it is not the general case, as example 7.4 shows below. If this is the case, we proceed as follows. By lemma 7.2, for every k, we can find an admissible word with k first letters in V_k^i. Since the diameter of the sectors tends uniformly to zero, it is clear that all the letters converges to v_∞. More formally, we have that $\lim_k \rho(x_i^k, x_j^k) \leq \lim_k \mathrm{diam} V_k^i = 0$, so that (x_i^k) and (x_j^k) are in the same equivalence class in the completion of V, i.e, they are equal in the completion. Hence, we were able to find a sequence converging to the fixed point $(v_\infty, v_\infty, ...)$, and, in particular, the first implication is proved.

The second case remains under construction.

Reciprocally, suppose that $v_\infty \to v'_\infty$ are elements of ∂V and we have to prove that $v_\infty = v'_\infty$. By hypothesis, we can take Cauchy sequences of letters $(x_0^n)_n \to v_\infty$ and $(x_1^n)_n \to v'_\infty$ such that $x_0^n \to x_1^n$ for every $n \geq 1$. Without loss of generality, we can suppose that these sequences are of distinct elements (they cannot be of finite elements since they converge to some element of ∂V, so one can just take a converging subsequence throwing away the elements that repeat themselves). Now, fix some $k \in \mathbb{N}$ and take n_k big enough such that both $x_0^{n_k}$ and $x_1^{n_k}$ are in some (possibly different) sectors (such a number has to exists, since if not, then for all n_k such that $x_0^{n_k}$ is in some sector, we would have $x_1^{n_k}$ in the finite set $\{1, 2, ..., N_k\}$. Since there are infinitely many n_k's such that $x_0^{n_k}$ is in the sectors, this would imply that there are infinitely many elements in that finite set, since the sequence (x_1^k) is of distinct elements, a contradiction). But it turns out that, if $x_0^{n_k}$ and $x_1^{n_k}$ were in different sectors, $V_k^{i_1}$ and $V_k^{i_2}$, these sectors would be connected, contradiction.

So far we've extracted subsequences $x_0^{n_k}$ and $x_1^{n_k}$ which are in the same sector $V_k^{n_k}$ for every $k \geq 1$ and each converging to it's respective infinity point v_∞ or v'_∞. By the same argument in the last item, the equivalence class of the sequence $(x_0^{n_k})$, represented by v_∞, is the same of the sequence $(x_1^{n_k})$, represented by v'_∞, as the diameter of the sector goes to zero. so that $v_\infty = v'_\infty$. ∎

Corollary 7.5.1. *If we substitute "connected" by "directionally connected but changes the interpretation of "directionally disconnected" to "totally disconnected", the previous lemma remains true, with exception of the existence of a fixed point for every point in ∂V. It is not sufficient, however, if we not require the sectores to be* totally *disconnected.*

Proof. The first part is obvious, once we will use the hypothesis of connected sectors just to obtain sequences such that $x_n \to y_n$, and is clear that this is always possible.

To the second part, however, we had used the fact that the sectors are disconnected, and directionally disconnected does not implies in disconnected. In fact, example 7.2 have sectors that are directionally disconnected but not disconnected, and the thesis do not hold. The substitution of directionally disconnected by totally disconnected works because it is equivalent to disconnected, as it has been shown. ■

Lastly, we are going to study the ergodic measures in the boundary of shift whose alphabet is sectorially arranged.

Lemma 7.6. *If μ is an ergodic measure supported on $\partial\Sigma$ and V is sectorially arranged, then there exists a fixed point $x \in \partial\Sigma$ such that $\mu = \delta_x$. In other words, deltas on fixed points are the unique ergodic measures on the boundary*

Proof. By lemma 6.12 we already know that, for every $\sigma-$invariant measure μ supported in the boundary, $\mu(\partial\Sigma(0,0,\infty)) = 1$. By lemma 7.5, we know that $\partial\Sigma(0,0,\infty) = \partial V^{\mathbb{N}} = F$, the set of fixed points. Thus, $\mu(F) = 1$. Every measurable subset of F is invariant. Indeed, $(\sigma^{-1}(E))\Delta E = (\sigma^{-1}(E))\backslash E \subset \partial\Sigma(0,1,\infty)$, for $E \subset F$. Then, $\mu((\sigma^{-1}(E))\Delta E) = 0$

Now, for n natural, take a cover of F by balls of radius $1/n$. Once ρ is totally bounded, we can always take finitely many. Consider the finite collection of sets formed disjuncting these balls. Once every set of these collection is invariant, as shown, if μ is ergodic, then one and only one of then have measure 1. Let E_n be this set. Again, we can cover $\overline{E}_n$ with finitely many balls of radius $n+1$. Taking $(\overline{E}_n)_n$ recursively we end up with a sequence of nested closed measurable sets whose diameter tends to zero. Then:

$$\mu\left(\bigcap_n \overline{E}_n\right) = \lim_n \mu(\overline{E}_n) = 1$$

But $\bigcap_n \overline{E_n} = \{x\}$, for some x, then $\mu(\{x\}) = 1$. Further, $x \in F$, otherwise $\mu(\{x\}) = 0$. Hence, μ is a delta supported in some fixed point. ■

Corollary 7.6.1. *The entropy of the boundary is zero.*

Proof. By the lemma, every ergodic measure is supported on a periodic orbit. Thus, its entropy is always zero. As the topological entropy is the supremum of the entropy of ergodic measures, it is also zero. ■

7.1 Main Theorem

Lemma 7.7. *If V is sectorially arranged and ρ is a discrete totally bounded metric, then Σ is interior rich.*

Proof. (i) Since the entropy is always non-negative, and since the entropy in the boundary is always zero, we have automatically $h(\nu) > h(\mu') - \epsilon$ for every $\epsilon > 0$,$\mu' \in \mathcal{M}(\partial\Sigma)$ and $\nu \in \mathcal{M}(\Sigma)$, so we only have to check the second inequality of the definition.

More specifically, by lemma 7.6, $\mu = \delta_{x_v}$, for some $x_v = (v_\infty, v_\infty, ...) \in \partial\Sigma$, where $v_\infty \in \partial V$. Then, given $\epsilon > 0$, $x_v \in \partial\Sigma$ and $\varphi \in UC_d(\Sigma)$, it is enough to find a measure $\nu \in \mathcal{M}(\Sigma)$ such that

$\left|\int \varphi d\mu - \int \varphi d\nu\right| < \epsilon$. The rest of the proof will be concerned in finding such a measure.

The measure will be supported on the orbit of a fixed point, then, we will need to prove the following:

(ii) Given $k \geq 1$ and $n_0 \geq 1$, we can build a periodic point $z(n) = (z_0, z_1, ...) \in \Sigma$, where n is bigger or equal to n_0 and such that $z_i \in V_k^{i(k,v_\infty)}$ for all $0 \leq i \leq n-1$ and also $z_i \in \{1, 2, ..., N_k\}$, for all $n \leq i \leq M_k + 1$.

In fact, lemma 7.2 says that there is an admissible word, say $z_0 z_1 ... z_n$ of length n_0 in $V_k^{i(k,v_\infty)}$. Then, by lemma 7.3 there are letters $u_1, u_2 \in \{1, 2, ..., N_k\}$ such that z_n can be connected to u_1 passing only through letters of our sector and u_2 can be connected to z_0 in the same way. By the Topologically Mixing property, we can take M_k to be a integer that can connect any two letters in the finite set (in special, the letters u_1 and u_2) by an admissible word of length M_k (which is independent of n).

(iii) Given ϵ, we can find k and n_0 such that $|\varphi(z(k,n)) - \varphi(x_v))| < \epsilon$, for every $n > n_0$. Once φ is uniformly continuous, there is η such that $d(z(k,n), x_v) < \eta \implies |\varphi(z(k,n)) - \varphi(x_v))| < \epsilon$. Then, it is enough to prove that we can find k, n_0 such that $d(z(k,n)), x_v) < \eta$. In fact, just take k and n_0 such that $diam V_k^i = \delta_k$ and θ^{n_0} are both less than $\eta(1-\theta)(1-\theta^4)/2$. In fact:

$$d(z(k,n), x_v) = \sum_{j=0}^{n-1} \theta^j \rho(z_j, v_\infty) + \sum_{j=n}^{p-1} \theta^j \rho(z_j, v_\infty) + \sum_{j=p}^{p+n-1} \theta^j \rho(z_j, v_\infty) + \sum_{j=p+n}^{2p-1} \theta^j \rho(z_j, v_\infty) + \cdots$$

$$\leq \sum_{j=0}^{n-1} \theta^j \delta_k + \sum_{j=n}^{p-1} \theta^j + \sum_{j=p}^{p+n-1} \theta^j \delta_k + \sum_{j=p+n}^{2p-1} \theta^j + \cdots = \sum_{j=0}^{\infty} a_0 \theta^{jp}$$

Where $p = n + q$ and $q = M_k + 2$ and:

$$a_0 = \frac{(1-\theta^{n-1})\delta_k + \theta^n(1-\theta^{n+q})}{1-\theta}$$

Then:

$$d(z(k,n), x_v) \leq \frac{(1-\theta^{n-1})\delta_k + \theta^n(1-\theta^p)}{(1-\theta)(1-\theta^p)}$$

Without loss of generality, taking $n_0 > 2$, we have that:

$$d(z(k,n), x_v) \leq \frac{\delta_k + \theta^n}{(1-\theta)(1-\theta^4)}$$

And it is easy to see that, taking δ_k and n_0 as said, we get the desired inequality.

(iv) Furthermore, we are going to suppose that we can find δ_k and n_0 such that $d(\sigma^j(z(k,n)), \sigma^j(x_v)) < \eta$ for $0 \leq j < n$. Although we make this hypothesis, I have no hope that this is true, and am pretty sure that it is false.

(v) Then, with ϵ, ϕ and x_v fixed, we take $z(k,n)$ as said above and define the following measure $\mu_{z(n)}$ in Σ:

$$\mu_{z(n)} = \frac{1}{n + M_k + 2} \sum_{i=0}^{n+M_k+1} \delta_{\sigma^i(z(n))}$$

Since each $z(k,n)$ is periodic, this measure is invariant. Now, we know are going to prove that this sequence of measures differ from μ at most by ϵ. Indeed, recalling that $\int \varphi d\mu = \varphi(x_v)$ by the definition of μ, we have that:

$$\left|\int \varphi(x)d\mu_z(x) - \int \varphi d\mu\right| = \left|\int [\varphi(x) - \varphi(x_v)]d\mu_z(x)\right| \leq \int |\varphi(x) - \varphi(x_v)|d\mu_z(x)$$

$$= \frac{1}{n+M_k+2} \sum_{i=0}^{n+M_k+1} \int |\varphi(x) - \varphi(x_v)|d\delta_{\sigma^i(z)} < \frac{1}{n} \sum_{i=0}^{n+M_k+1} |\varphi(\sigma^i(z)) - \varphi(x_v)|$$

By definition, $|\varphi(\sigma^i(z)) - \varphi(x_v)| \leq 2||\varphi||_\infty$, for all i, hence:

$$\left|\int \varphi(x)d\mu_z(x) - \int \varphi d\mu\right| < \frac{1}{n} \sum_{i=0}^{n-1} |\varphi(\sigma^i(z)) - \varphi(x_v)| + \frac{2(M_k+2)}{n}||\phi||_\infty$$

By the hypothesis above:

$$\left|\int \varphi(x)d\mu_z(x) - \int \varphi d\mu\right| < \epsilon + \frac{2(M_k+2)}{n}||\phi||_\infty$$

What tells us that we can approximate the ergodic measures on the boundary as well as we want it, and this completes the proof.

■

Theorem 7.8. *Let (Σ, σ) be a finite entropy CMS with set of vertices V. Let ρ be a discrete, totally bounded metric of V and $d = d_{\rho,\theta}$ be the induced metric in Σ. Then, given ϕ uniformly continuous with respect to ϕ, it holds that:*

$$P_\Sigma(\phi) = P_{\overline{\Sigma}}(\phi)$$

if (V, ρ) is sectorially arranged.

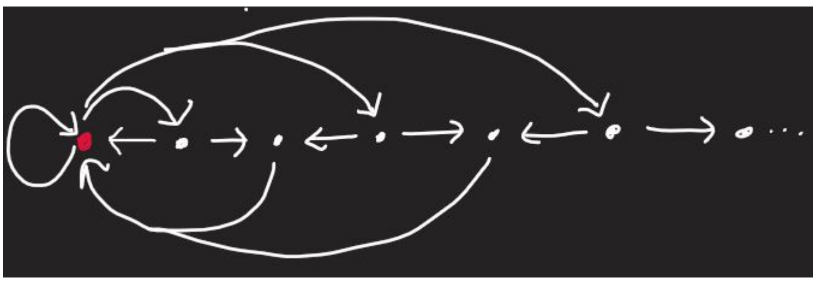
Proof. This follows directly from lemmas 6.14 and 7.7. ■

Corollary 7.8.1. *In this case, the pressure $P_\Sigma : UC_d(\Sigma) \to \mathbb{R}$ is Gateaux differentiable in a dense G_δ set of $UC_d(\Sigma)$ and the set of points at which the pressure is not Gateaux differentiable is an Aronszajn null set.*

Proof. It follows direct from the theorem and theorems 4.1 and 4.4. ■

7.2 Further Examples

Example 7.1. *In this example, we give a concrete case that has a possible sectorization such that we cannot extract arbitrarily big admissible words, from any sector. The entropy must be infinite, as indeed is the case. We have $A(1, n) = 1$ if n is even of $n = 1$; $A(n, 1) = 1$ if n is odd and $A(n, n \pm 1) = 1$ if n is even. The situation is shown in the next picture.*



Taking a vanishing metric, this example is clearly sectorially arranged if we take each sectorization having just one sector, namely $V_k = \{v; v > k\}$. The sector is always connected, its diameter *tends to zero as its complement tends to have infinite elements. One can easily seen that the largest admissible word that can be formed in some sector has length 1.*

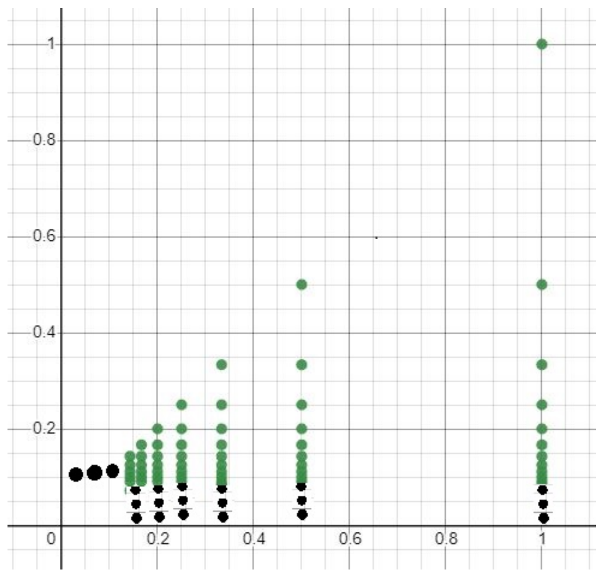
Nonetheless, we can find a different form of organising the sectors so that this not happens, for example, if we add 1 to each sector, we will be able to form arbitrarily big admissible words again. We may ask, then, if we can always find some sectorization that allows us to do so, even if the entropy is infinity.

Example 7.2 (Sectorially arranged and $+\infty \to -\infty$, but $+\infty \neq -\infty$)**.** *By notational convenience, we will denote $\mathbb{N}$ by $\mathbb{Z}$ and we'll consider the $\mathbb{Z} \times \mathbb{Z}$ transition matrix $(a_{i,j})$ defined by $a_{0,j} = 1$, for all $j \in \mathbb{Z}$, $a_{i,i-1} = 1$, for all $i \geq 1$, $a_{i,i+1} = 1$ for all $i \leq -1$, $a_{2i,-2i} = 1$ for all $i \geq 1$ and $a_{i,j} = 0$ otherwise. We'll consider the following metric in $\mathbb{Z}$:*

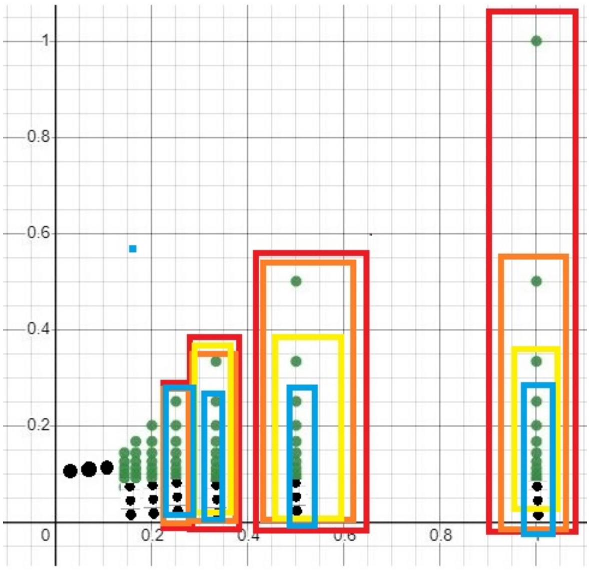
$$\rho(a,b) = \begin{cases} 0, & \text{if } a = b; \\ 1, & \text{if } ab \leq 0; \\ \left|\frac{1}{a} - \frac{1}{b}\right|, & \text{if } ab > 0 \end{cases}$$

If we restrict ourselves in $\mathbb{Z}_+$ or $\mathbb{Z}_-$, then ρ is of vanishing type as we've already proved, so there is exactly one infinity associated with each $\mathbb{Z}_\pm$, so we can say that $\partial\mathbb{Z}_\rho = \{+\infty, -\infty\}$ in this case. For each $k \geq 1$, we will define only two sectors by $V_k^1 = \{k, k+1, ...\}$ and $V_k^2 = \{-k, -k-1, ...\}$ and we'll check that this structure is sectorially arranged. Of course, for every $k \geq 1$ we have $\mathbb{Z} = \{-k-1, -k, ..., -1, 0, 1, 2, ..., k-1\} \cup V_k^1 \sqcup V_k^2$, every sector is trivially infinite, connected and the refinement property holds. To see why the diameter tends to zero, by bijecting $\mathbb{Z}$ with $\{\frac{1}{n} : n \in \mathbb{Z} \setminus \{0\}\}$ for instance we see that the diameter of V_k^1 is the diameter of the set $\{\frac{1}{k}, \frac{1}{k+1}, ...\}$ with the euclidean metric, which is just $\frac{1}{k} \to 0$, and the same holds for V_k^2. Finally, fixed k, we'll show that the sectors V_k^1 and V_k^2 are not connected with each other. In fact, for $k = 2i$ with $i \geq 1$, the only way to connect $2i \in V_{2i}^1$ with $-2i-1 \in V_{2i}^2$ is by going back to the origin first, since the only way to go directly from the first sector to the second without going to the origin first is by the point $2i \to -2i$ (since we are not allowed to go directly to any other points of the form $2(i+k)$ for $k > 0$ starting from $2i$), but since $a_{-2i,-2i-1} = 0$, we see that we cannot go from $2i$ to $-2i-1$ in this way. An analogous thinking for the odd sectors will get the result. Thus, this setting is sectorially arranged.

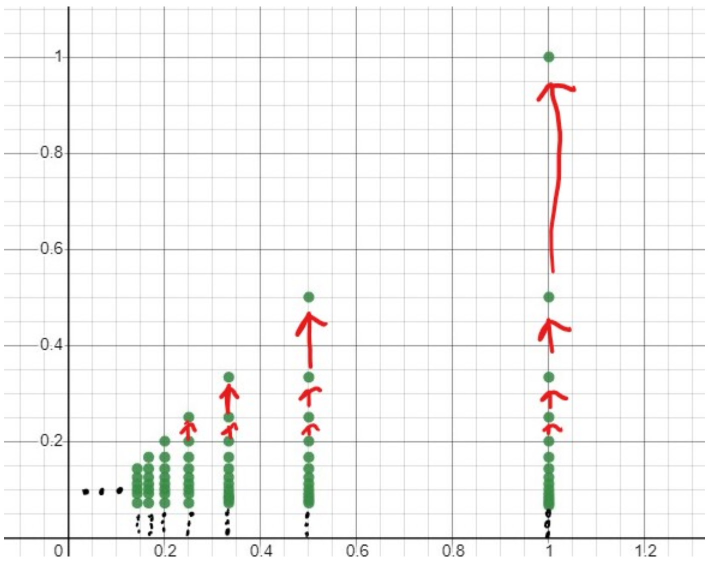
Example 7.3 ($v_\infty \in \partial V \setminus \bigsqcup_{i,k=0}^{+\infty} V_k^i$, but infinite entropy)**.** *Consider the set of vertices $V := \{\left(\frac{1}{m}, \frac{1}{n}\right) : m, n \in \mathbb{N}$ and $n \geq m\}$ endowed with the euclidean metric of $\mathbb{R}^2$ (in special, this metric space is totally bounded, as every bounded subset of $\mathbb{R}^2$ is and it also is obviously discrete). A picture is as follows:*



If $\overline{V}$ is the completion of V, we'll define it's boundary as $\partial V := \overline{V} \setminus V$. In our case, we have $\partial V = \{(\frac{1}{m}, 0) : m \in \mathbb{N}\} \cup \{0\}$ where $0 \in \partial V$ since it is the limit of the sequence $(\frac{1}{m}, \frac{1}{m})_{m \geq 1} \subset V$. We'll also define the sectors as $V_k^i := \{(\frac{1}{i}, \frac{1}{l} : l > k)\} \cap V$. A picture is as follows:



Of course, we have $V = \left[\{(\frac{1}{i}, \frac{1}{l}) : i \leq k \text{ and } l \leq k\} \cap V\right] \cup \dot{\bigcup}_{i=1}^{+\infty} V_k^i$ for each $k \geq 1$ and each set $\{(\frac{1}{i}, \frac{1}{l}) : i \leq k \text{ and } l \leq k\} \cap V$ is finite. Of course, the refinement property is trivially satisfied, each V_k^i is an infinite set and $diam V_k^i \leq \frac{1}{k} \to 0$. We'll now define the shift structure on the Markov Shift $\Sigma_A = V^{\mathbb{N}}$ by the transition matrix A defined by $A((1,1), v) = A(v, (1,1)) = 1$ (which already makes the shift topologically mixing), for all $v \in V$ and finally $A\left((\frac{1}{m}, \frac{1}{n}), \left(\frac{1}{i}, \frac{1}{j}\right)\right) = 1 \iff m = i$ and $n = j+1$, where this operation is defined for all $m, n \neq 1$ and $i, j \neq 1$. The structure is represented in the next figure:



This implies that each sector is connected (given two points inside a sector, there exists a path completely inside the sector which connects the points, no matter the direction) and the sectors are pairwise disconnected (to go from one sector to another, it is necessary to pass through the point $(1,1)$, which is outside of each sector). Thus, the shift space is sectorially arranged.

Now, for the contradiction: we already established that $0 \in \partial V$, but the origin is not in any sector, since the boundaries of the sectors are given by the points $(\frac{1}{i}, 0) \neq 0$.

However, the entropy of this example is clearly infinity. Indeed, we can find a different $2-$*periodic point for every column. Just take* $(1,1)$, *connect to the first point of each column and you can come back to* $(1,1)$

Based on the previous example, we can come up with a question. Is it possible to have a shift with infinitely many sectors, topologically mixing, but with finite entropy? The following example tries to give this an answer.

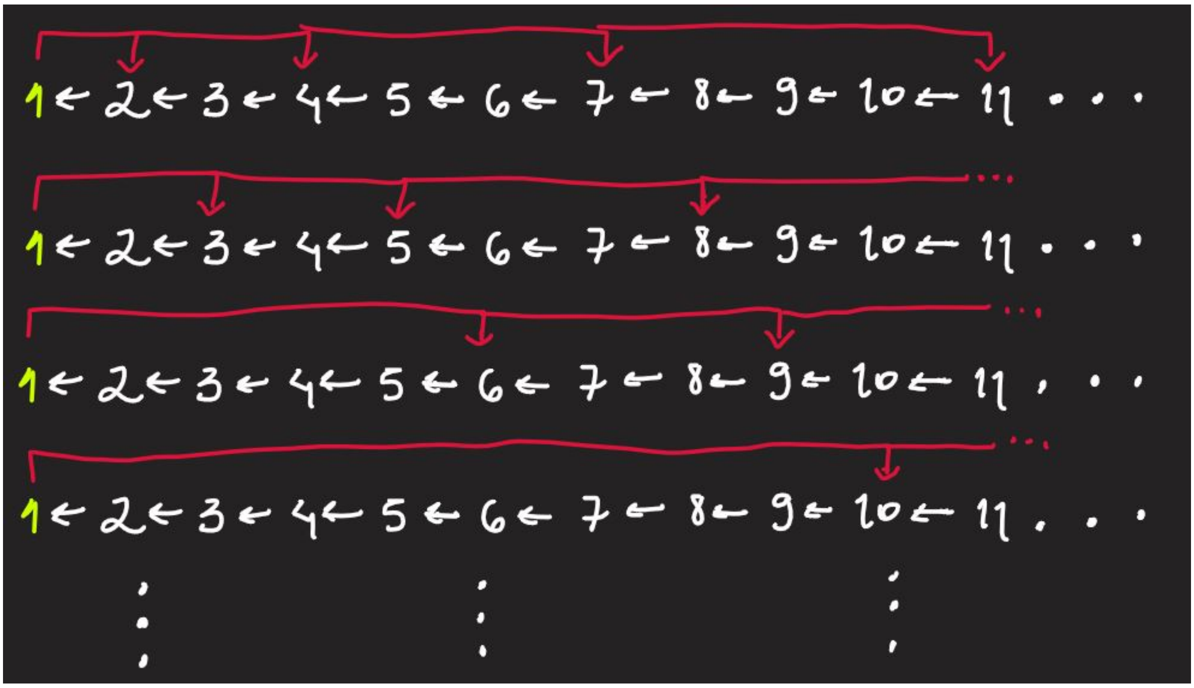
Example 7.4 (Fractal Triangle Modified)**.** *This example is very similar to the previous, with the exception that we make way less connections between* $(1,1)$ *and the points, to prevent that the entropy explodes. It is somewhat difficult to explain and (mainly) draw. So, we will forget by a moment the spacial and metric structure and pretend that each column is just like a Renewal shift, piled up. The connection between* $(1,1)$ *and the other points will only be allowed in one direction and the number of points between these connections will vary depending on the column, like the following image suggests.*

The number 1, in yellow, was repeated in every column to facilitate the visualization, but they all correspond to the same vertex. The other numbers, although repeated in every column, represent different vertices. As it can be seen, in the first column, 1 *connects to* $2, 4, 7, 11, ...$, *in the second,* 1 *connects to* $3, 5, 8, 12, ...$, *in the third,* 1 *connects to* $6, 9, 13, ...$, *in the fourth,* 1 *connects to* $10, 14, ...$ *and so on.*

The logic behind the construction is the following: in the first column we wanted a periodic point of length 2, so we connected 1 to 2. Then, we advance one column and made a periodic point of length 3 by connecting 1 to 3. Then, we go back to the first column and made a periodic point of length 4 by connecting 1 to 4, then to the second column to make a periodic point of length 5, and then we advance one column to make a periodic point of length 6. Then, we go back to the first column and repeat this process indefinitely, always adding one column after each step.

We shall now calculate the entropy of this space. It is not hard to see that there are exactly one word of length k *that passes through* 1, *so as in lemma 5.8 each* B_k *is one. Thus, the function* f *is just* $f(x) = x + x^2 + x^3 + ... = x(1 + x + x^2 + ...) = \frac{x}{1-x}$. *Of course, we have* $1 - f(x) = \frac{1-x}{1-x} - \frac{x}{1-x} = \frac{1-x-x}{1-x} = \frac{1-2x}{1-x}$. *Thus, we have* $\frac{f(x)}{1-f(x)} = \frac{x}{1-x}\frac{1-x}{1-2x} = \frac{x}{1-2x}$. *This has first derivative* $f'(x) = \frac{1}{(1-2x)^2}$ *so* $f'(0) = 1$. *After each higher derivative a factor of* 2 *comes from derivation by parts and each exponent drops (if we are taking the* n*'th derivative, there is another factor of* n *coming into play). Thus, if we define* $a_n = \frac{d^n}{dx^n}\big|_{x=0} f(x)$, *we get the recursive relation* $a_n = 2na_{n-1}$ *with initial condition* $a_1 = 1$.

By iterating this function n *times, we see that* $a_n = 2n.2(n-1)...2(2).2.1 = 2^{n-1}n!$. *Then, we get* $\frac{1}{n!}a_n = 2^{n-1}$, *and the entropy easily becomes* $\log(2)$.



Example 7.5 (A Cantor Set as a Boundary)**.** *First of all, we will consider the dynamic:*

$$\hat{\sigma} : \bigcup_{n=1}^{+\infty} \{1,3\}^n \to \{\epsilon\} \cup \bigcup_{n=1}^{+\infty} \{1,3\}^n$$

Which acts on the space of finite sequences of 1*'s and* 3*'s, where* ϵ *is the empty word. Here, the finite word* $x_0 x_1 ... x_{n-1} x_n 2$ *is mapped to* $x_1 ... x_n 2$ *and the word* x_0 *is mapped to* ϵ*. We'll use the standard concatenation notation to build new words and, in the case of the empty word, we have* $\underline{w}\epsilon = \epsilon\underline{w} = \underline{w}$*. Our (countable) alphabet* V *is itself a space of finite sequences of* 1*'s and* 3*'s as follows:*

$$V = \left\{ \langle \underline{w}2 \rangle : \underline{w} \in \{\epsilon\} \cup \bigcup_{n=1}^{+\infty} \{1,3\}^n \right\}$$

Where we only use the brackets $\langle\rangle$ *to indicate that we are dealing with letters in the space* V*, and not words. We will also consider the usual left shift* $\sigma : V^{\mathbb{N}} \to V^{\mathbb{N}}$ *acting in this space. We can turn this into a Markov Shift by allowing* $2 \to \langle \underline{w}2 \rangle$ *for any* $\underline{w} \in \{\epsilon\} \cup \bigcup_{n=1}^{+\infty} \{1,3\}^n$ *and* $\langle \underline{w}2 \rangle \to \langle \underline{w'}2 \rangle$ *only if* $\underline{w'} = \hat{\sigma}\underline{w}$*. This just means that each point* $\underline{x} \in \Sigma$ *is necessarily of the form:*

$$x = (\langle \underline{w}^1 2 \rangle, \langle \hat{\sigma}\underline{w}^1 2 \rangle, ..., \langle \hat{\sigma}^{|w^1|}\underline{w}^1 2 \rangle, \langle \underline{w}^2 2 \rangle, \langle \hat{\sigma}\underline{w}^2 2 \rangle, ...)$$

Where $\hat{\sigma}^{|w^i|}\underline{w}^i 2 = 2$ *and* $w^i \in \bigcup_{n=1}^{+\infty} \{1,3\}^n$*.*

The idea behind this construction is that we are thinking of 1, 2 *and* 3 *as the parts of the thirds of the unit closed interval* $[0,1]$*, so* 1 *represents the first third* $[0, \frac{1}{3}]$*,* 2 *represents* $[\frac{1}{3}, \frac{2}{3}]$ *and* 3 *represents* $[\frac{2}{3}, 1]$*. A sequence* $(3,1,1,2)$*, for example, means that, in the construction of the Cantor set, we first go to* $[0, \frac{1}{3}]$*, then we go to the first third of this interval, i.e,* $[0, \frac{1}{9}]$*, then in the next step we go to the third part of this interval. This way of going to the right or left at each step can well be represented by the dyadic tree as follows:*

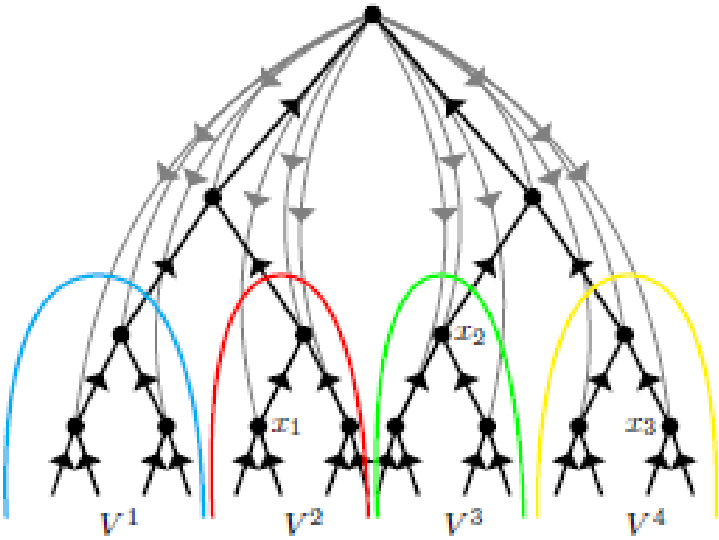


Figure 2: An example of sectors with a complex boundary with a renewal-like structure: the vertices are dots, the arrows given the Markov structure. The diagram continues downwards with infinitely many vertices and arrows in the same pattern. The arrows going from the base vertex are in grey so as not to obscure the structure too much. We pick out particular nodes x1, x2, x3 to show how the metric works in the text.

Note that we start counting the by the right, and not by the left, which is not usual. We do this, however, because we would like the shift acting in a point $\underline{x}$ to represent the same point, which approximates a point in the Cantor set, but with the last step in the construction of the Cantor set omitted (if we started counting from the left to right, the shift acting on a point $\underline{x}$ would result in a point $\hat{\sigma}\underline{x}$ approximating a point in the Cantor set which starts at the second step in the construction, which is not an interesting dynamic).

The metric which we will put here is defined as follows:

$$\rho(\langle \underline{w}2\rangle, \langle \underline{w}'2\rangle) = \begin{cases} 0, \text{ if } \underline{w} = \underline{w}'; \\ \frac{1}{1+\min\{i: \underline{w}_{|w|-i} \neq \underline{w'}_{|w'|-i}\}}, \text{ if otherwise.} \end{cases}$$

Of course, the minimum only makes sense if $|\underline{w}| = |\underline{w}'|$, but if $|\underline{w}| \neq |\underline{w}'|$ we can just complete one of the letters with empty words until say $|\epsilon\epsilon...\epsilon\underline{w}| = |\underline{w}'|$. This metric is good for us, since it starts counting differences from the right to the left, i.e, if two sequences of approximations of points in the Cantor set are equal for a long period of time, then their distance is small. First, we'll check that this is a discrete and totally bounded metric.

To see why it is discrete, take any vertex $\langle w_1 w_2 ... w_n 2\rangle$ and take $\epsilon = \frac{1}{n+2}$. Suppose first that we have a vertex $\underline{z} \neq \underline{w}$ with $|\underline{z}| \leq |\underline{w}|$. Since $\min\{i : \underline{z}_{|z|-i} \neq \underline{w}_{|w|-i}\} \leq |\underline{z}| < |\underline{w}|$ in this case, we get:

$$\rho(\langle \underline{z}2, \underline{w}2\rangle) = \frac{1}{1+\min\{i: \underline{w}_{|w|-i} \neq \underline{z}_{|z|-i}\}} \geq \frac{1}{1+|\underline{w}|} = \frac{1}{1+n} > \frac{1}{n+2} = \epsilon$$

Now, if $|\underline{z}| > |\underline{w}|$, then as we said before, we need to check where the differences between the vertexes occur by completing $\underline{w}$ with empty words until $|\underline{z}| = |\epsilon\epsilon...\epsilon\underline{w}|$. Of course, we naturally have $i \in \{i : \epsilon\epsilon...\epsilon\underline{w}_{|\epsilon\epsilon...\epsilon\underline{w}|-i} \neq \underline{z}_{|z|-i}\}$ for all $i \geq |\underline{w}|$, since starting to the point where the empty words come into play up to the last number (again, counting to the right to the left, so we are referring to the element $\epsilon\epsilon...\epsilon\underline{w}_0$ here) all the numbers of both vertexes differ, so that $\min\{i : \epsilon\epsilon...\epsilon\underline{w}_{|\epsilon\epsilon...\epsilon\underline{w}|-i} \neq \underline{z}_{|z|-i}\} < |\underline{w}|$ and we get back to the first case, and this proves the metric is discrete.

To see why the metric is totally bounded, take any $\epsilon > 0$ and consider the collection $A = \{\langle \underline{w}2 \rangle : |\underline{w}| \leq \frac{1}{\epsilon}\}$ (without loss of generality, we are supposing $\epsilon < 1$, so that $\frac{1}{\epsilon} > 1$ and this set is well defined). Of course, this set is finite and, given any letter $\langle \underline{z}2 \rangle$ we can suppose that $|\underline{z}| > \frac{1}{\epsilon} \geq N_0$ (otherwise our letter would already be part of some ball of the finite collection), where N_0 is the biggest integer less or equal to $\frac{1}{\epsilon}$. Thus, we can take the letter $\underline{w}$ of size N_0 such that $\underline{w}_{|w|-i} = \underline{z}_{|z|-i}$, for all $0 \leq i \leq N_0$ and note that $\underline{w} \in A$. Of course, we must have $\min\{i : \underline{w}_{|w|-i} \neq \underline{z}_{|z|-i}\} \geq N_0$, so that:

$$\rho(\langle \underline{z}2 \rangle, \langle \underline{w}2 \rangle) < \frac{1}{1+N_0}$$

Now, we must have $\frac{1}{1+N_0} < \epsilon$, otherwise we would have $\frac{1}{1+N_0} \geq \epsilon \implies N_0 + 1 \leq \frac{1}{\epsilon}$, so N_0 would not be the biggest integer less of equal to $\frac{1}{\epsilon}$, which is an absurd. Thus, we found an element $\langle \underline{w}2 \rangle$ of the finite set A such that $\langle \underline{z}2 \rangle \in B_\rho(\langle \underline{w}2 \rangle, \epsilon)$, and this finishes the proof of total boundness.

Also, by the geometry of the sectors, we clearly see that the vertex set is sectorially arranged.

At last, we will find an identification of the boundary of this dynamical system. Our best bet is $\{1,3\}^{\mathbb{N}}$, since this represents all the infinite sequences of 1's and 3's in our dyadic tree. We'll do this by finding a bijection between $\{1,3\}^{\mathbb{N}}$ and ∂V. First, we will define a function $f : \{1,3\}^{\mathbb{N}} \to \partial \Sigma$ given by $f((x_n)_n) = [(\langle x_0 2 \rangle, \langle x_1 x_0 2 \rangle, ...)]$. It is not hard to see that this function is well-defined, since the sequence of letters in the definition is Cauchy. Here, the infinite sequence is read from the right to the left as well, for example, $(...13132) = (..., x_4, x_3, x_2, x_1, x_0 2)$ and we omit the final number 2. It is possible to show that this function is injective, since if $[(\langle x_0 2 \rangle, \langle x_1 x_0 2 \rangle, ...)] = [(\langle y_0 2 \rangle, \langle y_1 y_0 2 \rangle, ...)]$, then:

$$\lim_{n \to +\infty} \rho(\langle x_n x_{n-1} ... x_0 2 \rangle, \langle y_n y_{n-1} ... y_0 2 \rangle) = 0$$

It is enough to show that $\langle x_n x_{n-1} ... x_0 2 \rangle = \langle y_n y_{n-1} ... y_0 2 \rangle$ for every n. If this were not the case, then there would be $N \geq 1$ such that $\langle x_N x_{N-1} ... x_0 2 \rangle \neq \langle y_N y_{N-1} ... y_0 2 \rangle$. In special, there would be some $0 \leq i_0 \leq N$ such that $y_{i_0} \neq x_{i_0}$ (without loss of generality, suppose this is the minimum index such that this happens). Thus, for every $m \geq N$ we would get $\rho(\langle x_m x_{m-1} ... x_0 2 \rangle, \langle y_m y_{m-1} ... y_0 2 \rangle) = \frac{1}{i_0+1}$, so the limit wouldn't be able to converge to zero. Thus, f is injective.

To see why it is surjective, take some $\underline{y} \in \partial \Sigma$, so there exists a Cauchy sequence:

$$(\langle \underline{w_n} 2 \rangle)_{n \geq 1} \subset \Sigma$$

Which converges to $\underline{y}$ under the complete metric we'll call d. It is enough to find a sequence $(x_n) \subset \{1,3\}^{\mathbb{N}}$ such that $f((x_n)_n) = [(\langle x_0 2 \rangle, \langle x_1 x_0 2 \rangle, ...)] = \underline{y}$, and for this it is enough to show that:

$$\lim_{n \to +\infty} \rho(\langle x_n x_{n-1} ... x_0 2 \rangle, \langle \underline{w_n} 2 \rangle) = 0$$

We will construct (x_n) as follows: for $n = 1$, there is $N_1 > 1$ such that $a \geq N_1 \implies \rho(\langle \underline{w_a} 2 \rangle, \langle \underline{w_{N_1}} 2 \rangle) \leq \frac{1}{1+1}$. In special, we have $\min\{i : \underline{w_a}_{|w_a|-i} \neq \underline{w_{N_1}}_{|w_{N_1}|-i}\} \geq 1$, so $\underline{w_a}$ and w_{N_1} have the same elements from 0 to 1, given that $a \geq N_1$. We can then define x_0 as the zero'th element of .

Then, for $n = 2$, there is $N_2 > N_1, 2$ such that $a \geq N_1 \implies \rho(\langle \underline{w_a} 2 \rangle, \langle \underline{w_{N_2}} 2 \rangle) \leq \frac{1}{1+2}$, so similarly $\underline{w_a}$ and $\underline{w_{N_2}}$ have the same number of elements from 0 to 2, given that $a \geq N_2$. Now, since $N_2 > N_1$ we get that $\underline{w_{N_2}}$ and $\underline{w_{N_1}}$ have the same elements from 0 to 1. We can now define x_2 as the second element from $\underline{w_{N_2}}$.

For $n = 3$, we can find $N_3 > N_2, 3$ such that $a \geq N_2$ implies that $\underline{w_a}$ and $\underline{w_{N_3}}$ have the same number of elements from 0 to 3 and in special $\underline{w_{N_3}}$ has the same elements from 0 to 2 as $\underline{w_{N_2}}$. We can then define x_3 as the third element from $\underline{w_{N_3}}$. We proceed inductively to find a sequence (x_n) of one and three's and we'll show this sequence will work. In fact, given $\epsilon > 0$, let N be the biggest integer less than $\frac{1}{\epsilon}$ and let N_{i_0} be the smallest of the N_i bigger or equal to N. Now, note that by construction we have that w_{N_n} and $\langle x_n x_{n-1} ... x_0 \rangle$ share the first N_n digits. If $n \geq \max\{i_0, N_{i_0}\}$, then $N_n > N_{i_0}$ so w_{N_n} share the first N_{i_0} digits with $w_{N_{i_0}}$ and since $n \geq N_{i_0}$ we also have that w_n and $w_{N_{i_0}}$ share the first N_{i_0} elements, and so $\langle x_n x_{n-1} ... x_0 \rangle$ and w_n share the first N_{i_0} digits. But this means that $\min\{i : \underline{w_n}_{|w_n|-i} \neq w_{\langle x_n x_{n-1}...x_0\rangle | \langle x_n x_{n-1}...x_0 \rangle|-i}\} \geq N_{i_0}$, so that:

$$\rho(\langle x_n x_{n-1} ... x_0 2 \rangle, \langle \underline{w_n 2} \rangle) \leq \frac{1}{N_{i_0} + 1}$$

Now, if $\frac{1}{N_{i_0}+1} \geq \epsilon$, then $N + 1 \leq N_{i_0} + 1 \leq \frac{1}{\epsilon}$, which is an absurd. Thus:

$$\rho(\langle x_n x_{n-1} ... x_0 2 \rangle, \langle \underline{w_n 2} \rangle) \leq \frac{1}{N_{i_0} + 1} < \epsilon$$

As required.

Finally, we are going to calculate the entropy of this example, again by means of 5.8. In this case, it is not difficult to convince yourself that:

$$f(x) = x + 2x^2 + 4x^3 + ... = \frac{x}{1 - 2x}$$

which is well defined at least in $(-\frac{1}{2}, \frac{1}{2})$. Clearly:

$$\frac{f(x)}{1 - f(x)} = \frac{x}{1 - 3x} = x + 3x^2 + 9x^3 + ...$$

In this case, with no much effort, we can see that:

$$\frac{1}{n!} \frac{d^n}{dx^n} \left(\frac{x}{1 - 3x} \right) \bigg|_{x=0} = 3^{n-1}$$

So $h_{top}(\sigma) = \log 3$. By slightly modifying the shift so that each vertex gives rise to k others, instead of 2, it can be easily shown that the entropy will be $\log(k + 1)$.

References

[BS00] Barreira, L. and Schmelin, J. *Sets of "non-typical" points have full topological entropy and full Hausdorff dimension.* Israel J. Math. 116 (2000), 29–70.

[BCL07] Béguin, F., Crovisier, S. and Le Roux, F. *Construction of curious minimal uniquely ergodic homeomorphisms on manifolds: the Denjoy-Rees technique.* Ann. Sci. Ecole Norm. Sup. (4) 40 (2007), no. 2, 251–308.

[BL00] Benyamini,Y. and Lindenstrauss, J. *Geometric nonlinear functional analysis.* Vol. 1. American Mathematical Society Colloquium Publications, 48. American Mathematical Society, Providence, RI, 2000.

[BEFR18a] Bissacot, R., Exel, R., Frausino, R., Raszeja, T. *Quasi-invariant measures for generalized approximately proper equivalence relations*, available at arxiv.org/abs/1809.02461

[BEFR18b] Bissacot, R., Exel, R., Frausino, R., Raszeja, T. *Thermodynamic Formalism for Generalized Markov Shifts on Infinitely Many States*, available at arxiv.org/abs/1808.00765

[Bow73] Bowen, R. *Topological entropy for noncompact sets.* Trans. Amer. Math. Soc. 184 (1973), 125–136.

[Bow08] Bowen, R. *Equilibrium states and the ergodic theory of Anosov diffeomorphisms.* Second revised edition. With a preface by David Ruelle. Edited by Jean-Ren´e Chazottes. Lecture Notes in Mathematics, 470. Springer-Verlag, Berlin 2008.

[DvE80] H.A.M. Daniëls and A.C.D. van Enter *Differentiability properties of the pressure in lattice systems.* Comm. Math. Phys. 71 (1980), no. 1, 65–76.

[FF95] D. Fiebig and U. Fiebig *Topological boundaries for countable state Markov shifts.* Proc. London Math. Soc. (3) 70 (1995), no. 3, 625–643.

[Fie13] D. Fiebig *Canonical compactifications for Markov shifts.* Ergodic Theory Dynam. Systems 33 (2013), no. 2, 441–454.

[GS98] B.M. Gurevich and S.V. Savchenko *Thermodynamic formalism for symbolic Markov chains with a countable number of states.* Uspekhi Mat. Nauk 53 (1998), no. 2(320), 3–106; translation in Russian Math. Surveys 53 (1998), no. 2, 245–344

[HK95] M. Handel and B. Kitchens *Metrics and entropy for non-compact spaces.* With an appendix by Daniel J. Rudolph. Israel J. Math. 91 (1995), no. 1-3, 253–271.

[IT13] G. Iommi and M. Todd *Transience in dynamical systems.* Ergodic Theory Dynam. Systems, 33 (2013) 1450–1476.

[IT18] G. Iommi, M.Todd and Aníbal Velozo *Upper Semi-Continuity of Entropy in Non-Compact Settings.* Available at: https://arxiv.org/abs/1809.10022

[IT20] G. Iommi and M.Todd *Differentiability of the Pressure in non-compact spaces.* Available at: https://arxiv.org/abs/2010.10250

[IP84] R.B. Israel and R.R. Phelps *Some convexity questions arising in statistical mechanics.* Math. Scand. 54 (1984), no. 1, 133–156

[Lim] Y. Lima *Symbolic dynamics for nonuniformly hyperbolic systems.* To appear in Ergodic Theory Dynam. Systems.

[MU96] R.D. Mauldin and M. Urbánski *Dimensions and measures in infinite iterated function systems.* Proc. London Math. Soc. (3) 73 (1996), no. 1, 105–154.

[PP84] Y.B. Pesin and B.S. Pitskel' *Topological pressure and the variational principle for noncompact sets*, Funktsional. Anal. i Prilozhen. 18 (1984) 50–63.

[Pes97] B. Pesin *Dimension theory in dynamical systems.* Contemporary views and applications. Chicago Lectures in Mathematics. University of Chicago Press, Chicago, IL, 1997. xii+304

[Phe] R.R. Phelps *Convex functions, monotone operators and differentiability.* Second edition. Lecture Notes in Mathematics, 1364. Springer-Verlag, Berlin, 1993. xii+117

[Rue78] D. Ruelle *Thermodynamic formalism. The mathematical structures of classical equilibrium statistical mechanics.* With a foreword by Giovanni Gallavotti and Gian-Carlo Rota. Encyclopedia of Mathematics and its Applications, 5. Addison-Wesley Publishing Co., Reading, Mass., 1978.

[Sar99] O. Sarig *Thermodynamic formalism for countable Markov shifts.* Ergodic Theory Dynam. Systems 19 (1999), no. 6, 1565–1593

[Sar01] O. Sarig *Phase transitions for countable Markov shifts.* Comm. Math. Phys. 217 (2001), no. 3, 555–577.

[Sar13] [S3] O. Sarig *Symbolic dynamics for surface diffeomorphisms with positive entropy.* J. Amer. Math. Soc. 26 (2013), no. 2, 341–426.

[Shw19] O. Shwartz *Thermodynamic formalism for transient potential functions.* Comm. Math. Phys. 366 (2019), no. 2, 737–779.

[Sin72] J.G. Sinai *Gibbs measures in ergodic theory.* Uspehi Mat. Nauk 27 (1972) 21–64.

[Tho11] D.J. Thompson, *A thermodynamic definition of topological pressure for non-compact sets.* Ergodic Theory Dynam. Systems 31 (2011) 527–547.

[Wal82] P. Walters *An introduction to ergodic theory.* Graduate Texts in Mathematics, 79. Springer-Verlag, New York-Berlin, 1982. ix+250

[Wal92] P. Walters *Differentiability properties of the pressure of a continuous transformation on a compact metric space.* J. London Math. Soc. (2) 46 (1992), no. 3, 471–481

[Zar86] A.S. Zargaryan *A variational principle for the topological pressure in the case of a Markov chain with a countable number of states.* Mat. Zametki 40 (1986), no. 6, 749–761, 829.