Research Article

Francisco Balibrea*

On the origin and development of some notions of entropy

DOI 10.1515/taa-2015-0006
Received June 2, 2015; accepted August 13, 2015

Abstract: Discrete dynamical systems are given by the pair \((X, f)\) where \(X\) is a compact metric space and \(f : X \to X\) a continuous map. During years, a long list of results have appeared to precise and understand what is the complexity of the systems. Among them, one of the most popular is that of topological entropy. In modern applications other conditions on \(X\) and \(f\) have been considered. For example \(X\) can be non-compact or \(f\) can be discontinuous (only in a finite number of points and with bounded jumps on the values of \(f\) or even non-bounded jumps). Such systems are interesting from theoretical point of view in Topological Dynamics and appear frequently in applied sciences such as Electronics and Control Theory.

In this paper we are dealing mainly with the original ideas of entropy in Thermodynamics and their evolution until the appearing in the twenty century of the notions of Shannon and Kolmogorov-Sinai entropies and the subsequent topological entropy. In turn such notions have to evolve to other recent situations where it is necessary to give some extended versions of them adapted to the new problems.

Keywords: Clausius; Boltzmann; Gibbs; Shannon; Kolmogorov-Sinai entropies; topological entropy

1 Introduction

Introduced by Rudolf Clausius, the notion of entropy appeared for the first time in the setting of Physics but has been adopted in other fields with different meanings. Here we present some of them.

The first idea to be considered is that given a physical system, the energy contained in it is comparable with the water contained in lakes, rivers or the sea. In such physical systems only surface waters are the unique which can be used to be transformed into work mechanical work or simply work (for example causing the rotation of a turbine).

In Classical Physics, entropy is seen as a magnitude which in every time is proportional to the quantity of energy that at such time can not be transformed into mechanical work.

Using the above interpretation, entropy plays a central role in the formulation of the Second Law of Thermodynamics which states that in an isolated physical system, any transformation of it leads to an increase of its entropy.

In Probability Theory, the entropy of a random variable measures the uncertainty over the values which can be reached by the variable.

In Information Theory, the entropy associate to the compression of a message (for example, of a file from a computer), quantifies the contain of information of the message to have the minimum lost of information in the compression process previous to its transmission.

In Dynamical Systems, the entropy measures the exponential complexity of the system or the average flow of information per unit of time.

*Corresponding Author: Francisco Balibrea: Departamento de Matemáticas, Universidad de Murcia, 30100-Murcia, Spain, E-mail: balibrea@um.es
In real life, the word entropy is understood as disorder or chaos.

Concerning the evolution of the notion of entropy, it is necessary to mention the outstanding paper [26] concentrated in the ergodic point of view and Kolmogorov’s new ideas on dynamical systems and the book [17] on comparison between different types of entropy and many topics associate to the topological entropy.

2 Prehistory of the entropy

The charge of putting the last stone in the thermodynamics science or simply Thermodynamics, was Rudolph Clausius in 1850 in the paper [15], where he coined the term entropy taking the Greek word \( \text{εντρπια} \) (trope) which means tranformation and trying to imitate the sound of the word energy. For him the meaning of entropy was the form of energy that eventually and inevitably turns into a useless heat.

Such meaning was inspired by an earlier formulation made by the French physician and mathematician Sasi Carnot (see [14]) of what is known his formulation of the Second Law of Thermodynamics: *entropy represents the energy no longer capable to perform work. In isolated systems it can only grow.* Sadi Carnot stated that in an ideal engine (which does not interchange heat with its outside), the entropy would be constant.

In Clausius time, were known previous experiments made by Joule in the sense of proving without doubt that mechanical work could always be transformed into heat. The reasoning to reach it was as follows. Carnot assumed that in his engine could not have losses of heat, according with the opinions and tendency of that time. Nevertheless, the experiments made by Joule indicated that heat could be created and in fact, Joule gave a precise equivalence between mechanical work and heat. If heat could be created, then could be also destroyed; therefore the claim of Carnot was false. As a consequence, Clausius wondered what was the origin of the energy necessary for a machine of Carnot to create mechanical work. For him the answer was clear, one part of the exchange of heat between the two sources composing the Carnot machine is the mechanism by which it is ceated mechanical work. Therefore heat could be created and destroyed from nothing but always transformed into the equivalent quantity of mechanical work. In this way the total energy had to maintain constant. This reasoning leads to the First Law of Thermodynamics: *the absorbed heat made by a physical system is equivalent to the mechanical work made by it, or increases the internal energy or is a combination of both processes.* In 1856, Clausius refined it by a more or less heuristic formulation, using differential calculus, earning precision but loosing intuition. Besides, around 1862 he assumed the atomic hypothesis of the materia and introduced the idea of how molecules separate. Finally in 1865 he introduced his notion of entropy and formulated the Second Law of Thermodynamics.

Clausius introduced the formula

\[
\Delta S = \frac{\Delta Q}{T}
\]

where \( \Delta S \) denotes the increment of entropy of a system and \( \Delta Q \) denotes the increment in heat of it at a temperature \( T \). That is, the increasing of entropy is proportional to the increasing in heat and inverse to the temperature. He proved, for example, that the sum of all increments of entropy during a complete Carnot cycle is zero, which means that the system increases the entropy when a systems receives heat in the same quantity that it losses in the part of the cycle when the systems cools. Since Carnot engine was ideal then it was a mechanism of a maximum efficiency. Nevertheless a real engine has unavoidable losses. Finally bouth things mean that

\[
\Delta S \geq 0
\]

which is the Second Law of Thermodynamics. That is, *when a system is transformed, then the entropy increases.*

The ideas of Clausius were refined by Ludwig Boltzmann. In first place Boltzmann was a follower of the cinetic theory of gases introduced by Daniel Bernouilli who considered that all fluids are agregates of molecules moving continuously. Temperature could be interpreted as a measure of the energy of particles. The termal energy of a gas is identified with the cinetic energy of the individual molecules which explains that
heat and mechanical work were different transmission forms of energy. More precisely, Boltzmann identified the temperature of a gas with the mean cinetic energy of its molecules. In an equilibrium state (there is no transmission between two substances of heat since both have the same temperature), there is no transmission of cinetic energy. Otherwise when two substances are not in equilibrium, the cinetic energy goes from the hotter to the colder. This means that average of cinetic energy behaves similarly than temperature and therefore both can be identified.

Boltzmann used a curious hypothesis. He assumed that the motion of molecules was periodic. That is, given sufficient time, a molecule is changing changed of level of energy until coming back to the initial level of it.

With the explanation of temperature in mechanics terms, the First Law of Thermodynamics was clarified. In fact heat and mechanical work were interchangeable because they are two forms of energy. But it remained to explain the Second Law which was difficult because the notion of entropy was not quite clear. Boltzmann tried to do the job using a mathematical formulation, very far from physics interpretations. He proved that heat (understood as a supplied energy) divided by temperature gave a quantity whose behavior was exactly that of entropy. Finally he gave thermodimamics macroscopic reasons without considering the molecular behavior, to prove the Second Law.

2.1 The ideas of Maxwell

James Maxwell was an impulsor of the cinetic theory of gases during the XIX century. His main idea was to describe the behavior of molecules in a gas by mean of a distribution function. He considered a great quantity of molecules of the gas and take knowledge of their velocities which is much better than considering individual molecules which is an intractable way from mathematical point of view. Such function had to indicate how were distributed the velocities in molecules. Even with such approach was possible to compute most of relevant properties of gases.

To get an adequate mechanical description of a fluid, Maxwell had to overcome two difficulties. By one way to find a distribution function adequate to every temperature and prove that such a function was unique. For the first problem, he claimed that the gaussian function represented adequately the distribution of velocities of molecules. The second problem was open until its solution by Boltmann. In 1868, Boltzmann gave a justification of using the gaussian function to describe a gas in most of particular cases and even he extended what was made by Maxwell in the sense of including gases submitted to a big range of forces.

2.2 Boltzmann’s contributions

In [9], Boltmann succeeded in proving Second Law of Thermodynamics using principles of Mechanics. In fact, such paper was the starting point of the Statistical Physics. It contained two important innovations; by one hand, the introduction of the currently known as the Boltzmann equation which models the behavior of a gas in different situations; on other hand, his first proof of Second Law is a consequence of the atomic theory and of the probability theory, stating what it is known as H theorem. In his formulation he used two hypothesis of simplification. The first is the gas has a uniform space distribution. The second velocities in each direction are equiprobables.

Boltzmann equation is a description of the evolution of the probability distribution of a dilute gas depending of several factors. In its simplest form is:

$$\frac{\partial F(t,x,v)}{\partial t} = \left(\frac{\partial F(t,x,v)}{\partial t}\right)_{\text{force}} + \left(\frac{\partial F(t,x,v)}{\partial t}\right)_{\text{diffusion}} + \left(\frac{\partial F(t,x,v)}{\partial t}\right)_{\text{collisions}}$$

In this case, $F(t,x,v)$ denotes and describes the density function of particles in the phase space. $F$ can be called more accurately the empirical measure. According with [21] and [22], modernly the equation can be written as
\[
\frac{\partial F}{\partial t} + \mathbf{v} \cdot \nabla_x F = \Omega(F, F)
\]

where \(\Omega(F, F)\) is the Boltzmann collision operator acting only on the velocity variable \(\mathbf{v}\) and is local in \((t, x)\) as

\[
\Omega(F, F)(\mathbf{v}) = \left( \int_{\mathbb{R}^3} d\mathbf{v}_*[\mathbb{B}(\mathbf{v} - \mathbf{v}_*), \sigma] [F, F_* - F, F] \right)
\]

where we are using the shorthand \(F = F(\mathbf{v}), F_* = F(\mathbf{v}_*), F = F'(\mathbf{v}), F_* = F'(\mathbf{v}_*)\). In this formula, \(\mathbf{v}, \mathbf{v}_*\) and \(\mathbf{v}, \mathbf{v}_*\) are the velocities of a pair of particles before and after collision that result from parametrizing over the sphere \(S^2\) the physical law of elastic collisions:

\[\mathbf{v} + \mathbf{v}_* = \mathbf{v}' + \mathbf{v}'_* \quad |\mathbf{v}|^2 + |\mathbf{v}_*|^2 = |\mathbf{v}'|^2 + |\mathbf{v}'_*|^2\]

A deep prediction of Boltzmann’s equation is given by the cited H-Theorem which says that the solutions of the equation satisfy

\[
\frac{d}{dt} H = \frac{1}{2} \int_{\mathbb{R}^3} dx \int_{\mathbb{R}^3} d\mathbf{v} F \log F \geq 0
\]

This \(H\)-functional is increasing with respect to time and all of it maximizers are exactly the Maxwellian states

\[
\mu(\mathbf{v}) = (2\pi)^{-\frac{3}{2}} e^{-|\mathbf{v}|^2/2}
\]

This Boltmann’s H-theorem proves the Second Law of Thermodynamics, meaning that physical entropy of an isolated system does not decrease with time. In the setting of Statistical Physics, it has been considered the most important Boltmann’s contribution.

For the former Boltmann’s equation, recently has been proved that it has classical solutions holding some relevant additional conditions (see [21] and [22]) which has been an open and difficult problem in Mathematics for 140 years.

In the deduction of his equation, he used what is known in the literature as Stosszahlansatz or molecular chaos. According to it, the molecules or atoms (particles) of a gas move colliding among them. Boltmann assumed that before a collision, velocities of particles are not interrelated, that is, they were moving completely at random. This does not happen after the collision, since the direction of moving of a particle depends on where were the collided particle. This assumption provokes a temporal asymmetry in the mathematical analysis, since it is necessary to distinguish between the past (there is no correlation) and the future. In fact this hides the notion of thermodynamics irreversality. Boltzmann’s equation proves that the change in the function \(F\) is only due to external forces, collisions among particles and diffusion phenomenon which is the statistical tendency of particles located in a region to expand trying to deal with all the allowed space.

Boltzmann proved also that the Maxwell distribution given before is a solution of the equation. After he proved that if a gas reaches such a distribution, then it will not change, that is the internal collisions among molecules do not change its state. Using his equation, he proved also that any gas in any state will tend to reach the Maxwell distribution. But the most relevant result was that the proof of the Second Law can be obtained from mechanics principles which means to do a mechanics interpretation of such Law. This result is again the Boltzmann H-theorem. For it, he considered instead of \(F\) the average of it, in fact its logarithm. It is the value \(H\) (originally denoted by \(E\) by Boltzmann in his papers). \(H\) had to be constant or decrease during any physical process. Changing the sign of \(H\), Boltzmann found a mechanics equivalent of the entropy (it remains constant or increases in any physical process). This fact represents a microscopic interpretation of the Second Law. Additionally, the Clausius’s entropy is valid only for systems in equilibrium, but the Boltzmann approach does not depend on it, and it is valid in any situation. Today the physics community disposes of entropy definitions valid for quantic systems and relativists due to the versality of Boltzmann formulation. But [9] contained an important additional result. Previously to it, Boltzmann assumed the ergodic hypothesis, that is, the fact that any particule has to reach all level of energy before reaching the initial one. But this point
of view was very complicate to considered in applications. He changed it and considered that had to be only a finite number of energy levels, multiples of one number. Then to proof his theorem, he used energy instead of velocities and discretized it. As a consequence, calculus became easier and when energies are transformed it is clear that given sufficient time, all particules will reach all levels of such discretization.

Boltmann distribution applied to the black body problem, reproduces exactly the Plank results. Some time after, Einstein explained the photoelectric effect, that is, the creation of a electric current from the incidence of light on a metal using a similar hypothesis. He assumed that light was composed of particles whose energy could not reach any value since was discretized. Such point of view was very important for the developing of Quantic Mechanics.

In order to overcome some criticisms to the previous ideas, in [11] Boltzmann did not consider a velocity distribution of a gas, but he thought on the probability that the gas would reach a state if it is known all the possible states. This meant to do an inventory of all configurations of the gas and then compute their probabilities. The state of maximal probability would correspond with that observed at a macroscopic scale. It was made in the following way contained in [11]. Boltzmann did not consider the distribution of velocities of a gas, but the probability that it would be in a state chosen among all possible states. But this means to have an inventory of all possible configurations of the system and its number to obtain the probabilities and the state of maximal probability. First he obtained the number of molecules having each discrete level of energy for a total fixed energy. The state of the system is given from macroscopic point of view by such number independently of what individual molecules has such level of energy.

Boltzmann named each individual state as complexion or in modern terminology microstate since such state is not observable. Distributions of energy where the only necessary is to know the number of molecules in each level of energy are known as macrostates since they are macroscopically observable. Then he introduced the number \( B \) that allows to obtain a new expression for entropy. \( B \) is the number of microstates supplying the same distribution of energy. Now calculate \( B \) for all distributions and compare them. The proportion between \( B \) and the total number of microstates is the probability that the system has in a state described by \( B \). After obtaining all possible distributions, the following step is to count how many microstates there is in each possible state. This was named by him as permutability derived from permutation and represented by \( B \). After some computations it can be observed that permutability was considerable greater in intermediate distributions, that is, in those in which energy is distributed more or less homogeneously, even similarly to Boltzmann distribution. After doing the above computations he obtained a general expression for permutability of a distribution, first assuming that number of molecules were very big and then that energy reached continuous values. Then he named degree of permutability to the logarithm of permutability.

Finally he observed that the expression for the degree of permutability was equal to the value \( H \) obtained in [9] with inverted sign. It is relevant since \( H \) was equal to entropy with negative sign. The conclusion is that \( \log B \) could be used as a measure of the entropy of the system. Additionally, this method can be used not only for gases but also for any other substance, monoatomic o poliatomic, liquid or solid. In [11], the entropy was finally introduced as \( 2/3 \) of the measure of permutability. In the modern nomenclature, such factor is incorporated to what is known was as Boltzmann’s constant although he never used it. Since permutability and number of microscopic states compatible with a distribution are proportionals, nowadays it is used the last number instead of permutability. Finally, entropy is proportional to the logarithm of the microscopic states compatible with the observed macroscopic state (such as temperature, pressure, volume, etc).

\[
S = K \log W \quad (1)
\]

the logarithm is used because simplifies the computations and reproduces the property of additivity of entropy, in the sense that the entropy of two systems sums instead of multiplies as does the permutability. All this modern terminology and formulas were introduced by J.Gibbs as part of his formalization of what is known as Statistical Physics. Former formula is graved in Boltzmann’s tomb. Einstein named this formula as the Boltzmann’s principle. In (1), the entropy \( S \) increases when \( W \) does. The more microstates, the more disorder and more entropy. Besides for only one possible microstate, entropy is zero.
The notion of disorder is an intuitive notion depending on the system to be considered. In the case of a gas, it is considered as an ordered state if its molecules have a distribution of energies or positions very different of those random which means the Boltzmann distribution. The most disordered states are the most probable and as consequence have the most entropy. Another consequence is that in the universe the disorder tends to increase.

3 Shannon’s entropy

The definition of Boltzmann’s entropy is quite a lot general and its idea has been used to introduce others notions of entropy in Mathematics, Computation Sciences and other fields. One of them was proposed by C. Shannon (see [36]) to evaluate the quantity of information contained in a message. A message is a sequence of symbols containing some type of information and which have to be transmitted. Let us suppose we have a sequence composed only of two symbols $\{0, 1\}$. If the frequency of zeros and ones is not random and there are some tendencies towards more zeros or ones, an observer receiving and reading a message of zeros and ones could predict the following symbol after having received a finite number of them. The sequence $1, 1, 1, \ldots$ is further predictable and after receiving one hundred of ones, it is likely the following symbol would be another 1. In this situation the information given by the message is very poor since one knows what needs before receiving the message. In such case the Shannon’s entropy would be minimum. Oppositely, the Shannon’s entropy has to be maximal when the sequence of symbols is a random sequence of zeros and ones. In this case after hundred of transmitted symbols, the only way to know the new one is just receiving it. In practice, most of messages using alphabets of languages, for example English and Spanish, has entropy relatively low, due to statistical preponderance of some letters in such languages. This made to have a low information in such messages and it facilitates the compression of them. Shannon’s entropy was inspired on Boltzmann’s and has been during years an interesting and important problem in Mathematics and Physics to understand their relations.

Inspired also in Boltzmann’s entropy, in Linux system (the operational system of free code producing an Android) the term entropy is used to specify the random data taken by the system from movements of mouse and keyboard produced when executing special instructions.

In probability theory, a probability vector $p$ is a sequence of finitely many non-negative numbers $\{p_1, p_2, \ldots, p_n\}$ holding $\sum_{i=1}^{n} p_i = 1$. We define the Shannon’s entropy of $p$ as:

$$H(p) = - \sum_{i=1}^{n} p_i \log_2 p_i$$

where we take $0 \log_2 0 = 0$. This definition is associate to the situation in which we have a finite partition of a probability space. If we consider an abstract space $\Delta$ equipped with a probability measure $\mu$ and $\mathcal{P}$ is a finite partition of $\Delta$, that is, the family of subsets $\{A_1, A_2, \ldots, A_n\}$ where they are pairwise disjoint and $\mu$-measurable and their union is $\Delta$. Then the probabilities $p_i = \mu(A_i)$ form a probability vector $P(\mathcal{P})$. Then we define

$$H_\mu(\mathcal{P}) = H(P(\mathcal{P}))$$

this example can be related with the notion of information. In former setting, given a measurable set $A$, we define the information associate with it, as

$$I(A) = - \log_2 (\mu(A))$$

and the information function $I_\Omega$ associate with a partition $\{A_1, A_2, \ldots, A_n\}$ of the space $\Omega$ as a function of $\omega \in \Omega$ given by
where it is assumed that as a function of \( \omega \), \( \mu(A_i)(\omega) \) is constant in \( A_i \) and \( \chi_A \) is the characteristic function of \( A_i \). It is immediate that the expected value of the information function with respect to \( \mu \) has to be just \( H_\mu(\mathcal{A}) \).

In [17] it is given two interpretations of Shannon entropy and in terms of information function and uncertainty. In the first case, the interpretation is based in the fact that given \( \omega \in \Omega \) and the above partition of \( \Omega \), the information gives an answer to the question: in what \( A_i \) we are?. Such a question is not binary, but can be replaced by a finite amount of binary questions necessary to determine where is \( \omega \). Depending on the arrangement of the sets \( A_i \), we obtain different results to locate the belonging of \( \omega \). If we denote by \( N(\omega) \) the smaller number to do such process we have (see [17]):

\[
I_\mathcal{A}(\omega) = N(\omega) - I_\mathcal{A}(\omega) + 1 \quad \text{for almost every } \omega.
\]

The real number \( I_\mathcal{A}(\omega) \) can be interpreted as a precise value and then entropy can be interpreted as the expected amount of information needed to locate a point in the given partition of \( \Omega \).

Let \( X \) be a random variable in the probability space \( \Omega \) reaching values in a finite set \( \{x_1, x_2, ..., x_n\} \). Associate to \( X \) we have the partition of \( \Omega \) into the sets \( A_i = \{ \omega \in \Omega : X(\omega) = x_i \} \). The probabilities \( p_i = \mu(A_i) = \text{Prob}(X = x_i) \) form a probability vector (distribution of \( X \)). Suppose now that an experimenter knows the distribution of \( X \) before performing a experiment, that is, before choosing an \( \omega \in \Omega \) and obtaining the value \( X(\omega) \). His/her uncertainty on the outcome is the expected value of the information he/her is missing to be certain. According to the above, such a value is the value of the entropy \( H_\mu(\mathcal{A}) \).

### 3.1 Comparation between Boltzmann and Shannon entropy (see [17])

Connections between Boltzmann and Shannon entropy has been a matter of controversy during years. At first view both notions are connected by their definitions, both refer to probability, but their analogy is far of being obvious.

The interpretation of analogy and differences rely on the differences between macroscopic states associate to Thermodynamics and microscopic states to Statistical Mechanics. That is, a macroscopic state is associate to a thermodynamical one described by probability distributions of physical magnitudes such as pressures, temperatures, volumes, etc, which can be explained in several ways, while in microscopic states one distinguishes all individuals particles taking into account their positions or velocities. Given a thermodynamical state \( \Lambda \) we have mentioned that the difference \( S(\Lambda) - S_{\text{max}} \) (where \( S_{\text{max}} \) denotes the maximum value of the entropy of all states) is proportional to \( \log_2(\text{Prob}(\Lambda)) \), that is, the logarithm of the probability of the macroscopic state \( \Lambda \) in the space \( \Omega \) of all microscopic states \( \omega \). Finally we find the following formula

\[
S_{\text{max}} - S(\Lambda) = kI(\Lambda) \quad (2)
\]

where \( I(\Lambda) \) denotes the probabilistic information associate to \( \Lambda \subset \Omega \). The conclusion suggested by last formula is that Boltmann’s entropy is strongly connected to Shannon’ information rather than with Shannon’s entropy. The above formula has also a controversial interpretation, since being \( S(\Lambda) \) negative it is reversed the sense of monotonicity. The more information associated to \( \Lambda \) the smaller is its Boltmann’s entropy we have. Such fact is explained at light of interpreting the meaning of information associate to a state.

Information on a state in a system is the information received by an outside observer. This means that it is reasonable to assume that such information escapes from the system and as consequence it will have a negative sign. It is the knowledge on the system by an observer what gives the degree of usefulness of the energy contained in the system to produce physical work, that it, decreasing the entropy of the system.

With the former ideas it is possible to do a complete interpretation of the above notions. The key idea is that every microstate in a system appreciated by an observer belongs to a macrostate \( \Lambda \), hides the information on its identity. Let us denote by \( I_\Lambda(\Lambda) \) the joint information hiding in the system in the state identified by the observer as \( \Lambda \). This entropy is maximal at the maximal and then it is calculated by \( \frac{S_{\text{max}}}{k} \). In the state \( \Lambda \), it is diminished by \( I(\Lambda) \) the information already taken by the observer. Then we have
\( I_h(A) = \frac{S_{\text{max}}}{k} - I(A) \)

and together with (2) we obtain

\[
S(A) = kI_h(A)
\]

which gives us a new interpretation of Boltzmann's entropy. In fact such entropy is proportional to the information still hiding in the system if the macrostate state \( A \) has been detected. The Boltzmann's entropy is determined up to an additive constant and then we can compute the change of entropy from one state to another. It is really hard to obtain the adequate constant, since the maximal state depends on the level of precision of the microstates composing a macrostate. If the space \( \Omega \) is infinite, then also it is infinite the maximal entropy. According to what was commented in previous sections, it is necessary to do a quantum approach, that is considered that \( \Omega \) is composed only of a finite number of points. The advantage is that in such case, Boltzmann's entropy has a new interpretation in direct terms of Shannon's entropy (not depending of information function). The highest possible Shannon's entropy \( H_\mu(Q) \) is achieved when \( Q = \Omega \) which denotes the partition of \( \Omega \) into singles states \( \omega \) and \( \mu \) is the uniform measure on \( \Omega \), that is, in the case that each state has probability \( 1/|\Omega| \), where \( |\Omega| \) denotes the cardinality of \( \Omega \). In such case we have

\[
S_{\text{maximal}} = kH_\mu(\Omega) = k\log_2|\Omega|
\]

We detect that the system is in state \( A \) if the information is \( I(A) = -\log_2(\mu(A)) = -\log_2\left(\frac{\mathcal{A}}{|\Omega|}\right) \). Using (2), we obtain

\[
S(A) = k\left(-\log_2\mathcal{A} + \log_2\left(\frac{\mathcal{A}}{|\Omega|}\right)\right) = k\log_2\mathcal{A}
\]

which is \( k \) times the Shannon's entropy of \( \mu_A \), the normalized uniform measure restricted to \( A \). Both entropies are compared by mean of the above formula and one obtains the unknown additive constant.

### 4 On Tsallis entropy

It is well known that Statistical Mechanics is essentially constructed with two ingredients, Mechanics (including electromagnetis forces) and Probability Theory. From it we may construct notions such as energy and entropy and obtain for them mathematical expressions. In what concerns entropy, guided by Thermodynamics we obtain that entropies from Clausius and Boltzmann are extensive, which means that the entropy \( S(N) \) of a macroscopic system of \( N \) elements asymptotically behaves as \( N \), which is denoted by

\[
S(N) \propto N
\]

when \( N \to \infty \) One expects to keep such property for both short (for example for collissions between paticules) and long-range interactions. For short-range it is the case using strictly thermodynamic tools. But for long-range the former approach is not evident. In such cases, total energy \( U \) of the system behaves non-extensively (more precisely, \( U(N) \) increases faster than \( N \) for large \( N \). Nevertheless, it can be seen that Clausius and Boltzmann entropies remain extensive (see [39]).

The next relevant question is to be able to introduce a mathematical expression in probability terms of a entropy for long-range interations, taking into account the admissible microscopic configurations given by correlations between the \( N \) elements of the system, to obtain that \( S(N) \) is extensive.

Assume we have a system whose total number \( T(N) \) of admissible configurations are equally probable and satisfies

\[
T(N) \propto \mu^N, \quad (N \to \infty, \mu > 1) \quad (3)
\]
Such assumption corresponds to probabilistic independence of the $N$ elements of the system and $\mu$ is a factor of proportionality, that is, it is assumed that $T(N+1) \sim \mu T(N)$. Using now Shannon’s entropy of a probabilistic vector

$$S_{sh} = -k \sum_{i=1}^{T(N)} p_i \log_2 p_i$$

where $\sum_{i=1}^{T(N)} p_i = 1$. For equal probabilities, we have

$$S_{sh} = k \log_2 T(N)$$

using (3) we have

$$S_{sh}(N) = k \log_2 T(N) \propto N$$

and the Shannon entropy results extensive.

Let us assume a system whose elements are strongly correlated in such a way that the number of equally probable admissible configurations (configurations whose probability are non-zero) holds

$$T(N) \propto N^\rho \quad (4)$$

when $N \to \infty$ and $\rho > 0$. In this case is not adequate using Shannon’s entropy since $S_{sh}(N) \propto \log_2 N$ and it violates Thermodynamics. Instead of we use the following generalized expression (see [39])

$$S_q = k \left[ \frac{\sum_{i=1}^{T} p_i^q}{q - 1} \right] = -k \sum_{i=1}^{T} p_i^q \log_2 p_i = k \sum_{i=1}^{T} p_i \log_2 \left( \frac{1}{p_i} \right)$$

with $\sum_{i=1}^{T} p_i = 1$. where we have used that $\log_2 z = \frac{z^{\frac{1}{q}} - 1}{q}$ and taken $\log_2 z = \log z$. It is evident that $S_1 = S_{sh}$ (shannon entropy) and that for equal probabilities is

$$S_q = k \log_2 T$$

Using (4) we obtain

$$S_q(N) = k \left[ \frac{T(N)^{1-q} - 1}{1-q} \right] \propto N^{\rho(1-q)}$$

Therefore, choosing $q = 1 - \frac{1}{\rho}$ we obtain that $S_{\frac{1}{\rho}}(N) \propto N$ and this is in accordance with Thermodynamics. As a consequence, Shannon’s entropy can be replaced by Tsallis notion having extensivity, changing adequately the index parameter, when there exists strong correlations in the system.

At this point it is necessary to distinguish between the properties of non-extensivity and non-additivity. A notion of entropy $S$ is additive with respect to a variable, if for two probabilistic independent system $A$ and $B$ it is held $S(A \cup B) = S(A) + S(B)$. Using the expression of Tsallis entropy, we have

$$\frac{S_q(A \cup B)}{k} = \frac{S_q(A)}{k} + \frac{S_q(B)}{k} = (1-q) \frac{S_q(A) S_q(B)}{k}$$

which means that $S_{sh}$ is additive but $S_q$ is non-additive for $q \neq 1$.

In the case of having a continuous random variable $X$ with density of probability distribution function $p(x)$ in a domain $D \subseteq \mathbb{R}^3$, it is possible to obtain expressions for the Tsallis’s entropy, (see [39], [40]).

$$S_q = k \left[ \frac{1}{q - 1} \right] \int_D [p(x)]^q q \log_2 p(x) dx = -k \int_D [p(x)]^q \log_2 p(x) dx = k \int_D [p(x)]^q \log_2 \left( \frac{1}{p(x)} \right) dx$$

where it is held $\int_D p(x) dx = 1$.

Phenomena such that the distribution of motion of cold atoms in dissipative optical lattices, fluctuations of the magnetic field in the solar wind, etc (see [39], [40]) are examples where there is internal strong
correlations. When \( p(x_i) = 0 \) for some \( i \), then the corresponding summand \( 0 \log 0 \) is taken to be 0 consistently with

\[
\lim_{p \to 0^+} p \log(p) = 0
\]

\section{Kolmogorov-Sinai entropy}

One of the key notions in Ergodic Theory of Dynamical Systems was introduced and developed by Andrei Nikolaevich Kolmogorov in 1953 [30] and complete and improved by students of him, by Sinai, Rokhlin, Pinsker and Abramov years after during the period from 1958 till 1962. It is currently known as Kolmogorov-Sinai entropy (KS-entropy or simply KS). It refers to Shannon entropy but differs greatly because it is introduced in a different setting, that of measure-preserving transformations on Dynamical Systems.

It is the entropy of a transformation \( f \) which preserves the measure in a probabilistic space given by the triad \( (X, \mathcal{B}, m) (m(f^{-1}(B)) = m(B), \ B \in \mathcal{B}) \). It is introduced in three steps

1. Entropy of a finite sub-\( \sigma \)-algebra of \( \mathcal{B} \)
2. Entropy of a transformation \( f \) in relation to a finite sub-\( \sigma \)-algebra of \( \mathcal{B} \)
3. Finally, we can obtain the entropy of \( f \)

\textbf{Definition 5.1.} Let \( \mathcal{A} \) be a finite sub-algebra of \( \mathcal{B} \) and \( E(\mathcal{A}) = \{A_1, \ldots, A_n\} \). The entropy of \( \mathcal{A} \) is the number

\[
H(\mathcal{A}) = - \sum_{i=1}^{n} m(A_i) \log m(A_i)
\]

\( H(\mathcal{A}) \) is a measure of the avoid uncertainty (or the gained information) when it is delivered an experiment whose results resultados are \( \{A_1, \ldots, A_n\} \)

\textbf{Definition 5.2.} Let \( f : X \to X \) be a transformation that preserves the measure in a probability space \( (X, \mathcal{B}, m) \).

If \( \mathcal{A} \) is a finite sub-\( \sigma \)-algebra of \( \mathcal{B} \), then

\[
h(f, \mathcal{A}) = \lim_{n \to \infty} \frac{1}{n} H(\bigcup_{i=0}^{n-1} f^{-i} \mathcal{A})
\]

is called the entropy of \( f \) related to \( \mathcal{A} \)

Looking for a meaning, we can think on \( f \) as the change of one day to the next. Then \( \bigcup_{i=0}^{n-1} f^{-i} \mathcal{A} \) represents the combined relationship of an experiment, which is represented by \( \mathcal{A} \), during \( n \) consecutive days.

Let \( f : X \to X \) be a preserving measure transformation on the measure of a probabilistic space \( (X, \mathcal{B}, m) \), then

\[
h(f) = \sup_{\mathcal{A}} h(f, \mathcal{A})
\]

where the supremum is taken on all sub-algebras \( \mathcal{A} \) of \( \mathcal{B} \) is called the entropy of \( f \)

If we think of \( f \) as the time of transition of a one day to the next, then \( h(f) \) is the maximum average per día obtained doing dayly the same experiment.

\subsection{On topological entropy}

Let \( (X, f) \) be a topological dynamical system, that is, let \( X \) be a non-empty Hausdorff space and \( f : X \to X \) a continuous map. In this setting the topological entropy is a number that measures the complexity of the system,
that is how complex is the behavior of all orbits of all points in the space. Approximately it measures the
rate of exponential increasing of the number of orbits that can be distinguishable when time goes on.

The original definition of topological entropy was introduced by Adler, Konheim and McAndrew (see
[1]). Their idea was to assign a number to every cover of a compact space measuring its size and was clearly
inspired in a Kolmogorov-Thomirov’s paper of 1961. Once clarified such initial notion, the rest is an imitation
of the Kolmogorov-Sinai’s definition.

If the space $X$ is metric not necessarily compact, a new definition was introduced by Bowen in 1971 and
independently by Dinaburg in 1970. Such definition uses the notion of $\varepsilon$-separate points. When the space is
compact, Bowen’s definition coincides with Adler, Konheim and McAndrew one. Here we do not go through
formulas since they have been introduced in several papers and books, for example [37], [27] and [17].

What is really interesting and relevant is the relationship between topological entropy and Kolmogorov-
Sinai entropy. It is called a Variational Principle that was formulated around 1970 by Dinaburg, Goodman and
Goodwin.

Let $f : X \rightarrow X$ be a continuous map of a compact metric space to itself. Then

$$h(f) = \sup \{ h_\mu(f) : \mu \in M(X, f) \}$$

where with $M(X)$ we denote the set of all Borel measures on $X$ and $M(X, f)$ the subset of $M(X)$ composed of
all measures on $X$ for which $f$ preserves the measure.

The following properties supplies a method to choose elements of $M(X, f)$.

Let $X$ be a compact metric space, then it is said of $\mu \in M(X, f)$ that it is a measure of maximal entropy of
$f$, if

$$h_\mu(f) = h(f)$$

If now we denote by $M_{\text{max}}(X, f)$ the set of all measures of maximal entropy of $f$, then we have
1. If $h(f) < \infty$, then $M_{\text{max}}(X, f)$ is composed exactly of all ergodic measures.
2. If $h(f) < \infty$ and $M_{\text{max}}(X, f) \neq \emptyset$, then $M_{\text{max}}(X, f)$ contains some ergodic measure.
3. If $h(f) = \infty$, then $M_{\text{max}}(X, f) \neq \emptyset$
4. If the function of entropy, $h(f)$ is upper semicontinuous, then $M_{\text{max}}(X, f)$ is non-empty and compact.
5. When $(X, \sigma)$ is a subshift (a symbolic systems), that is, $\sigma$ is the shift application of a closed and invariant
shift by the application composed by symbols of a finite alphabet, then

$$h(\sigma) = \lim_{n \rightarrow \infty} \frac{\log N(\mathcal{B}_n)}{n}$$

where $\mathcal{B}_n$ denotes the family of all words of length $n$ of $X$
6. In subshifts of finite type, the topological entropy coincides with the logarithm of spectral radius of the
transition matrix la matriz.
7. For piecewise real monotone maps, $f$ is

$$h(f) = \lim_{n \rightarrow \infty} \frac{\log c_n}{n}$$

where $c_n$ denotes the number of pieces of monotonocity of $f^n$.
8. In most cases introduced by Bowen, the topological entropy can be computed counting the of distinct
periodic orbits of period $n$ and then taking the upper limit after taking logarithms and divide by $n$.

Let $F : X \rightarrow X$ a continuous transformation of a compact metric space $(X, d)$. Let $C(X, \mathbb{R})$ denote the Banach
algebra of real-valued continuous functions with the supremum norm. The topological pressure of $F$ is a map

$$P(F, \cdot) : C(X, \mathbb{R}) \rightarrow \mathbb{R} \cup \{ \infty \}$$

having good properties relative to the structures of $C(X, \mathbb{R})$. It contains topological entropy in the sense that
$P(F, 0) = h(F)$ where $0$ denotes the zero function of $C(X, R)$. 

Here, a generalization of the Variational Principal can be proved and sometimes it gives a natural way of choosing important members of $M(X,f)$. In this theory ideas from mathematical statistical mechanics are used and the theory has another applications in differentiable dynamical systems.

For $g \in C(X,\mathbb{R})$, $n \geq 1$, and $\epsilon > 0$ put

$$P_n(F, g, \epsilon) = \sup \{ \sum_{x \in E} e^{S_n(x)} : E \text{ is a (n, } \epsilon \text{) separated subset of } X \}$$

where

$$(S_n(g))(x) = \sum_{i=0}^{n-1} g(F^i x)$$

Finally

$$P(F, g, \epsilon) = \limsup_{n \to \infty} \log P_n(F, g, \epsilon)$$

In the case of symbolic system, the entropy can be interpreted in terms of information content (see \cite{?}.)

Let $(X, \sigma)$ be a subshift of finitely many symbols with entropy, $h(\sigma)$. Let $M$ be the smallest integer strictly larger than $2^{h(\sigma)}$. Then $(X, \sigma)$ is conjugate (via a sliding block code) to a subshift $(Y, \sigma)$ on $M$ symbols and $M$ is the smallest such number (with the exception when $(X, \sigma)$ is conjugate to the full shift on $M$ symbols, in which case is $M = 2^{h(\sigma)}$)

In other words, $h(\sigma)$ informs on the minimal number of symbols sufficient to encode the system in real time (that is, without rescaling the time). If the original subshift uses some $Q > M$ symbols, it can be compressed (by reducing the alphabet to $M$ symbols) without loss of information.

Interpretations of topological entropy in other settings are

1. **Topological entropy for flows**
   Given a flow: $\phi : \mathbb{R} \times X \to X$, topological entropy is defined as the entropy of the time-one map

$$h(\phi) = h(\phi^1)$$

since in flows is used better the notion of equivalence than conjugacy, which admits rescaling of time. Topological entropy essentially distinguishes among flows with zero, positive finite and infinite entropy.

2. **Topological entropy for actions** of more general groups, like $\mathbb{Z}^d$, $\mathbb{R}^d$ and more generally amenable groups or even amenable semigroups. Variational Principle holds in those cases.

### 5.2 Entropy for non-autonomous systems

Such systems are described by the pair $(X, f_\infty)$ where $f_\infty = (f_n)_{n=0}^\infty$ and $f_n \in C(X, X)$ for all $n$ and $X$ is a metric space. With this formulation, they were introduced in the nineties by Kolyada and Snoha in \cite{29} and currently it is an active topic of research.

It is convenient for next definitions to introduce the notation

$$f_i^n = f_{i+(n-1)} \circ f_{i+(n-2)} \circ \ldots \circ f_{i+2} \circ f_{i+1} \circ f_i$$

with $i \geq 0$, $n > 0$ and $f_i^0$ = Identity on $X$ and $Tr_{f_\infty}(x_0) = (f_0^n(x_0))_{m=0}^\infty = (x_n)_{n=0}^\infty$

For any pair $x, y \in X$ and $n \geq 0$ we define

$$\rho_n(x, y) = \max_{i=0,\ldots,n-1} d(f_i^n(x), f_i^n(y))$$

A set $E \subset X$ is $(n, \epsilon, f_\infty)$-separate if $\rho_n(x, y) > \epsilon$ for every pair of distinct points $x, y \in E$.

The topological entropy of a non-autonomous system, $h(f_\infty)$ is given by

$$\lim_{\epsilon \to 0} \left( \limsup_{n \to \infty} \frac{1}{n} \log(s_n(f_\infty, \epsilon)) \right)$$
where $s_n(f_\infty, e)$ denotes the maximal cardinality of $(n, e, f_\infty)$-separate sets.

When $f_n = f$ for all $n$ the former formula is the Bowen’s expression for it.

The following result was proved by Kolyada-Snoha in their broader paper [29].

Let $X$ be a compact metric space and $(X, f_\infty)$ verifies that $f_i$ converges uniformly to a map $f : X \to X$, then is $h(f_\infty) \leq h(f)$.

The next result shows that uniform convergence is an essential assumption in the former result. For every $f \in C(I)$ there exists a non-autonomous system $(I, f_\infty)$ such that $f_i$ converges point-wise to $f$ and $h(f_\infty) = \infty$ (see [4]).

C.Kawan in [28], has introduced the notion of metric entropy for such systems, generalizing the classic metric entropy of Kolmogorov-Sinai. It is related via a variational inequality to the topological entropy introduced before. Moreover, such metric entropy, shares several properties with that of Kolmogorov-Sinai. In particular it is invariant with respect to appropriately defined isomorphisms, a power rule and a Rokhlin-type inequality.

1. Sequence topological entropy
2. Topological entropy in the setting of fuzzy spaces
3. Algebraic topology

6 Topological entropy for discontinuous maps

The topological entropy is difficult to compute if we use the former definition or other equivalent, for example, that defined by $(n, e)$–spanning sets (see [12]). Independently of the definition, in all cases to compute strictly $h(f)$ we must calculate a limit.

When we deal with discontinuous maps, the former approaches are not possible and it is necessary to introduce new notions of entropy in such a way that when applied to the continuous case, coincide with the former notions. In the cases we are considering in this paper, we always suppose that it will be possible to construct a symbolic dynamical system associate to the original system with discontinuities, where the influence of such discontinuities has to be taken into consideration. This can be seen by the existence of the forbidden set of the system, that is, the set of preimages of all points of discontinuity. The orbits of such points are not defined and as a consequence must be excluded of the symbolic consideration. Then the corresponding shift space is a closed subset of a full shift of a finite of symbols [31] which will be the usual situation we deal.

Now we are explaining such process. Let $\mathfrak{A}$ be a finite set of symbols and $\mathfrak{A}^\mathbb{Z} = X$ be the full-shift on it (the set of all bi-sequences constructed with the symbols of $\mathfrak{A}$). Consider now the shift given by $(X, \sigma_X)$ and the corresponding shift system $(X_\mathfrak{A}, \sigma_X | X_\mathfrak{A})$ where $X_\mathfrak{A} \subset X$ is closed and the bi-sequences belonging to it does not contains the blocks (words) associate to the forbidden set of the system. In most cases, it is easier to describe a shift space by specifying which finite blocks of symbols $(n$-blocks) are allowed, rather than which are forbidden.

If $x \in \mathfrak{A}^\mathbb{Z}$ and $\omega$ is a block over $\mathfrak{A}$, then $\omega$ occurs in $x$ if there are indices $i < j$ so that $\omega = x_i x_{i+1} \ldots x_j$. Let us denote by $\mathcal{B}_n(X)$ the set of all $n$-blocks that occur in points of $X$. $|\mathcal{B}_n|$ means the number of $n$-blocks which appear in points of $X$ and gives us an idea of the dynamical complexity of $X$. Instead of using the numbers $|\mathcal{B}_n|$ for $n \in \mathbb{N}$ it is better to consider their growth rate. In [31] it is proved that $|\mathcal{B}_n|$ behaves approximately like $2^{kn}$ being the constant $k$ such growth rate. It is also proved that $k$ can be computed approximately by $(1/n) \log_2 |\mathcal{B}_n|$ when $n$ is large (from now the $\log_2$ will be denote simply by $\log$). In the following we introduce a definition of entropy very useful in the setting of shift spaces.

Let $X$ be a shift space. The entropy of $X$ is

$$h(X) = \lim_{n \to \infty} \frac{1}{n} \log |\mathcal{B}_n(X)|$$
When \( X \) is a shift space with shift map \( \sigma_X \), then the topological entropy \( h(\sigma_X) \) equals \( h(X) \). To see it, consider the shift space \( X \) with the metric \( \rho(x, y) = \sum_{n \in \mathbb{Z}} \frac{|x_n - y_n|}{2^n(1 + |x_n - y_n|)} \) for \( x, y \in X \). If we apply the formula of topological entropy given before to \( \sigma_X \), it is easy to see that \( r_n(\sigma_X, 2^{-k}) = |\mathcal{B}_{n+2k}(X)| \). From this and such formula we obtain that \( h(\sigma_X) = h(X) \).

Further we are considering some different cases of real discontinuous maps and recall different techniques used to give sense to the computation of the entropy.

### 6.1 On Lorenz maps

In what follows, we will denote the unit real interval \([0, 1]\) by \( I \). Lorenz maps can arise as when we consider return maps to a cross section of semiflows on two dimensional branched manifolds. The point \( c \) appears when we conclude that there are points from which flow lines never returns to \( I \) (see the description given in [19], [20] or [13]).

**Definition 6.1.** A map \( f : I \to I \) is a Lorenz map if there is a point \( c \in (0, 1) \) such that

1. \( f \) is continuous and strictly increasing in \([0, c)\) and \((c, 1]\)
2. \( \lim_{x \to c^-} = 1 \) and \( \lim_{x \to c^+} = 0 \)

If the set of preimages of \( c \) is dense in \( I \), then under such a topological expanding condition, the map is called **expansive Lorenz map**. When a Lorenz map is not expanding, then it must have **homtervals** (\( H \) is a hominterval if \( f^n|H \) is a homeomorphism for every \( n \in \mathbb{N} \) or equivalently, the kneading sequence is constant (see [33]). It is evident that a Lorenz map is discontinuous in \( c \).

To calculate the entropy, we use a cutting invariant idea (see for example [19]) approach. It is evident that the set composed of \( c \) and all preimages belong to the **forbidden set** (\( \mathcal{F}(f) \)) of these map, e.g., \( y \in \mathcal{F}(f) \) if \( \text{Orb}_f(y) \) is not defined. This happens because the discontinuity of the map.

For each \( n \in \mathbb{N} \) we put

\[
\Lambda_n = \{x \in I : f^n(x) = c\}
\]

given any finite union of closed intervals, \( Z \subset I \) we introduce the \( n \)-th cutting number of \( f \) on \( Z \) (denoted by \( \gamma^n_{\mathcal{F}} \)) by

\[
\text{card}(\Lambda_n \cap \text{Int}(Z))
\]

such formula measure the number of times that \( Z \) is cut up by the discontinuity produced in the iterates of \( f \). If besides \( f(Z) \subseteq Z \) then we introduce an entropy for this case by

\[
h(f|Z) = \log^Z = \log \limsup_{n \to \infty} (\gamma^n_{\mathcal{F}})^{1/n}
\]

where \( s^Z \) is a grow rate of presence of numbers of the forbidden set of \( f \) in \( X \).

This growth rate is the reciprocal of the radius of convergence of the formal power series given by

\[
\gamma^Z = \sum_{n=0}^{\infty} \gamma^n_{\mathcal{F}} t^n
\]

which it is called the cutting invariant of \( f \) on \( Z \). By definition of Lorenz maps, it must be \( s^Z \leq 2 \) for any closed set \( Z \), since each point has in \( I \) at most two preimages.

If in the definition of a Lorenz map, we add \( f(c) = b \) with \( 0 < b < 1 \), then the forbidden set is empty since the images of any \( x \in I \) exist. Nevertheless the computation of the entropy is the same than in the previous case, since the shift space to be considered is the same when taken into account the discontinuity.

In the setting of Lorenz maps, more things have been obtained. For continuous unimodal maps, the topological entropy is related to the smallest positive zero \( s \) of the kneading invariant of \( f \) by the formula

\[
h(f) = \log(1/s)
\]
(see for example [31]) and for each positive entropy basic set $\Omega_i$ in the renormalization decomposition of the non-wandering set of $f$, there is a real zero $s_i$ of the kneading invariant such that $h(f|\Omega_i) = \log(1/s_i)$. In the setting of Lorenz maps we obtain a similar result, except in the fact that it is possible to obtain two basis set in the normalization decomposition having the same entropy. In this case it is proved ([19]) that it appears a double zero in the kneading invariant.

As an easy example let us define the map

$$f(x) = \begin{cases} 2x, & \text{if } 0 \leq x < \frac{1}{2}; \\ 2x - 1, & \text{if } \frac{1}{2} < x \leq 1 \end{cases}$$

is a Lorenz map but it is not expansive since the preimages of $\frac{1}{2}$ is not dense in $I$, in other words, $\beta(f)$ is composed of two convergent sequences of points convergent to 0 and $\frac{1}{2}$. It is easy to see that in such case, using the cutting method, $h(f) = \log 2$. Using the symbolic approach, the shift space of this problem is the full shift space for three symbols $\{L, C, R\}$.

In [13] has been introduced a class of Lorenz maps called contractive Lorenz maps whose dynamics is different to the expansive case and more complicated. A graph of one such maps can be seen in Table 1. The above paper is a deep analysis of the dynamics of such maps including the way in which the nonwandering set associate to them decomposes.

It is a pending task to compute the entropy of contractive Lorenz maps in the line of the Kneading Theory, completing the results of [19].

6.2 Other discontinuous maps

Now we introduce some families of non-Lorenz maps, even with two points of discontinuity.

6.2.1 Piecewise parabolic and piecewise linear real maps

Let $f : I \to I$ be the following family of discontinuous piecewise parabolic maps

$$f(x) = \begin{cases} 4x(1 - x), & \text{if } 0 \leq x < \frac{1}{2}; \\ rx(1 - x), & \text{if } \frac{1}{2} < x \leq 1; \ (2 \leq r < 4) \end{cases}$$

Since the members of such family are not Lorenz maps the dynamics behavior are different of them and the same for the method to calculate the topological entropy.

Using the kneading theory of [33] several algorithms have appeared to calculate the topological entropy for continuous piecewise monotone maps on the interval (see [16], [6], among others). In the case of piecewise linear Markov maps, can be seen ([5]) that the topological entropy is the maximal eigenvalue of a induced matrix composed only of zeros and one’s related to the intersection or not of the images of all subintervals of the Markov partition. Such idea allows us to compute by approximation by linear Markov maps, the topological entropy of general piecewise monotone maps. In fact if $f$ is piecewise monotone and $(g_n)_{n=1}^\infty$ a sequence of piecewise linear Markov maps, each with the same pieces of monotonocity than $f$ and converging to it in the $C_0$-topology, then in [23] is proved that

$$\lim_{n \to \infty} h(g_n)$$

When $f : I \to I$ is discontinuos with a finite number of discontinuity points, then the definition of entropy for such map the following, introduced in [23]. $f$ has topological entropy $h(f)$ if there is a convergent sequence of piecewise linear Markov maps $(g_n)_{n=1}^\infty$, all with the same number of pieces of continuity than $f$ and converging to $f$ in the $C_0$-topology if $(\log \lambda_n)_{n=1}^\infty$ converges to $h(f)$, where every $\lambda_n$ is the maximal eigenvalue of the induced matrices associated to $g_n$.

At this point, it is necessary to prove that the above notion is well defined and that under the dynamical point of view it is invariant by topological conjugations. Both things are made in [23]. In the case of piecewise
parabolic maps, the piecewise Markov maps used as approximations must be also discontinuous. One possibility is the construction of the maps shown in center of . In each step, the topological entropy is given by the exponential rate of increase with n of the monotone upper bounds of any of these quantities:
1. the number of turning points of \( f^n \)
2. the length of the graph of \( f^n \) and
3. the number of periodic points of \( f^n \)

(see [32]).

A paradigmatic family is that of tent maps, given by:

\[
f(x) = \begin{cases} 
  r(x-1), & \text{if } 0 \leq x \leq \frac{r-1}{r}; \\
  r(1-x), & \text{if } \frac{r-1}{r} \leq x \leq 1 
\end{cases}
\]

applying the length of the graph of \( f^n \) the topological entropy is \( \pm \log r \).

In [8] it is developed an algorithm of fast convergence and provides upper and lower bounds of the topological entropy. This method can be used in the coverage by Markov maps of general piecewise monotone maps including the discontinuous cases. The method is based in the construction of topological transfer matrices \( M_f \) associate to the map \( f \).

Let \( (x_0 = 0 < x_1 < \ldots < x_{p+1} = 1) \) a partition of the unit interval. The interval \([x_0, x_{p+1}]\) is divided into subintervals \( I_i = [x_i, x_{i+1}] \) with \( i = 0, 1, \ldots, p \). While the subintervals \( I_i \) are not necessarily laps, some of them can contain more than one lap. Then we add the turning point to the former partition and the notation will be \( I_i \) for the \( i \)th lap contained in \( I_i \). In [8] it is introduced the term \( m_{ij} \) as

\[
m_{ij} = \frac{|f(I_i) \cap I_j|}{|I_i|} \tag{3}
\]

where \(|I_i|\) denotes the length of the interval \( I_i \), \( m_{ij} \) means the proportion in which the image of \( I_i \) covers \( I_j \).

In the case of the family sawtooth maps with slopes \(+(-)\) it is immediate that the matrix \( M = (m_{ij}) \) has as eigenvalue \( s \) which is the antilogarithm of the topological entropy corresponding to the partition of \( I \) given by the three intervals \( I_1, I_2 \). This conclusion comes from the application of Perron-Frobenius theorem that such eigenvalue is maximal (see [18]).

In most cases, the matrix \( M \) can be complicated due to multiple covering and getting the maximal eigenvalue could be hard. One easier case is when the partition is Markov or when we have a Markov piecewise monotone map and a Markov partition can be chosen composed of the turning points and their orbits. In this case the terms \( m_{ij} \) of \( M \) are only zeros or one’s and looking for the maximal eigenvalue is less difficult. In this case it is known (see for example [25]), in cases of having points of finite discontinuities, such points have to be elements of the partition.

When we are dealing with non-markovian maps, in (3) we can have elements \( m_{ij} \) different of zero and one and most of them having such values. The effect of the former computation is that the \( \max 0, \log, E \) where \( E \) denotes the largest eigenvalue of \( M \) will be an approximation of the topological entropy having an error. Such error decreases with the number of subintervals of the partition. But what is more interesting is that we can introduce some bounds of the entropy, choosing adequately the partition.

To construct it, we take the subcolumn \((m_{ij}, m_{i+1,j}, \ldots, m_{i+k,j})^T\) of length \( k \), change it by a \( k \)-simplex, choosing a set of points \((P_0, \ldots, P_k) \in \mathbb{R}^{k+1}\) holding \( P_n \geq 0 \) and \( \sum_{n=0}^{n=k} |P_n| = 1 \). The end points of the simplex are the points \( P_0 = (1, 0, \ldots, 0)^T, P_1 = (0, 1, \ldots, 0)^T, \ldots, P_k = (0, 0, \ldots, 1)^T \). For \( k = 0 \) we choose as \( P_0 = (0, 0, \ldots, 0) \) and \( P_1 = (1, 1, \ldots, 1) \).

Applying this procedure, we replace each subcolumn of \( M \) following the above rules choosing adequately the points \( P_n \). Then compute the largest eigenvalue of all matrices so obtained. We get a set of numbers whose extrema give us lower and upper bounds of the largest eigenvalue of \( M \) given by the application of the Perron-Frobenius theorem. As a consequence of the Milnor-Thurston kneading theory ([33]), such extrema will be in turn, upper and lower bounds of the entropy. Such bounds can be improved using sufficiently long orbits of the turning points of the partitions we have considered, including in the partition the discontinuities points of the map (in case of having them). In all cases, we must justify the adequate procedures to be consistent.
In references [3], [6], [8], [16], [23], [25], [31] [34] it is possible to see computations in different particular examples, using the former methods.

### 6.2.2 Discontinuous maps with more than one point of discontinuity

In the more general setting we can have maps with a finite number of discontinuity points of finite jump and a finite number of turning points. The procedures of the former paragraph can be applied but the computation can be very difficult. Therefore some simplifications are valuable, for example on the line of what is made in [3] for the case of continuous maps with a finite number of turning points. If symbolic methods are used, always the discontinuity points must be points the corresponding partitions. For a treatment of maps with odd discontinuous points can be seen [34] and for the graph of an example of symmetric discontinuous for polynomialwise maps see in Figure 1. It is a pending to introduce an adequate notion of entropy for the case of maps with discontinuities of infinite with infinite jumps.

### 6.2.3 Transitive discontinuous real maps

It is well known that if a continuous interval map \( f : I \rightarrow I \) is transitive (recall that it is the case when for every pair of open sets \( U, V \), there exists \( n \in \mathbb{N} \) such that \( f^n(U) \cap V \neq \emptyset \), then \( h(f) \geq \log_2 2 \) (see [7]). The proof of such result is based in the fact that when \( X \) is a Baire separable metric space and \( x \) continuous, transitivity implies the existence of a point \( x \in X \) whose orbit is dense in \( X \). When \( f \) is discontinuous we can not assure the validity of the equivalence. Nevertheless, we have the following result, whose proof can be seen in [35].

Let \( X \) be a Baire separable metric space and \( f : X \rightarrow X \) be a transitive map with only a unique point of discontinuity. Then there is \( x \in X \) such that \( \text{Orb}_f(x) \) is dense in \( X \).

Even in [35] it is proved that it is possible to construct counterexamples with transitive maps having two points of discontinuity but for which the above property of having a point of dense orbit is not fulfilled.

### 7 Conclusions

In this paper we have reviewed the physics origins of the notion of entropy and its different meanings mainly throughout the work made by Clausius and Boltzmann. Keeping such ideas in mind was introduced the Shannon and Kolmogorov-Sinai entropies in the mathematical setting and then the important notion of topological entropy and its variants. In last paragraphs we have stated the necessity of extending such notions to non-autonomous systems and discrete dynamical systems described by means of discontinuous maps.

**Acknowledgement:** The research has been supported by the Proyecto MTM2014-51891-P from Spanish MINECO and from Proyecto Séneca de la Comunidad Autónoma de la Región de Murcia, 19294/PI/14.

### References


