Physica A 440 (2015) 110–117

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Substitution systems and nonextensive statistics

V. García-Morales *

Institute for Advanced Study, Technische Universität München, Lichtenbergstr. 2a, D-85748 Garching, Germany
Departament de Termodinàmica, Universitat de València, E-46100 Burjassot, Spain

HIGHLIGHTS

- Universal maps are derived for all deterministic 1D substitution systems.
- The map is also valid for non-constant-length substitution systems.
- The connection of substitution systems with nonextensive thermostatistics is pointed out.
- Substitution systems are shown to satisfy a 'Second Law' in the thermodynamic limit.

ARTICLE INFO

Article history:
Received 17 April 2015
Received in revised form 16 June 2015
Available online 11 August 2015

Keywords:
Symbolic dynamics
Fractals
Tilings
Tsallis entropy
Complexity
Irreversibility

ABSTRACT

Substitution systems evolve in time by generating sequences of symbols from a finite alphabet: At a certain iteration step, the existing symbols are systematically replaced by blocks of N_k symbols also within the alphabet (with N_k, a natural number, being the length of the kth block of the substitution). The dynamics of these systems leads naturally to fractals and self-similarity. By using $\mathcal{B}$-calculus (García-Morales, 2012) universal maps for deterministic substitution systems both of constant and non-constant length, are formulated in 1D. It is then shown how these systems can be put in direct correspondence with Tsallis entropy. A 'Second Law of Thermodynamics' is also proved for these systems in the asymptotic limit of large words.

1. Introduction

Substitution dynamical systems [1,2] are of great relevance to many branches of Physics and Mathematics including geometry, combinatorics, chaos and ergodic theory, spectral analysis, number theory, fractals and tilings. The history of substitution systems dates back to 1906 with the following construction by A. Thue [3]

$$a \to ab \to abba \to abbabaab \to abbabaabbaababba \to \cdots. \quad (1)$$

This sequence, later rediscovered by M. Morse [4] and therefore called the Thue–Morse sequence, can be easily obtained by the following iterative process: at each iteration replace each previous a in the sequence by ab and each b by ba. This is probably one of the simplest examples of a substitution system.

Another well-known example is suggested by the familiar construction leading to the Cantor set shown in Fig. 1. This construction is related to a sequence of words containing only '0's and '1's obtained from a substitution system: The replacements that are made at each iteration step t are $0 \to 000$ and $1 \to 101$. At $t = 0$ one starts with '1' and at successive iteration steps one obtains the sequence of words 101, 101000101, 101000101000000000101000101 and so on. Each of these words can thus be mapped to an odd natural number written in the binary base.

* Correspondence to: Departament de Termodinàmica, Universitat de València, E-46100 Burjassot, Spain. Tel.: +49 15224020629.
E-mail address: garmovla@uv.es.

http://dx.doi.org/10.1016/j.physa.2015.07.035
0378-4371/

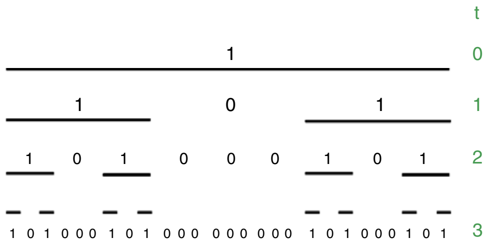

Fig. 1. Construction of the Cantor set through a substitution process. At each iteration step t the original segment is replaced by three segments that are one third shorter each, and the central one is removed. If the segments that remain at each iteration step are labeled by '1' and the segments that are removed are labeled by '0', this process has an associated binary sequence obtained by a simple substitution system.

Both substitution systems are called of *constant length* N, because the blocks of symbols used to replace the previous symbols have all the same size ($N = 2$ in the Thue–Morse sequence and $N = 3$ in the Cantor sequence). Thus, after t iteration steps, the sequence contains N^t symbols. Entire monographs have been dedicated to substitution systems of constant length (see for example Ref. [1]). They are closely related to geometric tiling substitutions [5,6], which are used to construct infinite tilings with a finite number of tile types. Since a single substitution rule is used in every iteration, substitution systems are also naturally related to fractals [7–11]: If each symbol denotes a segment of, say, a given color, and the segment is systematically replaced by N smaller segments that cover the previous one, the figure that will emerge as t grows is expected to exhibit self-similarity, as the example of the Cantor sequence shows.

Whereas there exist an extensive literature of constant-length substitution systems, the ones with non-constant length have also a major importance and are receiving increased interest [5]. The former are related to *geometric* substitution tilings and the latter to *combinatorial* substitution tilings [5]. To give a simple illustration of a substitution system with non-constant length let us, for example, consider the following substitution rule of the symbols a and b. At each step, one introduces the replacements $a \to ab$ and $b \to a$. Then, if we begin with a we get [5]

$$a \to ab \to aba \to abaab \to abaababa \to abaababaabaab \to \cdots. \tag{2}$$

The length of the words is 1, 2, 3, 5, 8, 13, ..., i.e. the Fibonacci sequence. This length is no longer simply equal to N^t at step t as is the case in constant-length substitution systems but is, generally, more complicated. Still, a general expression for these systems can be found.

The outline of this article is as follows. In Section 2 we construct a universal map for constant-length substitution systems. In Section 3 we show how these systems are naturally linked to entropy. Specifically, we show how Boltzmann and Tsallis entropies quite naturally arise from the dynamical evolution. Then, in Section 4 we tackle the nontrivial problem of substitution systems with non-constant length and the article finishes with some considerations on the thermostatistics of such systems.

2. $\mathcal{B}$-calculus and a universal map for substitution systems of constant length

The main idea behind $\mathcal{B}$-calculus is to use a most elementary mathematical structure involving sums and/or products of so-called $\mathcal{B}$-functions as the basic building block to model computational processes [12]. This view proves quite useful to describe any rule-based dynamical system, as cellular automata [12–14]. The $\mathcal{B}$-function for any real numbers x, y is defined as [12]

$$\mathcal{B}(x, y) \equiv \frac{1}{2}\left(\frac{x+y}{|x+y|} - \frac{x-y}{|x-y|}\right) = \frac{1}{2}\left(\mathrm{sign}(x+y) - \mathrm{sign}(x-y)\right) \tag{3}$$

with $\mathrm{sign}(x) \equiv \frac{x}{|x|}$ being the sign function. Since $\mathrm{sign}(0) \equiv 0$, at $x = \pm y$ (singular borders), $\mathcal{B}(\pm y, y) = \frac{y}{2|y|} = \mathrm{sign}(y)/2$ and at the origin $\mathcal{B}(0, 0) \equiv 0$.

Let S be the set of integers in the interval between 0 and $p - 1$, with $p > 1 \in \mathbb{N}$ denoting the alphabet size. We now consider an operator which replaces each possible value of a quantity $x \in S$ by a block of N values $y_0, y_1, \ldots, y_{N-1}$ all in S. Thus, such an operator maps S to S^N. Hence, if h runs from 0 to $N - 1$, the substitution operator is defined as

$$y_h = \mathrm{Subs}_{N;p}(x, h) \equiv \sum_{n=0}^{p-1}\sum_{m=0}^{N-1} a_{m+nN}\,\mathcal{B}\left(h - m, \frac{1}{2}\right)\mathcal{B}\left(n - x, \frac{1}{2}\right) = a_{h+Nx} \tag{4}$$

where all $a_{m+nN} \in S$. The operator is uniquely specified by a Wolfram code $\mathrm{Subs}_{N;p}$ with

$$\mathrm{Subs} \equiv \sum_{n=0}^{p-1}\sum_{m=0}^{N-1} a_{m+nN}\,p^{m+nN} \tag{5}$$

being a non-negative integer.

A substitution system is a substitution operator being iterated on a lattice that is also iteratively refined to accommodate each new output block: Each position on the lattice is labeled by a nonnegative integer $j \in [0, N^{t+1} - 1]$ and has a dynamical