Proceedings of the Institute of Mathematics and Mechanics,
National Academy of Sciences of Azerbaijan
Volume 46, Number 2, 2020, Pages 360–393
https://doi.org/10.29228/proc.37

FRACTAL SIMILARITIES BETWEEN THE DISTRIBUTION OF PRIMES AND NUCLEOTIDES IN DNA

CARLO CATTANI

Abstract. In this paper, a simple method for the analysis of auto-correlation on symbolic-numerical 1D time series is shortly described. This method is based on the definition of a suitable indicator matrix (of correlation) and the corresponding 2D binary image, which is a special kind of recurrence plot. The main parameters of complexity and multi-fractality are defined on the binary images and will be used to single out the main properties of the 1D time series by characterizing the fractal parameters computed on the corresponding 2D images. The short (window) wavelet transform will be also discussed by showing that clusters of wavelet coefficients might be used to single out some more information about the existence of possible hidden rule concerning the sequence distribution. As application of this method, the multifractal analysis of the prime number distribution and nucleotide distribution in DNA will be given to single out their fractal nature and the main similarities and distinctions in between the two.

1. Introduction

In recent years, there has been a rising interest for the statistical analysis of time series which show some fractal characteristic, such as DNA sequences and prime number distributions. DNA sequences were studied by using not only methods from signal analysis and frequency analysis but also based on fractal analysis [21, 22, 36, 39, 2-5,7-14,30-32,41-44,48-53,59-62,67-74,82-97]. Almost all these papers are aiming to single out some hidden characteristics of the time series, mostly related to complexity and multi-fractality, and to visualize the existence of regular (recursive) patterns in the data distributions (see e.g. [1, 2, 6, 7, 32, 36, 62, 91, 94, 9-12,21-24,41-45,48-52,67-69,82-85]). The existence of some hidden rules, might be detected by showing the existence of long range (auto or cross)-correlation on data [3, 5, 8, 11, 13, 14, 30, 31, 50, 59, 60, 61, 73, 74, 88, 89, 90, 91, 95, 97]. However, this method depends on the numerical representation (with or without redundancy), on the length of the sequence, and on the intrinsic nature of data as well. Therefore, some interesting attempts to add more information about data distributions and the existence of patterns

2010 *Mathematics Subject Classification.* 28A80, 42B25, 65T60, 11A41.

Key words and phrases. Binary Image Analysis, Wavelet Analysis, Fractal Analysis, Complexity, DNA, prime numbers.

might be given also by the fractal and wavelet analysis (see e.g. [2, 5, 6, 7, 36, 71, 86, 20-24]). Wavelet theory (see e.g. [18, 33, 75, 92]), due to the localization properties of wavelets, is a powerful method for the local analysis of signals thus being an expedient tool for singling out the dependence of each term of the time series with their closest. Also multifractal nature of the time series [56, 57, 58] can be easily detected by wavelet analysis [92].

An alternative original method for the analysis of time series is to convert the 1D signal into a suitable corresponding 2D image, which preserves the autocorrelation of the elements. It was originally suggested by Eckmann et Al. [37, 40, 65, 66] for the analysis of nonlinear dynamical systems by the so-called recurrence plots. A recurrence plot is simple the graphical representation of a binary map which show the relation (auto-correlation) among elements. On this plot the main parameters of complexity and multi-fractality can be computed and they will be taken as measure of the complexity of the original 1D image. Frequency analysis, it is used to easily visualize some correlations. By using the frequency we can define, for large sequences, same complexity parameters like randomness, fractal dimension, complexity, entropy. These parameters enable us to classify sequences when we compare one to another. Wavelet analysis as well can be done on the 2D image and the qualitative results obtained on this image can be also used to classify the original 1D time-series. The main advantage of this method is that, in the 2D image, there surprisingly appear some typical textures corresponding to each complexity state: chaotic, trend, periodic, disrupted state. Wavelet analysis, due to its localization property, will be used to show that some intrinsic properties of the sequence can be better singled out by analyzing the short wavelet transform coefficients. Some unexpected characteristic of the 1D sequence will become more evident by clustering the wavelet coefficients of 2D representation. We will see that by using fractal analysis (mainly based on the computation of the fractal dimension) and wavelet analysis of the 2D binary images we can obtain some interesting results for time-series, whose behavior and auto-correlation is still unknown such as the prime number distribution and nucleotide distribution in DNA.

This paper is organized as follows: the binary map is defined in section 2 while the main parameters for measuring the complexity of fractal sets will be defined in section 3. Section 4 resumes some fundamentals remarks on wavelet analysis and the segmented wavelet analysis. Binary plots on dynamical systems will be discussed in section 5. Section 6 shows the application of the fractal analysis on the prime number distribution and their representation with Ulam method is given in 7. The fractal analysis of DNA and the distribution of nucleotides is described in section 8. Random walks are discussed in section 9. In the conclusion some comments on future perspectives are given.

2. Indicator function

The existence of patterns or typical distributions in a time series can be singled out by the existence of some auto-correlation among the elements of the sequence. The auto-correlation can be computed by some classical methods and it measures the relationship of an element with the remaining elements of the sequence. A

simple method to visualize the auto-correlation is based on the indicator function and the corresponding correlation matrix as follows.

Let $S = \{x_k\}_{k=1,...,N}$, $T = \{y_k\}_{k=1,...,N}$ be two given sequences and R a binary relation, such that

$$x_h R y_k = \text{TRUE} \vee \text{FALSE} \qquad (h, k = 1, ..., N) \ ,$$

the indicator function (map) is the binary map

$$u : S \times T \to \{0 \ , 1\}$$

such that for $x_h \in S, \ y_k \in T$,

$$u^R_{h,k} \stackrel{\text{def}}{=} u^R(x_h, y_k) = \begin{cases} 1 & \text{if} \quad x_h R y_k \ = \ \text{TRUE} \\ 0 & \text{if} \quad x_h R y_k \ = \ \text{FALSE} \end{cases} \ . \tag{2.1}$$

We call $u^R(h, k)$ the cross-correlation matrix, auto-correlation matrix when $T = S$.

In this case, the indicator function is the symmetric map

$$u : S \times S \to \{0 \ , 1\}$$

such that for $h \in S$, $k \in S$

$$u_{hk} \stackrel{\text{def}}{=} u(h, \ k) = \begin{cases} 1 & \text{if} \quad x_h R x_k \ = \ \text{TRUE} \\ 0 & \text{if} \quad x_h R x_k \ = \ \text{FALSE} \end{cases} \ . \tag{2.2}$$

According to (2.2), the indicator of a N-length sequence can be easily represented by the $N \times N$ sparse symmetric matrix $\{u_{hk}\}$ of binary values $\{0, 1\}$, as the following table, where the relation R is the following

$$x_h R x_k = \text{TRUE iff } x_h \text{ is a prime} \wedge x_k \text{ is a prime}$$

so that for the first 11 numbers we have the table

$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$
11	1	1	0	1	0	1	0	0	0	1	...
10	0	0	0	0	0	0	0	0	0	0	...
9	0	0	0	0	0	0	0	0	0	0	...
8	0	0	0	0	0	0	0	0	0	0	...
7	1	1	0	1	0	1	0	0	0	1	...
6	0	0	0	0	0	0	0	0	0	0	...
5	1	1	0	1	0	1	0	0	0	1	...
4	0	0	0	0	0	0	0	0	0	0	...
3	1	1	0	1	0	1	0	0	0	1	...
2	1	1	0	1	0	1	0	0	0	1	...
u_{hk}	2	3	4	5	6	7	8	9	10	11	...

where both on bottom and on left there is the sequence S, and the composition table is done according to the indicator values u_{hk}.

This table (2.2) can be plot in 2 dimensions (Fig. 1) by putting a dot where $u_{hk} = 1$ and white blank when $u_{hk} = 0$ thus giving rise to the so-called dot-plots, recurrence plots [37, 40], binary images (see e.g. [17, 26, 28] and references therein). Our aim is to characterize the intrinsic properties of the original time

series by analysing the map (2.2), through some parameters of complexity-multifractality computed on the binary image Fig. 1. .

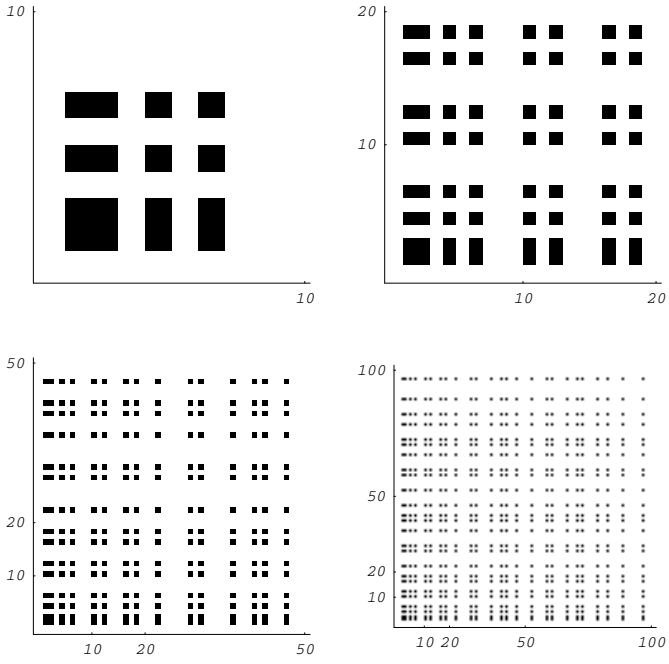

FIGURE 1. *Indicator matrix with* $n \leq 10,\ n \leq 20$ *(top) and* $n \leq 50,\ n \leq 100$ *(bottom) on the distribution of primes.*

3. Parameters of complexity and fractality

In this section, will be reviewed some of the most popular parameters based on frequency distribution, which can measure the complexity of a sequence, (see e.g. [24] and references therein). Their computation can be simplified if done on the binary image thus enabling also the measure of fractality as ratio of filling a 2D space domain.

Let $v_x(n)$ the frequency of the element x among the first n elements of the N-length sequence S, with $n \leq N$ and $p_x(n)$ the corresponding probability, i.e. $p_x(n) = v_x(n)/n$. The probability and the frequency count can be extended to a set of points in $\mathbb{R}^2$ by the ratio

$$p_1(r) = v_1(r)/r^2$$

being r the size of a gliding square.

In particular, for a given n-length sequence, the most popular parameters are as follows (see also [21, 22, 24, 23]).

R: *Randomness*:

$$R \stackrel{\text{def}}{=} 1 - \sigma(\nu_1(n),\ \nu_2(n),\ \ldots,\ \nu_{M_\ell}(n))$$

being σ the variance, so that $R = 1$ for random sequences and $R = 0$ for a non-random sequences

K**:** *Complexity*:

$$K = \frac{1}{n} \log \frac{n!}{\nu_1(n)!\nu_2(n)! \dots \nu_{M_\ell}(n)!}$$

H**:** *Normalized Shannon entropy* [81, 96]: is defined, on a numerable finite set, as

$$H(n) = -\frac{1}{\log \ell} \sum_{w \in A_\ell} p_w(n) \times \begin{cases} \log p_w(n) & \text{if} \quad p_w(n) \neq 0 \\ 0 & \text{if} \quad p_w(n) = 0 \end{cases} \tag{3.1}$$

where $p_w(n)$ should be computed for large sequences. According to Eq. (2.2), we will approximate its value with

$$p_w(n) \cong \frac{1}{n} \sum_{i=1}^{n} u_{wi}, \qquad (w \in S,\ 1 \leq n \leq N) \ .$$

D**:** *Fractal dimension*: is computed on the dot-plot, by the box counting algorithm [11, 12], as the average of the number $p(n)$ of $1's$ in the randomly taken $n \times n$ minors of the $N \times N$ indicator matrix u_{hk}

$$D = \frac{1}{2N} \sum_{n=2}^{N} \frac{\log p(n)}{\log n} \ .$$

D**:** *Lacunarity*: is also computed on the dot-plots, and it is the measure of gaps in the distribution (see e.g. [28] and references therein). It can be easily computed by the ratio of the second and first moment of the distribution

$$\Lambda(r) \ = \ \frac{\sum_{k=1}^{N} [p_r(k)]^2}{\left[\sum_{k=1}^{N} p_r(k)\right]^2}$$

as a function of the gliding box size r on the binary image.

In two dimensions this parameter can be easily computed on the binary image, by using the indicator function on a squared gliding box with r-length side, so that let u_{hk} $(h, k = 1, ...N)$ be the indicator matrix which gives rise to the binary plot. On the binary image we take a squared gliding box with r-length side, so that

$$\mu_r(h,\, k) = \sum_{s=h}^{h+r-1} \sum_{t=k}^{k+r-1} u_{st}$$

is the frequency of "1" within the box. In other words in each square we compute the number of "1". The corresponding probability is

$$p_r(h, k) = \frac{1}{r^2} \sum_{s=h}^{h+r-1} \sum_{t=k}^{k+r-1} u_{st}$$

Then the box moves over the binary image in order to cover different pieces of the image and to obtain the probability distribution

$$\{p_r\,(h,\,k)\}_{\,h,k=1,..,N}$$

As before the lacunarity is defined as a function of the square side, by the ratio of the second and first moment of the distribution

$$\Lambda\,(r) \;=\; \frac{\sigma^2\,[p_r\,(h,\,k)]}{\left[\overline{p_r\,(h,\,k)}\right]^2} \;+\; 1$$

D: *Succolarity*: is computed on the dot-plot, by the box counting algorithm [11, 12], as the average of the number $p(n)$ of $1's$ in the randomly taken $n \times n$ minors of the $N \times N$ indicator matrix u_{hk}.

The succolarity of a fractal set is a parameter which quantifies the capacity of flooding through the set. Thus, succolarity depends on the obstacles along a fixed direction. The succolarity as a fractal measure in image analysis has found some interesting applications [34, 35]. In particular, it gives a simple algorithm to evaluate the succolarity on a binary image.

Following the method suggested in [34, 35] the algorithm is based on the box counting method. However, in order to take into account the physical concept of flow and the corresponding, the pixels of the images have to be rearranged. The image is divided into equal box sizes $BS(k)$ of side length k. The succolarity, along a given direction $\vec{d}$, then is the product of the percentage of the occupied box OP by the pressure PR [34, 35], that is

$$SU\left(BS\,(k)\,,\,\vec{d}\right) \;=\; \frac{\sum\limits_{k=1}^{n} OP\,(BS\,(k))\;PR\,(BS\,(k)\,,\,pc)}{\sum\limits_{k=1}^{n} PR\,(BS\,(k)\,,\,pc)} \tag{3.2}$$

where, OP is the occupation percentage, with respect the full image, k is an index ranging from 1 to n, which corresponds to the number of box. The occupation percentage of the k-th box is $OP(BS(k))$ and $PR(BS(k), pc)$ is the pressure applied on the centroid of the k-th box, being pc the centroid of the box.

Example: For the computation of the succolarity we apply the algorithm (3.2) on the following indicator matrix:

$$\begin{pmatrix}
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \iddots \\
0 & 0 & 0 & 0 & 0 & 1 & \dots \\
1 & 0 & 1 & 0 & 1 & 0 & \dots \\
0 & 0 & 0 & 1 & 0 & 0 & \dots \\
1 & 0 & 1 & 0 & 1 & 0 & \dots \\
0 & 1 & 0 & 0 & 0 & 0 & \dots \\
1 & 0 & 1 & 0 & 1 & 0 & \dots
\end{pmatrix}$$

The flow along the direction from top left corner to bottom right corner changes the previous matrix into the follwoing (only two barriers are shown)

$$\begin{pmatrix} 0 & 0 & 0 & \vdots & 0 & \vdots & \ddots \\ 0 & 0 & 0 & 0 & 0 & 1 & \dots \\ \mathbf{1} & 0 & 1 & 0 & \mathbf{1} & 0 & \dots \\ 0 & \mathbf{1} & 0 & 1 & 0 & \mathbf{1} & \dots \\ 1 & 0 & \mathbf{1} & 0 & 1 & 0 & \dots \\ 0 & 1 & 0 & \mathbf{1} & 0 & 1 & \dots \\ 1 & 0 & 1 & 0 & \mathbf{1} & 0 & \dots \end{pmatrix}$$

so that the pressure is 6. While the flow along the direction of the main diagonal (i.e. bottom left corner to the top right corner), changes the indicator matrix into (only two barriers are shown)

$$\begin{pmatrix} \vdots & \vdots & \mathbf{1} & \vdots & 1 & \vdots & \mathbf{1} \\ 0 & \mathbf{1} & 0 & 1 & 0 & \mathbf{1} & \dots \\ \mathbf{1} & 0 & 1 & 0 & \mathbf{1} & 0 & 1 \\ 0 & 1 & 0 & \mathbf{1} & 0 & 1 & \dots \\ 1 & 0 & \mathbf{1} & 0 & 1 & 0 & 1 \\ 0 & \mathbf{1} & 0 & 1 & 0 & 1 & \dots \\ \mathbf{1} & 0 & 1 & 0 & 1 & 0 & \dots \end{pmatrix}$$

where the pressure is 5 .

The bolded font shows the direction of flow (the first from top to bottom and the second one from bottom to top). From the definition of succolarity we expect that this parameter tends to zero when n largely increases, in fact the flow tends to zero by increasing the number of barriers .

4. Wavelet Analysis

Another expedient method for the analysis of auto-correlation in a sequence is the Wavelet analysis. Focusing on the discrete wavelet transform, this method is based on the interpretation of the variability of wavelet coefficients. These coefficients, in fact, give a description of local abrupt changes and variance. However, since their value decay to zero quite rapidly as the scale factor goes to infinity, it has been proposed a short (window) discrete wavelet transform, where only the first 4 coef-ficients are taken into consideration. Wavelet analysis has been already extensively applied to the analysis of biological signals [5, 7, 60, 97] focussing on the complexity and heterogeneity.

We will consider in the following the Haar wavelet basis (see e.g. [21, 22, 24]) made by the so-called scaling functions:

$$\begin{cases} \varphi_k^n(x) \stackrel{\text{def}}{=} 2^{n/2}\varphi(2^n x - k) \ , \quad (0 \le n \ , \ 0 \le k \le 2^n - 1) \ , \\ \\ \varphi(2^n x - k) = \begin{cases} 1 \quad , x \ \in \Omega_k^n \quad , \\ 0 \quad , x \ \notin \Omega_k^n \ , \end{cases} \qquad \Omega_k^n \stackrel{\text{def}}{=} \left[\frac{k}{2^n}, \frac{k+1}{2^n} \right) \ , \end{cases} \tag{4.1}$$

and by the *Haar wavelets*:

$$\left\{\begin{array}{l} \psi_k^n(x) \stackrel{\text{def}}{=} 2^{n/2}\psi(2^n x - k)\ , \quad ||\psi_k^n(x)||_{L^2} = 1\ , \\ \\ \psi(2^n x - k) \stackrel{\text{def}}{=} \left\{\begin{array}{ll} -1\ , & x \in \left[\dfrac{k}{2^n}, \dfrac{k+1/2}{2^n}\right)\ , \\ 1\ , & x \in \left[\dfrac{k+1/2}{2^n}, \dfrac{k+1}{2^n}\right)\ , \quad (0 \le n\ ,\ 0 \le k \le 2^n - 1)\ , \\ 0\ , & \text{elsewhere}\ . \end{array}\right. \end{array}\right. \tag{4.2}$$

The *discrete Haar wavelet transform* is the $N \times N$ matrix $W^N : \mathbb{K}^N \subset \ell^2 \to \mathbb{K}^N \subset \ell^2$ which maps the time series $\boldsymbol{Y}$ into the vector of *wavelet coefficients* $\boldsymbol{\beta}_N = \{\alpha\ ,\ \beta_k^n\}$:

$$\left\{\begin{array}{rcl} W_N \boldsymbol{Y} & = & \boldsymbol{\beta}_N \\ \\ \boldsymbol{\beta}_N & \stackrel{\text{def}}{=} & \{\alpha, \beta_0^0, \ldots, \beta_{2^{M-1}-1}^{M-1}\}\ , \\ \\ \boldsymbol{Y} & \stackrel{\text{def}}{=} & \{Y_0,\ Y_1,\ \ldots,\ Y_{N-1}\}\ , \quad (2^M = N)\ . \end{array}\right. \tag{4.3}$$

The matrix W_N can be easily computed by some recursive product [15, 16, 21, 22, 24] so that with $N = 4\ , M = 2$, we have [21, 22, 24]

$$W_4 = \left(\begin{array}{cccc} \dfrac{1}{2} & \dfrac{1}{2} & \dfrac{1}{2} & \dfrac{1}{2} \\ \\ -\dfrac{1}{2} & -\dfrac{1}{2} & \dfrac{1}{2} & \dfrac{1}{2} \\ \\ -\dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} & 0 & 0 \\ \\ 0 & 0 & -\dfrac{1}{\sqrt{2}} & \dfrac{1}{\sqrt{2}} \end{array}\right)\ . \tag{4.4}$$

From Eq. (4.3) with $M = 2, N = 4$, by explicit computation, we have

$$\alpha = \frac{1}{4}\left(Y_0 + Y_1 + Y_2 + Y_3\right)$$

and [19, 20, 21]

$$\left\{\begin{array}{rl} \beta_0^0 & = \dfrac{1}{2}\left(Y_2 - Y_0 + Y_3 - Y_1\right) \\ \\ \beta_0^1 & = \dfrac{1}{\sqrt{2}}\left(Y_0 - Y_1\right)\ , \\ \\ \beta_1^1 & = \dfrac{1}{\sqrt{2}}\left(Y_3 - Y_2\right)\ . \end{array}\right.$$

So that the first wavelet coefficient α represents the average value of the sequence and the other coefficients β, also called detail coefficients, are strictly connected with the first order properties (jumps and variance) of the discrete time-series.

This analysis, performed in 1D-sequences can be extended to 2D-sequences as follows. Let $\boldsymbol{X},\boldsymbol{Y}$, be two N-length 1D sequences and let us define the p-parameter short wavelet transform which consists in the subdivision of the given sequence into p-length segments and apply the wavelet transform to each segment [19, 20, 21, 22, 23, 24]. With $p = 4$, from the $N = 2^M$-length complex vector $\boldsymbol{Y}$, which is subdivided into 2^{M-2} segments, the 4-parameter short Haar wavelet transform, locally performed by the matrix (4.4), gives the complex vector

$$W_4(\boldsymbol{Y}) = \boldsymbol{\beta} + i\boldsymbol{\beta}^*$$

with

$$\boldsymbol{\beta} = W_4\Re(\boldsymbol{Y}) = \left(\alpha\ ,\beta_0^0\ ,\beta_0^1\ ,\beta_1^1\right) \quad , \quad \boldsymbol{\beta}^* = W_4\Im(\boldsymbol{Y}) = \left(\alpha^*\ ,{\beta^*}_0^0\ ,{\beta^*}_0^1\ ,{\beta^*}_1^1\right)\ .$$

From there we obtain the cluster of points

$$(W_4\Re(\boldsymbol{Y}^s), W_4\Im(\boldsymbol{Y}^s))\,,\ s = 0,\ldots,2^M-2$$

in the 4-dimensional space

$$(\alpha\ ,\alpha^*)\times\left(\beta_0^0\ ,{\beta^*}_0^0\right)\times\left(\beta_0^1\ ,{\beta^*}_0^1\right)\times\left(\beta_1^1\ ,{\beta^*}_1^1\right)\ .$$

This algorithm enables us to construct clusters of wavelet coefficients and to study the correlation between the real and imaginary coefficients [21, 22, 24].

5. Recurrence plots for dynamical systems

In the qualitative analysis of a dynamical system a major roe is played by the computation of some physically meaningful statistical parameters such as the information dimension, entropy, Liapunov parameters and some more parameters related to complexity. These parameters might give some characterizing properties of the dynamical system that can help to classify the nature of the dynamical system. In [37] it was shown that this characterization can be realized by a simple binary plot (recurrence plot) associated with the orbit which describes the dynamical system. In doing so the analysis of the orbit (one-dimensional time series) is transferred into the analysis of a two-dimensional binary image. As stated in [37] "the information obtained from recurrence plots is often surprising, and not easily obtainable by other methods"

Let us assume, without restrictions, that the dynamical system is a one-dimensional system so that the solution is the n-length sequence $S = \{x_k\}_{k=1,\ldots,n}$ representing the solution $x(t)$ at time $t = k$, $(k = 1,\ldots,n < \infty)$. For a given distance δ we can define the boolean operator (indicator function)

$$u^\delta : S\times S \to \{0\ ,1\} \quad , \quad \delta > 0$$

such that for $x_h \in S$, $x_k \in S$

$$u_{hk}^\delta \overset{\text{def}}{=} u^\delta(x_h,\ x_k) = \begin{cases} 1\ , & \text{if} \quad |x_h - x_k| \le \delta \\ 0\ , & \text{if} \quad |x_h - x_k| > \delta\ . \end{cases} \tag{5.1}$$

Since $h,\ k$ correspond to two time instances, the recurrence plots are somehow a visualization of the auto-correlation in time of the sequence values.

Matrix (5.1) can be plotted in 2 dimensions (Fig. 1) by putting a dot where $u_{hk} = 1$ and white spot when $u_{hk} = 0$.

5.1. Example. For instance for the classical Cauchy problems of the elastic vibrations

$$\begin{cases} \dfrac{d^2x}{dt^2} + x = 0 \\ x|_{t=0} = 3 \quad , \quad \left.\dfrac{dx}{dt}\right|_{t=0} = 1 \end{cases}$$

and vibrations with damping

$$\begin{cases} \dfrac{d^2x}{dt^2} + 0.08\dfrac{dx}{dt} + x = 0 \\ x|_{t=0} = 3 \quad , \quad \left.\dfrac{dx}{dt}\right|_{t=0} = 1 \end{cases}$$

with

$$n = 100 \quad , \quad \delta = 0.2$$

we have the orbits and recurrence plots of Figure 2

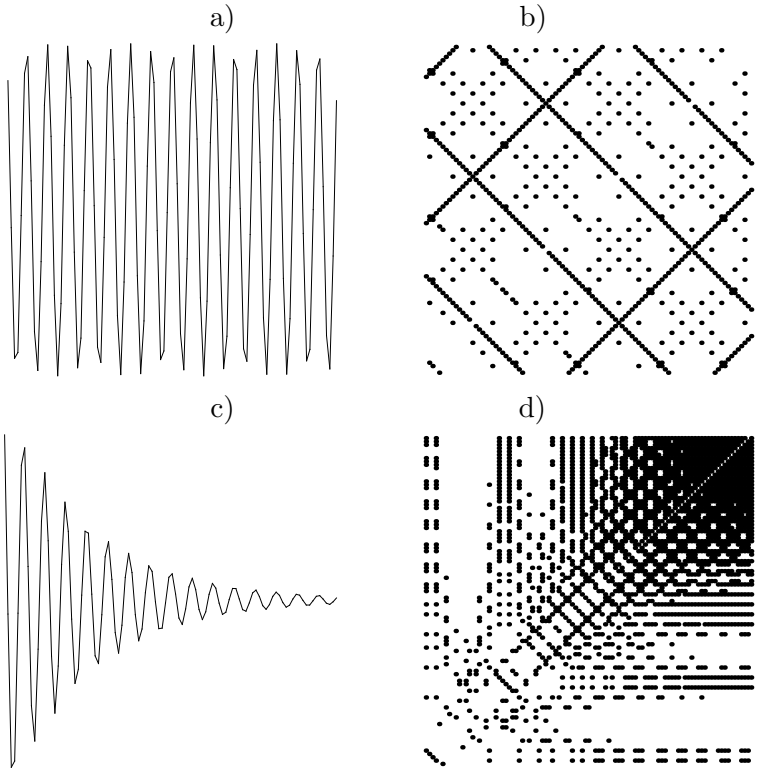

FIGURE 2. *Orbits (left column) and recurrence plots (right column) with* $\delta = 0.02$ *for the free vibration system* $a), b)$ *and vibration with damping* $c), d)$.

It can be seen that the free-vibration recurrence plot shows the periodic behavior while the recurrence plot for the vibration with damping shows the existence of the so-called trend [37, 40].

More in general some special textures have been identified and taken for comparing the behavior of other dynamical systems (Fig. 3).

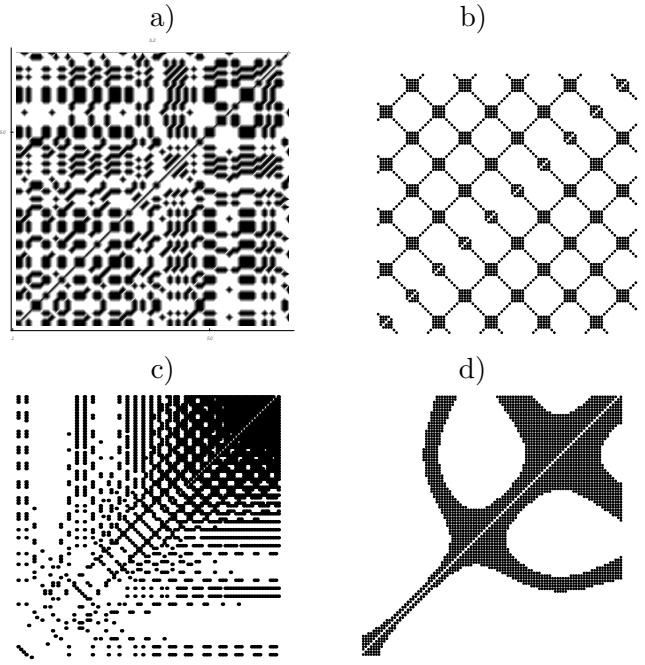

FIGURE 3. *Typical textures in recurrence plots: a) homogenous, b) periodic, c) disrupted and d) drift.*

6. Prime number distribution

In this section the binary map and binary plots will be defined for the distribution of primes. On the binary plots the main parameters of complexity and multifractality will be computed to show the fractal nature of primes.

6.1. Preliminary remarks on Primes. Given two positive integers a and n we say that a divides n (n is divisible by a) if and only if there exists a positive integer b such that

$$n = ab \ .$$

In other words, given two positive integers $a,\ n \in \mathbb{N}$

$$a|n \Leftrightarrow \exists\, b \in \mathbb{N} : n = ab \ ,$$

in this case it is said that a is a divisor of n (or a divides n).

Let

$$Div(n) \stackrel{\text{def}}{=} \{m \in \mathbb{N} : m|n\}$$

be the set of positive divisors of n and $|Div(n)|$ the cardinality of the set, we say that a positive integer p is a prime if $|Div(p)| = 2$, so that an integer p is a prime if its only positive divisors are 1 and p. The set of all primes is

$$\mathbb{P} \stackrel{\text{def}}{=} \{p \in \mathbb{N} : |Div(p)| = 2\}$$

and a fundamental elementary theorem of Arithmetic states that every integer larger than 1 can be expressed as a product of primes, so that any positive integer has a unique prime factorization (up to a suitable ordering), and primes play the role of atoms for the positive integers.

It was known, already by Euclid's time, that the number of primes is infinite, i.e. $|\mathbb{P}| = \infty$, however it is still unknown how they are distributed within $\mathbb{N}$.

If we define the counting function, $\pi(x) : \mathbb{R} \Rightarrow \mathbb{N}$, as

$$\pi(x) = |\mathbb{P}_x| \quad , \quad \mathbb{P}_x \stackrel{\text{def}}{=} \{p \in \mathbb{P} : p \leq x\} \quad , \quad \mathbb{P}_x \subseteq \mathbb{P}$$

it has been conjectured by Gauss that $\pi(x)$ asymptotically tends to $x/\log x$, i.e.

$$\pi(x) \sim x/\log x \tag{6.1}$$

so that the prime number theorem

$$\lim_{x\to\infty} \frac{\pi(x)}{x/\log x} = 1$$

holds.

6.2. Global fractal estimate by the correlation matrix. Let us define the binary map on prime distribution as follow:

$$u_{hk} \stackrel{\text{def}}{=} u(h,\ k) = \begin{cases} 1 & \text{if} \quad \{h \in \mathbb{P}\} \wedge \{k \in \mathbb{P}\} \quad = \text{ TRUE} \\ 0 & \text{if} \quad \{h \in \mathbb{P}\} \wedge \{k \in \mathbb{P}\} \quad = \text{ FALSE} \end{cases} , \ (h, k \in \mathbb{N}) \ , \tag{6.2}$$

and let $p(x),\ x \in \mathbb{R}$ be the probability to find a prime at the natural number x. For large values of x it is

$$p(x) \cong \frac{\pi(x)}{x} \stackrel{(6.1)}{=} \frac{1}{\log x}$$

So that, according to Gauss conjecture, the possibility to find some primes is vanishing for higher values of x, so that for higher values of n we find much more primes, but this probability reduces to zero

Let us compute the number of 1 in the minor $2^m - n \times 2^m - n$ of the indicator matrix $2^m \times 2^m$.

If we count the number of 1 in the $n \times n$ indicator matrix as a function of n we have the plot of frequencies (Fig. 4) which is similar to a Cantor function, thus suggesting us that the primes are distributed (within the indicator matrix) as fractals.

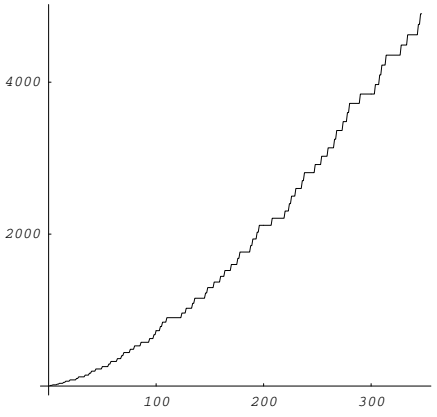

FIGURE 4. *Frequencies of 1 in the indicator matrix* ($n \leq 350$).

The binary plot of prime distribution (see Fig. 1) looks like the Cantor dust. So that we can assume that the binary plot of the distribution of primes is similar to a Cantor dust. We can compute the fractal dimension of the binary plot of primes and show that

Theorem 6.1. *The fractal dimension of the binary plot for the distribution of primes is* $\sqrt{2}$.

Proof: By using the indicator matrix it is possible to give a simple formula which enables us to estimate the fractal dimension as the average of the number $p(n)$ of 1 in the randomly taken $n \times n$ minors of the $N \times N$ correlation matrix u_{hk}

$$D = \frac{1}{N} \sum_{n=2}^{N} \frac{\log p(n)}{\log n} \ . \tag{6.3}$$

By a direct computation we obtain that the fractal dimension of the primes distribution in the binary matrix (binary plots of Fig. 1) is roughly $\sqrt{2}$. □

From this theorem there folows and interesting functional equation for the primes counting function $\pi(x)$:

Theorem 6.2. *The primes counting function fulfills the equation*

$$\pi(x) + \pi(x)^{\frac{1}{\sqrt{2}}} = 1 \qquad , \quad (x \to \infty, \ x \in \mathbb{N}) \tag{6.4}$$

Proof: Taking into account the definition of the fractal dimension (6.3), if we count the number of zeroes and the number of ones in the indicator matrix we have as a ratio

$$\frac{\log[1 - p(n)]}{\log p(n)} \cong \frac{1}{\sqrt{2}}$$

which is equivalent to

$$\frac{\log[1 - \pi(x)]}{\log \pi(x)} = \frac{1}{\sqrt{2}} \quad .$$

from where (6.4) follows.

□

6.3. Complexity. The existence of repeating motifs, periodicity and patchiness can be considered as a simple behavior of sequence. While non-repetitiveness or singularity might be taken as a characteristic feature of complexity. In order to have a measure of complexity, for an n-lenght sequence, it has been proposed [10] the following

$$K = \log \Omega^{1/n}$$

with

$$\Omega = \frac{n!}{\pi(n)!}$$

Thus we can show the following

Theorem 6.3. *The complexity for prime distribution is* $K = \dfrac{1}{3}$

Proof: By using a sliding n-window [10] over the full sequence $v(k)$ one can visualize the distribution of complexity on partial fragment of the sequence. It is interesting to notice that although there is an increasing complexity, for the first numbers of the sequence $n \leq 400$ there is a constant trend to complexity $K \cong 0.3333...$ which is given by the least square fit (Fig. 5). □

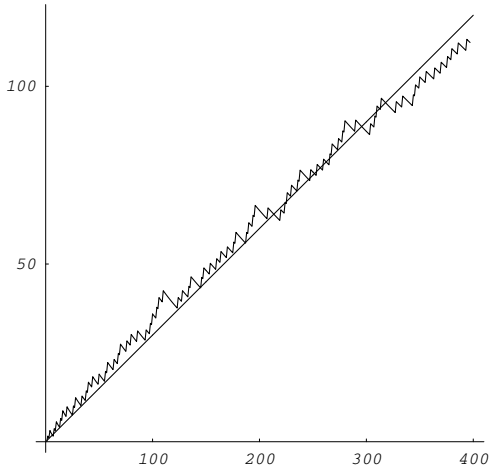

FIGURE 5. *Complexity for the first 400 primes and its corresponding least square fit.*

7. Statistics on Ulam Spiral

In this section we consider a 2D distribution of primes according to the so-called Ulam spiral [87]. In order to find some patterns in prime distribution integer can be arragend along a rectangular spiral. This is equivalent to map the 1D sequence of integers into a 2D sequence as follows (see Fig 6):

$$\begin{array}{cc} 1 & \{0,0\} \\ 2 & \{1,0\} \\ 3 & \{1,1\} \\ 4 & \{0,1\} \\ 5 & \{-1,1\} \\ 6 & \{-1,0\} \\ 7 & \{-1,-1\} \\ 8 & \{0,-1\} \\ 9 & \{1,-1\} \\ 10 & \{2,-1\} \\ 11 & \{2,0\} \\ \vdots & \vdots \end{array}$$

If we select from the above sequence only the prime numbers we obtain the so-called Ulam spiral (Fig. 6) so that primes seem to be distributed along some straight (mostly diagonal) lines in the plane.

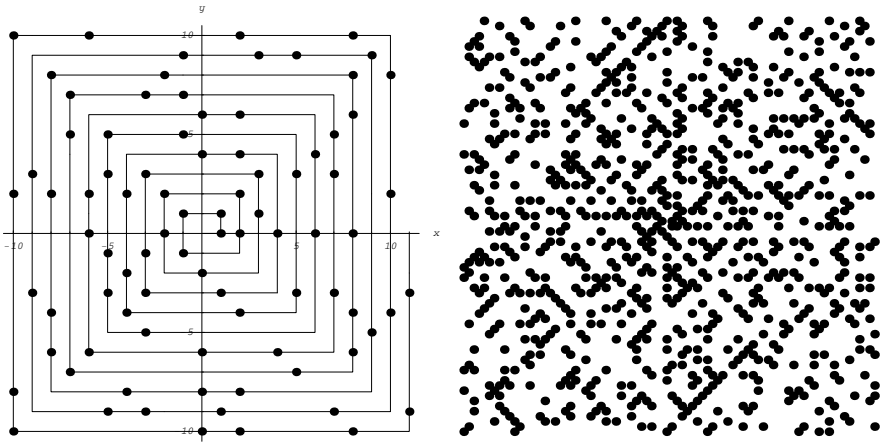

FIGURE 6. *Primes distribution on a rectangular spiral: first 87 primes (left) and 853 (right).*

It should be noticed that along the Ulam spiral, there is a ono-to-one map λ between $\mathbb{N}$ and the points of the spiral (with integer coordinates) in $\Re^2$

$$\lambda : \mathbb{N} \mapsto \gamma \subset \Re \times \Re$$

so that

$$\lambda(n) = (a,\ b) \qquad , \quad (n \in \mathbb{N}; (a,\ b) \in \gamma \subset \Re \times \Re;\ a \in \mathbb{Z},\ b \in \mathbb{Z})$$

and

$$\lambda^{-1}(a,\ b) = n\ .$$

This bijective map can be considered also between $\mathbb{N}$ and the complex space $\mathbb{C}$ so that each natural number corresponds to a complex number (with integer coefficients)

$$\lambda(n) = z \stackrel{\text{def}}{=} a + ib \qquad , \quad (n \in \mathbb{N};\ a,\ b \in \mathbb{Z}; z \in \mathbb{C})\ . \tag{7.1}$$

It has been observed that since primes are odd numbers (except 2) they are distributed (in the plane) along some alternate diagonal lines. However, it is still unclear how they are distributed along these lines. Some information, about the distribution of primes, can be obtained by the following theorem

Theorem 7.1. *In the Ulam spiral, if n is a prime number the remainder on division of $|\lambda(n)^2|$ by 2 is 0, i.e.*

$$\text{Mod}\ [|\lambda(n)^2|, 2] = 0 \qquad , n > 2 \tag{7.2}$$

so that, according to (7.1)

$$a^2 + b^2 = 2k \qquad , \qquad (k \in \mathbb{N}; a, b \in \mathbb{Z}) \tag{7.3}$$

Proof: It can be easily obtain by recursion. Let n_s define the set of primes belonging to the s-th spiral so that

$$\lambda(n_s) \stackrel{\text{def}}{=} \{\lambda(n)\} \qquad , \qquad (n \in s - th \text{ spiral})$$

It can be easily show that

$$n \in n_s \Longleftrightarrow |\Re[\lambda(n)]| \vee |\Im[\lambda(n)]| = s\ .$$

In particular it is

$$n_1 = \{2,\ 3,\ 5,\ 7\}$$

and

$$\lambda(n_1) \stackrel{\text{def}}{=} \{\lambda(2),\ \lambda(3),\ \lambda(5),\ \lambda(7)\}\ .$$

Let us show that when $s = 1$, the primes on the first circle of the Ulam spiral, but 2, fulfill (7.3). By a direct computation it can be seen that

$$\lambda(3) = \{1, 1\},\ \lambda(5) = \{-1, 1\},\ \lambda(7) = \{-1, -1\}\}$$

so that Eq. (7.2) is true , being

$$a^2 + b^2 = 2 \Rightarrow a = \pm 1,\ b = \pm 1\ .$$

Assume that Eq. (7.2) holds for the s-th spiral and show that it is true also for the $(s+1)$-th spiral. It is enough to show it for $n + 1 \in n_{s+1}$ and $n \in n_s$:

$$\lambda(n+1) = a \pm 1 + i(b \pm 1) = a + ib \pm (1 + i)$$

that is

$$\lambda(n+1) = \lambda(n) + \lambda(1)$$

□

As a consequence, the following theorem holds:

Theorem 7.2. *In the Ulam spiral, if n is an even number the remainder on division of $|\lambda(n)^2|$ by 2 is 1, i.e.*

$$\text{Mod}\ [|\lambda(n)^2|, 2] = 1 \qquad , n > 2 \tag{7.4}$$

so that, according to (7.1)

$$a^2 + b^2 = 2k + 1 \qquad , \qquad (k \in \mathbb{N}; a, b \in \mathbb{Z}) \tag{7.5}$$

Proof: Let n be an odd number for which Eq. (7.3) holds true and show that $n + 1$ (which is even number) fulfill Eq. (7.5). From $\lambda(n) = (a,\ b)$ it is $\lambda(n+1) = (a \pm 1,\ b)$ or $\lambda(n+1) = (a,\ b \pm 1)$. It is enough to show that these values fulfill (7.5). For instance, it is

$$(a \pm 1)^2 + b^2 = 2k + 1$$

i.e.

$$a^2 + b^2 = 2k \mp 2a$$

which is true according to (7.3).

□

Equation (7.3) is a necessary condition for a primality test, let us check on a few examples:

- $[k = 1]$ Eq. (7.3) becomes

  $$a^2 + b^2 = 2$$

  and the integer solutions are

  $$a = \pm 1\ ,\ b = \pm 1$$

  The corresponding points are

  $$z_1 = (-1, -1)\ ,\ z_2 = (-1, 1)\ ,\ z_3 = (1, -1)\ ,\ z_4 = (1, 1)\ .$$

By the inverse map we get the numbers of the first spiral ($s = 1$) among which there are prime numbers:

$$\lambda^{-1}(z_1) = 7\ , \lambda^{-1}(z_2) = 5\ , \lambda^{-1}(z_3) = 9\ , \lambda^{-1}(z_4) = 3\ ,$$

In this case the inverse map gives all primes of the first spiral together with 9 (which is not prime).

- $[k = 2]$ Eq. (7.3) becomes

$$a^2 + b^2 = 4$$

and the integer solutions are

$$a = \pm 2\ ,\ b = 0$$

and

$$a = 0\ ,\ b = \pm 2\ .$$

The corresponding points are

$$z_1 = (-2, 0)\ ,\ z_2 = (2, 0)\ ,\ z_3 = (0, -2)\ ,\ z_4 = (0, 2)$$

by the inverse map we get the prime numbers

$$\lambda^{-1}(z_1) = 19\ , \lambda^{-1}(z_2) = 11\ , \lambda^{-1}(z_3) = 23$$

and the integer (not prime) $\lambda^{-1}(z_4) = 15$.

- $[k = 3]$ Eq. (7.3) becomes

$$a^2 + b^2 = 6$$

there not exist integer solutions for a and b.

- $[k = 4]$ Eq. (7.3) becomes

$$a^2 + b^2 = 8$$

and the integer solutions are

$$a = \pm 2\ ,\ b = \pm 2\ .$$

The corresponding point are

$$z_1 = (-2, -2)\ ,\ z_2 = (-2, 2)\ ,\ z_3 = (2, -2)\ ,\ z_4 = (2, 2)$$

by the inverse map we get the prime numbers

$$\lambda^{-1}(z_2) = 7\ , \lambda^{-1}(z_4) = 13$$

and the integer (not prime) $\lambda^{-1}(z_1) = 21,\ \lambda^{-1}(z_3) = 25$.

A general solution of (7.3) can be found in some special cases:

- $[a = b\ , k = c^2]$ The solution is

$$a = b = \pm h \quad , \quad c \in \mathbb{N}\ .$$

- $[2k = c^2]$ We have a Pythagorean triple so that the solution is

$$a = 2k(m^2 - n^2)\ ,\ b = 4kmn\ ,\ c = 2k(m^2 + n^2) \quad , \quad (m > n\ , k \in \mathbb{N})$$

where m and n are coprime and exactly one of them is even.

Equation (7.3) can be used to test that to a given point in the plane doesn't correspond a prime and to find primes. For instance the number

$$z_0 = (10,\ -13)$$

doesn't correspond to a prime, in fact the sum

$$10^2 + 13^2 = 269$$

is not an even number. In fact, by the inverse map we have

$$\lambda^{-1}(z_0) = 726\ ,$$

which is not a prime.

Analogously, given an even number e.g. 648 let us find a couple of integers so that they sum up to 648

$$a^2 + b^2 = 648\ .$$

We have as solution

$$z_1 = (-18,\ 18),\ z_2 = (-18,\ -18),\ z_3 = (18,\ -18),\ z_4 = (18,\ 18)\ .$$

By the inverse map we get

$$\lambda^{-1}(z_1) = 1297$$

which is a prime, while the others are not

$$\lambda^{-1}(z_2) = 1333,\ \lambda^{-1}(z_3) = 1369,\ \lambda^{-1}(z_4) = 1261,$$

It should be noticed that since primes are distributed along some spirals, the absolute value of z grows within some fixed range (Fig. 7).

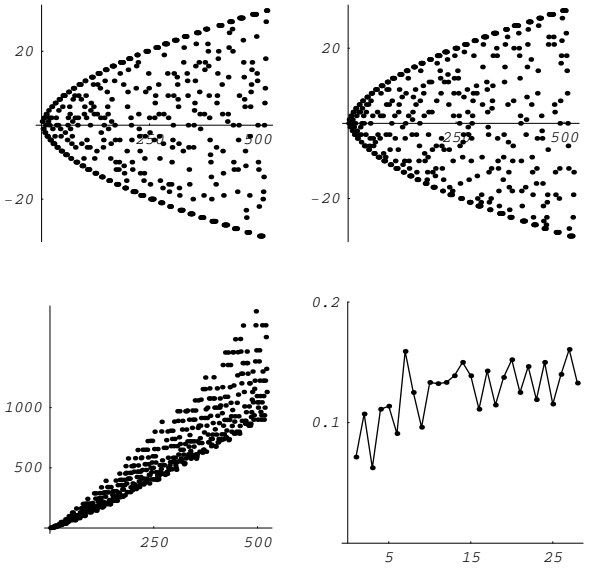

FIGURE 7. *On top, real (left) and imaginary (right) coefficients of $\lambda(p)$ along the first 30 spirals; on bottom, $|\lambda(p)^2|$ (left) and on the right the ratio of the number of primes within the max-min absolute value and those on the parabola $x = |\sqrt{y}|$ along the first 30 spirals (corresponding to the first 522 primes)*

It can be also easily shown by a direct computation that if $z = \lambda(n)$ is a prime belonging to the n_s spiral then (Fig. 8)

$$2k\,|z\overline{z}|^{1/4} = n_s \qquad , \qquad (k \in \mathbb{N}) \ .$$

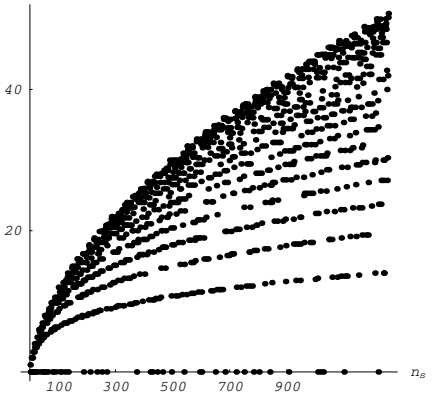

FIGURE 8. *Distribution of primes along the spirals*

7.1. Cluster analysis of the wavelet coefficients of the complex primes distribution. In this section the clusters of wavelet coefficients for the complex representation of primes (along the Ulam spiral) will be analized.

The cluster algorithm will be applied to the complex representation sequence of primes $\lambda(p)$, $p \in \Pi$, which is in the form

$$1, 1+i, -1+i, -1-i, 2, 2+2\,i, -2+2\,i, -2, -2\,i, 3+i, 3+3\,i, -3+3\,i, -3-i, \ldots$$

and to the random walk on $\lambda(p)$ (Fig. 9):

$$\sum_i \lambda(p_i) \qquad , \qquad (pi_i \in \mathbb{P}) \ .$$

that is the sequence

$$1, 2+i, 1+2\,i, i, 2+i, 4+3\,i, 2+5\,i, 5\,i, 3\,i, 3+4\,i, 6+7\,i, 3+10\,i, 9\,i, -3+6\,i, \ldots$$

For each complex representation of primes, along the spiral, there are 2 sets of wavelet coefficients which correspond to the real and complex coefficient of the complex values $\lambda(n)$. If we consider the first 4263 primes of the spiral and compute the 4-parameters discrete Haar wavelet transform, with the above clustering algorithm we have the patterns of Fig. 10. In Fig. 11 the 8-parameters transform is given.

It should be noticed that at the highest frequencies (Figs. 10,c,d , 11,f,g,h) the wavelet coefficients are distributed along the axes. In other words the difference between close complex numbers of the spiral (corresponding to prime numbers) is either real or pure imaginary. At the lower frequencies the wavelet coefficients have discrete values bounded by

$$\left|\beta_0^0 \pm \beta^{*0}_{\ 0}\right| \leq \pi \log_2 n \ .$$

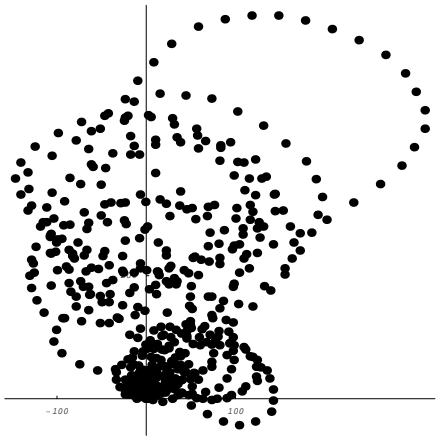

FIGURE 9. *Random walk on the first 522 primes of the Ulam spiral*

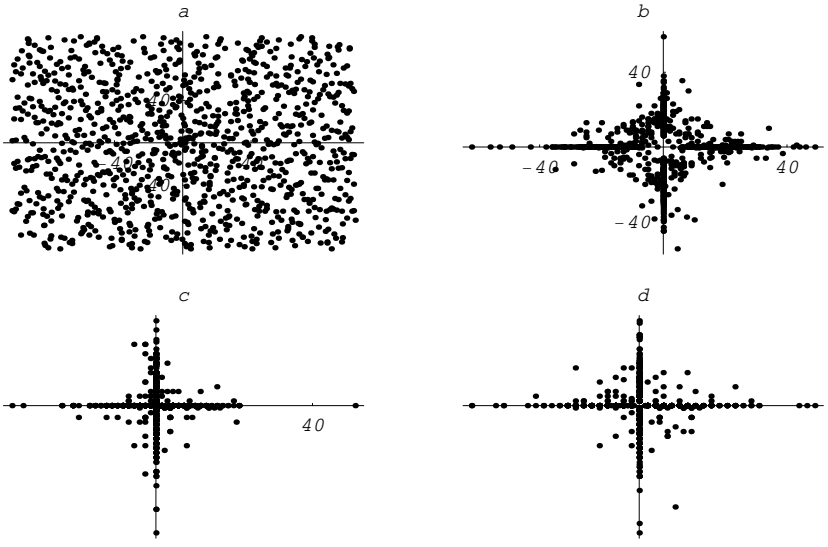

FIGURE 10. *Clusters of wavelet coefficients for the first 4263 primes of the Ulam spiral: a)* (α, α^*) *; b)* $\left(\beta_0^0, \beta^{*0}{}_0\right)$ *; c)* $\left(\beta_0^1, \beta^{*1}{}_0\right)$*; d)* $\left(\beta_1^1, \beta^{*1}{}_1\right)$.

8. Fractal analysis of DNA sequences

In this section the fractal analysis of DNA will be given. A DNA sequence is a double strand helix, where the nucleotide on each strand is paired with the other according to some complementary chemical rules. When one of the two strands is linearly stretched we have a sequence of symbols as

$$\{A,\ ,C,\ A,\ T,\ G, A\ ,T, ...\}$$

Since there exist 3-length subsequences (codons) having some chemical -biological meaning (related with proteins and then functionality) one of the main problem is to understand if there exist an underlying rule for the distribution of nucleotides.

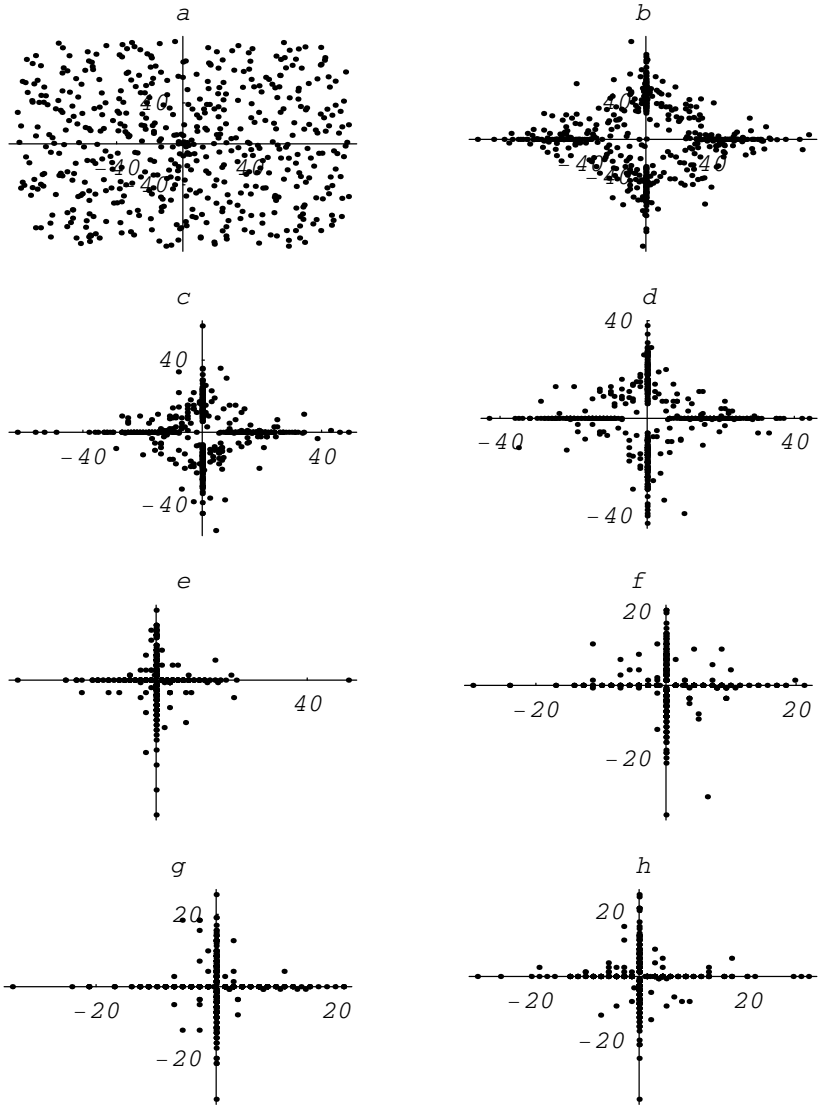

FIGURE 11. *Clusters of wavelet coefficients for the first 4263 primes of the Ulam spiral: a)* (α, α^*) *; b)* $\left(\beta_0^0, \beta^{*0}_{\ 0}\right)$ *; c)* $\left(\beta_0^1, \beta^{*1}_{\ 0}\right)$*; d)* $\left(\beta_1^1, \beta^{*1}_{\ 1}\right)$*, e)* $\left(\beta_0^2, \beta^{*2}_{\ 0}\right)$*, f)* $\left(\beta_1^2, \beta^{*2}_{\ 1}\right)$*, g)* $\left(\beta_2^2, \beta^{*2}_{\ 2}\right)$*, h)* $\left(\beta_3^2, \beta^{*2}_{\ 3}\right)$*,* .

The simplest analysis of DNA is based on the 4-alphabet of nucleotides

$$A_1 \stackrel{\text{def}}{=} \{A\ , C\ , G\ , T\} \tag{8.1}$$

being the nucleotides (nucleic acids): adenine (A), cytosine (C), guanine (G), thymine (T), (see e.g. [24, 23]) and their corresponding complex numerical representation [21, 22, 24, 23].

Let S_N be a N-length ordered linear sequence of nucleotides (8.1) the indicator function is defined as [21, 22, 24, 23])

$$u : (S_N) \times (S_N) \to \{0\ , 1\} \tag{8.2}$$

such that

$$u(x_h,\ x_k) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if} \quad x_h = x_k \\ \\ 0 & \text{if} \quad x_h \neq x_k \end{cases} \quad , \ (x_h \in (S_N)\ ,\ x_k \in (S_N))\ , \tag{8.3}$$

The indicator matrix ([21, 22, 27, 24, 23]), defined as $u_{hk} \stackrel{\text{def}}{=} u(x_h,\ x_k)$ can be used to obtain the two dimensional dot-plot (see e.g. Fig. 13).

Then the frequency of each symbol at the position k in the sequence can be computed as the marginal frequency

$$\nu_h(k) \stackrel{\text{def}}{=} \frac{1}{k} \sum_{j=1}^{k} u_{hj}^{*}\ , \tag{8.4}$$

so that

$$\sum_{h=1}^{M_\ell} \nu_h(k) = 1$$

As approximated value of probability to find the nucleotide $A,\ C,\ G,\ T$ at the position k in the sequence $D_\ell(S_N)$, can be taken the following

$$p_h(k) \cong \nu_h(k)\ . \tag{8.5}$$

It can be noticed that DNA sequences of a living organism resemble (Fig. 13) random sequences, with some short range influence, built on the same alphabet. This has been taken as an axiom of nucleotides distribution, so that DNA sequences are often considered as Markov chain [83]. However, there are some hidden rules in combining the nucleotides and these rules lead, during the evolution, to a steady distribution. In fact, the more primitive is the sequence the more randomly distributed are the nucleotides. It seems that, as a consequence of the evolution, nucleotides move from a disordered aggregation toward a more organized structure, shown by the growing islands in the dot plot. The biological evolution is such that the challenge for the self organization might follow from random permutations of a primitive disordered sequence so that the organization, i.e. the complexity, is only the result of many arbitrary permutations of randomness. During the challenge for complexity, DNA sequence becomes "less random" and it looses some kind of energy.

By a random permutation of the nucleotides form each sequence we can obtain a "new" sequence which has the same nucleotides but located in different places. It can be easily seen by a direct computation that

Theorem 8.1. *The fractal dimension computed on the binary image is invariant for random permutation of nucleotides.*

8.1. Spiral plot. Like for the distribution of primes, we can also visualize the distribution of nucleotides by mapping each sequence into the integer points of a spiral (Fig. 12) in $\mathbb{R}^2$, as the Ulam spiral [87]:

$$(S_N) \mapsto (x,\ y)\ ,\ (x,\ y \in \mathbb{Z})$$

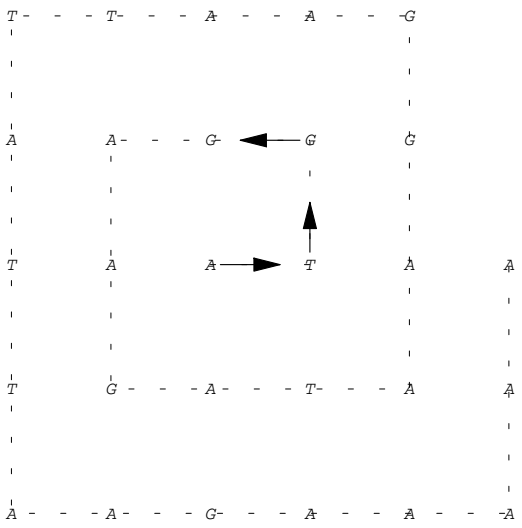

FIGURE 12. *Distribution of nucleotides on a rectangular spiral.*

8.2. Test Sequences. In the following, in order to single out the main features of biological sequences, we can compare the DNA sequence with some test sequences. Let $A_\ell = \{A,\ C,\ G, T\}$ be the alphabet, we have the following recursive sequences built on the alphabet

1.: Pseudo-random N-length sequence of nucleotides is the sequence $\{R_i\}_{i=1,\ldots,N}^{\ell}$ where r_i is a symbol randomly chosen in the alphabet A_ℓ, like e.g. $(\ell = 1)$:

$$\{A,\ C,\ A,\ G,\ T,\ A,\ T,\ G,\ G,\ A,\ T,\ T,\ A,\ C,\ C,\ ,G\ \ldots\}$$

2.: Pseudo-periodic N-sequence of nucleotides with period π is the direct sum of a given π-length pseudo-random sequence , such that $N = k\pi$, $(k \in \mathbb{N})$ and $R_i = R_{i+\pi}$, e.g.

$$\{A,\ C,\ A,\ G,\ A,\ C,\ A,\ G,\ A,\ C,\ A,\ G,\ A,\ C,\ A,\ G,\ldots\}\ ,\ (\pi = 4)$$

When $\pi = 1$ we have a pseudo-random sequence.

3.: Recursive sequence with highest dispersion. At each step, taken 4 nucleotides between each couple of nucletides are added the remaining, so that at the 2nd step we have:

$$\begin{array}{cccccccccc} \boldsymbol{A} & & & \boldsymbol{C} & & & \boldsymbol{G} & & & \boldsymbol{T} \\ \boldsymbol{A} & G & T & \boldsymbol{C} & A & T & \boldsymbol{G} & A & C & \boldsymbol{T} \end{array}$$

4.: Cantor sequence, is a recursive sequences as given in the previous item, where the middle pair is removed, that is

$$\begin{array}{cccccccccccccc} \boldsymbol{A} & & & \boldsymbol{C} & & & \boldsymbol{G} & & & \boldsymbol{T} & & & & \\ \boldsymbol{A} & G & T & \boldsymbol{C} & & & \boldsymbol{G} & A & C & \boldsymbol{T} & & & & \\ \boldsymbol{A} & T & C & \boldsymbol{G} & \boldsymbol{T} & A & G & \boldsymbol{C} & \boldsymbol{G} & C & T & \boldsymbol{A} & \boldsymbol{C} & G & A & \boldsymbol{T} \end{array}$$

5.: Cantor sequence, filled with random sequence. It is a Cantor sequences as given in the previous item, where the middle part is filled with a random sequence, that is

$$\begin{array}{cccccccccccc} \boldsymbol{A} & & & \boldsymbol{C} & & & & \boldsymbol{G} & & & \boldsymbol{T} \\ \boldsymbol{A} & G & T & \boldsymbol{C} & C & \ldots & A & \boldsymbol{G} & A & C & \boldsymbol{T} \end{array}$$

Alternatively the filled part can be sequences of repeats.

6.: Fibonacci sequence. If we define the sum of words on the ℓ-alphabet $\{a_0,\ a_1,\ \ldots,\ a_{\ell-1}\}$ as

$$a_i + a_j = a_{[i+j,|A_\ell|]},\ i,j = 0,1,\ldots,|A_\ell| - 1$$

being $[i+j,|A_\ell|]$ the remainder of the division by $|A_\ell|$, we can define as Fibonacci sequence the following:

$$x_1 = a_i\ ,\ x_2 = a_j\ ,\ x_{n+2} = x_{n+1} + x_n\ ,\ (n \geq 1\ ;\ 0 \leq i,j \leq |A_\ell| - 1) \qquad (8.6)$$

On the ℓ-alphabet of words we can choose the initial pair of values $x_1,\ x_2$, independently of their order, among 2-combinations with repetitions on a set of $|A_\ell|$ elements, for a total of

$$\left(\begin{array}{c} |A_\ell| + 1 \\ 2 \end{array} \right)$$

pairs.

Example 1.

On the 4-alphabet of nucleotides

$$a_0 = A\ ,\ a_1 = C\ ,\ a_2 = G\ ,\ a_3 = T$$

we can choose the initial pair among 10 multi-combinations. In this way we obtain some repeat sequences, as follows:

$$\begin{array}{c} x_1 = a_0 = A\ , x_2 = a_1 = C \\ \boldsymbol{0},\ \boldsymbol{1}, 1, 2, 3, 1, \boldsymbol{0},\ \boldsymbol{1}, 1, 2, 3, 1, \boldsymbol{0},\ \boldsymbol{1}\ \ldots \\ \boldsymbol{A},\ \boldsymbol{C},\ C,\ G,\ T,\ C,\ \boldsymbol{A},\ \boldsymbol{C}\ C,\ G,\ T,\ C,\ \ldots \end{array}$$

or

$$\begin{array}{c} x_1 = a_1 = C\ , x_2 = a_3 = T \\ \boldsymbol{1}, \boldsymbol{3}, 0, 3, 3, 2, \boldsymbol{1}, \boldsymbol{3}, 0, 3, 3, 2, \boldsymbol{1}, \boldsymbol{3}, \ldots \\ \boldsymbol{C},\ \boldsymbol{T},\ A,\ T,\ T,\ G,\ \boldsymbol{C},\ \boldsymbol{T},\ A,\ T,\ T,\ G,\ \ldots \end{array}$$

Example 2.

On the 20-alphabet of amino-acids

$$a_0 = M\ ,\ a_1 = E\ ,\ \ldots\ ,\ a_{19} = W$$

we can choose the initial pair among 210 multi-combinations.

For instance with

$$x_1 = a_0 = M\ ,\ x_2 = a_{19} = W$$

the first 30 terms of the Fibonacci sequence are

$$\begin{array}{c} 0, 19, 19, 18, 17, 15, 12, 7, 19, 6, 5, 11, 16, 7, 3, \\ 10, 13, 3, 16, 19, 15, 14, 9, 3, 12, 15, 7, 2, 9, 11, \ldots \end{array}$$

that is

$$M,\ W,\ W,\ Y,\ K,\ C,\ P,\ H,\ W,\ N,\ T,\ S,\ A,\ H,\ D,$$
$$L,\ F,\ D,\ A,\ W,\ C,\ I,\ G,\ D,\ P,\ C,\ H,\ Q,\ G,\ S,\ldots$$

Another sequence is as follows:

$$x_1 = a_0 = M\ ,\ x_2 = a_{18} = Y$$

the first 30 terms of the Fibonacci sequence are

$$0, 18, 18, 16, 14, 10, 4, 14, 18, 12, 10, 2, 12, 14, 6,$$
$$0, 6, 6, 12, 18, 10, 8, 18, 6, 4, 10, 14, 4, 18, 2, \ldots$$

that is

$$M,\ Y,\ Y,\ A,\ I,\ L,\ R,\ I,\ Y,\ P,\ L,\ Q,\ P,\ I,\ N,$$
$$M,\ N,\ N,\ P,\ Y,\ L,\ V,\ Y,\ N,\ R,\ L,\ I,\ R,\ Y,\ Q,\ldots$$

As last example we consider the sequence :

$$x_1 = a_6 = N\ ,\ x_2 = a_{12} = P$$

the first 30 terms of the Fibonacci sequence are

$$6, 12, 18, 10, 8, 18, 6, 4, 10, 14, 4, 18, 2, 0, 2,$$
$$2, 4, 6, 10, 16, 6, 2, 8, 10, 18, 8, 6, 14, 0, 14\ldots$$

that is

$$N,\ P,\ Y,\ L,\ V,\ Y,\ N,\ R,\ L,\ I,\ R,\ Y,\ Q,\ M,\ Q,$$
$$Q,\ R,\ N,\ L,\ A,\ N,\ Q,\ V,\ L,\ Y,\ V,\ N,\ I,\ M,\ I\ldots$$

8.3. Fractal analysis on binay plots. In this section we consider the dot-plots of some DNA sequnces.

If we plot the indicator matrix of some bacteria and compare it with a pseudo-random and periodic sequence, we can see that (Fig. 13)

(1) the main diagonal is a symmetry axis for the plot
(2) there are some motifs which are repeated at different scales like in a fractal;
(3) periodicity is detected by parallel lines to the main diagonal (Fig. 13, a2)
(4) empty spaces are more distributed than filled spaces, in the sense that the matrix u_{hk} is a sparse matrix (having more 0's than 1's);
(5) it seems that there are some square-like islands where black spots are more concentrated; these islands show the persistence of a nucleotide (Fig. 13,a2 and b1)
(6) the dot plot of archaea is very similar to the dot plot of a random sequence (Fig. 13, a1 and h3)

It can be noticed that DNA sequences of a living organism resemble (Fig. 13) random sequences, with some short range influence, built on the same alphabet. This has been taken as an axiom of nucleotides distribution, so that DNA sequences are often considered as Markov chain [83]. However, there are some hidden rules in combining the nucleotides and these rules lead, during the evolution, to a steady distribution. In fact, the more primitive is the sequence the more randomly distributed are the nucleotides. It seems that, as a consequence of the evolution, nucleotides move from a disordered aggregation toward a more

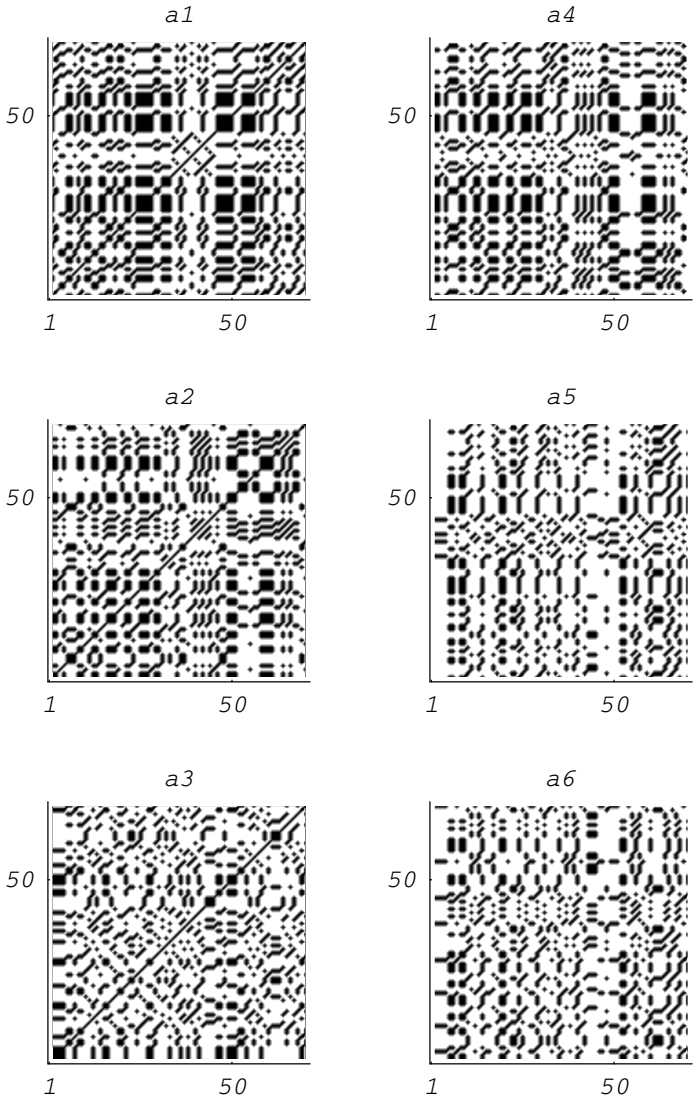

FIGURE 13. *Indicator matrix for:* $a1$) *pseudo-random* 70*-length sequence;* $a2$) *pseudo-periodic* 70*-length sequence with period* $\pi = 35$*;* $b1$) 70*-length Dna sequence of Mycoplasma KS1 bacter;* $h3$) 70*-length Dna sequence of Acidilobus Archaea.*

organized structure, shown by the growing islands in the dot plot. The biological evolution is such that the challenge for the self organization might follow from random permutations of a primitive disordered sequence so that the organization, i.e. the complexity, is only the result of many arbitrary permutations of randomness. During the challenge for complexity, DNA sequence becomes "less random" and it looses some kind of energy.

From the graphical representation of the indicator matrix for bacteria and amino acids we can see a more sparse matrix, but with some typical plots (Fig. 13)

These plots can immediately visualize the different distribution of nucleotide, for instance in b1, there is a major distribution of A, T while on the contrary in $h3$, the higher frequency belongs to C, G.

There is also another feature of these plots: they can be considered as a first attempt to give some 2D representation of the DNA sequence, thus opening new perspectives in a dynamical model representation of DNA. In the following section will be considered some 2D phase plots of DNA thus enabling us to define the recurrence plots as usually done with dynamical systems.

9. Random walks

In order to improve our analysis of correlation in DNA, we have to introduce a digitalization of the symbolic alphabet and apply the usual methods of signal analysis. In this section, the complex roots representation is proposed .

9.1. Complex Root Representation. The complex (digital) representation of a DNA sequence of words is the map of the symbolic sequence of words into a set of complex numbers and it is defined as

$$D_\ell(S_N) \xrightarrow{\rho} \mathbb{C}$$

such that for each $x_h \in D_\ell(S_N)$ it is $\rho(x_h) \in \mathbb{C}$. The complex root representation of $D_\ell(S_N)$ is the sequence of complex numbers $\mathbf{Y}_{M_\ell} = \{y_h\}_{h=1,\dots,M_\ell}$ defined as

$$y_h = \rho(x_h) \stackrel{\text{def}}{=} e^{2\pi i(j-1)/|A_\ell|} \quad , \quad (j = 1, \dots, |A_\ell|,\ h = 1, \dots, M_\ell) \tag{9.1}$$

with $i = \sqrt{-1}$ the imaginary unit and $M_\ell = |A_\ell|$. There follows that, independently on the alphabet, it is

$$|y_h| = |e^{2\pi i(j-1)/|A_\ell|}| = 1 \quad , \quad (\forall\, \ell;\ h = 1, \dots, M_\ell)$$

being all complex roots, of the unit, located on the unit circle of the complex plane $\mathbb{C}^1$. Therefore the complex representation of a DNA sequence is a sequence of complex numbers

$$y_h = \xi_h + \eta_h i \quad , \quad \xi_h = \Re(y_h)\ ,\ \eta_h = \Im(y_h)$$

with y_h given by (9.1).

9.2. Random walks. Random walk on the complex sequence $\mathbf{Y}_N$ is defined as the series $\mathbf{Z}_N = \{z_n\}_{n=1,\dots,N}$

$$z_n \stackrel{\text{def}}{=} \sum_{k=1,\dots,n} y_k \quad , \quad n = 1, ..., N \tag{9.2}$$

which is the cumulative sum

$$\left\{ y_1\ , y_1 + y_2\ , \dots , \sum_{s=1}^{n} y_s \dots , \sum_{s=1}^{N} y_s \right\}\ .$$

When $y_k = \rho(x_k)$ with $\mathbf{X}_k \in D(S_N)$ we will properly call these walks as DNA walk. When the y_k are randomly generated we will call them random walks.

It has been observed that DNA walks have the typical shape of fractals (see Fig. 14).

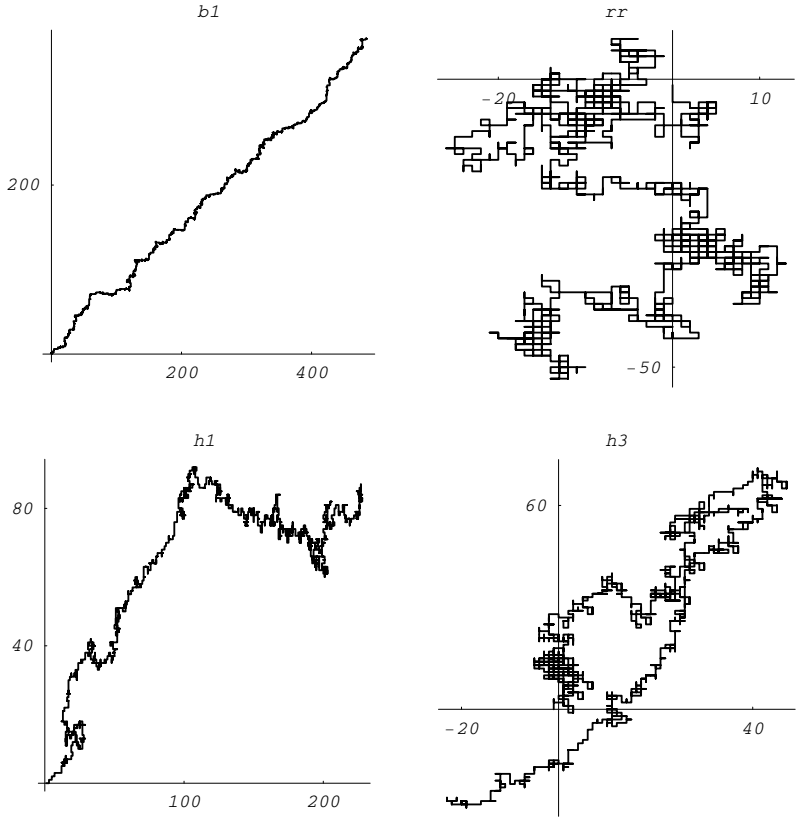

FIGURE 14. *Walks on the first 200 nucleotides: b1) Mycoplasma putrefaciens , b2) Mortierella verticillata , b3) Blattabacterium , h1) Aeropyrum pernix , h2) Acidianus hospitalis , h3) Acidilobus saccharovorans.*

9.3. Wavelet analysis on complex representation. As can be seen from Fig. 14, random walks and random sequences have a very special patterns.

The cluster algorithm for wavelet coefficients applied to the DNA walks, shows that the values of the wavelet coefficients belong to some discrete finite sets (Fig. 15). However, it should be noticed that this symmetry on detail coefficients is lost for wavelet transform on longer segments. There follows that DNA sequences have to be considered as a chain with short range dependence, in other words any acid nucleic is attached to the chain on the base of a (short) correlation with the previous acid nucleic.

The cluster algorithm applied to the DNA walks, shows that the values of the wavelet coefficients belong to some discrete finite sets (Fig. 15). However, it should be noticed that this symmetry on detail coefficients is lost for wavelet transform on longer segments. There follows that DNA sequences have to be considered as short range dependence, in other words any acid nucleic is attached

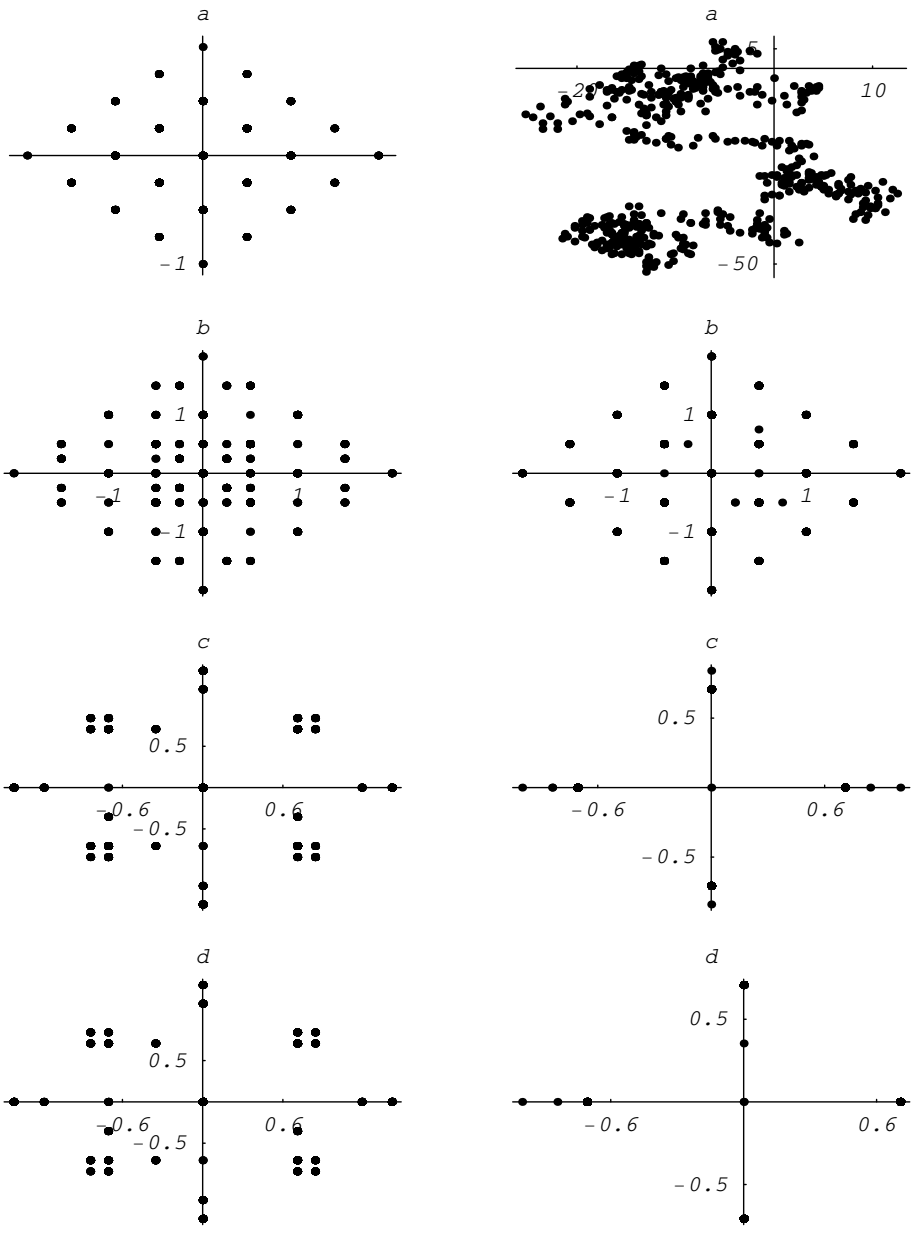

FIGURE 15. *Cluster analysis of the 4-th short Haar wavelet transform of a 4000-length random sequence (left) and its 2000-length random walk (right) : a)* (α, α^*) *; b)* $\left(\beta_0^0, \beta^{*0}_0\right)$ *; c)* $\left(\beta_0^1, \beta^{*1}_0\right)$ *; d)* $\left(\beta_1^1, \beta^{*1}_1\right)$.

to the chain on the base of a correlation of the previous acid nucleic. In other words, if we look for a dependence rule on the DNA nucleotides this dependence

might be summarized by a function as

$$x_{n+1} = f(x_n) \quad , \quad (n = 1, \ldots, N) \ .$$

It can be seen that the wavelet transform of $h3$ and its random permutation show that there are some similarities only at the level of detail coefficients however the more evolved sequences show a less energy.

10. Conclusions

In this paper some of the most popular methods of signal analysis for the analysis of complexity and multi-fractality of sequences have been shortly analyzed. They are particularly efficient for the analysis of binary images, by showing some properties that are difficult to single out on 1D sequences. Together with the clustering of wavelet coefficients this method has enabled us for the first time to characterize prime number distribution and nucleotide distribution.

References

[1] P.R. Aldrich, R. K. Horsley, S. M. Turcic, "Symmetry in the Language of Gene Expression: A Survey of Gene Promoter Networks in Multiple Bacterial Species and Non-σ Regulons", *Symmetry, 3* (2011) 1-20.

[2] M. Altaiski, O.Mornev, R.Polozov, "Wavelet analysis of DNA sequence", *Genetic Analysis, 12* (1996) 165-168.

[3] D. Anastassiou, "Frequency-domain analysis of biomolecular sequence", *Bioinformatics,* 16(12) (2000) 1073-1081.

[4] S. Ares, M. Castro, "Hidden structure in the randomness of the prime number sequence?", *Physica A* , 360 (2006) 285-296.

[5] A. Arneado, E.Bacry, P. V. Graves, J. F. Muzy, "Characterizing long-range correlations in DNA sequences from wavelet analysis", *Phys. Rev. Lett., 74* (1995) 3293-3296.

[6] A. Arneado, Y.D'Aubenton-Carafa, E.Bacry, P. V. Graves, J.F.Muzy, C.Thermes, "Wavelet based fractal analysis of DNA sequences?", *Physica D, 96* (1996) 291-320.

[7] A. Arneado, Y.D'Aubenton-Carafa, B.Audit, E.Bacry, J.F.Muzy, C.Thermes, "What can we learn with wavelets about DNA sequences?", *Physica A, 249* (1998) 439-448.

[8] B. Audit, C. Vaillant, A. Arneodo,Y. d'Aubenton-Carafa and C. Thermes, "Long range Correlations between DNA Bending Sites: Relation to the Structure and Dynamics of Nucleosomes", *J. Mol. Biol. 316* (2002) 903-918.

[9] Bai-Lin Hao, "Fractals from genomes-exact solutions of a biology inspired problem", *Physica A, 282* (2000) 225-246.

[10] J. A. Berger, S. K. Mitra, M. Carli, A. Neri, "Visualization and analysis of DNA sequences using DNA walks", *Journal of The Franklin Institutes, 341* (2004) 37-53.

[11] P. Bernaola-Galván, R. Román-Roldán, J. L. Oliver, "Compositional segmentation and long-range fractal correlations in DNA sequences", *Phys. Rev. E*, 53(5) (1996) 5181-5189.

[12] C. L. Berthelsen, J. A. Glazier, M. H. Skolnick, "Global fractal dimension of human DNA sequences treated as pseudorandom walks", *Phys. Rev. A*, 45(12) (1992) 8902-8913.

[13] B. Borstnik, D. Pumpernik, D. Lukman, "Analysis of apparent $1/f^{\alpha}$ spectrum in DNA sequences", *Europhys Lett. 23*, (1993) 389-394.

[14] S. V. Buldyrev, A. L. Goldberger, A. L. Havlin, C.-K. Peng, M. Simons, F. Sciortino, H. E. Stanley, "Long-range fractal correlations in DNA", *Phys. Rev. E 51*, (1995) 5084-5091.

[15] C. Cattani, "Haar Wavelet based Technique for Sharp Jumps Classification", *Mathematical Computer Modelling, 39* (2004) 255-279.

[16] C. Cattani, "Haar wavelets based technique in evolution problems", *Proc. Estonian Acad. of Sciences, Phys. Math.*, 53(1) (2004) 45-63.

[17] C.Cattani, "Fractal Patterns in Prime Numbers Distribution", Computational Science and its Applications, D. Taniar et Al.(Eds.) ICCSA, Springer-Verlag Berlin Heidelberg, LNCS 6017, Part 2, (2010), 164-176.

[18] C. Cattani, J.J. Rushchitsky, "Wavelet and Wave Analysis as applied to Materials with Micro or Nanostructure, Series on Advances in Mathematics for Applied Sciences", *World Scientific, Singapore*, 74 (2007).

[19] C. Cattani, "Complex representation of DNA sequences", Communications in Computer and Information Science, Proceedings of the "Bioinformatics Research and Development Second International Conference", BIRD 2008 Vienna, Austria, July 7-9, 2008, M. Elloumi *et al.* (Eds), Springer-Verlag Berlin Heidelberg, CCIS 13, (2008) 528–537.

[20] C.Cattani, "Harmonic Wavelet Approximation of Random, Fractal and High Frequency Signals", *Telecommunication Systems*, 43(3-4) (2010) 207–217.

[21] C.Cattani, "Wavelet Algorithms for DNA Analysis", Chapter 35, in *Algorithms in Computational Molecular Biology: Techniques, Approaches and Applications*, (Wiley Series in Bioinformatics) by Mourad Elloumi and Albert Y. Zomaya, John Wiley & Sons, (2010) 799–842.

[22] C.Cattani, "Fractals and Hidden Symmetries in DNA", *Mathematical Problems in Engineering*, vol. 2010, (2010) 1–31. doi:10.1155/2010/507056

[23] C.Cattani, "On the Existence of Wavelet Symmetries in Archaea DNA", *Computational and Mathematical Methods in Medicine*, vol. 2011, (2011) 1–21.

[24] C.Cattani, "Complexity and Simmetries in DNA sequences", Chapter 22, in *Handbook of biological discovery*, (Wiley Series in Bioinformatics) by Mourad Elloumi and Albert Y. Zomaya, John Wiley & Sons, (2012) 700–742.

[25] C. Cattani, "Wavelet algorithms for DNA analysis. In Algorithms in Computational Molecular Biology: Techniques, Approaches and Applications", M. Elloumi and A. Y. Zomaya, (Eds.), Wiley Series in Bioinformatics, John Wiley & Sons, (2013) 799–842.

[26] C. Cattani, A. Ciancio, "On the fractal distribution of primes and prime-indexed primes by the binary image analysis", *Physica A*, 460 (2016), 222–229.

[27] C. Cattani, G. Pierro, "Complexity on Acute Myeloid Leukemia mRNA Transcript Variant", *Mathematical Problems in Engineering*, vol. 2011, (2011), 1–16.

[28] C. Cattani, G. Pierro, "On the Fractal Geometry of DNA by the Binary Image Analysis", *Bullettin of Mathematical Biology*, (2013), 1–27.

[29] C.Cattani, G. Pierro, G. Altieri, "Entropy and multi-fractality for the myeloma multiple TET 2 gene", Mathematical Problems in Engineering, MPE vol. 2011, (2011) 1–17. doi:10.1155/2011/193761

[30] E. A. Cheever, D. B. Searls, W. Karanaratne, G. C. Overton, "Using signal processing techniques for DNA sequence comparison", *Proc. 15th Annu. Northeast Bioeng. Conf.*, (1989) 173–174.

[31] P. D. Cristea, "Large scale features in DNA genomic signals", *Signal Processing, 83*, (2003) 871–888.

[32] E. Coward, "Equivalence of two Fourier methods for biological sequences", *Journal of Mathematical Biology, 36*, (1997) 64–70.

[33] I. Daubechies, *Ten Lectures on wavelets.* SIAM, Philadelphia, PA, (1992).

[34] R.H.C. de Melo, A Conci, Succolarity: Defining a Method to calculate this Fractal Measure. System, Signals and Image Processing, (2008) 291–294.
[35] R.H.C. de Melo, A Conci, How Succolarity could be used as another fractal measure in image analysis. Telecommun. Syst, (2011) 1–13.
[36] G. Dodin, P. Vandergheynst, P. Levoir, C. Cordier, L. Marcourt, "Fourier and Wavelet Transform Analysis, a Tool for Visualizing Regular Patterns in DNA Sequences", *J. Theor. Biol., 206* (2000) 323–326.
[37] J.P. Eckmann, S.O. Kamphorst, D. Ruelle, "Recurrence Plots of Dynamical Systems", *Europhysics Letters, 5* (1987) 973–977.
[38] H. M. Edwards, Riemann's zeta-function, New York-London, Academic Press 1974.
[39] Farjamnia Gh., Gashti M.Z, Barangi H., Gasimov Y.S. "The study of support vector machine to classify the medical data", International Journal of Computer Science and Network Security, Vol.17, No.12, (2017), 145–150.
[40] P. Faure, A. Lesne, "Recurrence plots for symbolic sequences", *International Journal of Bifurcation and Chaos*, 20(6), (2010) 1731–1749.
[41] R. Ferrer-i-Cancho, N. Forns, "The self-organization of genomes", *Complexity, 15* (2010) 34–36.
[42] J.P. Fitch, B. Sokhansanj, "Genomic engineering: moving beyond DNA sequence to function", *Proc. IEEE 88*, (12) (2000) 1949–1971.
[43] M. A. Gates, "Simpler DNA sequence representations", *Nature 316*, 219 (18 July 1985) doi:10.1038/316219a0.
[44] M. A. Gates, "A simple way to look at DNA", *J. Theor. Biol., 119* (1986) 319–328.
[45] H. Gee, "A journey into the genome: what's there", *Nature*, 12 February 2001, http://www.nature.com/nsu/010215/010215-3.html.
[46] H. Herzel, E.N.Trifonov, O.Weiss, I.Grosse, "Interpreting correlations in biosequences", *Physica A 249*, (1998) 449-459.
[47] K. Hu, P. Ch. Ivanov, Z. Chen, P. Carpena, H. E. Stanley, "Effect of trends on detrended fluctuation analysis", *Phys. Rev. E 64*, (2001) 011114.
[48] X.-Y. Jiang, D. Lavenier, S. S.-T. Yau, "Coding Region Prediction Based on a Universal DNA Sequence Representation Method", *Journal of Computational Biology*, 15(10) (2008) 1237–1256.
[49] E. Hamori, J. Ruskin, "H Curves, A Novel Method of Representation of Nucleotide Series Especially Suited for Long DNA Sequences", *The Journal of Biological Chemistry*, 258(2) (1983) 1318–1327.
[50] H. Herzel, E.N.Trifonov, O.Weiss, I.Grosse, "Interpreting correlations in biosequences", *Physica A 249*, (1998) 449–459.
[51] J. L. Howland, *The Surprising Archaea*, Oxford University Press, New York and Oxford, 2000.
[52] S. Karlin, V. Brendel, "Patchiness and correlations in DNA sequence", *Science 259*, (1993) 677–680.
[53] A. Laghrib, "A comparative study between tv, tv2 , btv and combined models for the multi-frame super-resolution", Advanced Mathematical Models & Applications, Vol.5, No.1, (2020), 80–94.
[54] J. E. Littlewood, "Sur la distribution des nombres premieres", *C. R. Acad. Sci. Paris*, 158 (1914) 1869-1872.
[55] M. T. Madigan, B. L. Marrs, "Extremophiles", *Scientific American, 4* (1997), 82–87.
[56] M. Li, "Fractal Time Series-A Tutorial Review", *Mathematical Problems in Engineering*, vol. 2010 (2010) 1–26. doi:10.1155/2010/157264
[57] M. Li and J.-Y. Li, "On the predictability of long-range dependent series", *Mathematical Problems in Engineering*, vol. 2010, (2010) 1–9. doi:10.1155/2010/397454

[58] M. Li and S. C. Lim, "Power spectrum of generalized Cauchy Process", *Telecommunication Systems, 43* (3-4), (2010) 219–222.
[59] W. Li, "The complexity of DNA: the measure of compositional heterogenity in DNA sequence and measures of complexity", *Complexity 3* , (1997) 33–37.
[60] W. Li, "The study of correlation structures of DNA sequences: a critical review", *Computer Chem.*, 21(4) (1997) 257–271.
[61] W. Li and K. Kaneko, "Long-range correlations and partial $1/f^{\alpha}$ spectrum in a noncoding DNA sequence", *Europhys. Lett. 17* , (1992) 655–660.
[62] J.G. McNally, D. Mazza, "Fractal geometry in the nucleus", *EMBO J., 29* (2010) 2–3.
[63] M. Nachaoui, "Parameter learning for combined first and second order total variation for image reconstruction", Advanced Mathematical Models & Applications, Vol.5, No.1, 2020, 53–69.
[64] W. Narkiewicz, The development of prime number theory, Springer, 2000.
[65] N. Marwan, J. Kurths, Nonlinear analysis of bivariate data with cross recurrence plots. Physics Letters A, 30(5–6), (2002) 299–307.
[66] N. Marwan,M.C. Romano, M. Thiel, J. Kurths, Recurrrence Plots for the Analysis of Complex Systems. Physics Reports, 438(5-6), (2007) 237–329.
[67] K. Metze, I. Lorand-Metze, N.J. Leite, R.L. Adam, "Goodness-of-fit of the fractal dimension as a prognostic factor". *Cell Oncol., 31* (2009) 503–504.
[68] K. Metze, "Fractal dimension of chromatin and cancer prognosis", *Epigenomics, 2* (5), (2010) 601–604.
[69] T. Misteli, "Self-organization in the genome", *Proc. Natl. Acad. Sciences USA, 106* (2009) 6885–6886.
[70] Li Ming, Fractal time series — a tutorial review. Mathematical Problems in Engineering, Vol. 2010, (2010) 1–26.
[71] K. B. Murray, D. Gorse, J. M. Thornton, "Wavelet Transform for the characterization and detection of repeating motifs", *JMB, J. Mol. Biol. 316*, (2002) 341–363.
[72] National Center for Biotechnology Information, http://www.ncbi.nml.nih.gov/GenBank; Genome Browser, http://genome.ucsc.edu; European Informatics Institute, http://www.ebi.ac.uk; Ensembl, http://www.ensembl.org.
[73] C.-K. Peng, S.V. Buldryev, A.L. Goldberg, S. Havlin, F. Sciortino, M. Simons and H.E. Stanley, "Long-range correlatins in nucleotide sequences", *Nature 356* , (1992) 168–170.
[74] C.-K. Peng, S.V. Buldryev, S. Havlin, M. Simons, H.E. Stanley and A.L. Goldberg,, "Mosaic organization of DNA nucleotides", *Phys. Rev. E 49* , (1994) 1685–1689.
[75] D. B. Percival and A.T. Walden, *Wavelet Methods for Time Series Analysis*, Cambridge University Press, 2000.
[76] G. Pierro, Sequence complexity of Chromosome 3 in Caenorhabditis elegans. Advances in Bioinformatics, Vol. 2012, (2012) 1–12.
[77] R. E. Plotnick, R.H. Gardner, R.V. O'Neill, Lacunarity indices as measures of landscape texture. Landscape Ecology, 8(3), (1993) 201–211.
[78] R. E. Plotnick, R.H. Gardner, W.W. Hargrove, K. Prestegaard, M. Perlmutter, Lacunarity analysis: A general technique for the analysis of spatial patterns. Physical Review E, 53(5), (1996) 5461–5468.
[79] M. Schlesinger, "On the Riemann hypothesis: a fractal random walk approach", *Physica A* , 138 (1986) 310-319.
[80] C.E. Shannon, "A Mathematical Theory of Communication", *The Bell System Technical Journal, 27* (1948) 379–423, 623–656.

[81] H. H. Shapiro, Introduction to the Theory of Numbers, John Wiley & Sons , New York 1983.
[82] R.V. Solé, "Genome size, self-organization and DNA's dark matter", *Complexity, 16* (2010) 20–23.
[83] J. Szczepanski, T. Michalek, "Random Fields Approach to the Study of DNA Chains", *Journal of Biological Physics 29*, (2003) 39–54.
[84] M. Stein and S. M. Ulam, "An Observation on the Distribution of Primes", American Mathematical Monthly (Mathematical Association of America) 74, (1) (1967) 43–44,
[85] M. Takahashi, "A fractal model of chromosomes and chromosomal DNA replication", *J. Theor. Biol., 141* (1989) 117–136.
[86] A. A. Tsonis, P.Kumar, J.B.Elsner and P.A.Tsonis, "Wavelet Analysis of DNA sequences", *Physical Review E, 53* (1996), 1828–1834.
[87] M. Stein and S. M. Ulam, "An Observation on the Distribution of Primes", American Mathematical Monthly (Mathematical Association of America) 74, (1) (1967) 43–44.
[88] P. P. Vaidyanathan, B.-J. Yoon, "The role of signal-processing concepts in genomics and proteomics", *Journal of The Franklin Institute, 341* (2004) 111–135.
[89] R. F. Voss, "Evolution of Long-Range Fractal Correlations and $1/f$ Noise in DNA Base Sequences", *Physical Review Letters*, 68(25) (1992) 3805–3808.
[90] R. F. Voss, "Long-Range Fractal Correlations in DNA introns and exons", *Fractals, 2*, (1992) 1–6.
[91] O. Weiss, H. Herzel, "Correlations in protein sequences and property codes", *J. Theor. Biol. 190*, (1998) 341–353.
[92] G. Wornell, *Signal Processing with Fractals: A Wavelet-Based Approach*, Prentice Hall, (1996).
[93] C. R. Woese , G. E. Fox, "Phylogenetic structure of the prokaryotic domain: The primary kingdoms", *Proc. Natl. Acad. Sci. 74*, (1977) 5088–5090.
[94] S.S.-T., Yau, J. Wang, A. Niknejad, C. Lu, N. Jin, Y.-K. Ho, "DNA sequence representation without degeneracy", *Nucleid Acids Res., 31*, (2003) 3078–3080.
[95] Z. G. Yu, W. W. Anh, B. Wang, "Correlation property of length sequences based on global structure of the complex genome", *Phys. Rev. E, 63*, (2001), 011903.
[96] R. M. Yulmetyev, N. A. Emelyanova, F. M. Gafarov, "Dynamical Shannon entropy and information Tsallis entropy in complex systems", *Physica A, 341* (2004), 649–676.
[97] M. Zhang, "Exploratory analysis of long genomic DNA sequences using the wavelet transform: examples using polyomavirus genomes", *Genome Sequencing and Analysis Conference VI*, (1995) 72–85.

Carlo Cattani
Engineering School, DEIM, Tuscia University, Via del Paradiso 47, 01100 Viterbo, Italy
E-mail address: cattani@unitus.it

Received: September 9, 2020; Accepted: November 6, 2020