# Fractal Geometry of a Dendrogram

Francisco Casanova-del-Angel
SEPI ESIA Instituto Politécnico Nacional. México. fcasanova@ipn.mx

**Abstract**

The fractal structure shown by the poligonal created by unions of middle points of vertices, nodes or peaks of dendograms terminal classes is presented. Its generating fractal, the details of its construction, and the way to measure its segments are defined; its property of inverted scale, the type of meshing, its property of axial symmetry and a theorem on transformation of linear affinity are considered. This is exemplified by means of two applications with real data.

**Keywords**: dendrogram, fractal, fractal propagation, axial symmetry and generating curve.

## 1. Introduction

Classification is a mathematical technique used for factorial taxonomization and factorial description of data under study. In data analysis, classification, whether hierarchical or not, is highly present, and is often used in psychology, sociology, linguistics and archeology, that is, in Human Sciences. Classification methods are also widely used in biology, medicine, botanics, zoology and ecology, as well as in physics, economy and history. It is well known that the most important contribution of such techniques is their usefulness as complementary methodology for other statistical methods, since it helps fairly efficiently to interpret groups of homogeneus objects which a factorial analysis is not capable to establish. As we know, classification technices generate a dendrogramatic relationship called “tree”, which ends at a point where its aggregation index is not so high, thus providing real partitions and defining classes. A helpful technique for reading and interpreting a hierarchical dendrogram is presented below, consisting of characterizing fractally the poligonal created by unions of middle points of vertices, nodes or peaks of terminal classes of a hierarchical tree, which allows us to make sure that the hierarchy thus created is, if not optimal, a good taxonomization.

## 2. Characterization of dendrogramatic fractal

*Definition of generating curve*. Let $I_0$ be a unitary lenght line segment, contained in a closed interval, that is, $I_0 \subset [a, b]$. Let $I_1$ be a set with sectioned behavior, consisting in three segments of a straight line which create, based on starting point $a$ of $I_0$, two scalene triangles reflected regarding the middle point $c$ of $I_0$, obtained as follows: the first half of segment $I_1$ is substituted or removed by the sides of triangle which create an angle with $I_0$. This process is repeated for the second half, but with the sides reflected from middle point $c$. This process is known as *generator* which is called state 1. Construction of set $I_2$ is made applying the generator to every segment of $I_1$, which is called state 2. Thus, set $I_k$ is created applying generator $I_1$ on every segment of $I_{k-1}$, which is called state $k$, Figure 1.a. Figure 1.b shows the curve built on the plane through reflection of $I_k$ on every side.

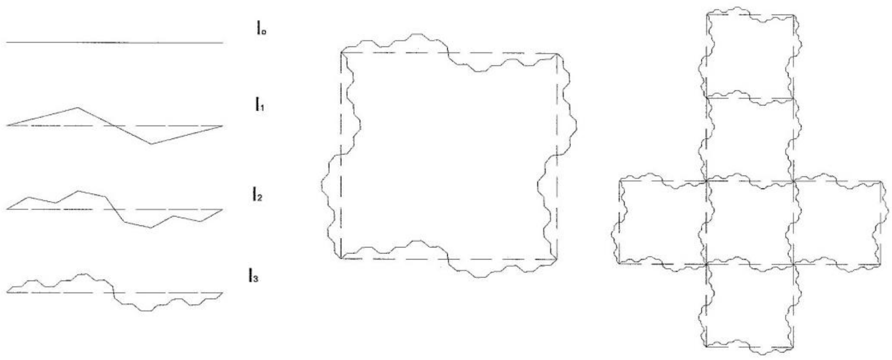
Curve $I$ is characterized by being of a similar scale since, based on a transformation $F$: $\Re^n \to \Re^n$ with $\lambda_i > 0\ \forall$ i $\exists_n\ a, b \in \Re^n$ such that $|F_i(a)-F_i(b)| = \lambda_i |a-b|$. Similarity of scale is present for triangles created with $I_0$, by generator $I_1$. In like manner, it also has the property to be affine since, based on transformation $F$ already defined, $F(a) = T(a) + \alpha$ with $T$ a non-singular linear transformation and $\alpha \in \Re^n$. It must be remembered that affinity is conceived as a shear transformation or resistant to cutting, and is a contracting or expanding effect, not necessarily in the same direction ***(Mandelbrot, 1999)***.

A curve type $l_k$ satisfies the *scale principle* if all its relative figures are linked to each other by a scale law. Let $I$ be a Borel's set such that $I = \{I^1, I^2, I^3\}$, where $I^j$ is a finite succesion of line segments creating the *generator* $\forall$ j = 1, 2, 3, in order that $I_i = \cup^3_{j=1} I^j_i$ $\forall$ i = 1, 2, … is a countable sequence of sets. Thus, measurement $\mu$, of segments $I_i$ is defined as:

$$\mu(\cup^{\infty}_{i=1} I_i) = \sum^{\infty}_{i=1} \mu(\cup^{3}_{j=1} I^j_i) = \mu(I_i) = 3\mu(I_{i-1}) \quad \forall \;\; i = 1, \ldots \qquad 1$$

When a geometrical discontinuity is of the fractal type, generated by a natural process, a uniform reticulate should be built. Let $(\chi, P(\chi), \mu)$ be a space with measures such that the sample space is $\chi = [0, 1]$ and $P(\chi)$ is a set of subsets of $\chi$, where the measurement is $\mu$. Since the system is dynamic, $\chi \subseteq \Re^p$ is the phase space. Let us consider a $\chi$ reticulate covered by $p$-dimensional boxes with radius $\delta_n$, where $B_{\delta n}(t)$ is the neighbor box containing the segment of straight line or point $t$. Succession of neighbor boxes has radius $\delta_n \rightarrow 0$ as $n \rightarrow \infty$.

State Measure




a) b) c)

Figure 1. a) Construction of the fractal curve $I$. In every state $I_k$ generator $I_1$ is applied on every segment of the curve. b) Fractal curve $I$ on the plane, and c) cubic form of fractal.

*Theorem (of inverted scale property)*. Let $F^{-1}$ be an inverted similar transformation of scale factor $k > 0$, such that $F^{-1}: \Re^n \rightarrow \Re^n$. Si $\exists\, II \subset \Re^n$ then $\lambda^k H^k(F^{-1}(II)) = H^k(II)$ $\forall\, \lambda > 0$.

*Dem*: since $\{I_i\}$ is a countable set of neighbors $\delta$ covering $I$, if we apply the inverted similar transformation: $F^{-1}\{I_i\} = \{F^{-1}(I_i)\}$ which is a $\delta/\lambda = \lambda^{-1}\delta$ neighbor of $F^{-1}(II)$, that is: $\lambda^k H^k_{\delta/\lambda}(F^{-1}(II)) \leq H^k(II)$ $\forall\, \lambda > 0$. On the limit when $\delta \rightarrow 0$, the above inequality becomes:

$$H^k(F^{-1}(II)) = \lambda^{-k} H^k(II)$$

**QED**

The theorem herein proven allows us to reduce the length of a fractal object.

**Meshing and definition of fractal outline**

Considering a random generation of $f(t)$, the original curve is rotated to different angles, preferably constant, in order to calculate $D_{\theta°}$ for every case.

$$D_{\theta°} = \frac{\sum_i \{D_{\theta° i} \ \forall \ i = 1, \ldots, n\}}{n} \qquad 2$$

In order to rotate the original curve a certain number of times, let us consider mapping $M_n$: $\chi \rightarrow \Re$, where $M_n(t) = -\log \mu[B_{\delta n}(t)]$, if $\mu[B_{\delta n}(t)] > 0$ then $C_n(t)$ is a re-scalated version of $M_n(t)$, that is: $C_n(t) = M_n(t)/(-\log \delta_n)$, where $C_n$ describes the local behavior of $\mu$ measurement.

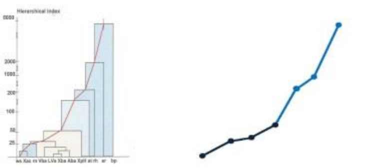
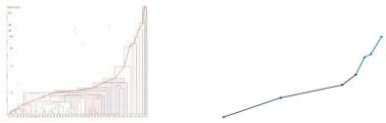
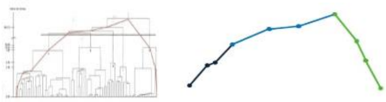
**Application**

The aim of this application is to define and statistically analyze the effects of atmospheric corrosion on structural steel of civil infrastructure in Mexico City, where four industrial sites are located: Vallejo, Azcapotzalco, Xalostoc and Tultitlán. Analyzed and used data comprises three data tables. A table of meteorological data, *IxJ1*, containing weekly average values, obtained from the Experimental Meteorological of the Instituto Politécnico Nacional, which comprises five variables: wind speed, air temperature, relative humidity, solar radiation and rainfall. A data table, *IxJ2*, containing weekly maximum average concentrations of two pollutants sulfur dioxide recorded in Vallejo, La Villa, Azcapotzalco and Xalostoc Stations, and chlorides or anions concentrations for maximum weekly values in Xalostoc Station, as well as Hp or hydrogen potential of rainfall water. The third data table, *IxJ3*, is related to values of corrosion speeds and contains values for both depositions.

After applying the theory of Data Analysis and having finally built a dendrogram, Table 1 was created, containing fifth columns. The first column takes a census of the fractal and its features; the second column contains the graphic of the fractal; the third column shows the calculated value of its fractal dimension, obtained in accordance with the theory presented and developed; the fourth column shows the standard deviation value of the equation value of the straight line with slope $D_{\theta°}$; and the fifth column contains the equation of the straight line with slope $D_{\theta°}$. Obtained values were checked using software Benoit 1.3 ***(Benoit 1.3. 2008)***. It must be noticed that $\sigma$ measures the probability that an observation is at a certain distance from the average observation, which is valid if the system analyzed is random.

Let us take the poligonals of application 1, which represent the sequence of incorporation of variables into the developed hierarchical analysis. Figures in Table 1 are not in the same hierarchical scale. For the first one, where the dendogram is built only for metereological variables, the poligonal is not enough to draft the trend of the final structure of the dendrogram under study. Figure 1, Table 1; where the variables of atmospheric pollution have been incorporated, shows the generator of the poligonal up to two times. The segments of the generator are built in a linear manner, from the base of a class to the terminal point of a higher class, such that the sequence of such set of classes is defining, in concept, part of the behavior of variables and that its hierarchical index shows how different is such set of classes from the remaining ones. This format does not reflect small cracks shown by the poligonal for such path within the original dendrogram.

Table 1. Parameters of fractal dimension of components of dendrograms in application.

| No. | Fractal | Fractal dimensión $D_{\theta°}$ | Standard deviation $\sigma$ | Equation of straight line with slope $D_{\theta°}$ |
|---|---|---|---|---|
| 1 |  | 1.96226 | 0.0081067 | 2.06 E+06 $X^{-1.96}$ |
| 2 |  | 1.96166 | 0.0079848 | 2.05 E+06 $X^{-1.96}$ |
| 3 |  | 1.96197 | 0.0080498 | 2.05 E+06 $X^{-1.96}$ |
| 4 | Generator $I_1$ o state 1  | 1.96267 | 0.0081833 | 2.06 E+06 $X^{-1.96}$ |
| 5 | Generator $I_2$ o state 2  | 1.96267 | 0.0081836 | 2.06 E+06 $X^{-1.96}$ |
| 6 | Generator $I_3$ o state 3  | 1.96266 | 0.0081803 | 2.06 E+06 $X^{-1.96}$ |

Every generator of the fractal curve defines a feature of data under study, which adds value to the generation of hierarchical dendrograms. For instance, fractal curve of Figure 3, Table 1, is created by three generating curves. The first generating curve is the segments of which define the stabilization of corrosion effect. The second generating curve defines the transition of corrosion effect, the effect of meteorology and the transition of the effect of exceeding humidity. The third generating curve defines the starting state of corrosion effect on samples studied ***(Casanova y Toquiantzi, 2008)***. It should be specially noticed that, as every generating curve defines a branch of the dendrogram, this happens also for the segments of the generating curve. For instance, the first segment of the generating curve adds gentle breeze, low radiation, average humidty and low $SO_2$. The second segment contains low rainfall, low salts, and pH lower than or equal to 5 units, while the third segment adds values of corrosion speeds, tube samples, 3/8” rods, Ts, channels, plates, and squares. Figure 4, Table 1, shows the ideal generator. Figures 5 and 6, Table 1, show the first two states of the generated ideal generator. It should be noticed that its calculated fractal dimension is the same for the three cases and that the four first decimal figures of the dimensions of all the fractal curves coincide.

**Conclusions**

The property of axial symmetry of the generating curve proves the affinity of triangles created by such, with state zero, Figure 2.b. Characterizing the generating curve as a countable sequence of sets, allows us to prove the existence of the transformation to linear affinity of such.

It has been specially developed herein a methodology for the fractal analysis of geometrical discontinuity or polygon created by the unions of middle points of vertices, nodes or peaks of dendrograms terminal classes: i) the outline of the dendrogram under study is drawn, joining middle points of vertices, nodes or peaks of terminal classes, verifying the sectioned behavior of the polygon; ii) every mirror behavior of any section of the curve is identified; iii) the curve showing the direction of the fractal propagation is isolated and a straight line is drawn from the starting point to the ending point, calculating the angle of propagation regarding the ordinates axis; iv) the scale factor of the fractal curve is calculated; v) the generator of the curve is scale-reproduced; vi) meshings are created for the polygon in order to calculate its fractal dimension; vii) fractal dimension is calculated for every different meshing and rotation, and the results are shown on a graph; and viii) the auto-affinity property of the polygon is checked.

It should be noticed that every branch of the generator or generating curve defines or implies a part of the hierarchy already built.

**Acknowledge**

This document was developed with part of the time devoted to the IPN-SIP 20100674 and IPN-SIP 20110437 research projects.

**References**

**Benoit 1.3. 2008**. TruSoft Int´l Inc. USA. www.trusoft.com

**Casanova-del-Angel, F and Toquiantzi Butrón, R. 2008**. Corrosion phases of structural shapes exposed to the atmosphere. Corrosion Science. Vol. 50, issue 8, August (2008) pp. 2288-2295. ISSN: 0010-938X doi: 10.1016/ j.corsi2008.05.015.

**Mandelbrot, B. B. 1999**. *Multifractals and 1/f noise*. Springer-Verlag. ISBN: 0-387-98539-5.