2012

Recurrence Relations, Fractals, and Chaos: Implications for Analyzing Gene Structure

Sarah. M. Harmon
Colby College

Recurrence Relations, Fractals, and Chaos:

Implications for Analyzing Gene Structure

Sarah M. Harmon

Colby College

Abstract

The “chaos game” is a well-known algorithm by which one may construct a pictorial representation of an iterative process. The resulting sets are known as fractals and can be mathematically characterized by measures of dimension as well as by their associated recurrence relations. Using the chaos game algorithm, is it possible to derive meaningful structure out of our own genetic encoding, and that of other organisms? In this paper, I will present one method of applying the chaos game to biological data and subsequently will discuss both the mathematical and biological implications of the results.

Keywords: chaos game, fractional dimension, gene structure

Recurrence Relations, Fractals, and Chaos: Implications for Analyzing Gene Structure

I. Introduction

In the past thirty years, research exploring the relationship between biology and chaos theory has intensified. Recursive sequences were initially used to model populations. Later, however, more and more biological processes and topics began to be modeled, explained, and predicted by using fractals and chaotic systems (e.g., Dastgheib et al., 2011; Devaney, 1990; Esteller et al., 1999; Fiser et al., 1994; Kuroda & Tsuda, 2004; Levy & Pollack, 2001; Paramanathan & Uthayakumar, 2007; Pollack, 1991; Tino, 1998; Werner, 2010). Some may be traced back to a single publication, which proposed the idea of using a method of fractal generation called the chaos game to describe gene sequences (Jeffrey, 1990). I will provide the reader with a general mathematical foundation regarding recurrence relations, fractals, and fractional dimension before describing this application and its consequences more specifically.

II. Recurrence Relations

Every infinite sequence of numbers

$$a_0, a_1, a_2, ..., a_n, ...$$

is associated with an infinite series

$$g(x) = a_0 + a_1x + a_2x^2 + ... + a_mx^m + ...,$$

which is called its *generating function*. For a finite sequence

$$a_0, a_1, a_2, ..., a_n,$$

the generating function simply becomes

$$g(x) = a_0 + a_1x + a_2x^2 + ... + a_mx^m.$$

The fact that every sequence has an associated generating function is useful when solving *recurrence relations*. A recurrence relation is of the form

$$x_{n+1} = f(x_n, x_{n-1}, ..., x_{n-m})$$

for some function *f* and integer *m*. As an example, consider the recurrence relation

$$a_n = a_{n-1} + a_{n-2}$$

given that $a_0 = 0$ and $a_1 = a_2 = 1$.* In order to find a *closed-form* (non-recursive) solution of this equation, consider

$$g(x) = a_0 + a_1 x + a_2 x^2 + ... + a_n x^n + ...$$

Then, since $a_0 = 0$,

$$xg(x) = a_1 x^2 + a_2 x^3 + a_3 x^4 + ... + a_{n-1} x^n + ..., \text{ and}$$

$$x^2 g(x) = a_1 x^3 + a_2 x^4 + a_3 x^5 + ... + a_{n-2} x^n + ...$$

Thus,

$$g(x) - xg(x) - x^2 g(x) = a_1 x + (a_2 - a_1)x^2 + (a_3 - a_2 - a_1)x^3 + ...$$
$$+(a_n - a_{n-1} - a_{n-2})x^n + ...$$
$$= x$$

by the given conditions. Hence, we conclude that since

$$(1 - x - x^2)g(x) = x$$

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

$$= \frac{1}{\sqrt{5}} \left(\frac{1}{1 - cx} - \frac{1}{1 - dx} \right) \text{ for } c = \frac{1 + \sqrt{5}}{2}, \; d = \frac{1 - \sqrt{5}}{2}$$

and subsequently that

* The observant reader may notice this recurrence relation describes the Fibonacci sequence.

$$\sqrt{5}g(x) = \frac{1}{1-cx} - \frac{1}{1-dx}$$

$$= \sum_{n=0}^{\infty} c^n x^n - \sum_{n=0}^{\infty} d^n x^n$$

$$= \sum_{n=0}^{\infty} (c^n - d^n) x^n$$

$$g(x) = \sum_{n=0}^{\infty} \frac{(c^n - d^n)x^n}{\sqrt{5}}$$

the closed-form solution of the original recurrence relation is

$$a_n = \frac{c^n - d^n}{\sqrt{5}}$$

$$= \frac{c^n - d^n}{c - d}$$

Notice that c and d are the *roots* to the *characteristic equation* $r^2 - r - 1 = 0$ that has the same coefficients as the original recurrence relation when written in the form $a_n - a_{n-1} - a_{n-2} = 0$.

Some recurrence relations, such as the *logistic equation*

$$x_{n+1} = ax_n + bx_{n,}^2$$

are nonlinear and cannot be solved analytically. This is related to the fact that nonlinear systems may be chaotic. Notably, a chaotic system is one that exhibits sensitivity to initial conditions, in that small changes to its initial conditions will produce drastically different results with no discernable pattern. Given an initial condition x_0, the sequence of points obtained ($x_0, x_1, \ldots$) is called the *orbit* of x. If the orbit is periodic, then the repeating sequence $p_1, p_2 \ldots p_k$ in the orbit is a *cycle* of length K. Then, in addition to

sensitive dependence on initial conditions, a chaotic system has the following properties (Cull et al., 2005):

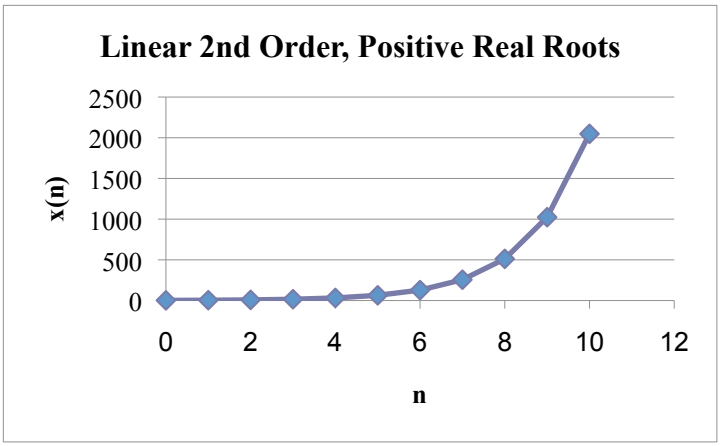
(1) the existence, depending on x_0, of cycles of every length,

(2) the existence of bounded but aperiodic orbits, and

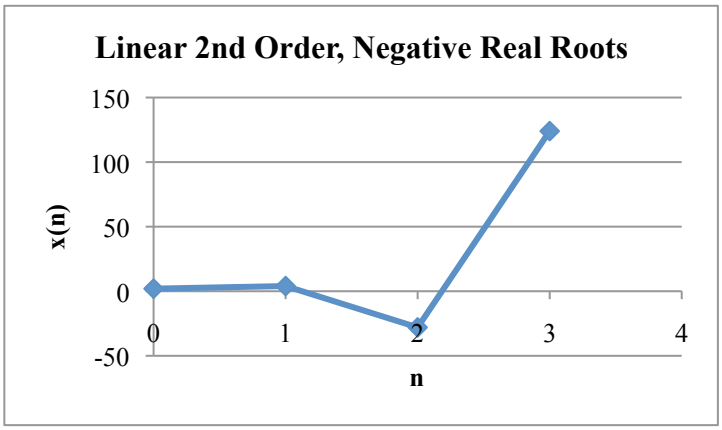
(3) for every open set A and every open set B there is an $x \in A$ such that if $x_0 = x$ then $x_k \in B$ for some k.

The graph of a recurrence relation is obtained by plotting values of x_n versus n, and the roots of the characteristic equation for a given recurrence relation are directly related to properties of its graph (Fig. 1). The magnitude of a root, for instance, determines the "steepness" of a graph's curve. Having all real positive roots tends to result in a smooth, asymptotic curve, whereas oscillations result from the presence of negative or complex roots. Trivially, negative roots will cause oscillations with a period of 2, whereas complex roots will generate oscillations with a period of 4. Initial conditions also affect the appearance of the graph of a recurrence relation; they may cause "initial oscillations" in the graph or help to determine whether the graph tends toward negative or positive infinity.

A



B



C

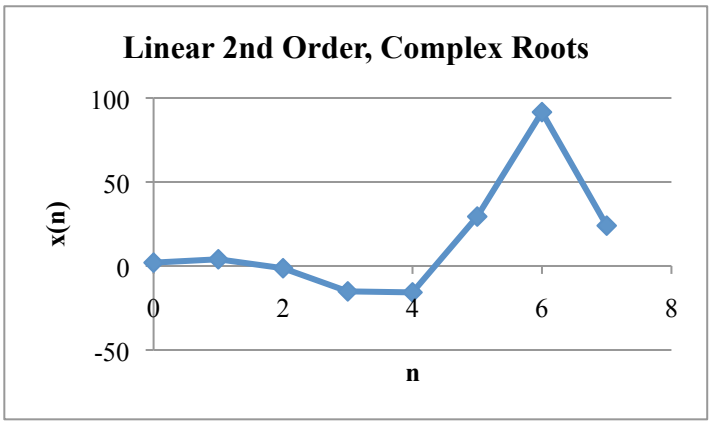


Figure 1. Sample graphs of simple recurrence relations. In (A), an equation with positive real roots ($x_n - 5x_{n-1} + 6x_{n-2} = 0$) is shown with closed-form solution $x_n = 2 \cdot 2^n$, whereas in (B), an equation with negative real roots ($x_n + 5x_{n-1} + 4x_{n-2} = 0$) is shown with closed-form solution $x_n = 2 \cdot (-4)^n$. In (C), an equation with complex roots ($3x_n - 4x_{n-1} + 10x_{n-2} = 0$) is shown, with closed-form solution $x_n = (1 + 2i\sqrt{\frac{2}{13}})(\frac{1}{3}(2 - i\sqrt{26}))^n + (1 - 2i\sqrt{\frac{2}{13}})(\frac{1}{3}(2 + i\sqrt{26}))^n$.

III. Fractals

A fractal is informally defined as a pattern with parts that are statistically characteristic of its entirety, i.e., a pattern that looks the same regardless of the degree of magnification. In contrast, a pattern is self-similar if it is congruent to a uniform scaling of itself. Fractals tend to be self-similar, but do not have to be. Some fractals are merely self-affine (Fig. 2), which means that their subunits are scaled by different amounts for different directions upon successive iterations.

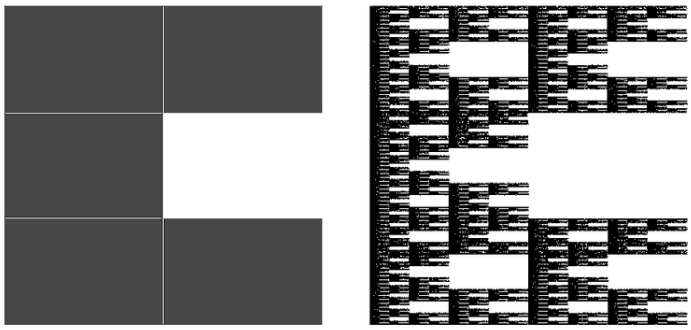


Figure 2. An example of a self-affine fractal (Prokofiev, 2009).[**]

The *Cantor set* has been described as the simplest fractal, and is constructed by iteratively removing the open middle thirds of a set of line segments. Mathematically, the n^{th} set X_n is expressed as

$$X_n = [\frac{1}{3}X_{n-1}] \cup [\frac{1}{3}X_{n-1} + \frac{2}{3}].$$

The Cantor set is defined as the intersection of this sequence of sets.

[**] Fractional dimension will be described in a later section, but it is worthwhile to note here that this fractal is an example of a set whose Hausdorff dimension and Minkowski-Bouligand dimension are not equal.

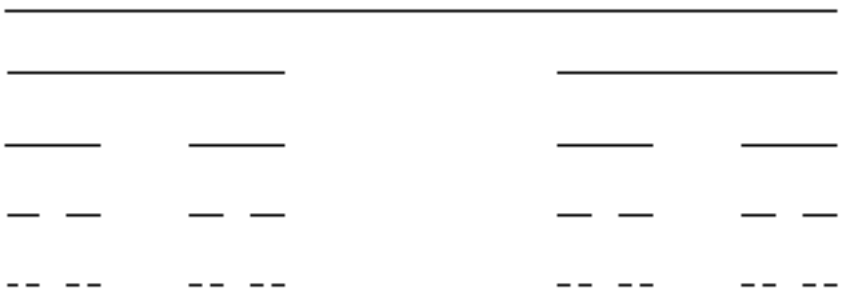


Figure 3. The first five iterations of the sequence that leads to the Cantor set.

The (quadratic) *Julia set* is also commonly known. Given some constant c, the forward iterations of a complex point x_0 are generated using the recurrence relation:

$$x_{n+1} = x_n^2 + c.^{***}$$

The set of points generated is the Julia set and can be plotted in the complex plane. Not all values of c will result in a fractal (Dufner et al. 1998). For instance, a fractal will not be produced if c is -2 or 0. When a fractal does result, it will be self-similar (Peitgen & Richter, 1986). The *Mandelbrot set* is the set of all complex c such that $|x_n|$ is bounded when starting with initial condition $x_0 = 0$ (Fig. 5).

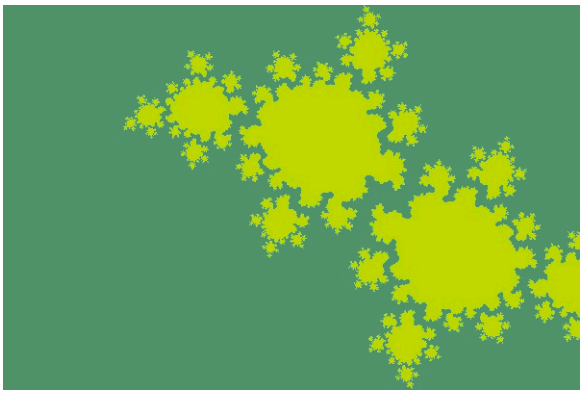


Figure 4. The quadratic Julia set implemented in Python, using $c = -0.4 + 0.6i$.

*** By inspection, the reader may notice this relation is comparable to the logistic equation mentioned earlier.

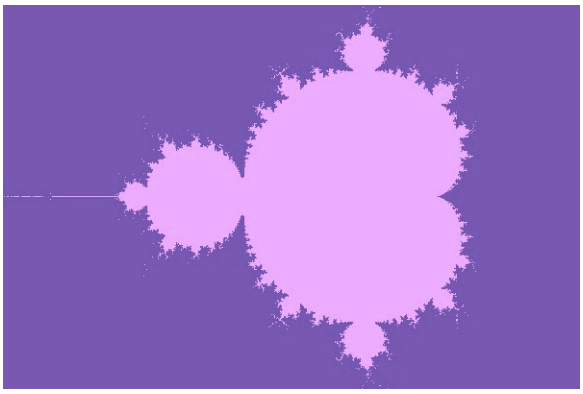


Figure 5. The Mandelbrot set implemented in Python, where $c = a + bi$ with the window $0.5 < a < 1$ and $-1 < b < -0.95$.

The *Sierpinski gasket*, or "Sierpinski triangle", can be easily compared to the Cantor set in that it is constructed by beginning with any closed triangle on a plane, splitting the triangle into four smaller triangles of equal size, and removing the open set of the middle triangle (Fig. 6). We will address how to construct the Sierpinski gasket via recurrence relations in the next section.

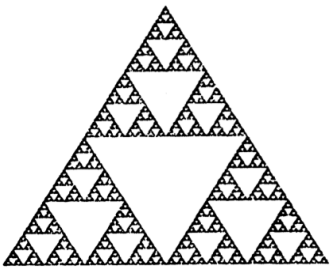


Figure 6. The well-known "Sierpinski gasket" (Jeffrey, 1990).

IV. Generating Fractals

Consider the following situation: start at the ½ mark on the unit interval [0,1] and toss a fair coin. If the coin lands heads up, plot a new point two-thirds of the way toward the endpoint 0. If the coin lands tails up, plot a new point two-thirds of the way toward the endpoint 1. As one continues to flip the coin and plot new points based on the location of

the previous point, the plotted points eventually begin to resemble the Cantor set.

To understand why this happens, imagine that any possible point in the interval could serve as an initial point. The two outcomes of the coin flip guarantee that the next point will be within a range of

$$\left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right],$$

leaving a prominent gap. Upon the second toss of the coin, there are four connected regions where the next possible point may land, namely,

$$\left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right].$$

Continuing this process, we notice that each time we are removing the open middle third of each remaining closed interval, which is exactly the procedure that leads to the construction of the Cantor set.[****]

We can play a similar "game" by starting at the center of a triangle and choosing a vertex at random towards which to move halfway. Continuing to choose and move towards new vertices at random from each preceding point will generate the Sierpinski gasket. Similar to playing the game on the unit interval, there are areas where points will never be plotted due to the nature of the game's conditions.

The situations just described are examples of using an *iterated function system* (IFS) to generate a fractal. An IFS can be expressed via linear equations that provide a recursive

[****] In fact, every point produced through this game is an *accumulation point* of the Cantor set. A point x is called an accumulation point of a set S if there exists a sequence x_n in S which converges to x (i.e., $\lim_{n\to\infty} x_n = x$). Often, fractals represent the accumulation points of iterated functions.

sequence (Barnsley & Demko 1985). For instance, the Sierpinski gasket can be produced using following equations, where x_n and y_n represent the new coordinates, x_{n-1} and y_{n-1} represent the preceding coordinates, and v_x and v_y represent the x- and y-coordinates of the chosen vertex, respectively:

$$x(n) = \frac{1}{2}\left(x_{n-1} + v_x\right) \qquad y(n) = \frac{1}{2}\left(y_{n-1} + v_y\right)$$

This method of approximate fractal generation is sometimes known as the *chaos game*.

When a pair of equations from the stated IFS is selected at random in the chaos game, we can make an assertion about whether a fractal will form with defined gaps. For instance, consider the fact that the process of moving halfway towards a vertex of a given figure maps the entire figure into one half of its height (Fig. 7). In any given odd-sided polygon, the point equidistant from each of the vertices is vertically lower than the central point located at half of the polygon's height. In an even-sided polygon, this central point is positioned at exactly one-half of the polygon's height. For these reasons, we do not observe "gaps" with even-sided polygons; after sufficient iterations, the entire shape becomes filled with points.

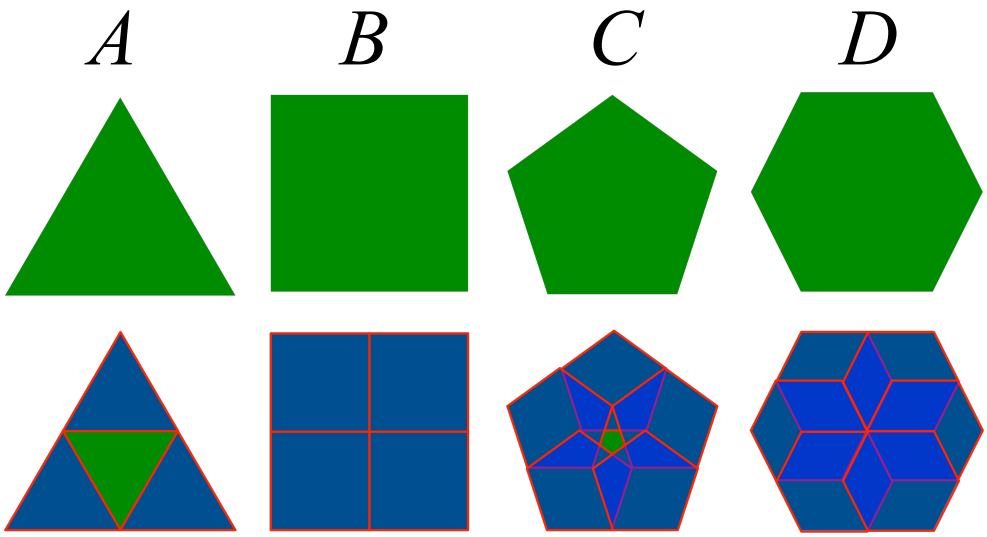


Figure 7. Odd-sided (A, C) and even-sided (B, D) polygons when they are mapped to polygons half their size, depending on which vertex is chosen in the chaos game. The point equidistant from each vertex in a given even-sided polygon is exactly its central halfway point, whereas the point equidistant from each vertex in an odd-sided polygon is lower than the halfway point. Hence, the odd-sided polygons display "gaps", while the even-sided polygons do not.

V. Fractional Dimension

The *dimension* of an object is, informally, the number of coordinates necessary to specify a point on the object. Thus, for example, a line is said to be 1-dimensional because only one variable is needed to specify position along a line. Similarly, a rectangle has dimension 2, and a cube has dimension 3. The dimension of some sets must be expressed as a fraction. In these instances, we say the sets have *fractional dimension.* Many fractals have fractional dimension. A fractal's dimension serves to describe its complexity and how its pattern changes with scale.

Two common types of dimension measures used are *Minkowski-Bouligand* (aka "box-counting") dimension and *Hausdorff* dimension. For many sets, the Minkowski-Bouligand and Hausdorff dimensions are equivalent. We will use the Minkowski-

Bouligand dimension in this paper, which is defined as follows for a set *S*. Given *x*, let *N*(*x*) be the number of boxes of side length *x* that it takes to cover *S*. Then, the Minkowski-Bouligand dimension of a set *S* is defined as

$$d(S) = \lim_{x \to 0} \frac{\ln(N(x))}{\ln(\frac{1}{x})}.$$

We can use this definition to verify the previously asserted dimension values for a line, a square, and a cube (Fig. 8). Further, we obtain the following when calculating the dimension of the Cantor set.

$$\begin{aligned} d(S) &= \lim_{x \to 0} \frac{\ln(N(x))}{\ln(\frac{1}{x})} \\ &= \lim_{n \to \infty} \frac{\ln(2^n)}{\ln(\frac{1}{\frac{1}{3^n}})} \\ &= \lim_{n \to \infty} \frac{\ln(2^n)}{\ln(3^n)} \\ &= \frac{\ln(2)}{\ln(3)} \\ &\approx 0.63 \end{aligned}$$

By the Minkowski-Bouligand definition, the Cantor set is less than 1-dimensional and has fractional dimension. We can similarly compute the dimension of the Sierpinski gasket.

$$\begin{aligned} d(S) &= \lim_{n \to \infty} \frac{\ln(N((\frac{1}{2})^n))}{\ln(\frac{1}{(\frac{1}{2})^n})} \\ &= \lim_{n \to \infty} \frac{\ln(3^n)}{\ln(2^n)} \\ &= \lim_{n \to \infty} \frac{n \cdot \ln(3)}{n \cdot \ln(2)} \\ &= \frac{\ln(3)}{\ln(2)} \\ &= 1.5849625 \end{aligned}$$

Thus, the Sierpinski gasket also has fractional dimension, but is of higher dimension relative to the Cantor set.

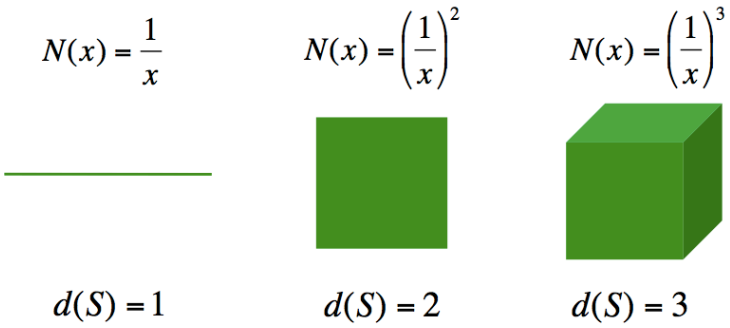


Figure 8. Calculation of the Minkowski-Bouligand dimension for a line, a square, and a cube.

VI. Chaos Game Representation (CGR)

Jeffrey proposed that the chaos game could be used to represent RNA or DNA sequences (1990). He used four vertices to represent the appropriate four nucleotides, and plotted points by reading through a variety of genetic sequences. One begins the game by starting in the center. As in the generation of the Sierpinski gasket (Section III), each ensuing point is halfway between the most recent preceding point and current the vertex being read. Thus, the square grid would display no prominent pattern if the gene sequence were completely random (refer to Fig. 7).

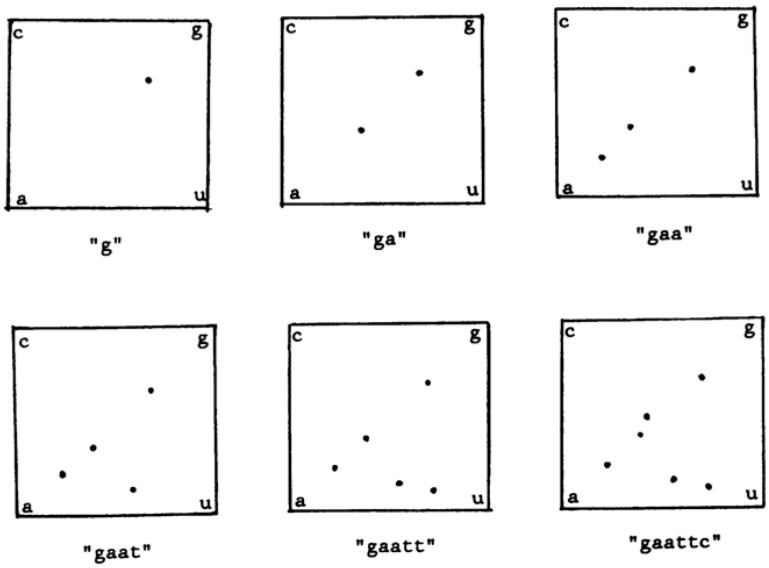


Figure 9. Chaos game representation of the first 6 bases of the Human Beta Globin Region on Chromosome 11, also known as HUMHBB (Jeffrey, 1990). Each vertex represents one nucleotide out of four: cytosine (c), guanine (g), adenine (a), and uracil (u). Biologically, HUMHBB is crucial for the creation of hemoglobin, which helps to carry oxygen from our lungs to the rest of our body.

Interestingly, a "double-scoop" pattern emerged in virtually all of the vertebrate sequences analyzed (Fig. 10). This pattern was not found in any group other than vertebrates aside from specific viruses (e.g., the HIV virus).

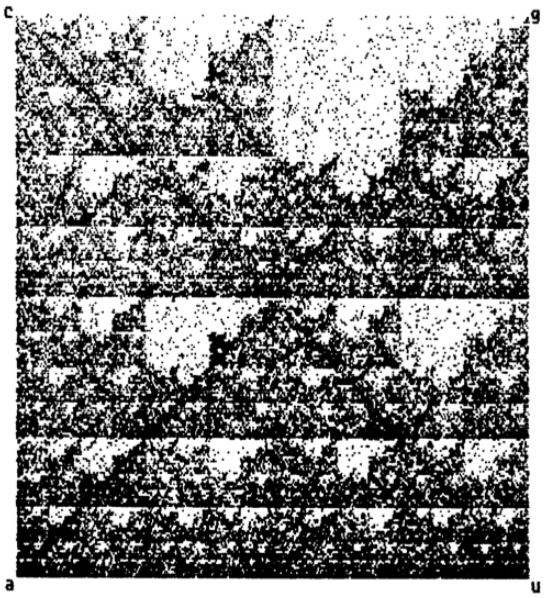


Figure 10. Jeffrey's completed CGR of HUMHBB after reading through 73,357 bases (1990). Note the repeating "double-scoop" pattern appearing throughout the figure, beginning in the upper right corner.

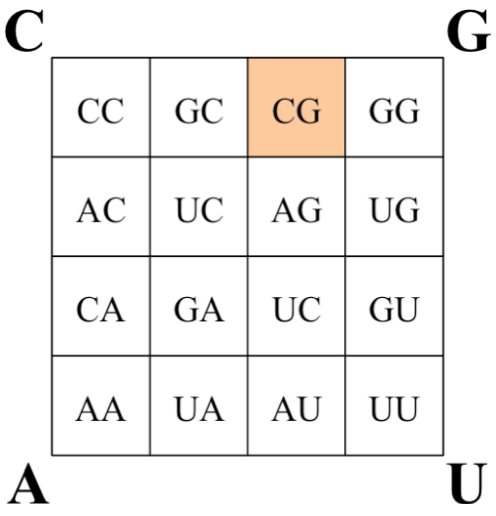


Figure 11. Subdivisions of the chaos game representation grid as presented by Jeffrey, with the "CG" subdivision highlighted (1990).

Consider dividing the CGR grid into four equivalent squares. Let the upper-left, upper-right, lower-left, and lower-right squares be termed the C, G, A and U quadrants of the grid respectively. Then, a point in the C quadrant represents a sequence ending with C, a point in the G quadrant represents a sequence ending with G, and so on. Let each of these quadrants be further divided into four squares, or "sub-quadrants", of equivalent size (Fig. 11). Then, for example, a point in the "CG" subquadrant is associated with a sequence ending in the dinucleotide CG. The grid can be further divided in this fashion, revealing subdivisions containing points that must be associated with sequences containing specific nucleotides.[*****]

Observe that a fractal with the double-scoop pattern is generated by removing all instances of guanine, "g", following cytosine, "c", in a sequence, but otherwise playing the game with randomly generated vertices. The set of "no CG" is obtained after infinitely many iterations (refer to Fig. 12).

If the entire grid is assumed to be a square with a side length of 4, the number of boxes (with a certain box side length) needed to cover this "no CG" set are as follows.

$$N(1) = 15$$

$$N(\tfrac{1}{2}) = 56$$

$$N(\tfrac{1}{4}) = 209$$

$$N(\tfrac{1}{8}) = 780$$

In counting the number of boxes of a certain side length *x* needed to cover a set, it is worth noting that only the approximation that includes missing squares of size *x* and

[*****] By this interpretation, the exact center of the square is associated with a null sequence.

larger is required. (Later approximations missing squares of side length less than x are unnecessary.)

Letting $x_n = N(1/2^n)$, the number of boxes to cover the set is described by the recurrence relation

$$N(\tfrac{1}{2^n}) = x_n = 4x_{n-1} - x_{n-2}.$$

However, we must be sure that this recurrence relation holds for all values of n, not just the values of n shown above. First, notice that because of self-similarity, the upper-left, lower-left, and lower-right quarter subdivisions of the grid are identical (refer to Fig. 12). Further, in any one of the aforementioned subdivisions, the upper-left, lower-left, and lower-right quarter subdivision is identical, and so on.

Next, refer to Fig. 12C for a visual example. Assume that when counting boxes of size $\frac{1}{2^{n-1}}$, the set with holes of sizes $\frac{1}{2^{n-1}}$ and larger, only, was used. If there were no gaps, then clearly $x_n = 4x_{n-1}$. However, box-shaped gaps of size $\frac{1}{2^n}$ exist. We must determine how many gaps there are and subtract this number from $4x_{n-1}$ to obtain the true value of x_n.

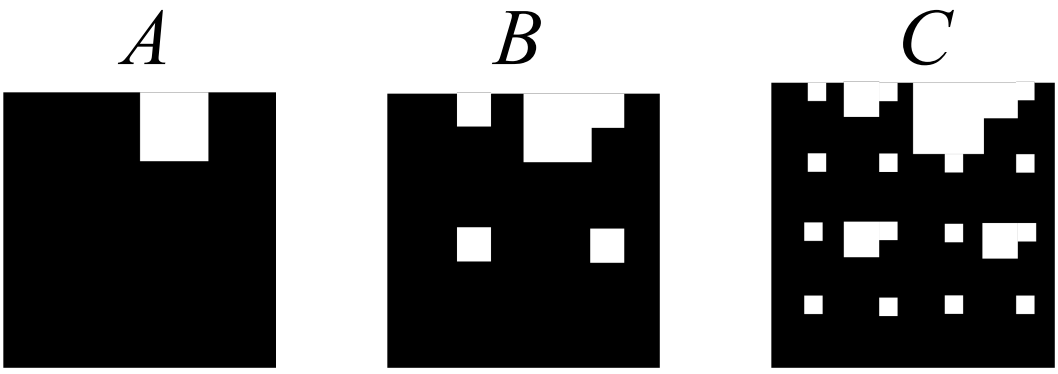


Figure 12. The first three iterations in the development of the set for no "CG".

Considering the smallest gaps in Fig. 12C to be of size $\frac{1}{2^n}$, the number of gaps of size $\frac{1}{2^n}$ *would* be one per black square of size $\frac{1}{2^{n-2}}$ if we did not retain the gaps formed in Fig. 12 A and B. However, there are exactly x_{n-2} not-all-white squares of size $\frac{1}{2^{n-2}}$. Because the number of gaps of size $\frac{1}{2^n}$ in Fig. 12C is equivalent to the number of black boxes of size $\frac{1}{2^{n-2}}$, in Fig. 12A, there are x_{n-2} gaps of size $\frac{1}{2^n}$. Thus, subtracting x_{n-2} from $4x_{n-1}$ will produce the true x_n.

By using this recurrence relation, we can obtain a closed-form solution.

$$N(\tfrac{1}{2^n}) = x_n = 4x_{n-1} - x_{n-2}$$

$$x_n = a(2-\sqrt{3})^n + b(2+\sqrt{3})^n$$

$$N(1) = a + b = 15 \Rightarrow a = 15 - b$$

$$N(\tfrac{1}{2}) = a(2-\sqrt{3})^1 + b(2+\sqrt{3})^1 = 56$$

$$b = \tfrac{15}{2} + \tfrac{13}{\sqrt{3}}$$

$$a = 15 - (\tfrac{15}{2} + \tfrac{13}{\sqrt{3}}) = \tfrac{15}{2} - \tfrac{13}{\sqrt{3}}$$

$$x_n = (\tfrac{15}{2} - \tfrac{13}{\sqrt{3}})(2-\sqrt{3})^n + (\tfrac{15}{2} + \tfrac{13}{\sqrt{3}})(2+\sqrt{3})^n$$

We can now calculate the dimension for the chaos game representation of a typical vertebrate gene sequence.

$$d(S) = \lim_{x\to 0} \frac{\ln(N(\frac{1}{2^n}))}{\ln(\frac{1}{\frac{1}{2^n}})}$$

$$= \lim_{n\to\infty} \frac{ln((\frac{15}{2} - \frac{13}{\sqrt{3}})(2-\sqrt{3})^n + (\frac{15}{2} + \frac{13}{\sqrt{3}})(2+\sqrt{3})^n)}{\ln(2^n)}$$

$$\approx 1.89997$$

VII. Verifying the Invariance of Fractional Dimension under Label Arrangement

Now that we possess a sufficient method to determine the fractional dimension of CGRs, an important question to answer is whether fractional dimension is invariant under the arrangement of the labeled vertices. We can resolve this inquiry by examining the fractional dimensions when other dinucleotides (namely, AG, UG, GG, AA, UU, CC, and GG) are eliminated but the sequence is otherwise random.

The "no UG" case is a trivial one, because a grid with the UG subdivision missing is identical to a grid with the CG subdivision missing, except rotated ninety degrees to the right and reflected around the central horizontal axis. The "no AG" case is less intuitive, but the same recurrence relation holds, and the proof arises much in the same way we derived the recurrence relation for no "CG". One may also confirm this equivalence between the two situations to some extent visually by noting that the same number of gaps exists in the no "AG" case as in the no "CG" case; they have simply shifted (Fig. 13).

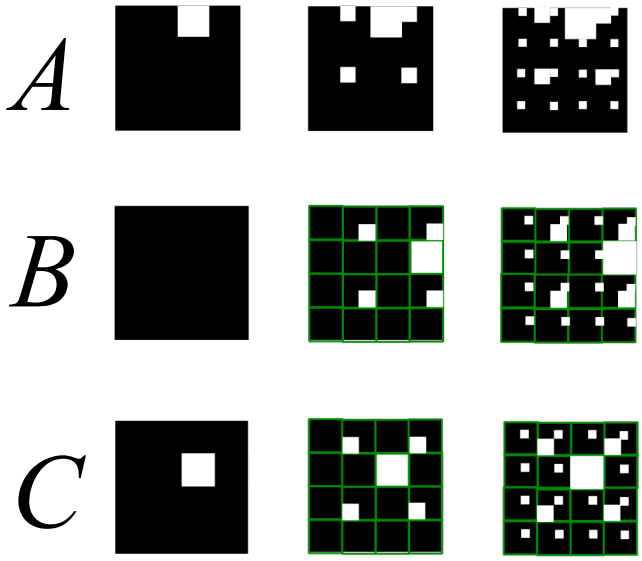


Figure 13. The first three iterations in the development of "no CG" (A), "no UG" (B), and "no AG", respectively. These situations share the same fractional dimension (≈ 1.89997) and can be represented by the same recurrence relation, $x_n = 4x_{n-1} - x_{n-2}$.

We now turn our attention to "no GG". If the entire grid is assumed to be a square with a side length of 4, the number of boxes (with a certain box side length) needed to cover this "no GG" set are

$$N(1) = 15$$

$$N(\tfrac{1}{2}) = 57$$

$$N(\tfrac{1}{4}) = 216$$

and so on. Letting $x_n = N(1/2^n)$, the number of boxes to cover the set can be described by the recurrence relation

$$N\left(\frac{1}{2^n}\right) = x_n = 3x_{n-1} + 3x_{n-2}.$$

To show this equation holds for all n, we must again assume that when counting boxes of size $\frac{1}{2^{n-1}}$, the set with holes of sizes $\frac{1}{2^{n-1}}$ and larger, only, was used (refer to Fig. 12C for a visual example). Define x_n to be the number of boxes of size $\frac{1}{2^n}$. If there were no gaps, then clearly $x_n = 4x_{n-1}$. However, box-shaped gaps of size $\frac{1}{2^n}$ exist. We must determine how many gaps there are and subtract this number from $4x_{n-1}$ to obtain the true value of x_n. In each not-all-white square of size $\frac{1}{2^{n-2}}$, gaps of size $\frac{1}{2^n}$ are located in all but three of the not-all-white squares of size $\frac{1}{2^{n-1}}$. The equation for the number of gaps then becomes

$$x_{n-1} - 3x_{n-2}$$

Subtracting this amount from $4x_{n-1}$ yields

$$\begin{aligned} x_n &= 4x_{n-1} - (x_{n-1} - 3x_{n-2}) \\ &= 3x_{n-1} + 3x_{n-2}. \end{aligned}$$

As before, we obtain a closed-form solution for the recurrence relation.

$$x_n = a\left[\left(\frac{1}{2}\right)(3-\sqrt{21})\right]^n + b\left[\left(\frac{1}{2}\right)(3+\sqrt{21})\right]^n$$

$$N(1) = a + b = 15 \Rightarrow a = 15 - b$$

$$N\left(\frac{1}{2}\right) = a\left[\left(\frac{1}{2}\right)(3-\sqrt{21})\right] + b\left[\left(\frac{1}{2}\right)(3+\sqrt{21})\right] = 57$$

$$b = \frac{15}{2} + \left(\frac{23\sqrt{\frac{3}{7}}}{2}\right)$$

$$a = 15 - \left(\frac{15}{2} + \left(\frac{23\sqrt{\frac{3}{7}}}{2} \right) \right)$$

$$x_n = \left(\frac{15}{2} - \frac{23\sqrt{\frac{3}{7}}}{2} \right) \left[\frac{1}{2}(3 - \sqrt{21}) \right]^n + \left(\frac{15}{2} + \frac{23\sqrt{\frac{3}{7}}}{2} \right) \left[\frac{1}{2}(3 + \sqrt{21}) \right]^n$$

We then use this solution to determine the dimension of dinucleotide repeats.

$$d(S) = \lim_{n \to \infty} \frac{\ln \left[\left(\frac{15}{2} - \frac{23\sqrt{\frac{3}{7}}}{2} \right) \left[\frac{1}{2}(3 - \sqrt{21}) \right]^n + \left(\frac{15}{2} + \frac{23\sqrt{\frac{3}{7}}}{2} \right) \left[\frac{1}{2}(3 + \sqrt{21}) \right]^n \right]}{\ln(2^n)}$$

$$\approx 1.92269$$

By inspection, we find that the remaining grids lacking dinucleotide repeats are identical to that of “no GG”. Visually, each of these grids is a “flipped” version of the “no GG” case (Fig. 14): “no AA” is a reflection about the diagonal between vertex *c* and *u*, “no CC” is a reflection about the line that divides the grid in half vertically, and “no UU” is a reflection about the line that divides the grid in half horizontally.

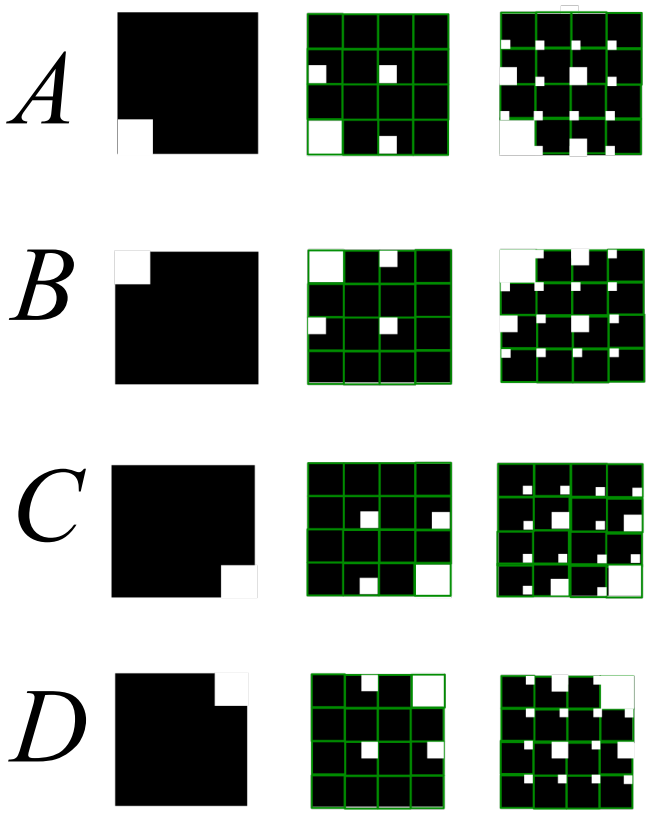


Figure 14. The first three iterations for dinucleotide repeats: "no AA" (A), "no CC" (B), "no UU" (C), and "no GG" (D). These situations share the same fractional dimension (≈ 1.92269) and are represented by the same recurrence relation, $N(1/2^n) = x_n = 3x_{n-1} + 3x_{n-2}$.

VIII. Conclusion

The "double-scoop" pattern described by Jeffrey (1990) can be approximated by a sequence that has no "CG" dinucleotides in the gene sequence but is otherwise randomly generated. It is likely that the consistency of this pattern among vertebrates is due to the selective disadvantage of CG dinucleotides, which are prone to methylation and subsequent mutation (Aswathi, 2009). However, now that we have reason to believe chaos game representation is a viable means of evaluating gene structure, many open questions remain. For instance, what happens to the CGR if we only examine the exons or the introns at one time? What additional methods could we use to quantify CGRs

mathematically, and what additional correlations can we make between patterns seen in CGRs and scientific taxonomy? Further, do other valid "types" of CGRs exist, and if so, what can they demonstrate about gene structure? Future research should seek to resolve these questions and more deeply explore the relationships between mathematics and biological science.

References

Aswathi, B. L. (2009). Chaos game representation. Centre for Bioinformatics, Dissertation.

Barnsley, M. F., & Demko, S. (1985). Iterated function systems and the global construction of fractals. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, Vol. 399, No. 1817, 243-275.

Bell, James. (No date). The Sierpinski Triangle: An Aesthestically Pleasing Limit Point. Dissertation.

Cull, P., Flahive, M. E., Robson R. O. (2005). Difference equations: from rabbits to chaos. New York, NY: Springer.

Dastgheib Z. A., Lithgow B., Moussavi Z. (2011). Application of fractal dimension on vestibular response signals for diagnosis of Parkinson's disease. *Conference Proceedings of the IEEE: Engineering in Medicine and Biology*, 7892-7895.

Devaney, R. L. (1990). Chaos, fractals, and dynamics: computer experiments in mathematics. Menlo Park, CA: Addison Wesley.

Dufner, J., Roser, A., and Unseld, F. (1998). *Fraktale und Julia-Mengen.* Verlag Harri Deutsch: Frankfurt.

Elaydi, S. (2005). An introduction to difference equations. New York: Springer.

Esteller, R., Vachtsevanos, G., Echauz, J., Henry, T., Pennell, P., Epstein, C., Bakay, R., Bowen, C., and Litt, B. (1999). Fractal dimension characterizes seizure onset in epileptic patients. *Conference Proceedings of the IEEE: Acoustics, Speech, and Signal Processing*, 4, 2343-2346.

Fiser, A., Tusnády, G. E., and Simon, I. (1994). Chaos game representation of protein structures. *Journal of Molecular Graphics and Modeling,* 12(4), 302-304.

Jeffrey, H. J. (1990). Chaos game representation of gene structure. *Nucleic Acids Research*, 18(8), 2163-2170.

Kuroda, S., and Tsuda, I. (2004). A complex systems approach to an interpretation of dynamic brain activity II: does Cantor coding provide a dynamic model for the formation of episodic memory? *Cortical Dynamics*, LNCS, 3146, 129-139.

Levy, S. D., and Pollack, J. V. (2001). Logical Computation on a Fractal Neural Substrate. *International Joint Conference on Neural Networks,* IEEE press.

Mandelbrot, B. (1967). How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science*, 156 (3775), 636–638.

Paramanathan, P., and Uthayakumar, R. (2007). Application of fractal theory in analysis of human electroencephalographic signals. *Computers in Biology and Medicine*, 38(3), 372-378.

Peitgen, H.-O., and Richter, P. H. (1986). The beauty of fractals: images of complex dynamical systems. Berlin: Springer-Verlag.

Pollack, J. B. (1991). The induction of dynamical recognizers. *Machine Learning*, 7, 227-252.

Prokofiev [Computer Programmer]. (2009). *Self-affine set* [Computer-generated image]. Retrieved April 30, 2012, from: http://upload.wikimedia.org/wikipedia/commons/0/0d/Self-affine_set.png

Tino, P. (1998). Spatial representation of symbolic sequences through iterated function systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 29, 386-393.

Werner, G. (2010). Fractals in the nervous system: conceptual implications for theoretical neuroscience. *Frontiers in Physiology*, 1(15), 1-28.