



# UNIVERSIDAD DE SONORA

"El Saber de Mis Hijos  
Hará Mi Grandeza"



El saber de mis hijos  
hará mi grandeza  
Acta: 64  
Foja: 67  
Libro: 2  
Expediente: 211202915

## ACTA DE EXAMEN PROFESIONAL

En la ciudad de Hermosillo, Sonora, siendo las 18:00 horas del día 18 de septiembre de 2015, se reunieron en el Departamento de Matemáticas de la Universidad de Sonora, los integrantes del jurado:

DR. FRANCISCO ARMANDO CARRILLO NAVARRO  
DRA. MARYSOL NAVARRO BURRUEL  
DR. DANIEL OLMOS LICEAGA

bajo la presidencia del primero y fungiendo como secretario el último, para realizar el examen profesional a:

**DIANA PATRICIA RIVERA SEGUNDO**

del programa: Licenciatura en Matemáticas

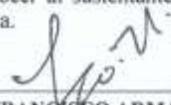
quien de acuerdo a la opción de titulación presentó tesis profesional titulada:

"FRACTALES EN LA BIOLOGÍA: EL CASO DEL DNA"

El jurado, después de debatir entre sí reservada y libremente, emitió el siguiente dictamen:

APROBADO POR UNANIMIDAD

El presidente del jurado dio a conocer al sustentante el resultado de su examen y para constancia se levantó la presente acta.

  
DR. FRANCISCO ARMANDO  
CARRILLO NAVARRO

Presidente

  
DR. DANIEL OLMOS LICEAGA

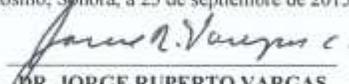
Secretario

  
DRA. MARYSOL NAVARRO  
BURRUEL

Vocal

DR. JORGE RUPERTO VARGAS CASTRO, jefe del Departamento de Matemáticas de la Universidad de Sonora, hace constar que las firmas que anteceden corresponden al jurado que intervino en el examen profesional.

Hermosillo, Sonora, a 25 de septiembre de 2015

  
DR. JORGE RUPERTO VARGAS  
CASTRO  
Jefe de Departamento



---

---

# UNIVERSIDAD DE SONORA

DIVISIÓN DE CIENCIAS EXACTAS Y NATURALES

Programa de Licenciatura en Matemáticas

Fractales en la biología: el caso del DNA

## T E S I S

Que para obtener el título de:

**Licenciado en Matemáticas**

Presenta:

Diana Patricia Rivera Segundo

Director de tesis: Dr. Pedro Miramontes

Hermosillo, Sonora, México,      14 de septiembre de 2015

## SINODALES

Dra. Marysol Navarro Burruel

Universidad de Sonora, Hermosillo, México

Mat. Arturo Fragozo Robles

Universidad de Sonora, Hermosillo, México

Dr. Daniel Olmos Liceaga

Universidad de Sonora, Hermosillo, México

Dr. Francisco Armando Carrillo Navarro

Universidad de Sonora, Hermosillo, México

*A mis abuelos*

*Eulogio y Balvaneda*

# *Agradecimientos*

Escribir los agradecimientos es lo que generalmente se hace hasta el final, y sin romper la pauta, espero que esto sea lo último que escriba sobre mi tesis de licenciatura y probablemente, lo único que algunos lean.

Esta tesis no hubiera sido posible sin la ayuda de mi director, Pedro Miramontes. En especial, quisiera agradecer por sus cursos y pláticas que me ayudaron a introducirme en el área de la biología matemática.

Gracias al Departamento de Matemáticas de la Universidad de Sonora, por su apoyo académico y económico durante el desarrollo de mis estudios. También quisiera agradecer a los profesores Martín Gildardo García Alvarado, Marysol Navarro Burruel, Jesús Francisco Espinoza Fierro y Ricardo Sáenz Casas, cuyas conversaciones y clases incrementaron mi gusto por la matemática. Al profesor Armando Carrillo, por la revisión de esta tesis.

Al profesor Arturo Fragozo Robles por siempre apoyarme y animarme a realizar mis proyectos.

Al profesor Daniel Olmos Liceaga, por guiarme en mis estudios de licenciatura, por su apoyo y confianza.

Gracias a mi familia; a mis abuelos, por ser siempre mi ejemplo a seguir; a mis padres, porque gracias a ellos toda mi vida es posible, literalmente; a mis tíos Israel y Luz del Carmen, quienes siempre estuvieron al pendiente de mis estudios; a mis tías Mirna Payán y Leticia Martínez.

También quisiera agradecer a mis amigos, Selene Alaíde López, Griselda Adilene Sánchez Cubedo, Eduardo Acuña Yeomans, Emanuel Gutiérrez Lira, Gabriel Adrian Salcedo Varela, José Alberto Jiménez Higuera, Luis Alberto Trujillo Ortega, Juan Diego Valdez Montes, Nohemy Palafox Lacarra. A lo largo de estos cuatro años me ayudaron a evitar cualquier indicio de soledad y tristeza; por las discusiones sobre este trabajo, de sus trabajos, de las matemáticas y cualquier otro tema en general.

Y por último: Gracias, Daniel.

# Índice general

Agradecimientos	II
Índice general	III
Índice de figuras	v
<b>1. Introducción</b>	<b>1</b>
<b>2. Fractales en la biología</b>	<b>3</b>
2.1. La geometría fractal de los seres vivos . . . . .	4
2.2. Series de tiempo . . . . .	9
2.3. Sistemas complejos . . . . .	13
<b>3. Fenomenología del DNA</b>	<b>15</b>
3.1. Estructura . . . . .	15
3.2. Estabilidad . . . . .	20
3.3. Replicación, transcripción y traducción . . . . .	23
3.4. Genes y genomas . . . . .	27
<b>4. Fractales</b>	<b>30</b>
4.1. Concepto . . . . .	31
4.2. Espacios métricos . . . . .	33
4.3. Dimensión Fractal . . . . .	43
4.4. Autosimilaridad . . . . .	48
4.5. Ejemplos . . . . .	55
<b>5. DNA y Sierpínski.</b>	<b>61</b>
5.1. Funciones lógicas . . . . .	61
5.2. Función XOR . . . . .	63
5.3. El Tetraedro de Sierpínski . . . . .	70
<b>6. Conclusiones</b>	<b>73</b>



# Índice de figuras

2.1. Variación de la longitud de la costa de Gran Bretaña según la escala de medición. [24]	4
2.2. Brocoli romanesco.	5
2.3. Árbol de Pitágoras.	5
2.4. Representación de los pulmones humanos.	7
2.5. Patrones formados por la bacteria <i>Paenibacillus vortex</i> bajo diferentes condiciones de crecimiento.	8
2.6. Diagrama del corazón.	11
2.7. Complejo QRS.	11
2.8. Intervalos cardiacos (RR) de una persona sana (derecha) y una persona con fallas en el corazón (izquierda).	13
3.1. Fotografía 51. Imagen obtenida por Rosalind Franklin que sirvió como fundamento para el modelo de la doble hélice del DNA de Watson y Crick.	16
3.2. Estructura de un nucleótido.	16
3.3. A la izquierda, la estructura de una hebra de DNA. A la derecha, puentes de hidrógeno entre los pares de bases según el modelo de Watson-Crick.	17
3.4. Puentes de hidrógeno entre los pares de bases según el modelo de Watson-Crick.	18
3.5. Estructura de las bases nitrogenadas.	19
3.6. Diferentes formas de la estructura secundaria del DNA. En la parte superior	21
3.7. Los valores que toma $\Delta H$ indican la estabilidad energética dada en kcal y los de $\Delta G$ indican la estabilidad relativa dados en kcal/mol.	22
3.8. La doble hélice del DNA puede ser desnaturalizada por calor, en medios muy básicos o muy ácidos.	23
3.9. Replicación del DNA.	25
3.10. Proceso de edición bajo el cual se obtiene el mRNA maduro	26
3.11. Tamaño del genoma de diferentes clases de organismos vivos.	27

3.12. Código genético. El primer nucleótido del triplete está indicado por la columna de la izquierda, el segundo por las columnas centrales y el tercero por la columna derecha. Los aminoácidos son codificados por el mRNA, donde las siglas indicadas en la tabla significan lo siguiente, ALA: Alanina, CYS: Cisteína, ASP: Ácido aspártico, GLU: Ácido glutámico, PHE: Felinalanina, GLY: Glicina, HIS: Histidina, ILE: Isoleucina, LYS: Lisina, LEU: Leucina, MET: Metionina, ASN: Asparagina, PRO: Prolina, GLN: Glutamina, ARG: Arginina, SER: Serina, TMR: Treonina, VAL: Valina, TPR: Triptofano, TYR: Tirosina, y STOP denota a los aminoácidos Pardo, Ámbar y Ocre que no codifican pero que sirven como signos de puntuación para finalizar el mensaje [3]. . . . .	29
4.1. Primera imagen del conjunto de Mandelbrot . . . . .	31
4.2. Curva de Takagi. . . . .	32
4.3. Conjunto de Cantor . . . . .	33
4.4. Gráfica de la función $\mathcal{H}^s(X)$ . . . . .	47
4.5. Árbol de Hata. . . . .	56
4.6. Procesión de los cangrejos. . . . .	57
4.7. Curva de Lévy. . . . .	58
4.8. Curva del dragón. . . . .	59
4.9. Triángulo de Sierpinski. . . . .	60
5.1. Tabla de verdad de NOT . . . . .	62
5.2. Tabla de verdad de AND . . . . .	63
5.3. Tabla de verdad de OR . . . . .	63
5.4. Tabla de verdad de XOR . . . . .	64
5.5. Patrición del cubo unitario en octantes. . . . .	66
5.6. A la izquierda, octantes permitidos al realizar la primera partición, a la derecha, zonas permitidas después de la segunda partición. . . . .	67
5.7. En la parte superior se muestra la regla 90 del autómata celular elemental, en la parte inferior el Triángulo de Sierpinski generado por la misma [23]. . . . .	67
5.8. Tetraedro obtenido a partir del gen <i>yoaD</i> de la <i>E. coli</i> . . . . .	68
5.9. Tetraedro generado a partir de una distribución binomial. . . . .	68
5.10. A la izquierda proyección de los datos de la E. Coli, a la derecha proyección de los datos aleatorios. . . . .	69
5.11. A la izquierda, proyección formada con un genoma humano. A la derecha, formada con el genoma de una rata. . . . .	69
5.12. A la izquierda vemos la proyección del genoma de la rickettsia, a la derecha, la proyección de datos aleatorios. . . . .	69
5.13. Tetraedro de Sierpinski. . . . .	71
5.14. Tetraedro formador por cubos. . . . .	72
5.15. Imagen en $\mathbb{R}^3$ de: (a) la función AND, (b) la función XOR, (c) la función OR. . . . .	72



# Capítulo 1

## Introducción

El que hacer de la ciencia siempre está inspirado en el mundo que le rodea. Así existen una gran cantidad de disciplinas que se enfocan al estudio de distintos aspectos del mismo, sin embargo, esto no quiere decir que estas disciplinas sean ajenas unas a otras, al contrario, se necesitan para lograr avances y desarrollarse. Hablemos del caso de la biología y la matemática, de manera que en esta primera sección es pertinente introducir el terreno donde nos manejaremos a lo largo de este trabajo de tesis: la biología matemática.

*Ian Stewart* en su libro *The mathematics of life* distingue cinco revoluciones en la biología que han cambiado la manera de pensar de los científicos sobre la vida y agrega una sexta que está en curso [17].

La primera de las cinco se da con la invención del microscopio por *Anton van Leeuwenhoek* hace más de trescientos años, la segunda inició con la clasificación sistemática de los seres vivos por *Carlos Linneo* en 1735, la tercera comienza con la tan controversial teoría de la evolución de *Charles Darwin* publicada en su libro *El origen de las especies* en 1859, la cuarta revolución se da a partir del descubrimiento de los genes por *Gregory Mendel* en 1865, y por último tenemos el descubrimiento de la estructura del DNA por *Francis Crick* y *James Watson* en 1953.

Adicionalmente, Stewart propone que la revolución en curso es referente a como la matemática se aplica a los procesos biológicos.

La matemática no sólo es usada para ayudar a manipular y manejar la información sobre la biología, si no que también provee de ciertos significados y ayuda a explicar como

funciona la vida.

Adicionalmente, podemos mencionar que la invención de la computadora y el surgimiento de las ciencias de la computación nos ayudan a manejar grandes cantidades de datos y hacer cálculos que humanamente no nos ha sido posible. Ramas de la biología y la matemática han sido creadas a partir de su existencia, por ejemplo, en el caso de la biología la bioinformática, y en la matemática, por mencionar algunas, las áreas referentes a los fractales y sistemas complejos.

En el capítulo 1 daremos una introducción de los fractales en la biología, con la intención de explicar los motivos que se tienen para aplicar esta área de las matemáticas a aspectos biológicos, comenzando por describir las formas de los seres vivos y finalizando con los sistemas complejos. En el capítulo 2 hablaremos de aspectos generales del DNA, con la finalidad de introducir al lector al estudio de este. Posteriormente en el capítulo 3 se formalizará matemáticamente el estudio de los fractales, en específico, los fractales auto-similares. Finalmente en el capítulo 4 se mostrará un algoritmo que nos permite hacer una conexión entre las dos áreas.

# Capítulo 2

## Fractales en la biología

En este capítulo tiene la intención de introducir la presencia de los fractales en la biología.

A lo largo de esta sección veremos ejemplos sobre como la geometría fractal esta presente en la morfología de los seres vivos, para posteriormente extender esta noción geométrica a una temporal y considerar fractales en el tiempo, que nos ayudan a analizar comportamientos irregulares en los sistemas biológicos. Después trataremos brevemente el tema de sistemas complejos, los cuales podemos identificar por una serie de propiedades que poseen, entre ellas una involucra a nuestro objeto de estudio. Una vez hayamos visto esto, nos resultará claro que el estudio que haremos del DNA solo es un caso particular.

Al comenzar a hablar de un tema siempre existen referencias obligadas, los ejemplos más comunes de fractales en la naturaleza son la costa de Gran Bretaña, donde la escala con la que se mida la costa dicta su longitud (Figura 2.1); el movimiento browniano, el cual se estudia el movimiento de las partículas; el brócoli romanesco, cuya forma es autosemejante, entre otros.

Pretendemos poco a poco nos irnos haciendo de una idea intuitiva de lo que es un fractal, que capítulos más adelante formalizaremos. Así mismo, mencionaremos conceptos relacionados al tema que no serán tratados a profundidad sino hasta el capítulo siguiente.

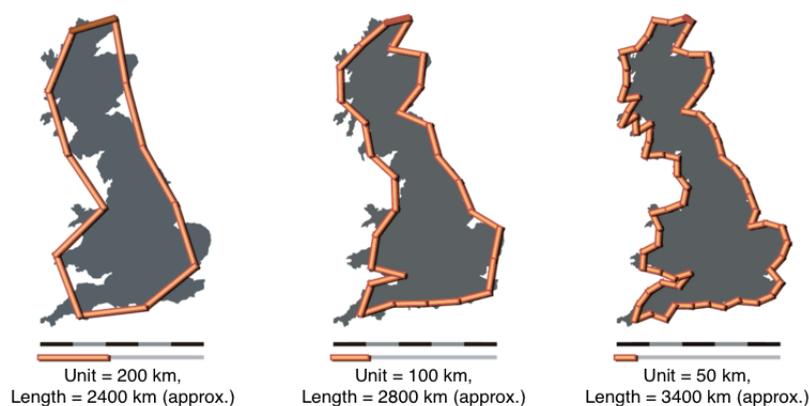


FIGURA 2.1: Variación de la longitud de la costa de Gran Bretaña según la escala de medición. [24]

## 2.1. La geometría fractal de los seres vivos

Las formas en que los seres vivos se presentan son diversas de algún modo, y el querer asociar estas formas a las que se conocen de la geometría clásica puede resultar artificial y tal vez demasiado simple en ciertos aspectos. Entonces nos cuestionamos si la geometría fractal es también otro artificio con el que forzamos a la matemática a describirlas. Recordemos que los modelos matemáticos son representaciones imperfectas de la realidad, y que el mejor modelo de la realidad que se conoce, es la realidad misma.

A continuación hablaremos de organismos en los que podemos identificar formas fractales de manera visual.

### Brócoli Romanesco

El Brócoli Romanesco es conocido por ser el fractal más común entre los seres vivos. En su estructura observamos racimos terminales que son semejantes a todo el brócoli en sí, es decir, por todo el brócoli podemos encontrar miniaturas de sí mismo. Esto nos indica que existe una autosemejanza en él.

El brocoli romanesco pertenece a la familia *Brassica oleracea*, del cual podemos observar fácilmente su geometría fractal (Figura 2.2). Podríamos decir que esta característica viene de familia, ya que a las plantas pertenecientes a esta también se les atribuyen geometría

fractal, ya que los cortes transversales de estas plantas se han sido asociados al Árbol de Pitágoras (Figura 2.3).

Una idea importante que surge de pensar en estas formas en términos matemáticos, es que nos sugiere una forma de cuantificar su forma utilizando su autosemejanza, como veremos más adelante.

Hasta este punto solamente hemos visto que los fractales nos ayudan a describir la apariencia de una planta, sin embargo, no hemos mencionado como estos se pueden relacionar con otros aspectos de los seres vivos.



FIGURA 2.2: Brocoli romanesco.

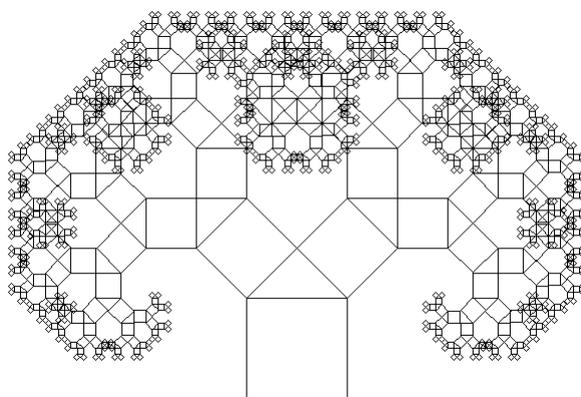


FIGURA 2.3: Árbol de Pitágoras.

## Pulmones

Veremos ahora el caso de los pulmones. Un aspecto destacable de las vías respiratorias es su geometría. Pensemos que en ellas se debe distribuir cierto volumen de aire inspirado en una superficie de intercambio, que a su vez se encuentra en un volumen acotado, con lo que podemos deducir que el diseño de un órgano no debería depender sólo de la forma, sino que también de su funcionamiento.

En los pulmones de un adulto aproximadamente existen 300 millones *alvéolos*, que son la parte final del árbol respiratorio y actúan como unidad primaria en el intercambio gaseoso, con un diámetro de  $187 \mu\text{m}$  en promedio. Para darnos una idea del tamaño que deben tener considerando estas características, se calcula que el área de los pulmones es aproximadamente de  $160 \text{ m}^2$ , lo mismo que mide una cancha de voleibol.

Para darnos una idea de la complejidad de su forma, las vías respiratorias en el pulmón bifurcan dicotómicamente (Figura 2.4) a través de 23 generaciones, las primeras 14 son los conductos bronquiales.

Para tratar a los pulmones matemáticamente, supondremos que el conjunto de todas las puntas de las ramas forman un conjunto *autosimilar*, al cual podremos medir su *dimensión fractal*  $d$ .<sup>1</sup>

Para darnos una idea de lo que queremos decir con un conjunto autosimilar en este caso, consideremos que el crecimiento del pulmón prenatal comienza con un brote, que evoluciona hasta formar un tubo, el cual forma a su vez dos brotes, y en cada uno de ellos se repite el proceso anterior, un proceso de crecimiento que es *autosemejante*. Notemos que no estamos tomando en cuenta una posición que nos describa qué es lo óptimo, sino que estamos pensando en términos de simplicidad: el programa regulador del crecimiento es más corto si cada paso es una repetición del anterior a una escala menor. De esta manera se nos sugiere el resultado del crecimiento queda completamente determinado por la razón grosor/longitud de las ramas y por un parámetro relacionado con el diámetro de las mismas y de alguna manera con su dimensión fractal. Llamaremos a este parámetro  $\delta$  [8].

---

<sup>1</sup> Asociamos el concepto de dimensión a números enteros, sin embargo, esta asociación no nos permite distinguir entre ciertos objetos. Por ejemplo, sabemos que la dimensión de una línea recta tanto como de una línea costera es 1. La dimensión fractal es un número real que no necesariamente es entero, lo cual nos brinda más información sobre nuestro objeto.

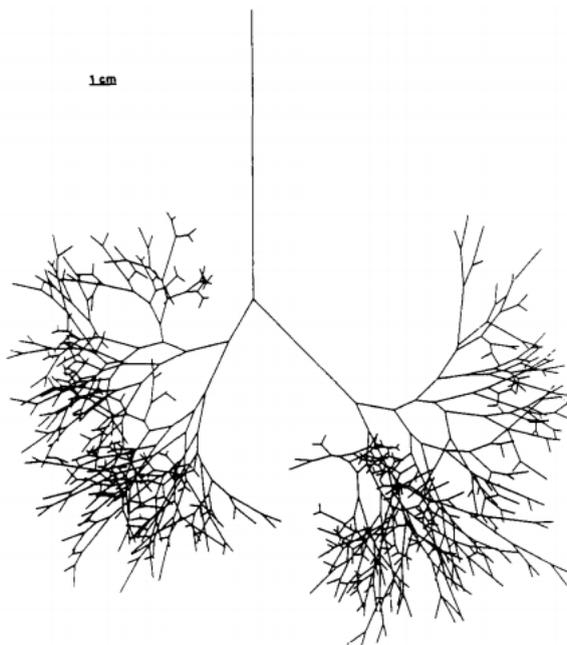


FIGURA 2.4: Representación de los pulmones humanos.

Ahora bien, dependiendo del valor que tome  $\delta$ , el crecimiento según estas reglas llega a uno de los tres resultados siguientes:

1. Después de un número finito de estos pasos, las ramas agotan el espacio disponible para el crecimiento.
2. Las ramas nunca llegan a llenar más que una parte del espacio accesible.
3. Resulta que el espacio disponible es exactamente el que necesitan.

Cuando  $\delta = 3$ , corresponde al límite de tubos infinitamente finos y sería el caso ideal, es decir, que está llenando todo el espacio disponible. Sin embargo sabemos que esto no es posible. Pero los tubos no son infinitamente pequeños, y lo más que se obtiene es que los valores de  $\delta$  sean ligeramente inferiores a 3, lo que concuerda con la evidencia empírica, pero esto se obtiene como una simple consecuencia adicional de la propiedad de autosemejanza.

## Bacterias

Las bacterias son organismos unicelulares, sin embargo no es común que existan como organismos aislados, sino que forman colonias en las que crecen y se dividen. Para sobrevivir, la colonia de bacterias debe extenderse en diferentes direcciones con el fin de llegar hasta donde se pueden obtener los nutrientes necesarios. Este es un comportamiento cooperativo que da lugar a que se presenten patrones complejos en la forma de las colonias. (Figura 2.5)

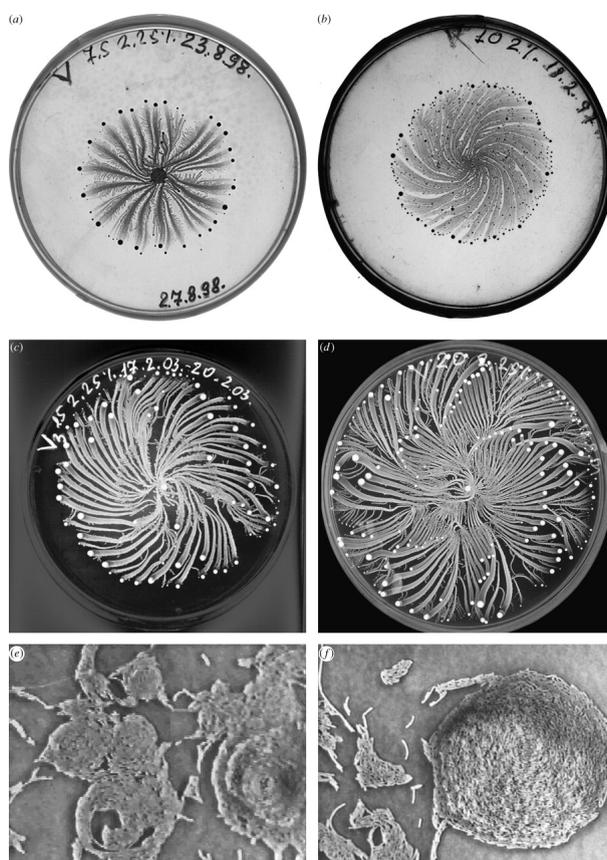


FIGURA 2.5: Patrones formados por la bacteria *Paenibacillus vortex* bajo diferentes condiciones de crecimiento.

Hay diferentes manera de explicar razones por las cuales las bacterias forman estos patrones durante su desarrollo y crecimiento. Por ejemplo, que esto se deba a moléculas en movimiento entre las bacterias, de manera que fuerzas adhesivas que las mantienen juntas, o que al dividirse se encuentran unas contra otras.

La familia de bacterias *Paenibacillus* mostrada en la imagen anterior presenta los patrones a los que nos refererimos. Los miembros de familia tienen la habilidad de formar

complejas estructuras multicelulares conforme crecen en placas de agar. Los patrones pueden ser ramificados o en forma de arbusto, arremolinados o quirales.

Para el caso de estas bacterias, la formación de estos patrones está asociada al fenómeno de mortalidad en enjambre, que es el comportamiento en cooperativo que mencionamos anteriormente. Esta conducta colectiva presentada en las *Paenibacillus* tiene aplicaciones. Se realizan simulaciones en computadora basadas en algunos principios derivados del estudio de esta familia. Estas simulaciones muestran interacciones adaptables entre un grupo de objetos en movimiento, cuya regla es dirigirse en dirección de mayor beneficio, lo cual ha dado un desempeño superior comparado con los modelos basados solamente en interacciones estáticas de repulsión y orientación.

## 2.2. Series de tiempo

Como vimos anteriormente, con un enfoque fractal podemos describir formas geométricas sin escala de longitud predominante. Esto puede ser extendido para analizar *series de tiempo* con fluctuaciones cuyas componentes, similarmente, no se manifiestan únicamente en una escala de tiempo específica.

Una serie de tiempo es una sucesión de observaciones medidas en determinados momentos, ordenados cronológicamente e igualmente espaciados.

Diremos que una serie de tiempo es fractal si presenta irregularidades a través de diferentes escalas temporales y resultan invariantes estadísticamente con respecto a la escala de temporal con la que se esté analizando.

Las series de tiempo fractales son originadas por procesos dinámicos que generalmente involucran interacciones no lineales. Lo que esperamos obtener a partir del análisis fractal es información acerca de la dinámica que generó a la misma y patrones similares entre este tipo de fenómenos.

Un ejemplo de estos patrones que exhiben las series tiempo fractales, es una característica o propiedad  $q$  que describe una ley de potencias estadística con respecto a la escala  $s$  que

se cuantifique el tiempo, es decir,

$$q = ps^\varepsilon$$

donde  $p$  es un factor de proporcionalidad y  $\varepsilon$  un parámetro conocido como el exponente fractal o de escala que puede ser relacionado con la dimensión fractal, de la cual se hablará capítulos más adelante.

Como ejemplos de series de tiempo podemos mencionar los registros de la temperatura hora a hora de Hermosillo, el número de nacimientos cada año de un país, la cantidad de manchas solares por año, el índice de contaminación registrado cada día en la Ciudad de México, por mencionar solo unos ejemplos. Desarrollaremos el caso de las series de tiempo relacionadas con la actividad cardiaca.

En los seres vivos, las funciones y estructuras celulares se mantienen y evolucionan debido a desequilibrios permanentes entre las células y su entorno. De esa manera, la actividad cardiovascular es regulada por múltiples sistemas que mantienen en niveles adecuados el flujo sanguíneo y las presiones capilares. Mediante el mantenimiento de la presión hidrostática en estos se promueven los desequilibrios con el líquido intersticial necesarios para conservar la actividad celular. De esta manera, el control fisiológico a corto plazo de la función cardiovascular, que abarca escalas temporales desde segundos a minutos, se dirige principalmente a la regulación de la presión arterial media.

Existen diversos factores que influyen en la actividad cardiaca, entre ellos se encuentra el control nervioso ocasionado por la descarga permanente de neurotransmisores. La importancia de estos consiste en que ejercen una acción especializada y modifican la frecuencia cardiaca.

A raíz de estas influencias, las perturbaciones externas, y en menor medida otros factores como descargas hormonales o la organización del tejido cardiaco, la frecuencia cardiaca y otras variables cardiovasculares exhiben fluctuaciones en diferentes escalas temporales.

En el caso del corazón, el análisis de fluctuaciones en intervalos cardiacos se realiza en mediciones de lo que conocemos como variabilidad en la frecuencia cardiaca (VFC).

La medición de la VFC típicamente se realiza a partir de señales electrocardiográficas, que se forman a partir de la actividad eléctrica de la aurícula y los ventrículos.

Explicaremos brevemente como se genera la actividad eléctrica. El nodo senoauricular (Figura 2.4) se encuentra en la aurícula derecha, cuyas células poseen un potencial transmembrana que decrece espontáneamente hasta alcanzar el potencial umbral, produciéndose la auto excitación. La propagación de la excitación a través de los ventrículos da origen al *complejo QRS* que equivale a la contracción ventricular.

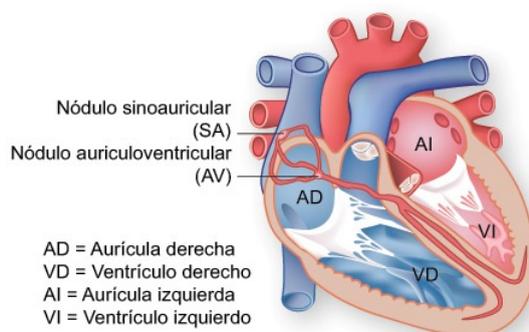


FIGURA 2.6: Diagrama del corazón.

El complejo QRS es un conjunto de ondas que representan la despolarización de los ventrículos. La onda Q es la primera onda negativa del complejo, la onda R es la primera onda positiva del complejo, mientras que la onda S es la segunda onda negativa del complejo.

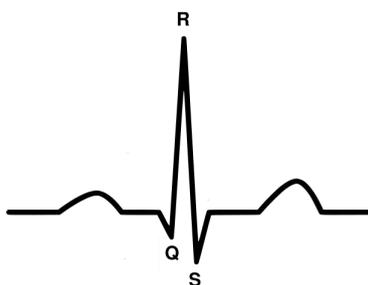


FIGURA 2.7: Complejo QRS.

Ahora que tenemos una manera de medir un fenómeno cardiaco, queremos estudiar las mediciones de alguna manera, para ello introduciremos una herramienta basada en el análisis fractal. Uno de los métodos fractales más utilizado para explorar series de tiempo es el *análisis de fluctuaciones sin tendencia* (DFA por sus siglas en inglés) que fue

introducido por Peng *et. al* en 1994, particularmente para el análisis de las fluctuaciones en la frecuencia cardiaca, siendo una modificación del análisis de la raíz cuadrática media para datos no estacionarios.

Esencialmente, en el análisis de fluctuaciones sin tendencia la serie de datos originales es integrada de acuerdo a la siguiente ecuación:

$$Y(k) = \sum_{i=1}^k [RR(i) - \bar{RR}]$$

donde  $Y(k)$  representa el  $k$ -ésimo valor del resultado de la integración donde  $k = \{1, \dots, L\}$ ,  $RR(i)$  es el  $i$ -ésimo valor o intervalo de latido y  $\bar{RR}$  es el valor medio de toda la serie original con  $L$  intervalos o muestras en total. Posteriormente la serie integrada es dividida en segmentos independientes del mismo tamaño que  $n$  que comprenden a un número de intervalos cardiacos, los cuales abarcan una escala de tiempo particular. La tendencia local en estos segmentos  $Y_n$  es estimada por un ajuste polinomial con el propósito de restarla localmente para reducir o eliminar las transiciones no estacionarias. Así, la fluctuación cuadrática media  $F$  en función de  $n$  puede ser calculada de la siguiente manera:

$$F(n) = \sqrt{\frac{1}{L} \sum_{k=1}^L [Y(k) - Y_n(k)]^2}$$

Este procedimiento se repite para un conjunto amplio de valores de  $n$ , obteniéndose una relación que se aproxima por el siguiente modelo o ley de potencia:

$$F(n) \approx n^\alpha$$

donde  $\alpha$  se estima por la pendiente obtenida a partir de una representación logarítmica entre las variables  $F(n)$  y  $n$ . En el caso de las fluctuaciones cardiacas, el exponente  $\alpha$  se estima para un intervalo de escalas  $n$  que abarca un corto plazo (menor a 11 intervalos cardiacos consecutivos) p, por el contrario un largo plazo. Esto a raíz de que el compartamiento a largo plazo de las escalas no es estrictamente uniforme sino que, básicamente, refleja dos dinámicas fractales principales asociadas a los diferentes mecanismos o factores de control cardiaco, que operan o influyen ya sea en un corto o largo plazo.

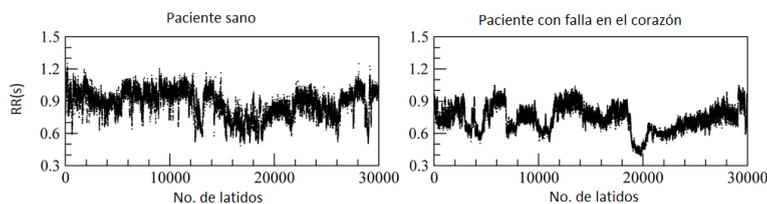


FIGURA 2.8: Intervalos cardiacos (RR) de una persona sana (derecha) y una persona con fallas en el corazón (izquierda).

## 2.3. Sistemas complejos

Esta última sección del capítulo la dedicaremos a hablar sobre los *sistemas complejos*. Existen dos enfoques clásicos en la ciencia con los que se tratan los fenómenos de la naturaleza: el determinista y el probabilista.

El enfoque determinista nos dice que, dada las condiciones iniciales de un problema, siempre nos será posible predecir el futuro de este. Por ejemplo, nos es posible calcular la posición de la Tierra en su órbita alrededor del Sol en determinada fecha, la trayectoria de un satélite alrededor de la Tierra, etc. Desde esta perspectiva, estos fenómenos se suelen modelar mediante ecuaciones diferenciales que hay que resolver para obtener la información que deseamos, sin embargo, existen casos que para obtener dicha información, la cantidad de ecuaciones es demasiado grande y no nos es posible hacerlo. Entonces llegamos al límite de este enfoque.

Por otro lado, el enfoque probabilista considera que el azar está presente en la naturaleza y debido a esto podremos estimar el posible futuro de los fenómenos. Como ejemplos de estos tenemos la cantidad de nacimientos y muertes en un año, el lanzamiento de una moneda, la posición de una partícula suspendida en un fluido, etc. Este enfoque es el utilizado por la mecánica estadística, donde la cantidad de componentes a tratar es de orden  $2^3$ , con lo cual no nos atrevemos siquiera a tratarlo mediante un sistema de ecuaciones. La información que obtenemos resulta un promedio de lo que ocurre en todas las componentes, sin embargo, para emplear los métodos de la mecánica estadística se requiere que la cantidad de objetos a tratar sea de dichos órdenes de magnitud.

Entonces, nos podemos cuestionar acerca de los fenómenos que quedan en medio, es decir, que tienen demasiadas componentes para ser tratado mediante ecuaciones deterministas, pero muy pocas para tratarlos mediante el método de la mecánica estadística. Los

sistemas complejos están formados por un conjunto grande de componentes individuales que interactúan entre sí y que pueden modificar sus estados internos como resultado de estas interacciones. La cantidad de estas componentes queda en el abismo que existe entre los dos enfoques.

En los sistemas complejos surgen espontáneamente propiedades a partir de las interacciones entre los elementos que la constituyen. A estas propiedades se les conoce como *propiedades emergentes*. También estos sistemas pueden situarse en regímenes críticos caracterizados por la presencia de fluctuaciones espaciales y temporales en todas las escalas posibles, lo que muestra que los fractales están presentes en los sistemas complejos.

En la naturaleza podemos encontrar casos de sistemas complejos. Por ejemplo, recientemente Per Bak y Kim Sneppen plantearon que las espontáneas extinciones catastróficas pueden estar relacionadas con una distribución conforme a leyes de potencias, mismas que se observan en modelos de *criticalidad autoorganizada*.

El concepto de criticalidad autoorganizada fue introducido por Bak *et al.* para explicar el patrón de comportamiento de diversos sistemas o fenómenos complejos existentes en la naturaleza. Estos sistemas complejos evolucionan hacia estados críticos, donde hay alteraciones que se producen sin la intervención de ningún agente externo, que desencadena fenómenos de cualquier tamaño. Estos, se dan a todas las escalas a las que se observa el sistema, es decir, son invariantes a la escala.

Estos conceptos que han surgido en estas últimas décadas han sido importantes aportaciones para entender los fenómenos de otra manera. Por ejemplo, la formulación de teorías y leyes en biología ha sido diferente que en la física, donde estas son expresadas mediante ecuaciones que relacionan cantidades que son físicamente observables[1].

# Capítulo 3

## Fenomenología del DNA

Todos los organismos del planeta, tanto si tienen o no un núcleo definido, poseen cromosomas, en los que se almacena el material genético transmitido de generación en generación. Las moléculas transmisoras de los caracteres hereditarios son, en todos los organismos, los ácidos nucleicos. Salvo en contadas excepciones, por lo demás secundarias, el ácido nucleico depositario de la información hereditaria es la molécula llamada DNA.

La importancia y significado del DNA siendo el material genético puede ser apreciada entendiendo su estructura. El caudal de información hereditario, viene determinado por las diferentes disposiciones de los cuatro tipos de nucleótidos. Así, este capítulo tiene la intención de presentar aspectos básicos sobre el DNA.

El DNA fue identificado por Friedrich Miescher y el modelo de la doble hélice fue propuesto por James Watson y Francis Crick en 1953 basado los datos obtenidos por Rosalind Franklin [18].

### 3.1. Estructura

La molécula del DNA<sup>1</sup> es un polímero lineal dado por una sucesión donde cada unidad básica es un monómero llamado *nucléotido*. Los nucleótidos son moléculas formadas por

---

<sup>1</sup>DNA corresponden a las siglas del ácido desoxirribonucleico en inglés. Aunque en español sea ADN, internacionalmente se acepta DNA como un símbolo, análogo a los elementos químicos en la tabla periódica [5].

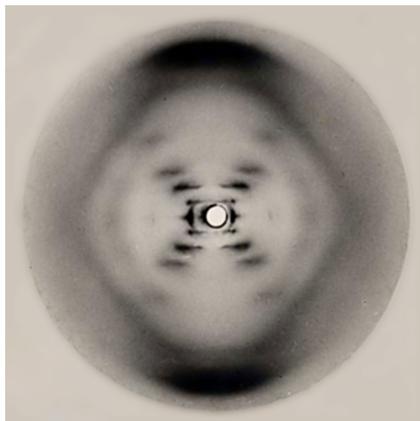


FIGURA 3.1: Fotografía 51. Imagen obtenida por Rosalind Franklin que sirvió como fundamento para el modelo de la doble hélice del DNA de Watson y Crick.

la unión covalente de tres componentes, que tienen una parte constante llamada *nucleósido*; la desoxirribosa, que es un monosacárido de cinco carbonos (pentosa) y una molécula de fosfato; y una parte variable constituida por cualquiera de las siguientes bases nitrogenadas: adenina (A), citosina (C), guanina (G) y timina (T). Es decir, las palabras en el lenguaje del DNA están caracterizadas por sucesiones lineales de As, Cs, Gs y Ts.

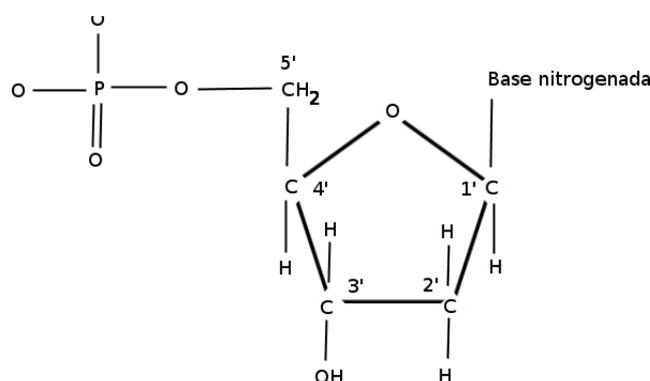


FIGURA 3.2: Estructura de un nucleótido.

Para formar esta secuencia es necesario que los monómeros se conecten entre sí de alguna forma. Primeramente debemos aclarar que por convención, a los azúcares y anillos heterociclos de los agrega un símbolo (') para evitar ambigüedades en la notación. Las componentes de los nucleótidos se encuentran unidas de la siguiente manera: la base nitrogenada se encuentra ligada a la posición 1' de la desoxirribosa, mientras que la molécula de fosfato se encuentra unida a la posición 5' de este monosacárido. A su vez, los nucleótidos se unen entre sí para formar el polímero, la posición 5' de la desoxirribosa de un nucleótido se conecta con la posición 3' del siguiente mediante una molécula de fosfato,

así por convención una cadena de nucleótidos corre en dirección  $5' \rightarrow 3'$  (Figura 3.3 izquierda).

Las dos hebras de DNA se encuentran unidas mediante *puentes de hidrógeno* entre cada una de las bases. Recordemos que un puente de hidrógeno es una atracción electrostática muy débil entre un átomo de hidrógeno unido mediante un enlace covalente y un átomo con un par de electrones no compartido (Figura 3.3 derecha).

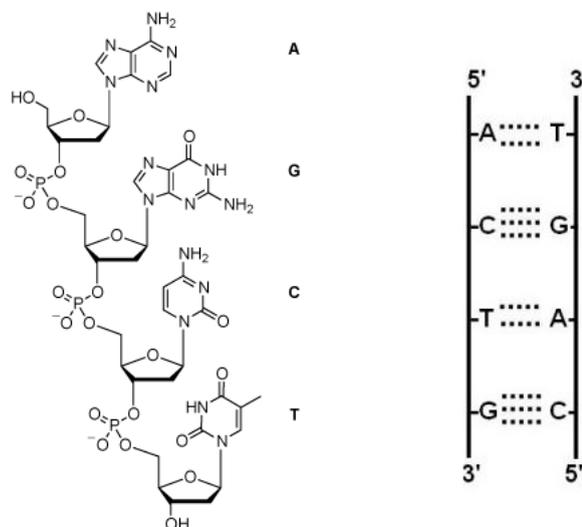


FIGURA 3.3: A la izquierda, la estructura de una hebra de DNA. A la derecha, puentes de hidrógeno entre los pares de bases según el modelo de Watson-Crick.

Por su parte, las bases nitrogenadas siguen una regla de asociación y tienen características que nos permitan trabajar con la secuencia de DNA de forma binaria, asignando ceros y unos a los nucleótidos, con base en lo que nos sugieren las siguientes clasificaciones:

1. Para que las dos hebras del DNA se unan entre sí, la condición que se tiene que cumplir es que adenina y timina siempre se correspondan, similarmente citosina y guanina. Esto nos permite conocer la información de la doble hélice basándonos solamente en lo que ocurre en una de las hebras puesto que esta condición nos indica que las bases son complementarias. Las parejas A-T están unidas por dos puentes de hidrógeno y C-G por tres. Así, con base en las conexiones de puentes de hidrógeno, A y T son débiles mientras que C y G son fuertes (Figura 3.4). Designamos a las primeras con  $W$  y a las segundas por  $S$ , de manera que:

$$WS = \begin{cases} \text{Fuertes (C, G)} & \rightarrow 1 \\ \text{Débiles (A, T)} & \rightarrow 0 \end{cases}$$

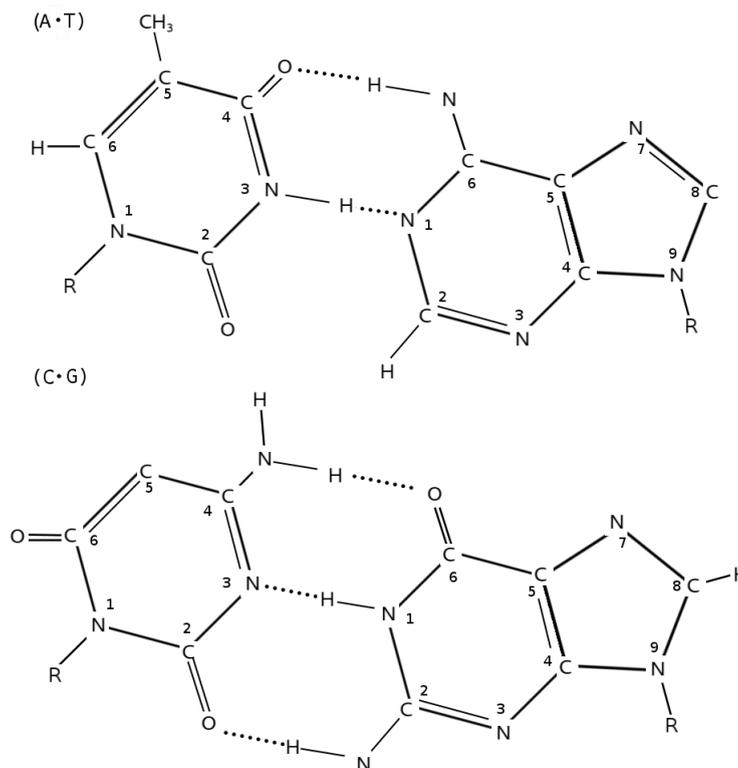


FIGURA 3.4: Puentes de hidrógeno entre los pares de bases según el modelo de Watson-Crick.

- En esta clasificación distinguimos a las bases mediante la cantidad de anillos compuestos de carbono hidrógeno presentes. Adenina y guanina son llamadas "grandes" ya que son moléculas de dos anillos y son llamadas purinas ( $R$ ). La Adenina tiene un grupo amino ( $-\text{NH}_2$ ) el carbono que se encuentra localizado en C6 (el carbono ubicado en la posición 6 en el anillo) y la guanina tiene un grupo amino en C2 y un grupo carbonilo en C6. Citosina y timina son llamadas "pequeñas" por ser de un anillo y son conocidas como pirimidinas ( $Y$ ). La citocina tiene un átomo de hidrógeno en C5 y un grupo amino en C4 mientras que la timina contiene un grupo metil en la posición C5 con grupos carbonilos en las posiciones C4 y C2 (Figura 3.5).

$$YR = \begin{cases} \text{Grandes (A, G)} & \rightarrow 1 \\ \text{Pequeñas (C, T)} & \rightarrow 0 \end{cases}$$

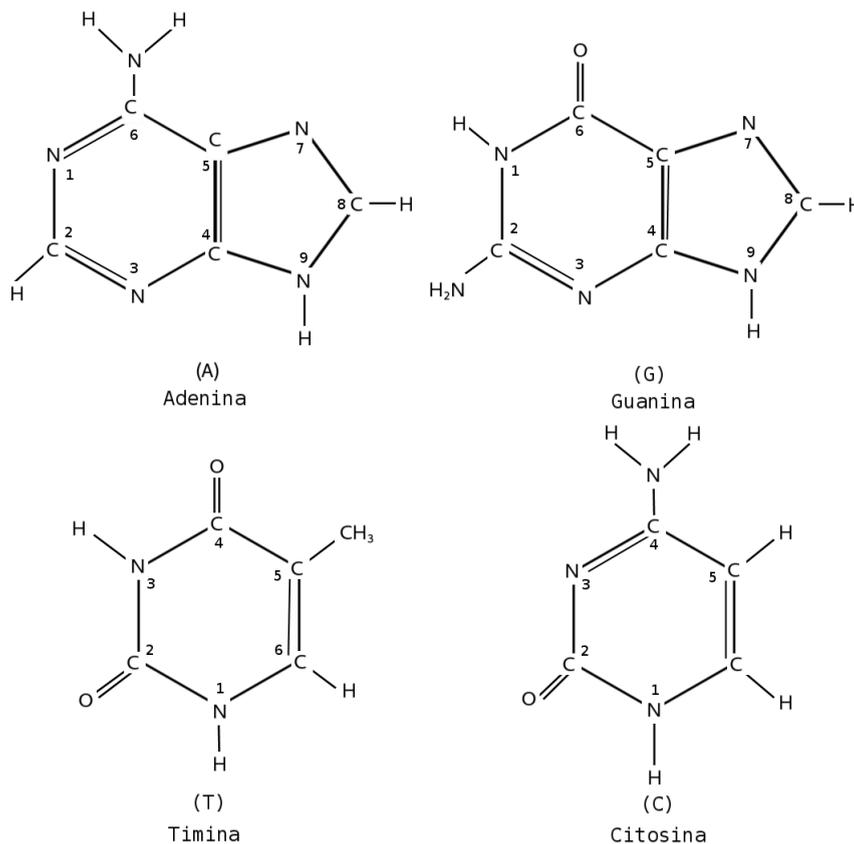


FIGURA 3.5: Estructura de las bases nitrogenadas.

3. La tercera y última clasificación separa a adenina y a citosina en un grupo amínico ( $M$ ), reflejado en los surcos mayores de la doble hélice, y a guanina y timina en uno cetónico ( $K$ ), lo que provoca diferencias estructurales, ya que el grupo cetónico tiene una carga ligeramente negativa, mientras que el grupo amínico ligeramente positiva. Cabe mencionar que esta combinación queda determinada a partir de las dos anteriores.

$$MK = \begin{cases} \text{Aminadas (A, C)} & \rightarrow 1 \\ \text{Cetónicas (G, T)} & \rightarrow 0 \end{cases}$$

Hasta este punto hemos explicado la estructura básica de la doble hélice al nivel de elementos que la conforman. Ahora, brevemente explicaremos la organización que tiene la cadena completa de DNA.

A todos los organismos conocidos los podemos incluir en dos grandes categorías: los *procariontes* y los *eucariontes*. Los primeros no tienen un núcleo definido y su molécula de DNA se encuentra en el citoplasma, organizado normalmente en un único cromosoma

circular. En los segundos el DNA se encuentra en el núcleo celular, y poseen uno o varios cromosomas lineales[5].

Además, en los organismos eucariontes el DNA existe una organización geométrica que consiste de una familia de proteínas llamas *histonas*. Esta familia está compuesta por las histonas H1, H2a, H2b, H3 y H4. Dos copias de H2a, H2b, H3 y H4 forman un núcleo de ocho moléculas alrededor del cual un hilo de 145 pares de bases nitrogenadas de la hélice del DNA se enrolla. La estructura es asegurada por una histona H1, que da estabilidad al complejo. Estas unidades de histonas con DNA son llamadas *nucelosomas*.

A su vez, los histonas se vuelven a enrollar formando la *cromatina*. Este enrollamiento puede ser comparado con los cables de teléfono, de esa manera el DNA se compacta en un factor de seis [5].

## 3.2. Estabilidad

Como veremos más adelante, algunas características importantes de la doble hélice del DNA son sus capacidades de replicación, transcripción y traducción, es por ello que nos concierne el estudio de los procesos de separación y unión.

Es importante que la información genética exista en forma de una doble helice. El principio de complementariedad del DNA nos provee de dos hebras que sirven de plantillas que pueden ser copiadas por la polimerasa del DNA, produciendo dos copias exactas del material hereditario, de manera que la forma de doble hélice es de suma importancia cuando una de las partes se ve dañada por químicos genotóxicos, ionización o radiación ultravioleta. Adicionalmente, gracias al empacamiento y condensación del DNA que vimos al final de la sección pasada, se asegura que la información genética es químicamente inerte.

Durante los procesos de replicación, transcripción y traducción, se puede observar al DNA de diferentes formas. A la secuencia de nucleótidos en una de las hebras, y que determina la complementaria se le llama *estructura primaria* del DNA. A la doble hélice formada por el torcimiento de la escalera de la estructura primaria del DNA y su complemento, recibe el nombre de *estructura secundaria*.

Estudios de difracción de rayos X han revelado que existe más de una posible conformación de la estructura secundaria, es decir, de las formas en la que se presenta la doble hélice. Estas formas pueden variar debido al porcentaje de bases nitrogenadas que lo conforman, a las condiciones de salinidad, etc. A alta salinidad la estructura dominante se llama A-DNA y a baja salinidad B-DNA, que es la que corresponde al modelo conocido por todos. Posteriormente se encontró una tercera conformación a la que se dio el nombre de Z-DNA, la cual queda caracterizada por una secuencia alternante de bases púricas y pirimidínicas (GCGCGC), debido a la cual sigue un curso de zig-zag.

En cuanto a su estabilidad termodinámica, se considera que la estructura de doble

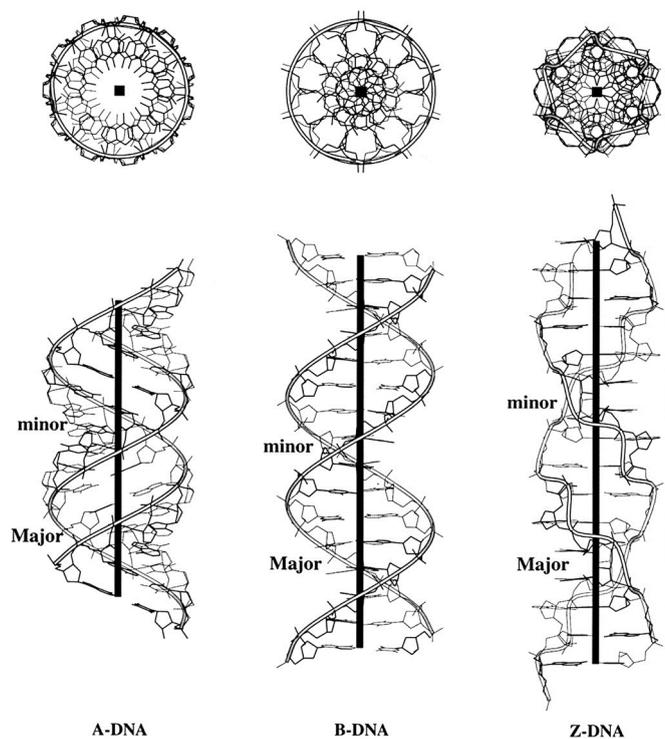


FIGURA 3.6: Diferentes formas de la estructura secundaria del DNA. En la parte superior

hélice notablemente estable, lo cual se debe principalmente a dos fuerzas químicas: los *puentes de hidrógeno* y al apilamiento de las bases nitrogenadas. Un puente de hidrógeno es un enlace que se establece entre moléculas capaces de generar cargas parciales.

Esto hace que la estabilidad del DNA depende de la composición de la sucesión que formen sus bases y de su longitud. Con esto, podemos deducir que la estabilidad del DNA depende de la interacción que existe entre los vecinos más cercanos en cada una de

sus componentes (termodinámica local), que puede ser considerada como la suma de las interacciones de los vecinos cercano y calculada mediante la siguiente tabla [2].

Termodinámica de local del DNA		
Interacción	$\Delta H$	$\Delta G$
AA/TT	9.1	1.9
AT/TA	8.6	1.5
TA/AT	6.0	0.9
CA/GT	5.8	1.9
GT/CA	6.5	1.3
CT/GA	7.8	1.6
GA/CT	5.6	1.6
CG/GC	11.9	3.6
GC/CG	11.1	3.1
GG/CC	11.0	3.1

FIGURA 3.7: Los valores que toma  $\Delta H$  indican la estabilidad energética dada en kcal y los de  $\Delta G$  indican la estabilidad relativa dados en kcal/mol.

Los pares de bases G-C tienen un puente de hidrógeno más que los pares A-T, y como podemos observar en los valores de la tabla, son más estables. De esa manera, con una mayor proporción de pares G-C que de pares A-T se necesita más energía para desnaturar al DNA completamente. Podríamos entonces pensar que la estructura primaria del DNA determina su estabilidad, sin embargo, es necesario tomar en cuenta otros factores, como las fuerzas de *van der Waals* que se originan de la interacción diagonal de las bases de una de las hebras con sus contrapartes en el complemento. Las fuerzas de van der Waals son atracciones entre átomos, moléculas y superficies; difieren de los enlaces covalentes y iónicos en el sentido de que son causadas por correlaciones en la fluctuación de las polarización de las partículas cercanas, como consecuencia de la dinámica cuántica.

La separación del DNA ocurre cuando se rompen los puentes de hidrógeno que mantienen unidas a las dos hebras, es decir, las dos fuerzas debidas a los puentes de hidrógeno y al apilamiento han sido superadas, sin embargo, los enlaces covalentes de la estructura primaria no se separan.

Los puentes de hidrógeno pueden ser separados por diversos motivos. Por ejemplo, existen condiciones bajo las cuales la doble hélice se rompe fácilmente de DNA en una incubación a  $\text{pH} > 12$  o  $\text{pH} < 2$  a causa de la ionización de las bases. Otra razón puede ser la ionización,

que resulta en un cambio en los puentes de hidrógeno, es decir, se alteran las cargas en los puentes de hidrógeno.

A su vez, también los medios muy ácidos degradan al DNA, es por ello que la desnaturalización del DNA se lleva a cabo con tratamientos alcalinos.

Incrementar la temperatura del DNA es otra manera de desestabilizarlo, lo que resulta en la separación de las dos hebras. El calor rompe los puentes de hidrógeno y destruye los escudos de hidratación del DNA, debilitando las fuerzas que mantienen unida a la doble hélice. La temperatura a la cual el 50 % de las cadenas de DNA se han desenrollado o desnaturalizado se le llama temperatura de fusión, denotada por  $T_m$ .

Como mencionamos al principio de la sección, la separación y unión del DNA es im-

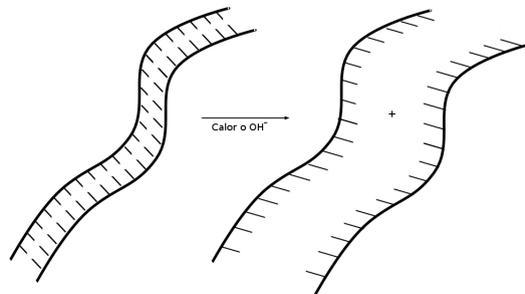


FIGURA 3.8: La doble hélice del DNA puede ser desnaturalizada por calor, en medios muy básicos o muy ácidos.

portante para los procesos de replicación, traducción y transcripción. Podemos ahora hacernos una idea de las condiciones bajo las cuales estos se llevan a cabo,

### 3.3. Replicación, transcripción y traducción

Anteriormente vimos que por el principio de complementariedad, basta conocer solamente una de las hélices del DNA. Algunos procesos requieren que la doble hélice del DNA sea separada en dos hebras de manera temporal.

## Replicación

En este proceso el DNA se duplica, en otras palabras, se forman dos copias idénticas, salvo por mutaciones debidas a sustituciones, supresiones o inserciones que ocurren durante el proceso de copiado, que quedarán insertadas en células distintas producto de la división celular.

La replicación del DNA comienza cuando este es desenrollado y sus puentes de hidrógeno son separados, siendo la helicasa la encargada de llevar acabo esta tarea. Así, cada una de las dos hebras de DNA sirve como plantilla para inserción de nuevos nucleótidos con base en el principio de complementariedad.

Después de haber sido separadas, un pequeño segmento de RNA llamado primer, provee de un punto de partida para la síntesis de DNA, que invariablemente se lleva en dirección  $5' \rightarrow 3'$ . Este primer es sintetizado previamente por una enzima llamada primasa.

Una enzima llamada polimerasa del DNA comienza a replicar al DNA añadiendo nucleótidos a la cadena original. Se conocen numerosos tipos de polimerasas, tanto en procariontes como en eucariontes, pero su mecanismo de acción es similar en todos los casos.

Una vez que la síntesis ha terminado, los primers son reemplazados con DNA, y cualquier hueco entre los segmentos del DNA sintetizado son sellados por enzimas.

Para asegurarse de que en el nuevo DNA sintetizado, la célula realiza correcciones de errores para evitar errores, supresiones o mutaciones, aunque también es importante mencionar que la acumulación o fijación de dichas mutaciones también depende de otros procesos. Sin embargo, lo cierto es que si la replicación fuera perfecta, no habría variación genética en las poblaciones, lo que implica que no habría evolución.

## Transcripción - traducción

La transcripción es un proceso en el cual la información de una cadena de DNA es copiada en una nueva molécula mensajera de RNA (mRNA), que determina la manera en que la célula produce los polipéptidos que eventualmente darán lugar a las proteínas. El RNA es idéntico al DNA, excepto que la timina es reemplazada por un uracilo (U).

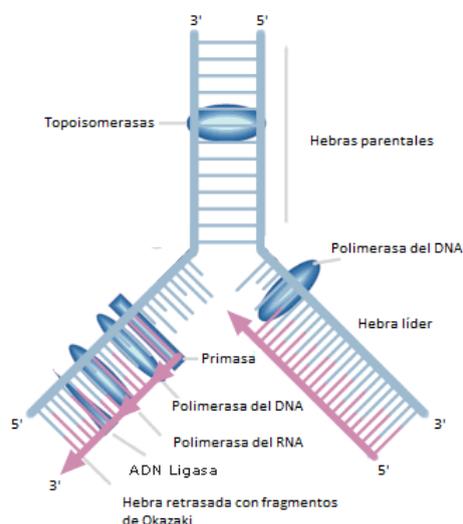


FIGURA 3.9: Replicación del DNA.

La transcripción es una de las etapas más importantes en el desarrollo de una célula o un organismo, ya que ella se definen a cuáles genes serán transferidos a la siguiente generación y con que frecuencia, es decir, la regulación génica.

El proceso de transcripción se inicia cuando una enzima llamada polimerasa del RNA se junta a una de las hebras que sirven de plantilla del DNA y comienza a catalizar la producción de RNA complementario. Un promotor es una secuencia de DNA que define donde la transcripción de un gen comienza.

De esa manera, el proceso de traducción comienza donde la polimerasa del RNA forma el RNA mensajero (mRNA) y lo utiliza en la síntesis de una proteína.

El mRNA se alarga, y una vez la síntesis haya terminado, la transcripción está completa.

Además del mRNA, durante la transcripción son creados el tRNA (RNA de transferencia) y el rRNA (RNA ribosomal). El tRNA se encarga de transportar los aminoácidos a los ribosomas y ordenarlos a lo largo del mRNA. Por su parte, el rRNA forma parte de los ribosomas. Estos se encargan de la síntesis de proteínas según la secuencia de nucleótidos.

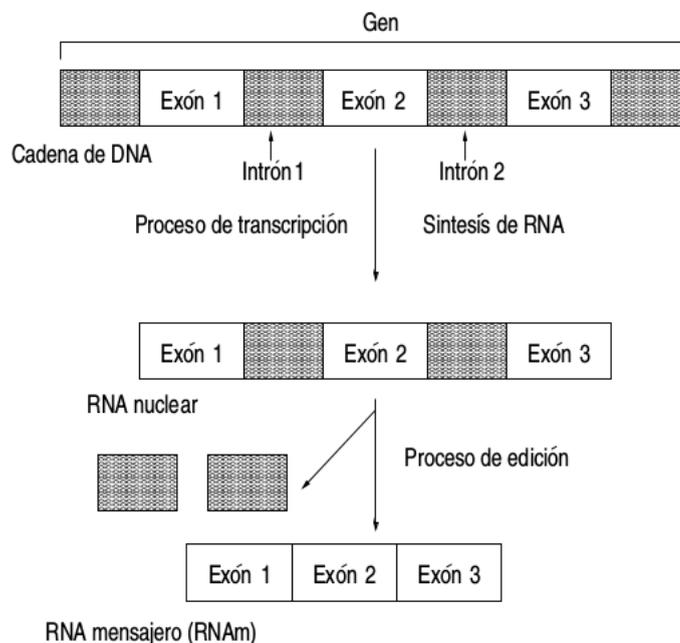


FIGURA 3.10: Proceso de edición bajo el cual se obtiene el mRNA maduro

Por su parte, la traducción es el proceso en que se sintetizan las nuevas proteínas a partir de la información contenida en el mRNA. Durante la traducción, el mRNA es traducido con base en el código genético, el cual define las distintas manera en que secuencias de tres nucleótidos son clasificadas, estas secuencias son llamadas comúnmente *codones*. En 1961 Crick y sus colegas mostraron que las secuencias deben leerse en tripletes no empalmados desde un punto fijo de comienzo. Esto quiere decir que un codón consiste de tres nucleótidos, los cuales pueden ser representados por trinucleótidos sucesivos.

La información codificable en proteínas de los genes de los organismos constituidos por células eucariontes está contenida en paquetes llamados exones, mientras que los segmentos no condificadores entre estos se les llama intrones. Los exones son las regiones representadas en el RNA maduro y los intrones son aquellos que son retirados de la secuencia recién transcrita del mRNA Figura 3.10 y sólo se encuentran en células eucariotas. Un gen comienza y termina con los exones.

La molécula de RNA mensajero que está libre de intrones, viaja a una estructura llamada *ribosoma*. El ribosoma es una molécula formada por moléculas de RNA ribosomal y proteínas que actúan como traductor del lenguaje del DNA y RNA al lenguaje de aminoácidos. En los ribosomas se lee secuencialmente el RNA y por cada tres bases se agrega un aminoácido a una cadena lineal conocida como polipéptido.

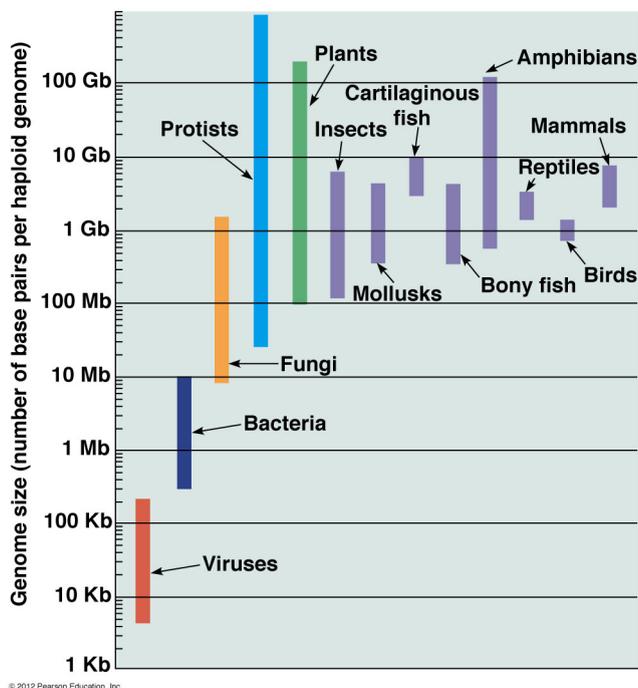


FIGURA 3.11: Tamaño del genoma de diferentes clases de organismos vivos.

### 3.4. Genes y genomas

Un gen estructural es un segmento de DNA que contiene la información necesaria para la síntesis de una proteína, es decir, un gen estructural consiste del DNA necesario para formar una proteína.

Al polímero completo, es decir, al total de DNA en un organismo se le llama *genoma*. De esa manera, un gen es una parte del del genoma. Entonces, el genoma se está dividido en cromosomas, que contienen a los genes que contienen al DNA. Por ejemplo, el genoma humano es la secuencia de DNA que está contenida en los 23 pares de cromosomas.

El DNA de un organismo puede ir desde los miles hasta los billones de nucleótidos, entre los primeros podemos mencionar a los virus y en el caso de los últimos a algunos anfibios.

Usualmente se piensa que los organismos simples son menos *complejos* y tienen menos funciones que realizar, y que de ahí que necesiten un menor acopio de información genética, sin embargo, la Figura 3.11 nos muestra claramente como esta idea es correcta. El número de funciones que un organismo está llamado a ejecutar en el curso de su vida permite hacerse de una idea de su grado de complejidad.

Cabe mencionar que en algunos casos presentados en los virus, el RNA funciona como el portador del material hereditario. El genoma de RNA puede ser de cadena sencilla, como la Hepatitis C o del virus de la inmunodeficiencia humana, o de doble cadena, como en el rotavirus.

En total, existen 64 combinaciones posibles a partir de ácidos nucleicos que forman codones, entre los que se encuentra un codón con el que se inicia a leer la secuencia (ATG), 61 combinaciones que se traducen en aminoácidos y otras 3 que indican lugares donde la traducción se detiene. Así, 61 codones solo pueden traducirse en 20 aminoácidos.

En la Figura 3.6 se muestra la equivalencia entre tripletes de DNA, representados por las cuatro bases nitrogenadas, y los aminoácidos. A la información contenida en esta tabla se le denomina *código genético* estándar. En la naturaleza únicamente son utilizados estos 20 aminoácidos para la construcción de proteínas, de lo que se infiere que existen codones que son sinónimos del mismo aminoácido, debido a esto se dice que el código genético es degenerado. Es importante recordar que no hay solapamientos, es decir, que cada ácido nucleico puede ser parte de a lo más un codón, por lo que se dice que es específico.

Debemos mencionar también que el código genético es casi universal, es decir, casi todos los organismos, organelos y virus conocidos comparten el mismo código con excepción de mitocondrias y algunos protozoos.

	T	C	A	G	
T	PHE	SER	TYR	CYS	T
T	PHE	SER	TYR	CYS	C
T	LEU	SER	STOP	STOP	A
T	LEU	SER	STOP	TRP	G
C	LEU	PRO	HIS	ARG	T
C	LEU	PRO	HIS	ARG	C
C	LEU	PRO	GLN	ARG	A
C	LEU	PRO	GLN	ARG	G
A	ILE	THR	ASN	SER	T
A	ILE	THR	ASN	SER	C
A	ILE	THR	LYS	ARG	A
A	MET	THR	LYS	ARG	G
G	VAL	ALA	ASP	GLY	T
G	VAL	ALA	ASP	GLY	C
G	VAL	ALA	GLU	GLY	A
G	VAL	ALA	GLU	GLY	G

FIGURA 3.12: Código genético. El primer nucleótido del triplete está indicado por la columna de la izquierda, el segundo por las columnas centrales y el tercero por la columna derecha. Los aminoácidos son codificados por el mRNA, donde las siglas indicadas en la tabla significan lo siguiente, ALA: Alanina, CYS: Cisteína, ASP: Ácido aspártico, GLU: Ácido glutámico, PHE: Felinalanina, GLY: Glicina, HIS: Histidina, ILE: Isoleucina, LYS: Lisina, LEU: Leucina, MET: Metionina, ASN: Asparagina, PRO: Prolina, GLN: Glutamina, ARG: Arginina, SER: Serina, TMR: Treonina, VAL: Valina, TPR: Triptofano, TYR: Tirosina, y STOP denota a los aminoácidos Pardo, Ámbar y Ocre que no codifican pero que sirven como signos de puntuación para finalizar el mensaje [3].

# Capítulo 4

## Fractales

La historia de los fractales no es reciente: podemos encontrar ejemplos de estos objetos matemáticos desde el siglo XIX como funciones continuas en todas partes, pero derivables en ninguna, sirviendo así como contraejemplos en el análisis matemático, lo que los llevó a ser vistos por algunos matemáticos como curvas monstruosas.

Entre estos podemos encontrar la función de Weierstrass, la curva de Koch, el conjunto de Cantor, el conjunto de Julia, entre otros, que datan de finales del siglo XIX y principios del siglo XX, sin embargo, no fue sino hasta finales del siglo XX cuando por primera vez se habló de *fractales*.

Tal vez fue que Gaston Julia, al menos en ese entonces, no pudo ver un conjunto de Julia, debido a que los fractales hicieron su aparición antes de que la tecnología estuviese lo suficientemente desarrollada para visualizarlos y que esto los llevara a convertirse en una disciplina propia. No fue sino hasta la década de los setenta, cuando Mandelbrot hizo ver que estos conjuntos tienen propiedades en común. Su visión está claramente plasmada en la siguiente:

*Las nubes no son esferas, las montañas no son conos, las costas no son círculos, y las cortezas de los árboles no son lisas, ni los relámpagos viajan en una línea recta.*

Mandelbrot planteó todo esto en una época donde las computadoras le permitieron realizar y ver imágenes de algunos fractales.

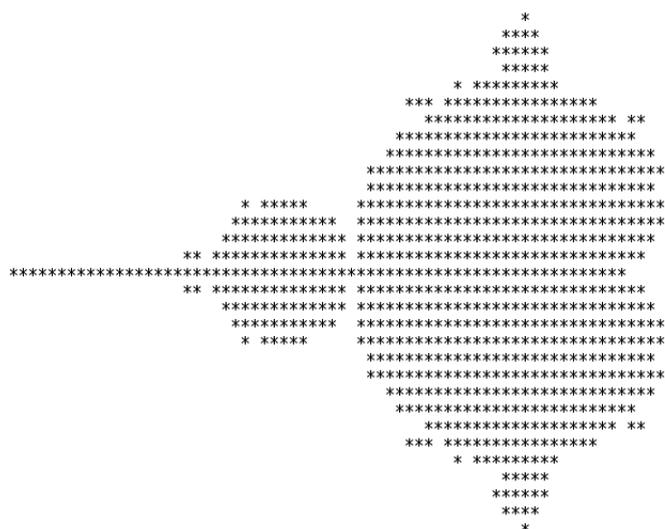


FIGURA 4.1: Primera imagen del conjunto de Mandelbrot

A lo largo de este capítulo explicaremos formalmente la teoría de los fractales. Comenzaremos hablando del concepto de fractal, posteriormente introduciremos algunos conceptos básicos para el seguimiento del capítulo para finalmente abordar la dimensión fractal y la autosimilaridad.

## 4.1. Concepto

La palabra fractal por si misma nos da ideas intuitivas de lo que es, incluso la definición que Mandelbrot dio y la mayoría de nosotros conoce es la siguiente:

**Definición 4.1.** *Decimos que un conjunto  $X$  en  $\mathbb{R}^n$  es un fractal si:*

$$\dim_T(X) < \dim_H(X)$$

Donde  $\dim_T(X)$  es la *dimensión topológica*<sup>1</sup> y  $\dim_H(X)$  es la dimensión fractal. Sin embargo, el aceptar esta definición nos privaría de algunos conjuntos como la Curva de Takagi, cuya dimensión de fractal es 1 y coincide con su dimensión topológica, lo que nos indica que este podría ser fractal. Por otro lado, tampoco podemos decir que los fractales son aquellos conjuntos que exhiben autosimilaridad, ya que bajo esta condición, una línea

<sup>1</sup>La dimensión topológica siempre toma valores enteros y se define inductivamente partiendo de que el conjunto vacío tiene dimensión -1.

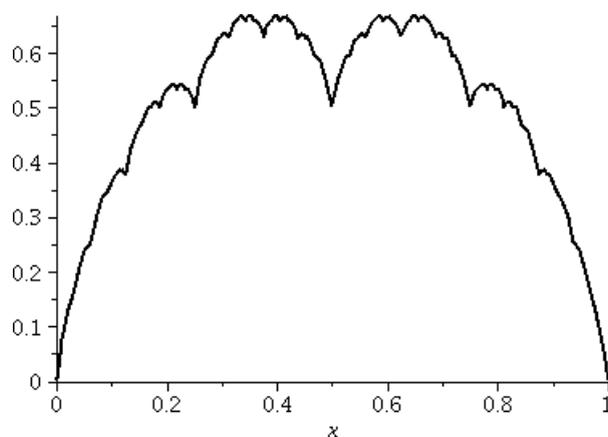


FIGURA 4.2: Curva de Takagi.

recta podría ser considerada como un fractal, ya que es invariante ante cualquier escala de observación.

Esto nos dice lo siguiente: no hay un acuerdo para definir un fractal (o fractal no tiene definición). Como menciona Falconer en [4], la definición de *fractal* debe ser considerada como la definición de *vida*: se le da un significado a la palabra *vida* a través de las características y propiedades de los seres vivos, sin embargo, con cualquier definición dejamos algunos casos fuera de ella. Análogamente, para nuestros fines consideraremos que un fractal es un conjunto  $F$  con las siguientes propiedades:

1.  $F$  tiene una estructura fina, es decir, detallada ante cambios de escala.
2.  $F$  es demasiado irregular para ser descrito en el lenguaje de la geometría tradicional, tanto local como globalmente.
3.  $F$  presenta autosimilaridad, ya sea aproximada o estadística.
4. La dimensión de Hausdorff de  $F$  es mayor que su dimensión topológica.
5.  $F$  puede ser descrito de manera sencilla, por ejemplo recursivamente.

Como un ejemplo de un fractal vamos a construir el conjunto de Cantor.

Como se muestra en la Figura 4.3 tomemos una línea recta, la cual consideraremos que tiene longitud 1 y la vamos a llamar  $\mathcal{C}_0$ . Ahora, dividamos esta línea en tres partes iguales y quitemos el segmento central, de manera que nos quedamos con dos segmentos, cada uno con longitud  $\frac{1}{3}$ , al conjunto formado por estos le llamamos  $\mathcal{C}_1$ . Enseguida repetimos el

proceso con estos dos segmentos, obteniendo así cuatro segmentos de longitud  $\frac{1}{9}$  y análogamente obtenemos  $\mathcal{C}_2$ . Similarmente repetimos el proceso una cantidad numerable de veces.

En cada paso vamos obteniendo una mayor cantidad de segmentos de longitud cada vez menor. Definimos entonces al conjunto de Cantor como

$$\mathcal{C} = \bigcap_{n=0}^{\infty} \mathcal{C}_n$$

Si comparamos el primer renglón y el segundo de la Figura 4.3 nos damos cuenta que son similares entre si, y que podemos hacer la misma comparación entre cualesquiera dos renglones. Esto nos indica que el conjunto de Cantor es un fractal, y al calcular su dimensión fractal, como veremos más adelante, obtenemos que esta es 0.63. Al ser su dimensión fractal mayor que 0 nos indica que el conjunto es más que un punto, y al ser menor que 1 nos dice que es menos que una línea.



FIGURA 4.3: Conjunto de Cantor

## 4.2. Espacios métricos

Primero, definiremos algunos conceptos que nos permitirán avanzar en la teoría de los fractales.

**Definición 4.2.** Sea  $\mathbf{X}$  un conjunto no vacío. Una función  $d: \mathbf{X} \times \mathbf{X} \rightarrow \mathbb{R}$  es llamada una métrica si satisface las siguientes condiciones:

1.  $d(x,y) \geq 0$  para toda  $x, y \in \mathbf{X}$
2.  $d(x,y) = 0$  si y sólo si  $x = y$
3.  $d(x,y) = d(y,x)$  para toda  $x, y \in \mathbf{X}$
4.  $d(x,y) \leq d(x,z) + d(z,y)$  para toda  $x, y, z \in \mathbf{X}$

**Definición 4.3.** Un espacio métrico es un par ordenado  $(\mathbf{X}, d)$  donde  $\mathbf{X}$  es un conjunto no vacío y  $d$  es una métrica en  $\mathbf{X}$ .

Los conjuntos que tienen estructura de espacio métricos son, en cierta manera, sencillos de trabajar, ya que la métrica nos da información acerca de la cercanía entre dos puntos del espacio.

**Definición 4.4.** Sea  $(\mathbf{X}, d)$  un espacio métrico y  $x_0 \in \mathbf{X}$ . La bola abierta de radio  $\varepsilon$  con centro en  $x_0$  es el conjunto

$$B_\varepsilon(x_0) = \{y \in \mathbf{X} : d(x_0, y) < \varepsilon\}$$

Es decir, la bola abierta de radio  $\varepsilon$  alrededor de  $x_0$  es el conjunto que queda determinado por los puntos cuya distancia a  $x_0$  es estrictamente menor que  $\varepsilon$ .

**Definición 4.5.** Sea  $U \subset \mathbf{X}$ , decimos que  $U$  es abierto si para todo  $x \in U$  existe  $\varepsilon > 0$  tal que  $B_\varepsilon(x)$  está completamente contenida en  $U$ .

En otras palabras, en un conjunto abierto, todos sus elementos son el centro de una bola abierta contenida totalmente en el conjunto.

A la colección de todos los conjuntos abiertos en  $(\mathbf{X}, d)$  se le llama una *topología*, la cual satisface las siguientes propiedades:

**Proposición 4.6.** Sea  $(\mathbf{X}, d)$  un espacio métrico.

1.  $\mathbf{X}$  y  $\emptyset$  son conjuntos abiertos.
2. Si  $\{U_\alpha\}$  es una colección arbitraria de conjuntos abiertos con  $\alpha \in I$  un conjunto de índices, entonces la unión de estos conjuntos,  $\bigcup_{\alpha \in I} U_\alpha$ , es un conjunto abierto.
3. Si  $U_1, U_2, \dots, U_n$  es una colección finita de conjuntos abiertos, entonces la intersección de ellos,  $\bigcap_{i=1}^n U_i$ , es un conjunto abierto.

**Demostración:**

1. Por definición, se tiene que si  $x$  está en  $\mathbf{X}$ , entonces  $B_1(x) \subset \mathbf{X}$ . Ahora bien, la proposición “si  $x$  es elemento de  $\emptyset$  entonces existe  $\varepsilon > 0$  tal que  $B_\varepsilon(x)$  está contenida en  $\emptyset$ ” es verdadera, porque el antecedente “ $x$  es elemento de  $\emptyset$ ” es falso. Por lo tanto,  $\mathbf{X}, \emptyset$  son abiertos.

2. Si  $x \in \bigcup_{\alpha \in I} U_\alpha$ , entonces  $x$  está en alguna de los  $U_\alpha$ , digamos  $U_{\alpha_0}$ . Como  $U_{\alpha_0}$  es abierto, entonces podemos encontrar  $\varepsilon > 0$  tal que  $B_\varepsilon(x) \subset U_{\alpha_0}$ . Así  $B_\varepsilon(x) \subset \bigcup_{\alpha \in I} U_\alpha$ , por lo que es abierto.
3. Si  $x \in \bigcap_i U_i$ , entonces  $x \in U_i$  para cada  $i = 1, 2, \dots, n$ . Como cada uno de los conjuntos  $U_i$  es abierto, existen  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$  mayores que 0 tales que  $B_{\varepsilon_i}(x) \subset U_i$ , para cada  $i$ . Si tomamos  $\varepsilon$  como el mínimo de todos los  $\varepsilon_i$ , entonces  $B_\varepsilon(x)$  está contenida en  $B_{\varepsilon_i}(x)$  que a su vez es subconjunto de  $U_i$  para cada  $i$ , y por lo tanto  $B_\varepsilon(x) \subset \bigcap_i U_i$ . Por lo tanto,  $\bigcap_{i=1}^n U_i$ , es un conjunto abierto.

□

**Definición 4.7.** Sea  $(\mathbf{X}, d)$  un espacio métrico,  $A \subset \mathbf{X}$  y  $x_0 \in \mathbf{X}$ . Decimos que  $x_0$  es un punto de acumulación (o punto límite) de  $A$ , si para todo  $\varepsilon > 0$  la bola  $B_\varepsilon(x_0)$  contiene puntos de  $A$  distintos de  $x_0$ , es decir

$$B_\varepsilon(x_0) \cap A \setminus \{x_0\} \neq \emptyset$$

Si  $x_0 \in A$  y no es punto de acumulación de  $A$ , entonces decimos que  $x_0$  es un punto aislado de  $A$ .

**Definición 4.8.** Sea  $(\mathbf{X}, d)$  un espacio métrico y  $E \subset \mathbf{X}$ . Decimos que  $E$  es cerrado si contiene a todos sus puntos de acumulación.

**Proposición 4.9.** Sea  $(\mathbf{X}, d)$  un espacio métrico. Si  $E \subset \mathbf{X}$  es cerrado entonces su complemento  $\mathbf{X} \setminus E$  es abierto.

**Demostración.** Si  $E$  es cerrado, entonces para cada  $x$  en el complemento, entonces existe un  $\varepsilon > 0$  tal que  $B_\varepsilon(x) \cap E = \emptyset$ , ya que  $x$  no es un punto de acumulación, es decir,  $B_\varepsilon(x)$  está completamente contenida en el complemento. Por lo tanto,  $\mathbf{X} \setminus E$  es abierto.

□

Al igual que en el caso de los conjuntos abiertos, con la colección de conjuntos cerrados se puede formar una topología, la cual satisface lo siguiente,

**Proposición 4.10.** Sea  $(\mathbf{X}, d)$  un espacio métrico. Entonces:

1.  $\mathbf{X}$  y  $\emptyset$  son conjuntos cerrados.

2. Si  $\{E_\alpha\}$  es una colección de conjuntos cerrados donde  $\alpha \in I$  un conjunto arbitrario de índices, entonces  $\bigcap_{\alpha \in I} E_\alpha$  es un conjunto cerrado.
3. Si  $E_1, E_2, \dots, E_n$  es una colección finita de conjuntos cerrados, entonces  $\bigcup_{i=1}^n E_i$  es cerrado.

**Demostración:**

1.  $\mathbf{X}$  y  $\emptyset$  son cerrados por ser complementos de los conjuntos abiertos  $\emptyset$  y  $\mathbf{X}$ , respectivamente.
2. Tomemos  $\{E_\alpha\}$  una colección de conjuntos cerrados y por las leyes de De Morgan,

$$\mathbf{X} \setminus \bigcap_{\alpha \in I} E_\alpha = \bigcup_{\alpha \in I} (\mathbf{X} \setminus E_\alpha).$$

Como  $\mathbf{X} \setminus E_\alpha$  es abierto y la unión arbitraria de conjuntos abiertos es un conjunto abierto, tenemos que  $\bigcap_{\alpha \in I} E_\alpha$  es un conjunto cerrado.

3. De manera similar, consideremos  $E_i$  un conjunto cerrado para  $i = 1, 2, \dots, n$ , así

$$\mathbf{X} \setminus \bigcup_{i=1}^n E_i = \bigcap_{i=1}^n (\mathbf{X} \setminus E_i)$$

Cómo la parte derecha de la igualdad es abierta por ser la intersección finita de conjuntos abiertos, la parte izquierda también lo es. De aquí  $\bigcup_{i=1}^n E_i$  es cerrado.

**Definición 4.11.** Una separación de  $(\mathbf{X}, d)$  es un par  $U, V$  de abiertos disjuntos no triviales de  $\mathbf{X}$  cuya unión es  $\mathbf{X}$ . Se dice que  $(\mathbf{X}, d)$  es conexo si no existe una separación de  $\mathbf{X}$ .

**Teorema 1.** La imagen continua de un espacio conexo bajo una función continua es conexa.

**Demostración:** Sea  $f : (\mathbf{X}, d) \rightarrow (\mathbf{Y}, d)$  una función continua. Queremos probar que  $f(\mathbf{X})$  es conexo.

Supongamos que  $U \cup V$  es una separación de  $f(\mathbf{X})$ , entonces  $f^{-1}(U) \cup f^{-1}(V)$  constituye una separación de  $\mathbf{X}$  en dos abiertos no vacíos, puesto que  $f$  es continua y sobreyectiva, lo que implica que  $(\mathbf{X}, d)$  no es conexo.

□

**Definición 4.12.** Sea  $A \subset \mathbf{X}$  un subconjunto de un espacio métrico  $(\mathbf{X}, d)$ . Entonces  $A$  es conexo por trayectorias si, para cada par de puntos  $x, y \in A$ , existe una función continua  $f : [0, 1] \rightarrow A$ , de  $[0, 1]$  con la métrica euclídea a  $(A, d)$ , tal que  $f(0) = x$  y  $f(1) = y$ . A esta función  $f$  se le llama una trayectoria de  $x$  a  $y$ .

**Definición 4.13.** Sea  $(\mathbf{X}, d)$  un espacio métrico. Una sucesión en  $\mathbf{X}$  es una función  $f: \mathbb{N} \rightarrow \mathbf{X}$ . Denotamos a la sucesión como  $(x_n)_{n=1}^{\infty}$ .

Decimos que  $(x_n)_{n=1}^{\infty}$  converge a  $x \in \mathbf{X}$  si, para todo  $\varepsilon > 0$ , existe  $N \in \mathbb{N}$  tal que, si  $n \geq N$ , entonces  $d(x_n, x) < \varepsilon$ , y lo denotamos como  $x_n \rightarrow x$ .

También, al punto  $x$  al cual la sucesión converge, es llamado el límite de la sucesión, y se le denota como

$$x = \lim_{n \rightarrow \infty} x_n$$

**Proposición 4.14.** Si una sucesión  $(x_n)_{n=1}^{\infty}$  converge, entonces su límite es único.

**Demostración:** Supongamos que  $(x_n)$  converge a  $x$  y que a su vez  $(x_n)$  converge a  $y$ . Así, existen constantes  $N_1, N_2 \in \mathbb{N}$  tales que si  $n_1 \geq N_1, n_2 \geq N_2$  entonces

$$d(x_{n_1}, x) < \varepsilon/2 \text{ y } d(x_{n_2}, x) < \varepsilon/2$$

Si escogemos  $N = \max\{N_1, N_2\}$  entonces

$$d(x, y) \leq d(x, x_N) + d(x_N, y) < \varepsilon/2 + \varepsilon/2 = \varepsilon$$

Al ser  $\varepsilon$  arbitrario, concluimos que  $d(x, y) = 0$  y por lo tanto  $x = y$ .

□

**Definición 4.15.** Una sucesión  $(x_n)_{n=1}^{\infty}$  en un espacio métrico se dice de Cauchy si, para  $\varepsilon > 0$ , existe  $N \in \mathbb{N}$  tal que

$$d(x_n, x_m) < \varepsilon$$

para toda  $n, m \geq N$ .

Es decir, decimos que los elementos de la sucesión se van acercando cada vez más entre sí. Las sucesiones de Cauchy juega un papel importante en los espacios métricos, como veremos más adelante.

**Definición 4.16.** Decimos que un espacio  $(\mathbf{X}, d)$  es completo si toda sucesión de Cauchy en el espacio converge.

**Definición 4.17.** Sea  $(\mathbf{X}, d)$  un espacio métrico y  $A \subset \mathbf{X}$ . Una cubierta de  $A$  es una familia  $\{U_\alpha\}$  con  $\alpha$  en un conjunto de índices  $I$ , de conjuntos abiertos en  $\mathbf{X}$  tales que

$$A \subset \bigcup_{\alpha \in I} U_\alpha$$

Es decir, la colección  $\{U_\alpha\}$  de conjuntos abierto cubre al conjunto  $A$ .

Por ejemplo, bajo esta definición siempre es posible cubrir cualquier conjunto con una colección de bolas abiertas de radio  $\varepsilon > 0$  de la siguiente manera:

$$A \subset \bigcup_{x \in A} B_\varepsilon(x),$$

**Definición 4.18.** Un subconjunto  $F$  de un espacio métrico  $\mathbf{X}$  es compacto si toda cubierta de  $F$  tiene una subcubierta finita.

De manera equivalente decimos que un conjunto es compacto si toda sucesión de elementos en el conjunto tiene una subsucesión convergente. Bajo esta equivalencia, al conjunto se le suele llamar secuencialmente compacto.

**Teorema 2.** Si  $Y$  es un subespacio de  $\mathbf{X}$ , entonces  $A \subset Y$  es compacto en  $Y$  si, y sólo si, es compacto en  $\mathbf{X}$ .

**Demostración:** Supongamos primero que  $A \subset Y$  es compacto en  $Y$ . Cualquier cubierta de  $A$  en  $Y$ , tiene una subcubierta finita. Consideremos  $\{U_\alpha\}$  una cubierta de  $A$  en  $\mathbf{X}$ . Definimos los conjuntos

$$V_\alpha = U_\alpha \cap Y,$$

los cuales forman una cubierta de  $A$  en  $Y$ , donde cada  $V_\alpha$  es abierto en  $Y$  y

$$A = A \cap Y \subset \bigcup_{\alpha} U_\alpha \cap Y = \bigcup_{\alpha} V_\alpha$$

Como  $A$  es compacto en  $Y$ ,  $\{V_\alpha\}$  tiene una subcubierta finita, digamos  $\{V_{\alpha_1}, V_{\alpha_2}, \dots, V_{\alpha_k}\}$ . Por lo tanto  $\{U_{\alpha_1}, U_{\alpha_2}, \dots, U_{\alpha_k}\}$  es una subcubierta finita de  $A$  en  $\mathbf{X}$ .

Para demostrar el otro sentido, supongamos que  $A \subset Y$  y que  $A$  es compacto en  $\mathbf{X}$ . Sea  $\{U_\alpha\}$  con  $\alpha \in I$  un conjunto de índices, una cubierta de  $A$  en  $Y$ . Entonces, para cada  $\alpha$ , existe un abierto  $V_\alpha$  en  $\mathbf{X}$  tal que

$$U_\alpha = Y \cap V_\alpha$$

Como  $A \subset \bigcup_{\alpha \in I} U_\alpha$  entonces

$$A \subset \bigcup_{\alpha \in I} V_\alpha$$

Por lo que  $\{V_\alpha\}$  es una cubierta de  $A$  en  $\mathbf{X}$ . Como  $A$  es compacto en  $\mathbf{X}$ , podemos seleccionar una subcubierta finita de los conjuntos  $V_\alpha$ , digamos

$$\{V_{\alpha_1}, \dots, V_{\alpha_k}\}$$

De manera que

$$\{U_{\alpha_1}, \dots, U_{\alpha_k}\}$$

es una subcubierta finita de  $A$  en  $Y$ , y por lo tanto es compacto en  $Y$ .

□

**Definición 4.19.** Sea  $(\mathbf{X}, d)$  un espacio métrico. Un subconjunto  $A$  es totalmente acotado si, para cada  $\varepsilon > 0$  existen  $x_1, x_2, \dots, x_n \in \mathbf{X}$  tales que

$$A \subset B_\varepsilon(x_1) \cup B_\varepsilon(x_2) \cup \dots \cup B_\varepsilon(x_n)$$

Al conjunto  $\{x_1, x_2, \dots, x_n\} \subset \mathbf{X}$  se le llama  $\varepsilon$ -red.

**Teorema 3.** Si  $A \subset (\mathbf{X}, d)$  es compacto, entonces es totalmente acotado.

**Demostración:** Supongamos que  $A$  no es totalmente acotado. Mostraremos entonces que  $A$  no puede ser secuencialmente compacto.

Como  $A$  no es totalmente acotado, existe un  $\varepsilon_0 > 0$  tal que  $A$  no puede ser cubierto por un número finito de bolas de radio  $\varepsilon_0$ . Construimos entonces la siguiente sucesión: consideramos  $x_1 \in A$  e inductivamente tomamos  $x_{n+1}$  tal que

$$x_{n+1} \in A \setminus \bigcup_{i=1}^n B_{\varepsilon_0}(x_i).$$

Esto es posible porque

$$A \not\subset \bigcup_{i=1}^n B_{\varepsilon_0}(x_i)$$

De manera que  $(x_n)$  no tiene subsucesiones convergentes porque

$$d(x_m, x_n) \geq \varepsilon_0$$

para cualesquiera  $m, n \in \mathbb{N}$ .

□

**Lema 4.20.** *Sea  $A \subset (\mathbf{X}, d)$  un espacio secuencialmente compacto y  $\{U_\alpha\}$  una cubierta  $A$ . Entonces existe  $\delta > 0$  tal que, para todo  $x \in A$ , existe un  $\alpha$  tal que*

$$B_\delta(x) \subset U_\alpha$$

**Demostración:** Supongamos que para cada  $\delta > 0$ , existe alguna bola  $B_\delta(x)$  que no está contenida en ningún  $U_\alpha$ . Entonces podemos construir una sucesión  $(x_n)$  en  $A$  tal que  $B_{1/n}(x_n)$  no está contenida en ningún  $U_\alpha$ . Como  $A$  es secuencialmente compacto,  $(x_n)$  tiene una subsucesión convergente, digamos  $x_{n_k} \rightarrow x$  en  $A$ . Como  $\{U_\alpha\}$  una cubierta,  $x \in U_{\alpha_0}$  para algún  $\alpha_0$ . Pero  $U_{\alpha_0}$  es abierto, por lo que

$$B_\varepsilon(x) \subset U_{\alpha_0}$$

para algún  $\varepsilon > 0$ . Ahora bien,  $x_{n_k} \rightarrow x$ , así que existe un entero  $N$  que podemos escoger mayor que  $2/\varepsilon$ , tal que,

$$d(x_N, x) < \varepsilon$$

Sin embargo, esto implica que

$$B_{1/N} \subset U_{\alpha_0}$$

porque, si  $z \in B_{1/N}$ , entonces  $d(z, x_N) < 1/N$ , y por la desigualdad de triángulo,

$$d(z, x) \leq d(z, x_N) + d(x_N, x) < \frac{1}{N} + \frac{\varepsilon}{2} = \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

y

$$z \in B_\varepsilon(x) \subset U_{\alpha_0}.$$

Que es una contradicción, por lo que debe existir un  $\delta$  con la propiedad deseada.

□

**Definición 4.21.** Sea  $(\mathbf{X}, d)$  un espacio métrico completo,  $x \in \mathbf{X}$  y  $B$  un conjunto compacto en  $\mathbf{X}$  distinto del vacío. Definimos

$$d(x, B) = \min\{d(x, y) : y \in B\}$$

$d(x, B)$  es llamada la distancia del punto  $x$  al conjunto  $B$ .

Ahora definimos el siguiente espacio:

$$\mathcal{C}(\mathbf{X}) = \{A : A \neq \emptyset \text{ y es un subconjunto compacto de } \mathbf{X}\}.$$

Lo siguiente es definir una métrica sobre  $\mathcal{C}(\mathbf{X})$ , la cual llamaremos la métrica de Hausdorff sobre  $\mathcal{C}(\mathbf{X})$ .

**Teorema 4.** Para  $A, B \in \mathcal{C}(\mathbf{X})$ , definimos

$$d_H(A, B) = \inf\{r > 0 : B \subset U_r(A) \text{ y } A \subset U_r(B)\}$$

donde  $U_r(A) = \{x \in \mathbf{X} : d(x, y) \leq r \text{ para alguna } y \in A\} = \bigcup_{y \in A} B_r(y)$ . Entonces  $d_H$  es una métrica sobre  $\mathcal{C}(\mathbf{X})$ . Más aun, si  $(\mathbf{X}, d)$  es completo, entonces  $(\mathcal{C}(\mathbf{X}), d_H)$  es completo.

**Demostración:** Primero mostraremos que en efecto  $d_H$  es una métrica, por lo que debemos verificar que se cumplan las cuatro propiedades que caracterizan a una métrica.

Notemos que, bajo la definición dada, es claro que  $d_H(A, B) = d_H(B, A)$  y  $d_H(A, A) = 0$ .

Para mostrar que  $d_H(A, B) = 0$  implica que  $A = B$ , observemos que para cualquier  $n$ ,  $A$  está contenido  $U_{1/n}(B)$ . Así, para cada  $x \in A$ , podemos escoger  $x_n \in B$  tal que  $d(x, x_n) \leq 1/n$ . Como  $B$  es cerrado,  $x \in B$  y obtenemos que  $A \subseteq B$ . De manera análoga podemos obtener la otra contención.

Para mostrar la desigualdad del triángulo consideremos dos números  $r$  y  $s$  tales que si

$$r > d_H(A, B) \text{ y } s > d_H(B, C),$$

entonces

$$C \subseteq U_{r+s}(A) \text{ y } A \subseteq U_{r+s}(C).$$

Lo anterior implica que  $r + s \geq d_H(A, C)$ . Por lo tanto,

$$d_H(A, B) + d_H(B, C) \geq d_H(A, C).$$

Ahora probaremos que  $(\mathcal{C}(\mathbf{X}), d_H)$  es completo. Para una sucesión de Cauchy  $\{A_n\}_{n \geq 1}$  en  $(\mathcal{C}(\mathbf{X}), d_H)$ , definimos

$$B_n = \overline{\bigcup_{k \geq n} A_k}$$

Primero, mostraremos que  $B_n$  es compacto. Como  $B_n$  es una sucesión monótona decreciente de conjuntos cerrados, basta con ver que  $B_1$  es compacto. Para cualquier  $r > 0$ , podemos escoger  $m$  tal que

$$A_k \subseteq U_{r/2}(A_m)$$

para toda  $k \geq m$ . Como  $A_m$  es compacto, existe una  $r/2$ -red  $P$  de  $A_m$ , donde  $P$  es el conjunto de puntos que conforman nuestra  $r/2$ -red. Como  $\bigcup_{x \in P} B_r(x)$  es cerrado, es fácil verificar que  $P$  también es una  $r$ -red de  $B_m$ . Añadiendo  $r$ -redes  $P$  de  $A_1, A_2, \dots, A_{m-1}$ , podemos obtener una  $r$ -red de  $B_1$ . Por lo tanto  $B_1$  es totalmente acotado. También,  $B_1$  es completo porque es un subconjunto cerrado de un espacio métrico completo  $\mathbf{X}$ . De aquí se sigue que  $B_n$  es compacto.

Ahora, como  $\{B_n\}$  es una sucesión monótona decreciente de compactos no vacíos,

$$A = \bigcap_{n \geq 1} B_n$$

es compacto y no vacío. Para cualquier  $r > 0$ , podemos escoger  $m$  tal que  $A_k \subseteq U_r(A_m)$  para toda  $k \geq m$ . Entonces

$$A \subseteq B_m \subseteq U_r(A)$$

para  $m$  lo suficientemente grande. Por lo que tenemos  $d_H(A, A_m) \leq r$  para  $m$  lo suficientemente grande. Entonces  $A_m$  converge a  $A$  con la métrica de Hausdorff. Por lo tanto,  $(\mathcal{C}(X), d_H)$  es completo.

□

Este espacio métrico que definimos es precisamente el espacio métrico de los fractales, es decir, consideraremos que los fractales son conjuntos compactos de un espacio métrico. Como mencionamos al inicio de capítulo, buscamos que los fractales puedan ser expresados de manera sencilla, por lo que introducimos el siguiente concepto.

**Definición 4.22.** *Un sistema de funciones iteradas (IFS) consiste de una familia finita de contracciones en un espacio métrico completo.*

Los sistemas de funciones iteradas jugarán un papel importante más adelante, cuando hablemos de conjuntos *autosimilares*.

### 4.3. Dimensión Fractal

Estamos familiarizados con el concepto de dimensión gracias a la geometría y al álgebra lineal. De la primera aprendimos que un punto tiene dimensión cero, una línea dimensión uno y un cuadrado dimensión dos, sin embargo, la asociación que hacemos entre estos objetos y la naturaleza es un tanto artificial y en algunas ocasiones no es correcta. De la segunda obtenemos un concepto más general de dimensión, aplicado a los espacios vectoriales, donde la dimensión nos ayuda a clasificar a todos los espacios que son esencialmente los mismo, pero eso sigue sin ser lo que deseamos que la dimensión nos ayude a conocer.

En capítulos anteriores dimos una idea intuitiva de lo que es la dimensión fractal, misma que formalizaremos a lo largo de este apartado.

Debemos considerar que la dimensión de un conjunto  $\mathbf{X} \in \mathbb{R}^n$  satisfaga las siguientes propiedades:

1. Que para el conjunto unitario  $\{p\}$ ,  $\dim(\{p\}) = 0$ , para la recta  $l$ , para el intervalo unitario  $I^1$ ,  $\dim(I^1) = 1$ , y en general, para el hipercubo  $m$ -dimensional  $I^m$ ,  $\dim(I^m) = m$ .

2. (*Monotonía*) Si  $\mathbf{X} \subset \mathbf{Y}$ ,

$$\dim(\mathbf{X}) \leq \dim(\mathbf{Y})$$

3. (*Estabilidad numerable*) Si  $\{X_j\}$  es una sucesión de subconjuntos cerrados de  $(\mathbb{R})^n$ , entonces

$$\dim \left( \bigcup_{j=1}^{\infty} (X_j) \right) = \sup_{j \geq 1} \dim (X_j)$$

4. (*Invarianza*) Para una función arbitraria  $\psi$  que pertenece a alguna familia de homeomorfismos de  $(\mathbb{R})^n$  a  $(\mathbb{R})^n$

$$\dim(\psi(X)) = \dim(X)$$

Las condiciones 1 y 2 resultan bastantes naturales, que no es el caso de la condición 3, incluso en algunas ocasiones en lugar de considerar estabilidad numerable se considera una condición similar, la estabilidad finita y el inciso 3 quedaría escrito como

- 3'. Si  $X_1, X_2, \dots, X_m$  son subconjuntos cerrados de  $(\mathbb{R})^n$ , entonces

$$\dim \left( \bigcup_{j=1}^m (X_j) \right) = \max_{m \geq j \geq 1} \dim (X_j)$$

Brevemente introduciremos algunos conceptos de teoría de la medida que son esenciales para entender el concepto de dimensión y sus variantes.

**Definición 4.23.** Sea  $X$  un conjunto, decimos que  $\mathcal{A}$  es un álgebra en  $X$  si es una colección de subconjuntos de  $X$  que satisface la siguientes condiciones:

1.  $X \in \mathcal{A}$ .

2. Si  $B \in \mathcal{A}$ , entonces  $X \setminus B \in \mathcal{A}$ .

3. Si  $\{B_n\}$  es una colección finita de conjuntos en  $\mathcal{A}$ , entonces

$$\bigcup_{i=1}^n B_i,$$

pertenece a  $\mathcal{A}$ .

En particular, decimos que  $\mathcal{A}$  es  $\sigma$ -álgebra si  $\mathcal{A}$  es álgebra y si  $\{B_n\}$  es una colección numerable de conjuntos en  $\mathcal{A}$ . Entonces

$$\bigcup_{i \geq 1} B_i,$$

también está en  $\mathcal{A}$ .

**Definición 4.24.** Diremos que la pareja  $(\mathbf{X}, \mathcal{A})$  es un espacio de medida si  $\mathcal{A}$  es una  $\sigma$ -álgebra en  $\mathbf{X}$ .

**Definición 4.25.** Sea  $(\mathbf{X}, \mathcal{A})$  un espacio de medida. Una medida en  $\mathbf{X}$  es una función  $\mu : \mathcal{A} \rightarrow \mathbb{R}$  tal que

1.  $\mu(\emptyset) = 0$ .
2. Si  $B \in \mathcal{A}$ , entonces  $\mu(B) \geq 0$ .
3. Si  $\{B_n\}$  es una colección numerable de conjuntos en  $\mathcal{A}$ , tal que  $B_i \cap B_j = \emptyset$  para  $i \neq j$ , entonces

$$\mu\left(\bigcup_{i \geq 1} B_i\right) = \sum_{i \geq 1} \mu(B_i).$$

**Definición 4.26.** Una medida exterior  $\mu^*$  es una función cuyo dominio son los subconjuntos de  $\mathbf{X}$  y que cumple las siguientes condiciones:

1.  $\mu^*(\emptyset) = 0$
2. (Monotonía) Si  $A \subset B$ , entonces  $\mu^*(A) \leq \mu^*(B)$
3. (Subaditividad) Para toda sucesión de subconjuntos  $\{A_n\}$  de  $\mathbf{X}$  se tiene que

$$\mu^*\left(\sum_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$$

En particular, es de nuestro interés la que es llamada la medida exterior  $s$ -dimensional de Hausdorff, ya que con base en ella construiremos la dimensión de Hausdorff. Dicha dimensión es a la que queremos llegar, puesto que cumple con las condiciones mencionadas al inicio de la sección y es la que tomaremos sobre los fractales.

**Definición 4.27.** Decimos que el diámetro de un conjunto  $U$  es

$$|U| = \sup_{x,y \in U} \|x - y\|.$$

Así, una  $\delta$ -cubierta de un conjunto dado  $X$  es una sucesión numerable de conjuntos  $\{U_i\}$  que satisfacen

$$X \subset \bigcup_{i=1}^{\infty} U_i$$

con  $0 \leq |U_i| \leq \delta$ . Observemos también que la definición depende únicamente del diámetro de las  $\delta$ -cubiertas y no la manera en la que  $X$  es cubierto.

Anteriormente definimos el espacio métrico  $(\mathcal{C}(\mathbf{X}), d_H)$ , cuyos elementos son conjuntos compactos, por lo que nos enfocaremos a trabajar con ellos.

**Definición 4.28.** Sean  $X$  un subconjunto de  $(\mathcal{C}(\mathbf{X}), d_H)$ ,  $\delta > 0$  y  $\mathcal{U}$  la colección de todas las  $\delta$ -cubiertas  $\{U_i\}_{i \in I}$  posibles para  $X$ , definimos la medida exterior  $s$ -dimensional de Hausdorff como

$$\mathcal{H}_\delta^s(X) = \inf_{\mathcal{U}} \sum_{i=1}^{\infty} |U_i|^s$$

donde  $s$  es un real no negativo.

Para  $X$ , tenemos que  $\mathcal{H}_\delta^s(X)$  es una función de dos variables,  $s$  y  $\delta$ . Podemos pensar que esta función toma valores en  $[0, \infty]$ .

Si fijamos  $s$ , obtenemos que  $\mathcal{H}_\delta^s(X)$  es una función que depende del parámetro  $\delta$  la cual crece conforme  $\delta$  decrece. Como incluimos al infinito dentro del conjunto de valores que la medida exterior  $s$ -dimensional de Hausdorff puede tomar, siempre podemos obtener el siguiente límite

$$\mathcal{H}^s(X) = \lim_{\delta \rightarrow 0^+} \mathcal{H}_\delta^s(X)$$

Si consideramos  $s, t$  valores positivos tales que  $s < t$  y cualquier  $\delta$ -cubierta  $\{U_i\}$  de  $X$  tenemos

$$\sum_{i=1}^{\infty} |U_i|^t = \sum_{i=1}^{\infty} |U_i|^s |U_i|^{t-s} \leq \delta^{t-s} \sum_{i=1}^{\infty} |U_i|^s$$

Esto es,

$$\mathcal{H}_\delta^t \leq \delta^{t-s}(X) \mathcal{H}_\delta^s(X)$$

En particular, si  $\mathcal{H}_\delta^s(X) < \infty$  entonces  $\mathcal{H}_\delta^t(X) = 0$  y si  $\mathcal{H}_\delta^t(X) > 0$  entonces  $\mathcal{H}_\delta^s(X) = \infty$ .

En otras palabras, la gráfica de  $\mathcal{H}_\delta^s(X)$  es una función escalón con a lo más un punto de discontinuidad como se muestra en la Figura 4.4.

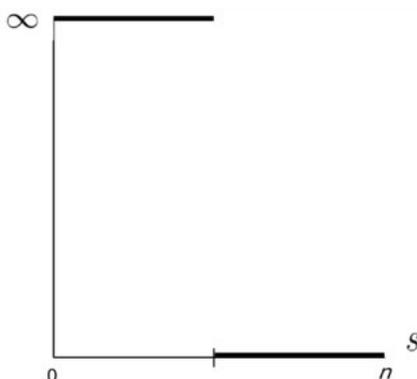


FIGURA 4.4: Gráfica de la función  $\mathcal{H}^s(X)$

**Definición 4.29.** Al valor donde la función  $\mathcal{H}^s(X)$  es discontinua, le llamamos la *dimensión de Hausdorff* de  $X$  y lo denotamos por  $\dim_H(X)$ .

Observemos que, a partir de lo observado en la Figura 4.4 se deducen las siguientes igualdades:

$$\dim_H(X) = \sup_s \{\mathcal{H}^s(X) = \infty\} = \inf_s \{\mathcal{H}^s(X) = 0\}$$

La dimensión de Hausdorff cumple todas las condiciones deseadas para una dimensión, sin embargo por como la definimos, no siempre es posible calcular, es por ello que introducimos otras dimensiones no enteras las cuales nos sirven para los fractales.

**Definición 4.30.** Sea  $X \subset \mathbb{R}^n$  y  $s \geq 0$ . La premedida  $s$ -dimensional de empaque de  $\mathbf{X}$  se define como

$$P_0^s(X) = \limsup_{\delta \rightarrow 0} \left\{ \sum_{i \in I} |B_i|^s \right\}$$

**Definición 4.31.** Sea  $F(r)$  el área trazada por círculos de radio  $r$  a lo largo de una curva fractal. Entonces,

$$D_M = \lim_{r \rightarrow 0} \frac{\ln F(r)}{-\ln r} + 2$$

existe y se le conoce como la dimensión de Minkowski - Bouligand.

**Definición 4.32.** Definimos la integral de correlación como

$$C(\varepsilon) = \lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i,j=1}^{\infty} H(\varepsilon - |x_i - x_j|),$$

donde  $H$  es la función escalón. Cuando el siguiente límite existe, la dimensión de correlación se define como

$$d_{cor} = \lim_{\varepsilon, \varepsilon' \rightarrow 0^+} \frac{\ln \frac{C(\varepsilon)}{C(\varepsilon')}}{\ln \frac{\varepsilon}{\varepsilon'}}$$

## 4.4. Autosimilaridad

De manera breve, diremos que un conjunto autosimilar es un conjunto que consiste de miniaturas de si mismo. Como hemos vistos en la sección anterior, hay aspectos que son tan sencillos alrededor de los fractales, es por ello que restringiremos nuestro estudio solamente a un tipo específico de fractales, es que poseen autosimilaridad, como el Triángulo de Sierpinski.

A lo largo de esta sección, mostraremos la existencia de conjuntos autosimilares y algunas de sus propiedades.

**Definición 4.33.** Sean  $(\mathbf{X}, d)$  un espacio métrico y  $f: \mathbf{X} \rightarrow \mathbf{X}$ . Un punto  $x \in \mathbf{X}$  para el cual  $f(x) = x$  es llamado un punto fijo.

**Definición 4.34.** Sean  $(\mathbf{X}, d_{\mathbf{X}})$  y  $(\mathbf{Y}, d_{\mathbf{Y}})$  espacios métricos. Un mapeo  $\mathbf{X} \rightarrow \mathbf{Y}$  se dice ser (uniformemente continuo) Lipschitz sobre  $\mathbf{X}$  con respecto a  $d_{\mathbf{X}}, d_{\mathbf{Y}}$  si

$$L = \sup_{x,y \in \mathbf{X}, x \neq y} \frac{d_{\mathbf{Y}}(f(x), f(y))}{d_{\mathbf{X}}(x, y)} < \infty$$

La constante  $L$  mencionada arriba es llamada la constante de Lipschitz de  $f$  y se denota por  $L = \text{Lip}(f)$ .

**Definición 4.35.** Sea  $(\mathbf{X}, d)$  un espacio métrico. Si  $f: \mathbf{X} \rightarrow \mathbf{X}$  es Lipschitz sobre  $\mathbf{X}$  con respecto a  $d$  y  $\text{Lip}(f) < 1$ , entonces  $f$  es llamada una contracción con radio  $r$ , donde  $r$  es la constasnte de Lipschitz.

En particular, una similaridad en  $\mathbf{X}$  es una función  $f: \mathbf{X} \rightarrow \mathbf{X}$  tal que

$$d(f(x), f(y)) = rd(x, y)$$

Como estamos interesados en trabajar con conjuntos autosimilares, formalmente los definimos de la siguiente manera:

**Definición 4.36.** Sea  $A \subset M$ , decimos que  $A$  es autosimilar si existen  $f_1, f_2, \dots, f_k$  contracciones tales que:

$$A = \bigcup_{i=1}^k f_i(A)$$

Ahora bien, dado un conjunto de contracciones, nos cuestionamos si siempre existe un conjunto autosimilar bajo las condiciones planteadas en la definición y además es único, lo cual constestaremos con los siguientes:

**Teorema 5. (mapeo contractivo).** Sean  $(\mathbf{X}, d)$  un espacio métrico completo y  $f: \mathbf{X} \rightarrow \mathbf{X}$  una contracción con respecto a la métrica  $d$ . Entonces existe un único punto fijo. Más aun, si  $x^*$  es punto fijo de  $f$ , entonces  $\{f^n(a)\}_{n \geq 0}$  converge a  $x^*$  para toda  $a \in \mathbf{X}$ , donde  $f^n$  es la  $n$ -ésima iteración de  $f$ .

**Demostración:** Si  $r$  es el radio de contracción de  $f$ , entonces, para  $m > n$

$$\begin{aligned} d(f^n(a), f^m(a)) &\leq d(f^n(a), f^{n+1}(a)) + \dots + d(f^{m-1}(a), f^m(a)) \\ &\leq (r^n + \dots + r^{m-1})d(a, f(a)) \leq \frac{r^n}{1-r}d(a, f(a)) \end{aligned}$$

Por lo tanto  $\{f^n(a)\}_{n \geq 0}$  es una sucesión de Cauchy. Como  $(\mathbf{X}, d)$  es completo, entonces existe  $x^* \in X$  tal que  $f^n(a) \rightarrow x^*$  cuando  $n \rightarrow \infty$ . Como  $f^{(n+1)}(a) = f(f^n(a))$  y  $f$  es continua,  $x^* = f(x^*)$ .

Ahora, si  $f(x) = x$  y  $f(y) = y$ , entonces  $d(x, y) = d(f(x), f(y)) \leq rd(x, y)$ . Entonces  $d(x, y) = 0$  y  $x = y$ , con lo que probamos la unicidad de los punto fijo.

□

**Teorema 6.** *Dada una familia de contracciones  $\{\varphi_1, \varphi_2, \dots, \varphi_m\}$ ,  $m \geq 2$  en  $\mathcal{C}(\mathbf{X})$ , existe un único conjunto autosimilar  $V$ .*

**Demostración:** Primero definimos un mapeo  $\Phi: \mathcal{C}(\mathbf{X}) \rightarrow \mathcal{C}(\mathbf{X})$  tal que

$$\Phi(A) = \bigcup_{i=1}^m \varphi_i(A)$$

Como cada  $\varphi_i$  es continua, aseguramos que  $\Phi$  mapea  $\mathcal{C}(\mathbf{X})$  en  $\mathcal{C}(\mathbf{X})$ . Como  $\mathcal{C}(\mathbf{X})$  es un espacio métrico completo, lo que queda es mostrar que  $\Phi$  es una contracción en  $\mathcal{C}(\mathbf{X})$ .

Para conjuntos arbitrarios  $A_0, A_1, A_2, A_3 \in \mathcal{C}(\mathbf{X})$  se siguen las siguientes propiedades:

1.  $d_H(\varphi_i(A_0), \varphi_i(A_1)) \leq \text{Lip}(\varphi_i)d_H(A_0, A_1)$
2.  $d_H(A_0 \cup A_1, A_2 \cup A_3) \leq \text{máx}\{d_H(A_0, A_2), d_H(A_1, A_3)\}$

Para demostrar 1, hacemos  $s = d_H(A_0, A_1)$ ; entonces para cualquier  $x \in A_0$  existe  $y \in A_1$  tal que  $d(x, y) \leq s$ . Esto implica que

$$d(\varphi_i(x), \varphi_i(y)) \leq \text{Lip}(\varphi_i)d(x, y) \leq \text{Lip}(\varphi_i)s,$$

definimos  $N_\varepsilon = \{x \in \mathbf{X} : d(x, A) \equiv \min_{y \in A} d(x, y) \leq \varepsilon\}$  y observemos que  $\varphi_i(x) \in N_t(\varphi_i(A_1))$ , donde  $t = \text{Lip}(\varphi_i)s$ . De aquí que  $\varphi_i(A_0) \subset N_t(\varphi_i(A_1))$ , debido a que escogimos  $x$  de manera arbitraria. Para probar 2 establecemos  $s = d_H(A_0, A_2)$  y  $t = d_H(A_1, A_3)$ ; entonces las inclusiones  $A_0 \subset N_s(A_2)$  y  $A_1 \subset N_t(A_3)$  nos dan las inclusiones

$$A_0 \cup A_1 \subset N_s(A_2) \cup N_t(A_3) \subset N_r(A_2 \cup A_3)$$

donde  $r = \max\{s, t\}$ . De manera análoga se muestra la segunda inclusión.

Para cualquier  $A, B \in \mathcal{C}(\mathbf{X})$  usamos repetidamente la desigualdad 2 para obtener

$$\begin{aligned} d_H(\Phi(A), \Phi(B)) &= d_H\left(\bigcup_{i=1}^m \varphi_i(A), \bigcup_{i=1}^m \varphi_i(B)\right) \\ &\leq \max_{1 \leq i \leq m} d_H(\varphi_i(A), \varphi_i(B)) \end{aligned}$$

A su vez, de la propiedad 1, se sigue

$$\max_{1 \leq i \leq m} d_H(\varphi_i(A), \varphi_i(B)) \leq \left( \max_{1 \leq i \leq m} \text{Lip}(\varphi_i) \right) d_H(A, B)$$

de las cuales, obtenemos

$$\max_{1 \leq i \leq m} \text{Lip}(\varphi_i) < 1$$

Lo que nos permite concluir que  $\Phi: \mathcal{C}(\mathbf{X}) \rightarrow \mathcal{C}(\mathbf{X})$  es una contracción.

□

**Definición 4.37.** Sea  $\{\psi_1, \psi_2, \dots, \psi_m\}$  un conjunto de contracciones, definimos la dimensión de similaridad del conjunto autosimilar  $V$  como la posible raíz de la siguiente ecuación en  $d$

$$\sum_{i=1}^m (L(\psi_i))^d = 1$$

la cual depende únicamente de la constante de Lipschitz de las contracciones. Denotamos este valor por  $\dim_s(V)$ .

**Lema 4.38.** Para un conjunto autosimilar  $V$ ,

$$\dim_H(V) \leq \dim_s(V)$$

**Demostración:** Es suficiente con mostrar que  $\mathcal{H}^s(V) < \infty$ , donde  $s = \dim_s(V)$ . Dado un subconjunto  $A$  de  $\mathbb{R}^n$  y una sucesión,

$$i_1 i_2 \dots i_k \quad (1 \leq i_j \leq m, j = 1, 2, \dots, k)$$

de  $k$  números que van desde 1 hasta  $m$  y por simplicidad usamos la siguiente abreviación:

$$A_{i_1 i_2 \dots i_k} \equiv \psi_1 \circ \psi_2 \circ \dots \circ \psi_{i_k}(A).$$

Entonces, sabemos que existe un único conjunto tal que

$$A_{i_1 i_2 \dots i_k} = V$$

Además, tenemos

$$|V_{i_1 i_2 \dots i_k}| \leq |V| \prod_{j=1}^k L(\psi_{i_j}) \leq \lambda^k |V|$$

donde  $\lambda$  es el mayor entre las constantes de Lipschitz  $L(\psi) < 1$ . Sea  $\delta = \lambda^k |V|$ . Entonces el conjunto  $\{V_{i_1 i_2 \dots i_k}\}$  es una  $\delta$ -cubierta de  $V$ , y entonces tenemos

$$\begin{aligned} \mathcal{H}_\delta^s(V) &\leq \sum_{i_1=1}^m \dots \sum_{i_k=1}^m |V_{i_1 i_2 \dots i_k}|^s \leq |V|^s \sum_{i_1=1}^m \dots \sum_{i_k=1}^m \prod_{j=1}^k (L(\psi_{i_j}))^s \\ &\leq |V|^s \left( \sum_{i=1}^m (L(\psi_i))^s \right)^s = |V|^s \end{aligned}$$

Ahora podemos disminuir el valor de  $\delta$  tanto como queramos tomando  $k$  lo suficientemente grande. De aquí se sigue  $\mathcal{H}^s(V) \leq |V|^s$ .

□

**Observación:** Decimos que un conjunto de contracciones  $\{\varphi_1, \varphi_2, \dots, \varphi_m\}$  satisface la condición del conjunto abierto si existe un conjunto abierto no vacío  $U \subset \mathbf{X}$  tal que

$$\varphi_i(U) \subset U, \quad 1 \leq i \leq m, \quad \text{y} \quad \varphi_i(U) \cap \varphi_j(U) = \emptyset, \quad i \neq j.$$

**Teorema 7. Hutchinson** Sean  $\{\varphi_1, \varphi_2, \dots, \varphi_m\}$  contracciones similares en  $\mathbf{X}$  que satisfacen la condición del conjunto abierto. Sean  $\alpha_1, \alpha_2, \dots, \alpha_m$  las constantes de Lipschitz de cada  $\varphi_i$ , y sea  $V \neq \emptyset$  el autosimilar correspondiente.

Entonces  $d = \dim_H(V)$  es el único número tal que

$$\alpha_1^d + \alpha_2^d + \dots + \alpha_m^d = 1$$

**Demostración:** Supongamos que el conjunto acotado  $U$  que satisface la condición del conjunto abierto contiene una bola cerrada de radio  $\alpha$  y a su vez está contenida en una bola de radio  $\beta$ . Establecemos

$$\gamma = \min_{1 \leq i \leq m} L(\varphi_i) \in (0, 1)$$

Por contradicción mostraremos que

$$H^s(V) \geq \left( \frac{\alpha\gamma}{2\beta + 1} \right)^s$$

Con esta desigualdad la prueba quedará completa, ya que  $\mathcal{H}^s(V) < |V|^s$  por el lema anterior, incluso, habremos mostrado que  $V$  es un  $s$ -conjunto.

Denotamos por  $\tau$  el valor del lado derecho de la desigualdad y supongamos que  $\mathcal{H}^s(V) < \tau$ . Entonces, para ciertas  $\delta$  pequeñas y una  $\delta$ -cubierta  $\{W_i\}$  de  $V$  la desigualdad

$$\sum_{i=1}^{\infty} |W_i|^s < \tau$$

se sigue. Aquí, podemos extender cada  $W_i$  ligeramente mientras conservemos la desigualdad, así suponemos que cada  $W_i$  es un conjunto abierto. Al ser  $V$  compacto, un número finito de conjuntos abiertos  $\{W_i\}$  cubre  $V$ .  $\{W_i\}$  cubre  $V$ . Por lo tanto, podemos suponer que  $V$  tiene una cubierta abierta, finita y acotada  $\{W_i\}$ , tal que para algún natural  $N$  tenemos

$$V \subset \bigcup_{i=1}^N W_i \quad \text{y} \quad \sum_{i=1}^N |W_i|^s < \tau \tag{4.1}$$

Para cada  $k \geq 1$  y un conjunto arbitrario  $A \in \mathbb{R}^n$ , sea

$$\mu(A) = \sum \prod_{j=1}^k (L(\psi_{ij}))^s$$

donde la suma corre a través de cada sucesión finita  $i_1 i_2 \dots i_k$  la cual satisface la condición  $A \cap \bar{U}_{i_1 i_2 \dots i_k} \neq \emptyset$ . Con  $A$  fijo, la sucesión  $\{\mu_k(A)\}$  es monótona decreciente, ya que

$$U_{i_1 i_2 \dots i_j} \subset U_{i_1 i_2 \dots i_k}, \quad 1 \leq j \leq m, \quad \text{y} \quad \sum_{j=1}^m (L(\psi_j))^s = 1(*).$$

Esto es, el límite

$$\mu(A) = \lim_{k \rightarrow \infty} \mu_k(A)$$

existe. En particular, la inclusión  $V \subset \bar{U}$  implica que  $\mu(V) = 1$ . También tenemos las propiedades:

1. Si  $A \subset B$  entonces  $\mu(A) \leq \mu(B)$ .
2.  $\mu(A \cup B) \leq \mu(A) + \mu(B)$ .

Y usando esas proposiciones obtenemos las siguientes desigualdades,

$$1 = \mu(V) \leq \mu\left(\bigcup_{j=1}^N W_j\right) \leq \sum_{j=1}^N \mu(W_j).$$

Lo siguiente es evaluar  $\mu(W_j)$ . Como hemos supuesto que  $|W_j|$  es menor que  $\delta$ , existe al menos un número natural  $k \geq 2$  para el cual las desigualdades

$$L(\psi_{i_j}) \dots L(\psi_{i_{k-1}}) > |W_j| > L(\psi_{i_j}) \dots L(\psi_{i_k})$$

se siguen. Sea  $A_j$  la familia de sucesiones  $\{i_1 \dots i_k\}$  la cual satisface las relaciones mencionadas anteriormente. La longitud de las sucesiones en  $A_j$  no es uniforme en general. Sea  $l$  el valor máximo de la longitud de estas sucesiones. Entonces, por (\*), tenemos

$$\mu_l(W_j) \leq \sum (L(\psi_{i_1}) \dots L(\psi_{i_k}))^s$$

donde la suma va sobre  $i_1 \dots i_k$  con  $W_j \cap \bar{U}_{i_1 \dots i_k} \neq \emptyset$ . Esto es porque la condición  $W_j \cap \bar{U}_{i_1 \dots i_k \dots i_l} \neq \emptyset$  implica que  $W_j \cap \bar{U}_{i_1 \dots i_k} \neq \emptyset$ . Por lo tanto, de aquí se sigue que

$$\mu(W_j) \leq \mu_l(W_j) \leq p|W_j|^s,$$

donde  $p$  es el número de sucesiones en  $\Lambda_j$  con  $W_j \cap \bar{U}_{i_1 \dots i_k} \neq \emptyset$ .

Si establecemos

$$r = L(\psi_{i_1}) \dots L(\psi_{i_k})\alpha \text{ y } R = L(\psi_{i_1}) \dots L(\psi_{i_k})\beta;$$

entonces cada  $\psi_i$  es una contracción similar, podemos ver que  $U_{i_1 \dots i_k}$  contiene una bola cerrada de radio  $r$  y a su vez está contenido en una bola cerrada de radio  $R$ . Si  $x$  es un

punto arbitrario de  $W_j$ , la bola cerrada de radio  $|W_j|$  centrada en  $x$  contiene a  $W_j$ , y así  $U_{i_1 \dots i_k}$  está contenido en alguna de las bolas cerradas de radio  $2R + |W_j|$ . También tenemos que

$$r \geq \alpha\gamma L(\varphi_i) \dots L(\varphi_{k-1}) > \alpha\gamma|W_j| \text{ y } R \leq \beta|W_j|;$$

por lo tanto, resulta que  $U_{i_1 \dots i_k}$  se encuentra en alguna de las bolas cerradas de radio  $\alpha\gamma|W_j|$ . Recordemos que las bolas de radio  $(2\beta + 1)|W_j|$  no dependen de las sucesiones en  $\Lambda_j$ . Por otra parte, los  $U_{i_1 \dots i_k}$  son mutuamente disjuntos, y comparando su volúmen  $n$ -dimensional, obtenemos

$$p(\alpha\gamma|W_j|)^n \leq ((2\beta + 1)|W_j|)^n,$$

o equivalentemente

$$p \leq \left(\frac{2\beta + 1}{\alpha\gamma}\right)^n = \tau^{-1}$$

lo que implica que  $\mu(W_j) \leq \tau^{-1}|W_j|^s$ , con la cual tenemos

$$1 \leq \sum_{j=1}^N \mu(W_j) \leq \tau^{-1} \sum_{j=1}^N |W_j|^s;$$

lo que contradice (3.1).

□

## 4.5. Ejemplos

Para concluir con este capítulo, exhibiremos algunos ejemplos de fractales que son auto-similares, a los cuales también les calcularemos su dimensión de Hausdorff haciendo uso del teorema de Hutchinson.

Como primer ejemplo de un fractal autosimilar tenemos al árbol de Hata (Figura 4.5), el cual lo denotaremos por  $\mathcal{A}$  y queda determinado por las siguientes dos contracciones en  $\mathbb{C}$ ,

$$\varphi_1(z) = \omega\bar{z}, \quad \varphi_2(z) = \frac{2\bar{z}+1}{3}$$

De las cuales podemos obtener

$$\mathcal{A} = \varphi_1(\mathcal{A}) \cup \varphi_2(\mathcal{A})$$

y

$$\text{Lip}(\varphi_1) = \frac{1}{\sqrt{3}} \text{ y } \text{Lip}(\varphi_2) = \frac{2}{3}$$

Así, por el teorema de Hutchison tenemos que si  $d = \dim_H(\mathcal{A})$ , entonces

$$\left(\frac{1}{\sqrt{3}}\right)^d + \left(\frac{2}{3}\right)^d = 1$$

Con lo que obtenemos que  $d \approx 1.43655$ .

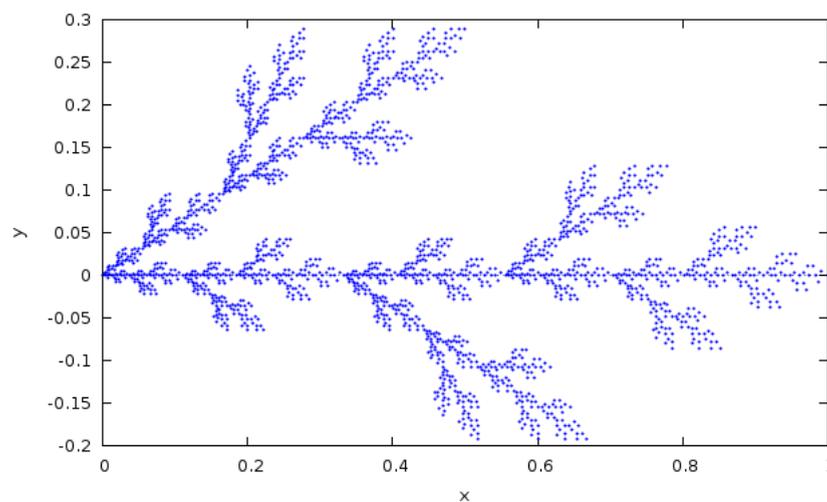


FIGURA 4.5: Árbol de Hata.

Otro conjunto autosimilar es la procesión de los cangrejos (Figura 4.6), el cual denotaremos por  $\mathcal{P}$  se define a partir de las siguientes contracciones

$$\varphi_1(z) = \frac{1+i}{2}z \quad \varphi_2(z) = \frac{1-i}{2}\bar{z}$$

Para las cuales

$$\mathcal{P} = \varphi_1(\mathcal{P}) \cup \varphi_2(\mathcal{P})$$

y

$$Lip(\varphi_1) = \frac{\sqrt{2}}{2} \text{ y } Lip(\varphi_2) = \frac{\sqrt{2}}{2}$$

Sea  $d = \dim_H(\mathcal{P})$ ,

$$\left(\frac{\sqrt{2}}{2}\right)^d + \left(\frac{\sqrt{2}}{2}\right)^d = 1$$

Con lo que tenemos que  $d = 2$ .

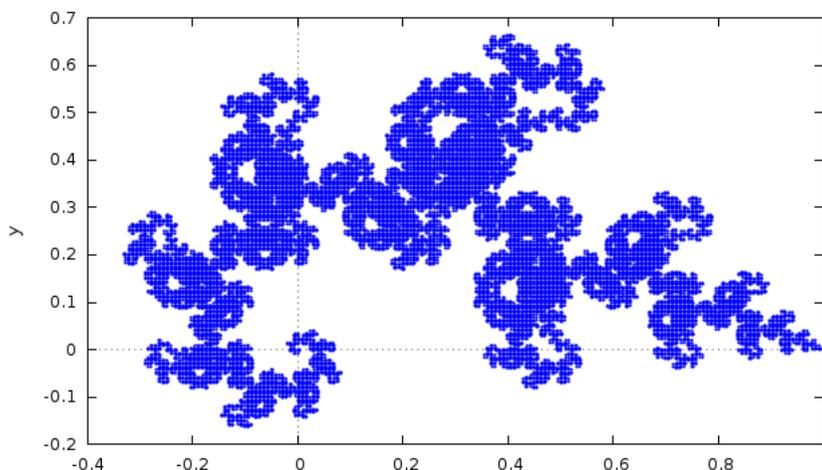


FIGURA 4.6: Procesión de los cangrejos.

El siguiente ejemplo es la curva de Lévy ( $\mathcal{L}$ ) (Figura 4.7), la cual es conjunto atractor de las siguientes contracciones

$$\varphi_1(z) = \frac{1+i}{2}z \quad \varphi_2(z) = \frac{1-i}{2}z$$

Para las cuales

$$\mathcal{L} = \varphi_1 \cup \varphi_2(\mathcal{L})$$

y

$$\text{Lip}(\varphi_1) = \frac{\sqrt{2}}{2} \text{ y } \text{Lip}(\varphi_2) = \frac{\sqrt{2}}{2}$$

Sea  $d = \dim_H(\mathcal{L})$ ,

$$\left(\frac{\sqrt{2}}{2}\right)^d + \left(\frac{\sqrt{2}}{2}\right)^d = 1$$

Con lo que tenemos que  $d = 2$ .

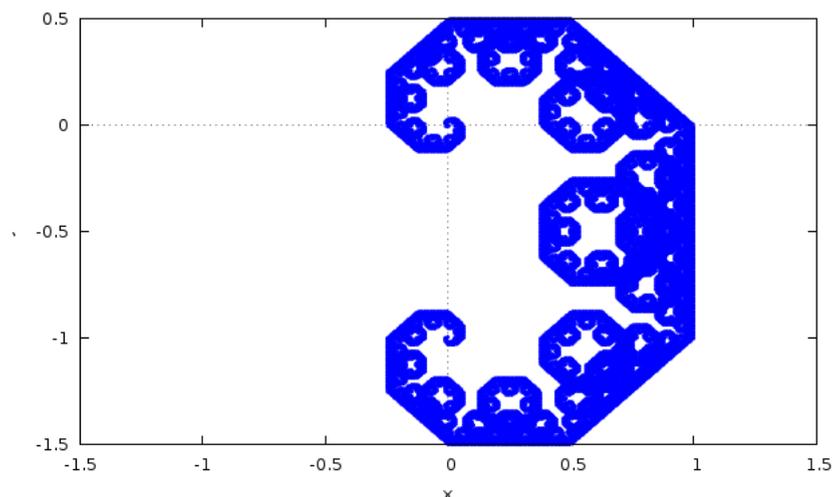


FIGURA 4.7: Curva de Lévy.

Otro fractal autosimilar es el conocido como la curva del dragón ( $\mathcal{D}$ ) (Figura 4.8), el cual es conjunto atractor de las siguientes contracciones

$$\varphi_1(z) = \frac{1+i}{2}z \quad \varphi_2(z) = \frac{1+i}{2}z$$

Para las cuales

$$\mathcal{D} = \varphi_1 \cup \varphi_2(\mathcal{D})$$

y

$$Lip(\varphi_1) = \frac{\sqrt{2}}{2} \text{ y } Lip(\varphi_2) = \frac{\sqrt{2}}{2}$$

Sea  $d = \dim_H(\mathcal{D})$ ,

$$\left(\frac{\sqrt{2}}{2}\right)^d + \left(\frac{\sqrt{2}}{2}\right)^d = 1$$

Con lo que tenemos que  $d = 2$ .

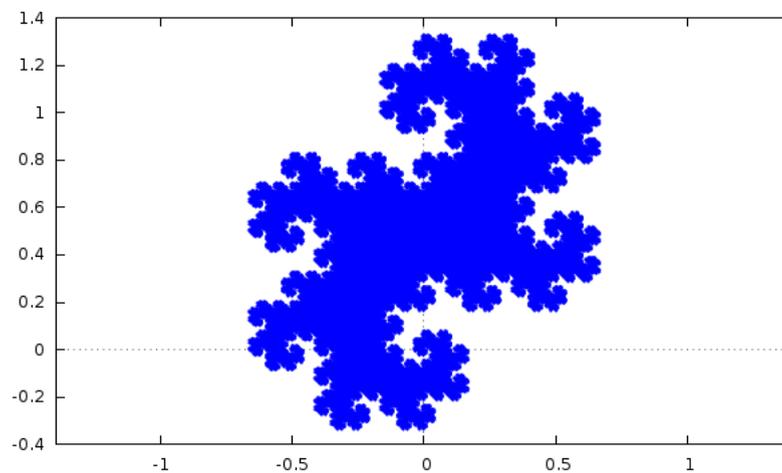


FIGURA 4.8: Curva del dragón.

El triángulo de Sierpinski ( $\mathcal{T}$ ) (Figura 4.9) es el conjunto atractor de las siguientes contracciones

$$\varphi_1(z) = \frac{1}{2}(z + p_1) \quad \varphi_2(z) = \frac{1}{2}(z + p_2) \quad \varphi_3(z) = \frac{1}{2}(z + p_3)$$

donde  $p_1, p_2, p_3$  son los vértices de un triángulo y para las cuales

$$\mathcal{T} = \bigcup_{i=1}^3 \varphi_i(\mathcal{T})$$

con

$$\text{Lip}(\varphi_1) = \frac{1}{2}, \text{Lip}(\varphi_2) = \frac{1}{2}, \text{Lip}(\varphi_3) = \frac{1}{2}$$

Sea  $d = \dim_H(\mathcal{T})$ ,

$$\left(\frac{1}{2}\right)^d + \left(\frac{1}{2}\right)^d + \left(\frac{1}{2}\right)^d = 1$$

Con lo que tenemos que  $d = \frac{\log(3)}{\log(2)}$ .

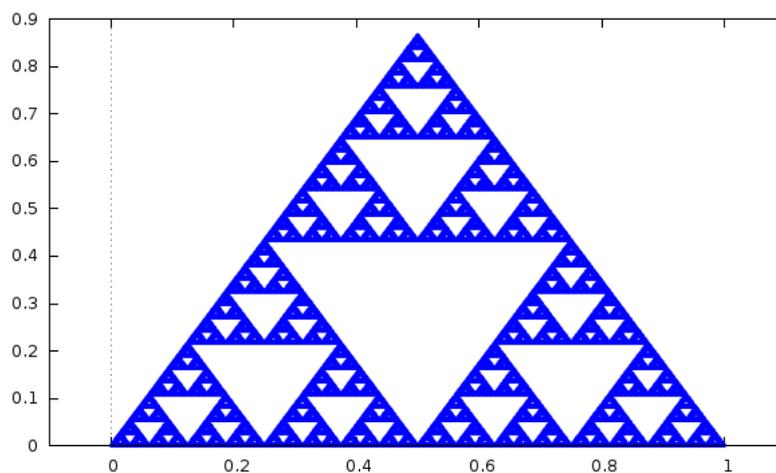


FIGURA 4.9: Triángulo de Sierpinski.

# Capítulo 5

## DNA y Sierpínski.

El lenguaje de la herencia está escrito en un reducido alfabeto de cuatro letras (A, C, G, T), es decir, todo el caudal de información hereditaria viene determinado por las diferentes disposiciones de los cuatro tipos de nucleótidos.

En este capítulo usaremos esta idea para tratar al DNA como un conjunto de cuatro elementos, por lo que podemos tratarlo matemáticamente. Primeramente introduciremos los operadores lógicos binarios, en específico, hablaremos de la *función XOR* que será la que nos permitirá hacer la conexión entre el DNA y los fractales. Posteriormente, explicaremos el algoritmo para obtener un tetraedro de Sierpínski a partir de una cadena de DNA, para después observar las diferencias entre una cadena aleatoria y una cadena que pertenece a un organismo. Finalmente, tomaremos la proyección de los tetraedros de Sierpínski de diferentes seres vivos.

### 5.1. Funciones lógicas

El volumen informativo de cada mensaje se suele representar mediante bits (que es la contracción de *binary digits*), y el DNA no es la excepción. Toda pregunta puede ser formulada de tal manera que sea susceptible de ser respondida por un "no." por un "sí", es decir, 0 o 1.

Pensemos en lo siguiente: al representar al DNA con solo dos símbolos, en lugar de cuatro, el número de bits contenido en una molécula de DNA equivaldría al doble de pares de nucleótidos. A su vez, existen cuatro clases distintas de nucleótidos, por lo que el número

de bits contenidos es el cuádruple del número de pares de nucleótidos. Para darnos una idea, el genoma humano contiene  $3 \times 10^9$  pares de bases, por lo que contiene  $1.2 \times 10^{10}$  bits de información.

Esta información codificada en ceros y unos puede ser tratada mediante operadores lógicos. Un operador lógico asigna un valor, ya sea cero o uno, a la combinación de condiciones entre las variables, en este caso, la información codificada. En esta sección introduciremos tres operadores lógicos: NOT, AND y OR, y en la siguiente trataremos más detalladamente al operador XOR.

## Función NOT

La operación NOT consiste en negar el estado de una variable, es decir, cambiar el cero por el uno y viceversa. Así, la función NOT es una función de una sola variable que invierte el valor de esta y cuya tabla de verdad es la correspondiente a la Figura 4.1.

A	!A
0	1
1	0
0	1
1	0

FIGURA 5.1: Tabla de verdad de NOT

## Función AND

**Definición 5.1.** *La conjunción o “AND” es un operador lógico que toma el valor verdadero si las condiciones son verdaderas y es falsa cuando al menos una de las condiciones es falsa.*

Lo denotaremos por  $\wedge$ .

Podemos observar que la función AND es una multiplicación lógica, ya que como podemos observar en la tabla, el resultado consiste de la multiplicación de las dos variables.

A	B	$A \wedge B$
0	0	0
0	1	0
1	0	0
1	1	1

FIGURA 5.2: Tabla de verdad de AND

## Función OR

**Definición 5.2.** *La disyunción inclusiva o “OR” es un operador lógico que toma el valor verdadero al menos una de las condiciones es verdadera y es falso solamente si todas las condiciones lo son.*

Lo denotaremos por  $\vee$ .

A	B	$A \vee B$
0	0	0
0	1	1
1	0	1
1	1	1

FIGURA 5.3: Tabla de verdad de OR

A continuación introduciremos la función lógica con la que trabajaremos la información de una cadena de DNA.

## 5.2. Función XOR

En el capítulo 1 mostramos como toda cadena de DNA puede ser codificada mediante una sucesión de ceros y unos de tres maneras distintas, además, como veremos, la función XOR determina la tercera forma en que una secuencia de DNA está en función de las otras dos.

**Definición 5.3.** *La disyunción exclusiva o “XOR” es un operador lógico que toma el valor verdadero si exactamente una de las dos condiciones es verdadera. Este operador es idéntico a la no equivalencia.*

Lo denotaremos por  $\oplus$  y puede ser visualizado utilizando los operadores AND y OR de la siguiente manera:

$$\begin{aligned} A \oplus B &= (A \wedge !B) \vee (!A \wedge B) \\ &= (A \wedge B) \vee (!A \wedge !B) \end{aligned}$$

La tabla de verdad del operador XOR se muestra en la Figura 4.4.

A	B	$A \oplus B$
0	0	0
0	1	1
1	0	1
1	1	0

FIGURA 5.4: Tabla de verdad de XOR

Para relacionar el DNA con estos operadores, consideramos un segmento de DNA constituido por nucleótidos y lo interpretamos como una sucesión de elementos que pueden tomar valores en el alfabeto mencionado.

Por ejemplo, supongamos que la siguiente sucesión representa un segmento de DNA del algún organismo:

$$S_n = \text{ATCGTTACCGGCAT} \dots$$

Transformamos  $S_n$  en una sucesión binaria mediante el criterios  $WS$ ,  $YR$  y  $MK$ . Recordemos que bajo el criterio  $WS$  a C y G les corresponde el valor 1, mientras que a A y T el valor 0; el criterio  $YR$  asigna 1 a A, G y 0 a C y T; por último, el criterio  $MK$  asigna 1 a A, C y 0 a G y T. Así, obtenemos la siguiente tabla:

Criterio	A	T	C	G	T	T	A	C	C	G	G	C	A	T	...
$WS$	0	0	1	1	0	0	0	1	1	1	1	0	0	0	...
$YR$	1	0	0	1	0	0	1	0	0	1	1	0	1	0	...
$MK$	1	0	1	0	0	0	1	1	1	0	0	0	1	0	...

Ahora que hemos encontrado la manera de representar al DNA como una sucesión de 0 y 1. De la anterior tabla podemos hacer una comparación entre las filas y notar que al

sumar dos de ellas en base dos, se obtiene la tercera fila, justo como en el criterio XOR.

Adicionalmente, partir de esta representación podemos hacer una identificación de una secuencia de DNA con elementos de  $\mathbb{R}^3$ .

El genoma de un organismo contiene un gran número de nucleótidos, por lo que hacemos una partición del genoma en unidades más pequeñas. Por ejemplo, el genoma humano contiene 3200 millones de pares de nucleótidos, así que para trabajar con él podríamos utilizar unidades de 10 nucleótidos y reducir un orden de magnitud. Como decidimos usar unidades de 10 nucleótidos, lo que implica que nuestra sucesión de ceros y unos tendrá 320 millones de segmentos conformados por 10 elementos.

Identificamos la sucesión determinada por cada criterio con un número en el intervalo  $[0,1]$ , donde la cantidad de cifras después del punto decimal está dada por la cantidad de elementos que decidimos que tendrá cada segmento. En el caso del ejemplo del genoma humano, las coordenadas del elemento en  $\mathbb{R}^3$  tendrían 10 cifras decimales cada una.

Para visualizar esto mejor, tomemos el segmento ATCGTTACCGGCAT al cual previamente le asignamos sus representaciones, entonces, sus valores según cada criterio son,

Criterio	ATCGTTACCGGCAT
<i>WS</i>	0.00110001111000
<i>YR</i>	0.10010010011010
<i>MK</i>	0.10100011100010

y el punto que le corresponde en  $\mathbb{R}^3$  es

$$(0.00110001111000, 0.10010010011010, 0.10100011100010)$$

Sin embargo, esta representación está en base dos, y lo deseable es expresarla en base diez. Observe que en la representación *WS* la primera cifra decimal es 0, lo que quiere decir que el valor que le corresponde al segmento bajo este criterio es menor que 0.5. Bajo los otros dos criterios, tenemos que el primer decimal es 1, por lo tanto, el valor que le corresponde bajo *YR* y *MK* es mayor que 0.5.

Podemos pensar en que estamos considerando una partición en octantes del cubo unitario, ya que cada coordenada de cada punto en  $\mathbb{R}^3$  está contenida en  $[0,1]$  y cada criterio

corresponde a un eje. Seleccionamos el octante correspondiente a la primera cifra decimal (Figura 5.5). Notemos que solamente cuatro de los ocho octantes corresponden a zonas donde se pueden los puntos correspondientes a segmentos, ya que los valores que toma DNA bajo  $WS$  y  $YR$  determinan los octantes donde podremos encontrar los puntos correspondientes.

Similarmente, el octante seleccionado será dividido en ocho partes, y de la misma manera, solo cuatro de ellos corresponden a zonas donde es posible encontrar puntos. Así podemos encontrar los puntos de la secuencia de DNA en  $\mathbb{R}^3$  de una manera intuitiva.

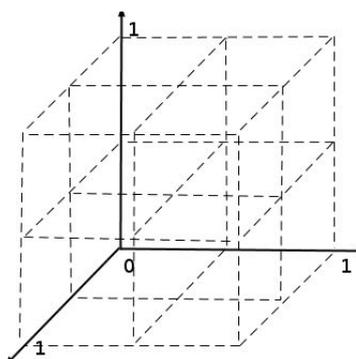


FIGURA 5.5: Partición del cubo unitario en octantes.

Formalmente, lo que haremos es hacer una partición de la secuencia del DNA de un organismo en segmentos de longitud fija y asignarles tres números reales a través de los criterios  $WS$ ,  $YR$  y  $MK$ . Así, tomamos como dominio el conjunto de los puntos en  $[0,1]$  determinados por los criterios  $WS$ ,  $YR$  y codominio el conjunto de los puntos en el mismo intervalo pero que están determinados por el criterio  $MK$ . De esa manera, consideramos  $f : [0, 1] \times [0, 1] \rightarrow [0, 1]$ , donde  $f$  es la función XOR.

Debido a la definición de la función, existen zonas restringidas donde se puede encontrar la imagen de nuestro punto. Por ejemplo, sabemos que en la zona correspondiente a la segunda mitad de los intervalos no se encuentra la imagen del punto, puesto que la función asigna la primera mitad del intervalo en el eje  $z$  cuando al punto le corresponde la segunda parte en los otros dos ejes.

Notemos que al aplicar la función, obtenemos copias autosimilares de la figura en una escala más pequeña, en particular, la gráfica de la función corresponde a un Tetrahedro de Sierpinski. Vale la pena mencionar que la función XOR corresponde a la regla 60 del

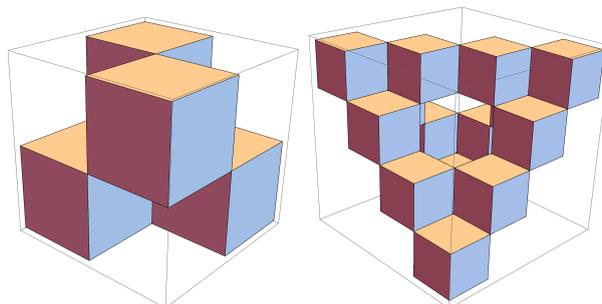


FIGURA 5.6: A la izquierda, octantes permitidos al realizar la primera partición, a la derecha, zonas permitidas después de la segunda partición.

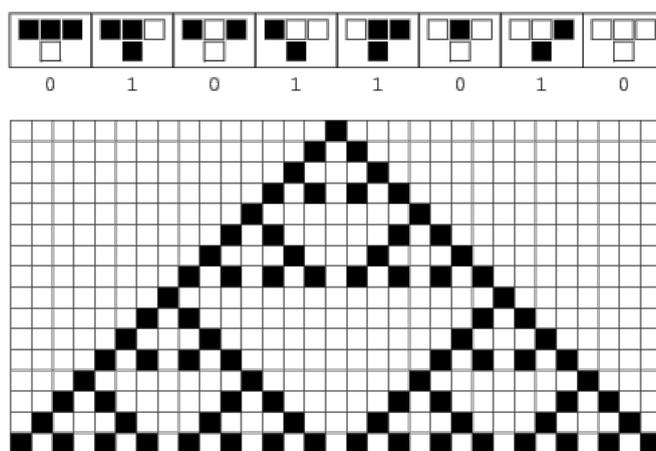


FIGURA 5.7: En la parte superior se muestra la regla 90 del autómata celular elemental, en la parte inferior el Triángulo de Sierpinski generado por la misma [23].

autómata celular elemental (Figura 5.7), a la cual le corresponde el triángulo de Sierpinski.

Ahora aplicaremos el algoritmo visto en la cepa K-12 de la *Escherichia coli*<sup>1</sup>, una entera bacteria procariota cuyo único cromosoma circular tiene 4 639 221 pares de bases nitrogenadas de donde consideramos 15 000 pares de bases del gen *yoaD*<sup>2</sup>, asociado a la regulación de la celulosa. Al implementar la función en *GNU Octave*, se observa el Tetraedro descrito por el proceso mencionado (Figura 5.8).

Además, generamos también una sucesión aleatoria con una longitud de 15000 elementos utilizando unidades de 4 elementos, representando así un segmento de DNA. Para ello,

<sup>1</sup>La *Escherichia coli* es la bacteria más estudiada por la humanidad.

<sup>2</sup>Los datos fueron tomados de [19]

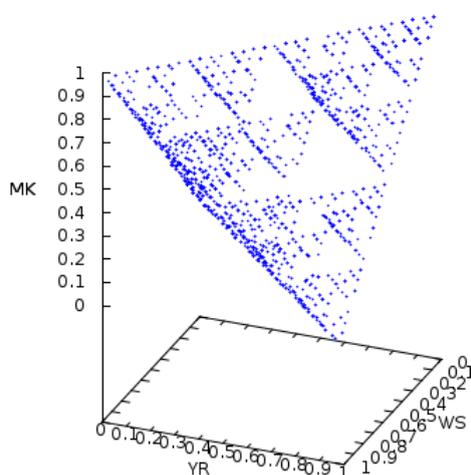


FIGURA 5.8: Tetraedro obtenido a partir del gen *yoaD* de la *E. coli*.

consideramos una distribución binomial con parámetros  $n = 4$  y  $p = 0.25$ , de la que se obtuvo el tetraedro de la Figura 5.8.

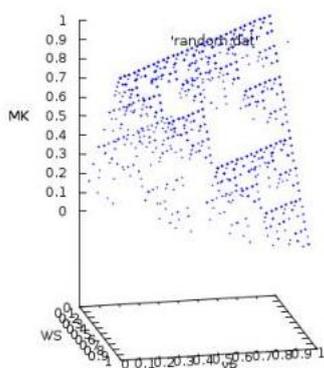


FIGURA 5.9: Tetraedro generado a partir de una distribución binomial.

Podemos ver que los tetraedros generados por la sucesión correspondiente a la *E. coli* y a la sucesión aleatoria son similares, entonces nos cuestionamos acerca de la presencia del azar en el orden en que aparecen las bases nitrogenadas en la secuencia del DNA de un ser vivo. Para ver la diferencia entre los datos de la *E. Coli* y los generados aleatoriamente, nos fijamos en la proyección de los tetraedros en los ejes *WS* y *YR*.

Notemos que en la proyección de los datos de la *E. Coli* podemos encontrar que hay una mayor concentración de puntos en ciertas zonas, mientras que evaden otras. En el

caso de los datos generados por la distribución binomial, observamos que el patrón formado por la esta no es similar al que se obtuvo del organismo.

Podemos pensar en esto como una especie de firma genómica, que está caracterizada por la densidad de los iterados, es decir, de las zonas de mayor concentración. En las siguientes imágenes vemos las proyecciones de los Tetraedros de Sierpinski de diferentes especies.

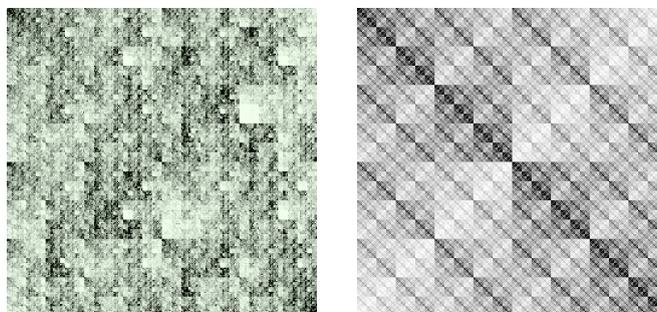


FIGURA 5.10: A la izquierda proyección de los datos de la *E. Coli*, a la derecha proyección de los datos aleatorios.

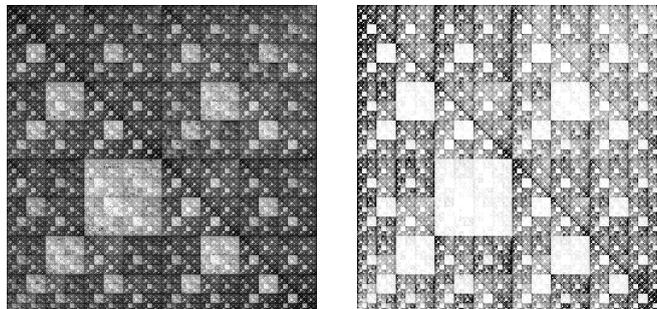


FIGURA 5.11: A la izquierda, proyección formada con un genoma humano. A la derecha, formada con el genoma de una rata.

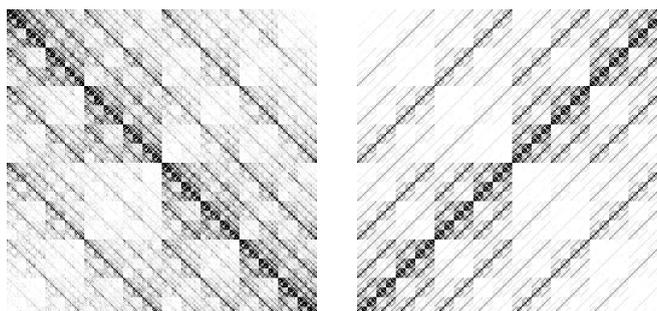


FIGURA 5.12: A la izquierda vemos la proyección del genoma de la rickettsia, a la derecha, la proyección de datos aleatorios.

### 5.3. El Tetraedro de Sierpinski

En esta última sección hablaremos del tetraedro de Sierpinski, el fractal obtenido a partir del algoritmo implementado.

Definiremos el tetraedro de Sierpinski como un conjunto autosimilar de  $\mathbb{R}^3$ . Sean

$$p_1 = (0, 0, 0)$$

$$p_2 = (1, 1, 0)$$

$$p_3 = (1, 0, 1)$$

$$p_4 = (0, 1, 1)$$

y para  $1 \leq i \leq 4$  introducimos las funciones  $f_i : [0, 1] \times [0, 1] \rightarrow [0, 1]$  tales que

$$f_i(x) = \frac{1}{2}(x + p_i)$$

Estas funciones son similaridades con  $Lip(f_i) = \frac{1}{2}$  y por el Teorema 6, definen un único conjunto autosimilar.

**Definición 5.4.** *Sea  $\Phi$  la función*

$$\Phi(X) = \bigcup_{i=1}^4 f_i(X)$$

donde  $X$  es un subconjunto de  $\mathbb{R}^3$ . Entonces definimos  $S$  como el único conjunto que satisface:

$$\Phi(S) = S$$

Llamamos a  $S$  el Tetraedro de Sierpinski.

**Proposición 5.5.** *La dimensión de Hausdorff del Tetraedro de Sierpinski es 2.*

**Demostración:** De cada  $f_i$ , tenemos que el conjunto de contracciones similares es  $\{\frac{1}{2}\}$  y por el Teorema de Hutchinson (Teorema 7), la dimensión de similaridad  $d$  satisface la siguiente igualdad,

$$\frac{1}{2^d} + \frac{1}{2^d} + \frac{1}{2^d} + \frac{1}{2^d} = 1$$

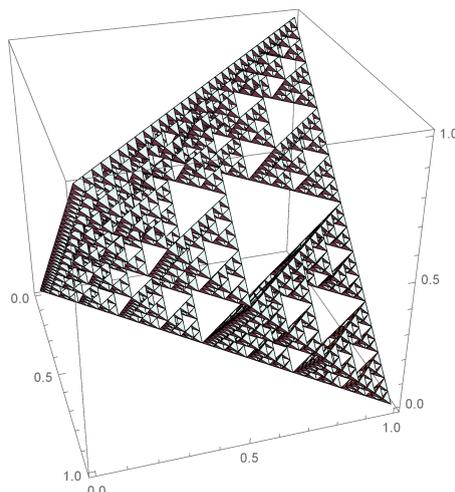


FIGURA 5.13: Tetraedro de Sierpinski.

Así, obtenemos que la igualdad se satisface para  $d = 2$ , y como el Tetraedro de Sierpinski es un fractal autosimilar, su dimensión de Hausdorff y de similaridad coinciden. Por lo tanto,  $\dim_H(S) = 2$

□

**Definición 5.6.** Sea  $m$  un entero no negativo. Llamamos  $S_0$  al tetraedro inicial formado por los vértices  $\{p_1, p_2, p_3, p_4\}$ . Análogamente,

$$S_{m+1} = \Phi(S_m)$$

Extendemos la definición anterior al no solo considerar el tetraedro formado por  $\{p_1, p_2, p_3, p_4\}$ . En lugar de considerar  $S_0$  como dicho tetraedro, consideramos  $S_0$  como el cubo unitario, y al aplicarle la función  $\Phi$ , el resultado seguirá siendo el Tetraedro de Sierpinski (Figura 5.14).

Además, para el Tetraedro de Sierpinski tenemos las siguientes propiedades:

1. El Tetraedro de Sierpinski es un conjunto compacto.
2. El Tetraedro de Sierpinski es conexo.
3. El Tetraedro de Sierpinski es conexo por trayectorias.

El inciso 1) se tiene a partir de la definición del tetraedro de Sierpinski, para ver la demostración de 2) y 3) se sugiere revisar [6].

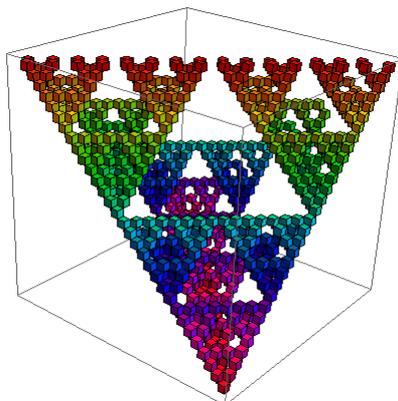


FIGURA 5.14: Teraedro formador por cubos.

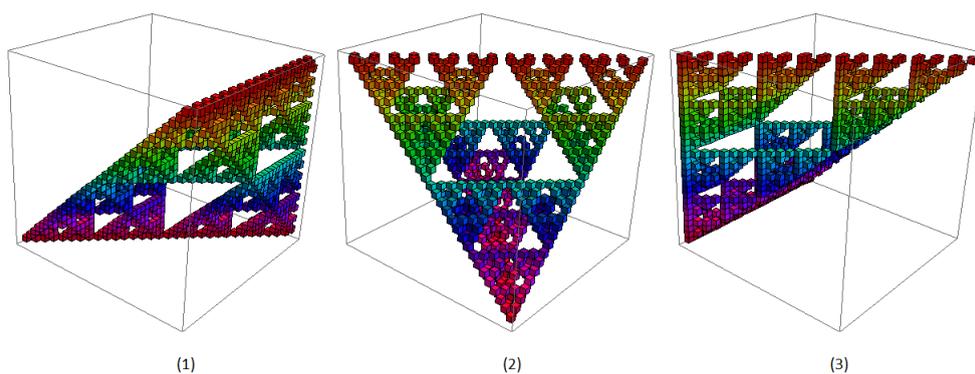


FIGURA 5.15: Imagen en  $\mathbb{R}^3$  de: (a) la función AND, (b) la función XOR, (c) la función OR.

Algo importante que no debemos olvidar, es que el tetraedro de Sierpinski obtenido a partir del DNA es una característica de la función XOR y de éste. En la primera sección introducimos las funciones lógicas y todas ellas tienen como imagen tetraedros de Sierpinski, sin embargo, las funciones AND y OR no representan alguna característica biológica de nuestro interés por el momento.

# Capítulo 6

## Conclusiones

Hoy en día se tiene una gran cantidad de información de DNA, sin embargo, aún no se conocen muchos aspectos de ella, por ejemplo, el término DNA no codificador se refiere a todo el DNA que no está involucrado en la codificación de un producto de mRNA maduro, el DNA no codificante es también llamado DNA basura. Puesto que no se tiene entendimiento de que es lo que hace, muchos suponen que simplemente no hace nada.

Con este trabajo de tesis no se tiene la intención de descubrir algo sobre eso, sin embargo, el ejemplo ilustra que aún queda mucho camino que recorrer para comprender todo el genoma de un organismo, y más aún, la semejanza y la diferencia entre los genomas de diferentes seres vivos.

La intención es que la matemática nos ayude al comprender los fenómenos biológicos. Podríamos preguntarnos acerca de la densidad de iterados que aparecen en las proyecciones de los tetraedros de Sierpínsiki: sobre si las zonas sin puntos corresponden a palabras prohibidas en el genoma, o sobre que es lo que origina las zonas más densas, también podemos cuestionarnos acerca de la similitud entre algunas proyecciones correspondientes a diferentes especies. Esto abre la posibilidad a un trabajo futuro para responder a estas cuestiones.

# Bibliografía

- [1] BAK, P., *How nature works: the science of self-organized criticality*, Springer, 1996.
- [2] BRESLAUER, K., FRANK, R., ET AL, *Predicting DNA duplex stability from the base sequence*, Proc.Natl. Acad. Sci. USA 83, 1986.
- [3] CARREÓN, G., *El juego circular del caos en el ADN y compresion fractal de imágenes*, Tesis de licenciatura, Universidad Nacional Autónoma de México, 2004.
- [4] FALCONER, K., *Fractal geometry: mathematical foundatios and applications*, John Wiley e hijos. 1990.
- [5] GARCÍA-COLÍN, L., DAGDUG, L., PICQUART, M., VÁZQUE, E. COORDINADORES, Introducción a la física biológica 1, Colegio Nacional, México 2010. *Fractal geometry: mathematical foundatios and applications*, John Wiley e hijos. 1990.
- [6] HATTORI, K., HATTORI, T., KUSUOKA, S. , *Self-avoiding Paths on the Three Dimensional Sierpinski Gasket*, Publ. RIMS, Kyoto Univ., 1993.
- [7] KIGAMI, J., MASAYA, Y., MASAYOSHI, H. , *Mathematics of fractals*, Translations of Mathematical Monographs, 167, American Mathematical Society, 1997.
- [8] MANDELBROT, M. , *La geometría fractal de la naturaleza*, Tusquets Editores, 2009.
- [9] MARQUEZ, T., *Una red neuronal de votación para análisis genómico*, Tesis de licenciatura, Universidad Nacional Autónoma de México, 2000.
- [10] MIRAMONTES P., COCHO, G., *DNA dimer correlations reflect in vivo conditions and discriminate among nearest-neighbor base pair free energy parameter measures*, Physisca A 321, 2003.
- [11] MIRAMONTES P., *Río de tiempo y agua. Procesos y estructura de la Ciencia en nuestros días.*, Facultad de Ciencias de la Universidad Nacional Autónoma de México, CopIt ArXives, 2010.

- 
- [12] NOGUERA, R., RUIZ, R., *El proyecto genoma humano*, Ciencias, 58, abril - junio 2000.
- [13] SAENZ, R., *Análisis Real: Primer Curso*, Universidad de Colima.
- [14] SAGAN, C. , *Los dragones del Edén. Especulaciones sobre la evolución de la inteligencia humana*, Grijalvo, S.A. de S.V., 1984.
- [15] SCHRÖDINGER, E. , *¿Qué es la vida?*, Tusquets, 2014.
- [16] SINDEN, R., *DNA Structure and Function*, Academic Press. 1994.
- [17] STEWART, I., *Mathematics of Life*, Basic Books, 2001.
- [18] DNA LEARNING CENTER, recuperado el 02 de marzo de 2015 en <http://www.dnai.org/timeline/>.
- [19] ECOGENE, recuperado el 26 de abril de 2015 en <http://www.ecogene.org/?q=gene/EG13516>.
- [20] LA WEB DEL ELECTROCARDIOGRAMA, recuperado el 29 de abril de 2015 en <http://www.my-ekg.com/generalidades-ekg/ondas-electrocardiograma.html>
- [21] SISTEMA DE DIAGNÓSTICO A DOMICILIO DE APNEA, recuperado el 29 de abril de 2015 en <http://catedratelefónica.ulpgc.es/blog/sistemadiagnostico domicilioapnea /2013/03/27/fundamentos-de-la-senal-electrocardiografica/>
- [22] NATURE EDUCATION, recuperado el 26 de febrero de 2015 en <http://www.nature.com/scitable/>.
- [23] WOLFRAM MATHWORLD, recuperado el 26 de abril de 2015 en <http://mathworld.wolfram.com/Rule90.html>.
- [24] COMPUTATIONAL LEGAL STUDIES, recuperado el 18 de mayo de 2015 en <http://computationallegalstudies.com/wp-content/uploads/2009/08/Picture-51.png>