

16



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO 2E

FACULTAD DE CIENCIAS

GEOMETRIA EN BIOLOGIA

(SIMULACION DEL CRECIMIENTO DE ALGUNOS TIPOS DE PLANTAS)

T E S I S  
QUE PARA OBTENER EL TITULO DE:  
M A T E M A T I C O  
P R E S E N T A:

VINICIO ANTONIO GOMEZ GUTIERREZ



MEXICO, D. F.

1995

FALLA DE ORIGEN



Universidad Nacional  
Autónoma de México



## **UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso**

### **DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**TESIS SIN PAGINACION**

**COMPLETA LA INFORMACION**

CIUDAD UNIVERSITARIA



UNIVERSIDAD NACIONAL  
AVENIDA DE  
MEXICO

FACULTAD DE CIENCIAS  
División de Estudios  
Profesionales  
Exp. Núm. 55

M. EN C. VIRGINIA ABRIN BATULE  
Jefe de la División de Estudios Profesionales  
Universidad Nacional Autónoma de México.  
P r e s e n t e .

Por medio de la presente, nos permitimos informar a Usted, que habiendo  
revisado el trabajo de tesis que realizó el pasante \_\_\_\_\_

Gómez Gutiérrez Vinicio Antonio

con número de cuenta 8635454-1 con el título: \_\_\_\_\_

"Geometría en Biología (Simulación del crecimiento de algunos tipos  
de plantas)"

Consideramos que reúne los méritos necesarios para que pueda conti-  
nuar el trámite de su Examen Profesional para obtener el título de -  
Matemático.

GRADO NOMBRE Y APELLIDOS COMPLETOS

FIRMA

M. en C. Guillermo Gómez Alcaraz

Director de Tesis

Actuario Javier Fernández García

Doctor Oscar Alfredo Palma Velasco

Bióloga Rosa María Bayona Cejls

Suplente

Doctora Ligia Collado Vides

Suplente

*RUBEN*  
*Ligia Collado Vides*

GEOMETRIA EN BIOLOGIA  
(SIMULACION DEL CRECIMIENTO DE ALGUNOS TIPOS DE PLANTAS)

INDICE GENERAL.

INTRODUCCION. ESTUDIO DE LA GEOMETRIA EN ALGUNAS PLANTAS.

CAPITULO I. SISTEMAS LINDENMAYER Y SISTEMAS DE FUNCIONES ITERADAS

- A) QUE SON LOS SISTEMAS LINDENMAYER
- B) APLICACIONES
- C) QUE SON LOS IFS
- D) RELACION ENTRE LOS IFS Y LOS SISTEMAS LINDENMAYER

CAPITULO II. REDES NEURONALES Y APRENDIZAJE.

- A) INTRODUCCION
- B) APRENDIZAJE

CAPITULO III. "APROXIMACION" DE FRACTALES.

- A) METODO DEL PUNTO LEJANO
- B) ESTABILIDAD DE UN IFS. CONCLUSIONES.

## INTRODUCCION.

En el estudio de la naturaleza nos encontramos con fenómenos de gran belleza y complejidad, uno de ellos es la forma en que se desarrollan las algas marinas. Consideremos el problema de hacer modelos que presenten un comportamiento que recoja las características fundamentales de este fenómeno. Podemos abordarlo desde diversas hipótesis explicativas en cuanto a qué rige fundamentalmente el crecimiento: desde el punto de vista de la genética, en el ADN ya está la información necesaria, otro enfoque es considerar que la influencia de la luz, especies depredadoras, los nutrientes y contaminantes en el medio ambiente tienen un papel decisivo.

En el presente trabajo vamos a estudiar dos formas de recrear la estructura geométrica de un alga: los sistemas Lindenmayer y los sistemas de funciones iteradas (IFS). Nos limitaremos a mostrar cómo pueden reproducir el modo de ramificación del alga *Bostrychia radicans*, así como algunas técnicas susceptibles de generalización, y algunas propiedades que vinculan a los sistemas Lindenmayer, con los IFS. Además estudiamos cómo es que el modelo depende de los parámetros utilizados.

En el estudio de los fractales puede darse el caso de que un IFS no corresponda exactamente al fractal que queremos. Con el propósito de contar con elementos que nos indiquen en qué dirección debemos hacer los cambios de los parámetros para ajustar un IFS, estudiaremos las redes neuronales. También se analizan algunas limitaciones que no se pueden superar solo cambiando los parámetros. En fin, sirva esta introducción para invitar al lector a enfrentarse a unas matemáticas poco conocidas, pero estrechamente relacionadas con temas tan actuales e interesantes como los fractales y la "complejidad" de los sistemas dinámicos discretos involucrados.

## CAPITULO I.

### SISTEMAS LINDENMAYER Y SISTEMAS DE FUNCIONES ITERADAS (IFS)

#### A) QUE SON LOS SISTEMAS LINDENMAYER

Los sistemas Lindenmayer [10] son modelos con los que podemos simular (imitar) el crecimiento de plantas haciendo énfasis en las relaciones entre las unidades de crecimiento y su proceso de ramificación. Podemos pensar a los sistemas Lindenmayer como sistemas dinámicos discretos definidos sobre conjuntos no vacíos de palabras. Supongamos que tenemos una palabra inicial (axioma) y que podemos transformar cada palabra (con ciertas reglas de transformación). Así podemos generar una sucesión de palabras. Esto sería un sistema de reescritura.

Los sistemas de reescritura pueden ser usados de dos maneras distintas. En la primera, el sistema de reescritura opera directamente sobre objetos bidimensionales tales como arreglos [4], gráficas [12] y formas [5]. En la segunda las cadenas gramaticales (en sentido amplio, esto es, incluyendo el sistema de reescritura paralela) se utilizan para definir cadenas de símbolos. En esta sección se describe un método para generar imágenes, basado en este segundo enfoque. Trataremos de que el modelo matemático resultante pueda aplicarse para crear una gama de curvas fractales (a través de aproximaciones) que parten desde ejemplos clásicos de principios de este siglo, hasta figuras que reflejen la estructura geométrica de ciertas plantas.

Un ejemplo de un sistema de reescritura sería la *curva de Koch* o *copo de nieve*. En ésta el objeto inicial es un triángulo equilátero, y la regla es que a cada segmento lo dividimos en tres, le quitamos la parte del medio y le ponemos en el hueco un pico formado por dos segmentos, cada uno de la misma longitud que la parte que quitamos.

Sean  $s$  un segmento,  $\alpha$  un ángulo de  $\pi/3$  en sentido positivo y  $\beta$  el mismo ángulo en sentido negativo. Así podemos representar al triángulo original por la palabra generadora

$saasaaas$

Definamos la función

$F: \{s, \alpha, \beta\} \rightarrow \{\text{palabras formadas con } s, \alpha \text{ y } \beta\}$   
 como sigue:

$$F(s) = \left(\frac{s}{3}\right)\beta\left(\frac{s}{3}\right)\alpha\left(\frac{s}{3}\right)\beta\left(\frac{s}{3}\right)$$

$$F(\alpha) = \alpha$$

$$F(\beta) = \beta$$

Partiendo de  $saasaaas$ , la reescritura sería:

$$\left(\frac{s}{3}\right)\beta\left(\frac{s}{3}\right)\alpha\left(\frac{s}{3}\right)\beta\left(\frac{s}{3}\right)\alpha\left(\frac{s}{3}\right)\beta\left(\frac{s}{3}\right)\alpha\left(\frac{s}{3}\right)\beta\left(\frac{s}{3}\right)\alpha\left(\frac{s}{3}\right)\beta\left(\frac{s}{3}\right)\alpha\left(\frac{s}{3}\right)\beta\left(\frac{s}{3}\right)\alpha\left(\frac{s}{3}\right)\beta\left(\frac{s}{3}\right)$$

y correspondería a las figuras 1 y 2:



Fig. 1

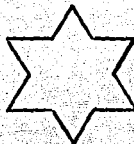


Fig. 2

Podemos aplicar otra vez la función de reescritura, generando otra figura, y crear así una sucesión de curvas cuyo "límite" es la curva de Koch. Una posible definición recursiva sería definir  $A_0 = X_0 \cup Y_0 \cup Z_0$  donde  $X_0$  es el lado de abajo del triángulo,  $Y_0$  el lado derecho y  $Z_0$  el lado izquierdo; si cada uno tiene longitud  $s$  y le aplicamos la función  $F$  definida anteriormente, podemos hacer que  $A_{n+1} = X_{n+1} \cup Y_{n+1} \cup Z_{n+1}$  donde  $X_{n+1} = F(X_n)$ ,  $Y_{n+1} = F(Y_n)$  y  $Z_{n+1} = F(Z_n)$



La historia explícitamente reconocida de los sistemas Lindenmayer se remonta a unos cuantos años, luego de que Noam Chomsky estableciera el concepto de gramática con estructura de frases. En los trabajos [11] y [8] se propone describir figuras usando lenguajes formales (cadenas).

La descripción del crecimiento de organismos vivos es introducida por Aristid Lindenmayer [9] a través de sistemas de reescritura en paralelo (como sucede en la división celular). La teoría este tipo de sistemas ha continuado desarrollándose [10].

Ahora daremos una definición más formal de los sistemas Lindenmayer.

Consideremos un conjunto  $V$  formado por símbolos susceptibles de interpretación geométrica o biológica, nos referiremos a este conjunto como *alfabeto*. Sea  $V^*$  el conjunto de todas las palabras finitas no vacías formadas por elementos de  $V$ .

Un *sistema Lindenmayer* de cadenas es una terna ordenada  $\langle V, w, P \rangle$  donde  $V$  es el *alfabeto*,  $w \in V^*$  es una palabra no vacía conocida como el *axioma*, y  $P \subseteq V \times V^*$  un conjunto finito de *producciones*, las cuales son parejas  $(a, X)$  donde a la letra  $a$  se le asocia la palabra  $X$  y se denota  $a \rightarrow X$ .

Se supone que para toda  $a \in V$  existe al menos una palabra  $X$  tal que  $a \rightarrow X$ . Si no está definida explícitamente, se sobreentiende que le corresponde la identidad  $a \rightarrow a$ .

Se dice que un sistema Lindenmayer es *determinístico* si a cada letra siempre le asocia una única palabra. Si no es única, deberá haber una regla que permita asociar a cada letra una palabra, pero esta regla puede ser aleatoria o puede depender del contexto (es decir, de las letras vecinas) [10].

**B) APLICACIONES.**

Los sistemas Lindenmayer se pueden aplicar al estudio de diferentes tipos de crecimiento, por ejemplo en un filamento se pueden presentar los siguientes tipos de crecimiento:

**a) CRECIMIENTO EXPONENCIAL** (Todas las células se dividen).

Consideremos  $V=(x)$  con el axioma  $x$  y la regla  $x \rightarrow xx$

<p><math>x</math>  <math>xx</math>  <math>xxxx</math>  <math>xxxxxxxx</math>  <math>xxxxxxxxxxxxxxxx</math></p>	<p>Obsérvese (fig. 3) que cada <math>x</math> desde el punto de vista geométrico representa una célula.</p> <p>El número de células cumple la ecuación en diferencias:</p> $N_{n+1} = 2N_n,$ <p>por lo tanto, en la <math>n</math>-ésima iteración hay <math>2^n</math> células.</p>
---	--

**b) CRECIMIENTO APICAL A RAZON CONSTANTE** (Sólo se divide una célula).

Sea  $V=(a,b)$  con el axioma  $ba$  y las reglas  $a \rightarrow ba$  y  $b \rightarrow b$

Los cuadros negros representan las células  $b$  y los punteados representan las células  $a$  (fig. 4).

<p><math>ba</math>  <math>bba</math>  <math>bbba</math>  <math>bbbba</math>  <math>bbbbba</math></p> <p>...</p>	<p><math>\{\bullet\bullet\}</math>  <math>\{\bullet\bullet\bullet\}</math>  <math>\{\bullet\bullet\bullet\bullet\}</math>  <math>\{\bullet\bullet\bullet\bullet\bullet\}</math>  <math>\{\bullet\bullet\bullet\bullet\bullet\bullet\}</math></p>	<p>Aquí el número de células satisface la ecuación en diferencias:</p> $N_{n+1} = 1 + N_n,$ <p>en la <math>n</math>-ésima iteración hay <math>n+1</math> células</p>
---	--	--

c) **CRECIMIENTO APICAL A RAZON DECRECIENTE.**

Igual que el anterior, pero ahora con la regla:  $a \rightarrow \frac{1}{2}ba$   
 Haciendo la convención de que se multiplican por 1/2 todas las longitudes de lo que esté a la derecha del 1/2.  
 Entonces las reescrituras en las diferentes etapas son:  
 (fig. 5)

$ba$	$\rightarrow$	$\rightarrow$
$b(\frac{1}{2})ba$	$\rightarrow$	$\rightarrow \rightarrow$
$b(\frac{1}{2})b(\frac{1}{2})ba$	$\rightarrow$	$\rightarrow \rightarrow \rightarrow$
$b(\frac{1}{2})b(\frac{1}{2})b(\frac{1}{2})ba$	$\rightarrow$	$\rightarrow \rightarrow \rightarrow \rightarrow$

En la n-ésima iteración mide  $b \sum_{k=0}^n (\frac{1}{2})^k a = baS_n$

Este tipo de crecimiento es acotado, a diferencia de los otros, ya que formalmente, en el límite, cuando  $n \rightarrow \infty$ ,  $\lim_{n \rightarrow \infty} baS_n = 2ab$ .

A una función  $g: V^* \rightarrow \{ASR\}^2$ , que le asocia una figura a cada cadena, se le conoce como la *interpretación gráfica* de dichas cadenas. Podemos aprovechar esto para modelar el crecimiento en el plano usando la computadora, en particular en un ambiente tipo lenguaje LOGO se dispone de las instrucciones siguientes:

- f Pinta 1 "pixel" y avanza d pixeles.
- + Gira la dirección un cierto ángulo  $\alpha$  contra reloj.
- Gira la dirección el ángulo  $\alpha$  a reloj.
- [ Guarda en memoria la dirección y posición al inicio del corchete.
- ] Regresa a la dirección y posición almacenadas al inicio del corchete.
- @n Multiplica por el número n

En particular, si  $v$  es una cadena de símbolos,  $(x_0, y_0, \alpha_0)$  el estado inicial de la "tortuga", (con el número  $d$  de pixeles y el ángulo  $\alpha$  fijos), la figura dibujada por la tortuga se llama la *interpretación logo* de  $v$ .

Usando la simbología de este lenguaje, podemos estudiar procesos de ramificación como los siguientes:

**d) RAMIFICACION OPUESTA DISTICA.**

Con el axioma  $fa$  y un ángulo  $\alpha$  consideremos las siguientes producciones:

$a \rightarrow fb$

$b \rightarrow fc$

$c \rightarrow [+d][-d]fa$

$d \rightarrow fd$

entonces el crecimiento quedará dado mediante las iteraciones sucesivas como se observa en la fig. 6:

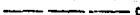
$fa$



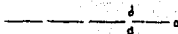
$ffb$



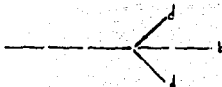
$fffc$



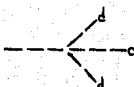
$fff[+d][-d]fa$



$fff[+fd][-fd]ffb$



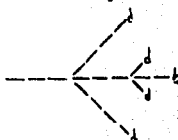
$fff[+ffd][-ffd]fffc$



$fff[+fffd][-fffd]fff[+d][-d]fa$



$fff[+fffd][-fffd]fff[+fd][-fd]ffb$



**e) RAMIFICACION OPUESTA ALTERNADA.**

Con el axioma  $fa$  y un ángulo  $\alpha$  se proponen las siguientes producciones:

$a \rightarrow f[+x]fb$

$b \rightarrow f[-y]fa$

$x \rightarrow a$

$y \rightarrow b$

entonces el crecimiento será así:

(fig. 7)

$fa$



$ff[+x]fb$



$ff[+a]ff[-y]fa$



$ff[+f[+x]fb]ff[-b]ff[+x]fb$

$ff[+f[+a]ff[-y]fa]ff[-f[-y]fa]ff[+a]ff[-y]fa$

#### f) DESCRIPCION DE BOSTRYCHIA RADICANS.

*Bostrychia radicans* es la especie de alga cuyo crecimiento vamos a tratar de recrear con sistemas Lindenmayer. Por simplicidad, en adelante nos referiremos a ella como *B. radicans*. El crecimiento de un individuo de esta especie, a grandes rasgos se desarrolla de la siguiente manera:

Hay una etapa inicial, en la cual la espora va convirtiéndose en un segmento de células dispuestas de manera cilíndrica, que en un extremo tienen una prolongación que le sirve para afianzarse al medio (como una raíz), y en el otro extremo una célula de la cual depende el crecimiento posterior, la célula *apical*.

Esta célula a su vez se divide en otras dos, una que formará parte del segmento anterior (o de un nuevo segmento) y otra que continuará el crecimiento en la misma dirección, formando un estolón. "Periódicamente", después de cierto número de segmentos (que no siempre es el mismo, pero se mantiene en un rango promedio), la célula *apical* se divide en tres células, cada una de las cuales, da origen a un eje de crecimiento distinto.

- 1) El eje que continúa en la dirección de los segmentos anteriores, es de tipo apical *indeterminado*.
- 2) El eje de una rama, que presentará un tipo de crecimiento apical *determinado*, y que además se va ramificando alternadamente.
- 3) El eje del háptera (es como la parte de donde se agarran las enredaderas, no es raíz, pero sirve para fijarse al substrato)

(Ver fig. 8)

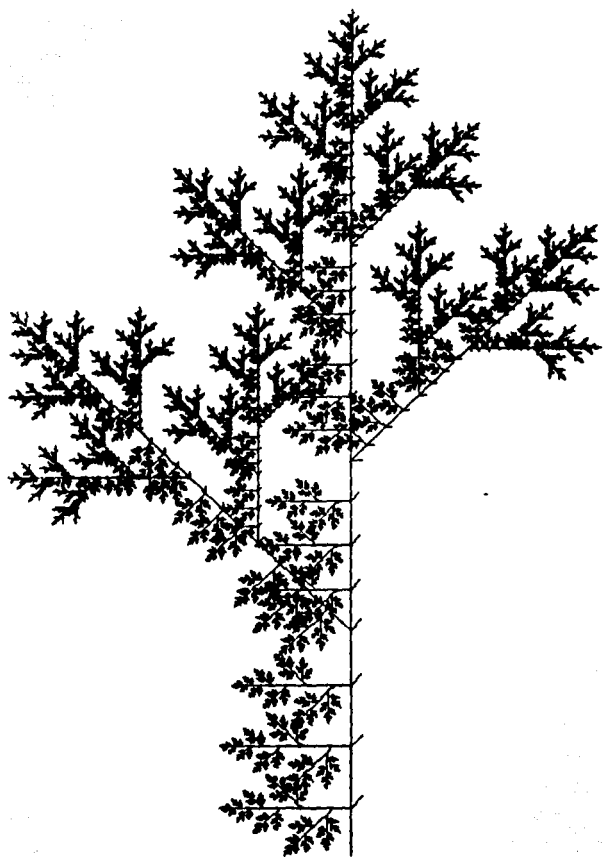


En términos biológicos, el crecimiento *determinado* es aquél donde el número de células no crece indefinidamente, y el tamaño de las mismas tampoco crece más allá de cierto rango. En contraposición, el crecimiento *indeterminado*, es aquél donde el proceso de división celular se desarrolla permanentemente, presentando un tipo de crecimiento potencialmente infinito. Sin embargo, eventualmente una célula apical de las ramas puede empezar a generar crecimiento indeterminado, de manera que el individuo desarrolle estolones en varias direcciones.

Se observa que cada rama es genéticamente igual al anterior, es decir, se trata de un crecimiento *clonal* [3]. Esta característica es la que hace más plausible la modelación con sistemas Lindenmayer, porque el crecimiento clonal se adapta a la iteración de la forma geométrica básica de la rama, y se puede hacer de la siguiente manera, en particular en el sistema *fractint*. (fig. 9)

$\alpha=8$	{Especificamos el ángulo $2\pi/\alpha$ }
axiom c	{La palabra generadora será c}
c → ff@1.1rffrff@.9rffhs	{c,s son las células apicales,
s → ff@1.1rffrff@.9rffkc	pero por razones de ahorro de
	iteraciones, compactamos varias
	etapas en una}
h → [+s][-d]	{h,k son las células ramet que se
k → [-c][--d]	transforman en estolones}
d → f	
r → [++a][-d]	{r es la célula generadora de
	ramet}
a → f[+x]fb	{avanza y se ramifica a la izquierda}
b → f[-y]fa	{avanza y se ramifica a la derecha}
x → a	
y → b	{juntos producen ramificación alterna}
f → @1.36f@1/1.36	{escalamiento proporcional}



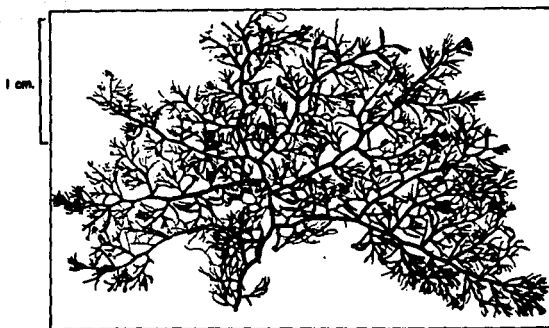


La principal *utilidad* de este modelo para el *estudio biológico* está en que permite plantear y discutir mejor las estrategias que usa la planta para ocupar el espacio.

Por su regularidad y sencillez, este tipo de sistemas puede además auxiliar el *trabajo educativo* y de difusión científica, en cuanto representa una relación interdisciplinaria entre la Biología y la Matemática.

Las *limitaciones* de este modelo aparecen comparando con las imágenes reales del alga:

- 1) El alga real no crece en un plano, sino en la superficie de las raíces de los manglares y bajo la influencia de las mareas. En consecuencia, la disposición de las unidades de crecimiento no será en línea recta como se supuso en el modelo.



2) En este modelo se supone que cada tres ramas aparece un nuevo estolón. Sin embargo, todavía no se conocen los factores que determinan cuando una célula apical de rama se transforma de célula apical en estolón. En todo caso, otra opción sería usar reglas aleatorias para reescribir algunas palabras en el sistema.

Para futuras investigaciones haría falta generalizar el modelo al espacio, implementar reglas aleatorias y calcular la dimensión fractal del alga.

### C) IFS

Definición:

Un sistema de funciones iteradas (IFS) es un espacio métrico completo  $(X, d)$  con un conjunto finito de mapeos  $w_n: X \rightarrow X$  cada uno con factor de contracción  $s_n < 1$ .

Se denota por  $\{X; w_n, n=1, 2, \dots, N\}$  y  $s = \max\{s_n\}$  al sistema de funciones iteradas y al factor de contracción respectivamente.

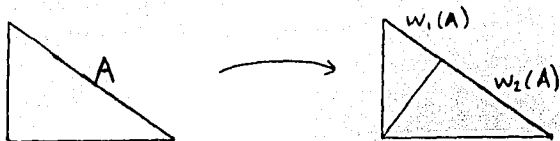
Dado un IFS, podemos asociarle la función conjuntista  $W: \{A \subseteq X\} \rightarrow \{A \subseteq X\}$  definida como sigue

$$W(A) = \bigcup_{n=1}^N w_n(A)$$

es decir, la unión de las imágenes directas de  $A$  bajo las  $w_n$ .

Sea  $H(X)$  = el espacio de los subconjuntos compactos no vacíos de  $X$ , (en el apéndice vemos que es un espacio métrico). Si restringimos la función  $W$  a  $H(X)$ , resulta que es una contracción, y tiene un único punto fijo  $A \in H(X)$ , se llama el atractor del IFS y satisface

la ecuación  $W(A) = A$ , es decir  $A = \bigcup_{n=1}^N w_n(A)$ .



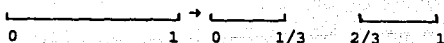
Esto significa geométicamente que el conjunto compacto  $A$  es igual a la unión de sus imágenes bajo las funciones del IFS, de hecho, cada contracción cumple que  $w_n(A) \subseteq A$ , y cubren a  $A$  como un rompecabezas donde se pueden intersectar las piezas.

Por ejemplo, si consideramos  $\mathbb{R}$  con la métrica usual, y elegimos las funciones  $w_1(x) = \frac{x}{2}$ ,  $w_2(x) = \frac{x}{2} + \frac{1}{2}$ , cada una con factor de contracción  $\frac{1}{2}$ , y nos fijamos en el intervalo  $[0,1]$ , resulta que  $w_1([0,1]) = [0, .5]$  y por otra parte  $w_2([0,1]) = [.5, 1]$ , notamos que  $[0,1] = w_1([0,1]) \cup w_2([0,1])$ , entonces  $A = [0,1]$  es el atractor de este IFS. (fig. 11)



Si ahora cambiamos  $w_1(x) = \frac{x}{3}$ , y  $w_2(x) = \frac{x}{3} + \frac{2}{3}$ , el  $[0,1]$  ya no será el atractor, pero podemos fijarnos en sus imágenes:  
 $w_1([0,1]) = [0, 1/3]$ ,  $w_2([0,1]) = [2/3, 1]$ , así  
 $W([0,1]) = w_1([0,1]) \cup w_2([0,1]) = [0, 1/3] \cup [2/3, 1]$

fig. 12



Si aplicamos  $W$  otra vez, resulta:

$$w_1([0,1/3] \cup [2/3,1]) = [0,1/9] \cup [2/9,1/3]$$

$$\text{y } w_2([0,1/3] \cup [2/3,1]) = [2/3,7/9] \cup [8/9,1]$$

Entonces  $W^2([0,1]) = W(W([0,1]))$

$$W([0,1/3] \cup [2/3,1]) = [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1]$$

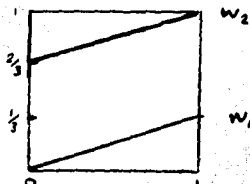
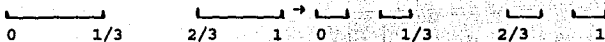


fig. 13



Se puede ver que  $W^n([0,1])$  "tiende" al conjunto de Cantor  $C$ . De hecho,  $h(W^n([0,1]), C) < (1/3)^n \rightarrow 0$ , cuando  $n \rightarrow \infty$ . Es más, se puede ver que el conjunto de Cantor satisface la ecuación conjuntista  $W(C) = C$ , es decir, es el punto fijo y en consecuencia, es el atractor.

Pero vamos a enfocar nuestra atención a IFS en  $\mathbb{R}^2$  con *mapeos afines* de la forma

$$w_n(x, y) = \begin{pmatrix} a_n & b_n \\ c_n & d_n \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e_n \\ f_n \end{pmatrix},$$

que son *contracciones, rotaciones y traslaciones*.

Se puede ver que una condición para que sean contracciones es que los valores propios sean menores que 1, o equivalentemente, que  $(a+d)^2 - 4 < (a-d)^2 + 4bc < 4$ . (ver apéndice).

Un ejemplo análogo al IFS del conjunto de Cantor es el triángulo de Sierpinski, a saber:

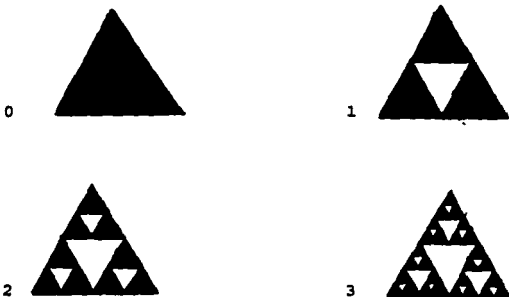
Sean

$$w_1 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} .5 & 0 \\ 0 & .5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

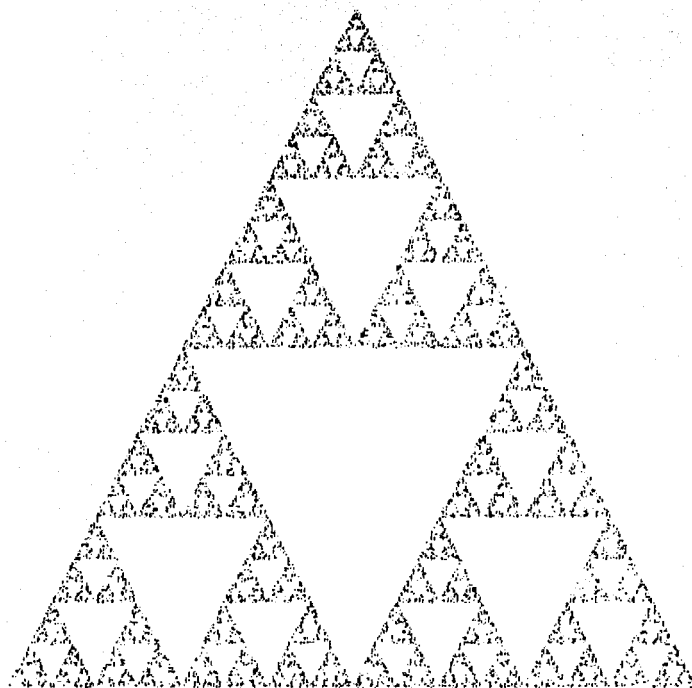
$$w_2 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} .5 & 0 \\ 0 & .5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$w_3 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} .5 & 0 \\ 0 & .5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} .5 \\ \sqrt{3}/4 \end{pmatrix}$$

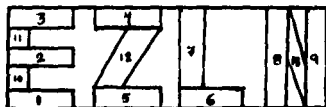
Este IFS mapea el triángulo equilátero con vértices en  $(0,0)$ , en  $(1,0)$  y en  $(.5, \sqrt{3}/4)$  en él mismo, generando tres triángulos, pero dejando un "hueco triangular" en cada etapa. Varias iteraciones de este procedimiento se ilustran en la fig. 14:



De este modo se pueden generar figuras con propiedades de autosemejanza, y dentro de éstas, las que son semejantes a sí mismas a escalas tan pequeñas como queramos, es decir, los fractales.



A manera de ejemplo consideremos la palabra EZLN, tiene 12 segmentos de "recta" (en realidad usaremos 13 paralelogramos que no se intersectan), Encerremos la palabra en un rectángulo de 15 por 5 unidades, y numeremos cada uno de los paralelogramos (figura 15)



Vamos a encontrar transformaciones de la forma

$W_n(x, y) = \begin{bmatrix} a_n & b_n \\ c_n & d_n \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} e_n \\ f_n \end{bmatrix}$  cuyas imágenes queden contenidas en cada una de las regiones numeradas.

Construcción de las  $W_n(\vec{x}) = A_n \vec{x} + k_n$

Vamos a determinar cada  $W_n$  por la asignación de dos vectores linealmente independientes. Estas originan sistemas de ecuaciones cuya solución son los parámetros buscados.

$W_1$  tal que  $A_1(15, 0) = (3, 0)$ ,  $A_1(0, 5) = (0, 1)$  y  $k_1 = (0, 0)$ .

$W_2$  con  $A_2 = A_1$  y  $k_2 = (0, 2)$

$W_3$  con  $A_3 = A_1$  y  $k_3 = (0, 4)$

$W_4$  con  $A_4 = A_1$  y  $k_4 = (4, 4)$

$W_5$  con  $A_5 = A_1$  y  $k_5 = (4, 0)$

$W_6$  con  $A_6 = A_1$  y  $k_6 = (8, 0)$

$W_7$  tal que  $A_7(15, 0) = (0, 4)$ ,  $A_7(0, 5) = (-1, 0)$  y  $k_7 = (9, 1)$

$W_8$  tal que  $A_8(15, 0) = (0, 5)$ ,  $A_8(0, 5) = (-1, 0)$  y  $k_8 = (13, 0)$

$W_9$  con  $A_9 = A_8$  y  $k_9 = (15, 0)$

$W_{10}$  tal que  $A_{10}(15, 0) = (0, 1)$ ,  $A_{10}(0, 5) = (-1, 0)$  y  $k_{10} = (1, 1)$

$W_{11}$  con  $A_{11} = A_{10}$  y  $k_{11} = (1, 3)$

$W_{12}$  tal que  $A_{12}(15, 0) = (2, 3)$ ,  $A_{12}(0, 5) = (-1, 0)$  y  $k_{12} = (5, 1)$

$W_{13}$  tal que  $A_{13}(15, 0) = (1, -3)$ ,  $A_{13}(0, 5) = (0, 2)$  y  $k_{13} = (13, 3)$ .



THE UNIVERSITY OF CHICAGO

PHYSICS DEPARTMENT

PHYSICS 311

LECTURE 1

LECTURE 2

#### D) RELACION ENTRE LOS IFS Y LOS SISTEMAS LINDENMAYER

Dado un sistema Lindenmayer, consideremos el problema de construir un sistema de funciones iteradas (IFS) y un conjunto inicial tales que las imágenes sucesivas de dicho conjunto correspondan a cada iteración del sistema sobre el axioma (palabra generadora o condición inicial).

Consideremos un sistema que modela un crecimiento determinado, de modo que las imágenes de la imagen inicial se aproximen a un conjunto límite.

Propondremos un procedimiento que utilice el hecho de que el atractor  $A$  de un IFS cumple la ecuación

$$A = \bigcup_{n=1}^N v_n(A)$$

tratando de que el atractor del IFS coincida con el conjunto límite del sistema Lindenmayer.

Sea  $A_n$  el conjunto correspondiente a la  $n$ -ésima iteración del IFS, y  $P_n$  la palabra correspondiente del sistema Lindenmayer. Si podemos plantear una ecuación recursiva en las palabras,  $P_{n+1} = f(P_n)$ , y de ésta formular una ecuación recursiva en los conjuntos  $A_{n+1} = F(A_n)$ , y si en un proceso al límite, los conjuntos  $A_n$  tienden a un conjunto  $A$  cuando  $n$  tiende a infinito, para este conjunto  $A$ , la ecuación recursiva se transforma en una igualdad entre  $A$  y una unión de transformaciones del mismo  $A$ , (por ejemplo, copias reducidas, trasladadas y rotadas), estas transformaciones de  $A$  determinarán las funciones del IFS que construiremos. Dependiendo del axioma del sistema Lindenmayer haremos la elección del conjunto inicial del IFS.

Podemos plantear la ecuación recursiva en los conjuntos  $A_n$

$$A_{n+1} = f(s) \cup T_1(A_n) \cup T_2(A_n) \cup T_3(A_n)$$

Donde cada  $T_i: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  de la forma

$$T_i(x) = M_i x + b$$

con  $x, b \in \mathbb{R}^2$ , y  $M_i$  una matriz que contrae por un factor de  $1/2$  y puede rotar un cierto ángulo en sentido positivo o negativo.

Tomando límites en la ecuación:

$$A = f(s) \cup T_1(A) \cup T_2(A) \cup T_3(A)$$

$$\text{donde } A = \lim_{n \rightarrow \infty} A_n$$

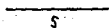
Tenemos una ecuación del tipo  $A = F(A)$  de "conjunto fijo" con contracción, por lo tanto, partiendo de un conjunto inicial, podemos mapear sus imágenes sucesivamente y deberían acercarse al conjunto  $A$ .

En este ejemplo proponemos a  $f(s)$  como el conjunto inicial y a las  $T_i$  como funciones del IFS. El comportamiento geométrico que presenta es como sigue:

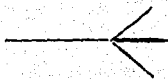
Se empieza con un segmento de longitud  $s$ , multipliquémoslo por una razón constante  $r$  menor que 1, (en este caso  $1/2$ ) y formemos tres copias, cada una de las cuales pegaremos con el segmento original, una en línea recta y las otras dos formando un ángulo  $\alpha$  con los otros segmentos. Después cada una de estas "ramas" se van ramificando de manera autosemejante, y tienden a un objeto como un alga con ramificación paralela.

(fig. 17)

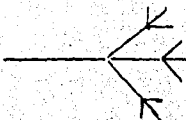
etapa 0



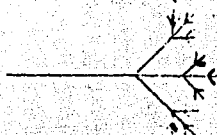
etapa 1



etapa 2



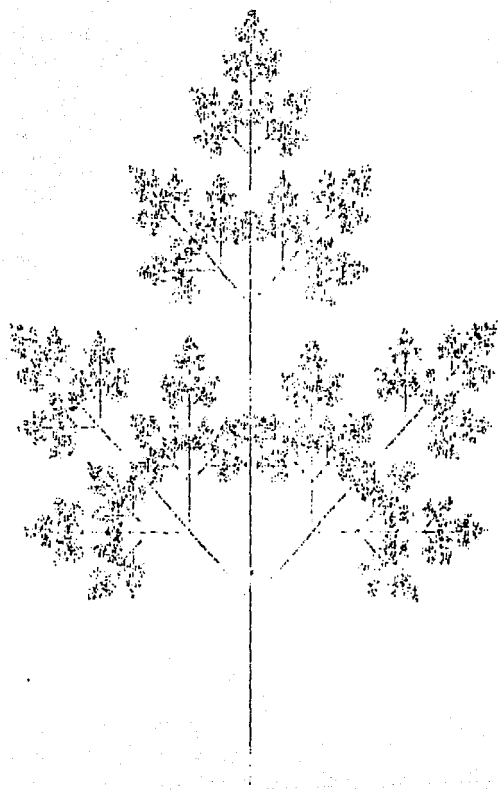
etapa 3

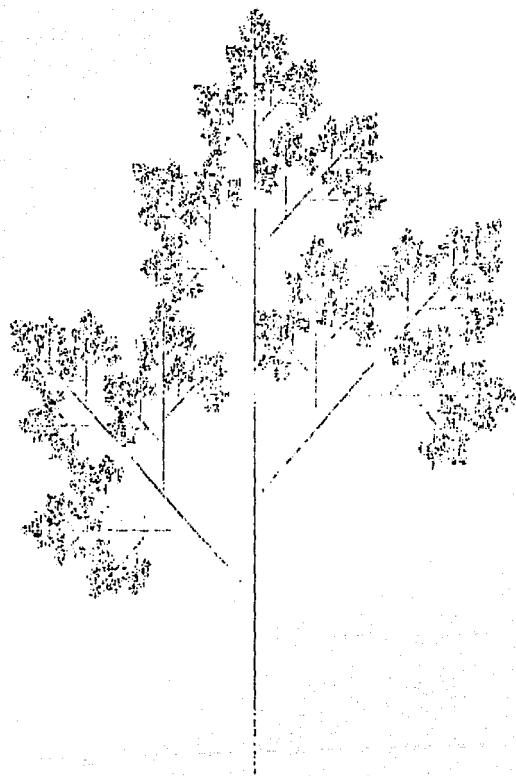


(El "límite" se anexa en la fig. 18)

Si hacemos algunas modificaciones en las traslaciones ahora tendremos una ramificación alternada como el de *radicans* (ver la figura 19). Sin embargo, no podremos reconstruir el estolón porque los atractores de los IFS tienen la propiedad de la autosemejanza, la cual no cumple el sistema Lindenmayer usado para *radicans*.

fig 18 y 19





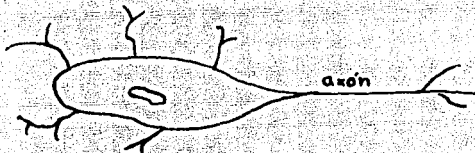
## CAPITULO II. REDES NEURONALES Y APRENDIZAJE.

### A) INTRODUCCION.

Las *redes neuronales* son modelos matemáticos inspirados en el funcionamiento de las neuronas reales, enfocados al aspecto de la transmisión de las *señales*.

Este tipo de modelos tienen una amplia gama de aplicaciones en tareas de clasificación, optimización, "aprendizaje", y otras donde sirva el procesamiento paralelo de información [6].

La unidad básica de una red neuronal es, naturalmente, la neurona. Una neurona biológica es una célula que consta de un cuerpo (soma), varias ramificaciones llamadas *dendritas*, y una prolongación principal conocida como el *axón*, por medio del cual puede transmitir las señales (fig. 1).



Cuando el axón de una neurona se conecta con las dendritas de otra, se dice que forman *sinapsis* y en este caso, la neurona recibe las señales a través de las dendritas. Estas señales influyen en el potencial eléctrico, de modo que si alcanza cierto valor umbral, un pulso o *señal* es emitida.

*McCulloch* y *Pitts* propusieron un modelo según el cual las señales de entrada se suman ponderadamente, disparando con cierta intensidad si la suma rebasa el umbral, en caso contrario no dispara *señal* alguna [6].

Así, si representamos el estado  $n_i$  de la  $i$ -ésima neurona por 1 si dispara y 0 si no lo hace, tendríamos que este estado está en función de las señales recibidas de la manera siguiente:

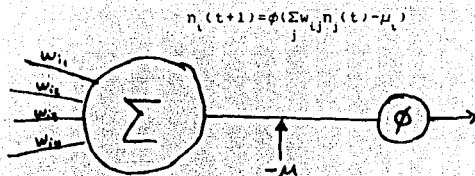


fig. 2

Donde  $\phi(x) = 1$  si  $x \geq 0$  y  $\phi(x) = 0$  si  $x < 0$ .

$\mu_i$  es el valor umbral de la neurona  $i$ -ésima, cada  $w_{ij}$  representa la fuerza de las sinápsis que vincula la  $j$ -ésima neurona con la  $i$ -ésima.  $w_{ij} > 0$  si la sinápsis es excitatoria,  $w_{ij} < 0$  si la sinápsis es inhibitoria, y  $w_{ij} = 0$  si no hay sinápsis entre ambas neuronas.

Este modelo supone además un tratamiento discreto del tiempo, en donde el estado próximo ( $t+1$ ) de una neurona depende de los estados actuales ( $t$ ) de las neuronas que inciden en ella.

Algunas limitaciones de este modelo en particular son las siguientes:

- Las neuronas reales no siempre disparan sus señales con la misma intensidad, pueden hacerlo incluso gradualmente.
- Las entradas no necesariamente influyen de manera lineal.
- La emisión de las señales no tiene por qué ser simultánea, ya que cada neurona puede tardar un tiempo diferente en cada paso.



Una generalización del modelo presentado sería:

$$n_i = g\left(\sum_{j=1}^n w_{ij} n_j - \mu_i\right)$$

donde  $g$  sea una función no lineal continua, creciente y acotada, llamada función de activación o función de transferencia.

A continuación presentamos algunos ejemplos de las aplicaciones de las redes neuronales.

### 1) Clasificación simple.

Consideremos el conjunto siguiente:

$\{(1,1,1), (1,1,-1), (1,-1,1), (1,-1,-1), (-1,-1,1), (-1,-1,-1)\}$

Supongase que para cualquiera de los elementos de este conjunto se tiene que definir si se parece más al vector

$$\vec{c} = (-1, 1, 1)$$

o al

$$-\vec{c} = (1, -1, -1).$$

Una manera de hacerlo es construir una neurona con la estructura  $f(\vec{x}) = \text{sgn}(\vec{c} \cdot \vec{x})$ , decidiéndose por  $\vec{c}$  si sale 1 y decidiendo  $-\vec{c}$  si sale -1. (Aquí  $\vec{c} \cdot \vec{x} = (-1, 1, 1) \cdot (x, y, z) = -x + y + z$ ).

Consideremos  $\vec{c} \cdot \vec{x} = 0$ , es decir,  $-x + y + z = 0$ , es la ecuación de un plano perpendicular a  $\vec{c}$  que pasa por el origen, si  $\vec{c} \cdot \vec{x} > 0$ ,  $\vec{x}$  está del mismo lado que  $\vec{c}$ , mientras que cuando  $\vec{c} \cdot \vec{x} < 0$ ,  $\vec{x}$  está del lado contrario.

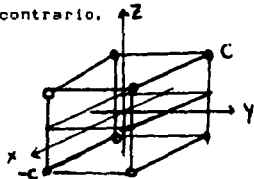
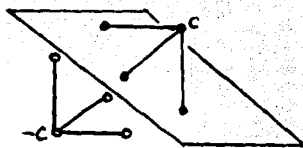


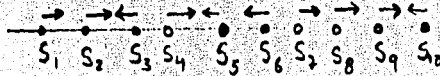
fig. 3



Podríamos preguntarnos si alguno de los vectores del conjunto que queremos clasificar queda en la región indecidible, es decir, en el plano  $-x + y + z = 0$ , pero como estamos trabajando con vectores con entradas que son 1 ó -1, la suma y resta de tres de ellos no da cero.

## 2) Aplicaciones a la física.

Existe una analogía entre las redes de neuronas de Hopfield y algunos modelos de materiales magnéticos. Una descripción simple de un material magnético es que consiste de un conjunto de imanes atómicos arreglados en una latiz regular (por ejemplo, un arreglo rectangular) que representa la estructura cristalina del metal.



Usaremos el término "spin" (giro) para estos imanes atómicos, nombre tomado de la mecánica cuántica. Los spines pueden tener varias direcciones, dependiendo del material estudiado. El caso más simple es cuando solo hay dos direcciones posibles. Estos se representan en un modelo de Ising por una variable binaria que nos indica el estado  $S_i$  para cada sitio  $i$  de la latiz con valores  $+1$ , si el spin gira en una dirección; ó  $-1$ , si el spin gira en la otra dirección.

Este tipo de modelos se pueden usar en otros contextos, por ejemplo, en el estudio de un gas podemos describir la posición de las partículas especificando una variable de Ising (lleno/vacío) en cada celda de una fina red que cubre el espacio donde está el gas.

Un modelo de Ising no está completamente especificado hasta que las dinámicas de interacción de los spin estén dadas.

En el caso del material magnético, cada spin está influido por el campo magnético  $h$ . Este campo consiste de un campo externo  $h^{ext}$  más un campo interno producido por los otros spines.

Si suponemos que la contribución de un átomo al campo magnético en otro lugar, es proporcional a su propio spin, sumando las contribuciones de los átomos vecinos tenemos un campo magnético que influye en la unidad  $S_i$

$$h_i = \sum_j w_{ij} S_j \cdot h^{ext}$$

donde  $w_{ij}$  mide la fuerza de la influencia de  $S_j$  en  $S_i$ .

A baja temperatura, un spin tiende a alinearse paralelamente al campo local  $h_i$  que actúa en él, esto lo describe la ecuación

$$S_i = \text{sgn}(h_i)$$

que es análoga a la de las neuronas.

Otro camino es dar una energía potencial correspondiente a las interacciones, por ejemplo

$$H = -\frac{1}{2} \sum_{i,j} w_{ij} S_i S_j - h^{ext} \sum_i S_i$$

donde  $h^{ext}$  representa el umbral.

Podemos ver a  $H$  como a una función de energía, es decir, una función con la propiedad de que le impone al sistema una regla dinámica según la cual decrece dicha función; es decir, se comporta como los sistemas físicos que evolucionan tendiendo a la energía mínima.

En los problemas anteriores las fuerzas de las conexiones son simétricas, (esto quiere decir que cada  $w_{ij}$  es igual al  $w_{ji}$  correspondiente) y la regla dinámica elegida fue

$$S_i := \text{sgn}(\sum_j w_{ij} S_j)$$

Podemos ver que así decrece la función

$$H = K - \sum_{i,j} w_{ij} S_i S_j$$

con  $K$  una constante.

Vamos a ver si la dinámica propuesta realmente hace decrecer esta función, comparando el valor de la función en la configuración  $(S_1, S_2, \dots, S_i, \dots, S_N)$ , con el valor de  $H$  en  $(S_1, \dots, S_i', \dots, S_N)$ . Aplicando la regla dinámica que a  $S_i$  le asigna  $S_i'$ ,

Para una unidad particular  $i$ , sea

$$S_i' = \text{sgn}(\sum_j w_{ij} S_j)$$

Si  $S_i' = S_i$ , la energía no cambia.

Si  $S_i' \neq S_i$ , entonces  $S_i' = -S_i$

y entonces comparando los términos que involucran a  $S_i$  tenemos que:

$$\begin{aligned} H' - H &= - \sum_{i,j} w_{ij} S_i' S_j + \sum_{i,j} w_{ij} S_i S_j \\ &= 2S_i \sum_{j \neq i} w_{ij} S_j \\ &= 2S_i \sum_j w_{ij} S_j - 2w_{ii} \\ &= -2S_i \sum_{j \neq i} w_{ij} S_j - 2w_{ii} \end{aligned}$$

El primer término es negativo, y por la ecuación de la *Regla de Hebb*  $w_{ii} > 0$ . Entonces  $H' - H$  es negativo y por lo tanto decreció  $H$ .

A alta temperatura, las fluctuaciones térmicas tienen el efecto de voltear los spin, y choca con la tendencia del campo magnético a alinearlos. Una manera de describir este efecto es introduciendo unidades estocásticas, por ejemplo, con la *dinámica de Glauber*, definida por la regla estocástica

$$S_i = +1 \text{ con probabilidad } g(h_i)$$

y

$$S_i = -1 \text{ con probabilidad } 1 - g(h_i),$$

$$\text{donde } g(h) = f_b(h) = 1/(1 + e^{-2bh})$$

Está bien definida porque  $0 < g(h) < 1$  para toda  $h$ .

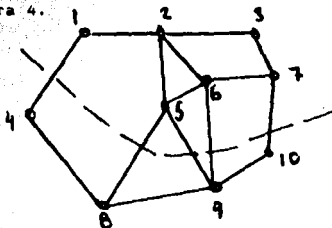
Observemos que cuando  $b \rightarrow \infty$ ,  $f_b(h) \rightarrow \text{sgn}(h)$ , esto quiere decir que  $P(S_i = 1 | h > 0) = 1$ ,  $P(S_i = -1 | h > 0) = 0$ ,  $P(S_i = 1 | h < 0) = 0$  y que  $P(S_i = -1 | h < 0) = 1$ . En otras palabras, se pierde la aleatoriedad.

Por otra parte,  $f_b(h) \rightarrow 1/2$  si  $b \rightarrow 0$ , lo cual implica que  $P(S_i = 1 | h > 0) = 1/2$ ,  $P(S_i = -1 | h > 0) = 1/2$ ,  $P(S_i = 1 | h < 0) = 1/2$ ,  $P(S_i = -1 | h < 0) = 1/2$ , lo que significa que la aleatoriedad domina totalmente al campo. Al parámetro  $k = 1/b$  podemos interpretarlo como una temperatura, de modo que abre una vía de generalización del modelo de red neuronal que hemos trabajado, introduciendo unidades donde determinemos aleatoriamente el umbral, según la distribución  $f_b(\cdot)$ .

#### 4) Bipartición de gráficas.

En este problema, una gráfica es un conjunto de  $n$  vértices y un conjunto de aristas que los unen. Consideraremos gráficas conexas con un número par de vértices. El problema de la bipartición consiste en dividir los vértices en dos conjuntos A y B de igual tamaño minimizando el número de aristas entre ambos conjuntos.

Ver figura 4.



Definimos  $C_{ij} = 1$  si los vértices  $i$  y  $j$  están conectados; y  $C_{ij} = 0$  si no lo están.

Definimos además  $S_i = 1$  si está en el conjunto A y  $S_i = -1$  si está en B. Entonces queremos minimizar  $L = -\sum_{(i,j)} C_{ij} S_i S_j$ , sujeto a la restricción

$\sum_j S_j = 0$  (esta es una forma de decir que tengan el mismo tamaño A y B)

Una manera de insertar la restricción suavemente es sumarla como un término penalizador; por ejemplo  $H = -\sum_{(i,j)} C_{ij} S_i S_j + \mu (\sum_i S_i)^2$ .

Como

$$\begin{aligned} (S_1 + \dots + S_n)(S_1 + \dots + S_n) &= S_1^2 + S_1 S_2 + \dots + S_1 S_n \\ &+ S_2 S_1 + S_2^2 + \dots + S_2 S_n \\ &\dots \\ &+ S_n S_1 + S_n S_2 + \dots + S_n^2 \\ &= \sum_i S_i^2 + 2 \sum_{(i,j)} S_i S_j \end{aligned}$$

Simplificando queda

$$\begin{aligned} H &= -\sum_{(i,j)} C_{ij} S_i S_j + n\mu - 2\mu \sum_{(i,j)} S_i S_j \\ &= n\mu - \sum_{(i,j)} (C_{ij} - 2\mu) S_i S_j \end{aligned}$$

Es de la forma

$$K = \sum_{i,j} w_{ij} S_i S_j$$

donde  $K = n\mu$  y  $w_{ij} = C_{ij} - 2\mu$ .

Así pues, podemos usar la regla

$$S_i' = \text{sgn}(\sum_j w_{ij} S_j)$$

y hará decrecer  $H$  como a una función de energía.

## B) APRENDIZAJE.

Si tenemos una red neuronal que a cada vector de señales de entrada  $\vec{x}$ , le asocia un único vector de señales de salida  $\vec{y}$ , podemos pensarla como una función que va de  $E$  (el conjunto de las posibles configuraciones de entrada) en  $S$  (el conjunto de las posibles salidas).

Dada una función  $f: A \rightarrow S$ , ASE, podemos plantear el problema de construir una red neuronal  $R: E \rightarrow S$  tal que  $R(a) = f(a)$  para toda  $a \in A$ , en otras palabras, tal que  $R|_A = f$ . Este planteamiento trae consigo dos cuestiones, una es la existencia de una red neuronal que satisfaga el requerimiento, la otra es cómo construirla.

La primera cuestión depende básicamente de la función  $f$  y de su dominio  $A$ , más adelante la discutiremos someramente para el caso de redes organizadas en capas de neuronas, donde cada elemento de una capa solamente puede recibir señales de la capa anterior, e incidir en neuronas de la siguiente capa. Sobre la otra cuestión, suponiendo que la red tiene el tipo de organización necesaria, así como el número de neuronas suficientes, el problema se reduce a determinar los parámetros (los pesos de las conexiones y los umbrales de cada neurona). Una manera de hacer esto es comparar las salidas buscadas  $f(a)$  con las salidas obtenidas  $R(a)$  para cada  $a \in A$ , y proponer un mecanismo mediante el cual se modifiquen los parámetros tratando de no cometer el mismo error, en tal caso diremos que la red está aprendiendo la asociación  $a \rightarrow f(a)$ .

Comencemos por las situaciones más sencillas.

Ejemplo 1: Función binaria OR.

Consideremos la siguiente asociación:

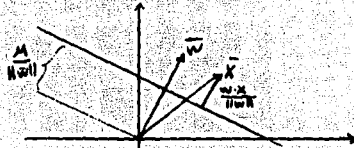
(0,0) ---> 0  
(1,0) ---> 1  
(0,1) ---> 1  
(1,1) ---> 1

Vamos a tratar de recrear esta función con una red consistente de una neurona de la forma

$$N(\bar{x}) = \phi(\bar{w} \cdot \bar{x} - \mu)$$

donde  $\phi(t) = 1$  si  $t \geq 0$ , y  $\phi(t) = 0$  si  $t < 0$ .

Notamos que el semiplano ( $\bar{x} \in \mathbb{R}^2$  tales que  $\bar{w} \cdot \bar{x} - \mu \geq 0$ ) va a dar el 1, y que su complemento (otro semiplano) va a dar el 0 bajo  $N(\bar{x})$ . Geométricamente la recta que los divide es perpendicular a  $\bar{w}$  y está a una distancia  $\mu / \|\bar{w}\|$  del origen.



En general, una neurona del tipo  $\phi(\bar{w} \cdot \bar{x} - \mu)$  así, divide el espacio  $n$ -dimensional en dos regiones con un hiperplano que asigna 0 a los vectores  $\bar{x}$  cuya proyección  $(\bar{w} \cdot \bar{x}) / \|\bar{w}\|$  sobre  $\bar{w}$  es menor que  $\mu / \|\bar{w}\|$ , y asigna 1 a aquellos vectores cuya proyección sobre  $\bar{w}$  sea mayor que  $\mu / \|\bar{w}\|$ .

Para el problema planteado por la función OR, podemos proponer a  $w=(1,1)$  y  $\mu=1/2$ , y con una sola neurona resolvemos el problema.

$$(1,1) \cdot (u,v) = u+v$$

tenemos que  $\phi(wx-\mu) = \phi(u+v-1/2)$

$$\begin{array}{ll} (0,0) \rightarrow 0 & (0,1) \rightarrow 1 \\ (1,0) \rightarrow 1 & (1,1) \rightarrow 1 \end{array}$$

*Ejemplo 2:* Sea  $E=\mathbb{R}^n$ ,  $S=(0,1)$ ,  $A$  un subconjunto finito de  $\mathbb{R}^n$ , y  $f:A \rightarrow S$  una función binaria.

Se dice que el problema es linealmente separable, cuando existe una recta (o hiperplano)  $w \cdot x = \mu$  que genera una partición de  $A$  en dos conjuntos ajenos  $B = \{x \mid w \cdot x \geq \mu\} \cap A$ , y  $C = \{x \mid w \cdot x < \mu\} \cap A$ , tales que  $B = f^{-1}(1)$  y  $C = f^{-1}(0)$ . En este caso una neurona  $N(x) = \phi(w \cdot x - \mu)$  resuelve el problema, pues para  $b \in B$ ,  $N(b) = 1 = f(b)$  y para  $c \in C$ ,  $N(c) = 0 = f(c)$ , entonces  $N|_A = f$ .

Pero no siempre es fácil elegir a  $w$  y  $\mu$  adecuadamente. Sin embargo, podemos desarrollar técnicas para ajustar estos parámetros, en este trabajo presentaremos el método del gradiente descendente [6],[7].

Primero veremos cómo podemos suponer, sin pérdida de generalidad, que  $\mu=0$ , porque lo podemos incluir como una entrada más del vector  $w$ , si agregamos la entrada  $-1$  a cada uno de los vectores entrada originales, por ejemplo en el problema de la función OR, tendríamos que separar los vectores  $(0,0,-1)$ ,  $(0,1,-1)$ ,  $(1,0,-1)$  y  $(1,1,-1)$  con un plano que pase por el origen.

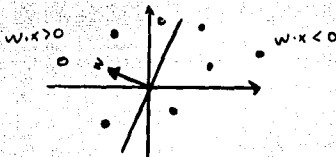


En general replantearíamos el problema del siguiente modo:  
Sean

$$A' = Ax(-1), E' = ExR, F: A' \rightarrow S \text{ definida por } F(a, -1) = f(a)$$

El hiperplano  $\bar{w} \cdot \bar{x} - \mu = 0$  está contenido en otro hiperplano mayor cuya ecuación es  $(\bar{w}, \mu) \cdot (\bar{x}, t) = 0$  que si pasa por el origen. De hecho, haciendo  $t = -1$ , queda la misma ecuación. Por lo tanto podemos enfocar nuestra atención a neuronas de la forma  $N(\bar{x}) = g(\bar{w} \cdot \bar{x})$ . También podemos suponer que  $g(t) = \text{sgn}(t)$  [7].

Observación. Un vector  $\bar{w}$  es solución del problema siempre y cuando  $\bar{w} \cdot \bar{x} > 0$  para todos los  $\bar{x} \in \bar{X}^{-1}(1)$  y  $\bar{w} \cdot \bar{x} < 0$  para los  $\bar{x} \in \bar{X}^{-1}(0)$ , en otras palabras, para cada  $\bar{x}$  fijo,  $\bar{w}$  debe estar en un semiplano determinado por  $\bar{x}$ . Y viceversa, si  $\bar{w}$  está en todos estos semiplanos, es solución. Por lo tanto, el conjunto de  $\bar{w}$  que son solución es la intersección de dichos semiplanos. (Ver figura).



Hechas las consideraciones anteriores, establezcamos una medida del error, por ejemplo, la siguiente:

$$E(\bar{w}) = \sum_{\bar{a} \in A} (f(\bar{a}) - \bar{w} \cdot \bar{a})^2$$

$f(\bar{a})$  es la salida que deseamos que corresponda a  $\bar{a}$ , y  $N(\bar{a}) = \text{sgn}(\bar{w} \cdot \bar{a})$

Queremos modificar  $\bar{w}$  en las dirección que reduzca el error. Sabemos que el gradiente de una función indica la dirección en que crece más, por lo tanto, el error  $E(\bar{w})$  se reduce en la dirección contraria a su gradiente.

Las entradas del gradiente de E son:

$$\frac{\partial E}{\partial w_i} = 2 \sum_{\bar{a} \in A} (f(\bar{a}) - \bar{w} \cdot \bar{a}) (-a_i) \quad \bar{a} = (a_1, \dots, a_n)$$

Por lo tanto,  $-\eta \nabla E = 2\eta \sum_{a \in A} (f(\bar{a}) - \bar{w} \cdot \bar{a}) \bar{a}$

Entonces  $\Delta \bar{w}_a = 2\eta (f(\bar{a}) - \bar{w} \cdot \bar{a}) \bar{a}$

Y la propuesta sería modificar el vector de pesos con la regla:  
 $\bar{w}_{\text{nuevo}} = \bar{w} + \Delta \bar{w}_a$  si  $\bar{w} \cdot \bar{a}$  no tiene el signo deseado.

Podemos ilustrar el procedimiento:

### Ejemplo 3.

Supongamos que se quiere una neurona  $N(\bar{x}) = \text{sgn}(\bar{w} \cdot \bar{x})$  que produzca las siguientes respuestas:

$a = (3, 1) \rightarrow 1 = f(a)$

$b = (3, -2) \rightarrow 1 = f(b)$

$c = (-2, -1) \rightarrow -1 = f(c)$

Empecemos, por ejemplo, con  $\bar{w}_0 = (0, 1)$  y sea  $\eta = 1$

Probemos con a:

$\bar{w}_0 \cdot a = (0, 1) \cdot (3, 1) = 1$ , como queremos.

Ahora probamos con b:

$\bar{w}_0 \cdot b = (0, 1) \cdot (3, -2) = -2$ , tenemos que cambiar

$\bar{w}_1 = \bar{w}_0 + \Delta \bar{w}_b$  como  $\Delta \bar{w}_b = \eta (f(b) - \bar{w}_0 \cdot b) b = 1(1 + 2)(3, -2) = (9, -6)$   
 $= (0, 1) + (9, -6) = (9, -5)$

Probamos con c:

$\bar{w}_1 \cdot c = (9, -5) \cdot (-2, -1) = -13$ , vamos bien

Checamos

$\bar{w}_1 \cdot a = (9, -5) \cdot (3, 1) > 0$  como se pide

$\bar{w}_1 \cdot b = (9, -5) \cdot (3, -2) > 0$  como debe ser.

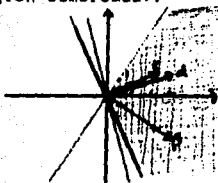
∴ La neurona  $N$  con  $\bar{w}_1 = (9, -5)$  sí reprodujo la asociación deseada.

Aprovechando el ejemplo, quisiéramos mostrar una representación alternativa que nos ayude a visualizar una prueba para el algoritmo. Definamos  $\alpha = f(a) = 1(3, 1) = (3, 1)$

$\beta = f(b) = 1(3, -2) = (3, -2)$

$\gamma = f(c) = -1(-2, -1) = (2, 1)$

Ahora la condición  $\text{sgn}(\bar{w} \cdot a) = f(a)$  se transforma en  $\bar{w} \cdot a > 0$ . Entonces en este planteamiento el problema es linealmente separable si y sólo si las nuevas entradas  $a$  quedan todas del mismo lado de una recta que pasa por el origen (o plano o hiperplano, según el problema). En este caso está determinado por un vector perpendicular, el cual está en un cono o pirámide (en el ejemplo, en la región sombreada).



El método se traduciría como sigue:

Supongamos que queremos una red neuronal que responda si, (+1), a un conjunto de entradas  $F^+$ , y que responda no, (-1), a otro conjunto de entradas  $F^-$ . Supongamos también que el problema es linealmente separable.

Primero hagamos  $F = F^+ \cup G$ , donde  $G = \{\bar{x} = -a \text{ tales que } a \in F^-\}$  el problema se trata de encontrar un vector de pesos  $\bar{w}$  tal que  $\bar{w} \cdot \bar{x} > 0$  para todos los  $\bar{x} \in F$ .

A continuación utilizaríamos el algoritmo siguiente [7]:

INICIO: Elegir un  $\bar{w}_0$  y un  $\bar{x} \in F$  cualquiera.

PRUEBA: Si  $\bar{w}_i \cdot \bar{x} > 0$  se elige otro  $\bar{x}$  y regresa a PRUEBA hasta que pasen todos los elementos de  $F$ . Si no es así ve a CAMBIO.

CAMBIO:  $\bar{w}_{i+1} = \bar{w}_i + \eta \bar{x}$ .

donde  $\eta > 0$  y se llama el paso del algoritmo.

¿Cómo sabemos que este algoritmo funciona?

Sea  $\bar{u}$  un vector solución unitario, sea  $\delta = \min_{x \in F} (\bar{u} \cdot \bar{x}) / 2$ ,  $\delta > 0$  porque  $F$  es finito.

Sea  $G(\bar{w}) = \bar{u} \cdot \bar{w} / \|\bar{w}\| = \cos \alpha$

donde  $\alpha$  es el ángulo entre  $\bar{u}$  y  $\bar{w}$ .

(Observamos que  $\cos \alpha \rightarrow 1 \Leftrightarrow \alpha \rightarrow 0 \Leftrightarrow \bar{w}$  se acerca al cono donde está  $\bar{u}$  y por lo tanto tarde o temprano entra en dicho cono).

Cada vez que incrementamos  $\bar{w}$  sucede que:

$$\bar{u} \cdot \bar{w}_{m+1} = \bar{u} \cdot (\bar{w}_m + \eta \bar{x}) = \bar{u} \cdot \bar{w}_m + \eta \bar{u} \cdot \bar{x} > \bar{u} \cdot \bar{w}_m + \eta \delta$$

De la desigualdad anterior se sigue que

$$\bar{u} \cdot \bar{w}_n > n \eta \delta \quad (1)$$

Por otra parte, como

$$\|\bar{w}_{m+1}\|^2 = \|\bar{w}_m + \eta \bar{x}\|^2 = \|\bar{w}_m\|^2 + 2\eta \bar{w}_m \cdot \bar{x} + \eta^2 \|\bar{x}\|^2$$

Pero para cambiar  $\bar{w}_m$  a  $\bar{w}_{m+1}$  es necesario que  $\bar{w}_m \cdot \bar{x} < 0$ .

Entonces  $\|\bar{w}_{m+1}\|^2 = \|\bar{w}_m\|^2 + 2\eta \bar{w}_m \cdot \bar{x} + \eta^2 \|\bar{x}\|^2 \leq \|\bar{w}_m\|^2 + (M\eta)^2$

donde  $M = \max_{x \in F} \|\bar{x}\|$ .

Si elegimos  $0 < \eta \leq 1/M$  resulta que  $\|\bar{w}_{m+1}\|^2 \leq \|\bar{w}_m\|^2 + 1$

y entonces  $\|\bar{w}_n\|^2 \leq n \quad (2)$

Combinando (1) y (2) obtenemos

$$G(\bar{w}_n) = \bar{u} \cdot \bar{w}_n / \|\bar{w}_n\| \geq n \eta \delta / \|\bar{w}_n\| \geq n \eta \delta / (n)^{1/2} = \eta \delta (n)^{1/2}$$

Pero

$$G(\bar{w}) = \bar{u} \cdot \bar{w} / \|\bar{w}\| = \cos \alpha \leq 1 \text{ por lo tanto } n \leq 1/(\eta \delta)^2$$

En consecuencia el proceso de hallar  $\bar{w}$  solución es finito.

**Ejemplo 4.**

**Función XOR.**

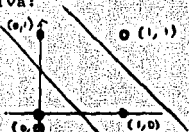
(0,0) ----> 0

(1,0) ----> 1

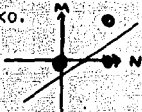
(0,1) ----> 1

(1,1) ----> 0

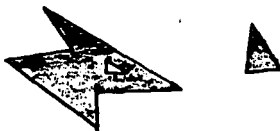
El problema se convierte en dividir el plano en dos regiones ajenas A y B, de modo que A contenga a (0,1) y (1,0); mientras B contenga a (0,0) y (1,1). Notemos que no es linealmente separable, y aunque no se pueden separar con una sola recta, se puede con dos. Esto sugiere la idea de usar dos neuronas cada una determinando una recta, y otra neurona más para hacer la clasificación decisiva:



Con las neuronas  $N = \phi(\bar{w}_1 \cdot \bar{x} - 0.5)$   $M = \phi(\bar{w}_2 \cdot \bar{x} - 1.5)$   $\theta = \phi(\bar{w}_3 \cdot \bar{y} - 0.5)$  donde  $\bar{w}_1 = (1, 1)$ ,  $\bar{w}_2 = (1, -2)$ ,  $\bar{x} \in (0, 1) \times (0, 1)$ ,  $\bar{y} = (N, M) \in (0, 1) \times (0, 1)$ ,  $\phi(x) = 1$  si  $x > 0$  y  $\phi(x) = 0$  si  $x < 0$ .



Usando dos hileras de neuronas, podemos construir regiones de decisión convexas delimitadas por rectas (intersectando semiplanos, cada uno de los cuales está determinado por una neurona) y usando tres hileras, se puede formar cualquier región cuya frontera la formen un conjunto finito de segmentos de recta [7].



¿De qué manera podemos generalizar el algoritmo de aprendizaje para redes de varias capas? Una manera es con el método de retropropagación [6], [7].

Consideremos la red de tres capas

$$\bar{x}, u_i \in \mathbb{R}^m, \bar{y}, v_j \in \mathbb{R}^n, \bar{z}, w \in \mathbb{R}^p$$

donde

$$y_i = g(u_i \cdot \bar{x}), \quad i=1, \dots, n$$

$$z_j = g(v_j \cdot \bar{y}), \quad j=1, \dots, p$$

$$\theta = g(\bar{w} \cdot \bar{z})$$

donde  $g$  es la función de activación

Si denotamos por  $G_S: \mathbb{R}^n \rightarrow \mathbb{R}^n$  la función  $(g(x_1), \dots, g(x_n))$

$A = (u_{ik})$ ,  $B = (v_{jl})$ , las matrices de los pesos de las primeras dos capas, y  $\bar{w}$  el vector de pesos de salida, podemos ver la red como una composición:

$$\bar{x} \rightarrow G_n(A\bar{x}) = \bar{y} \rightarrow G_p(B\bar{y}) = \bar{z} \rightarrow g(\bar{w} \cdot \bar{z}) = \theta$$

O si se prefiere, podemos escribir directamente

$$R(\bar{x}) = g\left(\sum_{j=1}^p w_j g\left(\sum_{i=1}^n v_{ji} g\left(\sum_{k=1}^m u_{ik} x_k\right)\right)\right)$$

Aclarada la dependencia de  $R$  respecto de los pesos y las entradas, podemos plantear la función de energía que queremos minimizar:

$$E = \frac{1}{2} \sum_{\bar{a} \in A} (f(\bar{a}) - R(\bar{a}))^2$$

donde  $f(\bar{a})$  es la salida deseada para  $\bar{a}$  y  $R(\bar{a})$  es la salida obtenida con la red usando los parámetros  $A$ ,  $B$  y  $\bar{w}$ .

Queremos saber en qué dirección modificar los vectores  $\bar{w}$ ,  $\bar{v}_j$  y  $\bar{u}_i$  de manera que disminuya la "energía".

Para modificar  $\bar{w}$ :

$$\frac{\partial E}{\partial w_j} = \sum_{\bar{a} \in A} (f(\bar{a}) - R(\bar{a})) (-g'(\bar{w} \cdot \bar{z})) z_j$$

Si llamamos  $\delta_a$  a  $(f(\bar{a}) - R(\bar{a})) g'(\bar{w} \cdot \bar{z})$ , el incremento sería

$$\Delta_a \bar{w} = \delta_a \bar{z}$$

y entonces la entrada  $\bar{a}$  cambia a  $\bar{w}$  en la dirección de  $\bar{z}$ .

Para modificar cada  $\bar{v}_j$

$$\frac{\partial E}{\partial v_{ji}} = \sum_{a \in A} (f(\bar{a}) - R(\bar{a})) (-g'(\bar{w} \cdot \bar{z})) \frac{\partial \bar{w} \cdot \bar{z}}{\partial v_{ji}}$$

pero

$$\bar{w} \cdot \bar{z} = \sum_{j=1}^p w_j g(\bar{v}_j \cdot \bar{y})$$

entonces

$$\frac{\partial \bar{w} \cdot \bar{z}}{\partial v_{ji}} = w_j \frac{\partial}{\partial v_{ji}} g(\bar{v}_j \cdot \bar{y}) = w_j g'(\bar{v}_j \cdot \bar{y}) y_i$$

Por lo cual

$$\frac{\partial E}{\partial v_{ji}} = \sum_{a \in A} \delta_a w_j g'(\bar{v}_j \cdot \bar{y}) y_i$$

$$\text{Así, si } \delta_{aj} = \delta_a w_j g'(\bar{v}_j \cdot \bar{y}) y_i$$

el incremento será  $\Delta_a \bar{v}_j$  y  $\delta_{aj} \bar{y}$

Por último, para modificar  $\bar{u}_l$

$$\frac{\partial E}{\partial u_{lk}} = \sum_{a \in A} (f(\bar{a}) - R(\bar{a})) (-g'(\bar{w} \cdot \bar{z})) \frac{\partial \bar{w} \cdot \bar{z}}{\partial u_{lk}}$$

$$\begin{aligned} \text{como } \frac{\partial \bar{w} \cdot \bar{z}}{\partial u_{lk}} &= \sum_{j=1}^p \frac{\partial \bar{w} \cdot \bar{z}}{\partial z_j} \frac{\partial z_j}{\partial y_l} \frac{\partial y_l}{\partial u_{lk}} \\ &= \sum_{j=1}^p w_j g'(\bar{v}_j \cdot \bar{y}) y_l g'(\bar{u}_l \cdot \bar{x}) x_k \end{aligned}$$

Por lo tanto

$$\frac{\partial E}{\partial u_{lk}} = \sum_{a \in A} \sum_{j=1}^p \delta_{aj} y_l g'(\bar{u}_l \cdot \bar{x}) x_k$$

$$\text{Si denotamos } \delta_{ajl} = \sum_{j=1}^p \delta_{aj} y_l g'(\bar{u}_l \cdot \bar{x})$$

Entonces se propone  $\Delta_a \bar{u}_l = \delta_{ajl} \bar{x}$

### CAPITULO III "APROXIMACION" DE FRACTALES.

Consideremos el siguiente problema:

"Dado un conjunto compacto  $A \subseteq \mathbb{P}^n$  con la propiedad de ser igual a la union de un numero finito de subconjuntos  $A_k$  semejantes a  $A$ : encontrar un IFS cuyo atractor sea  $A$ ."

En este contexto, decimos que  $A$  es semejante a  $B$  si  $A$  es la imagen directa de  $B$  bajo alguna transformación afin invertible. Los fractales suelen tener la propiedad de autosemejanza, que quiere decir que no importa que tan pequeña sea la escala a la que estemos estudiando el fractal, sigue siendo igual que a la escala original.

Si no podemos construir un IFS cuyo atractor sea idéntico a  $A$ , pero que "se le parezca" tanto como queramos, diremos que estamos "aproximando" a  $A$ .

#### A) METODO DEL PUNTO LEJANO.

Esta es una propuesta de procedimiento que permita resolver el problema anterior, se elaboró basándose en las técnicas de las redes neuronales, y se puso a prueba con un fractal muy sencillo (para ser fractal), el conjunto de Cantor. A continuación lo presentamos.

Sea  $C = \bigcup_{k=1}^N C_k$  el conjunto que queremos que sea el atractor del IFS.

Para cada  $C_k \subseteq \mathbb{P}^n$  tomemos una transformación  $w_k: \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $w_k(x) = A_k x + b_k$ .

Si podemos determinar  $A_k$  y  $b_k$  de manera que  $w_k(C) = C_k$ , entonces el conjunto  $C$  satisficará  $C = \bigcup_{k=1}^N w_k(C)$ , que es la ecuación que determina al atractor de un IFS.

Vamos a desarrollar un método en el cual partiremos de una estimación "a priori" de cada  $A_k$  y  $b_k$  compararemos  $w_k(C)$  con  $C_k$  y con esta información estimaremos "a posteriori" los parámetros, con el objetivo de reducir el error.



Para esto necesitamos establecer una medida del error correspondiente a cada elección de los parámetros.

Sea  $E = \sum_{k=1}^N h(w_k(C), C_k)$ , donde  $h(A, B) = \max(D(A, B), D(B, A))$ , y  $D(A, B)$  es la distancia "del punto más lejano de A a B", y  $D(B, A)$  es la distancia "del punto más lejano de B a A". Se puede demostrar que  $h$  es una métrica sobre el conjunto de los subconjuntos compactos no vacíos de  $\mathbb{R}^n$  (ver el apéndice).

Entonces el problema se resuelve si minimizamos una función que depende solamente de los parámetros (Pues  $C$  está fijo), como cada  $A_k \in \mathbb{R}^n \times \mathbb{P}^n$ ,  $b_k \in \mathbb{P}^n$ , y  $k=1, \dots, N$  tenemos que  $E: \mathbb{P}^{n(n+1)N} \rightarrow \mathbb{R}$ .

Sin embargo, como no podemos calcular su derivada (en caso de que exista), no le podemos aplicar los métodos clásicos del cálculo diferencial, de ahí la necesidad de enunciar este problema desde otra óptica.

Si se trata de hacer que cada  $w_k(C)$  coincida con  $C_k$  o en otras palabras, de que a cada entrada  $x \in C$  le corresponda una cierta salida  $y \in C_k$  estamos ante un problema de "aprendizaje" en el cual buscamos que  $y = A_k x + b_k$ .

Si desglosamos la ecuación vectorial anterior en cada componente, tenemos que  $y_j = \sum_{i=1}^n a_{ij} x_i + b_j$ , entonces podemos ver a  $y_j$  como la salida de una neurona que tiene  $x_1, \dots, x_n$  por  $n$  entradas,  $\vec{a}_j = (a_{j1}, \dots, a_{jn})$  como vector de pesos, un umbral  $-b_j$  y a  $g(x) = x$  como función de activación. Entonces el problema de ajustar los pesos y los umbrales de manera que se produzca la asociación deseada, es el mismo que el problema de minimizar  $E$ . Parecería que no podemos aplicar la técnica del gradiente descendente porque el conjunto de entradas  $C$  es infinito (si  $C$  fuera finito sería trivial) y porque no tenemos especificada para cada  $x \in C$  cual  $z \in C_k$  asociarle, pero al evaluar  $h$  tenemos que  $h(w_k(C), C_k) = d(y, z)$  para algunos  $y \in w_k(C)$ ,  $z \in C_k$ .

Por simplicidad denotamos  $v_j = (\bar{a}_j, b_j) \cdot (x, 1)$ , donde  $x = (x_1, \dots, x_n)$ ,  $y = (y_1, \dots, y_n)$ ,  $z = (z_1, \dots, z_n)$

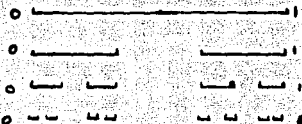
Para aplicar el método del gradiente descendente, nos fijamos en los puntos  $y, z$  mencionados anteriormente y comparamos  $z_j - y_j$ . Entonces procede modificar  $(\bar{a}_j, b_j)$  en la dirección de un vector  $\Delta = c(x, 1)$

Para elegir  $c$  de modo que aproxime lo más posible  $y$  a  $z$ , hacemos

$$\begin{aligned}(\bar{a}_j + c x, b_j + c) \cdot (x, 1) &= z_j \\ (\bar{a}_j, b_j) \cdot (x, 1) + c(x, 1) \cdot (x, 1) &= z_j \\ y_j + c(x_1^2 + \dots + x_n^2 + 1) &= z_j \\ c &= (z_j - y_j) / (x_1^2 + \dots + x_n^2 + 1).\end{aligned}$$

Y entonces transformamos  $(\bar{a}_j, b_j)$  en  $(\bar{a}_j, b_j) + \Delta$ .

Ilustraremos el procedimiento aplicándolo en uno de los casos más sencillos, el conjunto de Cantor. Recordemos que es la intersección de conjuntos de la forma  $C_0 = [0, 1]$ ,  $C_1 = [0, 1/3] \cup [2/3, 1]$ ,  $C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$ , ... donde el siguiente conjunto se puede construir a partir del anterior "quitándole" la parte de enmedio de cada uno de los intervalos que lo componen.



Observamos que es igual a dos copias reducidas de  $e_1$ . Por eso trabajaremos con dos transformaciones de la forma  $x \rightarrow ax + b$ . Sean  $T_1(x) = 4x$ ,  $U_1(x) = 3x + 7$ .

Como el proceso es independiente en cada función, lo aplicaremos primero a  $T_1$ .

Comparando  $C_1$  y  $T_1(C)$  comprobamos que  $z=.33$ ,  $x=1$ ,  $y=T_1(x)=.4$

$$\Delta b = (.33 - .4) / 2 = -.07 / 2 = -.035$$

$$\Delta a = x \Delta b = -.035$$

$$\text{Entonces } T_2(x) = .37x - .035$$

Comparando  $C_2$  y  $T_2(C)$  tenemos que  $z=0$ ,  $x=0$ ,  $T_2(x)=-.035$

$$\Delta b = (0 - (-.035)) / 1 = .035$$

$$\Delta a = x \Delta b = 0$$

$$\text{Entonces } T_3(x) = .37x$$

Comparando  $C_3$  y  $T_3(C)$  obtenemos  $z=.33$ ,  $x=1$ ,  $T_3(x)=.37$

$$\Delta b = (.33 - .37) / 2 = -.04 / 2 = -.02$$

$$\Delta a = x \Delta b = -.02$$

$$\text{Por lo cual } T_4(x) = .35x - .02$$

Pasemos ahora con  $C_4$  y  $W_4(C)$ , la distancia la evaluamos en

$$z=.66, x=0, W_4(x)=.7$$

así,  $\Delta b = (.66 - .7) / 1 = -.04$ ;  $\Delta a = x \Delta b = 0$ .

$$\text{Entonces, } W_5(x) = .3x + .66$$

Comparando  $C_5$  con  $W_5(C)$ , resulta que  $z=1$ ,  $x=1$ ,  $W_5(x) = (.66 - .7) / 2$ ,

entonces

$$\Delta b = (1 - .96) / 2$$

$$\Delta a = x \Delta b$$

$$\text{Por lo cual, } W_6(x) = .32x + .66$$

Si ahora comparamos  $C_6$  y  $W_6(C)$ , hallamos que  $z=.66$ ,  $x=0$ ,  $W_6(x)=.66$ .

entonces

$$\Delta b = (.66 - .68) / 1 = -.02$$

$$\Delta a = x \Delta b = 0$$

$$\text{De donde } W_7(x) = .32x + .66$$

Al considerar cada vez más iteraciones, vemos que la sucesión de funciones  $W_n$  converge a  $W(x) = .33x + .66$  y  $T_n \rightarrow T(x) = .33x$ , que conforman el IFS buscado.

## B) ESTABILIDAD DE UN IFS. CONCLUSIONES.

Para finalizar analizaremos la dependencia de los IFS respecto a pequeños cambios en los parámetros. Comenzaremos retomando un teorema que nos muestra la dependencia de los atractores (que son "puntos fijos") respecto de los parámetros.

**Teorema.** "Los puntos fijos dependen continuamente de los parámetros"  
Sea  $(X, d)$  un espacio métrico completo (como  $(\mathbb{R}^n, \|\cdot\|)$ , véase el apéndice) y sea  $(P, d_p)$  otro espacio métrico (por ejemplo, el de los parámetros).

Sea  $w: P \times X \rightarrow X$  una familia de contracciones en  $X$  cada una con factor de contracción  $0 \leq s < 1$ . (Esto significa que cada  $w(p, \cdot)$  es una contracción en  $X$ )

Si para cada  $x \in X$  fijo,  $w$  es continua en  $P$ , entonces el punto fijo correspondiente a  $w$  depende continuamente de  $p \in P$ . En otras palabras, la función  $x_f: P \rightarrow X$  que a cada vector de parámetros  $p$  le asocia el punto fijo de la contracción  $w(p, \cdot)$ , esta función es continua.

Por demostrar:  $\forall \epsilon > 0$  existe  $\delta > 0$  tal que  
 $d_p(p, q) < \delta \Rightarrow d(x_f(p), x_f(q)) < \epsilon$ .

Demostración.

Sean  $p, q \in P$  y  $\epsilon > 0$  cualesquiera.

Sea  $x_f(p)$  el punto fijo de  $w$  dado  $p$ .

Entonces para toda  $q \in P$  se tiene que

$$\begin{aligned} d(x_f(p), x_f(q)) &= d(w(p, x_f(p)), w(q, x_f(q))) \text{ por ser puntos fijos} \\ &\leq d(w(p, x_f(p)), w(q, x_f(p))) + d(w(q, x_f(p)), w(q, x_f(q))) \\ &\quad \text{por la desigualdad del } \Delta \\ &\leq d(w(p, x_f(p)), w(q, x_f(p))) + s d(x_f(p), x_f(q)) \\ &\quad \text{por ser } w(q, \cdot) \text{ contracción} \end{aligned}$$

despejando se obtiene que

$$\begin{aligned} d(x_f(p), x_f(q))(1-s) &\leq d(w(p, x_f(p)), w(q, x_f(p))) \\ d(x_f(p), x_f(q)) &\leq \frac{1}{1-s} d(w(p, x_f(p)), w(q, x_f(p))) \end{aligned}$$

Por lo tanto, si  $q$  está suficientemente cercano a  $p$ , por ser  $w$  continua en  $P$  cuando  $x$  está fijo, podemos hacer  $d(w(p, x_f(p)), w(q, x_f(p))) < \epsilon(1-s)$  y entonces existe  $\delta > 0$  tal que si  $d(p, q) < \delta$ , entonces  $d(x_f(p), x_f(q)) < \epsilon$ .

□

A manera de ejemplo, podemos considerar  $w(x) = Ax + b$ , sea  $x_f$  punto fijo, es decir,  $Ax_f + b = x_f$ , o equivalentemente,  $(A - I)x_f = -b$ . Si  $w$  es contracción,  $\lambda = 1$  no es valor propio y por lo tanto  $\det(A - I)$  es distinto de cero, entonces  $(A - I)$  es invertible y podemos despejar al punto fijo  $x_f = -(A - I)^{-1}b$ . Claramente depende continuamente de los parámetros.

Más aún, en los IFS compuestos por funciones del tipo  $w(x) = Ax + b$ , con  $A$  matriz,  $x, b$  vectores, el atractor es función continua de los parámetros.

Si aplicamos el teorema anterior haciendo  $(X, d) = (H, h)$ ,  $(P, d_p) = \mathbb{R}^{Nn(n+1)}$ ,  $x_f =$  atractor del IFS, podemos concluir que si perturbamos ligeramente los componentes de las matrices y vectores de un IFS de transformaciones afines, los atractores no se diferenciarán mucho en el sentido de la métrica  $h$ .

Más aún, si tomamos en cuenta que  $d(x, f(x)) \leq d(x, x_f) + d(x_f, f(x))$   
 $\leq d(x, x_f) + sd(x_f, x)$   
 $\leq (1+s)d(x, x_f)$   
 $\leq 2d(x, x_f)$

podemos concluir que  $d(x, f(x)) < \epsilon$  cuando  $d(x, x_f) < \delta = \epsilon/2$

Por lo tanto el error  $E = \sum_{k=1}^N h(w_k(C), C_k)$  como función de  $A_1, \dots, A_n, b_1, \dots, b_n$  es continuo en el punto donde alcanza su mínimo.

Esto nos da una base para justificar que podemos reducir el error tanto como queramos si los parámetros que proponemos están suficientemente cercanos a la solución.

En conclusión:

1) Los atractores de los IFS son localmente estables en el sentido de que no son muy sensibles a pequeños cambios en los parámetros.

2) Es plausible la convergencia (no demostrada) del método que propusimos en la sección anterior, bajo la hipótesis de que el fractal con el cual trabajamos es el atractor de algún IFS de transformaciones afines.

Queda abierta la posibilidad de encontrar nuevos vínculos entre los fractales, los sistemas dinámicos (en especial los de Lindenmayer) y las técnicas inspiradas en las redes neuronales.

## APENDICE

### 1 ESPACIOS METRICOS

**ESPACIO METRICO.** Sea  $E$  un conjunto donde está definida una función  $d: E \times E \rightarrow \mathbb{R}$  con las siguientes propiedades para todo  $x, y, z \in E$ :

- 1)  $d(x, y) \geq 0$  y  $d(x, y) = 0 \Leftrightarrow x = y$
- 2)  $d(x, y) = d(y, x)$
- 3)  $d(x, y) \leq d(x, z) + d(z, y)$

Se dice que  $(E, d)$  es un espacio métrico y  $d$  se llama métrica.

Por ejemplo,  $\mathbb{R}^3$  con la distancia euclídeana, es un espacio métrico.

En un espacio métrico tienen sentido algunas nociones topológicas:

**Vecindad.** El conjunto  $V_c(x_0) = \{y \in E \text{ tales que } d(x_0, y) < c\}$  está bien definido en cualquier espacio métrico  $(E, d)$ , y se llama vecindad de radio  $c$  centrada en  $x_0$  (siempre y cuando  $x_0 \in E$ ).

**Conjunto acotado.** Se dice que un conjunto es acotado si existe alguna vecindad que lo contiene.

**Abierto.** Se dice que un conjunto  $A$  es abierto si para todo  $x \in A$  existe alguna vecindad  $V_c(x)$  contenida completamente en  $A$ .

**Cerrado.** Se dice que un conjunto es cerrado si su complemento es abierto.

**Convergencia de sucesiones.** Sea  $(E, d)$  un espacio métrico y  $L \in E$ . Una sucesión  $x_n: \mathbb{N} \rightarrow E$  converge a  $L$  si para cualquier  $\epsilon > 0$  existe un término  $x_n$  (que depende de  $\epsilon$ ) a partir del cual todos los términos siguientes de la sucesión están en  $V_\epsilon(L)$ .  $L$  se llama el límite de la sucesión.

**Sucesiones de Cauchy.** Una sucesión  $x_n: \mathbb{N} \rightarrow E$ , es de Cauchy si para toda  $\epsilon > 0$  existe un término  $x_n$  (que depende de  $\epsilon$ ) a partir del cual los términos siguientes distan entre sí menos que  $\epsilon$ .

**Conjunto compacto.** Un conjunto  $K$  es compacto si toda sucesión contenida en  $K$  tiene alguna subsucesión convergente.

**Conjunto totalmente acotado.** Un conjunto  $A$  está totalmente acotado si para toda  $\epsilon > 0$  existe un conjunto finito  $\{y_1, \dots, y_n\} \subseteq A$  tales que  $A \subseteq \bigcup_{i=1}^n V_\epsilon(y_i)$ .

**Función continua.** Sean  $(X, d_1)$ ,  $(Y, d_2)$  espacios métricos. Una función  $f: X \rightarrow Y$  es continua si para todo conjunto abierto  $A \subseteq Y$ , el conjunto  $f^{-1}(A) = \{x \in X \text{ tales que } f(x) \in A\}$  es abierto.

Algunos resultados de espacios métricos nos indican que:

- Si una función  $f: K \rightarrow \mathbb{R}$ , es continua en  $K$  un conjunto compacto, alcanza su valor máximo y su mínimo.
- Un conjunto  $K$  es compacto  $\Leftrightarrow$  es cerrado y totalmente acotado.
- Un conjunto  $F$  es cerrado  $\Leftrightarrow$  para toda sucesión convergente que consiste de elementos de  $F$ , el límite está en  $F$ .
- Toda sucesión convergente es de Cauchy.

Se dice que un espacio métrico  $(X, d)$  es **completo** si toda sucesión de Cauchy converge a algún elemento de  $X$ .

-Si en un espacio métrico completo  $(X, d)$ ,  $f: X \rightarrow X$  tiene la propiedad de que  $d(f(x), f(y)) < kd(x, y)$  con  $0 < k < 1$  para todo  $(x, y) \in X \times X$ , entonces la sucesión  $x_0 = a$ ,  $x_{n+1} = f(x_n)$  converge sin importar cual sea  $x_0$ ; y además el límite es el único punto fijo de  $f$ .



## 2 A) $(H(X), h(d))$

Si  $(X, d)$  es un espacio métrico, definimos  $H(X)$  como el conjunto de los subconjuntos compactos no vacíos de  $X$ , y si  $A, B \in H(X)$ , sean  $d(x, B) = \inf\{d(x, y) \mid y \in B\}$ ,  $D(A, B) = \sup\{d(x, B) \mid x \in A\}$  y  $h(A, B) = \max\{D(A, B), D(B, A)\}$ . Observamos que se alcanzan el ínfimo y el supremo porque  $A$  y  $B$  son compactos.

Parecería suficiente  $D$  para definir una distancia, pero en  $D(A, B)$  puede no ser igual que  $D(B, A)$ , por ejemplo si  $(X, d)$  son los números reales con la distancia usual, si  $A = [0, 2]$  y  $B = [3, 4]$ ,  $d(x, B) = \min\{|x - y| \mid y \in B\} = 3 - x$  si  $x < 3$ . Entonces  $D(A, B) = \max\{d(x, B) \mid x \in A\} = \max\{3 - x \mid x \in A\} = 3$ . Sin embargo,  $d(y, A) = \min\{|y - a| \mid a \in A\} = y - 2$  si  $y > 2$ . Entonces  $D(B, A) = \max\{d(y, A) \mid y \in B\} = \max\{y - 2 \mid y \in B\} = 2$ .

## 2 B) $(H(X), h)$ ES ESPACIO METRICO

Para esto tenemos que probar las siguientes propiedades:

- i)  $h(A, B) \geq 0$  y  $h(A, B) = 0 \iff A = B$ .
- ii)  $h(A, B) = h(B, A)$
- iii)  $h(A, B) \leq h(A, C) + h(C, B)$

Para cualesquiera  $A, B, C \subseteq X$  compactos y no vacíos.

- i)  $d(x, y) \geq 0$  implica que  $d(x, B) \geq 0$ , entonces  $D(A, B) \geq 0$  y  $h(A, B) \geq 0$ .

Si  $A = B$ , para cualquier  $x \in A$  sucede que  $d(x, B) = 0$  y  $d(x, A) = 0$  pues  $d(x, x) = 0$ , entonces  $D(A, B) = \max\{d(x, B) \mid x \in A\} = \max\{0\} = 0$  y  $D(B, A) = \max\{d(x, A) \mid x \in B\} = \max\{0\} = 0$  y entonces  $h(A, B) = 0$ .

Si  $h(A, B) = 0$ , entonces  $D(A, B) = 0$  y  $D(B, A) = 0$ .

Vamos a demostrar que  $A = B$ , es decir, que  $A \subseteq B$  y  $B \subseteq A$ .

ESTA TESIS NO DEBE  
SALIR DE LA BIBLIOTECA

2 C) SI  $(X, d)$  ES UN ESPACIO METRICO COMPLETO,  $(h(X), h)$  TAMBIEN

Para este resultado utilizaremos algunos lemas preeliminarios.

Sean  $S \subseteq X$  y  $\alpha > 0$ , el conjunto  $\{x \in X : d(x, y) \leq \alpha \text{ para alguna } y \in S\}$  se llama la dilatación de  $S$  por una bola de radio  $\alpha$  y lo denotamos por  $Dil_\alpha(S)$ . Nótese que para  $\alpha = 0$ ,  $SDil_\alpha(S)$ .

Lema 1.

Para cualesquiera  $A, B \in H(X)$ ,  $h(A, B) \leq \alpha \iff A \subseteq SDil_\alpha(B)$  y  $B \subseteq SDil_\alpha(A)$ .

Demostración.

$\max\{D(A, B), D(B, A)\} = h(A, B) \leq \alpha \iff D(A, B) \leq \alpha \text{ y } D(B, A) \leq \alpha$ .

Basta con demostrar que  $D(A, B) \leq \alpha \iff A \subseteq SDil_\alpha(B)$ .

$D(A, B) \leq \alpha \iff \max\{d(a, B) : a \in A\} \leq \alpha$

$\iff d(a, B) \leq \alpha \text{ para toda } a \in A$ .

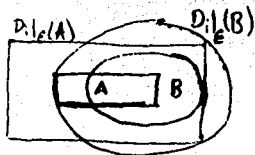
Pero  $d(a, B) = \min\{d(a, b) : b \in B\} = d(a, y)$  para algún  $y \in B$

Esto quiere decir que toda  $a \in A$  está en  $Dil_\alpha(B)$ .

En otras palabras  $A \subseteq SDil_\alpha(B)$ .

Análogamente,  $D(B, A) \leq \alpha \iff B \subseteq SDil_\alpha(A)$ .

Entonces  $h(A, B) \leq \alpha \iff A \subseteq SDil_\alpha(B)$  y  $B \subseteq SDil_\alpha(A)$ .



Notamos que este lema nos da una interpretación gráfica de  $h$ .

Para demostrar que  $A \subseteq B$ , sea  $x \in A$ .

$0 \leq d(x, B) = \max\{d(x, b) \mid b \in B\} = D(A, B) = 0$ , entonces  $d(x, B) = 0$ .

$d(x, B) = \min\{d(x, b) \mid b \in B\} = 0 \rightarrow$  Existe un  $b \in B$  tal que  $d(x, b) = 0$ , pero como  $d$  es métrica, entonces  $x = b$ . Por lo tanto,  $x \in B$ .

Para probar que  $B \subseteq A$ , sea  $y \in B$ .

$0 \leq d(y, A) = \max\{d(y, a) \mid a \in A\} = D(B, A) = 0$ , entonces  $d(y, A) = 0$ .

$d(y, A) = \min\{d(y, a) \mid a \in A\} = 0 \rightarrow$  Existe un  $a \in A$  tal que  $d(y, a) = 0$ , pero por ser  $d$  una métrica, entonces  $y = a$ . Por lo tanto  $y \in A$ .

Entonces  $h(A, B) = 0 \rightarrow A = B$ .

ii)  $h(A, B) = \max\{D(A, B), D(B, A)\} = \max\{D(B, A), D(A, B)\} = h(B, A)$ .

iii) Por demostrar:  $h(A, B) \leq h(A, C) + h(C, B)$ .

Para cualesquiera  $a \in A$  y  $c \in C$  fijos, ocurre que

$$\begin{aligned} d(a, B) &= \min\{d(a, b) \mid b \in B\} \\ &\leq \min\{d(a, c) + d(c, b) \mid b \in B\} \text{ porque } d \text{ es métrica} \\ &= d(a, c) + \min\{d(c, b) \mid b \in B\} \text{ pues } a \text{ y } c \text{ están fijos} \\ &= d(a, c) + d(c, B) \end{aligned}$$

En particular para  $c = c_0$  donde  $d(a, c)$  alcanza su mínimo dejando

fijo  $a$ . Como  $d(a, C) = \min\{d(a, c) \mid c \in C\} = d(a, c_0)$ , entonces

$$d(a, B) \leq d(a, c_0) + d(c_0, B) = d(a, C) + d(c_0, B)$$

Si elegimos el  $a \in A$  donde  $d(a, B)$  alcanza su máximo, entonces

$$\begin{aligned} D(A, B) &= d(a_0, B) \leq d(a_0, C) + d(c_0, B) \\ &\leq \max\{d(x, C) \mid x \in A\} + \max\{d(z, B) \mid z \in C\} \\ &= D(A, C) + D(C, B) \\ &\leq h(A, C) + h(C, B) \end{aligned}$$

Análogamente  $D(B, A) \leq D(B, C) + D(C, A) \leq h(B, C) + h(C, A)$

Entonces  $\max\{D(A, B), D(B, A)\} \leq h(A, C) + h(C, B)$

$$\therefore h(A, B) \leq h(A, C) + h(C, B)$$

Lema 2.

Sean  $\{A_n\}$  una sucesión de Cauchy en  $M(X)$  y  $\{A_{n_j}\}$  una subsucesión de la primera. Si existe  $\{x_{n_j}\}$ , una sucesión de Cauchy en  $X$  tal que  $x_{n_j} \in A_{n_j}$ , entonces podemos extenderla a una sucesión de Cauchy  $\{\hat{x}_n\}$  tal que  $\hat{x}_n \in A_n$  y  $\hat{x}_{n_j} = x_{n_j}$ .

Demostración.

Vamos a construir  $\{\hat{x}_n\}$  eligiendo  $\hat{x}_n = x_{n_j}$  para  $n = n_j$ , y adecuadamente para  $n \neq n_j$ .

Para  $n \in \{1, \dots, n_1 - 1\}$  elegimos  $\hat{x}_n$  en  $\{x \in A_n \mid d(x, x_{n_1}) = d(x_{n_1}, A_n)\}$ , este conjunto es no vacío porque es el conjunto de puntos donde  $f: A_n \rightarrow \mathbb{R}$ ,  $f(x) = d(x_{n_1}, x)$  alcanza su valor mínimo.

Similarmente para  $j > 1$  y  $n \in \{n_{j-1} + 1, n_{j-1} + 2, \dots, n_j - 1\}$  elegimos  $\hat{x}_n$  en  $\{x \in A_n \mid d(x, x_{n_j}) = d(x_{n_j}, A_n)\}$  que se no vacío porque es donde  $d(x, x_{n_j})$  como función de  $A_n$  alcanza su mínimo.

Afirmamos que  $\{\hat{x}_n\}$  cumple la propiedad prometida.

Por construcción  $\hat{x}_{n_j} = x_{n_j}$ , y  $\hat{x}_n \in A_n$ .

Para ver que es una sucesión de Cauchy

Sea  $\epsilon > 0$ ; entonces existen

$N_1$  tal que  $n, m > N_1 \Rightarrow d(x_{n_j}, x_{m_j}) < \epsilon/3$  (por ser  $\{x_{n_j}\}$  de Cauchy)

$N_2$  tal que  $m, n > N_2 \Rightarrow h(A_m, A_n) < \epsilon/3$  (por ser  $\{A_n\}$  de Cauchy)

Sea  $N = \max\{N_1, N_2\}$ .

Si  $n, m > N$ , y elegimos  $j, k$  tales que  $m \in \{n_{j-1} + 1, \dots, n_j\}$  y  $n \in \{n_{k-1} + 1, \dots, n_k\}$  entonces

$$\begin{aligned} d(\hat{x}_m, \hat{x}_n) &\leq d(\hat{x}_m, x_{n_j}) + d(x_{n_j}, x_{m_j}) + d(x_{m_j}, \hat{x}_n) \\ &< d(\hat{x}_m, x_{n_j}) + d(x_{n_j}, \hat{x}_n) + \epsilon/3 \end{aligned} \quad (*)$$

Falta argumentar por qué  $d(\hat{x}_m, x_{n_j}) < \epsilon/3$ .

Sabemos que  $h(A_m, A_n) < \epsilon/3$ ,

aplicando el lema 1 tenemos que  $A_n \subseteq \text{Dil}_{\epsilon/3}(A_m)$

esto quiere decir que todo  $a \in A_n$ , cumple que  $d(a, a_m) \leq \epsilon/3$  para alguna  $a_m \in A_m$ . En particular  $x_{n_j} \in A_n$ , y por lo tanto  $d(x_{n_j}, b_m) < \epsilon/3$  para alguna  $b_m \in A_m$ .

Entonces  $d(x_{n_j}, A_m) = \min\{d(x_{n_j}, a_m) \mid a_m \in A_m\} \leq d(x_{n_j}, b_m) < \epsilon/3$

Pero  $\hat{x}_m$  fue elegido de modo que  $d(\hat{x}_m, x_{n_j}) = d(x_{n_j}, A_m)$ .

Por lo tanto,  $d(\hat{x}_m, x_{n_j}) < \epsilon/3$ .

Similarmente  $d(x_{n_k}, \hat{x}_n) < \epsilon/3$ .

Aplicando esto a (\*) obtenemos que  $d(\hat{x}_m, \hat{x}_n) < \epsilon$  si  $m, n > N$ .

QED.

lema 3.  $\text{Dil}_{\epsilon}(A)$  es cerrado si  $A$  es compacto.

Demostración.

Supongamos que  $\text{Dil}_{\epsilon}(A)$  no es cerrado,

entonces  $(\text{Dil}_{\epsilon}(A))^c$  no es abierto,

esto quiere decir que existe un  $x \in \text{Dil}_{\epsilon}(A)$  tal que en toda vecindad hay elementos de  $\text{Dil}_{\epsilon}(A)$ .

En particular, para las vecindades de radio  $1/n$ .

Entonces existen  $y_n$  tales que  $y_n \in \text{Dil}_{\epsilon}(A)$  y  $y_n \in V_{1/n}(x)$ .

Es decir, existe una sucesión  $\{y_n\}$  tal que para cada  $n$ ,

$d(y_n, x) \leq 1/n$  y  $d(y_n, z_n) \leq \epsilon$  para algún  $z_n \in A$ .

Por la desigualdad del triángulo,

$d(x, z_n) \leq d(x, y_n) + d(y_n, z_n) \leq \epsilon + 1/n$

Por lo tanto,  $\inf\{d(x, z) \mid z \in A\} \leq \epsilon$ , pero como  $A$  es compacto, y  $f_x: A \rightarrow \mathbb{R}$  definida por  $f_x(z) = d(x, z)$  es continua, alcanza su mínimo y

entonces existe  $z \in A$  tal que  $d(x, z) \leq \epsilon$ .

En consecuencia,  $x \in \text{Dil}_{\epsilon}(A)$  lo cual contradice la suposición inicial. Por lo tanto,  $\text{Dil}_{\epsilon}(A)$  es cerrado. QED.

Ahora demostraremos que  $(M(X), h)$  es un espacio métrico completo si  $(X, d)$  lo es. Más aún, si  $\{A_n\} \subseteq M(X)$  es una sucesión de Cauchy, entonces

$$A = \lim_{n \rightarrow \infty} (A_n) \in M(X)$$

y se puede caracterizar como sigue:

A =  $\{x \in X \mid \text{existe una sucesión } \{x_n\} \text{ que converge a } x \text{ con } x_n \in A_n\}$

Demostración.

Sea  $\{A_n\}$  una sucesión de Cauchy en  $M(X)$ . Probaremos que:

i)  $A \neq \emptyset$ .

ii) A es cerrado.

iii)  $\forall \epsilon > 0 \exists N \in \mathbb{N}$  tal que  $n \geq N \Rightarrow A \subseteq \text{Dil}_\epsilon(A_n)$ .

iv) A es totalmente acotado y (por ii) compacto.

v)  $\lim_{n \rightarrow \infty} (A_n) = A$ .

i) Basta demostrar que existe una sucesión  $\{a_n\}$  de Cauchy en X tal que  $a_n \in A_n$ , porque como X es completo,  $\{a_n\}$  será convergente y su límite estará en A. ( $\therefore A \neq \emptyset$ ).

Sean  $N_1 < N_2 < \dots$  tales que  $h(A_n, A_m) < \frac{1}{2^i}$  para  $n, m \geq N_i$

Sea  $x_{N_1} \in A_{N_1}$ , como  $h(A_{N_1}, A_{N_2}) < \frac{1}{2}$  podemos encontrar un  $x_{N_2} \in A_{N_2}$  tal

que  $d(x_{N_1}, x_{N_2}) < \frac{1}{2}$ . (Existe un elemento así, porque si no,

$d(x_{N_1}, a) \geq \frac{1}{2}$  para todos los  $a \in A_{N_2}$  y entonces  $d(x_{N_1}, A_{N_2}) =$

$\min\{d(x_{N_1}, a) \mid a \in A_{N_2}\} \geq \frac{1}{2}$ . Entonces  $D(A_{N_1}, A_{N_2}) = \max\{d(b, A_{N_2}) \mid b \in A_{N_1}\} \geq \frac{1}{2}$ .

Y como  $h(A_{N_1}, A_{N_2}) = \max\{D(A_{N_1}, A_{N_2}), D(A_{N_2}, A_{N_1})\}$ ,  $h(A_{N_1}, A_{N_2}) \geq \frac{1}{2}$  lo cual contradice la elección de  $N_1$ .

Inductivamente, si ya elegimos  $x_{N_1}, x_{N_2}, \dots, x_{N_k}$  con cada  $x_{N_i} \in A_{N_i}$  los cuales cumplen  $d(x_{N_i}, x_{N_{i+1}}) < \frac{1}{2^i}$ , podemos añadir uno más a la lista, pues  $h(A_{N_k}, A_{N_{k+1}}) < \frac{1}{2^k}$  y  $x_{N_k} \in A_{N_k}$  implican que existe un  $x_{N_{k+1}} \in A_{N_{k+1}}$  tal que  $d(x_{N_k}, x_{N_{k+1}}) < \frac{1}{2^k}$  (De no ser así, mediante un razonamiento como el anterior llegaríamos a una contradicción). Entonces podemos construir una sucesión  $\{x_{N_i}\}$  con  $x_{N_i} \in A_{N_i}$  y  $d(x_{N_i}, x_{N_{i+1}}) < \frac{1}{2^i}$ .

Vamos a ver que ésta es una sucesión de Cauchy en  $X$ .

Sea  $\epsilon > 0$ , elegimos  $N$  tal que  $\frac{1}{2^{N-1}} < \epsilon$ .

Para  $n > m > N$  sucede que

$$d(x_{N_m}, x_{N_n}) \leq d(x_{N_m}, x_{N_{m+1}}) + d(x_{N_{m+1}}, x_{N_{m+2}}) + \dots + d(x_{N_{n-1}}, x_{N_n}) \\ < \sum_{i=m}^{n-1} \frac{1}{2^i} = \frac{1}{2^{m-1}} < \epsilon.$$

Aplicando el lema 2, existe  $\{a_i\}$  una sucesión con las propiedades siguientes: 1)  $a_i \in A_i$ ,

2)  $a_{N_i} = x_{N_i}$

3) es sucesión de Cauchy en  $X$  ( $\therefore$  converge)

ii) Para mostrar que  $A$  es cerrado, sea  $\{a_n\} \rightarrow a$  con  $a_n \in A$ .

Por demostrar: Que  $a \in A$ .

Como  $\{a_n\} \rightarrow a$ , podemos construir una sucesión creciente de enteros  $N_1 < N_2 < \dots$  de manera que  $d(a_{N_k}, a) < \frac{1}{k}$ .

Como cada  $a_{N_k} \in A$ ,  $a_{N_k} = \lim_{m \rightarrow \infty} (a_{N_k, m})$  para alguna sucesión convergente tal que cada  $a_{N_k, m} \in A_m$ , podemos hallar una sucesión creciente de enteros  $m_1 < m_2 < \dots$  tal que  $d(a_{N_k, m_k}, a_{N_k}) < \frac{1}{k}$ .

Entonces  $d(a_{N_k, m_k}, a) \leq d(a_{N_k, m_k}, a_{N_k}) + d(a_{N_k}, a) < \frac{2}{k}$ .

Si hacemos  $a_{n_1, n_2} = x_{n_1}$ , cada  $x_{n_1} \in A_{n_1}$  y  $\{x_{n_1}\} \rightarrow a$ .

Por el lema 2 podemos extenderla a una sucesión donde cada  $z_n \in A_n$ ,  $z_{n_1} = x_{n_1}$  y  $\{z_n\}$  resulta ser una sucesión de Cauchy con una subsucesión convergente a "a", por lo tanto,  $\{z_n\} \rightarrow a$ , y por lo tanto,  $a \in A$ .

iii)  $\forall \epsilon > 0 \exists N$  tal que  $n \geq N \Rightarrow a \in \text{Dil}_\epsilon(A_n)$

Sea  $\epsilon > 0$ ;

por ser  $\{A_n\}$  de Cauchy,  $\exists N_1$  tal que  $n, m \geq N \Rightarrow h(A_n, A_m) \leq \epsilon$

aplicando el lema 1,  $A_n \in \text{Dil}_\epsilon(A_n)$ .

Por demostrar:  $a \in \text{Dil}_\epsilon(A_n)$ .

Sea  $a \in A$ .

Entonces existe una sucesión  $\{a_k\} \rightarrow a$  con  $a_k \in A_k$ .

De modo que para esta  $\epsilon$ , existe una  $N_2$  tal que  $m \geq N_2 \Rightarrow d(a, a_m) < \epsilon$ .

Si elegimos  $N = \sup\{N_1, N_2\}$ , para  $m \geq N$  sucede que  $a_m \in \text{Dil}_\epsilon(A_n)$ . Como  $A_n$  es compacto,  $\text{Dil}_\epsilon(A_n)$  es cerrado (por el lema 3).

Ya que  $\{a_m\}$  es una sucesión convergente de elementos de un cerrado, y  $\{a_m\} \rightarrow a$ , entonces  $a \in \text{Dil}_\epsilon(A_n)$ .

iv) A es totalmente acotado.

Supongamos que A no es totalmente acotado.

Entonces para alguna  $\epsilon_0 > 0$  no existe ninguna  $\epsilon_0$ -red finita.

Por lo tanto existe una sucesión  $\{x_n\}$  en A tal que  $d(x_i, x_j) \geq \epsilon_0$  para  $i \neq j$ . Por iii) existe N tal que  $A \in \text{Dil}_{\delta}(A_n)$  con  $\delta = \epsilon_0/3$ .

Para cada  $x_i$  existe una  $y_i \in A_n$  tal que  $d(x_i, y_i) < \delta$ .

Como  $A_n$  es compacto, alguna subsucesión  $\{y_{n_i}\}$  de  $\{y_n\}$  converge, por lo que existen puntos en  $\{y_i\}$  tan cercanos como se quiera, en particular,  $y_{n_1}, y_{n_2}$  que distan menos que  $\delta = \epsilon_0/3$

Pero  $d(x_{n_1}, x_{n_2}) \geq d(x_{n_1}, y_{n_1}) + d(y_{n_1}, y_{n_2}) + d(y_{n_2}, x_{n_2}) < \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon$

Lo que contradice la elección de las x's.

Entonces A es totalmente acotado.

(Y aplicando ii) es compacto).



v)  $\lim_{n \rightarrow \infty} (A_n) = A$ .

Por iv) y i)  $A \in H(X)$ .

Ahora bien,  $\lim_{n \rightarrow \infty} (A_n) = A \Rightarrow \forall \epsilon > 0 \exists N$  tal que  $n > N \Rightarrow h(A, A_n) \leq \epsilon$ .

$\Rightarrow \forall \epsilon > 0 \exists N$  tal que  $n > N \Rightarrow A \subseteq \text{Dil}_\epsilon(A_n)$  y  $A_n \subseteq \text{Dil}_\epsilon(A)$ .

Por iii) ya tenemos una contención.

Sea  $\epsilon > 0$ , por ser  $\{A_n\}$  de Cauchy, existe una  $N$  tal que si  $n, m \geq N$ , entonces  $h(A_n, A_m) \leq \epsilon/2$  y por el lema 1,  $A_n \subseteq \text{Dil}_{\epsilon/2}(A_m)$ .

Por demostrar:  $A_n \subseteq \text{Dil}_\epsilon(A)$ .

Para  $\delta = \epsilon/4$ , existe  $N_1 > n$  tal que si  $m, k \geq N_1$ , entonces  $A_m \subseteq \text{Dil}_\delta(A_k)$ .

Para  $\delta = \epsilon/8$ , existe  $N_2 > N_1$  tal que  $m, k \geq N_2 \Rightarrow A_m \subseteq \text{Dil}_\delta(A_k)$ .

...

Para  $\delta = \frac{\epsilon}{2^{j+1}}$ , existe  $N_j > N_{j-1}$  tal que  $m, k \geq N_j \Rightarrow A_m \subseteq \text{Dil}_\delta(A_k)$ .

Observemos que  $A_n \subseteq \text{Dil}_{\epsilon/2^j}(A_{N_j})$  para todas las  $j$ 's. (\*\*)

Sea  $x \in A_n$ ,

Por (\*\*) existe  $x_{N_j} \in A_{N_j}$  tal que  $d(x, x_{N_j}) < \epsilon/2$ .

Como  $N_2 > N_1 \geq N_1$ ,  $A_{N_1} \subseteq \text{Dil}_{\epsilon/4}(A_{N_2})$  y entonces existe  $x_{N_2} \in A_{N_2}$  tal que  $d(x_{N_2}, x_{N_1}) < \epsilon/4$ .

De esta manera podemos construir una sucesión  $\{x_{N_j}\}$  tal que

$x_{N_j} \in A_{N_j}$  y  $d(x_{N_j}, x_{N_{j-1}}) < \frac{\epsilon}{2^{j+1}}$ .

Entonces

$$\begin{aligned} d(x, x_{N_j}) &\leq d(x, x_{N_1}) + d(x_{N_1}, x_{N_2}) + \dots + d(x_{N_{j-1}}, x_{N_j}) \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{4} + \dots + \frac{\epsilon}{2^{j+1}} < \epsilon \end{aligned}$$

Como  $\{x_{N_j}\}$  es de Cauchy,  $\{x_{N_j}\} \rightarrow a \in A$ .

Más aún, como  $d(x, x_{N_j}) < \epsilon$  para toda  $j$ , ocurre que  $d(x, a) \leq \epsilon$ .

Entonces  $A_n \subseteq \text{Dil}_\epsilon(A)$  para  $n > N$  (sirve la misma  $N$  que la que existe por ser  $\{A_n\}$  sucesión de Cauchy).

Entonces  $\lim_{n \rightarrow \infty} (A_n) = A$  para cualquier sucesión de Cauchy en  $(H(X), h)$ .

En otras palabras,  $(H(X), h)$  es completo.

### 3 UNA CONDICION PARA QUE UNA TRANSFORMACION LINEAL SEA CONTRACCION.

Las transformaciones lineales que van del plano en el plano tienen asociados ciertos valores  $\lambda$ , las soluciones de la ecuación  $\det(A-\lambda I)=0$ , se llaman valores propios, tienen la propiedad de que  $\|T(x)\| \leq \|x\| \max|\lambda_i|$ , por lo tanto, si cada  $|\lambda_i| < 1$  entonces  $T$  es contracción.

$$\det(A-\lambda I) = \begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix} = (a-\lambda)(d-\lambda) - bc = 0$$

Despejando y simplificando encontramos que  $\lambda = \frac{a+d \pm \sqrt{4bc + (a-d)^2}}{2}$

Por lo tanto  $|\lambda| < 1$  si  $\frac{(a+d)^2 + |(a-d)^2 + 4bc|}{4} < 1$ .

Esta es una condición para que  $T(x) = Ax + b$  sea contracción.

## REFERENCIAS.

- [1]. Abelson, H. y di Sessa, A. (1982): "Turtle Geometry". M.I.T. Press.
- [2]. Barnsley, M. (1988): "Fractals Everywhere". Academic Press.
- [3]. Collado Vides, Ligia (1992): "Estudio Fisionómico-Arquitectónico de las algas del sistema lagunar de Nichupté, Q. Roo, México". Tesis doctoral (Biología). Fac. de Ciencias, UNAM.
- [4]. Dacey, M.F. (1970): "The syntax of a triangle and some other figures". Pattern Recognition, 2, 11-31.
- [5]. Gips, J. (1975): "Shape grammars and their uses: artificial perception, shape generation and computer aesthetics". Birkhäuser-Verlag.
- [6]. Hertz J., Krog A. y Palmer, Richard (1991): "Introduction to the theory of Neural Networks". Santa Fe Institute Studies in the Sciences of Complexity.
- [7]. Khana, Tarun (1990): "Foundations of Neural Networks". Addison-Wesley.
- [8]. Ledley, R. (1964): "High-speed automatic analysis of biomedical pictures". Science 146, Núm. 3641, 216-223.
- [9]. Lindenmayer, A. (1968): "Mathematical models for cellular interaction in development", Parts I and II. Journal of Theoretical Biology 18, 280-315.
- [10]. Prusinkiewicz, P. y Lindenmayer, A. (1990): "The algorithmic Beauty of Plants". Springer Verlag, N.Y.
- [11]. Narasimhan, R. (1962): "A linguistic approach to pattern recognition". Rep. 21, Digital Computer Laboratory, University of Illinois, Urbana.
- [12]. Rosenfeld, J. y Milgram, D. (1972): "Web automata and web grammars, Machine Intelligence, 7, 307-324.