Orden y caos en sistemas complejos
Orden y caos
en sistemas complejos

Ricard V. Solé
Susanna C. Manrubia

Esta obra fue galardonada por la UPC en 1993
A nuestros compañeros del Grupo de Sistemas Complejos: Jordi Bascompte, Jordi Delgado y Bartolo Luque.

Por todos los momentos de amistad y complejidad.
Prólogo

¿Es el mundo predecible? ¿Existe un orden oculto detrás de los torbellinos de la turbulencia? ¿Es estable el sistema solar? ¿Qué es el desorden? La respuesta a estas preguntas, planteadas por los científicos desde hace mucho tiempo, ha sido motivo de profundas discusiones y, no obstante, se han derivado pocas conclusiones. Sin embargo, desde principios de los años 70, una nueva, revolucionaria y sorprendente solución apareció bajo el nombre de caos. Las consecuencias de este descubrimiento fueron enormes. Gran parte del desorden que nos rodea resultó ser sólo aparente. Detras de él se oculta un orden que podemos llevar a traducir en modelos matemáticos simples, los cuales han modificado por completo la visión clásica de orden-desorden como conceptos opuestos.

El caos determinista es, sin embargo, sólo una pieza (aunque especialmente bien comprendida) de un enorme conjunto de nuevos conceptos que, genéricamente, se agrupan bajo lo que conocemos como teoría de los sistemas complejos. La búsqueda de las leyes de lo complejo se ha convertido, en el final del siglo XX, en el objetivo de estudiosos procedentes de campos muy diversos. Esta búsqueda ha sido muy difícil, pero en su curso se han generado nuevas teorías y modelos. Las redes neuronales, los autómatas celulares, los objetos fractales o la criticalidad autoorganizada han permitido formular, de forma simple, las primeras hipótesis generales. En este libro hemos intentado recopilar (de forma introductoria) la mayor parte de los métodos e ideas implicados. Es un libro especial del que, hasta la fecha, no existe equivalente. Se ha procurado que el lector disponga de las herramientas de partida para comprender los elementos básicos de la teoría. Este tratamiento no ha sido exhaustivo (con objeto de no duplicar el tamaño del volumen) y, en este sentido, la bibliografía citada al final de cada capítulo ha sido cuidadosamente escogida a fin de llenar los posibles huecos.

La teoría de la complejidad es una teoría aún en fase de crecimiento. Su completo desarrollo requerirá décadas, pero no cabe duda de que es mucho lo que ya se ha logrado. Aunque el camino a recorrer es largo y los retos muy numerosos, de algo sí podemos estar seguros: el primer asalto a la fortaleza de la complejidad ya ha empezado.
Los autores desean agradecer a la Universitat Politècnica de Catalunya la concesión de una ayuda para la elaboración de este texto en abril de 1994.

Son muchas las personas que han contribuido a nuestra formación y con las que hemos tenido el placer de compartir el entusiasmo por la teoría de la complejidad. Nuestro más sincero agradecimiento a Montse Aguadé, Kosthya Anokhin, Jaume Baguñá, Per Bak, Michael Benton, Adolfo Borraz, Vera Calenbuhr, Germinal Cocho, Alvaro Corral, Albert Díaz-Guilera, Esteban Domingo, Jordi Flos, Nigel Franks, Marta Ginovart, Charles Godfray, José Manuel Gómez-Vilar, Brian Goodwin, Deborah Gordon, Emilia Gutiérrez, Hermann Haken, Michael Hassell, Christian Holscher, Kunihiko Kaneko, Daniel López, Ramon Margalef, Norbert Martínez, Robert May, Liset Menéndez de la Prida, Octavio Miramontes, Pedro Miramontes, M. E. J. Newman, Alexander Mikhailov, Sundaram Parthasarathy, Conrad Pérez Vicente, Steven Rose, Miguel Rubí, Hernán Ruiz Bonet, Joan Valdés, Ton Sales, Juan Manuel Sánchez, Jonathan Silvertown, Joan Manel Solé, George Sugihara, Joaquim Valls, Esteban Vegas y Jorge Wagensberg. Y muy especialmente a las personas con quienes pasamos la mayor parte del día: a nuestros compañeros del Grupo de Sistemas Complejos, a nuestras familias y a Ramon e Isabel, de quienes hemos tomado tiempo para escribir esta obra.
Indice

1 Entropía, Información y Complejidad
   1.1 Random walkers ........................................ 19
   1.2 Entropía y Complejidad .................................. 20
   1.3 Entropía máxima y principios variacionales
       1.3.1 Distribución uniforme ................................ 24
       1.3.2 Distribución de Boltzmann .................................. 25
       1.3.3 Caso general (en ligaduras) .......................... 26
   1.4 Sistemas alejados del equilibrio .......................... 27
   1.5 Información Conjunta ...................................... 28
   1.6 Información en canales con ruido .......................... 30
   1.7 Determinación de la capacidad .............................. 32
   1.8 Canal binario .............................................. 34
   1.9 Información mutua y función de correlación ............ 35
   1.10 Complejidad: algunos comentarios .................... 37
   1.11 Apéndice. Procesos estocásticos ....................... 39

2 Sistemas Dinámicos ........................................ 43
   2.1 Sistemas dinámicos continuos ............................ 46
       2.1.1 Sistemas lineales autónomos en $\mathbb{R}^n$ ............ 46
       2.1.2 Sistemas lineales autónomos en $\mathbb{R}^2$ ............. 47
       2.1.3 Ejemplos en $\mathbb{R}^3$ ................................ 49
       2.1.4 Estabilidad en sistemas no lineales .................. 52
   2.2 El Principio de Control (Slaving Principle) ............ 61
       2.2.1 Organización ........................................... 61
       2.2.2 Autoorganización ...................................... 62
   2.3 Funciones de Lyapunov .................................... 65
   2.4 Sistemas gradiente ....................................... 67
   2.5 Sistemas discretos ....................................... 69

3 Fractales .................................................. 77
   3.1 Caracterización de los objetos fractales .................. 78
       3.1.1 Dimensión de “box-counting” ............................ 80
       3.1.2 Ejemplos ............................................... 82
   3.2 Fundamentos Matemáticos de la Geometría Fractal ........ 86
       3.2.1 Teoría básica de conjuntos ............................ 86
       3.2.2 Funciones y Límites ................................... 90
       3.2.3 Medidas y Distribuciones de Masa .................... 93
   3.3 Sistemas de funciones iteradas (Iterated function systems, IFS) .................. 96
<table>
<thead>
<tr>
<th>Sección</th>
<th>Título</th>
<th>Página</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3.1</td>
<td>Transformaciones de semejanza en $\mathbb{R}^2$</td>
<td>97</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Ejemplos</td>
<td>98</td>
</tr>
<tr>
<td>3.3.3</td>
<td>El teorema del Collage</td>
<td>99</td>
</tr>
<tr>
<td>3.4</td>
<td>Los conjuntos de Julia y de Mandelbrot</td>
<td>100</td>
</tr>
<tr>
<td>3.4.1</td>
<td>Algebra elemental de los números complejos, $\mathbb{C}$</td>
<td>100</td>
</tr>
<tr>
<td>3.4.2</td>
<td>Los conjuntos de Julia</td>
<td>104</td>
</tr>
<tr>
<td>3.4.3</td>
<td>El conjunto de Mandelbrot</td>
<td>106</td>
</tr>
<tr>
<td>3.5</td>
<td>Fractales no deterministas</td>
<td>110</td>
</tr>
<tr>
<td>3.5.1</td>
<td>Multifractales</td>
<td>112</td>
</tr>
<tr>
<td>3.5.2</td>
<td>Agregación limitada por difusión (DLA)</td>
<td>116</td>
</tr>
<tr>
<td>4</td>
<td>Atractores Periódicos y Cuasiperiódicos</td>
<td>121</td>
</tr>
<tr>
<td>4.1</td>
<td>Bifurcaciones</td>
<td>122</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Un único valor propio nulo</td>
<td>122</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Bifurcación de Poincaré-Andronov-Hopf</td>
<td>126</td>
</tr>
<tr>
<td>4.2</td>
<td>La aplicación de Poincaré</td>
<td>129</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Función de desplazamiento</td>
<td>131</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Análisis cualitativo y numérico de la SP</td>
<td>133</td>
</tr>
<tr>
<td>5</td>
<td>Caos Determinista</td>
<td>147</td>
</tr>
<tr>
<td>5.1</td>
<td>Atractores extraños</td>
<td>148</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Lorenz: puntos críticos y estabilidad</td>
<td>148</td>
</tr>
<tr>
<td>5.2</td>
<td>Duplicación de periodo: $f_\mu(x) = \mu x(1 - x)$</td>
<td>152</td>
</tr>
<tr>
<td>5.3</td>
<td>Caos en sistemas discretos</td>
<td>156</td>
</tr>
<tr>
<td>5.4</td>
<td>Exponentes de Lyapunov</td>
<td>160</td>
</tr>
<tr>
<td>5.5</td>
<td>La aplicación triangular</td>
<td>162</td>
</tr>
<tr>
<td>5.6</td>
<td>Sistemas discretos: $d &gt; 1$</td>
<td>163</td>
</tr>
<tr>
<td>5.7</td>
<td>El modelo de Hénon</td>
<td>165</td>
</tr>
<tr>
<td>5.8</td>
<td>La transformación del panadero</td>
<td>168</td>
</tr>
<tr>
<td>5.9</td>
<td>Mixing y ergodicidad</td>
<td>172</td>
</tr>
<tr>
<td>5.10</td>
<td>Mixing en la ecuación logística</td>
<td>174</td>
</tr>
<tr>
<td>5.11</td>
<td>Caos determinista: definición</td>
<td>175</td>
</tr>
<tr>
<td>5.12</td>
<td>Dinámica simbólica</td>
<td>176</td>
</tr>
<tr>
<td>5.13</td>
<td>Caos en el operador $s(x)$</td>
<td>179</td>
</tr>
<tr>
<td>5.13.1</td>
<td>Sensibilidad a las condiciones iniciales</td>
<td>179</td>
</tr>
<tr>
<td>5.13.2</td>
<td>Puntos periódicos densos</td>
<td>180</td>
</tr>
<tr>
<td>5.13.3</td>
<td>Mixing</td>
<td>181</td>
</tr>
<tr>
<td>5.14</td>
<td>Caos en la aplicación triangular</td>
<td>182</td>
</tr>
<tr>
<td>5.14.1</td>
<td>Puntos periódicos densos</td>
<td>183</td>
</tr>
<tr>
<td>5.14.2</td>
<td>Sensibilidad a las condiciones iniciales</td>
<td>184</td>
</tr>
<tr>
<td>5.14.3</td>
<td>Mixing</td>
<td>185</td>
</tr>
<tr>
<td>5.14.4</td>
<td>Consecuencias: Caos en $\mu x(1 - x)$</td>
<td>186</td>
</tr>
<tr>
<td>5.15</td>
<td>La herradura de Smale</td>
<td>186</td>
</tr>
<tr>
<td>5.16</td>
<td>Universalidad en aplicaciones cuadráticas</td>
<td>189</td>
</tr>
<tr>
<td>5.17</td>
<td>Universalidad: aproximación de May-Oster</td>
<td>192</td>
</tr>
</tbody>
</table>
6  **Análisis de Fenómenos Caóticos** .............................................. 229
6.1 Función de autocorrelación ............................................. 231
6.2 Transformada de Fourier ..................................................... 233
6.3 Teorema de Whitney y reconstrucción ................................. 239
   6.3.1 Elección de \(\tau\) para reconstruir ............................... 244
6.4 Dimensión de correlación ............................................... 245
6.5 Atractores extraños en electrocardiogramas ......................... 249
6.6 Limites fundamentales en \(\nu\) y \(\lambda_L\) ............................. 250
6.7 Exponentes de Lyapunov: método de Wolf ......................... 255
6.8 La conjetura de Kaplan-Yorke ........................................... 256
6.9 Detección de determinismo ............................................. 258
6.10 Control del caos ......................................................... 260
   6.10.1 El método OGY ...................................................... 261
   6.10.2 Control de la aplicación de Hénon por el método OGY .... 264
   6.10.3 El método GM ...................................................... 265
   6.10.4 Control de la aplicación de Hénon por el método GM .... 265

7  **Fenómenos Críticos** ......................................................... 271
7.1 El Modelo de Ising ....................................................... 274
   7.1.1 El Modelo de Ising ............................................... 275
   7.1.2 Exponentes críticos y universalidad ........................... 279
   7.1.3 Ising en 1 dimensión: Grupo de Renormalización .......... 282
   7.1.4 Ising en 2 dimensiones: Teoría de Campo Medio ............ 286
   7.1.5 El modelo de Ginzburg-Landau ................................. 288
   7.1.6 La teoría de Landau ............................................. 289
   7.1.7 Ising en 2 dimensiones: Renormalización en el Espacio Real 290
   7.1.8 Simulación del modelo de Ising ............................... 295
7.2 Percolación .............................................................. 296
   7.2.1 Solución exacta en una dimensión ................................ 298
   7.2.2 Exponentes críticos ............................................. 300
   7.2.3 Percolación en dos dimensiones: renormalización en el espacio real 303
7.3 Conclusiones e implicaciones .......................................... 305

8  **Sistemas Críticos Autoorganizados** .................................... 305
8.1 Leyes de escala .......................................................... 306
8.2 Sistemas críticos autoorganizados (SOC) .......................... 309
   8.2.1 La pila de arena .................................................. 310
8.3 El bosque en llamas (Forest Fire) ................................... 313
8.4 Terremotos ............................................................... 317
   8.4.1 Teoría de Campo Medio para el tiempo de retorno .......... 318
   8.4.2 Un modelo sencillo ............................................. 321
8.5 El Juego del Bosque .................................................... 322
   8.5.1 El modelo .......................................................... 323
   8.5.2 Resultados .......................................................... 326
8.6 Un modelo de modelos ................................................... 332
8.7 La predicción en SOC. Conclusiones ................................... 334
9 Autómatas Celulares
  9.1 Autómatas celulares deterministas .................................................. 338
  9.2 Shigamare: ondas en el bosque ............................................................. 340
  9.3 Caracterización cualitativa ........................................................................ 341
  9.4 Caracterización cuantitativa ...................................................................... 342
  9.5 Computación, autómatas y lenguajes formales ........................................... 348
  9.6 Life: computación universal ...................................................................... 350
  9.7 Parámetro λ de Langton ............................................................................ 353
  9.8 Autómatas celulares y medios excitables ................................................ 355

10 Estructuras de Turing y Caos Espaciotemporal ............................................ 361
  10.1 Procesos de difusión ............................................................................... 363
  10.2 La ecuación de difusión ........................................................................... 365
  10.3 Soluciones para $\partial_t u = D \partial_x^2 u$ ......................................................... 367
  10.4 Estabilidad de las soluciones .................................................................... 369
  10.5 Modelos de reacción-difusión .................................................................. 369
      10.5.1 Estructuras disipativas: el Brusselator ................................................. 371
      10.5.2 Gradientes y polaridad ....................................................................... 374
  10.6 Bifurcación de estructuras estacionarias .................................................. 375
  10.7 Modelo de Gierer-Meinhardt ................................................................... 376
  10.8 Estructuras bidimensionales ..................................................................... 380
  10.9 Redes acopladas y caos espaciotemporal ................................................ 384
  10.10 Redes logísticas ..................................................................................... 385
  10.11 Bifurcaciones: análisis formal ................................................................ 388
  10.12 Exponente de Lyapunov espaciotemporal ................................................ 390
  10.13 Supertransitorios y caos espacial ............................................................ 393
  10.14 Competencia y caos espaciotemporal ...................................................... 396
  10.15 Ondas espirales en redes acopladas ....................................................... 400

11 Redes de Kauffman ....................................................................................... 407
  11.1 Control de la expresión genómica ............................................................ 408
  11.2 Regulación compleja, modelos simples ................................................... 410
  11.3 Redes de Kauffman .................................................................................. 413
  11.4 Propiedades dinámicas ............................................................................ 415
      11.4.1 Redes $K = N$ ................................................................................ 415
      11.4.2 Redes $K \geq 5$ ............................................................................ 416
      11.4.3 Redes $K = 1$ ............................................................................. 417
      11.4.4 Redes $K_c = 2$ (Orden colectivo espontáneo) .................................... 417
  11.5 Mecánica estadística: método de Derrida ................................................ 419
  11.6 Percolación: red bidimensional .............................................................. 421
  11.7 Redes de Kauffman generalizadas ............................................................ 422

12 Evolución, Criticalidad y Extinciones ........................................................ 427
  12.1 Extinciones y macroevolución ................................................................... 429
  12.2 La hipótesis de la Reina Roja ................................................................ 432
  12.3 Criticalidad, fractales y evolución ............................................................ 436
  12.4 Modelo de Kauffman ............................................................................... 438
  12.5 Modelo de Bak-Sneppen ........................................................................... 440
      12.5.1 Teoría de campo medio ................................................................. 442
  12.6 Modelos con extinción explícita .............................................................. 445
13 Retrovirus y Cuasiespecies: Entre el Orden y el Caos

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.1 Información genética</td>
<td>455</td>
</tr>
<tr>
<td>13.2 Variabilidad en retrovirus</td>
<td>456</td>
</tr>
<tr>
<td>13.3 Dinámica de replicación molecular</td>
<td>457</td>
</tr>
<tr>
<td>13.4 Replicación con error: cuasiespecies</td>
<td>462</td>
</tr>
<tr>
<td>13.5 La catástrofe de error</td>
<td>465</td>
</tr>
<tr>
<td>13.6 Virus y la organización del sistema inmunitario</td>
<td>468</td>
</tr>
<tr>
<td>13.7 SIDA: en el umbral de diversidad</td>
<td>470</td>
</tr>
<tr>
<td>13.8 Dinámica básica y umbral de diversidad</td>
<td>471</td>
</tr>
<tr>
<td>13.9 $D(v_1, \ldots, v_n)$ como función de Lyapunov</td>
<td>474</td>
</tr>
<tr>
<td>13.10 SIDA y evolución de poblaciones CD4</td>
<td>475</td>
</tr>
<tr>
<td>13.11 Hiperciclos y evolución molecular</td>
<td>476</td>
</tr>
</tbody>
</table>

14 Biodiversidad, Fragmentación del Hábitat y Extinción

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.1 Modelo de Levins</td>
<td>481</td>
</tr>
<tr>
<td>14.2 Competencia entre dos especies</td>
<td>482</td>
</tr>
<tr>
<td>14.3 Competencia multispecífica</td>
<td>483</td>
</tr>
<tr>
<td>14.4 Destrucción del hábitat y coexistencia</td>
<td>485</td>
</tr>
<tr>
<td>14.5 Fragmentación y fenómenos críticos</td>
<td>487</td>
</tr>
<tr>
<td>14.6 La deuda de la extinción</td>
<td>492</td>
</tr>
</tbody>
</table>

15 Neurodinámica

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>15.1 Atractores extraños en sistemas neurales</td>
<td>496</td>
</tr>
<tr>
<td>15.2 Sistemas neurales y duplicación de periodo</td>
<td>500</td>
</tr>
<tr>
<td>15.3 Oscilaciones y caos en el cortex cerebral</td>
<td>501</td>
</tr>
<tr>
<td>15.4 Control de caos en el cerebro</td>
<td>506</td>
</tr>
<tr>
<td>15.5 Control de caos en redes neurales</td>
<td>508</td>
</tr>
<tr>
<td>15.6 Modelo de Hopfield</td>
<td>510</td>
</tr>
<tr>
<td>15.6.1 Modelo teórico: dinámica</td>
<td>511</td>
</tr>
<tr>
<td>15.6.2 Función energía</td>
<td>515</td>
</tr>
<tr>
<td>15.6.3 Red de Hopfield estocástica</td>
<td>517</td>
</tr>
<tr>
<td>15.7 Capacidad de la red estocástica</td>
<td>518</td>
</tr>
<tr>
<td>15.8 Retropropagación (back propagation)</td>
<td>521</td>
</tr>
<tr>
<td>15.9 La máquina de Boltzmann</td>
<td>524</td>
</tr>
<tr>
<td>15.10 Redes con intermediarios</td>
<td>529</td>
</tr>
<tr>
<td>15.11 Transiciones de fase en el cerebro</td>
<td>532</td>
</tr>
</tbody>
</table>

16 Redes Neurales Fluidas

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>16.1 Dinámica de la distribución colectiva</td>
<td>544</td>
</tr>
<tr>
<td>16.2 Comportamiento probabilista: la estrategia del error</td>
<td>545</td>
</tr>
<tr>
<td>16.3 Termitas y orden por fluctuaciones</td>
<td>548</td>
</tr>
<tr>
<td>16.4 Oscilaciones y redes neurales fluidas</td>
<td>551</td>
</tr>
<tr>
<td>16.5 Información y transiciones de fase</td>
<td>555</td>
</tr>
<tr>
<td>16.6 Hormigas y máquinas de Turing</td>
<td>559</td>
</tr>
</tbody>
</table>
17 Caos Hamiltoniano 565
17.1 La mecánica de Hamilton y Jacobi ............................. 565
17.2 Sistemas dinámicos integrables .................................. 569
17.3 Teoría de perturbaciones ......................................... 572
17.4 Resonancias y el teorema KAM ................................. 575
17.5 El teorema de Poincaré-Birkhoff ................................. 576
17.6 Caos en el Sistema Solar ......................................... 580
  17.6.1 El cinturón de asteroides .................................. 580
  17.6.2 Los anillos de Saturno ....................................... 582
  17.6.3 El movimiento de Hiperión ................................. 583
Capítulo 1

Entropía, Información y Complejidad

La entropía crece sin cesar. El segundo principio de la termodinámica predice el decaimiento de todas las estructuras con el tiempo. Lo ordenado dejará de serlo, tarde o temprano, dando paso al desorden. Pero aunque este principio es ciertamente general, a nuestro alrededor se agitan miles de sistemas complejos que, en una forma u otra, exhiben un alto grado de orden. La vida es el ejemplo preeminentemente, pero incluso en los sistemas no vivos puede darse la aparición de orden en las situaciones más inesperadas. Imagínemos una reacción química en la que mezclamos sobre una superficie ciertos reactivos. La imagen clásica de la termodinámica nos dice que este sistema evoluciona hacia una situación de equilibrio caracterizada por la máxima entropía y la homogeneidad. Una vez terminada la reacción, nada ocurrió de nuevo: veremos una disolución homogénea, del mismo color, y nada más.

Sin embargo, las cosas no siempre son así. Ciertas reacciones químicas generan estructuras espaciales de enorme complejidad, como la que se indica en la figura 1.1. La superficie nos define para cada punto del espacio la concentración local de uno de los componentes de la reacción (Nicolis y Prigogine, 1977, 1988). Partiendo de una concentración espacial homogénea de los reactivos (que habremos agitado previamente) se van creando ondas macroscópicas de gran tamaño, que forman espirales en rotación. Estas ondas son visibles a simple vista y por lo tanto afectan a billones de moléculas que se han “autooroganizan” espontáneamente para dar lugar a una estructura ordenada. Este resultado fue recibido con enorme escepticismo. El químico Boris Belousov descubrió en 1950, en su laboratorio de biofísica de Moscú, una de estas reacciones que aparentemente contradecían el segundo principio. En 1951 vio cómo su primer artículo acerca de este resultado era rechazado por el editor de una revista científica. Dicho editor le señaló que su “descubrimiento supuestamente descubierto” era del todo imposible (véase Coweney y Highfield, 1992, para un relato pormenorizado de esta historia). Más tarde otros científicos darian con resultados similares y Anatóly Zhabotinsky llevó a cabo un estudio pormenorizado que acabó de convencer a los escépticos. Belousov fue finalmente reconocido... póstumamente.

A lo largo de este texto veremos la aparición de complejidad en sistemas de todo tipo. Pese a la aparente contradicción con la segunda ley, que se aplica a sistemas cerrados, los sistemas que nos interesan son sistemas abiertos que intercambian energía y materia con el exterior. Este intercambio tiene a veces un aspecto especial: lo que se intercambia es, de hecho, información. A partir de sistemas formados por elementos simples, alejados del equilibrio, la vida se autoorganiza de formas sorprendentes. La segunda ley siempre acaba ganando la partida, pero durante ésta muchas son las cosas que pueden ocurrir. Una de ellas es la emergencia espontánea de lo complejo.
Comprender los orígenes de la complejidad no es una tarea fácil. El punto de partida tampoco lo es: no disponemos de una definición simple y diáfana de lo “complejo”. En el presente texto intentaremos abordar esta pregunta desde sus fundamentos y volveremos a ella al final del libro. Nuestro punto de partida en este capítulo será de carácter macroscópico, más aún, de carácter probabilista. Partiremos de la idea de entropía e intentaremos analizar la complejidad desde esta magnitud y otras que surgen de la teoría de la información.

Puede resultar extraño que, para analizar la emergencia de la complejidad, empleemos herramientas matemáticas típicas del análisis de los sistemas desordenados. Sea cual sea la definición que acabemos empleando, lo complejo se halla a medio camino entre lo ordenado (un cristal, por ejemplo) y lo desordenado (un gas). En la figura 1.2 se muestran tres ejemplos de sistemas, dos de ellos en los extremos de la complejidad y uno intermedio. En el caso (a), tenemos una estructura ordenada, fácilmente predecible (basta con observar una pequeña parte para hacerse una idea del comportamiento global) y lo mismo ocurre en (c), aunque ahora se trate de un sistema totalmente desordenado. En (b) podemos ver un ejemplo de estructura compleja. Existen elementos de desorden (al menos aparentemente) que hacen difícil predecir la estructura global a partir de fragmentos de la misma. Sin embargo, está claro que existe un orden subyacente dentro de esta estructura. Hay regularidades que podemos intuir, aunque por ahora no sepamos cómo medirlas. Pero puesto que hemos hablado de orden y de desorden, la entropía puede servirnos de punto de partida.

La entropía juega en física un papel preponderante en nuestra exploración de los fenómenos dinámicos. Es bien conocida la segunda ley de la termodinámica, la cual afirma que la entropía siempre aumenta. Será por tanto una magnitud a tener en cuenta en nuestro estudio, que tratará de hecho de las propiedades de la evolución temporal de sistemas muy diversos.

Para medir esta magnitud disponemos de una aproximación que posee una tremenda generalidad. Sea un sistema cualquiera (un conjunto de átomos, por ejemplo) sobre el que hemos definido cierta cantidad, de forma que podamos contar cuántos elementos tienen cada valor, y establecer así las probabilidades de tener un objeto escogido al azar en cada estado. Sea \( \{p_1, \ldots, p_n\} \)
dicho conjunto. Estas probabilidades verifican obviamente la condición \( p_j \in [0, 1] \), así como la normalization

\[
\sum_{j=1}^{N} p_j = 1
\]

La entropía \(^1\) se define por:

\[
H = - \sum_{i=1}^{N} p_i \log p_i
\]

siendo \( N \) el número de estados posibles (las posibles energías de los átomos).

Veamos ahora cómo justificar de manera intuitiva la definición de entropía a partir de criterios de información. Consideremos un conjunto de "sucesos" \( \{A_1, ..., A_n\} \) definibles sobre un problema dado \( \Gamma \) (el conjunto de posibles resultados del lanzamiento de un dado, por ejemplo), de tal modo que formen una partición, esto es,

\[
\begin{align*}
(a) & \quad A_i \cap A_j = \emptyset \quad \forall i, j = 1, ..., N \\
(b) & \quad \bigcup_{j=1}^{N} A_j = \Gamma
\end{align*}
\]

Existirán en general múltiples particiones posibles sobre las que definir probabilidades. Imaginemos que deseamos definir una medida de la información proporcionada por un suceso dado. Intuitivamente, esta medida debería verificar algunos requisitos. En particular:

- Un suceso más improbable (con baja probabilidad) nos da más información\(^2\). Esperaremos por lo tanto encontrar una medida de información que dependa de la probabilidad en la forma:

\[
I(A_k) = f(1/p_k)
\]

siendo \( f(x) \) una función creciente.

---

\(^1\)El símbolo \( H \) se debe a Ludwig Boltzmann.

\(^2\)Por ejemplo, al resolver un crucigrama en castellano, la letra \( Z \) o la \( W \) restringen más las posibilidades y en ese sentido nos dan más información.
Supongamos ahora \( n \) sucesos equiprobables para los que

\[
p_j = \frac{1}{n} \quad j = 1, \ldots, n
\]

Si consideramos \( m \) realizaciones independientes en un mismo instante, el número total de posibilidades es \( n^m \). Además, cada \( m \)-pla tendrá una probabilidad de ocurrir de \( 1/n^m \) y en este caso parece razonable que la incertidumbre en la realización de \( m \) sucesos sea \( m \) veces la incertidumbre asociada a un único suceso, esto es:

\[
I(A_{i1}, \ldots, A_{im}) = f(n^{-m}) = mf(A_{ik}) = mf\left(\frac{1}{n}\right)
\]

Una función \( f(x) \) que satisface ambas propiedades es la función logaritmo, esto es:

\[
I(A_k) = \log \left(1/P(A_k)\right) = -\log P(A_k)
\]

Tendremos así la siguiente definición de la entropía asociada a un conjunto de probabilidades:

**Definición**

La información (autoinformación) de un suceso \( A_k \) se define como

\[
I(A_k) = -\log (p_k) \tag{1.1.1}
\]

Puesto que en general tendremos un conjunto de sucesos (letras de un alfabeto, símbolos, etc.) sobre los que definiremos un conjunto de probabilidades, podríamos preguntarnos cuál será la información promedio de todo el sistema. Si empleamos la definición general de magnitud promedio (la media) de un sistema dado, ésta viene definida por:

\[
< f > = \sum_{i=1}^{n} p_k f_k
\]
siendo aquí $f_k$ el valor asociado al $k$-ésimo "estado" y $p_k$ la probabilidad de que dicho valor se observe. Llegamos así a la definición de entropía:

**Definición**

La entropía $H$ es el valor medio de la autoinformación, esto es,

$$H = - \sum_{i=1}^{N} p_i \log p_i$$

(1.1.2)

definida para un sistema cualquiera de probabilidades ³ ($p_i$).

Esta definición probabilista de $H$ fue formulada por vez primera por el genial físico austriaco Ludwig Boltzmann. Boltzmann realizó contribuciones cruciales al desarrollo de la mecánica estadística. Durante toda su vida buscó la solución a un problema fundamental: la explicación de la irreversibilidad de los procesos naturales y específicamente de la irreversibilidad expresada en la segunda ley de la termodinámica. Buscó una explicación mecánica, de carácter microscópico, para la existencia de una flecha del tiempo. La fórmula de la entropía aparece sobre la lápida de su tumba en el cementerio de Viena.

Puesto que $I(x)$ mide la incertidumbre, $H$ nos dará un valor medio de la incertidumbre sobre el sistema. Puesto que $p_k \geq 0$, se tiene $\log(p_k) \leq 0$ y $H \geq 0$. Vemos claramente que para un sistema en el que $p_j = 1$, y en consecuencia las demás probabilidades sean nulas ($p_k, k \neq j = 0$), se tiene incertidumbre nula ($H = 0$), como cabía esperar. El límite superior es por lo tanto evidente. El límite superior puede probarse mediante el siguiente

**Teorema**

La entropía de Boltzmann $H$ verifica $H \leq \log(n)$, siendo $H = \log(n)$ si y sólo si tenemos equiprobabilidad, esto es $p_j = 1/n$, $\forall j = 1, \ldots, n$.

Como caso particular, consideremos la entropía definida para un sistema con sólo dos estados, i.e. $\Gamma = \{A_1, A_2\}$. Dado que podemos escribir $p_1 = p$ y $p_2 = 1 - p$, se tiene:

$$H(p) = - \left[ p \log(p) + (1 - p) \log(1 - p) \right]$$

que se representa en la figura 1.3, y que posee un máximo en $p = 1/2$, como establece el teorema anterior. Si uno de los sucesos ocurre con probabilidad unidad, $H(p) = 0$.

### 1.1 Random walkers

A título de ejemplo, consideremos un conjunto de partículas que se desplazan al azar o *random walkers* (RW), sobre un retículo (rejilla) de lado $L$. Tenemos así $L^2$ posiciones accesibles. En un instante dado, cada uno de los elementos se desplaza al azar a una de sus posiciones vecinas más próximas (o bien permanece en su propia posición). Si tomamos una red unidimensional y un único elemento, la trayectoria que seguiría se ilustra en la figura 1.4, en la que en el eje horizontal se indica el tiempo (que asumimos discretizado) y en el eje vertical la posición del objeto.

Supongamos ahora que empleamos una red unidimensional, de manera que un elemento pueda saltar a cualquiera de sus dos posiciones vecinas con probabilidad 1/3 o permanecer en ella con la misma probabilidad. Un punto de la red puede estar ocupado por más de un elemento, e indicaremos por

³Observemos que podemos tener sucesos de probabilidad nula. En este caso, la existencia del límite $\lim_{x \to 0} x \log(x) = 0$ evita cualquier problema.
Figura 1.4: Trayectoria de un *random walker* en un espacio unidimensional. La posición de la partícula se indica en el eje vertical.

\[ \{ p_t(j) \} ; \quad j = 1, 2, ..., L \]

la probabilidad (para un instante \( t \)) de encontrar un RW en la posición j-ésima. Imaginemos que inicialmente todos los elementos se hallan en el punto central, i.e.

\[ p_0(\frac{L}{2}) = 1 \]

y cero para las restantes \( p_0(j) \). Supongamos que, a partir de ese instante, los objetos pueden desplazarse, y que seguimos a lo largo del tiempo la evolución de \( \{ p_t(i) \} \). En la figura 1.5 se resume el resultado de este experimento simulado. Al principio, los elementos se concentran en el punto central, pero con el tiempo se van dispersando a lo largo de la red dando lugar a una campana de Gauss muy achatada. Si esperamos lo suficiente, el resultado final es una distribución homogénea.

La tendencia hacia este estado de máximo desorden se puede también visualizar con una gráfica de la evolución de \( H(t) \), como la que se muestra en la figura 1.5 (b). Vemos que, salvo pequeñas fluctuaciones asociadas al tamaño finito de nuestro sistema, es una función claramente creciente en el tiempo. Al alcanzar el estado de equilibrio final, la entropía alcanza su valor máximo, en este caso.

\[ H(\infty) = - \sum_{j=1}^{L} p_\infty(j) \log(p_\infty(j)) = - \sum_{j=1}^{L} \frac{N}{L} \log \left( \frac{N}{L} \right) = \log(L) \]

Debemos indicar que, estrictamente, tendremos fluctuaciones cercanas al valor asintótico, tanto más importantes cuanto menor sea el tamaño del sistema (el número de RW implicados).

### 1.2 Entropía y Complejidad

La entropía de un sistema físico proporciona una primera aproximación en nuestra búsqueda de una medida de complejidad. Su empleo en ecología (Margalef, 1987) es de hecho un método
Figura 1.5: (a) Distribución de probabilidad \{p(j)\} obtenida para un conjunto de \( N = 8000 \) random walkers sobre una red de \( L = 31 \) puntos. (b) Entropía asociada a la evolución del sistema anterior.

generalizado de medir la diversidad de especies de un ecosistema. ¿No es entonces una medida adecuada de la complejidad del sistema? La respuesta es negativa.

El uso de la entropía de Boltzmann como medida de complejidad ha sido sin embargo habitual en ciencias de la computación. Imaginemos un programa que ejecuta cierto número de instrucciones, y supongamos (razonablemente) que la salida final está formada por un conjunto de ceros y unos. Nuestra intuición nos dice que cuanto más complejo sea el algoritmo implicado en esta salida, tanto más “compleja” será. Aquí entendemos por complejidad la dificultad que supone generar la secuencia. En términos más simples diríamos que un objeto es complejo si contiene “información difícil de obtener” (Ruelle, 1993).

Más específicamente, imaginemos un ordenador que dispone de un algoritmo de cierta longitud que le permite generar una secuencia dada. Definiremos a continuación la \textit{complejidad algorítmica}, que medirá la dificultad de generar una secuencia de bits mediante un algoritmo. Si la secuencia es regular, como la que vemos en la figura 1.6(a), el programa necesario para generarla es simplemente “escribe 110” junto con la repetición de esta afirmación. En el caso (b), en el que hemos generado una secuencia al azar, el programa será tan largo como la propia secuencia, y la complejidad algorítmica será la mayor posible. Vemos por lo tanto que la medida de complejidad que nos proporciona la complejidad algorítmica es inadecuada para nuestra intuición de lo que entendemos por complejidad. Desde el punto de vista de esta medida, un gas ideal o cualquier otro sistema en equilibrio termodinámico serían los objetos más complejos.

Como indicábamos al principio de esta sección, la entropía de Boltzmann fue introducida en ecología teórica por Margalef (1968), dándole el nombre de diversidad ecológica o simplemente \textit{diversidad}. La observación de ecosistemas complejos nos muestra cierto número de regularidades relevantes. Una de ellas es la distribución de especies ordenadas de más a menos abundante.
Figura 1.6: Complejidad algorítmica. Cuanto más desordenada sea la secuencia que aparece, mayor longitud deberá poseer el algoritmo que la genera. (a) Secuencia regular. (b) Secuencia aleatoria.

Al realizar esta ordenación descubriremos una relación decreciente muy típica. Los ecosistemas reales no están constituidos por una sola especie \( H = 0 \) ni por una distribución uniforme de individuos de cada especie, como ocurriría en un museo \( H \) máxima). A medio camino entre ambos extremos los ecosistemas reales parecen encontrar un balance entre ambas posibilidades. La vida genera constantemente diversidad, y por tanto no debemos esperar encontrar sistemas de gran simplicidad (a menos que el medio ambiente lo imponga así). Por otra parte, un ecosistema muy diverso puede tener problemas funcionales. Una observación generalizada es que los ecosistemas naturales muestran una diversidad acotada en un intervalo bien definido, superando raramente los 5 bits.

Como ha señalado Ramon Margalef la diversidad es una expresión de la estructura resultante de la forma en la que interactúan los elementos (especies) del sistema. La diversidad es sin lugar a dudas un elemento necesario para mantener una estructura compleja. Si \( H \) es reducida, las posibilidades de mantener una estructura compleja se reducen; si \( H \) es muy elevada, será difícil mantener la funcionalidad, a menos que otras propiedades se modifiquen adecuadamente.

La relación entre \( H \) y la forma en que las distintas partes del sistema se relacionan entre sí puede expresarse mediante una medida de conectividad. En la figura 1.7(a) vemos una gráfica de esta medida junto a la entropía correspondiente. Cada punto corresponde a un circuito electrónico funcional, para el que se han medido las probabilidades de cada tipo de elemento (díodos, resistencias, etc.) así como el número de conexiones promedio que cada elemento posee con los demás. Podemos observar que existe una clara relación decreciente: a mayor diversidad, menor conectividad. Otra gráfica útil se muestra en la figura 1.7(b), en la que también observamos una relación aún más acusada entre el número de piezas distintas (tipos de elementos) y la conectividad. Estas relaciones expresan el balance existente entre el número de distintos elementos que forman el sistema (ya sean especies o componentes electrónicos) y el grado de relación directa que existe entre dichos elementos. Para lograr un buen funcionamiento, si el número de elementos posibles se hace mayor, la flexibilidad necesaria se obtiene haciendo menos rígida la relación entre las partes o,
Figura 1.7: (a) Diagrama de Conectividad-Diversidad (entropía) para un conjunto de 78 circuitos electrónicos comerciales (datos tomados de Margalef y Gutiérrez, 1983). (b) Diagrama del número de piezas distintas empleadas versus Conectividad para el conjunto de circuitos anterior (op. cit.). Obsérvese la relación potencial entre ambas cantidades, característica de un gran número de sistemas complejos, incluidos los ecosistemas reales.

lo que es lo mismo, reduciendo la conectividad. Estos resultados, de validez general, nos indican que para comprender adecuadamente la complejidad de un sistema necesitamos alguna medida en la que se introduzca el grado de relación entre las partes. Veremos más adelante una medida adecuada para caracterizar esta propiedad y la complejidad del sistema.

1.3 Entropía máxima y principios variacionales

Hemos visto en el caso de los "random walkers" que la entropía del sistema crece asintóticamente hasta alcanzar su valor máximo. La evolución espontánea del sistema lo conduce a un estado de máximo "desorden". En situaciones algo más interesantes, la distribución de probabilidad no es, sin embargo, tan trivial. Si observamos por ejemplo la distribución de energías de los átomos de un gas en equilibrio, veremos que sigue una forma exponencial, del tipo \( P_i \approx e^{-\beta E_i} \), siendo \( E_i \) la energía del nivel \( i \)-ésimo. A mayores energías, encontraremos pocos átomos, mientras que lo más probable será encontrarlos en el estado de mínima energía. El sistema evoluciona en este caso hacia un estado de máxima entropía, pero la distribución final no es uniforme. Existe un método general de encontrar dicha distribución si conocemos de antemano las "ligaduras" que operan sobre dicho sistema. Por ligaduras entendemos cualquier restricción de tipo macroscópico sobre el conjunto de probabilidades \( \{p_i\} \), y habitualmente las escribiremos como:

\[
L_k(p_1, \ldots, p_n) = C_k
\]

Entre otros posibles ejemplos, estas ligaduras pueden ser tan triviales como la propia normalización de probabilidades,

\[
L_1(p_1, \ldots, p_n) = \sum_j p_j = 1
\]

(1.1.3)

o bien el hecho de que el valor medio de la magnitud relevante (la energía, por ejemplo) sea constante. Si en nuestro sistema tenemos que la cantidad \( f_k \) se presenta con probabilidad \( p_k \), su
valor medio será:

\[ L(f) = \sum_j p_k f_k = \langle f \rangle \]

Para encontrar, dadas las ligaduras, la distribución más probable, basta con hallar el máximo de la función entropía \( H \) restringida por el conjunto \( L_k \). Este método, conocido como formalismo del Maxent (de "maximum entropy formalism") se basa en la resolución de la ecuación variacional:

\[ \delta \left\{ H([p_j]) - \sum_k \alpha_k [L_k - C_k] \right\} = 0 \]

La solución de este problema no es sino la distribución de probabilidad buscada, \( \{p_j\} \). Los valores \( \alpha_k \) son parámetros a determinar, y se denominan multiplicadores de Lagrange. Ilustraremos el método con dos ejemplos típicos.

### 1.3.1 Distribución uniforme

Consideremos un sistema carente de toda restricción acerca de su posible comportamiento (no está limitado por la energía, etc.). Un ejemplo de este tipo serían los "random walkers" ya analizados, que pueden ocupar un cierto número de posiciones (estados) con probabilidad \( p(j) \). No existe ningún tipo de interacción entre ellos ni tampoco existe ninguna magnitud conservada, excepto el número de partículas. Tenemos así que hallar la distribución asociada al máximo de \( H \) con la única restricción de normalización de probabilidades, 1.1.3. La ecuación variacional será entonces:

\[ \frac{\partial}{\partial p_j} \left\{ - \sum_j p_j \log (p_j) - \alpha \left[ \sum_j p_j - 1 \right] \right\} = 0 \]

que nos da como resultado

\[ -\log(p_j) - 1 - \alpha = 0 \]

luego las probabilidades serán de la forma:

\[ p_j = e^{-(1+\alpha)} \]

esto es, iguales entre sí. Ahora sólo nos queda evaluar la constante \( \alpha \). Para ello emplearemos la ligadura de normalización:

\[ \sum_j p_j = \sum_j e^{-(1+\alpha)} = \frac{e^{-(1+\alpha)}}{e^{-(1+\alpha)}} = 1 \]

o, lo que es lo mismo, \( e^{-(1+\alpha)} = 1/n \), como cabía esperar.

Hemos obtenido por lo tanto la distribución de probabilidades que hace máxima la entropía, esto es,

\[ p_j = \frac{1}{n} \]

(como ocurre en los "random walkers") y la entropía es \( H = \log(n) \), que se corresponde con la máxima posible, como ya indicábamos anteriormente.
1.3.2 Distribución de Boltzmann

Imaginemos ahora un sistema aislado (como un gas dentro de una caja) en el que el número de elementos se conserva así como la energía total (u otra cantidad en el caso de un sistema distinto). Supondremos nuevamente que la interacción entre elementos no es relevante para nuestro análisis. Por tanto, nuestro sistema está sometido a una ligadura tal y como

\[ L(f) = \sum_k p_k f_k = < f > \]

a la que añadiremos la ligadura 1.1.3.

En este nuevo caso, la ecuación variacional añadirá un término nuevo, es decir,

\[ \frac{\partial}{\partial p_j} \left\{ -\sum_j p_j \log(p_j) - \alpha \left[ \sum_j p_j - 1 \right] - \beta \left[ \sum_j p_j f_j - < f > \right] \right\} = 0 \]

La ecuación nos da en este caso

\[ p_j = e^{-\alpha - \beta f_j} \]

donde hemos reescrito \( \alpha \) en lugar de \( 1 + \alpha \). La aplicación de las ligaduras nos permitirá calcular las constantes. La condición de normalización proporciona

\[ \sum_j p_j = \sum_j e^{-\alpha - \beta f_j} = e^{-\alpha} \sum_j e^{-\beta f_j} = 1 \]

que podemos escribir en la forma \( e^{-\alpha} = \frac{1}{Z} \) siendo \( Z \) la llamada función de partición,

\[ Z = \sum_{j=1}^{n} e^{-\beta f_j} \]

Tenemos así una dependencia exponencial de las probabilidades respecto de los valores \( f_j \). Podemos calcular explícitamente el segundo multiplicador empleando la ligadura correspondiente al valor medio. Para ello, pasamos al continuo reemplazando los sumatorios por integrales. Así,

\[ Z = \int_0^\infty e^{-\beta f} df \]

y entonces las probabilidades pasan a ser funciones de densidad de probabilidad:

\[ p(f) = \frac{e^{-\beta f}}{Z} \]

La segunda ligadura será:

\[ \int_0^\infty p(f) f df = < f > \]

luego tendremos \(^4\):

\[ \int_0^\infty f e^{-\beta f} df = \frac{\Gamma(2)}{\beta^2} = \frac{1!}{\beta^2} \]

\(^4\)La función gamma \( \Gamma(n) \) verifica las propiedades siguientes:

\[ \int_0^\infty x^n e^{-ax} dx = \frac{\Gamma(n+1)}{a^{n+1}} \quad \Gamma(n+1) = n! \quad n = 0, 1, 2, \ldots \]
y puesto que
\[ \int_{0}^{\infty} e^{-\alpha x} dx = \frac{1}{\beta} \]
se tiene entonces
\[ \beta = \frac{1}{\langle f \rangle} \]
luego la distribución explícita depende de una única magnitud macroscópica, en este caso el valor promedio:
\[ Z = \sum_{j=1}^{n} e^{-f_{j}/\langle f \rangle} \]

Aunque la distribución que hemos obtenido es típica de sistemas físicos en equilibrio (bajo conservación de la energía dentro del sistema) ha sido observada en sistemas abiertos, tales como poblaciones de peces (Lurie et al., 1983). En este caso, se midieron las biomasa de los individuos y se calcularon las frecuencias de cada clase de masa (esto es, se obtuvo un histograma de frecuencias):
\[ \{p(m_1), p(m_2), ..., p(m_n)\} \]
y la distribución observada era efectivamente la de Boltzmann:
\[ p(m_j) = \frac{1}{Z} \exp \left( -\frac{m_j}{\langle m \rangle} \right) \]
siendo \( \langle m \rangle \) la masa media. Otras aplicaciones a sistemas no-lineales alejados del equilibrio también han mostrado la posibilidad de obtener esta distribución (Solé y Luque, 1994) cuando la ligadura de conservación hace referencia a condiciones de estabilidad estructural de un sistema dinámico.

### 1.3.3 Caso general (n ligaduras)

En el caso más general, en el que junto a la condición de normalización poseemos otras \( N \) ligaduras adicionales, nuestro objetivo será resolver la ecuación variacional dada por:
\[ \delta \left[ H - (\lambda - 1) \sum_{i} p_i - \sum_{k} \lambda_k \sum_{i} p_i f^{(k)}_i \right] = 0 \]
donde hemos indicado por \( \{\lambda_k\} \) el conjunto de multiplicadores de Lagrange asociados a cada ligadura \( f^{(k)}_i \). Emplearemos en la normalización el símbolo \( (\lambda - 1) \) por simplicidad en el cálculo. La derivación respecto a \( p_i \) nos da:
\[ -\log(p_i) - 1 - (\lambda - 1) - \sum_{k} \lambda_k f^{(k)}_i = 0 \]
esto es,
\[ p_i = \exp \left[ -\lambda - \sum_{k} \lambda_k f^{(k)}_i \right] \]
Si ahora aplicamos a este conjunto de probabilidades la ligadura de normalización, obtenemos:
\[ \sum_{i} p_i = e^{-\lambda} \left\{ \sum_{i} \exp \left[ -\sum_{k} \lambda_k f^{(k)}_i \right] \right\} = 1 \]
y como antes indicaremos la función de partición $Z$ como:

$$Z(\{\lambda_i\}) = \sum_i \exp \left( - \sum_k \lambda_k f_i^{(k)} \right)$$

es decir,

$$e^\lambda = Z$$

$$\lambda = \ln(Z)$$

lo cual nos permite determinar $\lambda$ una vez conozcamos $\{\lambda_k\}$.

Para hallar las ecuaciones para los multiplicadores $\lambda_k$, insertamos $p_i$ dentro de las ecuaciones que definen las ligaduras, esto es, en

$$\sum_i p_i f_i^{(k)} = < f_i^{(k)} >$$

Tenemos entonces

$$< f_i^{(k)} > = e^{-\lambda} \sum_i \exp \left[ - \sum_k \lambda_k f_i^{(k)} \right] f_i^{(k)}$$

que podemos reescribir, obteniendo

$$< f_i^{(k)} > = Z^{-1} \frac{\partial}{\partial \lambda_k} \sum_i \exp \left[ - \sum_k \lambda_k f_i^{(k)} \right]$$

$$= \frac{\partial}{\partial \lambda_k} \ln(Z(\lambda_1, ..., \lambda_n))$$

Si insertamos las $p_i$ deducidas del principio variacional en $H$, obtenemos la máxima entropía compatible con las ligaduras, $H_{max}$:

$$H_{max} = - \sum_i p_i \log(p_i)$$

$$= \sum_i \exp \left[ - \lambda - \sum_k \lambda_k f_i^{(k)} \right] \left[ - \lambda - \sum_k \lambda_k f_i^{(k)} \right]$$

$$= \lambda \sum_i p_i + \sum_k \lambda_k \sum_i p_i f_i^{(k)}$$

$$= \lambda + \sum_k \lambda_k f_k$$

Vemos así que la máxima entropía puede ser representada por los valores medios y los multiplicadores de Lagrange.

### 1.4 Sistemas alejados del equilibrio

La elección del conjunto apropiado de ligaduras puede no ser evidente. Tampoco lo es, en principio, el conjunto de restricciones que deben considerarse acerca del sistema empleado. El hecho de que tratemos como ejemplos estándar sistemas en equilibrio termodinámico y, por tanto, poco proclives a mostrar complejidad de algún tipo, podría hacernos creer que el principio anterior
queda restringido a sistemas en equilibrio. Sin embargo, pueden obtenerse excelentes resultados para sistemas alejados del equilibrio eligiendo las ligaduras apropiadas. Un ejemplo particularmente brillante de aplicación del método variacional a un sistema físico alejado del equilibrio es el estudio de Haken (1988). Haken ha demostrado que, empleando el Maxent con ligaduras que introducen las correlaciones macroscópicas a través de los momentos de orden superior, pueden obtenerse de forma exacta las distribuciones de probabilidad asociadas a las medidas experimentales obtenidas en experimentos con láseres (y en general, para sistemas alejados del equilibrio cercanos a puntos críticos). La aproximación general (tratada por Haken en su libro, 1988) consiste en la siguiente idea. Partimos de un sistema descrito por el vector de estado

\[ q = (q_1, q_2, ..., q_N) \]

cuyas componentes son medibles por el experimentador. Aquí el subíndice \( i \) de \( q_i \) puede ser una célula o distintos tipos de cantidades físicas o químicas. Asumiremos que los promedios estadísticos sobre las \( q_i \) y sus momentos hasta orden cuatro son conocidos. Introducimos entonces como ligaduras:

\[ f_i = < q_i > \]
\[ f_{ij} = < q_i q_j > \]
\[ f_{ijk} = < q_i q_j q_k > \]
\[ f_{ijkl} = < q_i q_j q_k q_l > \]

Con las que llevaremos a cabo la maximización de la entropía. Puede comprobarse que la distribución de probabilidad \( p(q) \) viene dada por:

\[ p(q) = \exp \left[ V(\lambda, q) \right] \]

siendo \( V(\lambda, q) \) una función lineal de los multiplicadores de Lagrange y no-lineal de las componentes del vector de estado \( q \), definida por

\[ V(\lambda, q) = \lambda + \sum_i \lambda_i q_i + \sum_{i,j} \lambda_{ij} q_i q_j + \sum_{i,j,k} \lambda_{ijk} q_i q_j q_k + \sum_{i,j,k,l} \lambda_{ijkl} q_i q_j q_k q_l \]

donde \( V(\lambda, q) \) cumplirá, mediante un adecuado cambio de coordenadas, el requisito

\[ \frac{\partial V(\lambda, q)}{\partial q_i} = 0 ; \quad i = 1, 2, ..., N \]

como esperaríamos en el marco de la teoría de transiciones de fase de no-equilibrio (Haken, 1987; 1988). Este estudio va más allá de nuestras pretensiones en esta introducción. Aconsejamos al lector las monografías de Haken acerca de esta aplicación y posibles extensiones.

1.5 Información Conjunta

Un sistema en el que no hay interacción entre sus elementos ni entre éstos y el entorno (un sistema aislado) evoluciona al estado de máxima entropía. Esta situación ha quedado manifiesta en el ejemplo anterior de los "random walkers". Sin embargo, el intercambio de información se halla siempre presente en los sistemas complejos. En ocasiones este intercambio es muy simple (el enlace entre átomos) y en ocasiones más sutil (la comunicación entre neuronas). La emisión, recepción y elaboración de los mensajes está detrás de la mayoría de los fenómenos que nos rodean. La teoría de la información intenta cuantificar estas magnitudes y nos será muy útil en nuestro desarrollo
posterior de algunas ideas fundamentales. Definiremos a continuación algunas cantidades de interés así como nuevas medidas de entropía, para terminar definiendo la información transmitida.

Supongamos dos variables aleatorias $X, Y$ definidas con valores sobre los conjuntos

$$S = \{A_1, ..., A_n\}$$

$$R = \{B_1, ..., B_m\}$$

tales que las respectivas probabilidades vienen dadas por $\{p(A_i)\}$ y $\{p(B_j)\}$. Ambas están normalizadas, obviamente. Sea

$$P(A_i \cap B_j) = p(A_i, B_j)$$

la probabilidad conjunta del par $(A_i, B_j)$, esto es, de que ambos sucesos se den simultáneamente. Definiremos también las probabilidades condicionadas $P(A_i|B_j)$ será la probabilidad de que se dé $A_i$ si sabemos que se ha dado $B_j$, y análogamente definiremos $P(B_j|A_i)$. Estas probabilidades están relacionadas entre sí a través de las igualdades:

$$p(A_i, B_j) = P(A_i)P(B_j|A_i)$$

$$P(A_i, B_j) = P(B_j)P(A_i|B_j)$$

Podemos entonces definir la entropía condicionada por $Y = B_j$ como la suma:

$$H(X|Y = B_j) = - \sum_{i=1}^{n} P(A_i|B_j) \log P(A_i|B_j)$$

que será de hecho el grado de incertidumbre existente sobre el conjunto $S$ si se ha dado $B_j \in R$. En términos de un canal de comunicación, en el que $S$ sería el conjunto de símbolos de entrada y $R$ el conjunto de símbolos del receptor ($n = m$ y supondremos los mismos símbolos), las probabilidades condicionadas nos dan de hecho la fiabilidad del canal de comunicación. La entropía condicionada, sin más, será el promedio de estos valores, es decir, la suma ponderada sobre los $A_i$:

$$H(X|Y) = - \sum_{j=1}^{m} P(B_j)H(X|Y = B_j)$$

$$= \sum_{j=1}^{m} \sum_{i=1}^{n} P(B_j)P(A_i|B_j) \log P(A_i|B_j)$$

la cual medirá la incertidumbre sobre $S$ conocido $R$.

Si definimos también la entropía conjunta como la entropía asociada a las probabilidades $P(A_i, B_j)$, esto es

$$H(X, Y) = - \sum_{j=1}^{m} \sum_{i=1}^{n} P(A_i, B_j) \log (P(A_i, B_j))$$

algunas relaciones y desigualdades pueden ser obtenidas de manera simple:

**Propiedades**

(i) $H(X, Y) = H(X) + H(Y|X)$

(ii) $H(X, Y) \leq H(X) + H(Y)$
De especial importancia es la desigualdad \( H(X|Y) \leq H(X) \), que establece que la entropía de \( X \) condicionada a \( Y \) no puede exceder la de la propia \( X \).

Llegamos ahora a la definición clave que nos servirá, más adelante, como medida de complejidad: la información conjunta. Transferencia de información o, simplemente, información:

**Definición**

La información que aporta el conocimiento de \( Y \) sobre \( X \), \( I(X,Y) \), se define por:

\[
I(X,Y) = H(X) - H(X|Y) \geq 0
\]

De forma inmediata, se tiene que \( I(X,Y) \) se mide también por:

\[
I(X,Y) = H(X) + H(Y) - H(X,Y)
\]

que emplearemos en nuestro análisis cuantitativo de la definición de complejidad. Vemos que

\[
I(X,Y) \leq H(X) + H(Y)
\]

lo que en la práctica equivale a decir que en ausencia de correlaciones entre ambos conjuntos (si las probabilidades conjuntas son \( P(A_i, B_j) = \delta_{ij} \)) la información es la suma simple de entropías independientes. Este tipo de medidas conjuntas se ha empleado en distintos campos. Una posible aplicación a la estructura de las redes de transferencia de energía en ecosistemas parece indicar que los valores observados de información y entropía conjuntas están acotados en un estrecho margen (Wagensberg et al., 1991).

Como veremos más adelante, numerosos sistemas parecen presentar, cuando evolucionan en el tiempo, un comportamiento especial (¿la complejidad?) que podríamos situar a medio camino entre los estados de orden y los de “caos”: Orden → Complejidad → Caos. Para esta ordenación, veremos que la información nos servirá, en general, de medida de complejidad.

### 1.6 Información en canales con ruido

Consideremos un canal en el que se emplea un alfabeto de entrada \( A = \{A_1, \ldots, A_n\} \) y de salida \( B = \{B_1, \ldots, B_m\} \) así como cierto conjunto de relaciones estadísticas entre entradas y salidas. Está claro que la salida (un símbulo dado de \( B \)) será en general una función del conjunto de símbolos de entrada así como, eventualmente, de los símbolos de salida que le precedieron, y del “estado” del canal. En esta sección daremos algunas definiciones básicas y analizaremos el problema de los canales con ruido.

**Definición**

Un canal de comunicación (CC) se denomina canal sin memoria (CSM) cuando la aparición de un símbulo dado en la salida depende únicamente del símbulo de entrada presente en ese instante, y no de los símbolos precedentes.

Aquí nos limitaremos a este tipo de canal, y las relaciones de tipo estadístico a las que aludíamos serán el conjunto de \( n \times m \) probabilidades condicionadas

\[
p_{ij} = P(B_j|A_i) : \quad B_j \in B; \quad A_i \in A
\]

de obtener el símbulo \( B_j \in B \) cuando se ha emitido \( A_i \in A \). Las definiciones previas de entropías (condicionada, conjunta, etc.) así como la información \( I(A,B) \) son aplicables. En particular, \( I(A,B) \) nos dará una medida de la cantidad media de información sobre la entrada recibida a la
salida. La información depende no sólo de las probabilidades condicionadas \( p_{ij} \) sino también de las probabilidades de entrada \( \{P(A_i)\} \).

Para una matriz \( \{p_{ij}\} \) dada, existirá en general una distribución de entrada que haga máxima la información. Este valor,

\[
C \equiv \max_{P(A_i)} [I(A, B)]
\]

se denominaidad de canal.

A continuación definiremos varios tipos de canal de comunicación, que verifican algunas propiedades de interés.

**Canal sin pérdidas**

Un CC se denomina canal sin pérdidas si \( H(A|B) = 0 \) para cualquier distribución de entrada \( \{P(A_i)\} \). En este caso, se tiene:

\[
C \equiv \max_{P(A_i)} [I(A, B)] = \max_{P(A_i)} [H(A)] = \log (n)
\]

y como vemos en este caso la salida determina la entrada de forma única.

**Canal determinista**

Un CC se denomina determinista si \( H(B|A) = 0 \) para cualquier distribución de entrada \( \{P(A_i)\} \). Esta restricción nos lleva a:

\[
C = \max_{P(A_i)} [H(B)] = \log (m)
\]

y ahora la entrada determina la salida de forma única.

**Canal sin ruido**

Un CC se denomina canal sin ruido si cumple las dos condiciones anteriores: es por lo tanto un canal sin pérdidas y determinista. Ahora tenemos una relación bidireccional entre entrada y salida, y obviamente \( n = m \).

**Canal independiente**

Si la entrada y la salida son independientes, entonces \( I(A, B) = 0, \forall \{P(A_i)\} \), su capacidad es cero y no permite la transmisión de información.

**Canal simétrico respecto a la entrada**

Se llama así al canal en el que cada fila de la matriz de probabilidades \( P(B_j|A_i) \) contiene (en algún orden) los mismos elementos. Un ejemplo sería \( n = 2, m = 3 \):

\[
\Pi = \begin{pmatrix}
1/3 & 1/2 & 1/6 \\
1/6 & 1/2 & 1/3
\end{pmatrix}
\]

Observemos que en estos canales se cumple:

\[
H(B|A_i) = - \sum_{j=1}^{m} P(B_j|A_i) \log P(B_j|A_i) = H(B|A_i) = H_0
\]
siendo $A, \in A$ cualquier otro símbolo del alfabeto $A$. Aquí $H_0$ es independiente de $r$, y la entropía condicionada es:

$$H(B|A) = - \sum_{j=1}^{m} P(A_i)H(B|A_i) = H_0 \sum_{i=1}^{n} P(A_i) = H_0$$

y por lo tanto no depende de las probabilidades de entrada.

**Canal simétrico respecto a la salida**

Un canal de este tipo nos presenta los mismos elementos en todas las columnas. Un ejemplo sería:

$$\Pi = \begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \\ 2/3 & 1/3 \end{pmatrix}$$

Para este canal puede probarse que si $P(A_i) = 1/n$ entonces $P(B_j) = 1/m$.

**Canal simétrico**

Un canal de comunicación se llama **totalmente simétrico** (o simplemente **simétrico**) si es simétrico en la entrada y la salida. Por ejemplo:

$$\Pi = \begin{pmatrix} 1/3 & 1/3 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/3 & 1/3 \end{pmatrix}$$

Para este canal, se tiene:

$$H(B|A_i) = H(B|A_k) = H(B|A) = H_0$$

y por lo tanto

$$I(A, B) = H(B) - H_0$$

Puede probarse entonces (empleando la desigualdad $H(B) \leq \log(m)$) que

$$C = \max_{P(A_i)} [H(B) - H_0] = \log (m) - H_0$$

lo cual se obtiene para entradas equiprobables.

### 1.7 Determinación de la capacidad

Para canales simétricos, el cálculo de la capacidad no ofrece grandes dificultades. El caso más general, asimétrico, no siempre es analíticamente resoluble, pero existe un procedimiento general analítico para el caso de matrices II cuadradas no singulares.

Se trata de encontrar el máximo de la función

$$I(A, B) = H(B) - H(B|A) \geq 0$$

con las siguientes restricciones:

$$P(A_i) \geq 0 \; \forall A_i \in A$$

$$\sum_{i=1}^{n} P(A_i) = 1$$
El cálculo puede llevarse a cabo mediante el método de los multiplicadores de Lagrange, buscando el máximo de $C$ bajo las ligaduras anteriores. Para simplificar la notación, indicaremos:

$$x_i = P(A_i)$$

$$y_j = P(B_j) = \sum_{i=1}^{n} P(A_i, B_j) = \sum_{i=1}^{n} P(A_i)P(B_j|A_i) = \sum_{i=1}^{n} p_{ij}x_j$$

e indicaremos $H(B|A_i)$ en la forma:

$$H_i = H(B|A_i) = -\sum_{j=1}^{n} P(B_j|A_i) \log (P(B_j|A_i)) = -\sum_{j=1}^{n} p_{ij} \log (p_{ij})$$

Empleando esta notación, construimos la función:

$$G(x_1, ..., x_n) \equiv I(A, B) + \lambda \left( \sum_{i=1}^{n} x_i - 1 \right) = H(B) - H(B|A) + \lambda \left( \sum_{i=1}^{n} x_i - 1 \right)$$

y trataremos de resolver la ecuación variacional

$$\delta G(x_1, ..., x_n) = 0$$

para lo cual debemos calcular

$$\frac{\partial H(B)}{\partial x_k} = \sum_{j=1}^{n} \frac{\partial H(B)}{\partial y_j} \left( \frac{\partial y_j}{\partial x_k} \right) = -\sum_{j=1}^{n} p_{kj} \left[ \log (y_j) + \log e \right]$$

y además:

$$\frac{\partial H(B|A)}{\partial x_k} = H_k = -\sum_{j=1}^{n} p_{kj} \log (p_{kj})$$

lo que nos da

$$\frac{\partial G}{\partial x_k} = \lambda - H_k - \sum_{j=1}^{n} p_{kj} \left[ \log (y_j) + \log e \right] \quad k = 1, ..., n$$

Si empleamos la condición de normalización, $\sum_j p_{kj} = 1$, y la notación

$$\Lambda_j = -\log(y_j) - \log e + \lambda$$

tendremos un sistema de ecuaciones:

$$\sum_{j=1}^{n} p_{kj} \Lambda_j = H_j \quad j = 1, ..., n$$

Puesto que la matriz es no-singular, podemos hallar la solución del sistema anterior. $\Lambda_j = b_j$, que nos permita escribir:

$$-\log(y_j) - \log e + \lambda = b_j$$

o lo que es lo mismo,

$$y_j = \frac{1}{e} e^{\lambda - b}.$$
Ahora haremos uso de las ligaduras impuestas en la búsqueda de $C$. La normalización es ahora

\[
\frac{1}{e} \sum_{j=1}^{n} 2^{\lambda - b_j} = 1
\]

lo que nos lleva a obtener el multiplicador de Lagrange,

\[
\lambda = \log \left[ \frac{e}{\sum_{j=1}^{n} 2^{-b_j}} \right]
\]

que permitirá calcular $C$. Volviendo a la ecuación variacional tenemos que:

\[
\log e - \lambda = - \sum_{j=1}^{n} p_{kj} \log (y_j) - H_k
\]

que, después de multiplicar por $x_k$ y sumar, nos da:

\[
\sum_{k=1}^{n} x_k [\log e - \lambda] = - \sum_{k=1}^{n} \sum_{j=1}^{n} x_k p_{kj} \log (y_j) - \sum_{k=1}^{n} x_k H_k
\]

esto es

\[
\log e - \lambda = \max \{ B(B) - H(B|A) \} = C
\]

Sustituyendo el valor hallado de $\lambda$, obtenemos la capacidad del canal,

\[
C = \log \left( \sum_{j=1}^{n} 2^{-b_j} \right)
\]

El resultado que hemos obtenido será válido ($C > 0$) sólo si la desigualdad $\log e - \lambda > 0$ se cumple.

### 1.8 Canal binario

Este es un caso particularmente simple del problema analizado en la última sección, para un canal con matriz $\Pi$ de $n = 2$. Será de la forma:

\[
\Pi = \begin{pmatrix}
1 - p & p \\
q & 1 - q
\end{pmatrix}
\]

y en este caso debemos resolver el sistema:

\[
\Pi \left( \begin{array}{c}
\Lambda_1 \\
\Lambda_2
\end{array} \right) = \left( \begin{array}{c}
H_1 \\
H_2
\end{array} \right)
\]

donde ahora tendremos:

\[
y = P(B_1) = x(1 - p) + (1 - x)q
\]

\[
1 - y = P(B_2) = xp + (1 - x)(1 - q)
\]

\[
H(B) = -y \log (y) - (1 - y) \log (1 - y)
\]

\[
H(B|A) = x h(p) + (1 - x)h(q)
\]

Con lo que:
\[ \Pi^{-1} = \frac{1}{1-p-q} \begin{pmatrix} 1-q & -p \\ -q & 1-p \end{pmatrix} \]

y la solución se obtiene de resolver la ecuación

\[
\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \Pi^{-1} \begin{pmatrix} H_1 \\ H_2 \end{pmatrix}
\]

lo que nos da:

\[
b_1 = \frac{(1-q)h(p) - ph(q)}{1-p-q}
\]

\[
b_2 = \frac{-qh(p) - (1-p)h(q)}{1-p-q}
\]

luego la capacidad para este canal es:

\[
C = \log \left( 2^{-b_1} + 2^{-b_2} \right)
\]

### 1.9 Información mutua y función de correlación

La información conjunta, también conocida como información mutua, es una medida de la dependencia entre variables. Es además una medida de carácter general, que tiene en cuenta correlaciones de todo tipo (y no sólo lineales) y que emplearemos en distintos lugares de este texto. En cualquier caso, si las variables son independientes, \( I \) será cero, y su valor será tanto mayor cuanto mayor sea la correlación entre ambas (con algunas precisiones).

Otro tipo de magnitud muy empleada en física (y en el estudio de los sistemas complejos) es la denominada función de correlación. Es también (como su nombre indica) una medida de dependencia (correlación) entre variables, aunque en este caso la correlación es de tipo lineal. Junto a esta diferencia importante, existe otra aún más relevante en el campo de estudio que nos ocupa: la información mutua puede ser empleada sobre secuencias de símbolos y sobre secuencias numéricas, mientras que la función de correlación sólo puede aplicarse en el segundo caso. Dado que en la mayoría de los casos dispondremos de cierta información parcial acerca de nuestro sistema, que aparecerá como una colección ordenada de símbolos (por ejemplo, binarios), el interés de la información mutua sobrepasa al de la función de correlación.

En esta sección compararemos ambas medidas con el objeto de comprender sus propiedades y su relación. Supongamos para empezar que nuestro conjunto es una secuencia finita

\[ \{x_i\} \ ; \ i = 1, 2, ..., N \]

dentro de un conjunto de estados posibles \( \Sigma \):

\[ x_i \in \Sigma \equiv \{a_j\} \ ; \ j = 1, 2, ..., K \]

La función de correlación (capítulo 7) se definirá aquí por:

\[
\Gamma(r) \equiv \sum_{j}^{K} \sum_{i}^{K} a_i a_j P_{ij}(r) - \left( \sum_{j}^{K} a_j P_j \right)^2
\]

donde las probabilidades de cada símbolo \( P_j \) y la probabilidad conjunta de tener los símbolos \( a_i \) y \( a_j \), separados una distancia \( r \), esto es, \( P_{ij}(r) \), se calculan sobre la secuencia. La información mutua, definida también sobre pares separados una distancia \( r \), puede escribirse en la forma:
\[ I^{(1)}(r) \equiv \sum_{j} \sum_{i} P_{ij}(r) \log \left( \frac{P_{ij}(r)}{P_{i} P_{j}} \right) \]

que puede ser fácilmente generalizada (Li, 1990) para conjuntos de “bloques” (secuencias de tamaño \( L \)) en la forma:

\[ I^{(L)}(r) \equiv \sum_{j} \sum_{i} P_{ij}(r) \log \left( \frac{P_{ij}(r)}{P_{i} P_{j}} \right) \]

Podemos restringir nuestra comparación a secuencias binarias (esto es \( \Sigma = \{0, 1\} \)) y analizar la relación entre la función de correlación y la información \( I^{(1)}(r) \), que de ahora en adelante indicaremos simplemente por \( I(r) \). Para nuestra cadena de símbolos binarios, el número de probabilidades conjuntas independientes se reduce de cuatro a una. Como consecuencia, la función de correlación puede relacionarse de forma directa con la información mutua.

Puesto que sólo los valores no nulos pueden contribuir a la función de correlación, tendremos que:

\[ \Gamma(r) = P_{11}(r) - P_{1}^{2} \]

siendo \( P_{11}(r) \) la probabilidad conjunta de tener dos símbolos “1” separados una distancia \( r \), y \( P_{1} \) la probabilidad de observar dicho símbolo. Podemos encontrar una relación formal entre ambas funciones del siguiente modo (Li, 1990). En primer lugar, supongamos que la secuencia no posee una dirección particular, esto es, la secuencia puede ser analizada de derecha a izquierda y viceversa. Esta restricción de simetría nos lleva a la igualdad:

\[ P_{ij} = P_{ji}, \]

para \( i, j \in \{0, 1\} \). En segundo lugar, de la definición de probabilidad conjunta tenemos que:

\[ P_{1} = \sum_{j=0,1} P_{ij}(r) \]

Y finalmente una condición de normalización dada por:

\[ \sum_{i=0,1} \sum_{j=0,1} P_{ij}(r) = 1 \]

que es equivalente a la condición \( \sum_{j} P_{j} = 1 \) y no proporciona de hecho más restricciones al número de valores independientes de \( P_{ij} \). El número de variables independientes se reduce a una. Tenemos:

\[ P_{01}(r) = P_{10}(r) = P_{1} - P_{11}(r) \]
\[ P_{00}(r) = (1 - 2P_{1}) + P_{11}(r) \]

que, en términos de la función de correlación, nos dan:

\[ P_{11}(r) = \Gamma(r) + P_{1}^{2} \]
\[ P_{00}(r) = \Gamma(r) + P_{0}^{2} \]
\[ P_{01}(r) = P_{10}(r) = -\Gamma(r) + P_{0} P_{1} \]

Así, la relación general entre la información mutua y la función de correlación entre variables binarias es:
\[ I(r) = \Gamma(r) \log \left( \frac{(1 + \Gamma(r)/P_0^2)(1 + \Gamma(r)/P_1^2)}{(1 - \Gamma(r)/P_0^2)^2} \right) + P_0^2 \log \left( 1 + \frac{\Gamma(r)}{P_0^2} \right) + P_1^2 \log \left( 1 + \frac{\Gamma(r)}{P_1^2} \right) + 2P_0P_1 \log \left( 1 - \frac{\Gamma(r)}{P_0P_1} \right) \]

De esta expresión podemos obtener una aproximación útil cuando la función de correlación decrece a cero para valores grandes de \( r \) y \( \Gamma(r)/P_0, P_1 \) son pequeños. En este límite los términos de primer orden de \( \Gamma(r) \) se cancelan y sólo nos quedan los de segundo orden:

\[ I(r) \approx \frac{\Gamma^2(r)}{2} \left( \frac{1}{P_0^2} + \frac{1}{P_1^2} + \frac{2}{P_0P_1} \right) = \left( \frac{\Gamma(r)}{P_0P_1} \right)^2 \]

De esta última expresión vemos que la información decrece a cero con mayor rapidez que la función de correlación. Así, si \( \Gamma(r) \propto r^{-\beta} \) entonces tendremos \( I(r) \propto r^{-2\beta} \). Como veremos más adelante, la información mutua nos servirá de medida efectiva de complejidad. Sus propiedades y especialmente su generalidad (no considera sólo correlaciones lineales) la hacen especialmente apta en nuestro estudio.

### 1.10 Complejidad: algunos comentarios

Nuestro interés por el valor máximo de \( I(A, B) \) es debido a la propuesta teórica (realizada por diversos autores) de esta cantidad como medida de complejidad. Aunque hemos hablado de canales de comunicación, estas medidas son aplicables a sistemas complejos de cualquier tipo. En estos sistemas, podemos medir la información transmitida entre dos subsistemas (de algún tipo) a partir de un conjunto adecuado de probabilidades. Este hecho hace particularmente útil este conjunto de medidas.

Esta propuesta tiene mucho que ver con la aparición de cierto tipo de estructuras complejas (los objetos fractales) y de correlaciones de gran alcance en puntos críticos, bien conocidos en física (véase capítulo 7). Observemos que la información \( I(A, B) \) puede ser pequeña en dos situaciones extremas, de gran interés en nuestro estudio. La primera es la más obvia: si \( H \) corresponde a la matriz de un sistema en el que la entrada y la salida son totalmente independientes (de forma que las partes que intercambian información no influyen de hecho entre sí) entonces \( H(A, B) = H(A) + H(B) \), y la información es nula. Si imaginamos un sistema particular como una colonia de insectos sociales (fig. 1.8), la idea sería que el cambio de estado de un elemento, aunque pueda transmitirse, lo hace tan defectuosamente que de hecho no es "entendido" por el sistema, y los elementos son básicamente independientes.

Otra situación, menos evidente, aparece cuando los elementos del sistema se comportan exactamente de la misma forma (por ejemplo: permanecen todos en el mismo estado o cambian exactamente igual). En este caso los elementos de la matriz serían:

\[ \Pi_{ij} = \delta_{ij} \quad \forall \ i, j = 1, 2, ..., n \]

de manera que la información queda reducida a la entropía de un subsistema, \( I(A, B) = H(A) \). Así, si \( n = 2 \), y una de las probabilidades (en el alfabeto) es muy pequeña, tendremos nuevamente \( I(A, B) \approx 0 \).

Los dos casos indicados son de interés por un buen motivo: corresponden a dos situaciones extremas dentro de las posibilidades dinámicas de un sistema complejo. La primera corresponde...

a una situación de azar, de independencia entre elementos. En el segundo caso no tenemos (necesariamente) una transmisión real de información, sino que los elementos del sistema presentan una sincronía asociada a una dinámica uniforme para todos ellos o bien al seguimiento de una señal externa. En el primer caso, puede ocurrir que los elementos del sistema envíen información, pero ésta se pierde de alguna forma y por lo tanto no puede almacenarse. En el caso opuesto, la sincronización (que puede deberse a un fenómeno dinámico interno, como veremos) es tan fuerte que todos los elementos se igualan entre sí. Puede decirse que el sistema almacena información pero cualquier intento de transmitirla se enfrenta con la rigidez del propio conjunto (Solé y Miramontes, 1995).

El motivo de esta discusión es que los sistemas complejos, que almacenan y transmiten información, no parecen ocupar ninguno de los extremos antes expuestos. Es necesario transmitir información para procesar las señales procedentes del exterior del sistema (estamos suponiendo, naturalmente, sistemas abiertos) así como del propio sistema. Pero también es necesario conservarla en alguna forma, ya sea permanente o temporalmente. Tomemos nuevamente un grupo de insectos sociales, a los que volveremos más adelante. Si uno de ellos descubre una fuente de alimento o percibe un cambio externo, esta información debe ser transmitida de modo efectivo a través de la colonia, alcanzando a un número mayor o menor de individuos. Por otra parte, la señal debe ser mantenida (esto es, almacenada) durante cierto tiempo. Esta memoria a corto plazo permite responder a la señal. La complejidad emerge por lo tanto a medio camino entre el orden y el desorden (más adelante hablaremos de orden y caos) y ambos son necesarios. ¿Es entonces cualquier punto intermedio igualmente probable? o, ¿existe algún punto especial en el que la complejidad suela aparecer? La segunda pregunta tiene una respuesta afirmativa (Solé et al., 1996).
1.11 Apéndice. Procesos estocásticos

El ejemplo paradigmático de proceso dinámico al azar viene perfectamente representado por el llamado paseo al azar que, en la literatura especializada, suele indicarse a menudo como random walk (introducido anteriormente). Obtendremos ahora analíticamente la ecuación asociada al proceso de difusión y su generalización, la ecuación de Fokker-Planck.

Supongamos dado un espacio unidimensional, en el cual una partícula puede desplazarse al azar a derecha o izquierda con igual probabilidad. Supongamos que la probabilidad de que la partícula lleve a cabo un salto durante el intervalo temporal \((t, t + \delta t)\) es

\[
P(t, t + \delta t) = \beta \delta t + O(\delta t)
\]

donde \(O(\delta t)\) indica (posibles) términos de orden superior. Si los saltos son de cierta longitud fija, \(\delta x\), entonces las probabilidades de transición serán:

\[
P(x \rightarrow x + \delta x) = \frac{1}{2}
\]

\[
P(x \rightarrow x - \delta x) = \frac{1}{2}
\]

Sea \(P_i(t) \equiv P[x(t) = i \delta x]\). A partir de las hipótesis anteriores, tenemos que:

\[
P_i(t + \delta t) = (1 - \beta \delta t)P_i(t) + \frac{1}{2} \beta \delta t \left[ P_{i-1}(t) + P_{i+1}(t) \right] + O(\delta t)
\]

Esto es, podemos escribir

\[
P_i(t + \delta t) = -P_i(t) - \beta \delta t P_i(t) + \frac{1}{2} \beta \delta t \left[ P_{i-1}(t) + P_{i+1}(t) \right] + O(\delta t)
\]

luego dividiendo por \(\delta t\) y tomando el límite \(\delta t \rightarrow 0\), obtenemos:

\[
\frac{dP_i(t)}{dt} = -\beta P_i(t) + \frac{1}{2} \beta \left[ P_{i-1}(t) + P_{i+1}(t) \right]
\]

A continuación eliminaremos la dependencia en la variable \(x\) respecto de la discretización \(\delta x\). Si empleamos la notación \(P_i(t) = f(x, t)\), podemos escribir

\[
\frac{\partial f(x, t)}{\partial t} = -\beta f(x, t) + \frac{1}{2} \beta \left[ f(x - \delta x, t) + f(x + \delta x, t) \right]
\]

Desarrollando en serie de Taylor los términos en diferencias, obtenemos ahora:

\[
\frac{\partial f(x, t)}{\partial t} = -\beta \frac{\partial f(x, t)}{\partial x} \delta x + \frac{\beta}{2} \frac{\partial^2 f(x, t)}{\partial x^2} \left( \frac{\delta x}{2} \right)^2 +
\]

\[
+ \frac{\beta}{2} \frac{\partial f(x, t)}{\partial x} \delta x + \frac{\beta}{2} \frac{\partial^2 f(x, t)}{\partial x^2} \left( \frac{\delta x}{2} \right)^2 + ...
\]

\[
= \frac{\beta}{2} \frac{\partial^2 f(x, t)}{\partial x^2} + ...
\]

Aquí \(\beta\) se tomará proporcional a \(1/(\delta x)^2\). Si definimos \(a \equiv \beta(\delta x)^2\), obtenemos finalmente la ecuación estocástica:

\[
\frac{\partial f(x, t)}{\partial t} = \frac{1}{2} a \frac{\partial^2 f(x, t)}{\partial x^2}.
\]
que es una ecuación de difusión. Sus soluciones nos proporcionarán la densidad de probabilidad $f(x, t)$ de encontrar la partícula en algún lugar del espacio. Puede comprobarse que una solución posible de esta ecuación, compatible con la condición inicial $v(x, 0) = \delta(0)$, es con una partícula situada en el origen en el instante inicial, será:

$$f(x, t) = \frac{1}{\sqrt{2\pi at}} \exp \left[ -\frac{x^2}{2at} \right]$$

Si comparamos esta solución con la obtenida anteriormente de forma numérica para un conjunto de partículas moviéndose al azar, obtenemos, como cabía esperar, un comportamiento muy similar. La probabilidad de encontrar una partícula dada $f(x, t)$ debe coincidir con las frecuencias de encuentro asociadas al colectivo de partículas independientes.

A continuación, consideraremos un resultado teórico general de gran importancia. Derivaremos las ecuaciones de difusión generales en el caso en que las probabilidades de transición dependan de la interacción de una forma no-trivial. Partiremos de $x(t)$, un proceso estocástico. Sea la probabilidad de transición (infinitesimal) definida por:

$$q(x, \xi, \delta t) = P[\xi \leq \delta x \leq \xi + d\xi \mid x(t) = x]$$

Donde como sabemos debe cumplirse

$$\int q(x, \xi, \delta t) d\xi = 1$$

Luego tendremos:

$$f(x, t + \delta t) = \int f(x - \xi, t)q(x - \xi, \xi, \delta t) d\xi + O(\delta t) \quad (1.A.1)$$

$f(x, t)$ puede ser representada en la forma

$$f(x, t) = \int q(x, \xi, \delta t) d\xi \quad (1.A.2)$$

Si restamos 1.A.2 de 1.A.1, obtenemos:

$$\delta f(x, t) = f(x, t + \delta t) - f(x, t) = \int \left[ (fq)_{x-\xi, t} - (fq)_{x, \xi, t} \right] q d\xi$$

Desarrollando el integrando en serie de Taylor, obtenemos:

$$\delta f = \int \left[ -\xi \partial_x (fq) + \frac{1}{2} \xi^2 \partial_x^2 (fq) + \ldots \right] d\xi$$

$$\delta f = -\partial_x \int \xi (fq)_{x, \xi, t} d\xi + \frac{1}{2} \partial_x^2 \int \xi^2 (fq)_{x, \xi, t} d\xi + \ldots$$

$$\delta f = -\partial_x \left[ f(x, t) \int \xi q(x, \xi, \delta t) d\xi \right] + \frac{1}{2} \partial_x^2 \left[ f(x, t) \int \xi^2 q(x, \xi, \delta t) d\xi \right] + \ldots$$

Ahora realizaremos dos hipótesis acerca de los momentos de la distribución $f(x, t)$. Estas son:

$$E[\delta x \mid x(t) = x] = \int \xi q(x, \xi, \delta t) d\xi = b(x)\delta t + O(\delta t)$$
\[ E[\delta x^2 | x(t) = x] = \int \xi^2 q(x, \xi, \delta t) d\xi = a(x) \delta t + O(\delta t) \]

De estas expresiones se sigue que:

\[ \delta f(x, t) = \delta t \left[ -\frac{\partial}{\partial x} (b(x) f(x, t)) + \frac{1}{2} \frac{\partial^2}{\partial x^2} (a(x) f(x, t)) \right] \]

esto es,

\[ \frac{\partial f(x, t)}{\partial t} = -\frac{\partial}{\partial x} (b(x) f(x, t)) + \frac{1}{2} \frac{\partial^2}{\partial x^2} (a(x) f(x, t)) \]

expresión que se conoce como ecuación de Fokker-Planck. Existe un tratamiento general de los problemas de autoorganización (planteados a lo largo de este libro) basado en el empleo de este formalismo (Nicolis y Prigogine, 1977; Haken, 1988).

Bibliografía

Capítulo 2

Sistemas Dinámicos

En este capítulo estudiaremos en qué forma podemos representar, analizar e interpretar los sistemas físicos que presentan una variación de las magnitudes que los definen (la posición, la velocidad, el número de elementos activos, la presión, o variables más abstractas, que pueden ser incluso combinación de las anteriores) en el tiempo. De forma genérica, estos sistemas se denominan sistemas dinámicos. Las ecuaciones que los representan serán de la forma

$$\frac{dx}{dt} = \dot{x} = F(x, t; \mu)$$

y entonces diremos que el sistema dinámico es continuo y está representado por una ecuación diferencial ordinaria (EDO), o bien

$$x \longrightarrow g(x; \mu)$$

caso en que se denomina discreto y la ecuación lo representa es entonces una aplicación, o ecuación en diferencias. El conjunto de variables del sistema está representado por $x$, que es un punto de un conjunto abierto en $\mathbb{R}^n$: $x \in U \subseteq \mathbb{R}^n$. El tiempo se representa usualmente por $t \in \mathbb{R}$, aunque se puede considerar, de forma más general, como la variable encargada de parametrizar las trayectorias (o soluciones), descritas por el vector $x$. El vector $\mu$ representa una colección de $p$ parámetros, $\mu \in \mathbb{R}^p$ que son fijados en cada evolución del sistema, como por ejemplo la cantidad de energía suministrada o perdida cada tiempo $T$, la viscosidad de un fluido, una diferencia de temperaturas o muchos otros. Finalmente, $F \in \mathbb{R}^n$ y puede ser en general cualquier función arbitraria.

Una solución de una ecuación diferencial tiene la interpretación geométrica de una curva en $\mathbb{R}^n$, y la ecuación diferencial proporciona el vector tangente en cada punto. Debido a ello, a una ecuación diferencial ordinaria se la denomina también campo de vectores. El espacio $\mathbb{R}^n$, donde no aparece el tiempo $t$, sino las variables $x$ dependientes se denomina espacio de fases. La geometría de las soluciones en el espacio de fases nos proporcionará toda la información sobre la dinámica del sistema.

Cuando integremos una EDO aparecerá una constante $C$ que puede ser fijada de forma arbitraria. Usualmente se debe dar, junto con la EDO, lo que se llama condición inicial a fin de obtener una solución única para la dinámica del sistema. La condición inicial (c.i.) consiste en la especificación del punto $x_0$ en donde se encuentra el sistema en un instante $t_0$ determinado. Veamos que la condición inicial puede hacer que el sistema evolucione hacia un estado estacionario.

---

1 Para una descripción breve de la teoría de conjuntos requerida en este capítulo véase la sección 3.2.
Figura 2.1: Variación temporal de la posición y la velocidad en el movimiento circular uniforme. A la derecha, representación del espacio de fases, independiente del tiempo.

fijo en el tiempo, o bien oscile, o escape al infinito, en una forma que será especificada a lo largo del capítulo.

Ilustraremos con un ejemplo sencillo las definiciones dadas. Consideremos el siguiente sistema:

\[ \begin{align*}
    \dot{x} &= v \\
    \dot{v} &= -x
\end{align*} \]

En este caso, \( n = 2 \), y ésta es la dimensión del espacio de fases. Si tomamos la c.i. \( t_0 = 0, (x(t_0), v(t_0)) = (0, 1) \), la solución del sistema en función del tiempo es

\[ (x(t), v(t)) = (\sin(t), \cos(t)) \]

Para obtener una ecuación independiente del tiempo (trayectoria en el espacio de fases) es suficiente, en este caso, con elevar ambas variables al cuadrado,

\[ x^2 + v^2 = 1 \]

que es la ecuación de una circunferencia. La representación de cada variable en función del tiempo y el correspondiente espacio de fases se puede ver en la figura 2.1.

En todos los sistemas dinámicos tratados en este libro se supone que la solución de la EDO correspondiente existe. Obsérvese que prácticamente todas las EDOs que veremos pertenecerán a sistemas físicos, reales, que efectivamente tienen una cierta evolución temporal, por tanto las soluciones de las ecuaciones que los describen deben existir, ya sea en forma analítica o numérica. El lector interesado en la existencia y unicidad de solución de las EDOs puede consultar Arnold (1995) o Boyce y DiPrima (1983), por ejemplo.

En cuanto a las aplicaciones, \( x \rightarrow g(x; \mu) \), rara vez seremos capaces de encontrar una solución cerrada para la iteración \( m \)-ésima en función de la condición inicial. La aplicación también se escribe usualmente en la forma

\[ x_{m+1} = g(x_m; \mu) \]
donde se indica que el punto $x_{m+1}$ depende de la localización del sistema en la iteración anterior ($x_m$) y del conjunto de parámetros $\mu$. La dinámica de las aplicaciones no es una curva continua, sino una colección discreta de puntos. Entre un punto y el siguiente dista una iteración, que representa cierto intervalo de tiempo que depende de cada aplicación. Este intervalo puede ser el tiempo de respuesta de un aparato de medida, por ejemplo, o en un caso que veremos en todo detalle, el tiempo entre generaciones en una especie animal, descrito por la llamada aplicación logística.

Consideremos un caso particular sencillo de aplicación discreta en una dimensión:

$$x_{m+1} = \mu x_m$$

Si partimos de cierto punto inicial $x_0$, las iteraciones sucesivas serán $x_1 = \mu x_0$, $x_2 = \mu x_1 = \mu^2 x_0$, $x_3 = \mu x_2 = \mu^2 x_1 = \mu^3 x_0$, y en general,

$$x_k = \mu^k x_0$$

Para este caso sencillo si podemos obtener una solución analítica para cualquier iteración conociendo el punto inicial. Además, en función del valor de $\mu$ podemos describir rápidamente la dinámica del sistema:

1. si $|\mu| < 1$ el sistema tiende al punto final $x_f = 0$, de forma monótona decreciente si $\mu > 0$ y de forma oscilante si $\mu < 0$,

2. si $|\mu| > 1$ el sistema diverge a infinito para $k \to \infty$, de forma monótona creciente u oscilante dependiendo del signo de $\mu$, como en el caso anterior,

3. si $|\mu| = 1$ el sistema permanece en el punto inicial $x_0$ para $\mu > 0$, y oscila entre $x_0$ y $-x_0$ si $\mu < 0$.

La figura 2.2 representa la evolución temporal del sistema para $\mu = -0.9$. El espacio de fases es en este caso unidimensional.

Cuando la aplicación sea invertible, es decir, cuando exista la función inversa de $g(x;\mu)$, $g^{-1}(x;\mu)$, será posible estudiar la secuencia doblemente infinita.
\{ \ldots g^{-n}(x_0; \mu), \ldots, g^{-1}(x_0; \mu), x_0, g(x_0; \mu), \ldots, g^n(x_0; \mu), \ldots \}

donde \( g^n \) se define de forma inductiva como

\[ g^n(x_0; \mu) \equiv g(g^{n-1}(x_0; \mu)) \]

Si \( g(x; \mu) \) no es invertible, sólo es posible estudiar la secuencia

\[ \{ x_0, g(x_0; \mu), g^2(x_0; \mu), \ldots, g^n(x_0; \mu), \ldots \} \]

2.1 Sistemas dinámicos continuos

2.1.1 Sistemas lineales autónomos en \( \mathbb{R}^n \)

Un sistema \textit{autónomo} de ecuaciones diferenciales es aquel en el que el tiempo no aparece de forma explícita en las relaciones que verifican las derivadas temporales de las variables del sistema, es decir, el sistema de ecuaciones diferenciales es de la forma

\[
\frac{dx}{dt} = F(x(t); \mu) \tag{2.2.1}
\]

donde \( x \in \mathbb{R}^n \), \( \mu \) representa los parámetros del sistema y el tiempo \( t \) sólo está representado de forma implícita en la colección de variables \( x = \{x_1, x_2, \ldots, x_n\} \), y no aparece explícitamente en las ecuaciones. Además, diremos que el sistema de ecuaciones diferenciales es \textit{lineal} si en las funciones \( F \) sólo aparecen términos de la forma \( \beta_j x_i \) en la definición de la variable \( x_j \); una variable depende de las demás a través de términos multiplicativos constantes, y no a través de productos de dos o más variables o de funciones de éstas. Así, si el sistema de ecuaciones diferenciales es autónomo y lineal tiene la siguiente expresión sencilla:

\[ \dot{x} = Ax \]

donde \( A \) es una matriz de coeficiente constantes \(^2\). De forma más explícita:

\[
\begin{pmatrix}
\dot{x}_1 \\
\dot{x}_2 \\
\vdots \\
\dot{x}_n
\end{pmatrix} =
\begin{pmatrix}
\beta_{11} & \beta_{12} & \cdots & \beta_{1n} \\
\beta_{21} & \beta_{22} & \cdots & \beta_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
\beta_{n1} & \beta_{n2} & \cdots & \beta_{nn}
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n
\end{pmatrix}
\]

En este caso, la solución del sistema se puede hallar analíticamente y es posible conocer el comportamiento del sistema (dadas unas determinadas condiciones iniciales) en todo el espacio de fases del mismo.

No vamos a analizar en este libro cuáles son las técnicas que se pueden emplear para solucionar en general sistemas de ecuaciones diferenciales. Cuando tratábamos con sistemas complejos, que presentan comportamientos “complicados” en su espacio de fases, no es habitual que se pueda encontrar una solución analítica para la evolución del sistema. Esto obliga a utilizar técnicas de estudio del sistema a nivel local, esto es, cerca de los denominados \textit{puntos fijos}. Iniciamos el estudio de estos puntos fijos y del tipo de estabilidad que presentan con los sistemas lineales anteriores.

\(^2\)En principio podría aparecer un término independiente que simplemente significa una traslación del origen de coordenadas, de forma que el sistema tendría la forma más general \( x = Ax + p \). Este término independiente \( p \in \mathbb{R}^n \) puede ser reabsorbido mediante un cambio adecuado de coordenadas, como se verá más adelante.
Un punto fijo es aquel punto del espacio de fases en el que las variables del sistema presentan una variación nula (derivada nula) en ausencia de perturbaciones. Por tanto, pueden ser hallados como soluciones de la ecuación

$$Ax = 0$$

### 2.1.2 Sistemas lineales autónomos en $\mathbb{R}^2$

Utilicemos un ejemplo sencillo en $\mathbb{R}^2$ que nos permita ver de forma intuitiva cuál será el comportamiento del sistema cuando nos separemos muy ligeramente de un punto fijo. Consideremos el sistema

$$\frac{dx}{dt} = Ax = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

y supongamos que el determinante de $A$ es no nulo:

$$\det(A) = \beta_{11}\beta_{22} - \beta_{12}\beta_{21} \neq 0$$

En este caso, el origen $(x_1, x_2) = (0, 0)$ es de forma trivial un punto fijo del sistema. La soluciones de este sistema lineal están determinadas por los valores propios de la matriz $A$. La expresión

$$\det(A - \lambda I) = 0$$

proporciona en este caso dos valores para $\lambda$, los valores propios de $A$. Aquí, $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ es la matriz identidad en $\mathbb{R}^2$. Obtenemos

$$\begin{vmatrix} \beta_{11} - \lambda & \beta_{12} \\ \beta_{21} & \beta_{22} - \lambda \end{vmatrix} = (\beta_{11} - \lambda)(\beta_{22} - \lambda) - \beta_{12}\beta_{21} = 0$$

que proporciona la ecuación de segundo grado

$$\lambda^2 - TrA + \det A = 0$$

donde $TrA$ es la traza de $A$, $TrA = \beta_{11} + \beta_{22}$. Obtenemos, pues

$$\lambda_{\pm} = \frac{TrA \pm \sqrt{(TrA)^2 - 4\det(A)}}{2}$$

Consideremos también los vectores propios de la matriz $A$, que estarán dados por la ecuación

$$Ax_{\pm} = \lambda_{\pm}x_{\pm} \quad (2.2.2)$$

En este caso serán dos vectores propios, cada uno correspondiente a uno de los valores propios anteriores. El sistema lineal admite finalmente una solución exacta de la forma

$$x = c_1e^{\lambda_+t}x_+ + c_2e^{\lambda_-t}x_-$$

Las dos constantes de integración $c_1$ y $c_2$ dependen de las condiciones iniciales del sistema. Se puede observar en esta expresión general que la tendencia del sistema a largo término depende del signo de los valores propios, $\lambda_{\pm}$:
1. Si \( \Re(\lambda_\pm) < 0 \), para los dos valores propios, entonces el punto \( x_0 = (0, 0) \) es asintóticamente estable \(^3\).

2. Si existe al menos un valor propio con parte real positiva, entonces \( x_0 \) es inestable.

3. Si \( \Re(\lambda_\pm) = 0 \) para los dos valores propios, o bien la parte real es cero para uno y negativa para el otro, entonces \( x_0 \) es estable, pero no asintóticamente.

Ilustraremos lo anteriormente visto con un ejemplo concreto.

**Ejemplo**

\[
\begin{align*}
\dot{x}_1 &= x_1 + x_2 \\
\dot{x}_2 &= 4x_1 + x_2
\end{align*}
\]

En este caso, \( A = \begin{pmatrix} 1 & 1 \\ 4 & 1 \end{pmatrix} \), y sus valores propios están dados por

\[
\det(A - \lambda I) = \begin{vmatrix} 1 & 1 \\ 4 & 1 - \lambda \end{vmatrix} = \begin{vmatrix} 1 - \lambda & 1 \\ 4 & 1 - \lambda \end{vmatrix}
\]

\( (1 - \lambda)(1 - \lambda) - 4 = 0 \) proporciona la ecuación de segundo grado \( \lambda^2 - 2\lambda - 3 = 0 \), la cual tiene como soluciones

\[
\lambda_+ = 3, \quad \lambda_- = -1
\]

Los vectores propios asociados se calculan a partir de (2.2.2), o bien \( (A - \lambda I)x_\pm = 0 \). Para \( \lambda_+ = 3 \) obtenemos

\[
\begin{pmatrix} -2 & 1 \\ 4 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\]

que tiene como solución \( x_+ = (1, 2) \)^4. Para \( \lambda_- = -1 \),

\[
\begin{pmatrix} 2 & 1 \\ 4 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}
\]

que proporciona \( x_- = (1, -2) \). Así que la solución del sistema es

\[
x = c_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} e^{3t} + c_2 \begin{pmatrix} 1 \\ -2 \end{pmatrix} e^{-t}
\]

Si tomamos como condición inicial que en el instante \( t = 0 \) el sistema se hallaba en el punto \( x(t = 0) = (1, 1) \) podemos determinar las constantes de integración por simple sustitución,

\[
\begin{pmatrix} 1 \\ 1 \end{pmatrix} = c_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + c_2 \begin{pmatrix} 1 \\ -2 \end{pmatrix}
\]

que inmediatamente proporciona \( c_1 = 4/3 \) y \( c_2 = -1/3 \). Incluyendo la condición inicial escogida,

\[
x = \frac{4}{3} \begin{pmatrix} 1 \\ 2 \end{pmatrix} e^{3t} - \frac{1}{3} \begin{pmatrix} 1 \\ -2 \end{pmatrix} e^{-t}
\]

---

\(^3\) \( \Re(\lambda) \) significa **parte real** de \( \lambda \). Véase la sección 3.5.1 si se requiere alguna explicación adicional.

\(^4\) Y cualquier combinación lineal de este vector.
Los valores propios de la matriz de coeficientes constantes del sistema dinámico permiten una clasificación de los diferentes puntos fijos que se pueden hallar en \( \mathbb{R}^2 \). Utilizaremos la siguiente notación: \( p \equiv \text{Tr}(A) \), \( q \equiv \text{det}(A) \).

- **a.** Raíces reales y distintas: \((p^2 - 4q) > 0\).
  
  1. Si \( q > 0 \), las raíces tienen igual signo. Si \( \lambda_+ < 0 \), \( x_0 \) es estable.
  
  2. Si \( \lambda_+ > 0 \), \( x_0 \) es inestable.

  En estos dos casos el punto fijo se denomina **nodo**. Será estable en el primer caso (las trayectorias se moverán acercándose al punto fijo) e inestable en el segundo (las trayectorias se alejarán del punto fijo).

  3. Si \( q < 0 \), las raíces pueden tener signo diferente. En el caso de que así suceda se denomina al punto fijo **punto de silla**.

- **b.** Raíces iguales: \((p^2 - 4q) = 0\).

  La solución general de la ecuación no es la dada anteriormente (donde se suponía que los valores propios eran diferentes). El punto fijo se denomina en este caso **nodo propio**, si las trayectorias se aproximan con pendientes diferentes al punto fijo, o bien **nodo impropio**, si todas las trayectorias acaban llegando al punto fijo con la misma inclinación.

- **c.** Raíces complejas: \((p^2 - 4q) < 0\).

  Si \( p \neq 0 \), se obtiene un **foco** estable para \( p < 0 \) y un foco inestable para \( p > 0 \). Las trayectorias convergen a o divergen del punto fijo \( x_0 \) en forma espiral.

- **d.** Raíces complejas puras: \( p = 0, q > 0 \).

  Las trayectorias se cierran en este caso alrededor de \( x_0 \), y el punto fijo se denomina **centro**.

Un último caso, bastante excepcional, se puede encontrar cuando \( \text{det}(A) = 0 \), es decir, la matriz \( A \) es degenerada. En este caso, las soluciones poseen un grado de libertad: son trayectorias paralelas que convergen a una única recta (recta degenerada). La solución al sistema propuesto con esta particularidad no es única.

La clasificación anterior se puede resumir en un único esquema donde podemos situar todos los puntos fijos posibles en función de los valores que toman \( p \) y \( q \). El esquema se representa en la figura 2.3.

### 2.1.3 Ejemplos en \( \mathbb{R}^3 \)

Cuando pasamos al estudio de la estabilidad de los puntos fijos en \( \mathbb{R}^n \), podemos utilizar una total analogía con el análisis efectuado en \( \mathbb{R}^2 \). Siempre será necesario obtener el valor del punto(s) fijo(s) \( x_0 \in \mathbb{R}^n \) y estudiar el sistema de ecuaciones diferenciales en un entorno de este punto. Serán los valores propios los que permitirán de nuevo determinar la estabilidad y la tendencia de las trayectorias a los puntos fijos, independientemente de las dificultades que podamos encontrar en la obtención de la solución completa del sistema (si es que existe). Utilizaremos dos ejemplos concretos en \( \mathbb{R}^3 \) para ilustrar la forma de proceder en el estudio de la estabilidad de los puntos fijos.
Figura 2.3: Clasificación de los puntos fijos de los sistemas de ecuaciones diferenciales en dos dimensiones; \( p \equiv \) traza de \( A \), \( q \equiv \) determinante de \( A \).

**Ejemplo 1**

\[
\begin{align*}
    \dot{x} &= -ax \\
    \dot{y} &= x - 3y \\
    \dot{z} &= -(z + y)
\end{align*}
\]

Este sistema es de la forma \( \dot{x} = Ax \), con

\[
A = \begin{pmatrix}
    -a & 0 & 0 \\
    1 & -3 & 0 \\
    0 & -1 & -1
\end{pmatrix}
\]

Los puntos de equilibrio corresponden a la solución de \( \dot{x} = 0 \):

\[
    ax = 0 \\
    x - 3y = 0 \\
    z + y = 0
\]

que tiene como única solución el origen, \( x^* = (0, 0, 0) \) cuando \( a \neq 0 \). Los valores propios de \( A \) asociados a \( x^* \) serán

\[
\det(A - \lambda I) = 0 \iff \begin{vmatrix}
    -a - \lambda & 0 & 0 \\
    1 & -3 - \lambda & 0 \\
    0 & -1 & -1 - \lambda
\end{vmatrix} = 0
\]
que proporciona

\[ \lambda_1 = -a, \quad \lambda_2 = -3, \quad \lambda_3 = -1 \]

Dependerá del signo del parámetro \( a \) que el punto fijo sea estable o inestable. Para \( a > 0 \), \( x^* \) será asintóticamente estable (nodo estable), y para \( a < 0 \) será inestable (punto de silla), con dos direcciones estables, dadas por los vectores propios asociados a \( \lambda_2 \) y \( \lambda_3 \), y una inestable, caracterizada por el vector propio asociado al valor propio positivo.

**Ejemplo 2**

\[
\begin{align*}
\dot{x} &= x - y + 4z + 2 \\
\dot{y} &= 3x + 2y - z + 1 \\
\dot{z} &= 2x + y - z + 1
\end{align*}
\]

El punto fijo resulta ser \( x^* = (-1, 1, 0) \). Mediante este ejemplo veremos que un simple cambio lineal de coordenadas podemos trasladar el punto fijo al origen y estudiar el sistema como en los casos anteriores. Realicemos el cambio siguiente: \( \tilde{x} = x + 1, \tilde{y} = y - 1, \tilde{z} = z \). Derivando estas tres expresiones y sustituyendo las variables \( x, y \) y \( z \) por las nuevas con tilde se obtiene:

\[
\begin{align*}
\dot{\tilde{x}} &= \tilde{x} = x - y + 4z + 2 = \ldots = \tilde{x} - \tilde{y} + 4\tilde{z} \\
\dot{\tilde{y}} &= \tilde{y} = 3x + 2y - z + 1 = \ldots = 3\tilde{x} + 2\tilde{y} - \tilde{z} \\
\dot{\tilde{z}} &= \tilde{z} = 2x + y - z + 1 = \ldots = 2\tilde{x} + \tilde{y} - \tilde{z}
\end{align*}
\]

Tomemos pues el segundo sistema, sin términos independientes, y estudiemos el comportamiento del origen, una vez hemos visto que es equivalente al primero. La matriz lineal asociada es

\[ A = \begin{pmatrix} 1 & -1 & 4 \\ 3 & 2 & -1 \\ 2 & 1 & -1 \end{pmatrix} \]

El estudio de \( \det(A - \lambda I) = 0 \) proporciona los valores propios

\[ \lambda_1 = 1, \quad \lambda_2 = -2, \quad \lambda_3 = 3 \]

Así que el origen es un punto de silla estable en una dirección e inestable en otra. Estas direcciones están dadas (y se pueden calcular fácilmente) por los vectores propios asociados a cada uno de los valores propios, como se ha dicho anteriormente. Los vectores propios son los que permiten diagonalizar la matriz \( A \) (mediante un cambio de coordenadas dado por estos), y podríamos hacer coincidir las direcciones de estabilidad e inestabilidad con las de los ejes coordenados.

**Variedad estable, inestable y centro de los puntos fijos**

Aprovechamos la discusión iniciada sobre el papel de los vectores propios en la determinación de las direcciones de estabilidad e inestabilidad para introducir la idea de *variedad* asociada a un punto fijo.
Teorema

Consideremos el sistema lineal $\dot{x} = Ax$. Entonces $R^n = E^s \oplus E^c \oplus E^u$, donde $E^s$, $E^c$ y $E^u$ son subespacios invariantes, correspondientes a los valores propios $\lambda$ de $A$ con $\Re(\lambda) < 0$, $\Re(\lambda) = 0$, $\Re(\lambda) > 0$, respectivamente, que verifican:

$$y_0 \in E^s \implies \lim_{t \to -\infty} y(t, y_0) = 0 \quad \text{y} \quad \lim_{t \to -\infty} y(t, y_0) = \infty, \text{ exponencialmente}$$

$$y_0 \in E^u \implies \lim_{t \to -\infty} y(t, y_0) = \infty \land \lim_{t \to -\infty} y(t, y_0) = 0, \text{ exponencialmente}$$

Los subespacios $E^s$, $E^c$ y $E^u$ se denominan *variedad invariante lineal estable, centro e inestable*, respectivamente.

En el último ejemplo (2) estudiado, dados los valores propios, podemos ver que $E^s$ será un espacio de dimensión 1 (una recta) y $E^u$ un espacio de dimensión 2 (un plano). Se puede verificar que los dos vectores propios que determinarán el plano y el vector propio que determinará la recta estable son linealmente independientes, y por tanto $R^3 = E^s \oplus E^u$, el espacio de dimensión 3 se podrá descomponer como suma directa de los dos subespacios, o, dicho de otra manera, los tres vectores propios serán una base de $R^3$.

Veámoslo de forma explícita. Los vectores propios asociados a cada uno de los valores propios serán $v_i = (x_i, y_i, z_i)$, con $i = 1, 2, 3$, y se obtienen de la resolución explícita de la ecuación $(A - \lambda_i I)v_i = 0$:

$$\lambda_1 = 1 \implies \begin{pmatrix} 0 & -1 & 4 \\ 3 & 1 & -1 \\ 2 & 1 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \implies v_1 = (1, -4, -1)$$

$$\lambda_2 = -2 \implies \begin{pmatrix} 3 & -1 & 4 \\ 3 & 4 & -1 \\ 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \implies v_2 = (1, -1, -1)$$

$$\lambda_3 = 3 \implies \begin{pmatrix} -2 & -1 & 4 \\ 3 & -1 & -1 \\ 2 & 1 & -4 \end{pmatrix} \begin{pmatrix} x_3 \\ y_3 \\ z_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \implies v_3 = (1, 2, 1)$$

Una base de $E^s$ es el vector $v_2 = (1, -1, -1)$. Una base de $E^u$ la constituyen los vectores $v_1 = (1, -4, -1)$ y $v_3 = (1, 2, 1)$. Se puede comprobar que son linealmente independientes, aunque no ortogonales. Véase la figura 2.4, donde se han representado los dos subespacios determinados por los vectores propios.

### 2.1.4 Estabilidad en sistemas no lineales

Quizá es ahora el momento de formalizar algunas de las cuestiones relativas a la estabilidad de los puntos fijos de un sistema. Ya hemos visto que los valores propios de la matriz de coeficientes, en el caso lineal, proporcionan en criterio de estabilidad para los puntos de equilibrio. Dada la relativa facilidad con que podemos calcular la solución explícita de un sistema lineal, es fácil demostrar cómo los valores propios determinan la convergencia o la divergencia de pequeñas perturbaciones alrededor del punto fijo. Pero cuando comenzamos a trabajar con sistemas no lineales, es lícito cuestionarse si podemos adoptar sin ningún problema el mismo criterio de estabilidad. En esta sección justificaremos el uso de ciertos criterios y veremos cuándo el estudio del sistema lineal es suficiente para determinar el comportamiento del sistema no lineal.
Figura 2.4: Representación de los subespacios invariantes del sistema resuelto en el ejemplo 2 de la sección 2.2.3. El plano rayado es una variedad inestable, y la recta que lo intersecta es la variedad estable.

Consideremos, en primer lugar y a modo de introducción (también útil para fijar conceptos ya vistos), el comportamiento de un sistema no lineal en una dimensión alrededor de un punto de equilibrio. Supongamos una ecuación diferencial de la forma

\[ \frac{dx}{dt} = f_\mu(x) \]

Los puntos de equilibrio satisfacen \( f_\mu(x) = 0 \). Supongamos que la solución de esta ecuación proporciona el punto fijo \( x^* \), y realicemos un pequeño desplazamiento alrededor de él.

\[ y(t) = x(t) - x^* \]

con \( |y(t)| << 1 \), lo que significa que la trayectoria \( x(t) \) será muy cercana al punto fijo. Podemos entonces considerar un desarrollo

\[ \frac{dy}{dt} = f_\mu(x^* + y) = f_\mu(x^*) + y \frac{\partial f_\mu(x)}{\partial x} \bigg|_{x^*} + O(2) \]

donde indicamos por \( O(2) \) las derivadas de orden superior respecto de \( x \). Considerando únicamente el desarrollo a primer orden

\[ \frac{dy}{dt} = \frac{\partial f_\mu(x)}{\partial x} \bigg|_{x^*} y \]

que puede ser inmediatamente resuelto:

\[ y(t) = y(0)e^{\lambda t} \]

Este resultado da un crecimiento de la perturbación para

\[ \lambda = \frac{\partial f_\mu(x)}{\partial x} \bigg|_{x^*} > 0 \]
y un decrecimiento en caso contrario ($\lambda < 0$). Veremos que, en el caso marginal $\lambda = 0$ la estabilidad ha de ser decidida mediante el estudio de los términos de orden superior, ya que, si el sistema es no lineal, serán éstos los que influirán decisivamente sobre este caso de estabilidad indecible. Podemos tener varios puntos de equilibrio $x^*$, y su estabilidad puede analizarse gráficamente con mucha facilidad simplemente inspeccionando la pendiente local de la gráfica de $f_\mu(x)$ en los puntos de corte con el eje $x$, que corresponden obviamente a los puntos críticos.

Consideremos de nuevo un sistema en $\mathbb{R}^n$, ahora no lineal, de la forma general
\[
\frac{dx}{dt} = F_\mu(x(t))
\]
(2.2.3)
donde $F_\mu$ es una función arbitraria de las coordenadas del sistema. Obtendremos los puntos estacionarios como las soluciones de $F_\mu(x(t)) = 0$. En general, tendremos un conjunto de puntos de equilibrio
\[
\Gamma_\mu = \{x^*_k : F_\mu(x^*_k) = 0, \quad k \in \mathbb{N}\}
\]
que verifican la condición.

Daremos a continuación las definiciones básicas que posibilitan la definición y la clasificación de los diferentes tipos de estabilidad.

**Definición**

Sea $x_0 \in W$ un punto de equilibrio de $F_\mu(x(t)) \in C^1(W)$. Diremos que $x_0$ es un punto de equilibrio estable si y sólo si para todo entorno $U$ de $x_0 \in W$, existe un entorno $U_1 \in U$ de $x_0$ tal que cualquier solución $x(t)$ que parte de $U_1$ está definida y permanece en $U$, $\forall t > 0$.

O bien podemos considerar la definición alternativa:

Una solución $x(t)$ del problema (2.2.3) es estable en el sentido de Lyapunov si:
\[
\forall \varepsilon > 0, \forall t > 0, \exists \eta > 0; ||x_0 - x'_0|| < \eta \implies \forall t > t_0, ||x - x'|| < \varepsilon
\]
donde $x'(t)$ es solución de 2.2.3 con condición inicial dada por $x'_0$.

**Definición**

Una solución $x(t)$ del problema (2.2.3) es asintóticamente estable (en el sentido de Lyapunov) si:
\[
\forall t \in \mathbb{R}, \exists \eta > 0; ||x_0 - x'_0|| < \eta \implies \lim_{t \to \infty} ||x - x'|| = 0
\]

La siguiente definición es particularmente importante, y hace referencia a la estabilidad cualitativa global de las trayectorias del sistema.

**Definición**

El sistema definido por 2.2.3 es estructuralmente estable si se verifica que
\[
\forall \varepsilon > 0, \forall t \in \mathbb{R}, \exists \eta > 0; ||\mu - \mu'|| < \eta \implies ||x - x'|| < \varepsilon
\]
donde $x'$ es una solución de 2.2.3 con parámetro $\mu'$.

En otras palabras, 2.2.3 es un sistema estructuralmente estable si permanece equivalente a sí mismo tras una pequeña perturbación del campo vectorial.

\textsuperscript{5} $C^1$ significa que la primera derivada de una función es continua en el dominio $W$ requerido.
Ejemplo

La ecuación que describe el movimiento de un péndulo físico de longitud $l$ y masa $m$ en presencia de rozamiento (con coeficiente $\mu$) para pequeñas oscilaciones y sometido a la acción de la gravedad $g$ es

$$mgl \frac{d^2\theta}{dt^2} + \mu \frac{d\theta}{dt} + mg\theta = 0$$

para $\mu > 0$ y $\theta \approx 0$, el ángulo que el péndulo forma con la vertical. Si realizamos el cambio

$$\frac{d\theta}{dt} \equiv \frac{dx_1}{dt} = x_2$$

$$\theta = x_1$$

obtenemos

$$\frac{dx_2}{dt} = -\frac{g}{l} x_1 - \frac{\mu}{ml} x_2$$

$$\frac{dx_1}{dt} = x_2$$

dependiente de una función de la forma $x_1 = F(x_1, x_2, t)$, donde $F$ es una función no lineal. Esta ecuación puede ser escrita en el espacio de fases $(x_1, x_2)$ como un sistema no lineal

$$\dot{x}_1 = F_1(x_1, x_2, t)$$

$$\dot{x}_2 = F_2(x_1, x_2, t)$$

que representa un sistema autónomo lineal en $\mathbb{R}^2$. En ausencia de rozamiento ($\mu = 0$) el péndulo describe trayectorias cerradas en su espacio de fases. Sin embargo, un cambio infinitesimal en el valor del parámetro $\mu$ da lugar a un nuevo comportamiento del sistema. Aparece un amortiguamiento que provoca un cambio de órbitas cerradas a un punto como atractor del sistema. Si una vez $\mu \neq 0$ aumentamos su valor, obtendremos un cambio cuantitativo (suave) en la velocidad con que el sistema alcanza el punto crítico, pero no hallaremos ningún cambio cualitativo (como el paso anterior, de centro a foco) más. Aplicando la noción de estabilidad estructural podemos decir que el péndulo con rozamiento es estructuralmente estable, mientras que en ausencia de rozamiento (estado marginal) no lo es.

Teorema

Sea $W \subset \mathbb{R}^n$ abierto, y sea el sistema dinámico 2.2.3. Supongamos que $x_0$ es un punto de equilibrio estable de 2.2.3. Entonces, ningún valor propio de la matriz $\text{DF}_\mu(x_0)$ tiene parte real positiva. La matriz $\text{DF}_\mu(x_0)$ es la llamada matriz jacobiana de la función $F_\mu$, y sus elementos son las derivadas

$$[\text{DF}_\mu]_{ij} = \frac{\partial F_\mu}{\partial x_j}$$

Definición

Diremos que un punto de equilibrio es hiperbólico si $\text{DF}_\mu(x_0)$ no tiene valores propios con parte real nula.

Como consecuencia del teorema y la definición anteriores, un punto de equilibrio hiperbólico es o bien inestable o bien asintóticamente estable.

Cualquier debamos analizar la estabilidad de los puntos de equilibrio de un sistema de EDOs no lineal, hallaremos primero cuáles son estos puntos fijos y seguidamente efectuarás una linealización del sistema alrededor de éstos. Consideraremos $x_0$ como la solución de equilibrio del sistema 2.2.3 y la aproximación definida por
\[ F_\mu(x_0 + x') = F_\mu(x_0) + \left( \frac{\partial F_\mu}{\partial x} \right)_{x_0} x + \frac{1}{2} \left( \frac{\partial^2 F_\mu}{\partial x \partial x} \right)_{x_0} xx + \ldots \]

Esta expresión podría ser muy compleja, pero es posible simplificarla utilizando el principio de estabilidad lineal, que establece la relación entre el sistema no lineal y el linealizado, en el que se omiten los términos de orden superior. Esta relación será inmediatamente formalizada. Utilicemos:

\[ L_\mu = \left( \frac{\partial F_\mu}{\partial x} \right)_{x_0} \quad ; \quad h_\mu = \frac{1}{2} \left( \frac{\partial^2 F_\mu}{\partial x \partial x} \right)_{x_0} xx + \ldots \]

para los términos lineal y no lineal, respectivamente, con lo cual los sistemas

\[ \frac{\partial x}{\partial t} = L_\mu x + h_\mu(x) \quad y \quad \frac{\partial x}{\partial t} = L_\mu x \]

representan el sistema no lineal y el lineal asociado, admitiendo ambos como solución el punto fijo \( x_0 = 0 \). Los teoremas que los relacionan son los siguientes:

**Teorema de Hadamart-Perron (puntos hiperbólicos)**

Si \( F \in C^r \) y \( x_0 \) es un punto hiperbólico, entonces existen dos variedades invariantes, \( W^s \) y \( W^u \) que pasan por \( x_0 \) y son tangentes a las variedades invariantes lineales estable e inestable, respectivamente. Se verifica:

\[ x \in W^s \implies \lim_{t \to \infty} \Phi(t, x) = x_0, \text{ exponencialmente} \]
\[ x \in W^u \implies \lim_{t \to -\infty} \Phi(t, x) = x_0, \text{ exponencialmente} \]

donde \( \Phi(t, x) \) es una solución del sistema 2.2.3.

**Teorema de Hartman-Grobman**

Sea \( U \subset \mathbb{R}^n \) un subconjunto abierto que contenga el origen, sea \( F_\mu(x, t) \in C^1(U) \) y \( \Phi_t \), el flujo del sistema no lineal \( \partial_t x = F_\mu(x, t) \). Supongamos que \( F(0) = 0 \) y que la matriz \( A = DF(0) \) no posee ningún valor propio con parte real cero. Existe un homeomorfismo \(^6 H \) de un conjunto abierto \( V \) con \( 0 \in V \) en un conjunto abierto \( W \) (también con \( 0 \in W \)) tal que para cada punto \( x_0 \in V \), existe un intervalo abierto \( I_0 \in \mathbb{R} \) que contiene el cero tal que \( \forall x_0 \in V \) y \( \forall t \in I_0 \) se verifica

\[ H \circ \Phi_t(x_0) = e^{At}H(x_0) \]

es decir, \( H \) aplica las trayectorias del sistema no lineal cerca del origen en las trayectorias del sistema lineal, también cerca del origen, y preserva la parametrización.

Queda finalmente por establecer cuándo se dan las condiciones necesarias para que aparezca una variedad centro, ya que acabamos de ver la correspondencia que es posible establecer entre las variedades lineales estable \((E^s)\) e inestable \((E^u)\) y las variedades no lineales correspondientes \((W^s \text{ y } W^u)\), respectivamente, esto es, en los casos en que la parte real de los valores propios de la matriz lineal es no nula.

\(^6\)Véase la sección 3.2.3 para la definición de homeomorfismo.
**Teorema.** Existencia de variedad centro

Si $F \in C^{r+1}$ y $x_0$ es un punto fijo:

i) Existen tres variedades invariantes $W^u$, $W^s$ (que son $C^{r+1}$) y $W^c$ (que es $C^r$) tangentes a las variedades lineales estable, inestable y centro, respectivamente. El comportamiento de las soluciones en $W^u$ y $W^s$ es el mismo que en el teorema de Hartman-Grobman. Se puede demostrar que $W^c$ contiene dos subvariedades invariantes, $W^{cu}$ en $C^r$ y $W^{cs} \in C^r$ que satisfacen

$$x \in W^{cs} \implies \lim_{t \to \infty} \Phi(t, x) = x_0$$

$$x \in W^{cu} \implies \lim_{t \to -\infty} \Phi(t, x) = x_0$$

ii) Existe un homeomorfismo $h$ en un entorno de $x_0$ tal que la ecuación transformada se escribe

$$\dot{x} = -x, \text{ para } h^{-1}(x) \in W^s$$

$$\dot{y} = y, \text{ para } h^{-1}(x) \in W^u$$

$$\dot{z} = f(z), \text{ para } h^{-1}(x) \in W^c$$

Las variedades $W^s$, $W^u$ y $W^c$ se denominan variabilidad invariante estable, centro e inestable, respectivamente. La variedad centro no es única.

El comportamiento de las soluciones sobre la variedad centro depende de los términos no lineales de la ecuación, y por tanto los límites para $t \to \pm \infty$ no serán en general exponenciales. Esto implica que, usualmente, será posible observar la dinámica sobre la variedad centro, dado que $W^s$ y $W^u$ producen una variación mucho más rápida sobre las soluciones $\Phi(x, t)$, y éstas caen a $W^c$. Veremos claramente cómo se produce esta pérdida de importancia de unas variables respecto de otras en la próxima sección, donde estudiaremos el llamado principio de control ("slaving principle").

Con los teoremas anteriores hemos demostrado que, en el caso de tener valores propios con parte real no nula, podemos estudiar y determinar la estabilidad del sistema no lineal utilizando su linearización correspondiente alrededor del punto fijo. Los ejemplos siguientes ilustran la forma general en que debemos proceder.

**Ejemplo 1**

Estudiaremos brevemente un sistema dinámico bastante conocido, el *Brusselator*. Este sistema proviene de un modelo de reacción química que, para ciertos valores de los parámetros, puede producir oscilaciones en la concentración de los componentes (cíclos límite). No llegaremos al cálculo de estos ciclos límite, que serán introducidos en el capítulo sobre atractores periódicos, pero clasificaremos la estabilidad de sus puntos fijos ateniendo al criterio enunciado en la sección 2.2.2.

La reacción química que da lugar al modelo del *Brusselator* consta de dos pasos intermedios acoplados y de un paso autocatalítico, y se la trata habitualmente como reacción irreversible:

$$a \rightarrow x$$

$$b + x \rightarrow y + d$$

$$2x + y \rightarrow 3x$$

$$z \rightarrow f$$

Se supone que las concentraciones de $a$ y $b$ se pueden gobernar desde el exterior y que el sistema es espacialmente homogéneo, con lo cual las ecuaciones correspondientes a las velocidades de reacción son
\[
\frac{dx}{dt} = a - bx + x^2y - x \\
\frac{dy}{dt} = bx - x^2y
\]
con \(a, b > 0\), lo cual corresponde a las soluciones con significado físico. Cuando igualamos ambas derivadas a cero obtenemos el único punto fijo del sistema (para cada pareja de parámetros),

\[
x^* = \left( \frac{b}{a}, \frac{b}{a} \right)
\]

La estabilidad de \(x^*\) puede estudiarse a partir de la traza y el determinante de la matriz lineal en el punto:

\[
DF(x^*) = \begin{pmatrix}\frac{b-1}{a} & \frac{a^2}{a} \\ -b & -a^2 \end{pmatrix}
\]
a partir de la cual obtenemos

\[
q \equiv \det(DF) = a^2 > 0, \quad p \equiv Tr(DF) = b - 1 - a^2
\]

El determinante de la matriz lineal es positivo para cualquier combinación de parámetros, mientras que el signo de la traza y su valor permitirán una clasificación del espacio \((a, b)\) en diferentes dominios con diferente estabilidad para el punto fijo. Podemos establecer lo siguiente:

- \(p > 0\) y \(p^2 > 4q\) \(\implies b > 1 + a^2\) y \(b > a^2 + 2a + 1\), y tendremos en este dominio nodos inestables.
- \(p > 0\) y \(p^2 < 4q\) \(\implies b > 1 + a^2\) y \(b < a^2 + 2a + 1\), dominio de focos inestables.
- \(p < 0\) y \(p^2 > 4q\) \(\implies b < 1 + a^2\) y \(b < a^2 - 2a + 1\), donde encontramos nodos estables.
- \(p < 0\) y \(p^2 < 4q\) \(\implies b < 1 + a^2\) y \(b > a^2 - 2a + 1\), correspondiendo a focos estables.

Las curvas sobre las que se da la igualdad en los casos anteriores corresponden a puntos particulares, como centros, que no hallaremos en realidad, ya que resultan destruidos en este caso por los términos no lineales. La figura 2.5 muestra el espacio de parámetros del sistema con los dominios correspondientes a cada tipo de punto fijo.

**Ejemplo 2**

\[
\dot{x} = -ax + x^2 \\
\dot{y} = x - 3y \\
\dot{z} = -(x + y)
\]
que podemos escribir como \(\dot{x} = A + F(x, y, z)\), con

\[
A = \begin{pmatrix}-a & 0 & 0 \\ 1 & -3 & 0 \\ 0 & -1 & -1\end{pmatrix} \quad F(x, y, z) = \begin{pmatrix}x^2 \\ 0 \\ 0\end{pmatrix}
\]

En este caso los puntos de equilibrio corresponden a la solución de \(\dot{x} = 0\),

\[-ax + x^2 = 0\]
Figura 2.5: Espacio de parámetros del Brusselator, donde se muestra el tipo de estabilidad que presenta el punto fijo en función de los parámetros $a$ y $b$.

$$
x - 3y = 0
-(x + y) = 0
$$

que tiene como solución los puntos $x_1^* = (0, 0, 0)$ y $x_2^* = (a, a/3, -a/3)$, para $a \neq 0$, que es el parámetro del sistema. Tenemos, por tanto, dos puntos alrededor de los cuales linearizar. En general $^7$

$$
L = \begin{pmatrix}
-a + 2x & 0 & 0 \\
1 & -3 & 0 \\
0 & -1 & -1
\end{pmatrix}
$$

Para $x_1^* = (0, 0, 0)$.

$$
Lx_1^* = \begin{pmatrix}
-a \\
-3 \\
-1
\end{pmatrix}
$$

y obtenemos los valores propios de

$$
det (Lx_1^* - \lambda I) = \begin{vmatrix}
-a - \lambda & 0 & 0 \\
1 & -3 - \lambda & 0 \\
0 & -1 & -1 - \lambda
\end{vmatrix} = 0
$$

que proporciona $\lambda_1 = -a$, $\lambda_2 = -3$, $\lambda_3 = -1$.

Para $x_2^* = (a, a/3, -a/3)$

$$
Lx_2^* = \begin{pmatrix}
a & 0 & 0 \\
1 & -3 & 0 \\
0 & -1 & -1
\end{pmatrix}
$$

$^7$El cambio de coordenadas que se ha realizado anteriormente para trasladar un punto fijo arbitrario al origen es equivalente a considerar la matriz $L$ de derivadas primeras del sistema y sustituir en ella el valor de cada uno de los puntos fijos, como el lector puede fácilmente comprobar. En el caso de los sistemas no lineales, el uso de la matriz de derivadas es más simple.
y sus valores propios son $\lambda_1 = a$, $\lambda_2 = -3$, $\lambda_3 = -1$.

Observemos que depende del signo del parámetro $a$ que los puntos sean estables o inestables. Si $a > 0$, $x_1^*$ es asintóticamente estable (todos sus valores propios tienen parte real negativa) y $x_2^*$ es inestable, por existir un valor propio con parte real positiva. Si $a < 0$, la estabilidad de los puntos fijos se intercambia. En el caso $a = 0$ la estabilidad es indeterminada, y está determinada por el término de orden superior $x^2$. En este caso, sin embargo, la estabilidad se puede decidir por simple inspección. Observemos que el término $x^2$ (el único presente en la ecuación de $\dot{x}$ cuando $a = 0$) implica en cualquier caso un crecimiento de la variable $x$ ($x^2 \geq 0 \Rightarrow \frac{d}{dt} x^2 \geq 0$), así que cualquier pequeña variación alrededor del punto $x = (0,0,0)$ implica que éste escapará en la dirección del vector $v^* = (3,1,-1)$ (es el vector asociado al valor propio nulo, que determina la dirección tangente a la variedad centro en el punto fijo), inicialmente. La variedad, ahora, sólo es lineal localmente, así que la trayectoria se desviará de la recta que marca $v^*$.

**Ejemplo 3**

\[
\begin{align*}
\dot{x} &= a(e^x - 1) - \sin(y) \\
\dot{y} &= ay - z \\
\dot{z} &= \frac{a}{1 - z} - \frac{x + a}{1 + y^2}
\end{align*}
\]

El único punto fijo de este sistema es el origen de coordenadas. Para encontrar la parte lineal de este sistema, desarrollaremos las funciones que aparecen en series de Taylor alrededor del origen

\[
\begin{align*}
\sin(y) &= y - \frac{y^3}{3!} + \frac{y^5}{5!} - \ldots \\
\frac{1}{1 - z} &= 1 + z + z^2 + z^3 + \ldots \\
\frac{1}{1 + y^2} &= 1 - y^2 + y^4 - y^6 + \ldots \\
\end{align*}
\]

y sustituyendo obtenemos

\[
\begin{align*}
\dot{x} &= ax + \frac{ax^2}{2} - y + \frac{y^3}{6} + O(4) \\
\dot{y} &= ay - z \\
\dot{z} &= a(1 + z + z^2) - (x + a)(1 - y^2) + O(3)
\end{align*}
\]

El sistema a primer orden (parte lineal) queda simplemente

\[
\begin{align*}
\dot{x} &= ax - y \\
\dot{y} &= ay - z \\
\dot{z} &= az - x
\end{align*}
\]

El cálculo de los valores propios en la forma ya conocida proporciona

\[
\lambda_1 = a - 1, \quad \lambda_2 = \frac{1}{2} + a + i\sqrt{3}, \quad \lambda_3 = \frac{1}{2} + a - i\sqrt{3}
\]

Debido a que existen dos valores propios con parte imaginaria, las trayectorias siempre entrarán al (o saldrán del) punto fijo "oscilando", en forma de espiral. Un razonamiento sencillo cualitativo permite ver por qué esto es así. Pensemos que la solución del sistema de ecuaciones diferenciales presenta términos de la forma $e^{\lambda t}$, en función de los valores propios. Si $\lambda$ presenta una parte compleja, basta con recordar la notación exponencial de las funciones trigonométricas ($\cos(\theta) = \ldots$).
\((e^{i\theta} + e^{-i\theta})/2,\) por ejemplo) para relacionar los valores propios complejos con el movimiento circular. Las partes no lineales de las ecuaciones desviaran la trayectoria de la espiral perfecta, pero la topología no se verá afectada.

Clasifiquemos las diferentes situaciones que podemos encontrar en función de los valores de \(\alpha:\)

- \(\alpha < -\frac{1}{2}:\) en este caso, el origen es estable. Sólo existe \(W^s \subset \mathbb{R}^3.\) El punto fijo es un foco (por la parte compleja de los valores propios) estable.

- \(\alpha = -\frac{1}{2}:\) Es este un punto singular, en donde el punto fijo pierde su estabilidad y "emite" una órbita periódica. Se produce entonces lo que se llama una bifurcación de Hopf, que será descrita en el capítulo 4. Aparece para este valor del parámetro la variedad centro, \(W^c \subset \mathbb{R}^2,\) mientras que \(W^s \subset \mathbb{R}.\)

- \(\frac{1}{2} < \alpha < 1: \) En este caso, \(\Re(\lambda_\pm) > 0\) y tenemos \(W^u \subset \mathbb{R}^2\) y \(\Re(\lambda_1) < 0,\) y resulta \(W^s \subset \mathbb{R}.\) Existen por tanto dos direcciones inestables dadas localmente por los vectores propios asociados, y una dirección estable. Las variedades invariantes ya no son lineales, y por tanto sólo podemos pensar en planos o rectas en un entorno infinitesimal del punto fijo. Al alejarnos, los términos no lineales deforman los subespacios \(W^u.\)

- \(\alpha = 1: \) Existe aquí una nueva bifurcación, aparece otra variedad centro de dimensión 1 y tenemos también una variedad inestable de dimensión 2.

- \(\alpha > 1.\) El punto fijo es ahora inestable en todas las direcciones: \(W^u \subset \mathbb{R}^3.\)

### 2.2 El Principio de Control (Slaving Principle)

El principio de control, que describiremos a continuación, fue enunciado por Hermann Haken (1977). El principio se basa en la separación de escalas temporales que se puede realizar entre las \(n\) variables de un sistema dinámico, y permite entender porqué un sistema que aparentemente posee muchos grados de libertad puede ser descrito en ocasiones con muy pocas variables. Se recomienda que esta sección se relacione con lo anteriormente visto sobre reducción de la dinámica a la variedad centro y con las secciones del libro en las que se habla de autoorganización. De hecho, como veremos seguidamente, la autoorganización puede también ser entendida considerando las fuerzas como parte del sistema dinámico, y aplicando en el citado sistema el principio de control.

#### 2.2.1 Organización

Podemos definir organización como la forma coherente en que actúa un sistema bajo unas órdenes externas dadas. Se entiende que la finalidad de este comportamiento regulado es la producción de cierto resultado o producto.

Traducimos lo anterior al lenguaje de las matemáticas. Consideremos el caso en que el "efecto", o la acción de las "órdenes" externas pueda ser representado por una variable \(q,\) que cambiará en un intervalo de tiempo \(\Delta t\) en una cantidad proporcional a la causa \(F.\) Así que consideraremos ecuaciones de la forma

\[
\dot{q}(t) = F_0(q(t); t)
\]  
(2.3.1)

En ausencia de fuerzas externas suponemos \(q = 0,\) y requerimos que el sistema regrese a este estado cuando, una vez aplicada, la fuerza externa desaparezca. La ecuación más sencilla que cumple ambas exigencias es

\[
\dot{q} = \gamma q
\]
con $\gamma > 0$. Si se añade una fuerza externa, obtenemos

$$\dot{q} = \gamma q + F(t)$$

La solución formal de esta ecuación se puede escribir como

$$q(t) = \int_0^t e^{-\gamma(t-\tau)} F(\tau) d\tau$$

(2.3.2)

despreciando los transitorios. En principio, la respuesta del sistema en el instante $t$ depende de toda la historia anterior (obsérvese que la integral se realiza sobre todos los valores de $\tau$ hasta el instante $t$ considerado, y cuando $t \to \infty$ el límite superior de la integral diverge). Consideraremos en lo sucesivo únicamente sistemas que presenten una respuesta instantánea a la acción de la fuerza. Para aclarar el significado de esta suposición, consideremos un caso particular de fuerza.

$$F(t) = ae^{-bt}$$

que permite la realización inmediata de la integral 2.3.2, con el resultado

$$q(t) = \frac{a}{\gamma - \delta}(e^{-\delta t} - e^{-\gamma t})$$

Podemos ahora expresar cuantitativamente la condición de que $q$ responda instantáneamente a la fuerza:

$$\gamma >> \delta$$

caso en el cual

$$q(t) \approx \frac{a}{\gamma} e^{-\delta t} = \frac{1}{\gamma} F(t)$$

Es decir, la constante de tiempo del sistema, $t_0 = 1/\gamma$, debe ser mucho más corta que la constante de tiempo que caracteriza a la fuerza, a las “órdenes”, $t' = 1/\delta$. Esta suposición se denomina aproximación adiabática.

Observemos que habríamos llegado al mismo resultado de haber supuesto desde el comienzo $\dot{q} = 0$, con lo cual habríamos obtenido

$$0 = -\gamma q + F(t)$$

Estos resultados son inmediatamente generalizables a un sistema de $n$ ecuaciones, en donde la suposición de respuesta instantánea significaría considerar $q = 0$, para todas las variables.

### 2.2.2 Autoorganización

Podemos definir autoorganización como el trabajo conjunto y coherente realizado por un sistema en ausencia de órdenes externas, simplemente como resultado del entendimiento mutuo entre las partes que lo forman.

Debe resultar evidente que las fuerzas que actúan sobre un sistema no son nunca fuerzas externas e inamovibles. Cuando tratamos con sistemas abiertos (como son la mayoría de los sistemas naturales, fuera del laboratorio) necesitamos introducir las fuerzas como constituyentes dinámicos del sistema: son las fuerzas las responsables de las acciones del sistema, pero pueden verse modificadas por la respuesta que el sistema presente. Pongamos un ejemplo mecánico. Imaginemos una gran montaña (el sistema) sobre la que comienza a incidir un viento fuerte (la fuerza) de, digamos,
80 km/h. El viento es el responsable de la lenta erosión de la montaña, para acabar convertida en llano (en ausencia de nuevas elevaciones geológicas del terreno). Por otra parte, el cambio del paisaje tiene una influencia clara en la velocidad que este viento tomará en los distintos puntos del relieve. Las montañas apantallan el viento, los llanos no. Por tanto, sistema y fuerza no son dos entidades independientes, sino partes de la descripción global del supersistema que constituyen.

Llamemos en vista de lo anterior \( q_1 \) a la fuerza \( F \) anterior, y \( q_2 \) a la variable \( q \) original, y consideremos el siguiente ejemplo,

\[
\dot{q}_1 = \gamma_1 q_1 - a q_1 q_2 \quad (2.3.3)
\]
\[
\dot{q}_2 = \gamma_2 q_2 + b q_1^2 \quad (2.3.4)
\]

Supondremos de nuevo que \( q_2 \) tenderá a 0 en ausencia de \( q_1 \), con lo cual \( \gamma_2 = 0 \). Supongamos que, por analogía con el caso anterior,

\[ \gamma_2 \gg \gamma_1 \]

lo cual permite la resolución aproximada del sistema 2.3.4, utilizando \( \dot{q}_2 = 0 \) \(^8\), para obtener

\[
\dot{q}_2(t) \approx \gamma_2^{-1} b q_1^2(t) \quad (2.3.5)
\]

Se dice, en vista de la respuesta inmediata del sistema \( q_2 \) a la acción de \( q_1 \), que \( q_2 \) controla la dinámica. Sin embargo, el sistema controlado 2.3.4 también actúa sobre 2.3.3. Si sustituimos 2.3.5 en 2.3.3 obtenemos

\[
\dot{q}_1 = -\gamma_1 q_1 - \frac{ab}{\gamma_2} q_1^2
\]

Este sistema se verá con detalle en el capítulo sobre atractores periódicos, ya que las ecuaciones dinámicas de este tipo presentan dos tipos de solución a largo plazo, dependiendo del signo de \( \gamma_1 \): si \( \gamma_1 > 0 \), \( q_1 = 0 \), y por tanto también \( q_2 = 0 \), así que no se produce ninguna respuesta en el sistema; si \( \gamma_1 < 0 \), entonces existe una solución para el estado estacionario no nula,

\[
q_1 = \pm \left( \frac{b}{a} \frac{\gamma_1}{\gamma_2} \right)^{1/2}
\]

y por tanto también \( q_2 \neq 0 \). Este es un caso particular de lo que genéricamente se llama rotura de simetría.

En el capítulo 7, cuando tratemos con los fenómenos críticos, se entenderá la denominación parámetro de orden que se da a la variable \( q_1 \), la cual, a riesgo de resultar redundante, cuantifica el grado de orden existente en el sistema \(^9\).

En general, podemos denominar a las variables que controlan el sistema parámetros de orden, si la dinámica en otros subsistemas está condicionada a la que presenten estos parámetros. Generalizemos las ideas anteriores a sistemas con \( n \) variables. Supongamos que tenemos un sistema de la forma

\[
\dot{q}_i = -\gamma_i q_i + g_i(q_1, \ldots, q_n) \quad (2.3.6)
\]

\(^8\) Nótese que es necesario suponer que las variables \( q_i \) permanecen acotadas para realizar esta aproximación sin problemas.

\(^9\) Compárese con la transición crítica que se produce en los sistemas magnéticos de para- a ferromagnéticos. La variable \( q_1 \) corresponde directamente a la magnetización \( m \).
con $i = 1, \ldots, n$, y que podemos realizar dos subgrupos. En uno de ellos, en el que $i = 1, \ldots, m$, colocaremos los "modos lentos" (que serán los parámetros de orden del sistema), es decir, aquellos en los que $\gamma_i < 0$ pero de valor absoluto pequeño, y en el otro, con $s = m + 1, \ldots, n$, los "modos rápidos". Si utilizamos la aproximación adiabática podemos suponer que $\dot{q}_s = 0$ para $s = m + 1, \ldots, n$, y supondremos que, debido a la separación que el valor de $\gamma$ ha permitido, también $|q_s| >> |q_i|$, aunque esta suposición debe de ser comprobada en cada caso. Podremos en consecuencia solucionar el sistema 2.3.6 para $q_s$, $s = m + 1, \ldots, n$, y por tanto la solución del sistema

$$\dot{q}_i = -\gamma_i q_i + g_i(q_1, \ldots, q_m; q_{m+1}(q_i), \ldots, q_n(q_i))$$ (2.3.7)

establecerá la posibilidad de un "resultado" nulo en el sistema.

El resultado anterior es generalizable a sistemas en los que, en principio, no sea posible realizar la separación en modos lentos y modos rápidos que se ha supuesto inicialmente. Es suficiente con que exista una ordenación jerárquica del tipo

$$\gamma^{(1)} >> \gamma^{(2)} >> \gamma^{(3)} >> \ldots$$

para realizar una eliminación adiabática de variables una por una, comenzando por $\gamma^{(1)}$.

La consecuencia más importante de todo lo anterior es que, si los modos rápidos caen bajo el control de los modos lentos adiabáticamente, el comportamiento global del sistema estará controlado por la dinámica de unos pocos parámetros de orden. Por tanto, incluso sistemas dinámicos con un gran número de variables pueden presentar un comportamiento coherente y bien regulado. Además, el hecho de que puedan darse roturas de simetría en el sistema implica que éste puede operar en diversos estados, bien definidos por el comportamiento de los parámetros de orden.
2.3 Funciones de Lyapunov

Hemos visto un cierto estudio que podemos realizar sobre un sistema de EDOs para determinar la estabilidad de sus puntos fijos. También hemos podido apreciar cómo aparecen ciertas dificultades cuando tenemos valores propios con parte real nula. Describiremos en esta sección una aproximación muy útil, debida a Lyapunov, que permite considerar ciertas funciones para garantizar la estabilidad de los puntos de equilibrio, y que en algunos casos puede resultar más sencilla de utilizar que toda la álgebra implicada en el tratamiento de la matriz lineal o de los términos no lineales.

Definición

Si \( F \in C^1(E) \), \( V \in C^1(E) \) y \( \Phi_t \) es el flujo de la ecuación diferencial 2.2.3, entonces para \( x \in E \) la derivada de una función \( V(x) \) sobre las soluciones \( \Phi_t \) es

\[
\dot{V}(x) = \frac{d}{dt} V(\Phi_t(x))|_{t=0} = D V(x) F(x)
\]

La última igualdad resulta de la aplicación de la regla de la cadena. Si \( \dot{V}(x) \) es negativa en \( E \), entonces \( V(x) \) decrece sobre la solución \( \Phi_t(x_0) \) a través de \( x_0 \in E \) en \( t = 0 \). Además, en \( \mathbb{R}^2 \), si \( \dot{V}(x) \leq 0 \), con la igualdad sólo en \( x = 0 \), entonces, para una constante \( C > 0 \), la familia de curvas \( V(x) = C \) constituye una familia de curvas cerradas que rodean el origen y las trayectorias de 2.2.3 cruzan estas curvas del exterior al interior con \( t \) creciente; es decir, el origen es un punto asintóticamente estable.

Una función \( V : \mathbb{R}^n \to \mathbb{R}^n \) que satisfaga las condiciones del teorema siguiente se denomina función de Lyapunov.

Teorema

Sea \( E \) un subconjunto abierto de \( \mathbb{R}^n \) que contenga \( x_0 \). Supongamos que \( F \in C^1(E) \) y que \( F(x_0, \mu) = 0 \). Supongamos también que existe una función \( V \in C^1(E) \) que satisface \( V(x_0) = 0 \) y \( V(x) > 0 \) si \( x \neq x_0 \). Entonces

- Si \( \dot{V}(x) \leq 0 \), \( \forall x \in E \), \( x_0 \) es estable.
- Si \( \dot{V}(x) < 0 \), \( \forall x \in E - \{x_0\} \), \( x_0 \) es asintóticamente estable.
- Si \( \dot{V}(x) > 0 \), \( \forall x \in E - \{x_0\} \), \( x_0 \) es inestable.

La existencia de una función de Lyapunov es especialmente relevante en muchos casos reales, ya que, de hecho, define un potencial para la evolución del sistema.

Si bien no es necesario resolver el sistema de ecuaciones diferenciales para aplicar el teorema de Lyapunov, debe tenerse presente que no existe un método general para hallar \( V \). El el caso de muchos sistemas físicos, \( V \) será la energía potencial del sistema. Veamos algunos ejemplos.

Ejemplo 1

Consideremos el sistema dinámico definido por

\[
\frac{dx}{dt} = 2y(z - 1)
\]
\[ \frac{dy}{dt} = -z(z - 1) \]
\[ \frac{dz}{dt} = -z^3 \]

El único punto de equilibrio es el origen. Intentemos determinar su estabilidad como habíamos aprendido en la sección anterior:

\[ L(x_0) = \begin{pmatrix} 0 & -2 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \]

que proporciona

\[ \lambda_1 = 0, \quad \lambda_2 = +i\sqrt{2}, \quad \lambda_3 = -i\sqrt{2} \]

Los tres valores propios tienen parte real nula, así que no podemos decir nada sobre la estabilidad del punto fijo sin considerar los términos no lineales. Intentemos hallar una función de Lyapunov para este sistema. Lo habitual es suponer una cierta función con términos de orden 2 y coeficientes indeterminados. Los términos de orden 2 (o, en el caso de no ser estos suficiente, los de cualquier orden par), aseguran \( V(x) > 0 \) en un entorno de cualquier punto fijo real. Los coeficientes se determinan utilizando el flujo del sistema e intentando que se cumpla \( \dot{V}(x) \leq 0 \) en el dominio estudiado. Probemos en este caso la función

\[ V(x, y, z) = ax^2 + by^2 + cz^2 \]

con \( a, b, c > 0 \). Se cumple

\[ \dot{V}(x, y, z) = 2ax\dot{x} + 2by\dot{y} + 2cz\dot{z} = 4axy(z - 1) - 2byz(z - 1) - 2cz^4 \]

donde se ha utilizado el valor de las derivadas que proporciona el flujo de cada una de las ecuaciones diferenciales. Observe que es posible cancelar los dos primeros términos si escogemos \( 2a = b \) (los únicos que podrían ser positivos), y tomando \( c > 0 \) (como ya se había supuesto) aseguramos \( \dot{V}(x) \leq 0 \) en todo el dominio. Para el caso particular \( a = 1, b = 2, c = 1 \), obtenemos

\[ V(x, y, z) = x^2 + 2y^2 + z^2 \]
\[ \dot{V}(x, y, z) = -z^4 \]

sobre el flujo del sistema, con lo cual se cumplen las condiciones del teorema y podemos asegurar que el punto fijo \( x_0 = (0, 0, 0) \) es asintóticamente estable, en este caso en todo el dominio \( \mathbb{R}^3 \), dado que no hay ningún otro punto fijo.

Ejemplo 2

Consideremos la ecuación diferencial de segundo orden \( \ddot{x} + p(x) = 0 \), donde la función continua \( p(x) \) satisface \( xp(x) > 0 \) para \( x \neq 0 \). Obsérvese que esta es la ecuación dinámica de un cuerpo sometido a una fuerza total dependiente de la posición (fuerza \( \equiv F(x) \equiv -p(x) \)) según viene dada por la segunda ley de Newton. Escribamos el sistema en la forma

\[ \dot{x}_1 = x_2 \]

\[ \dot{x}_2 = -x_1 \]

\[ ^{10} \text{Con el cambio de variables momento cinético } \equiv z = mv. \]


\[ \dot{x}_2 = -p(x_1) \]

donde \( x_1 \equiv x \). La energía total del sistema es

\[ V(x) = \frac{x_1^2}{2} + \int_0^{x_1} p(s)ds \]

simplemente la suma de la parte cinética, \( \frac{1}{2} \dot{x}_1^2 \), y la parte potencial (suponiendo que estamos tratando con una fuerza conservativa, que por tanto deriva de una cierta función potencial). La energía total del sistema proporciona en este caso directamente una función de Lyapunov. Observemos

\[ \dot{V}(x) = p(x_1)x_2 + x_2(-p(x_1)) \equiv 0 \]

Las curvas solución están dadas por \( V(x) = c \) (significa que la energía es constante sobre las trayectorias del sistema) y el origen es un punto de equilibrio estable.

Una aplicación directa la podría proporcionar la función \( p(x) = \frac{L}{m} \sin(x) \), para \( x \in (0, \pi) \). Se obtiene integrando sin dificultad

\[ V(x) = \frac{x_1^2}{2} + \frac{f}{m}(\cos(x_1) - 1) \]

\[ \dot{V} = x_2 \frac{f}{m} \sin(x_1) - \frac{f}{m} \sin(x_1)x_2 = 0. \]

### 2.4 Sistemas gradiente

Se denomina *sistema gradiente* en un conjunto abierto \( U \in \mathbb{R}^2 \) a un sistema dinámico definido en la forma

\[ \frac{dx}{dt} = -\nabla V_\mu x \]

siendo \( V_\mu : U \rightarrow \mathbb{R} \) una función \( C^2 \) y \( \nabla \) el operador gradiente habitual, definido por

\[ \nabla \equiv (\partial_{x_1} V_\mu, \ldots, \partial_{x_n} V_\mu) \]

que da el campo vectorial \( \nabla V_\mu : \rightarrow \mathbb{R}^n \) asociado a \( V_\mu \). Estos sistemas gradiente poseen propiedades especiales que hacen sus flujos particularmente simples. Daremos a continuación unos teoremas y dos ejemplos ilustrativos.

**Teorema**

Sea \( \partial_\mu V_\mu(x) \leq 0 \), \( \forall x \in U \), y supongamos que \( V_\mu(x_0) = 0 \). Entonces, \( x_0 \) es un punto de equilibrio del sistema dinámico.

**Corolario**

Sea \( x_0 \) un mínimo aislado de \( V_\mu \). Entonces, \( x_0 \) es un punto de equilibrio asintóticamente estable.

**Teorema**

En los puntos regulares, esto es, los \( x \) tales que \( V_\mu(x) \neq 0 \), el campo vectorial definido por \( -\nabla V_\mu \) es perpendicular a las superficies de nivel de \( V_\mu \) del sistema dinámico.
Teorema

Dado un sistema dinámico gradiente, en los puntos regulares las trayectorias cruzan las superficies de nivel ortogonalmente. Los puntos no regulares son de equilibrio.

Ejemplo 1

Consideremos el sistema dinámico definido por

\[
\frac{dx}{dt} = -2x(x-1)(2x-1) \\
\frac{dy}{dt} = -2y
\]

Esto puede ser reescrito en la forma de un sistema gradiente, esto es, \( \dot{x} = -\nabla V \), si tomamos como función potencial

\[ V(x, y) = x^2(x-1)^2 + y^2 \]

con lo que el gradiente viene dado por

\[
-\nabla V = (-\partial_x V, -\partial_y V) = (-2x(x-1)(2x-1), \ -2y)
\]

Los puntos de equilibrio obtenidos de igualar a cero las ecuaciones de partida son \( x_1^* = (0, 0) \), \( x_2^* = (1/2, 0) \), \( x_3^* = (1, 0) \). Tenemos entonces

\[
DF = \begin{pmatrix}
-\partial_x \{-2x(x-1)(2x-1)\} & 0 \\
0 & \partial_y \{-2y\}
\end{pmatrix}
\]

que da como resultado

\[
DF = \begin{pmatrix}
-2(6x^2 - 6x + 1) & 0 \\
0 & -2
\end{pmatrix}
\]

Un estudio de estos valores revela que los puntos \( x_1^* \) y \( x_3^* \) son puntos de equilibrio estable (también llamados en ocasiones sumideros) y \( x_2^* \) es un punto de silla, y por tanto inestable.

Ejemplo 2

Consideraremos ahora un modelo de sistema neural compuesto por una población de neuronas activadoras e inhibidoras. Si bien las neuronas son elementos discretos y en general altamente conectados entre sí, la aproximación continua es adecuada en ciertos casos. Las ecuaciones que damos a continuación suponen conectividad total, y el modelo queda definido por

\[
\frac{du_1}{dt} = -u_1 + 2\epsilon_{11}\Phi(u_1) - 2\epsilon_{12}\Phi(u_2) + w_1
\]

\[
\frac{du_2}{dt} = -u_2 + 2\epsilon_{21}\Phi(u_1) - 2\epsilon_{22}\Phi(u_2) + w_2
\]

donde \( \Phi(x) \) es la función que representa la interacción entre neuronas, y se define en este modelo por

\[ \Phi(x) = (1 + e^{-\langle x - \theta \rangle})^{-1} \]

El sistema lineal correspondiente está dado por la matriz L,
\[ L = \begin{pmatrix} -1 + 2\varepsilon_{11} & -2\varepsilon_{12} \\ 2\varepsilon_{21} & -(1 + \varepsilon_{22}) \end{pmatrix} \]

Puede comprobarse que \((u_1, u_2) = (0, 0)\) será un punto de silla si se verifica la condición
\[
\frac{\varepsilon_{12}\varepsilon_{21}}{\varepsilon_{22} + 2} < \varepsilon_{11} - 2
\]
y será un nodo o un foco en el caso contrario. Las ecuaciones del modelo admiten la siguiente representación:
\[
\begin{align*}
\partial_t u_1 &= -\nabla_{u_1} V_1(u_1, u_2) - \nabla_{u_2} V_2(u_1, u_2) \\
\partial_t u_2 &= -\nabla_{u_1} V_1(u_1, u_2) - \nabla_{u_1} V_2(u_1, u_2)
\end{align*}
\]
siendo \(V_1\) y \(V_2\) las funciones potencial definidas por
\[
V_1 = \frac{1}{2} u_1^2 - 2\varepsilon_{11} \left[u_1 + \ln \frac{1}{2} (1 + e^{-u_1})\right] - w_1 u_1 +
+ \frac{1}{2} u_2^2 - 2\varepsilon_{22} \left[u_2 + \ln \frac{1}{2} (1 + e^{-u_2})\right] - w_2 u_2
\]
\[
V_2 = 2\varepsilon_{21} \left[u_1 + \ln \frac{1}{2} (1 + e^{-u_1})\right] + 2\varepsilon_{12} \left[u_2 + \ln \frac{1}{2} (1 + e^{-u_2})\right]
\]
Para el caso especial \(\varepsilon_{12} = \varepsilon_{21} = 0\) se tiene que el modelo se define así
\[
\begin{align*}
\frac{du_1}{dt} &= -\nabla_{u_1} V_2 \\
\frac{du_2}{dt} &= \nabla_{u_1} V_2
\end{align*}
\]
lo cual define un sistema conservativo con hamiltoniano (efectivo) \(V_2\).

### 2.5 Sistemas discretos

Consideraremos en esta sección sistemas discretos unidimensionales, definidos mediante ecuaciones de recurrencia del tipo
\[
x_{n+1} = F_\mu(x_n)
\]
donde \(x_n\) es por ejemplo la población de una especie en la generación \(n\)-ésima, habitualmente normalizada entre 0 y 1, utilizando
\[
x_n = \frac{n_0 \text{ de individuos}}{n_0 \text{ máximo posible}}
\]
y \(\mu\) indica un parámetro o conjunto de parámetros. Ejemplos de este tipo de sistema serían
\[
\begin{align*}
x_{n+1} &= \mu x_n (1 - x_n) \\
x_{n+1} &= r_n e^{\mu (1 - x_n / K)} \\
x_{n+1} &= -\alpha x_n^2 + \beta z_n^2
\end{align*}
\]
Figura 2.7: Diagrama de bifurcación (véase el capítulo sobre caos) correspondiente a la aplicación discreta 2-dimensional 2.6.6, con $\mu = \beta$.

Los valores sucesivos se obtienen del anterior por iteración de la ecuación $F_\mu(x)$. Formalicemos las ideas básicas referentes a las aplicaciones discretas.

**Definición**

Sea la aplicación

$$x_{n+1} = F_\mu(x) \tag{2.6.4}$$

con $x \in U \subset \mathbb{R}$, $n \in \mathbb{Z}$ y $\mu \in \mathbb{R}$, donde $U$ y $V$ son conjuntos abiertos. Diremos que 2.6.4 es un sistema dinámico discreto.

La definición anterior se generaliza trivialmente para sistemas de dimensión $N$, para los que tendríamos la ecuación vectorial

$$\mathbf{x}_{n+1} = F_\mu(\mathbf{x}) \tag{2.6.5}$$

siendo $\mathbf{x}_n = (x_1, x_2, \ldots, x_N)$. Un ejemplo de este caso lo proporcionan, para $N = 2$, las ecuaciones de Lotka-Volterra:

$$x_{n+1} = \mu x_n (1 - x_n - y_n)$$

$$y_{n+1} = \beta x_n y_n \tag{2.6.6}$$

donde $x_n$ sería la población de presas en la $n$-ésima generación y $y_n$ la correspondiente a los depredadores. Un modelo de este tipo será analizado más adelante.

Estudiaremos en primer lugar sistemas de una dimensión para generalizar posteriormente a $N$ dimensiones.

El primer paso en el estudio de un sistema como 2.6.4 es hallar sus puntos de equilibrio, tal y como se ha hecho con los sistemas continuos. Consideremos el ejemplo 2.6.2. Si imponemos la condición $x_{n+1} = x_n = x^*$, podemos determinar los valores de la población que no varían si son alcanzados. Obtenemos
Figura 2.8: (a) Equilibrio inestable. (b) Equilibrio indiferente. (c) Equilibrio estable.

\[ x^* = x^* \exp \left( \mu \left( 1 - \frac{x^*}{K} \right) \right) \Rightarrow x_1^* = 0, \quad x_2^* = K \]

Tenemos dos puntos de equilibrio. Si el sistema se encuentra exactamente en uno de estos estados, permanecerá en él indefinidamente y sin alteración. Sin embargo, la estabilidad de cada punto puede ser distinta. Para entender esta afirmación, observemos la figura 2.8, en la que representamos tres situaciones que implican puntos críticos. Evidentemente, este sujeto mecánico también se aplica a los puntos de equilibrio de los sistemas dinámicos continuos. En cada caso, la bola permanecerá en su posición si no actúa ninguna perturbación. Físicamente, sabemos que las fluctuaciones siempre están presentes, y vemos claramente que el resultado de desplazar la bola una distancia muy pequeña respecto de la posición de equilibrio tiene efectos totalmente distintos en cada caso.

Un punto de equilibrio será estable o inestable si esta desviación respecto de \( x^* \) disminuye o aumenta con el tiempo, respectivamente. Para analizar este punto, consideremos también en estos sistemas discretos un desarrollo de la función \( F_\mu(x) \) cerca del punto crítico, es decir, \( x_n = x^* + y_n \) \(^{11}\), siendo \( y_n \) una pequeña perturbación de \( x^* \), estudiaremos como crece ésta a medida que el tiempo transcurre. Tendremos

\[ y_{n+1} = x_{n+1} - x^* = F_\mu(x^*) + \left. \frac{\partial F_\mu(x)}{\partial x} \right|_{x^*} (x_n - x^*) \]

es decir

\[ y_{n+1} = \left. \frac{\partial F_\mu(x)}{\partial x} \right|_{x^*} y_n \]

que nos ha conducido a una ecuación lineal para el comportamiento de las perturbaciones cerca del punto de equilibrio:

\[ y_{n+1} = \left. \frac{\partial F_\mu(x)}{\partial x} \right|_{x^*} y_n \]

\(^{11}\) El hecho de que el sistema tenga una dinámica discreta no significa que no podamos tomar dos condiciones iniciales tan próximas como desemos, y de valor real. A partir de esta condición inicial, los puntos por los que pasa el sistema son discretos y determinados.
Esta iteración no es más que una progresión geométrica, donde la derivada $x^*$ hace de razón. De esta expresión obtenemos inmediatamente la condición que define la estabilidad. Si definimos

$$\lambda = \left| \frac{\partial F_\mu(x)}{\partial x} \right|_{x^*}$$

entonces tendremos

1. Estabilidad: $|\lambda| < 1$
2. Inestabilidad: $|\lambda| > 1$
3. Estabilidad marginal: $|\lambda| = 1$

La condición 3 establece las fronteras entre dominios de estabilidad e inestabilidad. En general, esta frontera será una función de los parámetros del sistema \(^{12}\).

Para el ejemplo 2.6.2, tendremos:

$$\frac{\partial F_\mu(x)}{\partial x} = \frac{\partial}{\partial x} \left\{ x e^{-\mu(1-x/K)} \right\} = \left( 1 - \frac{\mu x}{K} \right) e^{\mu(1-\frac{x}{K})}$$

y la condición de equilibrio para nuestro sistema en relación al punto $x^* = K$ resulta

$$\left| \frac{\partial F_\mu(x)}{\partial x} \right|_{x^*} = |1 - \mu| < 1$$

que define un dominio de estabilidad dado por $\mu \in (0, 2)$. Notemos que la forma en que nos aproximamos al punto de equilibrio depende del signo de la derivada. Si

$$0 < \left| \frac{\partial F_\mu(x)}{\partial x} \right|_{x^*} < 1$$

tenemos una aproximación monótona hacia el punto de equilibrio. Si su valor es negativo, es decir, si

$$-1 < \left| \frac{\partial F_\mu(x)}{\partial x} \right|_{x^*} < 0$$

la aproximación se da a través de oscilaciones de tamaño cada vez menor (véase la figura 2.2). Es importante señalar que este último tipo de dinámica no se presenta en ningún caso cuando tratamos con sistemas continuos.

**Ejemplo**

Analizaremos a continuación un ejemplo paradigmático de aplicación discreta: la aplicación logística (R. May, 1979). Esta aplicación está definida por

$$x_{n+1} = \mu x_n (1 - x_n)$$

con $\mu \in [0, 4]$ si se pretende mantener el valor de $x_n$ acotado entre 0 y 1. Esta ecuación es el sistema unidimensional no lineal más simple, y su importancia (como veremos) es fundamental.

---

\(^{12}\)Obsérvese que se exige, en el caso discreto, que $\left| \frac{\partial F_\mu(x^*)}{\partial x^*} \right| < 1$, mientras que en el caso continuo se exigía que esta misma derivada fuese menor que 0. Pensemos que, ahora, simplemente multiplicamos el valor de $x_n$ por el valor de $\lambda$ (que antes aparecía en una expresión exponencial). Así que, si $|\lambda| < 1$, los valores se aproximarán sucesivamente y de forma discreta al punto fijo.
El comportamiento cualitativo de 2.6.1 respecto del parámetro \( \mu \) tiene las mismas propiedades para todas las aplicaciones que presenten un único máximo en el intervalo unidad. Busquemos los puntos fijos de 2.6.1. Verifican:

\[
x^* = \mu x^*(1 - x^*)
\]

que nos proporciona una sencilla ecuación de segundo orden. Sus raíces son

\[
x_1^* = 0, \quad x_2^* = 1 - \frac{1}{\mu}
\]

para las que analizaremos la estabilidad de las trayectorias. Tendremos, por tanto:

\[
\lambda(x_1^*) = \left( \frac{\partial F(x_n)}{\partial x_n} \right)_{x=x_1^*} = \mu
\]

\[
\lambda(x_2^*) = \left( \frac{\partial F(x_n)}{\partial x_n} \right)_{x=x_2^*} = -\mu + 2
\]

El origen será estable para \( \lambda < 1 \) e inestable en cualquier otro caso, mientras que para el segundo punto fijo, obtendremos estabilidad para \( \lambda < 3 \) y se inestabilizará cuando se cruce el valor \( \lambda = 3 \). En este punto se produce una bifurcación, y la aplicación deja de tener puntos estables de período 1, y aparece una órbita estable de período 2: existe una alternancia entre dos puntos, que son puntos fijos de la aplicación doblemente iterada.

La aplicación logística posee propiedades sorprendentes, entre ellas la alta complejidad que pueden generar sus evoluciones, cuando los valores del parámetro \( \mu \) son lo bastante elevados. El análisis de esta aplicación concreta se continuará en el capítulo sobre caos.

Ejemplo

Analizaremos a continuación una aplicación discreta en dos dimensiones de Lotka-Volterra que es adecuada para modelizar una interacción del tipo presa-depredador:

\[
x_{n+1} = \mu x_n (1 - x_n - y_n)
\]

\[
y_{n+1} = \beta x_n y_n
\]

donde \( x_n \) e \( y_n \) son las poblaciones normalizadas de las presas y de los depredadores, respectivamente. Para cada una de las poblaciones, \( \mu \) y \( \beta \) son los ritmos de crecimiento. Analizando primeramente los puntos fijos obtenemos que el sistema presenta dos valores estacionarios,

\[
x_1^* = (0, 0), \quad x_2^* = \left( \frac{1}{\beta}, 1 - \frac{(\beta + \mu)}{\beta \mu} \right)
\]

El estudio de la matriz de derivadas parciales en cada uno de los puntos proporcionará de nuevo el tipo de estabilidad correspondiente:

\[
DF = \begin{pmatrix}
\mu(1 - 2x - y) & -\mu x \\
\beta y & \beta z
\end{pmatrix}
\]

Considerremos por simplicidad el caso particular \( \beta = \mu \), con lo cual obtenemos para el punto fijo no trivial \( x_2^* \):

\footnote{Vease el capítulo 4.}
Figura 2.9: Aspecto del espacio de fases del modelo de Lotka-Volterra en dos dimensiones para diversos valores del parámetro $\mu$. De izquierda a derecha y de arriba a abajo, $\mu = 3.41, 3.42, 3.43, 3.45$. La situación dinámica del sistema en cada caso se deduce del diagrama de bifurcación correspondiente (figura 2.7).
\[ \mathbf{DF}(x_3^*) = \begin{pmatrix} 0 & -1 \\ \mu - 2 & 1 \end{pmatrix} \]

que tiene asociada la ecuación de valores propios

\[ \det(\mathbf{DF}(x_3^*)) = \det \begin{pmatrix} -\lambda & -1 \\ \mu - 2 & 1 - \lambda \end{pmatrix} = 0 \]

El punto fijo será estable si el módulo de ambos valores propios es menor que 1. Esto conduce al dominio de estabilidad

\[ S(x_3^*) = \{ \mu : \mu \in (2, 3) \} \]

Cuando se aumenta el valor de \( \mu \) más allá de 3 se obtiene un escenario de bifurcación \(^{14}\) donde el caos aparece a partir de \( \mu \approx 3.43 \).

**Bibliografía**


\(^{14}\) Véase el capítulo sobre caos.
Capítulo 3
Fractales

Introducimos en este capítulo la geometría de los objetos fractales, la cual de algún modo es la geometría de los sistemas complejos, nuestro objeto de estudio.

Los fractales, con el nombre de curvas no derivables, o no rectificables, aparecieron en las matemáticas hacia finales del siglo XIX. En un principio, los fractales fueron ejemplo de objetos curiosos. Eran curvas o superficies infinitamente plegadas, líneas infinitas compactificadas de forma regular en una superficie finita, superficies no derivables en ningún punto, conjuntos de puntos aislados isomoros a la recta real, ejemplos en resumen de objetos no rectificables. Los primeros nombres relacionados con esta disciplina (insistimos, aún no se llamaban fractales) son probablemente de todos conocidos: Cantor, Peano, Hilbert, Hausdorff, Sierpinski, von Koch. Y aún muchos otros que no estuvieron directamente involucrados, pero que sentaron las bases de la teoría de la medida necesaria para realizar una descripción matemáticamente correcta de los objetos de que hablamos: Lebesgue, Poincaré, Menger y algunos de los anteriormente citados, entre otros.

La difusión que actualmente han tenido los objetos fractales está claramente motivada por la estética que con ellos se relaciona. No fue hasta que aparecieron los ordenadores que objetos maravillosos como los conjuntos de Gaston Julia pudieron ser visualizados.

El nombre al que actualmente va ligado cualquier objeto fractal es el de Benoit B. Mandelbrot. quien les dio el nombre actual. fractales (del latín fractus, interrumpido o irregular) y popularizó la nueva geometría. Mandelbrot vio la naturaleza a través de la geometría fractal. Como él mismo dijo

"... las montañas no son conos, las nubes no son esferas, ni la corteza de los árboles es lisa."

Son innumerables los ejemplos de objetos fractales que se han encontrado en todas las ramas del conocimiento, y no únicamente en las relacionadas con la física, las matemáticas o la biología. El mundo microscópico, el orgánico pero también el inorgánico, está repleto de objetos fractales. Las fracturas, por ejemplo, presentan esta propiedad. La superficie de las células, la estructura de nuestros pulmones o del aparato circulatorio, las formaciones de nubes, las montañas, quizá la distribución de materia en la galaxia, las fluctuaciones en la intensidad de radiación de un quasar, los árboles, los líquenes, los relámpagos, ... son parte de la inacabable serie.

Se ha dirigido la atención incluso hacia la forma de crecimiento de las ciudades. Existe actualmente amplio acuerdo en que el crecimiento urbano no se produce de forma que el espacio disponible se llene de forma compacta. Dado que la dimensión vertical de las ciudades es despreciable frente a las dos dimensiones horizontales, se puede caracterizar la geometría de los asentamientos urbanos...
con una dimensión que varía entre 1 y 2. Algunos cálculos realizados en más de 30 ciudades asignan una dimensión entre 1.6 y 1.8 a la mayoría de ellas. Un claro reflejo de la fractalidad de la ciudad se puede hallar en la autosimilaridad de la red de transportes urbanos (M. Batty y P. Longley, 1994).

En las páginas siguientes, intentaremos describir esta nueva geometría de la naturaleza.

### 3.1 Caracterización de los objetos fractales.

Fractal es todo objeto que posea autosimilaridad. Esto es, a todas las escalas a las que podamos contemplarlo, descubrimos que la parte es semejante al todo. En consecuencia, una primera imagen intuitiva corresponde a un objeto infinitamente doblado sobre sí mismo, con infinitos pliegues, con infinita estructura. Con una cierta práctica es fácil reconocer a simple vista los objetos fractales. Pero también necesitamos una descripción cuantitativa, un valor numérico que pueda ser asociado a cada uno de ellos, que describa su estructura y que permita diferenciarlos.

Esta caracterización la proporciona, en primer lugar, la llamada dimensión fractal del conjunto. Introduciremos este concepto con la ayuda de un caso particular: la curva de Helge von Koch.

La matemática sueca Helge von Koch fue quien, a principios de este siglo, estudió la geometría de los conjuntos actualmente llamados fractales. Su curva apareció como contrapartida a las curvas rectificables, “suaves”, que tan importantes son en muchas ramas de la matemática. La curva de von Koch data de 1906, y es un ejemplo de curva no diferenciable en ningún punto, es decir, no existe lugar en donde su tangente pueda ser definida o trazada. Es de construcción sencilla, pero proporcionó una inspiración fundamental para generalizaciones posteriores. Se puede ver la curva de von Koch representada en la figura 3.2.

Se construye de la forma siguiente: tomemos un segmento de recta de longitud unidad \( \Gamma_0 \), normalizado, y sustituymos su tercio central por dos segmentos de longitud 1/3 que formen un triángulo equilátero con la base central anterior, que debe de ser eliminada. Este primer paso proporciona el objeto \( \Gamma_1 \), llamado generador. El segundo paso consiste en repetir el proceso anterior...
Figura 3.2: Primer paso ($\Gamma_0$), generador ($\Gamma_1$), y algunos estadios posteriores de la construcción de la curva de von Koch.

en cada uno de los segmentos de longitud $1/3$ restantes (ahora 4): sustituimos su tercio medio (de longitud $1/9$ respecto de la unidad inicial) por otros dos segmentos también de longitud $1/9$, en la misma disposición anterior. Este proceso debe iterarse infinitas veces, a fin de conseguir el fractal matemático real llamado curva de von Koch.

Estudiemos ahora cuál es la longitud de la curva poligonal a medida que avanzan las iteraciones:

<table>
<thead>
<tr>
<th>Iteración</th>
<th>$\Gamma_0$</th>
<th>$\Gamma_1$</th>
<th>$\Gamma_2$</th>
<th>$\Gamma_3$</th>
<th>$\ldots$</th>
<th>$\Gamma_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Longitud</td>
<td>1</td>
<td>$4^1 (1/3)$</td>
<td>$4^2 (1/3)^2$</td>
<td>$4^3 (1/3)^3$</td>
<td>$\ldots$</td>
<td>$4^n (1/3)^n$</td>
</tr>
</tbody>
</table>

Esta sucesión proporciona la siguiente longitud para la curva estudiada:

$$\lim_{n \to \infty} \left( \frac{4}{3} \right)^n \to \infty$$

Por tanto, el objeto que estamos considerando es no acotado, tiene una longitud infinita, y ello nos proporciona una primera idea de la complejidad de un viaje a lo largo de su perímetro: giros en el interior de los giros, entradas y salidas incesantes. Pensemos también que, por causa de la autosimilaridad, la distancia entre cualquier par de puntos en esta curva es infinita.

Necesitamos alguna forma de caracterizar el objeto descrito. Antes de dar ningún resultado, véase la figura 3.3, en la que se repasa brevemente la forma en que medimos curvas, superficies y volúmenes.

En los casos particulares allí representados, es un número entero ($D = 1, 2, 3$) el que se ocupa de establecer la ley de escala existente entre el número de elementos ($N$) con una cierta longitud lineal característica ($r$), necesarios para recubrir una recta, una superficie o un volumen, de forma
Figura 3.3: Relación entre las partes realizadas de un objeto $D$-dimensional y la dimensión del mismo, suponiendo que el hipervolumen está normalizado a valor 1 en cada espacio de trabajo. De cada uno de los objetos se realizan $N$ partes. El factor lineal de escala es, en cada caso (de arriba a abajo) $r_x = 1/N$, $r_y = 1/N^{1/2}$ y $r_z = 1/N^{1/3}$. Obsérvese que se mantiene la longitud, la superficie o el volumen totales si se verifica $N r^D = 1$, donde $D = 1, 2, 3$ es la dimensión topológica de cada uno de los objetos.

que el producto $N r^D$ se mantenga acotado. Este producto toma valor 1 si hemos partido de un elemento original normalizado. Vistos los ejemplos anteriores, se nos ocurre una generalización inmediata a aplicar a la curva de von Koch. ¿Cuál sería el valor de $D$ que mantendría el producto $N r^D$ finito? Recordemos que el número de elementos recubridores crecía en cada iteración en un factor 4, en tanto que su longitud lo hacía en un factor 1/3. Buscamos, pues, un valor $D$ tal que

$$
\lim_{n \to \infty} N r^D = \lim_{n \to \infty} 4^n \left( \frac{1}{3} \right)^n = 1
$$

Esto sólo ocurrirá si $4/3^D = 1$, ya que en caso de no producirse la igualdad, el límite $n \to \infty$ provocaría, o bien la divergencia (si $4/3^D > 1$), o bien un valor nulo (si $4/3^D < 1$). Obtenemos, por tanto,

$$
D = \frac{\ln 4}{\ln 3} \approx 1.2619
$$

Este valor se denomina dimensión de medida ($D_M$), y coincide con la dimensión fractal ($D_F$) siempre y cuando sea posible establecer el límite $n \to \infty$ (o, de forma equivalente, $1/r \to 0$). El ejemplo y la definición anteriores proporcionan las bases del método más habitualmente utilizado para calcular la dimensión fractal de un objeto: es el denominado "box-counting" (recuento por cajas), que describiremos más adelante.

Recordemos ahora el concepto de dimensión topológica, que unido a la dimensión medida definida proporcionará la primera caracterización cuantitativa de los objetos fractales. La dimensión topológica ($D_T$) de un objeto es un valor entero ($0, 1, 2, 3...$) que coincide con la dimensión del espacio de soporte. Por ejemplo, un punto tiene dimensión topológica 0, y una curva,
sea del tipo que sea, posee $D_T = 1$, tanto si se trata de una recta como de la curva de von Koch. Cualquier superficie presenta $D_T = 2$, y para los volúmenes, $D_T = 3$, como casos particulares contenidos en el espacio habitual de 3 dimensiones. Tenemos así la siguiente:

**Definición**

Se denomina *fractal* a todo objeto tal que posea dimensión topológica de valor menor que su dimensión de medida \(^1\):

$$D_T < D_M$$

$D_M$ proporciona, pues, un valor numérico que describe la estructura del fractal estudiado. En el caso particular de una curva, con $D_T = 1$, observaremos que la dimensión de medida $D_M$ puede tomar muy diversos valores, dependiendo de cada caso particular. Por ejemplo, veremos que para curvas fractales contenidas en el plano, $1 < D_M \leq 2$, con cualquier valor posible entre estos dos. El caso límite, $D_T = 1, D_M = 2$, lo proporcionarán las llamadas *curvas de Peano*. Pero no existe un valor máximo que $D_M$ pueda tomar. El movimiento browniano, por ejemplo, descrito en el capítulo 1, ocupa todo el espacio que tenga disponible (en tiempo infinito). Por tanto, aún con $D_T = 1$, es decir, tratándose de una curva unidimensional, $D_M > 1$, y si el espacio que contiene el movimiento browniano corresponde a un caso teórico de dimensión 5, obtendremos $D_M = 5$.

### 3.1.1 Dimensión de “box-counting”

Introduciremos en este apartado la forma práctica más utilizada para estimar la dimensión fractal de un objeto. Definimos, primeramente, la dimensión fractal como:

$$D_F = -\lim_{\delta \to 0} \frac{\ln N(\delta)}{\ln \delta}$$

donde $N(\delta)$ es el número de elementos de longitud característica $\delta$ necesarios para recubrir el conjunto estudiado. En la práctica, es en general imposible la realización del límite $\delta \to 0$, debido a la inexistencia, en la mayoría de los casos, de una expresión analítica que proporcione $N(\delta)$ en función de $\delta$, lo cual impide el cálculo teórico del límite anterior. La solución para estimar $D_F$ la aporta el método que nos ocupa. Observemos que es posible escribir:

$$\ln N(\delta) = D_F' \ln \left(\frac{1}{\delta}\right)$$

Si representamos $\ln(1/\delta)$ en el eje $x$ y $\ln N(\delta)$ en el eje $y$, debemos obtener aproximadamente una recta de pendiente $D_F' \equiv D_{BC} \ ^2$. Este valor, $D_{BC}$, es la llamada dimensión de “box-counting”, y en la mayoría de los casos coincide con $D_F$, o bien la aproxima de forma acertada. El cálculo práctico de $D_{BC}$ consistirá, pues, en conseguir unos cuantos puntos experimentales para realizar una buena estimación de la pendiente de la recta. Centrémonos en un caso concreto: el conjunto de la figura 3.5. Este conjunto es un atractor extraño procedente de un sistema dinámico del tipo Lotka-Volterra, que será descrito en el capítulo sobre sistemas dinámicos. Inicialmente, debe colocarse el atractor en el interior de un cuadrado de lado unidad (la normalización es necesaria). Este cuadrado se divide en subunidades de longitudes características $1/4, 1/8, 1/16, 1/25, 1/50, 1/100$ y $1/200$, en el ejemplo. Debemos contar, a cada escala, el número de cajas necesarias para contener el fractal. La representación en ejes logarítmicos de las dos series obtenidas permite la

---

\(^1\)En el caso de objetos no fractales, siempre se verifica la relación $D_T = D_M$.

\(^2\)Escribimos $D_F'$, en lugar de $D_F$, debido a que en principio no tienen porque ser valores coincidentes. Uno proviene de un límite infinito, y otro se obtiene a partir de una recta de regresión.
realización de una recta de regresión y la determinación de $D_{BC}$. Los resultados están representados en la figura 3.5.

Uno de los primeros ejemplos de cálculo de dimensión fractal con el método del "box-counting" lo dio Fry Richardson. Fue el cálculo de la $D_F$ correspondiente a la costa de Gran Bretaña. Richardson se dio cuenta de que la longitud de la costa variaba en función del detalle con que fuese dibujada, o, de forma equivalente, dependiendo de la longitud de la 'regla' elemental utilizada para medirla. Postuló que, en caso de disponer de precisión infinita, la longitud de la costa no sería una cantidad acotada, y estableció para ella una dimensión $D_F \approx 1.3$.

A partir de ahora no distinguiremos entre dimensión de medida, dimensión de "box-counting" y dimensión fractal, y en todos los casos descritos en esta sección nos referiremos a esta última.

### 3.1.2 Ejemplos

Veremos a continuación unos cuantos casos particulares de objetos fractales. Detallaremos su construcción y calcularemos su dimensión fractal. Son fractales de construcción geométrica bastante sencilla, y no será necesario el uso del método del "box-counting" para determinar $D_F$. Este método tendrá amplia utilización en el capítulo reservado a dinámica caótica, en particular en relación a los atractores extraños que allí aparecen.

Los fractales siguientes se han escogido por su carácter paradigmático y por la relevancia histórica que han tenido. Podríamos decir que son ya clásicos de la geometría fractal.

**Polvo de Cantor (Cantor Dust)**

George Cantor (1845-1918), matemático alemán, centró su trabajo principalmente en lo que actualmente llamamos teoría de conjuntos. Fue el primero en distinguir diferentes tipos de infinitos, darles un nombre y establecer un álgebra para ellos. El conjunto que ahora descubriremos fue publicado por vez primera en 1883, y es un ejemplo de toda una clase de conjuntos patológicos y excepcionales.

El conjunto se denomina *polvo de Cantor* (o simplemente *conjunto de Cantor*, $C_\infty$) debido a que está formado por una cantidad infinita (y no numerable) de puntos disjuntos. Se construye de la forma siguiente: de un segmento de longitud inicial unida se elimina el tercio central; en los dos segmentos restantes, ahora de longitud 1/3, se realiza la misma operación, y esta se repite sobre los segmentos que queden *ad infinitum*. El resultado final es un conjunto de puntos aislados en el
Figura 3.5: Atractor extraño producido por un modelo de Lotka-Volterra. Para el cálculo de la dimensión de "box-counting" se divide el cuadrado unidad que contiene el fractal en cajas iguales de diversos tamaños, que proporcionarán los datos experimentales para realizar la recta de regresión. En la parte inferior derecha se puede ver la recta de regresión calculada a partir de las divisiones del conjunto especificadas en el texto. La dimensión de "box-counting" del atractor es $D_{BC} = 1.687$

Figura 3.6: Detalle de los primeros pasos de la construcción del polvo de Cantor descrito en el texto. A medida que las iteraciones avanzan, el conjunto se hace cada vez más difuso, hasta que es imposible de distinguir, debido a su naturaleza puntual (en el límite $n \to \infty$).
Figura 3.7: Las tres primeras iteraciones en la construcción de una de las curvas de Peano. A medida que las iteraciones avanzan se hace más difícil recordar que habíamos partido de una curva, y el objeto resultante cada vez parece más una porción sólida del plano.

intervalo real [0,1]. Evidentemente, el conjunto límite es imposible de representar. Los primeros pasos de la construcción se pueden ver en la figura 3.6. Observemos que, debido a su naturaleza de conjunto formado por puntos discretos, \( D_T = 0 \) para el polvo de Cantor. Calculemos su dimensión fractal. Para ello sólo es necesario apreciar que, en el paso \( n \) de su construcción, son necesarios \( 2^n \) segmentos para recubrirlo, que tendrán una longitud \( (1/3)^n \). Por tanto, la dimensión fractal del conjunto de Cantor es

\[
D_F = \frac{\ln 2}{\ln 3} \approx 0.6309
\]

Verificamos aquí que \( D_T < D_F \), y \( D_F \) mide la forma en que el conjunto recubre el espacio de dimensión 1, en este caso.

Comentemos finalmente que la intersección de la curva de von Koch con una recta trazada en su base proporciona exactamente el descrito polvo de Cantor.

**La Curva de Peano**

Las curvas de Giuseppe Peano (1858-1912) representan un ejemplo extremo de cómo un objeto de dimensión topológica 1, como es una recta, puede doblarse sobre el mismo de tal forma que llegue a cubrir totalmente el plano. Esto provoca que su dimensión fractal alcance el valor 2, como calcularemos. Inmediatamente después de Peano fue David Hilbert (1862-1943) quien se dedicó al estudio de estas curvas recubridoras del plano, y también propuso unos cuantos ejemplos. Su trabajo en lo que respecta a los fractales es, sin embargo, prácticamente insignificante, comparado con las excelentes contribuciones que realizó en otros muchos campos de las matemáticas.

En la figura 3.7 podemos observar los pasos 0, 1 (generador) y 2 en la construcción de una curva de Peano.

De nuevo es el tercio central el que se sustituye, en este caso por siete nuevos segmentos, en el orden especificado en la figura. En el paso \( n \) de la construcción tendremos, pues, un total de \( 9^n \) segmentos, escalados en una fracción \( (1/3)^n \) del total. Esto proporciona la dimensión fractal

\[
D_F = \frac{\ln 9}{\ln 3} = 2
\]

para la curva de Peano.
Figura 3.8: Primeros estadios de la construcción del triángulo de Sierpiński. La eliminación de triángulos, llevada al límite, provoca que el objeto tenga superficie nula. Sus paredes se convierten en líneas de grosor cero.

Por extensión, y en atención al primer constructor de estas curvas, todos los fractales con $D_T = 1$ y $D_F = 2$ se denominan curvas de Peano.

El triángulo de Sierpiński y la esponja de Sierpiński-Menger

Waclaw Sierpiński (1882-1969), matemático polaco, introdujo en 1916 otro fractal clásico, el triángulo de Sierpiński. Fue un matemático muy reconocido en su tiempo, en el mundo entero, tanto que hasta uno de los cráteres de la luna fue bautizado con su nombre. La construcción de su triángulo se realiza de la forma que se puede ver en la figura 3.8. Dado un triángulo equilátero, se elimina el triángulo interior delimitado por la unión de los puntos medios de cada lado. Esto proporciona el objeto generador. Al igual que en los casos anteriores, el proceso se repite en cada una de las subunidades restantes, hasta el infinito. Este fractal fue publicado en un artículo de 1916, llamado "Sobre una curva en la cual todo punto es un punto de ramificación", título que describe la principal característica del conjunto: la no existencia de puntos aislados y el contacto de cada punto con al menos otros tres del conjunto. Para calcular la dimensión fractal, observemos que en el estadio de construcción $n$, necesitamos $3^n$ triángulos de tamaño lineal $(1/2)^n$, con lo cual

$$D_F = \frac{\ln 3}{\ln 2} \approx 1.585$$

Otro de los objetos creados por Sierpiński, la alfombra de Sierpiński, inspiró a Karl Menger un fractal en 3 dimensiones (hasta ahora sólo habíamos visto fractales contenidos en el plano), la esponja de Sierpiński-Menger. La alfombra de Sierpiński es simplemente una proyección en 2 dimensiones de este último. Este objeto, uno de los pocos ejemplos de fractales geométricos en 3 dimensiones, se construye de la forma siguiente: de un cubo de arista unidad se elimina el cubo central de arista $1/3$ y los cubos centrales de las seis caras, es decir, 7 en total. El proceso se repite sobre cada uno de los 20 cubos de arista $1/3$ restantes, de nuevo hasta el infinito. El resultado final es un fractal de hojaldre, con una $D_T = 2$, ya que el proceso de vaciado ha reducido su grosor a cero. La dimensión fractal es

3Debe recordarse que en el proceso de eliminación de triángulos para formar la figura final se deja una superficie de grosor cero, es decir, el triángulo inicial se ha reducido a una serie infinita de líneas, por tanto con dimensión topológica 1.
\[ D_F = \frac{\ln 20}{\ln 3} \approx 2.727 \]

### 3.2 Fundamentos Matemáticos de la Geometría Fractal

El lector que posea conocimientos básicos de teoría de funciones, topología y teoría de la medida puede saltarse esta sección. Además, de los conceptos estrictamente necesarios para caracterizar correctamente los objetos fractales, introduciremos algunos otros que no son necesarios aquí, pero a los que se hará referencia desde otros capítulos del libro, y que por continuidad se ha creído conveniente definir aquí de forma agrupada.

La caracterización rigurosa de un objeto fractal pasa necesariamente por ciertos conceptos generales relacionados con la teoría de conjuntos, la topología, con funciones y límites de funciones, unas dosis de teoría de la medida y algunas nociones de teoría de la probabilidad. Intentaremos trabajar las bases matemáticas a nivel elemental (pero no trivial) para producir un primer acercamiento a la geometría fractal. Todas las definiciones y conceptos que veremos en este apartado se pueden encontrar de forma más extensa en Falconer (1990) y Apostol (1974), entre otras muchas más referencias de topología y análisis general.

Fijaremos nuestras ideas en los objetos fractales, que son el propósito de este capítulo, e intentaremos que todas las definiciones dadas en los apartados siguientes estén reforzadas visualmente por alguno de estos conjuntos. Aprovecharemos para ello las ideas semi intuitivas que hemos dado en la primera sección del capítulo. Supondremos conocidas ciertas nociones matemáticas a nivel elemental como espacio euclídeo, conjunto vacío, etc.

#### 3.2.1 Teoría básica de conjuntos

Los objetos fractales están localizados en el espacio euclídeo \( n \)-dimensional, \( \mathbb{R}^n \). Una parte muy importante de ellos tiene una representación geométrica sencilla donde \( n = 1, 2, 3 \). A esta clase pertenecen todos los fractales que iterativamente se han ido construyendo en la sección anterior. Nuestros conjuntos serán en general subconjuntos de \( \mathbb{R}^n \). Por ejemplo, consideremos la curva de von Koch, que podemos denotar por \( \Gamma_\infty \). Escribiremos \( \Gamma_\infty \subseteq \mathbb{R}^2 \) para indicar que esta curva es un objeto contenido en el plano euclídeo. Los puntos de \( \mathbb{R}^n \) serán de la forma \( x, y, z, \ldots, \) \( n \)-vectors \( n \)-dimensionales, de \( n \) componentes, y escribiremos \( r \in \mathbb{R}^n \) para indicar la pertenencia del punto al espacio. Una bola cerrada de centro \( x \) y radio \( r \) se define como

\[ B_r(x) = \{ y : |y - x| \leq r \} \]

es decir, el conjunto de puntos \( y \) tales que la distancia entre el centro \( x \) y los puntos del conjunto \( B_r(x) \) es menor o igual que \( r \). \( |y - x| \) es la distancia euclidiana habitual,

\[ |y - x| = \left( \sum_{i=1}^{n} |x_i - y_i|^2 \right)^{\frac{1}{2}} \]

---

1. Recuerdemos que una curva tiene dimensión topológica 1. Si la curva es fractal, su dimensión de medida es mayor que 1. Esto significa que la dimensión mínima de un espacio euclídeo que contenga este objeto es 2: \( D_T < D_M < n \), donde \( n \) es el primer entero mayor que \( D_M \).

2. Es habitual designar las cantidades vectoriales, pertenecientes a \( \mathbb{R}^n \) en negrita, \( x \in \mathbb{R}^n \), tal y como se ha visto anteriormente (por ejemplo en el capítulo dedicado a los sistemas dinámicos). La norma en todo este libro será el uso de la negrita para designar las cantidades vectoriales. Únicamente en este capítulo no lo haremos así, tal como se define, para no sobrecargar la notación. Dado que se especifica en cada caso, no existe posible confusión.
Figura 3.9: La esponja de Sierpiński-Menger a la izquierda, en tres de sus primeros pasos de construcción. A la derecha, otro fractal en 3 dimensiones, construido de forma geométrica sencilla por uno de los autores (S.C.M.). La dimensión fractal de este último es 2.807, como el lector podrá comprobar.
Una bola abierta es

$$B_r^o(x) = \{ y; |y - x| < r \}$$

Las bolsas cerradas contienen la frontera que las separa del resto de $$\mathbb{R}^n$$, no así las bolsas abiertas\(^6\). Como ejemplo consideremos el conjunto de Cantor de la figura 3.6. En su construcción eliminamos el tercio central, que es una bola abierta $$B_r^o \subset \mathbb{R}^1$$, y los conjuntos que quedan son bolsas cerradas de $$\mathbb{R}^1$$. Una inspección cuidadosa de este fractal revela que, en el límite, únicamente los puntos situados en los vértices de los segmentos restantes tras cada iteración permanecen en el conjunto. Si en lugar de sustraer bolsas abiertas al segmento inicial sustrajésemos bolsas cerradas, el resultado final sería un conjunto vacío.

En el caso particular de $$\mathbb{R}^1$$, denominamos a las bolsas abiertas intervalos abiertos, $$\{ x; a < x < b \}$$, con $$a < b$$, y a las bolsas cerradas intervalos cerrados, $$\{ x; a \leq x \leq b \}$$. Se define por extensión como conjunto abierto aquel que no contiene ninguno de sus puntos frontera, y como conjunto cerrado aquel que contiene la frontera.

Denominaremos $$\cup_\alpha A_\alpha$$ a la unión de una colección arbitraria de conjuntos $$\{ A_\alpha \}$$ y $$\cap_\alpha A_\alpha$$ a su intersección. Recordemos de nuevo que, por ejemplo, la intersección de la curva de von Koch ($$\Gamma_\infty$$) con $$\mathbb{R}^1$$ situado en su base proporciona el conjunto de Cantor ($$C_\infty$$) descrito:

$$\Gamma_\infty \cap \mathbb{R}^1 = C_\infty$$

La diferencia entre dos conjuntos $$A$$ y $$B$$ ($$A - B$$) consta de los puntos contenidos en $$A$$ que no pertenecen a $$B$$. El conjunto $$\mathbb{R}^n - A$$ se denomina complementario de $$A$$, y lo denotaremos $$\dot{A}$$. El complementario de un conjunto abierto es cerrado, y viceversa. Un conjunto infinito es numerable si a sus elementos se les puede asignar un lugar concreto en una serie $$x_1, x_2, x_3, ...$$, es decir, si existe una correspondencia entre los números naturales y los elementos del conjunto. Si esta correspondencia no puede ser establecida, el conjunto se denomina no numerable. Ejemplos de conjuntos numerables son los números naturales $$\mathbb{N}$$, los racionales $$\mathbb{Q}$$, o el número de puntos

\(^6\)La frontera es el conjunto $$F = \{ y; |x - y| = r \}$$, y tanto es frontera de la bola abierta como de la bola cerrada.
dibujos de una curva de Peano (aquéllos por los que pasamos dos veces cuando construimos el fractal utilizando una línea continua). Ejemplos de conjuntos no numerables son los números reales \( \mathbb{R} \), el número de vértices de la curva de Helge von Koch o el número de puntos del conjunto de Cantor \( C_\infty \).

Si \( A \) es un conjunto de números reales, el *supremo* de \( A \) es el menor número \( m \in \mathbb{R} \) tal que \( m \leq x, \forall x \in A \). Escribiremos \( m = \sup(A) \). El *infinito* será el mayor número \( m \) tal que \( m \leq x, \forall x \in A \): \( m = \inf(A) \). El *diámetro* de un subconjunto \( A \subset \mathbb{R} \) será la mayor distancia entre dos puntos contenidos en \( A \). Un conjunto se llama *acotado* si tiene diámetro finito. Una sucesión \( \{x_k\} \) de \( \mathbb{R}^n \) converge a un punto \( x \in \mathbb{R}^n \) cuando \( k \to \infty \) si, dado \( \epsilon > 0 \),

\[ \exists K; \quad |x_k - x| < \epsilon \quad \forall k > K \]

es decir, si \( |x_k - x| \) tiende a cero. \( x \) es el límite de la sucesión,

\[ \lim_{k \to \infty} x_k = x \]

Un conjunto \( B \) es un subconjunto denso en \( A \) si siempre hay un punto de \( B \) arbitrariamente cercano a un punto de \( A \). El conjunto \( Q \) de los números racionales es denso en \( \mathbb{R} \), el conjunto de los reales. Un conjunto \( A \) se denomina *compacto* si cualquier colección de conjuntos abiertos que recubran a \( A \) contiene una subcolección finita que también sea recubridora de \( A \).

Se puede demostrar que cualquier subconjunto compacto de \( \mathbb{R}^n \) es a la vez cerrado y acotado. Por esta razón, todos los fractales geométricos que hemos visto son conjuntos compactos. La intersección de una colección arbitraria de conjuntos compactos es un conjunto compacto. Se puede comprobar que si \( A_1 \supset A_2 \supset \ldots \) es una secuencia decreciente de conjuntos compactos, la intersección

\[ \bigcap_{i=1}^{\infty} A_i \]

es no-vacia. Este teorema es importante en la construcción de fractales de forma iterativa por eliminación de conjuntos abiertos. Consideremos, por ejemplo, la alfombra de Sierpiński. Observamos que cada nueva iteración es un conjunto compacto \( A_i \) (cerrado, por contener a su frontera, y acotado, por existir un conjunto finito mayor que lo contiene), contenido en la iteración anterior \( A_{i-1} \). El conjunto límite se construye como la intersección de los conjuntos resultantes de iterar un número infinito de veces el proceso. El teorema anterior nos asegura que el conjunto límite existe, y es no-vacio.

Un subconjunto \( A \subset \mathbb{R}^n \) es conexo si no existen conjuntos abiertos \( U \) y \( V \) tales que \( U \cup V \) contenga el conjunto \( A \) y se verifique que las intersecciones \( A \cap U \) y \( A \cap V \) son disjuntas y no vacías. De forma intuitiva (y como el nombre indica) pensaremos en conjuntos conexos como aquellos formados por una sola pieza. El mayor subconjunto de \( A \) que contenga un cierto punto \( x \) se denomina *componente conexa* de \( x \). Un conjunto \( A \) es totalmente desconexo si la componente conexa de cada punto \( x \) es únicamente ese punto, es decir, \( A \) es un conjunto formado por puntos aislados.

En el caso de los objetos fractales, es importante observar que, por causa de la autosimilaridad, todo fractal puro es, o bien totalmente conexo (consta de una única pieza conexa), o bien totalmente desconexo (está formado por un conjunto discreto de puntos). Un repaso a los fractales hasta ahora

---

7En algunos casos, cuando utilizamos volúmenes geométricos \( V \subset \mathbb{R}^n \), nos referiremos a la *longitud característica* (6), definida como \( V^\frac{1}{n} \). Cuando hablamos de recubrimientos para calcular dimensiones fractales, veremos que utilizar el diámetro o la longitud característica de los elementos recubridores no varía los resultados cuantitativamente. En el caso de \( n = 2 \), tendremos superficies recubridoras, y \( \delta = \sqrt{5} \).
definidos corroborará esta afirmación. Volveremos sobre la conectividad más adelante, cuando trataremos los llamados conjuntos de Julia.

La noción de conectividad nos proporciona una definición rigurosa de dimensión topológica: un conjunto tiene dimensión topológica \( n \) si para separarlo en dos partes necesitamos extraer un conjunto de dimensión \( n - 1 \). Así, como que para desconectar una curva nos basta con restarle un punto \( (D_1 = 0) \), una curva es un objeto con \( D_1 = 1 \). Para desconectar un plano \( (D_2 = 2) \), necesitamos restarle una curva \( (D_2 = 1) \), etc.

Definiremos finalmente conjunto de Borel o boreliano. La clase de conjuntos de Borel es la menor subcolección de subconjuntos de \( \mathbb{R}^n \) con las siguientes propiedades:

1. Todo conjunto abierto y todo conjunto cerrado es un boreliano

2. Tanto la unión como la intersección de una colección arbitraria finita o numerable de conjuntos de Borel es un boreliano.

### 3.2.2 Funciones y Límites

Los conceptos que se van a definir en esta sección serán en particular aplicados en una sección posterior dedicada a los llamados sistemas de funciones iteradas (IFS), en donde ciertas transformaciones bien definidas se aplicarán sobre conjuntos cerrados para producir objetos fractales.

Consideremos dos conjuntos, \( X \) e \( Y \). Una aplicación, función o transformación \( f \) de \( X \) a \( Y \) es cierta regla que asocia un punto \( f(x) \in Y \) a cada punto \( x \in X \). Se escribe

\[
f : X \rightarrow Y
\]

\( X \) es el dominio de \( f \). \( f(x) \) es la imagen de \( x \), y ésta es a su vez antiimagen de \( f(x) \). En general, la imagen y la antiimagen no están definidas necesariamente de forma única. Escribiremos \( f(A) \) para denotar la imagen de un subconjunto \( A \subset X \) y \( f^{-1}(B) \) para referirnos a la antiimagen de \( B \subset Y \).

Una función \( f : X \rightarrow Y \) es inyectiva si \( f(x) \neq f(y) \), cuando \( x \neq y \). \( f \) se denomina sobreyectiva si \( \forall y \in Y \), existe un elemento \( x \in X \) con \( f(x) = y \), es decir, todo elemento de \( Y \) es imagen de algún punto de \( X \). Una función que sea de forma simultánea inyectiva y sobreyectiva se denomina biyectiva (correspondencia uno a uno). En este caso se puede definir la función inversa \( f^{-1} : Y \rightarrow X \), tomando como \( f^{-1}(y) \) el único elemento de \( X \) tal que \( f(x) = y \). Entonces,

\[
f^{-1}(f(x)) = x, \quad \forall x \in X
\]

\[
f(f^{-1}(y)) = y, \quad \forall y \in Y
\]

La composición de las funciones \( f : X \rightarrow Y \) y \( g : Y \rightarrow Z \) es la función

\[
g \circ f : X \rightarrow Z
\]

dada por

\[
(g \circ f)(x) = g(f(x))
\]

Se puede componer un número arbitrario de funciones, generalizando de forma obvia la definición anterior. Algunas funciones que se aplican de \( \mathbb{R}^n \) en \( \mathbb{R}^n \) tienen un significado geométrico importante, y son las siguientes:
Una isometría $S : \mathbb{R}^n \to \mathbb{R}^n$ es una aplicación que preserva la distancia entre dos puntos:

$$|S(x) - S(y)| = |x - y|, \quad \forall x, y \in \mathbb{R}^n$$

También se las denomina congruencias. Son casos particulares de congruencias:

- Las traslaciones, $S(x) = x + a$, en donde los puntos son desplazados una distancia $|a|$ paralelos al vector $a$.
- Las rotaciones, $|S(x) - a| = |x - a|$, en donde $a$ es el centro de rotación.
- Las reflexiones, en las que los puntos se convierten en su imagen especular respecto de una hipersuperficie $(n - 1)$-dimensional.

Una isometría directa es suma de traslaciones y rotaciones (no reflexiones). Una transformación $S : \mathbb{R}^n \to \mathbb{R}^n$ se llama de semejanza si existe una constante $c$ que verifique

$$|S(x) - S(y)| = c|x - y|, \quad \forall x, y \in \mathbb{R}^n$$

Una aplicación $T : \mathbb{R}^n \to \mathbb{R}^n$ es lineal si

$$T(x + y) = T(x) + T(y) \quad y \quad T(\lambda x) = \lambda T(x), \quad \forall x, y \in \mathbb{R}^n$$

y $\lambda \in \mathbb{R}$. $T$ se puede representar mediante una matriz, y es no-singular si $T(x) = 0$ si y sólo si $x = 0$.

Una afinidad o transformación afin verifica

$$S(x) = T(x) + a$$

donde $T$ es una transformación lineal no-singular y $a$ es un punto de $\mathbb{R}^n$.

Una función $f : X \to Y$ es una función de Hölder de exponente $\alpha$ si

$$|f(x) - f(y)| \leq c|x - y|^{\alpha}, \quad x, y \in X$$
Figura 3.12: Aplicaciones afines sobre el conjunto representado. Las homotecias contraen en una de las dos direcciones o en ambas al mismo tiempo, proporcionando figuras que no son isométricas a la inicial. La reflexión (en este caso según un eje vertical) y la rotación si conservan la distancia entre puntos, y son por tanto transformaciones de isometría.

Para alguna constante $c$, $f$ se denomina función de Lipschitz si $c = 1$, y bi-Lipschitz si

\[ c_1|x - y| \leq |f(x) - f(y)| \leq c_2|x - y|, \quad x, y \in X \]

para $0 < c_1 \leq c_2 < \infty$.

Consideremos $X \subseteq \mathbb{R}^n$ y $Y \subseteq \mathbb{R}^m$. Sea $f : X \to Y$ cierta aplicación y $a \in \bar{X}$ ($\bar{X} = X \cup F_X$, es decir, el conjunto $X$ más su frontera)\(^8\). Diremos que $f(x)$ tiene límite $y$ (o tiende a $y$) cuando $x \to a$ si, dado $\epsilon > 0$, \(\exists \delta > 0\) tal que $|f(x) - y| < \epsilon$, $\forall x \in X$, con $|x - a| < \delta$. En general lo escribiremos como:

\[ \lim_{x \to a} f(x) = y \]

$f(x)$ tiende a infinito si, dado $M$, \(\exists \delta > 0\) tal que $f(x) > M$, para $|x - a| < \delta$. Entonces:

\[ \lim_{x \to a} f(x) = \infty \]

y de forma similar para $f(x) \to -\infty$. Como casos particulares, en relación a los cálculos que realizaremos con objetos fractales, estaremos especialmente interesados en los límites $x \to 0$.

Supongamos $f : \mathbb{R}^+ \to \mathbb{R}$. Es posible que el límite no exista, que la función presente infinitas oscilaciones cuando nos acerquemos a este punto. Definimos en estos casos el límite superior.

\[ \limsup_{x \to 0} f(x) \equiv \lim_{x \to 0}(\sup\{f(x) : 0 < x < r\} \]

$\sup\{f(x) : 0 < x < r\}$ es o bien $\infty$ $\forall r > 0$, o bien decrece cuando $r$ decrece, y por tanto el límite superior siempre existe. El límite inferior se define de forma similar.

\(^8\) \(\bar{X}\) se denomina en general adherencia de $X$, y es el menor conjunto cerrado que contiene $X$. Si $X$ es cerrado, $\bar{X} = X$. 
Figura 3.13: La aplicación representada es una función continua en los valores reales de $x$ y discontinua en los racionales. A cada valor de $x \in [0,1]$ racional con la forma fraccionaria irreductible $p/q$ se le asocia el valor de $y = 1/q$. En los valores irracionales vale 0. La función tiene propiedades fractales claramente visibles.

$$\liminf_{x \to 0} f(x) \equiv \lim_{x \to 0} \left( \inf \{f(x) : 0 < x < r\} \right)$$

Si el límite superior y el inferior coinciden, $\lim_{x \to 0} f(x)$ existe, y $\lim_{x \to 0} f(x) = \liminf f(x) = \limsup f(x)$.

Una función $f : X \to Y$ es continua en un punto $a \in X$ si $f(x) \to f(a)$ cuando $x \to a$, y es continua en $X$ si es continua en todos los puntos del conjunto. Si $f : X \to Y$ es una función biyectiva y continua con inversa $f^{-1} : Y \to X$ continua, se denomina homeomorfismo, y en este caso $X$ y $Y$ son conjuntos homeomorfos.

La función $f : \mathbb{R} \to \mathbb{R}$ es diferenciable en $x$ con el valor $f'(x)$ como derivada si

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h} = f'(x)$$

existe. Será diferenciable en continuidad si $f'(x)$ es continua en $x$. En general, $f : \mathbb{R}^n \to \mathbb{R}^n$ es diferenciable en $x$ con la aplicación lineal $f'(x) : \mathbb{R}^n \to \mathbb{R}^n$ como derivada si se verifica

$$\lim_{|h| \to 0} \frac{|f(x+h) - f(x) - f'(x)h|}{|h|} = 0$$

Diremos que una secuencia de funciones $\{f_k\}$, $f_k : X \to Y$, $X \subset \mathbb{R}^n$, $Y \subset \mathbb{R}^m$, converge punto a punto a una función $f : X \to Y$ si $f_k(x) \to f(x)$ cuando $k \to \infty$, $\forall x \in X$. La convergencia es uniforme si

$$\sup_{x \in X} |f_k(x) - f(x)| \to 0 \ (k \to \infty)$$

Si las funciones $f_k$ son continuas y convergen uniformemente a $f$, entonces $f$ es continua.

### 3.2.3 Medidas y Distribuciones de Masa

El concepto de medida será de especial utilidad en la sección dedicada a multifractales (objetos que, debido a su complejidad, necesitan ser caracterizados por un espectro continuo de dimensiones fractales). Unos ligeros esbozos de la teoría de la medida serán suficientes para nuestros propósitos.
Sólo nos referiremos a medidas definidas sobre subconjuntos de \( \mathbb{R}^n \). La idea será asociar una cantidad numérica (la medida) a cada una de las partes en que el conjunto pueda ser dividido, o de las que esté formado.

Llamamos a \( \mu \) medida en \( \mathbb{R}^n \) si a cada subconjunto le asocia un valor no negativo, de la forma siguiente:

1. \( \mu(\emptyset) = 0 \), es decir, el conjunto vacío tiene medida nula.
2. \( \mu(A) \leq \mu(B) \) si \( A \subset B \)
3. Si \( A_1, A_2, \ldots \) es una secuencia finita (o numerable) de conjuntos, entonces:

\[
\mu \left( \bigcup_{i=1}^{\infty} A_i \right) \leq \sum_{i=1}^{\infty} \mu(A_i)
\]

La igualdad se cumple cuando la colección \( \{A_i\} \) está formada por borelianos disjuntos \( (A_i \cap A_j = \emptyset, \forall i, j) \).

\( \mu(A) \) es la medida del conjunto \( A \). Dos conjuntos del mismo tamaño (en el sentido geométrico) pueden tener medida diferente. Veremos que a un conjunto se le puede asociar una medida mayor (numéricamente) que a otro, aún cuando su tamaño geométrico (la “superficie” que ocupa) sea menor. Cuando la medida de un conjunto es proporcional a su tamaño geométrico, diremos que trabajamos con una medida de masa. En este caso particular, si tomamos un conjunto acotado de \( \mathbb{R}^n \), \( 0 < \mu(\mathbb{R}^n) < \infty \), la medida de masa se puede asociar directamente a la “masa” del conjunto: si lo dividimos en partes, cada una de ellas tiene un valor coincidente con el valor de su masa (o proporcional a ella).

Llamaremos soporte de una medida, \( M \), al conjunto sobre el cual ésta se concentra: es el menor conjunto cerrado \( X \) tal que

\[
\mu(\mathbb{R}^n - X) = 0
\]

El soporte \( M \) de \( \mu \) siempre es cerrado, y \( x \in M \) si y sólo si

\[
\mu(B_r(x)) > 0, \quad \forall r > 0
\]

Diremos que \( \mu \) es una medida sobre un conjunto \( A \) si éste contiene el soporte de \( \mu \). En todos los casos que contemplaremos efectuaremos siempre una normalización sobre la medida (en ese caso, podremos pensar en el valor asociado a cada parte \( A_i \) de nuestro conjunto \( A \) como en su probabilidad de ocurrencia). Normalizar la medida significa que \( \mu(A) = 1 \), y si

\[
A = \bigcup_{i=1}^{\infty} A_i
\]

siempre \( \mu(A_i) \leq 1 \). En los casos que nos ocuparán, la división que haremos del conjunto \( A \) será en \( n \) piezas disjuntas, y por tanto

\[
\mu(A) = \mu(A_1) + \mu(A_2) + \ldots + \mu(A_n) = 1
\]

Tras esta introducción a las nociones básicas de conjuntos y funciones y la definición dada de medida estamos en condiciones de introducir la medida y la dimensión de Hausdorff. La dimensión de Hausdorff asocia un valor numérico no necesariamente entero a cualquier subconjunto de \( \mathbb{R}^n \). Aunque introducida muchos años antes de la “invasión” de los fractales, con estos se reveló de especial utilidad, ya que permite distinguir, en una primera aproximación, el objeto con el que estamos trabajando.
Medida de Hausdorff

Consideremos $F \subseteq \mathbb{R}^n$ y una colección $\{U_i\}$ de conjuntos de diámetro máximo $\delta$ que recubren $F$:

$$F \subseteq \bigcup_{i=1}^{\infty} U_i$$

con $0 < |U_i| \leq \delta$, $\forall i$, diremos que $\{U_i\}$ es un $\delta$-recubrimiento de $F$. Definimos, $\forall \delta > 0$, $s \geq 0$,

$$\mathcal{H}_{\delta}^s(F) = \inf \left\{ \sum_{i=1}^{\infty} |U_i|^s ; \{U_i\} \text{ es un } \delta \text{-recubrimiento de } F \right\}$$

Estamos buscando el mínimo de la suma de las potencias $s$-ésimas del diámetro del $\delta$-recubrimiento. Si $\delta \to 0$, el número de recubrimientos permitidos disminuye, y se acerca a un límite,

$$\mathcal{H}^s(F) = \lim_{\delta \to 0} \mathcal{H}_{\delta}^s(F)$$

Esta es la medida de Hausdorff s-dimensional de $F$. Cuando los conjuntos medidos son los volúmenes, superficies y curvas habituales, $s$ toma valores enteros que se corresponden con las dimensiones topológicas conocidas. La medida de Hausdorff posee una propiedad de escala fundamental:

$$\mathcal{H}^s(\lambda F) = \lambda^s \mathcal{H}^s(F)$$

es decir, cuando el conjunto $s$-dimensional sobre el que aplicamos la medida se multiplica en un factor $\lambda$, la medida del conjunto se ve modificada en un factor $\lambda^s$. Recordemos en este punto la primera sección de este capítulo, en donde dividíamos una recta, un cuadrado y un cubo en subunidades y obteníamos un factor de escala $D$, en aquel caso un número entero. Para un objeto fractal, a continuación, $D$ tomaba un valor irracional. Aquel valor $D$ es lo que ahora llamamos $s$, dimensión de la medida de Hausdorff.

Dimensión de Hausdorff

Volvamos sobre $\mathcal{H}^s_{\delta}(F)$, definida en el párrafo anterior. Para cualquier conjunto $F y \delta < 1$, $\mathcal{H}_{\delta}^s(F)$ es una función no-creciente con $s$, así que $\mathcal{H}^s(F)$, según su definición, también lo es. Si consideramos $t > s$ y $\{U_i\}$ un $\delta$-recubrimiento de $F$, se verifica

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

Tomando mínimos,

$$\mathcal{H}^s_{\delta}(F) \leq \delta^{t-s} \mathcal{H}^s_{\delta}(F)$$

Haciendo el límite $\delta \to 0$, se observa que si $\mathcal{H}^s(F) < \infty$, entonces $\mathcal{H}^s(F) = 0$, para $t > s$. Si representamos $\mathcal{H}^s(F)$ en función de $s$, observaremos que existe un valor crítico $s^*$ en el que $\mathcal{H}^s(F)$ cambia su valor de $\infty$ a 0. Este valor crítico $s^*$ se denomina dimensión de Hausdorff de $F (\dim_H F)$, o también dimensión de Hausdorff-Besicovitch.

Recordemos de nuevo que, cuando definimos la dimensión de medida, pediamos que el producto $N \tau^D$ se mantuviera finito cuando $n \to \infty$ (número de elementos recubridores). Obtenemos un claro paralelismo con las nociones que ahora estamos definiendo: sólo existe un valor de $D$ (ahora $s$) que en el límite $n \to \infty$ (en la mayoría de los casos equivalente a reducir a cero el diámetro del

---

9 Para más detalles, y sobre la importancia de las funciones de escala, véase la sección 8.2.
Figura 3.14: Representación cualitativa del valor de $s^*$ (dimensión de Hausdorff-Besicovitch) que hace $\mathcal{H}^s(F)$ finito.

recubrimiento, $\delta \to 0$) produzca un valor finito de $N^s(\delta)$ (cantidad análoga a la medida de Hausdorff de dimensión $s$), y ese valor es la dimensión de nuestro conjunto.

El cálculo exacto de la dimensión de Hausdorff no siempre es posible, y con frecuencia resulta muy complicado. Esta dimensión es el mejor estimador de la dimensión de los objetos fractales, pero muy a menudo se utilizan definiciones alternativas de dimensión, mucho más apropiadas en la práctica. Entre éstas, la más utilizada es la ya descrita dimensión de "box-counting". Remarquemos que en ese caso no se calcula un límite analítico, sino que se interpola una recta para valores de $\delta$ no necesariamente pequeños. En ocasiones, esto puede provocar la aparición de dimensiones dependientes de la escala, es decir, puede ser que la pendiente de la recta que debería dar la dimensión fractal no se mantenga constante. Esta dependencia en $\delta$ es con frecuencia importante, y no casual. Volveremos sobre ella al referirnos a los objetos multifractales. Otra característica importante del método del "box-counting" es que el $\delta$-recubrimiento se puede escoger de diversas formas, y en todos los casos se puede demostrar que el valor de la dimensión obtenido es, en el límite, el mismo. Así, sería equivalente tomar bolas de radio $\delta$, cubos de arista $\delta$, una red $n$-dimensional de cubos de lado $\delta$ o conjuntos irregulares de diámetro máximo $\delta$, entre otros. Recordemos que, para obtener la dimensión de Hausdorff, este recubrimiento debe ser el mínimo, y en general, con el método de "box-counting", se produce una sobreestimación del recubrimiento necesario.

Muchas otras definiciones de dimensiones alternativas pueden ser halladas en la literatura especializada. Para nuestros propósitos, lo visto hasta ahora será suficiente.

3.3 Sistemas de funciones iteradas (Iterated function systems, IFS)

Estudiaremos en esta sección un tipo de fractales deterministas, definidos por transformaciones afines y comúnmente denominados IFS (sistemas de funciones iteradas). La idea general consiste en aplicar cierto conjunto de transformaciones afines contractivas (serán de semejanza, con $c < 1$ \footnote{Véase la sección anterior.}), a un conjunto cerrado arbitrario de $\mathbb{R}^n$. Cuando el sistema de funciones se ha aplicado un número infinito de veces, llegamos a lo que se llama punto fijo de la transformación, y éste es, ni
más ni menos, un conjunto fractal. Estas nociones serán especificadas más adelante. La forma en que los fractales aquí aparecen, como punto fijo de un conjunto de transformaciones afines, nos permitirá ilustrar brevemente el problema inverso: dado un conjunto compacto autosimilar de \( \mathbb{R}^n \), ¿cómo sería posible determinar el conjunto de transformaciones que lo definen? Esta idea puede ser de gran importancia aplicada a la codificación de imágenes. Se conseguiría una gran compresión de la información si en lugar de ser necesaria la especificación de un color para cada pixel de una imagen (unos \( 10^4 \) parámetros), ésta pudiese ser especificada por unas diez transformaciones de afinidad (menos de \( 10^2 \) parámetros serían suficientes). Una descripción más exhaustiva de los IFS se puede encontrar en los libros de Barnsley (1988) y Peitgen (1992).

### 3.3.1 Transformaciones de semejanza en \( \mathbb{R}^2 \)

En el plano real \( \mathbb{R}^2 \), las transformaciones afines toman una forma sencilla cuando se aplican a un punto arbitrario \( x = (x_1, x_2) \):

\[
\begin{pmatrix}
a & b \\
c & d
\end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix}
+
\begin{pmatrix}
e \\
f
\end{pmatrix}
= Ax + t \equiv x'
\]

\( x' \) es el punto transformado de \( x \), tras aplicar la matriz \( A \) (que representa homotecias, reflexiones y rotaciones), y el vector \( t \), que define una traslación. Estamos interesados en ciertas formas particulares de las transformaciones afines, que son las transformaciones de semejanza.

Una reflexión, \( R \) toma la forma particular

\[
R
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix}
= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix}
\]

Observemos que, simplemente, los puntos de la forma \((x_1, x_2)\) pasan a \((x_1, -x_2)\), bajo la acción de la matriz \( R \).

Una rotación de ángulo \( \theta \) es de la forma

\[
R_\theta
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix}
= \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix}
\]

es decir, el punto \((x_1, x_2)\) se transforma en

\[
(x_1 \cos \theta - x_2 \sin \theta, x_1 \sin \theta + x_2 \cos \theta)
\]

que es simplemente una rotación de ángulo \( \theta \) en sentido antihorario (considere el lector los casos particulares \( x_1 = 0 \) y \( x_2 = 0 \)). Una homotecia \( H_c \) es

\[
H_c
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix}
= \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix}
\begin{pmatrix}
x_1 \\
x_2
\end{pmatrix}
\]

en donde simplemente multiplicamos las dos coordenadas del punto \( x \) por un único factor \( c \). En las aplicaciones que veremos, \( c < 1 \) siempre, y por tanto las llamaremos contracciones. Únicamente estudiaremos los casos en que el factor de contracción \( c \) es el mismo para los dos ejes coordenados. Casos más generales contemplan la posibilidad de que exista un factor \( c_1 \) en el eje de abscisas y otro, \( c_2 \neq c_1 \) en el eje de ordenadas. De hecho, éste es un caso habitual que se da, por ejemplo, en la transformación del panadero descrita en el capítulo sobre caos.

Estas transformaciones, \( R, R_\theta \) y \( H_c \), se pueden componer para ser aplicadas sucesivamente sobre un punto. Solamente veremos composición de \( R_\theta \) y \( H_c \), que escribiremos

\[\text{...}\]

\[\text{...}\]

\[\text{...}\]
\[ H_c R_\theta = \begin{pmatrix} c & 0 \\ 0 & c \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} = \begin{pmatrix} c \cos \theta & -c \sin \theta \\ c \sin \theta & c \cos \theta \end{pmatrix} \]

operación que es conmutativa:

\[ H_c R_\theta = R_\theta H_c \]

así que tanto es contraer primero y rotar después como hacerlo en orden inverso.

Dado que además de ser una operación conmutativa existe el elemento neutro (la matriz identidad, \( I_{2 \times 2} \)), el elemento inverso (siempre que \( c \neq 0 \)) y la composición de rotaciones y contracciones sucesivas se puede expresar como una única operación, estas dos operaciones tienen una estructura de grupo. Este grupo puede ser ampliado con las reflexiones \( R \) y las traslaciones \( t \) \( (z' = x + t) \), para constituir el grupo de las transformaciones de afinidad.

### 3.3.2 Ejemplos

Pasemos ya a describir algunos de los fractales conocidos en términos de transformaciones afines. Dado que trabajaremos con contracciones \( c < 1 \), una única transformación nunca será suficiente para describir un objeto finito: la transformación debe de ser iterada infinitas veces antes de encontrar su punto fijo \(^{12} \), y si ésta es una contracción, el punto fijo en \( \mathbb{R}^2 \) siempre será el origen, \((0, 0)\). Necesitaremos, por tanto, una serie de transformaciones afines, \( \mathcal{W} \)

\[ \mathcal{W} = \{ w_1, w_2, \ldots, w_n \} \]

 tales que, al ser sucesivamente aplicadas a un conjunto cerrado arbitrario de \( \mathbb{R}^2 \) generen el objeto deseado.

Comencemos con un caso sencillo: el conjunto de Cantor. Partamos (como ya hemos hecho anteriormente) del segmento unidad, \([0, 1] \), y busquemos cuáles son las transformaciones que proporcionan los dos segmentos \([0, 1/3] \) y \([2/3, 1] \) correspondientes a la primera iteración. Estos dos segmentos son idénticos al original, pero \( 1/3 \) menores. El factor de escala, \( c \), será por tanto \( c = 1/3 \). Su orientación respecto del segmento inicial no ha variado, por tanto el ángulo de rotación es \( \theta = 0 \). Por otra parte, ahora tenemos dos segmentos, uno situado en el origen, que por tanto no necesita ser trasladado respecto del segmento original, y otro con origen en el punto \( P = 2/3 \), que requiere por tanto una traslación \( t = (2/3, 0) \) en \( \mathbb{R}^2 \). Las dos transformaciones que definen esta iteración son, pues

\[ w_1(x) = \begin{pmatrix} 1/3 \\ 0 \end{pmatrix} x + \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \quad w_2(x) = \begin{pmatrix} 2/3 \\ 0 \end{pmatrix} x + \begin{pmatrix} 0 \\ 3 \end{pmatrix} \]

En este caso sencillo no es necesario trabajar en \( \mathbb{R}^2 \), puesto que el conjunto de Cantor está contenido en \( \mathbb{R}^1 \). Por tanto, podemos reducir las transformaciones a

\[ w_1 = \frac{1}{3} x; \quad w_2 = \frac{1}{3} x + \frac{2}{3} \]

Si aplicamos de nuevo \((w_1 \circ w_2)\) al conjunto resultante, \([0, 1/3] \cup [2/3, 1]\), obtenemos

\[
(w_1 \circ w_2) \left\{ \left[ \begin{array}{c} 0 \\ \frac{1}{3} \end{array} \right] \cup \left[ \begin{array}{c} \frac{2}{3} \\ 1 \end{array} \right] \right\} = \left[ \begin{array}{c} 0 \\ \frac{1}{9} \end{array} \right] \cup \left[ \begin{array}{c} \frac{2}{9} \\ \frac{3}{9} \end{array} \right] \cup \left[ \begin{array}{c} \frac{6}{9} \\ \frac{7}{9} \end{array} \right] \cup \left[ \begin{array}{c} \frac{8}{9} \\ 1 \end{array} \right]
\]

\(^{12}\)La idea de punto fijo es intuitiva: el punto fijo \( x^* \) de una función \( f(x) \) será aquel que no varíe bajo la aplicación de la función: \( f(x^*) = x^* \). Por ejemplo, el punto fijo (en este caso los puntos fijos) de la función \( f(x) = x^2 + x - 1 \) es la solución de \( (x^*)^2 + x^* - 1 = x^* \) y \( x^* = \pm 1 \).
Las aplicaciones \( w_1 \) y \( w_2 \), iteradas infinitas veces, proporcionan el conjunto de Cantor. Este conjunto queda en consecuencia completamente descrito por únicamente tres parámetros: \( c = 1/3 \), \( t_1 = 0 \), \( t_2 = 2/3 \), correspondientes al factor de contracción y a las dos traslaciones.

Describamos otro ejemplo, ya en \( \mathbb{R}^2 \), algo más complicado: la curva de von Koch. En la primera iteración, el conjunto cerrado inicial se transforma en cuatro, según ciertas transformaciones que contraen en un factor \( c = 1/3 \), todas por igual, y además:

- \( w_1 \) sitúa el conjunto en el origen, \( x_1 = 0, y_1 = 0 \), sin cambiar su orientación, por tanto \( t_1 = (0, 0) \) y \( \theta_1 = 0 \).
- \( w_2 \) rota \( 60^\circ \) y sitúa el conjunto en \( x_2 = 1/3, y_2 = 0 \), por tanto \( t_2 = (1/3, 0) \) y \( \theta_2 = 60^\circ \).
- \( w_3 \) rota \( -60^\circ \) y sitúa el conjunto en \( x_3 = 1/2, y_3 = \sqrt{3}/6 \), por tanto \( t_3 = (1/2, \sqrt{3}/6) \) y \( \theta_3 = 300^\circ \).
- \( w_4 \) no rota y sitúa el conjunto en \( x_4 = 2/3, y_4 = 0 \), por tanto \( t_4 = (2/3, 0) \) y \( \theta_4 = 0 \).

Así que la curva de von Koch está definida por

\[
\begin{align*}
  w_1(x) &= \left( \frac{1}{3}, 0 \right) x + \left( 0, 0 \right) ; \quad w_2(x) = \left( \frac{1}{6}, \frac{-\sqrt{3}}{6} \right) x + \left( \frac{1}{2}, 0 \right) \\
  w_3(x) &= \left( \frac{1}{6}, \frac{\sqrt{3}}{6} \right) x + \left( \frac{1}{2}, \frac{\sqrt{3}}{6} \right) ; \quad w_4(x) = \left( \frac{1}{3}, 0 \right) x + \left( \frac{2}{3}, 0 \right)
\end{align*}
\]

según la matriz composición de homotecias y rotaciones descrita en la sección anterior. El número de parámetros necesarios en este caso es de 13: un factor de escala \( c \) y cuatro transformaciones \( w_i \), en donde cada una necesita ser especificada, por un ángulo \( \theta_i \) y dos valores para cada trasladación, \( t_i \).

De forma compacta se designa por \( W(A) \) el conjunto de las transformaciones que iterativamente se aplican sobre el conjunto cerrado inicial \( A \) para generar el fractal deseado:

\[
W(A) = \bigcup_{n=1}^{n} w_n(A)
\]

Se puede demostrar que una transformación contractiva (como lo es \( W \)) en el espacio métrico completo en el que trabajamos (\( \mathbb{R}^2 \) con la distancia euclidiana) tiene un único punto fijo, y por tanto un único objeto aparece como límite de las iteraciones. ¿Podría ahora el lector determinar cuál es el conjunto de transformaciones que produce la alfombra de Sierpiński o la curva de Peano?

### 3.3.3 El teorema del Collage

El paso siguiente en la construcción de objetos fractales a partir de la transformación genérica \( W(A) \) es la resolución del problema inverso. Consideremos un objeto real, por ejemplo la hoja de un árbol, o el árbol mismo, que de alguna forma pueda ser visto como un conjunto autosimilar. ¿Seríamos capaces de construir \( W(A) \) de forma que su iteración sobre un conjunto cerrado de \( \mathbb{R}^2 \) (que por sencillez podemos considerar un único punto) produzca la imagen real de forma más o menos aproximada? La respuesta es sí, y la solución viene dada a través del llamado teorema del collage, que puede ser enunciado de la forma siguiente:

Realicen diversas copias de diferentes tamaños del conjunto que se pretende reproducir. Recúbrase con ellas el conjunto, tan aproximadamente como sea posible. El tamaño de las copias respecto del conjunto original da el factor de contracción de cada aplicación. El número de copias
requeridas es el número de transformaciones \( w_i \) necesarias. El ángulo de rotación \( \theta_i \) de cada una viene dado por su orientación respecto del original. Su colocación exacta \( (t_i) \) se determina escogiendo unos ejes y, por ejemplo, situando nuestro objeto en el recuadro \([0,1] \times [0,1]\).

Un ejemplo de la forma de proceder y de las aplicaciones obtenidas \( \{w_i\} \), así como del conjunto resultante de tomar las aplicaciones y generar a partir del punto \((0,0)\) el conjunto inicial se da a continuación.

Consideremos el grabado de una hoja representado en la figura 3.15, arriba a la izquierda. La figura de la derecha es una reproducción de la inicial en donde se han utilizado cinco copias de distintos tamaños. Los factores de escala vienen en este caso directamente determinados por el factor que daba la "fotocopiadora" utilizada. Si escogemos dos puntos de la imagen inicial que permitan trazar una recta guía, el ángulo que forme esta recta con cada una de sus réplicas menores proporcionará los ángulos de rotación de las transformaciones. Si se lleva a cabo este proceso, las transformaciones que se obtienen son las siguientes:

- \( w_1: \ r_1 = 0.75, \ \theta_1 = 0^\circ \)
- \( w_2: \ r_2 = 0.60, \ \theta_2 = 0^\circ \)
- \( w_3: \ r_3 = 0.40, \ \theta_3 = 0^\circ \)
- \( w_4: \ r_4 = 0.37, \ \theta_4 = 61^\circ \)
- \( w_5: \ r_5 = 0.27, \ \theta_5 = -50^\circ \)

Dados los valores de los factores de contracción de cada una de las aplicaciones generadoras del objeto fractal, se puede calcular de forma exacta la dimensión fractal de éste, siempre y cuando no haya zonas solapantes en el conjunto, sin necesidad de recurrir a métodos de recubrimiento como el "box-counting". El valor de la dimensión fractal para un objeto generado por una colección de \( k \) aplicaciones con factores de contracción \( \{r_1, \ldots, r_k\} \) se obtiene resolviendo la ecuación trascendente

\[
\sum_{i=1}^{k} r_i^D = 1
\]

donde \( D \) es la dimensión fractal. La expresión no es aplicable, por ejemplo, al dibujo en forma de hoja de la figura 3.15, ya que existen zonas de solapamiento, procedentes de distintas aplicaciones. Sí podría ser aplicado a los dos objetos de la figura 3.16 (IFS y ?), ya que, en éstas, no hay intersección entre las transformaciones afines.

Se puede adivinar las grandes posibilidades que el método ofrece, una vez que el usuario se ha familiarizado con él. Las imágenes representadas en la figura 3.16 ponen a prueba el grado de comprensión alcanzado en esta sección. ¿Puede el lector descubrir las aplicaciones que generan estos dos últimos fractales?

### 3.4 Los conjuntos de Julia y de Mandelbrot

#### 3.4.1 Algebra elemental de los números complejos, \( \mathbb{C} \)

Antes de iniciar el estudio de los conjuntos de Julia, como prólogo a la descripción del conjunto de Mandelbrot, es necesario poder manejar con cierta comodidad los números complejos, ya que los conjuntos que van a ser descritos se encuentran definidos en el plano complejo \( \mathbb{C} \), espacio de dos dimensiones. Empezamos, pues, esta sección con un repaso del álgebra elemental y de la notación usual de los números complejos.

Un número complejo, formado por el par \((x, y)\) con \(x, y \in \mathbb{R}\) es de la forma
Figura 3.15: Arriba a la izquierda, grabado de una hoja. A la derecha, arriba, reproducción de la figura inicial utilizando cinco réplicas de distintos tamaños. Su colocación, tamaño relativo y orientación proporcionan los parámetros de las transformaciones afines que reproducirán el dibujo original. Abajo, a la izquierda, se representan las aplicaciones afines generadoras. Los rectángulos son las imágenes del rectángulo mayor exterior bajo una iteración de cada una de las cinco aplicaciones. A la derecha, abajo, se puede ver la imagen obtenida tras unas decenas de iteraciones.
Figura 3.16: Para generar estas dos imágenes ha sido necesario permitir la existencia de dos factores de contracción diferentes en las direcciones x e y, como se puede apreciar de inmediato. La primera figura (IFS) requiere 10 transformaciones de afinidad, y 7 son suficientes para generar la segunda (?).

\[ z = x + yi \]

donde \( x \) se denomina parte real, \( \Re(z) = x \), e y es la parte imaginaria, \( \Im(z) = y \), que siempre va acompañada de la unidad imaginaria \( i \). Esta se define como:

\[ i = \sqrt{-1} \quad \Rightarrow \quad i^2 = -1 \]

Su introducción en las matemáticas fue debida al deseo de proporcionar una solución formal a ecuaciones en las que aparecían raíces pares de números negativos. Con la introducción de \( i \) y el álgebra de los números complejos se avanzó en el estudio teórico de expresiones del tipo \( z^2 + 1 = 0 \), a la vez que se abrieron campos nuevos de investigación, como el de la variable compleja o el que nos ocupa, el de los conjuntos fractales en el plano complejo.

Dados dos números complejos, \( z = x + yi \) y \( w = u + vi \), su suma es la cantidad

\[ z + w = (x + yi) + (u + vi) = (x + u) + (y + v)i \]

es decir, las partes reales se suman para dar la parte real y las imaginarias para proporcionar la parte imaginaria. El producto de \( z \) y \( w \) es

\[ z \cdot w = (x + yi)(u + vi) = xu + yui + xvi + yvi^2 = (xu - yv) + (xv + uy)i \]

donde se aplica la definición de \( i \). El cociente entre \( z \) y \( w \) se obtiene como

\[ \frac{z}{w} = \frac{x + yi}{u + vi} = \frac{x + yi}{u + vi} \cdot \frac{u - vi}{u - vi} = \frac{(ux + yv) + (yu - vx)i}{u^2 + v^2} = \frac{ux + yv}{u^2 + v^2} + \frac{yu - vx}{u^2 + v^2}i \]

Se multiplica el denominador por su conjugado (que consiste únicamente en cambiar el signo a la parte imaginaria) para eliminar \( i \) del denominador, proporcionando así una parte real y una compleja separadas como resultado final. La interpretación habitual de los números complejos es la
Figura 3.17: Representación geométrica de los números complejos. En el eje $x$ se representa la parte real, y en el eje $y$ la imaginaria. Se indica la notación en coordenadas cartesianas, $(x, y)$ y en coordenadas polares, $(r, \phi)$.

de puntos en un plano. En el eje $x$ se representa la parte real y en el eje $y$ la imaginaria. Es también frecuente pensar en el número complejo como en un vector situado en el origen de coordenadas. Una notación alternativa y muy útil de los números complejos es su forma trigonométrica. Utilizando la transformación de coordenadas cartesianas a coordenadas polares

$$z = x + yi$$

y la transformación inversa

$$z = r \cos(\phi), \quad y = r \sin(\phi)$$

se puede escribir

$$z = r(\cos \phi + i \sin \phi)$$

donde $r$ es el módulo, que da la distancia al origen de coordenadas, y $\phi$ es el ángulo que el "vector" forma con el eje $x$, llamado argumento de $z$, $\phi = \text{arg}(z)$.

Esta representación se puede transformar aún, por medio de la notación de Euler (o forma polar), que establece la equivalencia

$$e^{i\phi} = \cos \phi + i \sin \phi$$

La igualdad se puede verificar desarrollando las tres funciones en serie de Taylor. Por tanto, el número complejo $z = x + yi$ se puede denotar por

$$z = re^{i\phi}$$

con las definiciones ya dadas para $r$ y $\phi$. Esta notación simplifica sobremanera el producto y la división, y permite un cálculo rápido de las raíces $n$-ésimas de un número complejo. El producto, dados $z = re^{i\phi}$ y $w = se^{i\psi}$ se obtiene como

$$zw = rs e^{i(\phi + \psi)}$$

y el cociente


\[
\frac{z}{w} = \frac{r}{s} e^{i(\phi - \psi)}
\]

Las \(n\) soluciones \(\{a_n\}\) de la raíz \(n\)-ésima de un número complejo \(z\) se obtienen también fácilmente de la forma siguiente:

\[
a_k = \sqrt[n]{r} \exp \left\{ i \left( \phi + 2\pi \frac{k}{n} \right) \right\}
\]
donde \(k = 1, 2, \ldots, n\). Es decir,

\[
a_1 = \sqrt[n]{r} e^{i(\phi + 2\pi/n)} \quad a_2 = \sqrt[n]{r} e^{i(\phi + 4\pi/n)} \quad a_3 = \sqrt[n]{r} e^{i(\phi + 6\pi/n)} \quad \ldots \quad a_n = \sqrt[n]{r} e^{i\phi}
\]

### 3.4.2 Los conjuntos de Julia

Gaston Julia (1893-1978) fue uno de los fundadores de la actual teoría de sistemas dinámicos. Con tan sólo 25 años publicó su trabajo “Mémoire sur l'itération des fonctions rationelles”\(^{13}\), que le valió reconocimiento internacional en la década de 1920. La mayoría de sus estudios fueron llevados a cabo en un hospital, mientras se encontraba convaleciente de las heridas que, durante la Primera Guerra Mundial, le produjeron la pérdida de la nariz.

Julia trabajaba con los polinomios definidos en el plano complejo. Daremos seguidamente las ideas que proporcionan la definición de los llamados conjuntos de Julia. Consideremos el polinomio de segundo grado más sencillo en el plano complejo,

\[
f(z) = z^2
\]
y estudiamos el comportamiento de todos los puntos del plano complejo, \(w_0 = u + vi\) bajo infinitas iteraciones de este polinomio. Por sencillez, escribamos \(w_0 = re^{i\phi}\), en general, y observemos que las iteraciones sucesivas serán de la forma

\[
w_0 = re^{i\phi}, \quad w_1 = w_0^2 = r^2e^{2i\phi}, \quad w_2 = w_1^2 = w_0^4 = r^4e^{4i\phi}, \ldots, \quad w_k = (re^{i\phi})^{2k}
\]

Es sencillo estudiar el comportamiento del punto \(w_0\) cuando \(k \to \infty\). Observemos que, si \(r < 1\), la potencia \(r^{2k}\) tenderá a cero para \(k \to \infty\), y por tanto, independientemente del ángulo \(\phi\), todos los puntos que inicialmente se encontraban en el interior del círculo de radio unidad tienen el origen como atractor. Por otra parte, si \(|w_0| = r > 1\), cuando \(k \to \infty\) estos puntos crecen sin límite, y acaban escapando al infinito. Este es el conjunto de los puntos de escape, que denotaremos por

\[
\mathcal{E} = \{w_0; \quad |w_0| \to \infty \text{ si } k \to \infty\}
\]

El complementario de \(\mathcal{E}\) es el conjunto de los puntos prisioneros, es decir, los que mantienen un módulo finito cuando \(k \to \infty\):

\[
\mathcal{P} = \{w_0; \quad w_0 \notin \mathcal{E}\}
\]

Este conjunto \(\mathcal{P}\) contiene tanto los puntos cuyo atractor es el origen (\(|w_0| < 1\)) como aquellos que constituyen su frontera, \(\mathcal{F}_P\), que siempre permanecen en ella y que están caracterizados por tener módulo unidad, \(r = 1\). El conjunto frontera de \(\mathcal{P}\), \(\mathcal{F}_P\), es el conjunto de Julia para el polinomio \(f(z) = z^2\). Los conjuntos de Julia habituales son objetos fractales, y provienen de polinomios de la forma \(f(z) = z^2 + c\), en donde \(c\) es un número complejo arbitrario en cada caso y responsable de los infinitos conjuntos de Julia existentes (uno para cada valor de \(c\)). El caso

---

\(^{13}\)Gaston Julia, “Mémoire sur...”, Jour. de Math. Pure et Appli. 8 (1918), 47-245.
Figura 3.18: Cálculo numérico de conjuntos de Julia a partir de un conjunto cerrado inicial. La secuencia de figuras es de izquierda a derecha, y de arriba abajo. Inicialmente, se toma la figura en forma de “L” y se itera la aplicación inversa de un polinomio complejo de grado 2, con \( c = -0.5 + 0.3i \). Se representan las cuatro primeras iteraciones y el conjunto de Julia final. Este mismo conjunto (recuadro 6) es el conjunto cerrado inicial para la creación del conjunto de Julia correspondiente a \( c = 0.5i \). Se han representado los dos primeros pasos y el conjunto final correspondiente (última figura).

anterior, que nos ha servido para definir el conjunto de los puntos prisioneros y de los puntos de escape (de ahora en adelante \( \mathcal{P}_c \) y \( \mathcal{E}_c \), dependientes del parámetro \( c \)) corresponde al caso \( c = 0 \).

Podría parecer que la obtención de los puntos pertenecientes a un cierto conjunto de Julia no es excesivamente complicada. Sin embargo, no es así. Excepto en el caso trivial \( c = 0 \), es imposible obtener expresiones compactas que designen \( \mathcal{P}_c \).

Una primera aproximación numérica al conjunto consistiría en dividir el plano complejo en una red discreta de puntos y calcular para cada uno de ellos un cierto número de iteraciones, a fin de determinar la convergencia o la divergencia de dicho punto. Este método es terriblemente tedioso, y requeriríamos gran número de iteraciones para llegar a una imagen más o menos cercana al conjunto final. Existe un método mucho más sencillo, que aprovecha la invertibilidad del polinomio \( f(z) \). En efecto, dado \( f(z) = z^2 + c \), consideremos la aplicación inversa,

\[
f^{-1}(z) = \pm \sqrt{z - c}
\]

Si, dado cualquier conjunto cerrado de \( \mathbb{C} \) realizamos sucesivas iteraciones de \( f^{-1}(z) \), el conjunto de Julia aparecerá como punto fijo de esta transformación. Por otra parte, se puede demostrar que todos los puntos \( w_0 \) tales que su módulo \( |w_0| \) sea mayor que el valor \( r(c) \), donde
Figura 3.19: Cálculo de dos conjuntos de Julia utilizando las antíimágenes sucesivas de la circunferencia de radio 2. En cada antíimagen se produce un cambio de color. Se han representado las cuatro primeras iteraciones y en el interior el conjunto de Julia final correspondiente (tras unas 25 iteraciones, que son suficientes en estos casos para dar una imagen indistinguible del conjunto límite a esta escala). A la derecha, \( c = -0.6 + 0.4i \), y a la izquierda, \( c = -1.3 \).

\[
r(c) = \max(|c|, 2)
\]

pertenecen a \( E_c \), y por tanto no están en el conjunto de Julia. Según esta condición, las iteraciones de \( f^{-1} \) que permiten encontrar el correspondiente conjunto de Julia se realizan (de forma muy conveniente) a partir del conjunto compacto de \( C \) formado por el círculo de radio \( r(c) \). En pocas iteraciones, unas 10 ó 15, la imagen del conjunto de Julia aparece definida con considerable precisión.

El conjunto de Julia correspondiente a \( f(z) = z^2 + c \) es invariante bajo la acción de \( f(z) \). Por tanto, los puntos pertenecientes al conjunto deben de seguir perteneciendo a él tras aplicar la transformación \( w_0 \rightarrow w_0^2 + c \). El conjunto de Julia posee puntos de todas las periodicidades y puntos que siguen dinámica caótica sobre \( F_{F_2} \). Esto implica la existencia de sensibilidad a las condiciones iniciales, en el sentido siguiente: la iteración aplicada a los puntos incluidos en una pequeña porción del conjunto de Julia provoca que éstos acaben repartidos por todo el conjunto tras unas pocas iteraciones.

### 3.4.3 El conjunto de Mandelbrot

Benoit B. Mandelbrot nació en Polonia en 1924, pero a los 12 años se trasladó a Francia con su familia. Su tío Szolem Mandelbrojt, profesor del Collège de France, lo introdujo en los trabajos de Julia y Fatou. B. B. Mandelbrot era un apasionado de la geometría, aun cuando durante su juventud ésta parecía estar en franca regresión. Recuerdenos que el trabajo de Julia sobre los polinomios de grado dos fue publicado en 1918, pero en aquellos años, sin ordenadores, la visualización de los puntos fijos de las aplicaciones, de los conjuntos de Julia, era imposible. Un
Figura 3.20: Ejemplo de la sensibilidad a las condiciones iniciales en el conjunto de Julia correspondiente al caso trivial $c = 0$. Inicialmente, se escoge un pequeño grupo de puntos, comprendidos entre 2 y 3 grados de arco. Tras tan sólo siete iteraciones de la función $z \rightarrow z^2 + c$, pasan a ocupar la zona del conjunto comprendida entre las líneas punteadas.

primer esquema del aspecto que éstos deberían presentar se dio en 1925 \(^{14}\), pero la imagen aún era sumamente tosca. Recuerden también que los conjuntos de Julia podían ser conexos o desconexos, dependiendo del valor del parámetro $c$, y debido a su carácter autosimilar. Mandelbrot, entre 1979 y 1980, se propuso la clasificación rigurosa de los conjuntos según $c$, es decir, calculó los valores de $c$ que producían conjuntos conexos y los representó \(^{15}\). En principio, podía suceder que la zona del plano complejo que contuviese los valores de $c$ responsables de conjuntos de Julia conexos tuviese una forma geométrica bien definida, podría ser un círculo o un polígono sencillo, o bien una colección de piezas desconectadas repartidas por el plano. No obstante, el citado grupo de puntos resultó ser de exquisita geometría y extrema complicación. Se le ha calificado como "el objeto más bello y complicado jamás visto". La frontera del conjunto de Mandelbrot posee autosimilaridad a todas las escalas, estructura infinita. La autosimilaridad no es estricta: podemos obtener copias ligeramente deformadas del conjunto original a todo lo largo de la frontera, copias de cualquier tamaño, pero no exactas (en el sentido de que exista una transformación isométrica que convierta una en otra).

Cualquier valor de $c$ que escojamos en el interior del conjunto proporcionará un conjunto de Julia conexo, y cualquier valor del exterior producirá uno desconexo. En consecuencia, serán los valores de $c$ cercanos a la frontera los que producirán los conjuntos de Julia con mayor estructura (asociada en este caso a una mayor dimensión fractal). Parecía, pues, que B. B. Mandelbrot había conseguido la unificación de su pasión (la geometría) con un campo nuevo de las matemáticas que le iba a proporcionar trabajo y distracción durante toda su vida (y hasta el momento, así ha sido).

La descripción del conjunto dada hasta ahora permite una primera definición rigurosa de los


Figura 3.21: El conjunto de Mandelbrot en el plano complejo. Se ha representado la circunferencia de radio 2 (que contiene totalmente al conjunto) y se han realizado divisiones cada 0.5 unidades, para facilitar la identificación de la zona ocupada.

puntos que lo constituyen. Denotemos el conjunto de Mandelbrot por $\mathcal{M}$ (nomenclatura usual):

$$\{M = c \in C : P_c \text{ es conexo}\}$$

Recordemos que $P_c$ representaba el conjunto prisionero para el parámetro $c$ (llamado a veces conjunto de Fatou), y su frontera es el conjunto de Julia para $c$. Si bien la definición dada de $\mathcal{M}$ lo determina completamente y es rigurosa, el cálculo de los puntos $c \in \mathcal{M}$ es extremadamente largo, puesto que pasa por establecer la conectividad de todos y cada uno de los conjuntos de Julia existentes. Afortunadamente, algunas propiedades de $\mathcal{M}$ ayudan en el cálculo. Por ejemplo, en 1982 fue posible demostrar que $\mathcal{M}$ es un conjunto conexo $^{16}$, con lo cual es suficiente con preocuparse de su frontera. La empresa sería ahora fácil de no ser por la infinita sinuosidad del conjunto de Mandelbrot: resulta imposible "seguir" la frontera, ya que ésta tiene una longitud infinita. Imagine el lector cuán complejo es el conjunto que hasta 1991 no fue posible determinar su dimensión de Hausdorff de forma rigurosa, y cuando se consiguió, resultó ser $D_H = 2!^{17}$

Un argumento de carácter geométrico permite una definición más sencilla (operativamente) de $\mathcal{M}$. Consideremos un conjunto de Julia desconexo. Cuando éste se calcula utilizando las antíimagines sucesivas de la circunferencia de radio $r = 2$, en cierta iteración debe de aparecer la desconexión, que se manifiesta por una figura en forma de "8", es decir, una figura de tipo circular.


$^{17}$M. Shishikura "The Hausdorff dimension of the boundary of the Mandelbrot set and Julia sets", *SUNY Stony Brook, Institute for Mathematical Sciences, Preprint # 1991/7.*
Figura 3.22: Dos conjuntos de Julia desconexos en los que aparece la característica figura en forma de “8” en una de las antiimágenes. En un caso la desconexión aparece en la tercera iteración (para $c = -1 + i$), y en otro en la cuarta (para $c = -1.1i$).

que incluye dos rosquillas de este tipo, separadas. Observemos que los conjuntos de Julia presentan una simetría de rotación según un ángulo de 180$^\circ$, debida a que los puntos del conjunto son las soluciones de una raíz cuadrada, como se ha visto. La existencia de esta iteración desconexa, juntamente con la simetría de rotación, implica automáticamente que el punto $z_0 = 0 + 0i$, el origen, no pertenece a ningún conjunto de Julia desconexo, y por tanto podemos establecer que, si el origen pertenece a $E_c$ (conjunto de los puntos de escape para un parámetro $c$ dado), el conjunto de Julia será desconexo, y $c$ no pertenecerá a $M$.

Esta propiedad proporciona la siguiente definición alternativa de $M$, que ya fue de hecho utilizada en 1979 por Mandelbrot para el cálculo del conjunto:

$$M = \{c \in \mathbb{C} : c \to c^2 + c \to (c^2 + c)^2 + c \to \ldots \text{ es finito}\}$$

Establecer el carácter no divergente de la órbita crítica $0 \to c \to c^2 + c \ldots$ no es tampoco inmediato, ya que implica el cálculo de un límite infinito. Usualmente se calcula un cierto número de iteraciones, del orden de $10^2$, y esto da una aproximación bastante buena del conjunto. Si la definición de la imagen o la precisión del cálculo requerido es mayor, mayor también será el número de iteraciones que deben de ser realizadas. También se pueden calcular aproximaciones sucesivas de $M$: sólo es necesario considerar el conjunto de valores de $c$ que escapan, bajo la iteración anterior, en un paso (de iteración), en dos pasos, en tres... de la circunferencia de radio $r = 2$, y progresivamente se van produciendo imágenes cada vez más cercanas a $M$.

En la figura 3.21 se representa el conjunto de los puntos que escapan tras 25 iteraciones (de $z_0 \to z_0^2 + c$, comenzando con $z_0 = 0 + 0i$), y la imagen es una muy buena aproximación de $M$, prácticamente indistinguible a esta escala del representado. $M$ se extiende desde $-2$ a 0.75 en el eje real, de forma exacta, y desde $-1.25$ hasta 1.25 en el eje imaginario, de forma aproximada, como se puede ver en la figura. Todos los valores de $c$ que se encuentran en su interior (sin considerar la frontera) producen conjuntos de Julia conexos con interior no vacío, es decir, en el conjunto de Julia, la eliminación de la frontera no produce su desconexión. Los valores de $c$ localizados en la frontera de $M$ producen conjuntos de Julia que en algún lugar mantienen la conexión mediante un
único punto. Los valores de \( c \) en el exterior de \( M \) producen conjuntos de Julia desconexos, tanto más cuanto mayor es la distancia a \( M \).

Observando con atención el conjunto de Mandelbrot se puede descubrir la existencia de zonas con interior que están conectadas mediante un único punto al cuerpo principal de \( M \). Cuando tomamos un valor de \( c \) en el interior de una de estas zonas, el conjunto de Julia correspondiente reproduce esta estructura: existen zonas de interior no vacío conectadas a través de puntos. En particular, en la “aguja” de \( M \) existen muchas de estas zonas, visibles a la escala en la que \( M \) se ve completo, y su número aumenta infinitamente con el aumento de la imagen. En particular, el “cuello” que une el cuerpo principal con el apéndice mayor de \( M \) (el círculo de centro \(-1\) y radio \(0.25\)) consta de un solo punto, \( u_0 = -0.75 + 0 i \).

Es posible aún una clasificación mucho más exhaustiva de los conjuntos de Julia atendiendo a la estructura de \( M \). En particular, se pueden establecer órbitas de período \( p \) para la dinámica de los puntos sobre el conjunto de Julia, y cada período puede ser asociado a extremidades o apéndices del cuerpo principal de \( M \).

### 3.5 Fractales no deterministas

Hemos considerado en las secciones anteriores un gran número de fractales que podemos denominar *deterministas*, puesto que existe una regla de construcción del objeto que permite su reproducción *exacta* tantas veces como se desee. Los fractales vistos hasta ahora son interesantes principalmente como objetos matemáticos, pero la relación caos-fractales-sistemas físicos aún no ha aparecido de forma clara. Hemos reservado para esta última sección los fractales *no deterministas*, que están relacionados con procesos físicos y con su modelización, y que siempre poseen cierto grado de aleatoriedad en su construcción. El hecho de que sea un proceso físico el responsable de la generación de una estructura con distribución espacial (el fractal no determinista) provoca que, si bien cada objeto generado será único, debido a los términos aleatorios, existirán propiedades comunes a todos ellos, originadas por el proceso físico, que es el representante de un orden superior (véase el capítulo sobre fenómenos críticos). Cantidades como la dimensión fractal, la densidad de masa, la conectividad o el espectro multifractal (que será definido más adelante) se revelarán como invariantes de los fractales generados mediante un mismo sistema. Estas magnitudes comunes se deben en ocasiones a la existencia de cierto tipo de universalidad, de propiedades que son comunes a diferentes sistemas y que representan un nivel de organización superior 18.

Otro punto de interés en los fractales físicos es la existencia de escalas límite para la presencia de autosimilaridad. En los fractales deterministas vistos, descritos por una ecuación (o varias), no hay ningún problema en calcular la invariancia del objeto, desde lo infinitamente pequeño hasta lo infinitamente grande (de hecho, es suficiente que exista el límite para uno de los dos extremos y nos situemos a medio camino), utilizando su definición misma. Cuando tratamos con sistemas físicos, se hace evidente la imposibilidad de alcanzar estos dos límites. Por una parte, la mayor escala de autosimilaridad posible es la determinada por el tamaño del sistema, por el espacio físico que ocupa. Con frecuencia se simulan ciertos sistemas sobre redes con \( N \times N \) elementos. En este caso, la escala dada por la longitud \( l_M = N \) es una escala crítica de corte para la existencia de fractalidad. Por otra parte la escala menor, \( l_m \) estará acotada por la escala de definició de la interacción elemental. Supongamos que utilizamos un sistema con un grupo de hormigas que interaccionan. Entonces la escala mínima es la propia del individuo, la hormiga, y las partes de ésta de ninguna manera podrán poseer la invariancia de escala del sistema. Por último, encontraremos sistemas que posean alguna otra escala característica entre \( l_M \) y \( l_m \), y éstas nos darán pistas

18 Véase el capítulo 7.
Figura 3.23: Diversos conjuntos de Julia. Unos se han representado con las antiimágenes sucesivas de la circunferencia de radio 2, y otros sin ellas. El lector puede utilizar el conjunto de Mandelbrot para clasificarlos y comprobar su conectividad. De izquierda a derecha y de arriba abajo: \( c = -1.47 \), \( c = -0.098555 - 0.7879i \), \( c = -1.25 + 0.25i \), \( c = i \).
sobre los fenómenos físicos subyacentes. Esta escala (o escala) intermedia puede manifestarse, por ejemplo, en un cambio brusco en la dimensión fractal del sistema, como se ha observado en los arrecifes de coral, entre otros casos.

### 3.5.1 Multifractales

Introducimos ahora una nueva herramienta analítica con clara interpretación física: el cálculo del espectro multifractal de un objeto autosimilar. La definición de multifractal constituye una extensíon de la idea de fractal. Data de 1974, y se debe de nuevo a B. B. Mandelbrot, quien introdujo el concepto a fin de proporcionar una descripción más precisa del fenómeno de la turbulencia.

Comencemos con una visión intuitiva de la multifractalidad. Imaginemos para ello un fractal, en el que se pueda reconocer la existencia de la autosimilaridad, como el objeto de la figura 8.13, pero que no esté definido por ninguna regla sencilla de construcción a todas las escalas.

Requerimos en este punto la ayuda del método del “box-counting”, descrito en la sección 3.1.1. Imaginemos que recubrimos el conjunto con cajas de distintos tamaños (δ variable, como el método requiere), e intentamos conseguir la representación de una recta en la gráfica de \( \ln(N(\delta)) \) frente a \( \ln(\delta) \). En ocasiones, los puntos no se ajustan a una recta de forma aproximada, sino que aparentemente la dimensión de “box-counting” (la pendiente de la gráfica) varía con la escala. De aquí se puede deducir inmediatamente que la medida del conjunto no es autosimilar. En la inmensa mayoría de los casos, esto implica que el conjunto que estamos intentando describir no puede ser caracterizado por una única dimensión fractal. Si la medida está desigualmente repartida, con zonas más y menos densas en el mismo conjunto, necesitaremos valores diferentes de la dimensión fractal para caracterizar cada una de estas regiones.

Estaremos pues de acuerdo en reconocer que habrá sistemas que no podrán ser caracterizados por un único número, la dimensión fractal (calculada mediante “box-counting”), puesto que cualquier valor para \( D_B \) será el obtenido a cierta escala, y diferente del que se calcule a una escala distinta. La solución consiste, precisamente, en no dar una única dimensión fractal, sino todo un espectro continuo, que caracterizará con infinita mayor precisión el sistema. A los objetos fractales que necesitan ser descritos por este conjunto infinito de exponentes se les denomina multifractales. Pasemos a su caracterización rigurosa.

El primer paso consiste en la definición de una densidad de probabilidad sobre el sistema. Supongamos (siempre pensando, ahora, en la distribución espacial de cierta cantidad) que el sistema se divide en \( m \) porciones \( \{a_1, a_2, \ldots, a_m\} \), y que a cada una se le asocia una probabilidad \( p_i \), \( i = 1, 2, \ldots, m \). Para fijar ideas, vamos a concentrarnos únicamente en medidas de masa: si el sistema tiene un volumen \( V \) (en \( n \) dimensiones) y cada una de sus \( m \) partes tiene un volumen \( v_i \), entonces la probabilidad (medida) de cada una de estas partes es:

\[
p_i = \frac{v_i}{V}
\]

y lo único que representamos es la proporción de volumen contenido en cada parte respecto del tamaño total del sistema. En la bibliografía del final del capítulo se pueden encontrar definiciones más generales. Es fácil considerar una imagen geométrica de esta probabilidad: es, precisamente, la probabilidad de que al escoger al azar un punto del sistema, este pertenecer a la porción \( a_i \).

Escojamos ahora una escala en el sistema, \( l \), que nos permita realizar \( l^{-2} \) divisiones (al igual que dividimos en cajas para calcular la dimensión fractal, normalizaremos el sistema a tamaño unidad, y en principio debemos realizar una regresión lineal para calcular cada una de las dimensiones que van a definirse, con lo cual \( l \) será variable). Cada división lleva asociada una probabilidad \( p_l \), pero puede ser que alguna de ellas tenga probabilidad nula, debido a que no recubre ninguna parte
del sistema. En la suma que ahora definiremos, sólo cuentan las cajas con \( p_i \neq 0 \). Consideramos la cantidad

\[
\chi(q) = \sum_{i=1}^{m} p_i^q
\]

que permite definir las llamadas \text{dimensiones de correlación}.

\[
D(q) = \lim_{l \to 0} \left[ \frac{1}{q-1} \frac{\ln \chi(q)}{\ln l} \right]
\]

\( D_0 \) es simplemente la dimensión fractal (de “box-counting”, por la forma en que se está definiendo el cálculo) del sistema. Efectivamente, cuando \( q = 0 \) obtenemos

\[
D(0) = -\lim_{l \to 0} \frac{\ln \sum_{i=1}^{m} p_i^0}{\ln l}
\]

pero

\[
\sum_{i=1}^{m} p_i^0 = \sum_{i=1}^{m} 1 = m
\]

que es en este caso el número de cajas necesarias para recubrir el sistema.

\( D_1 \) se denomina \text{dimensión de información}, y se puede ver que cuando \( q = 1 \) (realizando el correspondiente límite),

\[
D(1) = -\lim_{l \to 0} \frac{\sum_{i=1}^{m} p_i \ln p_i}{\ln l}
\]

Obsérvese que

\[
S(l) = -\sum_{i} p_i \ln p_i
\]

es simplemente la entropía de Shannon a la escala \( l \). \( D(2) \) es la llamada dimensión de correlación,

\[
D(2) = \lim_{l \to 0} \frac{\ln \sum_{i=1}^{m} p_i^2}{\ln l}
\]

que evalúa en cierto modo la probabilidad conjunta de dos partes del sistema (en el término \( p_i^2 \)). Para \( q > 2 \), \( D(q) \) ya no tiene un nombre específico: es simplemente la \text{dimensión de correlación (o dimensión generalizada) de orden} \( q \). Aunque son los valores enteros de \( q \) los que tienen una interpretación directa, \( D(q) \) es una función perfectamente analítica para \( q \in \mathbb{R} \) (excepto casos singulares que no describiremos\(^\text{19}\)) y por tanto se extiende su definición para todo valor de \( q \) real, no sólo entero.

A partir de la definición de \( D(q) \) se puede obtener de forma sencilla el espectro de dimensiones fractales, que llamaremos \( f(\alpha) \). Definimos

\[
\alpha(q) = \frac{d}{dq} [(q-1)D(q)]
\]

y

\[
f(\alpha) = q \alpha(q) - D(q)(q - 1)
\]

Figura 3.24: Forma genérica de las tres funciones que caracterizan la multifractalidad. Las asintotas de \( \tau(q) \) permiten encontrar el valor de los puntos extremos del espectro multifractal, \( f(\alpha) \), según se indica en la figura. El valor de \( D(q) \) en \( q = 0 \) coincide con el máximo de \( f(\alpha) \) y da la dimensión fractal habitual.

Las funciones \( D(q) \) y \( f(\alpha) \) tienen ciertas propiedades generales, aparte de las ya citadas:

\[
D(q) \geq D(q'), \quad \forall q < q'
\]

\( D(q) \) es una función decreciente. Sólo se da la igualdad (que se cumple entonces para todo valor de \( q \)) si el objeto no es un multifractal, sino lo que se denomina un fractal puro, perfectamente descrito por un único valor de \( D(q) \), por \( D(0) \).

\( f(\alpha) \) es una función convexa de \( \alpha \). Su máximo coincide con la dimensión fractal del conjunto, para \( q = 0 \),

\[
f(\alpha(0)) = D(0)
\]

\( f(\alpha) \) siempre es tangente a la recta de pendiente unidad en el punto \( q = 1 \) \((D(1))\):

\[
f(\alpha(1)) = \alpha(1)
\]

Por último, \( f(\alpha) \) se encuentra definida entre dos valores de \( \alpha \), \( \alpha_M \) y \( \alpha_m \), y el valor de \( f(\alpha_M) \) y \( f(\alpha_m) \) (que puede ser cero) depende del soporte de la medida de probabilidad definida en el conjunto.

Definimos una última cantidad:

\[
\tau(q) = D(q) \cdot (q - 1)
\]

\( \tau(q) \) tiene también propiedades que ayudan a caracterizar la multifractalidad y proporcionan algunos valores concretos de \( D(q) \) y \( f(\alpha) \). La forma genérica de las tres funciones \( \tau(q) \), \( f(\alpha) \) y \( D(q) \) está representada en la figura 3.24.

Hemos introducido formalmente las expresiones que permiten el cálculo de los exponentes multifractales de un conjunto. Las funciones \( \chi(q), \tau(q) \) y \( f(\alpha) \) tienen, por otra parte, una interpretación concreta en el marco de la mecánica estadística (ME). Vamos a dar esta formulación alternativa a fin de profundizar un poco más en el significado del infinito espectro de dimensiones multifractales que un objeto puede presentar.
Consideremos de nuevo el conjunto de probabilidades \( \{ p_i \} \) definidas sobre nuestro multifractal y el histograma \( H(\ln p_i) \), es decir, la función distribución no de \( \{ p_i \} \), sino de \( \{ \ln p_i \} \). Como toda función de distribución, ésta puede ser caracterizada por sus momentos, que en este caso se escriben

\[
Z_\beta = \sum_i \ln p_i H(\ln p) \exp \{-\beta(\ln p)\}
\]

Observemos que \( Z_\beta \) es absolutamente análoga a una función de partición de la ME. La expresión anterior se puede transformar en

\[
Z_\beta = \sum_i p_i^\beta
\]

donde la suma se realiza ahora sobre todas las configuraciones del sistema (sobre todas sus partes), y no sobre las diferentes probabilidades, como en la primera definición. Dada la función de partición, se acostumbra a definir la energía libre del sistema, que es la encargada de identificar la posible existencia de transiciones de fase (véase el capítulo 7),

\[
F(\beta) = -\frac{\ln Z_\beta}{\ln L}
\]

que se puede transformar en

\[
Z_\beta = L^{-F(\beta)}
\]

Así, podemos establecer una dependencia potencial de la función de partición \( Z_\beta \) con la escala, \( L \): \( F(\beta) \) resulta ser el exponente que determina esta relación. Otras magnitudes tienen también su interpretación termodinámica: \( \beta \) es, como el lector quizá habrá ya imaginado, el inverso de la temperatura (o, de forma más general, del parámetro que determina el grado de orden existente en el sistema), y \( \ln p_i/\ln L \) es la magnitud análoga de la energía. Se hace evidente ahora el interés en usar \( H(\ln p) \) como magnitud equivalente a la distribución de energías en el sistema a una escala \( L \) fija.

Finalmente, podríamos hallar la entropía del sistema como la transformada de Legendre de la energía libre,

\[
S(\langle E \rangle) = F(\beta) - \beta \frac{dF(\beta)}{d\beta}
\]

donde

\[
\langle E \rangle = -\frac{dF(\beta)}{d\beta}
\]

La equivalencia entre ME y funciones multifractales queda, pues, de la forma siguiente,

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>( q )</th>
<th>( F(\beta) )</th>
<th>( \tau(q) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( E )</td>
<td>( \alpha )</td>
<td>( S(E) )</td>
<td>( f(\alpha) )</td>
</tr>
</tbody>
</table>
La demostración definitiva de la multifractalidad de un objeto consistiría en observar, por ejemplo, un cambio de pendiente en la función $\tau(q)$ cuando nos encontramos cerca de $q = 0$. Este cambio de pendiente es el responsable del cambio en $D(q)$ (entre $q \to -\infty$ y $q \to \infty$) y de la dispersión en $f(\alpha)$. Si el objeto tratado es un fractal puro, $\tau(q)$ no presenta cambio de pendiente, $D(q)$ es una línea recta horizontal y $f(\alpha)$ está formada por un único punto.

Existe actualmente una extensa literatura sobre multifractales. En cualquiera de los libros citados como bibliografía del capítulo es posible encontrar una ampliación de los conceptos vistos.

### 3.5.2 Agregación limitada por difusión (DLA)

Las estructuras que describiremos a continuación constituyen el primer ejemplo de fractal no determinista. El aspecto general que presentan se puede ver en la figura 3.25. Su forma de construcción es la siguiente: consideremos una red de $N \times N$ puntos. Inicialmente, todas las celdas están desocupadas excepto una en el centro de la red. Aleatoriamente se libera un “random walker” a una distancia $r$ grande del centro. Este “random walker” se mueve hasta que localiza una de las celdas ocupadas, momento en el cual se adhiere irreversiblemente a estas, ocupando el la última en que se hallaba. La probabilidad de que las partículas se adhieran a las puntas de la estructura en formación es siempre mayor que la probabilidad de que se adhieran a los lados. Estas probabilidades dependen de la estructura del agregado, y son las responsables de la formación de fiordos muy profundos que tienen una probabilidad de crecimiento varios órdenes de magnitud inferior a la de los extremos.

La distribución de las probabilidades de crecimiento en las celdas desocupadas de la red en los agregados limitados por difusión (DLA) caracteriza su estructura, y se ha observado que presenta espectro multifractal. En este caso, la distribución de probabilidades define directamente una medida en el sistema, igual que habíamos definido la medida de masa. Obsérvese, además, que ambas pueden entenderse como probabilidades. En ocasiones, este espectro multifractal no está definido para $q < 0$ (de hecho, $q(-\infty)$ diverge cuando el tamaño del agregado tiende a infinito),

---

20 Esta es la regla más elemental de formación de un agregado de este tipo. Existen numerosas variantes, en las que el caminante podría adherirse con una cierta probabilidad, o en las que las conexiones formadas podrían eventualmente romperse.
debido a propiedades que no describiremos. Por otra parte, si intentamos realizar una medida de masa sobre el sistema y calcular, por ejemplo, la función \( f(\alpha) \), se observa que nos encontramos ante un fractal puro, con un sólo valor para su dimensión fractal espacial, que resulta ser

\[ D_0(DLA) = 1.70 \pm 0.01 \]

La ausencia de multifractalidad en la distribución de masa se debe a la existencia de una medida laplaciana subyacente a la generación del fractal, y que comparten todos los sistemas que presentan DLA. Algunos de éstos son la deposición electroquímica, la solidificación de dendritas (en la formación de cristales de nieve o en el crecimiento de líquenes, por ejemplo), la producción de relámpagos (por ionización del aire o por exceso de potencial aplicado a un dieléctrico), la digitación viscosa (invasión de una sustancia por otra de diferente densidad, por ejemplo agua en petróleo para realizar el vaciado de pozos), o la cristalización rápida de lava, entre más de 50 sistemas que forman el mismo tipo de estructuras.

Entre los sistemas que poseen un campo eléctrico como generador (por ejemplo relámpagos y deposiciones eléctricas), \( E \) es una cantidad con gradiente nulo (cumple \( \nabla E = 0 \), y si la cantidad no depende del tiempo es además una magnitud conservada) y por tanto se verifica la ecuación de Laplace para el potencial eléctrico,

\[ \nabla^2 \phi = 0 \]

En los sistemas que provienen del campo de la mecánica de fluidos (digitación viscosa), se verifica una ecuación de Laplace para la presión del fluido,

\[ \nabla^2 P = 0 \]

Para los sistemas que forman dendritas y solidifican, constituyendo el agregado (crustales de nieve, lava, neuronas de la retina), es la concentración del compuesto, \( c \) (la variación de la cual da el ritmo de crecimiento del DLA), la que verifica una ecuación laplaciana,

\[ \nabla^2 c = 0 \]

La tabla siguiente proporciona la equivalencia entre estos sistemas y las magnitudes que los caracterizan (\( r \) denota la posición, y \( t \) el tiempo):

<table>
<thead>
<tr>
<th>Sistemas eléctricos</th>
<th>S. Mecánica de fluidos</th>
<th>Solidificación dendrítica</th>
</tr>
</thead>
<tbody>
<tr>
<td>Potencial electrostático ( \phi(r,t) )</td>
<td>Presión ( P(r,t) )</td>
<td>concentración ( c(r,t) )</td>
</tr>
<tr>
<td>Campo eléctrico ( E \propto -\nabla \phi(r,t) )</td>
<td>Velocidad ( \nu \propto -\nabla P(r,t) )</td>
<td>Ritmo de crecimiento ( \nu \propto -\nabla c(r,t) )</td>
</tr>
<tr>
<td>Gradiente nulo ( \nabla E = 0 )</td>
<td>( \nabla \nu = 0 )</td>
<td>( \nabla \nu = 0 )</td>
</tr>
<tr>
<td>Ecuación de Laplace ( \nabla^2 \phi = 0 )</td>
<td>( \nabla^2 P = 0 )</td>
<td>( \nabla^2 c = 0 )</td>
</tr>
</tbody>
</table>

Todos los sistemas con medidas laplacianas producen fractales puros en la distribución espacial. Intuitivamente, se puede entender que un campo eléctrico, una velocidad o un ritmo de crecimiento uniformes provoquen que la distribución de la masa sea en cierto modo también "uniforme", que
aqui significa independiente de la zona espacial considerada, con lo cual no es posible generar áreas más o menos densas que estarían asociadas a los distintos valores de $f(\alpha)$, a las distintas dimensiones fractales.

Recientemente, la estructura de los agregados limitados por difusión se ha encontrado en colonias de bacterias, en el crecimiento de neuronas e incluso en la configuración del Metro de París. Todos estos sistemas comparten la misma geometría, característica de la cual es el valor 1,70 de todas sus dimensiones de correlación. La ubicuidad de estos agregados induce a pensar en alguna razón de fondo, aún no determinada, que justificaría de forma natural la ventaja de utilizar este tipo de morfología. Si incluso la evolución parece haber escogido estas estructuras como forma óptima de construcción, la pregunta abierta es, ¿porqué? De momento, se puede afirmar que estos agregados proporcionan un “llenado” del espacio con una conectividad óptima para los propósitos del sistema.

Hay muchos otros sistemas físicos que producen fractales en el espacio. Superpuestos a la dinámica determinista encontramos en ocasiones términos de ruido que destruyen la autosimilaridad exacta, verificándose entonces una autosimilaridad de tipo estadístico.

Pensemos que cualquier sistema que incorpore algún término aleatorio en su dinámica provoca una independencia de escala en su estructura, debido a que estos términos no deterministas no poseen un nivel específico de dehiscencia, sino que producen fluctuaciones y desviaciones a todas las escalas. Un ejemplo interesante lo constituye la percolación, que será ampliamente descrita en el capítulo dedicado a fenómenos críticos, así como algunos otros sistemas físicos y los fractales asociados correspondientes, que serán descritos en secciones específicas de este libro.

Veremos también en qué consisten los sistemas críticos autoorganizados, los cuales generan de forma espontánea estructuras fractales, y no sólo en el espacio, sino también en el tiempo. Los atractores extraños aparecerán a todo lo largo del libro, y prácticamente todos serán fractales (en muchas ocasiones multifractales), con estructura y dimensión dependientes del sistema dinámico generador (sea o no determinista). En este caso, la generación de fractalidad es debida a la propia estructura del espacio de fases, a los infinitos estiramientos y plegamientos que sufre el soporte de las órbitas. En el capítulo dedicado a caos se describe este caso extensamente.

Los autómatas celulares (descritos en el capítulo correspondiente) son sistemas generadores en ocasiones de fractales (en función de ciertos parámetros) generalmente deterministas. El fractal no determinista aparece como consecuencia de la introducción de ruido en la dinámica, o de probabilidades de transición entre estados. En cualquier caso, debido a términos aleatorios, no controlados.

Bibliografía


Capítulo 4

Atractores Periódicos y Cuasiperiódicos

Se ha visto en el capítulo 2 la forma en que se puede iniciar el estudio de un sistema de ecuaciones no lineal. Es posible conocer ciertas características del equilibrio a nivel local, en los casos en los que, como se ha descrito, la estabilidad es asintótica. Sin embargo, aún no somos capaces de decir nada sobre el sistema a nivel global. Por ejemplo, sabemos a partir de la experiencia que algunos sistemas presentan lo que llamamos ciclos límite, que son, como su nombre indica, trayectorias cerradas a las cuales todas las trayectorias del sistema pueden acabar convergiendo. La existencia y la determinación de estos ciclos límite (que también pueden ser repulsores), de las llamadas órbitas homoclínicas (órbitas que unen un punto fijo con él mismo) o de las órbitas heteroclínicas (unen dos puntos fijos diferentes) permite el estudio de los sistemas dinámicos a nivel global. Este estudio (global) es un poco más complicado que el de los puntos fijos (local), pero es capaz también de decir más sobre el sistema. Por ejemplo, si fuésemos capaces de determinar que en cierto sistema dinámico, para cierto valor de los parámetros, aparecen dos órbitas homoclínicas, podríamos asegurar que el sistema va a tener, en este caso, dinámica caótica.

La aparición del tipo de órbitas citado va ligada a la existencia de la variedad centro (valores propios con parte real nula en la matriz lineal). Llevaremos en este capítulo un poco más allá el estudio de los sistemas dinámicos con valores propios nulos. Entre las situaciones periódicas y el caos veremos también que podemos encontrarnos con los llamados atractores cuasiperiódicos (sobre los que el sistema describe órbitas cuasiperiódicas). En este caso, el movimiento posee más de una frecuencia característica, y tiene lugar sobre los llamados toros $n$-dimensionales. Un toro en dos dimensiones es cualquier curva cerrada, y en tres puede ser la cámara de un neumático, o un donut, o una taza (los tres tienen un único agujero, se dice que son topológicamente equivalentes).

Trataremos en este capítulo aún con sistemas disipativos. En el capítulo 10, sobre caos hamiltoniano, describiremos sistemas que exhibirán algunas propiedades idénticas a los aquí descritos. En particular, hallaremos también caos y órbitas cuasiperiódicas: a diferencia de los sistemas disipativos, los sistemas hamiltonianos son conservativos, y en consecuencia no poseen atractores de la dinámica en el sentido de “lugar o lugares del espacio de las fases al que el sistema tiende en tiempo infinito asintóticamente, para cualquier condición inicial”.

Iniciaremos el capítulo con el estudio de la forma en que aparecen los ciclos límite, primer paso hacia el estudio de la dinámica a nivel global.
4.1 Bifurcaciones

Definición

Supongamos una EDO de la forma

\[ \dot{x} = F(x, \mu) \]

\[ x \in \mathbb{R}^n, \mu \in \mathbb{R}^p. \] Los puntos fijos

\[ x_0^i; \quad F(x_0^i, \mu) = 0, \quad i \in \mathbb{N} \]

dependerán del valor de los parámetros \( \mu \). Cuando estos varíen, tanto el número de puntos fijos como su naturaleza (estables, inestables o indecidibles; nodos, focos, puntos de silla, ...) pueden variar. Cuando esto sucede se dice que el sistema sufre una bifurcación.

Para tener un cambio en las características de un punto fijo, y por tanto un cambio cualitativo en el espacio de fases del sistema de EDOs, es necesario que se anule alguno de sus valores propios. Pequeños cambios en el caso de tener estabilidad asintótica (\( \Re(\lambda_i) \neq 0, \forall i \)) no implican cambios en la estabilidad. Únicamente cuando los parámetros del sistema permiten cruzar de \( \Re(\lambda_i) > 0 \) a \( \Re(\lambda_j) < 0 \) (o en sentido contrario) para algunos \( \lambda_j \), aparecen puntos fijos cualitativamente diferentes, y por tanto existe la posibilidad de variar la dinámica asintótica del sistema.

Describiremos a continuación brevemente algunos de los casos que podemos encontrar, en función del número de valores propios con parte real nula y de los términos de orden superior. Veremos que, como consecuencia del cambio cualitativo en la dinámica, el sistema es estructuralmente inestable cuando el parámetro (o conjunto de parámetros) toma el valor que provoca la anulación de la parte real de uno o varios de los valores propios. Es decir, pequeñas perturbaciones del sistema en este estado cambian su retrato de fases.

4.1.1 Un único valor propio nulo

Si existe un único valor propio nulo debe de tener parte imaginaria también nula, ya que de no ser así debería presentarse su pareja compleja conjugada, y el valor propio nulo no sería único. Por tanto, debe ser

\[ \lambda = 0 \]

En este caso, la variedad centro asociada al valor propio es unidimensional, y podemos escribir todas las variables del sistema como funciones de la variable asociada al valor propio nulo, en un entorno del punto fijo. Véase en el apéndice correspondiente a este capítulo la fórmula en que es posible reducir el sistema de EDOs a tantas ecuaciones como valores propios nulos aparezcantan, y por tanto tratar el sistema reducido a la variedad centro (siempre en un entorno del punto de equilibrio). Para más información, consultese la bibliografía citada al final del capítulo.

Bifurcación silla-nodo

En este primer caso, el sistema pasa de la no existencia de puntos fijos a la aparición de un punto de silla para el valor exacto del parámetro en la bifurcación. Este punto pasa a convertirse inmediatamente en una pareja de dos nodos que presentan estabilidades contrarias. Veámoslo con un ejemplo. Consideremos la EDO en una dimensión

\[ \dot{x} = (\mu - \mu_c) - x^2 \]
Figura 4.1: Esquema de una bifurcación silla-nodo en una dimensión en función del parámetro del sistema. En todas las figuras sobre bifurcaciones, la línea continua significa que el punto fijo es estable, y la discontinua, inestable. Los puntos fijos se marcan con un pequeño círculo. Prestése atención en los casos en que existan puntos fijos sobre el eje x.

y calculemos sus puntos fijos, soluciones de \((\mu - \mu_c) - x^2 = 0\). En general, \(x_{\pm} = \pm \sqrt{\mu - \mu_c}\). Observamos que para \(\mu < \mu_c\) el sistema no presenta ningún punto fijo. En el punto exacto \(\mu = \mu_c\) tenemos una única solución, \(x^* = 0\), (a la cual le corresponde un valor propio nulo) y que resulta ser un punto de silla. Cuando \(\mu > \mu_c\), tenemos dos soluciones, \(x_{\pm} = \pm \sqrt{\mu - \mu_c}\), que corresponden a un nodo estable (valor positivo de la raíz) y a un nodo inestable (-). La figura 4.1 contiene un esquema de la forma en que se produce este tipo de bifurcación.

En el punto crítico, \(\mu = \mu_c\), y sólo existe variedad centro, \(W^c\), que contiene toda la recta real. Cuando \(\mu > \mu_c\), las variedades estable e inestable son

\[
W^s(\sqrt{\mu - \mu_c}) = (-\sqrt{\mu - \mu_c}, \infty) \quad W^u(-\sqrt{\mu - \mu_c}) = (-\infty, \sqrt{\mu - \mu_c})
\]

Obsérvese el solapamiento de las variedades en el intervalo

\[
S = (-\sqrt{\mu - \mu_c}, \sqrt{\mu - \mu_c})
\]

Este intervalo pertenece a la variedad estable para el nodo atractor, ya que los puntos en esta zona "caen" a \(\sqrt{\mu - \mu_c}\). En cambio, pertenece a la variedad inestable para el nodo repulsor, ya que se alejan de \(-\sqrt{\mu - \mu_c}\). Podemos decir que el campo de vectores \(F(x) = -x^2\) es estructuralmente inestable, puesto que cualquier pequeña perturbación introducida por un parámetro \(\mu\) hace desaparecer el punto de silla y lleva al sistema bien a no tener ningún punto fijo, bien a la aparición de dos nodos de estabilidad opuesta.

Si consideramos un desarrollo del campo \(F(x, \mu)\) que define la EDO,

\[
F(x, \mu) = F_{\mu}(0, 0)\mu + \frac{1}{2} F_{xx}(0, 0)x^2 + O(\mu^2, \mu x, x^3)
\]

la condición para que aparezca una bifurcación del tipo citado es

\[
F_{\mu}(0, 0) \neq 0, \quad F_{xx}(0, 0) \neq 0\]

**Bifurcación transcritica**

Consideremos la ecuación en una dimensión
Figura 4.2: Esquema de la generación de una bifurcación transcritica. La estabilidad de los puntos fijos se intercambia en el punto de bifurcación.

\[ \dot{x} = (\mu - \mu_c)x - x^2 \]

Los puntos fijos son soluciones de

\[ x \{ (\mu - \mu_c) - x \} = 0 \implies x_1 = 0, \quad x_2 = \mu - \mu_c \]

El valor \( \mu = \mu_c \) representa de nuevo un punto de bifurcación en el cual, en este caso, simplemente se intercambia la estabilidad de los dos puntos fijos existentes. Obsérvese el diagrama de bifurcación en la figura 4.2.

**Bifurcación en horquilla (pitchfork)**

Consideremos

\[ \dot{x} = (\mu - \mu_c)x - x^3 \]

En este caso, los puntos fijos son

\[ x_1 = 0, \quad x_2 = \pm \sqrt{\mu - \mu_c} \]

En el punto \( \mu = \mu_c \) se pasa de un único punto fijo estable (que existe en todo el dominio \( \mu \leq \mu_c \)) a la existencia de tres puntos fijos: el origen pierde su estabilidad y aparece una pareja de puntos fijos estables.

Podemos entender esta bifurcación físicamente si tenemos en cuenta que la ecuación de partida puede expresarse como la derivada de una función potencial

\[ \frac{dx}{dt} = -\frac{\partial V_\mu}{\partial x} \]

que en nuestro ejemplo es

\[ V_\mu = \frac{(\mu_c - \mu)}{2} x^2 + \frac{1}{4} x^4 \]

y gráficamente se puede ver que presenta un único mínimo en \( x = 0 \) para \( \mu > \mu_c \) y dos mínimos para \( \mu < \mu_c \) (véase la figura 8 del capítulo 7). Este esquema es muy útil para comprender la
naturaleza de la bifurcación. Observamos que en todo momento el potencial es simétrico, con una única opción (un atractor único) estable que se bifurca en dos nuevos atractores.

Si imaginamos que el estado del sistema se representa al principio por una bola que puede rodar situada en el origen \( x = 0 \), observamos que cualquier perturbación alrededor de este punto cuando \( \mu - \mu_c < 0 \) no aleja el sistema del origen: el punto de equilibrio es estable. En cambio, cuando \( \mu - \mu_c > 0 \), una perturbación de la bola implica que ésta rodará hacia uno de los dos mínimos posibles. Pese a la existencia de simetría, el sistema se ve forzado a efectuar una elección. Esta bifurcación, en la que un sistema dinámico no lineal realiza una elección entre dos alternativas posibles después de atravesar un punto crítico recibe el nombre de \textit{bifurcación con rotura de simetría}.

Los fenómenos no lineales que muestran rotura de simetría juegan un papel esencial en numerosas ideas físicas (véase el capítulo sobre estructuras de Turing) pero también en otras áreas. El economista Brian Arthur, que ha aplicado estas ideas en su estudio de las retroalimentaciones (\textit{feedback}) positivas en economía, nos da algunos ejemplos. Uno de ellos es la adopción de relojes “horarios” en lugar de los “antihorarios”. El reloj de la catedral de Florencia es un ejemplo del segundo caso. Fue diseñado en 1443 por Paolo Vicello, en una época en la que la construcción actual no estaba aún definida. Pero, tal y como señala Arthur, a medida que algunos relojes de sentido horario fueron ganando al número de sus contrapartidas antihorarias, una retroalimentación positiva precipitó el resultado de esta competencia en una única dirección. Estas roturas de simetría juegan probablemente un papel destacado en economía. Podemos pensar en otro ejemplo más actual, como la imposición del sistema de video VHS sobre el sistema Beta. Obsérvese que prácticamente no ha quedado lugar para la coexistencia.

La bifurcación transcritica y la bifurcación en horquilla no son bifurcaciones \textit{genéricas}. Esto quiere decir que cierto tipo de perturbaciones puede provocar que desaparezcan y se conviertan en una bifurcación silla-nodo, que sí resulta ser genérica. Una perturbación del tipo \( \epsilon \) no las destruiría, ya que simplemente implica una traslación del punto en el cual se produce la bifurcación. En lugar de ser en \( \mu = \mu_c \) sería en \( \mu = \mu_c - \epsilon \). En cambio, el hecho de añadir un parámetro \( \epsilon \), por ejemplo, cambia los puntos fijos. Consideremos la bifurcación transcritica. Si el sistema fuese de la forma

\[
F(x, \mu, \epsilon) = (\mu - \mu_c)x - x^2 + \epsilon
\]

obtenemos
Para ε > 0 no existe cruce entre las curvas de puntos fijos, no hay intercambio de estabilidad ni, por tanto, bifurcación (ya que el número de puntos fijos no varía). Para ε < 0 existe un dominio 
((μ - μ_c)^2 < 4ε) en el que no hay solución, y no existen puntos fijos.

Una situación semejante aparece en la bifurcación en horquilla para el mismo tipo de perturbación. Sin embargo, existen mecanismos que permiten que, por ejemplo, la bifurcación en horquilla se observe de forma genérica. Tal es el caso de las ecuaciones de Lorenz, que, debido a su simetría (implicita en el problema físico que se está tratando) no admitirían una perturbación del tipo +ε, y por tanto la bifurcación en horquilla resulta ser la bifurcación genérica para este sistema.

Observemos que en estas dos últimas bifurcaciones no se cumple F_μ(0, 0) ≠ 0. En el caso de la bifurcación transcritical tenemos F_μ(0, 0) ≠ 0, y en la bifurcación en horquilla F_μ(0, 0) = 0.

Podríamos describir muchas otras bifurcaciones, en las que, dependiendo del orden de la parte no lineal (digamos m) y de sus coeficientes podríamos obtener hasta m puntos fijos diferentes tras la bifurcación, que presentarían estabilidad o inestabilidad. También podríamos analizar qué tipo de perturbaciones podrían destruir la bifurcación o parte de ella, o bien si ésta es genérica. Sin embargo, lo hasta ahora expuesto es suficiente e ilustra los tipos principales de bifurcaciones producidas en sistemas con un único valor propio nulo.

### 4.1.2 Bifurcación de Poincaré-Andronov-Hopf

Esta bifurcación se produce cuando el sistema de EDOs considerado posee dos valores propios con parte real nula y con parte imaginaria no nula, siendo por tanto uno de estos valores el complejo conjugado del otro. El resto de valores propios del sistema se suponen, evidentemente, con parte real no nula. Podemos decir que una bifurcación de Poincaré-Andronov-Hopf, o muy frecuentemente sólo bifurcación de Hopf, genera un ciclo límite a partir de un punto fijo, cuando este se inestabiliza. Se comprende que la parte real de los valores propios nulos sufre pues un cambio, de negativo a positivo. Las definiciones siguientes, junto con el teorema en el cual son necesarias, establecen las condiciones de existencia de una bifurcación de Hopf.

Consideremos el sistema en dos dimensiones

\[
\begin{align*}
\dot{x} &= \mu x - y + p(x, y) \\
\dot{y} &= x + \mu y + q(x, y) \\
\end{align*}
\]

(4.2.1)

donde p y q son funciones analíticas que admiten un desarrollo de la forma

\[
\begin{align*}
p(x, y) &= \sum_{i+j\geq 2} a_{ij} x^i y^j \\
q(x, y) &= \sum_{i+j\geq 2} b_{ij} x^i y^j \\
\end{align*}
\]

= \(a_{20}x^2 + a_{11}xy + a_{02}y^2 + a_{30}x^3 + a_{21}x^2y + a_{12}xy^2 + a_{03}y^3\) + ...

= \(b_{20}x^2 + b_{11}xy + b_{02}y^2 + b_{30}x^3 + b_{21}x^2y + b_{12}xy^2 + b_{03}y^3\) + ...
Definición

El número de Lyapunov correspondiente al sistema anterior es

\[
\sigma = \frac{3\pi}{2} \left[ 3(a_{30} + b_{03}) + (a_{12} + b_{21}) - 2(a_{20}b_{21} - a_{02}b_{02}) + a_{11}(a_{02} + a_{20}) - b_{11}(b_{02} + b_{20}) \right]
\]

En particular, si \( \sigma \neq 0 \) el origen es un foco estable para \( \sigma < 0 \) e inestable para \( \sigma > 0 \), y en el punto \( \mu = 0 \) tiene lugar una bifurcación de Hopf, tal como enuncia el siguiente teorema.

Teorema

Si \( \sigma \neq 0 \), en el origen del sistema plano analítico 4.2.1 tiene lugar una bifurcación de Hopf para el parámetro \( \mu = 0 \). En particular, si \( \sigma < 0 \), entonces 4.2.1 tiene un ciclo límite estable único para \( \mu > 0 \) y ningún ciclo límite para \( \mu \leq 0 \); si \( \sigma > 0 \), 4.2.1 tiene un único ciclo límite inestable para \( \mu < 0 \) y ningún ciclo límite para \( \mu \geq 0 \). Si \( \sigma < 0 \), el mapa de fases local de 4.2.1 es topológicamente equivalente al que se muestra en la figura 4.4 y existe una superficie de órbitas periódicas que corta tangencial y cuadráticamente el plano \((x, y)\) en el origen en \(\mathbb{R}^2 \times \mathbb{R}\).

Para un sistema plano analítico de la forma

\[
\dot{x} = ax + by + p(x, y) \\
\dot{y} = cx + dy + q(x, y)
\]

con \( \Delta = ad - bc > 0 \), \( a + b = 0 \) y \( p(x, y) \) y \( q(x, y) \) funciones analíticas de la forma antes indicada, la matriz \( \begin{pmatrix} a & b \\ c & d \end{pmatrix} \) posee una pareja de valores propios compleja conjugada y el origen será un foco. El número de Lyapunov \( \sigma \) está entonces dado por

\[
\sigma = \frac{-3\pi}{2b\Delta^{3/2}} \left\{ [ac(a_{11}^2 + a_{11}b_{02} + a_{02}b_{11}) + ab(b_{11}^2 + a_{20}b_{11} + a_{11}b_{02}) + \\
bc^2(11a_{02} + 2a_{02}b_{02}) - 2ac(b_{02}^2 - a_{20}a_{02}) - 2ab(a_{20}b_{20} - b_{20}a_{02}) - \\
b^2(2a_{20}b_{20} + b_{11}b_{20}) + (bc - 2a^2)(b_{11}b_{02} - a_{11}a_{02})] - \\
(a^2 + bc) [3(c_{03} - b_{30}) + 2a(a_{21} + b_{12}) + (ca_{12} - bb_{21})] \right\}
\]

Cuando las condiciones anteriormente exigidas no se satisfacen (en particular, si \( \sigma = 0 \)) se puede producir más de un ciclo límite en el punto de bifurcación \(^1\).

Ilustraremos con un ejemplo la bifurcación de Hopf.

Ejemplo

\[
\dot{x} = -y + x(\mu - x^2 - y^2) \\
\dot{y} = x + y(\mu - x^2 - y^2)
\]

El único punto crítico de este sistema corresponde al origen, \( x = (0, 0) \). Si consideramos la parte lineal del sistema,

\(^1\) Esto depende concretamente de la multiplicidad del foco que se inestabiliza. En el caso de la bifurcación de Hopf, el foco tiene multiplicidad uno, y emite un único ciclo límite. Si tuviese multiplicidad \( k \), podría emitir hasta \( k \) ciclos límite. Véase la siguiente sección, 4.3.
\[ DF(0, 0) = \begin{pmatrix} \mu & -1 \\ 1 & \mu \end{pmatrix} \]

y calculamos los valores propios asociados \( \lambda \), obtenemos

\[ \lambda _{\pm} = -\mu \pm i \]

Si \( \mu > 0 \) el origen es un foco estable, y si \( \mu < 0 \) es un foco inestable. Para \( \mu = 0 \) tiene lugar una bifurcación. Calculemos \( \sigma \) para comprobar que es una bifurcación de Hopf (obsérvese que la forma del sistema corresponde a la forma genérica dada en el enunciado del teorema sobre esta bifurcación). En este caso tenemos

\[ p(x, y) = -x^3 - xy^2 \quad q(x, y) = -yx^2 - y^3 \]

así que los coeficientes no nulos del desarrollo de \( p \) y \( q \) son

\[ a_{30} = -1, \quad a_{12} = -1 \]

para \( p \) y

\[ b_{21} = -1, \quad b_{03} = -1 \]

para \( q \). Por tanto,

\[ \sigma = \frac{3\pi}{2} \left[ 3(a_{30} + b_{03}) + (a_{12} + b_{21}) \right] = -12\pi \neq 0 \]

Confirmanos que el origen es un foco estable \( (\sigma < 0) \) y mediante el uso del teorema visto podemos afirmar que tras la bifurcación el sistema presentará un ciclo límite único y estable (para \( \mu < 0 \)).

**El cambio a coordenadas polares**

Una forma alternativa, útil y muy intuitiva de comprobar la existencia del ciclo límite consiste en escribir las ecuaciones del sistema en coordenadas polares. En el caso anterior, realicemos el cambio

\[ x = r \cos \theta \]
\[ y = r \sin \theta \]

Si consideramos el cambio inverso,

\[ r = \sqrt{x^2 + y^2} \]
\[ \theta = \arctan \frac{y}{x} \]

y derivamos, resulta

\[ 2r\dot{r} = 2x\dot{x} + 2y\dot{y} \]
\[ \dot{\theta} = \frac{1}{1 + \left( \frac{x}{y} \right)^2} \frac{\dot{x}}{y} + \frac{1}{1 + \left( \frac{x}{y} \right)^2} \frac{\dot{y}}{y} \]

En estas últimas expresiones debemos sustituir \( x, y \) en función de \( r \) y \( \theta \) y la definición de nuestro sistema,
Figura 4.4: Tres estados de la bifurcación de Hopf en el sistema del ejemplo 1. Para \( \mu < 0 \) no existe ciclo límite. El origen es un foco estable. En \( \mu = 0 \) se produce la bifurcación. La tendencia de las trayectorias al origen es muy lenta. Cuando \( \mu > 0 \) aparece un ciclo límite. El origen se ha convertido en un foco inestable.

\[
\dot{x} = -r \sin \theta + r \cos \theta (\mu - r^2) \\
\dot{y} = r \cos \theta + r \sin \theta (\mu - r^2)
\]

Se hace uso únicamente de \( x^2 + y^2 = r^2 (\cos^2 \theta + \sin^2 \theta) = 1 \), lo cual permite llegar, sin excesivo esfuerzo, a

\[
\dot{r} = r (\mu - r^2) \\
\dot{\theta} = 1
\]

Hemos separado el sistema en dos ecuaciones independientes: una para la parte radial y otra para la parte angular. La última tiene una solución trivial, \( \theta(t) = \theta_0 + t \), lo cual indica que los puntos sobre las trayectorias se desplazan a velocidad angular constante alrededor del origen. Por otra parte, para \( \mu > 0 \) tenemos una solución para la parte radial que resulta ser estable, \( r = +\sqrt{\mu} \). La solución negativa de la raíz no tiene sentido, ya que \( r \geq 0 \) en coordenadas polares.

En la figura 4.4 se representan las trayectorias del sistema para diversas condiciones iniciales antes, en y después de la bifurcación. Para \( \mu = 0 \) las trayectorias tienden muy lentamente al origen (o al ciclo límite de radio nulo), así que sólo se ha representado una parte de las mismas.

### 4.2 La aplicación de Poincaré

La aplicación de Poincaré es una herramienta básica y geométricamente sencilla para estudiar la estabilidad y las bifurcaciones de órbitas periódicas. La aplicación de Poincaré, o aplicación de primer retorno, fue presentada por Henri Poincaré en 1881. Se basa en una idea sencilla que podemos enunciar de la forma siguiente: si \( \Gamma \) es una órbita periódica del sistema

\[
\dot{x} = F(x)
\]

a través del punto \( x_0 \) y \( \Sigma \) es un hiperplano perpendicular a \( \Gamma \) en \( x_0 \), entonces para cualquier punto \( x \in \Sigma \) suficientemente próximo a \( x_0 \), la solución de 4.3.1 que pasa por \( x \) en \( t = 0 \), \( \Phi_t(x) \), atravesará \( \Sigma \) de nuevo en un punto \( P(x) \) cercano a \( x_0 \). La aplicación
se denomina \textit{aplicación de Poincaré}.

El siguiente teorema establece la existencia y continuidad de la \textit{aplicación de Poincaré}, \( P(x) \) y de su primera derivada, \( DP(x) \).

\textbf{Teorema}

Sea \( E \) un subconjunto abierto de \( \mathbb{R}^n \), y \( F \in C^1(E) \). Supongamos que \( \Phi_t(x) \) es una solución periódica de 4.3.1, de periodo \( T \), y que el ciclo
\[
\Gamma = \{ x \in \mathbb{R}^n; x = \Phi_t(x_0), 0 \leq t \leq T \}
\]
está contenido en \( E \). Sea \( \Sigma \) el hiperplano ortogonal a \( \Gamma \) en \( x_0 \).
\[
\Sigma = \{ x \in \mathbb{R}^n; (x - x_0).F(x_0) = 0 \}
\]
En este caso, existe \( \delta > 0 \) y una función única \( \tau(x) \), definida y derivable en continuidad para \( x \in N_{\delta}(x_0) \) tal que
\[
\tau(x_0). \Phi_t(x)(x) \in \Sigma, \quad \forall x \in N_{\delta}(x_0)
\]

Según las definiciones dadas en el teorema anterior, y para \( x \in N_{\delta}(x_0) \cap \Sigma \), la función
\[
P(x) = \Phi_t(x)(x)
\]
se denomina \textit{aplicación de Poincaré} para \( \Gamma \) en \( x_0 \). Los puntos fijos de la \textit{aplicación de Poincaré}, es decir, \( x \in \Sigma \) que satisfacen \( P(x) = x \) corresponden a órbitas periódicas.

Si realizamos el cambio \( t \rightarrow -t \) en el sistema 4.3.1, se puede ver que \( P(x) \) tiene una función inversa \( P^{-1}(x) \) derivable en continuidad, y por tanto \( P \) es un difeomorfismo ("función suave con inversa suave"). En ciertas ocasiones es posible calcular analíticamente la expresión de la \textit{aplicación de Poincaré}. Consideremos como ejemplo el sistema estudiado en la sección anterior,
\[
\dot{x} = -y + x(\mu - x^2 - y^2)
\]
\[
\dot{y} = x + y(\mu - x^2 - y^2)
\]
Este sistema tiene, para \( \mu > 0 \), un ciclo límite como solución,
\[
\gamma(t) = (\sqrt{\mu} \cos t, \sqrt{\mu} \sin t)^T
\]
La aplicación de Poincaré se puede determinar en este caso solucionando el sistema escrito en coordenadas polares. Recordemos que habíamos obtenido
\[
\dot{r} = r(\mu - r^2)
\]
\[
\dot{\theta} = 1
\]
Consideremos las condiciones iniciales \( r(0) = r_0 \) y \( \theta(0) = \theta_0 \). La primera ecuación tiene una solución de la forma
\[
r(t, r_0) = \left[ \frac{1}{\mu} + \left( \frac{1}{r_0^2} - \frac{1}{\mu} \right) e^{-2\mu t} \right]^{-1/2}
\]
lo cual se puede comprobar por simple sustitución, y la segunda

$$\theta(t, \theta_0) = t + \theta_0$$

Si $\Sigma$ es la recta que tiene pendiente $\theta = \theta_0$ y pasa por el origen, entonces $\Sigma$ es perpendicular a $\Gamma$ y la trayectoria a través del punto $(\tau_0, \theta_0) \in \Sigma \cap \Gamma$ en $t = 0$ intersecta $\Sigma$ de nuevo en $t = 2\pi$.

La aplicación de Poincaré estará dada por

$$P(\tau_0) = \left[ \frac{1}{\mu} + \left( \frac{1}{r_0^2} - \frac{1}{\mu} \right) e^{-4\mu\pi} \right]^{-1/2}$$

con lo cual estamos simplemente considerando los puntos de la trayectoria correspondientes a intervalos discretos de tiempo con un incremento $\Delta t = 2\pi$. La aplicación de Poincaré proporciona en cierto modo una visión "estroboscópica" de la dinámica. Si buscamos los puntos fijos de la aplicación ($P(\tau_0) = \tau_0$), que corresponden al ciclo límite anterior $\gamma(t)$, obtenemos $\tau_0 = \sqrt{\mu}$, como era de esperar. La derivada de $P(r)$ en $\tau_0 = \sqrt{\mu}$ proporciona

$$P'(\tau_0) = \frac{1}{\tau_0^2} \left[ \frac{1}{\mu} + \left( \frac{1}{r_0^2} - \frac{1}{\mu} \right) e^{-4\mu\pi} \right] e^{-4\pi\mu} = e^{-4\pi\mu}$$

Como se vio en el capítulo 2, en referencia a las aplicaiciones discretas y su estabilidad, podemos utilizar ahora aquellos criterios y estudiar la estabilidad del ciclo límite $\gamma(t)$. Este será estable si

$$P'(\tau_0) < 1 \implies e^{-4\pi\mu} < 1$$

En particular, tendremos la igualdad cuando $\mu = 0$, pero $\forall \mu > 0$ la función $e^{-4\pi\mu}$ será menor que 1. (Obsérvese que es una función monótona decreciente con el máximo precisamente en $\mu = 0$.)

### 4.2.1 Función de desplazamiento

Relacionada con la función de Poincaré introduciremos ahora la función de desplazamiento, que permite caracterizar la multiplicidad de un ciclo límite y de los focos que experimentan la bifurcación de Hopf. Los resultados que enunciamos a continuación se aplican solamente a sistemas en 2 dimensiones.

Supongamos que (sin pérdida de generalidad) trasladamos el origen al punto $x_0 \in \Gamma \cap \Sigma$. En este caso, $\Sigma$ siempre pasará por el origen. Ahora, $0 \in \Gamma \cap \Sigma$ divide a $\Sigma$ en dos subsegmentos abiertos, $\Sigma^+$ y $\Sigma^-$. $\Sigma^+$ pertenece por completo al exterior de $\Gamma$.

Sea $s$ la distancia al origen de los puntos que interceptan $\Sigma$, y escojamos $s > 0$ para los puntos $x_+ \in \Sigma^+$ y $s < 0$ para $x_- \in \Sigma^-$. Según el teorema sobre la aplicación de Poincaré, ésta está definida para $|s| < \delta$, y hemos tomado $P(0) = 0$. Introducimos ahora la función de desplazamiento como

$$d(s) = P(s) - s$$

Entonces,

$$d(0) = 0, \quad d'(s) = P'(s) - 1$$

Observemos que $d(s)$ permite caracterizar la estabilidad de las órbitas periódicas de la forma siguiente: si $d'(0) > 0$, el ciclo límite es instable; si $d'(0) < 0$, el ciclo límite es estable.
Definición

Sea $P(s)$ la aplicación de Poincaré para un ciclo $\Gamma$ de un sistema analítico en dos dimensiones, y sea $d(s) = P(s) - s$ la función de desplazamiento. Si $d(0) = d'(0) = \ldots = d^{(k-1)}(0) = 0$ y $d^{(k)}(0) \neq 0$, $\Gamma$ se denomina ciclo límite múltiple de multiplicidad $k$. Si $k = 1$, $\Gamma$ se denomina ciclo límite simple.

Observemos que las bifurcaciones que habíamos estudiado para los puntos fijos focos podemos aplicarlas ahora a puntos fijos de la aplicación de Poincaré que sean focos. La generalización inmediata implica que un ciclo límite puede bifurcarse en hasta $k$ ciclos límite, si $k$ es su multiplicidad.

Si el sistema 4.2.1 tiene un foco en el origen, entonces es linealmente equivalente al sistema

$$
\begin{align*}
\dot{x} &= ax - by + p(x, y) \\
\dot{y} &= bx + ay + q(x, y)
\end{align*}
$$

con $b \neq 0$.

Teorema

Sea $P(s)$ la aplicación de Poincaré para un foco en el origen del sistema analítico en dos dimensiones 4.3.2, con $b \neq 0$ y supongamos que $P(s)$ está definida para $0 < s < \delta_0$. Entonces, existe $\delta > 0$ tal que $P(s)$ se puede extender a una función analítica definida para $|s| < \delta$. Además,

$$
P(0) = 0, \quad P'(0) = \exp\left\{\frac{2\pi i a}{|b|}\right\}
$$

y si $d(s) = P(s) - s$, entonces el producto

$$
d(s) d(-s) < 0
$$

para $0 < |s| < \delta$. El hecho de que $d(s) d(-s) < 0$ para $0 < |s| < \delta$ se puede utilizar para demostrar que si

$$
d(0) = d'(0) = \ldots = d^{(k-1)}(0) = 0 \quad y \quad d^{(k)}(0) \neq 0
$$

entonces $k$ es impar, es decir, $k = 2m + 1$. El entero $m = (k - 1)/2$ se denomina multiplicidad del foco. Si $m = 0$ tenemos un foco simple, y se sigue del teorema anterior que el sistema 4.3.2 con $b \neq 0$ tiene un foco simple en el origen si, y sólo si, $a \neq 0$.

El signo de $d'(0)$, es decir, el signo de $a$, determina la estabilidad del origen en este caso. Si $a < 0$ el origen es un foco estable, y si $a > 0$ es inestable. Si $d'(0) = 0 (a = 0)$, entonces 4.3.2 tiene un foco múltiple o un centro en el origen. Si $d'(0) = 0$, entonces la primera derivada no nula

$$
\sigma \equiv d^{(k)}(0) \neq 0
$$

se llama número de Lyapunov para el foco. Si $\sigma < 0$ el foco es estable, y si $\sigma > 0$ el foco es inestable.

Es con frecuencia muy complicado llegar a una expresión analítica para la aplicación de Poincaré, y muy a menudo ni siquiera existe (analíticamente). Sin embargo, la imagen geométrica que la aplicación de Poincaré aporta es de suma utilidad en el estudio de los sistemas dinámicos, especialmente cuando la dimensión del sistema es elevada. Obsérvese que es necesario resolver
las ecuaciones diferenciales si se pretende obtener $P(s)$ analíticamente. No obstante, también se puede resolver las ecuaciones numéricamente y estudiar entonces la sección de Poincaré, a fin de obtener información sobre el sistema dinámico. Es suficiente considerar en este caso un hiperplano o una hypersuperficie que corte las trayectorias del sistema. No es necesario que la intersección se realice exactamente en forma perpendicular. Se dice entonces que la hypersuperficie (en general) $\Sigma$ corta las trayectorias de forma transversal. La sección de Poincaré resultante de una intersección transversal es topológicamente equivalente a la que resulta de una intersección perpendicular, así que las características cualitativas del sistema permanecen invariantes.

4.2.2 Análisis cualitativo y numérico de la SP

Realicemos una breve descripción del aspecto de la sección de Poincaré y lo que de ella se deduce. Si observamos una serie de puntos que convergen a uno fijo, independientemente de la condición inicial, es inmediato concluir que nuestro sistema dinámico posee un ciclo límite estable. También podríamos observar que los puntos de la sección oscilan alternativamente entre dos, tres, cuatro o más puntos, y los repiten siempre en el mismo orden. En este caso, estaríamos ante órbitas de periodo dos, tres, cuatro o superior, respectivamente. En algunas ocasiones los puntos de la sección pueden acabar llenando una curva cerrada. Nos hallamos en este caso ante una órbita cuasiperiódica. Las trayectorias del sistema se moverán sobre un toro $n-$dimensional. Este tipo de dinámica será descrito en las próximas secciones.

Veamos con un ejemplo simple cómo el estudio de la sección de Poincaré puede proporcionar considerable información sobre la dinámica de un sistema. Consideremos un péndulo unidimensional forzado por una fuerza exterior periódica que depende de la posición del péndulo,

$$\frac{d^2 x}{dt^2} + d \sin(x) = \cos(kx - wt)$$

d = g/l, donde g es la constante de la gravedad y l es la longitud del péndulo. Si se escoge $k \equiv 1$ y $d \equiv 10$, por ejemplo, se puede utilizar la frecuencia $w$ como parámetro libre del sistema, y estudiar la variación de la dinámica en función de él. En particular, es posible obtener transiciones caos-cuasiperiódico-periódico y también en sentido contrario. En la figura 4.5 se han representado cuatro secciones de Poincaré para diversos valores de $w$. En este caso, se resolvió la ecuación diferencial numéricamente y se representaron en la sección de Poincaré los puntos que cumplían las condiciones

$$|x| < \delta, \quad x < 0$$

con $\delta \approx 0.001$. Con este convenio nos aseguramos que los puntos sean prácticamente los correspondientes a la intersección de la trayectoria con el plano $x = 0$ y consideramos únicamente los que atraviesan dicho plano en una única dirección.

En algunos casos, la sección de Poincaré está formada por un conjunto de puntos que ocupan con densidad variable toda la zona de la sección, y en un orden aparentemente impredecible. Estamos entonces, muy probablemente, ante un sistema dinámico que exhibe caos determinista. Con frecuencia, la sola observación de una trayectoria (solución de un sistema dinámico) en $\mathbb{R}^3$ (por ejemplo) no es capaz de distinguir entre cuasiperiódico y caos. La sección de Poincaré, en cambio, discrimina perfectamente los dos casos. Es recomendable, en el caso de realizar estudios numéricos, observar el comportamiento del sistema a largo plazo. La representación gráfica que se realice de la dinámica puede ser poco clara si se representa el estado transitorio, esto es, los

---

2En general, sin embargo, se exige alguna otra evidencia para afirmar que un sistema presenta caos determinista.
Figura 4.5: Secciones de Poincaré para el sistema $\frac{d^2z}{dt^2} + d \sin(z) = \cos(kx - wt)$, con $k = 1, d = 10$ y, de izquierda a derecha y de arriba abajo, $w = 2.95, 2.98, 2.99, 3.00$. En este caso, el incremento en la frecuencia externa provoca una transición del caos a la cuasiperiodicidad. El atractor correspondiente a la órbita total del sistema sería topológicamente equivalente a la superficie engendrada por la rotación de la sección de Poincaré alrededor, por ejemplo, del eje $y$ (en el caso del atractor caótico esto no sería del todo cierto, como el lector puede imaginar).
Atractores Periódicos y Cuasiperiódicos

estados por los que el sistema pasa antes de llegar a un punto fijo, a un ciclo límite, etc. Por tanto, debe eliminarse un cierto intervalo inicial de tiempo en la simulación a fin de determinar el estado más probable en que el sistema será realmente observado. Considerése, por ejemplo, el caso de la aplicación logística. Cerca de los puntos donde se producen las bifurcaciones de periodo $n$ a periodo $2n$, se observa que la dinámica conduce al sistema lentamente a un atractor de periodo $n$. El estado transitorio es largo, y un estudio que analizase pocos pasos de tiempo podría conducir a conclusiones erróneas sobre la dinámica.

4.3 Teorema de Poincaré-Bendixson

El teorema de Poincaré-Bendixson fue enunciado por estos autores a principios del siglo. Sólo es válido para sistemas en $\mathbb{R}^2$, y permite demostrar la existencia de un ciclo límite sin necesidad de hallar explícitamente la expresión de éste.

Definamos primeramente lo que se entiende por conjuntos $\alpha-$ y $\omega-$límite de un sistema dinámico.

Consideremos el sistema autónomo

$$\dot{x} = F(x)$$ (4.4.1)

con $F \in C^1(E)$ y $E$ un subconjunto abierto de $\mathbb{R}^2$. Llamaremos $\Phi(t, x)$ a la solución del sistema dinámico 4.4.1 en $E$. Para $x \in E$, la función $f(t, x) : \mathbb{R} \to E$ define una curva solución, trayectoria $u$ órbita de 4.4.1 que pasa por $x$ en $E$.

Definición

Un punto $p \in E$ es un punto $\omega-$límite de la trayectoria $\Phi(\cdot, x)$ del sistema 4.4.1 si existe una secuencia $t_n \to \infty$ tal que

$$\lim_{n \to \infty} \Phi(t_n, x) = p$$

Igualmente, si existe una secuencia $t_n \to -\infty$ tal que

$$\lim_{n \to -\infty} \Phi(t_n, x) = q$$

y el punto $q \in E$, entonces $q$ se denomina punto $\alpha-$límite de la trayectoria $\Phi(\cdot, x)$ del sistema 4.4.1.

El conjunto de todos los puntos $\omega-$límite de una trayectoria de 4.4.1 ($\Gamma$) se denomina conjunto $\omega-$límite (de la trayectoria) y se denota por $\omega(\Gamma)$. El conjunto de todos los puntos $\alpha-$límite de una trayectoria es el conjunto $\alpha-$límite, $\alpha(\Gamma)$. El conjunto unión

$$\alpha(\Gamma) \cup \omega(\Gamma)$$

se llama conjunto límite de $\Gamma$.

Se puede demostrar que $\alpha(\Gamma)$ y $\omega(\Gamma)$ son subconjuntos cerrados de $E$, y que si la trayectoria $\Gamma$ está contenida en un subconjunto compacto de $\mathbb{R}^n$, entonces $\alpha(\Gamma)$ y $\omega(\Gamma)$ son subconjuntos de $E$ no vacíos, compactos y conexos.
Teorema

(de Poincaré-Bendixson)

Supongamos que \( F \in C^1(E) \), donde \( E \) es un subconjunto abierto de \( \mathbb{R}^2 \) y que 4.4.1 tiene una trayectoria \( \Gamma \) contenida en un subconjunto compacto \( F \subset E \). Entonces, si \( \omega(\Gamma) \) no contiene ningún punto crítico de 4.4.1, \( \omega(\Gamma) \) es una órbita periódica de 4.4.1.

Veamos a continuación con un ejemplo cómo es posible aplicar el teorema anterior para determinar la existencia de un ciclo límite en el espacio de fases de un sistema dinámico.

Ejemplo

Consideremos el sistema

\[
\begin{align*}
\dot{x} &= x - y - x^3 \\
\dot{y} &= x + y - y^3
\end{align*}
\]

Su único punto fijo es el origen, \( x_0 = (0,0) \). Excluyendo, pues, el origen, podemos intentar encontrar una zona de \( \mathbb{R}^2 \) que contenga \( \omega(\Gamma) \). Escibamos primeramente (como empieza a ser habitual en el sistema en coordenadas polares. Sustituyendo en

\[
\begin{align*}
\dot{\theta} &= \frac{1}{r} [\dot{x} \cos \theta - \dot{y} \sin \theta] \\
\dot{r} &= \dot{x} \cos \theta + \dot{y} \sin \theta
\end{align*}
\]

(expressiones válidas en general para cualquier cambio a coordenadas polares). con \( x = r \cos \theta \), \( y = r \sin \theta \), obtenemos para el sistema anterior

\[
\begin{align*}
\dot{\theta} &= -1 - r^2 \sin \theta \cos \theta \\
\dot{r} &= r - r^3 (\cos^4 \theta + \sin^4 \theta)
\end{align*}
\]

La función \( f(\theta) = (\cos^4 \theta + \sin^4 \theta) \) está acotada entre los valores \( \frac{1}{2} \) y 1, es decir \( f(\theta) \in \left[ \frac{1}{2}, 1 \right] \). Utilizamos estos dos valores para determinar si, para algunos valores de \( r \), sería posible encontrar \( \dot{r} < 0 \) en cierta circunferencia \( r > r_0 \) en otra. Obsérvese que, si es posible determinar una zona con \( \dot{r} < 0 \), esto significa que todas las trayectorias "entrán" en este dominio. De forma equivalente, si encontramos un dominio con \( \dot{r} > 0 \), sabemos que las trayectorias del sistema "salen" de la zona establecida.

Consideremos, pues, los valores extremos que \( \dot{r} \) puede tomar:

\[
\dot{r} = r - \frac{r^3}{2}, \quad \dot{r} = r - r^3
\]

Busquemos una zona en la que se cumplan las dos desigualdades \( r - r^3/2 < 0 \) y \( r - r^3 < 0 \) simultáneamente. Para \( \dot{r} < 0 \), obtenemos \( r > \sqrt{2} \) en el primer caso y \( r > 1 \) en el segundo. Esto significa que en la circunferencia \( r = \sqrt{2} + \varepsilon \), con \( \varepsilon \) arbitrariamente pequeño, \( \dot{r} \) es siempre negativo para cualquier valor de \( \theta \), con lo cual podemos asegurar que el conjunto \( \omega(\Gamma) \) se halla en el interior de la circunferencia limitada por \( r = \sqrt{2} + \varepsilon \). Por otra parte, cambiando el signo de las desigualdades obtenemos que \( \dot{r} > 0 \) siempre para \( r = 1 - \varepsilon \), de nuevo con \( \varepsilon \) arbitrariamente pequeño. Así que todas las trayectorias salen de la zona \( r = 1 - \varepsilon \). En consecuencia, tenemos \( \omega(\Gamma) \) confinado a la corona definida por

---

\(^3\)Estos valores se obtienen calculando los extremos de \( f(\theta) \) y viendo que tiene dos máximos, en \( \theta = 0 \) y \( \theta = \pi/2 \), de valor \( 1 \) y un mínimo en \( \theta = \pi/4 \), de valor \( 1/2 \).
Figura 4.6: Ciclo límite correspondiente al ejemplo de la sección 4.4. Las líneas discontinuas señalan las dos circunferencias de las cuales las trayectorias "entran" o "salen", que pueden ser calculadas analíticamente y permiten la aplicación del teorema de Poincaré-Bendixson.

\[ (r_{\text{min}}, r_{\text{max}}) = (1 - \epsilon, \sqrt{2} + \epsilon). \forall \theta \]

Como el único punto fijo del sistema es el origen de coordenadas, no hay ningún punto fijo en la corona anterior, así que por el teorema de Poincaré-Bendixson podemos asegurar que existe un ciclo límite en esta área. Véase el mapa de fases del sistema dinámico en la figura 4.6.

Todo lo hasta aquí expuesto sobre la existencia de ciclos límite y cómo pueden ser determinados tiene una traducción muy sencilla que resulta ser correcta la mayor parte de las veces cuando tratamos con sistemas físicos. Imaginemos que en cierto sistema real sabemos que las trayectorias no pueden escapar al infinito, es decir, la solución está forzosamente acotada a cierta zona del espacio de fases. Si, además, en esta zona el único punto de equilibrio que tenemos es inestable, podemos casi asegurar que la configuración descrita del sistema físico lleva a la existencia de un ciclo límite para la dinámica, aunque en ocasiones el supuesto ciclo límite puede ser una órbita periódica, cuasiperiódica o un atractor extraño. Pongamos un ejemplo. Supongamos que disponemos de un péndulo en el extremo del cual colocamos un imán que oscilará con él. Supongamos también que en la vertical del péndulo, allí donde habría un punto de equilibrio estable (debido al rozamiento con el aire el péndulo acabaría deteniéndose en posición vertical) colocamos un segundo imán con polaridad opuesta al primero. El punto de equilibrio estable es de esta forma inestabilizado. Y, además, el péndulo no puede escapar al infinito (como es fácil de imaginar). La intuición nos dice que posiblemente el péndulo oscilará alrededor de este punto inestable. Así, efectivamente, sucede. Y como éste, muchos otros sistemas: reacciones químicas (las concentraciones de los reactivos son finitas y algunas tienen valores fijos inestables), neuronas (potencial eléctrico finito, estado de reposo inestable), ...

### 4.4 Atractores cuasiperiódicos

Se ha visto, en el capítulo sobre sistemas dinámicos y en este mismo, la dinámica a largo término que pueden presentar los sistemas dinámicos: puntos fijos y ciclos límite. En la escala que acaba
lllevando a los atractores extraños y a la dinámica caótica nos falta el caso que nos ocupa en esta sección: los atractores (la dinámica) cuasiperiódicos.

Veremos que este tipo de movimiento tiene un papel esencial en el capítulo que dedicaremos al caos hamiltoniano, donde trataremos con sistemas conservativos. En las páginas siguientes veremos algunos sistemas dissipativos que presentan dinámica cuasiperiódica, en los cuales, por tanto, la dinámica es atractiva, y no se da sobre (como veremos) un toro n-dimensional desde el principio, sino que tiende asintóticamente hacia él.

Podemos pensar en el movimiento cuasiperiódico como en una composición de movimientos periódicos, de diversas frecuencias cuya superposición (en ciertos casos) podrá conducir al sistema a los citados atractores. Hablaremos de N-cuasiperiodicidad cuando el número de frecuencias Ωi, constituyentes del movimiento cuasiperiódico sea N. Las variables dinámicas del sistema de la forma f(t) estarán representadas por funciones de N variables independientes G(t₁, t₂, ..., tₙ), tales que serán funciones periódicas en cada una de sus variables,

\[ G(t₁, t₂, ..., tₙ + Tᵢ, ..., tₙ) = G(t₁, t₂, ..., tₙ) \]

con un periodo Tᵢ correspondiente a cada una de las variables. Las N frecuencias implicadas deben de ser incommensurables, es decir ninguna de ellas debe de poder ser expresada como combinación lineal de las demás. Se cumplirá por tanto que

\[ m₁Ω₁ + m₂Ω₂ + ... + mᵩΩᵩ = 0 \]

donde \( mᵩ ∈ ℤ \) no debe poseer ninguna solución excepto la trivial \( mᵩ = 0, ∀ᵩ \). En términos de G, las variables del sistema se podrán representar como

\[ f(t) = G(t₁, t₂, ..., tₙ) \]

de decir, f(t) corresponde al valor de G con todas sus variables \( tᵩ = t \). Debido a su periodicidad, G podrá ser fácilmente expresada como una serie de Fourier de N factores,

\[ G = \sum_{n₁,n₂, ..., nₙ} a_{n₁...nₙ} \exp \{ i(n₁Ω₁t₁ + n₂Ω₂t₂ + ... + nₙΩₙtₙ) \} \]

Tomando \( tᵩ = t \) y realizando la transformada de Fourier sobre la función G obtenemos

\[ \hat{f}(w) = 2π \sum_{n₁,n₂, ..., nₙ} a_{n₁...nₙ} \delta (w - (n₁Ω₁t₁ + n₂Ω₂t₂ + ... + nₙΩₙtₙ)) \]

Así que la transformada de Fourier de una variable dinámica f(t) está formada por todas las combinaciones lineales enteras de las N frecuencias fundamentales, Ω₁, ..., Ωₙ. El espectro de potencias (el cuadrado de la transformada de Fourier) correspondiente a un movimiento cuasiperiódico presentará, por tanto, unos picos característicos a estos valores concretos. Los picos son más agudos cuanto menor es el valor de los coeficientes \( n₁, ... nₙ \). A medida que éstos aumentan de valor, los picos disminuyen de amplitud hasta que, si trabajamos con un sistema real, serán indistinguibles del ruido inherente al dispositivo.

4.4.1 2-cuasiperiodicidad

Consideremos brevemente el caso particular en el que sólo dos frecuencias están implicadas en el movimiento del sistema, Ω₁ y Ω₂. Supongamos que la solución de la dinámica del sistema, x, tiene componentes vectoriales dados por la ecuación

\[ x^{(i)}(t) = G^{(i)}(t₁, t₂) \]
Figura 4.7: Espectro de potencias correspondiente a un movimiento periódico, a uno cuasiperiódico con dos frecuencias características y a uno caótico. Los picos limpios que se aprecian en el segundo caso desaparecen en el segundo, ya que en el caso de tener dinámica caótica se obtiene un espectro amplio, con todos los valores de las frecuencias representados.
Considerando que \( t_1 \) y \( t_2 \) son funciones periódicas con periodos \( T_1 \) y \( T_2 \), respectivamente, sólo será necesario especificar el valor de \( t \), módulo \( T_1 \), es decir, las variables \( t \), pueden ser consideradas como ángulos, y por tanto definimos las variables angulares

\[
\theta_j = \Omega_j t_j \pmod{2\pi}
\]

El estado del sistema puede ser especificado dando únicamente dos ángulos. Desde el punto de vista geométrico, la especificación de un ángulo se puede pensar como la especificación de un punto sobre un círculo. La especificación de dos ángulos es equivalente a la especificación de un punto sobre un toro 2-dimensional \( (T^2) \), así que podemos pensar que las trayectorias cuasiperiódicas están confinadas (asintóticamente) a la superficie de este toro. Efectivamente, el movimiento cuasiperiódico tiene lugar sobre una superficie 2-dimensional topológicamente equivalente a este toro.

Observemos que se ha exigido anteriormente que las frecuencias fuesen incommensurables a fin de obtener un movimiento realmente cuasiperiódico. Si el cociente

\[
R = \frac{\Omega_1}{\Omega_2}
\]

llamado número de rotación fuese un número racional de la forma \( p/q \), entonces las trayectorias sobre el toro acabarían cerrándose sobre ellas mismas, y el movimiento sería periódico, con periodo

\[
pT_1 = qT_2
\]

Cuando \( R \) es irracional, el movimiento es cuasiperiódico, y las órbitas acaban llenando densamente la superficie toroidal.

En el caso de que el movimiento fuese \( N \)-cuasiperiódico, podríamos igualmente considerar \( N \)-toros \( (T^N) \), sobre los cuales se daría ahora el movimiento que estaría especificado por \( N \) variables angulares.

**4.4.2 La aplicación circular**

Consideremos el sistema dinámico continuo

\[
\frac{d\theta^{(1)}}{dt} = \Omega_1
\]
\[
\frac{d\theta^{(2)}}{dt} = \Omega_2
\]

donde las variables angulares son

\[
\theta^{(1)}(t) = \Omega_1 t + \theta_0^{(1)}
\]
\[
\theta^{(2)}(t) = \Omega_2 t + \theta_0^{(2)}
\]

Si tomamos una sección de la trayectoria, la correspondiente por ejemplo a \( \theta^{(2)} \) módulo \( 2\pi \) = \( ct \), obtenemos una aplicación unidimensional para \( \theta_n = \theta^{(1)}(t_n) \) módulo \( 2\pi \),

\[
\theta_{n+1} = (\theta_n + w) \pmod{2\pi}
\]

con \( w = 2\pi \Omega_1 / \Omega_2 \). Siempre que las frecuencias sean incommensurables, y para cualquier condición inicial, la órbita que se obtendrá llenará densamente el círculo unidad, mientras que si \( \Omega_1 / \Omega_2 = p/q \), un valor racional, entonces la órbita será periódica con periodo \( q \). Las órbitas periódicas sólo se
obtienen para un conjunto de valores de \( w \) con medida de Lebesgue cero, lo cual significa que estos valores tienen una probabilidad a priori cero de aparecer.

Esta imposibilidad práctica de obtener órbitas periódicas cambia cuando se introduce algún tipo de término no lineal, algún acoplamiento entre las frecuencias que puede inducir un tipo de resonancia y provocar en consecuencia órbitas periódicas con probabilidad no nula. A fin de estudiar el efecto del acoplamiento en el sistema, Arnold introdujo en 1965 la siguiente aplicación modificada

\[
\theta_{n+1} = (\theta_n + w + k \sin (\theta_n)) \mod 2\pi
\]

llamada aplicación circular. Los parámetros \( w \) y \( k \) juegan un papel fundamental en la dinámica que presenta esta aplicación, así como el número de rotación, que en este caso está dado por

\[
R = \frac{1}{2\pi} \lim_{m \to \infty} \frac{1}{m} \sum_{n=0}^{m-1} \Delta \theta_n
\]

donde \( \Delta \theta_n = w + k \sin \theta_n \). Se plantea ahora el problema de distinguir cuáles serán los valores de \( k \), que se utiliza como parámetro variable en el sistema habitualmente, que producirán dinámica cuasiperiódica, es decir, un valor irracional para \( R \). La obtención de estos valores en el dominio \( k \approx 0 \) plantea el llamado problema de los denominadores pequeños, con el que nos volveremos a encontrar en el capítulo sobre caos hamiltoniano. El problema en cuestión se produce cuando intentamos realizar un desarrollo para \( k \) pequeña, alrededor de cero. En este desarrollo aparecen términos de la forma

\[
\sum_m \frac{A_m}{m} \exp(i m \theta)
\]

Para cualquier valor de \( R \) irracional podemos encontrar un valor de \( m \) que nos haga el denominador anterior tan pequeño como deseemos, ya que cuando el producto \( Rm \approx n \), con \( n \) un entero, la parte exponencial se acerca arbitrariamente a 1. El problema que entonces se presenta es el de establecer la convergencia de la serie que se estudia. Si efectivamente \( Rm \approx n \), obtenemos

\[
1 - \exp(i m \theta) \approx -2\pi i (mR - n)
\]

así que la contribución de este término a la serie es aproximadamente

\[
\frac{1}{2\pi m} \left| \frac{A_m}{m} \right| \approx |A_m| R - n/m
\]

Estudiemos el caso en que se producirá movimiento cuasiperiódico, esto es, cuando \( R \) sea irracional. Cuando el valor de \( R \) satisface

\[
|R - \frac{n}{m}| > \frac{K}{m^{2+\epsilon}}
\]

para algún valor positivo de \( K \) y \( \epsilon \) y para cualquier valor de \( m \) y \( n \), \( m \neq 0 \) se dice que \( R \) es un número mal aproximable por racionales. Sin embargo, los coeficientes \( |A_m| \) se obtienen del desarrollo en serie de Fourier de una función analítica, con lo cual presentan un decaimiento exponencial con \( m \),

\[
\frac{1}{2\pi m} \left| \frac{A_m}{m} \right| < O(m^{1+\epsilon}) \exp\left(-\sigma|m|\right)
\]

lo que implica que la suma convergerá para todos los valores de \( R \) que sean mal aproximables por racionales. Por desgracia, la convergencia de esta suma no implica la convergencia de la serie perturbativa.
Se puede obtener que la medida de Lebesgue para los valores de \( w \) que presentan cuasiperiodicidad en el intervalo \( w \in [0, 2\pi] \) es diferente de cero para \( k \) pequeña, y que tiende a \( 2\pi \) cuando \( k \to 0 \). Para \( k \) pequeña se sigue dando cuasiperiodicidad, y los valores de \( w \) para los que se obtiene ocupan la mayor parte de la medida de Lebesgue en el intervalo de valores estudiado.

### 4.5 Apéndice

Incluimos en este apéndice un caso más detallado de análisis de un sistema dinámico. Creemos que el ejemplo ilustra muy bien la forma en que se puede calcular explícitamente la ecuación de una variedad centro, y la dificultad de cálculo no es excesiva.

Consideraremos el sistema

\[
\begin{align*}
\dot{x} &= xz - wy \\
\dot{y} &= wx + yz \\
\dot{z} &= p + z - \frac{1}{3} z^3 - (x^2 + y^2)(1 + qx + ez)
\end{align*}
\] (A.1)

que proviene del modelo de Langford para el flujo de Couette-Taylor \(^4\). Los valores de las variables están fijados a \( w = 10, e = 0.5, q = 0.7 \), y \( p \) actúa como parámetro de bifurcación.

El primer paso es el cálculo de los puntos fijos del sistema. Las dos primeras ecuaciones proporcionan

\[
x = 0, \quad y = 0
\]

y de la tercera obtenemos una curva que dará el valor de \( z \) en función del parámetro \( p \),

\[
p + z - \frac{1}{3} z^3 = 0
\] (A.2)

Esta ecuación tiene diferente número de soluciones (reales, que son las que nos interesan) dependiendo de \( p \). Efectivamente,

- Para \( p < -2/3 \) o \( p > 2/3 \), existe un único punto fijo.
- Para \( p = \pm 2/3 \) existen dos puntos fijos.
- Para \(-2/3 < p < 2/3 \) existen tres puntos fijos.

En la figura 4.8 se representa la curva de soluciones para \( z \) en función del parámetro.

Estudiemos la matriz lineal para conocer la estabilidad de los puntos fijos. En general, para un valor de \( p \) dado, el punto fijo será de la forma

\[
x^*(p) = (0, 0, z_0)
\]

con lo cual la matriz lineal en este punto correspondiente al sistema A.1 quedará de la forma

\[
A = \begin{pmatrix}
z_0 & -w & 0 \\
w & z_0 & 0 \\
0 & 0 & 1 - z_0
\end{pmatrix}
\]

Figura 4.8: Soluciones para la variable del sistema A.1 en función del parámetro $p$.

que tiene como valores propios

$$
    \lambda_1 = z_0 + wi, \quad \lambda_2 = z_0 - wi, \quad \lambda_3 = 1 - z_0^2
$$

y como vectores propios

$$
    v_1 = (i, -1, 0), \quad v_2 = (-i, -1, 0), \quad v_3 = (0, 0, 1)
$$

Los valores propios permiten clasificar los puntos fijos según su estabilidad, dependiendo del parámetro $p$.

- **$p < -2/3 \implies z_0 < -2$, y en este caso existe un único punto fijo, para el cual resulta $\Re(\lambda_1) < 0, \Re(\lambda_2) < 0, \lambda_3 > 0$, así que el punto fijo es inestable.**

- **$p = -2/3 \implies z_0 = 1, -2$ (existen dos puntos fijos ahora, el de valor $z_0 = 1$ es un punto fijo doble), para los cuales resulta:**
  1. $z_0 = 1$: $\lambda_1 = 1 + wi, \lambda_2 = 1 - wi, \lambda_3 = 0$, el punto fijo es inestable
  2. $z_0 = -2$: $\lambda_1 = -2 + wi, \lambda_2 = -2 - wi, \lambda_3 = -3$, el punto fijo es estable.

- **$-2/3 < p < 0$ tenemos los siguientes dominios de estabilidad, correspondientes a las diferentes soluciones de A.2:**
  1. $-2 < z_0 < -\sqrt{3}$: $\Re(\lambda_1) < 0, \Re(\lambda_2) < 0, \lambda_3 < 0$, el punto fijo será estable
  2. $1 < z_0 < \sqrt{3}$: $\Re(\lambda_1) > 0, \Re(\lambda_2) > 0, \lambda_3 < 0$, el punto fijo será inestable
  3. $0 < z_0 < 1$: $\Re(\lambda_1) < 0, \Re(\lambda_2) < 0, \lambda_3 > 0$, el punto fijo será también inestable, aunque la dimensionalidad de las variedades asociadas cambia.

- **$p = 0$ obtenemos tres puntos fijos,**
  1. $z_0 = 0$: $\lambda_1 = iv, \lambda_2 = -iv, \lambda_3 = 1$, punto fijo inestable
  2. $z_0 = \sqrt{3}$: $\Re(\lambda_1) > 0, \Re(\lambda_2) > 0, \lambda_3 < 0$, punto fijo inestable
  3. $z_0 = -\sqrt{3}$: $\Re(\lambda_1) < 0, \Re(\lambda_2) < 0, \lambda_3 < 0$, punto fijo estable
\[ 0 < p < 2/3 \] tenemos otros tres dominios de estabilidad, también correspondientes a las soluciones de A.2:

1. \( -\sqrt{3} < z_0 < -1 \): \( \Re(\lambda_1) < 0, \Re(\lambda_2) < 0, \lambda_3 < 0 \), el punto fijo será estable
2. \( \sqrt{3} < z_0 < 2 \): \( \Re(\lambda_1) > 0, \Re(\lambda_2) > 0, \lambda_3 < 0 \), el punto fijo será inestable
3. \( -1 < z_0 < 0 \): \( \Re(\lambda_1) < 0, \Re(\lambda_2) < 0, \lambda_3 > 0 \), el punto fijo será también inestable

\[ p > 2/3 \implies z_0 > 2 \] y resulta en este último caso \( \Re(\lambda_1) > 0, \Re(\lambda_2) > 0, \lambda_3 < 0 \), y el punto fijo será inestable.

Estudiaremos a continuación la variedad centro que se produce cuando el parámetro \( p \) pasa de ser positivo a ser negativo. Obsérvese en la relación anterior que cuando \( p = 0 \) se obtiene que el punto fijo \( x^*(p = 0) = (0, 0, 0) \) tiene asociados dos valores propios nulos (con cual aparecerá una variedad centro de dimensión dos) y un valor propio positivo (correspondiente a una variedad inestable unidimensional).

Necesitamos en primer lugar escribir el sistema A.1 en forma de Jordan (o forma normal), para lo cual utilizaremos el cambio que proporcionan los vectores propios y aplicaremos una traslación en la variable \( z \),

\[
\begin{align*}
    x' &= ix - y \\
    y' &= -ix - y \\
    z' &= z + p
\end{align*}
\]

y así podemos escribir el sistema A.1 en la forma

\[
\begin{align*}
    \dot{z} &= (iw - p)x + xz \\
    \dot{y} &= (-iw - p)y + yz \\
    \dot{z} &= z - \frac{1}{3}(z - p)^3 - yx \left( 1 + i\frac{q}{2}(y - z) + e(z - p) \right)
\end{align*}
\]  

 donde se han suprimido las primas por comodidad. Vemos que la variedad centro está determinada por las dos primeras variables, \( x \) y \( y \). Localmente, sabemos que en la variedad centro se puede escribir la variable \( z \) como función de \( x \) y de \( y \). Consideremos un desarrollo de \( z \) a segundo orden como función de \( x \) e \( y \) para encontrar, siguiendo las ecuaciones dinámicas A.3, la ecuación a segundo orden que corresponderá a la variedad centro,

\[
\begin{align*}
    z &= ap^2 + bpz + cp + dz^2 + fxy + gy^2 + O(3)
\end{align*}
\]

 donde \( a, b, c, d, f, g \) son coeficientes que han de ser determinados. Por sustitución directa en

\[
\begin{align*}
    \dot{z} &= z - \frac{1}{3}(z - p)^3 - yx \left( 1 + i\frac{q}{2}(y - z) + e(z - p) \right)
\end{align*}
\]

 e igualando los términos con las mismas variables obtenemos ecuaciones que nos permiten calcular los coeficientes del desarrollo de \( z \) (indicamos al principio el término que produce cada ecuación),

- \( p^2 \): \( a = 0 \)
- \( px \): \( iwb = b \implies b = 0 \)
- \( py \): \( -iwc = c \implies c = 0 \)
\begin{itemize}
  \item $x^2$: $2diw = d \Rightarrow d = 0$
  \item $y^2$: $-2giw = g \Rightarrow g = 0$
  \item $xy$: $fiw = fiw = f - 1 \Rightarrow f = 1$
\end{itemize}

Así que obtenemos que, localmente (para $p \approx 0$), y a segundo orden

$$z = xy$$

y podemos escribir las ecuaciones de la dinámica reducidas a la variedad centro como

$$\dot{z} = (iw - p)x + x^2y$$
$$\dot{y} = (-iw - p)y + xy^2$$

En este caso particular se puede ver que no es posible determinar la estabilidad del punto fijo con la aproximación a segundo orden, lo cual haría necesario el cálculo de la variedad centro a un orden superior si éste fuese nuestro propósito. Se ha visto el mecanismo esencial: desarrollo de $z$ en función de $x$, $y$ y $p$ al orden deseado y, siguiendo las ecuaciones dinámicas, cálculo de la variedad centro. Se puede encontrar en la bibliografía especializada los teoremas que permiten establecer la estabilidad del punto fijo en cada caso.

### Bibliografía

Junto con algunos de los textos citados en el capítulo 2, se puede consultar


Capítulo 5

Caos Determinista

Hemos analizado en capítulos anteriores el comportamiento de sistemas que exhiben atractores periódicos o puntuales. Los ejemplos dados hacen referencia por lo tanto a sistemas intrínsecamente predecibles. En este sentido, se asumió desde el principio la validez de la implicación:

Determinismo $\implies$ Predicción

Las herramientas básicas presentadas anteriormente nos permiten en muchos casos conocer la dinámica de las soluciones, cuento menos en las proximidades de los puntos críticos. A la pregunta de si son éstas todas las soluciones posibles o suficientes en nuestra descripción de la realidad, la respuesta es, como veremos, decididamente negativa.

La ciencia ha tendido a clasificar los fenómenos naturales en dos grandes grupos, atendiendo a nuestra capacidad de explicar su comportamiento. En uno de ellos se hallan fenómenos que exhiben, en una u otra forma, comportamientos simples, ya sean estos estados estacionarios o periódicos. Para estos problemas se dispone de un formalismo como el anterior, y podemos predecir sin dificultad el futuro del sistema. Aparece aquí una segunda implicación, intuitivamente clara:

Dinámica simple $\implies$ Modelo simple

que, implicitamente, hizo válido que se asumiera la implicación contraria:

Modelo simple $\implies$ Dinámica simple

Por otra parte están aquellos sistemas que se comportan de forma compleja y que estudiamos empleando fundamentalmente herramientas de tipo estadístico. Para estos sistemas renunciamos a una descripción detallada y nos contentamos con promedios que, aunque pobres en detalle, nos dan, cuento menos algún tipo de información. La conclusión implícita en esta separación fue que la observación de una dinámica compleja e irregular en un sistema cualquiera era debida a una complejidad interna del mismo sistema. Si observamos un fluido turbulento, la compleja estructura de los remolinos nos da un ejemplo de esta idea. Las oscilaciones irregulares de poblaciones naturales, los cambios en el clima y el registro de la actividad cerebral serían buenos ejemplos adicionales. Tendremos oportunidad de volver a ellos más adelante. Lo que nadie esperaba, y que en gran medida ha transformado nuestra forma de observar y comprender el universo, es que la mayoría de los sistemas “complejos”, que se habían supuesto como únicamente descriptibles mediante métodos
5.1 Atractores extraños

Partamos nuevamente de nuestro sistema dinámico que, como sabemos, puede ser descrito geométricamente a través de su atractor asociado. Consideremos un ejemplo clásico que nos permitirá introducir el concepto de caos y un nuevo atractor. Este sistema, conocido como modelo de Lorenz, se define por:

\[
\frac{dx}{dt} = -\sigma(x - y) \tag{5.1.1}
\]
\[
\frac{dy}{dt} = r x - y - xy \tag{5.1.2}
\]
\[
\frac{dz}{dt} = b z + xy \tag{5.1.3}
\]

y describe de manera esquemática la dinámica de una capa de fluido que presenta una diferencia de temperatura \(\Delta T\) entre sus superficies inferior y superior. Este sistema es bien conocido clásicamente: para \(\Delta T\) bajos, el fluido transporta el calor por conducción. Para un cierto valor crítico, el fluido entra en régimen convectivo y aparece una estructura espacial (ya indicada en el capítulo anterior) que es reemplazada finalmente por un comportamiento turbulento para una diferencia aún mayor. Este comportamiento desordenado aparece en el mismo sistema que da lugar a los otros comportamientos. No hemos modificado la estructura del sistema (que sigue siendo simple) y sin embargo algo nuevo (y aparentemente muy complejo) tiene lugar.

Estudiaremos en primer lugar la estabilidad de las soluciones. Notemos que el sistema es siempre disipativo (véase el capítulo 1) puesto que para un volumen arbitrario \(V\), se tiene:

\[
\partial_t V = \int_V \text{div} F = \int_V \left[ \frac{\partial}{\partial x} \dot{x} + \frac{\partial}{\partial y} \dot{y} + \frac{\partial}{\partial z} \dot{z} \right] = -(\sigma + b + 1) V < 0
\]

puesto que las constantes se definen positivas. El volumen se contrae por lo tanto exponencialmente con el paso del tiempo.

5.1.1 Lorenz: puntos críticos y estabilidad

En primer lugar estudiaremos los puntos críticos del modelo de Lorenz y su estabilidad local. Partiendo de las ecuaciones del modelo 5.1.1, 5.1.2 y 5.1.3, los puntos fijos del sistema son:

\[
X_1 = (0, 0, 0)
\]
\[
X_{\pm} = (\pm \sqrt{b(r - 1)}, \mp \sqrt{b(r - 1)}, r + 1)
\]

El primer punto corresponde físicamente a la situación de fluido estable, sin movimiento y con transporte de calor por conducción. La matriz asociada al sistema lineal será:
Figura 5.1: Orden y caos. Clásicamente el orden queda ejemplificado por el ritmo regular de un reloj (que no es más que un oscilador armónico simple). El caos, representado por la tormenta, fue entendido como el azar imprevisible. Para comprenderlo, deberíamos ser capaces de conocer en detalle todas las variables implicadas, algo virtualmente imposible de realizar. El caos determinista, sin embargo, ha modificado completamente esta visión.

\[
L(X_1) = \begin{pmatrix}
-\sigma & \sigma & 0 \\
\sigma & -1 & 0 \\
0 & 0 & -b
\end{pmatrix}
\]

que nos proporciona los valores propios:

\[
\lambda_{1,2} = -\frac{\sigma + 1}{2} \pm \frac{1}{2} \sqrt{(\sigma + 1)^2 + 4(r - 1)\sigma}
\]

\[
\lambda_3 = -b
\]

\(X_1\) será estable si \(\Re(\lambda) < 0\), lo cual implica \(r \in (0, 1)\), e inestable en consecuencia cuando \(r > 1\). Los valores propios evidencian un punto crítico que se comporta, para \(r > 1\), como un punto de silla tridimensional.

La convección de Bénard (su equivalente en este modelo) se inicia en \(r = 1\), donde \(\lambda_1 = 0\), y aquí \(X_\pm\) entran en escena. Haremos a continuación un análisis más detallado del comportamiento de estos puntos. Realizaremos el cambio de coordenadas (véase el capítulo 2)

\[
x' = x \mp \{b(r - 1)\}^{1/2}
\]

\[
y' = y \mp \{b(r - 1)\}^{1/2}
\]

\[
z' = z - (r - 1)
\]

con lo cual las ecuaciones linealizadas son ahora:

\[
\frac{dx}{dt} = -\sigma(x - y)
\]
\[ \frac{dy}{dt} = z - y + \epsilon \]
\[ \frac{dz}{dt} = \pm \epsilon (x + y) - bz \]

donde se han suprimido las primas por comodidad y se ha definido \( \epsilon = \sqrt{b(r - 1)} \).

Los valores propios se obtendrán del polinomio característico:

\[ P(\lambda) = \lambda^3 + (\sigma + b + 1)\lambda^2 + b(\sigma + r)\lambda + 2b\sigma(r - 1) = 0 \]

Para analizar la estabilidad de esta ecuación podemos recurrir a métodos numéricos o emplear cierta álgebra simple (aunque algo intrincada). Notemos que todos los coeficientes de \( P(\lambda) \) son positivos, ya que imponemos \( r > 1 \) para que \( X_+ \) y \( X_- \) existan. Tenemos así \( P(\lambda) > 0 \) para cualquier \( \lambda > 0 \). Tendremos inestabilidad (es decir, \( \Re(\lambda) > 0 \)) únicamente si existen soluciones \( \lambda \in \mathbb{C} \) complejas conjugadas en \( P(\lambda) = 0 \). Es fácil ver que si \( r = 1 \),

\[ \lambda_1 = 0, \quad \lambda_2 = -b, \quad \lambda_3 = -(\sigma + 1) \]

y tendremos estabilidad marginal. Para \( r > 1 \) tenemos inestabilidad para algún \( \omega \) en el que \( \Re(\lambda) = 0 \), y por tanto \( \lambda_1 = i\omega \) y \( \lambda_2 = -i\omega \). Ahora bien, la suma de los tres ceros de la ecuación cúbica es:

\[ \lambda_1 + \lambda_2 + \lambda_3 = -(\sigma + b + 1) < 0 \]

y por lo tanto

\[ \lambda_3 = -(\sigma + b + 1) < 0 \]

En el punto de inestabilidad, \( \lambda_{1,2} = \pm i\omega \), y puede probarse en ese punto que

\[ r \equiv r_c = \frac{\sigma(\sigma + b + 3)}{\sigma - b - 1} \]

y tendremos inestabilidad para aquellos pares \( \{\sigma, b\} \) tales que \( r_c > 1 \). Los puntos críticos \( X_+ \) serán estables si

\[ \sigma < b + 1 \quad y \quad r > 1 \]

o bien si

\[ \sigma > b + 1 \quad y \quad r_c > r > 1 \]

Si se da la inestabilidad, sucede lo siguiente: en medida que \( r \) aumenta desde 1: \( \lambda_1 \) decrece, desde cero hasta que “colisiona” con \( \lambda_2 \) (cuando \( \lambda_1 = \lambda_2 < 0 \)), se convierten en una pareja compleja conjugada, y finalmente su parte real se hace positiva. \( \lambda_3 \) permanece negativo para todos los valores \( r > 1 \). Cada uno de los dos puntos \( X_+ \) y \( X_- \) tiene un valor propio negativo y dos con parte real positiva y complejos conjugados, cuando nos hallamos en la zona de inestabilidad. La dinámica se localiza sobre órbitas que se acercan al punto fijo en la dirección determinada por la variedad estable asociada a \( \lambda_3 \) y se alejan en forma espiral sobre la variedad inestable que determinan \( \lambda_1 \) y \( \lambda_2 \).

Podemos, pues, hallar ciertas combinaciones de parámetros que arrojarán inestabilidad para los tres puntos fijos, \( X_1 \), \( X_+ \) y \( X_- \), y entonces, dado que la dinámica está acotada en el espacio de fases y no divergirá a infinito, se presentarán situaciones dinámicas complejas, como en el caso siguiente, lo cual hace necesario recurrir a partir de aquí a un análisis global de la dinámica.
Supongamos que tomamos las siguientes constantes, $r = 28$, $\sigma = 10$, $b = 8/3$ y analizamos numéricamente el comportamiento de las soluciones. Nuestro sistema es determinista y, sin embargo, exhibe una dinámica de gran complejidad, que, de hecho, es totalmente aperiódica. La existencia de determinismo (y de un número reducido de dimensiones) queda patente cuando estudiamos la estructura del atractor, que vemos en la figura 5.2. Si seguimos la trayectoria de una condición inicial dada a lo largo del tiempo, veríamos que se mueve sobre este objeto de una forma errática, pasando de un lado a otro de forma aparentemente caprichosa. El atractor tiene volumen cero, y su dimensión estará por debajo de $d = 3$. ¿Se trata de algún ciclo límite muy complicado, restringido a un subespacio de dimensión $d = 2$? No: el atractor de Lorenz es un atractor extraño, un nuevo objeto de una estructura muy intrincada. En su interior encontraremos fractales, extrañas propiedades y, sobre todo, un nuevo fenómeno que nos acercará a uno de los problemas más apasionantes de la complejidad: el caos determinista. Antes de volver al atractor de Lorenz, daremos un rodeo alrededor de los sistemas discretos.

Señalamos que Edward Lorenz partió en su estudio inicial de sistemas de mayor complejidad, formados por 14 ecuaciones (Lorenz, 1995), que fue simplificando progresivamente. En 1971, tras asistir a un congreso acerca del problema de la turbulencia y oír a David Ruelle, que acababa de publicar un artículo junto con F. Takens (titulado “sobre la naturaleza de la turbulencia”. Ruelle y Takens, 1971) en el que por primera vez aparecía la expresión “atractor extraño”, Lorenz llevó a cabo la última simplificación que convertiría su sistema en un modelo tridimensional (véase apéndice). Como señala Lorenz, ésta una representación muy simplificada de la realidad, pero sí este modelo exhibiera un comportamiento que, en alguna forma, fuera impredecible, entonces tal vez el clima real pudiera serlo por motivos distintos a los que supondríamos (esto es, un conjunto enorme de variables). Lorenz empleó el ordenador para resolver el sistema, y vio algo realmente sorprendente: su sistema, totalmente determinista, era completamente aperiódico. Jamás se repetía y, lo que era aún más sorprendente, era intrínsecamente impredecible. El caos determinista había entrado en escena una vez más. Esta vez, de forma definitiva.
5.2 Duplicación de periodo: $f_\mu(x) = \mu x(1 - x)$

En esta sección nos detendremos a explorar el comportamiento de uno de los sistemas dinámicos más simples conocidos: la ecuación logística (May, 1976). Este sistema discreto unidimensional, que ya fue introducido en el capítulo de sistemas dinámicos, está definido por:

$$x_{n+1} = f_\mu(x_n) = \mu x_n(1 - x_n)$$  \hspace{1cm} (5.2.1)

siendo $\mu \in [0, 4]$ un parámetro y $x_n \in [0, 1]$ (aunque algunos textos exploran el comportamiento para valores mayores, véase Devaney, 1986). El sistema logístico puede generar comportamientos dinámicos de extraordinaria complejidad.

Ya vimos que el punto fijo $x^* = 0$ es stable para $\mu < 1$ e inestable en caso contrario. En $\mu = 1$ aparece un nuevo punto crítico, $x^* = 1 - 1/\mu$ que permanece estable dentro del intervalo $\mu \in (1, 3)$. Más allá de $\mu = 3$, este punto deja de ser estable. La estabilidad de estos puntos puede analizarse empleando diagramas de recurrencia $(x_n, x_{n+1})$ como los que se indican en la figura 5.3. Dibujamos la curva logística $f_\mu(x)$ dentro del intervalo unidad, así como la recta bisectriz $x_{n+1} = x_n$. Observemos que dicha recta intersecta a la curva logística en dos puntos, que son de hecho los puntos fijos antes indicados. El procedimiento para iterar sobre este diagrama es el siguiente:

- Partimos de una condición inicial dada $x_0$ (un punto del eje horizontal, $x_n$).
- Trazamos la recta que va hasta el valor de la imagen de $x_0$, esto es, un punto sobre la curva logística de valor $x_1 = f_\mu(x_0)$.
- El nuevo valor sirve de condición nueva sobre la que iterar. En lugar de volver al eje horizontal, simplemente trazamos la recta horizontal que va del último punto a la recta bisectriz, y repetimos la iteración para alcanzar de nuevo la curva en $x_2 = f_\mu(x_1)$.
- Iteramos siguiendo los pasos anteriores.

En la figura 5.3 podemos ver el resultado de nuestro experimento numérico: la trayectoria forma un diagrama que converge hacia el punto fijo $x^* = 1 - 1/\mu$. Este punto es estable, como queda de manifiesto en el comportamiento de las iteraciones sucesivas.
Figura 5.4: (a) Diagrama de iteración para \( \mu = 3.2 \), que muestra comportamiento periódico. En las proximidades del punto fijo \( z^*_1 \), la órbita se desplaza alejándose de este valor (que es inestable) para alcanzar alternativamente los puntos \( x_1 \) y \( x_2 \). (b) Diagrama de bifurcación correspondiente.

Para \( \mu > 3 \), el punto anterior se hace inestable. Si empleamos el diagrama de recurrencia anterior, encontramos que esta inestabilidad conlleva la aparición de dos nuevos puntos sobre la curva logística, que dan lugar de hecho a una trayectoria periódica, a una órbita de periodo dos

\[
O^{(2)}_\mu = \{ x_1, x_2 \}
\]

esto es, a un par de puntos tales que:

\[
x_1 = f_\mu(x_2)
\]

\[
x_2 = f_\mu(x_1)
\]

o, lo que es lo mismo.

\[
x_1 = f^{(2)}_\mu(x_1) = f_\mu[f_\mu(x_1)]
\]

\[
x_2 = f^{(2)}_\mu(x_2) = f_\mu[f_\mu(x_2)]
\]

El punto \( \mu_c = 3 \) es un punto de bifurcación y a este tipo particular se la denomina bifurcación con duplicación de periodo. En general, los puntos fijos de una órbita \( p \)-periódica forman un conjunto:

\[
O^{(p)}_\mu = \{ x^{*(p)}_i \ ; \ i = 1, \ldots, p \ | \ x^{*(p)}_i = f^{(p)}_\mu(x^{*(p-1)}_i) \}
\]

que, para el caso que nos ocupa \( (p = 2) \), se obtienen de la ecuación:

\[
x^* = f^{(2)}_\mu(x^*) = f_\mu[f_\mu(x^*)]
\]

esto es, para el caso logístico,

\[
x^* = f^{(2)}_\mu(x^*) = \mu x^*(1 - x^*)[1 - \mu x^*[1 - x^*]]
\]

lo que nos da una ecuación de cuarto grado (haremos \( x^* \equiv x \)):

\[
-\mu^3x^4 + 2\mu^2x^3 - (\mu^2 + \mu^3)x^2 + (\mu^2 - 1)x = 0
\]
Figura 5.5: (a) Diagrama de iteración para $f^{(2)}(x)$ con (a) $\mu = 2.8$ y (b) $\mu = 3.2$. Vemos la aparición de dos nuevos puntos fijos correspondientes a la órbita periódica.

aunque esta ecuación sería en principio difícilmente resoluble, dada que sus soluciones son ya conocidas: $x^*_1 = 0$ y $x^*_2 = 1 - 1/\mu$, puesto que ambas verifican (trivialmente) las condiciones de la ecuación de partida. La primera solución permite obtener:

$$-\mu^3 x^3 + 2\mu^3 x^2 - (\mu^2 + \mu^3) x + (\mu^2 - 1) = 0$$

Mientras que, dividiendo esta ecuación por $(x - x^*_1)$, obtenemos una ecuación cuadrática:

$$-\mu^3 x^2 + 2\mu^3 x - (\mu^2 + \mu^3) = 0$$

lo que nos lleva a obtener las dos nuevas soluciones a partir de

$$x^2 = \frac{1}{\mu} + \frac{1 + \mu^2}{\mu^2}$$

esto es,

$$x_{1,2} = \frac{(\mu + 1) \pm \sqrt{\mu^2 - 2\mu - 3}}{2\mu}$$

En la figura 5.4 (a) representamos el diagrama de recurrencia para $\mu = 3.2$ junto con el correspondiente diagrama de bifurcación (figura 5.4 (b)). Como es habitual, indicamos con líneas discontinuas las soluciones inestables y mediante línea continua las estables. No olvidemos que ahora las ramas que aparecen a ambos lados de $x^* = 1 - 1/\mu$ forman parte de la órbita periódica.

Podemos representar esta situación mediante un nuevo diagrama de recurrencia, ahora el correspondiente a los puntos $(x_n, x_{n+2})$ (figura 5.5). La curva es ahora $f^{(3)}_\mu(x)$ que, como vemos, tiene un comportamiento algo más complicado. Los puntos fijos de la aplicación $f^{(2)}_\mu(x)$ aparecen ahora en las intersecciones entre la nueva bisectriz $x_{n+2} = x_n$ y la curva. Para valores inferiores a $\mu = 3$, sólo aparecen dos puntos que corresponden a los dos puntos fijos de $f_\mu(x)$. Para $\mu > 3$, surgen las nuevas soluciones que se mantendrán estables hasta un nuevo valor $\mu'_\mu$.

A medida que $\mu$ aumenta, la amplitud de las oscilaciones en la ecuación logística crece (como vemos en el diagrama de recurrencia). Más adelante se da una nueva pérdida de estabilidad, la cual genera una órbita de periodo cuatro. Estas bifurcaciones con duplicación de periodo se
dan sucesivamente y con mayor rapidez, de forma que aparecen comportamientos periódicos de periodos 8, 16, 32, ..., $2^n$. Las órbitas que se han hecho inestables serán, como antes, soluciones de las nuevas ecuaciones de los puntos fijos de las nuevas órbitas, de forma que, a medida que aumenta el valor del parámetro, el número de órbitas periódicas inestables va creciendo con rapidez. En la figura 5.6 vemos el diagrama de bifurcación con duplicación de periodo (descrito por primera vez por May y Oster, 1976) en el que se aprecian las propiedades ya indicadas.

La estabilidad de las órbitas de periodo arbitrario puede obtenerse a partir del estudio de la aproximación lineal a la aplicación $f^{(2)}_\mu$, de manera similar a como se estudiaba la estabilidad de los puntos fijos del sistema de partida. Para $p = 2$, el criterio de estabilidad se establece a partir de los puntos fijos que forman la órbita, es decir, $x^*_i = f^{(2)}_\mu(x^*_i)$ (con $i = 1, 2$).

$$\lambda^{(2)} = \frac{\partial f^{(2)}_\mu(x^*_i)}{\partial x^*}$$

esto es, de la desigualdad $|\lambda^{(2)}| < 1$. Podemos obtener una expresión más manejable empleando la regla de la cadena:

$$\lambda^{(2)} = \left( \frac{\partial f^{(2)}_\mu(x)}{\partial x} \right)_{x^*_i} \left( \frac{\partial f^{(2)}_\mu(x)}{\partial x} \right)_{f^{(2)}_\mu(x)} \left( \frac{\partial f^{(2)}_\mu(x)}{\partial x} \right)_{x^*_i} \left( \frac{\partial f^{(2)}_\mu(x)}{\partial x} \right)_{x^*_i}$$

Esta expresión nos permite escribir la forma general del criterio de estabilidad para una órbita $p$-periódica arbitraria,

$$\lambda^{(p)} = \prod_{j=1}^{p} \left( \frac{\partial f^{(2)}_\mu(x)}{\partial x} \right)_{x^*_i}$$

para la que igualmente tendremos estabilidad siempre que $|\lambda^{(p)}| < 1$. 
Un resultado interesante que ya fue inicialmente observado por distintos autores (May, 1976) es el hecho de que estas órbitas de periodicidad creciente (que aparecen a través de un proceso de duplicación de periodo) se dan en un gran número de sistemas dinámicos discretos unidimensionales. Así, los siguientes sistemas muestran este comportamiento:

\[ x_{n+1} = x_n \exp\left(\mu(1 - x_n)\right) \]

\[ x_{n+1} = \frac{\alpha x_n}{1 + x_n^3} \]

\[ x_{n+1} = \sin(\pi x_n) \]

aunque la aparición de las bifurcaciones tenga lugar en valores distintos y las amplitudes difieran de las anteriores. Todas estas funciones comparten la existencia de un único máximo en el intervalo de definición de \( x \) así como su continuidad sobre dicho intervalo (más adelante veremos que existe un conjunto bien definido de puntos en común que las agrupa en una clase de universalidad, según definiremos en el capítulo 7).

Un análisis de la aparición de las órbitas periódicas sucesivas, que generan órbitas de periodo \( 2^n+1 \) a partir de la última órbita de periodo \( 2^n \), indica que el límite

\[ \delta = \lim_{n \to \infty} \frac{\mu_n - \mu_{n-1}}{\mu_{n+1} - \mu_n} \]

(que no es sino el cociente entre las distancias sucesivas en los valores de \( \mu \) que dan lugar a bifurcaciones con duplicación de periodo), converge a un valor constante dado por \( \delta \approx 4.6692 \). ... Este es un número trascendente, conocido como constante de Feigenbaum (Feigenbaum, 1978).

Observemos que, dado un primer par de valores del parámetro \( \mu \) en los que aparecen bifurcaciones consecutivas, podemos llevar a cabo una estimación de los puntos sucesivos \( \mu_n \) en los que tiene lugar una bifurcación a partir de la recurrencia:

\[ \mu_{n+1} = \mu_n + \frac{\mu_n - \mu_{n-1}}{\delta_n} \]

donde \( \delta_n \) será:

\[ \delta_n = \frac{\mu_{n-1} - \mu_{n-2}}{\mu_n - \mu_{n-1}} \]

lo que nos da, empleando el método de Newton, un valor límite para \( \mu_n \) dado por

\[ \mu_\infty \approx 3.8699456... \]

que es, como vemos, un punto de acumulación para esta serie.

Debemos preguntarnos a continuación qué puede ocurrir más allá de este valor \( \mu_\infty \). El número de órbitas periódicas inestables ha crecido indefinidamente, de forma que las trayectorias se han hecho cada vez más complicadas.

### 5.3 Caos en sistemas discretos

Más allá del punto \( \mu_\infty \) aparece el caos determinista. En la figura 5.7 vemos dos ejemplos de diagramas de recurrencia en este dominio, para \( \mu = 3.7 \) y \( \mu = 4 \). Podemos ver el caos en acción y su efecto sobre la evolución futura del sistema dinámico empleando dos condiciones iniciales muy próximas y siguiendo sus cambios a lo largo del tiempo. Tomemos la ecuación logística con \( \mu = 4 \) y las condiciones iniciales \( x_0 = 0.1000 \) y \( x'_0 = 0.1001 \), que sólo difieren entre sí en el cuarto decimal.
Figura 5.7: Caos: aplicación logística en el dominio caótico. Vemos el diagrama de recurrencia para (a) $\mu = 3.7$ y (b) $\mu = 4$.

La separación entre ambas es, por lo tanto, $\epsilon = 10^{-4}$. No debemos olvidar que el sistema dinámico que empleamos es determinista así que, en principio, tal vez esperaríamos que la evolución de ambas condiciones iniciales fuera básicamente idéntica al cabo de mucho tiempo. Podemos jir más allá: supongamos que estamos modelizando un sistema real perfectamente descrito por la ecuación logística. Imaginemos que nuestro modelo parte de $x'_0 = x_0 + \epsilon$ que, de hecho, puede imaginarse como una estimación de $x_0$ (el estado "real") con un error $\epsilon$. La pregunta es en qué medida se verá afectada nuestra predicción por el error inicial.

En la figura 5.8 (a) vemos el resultado de nuestro experimento numérico. Como era de esperar, inicialmente ambas trayectorias se solapan entre sí, sin prácticamente ninguna diferencia apreciable. En esta escala temporal podemos decir que nuestra predicción de la evolución del sistema es buena, y este resultado encaja con nuestras expectativas acerca de lo que implica el determinismo. Pero muy poco tiempo después observamos que la separación entre ambas trayectorias crece enormemente. Podemos dar una medida de la separación absoluta mediante la variable:

$$z(t) = |f_\mu^n(x_0 + \epsilon) - f_\mu^n(x_0)|$$

que también hemos representado, en la figura 5.8 (b). Observamos que, aunque $z(t)$ fluctúa enormemente (y por tanto cada trayectoria es básicamente independiente) su valor promedio es relativamente alto. Esta situación es general en los sistemas dinámicos que exhiben caos: son sistemas que presentan sensibilidad a las condiciones iniciales (SCI). Como veremos, existe una conexión profunda entre las propiedades topológicas de los atractores extraños e esta SCI.

Formalmente, podemos definir esta SCI en la forma:

**Definición**

Diremos que la aplicación $f_\mu : U \to U$ tiene sensibilidad a las condiciones iniciales (SCI) si $^1$:

$$\exists \delta > 0 \ ; \ \forall x \in U, \ \forall B(x) \ \exists \{ y \in B(x) \mid n > 0 \}$$

$^1B(x)$ es una bola abierta que contiene a $x$. Véase la sección 3.3.
Figura 5.8: Sensibilidad a las condiciones iniciales para $\mu = 4$. (a) Se indican dos órbitas que difieren en el valor de sus condiciones iniciales. Aquí, la primera parte de $x_0 = 0.1000$ (línea continua) y la segunda de $x^* = 0.1001$ (trayectoria discontinua). (b) Diferencia (en valor absoluto) entre ambas trayectorias. Esta diferencia, que fluctúa enormemente con el tiempo, define la existencia de sensibilidad a las condiciones iniciales.

 tales que

$$|f^n_\mu(x) - f^n_\mu(y)| > \delta$$

En forma intuitiva: la aplicación es sensible a las condiciones iniciales si existen puntos arbitrariamente próximos a $x$ que se separan al menos en $\delta$ bajo $n$ iteraciones de la función $f_\mu$.

¿Cómo se traduce este comportamiento en el diagrama de bifurcación de la aplicación logística? En la sección anterior nos habíamos detenido en los límites de la secuencia de bifurcaciones que, como vimos, posee un punto de acumulación para cierto valor $\mu_\infty$. Si continuamos nuestra representación de las órbitas del sistema para valores superiores, vemos por primera vez la estructura del diagrama en la región caótica (figura 5.9 (a)), que posee una extraordinaria complejidad. Las regiones sombreadas corresponden a órbitas aperiódicas (caóticas) que aparecen salpicadas por ventanas de periodicidad de mayor o menor tamaño. Si llevamos a cabo una ampliación de un detalle de una de estas ventanas (figura 5.9 (b)) nos encontramos con una sorpresa: el diagrama de bifurcación anteriormente obtenido aparece nuevamente en esta escala de detalle y, de hecho, puede ser obtenido una y otra vez, con toda su complejidad, a partir de detalles de diagramas sucesivos. Pese a su simplicidad, la ecuación logística esconde una complejidad asombrosa.

Podemos ilustrar esta definición con un ejemplo formal simple. Consideremos el sistema dinámico (Martín et al., 1994)

$$x_{n+1} = f(x_n) = x_n^2$$

demostremos que, dentro del intervalo $U = [1, \infty)$ este sistema simple (pero no lineal) exhibe SCI. Para el intervalo $U$ empleado, la órbita que parte de cualquier punto $x_0 > 1$ diverge a infinito. Tenemos:

$$\lim_{n \to \infty} f^n(x) = \lim_{n \to \infty} x^{2^n} = \infty ; \ x_0 > 1$$
Figura 5.9: Diagrama de Feigenbaum. (a) Diagrama para distintos valores de \( \mu \) desde el dominio periódico a los límites del dominio caótico. En (b) indicamos una ampliación de un pequeño detalle del primer diagrama, obtenido de una parte de la ventana periódica.

(y toma el valor constante 1 si \( x_0 = 1 \)).

Sea \( \delta > 0, \; 1 > \epsilon > 0 \) y \( x \in [0, \infty) \). Cualquier punto arbitrario \( y \in (x, x + \epsilon) \) verifica la desigualdad

\[
|y - x| < \epsilon
\]
y, por el teorema del valor medio, tenemos que

\[
|f^n(y) - f^n(x)| = |\partial_x f^n(x)\xi|y - x|
\]
donde \( \xi \in (x, x + \epsilon) \) y es un punto arbitrario en este intervalo, que existe tal que la condición anterior se verifica. Aplicando la regla de la cadena,

\[
\left. \frac{\partial f^n(x)}{\partial x} \right|_{\xi} = f'(\xi)f'(f(\xi))f'(f^2(\xi))...f'(f^{n-1}(\xi))
\]

Ahora, puesto que \( f'(x) = 2x \geq 2 \) (ya que \( x \geq 1 \)) y dado que

\[
\{f^k(\xi) ; \; k = 0, ..., \infty\} \subset [1, \infty)
\]

se verifica que \( (f^n)'(\xi) \geq 2^n \), luego

\[
|f^n(y) - f^n(x)| \geq 2^n |y - x|
\]
y se verifica \( 2^n |y - x| > 1 \) si

\[
n > \frac{\ln |y - x|^{-1}}{\ln(2)}
\]

Por tanto, para el valor de \( n \) que verifica la condición anterior se obtiene una separación entre las dos trayectorias superior a \( \epsilon \),
\[ |f^n(y) - f^n(x)| \geq 1 > \epsilon \]

esto es, el sistema anterior presenta sensibilidad a las condiciones iniciales, como queríamos demostrar.

Debe quedar claro del último ejemplo que la sensibilidad a las condiciones iniciales es característica de los sistemas que exhiben caos determinista (concepto que iremos definiendo). Aunque el ejemplo anterior es ilustrativo, no lo es de manera completa: el sistema dinámico anterior no está acotado como lo estarán los sistemas dinámicos de interés que analizaremos (como es el caso de la ecuación logística). La existencia de un atractor acotado a cierto dominio del espacio de fases será un requisito claro a exigir.

### 5.4 Exponentes de Lyapunov

Hemos definido con anterioridad el concepto de atractor extraño ligando la geometría de estos objetos a la sensibilidad a las condiciones iniciales. El carácter disipativo de la dinámica garantiza la convergencia de las trayectorias dentro de la cuenca de atracción hacia cierta región acotada del espacio de las fases que, como ya vimos, define la dinámica del sistema y nos da una imagen del orden (y del determinismo) subyacente. Por otra parte, la sensibilidad a las condiciones iniciales hace que dos puntos inicialmente próximos situados sobre el atractor se alejen de manera exponencial con el paso del tiempo. El resultado de ambas tendencias da lugar a un proceso de estratificación-plegado que origina las propiedades fractales ya comentadas. Veremos ahora cómo cuantificar, desde el punto de vista de la dinámica del sistema, la divergencia (y por tanto la inestabilidad local) de las trayectorias. Emplearemos para ello los exponentes de Lyapunov, que nos permitirán conectar con la geometría del atractor a través de su proceso de formación.

Para introducir el concepto de exponente de Lyapunov, \( \lambda_L \), emplearemos como ejemplo una aplicación unidimensional que extenderemos a continuación a un ejemplo bidimensional. Consideremos la aplicación \( x_{n+1} = F_\mu(x_n) \) con la expresión \( F_\mu = \mu x(1-x) \) y definida sobre el intervalo unidad \( U = [0, 1] \). Dadas dos condiciones iniciales muy próximas, el comportamiento de ambas diverge, en el régimen caótico, de una forma rápida que podemos cuantificar.

El ejemplo numérico anterior, que estudiábamos para la ecuación logística con \( \mu = 4 \), tiene sus contraparidas equivalentes en otros sistemas dinámicos caóticos disipativos, como el modelo de Lorenz. Consideremos ahora el problema en términos generales, manteniendo por ahora la aproximación 1-dimensional. Tomemos dos condiciones iniciales \( x_0^1, x_0^2 \in U \) separadas una distancia \( \epsilon \ll 1 \), con lo que podemos escribir de hecho \( x_1^1 = x_0^1 \) y \( x_1^2 = x_0^2 + \epsilon \). Tras \( n \) iteraciones de \( F_\mu \), estas condiciones iniciales habrán evolucionado hasta los nuevos puntos definidos por

\[
x_1 = F_\mu^n(x_0) \\
x_2 = F_\mu^n(x_0 + \epsilon)
\]

La separación inicial \( \epsilon \) habrá incrementado su valor en un cierto factor, digamos

\[
\epsilon_n = \epsilon G(x_0, n)
\]

que indicaremos en la forma:

\[ G(x_0, n) = \epsilon^{\lambda(x_0)} \]

siendo \( \lambda(x_0) \) el llamado exponente de Lyapunov. Se tiene por tanto que:

\[
\epsilon \epsilon^{\lambda(x_0)} = |F_\mu^n(x_0 + \epsilon) - F_\mu^n(x_0)|
\]
que nos da, tomando el límite para \( n \to \infty \) y \( \epsilon \to 0 \):

\[
\lambda(x_0) = \lim_{n \to \infty} \lim_{\epsilon \to 0} \frac{1}{n} \ln \left| \frac{F_{\mu}^n(x_0 + \epsilon) - F_{\mu}^n(x_0)}{\epsilon} \right|
\]

que obviamente se reduce a la expresión:

\[
\lambda(x_0) = \lim_{n \to \infty} \frac{1}{n} \ln \left| \frac{\partial_x F_{\mu}^n(x_0)}{\partial x_0} F_{\mu}^n(x_0) \right|
\]

En este contexto, \( e^\lambda \) es el factor (promediado) de divergencia entre condiciones iniciales próximas tras \( n = 1 \) iteraciones. Una expresión más simple se obtiene manipulando la expresión anterior:

\[
\lambda(x_0) = \lim_{n \to \infty} \frac{1}{n} \ln \left| \frac{\partial_x F_{\mu}^n(x_0)}{\partial x_0} F_{\mu}^n(x_0) \right| = \lim_{n \to \infty} \frac{1}{n} \ln \left| \frac{\partial_x F_{\mu}^n(F_{\mu}(\ldots(F_{\mu}(x_0))))}{\partial x_0} F_{\mu}(x_0) \right|
\]

y dado que

\[
(\partial_x F_{\mu}^2(x_0))_{x_0} = F_{\mu}'(x_0) F_{\mu}'(x_0)
\]

con \( x_1 = F_{\mu}(x_0) \), reescrivimos \( \lambda(x_0) \) en la forma:

\[
\lambda(x_0) = \lim_{n \to \infty} \frac{1}{n} \ln \left| \prod_{i=1}^{n-1} \partial_x F_{\mu}^i(x_i) \right|
\]

o lo que es lo mismo:

\[
\lambda(x_0) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \ln |\partial_x F_{\mu}^i(x_i)|
\]  \hspace{1cm} (5.4.1)

Si bien esta última expresión no nos permitirá (en general) obtener resultados analíticos, es de fácil manejo en lo que al cálculo numérico se refiere, y permite observar propiedades cualitativa y cuantitativamente relevantes. Como ejemplo (figura 5.10) tomemos la aplicación logística y calculemos el valor de \( \lambda \) para distintos valores de \( \mu \), promediando \( N = 5000 \) iteraciones tras eliminar los primeras \( \tau = 1000 \). El procedimiento de cálculo seguido, explícitamente, es:

- Damos un primer valor de \( \mu \) y una condición inicial (digamos \( x = 0.1 \)) que emplearemos para obtener una serie de valores.

- Iteramos la aplicación \( f_{\mu}(x_n) = \mu x_n(1 - x_n) \) durante \( N + \tau \) pasos de tiempo, descartando los primeros \( \tau \) (para evitar incluir el comportamiento transitorio). Tenemos entonces una serie

\[ S_{\mu} = \{x_{\tau+1}, \ldots, x_{\tau+N} \} \]

- Para cada uno de los valores obtenidos \( x_j \in S_{\mu} \), calculamos el valor de

\[ \ln |\partial_x f_{\mu}(x_j)| = \ln |\mu x_j(1 - x_j)| \]

valor que iremos sumando para todos los \( x_j \in S_{\mu} \) hasta obtener el promedio temporal, lo que nos dará una estimación del exponente de Lyapunov.

- Repetimos el procedimiento anterior para un nuevo \( \mu \).
Figura 5.10: Aplicación logística: Exponentes de Lyapunov. (a) Exponentes obtenidos para \( \mu \in (3.4, 4.0) \) en los que apreciamos una transición hacia valores positivos para \( \mu > \mu_\infty \) que aparecen salpicados de un gran número de caídas hacia valores negativos o nulos que indican la presencia de órbitas periódicas. (b) Ampliación de la gráfica anterior, para \( \mu \in (3.85, 3.95) \).

Podemos observar la existencia de una zona de exponentes positivos que corresponde a los valores de \( \mu \) del dominio caótico. Previamente, podemos observar las últimas etapas del escenario de Feigenbaum. Los puntos para los que \( \lambda = 0 \) corresponden a los puntos de bifurcación (de estabilidad marginal). Una propiedad especialmente interesante es la existencia de “ventanas” periódicas entremezcladas en la zona caótica. De hecho, si ampliamos un detalle cualquiera de este dominio (véase fig. 5.10 (b)), aparecen otras ventanas. Este hecho conlleva la inestabilidad estructural del caos en este sistema: cualquier entorno de un \( \mu \) con un exponente asociado positivo contiene puntos periódicos. Este fenómeno, como veremos, no tiene contrapartida en los modelos que incluyen el espacio físico en forma explícita.

### 5.5 La aplicación triangular

Existe un sistema dinámico discreto unidimensional especialmente interesante debido a su simplicidad y a la posibilidad de obtener, de forma análoga, su exponente de Lyapunov asociado. El sistema en cuestión está definido por la aplicación:

\[
x_{n+1} = f_\mu(x_n) = \mu \left( 1 - 2\left| \frac{1}{2} - x_n \right| \right)
\]

donde \( \mu \in (0, 1) \). Podríamos reescribir esta aplicación en dos partes, esto es:

\[
x_{n+1} = 2\mu x_n \quad \text{para} \quad x_n < 1/2 \quad \text{y} \quad x_{n+1} = 2\mu (1 - x_n) \quad \text{para} \quad x_n > 1/2
\]

Esta función se representa en la figura 5.11, en la que hemos indicado además la trayectoria del sistema para cuatro valores distintos de \( \mu \). Excepto para el primer caso, los otros tres muestran una dinámica compleja. Demostraremos que, efectivamente, se trata de caos determinista (exponente de Lyapunov positivo).
Figura 5.11: Aplicación triangular: diagramas de recurrencia para \( \mu = 0.7 \) y para \( \mu = 1 \).

Para esta función, el cálculo de la derivada nos da, para cada subintervalo \((0, 1/2), (1/2, 1)\), los valores \(\mu\) y \(-\mu\), respectivamente. Si empleamos la expresión del exponente de Lyapunov 5.4.1, obtenemos

\[
\lambda_L = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \ln |\partial_x f_\mu(x_i)|
\]

\[
= \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \ln |\mu| = \ln(\mu)
\]

que será positivo si \(\mu > 1/2\) y negativo en caso contrario. Así, más allá de \(\mu_c = 1/2\) este sistema siempre mostrará caos. Aunque la aplicación triangular es un sistema dinámico muy simple, sus propiedades (así como las de otros sistemas similares) la hacen, como veremos, de enorme interés en nuestro estudio.

5.6 Sistemas discretos: \(d > 1\)

Los resultados previos pueden obtenerse en sistemas discretos de mayor dimensión. Para estos sistemas, definidos ahora por un conjunto de ecuaciones discretas del tipo

\[
x_{n+1} = f_\mu(x_n)
\]

con \(x_n = (x_n^1, ..., x_n^d)\) y \(f_\mu = (f_\mu^1, ..., f_\mu^d)\) tendremos también combinaciones de parámetros que darán lugar a dinámicas enormemente complejas. Un ejemplo de éstas ya se vio en el capítulo 2, en el que se mostraba la dinámica del modelo de Lotka-Volterra

\[
x_{n+1} = \mu x_n (1 - x_n - y_n)
\]

\[
y_{n+1} = \mu x_n y_n
\]

el cual exhibe atractores de gran complejidad para ciertos intervalos de valores de \(\mu\).
Figura 5.12: Deformación de un entorno de condiciones iniciales bajo la acción del flujo de sistema.

La estabilidad de los puntos de equilibrio se analiza de forma similar a lo que ya se vio con anterioridad. El sistema lineal asociado en las proximidades de un punto fijo dado será:

$$x_{n+1}^i = \sum_{j=1}^{d} \left( \frac{\partial F^i}{\partial x_j} \right)(x_n)x_n^j$$

y, por analogía con lo anterior, y empleando notación vectorial, tenemos

$$x_n = \left( \prod_{k=0}^{n-1} J_k(\partial_j F^i) \right) x_0$$

siendo $J_k(\partial_j F^i)$ la matriz de Jacobi calculada en el instante $k$. Es importante apuntar que el producto matricial que aparece entre paréntesis debe mantener el orden indicado

$$J_0 J_1 \ldots J_{n-1}$$

ya que en general $J_l J_m \neq J_m J_l$.

Podemos ahora explorar las soluciones caóticas de forma cuantitativa midiendo ciertas cantidades (de carácter estadístico) que se corresponden con la definición anterior de exponente de Lyapunov. Geométricamente, y restringiéndonos al caso $d = 2$ (sin pérdida de generalidad) podemos comprender el significado de los exponentes de Lyapunov $\{\lambda_k; k = 1, 2\}$ a partir del esquema de la figura 5.12.

Una bola de condiciones iniciales de radio $r$ se deforma bajo la acción de $F_\mu$ tal y como se indica en el esquema. En una dirección se da estrangulamiento del semieje y en la ortogonal contracción. En esta forma, los nuevos semiejes $a_k$ y $b_k$ de la elipse tienen un valor en la $k-$ésima iteración dado por $a_k(x, y) \approx r e^{\lambda_1 k}$ y $b_k(x, y) \approx r e^{\lambda_2 k}$, siendo $\lambda_1, \lambda_2$ las tasas de crecimiento. Los exponentes de Lyapunov se definen entonces como el logaritmo del promedio de estas tasas,

$$\lambda_1(x, y) = \lim_{k \to \infty} \frac{1}{k} \ln a_k(x, y)$$

$$\lambda_2(x, y) = \lim_{k \to \infty} \frac{1}{k} \ln b_k(x, y)$$

De forma general, para una aplicación $d-$dimensional, el procedimiento a seguir para calcular el espectro de exponentes de Lyapunov será diagonalizar la matriz Jacobiana asociada a la $T-$ésima
iteración \((L_{\mu})(T)\) (el producto de \(T\) veces la matriz calculada para cada iteración) y obtener los valores propios asociados, que denominaremos (para evitar confusiones):

\[
\{\Lambda_i \ ; \ i = 1, 2, \ldots, d\}
\]

y los ordenaremos de forma que

\[
|\Lambda_j| \geq |\Lambda_{j+1}| \ ; \ \forall j = 1, \ldots, d
\]

Entonces obtendremos los exponentes de Lyapunov llevando a cabo el límite

\[
\lambda_j = \lim_{n \to \infty} \frac{1}{n} \ln |\Lambda_j|
\]

siempre que dicho límite exista. Para el conjunto de puntos que convergen por iteración al atractor (los que pertenecen a la cuenca de atracción del mismo) este límite será independiente de la condición inicial escogida.

Veremos más adelante cómo calcular de forma general los exponentes de Lyapunov. Notemos de momento que, a partir de la propiedad

\[
\prod_{j=1}^{p} \Lambda_j = \det(L_{\mu}(T))
\]

en muchos casos de interés (como el que veremos a continuación) se tiene que \(\det(L_{\mu}(T)) = C\), una constante independiente de las variables dinámicas. Bajo estas condiciones, tendremos que \(\det(L(T)) = C^p\). En este caso particular,

\[
\sum_{j=1}^{p} \lambda_j = \ln |C|
\]

Para el caso especial en que el volumen no cambia con el tiempo, tendremos que \(|C| = 1\) y por lo tanto para sistemas dinámicos discretos que preservan el volumen (el área en dos dimensiones)

\[
\sum_{j=1}^{p} \lambda_j = 0
\]

Una consecuencia adicional del resultado anterior es que \(\lambda_p \leq \ln |C|/p\).

### 5.7 El modelo de Hénon

Uno de los sistemas discretos bidimensionales mejor conocidos es el sistema de Hénon, que puede escribirse en la forma

\[
y_{n+1} + B y_{n-1} = 2 y_n + 2 y_n^2
\]

y que, para \(|B| < 1\), es disipativo. Para \(B = 0\) la aplicación anterior se reduce a la logística, mientras que si \(B = 1\), tenemos un sistema dinámico conservativo. De manera equivalente, podemos escribir el sistema de Hénon en la forma:

\[
x_{n+1} = y_n + 1 - ax_n^2
\]
\[
y_{n+1} = bx_n
\]
Figura 5.13: (a) Atractor de Hénon \( a = 1.4, \ b = 0.3 \), y (b) exponente de Lyapunov máximo para la aplicación de Hénon.

con Jacobiano asociado (ahora \( \mu \equiv (a, b) \)),

\[
L_\mu = \begin{pmatrix}
-2ax_n \\ b \\
1 \\ 0
\end{pmatrix}
\]

lo que nos da un determinante \( \det(L_\mu) = -b \). Dado que es constante, podemos emplear el resultado obtenido anteriormente acerca de la suma de los exponentes de Lyapunov:

\[
\lambda_1 + \lambda_2 = \ln |b|
\]

Y por lo tanto en este caso no es preciso calcular más que el exponente de Lyapunov máximo. En la figura 5.13 se representa el atractor de Hénon, así como el mayor exponente \( \lambda_1 \) obtenido para el modelo de Hénon, para \( b = 0.3 \).

Podemos analizar la estabilidad de los puntos fijos del sistema de Hénon. Estos son \( X^* = (x^*_+, bx^*_+) \) y \( X^- = (x^-*, bx^-*) \), siendo

\[
x_\pm = \frac{1}{2a} \left( -(1-b) \pm \sqrt{(1-b)^2 + 4a^2} \right)
\]

valores reales para

\[
a \geq a_c \equiv -\frac{1}{4} (1-b)^2
\]

Para \( a > a_c \), los valores propios de la matriz de Jacobi nos dan:

\[
\Lambda_\pm = -ax_\pm \pm \sqrt{a^2(x_\pm^*)^2 + b}
\]

con lo que tendremos estabilidad para

\[
|x_\pm^*| < \frac{1-b}{2a}
\]

Este conjunto de valores de \( x \) constituye la zona contractiva. Tenemos:
Figura 5.14: Atractor (extraño) de Hénon, para $a = 1.4$ y $b = 0.3$. Podemos apreciar el efecto del estiramiento y plegado de las trayectorias (véase el texto). Al llevar a cabo sucesivas ampliaciones de una zona del atractor, descubrimos que se trata de un objeto fractal.

- (i) Para $x^+_+$, la desigualdad anterior permite calcular la frontera de la zona contractiva, dada por $s_1 = 3(1 - b)^2/4$. Para $a \in (0, a_1)$ el punto fijo $x^+_+$ es asintóticamente estable.

- (ii) Para $x^-_+$, puede comprobarse que es un punto fijo inestable $\forall a$.

Cuando $a > a_1$, el punto fijo $x^+_+$ deja de ser un atractor estable. Al igual que ocurriría con la función logística discreta, nuestro sistema abandona la estabilidad para dar lugar a trayectorias periódicas. Hénon (1976) (véase también el exhaustivo trabajo de Simó (1979)) estudió el comportamiento de este sistema (que es de hecho una aproximación a ciertos problemas físicos más cercanos al representado por el atractor de Lorenz) observando la aparición de comportamientos de gran complejidad, como es de esperar a partir de lo anteriormente visto con el cálculo del exponente máximo de Lyapunov. Para $a > a_1 = 0.7096$ aparecen bifurcaciones con duplicación de periodo que alcanzan el régimen caótico para $a_\infty \approx 1.06...$ dando lugar a atractores extraños. En la figura 5.14 (a-c) vemos una representación gráfica de este atractor junto con la ampliación sucesiva de un detalle del mismo. Podemos apreciar con claridad la repetición de la misma estructura a cualquier escala.

Podemos aprovechar este ejemplo para comprobar la riqueza de estructuras asociadas a la aparición de un atractor extraño, representando gráficamente la cuenca de atracción de esta solución caótica (fig. 5.15). El estudio de estas cuencas revela una enorme complejidad en su estructura, que a menudo es fractal (Grebeni et al., 1987). En el capítulo sobre fractales se ha representado numerosos ejemplos de conjuntos de Julia. Desde el punto de vista dinámico, las ecuaciones iterativas que dan lugar a estos conjuntos se pueden interpretar como aplicaciones dinámicas, del mismo tipo que las que se están discutiendo, aunque con valores en el plano complejo. Dábamos en aquel capítulo la forma en que se calculaban los puntos $z \in C$ pertenecientes a cada uno de los conjuntos de Julia: eran, de hecho, los que separaban la frontera entre los puntos del plano complejo que tenían como atractor el origen y los que escapaban al infinito bajo iteración de la función $z \rightarrow z^2 + c$ ($c \in C$). Por tanto, cada uno de los conjuntos de Julia se puede interpretar como la cuenca de atracción de la función anterior, para cada valor de $c$. Los puntos sobre los conjuntos de Julia presentan sensibilidad a las condiciones iniciales bajo la iteración del polinomio $z^2 + c$. 
5.8 La transformación del panadero

Estudiaremos ahora un caso de aplicación bidimensional especialmente interesante por su simplicidad y comportamiento. Se trata de la conocida transformación del panadero, definida por el sistema dinámico discreto

\[
F_{\mu}(x, y) = \begin{cases} 
(2x, 2y) & \text{si } x \in [0, 1/2] \\
(2x - 1, \mu y + 1/2) & \text{si } x \in (1/2, 1] 
\end{cases}
\]

que se aplica sobre los puntos del cuadrado unidad \( E = [0, 1] \times [0, 1] \), y \( 0 < \mu < 0.5 \). El efecto de esta aplicación \( F_{\mu} : E \rightarrow E \) sobre \( E \) puede verse en la figura 5.16.

Esta aplicación "estira" el conjunto \( E \), dando lugar a un rectángulo de área \( 2 \times \mu \), que es a continuación "cortado" por su mitad dando dos partes de área \( 1 \times \mu \) que se unen (se pliegan) con un vacío de lado \( \frac{1}{2} - \mu \) entre ambas. Al aplicar sucesivamente la transformación (similar a la que realiza un panadero al amasar), el conjunto inicial se transforma en \( E_k = F_{\mu}^k(E) \), formado por una colección de \( 2^k \) bandas horizontales de grosor \( \mu^{-k} \) y separadas por vacíos de longitud vertical \( l_k \geq (0.5 - \mu)\mu^{1-k} \). Puesto que \( F_{\mu}(E_k) = E_{k+1} \), el conjunto compacto \( G \) definido por

\[
G = \bigcap_{k=0}^{\infty} E_k
\]

satisfiere claramente el requisito de invarianza \( F_{\mu}(G) = G \). Esta aplicación genera una transformación de estirado-y-pliegado sobre el conjunto \( E \). El proceso de estiramiento introduce la inestabilidad que da lugar a la sensibilidad a las condiciones iniciales, y el pliegue introduce la adecuada disipación (confinando los puntos de \( E \) a este mismo dominio). Para este sistema el cálculo de los exponentes puede realizarse analíticamente. La matriz de Jacobi es ahora:

\[
L_{\mu} = \begin{pmatrix} 2 & 0 \\ 0 & \mu \end{pmatrix}
\]

y por lo tanto el producto matricial definido por
Figura 5.16: Transformación del panadero: la idea intuitiva de cómo se origina la sensibilidad a las condiciones iniciales aparece bien ejemplificada por el procedimiento de amasar pan. Al aplastar la masa en una dirección, separamos puntos cercanos (sensibilidad a las CI) mientras que al plegar la masa confinamos el sistema a una región limitada.

\[ L_{\mu}^k = \prod_{j=1}^{k} P_{\mu}(x, y) \]

es la matriz diagonal:

\[ L_{\mu}^k = \begin{pmatrix} 2^k & 0 \\ 0 & \mu^k \end{pmatrix} \]

los semiejes serán por lo tanto

\[ a_k(x, y) = 2^k ; \quad b_k(x, y) = \mu^k \]

y los exponentes de Lyapunov correspondientes serán:

\[ \lambda_1 = \ln 2 \quad (> 0) \quad \lambda_2 = \ln \mu \quad (< 0) \]

Este ejemplo, de gran simplicidad, ilustra con enorme claridad el fenómeno que subyace a la aparición del caos determinista en sistemas dinámicos disipativos. El objeto obtenido por iteración de \( F_{\mu} \) muestra sensibilidad a las condiciones iniciales y propiedades fractales: ambos fenómenos son caras de la misma moneda.

Para el conjunto invariante generado por este sistema, podemos calcular sin dificultad su dimensión fractal. Si tomamos tres iteraciones sucesivas del sistema dinámico (fig. 5.17), podemos calcular el número de objetos de tamaño vertical dado (en la horizontal tenemos siempre la misma longitud) que se requieren para recubrir el conjunto generado. En la figura se indican claramente las longitudes e consecutivas, así como el número de objetos necesarios para obtener el recubrimiento. Tenemos entonces que el conjunto invariante final, resultante de la transformación, \( A_\infty \),
Figura 5.17: Iteraciones sucesivas de la transformación del panadero. Se indican las longitudes del los lados de las sucesivas imágenes obtenidas a través de la iteración. Se indica también el número de objetos de lado dado necesarios para completar el recubrimiento. Este cálculo permite determinar la dimensión fractal asociada al objeto obtenido (el atractor).

tendrá dimensión (sumamos una unidad para introducir la dimensión horizontal):

\[
D(\Lambda_\infty) = 1 - \lim_{n \to \infty} \frac{\ln[N_n(\epsilon)]}{\ln(\epsilon)} = 1 - \lim_{n \to \infty} \frac{\ln[2^n]}{\ln(\mu^n)} = 1 - \frac{\ln(2)}{\ln(\mu)} < 2
\]

Para terminar, estudiaremos brevemente las propiedades estadísticas de la aplicación del panadero para \( \mu = 2 \), que define un sistema dinámico conservativo. Para analizar este problema recurriremos a algunas herramientas de tipo estadístico como la ecuación maestra (capítulo 1). Llamemos \( p_n(x_n, y_n) \) a la densidad de probabilidad asociada al sistema dinámico. Tendremos que:

\[
p_n(x_n, y_n) \, dx_n \, dy_n = p_0(x_0, y_0) \, dx_0 \, dy_0 = p_{n-1}(x_{n-1}, y_{n-1}) \, dx_{n-1} \, dy_{n-1}
\]

Ahora bien, puesto que el sistema es conservativo, el Jacobiano asociado a la transformación es unitario, y por lo tanto tenemos conservación de la densidad de probabilidad,

\[
p_n(x_n, y_n) = p_{n-1}(x_{n-1}, y_{n-1}) = \cdots = p_0(x_0, y_0)
\]

La inversa de la transformación será

\[
G(x_n, y_n) = \left( \frac{x_n + 1}{2}, 2y_n \right) \quad \text{si} \quad x_n \in \left[ 0, \frac{1}{2} \right]
\]

\[
G(x_n, y_n) = \left( \frac{x_n + 1}{2}, 2y_{n+1} - 1 \right) \quad \text{si} \quad x_n \in \left( \frac{1}{2}, 1 \right]
\]

o, lo que es lo mismo,

\[
G(x_n, y_n) = \left( \frac{x_n + 1}{2}, 2y_{n+1} \right) \quad \text{si} \quad y_n \in \left[ 0, \frac{1}{2} \right]
\]

\[
G(x_n, y_n) = \left( \frac{x_n + 1}{2}, 2y_{n+1} - 1 \right) \quad \text{si} \quad y_n \in (1/2, 1]
\]
Si reemplazamos \(-1\) por \((x_{n-1}, y_{n-1})\) por sus expresiones en términos de \((x_n, y_n)\) a izquierda y derecha de la ecuación \(p_n(x_n, y_n) = p_{n-1}(x_{n-1}, y_{n-1})\) y haciendo \((x_n, y_n) \equiv (x, y)\), obtenemos la ecuación maestra:

\[
p_n(x, y) = p_{n-1}\left(\frac{x}{2}, 2y\right) \quad \text{si} \quad y \in \left[0, \frac{1}{2}\right]
\]

\[
p_n(x, y) = p_{n-1}\left(\frac{x + 1}{2}, 2y - 1\right) \quad \text{si} \quad y \in \left(\frac{1}{2}, 1\right]
\]

Este sistema dinámico presenta una aproximación al equilibrio estadístico (en el sentido ya discutido). De esta ecuación podemos obtener los promedios estadísticos de nuestro sistema. En particular, podemos expresar el promedio de cualquier variable \(f(x, y)\) como

\[
< f(x, y) > = \int_{0}^{1} \int_{0}^{1} p_n(x, y)f(x, y) \, dx \, dy = \int_{0}^{1} \int_{0}^{1} f(x, y) \, dx \, dy
\]

Podemos también obtener una ecuación maestra a partir de la ecuación de Perron-Frobenius.

\[
p_n(x, y) = \int_{0}^{1} \int_{0}^{1} p_{n-1}(x', y') \delta[x - f(x')] \, \delta[y - g(y')] \, dx' \, dy'
\]

donde \(f(x)\) y \(g(y)\) son las funciones

\[
f(x) = \begin{cases} 
2x & x \in [0, 1/2) \\
2x - 1 & x \in [1/2, 1]
\end{cases}
\]

\[
g(y) = \begin{cases} 
y/2 & y \in [0, 1/2) \\
1/2(y + 1) & y \in [1/2, 1]
\end{cases}
\]

Dado que

\[
\delta[x - f(x')] = \frac{1}{|f'(x')|} \delta[x' - f^{-1}(x)]
\]

y empleando el hecho de que el valor de \(x\) dentro del intervalo unidad determina el nuevo valor de \(y_{n+1}\), tendremos

\[
p_n(x, y) = \int_{0}^{1/2} dx' \int_{0}^{1/2} p_{n-1}(x', y') \delta\left(\frac{x'}{2}\right) \delta\left[y' - 2y\right] \, dy'
\]

\[
= \int_{1/2}^{1} dx' \int_{1/2}^{1} p_{n-1}(x', y') \delta\left(\frac{x' + 1}{2}\right) \delta\left[y' - (2y - 1)\right] \, dy'
\]

luego, si \(0 \leq y < 1/2\), obtenemos

\[
p_n(x, y) = p_{n-1}\left(\frac{x}{2}, 2y\right)
\]

y, para \(y \in [1/2, 1]\) obtenemos

\[
p_n(x, y) = p_{n-1}\left(\frac{x + 1}{2}, 2y - 1\right)
\]

como antes. Puede demostrarse (Ford. 1983) que, de hecho, las propiedades estadísticas de este sistema dinámico son las mismas que las del proceso de lanzamiento de una moneda.
5.9 Mixing y ergodicidad

Anteriormente, hemos introducido la llamada transformación del panadero, un sistema dinámico bidimensional bien conocido en la teoría de sistemas caóticos, y que nos permitirá entender algunos fenómenos básicos, así como el papel que el caos determinista puede jugar en ellos. Una propiedad especialmente relevante es la ergodicidad y el llamado mixing (o propiedad de mezcla), ambos relacionados entre sí. La propiedad de mezcla juega un papel muy importante en la definición de caos determinista, y es también conocida como transitiiedad topológica. En esta sección daremos algunas definiciones básicas.

Imaginemos dos fluidos (digamos tinta y agua) que se mezclan entre sí (figura 5.18). Esta experiencia simple permite acercarnos intuitivamente al problema de la evolución hacia el equilibrio en sistemas dinámicos. Inicialmente, el fluido oscuro ocupa una porción reducida del recipiente (indicada por $B'$ en la figura). Llamemos $\mathcal{R}$ a esta región, e imaginemos que ambos fluidos son incompresibles, de forma que el volumen de ambos se conserva y supongamos que son no-viscosos, de forma que podamos ignorar los efectos de la difusión molecular de un fluido a través de su frontera con el otro.

A medida que el tiempo transcurre y el fluido es agitado, la región inicial se modifica, extendiéndose y deformándose. Indicaremos esta evolución en términos de un sistema dinámico:

$$\mathcal{R}_t = f^t(\mathcal{R})$$

donde $f$ es el operador que define la evolución temporal de $\mathcal{R}$. Para $t \to \infty$, los filamentos se van extendiendo uniformemente hasta que el fluido se vuelve rosado. Podríamos decir que, como observadores, llevamos a cabo un “promedio de grano grueso” que nos permite llegar a la obtención de una evolución irreversible hacia el equilibrio estadístico. Idealmente, sin embargo,
si observásemos el fluido a una escala microscópica\footnote{Más exactamente, a una escala microscópica y antes de que la difusión molecular haya tenido tiempo suficiente de actuar.} veríamos que la tinta estaría formando hilos finos de formas muy complejas.

La condición de mixing puede formularse matemáticamente como sigue: sea $B$ una región del fluido de volumen $\mu(B)$ y sea $A$ otra parte (por ejemplo, la cuarta parte del volumen de agua, situada en el fondo del recipiente, figura 5.18), de volumen $\mu(A)$. Como es habitual, definimos $\mu(C)$ por:

$$\mu(C) \equiv \int_C \mu(x) dx$$

(para $C = A, B$). Si $\Gamma$ indica el espacio de fases completo, debe darse la normalización

$$\int_\Gamma \mu(x) dx = 1$$

Bajo la acción de la mezcla, $B$ evoluciona en el tiempo. Podemos definir una dinámica mediante la aplicación $f : \mathbb{R} \rightarrow \mathbb{R}$, como antes (y sobre la que se define la medida $\mu$). Se dice que la dinámica es mixing (o que posee la propiedad de mezcla) si:

$$\lim_{k \to \infty} \mu\left(f^{-k}(B) \cup A\right) = \mu(A)\mu(B) \tag{5.9.1}$$

Si además $f$ es una aplicación uno a uno \footnote{Esto es, tal que si $x_1 \neq x_2$ se tenga $f(x_1) \neq f(x_2)$,}, obtendremos:

$$\lim_{k \to \infty} \mu\left(f^k(B) \cup A\right) = \mu(A)\mu(B)$$

En otros términos, tendremos que si $B_t = f^t(B)$, entonces

$$\lim_{t \to \infty} \frac{\mu(B_t \cup A)}{\mu(B)} = \frac{\mu(A)}{\mu(B)}$$

lo que no es otra cosa, físicamente, que el cociente entre los volúmenes de $A$ y $B$. La fracción de tinta que se esparce por la región $A$ deberá, con gran probabilidad, igualar la fracción del espacio total ocupado por $A$.

En particular, se puede demostrar que un sistema mixing posee la siguiente propiedad: si $A$ y $B$ corresponden a la misma región, entonces se tiene:

$$\lim_{t \to \infty} \int_\Gamma \mu(x) f^t(x) x dx \equiv < x_t x_0 >$$

$$= \left[\int_\Gamma \mu(x) x dx \right]^2 = < x_0 >^2$$

esto es, la propiedad de mixing implica que la función de autocorrelación, definida por

$$C(t) = < (x_t - < x_0 >)(x_0 - < x_0 >) > = < x_t x_0 > - < x_0 >^2$$

decae a cero (el sistema se relaja al estado de equilibrio termodinámico).

La propiedad de mixing es una caracterización fuerte de caos. En particular, se tiene que:

$$\text{mixing} \implies \text{caos}$$
(en el sentido de Li y Yorke, como se verá más adelante) pero

\[ \text{Caos} \Leftrightarrow \text{mixing} \]

Del mismo modo, se tiene que

\[ \text{mixing} \implies \text{ergodicidad} \]

\[ \text{ergodicidad} \implies \text{mixing} \]

Podemos dar un ejemplo sencillo de sistema dinámico que no exhibe la propiedad de mezcla. Empleando otra vez el sistema

\[ x_{n+1} = f(x_n) = x_n^2 \]

en \( x \in [0,1] \), distinto del anteriormente considerado. Tomemos los subintervalos

\[ I = \left[ \frac{1}{2^5}, \frac{1}{2^4} \right] \quad J = \left[ \frac{1}{2^7}, \frac{1}{2^6} \right] \]

Podemos ver fácilmente que, \( \forall n > 1 \), se tiene

\[ f^n(I) \cup J = \left[ \frac{1}{2^{10n}}, \frac{1}{2^{9n}} \right] \cup \left[ \frac{1}{2^7}, \frac{1}{2^6} \right] = \emptyset \]

\[ f^n(I) \cup I = \left[ \frac{1}{2^{14n}}, \frac{1}{2^{13n}} \right] \cup \left[ \frac{1}{2^7}, \frac{1}{2^6} \right] = \emptyset \]

luego aquí no se da mezcla, como queríamos demostrar.

### 5.10 Mixing en la ecuación logística

Hemos visto cómo para valores de \( \mu \) adecuados la ecuación logística (y otras de su familia) muestran un comportamiento altamente inestable, que hemos denominado sensibilidad a las condiciones iniciales. La aplicación logística muestra, en el régimen caótico, la propiedad de mixing: si tomamos cualquier intervalo \( I \subset [0,1] \) como punto de partida (tomamos \( x_0 \in I \) veremos que más tarde o más temprano las trayectorias que salen de \( I \) terminan visitando cualquier otro subintervalo \( J \subset [0,1] \) que elimamos.

Para comprobarlo, tomemos \( \mu = 4 \) y una partición \( P \) del intervalo unidad en \( N \) subintervalos:

\[ P = \left\{ I_k \mid I_k \equiv \left[ \frac{k-1}{N}, \frac{k}{N} \right] : k = 1, \ldots, N \right\} \]

y tomemos cualquiera de ellos como punto de partida de nuestras condiciones iniciales \( x_0 \in I_i \in P \). ¿Llegaremos a cualquier otro \( I_j \) por iteración de la ecuación logística? En la figura 5.19 damos un ejemplo de esta situación. Partiendo de \( I_i \) (a la izquierda) alcanzamos \( I_j \) al cabo de cierto número de pasos de tiempo.

Podemos analizar este resultado de forma más sistemática. Siguiendo el tratamiento de Peitgen y otros (1992), tomamos un conjunto de \( 10^4 \) condiciones iniciales \( x_0^{(i)} \in I \subset [0,1] \), siendo

\[ I = [0.2, 0.2 + 10^{-11}] \]

y tomamos un segundo intervalo \( J \) de llegada definido por

\[ J = [0.68, 0.69] \]
Figura 5.19: (a) Mixing en la aplicación logística \((μ = 4)\): partiendo de un intervalo \(I\) dado, alcanzamos, al cabo de cierto número de iteraciones, otro intervalo \(J\) previamente elegido. (b) Dinámica del número de condiciones iniciales que parten de un subintervalo \(I = [0.2, 0.2 + 10^{-11}]\) y que aún no han caído en el intervalo de llegada \(J = [0.65, 0.69]\) al cabo de \(n\) iteraciones.

Seguimos las órbitas que parten de cada condición inicial \(x_0^{(n)} \in I\) y contamos, en cada iteración, el número de "supervivientes" \(S(n)\), esto es, cuántas trayectorias no han caído aún en \(J\) al cabo de \(n\) iteraciones. En la figura 5.19 se muestra el comportamiento de \(S(n)\) en función del número de iteraciones \(n\). Vemos que se da un decaimiento exponencial

\[
S(n) \approx e^{-n/\tau}
\]

siendo \(\tau\) el tiempo característico necesario para reducir el número de supervivientes en un factor \(1/e \approx 0.368\).

5.11 Caos determinista: definición

Hemos establecido ya algunas propiedades básicas que describen, a uno u otro nivel, la presencia de dinámicas aperiódicas en sistemas deterministas. Las propiedades de mezcla (mixing) y sensibilidad a las condiciones iniciales, junto con la presencia de un conjunto infinito de órbitas periódicas, son los ingredientes de la siguiente definición:

**Definición** (*Devaney, 1986*)

Un sistema dinámico \(x_{n+1} = f_\mu(x_n)\) definido sobre el intervalo \(U\) es caótico si verifica las tres propiedades siguientes:

1. Los puntos periódicos de \(f_\mu(x)\) son densos en \(U\).
2. Es topológicamente transitivo (presenta mezcla).
3. Es sensible a las condiciones iniciales.
Señalamos que esta definición, dada inicialmente por Devaney en 1986, ha sido analizada posteriormente por diversos autores. Banks y sus colaboradores demostraron en 1992 que las propiedades 1 y 2 implican 3. Explicitamente, obtuvieron el siguiente resultado:

**Teorema**

Sea \( f_\mu : U \to U \) topológicamente transitiva y supongamos que los puntos periódicos de \( f_\mu \) son densos en \( U \). Si \( U \) contiene un número infinito de elementos, entonces \( f_\mu \) exhibirá sensibilidad a las condiciones iniciales.

Más adelante, en 1994, Vellekoop y Berglund demostraron que, bajo la hipótesis de continuidad de \( f_\mu \) sobre el intervalo de definición de ésta, la propiedad de transitividad topológica implica las otras dos propiedades. En consecuencia, si \( U \subseteq \mathbb{R} \) es un intervalo (no necesariamente finito) sobre el que está definida \( f_\mu \) y ésta es continua, entonces el sistema dinámico anterior será caótico si y solo si tiene la propiedad de mezcla (transitividad topológica).

Pese a estos resultados, siguiendo el tratamiento de otros textos (Peitgen et al., 1992) seguiremos empleando la definición de caos anterior, dado que las propiedades implicadas son de gran interés por sí mismas. A continuación analizaremos algunas ideas suplementarias así como algún ejemplo.

### 5.12 Dinámica simbólica

En esta sección definiremos una aproximación distinta a la dinámica determinista. El enfoque de fondo consiste en sustituir, en cierta forma, la dinámica definida mediante variables continuas por un conjunto de símbolos ( pertenecientes a un conjunto discreto) que permiten tratar las non-linealidades, y el caos en particular, de un modo especialmente útil (Bai-Lin, 1989).

Consideraremos en primer lugar el llamado *espacio de secuencias binarias* \( \Sigma_2 \), definido por:

\[
\Sigma_2 = \left\{ s = (s_0s_1s_2...) \mid s_i \in \{0,1\} \right\}
\]

(de forma general, podríamos emplear \( \Sigma_n \), donde \( s_j \in \{0,1,2,...,n-1\} \)). Los elementos de \( \Sigma_2 \) serán, por definición, secuencias infinitas.

Podemos definir una métrica sobre \( \Sigma_2 \) como sigue. Sean las secuencias \( s, r \in \Sigma_2 \):

\[
s = (s_0s_1s_2...) \\
r = (r_0r_1r_2...)
\]

cualquieras (habitualmente se toma \( r_0, s_0 = 0 \)). La distancia entre estas secuencias se define por:

\[
d[s, r] = \sum_{i=0}^{\infty} \frac{|s_i - r_i|}{2^i}
\]

Puesto que \( |s_i - r_i| \in \{0,1\} \), esta serie infinita estará acotada por la serie geométrica

\[
\sum_{i=0}^{\infty} \frac{1}{2^i} = 2
\]

y por lo tanto es convergente.

Tomemos por ejemplo \( s = (000...) \) y \( r = (1111...) \). Su distancia es, directamente, \( d[s, r] = 2 \). Si empleáramos \( r = (1010...) \), entonces
\[ d[s, r] = \sum_{i=0}^{\infty} \frac{1}{2^i} = \frac{1}{1 - 1/4} = \frac{4}{3} \]

Esta distancia define una métrica sobre el conjunto \( \Sigma_2 \). Este resultado puede comprobarse con facilidad. En primer lugar, debemos tener (y así ocurre):

\[ d[s, r] \geq 0 \quad \forall \; s, r \in \Sigma_2 \]

siendo la igualdad cierta si y sólo si \( s_i = r_i, \forall i \). Además, se da la condición de simetría \( d[s, r] = d[r, s] \) (puesto que \( |s_i - r_i| = |r_i - s_i| \)) y finalmente se tiene que, para cualquier terna \( s, r, t \in \Sigma_2 \), se da la desigualdad

\[ d[r, s] + d[s, t] \geq d[r, t] \]

que tiene lugar, puesto que

\[ |r_i - s_i| + |s_i - t_i| \geq |r_i - t_i| \]

Esta distancia nos permitirá decidir qué subconjuntos de \( \Sigma_2 \) son abiertos o cerrados y, especialmente, la proximidad (distancia) entre secuencias. Un resultado especialmente importante viene dado por el siguiente teorema.

**Teorema**

Sean \( s, t \in \Sigma_2 \) y supongamos que \( s_i = t_i \) para \( i = 0, 1, ..., n \). Entonces, \( d[s, t] \leq 1/2^n \). Recíprocamente, si \( d[s, t] \leq 1/2^n \) entonces \( s_i = t_i \) para \( i \leq n \).

La demostración es simple. Si \( s_i = t_i \) para \( i \leq n \), entonces

\[ d[s, t] = \sum_{i=0}^{n} \frac{|s_i - s_i|}{2^i} + \sum_{i=n+1}^{\infty} \frac{|s_i - t_i|}{2^i} \]

\[ \leq \sum_{i=n+1}^{\infty} \frac{1}{2^i} = \frac{1}{2^n} \]

Por otra parte, si \( s_i \neq t_i \) para algún \( i \leq n \), entonces tendremos

\[ d[s, t] \geq \frac{1}{2^i} \geq \frac{1}{2^n} \]

luego \( d[s, t] < 1/2^n \) y por lo tanto, \( s_i = t_i \) para \( i \leq n \).

Este resultado reviste importancia en la medida en que podremos decidir rápidamente si dos secuencias son o no similares (esto es, próximas en el espacio métrico) siempre que sus primeros símbolos (o bits) coincidan.

Un ingrediente esencial para el estudio de la dinámica simbólica es la aplicación (o operador) *shift* (desplazamiento). Se define por:

\[ \sigma : \Sigma_2 \rightarrow \Sigma_2 \]

\[ (s_0 s_1 s_2 ...) \rightarrow \sigma(s_0 s_1 s_2 ...) = (s_1 s_2 s_3 ...) \]

También se indica por \( 0.s_0 s_1 s_2 ... \rightarrow 0.s_1 s_2 s_3 ... \), y su dinámica es, como se ve, muy simple: \( \sigma \) desplaza hacia la izquierda la secuencia, “olvidando” el primer símbolo. Esta es una aplicación

Esta codificación binaria permite además definir una partición ordenada sobre el intervalo unidad. Si partimos el intervalo $[0, 1]$ en dos mitades iguales, esto es $[0, 1/2)$ y $[1/2, 1]$, los números $x \in [0, 1/2)$ tendrán una CB que empieza por 0.0 y para $x \in [0, 1/2)$ tendremos una CB que empieza por 0.1. Si seguimos dividiendo estos intervalos por la mitad, obtendremos una partición en cuatro subconjuntos para los que tendremos una CB que empezará por:

$$
\begin{align*}
[0, \frac{1}{4}) & \rightarrow 0.00 & \left[\frac{1}{4}, \frac{1}{2}\right) & \rightarrow 0.01 \\
\left[\frac{1}{2}, \frac{3}{4}\right) & \rightarrow 0.10 & \left[\frac{3}{4}, 1\right) & \rightarrow 0.11
\end{align*}
$$

y así sucesivamente.

Esta aplicación es enormemente simple, y el hecho de que esté definida linealmente a trozos (de forma similar a la aplicación triangular que vimos con anterioridad) la hace especialmente interesante. Este operador/función exhibe caos, y permite estudiar las propiedades básicas del caos determinista de forma exhaustiva.

### 5.13 Caos en el operador $\sigma(x)$

En esta sección demostraremos que el sistema dinámico definido a partir del operador shift es caótico, empleando para ello la definición de Devaney anteriormente dada.

#### 5.13.1 Sensibilidad a las condiciones iniciales

De la definición antes dada, debemos demostrar que existe $\delta > 0$ tal que $\forall x \in [0, 1)$, $\forall \epsilon > 0$, existe un punto $y \in [0, 1)$ y un entero $k > 0$ tales que se cumpla

$$
|x - y| < \epsilon
$$

$$
|\sigma^k(x) - \sigma^k(y)| > \delta
$$

En nuestra demostración emplearemos la codificación binaria introducida previamente. Para un $\epsilon$ dado (con $0 < \epsilon \ll 1$) y un $x$ definido por:

$$
x = 0.a_1a_2a_3... \in [0, 1)
$$

Podemos elegir un valor de $k \geq 1$ tal que la desigualdad $2^{-k} < \epsilon$ se cumpla. Elegiremos a continuación el número $y$ como aquel tal que coincide, en su codificación binaria, con todos los dígitos de $x$ excepto aquel que ocupa la posición $k + 1$. Se tiene entonces que

$$
d(x, y) = \frac{|a_{k+1} - b_{k+1}|}{2^{k+1}} = \frac{1}{2^{k+1}} < 2^{-k} < \epsilon
$$

y por otra parte tenemos que:

$$
d(\sigma^k(x), \sigma^k(y)) = d(0.a_{k+1}a_{k+2}... - 0.b_{k+1}b_{k+2}...) = 0.1 - \frac{1}{4} > \epsilon
$$

luego el sistema dinámico es SCL, como queríamos demostrar.
Figura 5.21: Orbitas periódicas obtenidas empleando el sistema dinámico definido por el operador $\Sigma(x)$ (shift), partiendo de dos C1 distintas: $x_1 = 1/3$ y $x_2 = 6/7$.

### 5.13.2 Puntos periódicos densos

Para demostrar esta propiedad, notemos que aquellos puntos sobre el intervalo unidad tales que posean una CB que posea un ciclo periódico serán puntos periódicos de la aplicación $\sigma(x)$. Esto es, si

$$x = 0.a_1a_2a_3\ldots a_k$$

entonces $x = \sigma^k(x)$ (como requiere la definición de órbita periódica de un sistema dinámico). Podemos comprobar que así es aplicando sucesivamente el operador sobre $x$:

$$\sigma(x) = 0.a_2a_3\ldots a_ka_1a_2\ldots a_k$$

$$\sigma^2(x) = 0.a_3\ldots a_ka_1a_2\ldots a_k$$

$$\vdots$$

$$\sigma^{k-1}(x) = 0.a_ka_1a_2\ldots a_k$$

$$\sigma^k(x) = 0.a_1a_2\ldots a_k$$

Podemos dar ejemplos de órbitas periódicas partiendo de condiciones iniciales dadas. En la figura 5.21 se muestra un ejemplo de órbitas periódicas obtenidas partiendo de los puntos $x = 1/3$ y de $x = 6/7$, cuyas codificaciones binarias son $0.0\overline{1}$ y $0.1\overline{1}$, respectivamente.

Para demostrar que estos puntos son densos, debemos ver que $\forall x \in [0,1)$ existen puntos periódicos tan cercanos a $x$ como se quiera, esto es, para cualquier entorno de $B_\epsilon(x)$ de radio $\epsilon$ podemos encontrar un punto periódico $y$ tal que

$$d(x,y) < \epsilon$$

Este resultado puede probarse con facilidad. Supongamos que para un $\epsilon$ dado $x = 0.a_1a_2a_3\ldots$, podemos encontrar un entero $k \geq 1$ tal que $2^{-k} < \epsilon$. Consideremos el punto $y$ cuya codificación binaria es periódica, siendo sus $k$ dígitos (que se repiten) idénticos a los $k$ primeros dígitos de $x$:

$$y = 0.\overline{a_1a_2a_3}$$
que es un punto periódico de periodo menor o igual que $k$ y que, como sabemos, 

$$d(x, y) \leq 2^{-k} < \epsilon$$

luego hemos demostrado que los puntos periódicos son densos en el espacio fásico (el intervalo unidad).

5.13.3 Mixing

Como hemos indicado anteriormente, los números irracionales poseen codificaciones binarias formadas por cadenas infinitas, no periódicas, de dígitos. Podemos intuir que la órbita asociada a un irracional recorrerá el espacio de fases dando lugar a un conjunto de puntos denso en él. Para demostrarlo, consideremos dos subintervalos $I, J \subset [0, 1)$ arbitrariamente pequeños. Queremos ver que siempre podremos encontrar un entero $k \geq 1$ tal que

$$\sigma^k(I) \cap J \neq \emptyset$$

Llamemos $\mu(I), \mu(J)$ a la longitud (medida) de estos intervalos. Podemos elegir un $k \geq 1$ tal que

$$\mu(I) > \frac{1}{2^{k-1}}$$

y supongamos que

$$x_0 = 0.a_1a_2a_3... \in I$$

es el punto medio de este intervalo. Sea

$$y = 0.b_1b_2b_3... \in J$$

un punto arbitrario dentro del segundo intervalo.

Consideremos un tercer punto $x \in [0, 1)$ tal que su CB sea la formada por los $k$ primeros dígitos de $x_0 \in I$ seguidos a continuación por los de $y \in J$:

$$x = 0.a_1a_2a_3...a_kb_1b_2b_3...$$

Ahora, puesto que $x$ y $x_0$ coinciden en sus primeros $k$ dígitos, su distancia será, como sabemos, inferior a $2^{-k}$ y tendremos:

$$d(x, x_0) \leq 2^{-k} = \frac{1}{2^{k-1}} < \frac{1}{2^k} \mu(I)$$

o lo que es lo mismo: la distancia de $x$ al centro de $I$ es inferior que la mitad de la longitud de $I$. Se tiene además que:

$$\sigma(x) = 0.a_2a_3...a_kb_1b_2b_3...$$

$$\sigma^2(x) = 0.a_3a_4...a_kb_1b_2b_3...$$

$$\vdots$$

$$\sigma^{k-1}(x) = 0.a_kb_1b_2b_3...$$

$$\sigma^k(x) = 0.b_1b_2b_3... = y \in J$$

Luego hemos demostrado que existe un $x \in I$ tal que $\sigma^k(x) \in J$, esto es, tal que

$$\sigma^k(I) \cap J \neq \emptyset$$

con lo que queda demostrado.
5.14 Caos en la aplicación triangular

En una sección anterior hemos introducido la denominada aplicación triangular, que, como vimos, exhibe caos determinista para una amplia región de su espacio de parámetros. El exponente de Lyapunov asociado era calculable de forma analítica y tomaba valores positivos para dicho intervalo. Esta aplicación estaba definida por:

\[ x_{n+1} = \Phi(x_n) = \mu \left( 1 - 2 \left| \frac{1}{2} - x_n \right| \right) \]

y en esta sección demostraremos, empleando la definición anterior, que para \( \mu = 1 \), esto es, para

\[ x_{n+1} = \begin{cases} 
2x & \text{si } x \in [0, 1/2) \\
2(1 - x) & \text{si } x \in [1/2, 1) 
\end{cases} \]

la dinámica de esta aplicación exhibe caos determinista (seguiremos, como antes, el desarrollo presentado por Martínez et al., 1994). Para llevar a cabo esta demostración emplearemos el formalismo anterior basado en la dinámica simbólica así como el operador shift previamente definido. Dicho operador, \( \sigma(x) = \text{frac}(2x) \), tiene propiedades muy útiles en el estudio del caos.

Señalamos, en primer lugar, que la función \( \text{frac}(x) \) cumple, para cualquier \( m \in \mathbb{N} \), las siguientes propiedades:

\[ \text{frac}(x + m) = \text{frac}(x) \]
\[ \text{frac}(mx) = \text{frac}(m \text{frac}(x)) \]

por inducción, puede probarse además que:

\[ \sigma^k(x) = \text{frac}(2^k x) \]

Para continuar, veremos qué relación guardan las funciones \( \Phi \) y \( \sigma \). Si tomamos \( \sigma(1) = 1 \) (que es el valor que obtendríamos al aplicar dicho operador sobre \( x = 1 \)), entonces se tiene que

\[ \Phi^{k+1} = \Phi \circ \sigma^k \quad \forall k \geq 1 \]

o, lo que es lo mismo, \( \Phi^{k+1}(x) = \Phi(\sigma^k(x)) \) para cualquier valor de \( x \) sobre el intervalo unidad.

Podemos demostrar este resultado por inducción. Para el caso \( k = 1 \), se tiene que \( \sigma(x) = \Phi(x) \) luego \( \Phi^2(x) = \Phi(\sigma(x)) \). Si \( x \in [1/2, 1) \), se tiene que \( \Phi(x) = 2(1 - x) \) y \( \sigma(x) = 2x - 1 \) y dado que \( \Phi^2(x) + \sigma(x) = 1 \) y \( \Phi(x) \) es simétrica sobre el intervalo se cumplirá que \( \Phi(2 - 2x) = \sigma(2x - 1) \), es decir, \( \Phi^2(x) = \Phi(\sigma(x)) \).

Siguiendo el procedimiento de inducción, asumiremos que se cumple para \( k = n \) y, empleando ambos resultados (para \( k = 1 \) y \( k = n \)) tendremos que:

\[ \Phi^{n+2} = \Phi \circ \Phi^{n+1} = \Phi \circ \Phi \circ \sigma^n = \Phi \circ \sigma \circ \sigma^n = \Phi \circ \sigma^{n+1} \]

con lo que queda demostrado.

Finalmente, antes de estudiar las condiciones para demostrar la existencia de caos, transformaremos la aplicación triangular \( \Phi \) en términos de codificaciones binarias. Sea ahora \( x = a_1a_2a_3... \in [0, 1) \). Como sabemos, si \( a_1 = 0 \), el punto \( x \) pertenecerá a la primera mitad del intervalo unidad, i.e. \( x \in [0, 1/2) \). Tendremos entonces:

\[ \Phi(x) = \sigma(x) = 0.a_2a_3a_4... \]

luego ambos operadores actúan en igual forma sobre los puntos de la primera mitad. En cambio, si \( x \in [1/2, 1) \), tendremos:

\[ \Phi(x) = 2(1 - x) = 1 - \sigma(x) \]


\[ 1 - 0.a_2a_3a_4... = 0.a_2a_3a_4... \]

siendo $a^* = 1$ si $a = 0$ y $a^* = 0$ en caso contrario. Si tenemos en cuenta que

\[ 0.a_2a_3a_4... + 0.a_2^*a_3^*a_4^*... = 0.111... = 1 \]

el operador correspondiente a la aplicación triangular se escribirá como

\[ \Phi(0.a_1a_2a_3...) = \begin{cases} 
0.a_2a_3a_4 & \text{si } a_1 = 0 \\
0.a_2^*a_3^*a_4^* & \text{si } a_1 = 1 
\end{cases} \]

a continuación demostraremos que esta aplicación es caótica.

### 5.14.1 Puntos periódicos densos

Sea $x$ tal que $\sigma^k(x) = x$ (un punto $k$-periódico del operador shift. Se tiene entonces que:

\[ \Phi(x) = \Phi(\sigma^k(x)) = \Phi^{k+1}(x) = \Phi^k(\Phi(x)) \]

y por lo tanto $\Phi(x)$ será un punto $k$-periódico de la aplicación triangular. Vimos anteriormente que aquellos puntos tales que su CB estaba formada por un ciclo $k$-periódico, esto es, de la forma

\[ 0.a_1a_2...a_k \]

son puntos $k$-periódicos del operador shift. En consecuencia, todas las codificaciones binarias

\[ \Phi(0.a_1a_2a_3...a_k) = \begin{cases} 
0.a_2a_3...a_k0 & \text{si } a_1 = 0 \\
0.a_2^*a_3^*...a_k^*0 & \text{si } a_1 = 1 
\end{cases} \]

son puntos periódicos de periodo $k$ para la aplicación triangular. Los puntos $x$ tales que su codificación binaria se escriba en la forma

\[ x = 0.a_1a_2...a_k0 \]

serán puntos $k$-periódicos de $\Phi(x)$.

Consideremos a continuación el punto

\[ x = 0.a_1a_2a_3... \in [0, 1) \]

y sea un $\epsilon > 0$. Elegiremos un $k \geq 1$ tal que $2^{-k} < \epsilon$ y consideraremos el punto periódico de periodo $k + 1$ dado por:

\[ y = 0.a_1a_2a_3...a_k0 \in [0, 1) \]

que coincide en los primeros $k$ dígitos con $x$, luego su distancia será tal que

\[ d(x, y) \leq 2^{-k} < \epsilon \]

Con lo que queda demostrada la propiedad.

En la figura 5.22 (a,b) se muestran las órbitas correspondientes a los puntos $0.110 = 6/7$ y $0.0010 = 2/15$. 
Figura 5.22: Orbitas periódicas para la aplicación triangular $\Phi(x)$. Los puntos iniciales son: (a) $x = \frac{6}{7}$ y (b) $x = \frac{2}{15}$.

### 5.14.2 Sensibilidad a las condiciones iniciales

Sea $\epsilon > 0$, tomemos $\delta = \frac{1}{4}$ y sea

$$x = 0.a_1a_2a_3... \in [0,1)$$

un punto cualquiera. Sea dado un $k$ tal que $2^{-k} < \epsilon$ y a continuación elegimos como $y$ un punto del intervalo unidad que sólo difiera de $x$ en el digito $k+1$-ésimo:

$$y = 0.a_1a_2a_3...a_k\hat{a}_{k+1}a_{k+2}a_{k+3}...$$

con lo que

$$d(x,y) \leq 2^{-k} < \epsilon$$

Por otra parte, si $a_k = 0$, tendremos:

$$|\Phi^k(x) - \Phi^k(y)| = |\Phi(\sigma^{k-1}(x) - \Phi\sigma^{k-1}(y)| =$$

$$|\Phi(0.a_k\sigma_{k+1}a_{k+2}a_{k+3}...) - 0.a_k\hat{a}_{k+1}a_{k+2}a_{k+3}...)|$$

$$|0.a_{k+1}a_{k+2}a_{k+3}...a_{k+1}a_{k+2}a_{k+3}...| = 0.1 = \frac{1}{2} > \delta$$

y, si $a_k = 1$, se tiene:

$$|\Phi^k(x) - \Phi^k(y)| = |0.a_{k+1}a_{k+2}a_{k+3}...a_{k+1}a_{k+2}a_{k+3}...| = 0.1 = \frac{1}{2} > \delta$$

con lo que queda demostrado.
5.14.3 Mixing

Finalmente, probaremos que la aplicación triangular posee la propiedad de mezcla. Consideremos, siguiendo la definición, dos subintervalos \( I, J \subset [0, 1) \) arbitrariamente pequeños. Queremos ver que siempre podremos encontrar un entero \( k \geq 1 \) tal que

\[ \sigma^k(I) \cap J \neq \emptyset \]

El procedimiento será idéntico básicamente al desarrollado para el operador shift.

Llamemos \( \mu(I), \mu(J) \) a la longitud (medida) de estos intervalos. Podemos elegir un \( k \geq 1 \) tal que

\[ \mu(I) > \frac{1}{2^k - 1} \]

y sean ahora los puntos:

\[ x_0 = 0.a_1a_2a_3\ldots \in I \]

(en el centro de este intervalo) y

\[ y_0 = 0.b_1b_2b_3\ldots \in J \]

un punto arbitrario sobre \( J \). Si un punto \( y \) posee los mismos \( k \) primeros dígitos que \( x_0 \), entonces \( x \in I \), ya que si

\[ x = 0.c_1c_2c_3\ldots c_{k+1}c_{k+2}\ldots \]

entonces

\[ d(x, x_0) \leq 2^{-k} = \frac{1}{2} \mu(I) \]

y un razonamiento idéntico puede hacerse para un punto \( y \) que tenga la misma proximidad a \( y_0 \). Si \( a_k = 0 \) y elegimos

\[ x = 0.a_1a_2\ldots a_kb_1b_2\ldots b_k \in I \]

se tiene que

\[ \Phi^k(x) = \Phi(\sigma^{k-1}(x)) = \Phi(0.a_kb_1b_2\ldots b_k) = 0.b_1b_2\ldots b_k \in J \]

Si, por otra parte, \( a_k = 1 \) elegimos

\[ x = 0.a_1a_2\ldots a_kb_1^*b_2^*\ldots b_k^* \in I \]

y tendremos ahora, del mismo modo:

\[ \Phi^k(x) = \Phi(\sigma^{k-1}(x)) = \Phi(0.a_kb_1^*b_2^*\ldots b_k^*) = 0.b_1b_2\ldots b_k \in J \]

Luego en cualquier caso se cumple que, dado un punto \( x \in I \) encontramos un \( k \) para el que se cumple \( \Phi^k(x) \in J \), y por tanto

\[ \Phi^k(I) \cup J \neq \emptyset \]
5.14.4 Consecuencias: Caos en $\mu x(1 - x)$

Los resultados anteriores son extensibles a la aplicación logística $f_\mu(x) = \mu x(1 - x)$ para el caso límite $\mu = 4$. Para demostrar la caosidad de este sistema dinámico podemos recurrir a la aplicación triangular, que hemos estudiado en la sección anterior. Consideremos la siguiente función

$$y = H(x) = \sin^2\left(\frac{\pi x}{2}\right)$$

que define una aplicación biyectiva sobre los puntos del intervalo unidad. Puede demostrarse fácilmente (Holmgren, 1994; Devaney, 1986) que

$$f_\mu(H(x)) = H(\Phi(x))$$

y a partir de esta relación puede probarse que si $x$ es un punto $k$-periódico de la aplicación triangular, entonces $H(x)$ es un punto $k$-periódico de la aplicación logística (y viceversa). Este resultado permite demostrar el comportamiento caótico de la aplicación logística.

Señalemos, finalmente, que la partición del intervalo unidad en términos de codificación binaria de símbolos permite introducir una conexión entre las secuencias de dígitos que resultan de la dinámica de la aplicación logística para $\mu = 4$ y la de un proceso aleatorio como es el lanzamiento de una moneda. En 1973, Metrópolis, Stein y Stein emplearon la partición binaria del intervalo para analizar esta relación. Una secuencia de caras y cruces generadas por el lanzamiento de una moneda se denomina secuencia de Bernoulli. Una posible definición de caos determinista (Jackson, 1991) es que, para cualquier secuencia de Bernoulli dada, existe una condición inicial que permite generar dicha secuencia.

5.15 La herradura de Smale

Un sistema dinámico muy bien estudiado y que volverá a aparecer más adelante cuando exploraremos el problema del caos en sistemas continuos es la denominada aplicación de Smale. En la figura 5.23 vemos el procedimiento básico de actuación de este sistema dinámico. Partiendo del cuadrado unidad $U \equiv [0, 1] \times [0, 1]$, llevamos a cabo un estramamiento del mismo hasta formar la denominada herradura de Smale, que solapamos entonces encima de $U$. A continuación, nos quedamos sólo con aquellos puntos que se hallan en la intersección entre ambos objetos, y repetimos la misma iteración.

De forma similar a lo que vimos en la sección de la transformación del panadero, podemos imaginar que esta dinámica queda definida por cierta aplicación bidimensional

$$f_\mu : U \rightarrow U$$

que actúa sobre los puntos de $U$. Asumiremos (razonablemente, a partir de la definición anterior) que esta aplicación es continua e invertible. La aplicación inversa $f_\mu^{-1}$ transformará la herradura $f_\mu(U)$ en el cuadrado invirtiendo los pasos anteriores.

Si llevamos a cabo iteraciones sucesivas, veremos que el resultado de aplicar $f_\mu$ es la generación de bandas verticales de grosor cada vez menor. Si indicamos las bandas verticales resultantes de la primera iteración por $V_1$ y $V_2$, tendremos que:

$$U \cap f_\mu(U) = V_1 \cup V_2$$

Si ahora llevamos a cabo una segunda iteración, las bandas anteriores se transforman en cuatro bandas nuevas de menor grosor (figura 5.24) y que indicaremos por

$$V_{11}, V_{12}, V_{21}, V_{22}$$
Figura 5.23: (a) Sistema dinámico que genera la herradura de Smale: a partir del cuadrado unidad $U$, llevamos a cabo un estramiento-plegado del mismo y solapamos la herradura con el cuadrado unidad, reteniendo sólo aquellos puntos que pertenecen a la intersección. (b) Repetimos este procedimiento e indicamos los conjuntos de puntos resultantes.

Figura 5.24: Bandas verticales y horizontales generadas por la aplicación de Smale.
Figura 5.25: Intersección entre bandas verticales y horizontales. La aplicación reiterada de la aplicación de Smale da lugar, asintóticamente, a un conjunto de Cantor.

escribiremos ahora

\[ U \cap f_{\mu}(U) \cap f_{\mu}^2(U) = V_{11} \cup V_{12} \cup V_{22} \cup V_{21} \]

y, de manera similar, tendremos que las bandas horizontales obtenidas por aplicación de la inversa \( f_{\mu}^{-1} \) serán tales que:

\[ U \cap f_{\mu}^{-1}(U) = H_1 \cup H_2 \]

y

\[ U \cap f_{\mu}^{-1}(U) \cap f_{\mu}^{-2}(U) = H_{11} \cup H_{12} \cup H_{22} \cup H_{21} \]

Observemos que:

\[ f_{\mu}(H_i) = V_i \quad i = 1, 2 \]

\[ f_{\mu}^2(V_{ij}) = H_{ij} \quad i, j = 1, 2 \]

y que por iteración del sistema de Smale vamos obteniendo, después de \( k \) iteraciones, \( 2^k \) bandas verticales en la intersección \( U \cap f_{\mu}^k(U) \) (\( k = 1, 2, \ldots \)) y de forma similar la inversa nos dará, después de \( k \) iteraciones, \( 2^k \) bandas horizontales en la intersección \( U \cap f_{\mu}^{-k}(U) \) (\( k = 1, 2, \ldots \)).

La mayor parte de los puntos de \( U \) abandonarán el cuadrado unid al iterar la aplicación o su inversa. Si olvidamos estos puntos, podemos preguntarnos qué estructura tendrá el conjunto invariante \( \Lambda \) que queda en el cuadrado después de iterar indefinidamente, esto es,

\[ \Lambda = \left\{ x \in U ; f_{\mu}^k(x) \in U, \forall -\infty < k < \infty \right\} \]

Este conjunto es de hecho la intersección infinita entre las iteraciones empleadas:

\[ \Lambda = \cdots \cap f_{\mu}^{-k}(U) \cap \cdots \cap f_{\mu}(U) \cap f_{\mu}^{-2}(U) \cap f_{\mu}^{-1} \cap U \cap f_{\mu}(U) \cap f_{\mu}^2(U) \cap \cdots \cap f_{\mu}^k(U) \cap \cdots \]
Este conjunto será un objeto fractal. Para la primera iteración de ambas aplicaciones, el conjunto invariante deberá estar incluido en la intersección dada por:

$$f^{-1}_\mu(U) \cap U \cap f_\mu(U)$$

que estará formado por cuatro cuadrados (figura 5.25 (a)). Iterando por segunda vez, la intersección en la que se encontrará incluido Λ será ahora un conjunto de dieciséis cuadrados menores (figura 5.25 (b)) definido por:

$$f^{-2}_\mu(U) \cap f^{-1}_\mu(U) \cap U \cap f_\mu(U) \cap f^2_\mu(U)$$

El conjunto resultante es obviamente auto-similar (un conjunto de Cantor).

Podemos emplear el formalismo de la dinámica simbólica para analizar sus propiedades (Wiggins, 1990; Kuznetsov, 1995). Puede demostrarse que existe una biyección entre Λ y el conjunto $\Omega_2$ de las secuencias binarias bi-infinitas, esto es, el conjunto de secuencias de la forma:

$$\omega = \ldots a_{-2}a_{-1}a_0a_1a_2\ldots$$

donde

$$\omega_k = \begin{cases} 1 & \text{si } f^k_\mu(x) \in H_1 \\ 0 & \text{si } f^k_\mu(x) \in H_2 \end{cases}$$

para $k = 0, \pm 1, \pm 2, \ldots$ y donde $f^0_\mu$ es la aplicación identidad.

Esta equivalencia permite, de forma algo más complicada que en los ejemplos anteriores, demostrar la presencia de caos en este sistema dinámico. El siguiente teorema (Smale, 1963) resume los resultados más importantes acerca del comportamiento de esta aplicación:

**Teorema (Smale, 1963)**

La aplicación de Smale $f_\mu$ posee un conjunto invariante Λ que contiene un conjunto numerable de órbitas periódicas de periodicidad arbitrariamente grande, y un conjunto no-numerable de órbitas no periódicas, entre las cuales están aquellas que pasan arbitrariamente cerca de cualquier punto de Λ.

Sin entrar en más detalles (véase Wiggins, 1990) indiquemos que la aparición de este tipo de estructuras en herradura (que reencontraremos más adelante) sirve para detectar la aparición de caos así como para caracterizar formalmente los atractores extraños en sistemas disipativos y conservativos (véase el capítulo sobre caos hamiltoniano, 17).

Otra propiedad importante del ejemplo de la herradura de Smale es que podemos perturbar ligeramente la aplicación sin que tengan lugar cambios cualitativos de importancia sobre la dinámica, lo que garantiza la estabilidad estructural del comportamiento.

### 5.16 Universalidad en aplicaciones cuadráticas

La existencia de una constante universal que caracteriza las propiedades globales del escenario de bifurcación con duplicación de periodo es de enorme importancia. La constante de Feigenbaum es universal en un sentido formal bien definido. Sea

$$\mathcal{F}_\mu = \{ f_\mu(x) \}$$

el conjunto de funciones uniparamétricas (conocidas como aplicaciones cuadráticas) que verifican las siguientes condiciones:

1. $f_\mu(x) \in C^1[0, 1]$
Figura 5.26: Orbitas superestables para la aplicación logística. Se indica los puntos $\mu'_i$ en los que aparecen.

2. Existe un punto $x_m$ tal que $f''(x_m) \neq 0$
3. $f_\mu(x)$ es monótona en $[0, x_m)$ y en $(x_m, 1]$
4. $f_\mu(x)$ posee una derivada de Schwarz $S(f_\mu)$ negativa, esto es,

$$S(f_\mu) = \frac{f'''_\mu(x)}{f'_\mu} - 3 \left( \frac{f''_\mu(x)}{f'_\mu} \right)^2 < 0$$

Esta clase de funciones exhibirá un escenario de Feigenbaum caracterizado por la misma constante universal $\delta$:

$$\delta = \lim_{k \to \infty} \frac{\mu_{k+1} - \mu_k}{\mu_{k+2} - \mu_{k+1}} \approx 4.6692$$

Mitchel Feigenbaum (1978) demostró esta universalidad empleando el método del grupo de renormalización (discutido, en otro contexto, en el capítulo 7). Una primera constante universal es $\delta$, que puede escribirse como varible en la ley asintótica de escala (es decir, para $n \to \infty$),

$$\mu_n \approx \mu_\infty - c\delta^{-n}$$

siendo $c$ una constante.

La segunda constante universal está relacionada con la aparición de órbitas superestables, esto es, aquellas órbitas $O^{(p)}_\mu$ tales que

$$|\lambda^{(p)}_\mu| = 0$$

que aparecerán sucesivamente en los puntos $\mu'_1$, $\mu'_2$, $\mu'_3$, ... que se indican en la figura 5.26.

Los ciclos superestables obedecen una relación $d_n/d_{n+1} \approx -\alpha$. Estos ciclos incluyen el punto $x = 0.5$, y $d_n$ es la distancia
entre \( x = 1/2 \) y el punto de \( O_{\mu^*} \), más cercano a éste. Tendremos por lo tanto una nueva ley de escala,

\[
d_n \approx d_\infty - c^\alpha \alpha^{-n}
\]

como la anterior para \( n \to \infty \). El exponente \( \alpha \) se interpreta de forma análoga a como se ha hecho con los exponentes de la teoría de transiciones de fase (capítulo 7). Empleando las ecuaciones del grupo de renormalización, Feigenbaum demostró que

\[
\alpha = 2.5029078751... \\
\delta = 4.6692016091... \\
\mu_{\infty} = 3.5699456...
\]

En la teoría de fenómenos críticos, la universalidad se obtiene debido a que el método del grupo de renormalización proporciona un conjunto de ecuaciones cuyos puntos fijos describen cierto tipo de transiciones de fase. A medida que nos aproximamos a este punto fijo estable, los detalles de la condición inicial de la que partimos se pierden. Por lo tanto, la universalidad de los exponentes críticos surge si el punto fijo posee una cuenca de atracción finita.

La idea de Feigenbaum fue encontrar una descripción iterativa del proceso de duplicación de periodo tal que, en el límite, proporcionara la misma solución (punto fijo) para todas las aplicaciones cuadráticas definidas sobre el intervalo, con independencia de sus detalles particulares (véase McCauley, 1994, para una discusión más general).

Feigenbaum parte de la composición entre funciones (que son ahora los objetos matemáticos de interés) en la forma

\[
f^{(2^n)}_{\mu^*} \circ f^{(2^n)}_{\mu^*} = f^{(2^{n+1})}_{\mu^*}
\]

(5.16.1)

e introduce la siguiente hipótesis de escala,

\[
f^{(2^{n+1})}_{\mu^*}(x) \approx (-\alpha)^n f^*_\mu ((-\alpha)^n x)
\]

(5.16.2)

siendo \( f^*_\mu \) el punto fijo de 5.16.1. Si combinamos 5.16.1 y 5.16.2, tenemos que

\[
f^{(2^n)}_{\mu^*} \circ f^{(2^n)}_{\mu^*}(x) = (-\alpha)^{-n} f^*_\mu [(-\alpha)^n (-\alpha)^{-n} f^*_\mu ((-\alpha)^n x)]
\]

(5.16.3)

y dado que también se tiene

\[
f^{(2^{n+1})}_{\mu^*}(x) \approx (-\alpha)^{-n-1} f^*_\mu [(-\alpha)^{n+1} x]
\]

(5.16.4)
podemos combinar 5.16.1 y 5.16.3 para obtener

\[
(-\alpha)^{-n} f^*_\mu ((-\alpha)^{n+1} x) \approx (-\alpha)^n f^*_\mu ((\mu^*)^{-n} x)
\]

así como

\[
f^*_\mu (x) = -\alpha f^*_\mu (x/\alpha)
\]
todo ello definido en el espacio de aplicaciones cuadráticas, no invertibles.
Dada una solución \( f^*(x) \), es fácil probar que

\[
g(x, \mu) = \mu f^* \left( \frac{x}{\mu} \right)
\]
es también solución para cualquier \( \mu \). Podemos establecer la escala tomando \( \mu f^*(0) = 1 \). Escribiremos las soluciones de

\[
g(x) = -\alpha g \left( \frac{x}{\alpha} \right)
\]
en la forma

\[
g(x) = 1 + bx^2 + \ldots
\]
con \( b < 0 \). Este desarrollo puede resolverse aproximadamente (a orden \( x^2 \)) dando \( 1 = -\alpha(1 + b) \) y \( b = -\alpha/2 \). La condición \( b < 0 \) da

\[
\alpha = 1 + \sqrt{3} \approx 2.723
\]
que es una buena aproximación para \( \alpha = 2.5029 \ldots \)

Un desarrollo formal basado en esta aproximación permite obtener de forma análoga los valores de \( \delta \) y \( \mu_\infty \) (Schuster, 1989). En la siguiente sección daremos un procedimiento distinto, desarrollado por R. May y G. Oster.

### 5.17 Universalidad: aproximación de May-Oster

Podemos obtener una buena estimación de la constante \( \delta \) de Feigenbaum siguiendo un procedimiento distinto al del grupo de renormalización. Este planteamiento fue desarrollado por May y Oster (1976).

Sea \( f^{(k)}_{\mu}(x) \) la \( k \)-ésima iteración de la aplicación \( f_{\mu}(x) \). Un ciclo de periodo \( p \) queda definido, como ya sabemos, por el conjunto

\[
O^{(k)}_{\mu} = \left\{ x^{*(k)}_i, \quad i = 1, \ldots, k; \quad x^{*(k)}_i = f^{(k)}_{\mu}(x^{*(k)}_i) \right\}
\]

Sea \( \lambda^{(k)}(\mu) \) el valor asociado a la estabilidad de dicha órbita,

\[
\lambda^{(k)}(\mu) = \left| \prod_{i=1}^{k} \partial_{x} f^{(k)}_{\mu}(x^{*(k)}_i) \right|
\]
es decir, la pendiente de la \( k \)-ésima iteración de la aplicación. Cualquier ciclo estable de periodo \( k \) aparece, como sabemos, para cierto valor \( \lambda^{(k)} = 1 \) y pierde su estabilidad (dando lugar a una órbita de periodo \( 2k \), i.e. \( O^{(k)}_{\mu} \)) cuando \( \lambda^{(k)} = -1 \).

Para ir de un límite al otro (de \( \lambda^{(k)} = +1 \) a \( \lambda^{(k)} = -1 \)) el parámetro \( \mu^{(k)} \) sufre cierta variación, que indicaremos por \( \Delta \mu^{(k)} \). En lo que sigue, estudiaremos el comportamiento del límite

\[
\delta = \lim_{k \to \infty} \frac{\Delta \mu^{(k)}}{\Delta \mu^{(2k)}}
\]
esto es, del cociente entre la variación de \( \mu \) en dos intervalos de duplicación de periodo consecutivos.

Llamaros \( \mu^{(k)}_0 \), al valor de \( \mu \) tal que

\[
\lambda^{(k)}(\mu^{(k)}_0) = +1
\]
Figura 5.27: Bifurcaciones con duplicación de periodo. Se indica la notación empleada en el cálculo de la constante de Feigenbaum \( \delta \) mediante el método de May-Oster.

esto es, al valor de \( \mu \) para el que aparece un ciclo de periodo \( k \). Indicaremos (figura 5.27) los siguientes valores de \( \mu \) por:

\[
\mu = \mu_0^{(k)} + \epsilon
\]

Un desarrollo de Taylor nos da, para valores cercanos a \( \mu \),

\[
\lambda^{(k)}(\mu) = 1 + \epsilon \left( \frac{\partial \lambda^{(k)}}{\partial \mu} \right)_{\mu_0^{(k)}} + O(\epsilon^2)
\]

Indicaremos la primera derivada mediante la siguiente notación:

\[
A_0^{(k)} = \left( \frac{\partial \lambda^{(k)}}{\partial \mu} \right)_{\mu_0^{(k)}} = \left[ \frac{\partial^2 \mu f_\mu^{(k)}(x)}{\partial \mu^2} \right]_{\mu_0^{(k)}}
\]

Empleando estas definiciones, podemos estimar el valor de \( \delta \) mediante el cálculo del cociente entre \( A_0^{(2k)} \) y \( A_0^{(k)} \). Sin embargo, \( A_0^{(k)} \) no puede ser, en general, calculado analíticamente. Sin embargo, podemos relacionar la pendiente \( \lambda^{(2k)} \) asociada con \( f_\mu^{(2k)}(x) \) con \( \lambda^{(k)} \), asociada a \( f_\mu^{(k)}(x) \).

La órbita de periodo \( 2k \) aparece con \( \lambda^{(k)} = 1 \) y se hace inestable para \( \lambda^{(k)} = -1 \), en el punto para el cual emerge la órbita \( 2k \)-periódica, esto es, para \( \lambda^{(2k)} = 1 \). Esta órbita a su vez se hará inestable para \( \lambda^{(2k)} = -1 \). En este punto, el escalar \( \lambda^{(k)} \) tomará cierto valor negativo, que indicaremos por \( \lambda^{(k)}_c \).

Si sustituimos en el desarrollo en Taylor anterior para \( \lambda^{(k)}(\mu) \) e ignoramos los términos de segundo orden, obtenemos:

\[
\Delta \mu(k) = -\frac{2}{A_0^{(k)}}
\]

\[
\Delta \mu(2k) = -\frac{\lambda^{(k)}_c + 1}{A_0^{(k)}}
\]
Figura 5.28: (a) La aplicación $f^{(k)}_\mu$ se muestra en las proximidades de uno de sus puntos $k$-periódicos, $x^{*(k)}_i$. Para la curva discontinua, la pendiente de $f^{(k)}_\mu$ en la intersección con la bisectriz es tal que $|f^{(k)}_\mu| < 1$, y el punto fijo es estable. Para la curva continua, se tiene $|f^{(k)}_\mu| > 1$, esto es, el punto fijo es inestable. (b) La aplicación correspondiente a la bifurcación consecutiva, $f^{(2k)}_\mu$, se muestra en las proximidades del mismo punto fijo $k$-periódico. Como se discute en el texto, la aparición de la instabilidad de la órbita $k$-periódica viene acompañada por una bifurcación en la aplicación $f^{(2k)}_\mu$ dando lugar a dos nuevos puntos fijos de periodo $2k$, indicados por $x^{*(2k)}_i$, a cada lado de $x^{*(k)}_i$.

de donde llegamos al siguiente resultado:

$$\delta = \lim_{k \to \infty} -\frac{2}{\lambda^{(k)}_c + 1}$$

Queda por encontrar una relación asintótica entre los valores $\lambda^{(k)}_1$ y $\lambda^{(2k)}_c$, con los que podremos determinar $\lambda^{(k)}_c$. Podemos emplear para ello, siguiendo a May-Oster, un polinomio cúbico (véase la figura 5.28) que se compone de la forma en que $f^{(2k)}_\mu$ exhibe la bifurcación cuando el ciclo de periodo $k$ se hace inestable.

Supongamos que desarrollamos $f^{(2k)}_\mu$ en serie de Taylor hasta tercer orden alrededor de $x^{*(k)}_i$:

$$f^{(2k)}_\mu(x^{*(k)}_i + \xi) \approx x^{*(k)}_i + A\xi + \frac{1}{2} B\xi^2 + \frac{1}{6} C\xi^3 + O(\xi^4)$$

Necesitamos llevar a cabo este desarrollo hasta términos cúbicos (al menos) para tener en cuenta, de forma apropiada, la forma cualitativa de $f^{(2k)}_\mu$ en las proximidades de $x^{*(k)}_i$, esto es, para describir la bifurcación en un par de nuevos puntos, cada uno a cada lado de $x^{*(k)}_i$. Recordemos que

$$f^{(2k)}_\mu(x^{*(k)}_i + \xi) = f^{(k)}_\mu(f^{(k)}_\mu(x^{*(k)}_i + \xi)) = f^{(k)}_\mu(x^{*(k)}_i + \lambda^{(k)}_c \xi + ...$$

Podemos expresar los coeficientes del desarrollo ($A$, $B$, ... ) en términos de las derivadas de la aplicación de orden inferior, $f^{(k)}_\mu$. En este caso encontramos:

$$A = (\lambda^{(k)}_c)^2$$

$$B = (\lambda^{(k)}_c) \left[ (1 + \lambda^{(k)}_c) \frac{\partial^2 f}{\partial x^2} \right]_{x=x^{*(k)}_i}$$
Notemos que los puntos fijos de periodo \( k \) se bifurcan en \( \lambda^{(k)} = -1 \). Por tanto, para \( \mu \) justo por encima de este valor, \( A \approx 1 \) y \( B \approx 0 \). Para los puntos fijos de periodo \( 2k \), escribiremos
\[ x^{(2k)} = x^{(k)} + \xi^* \]. Los valores de \( \xi^* \) se determinan a partir de:
\[ \xi^* \approx A\xi^* + \frac{1}{2} B\xi^* + \frac{1}{6} C\xi^* \]

Aparte de la solución \( k \)-periódica degenerada \( (\xi^* = 0) \). existe un par de soluciones dadas por
\[ 0 \approx (A - 1) + \frac{1}{2} B\xi^* + \frac{1}{6} C\xi^* \]

Notemos que estas soluciones serán reales si y sólo si \( A \geq 1 \) (esto es, si \( |\lambda^{(k)}| \geq 1 \)) de acuerdo con lo expuesto anteriormente. A continuación, notemos que la pendiente en estos puntos \( 2k \)-periódicos es:
\[ \lambda^{(2k)} = \left[ \frac{\partial^2 f^{(2k)}}{\partial \xi} \right]_{\xi = \xi^*} \approx A + B\xi^* + \frac{1}{2} C\xi^* \]

Empleando las relaciones anteriores, podemos simplificar esta expresión para obtener
\[ \lambda^{(2k)} \approx (3 - 2A) - \frac{1}{2} B\xi^* \]

La cantidad \( \xi^* \) representa la distancia entre el punto periódico inestable de periodo \( k \) y el punto inicialmente estable asociado a la órbita de periodo \( 2k \), y será una cantidad pequeña incluso para valores bajos de \( k \). Además, el coeficiente \( B \) es cero cuando tiene lugar la primera bifurcación, luego podemos despreciar el término \( B\xi^* / 2 \) de la última igualdad.

Con estas aproximaciones llegamos a la expresión para el valor de \( A \) en el punto para el que \( \lambda^{(2k)} \to -1 \) (donde la \( 2k \)-órbita se hace inestable): \( 2A \approx 4 \), esto es, \( A \approx 2 \). Si empleamos ahora la ecuación anterior que relacionaba \( A \) con \( \lambda^{(k)} \), tenemos para el valor correspondiente, (que habíamos llamado \( \lambda^{(k)} \)).
\[ \lambda^{(k)} \approx -\sqrt{2} + O(\delta^{-1}) \]

Sustituyendo este resultado en la ecuación para \( \delta \), llegamos por fin al resultado analítico para la aproximación a la constante de Feigenbaum:
\[ \delta \approx 2(1 + \sqrt{2}) \approx 4.828... \]

Comparando este valor con el valor exacto, vemos que existe un acceptable error del dos por ciento.

### 5.18 Período tres implica caos

En 1975 apareció un artículo de Tien Li y Jim Yorke encabezado por un sugestivo título: "Period Three Implies Chaos" (Periodo tres implica caos) en el que daban una elegante demostración, para sistemas dinámicos unidimensionales, acerca de la existencia de puntos periódicos de distintas periodicidades. Este teorema está muy directamente relacionado con otro muy simple, pero fundamental, de la teoría de sistemas dinámicos demostrado por A. Sarkovsky (Devaney, 1986). El teorema, enunciado explícitamente, dice:
Teorema (Li y Yorke, 1975)

Sea \( x_{n+1} = f_\mu(x_n) \) un sistema dinámico definido sobre un intervalo \( X \subset \mathbb{R} \) que presenta un punto periódico de período tres. Entonces, este sistema dinámico tendrá puntos periódicos de todos los periodos posibles.

Para demostrar este teorema, seguiremos el procedimiento de Martín et al. (1993), dando en primer lugar cuatro lemas básicos (que no demostraremos, véase Martín et al. 1995). Estos son:

Lema 1. Sea \( x_{n+1} = f_\mu(x_n) \) un sistema dinámico definido sobre un intervalo \( U \subset \mathbb{R} \). Si este sistema posee un punto periódico de período tres, entonces

\[
\exists a \in U \ ; \ f_\mu^3(a) = a < f_\mu(a) < f_\mu^2(a)
\]

o bien

\[
f_\mu^2(a) < f_\mu(a) < a < f_\mu^3(a)
\]

Lema 2. Sea \( I \subset \mathbb{R} \) y \( g: I \to \mathbb{R} \) una función continua. Para todo compacto \( I_1 \subset g(I) \) existe un compacto \( Q \subset I \) tal que \( g(Q) = I_1 \).

Lema 3. Sea \( J \) un intervalo, \( f: J \to J \) una función continua, y sea

\[
\{ I_n \ ; \ n = 0, 1, ... \}
\]

una sucesión de intervalos compactos con

\[
I_n \subset J \mid I_{n+1} \subset f(I_n) \ \forall n \geq 0
\]

Entonces existe una sucesión de intervalos compactos

\[
\{ Q_n \ ; \ n = 0, 1, ... \}
\]

tales que, para todo \( n \geq 0 \), se cumple que:

1. \( Q_{n+1} \subset Q_n \subset I_0 \)
2. \( f^n(Q_n) = I_n \)
3. \( f^n \left( \bigcap_{n=0}^{\infty} Q_n \right) \subset I_n \)

Lema 4. Sea \( J \subset \mathbb{R} \) y \( g: J \to \mathbb{R} \) una función continua, e \( I \subset J \) un intervalo compacto tal que \( I \subset g(I) \). Entonces \( \exists p \in I \) tal que \( g(p) = p \).

Estos lemas nos permitirán demostrar el teorema de Li y Yorke, tal y como se enunció más arriba.

Demostración del Teorema de Li-Yorke

Dado que nuestro sistema dinámico \( x_{n+1} = f(x_n) \) posee un punto periódico de período tres (omitimos el subíndice \( \mu \) por simplicidad), aplicaremos el lema 1 para encontrar un punto \( a \in X \) tal que \( d = a < b < c \) o bien \( c < b < a = d \). siendo \( b = f(a), c = f(b) = f^2(a) \) y \( d = f(c) = f^3(b) = f^3(a) = a \). Si se cumple la primera serie de desigualdades, esto es, \( d = a < b < c \) y si
$K = [a, b], L = [b, c]$, para cualquier $k \geq 1$ demostraremos que existe un punto periódico de periodo $k$.

Definamos para ello una sucesión

$$
\{I_n : n = 0, 1, \ldots\}
$$
de intervalos compactos como sigue. Si $k > 1$, tendremos:

$$I_n = \begin{cases} 
L & \text{si } 0 \leq n \leq k - 2 \\
K & \text{si } n = k - 1 \\
I_{n-k} & \text{si } n \geq k
\end{cases}
$$
y si $k = 1$, $I_n = L$ para todo $n \geq 0$.

Es decir, tendremos una colección de intervalos de la forma siguiente:

$$
\{I_n\}_{n=0}^{\infty} = \{L, L, L, \ldots\} \text{ si } k = 1 \\
\{I_n\}_{n=0}^{\infty} = \{L, L, \ldots, L, K, L, L, \ldots, L, K, \ldots\} \text{ si } k > 1
$$

Teniendo en cuenta que se dan las siguientes inclusiones:

$$f(L) = f([b, c]) \supseteq [f(c), f(b)] = [a, c] = L \cup K$$

$$f(K) = f([a, b]) \supseteq [f(c), f(b)] = [b, c] = L$$

tendremos que $\{I_n\}_{n=0}^{\infty}$ verifica las hipótesis del lema 3 y por lo tanto podemos hallar una sucesión $\{Q_n\}_{n=0}^{\infty}$ de intervalos compactos tal que

$$
\ldots \subseteq Q_{n+1} \subseteq Q_n \subseteq \ldots \subseteq Q_0 \subseteq I_0 = L
$$

con:

$$f^n(Q_n) = I_n$$

$$f^n \left( \bigcap_{n=0}^{\infty} Q_n \right) \subseteq I_n$$

Entonces tendremos que:

$$Q_k \subseteq Q_0 = I_0 = L$$

$$f^k(Q_k) = I_k = L$$

Luego

$$Q_k \subseteq f^k(Q_k)$$

y, aplicando ahora el lema 4 a la función $f^k : Q_k \to \mathbb{R}$ existirá un punto $p_k \in Q_k$ tal que

$$f^k(p_k) = p_k$$. Si el periodo de $p_k$ fuera menor que $k$, y dado que se tiene

$$f^n(Q_n) = I_n = \begin{cases} 
L & \text{si } n \leq k - 2 \\
K & \text{si } n = k - 1 \text{ y } L \cup K = \{b\} \\
L & \text{si } n \geq k
\end{cases}
$$

se debería cumplir entonces que $f^{k-1}(p_k) = b$ y por lo tanto que

$$f^{k+1}(p_k) = f^2(b) = d \notin L$$
pero tenemos que:

$$f^{k+1}(p_k) = f(p_k) \in f(Q_0) = I_0 = L$$

lo que introduce una contradicción. Es decir: $p_k$ es un punto periódico de periodo exactamente $k$. Dado que esto es válido para cualquier $k \geq 1$, el teorema está demostrado.

El teorema de Li-Yorke introduce la posibilidad de hallar órbitas de enorme complejidad en sistemas discretos unidimensionales en presencia de órbitas en periodo 3 aunque este resultado no es general: no es aplicable a sistemas de dimensión superior ni la existencia de cualquier periodicidad tiene iguales consecuencias. Este teorema nos remite en realidad las consecuencias del teorema de Sharkovsky (1964), cuyo enunciado es

**Teorema (Sharkovsky, 1964)**

Sea $\Omega$ el conjunto ordenado

$$\Omega = \{3 < 5 < 7 < \ldots < 2 \times 3 < 2 \times 5 < 2 \times 7 < \ldots \}$$

$$\ldots < 2^2 \times 3 < 2^2 \times 5 < 2^2 \times 7 < \ldots < 2 \times 4 < 2 \times 1 \}$$

Sea

$$f_\mu : U \rightarrow U$$

con $U \equiv [0, 1]$, una aplicación continua tal que $f(0) = f(1) = 0$ y que posee un único punto crítico. Si $m < n$, y $f_\mu$ posee un punto periódico de periodo $m$, entonces $f_\mu$ también posee un punto periódico de periodo $n$.

Así, para la aplicación logística, existe una órbita de periodo seis en $\mu = 3.627\ldots$ (el comienzo de una ventana periódica). De acuerdo con el teorema, existirán soluciones periódicas (inestables) de periodos $2 \times 5$, $2 \times 7$, etc., siguiendo la ordenación dada por $\Omega$ (notemos que el primer elemento de $\Omega$ es, precisamente, el valor 3). (Para una revisión introductoria de este teorema véase Kaplan, 1987.)

### 5.19 Caos en sistemas dinámicos continuos

En las secciones previas hemos explorado con cierto detalle la aparición de caos en sistemas discretos. La naturaleza, en general, se nos presenta en forma de variables (aproximadamente) continuas, de forma que la descripción natural será la de las ecuaciones diferenciales.

$$\frac{dx}{dt} = F_\mu(x(t))$$

(esta expresión en realidad sólo incluye sistemas autónomos, sin dependencia explícita con el tiempo, de hecho, el tipo de sistemas que trataremos). Notemos sin embargo que ello no impide que empleemos sistemas discretos como modelos adecuados de cierto tipo de fenómenos, como puede ser la dinámica de ciertas poblaciones de insectos (May, 1976) que, de forma natural, poseen generaciones separadas y una escala real de tiempo que podemos considerar discreta (un año, por ejemplo). Además, aunque en esta sección introduzciremos algunos modelos continuos que parecen alejados de los resultados obtenidos en los modelos basados en ecuaciones cuadráticas, veremos cómo recuperamos los resultados anteriores empleando secciones de Poincaré. El modelo de Lorenz, presentado en la introducción, es un ejemplo de sistema dinámico tridimensional que exhibe caos determinista. Analizaremos este modelo algo más adelante. Antes, detengámonos en un modelo distinto conocido como modelo de Rössler.
Figura 5.29: Bifurcaciones sucesivas del modelo de Rössler, para distintos valores del parámetro $\mu$. 
Este modelo se basa en un conjunto de tres ecuaciones diferenciales ordinarias:

\[
\begin{align*}
\frac{dx}{dt} &= -(y + z) \\
\frac{dy}{dt} &= x + \frac{1}{3}y \\
\frac{dz}{dt} &= 1 + (x - \mu)z
\end{align*}
\]

y es, de hecho, un *modelo del modelo* de Lorenz (Rössler, 1976), un *metamodelo*. Este modelo permite simplificar la estructura del atractor de Lorenz (que, recordemos, poseía dos partes simétricas) en forma similar a si estudiáramos sólo una parte del atractor de Lorenz.

El primer resultado interesante obtenido al explorar el parámetro \( \mu \) desde cero a valores crecientes, es la aparición de un escenario de bifurcación que, pese a las diferencias obvias con el modelo ya estudiado, identificaremos rápidamente como un escenario de Feigenbaum. Para valores crecientes del parámetro de bifurcación, el modelo de Rössler atraviesa distintos valores críticos en los que el sistema duplica su periodicidad. En la figura 5.29 vemos un ejemplo de estas bifurcaciones. Del mismo modo que ocurriera en los sistemas discretos anteriormente estudiados, las bifurcaciones aparecen cada vez con mayor rapidez, y finalmente alcanzamos un régimen dinámico aperiódico: el modelo muestra atractores extraños para valores suficientemente grandes del parámetro analizado.

¿Cómo construir este modelo a partir de argumentos geométricos relevantes? Supongamos que partimos del sistema bidimensional lineal definido por las ecuaciones:

\[
\begin{align*}
\frac{dx}{dt} &= -y - c \\
\frac{dy}{dt} &= x + ay
\end{align*}
\]

donde \( a, c \) son constantes y \( a < 2 \). Este sistema daría un comportamiento ya conocido: una espiral inestable en el plano \((x, y)\). Supongamos que el parámetro \( c \) fuera en realidad una variable dinámica, esto es, \( c = c(t) \). Puede demostrarse que la solución para el sistema autónomo inicial (con \( c \) constante) es de la forma:

\[
\begin{align*}
x &= ac + R[A^t e^{it}] e^{at/2} \\
x &= -c - R[(\frac{a}{2} + i\omega)A^t e^{it}]
\end{align*}
\]

(siendo \( \omega = \sqrt{1 - (a/2)^2} \)). Si \( c = c(t) \), el sistema dejó de ser autónomo y podemos visualizar el efecto de \( c(t) \) introduciendo un nuevo eje de coordenadas. En la figura 5.30 representamos esta situación. Si \( c \) crece, el centro de la espiral tiende a desplazarse hacia abajo en el plano \((x, y)\). Si \( c \) va variando y vuelve más adelante a tomar el mismo valor que al principio, el centro de la espiral también es el mismo y el movimiento es “reinyectado” sobre el plano \((x, y)\). Pero aquí está el punto clave: el lugar de reinyección dependerá *en general* de la historia de \( c(t) \) y por lo tanto el movimiento sobre el plano puede ser muy distinto del anterior. La idea de Rössler es precisamente convertir \( c(t) \) en una variable dinámica \( z(t) \), recuperando así el carácter autónomo de la dinámica. Llegamos así al modelo general

\[
\begin{align*}
\frac{dx}{dt} &= -y - z \\
\frac{dy}{dt} &= x + ay
\end{align*}
\]
Un conjunto de parámetros habitual es $\alpha = b = 0.2$, y se suele tomar como valor típico para generar caos $\mu = 5.7$. Vemos que si $x < \mu$, entonces $z(t)$ se aproxima a $b/(\mu - x)$ (tanto más rápidamente cuanto mayor sea $\mu - x$) mientras que, para $x > \mu$, $z(t)$ crece exponencialmente. El único término no-lineal, que aparece en la tercera ecuación (esto es, $xz$) da lugar a un estiramiento $+ \text{reinyección de } z(t)$ relativo al movimiento “lineal” de las variables del plano $(x, y)$.

De forma similar a los sistemas anteriores, el atractor extraño presenta una estructura geométrica altamente ordenada que, nuevamente, exhibe fractalidad. Podemos estudiar las propiedades del atractor en varias formas. Una primera imagen del proceso dinámico que genera caos en este sistema puede obtenerse a partir del estudio de su topología. Para generar sensibilidad a las condiciones iniciales, el atractor debe poseer una zona de “estiramiento” de puntos cercanos. A la vez, este objeto debe estar confinado en una zona del espacio de fases: deberá por lo tanto existir un proceso adicional de “plegamiento”. Ambas propiedades pueden de hecho observarse si miramos atentamente uno de estos atractores extraños. En la figura 5.30 mostramos la imagen del atractor obtenida a partir del sistema dinámico antes definido junto con una imagen aproximada de su estructura topológica.

La topología del atractor es reveladora. Observamos una zona $(S)$ de la “superficie” (estrictamente, no lo es) en la que puntos cercanos pueden separarse, con lo que tendremos sensibilidad a las condiciones iniciales. Volviendo al modelo planteado antes, la espiral inestable asociada al comportamiento sobre el plano $(x, y)$ introduce la separación. Esta separación coexiste con otra zona del atractor $(F)$ en la que las órbitas que previamente se han separado se pliegan para confinar el flujo en una región finita (la reinyección de la que hablábamos antes). El resultado es, de hecho, un proceso de estiramiento-plegamiento que se da una y otra vez. Podemos visualizar el proceso de dos formas: por una parte, podemos ver de qué forma un conjunto de condiciones iniciales sobre la región $S$ se desplaza alrededor del atractor para regresar cerca de su estado anterior. Este proceso se indica en la figura 5.31 (a-c) en tres etapas, que indican cómo un conjunto de condiciones iniciales distribuidas sobre un segmento de línea son estiradas y plegadas. Al volver hacia la zona inicial, vemos que nuestra línea se ha estirado y plegado dando lugar a una forma que nos es familiar: una herradura. Al dar otra vuelta, esta herradura se pliega otra vez sobre si
Figura 5.31: Generación de autosimilaridad en el atractor de Rössler. (a) Partiendo de un segmento de condiciones iniciales $\Theta$, empleamos el modelo para generar la dinámica y vemos que se forma una herradura como consecuencia del proceso de estiramiento-plegado característico de los sistemas caóticos. (b,c) Si repetimos el proceso, obtenemos herraduras plegadas sobre sí mismas.

misma generando una herradura doblada dos veces. Si el proceso continúa, lo que ocurrirá es que el atractor presentará propiedades fractales, tal y como esperaríamos.

Existe una segunda forma de explorar este sistema, basada en el empleo de secciones de Poincaré (véase también el capítulo 4). Supongamos que cortamos el atractor por medio de una sección de Poincaré $\Sigma$ tal y como indica la figura 5.32 (a). En esta gráfica ya vemos que la sección de Poincaré intersectará la trayectoria caótica a lo largo de un conjunto cuasilineal de puntos. Supongamos que este conjunto viene dado, considerando sólo las coordenadas según $x$, por:

$$X(\Sigma) = \{x_1, x_2, x_3, \ldots, x_n, \ldots\}$$

Ya sabemos que podemos obtener una imagen simplificada del fenómeno recurriendo al estudio de la aplicación $x_{n+1} = f_\mu(x_n)$, si es conocida, y que ésta conserva la información relevante acerca del sistema. Si representamos el histograma de retornos $(x_n, x_{n+1})$ para esta serie de puntos, el resultado, que se muestra en la figura 5.32 (b), no es otro que el que esperaríamos de un sistema dinámico discreto unidimensional, descrito por una aplicación cuadrática.

Sorprendentemente, nos acabamos de encontrar con que la dinámica subyacente a un sistema continuo tridimensional tiene profundas conexiones con un sistema mucho más simple: la familia de aplicaciones uniparamétricas que analizábamos en nuestro estudio del escenario de Feigenbaum. De hecho, podemos obtener una información equivalente representando una serie distinta, obtenida a partir de la secuencia de máximos sucesivos $x(M) = \{M_1, M_2, \ldots, M_n, \ldots\}$, empleando para ello cualquiera de las variables. Un diagrama $(M_n, M_{n+1})$ nos dará una parábola similar a la anterior. Si en lugar de representar esta gráfica exploramos los distintos valores de $\mu$ y, para cada uno, representamos los máximos sucesivos (obtenidos, por ejemplo, de $x(t)$), lo que obtendremos (fig. 5.33) no es sino un diagrama de Feigenbaum en el que apreciamos sin dificultad las sucesivas bifurcaciones con duplicación de periodo. Este resultado permite aplicar los resultados anteriores a sistemas reales muy diversos, en los que podemos detectar el mismo tipo de transiciones. En reacciones químicas, por ejemplo, podemos obtener escenarios de bifurcación de distintos tipos.
Figura 5.32: (a) Sección de Poincaré sobre el atractor de Rössler. (b) Diagrama $(x_n, x_{n+1})$ obtenido a partir de los valores que intersectan $\Sigma$.

Figura 5.33: Escenario de Feigenbaum obtenido para el atractor de Rössler empleando la serie resultante de considerar los máximos sucesivos en el atractor correspondiente.
Para la conocida reacción de Belousov-Zhabotinsky (Scott, 1994) es posible demostrar que la dinámica de los máximos obedece una aplicación unidimensional. Podemos caracterizar estos atractores empleando distintos tipos de medidas cuantitativas que se discutirán, en un contexto general, en siguiente capítulo.

5.20 Caos en sistemas no-autónomos

En todos nuestros ejemplos (y en el desarrollo general) hemos limitado nuestra atención a sistemas dinámicos autónomos, sin una dependencia temporal explícita en las funciones no lineales empleadas. Pero, como podemos imaginar, la introducción del tiempo podría relajar la restricción de una dimensionalidad mínima \( d = 3 \) para obtener caos. El tiempo actúa como una variable en un sistema bidimensional (en el que por tanto pasamos a tener tres variables, y con ello la posibilidad de caos). Un ejemplo es el modelo del Brusselator forzado periódicamente,

\[
\frac{dx}{dt} = a[1 + \cos(w_f t)] - bx + x^2 y - x
\]
\[
\frac{dy}{dt} = bx - x^2 y
\]

(Scott, 1984), el cual exhibe escenarios de bifurcación hacia el caos de una gran riqueza. Así, para el sistema no perturbado \( (a = 0) \) con \( a = 0.4, b = 1.2 \), el Brusselator exhibe un ciclo límite con una frecuencia natural \( w_0 = 0.3776 \) (periodo \( T = 16.64 \)). Para distintos valores de \( a \) y de la frecuencia \( w_f \), o más exactamente del cociente \( w_f/w_0 \), el sistema experimenta bifurcaciones de distintos tipos.

Otro comportamiento caótico que se presenta en sistemas no autónomos está propiciado por la introducción de un tiempo de retardo \( \tau \), por ejemplo en

\[
\frac{dx}{dt} = f(x; t, \tau) - \gamma z(t) = \frac{\lambda \theta^n z(t - \tau)}{\theta^n + z(t - \tau)} - \gamma z(t)
\]

que ha sido empleado como modelo de la dinámica de poblaciones de glóbulos rojos \( (z(t)) \) en sangre (Mackey y Glass, 1977). Esta ecuación es biológicamente interpretable en términos simples: las células sanguíneas poseen una vida media y se produce una renovación constante. Dado que se requieren unos cuatro días para que las nuevas células maduren, existirá un retardo en la dinámica que controla la producción de glóbulos maduros.

Los sistemas fisiológicos de control deben mantener las cantidades de células en niveles constantes (de una forma similar, aunque no igual, a lo que ocurre con un termostato). Para tiempos de retardo del orden de \( \tau = 2 \) y parámetros \( \gamma = 1, \theta = 1, \lambda = 2 \), el sistema exhibe bifurcaciones sucesivas para valores crecientes de \( n \), parámetro que da una medida de la no linealidad de la función \( f(z) \) del sistema. Por ejemplo, el caso

\[
f(x; t, \tau) = \frac{\lambda \theta^n z(t - \tau)}{\theta^n + z(t - \tau)}
\]

genera atractores extraños y periódicos, como se indica en la figura 5.34. Estas dinámicas en las que aparecen excitaciones complejas han sido de hecho observadas en algunas situaciones patológicas, como ciertos tipos de leucemia (Mackey y Glass, 1977).

En otros casos, los retardos temporales afectan a la dinámica de poblaciones de insectos (Gurney et al., 1980), que exhiben comportamientos complejos, y lo mismo ocurre en algunas poblaciones de mamíferos (May, 1983).
Figura 5.34: Atractores correspondientes al sistema con retardo 5.20.1, para (a) \( n = 7 \) y (b) \( n = 10 \).

Un ejemplo muy distinto procede del estudio de colectivos de elementos que interactúan entre sí y se localizan en un ambiente externo del que poseen cierta información. Estos elementos, conocidos como *agentes predictivos*, evalúan el estado del sistema y de los recursos y llevan a cabo una decisión (o predicción) basada en el estado previo del sistema (el pasado sirve para predecir el futuro). Puede demostrarse que, si \( f(t) \) es el número de agentes que emplean un recurso dado, la dinámica del número promedio de agentes obedece una ecuación con retardo (Keptart et al., 1990)

\[
\frac{d < f(t) >}{dt} = \alpha [\rho (f^*(t) - < f(t) >)]
\]

siendo \( \rho (f^*(t)) \) la predicción llevada a cabo por el sistema partiendo del valor de \( < f(t - \tau) > \). Estos modelos incorporan de manera natural retardos temporales en la capacidad de predicción (fiable) y dan lugar a comportamientos muy complejos, de importantes implicaciones para sistemas económicos (Huberman, 1988). Típicamente, estos sistemas exhiben un comportamiento complejo que, en algunos casos, da lugar a periodos de estabilidad seguidos de repentinos cambios, eventualmente catastróficos (Glance y Huberman, 1993).

### 5.21 Caos homoclínico

El atractor de Rössler y muchos otros sistemas dinámicos como las ecuaciones de Lotka-Volterra tridimensionales (se verá más adelante) pertenecen a un conjunto amplio de sistemas que presentan ciertas propiedades comunes entre las que está implicada la existencia de órbitas homoclínicas.

Recordemos que una órbita \( \Gamma_0 \) que parte de cierto punto \( x \in \mathbb{R}^n \) se denomina homoclínica respecto al punto fijo \( x_0 \) del sistema dinámico si \( \phi^t_i (x) = x_0 \), para \( t \to \pm \infty \).

La órbita homoclínica \( \Gamma_0 \) pertenece a la intersección entre las variedades estable e inestable:

\[
\Gamma_0 \subseteq [W^s(x_0) \cap W^u(x_0)]
\]

Mostramos en la figura 5.35 dos ejemplos para sistemas de dimensión dos y tres.
Figura 5.35: Orbitas homoclínicas en dos y tres dimensiones.

Las órbitas homoclínicas de puntos de equilibrio hiperbólicos son de gran interés en nuestro estudio debido a que presentan inestabilidad estructural. En particular, puede demostrarse (Kuznetsov, 1995) el siguiente lema:

**Lema.** Una órbita homoclínica $\Gamma_0$ perteneciente a un punto de equilibrio hiperbólico $x_0$ del sistema dinámico

$$\frac{dx}{dt} = f_\mu(x), \quad x \in \mathbb{R}^n$$

es estructuralmente inestable.

Este lema implica que podemos perturbar un sistema que posea una órbita homoclínica $\Gamma_0$ respecto de $x_0$ de tal forma que el retrato de fases en las proximidades de $\Gamma_0 \cup x_0$ es topológicamente no-equivalente al original. De hecho (Kuznetsov, 1995) la órbita homoclínica simplemente desaparece genéricamente para perturbaciones $C^1$ del sistema original, lo que representa una bifurcación (capítulo 4).

El siguiente teorema, debido a Shilnikov, hace referencia a las implicaciones de la existencia de una órbita homoclínica en sistemas dinámicos 3-dimensionales del tipo silla-foco (figura 5.35).

**Teorema de Shilnikov**

Sea el sistema dinámico

$$\frac{dx}{dt} = f_\mu(x), \quad x \in \mathbb{R}^3$$

que supondremos puede escribirse en la forma

$$\frac{dx}{dt} = \rho x - wy + P(x, y, z)$$

$$\frac{dy}{dt} = \rho y + wz + Q(x, y, z)$$

...
\[ \frac{dz}{dt} = \lambda z + R(x, y, z) \]

donde \( P, Q \) y \( R \) son funciones analíticas tales que \( P(0) = Q(0) = R(0) = 0 \) y \( P'(0) = Q'(0) = R'(0) = 0 \) en el origen \((0, 0, 0)\). El origen es un punto fijo del tipo silla-foco, es decir, los valores propios son de la forma

\[ \lambda_+ = \rho \pm wi, \quad \lambda_3 = \lambda \in \mathbb{R}^- \]

Supongamos que el origen posee una órbita homocinética \( \Gamma_0 \). Entonces

1. Si \( \lambda > -\rho > 0 \) (o si \( -\lambda > \rho > 0 \)), cualquier entorno de \( \Gamma_0 \) contiene un conjunto numerable de soluciones periódicas inestables de tipo silla.

2. En un entorno de \( \Gamma_0 \) existe un subconjunto de trayectorias que muestran comportamiento aperiódico, en el sentido de que existe una correspondencia biyectiva con el operador \( \sigma \) (“shift”) con un número infinito de símbolos.

Las consecuencias del teorema son de gran importancia: se desprende de sus conclusiones que existirá (de forma muy general) una relación entre trayectorias homocinéticas y caos determinista. Esperaremos por lo tanto, bajo las condiciones anteriores, observar atractores extraños (del tipo del atractor de Rössler).

Existen distintos enunciados alternativos del teorema (véase, por ejemplo, Guckenheimer y Holmes, 1983). Uno de ellos hace referencia al comportamiento de las trayectorias en un entorno próximo al punto \( x_0 \). Si tomamos una superficie cilíndrica \( \Sigma_0 \) alrededor de \( x_0 \) (figura 5.36), que fuese cubierta, arriba y abajo, por círculos \( \Sigma_1 \), y si indicamos por \( \Sigma^*_0 \) la parte inferior, podemos analizar el comportamiento de la aplicación de Shilnikov,

\[ \Pi : V \rightarrow \Sigma_0 \cup \Sigma^*_0 \]

siendo \( V \subset \Sigma_0 \) una pequeña superficie sobre el cilindro superior. Podemos considerar \( V \subset \Sigma_0 \) como un conjunto de condiciones iniciales. Indicamos dos aplicaciones definibles,

\[ \Psi_0 : \Sigma_0 \rightarrow \Sigma_1 \]

que caracteriza el comportamiento local de \( x_0 \), mientras que la segunda

\[ \Psi_H : \Sigma_1 \rightarrow \Sigma_0 \cup \Sigma^*_0 \]

tiene en cuenta el comportamiento local de los puntos próximos a la órbita homocinética \( \Gamma_0 \) que no están en las proximidades de \( x_0 \).

El teorema afirma (en estos términos) que existirá un conjunto numerable de herraduras de Smale definidas por \( \Psi_H \) sobre \( \Sigma_0 \cup \Sigma^*_0 \) (indicamos una de estas por \( \Psi_H(V) \)).

De forma similar a lo analizado anteriormente, el empleo de una sección de Poincaré nos permite visualizar el comportamiento en términos más simples, en este caso en términos del sistema de Smale que, como sabemos, es caótico (Guckenheimer y Holmes, 1983; Wiggins, 1990; Kuznetsov, 1995).

Un problema altamente no trivial en la aplicación de este teorema reside precisamente en probar que \( \Gamma_0 \) existe. Aún así, para sistemas que exhiben caos, podemos explorar la posible aparición de \( \Gamma_0 \) a partir de argumentos cualitativos simples (Arneodo et al., 1980). Consideremos el sistema de Lotka-Volterra definido en general por el sistema \( n \)-dimensional

\[ \frac{dN_i}{dt} = N_i \left[ \gamma_i - \sum_{j=1}^{n} \mu_{ij} N_j \right], \quad i = 1, \ldots, n \]
Figura 5.36: Orbitas homoclínicas y herraduras de Smale en el caos tipo Shilnikov.

que pertenece al conjunto genérico de la forma

$$\frac{dN_i}{dt} = N_i f^{(i)}_\mu ([N_j]), \quad i, j = 1, \ldots, n$$

para el caso particular $f^{(i)}_\mu = \gamma_i - \sum_{j} \mu_{ij} N_j$. Estos sistemas presentan la particularidad de que todos los hiperplanos

$$\{N_j\} = 0, \quad j \in J, \quad J \subseteq \{1, \ldots, n\}$$

son globalmente invariantes. En $n = 3$, y eligiendo el punto $N_1 = N_2 = N_3 = 1$ como punto central de equilibrio (sin pérdida de generalidad) las ecuaciones pueden escribirse en la forma

$$\frac{dN_i}{dt} = N_i \sum_{j=1}^{3} \alpha_{ij} (1 - N_j)$$

y podemos analizar su comportamiento para distintos valores de la matriz $(\alpha_{ij})$. Siguiendo a Arneodo tomemos la matriz

$$\begin{pmatrix}
(\alpha_{ij})_\mu = \\
0.5 & 0.5 & 0.1 \\
-0.5 & -0.1 & 0.1 \\
\mu & 0.1 & 0.1
\end{pmatrix}$$

en la que sólo $\alpha_{13} \equiv \mu$ es variable. Para valores crecientes de $\mu$, este sistema experimenta sucesivas duplicaciones de periodo hasta alcanzar el régimen caótico. En la figura 5.37 (a,b) mostramos dos atractores para distintos valores de $\mu$.

Arneodo et al. obtuvieron un argumento heurístico de validez general (para este tipo de sistemas) que permite conjeturar la existencia de una órbita homoclínica $\Gamma_0$. Supongamos que consideramos una familia $\mu$-paramétrica de ecuaciones diferenciales tales que

1. Para $\mu > \mu_0$, existe un punto fijo $x_0$ del tipo silla-foco que satisface las condiciones del teorema de Shilnikov (es decir, $\lambda > -\rho > 0 \ o \ -\lambda > \rho > 0$).
Figura 5.37: Atractores para los valores de $\mu$ (a) $\mu = 1.43$ y (b) $\mu = 1.65$ del sistema de Arneodo. 
(c,d) Conexión entre el ciclo límite (que aumenta de tamaño con $\mu$) y el punto silla-foco inestable. 
(e) Formación de la órbita homoclínica $\Gamma_0$ y (f) interpretación de $\Gamma_0$ para el sistema de Lotka-Volterra.
Figura 5.38: Ventana de periodo tres. (a) Diagrama esquemático de la estructura de las bifurcaciones en la aplicación logística. (b) Ampliación de la ventana de periodo tres, cerca de la cual se da intermitencia.

2. Para $\mu = \mu_H > \mu_0$, el sistema experimenta una bifurcación de Hopf supercrítica en otro punto $x_1$, que genera una órbita periódica $\gamma$ estable para $\mu > \mu_H$ (figura 5.37 (c)).

3. Para $\mu > \mu_H$, la variedad estable del punto $x_0$ converge hacia $\gamma$ cuando ésta es estable.

4. Si $\mu$ crece, el tamaño de $\gamma$ crece con mayor rapidez que la distancia $||x_0 - x_1||$ (fig. 5.37 (d)).

Bajo estas condiciones podemos esperar que la variedad inestable de $x_0$ se acerque lo bastante a $\gamma$ como para dar lugar a la órbita homoclínica $\Gamma_0$ considerada por el teorema (fig. 5.37 (e)).

Este procedimiento heurístico ha dado buenos resultados en varios modelos generales de gran interés, como el modelo de Lotka-Volterra anterior, para el que representamos en la figura 5.37 (f) la órbita homoclínica asociada al sistema para $\mu = 1.708$.

El caos de "tipo Shilnikov", como se conoce también, ha sido detectado en multitud de sistemas físicos y biológicos, que van desde láseres hasta, como veremos, la dinámica cerebral (véase el capítulo 15).

5.22 Intermitencia

El escenario de Feigenbaum que hemos analizado es una ruta hacia el caos muy importante, aunque no la única posible. Existen dos escenarios distintos que juegan un papel muy importante en la aparición de caos en sistemas naturales tan distintos como los fluidos o el tejido cardíaco. Los analizaremos en esta sección y la siguiente.

El primero se conoce con el nombre de intermitencia. La denominación procede de la forma peculiar en la que irrumpen el caos a lo largo de la dinámica temporal del sistema: en la misma evolución temporal se alternan comportamientos regulares (básicamente periódicos) interrumpidos por ráfagas desordenadas más o menos cortas. En esta sección presentaremos un tipo de escenario intermitente (no el único, véase Schuster, 1989) conocido como intermitencia de tipo-I.
La aparición de intermitencia tiene lugar, por ejemplo, en sistemas dinámicos discretos que presentan escenario de Feigenbaum. En el caso de la aplicación logística, esto ocurre en las proximidades de la ventana de periodo tres que se observa en el interior del dominio caótico (figura 5.38) por encima de \( \mu_c = 1 + \sqrt{8} \).

La dinámica de la aplicación logística cerca de \( \mu_c \), o más explícitamente, para valores de \( \mu \) tales que \( 0 < \mu_c - \mu \ll 1 \), es la característica del régimen intermitente. En la figura 5.39 (a) vemos un ejemplo de esta dinámica para \( \mu_c - \mu = 0.002 \). Esta dinámica aparece como una combinación de oscilaciones cercanas a una órbita de periodo tres, interrumpidas por ráfagas de tipo no-periodico. La explicación para este comportamiento puede obtenerse analizando la figura 5.39 (b), en la que representamos la tercera iteración de \( f_\mu(x) \) esto es, \( f_\mu^{(3)}(x) \), frente a \( x \), para un valor de \( \mu \) algo por debajo de \( \mu_c = 1 + \sqrt{8} \). Si iteramos esta función (recordemos que ahora los puntos fijos representan elementos de la órbita de periodo tres) veremos que, en las proximidades de la tangencia indicada, la dinámica del sistema pasará muy cerca de la órbita 3-periodica, que es inestable. Tras abandonar esta región visitará otras zonas de \( f_\mu^{(3)}(x) \) dando lugar a nuevas ráfagas desordenadas para volver a la misma región.

Existen diversas propiedades del régimen de intermitencia que permiten caracterizar el comportamiento de estos sistemas en forma cuantitativa. Una de éstas es la longitud promedio \( \langle L \rangle \) de la denominada "región laminar", en función del parámetro \( \epsilon = \mu - \mu_c \). El término "laminar" designa a las regiones ordenadas en la dinámica del modelo considerado, y procede de la dinámica de fluidos, en la que el régimen laminar corresponde al fluido no turbulento que fluye de forma suave y continua. El fenómeno de la intermitencia es de hecho bien conocido en dinámica de fluidos.

Si desarrollamos \( f_\mu^{(3)}(x) \) en las proximidades de \( x_c \) y \( \mu_c \), definidos por:

\[
\frac{\partial}{\partial x} f_\mu^{(3)}(x_c) = 1
\]

\[
f_\mu^{(3)}(x_c) = x_c
\]

obtenemos:

\[
f_\mu^{(3)}(x_c) = f_\mu^{(3)}(x_c + (x - x_c)) = x_c + (x - x_c) + \alpha_c (x - x_c)^2 + \beta_c (x - x_c)^3
\]

siendo \( \alpha_c \) y \( \beta_c \) dos constantes definidas por:

\[
\alpha_c = \frac{1}{2} \left( \frac{\partial^2}{\partial x^2} f_\mu^{(3)} \right)_{\mu_c, x_c}
\]

\[
\beta_c = \left( \frac{\partial}{\partial \mu} f_\mu^{(3)} \right)_{\mu_c, x_c}
\]

Definiendo \( y \equiv (x - x_c)/\beta_c \) y \( \sigma \equiv \alpha_c \beta_c > 0 \), la aplicación \( x_{n+1} = f_\mu^{(3)}(x_n) \) nos da, en las proximidades de \( x_c \), una expresión:

\[
y_{n+1} = y_n + \sigma y_n^2 + \epsilon
\]

Y las regiones laminares se definirán ahora con un criterio intuitivamente claro: las iteraciones consecutivas del sistema deben cambiar muy poco, lo cual puede indicarse formalmente por la desigualdad

\[
|y_n| < \theta \ll 1
\]
Figura 5.39: (a) Intermitencia en la ecuación logística: para $\mu_c - \mu = 0.002$ observamos regiones de comportamiento laminar interrumpido por ráfagas cortas de comportamiento irregular. Para visualizar las regiones laminares, sólo indicamos los valores de la aplicación cada tres iteraciones. (b) Longitud promedio de la región laminar (aquí $\theta = 0.01$).

esto es, la distancia a $x_c$ debe ser inferior a cierto umbral $\theta$. Si aproximamos la ecuación en diferencias para $y_{n+1}$ por la ecuación diferencial correspondiente, tendremos:

$$\frac{dy}{dt} = ay^2 + \epsilon$$

donde $l$ hace referencia a las iteraciones sobre la región laminar. Integlando entre los límites de la región ($L_o$ y $L_i$), obtenemos:

$$l(y_0, y_i) = \frac{1}{\sqrt{\alpha}} \left[ \arctan \left( \frac{y_0}{\sqrt{\epsilon/a}} \right) - \arctan \left( \frac{y_i}{\sqrt{\epsilon/a}} \right) \right]$$

A continuación, encontraremos la longitud promedio $<l>$ suponiendo que existe $P(y_i)$, la probabilidad de que, después de una ráfaga caótica, la órbita sea reinyectada cerca del punto $x_c$, de forma que se verifiquen las hipótesis anteriores. Esta probabilidad es simétrica respecto del punto $x_c$: $P(y_i) = P(-y_i)$, y tendremos:

$$<l> = \int_{-\theta}^{\theta} P(y_i)l(y_i, y_i)dy_i = \frac{1}{\sqrt{\alpha}} \left[ \arctan \left( \frac{\theta}{\sqrt{\epsilon/a}} \right) \right]$$

y, para

$$\frac{\theta}{\sqrt{\epsilon/a}} \ll 1$$

obtenemos una expresión asíntotica para la longitud promedio de la región laminar:

$$<l> \propto \epsilon^{-1/2}$$

La ley de escala anterior es muy general en estos sistemas. Se ha propuesto además que el mecanismo de intermitencia podría proporcionar un origen para el "ruído $1/f$" (Schuster, 1989) distinto del de la criticalidad autoorganizada, descrito en el capítulo 8.
5.23 Escenario de Ruelle-Takens-Newhouse

Un tercer escenario de bifurcación que desemboca en la generación de caos determinista tiene relación con la aparición de comportamientos cuasiperiódicos en sistemas dinámicos no-lineales. La cuasiperiodicidad fue estudiada en el capítulo 4, junto con las propiedades de los sistemas dinámicos que exhiben ciclos límite.

Este escenario fue introducido por primera vez en 1978 por Ruelle, Takens y Newhouse, y también es conocido como escenario RTN. Estos autores analizaron un estudio llevado a cabo por el célebre físico ruso Lev Landau en 1944, concerniente al problema de la aparición de la turbulencia en fluidos (Landau, 1986).

La hipótesis de Landau consiste en suponer que el comportamiento turbulento es el resultado de la aparición de un número infinito de frecuencias independientes asociadas a los remolinos de distintas escalas que vemos experimentalmente, y que aumentan en número y complejidad a medida que nos adentramos en el régimen turbulento. A medida que la velocidad del fluido aumenta (en alguna forma) aparecen frecuencias de oscilación que van sumándose hasta alcanzar un espectro continuo (capítulo siguiente). Existen algunos experimentos clásicos que permiten observar estas transiciones hacia el régimen turbulento. Uno de ellos es el experimento de Couette-Taylor, en el que un fluido se halla confinado entre dos cilindros concéntricos (figura 5.40) con el interior móvil que gira a una frecuencia ajustable por el experimentador. Si el cilindro gira lentamente, nada ocurre. Existen varios parámetros que controlan la dinámica cualitativa de este sistema, como el cociente \( \eta = r_1/r_2 \) entre los radios del cilindro externo, \( r_2 \), y del cilindro interno, \( r_1 \), o el número de Reynolds, que es de hecho el valor explorado habitualmente. Este número adimensional se define como

\[
R = \Omega r_1 d/\nu
\]

siendo \( \Omega \) la velocidad angular del movimiento del cilindro interno, \( d = r_2 - r_1 \) y \( \nu = \rho / \mu \) es la viscosidad cinemática del fluido. Dado que los parámetros \( r_1 \), \( d \) y \( \nu \) están fijados, \( R \) refleja la velocidad de rotación del cilindro interno.

Para valores de \( R \) crecientes aparece, para cierto valor crítico \( R_c \), una ordenación espontánea del fluido en forma de estructuras macroscópicas regulares con dinámica periódica (ciclo límite)
Figura 5.41: Transiciones entre distintos comportamientos dinámicos para el experimento de Couette-Taylor.
que, para velocidades algo mayores, presenta una segunda componente de oscilación. Para valores aún mayores, podemos ver una transición súbita (al menos en apariencia) hacia un comportamiento desordenado: la turbulencia.

En la figura 3.41 vemos unos ejemplos de estas dinámicas, medidas experimentalmente empleando un registro de la velocidad local del fluido. El sistema presenta oscilaciones que van ganando complejidad a medida que el número de Reynolds, que presentamos en la forma $R/R_c$, aumenta. Estos resultados sugieren que la conjetura formulada por Landau sería parcialmente válida, aunque ello no es claro a partir de los experimentos, en especial porque la transición hacia el régimen turbulento parece más rápida de lo que esperaríamos a partir de la hipótesis de Landau.

Ruelle, Takens y Newhouse demostraron que, de hecho, el escenario mediante el que un fluido desemboca en el régimen turbulento puede ser considerablemente distinto. En particular, estos autores (que introdujeran por primera vez el término atractor extraño) demostraron que un comportamiento aperiódico puede obtenerse como resultado de un escenario de bifurcación en el que, tras la aparición de las primeras bifurcaciones en las que aparecen nuevas frecuencias, los atractores cuasiperiódicos se “rompen” y pueden dar lugar a un atractor extraño.

Este escenario ha sido investigado empleando sistemas dinámicos discretos bidimensionales, que permiten explorar en detalle las propiedades de universalidad de esta transición. En primer lugar, recordemos que consistía el comportamiento cuasiperiódico. En la figura 3.42 vemos un toro en $\mathbb{R}^3$ sobre el cual se mueve una órbita que indicamos por una curva. Las dos frecuencias presentes en el toro se indican mediante las variables angulares $\phi$ y $\theta$. En 3.42 (a) la curva se cierra sobre sí misma (es una órbita periódica). Al cociente entre frecuencias, $\Omega = f_1/f_2$ se lo conoce como número de rotación, y cuenta el número de rotaciones en la coordenada $\theta$ por cada rotación según $\phi$. Para el caso de la figura, tenemos $\Omega = 3$. En general, si $\Omega$ es racional, la órbita se cerrará sobre sí misma, mientras que para valores irracionales, no se cerrará nunca y tendremos cuasiperioididad. Como ocurre en otros casos (se ha mencionado en el capítulo 4) el empleo de una sección a través de la trayectoria puede simplificar el estudio de la dinámica resultante. Podemos llevar cabo esta sección tal y como se indica en la figura 3.42 (b). Cada punto de intersección define un ángulo $\theta_n$ y podemos analizar el comportamiento de las órbitas analizando sus intersecciones con $\Sigma$.

Empleando la sección de Poincaré podemos definir una aplicación bidimensional asociada al sistema dinámico en la forma:

$$\theta_{n+1} = g_{\mu}(\theta_n)$$

que, para un caso más general en el que el movimiento se desplace en la superficie de un toro deformado (con un radio característico variable), deberá completarse con una iteración para la distancia $r$.

Un sistema dinámico discreto especialmente bien conocido es la siguiente aplicación disipativa sobre el círculo:

$$\theta_{n+1} = \theta_n + \Omega - \frac{K}{2\pi} \sin(2\pi \theta_n) + b r_n \quad (mod \ 1)$$

$$r_{n+1} = b r_n - \frac{K}{2\pi} \sin(2\pi \theta_n)$$

Esta expresión se obtiene (Schuster, 1989) a partir del estudio de una partícula en movimiento circular (en dos dimensiones) golpeada periódicamente. La ecuación de este sistema es:

$$\frac{d^2 \phi}{dt^2} = - \frac{d \phi}{dt} + Kf(\phi) \sum_{n=0}^{\infty} \delta(t - nT)$$

(con $n \in \mathbb{Z}$), y donde $\phi$ indica la posición (angular) de la partícula, $\Gamma$ es la constante de amortiguamiento y el segundo término de la derecha introduce la fuerza, aplicada de forma periódica ($T$ es el periodo). Un análisis de este sistema dinámico (Schuster, 1989) permite reducir este sistema
a una aplicación bidimensional que, bajo un cambio de variables adecuado, nos lleva a nuestro sistema discreto.

El parámetro $K$ mide la importancia de la no-linealidad dada por el término $\sin(2\pi \theta_n)$. Este sistema muestra cuasiperiodicidad para algunos valores de $K$, en tanto que para valores mayores se produce una transición hacia el caos a través de la rotura del toro invariante. En la figura 5.42 (a,b) se muestra un ejemplo de esta transición, para $\Omega = 0.612$ y $b = 0.25$, donde podemos ver la aparición de cuasiperiodicidad seguida de caos. La deformación del toro es, típicamente, el anuncio de su rotura y de la consiguiente aparición de comportamientos irregulares. Si volviéramos al sistema continuo de partida podríamos visualizar este proceso tal y como indica la figura 5.42 (a-c) para un sistema dinámico hipotético. Presentamos tres situaciones diferentes, en las que el sistema exhibe, para cierto valor de un parámetro de bifurcación $\mu$, un atractor periódico (al que puede haber llegado, por ejemplo, a través de una bifurcación de Hopf). Si $\mu$ aumenta, puede ocurrir que aparezca una frecuencia adicional, con lo que las órbitas del sistema se desplazarán sobre la superficie de un toro, mostrando eventualmente cuasiperiodicidad. Finalmente, para valores crecientes del parámetro puede darse deformación del toro seguida de su rotura. Este escenario está formalmente descrito por la teoría de Ruelle, Takens y Newhouse.

Para sistemas altamente disipativos, donde $b \to 0$, la componente radial de la trayectoria desaparece en el sistema dinámico antes descrito y se reduce a la llamada aplicación circular:

$$\theta_{n+1} = f(\theta_n) \equiv \theta_n + \Omega - \frac{K}{2\pi} \sin(2\pi \theta_n) \quad (mod\ 1)$$

que describe la transición cuasiperiodicidad $\to$ caos en términos (únicamente) de la variable angular $\theta_n$. Esta aplicación permite estudiar en forma muy simple (pero general) la transición al atractor extraño por rotura del toro invariante. Señalaremos que (análogamente a lo que ocurrirá con la aplicación logística en relación con el escenario de Feigenbaum) la forma particular de la función $f(\theta)$ es poco relevante (tendremos otra vez universalidad). En particular, $f(\theta)$ presenta las siguientes propiedades:

- $f(\theta + 1) = 1 + f(\theta)$
- Si $|k| < 1$, $f(\theta)$ y su inversa existen y son derivables (esto es, $f(\theta)$ es un difeomorfismo).
- Para $k = 1$, la inversa se hace no diferenciable, y si $|k| > 1$, deja de ser única.

Estos comportamientos quedan ilustrados en la figura 5.43.

Para $k < 1$ tendremos órbitas no-caóticas. Cuando $f(\theta)$ se hace no invertible, aparece la posibilidad de hallar órbitas caóticas, como las que muestra la figura 5.44.

Una enorme cantidad de información acerca de la estructura de las transiciones que tienen lugar en este sistema dinámico (y que nos da una buena idea de su complejidad) se obtiene explorando numéricamente el valor del exponente de Lyapunov $\lambda_L(\Omega, k)$ para cada par de valores $\Omega$, $k$.

Para $|k| < 1$, observamos un tipo de estructura característica, en forma de triángulos deformados, que recibe el nombre de lenguas de Arnold (figura 5.45). En cada una de estas bandas, el número de rotación es constante y racional. Estas estructuras son densas en el espacio de parámetros y no se cortan, formando un conjunto de Cantor. Los límites superiores (donde los triángulos se ensanchan) están ligados a la transición que convierte a la función en no-invertible.

### 5.23.1 Cuasiperiodicidad y caos en tejido cardíaco

Existe un estudio experimental acerca de la aparición de atractores extraños en cultivos de células del corazón que demuestra claramente la aparición de caos a través del escenario RTN (Guevara et al., 1981; Glass et al., 1984; Holden, 1987). La existencia de caos determinista en el corazón humano
Figura 5.42: Transición cuasiperiodidad-caos. (a) El sistema converge a una orbita periódica. Aquí, la superficie representa el flujo de una línea continua de puntos iniciales bajo la acción de la dinámica. (b) Aparece una segunda frecuencia que lleva al sistema a moverse sobre la superficie de un toro. (c) El toro invariante $T^2$ se deforma, sufre rotura, y surge el caos. Observe que la sección del toro revela con claridad la deformación a que da lugar la transición al caos. En este sentido, las secciones de Poincaré permiten obtener información suficiente acerca de la naturaleza de dichas transiciones (véase texto).
Figura 5.43: Comportamiento de la aplicación $f(\theta)$ para distintos valores de $k$. Para $k > 1$, vemos que $f(\theta)$ se hace no invertible.

Figura 5.44: Orbea cuasiperiódica y orbea caótica en la aplicación circular. Los parámetros son los mismos que en la figura anterior.
será analizada en el siguiente capítulo. Los experimentos de Glass y sus colaboradores consisten en estimular eléctricamente de forma periódica grupos de células de tejido cardíaco. Estas células se unen de forma que “laten” espontáneamente, en forma coordinada. El estímulo de estas células se lleva a cabo mediante el empleo de un microelectrodo, a través del cual podemos suministrar pequeños pulsos de corriente. Por medio de estas manipulaciones se ha obtenido escenarios de bifurcación muy diversos, que incluyen intermitencia, duplicación de periódos y cuasiperiodicidad.

En la figura 5.46 vemos un ejemplo del potencial registrado (periódico) que posee periodicidad $T_0$, junto con una señal súbita que indica la introducción de un estímulo. 8 unidades de tiempo después del comienzo de un ciclo. Tras la perturbación, el grupo de células vuelve a su ciclo anterior. Sin embargo, la introducción de secuencias periódicas de pulsos permite generar nuevas dinámicas, en particular, caos a través de un escenario de RTN.

Para bajas intensidades del estímulo, se obtiene dinámica cuasiperiódica, como la que se indica en la figura 5.47 (Holden, 1987), en la que la introducción de estímulos se indica por flechas verticales. En (b) se indica un detalle de dicha dinámica. Si representamos en un diagrama $(\theta_n, \theta_{n+1})$ la aplicación de primer retorno, veremos una secuencia monótona de puntos, casi paralela a la bisectriz. El estímulo periódico da lugar a una transición a la cuasiperiodicidad.

Cuando aumenta la intensidad del estímulo, la aplicación $(\theta_n, \theta_{n+1})$ presenta un cambio evidente: se hace no-invertible (fig. 5.48 (b)) y aparece el caos.

Este tipo de comportamiento abunda en sistemas fisiológicos sometidos a estímulos periódicos (Holden, 1987) y muy posiblemente se dan en condiciones naturales. Veremos en el próximo capítulo que incluso el corazón humano muestra claros signos de caos determinista en su dinámica intrínseca.
Figura 5.46: Registro de la actividad de un grupo de células cardíacas (Guevara et al., 1981). El ciclo intrínseco es de duración $T_0$, y no se ve afectado por la perturbación.

Figura 5.47: (a) Dinámica cuasiperiódica. La introducción de estímulos se indica por una flecha vertical. (b) Detalle de la dinámica.
Figura 5.48: (a) Diagrama \((\theta_n, \theta_{n+1})\) obtenido experimentalmente para el tejido cardíaco sometido a estimulación periódica, que indica la presencia de periodicidad; (b) Dinámica caótica cuando la intensidad del estímulo aumenta. Vemos cómo la aplicación obtenida experimentalmente se hace no-invertible.

## 5.24 Apéndice A: Caos y principios variacionales

En esta sección aplicaremos el formalismo que fue introducido en el primer tema del libro: el método de la máxima entropía. Recordemos que la idea básica era encontrar la distribución de probabilidad asociada a ciertas restricciones genéricas sobre las propiedades estadísticas del sistema. Calculábamos el valor máximo de la entropía compatible con dichas ligaduras, y el método variacional nos proporcionaba la distribución más probable. En este caso, nos será útil, entre otras cosas, para calcular el valor del exponente de Lyapunov.

Bajo la hipótesis de ergodidad, podemos suponer que los momentos de orden \(n\) de un sistema dinámico \(x_{n+1} = f_\mu(x_n)\), definido sobre el intervalo unidad \([0, 1]\), están relacionados con su densidad de probabilidad por:

\[
<x^n_t> = \int_0^1 y^n \rho(y) dy
\]

Donde \(\rho(x) > 0\) para \(x \in [0, 1]\) y

\[
\int_0^1 \rho(y) dy = 1
\]

La evolución temporal de la densidad de probabilidad viene dada por

\[
\rho_{t+1}(x) = \int_0^1 \rho_t(y) \delta(x - f(y)) dy = 1
\]

Para \(t = 0, 1, ..., \) y \(\rho_0\) es la densidad inicial.

La densidad de probabilidad asintótica se sigue de la resolución de la ecuación integral de Perron-Frobenius, definida por:

\[
\rho(y) = \int_0^1 \rho(x) \delta(y - f(x)) dx
\]
que, como hemos visto, da resultados analíticos para algunos ejemplos de interés. Por otra parte, si
la densidad de probabilidad es conocida, podremos evaluar el exponente de Lyapunov $\lambda_L$ mediante

$$\lambda_L = \int_0^1 \rho(x) \ln \left| \frac{df}{dx} \right| dx$$

Podemos emplear el principio de máxima entropía para obtener $\rho(x)$ empleando como ligaduras $N$
momentos de la distribución. Siguiendo el estudio de Steeb (Steeb et al., 1994), emplearemos
$N = 2$. Como vimos previamente, la entropía de Boltzmann continua era

$$H = \int_0^1 \rho(x) \ln \rho(x) dx$$

Tomaremos los momentos de orden inferior de las variables dinámicas que, para sistemas
ergódicos, son iguales a los momentos de la distribución de probabilidad para la evolución
temporal. Siguiendo los pasos habituales, tendremos, para el caso general, una ecuación variacional
asociada a:

$$\Gamma = - \int_0^1 \rho(x) \ln \rho(x) dx + \beta_0 \left( 1 - \int_0^1 \rho(x) dx \right) +
+ \sum_{n=1}^N \beta_n \left( < x^n > - \int_0^1 x^n \rho(x) dx \right)$$

siendo $\beta_n$ el conjunto habitual de multiplicadores de Lagrange asociados a las $N + 1$ ligaduras. La
ecuación variacional nos da en este caso:

$$\rho(x) = \exp \left( -1 - \sum_{n=0}^N \beta_n x^n \right) \equiv \frac{1}{Z} \exp \left( \sum_{n=0}^N \beta_n x^n \right)$$

Aqui $Z \equiv e^{1+\beta_0}$ queda determinada mediante la normalización de la densidad de probabilidad,
de forma que

$$\frac{1}{Z} \int_0^1 \exp \left( \sum_{n=0}^N \beta_n x^n \right) dx = 1$$

Los restantes multiplicadores de Lagrange se obtienen resolviendo el conjunto de $N$ ecuaciones
triangulares acopladas:

$$< x^m > = \frac{1}{Z} \int_0^1 x^m \exp \left( \sum_{n=0}^N \beta_n x^n \right) dx$$

para $m = 1, 2, \ldots, N$.

Consideremos dos ejemplos resolubles analíticamente. El primero será la aplicación triangular,
con $\mu = 2$. Este sistema dinámico es mixing y ergódico. La distribución exacta de probabilidad es
$\rho(x) = 1$, así que sus momentos asociados son

$$< x^n > = \frac{1}{n + 1}$$

Tomando los dos primeros momentos, $< x > = 1/2$ y $< x^2 > = 1/3$, la distribución será

$$\rho(x) = \frac{1}{Z} \exp \left( -\beta_1 x - \beta_2 x^2 \right)$$
Debemos entonces resolver las ecuaciones no lineales:

\[ 1 = \int_0^1 \exp \left[ -1 - \beta_0 - \beta_1 x - \beta_2 x^2 \right] \equiv I \]

\[ \frac{1}{2} = \int_0^1 x \exp \left[ -1 - \beta_0 - \beta_1 x - \beta_2 x^2 \right] \equiv -\frac{dI}{d\beta_1} \]

\[ \frac{1}{3} = \int_0^1 x^2 \exp \left[ -1 - \beta_0 - \beta_1 x - \beta_2 x^2 \right] \equiv -\frac{dI}{d\beta_2} \]

para hallar los multiplicadores de Lagrange. Se obtiene \( \beta_0 = -1, \beta_1 = 0, \beta_2 = 0 \). Así que obtenemos \( \rho(x) = 1 \), en concordancia con la solución exacta.

Finalmente, tomemos la aplicación logística para \( \mu = 4 \). La densidad de probabilidad es (Schuster, 1989),

\[ \rho(x) = \frac{1}{\pi \sqrt{x(1-x)}} \]

Los momentos toman ahora la forma

\[ < x^n > = \frac{1}{2^m} \left( \frac{2n!}{n!n!} \right) \]

Luego los primeros momentos serán \( < x > = 1/2 \) y \( < x^2 > = 3/8 \). Las nuevas ecuaciones son ahora:

\[ 1 = \int_0^1 \exp \left[ -1 - \beta_0 - \beta_1 x - \beta_2 x^2 \right] \]

\[ \frac{1}{2} = \int_0^1 x \exp \left[ -1 - \beta_0 - \beta_1 x - \beta_2 x^2 \right] \]

\[ \frac{3}{8} = \int_0^1 x^2 \exp \left[ -1 - \beta_0 - \beta_1 x - \beta_2 x^2 \right] \]

Resolviendo numéricamente, encontramos los valores

\[ \beta_0 = +2.69242 \]
\[ \beta_1 = -6.76825 \]
\[ \beta_2 = -\beta_1 \]

Empleando estos valores, la integración numérica de la expresión integral para el exponente de Lyapunov nos da un valor \( \lambda_L = 0.72 \), muy próximo al valor exacto.

### 5.25 Apéndice B: Derivación de las ecuaciones de Lorenz

Siguiendo la exposición presentada por Schuster (Schuster, 1989) describiremos aquí el procedimiento básico de obtención de las ecuaciones de Lorenz a partir del modelo basado en las ecuaciones básicas del movimiento de un fluido convectivo.

La idea de partida, descrita por el famoso experimento de Bénard, es estudiar el comportamiento de una capa de fluido emplazada entre dos superficies planas y paralelas, separadas una distancia \( h \) (en el eje vertical, figura 5.49). Si calentamos homogéneamente por debajo, se producirá un transporte de calor entre esta superficie y la de arriba (que estará más fría), el cual, en principio, tendrá lugar por conducción. Sin embargo, si esta diferencia aumenta se produce un
Figura 5.49: Fluido convectivo: aparecen rollos convectivos de forma espontánea.

nuevo fenómeno de autoorganización: el fluido presenta transporte convectivo y junto a éste se da la organización espacial en forma de rollos convectivos. Si la temperatura sigue subiendo, esta situación se hace inestable y aparecen nuevas frecuencias que, eventualmente, desembocan en el caos.

El fluido estará caracterizado por un campo de velocidades \( v(r, t) \) y un campo de temperaturas, \( T(r, t) \). Las ecuaciones básicas que describen el modelo serán:

- La ecuación de Navier-Stokes:

\[
\rho \frac{\partial v}{\partial t} = F - \nabla p + \mu \nabla^2 v
\]

- La ecuación de Fourier para la conducción de calor:

\[
\frac{dT}{dt} = \kappa \nabla^2 T
\]

- La ecuación de continuidad:

\[
\frac{\partial \rho}{\partial t} + \text{div}(\rho v) = 0
\]

con las condiciones de contorno:

\[
T(x, y, z = 0, t) = T_0 + \Delta T
\]

\[
T(x, y, z = h, t) = T_0
\]

Siendo \( \rho \) la densidad del fluido, \( \mu \) su viscosidad, \( p \) la presión, \( \kappa \) es la conductividad térmica y \( F \) indica la fuerza asociada a la gravedad, dada por \( F = \rho g k \), siendo \( k \) el vector unitario en la dirección negativa del eje \( z \). Las no-linealidades de este sistema de ecuaciones provienen de la ecuación de Navier-Stokes (que es cuadrática en la velocidad).
Supondremos a continuación que el sistema es invariante bajo traslación en la dirección $y$, de manera que asumimos que los roles de convección son de longitud infinita. Supondremos adicionalmente que todos los coeficientes son constantes, excepto la densidad, que depende linealmente del gradiente de temperatura en la forma:

$$\rho = \rho^*(1 - \alpha \Delta T)$$

La ecuación de continuidad será entonces:

$$\frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} = 0$$

con $w = v_z$ y $u = v_x$. Podemos introducir la función $\psi(x, z, t)$ tal que:

$$u = -\frac{\partial \psi}{\partial z}; \quad w = -\frac{\partial \psi}{\partial x}$$

que satisface automáticamente las anteriores igualdades.

El paso siguiente es introducir una perturbación $\theta(x, z, t)$ del campo lineal de temperatura, por medio de la expresión:

$$T(x, z, t) = T_0 + \Delta T - \frac{\Delta t}{h} z + \theta(x, z, t)$$

lo que nos lleva al siguiente sistema de ecuaciones:

$$\frac{\partial}{\partial t} \nabla^2 \psi = - \frac{\partial \psi}{\partial t} \nabla^2 \psi + \nu
\nabla^4 \psi + g_0 \frac{\partial \theta}{\partial x}$$

$$\frac{\partial}{\partial t} \theta = - \frac{\partial \psi}{\partial t} \frac{\partial \theta}{\partial x} + \frac{\Delta t}{h} \frac{\partial \psi}{\partial x} + \kappa \nabla^2 \theta$$

Siendo $\nu \equiv \mu/\rho^*$ la viscosidad cinemática y en las que hemos empleado la siguiente notación abreviada:

$$\frac{\partial(a, b)}{\partial(x, z)} \equiv \frac{\partial a}{\partial x} \frac{\partial b}{\partial z} - \frac{\partial a}{\partial z} \frac{\partial b}{\partial x}$$

Asumiremos finalmente las condiciones de contorno empleadas por Lorenz:

$$\theta(0, 0, t) = \theta(0, h, t) = \psi(0, 0, t) = \psi(0, h, t) = \nabla^2 \psi(0, 0, t) = \nabla^2 \psi(0, h, t)$$

y, llevando a cabo un desarrollo de Fourier para $\psi$ y $\theta$, retendremos sólo los términos de orden inferior y emplearemos las relaciones:

$$\frac{a}{1 + a^2 \kappa} \frac{1}{\Delta T} = \sqrt{2X(t)} \sin \left(\frac{\pi x}{h}\right) \sin \left(\frac{\pi z}{h}\right)$$

$$\frac{\pi R}{R_c \Delta T} \theta = \sqrt{2Y(t)} \cos \left(\frac{\pi x}{h}\right) \sin \left(\frac{\pi z}{h}\right) - Z(t) \sin \left(\frac{2\pi z}{h}\right)$$

Donde ahora $R \equiv (g a h^3 / \nu) \Delta T$ es el número de Rayleigh, $a$ está definida por: la figura 5-49 (expresa el tamaño horizontal relativo del rollo convectivo respecto del grosor del líquido) y $R_c \equiv \pi^4 a^2 (1 + a^2)^3$. Insertando las relaciones anteriores en las ecuaciones reducidas para $\psi$ y $\theta$, obtenemos finalmente el modelo de Lorenz:

$$\frac{dx}{dt} = -\sigma(x - y)$$
\[
\frac{dy}{dt} = rx - y - zx \\
\frac{dz}{dt} = b(z + ry)
\]
donde las derivadas se llevan a cabo en realidad sobre el tiempo normalizado, \(\sigma = \nu/\kappa\) es el llamado número de Prandt, \(b = 4(1+a^2)^{-1}\) y \(\tau = \beta/R_o\) es el parámetro de bifurcación que hemos explorado al principio y que es proporcional a la diferencia de temperatura.

Bibliografía


Capítulo 6

Análisis de Fenómenos Caóticos

Hemos definido con anterioridad el concepto de atractor extraño y vimos que la geometría de estos objetos es, al menos en parte, resultante de la sensibilidad a las condiciones iniciales. El carácter disipativo de la dinámica garantiza la convergencia de las trayectorias dentro de la cuenca de atracción hacia cierta región acotada del espacio de las fases que define la dinámica del sistema y nos da una imagen del orden (y del determinismo) subyacente. Por otra parte, la sensibilidad a las condiciones iniciales hace que dos puntos inicialmente próximos situados sobre el atractor se alejen de manera exponencial con el paso del tiempo. El resultado de ambas tendencias da lugar a un proceso de estiramiento y plegado que origina las propiedades fractales ya comentadas. Veremos ahora cómo cuantificar, desde el punto de vista de la dinámica del sistema, la divergencia (y por tanto la inestabilidad local) de las trayectorias.

Ya hemos visto que podemos calcular la dimensión fractal de los conjuntos atractores, al menos en algunos casos. Hemos obtenido también expresiones para los exponentes de Lyapunov. Ahora bien, la importancia del caos determinista sería reducida si no fuéramos capaces de aplicar las ideas anteriores a la realidad. Y, aunque ciertamente las matemáticas son el lenguaje del universo, la realidad no se presenta anunciándose con ecuaciones diferenciales. Aunque seamos capaces de intuir cierto orden bajo el azar aparente, debemos medir cantidades relevantes o bien extraer el orden intrínseco mediante procedimientos formales adecuados.

En algunos casos podemos aplicar de forma directa algunas ideas ya expuestas. Así, la dimensión fractal de un atractor extraño generado por una aplicación bidimensional puede ser calculada mediante el procedimiento de box-counting (ver capítulo 3). Este hecho sugiere que tal vez podríamos emplear la medida de forma general para cualquier conjunto atractor, pero de hecho se ha probado que no es así (Greenside et al., 1982). Para sistemas dinámicos con tres o más grados de libertad, la convergencia de estas medidas suele ser mala, si la hay. Debemos emplear nuevas medidas de dimensión fractal. Pero esta no es la mayor dificultad. Los problemas vienen del hecho ya indicado de que no todas las variables son conocidas o accesibles.

Imaginemos una población de depredadores de la que conocemos (mediante estimas de algún tipo) su dinámica (fig. 6.1). Este sería el caso de las fluctuaciones de las poblaciones de los linces del Canadá, registrados por la Hudson Bay Company, que comerciaba con sus pieles. A lo largo de casi 200 años, el número de pieles contadas permitió disponer de una estimación de las fluctuaciones poblacionales. Los linces no son, sin embargo, un sistema aislado. Se alimentan de liebres árticas y éstas a su vez lo hacen de la vegetación. Al menos tres variables (si no más) se hallan en juego. Aunque existen también algunos registros de las fluctuaciones de liebres, que también exhiben ciertas periodicidades, no disponemos de series temporales equivalentes de todas las variables. Sólo de una. Hay casos aún más dramáticos, como es el cerebro. Hace mucho tiempo que se emplean los electroencefalogramas como fuente de información para el diagnóstico médico. En la
Figura 6.1: Oscilaciones temporales de la población de lince del Canadá (*Lynx canadiensis*). Se observa una periodicidad de unos 10-11 años con una gran variabilidad en la amplitud.

Figura 6.2: Oscilaciones temporales de la actividad cerebral registradas a partir de medidas de potencial en un punto del cuero cabelludo.
figura 6.2 vemos algunos ejemplos de estas señales, obtenidas mediante electrodos colocados sobre el cuero cabelludo. Estas señales son irregulares aunque, desde luego, no totalmente aleatorias. ¿Podrían estar asociadas con un fenómeno caótico? De ser así, parece difícil demostrar semejante hipótesis, ya que no disponemos de una idea a priori de cuántas variables están implicadas y ni siquiera la medida que llevamos a cabo nos permite saber exactamente qué tipo de variable estamos manipulando. Estos ejemplos nos dan una idea clara de las dificultades a las que nos enfrentamos. Nos preguntamos ahora ¿es posible extraer información de estas medidas? o, más explícitamente, ¿podemos conocer cuántas variables están implicadas y deducir la presencia de un atractor? Por sorprendente que parezca, la respuesta es afirmativa.

En primer lugar estudiaremos en este capítulo el análisis de datos mediante la función de autocorrelación y la transformada de Fourier. A continuación veremos cómo resolver el problema de medir la dimensión fractal de un atractor extraño cualquiera. Posteriormente discutiremos cómo estimar el exponente (o exponencia) de Lyapunov de un sistema dinámico sin recurrir a sus ecuaciones y finalmente veremos cómo es posible reconstruir la dinámica subyacente y determinar el número de dimensiones de un sistema experimental partiendo de una sola variable.

### 6.1 Función de autocorrelación

Una primera medida, fácilmente realizable, y que nos da cierta información acerca de la serie, se obtiene a partir de la llamada función de autocorrelación \( C(\tau) \). Supongamos una serie temporal de datos \( S = \{x_1, x_2, \ldots, x_N\} \). Sea \( \hat{z}_i = f_\mu(x_0) - < x > \). Entonces, la función de autocorrelación de la serie se define en la forma

\[
C(\tau) = \lim_{N \to \infty} \sum_{i=0}^{N-1} \hat{z}_{i+\tau} \hat{z}_i = \lim_{N \to \infty} \sum_{i=0}^{N-1} < x_{i+\tau} - < x > < x_i - < x >
\]

donde, para una aplicación unidimensional, en la que \( x_{i+1} = f_\mu(x_i) \), se tendría

\[
< x > = \lim_{N \to \infty} \sum_{i=0}^{N-1} f_\mu'(x_0)
\]

De esta definición se sigue que la función de autocorrelación proporciona una medida de la irregularidad de la secuencia de valores de la serie \( S \). En la definición vemos que se emplean los productos de las desviaciones respecto de la media \( < x > \), luego si la serie es periódica veremos que, para ciertos valores de \( \tau \), en los que el sistema presenta correlaciones (productos del mismo signo), tendremos máximos/mínimos en el valor estadístico de \( C(\tau) \).

Si disponemos de una densidad de probabilidad invariable \( \rho(x) \) para la aplicación dada (sobre el intervalo unidad), la función de autocorrelación puede escribirse en la forma:

\[
C(\tau) = \int_0^1 \rho(x)z f_\mu'(x)dz - \left[ \int_0^1 \rho(x)zdz \right]^2
\]

Para la aplicación triangular, por ejemplo, se tiene:

\[
C(\tau) = \int_0^1 x f_\mu'(x)dz - \left[ \int_0^1 xdz \right]^2
\]

\[
= \int_{-1/2}^{1/2} y f_\mu'(y + \frac{1}{2})dy + \frac{1}{2} \int_{-1/2}^{1/2} f_\mu'(y + \frac{1}{2})dy - \frac{1}{4} = \frac{1}{12} \delta_{\tau,0}
\]
**Figura 6.3:** Función de autocorrelación para (a) una serie temporal obtenida a partir de la aplicación logística con $\mu = 3.8$ y (b) electroencefalograma patológico (enfermedad de Creutzfeld-Jacob).
y por lo tanto las iteraciones están delta-correlacionadas. En la figura 6.3 se muestran dos ejemplos de funciones de autocorrelación obtenidas para la ecuación logística en el dominio caótico y para una serie temporal de un electroencefalograma.

### 6.2 Transformada de Fourier

Cualquier señal (o serie temporal, que podemos imaginar como un muestreo de cierta función) puede ser representada como la superposición de un número infinito de funciones periódicas. Para ser más explícitos, si \( f(t) \) es una cierta función del tiempo, admite una representación en la forma:

\[
f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} a_k \cos \left( \frac{k\pi}{L} t \right) + b_k \sin \left( \frac{k\pi}{L} t \right)
\]

conocida habitualmente como serie de Fourier. Los términos dentro de los paréntesis asociados a las funciones trigonométricas introducen los valores de las frecuencias asociadas a cada modo de Fourier, siendo los coeficientes de Fourier \( a_k \) y \( b_k \) las amplitudes asociadas a dichos modos. En cierta forma, estas amplitudes nos dan una medida de la importancia del término periódico al que están asociadas (véase el capítulo sobre neurodinámica).

Los coeficientes de la serie pueden determinarse a partir de las integrales definidas por:

\[
a_k = \int_{-L}^{+L} \cos \left( \frac{k\pi}{L} t \right) f(t) dt \quad k = 0, 1, ...
\]

\[
b_k = \int_{-L}^{+L} \sin \left( \frac{k\pi}{L} t \right) f(t) dt \quad k = 0, 1, ...
\]

La descomposición de una función en serie de Fourier admite una representación compleja que es la empleada habitualmente. De la notación anterior, podemos ver que:

\[
a_k \cos(kx) + b_k \sin(kx) = a_k \left( \frac{e^{ikx} + e^{-ikx}}{2} \right) + b_k \left( \frac{e^{ikx} - e^{-ikx}}{2i} \right) = c_k e^{ikx} + c_{-k} e^{-ikx}
\]

donde (suponiendo \( b_0 = 0 \)) se tiene

\[
c_k = \frac{a_k - ib_k}{2} \quad c_{-k} = \frac{a_k + ib_k}{2}
\]

cual con lo cual

\[
c_k = \frac{1}{2\pi} \int_{0}^{2\pi} (\cos kt - i \sin kt) f(t) dt = \frac{1}{2\pi} \int_{0}^{2\pi} f(t)e^{-ikt} dt
\]

\[
c_{-k} = \frac{1}{2\pi} \int_{0}^{2\pi} (\cos kt + i \sin kt) f(t) dt = \frac{1}{2\pi} \int_{0}^{2\pi} f(t)e^{ikt} dt
\]

Podemos resumir lo anterior en una sola expresión,

\[
c_k = \frac{1}{2\pi} \int_{0}^{2\pi} f(t)e^{-ikt} dt \quad k = 0, \pm 1, \pm 2, ...
\]

Señalemos que, si \( f(x) \) es una función real, entonces \( a_k \) y \( b_k \) son reales y los números \( c_k, c_{-k} \) serán en general complejos, mutuamente conjugados.
Figura 6.4: Espectro de Fourier para el atractor de Lorenz. Pese al carácter determinista de la dinámica, aparecen infinitas frecuencias representadas en forma de una banda continua, que sugiere en primera instancia algún tipo de ruido.

El método de la transformada de Fourier (TDF) ha sido descrito en numerosos textos especializados en el contexto del caos determinista, y aquí nos limitaremos a una revisión somera. Para una descripción más detallada, véase la lectura del capítulo III de Berge et al. (1988). Para el planteamiento detallado del cálculo numérico del problema, véase Press et al. (1992). La idea básica que subyace a la aplicación de este método es la posibilidad de detectar, a partir de un conjunto de medidas temporales, las frecuencias básicas que participan en el fenómeno y su importancia relativa (en términos de la amplitud asociada). Si el fenómeno es totalmente periódico y sólo aparece una frecuencia, como sería el caso de una señal sinusoidal, entonces esta frecuencia nos da la información básica acerca de la serie en relación a su periodicidad. Si la serie es biperiódica, tendremos dos frecuencias, aunque cada una puede poseer una amplitud asociada distinta. El método de la transformada de Fourier nos dará los valores de las frecuencias y su importancia en relación con las amplitudes. En ocasiones, debido a la naturaleza “ruídos” de la señal, el método permite detectar frecuencias no evidentes. En otros casos, no aparecen frecuencias características sino una distribución continua de éstas, lo que puede estar asociado a ruido o a caos.

Supongamos una serie temporal (obtenida a partir de una señal dada) \( \{x(t_j)\} \) siendo \( t_j = \Delta t \), \( j = 1, 2, \ldots, N \). La cantidad \( \Delta t \) es el intervalo de medida con el que se ha muestreado la señal (o en su caso el intervalo de integración numérica de las ecuaciones) o bien el tiempo característico asociado al sistema, como el que tenemos en sistemas dinámicos discretos.

Los inversos de los tiempos, esto es, \( f \equiv t_j^{-1} \) son frecuencias, y una de ellas, la frecuencia de Nyquist,

\[
f_c = \frac{1}{2\Delta t}
\]

juega un papel especialmente importante. Si las frecuencias características del fenómeno estudiado se limitan a valores inferiores a \( f_c \), puede garantizarse que la serie finita contiene la información suficiente acerca del problema (en cuanto a sus periodicidades). Dicho de otro modo, la serie incluye el máximo periodo dentro del cual la dinámica se repite. Si la serie no incluye una muestra lo bastante completa o bien si el muestreo que se realiza se hace con un \( \Delta t \) demasiado grande,
tendremos una serie distorsionada por los componentes de la señal no representados en \( \{x(t_j)\} \). Esta distorsión se denomina aliasing y puede detectarse por el hecho de que la TDF, que definiremos algo más adelante, no tiene a cero para \( f \rightarrow f_c \). En caso de darse esta situación, la serie tendrá poca utilidad, aunque existen formas de salvar este obstáculo, al menos en algunos casos.

La TDF de un conjunto finito \( \{x(t_j)\} \) (en el que el tiempo está muestreado en la forma: \( t_j \equiv j \Delta t, (j = 1, 2, ..., N) \)) se define a partir de la relación

\[
x(t_j) \equiv x(j \Delta t) = \frac{1}{N} \sum_{k=1}^{N} \hat{x}_k e^{-i 2\pi k j / N}
\]

\((i \equiv \sqrt{-1})\). Aquí \( \{\hat{x}_k\} \) es la TDF del conjunto \( \{x(t_j)\} \). Si empleamos el hecho de que:

\[
\sum_{k=1}^{N} \left[ \hat{x}_k e^{-i 2\pi k j / N} \right] \left[ \hat{\chi}_k e^{-i 2\pi \gamma j / N} \right]^* = N \delta_{k \gamma}
\]

(donde el asterisco indica la función compleja conjugada), podremos escribir la expresión de la TDF a partir de la serie de partida:

\[
\hat{x}_j = \sum_{k=1}^{N} \hat{x}_k e^{i 2\pi k (j / N)}
\]

y tenemos además, de 6.2.1 la siguiente identidad (teorema de Parseval)

\[
\sum_{j=1}^{N} |x(t_j)|^2 = \frac{1}{N} \sum_{j=1}^{N} |\hat{x}_k|^2
\]

Si empleamos ahora la función de autocorrelación,

\[
C_\gamma = \sum_{j=1}^{N} x(t_j) x(t_{j+r})
\]

puede demostrarse que \( C_\gamma \) es:

\[
C_\gamma = \sum_{k=1}^{N} |\hat{x}_k|^2 e^{-i 2\pi \gamma k / N}
\]

y la inversa de esta función nos da una expresión de gran utilidad:

\[
|\hat{x}_k|^2 = \sum_{\gamma=1}^{N} C_\gamma e^{-i 2\pi \gamma k / N}
\]

La gráfica de \( P(f_k) = |\hat{x}_k|^2 \) en función de la frecuencia asociada \( f_k = k / (N \Delta t) \) se denomina espectro de potencia y es el más habitualmente empleado. La denominación viene de la asimilación de \( P(f) \) a una potencia, esto es, a cierta cantidad de energía por unidad de tiempo. Consideraremos a continuación algunos ejemplos típicos.

El espectro de Fourier, igual que la función de correlación antes introducida, no es una medida apropiada para establecer si un fenómeno dado es o no caótico. Los sistemas que exhiben estocasticidad (cierto tipo de aleatoriedad) también exhiben espectros con una banda amplia de frecuencias, así que esta observación (por otra parte cualitativa) por sí sola es poco informativa. Sin embargo, para explorar sistemas experimentales bien controlados, en los que los efectos del
Figura 6.5: Espectros de Fourier el modelo discreto de Lotka-Volterra, con distintos valores del parámetro de bifurcación. Para las órbitas que exhiben cierta periodicidad, obtenemos espectros de potencia con picos bien definidos, mientras que para los atractores caóticos obtenemos bandas de frecuencia casi continuas.
Figura 6.6: Serie temporal y espectro de Fourier obtenidos de experimentos de dinámica de fluidos (Couette-Taylor). La aparición de nuevas bifurcaciones va acompañada de la aparición de nuevos armónicos. En esta figura estamos en las proximidades de la transición hacia el caos (régimen cuasiperiódico). Se indica también una sección de Poincaré (Brandstater y Swinney, 1987)
Figura 6.7: Serie temporal y espectros de Fourier correspondientes al experimento de Couette-Taylor, ya en la zona de caos desarrollado (compárese con la figura 6.6). La aparición de nuevas bifurcaciones va acompañada de la aparición de nuevos armónicos y de una banda continua de frecuencias. La sección de Poincaré indica la destrucción del toro invariante (Brandstater y Swinney, 1987).
Figura 6.8: Partiendo de un toro (izquierda) y deformándolo de forma suave y progresiva, podemos obtener distintos objetos topológicamente equivalentes.

ruido sean conocidos, la aparición de bifurcaciones puede ser adecuadamente caracterizada y proporcionar información útil. Un trabajo clásico en esta dirección lo llevaron a cabo Harry Swinney y sus colaboradores en relación con la aparición de caos en fluidos turbulentos. Empleando un sistema de estudio estándar, conocido como el experimento de Couette-Taylor, en el que un fluido gira entre dos cilindros concéntricos en rotación (Brandstater y Swinney, 1987; capítulo 5) estos físicos analizaron las series temporales obtenidas a partir de las mediciones de velocidad del fluido, que vemos en la figura 6.6. La evolución hacia el caos sigue, como ya vimos, un escenario de Ruelle-Takens. Los espectros de Fourier correspondientes se muestran a la derecha. La aparición de armónicos sucesivos nos indica la aparición de dinámicas más complejas con un espectro de Fourier que muestra una banda continua de frecuencias junto a los armónicos dominantes.

6.3 Teorema de Whitney y reconstrucción

Hasta ahora hemos supuesto casi siempre que la serie temporal que analizamos proviene de un sistema dinámico conocido, con variables también conocidas de antemano. Sin embargo, lo más probable es que, ante un problema dado, sólo uno de los componentes del sistema completo sea accesible a nuestro conocimiento. Así ocurre en los ejemplos de la introducción. En el primero, en el que mostrábamos la serie de cambios temporales de las capturas de linces en Canadá, disponemos de información acerca de uno de los componentes de una cadena trófica compleja. Podemos conjugar al menos dos variables adicionales: las presas (básicamente liebres árticas) y la vegetación de la que éstas se alimentan. Si sólo disponemos de una variable, ¿cómo se puede extraer información relevante acerca del sistema dinámico completo? En algunos casos, la situación es aún peor: la medida de los potenciales locales en el cerebro nos proporciona una serie de temporal para la que ni siquiera tenemos una variable o conjunto de variables de referencia bien definidas.

Si no existiera alguna forma de extraer propiedades relevantes a partir de estas series, que serán las que encontraremos en nuestros experimentos o mediciones de campo, poco sería lo que pudiéramos decir acerca del caos determinista. Existe, sin embargo, un procedimiento efectivo para recuperar, a partir de la serie temporal única, las propiedades relevantes del sistema completo y, lo que es aún más importante: de conocer cuántas variables precisamos para describirlo.

Antes de introducir este procedimiento, detengámonos a definir un concepto matemático que emplearemos más adelante: la idea de equivalencia topológica. Supongamos que partimos de un objeto $\Omega \subset \mathbb{R}^3$ tal y como un toro (figura 6.8). Imaginemos que podemos deformarlo de forma continua (sin cortar ni pegar) hasta convertirlo en $\Omega'$. Este segundo objeto comparte con el primero su superficie suave, continua, y un agujero. Podríamos seguir deformando $\Omega'$ hasta obtener $\Omega''$ (figura 6.8), que sigue manteniendo las mismas propiedades. Está claro también que podemos ir
de un lado a otro de forma reversible. Los objetos $\Omega$, $\Omega'$ y $\Omega''$ (y todas sus formas intermedias) son 	extit{topológicamente equivalentes}. Un toro y una esfera, por ejemplo, no son topológicamente equivalentes: no podemos transformar el uno en otro en la forma indicada. Más formalmente, la equivalencia topológica se define, como veremos, por medio de un difeomorfismo que describe en términos matemáticos la transformación de un objeto en otro. Los atractores extraños son objetos geométricos a los que podremos aplicar esta idea.

Supongamos dada una serie de $N$ datos ordenados cronológicamente y correspondientes a una sola variable, digamos:

$$\Gamma(N) = \{x(t_1), x(t_2), ..., x(t_N)\}$$

La idea intuitiva que analizaremos a continuación es relativamente fácil de explicar. La serie de puntos anterior corresponde a una de las variables (digamos la $j$-ésima) del sistema $D$-dimensional subyacente. El comportamiento de los puntos del conjunto $\Gamma(N)$ es el resultado de la interacción con las restantes $(D-1)$ variables. Luego, de alguna forma, la información contenida en $\Gamma(N)$ debe retener las propiedades del sistema $D$-dimensional.

Construyamos un nuevo conjunto de puntos a partir de $\Gamma(N)$, de dimensión $D$:

$$\Gamma(N, D, \tau) = \{x_i = (x(t_i), x(t_i + \tau), ..., x(t_i + (D-1)\tau))\}$$

siendo $\tau > 0$. La elección de $\tau$ y $D$ es por ahora arbitraria, si bien veremos que existen criterios para determinar su valor adecuado. La nueva serie de puntos $D$-dimensional $\Gamma(N, D, \tau)$ se obtiene por tanto tomando como coordenadas de $x_i$ los valores de $x(t_i)$ y los correspondientes $\tau$-desplazamientos. Parece intuitivamente claro que el valor de $x(t_i + n\tau)$ y el correspondiente $x(t_i + (n+1)\tau)$ se hallarán correlacionados, puesto que ambos han sido generados a través de cierta ley determinista. También parece claro que $\tau$ no puede ser en la práctica totalmente arbitrario: si $\tau$ es muy reducido, los puntos $x(t_i + n\tau)$ y $x(t_i + (n+1)\tau)$ se hallarán triviaslmente correlacionados (con una clara dependencia lineal) y si $\tau$ es excesivamente grande, la correlación entre ambos puntos puede ser de hecho nula.

Lo que se desprende de lo anterior es que existe una relación topológica entre el conjunto $\Gamma(N, D, \tau)$ y el correspondiente (y desconocido) conjunto

$$\Gamma(D) = \{x_i = (x_1(t_i), x_2(t_i), ..., x_D(t_i))\}$$

que obtendríamos de conocer las ecuaciones del sistema. Existe de hecho la equivalencia

$$\Gamma(N, D, \tau) \iff \Gamma(D)$$
y se basa en un teorema de la topología algebraica conocido como teorema de Whitney.

**Teorema (Whitney)**

Si $M$ es una variedad $N$-dimensional, existe una inmersión

$$\Phi : M \rightarrow \mathbb{R}^{2N}$$

(véase Takens, 1980).

Por inmersión se entiende un difeomorfismo $\Phi$ que aplica $M$ sobre una variedad compacta $N$-dimensional, $M' \subset \mathbb{R}^{2N+1}$. (Un ejemplo trivial de inmersión nos lo da un círculo: se trata de una variedad 1-dimensional que puede sumergirse en $\mathbb{R}^2$ pero no en $\mathbb{R}^1$, otras dificultades pueden aparecer en función de las propiedades geométricas del objeto, como es el caso de la botella de Klein (figura 6.9). La versión dada por Takens posteriormente es (Takens, 1980):

**Teorema (Whitney-Takens)**

Si $M$ es un conjunto compacto y

$$\Phi : M \rightarrow \mathbb{R}^{2N}$$

tal que $\Phi \in C^2$, entonces $\Phi$, genéricamente, es una inmersión.

De forma más general, si el objeto en cuestión es un atractor generado a partir de un sistema dinámico y en consecuencia podemos hablar de un campo vectorial (definido por el flujo del sistema dinámico), diremos que dos campos $C^r$-vectoriales $F$, $G$ son $C^k$-equivalentes (donde $k \leq r$) si existe un $C^k$-difeomorfismo $\Phi$ que transforma las órbitas $\phi_t(y)$ de $F$ en órbitas $\Phi(\phi_t(y))$ de $G$ de forma que preservan la orientación. De forma intuitiva, podemos imaginar $\Phi$ como un cambio de coordenadas invertible y (posiblemente) no-lineal que, aunque distorsiona el flujo, lo hace de forma suave y no confunde el orden en que las trayectorias visitan distintos puntos del atractor.

A partir de este resultado, se desprende de lo anterior que, dado un sistema dinámico

$$\frac{dx}{dt} = F_\mu(x(t))$$

con $x \in \mathbb{R}^D$, y siendo $\Gamma_\mu(D) = \{x = (x_1, ..., x_D)\}$, si tomamos la $j$-ésima variable y la serie correspondiente $\Gamma^{(k)} = \{x_k(t_1), ..., x_k(t_n), \ldots\}$, ésta nos permite la construcción de la serie $D$-dimensional $\gamma^{(k)}(D)$ definida por:

$$\gamma^{(k)}(D) = \{x_j = (x_k(t_j), x_k(t_j + \tau), ..., x_k(t_j + (D - 1)\tau)) \subset \mathbb{R}^D\}$$

que es topológicamente equivalente a $\Gamma_\mu(D)$.

Este resultado es de una importancia trascendental. Si el sistema dinámico que subyace a la serie temporal es caótico y posee un atractor extraño de baja dimensión, las propiedades de éste se conservarán en el atractor reconstruido empleando el método descrito. Podremos, en consecuencia, disponer de una fuerte evidencia de caos determinista partiendo únicamente de la serie temporal medida. Podemos ilustrar este resultado empleando como ejemplo el atractor de Rössler, que ya hemos analizado con anterioridad. En el dominio caótico, este atractor presenta una topología bien definida, confinada (aproximadamente) a una superficie bidimensional en la que tiene lugar el estiramiento que produce la sensibilidad a las condiciones iniciales y el plegamiento que permite confinar las trayectorias y generar la fractalidad. En la figura 6.10 indicamos este sistema completo, en el que vemos la proyección en el plano $(x, y)$ y la geometría básica. Podemos reconstruir el atractor empleando la variable $z(t)$, que nos permite reconstruir el atractor en el espacio $(x(t), z(t + \tau), x(t + 2\tau))$. El resultado de esta reconstrucción se indica debajo: la proyección sobre el
Figura 6.10: Equivalencia topológica entre (a) el atractor de Rössler (indicamos su proyección y su topología tridimensional) y (b) su contrapartida reconstruida. Pese a las diferencias, es evidente que poseen la misma estructura topológica, que, entre otras cosas, refleja la existencia de un plegamiento que origina la fractalidad estadística del atractor.
plano \((x(t), x(t + \tau))\) es claramente consistente con la anterior proyección, y la topología del atractor reconstruido nos muestra las mismas propiedades que el original: divergencia de condiciones iniciales en una región y plegamiento. Ambos atractores son topológicamente equivalentes.

Volviendo al ejemplo de los lince del Canadá, William Schaffer realizó estudios pioneros acerca de la reconstrucción de atractores basada en esta idea. En la figura 6.11 mostramos el resultado de reconstruir la serie anterior \(\{L(t_i)\}\) en un espacio tridimensional de coordenadas

\[
\{L(t), L(t + \tau), L(t + 2\tau)\}
\]

siendo \(\tau = 3\) años (Schaffer, 1984). Esta reconstrucción sugiere la existencia de un sistema determinista de baja dimensión. El estudio de Schaffer indica, de hecho, que las evidencias apuntan hacia un atractor extraño. En la misma figura indicamos la topología del atractor sugerida por la reconstrucción. Como vemos, aparece la estructura del atractor de Rössler.

Este tipo de tratamiento ha sido extensamente empleado en muchos sistemas, algunos de los cuales se analizan en este texto. Uno de los problemas importantes a tener en cuenta consiste en elegir adecuadamente el espacio en el que se reconstruye el atractor así como el valor adecuado del tiempo de retardo \(\tau\). El primer problema requiere conocer la dimensionalidad del sistema bajo estudio (problema que se analizará algo más adelante). El segundo se analizará en la siguiente sección, pero podemos adelantar algunas ideas. En la figura 6.12 indicamos algunos ejemplos de reconstrucción del atractor de Lorenz, empleando como variable \(x(t)\). Vemos que el efecto de \(\tau\) sobre el aspecto del atractor reconstruido es considerable. Para valores de \(\tau\) muy pequeños, los puntos están muy correlacionados, de forma que tenderán a tomar valores muy similares, y el atractor resultante quedará confinado a un dominio longitudinal muy reducido. Para \(\tau\) muy grande, los valores de \(x(t)\) son prácticamente independientes (no olvidemos que se trata de un sistema caótico) y como resultado la serie presenta valores descorrelacionados. Está claro por tanto que deberemos emplear valores intermedios de \(\tau\) tales que permitan un desacoplamiento significativo entre los valores consecutivos sin llegar a su independencia completa.
Figura 6.12: Reconstrucción del attractor de Lorenz empleando varios valores de $\tau$ (expresado como múltiplo del intervalo de integración numérica $\Delta t$). Los valores pequeños nos dan atractores comprimidos en una zona estrecha y los valores grandes indican claramente la descorrelación de la serie a grandes distancias temporales.

6.3.1 Elección de $\tau$ para reconstruir

Ya hemos visto que la calidad de la reconstrucción depende de la elección del tiempo de retardo $\tau$. Debemos disponer de un criterio claro de elección del intervalo de retardo que no se base en argumentos cualitativos y en esta sección veremos cómo estimar $\tau$ a partir de argumentos basados en la teoría de la información (Abarbanel, 1993).

Ya vimos con anterioridad cómo definir la información mutua entre dos (sub-) sistemas $A$ y $B$. En un contexto general, podemos llevar a cabo un conjunto de medidas $a_i \in A$ y $b_j \in B$. Si podemos obtener una distribución de probabilidad para las medidas sobre $A$ y $B$, esto es, $P_A(a_i), P_B(b_j)$ así como para las medidas simultáneas,

$$P_{AB}(a_i, b_j) \equiv P[a_i \in A \cap b_j \in B]$$

entonces, como sabemos, la información conjunta $I = I(A, B)$ puede calcularse a partir de la información acerca de una medida de $a_i \in A$ y $b_j \in B$, esto es,

$$I(a_i, b_j) = \log_2 \left[ \frac{P_{AB}(a_i, b_j)}{P_A(a_i)P_B(b_j)} \right]$$

(que es obviamente simétrica). La información mutua será el promedio sobre estas cantidades,

$$I = \sum_{a_i \in A} \sum_{b_j \in B} P_{AB}(a_i, b_j) I(a_i, b_j)$$

coincidente con las definiciones de información dadas anteriormente. Si es posible definir una distribución continua, entonces se tiene una expresión integral,

$$I = \int \int P(x, y) \log_2 \left[ \frac{P(x, y)}{P(x)P(y)} \right] dxdy$$
Análisis de Fenómenos Caóticos

Debemos trasladar ahora esta definición al problema de las observaciones realizadas sobre un sistema cualquiera del que poseemos información temporal en forma de una serie de valores \( \{s(t)\} \). Ahora, llamaremos \( A \equiv \{s(t)\} \) y \( B \equiv \{s(t + \tau)\} \). La información mutua será la información asociada a las observaciones realizadas en \( t \) con las obtenidas en \( t + \tau \). Así, para un \( \tau \) dado, tendremos:

\[
I(\tau) = \sum_{t=1}^{N} P(s(t), s(t + \tau)) \log_2 \left( \frac{P(s(t), s(t + \tau))}{P(s(t))P(s(t + \tau))} \right)
\]

Esta medida constituye de hecho una generalización, ahora para cualquier sistema no-lineal, de la función de correlación (empleada típicamente en sistemas lineales). Está claro que si las medidas temporales son independientes tendremos \( I(\tau) = 0 \).

Para evaluar \( I(\tau) = 0 \) a partir de \( \{s(t)\} \), calcularemos en primer lugar el histograma con una precisión \( \epsilon \) para las frecuencias de aparición del suceso \( s(t) \). Si la serie es larga y podemos conjeturar estacionariedad, tendremos \( P(t) = P(t + \tau) \). Para obtener la distribución conjunta, contaremos las frecuencias de aparición de los pares \( (s(t), s(t + \tau)) \) según el mismo criterio.

El problema a continuación es decidir qué propiedad de \( I(\tau) \) deberíamos emplear para establecer el valor de \( \tau \) más adecuado. Como criterio, Fraser y Swinney (1986) propusieron emplear el primer mínimo de \( I(\tau) \) como valor adecuado del retardo temporal.

### 6.4 Dimensión de correlación

Llegamos en esta sección a otro punto de enorme importancia en nuestra discusión: estimar el número de dimensiones \( d \) mínimo necesario para caracterizar nuestro sistema caótico. Una vez más, partimos de nuestra serie temporal única. Aunque hemos visto que es posible extraer una primera información cualitativa acerca de la posible existencia de caos por medio del método de reconstrucción, no sabemos en realidad cuántas dimensiones debemos emplear. Este problema es clave: si pudiéramos conocer el valor de \( d \) y este valor fuera reducido, dispondríamos de la evidencia necesaria para atacar el problema de construir un modelo. En definitiva: la existencia o no de una baja dimensionalidad marca el punto de inflexión que decide si existe o no un modelo simple del fenómeno. Como veremos, el método de reconstrucción nos permitirá abordar este problema y resolverlo satisfactoriamente.

El método que descriptremos en esta sección fue desarrollado por Grassberger y Procaccia (G. y P., 1983a) y es aplicable (bajo ciertas restricciones que estudiarémos) a cualquier sistema dinámico \( d \)-dimensional. Señálemos antes de continuar que el método de “box-counting”, que se introdujo con anterioridad, no es en general aplicable para dimensiones mayores que \( d = 2 \) (Greenside et al., 1982). Esta medida constituye una cota inferior de la dimensión fractal previamente definida y es fácilmente calculable a partir de un conjunto de puntos.

Sea dado este conjunto de puntos \( d \)-dimensional:

\[
\Omega_d = \{X_i \in \mathbb{R}^d, i = 1, 2, \ldots, N\}
\]

donde suponemos que los \( X_i \) se hallan sobre el atractor. La integral de correlación para \( r > 0 \) se define como

\[
C(r) = \lim_{N \to \infty} \frac{1}{N(N - 1)} \sum_{i < j} H\left(r - ||X_i - X_j||\right)
\]

siendo \( H(z) \) la función de Heaviside, para la cual \( H(z) = 1 \) si \( z \geq 0 \) y \( H(z) = 0 \) en caso contrario. El sumatorio puede interpretarse fácilmente: es el número de pares \( (X_i, X_j) \) tales que su distancia
Figura 6.13: (a) Determinación de $C(r)$ sobre el atractor de Lotka-Volterra. (b) Detalle del método.

La distancia euclidéa es inferior a $r$. Esta condición vendrá dada por tanto por

$$||X_i - X_j|| = \left[ \sum_{k=1}^{d} (x_i^k - x_j^k)^2 \right]^{1/2} < r.$$  

Para un objeto dado, el comportamiento de $C(r)$ con $r$ sigue una ley exponencial,

$$C(r) \approx r^{-\nu}$$

siendo el exponente $\nu$ la magnitud que denominaremos dimensión de correlación, estrechamente relacionada con $D$. Se define como:

$$\nu = -\lim_{r \to 0} \frac{\ln(C(r))}{\ln(r)}$$

Nuevamente, se evalúa $\nu$ trazando la curva definida por los pares de puntos $(\log(r), \log(C(r)))$ y se estima la pendiente en la región lineal (si esta existe) tal y como se indica en la figura 6.14. En esta zona lineal, las propiedades de invarianza de escala se ponen claramente de manifiesto. Los experimentos numéricos demuestran que $\nu$ se halla en general por debajo de $D$ si bien se aproxima mucho a su valor.

Supongamos que cubrimos nuestro atractor dado por el conjunto de puntos $\Omega_d$ con $d$-hipercubos de lado $\epsilon$. Si se trata de un objeto fractal, el número de cubos $N(\epsilon)$ de estas dimensiones que contienen puntos del atractor sigue una ley

$$N(\epsilon) \approx \epsilon^{-D}$$

Llamemos $n_i$ ($i = 1, 2, ..., N$) al número de puntos de $\Omega_d$ que se hallan en el interior del $i$-ésimo cubo no vacío. Podemos escribir aproximadamente

$$C(\epsilon) \approx \frac{1}{N^2} \sum_{i=1}^{M(\epsilon)} n_i^2 = \frac{M(\epsilon)}{N^2} < n^2 >$$
Figura 6.14: Función de correlación $C(r)$ para el modelo de Lotka-Volterra discreto.

siendo $<n^2>$ el promedio sobre todos los hiper-cubos ocupados. Empleando la desigualdad de Schwarz, se tiene:

$$C(\varepsilon) \geq \frac{M(\varepsilon)}{N^2} <n^2> = \frac{1}{N^2 M(\varepsilon)} \left[ \sum_{i=1}^{M(\varepsilon)} n_i \right]^2 = \frac{1}{M(\varepsilon)} \approx \varepsilon^{-D}$$

donde hemos empleado el hecho de que

$$N = \sum_{i=1}^{M(\varepsilon)} n_i$$

De este último resultado se sigue, como queríamos ver, que:

$$\nu \leq D$$

Para entender mejor la relación anterior, introduciremos una vez más la entropía de Boltzmann, dada por

$$S(\varepsilon) = -\sum_{i=1}^{M(\varepsilon)} p_i \log p_i$$

siendo $\varepsilon$ la precisión (como antes). Aquí $p_i$ es la probabilidad de que un punto pertenezca al $i$–ésimo (hiper–)cubo. Es decir,

$$p_i = \frac{\mu_i}{N}$$

para $N \to \infty$. Para un recubrimiento uniforme del espacio de fases, tenemos

$$p_i = \frac{1}{M(\varepsilon)}$$
y la entropía será en este caso particular, como ya sabemos,

$$S_0(\epsilon) = \log[M(\epsilon)] = C - D \log(\epsilon)$$

Pero en general tendremos que $S(\epsilon) < S_0(\epsilon)$. Si suponemos que la entropía puede escribirse en la forma:

$$S(\epsilon) = S_0 - \sigma \log(\epsilon)$$

($\sigma$ se denomina dimensión de información), obtenemos:

$$\sigma \leq D$$

Ahora, para probar que $\nu \leq \sigma$, consideremos dos recubrimientos (encajados entre sí) de precisión $\epsilon$ y $2\epsilon$, respectivamente. Tenemos obviamente que

$$M(\epsilon) = 2^D M(2\epsilon)$$

y si $p_i$ es la probabilidad de que un punto caiga en el interior del $i$-ésimo cubo de la $\epsilon$-partición y $P_j$ la de que caiga en el $j$-ésimo cubo de la $2\epsilon$-partición, tendremos:

$$P_j = \sum_{i \in j} p_i$$

Definamos ahora $\omega_i$ como $p_i = \omega_i P_j$, de donde se sigue que

$$\sum_{i \in j} \omega_i = 1$$

De acuerdo con la ecuación para $C(\epsilon)$ previamente obtenida, se obtiene (a primer orden):

$$C(\epsilon) \approx \sum_{i \geq 1} p_i^2 \leq \sum_{i \geq 1} P_j^2 \sum_{i \in j} \omega_i^2$$

y si asumimos que $\omega_i$ es independiente de $j$.

$$\frac{C(\epsilon)}{C(2\epsilon)} = \frac{\omega^2}{\omega}$$

Si calculamos la diferencia entre las entropías correspondientes a las dos particiones, vemos que

$$S(2\epsilon) - S(\epsilon) = \sum_j P_j \sum_{i \in j} \omega_i \log[\omega_i] = \frac{\omega \log[\omega]}{\omega}$$

y si definimos $W = \omega / \omega >$, podemos emplear la desigualdad

$$< x^2 > \geq \exp( < x \log(x) >)$$

para ver que, efectivamente, $\nu \leq \sigma$. Combinando a la vez $\nu$ y $D$, obtenemos una excelente estimación del contenido en información del atractor por medio de las desigualdades

$$\nu \leq \sigma \leq D$$

con igualdad sólo para el caso trivial de recubrimiento uniforme.

Tal y como han hecho notar Grassberger y Procaccia (1983) el valor de $\nu$ tiene un interes suplementario: es en cierta medida una cantidad más interesante que $D$. ya que es sensible al proceso dinámico subyacente. Remítimos al lector al artículo original para un estudio más detallado de la cuestión.

Queda un punto abierto en este apartado: la aplicación de esta medida a un conjunto de puntos experimentales arbitrario si desconocemos la mayoría de las variables que intervienen únicamente disponemos de una de ellas.
Figura 6.15: (a) Estructura básica del corazón humano. (b) Patrón normal de actividad cardiaca, con el complejo tónico de ondas de despolarización-repolarización.

6.5 Atractores extraños en electrocardiogramas

Los ejemplos de sistemas reales en los que se ha detectado caos determinista por medio de los métodos antes descritos son muy numerosos e incluyen turbulencia débil en fluidos (Brandstater et al., 1986), reacciones químicas caóticas (Roux, 1983); láseres (Izeda, 1980; Haken, 1975), epidemias (Schaffer y Kot., 1985) y dinámica de tejido cardiaco (Guevara et al., 1981). Entre los muchos ejemplos a considerar, está el caso de la dinámica del latido del corazón así como sus distintas patologías.

Si bien el estudio de los electrocardiogramas (ECG) tiene una larga tradición en medicina, la exploración detallada de la dinámica no-lineal del corazón no se inició hasta la aparición de la teoría del caos (véase Glass y Mackey, 1988, para una revisión de éste y otros sistemas fisiológicos de interés).

La fisiología del latido cardiaco es bien conocida. En el corazón humano (figura 6.15 (a)) encontramos cuatro cavidades: dos aurículas y dos ventrículos. La sangre venosa entra por la aurícula derecha, de ésta se desplaza al ventrículo derecho que, al contraerse, la envía a los pulmones en los que se llevará a cabo el intercambio gaseoso. A continuación esta sangre arterial volverá de nuevo a la aurícula izquierda, al ventrículo izquierdo y de éste (por contracción) a todo el cuerpo. El mecanismo de contracción muscular viene controlado por un marcapasos situado en la aurícula derecha, llamado nodo sinoauricular. Es en realidad una pequeña tira de músculo especializado en la que se originan los potenciales de acción que se transmiten hacia las aurículas. La contracción de éstas se aprecia en la onda del electrocardiograma sano y se denomina onda P (figura 6.15 (b)). Esta onda alcanza entonces un segundo marcapasos, llamado nodo auriculo-ventricular a partir del cual se propaga la onda a continuación a través del haz de Hiss-Purkinje (cuya estructura es fractal) dando lugar a la contracción de los ventrículos. La onda se transmite desde la cara interna hacia la periferia. La contracción ventricular es la que da lugar al llamado complejo QRS (figura 6.15 (b)) que incluye tres ondas distintas de despolarización. La relajación posterior del ventrículo introduce el último elemento en la dinámica: la onda T de repolarización.

Las distintas cardiopatías darán lugar a modificaciones importantes de esta situación. Parece
evidente que la situación normal, que identificamos con un corazón sano, corresponde a un sistema totalmente periódico. Esta regularidad está por supuesto presente pero, de hecho, el corazón sano es caótico. Este caos determinista puede detectarse de dos formas básicas, que veremos a continuación. Pero antes, vamos a ver cómo aparecen los atractores reconstructos a partir de distintas señales obtenidas de un sujeto sano así como de enfermos que padecen distintas cardiopatías (figura 6.16).

Cualquier cambio en la transmisión del impulso a través del corazón puede dar lugar a modificaciones que alteren la forma de las ondas del electrocardiograma. En ocasiones estos cambios tienen que ver con lesiones del tejido, pero también pueden producirse por arritmias de diverso origen, como el debilitamiento de miocardo. El electrocardiograma puede llegar a ser enormemente desordenado, como ocurre en la fibrilación ventricular, en la que se observan cambios irregulares espasmódicos de los potenciales. Si analizamos el ECG normal, podemos aplicar las técnicas antes mencionadas para determinar su dimensionalidad. Partiendo del registro, llevamos a cabo la reconstrucción del atractor (tal y como se muestra en la figura 6.16 (a) para $d_e = 2$). Aunque la periodicidad está claramente presente (tenemos un órbita compleja que se repite) esta representación ya sugiere cierta variabilidad que podría indicar la existencia de una dinámica más compleja que un simple ciclo límite deformado.

Babloyantz y sus colaboradores (Babloyantz y Destiche, 1988) estudiaron este sistema y obtuvieron como resultado una clara prueba de caos de baja dimensionalidad en la dinámica del corazón sano. Explorando mediante las técnicas anteriores los electrocardiogramas de individuos sanos, encontraron que los atractores reconstructos daban un valor de dimensión de correlación $\nu \in (4,5)$ y otras medidas (incluyendo los diagramas de Poincaré) eran consistentes con dicho resultado.

Una segunda evidencia en favor de dinámicas complejas en el corazón nos la da el estudio de una serie de tiempo distinta: la que obtenemos a partir de los tiempos entre latidos, y que se denomina frecuencia cardíaca. Un ejemplo de esta serie para un corazón sano, mostrada en la figura 6.17, indica la existencia de autosimilaridad en su estructura temporal (confirmada por distintas medidas). Incluso en situación de reposo, el ritmo cardíaco fluctúa de forma irregular, lo que contradice la apariencia de regularidad que antes mencionábamos.

El estudio demuestra que las propiedades dinámicas del corazón sano son consistentes con un sistema débilmente caótico. El espectro de Fourier da una señal del tipo $P(f) \propto f^{-\beta}$ siendo $\beta \approx 1$, lo cual, como veremos, tiene implicaciones considerables en el estudio de los sistemas complejos. Por el contrario (y en contra de lo que sugiere la intuición) numerosas cardiopatías dan lugar a situaciones dinámicas de alta periodicidad. La muerte subita es un ejemplo claro de esta situación, en la que el corazón se desplaza de una situación de caos a una situación periódica que, en este caso, es patológica desde el punto de vista fisiológico.

Estos resultados han sido extendidos a otras situaciones relacionadas con la fisiología humana y distintas patologías (Glass y Mackey, 1988). El hecho de que podamos caracterizar mediante sistemas dinámicos simples situaciones patológicas aparentemente complicadas ha hecho acuñar el término enfermedades dinámicas a los autores de estos estudios. Existen numerosos ejemplos que van desde las fluctuaciones de glóbulos blancos en sangre en ciertas leucemias hasta desórdenes respiratorios como la conocida por respiración de Cheyne-Stokes en la que el ritmo de ventilación se hace muy irregular.

### 6.6 Límites fundamentales en $\nu$ y $\lambda_L$

Desde el inicio de los estudios acerca de la existencia de caos determinista en sistemas reales (especialmente mediante técnicas de reconstrucción) los análisis de series temporales han sido numerosísimos. En un buen número de casos, como los sistemas químicos, láseres o electrocardiogra-
Figura 6.16: Atractores reconstruidos a partir de electrocardiogramas obtenidos (a) de un sujeto sano (b) de un sujeto con una cardiopatía congénita (c) durante una taquicardia ventricular y (d) de un sujeto que ha padecido en alguna ocasión una taquicardia ventricular.

Figura 6.18: Serie temporal del cociente $O^{18}/O^{16}$ (obtenido por interpolación de Grassberger (1986).
Figura 6.19: (a) Función de correlación \( C(r) \) para la serie anterior y (b) saturación aparente de \( \nu \).

mas, los resultados han sido muy claros. Pero estos resultados no siempre han sido transparentes, debido a la longitud de las series (a menudo muy sugerentes, pero claramente insuficientes), al ruido asociado a éstas o a ambos factores. Un caso particularmente confuso nos lo dan los estudios de Nicolis y Nicolis (1988) acerca de la existencia de un atractor de baja dimensión en la dinámica a gran escala del clima terrestre.

Para obtener la información necesaria, se emplearon los registros de las muestras de profundidad de hielos polares. A partir de éstas, se estimó el cociente \( O^{18}/O^{16} \) entre isótopos de oxígeno, que proporciona una estimación del volumen de hielo total e indirectamente del calentamiento global del planeta. En la figura 6.18 se muestra un ejemplo de estas series (de unos 500 puntos en los trabajos de Nicolis y posteriores) que nos da las fluctuaciones a lo largo de unos \( 8 \times 10^5 \) años. La serie temporal es irregular, pero por otra parte observamos algunos signos de regularidad que sugerirían algún tipo de estructura subyacente. Podemos reconstruir el posible atractor, y a continuación estimar la dimensión fractal del conjunto.

Para los datos anteriores, el diagrama \( (\log(C(r)) - \log(r)) \) así como la relación entre \( \nu \) y la dimensión de inmersión \( d_e \) se dan en la figura 6.19. Vemos una aparente saturación hacia \( \nu \approx 3.1 \). El estudio de Nicolis y Nicolis concluía en que sus resultados eran compatibles con un atractor de baja dimensión. De ser válido, este resultado indicaría que las propiedades esenciales del clima a gran escala quedarían adecuadamente descritas mediante un sistema dinámico con cuatro grados de libertad. La conclusión es fuerte, pero el número de datos es reducido. Los análisis posteriores (ver Grassberger, 1986) cuestionaron la validez absoluta de estos resultados, especialmente la existencia de una adecuada saturación en \( \nu \). No basta con que la gráfica \( \nu - d_e \) se desvía claramente de la esperada por un sistema con ruido, sino que debemos exigir una clara saturación. De forma general, necesitamos de un criterio para disponer de una cota en el número de puntos empleados y las dimensiones estimables.

Eckmann y Ruelle (1993) propusieron una estimación de los límites fundamentales asociados al cálculo de \( \nu \) y del mayor exponente de Lyapunov \( \lambda_L \). Expondremos aquí su argumentación básica. Supondremos que se tiene una serie temporal \( \{x_1, x_2, ..., x_N\} \) sobre la que se aplicará el método de reconstrucción. En \( d_e = m \), tenemos una trayectoria \( m \)-dimensional dada por el conjunto de vectores

\[
X_n = (x_n, x_{n+1}, ..., x_{n+m-1})
\]
El método de GP consiste, como ya sabemos, en calcular la función de correlación \( C(r) \) que está acotada según

\[
0 \leq C(r) \leq \frac{1}{2} (N-m)(N-m+1) \approx \frac{N^2}{2}
\]

y a continuación estimamos \( \nu \). El método asume la relación \( C(r) \approx r^\nu \), y por tanto si \( d_a \) es el diámetro del atractor reconstruido, tendremos aproximadamente:

\[
C(d_a) \approx \frac{N^2}{2}
\]

luego

\[
C(r) \approx \frac{N^2}{2} \left( \frac{r}{d_a} \right)^\nu
\]

Como hemos visto, la estimación de la pendiente de \( \log[C(r)] \) se lleva a cabo para valores pequeños de \( r \), comparados con \( d_a \). Por otra parte, \( C(r) \) debe ser grande así que debemos imponer:

\[
\frac{N^2}{2} \left( \frac{r}{d_a} \right)^\nu \gg 1
\]

y

\[
\frac{r}{d_a} = \rho \ll 1
\]

Tomando logaritmos, obtenemos finalmente la restricción:

\[
2 \log(N) \geq \nu \log \left( \frac{d_a}{r} \right)
\]

Si tomamos el caso límite (la igualdad) se sigue que el algoritmo GP no permite estimar dimensiones \( \nu \) mayores que

\[
\nu = \frac{2 \log(N)}{\log \left( \frac{d_a}{r} \right)}
\]

Empleando logaritmos decimales, y si \( \rho = 0.1 \), vemos que para \( d_e \leq 6 \) necesitaremos \( N \approx 10^3 \), mientras que para \( d_e \leq 10 \) serán necesarios \( N \approx 10^5 \).

Un argumento similar puede extraerse para el exponente máximo de Lyapunov. Cualquier método que se emplee para estimar \( \lambda_L \) requiere que cerca de ciertos puntos \( X_n \) puedan hallarse otros puntos \( X_{n+k} \) (para algún \( k \)) de forma que pueda llevarse a cabo una estimación de la divergencia promedio de las órbitas. El número de puntos en una bola de radio \( r \) alrededor de \( X \) es \( \mathcal{N}(r) \approx r^\nu \) con \( \mathcal{N}(d_a) = N \), luego tenemos que

\[
\mathcal{N}(d_a) \approx D \left( \frac{r}{d_a} \right)^\nu
\]

Este resultado nos lleva nuevamente a imponer las desigualdades,

\[
\mathcal{N} \left( \frac{r}{d_a} \right)^\nu \gg 1
\]

y

\[
\frac{r}{d_a} = \rho \ll 1
\]
Debemos por tanto exigir que

\[ \log(N) > \nu \log \left( \frac{d_a}{r} \right) = \nu \log \left( \frac{1}{p} \right) \]

lo que nos lleva al resultado de que el número de puntos \( N \) necesarios para estimar \( \lambda_L \) es aproximadamente la raíz cuadrada de los empleados en el cálculo de \( \nu \).

Finalmente, debemos hacer notar que las desigualdades previas son fuertemente restrictivas, y se han desarrollado métodos de estimación adicionales de \( \nu \), con el objetivo de reducir los requerimientos en el número de datos manejados (Abraham et al., 1986; Ellner, 1988). Siempre que sea posible, deben aplicarse todos los métodos aquí analizados, unidos al mayor conocimiento accesible del problema en cuestión. Los modelos, si se construyen con la información suficiente, pueden ser valiosísimos y complementar en mayor o menor grado las carencias de datos.

### 6.7 Exponentes de Lyapunov: método de Wolf

Una medida adicional que podemos llevar a cabo es la estimación numérica de los exponentes de Lyapunov a partir de una serie temporal. Supongamos que nuestro sistema dinámico posee dimensión \( d \) (ya sabemos cómo determinarla). Si el sistema es caótico, al menos uno de los \( d \) exponentes de Lyapunov será positivo. Si los ordenamos de mayor a menor, un exponente máximo de Lyapunov positivo nos dará evidencia adicional de caos. Existen distintos métodos de cálculo de exponentes de Lyapunov (ver Ott et al., 1994) y aquí sólo nos detendremos en el más simple, desarrollado por Wolf y colaboradores en 1985.

Supongamos que nuestro atractor \( \Omega \) (típicamente reconstruido en un espacio de dimensión \( d \) a partir de una serie temporal) está formado por un conjunto de puntos \( X(t) \in \Omega \). Tomemos uno de estos puntos \( X_1(0) \in \Omega \) y busquemos el punto más cercano \( X_2(0) \) tal que su distancia sea

\[ L(0) = ||X_1(0) - X_2(0)|| > \epsilon \]

esto es, el punto más cercano por encima de cierta separación mínima \( \epsilon \) que corresponde a la amplitud del (posible) ruido asociado a la medida. Dado que a menudo este valor no se conoce, el método requiere ciertas pruebas empíricas (Wolf et al., 1985).

A continuación, calculamos la distancia entre las imágenes obtenidas al cabo de cierto tiempo \( T \), esto es,

\[ L(T) = ||X_1(T) - X_2(T)|| \]

a partir de las cuales podemos calcular la divergencia

\[ \lambda_L = \frac{1}{T} \log_2 \left[ \frac{L(T)}{L(0)} \right] \]

que será positiva si las trayectorias divergen. El proceso se repite a continuación empleando \( X_1(T) \) y buscando otra vez el punto más cercano, etc. Promediando, se obtiene la estimación del exponente de Lyapunov

\[ \lambda_L = \frac{1}{rT} \sum_{t=1}^{r} \log_2 \left[ \frac{||X_1(t + T) - X_2(t + T)||}{||X_1(t) - X_2(t)||} \right] \]

promediado sobre \( r \) parejas de puntos que cumplen las desigualdades antes indicadas.

El método de Wolf posee algunos inconvenientes, entre ellos el hecho de que precisa de varios parámetros para ser empleado. En otros métodos más sofisticados estos parámetros quedan reducidos a unos pocos.
6.8 La conjetura de Kaplan-Yorke

Los atractores extraños, como hemos visto, presentan típicamente propiedades geométricas propias de objetos fractales. Este resultado se sigue de la coexistencia simultánea de los procesos de estiramiento y plegado. La disipación conlleva el plegado y la sensibilidad a las condiciones iniciales el estiramiento. Los exponentes de Lyapunov miden el estiramiento de las trayectorias y la dimensión fractal las propiedades geométricas resultantes de ambos procesos. Parece natural esperar que exista una relación (tal vez exacta) entre ambas cantidades. De ser así, podría ser muy útil en el estudio de sistemas experimentales a la vez que permitiría comprobar el resultado de ambas medidas. En esta sección daremos argumentos cualitativos que conducen a dicha relación, que ha sido formulada como una conjetura (Jackson, 1991).

Consideremos para empezar una aplicación bidimensional \( f_\mu(x, y) \) (que supondremos en el dominio caótico) y que apliquemos ésta al conjunto de puntos \( U \) definido por el cuadrado unidad

\[
U \equiv [0, 1] \times [0, 1]
\]

El resultado de aplicar una vez \( f_\mu(x, y) \) a \( U \) es un nuevo conjunto \( S_1 \in U \) en el que se ha producido estiramiento en una dirección y plegamiento en la otra. Obtenemos así un nuevo objeto con dimensiones características dadas por \( L_1 > 1 \) (dirección de estiramiento) y \( L_2 < 1 \) (plegado), como se indica en la figura 6.20.

Asumiendo disipación, tenemos que \( L_1 L_2 < 1 \). El área por tanto decrece y obtenemos una “herradura” dentro de \( U \) (recordemos la transformación del panadero, estudiada en el capítulo anterior). Podemos recubrir \( S_1 \) con cuadrados de lado \( L_2 \). El número de éstos será:

\[
N(L_1) \approx \frac{L_1}{L_2}
\]

Al repetir de nuevo la operación, obtenemos una nueva herradura de longitud \( L_1^2 \) y grosor \( L_2^2 \). Ahora necesitaremos cuadrados de lado \( L_2^2 \) para recubrir el nuevo conjunto \( S_2 \); su número será ahora:

\[
N(L_2^2) \approx \left[ \frac{L_1}{L_2} \right]^2
\]

Para la \( k \)-ésima iteración, obtendremos

\[
N(L_2^k) \approx \left[ \frac{L_1}{L_2} \right]^k
\]
elementos.

Consideremos ahora la conocida expresión para la dimensión fractal, dada por

$$\lim_{\epsilon \to 0} N(\epsilon) \approx \epsilon^{-D}$$

De esta obtenemos, empleando

$$N(L^k) \approx \left[ \frac{L_1}{L_2} \right]^k = (L_2^k)^{-D}$$

una relación para la dimensión fractal y los valores $L_1, L_2$:

$$D = 1 + \frac{\log(L_1)}{|\log(L_2)|}$$

y, puesto que los $L_k$ como sabemos están relacionados con los exponentes de Lyapunov por $L_k = e^{\lambda_k}$, se sigue la relación entre dimensión y exponentes de Lyapunov

$$D = 1 + \frac{\lambda_1}{|\lambda_2|}$$

Esta sería una expresión particular de la conjetura K-Y. ¿Es este resultado general? Kaplan y Yorke propusieron en 1970 que efectivamente así era. Podemos esbozar el razonamiento como sigue (Jackson, 1991). Si ordenamos los exponentes de Lyapunov en orden decreciente, esto es $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$, y $M \in \mathbb{Z}$ es el mayor entero tal que

$$\sum_{j=1}^{M} \lambda_j \geq 0$$

(que equivale a decir que $L_1 L_2 \ldots L_M > 1$), Kaplan y Yorke conjecturaron que la dimensión del atractor sería

$$D_{ky} = M + \frac{\sum_{j=1}^{M} \lambda_j}{|\lambda_{M+1}|}$$

Esta igualdad es clara en el caso 2-dimensional, donde $\lambda_1 + \lambda_2 < 0$, $M = 1$ y por tanto $D_{ky} = D$. Para ver que la igualdad es cierta para otros casos (aunque no en todos, como demostraron Grassberger y Procaccia, 1983b) analizaremos el caso tridimensional.

Sean ahora las longitudes características $L_1, L_2, L_3$ y supongamos que $L_1 L_2 L_3 < 1$, de forma que $L_3 < 1$. Consideraremos el caso $L_1 > L_2 > 1$ de manera que tenemos expansión en dos direcciones. Un ejemplo de este caso se muestra en la figura 6.21. Para recubrir el atractor con cubos de lado $L_3$, precisaremos

$$N(L_3) = \left[ \frac{L_1}{L_3} \right] \left[ \frac{L_2}{L_3} \right]$$

cubos. Si iteramos $k$ veces, necesitaremos

$$N(L_3^k) = \left[ \frac{L_1 L_2}{L_3^2} \right]^k$$

cubos. En el límite para $k \to \infty$, se obtiene
Figura 6.21: Caso 3D, con $L_1 > L_2 > 1$ (véase el texto).

$$\left[\frac{L_1 L_2}{L_3^2}\right]^k = (L_3^k)^{-D}$$

y de aquí,

$$D = 2 + \frac{\log(L_1 L_2)}{|\log(L_3)|} = 2 + \frac{\lambda_1 + \lambda_2}{|\lambda_3|}$$

Ahora bien, dado que $\lambda_1 + \lambda_2 + \lambda_3 < 1$, tenemos $M = 2$ y por tanto $D_{ky} = D$.

Ahora consideraremos una argumentación adicional para la conjetura KY. Consideremos un subespacio $M$-dimensional de $\mathbb{R}^n$ que se expande y hacia el cual convergen todas las trayectorias. La dirección de aproximación más lenta se da a lo largo del eje $M + 1$, luego si tomamos hipercubos de lado $L_{M+1}^k$, podremos recubrir el attractor empleando un número

$$N(k) = \left[\prod_{j=1}^{M} \frac{L_j}{L_{M+1}^k}\right]^k$$

Si utilizamos este $N(k)$ para estimar el número mínimo de cubos necesario para recubrir el attractor, entonces tendremos

$$N_{ky}(L_{M+1}^k) = \left[\prod_{j=1}^{M} \frac{L_j}{L_{M+1}^k}\right]^k = [L_{M+1}^k]^{-D_{ky}}$$

Hay una fuente posible de error al realizar la estimación, que se da debido al hecho de que podíamos encontrar una colección de hipercubos capaz de recubrir el attractor y tal que su número fuera inferior a $N_{ky}$. Que ello sea así o no dependerá de cómo tenga lugar el plegamiento definido por la dinámica. Un ejemplo simple (Grassberger y Procaccia, 1983b) nos da el caso $L_1 > 1 > L_2 > L_3$. En la figura 6.22 vemos el plegamiento en la dirección 3. Es obvio que los $L_3^k$-cubos pueden incluir muchas capas generadas en el proceso de plegamiento a lo largo de la dirección 3. Tendríamos así $N(L_3^k) < N_{ky}(L_3^k)$, esto es, $D_{ky} > D$.

### 6.9 Detección de determinismo

En presencia de una serie temporal irregular (con un espectro de Fourier sugerente de aperiodicidad de algún tipo) una posible forma de detectar caos determinista y distinguir el fenómeno de ruido,
consistentemente simplemente si la serie está o no generada por un sistema determinista.
La detección de determinismo puede llevarse a cabo de distintas formas (Kaplan y Glass, 1991: Wayland et al., 1993) pero todas ellas descansan en la hipótesis de continuidad de las órbitas en el espacio de fases. Este hecho se considera en estos métodos como evidencia de determinismo.

Supongamos que partimos de una serie temporal \( \{x(t)\} \) con \( t = 1, \ldots, N \). Construimos en primer lugar la serie \( E \)-dimensional

\[
X(j) = \{z(j), z(j + \tau), \ldots, z(j + (E - 1)\tau)\}
\]

que, como sabemos, quedará adecuadamente aproximada por cierta función continua \( f \) (que representa el campo vectorial asociado al sistema dinámico, si éste está definido). Para un \( X(i) \) dado arbitrario, un método de detección de determinismo (Wayland et al., 1993) es buscar \( k \) vectores \( \{X(j)\} \) tales que

\[
\|X(i) - X(j)\| < \varepsilon
\]
más concretamente, el conjunto de \( k \) puntos más cercanos (según un entorno de radio \( e \)) a \( \mathbf{X}_i \). A continuación calculamos las imágenes de estos puntos, que llamaremos \( \mathbf{Y}_j \) y calcularemos el vector desplazamiento:

\[
\mathbf{V}(j) = \mathbf{Y}(j) - \mathbf{X}(j)
\]

Si el sistema es determinista, deberíamos esperar que estos vectores fuesen aproximadamente iguales. Si \( < \mathbf{V} > \) es la media de estos desplazamientos,

\[
< \mathbf{V} > = \frac{1}{k + 1} \sum_{j=0}^{k} \mathbf{V}(j)
\]

finalmente, podemos calcular el error de traslación \( \epsilon_t \), definido por:

\[
\epsilon_t = \frac{1}{k + 1} \sum_{j=0}^{k} ||\mathbf{V}(j) - < \mathbf{V} > ||^2
\]

empleamos la distancia euclídea). El error de traslación mide la dispersión experimentada por \( \mathbf{X}_t \) relativa al desplazamiento promedio \( < \mathbf{V} > \). Si la serie temporal es determinista, los \( \mathbf{V}(t) \) serán aproximadamente iguales y el error de traslación muy pequeño. En la figura 6.23 vemos un ejemplo de cálculo de \( \epsilon_t \), donde se ha promediado sobre \( N_r \) puntos elegidos al azar sobre el atractor de Hénon (parámetros estándar) y para una serie de ruido blanco. El promedio \( < \epsilon_t > \) se ha calculado empleando \( N = 1024 \) puntos, con \( r = 2 \), \( k = 4 \) y \( N_r = 100 \) puntos al azar sobre cada serie. El sistema de Hénon exhibe un crecimiento exponencial en \( < \epsilon_t > \) al aumentar \( E \), mientras que se mantiene aproximadamente constante para un sistema aleatorio.

Existen otras medidas de interés que permiten descubrir la presencia de caos determinista en series temporales. Algunas de ellas se basan en representaciones apropiadas del sistema dinámico empleando gráficas que permitan detectar las recurrencias del sistema (Eckmann et al., 1987). Otros, de enorme interés, se basan en la existencia de predictibilidad a corto plazo característica de los sistemas que exhiben caos determinista. Uno de los métodos más interesantes es el método de predicción no-lineal, desarrollado por G. Sugihara y R. May en 1990. Consiste en dividir la serie temporal en dos mitades, emplear la primera como fuente de datos conocidos y a partir de esta primera mitad llevar a cabo la predicción sobre la segunda parte. Para cierta dimensión de inmersión dada reconstruimos el atractor del sistema y, siguiendo una filosofía similar a la del método anterior, tomamos puntos de la segunda mitad de la serie y buscamos aquellos de la primera mitad que están más cerca (con cierto criterio de proximidad). La proyección posterior en el tiempo de estos puntos (empleando un criterio geométrico dado) se compara con el valor esperado conocido en la segunda mitad de la serie. La bondad de la predicción de mide y promedia. Este método se aplicó con muy buenos resultados a algunas series bien conocidas de epidemias, confirmando la existencia de caos que ya se había conjeturado por otros métodos.

Existen diversas monografías acerca de estas (y otras) aproximaciones (Ott et al., 1994; Abarbanel et al., 1993; Bascompte, 1995), que aquí sólo nos hemos limitado a resumir.

### 6.10 Control del caos

Tras haber analizado una cantidad considerable de sistemas caóticos, probablemente se haya llegado a la conclusión de que la dinámica caótica es en cierto sentido mucho más desordenada que la dinámica periódica o cuasiperiódica. Por ejemplo, si consideramos un péndulo simple no disipativo, sabemos que éste oscila eternamente con una amplitud fijada que depende de la condición inicial. Si variáramos ligeramente la posición o la velocidad iniciales, el péndulo cambiará su amplitud. Y
nada más. Si contemplamos su espacio de fases, veremos que antes se movía en una órbita y ahora lo hace en otra. Ambas órbitas son estables e independientes. Dadas las condiciones iniciales, el movimiento siempre tiene lugar en un lugar del espacio de fases limitado, y no se puede acceder al resto. El movimiento, efectivamente, es altamente ordenado e infinitamente predictable, y pequeños cambios en las condiciones iniciales únicamente inducen pequeños cambios en la dinámica a largo término.

Pensemos ahora en un sistema caótico. Para empezar, si un sistema es caótico resulta que probablemente también será ergódico, con lo cual su trayectoria puede acabar visitando todo el espacio de fases del sistema. Si variamos ligeramente la condición inicial, ya sabemos que obtendremos divergencia exponencial de las trayectorias, pero los lugares del espacio de fases visitados por el sistema serán los mismos. El sistema, ciertamente, puede parecer mucho más desordenado, pero también es mucho más flexible. Si en este sistema realizamos un cambio arbitrariamente pequeño en las condiciones iniciales, la dinámica a largo término sufre un cambio enorme, que de hecho imposibilita la predicción. En este sentido, el estudio de la respuesta de un sistema dinámico a pequeñas perturbaciones puede ser una medida de la existencia de dinámica caótica en el sistema.

Recordemos el escenario de Feigenbaum hacia el caos. La dinámica del sistema, variada en función de un parámetro, presenta una serie de bifurcaciones sucesivas que tienen un punto de acumulación, a partir del cual la dinámica es caótica. En cada una de estas bifurcaciones, una órbita de periodo $p$ se inestabiliza (no desaparece, sino que ya no es un atractor estable de la dinámica) y da lugar a una órbita de periodo $2p$. Cuando nos adentramos en el dominio caótico, las órbitas periódicas aún existen, pero son inestables. En particular, un atractor extraño puede ser caracterizado mediante estas órbitas periódicas (Lathrop y Kostelich, 1989), y debido a la ergodicidad del sistema, puntos infinitamente cercanos a cualquiera de ellas van a ser visitados por el sistema, en un cierto tiempo característico que puede ser estimado. Más que eso, arbitrariamente cerca de cualquier punto del espacio de fases, existe una órbita periódica. Esto es, las órbitas periódicas son densas en el atractor extraño.

La flexibilidad de los sistemas caóticos posibilita el uso de técnicas de control que estabilicen algunos de las infinitas órbitas periódicas presentes en un atractor extraño, mediante pequeñas perturbaciones en las variables o en los parámetros del sistema. Esta es la esencia de la teoría del control del caos. Los primeros pasos los dieron en 1990 E. Ott, C. Grebogi y J. Yorke, quienes introdujeron lo que actualmente se conoce como método OGY, y que se basa en la variación controlada de los parámetros del sistema. Para la aplicación de este método es necesario conocer la expresión analítica de la órbita que se desea estabilizar, o cuanto menos el punto por donde intersecta la sección de Poincaré (el punto fijo de la aplicación correspondiente). También es posible el cálculo a partir de precisas estimaciones numéricas, con lo cual en principio no es necesario conocer a priori la dinámica del sistema de forma analítica.

Rápidamente se intentó aplicar este tipo de control a dispositivos experimentales, y los resultados interesantes no se hicieron esperar. Los ejemplos más inmediatos fueron sistemas clasificables dentro de lo que llamariamos osciladores caóticos, sencillos de recrear en el laboratorio y también (como se vio) sencillos de controlar (Ditto, Rauseo y Spano, 1990; Singer, Wang y Bau, 1991).

Un segundo método de control, que relajaba las condiciones exigidas por el método OGY, fue propuesto por Güémez y Matías (método GM) en 1994. En este caso, era posible llegar a estabilizar órbitas de periodo arbitrario $p$ en un sistema caótico aplicando perturbaciones periódicas con este mismo periodo. La ventaja de este nuevo método es que no es necesario conocer a priori la dinámica del sistema, ni analíticamente ni numéricamente. Veamos estos métodos y sus aplicaciones.

6.10.1 El método OGY

Seguiremos en esta sección básicamente el artículo de F. J. Romeiras et al. (1992), ya que es la referencia en donde este método está más ampliamente explicado.
Consideraremos, pues, sistemas dinámicos discretos del tipo

\[ x_{i+1} = F(x_i, p) \]  

(6.10.1)
donde \( p \) es el parámetro variable del sistema, de valor real pero restringido a un intervalo relativamente estrecho en este caso. En principio podríamos tratar con sistemas continuos, pero nos restringiríamos a su sección de Poincaré, con lo cual estos sistemas se incluyen en el caso 6.10.1.

El valor de \( p \) está restringido porque es el parámetro que se utilizará para controlar el sistema, así que únicamente se permitirán pequeñas variaciones alrededor del valor \( \bar{p} \), que es el que da la dinámica real del sistema, \( |p - \bar{p}| < \delta \). Supondremos que trabajamos en el régimen caótico, y por tanto cuando \( p = \bar{p} \) debemos obtener un atractor extraño con infinitas órbitas periódicas inestables arbitrariamente próximas a cualquier punto. Debemos decidir en primer lugar cuál de estas órbitas inestables deseamos estabilizar. Será necesario considerar la aproximación de primer orden (lineal) del sistema alrededor del punto fijo de la aplicación de Poincaré (correspondiente a la órbita periódica). Entonces, de forma local, en un entorno de este punto, se aplicará el método de control. La dinámica caótica implica que el sistema entrará en este dominio en tiempo finito.

Nos centraremos en la estabilización de órbitas de periodo 1, como ejemplo de la forma de trabajar con el método de control OGY. Llamemos \( x^*(p) \) al punto fijo inestable de la sección de Poincaré. Para valores de \( p \) cercanos a \( \bar{p} \) y en un entorno del punto fijo \( x^*(\bar{p}) \), la aplicación 6.10.1 se puede aproximar linealmente por

\[ x_{i+1} - x^*(\bar{p}) = A [x_i - x^*(\bar{p})] + B (p - \bar{p}) \]  

(6.10.2)
donde \( A \) es una matriz jacobiana \( n \times n \) y \( B \) es un vector \( n \)-dimensional,

\[ A = D_{x_i} F(x_i, p), \quad B = D_p F(x, p) \]

con las derivadas parciales evaluadas en los puntos fijos, \( x^*(\bar{p}) \) y \( \bar{p} \). Supondremos que el parámetro \( p \) es una función lineal de la variable \( x_i \) (así es como se introduce el control en el sistema) de la forma

\[ p - \bar{p} = -K^T [x_i - x^*(\bar{p})] \]  

(6.10.3)
La matriz \( 1 \times n \)-dimensional \( K^T \) debe de ser determinada de forma que el punto fijo \( x^*(p) \) sea estable. Sustituyendo 6.10.3 en 6.10.2 se obtiene

\[ x_{i+1} - x^*(\bar{p}) = (A - BK^T) [x_i - x^*(\bar{p})] \]  

(6.10.4)
que muestra que el punto fijo será estable si la matriz \( A - BK^T \) es asintóticamente estable, es decir, como se ha visto en el capítulo 2, si todos sus valores propios tienen módulo menor que la unidad.

**Técnica del emplazamiento de los polos**

El problema de determinar el valor del vector \( K^T \) para que la matriz \( A - BK^T \) tenga valores propios previamente especificados se soluciona mediante la técnica conocida como emplazamiento de los polos. Los valores propios de la matriz \( A - BK^T \) se denominan polos reguladores.

Si suponemos que deseamos obtener el conjunto \( \{\mu_1, \ldots, \mu_n\} \) de valores propios para la matriz \( A - BK^T \), los resultados que siguen garantizan la existencia y la unicidad de la solución, y proporcionan el método para obtenerla (método de Ackermann).
1. El problema del emplazamiento de los polos tiene una solución única si y sólo si la matriz $n \times n$

\[ C = (B^T A B^T A^2 B^T \ldots A^{n-1} B) \]

es de rango $n$. $C$ se llama matriz de controlabilidad.

2. La solución al problema del emplazamiento de los polos está dada por

\[ K^T = (a_n - a_{n-1}, \ldots, a_1 - a_1) T^{-1} \]

con $T = CW$, y

\[
W = \begin{pmatrix}
    a_{n-1} & a_{n-2} & \ldots & a_1 & 1 \\
    a_{n-2} & a_{n-3} & \ldots & 1 & 0 \\
    \vdots & \vdots & \ddots & \vdots & \vdots \\
    a_0 & 1 & \ldots & 0 & 0 \\
    1 & 0 & \ldots & 0 & 0
\end{pmatrix}
\]

donde $\{a_1, a_2, \ldots, a_n\}$ son los coeficientes del polinomio característico de $A$,

\[ |sI - A| = s^n + a_1 s^{n-1} + \ldots + a_n \]

y $\{a_1, a_2, \ldots, a_n\}$ son finalmente los coeficientes del polinomio característico de $A - BK^T$,

\[ \prod_{j=1}^{n} (s - \mu_j) = s^n + a_1 s^{n-1} + \ldots + a_n \]

El parámetro del control

Hasta el momento, se ha limitado la aplicación del control a un entorno del punto fijo $x^*$. Por otra parte, dado que se ha limitado el rango en el que se puede variar el parámetro de control $p$, obtenemos una restricción

\[ |K^T [x_i - x^*(\tilde{p})]| < \delta \]

que define una franja de anchura $2\delta/|K^T|$. Se activará el control sólo para valores de $x_i$ en el interior de este dominio, y el parámetro tendrá el valor real $\tilde{p}$ cuando el sistema se halle fuera de la franja determinada. En resumen, se determina el control por

\[ p - \tilde{p} = -K^T [x_i - x^*(\tilde{p})] u \left( \delta - |K^T [x_i - x^*(\tilde{p})]| \right) \]

para valores arbitrarios de $x_i$, no necesariamente cercanos al punto fijo, y donde $u$ es la función escalón o de Heaviside (ya vista).

En principio, cualquier elección de los polos reguladores en el interior del círculo unidad sería válida. Sin embargo, si se conoce el punto fijo y el valor de los vectores propios que definen las direcciones de las variedades asociadas a este punto fijo, se puede imponer por ejemplo que el vector $K^T$ tenga dirección paralela a la variedad estable, lo cual optimiza el tiempo de estabilización necesario.
6.10.2 Control de la aplicación de Hénon por el método OGY

Describiremos explícitamente cómo calcular las magnitudes anteriores en el caso concreto de la aplicación de Hénon (véase el capítulo sobre caos). Recordemos que esta aplicación 2-dimensional está dada por

\[
x_{n+1} = a - x_n^2 + by_n
\]
\[
y_{n+1} = x_n
\]  

(6.10.5)

Fijaremos el valor del parámetro \( b = 0.3 \) y consideraremos pequeñas variaciones alrededor del valor de \( a = 1.4 \) a fin de controlar la órbita periódica del sistema. El punto fijo de la aplicación es

\[
x^* = (x_0, y_0) = q(1,1), \quad q = -c + (c^2 + a)^{1/2}, \quad c = \frac{1}{2}(1 - b)
\]

para \( a > -c^2 \). La matriz jacobiana del sistema es

\[
\text{D}_x F(x_i) = \begin{pmatrix} -2x & b \\ 1 & 0 \end{pmatrix}
\]

La estabilidad del punto fijo está determinada por las raíces de la ecuación característica

\[
|\text{D}_x F(x^*, a) - sI| = 0
\]

El punto fijo será estable si \(-c^2 < a < 3c^2\) e inestable si \(a > 3c^2\) (por tanto inestable cuando \( b = 0.3 \)). Las cantidades relevantes que necesitamos son, calculadas a partir de las definiciones dadas en la sección anterior,

\[
A = \begin{pmatrix} -2x_0 & b \\ 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad C(B^t AB) = \begin{pmatrix} 1 & -2x_0 \\ 0 & 1 \end{pmatrix}
\]
\[
W = \begin{pmatrix} 2x_0 & 1 \\ 1 & 0 \end{pmatrix}, \quad T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad K^T = (\alpha_1 - a_1, \alpha_2 - a_2)
\]

donde

\[
\alpha_1 = 2q = -(\lambda_u + \lambda_s), \quad \alpha_2 = -b = \lambda_u \lambda_s
\]

y

\[
\alpha_1 = -(\mu_1 + \mu_2), \quad \alpha_2 = \mu_1 \mu_2
\]

Los valores propios de la matriz \( A \) son \( \lambda_s \) y \( \lambda_u \), correspondientes a las soluciones positiva y negativa, respectivamente, de

\[
\lambda_{u,n} = -q \pm (q^2 + b)^{1/2}
\]

Las cantidades \( \mu_1 \) y \( \mu_2 \) son los polos reguladores. Una posible elección de estas dos cantidades es

\[
\mu_1 = 0, \quad \mu_2 = \lambda_s
\]

y en este caso controlaremos el sistema privilegiando la dirección de la variedad estable. La matriz \( K^T \) será

\[
K^T = \lambda_u (1, -\lambda_s)
\]
Cuando el sistema entra en la zona

\[ |K^T x_i - x_i' (\tilde{a})| < \delta \]

activamos el control. El tiempo que el sistema tarda en estabilizarse en la órbita elegida es mínimo para la elección realizada de los pares reguladores.

### 6.10.3 El método GM

El método de Gámez y Matías (GM) de control del caos es más sencillo de aplicar que el anterior, ya que no implica ningún conocimiento sobre las órbitas periódicas del sistema. La idea es muy sencilla: si deseamos estabilizar una órbita de periodo \( p \) en un sistema caótico, debemos perturbar el sistema con pulsos periódicos proporcionales al estado que el sistema presenta. Para algún valor de la intensidad de la perturbación, que debe ser determinado en cada caso y para cada sistema, la órbita deseada se estabiliza. En principio, si se varía la intensidad de la perturbación, \( \gamma \), se obtiene todo un diagrama de bifurcación. Para algún valor de \( \gamma \), en particular, será posible obtener la órbita del periodo deseado.

Este método es más apropiado para controlar sistemas en los que no se conozcan los parámetros que los guían, pero en los que se pueda realizar una medición sobre las variables. Conociendo las variables y actuando sobre ellas se estabiliza el sistema. En particular, podemos pensar en sistemas químicos, sobre los que se podría añadir fácilmente un reactante externamente, o en ecosistemas, con parámetros difíciles de localizar, pero en los que se pueden evaluar adecuadamente las densidades de población, por ejemplo.

La aplicación del método es muy sencilla. Consideremos una aplicación \( n \)-dimensional del tipo

\[ x_{n+1} = F(x_n, \mu) \]

donde \( \mu \) es un parámetro del sistema. Si se desea estabilizar una órbita \( p \)-periódica, aplicaremos una perturbación a intervalos regulares, en la forma que se ha dicho, y de hecho tendremos una nueva aplicación de la forma

\[ x_{j+1} = F_j(x_j, \mu) \quad j = 1, \ldots, i - 1, i + 1, \ldots n \]

\[ x_{i+1} = F_i(x_i, \mu) \times (1 + \gamma \delta_{i,p}) \]

donde como se puede ver la perturbación se aplica a una sola de las variables, y \( \gamma \) es un parámetro en esta nueva aplicación. La variación de este nuevo parámetro puede inducir, como ya se ha comentado, una cascada de bifurcaciones sucesivas hacia el caos. El término \( \delta_{i,p} \) significa que sólo debemos multiplicar el estado actual del sistema por \( 1 + \gamma \) cada \( p \) pasos de tiempo.

### 6.10.4 Control de la aplicación de Hénon por el método GM

Consideremos de nuevo la aplicación de Hénon 6.10.5 y aplicaremos esta vez el método GM para controlar una zona caótica de esta aplicación. Tomaremos en particular los mismos parámetros que en la sección 6.10.3, \( a = 1.4 \) y \( b = 0.3 \), pero ahora perturbaremos periódicamente una de las variables. Hemos escogido la variable \( x \), que será perturbada cada paso de tiempo, con lo cual tenemos ahora la aplicación

\[ x_{n+1} = \left( a - x_n^2 + by \right) \times (1 + \gamma) \]

\[ y_{n+1} = x_n \]  

(6.10.6)
Figura 6.24: Control de la aplicación de Hénon según el método GM. Los parámetros de la aplicación original son $a = 1.4$ y $b = 0.3$. En el eje $x$ se varía el valor del parámetro perturbador $\gamma$, lo cual permite recuperar un diagrama de bifurcación completo.

Donde $-1 < \gamma < 0$, para mantener las variables dentro de su dominio de definición y evitar divergencias. Obsérvese en la figura 6.24 el nuevo diagrama de bifurcación correspondiente a la aplicación perturbada 6.10.6.

Como se puede ver, distintos valores del parámetro $\gamma$ provocan la estabilización de órbitas de diferentes periodos. Por ejemplo, podríamos estabilizar una órbita de periodo 2 utilizando valores de $\gamma$ entre -0.4 y -0.2, aproximadamente (figuras 6.24 y 6.25).

Los sistemas caóticos son controlables mediante pequeñas perturbaciones, precisamente porque son caóticos. Un sistema periódico, más regular, no posee esta flexibilidad. El control de sistemas regulares implica grandes perturbaciones, tan grandes como el efecto que se desea obtener. Parece claro que esta propiedad será una nueva característica a explorar y explotar en los sistemas caóticos. También funcionará el caso contrario, la desestabilización de dinámicas periódicas. En una dinámica originariamente caótica que haya sido estabilizada por algún tipo de malfunción, se puede toruar al caos original mediante pequeñas perturbaciones. Aunque es quizá demasiado pronto para haber obtenido resultados espectaculares, la idea del control del caos debería de tener amplia aplicación en medicina: nuestro sistema nervioso y nuestro corazón, los dos motores principales de nuestro cuerpo, son débilmente caóticos. Veremos que la periodicidad puede ser patológica (véase el capítulo sobre neurodinámica).

A nivel teórico se ha iniciado la aplicación de los métodos de control a redes neuronales, ya sea en sistemas continuos (Sepulchre y Babloyantz, 1993) o discretos (Solé y Menéndez de la Prida, 1995), en cualquier caso con resultados espectaculares y muy prometedores. A otro nivel, ya se empieza a hablar de la esperanza que estos métodos ofrecen a disfunciones periódicas, como la epilepsia (Glanz, 1994).

Los primeros experimentos de control en sistemas experimentales ya se han sucedido: se ha eliminado la turbulencia (manteniendo por tanto el régimen laminar) en fenómenos de convección térmica (Singer et al., 1991); se ha estabilizado un oscilador sometido a la acción gravitatoria y a una fuerza magnética (Ditto et al., 1990); se han eliminado arritmias quirúricamente inducidas en
Figura 6.25: Estabilización de una órbita de periodo 2 en la aplicación de Hénon mediante el método GM. El valor de la perturbación (aplicada cada paso de tiempo) es $\gamma = -0.2$. Se comienza a perturbar en el paso $t_1 = 100$, y se elimina el control en $t_2 = 200$.

el ventrículo de un conejo (Garfinkel et al., 1992), e incluso se ha empezado a trabajar ya en tejido neuronal cultivado in vitro (Schiff et al., 1994). Sin duda, estos ejemplos representan la punta del iceberg.

**Bibliografía**


Capítulo 7

Fenómenos Críticos

En capítulos anteriores hemos explicado la existencia de comportamientos dinámicos de distintos tipos. Hemos encontrado soluciones periódicas, cuasiperiódicas y caóticas, de orden y desorden. Ya en la introducción conjeturábamos que la complejidad debía ser un fenómeno emergente a medio camino entre las estructuras ordenadas y las desordenadas. Diversas propiedades clave, como son la existencia de fractalidad, una elevada capacidad de transferir información o la habilidad de algunos sistemas para responder a un entorno variable, necesitan cierto grado de orden y cierto grado de desorden. Hasta ahora no hemos profundizado en el problema de delimitar (si lo hay) el dominio en el que la complejidad aparece más a menudo. Una primera aproximación, enormemente valiosa, nos la proporciona la teoría de los fenómenos críticos.

Los fenómenos críticos representan un caso particular de las denominadas transiciones de fase. Siempre que un sistema físico pasa de una fase, o estado, a otro, decimos que experimenta una transición. En esta, las propiedades físicas del sistema cambian. En la vida cotidiana observamos numerosos ejemplos de transiciones de fase: el agua que hiere, la naftalina que se evapora, el hielo que se funde. Está claro en estos casos que la geometría de nuestro sistema y sus propiedades cambian debido a la transición. Las dos fases del sistema (por ejemplo, agua líquida antes y vapor de agua después), están claramente diferenciadas. Si la transición se realiza en condiciones cotidianas, habituales. Podemos encontrar unas condiciones habituales en la cocina de casa, donde, si la presión es de una atmósfera, el cambio tendrá lugar a una temperatura de 100°C. Como el vapor de agua es menos denso siempre se coloca sobre el agua líquida. A fin de que la transición tenga lugar debemos suministrar una cierta cantidad de energía al agua líquida. Esta energía se emplea en dotar a las moléculas de energía cinética, de modo que puedan romperse los enlaces por puente de hidrógeno, responsables del estado líquido habitual del agua. Esta energía suministrada se denomina en este caso calor latente de vaporización, y está claro que será una cantidad no nula en las condiciones habituales.

Supongamos ahora que, paulatinamente, vamos aumentando la presión a la que se produce la transición. Si la presión es superior a 1 atmósfera, la temperatura a la que el agua hiere aumenta a su vez. El gas es mucho más compresible que el agua, así que su densidad aumentará rápidamente cuando se lo someta a presión. ¿Qué sucede si aumentamos aún más la presión, hasta el punto en que gas y líquido presenten la misma densidad? Esto es posible, y cuanto más nos acercamos a este punto, más se asemejan las dos fases. Cuando las densidades se igualan, gas y líquido no son dos fases aisladas y distinguibles, sino que se hallan intimamente mezcladas: hay burbujas de líquido dentro del gas, que contienen a su vez más burbujas de gas menores, rellenas de pequeñas burbujas de líquido... y así hasta el nivel molecular. En este punto, para estas condiciones, donde la temperatura es de $T_e = 647°C$ y las densidades se han igualado a $ρ_e = 0.323g/cm^3$, ya no es necesario absorber calor para pasar a la fase gaseosa, así que el calor latente de vaporización se
Figura 7.1: Diagrama de fases del agua. Las líneas continuas representan las zonas de coexistencia entre fases. La línea de coexistencia líquido-gas acaba en un punto crítico. Rodeando esta zona se puede obtener un cambio de fase continuo (línea de puntos, entre A y B).

anula en este punto. Aquí es donde se produce una transición crítica entre las dos fases. En la figura 7.1 se puede ver el diagrama de fases del agua.

En el caso de la transición sólido-líquido para el agua, la coexistencia crítica de las dos fases (donde el calor latente de fusión sería nulo y las densidades se igualarían) no se produce para ningún valor de la presión. Todas las transiciones de este tipo son denominadas de primer orden, con calor específico de fusión no nulo y separación explícita de las fases.

El ejemplo de la transición crítica líquido-gas del agua nos sirve para introducir las medidas cuantitativas que pueden ser aplicadas a las transiciones de fase y, en particular, las utilizadas en el caso de los fenómenos críticos.

<table>
<thead>
<tr>
<th>Transiciones de fase de primer orden</th>
<th>Transiciones de fase de orden superior (críticas)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calor latente ≠ 0</td>
<td>Calor latente = 0</td>
</tr>
<tr>
<td>Variables del sistema discontinuas</td>
<td>Las variables del sistema cambian suavemente (las derivadas no)</td>
</tr>
<tr>
<td>Reordenación radical de la estructura del material</td>
<td>Reordenación suave de la estructura (p. de orden)</td>
</tr>
</tbody>
</table>

En la tabla anterior se resumen esquemáticamente las diferencias entre las transiciones discontinuas o de primer orden (calor latente, reordenación del material,...) y las transiciones continuas, críticas o de orden superior (calor latente nulo, variaciones suaves,...).

Una de las variables importantes del sistema, que ayuda a definir el tipo de transición, es en este caso la diferencia de densidad entre la fase líquida ($\rho_l$) y la fase vapor ($\rho_g$), $\Delta \rho = \rho_l - \rho_g$. Esta diferencia, finita antes de alcanzar el punto crítico, se anula al llegar a éste, y sigue siendo cero después de la transición (para temperaturas superiores a $T_c$). La magnitud que provoca el cambio puede ser la temperatura o la presión, pero obsérvese que en cualquier caso nos moveremos sobre la curva de coexistencia de las dos fases. La magnitud $\Delta \rho$ es el primer ejemplo de lo que se denomina parámetro de orden. Como su nombre indica, es una especie de medida del orden.
Figura 7.2: Comportamiento cualitativo del parámetro de orden en la transición líquido-gas del agua, en función de la temperatura.

Existe un sistema. Siempre tiene valor cero antes de la transición y se hace positivo de forma continua después (o viceversa). No existe, sin embargo, ningún método que nos indique cómo hallar, en general, un parámetro de orden. Para la transición líquido-gas del agua, el parámetro es la diferencia de densidades. Veremos que para los fenómenos paramagnéticos juega este papel la magnetización, pero para otros sistemas hallaremos una enorme variedad: el número de celdas en la máxima agrupación para una transición del tipo de la percolación (se verá más adelante), la amplitud cuántica de un cierto suceso para una transición a la superconductividad o la diferencia entre ciertas densidades molares en las mezclas de fluidos binarios, entre otros muchos.

Resulta esencial en la descripción y comprensión de los fenómenos críticos el estudio del comportamiento de ciertas magnitudes en el tiempo y en el espacio. Ya hemos visto que las dos fases líquido-gas del agua, antes de la transición crítica, están perfectamente diferenciadas. Sin embargo, en la transición y después de ésta, están fuertemente mezcladas. Está claro que ha habido un cambio en la geometría, en el tipo de homogeneidad que el sistema presenta antes y después. En particular, justo en el punto en que tiene lugar la transición, la coexistencia de las dos fases presenta una geometría fractal, únicamente limitada por la escala molecular, por un lado, y por el tamaño del sistema, por otro. La coexistencia de las dos fases en el punto crítico es una coexistencia dinámica. Existen fluctuaciones que abarcan todas las escalas de tiempo y también de espacio. En este sentido, se hace terriblemente difícil establecer valores exactos predeterminados, sea en el tiempo o en el espacio, debido a estas fluctuaciones tan violentas. Por ejemplo, en el caso de la transición crítica del agua, se produce un fenómeno óptico conocido con el nombre de opalescencia crítica. El agua pierde su aspecto transparente debido a la dispersión de la luz provocada por las fluctuaciones de todos los tamaños (dominios de gas o líquido que pasan de una fase a otra), las cuales le otorgan una apariencia turbia. De ahí el nombre de opalescencia.

Este tipo de geometría autosimilar es lo que confiere a los sistemas críticos su mayor interés. No hay ni un tiempo ni una longitud característica en el sistema. Los sucesos espaciales y temporales se repiten a cada escala a la que estos sistemas responden.

En las secciones siguientes formalizaremos todas las ideas descritas y veremos cómo precisamente los valores de los exponentes que caracterizan las leyes de escala, los llamados exponentes críticos, permiten agrupar bajo clases comunes diversos fenómenos críticos que se producen en sistemas muy diferentes a nivel elemental, y que no podrían ser comparados lejos de estas transi-
ciones críticas tan particulares. Este hecho será de gran importancia en nuestro intento de hallar propiedades generales comunes a distintos sistemas complejos.

7.1 El Modelo de Ising

El modelo de Ising representa, sin duda alguna, el ejemplo más utilizado para describir una transición crítica, en este caso entre un estado (o fase) paramagnético y otro ferromagnético. Fue ideado por W. Lenz en 1920, pero la primera solución exacta, para el modelo en una dimensión, la dio E. Ising en 1925. El modelo en dos dimensiones no fue solucionado hasta 1944 (Onsager), y a pesar de casi 80 años de esfuerzo, el modelo en tres dimensiones no ha podido ser resuelto exactamente.

Cuando hablamos de "resolver un modelo" nos estamos refiriendo, en este caso, a hallar analíticamente y de forma exacta, su función de partición. A partir de ella, se calculará el punto en que tiene lugar la transición (en caso de que exista) y los exponentes críticos que la caracterizan, si esto es posible.

Resolveremos explícitamente el modelo en una dimensión (veremos que no presenta transición crítica) para introducir las ideas básicas de las técnicas de la renormalización. Estas técnicas aprovechan la invarianza de escala que el punto crítico presenta para plantear ecuaciones de recurrencia, precisamente a escalas diferentes. Si se pueden solucionar, el punto fijo de la recurrencia proporciona el punto invariante de escala, el punto crítico. Por desgracia, estas ecuaciones son en general no resolubles, aún en el caso de tratar con modelos (como el de Ising) extremadamente simplificados.

7.1.1 El Modelo de Ising

Consideremos una red en n-dimensiones, y coloquemos en el centro de cada una de sus celdas un "espín". La representación más simple de este espín consiste en asignar un estado \( s_i = \pm 1 \) a la celda. Se supone que la configuración de energía menor se obtiene cuando todos los espines están alineados en la misma dirección, paralelos (esto es con idéntico valor). Sin embargo, esta ordenación sólo sería posible en ausencia de agitación térmica, a temperatura cero. En general, el estado observado del sistema representará un compromiso entre la situación de mínima energía (espines paralelos) y las fluctuaciones provocadas por el hecho de tener temperatura no nula. Al sistema le puede añadir un campo magnético externo en el caso más general, con lo cual se marca una dirección preferente de orientación. El tratamiento analítico del sistema comienza planteando su hamiltoniano \( (H) \). Siempre que trabajemos con sistemas conservativos, el hamiltoniano coincide con la energía del sistema, y en el caso del modelo de Ising esta función es

\[
H = \frac{1}{2} \sum_{i,j} J_{ij} s_i s_j - B \sum_i s_i
\]

\[J_{ij} = \begin{cases} 
J & \text{si } i \text{ y } j \text{ son vecinos} \\
0 & \text{en otro caso}
\end{cases}
\]

con \( J < 0 \), como corresponde a un sólido ferromagnético (la mínima energía se obtiene con espines paralelos) \(^1\). El caso \( J > 0 \) representaría un sólido anti-ferromagnético. El cambio no aporta nada nuevo, ya que las propiedades termodinámicas son idénticas al anterior.

\(^1\)Los productos entre vectores \( s, s_j \) y \( B_s \) son productos escalares, en general. Cuando los estados de los espines son dos, como en los casos que trataremos, resultan ser siempre paralelos, con lo que el producto escalar se convierte simplemente en el producto de los estados. En el caso del campo externo \( B \), escribiremos en ocasiones simplemente \( B \), para referirnos al módulo \( |B| = B \) cuando sea este valor el que se puede utilizar en la expresión.
Figura 7.3: Ejemplo de ordenación de los espines en el modelo de Ising 2-dimensional. Se ha marcado el mayor dominio correlacionado en la dirección del campo $B$ según el vecinaje de von Neumann.

El número de vecinos en el modelo de Ising es de 2 en una dimensión (derecha e izquierda), los cuatro conectados por las aristas en dos dimensiones (vecinaje de von Neumann), y serían 6 en tres dimensiones. $B$ representa el campo externo que tenderá a alinear los espines en su dirección. A partir de $H$ definimos la función de partición del sistema. En general, se define la función de partición $Z$ como

$$Z = \sum_{\text{estados}} e^{-\beta H} = \sum_i e^{-\beta E_i}, \quad \beta = \frac{1}{k_B T}$$

$k_B$ es la constante de Boltzmann, $T$ es la temperatura, y la suma se extiende a todos los estados o configuraciones del sistema, caracterizados por una energía $E_i$. A partir de $Z$ se pueden derivar todas las funciones termodinámicas del sistema. Para el modelo de Ising, $Z_I = \sum_{\{s_i\}} \exp \left[ \beta (B \sum_i s_i - \frac{1}{2} \sum_{ij} J_{ij} s_i s_j) \right]$

donde $\{s_i\}$ indica que la suma debe ser realizada sobre todas las posibles asignaciones ±1 a las celdas del sistema.

### 7.1.2 Exponentes críticos y universalidad

Ya hemos comentado la competencia existente en el sólido ferromagnético entre la tendencia a la mínima energía (alineación de los espines en una sola dirección) y el desorden introducido por la agitación térmica. Está claro que tenemos dos casos extremos: para $T = 0K$ el orden es total; todos los espines miran hacia "arriba" o hacia "abajo". Por otra parte, si la energía térmica es suficientemente grande, el sistema presentará una estructura muy desordenada, aleatoria, en la que el estado de una celda no proporcionará ninguna información sobre el de sus vecinas. Existe una temperatura ($T_c$), la temperatura de Curie, entre estos dos extremos, para la que tiene lugar una

---

2 Con frecuencia se habla de $H$ como de la energía de un sistema, y en ocasiones se trata al hamiltoniano como si lo fuese. De hecho, para ser rigurosos, deberíamos diferenciar entre ambas cosas, y hablar de $H$ como de un operador matemático, del cual las cantidades $E_i$ son los valores propios, los observables.
transición crítica, según la fenomenología descrita en la introducción. En este caso, el parámetro de orden ($\Phi$, en general), es una magnitud bastante intuitiva. Observemos: para $T$ pequeñas, de hecho simplemente por debajo de la temperatura crítica $T_c$, existe una magnetización neta en el sistema, el orden gana al desorden y hallamos una mayoría de espines apuntando en una de las dos direcciones, $+1$ o $-1$; para temperaturas elevadas, con un sistema aleatorio, la magnetización total se anula. Es sencillo, pues, definir el parámetro de orden como la magnetización normalizada del sistema a cada temperatura

$$\Phi \equiv m = \frac{1}{N^d} \sum_i s_i$$

$N$ es la medida lineal del sistema, y $d$ la dimensión del espacio en que trabajemos. Se observa experimentalmente que el parámetro de orden se hace cero como una potencia de la temperatura reducida

$$t = \frac{T - T_c}{T_c}$$

la cual mide la desviación de la temperatura respecto de su valor crítico:

$$\Phi \propto t^\beta \quad (t < 0)$$

Para todo sistema crítico, se define el exponente $\beta$ como el que caracteriza la forma en que el parámetro de orden tiende a cero. Otras magnitudes presentan singularidades, o divergencias, cuando $T \rightarrow T_c$, o bien en el mismo punto crítico. Por ejemplo, el calor específico tiene una singularidad del tipo

$$C_B \propto t^{-\alpha}$$

lo cual define un segundo exponente crítico, $\alpha$.

Definamos la longitud de correlación, $\xi$, como la medida de la dimensión lineal típica de la pieza más grande correlacionada en el sistema. Por “pieza correlacionada” entenderemos que todas las celdas que la constituyen están conectadas (siempre según el criterio de vecindad que se tome). Obsérvese en la figura 7.4 cómo aumenta esta longitud a medida que nos acercamos a la temperatura crítica.

De hecho, también presenta una singularidad, caracterizada por un tercer exponente:

$$\xi \propto t^{-\nu}$$

La longitud de correlación proporciona el tamaño máximo hasta el cual el sistema tiene propiedades autosimilares. El hecho de que sea una cantidad divergente en el punto crítico implica que la autosimilaridad se puede extender al sistema completo.

Un cuarto exponente, $\eta$, corresponde a la función de correlación conexa de dos puntos, $G_c^{(2)}(r)$. Para el caso de un espin $s_i$, la función de correlación de dos puntos se define como

$$G^{(2)}(i, j) \equiv < s_i s_j >$$

donde $<>$ indica un promedio térmico sobre la pareja de espines situados en las posiciones $i$ y $j$. En general, $G^{(2)}(r)$ se define en función del parámetro de orden $\Phi(r)$ 3 como

3El parámetro de orden no es necesariamente una cantidad escalar. En general, tendrá dimensión $D$, y siempre puede ser definido localmente. De ahí la dependencia explícita anterior del punto del espacio considerado.
Figura 7.4: Divergencia del tamaño medio de las agrupaciones en la transición de percolación. El tamaño medio $S$ es proporcional a la longitud de correlación $\xi$ (véase la sección 7.3). En el caso teórico la divergencia es estricta. En el caso experimental, se convierte en un máximo debido al corte impuesto por el tamaño finito del sistema.

$$G^{(2)}(r) \equiv \langle \Phi(0)\Phi(r) \rangle$$

A partir de aquí, la función de correlación conexa de dos puntos, que sólo considera las fluctuaciones en el parámetro de orden, es:

$$G^{(2)}_\xi(r) \equiv \langle \Phi(0)\Phi(r) \rangle - |\langle \Phi(r) \rangle|^2$$

Cuando $T = T_c$, se observa experimentalmente que $G^{(2)}_\xi(r)$ presenta la forma asintótica

$$G^{(2)}_\xi(r) \propto \frac{1}{r^{d-2+\eta}}$$

para distancias $r$ grandes comparadas con las distancias entre elementos, siendo $d$ la dimensión del sistema. $G^{(2)}_\xi(r)$ se relaciona con la longitud de correlación

$$G^{(2)}_\xi(r) \propto \exp \left\{ \frac{r}{\xi} \right\}$$

para $r$ grande, $\gamma$

$$0 \neq \frac{|T - T_c|}{T_c} \ll 1$$

El parámetro de orden fluctúa en bloques de todos los tamaños hasta $\xi$. Las fluctuaciones divergen cuando $T \to T_c$.

Se pueden definir dos exponentes más aparte de los cuatro presentados. Se define la susceptibilidad térmica de un material como la variación de la magnetización cuando se varía un campo externo a temperatura constante:

$$\chi_T \propto \left( \frac{\partial m}{\partial B} \right)_T$$
y se observa que $\chi_T$ diverge en el punto crítico según un exponente $\gamma$. Por otra parte, si el campo $B$ aplicado no es nulo pero es pequeño, se puede calcular el valor de $m$ cuando $B \to 0$, y se encuentra que, para $T = T_c$,

$$m \propto B^{1/\delta}$$

lo cual define un sexto y último exponente crítico, $\delta$.

En la tabla siguiente se resumen todos los exponentes que caracterizan los sistemas magnéticos. Obsérvese que todos se definen para un valor nulo del campo magnético ($B = 0$), excepto $\delta$.

<table>
<thead>
<tr>
<th>Exponente</th>
<th>Definición</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>$c_B \propto \alpha^{-1} \left( \left</td>
</tr>
<tr>
<td>$\beta$</td>
<td>$m \propto (T_c - T)^\beta, \ T \to T_c^-, \ B = 0$</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>$\chi_T \propto \left</td>
</tr>
<tr>
<td>$\delta$</td>
<td>$m \propto B^1, \ B \to 0, \ T = T_c$</td>
</tr>
<tr>
<td>$\nu$</td>
<td>$\xi \propto \left</td>
</tr>
<tr>
<td>$\eta$</td>
<td>$G^{(2)}(r) \propto \frac{1}{r^{\eta - 2\nu}}, \ T = T_c, \ B = 0$</td>
</tr>
</tbody>
</table>

Los exponentes anteriores no son todos independientes. Verifícan relaciones que hacen que sea suficiente conocer dos de ellos para determinar todos los demás. Las relaciones que cumplen son las siguientes

- $\alpha + 2\beta + \gamma = 2$ (Ley de Rushbrooke)
- $\alpha + \beta(\delta + 1) = 2$ (Ley de Griffiths)
- $\nu d = 2 - \alpha$ (Ley de Josephson)
- $(2 - \eta)\nu = \gamma$ (Ley de Fisher)

Los exponentes críticos caracterizan totalmente el comportamiento de las magnitudes del sistema cuando nos acercamos al punto crítico. Estos exponentes, de forma rigurosa, sólo estarán bien definidos en el límite termodinámico, cuando el tamaño del sistema tiende a infinito. En cualquier tipo de simulación, está claro que será necesario restringir esta condición y trabajar con sistemas finitos. Entonces, las divergencias se suavizan y pasan a convertirse en máximos, y el parámetro de orden resulta ser una función suave, con derivada continua, a diferencia del caso límite. Existen técnicas que permiten esquivar esta dificultad y proporcionan muy buenas estimaciones numéricas. El análisis del sistema a tamaños finitos pero a muchos diferentes, la introducción de funciones de corte que simularien el efecto de las fronteras, o la idea de utilizar escalas diferentes en una representación determinada del sistema son algunas de estas técnicas. También se han hecho aproximaciones analíticas (grupo de renormalización) que en ocasiones han permitido calcular tanto el punto en el que se produce la transición como los exponentes que la caracterizan exactamente, y en otras con gran aproximación. Daremos algunos ejemplos en las secciones siguientes.

4Aunque en algunas simulaciones se ha llegado a trabajar con redes de tamaño hasta $10^8 \times 10^8$, esto aún queda muy lejos de infinito. En mecánica estadística puede considerarse que el infinito se alcanza cuando el número de partículas es de $10^{23}$. 
El hecho de que sean los exponentes críticos los que caracterizan estas transiciones ha permitido definir lo que se conoce con el nombre de *universalidad*. Se ha constatado que estos exponentes tienen valores que son *marcadamente insensibles a los detalles específicos del sistema*. Por ejemplo, los exponentes no dependen de la topología de la red sobre la que coloquemos el sistema: triangular, cuadrada, hexagonal ... pero sí de su dimensión. La insensibilidad llega a tal punto en ciertos casos que, por ejemplo, sistemas hidrodinámicos y sistemas magnéticos presentan los mismos exponentes, dentro del margen de error con que pueden ser determinados. Cuando esto sucede, se dice que los sistemas pertenecen a la misma *clase de universalidad*: cerca del punto crítico se comportan del mismo modo, y por tanto, si encontramos ante una "física universal", en la que los detalles microscópicos se hacen irrelevantes debido a la aparición de la estructura a gran escala, emergente, que el sistema presenta.

### 7.1.3 Ising en 1 dimensión: Grupo de Renormalización

Hasta principios de la década de los 70 no se había conseguido ninguna teoría exacta que posibilitase la descripción de un sistema cerca de su punto crítico. Se había usado descripciones aproximadas (principalmente teorías de campo medio, como la de van der Waals o la de Landau) que fallaban cerca de las transiciones críticas. Sin embargo, en 1966, Kadanoff publicó un artículo en donde daba las ideas generales que, argumentaba, permitirían el cálculo de los exponentes críticos de un sistema sin necesidad de conocer exactamente su función de partición. Kadanoff puso el énfasis en las configuraciones del sistema, en su aspecto, y en cómo la invarianza de escala podría ser utilizada para la extracción de los exponentes críticos. Las ideas de Kadanoff tomaron cuerpo matemático en un trabajo posterior de Kenneth G. Wilson (1971) y más tarde gracias a la aparición de las técnicas del ahora llamado *grupo de renormalización* (Wilson y Kogut, 1974). Actualmente, se divide a estas técnicas en dos subgrupos: unas se aplican directamente al espacio real y utilizan las coordenadas naturales que lo describen (más en la línea de las ideas originales de Kadanoff) y otras se aplican en el "k−espacio", y son más usadas en teoría cuántica de campos, donde utilizan como variables de descripción no las coordenadas, sino los momentos (k). Cuando utilicemos las primeras hablarremos de *renormalización en el espacio real*.

El *grupo de renormalización* (GR) tiene como propósito calcular exactamente los exponentes que caracterizan una transición crítica, α y β, por ejemplo. Sería también interesante poder obtener la función de partición Z a cualquier temperatura, aunque esto no es en general posible. Sin embargo, un caso especial, el modelo de Ising en una dimensión, puede ser totalmente resuelto utilizando las herramientas que el GR proporciona. Consideremos la función de partición para el caso unidimensional en ausencia de campo externo (esto es, B = 0)

\[
Z = \sum_{\{s_i\}} \exp \left[ K \sum_{<ij>} s_i s_j \right]
\]

donde la suma se realiza sobre próximos vecinos y \(K \equiv -\beta J/2\), llamada constante de acoplamiento. Si desarrollamos la suma

\[
Z = \sum_{\{s_i\}} \exp[K(s_1 s_2 + s_2 s_3 + s_3 s_4 + s_4 s_5 + \ldots)]
\]

y escribimos

\[
Z = \sum \ldots e^{K(s_1 s_2 + s_2 s_3)} e^{K(s_3 s_4 + s_4 s_5)} \ldots
\]

s_i aparece únicamente en la primera exponencial. Sus dos estados posibles son ±1, así que podemos efectuar la suma sobre este espin en concreto:
Figura 7.5: Ilustración de cómo se produce la reducción del número de espines en 1 dimensión según las ideas del grupo de renormalización.

\[ Z = \sum \ldots \left[ e^{K(s_1 + s_3)} + e^{-K(s_1 + s_3)} \right] e^{K(s_2 + s_4)} \ldots \]

Podemos seguir sumando sobre todos los espines pares, y obtener

\[ Z = \sum \ldots \left[ e^{K(s_1 + s_3)} + e^{-K(s_1 + s_3)} \right] \left[ e^{K(s_2 + s_4)} + e^{-K(s_2 + s_4)} \right] \ldots \]

La suma queda ahora planteada únicamente sobre espines impares. La idea siguiente, ya propia del grupo de renormalización, es tratar de encontrar un nuevo valor para la constante de acoplamiento, \( K' \), y una función \( f \) tales que

\[ e^{K(s_1 + s_3)} + e^{-K(s_1 + s_3)} = f(K') e^{K'(s_1 + s_3)} \]

para todos los valores posibles de \( s_1 \) y \( s_3 \). \( f(K) \) es independiente de los espines. Consideremos los casos particulares \( s_1 = s_3 = 1 \) y \( s_1 = -s_3 = 1 \) (los otros dos son análogos). Entonces

\[ e^{2K} + e^{-2K} = f(K) e^{K'} \]

\[ 2 = f(K) e^{-K'} \]

Dividiendo las dos ecuaciones obtenemos

\[ K' = \frac{1}{2} \ln \left[ \cosh (2K) \right] \]

y sustituyendo en la segunda

\[ f(K) = 2 \cosh^{\frac{1}{2}}(2K) \]

Así que ahora podemos escribir la función de partición como

\[ Z = \sum \ldots f(K) e^{K'(s_2 + s_4)} \ldots = f(K')^\frac{N}{2} \sum e^{K'(s_2 + s_4 + s_6 + \ldots)} \]

Esta es la misma función de partición anterior si hubiésemos tenido solamente \( N/2 \) espines y la constante de acoplamiento fuese \( K' \). Así obtenemos la relación

\[ Z(N, K) = f(K')^\frac{N}{2} Z \left( \frac{N}{2}, K' \right) \] (7.2.1)
La ecuación de recurrencia para $K$ proporciona únicamente dos valores triviales para el punto fijo ($K = K' = K^*$):

$$K_1^* = 0, \quad K_2^* = \infty$$

Si recordamos que $K \propto T^{-1}$, vemos que el modelo de Ising en una dimensión no presenta transición crítica para ningún valor finito de la temperatura. La función de partición puede ser calculada en este caso, sin embargo, con notable precisión. Recuérdese que la energía libre $F$ en un sistema termodinámico es una función directamente proporcional al tamaño del sistema y, excepto una constante, igual al logaritmo de la función de partición. Así,

$$\ln Z = N\zeta$$

donde $\zeta$ depende de $K$ pero no de $N$. Por sustitución en 7.2.1,

$$\zeta(K) = \frac{1}{2} \ln [f(K)] + \frac{1}{2} \zeta(K')$$

Invirtiendo la relación y sustituyendo $f(K)$ por su valor,

$$\zeta(K') = 2 \zeta(K) - \ln \left[ 2 \cosh^{\frac{1}{2}}(2K) \right]$$

Esta ecuación, junto con

$$K = \frac{1}{2} \cosh^{-1}(e^{2K'})$$

son los resultados básicos del grupo de renormalización. Son ecuaciones de recurrencia que permiten calcular el punto fijo $K^*$ (trivial en este caso) y la función de partición $\zeta$ (excepto una constante) con precisión arbitraria. Ahora, dado cualquier valor de $K$, la ecuación para $\zeta$ proporciona el valor de la función de partición. La iteración sucesiva alternada de las dos ecuaciones define además un "flujo" en el espacio ($\zeta, K$), que se mueve hacia un punto fijo, hacia el punto crítico invariante por la transformación.
Veamos cual sería la forma de proceder. Si $K'$ es muy pequeña (digamos $K' \sim 0.01$), eso implica una temperatura elevada, y una gran agitación térmica en el sistema. La interacción entre los espines es despreciable y en ese caso la función de partición corresponde al número de estados diferentes que el sistema puede presentar:

$$Z \sim 2^N$$

mientras que

$$\zeta(0.01) \sim \ln(2)$$

Ahora calculamos $K$ partiendo de la ecuación 7.2.3, $K = 0.100334$. Sustituyendo en la expresión 7.2.2 obtenemos $\zeta = 0.698147$. Repetimos ahora el proceso con $K' = 0.100334$, que proporciona $\zeta = 0.745814$, $y$ de nuevo $K' = 0.327447$. Así podríamos seguir indefinidamente. Las ecuaciones del grupo de renormalización llevan a $K'$ y a $\zeta$ a los valores que les corresponden en el punto fijo. Como sólo $K' = 0, \infty$ son valores invariantes de las ecuaciones, acabaríamos en $K' = \infty$, dado que la ecuación utilizada provoca un flujo desde 0 hacia $\infty$. Si realizásemos las iteraciones aislando $K'$ en lugar de $K$, el flujo nos llevaría en sentido contrario, hacia $K' = 0$.

Para cada valor de $K$ que hayamos ido colocando en las ecuaciones, correspondiente a una cierta temperatura, habremos obtenido el valor de $\zeta$ asociado. Estos valores de $\zeta$, aunque no corresponden al punto fijo, proporcionan el valor numérico de la función de partición con una precisión $10^{-6}$ respecto de su valor exacto. Este valor se calcula mediante la solución exacta de Ising,

$$Z = \left[2 \cosh \left( \frac{J}{\kappa B T} \right) \right]^N$$

7.1.4 **Ising en 2 dimensiones: Teoría de Campo Medio**

A diferencia del modelo de Ising en una dimensión, el modelo en dos dimensiones sí presenta transición de fase. No obstante, su solución tardó 20 años en llegar (respecto a la del modelo en una dimensión, lo cual debe proporcionar una idea de la dificultad que entraña el aumento de dimensión), y fue dada por Onsager en 1944, quien halló que el valor del umbral de transición era

$$K_c = \frac{1}{2} \sinh^{-1}(1) = 0.4069$$

No es sencillo derivar la solución exacta de Onsager, así que trataremos el modelo en dos dimensiones de forma aproximada (en esta sección) y daremos una guía para el posible tratamiento numérico (más adelante). Comenzamos introduciendo las teorías de campo medio y aplicándolas en particular al caso que nos ocupa.

La idea básica de las teorías de campo medio consiste en eliminar los términos interactivos a nivel local, y que dependerán de cada punto en el sistema, para sustituirlas por cantidades medias que consideren la influencia global de toda la colectividad sobre cada uno de los elementos que la forman. De hecho, cualquier teoría de campo medio es equivalente a considerar que la interacción entre los componentes del sistema tiene alcance infinito (cuando en la mayoría de los casos simplemente se estará considerando que la interacción es local y a primeros vecinos). De hecho, durante mucho tiempo se pensó que la teoría de campo medio era exacta en esencia, debido a poderosos argumentos de L. D. Landau (1908-1968). Ahora se sabe que esto no es así, excepto para sistemas que efectivamente tienen alcance infinito, o cuando la dimensionalidad de estos es suficientemente grande. La técnica del campo medio fue ideada por P. E. Weiss (1865-1940) como
teoría del magnetismo, y durante mucho tiempo fue la única que se tenía para las transiciones de fase.

La teoría de campo medio es la aproximación más simple a un modelo de modelos, o metamodelo llamado de Ginzburg-Landau, el cual contiene toda la descripción de la física que acontece en una transición de fase continua.

Consideremos el modelo de Ising sometido a un campo externo uniforme \( B \). Calculemos el promedio térmico de un cierto espín \( s_i \):

\[
< s_i > = \frac{\exp \left[ -\beta \left( \sum_j s_j J_{ij} - B \right) \right] - \exp \left[ \beta \left( \sum_j s_j J_{ij} - B \right) \right]}{\exp \left[ -\beta \left( \sum_j s_j J_{ij} - B \right) \right] + \exp \left[ -\beta \left( \sum_j s_j J_{ij} - B \right) \right]}
\]

\[
= -\tanh \left[ \beta \left( \sum_j J_{ij} s_j - B \right) \right]
\]

El numerador pesa la probabilidad de que \( s_i \) sea positivo o negativo, y el denominador normaliza el valor resultante (la función "tangente hiperbólica" está comprendida entre -1 y +1). Ahora deberíamos continuar el proceso de cálculo del promedio térmico considerando los valores que los vecinos \( s_j \) irán tomando, promediar y considerar de nuevo los vecinos de los \( s_j \), ... A fin de romper esta serie infinita se toma una aproximación simple. Escribamos

\[
H_i = \left( \sum_j J_{ij} s_j - B \right) s_i
\]

La cantidad entre paréntesis actúa simplemente como un término efectivo para la interacción del espín \( s_i \). Simplifiquemos el término de interacción considerando únicamente el promedio espacial de los valores de \( s_j \),

\[
< s_j > = \frac{1}{N} \sum_j s_j
\]

con lo cual

\[
H_i = \left( \sum_j J_{ij} < s_j > - B \right) s_i
\]

Definamos

\[
\sum_j J_{ij} \equiv - J_0 q
\]

donde \( q \) es el número de coordinación, de próximos vecinos, y tendremos

\[
H = \sum_i H_i = - \sum_i s_i [q J_0 < s_j > + B]
\]

El promedio térmico sobre \( s_i \) queda ahora

\footnote{El factor 1/2 que aparece en el término de interacción en el hamiltoniano puede ser reabsorbido en la definición de la constante \( J_{ij} \).}

\footnote{Esta sustitución es la que da a la teoría de campo medio su nombre.}
Figura 7.7: Solución gráfica de la ecuación \( m = \tanh (\beta qJ_0 m) \). Cuando \( \beta qJ_0 < 1 \), la única solución es \( m = 0 \). Cuando \( \beta qJ_0 > 1 \), existe otra solución no nula, dada por la intersección de las dos funciones \( y = m \), \( y = \tanh (\beta qJ_0 m) \).

\[ < s_i > = \tanh [\beta (qJ_0 < s_j > + B)] \]

o en términos de la magnetización,

\[ m = \tanh [\beta (qJ_0 m + B)] \]

ya que \( m \equiv < s_i >= < s_j > \), dado que el sistema es invariante por traslación y únicamente consideramos promedios. Esta ecuación trascendente puede resolverse numéricamente o gráficamente. Consideremos el caso \( B = 0 \). La función tangente hiperbólica tiene pendiente menor que 1 cuando el coeficiente \( (\beta qJ_0) \) que multiplica a la variable \( (m) \) es menor que 1. Esto implica que, cuando \( \beta qJ_0 < 1 \), la única solución de la ecuación 7.2.3 es \( m = 0 \), el origen de coordenadas, con lo cual, si la magnetización del sistema es nula, estamos en la zona de \( T > T_c \). Para \( \beta qJ_0 > 1 \) existe una solución para \( m \) diferente de cero (por tanto en la región \( T < T_c \)) con lo cual

\[ \beta qJ_0 = 1 \]

define el umbral crítico de transición.

La teoría de campo medio predice pues la existencia de transición para el modelo de Ising en \( d \) dimensiones (obsérvese que aún no se ha hecho ninguna suposición sobre este particular), entre las cuales se halla \( d = 1 \), aunque hemos visto anteriormente (mediante una derivación exacta) que tal transición no existe. Para \( d = 2 \) tenemos \( q = 4 \), y la teoría de campo medio predice una transición en \( \beta_c = 0.25/J_0 \), a comparar con el valor exacto \( \beta_c = 0.4407/J_0 \). En \( d = 3 \), la predicción de campo medio es \( \beta_c = 0.133/J_0 \), y el valor numérico más aproximado \( \beta_c = 0.222/J_0 \). El campo medio siempre sobreestima el valor de \( T_c \), y esta es una propiedad general de la teoría.

Evaluemos a continuación los exponentes que esta aproximación predice. Reescribamos la ecuación para la magnetización usando variables reducidas,

\[ t \equiv \frac{T - T_c}{T_c}, \quad b \equiv \beta B \]

\[ m = \tanh \left( \frac{m}{1 + t} + b \right) \]  

(7.2.4)
Donde se utiliza que $\beta Q J_0 = 1$. Examinemos la magnetización espontánea ($m$) en ausencia de campo ($B = 0$). El argumento de la función tanh es pequeño, para $T \approx T_c$, y podemos considerar un desarrollo de la forma

$$\tanh (x) = x - \frac{1}{3} x^3 + O(x^5)$$

Así que

$$m \approx \frac{m}{1 + t} - \frac{1}{3} \left( \frac{m}{1 - t} \right)^3$$

La solución no nula corresponde a

$$m^2 = -3t(1 - t)^2$$

y por tanto, cuando $|t| \approx 0$, $m \propto |t|^{1/2}$, lo cual proporciona el valor $\beta = 1/2$ para el exponente que caracteriza el parámetro de orden.

Calculemos ahora la susceptibilidad térmica para campo cero. Consideremos un desarrollo de $\tanh(x)$ a primer orden para $T > T_c$ (suficiente, ya que impondremos $B = 0$, y $m$ es nula) para obtener

$$m = \frac{m}{1 + t} + b$$

y la susceptibilidad $\chi$

$$\chi \equiv \beta \left. \frac{\partial m}{\partial b} \right|_{b=0} \propto \frac{1}{t}$$

lo cual proporciona un segundo exponente, $\gamma = 1$. Para $T < T_c$ debemos considerar también el término de tercer orden en el desarrollo, y sustituir el valor de $m$, que ahora no es nulo, para obtener $\gamma' = 1$, también.

La magnetización que resulta justo en el punto crítico cuando se aplica un campo $B$ se obtiene desarrollando a tercer orden la ecuación 7.2.4,

$$m = m + b - \frac{1}{3} m^3 - \frac{1}{3} b^3 + O(m^5, b^5)$$

con lo cual, para $m$ y $b$ pequeños,

$$b \propto m^3$$

lo cual implica $\delta = 3$.

Con algo más de esfuerzo es posible encontrar $\alpha = \alpha' = 0$ (divergencia del calor específico), $\nu = 1/2$ (divergencia de la longitud de correlación) y $\eta = 0$ (divergencia de la función conexa de dos puntos en $T = T_c$).

El problema que presenta la teoría de campo medio, y que hace que sus estimaciones, tanto de los umbrales de transición como de los exponentes críticos, estén alejadas de los valores reales, es precisamente la suposición de la uniformidad del parámetro de orden cerca de la transición. Localmente existen fuertes correlaciones en el sistema que alejan los pequeños (o grandes) grupos correlacionados del comportamiento de la media. La consideración de las fluctuaciones en el punto crítico y de las correlaciones a gran escala que allí aparecen vino de la mano del modelo de Ginzburg-Landau, que describiremos brevemente en la próxima sección.

Acabamos con un resumen de los exponentes críticos para la teoría de campo medio y para el modelo de Ising en 2 y 3 dimensiones. Los del modelo 2 dimensional son exactos (dados por
la solución de Onsager), mientras que los del modelo 3-dimensional son aproximados. De ahí que estén acompañados de un error. Compárese estos últimos con los exponentes obtenidos para un sistema real, una pieza de hierro, de forma experimental.

<table>
<thead>
<tr>
<th>Exponente</th>
<th>C. Medio</th>
<th>Ising $d = 2$</th>
<th>Ising $d = 3$</th>
<th>Fe</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>0</td>
<td>0</td>
<td>0.119 ± .006</td>
<td>-0.03 ± .12</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1/2</td>
<td>1/8</td>
<td>0.326 ± .004</td>
<td>0.37 ± .01</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>1</td>
<td>7/4</td>
<td>1.239 ± .003</td>
<td>1.33 ± .015</td>
</tr>
<tr>
<td>$\delta$</td>
<td>3</td>
<td>15</td>
<td>4.80 ± .05</td>
<td>4.3 ± 1</td>
</tr>
<tr>
<td>$\eta$</td>
<td>0</td>
<td>1/4</td>
<td>0.024 ± .007</td>
<td>0.07 ± .04</td>
</tr>
<tr>
<td>$\nu$</td>
<td>1/2</td>
<td>1</td>
<td>0.627 ± .002</td>
<td>0.69 ± .02</td>
</tr>
</tbody>
</table>

### 7.1.5 El modelo de Ginzburg-Landau

Esta sección y la siguiente pueden resultar excesivamente áridas para el lector que no esté familiarizado con los métodos de la mecánica estadística e incluso de la teoría de transiciones críticas. Con ellas se pretende dar una explicación analítica a estas transiciones, pero si el lector lo cree conveniente puede ignorarlas y saltar sin mayor problema a la sección 7.2.8.

Describiremos brevemente en qué consiste este metamodelo, y cómo incorpora detalles que no estaban considerados en las teorías de campo medio. La esencia del modelo es la introducción de las fluctuaciones en el parámetro de orden. Se hace explícita en este modelo la dependencia de la dimensión $d$ del espacio de trabajo y también de la dimensionalidad $D$ del parámetro de orden, mientras que los detalles topológicos de la red, por ejemplo, no aparecen como elementos relevantes en la transición.

El modelo de Landau-Ginzburg (LG) considera que el parámetro de orden es un campo continuo de $D$ componentes, $\Phi(x)$: una magnitud vectorial que depende (en principio) de cada punto del espacio. De esta forma será capaz de considerar las fluctuaciones locales. Vayamos introduciendo modificaciones que nos conducirán finalmente al modelo de LG. Consideremos en primer lugar un problema de magnetización en dimensión $d$ arbitraria. El parámetro de orden $\Phi$ será la magnetización local media, y entenderemos que trabajamos con un volumen de tamaño menor que la resolución que sea posible conseguir con un cierto aparato de medida, pero que es capaz de contener muchos átomos (en ocasiones se denomina a esta escala de trabajo “escala o aproximación mesoscópica”, entre macro y microscópica). En general $\Phi$ es un vector que puede apuntar en cualquier dirección. Si $\Phi$ no varía de punto a punto, entonces la energía del sistema sólo dependerá del módulo de $\Phi$, así que se podrá expresar la densidad hamiltoniana\(^7\) como

$$ h = h_0 + \frac{1}{2} \mu' |\Phi|^2 + \frac{1}{4} \lambda'(|\Phi|^2)^2 + \ldots $$

donde $\mu'$ y $\lambda'$ son parámetros a determinar. Si, de forma más general, $\Phi$ varía de punto a punto, aparecerán términos dependientes del módulo de su gradiente espacial, $|\nabla \Phi|^2$. Esta dependencia se puede incluir en $h_0$:

$$ h_0 = \frac{1}{2} \alpha^2 |\nabla \Phi|^2 + \ldots $$

\(^7\)Se llama a la expresión siguiente densidad porque aparecerá integrada sobre todas las configuraciones del sistema en la expresión de la probabilidad de cierto suceso.
donde $\alpha'$ es una constante real. Tomando los primeros términos de los desarrollos anteriores podemos expresar la energía $H$ de un microestado compatible con $\Phi(x)$ como

$$H = \int d^4 x \, h(x) = \int d^4 x \left[ \frac{1}{2} \alpha' |\nabla \Phi|^2 + \frac{1}{2} \mu' |\Phi|^2 + \frac{1}{4!} \lambda' (|\Phi|^2)^2 \right]$$

Debemos ahora sumar sobre todos los microestados compatibles con $\Phi(x)$. El número $N$ de tales microestados es la suma sobre todos los volúmenes para los que está definida la magnetización media $\bar{\Phi}$ de las posibles ordenaciones dentro de estos volúmenes que dan $\bar{\Phi}$ como resultado. Este número es

$$N = \int d^4 x \, \exp \left\{ w(|\Phi|^2) \right\}$$

Si desarrollamos $w$,

$$w(|\Phi|^2) = ct + \frac{1}{2} w_2 |\Phi|^2 + \frac{1}{4!} w_4 (|\Phi|^2)^2 + \ldots$$

Tomando sólo los primeros términos de este desarrollo se puede expresar la probabilidad de cierto campo de magnetización media, $\Phi(x)$ como

$$P[\Phi] \propto \exp \left\{ - \int d^4 x \left( \beta h - w \right) \right\} \propto \exp \left\{ - \int d^4 x \, \mathcal{H}_{LG} \right\}$$

donde $\mathcal{H}_{LG}$, el hamiltoniano de Landau-Ginzburg es

$$\mathcal{H}_{LG} = \frac{1}{2} \alpha^2 |\nabla \Phi|^2 + \frac{1}{2} \mu^2 |\Phi|^2 + \frac{1}{4!} \lambda (|\Phi|^2)^2$$

$$\alpha \equiv \sqrt{\beta \alpha'}; \quad \mu^2 \equiv \beta \mu'; \quad \lambda \equiv \beta \lambda' - w_2$$

Podemos además generalizar para incluir un posible campo externo, el cual dará lugar a un incremento $-B \Phi$ en la densidad de energía,

$$\mathcal{H}_{LG} = \frac{1}{2} \alpha^2 |\nabla \Phi|^2 + \frac{1}{2} \mu^2 |\Phi|^2 + \frac{1}{4!} \lambda (|\Phi|^2)^2 - \beta B \Phi$$

El funcional de probabilidad $P[\Phi]$ y la densidad hamiltoniana anterior definen el modelo de Ginzburg-Landau. Se observa la dependencia de tres parámetros fenomenológicos: $\alpha$, una longitud característica que determina la importancia del término gradiante; $\mu^2$, cuya variación conduce a la transición crítica. y $\lambda$ (que debe ser positivo para que $P$ sea normalizable).

A través de este modelo se puede ver que la clase de universalidad de un sistema depende de su dimensión, $d$, y de la dimensión del parámetro de orden, $D$, pero, además, también de los grupos de simetría bajo los que $\mathcal{H}_{LG}$ sea invariante. Por ejemplo, la introducción de un término de la forma

$$\sum_{a=1}^{D} d_a^4$$

que no es invariante cuando $\Phi$ es rotado en su espacio $D$-dimensional podría provocar un cambio en alguno de los exponentes críticos.
7.1.6 La teoría de Landau

La teoría de Landau para las transiciones críticas puede ser derivada de la densidad hamiltoniana de Ginzburg-Landau. Es suficiente con considerar en la función de partición de este modelo,

\[ Z_{LG} = \int \mathcal{D}\Phi \exp\left\{-\int d^4 x \mathcal{H}_{LG}(\Phi)\right\} \]

el valor de \( \Phi, \Phi_0 \), que minimiza la integral en la exponencial, y suponer que la integral sobre \( \mathcal{D}\Phi \) está totalmente dominada por el valor máximo. Es decir:

\[ Z_{LG} \approx c \varepsilon \ Z_L, \quad Z_L \equiv \exp\left\{-\int d^4 x \mathcal{H}_{LG}(\Phi_0)\right\} \]

El caso \( \Phi = \varepsilon \) (esto implica que \( \nabla\Phi = 0 \)) hace la integral mínima, así que ya podemos prescindir directamente de este término. Además, si suponemos que la dirección de \( \Phi \) está determinada por \( \textbf{B} \), entonces podemos fijarnos únicamente en el módulo, \( \phi = |\Phi| \), con lo cual

\[ \int d^4 x \mathcal{H}_{LG}(\Phi_0) = \frac{1}{2} \mu^2 \phi_0^2 + \frac{1}{4!} \lambda \phi_0^4 \]

En ausencia de campo externo podemos escribir la energía libre del sistema como

\[ f = f_0 + a_2 m^2 + a_4 m^4 \]

Consideremos la energía libre como función de \( a_2 \). Si \( a_2 > 0 \), resulta que el mínimo de \( f \) está en \( m = 0 \) (fase paramagnética). Si \( a_2 < 0 \), el mínimo de \( f \) se da para un valor finito de \( m, m_0 \) (de hecho, de forma simétrica, para \( \pm m_0 \), lo que refleja la rotura de simetría para privilegiar una u otra orientación), con lo cual estamos en una fase ferromagnética. \( a_2 = 0 \) corresponde a la temperatura crítica a la cual aparece una magnetización espontánea.

Escribamos \( a_2 = a_2^t \) (es de nuevo la temperatura reducida). Como resulta que la magnetización se anula de forma continua, estamos ante una transición crítica, y podemos calcular los exponentes que la caracterizan.
La magnetización de equilibrio corresponde al mínimo de $f$:

$$\frac{\partial f}{\partial m} = 2\tilde{a}_2 tm + 4a_4 m^3 = 0$$

Cuando estamos en la zona $t < 0$, resulta $m \propto (-t)^{1/2}$, y por tanto obtenemos un primer exponente, $\beta = 1/2$. Obtendremos $\alpha$ derivando dos veces $f$ respecto de la temperatura ($C_B \propto \partial^2 f/\partial t^2$). Observaremos que $C_B \rightarrow ct$ cuando $t \rightarrow 0^-$, y que $C_B = 0$ cuando $t > 0$, con lo cual concluimos que existe una discontinuidad finita en el punto de transición, y por tanto $\alpha = 0$.

Añadimos un campo externo $h$ para calcular los exponentes $\gamma$ y $\delta$:

$$f = f_0 - hm + \tilde{a}_2 tm^2 + a_4 m^4$$

En el equilibrio,

$$\frac{\partial f}{\partial m} = -h + 2\tilde{a}_2 tm + 4a_4 m^3 = 0$$

y sobre la isotermica critica $t = 0$, con lo cual $m^3 \propto h$ y resulta $\delta = 3$. También se obtiene $\gamma = 1$, y los restantes exponentes idénticos a los de la teoría de campo medio que ya hemos visto. La teoría de Landau es, en consecuencia, una teoría de campo medio, y no resultaría complicado establecer la correspondencia exacta que existe entre ambas.

### 7.1.7 Ising en 2 dimensiones: Renormalización en el Espacio Real

Las técnicas de renormalización en el espacio real sólo se aplican a modelos realizados sobre una red que posea ciertas características geométricas: debe presentar simetría discreta de escala. Esto significa que, si la miramos a diferentes resoluciones (con un aumento $b$ discreto) debemos observar el mismo tipo de geometría que en la red original. Si realizamos bloques iguales con las celdas unidad, y reemplazamos estos bloques por nuevas celdas, ahora a otra escala (superceldas) debemos obtener una red idéntica a la de partida, excepto por el incremento producido en el parámetro $a$ de la red (distancia entre celdas), ya que

$$a \rightarrow a' \equiv ba$$

El proceso de renormalización en la red acaba cuando todas las magnitudes han podido ser redefinidas de acuerdo con el factor $b$. La sección 7.3.4 muestra dos criterios de renormalización, uno para una red triangular y otro para una red cuadrada. En general, si agrupamos $p$ celdas para formar una supercelda, el factor de escala $b$ es $b = \sqrt[p]{p}$, donde $d$ es la dimensión del espacio de trabajo.

Veamos cómo definiríamos en el caso del modelo de Ising en dos dimensiones unas variables sobre los bloques, y cómo la invarianza de su distribución (por ejemplo) nos señala el lugar exacto del punto crítico del sistema (punto fijo del proceso de renormalización por bloques).

Fijemos ideas considerando un modelo las variables del cual sean espines, definidos en cada celda de la red. Supongamos que agrupamos $p$ celdas en un bloque $k$, para el cual definiremos una nueva variable, $\sigma_k^{(1)}$ (variable de bloque) en función de los espines $s_i$ de las $p$ celdas que contiene. Si $S_k$ es el conjunto de $p$ celdas en el bloque $k$, tendríamos en general que

$$\sigma_k^{(1)} = f(\{s_i\}), \quad \forall i \in S_k$$

Tras la aplicación de esta transformación llegaremos a la misma red original, excepto por el hecho de que tendremos $N/p$ celdas, en lugar de las $N$ de partida. De forma recursiva es posible seguir
definiendo variables para los bloques en cada nueva escala, que vendrá determinada por factores \(b, b^2, b^3, \ldots, b^n, b^{n+1}, \ldots\) según

\[
\sigma_k^{(n+1)} = f(\{\sigma_k^{(n)}\})
\]

Existen diversas formas de escoger la función \(f(x)\), que en cualquier caso debe ser representativa de los \(p\) espines \(s_i\) que el bloque \(k\) contiene. Un criterio habitual es la llamada regla de la mayoría:

\[
\sigma_k^{(n+1)} = \begin{cases} 
\text{sgn} \left( \sum \sigma_k^{(n)} \right) & \text{si } \sum \sigma_k^{(n)} \neq 0 \\
+1 \quad 0 \quad -1 & \text{si } \sum \sigma_k^{(n)} = 0
\end{cases}
\]

La variable \(\sigma_k^{(n+1)}\) vale +1 si la suma en la iteración anterior es positiva (mayoría de espines +1) o bien -1, si la suma es negativa. En el caso de que la suma de los espines en el bloque sea 0 (tantos en un estado como en el opuesto) entonces se le asigna aleatoriamente uno de los dos valores.

Otro criterio adecuado en ocasiones consiste en escoger arbitrariamente el valor de uno de los espines del bloque de la iteración anterior ("diezmación")

\[
\sigma_k^{(n+1)} = \sigma_i^{(n)}, \quad i \in S_k
\]

Aunque estadísticamente los más abundantes siempre estarán más representados cuando la muestra sea suficientemente amplia, en ocasiones este criterio da resultados notablemente peores que los obtenidos con la regla de la mayoría, por ejemplo.

Una definición adecuada en ocasiones puede ser la de considerar el valor medio de los elementos en cada bloque:

\[
\sigma_k^{(n+1)} = \frac{1}{p} \sum_i \sigma_k^{(n)}, \quad i \in S_k
\]

Si utilizamos en el caso de tener sólo los estados ±1 esta definición, está claro que obtendremos nuevos estados que no aparecían en la red original, lo cual podríamos pensar que es un problema, en principio. Sin embargo, este criterio es capaz de agrupar todos los sistemas magnéticos, con valores arbitrarios para el estado de los espines, en el mismo punto crítico. Considerando la distribución de los valores de los \(\sigma_k^{(n+1)}\), obtendríamos una curva invariante, con un pico en el valor +1 y otro en el valor -1, que no depende del espín original del sistema. Esta distribución invariante refleja precisamente la pertenencia de todos estos sistemas a la misma clase de universalidad.

Otras reglas adecuadas, que reflejan el estado de los espines y la invariancia de la red pueden ser utilizadas. No existe un criterio universal, y únicamente debemos pensar en el estado de los elementos del sistema y en las propiedades que queremos preservar.

### 7.1.8 Simulación del modelo de Ising

Daremos finalmente unas ideas sobre la implementación de un algoritmo que permite simular en un ordenador la dinámica del modelo de Ising. Es interesante constatar que, a pesar de la gran dificultad que entraña el tratamiento analítico de las transiciones críticas, algunas simulaciones sencillas nos acercan notablemente a los resultados exactos. En este caso, el problema se reduce

---

8. No únicamente aplicaremos este tipo de reglas a espines con estados discretos y tanto positivos como negativos. Podemos tener todo un continuo de estados, y en particular podrían ser todos positivos. En cada caso, la definición debe de ajustarse al sistema.
Figura 7.9: Renormalización por bloques para una cierta configuración de percolación antes, cerca y después del punto crítico. Se ha utilizado la regla de la mayoría. Se representa en cada caso una porción de tamaño $25 \times 25$ de redes de lado 400, 200, 100 y 50, de arriba a abajo. La probabilidad de ocupación, de izquierda a derecha, es $p = 0.8$, 0.59 y 0.2.
a considerar, para cada uno de los espines del sistema tratado, la influencia que sobre él ejercen sus vecinos, lo que se denomina su campo local. Considerando además la temperatura del sistema podemos definir una probabilidad para cada espín de hallarse en el estado +1 o en el -1, con lo cual, mediante algún método de generación de números aleatorios, y escogiendo espines al azar para su actualización, podemos observar la relajación del sistema hasta el estado de equilibrio (estadístico) y evaluar las fluctuaciones en este estado.

Consideremos de nuevo el hamiltoniano del modelo de Ising, y supongamos que representa electrones, los estados de espín de los cuales únicamente pueden ser $s_i = +\frac{1}{2}$ o $-\frac{1}{2}$,

$$H = -J \sum_{<i,j>} s_i s_j + \gamma B \sum_i s_i$$

y ahora nos fijaremos en el valor exacto de los parámetros y en su dimensión, para cuantificar exactamente los resultados. El valor del campo magnético se dará en $T$ (teslas), $\gamma = e\hbar/m_e$, donde $e = 1.602.10^{-19} C$ es la carga del electrón, $\hbar = h/2\pi$, con la constante de Planck $h = 6,626.10^{-34} J.s$ y $m_e = 9,109.10^{-31} kg$ es la masa del electrón. Para $J$ se puede tomar un valor del orden de $J = 17. k_B (J/k_B = 17K)$, lo cual concuerda bastante bien con la interacción estimada en casos reales. $k_B = 1.381.10^{-23} J/K$ es la constante de Boltzmann. Los valores exactos de los parámetros no son necesarios para calcular las probabilidades de cada estado de espín, la magnetización media, las correlaciones entre estados o la información conjunta, magnitudes todas ellas que veremos seguidamente.

La energía del espín $i$-ésimo es

$$E_i(s_i) = -\left(J \sum_j s_j - \gamma B\right) s_i = H_i s_i$$

donde la suma $j$ se extiende a los vecinos próximos (cuatro si $d = 2$, seis si $d = 3$). La probabilidad de que la celda $i$ tenga componente de espín $s_i$ es proporcional al factor de Boltzmann,

$$p(s_i) \propto \exp\left\{ -E_i(s_i) \right\}/k_B T$$

a temperatura $T$ (en Kelvin).

Modelizaremos el sistema comenzando con una red de $N^d$ elementos con estados de espín asignados al azar. El campo de interacción

$$H_i = J \sum_j s_j - \gamma B$$

debe ser evaluado para cada celda, utilizando condiciones periódicas de contorno (los vecinos por la derecha de las celdas $N$ son las celdas 1, y viceversa) para minimizar el efecto de la frontera. Se evalúan entonces las probabilidades relativas,

$$P_i(s_i) = \frac{\exp\left\{ H_i s_i \right\}}{\sum_{s_i} \exp\left\{ H_i s_i \right\}}$$

El intervalo $[0,1]$ se divide entonces en dos subintervalos (correspondientes a $s_i = 1/2$, $s_i = -1/2$) con longitudes iguales a las probabilidades de cada estado de espín. Se elige un número aleatorio entre 0 y 1 y se asigna a la celda el valor del espín determinado por este número. El proceso se repite hasta que se alcanza el equilibrio termodinámico. Si actualizamos una celda en cada iteración, son necesarias unas $10N^d$ iteraciones para alcanzar el equilibrio. Si $T \approx T_c$, el número
Figura 7.10: Fluctuaciones en la magnetización media de los espines en el modelo de Ising en dos dimensiones cerca del punto crítico. El tamaño del sistema es $N = 90$.

de iteraciones requerido puede ser mucho mayor, debido a las enormes fluctuaciones que aparecen cerca de la transición crítica. Una vez alcanzado el equilibrio se pueden calcular todas las variables termodinámicas del sistema. Por ejemplo, $\langle m \rangle$, la magnetización media será simplemente

$$\langle m \rangle = \frac{1}{N^d} \sum_{i=1}^{N^d} s_i$$

y debemos obtener $\langle m \rangle \approx 0$ para $T > T_c$ y $\langle m \rangle \neq 0$ para $T < T_c$, con lo cual podemos estimar el punto de transición.

Si llamamos $\langle m_c \rangle$ a la media de las correlaciones entre vecinos,

$$\langle m_c \rangle = \frac{1}{dN^d} \sum_{\langle i,j \rangle} s_is_j$$

donde el factor $1/d$ descuenta las parejas repetidas, tenemos que la energía interna por celda será

$$\langle u \rangle = -\frac{Jd\langle m_c \rangle + \gamma B \langle m \rangle}{k_B}$$

d incluye ahora el hecho de que tenemos $N$ celdas, y cada una tendrá $2d$ vecinos, así que existen $dN$ parejas distintas de vecinos $\langle i,j \rangle$. Si realizamos además medidas a diferentes temperaturas, podremos calcular $\langle C_B \rangle$, el calor específico por celda a $B = cT$.

$$\langle C_B \rangle = \left( \frac{\partial \langle u \rangle}{\partial T} \right)_B$$

la entropía media por celda,

$$\langle s \rangle = \int_0^T \langle C_B(T') \rangle \frac{dT'}{T}$$

y la energía libre de Helmholtz por celda,

$$\langle f \rangle = \langle u \rangle - T \langle s \rangle$$

\footnote{Aunque podríamos haber cancelado este factor $d$ con el que aparece en la definición de la función $\langle m_c \rangle$, aquella es su definición correcta, así que hemos considerado oportuno explicitar la dependencia.}
La Información Conjunta

Existe otra medida interesante que puede ser realizada no sólo en este sistema, sino también en cualquier otro en el que podamos conocer los estados por los que pasan los elementos del sistema a lo largo del tiempo. Es la llamada información conjunta, \( I_{<i,j>} \), que se define como

\[
I_{<i,j>} = S_i + S_j - S_{i,j}
\]

donde \( S_i \), \( S_j \) son las entropías asociadas a los estados que \( i \) y \( j \) (vecinos mutuos) presentan a lo largo del tiempo, y \( S_{i,j} \) es la entropía asociada a los estados conjuntos de estos dos mismos elementos. En general,

\[
S_k = - \sum_e p_e \log p_e
\]

donde \( e \) indica suma sobre todos los estados que el elemento \( s_k \) presenta a lo largo del tiempo, y \( p_e \) representa la probabilidad de cada uno de estos estados. Para la entropía conjunta,

\[
S_{kl} = - \sum_{ee} p_{ee} \log p_{ee},
\]

donde \( p_{ee} \) es ahora la probabilidad de cada estado o configuración simultánea de la pareja de vecinos. La suma se extiende a todas las configuraciones que se han presentado a lo largo del tiempo.

Pongamos un ejemplo. Supongamos que hemos obtenido la siguiente secuencia para dos espines que son vecinos próximos:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

El espín \( s_1 \) ha sido hallado 6 veces en el estado +1, y 9 veces en el estado -1. Por tanto

\[
S_i = - \sum_e p_e \log p_e = - \frac{6}{15} \log \frac{6}{15} - \frac{9}{15} \log \frac{9}{15} \approx 0.2929
\]

Para el espín \( s_2 \), por otra parte, obtenemos

\[
S_j = - \frac{8}{15} \log \frac{8}{15} - \frac{7}{15} \log \frac{7}{15} \approx 0.3001
\]

La entropía conjunta \( S_{i,j} \) corresponde a las situaciones relativas en que los espines pueden ser hallados en cada paso de tiempo. Tenemos cuatro opciones en este caso: (1,1), (1,-1), (-1,1) y (-1,-1). Si miramos la serie, obtenemos 4 veces la primera combinación, 2 veces la segunda, 4 veces la tercera y 5 veces la última, lo cual conduce a

\[
S_{i,j} = - \frac{4}{15} \frac{4}{15} - \frac{2}{15} \frac{2}{15} - \frac{5}{15} \frac{5}{15} \approx 0.5819
\]

y finalmente, \( I_{<i,j>} = 0.0111 \) para esta serie particular. Para obtener una buena estadística necesitamos hacer promedios durante tiempos muy largos (en ocasiones del orden de \( 10^5 \) a \( 10^4 \) o más iteraciones), ya que sólo tomamos una pareja de elementos en cada paso de tiempo, y de esta seguimos la evolución, siempre de la misma, con lo cual una buena media significará en este caso un tiempo largo, y no un sistema grande.
La característica más interesante de esta medida es que presenta un máximo en el punto donde se produce la transición. La podemos interpretar como una medida de las correlaciones que se crean en el sistema, de la dependencia del estado de un elemento del que presente su vecino. La gráfica de la información conjunta para el modelo de Ising se representa en la figura 7.11.

7.2 Percolación

Imaginemos una porción de roca porosa. Supongamos que queremos saber en qué condiciones un fluido será capaz de penetrar en la roca por un extremo y reaparecer por el opuesto. Nuestro problema se limita a conocer si existe un camino en el interior de la roca que conecte dos caras opuestas. Si así ocurre, diremos que los poros, o los huecos, en la roca percolan en ella. Tomemos otro ejemplo. Consideremos una superficie aislante sobre la que aleatoriamente vamos depositando pequeños trozos de un cierto material conductor. Si sometemos esta superficie a una diferencia de potencial, está claro que no habrá paso de corriente hasta que las deposiciones metálicas consigan formar un camino conexo entre los extremos. Observemos que la conducción es nula antes de la aparición de dicho camino y siempre positiva después. Supongamos, por último, que deseamos establecer el tiempo de extinción natural de un fuego en un bosque. Este tiempo dependerá del tamaño de la agrupación conexa de árboles en la que el fuego se ha iniciado. Resulta ser que esta longitud temporal, además, no es una función linealmente creciente con la densidad de árboles, como en principio se podría suponer. Los grupos de árboles son bastante pequeños para densidades bajas, y cuando crece la densidad, después de un cierto umbral crítico, la conexión de todos los árboles del bosque se produce con asombrosa rapidez.

Los tres ejemplos descritos han sido estudiados a partir de modelos simplificados de percolación sobre todo tipo de redes y en muchas dimensiones diferentes: desde 1 hasta $\infty$, aunque los casos más difíciles siempre están comprendidos entre 3 y 5, ambas incluidas (4).

En el caso más sencillo, y probablemente también más ilustrativo, consideraremos una red cuadrada 2-dimensional, las celdas de la cual se irán llenando aleatoriamente con una cierta probabilidad $p$. Si pudiésemos estudiar una red infinita (a fin de evitar las desviaciones estadísticas que...
Figura 7.12: Celdas ocupadas en el modelo de la percolación en una red cuadrada para diversos valores de la probabilidad de ocupación. De izquierda a derecha, $p_c > p = 0.2$, $p_c \approx p = 0.59$, $p_c < p = 0.8$. Obsérvese la geometría del camino que conecta los extremos de la red.

Figura 7.13: Una posible cadena de celdas ocupadas y vacías para el modelo de la percolación en una dimensión. Se ha escogido $p = 0.6 < p_c$.

cualquier red finita contiene) veríamos que, para probabilidades $p$ pequeñas, por debajo de un cierto umbral crítico $p_c$ no existe ningún camino que conecte los dos extremos de la red. Si $p > p_c$, este camino existe siempre, y además, como las partículas se unen unas a otras muy rápidamente poco después de cruzar el umbral, tiende a ser bastante directo, recto. En cambio, cuando $p \approx p_c$, cuando el camino justo acaba de aparecer por primera vez, presenta muchas “curvas”. Concretamente, si fuésemos capaces de trabajar con redes infinitas, obtendríamos que la longitud de este camino se hace infinita para $p = p_c$. Véase la figura 7.12.

La magnitud $p_c$ representa para la percolación lo que la temperatura de Curie para los sistemas magnéticos. Es el umbral crítico, en el que el parámetro de orden se hace positivo (con derivada discontinua) y gran parte de las cantidades que describen el sistema se anulan o divergen de forma potencial, según un cierto exponente que las caracteriza. Esto es lo que veremos en las próximas secciones.

7.2.1 Solución exacta en una dimensión

Estudiaremos el modelo en una dimensión debido a que posee una solución exacta y a que, aunque muy simplificada, ilustra algunos aspectos de la percolación en dimensiones mayores.

Llamaremos agrupación (“cluster”) a un conjunto conexo de celdas ocupadas (recordemos que previamente, en función de una probabilidad $p$, estas celdas han sido definidas como llenas o vacías). Véase en la figura 7.13 un ejemplo con agrupaciones de diversos tamaños.

Definiremos el número de agrupación, $n_s$, como la probabilidad de que en nuestra cadena de longitud $L$ ($L \to \infty$) encontremos una agrupación de $s$ celdas llenas. Observemos que, según la definición, la probabilidad de tener $s$ celdas consecutivas ocupadas será $p^s$ (para sucesos estadísticamente independientes, como se está suponiendo) y además debemos pedir que los dos
extremos están vacíos (lo cual sucede, para cada extremo, con probabilidad \((1 - p)\)). Así,

\[
n_s = p^s (1 - p)^2
\]

\(n_s\) tiene el mismo valor que la probabilidad (en una cadena infinita) de que una celda determinada sea el extremo izquierdo (igual se podría tomar el derecho) final, vacío, de la agrupación. La probabilidad de que una determinada celda pertenezca a una agrupación de tamaño \(s\) será \(s n_s\), donde estamos suponiendo indistinguibilidad.

El umbral de percolación para el caso 1-dimensional resulta ser \(p_c = 1\). Si \(p < p_c = 1\), siempre quedará alguna celda vacía (en una red infinita) así que no existirá camino conexo entre los extremos. En una dimensión, pues, no podemos acceder a la zona \(p > p_c\). Trabajemos con lo que tenemos. La probabilidad de que una celda esté ocupada es \(p\), pero también coincide con la probabilidad de que pertenezca a una agrupación arbitraria. Por tanto,

\[
\sum_{r=1}^{\infty} n_s s = p (p < p_c)
\]

Esta suma también está restringida a \(p < p_c\), para dimensiones mayores, ya que las celdas pertenecientes a la agrupación máxima (que contiene el camino conexo) deben ser consideradas independientemente.

Calculemos cuál es el tamaño medio de las agrupaciones del sistema. La probabilidad de que una celda arbitraria (esté o no ocupada) pertenezca a una agrupación de tamaño \(s\) es \(n_s s\), mientras que la probabilidad de que pertenezca a cualquier agrupación finita es \(\sum_s n_s s\), así que

\[
w_s = \frac{n_s s}{\sum_s n_s s}
\]

es la probabilidad de que al escoger arbitrariamente una celda del sistema, la agrupación a la que esta pertenece tenga tamaño \(s\). El tamaño medio de las agrupaciones, \(S\), en el sistema será en consecuencia

\[
S = \sum_s w_s s = \sum_s \left( \frac{n_s s^2}{\sum_s n_s s} \right)
\]

La agrupación infinita está excluida de la suma.\(^{10}\) Calculemos explícitamente. El denominador anterior es simplemente \(p\), y el numerador, por la definición de \(n_s\),

\[(1 - p^2) \sum_s s^2 p^s = \ldots = (1 - p^2) \left( \frac{d}{dp} \right)^2 \sum_s p^s \]

donde se ha utilizado dos veces que

\[
\frac{p}{s} \frac{dp}{dp} = p
\]

lo cual permite calcular la última suma, ahora convertida en una progresión geométrica sencilla, para proporcionar

\(^{10}\)Existen algunas definiciones alternativas. Si considerásemos que son las celdas, y no las agrupaciones, las que se escogen con equiprobabilidad, entonces

\[
\hat{S} = \sum_s \frac{n_s s}{\sum_s n_s}
\]

por ejemplo.
\[
S = \frac{1 + p}{1 - p} \quad (p < p_c)
\]
Observemos que esta cantidad diverge cuando \( p \to p_c = 1 \). La divergencia se producirá igualmente en cualquier dimensión, aunque el umbral crítico no será el mismo.

Podemos definir la función de correlación \( g(r) \), como la probabilidad de que una celda situada a distancia \( r \) de una celda ocupada pertenezca a la misma agrupación. Sencillamente, el suceso "estar ocupada" debe darse entre la celda origen y la que está situada a distancia \( r \) sin interrupción:

\[
g(r) = p^r, \quad \forall p \quad \forall r
\]

Para \( p < p_c \), \( g(r) \) tiende a cero de forma exponencial cuando \( r \to \infty \)

\[
g(r) \propto \exp \left( -\frac{r}{\xi} \right)
\]

donde

\[
\xi = -\frac{1}{\ln(p)} \approx \frac{1}{(p_c - p)}
\]

cuando \( p \to p_c = 1 \), y utilizando \( \ln(1 - z) \approx -z \) para \( z \) pequeña. \( \xi \) es la longitud de correlación, también divergente en el punto crítico. Intuitivamente se deduce que \( \xi \) será proporcional al tamaño medio de las agrupaciones, \( S \), cerca del punto crítico:

\[
S \propto \xi \quad (p \to p_c)
\]

En general, \( \xi \) es proporcional al diámetro típico de las agrupaciones. Sólo en una dimensión se verifica la proporcionalidad anterior, que en el caso general será \( S \propto \xi^{\nu} \). En una dimensión, obtenemos \( \nu = 1 \).

De las relaciones anteriores, que dan resultados exactos en una dimensión, vemos que existen cantidades divergentes que pueden ser caracterizadas por simples potencias, como \((p_c - p)^{-1}\), cuando \( p \to p_c \).

### 7.2.2 Exponentes críticos

El modelo de percolación en una dimensión ilustra el comportamiento de algunas de las cantidades relevantes en los sistemas que experimentan esta transición crítica. Sin embargo, dado que no podemos acceder a la zona \( p > p_c \), cantidades tan importante como el parámetro de orden, por ejemplo, no pueden ser definidas. Tanto éste como los exponentes críticos más relevantes en los sistemas percolantes van a ser definidos en esta sección. Como guía, podemos tener siempre presente la imagen de la percolación en una red cuadrada de dos dimensiones.

¿Cómo se define el parámetro de orden? Sabemos que antes de \( p_c \) no existe camino conexo, mientras que éste está siempre presente cuando \( p \geq p_c \). La primera vez que aparece, son muy pocas las celdas que forman la agrupación mayor (que claramente será la que contiene el camino de percolación, el que atraviesa todo el sistema), mientras que, a medida que \( p \) crece, más y más celdas se unen a ella. Con esta idea en mente, definimos el parámetro de orden \( \Phi \) (con frecuencia llamado \( P \) o \( P_\infty \) para la percolación) como el número de celdas contenidas en la agrupación de percolación:

\[
\Phi = \frac{\# \text{ celdas en la agrupación de percolación}}{L^2 p}
\]
donde $L^2p$ es el número total de celdas ocupadas. Cuando $p \rightarrow p_c$, el parámetro de orden presenta una tendencia potencial,

$$\Phi \propto (p - p_c)^3$$

3 se puede interpretar como la forma en que la conectividad de la agrupación máxima tiende a cero.

En general, el tamaño medio de las agrupaciones divergerá cuando nos acerquemos al punto crítico por la derecha o por la izquierda (recordemos que la agrupación máxima queda eliminada de la suma sobre agrupaciones para $p \geq p_c$), según un exponente $\gamma$:

$$S(p) \propto |p - p_c|^{-\gamma}$$

La longitud de correlación, ya definida, sigue una ley

$$\xi(p) \propto L \propto |p - p_c|^{-\nu}$$

En general no será posible utilizar redes infinitas en las descripciones de nuestros sistemas, ya que éstas sólo pueden ser tratadas de forma analítica, y para las dimensiones más interesantes ($d = 2, 3$) es casi siempre imposible obtener resultados exactos (en ocasiones sí se puede hacer para $d = 2$). Por tanto, nos veremos con frecuencia obligados a realizar simulaciones numéricas. Si intentamos hallar alguno de los exponentes citados por simple simulación directa, veremos que es casi imposible obtener buenos resultados si trabajamos con redes que en principio nos parecerían razonablemente grandes ($L \sim 10^2 - 10^3$). Sin embargo, podemos aprovechar la proporcionalidad entre $\xi$ y $L$ para realizar el denominado análisis de tamaño finito. Si invertimos la última relación, obtenemos $|p - p_c| \propto L^{-1/\nu}$. Podemos sustituir ahora $|p - p_c|$ en alguna de las funciones anteriores, por ejemplo

$$\Phi(p = p_c) \propto L^{\beta/\nu}$$

y realizar un estudio de redes de tamaños diferentes, finitos, en el punto crítico. Una interpolación logarítmica permite obtener el cociente $\beta/\nu$ con una aproximación mucho mayor de lo que posibilitaría determinación independiente de los exponentes.

Definiremos dos exponentes críticos más, esta vez relacionados con la geometría de las configuraciones (de las agrupaciones) en el punto crítico, para $p = p_c$. Por ejemplo, ¿cual es la distribución de tamaños $s$ de las agrupaciones en el punto de transición? Esta distribución sigue también una ley potencial en $p = p_c$, que se desvía (de hecho solo una parte de la distribución se ajusta a ella) para $p \neq p_c$:

$$n_s \propto s^{-\sigma} e^{-cs}$$

con

$$c \propto |p - p_c|^{1/2}$$

lo cual define el exponente $\sigma$, que caracteriza el límite de validez de la ley potencial, ya que la función exponencial $e^{-cs}$ actúa como función de corte, dado que tiende más rápidamente a cero que la función potencial. Obsérvese que la expresión anterior se reduce a $n_s \propto s^{-\tau}$ para $p = p_c$. El último exponente que definiremos es la dimensión fractal de la agrupación máxima.

---

11El límite para el parámetro de orden sólo está definido por uno de los dos lados, ya que en el otro es siempre cero por definición. No distinguiremos si el límite se toma por la derecha o por la izquierda.
para \( p = p_c \). Si \( p < p_c \), la fractalidad no se extiende hasta el tamaño total del sistema, puesto que esta agrupación no lo ocupa, aunque existe autosimilaridad (hasta la longitud de correlación \( \xi \)). Si \( p > p_c \), la geometría experimenta un cambio hacia la homogeneidad: la agrupación máxima está "excesivamente llena", ya que para \( p > p_c \) muchas otras celdas ocupadas se conectan a ella rápidamente. Si \( M \) es la cantidad de celdas que forman la agrupación de percolación,

\[
M(l) \propto l^D
\]

donde \( l \) es la medida lineal de las cajas que recubren el sistema, y \( D \) la dimensión fractal correspondiente a la estructura (véase el capítulo 3).

De nuevo los exponentes definidos no son independientes. Se ha hallado que verifican las siguientes relaciones:

- \( \beta \sigma = \tau - 2 \)
- \( \gamma \sigma = 3 - \tau \)
- \( \frac{1}{\beta} = \sigma \nu \)
- \( D = d - \frac{\beta}{\nu} \)

La última relación de esta lista pertenece al grupo de las llamadas relaciones de hiperescala, debido a que contiene la dimensión espacial \( d \) del sistema. Así que en el caso de la percolación, como ya se había visto para las transiciones magnéticas, el conocimiento de dos exponentes es suficiente para dar el valor de todos los demás.

Detallamos a continuación el valor de los exponentes críticos para la percolación en 2 y 3 dimensiones. De nuevo, en 2 dimensiones son valores exactos (análíticos), y en 3 dimensiones son numéricos. Se dan también los umbrales de percolación para redes con diferente geometría en dimensión 2. Dada la dimensión del sistema, el umbral de percolación puede depender de la topología de la red, no así los exponentes críticos.

<table>
<thead>
<tr>
<th>( d )</th>
<th>( \beta )</th>
<th>( \gamma )</th>
<th>( \nu )</th>
<th>( \tau )</th>
<th>( \sigma )</th>
<th>( D )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>5/36</td>
<td>43/18</td>
<td>4/3</td>
<td>187/91</td>
<td>36/91</td>
<td>91/48</td>
</tr>
<tr>
<td>3</td>
<td>0.41</td>
<td>1.80</td>
<td>0.88</td>
<td>2.18</td>
<td>0.45</td>
<td>2.53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>cuadrada</th>
<th>cúbica</th>
<th>triangular</th>
<th>hexagonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_c )</td>
<td>0.592746</td>
<td>0.3116</td>
<td>0.50000</td>
<td>0.6962</td>
</tr>
</tbody>
</table>

### 7.2.3 Percolación en dos dimensiones: renormalización en el espacio real

La idea básica de la renormalización consiste en aprovechar la autosimilaridad que los sistemas presentan en una transición crítica. El límite para esta autosimilaridad viene dado por la longitud de correlación, \( \xi \). Para todo valor de \( p \), \( \xi \) representa la máxima longitud a la que el sistema repite estructura. Así que sólo cuando \( \xi \) diverge podemos aprovechar la repetición para plantear ecuaciones a distintas escalas de longitud.
En la renormalización en el espacio real, cada conjunto de celdas es reemplazado por una supercelda a la que, en principio, debería serle asignado un estado que fuese representativo de los estados que presentaban las celdas a las que sustituye. Cada una de estas superceldas tendrá una dimensión lineal b y sustituirá b^d celdas de la escala menor anterior. Esta renormalización de celdas requiere ciertas restricciones a fin de producir resultados interesantes. Por ejemplo, en general, la densidad de ocupación de las superceldas será \( p' \neq p \), siendo \( p \) la densidad de las celdas elementales. Unicamente para \( p = p_c \) se obtendrá \( p' = p = p_c \) (siempre suponiendo que hemos sabido tomar un criterio hábil de asignación de estados a la escala superior, ya veremos exactamente cómo), como resultado de la invariancia de escala en el punto crítico. Si en la red de celdas inicial teníamos

\[
\xi = C \left| \rho - p_c \right|^{-\nu}
\]

en la red renormalizada de superceldas tendremos

\[
\xi' = C \left| \rho' - p_c \right|^{-\nu}
\]

\((C = \text{constante})\). En nuestra nueva escala, \( \xi' \) será \( \xi' = \xi/b \), según el reescalamiento efectuado. Así pues,

\[
\frac{b}{\rho' - p_c} = \frac{b}{\rho - p_c}
\]

que es la ecuación básica de la renormalización en espacio real. Si tomamos el logaritmo a ambos lados,

\[
\frac{1}{\nu} = \frac{\log \left( \frac{\rho' - p_c}{\rho - p_c} \right)}{\log b} \equiv \frac{\log \lambda}{\log b} \; ; \; \lambda = \frac{\rho' - p_c}{\rho - p_c} = \frac{dp'}{dp} \bigg|_{p=p_c}
\]

en el límite \( p, \rho' \rightarrow p_c \).

Ilustraremos estas ideas con dos ejemplos prácticos. Consideremos en primer lugar una red triangular en dos dimensiones. Véase la figura 7.14, que ilustra la forma en que se renormaliza la red original para que se conserve la geometría y cada celda elemental pertenezca a una única supercelda renormalizada.

¿Cuándo diremos que una supercelda está ocupada? Es necesario establecer un criterio razonable para hallar \( \rho' \).

1. La supercelda estará ocupada si las tres celdas que la forman lo estaban. Esto sucede con una probabilidad \( p^3 \).
2. Estará también ocupada si dos de las celdas constituyentes lo estaban. La probabilidad de este suceso es \( p^2(1-p) \) y puede ocurrir de tres formas diferentes (que corresponden a la extracción de cada uno de los tres vértices del triángulo).
3. Estará desocupada si sólo una de las celdas estaba ocupada o si las tres estaban vacías.

Este criterio (estará el lector de acuerdo en que es sumamente razonable) permite escribir

\[
\rho' = p^3 + 3p^2(1-p)
\]

Nos interesa el punto fijo \( p^* \) de esta ecuación \((p = p')\). Tenemos tres soluciones para la condición exigida:

\footnote{En general se supondrá que la separación entre las celdas originales es la distancia unitad.}
Figura 7.14: Renormalización en el espacio real de una red triangular en dos dimensiones.

\[ p^* = 0, \frac{1}{2}, 1 \]

0 y 1 son soluciones triviales a la condición de renormalización impuesta. Obtenemos, no obstante, un resultado no trivial, \( p^* = 1/2 \), que resulta ser el umbral conocido para la percolación en la red triangular. Calculemos \( \lambda \):

\[ p' = p^* + \lambda(p - p^*) + O((p - p^*)^2) \]

realizando un desarrollo alrededor del punto fijo. Simplemente derivando

\[ p' = 6p - 6p^2 = \frac{3}{2} \quad \text{en} \quad p = p^* = \frac{1}{2} \]

En la red triangular, \( b = \sqrt{3} \), así que

\[ \nu = \frac{\log b}{\log \lambda} = \frac{\log \sqrt{3}}{\log \frac{3}{2}} = 1.355 \]

Este resultado está en buen acuerdo con el valor supuestamente correcto \( \nu = \frac{4}{3} \), aunque, a diferencia del valor del umbral de percolación, no proporciona el valor exacto.

Utilicemos el mismo procedimiento con una red cuadrada. El convenio de renormalización se representa en la figura 7.15, lo cual conduce a la ecuación

\[ p' = p^4 + 3p^3(1 - p) \]

con las soluciones en el punto fijo \( p' = p = p_c \)

\[ p^* = 0, 1, \frac{1 \pm \sqrt{13}}{6} \]
y la única solución no trivial \( p^* = 0.768 \).

Esta solución está más alejada del umbral conocido, \( p_c = 0.5927 \ldots \). En este caso, \( b = 2 \) y \( \lambda = 9p^2 - 5p^3 \),

\[
\nu = \frac{\log 2}{\log \lambda} \approx 1.40
\]

proporcionando un valor razonablemente cercano a \( \frac{2}{3} \).

Es necesario decir que este desarrollo en unas pocas líneas, aunque no exacto, es probablemente mucho mejor que cualquier simulación que se pueda hacer en el término de unos pocos días, lo cual debería dar una idea de la potencia de las técnicas de la renormalización. Los resultados serán tanto más exactos cuanto mayor sea la supercelda renormalizada, esto es, cuanto mayor sea \( b \). Sin embargo, aumentar \( b \) significa multiplicar enormemente el número de configuraciones que deben ser analizadas como posibles contribuciones a la supercelda. De hecho, en una red cuadrada deben ser analizadas \( 2^{b^2} \) configuraciones. Simulaciones del tipo Montecarlo han llegado a renormalizar celdas con \( b = 10^4 \), produciendo para \( \nu^{-1} \) un valor 0.75, exacto hasta las centésimas. Para \( b \approx 2 \), queda pues claro que las ecuaciones obtenidas son únicamente aproximadas, y el hecho de haber obtenido exactamente uno de los valores que se pretendía calcular no debe llevarnos a engaño. Hemos obtenido también, y por otra parte, dos valores muy próximos para \( \nu \) considerando que partíamos de dos geometrías absolutamente diferentes. Aquí vemos claramente reflejado el término universalidad de los exponentes críticos: la técnica de la renormalización ignora las diferencias locales menores y aplica todos los problemas críticos en el entorno del punto fijo.

### 7.3 Conclusiones e implicaciones

Tras los dos modelos vistos (Ising y percolación) se hacen necesarias algunas puntualizaciones sobre los fenómenos críticos, sobre la aparición de la física independiente de la escala de observación.

La conclusión principal que la no dependencia del tamaño de descripción aporta es el fortalecimiento y la validez de las descripciones realizadas con modelos en espacio o tiempo discreto
cuando el sistema evoluciona cerca de un punto crítico. Efectivamente, hemos visto cómo el método de la renormalización (sea en el espacio de parámetros o en el espacio real) elimina los detalles microscópicos del sistema, proporcionando no obstante resultados exactos. Nuestros modelos discretos de fenómenos críticos deben ser vistos ahora desde una nueva perspectiva: la de la posible exactitud. No es en absoluto necesario tener en cuenta los más ínfimos detalles de la realidad en su descripción, ya que puede ser que la aparición de las estructuras a gran escala los borre completamente. Probablemente las pruebas más contundentes las aportan los resultados cuantitativos para los exponentes críticos: sistemas magnéticos e hidrodinámicos, por una parte, o formación de caminos conexos en una roca y conducción de corriente en una superficie, por otra, resultan ser fenómenos descritos por los mismos números mágicos. Los detalles del sistema, como el número de vecinos de cada partícula, si ésta es fija o fluida (lucero y agua, por ejemplo), si la interacción es mediante puentes de hidrógeno o mediante estados de espín, son irrelevantes en la transición crítica: la emergencia de una estructura que ocupa toda la escala del sistema, el fenómeno cooperativo que aparece a una escala mucho mayor que la del elemento individual, los hace irrelevantes.

El criterio de validez de la descripción discreta está, en cualquier caso, acotado cuantitativamente por la longitud de correlación, $\xi$. Junto con ella, sólo otras dos longitudes son relevantes en el sistema: una longitud máxima que sería la del sistema total propiamente y una longitud mínima, a la cual se definen las interacciones entre los constituyentes elementales. Es interesante también constatar que la aparición de la estructura a gran escala no está en absoluto impuesta por las leyes dinámicas o por las interacciones, que se definen a nivel local. La superestructura emerge en el sistema, es la respuesta macroscópica a unas ligaduras externas que, sin embargo, ni la contienen ni la condicionan.

En el próximo capítulo descrebiremos con mayor extensión cuál es la peculiaridad que las ubican funciones potenciales exhiben, e introduciremos un conjunto de sistemas y algunos modelos que, dejados a su libre evolución, parecen dirigirse hacia puntos críticos. Hemos visto en este capítulo que las transiciones críticas son forzadas por la introducción de unas condiciones termodinámicas (presión, temperatura, campo magnético, ...) muy precisas. Es necesario, parece ser, forzar el sistema, empujarlo, si pretendemos observar un cambio de fase crítico. Sin embargo, en el capítulo siguiente veremos que el punto crítico puede ser alcanzado de forma espontánea. ¿Porqué? ¿Qué ventajas representan una estructura espacial y una aparición de sucesos temporales independientes ambas de la escala de observación? Intentaremos responder a estas preguntas con la introducción de los sistemas críticos autoorganizados.

Bibliografía

Capítulo 8

Sistemas Críticos
Autoorganizados

Llegamos a uno de los temas más interesantes del libro, donde se agrupan multitud de conceptos que hemos ido explorando en capítulos precedentes, como la geometría fractal, la dinámica de sistemas, los autómatas celulares, los fenómenos críticos y la frontera del caos. Los sistemas que presentaremos son ejemplos de un equilibrio delicado donde orden y desorden coexisten. Es el balance entre una cierta aleatoriedad y la autoorganización del sistema (capaz de borrar los detalles locales o microscópicos) lo que hace de los sistemas críticos autoorganizados una de las líneas de investigación más interesantes e innovadoras de las dos últimas décadas. No sólo los ejemplos que se irán sucediendo en este capítulo corresponden a los sistemas citados. En otras partes de este libro se ha hablado, o se hablará, de virus (y la catástrofe de error asociada), de evolución (y la autosimilaridad temporal de las extinciones) y se harán algunas sugerencias en esta dirección cuando se hable de neurodinámica.

Iniciaremos el capítulo profundizando en el significado de las funciones potenciales, las que representan matemáticamente la autosimilaridad. Inmediatamente pasaremos a la descripción de una serie de modelos y sistemas reales sobre los que se conjetura que evolucionan cerca de un punto crítico. En la última sección analizaremos el tipo de predicción que los sistemas críticos permiten, y los compararemos con la predicción en los sistemas caóticos.

8.1 Leyes de escala

Se ha visto en el capítulo anterior que las leyes de escala, o leyes potenciales, aparecen cerca de los puntos críticos. Cuando un sistema sigue una ley de escala, se ha dicho que sus propiedades se hacen independientes de la escala de observación. ¿Qué significa esto, exactamente? Intentaremos, en esta sección, profundizar un poco más en el significado de las funciones potenciales, a la vez que describiremos numerosos ejemplos en los que se ha hallado relaciones de este tipo. No es únicamente la estructura geométrica la que posee leyes de escala (caso en el cual hablamos de fractales) sino que también la ocurrencia temporal de algunos fenómenos puede ser invariante bajo cambios de escala (hablamos entonces de ruido $1/f^\beta$). Veremos que, en el caso del tiempo, esta invariancia tiene implicaciones en la predicción.

Lo primero que debemos analizar es porqué las leyes potenciales están libres de cualquier escala característica, y porqué ningún otro tipo de función (polinómica, exponencial, trigonométrica, ...) posee esta propiedad.
Por razones dimensionales, las únicas magnitudes que pueden ser elevadas a potencias arbitrarias son adimensionales. Esto implica que, en presencia de un exponente \( \eta \in \mathbb{R} \) arbitrario, una ley potencial será de la forma

\[
g_1(x) = \left( \frac{x}{x_0} \right)^\eta
\]

donde \( x_0 \) es alguna medida patrón con las mismas dimensiones que \( x \). Por otra parte, podemos considerar una función exponencial, por ejemplo. De nuevo, sólo cantidades adimensionales pueden aparecer como argumento de funciones trascendentales, así que tendremos

\[
g_2(x) = \exp \left\{ \frac{x}{x_0} \right\}
\]

¿Cuál es, entonces, la diferencia esencial entre una y otra función? Supongamos que medimos nuestra función \( g_2 \) para valores de \( x \) en el intervalo \((x_0/2, 2x_0)\), es decir, un intervalo logarítmicamente centrado en \( x_0 \). Consideremos el cociente entre el valor máximo y el valor mínimo que la función presenta en este intervalo: \( e^x / e^{1/2} = e^{3/2} \). Tomemos ahora un intervalo diferente, esta vez centrado en 10\(x_0 = 10x_0 \), y midamos en cociente entre el valor máximo y el mínimo en \((x_0/2, 2x_0)\). Obtendremos ahora un cociente \( e^{15} \). Si centramos el intervalo en 100\(x_0 \), la razón será \( e^{150} \). Así que las gráficas de \( g_2 \) en estos tres intervalos no son modelos a escala unas de otras: no es posible superponerlas mediante un simple cambio lineal (lo que significa bien una contracción bien una dilatación) de escala en los ejes, no son gráficas semejantes.

Realicemos la misma exploración en la función \( g_1 \). La razón entre el valor máximo y el mínimo que la función presenta en los mismos tres intervalos anteriores es siempre \( 4^9 \), así que ahora sí será posible superponer las tres partes de la gráfica realizando un simple cambio lineal. En este sentido, un fenómeno que obedece leyes de escala tiene el mismo aspecto, independientemente de la escala a la que lo contemplemos. Observemos, finalmente, que debido a la variación del cociente considerado, es posible detectar el valor de \( x_0 \) en una función exponencial. Sin embargo, la equivalencia de todos los cocientes en una ley potencial borra las huellas de \( x_0 \). Aunque se utilice para hacer el argumento de la función adimensional, la simple exploración de la función no revela qué medida tiene este patrón, ya que lo único que el observador puede detectar es el valor relativo a escalas diferentes, la relación entre dos escalas de observación, es decir el cociente, y éste es constante. Sólo una ley potencial está libre de escala.

¿Dónde encontramos leyes potenciales? Aparecen en una gran cantidad de fenómenos. De forma muy general, podríamos decir que las encontramos cuando una gran cantidad de elementos interaccionan entre ellos para producir una estructura a un nivel superior. Estos elementos constituyentes del sistema poseen una historia común, a lo largo de la cual las interacciones locales han podido extenderse a lo largo de todo el sistema. Es decir, ha habido una transmisión de información a escala global. Estos sistemas evolucionan lejos del equilibrio, y habitualmente son sistemas altamente dissipativos. Esta descripción general es probablemente un poco vaga. Los ejemplos siguientes ayudarán a concretar su significado.

El primer descubrimiento sistemático de leyes de escala en muchos sistemas diferentes se remonta a la década de 1930. Fue George Kingsley Zipf, un profesor de alemán de Harvard quien descubrió una ley fenomenológica, ahora llamada ley de Zipf, que relacionaba un cierto fenómeno cuantificable con la frecuencia de aparición de cada valor de dicho fenómeno.

Una de las primeras medidas realizadas tiene que ver con las lenguas. Zipf midió la longitud de las palabras en un texto y la frecuencia con que cada longitud aparecía. El resultado fue una ley

\[1\text{Véase la sección 3.3.2.}\]
Figura 8.1: Representación del número de habitantes de algunas ciudades americanas en función del rango que ocuparían, según la ley de Zipf.

potencial, y constató que todos los idiomas modernos seguían leyes de escala del mismo tipo. Leyes de potencia fueron observadas en sistemas tan dispares como el tamaño de las áreas metropolitanas estadounidenses, las mayores empresas por volumen de negocio o las exportaciones anuales de un país. Zipf asignó un rango 1 al mayor ejemplo de cada sistema, 2 al segundo, y así sucesivamente, a la vez que contaba cuántos ejemplares tenía en cada rango. En la ley de Zipf original, la magnitud considerada (por ejemplo el número de habitantes en una ciudad) era inversamente proporcional al rango ocupado.

Puntualicemos lo siguiente: podría pensarse que es casi trivial que haya pocos ejemplos de, digamos, ciudades muy grandes (para fijar ideas), bastante más de ciudades medianas y muchas ciudades pequeñas. De acuerdo, es trivial. Pero existe una infinidad de funciones monótonas decrecientes. Así que lo que ya no es trivial es que la escogida sea la única de entre ellas invariante de escala.

Benoit Mandelbrot ha generalizado la ley de Zipf original de forma que dé cabida a leyes igualmente invariants de escala pero más generales. Si el rango de un suceso es $r$, esta ley se puede escribir

$$f(r) = \frac{1}{(r + a) ^ \eta}$$

donde $a$ es una constante y $\eta$ el exponente que caracteriza la ley. En el caso de la ley de Zipf original, $a = 0$ y $\eta = 1$.

Estos ejemplos son sencillos y muy claros. Es quizá un poco menos evidente el estudio de la autosimilaridad temporal, de las leyes de escala del tipo $1/f^\eta$. Consideremos un ejemplo que ilustra bien la repetición de la dinámica a escalas de tiempo diferentes. Supongamos que en un cierto instante dejamos partir un móvil del origen de coordenadas y permitimos que se desplace una unidad de longitud hacia la derecha o hacia la izquierda (en una dimensión) a intervalos discretos de tiempo. La gráfica 8.2 ilustra el desplazamiento del móvil. Una ampliación de una parte de la trayectoria reproduce (no de forma exacta, sino estadística) el aspecto de la totalidad.

Este movimiento es el de un random walker en 1 dimensión (véase el capítulo 1).
Figura 8.2: Evolución temporal de un móvil que se mueve una unidad temporal hacia la derecha o hacia la izquierda (en el eje horizontal) en cada paso de tiempo (en el eje vertical). La figura de la derecha es una ampliación del recuadro pequeño de la izquierda.

La forma habitual de describir la estructura de una señal temporal consiste en estudiar su espectro de Fourier. La figura 8.3 representa, de forma muy esquemática, cómo la superposición de cinco armónicos (funciones sinusoidales puras, de frecuencia única) puede dar lugar a una señal más complicada, no periódica. En los sistemas que presentan leyes potenciales en su dinámica temporal, no existe ningún tipo de repetición exacta.

Obsérvese que la periodicidad implica existencia de un tiempo característico (el periodo), y estos sistemas están por definición libres de él. La no periodicidad también implica que, cuanto más tiempo midamos, más estructura diferente hallaremos en el sistema, y mayor será la cantidad de armónicos de baja frecuencia necesaria para describirlo. Cuando, finalmente, mediante el estudio del espectro de Fourier, constatamos que un sistema posee una dinámica del tipo $1/f^\beta$, el valor del exponente $\beta$ proporciona información sobre la dinámica que representa. Por ejemplo, si $\beta = 0$, el espectro de Fourier contiene todas las frecuencias en la misma cantidad, con la misma amplitud. Es lo que llamamos ruido blanco, y lo podemos encontrar en la agitación térmica de cualquier sustancia. En otro extremo tendríamos $\beta = 2$. Es un random walker el que posee este exponente. La dinámica presenta fuertes correlaciones, provocadas por la prohibición de desplazarse más de una unidad de longitud respecto de la posición anterior, pero tiene un componente aleatorio: el sistema sólo tiene memoria de un estado, el inmediatamente anterior, para decidir el actual. Las situaciones más interesantes aparecen entre los dos casos anteriores. Para $\beta = 1$ tenemos el llamado ruido $1/f$, que parece ser, al igual que la ley de Zipf y sus variantes, oblicuo en la naturaleza. Podríamos decir que se trata de una dinámica en la que las correlaciones se alternan con las sorpresas. Se ha encontrado este tipo de “ruído” en la música clásica, especialmente en la de Mozart o Bach, por ser las que presentan mayor velocidad de variación, pero también en el jazz o en otros tipos de música moderna, en el ruido del agua que cae por una cascada, en la sucesión de terremotos de diferente intensidad, en las variaciones de las extinciones en el registro fósil, ...

Se supone que la aparición del ruido $1/f$ tiene como principal característica (implícita en su invarianza de escala) la aparición de correlaciones que se extienden en un amplio rango de escalas temporales, como consecuencia de algún tipo de fenómeno cooperativo. Esta fluctuación temporal de tipo $1/f$ se cree que es el resultado de la dinámica intrínseca de los sistemas críticos autoorganizados. Estos son muy abundantes en la naturaleza: la luz de los cuásares, la intensidad
de las manchas solares, el paso de corriente por una resistencia, el flujo de arena en un reloj, el agua en un río como el Nilo, la dinámica de la bolsa, ...

Con los modelos que serán descritos en las próximas secciones veremos cómo las leyes potenciales en el tiempo y en el espacio tienen una descripción unificada, cómo unas y otras se entienden simultáneamente y cómo se complementan. Intentaremos entender porqué aparecen y daremos las implicaciones que tienen en la predicción y en la descripción de numerosos fenómenos naturales. Veremos que estos modelos reproducen por otro lado muy fielmente las leyes halladas en la realidad. Las razones expuestas en la última sección del capítulo anterior sobre fenómenos críticos avalan la validez de las conclusiones que de estos modelos simplificados se deriven. En las próximas secciones discutiremos la conjetura provocada por la omnipresencia de los fenómenos libres de escala: ¿tienden algunos sistemas dinámicos de forma natural a los puntos críticos?

8.2 Sistemas críticos autoorganizados (SOC)

La idea germinal sobre la posible tendencia espontánea de algunos sistemas a un punto crítico data de 1987, y se debe al danés Per Bak y a dos de sus colaboradores. Chao Tang y Kurt Wiesenfeld. Estos autores constataron la ubicua presencia simultánea de fractalidad y ruido $1/f$ en sistemas dinámicos con extensión espacial. Sugerieron que la aparición de un estado crítico autoorganizado, que el propio sistema escoge, proporcionaba una conexión entre la dinámica no lineal, la aparición de autosimilaridad espacial y el ruido $1/f$ de una forma natural y robusta.

La aproximación a algunos problemas que proponía esta teoría era innovadora. Existe toda una escuela sobre la reducción de los problemas en física. Es muy usual que un problema con muchos grados de libertad se trate de forma perturbativa, que se intente disminuir el número de variables y que se reduzca la interacción de los términos de orden superior. Esta aproximación ha proporcionado resultados brillantes en numerosos casos. Pensemos en la termodinámica (donde la partícula individual es indistinguible de las demás) o en la física de partículas, donde los desarrollos pertubativos son la base de prácticamente todo cálculo. Sin embargo, no podemos aplicar el método a, por ejemplo, un ecosistema. Para empezar, la especie individual deja de tener sentido
aislada del medio. De él depende, pero a su vez lo hace cambiar, en un círculo continuo, de forma que no se puede separar el nivel de estudio de la especie del nivel superior del ecosistema. Esta misma interdependencia de las especies hace que el sistema sea muy sensible a pequeños cambios, al ruido. Sin embargo, debe de ser suficientemente robusto como para haber llegado hasta el presente. Es el balance entre la sensibilidad a las perturbaciones y a la vez la robustez del sistema para superarlos o integrarlos lo que hace de estos sistemas **sistemas críticos**.

La autosimilaridad espacial y temporal está muy bien entendida en los fenómenos cooperativos que aparecen en las transiciones críticas de equilibrio en algunos sistemas físicos (véase el capítulo anterior). Sin embargo, el punto crítico sólo puede ser alcanzado por medio de un parámetro finamente controlado, y por tanto su aparición en la naturaleza debería ser accidental. El hecho de que sean muchas las observaciones que apoyan la existencia natural de sistemas en o cerca de un punto crítico implica que debe de existir algún mecanismo que los empuje a este punto como resultado de la dinámica intrínseca, de las interacciones, de la relación con el medio. Los sistemas naturales que se observan en los puntos críticos no operan en el equilibrio termodinámico. Al contrario, cualquier sistema vivo está lejos del equilibrio por definición: tiene una entrada constante de energía (imprescindible para mantenerse vivo) y una salida también constante de energía transformada. Son sistemas altamente disipativos. En algunos de ellos (como la pila de arena, que describiremos a continuación) existe una conservación local de la “energía”, mientras que en otros (el fuego del bosque o algún modelo de terremotos) esta energía que puede ser definida no resulta conservada ni siquiera localmente. La conservación global se produce en todos los sistemas críticos autoorganizados.

### 8.2.1 La pila de arena

El modelo sencillo que describiremos a continuación fue el primer metamodelo que intentaba aclarar algunas de las cuestiones relativas a la autoorganización en un punto crítico y junto con el cual se introdujo la criticalidad autoorganizada (**selforganized criticality, SOC**). Aunque elemental, ayuda sin duda a comprender cómo se genera y se mantiene el estado crítico en un sistema disipativo de no equilibrio, cómo perturbaciones mínimas pueden propagarse a todo el sistema y cómo se genera
la invariancia de escala. Describiremos el modelo en dos dimensiones, porque es probablemente el más ilustrativo, aunque se ha formulado también en una dimensión (para la que hay una solución exacta) y se ha simulado también en tres dimensiones.

Nuestra más que probable experiencia con pilas de arena reales debe indicar por qué un modelo sobre ellas puede ejemplificar un estado crítico. Si depositamos cuidadosamente arena en una montañita, observaremos que la pendiente crece hasta llegar a un valor que se mantiene, a pesar de que se siga añadiendo arena al montón. También podríamos realizar una especie de “castillo” con arena mojada. A medida que ésta se fuera secando, veríamos cómo se iría desprendiendo hasta alcanzar de nuevo la misma pendiente crítica anterior. Este estado es un atractor del sistema. Observemos que no representa la situación más estable que se podría alcanzar. Sería más estable el caso en que los gramos de arena se distribuyan uniformemente en una superficie plana, con lo cual se evitarían las avalanchas de arena por los lados. El estado de equilibrio dinámico que el sistema mantiene es un estado metaestable, en el que la adición de un único grano puede provocar desde la caída de un solo grano hasta la precipitación de todo el montón.

Una forma de modelizar estas avalanchas en un montón de arena es la siguiente. Consideremos una red cuadrada en dos dimensiones. Cada una de las celdas tendrá una altura asignada, \( z(x, y) \), que podrá tomar valores discretos entre cero y cuatro:

\[
z(x, y) \in \{0, 1, 2, 3, 4\}
\]

El valor \( z_c = 4 \) es el valor crítico para la altura. Si una celda alcanza el tamaño \( z_c \), reparte su carga entre sus cuatro vecinos por igual: a cada uno le corresponde “una unidad de arena”. Exteriormente se deposita una unidad en cada paso de tiempo sobre una celda que puede ser siempre la misma o puede estar escogida al azar.

\[\text{(2)}\]

En el caso de utilizar siempre la misma celda para introducir energía en el sistema, se generan configuraciones simétricas (lógicamente) en las que se puede apreciar fácilmente la disposición fractal de las alturas. Si se escogen las celdas al azar, nos enfrentamos a una configuración aleatoria, en la que la fractalidad es más difícil de detectar. Los resultados esenciales son sin embargo los mismos.
Figura 8.6: Aspecto posible de una instantánea del modelo de la pila de arena.

El modelo tiene condiciones de frontera abierta: los “granos de arena” que se pierden desde las celdas de la frontera caen fuera del sistema.

Las reglas del modelo se pueden resumir en la forma siguiente:

\[ z(x', y') \rightarrow z(x', y') + 1 \]

lo que significa que se ha escogido una celda aleatoria \((x', y')\) para depositar una unidad. Si alguna celda alcanza el tamaño crítico,

\[ z_c(x, y) \rightarrow z(x, y) - 4 \]
\[ z(x \pm 1, y) \rightarrow z(x \pm 1, y) + 1 \]
\[ z(x, y \pm 1) \rightarrow z(x, y \pm 1) + 1 \]

y el reparto prosigue hasta que \(z(x, y) < z_c, \forall x, y\), y se pierde la energía (o granos) sobrante que alcanza la frontera, cuando \(x, y = 1, N\). Dada la definición del modelo, no es posible establecer una analogía clara entre la pendiente crítica que se observa en el caso real y otra cantidad que pueda ser definida en el modelo. Sin embargo, hay algunas magnitudes interesantes que sí es posible medir con facilidad.

Definimos como \textit{avalancha} la cantidad de celdas afectadas en una actualización, debido a la adición de un único grano de arena. Cuando éste se ha depositado, dejamos que el sistema alcance de nuevo el estado de equilibrio en el que ninguna de las celdas tiene una altura superior a la crítica, \(z_c\), antes de depositar un nuevo grano. Todas las celdas que han sido actualizadas definen el tamaño de la avalancha. También podemos preguntarnos por el tiempo que el sistema tarda en recuperar el equilibrio. Observemos que el tamaño de una avalancha y este tiempo no son la misma cantidad: puede ser que durante la actualización del sistema pasemos varias veces por la misma celda, y también que en un único paso de tiempo varias celdas resulten actualizadas. Y ahora podríamos preguntarnos, ¿cuál es la distribución de tamaños de avalanchas provocada por la adición de un único grano de arena al sistema? o bien, ¿cuál es la distribución de tiempos asociada a las actualizaciones?

Ambas cantidades siguen leyes potenciales, y en este caso, es sencillo ligar ambas dependencias. La ausencia de una escala de longitud característica conduce en este caso directamente a la ausencia de un tiempo característico para las fluctuaciones. Se sabe que una superposición de los pulsos de
Figura 8.7: Distribución del tiempo de duración de las avalanchas para el modelo de la pila de arena descrito en el texto.

una cantidad física que tenga una distribución de tiempos de vida del tipo $D(T) \propto T^{-\alpha}$ (pesada con el valor medio de la cantidad durante el pulso) lleva a un espectro de frecuencias potencial, $S(f) \propto f^{-\beta-\epsilon}$, así que un espectro de frecuencias del tipo $1/f$ es equivalente a una distribución potencial en los tiempos de vida. Los resultados proporcionados por el modelo se representan en las figuras 8.6 y 8.7.

Estos resultados han sido confirmados por experimentos reales con pilas de arena en dos dimensiones y pilas de otros materiales, como por ejemplo pequeñas piezas metálicas con diferentes geometrías (irregulares, esféricas, ...). También se han montado hábiles dispositivos con pilas de arroz en una dimensión, y se ha encontrado que el valor de los exponentes depende de la forma concreta que el grano de arroz presenta.

8.3 El bosque en llamas (Forest Fire)

El primer modelo de bosque en llamas se presentó en 1990, y se debe a P. Bak, K. Chen y C. Tang. Consiste en un automata celular sencillo en el que cada una de las celdas del automata (los “árboles”) puede estar en tres estados diferentes: conteniendo un árbol vivo, un árbol ardiente o sin árbol (celda vacía). El término “bosque” es un tanto exagerado, dado que en ningún momento se considera que los árboles interaccionen entre ellos o crezcan, dos suposiciones básicas en un modelo de dinámica forestal. Sin embargo, la pretensión del modelo es otra, y la imagen del bosque sirve a sus fines. La única forma que un árbol tiene de desaparecer en este bosque es quemado. Los fuegos deben autoproporcionarse para que la dinámica no desaparezca.

La simulación del modelo comienza con una distribución arbitraria de celdas vacías, ocupadas por un árbol vivo u ocupadas por un árbol en llamas (un número pequeño de estos últimos en relación a los primeros). El sistema se actualiza de forma paralela según las siguientes reglas:

1. Un árbol en llamas se convierte en una celda vacía.
2. Un árbol vivo prende fuego si alguno de sus 4 vecinos está ardiente.
3. En una celda vacía aparece un árbol vivo con una probabilidad $p$. 
Figura 8.8: Tres instantes consecutivos de la propagación del frente de fuego en el modelo del bosque en llamas. Los cuadrados representan árboles que arden, y las cruces árboles vivos. El resto son celdas vacías, en las que puede nacer un árbol en cada paso de tiempo con una probabilidad que es en este caso $p = 0.09$.

En el caso de que el tamaño del sistema sea mayor que la longitud de correlación del fuego, el modelo alcanza un estado estacionario con una densidad finita de árboles en llamas. En la figura 8.8 se representan tres pasos de tiempo consecutivos en el modelo.

Se ha visto que este modelo de bosque en llamas no presenta ni una dinámica ni una geometría críticas en el sentido que hasta ahora se ha descrito. En 1992 se realizó una pequeña pero no trivial modificación a este modelo, con lo cual se pudo simular un estado realmente crítico. Este modelo modificado se debe a Barbara Drossel y Franz Schwabl (véase también el trabajo de C. L. Henley, 1993). La modificación consiste en introducir una pequeña probabilidad, $f$, de ignición espontánea para los árboles vivos. Así, se introduce una cuarta regla en el modelo,

4. Un árbol vivo sin vecinos en llamas se quema con una probabilidad $f$.

La introducción de este nuevo parámetro permite trabajar con cualquier valor para la probabilidad de nacimiento $p$, sin el problema de la posible extinción total del fuego que el modelo inicial presentaba para $p$ pequeña. Ahora es posible que el fuego se produzca en pequeños grupos aislados de elementos. Este nuevo modelo tiene una estrecha relación con el problema de la percolación. Obsérvese que cuando se incendia un árbol, todos los que están conectados a él también se incendiarán, y habremos eliminado toda la agrupación de la que el árbol afectado formaba parte. Ahora, a diferencia del caso $f = 0$, pequeñas zonas aisladas de bosque se pueden quema, y no es necesario que este fuego se mantenga (extendiéndose en consecuencia a la escala total del sistema) para que la dinámica no evolucione hasta el punto trivial de ausencia de fuego. Si pensamos en el tipo de estructuras que se formarán para $f \neq 0$, y tenemos presente el modelo de la percolación, parece sencillo deducir que los puntos en ignición tendrán una distribución fractal sobre el sistema para ciertas combinaciones de los parámetros $f$ y $p$. Esto no sucedía con el caso $f = 0$. La exigencia de que el fuego se autoperpetúase en el sistema llevaba en algunos casos a la formación de macroestructuras con longitudes características. Véase en la figura 8.9 cómo la introducción de un valor fijo para $f$, incluso en el caso de que éste sea muy pequeño, produce la progresiva desaparición de estas estructuras.

Profundicemos un poco más en el último modelo, ya que permite algunos cálculos analíticos que aseguran su estado crítico. Consideremos una simplificación del modelo, únicamente para efectuar
Figura 8.9: Dos instantáneas del bosque en llamas para una red cuadrada de tamaño $300 \times 300$ en el estado estacionario. A la izquierda, $f = 0$, y a la derecha, $f = 0.0005$. En ambos casos, $p = 0.06$. Sólo se ha representado, para mayor claridad, los árboles en llamas.

el siguiente razonamiento. Supongamos que, cuando un árbol se ve afectado por la probabilidad de ignición $f$, todos los árboles que están conectados a él arden en el mismo paso de tiempo. Supongamos que la densidad media estacionaria (independiente del tiempo) de árboles vivos en el bosque es $\bar{\rho}$. Entonces, en un intervalo de $t$ pasos de tiempo habrán ardido

$$tf\bar{\rho}L^2$$

árboles, siendo $L$ la medida lineal del sistema y suponiendo que trabajamos en dimensión 2. Por otra parte, el número de árboles que nacerán en los mismos $t$ pasos de tiempo será

$$tp(1 - \bar{\rho})L^2$$

Por tanto el número medio de árboles destruidos por uno incendiado (el tamaño medio de las agrupaciones del sistema, de hecho) es

$$\bar{s} = \frac{p(1 - \bar{\rho})}{t}$$

independiente de $L$. Obsérvese que, si $\bar{\rho} < 1$ en el límite $f/p \to 0$, entonces $\bar{s} \propto p/f$, y obtenemos una primera ley potencial con exponente 1 para la divergencia de la relevante cantidad $\bar{s}$.

En el modelo original, la ignición del grupo no es instantánea, pero la ecuación para $\bar{s}$ será aún válida si, en el tiempo $T(s)$ que tarda en quemarse una agrupación de tamaño $s$, la cantidad de árboles que se ha unido a su frontera es despreciable. Esto se cumplirá siempre que $p^{-1} \gg T(\bar{s})$. Obsérvese que $1/p$ proporciona el tiempo característico en que vuelve a aparecer un árbol tras un frente de fuego (que deja la celda vacía).  

Estamos suponiendo que en el estado estacionario aparecen tantos árboles como se destruyen, es decir, si cada árbol directamente afectado por la probabilidad de ignición arrastra de media $\bar{s}$ árboles, entonces se verifica

$$tf\bar{\rho}L^2\bar{s} = tp(1 - \bar{\rho})L^2$$

de donde recuperamos 8.4.1.

En el caso $f = 0$, $p^{-1}$ da la distancia característica entre los frentes de fuego que se forman en el sistema.
Podemos calcular las relaciones de escala y los exponentes críticos que se deducen de la suposición de invariancia de escala. Definimos el radio de una agrupación en el bosque, $R(s)$, como la raíz cuadrada media de las celdas de la agrupación al centro de masas de la misma. La longitud de correlación es $\xi = R(\hat{s})$, y se definen los exponentes $\nu$ y $\tau$ como

$$\xi \propto \left(\frac{f}{p}\right)^{-\nu} \quad T(s) \propto \xi^\tau$$

La condición anterior sobre la rapidez de ignición de las agrupaciones se lee ahora $\nu \ll (f/p)^{\nu \tau}$.

Sea $N(s)$ el número de agrupaciones de $s$ árboles en un sistema de cierto tamaño. Introducimos los momentos de la distribución no normalizada, $N(s)$ como

$$m_n = \int_1^\infty s^n N(s)ds$$

Por la expresión 8.4.1, $N(s)$ no puede decaer más rápidamente que una función potencial en el límite $f/p \to 0$. Como la densidad de árboles en el sistema $\rho$, es finita, $N(s)$ decaer como máximo como $s^{-2}$ (para asegurar la convergencia de la integral). Por tanto, $N(s)$ seguirá una ley potencial cuando $f/p \to 0$. Bajo la transformación de escala

$$\hat{s} \to \frac{\hat{s}}{b}, \quad \frac{f}{p} \to b^{1/\nu} \frac{f}{p}$$

tanto la distribución normalizada de agrupaciones,

$$\frac{N(s)ds}{m_0}$$

como la función $R(s)$ deben ser invariantes. Esto implica las siguientes leyes de escala:

$$N(s) \propto s^{-\tau} \times \begin{cases} C(s/s_{max}) & si \quad \tau > 2 \\ C(s/s_{max}) \ln^{-1}(s/s_{max}) & si \quad \tau = 2 \end{cases}$$

$$R(s) \propto s^{1/\mu} \hat{C}(s/s_{max})$$

con $s_{max} \propto (f/p)^{-1}$. $\lambda$ y $\tau$ son nuevos exponentes críticos, y las funciones de corte $C$ y $\hat{C}$ decrecen monótonamente desde 1 para $x \ll 1$ hasta 0 para $x >> 1$. La corrección logarítmica para $N(s)$ en el caso $\tau = 2$ garantiza una densidad media finita para $f/p \to 0$. El exponente $\mu$ es la dimensión fractal de las agrupaciones. De todo lo anterior se pueden deducir las siguientes relaciones:

$$\lambda = \nu \mu \quad d = \mu(\tau - 1)$$

donde $d$ es en general la dimensión del espacio de trabajo, que antes se había supuesto de valor 2. Como por geometría $\mu \leq d$, resulta $\tau \geq 2$, como ya se había supuesto. Por tanto, $\hat{s}$ es ahora

$$\hat{s} = \frac{m_2}{m_1} \propto \begin{cases} s_{max}^{3-\tau} & 3 > \tau > 2 \\ s_{max}/\ln(s_{max}) & \tau = 2 \end{cases}$$

Ya tenemos $\hat{s} \propto |p/f|$. Como existe una única escala divergente en el sistema, ésta deberá ser $s_{max}$, con lo cual

$$\hat{s} \propto s_{max}^{1/\lambda}$$

Junto con las relaciones anteriores obtenemos para los exponentes críticos
\[ \lambda = X, \quad 3 - \tau = \frac{1}{X}, \quad \mu = d, \quad \nu = \frac{X}{d} \left( 2 - \frac{1}{X} \right) \]

en función únicamente de \( X \), y para la dimensión \( d = 2 \)

\[ \mu = \frac{2}{2 - 1/X}, \quad \nu = X - \frac{1}{2} \]

Las simulaciones numéricas proporcionan los siguientes valores de los exponentes

\[ X = 1.180, \quad \tau = 2.150, \quad \lambda = 1.167, \quad \mu = 1.95, \quad \nu = 0.58 \]

Las relaciones entre ellos que se habían derivado se cumplen dentro del margen de error experimental. Se puede apreciar que estos exponentes, aunque definitivamente diferentes, no presentan valores muy alejados de sus análogos en el caso de la percolación pura (\( \tau \) se corresponde con el exponente del mismo nombre, \( \mu \) se relaciona con la dimenión fractal de la agrupación de percolación, ...). Las simulaciones permiten explorar regiones alejadas del caso límite. Por ejemplo, se puede calcular la densidad estacionaria en el sistema, \( \rho \), que resulta ser \( \rho = 0.39 \) para el sistema en 2 dimensiones. Este valor cumple una propiedad de extremo: es el mínimo de todos los valores compatibles con los parámetros. Se puede interpretar como que el sistema se organiza en un estado maximamente destructivo: el fuego quema tantos árboles como es posible, haciendo que la cantidad de energía disipada en el sistema sea también máxima.

### 8.4 Terremotos

Fue en el año 1956 cuando B. Gutenberg y C. F. Richter obtuvieron una ley potencial empírica que relacionaba la frecuencia de los terremotos con la energía que se liberaba en cada evento. Si \( s \) es el momento sísmico (o energía \( E \)) correspondiente a un cierto terremoto, el número total \( N(s) \) seguiría una ley

\[ N(s) \propto s^{-1-\beta} \]

Los datos reales fueron extraídos del catálogo de Harvard de terremotos. Un dato muy curioso es que el exponente \( \beta \) es prácticamente independiente del área geográfica observada. Es decir, aunque es evidente y bien sabido que existen ciertas regiones de la corteza terrestre (por ejemplo las zonas de fallas) donde los movimientos sísmicos de alta intensidad son mucho más probables, y otras donde los terremotos de tierra aparecen rara vez y son de mucha menor intensidad, todos los eventos se ajustan a la misma ley, con el mismo exponente. Esta evidencia lleva a suponer que la estructura global que actualmente presenta la corteza terrestre es el resultado de muchos millones de años de evolución conjunta. Ha transcurrido tiempo suficiente como para que la información correspondiente a cada uno de los puntos de la corteza se haya transmitido al sistema global. Es la historia común la causa de que las interacciones locales hayan llevado al sistema al estado supuestamente crítico que actualmente observamos.

El hecho de que exista la ley universal de Gutenberg-Richter implica que la frecuencia de los grandes terremotos se puede extrapolar a partir de la frecuencia de los pequeños, lo cual indica la existencia de un mecanismo común subyacente. Esta ley puede ser entendida formalmente como la consecuencia de un proceso en cadena. Supongamos que en algún punto de la corteza se inicia una actividad sísmica como resultado de una cierta inestabilidad. Esta actividad puede propagarse a las zonas contiguas, seguir inalterada o extinguirse. Finalmente habrá toda una zona afectada, y la intensidad del terremoto puede medirse por su extensión. Sólo si la probabilidad de propagación es exactamente igual a la probabilidad de que la actividad se extinguiga sigue la distribución de
frecuencias de terremotos una ley potencial. La reacción en cadena debe de ser precisamente crítica. Se supone que la tierra ha tenido tiempo suficiente durante toda su evolución geológica como para alcanzar el estado actual que observamos, de no equilibrio y altamente excitado, de forma que se ha podido llegar a un punto de estado estacionario, estadísticamente hablando.

Existe una evidencia empírica independiente que corrobora la idea de que la corteza está organizada en un punto crítico: la distribución sobre ésta de los epicentros de los terremotos. Si medimos el reparto de la actividad sísmica sobre la tierra, observaremos que existe un claro agrupamiento de seísmos. En absoluto presentan una distribución uniforme, sino que se sitúan de forma que se puede asociar una dimensión fractal, entre 1 y 2, a la geometría de los epicentros. Las dos razones expuestas, autosimilaridad en el tiempo (en la intensidad de los terremotos) y autosimilaridad en el espacio (existencia de una distribución fractal de epicentros) son suficientes como para intentar una aproximación al sistema mediante un modelo simplificado, si conjeturamos que estamos de nuevo ante un sistema crítico autoorganizado.

8.4.1 Teoría de Campo Medio para el tiempo de retorno

Tras la propuesta de la existencia de sistemas críticos autoorganizados (1987), no tardó en sugerirse que la idea encajaba bien con los datos empíricos de los registros sísmicos (A. Sornette y D. Sornette, 1989). Los modelos habituales de la corteza consideran que se trata de un material plástico que puede soportar ciertos esfuerzos (fuerzas) hasta el umbral en el que se produce una ruptura, como consecuencia de la superación del límite de deformación que el material constituye. Se considera habitualmente que la colisión de las placas continentales provoca la aparición creciente de esfuerzos en los bordes. Estas tensiones se difunden “suavemente” hacia el interior de la placa, distribuyendo la energía de acuerdo con las leyes de la elasticidad, y perdiéndola en el punto donde se produce la ruptura. La energía perdida se transforma en la realidad en ondas sísmicas.
En esta sección describiremos una aproximación de campo medio realizada por A. Sornette y D. Sornette que permite estimar el tiempo característico de retorno de un terremoto de una cierta magnitud. Esta derivación es muy interesante desde el punto de vista de la predicción. Como veremos hacia el final del capítulo, los sistemas críticos autoorganizados permiten un tiempo de predicción muy superior al de los sistemas caóticos (de hecho el tiempo de predicción podría tender a infinito si tuviésemos suficiente información - no necesariamente infinita - sobre el estado del sistema). Esta sección es un primer paso en esa dirección.

Retomemos la citada ley de Gutenberg-Richter, según la cual el número $N$ de terremotos de una magnitud $M$ se puede expresar como

$$ \log N = a - b \, M $$

y es válida en un rango de magnitudes $2 \leq M \leq 8$. Las constantes $a$ y $b$ pueden depender ligeramente de la región particular de observación. $b$ es considerablemente universal ($0.8 \leq b \leq 1.1$) para todos los terremotos con $2 \leq M \leq 6.5$, en cualquier parte del mundo. Nos concentraremos en la zona citada, donde se verifica la relación anterior. La magnitud del terremoto se relaciona con la energía que éste libera

$$ \log E = c + d \, M $$

con lo cual se obtiene la siguiente ley para el número $N$ de terremotos que liberan una energía $E$:

$$ N \propto E^{-\tau + 1} $$

y el valor observado es $\tau \approx 2$. Sólo si $\tau > 2$ se obtiene un número finito de terremotos, definido como

$$ \int_{E_0}^{\infty} N(E')dE' < \infty $$

así que restringimos la siguiente discusión a esta condición. La pregunta que ahora nos planteamos es: ¿cuánto tiempo se debe esperar para observar un terremoto de energía $E$ después de un evento de la misma magnitud? Incluso en el caso de que la respuesta esté afectada de error, la posibilidad de estimar el tiempo de recurrencia de un seismo de gran magnitud es terriblemente interesante.

Supongamos que, en promedio, la energía que se almacena en las tensiones entre placas continentales en el sistema es $e_0$ para cada intervalo de tiempo $t_0$. En un tiempo $t$ se almacenará una energía $e = te_0/t_0$. Sea

$$ \int_{E_{\max}}^{\infty} N(E')dE' $$

la probabilidad de que un terremoto determinado tenga una energía superior a $E_{\max}$. Supongamos que en un intervalo de tiempo $t$ se producen $n(t)$ terremotos de energía arbitraria. La condición

$$ n(t) \int_{E_{\max}}^{\infty} N(E')dE' \approx 1 $$

significa que, como máximo, se ha producido un terremoto de energía $E_{\max}$ en el intervalo de tiempo $t$. Por la expresión de $N(E)$ obtenemos, integrando

$$ n(t) \propto E_{\max}(t)^{\tau - 2} $$

Esta última expresión da el valor del terremoto típico mayor observado tras un tiempo de espera $t$: $E_{\max} \propto n(t)^{1/(\tau - 3)}$. 

Sistemas Críticos Autoorganizados
También podríamos calcular la probabilidad \( P(t, E) \) de que todos los terremotos que se han producido entre tiempo 0 y tiempo \( t \) tengan una energía menor que \( E \):

\[
P(t, E) \propto \exp \left[ -n(t) \int_{E}^{\infty} N(E')dE' \right] \propto \exp \left[ n(t) E^{-(r-2)} \right]
\]

con \( C \) constante. Definimos el flujo medio de energía como

\[
J(t) \propto \frac{n(t)}{t} < E > (t)
\]

Entonces, conocido \( n(t) \), el tiempo de espera correspondiente a cada valor de la magnitud se deduce de la condición de conservación del flujo \( J(t) \).

La energía media que se libera en un tiempo \( t \) es

\[
< E > (t) = \int_{1}^{E_{\text{max}}(t)} E' N(E')dE'
\]

e integrando (con \( N \propto E^{1-r} \)) resulta

\[
< E > (t) \propto n(t)^{\frac{r-2}{r}}
\]

La condición de que la energía que entra en el sistema se disipe en la misma cantidad debido a los seísmos implica que el flujo de energía \( J \) es constante, y por tanto

\[
n(t) \propto t^{r-2}
\]

La misma condición de estacionariedad implica que \( n(t) \) debe ser proporcional a \( t \), ya que el número medio de terremotos por unidad de tiempo \( \frac{dQ}{t} \) es constante. Esto implica finalmente que el valor de \( r \) según esta teoría de campo medio es

\[
\tau_{CM} = 3
\]

La coincidencia numérica con el valor observado no es excesivamente buena, ya que se ha eliminado toda la dependencia en las fluctuaciones espacio-temporales.

La condición \( n(t) \propto t \) implica

\[
E_{\text{max}} \propto t,
\]

divergencia que se verá truncada por los efectos de tamaño finito. Como es habitual, el tamaño del sistema impone un límite superior a la autoalinead, como ya se había comentado en general.

La última ecuación permite la predicción del tiempo de retorno típico de los terremotos de una magnitud dada:

\[
t = t_0 \exp [dM]
\]

El criterio que da potencia a esta expresión es nuevamente el de comparación de dos cocientes (véase la sección 8.2). Supongamos que en una cierta región se produce un terremoto de magnitud \( M_1 = 3 \) cada \( t_{M_1} = 1 \) mes. Entonces, el tiempo típico de retorno de un terremoto de magnitud \( M_2 \) en la misma región será

\[
t_{M_2} = \exp [1.5M_2 - M_1] / t_{M_1} \approx 1800 \text{ meses}
\]

Esta aproximación de campo medio conduce directamente a la predicción de ruido \( \frac{1}{t} \) en los periodos entre grandes terremotos. Consideremos de nuevo la probabilidad \( P(t, E_{\text{max}}) \approx ct., \) de
que todos los terremotos en un intervalo $t$ tengan una energía inferior a $E_{max}$. La probabilidad de que en este intervalo no aparezca ningún terremoto de energía $E_{max}$ y si lo haga en el intervalo siguiente, $t + dt$ es

$$p(t) \approx P(t, E_{max}(t)) \left\{ \int_{E_{max}}^{\infty} N(E')dE' \right\} \frac{dn(t)}{dt} \approx Cl^{-1}$$

que es la ley potencial que caracteriza el ruido $\frac{1}{f}$, ya que es invariante bajo el cambio $t \rightarrow \lambda t$.

### 8.4.2 Un modelo sencillo

Uno de los modelos más sencillos para describir la dinámica de un terremoto es el que descriptibremos a continuación. Consideraremos una red cuadrada en la que cada celda representará un bloque capaz de interaccionar con los bloques próximos. Sobre cada uno de estos bloques en la posición $(i, j)$ actúa una fuerza de módulo $F_{ij}$, en una dirección no especificada. Inicialmente, $F_{ij}$ tiene un valor pequeño aleatorio. A cada uno de los bloques del sistema se le asigna un umbral crítico de ruptura, $F_c$. En cada paso de tiempo se aumenta la fuerza sobre cada bloque de manera uniforme. Cuando en uno de los bloques se llega al umbral de ruptura, se realiza la siguiente actualización:

$$F_{ij} \rightarrow 0$$

$$F_{nn} \rightarrow F_{nn} + \alpha F_{ij}$$

donde $F_{nn}$ representa la fuerza sobre los vecinos próximos, $(i \pm 1, j$ y $(i, j \pm 1)$. $\alpha$ es un parámetro ajustable del modelo. Si $\alpha = 1/4$, el sistema será localmente conservativo (la conservación global depende de las condiciones de contorno) y si $\alpha < 1/4$ es disipativo, y representa mejor el caso de los terremotos, en los que hay una transformación de energía de deformación plástica a energía mecanica (ondas). No requiere más explicación la forma en que se puede propagar la inestabilidad y dar lugar a una reacción en cadena. El mecanismo es similar al de la pila de arena, aunque ahora no es localmente conservativo (A. Corral et al., 1995).

La intensidad de un terremoto se medirá como el número total de celdas actualizadas cuando una de ellas ha sobrepasado su umbral crítico y se ha dejado evolucionar al sistema hasta un nuevo estado quiescente. Durante la actualización no se aumenta la fuerza en ninguna celda: existe la suposición implícita de que las escalas geológicas (en las que se producen los incrementos en las $F_{ij}$) son mucho mayores que el tiempo de ejecución de un terremoto, así que este último suceso se produce instantáneamente con respecto del primero. Este modelo fue sugerido por Olami, Feder y Christensen en 1992, quienes lo relacionaron con modelos previos en los que los bloques se conectaban por muelles, algo más complicados. Este modelo presentó una novedad interesante. Se había pensado que si no existía conservación local de la energía, aparecía en cualquier sistema una longitud, una escala característica, y esta violación de la conservación se creía que llevaba al sistema lejos de ser crítico. Sin embargo, se obtuvo en este caso que el sistema se comportaba como crítico para valores de $\alpha$ hasta 0.65, es decir, con una conservación únicamente del 20% (lo cual significa una pérdida del 80%).

El valor que este modelo da para el exponente real $b$ de Gutenberg y Richter depende del grado de disipación en el sistema, así que $b$ no es un exponente universal en el caso no conservativo para este modelo.

Existen variantes que modelizan también terremotos. Se ha considerado la posible relevancia de los desplazamientos verticales en los movimientos tectónicos, lo que implicaría considerar una tercera dimensión en el modelo. También se puede afinar más la mecánica de propagación del
Figura 8.11: Representación simple de la corteza terrestre, como bloques conectados por resortes. Sobre el bloque situado en la posición \((i, j)\) actúa una fuerza \(F_{ij}\).

El terremoto y la relajación de la ruptura en el modelo, siempre intentando aumentar la conexión con los casos reales. No obstante, de nuevo debemos tener presente que, si el sistema es crítico, los detalles menores en su descripción no afectan el resultado global. Sin embargo, el conocimiento del sistema puede ser determinante a nivel local. Como mínimo, una posible predicción local dinámica necesita de información sobre las condiciones de prácticamente todos los puntos del sistema.

8.5 El Juego del Bosque

Presentamos a continuación un modelo reciente (R. V. Solé y S. C. Manrubia, 1995) que describe la dinámica temporal y la estructura espacial de las selvas tropicales. La coincidencia no es únicamente cualitativa, sino también cuantitativa, ya que los resultados del modelo han podido ser comparados con datos de campo reales, que también presentaremos. Se puede discutir en este caso si el modelo realmente se encuentra en un punto crítico. Algunas magnitudes así lo atestiguan, en tanto que otras resultan ser autosimilares en zonas que no se corresponderían con los sistemas reales, con las medidas de campo. Puede ser que las mismas restricciones biológicas que impiden que todos los árboles de un bosque mueran a la vez (por ejemplo) alejen en cierto modo al sistema de los puntos críticos hasta ahora descritos. Veremos que esta separación se traduce en la existencia de escalas de corte, o longitudes características, a partir de las cuales se pierde la autosimilaridad estricta.

El trabajo principal realizado (sobre el sistema real y también sobre el modelo) corresponde al estudio de la distribución de claros en el bosque. Un claro es una extensión de bóveda baja, relacionada con la reciente caída de un árbol. Estos claros tienen una gran importancia en la dinámica de la selva, ya que son los responsables de ofrecer oportunidades de colonización a nuevos individuos. Clásicamente se divide a los árboles de las selvas tropicales en dos grandes grupos. Uno corresponde a las especies denominadas oportunistas. Estas aprovechan precisamente estos claros para colonizar rápidamente la zona abierta. Crecen a considerable velocidad para evitar la competición con sus vecinos (producen en consecuencia madera blanda, de baja calidad) y necesitan de gran cantidad de luz. No pueden crecer bajo la bóveda, por la escasa cantidad de luz.
solar que llega al suelo. El otro grupo está representado por los árboles que resisten bien la escasez de luz, pueden crecer bajo la bóveda y lo hacen lentamente, proporcionando maderas nobles de gran calidad. Esta distinción no está contemplada en el modelo, pero si se incorpora el mecanismo principal de desarrollo de nuevos individuos: la formación continua de claros.

A nivel práctico, se trabaja con una red cuadrada en la que cada punto dibujado corresponde a una zona de bóveda baja o, en el caso del modelo, a un “árbol” de altura nula. En la figura 8.13 se presenta una imagen correspondiente a un caso real. Es un área de 50 hectáreas (0.5 km²) de la selva de Barro Colorado (BCI), situada en una pequeña isla aislada en el canal de Panamá y preservada de la intervención humana. Corresponde a la mayor extensión sobre la que se ha realizado un estudio de esta clase.

La imagen representada es el resultado de dos años de trabajo de equipo desde el suelo de la selva. Se midió la altura de todos los árboles, arbustos y lianas de más de 2 m de altura. Cada punto negro de bóveda baja representa un área de 25 m², y se tomó como tal cuando la altura promediada de las esquinas no superaba los 10 m. Debe tenerse en cuenta que la altura de los árboles en la bóveda alcanza los 60 m, con lo que 10 m puede considerarse “un claro”.

8.5.1 El modelo

El *Juego del Bosque* es un autómata celular estocástico modelado sobre una red cuadrada 2-dimensional de tamaño $L \times L$. El término estocástico se refiere a la existencia de variables dependientes de una probabilidad, y es aplicable a todos los modelos descritos hasta ahora en este capítulo. Se ha considerado condiciones periódicas de contorno, pero en este modelo los resultados serían idénticos si considerásemos condiciones abiertas, debido a la disipación a nivel local, y no en las fronteras, como se producía por ejemplo en el autómata de la pila de arena.

Cada punto de la red representa un árbol que puede crecer desde una altura mínima $s_c$ hasta una máxima permitida, $s_m$. El ritmo de crecimiento depende del grado de interacción con sus vecinos. La dinámica del modelo se basa en la competición por los recursos del medio que se dan de forma natural entre los árboles. El estado de cada uno de ellos en cada instante de tiempo está representado por la función altura, $s(i,j)$, asignada a cada punto de la red.

El modelo está definido por cuatro reglas simples:

A Crecimiento. Un árbol dado crece si el apantallamiento provocado por sus vecinos no supera un cierto umbral. La altura del árbol se actualiza cada paso de tiempo $n$ de acuerdo
Figura 8.13: Imagen de 50 hectáreas de la isla de Barro Colorado, un bosque tropical situado en el canal de Panamá. Los puntos negros representan los lugares donde la bóveda era inferior a 10 m en 1982, 83, o en ambos años (los que duró el estudio realizado). En la parte inferior, distribución de claros que proporciona el Juego del Bosque para los parámetros $P_h = 0.5$, $P_d = 0.013$, $\gamma = 2.5$ y $\mu = 1$. 
con la siguiente regla:

\[ s_{n+1}(i,j) = s_n(i,j) + \Delta_n(i,j) \]

\( \Delta_n(i,j) \) considera la forma de interacción entre vecinos, que en este modelo se ha supuesto que son los ocho árboles que rodean a cada uno. Hemos considerado un apantallamiento sencillo y dependiente de la altura (el estado) de estos ocho árboles. \( \Delta_n(i,j) \) se define como

\[ \Theta(z) = 2 \text{ si } z > 0, \quad \text{y } \Theta(z) = 0 \text{ si } z < 0; \text{ si el apantallamiento es demasiado elevado, no se permite el crecimiento.} \quad < r, s > \text{ indica la restricción de la suma a los ocho próximos vecinos. Hemos tomado el mismo ritmo de crecimiento para todos los árboles, } \mu = 1. \gamma \text{ representa la fuerza de la interacción, y ha resultado ser el parámetro más importante del modelo, ya que es el único que introduce no linealidades, y por tanto el comportamiento más interesante. Si } \gamma = 0 \text{ no existe interacción (los árboles crecen de forma independiente unos de otros), y cuando } \gamma \to \infty, \text{ el crecimiento se imposibilita en presencia de vecinos, y el bosque experimenta una transición hacia un estado congestionado, en el que todos los elementos presentan el tamaño mínimo cuando se supera cierto umbral de percolación.} \]

\( \Delta_n \) introduce cualquier tipo de competición que se produzca entre los árboles: la carrera por la luz, por los nutrientes, por el espacio, y la posible interacción con otras plantas o animales que pudiesen jugar algún papel en el crecimiento. \( \Delta_n \) es una función sencilla, pero si la selva, como sistema dinámico, opera cerca de un punto crítico, sabemos que la invarianza de escala implica independencia de los detalles finos del sistema, y por tanto esta función debería ser suficiente como para considerar los aspectos esenciales de todos los factores que participan en la interacción.

\section*{B Muerte.} Un árbol cae (muere) cuando alcanza la altura máxima \( s_c \) (hemos considerado \( s_c = 30 \)), o bien en cualquier instante de tiempo, de acuerdo con cierta probabilidad \( p_d \), que se mantiene fija. Inmediatamente después de su muerte, su altura toma el valor cero. La suposición de que la probabilidad de muerte es constante, y por tanto independiente de la altura o de la edad de un árbol está apoyada por estudios de campo en selvas reales, que además sitúan este valor entre el 1 y el 2% anual.

\section*{C Formación de claros.} Cuando un árbol cae, no sólo él es eliminado, sino también los árboles situados a su alrededor en un radio \( R \), de acuerdo con la regla siguiente:

\[ \sum_{r,s} s_n(r,s) \leq s_n(i,j) \]

\[ ((r-i)^2+(s-j)^2 \leq i^2+j^2) \text{.} \] Esta regla tiene un fundamento biológico: durante el crecimiento de los árboles, son muy numerosas las llanas que los ligan, y que fuerzan pequeños "efectos dominó" cuando uno de estos gigantes se desploma. La biomasa de vecinos eliminada (como se deduce de la expresión anterior no está restringida a los ocho próximos vecinos) depende de la biomasa del árbol que cae. \( R \) se determina a partir de la desigualdad anterior. Cuando el árbol afectado por llegar al tamaño máximo \( s_c \) o por la probabilidad de muerte \( p_d \) cae, sólo él puede arrastrar a los árboles de su alrededor, pero éstos no pueden a su vez eliminar a otros. Esto evita un efecto dominó de alcance infinito, en el que todo el bosque se podría ver implicado. Si comparamos con el caso de las avalanchas en la pila de arena o con los
terremotos descritos en la sección anterior, deducimos sin dificultad que la eliminación de eventos del tamaño del sistema podría llevarlo lejos del punto crítico (estamos introduciendo una escala máxima para la autosimilaridad de la distribución de claros). Sin embargo, esta regla es muy razonable desde el punto de vista biológico, ya que en un bosque real no se forman ni árboles ni claros arbitrariamente grandes.

D Nacimiento. Un árbol nuevo puede aparecer en una celda vacía con una probabilidad \( p_b \) y una altura \( s_0 \). Se tomó \( s_0 = 0.1 \) como tamaño mínimo en todas las simulaciones.

La red se actualiza asincrónicamente. Cada paso de tiempo se divide en dos actualizaciones, cada una considerando \( L \times L \) celdas escogidas al azar. La primera actualización calcula el incremento de altura en todos los árboles del sistema (celdas ocupadas por un árbol) y la proporción de nuevos individuos (en las celdas que estaban ocupadas por un claro). La segunda actualización se encarga de eliminar los árboles que mueren y formar los claros. Cada paso de tiempo (con estas dos actualizaciones) corresponde aproximadamente a un año de tiempo en un bosque real.

Como se habrá podido ver, a pesar de que el modelo es algo más sofisticado que los descritos anteriormente, sólo las propiedades esenciales de lo que debe ser un bosque real se han introducido como reglas del modelo, y no se han considerado ni detalles finos sobre la estructura (como por ejemplo la diferenciación de especies, más de 300 en BCI) ni sobre las interacciones, siguiendo con la suposición de que el sistema está cerca de un estado crítico.

### 8.5.2 Resultados

En todas las simulaciones que presentaremos se mantuvo la propiedad de nacimiento en un valor \( p_b = 0.5 \). Otros valores simplemente desplazaban en el espacio de fases los comportamientos genéricos que describiremos. La interacción \( \gamma \) y la probabilidad de muerte, \( p_d \), se escogieron como parámetros del modelo, aunque esta última debe mantenerse entre los márgenes antes citados para representar un sistema real. El único parámetro realmente libre es \( \gamma \), ya que, dada la cantidad de elementos que agrupa, se hace imposible realizar una estimación de él en un caso real. Los valores extremos de la altura de los árboles, \( s_0 \) y \( s_e \) podían haber sido escogidos de otra forma. Para cada elección, se obtiene que el sistema se autoorganiza de forma tal que las leyes de escala y el aspecto cualitativo del espacio de parámetros no varían.

En la figura 8.14 se representa el espacio de parámetros del modelo, en el que se distinguen tres dominios bien definidos.

A. Oscilaciones, DO. Una pequeña parte del espacio de parámetros presenta oscilaciones en la biomasa (nótese que la escala de la figura 8.14 es logarítmica). Debido a la débil interacción \( \gamma \) y a la pequeña probabilidad de muerte, los árboles inician su crecimiento simultáneamente y al mismo ritmo. Mueven cuando alcanzan el tamaño crítico, lo que inicia de nuevo el ciclo de nacimiento y crecimiento.

El estudio de la transformada de Fourier de las fluctuaciones en la biomasa presenta un pico en la frecuencia de repetición de la dinámica descrita. También se ha medido la distribución de tamaños de árboles, \( F(h) \), que presenta picos a tamaños aproximadamente enteros (véase la figura 8.17), lo cual puede ser entendido siguiendo el análisis siguiente. Calculemos en este dominio el tamaño de un árbol en el paso \( n \)-ésimo de crecimiento. Supongamos que \( \bar{s} \) es el tamaño medio de los árboles. Si \( \gamma \) es suficientemente pequeña, el crecimiento en cada paso de tiempo se puede escribir como

\[
\Delta = 1 - \gamma \bar{s}
\]

donde se supone que \( \gamma \bar{s} < 1 \). Entonces, la altura de un árbol en el \( n \)-ésimo paso de crecimiento es
Figura 8.14: Espacio de parámetros del juego del bosque, donde se distinguen tres áreas de comportamientos cualitativamente distintos: el bosque aleatorio (RF), la zona de oscilaciones (DO) y la zona compleja (CF). La probabilidad de nacimiento es \( p_b = 0.5 \).

\[
s_n = 1 + (1 - \gamma \delta) + (1 - \gamma \delta)^2 + (1 - \gamma \delta)^3 + \ldots + (1 - \gamma \delta)^{n-1} + (1 - \gamma \delta)^n \frac{1}{\gamma \delta} s_0
\]

Esta expresión se puede sumar fácilmente, ya que se trata de una simple progresión geométrica,

\[
s_n = \frac{1}{\gamma \delta} + (s_0 - \frac{1}{\gamma \delta})(1 - \gamma \delta)^n
\]

Si \( \gamma \delta \ll 1 \), entonces \( \Delta \approx 1 \), y sólo hallaremos alturas de valor entero (más \( s_0 \)). Si \( \gamma \delta \) es una cantidad pequeña, pero no despreciable, la distribución de árboles presentará picos a alturas aproximadamente enteras, "moduladas" por el retraso que introduce \( \gamma \delta \). Este comportamiento se desvanece a medida que nos acercamos a las fronteras de este dominio.

**B. Bosque Aleatorio, RF.** La parte derecha del espacio de parámetros se denomina Bosque Aleatorio por la forma prácticamente independiente en que los árboles crecen en esta zona. No es el resultado actual de una interacción prácticamente nula, sino de la ausencia de vecinos, dado que la probabilidad de muerte toma ahora valores elevados. El crecimiento de cualquier árbol se ve rápidamente interrumpido, lo que conduce a una distribución casi estocástica de claros. La medida de la dimensión fractal proporciona un valor \( D_f = 2 \), reforzando la imagen anterior. Si nos mantenemos cerca de la frontera izquierda, aún es posible observar una cierta estructura (a pesar de la dimensión fractal de valor entero). Esta estructura se traduce en una dispersión no nula del espectro multifractal. Recordemos que, como se había visto en el capítulo dedicado a objetos fractales, los extremos de la función \( f(\alpha) \) (que daba las dimensiones fractales asociadas a los exponentes \( \alpha \) de divergencia de la medida) eran \( f(\infty) \) y \( f(-\infty) \), y corresponden a la dimensión fractal asociada a las zonas más y menos densas del conjunto, respectivamente. La diferencia entre los dos valores de \( \alpha \) en estos puntos,

\[
\Delta_{df} = \alpha(-\infty) - \alpha(+\infty)
\]

es una medida de la diversidad de tamaños disponibles en el sistema. Si una distribución es completamente aleatoria, \( \Delta_{df} = 0 \), mientras que cualquier inhomogeneidad se reflejará en un
valor no nulo, independientemente de la dimensión fractal del sistema. En el modelo se ha medido el espectro multifractal para los claros. El valor de $\Delta_{\mu}$ disminuye progresivamente a medida que $\mu$ aumenta. Cuando $\mu = 1$, el espectro multifractal colapsa en un valor único.

La distribución de árboles $F(h)$ que se obtiene en este dominio decae rápidamente, y los árboles grandes son prácticamente inexistentes.

C. Bosque Complejo, CF. Este es el dominio más interesante del espacio de parámetros. El criterio para acotar esta zona viene dado por la transformada de Fourier de las fluctuaciones en la biomasa. La figura 8.15 representa un ejemplo de estas variaciones.

Definimos la biomasa en cada paso de tiempo como la suma de las alturas totales de los árboles en el bosque,

$$B(t) = \sum_{(i,j)} s_t(i,j)$$

El espectro de Fourier es de la forma $f^{-\phi}$, con $\phi > 0.85$ en todo el dominio. Es decir, hemos seleccionado todos los parámetros que dan fluctuaciones cercanas a ruido $1/f$.

Las funciones multifractales identifican aquí las más amplias zonas de diversidad de tamaños, y la dimensión fractal es siempre no entera. Obsérvese el espectro multifractal en la figura 8.16.

Una medida muy útil de realizar sobre conjuntos de puntos, y que aporta interesante información sobre los mismos, es la función de correlación de dos puntos. En el caso que nos ocupa, se puede utilizar el criterio siguiente. Calcularemos la función de correlación de dos puntos como la media en el número de vecinos ($N_n$) a una distancia $d$ de una celda ocupada por un claro. $d = |r - r'|$, donde $r$ y $r'$ representan la posición de un punto y los situados a distancia $d$, respectivamente. La suma se realiza sobre todos los puntos ocupados por claros que se encuentren a una distancia $d$, de la frontera del sistema tal que verifique $d_b \geq d$. El número total de puntos que verifica esta condición es $N_P(d)$. El factor de normalización de la función, $N$, (habitualmente comprendida entre 0 y 1) es $N_P(d)$ veces el número máximo de vecinos ($N_{max}^n$) a esta distancia:
Figura 8.16: Dimensiones multifractales para la distribución de claros en el juego del bosque con los siguientes parámetros: $p_b = 0.5$, $p_d = 0.013$, $\gamma = 2.5$. La línea continua corresponde al caso real de Barro Colorado.

\[ C(d) = \frac{1}{N} \sum_{|x'-x|} \rho(x)\rho(x') \]

$\rho(x) = 1$ para una celda que contenga un claro, y 0 en otro caso.

La función $C(d)$ es potencial en todo el dominio. También se obtuvieron leyes de escala para la distribución de tamaños de claros y para la de tamaños de árboles (véanse las figuras 8.17, 8.18 y 8.19).

D. El área en blanco (WA) representa una transición entre los tres dominios previamente descritos. Se pueden hallar estructuras parcialmente fractales y correlaciones débiles, así como leyes potenciales en las fluctuaciones de biomasa que se desvían cada vez más del ruido $1/f$ a medida que nos alejamos de CF. La distribución de tamaños de árboles cambia también, de la ley potencial a una ley exponencial. Presenta pequeñas oscilaciones cuando nos acercamos a DO, y decae rápidamente al movernos hacia RF.

De los resultados previos se deduce que el modelo anterior puede presentar diversos estados de equilibrio dinámico. Existe en particular un dominio que presenta propiedades críticas. Afortunadamente, ha sido posible comparar nuestros resultados con datos obtenidos de selvas reales. El mismo estudio geométrico realizado sobre las distribuciones de claros del modelo se efectuó sobre el mapa de BCI mostrado en la figura 8.13. Se encontró una combinación de parámetros en particular que reproduce todos los valores cuantitativos que se habían obtenido para BCI. Estos son el conjunto

\[ \Omega^* = \{ p_b = 0.5, \ p_d = 0.013, \ \gamma = 2.5 \} \]

Véase la comparación en las gráficas 8.13, 8.16, 8.18, 8.19.

Además, el modelo permite la predicción del valor de algunas otras magnitudes que no es posible medir en el sistema real. Por ejemplo, si ajustamos todas las estructuras observadas con los parámetros del sistema, podemos esperar que la biomasa fluctúe como ruido $f^{-\phi}$, con $\phi \approx 1$. En particular, para el conjunto de parámetros $\Omega^*$ obtenemos $\phi = 1.02 \pm 0.03$, que es el valor más próximo a $1/f$ que proporciona el modelo. La distribución de tamaños de árboles no es de
Figura 8.17: Distribuciones de tamaños de árboles para diferentes combinaciones de parámetros en el modelo. De izquierda a derecha, $p_d = 0.001, \gamma = 0.01; p_d = 0.013, \gamma = 2.5; p_d = 0.6, \gamma = 0.1$.

Figura 8.18: Distribución de claros para BCI y para ciertos parámetros del juego del bosque ($\Omega^*$). Los exponentes coinciden dentro del margen de error.
Figura 8.19: Función de correlación de dos puntos para Barro Colorado y para los parámetros $\Omega^*$ del juego del bosque. Los cambios de pendiente a aproximadamente 30 y 60 m reales deben corresponder a un cambio en la geometría y posiblemente en la dinámica de las selvas.
Figura 8.20: Distribución de tamaños de árboles en una zona de 5 hectáreas en el Amazonas Central, con medidas tomadas durante 5 años consecutivos. El tamaño no se mide por la altura en este caso, sino por el diámetro del árbol a la altura del pecho (dbh, aproximadamente a 130 cm del suelo). Se sabe que dbh y tamaño siguen una relación de tipo potencial. El exponente que los relaciona permitiría comprobar si también los exponentes para las distribuciones de tamaños de árboles coinciden en los sistemas reales y en el modelo.

momento un dato disponible para BCI, pero otras selvas tropicales sí presentan un escalamiento potencial.

8.6 Un modelo de modelos

Presentamos finalmente un último ejemplo, un modelo minimal que se autoorganiza en un estado crítico. No pretendemos que represente ningún sistema real, pero es seguro que el lector encontrará reminiscencias de los modelos anteriores, e incluso es posible que halle algún sistema que podría ser representado por este modelo.

Supongamos de nuevo que trabajamos sobre una red cuadrada, las celdas de la cual pueden estar vacías u ocupadas, y en este último caso, activas o inactivas. Las reglas del juego son las siguientes:

1. Una celda activa en el instante \( t \) se desactiva en el instante \( t + 1 \).
2. Las celdas inactivas desaparecen (quedan vacías) cuando tienen un y sólo un vecino activo.
3. Las celdas activas se crean cuando tienen un único vecino activo con una celda inactiva en la posición opuesta.

Este es un juego sencillo de programar en un ordenador. La visión dinámica de la propagación de la actividad a través del sistema dice mucho más que la enumeración de las reglas. Por ejemplo, se observa una propagación de las parejas activa-inactiva, que se desplazan por el tablero como semáforos móviles. Cuando una de las celdas activas que se está desplazando con su cola inactiva "colisiona" con una celda inactiva solitaria, eso provoca la aparición de nuevas celdas activas, con frecuencia en direcciones diferentes a la original de propagación. Las avalanchas de diversos tamaños se entienden y se visualizan perfectamente en este modelo. La actividad a partir de un
Figura 8.21: En la parte superior se representa el número de celdas afectadas como resultado de una única activación. Esta señal es la imagen temporal típica correspondiente a un sistema crítico. En la parte inferior se representa la frecuencia de cada evento en función de su tamaño.
estado inicial arbitrario se disipa rápidamente. Para que no cese la actividad, se activa aleatoriamen
tente una celda del sistema cuando todas están inactivas. Esto permite calcular, igual que se había
hecho con la pila de arena, el tamaño de un suceso. Calculamos el número de celdas afectadas por
una única activación y representamos esta cantidad en función del tiempo, o bien calculamos su
distribución.

Es interesante hacer un salto de escala en estos sistemas y pensar que únicamente disponemos
de información local. Fijamos nuestra atención en una única celda del sistema y observamos la
sucesión de estados por los que pasa a lo largo de un cierto intervalo de tiempo. En la señal
obtenida no existe ningún tipo de periodicidad, y debemos advertir que la información de un solo
punto del espacio no permite, en los sistemas críticos, inferir nada sobre el comportamiento que el
sistema presenta globalmente.

8.7 La predicción en SOC. Conclusiones

Podemos decir, en general, que la predicción que nosotros podamos realizar en un cierto sistema
dependía de la forma en que una pequeña imprecisión inicial crece con el tiempo. Se ha visto en el
capítulo sobre caos que la desviación de dos condiciones iniciales próximas en el instante inicial es
de tipo exponencial,

$$\Delta \propto e^{\lambda t}$$

donde $\Delta = F(x(t)) - F(x(t))$ es la separación entre las imágenes de las dos condiciones $x(t_0)$ y
$x_0(t_0)$ en el instante $t$ según el flujo dinámico $F(x)$ considerado, e inicialmente $x(t_0) - x_0(t_0) = \delta$,
con $\delta \approx 0$. El valor $\lambda$ es el exponente de Lyapunov del sistema, y cuando $\lambda > 0$ el sistema es
caótico, presenta la llamada sensibilidad a las condiciones iniciales y se ha perdido la posibilidad
de predicción más allá del tiempo característico

$$\tau \approx \frac{1}{\lambda}$$

En los sistemas críticos autoorganizados, la desviación de las dos condiciones inicialmente
próximas se produce solamente como una ley potencial.
Figura 8.23: Separación de dos condiciones iniciales respecto de una trayectoria patrón (línea sólida). Las líneas punteadas se separan de ella según una ley potencial o exponencial, tal y como se indica. La separación potencial es infinitamente más lenta que la exponencial.

\[ \Delta \propto t^\lambda \]

caso en el que decimos que el sistema es débilmente caótico. Si medimos el exponente de Lyapunov para un sistema crítico, obtendremos un valor cero. Por otra parte, hemos visto que la dinámica no se repite nunca. Entendemos ahora que se diga que el sistema se debate entre el orden y el caos: su exponente de Lyapunov es cero (y uno pensaría inmediatamente que el sistema debe de ser periódico) pero nunca repite la trayectoria en el espacio de fases (con lo cual no es periódico). Existe una paradoja aparente, que no es tal, puesto que el exponente nulo es perfectamente compatible con la divergencia potencial.

Los sistemas críticos se hallan exactamente en la frontera del caos, donde las perturbaciones crecen linealmente, en vez de exponencialmente. Nuestra habilidad de predicción se pierde con el tiempo, pero mucho más lentamente de lo que se pierde en un sistema que presente caos desarrollado (\( \lambda > 0 \)). Otra implicación importante es que ahora no existe un tiempo característico a partir del cual la predicción se haga imposible. En principio, la información se pierde de forma suficientemente lenta como para que la predicción sea posible en cualquier instante futuro (si disponemos de la información adecuada sobre el instante actual).

Consideremos un ejemplo práctico. Supongamos que con las observaciones climáticas realizadas en cuatro observatorios fuese posible predecir el tiempo que hará durante dos días. Supongamos que deseamos predecir el tiempo de los cuatro días siguientes. Si el tiempo climático es crítico, esta observación sólo podrá realizar con ocho observatorios, y sería necesario disponer de diecisésis para predecir a ocho días. Parece razonable. Pero, ¿qué pasa si la divergencia es exponencial? Pues entonces necesitamos diecisésis observatorios para predecir a cuatro días, y sesenta y cuatro para ocho días. La serie es rápidamente divergente, y la predicción infinita es inabarcable físicamente.

Hemos visto que los sistemas críticos autoorganizados son invariantes bajo cambios de escala espacial y temporal. Habitualmente, es esta invariancia la única evidencia del estado crítico del sistema: en los casos que observamos en la naturaleza no disponemos de ningún parámetro de
control. Queda aún por determinar si estos sistemas reales podrán tambíem ser agrupados en clases de universalidad en función de su simetría y su dimensión, por ejemplo, al igual que se había hecho con los fenómenos críticos que presentaban sistemas clásicos de la física.

Animamos al lector a que explore todo lo que le rodea e intente identificar algunas de las características que hemos descrito. Aún queda mucho trabajo por realizar en este campo.

Bibliografía


Capítulo 9

Autómatas Celulares

Hemos analizado con anterioridad sistemas dinámicos discretos y continuos, viendo cómo la estabilidad de distintos comportamientos aparece y desaparece a través de puntos de bifurcación. Hemos visto que existen tres tipos básicos de comportamiento. En primer lugar, atractores puntuales, para los cuales la dinámica converge hacia un estado de equilibrio caracterizado por la constancia de todas las variables a lo largo del tiempo. En segundo lugar, comportamiento periódicos y cuasiperiódicos, completamente predictibles y que muestran cambios regulares a lo largo del tiempo. Finalmente, hemos realizado una introducción detallada a los sistemas dinámicos caóticos que, sorprendentemente, exhiben propiedades dinámicas enormemente complejas incluso cuando el número de variables implicadas es muy bajo. El caos determinista, caracterizable por cantidades tales como la dimensión de correlación del atractor y sus exponentes de Lyapunov, muestra, como hemos visto, una notable propiedad: su sensibilidad a las condiciones iniciales.

Formalmente, podríamos incluir a los sistemas dinámicos continuos dentro del conjunto de modelos del tipo:

\[ \frac{dX}{dt} = f(X) + \sigma(X)\xi \]

con \( X = (x_1, \ldots, x_n) \in \mathbb{R}^n \), siendo \( f \), en general, cierta función no-lineal y siendo \( \sigma(X)\xi \) un término que incorporaría factores estocásticos.

En este capítulo abordaremos de forma general una aproximación distinta, que ya ha aparecido con anterioridad (capítulos 7 y 8). Se trata de un tipo de modelos extensamente empleados en muchos campos, y que se caracteriza por introducir una discretización completa de todas las propiedades, que incluye el espacio, el tiempo e incluso el número de estados accesibles.

En este tipo de modelos, el sistema se halla compuesto por un cierto número de elementos que ocupan posiciones adyacentes sobre una red \( d \)-dimensional \( \lambda_d(L) = \{k\} \) (con \( k = (k_1, \ldots, k_d); 1 \leq k_j \leq L \)) y que evolucionarán en el tiempo siguiendo un conjunto \( T \) de reglas dado.

\[ S_{t+1}(r) = T \left( \{c_j\}; \{S_t(r')\} \right) \]

El conjunto de estados accesibles se halla limitado a \( k \) estados,

\[ S_t \in \Sigma = \{0, 1, \ldots, k - 1\} \]

Aquí \( T \) indica un conjunto de reglas que pueden ser deterministas o estocásticas. Los parámetros del sistema se designan por \( \{c_j\} \), y \( r' \) indica las posiciones de los autómatas vecinos a \( r \). Típicamente, el conjunto de puntos vecinos \( \mathcal{N}(r') \) estará distribuido simétricamente alrededor del punto de la red \( r \) considerado y los autómatas celulares capaces de simular sistemas reales verificarán ciertas restricciones naturales que veremos más adelante.
Figura 9.1: Reglas locales: cada elemento puede ser influenciado por sus vecinos más cercanos. (a) Autómatas unidimensionales. (b) Autómatas bidimensionales.

9.1 Automatas celulares deterministas

Los autómatas celulares deterministas (esto es, que no incorporan ningún tipo de estocasticidad en sus reglas dinámicas) fueron introducidos por primera vez en 1948 por los matemáticos húngaros von Neumann y Ulam, y más tarde analizados con cierto detalle por S. Wolfram (véase su recopilación de artículos, Wolfram, 1994). En esta sección seguiremos el tratamiento de Wolfram con el objeto de dar una visión general de este tipo de modelos.

Para simplificar consideraremos en primer lugar autómatas unidimensionales. El estado de uno de estos autómatas se indicará por \( a_i(t) \) para el elemento que ocupa la posición \( i \)-ésima (con \( i = 1, 2, ..., N \) elementos) en el instante (discreto) \( t \). La dinámica de dicho elemento se definirá en la forma

\[
a_{i}(t+1) = F[a_{i-r}(t), a_{i-r+1}(t), ..., a_i(t), ..., a_{i+r-1}(t), a_{i+r}(t)]
\]

donde \( F \) será cierta función que, para cada “vecindad” o entorno (figura 9.1)

\[
\{a_{i-r}(t), a_{i-r+1}(t), ..., a_i(t), ..., a_{i+r-1}(t), a_{i+r}(t)\}
\]

posible, definirá el nuevo estado adoptado por el autómata. Si no se dice lo contrario, supondremos que las condiciones de contorno, esto es, la definición de las reglas en los extremos del sistema, son de tipo periódico. Con ello indicamos que el vecino a la izquierda del primer autómata es \( N \) y el vecino de la derecha del autómata \( N \)-ésimo es \( N + 1 \equiv 1 \). Como vemos, cada autómata \( a_i \) interacciona con sus \( 2r \) vecinos más cercanos así como consigo mismo, luego necesitaremos especificar \( 2r + 1 \) estados de partida para definir la transición mediante \( F \). Cuando la interacción tiene lugar limitada a los dos vecinos más próximos (a derecha e izquierda, esto es, \( r = 1 \)) y sólo dos estados son posibles (\( k = 2 \)), hablaremos de autómatas elementales. Para definir completamente la dinámica de cada regla, debemos construir una tabla que especifique, para cada entorno, el nuevo estado adquirido por el elemento central. Así, para \( r = 1 \) y \( k = 2 \), indicamos las transiciones por

\[
a_{i-1}(t) a_i(t) a_{i+1}(t) \rightarrow a_i(t+1)
\]

que nos dicen cuál es el estado que adoptará el \( i \)-ésimo elemento si se da la terna \( a_{i-1}(t)a_i(t)a_{i+1}(t) \).
Una regla posible sería el conjunto de transiciones:

\[ 000 \rightarrow 0 \quad 001 \rightarrow 1 \quad 010 \rightarrow 0 \quad 011 \rightarrow 1 \]
\[ 100 \rightarrow 1 \quad 101 \rightarrow 0 \quad 110 \rightarrow 1 \quad 111 \rightarrow 0 \]

Una forma particular de definir una regla es suponer que el estado del automata viene dado por cierta función de la suma algebraica de los estados de los autómatas vecinos, a través de ciertas constantes enteras \( \{ \alpha_j \} \):

\[ a_i(t + 1) = f \left[ \sum_{j=-r}^{r} \alpha_j a_{i+j}(t) \right] \]

siendo el resultado de esta dinámica un valor entero \( a_j \in \Sigma \). Podemos recuperar las reglas antes descritas haciendo \( \alpha_j = k^{r-j} \). Los autómatas que siguen dinámicas de este tipo se denominan autómatas celulares totalizas.

Existe cierto tipo de restricciones generalmente empleadas en la definición de las reglas. Habitualmente se considera la configuración \([0, 0, ..., 0]\) como el estado fundamental de la dinámica, y se emplea la transición

\[ F[0, 0, ..., 0] = 0 \quad (9.2.1) \]

y, por lo tanto,

\[ f[0, 0, ..., 0] = 0 \quad (9.2.2) \]

Además, se suele considerar simetría en la condición definida por los vecinos. Más explícitamente, se utiliza la igualdad

\[ F[a_{i-r}, ..., a_{i+r}] = F[a_{i+r}, ..., a_{i-r}] \quad (9.2.3) \]

Los autómatas celulares que verifican las condiciones 9.2.1, 9.2.2 y 9.2.3 se denominan autómatas legales.

Las funciones \( F \) que definen las reglas pueden ser ordenadas especificando el número \( R(F) \) asociado a la regla, que viene dado por

\[ R(F) = \sum_{a_{i-r}, ..., a_{i+r}} F[a_{i-r}, ..., a_{i+r}] k^{s}, \quad s \equiv \sum_{i=-r}^{r} k^{r-i} a_{i+j} \]

y de forma similar las funciones \( f \) pueden ser especificadas por un código numérico que determina

\[ C(f) = \sum_{n=0}^{m} k^{n} f[n], \quad m \equiv (2r + 1)(k - 1) \]

En general, existirán \( k^{(2r+1)} \) posibles reglas, de las cuales (Wolfram, 1984)

\[ k^{k}, \quad k \equiv \left[ \frac{k^r + 1}{2} - 1 \right] k^{(2r+1)} \]

son legales (en el sentido previamente definido). Podemos ver que el número de reglas posibles aumenta con enorme rapidez (de forma exponencial) con \( r \).

Muy poco puede decirse, en general, del comportamiento de los autómatas celulares a partir únicamente de las reglas definidas. En este sentido, desde el punto de vista formal son difícilmente
tratables (cosa que no ocurre con muchos sistemas dinámicos) aunque numéricamente su implementación y estudio son muy simples. Los automatas celulares exhiben un conjunto reducido de tipos cualitativos de comportamiento, que han sido explorados en detalle y que resumiremos más adelante. Pero primero daremos un ejemplo de automata celular simple capaz de exhibir un comportamiento complejo que se observa en un sistema real: un bosque que muestra estructuras espaciales nada triviales.

### 9.2 Shigamare: ondas en el bosque

En ciertos bosques del Japón y de Norteamérica (bosques subalpinos de Abies) aparece un tipo de estructuras de gran tamaño que semejan ondas dentro del bosque. Estas ondas son en realidad frentes de árboles muertos. Su tamaño alcanza grandes dimensiones, y en cualquier caso afecta a centenares de árboles dispuestos a lo largo del frente. Estos frentes, además, se hallan en movimiento, a una velocidad de 0.5 - 3 metros por año. El frente de árboles muertos es ocupado rápidamente por semillas que regeneran la zona con árboles jóvenes de pequeño tamaño. Podemos observar en estos bosques que los árboles más viejos y altos se hallan adyacentes a la zona de regeneración, y una vez se sitúan en ella, se convierten en la siguiente remesa de árboles que morirán.

Esta observación podría sugerir varias causas posibles, todas ellas de origen físico, como por ejemplo la existencia de un terreno con ciertas particularidades especiales, que se reflejarían, en alguna forma, sobre la ordenación espacial de los árboles. Sin embargo, nada evidencia semejante dependencia de la geografía local. Sólo un factor introduce una causa física clara: la dirección de propagación de los frentes se da en la dirección de los vientos dominantes. Pero estas estructuras no se dan en otros bosques del mundo, en los que sin embargo también encontramos un viento dominante y un relieve similar. Tal vez debamos buscar las causas en el tipo de interacción que existe entre los árboles y en la forma en que el viento los afecta. Para ello, seguiremos el enfoque de Iwasa y sus colaboradores (Iwasa et al., 1991) que desarrollaron el modelo de ondas que presentaremos a continuación, y que llamaremos, siguiendo a estos autores, Shigamare (ondas de regeneración, en japonés).

El modelo parte de una red bidimensional sobre la que se disponen al azar los “áboles”, cuyo tamaño (en altura) indicaremos por una variable discreta $S_t(i,j) \in \mathbb{N}$ en la que $t = 0,1,2,\ldots$ indica el tiempo y el par $(i,j)$ indica la posición sobre la red $L \times L$ ($1 \leq i, j \leq L$). Supondremos que el viento posee una dirección dada, que incorporaremos de alguna forma a la interacción entre autómata/árbol. Siguiendo las observaciones de campo, debemos introducir el efecto observado de desecación de los árboles expuestos al viento. Tenemos por tanto una causa de mortalidad (posible) asociada con el viento y con la ausencia de protección frente a éste. Esta protección existirá si un árbol dado posee unos vecinos lo bastante grandes como para apantallar el viento y protegerlo.

Con esta información como punto de partida, podemos plantear el siguiente modelo. Cada árbol crecerá según

$$S_t(i,j) = S_t(i,j) + 1 \iff S_t(i,j) - S'_t(i,j) > d$$

siendo $S'_t(i,j)$ el siguiente promedio:

$$S'_t(i,j) = \frac{1}{1 + 2\alpha} \left[ S_t(i - 1, j) + \alpha \left( S_t(i + 1, j - 1) + S_t(i + 1, j + 1) \right) \right]$$

en caso contrario, se producirá la muerte del árbol:

$$S_t(i,j) = 0 \iff S_t(i,j) - S'_t(i,j) \leq d$$
Figura 9.2: Evolución temporal del autómata Shigamare, que genera espontáneamente ondas que se propagan. Indicamos el estado del sistema en un instante de tiempo determinado \((t = 300)\). Partiendo de una condición inicial arbitraria, el sistema alcanza un estado final caracterizado por la formación de frentes. La altura es proporcional al tono de gris, y el máximo de altura corresponde a las casillas en blanco.

Esta elección particular de los vecinos corresponde a un viento que va del oeste al este sobre la red (de izquierda a derecha). En términos intuitivos, la regla definida arriba nos dice que el árbol situado en la posición \((i, j)\) podrá crecer siempre y cuando su altura no sobrepase la media (antes definida) de los tamaños de sus vecinos posteriores (en relación al viento). En caso contrario, el árbol queda desprotegido y muere. La constante \(\alpha \in [0, 1]\) se introduce para dar un peso a la contribución de los árboles situados detrás del \((i, j)\)–ésimo y a ambos lados de éste. El árbol inmediatamente detrás de \(S(i, j)\) es el que le da la máxima protección. La constante \(d\) define el valor crítico de la diferencia entre las alturas que define la muerte.

El resultado de una simulación para \(\alpha = 0.5\) y \(d = 3\) se muestra, en tres instantes distintos, en la figura 9.2 (a-c). Partimos de una configuración inicial al azar en la que distribuimos árboles de tamaño arbitrario sobre la red, y que por lo tanto no define ninguna coherencia espacial. Con el curso del tiempo, vemos que aparecen estructuras bien definidas que terminan por hacerse muy coherentes, dando lugar a un frente que se propaga en la dirección del “viento” y que incluye a muchos elementos a la vez. La estructura y propiedades de esta dinámica se corresponden muy bien con las observadas en el bosque real (Iwasa et al., 1991). Vemos así cómo un modelo enormemente simple en el que reemplazamos los elementos reales (árboles complejos) por conjuntos de autómatas discretos que interactúan de forma igualmente simple generan patrones macroscópicos de forma autoorganizada. He aquí un ejemplo simple pero claro de las posibilidades de esta aproximación.

### 9.3 Caracterización cualitativa

Existen cuatro tipos cualitativos de dinámica asociados a los autómatas celulares. Aunque los patrones variarán en el espacio y el tiempo en distintas formas, específicas para cada función,
podemos clasificarlos en cuatro categorías, conocidas como *clases de Wolfram* (Wolfram, 1984). En todos estos casos, suponemos que el estado inicial es aleatorio (una secuencia escogida al azar) o bien se trata de una cadena de automatas en estado 0 y algunas posiciones escogidas al azar con un estado distinto. Se obtiene:

- **Clase I**: La evolución del sistema lleva a un estado homogéneo, sin estructuras espaciales ni temporales de ningún tipo.
- **Clase II**: la evolución del sistema da lugar a estructuras separadas de tipo estable o periódico.
- **Clase III**: la evolución da lugar a patrones caóticos. Espacialmente, surgen estructuras fractales y temporalmente observamos ciclos de longitud muy grande (dado el tamaño del sistema, no pueden ser de longitud superior a $k^N$).
- **Clase IV**: La evolución genera estructuras complejas localizadas, que se propagan a través de la cadena y cuya duración aumenta exponencialmente con el tamaño del sistema.

Las tres primeras clases se corresponden cualitativamente con los tres tipos de comportamientos observados en sistemas continuos y en este sentido vemos que los automatas celulares reflejan las mismas posibilidades dinámicas. En la figura 9.3 vemos distintos ejemplos de estos tipos de dinámicas.

En la tabla I damos las frecuencias de aparición de distintas clases para distintas combinaciones de $k$ y $r$. Al aumentar estos valores, la clase III es la que domina, siendo las dos primeras menos frecuentes. Como vemos, la clase IV es comparativamente rara (se halla ausente para $k = 2$ y $r = 1$).

<table>
<thead>
<tr>
<th>Clase</th>
<th>$k = 2, r = 1$</th>
<th>$k = 2, r = 2$</th>
<th>$k = 2, r = 3$</th>
<th>$k = 3, r = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.50</td>
<td>0.25</td>
<td>0.09</td>
<td>0.12</td>
</tr>
<tr>
<td>2</td>
<td>0.25</td>
<td>0.16</td>
<td>0.11</td>
<td>0.19</td>
</tr>
<tr>
<td>3</td>
<td>0.25</td>
<td>0.53</td>
<td>0.73</td>
<td>0.60</td>
</tr>
<tr>
<td>4</td>
<td>0.00</td>
<td>0.06</td>
<td>0.06</td>
<td>0.07</td>
</tr>
</tbody>
</table>

Tabla I: Frecuencias aproximadas de reglas de AC totalistas en cada una de las clases de Wolfram

Aunque aquí nos concentraremos en los automatas unidimensionales, señalemos que los estudios llevados a cabo sobre AC de distintas dimensionalidades confirman la clasificación previa (Wolfram, 1994). No nos hemos detenido a comentar la cuarta clase de la clasificación que, como veremos, juega un papel muy importante en nuestro estudio de la complejidad.

### 9.4 Caracterización cuantitativa

En esta sección exploraremos de forma cuantitativa, mediante medidas estadísticas, las clases de Wolfram introducidas en la sección anterior.

Aunque la condición inicial es aleatoria, típicamente (salvo algunos casos especiales) los automatas celulares evolucionan hacia configuraciones estacionarias caracterizadas cualitativamente por las propiedades ya mencionadas. Esta observación nos permite suponer, razonablemente, que estas configuraciones a las que tiende el automata poseerán propiedades estadísticas estacionarias.

Supongamos que queremos caracterizar los estados de la cadena en un instante dado. Imaginemos que estudiamos las secuencias de $X$ automatas consecutivos ("bloques" de tamaño $X$) y que queremos analizar la distribución de probabilidad asociada a distintas secuencias de este tamaño.
Figura 9.3: Autómatas celulares unidimensionales de (a) Clase I, (b) Clase II, (c) Clase III y (d) Clase IV.
Tenemos $k^X$ secuencias posibles, y podemos por tanto calcular la entropía espacial del conjunto o entropía topológica, $s^{(x)}(X)$, definida por:

$$s^{(x)}(X) = \frac{1}{X} \log_k \left( \sum_{j=1}^{k^X} \theta(p_j^{(x)}) \right)$$

donde el superíndice $x$ indica que los promedios se toman en el espacio (para un instante dado) y $p_j^{(x)}$ son las probabilidades de cada secuencia de tamaño $X$, esto es,

$$\sum_{j=1}^{k^X} p_j^{(x)} = 1$$

La función en la suma es $\theta(p) = 1$ si $p > 0$ y cero en caso contrario. Podemos también definir una entropía métrica (o medida espacial de entropía) $s^{(x)}_\mu(X)$ por medio de la entropía de Boltzmann de las cadenas de tamaño $X$:

$$s^{(x)}_\mu(X) = -\frac{1}{X} \sum_{j=1}^{k^X} p_j^{(x)} \log_k p_j^{(x)}$$

La primera entropía viene determinada directamente por el número total de bloques de longitud $X$, esto es, $N^{(x)}(X)$ generados por la evolución del autómata, de acuerdo con:

$$s^{(x)}(X) = \frac{1}{X} \log_k N^{(x)}$$

En la entropía métrica, cada bloque es promediado en términos de su probabilidad asociada, de forma que el resultado depende explícitamente de la distribución obtenida (de ahí el subíndice $\mu$, que indica la medida de probabilidad empleada). La entropía métrica da una medida del contenido de información por autómata.

Las definiciones anteriores introducen las desigualdades:

$$s^{(x)}_\mu(X) \leq s^{(x)}(X) \leq 1$$

donde la igualdad entre los dos primeros términos se cumple para el caso de equiprobabilidad en la distribución de bloques $\{p_j^{(x)}\}$ y la segunda igualdad se verifica en sistemas de secuencias aleatorias, en las que los $k^X$ secuencias de $X$-bloques sean equiprobables.

Además la entropía métrica verifica la condición de sub-adiitividad

$$(X_1 + X_2)s^{(x)}_\mu(X_1 + X_2) \leq X_1s^{(x)}_\mu(X_1) + X_2s^{(x)}_\mu(X_2)$$

siendo cierta la igualdad para bloques descorrelacionados (capítulo 1).

Para autómatas celulares que presenten estacionariedad en el sentido de poseer invariancia bajo traslación en las medidas de probabilidad, podemos encontrar restricciones aún más importantes que las anteriores. Así, la probabilidad $p_i^{(x)}[a_1, ..., a_X]$ para la secuencia $[a_1, ..., a_X]$ estará dada, en general, por

$$p_i^{(x)}[a_1, ..., a_X] = p^{(x)}[a_1, ..., a_{X-1}]p^{(x)}[a_X | a_1, ..., a_{X-1}]$$

siendo la última probabilidad la condicionada a tener $a_X$ en la posición $X$—ésima sabiendo que los valores adyacentes son $\{a_1, ..., a_{X-1}\}$.
Si definimos la entropía total por:

\[ S_{\mu}^{(x)}[a_1, \ldots, a_X] = - \sum_{[a_1, \ldots, a_X]} p^{(x)}[a_1, \ldots, a_X] \log_k \left[ p^{(x)}[a_1, \ldots, a_X] \right] \]

y la entropía condicionada (capítulo 1) por:

\[ S_{\mu}^{(x)}[a_X | a_1, \ldots, a_{X-1}] = - \sum_{[a_1, \ldots, a_X]} p^{(x)}[a_1, \ldots, a_{X-1}] \log_k \left[ p^{(x)}[a_X | a_1, \ldots, a_{X-1}] \right] \]

\[ \leq S_{\mu}^{(x)}[a_1, \ldots, a_X] \]

obtenemos la desigualdad

\[ X \delta_{\mu}^{(x)}(X) \leq S_{\mu}^{(x)}(X) \leq \frac{X - 1}{X} S_{\mu}^{(x)}(X - 1) + \frac{1}{X} s_{\mu}^{(x)}(X) \]

con lo que llegamos a

\[ X \delta_{\mu}^{(x)}(X) \leq s_{\mu}^{(x)}(X - 1) \]

lo cual implica un decrecimiento monótono de las entropías métricas con tamaños de bloque \( X \) decrecientes.

Podemos además construir diversas relaciones cuantitativas entre las entropías y la dimensión fractal de los patrones generados (típicamente en clase III). Recordemos el procedimiento básico de determinación de dimensiones fractales basado en el conteo de cajas. Supongamos que partimos del intervalo unidad \([0, 1]\), dentro del cual se halla un conjunto \( \Omega \) de puntos, que pueden haber sido generados a través de cierto proceso dinámico. Dibujamos el intervalo emplanteando \( k^b \) divisiones consecutivas de longitud \( k^{-b} \), que formarán una partición y sea \( N(\Omega, b) \) el número de intervalos que contienen elementos \( x \in \Omega \). Para una partición lo bastante fina, podemos escribir (capítulo 3)

\[ N(\Omega, b) \approx (k^b)^d \]

siendo \( d \) la dimensión de Kolmogorov, o capacidad del conjunto \( \Omega \) (Wolfram, 1984). Para un conjunto finito de puntos, tendremos \( d \to 0 \) (matemáticamente hablando), en tanto que un conjunto de Cantor nos dará cierto valor finito.

Esta definición permite construir la dimensión del conjunto, dada por

\[ d = \lim_{b \to \infty} \frac{1}{b} \log_k \left[ N(\Omega, b) \right] \]

que, excepto en casos patológicos, coincidirá con la dimensión de Hausdorff en el límite \( b \to \infty \). Esta última expresión puede ser aplicada de forma directa a las configuraciones obtenidas de la evolución de los autómatas celulares. Podemos tomar bloques de tamaño \( b \) que definirán el tamaño de la partición, y la definición de entropía topológica nos permite de hecho escribir la dimensión como

\[ d^{(x)} = \lim_{X \to \infty} s^{(x)}(X) \]

Para un autómata caótico, lo bastante desordenado como para que todas las posibles secuencias ocurran con probabilidad no nula, tendremos que \( d^{(x)} = 1 \) (como era de esperar). De forma similar, una dinámica que de estados homogéneos tendrá \( d^{(x)} = 0 \). Estos valores se obtendrán en general por medio de promedios para valores grandes de \( X \). Puede demostrarse que la expresión anterior
se expresa en una forma más apropiada (sobre todo si los conjuntos de bloques se obtienen para \( X \) pequeños)

\[
d^{(\varepsilon)}(x) = \lim_{\lambda \to -\infty} \frac{X s^{(\varepsilon)}(X)}{(X-1)s^{(\varepsilon)}(X-1)} = \lim_{\lambda \to -\infty} \log_k \left( \frac{N^{(\varepsilon)}(X)}{N^{(\varepsilon)}(X-1)} \right)
\]

La dimensión topológica (o de conjunto) puede emplearse para caracterizar el conjunto de configuraciones que aparecen en la evolución estacionaria del automata celular. Una dimensión métrica puede obtenerse en la forma

\[
d^{(\varepsilon)}_\mu = \lim_{X \to -\infty} s^{(\varepsilon)}_\mu(X)
\]

que introducirá una caracterización de las medidas de probabilidad sobre las configuraciones y que coincide con la información promedio por símbolo. Entre otras desigualdades, notemos que

\[
0 \leq d^{(\varepsilon)}_\mu \leq d^{(\varepsilon)} \leq 1
\]

Para una secuencia completamente aleatoria en la que todas las \( X \)-secuencias son equiprobables, se tendrá que \( d^{(\varepsilon)}_\mu = 1 \). Cualquier evolución que introduzca correlaciones (que es el caso más general) nos dará \( d^{(\varepsilon)}_\mu < 1 \).

Si nos interesa disponer de una medida de entropía espacial, en la que no tengamos en cuenta la evolución temporal, sino las estructuras generadas observadas en un instante dado, emplearemos las entropías espaciales que hemos descrito (entropías espaciales topológica y métrica). Podemos de forma similar definir unas entropías temporales que caracterizarán las secuencia de estados generados en una posición dada a lo largo de la evolución del automata. En este caso, si seguimos la evolución del sistema durante \( T \) iteraciones, las probabilidades formarán un conjunto \( \{p^{(j)}_j\} \) con \( k^T \) secuencias posibles. De forma análoga a lo visto al principio de esta sección, podemos definir la entropía topológica temporal en la forma

\[
s^{(t)}(T) = \frac{1}{T} \log_k \left( \sum_{j=1}^{k^T} \theta(p^{(t)}_j) \right)
\]

con la normalización \( \sum_{j=1}^{k^T} p^{(t)}_j = 1 \). La entropía métrica temporal correspondiente será

\[
s^{(t)}_\mu(T) = -\frac{1}{T} \sum_{j=1}^{k^T} p^{(t)}_j \log_k p^{(t)}_j
\]

Ambas entropías satisfacen desigualdades análogas a las ya obtenidas para las entropías espaciales. De la misma forma, existe una definición para las dimensiones,

\[
d^{(t)} = \lim_{T \to \infty} s^{(t)}(T) \quad d^{(t)}_\mu = \lim_{T \to \infty} s^{(t)}_\mu(T)
\]

Si la evolución del automata es periódica, de forma que cada elemento de la red toma valores que se repiten al cabo de cierto número de iteraciones, tendremos que

\[
d^{(t)} = d^{(t)}_\mu = 0
\]

Podemos, a partir de las definiciones anteriores, dar una tabla de valores característicos para las tres primeras clases de Wolfram:
<table>
<thead>
<tr>
<th>Clase (Wolfram)</th>
<th>Entropía espacial</th>
<th>Entropía temporal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clase I</td>
<td>$d_{x}^{i} = 0$</td>
<td>$d_{y}^{(i)} = 0$</td>
</tr>
<tr>
<td>Clase II</td>
<td>$d_{x}^{(i)} &gt; 0$</td>
<td>$d_{y}^{(i)} = 0$</td>
</tr>
<tr>
<td>Clase III</td>
<td>$d_{x}^{i} &gt; 0$</td>
<td>$d_{y}^{(i)} &gt; 0$</td>
</tr>
</tbody>
</table>

(estas cantidades se hallan indefinidas para los autómatas de clase IV, como veremos más adelante).

Una generalización de las anteriores entropías nos permite considerar ambas dimensiones (espacial y temporal). Sean ahora los conjuntos de probabilidades $p_{i}^{(t,x)}$ de las estructuras de longitud espacial $X$ y temporal $T$ (de las que tendremos $k^{TX}$ posibles). Podemos definir una nueva entropía topológica, dada por:

$$s^{(t,x)}(T;X) = \frac{1}{T} \log_{k} \left( \sum_{j=1}^{k^{TX}} \theta(p_{j}^{(t,x)}) \right)$$

con la normalización

$$\sum_{j=1}^{k^{TX}} p_{j}^{(t,x)} = 1$$

y la entropía métrica

$$s^{(t,x)}_{\mu}(T;X) = -\frac{1}{T} \sum_{j=1}^{k^{TX}} (p_{j}^{(t,x)}) \log_{k}(p_{j}^{(t,x)})$$

que verificarán, en particular

$$s^{(t)}_{\mu}(T) = s^{(t,x)}_{\mu}(T;1)$$

$$s^{(x)}_{\mu}(T) = \frac{1}{X} s^{(t,x)}_{\mu}(1;X)$$

Podemos definir otras muchas expresiones de interés (Wolfram, 1985). Una de ellas es el exponente de Lyapunov asociado a la evolución del autómata celular. Podemos ver de qué forma se define dicho exponente llevando a cabo un simple experimento numérico que ilustra la sensibilidad a las condiciones iniciales presentada por algunos de los autómatas celulares estudiados, en particular los de clase III. Tomemos autómatas con dos estados ($k = 2$) y, para distintas reglas, dos configuraciones iniciales al azar. $\{S_{0}(i)\}$ y $\{S_{0}'(i)\}$. Esta configuración de juegos diferirá en una posición, que elegiremos como posición central. Para dar un ejemplo específico, tomemos la regla 22 de los autómatas elementales, que podemos escribir en este caso en la forma compacta:

$$S_{t+1}(i) = \begin{cases} 1 & si \quad S_{t}(i-1) + S_{t}(i+1) + S_{t}(i) = 1 \\ 0 & si \quad S_{t}(i-1) + S_{t}(i+1) + S_{t}(i) \neq 1 \end{cases}$$

y que mostramos en la figura 9.4 (a) (es un autómata caótico). Ahora, representaremos a lo largo del tiempo la dinámica de las diferencias entre el estado de la primera y la segunda red (figura 9.4 (b)). Indicaremos por un punto blanco los lugares en los que $S_{t}(i) \neq S_{t}'(i)$ y por un punto negro el caso contrario. El exponente de Lyapunov $\lambda_{L}$ se define (Wolfram, 1983) como la velocidad media de propagación con la que se desplazan las diferencias a derecha e izquierda. Para la regla 22, se obtiene numéricamente un valor $\lambda \approx 0.76 > 0$ que nos indica la presencia de caos (véase Grassberger, 1986, para un estudio detallado de esta regla).
Figura 9.4: (a) Dinámica del autómata elemental con regla 22. (b) Propagación de las diferencias entre los estados de dos de estos autómatas cuyas configuraciones iniciales sólo difieren en un elemento.

Si exploráramos otros tipos de reglas, veríamos en general que para el autómata de clase II, la pequeña perturbación elegida se congela, sin propagarse. Para los autómatas de clase III, esta perturbación se propaga en ambas direcciones de forma que crea patrones con distintos grados de regularidad. Para los autómatas de clase IV, también tenemos propagación, pero ésta es mucho más impredecible: no se trata de una simple propagación constante sino que se crean estructuras complejas que mezclan porciones ordenadas y desordenadas junto con estructuras en propagación.

9.5 Computación, autómatas y lenguajes formales

Las características cuantitativas anteriores permiten una primera aproximación a las propiedades dinámicas de los autómatas celulares. Pero podemos obtener una descripción aún más completa recurriendo al formalismo de la informática teórica. Los conjuntos de configuraciones adoptados por los autómatas pueden contemplarse como lenguajes regulares, esto es, secuencias de símbolos (los valores dados en cada posición) que forman palabras de acuerdo con cierta gramática (Wolfram, 1984a; para una introducción excelente a la teoría de autómatas finitos, lenguajes formales y complejidad computacional, vease Hopcroft y Ullman, 1979). El conjunto de todas las posibles condiciones iniciales corresponde a un lenguaje formal trivial. El conjunto de configuraciones obtenidas después de un número finito de pasos de tiempo forma un lenguaje regular (Wolfram, 1984b), y las palabras que forman dicho lenguaje se corresponden con las posibles trayectorias sobre un grafo finito que representa un máquina finita.

Definir adecuadamente estos conceptos nos llevaría mucho tiempo, y no nos detendremos aquí a analizar este formalismo. Sin embargo, parece intuitivamente claro que los autómatas celulares pueden verse como computadores en los que las entradas quedan representadas por las configuraciones iniciales, que son “procesadas” a lo largo del tiempo por aplicación de las reglas previamente definidas. Bajo esta interpretación, el autómata celular incorpora la noción de datos de entrada (condiciones iniciales) y la de “programa” (regla empleada). Las reglas definidas representan el mecanismo básico de operación de un ordenador. En este sentido, podemos “ejecutar” distintos
Figura 9.5: Máquina de Turing básica: consta de un cinta de longitud infinita dividida en celdas y de un cabezal que lee una celda en cada paso.

programas (o evaluar distintas funciones) dando distintos conjuntos de condiciones iniciales. Este proceso es análogo a la evolución de una secuencia de símbolos manipulados por una Máquina de Turing (Turing, 1936; Hopcroft, 1984) que representa de hecho un modelo formal de ordenador. Pese a su simplicidad (figura 9.5) la Máquina de Turing (MT) es un modelo adecuado de cálculo. Pero en autómatas celulares, en lugar de considerar sólo un cabezal que modifica un elemento de la cadena en cada iteración, la evolución del autómata afecta simultáneamente a todas las posiciones en cada paso de tiempo. Puede demostrarse (véase Hopcroft y Ullman, 1979) que existe una Máquina de Turing Universal (MTU) que puede simular cualquier otra MT empleando para ello un programa de interpretación que describa la máquina que se pretende simular. En el mismo sentido, Wolfram conjetura que los autómatas de clase IV poseen capacidad de llevar a cabo computación universal. Un autómata celular universal (ACU) puede, en principio, simular cualquier otro autómata.

La gramática de un lenguaje formal proporciona un conjunto de reglas para generar y/o reconocer las palabras que pertenecen al lenguaje. Un ordenador idealizado (como es la máquina de Turing) puede en principio ser construido con la capacidad de implementar dichas reglas. Existe por ejemplo un autómata celular con $k = 18$ estados y $r = 1$ que es equivalente a una de las máquinas universales de Turing más simples que se conocen y, como veremos, un autómata bidimensional (el juego de la vida) que es también capaz de llevar a cabo computación universal. La demostración de que un autómata (u otro sistema dinámico) puede llevar a cabo computación universal puede obtenerse en dos formas: demostrando que podemos simular una Máquina de Turing Universal o, en una forma indirecta, demostrando que podemos construir, empleando las reglas definidas, puertas lógicas que, unidas, permiten llegar al mismo resultado. No todos los autómatas permiten llevar a cabo esta construcción. La clase IV es un candidato especialmente interesante en la medida en que estos autómatas generan estructuras de enorme complejidad que incluyen la propagación de señales (y por lo tanto de información) a lo largo del sistema. Siguiendo la segunda aproximación, nos detendremos en la posibilidad de llevar a cabo computación universal en el autómata conocido como juego de la vida.
9.6 Life: computación universal

Uno de los autómatas celulares bidimensionales más conocidos es sin duda el denominado juego de la vida (Life), ideado en 1970 por el matemático inglés John Conway. Es un buen ejemplo de cómo un conjunto de reglas muy simples posee la capacidad de generar una rica variedad de comportamientos. La idea de Conway perseguía la obtención del juego más simple posible (sin jugadores) que pudiera generar comportamientos imprevisibles. Conway y sus estudiantes llevaron a cabo un estudio exhaustivo y dieron finalmente con un modelo de enorme simplicidad en su definición formal pero de enorme complejidad en sus posibles evoluciones.

El juego de Conway está basado en un autómata celular definido sobre una red cuadrada y en la que el conjunto de estados posibles \( \Sigma \) consta de dos elementos: "vivo" (1) o "muerto" (0). Los vecinos a considerar son los cuatro más próximos. En el instante \( t \), algunos puntos de la red tendrán autómatas vivos; en el siguiente instante, el nuevo conjunto de autómatas en dicho estado quedará determinado por el siguiente conjunto de reglas:

- Un autómata vivo (1) que posea dos o tres vecinos vivos sobrevivirá en la próxima generación. En caso contrario, pasará al estado 0 (regla de muerte).
- Si un autómata en estado 0 posee exactamente tres vecinos vivos, en esa posición aparecerá un autómata vivo (regla de nacimiento).

Este sistema puede generar estructuras espaciales de gran complejidad. Algunos ejemplos de estas se muestran en la figura 9.6. Encontramos diversas estructuras características que pueden desplazarse por el sistema, conocidos como gliders, y representadas en la figura. Estas estructuras son especialmente importantes en la medida en que permiten la propagación de información a través del sistema.

Podemos demostrar que este autómata celular bidimensional posee las condiciones estructurales y dinámicas mínimas necesarias para llevar a cabo cálculo universal (Langton, 1991). Podemos construir puertas lógicas de distintos tipos (AND, OR y NOT en particular) que nos permitirían, una vez acopladas en alguna forma, construir una Máquina de Turing Universal. El primer elemento a emplear será el "glider", del que necesitaremos producir, de forma regular, un número indefinido. En la figura 9.7 (a) vemos un cañón de "gliders" (glider gun), una configuración periódica que produce un flujo constante de estos objetos con un espaciado adecuado entre ellos. Asociaremos a cada "glider" un "1" y un cero al espaciado entre ellos.

Podemos construir una puerta NOT empleando el cañón de "gliders", que interacciona con una cadena de dígitos de entrada que no son sino "gliders" (figura 9.7 (b)). Los "gliders" que definen la señal de entrada inciden perpendicularmente sobre el haz vertical. Cuando dos "gliders" colisionan, se aniquilan mutuamente, dejando tras de sí un vacío (cero), de forma que un 1 de entrada da un 0 en la salida. Cuando lo que incide es un cero (un espacio vacío) el "glider" que se desplaza hacia abajo sigue su camino, de manera que la salida al 0 es un 1: hemos obtenido una puerta NOT.

Esta estructura básica permite implementar las puertas AND y OR sin dificultad. Ahora precisamos definir dos entradas en lugar de una, pero una vez más emplearemos el cañón de "gliders" sobre el que haremos incidir las dos entradas, dadas a su vez por secuencias de "gliders" que colisionan perpendicularmente con el haz generado. En la figura 9.8 (a,b) mostramos las dos puertas construidas.

La puerta AND se implementa como sigue. Un primer haz de "gliders" incide sobre los "gliders" generados por el cañón. Como antes, el haz generado por el cañón está formado sólo por "gliders" con un cierto espaciado entre ellos, mientras que el haz B incluye "gliders" (1) y espacios en blanco (0). Como antes, si se produce colisión, tendremos aniquilación del "glider" y por lo tanto un cero en lugar de un uno. Algo más adelante incide el segundo haz que forma la señal de entrada, y nuestra salida se tomará como el resultado del haz A que atraviesa el haz de "gliders" (figura
Figura 9.6: Estructuras generadas por la dinámica del juego de la vida. Se incluye el nombre de algunas de ellas, entre las que destacan los gliders, que pueden propagarse a través del sistema.
Figura 9.7: (a) Cañón de "gliders". (b) Puerta NOT construida a partir del cañón anterior.

Figura 9.8: (a) Puerta lógica AND construida en el juego de la vida. (b) puerta OR.
9.8 (a)). No es difícil ver que la salida contiene un uno cuando ambos inputs eran uno y un cero en cualquier otro caso. Hemos obtenido por lo tanto una puerta AND. Finalmente, podemos construir una puerta OR empleando dos haces de "gliders" tal y como se indica en la figura 9.8 (b) (Langton, 1991). Si bien el último paso sería llevar a cabo la construcción efectiva de una máquina de Turing universal, los elementos de los que disponemos nos permiten hacer una conjetura de constructibilidad a partir de las puertas lógicas previamente obtenidas.

9.7 Parámetro $\lambda$ de Langton

En la clasificación anterior, hemos visto que las clases de autómatas de Wolfram cubren cuatro tipos de comportamiento cualitativamente distintos. Tres de ellos pertenecen a dinámicas que poseen una contrapartida en la descripción estándar de sistemas dinámicos: (a) Clase I: atractores puntuales, con estructura espacial nula (homogénea); (b) Clase II: atractores periódicos, con estructuras regulares que se repiten y (c) Clase III: autómatas caóticos, que presentan estructuras fractales en su evolución. La clase IV no posee una estructura simple que permita obtener promedios temporales debido precisamente a la aparición de estructuras en propagación. La definición de una numeración en las reglas no nos permite obtener una ordenación de tipo dinámico. Dos reglas adyacentes desde el punto de vista de su numeración pueden pertenecer a cualquiera de las cuatro clases. Sin embargo, sabemos que podemos definir, en sistemas dinámicos, parámetros que permitan detectar, en particular, las transiciones entre distintos tipos de comportamiento. Sería muy interesante disponer de un parámetro que permitiera ordenar los diferentes comportamientos de forma natural, que fueran de los más ordenados (clase I) a los más desordenados (clase III). Tal vez dicho ordenamiento permitiera, además, conocer cual es la posición de la clase IV respecto de las anteriores.

Dicha parametrización fue introducida por Chris Langton en 1990. Supongamos, redefiniendo el problema, una red $d$-dimensional sobre la que definimos un autómata celular que posee una vecindad $V$, donde indicaremos por $|V|$ el número de puntos de la red incluidos en dicha vecindad. Supondremos que el conjunto de estados es $\Sigma$ y que incluye $K$ estados posibles. Las reglas definidas para cada autómata se indicarán como una aplicación

$$F : \Sigma^N \rightarrow \Sigma$$

Sea ahora $D_N$ el conjunto de todas las posibles funciones $F$ para autómatas de $K$ estados y $N$ vecinos, dentro de la cual introduciremos la parametrización. El parámetro $\lambda$ de Langton se define en la siguiente forma (Langton, 1990; 1991). Sea cierto estado arbitrario $s_q \in \Sigma$ que llamaremos estado quietante. Dada una función de transición $F$, existirán $n_q$ transiciones hacia este estado en la función $F$. Supongamos que las restantes $K^N - n_q$ transiciones en $F$ se eligen al azar de entre los restantes $K - 1$ estados del conjunto $\Sigma - s_q$. El parámetro se define entonces como:

$$\lambda = \frac{K^N - n_q}{K^N}$$

De forma que podemos apreciar claramente lo que se espera obtener en casos extremos

- Si $n_q \approx K^N$, todas las transiciones en la función $F$ darán como resultado el estado $s_q$, y $\lambda \approx 0.0$. El autómata exhibirá comportamientos muy simples, y esperaremos observar autómatas de clase I o II.

- Cuando todos los estados finales están igualmente representados en las tablas de transiciones, tendremos $\lambda = 1 - 1/K$, que corresponderá a la situación de mayor heterogeneidad en la función $F$, de forma que esperaremos encontrar en este dominio los autómatas de clase III más desordenados.
Figura 9.9: Simulación de automatas celulares unidimensionales con 128 elementos, para $K = 4$, $N = 5$ y condiciones periódicas de contorno. Indicamos en cada caso el valor del parámetro $\lambda$ asociado.
Esta parametrización puede comprobarse sobre un ejemplo numérico. La adecuación del parámetro de Langton será tanto mejor cuanto mayor sea el número de estados y/o de vecinos empleados en la definición de la tabla (lo que nos permitirá llevar a cabo una estadística adecuada). En la figura 9.9 vemos el resultado de la simulación de un autómata de este tipo con $K = 4$, $N = 5$, 128 elementos y condiciones de contorno periódicas. En cada caso hemos empleado una regla distinta y se indica el valor del parámetro $\lambda$. Viendo estas simulaciones, un resultado evidente es la secuencia obtenida, que va desde dinámicas simples a dinámicas muy desordenadas. El segundo es mucho más sorprendente: en $\lambda = 0.5$ aparece la clase IV, justo en la frontera entre orden y desorden.

La aparición de complejidad en esta frontera nos es familiar: es lo que esperamos que ocurra en sistemas complejos situados en las proximidades de puntos críticos. De hecho, la ordenación obtenida a partir del parámetro $\lambda$ nos recuerda, de algún modo, lo que llamábamos una “temperatura” del autómata, en la medida en que introduce un mayor o menor grado de desorden. Podemos caracterizar cuantitativamente esta frontera en varias formas (Langton, 1990; 1991), entre las cuales está la longitud de los transitorios. Supongamos que introducimos una medida del tiempo necesario para caracterizar estatisticamente un autómata celular en relación a sus distribuciones de probabilidad estadionarias. Para autómatas de clase I y II, el estado estacionario se adquiere con rapidez, y algo similar ocurre con los autómatas de clase III. En cambio, los de clase IV, como ya hemos mencionado, presentan estructuras complejas que se propagan a través del sistema. Una misma zona puede permanecer oscilando regularmente y sufrir posteriormente una colisión con una estructura en propagación que la hace cambiar de forma caótica. Esta situación hace necesarios tiempos enormes para caracterizar apropiadamente el sistema. Esta situación no debe sorprendernos: a mayor complejidad, parece natural que debamos llevar a cabo un mayor número de medidas para caracterizar adecuadamente el estado del sistema. Para un sistema muy ordenado o desordenado, la situación se invierte. En la figura 9.10 (a) vemos un ejemplo de cálculo del tamaño de los transitorios, que es máximo en un sistema finito en las proximidades de $\lambda = 0.5$, tal y como esperaríamos de una transición de fase de segundo orden (crítica). El tamaño de estos transitorios diverge con el tamaño de la red, como vemos en la figura 9.10 (b). Nada de ello ocurre si estamos en los dominios caótico u ordenado.

Dado que el resultado final de esta parametrización es que los autómatas de clase IV aparecen en un punto crítico, Langton introduce una conjetura, que genéricamente denomina Computation en la frontera del Caos en la que establece la hipótesis de que la capacidad de cálculo emerge en la naturaleza en las proximidades de puntos críticos. En los que la computación universal se hace posible. Langton sugiere que la mezcla entre procesos que incorporan regularidades (necesarias para almacenar información) y factores que introducen desorden (necesario para manipular dicha información) justifican la aparición de esta propiedad, de manera natural, en el punto crítico.

Han surgido algunas críticas razonables a esta aproximación, que implican la inadecuación del parámetro para algunos casos explorados (Mitchell et al., 1993) y existen asimismo medidas alternativas más sofisticadas (Crutchfield y Young, 1990) que permiten caracterizar mejor las transiciones en términos computacionales. Sin embargo, no cabe duda de que el resultado de Langton es enormemente sugerente. Volveremos a él en el capítulo 16.

9.8 Autómatas celulares y medios excitables

Entre las múltiples aplicaciones dadas a los autómatas celulares, destaca la posibilidad de simular medios excitables. Un medio excitable es un sistema dinámico con capacidad para ser “excitado” por una perturbación que sobrepase cierto umbral. Después de la excitación, el sistema se vuelve “refractario”, de manera que requiere un cierto tiempo (durante el cual vuelve a su estado de reposo) para volver a ser excitado de nuevo. Esta situación se da en muchos sistemas químicos y también
Figura 9.10: (a) Dependencia de los transitorios (tiempo característico necesario para caracterizar estadísticamente el automata) con el parámetro $\lambda$. (b) Dependencia de los transitorios con el tamaño del sistema.

en sistemas vivos, en particular en las neuronas, el córtex cerebral y en el tejido cardíaco. En presencia de un espacio bidimensional o tridimensional, formado por múltiples sistemas excitables acoplados de alguna forma, observaremos la generación espontánea de ondas que se propagan. En algunas ocasiones estas ondas sirven para autoorganizar una estructura formada por miles de células, que contribuyen a formar la onda a la vez que son controladas por ésta (como indicábamos en el capítulo 2 en relación con el principio de control de H. Haken).

En el capítulo siguiente analizaremos la aparición de estructuras en sistemas de reacción-difusión, en los que espacio y tiempo son continuos. En esta sección nos limitaremos a presentar dos ejemplos de automatas celulares capaces de generar ondas espirales que se propagan.

Un primer modelo de este tipo (véase Mikhailov, 1990, para una revisión general y múltiples referencias escogidas) consiste en una red bidimensional en la que indicaremos por $\Phi_t(i,j)$ el estado de cada punto. Los nuevos estados se obtienen empleando el siguiente conjunto de reglas

$$
\Phi_{t+1}(i,j) = \begin{cases} 
\Phi_t(i,j) + 1 & \text{si} \quad 0 < \Phi_t(i,j) < \tau_e + \tau_r \\
0 & \text{si} \quad \Phi_t(i,j) = \tau_e + \tau_r \\
0 & \text{si} \quad \Phi_t(i,j) = 0 \quad y \quad u_t(i,j) < h \\
1 & \text{si} \quad \Phi_t(i,j) = 0 \quad y \quad u_t(i,j) \geq h 
\end{cases}
$$

El estado cero es el estado de reposo del automata, mientras que si

$$0 < \Phi_t(i,j) < \tau_e$$

se dice que el elemento se halla en estado excitado. Un estado refractario corresponde al conjunto

$$\tau_e < \Phi_t(i,j) < \tau_e + \tau_r$$

De acuerdo con las reglas anteriores, un elemento situado en el punto $(i,j)$ pasará del estado de reposo al estado excitado si cierta cantidad local $u_t(i,j)$ excede cierto umbral $h$, pudiendo incrementar progresivamente su estado hasta $\tau_e + \tau_r$, momento en el que el automata regresa al estado de reposo.
Figura 9.11: Autómata celular que simula un medio excitable. Los parámetros son: \( \tau_e = 4, \tau_r = 5, g = 0.5, h = 3, T = 15 \).

La cantidad \( u_t(i, j) \) se interpreta como una concentración de "activador" sobre el punto \((i, j)\). Un activador es producido por los elementos que se hallan en estado excitado, a la vez que se descompone a un cierto ritmo. Ambos efectos quedan descritos por la regla adicional definida a primeros vecinos

\[
u_{t+1}(i, j) = g u_t(i, j) + \sum_{k,l} C(k, l) I_t(i + k, j + l)
\]

donde indicaremos

\[
I_t(i + k, j + l) = \begin{cases} 
1 & \text{si } 0 < \Phi_t(i, j) < \tau_e \\
0 & \text{si } \Phi_t(i, j) = \tau_e + \tau_r \text{ o } \Phi_t(i, j) = 0 
\end{cases}
\]

Los coeficientes determinan el rango de la interacción. Para ser razonables en términos de los sistemas reales que modelizamos, deberán decayer a cero con la distancia de una forma relativamente rápida. En particular, podemos tomar interacción a primeros vecinos, con lo que restringiremos la suma anterior a los ocho vecinos más próximos. Empleando los parámetros \( \tau_e = 4, \tau_r = 5, g = 0.5, h = 3, T = 15 \), obtenemos como resultado de las reglas anteriores estructuras espaciales de gran complejidad, como las que se muestran en la figura 9.11 (a,b). Partiendo de una condición inicial con algunos lugares excitados distribuidos al azar, obtenemos al cabo de cierto tiempo ondas coherentes que se propagan por el espacio girando y colisionando.

Este tipo de sistemas ha sido estudiado en detalle por numerosos autores, que han extendido las propiedades anteriores a muchos casos reales de interés (Gerhardt y Schuster, 1989; Murray, 1989). Un ejemplo especialmente espectacular, por su capacidad de reproducción de las complejas estructuras tridimensionales que aparecen en reacciones químicas volúmicas, consiste en el siguiente conjunto de reglas (Markus y Hess, 1990). Consideremos una red bi- o tri-dimencional \((\Lambda_2, \Lambda_3)\) dividida en cuadrados (o cubos) de lado \(d\) (figura 9.12 (a)). Colocamos al azar un punto \((x, y) \in \Lambda_k\) dentro de cada uno de estos cuadrados y mantenemos esta distribución fija a lo largo de la simulación. Esta elección aleatoria se realiza con un objetivo concreto: evitar una elección particular de cierta geometría (cuadrada, hexagonal, etc.) que acabaría reflejándose en las estructuras macroscópicas formadas. El estado de cada celda, por otra parte, toma un conjunto discreto

de estados: (a) $S = 0$ (excitable); (b) $S = n + 1$ (excitado) y (c) $S \in \{n, n-1, \ldots, 1\}$ (estados refractarios). En cada celda, se dará la transición $S \rightarrow S - 1$ si no se produce excitación. Se introduce además un parámetro $S_m$ con el siguiente comportamiento: no puede darse excitación en una celda si $S > S_m$, pero sí cuando $0 \leq S \leq S_m$. La excitación de un punto se desencadena, como antes, por el estado de sus vecinos, que ahora incluyen las celdas cuyos puntos $(x, y)$ están a una distancia $\delta \leq R$ (figura 9.12 (a)). Un punto presentará excitación si el número $\nu$ de puntos excitados dentro del dominio circular (esférico) definido por $R$ es mayor que cierto umbral.

Las simulaciones de este sistema (véase Markus y Hess, 1990) dan lugar a estructuras enormemente complejas, como la que se muestra en la figura 9.12 (b). Las estructuras formadas son de gran realismo y permiten explorar la dinámica de reacciones químicas oscilantes o la propagación realista de ondas de excitación en el tejido cardíaco (véase Winfree, 1987, para una exposición detallada acerca de modelos de medios excitables y sus implicaciones). Otros modelos de medios excitables basados en autómatas celulares permiten, por ejemplo, estudiar la problemática de la turbulencia química, esto es, de la aparición de caos espaciotemporal, que abordaremos en el próximo capítulo (Oono y Kohmoto, 1985).

Para terminar, señalemos que los medios excitables, que podemos simular de varias formas (Mikhailov, 1990) pueden ser buenos candidatos a sistemas computacionales (Holden et al., 1991). Esta posibilidad de exploraremos más adelante (capítulo 16) es de gran relevancia, en la medida en que podría permitir comprender la forma en que ciertos sistemas naturales, como el cerebro, procesan información.

**Bibliografía**


Capítulo 10

Estructuras de Turing y Caos Espaciotemporal

Hemos visto hasta ahora numerosos ejemplos de sistemas que exhiben cambios en el tiempo con propiedades muy notables. Vimos cómo un mismo sistema dinámico puede, bajo ciertas condiciones, dar lugar a estados estacionarios (atractores puntuales) y bajo otras pasando por uno o varios puntos de bifurcación a oscilaciones de mayor o menor complejidad. Aunque en la discusión del modelo de Lorenz hemos introducido el fenómeno de la aparición de estructuras convectivas en una capa de fluido sometida a un gradiente de temperatura, los grados de libertad espaciales fueron hábilmente suprimidos por Lorenz para obtener un sistema de tres ecuaciones diferenciales ordinarias. En todo nuestro estudio previo, el espacio ha sido el gran ausente.

Pero a nuestro alrededor hay multitud de ejemplos en los que vemos la aparición de orden en forma de estructuras espaciales. El ejemplo de la convección de Bénard es uno de ellos, pero también lo son los patrones ordenados que emergen a lo largo de la morfogénesis en los organismos multicelulares. A partir de una sola célula inicial que experimenta divisiones consecutivas, éstas nuevas células van organizándose en el espacio y el tiempo. Aunque su número alcanza con facilidad miles o cientos de miles de elementos que se hallan en contacto entre sí en una escala local, éstas nuevas células muestran propiedades de organización que van mucho más allá de la escala de longitud de las células aisladas. En la superficie de alas de insectos o sobre las escamas de la piel de un pez observamos patrones ordenados de coloración resultantes de la diferenciación específica de grupos de células que solo parecen en contacto con sus vecinas más cercanas. Aún más espectaculares, los dibujos que observamos sobre las conchas de algunos moluscos (figura 10.1) desafían, aparentemente, cualquier explicación simple sobre su origen.

¿De dónde procede este orden? Para ser más explícitos, imaginemos una larva (hipotética) de un insecto, o el embrión temprano de algún mamífero. Ambos estarán formados por un gran número de células indiferenciadas que, a posteriori, adquieren unas propiedades específicas asociadas a su posición a lo largo del embrión. Con el paso del tiempo, aparecen estructuras ordenadas en forma de segmentos, de inicio de extremidades, de manchas más o menos regulares, etc. Pueden además llevarse a cabo experimentos simples pero enormemente reveladores que permiten acercarnos a los mecanismos de regulación del fenómeno. En la figura 10.2 mostramos un ejemplo especialmente interesante. En una fase temprana del desarrollo de un insecto, en la que no aparecen aún estructuras evidentes ya diferenciadas, atamos un hilo cerca del centro del embrión, que separa el insecto en dos mitades. El proceso de morfogénesis continúa, pero sólo se lleva a cabo en una mitad. En ésta, aparece el insecto completo aunque reducido en tamaño.

Este experimento es muy revelador de las capacidades de regulación implícitas en el proceso...
Figura 10.1: Patrones de coloración en conchas de moluscos.

Figura 10.2: Morfogénesis en un embrión de insecto (*Euscelis*). Llevamos a cabo una manipulación experimental (atando la parte media del embrión en formación) provocando una reestructuración del organismo, que aparece completo en una mitad del espacio normal.
de morfogénesis (al menos en algunos casos). Tenemos por lo tanto un sistema complejo capaz de organizarse en el espacio y el tiempo sin un control exterior. Las estructuras que aparecen son resultado de (1) sucesivas divisiones celulares que dan lugar a una estructura espacial básicamente indiferenciada (al menos en apariencia), y (2) la interacción de las células con sus células vecinas. ¿Cómo podemos interpretar teóricamente este fenómeno?

Una respuesta a la pregunta de cómo tiene lugar la aparición de estructuras en la morfogénesis fue dada por primera vez por el genial matemático inglés Alan Turing. Su artículo, Las bases químicas de la morfogénesis fue publicado en 1952 (Turing, 1952) y es uno de los clásicos de la ciencia del siglo XX. En él Turing atacó el problema planteándose la pregunta de cómo surgirían, en un medio químico homogéneo, estructuras ordenadas de forma regular. Turing empleó un modelo muy simple. Este modelo puede imaginarse como una tira de "células" dispuestas a lo largo de una línea, de forma que cada célula entra en contacto sólo con sus dos vecinas adyacentes. Imaginemos que dentro de cada célula se están sintetizando dos tipos de moléculas. Supongamos que estas moléculas, que interactúan entre sí en alguna forma, pueden difundirse hacia las células vecinas de forma pasiva. Podemos por lo tanto visualizar cada punto (célula) como un par de concentraciones de sustancias químicas (que Turing llamó morfógenos). Supongamos que, inicialmente, cada célula posee cierta concentración de morfógenos. Esta concentración será aproximadamente la misma para todas las células, si exceptuamos las pequeñas fluctuaciones inevitables en todo sistema físico. La difusión, como ya sabemos (capítulo 1) es un proceso que destruye correlaciones, haciendo el sistema más homogéneo aún. Los procesos de difusión eran bien conocidos por Turing. De hecho, su estudio es una parte prominente de la física matemática clásica. Puede parecer extraño que, para encontrar una teoría matemática acerca de la generación de estructuras, recurramos a un proceso que tiende a destruirlas. Turing demostró, sin embargo, que la difusión, junto con una interacción (no-lineal) entre los morfógenos, podía generar, de forma espontánea, orden macroscópico. Su intuición fue, en este sentido, excepcional.

10.1 Procesos de difusión

Antes de pasar al planteamiento de Turing haremos una breve revisión del fenómeno de la difusión pasiva así como de su formalización matemática. La difusión es un proceso fácilmente observable cuando mezclamos dos fluidos (digamos agua y tinta). Al poco tiempo, hemos obtenido una mezcla homogénea, resultante del movimiento al azar de las moléculas de ambos componentes. Como vimos en el capítulo 1, este movimiento puede representarse como un recorrido al azar (random walk) de las partículas dentro del espacio que éstas ocupan. El proceso de mezcla es irreversible (la entropía crece a medida que el sistema va desordenando) y el resultado final es un fluido en el que tenemos la misma concentración de ambas sustancias en cualquier volumen del espacio (siempre que éste no sea muy pequeño y alcance la escala de las fluctuaciones).

Para obtener un modelo del proceso de difusión, supongamos que tenemos un recipiente con un fluido en su interior. Imaginemos que la mitad derecha (A) contiene cierta sustancia coloreada (S), ausente en el lado izquierdo (B). Llamemos r e y a las concentraciones de S en los lados A y B (figura 10.3). Es fácil intuir que, cuanto mayor sea la diferencia de concentración entre ambos compartimentos, más rápido será el flujo de S. De aquí se sigue que la rapidez de cambio es proporcional a la diferencia de concentraciones entre ambos compartimentos. Si llamamos D a la constante de proporcionalidad, conocida como el coeficiente de difusión, tendremos las siguientes ecuaciones para la concentración en cada lado:

1A. Turing es particularmente conocido por sus aportaciones fundamentales a la la teoría de la computación. A él debemos el concepto de máquina de Turing Universal.
Figura 10.3: Difusión pasiva entre dos compartimentos.

\[
\frac{dx}{dt} = D(y - x)
\]

\[
\frac{dy}{dt} = D(x - y)
\]

pero, puesto que supondremos conservación en la cantidad de \( S \) total, que llamaremos \( C \), (esto es, \( x + y = C \)) podemos reducir el sistema anterior a un solo ecuación:

\[
\frac{dy}{dt} = D(C - 2y)
\]

que es fácilmente resoluble, proporcionando

\[
y(t) = \frac{C}{2} \left( 1 - e^{-2Dt} \right)
\]

Esta solución nos da un crecimiento asintótico en la concentración de \( S \) en el compartimento \( B \) hasta un valor final \( y(\infty) = C/2 \), es decir, la homogeneidad de concentración. Volvemos a encontrarnos con el resultado apuntado en el capítulo 1, cuando los random walkers se movían al azar desde su posición inicial (idéntica para todos) hasta llenar de forma homogénea todo el espacio.

El proceso de difusión ha destruido, como vemos, la diferencia inicial. El razonamiento anterior es fácilmente extensible a un sistema más complicado. Si consideramos un conjunto de \( N \) compartimentos conectados entre sí localmente y dispuestos a lo largo de una línea (figura 10.4), podemos generalizar el modelo anterior sin dificultad.

Sean ahora \( \{x_i\} (i = 1, \ldots, N) \) las concentraciones de \( S \) en cada punto. Ahora podemos imaginar el cambio en la concentración en el \( i \)-ésimo compartimento por:

\[
\frac{dx_i}{dt} = D(x_{i-1} - x_i) + D(x_{i+1} - x_i) = D(x_{i+1} + x_{i-1} - 2x_i)
\]

Estas ecuaciones serán válidas para \( i = 2, \ldots, N - 1 \). Observeemos que ahora, bajo la introducción de un espacio explícito, debemos definir qué es lo que ocurre en los límites de nuestro sistema: debemos establecer las condiciones de contorno. Dos condiciones típicas son: (a) flujo cero: los
límites del sistema (los compartimentos \( i = 1 \) e \( i = N \)) sólo están conectados por difusión con un vecino inmediato, esto es

\[
\frac{dx_1}{dt} = D(x_2 - x_1)
\]
\[
\frac{dx_N}{dt} = D(x_{N-1} - x_N)
\]

(b) condiciones periódicas de contorno: este caso (considerado por Turing en su artículo original) supone que los extremos están conectados entre sí (el sistema forma un anillo). Entonces tenemos:

\[
\frac{dx_N}{dt} = D(x_{N-1} + x_1 - 2x_N)
\]
\[
\frac{dx_1}{dt} = D(x_N + x_2 - x_1)
\]

Estas ecuaciones permiten simular numéricamente el comportamiento de las concentraciones iniciales a lo largo del tiempo. Es de hecho el equivalente, continuo en \( z \), del experimento de simulación numérica llevado a cabo en el capítulo 1. Partiendo de una condición inicial al azar, en la que tenemos fluctuaciones en las concentraciones de cada compartimento, la evolución dinámica lleva a una total eliminación de éstas.

10.2 La ecuación de difusión

El último paso antes de introducir las ecuaciones de Turing será la obtención de la llamada ecuación de difusión, una de las ecuaciones clásicas de la física matemática. Consideremos el siguiente problema: supongamos una línea continua a lo largo de la cual se define cierta variable \( C(x,t) \) para cada punto \( x \) y cada instante de tiempo \( t \). Esta variable puede ser una concentración de morfógeno o una temperatura (ambas obedecen el mismo tipo de ecuación). Observemos que lo que planteamos es de hecho la contrapartida de espacio continuo (figura 10.5) correspondiente a la versión anterior formada por \( N \) compartimentos.

Deduciremos a continuación la expresión matemática asociada a la difusión sobre un espacio continuo en el que supondremos que el transporte tiene lugar preferentemente a lo largo de un eje dado (digamos el eje \( x \)). La ecuación resultante contiene derivadas parciales respecto de la posición y del tiempo. Se trata por lo tanto de una ecuación diferencial en derivadas parciales.

El valor de \( C(x,t) \) será a partir de ahora la densidad (concentración) de partículas por unidad de volumen en una pequeña zona centrada en \( x \). Supondremos en adelante que \( C(x,t) \) es una función continua de \( t \) y de \( x \). Consideremos, siguiendo el esquema de la figura 10.5, que existe
cierto número de partículas en el interior de un paralelepípedo $P$ limitado, en el eje $x$, por los planos $z = x_0$ y $z = x_0 + L$. Supondremos además que se cumple la siguiente ley de balance: la tasa de cambio del número de partículas en $P$ es igual a la tasa de creación neta de partículas más la tasa neta de flujo de partículas a través de las fronteras de $P$.

Sea $Q(x, t)$ la tasa neta de creación, y llamemos $J(y, t)$ a la tasa de flujo de partículas a través del plano $x = y$. Tomaremos el flujo como positivo si se da en el sentido positivo de las $x$ y negativo en caso contrario. Bajo estas condiciones, la ley de balance se escribirá de forma general como:

$$\frac{\partial}{\partial t} \int_{x_0}^{x_0 + L} C(x, t) dx = J(x_0, t) - J(x_0 + L, t) + \int_{x_0}^{x_0 + L} Q(x, t) dx$$

$\forall x_0, L$. Se ha considerado que el flujo es básicamente unidimensional, y que tiene lugar entre los planos situados en $x = x_0, L$. Empleando el teorema integral del valor medio obtenemos

$$\frac{\partial}{\partial t} C(q_1, t) L = Q(q_2, t) L + J(x_0, t) - J(x_0 + L, t)$$

siendo $x_0 \leq q_1 \leq x_0 + L$ y $x_0 \leq q_2 \leq x_0 + L$. Para obtener la ecuación de difusión, debemos dividir por $L$ y tomar el límite para $L \to 0$. Obtenemos, para un $x$ arbitrario

$$\frac{\partial C(x, t)}{\partial t} = Q(x, t) - \frac{\partial J(x, t)}{\partial x}$$

Ahora, consideremos el caso más simple, en el que las partículas son "inertes", en el sentido de que no son creadas ni destruidas. No hay por lo tanto término de creación, y podemos tomar $Q = 0$. Ahora debemos hallar la expresión más simple asociada al movimiento pasivo de estas partículas. En primer lugar, consideremos el desarrollo en Taylor de $C(x + h, t)$

$$C(x + h, t) = C(x, t) + h \frac{\partial C(x, t)}{\partial x} + \frac{h^2}{2!} \frac{\partial^2 C(x, t)}{\partial x^2} + ...$$

Si suponemos (razonablemente) que el flujo $J(x, t)$ depende de las concentraciones cerca de $x$, vemos que de hecho este término deberá depender de la diferencia de concentraciones entre los puntos considerados, esto es, de $\partial_x C$. Por lo tanto, emplearemos

$$J = -D \frac{\partial C}{\partial x}$$

lo que nos da, bajo las hipótesis previas, la ecuación de difusión (ley de Fick)
\[ \frac{\partial C}{\partial t} = \frac{\partial}{\partial x} \left[ D(C) \frac{\partial C}{\partial x} \right] \]

Si suponemos, como ocurrirá a menudo, que el coeficiente de difusión \( D(C) \) es constante, entonces obtendremos la expresión usual de la ley de difusión

\[ \frac{\partial C}{\partial t} = D \frac{\partial^2 C}{\partial x^2} \]

(la cual se corresponde con la ecuación diferencial deducida al final del capítulo 1 en términos de procesos estocásticos).

La ecuación de difusión anterior puede ser fácilmente generalizada para dos y tres dimensiones. Tenemos en estos casos

\[ \frac{\partial C}{\partial t} = D \left[ \frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} \right], \quad (d = 2) \]

\[ \frac{\partial C}{\partial t} = D \left[ \frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} + \frac{\partial^2 C}{\partial z^2} \right], \quad (d = 3) \]

De forma compacta indicaremos los términos de difusión mediante el empleo del operador laplaciano \( \nabla^2 \), definido, en tres dimensiones, por

\[ \nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \]

notación que emplearemos de ahora en adelante.

Esta ecuación puede ser resuelta para un gran número de situaciones de interés, que quedarán especificadas por las condiciones iniciales empleadas junto con las propiedades elegidas para los contornos del sistema (véase más adelante). Tomemos por ejemplo la ecuación de difusión unidimensional y sea el dominio dado por \( 0 < x < \infty \). Supongamos que la concentración en el extremo del sistema se mantiene fija en alguna forma, esto es, \( C(0, t) = C_0 \) para \( t > 0 \) y que no había moléculas en el tubo al comienzo del tiempo, es decir, \( C(x, 0) = 0 \) para \( x > 0 \).

Puede demostrarse que, bajo estas condiciones, la solución de la ecuación de difusión unidimensional es

\[ C(x, t) = 2C_0 \left[ 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp \left( -\frac{s^2}{2} \right) ds \right] \]

donde \( z \) se ha definido como

\[ z = x(Dt)^{-1/2} \]

### 10.3 Soluciones para \( \frac{\partial}{\partial t} u = D \frac{\partial^2}{\partial x^2} u \)

Ya hemos visto con anterioridad que la difusión pasiva, en ausencia de otras variables, destruye las correlaciones. Las diferencias de concentración son reducidas hasta alcanzar un estado de equilibrio completamente homogéneo. En esta sección plantearemos esta cuestión en términos formales. La aproximación matemática que emplearemos será útil en el desarrollo de la teoría de Turing.

Partamos de la ecuación de difusión unidimensional. Tenemos

\[ \frac{\partial u}{\partial t} = D \frac{\partial^2 u}{\partial x^2} \]
Una solución general de esta ecuación puede obtenerse a partir de una superposición de modos de Fourier espaciales, dados por

\[ u(x, t) = \left[ a \cos(qx) + b \sin(qx) \right] e^{\omega t} \]

Calculando las derivadas que aparecen en la ecuación de difusión, obtenemos

\[ \frac{\partial u(x, t)}{\partial t} = \omega e^{\omega t} \left[ a \cos(qx) + b \sin(qx) \right] \]

\[ \frac{\partial^2 u(x, t)}{\partial x^2} = -q^2 e^{\omega t} \left[ a \cos(qx) + b \sin(qx) \right] \]

Sustituyendo éstas, obtenemos la siguiente expresión para \( u(x, t) \)

\[ u(x, t) = e^{-Dq^2t} \left[ a \cos(qx) + b \sin(qx) \right] \]

El siguiente paso será definir las condiciones de contorno del problema. Tenemos, como ya sabemos, dos posibilidades de interés (para un dominio finito). Formalmente, éstas se indican por:

(a) Condiciones de contorno de Von Neumann (flujo cero),

\[ \left( \frac{\partial u(x, t)}{\partial x} \right)_{x=0} = \left( \frac{\partial u(x, t)}{\partial x} \right)_{x=L} = 0 \]

(b) Condiciones de contorno de Dirichlet (valores constantes en los extremos),

\[ \left( \frac{\partial u(x, t)}{\partial x} \right)_{x=0} = \left( \frac{\partial u(x, t)}{\partial x} \right)_{x=L} = 0 \]

Si empleamos, por ejemplo, las condiciones de Von Neumann, obtenemos

\[ \left( \frac{\partial u(x, t)}{\partial x} \right)_{x=0,L} = \left[ q \, e^{-Dq^2t} \left( a \sin(qx) + b \cos(qx) \right) \right] = 0 \]

que será nula si \( q = 0 \) (esto es, la solución trivial independiente de \( x \)) o bien si el paréntesis del interior se hace cero en algún caso. Observemos que, para \( x = 0 \), la derivada nos da

\[ \left( \frac{\partial u(x, t)}{\partial x} \right)_{x=0} = a \sin(0) + b \cos(0) = b \implies b \equiv 0 \]

y, si empleamos este resultado y calculamos la expresión correspondiente al otro extremo \( (x = L) \), tenemos

\[ \left( \frac{\partial u(x, t)}{\partial x} \right)_{x=L} = a \sin qL = 0 \]

lo que nos proporciona un conjunto ordenado de valores de \( q \) posibles,

\[ q = \frac{n\pi}{L} \quad ; \quad n = 0, 1, 2, ... \]

luego la expresión de la solución \( u(x, t) \) será de la forma

\[ u(x, t) = a e^{-D \frac{x^2}{L^2} t} \cos\left( \frac{n\pi x}{L} \right) \quad ; \quad n = 0, 1, 2, ... \]

El parámetro \( a \) se obtendrá a partir de las condiciones iniciales. De forma general, las soluciones de la ecuación de difusión se escribirán en forma de un desarrollo de Fourier.
\[ u(x,t) = \sum_{n \geq 0} a_n e^{-D \frac{n^2 \pi^2}{L^2} t} \cos \left( \frac{n \pi}{L} x \right) \]

Para \( t = 0 \), tenemos de hecho

\[ u(x,t) = \sum_{n \geq 0} a_n \cos \left( \frac{n \pi}{L} x \right) \]

que representa la forma de la perturbación inicial. Para el segundo caso (condiciones de contorno de Dirichlet) un estudio equivalente nos dará un desarrollo en serie similar, sólo que en términos de funciones seno en lugar de coseno.

### 10.4 Estabilidad de las soluciones

¿Cómo se comportan estas soluciones? ¿Cuál es su estabilidad? En realidad, ya conocemos la respuesta a partir de las simulaciones llevadas a cabo anteriormente (capítulo 1). Sin embargo, podemos dar una respuesta formal empleando los desarrollos anteriores. Supongamos que partimos de una condición inicial homogénea \( u(x,0) = u_0 \) y que la perturbamos (de forma simular a cómo se había hecho en el estudio de la estabilidad), mediante una pequeña fluctuación, es decir, tomamos

\[ u(x,t) = u_0 + \delta u(x,t) \]

donde \( u(x,t) \ll |\delta u(x,t)| \). Tenemos entonces una ecuación para las perturbaciones (indicadas por \( \dot{y}(x,t) \equiv \delta u(x,t) \)), dada por

\[ \frac{\partial \dot{y}}{\partial t} = D \frac{\partial^2 \dot{y}}{\partial x^2} \]

Si consideramos las condiciones de flujo cero, sabemos que la solución general es de la forma

\[ Y(x,t) = \sum_{n \geq 0} a_n e^{-D \frac{n^2 \pi^2}{L^2} t} \cos \left( \frac{n \pi}{L} x \right) \]

y, puesto que \( D > 0 \), se tiene de forma inmediata que

\[ e^{-D \frac{n^2 \pi^2}{L^2} t} \to 0 \]

luego todas las perturbaciones de cualquier tamaño característico terminan por desaparecer y alcanzan el estado estacionario.

### 10.5 Modelos de reacción-difusión

Llegamos por fin a los modelos inspirados en la teoría de Turing. El escenario del que partimos emplea como sistema básico un espacio dentro del cual tenemos un conjunto de \( n \) morfógenos en interacción. Estos morfógenos, que asumiremos son el resultado de la actividad celular, pueden difundirse hacia el espacio circundante. En términos de células, cada una lleva a cabo la síntesis de los morfógenos, y éstos pueden difundir a las células vecinas.

En general, las ecuaciones serán del tipo

\[ \frac{\partial C_i}{\partial t} = f_i(C_1, \ldots, C_n) + D_i \nabla^2 C_i, \quad i = 1, 2, \ldots, n \]

para cada morfogéno \( C_i \).
Aquí \( f_i^n(C_j) \) denota, para cada \( C_i \), la dinámica de las interacciones entre distintas moléculas. Es el llamado término de reacción. Puesto que también incorporamos un término de difusión, estas ecuaciones reciben el nombre de modelos de reacción-difusión. En este capítulo nos limitaremos al caso más simple, en el que \( n = 2 \).

Resolveremos en primer lugar el problema en una dimensión espacial, y generalizaremos posteriormente. Partiremos del modelo

\[
\frac{\partial C_1}{\partial t} = f_1(C_1, C_2) + D_1 \frac{\partial^2 C_1}{\partial x^2}
\]

\[
\frac{\partial C_2}{\partial t} = f_2(C_1, C_2) + D_2 \frac{\partial^2 C_2}{\partial x^2}
\]

Observemos en primer lugar que este sistema admite una solución estacionaria trivial, correspondiente al estado espacialmente homogéneo obtenido a partir de las igualdades \( f_1(C_1, C_2) = 0 \), que proporcionan el punto fijo \( P = (C_1^*, C_2^*) \). La solución es obviamente estable, puesto que el operador (lineal) de difusión no puede alterar un estado homogéneo.

La cuestión ahora es: si en lugar de considerar el estado espacialmente homogéneo consideramos una situación similar, más real, en la que tenemos pequeñas fluctuaciones en las concentraciones alrededor de \( P \), ¿podría el efecto de los términos de reacción \( f_1(C_1, C_2) \) modificar esta estabilidad? Para obtener una respuesta, supondremos que dicha perturbación ha sido llevada a cabo, y que el estado homogéneo es ahora

\[
C_1(r, 0) = C_1^* + \delta C_1
\]

\[
C_2(r, 0) = C_2^* + \delta C_2
\]

donde como es habitual las perturbaciones son muy pequeñas \( |\delta C_i| << C_i^* \), para \( i = 1, 2 \). De forma similar a como se había planteado el estudio de la estabilidad de los sistemas dinámicos (capítulo 2) exploraremos ahora, en aproximación lineal, el comportamiento de las fluctuaciones. Para simplificar la notación, indicaremos \( c_i \equiv \delta C_i \). Las ecuaciones linealizadas cerca de \( P \) son

\[
\frac{\partial c_1}{\partial t} = L_{11} c_1 + L_{12} c_2 + D_1 \frac{\partial^2 c_1}{\partial x^2}
\]

\[
\frac{\partial c_2}{\partial t} = L_{21} c_1 + L_{22} c_2 + D_2 \frac{\partial^2 c_2}{\partial x^2}
\]

Como es habitual, \( L_{ij} \) son los elementos de la matriz de Jacobi, \( L_{ij} = \partial f_i(P)/\partial c_j \). Supongamos que el dominio espacial tiene longitud \( L \). Supongamos además que las condiciones de contorno son de flujo cero (sección 10.3). En este caso, descompondremos nuestras perturbaciones en modos de Fourier de la forma

\[
c_1(r, t) = A_n e^{\omega_n t} \cos \left( \frac{n \pi r}{L} \right)
\]

\[
c_2(r, t) = B_n e^{\omega_n t} \cos \left( \frac{n \pi r}{L} \right)
\]

Las derivadas de estas expresiones nos dan

\[
\frac{\partial c_1}{\partial t} = A_n \omega_n e^{\omega_n t} \cos \left( \frac{n \pi r}{L} \right)
\]

\[
\frac{\partial^2 c_1}{\partial r^2} = A_n \left( \frac{n^2 \pi^2}{L^2} \right) \omega_n e^{\omega_n t} \cos \left( \frac{n \pi r}{L} \right)
\]
(similarmente para \( c_2 \)). Si introducimos estas expresiones en las ecuaciones lineales anteriores, obtendremos un sistema de ecuaciones

\[
A_n \left[ L_{11} + D_1 \left( \frac{n^2 \pi^2}{L^2} \right) - \omega_n \right] + B_n L_{12} = 0
\]

\[
A_n L_{21} + B_n \left[ L_{22} + D_2 \left( \frac{n^2 \pi^2}{L^2} \right) - \omega_n \right] = 0
\]

Este sistema tendrá una solución no trivial si el determinante asociado es nulo, esto es, si

\[
\begin{vmatrix}
L_{11} + D_1 \left( \frac{n^2 \pi^2}{L^2} \right) - \omega_n & L_{12} \\
L_{21} & L_{22} + D_2 \left( \frac{n^2 \pi^2}{L^2} \right) - \omega_n
\end{vmatrix} = 0
\]

lo que nos da la ecuación de valores propios para \( \omega_n \)

\[
P(\omega_n) = \omega_n^2 + \left[ L_{11} + L_{22} + (D_1 + D_2) \frac{n^2 \pi^2}{L^2} \right] \omega_n + \]

\[
+ \left( L_{11} + D_1 \frac{n^2 \pi^2}{L^2} \right) \left( L_{22} + D_2 \frac{n^2 \pi^2}{L^2} \right) L_{12} L_{21} = 0
\]

Si \( \Re(\omega_n) > 0 \), tendremos una inestabilidad asociada al modo \( n \)-ésimo. Lo que indica este resultado es que aquellas perturbaciones tales que su longitud de onda asociada sea la del modo \( n \), serán amplificadas (el valor de la exponencial es positivo) y las que tengan valores \( \Re(\omega_n) < 0 \) serán amortiguadas. Si, bajo ciertas condiciones (que obviamente dependerán del tipo de funciones de reacción) algunos modos espaciales son favorecidos, deberíamos esperar observar estructuras macroscópicas con un tamaño característico del orden de las longitudes de onda asociadas a los modos inestables. En las siguientes subsecciones consideraremos algunos ejemplos.

10.5.1 Estructuras disipativas: el Brusselator

Nuestro primer ejemplo tiene como modelo de partida el conocido Brusselator, que fue desarrollado en el capítulo 4. Ya vimos que el modelo, introducido por I. Prigogine y sus colaboradores, era capaz de exhibir comportamientos complejos, en forma de oscilaciones periódicas (ciclos límite). Veremos ahora que la introducción de los términos de difusión (de un espacio físico, en definitiva) da lugar a fenómenos de gran complejidad.

El modelo de reacción-difusión será ahora

\[
\frac{\partial X(r, t)}{\partial t} = a - (b + 1) X + X^2 Y + D_1 \nabla^2 X
\]

\[
\frac{\partial Y(r, t)}{\partial t} = b X - X^2 Y + D_2 \nabla^2 Y
\]

(hemos mantenido la notación para las concentraciones de \( X \) e \( Y \), indicando la posición espacial mediante la notación \( r \)). Ahora por lo tanto \( \nabla^2 \equiv \frac{\partial^2}{\partial r^2} \). Las perturbaciones del estado homogéneo, dado por \( P^* = (a, b/a) \), quedarán indicadas en la forma

\[
X(r, t) = a + x(r, t)
\]

\[
Y(r, t) = \frac{b}{a} + y(r, t)
\]
Figura 10.6: Espacio paramétrico para el modelo de reacción-difusión para el Brusselator. Se indican los dominios de estabilidad para el sistema con estructuras espaciales estacionarias y para las inestabilidades espaciales que exhiben comportamiento periódico. Se indica también la posición de las soluciones caóticas.

siendo \( |x| << a \) y \( |y| << b/a \). Estudiaremos el comportamiento del sistema lineal (linealizamos cerca de la solución homogénea). Tenemos así un sistema dinámico

\[
\begin{align*}
\partial_t \begin{pmatrix} x \\ y \end{pmatrix} &= L_\mu \begin{pmatrix} x \\ y \end{pmatrix}
\end{align*}
\]

con \( L_\mu \) la matriz de Jacobi asociada, esto es

\[
L_\mu = \begin{pmatrix}
\partial_x f_1 + D_1 \nabla^2 & \partial_y f_1 \\
\partial_x f_2 & \partial_y f_2 + D_2 \nabla^2
\end{pmatrix}
\]

y donde \( f_i \) hace referencia a los términos de reacción para cada variable. Si evaluamos esta matriz en el punto correspondiente al estado espacialmente homogéneo, tenemos

\[
L_\mu(P^*) = \begin{pmatrix}
b - 1 + D_1 \nabla^2 & a^2 \\
-b & -a^2 + D_2 \nabla^2
\end{pmatrix}
\]

Sólo tenemos que aplicar el procedimiento general indicado anteriormente. El polinomio característico será en este caso

\[
P(\omega_n) = \omega_n^2 + \left[ \beta_n - \alpha_n \right] \omega_n + a^2 b - \alpha_n \beta_n = 0
\]

donde

\[
\alpha_n = b - 1 + D_1 \left( \frac{n^2 \pi^2}{L^2} \right)
\]

\[
\beta_n = a^2 b + D_2 \left( \frac{n^2 \pi^2}{L^2} \right)
\]
Figura 10.7: Oscilaciones espaciotemporales caóticas del Brusselator unidimensional.

Resolviendo, tenemos

$$\omega_n^\pm = \frac{1}{2} \left[ \alpha_n - \beta_n \pm \sqrt{(\alpha_n + \beta_n)^2 - 4a^2b} \right]$$

De aquí podemos obtener varios resultados. Supongamos que $$\omega_n \in \mathbb{R}$$. Si $$\omega_n < 0$$, cualquier perturbación inicial asociada al valor propio será amortiguada. Nos interesan, de hecho, los casos para los que $$\omega_n > 0$$, que estarán ligados a la amplificación de estructuras de cierto tamaño característico. Esto ocurrirá si $$\alpha_n\beta_n - a^2b > 0$$ esto es, si

$$b > 1 + \left( \frac{D_1}{D_2} \right) a^2 + \frac{a^2}{D_2n^2\pi^2} L^2 + \frac{D_1n^2\pi^2}{L^2}$$

Esta expresión define un dominio cuyo límite inferior viene dado por la curva de bifurcación (o de estabilidad marginal)

$$b_m^c = 1 + \left( \frac{D_1}{D_2} \right) a^2 + \frac{a^2}{D_2n^2\pi^2} L^2 + \frac{D_1n^2\pi^2}{L^2}$$

Si empleamos el plano $$(n, b)$$ para definir los dominios de estabilidad del problema, obtenemos la figura 10.6. Este diagrama es de gran interés: la superficie que define el conjunto de estados inestables (que mostrarán estructuras heterogéneas) posee un valor mínimo para cierto par $$(m, b_m)$$, dado por

$$m = \frac{aL^2}{\pi^2 \sqrt{D_1D_2}}, \quad b_m = \left( 1 + \frac{D_1}{D_2} a \right)^2$$

El valor de $$m$$ será real, pero, puesto que $$n \in \mathbb{N}$$, el primer punto crítico $$b_c$$ aparecerá para un valor $$n_c \in \mathbb{N}$$, como se indica en la figura. Para este valor diremos que aparece la primera
inestabilidad cuando \( b \) crece y cruza la línea por \((n, b)\). Veremos entonces que a lo largo del dominio espacial emergen estructuras coherentes con una longitud característica \( \lambda \approx 2L/n \).

En consecuencia, el análisis lineal de la estabilidad nos muestra que las pequeñas fluctuaciones del estado homogéneo son amplificadas (para ciertos valores de los parámetros), generando de forma espontánea estructuras macroscópicas con un tamaño característico. Desde el punto de vista de la física, este sistema, y otros del mismo tipo, pueden generar estructuras regulares debido a que son sistemas abiertos, esto es, que intercambian energía y materia con el exterior. Bajo ciertas circunstancias, algunos sistemas (entre los que se incluyen los sistemas vivos) pueden emplear la energía externa de forma apropiada de manera que se reduzca la entropía interna. El resultado es la generación de estructuras dissipativas (Nicolis y Prigogine, 1977), de las que el Brusselator, analizado en esta sección, sería un ejemplo.

10.5.2 Gradientes y polaridad

El estudio de la estabilidad de las soluciones dependientes del tiempo nos da la siguiente desigualdad

\[
\left( \frac{D_1}{D_2} \right)^{1/2} > \sqrt{1 + \frac{1 + \pi^2(D_1 + D_2)/L^2}{a^2}} - \frac{1}{a}
\]

Si exploramos el espacio de parámetros, veremos de hecho que son posibles una enorme cantidad de comportamientos, entre los cuales se incluyen estados caóticos (figura 10.7). Este resultado no debe sorprendernos, en la medida en que, aunque el sistema se describe por solo dos ecuaciones, es de hecho de dimensión infinita, dado que hemos incorporado el espacio.

Podemos obtener aún más información siguiendo el análisis de la estabilidad que hemos iniciado. En las proximidades del punto crítico, los vectores asociados a la matriz (operador) de Jacobi serán de la forma

\[
(u_m, v_m)^T = (C_1, C_2)^T \sin \left( \frac{m \pi r}{L} \right)
\]
\[(u_m, v_m)^T = (C_1, C_2)^T \cos \left( \frac{m \pi r}{L} \right)\]

Sustituyendo en el sistema lineal que hemos construido para el caso anterior (Brusselator), obtenemos el siguiente cociente

\[\frac{C_2}{C_1} = \frac{D_1 m^2 \pi^2 / L^2 + 1 - b_m}{a^2}\]

Para \(m = n_c\) nos da, aproximadamente,

\[\frac{C_2}{C_1} \approx \frac{1}{a} \left( \frac{D_1}{D_2} \right)^{1/2} \left[ 1 + a \left( \frac{D_1}{D_2} \right)^{1/2} \right] < 0\]

Y a continuación podremos obtener \(C_1\) (que está indeterminado). El resultado final (Nicolis y Prigogine, 1977) es un conjunto de soluciones en la forma

\[(x, y)^T = (C_1, C_2)^T e^{3(\omega_c)t} \sin \left( \frac{\pi r}{L} \right)\]

Para la primera bifurcación con condiciones de flujo cero obtenemos

\[(x, y)^T = (C_1, C_2)^T e^{3(\omega_c)t}\]

Una propiedad especialmente importante de la parte dependiente del espacio de las soluciones anteriores es que, para condiciones de contorno de flujo cero, permite obtener gradientes de polaridad de forma espontánea (figura 10.8), lo que explicaría de manera natural la aparición de una señal a lo largo del eje antero-posterior de un embrión que serviría a las células para determinar su posición a lo largo de dicho eje.

### 10.6 Bifurcación de estructuras estacionarias

Siguiendo el desarrollo de Nicolis y Prigogine, podemos dar un tratamiento general al problema de la obtención de la forma explícita de las soluciones estacionarias de las ecuaciones de reacción difusión una vez cruzado el punto de bifurcación \(b_c\). Supongamos que llevamos a cabo la aproximación

\[X(r, t) = a + x(r, t)\]
\[Y(r, t) = \frac{b}{a} + y(r, t)\]

y que introducimos dicha aproximación dentro de las ecuaciones dinámicas del Brusselator, reteniendo ahora las contribuciones de los términos no lineales. Descompongamos el operador lineal \(L\) en la forma siguiente:

\[L \equiv L_b = L_c + (L - L_c) = L_c + \begin{pmatrix} (b - b_c) & 0 \\ -(b - b_c)^2 & 0 \end{pmatrix}\]

donde \(L_c\) es el operador evaluado en el punto crítico de la primera bifurcación. Obtenemos entonces el sistema

\[L_c \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -h(x, y) \\ h(x, y) \end{pmatrix}\]

con

\[h(x, y) = (b - b_c)x + 2axy + \frac{b}{a} x^2 + x^2 y\]
y definimos de nuevo las condiciones de contorno como antes. Para calcular las soluciones \((x, y)^T\), debemos recordar que, en las proximidades del punto crítico, las correcciones a la solución \((a, b/a)^T\) (la rama termodinámica) deben ser pequeñas. Por lo tanto, podemos desarrollar tanto \((x, y)^T\) como \(\gamma = b - b_c\) en términos de un parámetro \(\epsilon\) (pequeño), en la forma

\[
\begin{pmatrix}
    x \\
    y
\end{pmatrix} = \epsilon \begin{pmatrix}
    x_0 \\
    y_0
\end{pmatrix} + \epsilon^2 \begin{pmatrix}
    x_1 \\
    y_1
\end{pmatrix}
\]

que introduciremos en la ecuación para \(L_c(x, y)^T\), identificando las potencias del mismo orden para \(\epsilon\). Obtenemos así un conjunto de expresiones de la forma

\[
L_c \begin{pmatrix}
    x_k \\
    y_k
\end{pmatrix} = \begin{pmatrix}
    -a_k \\
    a_k
\end{pmatrix}, \quad k = 0, 1, ...
\]

junto a las correspondientes condiciones de contorno, que serán de una de las formas siguientes:

\[
x_k(0) = x_k(L) = ... = 0
\]

\[
\frac{dx_k(0)}{dt} = \frac{dx_k(L)}{dt} = ... = 0
\]

Obtenemos así los primeros coeficientes en la forma

\[
a_0 = 0
\]

\[
a_1 = \gamma_1 x_0 + \frac{b_c}{a} x_0^2 + 2ax_0y_0
\]

\[
a_2 = \gamma_2 x_0 + \left(\gamma_1 + \frac{2b_c}{a} x_0 + 2ay_0\right)x_1 + 2ax_0y_1 + \gamma_1 \frac{x_0^2}{a} + x_0^2y_0
\]

y por lo tanto, como resultado tenemos que

\[
L_c \begin{pmatrix}
    x_0 \\
    y_0
\end{pmatrix} = 0
\]

esto es, este vector es proporcional al vector propio crítico asociado a \(L_c\).

**10.7 Modelo de Gierer-Meinhardt**

Aunque hemos mencionado repetidamente las estructuras de Turing y su importancia en el fenómeno de la morfogénesis, no hemos hecho demasiado hincapié en las hipótesis básicas que deberían suponerse como razonables para modelar dicho fenómeno. De hecho, las ecuaciones inicialmente propuestas por Turing carecían de estos elementos, si bien los resultados que obtuvo son genéricos. Fueron los investigadores alemanes Hans Meinhardt y Alfred Gierer quienes por primera vez exploraron este problema desde una perspectiva biológica. Estos autores estudianaron diversas formas de combinar la interacción de dos morfógenos con el objetivo de obtener estructuras ordenadas similares a las observadas en los seres vivos. Más aún: en algunos casos, los modelos fueron capaces de reproducir muy bien los resultados de experimentos de perturbación de embriones que daban lugar a individuos con diversas anomalías. Estas incluían la aparición de individuos con dos extremos idénticos a cada lado (dos cabezas o dos colas) o la reorganización de un embrión para dar lugar a un individuo menor, aunque completo (figura 10.2).

El modelo está basado en un mecanismo de activación-inhibición. Gierer y Meinhardt, así como Segel y Jackson de forma independiente, probaron que dos propiedades básicas eran especialmente
importantes para obtener estructuras espaciales. Estas propiedades son: (a) autocatálisis local y (b) inhibición de largo alcance. La primera propiedad es esencial para que las pequeñas fluctuaciones locales puedan ser amplificadas hasta adquirir un tamaño macroscópico. Un morfógeno se llamará autocalalítico si un pequeño incremento en su concentración local induce su crecimiento posterior. Esta catálisis no necesita ser directa: dos morfógenos pueden interactuar adecuadamente (como veremos) para producir el efecto deseado.

La autocatálisis, que amplifica inhomogeneidades iniciales, no es suficiente para crear patrones espaciales. Esta propiedad debe ser completada por la acción de una molécula antagonista que posea una alta tasa de disolución (con relación a la primera). Si llamamos $h$ a este morfógeno inhibidor, podemos formular un primer modelo que describa su interacción. Este es (Gierer y Meinhardt, 1972)

$$
\frac{\partial a}{\partial t} = \rho_a \frac{a^2}{(1 + \kappa_a a^2)} h - \mu_a a + \sigma_a D_a \nabla^2 a
$$

$$
\frac{\partial h}{\partial t} = \rho_h a^2 - \mu_h + \sigma_h + D_h \nabla^2 h
$$

Aunque veremos algunos ejemplos de la dinámica de este sistema para un dominio bidimensional, exploraremos en primer lugar la estabilidad del estado estacionario en una dimensión. Consideremos el caso particular $\kappa_a = \sigma_a = 0$. Llevaremos a cabo un cambio de variables con el objeto de obtener un sistema adecuadamente simplificado. Sean

$$
t = \mu_a t \quad ; \quad l = \sqrt{\frac{\mu_a}{D_h}}
$$

$$
a = \frac{\mu_a \rho_h}{\mu_h \rho_a} a \quad ; \quad lh = \frac{\mu_a^2 \rho_h}{\mu_h \rho_a^2} h
$$

lo que nos da el nuevo sistema

$$
\frac{\partial a}{\partial t} = \frac{a^2}{h} - a + \sigma D \nabla^2 a
$$

$$
\frac{\partial h}{\partial t} = \mu(a^2 - h) + \nabla^2 h
$$

donde hemos empleado la notación abreviada $D = D_a/D_h$, $\mu = \mu_h/\mu_a$, $\sigma = (\rho_h \sigma_a)/(\mu_h \rho_a)$ y $\nabla^2 = \partial^2/\partial x^2$. Además hemos omitido el simbolo anterior sobre las variables, lo que no genera confusión alguna.

Queremos resolver el sistema anterior, que supondremos acotado a cierto dominio espacial $\mathcal{D} = \{x ; \ 0 \leq x \leq L\}$. Supondremos además condiciones de contorno de flujo cero. El estado estacionario homogéneo queda, con esta notación, escrito en la forma $(a_0, h_0)$, con $a_0 = 1 + \sigma$ y $h_0 = a_0^2$.

Consideremos una perturbación de este estado,

$$
a = a_0 + \delta a
$$

$$
h = h_0 + \delta h
$$

donde, como antes, tenemos que $\delta a = f_a(x, t)$ y $\delta h = f_h(x, t)$, ambas muy pequeñas comparadas con los valores estacionarios de las variables. Si empleamos condiciones de contorno de flujo cero, las perturbaciones serán de la forma

$$
\delta a = f_a(x, t) = \delta a_0 e^{\omega t} \cos(2\pi k x)
$$
\[ \delta h = f_h(x, t) = \delta h_0 e^{i\omega t} \cos(2\pi k x) \]

con

\[ k_n = \frac{n\pi}{L}; \quad n = 0, 1, 2, \ldots \]

Cada uno de los valores de \( n \) está asociado a una de las "frecuencias" \( \omega_n \) (que pueden ser complejas). Las ecuaciones linealizadas cerca del punto de equilibrio se obtienen de la matriz de Jacobi

\[
L(a_0, h_0) = \begin{pmatrix}
\omega_n + Dk_n^2 + (1 - 2/a_0) & 1/a_0^2 \\
-2\mu a_0 & \omega_n + k_n^2 + \mu
\end{pmatrix}
\]

para la cual podemos calcular el determinante, que nos da en este caso

\[ P(\omega_n) = \omega_n^2 + \alpha \omega_n + \beta = 0 \]

donde \( \alpha \) y \( \beta \) indican

\[ \alpha = (1 + D)k_n^2 + 1 + \mu - 2/a_0 \]
\[ \beta = Dk_n^4 + (1 + \mu D - 2/a_0)k_n^2 + \mu \]

Una fluctuación asociada a la frecuencia \( \omega_n \) crecerá si su parte real es positiva. La longitud crítica de la perturbación \( L_c(n) \) queda definida por

\[ \Re(\omega_n) = 0 \]

y si \( \omega_n \) es compleja, entonces la condición es \( \alpha = 0 \). Empleando la restricción sobre los \( k_n \) definida por las condiciones de contorno, vemos que las perturbaciones oscillatorias pueden ser amplificadas en un sistema de tamaño \( L \) si y sólo si se satisface la desigualdad

\[ L > L_c(n) = n\pi \left[ \frac{1 + D}{2/a_0 - 1 - \mu} \right]^{1/2} \]

Vemos que la primera oscilación que puede obtenerse estará asociada a la frecuencia \( \omega_1 \), dado que \( L_c(1) < L_c(n) \) (\( n > 1 \)).

Ahora, detengámonos a estudiar lo que ocurre cuando \( \omega_n \) es real. En este caso, la condición \( \Re(\omega_n) = 0 \) se reduce a \( \beta = 0 \), con lo que tendremos amplificación de fluctuaciones si

\[ L > L_c(n) = n\pi \sqrt{2D} \left\{ \left( \frac{2}{a_0} - 1 - \mu D \right) + \left[ \left( \frac{2}{a_0} - 1 - \mu D \right)^2 - 4\mu D \right]^{1/2} \right\}^{-1/2} \]

y, nuevamente, \( L_c(1) < L_c(n) \) (\( n > 1 \)).

Tal y como señalan Koch y Memhardt (1994) la existencia de una longitud crítica tiene importantes consecuencias biológicas. Un organismo en crecimiento desarrollará estructuras no-uniformes si y sólo si su tamaño es superior a \( L_c(1) \).

Podemos obtener a continuación el diagrama de fases asociado a este modelo en el plano de parámetros \( (D, \mu) \). Supongamos que \( L \gg L_c(1) \)

esto es, el tamaño del sistema es claramente mayor que el de la longitud mínima característica. Supongamos (como hemos hecho con anterioridad) que \( k \) es una variable continua. Podemos calcular el valor de \( k_m \) del número de onda asociado a la mayor tasa de amplificación positiva,
Figura 10.9: Diagrama de estabilidad de las soluciones estacionarias del modelo de Meinhardt (véase texto).

\( \Re(\omega_n) \). En la aproximación lineal, \( k_m \) domina la evolución de las fluctuaciones, y lo determinaremos resolviendo las ecuaciones

\[
\frac{\partial}{\partial k} \Re[\omega(k)]|_{k=k_m} = 0
\]

\[
\frac{\partial^2}{\partial k^2} \Re[\omega(k)]|_{k=k_m} < 0
\]

Distinguiremos entre dos casos: real e imaginario. Si \( \omega \in \mathbb{C} \), entonces \( \Re[\omega(k)] = -\alpha/2 \), que cumple las expresiones anteriores para \( k_m = 0 \). Podemos verificar entonces que, si \( \sigma \in [0, 1] \), \( \omega(k_m = 0) \) es complejo con parte real positiva si

\[
\left( \frac{2}{\sqrt{a_0}} - 1 \right)^2 \leq \mu \leq \frac{2}{a_0} - 1
\]

Por otra parte, si \( \omega \) es real, el teorema de la función implícita aplicado a la ecuación característica \( P(\omega) \) demuestra que \( \partial_k \omega = 0 \) si

\[
\omega = -\frac{\partial \beta / \partial k}{\partial \alpha / \partial k} = \frac{k_m[2Dk^2_m + (1 + \mu D - 2/a_0)]}{k_m(1 + D)}
\]

pero, dado que el denominador de esta expresión es proporcional a \( k_m \), debemos considerar los casos \( k_m = 0 \) y \( k_m > 0 \).

Consideremos en primer lugar \( k_m > 0 \). Introduciendo la solución para \( \omega \) obtenida de la ecuación característica, obtenemos la siguiente expresión para \( k_m \)

\[
k_m = \left[ \frac{2/a_0 - 1 - \mu D}{2D} \right]^{1/2}
\]

para la cual \( \omega \) es máximo y \( \omega(k_m) > 0 \), con lo que \( \mu \) y \( D \) satisfacen la desigualdad

\[
\mu D < \left( \frac{2}{\sqrt{a_0}} - 1 \right)^2
\]
Finalmente, consideremos el caso $k_m = 0$, que requiere $\alpha^2 - 4\beta \geq 0$. La frecuencia poseerá un máximo positivo en $k = 0$ si $\mu$ cumple las desigualdades

$$0 \leq \mu < \left[ \frac{2}{\sqrt{\alpha}} - 1 \right]^2$$

Estos resultados se resumen en el diagrama de fases presentado en la figura 10.9. Los dominios obtenidos son:

- $0 < \mu < \left[ \frac{2}{\sqrt{\alpha}} - 1 \right]^2$: las fluctuaciones crecen exponencialmente con $k_m = 0$ (dominio $G_a$ en la figura).

- $\left[ \frac{2}{\sqrt{\alpha}} - 1 \right]^2 < \mu < 2/a_0 - 1$: las fluctuaciones asociadas con $k_m = 0$ oscilan con cierta frecuencia fija $\Im(\omega)$ y crecen exponencialmente a una tasa $\Re(\omega)$ (zona $G_b$ del diagrama).

- $\mu \geq 2/a_0 - 1$ y $\mu D < \left[ \frac{2}{\sqrt{\alpha}} - 1 \right]^2$: el estado espacialmente uniforme es inestable. Se indica en el diagrama por la zona $I$, y corresponde a la región interesante en términos de morfogénesis.

- $\mu > 2/a_0 - 1$ y $\mu D > \left[ \frac{2}{\sqrt{\alpha}} - 1 \right]^2$: el estado espacialmente uniforme es estable, y cualquier perturbación que llevemos a cabo es amortiguada y desaparece (región $H$).

Podemos explorar este modelo numéricamente, empleando la solución discretizada de las ecuaciones (véase el apéndice B en Koch y Meinhardt, 1994). Si empleamos las ecuaciones bidimensionales, esto es, con dos grados de libertad espaciales (cuya estabilidad general estudiaremos en la próxima sección), podemos obtener estructuras regulares estables, como las que vemos en la figura 10.10. En esta figura vemos de hecho la evolución desde el estado inicial, que es prácticamente homogéneo excepto por una perturbación aleatoria del orden de $10^{-3}$. A partir de esta perturbación, el sistema desarrolla una estructura espacial macroscópica: una estructura de Turing (Turing, 1952; Murray, 1989).

### 10.8 Estructuras bidimensionales

Introduciremos ahora un ejemplo de formación de patrones en dos dimensiones. Aunque en sí el tratamiento no difiere mucho de lo ya visto, estos modelos nos permitirán reconsiderar los resultados anteriores y nos servirán además para introducir algunas ideas acerca de cómo resolver numéricamente las ecuaciones de reacción-difusión. Además, está claro que la extensión del número de dimensiones es de importancia en la medida en que los sistemas reales a los que hemos hecho mención al principio son, al menos, bidimensionales (si pensamos por ejemplo en los patrones de coloración de la piel).

Consideremos el problema de la formación de los patrones de coloración típicos de la piel de los mamíferos. En la figura 10.11 vemos un ejemplo: las manchas del pelaje de un jaguar. Este es un ejemplo de los muchos dibujos de distintos tipos que van desde coloraciones homogéneas hasta patrones complejos formados por manchas y/o bandas regulares de distintos tamaños y distribución (Murray, 1990). El patrón de coloración que observamos se desarrolla hacia el final de
Figura 10.10: Generación de una estructura de Turing mediante el modelo de Meinhardt, para una red $40 \times 40$ y los parámetros $D_\alpha = 0.005$, $D_\beta = 0.2$, $\rho_n = 0.01$, $\rho = 0.02$, $\mu_\alpha = 0.01$, $\mu_\beta = 0.02$, $\sigma_\alpha = \kappa_\alpha = 0$. (a) $t=1$, (b) $t=500$, (c) $t=1500$, (d) $t=3000$. 
Figura 10.11: Ejemplo de patrón de coloración de la piel de un mamífero (un jaguar). Estos patrones pueden ser reproducidos mediante modelos de reacción-difusión.

la embriogénesis (el proceso de formación del embrión completo) lo que sugiere que su distribución ha sido previamente definida en alguna etapa anterior: existirá un pre-patrón. Este pre-patrón, que definirá en último término la configuración final, se genera en las primeras etapas de la gestación (para una cebra, por ejemplo, en los primeros 20-30 días sobre una gestación de un año).

Siguiendo las ideas básicas planteadas por el modelo de Turing, Jim Murray ha propuesto un modelo de formación de patrones de este tipo definido en la forma (Murray, 1990)

\[
\frac{\partial u}{\partial t} = \gamma f(u, v) + \nabla^2 u
\]

\[
\frac{\partial v}{\partial t} = \gamma g(u, v) + D \nabla^2 v
\]

donde los términos de reacción son

\[
f(u, v) = a - u - h(u, v)
\]

\[
g(u, v) = \alpha(b - v) - h(u, v)
\]

y la función \( h(u, v) \) que aparece en ambos es

\[
h(u, v) = \frac{\rho uv}{1 + u + Ku^2}
\]

Todas las constantes que aparecen son positivas y \( D > 1 \). El parámetro \( \gamma \) introduce un factor de escala, siendo de hecho una medida del tamaño del dominio en el que tiene lugar el fenómeno de reacción-difusión. Aunque la elección de la forma específica de las funciones será la indicada arriba, no debemos olvidar que el mecanismo de generación de estructuras de Turing descansa sobre un proceso de propiedades universales. Existe una familia infinita de funciones de interacción que darán los mismos resultados cualitativos (Murray, 1988).

El tratamiento del problema general es muy similar al que se había realizado en una dimensión. Ahora obtenemos una nueva relación entre dos valores naturales y el número de onda,

\[
k^2 = \pi^2 \left( \frac{n^2}{p^2} + \frac{m^2}{q^2} \right)
\]
Estructuras de Turing y Caos Espaciotemporal

Figura 10.12: Ejemplo de soluciones numéricas de las ecuaciones de Murray (arriba) comparadas con ejemplos de coloración reales, correspondientes a colas de mamíferos (Murray, 1988). Se han empleado los parámetros \( \alpha = 1.5, K = 1, \rho = 18.5, a = 92, b = 64 \), los cuales definen un estado estacionario \((u^*, v^*) = (10, 9)\). y, empleando la misma geometría, se utilizan distintos factores de escala: (a) \( \gamma = 9 \); (b) \( \gamma = 15 \) y (c) \( \gamma = 25 \). Se muestran los dibujos de las colas de: (d) guepardo, (e) jaguar adulto, (f) jineta (feto) y (g) leopardo.

siendo \( n, m \in \mathbb{Z} \). De forma similar a lo visto con anterioridad, los patrones espaciales poseerán una solución inestable dada por (Murray, 1990)

\[
w(x, y, t) = \sum_n \sum_m C_{nm} \exp\left(\lambda(k^2)t\right) \cos\left(\frac{n \pi x}{p}\right) \cos\left(\frac{m \pi y}{q}\right)
\]

siendo \( \lambda \) el valor propio asociado a \( k^2 \).

Este modelo permite explorar diversas situaciones entre las que se encuentran las distintas zonas del cuerpo que muestran patrones de manchas diferentes en función de la geometría. Así, por ejemplo, para explorar una zona con simetría aproximadamente cilíndrica (como la cola o las patas) podemos considerar condiciones periódicas de contorno y analizar el tipo de patrones obtenidos para distintas combinaciones de parámetros que generan estructuras. En la figura 10.12 vemos ejemplos de las soluciones obtenidas así como las correspondientes figuras observadas sobre la piel de distintos mamíferos.

El efecto más dramático mostrado por el modelo lo introduce el cambio de escala definido por el parámetro \( \gamma \). Si el tamaño del dominio \( \gamma \) es muy pequeño, no se dará ningún patrón espacial (dado que no habrá espacio suficiente para estabilizar patrones heterogéneos que surjan de modos de tamaño menor al del sistema. Esta situación es típica, y el modelo predice que habrá pocos mamíferos pequeños con dibujos, como así ocurre. A medida que \( \gamma \) crece, introducimos la posibilidad de obtener inestabilidades que generen estructuras disipativas, que van haciéndose de menor tamaño respecto al tamaño que define \( \gamma \). Sin embargo, para tamaños lo bastante grandes, nos encontramos con un efecto inesperado: volvemos a tener uniformidad. Sin embargo, así es (Murray, 1990) y, nuevamente, coincide con la observación de los grandes mamíferos (como el elefante) que carecen de patrón alguno de coloración.

Los resultados previos sugieren una explicación causal muy coherente de la formación de es-
Figura 10.13: Efecto del tamaño del dominio en el que se generan los patrones definidos por el modelo de Murray. Los parámetros son \( a = 1.5 \), \( K = 0.125 \), \( \rho = 13 \), \( a = 103 \), \( b = 77 \), \( d = 7 \). La dimensión de cada dominio (que aquí representamos del mismo tamaño) viene dada directamente por \( \gamma \). Los valores empleados son (a) \( \gamma < 0.1 \); (b) \( \gamma = 0.5 \); (c) \( \gamma = 25 \); (d) \( \gamma = 250 \); (e) \( \gamma = 1250 \); (f) \( \gamma = 3000 \); y (g) \( \gamma = 5000 \) (de Murray, 1990).

Las estructuras en sistemas biológicos, y especialmente a lo largo de la morfogénesis. Existen buenos ejemplos de sistemas reales en los que dichas estructuras han sido observadas, y los gradientes de sustancias claramente caracterizados. En otros casos, la observación directa del crecimiento del organismo (como es el caso de una especie de pez ángel) permite ver de forma directa la generación de una estructura de Turing, cuyo patrón final y proceso de formación encajan perfectamente en la teoría previamente descrita (Kondo y Asai, 1995). Otras evidencias apuntan a la existencia de procesos genéricos de creación de formas que serían básicamente universales (Goodwin, 1994; Alberch, 1980, 1989).

Del mismo modo, existen numerosas aplicaciones de este formalismo al problema de la formación de estructuras en ecosistemas, en los que a menudo apreciamos una heterogeneidad que no puede reducirse a las restricciones del ambiente. A menudo, de hecho, las evidencias apuntan con claridad hacia un origen interno, debido a las interacciones de poblaciones con el espacio (Bascompte y Solé, 1995).

### 10.9 Redes acopladas y caos espaciotemporal

Hemos mencionado brevemente la aparición de comportamientos irregulares en sistemas de reacción-difusión. También aquí, donde el espacio aparece representado, hallamos escenarios de bifurcación, de intermitencia o de Ruelle-Takens (capítulo 5). Ya hemos discutido con anterioridad este punto, cuando introdujimos el esquema de aparición de caos a través de la cuasiperiodicidad. Tanto en el experimento de Couette-Taylor como en el de convección de Bénard el sistema acababa generando atractores extraños. Un resultado de gran importancia es el hecho de que la dimensionalidad de estos atractores sigue siendo baja en muchos casos (excepto en la turbulencia desarrollada) pese a la dimensión potencial del sistema físico.

La mayoría de estudios del caos espaciotemporal (Kuramoto, 1984; Loskutov y Mikhailov,
1990) conllevan el estudio de sistemas de ecuaciones en derivadas parciales. Existe sin embargo un formalismo distinto, ideado por el físico Kunihiko Kaneko (Kaneko, 1983; Kaneko 1992) y que llamaremos redes de aplicaciones acopladas o, simplemente, aplicaciones acopladas (coupled map lattices, CML).

Estos sistemas dinámicos son discretos en el tiempo y en el espacio pero continuos en los estados que pueden adoptar las variables. Se trata de hecho de emplear una red d-dimensional sobre la que colocamos aplicaciones discertas (como las analizadas en los capítulos anteriores) conectadas entre sí localmente. Un modelo CML queda así definido por un conjunto de ecuaciones

\[ x_{n+1}(r) = f_\mu(x_n(r)) + C_n(\{r'\}, D) \]

donde \( x_n(r) \) es el valor local del sistema que podemos imaginar como una población en una zona del espacio para el tiempo \( n \), en el punto \( r \). La red tendrá \( L^d \) elementos, con \( L \in \mathbb{N} \). Análogamente a los ejemplos de sistemas de reacción-difusión presentados con anterioridad, tenemos un término de interacción \( f_\mu \) y un término de acoplamiento. Este término puede tomar distintas formas:

- **Acoplamiento lineal (difusión).** Esta es la contrapartida discreta del término de difusión que hemos introducido con anterioridad. Explicitamente, tendremos

\[ C_n(r', D) = D \left[ \sum_{r'} z_n(r') - z_n(r) \right] \]

con \( D \) el coeficiente de difusión. La suma se lleva a cabo sobre los \( q \) vecinos más próximos, cuya posición queda indicada por el conjunto \( \{r'\} \). El acoplamiento lineal es por lo tanto la contrapartida discreta del operador Laplaciano. Este acoplamiento induce inestabilidades que pueden hacer que se generen valores divergentes. Para evitarlo, se puede emplear un truncamiento, esto es, \( x_n(r) = 0 \) si \( x_n(r) < 0 \). Debemos tener en cuenta, sin embargo, que este truncamiento modifica la dinámica anterior, lo que hace aconsejable, en ocasiones, evitar su empleo.

- **Acoplamiento no-lineal (dispersión).** En este caso, el acoplamiento no consiste en una simple suma de términos, sino que toma la forma:

\[ C_n(r', D) = D \left\{ \sum_{r'} \left[ f_\mu(x_n(r')) - f_\mu(x_n(r)) \right] \right\} \]

y las ecuaciones completas se escribirán

\[ x_{n+1}(r) = (1 - D)f_\mu(x_n(k)) + \frac{1}{q} \sum_{r'} f_\mu(x_n(r')) \]

¿Cuál es el comportamiento cualitativo de estos sistemas dinámicos? Como es fácil imaginar, incluso el modelo más simple exhibirá comportamientos complejos, del mismo modo que la contrapartida discreta de la ecuación logística continua (que sólo posee un punto de equilibrio no trivial) muestra una enorme variedad de comportamientos dinámicos. Veremos a continuación un ejemplo.

### 10.10 Redes logísticas

Supongamos que acoplamos difusivamente un conjunto de aplicaciones logísticas a través del mecanismo anterior (Kaneko, 1989; Waller y Kapral, 1984; Solé y Valls, 1992; Bascompte y Solé,

El efecto de la difusión es muy importante: no sólo modifica el comportamiento de las soluciones, sino que tiene un efecto importante sobre su estructura en la región caótica. En este sentido, hemos visto en el capítulo dedicado al caos determinista que el escenario de bifurcación por duplicación de periodo terminaba alcanzando el régimen caótico y que éste aparecía interrumpido por un conjunto denso de “ventanas periódicas”, en las que la dinámica regresaba a un comportamiento regular y predecible. Aunque no se ha mencionado con anterioridad, debemos señalar que las ventanas periódicas son especialmente sensibles a los efectos del ruído. El efecto del espacio será de hecho, como veremos, similar estadísticamente al de un ruído añadido a la dinámica del sistema.

Consideremos una red unidimensional formada por \( i = 1, 2, ..., N \) aplicaciones unidimensionales, de forma que el estado de cada elemento de la red sea una función

\[
x_{n+1}(i) = F \left(x_n(i-1), x_n(i), x_n(i+1)\right)
\]

donde supondremos que la función que define la dinámica de cada elemento es una aplicación logística \( f_\mu = \mu x(1-x) \), con \( x \in [0,1] \) y \( \mu \in [0,4] \). Si empleamos acoplamiento por difusión, tendremos una dinámica dada por

\[
x_{n+1}(i) = f_\mu(x_n(i)) + D \left(x_n(i-1) + x_n(i) - 2x_n(i+1)\right)
\]

donde, como antes, \( D \) es el coeficiente de difusión. Supondremos que empleamos condiciones periódicas de contorno, esto es, \( x_N(N+1) = x_N(1) \) y \( x_N(0) = x_N(N) \). Restringiremos además el valor de \( D \) al intervalo \( D \in [0,1/2] \) (lo cual es razonable si tenemos en cuenta que el término de difusión empleado es un promedio del flujo de población entre puntos vecinos sobre la red). Deberemos hacer notar que este tipo de acoplamiento puede, en algunos casos, dar soluciones divergentes. Este hecho puede corregirse empleando un formalismo apropiado a cada caso, como por ejemplo el acoplamiento no-lineal antes introducido.

Es posible, como antes, analizar la estabilidad de las soluciones de la red con acoplamiento lineal. Tendremos ahora un conjunto de parámetros \( (c, D, N) \) para explorar. Podemos estudiar la estabilidad de las soluciones, siguiendo el procedimiento original del artículo de Turing (Turing, 1952) que nos permite reescribir las ecuaciones para las aplicaciones en términos de modos normales, tomando

\[
x_n(j) = \sum_{m=1}^{N} \xi_n(m) \exp \left(\frac{2\pi jm}{N}\right)
\]

Partiendo de esta transformación, y empleando la relación de ortogonalidad

\[
\sum_{j=1}^{N} \exp \left(\frac{2\pi j(m-k)}{N}\right) = N \delta_{km}
\]

(con \( \delta_{km} = 0 \) si \( k \neq m \) y \( \delta_{kk} = 1 \)), el \( m \)-ésimo modo normal (la amplitud del modo de Fourier asociado) será

\[
\xi_n(m) = \frac{1}{N} \sum_{j=1}^{N} x_n(j) \exp \left(-\frac{2\pi jm}{N}\right)
\]
y a partir de esta expresión podemos derivar la dinámica de las amplitudes, que toman la forma, en el sistema con acoplamiento lineal

\[
\xi_{n+1}(m) = \left[ \mu - 4D \sin^2 \left( \frac{\pi m}{N} \right) \right] \xi_n(m) - \mu \sum_{k=0}^{N-1} \xi_n(k) \xi_n(m-k)
\]

Hemos obtenido esta expresión sustituyendo 10.10.1 en términos de las amplitudes de Fourier en la ecuación dinámica para la red, multiplicando por \( \exp(-2\pi ijk/N) \) y llevando a cabo la suma para \( j = 1, ..., N \) (Waller y Kapral, 1984). De forma similar puede obtenerse la expresión para el acoplamiento no-linear, que nos da en este caso

\[
\xi_{n+1}(m) = \mu \left[ 1 - 4D \sin^2 \left( \frac{\pi m}{N} \right) \right] \left\{ \xi_n(m) \sum_{j=1}^{N} \xi_n(k) \xi_n(m-k) \right\}
\]

La difusión no acopla los modos normales en el primer caso, mientras que sí lo hace en el segundo.

Podemos ahora explorar la estabilidad de esta red para el sistema definido por la aplicación logística. Suponemos que el parámetro \( \mu \) se halla en el dominio de las soluciones estables (atractores puntuales) y que queremos estudiar la estabilidad de las soluciones espacialmente homogéneas correspondientes a \( z_k^* = 0 \) y \( z_k^* = 1 - 1/\mu \).

Siguiendo el procedimiento que ya conocemos, llevaremos a cabo la perturbación del estado espacialmente homogéneo

\[
x_n(i) = z^* + \delta x_n(i)
\]

Sustituyendo en 10.10.2 obtenemos una nueva condición sobre las perturbaciones

\[
\delta \xi_n(m) = z^* \delta m_0 + \delta \xi_n(m)
\]

donde

\[
\delta \xi_n(m) = \frac{1}{N} \sum_{j=1}^{N} \delta x_n(j) \exp \left( -\frac{2\pi ijm}{N} \right)
\]

Sustituyendo estas expresiones en las ecuaciones dinámicas para las amplitudes (y manteniendo sólo los términos lineales) obtenemos, en el caso de acoplamiento lineal,

\[
\delta \xi_n(m) = \left[ \mu(1-2z^*) - 4D \sin^2 \left( \frac{\pi m}{N} \right) \right] \delta \xi_0(m)
\]

La condición de estabilidad marginal para un componente de Fourier dado quedará definida por la condición (Waller y Kapral, 1984)

\[
\mu(1-2z^*) - 4D \sin^2 \left( \frac{\pi m}{N} \right) = \pm 1
\]

(para \( j = 0, ..., N-1 \) que, para \( z^* = 0 \) nos da

\[
\mu = \pm 1 + 4D \sin^2 \left( \frac{\pi m}{N} \right)
\]

mientras que, para \( z^* = 1 - 1/\mu \)

\[
\mu = (2 \pm 1) - 4D \sin^2 \left( \frac{\pi m}{N} \right)
\]

Las fronteras anteriores estarán asociadas con modos de longitud de onda \( N/j \) para un \( j \) dado. Tendremos dos casos límite. Por una parte, tendremos los modos de longitud de onda infinita, con
j = 0, cuyas bifurcaciones tienen lugar en la misma zona que ya se vio para los puntos fijos del sistema aislado (con \( D = 0 \)). Por otra parte, tendremos la longitud de onda de menor tamaño, que dividiremos en dos casos, en función de si la red tiene un número par o impar de elementos. Si \( N \) es impar, los dominios de estabilidad para los modos marginalmente estables quedarán definidos por las fronteras

\[
\mu = \pm 1 + 4D
\]
\[
\mu = (2 \mp 1) - 4D
\]

para \( x^* = 0 \) y para \( x^* = 1 - 1/\mu \), respectivamente. Este modo, de longitud de onda 2, corresponde a una red con elementos cuyos estados se hallan estrictamente alternados. En el segundo caso, un valor de \( N \) impar introduce un conjunto de modos con \( j = (N \pm 1)/2 \) que proporcionan las nuevas fronteras

\[
\mu = \pm 1 + 4D \sin^2 \left[ \frac{(N \pm 1)\pi}{2N} \right]
\]
\[
\mu = (2 \mp 1) - 4D \sin^2 \left[ \frac{(N \pm 1)\pi}{2N} \right]
\]

para \( x^* = 0 \) y para \( x^* = 1 - 1/\mu \), respectivamente. Estos modos dependen ahora de \( N \), aunque estas fronteras tienden a las del caso anterior para \( N \to \infty \), como es de esperar. Estos resultados, que permiten ver el comportamiento de las soluciones para distintos valores de la difusión, pueden ser generalizados, como veremos en la siguiente sección.

### 10.11 Bifurcaciones: análisis formal

Consideraremos en esta sección las propiedades formales de las redes con acoplamiento local definido por difusión, esto es,

\[
x_{n+1}(r) = f_\mu(x_n(k)) + D \sum_{r'} q \left[ x_n(r') - x_n(r) \right]
\]

donde por simplicidad emplearemos la notación

\[
\nabla^2 x_n(r) \equiv \sum_{r'} x_n(r') - q x_n(r)
\]

y supondremos condiciones periódicas de contorno. Consideraremos una perturbación del estado espacialmente homogéneo, dada por

\[
x_n(r) = x_0 + y_n(r)
\]

donde, como es habitual, suponemos que \(|y_n(r)| \ll x_0 \). El sistema dinámico original puede ser aproximado por

\[
x_{n+1}(r) = f_\mu(x_0 + y_n(r)) + D \left[ \sum_{r'} q \left( x_0 + y_n(r') \right) - (x_0 + y_n(r)) \right]
\]

y a primer orden escribiremos

\[
f_\mu(x_0 + y_n(r)) \approx f_\mu(x_0) + \left( \frac{\partial f_\mu}{\partial x} \right) x_0 y_n(r)
\]

luego se tiene que

\[
y_{n+1}(r) = \left( \frac{\partial f_\mu}{\partial x} \right) x_0 y_n(r) + D \nabla^2 y_n(r)
\]
donde hemos empleado \( x_0 = f_\mu(x_0) \).

Podemos por lo tanto definir un operador lineal, dado por

\[
\mathcal{L}_{\mu,D} = \left( \frac{\partial f_\mu}{\partial x} \right)_{x_0} + D \nabla^2
\]

que nos permite escribir las ecuaciones dinámicas para las perturbaciones en la forma

\[
y_{n+1}(r) = [\mathcal{L}_{\mu,D}] y_n(r)
\]

Exploraremos ahora el problema de las estabilidades de los estados estacionarios antes mencionados, extendiéndolo a órbitas \( p \)-periódicas de periodo arbitrario. Como sabemos (capítulo 5) estas órbitas están definidas por el conjunto \( O^{(p)}_\mu = \{x_0, x_1, \ldots, x_{p-1}\} \), de forma que \( f^{(p)}_\mu(x_j) = x_j \) para cualquier \( x_j \in O^{(p)}_\mu \).

Después de \( p \) iteraciones, tendremos

\[
y_{n+1}(r) = [\mathcal{L}_{\mu,D}]^{(p)} \ldots [\mathcal{L}_{\mu,D}] y_0(r)
\]

o. de forma más explícita, se dará la relación

\[
y_{n+1}(r) = \left[ \prod_{j=0}^{p-1} \left( \frac{\partial f_\mu}{\partial x} \right)_{x_j} + D \nabla^2 \right] y_0(r)
\]

Recordemos que el primer término entre parentesis, esto es,

\[
S^{(p)}_\mu = \prod_{j=0}^{p-1} \left( \frac{\partial f_\mu}{\partial x} \right)_{x_j}
\]

define la estabilidad de la órbita \( p \)-periodica (May and Oster, 1976; véase capítulo 5) para la aplicación aislada \( (D = 0) \). Específicamente, la órbita espacialmente homogénea será estable si y sólo si \( |S^{(p)}_\mu| < 1 \). A continuación, podemos llevar a cabo un desarrollo de Fourier,

\[
y_n(r) = \sum_k \xi_n k e^{ikr}
\]

donde \( k \) es el vector de onda, siendo cada componente de \( k \) proporcional a la inversa de la longitud de onda (en una dirección dada). Podemos obtener una ecuación para la evolución temporal de los modos de Fourier

\[
\xi_{p,k} = \prod_{j=0}^{p-1} \left[ \left( \frac{\partial f_\mu}{\partial x} \right)_{x_j} + u(k) \right] \xi_k
\]

Las fronteras de estabilidad para las órbitas espacialmente homogéneas quedarán especificadas por (Oppo y Kapral, 1986)

\[
\prod_{j=0}^{p-1} \left[ \left( \frac{\partial f_\mu}{\partial x} \right)_{x_j} + u(k) \right] = \pm 1
\]

siendo estas curvas, definidas en el plano \((\mu, \beta)\), dependientes del acoplamiento sólo a través de \( u(k) \). En una dimensión, tendremos (Kapral, 1986)

\[
u(k) = 4D \sin^2 \left( \frac{\pi k_1}{L} \right)
\]
y, en dos dimensiones,

\[ u(k) = 4D \left[ \cos\left( \frac{\pi(k_1 + k_2)}{L} \right) \cos\left( \frac{\pi(k_1 - k_2)}{L} \right) - 1 \right] \]

(en general \( k_i = 0, 1, \ldots, L/2 \) serán las componentes en cada dirección). Podemos estudiar detenidamente la estabilidad de las órbitas. El primer modo de Fourier que se hace inestable corresponde a la longitud de onda mínima (la que aparece asociada al patrón en tablero de ajedrez). Aquí \( k_i = L/2 \) para todo \( i \) (Kapral, 1985). Para este caso particular tenemos, en dos dimensiones,

\[
2^{n-1} \prod_{j=0}^{j=n-1} \left\{ \left( \frac{\partial f_n}{\partial x_j} \right) - 8D \right\} = \pm 1
\]

donde asumimos explícitamente la existencia de un escenario de bifurcación con duplicación de periodo. Esta expresión analítica nos permite obtener un resultado ya mencionado: la desestabilización de órbitas periódicas homogéneas, inducida por la difusión (Kapral, 1985; Stassinopoulos y Alstrem, 1992). Típicamente, el sistema generará estructuras espaciales más y más complejas a medida que la difusión aumenta. Señalemos para terminar que estas estructuras pueden estudiarse empleando distintas aproximaciones, que incluyen la transformación de la red acoplada inicial en un automa celular (Chaté y Manneville, 1989).

### 10.12 Exponente de Lyapunov espaciotemporal

Un problema importante que se ha mencionado en distintos lugares de este libro, en relación con las medidas de caos en sistemas reales (especialmente biológicos) es la dificultad de obtener series temporales lo bastante largas, con la garantía adicional de que éstas reflejen un proceso estacionario. Esta situación es especialmente relevante en el caso de los ecosistemas, en los que disponemos de series temporales de corta duración que hacen difícil la caracterización cuantitativa correcta del tipo de dinámica. Pero también lo es en otros casos, como en el cerebro, cuya dinámica parece exhibir caos pero para la cual no podemos garantizar estacionariedad.

El hecho de que los sistemas reales, tanto los ecosistemas como el cerebro, posean tipicamente grados de libertad espaciales (esto es, son sistemas distribuidos en el espacio) proporciona algunas posibilidades adicionales que pueden permitir llevar a cabo una medida cuantitativa del tipo de comportamiento dinámico observado. En este sentido, podemos desarrollar una medida similar al exponente de Lyapunov previamente introducido, pero que emplee la información espacial junto con la temporal (Solé y Bascompte, 1995).

Supongamos, restringiéndonos al caso discreto, que consideramos una red sobre la que interactúa un conjunto de \( s \) especies, definida por las ecuaciones

\[ x_{n+1}^j(k) = F_j^x(x_n(k)) + C_j^x(x_n(k)) \]

(con \( j = 1, \ldots, s \), \( x = (x_1, \ldots, x_s) \)) y donde las funciones \( F_j^x(x) \), \( C_j^x(x) \) indican los términos de interacción y difusión, respectivamente. En general, tendremos una red \( r \)-dimensional

\[
\Lambda^r(L) = \left\{ k = (k_1, \ldots, k_r) \mid 1 \leq k_j \leq L \right\}
\]

(con \( r = 2 \) en nuestros ejemplos). Partiremos de una condición inicial arbitraria y una vez alcanzado el régimen estacionario seguiremos la evolución de una de las variables implicadas, esto es, \( \{x_1^j(k)\} \).
Figura 10.14: Comparación entre dinámicas temporales locales y dinámica a corto término. En (a) vemos una red logística con $L = 10$, $\mu = 4$ y $D = 0.05$, en la que representamos, para un punto dado, las sucesivas iteraciones durante 200 pasos de tiempo. En (b) hemos representado, para cada punto de la red, cuatro iteraciones. Vemos claramente la similitud entre ambas gráficas, que sugiere que la información espacial puede ser empleada para evaluar el comportamiento temporal.

Figura 10.15: Caracterización del caos espaciotemporal: supongamos dos configuraciones consecutivas en el tiempo, obtenidas a partir de una red acoplada. En este ejemplo, se trata de un modelo de presa depredador (se verá más adelante) para el que mostramos las poblaciones locales correspondientes a las presas. Tomando dos puntos de la red cuyo estado inicial difiere en menos de un umbral dado $\epsilon$, seguimos su evolución y comparamos las divergencias posteriores para realizar la estimación del exponente de Lyapunov espaciotemporal.
Figura 10.16: Exponentes de Lyapunov espaciales, \( \lambda_i(d) \) obtenidos para una red logística \( 10 \times 10 \). (a) Valores del exponente para distintas combinaciones de parámetros y para dimensiones crecientes. (b) Empleando \( D = 0.1 \) y distintos valores de \( \mu \), obtenemos el conjunto de exponentes para distintos valores de \( \mu \).

Para cada punto de la red \( k \in A'(L) \) durante un número dado (típicamente corto) de iteraciones \( n = 1, \ldots, m \).

Sea pues el conjunto definido, para cada punto de la red, por

\[
\Gamma^j(k) = \left\{ x^j_1(k), \ldots, x^j_m(k) \right\}
\]

es decir, por la evolución local del sistema. Para cada uno de estos conjuntos \( \Gamma^j(k) \) podemos reconstruir el atractor local en un espacio \( d \)-dimensional, con lo que tendremos un nuevo conjunto de vectores

\[
\Gamma^j_2(k) = \left\{ X^j_1(k) = \left\{ x^j_1(k), \ldots, x^j_{d+1}(k) \right\} \right\} \subset \mathbb{R}^d
\]

con \( i = 1, \ldots, m - d + 1 \). Consideremos ahora el conjunto de todos los puntos formados por

\[
\Gamma_d(A) = \bigcup_{k \in A'(L)} \Gamma^j(k)
\]

que constará de \( (m - (d - 1))L^2 \) puntos. Si el sistema es de baja dimensionalidad, como ocurre en muchos casos, una representación de este conjunto coincide básicamente con la que obtendríamos si dispusiéramos de una larga serie temporal local, procedente de un punto de la red. Un ejemplo de esta idea se muestra en la figura 10.14, en la que vemos ambas series. En la primera se ha dibujado el conjunto de pares de valores \( (x_n, x_{n+1}) \) obtenidos de un punto de una red logística (con \( L = 10, \mu = 4, D = 0.05 \)) empleando una serie temporal larga. Vemos claramente la forma de la aplicación logística. En la otra figura vemos el resultado de superponer los estados locales consecutivos empleando sólo cuatro iteraciones. Vemos básicamente el mismo tipo de estructura.

El exponente de Lyapunov espaciotemporal se calcula como sigue. Para una dimensión \( d \) dada, tomamos cada vector \( X^j_i(k) \in \Gamma^j(k) \) (con \( i = 1, \ldots, m - d \)) para \( \forall k \in A'(L) \). A continuación, buscamos aquellos puntos de la red \( h \) tales que \( X_i(h) \in A'(L) \) y \( (h \neq k) \), sean tales que su
distancia cumpla la desigualdad

\[ \|X_i^j(k) - X_i^j(h)\| = \left( \sum_{u=1}^{i+d-1} (x_u^j(k) - x_u^j(h))^2 \right)^{1/2} < \epsilon \]

donde \( \epsilon \in (0, 1) \) es la máxima separación permitida (definida en forma similar a lo visto en el capítulo sobre medidas de caos). Para un instante de tiempo \( i \), indicaremos estos pares por \(< k, h >\). A continuación calcularemos la distancia entre el siguiente par de valores,

\[ \|X_{i+1}^j(k) - X_{i+1}^j(h)\| \]

y de estas cantidades podremos obtener el exponente \( \lambda_s(d) \) que se define en la forma

\[ \lambda_s(d) = \frac{1}{N_p} \sum_{i=1}^{m-d} \sum_{<k,h>} \ln \left( \frac{\|X_{i+1}^j(k) - X_{i+1}^j(h)\|}{\|X_i^j(k) - X_i^j(h)\|} \right) \]

siendo \( N_p \) el número total de pares \(< k, h >\) comparados. Tomando \( \epsilon = 0.1 \) y distintos valores de \( d \), podemos analizar además el efecto de la dimensión. En la figura 10.16 vemos el resultado de los cálculos empleando una red logística de \( 10 \times 10 \) elementos. En la primera 10.16 (a) indicamos el efecto de la dimensionalidad \( d \) sobre el valor del exponente obtenido para distintos valores de difusión y de \( \mu \). Vemos, entre otras cosas, que parece existir una región de saturación del valor de \( \lambda_s(d) \) que sugiere que la dimensión empleada es suficiente para caracterizar la dinámica. Podemos conjeturar razonablemente (Solé y Bascompte, 1995) que la dimensión inferior a la que el exponente se satura podría ser la dimensión de la dinámica. En la figura 10.16 (b), vemos, para una red con \( D = 0.1 \), el efecto de la dimensionalidad. Una vez más, el valor de los exponentes espaciales, sobreestimados para dimensión uno (tal y como esperaríamos) adquieren valores similares para dimensión \( d = 2 \) y \( d = 3 \).

Este método ha sido empleado en distintos casos, revelando su capacidad para detectar caos e, indirectamente, determinismo, en series obtenidas a partir de un sistema espacialmente distribuido. Dado que el exponente no requiere una gran cantidad de datos en el tiempo, podemos obtener estimaciones a corto plazo del valor de \( \lambda_s(d) \) que, eventualmente, nos podrían permitir detectar transiciones dinámicas en sistemas complejos como el cerebro.

### 10.13 Supertransitorios y caos espacial

Uno de los efectos más dramáticos de la introducción del espacio sobre la dinámica de estos sistemas es la destrucción de las ventanas periódicas, como ya se ha mencionado. Sabemos que el caos determinista es estructuralmente inestable en la región del diagrama de bifurcación en la que caos y órbitas periódicas coexisten. Estas últimas son densas en este dominio del espacio de parámetros, de forma que, dado un valor \( \mu^* \) en el que aparezca caos, existirá, para cualquier entorno de radio arbitrario alrededor de \( \mu^* \), alguna órbita periódica.

La introducción del espacio modifica drásticamente esta situación, como se puede ver en la figura 10.17, en la que comparamos el diagrama de bifurcación de la aplicación logística \( x_{n+1} = f_\mu(x_n) = 1 - \mu x_n^2 \) (que es de hecho una transformación bajo cambio de parámetros de la logística antes introducida) con el correspondiente para una red unidimensional con acoplamiento no-lineal (dispersión), que escribiremos

\[ x_{n+1} = (1 - D) f_\mu(x_n(i)) + \frac{1}{2} D \left\{ f_\mu(x_n(i+1)) + f_\mu(x_n(i-1)) \right\} \]
Figura 10.17: (a) Diagrama de bifurcación de la aplicación logística \( x_{n+1} = f_\mu(x_n) = 1 - \mu x_n^2 \), en el que observamos la existencia de múltiples ventanas periódicas dentro del dominio caótico. (b) Diagrama correspondiente a la red acoplada, con \( N = 64 \) y difusión \( D = 0.2 \). Las ventanas periódicas han desaparecido. El diagrama se ha obtenido empleando una condición inicial aleatoria.

En el diagrama de bifurcación mostrado, obtenido para una red con \( N = 64 \) elementos y \( D = 0.2 \), vemos que el dominio de caos es ahora continuo. Si representáramos el valor de los exponentes de Lyapunov, las caídas hacia valores negativos dentro del régimen caótico estarían ahora ausentes. El caos espaciotemporal es, por lo tanto, estructuralmente estable. Todas las ventanas periódicas son destruidas para virtualmente cualquier condición inicial. El mecanismo de destrucción, como ha demostrado Kaneko, se origina a partir de dos procesos: la intermitencia espaciotemporal y la existencia de un fenómeno enormemente interesante: los supertransitorios (Kaneko, 1990; Crutchfield y Kaneko, 1988).

Supongamos que partimos de un caso especial de ventana periódica (que ya hemos estudiado con anterioridad): la ventana de periodo tres. Para acoplamientos débiles, con \( D < D_c = 10^{-3} \), el sistema tiende a un estado final no-caótico, de periodo tres. Para valores superiores, la red presenta estructuras no-periódicas ("defectos") que se propagan y colisionan dando lugar a explosiones turbulentas sobre la red que se desplazan dentro de dominios periódicos: tenemos intermitencia espaciotemporal. Para redes de pequeño tamaño, estos defectos que introducen comportamientos aperiódicos acaban aniquilándose entre sí y obtenemos una red con dinámica 3-periódica. Sin embargo, a medida que aumentamos el tamaño de la red, vemos que el tiempo característico \( T_N \) necesario para alcanzar el estado periódico aumenta con enorme rapidez, de forma que tenemos

\[ T_N = Ce^{-N} \]

(Kaneko, 1990), siendo \( r \propto (D - D_c)^N \), con \( \gamma \approx 1 \). Esta divergencia exponencial tiene consecuencias muy importantes. Por una parte, la divergencia hace que el tiempo que debamos esperar para observar la caída final en el estado periódico crezca astronómicamente con el tamaño de la red. Por otra parte, la observación de estos sistemas comporta un resultado inesperado: si seguimos, por ejemplo, el comportamiento del exponente máximo de Lyapunov a lo largo del tiempo, veremos que su valor se mantiene aproximadamente constante durante el transitorio, cayendo de forma prácticamente subita cuando el atractor periódico es alcanzado (figura 10.18 (b)). Otras magnitudes estadísticas pueden calcularse de forma similar, y también éstas presentan dicho comportamiento estacionario.
Figura 10.18: (a) Dinámica de una red con $N = 50$ para $\mu = 1.752$ (en el dominio de la ventana de periodo tres). Los puntos negros indican $|x_n(i+1) - x_n(i)| > 0.3$. Vemos la propagación de estructuras a lo largo de la red; (b) caída súbita del exponente de Lyapunov obtenida para la red anterior (véase texto).

durante el transitorio. Este resultado es muy interesante, en la medida en que el comportamiento transitorio, que habitualmente hemos considerado irrelevante, es estadísticamente estacionario. En otras circunstancias, el transitorio es un régimen en el que no observamos estacionariedad de ninguna clase, dado que se supone que el sistema converge hacia el comportamiento realmente estacionario, caracterizado finalmente por un atractor. En este caso, sin embargo, el transitorio es estadísticamente estable y, a menos que dispusiéramos de un conocimiento muy detallado del sistema (como por ejemplo el espectro completo de exponentes de Lyapunov) no podríamos distinguirlo de un comportamiento caótico típico. Por este motivo, hablamos de supertransitorios, que serán, para un sistema incluso moderadamente pequeño, el único comportamiento realmente observable. Crutchfield y Kaneko, para otro sistema dinámico definido por la aplicación

$$x_{n+1} = \frac{1}{3} \left[ \Phi(x_n(i+1)) + \Phi(x_n(i-1)) + \Phi(x_n(i)) \right]$$

con $\Phi(x) = Ax + B$, y tomando $A = 0.91$, $B = 0.1$, que generan ciclos de periodicidad $p = 25$, han explorado numéricamente el comportamiento de los supertransitorios, obteniendo una relación

$$T_N = T_1 2^\Gamma$$

con

$$\Gamma = \left[ \frac{N - 1}{N} \right]$$

donde $N_c = 21.5 \pm 0.5$, $T_1 = 149.5 \pm 0.5$ y $\alpha = 3 \pm 0.3$. Este resultado nos dice que, para una red de tamaño moderado, digamos $N = 128$, se tiene $T \approx 10^{64}$, un número astronómico: empleando un ordenador que calcule cada nuevo estado de la red en $10^{-15}$ segundos, tardaríamos $10^{46}$ años en alcanzar el estado final sobre el atractor periódico.

Las consecuencias de este resultado son considerables: podría ocurrir que un estado dinámico aperiódico, observado en un sistema espacialmente distribuido, ya sea un ecosistema o un fluido turbulento, correspondiera, desde el punto de vista estrictamente matemático, a un estado
transitorio. Desde el punto de vista práctico, nunca observaremos el comportamiento final y lo verdaderamente estacionario será de hecho el supertransitorio.

### 10.14 Competencia y caos espaciotemporal

En esta sección consideraremos un modelo simple de interacción entre dos especies que compiten por un recurso en cierto dominio espacial. Supondremos que sus generaciones no se solapan y que por lo tanto podemos emplear la aproximación discreta antes introducida. Consideraremos dos poblaciones \( x_n \) e \( y_n \) de insectos competidores que seguirán las siguientes ecuaciones dinámicas

\[
\begin{align*}
  x_{n+1} &= F^{(1)}(x_n, y_n) = \mu_1 x_n (1 - x_n - \beta_1 y_n) \\
  y_{n+1} &= F^{(2)}(x_n, y_n) = \mu_2 y_n (1 - y_n - \beta_2 x_n)
\end{align*}
\]

(con \( \mu \in [0, 4] \) y \( \beta > 0 \)) las cuales permiten definir un modelo simple de competencia interespecífica (Begon y Mortimer, 1986; Solé, Bascompte y Valls, 1992). Observemos que, si \( \beta_i = 0 \), estas ecuaciones quedan desacopladas entre sí y tenemos una red logística para cada caso. El parámetro \( \beta \) introduce por lo tanto el término de competencia, que modifica la estabilidad de estas poblaciones.

Si analizamos el modelo, veremos que posee cuatro estados posibles de equilibrio:

- **Punto de equilibrio trivial**, \( P_0 = (0, 0) \), tipicamente inestable.
- **Puntos de exclusión competitiva**, definidos por
  \[
  P_1 = (1 - 1/\mu_1, 0) \quad P_2 = (0, 1 - 1/\mu_2)
  \]
  y que implican la eliminación de una de las especies a cargo del mejor competidor.
- **Punto de coexistencia competitiva**, dado por \( P_c = (x^*, y^*) \), \( x^*, y^* > 0 \) en el que ambas especies coexisten de manera estable pese a la competencia.

Consideremos aquí el caso simétrico, esto es, \( \beta_1 = \beta_2 \) y \( \mu_1 = \mu_2 \) para \( i = 1, 2 \). Para este caso especial, se tiene que \( x^* = y^* = (1 - 1/\mu)/(1 + \beta) \), y el análisis lineal de la estabilidad nos da los dominios \( S_i \), definidos por

\[
S(P_{1,2}) = \{ (\mu, \beta) \mid \beta > 1; \; \mu \in (1, 3) \}
\]

para los puntos de exclusión y

\[
S(P_c) = \{ (\mu, \beta) \mid \beta \in (0, 1); \; \mu \in (1, 3) \}
\]

para el de coexistencia.

Para un valor de \( \mu \) dado, tendremos coexistencia si \( \beta < 1 \) y exclusión en caso contrario. En este sentido, la situación simétrica \( \beta_1 = \beta_2 \) y \( \mu_1 = \mu_2 \) implica que, en función de las condiciones iniciales, uno de los competidores (el mejor situado poblacionalmente) sobrevivirá y el segundo será eliminado. La predicción de la teoría es por tanto que la competencia generará, de ser lo bastante intensa, una simplificación de la diversidad, en la que los peores competidores desaparecen o, en nuestro caso, uno de los dos en función de la condición inicial. Este problema es de hecho un ejemplo más de rotura de simetría (figura 10.19) en el que pasamos de un atractor estable (para \( \beta < 1 \)) a dos atractores simétricos (para \( \beta > 1 \)) entre los que las fluctuaciones eligen por rotura de simetría. Como veremos, sin embargo, este resultado cambia notablemente en el caso de incluir el espacio.
La contrapartida espaciotemporal nos da el sistema definido por la red bidimensional

\[ x_{n+1}(r) = \mu x_n(r)(1 - x_n(r) - \beta y_n(r)) + D_1 \nabla^2 x_n(r) \]

\[ y_{n+1}(r) = \mu y_n(r)(1 - y_n(r) - \beta x_n(r)) + D_2 \nabla^2 y_n(r) \]

con \( k = (i, j) \) y siendo el acoplamiento

\[ \nabla^2 x_n(r) = \sum_j x_n(j) - q x_n(r) \]

(si escogemos el acoplamiento no-lineal los resultados básicos del modelo no se alteran). Consideraremos aquí \( q = 4 \) vecinos. Emplearemos la regla de truncamiento \( x_n(r) = 0 \) si \( x_n(r) < 0 \), pero de hecho ésta es sólo necesaria en casos muy inestables.

Antes de proceder con estudio analítico, podemos llevar a cabo algunas simulaciones numéricas empleando las ecuaciones anteriores. Partiendo de una condición inicial al azar (una pequeña fluctuación alrededor del estado estacionario homogéneo, que tomaremos por \( x(i,j) = x^* + \delta x(i,j) \), \( y(i,j) = y^* + \delta y(i,j) \)) iteramos nuestro sistema para distintas combinaciones de parámetros. Tomemos por ejemplo una situación en la que \( \mu \) daría un estado estable (attractor puntual, \( \mu \in [1, 3] \)) de coexistencia, con \( \beta < 1 \). Una simulación nos demuestra que nada especialmente interesante ocurre: la coexistencia que predicen las ecuaciones se traslada de manera lineal a la red, sin más consecuencias. Pero lo sorprendente ocurre para \( \beta > 1 \). La predicción es que una de las especies eliminará a su competidora. Sin embargo, inmediatamente surge la pregunta de cómo afectará el espacio en cada caso, dado que en distintos puntos tendremos distintas condiciones iniciales. En la figura 10.19 vemos un ejemplo dramático del resultado de esta simulación: aparecen estructuras espaciales (de una escala claramente superior a la de interacción) en las que comprobamos la validez.
Figura 10.20: Cuatro ejemplos de estructuras finales obtenidas sobre una red de $30 \times 30$ nodos, con $D = 0.05$, $\mu = 2.5$ y $\beta = 1.1$. Para las cuatro, la condición inicial es una distribución uniforme con una pequeña fluctuación al azar añadida, distinta en cada caso.

Local del principio de exclusión. En algunos dominios, una especie ha excluido claramente a la otra. Las zonas más oscuras indican, para la especie representada, los lugares menos poblados por ésta, mientras que las zonas más claras indican la exclusión de la segunda. Localmente, se ha producido efectivamente rotura de simetría, pero globalmente, y como consecuencia de ese fenómeno, ambas especies coexisten. La aparición de una nueva fase heterogénea puede caracterizarse mediante la introducción de un parámetro de orden adecuado. Definamos en nuestro caso el siguiente:

$$\Omega_\beta = \frac{1}{L^2} \sum_i \sum_j |x(i, j) - y(i, j)|$$

que será cero si tenemos coexistencia y por lo tanto (pese a posibles fluctuaciones) los valores locales tienden a ser iguales, y positivo si el sistema crea estructuras y por lo tanto el valor absoluto de la diferencia es no nulo. En la figura 10.19 representamos un ejemplo de cálculo de este parámetro, que exhibe una brusca transición en $\beta_c = 1$, como era de esperar.

Además, podemos apreciar el efecto de la sensibilidad del patrón final respecto de las aleatoriedades iniciales llevando a cabo varias simulaciones en las que sólo variamos las fluctuaciones locales. El resultado de cuatro simulaciones distintas se muestra en la figura 10.20, en la que hemos representado el estado final.

Ahora, la estabilidad de las órbitas periódicas puede analizarse siguiendo los pasos anteriores. Consideremos la transformada de Fourier para las ecuaciones lineales asociadas a cada punto,

$$\delta \xi(k, t) = L(k, P^*) \delta \xi(k, t)$$

con la que llevaremos a cabo el cálculo de las bifurcaciones que generan patrones espaciales. Tendremos una ecuación característica dada por

$$P(\lambda) = \left| L_{\mu, \beta}(\partial_i F^{(j)}(P^*)) + D\Theta(k) - \lambda \right| \delta_{ij} = 0$$

donde $L_{\mu, \beta}(\partial_i F^{(j)}(P^*))$ denota los elementos de la matriz de Jacobi.
Introduciremos la dependencia explícita del espacio mediante el vector de onda,

\[ \Theta(k) = 4 \left\{ \cos \left( \frac{\pi (k + l)}{N} \right) \cos \left( \frac{\pi (k - l)}{N} \right) - 1 \right\}, \quad k = (k, l) \]

y de aquí podemos encontrar las fronteras de estabilidad de nuestro modelo. Para valores dados de \( D \), los dominiios del espacio de fases

\[ \Gamma = \{ (\mu, \beta) \mid \mu \in (0, 4) ; \beta > 0 \} \]

se obtendrán a partir de la condición crítica \( \lambda_i = \pm 1 \). Los valores propios correspondientes a los dos primeros puntos críticos serán

\[ \lambda_1(k) = 2 - \mu + D\Theta(k) \]
\[ \lambda_2(k) = \mu(1 - \beta) + \beta + D\Theta(k) \]

y para el punto de coexistencia, tendremos

\[ \lambda_1(k) = 2 - \mu + D\Theta(k) \]
\[ \lambda_2(k) = \mu \left[ 1 - (1 - 1/\mu)/(2 + \beta) \right] + D\Theta(k) \]

Nos restringiremos a los dos primeros, para los que podemos hallar las curvas correspondientes a los puntos de estabilidad marginal, dados por \( \lambda_j(k) = \pm 1 \) lo que proporciona las fronteras

\[ \mu_+ = 1 + D\Theta(k) \]
\[ \mu_- = 3 + D\Theta(k) \]
\[ \beta_+ = \frac{1}{\mu - 1} \left[ \mu - 1 + D\Theta(k) \right] \]
\[ \beta_- = \frac{1}{\mu - 1} \left[ \mu + 1 + D\Theta(k) \right] \]

Ahora, podemos elegir vectores \( k \) adecuados para obtener los dominios de estabilidad. Para \( k = (0, 0) \) (es decir, estados homogéneos) y para \( k = (N/2, N/2) \) (estados en tablero), tenemos \( \Theta(k) = 0 \) y \(-8\) respectivamente. Empleando estos valores, podemos hallar las fronteras para las bifurcaciones que rompen simetría dando lugar a heterogeneidades en el espacio, que serán

\[ \mu_0^+ = 1 ; \mu_0^+ = 3 \]
\[ \mu_0^{N/2} = \mu_0^+ - 8D ; \mu_0^{N/2} = \mu_0^+ - 8D \]
\[ \beta_0^+ = 1 ; \beta_0^+ = (\mu + 1)/(\mu - 1) \]
\[ \beta_0^{N/2} = \beta_0^+ - \frac{8D}{\mu - 1} \]
\[ \beta_0^{N/2} = \beta_0^+ - \frac{8D}{\mu - 1} \]

donde \( \xi_k^\pm \) y \( \xi_k^{N/2} \) son los valores \( (\xi = \mu, \beta) \) asociados a cada uno de los vectores de onda \( k = (0, 0) \) y \( k = (N/2, N/2) \), respectivamente. Como esperaríamos, la introducción de la difusión en un espacio explícitamente definido genera la aparición de inestabilidades. Para \( D = 0.05 \), la figura 10.20 muestra el espacio de fases obtenido para nuestro sistema, en el que algunas de las fronteras se han obtenido analíticamente con el procedimiento anterior y las restantes mediante métodos numéricos (como el cálculo del exponente de Lyapunov). El resultado más interesante de este estudio (aparte de la coexistencia global antes discutida) es la estabilidad de las estructuras de Turing generadas incluso en el dominio caótico. En esta zona observamos la presencia de comportamientos locales desordenados (figura 10.21 (a)) típicos del caos espaciotemporal que, sin embargo, no destruyen las fronteras que definen las distintas zonas. Estas estructuras han sido denominadas \textit{estructuras de Turing caóticas} y demuestran de manera evidente la posibilidad de que el caos temporal coexista sin dificultad con estructuras macroscópicas ordenadas, como las que vemos en la figura 10.21 (b).
Figura 10.21: (a) Dinámica local de una estructura de Turing caótica. Se indica, dentro de un dominio, la dinámica de la especie ganadora (línea continua) y la de la perdedora, que aparece a niveles bajos en línea discontinua. Pese a las grandes fluctuaciones poblacionales, que pueden incluso acercarla a la extinción local (indicada por una flecha) la estructura permanece. (b) Estructura de Turing caótica.

10.15 Ondas espirales en redes acopladas

Consideremos ahora la formación de un grupo distinto de estructuras, que ya mencionábamos en el capítulo anterior, y que podían ser generadas mediante autómatas celulares: las ondas espirales. Estas estructuras aparecen, como ya hemos mencionado, en multitud de sistemas complejos y son muy características de los modelos de reacción-difusión que describen medios excitable (Murray, 1990). Estas ondas (junto con otros tipos de estructuras) han sido halladas en modelos CML de distintos tipos como son el modelo de Lotka-Volterra discreto (Solé y Valls, 1991) y modelos de huésped-parasitoide (Hassell et al., 1991; Solé et al., 1992), que consideraremos en esta sección.

Supongamos que partimos del sistema dinámico descrito por el siguiente conjunto de ecuaciones

\[
x_{n+1} = \mu x_n (1 - x_n) F_\beta(x_n, y_n)
\]

\[
y_{n+1} = x_n (1 - F_\beta(x_n, y_n))
\]

(10.15.1)

que ha sido empleado, para ciertas elecciones de \( F_\beta \), en dinámica de poblaciones (Begon y Mortimer, 1986). Aquí \( x_n \) e \( y_n \) son un tipo especial de poblaciones de presa y depredador conocidas en la literatura como modelos huésped-parasitoide. Este modelo difiere del modelo de Lotka-Volterra debido a la forma especial de depredación que exhiben los parasitoides, que no matan a su víctima sino que la inmovilizan e introducen en la misma (que permanece viva) un huevo que se alimentará de ésta.

Asumiremos que el término de depredación es tal que \( F_\beta(x, \infty) = 0 \) y \( F_\beta(x, 0) = 1 \). Tomemos por ejemplo (Solé et al., 1992) \( F_\beta(x, y) = \exp(-\beta y) \). Este sistema dinámico posee un rico repertorio de comportamientos dinámicos, entre los cuales hallamos una transición hacia la cuasiperiodicidad, como podemos apreciar en el diagrama de bifurcación de la figura 10.22, en el que hemos tomado \( \mu = 4 \) (que emplearemos a lo largo del estudio de este sistema) y un intervalo de valores de \( \beta \in (1, 7) \). Para \( \beta < \beta_c = 1.75 \) aparece dinámica caótica debido a la inestabilidad (con extinción)

de la población de depredadores cuya ausencia genera comportamientos caóticos en la ecuación logística. Para \( \beta \in (\beta_c, 3.4) \), tenemos un colapso de esta solución en un atractor puntual que sufrirá posteriormente sucesivas bifurcaciones con órbitas de periodo 3 (puede demostrarse que no aparecen de otra periodicidad menor), atractores cuasiperiódicos y caos.

A continuación, podemos construir la contrapartida dependiente del espacio,

\[
x_{n+1}(k) = \mu x_n(k)(1 - x_n(k))e^{-\beta y_n(k)} + D_1 \nabla^2 x_n(r)
\]
\[
y_{n+1}(k) = x_n(k)(1 - e^{-\beta y_n(k)}) + D_2 \nabla^2 y_n(r)
\]

donde, nuevamente, tenemos

\[
\nabla^2 x_n(r) = \sum_{j} x_n(j) - qx_n(k)
\]

y, cuando fuera necesario, \( x_n(k) = 0 \) si \( x_n(k) < 0 \).

Supongamos que partimos de una situación en la que la población de presas se halla presente en todo el sistema y de un número reducido de puntos en los que se introducen inicialmente los depredadores. Tendremos por lo tanto

\[
\delta_1 < x_0(k) < \delta_2 \ \forall k \in \Lambda(L)
\]
\[
\delta_1 < y_0(k) < \delta_2 \ \forall k \in K_s
\]

siendo \( K_s = \{k_1, k_2, \ldots, k_s\} \) un número finito \( s < L^2 \) de puntos de la red escogidos al azar y \( k_j \in \Lambda(L) \). Tomaremos \( \delta_i \) tales que \( 0 < \delta_1 < \delta_2 < 1 \) (aquí \( \delta_1 = 0.3 \) y \( \delta_2 = 0.4 \)). El resultado de simular estos sistemas es la generación de una enorme variedad de estructuras, típicamente no-estacionarias, algunas de las cuales vemos en la figura 10.23. Las ondas espirales aparecen en abundancia en presencia de distintos tipos de dinámica, tanto periódica como cuasiperiódica o caótica. Pese a formarse en condiciones de inestabilidad, las ondas mantienen su coherencia en el régimen caótico, como vemos en la figura 10.23, para una red 512 × 512. Tenemos de hecho un amplio dominio del espacio de coeficientes de difusión en el que aparecen ondas espirales.
Figura 10.23: Estructuras generadas por el modelo huésped-parasitoide antes introducido (véase el escenario de la figura anterior). Observamos la formación de ondas espirales sobre una red de tamaño $512 \times 512$. Empleamos $\mu = 4$ y (a) $\beta = 4$ (cuaunperiódico), $D_1 = 0.001$ y $D_2 = 0.2$; (c) $\beta = 5.0$ (caótico), mismos $D_1$ y $D_2$.

Figura 10.24: Espacio $(D_1, D_2)$ de parámetros para el modelo de huésped-parasitoide. Vemos tres dominios cualitativamente distintos.
Figura 10.25: Ondas espirales obtenidas mediante un modelo de huésped-depredador con movimiento dispersivo y dependiente de la densidad de presas (Rohani y Miramontes, en prensa). Observamos la aparición de ondas espirales con distintos grados de coherencia

Para $\beta \in (3.5, \beta_h)$ esto es, en el régimen cuasisperiódico del sistema original, aparece una zona de caos inducido por difusión alrededor de

$$D_\epsilon = \{(D_1, D_2) \mid D_2 \approx \frac{1}{5} - D_1\}.$$ 

que de hecho separa dos dominios cualitativamente distintos, $D^-_\epsilon$ (para $D_2 < 1/5 - D_1$) y $D^+_\epsilon$ (en caso contrario). Para estos valores, el exponente de Lyapunov es cero y aparecen ondas espirales cuando nos encontramos en $D^-_\epsilon$. Por encima, no existen estructuras regulares de gran tamaño y sólo observamos dinámica caótica.

La aparición de ondas espirales en modelos de este tipo ha sido comprobada por otros estudios distintos (Rohani y Miramontes, en prensa) que han obtenido resultados similares, confirmando la existencia del escenario de bifurcación RTN. En la figura 10.25 vemos un ejemplo de los resultados obtenidos en estos modelos, que sugieren la universalidad de las ondas espirales en sistemas de presa-depredador como los antes descritos. El empleo de formas más sofisticadas de movimiento, empleando dispersión (acoplamiento no-lineal) e incorporando elementos más realistas como el movimiento de los depredadores en la dirección de los lugares más ricos en presas, confirma los resultados previos. De hecho, incluso la simulación de un ecosistema en el que detallamos el estado de cada individuo, incluyendo su edad, tamaño, sexo, movimiento al azar, alimentación, etc. (los llamados modelos orientados al individuo) nos permite ver cómo en un sistema de este tipo, en el que se incorporan estocasticidades de muchos tipos, se puede dar lugar a ondas coherentes (Bascompte et al., 1996). En este estudio, se emplea información acerca del ciclo de vida de cierta especie de liebre que habita en Canadá y que es depredada por los línces (este sistema fue introducido en el capítulo 5). Cada iteración es ahora un complejo conjunto de pequeñas reglas que definen el ciclo vital de cada especie en cada punto. Este modelo ha permitido dar una explicación plausible de la gran coherencia espaciotemporal de las oscilaciones de estas poblaciones, que serían debidas a un proceso de autoorganización en el que las ondas espaciales (claramente sugeridas a partir de los datos de campo) serían una cara de la moneda y las oscilaciones temporales (de las que se dispone de datos) la otra. El espacio y el tiempo son dos aspectos a considerar de forma
profundamente dependiente.

Bibliografía


Capítulo 11

Redes de Kauffman

En capítulos previos hemos estudiado distintas aproximaciones (básicamente deterministas) al problema de la dinámica de sistemas complejos. La existencia de propiedades emergentes, de propiedades macroscópicas nuevas, no presentes o reducibles a las de los elementos componentes, es el tema básico del discurso de la complejidad. Hemos comprobado además que la obtención de propiedades emergentes no requiere, en general, el empleo ni de interacciones ni de elementos básicos complicados. Los efectos no-lineales poseen la capacidad intrínseca de generar complejidad. Además, si lo complejo surge cerca de puntos críticos (véanse los capítulos 7 y 8) la descripción simplificada del sistema puede ser más que suficiente.

En este capítulo vamos a detenernos a explorar el comportamiento de uno de los modelos (o aproximaciones teóricas) más conocidos dentro del estudio de la complejidad: las redes Booleanas aleatorias, conocidas también como redes de Kauffman (RK) (revisadas en profundidad en Kauffman, 1993). Estos son sistemas dinámicos discretos en el espacio, en el tiempo y en el conjunto de estados \( \Sigma \) que, típicamente, está limitado a dos valores: \( \Sigma = \{0,1\} \). Estas propiedades son comunes con los automatas celulares, pero la forma en que los elementos se conectan entre sí no lo es. Para cada elemento \( S_i \in \Sigma \) (con \( i = 1, \ldots, N \) elementos formando la red) elegimos al azar exactamente \( K \) de entre los \( N \) elementos, los cuales servirán de “vecinos” o entradas con las que definiremos las reglas dinámicas. Así que ahora el entorno de un automata dentro de la red se define por un conjunto elegido al azar, lo que, necesariamente, dará lugar a una pérdida de correlaciones.

Las redes de Kauffman incorporan la aleatoriedad a un segundo nivel: la dinámica de la red podría ser homogénea, en el sentido de que la función que defina los cambios de cada elemento sea la misma para todos, y sólo dependa de las entradas particulares. Por el contrario, la dinámica vendrá definida por

\[
S_i(t + 1) = \Lambda_i \left[ S_{i_1}(t), S_{i_2}(t), \ldots, S_{i_K}(t) \right]
\]

siendo \( \Lambda_i \) una función Booleana elegida al azar de entre el conjunto \( \mathcal{F}_K \) de todas las posibles funciones booleanas con \( K \) entradas. En consecuencia, la red de Kauffman incorpora aleatoriedad a dos niveles muy importantes: en la forma de elegir los vecinos y en la forma en que éstos van a interactuar.

Pese a que la aleatoriedad de las RK es enorme, esta red posee propiedades de organización interna sorprendentes. Entre las aplicaciones de este formalismo (que son muchas) destaca por su interés la organización del genoma y el problema de la diferenciación celular, que analizaremos en este capítulo. Para entender el alcance de los resultados de Kauffman, daremos una breve introducción al problema del control de la expresión genómica.

407
Figura 11.1: Tipos celulares: en un organismo complejo, encontramos distintos tipos de células diferenciadas, que llevan a cabo tareas distintas. Su número es muy inferior al que esperaríamos de las combinaciones posibles de los genes presentes. ¿Por qué?

11.1 Control de la expresión genómica

Las células que componen los tejidos de un organismo superior son muy distintas entre sí, tanto en su morfología como en su función específica. En algunos casos están enormemente especializadas, de forma que llevan a cabo la síntesis de enormes cantidades de una o muy pocas proteínas. Este hecho nos indica claramente que algunos genes están muy activos: la diferenciación consistiría (al menos en parte) en la amplificación\(^1\) de algunas secuencias particulares de ADN que codifican las proteínas observadas.

Tipicamente, la diferenciación celular depende de cambios en la selección de secuencias de ADN. La naturaleza de esta selección hace que la diferenciación celular sea un fenómeno irreversible (en general) aunque no existan cambios realmente irreversibles en las propias secuencias.

Los resultados experimentales nos permiten concluir que:

- Todas las células de un organismo multicelular contienen exactamente el mismo ADN.
- Los distintos tipos celulares sintetizan distintos conjuntos de proteínas (estos conjuntos definen, en último término, el tipo celular observado).
- Células de tipos distintos transcriben distintos conjuntos de genes.

(para un resumen claro de los experimentos llevados a cabo para contrastar estos resultados, véase el capítulo 8 de Alberts et al., 1993).

Hoy sabemos que la activación o inactivación de genes tiene lugar a través de la acción de proteínas específicas que permiten activar o reprimir la actividad de genes concretos. El estudio de sistemas de regulación muy simples, como es el caso del bacteriófago λ, demuestra que incluso un conjunto pequeño de genes reguladores puede originar un patrón de comportamiento de gran

\(^1\)En Biología molecular, esta expresión hace referencia a la generación de copias múltiples de cierta estructura molecular. En nuestro caso, se refiere a la generación de proteínas específicas, resultantes de la activación selectiva de un gen.
Redes de Kauffman

complejidad, con muchas soluciones posibles. La consecuencia de esta observación es que si un sistema relativamente simple, como el fago λ, puede dar lugar a comportamientos complejos, las posibilidades de un genoma del tamaño de una célula compleja son, en principio, astronómicas.

Si el tipo celular depende de un patrón dado de activación genómica, entonces podríamos esperar observar una colección enorme de tipos celulares. Si, para simplificar, tomamos el estado de un gen como una variable binaria (véase más adelante), esto es, "0" para un gen inactivo y "1" para un gen activo, tendremos, siguiendo la nomenclatura de los automatas celulares, un conjunto de estados Booleano, $\Sigma = \{0, 1\}$, tal y como indicábamos en la introducción. Si nuestro sistema está formado por $N$ genes, entonces el conjunto de todos los posibles tipos celulares incluye

$$N_T = |\Sigma|^N = 2^N$$

tipos, lo que para valores grandes de $N$ (como es típico) nos lleva a una enorme diversidad potencial de células diferenciadas. Sin embargo, cuando nos acercamos al sistema real observamos que el número efectivo de células diferenciadas distintas es muy reducido: del orden de la raíz cuadrada del número estimado de genes

$$N_T = \sqrt{N}$$

. una cantidad enormemente inferior.

En el caso del genoma humano, para el que $N \approx 150000$ genes, tendríamos un potencial de no menos de $N_T \approx 10^{30000}$ pautas de expresión genómica. Sin embargo, el recuento de tipos nos da sólo 254 tipos celulares (cercano al que nos daría la raíz cuadrada del número de genes). ¿Cómo explicar este resultado?

Existe una propiedad adicional de gran importancia: la enorme estabilidad de los tipos celulares. Una vez se ha definido la identidad celular final (la célula diferenciada) este estado se mantiene estable. En condiciones normales, una célula diferenciada, sea ésta una neurona, célula hepática o macrófago, mantendrá su identidad hasta la muerte celular. Ocasionalmente, sin embargo, un error en la pauta adecuada de interacción puede destruir esta estabilidad. Algunos genes que habitualmente se hallan "dormidos" (o, más exactamente, reprimidos) vuelven a la actividad sin que un control adecuado pueda evitarlo. Un ejemplo lo dan los oncogenes, genes implicados en la aparición del cáncer. Por alguna causa (como puede ser la inserción de un retrovirus en las proximidades del gen) el oncogen, que normalmente se halla implicado en procesos básicos de diferenciación y/o división celular, y que estaría normalmente reprimido en la célula diferenciada, deja de estarlo. La célula, libre de mecanismos que lo impidan, se replica sin límites. Evitar esta situación y otras obliga al sistema de control a asegurar una adecuada diferenciación y su posterior estabilidad 2. Nuestra primera impresión es que existirá un mecanismo de control enormemente preciso capaz de coordinar la delicada red de interacciones (Tjian, 1995). Este mecanismo sería el resultado de un largo proceso de selección natural, en el que los sucesivos ensayos habrían generado una red dinámica de genes en interacción. En principio, una adecuada comprensión de esta red pasaría por un conocimiento detallado de todos y cada uno de sus componentes o al menos de gran parte de ellos, lo cual es una tarea irrealtizable. Pero, ¿es realmente así?

En este capítulo daremos una respuesta a la preguntas anteriores basada en un enfoque muy distinto del de la biología molecular (esencialmente reduccionista) basada en un enfoque teórico inspirado en las redes Booleanas. Emplearemos la aproximación introducida ya en 1967 por Stuart Kauffman (Kauffman, 1969). Como veremos, se trata de un modelo muy simple en el que las propiedades básicas de la diferenciación celular emergen de forma espontánea. Antes de introducir este modelo, revisaremos brevemente el problema del control de la actividad genómica así como una justificación formal de la elección de un modelo discreto basado en elementos binarios.

2Debemos señalar que el control no es sólo función de los mecanismos celulares de regulación genómica, sino también del "ambiente celular" cercano. Una misma célula puede dar un tipo celular u otro en función de las células cercanas con las que se comunica.
Figura 11.2: El modelo del operón: planteado por Jacob y Monod, considera el mecanismo de regulación de un grupo de genes en el que un gen represor (R) da lugar a un proteína que, unida al sitio operador (O) (cerca del promotor, (P)) puede reprimir la expresión de genes estructurales (SG).

11.2 Regulación compleja, modelos simples

¿En qué forma tiene lugar la regulación de la activida genómica? La existencia de mecanismos de regulación de la actividad de grupos de genes se puso de manifiesto a raíz de los estudios de los biólogos franceses Jacob y Monod. Estos propusieron el modelo del operón, cuya estructura básica se representa en la figura 11.2.

Las distintas partes del modelo se indican por: R = gen regulador, P = promotor (o centro promotor, al que se une el enzima que transcribe la cadena de ADN), O = operador y SG = genes estructurales. En este modelo, el gen regulador da lugar a una proteína que actúa de represor de los genes estructurales al interaccionar con el operador. El operador sería en este caso adyacente a los genes bajo control. La biología molecular de estos sistemas es bien conocida (a menos en algunos casos) y no vamos a profundizar en ella (para una revisión de estos aspectos, véase Watson et al., 1992). Pero de hecho los estudios posteriores han revelado una compleja maquinaria de control implicada en la actividad de los genes. Durante la transcripción, esto es, durante el proceso de lectura de una secuencia de ADN que codifica un gen, una polimerasa de ARN lee la cadena de ADN y “transcribe” esta información en una nueva cadena libre de ARN. Eventualmente, esta cadena de ARN será empleada en la construcción de una proteína o conjunto de éstas (es el proceso de traducción). Pero la transcripción, que es la etapa en la que el control se hace efectivo en una forma u otra, no consiste sólo en la lectura (o no) de un gen. Los estudios llevados a cabo (figura 11.3) indican claramente que otras moléculas participan en este proceso.

Con el objeto de que la polimerasa de ARN se una de forma fiable al promotor, ciertos factores (proteínas) deben unirse a ésta, formando un complejo molecular que logra que la unión tenga lugar. En bacterias, estos factores se conocen como factores sigma, pero éstos no se hallan presentes en células eucariotas (con núcleo). En células con núcleo encontramos una estructura fina de control mucho más sutil. En la secuencia de ADN aparecen, cerca del centro promotor, zonas intensificadoras (que estimulan la transcripción) y zonas silenciadoras, que la inhiben. Los genes eucariotas cuentan con distintas combinaciones de estas zonas. Existen distintos tipos de moléculas...
Figura 11.3: Mecanismo de transcripción: al ADN se unen no sólo la molécula de polimerasa que lleva a cabo la lectura de la cadena de ADN correspondiente, sino también un conjunto de moléculas que colaboran en una forma u otra en este proceso.

que se unen a estas zonas (activadores y represores, figura 11.3) que a su vez se ensamblan con otras moléculas. El resultado final de esta estructura es que el proceso de regulación de la transcripción (y en último término de la diferenciación) es muy complejo. Nuevamente, este resultado nos hace pensar en una descripción complicada del sistema global que, en principio, requeriría un conocimiento muy detallado de cada una de sus partes.

Incluso en el caso de que nos limitáramos a tratar un problema simple, la regulación depende de las concentraciones de proteínas presentes en la célula, y en consecuencia parece que la aproximación continua, en lugar de una formalización discreta, sería mucho más apropiada. Daremos sin embargo un ejemplo de modelo de regulación basado en un problema de naturaleza continua, que nos permitirá justificar la validez de la aproximación discreta.

Consideremos el siguiente modelo de comportamiento del bacteriófago lambda, un virus muy bien conocido que invade las bacterias Escherichia coli. Este virus posee dos tipos distintos de invasión: en uno de ellos (el ciclo lisogénico) termina integrado en el genoma de la célula huésped y por lo tanto se replicará cada vez que la bacteria lo haga. En el otro caso (el ciclo lítico) el virus se multiplica en el interior de la célula, dando muerte a ésta cuando la cantidad de partículas viricas es lo bastante grande. Una vez establecido, este ciclo se mantiene.

Podemos imaginar un modelo simple a partir del mecanismo de estas estrategias, en las que se hallan implicadas dos proteínas: el represor lambda, que bloquea el gen para la expresión de la proteína cro y ésta última, que a su vez bloquea la expresión del primero. Si el gen cro se expresa, el fago se multiplica en la célula y si no, se integra en el genoma.

Puede construirse una tabla muy simplificada para definir los estados del fago. Dado que está formado por $X = 2$ elementos (en nuestra aproximación, por supuesto) y que cada uno de ellos envía un input al segundo ($K = 1$) podemos ver que los atractores observados son (10) para el caso en el que el represor está activo y cro inactivo, y el segundo estado estable (01) que no es sino el complementario. La idea es entonces considerar los estados (00) y (11) como transitorios.
Figura 11.4: Plano de fases del sistema dinámico de inhibición mutua (para \( n > 2 \)) en el que vemos la existencia de un punto inestable (de silla) acompañado de dos atractores estables, situados en las proximidades de los puntos \((0,1)\) y \((1,0)\).

Este modelo es muy simple, y existen análisis cuidadosos más interesantes acerca del mecanismo de regulación del fago lambda (Kauffman, 1993; Watson et al., 1987 (vol. II, cap. 17)) que indican claramente que incluso un genoma tan simple como el de un virus bacteriano puede exhibir un alto grado de complejidad en sus estados.

Podemos construir un modelo continuo que nos permita, por una parte, dar cuenta del tipo de dinámica que exhibirá un sistema como este anterior, en el que se da represión cruzada (inhibición mutua) entre dos genes, y las relaciones entre dinámica continua y su contrapartida discreta.

Un modelo de inhibición mutua es

\[
\frac{dx}{dt} = f(x, y) = \frac{\beta^n}{\beta^n + y^n} - x
\]

\[
\frac{dy}{dt} = g(x, y) = \frac{\gamma^n}{\gamma^n + x^n} - y
\]

donde \( x, y > 0 \) son las variables que denotan las concentraciones de proteínas, \( n > 2 \) un entero y \( \beta = 1/2 \). Vemos que para que aparezca cada tipo de molécula es preciso que la otra se halle presente, aunque su efecto sea inhibitorio.

El estado estacionario de este sistema es \( x^* = y^* = 1/2 \). Si analizamos su estabilidad, obtenemos la siguiente matriz de Jacobi

\[
L_n(x^*, y^*) = \begin{pmatrix}
-1 & -\frac{n}{2} \\
-\frac{n}{2} & -1
\end{pmatrix}
\]

que proporciona los valores propios

\[
\lambda_1 = -1 + \frac{n}{2} \quad \lambda_2 = -1 - \frac{n}{2}
\]

cy cuya estabilidad vendrá determinada por el valor de \( n \). Para \( n > 2 \), el diagrama de fases del sistema se representa en la figura 11.4. Tenemos un punto inestable y dos puntos estables situados
Figura 11.5: Ejemplo de red de Kauffman con \( N = 3 \) y \( K = 2 \). Indicamos (a) la topología de la red empleada y las funciones booleanas asociadas a cada elemento, dadas por las tablas indicadas. (b) Tabla completa de transiciones, definida a partir de la red inicial. (c) Atractores del sistema: a partir de la tabla, podemos deducir las transiciones y los estados finales (los atractores del sistema).

en las proximidades de \((1,0)\) y \((0,1)\), esto es, cerca de lo que habíamos empleado como definición de atractores en el caso discreto. El modelo continuo captura la esencia del fenómeno de inhibición mutua (común a muchos sistemas biológicos, incluidos los ecosistemas) a la vez que prueba la adecuación de la aproximación discreta, más tratable cuando el sistema se hace grande.

### 11.3 Redes de Kauffman

Consideremos una red de elementos binarios que, de forma genérica, llamarémos generes. Estos elementos se supondrán conectados entre sí de forma aleatoria, tal y como se indicaba en la introducción. Cada gen recibirá entradas procedentes de exactamente \( K \) genes elegidos al azar de entre los \( N \) posibles. La interacción entre éstos puede ser más o menos compleja y dependerá de las propiedades específicas de conjunto. Para poder expresar esta complejidad de forma explícita, emplearemos un conjunto de funciones Booleanas \( \mathcal{F}_K = \{ \Lambda_i \} \), de forma que los estados de los elementos, \( S_i(t) \in \Sigma \) evolucionarán con el tiempo siguiendo un sistema dinámico descrito por

\[
S_i(t + 1) = \Lambda_i \left[ S_{i_1}(t), S_{i_2}(t), \ldots, S_{i_K}(t) \right]
\]

estos modelos son conocidos también con el nombre de modelos \( NK \).

De forma similar a lo que vimos en el capítulo acerca de los autómatas celulares, las funciones Booleanas quedan especificadas por una tabla en la que indicaremos los estados de cada uno de los elementos conectados y su nuevo valor. La tabla define el conjunto de posibles transiciones y a partir de ella podemos hallar los atractores del sistema y las cuencas de atracción del mismo. En la figura 11.5 damos un ejemplo de una red de tres elementos \((N = 3)\) conectados a pares \((K = 2)\).

Las funciones Booleanas asignadas a cada elemento (que son de hecho funciones lógicas simples, AND y OR) así como la tabla general que indica las transiciones para este ejemplo se indican también. Como resultado de esta elección, existen tres estados asintóticos a los que el sistema puede
Figura 11.6: Ejemplo de cuenca de atracción de una red de Kauffman con $N = 13$ y $K = 3$, en el que se indica de forma esquemática cómo distintas condiciones iniciales tienden a otros estados. Asintóticamente, el sistema alcanza un 7-ciclo (Wuensche, 1994).

acceder en función de las condiciones iniciales de partida. Estos estados, así como las transiciones que llevan a ellos, se indican también. Observemos que existen dos tipos de atractores: el estado 000, al que sólo podemos acceder desde él mismo, el atractor 111, al que llegamos partiendo de las condiciones iniciales 100, 110, 100 y 010 (tenemos una cuenca de atracción formada por cuatro estados transitorios más el estado 111) y finalmente un ciclo periódico formado por los estados 001 y 010, entre los que el sistema alterna.

A medida que el tamaño del sistema aumenta, las posibilidades dinámicas crecen, especialmente si el valor de $K$ aumenta. El número total de posibles funciones Booleanas $N_\Lambda(K)$ crece exponencialmente,

$$N_J(K) = (2^2)^K$$

y el número y complejidad de los atractores también puede crecer. Un ejemplo de la cuenca de atracción para un ciclo de periodo 7 de una red más compleja, con $N = 13$ y $K = 3$, se muestra en la figura 11.6 (para un estudio de la generación de estos diagramas, véase Wuensche, 1994).

El enfoque que requiere nuestro estudio extendido a redes de gran tamaño (que serían comparables a los genomas reales) requiere analizar las siguientes propiedades del comportamiento de estas redes (Kauffman, 1993):

- el número de estados que forman un atractor de la red, al que llamaremos longitud del ciclo. Este valor puede ir desde uno hasta $2^N$,
- el número de ciclos alternativos posibles. Al menos uno debe existir, pero pueden ser tantos como los posibles, esto es, $2^N$,
- el tamaño de las cuencas de atracción que desembocan en los atractores,
- la estabilidad de los atractores frente a perturbaciones minimales, que implican un cambio en un solo elemento ($S_i \rightarrow 0$ o viceversa),
los cambios observados en el comportamiento dinámico (lo que se denomina “daño”) causados por una alteración transitoria de una sola variable binaria,

- los cambios en los atractores y las cuencas de atracción debidos a mutaciones (cambios en un dígito) en las funciones Booleanas.

La hipótesis básica planteada por Kauffman es la siguiente: los distintos tipos celulares pueden identificarse como atractores de la dinámica de la red genómica que, en nuestra aproximación, identificamos como una red Booleana. La pregunta que nos formulamos es si, mediante la analogía con las redes de Kauffman, obtendremos, bajo ciertas condiciones, propiedades globales que nos aproximen en alguna forma a las preguntas formuladas al principio acerca del número de tipos celulares y su estabilidad.

11.4 Propiedades dinámicas

En lo que sigue, supondremos que la red considerada es grande y que la configuración inicial \( S_i(0) \) se elige al azar así como el conjunto de \( K \) vecinos de cada elemento y el conjunto de funciones \( A_i \in \mathcal{F}_k \) asociadas a estos. La elección al azar de estas cantidades permite aplicar hipótesis de naturaleza estadística, a la vez que garantiza la ausencia de correlaciones o artefactos espúreos. Si bajo esta elección aleatoria obtenemos resultados bien definidos, que indiquen la aparición de orden global de algún tipo, tal vez podremos comparar estas propiedades con las del genoma real.

A continuación, consideraremos las propiedades obtenidas del estudio de redes de Kauffman con distintas conectividades \( K \). Recordemos que las propiedades son de naturaleza estadística y que por tanto serán tanto más válidas cuanto mayor sea el tamaño de la red \( N \). Algunos resultados se han obtenido en forma numérica y otros a partir de aproximaciones analíticas.

11.4.1 Redes \( K = N \)

En estas redes, cada elemento recibe entradas de todos los elementos del sistema. Existe por lo tanto una sola forma de conectar la red (todos con todos). Cada elemento tendrá asignada una función Booleana \( A_i \in \mathcal{F}_N \). Este sistema presenta el mayor grado de desorden posible y el sucesor de un estado dado bajo la dinámica es aleatorio (uno cualquiera de los \( 2^N \) estados posibles).

El estudio de este sistema (Kauffman, 1993) da una longitud característica de los ciclos

\[
T_c \approx \frac{1}{2} \times 2^N
\]

donde \( T_c \) indica por lo tanto el número de iteraciones necesarias para repetir el mismo estado. El número de atractores del sistema es proporcional a \( N \),

\[
N_a \approx \frac{N}{e}
\]

Ahora debemos considerar la estabilidad de los ciclos, esto es, su estabilidad frente a perturbaciones. Hay dos tipos de perturbaciones que podemos introducir: (a) perturbaciones minimales y (b) perturbaciones estructurales. En el primer caso, modificamos únicamente un elemento (cambiando su estado externamente) y vemos si, como resultado de esta modificación, el sistema se desplaza hacia otro atractor o no. En el primer caso, el sistema será muy sensible a las perturbaciones externas (algo que no parece ocurrir en las células). Este tipo de estabilidad se denomina estabilidad homeostática, que es de hecho muy baja para redes fuertemente conectadas, como es el caso que nos ocupa. En el segundo caso, una función booleana es elegida al azar y una de
sus entradas modificada al azar. Esta perturbación puede ser absorbida por el sistema (que se mantendrá en el atractor previo) o puede dar lugar (como ocurre para $N = K$) a una transición entre atractores distintos. Hablamos en este caso de la propiedad de accesibilidad entre atractores, que es alta en esta red.

Estas redes se denominan caóticas. La longitud de sus ciclos es enorme: para $N = 200$, se tiene $T_c = 10^{30}$. Si tomamos dos condiciones iniciales que difieran únicamente en un bit, y seguimos la divergencia posterior entre ambas, medida como el número de dígitos distintos entre ambas cadenas, veremos que esta separación tiene lugar de forma exponencial. Aunque el carácter discreto de estos sistemas no permite hablar de exponenciales de Lyapunov en sentido estricto, el significado de “caos” es esencialmente el mismo.

Pese al alto grado de desorden de este sistema, el número de atractores $N_a$ es reducido: del orden de $N_a = 74$ para $N = 200$. Para esta red es posible obtener los resultados anteriores en forma analítica. Como hemos indicado anteriormente, los estados consecutivos obtenidos son estadísticamente independientes: tenemos $2^N = M$ posibles estados sucesores, que podemos considerar equiprobables (con probabilidad, por tanto, $1/M$).

Supongamos que partimos de la configuración inicial $c_0$ y observamos los estados sucesivos que presenta, $c_0 \to c_1 \to c_2 \to \ldots \to c_i \to \ldots \to c_{k-1}$, con $c_j \neq c_r$, $\forall 0 \leq j, r \leq k - 1$. La probabilidad de esta secuencia $P(k)$ es

$$P(k) = \left(1 - \frac{1}{M}\right) \left(1 - \frac{2}{M}\right) \ldots \left(1 - \frac{k-1}{M}\right) = \prod_{j=1}^{k-1} \left(1 - \frac{j}{M}\right),$$

que puede aproximarse por (Kauffman, 1993; Weissbuch, 1990)

$$P(k) \approx \exp\left[\frac{k(k-1)}{2M}\right]; \quad \forall k = 1, \ldots, M$$

De aquí tenemos que $p_k = P(k)/M$ es la probabilidad de que una configuración $c_k$ siga a la $c_{k-1}$. En particular, $P_k$ es la probabilidad de que $c_0$ pertenezca a un ciclo de longitud $k$. El número medio de ciclos de periodo $k$ será

$$n(k) = \frac{P(k)}{k}$$

La distribución quedará fuertemente desviada hacia los ciclos de corta duración. De esta expresión pueden deducirse algunas de las propiedades anteriores expuestas, en particular que el número de ciclos tiende a $(N \log 2)/2$.

Por otra parte, dada una configuración inicial $c_0$, podemos preguntarnos cuál será la probabilidad de completar un ciclo en $c_i$, dando una trayectoria de longitud $k - i$. Esta será $P(k)/M$, que es independiente de $i$. La longitud promedio $T_c$ se obtendrá de

$$T_c = \frac{1}{M} \int_0^{M-1} dk \int_0^{k-1} (k - i) \exp\left[\frac{k(k-1)}{M}\right] dk \approx \alpha 2^{N/2}$$

con $\alpha \approx 0.63$.

11.4.2 Redes $K \geq 5$

Las redes NK con $K \geq 5$ tienen una enorme cantidad de posibles conectividades alternativas entre sus $N$ elementos. Los atractores del sistema siguen siendo caóticos e incrementan en longitud de forma exponencial con $N$, más concretamente

$$T_c \approx \frac{1}{2} \times 2^k N$$
(con \( b > 1 \)), y el número de atractores sigue una relación

\[
N_a \approx \frac{1}{2} N \log \left( \frac{1}{2} \pm \alpha \right)
\]

que tiende a \( N/2 \) para \( K \rightarrow \infty \). Nuevamente, la estabilidad homeostática es baja y la accesibilidad entre atractores alta. Aquí, \( \alpha(K) = P(K) - 1/2 \), siendo \( P(K) \) una medida de la homogeneidad interna de las funciones Booleanas, de la que hablaremos más adelante.

### 11.4.3 Redes \( K = 1 \)

Este es el otro caso extremo a considerar: cada elemento de la red está conectado con otro elemento, y sólo uno. La estructura de la red es simple y debido al azar al azar muchos elementos quedan dispuestos formando "colas" que no controlan el comportamiento de la red. Este sistema también permite cálculos analíticos (Weissbuch, 1990). En promedio, unos

\[
N_b \approx N^{1/2} \ln N
\]

elementos de la red forman bucles cerrados y por lo tanto estos elementos son independientes y no pueden propagar su influencia a otras partes del sistema. La red es, esencialmente, un conjunto de subsistemas aislados. Ahora se tiene que

\[
T_c \approx \exp \left[ \frac{1}{8} \log^2(N) \right]
\]

\[
N_a \approx \left( \frac{2}{\sqrt{e}} \right)^{N(1+O(1))}
\]

Nuevamente, la estabilidad homeostática es baja y la accesibilidad entre atractores alta, con lo que de hecho desde el punto de vista de la estabilidad este caso extremo no difiere mucho de los anteriores, caracterizados por una alta conectividad. A este tipo de comportamiento se le denomina **fase congelada**.

### 11.4.4 Redes \( K_c = 2 \) (Orden colectivo espontáneo)

Contrariamente a lo que podría esperarse, el caso \( K = 2 \) no es un simple caso intermedio con propiedades similares a un extremo u otro. Para este valor crítico (que indicaremos por \( K_c \)), las redes de Kauffman experimentan un súbito cambio con profundas implicaciones (Kauffman, 1992; 1993). Podemos resumir las propiedades observadas en cuatro puntos:

- **Número de atractores**: \( N_a \approx \sqrt{N} \).
- **Los atractores son tipicamente estable frente a perturbaciones minimales.**
- **Los cambios en \( A_i \) en forma de alteraciones en un elemento sólo modifican débilmente el comportamiento de la red.**
- **La longitud de los ciclos es \( T_c \approx \sqrt{N} \).**

La enorme diferencia entre esta situación y las anteriores puede resumirse diciendo que, para este punto crítico, el número de atractores es muy reducido (precisamente del orden del número de tipos celulares observados) y que de forma espontánea estos atractores se hacen muy estables para
Figura 11.7: Comparación entre el número de células distintas en diversos organismos y el número de atractores obtenidos para el modelo de Kauffman con $K = 2$ (Kauffman, 1993).

esta conectividad. Para esta red, un sistema con $N = 10000$ elementos posee unos 100 atractores y la longitud de los ciclos es del orden de 100. Debemos añadir que en estos atractores muchos elementos se hallan “congelados” en uno de los dos estados posibles. Esta propiedad será, como veremos, fundamental.

Aunque la aparición de este comportamiento singular resulta extraño a una intuición lineal, ya sabemos que de hecho es natural en la aparición de la complejidad. El punto crítico obtenido, a medio camino entre las redes caóticas y las redes $K = 1$ (congeladas) es un verdadero punto crítico, como veremos a continuación. Una vez más, las propiedades de un sistema real (en este caso, el genoma) que denominaríamos sin dudar “complejo” podrían resultar de la existencia de un punto crítico.

Existen diversas medidas adicionales que pueden emplearse en la caracterización de la transición de fase y de cada tipo de dinámica. Uno de ellas (Weissbuch, 1990) es la magnetización. La magnetización de un elemento $s$, es el promedio temporal del estado del nodo, que se evalúa durante cierto periodo de tiempo $\tau$,

$$m(i) = \frac{1}{\tau} \sum_{t=1}^{\tau} s_i(t)$$

la cual, para periodos cortos, tomará valores fraccionarios ($1/4$, $1/3$, $2/3$, $1/2$,...) y en consecuencia tendremos, según el caso,

- Régimen ordenado: el histograma de magnetización, $P(m)$, nos dará un conjunto de picos bien definido.
- Régimen caótico: el histograma muestra una banda continua, con picos marcados en algunas zonas.

Este comportamiento nos recuerda, de hecho, el que veríamos en el espectro de Fourier de los sistemas no-lineales que experimentaban una transición hacia el caos determinista (capítulo 5) en la que típicamente se pasaba de un espectro dominado por algunas frecuencias a una banda de frecuencias continua.
11.5 Mecánica estadística: método de Derrida

Los resultados anteriores fueron obtenidos muy pronto por Kauffman y otros empleando simulaciones por ordenador, y para \( K = N \) y \( K = 1 \) de forma analítica. Sin embargo, existe un procedimiento para demostrar formalmente que este punto es efectivamente un punto crítico. El método formal fue desarrollado por Bernard Derrida y otros (Derrida y Pomeau, 1986; Derrida y Stauffer, 1986; Derrida y Weisbuch, 1986) y se conoce en la literatura como modelo **annealed** (MA). Esta aproximación se basa en introducir una nueva (y profunda) aleatoriedad adicional sobre la dinámica: en cada iteración, después de calcular el nuevo conjunto de estados, cambiamos todas las funciones Booleanas y las conexiones previamente empleadas por un nuevo conjunto, elegido al azar.

Por extraño que parezca, este modelo guarda una estrecha relación con el modelo planteado al principio, llamado modelo **quenched** (MQ). En el último, como sabemos, el sistema debe repetir su estado inicial en, como máximo, \( 2^N \) iteraciones. Por el contrario, en el caso del MA no esperaremos encontrar periodicidad temporal alguna. Sin embargo, la relación se hace más clara si imaginamos que \( N \) es muy grande, y en ese caso se supone que en mecánica estadística. Tomemos \( K = 2 \) y supongamos que queremos calcular el estado del elemento \( t \)-ésimo, \( S_t \), en el instante \( t + 1 \). Las conexiones están definidas y para llevar a cabo el cálculo debemos inspeccionar el estado de sus \( K = 2 \) vecinos (los que tenía en \( t \)), de forma que para calcular \( S_t(t + 2) \) habremos inspeccionado cuatro elementos y, en general, debemos analizar \( K^n \) valores con el objeto de determinar \( S_t(t + n) \) a partir de \( S_t(t) \). Luego este número crecerá exponencialmente con el tiempo. Este hecho hace plausible que ambos modelos sean estatisticamente equivalentes. A continuación, daremos la demostración de Derrida de que \( K \) es un punto crítico.

Consideremos dos secuencias de bits

\[
C_1(t) \equiv (S_1^{(1)}(t), ..., S_N^{(1)}(t))
\]

\[
C_2(t) \equiv (S_1^{(2)}(t), ..., S_N^{(2)}(t))
\]

que llamaremos **configuraciones** y que se han tomado al azar del conjunto \( C(N) \) de todas las posibles secuencias de tamaño \( N \) (con \( |C(N)| = 2^N \)). La **distancia de Hamming** entre secuencias, \( d_h(t) \) se define por

\[
d_h(t) = \sum_{j=1}^{N} |S_j^{(1)}(t) - S_j^{(2)}(t)|
\]

y mide la cantidad de elementos en estado distinto. El solapamiento entre estas secuencias, \( a_{12}(t) \) se define como el porcentaje de elementos en el mismo estado y es, obviamente, \( a_{12}(t) = (N - d_h(t))/N \). Derrida demostró que, para valores grandes de \( N \), las redes Booleanas del MA permiten obtener los mismos resultados cualitativos que los esperados para redes en el MQ. En particular, ambos modelos dan la misma evolución en \( a(t) \) entre dos configuraciones iniciales.

Explicitamente, el método de Derrida consiste en los siguientes pasos:

1. Tomamos dos configuraciones iniciales,

\[
C_1(t) \equiv (S_1^{(1)}(t), ..., S_N^{(1)}(t))
\]

\[
C_2(t) \equiv (S_1^{(2)}(t), ..., S_N^{(2)}(t))
\]

y, en ambas redes, empleamos el mismo conjunto de funciones Booleanas \( \Lambda_i(t) \in \mathcal{F}_K \), que habremos elegido previamente al azar de entre las que forman el conjunto \( \mathcal{F}_K \). Estas redes tendrán un solapamiento \( a(t) \).
2. Calculamos el nuevo estado de cada red, empleando las ecuaciones dinámicas previamente definidas.

3. Medimos el solapamiento $a(t + 1)$ entre los nuevos estados.

4. Manteniendo los valores de $C_1(t + 1)$ y $C_2(t + 1)$, escogemos al azar un nuevo conjunto de conexiones (elegimos un nuevo conjunto de $K$ vecinos para cada elemento) y un nuevo conjunto de funciones Booleanas, $A_i(t + 1) \in F_K$, y repetimos los pasos anteriores.

Lo que queremos ver es si el solapamiento en el MA tiende a cero (las perturbaciones son absorbidas por el sistema, que estará básicamente “congelado”) o no (crecen, propagándose como en el régimen caótico). La ecuación dinámica para el solapamiento es

$$ a_{12}(t + 1) = F_K(a_{12}(t)) = \frac{1}{2} [1 + (a_{12}(t))^K] $$

obtenida a partir del siguiente argumento. (i) Sea $N_{12}(t)$ el número de elementos en el mismo estado; cierta fracción de elementos de ambas configuraciones $C_i(t)$ tendrán exactamente las mismas entradas. El número esperado de dichos elementos en $t + 1$ será $N(a_{12}(t))^K$. (ii) El número de elementos $S_i(t)$ que tendrán al menos una de sus $K$ entradas distinta es $N(1 - (a_{12}(t))^K$ y puesto que los estados serán asignados al azar (pues las funciones booleanas son escogidas al azar) la mitad coincidirán en $t + 1$, luego el solapamiento evolucionará siguiendo el sistema dinámico discreto

$$ a_{12}(t + 1) = (a_{12}(t))^K = \frac{1}{2} [1 + (a_{12}(t))^K] $$

como queríamos ver.

Podemos transformar esta ecuación en una ecuación para la distancia de Hamming, y obtenemos entonces

$$ D_{t+1} = F_K(D_t) = \frac{1}{2} [1 - (1 - D_t)^K] $$

que nos da una medida de cómo evolucionará en el tiempo la distancia de Hamming de dos configuraciones iniciales $\{C_1(t), C_2(t)\}$. Esta aplicación tiene un punto fijo en $D^* = 0$, y, como sabemos, podemos investigar su estabilidad empleando la derivada de la aplicación en el punto fijo, esto es,

$$ \lambda(K) = \left| \frac{\partial F_K(D)}{\partial D} \right|_{D^*} $$

El modelo de Kauffman introducido al principio (MQ) presenta una fase congelada en la que $D^*$ es estable, mientras que para valores de conectividad altos, es inestable.

El punto crítico se obtendrá, como ya vimos, buscando la condición de estabilidad marginal de la aplicación anterior, que está caracterizada por la condición $\lambda(K) = 1$ que, como podemos comprobar, no es sino

$$ \left| \frac{\partial F_K(D)}{\partial D} \right|_{D^*} = K \frac{K}{2} [1 - D^*]^{K-1} = \frac{K}{2} = 1 $$

o, lo que es lo mismo, $K_c = 2$, como queríamos demostrar. Por encima de este valor, la distancia entre configuraciones crece (fase caótica) y por debajo el punto fijo $D^*$ es estable (fase congelada). Las propiedades de la red en el punto crítico son por lo tanto distintas de las de ambos lados, como cabía esperar, y como ya hemos visto anteriormente. En la figura 11.8 representamos el resultado de iterar la aplicación anterior por debajo y por encima del punto crítico.

En la próxima sección veremos un argumento distinto para la existencia de un punto crítico en la red de Kauffman, en la que consideraremos una situación particular: una red de Kauffman bidimensional cuadrada, con conectividad local similar a la que vimos en los autómatas celulares.
Figura 11.8: Dinámica de la distancia de Hamming obtenida a partir del modelo MA de Derrida (véase texto). (a) Dinámica para \( K = 1 < K_c \), que presenta convergencia hacia \( D^* = 0 \) (fase congelada). (b) Para \( K = 3 > K_c \), la red amplifica pequeñas desviaciones iniciales (fase caótica).

11.6 Percolación: red bidimensional

Consideremos una red de Kauffman cuadrada en la que, a diferencia de las definiciones anteriores, las conexiones entre elementos son estrictamente locales: cada unidad estará conectada sólo con los cuatro vecinos más cercanos (vecindad de Von Neumann), de manera que tendremos \( K = 4 \) sobre una red bidimensional de lado \( L \). Supondremos condiciones de contorno periódicas (Weissbuch y Stauffer, 1987). La dinámica será por lo tanto

\[
S_{ij}(t+1) = \Lambda_{ij} \left[ S_{i-1,j}(t), S_{i,j+1}(t), S_{i+1,j}(t), S_{i-1,j}(t) \right]
\]

siendo \( S_{ij}(t) \) el estado del elemento situado en las coordenadas \((i, j)\) de la red en el instante \( t \).

Al igual que ocurría en la red de Kauffman original, esta red presenta, para la conectividad empleada, comportamiento caótico. Para poder explorar otras posibilidades manteniendo la conectividad, podemos introducir un nuevo parámetro: el sesgo de las funciones booleanas \( \Lambda_{ij} \in \mathcal{F}_4 \) empleadas en este modelo. Si \( p = 0.5 \), indicamos que las salidas que dan las funciones booleanas son, con igual probabilidad, uno o cero. valores de \( p \) cercanos a \( p = 0 \) o a \( p = 1 \) indican que, con gran probabilidad, las funciones Booleanas darán el mismo valor de salida con independencia de las entradas. Es fácil ver que, en estas condiciones, la red tenderá a quedar congelada en forma de un conjunto de unos o de ceros. Así pues, el parámetro \( p \) nos servirá para obtener distintos tipos de comportamientos y posiblemente una transición de fase entre distintos tipos de dinámicas. Este parámetro se comportará obviamente de forma simétrica alrededor del valor \( p = 0.5 \), así que nos limitaremos a analizar su efecto en el intervalo \( p \in [0, 1/2] \).

Derrida y Stauffer (1986) demostraron que, efectivamente, la red bidimensional exhibe una transición de fase para un sesgo crítico \( p_c \approx 0.26 \) que separa una zona de comportamiento ordenado (congelada) cuando \( p < 0.26 \) del caótico para \( p > p_c \). En la figura 11.9 vemos dos ejemplos de simulaciones por ordenador de una red de Kauffman con \( 24 \times 24 \) elementos. El estudio de estas redes muestra que (de forma similar a las redes estándar) no todos los elementos de la red oscilan con la misma periodicidad. De hecho, la mayoría de ellos oscilarán con un periodo menor o estarán completamente congelados: este es el caso de la red (a), obtenida para un estado del
dominio ordenado, con \( p = 0.2 \). En este estado tenemos muchos grupos de elementos con una baja periodicidad, y sólo unos pocos grupos de periodicidad algo mayor. En cambio, para \( p = 0.3 \), en el dominio caótico, vemos en la figura 11.9 (b) que existe un conjunto conectado de elementos que se comporta caóticamente y que este grupo percola a través de la red, encerrando en su interior pequeñas islas congeladas. Los asteriscos indican de hecho que la longitud de los ciclos\(^3\) es superior a \( T_c = 999 \).

El punto crítico detectado en este modelo es interesante porque demuestra hasta qué punto las redes Booleanas aleatorias pueden ser capaces de exhibir orden en situaciones distintas. En este caso, el punto crítico aparece asociado a un fenómeno que ya hemos explorado con anterioridad: la percolación (capítulo 7). Diversos autores han explorado en detalle esta transición, encontrando parámetros de orden que permiten caracterizarla de manera adecuada (Weissbuch y Stauffer, 1987).

### 11.7 Redes de Kauffman generalizadas

Los resultados previos se han obtenido para una situación particular: una red en la que cada elemento está conectado exactamente con \( K \) entradas. Esta es una restricción importante, y deberíamos preguntarnos si una red que presentara una distribución de conectividades \( f(K) \) en lugar de un valor fijo presentaría propiedades similares y, en particular, un punto crítico.

La respuesta es afirmativa (Solé y Luque, 1995). Para demostrarlo, consideremos una red en la que cada elemento \( S_i \) recibe \( K_i \in \{1, 2, ..., K\} \) entradas que, como antes, se han elegido al azar de entre los \( N \) elementos. Ahora, las ecuaciones de evolución de esta red de Kauffman generalizada serán

\[
S_i(t + 1) = A_i \left[ S_{i_1}(t), S_{i_2}(t), ..., S_{i_{K_i}}(t) \right]
\]

\(^3\)En general, puede comprobarse que los grupos de elementos que oscilan en coordinación presentan periodicidades iguales o múltiplos enteros unas de otras. El ciclo del sistema es de hecho el mínimo común múltiplo de estas periodicidades.
Figura 11.10: Espacio de parámetros para la red de Kauffman generalizada. La curva crítica define dos dominios: la fase ordenada (zona sombreada) o “congelada” y la fase caótica, por encima de la curva.

y, nuevamente, empleamos el método de Derrida para detectar la presencia de un punto crítico. Sean dos configuraciones iniciales \( C_i(t) \in \mathcal{C}(N) \) y escojamos para cada elemento \( K_i \) vecinos al azar y las correspondientes funciones Booleanas, ahora pertenecientes al conjunto

\[
S(N, K_m) = \bigcup_{K_i=1}^{K_m} \mathcal{F}_{K_i}(N)
\]

siendo \( K_m \) la conectividad máxima (que tomaremos \( K_m = 10 \) en nuestras simulaciones).

Siguiendo el argumento del MA, obtenemos la siguiente ecuación para la evolución de los solapamientos

\[
a_{12}(t+1) = \frac{1}{2} \left[ 1 + \sum_{K_i=1}^{K_m} f(K_i) a_{12}(t)^{K_i} \right]
\]

o, en función de la distancia \( D_t \),

\[
D_{t+1} = G(D_t) = \frac{1}{2} \left[ 1 - \sum_{K_i=1}^{K_m} f(K_i)(1 - D_t)^{K_i} \right]
\]

Esta ecuación puede ser generalizada aún más si introducimos el sesgo \( p \) que, como sabemos, mide la probabilidad de que una función Booleana \( \Lambda_i \in S(N, K_m) \) dé como salida un valor “1”, dada una configuración de entrada. Empleando este parámetro, obtenemos

\[
D_{t+1} = 4p(1-p)G(D_t)
\]

La condición de estabilidad es ahora

\[
\left| 4p(1-p) \left[ \frac{\partial G(D^*)}{\partial D^*} \right] \right| = 1
\]
lo que nos lleva al resultado

$$2p(1-p) \sum_{K_i=1}^{K_m} f(K_i)K_i = 1$$

Obtenemos de aquí una línea crítica que relaciona la conectividad media $K$ y el sesgo de las funciones, $<K>$

$$<K> = \frac{1}{2p(1-p)}$$

Observamos que esta relación da, para $p = 0.5$, un valor crítico $K_c = 2$, que coincide con el que obtuvimos para redes aleatorias estándar. La condición crítica define dos fases en el espacio de parámetros ($p, <K>$), como vemos en la figura 11.10. Los valores por encima de la curva corresponden a las soluciones inestables que caracterizan la fase caótica. Por debajo, la región sombreada indica la fase congelada. Podemos ver que para conectividades altas, el orden está garantizado si el sesgo es lo bastante pequeño (y la homogeneidad interna de las funciones es, por tanto, alta).

Dado que el genoma de los organismos va incrementando en tamaño a lo largo de la evolución y que estos incrementos obedecen a duplicaciones de genes y otros procesos que tienen lugar al azar, podríamos preguntarnos cuál sería la distribución de conectividades $f(K_i)$ esperada.

La condición crítica obtenida puede servirnos para encontrar dicha distribución siguiendo el procedimiento variacional introducido en el capítulo 1. Recuerdemos que la idea de este método consiste en buscar la distribución de probabilidad que maximiza la entropía de Boltzmann bajo cierto conjunto de ligaduras (Haken, 1988). El procedimiento será emplear la entropía de Boltzmann para la distribución de conectividades,

$$S(\{f(K_i)\}) = -\frac{1}{K} \sum_{K_i=1}^{K} f(K_i) \log[f(K_i)]$$

y el problema, como sabemos, consiste en encontrar las ligaduras adecuadas para resolver la ecuación variacional mediante el método de los multiplicadores de Lagrange.

La primera ligadura es la normalización de probabilidades

$$\sum_{K_i=1}^{K_m} f(K_i) = 1$$

y para la segunda, emplearemos la condición crítica (Solé y Luque, 1995)

$$\sum_{K_i=1}^{K_m} f(K_i)K_i = \frac{1}{2p(1-p)}$$

que podemos justificar siguiendo la argumentación de Kauffman (Kauffman, 1993): el genoma está autoorganizado cerca del punto de transición, en el que el sistema está adecuadamente conectado para propagar efectivamente las señales y cambios a la vez que adquiere la mayor estabilidad (tanto homeostática como en términos de accesibilidad de atractores). Podemos llamar a esta ligadura la "ligadura de antiacos". La estabilidad genómica hará que los cambios que se produzcan al azar con resultado de un aumento de la conectividad y el tamaño deban obedecer esta restricción macroscópica.

Empleando estas ligaduras, resolveremos la ecuación variacional

$$\delta \left[ S(\{f(K_i)\}) - (\lambda - 1) \sum_{K_i=1}^{K} f(K_i) - \beta \sum_{K_i=1}^{K} f(K_i)K_i \right] = 0$$
Figura 11.11: Dinámica de un red de Kauffman generalizada con $N = 130$ correspondiente a tres redes que ocupan los tres puntos del espacio de parámetros indicado en la figura anterior. Aquí tenemos: $p = 2$ y (a) $< K >= 6$, (b) $< K >= 3.1$ y (c) $< K >= 2$.

que, como sabemos, nos da la ecuación correspondiente a la distribución de Boltzmann (vista en el capítulo 1)

$$f(K_i) = \exp[-\lambda - \beta K_i]$$

Aquí $\lambda$ y $\beta$ son los multiplicadores de Lagrange los cuales, una vez empleadas las ligaduras de partida, nos dan la distribución

$$f(K_i) = \frac{1}{Z} \exp[-K_i/ < K >]$$

o, lo que es lo mismo,

$$f(K_i) = \frac{1}{Z} \exp[-2p(1-p)K_i]$$

donde $Z$ es la función de partición

$$Z = \sum_{K_i=1}^{1K} \exp[-2p(1-p)K_i]$$

La distribución obtenida proporciona la distribución esperada de conectividades en el genoma. La mayoría de los elementos recibirá un pequeño número de entradas, mientras que sólo unos pocos genes estarán regulados por un número grande de otros genes. Así parece ocurrir de hecho en los genomas reales.

Podemos simular numéricamente una red generalizada. En la figura 11.11 se muestran tres ejemplos, que corresponden a tres puntos distintos del espacio de parámetros, representado en la figura 11.10. Tenemos por lo tanto tres redes con distribución exponencial de conectividades y parámetros correspondientes a las zonas (a) caótica ($< K >= 6$, $p = 0.2$), (b) crítica ($< K >= 3.1$, $p = 0.2$) y (c) congelada ($< K >= 2$, $p = 0.2$).
Bibliografía


Capítulo 12

Evolución, Criticalidad y Extinciones

“Nada tiene sentido en biología si no es a la luz de la evolución”. Esta afirmación de T. Dobzhansky nos remite al hecho evidente de que la estructura y función de células, organismos e incluso de ecosistemas enteros no es el fruto del azar, sino (al menos en parte) de un proceso. A este proceso lo denominaremos evolución, y constituye la espina dorsal de nuestra comprensión de la vida.

El nombre de Charles Darwin es indisoluble del de la teoría de la evolución. Darwin formuló de hecho una de las primeras teorías de la complejidad. La materia prima de la teoría era nada menos que la materia viva (la más compleja que conocemos) y la explicación de la emergencia de esta complejidad, simple y elegante. La pregunta de fondo (no contestada entonces) era la que Darwin llamaba “ese misterio de misterios: el origen de las especies”. ¿De dónde proceden las especies? Y lo que es también importante, ¿por qué y cómo se extinguen?

Tal vez las especies que Darwin y muchos otros habían observado en forma fósil desaparecieron simplemente porque no cabían en el arca de Noé. Pero semejante explicación (u otras suministradas por el creacionismo científico) no puede satisfacer a ningún observador escéptico y menos aún a un espíritu tan inquieto como el del joven Darwin. Los fósiles, que podemos fechar mediante distintas técnicas de datación, aparecen ordenados en patrones fácilmente intuibles. Las formas celulares simples preceden en el tiempo a las formas pluricelulares, que en el periodo cámbrico experimentaron una explosión de formas asombrosa. Darwin era consciente de unos hechos de gran importancia que sugerían una explicación coherente, una solución a la pregunta. La primera piedra la ponía la existencia de una variabilidad intrínseca, potencial, entre los individuos que forman parte de una población. Esta variabilidad puede verse en el proceso de selección artificial (que el propio Darwin ensayó con palomas) en el que, seleccionando individuos que poseyeran alguna propiedad de interés más acentuada (ya sea color, tamaño o forma) podemos obtener, a lo largo de generaciones, nuevos individuos que llegan a distanciarse notablemente de sus progenitores. Un buen ejemplo lo dan las numerosas razas de perros que conocemos, y que fueron obtenidas a lo largo de muchos años por selección artificial a partir de unas formas iniciales comunes. Hoy sabemos que la variabilidad tiene un origen físico (el ADN es una molécula sometida a cambios potencialmente transmisibles) aunque en la época de Darwin los trabajos de Mendel pasaron inadvertidos.

Darwin supo asociar a esta variabilidad (que debemos mirar como fenómeno poblacional) las restricciones impuestas por el medio. En su ensayo sobre los límites de la población, Malthus había trazado ya una parte del camino para la teoría de Darwin: en una población en la que los

\[^{1}\text{Aunque incompleta: el propio Darwin reconocía en su texto básico }\text{El Origen de las Especies}\text{ la ausencia de pruebas para ciertas afirmaciones clave.}\]
Figura 12.1: El nombre de Charles Darwin es indisoluble del de la evolución biológica. Su intuición extraordinaria le permitió llevar a cabo una síntesis genial de un cúmulo de información naturalista dispersa y de enorme diversidad. Su teoría de la evolución por selección natural transtornó la visión clásica de los orígenes de la complejidad biológica. Más allá de su impacto sobre la ciencia, se convirtió en un fenómeno social, que le deparó no pocas burlas, como la que muestra el grabado, publicado en el semanario satírico inglés “Punch”. Curiosamente, este diagrama que remeda el proceso evolutivo, empieza... ¡por el caos!
recursos son limitados, no todos los individuos potencialmente posibles pueden sobrevivir. Si existe variabilidad, podremos encontrar organismos que crecen más deprisa o que cuidan mejor de sus descendientes, organismos que toleran mejor ciertas temperaturas, etc. Los que mejor responden al ambiente externo, sobreviven. Esta es la idea esencial detrás de la teoría de la selección natural.

A lo largo de millones de años, este proceso, lento y progresivo, daría lugar a poblaciones de organismos de gran variedad, adaptados a los distintos ambientes de la Tierra. Si la variedad potencial es lo bastante grande, virtualmente cualquier estructura podría ser generada. El proceso de creación de complejidad es fundamentalmente histórico, aunque interviene un elemento de azar de gran importancia como es la mutación. El azar es entonces "canalizado" a través de la selección natural.

La teoría de Darwin, como todas las teorías geniales, era simple. Su impacto sobre el pensamiento del siglo XIX fue mucho más allá de lo que él mismo esperaba. Trascendió el ámbito científico y fue muy a menudo totalmente malinterpretada (para desgracia del propio Darwin). El origen de la complejidad aparece, en cualquier caso, como resultado de causas simples y comprensibles. La teoría de Darwin fue considerablemente elaborada a lo largo del siguiente siglo. El enorme desarrollo de la genética molecular y el estudio estadístico de las poblaciones dio pie a la construcción de un marco teórico de gran sofisticación: la genética de poblaciones. Esta teoría, de gran sofisticación matemática, junto con una enorme masa de datos paleontológicos y biogeográficos, permitió una nueva y poderosa síntesis, bautizada por sus autores como Neodarwinismo.

Entre las propiedades fundamentales del proceso evolutivo, la importancia de la historia es crucial. En particular, debemos señalar que la evolución es un largo proceso en el que las innovaciones han tomado forma partiendo de los elementos que la historia pasada habría dejado de si. En este sentido, la evolución no es un proceso simple de invención, sino una hábil chapuza. Tiene más que ver con el "bricolaje" que con la ingeniería. Por ejemplo: en nuestra sociedad industrializada, los inventos (consideremos los nuevos medios de locomoción, figura 12.2) pueden representar completas revoluciones de diseño respecto de anteriores aportaciones. El ingeniero no necesita aprovechar lo anterior. Puede innovar partiendo de elementos completamente distintos.

### 12.1 Extinciones y macroevolución

La evolución a gran escala, tal y como se nos muestra a partir del análisis del registro fósil, presenta algunas propiedades generales nada triviales. En primer lugar se han producido, a lo largo del curso de la vida sobre la Tierra, varios acontecimientos de extinción en masa que se han asociado de forma natural a causas externas (en nuestra discusión, no-biólogicas) como el impacto de un asteroide, una inusual actividad volcánica, invasiones del campo magnético o explosiones de supernovas. El caso de la frontera entre el Cretácico y el Terciario, que marcó el fin de la era de los dinosaurios, es tal vez la mejor conocida. En otros casos, la dinámica tectónica, que modificó en distintas ocasiones los niveles del mar, se ha considerado como una causa de extinción típica para la mayoría de los grupos de organismos marinos.

Pero no debemos olvidar que, pese a que estos acontecimientos podrían haber eliminado hasta el 96 por ciento de las especies en un período de tiempo corto (Raup, 1986), estas extinciones generalizadas sólo representan un 5 por ciento del total. Nos queda por lo tanto un 95 por cien, lo que se denomina genéricamente extinción de fondo, al que debemos dar una explicación. La primera observación sorprendente es que, pese a que esperaríamos un comportamiento muy distinto de ambos tipos de extinciones (dado que obedecían a causas en principio distintas), el estudio de la distribución de frecuencias de los acontecimientos de extinción nos muestra una curva continua que va, sin cambios aparentes, de las pequeñas extinciones a las extinciones en masa (figura 12.3 (a)) (Raup, 1986).

Este resultado contradice nuestra intuición: esperaríamos encontrar una distribución con dos
Figura 12.2: Innovación tecnológica: a diferencia de la evolución biológica, un nuevo invento técnico no precisa emplear los elementos antiguos para obtener una nueva solución. En el caso de la evolución biológica, las nuevas innovaciones se hallan acotadas por las restricciones adquiridas en el pasado, sobre las que hay que construir.

Figura 12.3: (a) Distribución de frecuencias de extinción para los 79 estados geológicos del Paleozoico, basados en los datos procedentes de 2316 familias de animales marinos (Raup, 1986). (b) Distribución de las duraciones de géneros fósiles (basado en Raup, 1993). En ambas figuras se muestra la interpolación exponencial y la potencial.
máximos, en lugar de una distribución continua. Otras medidas, como por ejemplo la duración de distintos géneros (en millones de años) muestra también un decamieno continuo (figura 12.3 (b)). Ambas figuras son de importancia en nuestra discusión: ¿qué tipo de distribución siguen? Es obvio que ambas muestran un fuerte decamieno, característico de muchas series de datos en biología. Si por ejemplo consideramos el problema de la ruina del jugador, descrito en el capítulo 1, y, partiendo de una condición inicial dada calculamos cuánto tiempo tardará en arruinarse y promediamos sobre miles de juegos, obtendremos una distribución decreciente.

En ambas figuras hemos interpolado las dos funciones decrecientes más típicas: una exponecional y una potencial. Para el número de extinciones de tamaño $s$, tenemos $N(S) \approx S^{-1.95}$ para el ajuste potencial y $N(S) \approx \exp(-0.028 S)$ para el ajuste exponecional. Para los tiempos de duración, $L$, tenemos $N(L) \approx L^{-1.84}$ y $N(L) \approx \exp(-0.043 L)$, respectivamente.

Ambas series se ajustan relativamente bien a ambos tipos de curvas. Sin embargo, como ya sabemos, la interpretación de ambas es muy distinta en términos de complejidad. Si una distribució sigue una ley potencial, entonces podemos conjeturar que (al menos hasta cierto punto) los mismos mecanismos operan a distintas escalas. Este hecho proviene, como ya vimos, de la fractalidad asociada a las leyes potenciales (capítulos 7 y 8). Si disponemos de una dependencia potencial para cierta cantidad $x$, esto es, si $N(x) = Cx^{-\alpha}$ ($C$ es una constante) entonces, si consideramos una escala dada, digamos $x$, y llevamos a cabo un cambio de escala $x' = \gamma x$, con $\gamma > 0$, tenemos que $N(x') = C(x')^{-\alpha}$. Pero, reordenando términos, vemos que esto no es más que la misma dependencia funcional, $N(x') = C'x'^{-\alpha}$ (con $C' = C\gamma^{-\alpha}$), y por lo tanto podemos en principio extrapolar de una escala a otra. Una situación completamente distinta aparece cuando consideramos una ley exponecional (característica, por otra parte, de los sistemas en equilibrio): la invariancia de escala no se da, y existen escalas de longitud características.

Como ya vimos en el capítulo 7, la primera situación se da en sistemas físicos situados en las proximidades de puntos críticos. La interacción entre elementos vecinos crea correlaciones que divergen y no existen escalas privilegiadas (Solé et al., 1996). En el segundo caso, las interacciones tienen poca relevancia y no poseen la capacidad de generar estructuras de escalas superiores a la que impone la distancia de interacción.

Consideremos a continuación otra observación llevada a cabo sobre los datos procedentes del registro fósil, y que no es menos sorprendente. Un análisis estadístico de grupos muy distintos revela que la probabilidad de extinción que puede obtenerse a partir de las curvas de supervivencia (como la que se muestra en la figura 12.4) es constante.

¿Cómo interpretar dicha constancia? Uno de los análisis que más influencia ha tenido sobre el estudio de la macroevolución fue llevado a cabo por Leigh Van Valen, quien analizó los datos antes mencionados. La hipótesis lanzada por este autor (Van Valen, 1973) es conocida como hipótesis de la Reina Roja, en honor al personaje que aparece en Alicia a través del espejo, de Lewis Carroll. Las gráficas obtenidas nos muestran, en contradicción con lo que esperaríamos, que la probabilidad de extinción dentro de cada grupo permanece básicamente constante a través del tiempo. Una especie puede desaparecer en cualquier momento (geológicamente hablando) con independencia de cuánto tiempo haya existido. Esperaríamos en realidad que aquellas especies que han vivido mucho tiempo fueran, en promedio, más duraderas. Este razonamiento se basa en la visión adaptacionista: los que han subsistido durante más tiempo lo han hecho porque estaban mejor adaptados, luego desaparecerán con menor probabilidad. Estas expectativas chocan frontalmente con las evidencias obtenidas. La hipótesis de Van Valen, que analizaremos en la siguiente sección, presenta una explicación teórica enormemente interesante y dinámica.

Debemos añadir, antes de continuar, un problema adicional de gran importancia: el ritmo al que tiene lugar la evolución. La visión Darwinista clásica (en la que el propio Darwin no creía del todo) nos expone el proceso evolutivo como un fenómeno lento y gradual, en el que los cambios se acumulan en el tiempo. Existe sin embargo una interpretación alternativa, basada en datos
observacionales, planteada por los paleontólogos Stephen Jay Gould y Niles Eldredge (véase Gould y Eldredge, 1993, y la bibliografía allí citada). Estos científicos han introducido la hipótesis de que los cambios en evolución tienen lugar de manera súbita. Veremos, en esta imagen, cambios nulos o muy pequeños que cubren grandes periodos de tiempo, interrumpidos por ráfagas de cambio en las que, durante un periodo de tiempo corto, se producen grandes modificaciones. Para esta visión de la evolución, la escala microevolutiva no puede extrapolarse a la macroevolutiva. En otras palabras: los procesos que tienen lugar al nivel de la genética de poblaciones individuales no generan los fenómenos observados a la escala macroevolutiva o, más exactamente, los segundos no son reductibles a los primeros.

12.2 La hipótesis de la Reina Roja

Las extinciones tienen lugar, entre grandes acontecimientos de extinción, de forma regular y constante. Como hemos visto en la sección anterior, un estudio de los datos disponibles revela una caída constante en cualquier grupo considerado y a cualquier nivel taxonómico. Van Valen postuló en 1973 una nueva “ley evolutiva” que introduzca una imagen dinámica de los ecosistemas y que, en este sentido, proporcionara una conexión entre ecología y evolución. Aunque semejante conexión parece evidente, raramente se tiene en cuenta en los estudios basados en genética de poblaciones. Sólo en casos contados se considera el efecto de la interacción entre especies. Típicamente, se aísla (en la teoría) una especie, se elige un carácter fácilmente analizable (aunque en ocasiones difícilmente interpretable en términos adaptativos) y se estudia la abundancia de los genes implicados.

Pero las especies no están únicamente en contacto con un medio natural formado por el clima y las particularidades geográficas; cada especie interactúa con cierto número de especies que forman el ecosistema en el que está inmersa. No es posible separar los efectos de las distintas especies ni reducir las interacciones a especies aisladas. Las propiedades del ecosistema son propiedades emergentes, resultantes de un proceso de autoorganización. Van Valen consideró esta situación como punto de partida: supongamos que cada especie está afectada por algunas de las que comp-
ponen el ecosistema. Para simplificar esta idea, tomemos un caso muy simple. Supongamos que consideramos una presa y un depredador y que nos limitamos a considerar sólo dos propiedades de cada especie. Para el depredador, podrían ser su velocidad y su tamaño. Para la presa, su velocidad y su habilidad para camuflarse. Supongamos que todas estas propiedades pueden medirse. Tendremos poblaciones de ambos, y estas poblaciones pueden, en principio, coevolucionar.

Imaginemos que consideramos, en un momento del tiempo, las posibles combinaciones de estas propiedades. Para un depredador, un tamaño superior al de su presa será una ventaja, aunque dentro de unos límites. La velocidad, por otra parte, no es una propiedad independiente del tamaño. Ambas dependen a su vez de las propiedades de la presa. Podemos imaginar que, para una población de presas dada (con cierto tamaño medio y cierta habilidad media de camuflarse) existe un conjunto potencial de posibilidades para el depredador algunas de las cuales son mejores que otras en el sentido de mejorar su adaptación. Si estas combinaciones y su grado de bondad fueran medibles, podríamos representar cuantitativamente el grado de "adaptación" (fitness) de la especie mediante un paisaje adaptativo como el que se indica en la figura 12.5. En los ejes indicamos por P1 y P2 las dos propiedades consideradas. La altura del relieve nos da una medida de la adecuación, o adaptación, de la especie a las condiciones dadas. Estos paisajes adaptativos son enormemente útiles como imágenes de la evolución, aunque estrictamente serán en general hiper superficies en un espacio de gran dimensión. Los cambios de la especie bajo las presiones del ambiente (físico y biótico) harán que la especie, colocada en un momento dado en alguna o algunas posiciones de esta superficie, se mueva hacia el pico más cercano. Estos máximos locales se conocen como picos adaptativos. Si la especie fuera capaz de variar con enorme rapidez, podría hallarse distribuida por toda la superficie y acceder rápidamente a los óptimos. Pero esta situación es claramente poco razonable biológicamente (excepto tal vez para los virus) y por lo tanto lo que tendremos será un punto sobre la superficie moviéndose (si es posible) hacia el pico adaptativo más cercano. Ahora, volvamos nuestra mirada hacia la presa. El depredador se ha movido hacia una situación en la que su presión sobre la presa ha aumentado, de forma que la presa (que
Figura 12.6: La Reina Roja y Alicia: al igual que estos personajes, las especies de un ecosistema deben cambiar con rapidez sólo para permanecer.

a su vez posecerá un paisaje adaptativo propio) deberá moverse simultáneamente, escapando del depredador. Y aquí es donde las cosas se complican: si la presa se mueve, no sólo se desplaza sobre el paisaje adaptativo, persiguiendo su máximo local, sino que a la vez estará modificando el paisaje adaptativo del depredador. Estos cambios sucederán en ambas direcciones, y el resultado será un paisaje siempre cambiante que obliga a ambos contendientes a cambiar. Si tenemos en cuenta que en realidad no será un par de especies, sino tal vez decenas, las que interactúan entre sí en formas sutiles y complejas, puede darse con facilidad una situación relevante en nuestra discusión: una especie puede ser incapaz de alcanzar las proximidades de su pico adaptativo más cercano y en consecuencia extinguirse. Esta es básicamente la propuesta de Van Valen. El entorno biótico es lo más importante: cada especie debe responder con rapidez suficiente a los cambios de las restantes para permanecer. Y aquí está lo más original de la teoría: en general, el cambio de las propiedades de cada especie no tiene lugar con el objeto de mejorar la adaptación, sino única y exclusivamente de permitir que la especie siga jugando su partida en el ecosistema. Si es incapaz de alcanzar un pico local, desaparecerá. Así pues, no se cambia para tratar de mejorar (en algún sentido) sino de permanecer. De aquí el nombre de la teoría de Van Valen: la Reina Roja (figura 12.6) en honor al personaje que aparece en el cuento de Lewis Carroll Alicia a través del espejo.

En una escena del libro, la Reina Roja toma a Alicia de la mano y emprenden una alocada carrera. Sin embargo, por mucho que corran, no hacen sino permanecer en el mismo lugar. La reina roja le dice a Alicia:

"Ahora, aquí, como ves, es preciso correr cuanto puedas para permanecer en el mismo lugar."

Una conclusión cualitativa de la hipótesis es que, muy posiblemente, esta dinámica dará como resultado una tasa de extinción constante. Los cambios en los paisajes adaptativos son muy complejos, así que una especie no puede prever ni la naturaleza y ni la magnitud de los picos de
su propio paisaje, que se arrugará o alisará de forma más o menos compleja. Esta incapacidad de previsión (por lo demás inevitable) hace que cualquiera de los "jugadores" pueda abandonar el juego con igual probabilidad.

La hipótesis de la reina roja ha sido formalizada por Maynard Smith y Stenseth (Maynard Smith y Stenseth, 1984; Stenseth, 1985) en un brillante estudio teórico de la hipótesis de Van Valen. Para obtener una expresión formal de las condiciones bajo las cuales tendrá lugar la "carrera de armamentos" entre los distintos contendientes, se definen ciertas cantidades básicas y se supone que el número de jugadores es constante.

Supongamos por lo tanto un ecosistema complejo formado por $N$ especies. Sea $W_i^*$ la mayor eficacia adaptativa (o simplemente adaptación) de la especie $i$-ésima en el ambiente dado, suponiendo que todas las especies tuviesen una composición genética óptima. Sea $W_i^m$ la adaptación media actual de la especie $i$-ésima. Definiremos a continuación la siguiente cantidad:

$$ L_i = \frac{W_i^* - W_i^m}{W_i^m}, \quad i = 1, 2, ..., N $$

la cual mide la distancia de la especie $i$-ésima al pico adaptativo óptimo. Sea $\beta_{ij}$ el cambio en $L_i$ que se producirá si se da un cambio en $L_j$. Entonces, si el promedio de las distancias es

$$ < L > = \frac{1}{N} \sum_i L_i $$

puede demostrarse fácilmente que:

$$ \frac{d < L >}{dt} = \frac{1}{N} \left[ \sum_j \left( L_i \sum_i \beta_{ij} - \sum_j L_j \right) \right] $$

y esta ecuación tendrá un punto fijo estable si

$$ \Gamma = \sum_i \beta_{ij} = 1 \quad \forall j $$

En caso contrario (Stenseth y Maynard Smith, 1984) puede demostrarse que $< L >$ disminuirá (evolución convergente) o bien aumentará (evolución divergente), en función de si $\Gamma < 1$ o $\Gamma > 1$.

Stenseth y Maynard Smith exploraron esta idea en detalle y llegaron a formular un conjunto de dos ecuaciones dinámicas que describen, en la escala de tiempo geológico, las variaciones del promedio de distancias así como el número de especies presente:

$$ \frac{d < L >}{dt} = \left[ a + b < L > + cN \right] < L > $$

$$ \frac{d N}{dt} = h + (d - e) < L > + (f - g)N $$

donde se asume que $b < 0$, $h > 0$ y $(f - g) < 0$. Analizando las distintas posibles combinaciones de parámetros, puede demostrarse que existen varios modos de evolución posibles:

- **Reina Roja estable**: comportamiento estable, caracterizado por una tasa uniforme de evolución, especiación y extinción, y un número constante de especies.

- **Estado estacionario**: el cambio evolutivo se frena progresivamente hasta que el sistema contiene el número máximo de especies posible.

- **Extinción**: todas las especies son eliminadas.
Figura 12.7: (a) Fluctuaciones en el número de familias del grupo Ammonoidea, representadas aquí desde su aparición hasta su extinción, cubriendo un periodo de 320 millones de años (basado en House, 1989). (b) Espectro de Fourier asociado, de tipo $1/f$.

Un cuarto caso implica divergencias no relevantes biológicamente. Dado que los ecosistemas existen, es de suponer que sólo dos de estas alternativas sean posibles: la Reina Roja o el estado estacionario. Para llegar algo más lejos, y en particular para explorar las propiedades antes expuestas acerca del registro fósil, debemos atacar el problema en otro frente.

### 12.3 Criticalidad, fractales y evolución

Entre los datos expuestos en la introducción, veíamos que existía evidencia (aunque no concluyente) de leyes potenciales, sugerentes de un posible estado crítico. Existen otras evidencias que dan soporte a esta posibilidad. Una de ellas (Solé y Bascompte, 1995) ha sido obtenida a partir del análisis de la dinámica de las fluctuaciones en el número de familias del grupo de los ammonites (Ammonoidea, figura 12.7). Este grupo de organismos, que dominó los mares de la Tierra (junto a los trióbitos) durante un largo periodo de 320 millones de años, desapareció completamente con la gran extinción de final del Cretácico (hace 65 millones de años).

Estos datos, procedentes de un estudio exhaustivo (House, 1989) han sido reanalizados (Solé y Bascompte, 1995) dando como resultado una evidencia adicional de criticalidad: el espectro de Fourier obtenido a partir de esta serie es muy cercano a $1/f$. Este resultado está notablemente alejado del que esperaríamos, en principio, para un "random walk" (que da $1/f^2$) y sugiere que un mismo mecanismo, no-lineal, podría estar operando a distintas escalas. De una forma parecida a lo que ocurre con el montón de arena del modelo de Bak, Tang y Wiesenfeld (capítulo 8) los cambios en las especies podrían generar extinciones a todas las escalas imaginables y, lo que es más interesante, establecería una conexión natural entre los acontecimientos grandes y pequeños. El origen de todas las extinciones sería el mismo, sólo que en la mayoría de los casos (como en las avalanchas de un montón de arena) serían pequeñas, y sólo en algunos casos implicarían un gran número de especies. Aunque un análisis pormenorizado de posibles alternativas no-críticas a este resultado, esto es, basadas en fenómenos puramente aleatorios, ha demostrado que un espectro $1/f$ puede eventualmente obtenerse a partir de series estocásticas (Solé y Bascompte, 1995) existen
Figura 12.8: (a) Distribución de frecuencias de aparición de nuevas especies, correspondientes a los datos de la figura anterior. (b) Distribución de extinciones asociada.

algunas evidencias adicionales que parecen dar credibilidad a la hipótesis del estado crítico. En particular, podemos analizar la distribución de frecuencias de aparición de nuevas familias (dentro de periodos de dos millones de años) que representamos en la figura 12.8 junto con las frecuencias de extinciones.

Encontramos distribuciones ajustables a una ley potencial \(N(S) \approx S^{-\eta}\). Este ajuste es claramente mejor que el ajuste exponencial. Sin embargo, aparece una diferencia interesante: el exponente para las frecuencias de extinción es ahora \(\eta \approx 1.35\), bastante alejado del obtenido por Raup para el conjunto total de grupos (véase más arriba), mientras que el de nuevas familias es \(\eta \approx 1.85\).

Esta información es enormemente sugerente, y existe una segunda fuente de datos consistente con la hipótesis de criticalidad, procedente del análisis detallado de la organización de los grupos taxonómicos. Estos grupos definen distintas formas de agrupar organismos: en especies, que se agrupan en géneros, que forman familias, órdenes y clases. Cada nivel contiene muchos grupos inferiores, y podríamos preguntarnos si las relaciones entre las frecuencias de, por ejemplo, especies dentro de géneros, o géneros dentro de familias, siguen algún tipo de distribución especial. En este sentido, la filogenia (esto es, el proceso que da lugar a las distintas estructuras taxonómicas) podría generar propiedades autosimilares compatibles con un estado crítico.

Existen evidencias cualitativas de que los árboles filogenéticos poseen propiedades fractales (Green, 1991). La cuantificación apareció a raíz de los estudios de Bruno Burlando acerca de la estructura de las relaciones entre taxones (Burlando, 1990; 1993). En este estudio, en el que se analizó un enorme conjunto de datos que contempla grupos taxonómicos muy diversos, incluyendo tanto grupos vivos como fósiles de periodos distintos, se demostró que dichas relaciones exhiben típicamente leyes potenciales muy bien definidas. La autosimilaridad en la estructura sería la contrapartida geométrica a la autosimilaridad temporal. Esta evidencia, sumada a la anterior, nos lleva a conjeturar, una vez más, que la complejidad de la evolución lleva a un estado crítico autoorganizado. Ahora, los modelos deben dar el siguiente paso.
12.4 Modelo de Kauffman

Un primer modelo en esta dirección fue propuesto por Kauffman (Kauffman y Johnson, 1990) y se basa en un modelo de redes aleatorias acopladas, con una filosofía parecida a la que se utilizó en el capítulo anterior. El punto de partida es el denominado paisaje adaptativo \( NK \), en el que definimos las especies como un conjunto de "genes" o propiedades que pueden tomar, como en las redes Booleanas, un valor binario \( S_i \in \Sigma = \{0, 1\} \). Ahora, en lugar de definir funciones Booleanas, asignamos a cada propiedad un número al azar, que definirá su adaptación. Más concretamente, supongamos que las "especies" de un ecosistema están definidas mediante conjuntos de \( N \) propiedades, esto es, \( \{S_1, S_2, ..., S_N\} \) con lo que asumimos que existen \( 2^N \) posibles combinaciones. Supondremos además que cada \( S_i \) depende exactamente de \( K \) propiedades adicionales. Dado que en principio la adaptación asociada a cada característica dependerá, como señalábamos con anterioridad, de las demás, podemos definir una tabla de valores \( F_i^{(K)} \) asociados a cada propiedad en presencia de las otras. Esta asignación se lleva a cabo, en el modelo de Kauffman, de forma aleatoria. Una vez definidos estos valores, podemos calcular la adaptación promedio, i.e.

\[
< F > = \frac{1}{N} \sum_{i=1}^{K} F_i^{(K)}(S_i; S_{i1}, ..., S_{iK})
\]

para cada configuración \( \{S_1, S_2, ..., S_N\} \) posible. Un ejemplo lo da la siguiente tabla, para un sistema con \( N = 3 \) propiedades y \( K = 2 \) conexiones:

<table>
<thead>
<tr>
<th>( (S_1, S_2, S_3) )</th>
<th>( F_1^{(2)} )</th>
<th>( F_2^{(2)} )</th>
<th>( F_3^{(2)} )</th>
<th>( &lt; F &gt; )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 0 0</td>
<td>0.6</td>
<td>0.3</td>
<td>0.5</td>
<td>0.47</td>
</tr>
<tr>
<td>0 0 1</td>
<td>0.1</td>
<td>0.5</td>
<td>0.9</td>
<td>0.50</td>
</tr>
<tr>
<td>0 1 0</td>
<td>0.4</td>
<td>0.8</td>
<td>0.1</td>
<td>0.43</td>
</tr>
<tr>
<td>0 1 1</td>
<td>0.3</td>
<td>0.5</td>
<td>0.8</td>
<td>0.53</td>
</tr>
<tr>
<td>1 0 0</td>
<td>0.9</td>
<td>0.9</td>
<td>0.7</td>
<td>0.63</td>
</tr>
<tr>
<td>1 0 1</td>
<td>0.7</td>
<td>0.2</td>
<td>0.3</td>
<td>0.40</td>
</tr>
<tr>
<td>1 1 0</td>
<td>0.6</td>
<td>0.7</td>
<td>0.6</td>
<td>0.63</td>
</tr>
<tr>
<td>1 1 1</td>
<td>0.7</td>
<td>0.9</td>
<td>0.5</td>
<td>0.70</td>
</tr>
</tbody>
</table>

Tabla 1: asignación de adaptaciones para un sistema con \( N = 3 \) genes (o propiedades dadas) y \( K = 2 \) interacciones por gen.

Supongamos que los cambios en esta especie tienen lugar en la dirección de incrementar la adaptación. Entonces, si representamos las posibles combinaciones de propiedades como los vértices de un cubo tridimensional, las posibles transiciones que implican el cambio en una propiedad (de \( S_i = 0 \) a \( S_i = 1 \) o de \( S_i = 1 \) a \( S_i = 0 \)) nos llevan, para el ejemplo anterior, a dos posibles estados finales que representan, en esta aproximación, los máximos locales del paisaje adaptativo (figura 12.9). Cada transición hacia un vértice vecino recibe el nombre de salto adaptativo. Para llevarla a cabo, la especie elige un vértice adyacente al azar y si la nueva adaptación promedio es mayor que la anterior, el salto tiene lugar. En caso contrario, se mantiene en su posición. El valor de \( K \) define la "rugosidad" del paisaje adaptativo, en la medida en que hace crecer en número (y variabilidad) a los picos adaptativos.

Si la especie está aislada, el resultado de la regla "aumentar \( < F > \) cambiando un estado" es alcanzar alguno de los picos adaptativos y permanecer en él de forma estable. El cambio, razonablemente, debe ser lento si suponemos que las propiedades no pueden cambiar (al menos con facilidad) de forma simultánea, sino de una en una.
Figura 12.9: Paisaje adaptativo correspondiente a la tabla I. Los vértices aparecen etiquetados por el estado del sistema (el conjunto de propiedades) y su adaptación. Las flechas indican la dirección de los movimientos hacia los máximos locales.

Pero imaginemos ahora que consideramos un conjunto de $S$ especies, cada una de ellas definida por una red $NK$ como la anterior, y que las conectamos entre sí por medio de un conjunto adicional de $0 < C < N$ conexiones. Más concretamente: elegimos al azar, para cada $S_j^i$, de cada $j$-ésima especie, $C$ propiedades de las demás, esto es

$$\{S_{j1}^1, \ldots, S_{jC}^C\}$$

(con $1 \leq ik \leq N$) con las que definiremos un nuevo conjunto de valores de adaptación (como en la tabla anterior). Este modelo recibe el nombre de red $NKC$. Hemos conseguido, de esta forma, conectar las $S$ especies formando un ecosistema en el que cada transición (salto adaptativo) llevado a cabo por una especie modificará el paisaje adaptativo de todas y cada una de las especies con las que se halla relacionada.

Típicamente, este sistema tiende a un equilibrio de Nash: las especies alcanzan óptimos locales en los que el sistema permanece estable. Señalemos que el tiempo requerido para alcanzar el equilibrio de Nash aumenta considerablemente con el número de conexiones entre especies, $C$. También ocurre que a lo largo de la evolución algunas especies forman subredes que prontamente no varían su estado, mientras que otras partes del sistema pueden cambiar con rapidez (Kauffman y Johnson, 1991).

Las observaciones de Kauffman y Johnson son numerosas pero la más relevante es la observación de las distribuciones de frecuencias de cambios. Imaginemos que, una vez alcanzado un equilibrio de Nash, el sistema empieza a cambiar: cada especie experimenta modificaciones en su paisaje adaptativo que la llevan a realizar saltos hacia posiciones de mayor $< F >$. Imaginemos que medimos el número de cambios hasta que el sistema alcanza de nuevo un equilibrio de Nash, lo que Kauffman y Johnson llaman avalanchas coevolutivas. De forma similar a lo que ocurrirá con las redes de Kauffman (capítulo 11) existen distintas fases de comportamiento, dependientes de los valores de $K$ y $C$, que corresponden a sistemas congelados (todas las especies en sus picos adaptativos locales) o a sistemas permanentemente cambiantes (fase caótica). Existe una zona de transición en la que tiene lugar el cambio de un estado al otro. Lo más importante de los resultados de Kauffman es precisamente la observación de que en este punto crítico se observan
Figura 12.10: (a) Distribución de avalanchas para sistemas $5 \times 5$ (empleamos una red cuadrada), para $N = 24$, $C = 1$, $K = 10$; (b) idénticos parámetros, pero $K = 14$; (c) idem, $K = 20$.

leyes potenciales en las distribuciones de avalanchas que, de forma genérica, Kauffman compara con los acontecimientos de extinción. En la figura 12.10 vemos algunos ejemplos de las distribuciones obtenidas.

La conjetura de estos autores es que la evolución puede entenderse como un proceso de criticalidad autoorganizada (aunque ellos no proporcionan un mecanismo para acceder al punto crítico). El modelo es enormemente sugestivo, aunque algo restringido en el hecho de fijar parámetros. El siguiente modelo, debido a Per Bak (capítulo 8) y sus colaboradores aborda el problema desde una aproximación mucho más simplificada pero no menos sugerente.

### 12.5 Modelo de Bak-Sneppen

Vamos a presentar en esta sección un modelo sencillo de macroevolución debido a Bak y Sneppen (1993), junto con su tratamiento analítico según la teoría del campo medio (véase el capítulo sobre fenómenos críticos), tal y como aparece en el artículo de Flyvbjerg, Sneppen y Bak de 1993.

Este modelo de evolución consiste en cierto número de especies en interacción, cada una de ellas caracterizada por un valor numérico que expresa el nivel de adaptación que la especie ha conseguido. Las especies se desplazan por medio de mutaciones en un paisaje adaptativo. En principio, mutaciones del códico genético de cada especie trasladarían aleatoriamente la posición que ocupa en este paisaje. Sin embargo, sólo se permiten mutaciones que incrementen la adaptación de la especie, de forma que se sitúa finalmente (y de forma rápida) en alguno de los máximos locales del paisaje. El movimiento adaptativo es rápido. Sólo es posible que se siga produciendo evolución si se permite la existencia de movimientos hacia peores adaptaciones con baja probabilidad. En este caso, la mayor parte del tiempo se encuentra a las especies situadas en máximos locales.

La estabilidad de cada especie se caracteriza por la altura de la barrera que separa su máximo local de estabilidad de otros máximos mejores. La altura de esta barrera es una medida del número de bits o de la cantidad de código genético que debe ser cambiado para alcanzar la mejor posición. Las mutaciones de un único bit se producen con cierta frecuencia, pero los movimientos coordinados
que resultan en una mejora complicada en la especie son mucho más raros. El tiempo de espera que se necesita para saltar a un máximo más pronunciado depende exponencialmente de la altura de la barrera. Cuando la adaptación es elevada, es difícil encontrar máximos próximos mejores, y por tanto la especie es bastante estable en su máximo local. Cuando la adaptación es baja, es mucho más probable encontrar estados próximos que mejoren la situación, y por tanto las barreras son bajas.

De entre todas las especies consideramos únicamente aquella que presenta la barrera menor, \( B_i \), (muy más probable de superar que cualquiera de mayor altura que ella). Las barreras son la medida de estabilidad en este modelo. El salto de la barrera se puede interpretar como una mutación de la especie o como su substitución por otra mejor en el nicho ecológico. Como consecuencia de que las barreras menores se corresponden mayoritariamente con una baja adaptación, y al revés sucede con las barreras mayores, las barreras son también una medida de la adaptación de la especie. Dado que las barreras menores son inestables, existe una colección de especies que no interaccionan y que acaban convergiendo a un estado "congelado" con las barreras más altas.

Se requiere un mecanismo que simule la interacción de las especies en el paisaje adaptativo. Supongamos que un movimiento de una de las especies provoca cambios en el paisaje, y por tanto inducirá mutaciones en las especies en él situadas. Una especie que posea una barrera muy alta puede ser incapaz de moverse por sí sola, pero finalmente será afectada por un vecino que al mutar reduzca su barrera y posibilite el cambio.

El modelo intenta representar las características expuestas, supuestas como más relevantes en el proceso de macroevolución, de la forma siguiente:

1. Se sitúa a \( N \) especies en una red unidimensional con condiciones periódicas de contorno.
2. Se asigna un valor aleatorio a la barrera, \( B_i \), con distribución homogénea en el intervalo \([0,1]\).
   En cada paso de tiempo, el sistema se actualiza de acuerdo con:
3. Se localiza la barrera menor y se asigna un nuevo valor aleatorio a esta celda.
4. Se cambia el paisaje adaptativo de los dos vecinos situados a la derecha y a la izquierda asignándoles de nuevo dos valores aleatorios a sus barreras.

La selección de la menor de las barreras se justifica por la separación exponencial entre las escalas de tiempo. Cuando se inicia la simulación, las actualizaciones sucesivas están muy descorrelacionadas, pero a medida que el valor de las barreras crece, es cada vez más probable que los vecinos próximos de una especie que acaba de mutar sean los siguientes en mutar, y se produce en consecuencia la correlación de los sucesos. Después de un largo transitorio, se llega a una distribución estacionaria.

Una de las medidas a realizar en el estado estacionario es la distribución \( C(x) \) de la distancia \( x \) entre mutaciones sucesivas. Se obtiene en este caso una distribución potencial,

\[
C(x) \propto x^{-3.15 \pm 0.05}
\]

lo cual confirma la conjetura de que el sistema se comporta de forma crítica. Esta distribución no depende de las condiciones iniciales, así que el estado estacionario es un atractor de la dinámica que se alcanza por autoorganización del sistema.

El cálculo de la distribución de los valores de barreras que mutan revela que todas las mutaciones tienen lugar para valores inferiores al crítico, \( B_C = 0.67 \pm 0.01 \). El umbral define el tiempo máximo de espera entre mutaciones sucesivas.

Si no existiesen correlaciones entre las especies que forman el sistema, todas alcanzarían el valor de máxima adaptación, \( B_i = 1 \), \( \forall i \), pero con extrema lentitud. Incluso si la regla de actualización
Figura 12.11: Equilibrio puntuado en el modelo de evolución descrito en el texto. Se ha representado la dinámica del número de “mutaciones” a lo largo del tiempo.

se cambia y, en lugar de sustituir en cada paso el valor de las barreras de los vecinos del elemento que tenía el valor menor de $B_i$, simplemente se elimina esta especie, el sistema resultante se encuentra aún lejos del equilibrio, es decir, del valor 1 para todas las barreras.

La actividad en el sistema está bien caracterizada utilizando “avalanchas evolutivas”. Se considera pertenecientes a la misma avalancha a todas las extinciones que han tenido lugar consecutivamente por debajo de un cierto umbral fijado. Cuando ha cesado la actividad durante un paso de tiempo, con el criterio definido por este umbral, se considera la avalancha finalizada. El número de mutaciones por debajo del umbral define una avalancha de tamaño $s$. La distribución de los tamaños, $N(s)$, sigue una ley potencial

$$N(s) \propto s^{-0.9\pm0.1}$$

indicando que se obtienen avalanchas coevolutivas a todas las escalas, incluyendo el tamaño total del sistema (véase el capítulo 8).

Si se asocia estabilidad con adaptación, la última es baja durante las grandes avalanchas, en tanto que se observa una alta adaptación durante los períodos de estasis con escasa actividad.

El mecanismo de evolución cerca de un estado crítico se puede interpretar como una búsqueda de una mejor adaptación local, que sucede raramente, pero que en ocasiones puede tener una influencia brutal en el sistema.

12.5.1 Teoría de campo medio

Supongamos que una avalancha evolutiva afecta a un total de $K$ especies, y que su duración temporal es despreciable frente al tiempo requerido para que el sistema evolucione espontáneamente en el estado estacionario (según las suposiciones realizadas en la sección anterior). El estado del ecosistema de $N$ especies se caracteriza por los valores

$$\{z_i\}, \quad i = 1, 2, \ldots, N$$

que representan las barreras efectivas hacia máximos mejores que el local en el que se encuentre la especie $i$. La dinámica consiste en seleccionar la especie con menor valor de barrera y cambiar
Figura 12.12: (a) Distribución de frecuencias en el número de mutaciones (actividad) y (b) frecuencias de duración de las avalanchas.

Esta cantidad, junto con la de $K - 1$ especies más. Los resultados del modelo dependen de cómo se decida escoger las $K$ especies. Una posible elección consiste en situar las especies en una red hiperesférica $d$-dimensional con interacción a primeros vecinos, con lo cual $K = 2d + 1$.

Seleccionamos por conveniencia matemática las $K - 1$ especies en interacción aleatoriamente entre las $N$ especies del sistema. Además, se supone que si una especie inicia varias veces una avalancha, las especies afectadas se escogen cada vez aleatoriamente entre todas las del sistema. Se puede realizar un análisis de campo medio despreciando las correlaciones entre los valores de las barreras. Si llamamos $p_i$ a la distribución del valor de barrera $x_i$, entonces la aproximación de campo medio consiste en suponer que

$$ p_i(x) = \frac{N!}{(i-1)!(N_i)!} p^{i-1}(x) p(x) Q^{N_i}(x) $$

de donde la distribución para la barrera menor será

$$ p_1(x) = N p(x) Q^{N_1}(x) $$

y donde se han definido

$$ P(x) = \int_0^x dx', p(x') \quad \quad Q(x) = \int_x^1 dx', p(x') $$

La normalización de $p$ proporciona

$$ \int_0^1 dx' p(x') = P(x) + Q(x) = 1, \quad \forall x $$

La ecuación dinámica para $P(x, t)$ es

$$ p(x, t + 1) = p(x, t) - \frac{1}{N} p_1(x, t) - \frac{K - 1}{N - 1} \left( p(x, t) - \frac{1}{N} p_1(x, t) \right) + \frac{K}{N} \quad (12.5.1) $$

La barrera menor se ha eliminado de esta expresión (segundo término de la derecha). El tercer término a la derecha representa la eliminación de $K - 1$ barreras de las $N - 1$ restantes tras
Eliminar la menor. Como estas $K-1$ pueden ser cualesquiera de las restantes, seguirán su misma distribución, es decir, $[Np(x,t) - p_1(x,t)]/(N-1)$. El último término representa la adición de $K$ nuevos valores equidistribuidos, que reemplazan a los eliminados por la avalancha. Observese que la ecuación total conserva la probabilidad.

La dinámica de campo medio es una aproximación a la ecuación maestra para el proceso de Markov de vecinos aleatorios, presentando ambos un único punto fijo atractivo. La ecuación 12.5.1 es una ecuación integral para $p(x)$, o bien una ecuación diferencial ordinaria para $Q(x)$. Su solución será la raíz positiva de la ecuación polinómica

$$(N-K)Q^N(x) + N(K-1)Q(x) + (N-1)K(z-1) = 0$$

(12.5.2)

En el límite $N \gg K > 1$, el primer término de esta ecuación es despreciable frente al segundo para los valores de $z$ donde $Q(z) < 1$ en un factor superior a $O(1/N)$. Tenemos por tanto

$$Q(z) = \frac{(N-1)K}{N(K-1)}(1-x) - \frac{N-K}{N(K-1)}Q^N(x) \approx \frac{K}{K-1}(1-x), \quad si \quad \frac{x-1}{K} \gg O(1/N) \quad (12.5.3)$$

En el caso $Q(z) \approx 1$ tendremos

$$Q(x) = \left( \frac{(N-1)K}{N-K}(1-x) - \frac{N(K-1)}{N-K}Q^N(x) \right)^{1/N} \approx (1-Kx)^{1/N}, \quad si \quad \frac{1}{K} - x \gg O(1/N)$$

(12.5.4)

considerando que $p(x) = -dQ(x)/dx$, se obtiene

$$p(x) \approx \frac{K}{N}, \quad para \quad \frac{1}{K} - x \gg O(1/N)$$

$$p(x) \approx \frac{K}{K-1}, \quad para \quad x - \frac{1}{K} \gg O(1/N)$$

La solución exacta para 12.5.2 se obtiene de forma sencilla numéricamente por iteración de las ecuaciones 12.5.3 y 12.5.4 para $x > 1/K$ y $x < 1/K$, respectivamente.

En el límite $N \to \infty$, $p(x)$ presenta una discontinuidad en $x = 1/K$: es cero por debajo de este valor y constante por encima. Se puede entender este resultado de forma aproximada. Supongamos que

$$p(x) \approx \frac{K}{N}, \quad para \quad 0 \leq x \leq \frac{1}{K}$$

y que

$$p(x) \approx \frac{K}{K-1}, \quad para \quad 0 \leq \frac{1}{K} \leq x$$

Entonces, el valor menor de las $N$ barreras, siguiendo la distribución definida por $P$, estará equidistribuido por debajo del umbral $1/K$, y los restantes $N-1$ valores serán típicamente mayores que $1/K$. Por tanto, cuando se elimina el valor menor, no queda ninguno por debajo del umbral, y los $K-1$ valores que se eliminan juntamente con éste han de ser escogidos por encima del umbral. Estos valores también están equidistribuidos: una vez eliminados, son reemplazados por otros $K$ que escogemos arbitrariamente de una distribución uniforme, en la cual, típicamente, uno de estos valores caerá por debajo del valor umbral. La probabilidad $p$ es estacionaria y se mantiene inalterada tras este proceso.
Este razonamiento nos conduce a otro aspecto de la dinámica asintótica del modelo. Si observamos a lo largo del tiempo cuál es la especie que desencadena cada avalancha, veremos que con gran frecuencia es una de las que acaban de participar en la última avalancha. Si aplicamos este razonamiento a la evolución real, parece que las especies más antiguas serán también las que tendrán mayores posibilidades de sobrevivir, lo cual, según los resultados expuestos en secciones previas, parece no ser cierto (Van Valen, 1973).

Se ha supuesto anteriormente que el número de elementos con valor de barrera bajo el umbral es 1. Se puede obtener una mejor estimación considerando la aproximación de campo medio, que proporciona

\[ NP \left( \frac{1}{R} \right) = \ln N - \ln(\ln N) - \ln(K - 1) + O \left( \frac{\ln(\ln N)}{\ln N} \right) + O \left( \frac{1}{\ln N} \right) + O \left( \frac{\ln N}{N} \right) \]

donde \( P(1/K) = 1 - Q(1/K) \), y \( Q(1/K) \) es la solución de la ecuación 12.5.2 con \( x = 1/K \). El valor medio del número de barreras bajo el umbral es \( NP(1/K) \). La fluctuación en este valor medio que se requiere para que finalice una avalancha se hace cada vez más rara a medida que \( N \) aumenta. Por tanto, el tamaño de las avalanchas, definido como el número de extinciones que contienen, crece con \( N \), y diverge cuando \( N \to \infty \).

Cuando \( N \to \infty \), se obtienen avalanchas de todos los tamaños, y la aproximación de campo medio permite determinar que las avalanchas mayores estarán distribuidas según una ley potencial,

\[ D(s) \propto s^{-3/2} \]

No existe valor medio para estas avalanchas críticas.

Existen algunos otros modelos recientes, algo más sofisticados, y que producen resultados más cercanos a los datos experimentales que se encuentran a partir del registro fósil. Uno de ellos se debe a Newman y Roberts (1995), que incluyen en su modelo tanto la interacción interespecífica, del mismo modo que Bak y Sneppen, como una nueva influencia, debida al medio. No describiremos extensivamente este modelo, pero el lector interesado puede consultar la referencia al final del capítulo.

En la siguiente sección describiremos un último modelo, con una sofisticación a medio camino entre los dos anteriores, y con resultados compatibles con los datos paleontológicos.

12.6 Modelos con extinción explícita

Los modelos anteriores emplean la idea de extinción en un sentido muy abstracto: no incorporan verdaderas extinciones ni la posibilidad de diversificación (esto es, de aparición de nuevas especies a partir de especies previas). Este es, sin duda, un punto débil de los modelos vistos (Maddox, 1994; 1995) especialmente en relación a las posibles comparaciones que podamos realizar con los datos procedentes del registro fósil. En este sentido, un modelo capaz de generar extinciones por medio de la interacción entre especies y que permita generar nuevas variantes a partir de las supervivientes sería deseable. Dicho modelo es posible (Solé, 1996) y permite comprobar la aparición de leyes potenciales asociadas a los acontecimientos de extinción.

En este modelo de macroevolución se pone especial énfasis en las conexiones entre las especies que componen la ecología, y son éstas las que desencadenan las extinciones. En un sistema con \( N \) especies, se supone que todas están conectadas con todas (por tanto influenciadas por todas), incluso con ellas mismas (lo cual puede interpretarse como la influencia que el estado interno de
la propia especie puede tener en su supervivencia). Esta restricción, suponer que conectividad es total, puede relajarse y hacerse variable (Solé, 1996).

Cuando la suma de las conexiones que una especie particular recibe cae por debajo de cierto umbral, la especie se extingue. Inmediatamente, el nicho que ha quedado vacío es ocupado por una nueva especie, que es una copia, con alguna mutación en las nuevas conexiones, de alguna de las especies supervivientes. Este hecho permite introducir la diversificación.

Consideremos un sistema con $N$ especies totalmente conectadas. El valor del enlace entre la especie $i$ y la especie $j$ es $w(i,j)$, con $w(i,j) \in [-1, 1]$, y la conexión es en principio un valor equidistribuido en el intervalo señalado. La simulación exacta del modelo sigue los pasos siguientes:

1. A cada paso de tiempo, y para cada una de las $i$ especies, consideramos la cantidad

   $$h_i = \sum_{j=1}^{N} w(i,j)$$

   que representa la entrada total sobre esta especie. Cuando este campo es menor que cero, $h_i < 0$, la especie se extingue.

2. Todas las conexiones $w(i,j)$ y $w(j,i)$ correspondientes a especies extinguidas toman el valor cero.

3. De entre las especies supervivientes, se escoje aleatoriamente una, digamos $k$. Los espacios que han quedado vacíos son ocupados por copias mutadas de esta especie, tal y como sigue. Para cada valor de las conexiones $w(k,l)$ y $w(l,k)$ se realiza la copia

   $$w(i,l) = w(k,l) + \epsilon r_{il}; \quad w(l,i) = w(l,k) + \epsilon r_{li}$$

   en el lugar que ocupaba la especie $i$ extinguida, donde $r$ es un número aleatorio con distribución uniforme en el intervalo $[-1, 1]$ que se escoge nuevamente para cada conexión y $\epsilon$ controla el tamaño de la mutación.
Figura 12.14: Variación del campo $h_i$ de una especie arbitraria a lo largo del tiempo. Cada vez que este valor cae bajo cero, la especie se ha extinguido. Obsérvese cómo, cuando tras la colonización de la celda el campo presenta un valor alejado del umbral crítico de extinción, este valor “deriva” lentamente hacia el valor cero.

4. Se escoge una conexión arbitraria en cada una de las $N$ especies en el sistema y se cambia al azar.

Y se repite el proceso.

Remarquemos que no existe ninguna condición de simetría en el valor de las conexiones. En efecto, la influencia que cierta especie $i$ pueda ejercer sobre la especie $j$ no implica nada sobre la influencia contraria (piénsese, sin ir más lejos, en los casos presa-depredador, coevolución o huésped-parasitoide).

Tras un periodo transitorio, el sistema alcanza un estado crítico, en donde se puede observar una distribución potencial de extinciones, entre las cuales aparecen ocasionalmente algunas incluso del tamaño total del sistema. Esperando un tiempo suficiente, se puede llegar incluso a la extinción total de todo el ecosistema. El tamaño $s$ de una extinción se calcula por medio de avalanchas, de la forma siguiente. Sea $n(t)$ el número de especies eliminadas en el paso de tiempo $t$ de la simulación. Entonces,

$$s = \sum_{t=t_i}^{t_f} n(t), \quad n(t_i) = 0, \quad n(t_f) = 0, \quad n(t) \neq 0, \quad \forall t_i < t < t_f$$

La distribución $N(s)$ de los valores de $s$ sigue una ley potencial con exponente $\alpha = 2.3 \pm 0.1$. Otra magnitud que presenta una distribución potencial es el tiempo de espera $T_e$ entre extinciones. Esta magnitud se define como $T_e = t_f - t_i$, con $n(t) = 0$, $\forall t_i \leq t \leq t_f$, y se obtiene una distribución $D(t_e)$ con exponente $\beta = 3.9 \pm 0.1$.

La dinámica macroscópica está guiada por el valor que presentan las conexiones. La conectividad tiende asintóticamente a ser total. Pensemos que es así como se permite que sea, pero que las extinciones y la eliminación de conexiones que conllevan alejan al sistema de esta posibilidad. Cada vez que se produce una extinción, y se produce la copia consiguiente, se aleja al sistema del estado crítico. En algunas ocasiones, una buena copia (lo cual significa que se copia una especie i con un campo $h_i \approx 1$) puede provocar un largo periodo de inactividad (en el sentido de ausencia de
extinciones) en el sistema. Por un proceso de difusión, provocado por los pequeños cambios aleatorios en las conexiones, el sistema retorna lentamente al estado crítico. En otras ocasiones, una mala copia (a copia ahora una especie con \( h_i \approx 0 \), aunque positivo, puesto que ha sobrevivido) puede inducir una extinción de mayores proporciones que la anterior. Además, puede también darse el caso de que cuando desaparece una especie que es “beneficiosa” para otras (en el sentido de que sus conexiones de salida son positivas), las que estaban sustentadas por ésta queden con un campo negativo que el cambio aleatorio de una única conexión sea incapaz de cambiar de signo. Podríamos ver aquí el papel de la contingencia en las extinciones: las malas copias encadenadas y las extinciones de especies de soporte son las que provocan las grandes extinciones en el modelo. Obsérvese la dinámica temporal de las extinciones y los “preludios” a las extinciones en masa.

La conclusión principal de este modelo y de los anteriores es la clara posibilidad de que la estructura del proceso de macroevolución sea el resultado de un proceso de autoorganización, que eventualmente se verá perturbado por causas físicas. Estas perturbaciones pueden desencadenar grandes extinciones, pero no tienen porqué ser el motivo estricto de éstas. Los factores bióticos podrían situar al sistema en el punto crítico y sería esta configuración (alcanzada por motivos de interacción entre especies) la que podría, eventualmente, hacer vulnerable el sistema.

Podemos resumir los resultados anteriores en la tabla siguiente (Solé y Bascompte, 1995):

<table>
<thead>
<tr>
<th>Propiedad</th>
<th>Observación</th>
<th>Proceso aleatorio</th>
<th>Criticalidad</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frecuencia de extinción</td>
<td>Decreciente(*)</td>
<td>Exponencial</td>
<td>Potencial</td>
</tr>
<tr>
<td>Dinámica (tasas)</td>
<td>Puntuada</td>
<td>Aleatoria (RW)</td>
<td>Puntuada</td>
</tr>
<tr>
<td>Dinámica (temporal)</td>
<td>( 1/f (**) )</td>
<td>( 1/f^2 ) (RW)</td>
<td>( 1/f )</td>
</tr>
<tr>
<td>Taxonomía</td>
<td>Fractal</td>
<td>Exponencial</td>
<td>Fractal</td>
</tr>
<tr>
<td>Extinciones en masa</td>
<td>Episódicas</td>
<td>Imposibles</td>
<td>Esperadas</td>
</tr>
</tbody>
</table>

Tabla II: Algunas propiedades generales de la macroevolución, tal y como se observan y como se predicen por medio de modelos aleatorios o críticos.

(*) Los datos accesibles son compatibles tanto con un ajuste exponencial como un potencial, aunque el ajuste es mejor en el segundo caso.

(**) Basado únicamente en los datos sobre ammonites (House, 1989)
Acabemos añadiendo que existe una crítica adicional, basada en el hecho de que la unidad de cambio en los modelos anteriores es la especie como tal, y no los individuos. Sin embargo, se asume en principio de forma natural que son los individuos las verdaderas unidades de selección. ¿Qué tipo de comportamiento esperaremos observar en un ecosistema complejo en el que tenemos individuos como entidades bien definidas formando poblaciones en interacción? Una forma de abordar este problema se lleva a cabo en el dominio de una disciplina denominada vida artificial (Levy, 1992, y referencias citadas). Este campo de investigación intenta plantear de forma muy general las propiedades y la dinámica de sistemas capaces de autoreplicarse, interaccionar y, en definitiva, evolucionar. Uno de los modelos más conocidos es el simulador Tierra, creado por Thomas Ray, de la Universidad de Delaware (Rennie, 1992; Adami, 1995). Este sistema está formado por una población de programas que compiten por la memoria del ordenador y que poseen la capacidad de mutar generando nuevos programas. El modelo Tierra es un buen ejemplo de evolución artificial: una vez en marcha, aparecen de forma espontánea parásitos, hiperparásitos, etc. Los parásitos de hecho influyen de forma decisiva en la generación de diversidad en el sistema. Y el modelo genera extinciones. Un ejemplo de la distribución de probabilidad de extinciones se muestra en la figura 12.15. Vemos una vez más una distribución potencial característica (con una cola final que se separa, debido a los efectos de la longitud finita de la simulación). La conclusión que extraemos de este resultado es que la extinción es, ciertamente, un fenómeno genérico, inevitable, en la evolución de un sistema complejo adaptativo.

12.7 Evolución, caos y contingencia

Entre las preguntas abiertas que podríamos formularnos acerca de la evolución está el problema de hasta qué punto es predecible, en algún sentido, la dinámica del proceso. Parece claro que, en principio, los detalles específicos son impredecibles. Los pequeños cambios producidos de forma fortuita pueden tener, a medio o largo plazo, grandes efectos. Si el escenario planteado por los modelos anteriores es adecuado, entonces la dinámica no-lineal hará virtualmente imposible predecir qué especies se hallarán presentes. En principio, lo único que posiblemente se conservará serán las propiedades macroscópicas relacionadas con la funcionalidad del ecosistema, como son
las relaciones entre diversidad y conectividad.

Stephen Jay Gould es el principal defensor de esta visión contingente del proceso de evolución (Gould, 1994):

"La historia incluye demasiado caos, o una dependencia extremadamente sensible de diferencias minúsculas e inconmensurables en las condiciones iniciales, lo que conduce a resultados divergentes sobre la base de disparidades mínimas en el origen. Y la historia comprende también la contingencia: los resultados actuales se deben a largas cadenas de estados previos impredecibles y no a la determinación inmediata por leyes eternas de la naturaleza."

Los resultados de Burlando acerca de la fractalidad de las relaciones taxonómicas nos sugieren la existencia de propiedades de autosimilaridad en la evolución futura que podrían reflejar la existencia de un estado crítico autoorganizado. Los datos obtenidos acerca de las fluctuaciones de las familias de ammonites son en principio consistentes con esta observación.

Hay sin embargo un factor de importancia a tener muy en cuenta. Las especies poseen una morfología (externa e interna) que las identifica como tales de forma específica. En la gran explosión de Cámbrico (Gould, 1991; 1994) se generaron, muy posiblemente, todos los tipos morfológicos que hoy subsisten (algunos de los aparecidos en la explosión del Cámbrico desaparecieron). Podríamos plantearnos la siguiente pregunta: ¿son posibles todas las morfologías imaginables? Esta cuestión está lejos de ser trivial. Intuitivamente, podríamos creer que, bajo la presión de la selección, se eligen cierto tipo de formas que proporcionan una mejor adaptación. Las formas observadas (y sobre las que podríamos, en principio, construir explicaciones de corte adaptacionista) no son arbitarias, pero deberíamos preguntarnos si existen restricciones de tipo genérico al conjunto de formas posibles. Algunos científicos opinan que, de hecho, el conjunto de formas está restringido a unas pocas combinaciones, y que estas restricciones al desarrollo son debidas a causas intrínsecas al proceso de morfológica (Goodwin, 1994).

Podemos visualizar las distintas alternativas en una imagen (figura 12.17) que resume un experimento mental. Imaginemos (como sugiere Gould, 1991) que pudieramos "rebobinar" la cinta de la película de la vida y volver hacia atrás, antes de la gran explosión del Cámbrico. Dejamos entonces que la película se ruede otra vez, teniendo en cuenta que ahora las pequeñas diferencias pueden ser amplificadas a lo largo del proceso. Pueden ocurrir básicamente tres cosas:

- Todo vuelve a ser exactamente igual. Los factores deterministas son tan fuertes que impiden de hecho que la historia se escriba de otra forma (figura 12.17 (a)).

- El escenario final es distinto, pero las restricciones al desarrollo hacen que, de hecho, veamos un conjunto de formas similares básicamente a las del caso original. Las limitaciones intrínsecas juegan un papel importante (figura 12.17 (b)).

- Todo es distinto. No existen limitaciones que impidan obtener formas de distintos tipos, y las pequeñas alteraciones de las condiciones iniciales han modificado drásticamente el resultado final (figura 12.17 (c)).

No sabemos si la verdad está incluida en una de estas opciones o si tal vez es una combinación de éstas. Pero no cabe duda de que ciertas restricciones operan a algunos niveles. En ecosistemas de distintos continentes encontramos una gran variedad de especies, pero no es raro hallar convergencias muy notables, como la presencia de osos hormigueros o ardillas voladoras, que sugieren la existencia de mecanismos comunes de generación de estructuras. Tal vez ocurre, en la evolución, aquello que ya hemos visto en nuestro estudio del caos determinista (Cohen y Stewart, 1991): existe contingencia y sensibilidad a las condiciones iniciales pero el proceso está confinado a un conjunto
Figura 12.17: Rebobinando la película. Supongamos que pudiéramos rebobinar la historia de la vida sobre la Tierra para volver a los comienzos. Imaginemos, por ejemplo, que estamos en el Precámbrico, antes de la aparición de los tipos morfológicos que conocemos. Podríamos encontrarnos con tres escenarios posibles para la evolución. En el primero (a) todo vuelve a ser como en la película anterior. Los factores deterministas juegan un papel tan preponderante que nada distinto puede tener lugar. Es un camino de una sola dirección. (b) El escenario es distinto, pero identificamos estructuras familiares, como espirales, segmentos, tubos, etc. Aunque podemos ver organismos muy distintos, su morfología comparte propiedades universales con las de otras posibles evoluciones. (c) El resultado de variar las condiciones iniciales da lugar a un escenario final donde emergen organismos completamente distintos. No hay restricciones importantes a la morfología y muchas formas son posibles.
Figura 12.18: Fluctuaciones temporales en un índice económico y espectro de Fourier asociado, del tipo $1/f$.

de posibilidades (el atractor) no arbitrario. Ambos procesos se hallan presentes, haciendo imposible la predicción del detalle, pero permitiendo una predicción racional de la naturaleza básica de las estructuras.

Cabe preguntarse, por último, si existen propiedades de tipo genérico en la dinámica de sistemas complejos adaptativos que comparten con la evolución biológica ciertos rasgos básicos. Este sería el caso de la economía. Aunque los modelos económicos clásicos son fundamentalmente modelos de equilibrio, los sistemas económicos reales son sistemas complejos adaptativos que también presentan fenómenos de contingencia y, tal vez, existe criticalidad en su dinámica. En la figura 12.18 vemos un ejemplo de fluctuaciones de un índice económico que sugiere claramente esta posibilidad, y existe de hecho un tratamiento formal basado en los modelos de Kauffman antes discutidos (Kauffman y Macready, 1995). De ser así, las grandes fluctuaciones que de vez en cuando generan grandes crisis económicas, serían el resultado de la evolución de un sistema complejo adaptativo en el filo del caos.

Bibliografía


Capítulo 13

Retrovirus y Cuasiespecies: Entre el Orden y el Caos

En este capítulo abordaremos un problema de especial importancia no sólo teórica sino (y muy especialmente) práctica. Sus protagonistas tienen mucho que ver con el orden, el caos y la complejidad. También representan formas de evolución rápida y a la vez la señal de que un sistema complejo, como es nuestro organismo, está expuesto a errores que pueden ellos mismos evolucionar hacia la complejidad. Durante la segunda mitad del siglo XX el desarrollo espectacular de la biología molecular ha conducido a un conocimiento antes impensable acerca del cáncer. A la vez, nos enfrentamos con nuevos tipos de virus, como el virus del SIDA, que plantean retos de gran dificultad a nuestra capacidad de responder a la enfermedad (Domingo, 1994; Domingo y Holland, 1995). Trataremos aquí de esbozar (muy esquemáticamente) las teorías actuales acerca del origen de los retrovirus y su forma de actuación. Veremos cómo los modelos nos acercan a aspectos del problema difíciles de abordar de otra forma, y cómo las no-linealidades vuelven a darnos sorpresas.

13.1 Información genética

Antes de entrar en el formalismo teórico, daremos en esta sección un repaso muy somero a los conceptos elementales de biología molecular relacionados con la información genética. Este resumen no pretende ser, obviamente, todo lo completo que debería. La información detallada puede encontrarse en cualquier buen texto de genética.

Todos los seres vivos almacenan la información necesaria para llevar a cabo su replicación en una o varias moléculas de ácido nucleico. Este puede ser ácido desoxirribonucleico (ADN) o ácido ribonucleico (ARN). En ambos casos cada molécula emplea cuatro símbolos básicos en los que se almacena la información (cuatro nucleótidos) y que constituyen el "alfabeto" del código. En las células (de cualquier tipo) el ADN almacena la información en secuencias de nucleótidos dispuestas a lo largo de la molécula y que, genéricamente, llamaremos genes. Cada gen debe ser "leído" por un enzima que, empleando el ADN como molde, construye una cadena de ARN (el llamado ARN mensajero). Este proceso se denomina transcripción. Este ARN será a continuación leído a su vez por los ribosomas para formar una proteína. Este proceso de lectura final se denomina traducción.

Tenemos así, para una célula típica, una secuencia de transferencia de información.

---

1 Empleamos esta definición poco rigurosa de forma consciente. La consulta de un texto de genética molecular mostrará al lector el motivo de nuestra elección. La definición de gen, a menudo planteada de forma muy simple, no lo es.
Figura 13.1: Estructura básica del virus del SIDA (VIH).

ADN $\rightarrow$ ARN $\rightarrow$ Proteína

esta secuencia de lectura fue conocida en biología molecular como el dogma central.

Existen sin embargo otras alternativas. Una de ellas es emplear el ARN como punto de partida, como ocurre en algunos virus. Una posibilidad es copiar el ARN (mediante una polimerasa adecuada) y realizar la traducción como antes. Sin embargo, una posibilidad al principio no considerada y que sin embargo también se da es el flujo de información inverso de ARN a ADN

ADN $\rightarrow$ ARN $\rightarrow$ Proteína

siendo el paso de ARN a ADN realizado mediante la denominada retrotranscriptasa o transcriptasa inversa. El descubrimiento de este enzima representó un paso gigantesco en nuestra comprensión de la interacción entre virus y células, así como de otros problemas, aparentemente distintos, como es el origen del cáncer.

13.2 Variabilidad en retrovirus

Los retrovirus son un grupo muy importante dentro de los virus de ARN, el grupo más extendido de parásitos intracelulares. En la figura 13.1 se muestra la estructura básica del más famoso de los retrovirus: el virus de la inmunodeficiencia humana (VIH) también conocido como virus del SIDA (por Síndrome de Inmunodeficiencia Adquirida). Su estructura básica es compartida por la mayoría de los virus: cierta cantidad de material genético (ARN en este caso) rodeada de una envoltura externa. En el interior encontramos también el enzima capaz de realizar la transcripción ARN $\rightarrow$ ADN, la denominada transcriptasa inversa.

El virus del SIDA apareció, según se ha estimado mediante técnicas de secuenciación junto con técnicas estadísticas especiales (Eigen, 1993) hace entre 600 y 1200 años. En este sentido,
es un virus “reciente”. Pertenece a la familia de los denominados virus emergentes. Aunque el término sugiere la aparición reciente de una entidad completamente nueva, de hecho sus orígenes son habitualmente lejanos en el tiempo. Lo que los hace nuevos es en realidad su invasión de un organismo con el que raramente tenían contacto (por ejemplo, un vertebrado distinto del huésped habitual, digamos el hombre) como consecuencia de la alteración del ecosistema en el que habita el organismo portador para el cual el virus raramente será letal.

Se ha estimado que alrededor del 70 por ciento de los virus que infectan organismos diferenciados son virus de ARN. Las células de los organismos complejos han evolucionado a lo largo del tiempo en compañía de estas entidades subcelulares. En el genoma de cualquier organismo multicelular encontraremos de hecho la evidencia contundente de esta historia compartida: cientos, miles de retrovirus (aparentemente “dormidos” o simplemente inactivos por alguna carencia) se hallan insertados en el genoma.

Para una célula compleja, existen numerosas limitaciones para la variabilidad en la expresión genómica. Más concretamente, una mutación en un sólo nucleótido puede hacer completamente errónea o simplemente anular la actividad de la proteína sintetizada. Existe una multitud de ejemplos de enfermedades humanas asociadas a una mutación en un gen determinado de los $10^5$ que componen el genoma. Es obvio que el sistema de replicación debe estar adecuadamente ayudado por mecanismos (reductantes) de comprobación. Se dice que el genoma posee un mecanismo de reparación que evita la aparición y acumulación de mutaciones, que a medio plazo serían letales. Puesto que el genoma de un organismo complejo es de gran tamaño, las tasas de mutación deben ser muy bajas. De hecho, las estimaciones nos dan valores de tasas de mutación por nucleótido de $10^{-8}$ a $10^{-11}$ por ciclo de replicación. Si experimentalmente suprimimos los mecanismos de reparación, estas tasas se disparan a valores de $10^{-4}$ sustituciones por nucleótido por ciclo de replicación. La estabilidad genómica requiere por lo tanto mecanismos que cuiden al máximo la fidelidad de copia.

Los virus de ARN, en cambio, nos muestran la otra cara de la moneda. Sus tasas de mutación enormes, junto con sus pequeños genomas, pueden dar lugar a una enorme cantidad de variantes en el curso de una infección. Los genomas de los virus de ARN son muy adecuados porque son simples, pequeños, se replican de forma eficiente pero, sobre todo, porque poseen, genéticamente, una plasticidad que los hace enormemente adaptables a las variaciones ambientales. Esta plasticidad surge del hecho de estar constituidos, poblacionalmente, por una distribución muy heterogénea. El mecanismo de transcripción inversa les permite de hecho disfrutar de las ventajas del mundo adaptable de ARN y del mundo de ADN, con sus oportunidades para la recombinación y la regulación.

Los virus de ARN (y ello no es sorprendente) poseen, como consecuencia de sus propiedades de variación, una enorme capacidad de evolucionar. Su variabilidad excede, con mucho, la de sus huéspedes. No es sorprendente que el sistema inmunitario haya desarrollado con la evolución unas frecuencias de mutación extraordinariamente altas en las regiones del genoma implicadas en la generación de la variabilidad de anticuerpos. Esta variabilidad es de hecho del mismo orden de magnitud que la variabilidad de los virus.

13.3 Dinámica de replicación molecular

Consideremos un modelo de dinámica de moléculas que se autoreplican, y que incluye a la vez la posibilidad de errores en la replicación. La introducción de mutantes será, como veremos, una pieza clave en nuestra comprensión de la evolución de ciertos virus así como de las características peculiares de algunas enfermedades causadas por éstos, como es el caso del SIDA. El modelo que

2Para una discusión detallada de estas propiedades y sus implicaciones, véase el libro de S. Morse (ed): “Emerging Viruses” (1993).
Figura 13.2: Replicación y mutación en el modelo de cuasiespecies. Las tasas de replicación aparecen indicadas por $A_k$ y las de degradación por $D_k$. Los elementos de la matriz de mutación se indican por $Q_{ij}$.

Consideraremos un sistema formado por $N$ moléculas capaces de autoreplicarse (podemos pensar en polinucleótidos). Sea $x_k(t) \geq 0$ (con $k = 1, 2, ..., N$) la concentración de estas moléculas. Supondremos que poseen $v(k)$ unidades (en general, ya sean nucleótidos, símbolos, bits, etc.). Sea entonces el conjunto

$$ x \in \mathbb{R}^N_+ \equiv \{ x \in \mathbb{R}^N \mid x_i \geq 0 \ (i = 1, ..., N) \} $$

La probabilidad de que cada unidad sea reproducida adecuadamente se representa por $q$. La probabilidad de que la molécula completa sea reproducida sin ningún error (suponiendo que las unidades son independientes) será

$$ Q = q^{v(k)} $$

que se denomina a menudo factor de calidad. Si la tasa de replicación de cada molécula se indica por $A_k$ y la de destrucción por $D_k < 1$, y si suponemos que los errores en la replicación dan lugar a copias distintas pero dentro del conjunto considerado (como así sucede en la realidad), tendremos una ecuación dinámica para las concentraciones de cada molécula,

$$ \frac{dx_k}{dt} = (A_k Q_k - D_k)x_k + \sum_{l \neq k} W_{kl} x_l $$

donde $W_{kl}$ es la tasa a la cual se genera una unidad $k$ a partir de una $l$. Occasionalmente, se puede añadir un término de flujo que introduciría una dilución en el sistema. Puede interpretarse como un término independiente de la cinética de degradación y asociado al hecho de que “salen” del sistema a un ritmo (de flujo) constante. De forma general los valores de $W_{kl}$ dependerán de lo próximas que se hallen las secuencias consideradas (de cuántas mutaciones separen una de otra).
Para un par de secuencias dadas \( I_k \) y \( I_l \), la probabilidad de obtener una a partir de la otra vendrá dada por la frecuencia:
\[
W_{kl} = q^r \left( \frac{1 - q}{q} \right)^{d(k,l)}
\]
donde \( d(k, j) \) es la distancia de Hamming entre las secuencias, definida como el número de dígitos distintos entre ambas.

Observamos que el término de replicación \( A_k \) está multiplicado por \( Q_k \), como era de esperar. Los errores en la replicación de la secuencia \( l \)-ésima deben dar lugar a variantes \( k \neq l \), luego debemos esperar que se dé alguna relación formal entre \( A_l, Q_l \) y los valores \( \{W_{kl}\} \). Podemos comprobar que, efectivamente, se tiene
\[
\sum_{k \neq l} W_{kl} = A_l (1 - Q_l)
\]
Si denominamos \( W_{kk} \) (tasa metabólica) a la cantidad
\[
W_{kk} = (A_k Q_k - D_k)
\]
entonces podemos escribir las ecuaciones dinámicas en la forma
\[
\frac{dx_k}{dt} = \sum_i W_{ki} x_i
\]
con lo que las propiedades del sistema vendrán dadas por la matriz de coeficientes
\[
L_{\mu} = \begin{pmatrix}
A_1 Q_1 - D_1 & W_{12} & W_{13} & \cdots & W_{1n} \\
W_{21} & A_2 Q_2 - D_2 & W_{23} & \cdots & W_{2n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
W_{n1} & W_{n2} & W_{n3} & \cdots & A_n Q_n - D_n
\end{pmatrix}
\]
Si suponemos que las tasas de mutación son despreciables, y tomamos \( Q = 1 \), es fácil comprobar que la dinámica se reduce a un conjunto de soluciones de la forma
\[
x_k(t) = x_k(0) \exp \left( (A_k - D_k) t \right)
\]
lo que nos da \( x_k \to 0 \) para aquellas especies tales que \( A_k > D_k \), y cero en caso contrario. Está claro que, en ausencia de selección de ningún tipo, la clase más representada será la que se replique con mayor rapidez. En general, indicaremos por \( E_l = A_l - D_l \) la llamada productividad de la copia \( l \)-ésima.

Una vez que disponemos de la forma lineal de la interacción entre las moléculas, podemos plantearnos la introducción de las condiciones de selección sobre éstas. Una restricción obvia está en el ambiente, que podemos incluir en varias formas. Impondremos, siguiendo el modelo de Eigen, la restricción de población constante (CP), esto es, \( \sum x_k(t) = C \) o, lo que es lo mismo,
\[
\frac{d}{dt} \left( \sum_k x_k(t) \right) = 0
\]
Esta condición, que puede obtenerse experimentalmente en un reactor adecuado (pero que es poco plausible en la naturaleza), se obtiene introduciendo un término de flujo, \( \Phi_k \) de forma que tenemos ahora
\[
\frac{dx_k}{dt} = \sum_{l} W_{kl} x_l + \Phi_k
\]
y la restricción de población constante se verificará siempre que

$$\sum_k \Phi_k = - \sum_k \sum_l W_{kl}x_l$$

Si tomamos $\Phi_k$ como proporcional a la fracción de unidades de tipo $k$, esto es, si

$$\Phi_k = \left(\frac{x_k}{C}\right)\Phi_t$$

donde $\Phi_t$ indica el flujo total, $\Phi_t = \sum_k \Phi_k$ entonces este último puede determinarse a partir de la expresión anterior,

$$\Phi_t = - \sum_k \sum_l W_{kl}x_l = - \sum_k W_{kk}x_k - \sum_{k \neq l} W_{kl}x_l = - \sum_k (A_k Q_k - D_k)x_k - \sum_l (A_l Q_l - D_l)x_l$$

Tenemos así

$$\Phi_t = - \sum_k (A_k Q_k - D_k)x_k \equiv - E_k x_k \equiv - < E(t) > \sum_k x_k$$

de donde puede verse que

$$< E(t) > = e^{-1} \sum_k \sum_l W_{kl}x_l$$

Para simplificar, volvamos al caso $Q = 1$ y $W_{ij} = 0$. El sistema de partida es por lo tanto

$$\frac{dx_i}{dt} = (A_i - D_i)x_i - \Phi x_i$$

con lo que ahora la condición CP se obtendrá de

$$\sum_{i=1}^n \frac{dx_i}{dt} = \sum_{i=1}^n [(A_i - D_i)x_i - \Phi x_i] = 0$$

lo que nos da en este caso

$$\Phi = \frac{1}{C} \sum_j E_j x_j = < E(t) >$$

Las ecuaciones se escribirán entonces en la forma compacta

$$\frac{dx_i}{dt} = (E_i - < E >)x_i$$

Observemos que, en ausencia de mutación, si sólo una variante se halla presente (digamos la $i$-ésima) entonces la dinámica se reduce a una sola ecuación,

$$\frac{dx_i}{dt} = E_i \left(1 - \frac{x_i}{C}\right) x_i$$

y por lo tanto el punto fijo definido por $\Omega_i^* \equiv (0, 0, ..., C, ..., 0)$, con un valor no-nulo en la $i$-ésima componente, es globalmente estable. Este resultado es obvio si consideramos que la única secuencia presente puede replicarse sin competencia hasta llenar el espacio disponible.
En este sistema \( n \)-dimensional, tenemos por lo tanto \( n \) posibles estados estacionarios \( \{ \Omega_i^* \} \), que son

\[
\Omega_1^* \equiv (C, 0, ..., 0), \quad \Omega_2^* \equiv (0, C, ..., 0), \quad ..., \quad \Omega_n^* \equiv (0, 0, ..., C)
\]

Ahora, consideremos una situación más general, en la que estudiaremos la estabilidad del \( i \)-ésimo punto fijo \( n \)-dimensional, esto es, \( \Omega_i^* \equiv (0, 0, ..., C, ..., 0) \), que implica la presencia de la secuencia \( I_i \) y la ausencia de todas las restantes. Supongamos que perturbamos este estado. La matriz de Jacobi será

\[
L_{ij} = \begin{pmatrix}
E_1 - E_i & 0 & \cdots & 0 \\
0 & E_2 - E_i & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
-\alpha_i & -\alpha_j & \cdots & \alpha_n \\
0 & 0 & \cdots & E_n - \alpha_i
\end{pmatrix}
\]

la cual posee una diagonal principal de elementos \( E_j - E_i \) y una fila no nula \( \{-\alpha_1, \ldots, -\alpha_n\} \) (la \( i \)-ésima). Los elementos de esta matriz se han obtenido como sigue,

\[
L_{ij}^\mu(\Omega_i^*) = \left( \frac{\partial f_j^\mu}{\partial x_j} \right) |_{\Omega_i^*} = \frac{\partial}{\partial x_j} \left[ \left( E_i - \frac{1}{C} \sum_{r=1}^{n} E_r x_r \right) \right] |_{\Omega_i^*}
\]

lo que nos da,

\[
L_{ij}^\mu(\Omega_i^*) = \begin{cases} 
E_i - E_j & i = j \\
-\alpha_j & i \neq j
\end{cases}
\]

Estos valores se obtienen de la expresión general

\[
L_{ij}^\mu(\Omega_i^*) = \left[ \delta_{ij} - \frac{1}{C} \sum_{r=1}^{n} \delta_{jr} E_r x_r \right] |_{\Omega_i^*}
\]

El polinomio característico asociado es, como puede verse,

\[
P(\lambda) = \prod_{j=1}^{n} \left( E_j - E_i - \lambda \right) = 0
\]

y los valores propios son por lo tanto \( \lambda_j = E_j - E_i \).

Para una alguna secuencia \( i \) dada, tendremos que \( E_1 > E_j \), mientras que para las restantes \( E_j > E_i \), lo que hará que el sistema tienda asintóticamente hacia el punto \( \Omega_i^* \). Con independencia de la condición inicial (excepto en casos triviales) la población tenderá al punto fijo homogéneo \( \Omega_i^* \).

Un ejemplo de esta situación se muestra en la figura 13.3, en la que hemos representado el resultado de una simulación numérica del sistema anterior para \( n = 7 \) (los parámetros se indican en el pie de figura).

Vemos que, para un sistema inicial de \( n = 6 \) secuencias, cuyas \( E_i \) están ordenadas de menor a mayor, la secuencia 6 crece con rapidez, pero la introducción en \( t = 100 \) de la secuencia 7, hasta entonces ausente y con una \( E_7 \) mayor que las demás, se acaba imponiendo. También vemos la gráfica de \( <E(t)> \), que se estabiliza al principio para volver a aumentar posteriormente (como consecuencia de la introducción de \( I_7 \)). Las secuencias que se replican más lentamente (con menor eficiencia) son, en cada caso, eliminadas. La aparición de variantes más eficientes desestabiliza los posibles estados de equilibrio dando lugar a la aparición de nuevos equilibrios.
Figura 13.3: Soluciones del sistema dinámico \( \frac{dx_i}{dt} = (E_i - < E >)x_i \), con \( n = 7 \). Se tiene: \( E_1 = 0.4 \), \( E_2 = 0.6 \), \( E_3 = 0.55 \), \( E_4 = 0.95 \), \( E_5 = 1.25 \), \( E_6 = 1.5 \), \( E_7 = 2.2 \). Las condiciones iniciales son \( x_i(0) = 1 \) para \( i = 1, \ldots, 6 \) y \( x_7(0) = 0 \). La producción promedio crece inicialmente hasta alcanzar \( < E > \approx 1.5 \), pero la introducción de la secuencia 7, \( x_7(100) = 0.01 \), genera un cambio en el sistema, dándose una nueva estabilización en \( < E > = 2.2 \). Tenemos un proceso simple de selección.

13.4 Replicación con error: cuasiespecies

Consideremos ahora un caso más general, para el que supondremos (ahora razonablemente) que las tasas de mutación \( W_{ij} \) no son cero. Consideremos por lo tanto la generalización del problema anterior, cuyas ecuaciones serán ahora de la forma

\[
\frac{dx_i}{dt} = (E_i - < E >) + \sum_{j \neq i} W_{ij}x_j \quad i = 1, 2, \ldots, n
\]

Podemos también escribirlas como

\[
\frac{dx_i}{dt} = \sum_j \psi_{ij}x_j - < E > x_i
\]

siendo los coeficientes de la suma \( \psi_{ii} = A_iQ_i - D_i \) para \( i = j \) y \( \psi_{ij} = W_{ij} \) en otro caso. Estas ecuaciones son altamente no-lineales, pero el sistema puede ser resuelto si suponemos que los valores \( \psi_{ij} \) no dependen de las concentraciones ni del tiempo. Se introduce entonces un nuevo conjunto de ecuaciones

\[
x_i(t) = q_i(t) h(t)
\]

siendo \( q_i(t) \) una función definida por

\[
h(t) = \exp \left( - \int_0^t < E(\tau) > d\tau \right)
\]

Si tenemos en cuenta que
\[
\frac{dh(t)}{dt} = - < E(t) > \exp \left( - \int_0^t < E(\tau) > d\tau \right)
\]

la introducción de este cambio de variable nos permite obtener el sistema lineal

\[
\frac{dq_i}{dt} = \sum_j \psi_{ij} q_i
\]

o, en notación vectorial,

\[
\frac{dq}{dt} = \Psi q
\]

con lo que podremos estudiar la estabilidad del sistema (ahora lineal) estudiando los valores propios de la matriz \( \Psi \),

\[
P(\lambda) = \det [\Psi - \lambda I] = 0
\]

Ahora, las soluciones de la ecuación característica, \( \{\lambda_i\} \) (con \( i = 1, \ldots, n \)) estarán asociadas a un vector propio \( v_i \) que, en general, poseerá más de una componente no nula. Los vectores propios satisfacen la conocida relación

\[
\frac{dv}{dt} = D_{\mu} v
\]

con \( D_{\mu} \) la matriz diagonal

\[
D_{\mu} = \begin{pmatrix}
\lambda_1 & 0 & \ldots & 0 \\
0 & \lambda_2 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & \lambda_n
\end{pmatrix}
\]

Los vectores \( q \) y \( v \) cumplen las relaciones \( q = U_{\mu} v \) o, lo que es lo mismo, \( v = U_{\mu}^{-1} q \), donde las matrices \( U \) y su inversa \( U_{\mu}^{-1} \) vienen determinadas por las componentes de los vectores propios de \( \Psi \). La matriz \( D_{\mu} \) es diagonal y puede resolverse con facilidad, por lo que se tendrá una solución para \( q \) y para \( x \).

Si sumamos las ecuaciones lineales, tenemos

\[
\sum_i \frac{dx_i}{dt} = \sum_i \sum_j \psi_{ij} x_j - < E > \sum_i x_i
\]

que, bajo la restricción \( CP \) nos da

\[
< E > = \frac{\sum_i \sum_j \psi_{ij} x_j}{\sum_i q_i}
\]

Si llevamos a cabo la suma sobre \( dq_i/dt \), tendremos por otra parte que

\[
\sum_i \frac{dq_i}{dt} = \sum_i \sum_j \psi_{ij} q_j = < E(t) > \sum_i q_i
\]

lo cual equivale a
\[ <E(t)> = \frac{\partial}{\partial t} \ln \left[ \sum_i q_i(t) \right] \]

Integrando esta última ecuación, obtenemos

\[ \sum_i q_i(t) = \beta \exp \left( \int_0^t <E(\tau)> d\tau \right) \]

siendo \( \beta \) una constante de integración. Si comparamos esta ecuación con la transformación inicial, vemos que

\[ h(t) = \frac{\beta}{\sum_i q_i(t)} \]

La constante de integración queda, por otra parte, determinada por las condiciones iniciales. Para \( t = 0 \), tenemos \( h(0) = 1 \), luego \( \beta = C \) y podemos resolver el sistema a partir de las ecuaciones

\[ x_i(t) = h(t)q_i(t) = C \frac{q_i(t)}{\sum_j q_j(t)} \]

\[ \frac{dq_i}{dt} = \sum_j \psi_{ij} q_j \]

Resolveremos estas ecuaciones para obtener la dependencia de \( x_j(t) \) respecto del tiempo (y de las restantes variables). Partiremos de las ecuaciones para los valores propios, que podemos escribir en la forma

\[ \sum_j \psi_{ij} u_{jk} = \lambda_k u_{ik} \]

Los \( \lambda_k \) son los valores propios de la matriz \( \Psi \) y \( u_{ik} \) son las componentes de los vectores propios correspondientes. Los valores propios se obtienen de la ecuación característica

\[ \det(\psi_{ij} - \lambda \delta_{ij}) = 0 \]

Si los valores propios son no-degenerados (son todos distintos) la solución para \( q_i(t) \) es

\[ q_i(t) = \sum_k \alpha_k u_{ik} e^{\lambda_k t} \]

con lo que, finalmente, obtenemos la siguiente expresión para las concentraciones

\[ x_i(t) = C \frac{\sum_k \alpha_k u_{ik} e^{\lambda_k t}}{\sum_j \sum_k \alpha_k u_{ik} e^{\lambda_k t}} \]  

(13.4.1)

Las constantes \( \alpha_j \) se pueden calcular a partir de las condiciones iniciales.

Vemos que las mutaciones conllevan un resultado final importante. En términos biológicos, en lugar de la selección de la mejor copia, hemos obtenido una distribución formada por un conjunto de poblaciones moleculares. Cada una de estas combinaciones es una cuasiespecie (Eigen, 1993; Mikhaylov, 1990; Domingo, 1994), la cual juega en realidad el mismo papel que las especies moleculares antes estudia das. Lo que se selecciona, de hecho, es la cuasiespecie.

La expresión 13.4.1 es general, pero poco manejable para estudiar las implicaciones de la introducción de las mutaciones en nuestra descripción. Analizaremos dichas implicaciones en la siguiente sección.
Figura 13.4: Representación intuitiva del espacio de secuencias: dada una secuencia maestra, que posee la tasa de replicación mayor, las restantes secuencias se agruparán alrededor de dicha secuencia. Podemos encontrar distintas situaciones, que indicamos aquí de forma muy esquemática.

Si la tasa de mutación es muy baja, prácticamente no habrá secuencias mutantes, y por lo tanto veremos una única secuencia dominante, dada por la maestra (a). Por el contrario, si la tasa de mutación es enorme, virtualmente cualquier secuencia será posible y la distribución llenará el espacio (b). Para valores intermedios tendremos una secuencia maestra acompañada de una nube de mutantes: tenemos una cuasiespecie (c).

13.5 La catástrofe de error

Hemos visto anteriormente que la fiabilidad en la reproducción de la cadena de bits (o moléculas) viene dada por el factor de calidad \( Q = q^\nu \). Este factor decrece con el número de bits en forma exponencial. En algún punto, al ser \( \nu \) muy grande, las ecuaciones de evolución que hemos obtenido dan únicamente un decaimiento de las cuasiespecies con el tiempo. Este punto de transición, más allá del cual la información deja de estar conservada y sólo tendríamos secuencias sin ningún significado, se conoce como catástrofe de error. La capacidad de variación, dada implícitamente por la fidelidad de la copia, tiene así un límite impuesto por la información genética (si estamos hablando de cadenas de ADN o ARN) que debe mantenerse. Es evidente que, a medida que hacemos crecer el tamaño del genoma que debe replicarse, debemos aumentar la fidelidad de la copia si queremos que la información se conserve. A igualdad de tasa de error, un genoma complejo sufrirá mayores problemas que uno simple. De la anterior expresión se sigue que, para una fidelidad dada, tendremos una longitud límite más allá de la cual se da la pérdida de la cuasiespecie.

Podemos dar un argumento teórico simple (Nowak, 1991) que implica sólo una secuencia mutante. Si \( x_1 \) es la concentración de la secuencia maestra y \( x_2 \) la mutante (que suponemos única), y si despreciamos la probabilidad de mutación de \( x_2 \rightarrow x_1 \), tenemos las ecuaciones

\[
\frac{dx_1}{dt} = A_1 Q x_1
\]

\[
\frac{dx_2}{dt} = A_1 (1 - Q) x_1 + A_2 x_2
\]

Las poblaciones convergen al cociente

\[
\frac{x_1}{x_2} \rightarrow \Omega = \frac{A_1 Q - A_2}{A_1 (1 - Q)}
\]
Figura 13.5: La catástrofe de error: para cierto valor crítico de la tasa de error $p$, tiene lugar un cambio súbito en las propiedades de la distribución de secuencias mutantes (la secuencia es de $\nu = 50$ unidades).

De esta expresión se sigue que la secuencia maestra sólo puede mantenerse si $\Omega > 0$, esto es, si

$$Q > Q_c = \frac{A_2}{A_1}$$

Si recordamos que $Q = q^\nu$, el umbral de error se obtiene entonces para

$$q > q_c = \left[ \frac{A_2}{A_1} \right]^{1/m}$$

Por debajo de este valor, la secuencia maestra se pierde aunque posea la tasa de replicación mayor. Este resultado conlleva, como veremos, una relación de gran importancia entre la fidelidad de copia y la longitud de la secuencia.

Podemos obtener una expresión más general de este umbral volviendo a las ecuaciones para la dinámica de replicación con error,

$$\frac{dx_k}{dt} = (W_k - <E>)x_k + \sum_{l \neq k} W_{kl}x_l$$

(con $W_k = A_kQ_k - D_k$). Supongamos que las tasas de mutación son muy pequeñas de forma que, en una primera aproximación, podamos despreciar su efecto. Tenemos entonces un sistema de ecuaciones

$$\frac{dx_k}{dt} = (E_k - <E>)x_k$$

con $E_k = A_k - D_k$, que proporciona el conjunto de puntos fijos
\[ \Omega_1 \equiv (x_1^0, 0, ..., 0), \quad \Omega_2 \equiv (0, x_2^0, ..., 0), \quad ..., \quad \Omega_n \equiv (0, 0, ..., x_n^0) \]

y, de manera similar a lo que ya vimos antes, encontraremos para estos puntos fijos una matriz jacobiana de la forma

\[
L_m = \begin{pmatrix}
E_1 - E_i & 0 & \cdots & 0 \\
0 & E_2 - E_i & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
-E_1 & -E_2 & \cdots & -E_n \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & E_n - E_i
\end{pmatrix}
\]

para la cual podemos obtener un criterio de estabilidad completamente análogo al que hemos realizado con anterioridad para el sistema sin mutación. En este caso, puede demostrarse (véase Montero y Morán, 1992, para una excelente introducción y desarrollo) que la productividad media \(< E >^*\) en el equilibrio de selección viene dada por el valor selectivo de la secuencia maestra

\[
< E >^* = W_m
\]

Y a partir de este resultado podemos determinar de forma simple la concentración de la secuencia maestra una vez alcanzado el equilibrio. Definamos en primer lugar la denominada \textit{productividad media residual} \(< E_j \neq m >\), dada por el cociente

\[
< E_j \neq m > = \frac{\sum_{j \neq m} E_j x_j}{\sum_{j \neq m} x_j}
\]

Esta se puede interpretar como la productividad media asociada a las restantes secuencias.

Podemos escribir, de forma general

\[
< E_j \neq m > = \frac{\sum_{j=1} \frac{E_j x_j - E_m x_m}{x_j - x_m}}{\sum_{j=1} \frac{x_j - x_m}{x_m}} = \frac{\sum_{j=1} \frac{E_j x_j - E_m x_m}{x - x_m}}{C - x_m}
\]

de donde es posible obtener la relación entre la concentración de la secuencia maestra y la productividad media poblacional,

\[
\frac{x_m}{C} = \frac{< E > - < E_j \neq m >}{E_m - < E_j \neq m >}
\]

Para el estado de equilibrio,

\[
\frac{x_m}{C} = \frac{W_m - < E_j \neq m >}{E_m - < E_j \neq m >}
\]

lo cual define una relación de gran importancia. Recordemos que \(Q \leq 1\), luego en general \(x_m/C < 1\). Para \(Q = 1\) la copia maestra es la única finalmente presente, pero en caso contrario obtendremos una cola de mutantes que ocupan la fracción \(1 - x_m/C\) restante: tenemos una cuasiespecie. Pero aquí aparece además la catástrofe de error: existirá un valor crítico \(Q_c\) en la fidelidad de copia por debajo de la cual la concentración de \(x_m\) se hará cero, con lo que se perderá toda la información del sistema.
13.6 Virus y organización del sistema inmunitario

Los virus, al igual que cualquier otra entidad extraña al organismo, desencadenan la respuesta inmunitaria encaminada a su eliminación. El sistema inmunitario, que se encarga de reconocer de forma específica moléculas e invasores extraños (a la vez que reconoce lo propio) es un sistema de enorme especificidad. ¿Cómo está organizado? En la figura 13.6 vemos un esquema simplificado del mecanismo celular básico de acción. Tenemos dos tipos celulares que llevan a cabo un reconocimiento específico de los cuerpos extraños: los linfocitos T (células T, que maduran, esto es, se diferencian, en el timo) y los linfocitos o células B (que lo hacen en el bazo). Ambos tipos celulares llevan a cabo una actividad dominada por su respuesta específica. Una vez que una entidad extraña (digamos un virus) ha penetrado en el torrente sanguíneo, el primer paso previo al reconocimiento es la ingestión de los virus libres por parte de un tercer componente de la respuesta inmune: los macrófagos. En su interior, los virus son fragmentados en pequeñas piezas (péptidos) que, finalmente, son presentados en la superficie celular. Técnicamente, cada tipo particular de secuencia recibe el nombre de epitopo. Esta “presentación” corre a cargo de un grupo de moléculas de enorme importancia que se conocen con el nombre de Complejo Mayor de Histocompatibilidad (MHC). Una vez presentadas en el exterior de la membrana celular, los pequeños fragmentos pueden ser reconocidos por las células T como moléculas extrañas (Engelhard, 1994).

A continuación se da, finalmente, el reconocimiento. Por el torrente sanguíneo circulan millones de linfocitos T que, en la superficie de sus membranas, poseen una molécula que puede reconocer un tipo de epitopo y sólo uno. Así, eventualmente una de éstas células se encuentra con el epitopo y lo reconoce. Este reconocimiento tiene lugar a través de un contacto químico que requiere que las moléculas puedan acoplarse adecuadamente. Una vez esto ocurre, la respuesta se desencadena: la célula T se divide (formará un clon de células T específicas para el marcador reconocido) a la vez que segrega moléculas que estimulan la respuesta (véase Nossal, 1993, y otros artículos incluidos en la misma revista). El hecho de que el reconocimiento sea específico es una ventaja relativa, como veremos. Los virus, que poseen una enorme capacidad de variar, son un verdadero problema para las defensas inmunitarias.

¿Cuál es, de hecho, la situación de los retrovirus naturales en relación con el comportamiento de las cuasispecies que acabamos de describir? Está claro para nuestra intuición que, puesto que el reconocimiento del sistema inmunitario reviste una enorme especificidad, un virus poco variable será una presa fácil para dicho sistema. Una vez reconocido, el virus es destruido. También parece claro que, puesto que el virus muta, la nueva variante es vista básicamente como una entidad distinta, la variación será intrínsecamente valiosa para el virus. Ahora bien: el umbral de error asociado a la catástrofe de error nos indica que, más allá de esta frontera, el virus desaparecerá como entidad biológica. ¿En qué lugar dentro de este espacio de tasas de mutación se encuentran los virus de ARN?

Martin Nowak ha llevado a cabo una estimación teórica de la tasa óptima que le permitiría a un retrovirus como el VIH escapar de la respuesta inmunitaria (Nowak, 1990). Si la fidelidad de replicación (promedio) de la transcriptasa inversa es q, la probabilidad de replicación del genoma vírico íntegro sin error será \( Q = q^r \), como ya sabemos. Para el VIH, \( r \approx 10^4 \). Sea \( m \) el número de posiciones genómicas tales que una mutación en al menos uno de estos lugares permita la producción de un mutante de escape. Tal y como indica Nowak, \( m \) podría ser la longitud de la región más variable del genoma vírico (el dominio V3), pero el valor de \( m \) no será importante si \( m << n \), lo cual es razonable.

La probabilidad de obtener un mutante de escape, esto es, la probabilidad \( P \) de obtener un mutante sin errores en \( n - m \) lugares del genoma, y al menos un error en los \( m \) restantes es

\[
P(n, m, q) = q^m(1 - q^m)
\]
Figura 13.6: Organización básica del sistema inmunatorio. El objetivo final de la respuesta inmunaria es un antígeno, molécula de ordinario ajena al organismo, procedente de una bacteria o virus (u otro invasor). Las células presentadoras de antígeno son altamente especializadas e ingieren antígenos, los fragmentan en pequeñas subunidades (péptidos) y los unen a un tipo de moléculas de enorme importancia: las moléculas del complejo principal de histocompatibilidad (MH). Así unidas son “presentadas” en la superficie de estas células. Es ahora cuando los linfocitos T que posean los receptores específicos para el antígeno concreto pueden activarse, dividirse (formando un clon que se amplifica a través de la respuesta al antígeno) y activar otros elementos del sistema inmunario. Mediante la secreción de linfocinas la detección del elemento extraño se propaga a través del sistema dando lugar a la activación de las células B, que segregan anticuerpos.
Esta probabilidad poseerá un máximo para cierto valor \( q^{*} \) de \( q \) que puede obtenerse fácilmente de

\[
\frac{\partial P(n, m, q)}{\partial q} = (n - m)(1 - q^{m})q^{n - m - 1} - mq^{n - 1} = q^{n - m - 1}(n - m)(1 - q^{m} - mq^{m}) = 0
\]

lo que nos da un valor

\[
q^{*} = \left(1 - \frac{m}{n}\right)^{1/m}
\]

aproximable, para \( m << n \), por

\[
q^{*} \approx 1 - \frac{1}{n}
\]

La tasa de error \( \epsilon^{*} \) (mutación) óptima será por lo tanto

\[
\epsilon^{*} = 1 - q^{*} = \frac{1}{n}
\]

Para el virus del SIDA obtenemos \( \epsilon^{*} \approx 10^{-4} \).

Este resultado puede mejorarse teniendo en cuenta las posiciones neutras del genoma vírico, dando entonces \( \epsilon^{*} \approx 10^{-3} - 10^{-4} \) (Nowak, 1990; 1992). Puesto que la hipótesis subyacente en esta aproximación es que el sistema inmune actúa constantemente sobre la población vírica seleccionando mutantes de escape, esta tasa (en concordancia con las medidas experimentales) reflejaría la existencia de una capacidad óptima del virus para sobrevivir.

13.7 SIDA: en el umbral de diversidad

El 1990 Nowak, May y Anderson (NMA) propusieron un modelo matemático para la interacción entre el conjunto de virus HIV-1 que participan en la infección (la cuasispecie) y las células del sistema inmunitario humano (una exposición introductoria excelente puede encontrarse en Nowak y McMichael, 1995). En 1991 Nowak et al. expusieron un modelo algo más detallado. El modelo es simple, parcialmente tratable de forma analítica y sus propiedades y predicciones muy notables. Los autores del modelo partieron de las ideas básicas acerca del comienzo y progresión de la infección. Como veremos, pusieron el énfasis en dar una explicación coherente y contrastable del origen del largo y variable periodo de latencia del virus posterior a la infección y anterior a los síntomas del SIDA. Revisaremos en primer lugar los puntos a tener en cuenta para dar a continuación el modelo y algunas de sus predicciones.

Posteriormente a la infección por HIV-1, se observa típicamente un prolongado periodo de incubación (la fase latente) altamente variable, en el que se da una lenta pero continuada caída de la población de células CD4+ 3. Durante la infección, el HIV-1 ataca a las células del sistema inmune pero también a una amplia variedad de tipos celulares, incluyendo el tracto intestinal, el hígado o el cerebro (entre otros).

El modelo de Nowak et al. (1991) introduce la variabilidad genética del HIV-1 (y su naturaleza de cuasispecie) como elemento clave para la aparición de la inmunodeficiencia. Como veremos, cuando la población virica alcanza cierto umbral de diversidad, la respuesta inmunitaria se viene abajo. El virus, como ya hemos visto en una sección anterior, posee un alto nivel de mutabilidad (añadido a la ausencia de mecanismos de replicación) lo que lleva a producir del orden de un nucleotido erróneo por ciclo de replicación. La población resultante será una cuasispecie en la que

---

3Este es el conjunto de células dominante entre los linfocitos T. Debe su nombre al receptor de la membrana celular específico para el virus.
Figura 13.7: Características temporales de la infección del VIH. Esta dinámica tiene un tiempo característico muy prolongado, que se extiende a lo largo de un periodo de varios años, durante los cuales el paciente no presenta síntomas externos. Durante el periodo asintomático, las poblaciones celulares permanecen estables, eliminando efectivamente las poblaciones sucesivas de virus. En algún punto, sin embargo, el sistema inmune pierde la batalla frente a la acumulación de variantes viricas por encima del umbral de diversidad. Al final, se da el colapso del sistema inmunitario con la aparición del cuadro clínico que identifica el síndrome de inmunodeficiencia.

Veremos un conjunto de secuencias que pueden presentar grandes diferencias en tasas de replicación, proteínas de la cubierta, etc. La variabilidad de estas últimas les confiere la capacidad de escapar de los sucesivos ataques del sistema inmune.

El modelo NMA introduce tres propiedades distintivas del HIV-1 en el proceso de infección:

- La aparición continuada de nuevas variantes viricas (mutantes de escape) con el objetivo de burlar al sistema inmunitario.

- Respuesta inmunológica contra el virus, tanto en forma específica (a través de células CD4+ que actúan sobre una variante específica) como por una reacción general (cruzada) contra todas las variantes.

- Cada variante del virus puede infectar (y matar) a cualquier célula CD4+.

Estos puntos permiten obtener un modelo más o menos sofisticado (aunque simple) de la dinámica del sistema inmunitario bajo la infección virica. Partiremos, en nuestro análisis, de los modelos más simples que nos permitirán captar la idea esencial y posteriormente comprobaremos su validez general en un modelo más realista.

13.8 Dinámica básica y umbral de diversidad

Analizaremos el modelo del umbral de diversidad partiendo de la aproximación teórica más simple posible. Pese a que representa una simplificación exagerada de la respuesta inmunitaria, las conclusiones básicas son completamente generales. Volveremos a encontrarnos con ellas cuando aná-
licemos un modelo más detallado de la interacción VIH-sistema inmunitario. Esta aproximación fue propuesta por primera vez por Nowak y May en 1991. Las ecuaciones son

\[
\frac{d v_i}{dt} = v_i (r - px_i) , \quad i = 1,2,\ldots,n
\]

\[
\frac{d x_i}{dt} = k v_i - u v x_i , \quad i = 1,2,\ldots,n
\]

(13.8.1)

Las variables \( v_i \) y \( x_i \) denotan las densidades de la \( i \)-ésima variante del virus y la densidad de células del sistema inmunitario dirigidas contra dicha variante, respectivamente. En este modelo se asume que la tasa de replicación del virus \( r \) es la misma para todas las variantes. La respuesta inmune específica está representada por el término \( px_i \). La producción de células \( x_i \) se supone proporcional a la densidad de virus de la variante \( i \)-ésima, y la indicaremos por \( k v_i \). La respuesta inmunitaria, finalmente, queda dañada por la acción de los virus, e indicamos esta destrucción mediante el término \( uv x_i \), donde \( v = \sum_i v_i \) es el total de virus. En esta aproximación tomaremos como constantes los parámetros \( r, p, k \) y \( u \), que serán idénticos para todos los mutantes del virus.

La dinámica de las poblaciones totales de virus y del sistema inmunitario se obtiene sumando las ecuaciones anteriores. Obtenemos

\[
\frac{dv}{dt} = v \left( r - p \sum_{i=1}^{n} \frac{x_i}{v} \right)
\]

\[
\frac{dx}{dt} = k v - u v x
\]

Hemos utilizado \( x = \sum_i x_i \). La población viral quedará bajo el control del sistema inmune si

\[
\frac{r}{p} < \sum_{i=1}^{n} \frac{x_i}{v}
\]

De la ecuación inicial para la dinámica de la población \( x_i \), vemos que la respuesta de las células del sistema inmune dirigidas a la variante \( i \) tiende a un estado estacionario,

\[
x_i = \frac{k v_i}{u v}
\]

Si empleamos este resultado y lo introducimos en la ecuación para la población viral total, obtenemos

\[
\frac{dv}{dt} = v \left( r - \frac{k D}{u} \right)
\]

donde \( D = \sum (v_i/v)^2 \) es el llamado índice de Simpson, empleado en ecología como una medida (inversa) de la diversidad. Para una población viral idéntica (totalmente homogénea, de diversidad mínima) el índice de Simpson toma su valor máximo, \( D = 1 \). Si la población es completamente heterogénea, con iguales cantidades de virus de cada clase, se tiene \( D = 1/n \).

De la última expresión se sigue un resultado de enorme importancia: si el índice de Simpson decrece por debajo de cierto valor crítico, dado en el presente modelo por

\[
D_c = \frac{r u}{p k}
\]

entonces \( dv/dt > 0 \) y la población de virus escapa al control del sistema inmunitario. Si tomamos una distribución de virus uniforme (con \( D = 1/n \)) umbral de diversidad puede expresarse en
Figura 13.8: Interacción VIH-sistema inmunitario. El virus del SIDA se replica en las células infectadas, de las que salen nuevas partículas. La respuesta inmunitaria sufre una asimetría de gran importancia en el desarrollo del síndrome: los virus pueden infectar (y replicarse a costa de) cualquier célula con los receptores adecuados (linfocitos T4, por ejemplo) mientras que la respuesta es altamente específica.
Figura 13.9: Dinámica de la replicación vírica en el modelo de Nowak-May con parámetros homogéneos. Consideramos cinco variantes mutantes del virus \( n = 5 \), descritas por la ecuación 13.8.1. (a) Concentración total de virus. (b) Concentraciones de las cuatro variantes viricas. (c) Inversa del índice de Simpson \( D \) como medida de la diversidad de virus. Parámetros empleados: \( r = 1, p = 4.98, k = 1, u = 1 \), estos valores implican un umbral de diversidad \( n_c = 1/D_c = 4.98 \approx 5 \). Las condiciones iniciales son: \( v_1 = 0.1, v_2 = 0.001, v_3 = 0.0001, v_4 = 0.00001, v_5 = 0.0000001; x_i = 0 \).

En términos del número de variantes víricas presentes. La población de virus escapará de la respuesta inmune si

\[
n > n_c = \frac{p k}{ru}
\]

En la figura 13.9 se muestra un ejemplo de simulación numérica para un sistema formado por \( n = 5 \) variantes. Para los parámetros elegidos, el umbral de diversidad es \( n_c = 5 \).

### 13.9 \( D(v_1, ..., v_n) \) como función de Lyapunov

Una versión aún más simple puede obtenerse si suponemos (llevando a cabo una aproximación adiabática) que la dinámica de la respuesta inmune es rápida comparada con la dinámica de la población del virus. Esto equivale a decir que las constantes \( u \) y \( k \) son grandes comparadas con \( r \) y \( p \). Podemos entonces reemplazar los valores estacionarios de \( z_i \) en la ecuación para la dinámica de las variantes víricas

\[
\frac{dv_i}{dt} = v_i \left( r - \frac{k}{u} \rho_i \right), \quad i = 1, 2, ..., n
\]

donde \( \rho_i \equiv v_i/v \) indica la frecuencia de la variante del virus.

Bajo estas condiciones, podemos demostrar que el índice de Simpson \( D = D(v_1, ..., v_n) \) es una función de Lyapunov (capítulo 2). Para demostrar esta afirmación, notemos en primer lugar que

\[
\frac{\partial D}{\partial t} = 2 \sum_{i=1}^{n} \left[ \left( \frac{dv_i}{dt} \right) \frac{v_i}{v^2} - \left( \frac{dv_i}{dt} \right) \frac{v_i^2}{v^3} \right]
\]
Lo que nos da, sustituyendo las ecuaciones para \( dt_i/dt \) y \( dv/dt \), el siguiente resultado:

\[
\frac{\partial D}{\partial t} = 2\frac{pk}{u} \left[ s_2^2 - s_3 \right]
\]

Hemos utilizado la notación

\[ S_k = \sum_i \rho_i^k \]

Podemos demostrar que \( \partial_t D \leq 0 \) y que \( \partial_t D = 0 \) únicamente si

\[ \rho_i = 1/n \quad \forall i = 1, ..., n \]

La demostración es una consecuencia inmediata de la desigualdad de Jensen,

\[ f \left[ \sum_i \alpha_i \rho_i \right] \leq \sum_i \alpha_i f(\rho_i) \]

donde la igualdad se da si todos los \( \rho_i \) son idénticos. Aquí \( f(z) \) es una función estrictamente convexa definida sobre un intervalo dado \( \Omega \), y las constantes \( \alpha_i > 0 \) son arbitrarias y tales que \( \sum_i \alpha_i = 1 \) y \( \rho_i \in \Omega \). Si elegimos \( f(z) = z^2 \) y \( \alpha_i = \rho_i \), obtenemos directamente

\[ \left[ \sum_i \rho_i^2 \right]^2 \leq \sum_i \rho_i^2 \]

La igualdad se da obviamente en el caso de valores de \( \rho_i \) iguales.

### 13.10 SIDA y evolución de poblaciones CD4

A continuación, consideraremos una de las versiones más detalladas y más completas del modelo del umbral de diversidad. Aquí se introducen como variables la población total de linfocitos CD4 (el blanco del virus), las variantes víricas y las respuestas específicas e inespecíficas. Las ecuaciones del modelo son ahora

\[ \frac{dX}{dt} = \Lambda - \mu X - u v X - k v X - k' v X \]

\[ \frac{dv_i}{dt} = v_i \left( r(X + z) - s z - p x_i \right), \quad i = 1, 2, ..., n \]

\[ \frac{dx_i}{dt} = k v_i X - u v x_i, \quad i = 1, 2, ..., n \]

\[ \frac{dz}{dt} = k' v X - u v z \]

En este modelo \( X \) indica la población total de células CD4, incluyendo tanto las que están dirigidas contra el VIH como las que no lo están. Estas células CD4 se producen a una tasa constante \( \Lambda \) y son eliminadas de forma natural a una tasa \( \mu X \). El VIH puede eliminar células CD4 a una tasa \( u v X \).

Las células CD4, como sabemos, son activadas bajo la exposición a los antígenos víricos, dando lugar al ataque contra el virus. Indicaremos estas poblaciones celulares por \( x_i \) (población que responde específicamente) y \( z \) (población con respuesta no específica). Tendremos esta vez nuevos términos de interacción dados por \( k v X \) y \( k' v X \). En este modelo partimos de cierto número de variantes (que pueden ser un sólo mutante) y generaremos nuevas variantes al azar con cierta
probabilidad $p$. De esta forma entran en acción nuevas cepas del virus, que eventualmente serán eliminadas por el sistema. El VIH se replica en las células CD4. La tasa de replicación vírica será por lo tanto proporcional al número de células CD4 presentes, esto es, $X + x + z$.

Para calcular el umbral de diversidad asociado a este modelo, partiremos de la ecuación para las poblaciones víricas, e imponemos que

$$
\frac{dv_i}{dt} = v_i \left( r(X + x + z) - sz - px_i \right) < 0 \quad \forall i = 1, 2, ..., n
$$

Sumando para todas las variantes, la condición de supresión del virus vendrá dada por

$$
\sum_{i=1}^{n} \left( r(X + x + z) - sz - px_i \right) = n[r(X + x + z) - sz] - px \leq 0
$$

con lo que el valor crítico $n_c$ se obtendrá para la igualdad a cero. El umbral de diversidad será entonces

$$
n_c = \frac{px}{r(X + x + z) - sz}
$$

Para una densidad dada de virus, $v$, el número de células CD4 converge a un valor

$$
X \to \frac{\Lambda}{\mu + vu'}
$$

siendo $u' = u + k + k'$. De forma similar, tendremos $x \to kX/u$ y $z \to k'X/u$. Obtenemos entonces, sustituyendo,

$$
n_c = \frac{pk}{(ru' - sk')}
$$

Podemos simular este modelo por ordenador. El resultado obtenido es cualitativamente equivalente al antes estudiado. El modelo de Novak explica satisfactoriamente la gran variabilidad asociada a la fase asintomática, así como el largo periodo de latencia. Modelos recientes indican que, de hecho, la aparición de dichas propiedades sería la consecuencia natural de un virus cuya dinámica se hallará en las proximidades de una transición de fase (Solé y González-García, 1996).

### 13.11 Hiperciclos y evolución molecular

La existencia del umbral de error que caracteriza la catástrofe de error introduce una crisis de información en los sistemas que comparten las propiedades básicas antes expuestas. Revisando las premisas asumidas en la construcción del modelo de Eigen y sus derivados, debemos recordar que hemos asumido linealidad en las interacciones, si excluimos los términos no-lineales (que podemos eliminar a través de un cambio de variable no-lineal) que permitían introducir restricciones. Las abundancias de las distintas poblaciones dependían básicamente de la producción de moléculas, con tasas proporcionales a las abundancias (por lo tanto, lineales) y sin términos de orden superior.

Manfred Eigen demostró que un sistema autocatalítico como el que hemos explorado con anterioridad podía superar la crisis de información mediante interacciones de orden superior entre distintas moléculas que permitieran el mantenimiento de varias cuasispecies (para una discusión detallada de las distintas aproximaciones, véase Montero et al., 1992)

El modelo, denominado *hiperciclo*, resuelve, en cierta forma, un problema de ingeniería: el problema de cruzar cierto umbral de complejidad (el que establece el umbral de error). El hiperciclo
Figura 13.10: Estructura básica de la organización en la catálisis molecular definida por el modelo del hiperciclo.

Está formado por un círculo cerrado de reacciones catalíticas en el que cada especie molecular cataliza la autoreplicación de la siguiente (figura 13.10).

Además, los hiperciclos permiten estudiar cuestiones de gran profundidad, como es la evolución prebiótica y el origen de la información genética (Eigen et al., 1986). La idea es la siguiente: imaginemos, para una población de moléculas autoreplicativas, que tenemos varias secuencias maestras, con sus respectivas colas de mutantes y supongamos que, en ausencia de las demás, cada cuasiespecie es estable. El contenido en información de todas las cuasiespecies rebasa el máximo permitido para una sola secuencia maestra en el umbral de error. Para que el conjunto sea estable colectivamente y retenga toda la información, deben darse tres condiciones:

- Cada cuasiespecie debe mantener su estabilidad, esto es, cada secuencia maestra debe competir con sus mutantes de modo que no se acumulen errores.
- Las distintas secuencias maestras, cada una con su propio valor selectivo, deben tolerarse entre sí, a través de reacciones cruzadas adecuadas (cierto tipo de cooperación molecular).
- El conjunto debe mantener estables las poblaciones de cada uno de sus miembros y competir con otros conjuntos similares.

En esta sección consideraremos un modelo simple de hiperciclo y sus propiedades básicas (Eigen et al., 1986; Morán y Montero, 1992). Consideremos un conjunto de $n$ especies moleculares con concentraciones $x_i$. La evolución temporal de una de ellas quedará descrita por un sistema dinámico

\[
\frac{dx_i}{dt} = G_i(x_1, ..., x_n) - x_i \Phi(x_1, ..., x_n)
\]

donde una vez más $\Phi$ es el término de dilución y $G_i$ describe el tipo específico de interacciones entre moléculas. Supondremos que $x = \sum_i x_i = 1$, con lo que

\[
\Phi = \sum_{i=1}^{n} G_i
\]
lo que nos da una dinámica para la población total

\[ \frac{dx_i}{dt} = \sum_{j=1}^{n} G_{ij} \left[ 1 - \sum_{j=1}^{n} x_j \right] = 0 \]

De este resultado sabemos que la dinámica del sistema quedará confinada al simplex \( S_n \) definido por el plano

\[ S_n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n ; \ x_i \geq 0; \sum_j x_j = 1\} \]

En un hiperciclo, la especie 1 actúa catalíticamente sobre la autoreplicación de la 2, la 2 sobre la 3, y así sucesivamente, hasta completar el ciclo con la acción de la especie n sobre la 1. La función más simple para las interacciones será de la forma \( G_i = k_i x_i x_{i-1} \), con \( k_i > 0 \) constante. Tendremos por lo tanto un conjunto de ecuaciones dinámicas

\[ \frac{dx_i}{dt} = x_i \left[ k_i x_i - \Phi \right] \quad i = 1, \ldots, n \]

con

\[ \Phi(x_1, \ldots, x_n) = \sum_{j=1}^{n} k_j x_j x_{j-1} \]

Observemos que para este modelo se tiene que, si \( x_i(0) = 0 \), esta especie estará ausente en cualquier instante posterior: \( x_i(t) = 0 \), ya que no hay términos de mutación que permitan la creación de nuevas especies. La condición de equilibrio nos da \( \Phi = k_i x_{i-1} \), luego los posibles puntos fijos deberán verificar la propiedad \( k_1 x_n = k_2 x_{n-1} = \ldots = k_n x_{n-1} \).

Ahora se trata de analizar la estabilidad de este estado. Aunque un cálculo directo de los valores propios de la matriz de Jacobi puede ser muy difícil, un truco simple nos permitirá obtener un resultado bastante sorprendente: la estabilidad del punto fijo depende únicamente de \( n \) y no de las tasas \( k_i \). Introduzcamos el siguiente cambio de coordenadas

\[ y_i = \frac{k_{i+1} x_i}{\sum_{j=1}^{n} k_{j+1} x_j} \]

Que nos permite obtener el sistema dinámico

\[ \frac{dy_i}{dt} = y_i \left[ y_{i-1} - \sum_{j=1}^{n} y_j y_{j-1} \right] \Omega(y_1, \ldots, y_n) \]

Siendo \( \Omega(y_1, \ldots, y_n) \) una función estrictamente positiva sobre el simplex \( S_n \). Si omitimos este término, simplemente llevamos a cabo un cambio en la velocidad de las reacciones. Obtenemos así el sistema dinámico

\[ \frac{dy_i}{dt} = y_i \left[ y_{i-1} - \sum_{j=1}^{n} y_j y_{j-1} \right] \]

en el que todas las constantes de reacción se han normalizado a la unidad. El hiperciclo es por lo tanto simétrico, y el punto de equilibrio situado en el interior del simplex es ahora

\[ P^* = \left( \frac{1}{n}, \ldots, \frac{1}{n} \right) \]

La matriz Jacobiana es obviamente cíclica.
\[
\begin{pmatrix}
c_0 & c_1 & c_2 & \cdots & c_{n-1} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
c_{n-1} & c_0 & c_1 & \cdots & c_{n-2} \\
c_1 & c_2 & c_3 & \cdots & c_0
\end{pmatrix}
\]
y los valores propios para una matriz cíclica nos dan
\[
\lambda_j = \sum_{k=0}^{n-1} c_k \exp \left( \frac{2i\pi k j}{n} \right), \quad j = 0, 1, \ldots, n - 1
\]

Los hiperciclos han servido de marco de referencia teórico a multitud de estudios acerca de los orígenes de la vida sobre la Tierra. Se ha explorado su dinámica así como su valor para comprender el comportamiento de algunos virus de ARN (como el qβ, Eigen et al., 1986). Entre los problemas teóricos abiertos, destaca el hecho de que la aparición de parásitos moleculares (los cuales se replican a expensas de la ayuda de algún componente del hiperciclo, pero no contribuyen al mantenimiento de éste) puede dar lugar a su desaparición.

Otra dificultad no menos importante surge con la aparición de dinámicas complejas para hiperciclos formados por \( N \geq 4 \) especies, que incluyen oscilaciones, cuasiperiodicidad y caos. A medida que \( N \) crece, las fluctuaciones temporales de las poblaciones de secuencias pueden ser muy grandes en amplitud, de forma que una especie sometida a una pequeña perturbación externa podría desaparecer, con lo que el hiperciclo podría colapsar. Aunque algunos de estos problemas se podrían resolver en presencia de compartimentos que aislaran el hiperciclo del entorno, esta solución está lejos de ser trivial. El problema podría quedar resuelto de un modo notablemente más simple con la introducción del espacio en nuestra descripción de la dinámica (Boerlijst y Hogeweg, 1991; May, 1991).

Podemos construir modelos de la dinámica espaciotemporal de un hiperciclo empleando distintos modelos, como los autómatas celulares o los modelos de reacción-difusión. En este segundo caso, podemos escribir las ecuaciones en distintas formas (Cronhjort y Blomberg, 1994). Un modelo específico sería (Cronhjort, 1993)

\[
\frac{\partial x_i}{\partial t} = k_i x_{i-1}M - g_i X_i + D_i \nabla^2 x_i, \quad i = 1, 2, \ldots, N; \quad x_0 = x_N
\]

\[
\frac{\partial P}{\partial t} = k_P x_N PM - g_P P + D_P \nabla^2 P
\]

\[
\frac{\partial M}{\partial t} = a - bM - \sum_{i=1,\ldots,N,P} L_i k_i X_{i-1} M + D_m \nabla^2 M
\]

Puede probarse que, al menos en algunos casos, la aparición de las estructuras espaciales, en forma de ondas espirales, permite evitar la acción de los parásitos moleculares. En cualquier caso, éste es un campo en plena exploración, del que posiblemente surgirán nuevas y sorprendentes ideas acerca de los orígenes de la vida y la evolución molecular.

**Bibliografía**


Capítulo 14

Biodiversidad, Fragmentación del Hábitat y Extinción

Los ecosistemas, al igual que hemos visto en otros ejemplos anteriores, son sistemas altamente no-lineales. Su dinámica viene regida por leyes alejadas de la visión reduccionista clásica. En el mismo sentido, su respuesta a las perturbaciones externas es también fuertemente no-lineal. Puesto que los ecosistemas (tanto terrestres como acuáticos) están experimentando grandes cambios asociados a actividades humanas de distinto tipo e intensidad, deberíamos preguntarnos qué tipo de fenómenos no-lineales pueden aparecer cuando un ecosistema es perturbado en alguna forma. Esta perturbación puede ser, por ejemplo, un proceso de fragmentación. Este proceso implica la destrucción de porciones locales de un ecosistema extendido en un espacio dado. Se trata de un tipo de perturbación de gran interés: una buena parte de las actividades humanas en las que se lleva a cabo la explotación de ecosistemas destruye áreas de mayor o menor tamaño que, muy a menudo, son irre recuperables. Esta destrucción no tiene porqué ser muy generalizada. Tal vez sólo una fracción del área total (digamos el 30 por ciento) ha sido perturbada. En este caso nuestra intuición podría sugerir que el daño global ha sido bajo, en la medida en que la mayor parte del ecosistema se ha mantenido intacta. Pero como ocurre a menudo con los sistemas complejos, esta intuición se ve trascindida por los modelos y, también, por la realidad. Como veremos, la destrucción del hábitat no necesita ser completa para producir un daño importante (a veces enorme) al ecosistema. Dadas las implicaciones de estos resultados, los modelos que vamos a presentar obligan a una reflexión acerca de nuestro conocimiento de la dinámica y conservación de los ecosistemas.

14.1 Modelo de Levins

Para empezar (Tilman, 1994), consideremos el siguiente modelo simple debido a Levins, en el que analizaremos la dinámica de una sola especie que ocupa un hábitat formado por un conjunto de sub-áreas o “zonas”. Imaginemos que se trata de plantas (sin pérdida de generalidad) y que cada zona se halla ocupada por un único individuo. La muerte de un individuo dejará un espacio vacío que otro podrá colonizar. Nuestro modelo es implícito: no consideramos un área real.

Sea $p$ la fracción de zonas ocupadas, que llamaremos abundancia. El modelo de Levins se define mediante la ecuación para la dinámica de $p$,

$$\frac{dp}{dt} = cp(1-p) - mp$$

481
donde $c$ es la tasa de colonización y $m$ la tasa de extinción (mortalidad) local. Vemos que la tasa de ocupación de zonas vacías, $cp$, está multiplicada por la fracción de lugares libres, dando como resultado la tasa de producción de lugares colonizados. Este modelo es en realidad muy similar al ya conocido de crecimiento logístico continuo. Puede demostrarse sin dificultad que el punto fijo

$$p^* = 1 - \frac{m}{c}$$

es globalmente estable. Vemos claramente (como era de esperar) que la especie persistirá si y sólo si $c > m$.

La propiedad más interesante del modelo anterior es que esta especie nunca ocupará completamente el hábitat (esto es, nunca obtendremos $p^* = 1$) a menos que $m = 0$ o $c \rightarrow \infty$. Ninguna de estas dos condiciones es realista biológicamente.

### 14.2 Competencia entre dos especies

Acabamos de ver que, según el modelo anterior, una especie aislada no ocupará todo el espacio accesible, y dejará eventualmente huecos en el mismo. Una especie adicional que sea un competidor inferior podría, bajo ciertas circunstancias, subsistir ocupando estos lugares vacíos. Esta claro que para lograrlo deberá ser, en algún sentido, un buen colonizador. El modelo presente se basa en el planteamiento de Hastings (Hastings, 1978) y en su análisis por Tilman (Tilman, 1994).

Imaginemos estas dos especies cuyas interacciones estructuraremos de forma jerárquica. Indicaremos al competidor superior mediante el subíndice “1” y al competidor inferior por “2”.

Asumiremos que el competidor superior siempre desplaza al inferior cuando ambos llegan a una zona dada, pero que el competidor inferior no puede ni invadir ni desplazar al competidor superior en este dominio. Estas hipótesis permiten formular el siguiente modelo

$$\frac{dp_1}{dt} = c_1p_1(1 - p_1) - m_1p_1 = f_1$$

$$\frac{dp_2}{dt} = c_2p_2(1 - p_1 - p_2) - m_2p_2 - c_1p_1p_2 = f_2$$

El término $-c_1p_1p_2$ introduce la competencia asociada a la aparición de ambos competidores en un territorio dado. Aquí $p_1$ interfiere a $p_2$ con una intensidad dada por $c_1$. Los puntos fijos de la dinámica se obtienen a partir de $dp_i/dt = 0$. El punto crítico no trivial de interés viene dado por las poblaciones

$$p_1^* = 1 - \frac{m_1}{c_1}$$

$$p_2^* = [1 - p_1^*] - \frac{m_2 - c_1p_1^*}{c_2}$$

Podemos analizar la estabilidad de este punto siguiendo el procedimiento habitual. La matriz de Jacobi resultante tiene un elemento nulo (concretamente $\partial f_1/\partial p_1 = 0$) lo cual facilita el análisis de la estabilidad. La estabilidad de $p_1^*$ no es distinta de la que vimos en el apartado anterior, y vendrá dada por la condición $m_1 < c_1$. Para $p_2^*$, la condición de estabilidad es

$$c_2(1 - p_1^* - 2p_2^*) - m_2 - c_1p_1^* < 0$$

lo que nos da una condición para la tasa de colonización del competidor inferior, si su población debe ser no nula

$$c_2 > c_1\left(\frac{p_1^*}{1 - p_1^*} + \frac{m_2}{m_1}\right)$$
Figura 14.1: Dinámica de dos especies competitivas con estructura jerárquica. La especie 1 es el mejor competidor. (a) Parámetros empleados: $c_1 = 0.2$, $m_1 = 0.1$ y $c_2 = 0.8$, $m_2 = 0.1$. Ambas especies parten de una condición inicial $p_i = 0.01$. (b) En este caso se tienen: $c_1 = 0.6$, $m_1 = 0.4$ y $c_2 = 0.6$, $m_2 = 0.1$, con las mismas poblaciones iniciales.

En otros términos, el punto fijo $(p_1^*, p_2^*)$ será estable si se cumplen las condiciones

\[ c_1 > m_1 \]
\[ c_2 > c_1 \left( c_1 + m_2 - m_1 \right) \]

14.3 Competencia multiespecífica

A continuación generalizaremos el argumento anterior al caso de $n$ especies competitivas (Tilman, 1994). Si mantenemos los argumentos anteriores y ordenamos secuencialmente los competidores de mejor a peor (con los criterios ya indicados), tendremos un conjunto de ecuaciones diferenciales no-lineales

\[ \frac{dp_i}{dt} = f_i(p_1, p_2, ..., p_n) = c_i p_i \left( 1 - \sum_{j=1}^{i-1} p_j \right) - m_i p_i - \sum_{j=1}^{i-1} c_j p_j p_i \]

(donde $i = 1, ..., n$). Un competidor superior puede invadir una zona ocupada por un competidor inferior (cualquiera de ellos dentro de la jerarquía) y desplazarlo. Como antes, no existe simetría: un competidor inferior no puede invadir ni desplazar a un competidor superior.

En el estado de equilibrio, tenemos un conjunto de valores poblacionales $P^* = (p_1^*, ..., p_n^*)$, donde

\[ p_i^* = 1 - \frac{m_i}{c_i} - \sum_{j=1}^{i-1} \left( 1 + \frac{c_j}{c_i} \right) p_j^* \]
Figura 14.2: Dinámica de cuatro especies competidoras con estructura jerárquica. La especie 1 es el mejor competidor y la 4 el peor. Todas las especies poseen iguales tasas de colonización, \( c_i = 0.5; i = 1, \ldots, 4 \) pero distintas tasas de mortalidad: \( m_1 = 0.4, m_2 = 0.225, m_3 = 0.11, m_4 = 0.05 \). Todas las especies parten de la condición inicial \( p_i = 0.01 \).

Podemos calcular, secuencialmente, las condiciones de estabilidad de este sistema. La matriz de Jacobi del sistema está formada por los elementos

\[
L_{ij} = \left( \frac{\partial f_i}{\partial p_j} \right) \mathbf{p}.
\]

Dada la estructura jerárquica que hemos establecido, la matriz de Jacobi es tal que \( L_{ij} = 0 \) para \( j > i \), esto es, se trata de una matriz triangular. Los valores propios, que definen la estabilidad de cada población, se obtienen entonces de forma directa a partir de los elementos de la diagonal principal. Tenemos así para cada especie

\[
L_{ii} = c_i - m_i - 2c_i p_i - \sum_{j=1}^{i-1} p_j (c_j + c_i)
\]

A fin de evaluar \( L_{ii} \) en \( \mathbf{P}^* \) iremos calculando los valores correspondientes de las tasas de colonización de forma secuencial. Para \( c_1 \) obtenemos

\[
c_1 = \frac{m_1}{1 - p_1^*}
\]

lo cual nos permite calcular \( c_2 \)

\[
c_2 = \frac{p_1^* m_1 + (1 - p_1^*) m_2}{(1 - p_1^*)(1 - p_1^* - p_2^*)}
\]

y nos lleva a la expresión general

\[
c_i = \frac{\sum_{j=1}^{i-1} p_j^* m_j + (1 - \sum_{j=1}^{i-1} p_j^*) m_i}{(1 - \sum_{j=1}^{i-1} p_j^*) (1 - \sum_{j=1}^{i} p_j^*)}
\]
Para calcular los valores propios \( \lambda_i \), introducimos los valores obtenidos para \( c_i \) y tenemos, reordenando

\[
\lambda_i = -m_i \left( 1 - \sum_{j=1}^{i-1} p_j \right) + \sum_{j=1}^{i-1} m_j p_j \left( 1 - \sum_{j=1}^{i} p_j \right)
\]

Si el hábitat contiene una sola especie, el único valor propio será \( \lambda_1 = -m_1 p_1/(1 - p_1) \) que es siempre negativo puesto que \( m_1 \) y \( p_1 \) son positivos, y \( p_1 < 1 \). Por lo tanto, siempre que la especie se halle presente en el hábitat, este estado será estable. Si sustituimos en la condición de equilibrio el valor de \( p_1 = 1 - m_1/c_1 \), obtenemos \( \lambda_1 = m_1 - c_1 \). Para un hábitat con dos especies, los valores propios son

\[
\lambda_1 = \frac{-m_1 p_1}{(1 - p_1)} \quad y \quad \lambda_2 = \frac{-m_2 [m_1 p_1 + m_2 (1 - p_1)]}{(1 - p_1)(1 - p_1 - p_2)}
\]

Ambos serán negativos para cualquier valor de los parámetros biológicamente plausible (esto es, siempre que \( 0 < m_1, m_2 < 1 \) y que \( 0 < p_1; p_2 < 1 \), con \( p_1 + p_2 = 1 \)) y por lo tanto el punto de equilibrio es localmente estable.

En definitiva, como indica la ecuación para \( \lambda_1 \), todos los valores propios serán negativos, lo cual implica que el equilibrio de este sistema con múltiples especies es siempre localmente estable.

### 14.4 Destrucción del hábitat y coexistencia

Consideramos a continuación una primera aproximación al problema de la destrucción del hábitat y su efecto sobre la coexistencia de especies competidoras. Este modelo fue introducido en 1992 por Sean Nees y Robert May y posteriormente reanalizado por distintos autores (Dytham, 1994). En él se explora el efecto de la eliminación de zonas de un ecosistema (su destrucción completa) sobre dos especies competidoras que, en condiciones normales, coexisten entre sí.

Al igual que en los modelos previos, consideramos una estructura jerárquica: el competidor inferior (B) es incapaz de invadir un área ocupada por el competidor superior (A). Por otra parte, si A invade un área ocupada por B, entonces B es eliminado (se extingue) inmediatamente. Supongamos que las tasas de colonización de los competidores superior e inferior son \( c_a \) y \( c_b \), respectivamente, y que sus tasas de extinción local son a su vez \( e_a \) y \( e_b \).

Consideraremos un entorno consistente en un gran número de áreas (conectadas entre sí de alguna forma) de las cuales una fracción \( h \) (habitables) es capaz de sostener una población, aunque puede estar vacía en un momento dado. Si indicamos por \( x, y, z \) las frecuencias para las áreas vacías, áreas ocupadas únicamente por A y áreas ocupadas por B, respectivamente, las ecuaciones que describen la dinámica de este sistema son

\[
\frac{dx}{dt} = -c_a xy + e_a y - c_b xz + e_b z
\]

\[
\frac{dy}{dt} = c_a xy - e_a y + c_d zy
\]

\[
\frac{dz}{dt} = c_b xz - e_b z - c_d zy
\]

que constituye en realidad un sistema bidimensional, dado que se cumple la conservación \( x + y + z = h \). El diagrama del modelo y las transiciones posibles se indican en la figura 14.3.

Estas ecuaciones poseen una solución no-trivial definida por el punto fijo \( P^* = (x^*, y^*, z^*) \) dado por
Figura 14.3: Estructura básica del modelo de Nee-May.

\[ x^* = \frac{1}{c}(hc_a - e_a + e_b) \]

\[ y^* = h - \frac{e_a}{c_a} \]

\[ z^* = \frac{e_a(c_a + c_b)}{c_a c_b} - \frac{e_b}{c_b} - \frac{hc_a}{c_b} \]

La condición necesaria para que el competidor inferior persista será

\[ \frac{c_b}{e_b} > \frac{c_a}{e_a} \]

y puede obtenerse si el competidor inferior tiene una estrategia de vida pionera ("fugitiva"), caracterizada por una alta tasa de colonización. Pero el competidor inferior persistirá si su tasa de colonización es baja siempre y cuando tenga simultáneamente una tasa de extinción menor.

Podemos analizar el efecto de la eliminación de áreas habitables, de la reducción de \( h \). Se observa que

- Incrementa el número total de áreas ocupadas por el competidor inferior (esto es, crece la proporción de áreas, tanto destruidas como ocupadas por el competidor inferior).

- Se reduce el número de áreas ocupadas por el competidor superior.

- Disminuye el número de áreas vacías (no destruidas).

Si el competidor inferior es un colonizador superior, entonces el número total de áreas ocupadas disminuye. Si es un colonizador inferior, el número de áreas ocupadas crece. Cuando \( h \) cae por debajo de cierto valor crítico, definido por

\[ h_c = \frac{e_a}{c_a} \]

el competidor superior no puede seguir persistiendo. Más allá de este punto, sólo el competidor inferior persiste. Finalmente, si \( h \) decrece por debajo de un segundo valor crítico

\[ h'_c = \frac{e_b}{c_b} \]
Figura 14.4: (a) Efecto de la eliminación de áreas accesibles. A medida que nos desplazamos hacia la izquierda, pasando de un sistema con todas sus áreas intactas a una destrucción progresiva, vemos que A decrece hasta extinguirse, y durante este proceso el competidor inferior aumenta su frecuencia en el sistema. Más tarde la abundancia del segundo competidor también empieza a declinar hasta extinguirse. (b) Efecto directo de la eliminación de áreas sobre ambas especies.

el competidor inferior también desaparece. Para el caso en que ambas especies poseen la misma tasa de extinción y con un competidor inferior que coloniza mejor, en la figura 14.4 vemos el resultado de un experimento numérico.

En la misma figura indicamos un diagrama del efecto directo de la eliminación de áreas accesibles sobre ambas especies. El efecto directo del competidor superior sobre el inferior se da de dos formas: por una parte, hace crecer su tasa de extinción en el sistema y, por otra parte, hace disminuir el número de áreas colonizables por B. De aquí se desprende que la destrucción del hábitat (la disminución de h) beneficia al competidor inferior disminuyendo la frecuencia de áreas ocupadas por el competidor superior.

El resultado de este estudio, confirmado por modelos más elaborados (Dytham, 1994) es de gran importancia: no es preciso destruir todo el hábitat accesible para provocar la extinción de las especies presentes. La estructura no-lineal de las interacciones puede dar lugar a la extinción de especies mucho antes de que el hábitat haya sido seriamente dañado. Este problema se trata en las siguientes secciones bajo distintas aproximaciones.

14.5 Fragmentación y fenómenos críticos

Hemos visto que para algunos valores críticos de la fracción de hábitat destruido se dan cambios súbitos en las abundancias. El cambio más notable es, sin duda, la extinción de una especie cuando aún queda una fracción importante del hábitat sin destruir. ¿Cómo interpretar de forma consistente este fenómeno, claramente no-lineal? Un ingrediente de importancia, que analizamos en esta sección, será la introducción explícita del espacio. Hasta ahora, el espacio sólo había aparecido de forma implícita en las ecuaciones. Como se ha probado recientemente (Bascompte y Solé, 1996), el fenómeno subyacente es un fenómeno crítico, del tipo discutido en el capítulo 7 (Solé et al., 1996).

Como punto de partida, consideremos una red bidimensional. Como en otros modelos an-
Figura 14.5: (a) Tamaño de la zona más grande de puntos no destruidos conectados entre sí, en función de la fracción de hábitat destruido. Se ha calculado para una red de 30 × 30 puntos. Se indican tres ejemplos (b, c y d) de dicha red para tres puntos representativos. Para valores bajos de \( D \) todos los lugares no destruidos están básicamente conectados entre sí, formando una sola zona, pero para cierto valor crítico se produce una transición brusca, y la zona anterior se fragmenta, presentando entonces una reducción brusca de tamaño.

...el espacio está por lo tanto discretizado. Al principio, todos los lugares (puntos de la red) se consideran habitables. Por lo tanto, la especie o especies bajo consideración pueden sobrevivir en cualquier parte. En lo sucesivo, indicaremos por un cuadrado blanco dichos puntos. A continuación, simularemos el proceso de destrucción del hábitat: eligiendo al azar un número dado de puntos de la red, los destruimos. La destrucción es permanente y ninguna especie puede colonizar (ni sobrevivir en) dichos puntos. Los indicaremos mediante cuadrados negros. Imaginemos que llevamos a cabo esta destrucción progresiva, y que indicamos por \( D \) la fracción de zonas destruidas.

Podemos, para cada valor de \( D \), calcular sobre la red simulada el tamaño de la zona más grande formada por puntos no destruidos. En la figura 14.5 se muestra la variación experimentada por esta cantidad. Como vemos, no sigue un cambio monótono, sino que presenta un decaimiento inicial lineal (como esperariamos) seguido de una caída rápida. También se indican tres ejemplos de redes en tres situaciones distintas, para tres valores de \( D \). Como podemos imaginar (capítulo 7) existe un cambio de gran importancia asociado a la aparición de un fenómeno de percolación. Para \( D_c \approx 1 - p_c = 0.41 \), esto es, cuando el umbral de percolación definido por la fracción \( p_c = 0.59 \) de lugares no destruidos se alcanza, se da una transición abrupta. La segunda red mostrada ocupa precisamente dicha posición en el diagrama. La fracción destruida de hábitat es ahora suficiente para dar lugar al aislamiento de distintas subzonas entre sí. La zona grande inicial se ha dividido por lo tanto en varias, y este proceso sigue repitiéndose después de alcanzar \( D_c \).

Para separar el efecto cuantitativo de pérdida de hábitat del efecto cualitativo de fragmentación, podemos intentar definir un parámetro de orden \( \Omega(D) \) que caracterice aún mejor la transición. La idea básica, como sabemos, es la existencia de dos fases separadas por un punto crítico. En nuestro sistema, la primera, \( D < D_c \), está asociada a un hábitat básicamente "conservado" en la medida en que los puntos no destruidos estarán conectados entre sí. A pesar de que a medida que nos
Figura 14.6: Comportamiento del parámetro $\Omega(D)$. Se ha considerado el mismo sistema empleado en la figura anterior, promediado sobre cinco repeticiones. Se observa una rápida transición en el punto crítico $D_c \approx 0.41$.

Aproximemos a $D_c$ el tamaño de dicha zona decrece, este proceso es continuo y, cualitativamente, no hay cambios de importancia. Por el contrario, para $D > D_c$, el sistema está claramente formado por un conjunto de (muchas) zonas de pequeño tamaño separadas (aisladas) entre sí. Definiremos el parámetro de orden por:

$$\Omega(D) = \frac{S_m}{\sum_i \sum_j \Theta(i,j)}$$

Donde hemos indicado por $S_m$ el tamaño de la zona conectada de mayor superficie, y la suma del denominador está realizada sobre toda la red, siendo $\Theta(i,j) = 1$ si el punto $(i,j)$ no está destruido y cero en caso contrario. Así, $\Omega(D)$ caracteriza la densidad de zonas destruidas a la cual tiene lugar la fragmentación del hábitat. En la figura 14.6 mostramos la gráfica de $\Omega(D)$ para una red de $50 \times 50$ (promediando sobre cinco sistemas) y para distintos valores de $D$. El parámetro refleja dramáticamente la aparición del punto crítico. Para valores por debajo de $D_c$, el tamaño de la zona mayor y el número de zonas no destruidas es aproximadamente el mismo, puesto que prácticamente toda la red de puntos habitables permanecerá conectada. Al alcanzar el punto crítico, esta relación deja de ser válida de forma súbita, puesto que la zona de mayor tamaño se fragmenta, y $\Omega(D)$ muestra una caída brusca.

La implicación inmediata para un sistema ecológico que deba ser manipulado por el hombre de forma racional es que la destrucción de un hábitat natural puede poseer efectos distintos en función de qué porcentaje se ha destruido previamente. Una destrucción de un 10 por ciento de un hábitat no perturbado tiene un efecto (en principio) cuantitativo (véase, sin embargo, la última sección). Pero si lo que hacemos es destruir un cinco por ciento de un hábitat en el que la destrucción previa era del 35 por cien los efectos pueden ser dramáticos.

Ahora, podemos incorporar al modelo anterior una población que ocupe las zonas habitables y que se propague a puntos vecinos con cierta probabilidad. Incluiremos en el modelo además
una probabilidad de extinción. Supongamos que el punto de partida es la red anterior con todas sus zonas accesibles, no destruidas. Vamos a estudiar a continuación el comportamiento de la población que ocupa dichas zonas. Supondremos que consideramos una sola especie, que puede extinguirse (de hallarse presente en un punto dado de la red) con una probabilidad \( p_e \). Un punto no ocupado puede ser colonizado por uno cualquiera de sus puntos vecinos (4 en nuestro estudio) con probabilidad \( p_c \), siempre que dicho punto se halle ocupado. En está definición del modelo vemos que la dispersión (y por lo tanto la colonización) está muy limitada.

Partiendo de una condición inicial arbitraria, esta situación inicial lleva al sistema a evolucionar hacia un estado estacionario en el que cierta fracción del total de puntos se halla colonizada. Esta fracción es fácilmente calculable a partir de un modelo determinista. Sea \( V \) la fracción de puntos ocupados. El modelo de Levins (sección 1) nos da la fracción de dichos puntos en el equilibrio. Empleando nuestra notación,

\[
V^* = 1 - \frac{p_e}{p_c}
\]

Este mismo modelo nos podría proporcionar una aproximación al problema que consideraremos a continuación: el efecto de la destrucción de una fracción \( D \) de zonas. Tendríamos entonces

\[
\frac{dV}{dt} = p_c V (1 - V - D) - p_e V
\]

que posee un nuevo punto de equilibrio dependiente de \( D \)

\[
V^* = 1 - D - \frac{p_c}{p_e}
\]

lo cual define una línea recta. En el plano \((D, V)\), esta recta corta el eje horizontal en un punto crítico \( D_c \) para el que \( V = 0 \),

\[
D_c = 1 - \frac{p_c}{p_e}
\]
y que denominamos umbral de extinción. Una vez más, pese a que el modelo permite la existencia de una cantidad no nula de lugares no destruidos, se produce la extinción de la especie. Aunque podemos afinar en la descripción del problema, este resultado parece ser enormemente robusto. Siguiendo el estudio de Lande (1988) podemos introducir ciertas modificaciones en el modelo. Consideremos la ocupación del hábitat $P$ dada por

$$P = 1 - \frac{1 - K}{1 - D}, \quad D < K$$

$$P = 0, \quad D \geq K$$

donde $K$ es el denominado potencial demográfico de la población, definido como la fracción de áreas ocupadas cuando todas las áreas son accesibles (esto es, $K = P$ para $D = 0$). Si realizamos una representación similar a la anterior para distintos valores de $K$, encontramos también un umbral de extinción para $D_e = K$.

¿Qué ocurre en el caso representado por el modelo con espacio explícito? Para valores pequeños de destrucción del hábitat, el modelo matemático y el modelo espacialmente explícito básicamente coinciden (figura 14.8 (a)). La pérdida de áreas colonizables no afecta grandemente al sistema, que simplemente refleja la disminución lineal del área total. Sin embargo, para valores progresivamente crecientes, el número de lugares ocupados cae con mayor rapidez, de forma no-lineal. Este hecho conlleva una situación aún peor que la reflejada por el modelo de Levins con destrucción del hábitat. A la pérdida de áreas accesibles se unen los efectos locales derivados de la fragmentación del hábitat. Si estudiamos el equivalente de la representación de Lande, dibujando la proporción de zonas ocupadas $P$ en función del hábitat destruido, vemos otra vez el mismo efecto.
Figura 14.9: La deuda de la extinción: para valores pequeños de \( q \), el valor de \( E \) crece con rapidez con la destrucción del hábitat.

14.6 La deuda de la extinción

Analizemos, por último, un resultado teórico de enorme importancia, basado en los modelos estudiados en las primeras secciones de este capítulo. Este resultado (Tilman et al., 1994) tiene grandes consecuencias en nuestra comprensión del efecto de la destrucción del hábitat sobre la evolución posterior (y las extinciones) que se producirá en el seno del ecosistema.

Las ecuaciones de partida son una modificación simple de las ecuaciones de Tilman. Concretamente, emplearemos

\[
\frac{dp_i}{dt} = f_i(p_1, p_2, \ldots, p_n; D) = c_i p_i \left(1 - D - \sum_{j=1}^{i} p_j \right) - m_i p_i - \sum_{j=1}^{i-1} c_j p_i p_j
\]

Aquí \( D \) indica que cierta fracción del hábitat ha sido permanentemente destruida. Los restantes parámetros, como vemos, son idénticos a los que hemos empleado anteriormente, y las especies se han ordenado también jerárquicamente, del mejor al peor competidor. La colonización tiene lugar sólo entre las zonas no destruidas. Así, aunque la variación que hemos introducido parece mínima, en realidad tiene en cuenta un factor ecológico muy importante.

Las poblaciones de equilibrio son ahora \( P^* = (p_1^*(D), \ldots, p_n^*(D)) \), donde

\[
p_i^*(D) = 1 - D - \frac{m_i}{c_i} - \sum_{j=1}^{i-1} \left(1 + \frac{c_j}{c_i}\right) p_j^*
\]

y nuevamente podemos obtener sucesivamente los valores de equilibrio para cada competidor (partiendo del primero). Tenemos entonces que

\[
p_i^*(D) = 1 - D - \frac{m_i}{c_i}
\]
Esta especie se extinguirá de forma determinista si la fracción de hábitat destruido iguala a su abundancia en el hábitat no perturbado, puesto que \( p_i^* (D) = 0 \) cuando \( D \geq 1 - m_1 / c_1 \) y \( p_i^* = 1 - m_1 / c_1 \) para \( D = 0 \). Si el competidor dominante ocupa, por ejemplo, un 10 por ciento del hábitat virgen, y destruimos al azar un 10 por ciento de dicho hábitat, la intuición nos dice que, puesto que aún queda un 90 por ciento de hábitat libre, la supervivencia del competidor queda asegurada. Sin embargo, la destrucción al azar de un porcentaje del hábitat tiene como resultado el mismo efecto que si hubiéramos destruido precisamente las zonas habitadas por dicho competidor: su extinción. De hecho, y en general cuanto más rara es una especie, menor es la fracción de hábitat que debe destruirse para generar su extinción.

Para ajustar mejor la distribución de especies a las que se observan en la naturaleza (de tipo geométrico), de la forma

\[
p_i = q(1 - q)^{i - 1}
\]

(\( q \) es la abundancia del mejor competidor, y asumimos \( m_i = m, \forall i \)), las tasas de colonización necesarias deben ser

\[
c_i^* = \frac{m}{(1 - q)^{2i - 1}}
\]

Si introducimos estos valores en \( p_i^* \), obtenemos

\[
p_i^* = 1 - D - \frac{m_i}{(m_i / (1 - q)^{2i - 1})} - \sum_{j=1}^{i-1} \left(1 + \frac{(1 - q)^{2i-1}}{(1 - q)^{2j - 1}}\right)
\]

\[
= 1 - D - (1 - q)^{2i - 1} - (i - 1) - (1 - q)^{2i-1} \sum_{j=1}^{i-1} \frac{1}{(1 - q)^{2j - 1}}
\]

lo que nos da

\[
p_i^* = 1 - (1 - q)^{2i - 1}
\]

Sustituyendo este resultado y resolviendo para \( D \), obtenemos que la extinción de la \( i \)-ésima especie tendrá lugar si

\[
D_i \geq 1 - (1 - q)^{2i - 1}
\]

Observemos que se tiene la ordenación

\[
D_1 < D_2 < D_3 < \ldots < D_i < \ldots < D_n
\]

con lo que las especies se extinguirán siguiendo el orden de mejor a peor competidor a medida que el hábitat sea destruido.

El número de competidores superiores que acaba extinguiéndose debido a la destrucción permanente del hábitat (la denominada deuda de la extinción, \( E \)) crece con rapidez con la destrucción del hábitat. Partiendo de la desigualdad \( D_i \geq 1 - (1 - q)^{2i - 1} \) y considerando el caso límite dado por

\[
\frac{\log(D - 1)}{\log(1 - q)} = 2i^* - 1
\]

obtenemos, resolviendo esta ecuación para \( i \) e indicando el valor que representa el número de especies extinguidas, por \( E = i^* \),

\[
E = \frac{\log[(1 - D)(1 - q)]}{2 \log[1 - q]}
\]
(figura 14.9). De esta expresión se sigue que una pequeña destrucción adicional del hábitat pone en peligro muchas más especies de lo que esperaríamos utilizando nuestra intuición lineal. Como resultado de este análisis, concluimos que la destrucción del hábitat pone en peligro más competidores superiores en una selva tropical que en un bosque templado, y a los grandes y raros vertebrados por encima de los pequeños, típicamente más abundantes. Cuanto más destruido está el hábitat, más importante es la respuesta a la fragmentación subsiguiente. Estos resultados teóricos han sido confirmados por simulaciones espacialmente explícitas realizadas por Tilman y sus colaboradores (Tilman et al., 1993).

En conclusión, aunque es bien sabido que la destrucción del hábitat causa la extinción de especies (Wilson, 1993), estos resultados indican claramente que dicha destrucción puede dar lugar a la extinción selectiva de los mejores competidores. Estas especies son a menudo las que emplean los recursos con mayor eficiencia y ejercen un control importante sobre la regulación del ecosistema. En consecuencia, la deuda de la extinción asociada a la destrucción del hábitat puede tener un efecto aún más perjudicial sobre los ecosistemas. Los efectos no-lineales son, en este sentido, un factor a tener muy en cuenta en estudios futuros de evaluación del impacto de las actividades humanas sobre los ecosistemas.

**Bibliografía**


Capítulo 15

Neurodinámica

Sin duda alguna, el cerebro es la estructura más compleja que conocemos. Diez mil millones de
neuronas forman la materia prima de este “órgano de la mente”. A lo largo de la evolución, los
primeros sistemas neurales simples, constituidos por unas pocas neuronas, fueron modificándose,
aumentando en tamaño y en complejidad. Con la aparición de mayores estructuras, la percepción
del mundo externo fue acompañada por un procesamiento cada vez más refinado y a menudo
constituido por varias etapas. Al explorar en la escala celular, descubrimos un sistema de notable
sofisticación. Los diagramas de Ramón y Cajal, (figura 15.1) que no han perdido su vigencia, nos
dan una idea de la delicada trama de relaciones entre las neuronas de la corteza cerebral.

Como ya mencionábamos con anterioridad, el cerebro es un buen ejemplo de cómo la información
se procesa a distintas escalas. Las señales enviadas por una neurona a sus vecinas con las que
conecta (de decenas a miles) pueden propagarse a distintas escalas hasta dar lugar a patrones
macroscópicos. En la figura 15.2 vemos un ejemplo de una secuencia de propagación de ondas de
potencial registradas sobre la superficie de la cabeza de un ser humano. El registro temporal, que
caracteriza el llamado electroencefalograma, nos da una clara idea de la existencia de elementos de
coherencia junto a pautas desordenadas. Esta coherencia a gran escala nos recuerda nuevamente,
que incluso en sistemas de gran complejidad, donde los elementos son en sí mismas estructuras
complejas (como es el caso de las neuronas) pueden aparecer estructuras macroscópicas regulares
que implican la autoorganización coherente de centenares de miles de unidades básicas.

Estas series de potenciales, por sí mismas, no nos dicen demasiado, aunque cualitativamente
son enormemente sugerentes. Como sabemos, podemos llevar a cabo un análisis de estas señales
y estudiar cuántas dimensiones implican, si la dinámica es o no caótica, etc. Es una situación
especialmente apasionante: el cerebro lleva a cabo procesos internos que conllevan funciones tan
sorprendentes como el pensamiento. El sustrato material es, en último término, la materia neural.
Pero es mucho menos claro en qué forma y a qué escalas tienen lugar los procesos cognitivos.
Un resultado sorprendente y de enormes consecuencias fue la detección de atractores extraños
en la dinámica cerebral, en sistemas distintos y en situaciones normales y patológicas. Como
veremos, es posible determinar el número de dimensiones implicadas y éste es, muy a menudo,
sorprendentemente pequeño (del orden de 10³). Las consecuencias de este resultado son aún hoy
inimaginables, pero está claro que, en nuestra comprensión futura del cerebro, los fenómenos no-
lineales y el caos jugarán un papel importante.
15.1 Atractores extraños en sistemas neurales

En los capítulos dedicados al caos determinista se mencionó la posibilidad de que la dinámica de los sistemas neurales complejos, como el cerebro, sea de hecho caótica. Empleando las técnicas de estudio de series temporales introducidas en el capítulo sobre análisis de fenómenos caóticos, podemos comprobar esta posibilidad.

Uno de los primeros estudios en esta dirección fue llevado a cabo por Agnes Babloyantz y sus colaboradores de la Universidad Libre de Bruselas (Babloyantz y Destexhe, 1986; Baçar, 1990). El primero de estos estudios consistió en analizar un fenómeno patológico: un tipo de epilepsia llamado petit mal. En estos episodios, conocidos como ausencias, la actividad cerebral presenta una enorme coherencia, de forma que abandonamos las irregulares del cerebro sano para observar un patrón espaciotemporal enormemente coherente (figura 15.3). La reconstrucción del atractor da un objeto claramente ordenado para el que podemos calcular la dimensión fractal mediante el método de Grassberger-Proccacia. El resultado de este cálculo fue un valor de la dimensión de correlación de \( D_2 = \nu = 2.05 \pm 0.09 \) que, junto con la estimación del exponente máximo de Lyapunov, que arrojó un valor de \( \lambda_L = 2.9 \pm 0.6 \), dieron respaldo a la evidencia de atractores extraños en la dinámica a gran escala del cerebro. Estudios posteriores, basados en series mucho más largas, han aumentado esta dimensionalidad hasta valores de \( \nu = 3.1 \pm 0.2 \).

Existen otras patologías que revelan una coherencia muy superior a la del cerebro humano normal. Una de ellas es la enfermedad de Creutzfeld-Jacob, asociada a la destrucción de las neuronas debida a la infección por un virus lento. En la figura 15.4 se presenta un ejemplo de EEG de esta patología, así como el espectro de Fourier asociado. Vemos que, pese a la alta periodicidad de las oscilaciones, aparece una amplia banda de frecuencias implicadas, típica de un sistema caótico.

El análisis de estos datos (Babloyantz y Destexhe, 1988) dio como resultado una dimensión de correlación de \( \nu = 4.4 \pm 0.4 \). El estudio de estados normales de actividad ha revelado un amplio espectro de dimensiones que parecen estar fuertemente influenciadas por la actividad mental. Algunos de estos resultados se resumen en la siguiente tabla:
Figura 15.2: Dinámica cerebral: podemos obtener una primera imagen macroscópica de lo que tiene lugar en la corteza cerebral (la parte evolutivamente más nueva del cerebro) a través de la exploración de su dinámica mediante los electroencefalogramas. En la figura indicamos las series de potenciales obtenidas en distintas zonas.
Figura 15.3: Electroencefalograma (EEG) obtenido durante un episodio de epilepsia (*petit mal*).

Figura 15.4: Electroencefalograma (EEG) de un enfermo de Creutzfeld-Jacob. Se muestra también el espectro de Fourier.
Figura 15.5: Reconstrucción de un atractor correspondiente a la actividad cerebral de un paciente con demencia, en dos puntos distintos de la cabeza.

<table>
<thead>
<tr>
<th>Autor</th>
<th>Estado/Sistema</th>
<th>Resultados</th>
</tr>
</thead>
<tbody>
<tr>
<td>Babloyantz et al., (1985)</td>
<td>Sueño (estado 2)</td>
<td>$D_2 = 5.03$</td>
</tr>
<tr>
<td></td>
<td>Sueño (estado 4)</td>
<td>$D_2 = 4.0 - 4.4$</td>
</tr>
<tr>
<td></td>
<td>Despierto (alfa)</td>
<td>$D_2 = 6.1$</td>
</tr>
<tr>
<td>Rapp et al., (1986)</td>
<td>Ojos cerrados</td>
<td>$D_2 = 2.4 - 2.6$</td>
</tr>
<tr>
<td>Babloyantz et al., (1986)</td>
<td>Creutzfeld-Jacob</td>
<td>$D_2 = 3.7 - 5.4$</td>
</tr>
<tr>
<td></td>
<td>Epilepsia</td>
<td>$D_2 = 2.05$</td>
</tr>
</tbody>
</table>

Tabla I: Dimensiones fractales obtenidas por diversos autores sobre distintos estados de EEG en humanos.

Otros trabajos experimentales se han basado en el empleo de animales de experimentación en los que se han llevado a cabo medidas intracraneales. En estos estudios se ha podido realizar un análisis más pormenorizado de la dimensionalidad a distintas escalas de la estructura cerebral. En particular, destacaremos los estudios de Baçar (1990) empleando gatos para los que obtuvo, en tres estructuras cerebrales distintas, las dimensiones fractales de los atractores. Estos datos, especialmente fiabes, se resumen en la siguiente tabla.

<table>
<thead>
<tr>
<th>Sistema neural</th>
<th>Dimensión del atractor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cortex</td>
<td>$D_2 = 5.0 \pm 0.1$</td>
</tr>
<tr>
<td>Hipocampo</td>
<td>$D_2 = 4.0 \pm 0.07$</td>
</tr>
<tr>
<td>Formación reticular</td>
<td>$D_2 = 4.4 \pm 0.07$</td>
</tr>
</tbody>
</table>

Existen numerosos estudios en los que se ha llevado a cabo un análisis de la actividad del cerebro en busca de evidencias de caos. Algunos de ellos se basan en otro tipo de medidas, como las que emplean la detección de predictibilidad en sistemas deterministas (Gallez y Babloyantz, 1991; Blinowska y Malinowski, 1991). Otros emplean métodos alternativos de caracterización, entre los que destaca el empleo de varios canales simultáneos para obtener atractores reconstruidos (véase Rombouts et al., 1995, y las referencias incluidas).
Entre los problemas asociados a estas medidas, está el hecho de la no-estacionariedad de las series obtenidas. Esta no-estacionariedad, que volverá a analizarse algo más adelante (en relación con la existencia de transiciones de fase) podría ser debida a la existencia de múltiples atractores en la dinámica del cerebro (figura 15.5) y el sistema llevará a cabo (posiblemente) transiciones entre éstos, sin necesidad de recurrir a un origen externo. Si ello ocurre, la dimensionalidad del EEG irá en aumento con la longitud del registro empleado (como sucede muy a menudo). A mayor tiempo, mayor número de posibles transiciones. Este problema podría resolverse empleando métodos de medida de caos que permitieran utilizar los grados de libertad espaciales, mediante registros multicanal del EEG, al estilo del exponente de Lyapunov espaciotemporal discutido en el capítulo 10 (Solé y Bascompte, 1995). Eventualmente, el método permitirá además detectar transiciones en la dinámica y, por lo tanto, la existencia de atractores múltiples.

15.2 Sistemas neuronales y duplicación de periodo

Los sistemas neuronales simples pueden exhibir comportamientos complejos de distintos tipos, que incluyen transiciones hacia el caos por cualquiera de los tres tipos básicos de escenario previamente analizados. Los resultados experimentales y sus contrapartidas teóricas son muy numerosos. Una de las manipulaciones típicas consiste en la estimulación periódica de una célula nerviosa o pequeños colectivos de neuronas, perturbación que permite observar la transición hacia el caos en diversas formas.

Un modelo simple que exhibe un escenario de bifurcación con duplicación de periodo fue analizado por Ermentrout (1984), quien demostró la existencia de atractores extraños. El modelo parte del siguiente conjunto de ecuaciones:

\[
\begin{align*}
\frac{dP}{dt} &= -P + a_{11}S_1(P) - a_{21}S_2(I) + a_{31}S_3(E) \\
\frac{dI}{dt} &= -\kappa I + a_{12}S_1(P) \\
\frac{dE}{dt} &= -\rho E + a_{12}S_1(P)
\end{align*}
\]

cuya arquitectura está limitada a una pequeña red formada por una neurona piramidal, la actividad de la cual indicamos por \( P \), y un par de interneuronas indicadas por \( I \) y \( E \) con actividades inhibidora y excitadora, respectivamente. La función \( S_i(z) \) será una función sigmoide que introduce el tipo específico de interacción. Un ejemplo podría ser \( S_i(z) = 1/(1+\exp(-\beta z)) \), pero otros, cualitativamente similares, son igualmente válidos.

Ermentrout demostró que este modelo (y toda una familia de modelos relacionados) admite una reducción a la forma normal (cápítulo 4, apéndice), proporcionando un sistema reducido del tipo

\[
\begin{align*}
\frac{dx}{dt} &= y - \mu x \\
\frac{dy}{dt} &= z - \nu x \\
\frac{dz}{dt} &= qz^2 - \gamma x
\end{align*}
\]

el cual posee dos puntos de equilibrio, \( x_0 = (0, 0, 0) \) y \( x_1 = (q/\gamma, \mu q/\gamma, \nu q/\gamma) \). Ambos puntos son estables de forma simultánea, así que podemos concentrarnos en estudiar la estabilidad del
Figura 15.6: Atractores en el modelo de Ermentrout: (a) atractor periódico y (b) atractor caótico (figuras obtenidas por J. M. Sánchez Ferrer).

primeramente. Esta estabilidad, como es habitual, puede obtenerse a partir del polinomio característico asociado a la matriz Jacobiana del sistema en el punto fijo,

$$P(\lambda) = \lambda^3 + \mu \lambda^2 + \nu \lambda + \gamma = 0$$  \hspace{1cm} (15.2.1)

lo que requiere, para la estabilidad, la condición:

$$\mu > 0 \ ; \ \nu > 0 \ ; \ \gamma > 0$$  \hspace{1cm} (15.2.2)

$$\mu \nu - \gamma > 0$$  \hspace{1cm} (15.2.3)

La segunda condición (obtenida a partir del criterio de Routh) define la aparición de valores propios complejos con parte real positiva. Si la desigualdad 15.2.3 se invierte, podemos obtener una bifurcación de Hopf. Supongamos que $\gamma = \mu \nu$. Entonces el polinomio característico poseerá raíces $\lambda = -\mu$ y $\lambda_\pm = \pm i \sqrt{\nu}$, y si las desigualdades anteriores se cumplen, estas raíces son imaginarias. Para $|\gamma - \mu \nu|$ pequeños, puede darse un ciclo límite (siguiendo los criterios presentados en el capítulo 4). De hecho así es, $\gamma$ al aumentar $\gamma$ (con $\mu = 1$, $\nu = 2$, $q = 3$) obtenemos un escenario de Feigenbaum. En la figura 15.6 mostramos una solución periódica (para $\gamma = 2.5$) y un atractor extraño (para $\gamma = 3.5$). Estos atractores son de hecho del mismo tipo que los observados para el modelo de Rössler.

Vemos por lo tanto que un modelo simple, muy general, de un conjunto reducido de neuronas, puede mostrar un escenario de bifurcación complejo. Este primer modelo permite ver que un sistema neural puede generar de manera espontánea oscilaciones. En la siguiente sección veremos que un mecanismo de este tipo, en el que intervienen neuronas inhibidoras y excitadoras, posee la capacidad de generar oscilaciones globales coherentes.

### 15.3 Oscilaciones y caos en el cortex cerebral

El córtex visual de los vertebrados, emplazado en el lóbulo occipital, recibe información de las neuronas de la retina. La información de cada ojo ocupa zonas adyacentes (pero claramente
Figura 15.7: Estructura básica simplificada del modelo de una columna cortical.

diferenciadas) en forma de las denominadas columnas de dominancia ocular (Kandel et al., 1991). La organización del córtex visual está especialmente bien entendida. Este hecho, y la importancia que juega la visión en nuestra percepción del mundo, convierten a esta región del cerebro en una zona de gran interés.

Las neuronas de cada columna se encuentran conectadas entre sí, y existe también cierto grado de conectividad entre columnas adyacentes. Un hallazgo experimental de gran importancia fue obtenido por Gray y Singer en 1989 al analizar la respuesta del córtex visual de animales de experimentación frente a estímulos externos. La frecuencia de estas oscilaciones es de unos 40-60 Hz y ha sido también observada por Freeman en el bulbo olfatorio (Freeman, 1975; Freeman et al., 1988). Dado que esta coherencia está ligada al acto de la percepción, parece claro que existe un proceso dinámico subyacente a dicha actividad. ¿De dónde procede esta fuente de organización espaciotemporal? H. Schuster y P. Wagner propusieron en 1990 un modelo matemático de las oscilaciones de una columna cortical, modelizadas como subpoblaciones de neuronas excitadoras e inhibidoras, así como una extensión del modelo a la interacción entre columnas. La unidad básica del modelo se muestra en la figura 15.7, en la que se indica por I y E las poblaciones celulares de neuronas inhibidoras y excitadoras, respectivamente.

Este modelo de columna única puede analizarse en profundidad. Las neuronas excitadoras poseen sinapsis excitadoras (de valor positivo) y las inhibidoras poseen sinapsis inhibidoras (de valor negativo). Cada neurona se describe por medio de cierta variable que indica la intensidad del estado y que, en términos experimentales, se mide como tasas de disparo. Las indicaremos por

$$\{e_k(t)\}, \quad k = 1, \ldots, N_e$$

$$\{i_l(t)\}, \quad l = 1, \ldots, N_i$$

para neuronas activadoras e inhibidoras, respectivamente. Si suponemos (razonablemente) que dentro de la columna cada neurona está conectada con todas la demás, el modelo de Schuster-Wagner queda descrito por el siguiente conjunto de ecuaciones

$$\frac{de_k}{dt} = -e_k + \Phi \left[ a_e \left( \frac{1}{N_e} \sum_{l=1}^{N_e} w_{lk} e_l - \frac{1}{N_i} \sum_{l=1}^{N_i} v_{lk} i_l - \theta_k^e + p_k \right) \right], \quad k = 1, 2, \ldots, N_e$$

$$\frac{di_l}{dt} = -i_l + \Phi \left[ a_i \left( \frac{1}{N_e} \sum_{k=1}^{N_e} w_{lk} e_k - \frac{1}{N_i} \sum_{l=1}^{N_i} v_{lk} i_l - \theta_i^l \right) \right], \quad k = 1, 2, \ldots, N_i$$
La función \( \Phi(x) = [1 + e^{-x}]^{-1} \) modeliza la respuesta sigmoideal, y vemos que si tanto las conexiones como los estímulos exteriores \( p_k \) (que sólo afectan, en este modelo, a las neuronas excitadoras) se anulan, todas las actividades caen a cero. Ahora nuestro objetivo es explorar el comportamiento de este sistema y estudiar su dinámica global. Pese al carácter multidimensional de este sistema dinámico formado por \( N_e + N_i \) ecuaciones, la simulación numérica indica que la dinámica global se sincroniza con facilidad dando lugar a oscilaciones regulares globales. Un estudio formal del modelo permite ver cómo la dinámica global puede de hecho extraerse de las ecuaciones anteriores, dando un modelo de dos ecuaciones básicas para la dinámica de las poblaciones globales.

Definamos las siguientes cantidades promedio,

\[
\begin{align*}
    c_1 &= \frac{1}{N_e^2} \sum_{k=1}^{N_e} \sum_{l=1}^{N_e} u_{kl} ; \quad c_2 = \frac{1}{N_e N_i} \sum_{k=1}^{N_e} \sum_{l=1}^{N_i} v_{kl} \\
    c_3 &= \frac{1}{N_e N_i} \sum_{k=1}^{N_e} \sum_{l=1}^{N_i} w_{kl} ; \quad c_4 = \frac{1}{N_i^2} \sum_{k=1}^{N_i} \sum_{l=1}^{N_i} z_{kl}
\end{align*}
\]

y también las actividades promedio de las poblaciones celulares excitadora e inhibidora,

\[
\begin{align*}
    E(t) &= \frac{1}{N_e} \sum_{j=1}^{N_e} e_j(t) \\
    I(t) &= \frac{1}{N_i} \sum_{l=1}^{N_i} i_l(t)
\end{align*}
\]

A partir de estas cantidades, llevaremos a cabo un estudio del sistema de partida, y obtendremos las ecuaciones globales de campo medio. Sumando sobre \( k \) las ecuaciones para \( \frac{d e_k}{dt} \) y dividiendo por \( N_e \), obtenemos

\[
\frac{1}{N_e} \sum_k \frac{d e_k}{dt} = \frac{d E}{dt} = -E + \frac{1}{N_e} \sum_{k=1}^{N_e} \Phi(\alpha_k^e)
\]

siendo \( \Phi(\alpha_k^e) \) el término que incluye los promedios sobre la función sigmoideal. Supongamos que podemos escribir los valores de las conexiones, actividades locales y otras cantidades como desviaciones respecto de los valores promedio, esto es,

\[
\begin{align*}
    u_{kl} &= c_1 + \delta u_{kl} ; \quad v_{kl} = c_2 + \delta v_{kl} \\
    e_k &= E + \delta e_k ; \quad i_k = I + \delta i_k
\end{align*}
\]

y finalmente, supongamos que definimos \( \alpha_k^{e,i} = A^{e,i} + \delta \alpha_k^{e,i} \), donde

\[
A^{e,i} = \frac{1}{N_e,i} \sum_l a_l^{e,i}
\]

Desarrollando en serie de Taylor, obtenemos

\[
\frac{d E}{dt} = -E + \frac{1}{N_e} \sum_{k=1}^{N_e} \left[ \Phi(A^e) \frac{\partial \Phi(A^e)}{\partial A^e} \delta e_k + \frac{1}{2} \frac{\partial^2 \Phi(A^e)}{\partial A^e} (\delta e_k)^2 + \ldots \right]
\]

Los términos de orden inferior no incluyen las fluctuaciones, y nos dan la ecuación de campo medio.
Figura 15.8: (a) Modelo de columnas acopladas; (b) acoplamiento efectivo entre osciladores acoplados, en función del estímulo externo $P$. Podemos apreciar una transición que presenta histéresis.

\[
\frac{dE}{dt} = -E + \frac{1}{N_e} \sum_{k=1}^{N_e} \Phi\left(\frac{1}{N_e} \sum_{l} a'_{kl}\right) = -E + \frac{1}{N_e} \sum_{k=1}^{N_e} \Phi(A)^e
\]

Los términos de primer orden tienden a cero por tener una media cero (por construcción). Los términos de la forma

\[
\frac{1}{N} \sum_{l} \delta_{kl} \delta_{e_l}
\]

se tratan del modo siguiente. Supongamos que las cantidades $\delta_{kl}$ tienen probabilidades asociadas $\rho(\delta_{kl})$ independientes de $k$ y $l$. El teorema central del límite nos garantiza que

\[
\frac{1}{N} \sum_{l} \delta_{kl} = O\left(\frac{1}{\sqrt{N}}\right)
\]

lo que nos lleva a las ecuaciones macroscópicas (de campo medio) dadas por

\[
\frac{dE}{dt} = -E + \Phi\left[\alpha_e(c_1 E - c_2 I - \Theta^e + P)\right]
\]

\[
\frac{dI}{dt} = -I + \Phi\left[a_i(c_3 E - c_4 I - \Theta^i)\right]
\]

siendo ahora

\[
\Theta^e = \frac{1}{N_e} \theta^e_k \quad \Theta^i = \frac{1}{N_i} \theta^i_k \quad P = \frac{1}{N_e} p_k
\]

La simulación numérica demuestra que la aproximación dada por el modelo SW es enormemente apropiada para describir el comportamiento macroscópico de los colectivos neurales del tipo definido al comienzo (Schuster y Wagner, 1990).
Podemos a continuación estudiar el comportamiento de dos columnas acopladas. Queremos ver si, bajo condiciones apropiadas, dichas columnas presentarán sincronización, tal y como vemos en el córtex visual normal. Tenemos un nuevo conjunto de ecuaciones,

\[
\frac{dE_k}{dt} = -E_k + \Phi \left[ A^e_k - \eta_a U_l \right]
\]

\[
\frac{dI_k}{dt} = -E_k + \Phi \left[ A^i_k - \eta_a V_l \right]
\]

para \((k, l = 1, 2; k \neq l)\) donde hemos empleado la siguiente notación abreviada

\[
U_l = a_1 E_l - a_2 I_l
\]

\[
V_l = a_3 E_l - a_4 I_l
\]

\[
A^e_k = a_e(c_1 E_k - c_2 I_k - \Theta^e + P_k)
\]

\[
A^i_k = a_i(c_3 E_k - c_4 I_k - \Theta^i)
\]

Desarrollando a primer orden el acoplamiento \(\eta\), (para acoplamientos débiles), lo cual significa

\[
\eta < < \max\{c_j\}_{j=1,...,4}
\]

e introduciendo las fluctuaciones

\[
x_k = E_k - E_{k0} \propto r_k \cos(\phi_k)
\]

\[
y_k = I_k - I_{k0} \propto r_k \sin(\phi_k)
\]

respecto del punto de equilibrio inestable \((E_{k0}, I_{k0})\), obtenemos (Schuster y Wagner, 1990) las ecuaciones para las fases

\[
\frac{d\phi_1}{dt} = \omega_1 - K_{12} \sin(\phi_1 - \phi_2)
\]

\[
\frac{d\phi_2}{dt} = \omega_2 - K_{21} \sin(\phi_2 - \phi_1)
\]

que describen el comportamiento de las columnas en el estado activo (oscilatorio). Aquí \(\omega\) son las frecuencias de los osciladores. Las constantes \(K_{kl}\) (en general se tendrá que \(K_{kl} \neq K_{lk}\) son proporcionales a \(\eta\). El desarrollo de las ecuaciones de partida respecto de \(\eta\),

\[
\frac{dE_k}{dt} = -E_k + \Phi(A^e_k) + \eta_a \Phi(A^e_k)U_l
\]

y una expresión equivalente para \(I_k\), dan lugar a unos términos de acoplamiento

\[
L^e_{kl} = \eta_a \Phi(A^e_k)U_l
\]

\[
L^i_{kl} = \eta_a \Phi(A^i_k)V_l
\]

que dependen fuertemente de las entradas sobre los dos osciladores, como se indica en la figura 15.8 (b). En la que se ha calculado la constante de acoplamiento promediada sobre el tiempo,

\[
< K >_t = \frac{1}{T} \int_0^T K(t) dt
\]

Podemos comprobar que el acoplamiento es débil si uno de los osciladores se halla inactivo (pasivo) mientras que la actividad de ambos hace crecer este acoplamiento hasta un factor \(\approx 10\). Bajo estas condiciones, esperaremos sincronización de columnas corticales.
El mismo modelo permite obtener atractores extraños para amplios conjuntos de combinaciones de parámetros, como veremos en la próxima sección. De hecho, las redes de muchos elementos con conexiones lo bastante numerosas (y asimétricas) exhiben, típicamente, comportamientos complejos. Eventualmente (de hecho, con mucha frecuencia) estos comportamientos son notablemente coherentes y pueden identificarse como caos de baja dimensión. Los estudios basados en modelos teóricos son numerosísimos (Kürten y Clark, 1986; Aihara et al., 1990). Este tipo de modelos ha sido enormemente desarrollado y se dispone de un marco teórico sólido (si bien no completo) de la dinámica de sistemas de osciladores acoplados. Destaquemos, entre los muchos resultados experimentales y teóricos obtenidos, los estudios de Walter Freeman acerca de la dinámica del bulbo olfatorio. La importancia del trabajo de Freeman es doble, en la medida en que su modelo responde a un estudio experimental (llevado a cabo por el propio Freeman y sus colaboradores) que permitió un conocimiento detallado del córtex olfativo y de la existencia de caos determinista en su dinámica. El modelo ofrece no sólo una aproximación detallada y acertada del sistema original, sino que la interpretación dada por Freeman tiene grandes implicaciones para la comprensión de cuál es la función del caos en la percepción.

El bulbo olfativo es una estructura neural organizada de una forma relativamente simple y la única que no posee ninguna estación de paso intermedia en su conexión con el córtex cerebral. Evolutivamente es predominante en los mamíferos primitivos (para los cuales jugaba un papel básico en un mundo dominado por los grandes reptiles) y fue haciéndose menos prominente a medida que, en algunos grupos, el neocórtex fue ganando terreno. En el cerebro humano, el córtex olfativo se reduce a una muy pequeña porción del córtex cerebral, mientras que en algunos mamíferos insectívoros su presencia es predominante.

El modelo de Freeman se basa en un diagrama simplificado de los tipos celulares y de las conexiones presentes en el córtex olfativo real, y permite comprender detalladamente el efecto de distintas perturbaciones (inducibles experimentalmente) sobre la dinámica del cerebro. Los EEGs obtenidos a partir del modelo son enormemente similares a los reales, y Freeman ha lanzado una hipótesis, basada en sus experimentos, sobre el posible papel dinámico de los atractores extraños en el cerebro (Freeman, 1991). Dos resultados son de especial importancia. El primero: existe una fuerte evidencia en favor de un almacenamiento del recuerdo de los olores en forma de atractores periódicos y estructuras de actividad neural espaciales. El segundo: el caos podría ser empleado como mecanismo de atención. Yendo aún más allá, el cerebro generaría, según Freeman, atractores extraños que le permitirían “dar sentido al mundo” (Skarda y Freeman, 1987).

### 15.4 Control de caos en el cerebro

Si los sistemas neuronales pueden exhibir caos, deberíamos preguntarnos si estos atractores extraños neurales pueden ser manipulados experimentalmente para inducir todo tipo de bifurcaciones. Aún más interesante, está el hecho de que el caos puede ser controlado (capítulo 6) y por lo tanto no es descabellado que dicho control pueda obtenerse en sistemas neuronales reales. De hecho, en 1994, un artículo publicado en *Nature* se titulaba así: “Control de caos en el cerebro”. Los autores de dicho artículo (Schiff et al., 1994) llevaron a cabo un experimento pionero en el que, actuando sobre cortes de tejido cerebral mantenidos en vivo, demostraron la posibilidad de controlar la dinámica irregular observada para obtener comportamientos regulares.

La región del cerebro de la que se obtuvo el tejido es el hipocampo, una estructura especial protegida en el interior de la cabeza, situada en el centro de la masa encefálica. El hipocampo es un componente enormemente importante del cerebro, que participa en los procesos de memoria, en la orientación espacial (permite construir un “mapa interno” del entorno) y en el cual se inician muy a menudo las ondas que desencadenan los ataques epilépticos. La estructura de la red neural que forma el hipocampo es bien conocida, y existen modelos de esta región muy bien estudiados. En la
Figura 15.9: Sección del hipocampo, mostrando la red neural que lo define. Este tipo de secciones pueden ser mantenidas in vivo experimentalmente, con lo que podemos manipular una red neural real.

La figura 15.9 indicanos un esquema de la sección típica de éste. Las neuronas del hipocampo forman una red que en los cortes queda reducida a una arquitectura en una capa. Tenemos un conjunto de neuronas reducido, con una conectividad conocida y que en condiciones normales muestra, en los individuos maduros, dinámicas irregulares.

El grupo de Schiff logró, empleando el método OGY de control (Shinbrot et al., 1993), controlar estas oscilaciones irregulares. El método consiste en construir las aplicaciones de primer retorno en las que como variables se emplea el tiempo \( I_n \) entre intervalos de oscilación. Cada cierto tiempo el sistema muestra, en el registro con microelectrodos, un pico en el potencial, y se requiere cierto tiempo \( I_n \) para generar el siguiente máximo. Si dibujamos en un plano \( (I_n, I_{n+1}) \) los pares obtenidos sucesivamente, obtenemos para el hipocampo normal un conjunto aparentemente desordenado de puntos. Sin embargo, una inspección cuidadosa, analizando el comportamiento de las órbitas a lo largo del tiempo cerca de la recta bisectriz \( I_{n+1} = I_n \) (sobre la que, recordemos, se hallarían los posibles puntos fijos de una supuesta aplicación \( I_{n+1} = f_\mu(I_n) \)), demuestra que el comportamiento de este sistema es de hecho consistente con el de un sistema caótico con un punto de silla instable \( I^* \). En la figura 15.10 vemos una de estas gráficas, sobre la que indicamos algunos puntos consecutivos cuya secuencia indica claramente la existencia de dos variedades que atraviesan \( I^* \). A lo largo de la variedad estable \( W^e \) los puntos tenderán a acercarse, mientras que a lo largo de la variedad inestable \( W^s \) tenderán a alejarse.

Este resultado posee un enorme valor en si mismo: la dinámica irregular del hipocampo, que en una inspección superficial del diagrama anterior sugeriría estocasticidad más que caos, es claramente de origen determinista, con variedades estables e inestables bien definidas. Una vez caracterizadas estas variedades, podemos aplicar el método OGY de control. Experimentalmente, se espera a que el sistema se aproxime al punto fijo inestable \( I^* \) (dentro de un radio \( r \)) a lo largo de la dirección definida por la variedad estable. A continuación, llevamos a cabo una perturbación que modifica la dinámica de forma que el sistema regrese de nuevo a un entorno próximo a \( I^* \). De esta manera empleamos la estructura interna del sistema dinámico (el punto de silla inestable) para devolver al sistema a la variedad estable una y otra vez.

Entre otros resultados, los autores de este estudio también llevaron a cabo algo sorprendente:
lo que ellos llamaron 

Control de caos en redes neurales

Se han llevado a cabo estudios acerca de control de caos en redes neurales que emplean distintas estructuras neurales y métodos de control. Estos estudios (Sepulchre y Babloyantz, 1993; Solé y De la Prada, 1995; De la Prada y Solé, 1996) han permitido comprobar hasta qué punto es posible controlar el caos. Algunos autores (Babloyantz, 1993) han sugerido que el mecanismo de control podría emplearse en tareas cognitivas complejas.

En un estudio reciente (Solé y De la Prada, 1995; De la Prada y Solé, 1996) se ha comprobado que es posible controlar el caos en sistemas neurales formados por conjuntos reducidos de neuronas manipulando un sólo elemento. Consideremos el modelo de Schuster-Wagner introducido anterior-
Figura 15.11: (a) Red neural de dos capas obtenida a partir de los osciladores neuronales del modelo de Schuster-Wagner. Presentamos un sistema de cinco columnas conectadas entre sí localmente. Este sistema exhibe típicamente caos, dando lugar a un atractor extraño que se indica en (b) en forma de líneas de trazos. Aplicando el método GM a una de las neuronas excitadoras, podemos controlar distintas órbitas periódicas, como la que se indica en línea continua.

mente. Este modelo, del que nos hemos limitado a estudiar los atractores periódicos, puede generar comportamientos caóticos para ciertas combinaciones de parámetros que reduzcan la simetría de las interacciones. Un ejemplo de red compleja se indica en la figura 15.11 (a), formada por cinco columnas conectadas localmente (para un estudio detallado, incluyendo los parámetros empleados, véase De la Prida y Solé, 1996).

Utilizando el método GM, introducido en el capítulo 6, actuamos sobre una sola neurona (o grupo de neuronas) de la capa excitadora. Sea \( n \) el número de unidades en la capa excitadora, e indiquemos por \( E_n \in \{1, 2, \ldots, n\} \) el estado de la unidad sobre la que aplicamos el control. El conjunto de ecuaciones empleado es por lo tanto

\[
\frac{dE_k}{dt} = \left\{ -E_k + \Phi \left[ a_e(c_1E_k - c_2I_k - \Theta^e + P_k) - \eta a_c \left( a_1 \sum_{j < k} E_j - a_2 \sum_{j < k} I_j \right) \right] \right\} \Gamma_k
\]

\[
\frac{dI_k}{dt} = -I_k + \Phi \left[ a_i(c_3E_k - c_4I_k - \Theta^i) - \eta a_i \left( a_3 \sum_{j < k} E_j - a_4 \sum_{j < k} I_j \right) \right]
\]

donde \( \Gamma_k \) es el término de control, dado por

\[
\Gamma_k = 1 + \epsilon \delta_{k, \mu} \delta(t, \tau)
\]

Aquí, empleamos la notación \( \delta_{k, \mu} = 1 \) si \( k = \mu \) y cero en caso contrario. Análogamente, \( \delta(t, \tau) = 1 \) si \( t = \tau \equiv m \Delta t \) y cero en caso contrario. El método se reduce por lo tanto a perturbar la \( \mu \)-ésima neurona cada cierto multiplo prefijado del paso de integración numérica.
Figura 15.12: Anticontrol de caos: a partir de una red neural periódica, se puede inducir un comportamiento irregular perturbiendo débilmente una de las unidades excitadoras; indicamos sólo dos registros sobre una red con cuatro columnas.

Δt. La notación $< j >$ indica vecinos más cercanos y $q \in \mathbb{N}$ dependerá, en general, de la topología de la red. El control se puede obtener de forma efectiva sin dificultad (siempre que la red no sea muy grande y dependiendo de su topología). En la figura 15.11 (b) vemos un ejemplo de control de un atractor extraño. La órbita controlada se indica en línea continua, y el atractor subyacente en línea discontinua.

El mismo procedimiento sirve para lograr el anticontrol. Mediante perturbaciones muy pequeñas actuando sobre una sola unidad, podemos desestabilizar una órbita periódica y obtener comportamiento caótico. El hecho de que experimentalmente se haya obtenido la modificación de la actividad global del hipocampo mediante la manipulación de una sola neurona (Miles y Wong, 1983) abre interesantes posibilidades de aplicación práctica de los resultados previos.

### 15.6 Modelo de Hopfield

Los modelos anteriores han sido presentados con el objetivo de mostrar cómo pueden ser reproducidos los patrones dinámicos básicos observados en sistemas neurales naturales. Hasta ahora, no hemos presentado ningún modelo en el que el proceso natural de reconocimiento de patrones y la existencia de memoria asociativa sean parte integrante de la dinámica. El modelo de Hopfield es un ejemplo particularmente simple de una red discreta, con dinámica sencilla, capaz de aprender (almacenar un conjunto de patrones) y de recordar con memoria asociativa. Este modelo (Hopfield, 1982) representa de hecho el punto de partida de un gran esfuerzo teórico basado, muy especialmente, en herramientas procedentes de la física de sistemas desordenados (para un enfoque de este tipo, con numerosas referencias, véase Amit, 1989).
15.6.1 Modelo teórico: dinámica

Supondremos que la red está formada por un conjunto de $N$ unidades (que, genéricamente y abusando del lenguaje se conviene en llamar neuronas) cuyos estados indicaremos por
\[ S_i(t) \in \Sigma\{-1, +1\}, \quad i = 1, 2, \ldots, N \]
y cuya dinámica quedará especificada por el conjunto de ecuaciones
\[ S_i(t + 1) = \Phi \left[ \sum_{j=1}^{N} W_{ij} S_j - \theta_i \right] \]
El conjunto de entradas definido por la suma dentro del paréntesis,
\[ h_i(t) = \sum_{j=1}^{N} W_{ij} S_j \quad (15.6.1) \]
se denomina campo local de la neurona $i$-ésima. Observemos que cada elemento está, en principio, conectado con todos los demás elementos. La conectividad es global. La función $\Phi(z)$ se toma aquí como la función signo, esto es, $\Phi(z) = \text{sgn}(z) = +1$ si $z \geq 0$ y $-1$ en caso contrario. Hemos indicado por $\theta_i$, cierto umbral que, en lo que sigue, y sin pérdida de generalidad, tomaremos como $\theta_i = 0$. La dinámica se reduce por lo tanto a
\[ S_i(t + 1) = \Phi \left[ \sum_{j=1}^{N} W_{ij} S_j \right] \quad (15.6.2) \]

Añadiremos que, tal y como se define aquí, la dinámica sería de tipo síncrona, esto es, todos los elementos perciben simultáneamente sus campos locales y cambian de estado simultáneamente en la misma iteración. En sistemas naturales, cierto grado de síncronía es más que clara (aunque no total). En el modelo de Hopfield emplearemos, por el contrario, una dinámica asíncrona, de forma que, en cada iteración, modificaremos el estado de sólo una neurona, para lo que habremos calculado el campo local previamente. Más específicamente, podemos implementar una dinámica asíncrona de dos formas:

- En cada paso de tiempo, seleccionamos al azar una unidad $i$ y aplicamos la dinámica definida por la ecuación 15.6.2.

- Dejamos que cada unidad pueda calcular su nuevo estado de acuerdo con cierta probabilidad por unidad de tiempo.

Ambas posibilidades son de hecho equivalentes.

En lo que sigue, supondremos que las memorias que se desea almacenar están formadas por secuencias de $N$ valores binarios, cuyas componentes indicaremos por
\[ \xi_i(t) \in \Sigma\{-1, +1\}, \quad i = 1, 2, \ldots, N \]
y serán por tanto de la forma
\[ \xi = (\xi_1, \xi_2, \ldots, \xi_N) \]
además, asumiremos que los valores $\xi_j$ se obtienen aleatoriamente, siguiendo cierta distribución de probabilidad.
Vemos que los elementos de la dinámica básica quedan definidos, excepto en lo que se refiere a las conexiones. La matriz $W_{ij}$ es, sin embargo, el elemento clave. En función de cómo definamos las conexiones, tendremos o no capacidad de almacenar la información adecuadamente. Como veremos, existe un procedimiento simple para almacenar información basado en una adecuada elección de las conexiones en función de los valores de entrada suministrados por los patrones $\xi$.

Para entender la forma de elegir los $W_{ij}$, consideraremos un caso muy simple, en el que la red debe almacenar sólo un patrón. Para que dicha memoria sea un estado estacionario de la dinámica (un punto fijo), debe cumplirse, como sabemos, la condición

$$sgn\left(\sum_{j=1}^{N} W_{ij} \xi_j\right) = \xi_i$$

Podemos ver que esto tiene lugar en particular si $W_{ij} \propto \xi_i \xi_j$, dado que $\xi_i^2 = 1$, y por lo tanto $sgn(\sum_j \xi_i \xi_j \xi_j) = sgn(\sum_j \xi_i) = \xi_i$.

Escogeremos en nuestro caso una conexión normalizada por el número de neuronas $N$

$$W_{ij} = \frac{1}{N} \xi_i \xi_j$$

Como vemos, de esta definición se sigue que las conexiones son simétricas, y habitualmente tomaremos $W_{ii} = 0$.

La estabilidad de este punto fijo viene claramente afectada por la matriz de conexiones. Notemos que, incluso si cierta cantidad de bits del patrón original están erróneamente representados, la suma sobre el total puede permitir que, aún así, el punto fijo retenga su estabilidad. En consecuencia, una configuración inicial que contenga algunos bits erróneos rápidamente será reemplazada, a través de la dinámica, por una colección de bits correctos: el estado inicial es "atraído" hacia la memoria $\xi$, que es un atractor del sistema\(^1\). Vemos por lo tanto, por primera vez, que el reconocimiento de un patrón puede asociarse a la existencia de atractores en un sistema dinámico.

El caso realmente interesante aparece cuando consideramos un conjunto de $p$ patrones a almacenar, esto es, $\{\xi^\mu\}$, con $\mu = 1, 2, ..., p$. Tenemos por lo tanto un conjunto de $p$ vectores $N$-dimensionales,

$$\xi^1 = (\xi^1_1, \xi^1_2, ..., \xi^1_N), \quad \xi^2 = (\xi^2_1, \xi^2_2, ..., \xi^2_N), \quad ..., \quad \xi^p = (\xi^p_1, \xi^p_2, ..., \xi^p_N)$$

Puede probarse que la forma anterior de elegir las conexiones puede emplearse de forma eficiente en este caso general. Lo que tendremos por lo tanto será una superposición de productos cruzados de la forma

$$W_{ij} = \frac{1}{N} \sum_{\mu=1}^{p} \xi^\mu_i \xi^\mu_j$$

Esta expresión se conoce como regla de Hebb. Su origen tiene cierta base neurofisiológica. La idea básica es fácil de entender: tomemos dos neuronas, digamos la $i$ y la $j$-ésima. Supongamos que estas neuronas reciben, durante la fase de aprendizaje, ciertas señales de valores $\xi^\mu_i$ y $\xi^\mu_j$, cuando "mostramos" a la red el patrón $k$-ésimo. La regla de Hebb hace que los estímulos correlacionados actúen reforzando la conexión entre neuronas, esto es, el valor de $W_{ij}$ que suma $1/N$ cada vez que dicha correlación se da. Por el contrario, las anticorrelaciones actúan debilitando el valor de la conexión. Observemos que esta regla tiene carácter local: sólo hacemos referencia a la información que llega a pares de neuronas conectadas entre sí, y no introducimos, a este nivel, más información

\(^1\)De hecho, este sistema posee dos atractores: el estado complementario a $\xi$, en el que invertimos el signo de todos los elementos, también verifica las mismas condiciones.
de una escala mayor. En este sentido, las propiedades globales del patrón no forman parte de las reglas locales: la capacidad de almacenar y recuperar la información en la red de Hopfield surge estrictamente de una propiedad global, asociada a la matriz de conectividad.

Ahora podemos explorar la estabilidad de la red bajo condiciones mucho más generales. Tomemos un patrón dado, digamos el \( \nu \)-ésimo. La condición de punto fijo estará ahora dada por

\[
sgn(h_i^\nu) = sgn \left[ \sum_{j=1}^{N} W_{ij} \xi_j^\nu \right] = \xi_i^\nu
\]

para todos los valores de \( i = 1, ..., N \). Empleando la definición de la regla de Hebb para las conexiones, el campo local se escribirá

\[
h_i^\nu = \sum_{j=1}^{N} W_{ij} \xi_j^\nu = \frac{1}{N} \sum_{\mu=1}^{p} \sum_{j=1}^{N} \xi_i^\mu \xi_j^\nu \xi_j^\mu \xi_j^\nu
\]

y se puede separar en dos términos, uno para \( \mu = \nu \) y otro para los demás valores de \( \mu \), con lo que tenemos

\[
h_i^\nu = \xi_i^\nu + \frac{1}{N} \sum_{\mu \neq \nu} \sum_{j=1}^{N} \xi_i^\mu \xi_j^\nu \xi_j^\mu \xi_j^\nu
\]

expresión a partir de la cual exploraremos la estabilidad. Notemos que, si el segundo término de la derecha fuera cero, podríamos garantizar inmediatamente la estabilidad del patrón. Pero de hecho ello ocurrirá también si el segundo término es lo bastante pequeño: si su magnitud es menor que 1, no podrá modificar el signo del campo local \( h_i^\nu \) y la condición de estabilidad se verificará. El término cruzado, esto es,

\[
\Omega_\nu = \frac{1}{N} \sum_{\mu \neq \nu} \sum_{j=1}^{N} \xi_i^\mu \xi_j^\nu \xi_j^\mu \xi_j^\nu
\]

será menor que 1 en muchos casos de interés si el número de patrones \( p \) es reducido. A continuación veremos cómo disponer de una estimación del número de patrones estables que la red puede almacenar en relación con el número total de neuronas \( N \), lo que se denomina capacidad de la red, \( \alpha = p/N \).

Consideremos la siguiente cantidad,

\[
\Gamma_\nu = -\xi_i^\nu \frac{1}{N} \sum_{j=1}^{N} \sum_{\mu \neq \nu} \xi_j^\mu \xi_j^\nu \xi_j^\mu \xi_j^\nu
\]

de hecho \(-\xi_i^\nu\) veces el término cruzado \( \Omega_\nu \). Si \( \Gamma_\nu \) es negativo, el término cruzado tendrá el mismo signo que la componente del patrón, \( \xi_i^\nu \), y por lo tanto no habrá problemas de inestabilidad. Pero si \( \Gamma_\nu \) es positivo y mayor que 1, cambiará el signo del campo local y el patrón \( \nu \)-ésimo se hará inestable.

\( \Gamma_\nu \) depende únicamente de los patrones almacenados. Si consideramos que éstos son puramente aleatorios, siendo la probabilidad de los valores \( +1 \) y \( -1 \) idéntica, con lo que

\[
P(\xi_j^\mu) = \frac{1}{2} \left[ \delta(\xi_j^\mu - 1) + \delta(\xi_j^\mu + 1) \right]
\]

e independiente de cada \( j \) y de cada \( \mu \), podemos estimar la probabilidad \( P_e \) de que obtengamos inestabilidad en algún bit concreto. La probabilidad de error será \( P_e = P(\Gamma_\nu > 1) \), y claramente deberá crecer con el número de patrones \( p \) almacenados. Si establecemos un criterio de error
Figura 15.13: Distribución de probabilidad de los valores de $\Gamma_\nu$. El área rayada indica la probabilidad de error.

tolerable, esto es, definimos un $1 >> \epsilon > 0$ tal que aceptamos el error en tanto que $P_\epsilon < \epsilon$, podremos estimar la capacidad del sistema. Obviamente, deberemos introducir cierta cota subjetiva para $\epsilon$.

Supongamos que tanto $N$ como $p$ son grandes. La distribución de $\Gamma_\nu$ es una binomial, de media cero y varianza $\sigma^2 = p/N$, pero puede ser aproximada por una distribución Gaussiana, con la misma media y varianza (figura 15.13). La probabilidad de error será la integral que define el área indicada en la curva

$$P_\epsilon = P[\Gamma_\nu > 1] = \frac{1}{\sqrt{2\pi\sigma}} \int_1^{\infty} e^{-x^2/2 \sigma^2} dx$$

$$= \frac{1}{2} \left[ 1 - \text{erf}(1/\sqrt{2\sigma^2}) \right] = \frac{1}{2} \left[ 1 - \text{erf}(1/\sqrt{N/2p}) \right]$$

con erf($x$) la función error, definida por

$$\text{erf}(x) \equiv \frac{2}{\sqrt{\pi}} \int_0^x e^{-u^2} du$$

Como es sabido, estos valores asociados a la distribución normal tipificada se hallan tabulados. En la siguiente tabla damos algunos valores indicativos.

<table>
<thead>
<tr>
<th>$P_\epsilon$ ($P$ error)</th>
<th>$p/N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.001</td>
<td>0.105</td>
</tr>
<tr>
<td>0.00036</td>
<td>0.138</td>
</tr>
<tr>
<td>0.010</td>
<td>0.185</td>
</tr>
<tr>
<td>0.050</td>
<td>0.37</td>
</tr>
<tr>
<td>0.100</td>
<td>0.61</td>
</tr>
</tbody>
</table>

Si elegimos, por ejemplo, $P_\epsilon < 0.01$, obtenemos que el máximo número de patrones almacenable será del orden de $p_{max} = 0.15N$. 
Tenemos por lo tanto una estimación aproximada de la capacidad de la red de Hopfield. Bajo las restricciones empleadas, está claro que sólo es una indicación de la estabilidad: una cota superior. Si el cambio afecta a muchos bits, puede ocurrir que el sistema (que es no-lineal) se desestabilice, yendo a parar a un patrón o atractor distinto. En cualquier caso nos indica en las proximidades de qué cocientes $p/N$ tendremos problemas. Señalemos además que, a medida que introduzcamos nuevas memorias en la red, ésta no presentará problemas hasta alcanzar las proximidades de la capacidad $\alpha_t$, en la que tiene lugar una transición brusca (Amit, 1989). De un modo similar, la red resiste la destrucción de conexiones (o elementos) relativamente bien hasta que pierde sus capacidades de forma rápida. No deja de ser interesante el hecho de que, en algunas patologías del cerebro de tipo degenerativo (como en algunas demencias) la degradación sea habitualmente lenta, pero se acelere en algunas etapas. Ello podría sugerir (sin que la extrapolación sea exacta) cierto papel de los fenómenos cooperativos que se pone de manifiesto en la aparición de estos cambios rápidos.

15.6.2 Función energía

Una de las propiedades más interesantes del modelo de Hopfield reside en la posibilidad de introducir una función energía $H$ dentro de la teoría de redes neurales. Para la red de Hopfield ésta se define por

$$F = H(S_1, ..., S_N; W_{ij}) = -\frac{1}{2} \sum_i \sum_j W_{ij} S_i S_j$$

Se dice que esta función energía define un "paisaje" o "relieve" de energía. Esta superficie es, lógicamente, de alta dimensión. Será además, típicamente, muy rugosa. La propiedad fundamental de esta función energía es que los mínimos corresponderán a los atractores/memorias de la dinámica del sistema. Puede demostrarse fácilmente que, de hecho, la dinámica definida previamente siempre conduce a valores decrecientes de $H$, como esperaríamos de cualquier sistema físico cerrado.

Efectivamente, supongamos que tiene lugar un cambio en una unidad, como consecuencia del campo local, esto es, obtenemos un valor $S'_i$ a partir del cálculo del campo local. El cambio en la energía que resulta de este cambio en una unidad es

$$\Delta H = E' - E = -\sum_j W_{ij} (S'_i - S_i) S_j = -2 \left( \sum_j W_{ij} \right) \text{sgn} \left( \sum_j W_{ij} \right) \leq 0$$

donde hemos considerado el hecho de que, si una unidad presenta un cambio, se tendrá $S'_i = -S_i$.

Este resultado puede representarse en una visión idealizada tal y como se muestra en la figura 15.14. Aquí hemos dibujado la función energía como una superficie en un espacio tridimensional, cuando en realidad tendríamos un espacio de alta dimensión. El estado inicial, que correspondería a un patrón conocido pero deteriorado (donde algunos bits estaban cambiados de signo), puede verse como un punto en el relieve de energía, dentro del valle de atracción del estado de memoria completo. La dinámica del sistema desplaza este valor espontáneamente hacia el fondo del valle, y a medida que ello ocurre los bits erróneos son reemplazados por los correctos. Una vez alcanzado el fondo, el atractor es estable, dado que $H$ es una función de Lyapunov, y la red recuerda el patrón completo.

La minimización local de la función energía permite de hecho recuperar de forma independiente la elección de la matriz de conectividad que define la regla de Hebb. Lo que deseamos es que la minimización de $H$ nos lleve a disminuir el solapamiento entre la configuración adoptada por la
Figura 15.14: Dinámica cualitativa asociada a la red de Hopfield. Existe una función energía $H$ asociada al conjunto de estados de la red que se comporta, cerca de los estados de memoria, como una función de Lyapunov del sistema. Si el estado inicial es un punto cercano (en el espacio multidimensional) a uno de los atractores del sistema, entonces cabe esperar que la dinámica espontánea del mismo lo conduzca al mínimo, siempre que estemos dentro de su valle de atracción. A medida que nos acercamos, vamos recuperando la información no presente inicialmente: el sistema presenta memoria asociativa.
el y el patrón almacenado \( \xi^\mu \). Escribamos

\[
H = -\frac{1}{2N} \left( \sum S_i \xi_i \right)^2
\]

Para el caso de muchos patrones, tenemos

\[
H = -\frac{1}{2N} \sum_{\mu=1}^{P} \left( \sum S_i \xi_i^\mu \right) = \frac{1}{2} \sum_{ij} \left( \frac{1}{N} \sum_{\mu=1}^{P} \xi^\mu_i \xi^\mu_j \right) S_i S_j
\]

que nos da, operando,

\[
H = -\frac{1}{2N} \sum_{i} \left( \sum S_i \xi_i \right) \left( \sum S_i \xi_i \right) = \frac{1}{2} \sum_{ij} \left( \frac{1}{N} \sum_{\mu=1}^{P} \xi^\mu_i \xi^\mu_j \right) S_i S_j
\]

precisamente la función energía definida por la regla de Hebb.

Señalemos que una vez cruzado el valor de la capacidad \( \alpha_c \), la red genera estados espúreos, que no son sino atractores de la dinámica no definidos por las memorias introducidas, sino por combinaciones de estas. En términos del paisaje de energía que hemos dibujado anteriormente, el crecimiento en el número de memorias almacenadas hace que aquél sea más rugoso. Para valores bajos de \( \alpha < \alpha_c \), tendremos valles (cuyo tamaño puede calcularse) grandes y bien definidos, que se harán de tamaño cada vez menor a medida que crezca \( p \). Para valores por encima de la capacidad, la rugosidad (y los estados espúreos) crecen con rapidez.

15.6.3 Red de Hopfield estocástica

Consideraremos ahora una generalización del modelo de Hopfield en el que introduciremos términos estocásticos. La dinámica se definirá de forma parecida al tratamiento dado al modelo de Ising en el capítulo de sistemas críticos. Siguiendo con el procedimiento anterior, ahora definiremos una probabilidad de transición por

\[
P(S_i \rightarrow -S_i) = \phi_{\beta}(\pm h_i) = \frac{1}{1 + \exp(-S_i \beta h_i)}
\]

De forma similar al modelo de Ising, el parámetro \( \beta \) actúa como una cantidad proporcional al inverso de la “temperatura” de la red \( \sigma \), de forma más general, del nivel de ruido de la misma.

Podemos tener una primera idea del comportamiento de la red estocástica mediante una aproximación de campo medio. Nos restringimos para ello al caso en que el número de patrones almacenados es muy inferior al número de neuronas presentes. Podemos obtener la siguiente expresión para \( < S_i > \)

\[
< S_i > = \tanh \left( \frac{\beta}{N} \sum_{j} \xi^\mu_i \xi^\mu_j < S_j > \right)
\]

Se trata de un sistema de ecuaciones no-lineal, no directamente resoluble, pero para el que podemos plantear la hipótesis (ansatz) de proporcionalidad: \( < S_i > = m \xi^\nu_i \), lo cual nos lleva a un nuevo sistema

\[
m \xi^\nu_i = \tanh \left( \frac{\beta}{N} \sum_{j} \xi^\mu_i \xi^\mu_j m \xi^\nu_j \right)
\]
que podemos (de forma similar al tratamiento anterior) separar en dos términos, uno proporcional a $\xi^\nu$ y un término cruzado que, para $p << N$, será despreciable. Encontramos entonces

$$m \xi^\nu = \tanh(\beta m \xi^\nu)$$

o, dado que la tangente hiperbólica cumple la relación $\tanh(-x) = -\tanh(x)$,

$$m = \tanh(\beta m)$$

(véase la teoría de campo medio para el modelo de Ising, capítulo 7) La última relación nos dice que los estados de memoria serán estables para temperaturas menores que 1. La temperatura crítica para la red estocástica $T_c$ es, en esta aproximación, $T_c = 1$. Tenemos un punto de transición de fase en el que el aumento de temperatura marca aquí la pérdida de memoria por la red, que tiene lugar de forma abrupta. Un estudio más detallado de esta transición requiere el formalismo de la mecánica estadística. En la próxima sección llevamos a cabo un estudio formal de la capacidad de la red estocástica (Amit et al., 1985; Hertz, 1991). El lector no interesado o no familiarizado con este formalismo puede prescindir de esta sección y seguir en 15.9.

### 15.7 Capacidad de la red estocástica

Partiremos en nuestro análisis empleando el parámetro $\alpha = p/N$ que ahora consideraremos del orden de la unidad (seguiremos los pasos indicados en Hertz, 1991). $N$ será grande, de forma que ignoraremos las diferencias entre $p$ y $p - 1$. Consideremos las variables de solapamiento, definidas por el conjunto

$$m_\nu = \frac{1}{N} \sum_i \xi_i^\nu \langle S_i \rangle$$

Supongamos que estamos analizando un patrón concreto, digamos el primero, con lo que supondremos que $m_1 \approx 1$ mientras que los restantes solapamientos son del orden de $\approx 1/\sqrt{N}$. En estas condiciones, la suma

$$r = \frac{1}{\alpha} \sum_{\nu \neq 1} m_\nu^2$$

será próxima a uno. El método que introducimos aquí requiere un cálculo autoconsistente de $r$ y de $m_1$. Partiremos de las ecuaciones de campo medio antes escritas, que introduciremos en la forma

$$m_\nu = \frac{1}{N} \sum_i \xi_i^\nu \tanh \left( \beta \sum_\mu \xi_i^\mu m_\mu \right)$$

si separamos los términos para $\mu = 1$ y para $\mu \neq 1$ tendríamos (para $\nu \neq 1$),

$$m_\nu = \frac{1}{N} \sum_i \xi_i^\nu \xi_i^1 \tanh \left[ \beta \left( m_1 + \xi_i^\nu \xi_i^1 m_\nu \right) + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right]$$

El primer término en la tangente hiperbólica es del orden de la unidad (tal y como hemos supuesto) y el último es grande dado que sumamos sobre $p$ términos. Pero el segundo término es pequeño (del orden de $1/\sqrt{N}$). Podemos por lo tanto llevar a cabo un desarrollo en Taylor

$$m_\nu = \frac{1}{N} \sum_i \xi_i^\nu \xi_i^1 \tanh \left[ \beta \left( m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right) \right] + \frac{1}{N} \sum_i \left( 1 - \tanh^2 \left[ \beta \left( m_1 + \sum_{\mu \neq 1, \nu} \xi_i^\mu \xi_i^1 m_\mu \right) \right] \right) m_\nu$$
A continuación, asumiremos que las solapamientos \( m_\nu \) son (ahora \( \mu \neq 1 \)) variables aleatorias independientes con media cero y varianza \( \sigma r / \rho \). Empleando el teorema central del límite, el promedio sobre \( i \) es un promedio sobre variables Gaussiana\( \xi^\mu_i \xi^\nu_1 m_\mu \) de varianza \( \sigma r \). Tendremos entonces

\[
m_\nu = \frac{1}{N} \sum_i \xi^\nu_i \xi^\nu_1 \tanh \left[ \beta \left( m_1 + \sum_{\mu \neq 1, \nu} \xi^\mu_i \xi^\nu_1 m_\mu \right) \right] + \beta m_\nu - \beta q m_\nu
\]

de donde

\[
m_\nu = \frac{N^{-1} \sum_i \xi^\nu_i \xi^\nu_1 \tanh \left[ \beta \left( m_1 + \sum_{\mu \neq 1, \nu} \xi^\mu_i \xi^\nu_1 m_\mu \right) \right]}{1 - \beta(1 - q)}
\]

con

\[
q = \int \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tanh \left[ \beta \left( m_1 + \sqrt{\sigma r} z \right) \right] dz
\]

Ahora procederemos al cálculo de \( r \). Haciendo el cuadrado de la última expresión para \( m_\nu \), tenemos que

\[
m^2_\nu = \left[ \frac{1}{1 - \beta(1 - q)} \right]^2 \frac{1}{N^2} \sum_i \sum_j \xi^\nu_i \xi^\nu_1 \xi^\nu_j \xi^\nu_1 \tanh \left[ \beta \left( m_1 + \sum_{\mu \neq 1, \nu} \xi^\mu_i \xi^\nu_1 m_\mu \right) \right] \times \tanh \left[ \beta \left( m_1 + \sum_{\mu \neq 1, \nu} \xi^\mu_i \xi^\nu_1 m_\mu \right) \right]
\]

y promediaremos el resultado sobre el conjunto de memorias. Dado que el \( \nu \)-ésimo patrón no aparece dentro de la tangente hiperbólica, los factores \( \xi^\nu_i \xi^\nu_1 \xi^\nu_j \xi^\nu_1 \) pueden ser promediados de forma separada y únicamente nos quedará el \( j = i \). En consecuencia, el promedio da un factor \( q \) y el resultado es independiente de \( \nu \):

\[
r = \frac{q}{\left[ 1 - \beta(1 - q) \right]^2}
\]

Necesitaremos también una ecuación para \( m_1 \) y, empleando la misma aproximación, partiendo de \( \nu = 1 \), podemos demostrar fácilmente que

\[
m_1 = \int \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tanh \left[ \beta \left( m_1 + \sqrt{\sigma r} z \right) \right] dz
\]

Ahora, las ecuaciones obtenidas para \( q \), \( r \) y \( m_1 \) pueden ser resueltas simultáneamente, lo cual se llevará a cabo en general de forma numérica. El caso de temperatura cero (\( \beta \to \infty \)) puede ser resuelto analíticamente. En este límite tenemos que \( q \to 1 \) pero \( C \equiv \beta(1 - q) \) es finita. El caso límite permite aproximar las integrales anteriores mediante el empleo de las identidades siguientes

\[
\int \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \left[ 1 - \tanh^2 \beta(az + b) \right] dz \approx \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \left. \tanh \beta(az + b) \right|_{\tanh \beta(az + b) = 0} \times \int \left[ 1 - \tanh^2 \beta(az + b) \right] dz
\]

\[
= \frac{1}{\sqrt{2\pi}} e^{-b^2/2a^2} \frac{1}{a\beta} \int \frac{\partial}{\partial z} \tanh \beta(az + b) dz = \frac{\sqrt{2}}{\pi} \frac{1}{a\beta} e^{-b^2/2a^2}
\]

y, por otra parte, tenemos que, para temperaturas cercanas a cero podemos aproximar la integral a la función error

\[
\int \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \tanh \left[ \beta(az + b) \right] dz \approx \int \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \operatorname{sgn} \left[ \beta(az + b) \right] dz
\]
Figura 15.15: Diagrama de fases del modelo de Hopfield. Presentamos los parámetros más relevantes: α y la temperatura T. Para temperatura cero, recuperamos el resultado de la capacidad crítica α ∼ 0.13 obtenido con anterioridad.

\[ = 2 \int_{b/a}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz - 1 = \operatorname{erf} \left( \frac{b}{\sqrt{2\alpha}} \right) \]

Y por lo tanto las tres ecuaciones iniciales se convierten en

\[ C \equiv \beta(1-q) = \sqrt{\frac{2}{\pi \alpha r}} \exp \left( -\frac{m^2}{2\alpha r} \right) \]

\[ r = \frac{1}{(1-C)^2} \]

\[ m = \operatorname{erf} \left( \frac{m}{\sqrt{2\alpha r}} \right) \]

Podemos calcular la capacidad de la red resolviendo estas ecuaciones. Si hacemos \( y = m/\sqrt{2\alpha r} \), obtenemos la siguiente ecuación

\[ y \left( \sqrt{2\alpha} + \frac{2}{\sqrt{\pi}} e^{-y^2} \right) = \operatorname{erf}(y) \]

que nos permite calcular el valor crítico \( \alpha_c \) para el cual las soluciones no triviales \( m \neq 0 \) desaparecen. Obtenemos numéricamente un valor \( \alpha_c \approx 0.138 \). El salto experimentado por \( m \) en este punto crítico es enorme: pasamos de \( m \approx 0.97 \) a \( m = 0 \) (tenemos de hecho una transición de fase de primer orden). Estas ecuaciones permiten, finalmente, obtener el espacio de parámetros del modelo de Hopfield, que mostramos en la figura 15.15. Observamos distintas regiones, de las cuales sólo la región sombreada corresponde a la de estados de memoria, en los que la red funciona correctamente y los estados de entrada corresponden a mínimos de la función energía. Las zonas restantes corresponden a una red desordenada (un vidrio de espin).
15.8 Retropropagación (back propagation)

Los modelos del tipo de la red de Hopfield se basan en la idea de aprendizaje no-supervisado: la red, formada por un sistema altamente interconectado, posee una forma particular de readaptar las conexiones en función de las entradas. Otros tipos de redes, típicamente formadas por varias capas, llevan a cabo lo que se denomina aprendizaje supervisado, en el que, dada una entrada a la red (que tiene lugar sobre una capa dada) el sistema procesa la información para dar una salida que comparamos con lo esperado. El aprendizaje está supervisado en la medida en que las modificaciones de la red se llevan a cabo en función de nuestra comparación.

El back propagation es un método de aprendizaje supervisado de redes neuronales que se aplica a sistemas con varios niveles de neuronas ocultas entre los niveles de entrada y de salida. Supongamos que un sistema capaz de aprender está formado por N niveles. El primero está constituido por neuronas de entrada que reciben una serie de estímulos \( \{I_j\} \). Las señales de salida \( \{O_j\} \) se leen en las neuronas del último nivel. Entre estas dos capas hay \( (N - 2) \) niveles de unidades ocultas. Sólo están conectadas entre sí las neuronas que corresponden a niveles vecinos. El número total de unidades en una capa puede variar de uno a otro nivel.

Cuando la neurona \( i \) en el nivel \( n \) recibe la señal \( z_i^{(n)} \), produce a su vez una señal de salida

\[
y_i^{(n)} = \left[ 1 + \exp \left( -z_i^{(n)} \right) \right]^{-1}
\]

(15.8.1)

La señal de entrada recibida por la unidad \( j \) en el nivel siguiente, \( (n + 1) \) se forma como una suma ponderada de las señales de salida en el nivel anterior, \( n \)

\[
z_j^{(n+1)} = \sum_i J_{ij} y_i^{(n)}
\]

(15.8.2)
Los pesos \( J_{ik}^{(n)} \) pueden tomar valores tanto positivos como negativos.

Los estímulos de entrada para las neuronas del primer nivel están dados por el conjunto de valores presentados para analizar, es decir \( x^{(1)}_j = I_j \). La salida de la última capa es el resultado de procesar estas entradas, \( y^{(n)}_j = O_j \).

Si los pesos de las conexiones, \( J_{ik} \), se establecen de forma aleatoria, la respuesta del sistema será también aleatoria. El propósito del método de aprendizaje por "back propagation" es ajustar los pesos de forma que se produzcan las salidas deseadas.

Formúlemos el problema de la forma siguiente. Supongamos que deseamos que un sistema clasifique todos los posibles conjuntos de estímulos de entrada en diversas categorías de acuerdo con un cierto criterio. Formamos entonces un conjunto suficientemente amplio de estímulos de entrenamiento, \( \{I_{ij,m}\} \), correspondientes a diferentes categorías, e indicamos para cada uno de ellos la categoría de salida esperada, \( \{D_{ij,m}\} \). Durante el proceso de aprendizaje, el sistema deberá de ajustar los pesos de forma que, finalmente, cada vez que se presente cierto estímulo se obtenga la salida deseada. Una vez finalizado el entrenamiento, si se presenta un estímulo que no estuviese incluido en el conjunto \( \{I_{ij,m}\} \), esperamos que el sistema sea capaz de clasificarlo en la categoría correspondiente. Esto sucedrá si el conjunto de estímulos de entrenamiento era suficientemente amplio y representativo. Definamos la función error como

\[
E = \frac{1}{2} \sum_{m=1}^{M} \sum_{j} (D_{ij,m} - O_{ij,m})^2
\]  
(15.8.3)

donde \( O_{ij,m} = y^{(N)}_{ij,m} \) son las salidas reales de los elementos del último nivel cuando el conjunto de valores de entrenamiento \( m \)-éximo, \( \{I_{ij,m}\} \) se ha presentado a las unidades de la primera capa. Evidentemente, esta función tendrá un mínimo cuando, para todos los \( M \) conjuntos de entrenamiento, las salidas reales coincidan con las salidas deseadas, \( D_{ij,m} \). Nuestro propósito ahora es encontrar los valores de los pesos \( J^{(n)}_{ij,m} \) que minimizan 15.8.3. El aprendizaje es, pues, un problema complejo de optimización.

A fin de minimizar \( E \) utilizando el método del steepest descent (descenso en gradiente, figura 15.17) debemos calcular las derivadas parciales de \( E \) respecto del peso de cada una de las conexiones. Cada una de las derivadas involucradas está dada por una suma de cantidades relacionadas con cada uno de los conjuntos de entrenamiento. Para un conjunto de entradas dado, estas derivadas pueden ser calculadas explícitamente mediante la técnica del "back propagation" (Rumelhart et al., 1986; Le Cun, 1985).

Obsérvese, para empezar, que se verifica

\[
\frac{\partial E}{\partial y^{(n-1)}_i} = \sum_j \frac{\partial E}{\partial y^{(n)}_j} \frac{\partial y^{(n)}_j}{\partial x^{(n)}_j} \frac{\partial x^{(n)}_j}{\partial y^{(n-1)}_i}
\]  
(15.8.4)

Se ha supuesto que uno de los conjuntos de entrenamiento se ha aplicado al nivel de entrada, así que \( x^{(n)}_j \) y \( y^{(n)}_j \) son las señales correspondientes a la entrada y a la salida de este ejemplo particular, es decir, \( y^{(n)}_j = y^{(n)}_{ij,m} \).

Utilizando 15.8.1 y 15.8.2, obtenemos de 15.8.4 una fórmula recurrente,

\[
\frac{\partial E}{\partial y^{(n-1)}_i} = \sum_j \frac{\partial E}{\partial y^{(n)}_j} y^{(n)}_j (1 - y^{(n)}_j) J^{(n-1)}_{ji}
\]  
(15.8.5)

Esta expresión nos permite calcular las derivadas \( \partial E/\partial y \) para un cierto nivel si se conocen las derivadas para el nivel siguiente. Para el último nivel se puede hallar fácilmente la expresión de las derivadas, de 15.8.3,
Figura 15.17: Descenso en gradiente: representamos una superficie de energía bidimensional (para una red neural se trata en realidad de un sistema multidimensional). Partiendo de un punto dado, el método expuesto en el texto nos permite escoger la dirección de disminución de la función error en la dirección del gradiente (la de variación más rápida).

\[
\frac{\partial E}{\partial y_i^{(N)}} = y_i^{(N)} - D_i
\]

Así que empezando por el nivel inferior y moviéndonos en sentido contrario es posible calcular, utilizando 15.8.5, todas las derivadas parciales requeridas para cada una de las capas. Además podemos utilizar la identidad

\[
\frac{\partial E}{\partial j_{ji}^{(n-1)}} = \frac{\partial E}{\partial y_j^{(n)}} \frac{\partial y_j^{(n)}}{\partial x_j^{(n)}} \frac{\partial x_j^{(n)}}{\partial j_{ji}^{(n-1)}}
\] (15.8.6)

que puede ser escrita como

\[
\frac{\partial E}{\partial j_{ji}^{(n-1)}} = \frac{\partial E}{\partial y_j^{(n)}} y_j^{(n)} (1 - y_j^{(n)}) y_j^{(n-1)}
\] (15.8.7)

aplicando 15.8.1 y 15.8.2.

Una vez que se han calculado todas las derivadas, esta última expresión permite hallar la variación de los pesos, \( \partial E / \partial J \), para cada conexión y para cada uno de los distintos conjuntos de entrenamiento. Estas derivadas permiten corregir los pesos tras la presentación de cada grupo de entradas. Podríamos también optar por la actualización de los pesos tras la presentación de todos los patrones de entrada: en este caso, sólo después de haberlos mostrado todos se realiza la actualización de los pesos, y la corrección introducida se corresponde con la suma de todas las correcciones individuales. En cualquiera de los dos casos, el procedimiento incluye presentación de los conjuntos de entrenamiento, cálculo de las derivadas de la función error \( E \) respecto de los pesos de las conexiones y una pequeña modificación de cada peso, mediante una corrección

\[
\Delta j_{ji}^{(n)} = -\epsilon \frac{\partial E}{\partial j_{ji}^{(n)}}
\] (15.8.8)
tras lo cual el ciclo completo de entrenamiento se repite.

El aprendizaje debe continuar hasta que se obtiene un valor aceptablemente pequeño del error 15.8.3. Si el proceso no permite alcanzar el error deseado, se puede repetir incrementando esta vez el número de unidades ocultas. En lugar de 15.8.8 es preferible usar el algoritmo de corrección

$$
\Delta J^{(n)}_{ji} = -\epsilon \frac{\partial E}{\partial J^{(n)}_{ij}} + \kappa \Delta J^{(n)}_{ji}(t - 1)
$$

(15.8.9)

donde el tiempo discreto $t$ aumenta en una unidad tras cada ciclo de aprendizaje, y el coeficiente $\kappa$, con valor entre 0 y 1, especifica la "inercia" del sistema que aprende.

Las capacidades de aprendizaje son mayores si las unidades del sistema presentan unas desviaciones permanentes, de forma que 15.8.2 sea reemplazada por

$$
J^{(n+1)}_{ji} = \sum_J J^{(n)}_{ji} y^{(n)}_i + b^{(n)}_j
$$

El valor de las desviaciones $b^{(n)}$ también puede ser ajustado durante el proceso de aprendizaje. La introducción de estos valores no implica ningún cambio esencial en los argumentos anteriores. Se pueden interpretar como la aparición de una conexión permanente $Y^{(n)}_{ji} = 1$ con peso $b^{(n)}_j$. En consecuencia, se las puede considerar en igualdad respecto de los pesos $J^{(n)}_{ji}$. De forma similar a 15.8.7 se encuentra

$$
\frac{\partial E}{\partial b^{(n-1)}_j} = \frac{\partial E}{\partial y^{(n)}_j} y^{(n)}_j (1 - y^{(n)}_j)
$$

Conocidas las derivadas, la corrección se puede introducir, por ejemplo, mediante

$$
\Delta b^{(n)}_j = -\epsilon \frac{\partial E}{\partial b^{(n)}_j}
$$

Una tarea que no se podría realizar sin unidades ocultas (es decir, conectando un nivel de entrada a uno de salida, sin más) es la determinación de series unidimensionales de datos simétricas respecto de su punto central, capicuas. En cambio, una solución a este problema se puede obtener con tan sólo dos unidades ocultas, y el método del "back propagation" es en este caso óptimo para determinar los valores de las conexiones.

La técnica del "back propagation" emplea un algoritmo de optimización determinista para encontrar los pesos requeridos. Existe en consecuencia el riesgo de que, moviéndonos estrictamente en la dirección del gradiente de la función error $E$, se alcance uno de los mínimos locales de esta función, y el proceso de aprendizaje se detenga a pesar de no haber llegado a ninguna solución genuina. Si esto sucede, se puede intertar añadir nuevos niveles de unidades ocultas. Esto incrementa la dimensionalidad del espacio de los pesos. Podemos esperar que aparezca algún valle nuevo que destruya las barreras que separan los mínimos locales indeseados del espacio de menor dimensionalidad. Sin embargo, no existe ninguna prueba de la convergencia del algoritmo, y se duda incluso de si ésta existe. En cualquier caso, el éxito de este tipo de red ha sido enorme, y se ha empleado en numerosas aplicaciones, que van desde el reconocimiento de voz hasta el análisis de imágenes, predicción de series temporales, etc. (Hertz et al., 1991).

### 15.9 La máquina de Boltzmann

La máquina de Boltzmann fue propuesta por Hinton y Sejnowski (1983), y es el sistema más universal con aprendizaje analógico. En general, el proceso del aprendizaje se puede contemplar
como la construcción, por parte de una máquina analógica, de algún tipo de representación interna, de modelo interno, que sea capaz de reproducir regularidades en las relaciones entre distintos esquemas en el "mundo exterior".

Supongamos que el "mundo exterior" contiene dos clases de elementos, los de la clase A y los de la clase B. Los elementos pertenecientes a la primera clase estarán especificados por un subíndice \( \alpha = 1, 2, \ldots, K \), y los de la segunda por \( \beta = 1, 2, \ldots, M \). Supongamos que la relación entre los elementos de cada una de estas clases es de naturaleza estadística, es decir, está determinada por las probabilidades conjuntas \( p(\alpha, \beta) \) de observación de varios elementos \( \beta \) y \( \alpha \). Podemos también definir probabilidades condicionadas,

\[
\pi(\beta | \alpha) = \frac{p(\beta, \alpha)}{p(\alpha)}
\]

que nos dice cual es la probabilidad de observar el elemento \( \beta \) si otro elemento \( \alpha \) está presente en el medio (y \( p(\alpha) \) es la probabilidad de observar a \( \alpha \)).

Supóngase además que un elemento \( \alpha \) corresponde a una cierta entrada de la máquina analógica, en tanto que los elementos \( \beta \) corresponden a las respuestas posibles. Diremos entonces que esta máquina tiene un modelo interno válido si las relaciones estadísticas entre sus entradas y sus respuestas son las mismas que entre los elementos correspondientes del mundo exterior. El aprendizaje puede ser interpretado como el proceso de construcción gradual de un modelo interno.

La máquina de Boltzmann representa una red de unidades bistables, cada una de ellas con dos estados especificados por una variable binaria, \( s_i \), que toma valores 0 y 1. Las unidades no se agrupan en niveles, sino que cada una de ellas está simétricamente conectada a las demás. Para cada patrón de actividad \( \{s_j\} \) de esta red, la energía \( E \) se define como

\[
E = -\frac{1}{2} \sum_{i,j} J_{ij} s_i s_j + \sum_i b_i s_i
\]

donde los pesos de las conexiones son \( J_{ij} \), con \( J_{ii} = 0 \), y \( b_i \) son las desviaciones.

Todas las unidades en una máquina de Boltzmann se dividen en visibles (unidades de entrada y de salida) y ocultas. Los estímulos externos se aplican a las unidades de entrada, y las respuestas correspondientes se obtienen en las unidades de salida.

La aplicación de un estímulo particular fija el valor de las unidades de entrada, pero deja libres todas las demás. Las unidades libres experimentan transiciones de acuerdo con el algoritmo que a continuación especificaremos. Al igual que en la sección anterior, donde hemos tratado con el algoritmo de "back propagation", las desviaciones \( b_i \) permanentes pueden ser interpretadas como conexiones entre la unidad \( i \) y una unidad de entrada adicional que estuviese continuamente en el estado activo. El peso de dicha conexión sería \( b_i \). En consecuencia, desde el punto de vista físico es suficiente con considerar una red sin desviaciones, es decir, con energía

\[
E = -\frac{1}{2} J_{ij} s_i s_j
\]

Cuando la unidad \( i \) cambia del estado inactivo al activo, la energía aumenta en una cantidad

\[
\Delta E_i = \sum_j J_{ij} s_j
\]

El algoritmo probabilístico de transición es tal que, para estados dados de todas las demás unidades, la unidad \( i \) pasa al estado activo \( (s_i = 1) \) en el siguiente paso de tiempo con una probabilidad

\[
p_i = \frac{1}{1 + \exp (\Delta E_i / \theta)}^{-1}
\]
donde $\theta$ juega el papel de la temperatura.

Mediante métodos generales de la física estadística, se puede ver que este tipo de transiciones probabilistas lleva finalmente al sistema a un estado de \textit{equilibrio térmico}, donde la probabilidad de observar una configuración particular \{$s_i$\} de la red está dada por la distribución de Boltzmann,

$$P(\{s_i\}) = Z^{-1} \exp \left[ -E(\{s_i\}) / \theta \right]$$

donde $Z$ es el factor de normalización.

Como se ha mencionado, una máquina de Boltzmann ha aprendido cuando ha conseguido un modelo interno correcto, que reproduce las relaciones probabilísticas entre los elementos de la clase A y B en el mundo exterior. Esto implica que algunos elementos de ambas clases deben ser previamente codificados en una serie de conjuntos binarios de actividad, $I^{(a)} = (I_1^{(a)}, \ldots, I_n^{(a)})$, y $O^{(b)} = (O_1^{(b)}, \ldots, O_m^{(b)})$ de unidades de entrada y de salida.

Cada modelo interno representa un conjunto particular de pesos $J_{ij}$ en la máquina de Boltzmann. Cuando todos estos pesos están fijados y las unidades de entrada se han especificado de acuerdo con una cierta configuración $I^{(a)}$, la red generará con cierta probabilidad diferentes configuraciones $O^{(b)}$ en las unidades de salida. Mediante la observación del comportamiento del sistema se puede calcular la probabilidad condicionada $\pi(\beta | \alpha)$ de que se dé una configuración de salida $\beta$ si la entrada $\alpha$ se había aplicado a las unidades de entrada. El proceso de aprendizaje consiste en el cambio gradual de los pesos $J_{ij}$ que acerca el valor de las probabilidades generadas por el modelo interno, $\pi'(\alpha | \beta)$, a las probabilidades condicionadas $\pi(\alpha | \beta)$ que existen en el mundo exterior.

El criterio de semejanza entre estos dos conjuntos de probabilidades puede estimarse por la medida de información siguiente

$$G = \sum_{\alpha, \beta} p(\beta, \alpha) \frac{\partial}{\partial J_{ij}} \ln \pi'(\beta | \alpha)$$

que juega el papel análogo al error en el algoritmo de “back propagation”. Debemos calcular la derivada logarítmica de $\pi'$ respecto de $J_{ij}$ para las configuraciones de actividad $I$ y $O$ de las unidades de entrada y de salida, respectivamente.

Llamemos $S = \{s_i\}$ a una configuración de actividad determinada de las unidades ocultas. Entonces, cuando se fija una cierta entrada $I$, la probabilidad de hallar los subsistemas de unidades ocultas y unidades de salida en los estados $S$ y $O$, respectivamente, está dada por la expresión

$$P_I(S, O) = Z_I^{-1} \exp \left[ -E_I(S, O) / \theta \right]$$

(15.9.1)

donde $Z_I$ se define como

$$Z_I = \sum_{S, O} \exp \left[ -E_I(S, O) / \theta \right]$$

Dado que la probabilidad condicionada $\pi'(O | I)$ se relaciona con $P_I(S, O)$ mediante

$$\pi'(O | I) = \sum_S P_I(S, O)$$

tenemos

$$\pi'(O | I) = Z_I^{-1} \sum_S \exp \left[ -E_I(S, O) / \theta \right]$$

(15.9.2)

Introduciendo la cantidad
\[ Z_{\text{LO}} = \sum_S \exp \left[ -E_I(S, O)/\theta \right] \]

que se calcula manteniendo fijos los estados de las unidades de entrada y de las unidades de salida, ambas simultáneamente, podemos reescribir 15.9.2 como

\[ \pi'(O|I) = \frac{Z_{\text{LO}}}{Z_I} \]

En consecuencia,

\[ \frac{\partial}{\partial J_{ij}} \ln \pi'(O|I) = \frac{1}{\theta} \left( <s_i s_j >_{\text{IO}} - <s_i s_j >_{\text{I}} \right) \]

\[ \frac{\partial}{\partial J_{ij}} \ln \pi'(O|I) = \frac{1}{\theta} \sum_{\alpha, \beta} \sum_{\gamma} p(\beta, \alpha) \left( <s_i s_j >_{\alpha} - <s_i s_j >_{\alpha, \beta(\alpha)} \right) \]

(15.9.3)

(15.9.4)

(15.9.5)

(15.9.6)

Si consideramos las definiciones dadas para \( Z_{\text{LO}} \) y \( Z_I \), así como la dependencia de la energía \( E \) en las conexiones, deducimos de la expresión anterior que

\[ \frac{\partial}{\partial J_{ij}} \ln \pi'(O|I) = \frac{1}{\theta} \sum_{\alpha, \beta} \sum_{\gamma} p(\beta, \alpha) \left( <s_i s_j >_{\alpha} - <s_i s_j >_{\alpha, \beta(\alpha)} \right) \]

(15.9.4)

(15.9.5)

(15.9.6)

(15.9.7)

Cómo se conocen las derivadas \( \partial G/\partial J_{ij} \), éstas pueden ser utilizadas para modificar los pesos \( J_{ij} \) a fin de disminuir el valor de \( G \). Un posible criterio para realizar lo especificado es

\[ \Delta J_{ij} = - \frac{\partial G}{\partial J_{ij}} \]

(15.9.5)

(15.9.6)

(15.9.7)

En un caso particular especialmente interesante, cuando cada elemento de entrada \( \alpha \) está unívocamente asociado a cada elemento de salida \( \beta \), \( \alpha = \alpha(\beta) \), y los elementos \( \alpha \) son equiprobables, la ecuación 15.9.4 toma la forma

\[ \frac{\partial G}{\partial J_{ij}} = \frac{1}{K \theta} \sum_{\alpha} p(\beta, \alpha) \left( <s_i s_j >_{\alpha} - <s_i s_j >_{\alpha, \beta(\alpha)} \right) \]

(15.9.6)

(15.9.7)

(15.9.8)

Damos primeramente la descripción del ciclo de aprendizaje en este caso sencillo. Consta de los pasos siguientes:

1. Las unidades de entrada son fijadas en los estados que corresponden a un elemento de entrada \( \alpha \), en tanto que las unidades de salida se fijan en los estados \( \beta(\alpha) \) impuestos por el de entrada. La red se lleva al equilibrio térmico y, promediando durante intervalos de tiempo largos, se calculan las correlaciones \( <s_i s_j >_{\alpha} \) para todas las parejas \( (i, j) \). Este paso se repite para todos los elementos \( \alpha \) en el conjunto de entrenamiento.

2. Se fijan las unidades de entrada de nuevo en la configuración especificada por el elemento \( \alpha \), pero las unidades de salida se dejan libres. Se lleva de nuevo la red al equilibrio térmico y se calculan las correlaciones \( <s_i s_j >_{\alpha} \) para todas las parejas. Este paso se repite para todos los elementos \( \alpha \).
3. Las correlaciones $<s_i s_j>_{\alpha,\beta(\alpha)}$ y $<s_i s_j>_{\alpha}$ calculadas se usan para determinar la derivada 15.9.6 y para modificar los pesos de acuerdo con 15.9.5.

Los ciclos anteriores deben repetirse hasta que se obtiene una frecuencia aceptable de asociaciones correctas. Si no acaba obteniéndose una buena convergencia, se repite el procedimiento entero con un número mayor de unidades ocultas.

A pesar de su aparente simplicidad, este algoritmo de aprendizaje tiene un punto sutil. Si deseamos construir una máquina de Boltzmann capaz de asociación determinista, se requiere una temperatura $\theta$ baja. De no ser así, las fluctuaciones estadísticas serán grandes y no se conseguirá una asociación clara. Cuando la temperatura es baja, la red se encontrará mayormente en las configuraciones de actividad de energía menor, una vez que se haya alcanzado el equilibrio térmico. Pero además de un mínimo central más profundo, la red debe poseer otros mínimos locales separados del principal por barreras suficientemente grandes. En el proceso de relajación, el sistema puede alcanzar algún mínimo local y permanecer allí hasta que una fluctuación suficientemente grande le permita sobrepasar la barrera. El tiempo de espera diverge exponencialmente con la disminución de la temperatura. Así que si empezamos con un valor de $\theta$ excesivamente bajo, necesitaremos un tiempo inaceptablemente largo para llegar a la red al equilibrio térmico supuesto en la expresión 15.9.6.

Esta dificultad puede ser eliminada mediante la aplicación del procedimiento del simulated annealing. En cada paso $A$ o $B$, debemos empezar con una temperatura suficientemente grande, y rebejarla de forma gradual de forma que la red permanezca en el equilibrio térmico. El valor de $\theta$ en la ecuación 15.9.6 es aquel en el cual este procedimiento acaba.

Se puede utilizar un proceso similar para aprender las asociaciones probabilísticas. Como se ha dicho, en este caso el elemento de entrada puede estar asociado a diferentes respuestas $\beta$, con frecuencias relativas determinadas por la probabilidad condicionada $\pi(\alpha|\beta)$ inicial. En el paso 1, fijamos las unidades de entrada y de salida mediante elementos arbitrariamente escogidos, $\alpha$ y $\beta$, y tras el simulated annealing, se determinan las correlaciones $<s_i s_j>_{\alpha,\beta}$ en el equilibrio térmico para un conjunto dado de pesos $J_{ij}$. El paso 2 no cambia. En el paso 3, se debe usar 15.9.4 en lugar de 15.9.6 para calcular las derivadas $\partial G/\partial J_{ij}$. Se puede modificar el proceso de aprendizaje de forma que todos los pesos cambien en una cantidad

$$\Delta J_{ij} = \frac{\epsilon}{\theta} \sum_{\alpha,\beta} \pi(\alpha,\beta) <s_i s_j>_{\alpha,\beta}$$

cada paso 1, y en una cantidad

$$\Delta J_{ij} = \frac{\epsilon}{\theta} \sum_{\alpha} \pi(\alpha) <s_i s_j>_{\alpha}$$

cada paso 2.

Por tanto, los pesos aumentan tras la aplicación de 1. y disminuyen tras 2. Podemos decir que en el primero se da un aprendizaje positivo, durante el cual la información del medio se almacena en los pesos de las conexiones. La segunda fase representa un “desaprendizaje”, durante el cual el sistema colecta estados aleatoriamente de acuerdo con la distribución de Boltzmann.

La mayor desventaja de la máquina de Boltzmann consiste en el tiempo tan largo requerido para el aprendizaje, debido al proceso de “annealing” y a las largas series necesarias para muchas parejas de elementos, a fin de producir una buena estadística.
15.10 Redes con intermediarios

La característica que distingue a los modelos de redes neuronales es su alta conectividad. Por ejemplo, en el modelo de Hopfield de memorias asociativas, cada uno de los elementos está conectado a todos los demás. Dado que el número total de conexiones requeridas crece como el cuadrado del número de elementos, aparece una seria dificultad cuando se intenta implementar estas redes a escala microscópica, digamos en dispositivos electrónicos moleculares. Desde el punto de vista tecnológico, es preferible utilizar redes a tener únicamente conexiones locales regulares. En esta sección mostraremos que muchos modelos de redes neuronales pueden ser reformulados en términos locales, con interacciones entre unidades transmitidas por intermediarios.

Consideremos para empezar un modelo de Hopfield probabilista con \( N \) espines \( S_i \). Supongamos que, en una unidad de tiempo, cualquier espín puede realizar un salto \( S_i \rightarrow S_i' \) con una probabilidad \( W(S_i \rightarrow S_i') \) que está determinada por las configuraciones de los espines en ese momento. Es decir, asumimos

\[
W(S_i \rightarrow S_i') = w \exp \left\{ h_i(S_i - S_i')/\theta \right\}
\]

donde, como antes, tenemos

\[
h_i = \sum_j J_{ij} S_j
\]

Cuando la temperatura \( \theta \) tiende a cero, sólo están permitidas las transiciones que disminuyen la energía total. El ritmo global de transiciones de los espines está especificado por un parámetro \( w \). Si utilizamos la regla hebbiana de aprendizaje, tendremos

\[
J_{ij} = \frac{1}{N} \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu
\]

El modelo local correspondiente con intermediarios se construye de la forma siguiente. Supongamos que los espines (en realidad unidades con dos estados) están inmersos en el medio, donde ciertas partículas llamadas intermediarios se hallan en difusión. Hay diversos tipos de intermediarios, y cada uno de ellos corresponde a una memoria \( \{ \xi_i^\mu \} \) particular. Por simplicidad, supondremos que la difusión de los intermediarios es suficientemente rápida como para asegurar una mezcla ideal (mixing), de forma que finalmente no se considerará ninguna distribución espacial de sucesos.

Cualquier espín, localizado en la celda \( R_i \), genera intermediarios del tipo \( \mu \) cuando \( S_i = \xi_i^\mu \), es decir, cuando su vector estado coincide con el sentido de este espín para la memoria almacenada \( \mu \)-ésima. Para una configuración fijaada de espines \( \{ S_i \} \), el ritmo de generación local \( \rho_\mu \) de intermediarios \( \mu \) en el punto \( r \) es

\[
\rho_\mu = \gamma \frac{V}{N} \sum_i (\xi_i^\mu S_i + 1) \delta(r - R_i)
\]  

(15.10.1)

donde \( V \) es el volumen del medio. Todos los intermediarios están sujetos a una desintegración a ritmo constante \( \gamma \) (lo cual previene su acumulación en el medio).

Los intermediarios de cada tipo \( \mu \) actúan sobre los espines, forzándolos a presentar el estado definido por la memoria \( \mu \)-ésima correspondiente. Los ritmos locales de transición entre estados de los espines están dados por

\[
\Phi_i(S_i \rightarrow S_i') = w \exp \left[ \frac{1}{\theta} \sum_\mu \xi_i^\mu (n_\mu(R_i) - 1)(S_i' - S_i) \right]
\]  

(15.10.2)
donde \( n_\mu \) es la densidad local de intermediarios \( \mu \). Comparando la última expresión con la probabilidad \( W \) de transición entre estados, observamos que los espines presentan transiciones térmicas en un campo efectivo

\[ h_i = \sum_\mu \xi'^\mu(n_\mu(R_i) - 1) \]

que se determina a partir de las densidades de los intermediarios en cada instante.

Las ecuaciones cinéticas para la densidad de intermediarios son simplemente

\[ \dot{n}_\mu = -\gamma n_\mu + \gamma \frac{V}{N} \sum_i (\xi'^\mu S_i + 1) \delta(r - R_i) + D \Delta n_\mu \]  \hspace{1cm} (15.10.3)

donde \( D \) es la constante de difusión del intermediario.

Si la difusión es suficientemente rápida, de forma que se garante una mezcla ideal, se mantendrá la distribución uniforme de intermediarios. Si además las vidas medias de éstos son mucho más cortas que el intervalo entre dos transiciones sucesivas de los espines, (es decir, \( \gamma >> \omega \)), entonces las densidades uniformes de los intermediarios se ajustan adiabáticamente a la configuración de los espines, de forma que

\[ n_\mu = \frac{1}{N} \sum_i (\xi'^\mu S_i + 1) \]

que sustituida en la expresión 15.10.2 proporciona

\[ \Phi_i(S_i \rightarrow S'_i) = \omega \exp \left[ \frac{1}{\theta N} \sum_j \sum_\mu \xi'^\mu \xi'^\mu_j (S'_j - S_i) \right] \]

Esta expresión coincide con el ritmo de transición de los espines en el modelo de Hopfield probabilístico con la regla de aprendizaje hebbiana. Así pues, en el límite \( \gamma >> \omega \) el modelo con intermediarios se reduce al modelo de Hopfield estándar.

Es interesante considerar el límite opuesto, \( \gamma << \omega \), en donde las transiciones de los espines entre estados son extremadamente rápidas. En este caso, son los espines los que se ajustan adiabáticamente a la población de intermediarios, la cual determina el campo efectivo. La distribución de probabilidad de Boltzmann para un espín \( S_i \) en un campo dado \( h_i \) a temperatura \( \theta \) es

\[ p(S_i) = \frac{\exp(h_i S_i/\theta)}{\exp(-h_i/\theta) + \exp(h_i/\theta)} \]

Dado que los tiempos característicos de los intermediarios son muy largos, éstos únicamente reaccionarán a los valores promedio de las variables de espín, las cuales coinciden con los promedios estadísticos \( < S_i > \) estimados a partir de la distribución de Boltzmann, es decir

\[ < S_i > = \sum_{S_i = \pm 1} S_i p(S_i) = \tanh \left( h_i / \theta \right) \]

Sustituyendo el valor medio en 15.10.3 y considerando la expresión dada para el campo efectivo, se obtiene la siguiente ecuación cinética para las concentraciones de intermediarios,

\[ \dot{n}_\mu = -\gamma n_\mu + \gamma \frac{V}{N} \sum_i \left\{ \xi'^\mu \left( \tanh \left[ \frac{1}{\theta} \sum_\nu \xi'^\mu_\nu (n_\nu (R) - 1) \right] \right) + 1 \} \delta(r - R_i) + D \Delta n_\mu \]  \hspace{1cm} (15.10.4)
Hemos derivado por tanto un sistema de ecuaciones de reacción-difusión multicomponentes que describen la evolución de las densidades de mediadores asociados con las diferentes memorias almacenadas. Discutiremos ahora sus soluciones en algunos casos particularmente relevantes.

Introduzcamos las nuevas variables $m_\mu = n_\mu - 1$ para escribir las ecuaciones 15.10.4 en la forma

$$
\dot{m}_\mu = -\gamma m_\mu + \gamma \frac{V}{N} \sum_i \xi_i^\mu \tanh \left[ \frac{1}{\theta} \sum_\nu \xi_i^\nu m_\nu(r) \right] \delta(r - R_i) \\
+ \gamma \left[ \frac{V}{N} \sum \delta(r - R_i) - 1 \right] + D \Delta m_\mu
$$

(15.10.5)

supongamos ahora que la difusión es muy rápida y que los intermediarios están uniformemente distribuidos en el medio. Entonces, 15.10.5 proporciona

$$
\dot{m}_\mu = -\gamma m_\mu + \frac{\gamma}{N} \sum_i \xi_i^\mu \tanh \left[ \frac{1}{\theta} \sum_\nu \xi_i^\nu m_\nu \right]
$$

A bajas temperaturas, este conjunto de ecuaciones diferenciales tiene muchos puntos fijos estables. En el límite $\theta \to 0$, tenemos tanh $(x/\theta) \approx sgn(x)$, y la ecuación anterior se convierte en

$$
\dot{m}_\mu = -\gamma m_\mu + \frac{\gamma}{N} \sum_i \xi_i^\mu \ sgn \left[ \sum_\nu \xi_i^\nu m_\nu \right]
$$

(15.10.6)

A continuación consideraremos el caso en el que todas las memorias almacenadas son ortogonales. En este caso, existen soluciones estacionarias sencillas para 15.10.6: $m_\mu = \delta_\mu \alpha$, con $\alpha = 1, 2, \ldots, M$. Estas soluciones son estables frente a pequeñas perturbaciones. Efectivamente, para $m_\mu = \delta_\mu \alpha + m'_\mu$ encontramos de 15.10.6

$$
\dot{m}'_\mu = -\gamma m'_\mu - \gamma \delta_\mu \alpha + \frac{\gamma}{N} \sum_i \xi_i^\mu \xi_i^\alpha = -\gamma m'_\mu
$$

Estos puntos fijos estables tienen una interpretación sencilla. En el equilibrio se cumple la igualdad

$$
m_\mu = \frac{1}{N} \sum_i \xi_i^\mu < S_i >
$$

y por tanto $m_\mu$ representa *solapamientos* entre la configuración $\{S_i\}$ y las diferentes memorias almacenadas, $\{\xi_i^\mu\}$. Cuando uno de los solapamientos iguale la unidad, la configuración de los espines coincide con la memoria almacenada correspondiente. Por tanto, el punto fijo $m_\mu = \delta_\mu \alpha$ corresponde a la configuración de la $\alpha$-ésima memoria. Dependiendo de las condiciones iniciales la evolución del sistema conduce a uno u otro punto fijo. Por tanto, en el límite $\gamma << w$ el sistema con intermediarios tiene las mismas propiedades dinámicas que el modelo de Hopfield.

De hecho, $m_\mu$ puede ser considerado como el *parámetro de orden* 2 del modelo de Hopfield.

El análisis anterior muestra que en ambos casos límite, $\gamma << w$ y $\gamma >> w$, el modelo local con intermediarios es efectivamente equivalente al modelo de Hopfield de memorias asociativas si la difusión es suficientemente rápida como para mantener una mezcla ideal en el volumen $V$ del sistema, es decir, si $D/\gamma >> V^{1/3}$. Un examen más sofisticado conduce a la equivalencia en las propiedades dinámicas asintóticas para cualquier relación entre $\gamma$ y $w$.

Es más, el modelo local con intermediarios se puede generalizar para otras reglas de aprendizaje que permitan escribir la matriz $[\rho_{ij}]$ en la forma

---

2Véase el capítulo sobre fenómenos críticos.
\[ J_{ij} = \sum_{\mu} \phi_{i\mu}^* \psi_{j\mu}^* \]

donde \( \{\phi_{i\mu}\} \) y \( \{\psi_{j\mu}\} \) son parejas de memorias. Entonces, \( \zeta_{i\mu}^* \) debe ser reemplazado por \( \psi_{i\mu}^* \) en 15.10.1 y por \( \phi_{j\mu}^* \) en 15.10.2.

Para resumir la discusión de esta sección, diremos que el análisis previo revela una relación íntima entre las redes neuronales y los medios activos descritos por ecuaciones de reacción-difusión multicomponentes.

### 15.11 Transiciones de fase en el cerebro

Para terminar, volvamos a la dinámica de gran escala del cerebro y a las conjeturas realizadas con anterioridad acerca de la complejidad y los puntos críticos. La dinámica de gran escala del cerebro es, como hemos visto, un fenómeno sorprendente. Su caracterización mediante medidas basadas en la teoría de sistemas no-líneales nos ha demostrado que la presencia de caos determinista es muy plausible y que, en consecuencia, una descripción teórica basada en un modelo dinámico con pocos grados de libertad (o parámetros de orden) es factible.

En la discusión acerca de las medidas de caos y dimensionalidad aplicadas a los EEG, ya introdujimos el problema de la no-estacionariedad de la actividad cerebral. Sería deseable que, para explorar las transiciones asociadas a los cambios entre (posibles) atractores, recurriéramos a situaciones experimentales adecuadamente controladas. Algunos resultados experimentales han arrojado una considerable luz acerca de la naturaleza de las transiciones entre distintos estados de actividad cerebral. Los trabajos del grupo de J. Scott Kelso y su grupo del Centro de Sistemas Complejos de la Universidad de Florida, en colaboración con Hermann Haken, del Instituto Max Planck en Stuttgart, han permitido ver cómo dichas transiciones pueden ser interpretadas consistentemente como transiciones de fase de no-equilibrio (Kelso et al., 1992; Fuchs et al., 1992; Kelso, 1995).

El experimento básico consiste en exponer al sujeo experimental a un estímulo acústico de tipo periódico y la tarea a realizar es apretar un botón entre dos tonos consecutivos. Se trata por lo tanto de llevar a cabo un patrón de coordinación sincopado. La frecuencia del estímulo al principio del experimento es de 1 Hz, incrementándose 0.25 Hz cada 10 repeticiones del estímulo. A una cierta frecuencia dada, alrededor de 2 Hz, el sujeto es incapaz de mantener la respuesta sincopada, pasando espontáneamente a pulsar el botón en fase con el estímulo, esto es: a un comportamiento sincronizado. Durante este experimento, el campo magnético fue registrado mediante un dispositivo SQUID de 37 canales dispuesto en forma circular (véase la figura 15.18).

Las propiedades dinámicas del sistema antes y después de la transición han sido exploradas en detalle por estos autores, así como las del punto en el que la transición tiene lugar. Kelso y sus colaboradores han analizado este experimento llevando a cabo distintos tipos de medidas que permiten probar la aparición de un súbito cambio en el comportamiento de la actividad cerebral. En la figura 15.19 mostramos un ejemplo de estas medidas, así como el resultado de su promedio sobre todos los canales. Se ha representado la diferencia de fase entre las señales de respuesta y las medidas obtenidas sobre la actividad cerebral (con el SQUID) relativas al estímulo de entrada. En todos los canales se percibe un cambio en cierto instante dado que es muy claro en el promedio.

Los datos experimentales anteriores, cuyo análisis ha revelado claros indicios de que la transición anterior es una verdadera transición de fase de no-equilibrio, permiten plantear un modelo teórico basado en un sistema no-lineal simple. El desarrollo de dicho modelo, llevado a cabo en otro contexto por Haken, Kelso y Bunz (Haken et al., 1985) se basa en el planteamiento del principio
Figura 15.18: Modelo de la cabeza del sujeto experimental y localización de la red de sensores. Los detectores se basan en un dispositivo denominado SQUID. El sistema incluye 37 canales, y detecta los campos magnéticos generados en el interior de la corteza cerebral. A diferencia de las medidas clásicas basadas en EEG, estos campos magnéticos pueden medirse sin ver afectado su valor por las distorsiones introducidas por el cráneo. En este tipo de medidas, el cráneo es transparente al campo magnético generado por las corrientes cerebrales. Podemos añadir que en este caso no es preciso recurrir a un punto de referencia dado (como ocurre en el caso de los EEG) para obtener una medida del campo local.

Figura 15.19: Diferencias de fase entre las señales del cerebro (registradas por el SQUID) y la respuesta manual relativas al estímulo (Fuchs et al., 1992).
de control, discutido en el capítulo 2. Recordemos que, en un sistema alejado del equilibrio, la descripción dinámica viene dada, en principio, por un conjunto de ecuaciones no-lineales

\[ \frac{dp}{dt} = f'_\mu(q_1, ..., q_n, t) \]

En las proximidades de puntos de bifurcación, sólo algunas de estas variables estarán afectadas por inestabilidades (los modos inestables) mientras que las restantes (los modos estables) serán eliminables a través de la aproximación adiabática. Las variables asociadas a las inestabilidades nos permitirán definir los parámetros de orden, que introducen la denominada causalidad circular: los parámetros de orden quedan determinados (y son creados) por la cooperación de cantidades microscópicas, pero al mismo tiempo los parámetros de orden gobiernan el comportamiento cualitativo de todo el sistema. Un problema central consiste, por tanto, en identificar cantidades macroscópicas adecuadas que nos permitan definir parámetros de orden (Haken y Stadler, 1989; para un estudio formal detallado, véase Haken, 1983).

Podemos escribir el sistema anterior en la forma

\[ \frac{dp}{dt} = N_\mu(q) + F(t) \]

donde \( q = (q_1, ..., q_n) \) y los términos \( N_\mu(q) \) y \( F(t) \) indican la interacción entre variables y el ruido asociado a las fluctuaciones internas, respectivamente. Variando el parámetro \( \mu \), un estado estable \( q_0 \) puede serse de serlo y ser reemplazado por un estado inestable que podemos representar en la forma general

\[ q = q_0 + v e^{\lambda t} \]

siendo \( v \) independiente del tiempo y determinado, como sabemos, por la ecuación lineal asociada

\[ L_\mu v = \lambda v \]

donde, en un caso general, \( v = v(r) \), (dependerá de las coordenadas espaciales), y de hecho tendremos cierta variedad de vectores propios \( \{v\} \) posibles.

El paso siguiente ahora consiste en considerar el estado (o solución) del sistema dinámico \( q(r, t) \) como una superposición

\[ q(r, t) = q_0(r) + \sum_n \xi_n(t) v_n(t) + \sum_s \xi_s(t) v_s(t) \]

en la que hemos separado los vectores inestables \( v_n(t) \), para los que \( \lambda > 0 \), de los estables, \( v_s(t) \) para los cuales \( \lambda < 0 \). Las amplitudes \( \xi_n(t) \) en la primera suma de la derecha son los parámetros de orden. Las amplitudes estables \( \xi_s(t) \) definen los modos estables. Tipicamente, lejos del equilibrio, el número de parámetros \( \xi_s \) suele ser grande mientras que el número de parámetros de orden \( \xi_u(t) \) será muy pequeño, quedando en ocasiones reducido a uno solo, tal y como predice el principio de control de Haken. En estas condiciones el comportamiento del sistema (en principio de alta dimensión) puede ser reducido a las ecuaciones para los parámetros de orden, i. e.

\[ \frac{d\xi_u}{dt} = N(\xi_u) + F \]

que describirán el sistema a una escala macroscópica (en lugar de la microscópica, definida por las ecuaciones de partida)³.

³Notemos que este análisis tiene profundas similitudes con el análisis de la estabilidad de las estructuras disipativas en sistemas de reacción-difusión, analizados en el capítulo 10.
Los parámetros de orden pueden competir o cooperar. Cuando compiten, sólo uno de ellos sobrevive. Así ocurre por ejemplo en el experimento de convección de Bénard (discutido en el capítulo 5, apéndice) cuando se forma un rollo convectivo. Cada rollo en el fluido posee un movimiento propio dominado por un parámetro de orden específico. El resultado final, en el que el sistema ha "decidido" por rotura de simetría la configuración final de rollos convectivos, es el resultado de la competencia entre parámetros de orden. Las fluctuaciones tienen, por lo tanto, un papel básico en la decisión final acerca de qué parámetros de orden subsisten.

El proceso de rotura de simetría puede representarse, como sabemos, mediante un potencial $V_\mu(\xi)$ que, bajo ciertas condiciones, experimentará un cambio cualitativo cuando el parámetro de orden cruce un punto de bifurcación $\mu_c$. Tendremos una ecuación dinámica (de tipo gradiente) para $\xi$ dada por

$$\frac{d\xi}{dt} = -\frac{\partial V_\mu(\xi)}{\partial \xi} + F(t)$$

la cual define un decaimiento de $\xi$ hacia el equilibrio cuando $\mu < \mu_c$. Para valores cercanos al punto crítico, las fluctuaciones asociadas al sistema se amplificarán hasta hacerse divergentes en $\mu_c$. Este fenómeno, conocido como relajación crítica (critical slowing down) caracteriza la aparición de cierto tipo de transiciones, y aparece en particular en los experimentos de Kelso. Más allá de $\mu_c$, el parámetro de orden puede tomar distintos valores por medio de procesos de rotura de simetría, que pueden hacerse más complejos a medida que atravesamos distintos puntos de bifurcación (Haken, 1983).

Kelso y sus colaboradores han empleado el conjunto de medidas obtenidas mediante el SQUID para estudiar detalladamente el comportamiento temporal de la dinámica de los modos espaciales de la actividad cerebral. La señal espaciotemporal, que indicaremos por $H(r,t)$ se descompone en funciones $\Phi_i(r)$ de tal forma que podemos escribir

$$H(r,t) = \sum_{i=1}^{N} \xi_i(t) \Phi_i(r)$$

siendo $\xi_i(t)$ las amplitudes asociadas (en cada instante) a las funciones $\Phi_i(r)$. Para una elección adecuada de $\{\Phi_i(r)\}$, la serie anterior puede truncarse a $N = 5$, dando una muy buena aproximación a los datos experimentales. Existen distintos procedimientos para elegir estas funciones espaciales, entre los que destaca el desarrollo estándar en serie de Fourier. Kelso emplea en su estudio la denominada descomposición de Karhunen-Loève (KL) que minimiza el error definido por la expresión integral

$$E_K = \int_T dt \int dr \left\{ H(r,t) - \sum_{i=1}^{N} \xi_i(t) \Phi_i(r) \right\}^2$$

para cada $K$ (que define el orden del truncamiento).

A lo largo del tiempo, las amplitudes asociadas a cada modo se van modificando con la dinámica. Las dos primeras, $\{\Phi_1(r), \Phi_2(r)\}$ dan cuenta del 75 por ciento de la señal obtenida y el primero de los modos representa de hecho el 60 por ciento de dicha señal. Tomando esta amplitud $\xi_1(t)$ como parámetro de orden, podemos caracterizar muy bien la transición. Un estudio del comportamiento de esta amplitud revela que presenta una caída dramática en el punto crítico, tal y como esperaríamos en una verdadera transición de fase, y experimenta relajación crítica cerca de ese punto.

Pero el análisis de la dinámica de esta amplitud aún ha resultado más sorprendente. Antes y después de la transición $\xi_1(t)$ presenta un comportamiento dinámico claramente distinto, como vemos en la figura 15.20 (a), en la que representamos las series temporales medidas (Kelso, 1995). Si
Figura 15.20: Comparación entre la dinámica de la amplitud $\xi_1(t)$ asociada al primer modo en la descomposición KL y el modelo de Kelso-Fuchs (Kelso, 1995). (a) Series temporales medidas y (b) simuladas antes de la transición así como las (c) posteriores a ésta y (d) junto al espectro de Fourier correspondiente. (e,f) Atractores reconstruidos a partir de los datos anteriores antes y Después de la transición ((e) datos experimentales; (f) atractores simulados).
representamos además los atractores reconstruidos, vemos también una clara transición dinámica entre el estado anterior y el posterior al fenómeno crítico. El segundo atractor es, de hecho, un buen ejemplo de caos de tipo Shilnikov (capítulo 5). A partir de estos datos, Kelso y sus colaboradores han obtenido un modelo teórico que reproduce notablemente bien no sólo las propiedades cualitativas de la transición, sino también algunas características cuantitativas, como el espectro de Fourier asociado (figura 15.20).

Estos resultados y otros obtenidos de forma independiente por otros autores indican que el cerebro es un sistema autoorganizado que opera en las proximidades de puntos críticos, permitiéndole posiblemente saltar de forma flexible y espontánea de un estado coherente (attractor) a otro (Kelso et al., 1992). Esta conjetura ha recibido un notable soporte teórico en el desarrollo llevado a cabo partiendo del formalismo de la Sinergética (Haken, 1983). Se ha obtenido un modelo simple (Jirsa et al., 1994) que permite interpretar los resultados anteriores en términos de la dinámica de un parámetro de orden que experimenta una transición de fase de no-equilibrio.

Para terminar, señalemos que esta aproximación se halla bastante alejada de otras, basadas implícitamente en la separación de la actividad del cerebro descrita como compartimentos separados. Así, la denominada tomografía por emisión de positrones (TEP) en la que se presenta la "actividad" del cerebro durante una determinada tarea (como leer, oir música o calcular) está basada en métodos de sustracción de dudosa significación. Así, por ejemplo, se compara la actividad del cerebro en dos casos: leer y hablar se compara con leer solamente. Se toman las imágenes correspondientes a ambas situaciones, y se lleva a cabo la diferencia. Pero la diferencia es una operación lineal, no lo olvidemos. Si bien es posible que ciertas zonas del cerebro se hallen más activas, las tareas cognitivas no están aisladas en compartimentos, sino distribuidas, en mayor o menor grado, en la actividad cerebral global (Kelso, 1995). Decir (como se hace a menudo) que se "ha fotografiado", pongamos por caso, el "centro de la música" suena muy bien, pero no es más que una cara nueva de un reduccionismo con rostro tecnológico.

Bibliografía


Capítulo 16

Redes Neurales Fluidas

A lo largo de los capítulos precedentes, hemos visto cómo los sistemas complejos exhiben propiedades emergentes. La interacción no-lineal entre elementos simples puede dar lugar a fenómenos en una escala de organización superior. Estos pueden ser la aparición de estructuras ordenadas cerca de puntos de bifurcación, de estructuras fractales o de comportamientos temporales periódicos o caóticos.

Existe cierta familia de procesos que, genéricamente, podemos agrupar bajo el nombre de *computación*. La idea de cálculo es fácilmente asimilable en términos de dispositivos artificiales: un ordenador lleva a cabo un cálculo procesando información de entrada y generando una salida, siguiendo para ello un conjunto de instrucciones especificadas. Este proceso de computación puede, en último término, reducirse a la manipulación de un conjunto de símbolos pertenecientes a un alfabeto binario $\Sigma = \{0, 1\}$. Como ya vimos en el capítulo dedicado a los autómatas celulares, podemos formalizar el problema de la computación mediante la máquina de Turing.

Detengámonos a reflexionar acerca de este problema. Parece claro (aunque tal vez no lo sea tanto) que en algunos sistemas naturales se llevan a cabo procesos dinámicos que podemos identificar, externamente, como procesos de cálculo. De forma esquemática, podemos imaginar estos sistemas como capaces de ser modificados por ciertas señales procedentes del medio y de cambiar posteriormente de estado para “responder” a la información de entrada. El cerebro procesa información, de alguna forma, y aunque dicho procesamiento parte -en principio- de la percepción del entorno, también se lleva a cabo sin necesidad de que dicha información se halle presente. Aún así, podemos quedarnos con la visión simplista de estímulo-respuesta en la que la percepción proporciona las entradas, que son procesadas y que, finalmente, generan una respuesta.

Existe otro tipo de sistemas complejos capaz de llevar a cabo procesos de computación que incluyen la memoria asociativa y el reconocimiento de patrones. A este grupo pertenecen el sistema inmunitario y los insectos sociales, de los que hablaremos en este capítulo (Wilson, 1971; Hölldobler y Wilson, 1990). Aunque popularmente los individuos de estos grupos son vistos como dotados de una inteligencia especial, la realidad nos muestra una evidencia bien distinta: los individuos como tales son realmente simples aunque, colectivamente, las sociedades puedan llevar a cabo actividades de gran complejidad.

De hecho, junto con la metamorfosis y el vuelo, la aparición de estas sociedades ha sido una de las grandes revoluciones experimentadas a lo largo de la evolución de los insectos. Las especies sociales aparecen en todos los hábitats del mundo bajo condiciones enormemente diversas. Encontramos hormigas en zonas desérticas y en lugares helados. En las selvas tropicales, donde su papel ecológico es muy importante, hormigas y termitas suman la tercera parte de la biomasa animal total existente.

¿Qué tiene de especial la sociedad por encima del comportamiento individual?

Parte de la respuesta tiene mucho que ver con las propiedades de los sistemas neurales, que
Figura 16.1: Hormigas y cerebros: aunque las diferencias entre ambos sistemas son patentes, comparten probablemente más puntos en común que diferencias.

hemos estudiado en el capítulo anterior, y con los que los insectos sociales comparten numerosas propiedades comunes. Las similitudes entre las hormigas (como colectivo) y el cerebro son muchas y han sido discutidas en el pasado por diversos autores (Hofstadter, 1979; Gordon et al., 1992; Solé et al., 1993a, 1993b, 1995; Adler y Gordon, 1992). Entre otras, destacaremos las siguientes:

- El comportamiento de los elementos aislados (hormigas o neuronas) no nos da prácticamente ninguna información acerca de cómo funciona el hormiguero o el cerebro. De alguna forma, los fenómenos de interés son generados por medio de la interacción, dando lugar a estructuras a una escala muy superior a la de los elementos implicados.

- Las propiedades computacionales observadas son el resultado de la interacción de los individuos no solo con su entorno, sino entre sí.

- La destrucción parcial de una fracción del colectivo (la desaparición de algunas hormigas o de algunas neuronas) no modifica, en general, la habilidad colectiva del sistema.

Parece claro que la colonia de hormigas puede visualizarse como un sistema distribuido, en el que la información procedente de distintos puntos del espacio es transmitida al sistema. Una vez (o simultáneamente) que la información llega, es procesada colectivamente y el colectivo responde de alguna forma. Algunos de estos procesos son simples: un individuo detecta una fuente de alimento, transmite esta información y el colectivo responde explotando dicha fuente. Sin embargo (en general) las situaciones habituales serán más complejas. La colonia puede tener que “decidir” entre varias fuentes de alimento distintas. Puede ser necesario llevar a cabo varios procesos de análisis de información externa de distintos tipos. Además, la propia estructura interna de la colonia requiere un “conocimiento” del estado interno para disponer de los recursos adecuadamente.

\[1\] Estrictamente hablando, deberíamos matizar esta afirmación. Por una parte, estamos suponiendo sociedades de insectos de cierto tamaño. Existen especies de hormigas con colonias de tamaño muy reducido en el que la complejidad de los individuos define en buena medida la del colectivo, si bien estos grupos son de hecho los menos evolucionados.
Figura 16.2: Hormigas y propiedades emergentes: la actividad del nivel inferior, en nuestro caso la escala individual de las hormigas, genera el parámetro de orden (la actividad global) que pasa a afectar a los propios individuos.

así como para mantener en buen estado el nido. El propio proceso de construcción de un nido implica la emergencia de un orden de gran escala que parte de la interacción entre elementos que reciben información básicamente local. Aquí, nuevamente, la idea de propiedad emergente aparece claramente definida.

En la figura 16.2 resumimos la existencia de comportamiento emergente en la estructura de una sociedad de insectos. Siguiendo la propuesta de Hermann Haken (1977, 1988; capítulo 2), el comportamiento colectivo, que podemos imaginar como la actividad del grupo, actúa como un parámetro de orden (adecuadamente definido) sostenido a partir de la actividad de los individuos, a la vez que controla el comportamiento de éstos. Aparece una vez más la causalidad circular en la que los elementos del sistema definen (y son modificados por) el comportamiento colectivo, que no es reducible al de los elementos.

En este caso, el comportamiento global que emerge de las interacciones locales se identifica con un proceso de computación. Este hecho ha llevado a acuñar la definición de computación emergente (Forrest, 1990). La premisa de la computación emergente es que pueden construirse sistemas computacionales interesantes y útiles explotando la interacción entre componentes simples. En algunos casos (como en la modelización del comportamiento inteligente) podría ser, de hecho, el único método factible.

Los requerimientos mínimos para la aparición de computación emergente parten de la presencia de cierto tipo de información colectiva que se halla ausente en los niveles inferiores. Los constituyentes serían (Forrest, 1990):

- Un grupo de agentes, cada uno capaz de seguir cierta colección de instrucciones (Ω) explícitamente definidas.

- Interacciones entre los agentes (de acuerdo con Ω) que generarían patrones globales a un nivel macroscópico, esto es, un epifenómeno.

- Una interpretación natural del epifenómeno en términos de computación.
Figura 16.3: Proceso de ordenamiento en el modelo de Denebourg (véase texto). A partir de un conjunto de objetos repartidos al azar, (a) las "hormigas" (no mostradas aquí) van desplazando los objetos siguiendo unas reglas muy simples, de tipo local. En pasos sucesivos, vemos cómo se van separando los distintos tipos de objetos en grupos bien definidos. Si esperamos lo suficiente, al final sólo se forman dos grupos.

Volviendo al tratamiento general de los procesos de autoorganización, la aparición de cálculo emergente implica la generación espontánea de orden a partir de un sistema inicialmente desordenado (al menos hasta cierto punto) y con elementos que, en general, no serán fiables. Aquí por fiabilidad entendemos que cometerán errores o mostrarán algún tipo de aleatoriedad. Podemos entonces emplear la idea de orden por fluctuaciones (Prigogine y Stengers, 1984; Nicolis y Prigogine, 1994) como punto de partida.

16.1 Dinámica de la distribución colectiva

Un primer ejemplo de modelo de comportamiento colectivo en grupos de insectos sociales, que presentaremos en esta sección, intenta explicar de forma simple y emergente cómo los insectos sociales, por ejemplo las hormigas, llevan a cabo una distribución efectiva de los distintos objetos que se encuentran en el interior del nido (Denebourg et al., 1992). En un hormiguero, hallaremos las larvas, huevos o semillas distribuidos en lugares concretos bien definidos. Aunque al observar un individuo que acarrea uno de estos objetos nos dará posiblemente la impresión de que lo hace con cierto grado de aleatoriedad (una hormiga puede de repente coger una semilla, desplazarla durante un ratito y dejarla en otra posición) es evidente que, de un modo u otro, el sistema logra ordenar espacialmente estos objetos de forma eficiente. Si experimentalmente manipulamos un nido de hormigas y mezclamos al azar los distintos objetos, veremos que, con el tiempo, las hormigas volverán a separar los objetos en sus clases anteriores.

El algoritmo de comportamiento que describiremos (Denebourg, 1992) no es tanto un modelo de la forma real en que las hormigas actúan, sino una demostración de que semejante algoritmo local es posible y efectivo. El modelo se basa en un conjunto de individuos que se desplazan al azar
sobre una superficie bidimensional. En el instante inicial un cierto número de hormigas/autómatas se halla distribuido al azar junto con un conjunto de elementos, también azarosamente repartidos. Estos elementos son de dos clases, y se indican por círculos blancos o negros según el caso. Cada individuo coge al azar uno de estos objetos si se encuentra aislado, llevándolo consigo y depositándolo otra vez en otro lugar. El modelo añade adicionalmente una memoria a corto plazo para cada individuo, de forma que éste puede "recordar" qué tipo de objetos ha ido encontrando durante las últimas $m$ iteraciones. Si indicamos por $A$ y $B$ los dos tipos, un individuo podría retener una secuencia digamos de $m = 10$ iteraciones, $00AB0AA0B0$ en la que los ceros indican ausencia de objetos. A partir de esta secuencia, el individuo posee una información aproximada (de carácter local) acerca de la abundancia de objetos del mismo tipo que ha ido encontrando. Para la secuencia anterior, la frecuencia de encuentros con objetos de tipo $A$ será $f_A = 3/10$ y la de objetos de tipo $B$ será $f_B = 2/10$. La probabilidad de coger un objeto si el autómata no lleva ninguno es

$$P_+ = \left[ \frac{k^+}{k^+ + f} \right]^2$$

siendo $f$ la fracción de lugares cercanos ocupados por el mismo tipo de objetos (definido previamente) y $k^+$ una constante. Vemos por lo tanto que esta probabilidad decrece con la cantidad de objetos cercanos, luego es tanto mayor cuanto más aislado esté el objeto y tanto menor cuanto más objetos del mismo tipo se hallen en la proximidades. La decisión de abandonar el objeto depende también del número de objetos de la misma clase que se encuentran, pero de forma que ahora, cuanto mayor sea dicho número, más probabilidades hay de que abandone el objeto. Específicamente, estos autores emplean

$$P_- = \left[ \frac{f}{k^- + f} \right]^2$$

como probabilidad de abandonar el objeto, siendo $f$ el número de objetos del mismo tipo y $k_-$ una nueva constante.

En la figura 16.3 vemos cuatro instantáneas del proceso de clasificación, generado a partir de las reglas locales anteriores. Vemos claramente que surgen estructuras diferenciadas que incluyen objetos de tipos distintos, y que su tamaño y agregación aumentan con el tiempo. Este es precisamente el proceso que observamos en colonias de hormigas que, eventualmente, podemos someter a manipulación en el laboratorio de forma que llevemos a cabo una homogeneización de objetos que los individuos deberán separar posteriormente. El proceso implica un feedback positivo (una vez más, fenómenos no-lineales) que permite la creación de estructuras a partir del desorden y las aleatoriedades que rigen a nivel microscópico. Una vez que algunas pequeñas fluctuaciones se han visto amplificadas, el proceso se autoencuentra hasta obtener los agregados observados.

### 16.2 Comportamiento probabilista: la estrategia del error

El error cometido por los elementos de un sistema complejo es considerado, en general, como una fuente de ruido no deseable. El ruido se identifica, en este sentido, con la incapacidad de llevar a cabo, de manera fiable, cierta tarea o ejecución de instrucciones. Aunque esta intuición parece muy clara, ciertas especies de hormigas explotan el error intrínseco a sus elementos como fuente de exploración del medio. En este sentido, como ha demostrado J. L. Deneubourg y sus colegas de la Universidad Libre de Bruselas (Deneubourg et al., 1986, 1989) el error asociado al ruido interno del sistema puede ser empleado como capacidad de exploración en un ambiente variable. La aleatoriedad puede tener una ventaja adaptativa para las hormigas.
Figura 16.4: (a) Evolución del número de hormigas en las proximidades de dos fuentes de alimento de distinta riqueza ($S_2 > S_1$) introducidas en instantes distintos. (b) Idem. para dos fuentes introducidas simultáneamente (Deneubourg et al., 1986).

Imaginemos un grupo de hormigas que lleva a cabo la búsqueda de alimento y su explotación una vez encontrado. En un ambiente predecible, donde existen fuentes estables de alimento, la mejor estrategia es localizar dichas fuentes y llevar a cabo su explotación de forma estable. La situación es más compleja si el ambiente (por tanto las fuentes de nutrientes) es más impredecible. En este caso, el colectivo debe llegar a un compromiso: debe explotar de forma efectiva las fuentes detectadas, a la vez que necesita mantener la exploración activa. Puesto que el segundo proceso, en el que se descubren nuevas fuentes a través de una búsqueda individual aleatoria, depende del azar, y el primero de una respuesta colectiva fiable, orden y desorden (una vez más) deben hallarse presentes.

Consideremos el siguiente problema: colocamos dos fuentes de alimento, S1 y S2, en las proximidades de una colonia de hormigas. Supongamos que primero colocamos una (S1) que es descubierta por las hormigas, las cuales inician su explotación. El número de hormigas cerca de esta fuente de alimento crecerá con el tiempo. Ahora, cuando la explotación de la fuente S1 ya se ha estabilizado, introducimos la fuente S2 que es más rica en alimento. Esta fuente será descubierta si existen hormigas que, de forma aleatoria, abandonan la hilera de individuos que se hallan explotando la fuente S1, de forma que puedan eventualmente descubrir otra fuente mejor. Cuando así ocurre, vemos experimentalmente (figura 16.4 (a)) que el número de hormigas en las proximidades de S2 aumenta rápidamente, mientras que cerca de S1 disminuye. Las hormigas de hecho depositan mayor cantidad de marcador químico (que las demás seguirán y reforzarán) en función de la riqueza de la fuente explotada. Así, pese a que los individuos llevan a cabo una interacción local, crean de hecho estructuras globales (la pista química que emerge a medida que por ella pasan más individuos) que a su vez pasan a controlar la actividad de los individuos. Si suponemos que S1 y S2 denotan además las concentraciones de cada fuente, vemos que las hormigas han llevado a cabo, colectivamente, un proceso de cálculo: decidir si $S_1 > S_2$, algo que los individuos obviamente no pueden realizar de forma individual. Se trata por lo tanto de un ejemplo simple pero claro de cálculo emergente.

Existe además un caso adicional muy interesante para ver el efecto de las no-linealidades en acción. Supongamos que mostramos al hormiguero dos fuentes exactamente iguales ($S_1 = S_2$)
Figura 16.5: Soluciones estacionarias para el modelo con dos fuentes, que nos da el número de hormigas en una fuente en función del tamaño de la colonia (en realidad, del número potencial de posibles recolectores, que será en principio proporcional al tamaño de la colonia).

situadas a distancias idénticas. En la figura 16.4 (b) vemos el resultado de la dinámica obtenida en el experimento: pese a la simetría del problema, las hormigas rompen la simetría y explotan preferentemente una de las fuentes de alimento (véase el capítulo 4, bifurcaciones). De hecho, el grado de explotación de la segunda nos dará una medida del grado de ruido presente en el sistema. Las pequeñas fluctuaciones, en todo caso, deciden fortuitamente cuál de las dos fuentes se explota. Vemos por lo tanto que el sistema es capaz de tomar una decisión, vía rotura de simetría, cuando las alternativas son idénticas.

Los resultados anteriores pueden resumirse mediante el siguiente modelo matemático

\[
\frac{dX_1}{dt} = a \cdot X_1 \cdot \mu(N - X - E) - b \cdot X_1 + c \cdot E
\]

\[
\frac{dX_2}{dt} = a \cdot X_2 \cdot \mu(N - X - E) - b \cdot X_2 + c \cdot E
\]

\[
\frac{dE}{dt} = a \cdot X(1 - \mu)(N - X - E) - p \cdot E - 2 \cdot c \cdot E
\]

donde indicamos por \( X = X_1 + X_2 \), siendo \( X_i \) la cantidad de hormigas sobre la fuente \( i \)-ésima, \( E \) es el número de "hormigas perdidas" (que no siguen las marcas químicas colectivas y se desplazan al azar) \( N \) es el número (potencial) de hormigas que se dedican a la recolección y \( a \) es el número de hormigas "reclutadas" por hormiga, por unidad de tiempo.

De las \( a \cdot X_i \cdot \mu(N - X - E) \) hormigas reclutadas, cierta fracción \( \mu \in [0, 1] \) alcanza la fuente, y una parte \( 1 - \mu \) se pierde. Las hormigas que alcanzan la fuente permanecen un tiempo promedio \( b^{-1} \) cerca de la fuente. Las hormigas perdidas, \( E \), tienen una probabilidad \( c \) de encontrar la fuente (por unidad de tiempo) o bien volverán al nido después de \( p^{-1} \) unidades de tiempo.

Si el reclutamiento depende de la densidad de elementos (como es razonable suponer), podemos introducir una dependencia natural en la forma

\[
\mu(X_i) = \frac{X_i}{g + X_i}
\]
que permite reproducir muy bien los resultados experimentales, incluyendo las roturas de simetría antes mencionadas. Su dependencia respecto del número de hormigas es especialmente interesante, como vemos en la figura 16.5. Indicamos por $S^+$ la solución simétrica, que se bifurca para cierto $N_c$ en dos ramas posibles, $A^+, A^-$. Esta figura se ha obtenido para $g = 24.3$, $b = 0.1$, $p = 0.033$, $a = 0.001$ y $c = 0.018$. Para valores pequeños de $N$, ambas fuentes se explotarían de forma simétrica, pero esta simetría se rompe más allá de cierto valor crítico $N_c$.

### 16.3 Termitas y orden por fluctuaciones

En 1977, Jean-Louis Deneubourg propuso el primer modelo de comportamiento colectivo en insectos sociales basado en un modelo matemático no-lineal (Deneubourg, 1977). Este problema es bien conocido por los estudiosos: el problema de la construcción de nidos en termitas. Los nidos de termitas pueden poseer dimensiones enormes, de varios metros de altura y varias toneladas de peso. En su interior, encontraremos cavidades de tamaños característicos muy superiores a los de los individuos que los construyeron, que interaccionan de forma local. El modelo que vamos a describir intenta explicar de forma simple el mecanismo mediante el cual semejantes estructuras coherentes pueden emerger de la interacción, aparentemente desordenada, entre los individuos.

Las hipótesis del modelo son las siguientes:

- El material que forma el termitero, que es manipulado por los individuos, adquiere como consecuencia de esta manipulación cierto olor o marca química característica (mediada por una feromona). Indicaremos por $P(r,t)$ la cantidad de material de construcción marcado en la posición $r$ y el instante $t$. La feromona puede disociarse de su soporte sólido; sea $H(r,t)$ la concentración de feromona libre, y supongamos que el material de construcción carece de las propiedades marcadoras de la feromona.

- $H$ puede difundirse libremente en el medio circundante de forma que se cree un gradiente de concentración local.

- Supondremos que el número de elementos “activos”, esto es, termitas que transportan material, presenta un flujo constante $\Phi$, y sea $C(r,t)$ su densidad.

- Los insectos portadores de material son atraídos positivamente por el gradiente de feromona de forma que tienden a desplazarse a las zonas de mayor concentración (presentan, por lo tanto, quimiotropismo). Esta orientación estará en principio compitiendo en todo momento con los efectos aleatorios del desplazamiento.

- El depósito de material en un punto del espacio es proporcional al número de insectos activos en dicho punto. Supondremos que no existe una interacción directa entre el material y las termitas.

Las hipótesis previas nos llevan a construir el siguiente modelo, basado en tres ecuaciones de reacción-difusión

$$\frac{\partial P(r,t)}{\partial t} = k_1 C - k_2 P$$

(16.3.1)

en ésta indicamos que la cantidad de material marcado depende de la abundancia local de individuos activos (que transportan material) proporcionalmente. El marcador se degrada (término de decaimiento local) según

$$\frac{\partial H(r,t)}{\partial t} = k_2 P - k_4 H + D_h \nabla^2 H$$

(16.3.2)
Figura 16.6: Estructura ordenada generada por interacción entre termitas, calculada a partir del modelo de Deneubourg (Bonabeau y Theraulaz, 1994).

que nos indica, razonablemente, que la cantidad de feromona libre crece con la cantidad de material marcado (de la que la feromona se disocia) y decrece por degradación proporcionalmente a su abundancia local. El último término indica la difusión pasiva de la feromona libre hacia posiciones vecinas. Finalmente, la ecuación

$$\frac{\partial C(r,t)}{\partial t} = \Phi - k_1 P + D_h \nabla^2 C + \gamma \frac{\partial}{\partial r} \left( C \frac{\partial H(r,t)}{\partial r} \right) \quad (16.3.3)$$

introduce la dinámica espaciotemporal del número de elementos activos, que posee un término de flujo constante $\Phi$, un término lineal de decaimiento (asociado a los individuos activos que dejan de serlo al abandonar el material) y finalmente un término de difusión junto a un término de movimiento no pasivo, que depende del gradiente de concentración de hormona $\partial H/\partial r$. El coeficiente $\gamma$ es especialmente importante, dado que introduce la intensidad con la que los individuos detectan y siguen el gradiente. Este coeficiente se llama constante quimiotáctica. Aquí el operador Laplaciano es $\nabla^2 \equiv \partial^2/\partial r^2$. Analizaremos el comportamiento de este modelo siguiendo el tratamiento que ya vimos en el capítulo sobre estructuras de Turing.

El modelo que constituyen 16.3.1. 16.3.2 y 16.3.3 posee una solución estacionaria, dada por:

$$(C_s, H_s, P_s) = \left( \frac{\Phi}{k_1}, \frac{\Phi}{k_4}, \frac{\Phi}{k_2} \right)$$

a partir de la cual consideraremos las perturbaciones del estado espacialmente homogéneo.

$$C(r,t) = C_s + c(r,t)$$
$$H(r,t) = H_s + h(r,t)$$
$$P(r,t) = P_s + p(r,t)$$
La matriz de Jacobi asociada es, \( L_{\mu} \), calculada en \((C_*, H_*, P_*)\),

\[
L_{\mu} = \begin{pmatrix}
k_1 + Dk^2 & \gamma k^2 C_* & 0 \\
0 & k_1 + D_h k^2 & -k_2 \\
-k_1 & 0 & k_2
\end{pmatrix}
\]

El determinante \( \det[L_{\mu} - I] = 0 \) proporciona la ecuación característica

\[
P(\lambda) = \lambda^3 + A\lambda^2 + B\lambda + E = 0
\]

La ecuación que sigue, para la amplificación de las perturbaciones, se obtendrá, tal y como vimos en el capítulo 10, de la condición:

\[
DD_h k^4 + k^2(D_h k_4 + D_h k_1 + \gamma \Phi) + k_1 k_4 = 0
\]

donde \( k = (2\pi n/L) \). Como sabemos, \( n \in \mathbb{Z} \) será un entero que nos dará el número de máximos observados en un sistema de longitud \( L \). Esta última expresión puede ser reescrita en la forma

\[
\gamma = -\frac{DD_h k^4 + k^2(D_h k_4 + D_h k_1) + k_1 k_4}{\Phi k^4}
\]

y presenta su máximo valor para

\[
k_c = \left( \frac{k_1 k_4}{DD_h} \right)
\]

Para este valor existirá un \( \gamma_c \) dado por

\[
\gamma_c \equiv \gamma(k_c) = -\frac{(D_h k_4)^{1/2} + (D_h k_1)^{1/2})^2}{\Phi}
\]

el cual define la frontera de aparición de estructuras espaciales macroscópicas. Un ejemplo de las estructuras obtenidas, calculado por integración numérica del sistema anterior, se muestra en la figura 16.6.

Así pues, para ciertas combinaciones de parámetros esperaremos encontrar estructuras macroscópicas que emergen de la interacción entre elementos que sólo obtienen información de forma local. Este es el principio básico en el que descansa la capacidad colectiva de construcción que observamos en los insectos sociales (Bonabeau y Theraulaz, 1994). Existen otros muchos ejemplos de interés en los que semejantes estructuras adquieren una gran sofisticación, como es por ejemplo la construcción de panales en las abejas. Existe una enorme regularidad en la generación tanto de las estructuras hexagonales que definen las celdas como en otras de mayor escala, como las estructuras paralelas en que se organizan los panales, de gran tamaño en relación al de los individuos. En este último caso, destaca el estudio llevado a cabo por el grupo de Deneubourg (Skarka et al., 1990) que considera el mecanismo de formación de panales basado en las interacciones entre individuos así como en las que se producen entre éstos y la cera. Todas estas interacciones son de naturaleza local. Al igual que ocurre en el caso de las termitas, el material actúa atraír a las abejas, y se produce un feedback positivo entre ambos mecanismos básicos. Las abejas depositan cera, cuya cantidad local depende de la interacción mutua entre individuos, que a su vez se ve modificada por la cera presente. Construyendo un sistema de tres ecuaciones de reacción difusión, estos autores obtienen soluciones para estas que muestran la estructura paralela antes mencionada (figura 16.7) y que permite interpretar distintos experimentos llevados a cabo sobre colmenas reales.
Figura 16.7: Crecimiento paralelo de paneles obtenido a partir del modelo de Skarka et al., (1990). A la derecha se indica la cantidad local (sobre un espacio bidimensional) de individuos (abejas) y a la izquierda la correspondiente cantidad de cera, lo cual proporciona una idea del tipo estructuras formadas.

Para terminar, señalemos que otros tipos de estructuras emergentes muy distintas, de tipo dinámico, pueden explicarse fácilmente a través de las aproximaciones anteriores. Entre ellas destaca la formación de patrones de búsqueda observados en las hormigas legionarias, muy comunes en distintas zonas de la Tierra. Estos insectos llevan a cabo la exploración masiva de su territorio cubriendo a lo largo de un día un área de hasta 1000 m². La forma del flujo de hormigas en exploración (que constituye toda la colonia, ya que no poseen ninguna estructura fija en la que refugiarse) es muy compleja, ramificándose en estructuras de gran detalle. Estas estructuras sugieren algún tipo de control central que, sin embargo, es totalmente inexistente, como debería deducirse de todo lo anteriormente expuesto. Las hormigas son prácticamente ciegas y sólo interaccionan entre sí mediante señales químicas locales y contactos de individuo a individuo. Sin embargo, un patrón macroscópico muy coordinado emerge de manera espontánea. Se han ideado modelos muy simples, basados en automatas que se desplazan siguiendo reglas mínimas, observadas en el estudio de los individuos en la selva. Dichos modelos proporcionan patrones de exploración muy similares, que nos muestran otra vez las capacidades de la autoorganización colectiva (Deneubourg et al., 1989).

16.4 Oscilaciones y redes neurales fluidas

Un fenómeno especialmente sorprendente de autoorganización en colonias de hormigas se da en el género *Leptothorax*. Estos hormigueros son de tamaño reducido (con unos 100-200 individuos como tamaño típico) y las especies de este género aparecen en todos los hábitats del mundo. El fenómeno en cuestión tiene que ver con la forma en que la actividad de los individuos se distribuye a lo largo del tiempo. El estudio del interior de los nidos llevado a cabo en el laboratorio demuestra que la actividad de la colonia es periódica, con una periodicidad de unos 20-30 minutos (Franks et al., 1990). En estos intervalos, se observa una oscilación aproximada del número total de hormigas activas (en movimiento, llevando a cabo algún tipo de tarea) de forma tal que en algunos intervalos
Figura 16.8: Oscilaciones colectivas de la actividad de hormigas del género *Leptothorax*. Los individuos aislados presentan un comportamiento desordenado, que no se observa en el colectivo. Para este caso, la colonia exhibe oscilaciones globales de actividad (Franks, 1990).

devimiento ningún elemento está activo mientras que en otros (tipicamente cortos) puede ocurrir que todos los individuos mantengan su actividad. Este resultado es sorprendente si tenemos en cuenta que, hasta ahora, los sistemas vivos en los que encontrábamos oscilaciones eran tales como el corazón, el cerebro, etc. ¿De dónde proceden estas oscilaciones? y, lo que es más difícil de explicar, ¿tiene alguna función?

Este comportamiento fue estudiado detalladamente por B. Cole, quien analizó además la dinámica de los individuos aislados. Podríamos esperar que las oscilaciones observadas fueran el resultado del acoplamiento entre elementos de por sí periódicos, pero no es así. Los individuos aislados, que podemos estudiar por separado, no exhiben oscilaciones, sino caos (Cole, 1991). Al añadir individuos a un nido artificial de superficie constante (de manera que vamos aumentando progresivamente la densidad de elementos) puede comprobarse que el comportamiento colectivo se va sincronizando progresivamente de forma que, para ciertas densidades, las oscilaciones se hacen visibles (en particular aparece un pico en el espectro de Fourier que se va haciendo más y más dominante a medida que aumentamos la densidad) y cada vez más regulares. Como consecuencia, la aparición de oscilaciones colectivas es claramente una propiedad emergente no reducible a la dinámica de los elementos aislados.

Para dar respuesta a las preguntas previamente formuladas, R. V. Solé introdujo un nuevo formalismo que permitiera analizar mediante modelos simples las propiedades de estos sistemas (Solé et al., 1992; Solé y Miramontes, 1993; Miramontes et al., 1993). El formalismo ha recibido el nombre de redes neuronales fluidas, y su inspiración, como se discutió en la introducción del capítulo, proviene de las analogías entre sistemas neuronales y grupos de insectos sociales. La idea es llevar a sus últimas consecuencias dichas analogías, considerando la colonia de hormigas como una verdadera red neural, solo que con una propiedad nueva: los elementos pueden desplazarse en un espacio dado, y en consecuencia las conexiones dejan de estar fijadas.

Consideremos por lo tanto un conjunto de $N$ “autómatas” (hormigas) definidos sobre una red bidimensional $\Lambda(L)$ de lado $L$, $\Lambda(L) = \{k = (i, j) \mid 1 \leq i, j \leq L\}$. El estado global $S_i(N)$ del colectivo será

$$S_i = \{S_i(k, t)\} \quad i = 1, \ldots, N.$$ 

Debido a la movilidad, las propiedades del entorno local de cada autómata (sus vecinos) se modificarán con el tiempo, y algunas correlaciones, por lo tanto, se destruirán (en este sentido, la red es “fluida”). Para explorar y reproducir las oscilaciones observadas, introduciremos el siguiente conjunto de reglas:
• Actividad: cada automata puede estar activo o inactivo. En el último caso, el elemento permanece inmóvil en su posición. Una vez activo, el elemento se mueve al azar hacia alguna de las ocho posiciones más próximas (si hay espacio accesible). Un elemento inactivo puede activarse ya sea por interacción con sus vecinos, o bien por medio de un mecanismo de activación espontánea al azar (o por medio de un proceso caótico subyacente, Solé et al., 1993).

• Red neural: los automatas se consideran como cierto tipo de “neuronas” y sus interacciones son análogas a las de una red neural típica. El estado de la neurona, que será en este modelo continuo, sigue una función sigmoide \( \Phi(x) \). Para un automata dado, con estado \( S_j(t) \) los nuevos estados se obtienen de

\[
S_j(t + 1) = \Phi \left( g \sum_{\langle i \rangle} J_{ij} S_j(t) \right)
\]

La suma se toma sobre los ocho vecinos más próximos (vecindad de Moore). También introducimos autointeracción \( (J_{ii} \neq 0) \), con lo que podemos escribir

\[
S_j(t + 1) = \Phi \left( g \left\{ J_{ii} S_i(t) + \sum_{i \neq j} J_{ij} S_j(t) \right\} \right)
\]

La matriz de conexiones no está fijada, sino que, tal y como ocurre en la realidad, depende del estado de los individuos que interactúan. Así, tendremos \( J_{ij} = f_\mu(S_i^t, S_j^t, \theta) \). Si el conjunto de posibles estados de cada automata es de tamaño \( M \), entonces \( (J_{ij}) \) tomará algún valor dentro de una matriz de posibles pares de estados, de tamaño \( M \times M \).

Tal y como indican los experimentos, los individuos aislados pueden activarse espontáneamente, así que deberemos introducir una regla asociada a esta propiedad. Los individuos inactivos, definidos a través de la desigualdad \( S(t) \leq \theta \), siendo \( \theta \) cierto valor umbral, pueden activarse espontáneamente (en ausencia de vecinos) con cierta probabilidad \( p_0 \). Una vez activados, alcanzan un estado \( S_0 \).

Aquí emplearemos \( \Phi(x) = \tanh(gx) \), así que las ecuaciones dinámicas serán de la forma

\[
S_i(t + 1) = \tanh \left( \mu S_i(t) \right)
\]

(con \( \mu = gJ_{ii} \), y \( J_{ii} > 0 \)). Puede demostrarse fácilmente que, para \( \mu < \mu_c = 1 \), sólo tenemos un punto fijo estable, \( S_0 = 0 \). Para \( \mu > \mu_c \), tenemos una bifurcación, con la aparición de dos estados simétricos \( S_{\pm} \). En aproximación lineal, tenemos que

\[
S(t + 1) \approx \frac{\partial \Phi(0)}{\partial S} S(t)
\]

para \( S = S_0 \), esto es, \( S(t + 1) = \mu S(t) \).

Para un individuo recién activado, se tiene \( S(0) = S_0 \), y su estado después de \( \tau \) iteraciones será \( S(\tau) = S_0 \mu^\tau \). De aquí podemos estimar el número de pasos de tiempo, \( \tau \), requeridos para alcanzar la inactivación, esto es, \( (S(\tau) \leq \theta) \). Nos da

\[
\tau = \frac{\log(\Gamma)}{\log(gJ_i)}
\]

con \( \Gamma = (\theta/S_0) \). Vemos que, para cierto \( \Gamma \), a medida que \( \mu = gJ_i \to \mu_c \) el tiempo de relajación \( \tau \) crece, con una singularidad en \( \mu = \mu_c \).
Figura 16.9: Oscilaciones colectivas de la actividad de la red neural fluida con activación espontánea. Las gráficas se han obtenido para $S_0 = 0.1$, $L = 10$, $g = 0.1$ y las densidades indicadas. Para densidades bajas, no se observa ningún patrón regular de comportamiento, tal y como esperamos a partir de las reglas empleadas. A medida que la densidad crece, la coherencia del sistema aumenta hasta generar oscilaciones muy bien definidas. A la derecha se representan los espectros de Fourier correspondientes.
Sean los subconjuntos de elementos activos e inactivos $N_+ = \sharp\{S_i > \theta\}$ y $N_- = N - N_+ = \sharp\{S_i \leq \theta\}$, respectivamente. Si no se introduce la regla de activación espontánea, el sistema fluido tiende hacia un atractor global dado por $S_i(\infty) = 0$; $\forall i = 1, ..., N$ para $\mu < \mu_c$, o bien hacia dos estados alternativos simétricos (por medio de un mecanismo de rotura de simetría) en caso contrario. El mecanismo de activación espontánea poseerá la capacidad necesaria para desplazar al sistema del atractor global y generar las oscilaciones observadas.

Observemos que, tal y como se ha definido el sistema, los elementos aislados no poseerán periodicidades intrínsecas. Se desactivarán al cabo de $r$ pasos de tiempo, y se activarán estocásticamente con cierta probabilidad $p_a$. Lo único que podremos decir por lo tanto, dinámicamente, es que la probabilidad de que, después de inactivarse, un autómata se active de nuevo en un instante $t$ seguirá una distribución $P_\text{nc}(t) = 1 - \exp(-p_a t)$. Al añadir elementos, la interacción puede generar nuevas propiedades. ¿Qué ocurre con el modelo que acabamos de definir? En la figura 16.9 podemos ver el resultado de variar la densidad de elementos $\rho = N/L^2$ sobre la dinámica del sistema. Podemos ver con claridad cómo van apareciendo oscilaciones regulares cada vez más claras a medida que aumentamos la densidad, tal y como veíamos en los experimentos de Franks y como evidencian los espectros de Fourier correspondientes.

### 16.5 Información y transiciones de fase

La transición entre la dinámica desordenada que vemos a bajas densidades y la dinámica coherente que se obtiene a densidad alta puede caracterizarse mediante el empleo de una medida estadística adecuada. Acudiremos una vez más a las cantidades basadas en la teoría de la información, que fueron introducidas anteriormente (capítulo 1). Analizaremos en primer lugar el comportamiento de la entropía de Boltzmann, $(H(\rho))$ asociada al número de elementos activos, y posteriormente estudiaremos la transferencia de información entre dos automatas.

Sea $T_j$ el número de iteraciones en las que $j$ elementos se encontraban simultáneamente activos. Si llevamos a cabo una simulación empleando $T$ iteraciones, la frecuencia relativa de dicho estado será $p(\rho, j) = T_j / T$, y la entropía de Boltzmann correspondiente

$$H(\rho) = -\sum_{i=0}^{N} p(\rho, i) \log p(\rho, i)$$

Como ya sabemos, $H(\rho)$ está acotada superiormente por $H^\text{max}(\rho) = \log(N + 1)$ e inferiormente por $H^\text{min}(\rho) = 0$. A medida que la densidad aumenta, la entropía crecerá, simplemente debido a que $N$ aumenta. Pero parece claro a partir de las simulaciones que, con la aparición de oscilaciones y por lo tanto de autoorganización, la entropía se verá acotada en alguna medida. Más allá de cierto punto, las correlaciones asociadas a la sincronización del sistema harán que éste preserve entropías reducidas.

Como vemos en la figura 16.10, aparece un máximo en cierta densidad $\rho_c$ que separa, por medio de una transición suave, el dominio de fluctuaciones irregulares del de oscilaciones colectivas. Aquí tenemos $S_0 = 0.1$, $p_a = 0.01$, $\theta = 10^{-6}$ y $g = 0.1$. El valor máximo depende de hecho básicamente de los parámetros $(\rho, g)$ y los empleamos como espacio paramétrico básico. En la figura 16.11 se muestra el resultado de analizar este espacio. Para valores muy pequeños de $g$, concretamente $g < g_c = 0.025$, la entropía antes definida no presenta ningún máximo, sino que crece aproximadamente según $H \approx \log(N)$. Los elementos son básicamente independientes y es en este sentido que la red fluida es aleatoria.

A continuación, consideraremos una medida aún más interesante, en tanto que, como hemos visto anteriormente, puede emplearse de forma efectiva como medida de complejidad: la información conjunta transmitida entre pares de elementos. También es importante en la medida en
Figura 16.10: (a) Entropía de Boltzmann $H(p)$ siendo $p(\rho, i)$ la probabilidad de encontrar exactamente $i$ individuos activos simultáneamente. Los parámetros son $L = 10$, $g = 0.08$. La gráfica muestra claramente la existencia de una densidad crítica en la que la entropía alcanza un máximo. En (b) se muestra la misma representación ($L = 10$) para valores distintos de densidad y del parámetro $g$.

Figura 16.11: Espacio paramétrico para la red neuronal fluida con activación espontánea. Se indican los distintos dominios correspondientes a distintos estados dinámicos.
Figura 16.12: (a) Información conjunta entre elementos en función de la densidad de los mismos. Vemos que presenta un máximo para cierta densidad crítica $\rho_c$. (b) Diagrama de información-entropía (véase texto).

que la información mutua es una cantidad directamente relacionada con la existencia de procesos implícitos de computación. Si tenemos en cuenta que lo que las hormigas manipulan colectivamente es precisamente información, deberíamos esperar que, para cierta densidad crítica especial, la información transmitida fuera máxima. El procedimiento, por otra parte, puede implementarse fácilmente en el estudio de otros sistemas.

Para calcular la información mutua $I(\rho)$ entre dos automatas arbitrarios, digamos el $i$-ésimo y el $j$-ésimo, debemos seguir su dinámica a lo largo del tiempo y de obtener las trayectorias

$$O_i(\rho) = \{S_i(t), \ldots, S_i(t + m)\} \quad y \quad O_j(\rho) = \{S_j(t), \ldots, S_j(t + m)\}$$

durante $m$ iteraciones. Nos restringiremos, más concretamente, a los conjuntos (asociados a los anteriores) $A_i(\rho), A_j(\rho)$, cuyos elementos se definen por $a_i(t) = 1$ si $S_i(t) > \theta$ y $a_i(t) = 0$ en caso contrario, y que corresponden de hecho a los estados observados experimentalmente. Tendremos por lo tanto

$$A_i(\rho) = \{a_i(t), \ldots, a_i(t + m)\} \quad y \quad A_j(\rho) = \{a_j(t), \ldots, a_j(t + m)\}$$

(con $a_i, a_j \in \Sigma = \{0, 1\}$), los cuales proporcionan una descripción basada en un alfabeto $\Sigma$.

Sea $p_i(r)$ la probabilidad de encontrar el $i$-ésimo automata en el estado $r \in \Sigma$ y sea $p_{ij}(r, s)$ la probabilidad conjunta de hallar simultáneamente al $i$-ésimo automata en el estado $r$ y al $j$-ésimo en el estado $s$: $p_{ij}(r, s) = P([a_i = r] \cap [a_j = s])$. La información mutua entre estos elementos es, como sabemos,

$$I(\rho)(A_i, A_j) = H(A_i) + H(A_j) - H(A_i, A_j)$$

con $H(A_k)$ la entropía de Boltzmann de un elemento dado,

$$H(A_k) = -\sum_r p_k(r) \log_2 p_k(r)$$
La entropía conjunta \( H(A_i, A_j) \) está definida por
\[
H(A_i, A_j) = - \sum_r \sum_s p_{ij}(r, s) \log_2 p_{ij}(r, s)
\]
donde las probabilidades \( p_i(r) \) y \( p_i(r, s) \) se han obtenido siguiendo a dos elementos dados a lo largo del tiempo, y han permitido obtener las entropías y la información transferida entre dichos autómatas (notemos que el canal de comunicación es ahora la propia colonia). En la figura 16.12 resumimos los resultados obtenidos. Para densidades bajas o altas, la información es baja, pero alcanza un máximo (para los parámetros indicados) en un valor \( \rho = \rho_c \) (\( \rho_c \approx 0.18 \)) que corresponde de hecho al punto en que aparecen las oscilaciones. También se muestra la gráfica información-entropía. En ésta observamos que el máximo de información se da a un valor intermedio de las entropías individuales, \( H^c_\mu \). Para \( H_\mu > H^c_\mu \), la transferencia de información es difícil dada la baja densidad del sistema, que impide la propagación de las activaciones. Para densidades altas, tenemos \( H_\mu < H^c_\mu \), ligada a la aparición de comportamiento colectivo coherente, que introduce efectos de memoria: los individuos propagan la actividad y además colaboran en mantenerla, lo cual genera intervalos de tiempo cada vez más largos en los que todos los elementos se hallan activos. En la frontera entre ambos comportamientos, esto es, en la frontera orden-caos, aparece la óptima transferencia de información. Vemos por lo tanto que estamos frente a un ejemplo de computación en la frontera del caos, concepto discutido en el capítulo sobre autómatas celulares. La evidencia experimental indica con claridad que las densidades observadas coinciden con las obtenidas por el modelo. Un estudio reciente (Delgado y Solé, 1995) indica claramente que la transición está correctamente definida en términos de las llamadas transiciones de fase inducidas por ruido.

Podemos obtener medidas adicionales de complejidad que permitan caracterizar el sistema. Por ejemplo, podemos calcular la longitud \( r \) de las series que necesitaremos para describir estadísticamente el sistema de forma correcta (tamaño del transitorio). Esta longitud diverge en los puntos críticos, así que debemos esperar que adopte un valor máximo para \( \rho_c \). Para medir este tiempo, recurriremos a la función:
\[
K_\mu(t) = \sum_r P(r; t) \log \left( \frac{P(r; t)}{P_s(r)} \right)
\]
denominada ganancia de información o información de Kullback (Haken, 1989; Solé y Miramontes, 1995) siendo \( \{ P_i(r) \} \) la distribución de probabilidad estacionaria y \( \{ P(r; t) \} \) la distribución calculada hasta el \( t \)-ésimo instante de tiempo. Como antes, \( r \in \{0, 1\} \), y tendremos
\[
\sum_r P(r; t) = 1, \quad \sum_r P_s(r) = 1
\]

La función \( K_\mu(t) \) posee una propiedad muy interesante (Nicolis y Prigogine): \( K_\mu(t) \geq 0 \), y tendremos igualdad si y sólo si \( P_s(r) = P(r; t) \) para todo valor de \( r \). La variación temporal de \( K_\mu(t) \) nos da la expresión
\[
\frac{\partial K_\mu}{\partial t} = \sum_{r, r'} \{ \omega(r|r')P(r'; t) - \omega(r'|r)P(r; t) \} \log \left( \frac{P(r; t)}{P_s(r)} \right)
\]
donde \( \omega(r|r') \) definen las probabilidades de transición. A partir de éstas, puede probarse que
\[
\frac{\partial K_\mu}{\partial t} \leq 0
\]
esto es, la ganancia de información siempre decrece con el tiempo.
Partiendo de una condición inicial arbitraria, calculamos en primer lugar la distribución estacionaria, \( \{ P_1(0), P_2(1) \} \) sobre \( A_\mu \). A continuación, definiremos el tamaño del transitorio \( \tau \) por el primer valor de \( t \) tal que:

\[
K_\mu(\tau) \leq \epsilon
\]

siendo \( \epsilon = 0.0025 \). El estudio de estos tiempos muestra un máximo bien definido en el punto crítico. Estudios recientes (Miramontes, 1995) han mostrado que las propiedades estadísticas temporales del modelo coinciden con las experimentales.

### 16.6 Hormigas y máquinas de Turing

En esta última sección introduciremos una propuesta teórica relacionada con el problema de la capacidad efectiva de cálculo de una red neural fluida (Solé y Delgado, 1995). Emplearemos como sistema formal el que hemos introducido previamente (con algunas modificaciones en las reglas específicas). Consideraremos ahora un conjunto de autónomas (hormigas) \( S_i(t) \equiv (S_i^1(t), ..., S_i^n(t)) \) que ahora adoptarán un conjunto discreto de estados, concretamente \( S_i \in \Sigma = \{0, 1\} \).

La dinámica de la red vendrá definida mediante un conjunto de probabilidades de transición,

\[
P(S_i \rightarrow -S_i) = \frac{1}{2} \{ 1 + \tanh \left[ \beta (h_i(t) S_i(t) - \Theta_i) \right] \}
\]

en las que \( h_i(t) \) indica el campo local (o campo externo) sobre los autónomas vecinos,

\[
h_i(t) = \sum_{j \in B_i(\rho)} J_{ij} S_j(t)
\]

siendo \( B_i(\rho) \) la vecindad considerada (a ocho vecinos en nuestro estudio). Introducimos además un umbral \( \Theta_i \). Al igual que antes, las conexiones serán en general dependientes del estado de los pares de individuos que interactúan: \( J_{ij} = \lambda(S_i(t), S_j(t)) \in \mathbb{R} \), y tendremos una tabla

\[
\Lambda = \begin{bmatrix}
\lambda_{++} & \lambda_{+-} \\
\lambda_{-+} & \lambda_{--}
\end{bmatrix}
\]

en la que indicamos que \( J_{ij} = \lambda_{++} \) si los elementos \( \{i, j\} \) son tales que \( S_i = S_j = +1 \), \( J_{ij} = \lambda_{+-} \) si \( S_i = +1 \) y \( S_j = -1 \), etc.

Existe un caso particular tratable, definido por la red fluida con conexiones unidad, i.e. \( \lambda_{ij} = 1 \) para todos los pares de interacciones. En este caso, podemos desarrollar una teoría de campo medio (capítulo 7) en la que empleamos el promedio \( m(S) = \sum_i S_i / N \) que evolucionará siguiendo la ecuación

\[
\Gamma \frac{\partial m}{\partial t} = -m + \tanh \left[ \beta \rho (m + \gamma) \right]
\]

donde \( \gamma \) es el campo externo.

Para \( \gamma = 0 \), el único atractor estable es \( m_0^* = 0 \) si \( J = \beta \rho < 0 = J_c \) y tendremos dos atractores \( m_\pm^* \neq 0 \)(con \( m_+^* = -m_-^* \)) para \( J > J_c \). Como ya sabemos, este resultado tiene lugar a través de un proceso de ruptura de simetría, que se da en \( J_c = 0 \). En términos de información (capítulo 1) podemos decir que el sistema ha "duplicado" la información existente al cruzar el punto crítico (Haken, 1988; Haken y Stadler, 1990). Estos atractores corresponderán a los mínimos de la función energía libre, dada por

\[
\frac{\partial \Phi(m^*, \gamma)}{\partial m^*} = 0
\]

y que podemos elegir en la forma

\[ \frac{1}{N} \Phi(m, \gamma) \equiv \phi(m, \gamma) = \frac{\beta \rho}{2} m^2 + \rho \ln \left( \cosh(\beta(\rho m + \gamma)) \right) \]

(también conocido como potencial de Ginzburg-Landau, véase el capítulo 7). De esta forma la dinámica del sistema viene dada, en la aproximación de campo medio, por un sistema gradiente,

\[ I \frac{dm}{dt} = -\frac{\partial \Phi(m, \gamma)}{\partial m} \]

Hemos introducido este desarrollo porque ilustra la conjetura realizada por Haken acerca de los mecanismos dinámicos de procesamiento de información en sistemas complejos (Haken, 1977; Haken y Stadler, 1990; Solé et al., 1993b). Para almacenar información, el sistema debe ser capaz de estabilizar sus atractores (que en nuestra aproximación corresponderían a estados dinámicos globales) en mínimos de un potencial, como aquellos definidos por \( \Phi(m, \gamma) \). Pero a la vez, el procesamiento de información requiere la capacidad de saltar de un atractor a otro, en caso de que sea necesario. Si, por medio de un mecanismo adecuado (autoregulado por el propio sistema) podemos modificar la función \( \Phi(m, \gamma) \) para acceder a nuevos mínimos, entonces el procesamiento de información y los procesos de cálculo serán posibles (figura 16.13).

Una red neural fluida puede presentar distintos atractores en varias formas. Una posibilidad simple sería una red con matriz ferromagnética, esto es,

\[ \Lambda_F = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \]

y tal que \( \beta \) pueda variar. Para \( \beta = 0 \) la mitad de los automatas estarán en el estado +1 y la otra mitad en el estado 0, que de ahora en adelante supondremos que representan dos tipos de
comportamiento (tareas). Cuando $\beta > 0$, tendremos rotura de simetría y dos nuevos atractores alternativos, \( \{m^*_4, m^*_5\} \).

Otra posibilidad es, por ejemplo, emplear un valor $\beta > 0$ pequeño y una matriz $A$ apropiada, de manera que las transiciones hacia un estado concreto sean más factibles. Por ejemplo, no es difícil ver que así ocurrirá por ejemplo con:

\[
A_0 = \begin{pmatrix}
1 & 1 - \epsilon \\
1 & 1 - \epsilon
\end{pmatrix}
\]

(con $\epsilon \in (0, 1)$), que hará que el sistema tienda a poseer todos sus elementos en el estado $S_i = +1$, y por lo tanto, $m \rightarrow m^*_4$.

La definición de la última matriz nos permite almacenar información, pero necesitamos, tal y como hemos indicado, un mecanismo de procesamiento que permita la transición entre atractores. Para ello requeriremos la formación de una señal colectiva, que el sistema pueda crear y automen- tener de forma estable, que tenga la capacidad de modificar la función energía y permitir así la transición. Siguiendo la observación de las colonias de hormigas, consideraremos la comunicación química como mecanismo básico. Nuestro deseo es conseguir que los elementos del sistema posean la capacidad de detectar señales externas y generar a su vez una señal que otros individuos puedan detectar y amplificar colectivamente, de manera que el sistema logre cambiar de atractor. Esto es posible, por ejemplo, si los individuos pueden dejar una marca química detectable por los demás, como ocurre con las hormigas.

Supongamos que un individuo detecta una señal externa, o sea una fuente de alimento local, y que emite una marca química dada. Esta señal tendrá una concentración local $C(i,j)$ (con $1 \leq i, j \leq L$, esto es, una red cuadrada) y supondremos que sigue una ecuación de difusión (capítulo 10),

\[
\frac{\partial C}{\partial t} = -\mu C + D \nabla^2 C
\]

do donde $\mu$ es la tasa de degradación, $D$ el coeficiente de difusión y $\nabla^2 = \partial^2_x + \partial^2_y$ el término de difusión pasiva.

Si no hay individuos, este campo de concentración se difunde y decrece hasta desaparecer. Si los elementos se hallan activos, detectarán el campo y, eventualmente, lo modificarán, siendo a su vez afectados por el campo (tenemos una propiedad emergente, como discutíamos al principio). Tanto el primero de los factores (la formación de un campo químico) como el segundo (la modificación del comportamiento individual a través de la acción del campo, véase Gordon, 1988) son conocidos y observables (Wilson, 1971; Hölldobler y Wilson, 1990).

Podemos ahora plantear un modelo completo, basado en los ingredientes anteriores. Nuestro objetivo será similar al que explorábamos en el capítulo 9 en relación con el problema de la Computación Universal (CU). La capacidad de CU (Hopcroft y Ullman, 1979) puede probarse, como ya hemos indicado, bien demostrando que una Máquina de Turing Universal (MTU) puede ser simulada por el sistema definido o, como hicimos en el caso del juego de la vida, demostrando que podemos emular los dispositivos lógicos necesarios para construir una MTU. Podemos demostrar que es posible construir una puerta NOR empleando una red neural fluida en la que los elementos generan, y se ven afectados por, un campo de concentración (Solé y Delgado, 1995).

La puerta NOR está definida por:

<table>
<thead>
<tr>
<th>Input 1 ($I_1$)</th>
<th>Input 2 ($I_2$)</th>
<th>$\Omega$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Figura 16.14: (a) Transición entre atractores en la red neural fluida descrita en el texto. Una vez introducida la señal \((C > 0)\), vemos que ésta es amplificada durante un tiempo, hasta que resulta eliminada y el sistema regresa a su atractor. Se indican \(\Omega\) y \(m\). (b) Espacio paramétrico \((\rho, \mu)\) con las distintas dinámicas, definidas a partir de la puerta lógica NOR.

en la que se indica por \(\Omega\) cierta respuesta colectiva, expresable en forma binaria. Supongamos que \(\beta > 0\) y que la matriz de conectividad es \(A_0\). Como consecuencia de esta elección, obtendremos cierta distribución de actividad definida por un conjunto de elementos que, mayoritariamente, estarán en el estado \(S_i = +1\). A continuación definimos dos entradas externas, \(I_1\) e \(I_2\). Estas serán dos señales emplazadas en dos puntos del espacio (el "ambiente" externo), que aquí tomamos como los vértices opuestos. Como señal, emplearemos una concentración fija \(C_0\) que puede hallarse \((I_1 = 1)\) o no \((I_1 = 0)\) presente (aunque existen muchas otras posibilidades). Cuando un automata detecta, en un punto dado de la red, cierta concentración \(C > \theta\) (\(\theta\) es un umbral), reforzará el campo en una cantidad \(\psi\). Finalmente, el campo local actuará sobre un individuo modificando la probabilidad de transición,

\[
h_{ij}(t, C_i) = \sum_{j \in E(\rho)} J_{ij} S_j(t) - C_i
\]

de forma que, en presencia de una concentración lo bastante grande, los elementos tenderán a cambiar de estado.

La "salida" colectiva se define por la función

\[
\Omega = H \left[ \sum_{j=1}^{N} S_j(t) \right]
\]

donde \(H(z)\) es la función de Heaviside, esto es, \(H(z) = 1\) para \(z > 0\) y \(H(z) = 0\) en caso contrario. Bajo estas condiciones, podemos comprobar que, efectivamente, el sistema puede llevar a cabo una transición colectiva desde el estado \(m_+\) al estado \(m_-\). Un ejemplo de esta transición se muestra en la figura 16.14 (a) para un conjunto dado de parámetros. Vemos cómo, una vez introducida la señal en un punto de la red, el sistema responde generando un cambio transitorio en el estado global, como vemos en la variación del número de individuos según el parámetro de salida \(\Omega\). Un estudio
del espacio de parámetros $(\sigma, \mu)$ (figura 16.14 (b)) nos muestra tres dominios bien definidos. En el primero (RN) tenemos una red fluida aleatoria, en la que las entradas no pueden ser amplificadas para dar lugar a una transición. En el otro extremo (zona sombreada) la amplificación es excesiva y el sistema es incapaz de regresar a su estado anterior. En la zona intermedia (zona blanca) aparece un dominio en el que la puerta NOR está correctamente definida y es reversible. Una vez el estímulo ha pasado, la amplificación descae y el sistema recupera su estado anterior (que, en un hormiguero, correspondería a la distribución de actividades normal). Para densidades intermedias (como ocurre en las colonias reales) el sistema posee la capacidad de almacenar y procesar información de forma estable.

Bibliografía


Capítulo 17

Caos Hamiltoniano

Este capítulo versará sobre una clase amplia de sistemas que presentan caos determinista: son los sistemas hamiltonianos. Para nosotros, un sistema hamiltoniano será aquel que conserva la energía total que posee, aunque en rigor el formalismo hamiltoniano puede ser aplicado también a sistemas cuya energía total dependa del tiempo.

Los sistemas conservativos hamiltonianos se hallan en la raíz del desarrollo del estudio de los sistemas caóticos. La primera vez que el caos apareció fue en la mecánica clásica, a finales del siglo pasado. (Por supuesto, el tipo de dinámica que ahora designamos con el nombre pomposo de caos determinista aún no se llamaba así. Y quizá la adopción de otra denominación hubiese evitado muchos malentendidos en la actualidad.) Henri Poincaré, quien trabajó durante algún tiempo en el problema de los tres cuerpos (preocupado por la estabilidad del sistema solar) fue el primero en encontrarse frente a la dinámica caótica desarrollada por un sistema conservativo. Es probablemente en el campo de la mecánica celeste donde el formalismo hamiltoniano se revela de mayor utilidad. Desde los tiempos de Poincaré han sido muchos los problemas no dissipativos estudiados en los que se ha encontrado caos determinista. Describiremos brevemente en este capítulo algunos de ellos, y comentaremos la influencia que la dinámica caótica puede tener en el movimiento de los asteroides, en la estabilidad de la resonancia rotacional de Júpiter y Saturno, en la variación del ángulo de rotación de la tierra, ...

Comenzaremos con una breve introducción al formalismo hamiltoniano (como es costumbre, en la bibilografía del final del capítulo se encontrará dónde buscar el tema desarrollado) para pasar a estudiar las peculiaridades de los sistemas conservativos, la aparición de resonancias y, no obstante, la posibilidad de encontrar caos. Posiblemente el resultado más brillante en relación a la estabilidad de las órbitas periódicas de los sistemas conservativos es el llamado teorema KAM (por Kolmogorov, Arnold y Moser), el cual nos garantiza la estabilidad de estas órbitas frente a perturbaciones que, en principio, podrían desestabilizar un sistema conservativo y, como consecuencia más catastrófica, nuestro propio sistema solar podría ser inestable. El teorema KAM permite que respiremos tranquilos por unos 100 millones de años...

17.1 La mecánica de Hamilton y Jacobi

Describamos en lo sucesivo nuestro sistema dinámico mediante dos conjuntos de variables: $q$ para las posiciones y $\dot{q}$ para las velocidades,

\[ q = (q_1, q_2, \ldots, q_n) \]
\[ \dot{q} = (\dot{q}_1, \dot{q}_2, \ldots, \dot{q}_n) \]
El número de coordenadas del sistema en el espacio de posiciones (o de velocidades), \( n \), se denomina número de grados de libertad del sistema. Es esencial conocer primeramente la llamada función lagrangiana o lagrangiante \( L \) de sistema. Si llamamos \( T \) a la energía cinética total y \( V \) a la energía potencial del sistema, \( L \) se define de la forma siguiente:

\[
L = T - V \tag{17.1.1}
\]

La energía cinética es siempre el producto

\[
T = \frac{m}{2} \dot{q}^2
\]

donde \( \dot{q}^2 = \dot{q}_1 \dot{q}_1 + \dot{q}_2 \dot{q}_2 + \ldots + \dot{q}_n \dot{q}_n \)

La energía potencial es una función que sólo depende de las posiciones,

\[
V = h(q_1, q_2, \ldots, q_n)
\]

Si definimos los momentos \( p_j \) asociados a las variables \( \dot{q}_j \) como las cantidades

\[
p_j = \frac{\partial L}{\partial \dot{q}_j} \tag{17.1.2}
\]

expresaremos las ecuaciones de Newton del movimiento, como

\[
\frac{dp_j}{dt} = -\frac{\partial V}{\partial q_j}
\]

A partir del lagrangiano antes definido podemos escribir las llamadas ecuaciones lagrangiiantes del movimiento, que representan la dinámica de cualquier sistema para el que conozcamos su función lagrangiana,

\[
\frac{d}{dt} \left( \frac{\partial L}{\partial \dot{q}_j} \right) - \frac{\partial L}{\partial q_j} = 0
\]

La siguiente transformación, que permite llegar a la función hamiltoniana, o hamiltoniano, \( H \) de un sistema, consiste en sustituir \( L(q, \dot{q}; t) \) por una nueva función \( H(q, p; t) \), de la forma siguiente:

\[
H(q, p; t) = \sum_{j=1}^{n} \dot{q}_j \frac{\partial L}{\partial \dot{q}_j} - L \tag{17.1.3}
\]

\( H \) es una función de \( q \) y \( p \), posiciones y momentos. Estos últimos están definidos por la expresión 17.1.2. En estas últimas líneas hemos incluido explícitamente la posible dependencia de \( L \) y \( H \) con el tiempo \( t \). Sin embargo, tratamos en este capítulo únicamente con sistemas conservativos y ello implica que ni \( L \) ni \( H \) pueden depender explícitamente del tiempo \( t \). Es decir, si nuestro sistema es conservativo,

\[
\frac{\partial H}{\partial t} = 0
\]

Esto implica que existe una cantidad conservada sobre las trayectorias del sistema, la energía \( E \),
\[ H(q, p) = E \]

El hamiltoniano permite escribir las ecuaciones del movimiento del sistema en la forma

\[
\frac{dp_i}{dt} = \frac{\partial H}{\partial q_i}, \quad \frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}
\]  \hspace{1cm} (17.1.4)

Estas son las llamadas ecuaciones del movimiento de Hamilton-Jacobi, las cuales forman un conjunto de 2n ecuaciones diferenciales de primer orden, en lugar de las n de segundo orden que proporcionaban las ecuaciones de Newton. Dado que tratamos con sistemas conservativos, no encontraremos regiones atractivas en el espacio de las fases (puntos fijos, ciclos límite o atractores extraños) como en los casos disipativos. El caso aparecerá aquí de forma diferente, con una topología diferente, como veremos.

Demostremos con un ejemplo sencillo la forma en que podemos llegar al hamiltoniano de un sistema.

**Ejemplo**

Consideremos el problema de los dos cuerpos en el caso simplificado en que uno de ellos (el de masa \( M \), y consideramos \( M > m \), con \( m \) la masa del segundo cuerpo) está fijo en un punto del espacio. El cuerpo menor, de masa \( m \), se mueve bajo la acción gravitatoria del primero, con una energía cinética \( T \) y una energía potencial gravitatoria \( V \) que son

\[ T = \frac{1}{2} m (\dot{x}^2 + y^2) \]
\[ V = -G \frac{mM}{\sqrt{x^2 + y^2}} \]

donde se ha considerado, sin pérdida de generalidad, que el movimiento tiene lugar en el plano \((x, y)\) \(^1\). Es más conveniente en este caso utilizar coordenadas polares,

\[ x = r \cos(\theta), \quad y = r \sin(\theta) \]

Derivando

\[
\begin{align*}
\dot{x} &= \dot{r} \cos(\theta) - \dot{\theta} r \sin(\theta) \\
\dot{y} &= \dot{r} \sin(\theta) + \theta \dot{r} \cos(\theta)
\end{align*}
\]

implica

\[ \dot{x}^2 + \dot{y}^2 = \dot{r}^2 + r^2 \dot{\theta}^2 \]

y considerando que \( x^2 + y^2 = r^2 \), la energía cinética se escribirá ahora como

\[ T = \frac{1}{2} m (\dot{r}^2 + r^2 \dot{\theta}^2) \]

y la energía potencial

\[ V = -G \frac{mM}{r} \]

El lagrangiano del sistema en coordenadas polares será pues \(^2\)

\[ L = T - V = \frac{1}{2} m (\dot{r}^2 + r^2 \dot{\theta}^2) + G \frac{mM}{r} \]
Figura 17.1: Esquema del problema integrable de los dos cuerpos tratado en la sección 17.1.

Para escribir el hamiltoniano del sistema recordemos que necesitamos la expresión de los momentos correspondientes a cada una de las variables del sistema. Las variables hamiltonianas son en este caso

\[ q = (r, \theta), \quad p = (p_r, p_\theta) \]

y los momentos se obtienen de 17.1.2, proporcionando

\[ p_r = \frac{\partial L}{\partial \dot{r}} = m \dot{r}, \quad p_\theta = \frac{\partial L}{\partial \dot{\theta}} = m r^2 \dot{\theta}. \]

El hamiltoniano será ahora

\[ H = \sum_i p_i \dot{q}_i - L = p_r \dot{r} + p_\theta \dot{\theta} - \frac{1}{2} m (\dot{r}^2 + r^2 \dot{\theta}^2) + G \frac{mM}{r} \]

que debe ser reescrito en función de coordenadas y momentos, lo cual, utilizando las relaciones 17.1.5 se reduce finalmente a

\[ H = \frac{1}{2m} \left[ p_r^2 + \frac{p_\theta^2}{r^2} \right] - G \frac{Mm}{r} \]

Las ecuaciones del movimiento que se obtendrán a partir de \( H \) y 17.1.4 serán

\[ \frac{\partial H}{\partial p_r} = \dot{r} = \frac{p_r}{m} \]

1 Si el potencial es central, el movimiento siempre tiene lugar en un plano, ya que el vector momento angular es una cantidad conservada, y representa el vector normal del plano que contiene la dinámica. Un simple giro de ejes permite situar el momento angular en el eje z y por tanto el movimiento en el plano \((x, y)\).

2 Obsérvese que, debido a que no utilizamos coordenadas cartesianas, aparece en la energía cinética el término \(r^2\), correspondiente a la métrica en coordenadas polares, para las cuales

\[ q^2 = (\dot{r}, \dot{\theta}) \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix} \begin{pmatrix} \dot{r} \\ \dot{\theta} \end{pmatrix} = \dot{r}^2 + r^2 \dot{\theta}^2 \]
\[
\frac{\partial H}{\partial \phi} = \dot{\phi} = \frac{p_\phi}{m r^2} \\
\frac{\partial H}{\partial r} = \dot{r} = G \frac{M m}{r^2} - \frac{1}{m} \frac{p_r^2}{r^3} \\
\frac{\partial H}{\partial \theta} = \dot{\theta} = 0 \implies m \dot{\theta} r^2 = ct
\]

De esta última relación se obtiene un resultado importante, que puede darnos una idea de la potencia del formalismo hamiltoniano: cuando el hamiltoniano no contiene explícitamente la coordenada \( q_i \), el momento conjugado de ésta, \( p_i \), es una cantidad conservada sobre las trayectorias del sistema. A \( q_i \) se le denomina entonces coordenada cíclica. Las cantidades conservadas juegan un papel fundamental en la resolución de estos sistemas dinámicos, como veremos a continuación.

### 17.2 Sistemas dinámicos integrables

Decimos que un sistema dinámico es integrable cuando podemos averiguar que existen soluciones para el sistema dependientes de funciones conocidas. Obsérvese que no decimos que seamos capaces de escribir esta solución, sino de determinar su existencia. De hecho, un sistema dinámico es integrable cuando existe un número suficiente de constantes del movimiento además de la energía, las cuales permiten realizar una predicción cualitativa del movimiento en el espacio de las fases. El ejemplo que hemos dado en la sección anterior se ha hallado casi de manera trivial: una cantidad conservada, el momento angular, además de la energía. No siempre es tan sencillo encontrar estas cantidades de forma explícita, aunque su presencia se puede reconocer cualitativamente porque generan toros invariantes en el espacio de las fases.

Si un sistema es integrable no puede ser caótico, así que la determinación de la existencia (o el resultado negativo de la no existencia) de constantes del movimiento es fundamental en el caso que nos ocupa. Si una cantidad \( F \) es una constante del movimiento de un sistema dinámico, verifica

\[
0 = \frac{d}{dt} F(p_i, q_i) = \sum_i \left( \frac{\partial F}{\partial p_i} \frac{dp_i}{dt} + \frac{\partial F}{\partial q_i} \frac{dq_i}{dt} \right) \\
= \sum_i \left( \frac{\partial H}{\partial p_i} \frac{\partial F}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial F}{\partial p_i} \right) = [H, F]
\]

La última línea define el paréntesis de Poisson, \([H, F]\), que se puede calcular para cualquier par de funciones en el espacio de fases. Si \([H, F] = 0\) en todo el espacio de fases, \( F \) es una constante del movimiento. La interpretación geometríca del este resultado lleva a la conclusión de que la dinámica del sistema tiene lugar sobre la intersección de las hipersuperficies \( F = ct\). \( H = ct\).

Utilizando el paréntesis de Poisson podemos escribir las ecuaciones de Hamilton en la forma

\[
\frac{dq_i}{dt} = [H, q_i] \\
\frac{dp_i}{dt} = [H, p_i]
\]

Un hamiltoniano con \( n \) grados de libertad será integrable cuando existan \( n \) funciones diferentes, a determinar, tales que

\[
[H, F_i] = 0, \quad i = 1, \ldots, n
\]
En general, sin embargo, en lugar de $n$ constantes del movimiento encontraremos, digamos, $k$ constantes, energía incluida, que restringirán el movimiento a una hipersuperficie $(2n-k)$-dimensional en el espacio de fases. La estructura interna de esta hipersuperficie no tiene porqué ser simple, en principio, a no ser que todos los paréntesis de Poisson entre las constantes del movimiento se anulen, caso en el cual se dice que estas constantes están en involución.

Supongamos, en el caso más favorable, que hemos hallado $n$ constantes del movimiento que están en involución. Entonces, se puede demostrar que el movimiento del sistema tiene lugar sobre un toro $n$-dimensional. Las trayectorias del sistema se localizan sobre estos toros invariantes, y el sistema es integrable. Lo más adecuado en estos casos es el uso de las llamadas variables acción-ángulo, sobre las que no nos extenderemos, pero que tratan de aprovechar la topología de estas hipersuperficies para describir la posición del sistema únicamente mediante un conjunto de ángulos (de forma análoga a lo visto en la parte dedicada a movimiento cuasiperiódico en el capítulo 4). El conjunto de variables ángulo, $(w_1, w_2, \ldots, w_n)$ define las frecuencias, $\Omega_i$ del sistema. El movimiento será periódico si las relaciones entre las frecuencias $\Omega_i$ son racionales, es decir, si éstas son commensurables. El conjunto complementario de variables, las variables de acción es $(I_1, I_2, \ldots, I_n)$, las cuales juegan el papel de momentos respecto de las variables $w_i$.

No daremos la forma en que se pueden hallar las transformaciones que pasan de $(p_i, q_i)$ a $(w_i, I_i)$. El lector interesado puede consultar Lanczos (1970) o Gantmacher (1975), entre muchos otros.

Supongamos que hemos hallado una función $S$, llamada transformación canónica que nos permite realizar el cambio de coordenadas citado, y de forma que, expresado en las nuevas coordenadas, el hamiltoniano del sistema resulte ser una función únicamente de los momentos,

$$H(q, p) \rightarrow H_0(I) = H_0 \left[ q_i, \frac{\partial S(q, I)}{\partial q} \right]$$

con

$$p = \frac{\partial S(q, I)}{\partial q}, \quad w = \frac{\partial S(q, I)}{\partial I}$$

así que las ecuaciones del movimiento quedan reducidas a

$$\dot{I}_i = -\frac{\partial H_0}{\partial w_i} = 0; \quad \dot{w}_i = \frac{\partial H_0}{\partial I_i} = w_i(I)$$

y pueden ser integradas para proporcionar la solución completa a la dinámica del sistema,

$$I_i = ct$$

$$w_i = \Omega_i t + \delta_i$$  (17.2.2)

Estas últimas ecuaciones son análogas a las que se obtendrían para un conjunto de $n$ osciladores no acoplados. La introducción de algún término de acoplamiento en el anterior hamiltoniano integrable podría provocar que las frecuencias $\Omega_i$ ya no fuesen todas independientes. Los acoplamientos pueden inducir resonancias y llevar las relaciones $\Omega_i/\Omega_j$ a cocientes racionales. Este es el mismo caso que se había visto para la cuasiperiodicidad en sistemas disipativos (capítulo 4). Las resonancias en el caso hamiltoniano pueden llegar a desestabilizar el sistema, como se verá en las próximas secciones.

El problema de hallar las constantes del movimiento de un sistema, a fin de saber si éste es integrable, y por tanto si presentará un movimiento regular, se convierte en la esencia del estudio de los sistemas dinámicos conservativos. En caso de que no existan $n$ constantes del movimiento en un sistema con $n$ grados de libertad, nos hallamos ante la posibilidad de la dinámica caótica.
El primer sistema en donde se descubrió la posibilidad de movimiento "irregular" fue en el problema de los tres cuerpos. Consideremos, para fijar ideas, el Sol, la Tierra y la Luna, y prescindamos de la influencia de los demás planetas. El espacio de fases de este sistema triple tiene 18 dimensiones. Habitualmente se considera el movimiento referido al sistema de referencia del centro de masas de los tres cuerpos, con lo cual las coordenadas espaciales requeridas son: 3 para el centro de masas del sistema, 3 para el vector que apunta desde el centro del Sol hasta el centro de masas del conjunto Tierra-Luna, y otras 3 para el vector que va de la Tierra a la Luna. Considerando los momentos correspondientes a cada uno de estos vectores obtenemos un total de 18 dimensiones.

En este sistema se pueden reconocer inmediatamente 10 constantes del movimiento:

1. Dado que el centro de masas de los tres cuerpos tiene un movimiento uniforme que no influye en los movimientos del Sol, la Tierra o la Luna, podemos considerar el sistema de referencia del centro de masas en reposo, y eliminar seis dimensiones (posición y momento del centro de masas).

2. El momento angular del sistema es constante. Su dirección determina un plano invariante a través del centro de masas, y nos permite eliminar otras tres dimensiones.

3. La energía del sistema es constante, lo cual permite en adición determinar el eje mayor del movimiento del centro de masas Tierra-Luna alrededor del Sol. La energía permite reducir en 1 la dimensión del sistema.

 Esto nos deja con un sistema de dimensión real 8. Y nos preguntamos, ¿existe alguna otra constante del movimiento, aparte de las citadas? La respuesta la dio el teorema de Bruns y Poincaré: no existe ninguna otra constante del movimiento que permita rebajar la dimensión del sistema.

El problema de los tres cuerpos fue el primero en descubrir todo un mundo de dinámica irregular. Poincaré encontró esta dinámica compleja, y concluyó hace ya 100 años que la no integrabilidad del problema de los tres cuerpos implicaba la posible inestabilidad del sistema solar (Peterson, 1994). Como aproximación al problema completo, Poincaré inició el estudio de las órbitas periódicas del sistema. Las siguientes palabras (Poincaré, 1892) describen la impresión que en él causó la "dinámica caótica" que acababa de descubrir:

"Lo que proporciona a estas órbitas periódicas su valor es que ofrecen, por decirlo de alguna manera, el único camino por el cual penetrar en esta fortaleza que tiene la reputación de ser inenjugable."

Sin embargo, las condiciones iniciales que sitúan la dinámica sobre órbitas periódicas tienen una medida de f 0 es cero entre el total de condiciones iniciales posible. Poincaré había hallado que, en consecuencia, el movimiento de los cuerpos en el sistema solar no se repetía casi nunca exactamente. No había recurrencia a la que aferrarse. El futuro, aunque determinista, se había convertido en impredecible.

Para entender cualitativamente lo que esto significa, consideremos una aplicación discreta sencilla y conservativa. Esta aplicación proviene de una simplificación del potencial de Hénon-Heiles que da lugar al siguiente hamiltoniano no integrable

\[ H = \frac{1}{2} p_1^2 + \frac{1}{2} p_2^2 + q_1^2 + q_2^3 + q_2q_3 \]

el cual simula la acción de la galaxia sobre una estrella. La aplicación resultante (Hénon, 1969) es la siguiente
\[ x_{n+1} = x_n \cos (\alpha) - (y_n - x_n^2) \sin (\alpha) \]
\[ y_{n+1} = x_n \sin (\alpha) + (y_n - x_n^2) \cos (\alpha) \] (17.2.3)

Si tomamos un conjunto suficientemente amplio de condiciones iniciales sobre el espacio de fases \((x, y)\) hallaremos una mezcla sorprendente de órbitas periódicas y órbitas caóticas. Véase la figura 17.2, la cual muestra claramente la variación de la dinámica con el aumento de energía en el sistema. Cuando la energía es pequeña, las órbitas son periódicas (centro del atráctor), en tanto que cuando aumenta, las órbitas periódicas quedan atrapadas en islas de estabilidad y rodeadas por un mar de órbitas caóticas.

La figura 17.3 representa concretamente tres órbitas correspondientes a este sistema. Las dos exteriores son caóticas y la interior es cuasiperiódica. Cada una de ellas corresponde a la elección de una condición inicial diferente. Si la órbita es caótica, la superficie ocupada por las imágenes sucesivas del punto inicial es extensa. Si la órbita fuese periódica ocuparía un conjunto discreto de puntos, y si es cuasiperiódica se situaría sobre una curva cerrada.

### 17.3 Teoría de perturbaciones

La mayoría de problemas clave en mecánica es integrable, entre ellos el movimiento de un planeta alrededor del sol (problema de los dos cuerpos) y el oscilador armónico. Resulta pues tentador intentar resolver los problemas más complicados partiendo de estos sencillos, perturbaando el hamiltoniano resoluble con el término no integrable. La teoría de perturbaciones que posibilita esta aproximación falla, sin embargo, cuando nos encontramos ante resonancias en las frecuencias que caracterizan un sistema, como veremos.

Recordemos que, para un sistema integrable, podíamos dar su solución en función del conjunto de variables acción-ángulo, y habíamos reducido la dinámica a una hipersuperficie topológicamente equivalente a un toro \(n\)-dimensional. Sobre esta variedad, \(T^n\), las posibles dinámicas se reducen a dos casos: periodicidad (cocientes racionales entre algunas de las frecuencias \(\Omega_i\)) o cuasiperiodicidad (frecuencias incommensurables).

Una aproximación sencilla a un problema no integrable, pero que posea un hamiltoniano \(H\) cercano al anterior \(H_0\), consiste en considerar la parte no integrable como una perturbación al caso integrable que se ha podido solucionar, en la forma

\[ H(I, w) = H_0(I) + \epsilon H_1(I, w) \]

con \(\epsilon\) pequeño. El hamiltoniano \(H\) se ha expresado mediante el conjunto de variables \((I, w, \ldots)\), pero para el simplemente representa un cambio de coordenadas, que, debido al término \(\epsilon H_1\), no permite ahora encontrar la solución en la forma antes indicada: ahora, las variables \(I_i\) no son constantes del movimiento. El hamiltoniano \(H_0(I)\) sí que corresponde a un problema integrable, escrito en coordenadas canónicas mediante las variables acción-ángulo para ese caso.

Intentemos encontrar un nuevo conjunto de variables acción-ángulo, \((w', I')\) mediante una transformación canónica \(S\) para el hamiltoniano perturbado \(H\), de forma que se verifique

\[ H \left[ \frac{\partial S}{\partial w_i}, w_i \right] = H'(I'_i) \]

que es la condición que debe cumplirse para obtener las constantes del movimiento \(I'_i\) y llegar a la integración total del sistema. Escribamos la nueva función \(S\) como

\[ S(I, w) = wI + \epsilon S_1(I, w) \]
Figura 17.2: Mapa de fases de la aplicación conservativa de Hénon 17.2.3, con \( \cos \alpha = 0.24 \), arriba a la izquierda. Las otras tres imágenes son ampliaciones del conjunto inicial, en las ventanas indicadas en cada una de las figuras.
Figura 17.3: Tres órbitas independientes para tres condiciones iniciales diferentes (correspondientes a tres valores diferentes de la energía) en la aplicación de Henon 17.2.3. La interior es cuasiperiódica y las dos exteriores son caóticas.

y con esta definición desarrollamos nuestro hamiltoniano a orden $\epsilon$:

$$
H_0(I') + \epsilon \frac{\partial H_0}{\partial I} \frac{\partial S_1(I', w)}{\partial w} + \epsilon H_1(I', w) + O(\epsilon^2) = H'(I')
$$

(17.2.4)

Como la expresión de la izquierda no puede depender de $w$, debemos exigir

$$
\Omega \frac{\partial S_1(I', w)}{\partial w} = -H_1(I', w)
$$

(17.2.5)

lo cual determina $S_1$, y donde

$$
\Omega = \frac{\partial H_0}{\partial I}
$$

son las frecuencias características del sistema no perturbado (ecuación 17.2.2). Dado que se supone $H_1$ (y por tanto también $S_1$) periódico en las componentes de $w$, podemos intentar resolver 17.2.4 utilizando desarrollos en serie de Fourier para $H_1$ y $S_1$.

$$
S_1(I', w) = \sum_{K \neq 0} S_{1,K}(I') e^{iKw}
$$

$$
H_1(I', w) = \sum_{K \neq 0} H_{1,K}(I') e^{iKw}
$$

con $K = 2\pi (m_1, m_2, \ldots, m_n)$, y $m_i$ valores enteros, $\forall i$. Por sustitución en 17.2.5 e igualando los coeficientes de los términos iguales en la serie de Fourier se obtiene

$$
S_1(I', w) = wI' + i\epsilon \sum_{K \neq 0} \frac{H_{1,K}(I')}{K \Omega(I')} e^{iKw}
$$
Por desgracia, obtendremos divergencias en la expresión anterior cuando el denominador se anule, es decir, cuando

$$K \Omega = 0 \Rightarrow m_1 \Omega_1 + m_2 \Omega_2 + \ldots + m_n \Omega_n = 0$$

lo cual quiere decir cuando existan resonancias entre las frecuencias, cuando éstas sean incommensurables.

Este es el conocido problema de los denominadores pequeños: el sistema no puede ser tratado mediante teoría de perturbaciones debido a la probable anulación de los denominadores de algunos términos de la serie (supuesta en principio formada por términos pequeños y convergente, a fin de que la teoría de perturbaciones sea útil). La situación es en realidad aún peor, ya que incluso en el caso en que $K \Omega \neq 0$, es decir, cuando las frecuencias son incommensurables, siempre es posible encontrar un conjunto de valores $m$ para el que

$$K \Omega < \delta$$

con $\delta$ arbitrariamente pequeño. Esta situación nos hace dudar de la convergencia de la serie incluso en ausencia de resonancias. Por tanto, nos encontramos ante el problema de la convergencia de la suma en $K$ y el de la convergencia del desarrollo en potencias de $\epsilon$.

Cuando se tratan ciertos problemas de mecánica celeste, se encuentra que algunas relaciones de frecuencias están alejadas de ser incommensurables para valores de $m_i$ pequeños (aunque otras relaciones están asombrosamente próximas a un racional, como se verá). Por ejemplo, la frecuencia de revolución de la Tierra alrededor del Sol es $11.86$ veces la de Júpiter. su principal perturbación. Podríamos considerar únicamente unos cuantos términos de la serie, en lugar de la serie completa, para predecir la dinámica durante un intervalo de tiempo finito, que puede ser muy largo a escala celeste. El éxito de estas predicciones, sin embargo, no soluciona el problema de la integrabilidad del hamiltoniano $H$ y de la existencia o no de toros $n$-dimensionales para su dinámica.

El problema de la predicción puede ser muy diferente dependiendo del tiempo característico del sistema, del valor de $\Omega_i$. Por ejemplo, en mecánica estadística se trabaja con unos “tiempos moleculares” del orden de $10^{13}$ revoluciones por segundo, lo cual corresponde a escala planetaria, a 1000 veces la edad del universo.

Tras este análisis, queda planteada la siguiente pregunta: ¿qué ocurre en el sistema original cuando una órbita con $K \Omega = 0$ se perturba mediante el término $\epsilon H$? o bien, ¿qué ocurre en el caso $K \Omega = \delta$ (con $\delta$ pequeño) bajo esta misma perturbación? El teorema KAM nos da parte de la respuesta.

### 17.4 Resonancias y el teorema KAM

Podríamos razonar que el problema de tener relaciones racionales entre las frecuencias de un sistema no debería ser tal, puesto que, si escogemos al azar un valor en la recta real, la probabilidad de que sea racional es cero a priori. Sin embargo, la existencia del fenómeno de la resonancia lleva a estos valores en principio arbitrarios a presentar efectivamente relaciones racionales. Probablemente la más conocida de estas resonancias a nivel celeste es el hecho de que la Luna siempre nos ofrece la misma cara. Es decir, su frecuencia de rotación y su frecuencia de traslación alrededor de la Tierra se encuentran en la relación 1:1. La explicación a este hecho, conocido desde la antigüedad, no fue hallada hasta 1764 por Lagrange. Otro caso, quizá no tan popular, es el de la resonancia de los periodos de traslación de Júpiter y Saturno alrededor del Sol. El primero invierte 12 años, y el segundo 30 en realizar una revolución completa. Cada 60 años estos dos planetas se encuentran exactamente en la misma posición relativa. La relación entre sus períodos es 2:5 con muy buena
aproximación. Así pues, parece ser que nos encontraremos efectivamente con la relación $K\Omega = 0$, o bien con un valor muy cercano a un racional.

El teorema KAM (Kolmogorov, 1954; Arnold, 1963; Moser, 1967) da una respuesta al comportamiento de los toros $n$-dimensionales bajo la acción de la perturbación $\epsilon H_1$. Consideremos el caso 2-dimensional en que las frecuencias implicadas en la dinámica sean $\Omega_1 < \Omega_2$. Entonces, si el cociente $\Omega_1/\Omega_2$ es suficientemente irracional, de forma que se verifique

$$\frac{|\Omega_1 - r|}{\Omega_2} > \frac{k(\epsilon)}{s^{2.5}}, \quad \forall \, r, s, \text{ enteros} \quad (17.4.1)$$

el toro 2-dimensional definido por $\Omega_1$ y $\Omega_2$ no se destruye: únicamente sufre una deformación topológica. El valor $k$ es independiente de $r$ y $s$ y verifica $k(\epsilon \to 0) \to 0$. Los toros 2-dimensionales no incluidos en 17.4.1 y que son en su mayor parte destruidos por la perturbación son los que satisfacen

$$\frac{|\Omega_1 - r|}{\Omega_2} < \frac{k(\epsilon)}{s^{2.5}} \quad (17.4.2)$$

para algún valor de $r$ y $s$ entero.

El conjunto de valores de $\Omega_1/\Omega_2$ para el que se verifica la relación 17.4.1 es no vacío, es decir, tras la perturbación, existen toros que “sobreviven” en el espacio de las fases. Calculemos aproximadamente la cantidad de toros (el intervalo total de frecuencias) que será destruida bajo la perturbación. Supongamos que el cociente de frecuencias está acotado entre 0 y 1 (sin pérdida de generalidad, $0 \leq \Omega_1/\Omega_2 < 1$). Entonces, suprimamos un intervalo de longitud $k/s^{2.5}$ alrededor de cada racional $r/s$ en el rango $[0, 1)$. La longitud total eliminada es

$$\sum_{s=1}^{\infty} \frac{k}{s^{2.5}} = k \sum_{s=1}^{\infty} \frac{1}{s^{1.5}} \approx k$$

cantidad que tiende a cero cuando $\epsilon \to 0$, y que proporciona un límite superior a la cantidad total de toros destruidos.

El teorema demuestra resultados análogos para $n$ dimensiones. Resumimos el resultado general: en un sistema perturbado, la mayoría de las órbitas se localizan sobre toros en el espacio de las fases. Las que no se mueven sobre estos toros forman un conjunto pequeño pero infinito, patológicamente distribuido en el espacio de fases cerca de cada toro no perturbado que soporta órbitas cerradas o parcialmente cerradas. Desde el punto de vista físico, los intervalos correspondientes a valores grandes de $s$ son difíciles de estudiar, ya que son muy estrechos, y pequeñas perturbaciones aleatorias (presentes en cualquier sistema real) añadidas a $\epsilon$ pueden llevar al sistema fuera de este intervalo, y colocarlo en el dominio de uno de los toros vecinos. Los intervalos de orden menor (s pequeño), que resultan de las resonancias de bajo orden entre las frecuencias no perturbadas, son relativamente amplios, y dan lugar a efectos observables y computables, como veremos en la sección 17.6.

17.5 El teorema de Poincaré-Birkhoff

El teorema KAM ha proporcionado la respuesta sobre el destino de los sistemas con relaciones de frecuencias casi racionales cuando son perturbados. Aún nos queda por determinar qué sucede con las relaciones racionales entre frecuencias. Veremos que, cuando esto sucede, el toro original perfectamente determinado se descompone en toros más y más pequeños, algunos de los cuales serán nuevamente estable y de acuerdo con el teorema KAM, en tanto que otros inestables seguirán
descomponiéndose. Entre los toros estables hallaremos una dinámica irregular, que en sistemas con \( n > 2 \) podrá cubrir todo el espacio de fases disponible.

Consideraremos en esta ocasión una sección de Poincaré de nuestros toros 2-dimensionales para visualizar mejor lo que sucede bajo perturbaciones. De nuevo \( H_0 \) es el hamiltoniano que hemos podido resolver exactamente, el cual posee dos frecuencias características que verifican una relación racional, y \( H_1 \) es el hamiltoniano perturbador. Los toros correspondientes al sistema integrable presentarán una sección de Poincaré formada por curvas concéntricas que en el caso más simple serán cerradas. Estas curvas se convierten en círculos mediante el uso de las variables acción-ángulo. El sistema integrable puede ser reducido a la aplicación 2-dimensional

\[
\begin{align*}
\rho_{i+1} &= \rho_i \\
\theta_{i+1} &= \theta_i + 2\pi\alpha(\rho_i)
\end{align*}
\]

o bien escrito en forma más compacta

\[
\begin{pmatrix}
\rho_{i+1} \\
\theta_{i+1}
\end{pmatrix} = T
\begin{pmatrix}
\rho_i \\
\theta_i
\end{pmatrix}
\]

donde \( T \) representa la transformación efectuada sobre el punto \((\rho, \theta)\). Se llama número de rotación a \( \alpha(\rho) \), y se puede ver que, para un sistema integrable, la aplicación anterior es la más general que representa su sección de Poincaré, y que \( \alpha(\rho) \equiv \Omega_1/\Omega_2 \) no depende del ángulo \( \theta \) (Berry, 1978).

Consideremos a continuación la aplicación perturbada

\[
\begin{align*}
\rho_{i+1} &= \rho_i + \epsilon f(\rho_i, \theta_i) \\
\theta_{i+1} &= \theta_i + 2\pi\alpha(\rho_i) + \epsilon g(\rho_i, \theta_i)
\end{align*}
\]

o, tal y como se ha hecho en el caso anterior,

\[
\begin{pmatrix}
\rho_{i+1} \\
\theta_{i+1}
\end{pmatrix} = T_t
\begin{pmatrix}
\rho_i \\
\theta_i
\end{pmatrix}
\]

donde las perturbaciones \( f \) y \( g \) dependerán evidentemente de \( H_1 \). Estamos suponiendo que el hamiltoniano \( H = H_0 + \epsilon H_1 \) es conservativo, así que la aplicación anterior también debe de serlo. Esto implica que las áreas obtenidas por aplicaciones sucesivas de \( T_t \) deben de ser iguales. Analicemos los puntos fijos de \( T_t \). Consideremos un círculo racional \( C \) no perturbado, sobre el cual

\[
\alpha(\rho) = \frac{r}{s}
\]

Todos los puntos \( P \in C \) son puntos fijos de \( T^p \) (resultado de aplicar \( T \) \( p \) veces), y el teorema KAM no nos dice nada acerca de lo que les ocurrirá cuando se aplique la perturbación. La respuesta la da el teorema de Poincaré-Birkhoff del punto fijo, que afirma que un múltiplo par de \( s \) de puntos fijos sobrevive a la perturbación.

Analicemos, pues, los puntos fijos de \( T_t \). Empecemos considerando dos círculos cercanos a \( C \), uno en el cual \( \alpha(\rho) > r/s \), al cual llamaremos \( C^+ \), y otro en el que \( \alpha(\rho) < r/s \), denominado \( C^- \) (véase la figura 17.3). La aplicación \( T^p \) aplica los puntos de \( C^- \) en el sentido de las agujas del reloj, y en sentido contrario en \( C^+ \). En \( C \), los puntos son fijos y por tanto no sufren ningún giro bajo la aplicación sucesiva de \( T_t \).

Cuando consideramos la aplicación \( T_t \), los giros relativos serán preservados cuando \( \epsilon \) sea suficientemente pequeño. Debemos por tanto poder localizar algún punto que no gire entre las dos curvas anteriores, algún punto que no cambie su coordenada angular bajo \( T_t \). Supongamos que todos los puntos para los que esto se verifica forman una curva cerrada \( R_c \) cercana a \( C \), para \( \epsilon \)
Figura 17.4: A la izquierda: curvas cerradas a las que se aplicará $T_\epsilon$. A la derecha: deformaciones sucesivas de los círculos anteriores. Las intersecciones corresponden a puntos que no giran cuando se les aplica $T_\epsilon$. Esta imagen geométrica permite deducir que los puntos hiperbólicos (h) y los puntos elípticos (e) aparecen en parejas necesariamente.

suficientemente pequeño. La imagen de esta curva será $T_\epsilon(R_\epsilon)$, y debe intersectar $R_\epsilon$ en un número par de puntos, ya que el área encerrada por ambas curvas debe ser la misma, por ser la aplicación conservativa.

Los puntos $P$ tales que $P \in \{R_\epsilon \cap T_\epsilon(R_\epsilon)\}$ deben ser puntos fijos de $T_\epsilon^n$. De la aplicación sucesiva de $T_\epsilon$ aparece una secuencia de puntos fijos elípticos e hiperbólicos. Veamos cuáles son sus características.

Consideremos la aplicación $T_\epsilon$ linealizada alrededor de sus puntos fijos, que podemos considerar situados en el origen de coordenadas (véase el capítulo sobre sistemas dinámicos). Tendremos

$$
\begin{pmatrix}
    r_{i+1} \\
    \theta_{i+1}
\end{pmatrix} =
\begin{pmatrix}
    T_{11} & T_{12} \\
    T_{21} & T_{22}
\end{pmatrix}
\begin{pmatrix}
    r_i \\
    \theta_i
\end{pmatrix} = A
\begin{pmatrix}
    r_i \\
    \theta_i
\end{pmatrix}
$$

La naturaleza del punto fijo está determinada por sus valores propios $\lambda_i$, así que, utilizando la matriz lineal, debe verificarse

$$
\left|\begin{array}{cc}
    T_{11} - \lambda & T_{12} \\
    T_{21} & T_{22} - \lambda
\end{array}\right| = 0
$$

Dado que la aplicación es conservativa, no existe contracción de las áreas en el espacio de las fases, y esta condición se traduce matemáticamente en $\det(A) = 1$, lo cual implica

$$
\lambda_2 = \frac{1}{\lambda_1}
$$

Por tanto los valores propios asociados a cada punto fijo deben ser o bien números reales que verifiquen 17.4.2, o bien números complejos conjugados en el círculo unidad. En el primer caso, el punto fijo se denomina hiperbólico, ya que las trayectorias en su vecindad pueden ser reducidas a hiperbolas, y en el segundo caso el punto fijo es elíptico, dado que las curvas invariantes son elipses (figura 17.4).

Los puntos fijos elípticos e hiperbólicos son lo que queda de los toros con frecuencias racionales tras la aplicación de la perturbación.
Figura 17.5: Estructura autosimilar formada por la perturbación. Se ha generado una cascada infinita de toros encajados.

Un punto fijo elíptico está rodeado por curvas cerradas (recordemos que estamos sobre la aplicación de Poincaré) que corresponden a nuevos toros menores, formados por la descomposición del anterior. A algunos de estos se les podrá aplicar el teorema KAM, que nos dará su estabilidad, y los restantes se descompondrán de nuevo de acuerdo con el teorema de Poincaré-Birkhoff, formando una estructura autosimilar infinita de toros encajados. El movimiento rotacional alrededor de los puntos elípticos es estable.

Los puntos fijos hiperbólicos son los responsables de la aparición de dinámica caótica. Para cada uno de estos puntos fijos tendremos dos variedades, una estable $W^s$ y una inestable $W^u$ (véase el capítulo sobre sistemas dinámicos) que tienen un comportamiento altamente irregular. Primeramente, no pueden autointersectarse, debido a la existencia y unicidad de las soluciones. En segundo lugar, sí pueden intersectarse una a otra. Esto sucede en los puntos llamados homoclínicos. Como la aplicación $T^p$ es continua, si un punto pertenece a la variedad estable debe permanecer en ella siempre, para cualquier valor de $p$. Lo mismo se aplica a los puntos que pertenecen a la variedad inestable. Así que, si un punto pertenece simultáneamente a ambas variedades, estas deben intersectarse un número infinito de veces antes de llegar a ningún punto fijo.

Este es el origen del caos en sistemas hamiltonianos. La imagen geométrica de los plegamientos y estiramientos que ambas variedades padecen se corresponde con la herradura de Smale, que se ha descrito en el capítulo sobre caos.

Esta fue la geometría del caos que descubrió Poincaré hace más de un siglo estudiando el problema de los tres cuerpos. Por desgracia, él no pudo ver las complicadas formas que ecuaciones tan sencillas como las anteriores generan. Este privilegio se hizo esperar hasta la aparición de los primeros ordenadores. Sin embargo, Poincaré pudo adivinar esta complejidad. Son sus propias palabras:

"... uno queda impactado por la complejidad de esta imagen, que ni siquiera intento dibujar. Nada puede darnos una idea mejor de la complejidad del problema de los tres"
Figura 17.6: Intersección de las variedades invariantes cuando se acercan a los puntos fijos.

cuerpos, y en general de todos los problemas de la dinámica...

Como resumen de esta sección diremos que, cuando las órbitas regulares sobre toros de un sistema integrable son perturbadas por un término no integrable, podemos obtener dinámica regular o dinámica completamente irregular, dependiendo de la condición inicial. Aunque la medida de las condiciones iniciales que llevan al movimiento regular no es nula (el teorema KAM lo explica) para cada relación racional de frecuencias (densamente distribuidas en la recta real) se obtienen toros más y más pequeños y órbitas irregulares debidas a los puntos fijos hiperbólicos. Por tanto, un cambio arbitrariamente pequeño en las condiciones iniciales conduce a comportamientos a largo término completamente diferentes. Hemos obtenido sensibilidad en las condiciones iniciales en sistemas hamiltonianos, la huella del caos.

17.6 Caos en el Sistema Solar

 Quizá las secciones anteriores, que llevan a la posibilidad de comportamientos altamente irregulares en sistemas conservativos, provoquen una sensación de irrealidad. ¿Podemos ver en realidad los toros estables, las zonas de caos con islas estables, las resonancias destructivas o los puntos hiperbólicos? La respuesta, afortunadamente, es sí. La mecánica celeste nos provee con gran cantidad de ejemplos sorprendentes. En esta sección aplicaremos los conceptos anteriores a casos reales. Algunos datos procedentes de sistemas reales serán la mejor confirmación del formalismo que hasta ahora hemos visto.

17.6.1 El cinturón de asteroides

Consideremos el movimiento de un pequeño cuerpo alrededor del Sol (de masa $M$), por ejemplo un asteroide, con masa $\mu$. Podrías imaginar que, en este problema de dos cuerpos con solución conocida desde Newton, aparece un elemento perturbador: Júpiter, el más grande de los planetas, con masa $m$.

Este nuevo problema resulta ser no integrable, como Poincaré demostró. El movimiento del asteroide podría pues inestabilizarse si la relación entre la frecuencia no perturbada del asteroide, $\Omega$, y la frecuencia angular de Júpiter, $\Omega_J$, fuese racional. Nos encontraríamos en el anillo de asteroides
con una gran cantidad de cuerpos que permiten verificar la validez de las teorías expuestas. La existencia de órbitas estables es en este caso una confirmación del teorema KAM. Véase la figura 17.8, donde se ha representado el número de asteroides catalogados en función del período orbital de Júpiter.

La escasez de asteroides con periodos que presenten relaciones simples racionales respecto del período de Júpiter fue advertida por Kirkwood en 1857. Su estudio inicial se basó en los 50 asteroides conocidos en ese momento.

Las relaciones racionales entre los periodos, las resonancias, son importantes porque significan que un asteroide y Júpiter (o cualquier pareja de cuerpos en general) se encuentran en la misma posición relativa a intervalos regulares. Estas coincidencias pueden acumularse, como si el movimiento fuese periódicamente perturbado por una fuerza externa, y el asteroide podría ser finalmente desplazado de su posición observada, llevándolo en algunos casos a cruzar la órbita de Saturno y a escapar del Sistema Solar, incluso. Por otra parte, las resonancias que observamos a nivel planetario no son coincidencias, sino el resultado de una disipación de energía lentísima que puede acabar situando dos cuerpos en resonancia. Por ejemplo, hace 2500 millones de años el período de rotación de la Tierra era de 20 horas, y la Luna se encontraba a una distancia media de 348.000 km del centro de la Tierra (a comparar con los 384.000 km actuales). Sin embargo, los efectos de marea (pérdida de energía por disipación viscosa) hicieron que la Tierra ralentizara su rotación, y paulatinamente la Luna se fue alejando. Finalmente, llegamos a la situación actual.

Puede ser que el lector esté más familiarizado con algunos problemas de osciladores forzados y amortiguados: un péndulo, por ejemplo, con cierta frecuencia de oscilación natural \( \omega_0 \), que pierda energía por rozamiento y al cual se aplica una fuerza periódica externa de frecuencia \( \omega \), acaba oscilando con esta misma frecuencia, en resonancia con la fuerza aplicada. El tiempo que tarda el movimiento en ser “capturado” por la fuerza externa depende del ritmo de disipación de energía en el péndulo.

Esta lenta disipación de energía no debe considerarse cuando plantearnos problemas a nivel celeste. Es totalmente despreciable en comparación con las escalas temporales a las que se intenta predecir la dinámica. Sería equivalente a intentar utilizar mecánica relativista para describir el movimiento de una bicicleta.

Existe un grupo de asteroides, los llamados Troianos, que parecen violar la escasez de asteroides
Figura 17.8: Esta representación del número de asteroides a varias distancias del Sol (lo cual corresponde a diferentes relaciones respecto del periodo de Júpiter) revela que muy pocos tienen órbitas con periodos correspondientes a fracciones simples del periodo orbital de Júpiter.

a relaciones racionales. Estos asteroides están situados en la misma órbita de Júpiter, por tanto la relación de frecuencias es 1:1. Sin embargo, representan un caso particular que debe ser tratado aparte. Los grupos de asteroides troyanos, el Sol y Júpiter forman los vértices de un triángulo equilátero. Lagrange descubrió en 1772 dos soluciones exactas al problema de los tres cuerpos: en los vértices de un triángulo equilátero y sobre una línea recta, a ciertas distancias. Lagrange consideró estas soluciones como una curiosidad, pero en 1906 se descubrió el primer asteroide en el grupo de los troyanos, que representan la única violación en la predicción que el teorema KAM realiza sobre los huecos que deben encontrarse en el cinturón de asteroides. Las posiciones que los troyanos ocupan se denominan puntos de Lagrange. Debemos remarcar que estos puntos no corresponden a toros, sino a órbitas cerradas aisladas.

17.6.2 Los anillos de Saturno

El problema de los tres cuerpos discutido aquí, con un cuerpo mayor $M$ como guía del movimiento, un cuerpo pequeño $\mu$ afectado por su influencia y una perturbación $m$, con $\mu << m << M$, es aplicable a otros subsistemas en el Sistema Solar. Un ejemplo típico lo dan los anillos de Saturno. Saturno guía la dinámica de las partículas materiales que constituyen sus anillos, y sus satélites actúan perturbando a estas últimas, de modo que, analógicamente a los huecos en el cinturón de asteroides, hallamos la división de Cassini en los anillos de Saturno, entre otros huecos.

Los anillos de Saturno son esencialmente "toros invariantes", y los espacios entre ellos pueden denominarse huecos KAM. La fuerza perturbadora principal es Mimas, la mayor luna de Saturno.

Presentamos a continuación un modelo muy simple que proporciona una aplicación conservativa capaz de reproducir la estructura de los anillos. Supongamos que tenemos un cuerpo de prueba con una velocidad angular $w$ situado a una distancia $r$ del centro del planeta. Sobre una órbita circular, la fuerza centripeta ($\propto w^2r$) iguale la fuerza gravitacional ($\propto r^{-2}$) con lo que la cantidad $w^2r^3$ será constante sobre todas las trayectorias circulares. Escojamos $r_0$ como radio de referencia. El tiempo que una partícula tarda en realizar una revolución situada a esta distancia se toma como
Figura 17.9: Posición relativa de los dos grupos de satélites troianos, el Sol y Júpiter.

La distancia unidad. Llamemos $\theta_n$ al ángulo que el radio vector forma con alguna dirección fijada. En general, el crecimiento de $\theta_n$ en cada paso de tiempo está dado por

$$
\theta_{n+1} = \theta_n + \frac{2\pi r_0^{3/2}}{r_n^{3/2}}
$$

para una órbita circular. Supondremos que las desviaciones respecto de ésta son tan pequeñas que la ecuación 17.6.1 es válida en general.

En el caso de Saturno, supongamos que Mimas se mueve en la órbita de referencia. Completaremos la aplicación exigiendo que sea conservativa y buscando la forma de actualizar también la distancia radial en cada paso de tiempo. La perturbación crea aceleraciones en la parte radial. Una aplicación que cumple todas las exigencias anteriores es

$$
r_{n+1} = 2r_n - h_n + \frac{A \cos(\theta_n)}{(r_0 - r_n)^2}
$$

$$
h_{n+1} = r_n
$$

junto con 17.6.1, para órbitas tales que $r_n < r_0$. Los resultados de este modelo para un gran número de partículas de prueba se muestran en la figura 17.10. Los anillos de Saturno reales presentan una estructura mucho más complicada de la que el presente modelo es capaz de proporcionar, pero el aspecto básico se reproduce. El modelo anterior se ha extraído de Frayland, 1992.

La constante $A$ se puede estimar a partir de magnitudes reales. Es directamente proporcional a la masa de Mimas, y se puede tomar como parámetro para variar la anchura de los huecos. Si se sigue con detalle el movimiento, se encuentran oscilaciones casi regulares en la dirección radial (figura 17.11), acotadas entre ciertos valores. Si se incrementa el valor de $A$, el período de estas oscilaciones decrece, en tanto que las cotas de oscilación prácticamente no varían.

17.6.3 El movimiento de Hiperión

Hiperión es uno de los satélites de Saturno, muy conocido por su forma irregular, muy lejana de la “perfección” de la esfera. Hiperión es dos veces más largo que ancho. Su diámetro menor es de unos 200 km.
Figura 17.10: Resultado de la aplicación 17.6.2 aplicada a un gran número de condiciones iniciales diferentes. La mayoría divergen en tiempos muy cortos, pero algunas son estables (no necesariamente regulares) y forman las zonas más densas del conjunto. Una parte no despreciable de los puntos representados puede pertenecer a estados transitorios, pero a nivel celeste estos transitorios son significativos en ocasiones. Los ejes se han pasado a coordenadas cartesianas con origen en el centro de Saturno, y sus unidades son de $10^3$ km.

Figura 17.11: Oscilaciones en el valor de la distancia radial al centro de Saturno para una partícula situada en el anillo más oscuro de la figura 17.10. El valor del parámetro $A$ es de $2 \times 10^{12}$ km$^3$. Aunque cerca de un movimiento periódico, las oscilaciones en el valor de $r$ no siguen ninguna pauta.
El movimiento de Hiperión presenta dos curiosidades. Por una parte, la rotación del satélite alrededor de su eje es caótica. Se conocen muchos otros cuerpos en el Sistema Solar que presentan la resonancia orbital-rotacional que hemos descrito en la Luna; siempre ofrecen la misma cara al planeta alrededor del cual orbitan. También hemos comentado que esta resonancia es posible debido a las fuerzas de marea que disipan energía y que permiten que finalmente el cuerpo quede atrapado en esta resonancia. Sin embargo, el tiempo necesario para que esto ocurra puede ser extremadamente largo. En el caso de Hiperión, se ha demostrado que este tiempo es del orden de la edad del Sistema Solar. En general, y hablando cualitativamente, se puede decir que cualquier sistema que intente llegar a un compromiso entre dos frecuencias competidoras (en el caso de Hiperión, por ejemplo, su periodo rotacional y su periodo traslacional), se encuentra empujado hacia el caos. En particular, todos los satélites de forma irregular han debido de pasar un cierto tiempo presentando dinámica caótica antes de ser atrapados en la resonancia orbital-rotacional.

La segunda curiosidad de Hiperión está relacionada con su aislamiento orbital. No se encuentra ningún otro cuerpo en sus proximidades, lo cual resulta sorprendente si consideramos que, en algún momento de su historia, habrá sido golpeado por cuerpos celestes que, si bien no lo han apartado de su órbita, pueden haber sido suficientemente fuertes como para arrancar fragmentos de su superficie. Es lo que se observa en muchos otros casos, fenómeno que provoca que haya toda una colección de pequeños fragmentos “satélite” orbitando alrededor de asteroides o satélites, por ejemplo. ¿Por qué no se observa esta agrupación de fragmentos alrededor de Hiperión? R. Bevilacqua et al. (1980) han demostrado que existe una zona caótica cercana a la órbita de Hiperión, la cual es la causante de la expulsión de estos pequeños fragmentos e impide su reagrupación. El sistema que origina esta zona caótica es el trio Saturno-Titán-Hiperión. Los fragmentos son expulsados tras un encuentro próximo con Titán.

Existen muchos otros ejemplos de caos real, detectado en series de medidas de cuerpos celestes, especialmente en el Sistema Solar. La discusión que se inició hace más de cien años, sobre la estabilidad de dicho sistema, sigue aún viva.

Actualmente se debaten ejemplos más concretos. Entre ellos, se habla del papel estabilizador de la Luna en la oblicuidad de la Tierra, y se apunta a la posible ausencia de vida en caso de no poseer un satélite del tamaño de la Luna (Laskar, 1994). Si el eje de rotación de la Tierra no hubiese poseído durante periodos de tiempo astronómicos una gran estabilidad, los cambios climáticos que se hubiesen derivado podrían haber obstaculizado el desarrollo de la vida, al menos en la forma en que hoy la conocemos. De hecho, los estudios realizados sobre las variaciones en la oblicuidad de otros planetas apuntan efectivamente a la existencia de grandes variaciones en ausencia de un elemento estabilizador (Laskar y Robutel, 1993).

Amplios estudios numéricos sobre el movimiento de Plutón, por ejemplo, parecen concluir que el planeta presenta movimiento caótico (Sussman y Wisdom, 1988). En estos análisis se han llegado a utilizar todos los planetas exteriores para evaluar el campo gravitacional total sobre Plutón. Este se halla en una resonancia 3:2 con el periodo orbital de Neptuno, y llega en ocasiones a cruzar su órbita. Estas perturbaciones son suficientemente grandes como para desestabilizar el planeta en un tiempo estimado de 20 millones de años.

Finalmente comentaremos la posible existencia de caos en el movimiento de la nube de cometas. En particular, es éste un proceso altamente dinámico, en el sentido de que algunos cometas pueden ser capturados por el Sistema Solar, especialmente debido a la influencia gravitacional de Júpiter, mientras que simultáneamente otros cuerpos son expulsados, originando un flujo constante en el aparentemente inmutable Sistema Solar. Los cometas tienen un tiempo de vida limitado y corto a las escalas que estamos tratando, a causa de la gran cantidad de materia que pierden en sus tránsitos cerca del Sol. Se ha hablado de la posible existencia de una nube de cometas alrededor del Sistema Solar, cuya causa sería una fuerte resonancia entre el periodo de Júpiter y el movimiento
de los cometas (Petrosky, 1987).

Bibliografía


Indice alfabético

ADN
- control 408
- expresión 455
- número de genes 410
agentes predictivos 204
agregación limitada por difusión 116
aliasing 235
ammonites
- fluctuaciones 436
análisis de tamaño finito 299
anticaos 424
aplicación 90
- circular 140
- de Poincaré 129
- lineal 91
- logística 72
- triangular 162, 182
ARN
- transcripción 410, 455
asteroides 580
- troyanos 581
attractor
- cuasiperiódico 137
- de Rössler 242
- extraño 148, 151
autómatas celulares 337
- computación 348
deterministas 338
elementales 338
entornos 338
- exponentes de Lyapunov 347
- legales 339
- totalistas 339
autoorganización 62
autosimilaridad 78
- estadística 118
azar 16
- y evolución 449

backpropagation 521
Bak-Sneppen, modelo de 440

teoría de campo medio 442
Barro Colorado 323
Belousov 15
Belousov-Zhabotinsky 202
bifurcación 153
- con rotura de simetría 124
- de Poincaré-Andronov-Hopf 126
- en horquilla 124
- genérica 124
- silla-nodo 122
- transcritica 123
bola
- abierta 88
- cerrada 86
Boltzmann 17, 25
- máquina de 524
brusselator 57, 371
- bifurcaciones 375
- estabilidad 372
calor
- específico 276
- latente 271
campo de vectores 43
caos 148, 150, 175
cardiaco 217
- en el sistema solar 580
- espaciotemporal 384
- hamiltoniano 565
genocíntico 205
- capacidad de un canal 31
catastrófe de error 465
células
- estimación del número 409
tipos 408
ciclo límite múltiple 132
cinturón de asteroides 580
clas de universalidad 279
clasificación de Wolfram 338
clima terrestre 253
- competencia 482
espacial 396
estabilidad 398
complejidad 15, 37
calorímetría 21
composición de funciones 90
computación universal 350
condiciones iniciales 43
conectividad 23, 89
congruencia 91
conjetura de Kaplan-Yorke 256
conjunto
abierta 88
acotado 89
cerrado 88
compacto 89
complementario 88
conexo 89, 107
de Borel 90
de Cantor
de Julia 104, 167
de Mandelbrot 106
de puntos de escape 104
de puntos prisioneros 104
diferencia 88
intersección 88
no numerable 88
numerable 88
totamente desconexo 89
unión 88
σ-límite 135
ω-límite 135
constante de Feigenbaum 156
control del caos 260
en el cerebro 506
en el hipocampo 508
en redes neurales 266, 508
método GM 265
método OG 261
conveción de Bénard 149, 223
convergencia uniforme 93
coordenadas 128
corazón humano
estructura 249
señal del 249 251
córtex cerebral
modelos 502
oscilaciones y caos 501
Courbet-Taylor 213
Creutzfeld-Jacob
atractores 499
señal 498
criticalidad autoorganizada 309
cuasiespecies 462
cuenca de atracción 167
en redes booleanas 414
determinismo 147
detección de 258
diagrama de fases 272
difusión 363
ecuación de 365
discreta 385
estabilidad 369
solución 367
dimensión
de boz-counting 113
de correlación 113, 245, 250
de orden q 113
de Hausdorff 95
de información 113
de Kolmogorov 345
de medida 80
fractal 100, 114, 169
topológica 80
dinámica simbólica 176
distancia 86
divergencia potencial 335
diversidad 21
dLA 116
duplicación de periodo 152
en sistemas neuronales 500
ecuación
de Fokker-Planck
de Hamilton-Jacobi del movimiento 567
de Mackey-Glass 204
de Newton del movimiento 566
de Perron-Frobenius 171
de reacción-difusión 369
electroencefalograma 495
dimensión fractal 496
y caos 496
entropía 15, 17, 19
condicionada 29
conjunta 29
métrica 344
topológica 344
epicentros 318
epilepsia 496
equilibrio
de Nash 439
puntuado 432
equivalencia topológica 239
ergodicidad 172
escenario
de Feigenbaum 190
de Landau 213
de Ruelle-Takens 212
espacio de fases 43
espectro
de dimensiones fractales 113
de potencia 235
estabilidad 48, 52, 65, 71
en redes booleanas 415
estructural 54
estado metaestable 311
evolución
azar en 429, 449
teoría darwiniana 427
y criticalidad 436
exponente de Lyapunov 160, 164, 219, 251, 255
espaciotemporal 390
exponentes críticos 273, 275, 286, 298, 316
relaciones entre 278
extinción
de fondo 429
deuda de la 492
modelos 445
y macroevolución 429
fago λ 411
regulación 412
Feigenbaum 156, 190
fenómeno crítico 271
fluctuaciones 277
flujo de Couette-Taylor 142
espectro de Fourier 237
forest fire 313
fractal 77, 81, 273, 309
ciudad 78
en taxonomía 437
estadístico 307
no determinista 110
frontera 88, 104
del caos 355
función
de autocorrelación 231
de correlación 35, 276, 298
de desplazamiento 131
de Hölder 91
de Lipschitz 92
de Lyapunov 65
de partición 25, 115, 275
diferenciable 93
en continuidad 93
genes
control 409
modelos 410
número de 408
genoma
control 408
gradientes 374
groupo de renormalización 279
hábitat
fragmentación 481
y coexistencia 485
y fenómenos críticos 487
hamiltoniano 274, 287, 566
Hebb, regla de 512
Hénon 165
herradura de Smale 186, 208
hiperciclos 476
dinámica espacial 479
modelo de Eigen de los 477
Hiperión, movimiento de 583
hipocampo
ccontrol de caos 508
estructura 507
hipótesis
de la frontera del caos 355
de la Reina Roja 432
homeomorfismo 93
homotecia 97
hormigas 542
comportamiento probabilista 545
oscilaciones colectivas 551
y distribución de tareas 544
y máquinas de Turing 559, 561
IFS 96
infimo 89
información 17
conjunta 28, 30, 244, 294
genética 455
inmersión 240
insectos sociales 541
intermitencia 210
templo abierto 88
isometría 91
directa 91

juego
de la vida
computación 350
reglas 349
del bosque 322
Julia, Gaston 104

Kauffman, redes de (véase redes de Kauff-
man)

lagrangiana 566
ley
de escala (o potencial) 305
de Gutenberg-Richter 317
de Zipf 306
ligaduras 23
límite
de una función 92
inferior 92
superior 92
longitud de correlación 276
Lorenz 148, 223

magnetización 276
Mandelbrot, Benoit B. 77, 106
máquina
de Boltzmann 524
de Turing 349
computación universal 561
y hormigas 559
maxent 23, 28
May, Robert 152
mecánica estadística 114
medida 93
de Hausdorff 95
de masa 94, 112
laplaciana 117
soporte de una 94, 114
medio excitable 355
método
de Derrida 419, 423
de Wolf 255
mezcla (mixing) 172, 185
modelo
de Bak-Sneppen 440
de córtex cerebral 502
de Eigen de replicación 458
de evolución con extinción explícita 445
de Gierer-Meinhardt 376
estabilidad 377
patrones 2D 380
de Ginzburg-Landau 286
de Hénon 165
control del 264
de Hopfield 510
capacidad 513, 518
función energía 515
de huésped parasitóide 400
de Ising 274
simulación del 290
de la pila de arena 310
de Levins 481
de Lorenz 148, 223
de los anillos de Saturno 582
de Lotka-Volterra 73, 163, 209
espectro de Fourier 236
de Murray 380
de Novak 470
de Rössler 198
de Schuster-Wagner 502
de terremotos 321
de virus 470
del bosque en llamas 313
del juego del bosque 322
HKF 534
NK 413
NKC 438
morfogénesis 361
multifractal 112, 327
multiplicadores de Lagrange 24
multiplicidad 132
mutación, matriz de 459

neurodinámica 495
número
de Lyapunov 127, 132
de rotación 140
mal aproximable por racionales 141

ondas espirales 400
operador shift 177
órbita
cuasiperiódica 121, 137
heteroclínica 121
homeoclínica 121, 205
Indice alfabético

periodica 129
orden por fluctuaciones 548
organización 61

paisaje adaptativo 433
parámetro
  de orden 63, 272, 276
  \( \lambda \) de Langton 353
paréntesis de Poisson 569
Peano, curva de 81, 84
percolación 295
  en redes de Kauffman 421
predicción 147, 157, 334
principio de control 61
probabilidad 112
problema
  de los denominadores pequeños 141, 575
  de los dos cuerpos 567
  de los tres cuerpos 571
procesos estocásticos 39
puerta lógica 350
punto de Lagrange 582
punto fijo 46, 49, 96, 98, 105, 281
  elíptico 578
  hiperbólico 55, 578
puntuacionismo 432

random walk 19, 24, 116, 307
recubrimiento 95
red de Hopfield 510
redes acopladas 384
  estabilidad estructural 394
  logísticas 385
  bifurcaciones 388
  estabilidad 386
redes de Kauffman
  booleanas 407, 413
  generalizadas 422
  mecánica estadística 419
  percolación 421
  red \( K = 1 \) 417
  red \( K = 2 \) 417
  red \( K > 5 \) 416
  red \( K = N \) 415
  transición de fase 421, 425
redes neurales
  con intermediarios 529
  control de caos 508
  fluidas 541, 552
  reflexión 91

región laminar 211
regla de Hebb 512
relaciones de hiperescala 300
relajación crítica 535
renormalización
  en el espacio real 279, 289, 300
  grupo de 279
resonancia 572, 575
retropropagación 521
retrovirus 456
  y mutabilidad 465
  y sistema inmune 468
rotación 91
rotura de simetría 63
  espacial 397
ruido \( 1/f \) 308, 328

salto adaptativo 438
Saturno, anillos de 582
sección de Poincaré 133, 198, 203
secuencia de Bernoulli 186
sensibilidad a las condiciones iniciales 106, 156, 179, 184, 580
Shigamare 340
SIDA 456
Sierpiński, Waclaw 85
  triángulo de 85
Sierpiński-Menger, esponja de 85, 87
sistema
  crítico autoorganizado 305, 309
  de funciones iteradas (IFS) 96
  débilmente caótico 335
sistema dinámico 43
  continuo 46
  integrable 569
  lineal autónomo 46
sistema dinámico discreto 43, 69
  de Lotka-Volterra 73, 163
  en 3D 209
sistema inmune
  y redes neurales 541
sistema no autónomo 204
slaving principle 61
SQUID, dispositivo 532
sucesión convergente 89
supertransitorios 393, 395
supremo 89
susceptibilidad térmica 277

temperatura crítica 275
teorema
de Li y Yorke 195
de Poincaré-Bendixon 135
de Poincaré-Birkhoff 576
de Sharkovsky 198
de Shilnikov 206
de Smale 189
de Whitney 239, 241
de Whitney-Takens 241
del Collage 99
KAM 575
teoría
de campo medio 282, 318
de Landau 288
de perturbaciones 572
terremotos 317
tiempos 121, 138
tenorados 579
transcripción 411, 455
transformación
afín 91, 97
de semejanza 91, 97
del panadero 168
transformada de Fourier 233
transición de fase 28, 271
critica 272
de primer orden 272
e información 555
en el cerebro 532
inducida por ruido 558
modelo HKF 534
transitividad topológica 172, 176
traslación 91
Turing, Alan 363
estructuras de 365
caos en las 397
umbral de diversidad 470
universalidad 279, 303
valores propios 47, 122
variables acción-ángulo 570
variedad 51, 57, 123, 142
centro 142
vectores propios 47
virus
de ARN 456