
Elementos de Matemáticas y Aplicaciones

M. Castrillón A. Díaz-Cano J. J. Etayo M. Folgueira
J. A. Infante L. M. Pozo J. M. Rey

VERSIÓN: 2014



Facultad de Matemáticas. Universidad Complutense de Madrid



Índice

Introducción	7
1. Números enteros. Dígitos de control y criptografía	9
1.1 Relaciones en los enteros	9
1.2 La división euclídea	10
1.3 El máximo común divisor	10
1.4 Los números primos	12
1.5 Aplicación: Dígitos de control	14
1.5.1 La letra del N.I.F.	14
1.5.2 Los dígitos de control en el D.N.I. electrónico	15
1.5.3 El Número de Registro Personal de los Funcionarios	17
1.6 Congruencias	17
1.7 Aplicación: Criptografía	25
1.7.1 Procedimiento aditivo	25
1.7.2 Procedimiento multiplicativo	25
1.7.3 Procedimiento exponencial	26
1.7.4 El procedimiento RSA	29
1.8 Problemas	29
2. Grupos de simetrías. Mosaicos	35
2.1 Las simetrías rotacionales de un tetraedro	35
2.2 Grupos. Isomorfismos y homomorfismos de grupos	35
2.2.1 Definición axiomática de grupo	35
2.2.2 Subgrupos. Grupos finitos. Orden	38
2.2.3 Isomorfismos: cuándo dos grupos son “iguales”	40
2.2.4 Homomorfismos de grupos. Núcleo e imagen	42
2.2.5 Subgrupos normales. El grupo cociente	42
2.3 Isometrías del plano y matrices ortogonales	43
2.4 Grupo diédrico e isometrías del plano	53
2.5 Grupos cristalográficos planos	56
2.6 Retículos y su clasificación	57
2.7 Aplicación: Clasificación de los mosaicos del plano	61
2.8 Problemas	66
3. Trigonometría plana y esférica. Aplicaciones	73
3.1 Introducción y contenidos	73
3.2 Trigonometría plana	73
3.2.1 Revisión	73
Funciones trigonométricas: sen, cos, tan, csc, sec, cot	73
Valores de las funciones trigonométricas para 0° , 90° , 180° y 270°	74
Relaciones fundamentales	74
Identidades	74
Funciones trigonométricas de dos ángulos	74
Valores de las funciones trigonométricas para 30° , 45° y 60°	76
Valores de las funciones trigonométricas para $-\alpha$, $90^\circ \pm \alpha$, $180^\circ \pm \alpha$, $270^\circ \pm \alpha$ y $k \cdot 360^\circ \pm \alpha$ siendo α un ángulo del primer cuadrante	76
Triángulos planos. Leyes y fórmulas generales	76
3.2.2 Aplicaciones de la trigonometría plana a la navegación y Geodesia	77



3.3	Trigonometría esférica	77
3.3.1	Geometría esférica	77
3.3.2	Triángulos esféricos	80
3.4	Aplicaciones de la trigonometría esférica a la navegación	88
3.5	Sistemas de coordenadas en la esfera celeste	90
3.5.1	La esfera celeste	90
3.5.2	Sistemas de coordenadas celestes	91
	Sistema horizontal (o altacimutal)	91
	Sistema ecuatorial horario	92
	Sistema de coordenadas ecuatorial absoluto	93
	Transformaciones entre sistemas de coordenadas celestes	94
3.6	El elipsoide terrestre	97
3.6.1	Forma y dimensiones de la Tierra. Reseña histórica	97
	Polémica sobre la forma de la Tierra	98
	Unidades de longitud y su relación con las dimensiones terrestres	98
	La Tierra representada por elipsoides	99
3.6.2	Posición de un observador en la superficie de la Tierra	99
	Coordenadas geocéntricas	100
	Coordenadas geodésicas	101
	Coordenadas astronómicas (o geográficas)	101
3.7	Posicionamiento utilizando satélites artificiales	102
	Componentes principales del sistema NAVSTAR–GPS	102
	Posicionamiento con GPS	104
3.8	Problemas	107
4.	Dinámica discreta	117
4.1	Sistemas dinámicos discretos. Terminología	117
4.2	Sistemas dinámicos lineales de primer orden	122
4.2.1	Puntos de equilibrio: estabilidad, análisis gráfico	123
4.2.2	Solución general y particular del sistema homogéneo	129
4.2.3	Solución general y particular del sistema afín	132
4.2.4	Aplicaciones	135
	Finanzas	135
	Datación mediante carbono 14	137
	Evolución genética	137
4.3	Sistemas dinámicos lineales de segundo orden	139
4.3.1	Sistemas homogéneos: Polinomio característico, solución general y particular	140
4.3.2	Sistemas afines: Solución general y particular	145
4.3.3	Aplicaciones	147
	Apuestas	147
	Endogamia	149
4.4	Sistemas dinámicos no lineales: Estabilidad y caos	151
4.4.1	Estabilidad	151
4.4.2	La ecuación logística	156
4.4.3	Ciclos y caos	160
4.5	Problemas	167
5.	Teoría de grafos y aplicaciones	175
5.1	Definiciones. Lema del apretón de manos	175
5.2	Isomorfismo de grafos	180
5.3	Matriz de adyacencia	182
5.4	Caminos eulerianos y hamiltonianos	186
5.4.1	Caminos eulerianos	186
5.4.2	Caminos hamiltonianos	189
5.5	Grafos planos	190
5.5.1	Fórmula de Euler	192
5.6	Coloración de grafos	193
5.7	Teoría espectral de grafos. Aplicación a Google	194
5.8	Problemas	199



Bibliografía 209





Introducción

La asignatura *Elementos de Matemáticas y aplicaciones* está encuadrada en el primer curso de los tres grados (Ingeniería Matemática, Matemáticas y Matemáticas y Estadística) que se imparten en la Facultad de Ciencias Matemáticas de la Universidad Complutense de Madrid. Se trata de una materia obligatoria, de carácter transversal y en la que se pretende dar una idea de la aplicabilidad de las Matemáticas en cuestiones cercanas a lo cotidiano.

Para ello, se han elegido algunos temas que forman parte de la que podríamos llamar *cultura general matemática*, cuya presentación se lleva a cabo tras haber estudiado la teoría matemática en la que se asientan, sin renunciar en ningún momento al rigor y precisión necesarios. En los cinco capítulos que constituyen el cuerpo de la asignatura se ha intentado destilar la parte teórica imprescindible, de forma que, siendo autocontenida, sirva para demostrar los resultados que vienen a constituir los cimientos de las distintas aplicaciones. Algunos de estos contenidos matemáticos constituyen un primer contacto con partes de la matemática que serán desarrolladas *in extenso* en asignaturas de cursos sucesivos.

A modo de resumen, podemos citar que la aplicación estrella del primer capítulo es el sistema de encriptado RSA; la *Aritmética modular* será la teoría necesaria para llegar hasta ella. El segundo capítulo tiene como fin último la clasificación de los mosaicos planos; para alcanzarlo, se introduce previamente la correspondiente *Teoría de grupos*. La Astronomía de posición y el posicionamiento por satélite (GPS) se presentan tras haber introducido la *Trigonometría esférica*. El estudio de la *Dinámica discreta* permite desarrollar algunas aplicaciones de tipo financiero, así como hacer una primera incursión en el caos a través de la ecuación logística. Finalmente, el algoritmo de ordenación de Google es la guinda que culmina la *Teoría de grafos* estudiada.

El presente trabajo es fruto de la experiencia de varios años impartiendo esta asignatura y ha sido realizado por un amplio grupo de profesores. Partiendo de diversos textos clásicos de las materias desarrolladas, hemos reelaborado el material con vistas a conseguir el enfoque pretendido que relacione la teoría con las aplicaciones. También hemos desarrollado ejemplos, gráficos e incluso un *software* propio, todo lo cual pensamos que facilita una mejor comprensión de los conceptos y resultados presentados.





1 Números enteros. Dígitos de control y criptografía

En este capítulo se presentan resultados relativos a números enteros y a la división euclídea y se estudian las congruencias lineales con vistas a definir la función de Euler. Todo ello permitirá analizar los sistemas de dígitos de control y los procedimientos de encriptación.

1.1 Relaciones en los enteros

Comenzamos recordando el concepto de relación de orden:

Definición 1.1.1 Dado un conjunto X diremos que \mathfrak{R} es una *relación de orden* en X si verifica las siguientes propiedades:

- a) *Reflexiva*: $x \mathfrak{R} x$ para todo $x \in X$.
- b) *Antisimétrica*: si $x, y \in X$ verifican $x \mathfrak{R} y$ e $y \mathfrak{R} x$ entonces $x = y$.
- c) *Transitiva*: si $x, y, z \in X$ verifican $x \mathfrak{R} y$ e $y \mathfrak{R} z$ entonces $x \mathfrak{R} z$. \square

En este capítulo vamos a trabajar con el conjunto de los números enteros, \mathbb{Z} , y su subconjunto \mathbb{N} de los números naturales. En ambos conjuntos existe una relación de orden “natural”, la determinada por

$$a \mathfrak{R} b \text{ si y sólo si } b - a \in \mathbb{N} \cup \{0\} \text{ (es decir, } a \leq b). \quad (1.1)$$

Definición 1.1.2 Sea X un conjunto y \mathfrak{R} una relación de orden en X . Diremos que \mathfrak{R} es un *orden total* en X si para todo par de elementos $x, y \in X$ se verifica que $x \mathfrak{R} y$ o $y \mathfrak{R} x$. \square

Ejercicio 1.1.3 Comprobar que (1.1) es una relación de *orden total* en \mathbb{N} y en \mathbb{Z} . \square

Definición 1.1.4 Sea X un conjunto y \mathfrak{R} una relación de orden en X . Diremos que \mathfrak{R} es una *buena ordenación* si para cada subconjunto no vacío Y de X existe un elemento $a \in Y$ de forma que a está relacionado con todos los elementos de Y . \square

Ejemplo 1.1.5 La relación (1.1) es una buena ordenación en \mathbb{N} , ya que el menor elemento de cualquier subconjunto Y está relacionado con todos los elementos de Y , sea éste cual sea. En cambio, la relación de orden (1.1) no es una buena ordenación en \mathbb{Z} , pues basta tomar un subconjunto en el que no exista “el menor número” (por ejemplo, basta tomar como Y los enteros negativos para que no haya ningún elemento relacionado con todos). De hecho, para cualquier $a \in \mathbb{Z}$, el conjunto

$$\{m \in \mathbb{Z} : a \leq m\}$$

está bien ordenado con la relación (1.1). \square

Observación 1.1.6 Todo buen orden es un orden total. \square

En el conjunto de números enteros se pueden definir otras muchas relaciones; una fundamental es la de divisibilidad:

Definición 1.1.7 Dados dos números enteros m y n , diremos que m divide a n si existe un entero d tal que $n = md$. Denotaremos esta situación por $m|n$. \square

Ejercicio 1.1.8 Comprobar que la relación de divisibilidad en \mathbb{N} es una relación de orden (no total) mientras que no lo es en \mathbb{Z} (porque no es antisimétrica). \square



1.2 La división euclídea

Vamos a probar que dados dos números enteros, uno de ellos distinto de 0, el otro se puede dividir entre éste obteniendo el cociente y el resto, y que éstos son únicos.

Teorema 1.2.1 Sean $n, m \in \mathbb{Z}$ con $m \neq 0$. Se verifica que existen unos únicos enteros q y r tales que $n = mq + r$, con $0 \leq r < |m|$. Los números q y r se denominan, respectivamente, cociente y resto de la división de n entre m .

DEMOSTRACIÓN.

a) Supongamos primero que $m > 0$. Nótese que para cualquier $q \in \mathbb{Z}$ se tiene que

$$n = mq + (n - mq).$$

La idea es encontrar un entero q de forma que $n - mq$ satisfaga la hipótesis para r . Consideramos el conjunto

$$S = \{n - mx, x \in \mathbb{Z}\},$$

el cual contiene números no negativos (compruébese como ejercicio). Sea S_0 el subconjunto formado por tales números, esto es,

$$S_0 = \{n - mx, x \in \mathbb{Z} \text{ y } n - mx \geq 0\}.$$

Puesto que $S_0 \neq \emptyset$, podemos considerar su menor elemento $r = n - mq$, con $q \in \mathbb{Z}$. Es obvio que $r \geq 0$ (pues $r \in S_0$). Supongamos que $r \geq m$ y lleguemos a una contradicción; en efecto, en tal caso se tendría que

$$0 \leq r - m = n - mq - m = n - m(q + 1) = r - m < r$$

y, por tanto, $n - m(q + 1)$ sería un elemento de S_0 menor que el primero r .

b) Si $m < 0$, basta aplicar el apartado anterior a $-m$, pues

$$n = (-m)q + r = m(-q) + r.$$

Para ver la unicidad, supongamos que hay dos cocientes y dos restos,

$$n = mq_1 + r_1 = mq_2 + r_2$$

con

$$0 \leq r_1 < |m| \text{ y } 0 \leq r_2 < |m|.$$

Si $q_1 \neq q_2$ entonces $|q_1 - q_2| \geq 1$ y $|m(q_1 - q_2)| = |m||q_1 - q_2| \geq |m|$. Por otra parte,

$$|m(q_1 - q_2)| = |r_1 - r_2| < |m|,$$

lo que contradice la desigualdad anterior. Por lo tanto, $q_1 = q_2$ y, en consecuencia, $r_1 = r_2$. \square

Observación 1.2.2 Nótese que en la división euclídea el resto es siempre no negativo. Por ejemplo, la división de -8 entre 3 proporciona como cociente -3 y resto 1 (y no cociente -2 y resto -2). \square

1.3 El máximo común divisor

Definición 1.3.1 Dados dos números enteros no nulos m y n , diremos que un número positivo d es un *máximo común divisor* de m y n si divide a ambos, y si cualquiera que divide a ambos divide a d . \square

Los siguientes resultados tratan sobre la existencia y unicidad del máximo común divisor de dos enteros no nulos, así como de la identidad de Bézout. En primer lugar, se presenta el conocido algoritmo de Euclides para el cálculo del máximo común divisor de dos números (garantizando de paso su existencia). Este algoritmo es particularmente interesante cuando los números son “grandes” pues, en tal caso, la conocida regla de los factores primos comunes no es practicable.



Teorema 1.3.2 (Algoritmo de Euclides) *Dados m y n enteros no nulos, denotando $r_0 = n$ y $r_1 = m$ se forma la secuencia de restos*

$$\begin{cases} r_0 = r_1 q_1 + r_2 \\ r_1 = r_2 q_2 + r_3 \\ r_2 = r_3 q_3 + r_4 \\ \dots \\ r_{i-2} = r_{i-1} q_{i-1} + r_i \\ \dots \\ r_{k-2} = r_{k-1} q_{k-1} + r_k \\ r_{k-1} = r_k q_k. \end{cases} \quad (1.2)$$

Entonces, el último resto no nulo r_k es un máximo común divisor de m y n .

DEMOSTRACIÓN. La sucesión anterior $\{r_i\}_{i=0}^k$ está bien definida al ser el proceso finito, ya que

$$|m| = |r_1| > r_2 > r_3 > \dots > r_i > \dots$$

puesto que el resto siempre es menor que el módulo del divisor. Es decir, cada resto es estrictamente menor que el anterior y, por lo tanto, necesariamente se llega a un resto 0 en, a lo sumo, $|m|$ etapas.

Veamos que, efectivamente, r_k es máximo común divisor de m y n . En primer lugar, recorremos la colección de igualdades (1.2) desde abajo. De la última resulta que r_k divide a r_{k-1} , ahora con esto de la penúltima deducimos que r_k divide a r_{k-2} y, siguiendo hasta arriba, llegamos a que r_k divide a m y a n . Por lo tanto, r_k es un divisor común de ambos números. Ahora supongamos que también j divide a ambos. Repetimos el recorrido, pero de arriba hacia abajo. De la primera igualdad resulta que j divide a r_2 , ahora usando esto, de la segunda obtenemos que j divide a r_3 y, reiterando el proceso, llegamos finalmente a que j divide a r_k . \square

Teorema 1.3.3 *Si $m, n \in \mathbb{Z}$ con $m, n \neq 0$, el máximo común divisor de m y n es único.*

DEMOSTRACIÓN. Si hubiera dos máximos comunes divisores, digamos d_1 y d_2 , como cada uno de ellos divide a los dos números m y n , cada uno divide también al otro máximo común divisor. Por lo tanto son dos números positivos que se dividen mutuamente y, en consecuencia, son iguales. \square

Notación 1.3.4 Desde ahora hablaremos del *máximo común divisor de dos números m y n* , y escribiremos $\text{mcd}(m, n)$. \square

Observación 1.3.5 Despejando r_k en la penúltima igualdad de (1.2), a continuación sustituyendo r_{k-1} por el resultado de despejarlo en la antepenúltima igualdad, y recorriendo todo el camino otra vez desde abajo hasta arriba, obtenemos una igualdad que expresa el máximo común divisor r_k como combinación entera de m y n . Se obtiene así lo que se conoce como una *identidad de Bézout*. El siguiente resultado proporciona un método eficiente para hallar una identidad de Bézout entre dos enteros no nulos. \square

Teorema 1.3.6 (Algoritmo extendido de Euclides. Identidad de Bézout) *Con la notación utilizada en el Teorema 1.3.2, consideramos las secuencias $\{x_i\}_{i=0}^k$ e $\{y_i\}_{i=0}^k$ definidas por*

$$\begin{cases} x_0 = 1, x_1 = 0 \\ y_0 = 0, y_1 = 1 \end{cases}$$

y

$$\begin{cases} x_i = x_{i-2} - x_{i-1} q_{i-1} \\ y_i = y_{i-2} - y_{i-1} q_{i-1} \end{cases}$$

para $i = 2, 3, \dots, k$. Entonces, se verifica que

$$r_i = x_i r_0 + y_i r_1, \quad i = 0, 1, \dots, k.$$

En particular, se tiene la identidad de Bézout

$$\text{mcd}(n, m) = r_k = x_k r_0 + y_k r_1 = x_k n + y_k m.$$



DEMOSTRACIÓN. Lo probamos por inducción completa en i :

- a) Es obvio que $r_0 = 1 \times r_0 + 0 \times r_1 = x_0 r_0 + y_0 r_1$ y que $r_1 = 0 \times r_0 + 1 \times r_1 = x_1 r_0 + y_1 r_1$.
- b) Para $i \geq 2$, suponemos cierto el resultado para los índices menores que i y lo probamos para i . Por la hipótesis de inducción, se verifica que

$$\begin{aligned} x_i r_0 + y_i r_1 &= (x_{i-2} - x_{i-1} q_{i-1}) r_0 + (y_{i-2} - y_{i-1} q_{i-1}) r_1 \\ &= (x_{i-2} r_0 + y_{i-2} r_1) - (x_{i-1} r_0 + y_{i-1} r_1) q_{i-1} \\ &= r_{i-2} - r_{i-1} q_{i-1} = r_i. \quad \square \end{aligned}$$

Observación 1.3.7 Nótese que, en general, la identidad de Bézout no es única. Por ejemplo, $\text{mcd}(6, 4) = 2$ y se tiene

$$6 \times 1 + 4 \times (-1) = 2 \quad \text{y} \quad 6 \times (-1) + 4 \times 2 = 2. \quad \square$$

Ejemplo 1.3.8 Si tomamos los números $n = 249$ y $m = 36$, a partir de la tabla

i	r_i	q_i	x_i	y_i
0	249		1	0
1	36	6	0	1
2	33	1	1	-6
3	3	11	-1	7
	0			

se tiene que $\text{mcd}(249, 36) = 3$ y la identidad de Bézout

$$3 = 249 \times (-1) + 36 \times 7. \quad \square$$

El concepto dual de máximo común divisor es el de mínimo común múltiplo.

Definición 1.3.9 Dados dos números enteros no nulos m y n , diremos que un número positivo l es un *mínimo común múltiplo* de m y n si ambos lo dividen, y divide a cualquier número t al que m y n dividan. \square

Observación 1.3.10 La unicidad del mínimo común múltiplo se prueba del mismo modo que la del máximo común divisor. Relegamos el procedimiento de cálculo (y, consecuentemente, la existencia) hasta el final de la Sección 1.4 relativa al estudio de los números primos. \square

1.4 Los números primos

Definición 1.4.1 Un número natural p mayor que 1 es un *número primo* si sus únicos divisores naturales son 1 y p . A los naturales mayores que 1 que no son primos los llamaremos *compuestos*. \square

Proposición 1.4.2 *Todo número natural n mayor que 1 tiene algún divisor primo.*

DEMOSTRACIÓN. Lo demostramos por inducción completa sobre n .

- a) Evidentemente 2 admite un factor primo, el propio 2.
- b) Ahora sea $n > 2$ y supongamos que hasta $n - 1$ todos los números admiten algún divisor primo. O bien n es primo, y ya tenemos su factor primo, o bien es compuesto, y entonces es divisible por algún natural t distinto de 1 y de n . Gracias a la hipótesis de inducción, t tiene algún divisor primo, que lo será también de n . \square

Proposición 1.4.3 *Todo número natural n mayor que 1 se puede expresar como producto de primos.*

DEMOSTRACIÓN. Nuevamente lo probamos por inducción.



- a) Para 2 se cumple.
- b) Sea ahora $n > 2$ y suponemos cierto el resultado hasta $n - 1$. Como n tiene algún divisor primo p , si éste coincide con n ya está; si no, se tiene que $n = kp$ con k menor que n y, por la hipótesis de inducción, producto de primos. En consecuencia, n también lo es. \square

Teorema 1.4.4 *El conjunto de los números primos es infinito.*

DEMOSTRACIÓN (Euclides). Supongamos que dicho conjunto fuera finito, digamos

$$P = \{p_1, p_2, \dots, p_n\}.$$

Entonces tomamos

$$k = 1 + p_1 p_2 \cdots p_n.$$

Ninguno de los primos p_i divide a k , pero éste tiene algún divisor primo, que por lo tanto tiene que ser alguno que no está en P . \square

Definición 1.4.5 Diremos que dos números $a, b \in \mathbb{Z} \setminus \{0\}$ son *primos entre sí* (o que a es primo con b , o b con a) si $\text{mcd}(a, b) = 1$. \square

Teorema 1.4.6 *Sean $a, b \in \mathbb{Z} \setminus \{0\}$ y $m \in \mathbb{Z} \setminus \{0\}$ primo con a . Si m divide a ab , entonces m divide a b .*

DEMOSTRACIÓN. Como m es primo con a , el máximo común divisor de ambos es 1 y, por lo tanto, tenemos una identidad de Bézout: existen $u, v \in \mathbb{Z}$ tales que

$$1 = mu + av.$$

Multiplicando por b , resulta

$$b = bmu + abv$$

y, como m divide a estos dos sumandos, divide a su suma que es b . \square

Corolario 1.4.7 *Sean $a, b \in \mathbb{Z} \setminus \{0\}$ y p un número primo. Si p divide a ab , entonces p divide a alguno de los dos.* \square

Observación 1.4.8 Por supuesto, el Corolario 1.4.7 se extiende a cualquier producto finito, por inducción sobre el número de factores. \square

Teorema 1.4.9 *La expresión de cada número natural n mayor que 1 como producto de primos es única (salvo el orden).*

DEMOSTRACIÓN. Vamos a probarlo por inducción sobre n .

- a) Por supuesto, el resultado es cierto para 2.
- b) Si $n > 2$, ya hemos visto que n es producto de primos. Supongamos que tenemos dos expresiones,

$$n = p_1 p_2 \cdots p_r = q_1 q_2 \cdots q_s,$$

donde los p_i y los q_j son primos y están ordenados de menor a mayor. Como p_1 divide a la primera expresión, divide a la segunda, y como es primo, divide a alguno de los factores, digamos q_j . Como éste también es primo, $p_1 = q_j$. Del mismo modo, q_1 divide a algún p_i y, por lo tanto, $q_1 = p_i$. Pero entonces

$$p_1 = q_j \geq q_1 = p_i \geq p_1.$$

Por lo tanto, $p_1 = q_1$. Ahora, dividiendo las dos expresiones por p_1 y q_1 , tenemos

$$p_2 p_3 \cdots p_r = q_2 q_3 \cdots q_s$$

y, aplicando la hipótesis de inducción, resulta que $r = s$ y cada p_i coincide con cada q_i . \square



Observación 1.4.10 Con este resultado queda garantizada la unicidad de *descomposición en factores primos*

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r} \text{ con } p_1 < p_2 < \cdots < p_r,$$

que es el modo razonable de expresar un número como producto de primos, agrupando los que se repiten, y ordenándolos de menor a mayor. \square

Observación 1.4.11 Volvemos ahora a traer a colación al máximo común divisor y al mínimo común múltiplo. Supongamos que tenemos dos naturales m y n mayores que 1, y expresemos ambos como producto de primos, permitiendo que los exponentes puedan ser 0 para incluir en ambas expresiones los mismos primos. Sean entonces

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_r^{\alpha_r} \text{ y } m = p_1^{\beta_1} p_2^{\beta_2} \cdots p_r^{\beta_r},$$

donde los exponentes $\alpha_i, \beta_i \geq 0$. Si para cada $i \in \{1, 2, \dots, r\}$ denotamos

$$\gamma_i = \min\{\alpha_i, \beta_i\} \text{ y } \delta_i = \max\{\alpha_i, \beta_i\},$$

es claro que

$$\text{mcd}(m, n) = p_1^{\gamma_1} p_2^{\gamma_2} \cdots p_r^{\gamma_r} \text{ y } \text{mcm}(m, n) = p_1^{\delta_1} p_2^{\delta_2} \cdots p_r^{\delta_r}. \quad \square$$

Observación 1.4.12 Las expresiones anteriores garantizan que para $m, n \in \mathbb{Z} \setminus \{0\}$ se tiene que

$$\text{mcd}(m, n) \text{mcm}(m, n) = |mn|.$$

En particular, si dos números son primos entre sí, su mínimo común múltiplo es el valor absoluto de su producto. \square

Ejemplo 1.4.13 Para hallar el mínimo común múltiplo de 249 y 36, puesto que $\text{mcd}(249, 36) = 3$ (como se vio en el Ejemplo 1.3.8), basta con calcular

$$249 \times \frac{36}{3} = 2988. \quad \square$$

1.5 Aplicación: Dígitos de control

Los *dígitos de control* se utilizan para comprobar que una colección de datos, que son los relevantes, son los correctos. Por lo tanto deben poder obtenerse de éstos de tal forma que la inconsistencia del conjunto implique que algún dato es erróneo. Se obtienen fundamentalmente calculando el resto de dividir un número deducido del relevante entre n , para algún n dado. Ejemplos bien conocidos en nuestra vida cotidiana son los siguientes:

1.5.1 La letra del N.I.F.

El *Número de Identificación Fiscal* está formado por el número del D.N.I. más una letra. Esta letra se obtiene calculando primero el resto de dividir el número del D.N.I. entre 23, y traduciendo luego cada uno de los 23 posibles restos a una letra predeterminada, según la siguiente tabla:

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
T	R	W	A	G	M	Y	F	P	D	X	B	N	J	Z	S	Q	V	H	L	C	K	E

Ejemplo 1.5.1 Z es la letra que le corresponde al D.N.I. 12345678, ya que 12345678 dividido entre 23 da de resto 14. \square

Ejercicio 1.5.2 Comprobar la letra de vuestro N.I.F. \square

Comprobemos que sirve como dígito de control.



Veamos cómo se obtienen los 4 dígitos de control anteriores:

- a) Dígito de control **a**: es la última cifra del número que se obtiene sumando los productos de las cifras del número de serie del soporte, respectivamente, por 7, 3, 1, 7, 3, 1, 7, 3, 1, una vez reconvertidas las tres primeras letras según la tabla

A	B	C	D	E	F	G	H	I	J	K	L	M
10	11	12	13	14	15	16	17	18	19	20	21	22
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
23	24	25	26	27	28	29	30	31	32	33	34	35

Como, en nuestro caso, $X = 33$, $Y = 34$ y $Z = 35$, debemos efectuar la operación

$$\begin{array}{r} 33 \quad 34 \quad 35 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \\ \times \quad 7 \quad 3 \quad 1 \quad 7 \quad 3 \quad 1 \quad 7 \quad 3 \quad 1 \\ \hline 231 \quad 102 \quad 35 \quad 7 \quad 6 \quad 3 \quad 28 \quad 15 \quad 6 \end{array}$$

Puesto que

$$231 + 102 + 35 + 7 + 6 + 3 + 28 + 15 + 6 = 433,$$

se tiene que $a = 3$.

- b) Dígito de control **b**: Se procede como en el caso anterior para el campo 7, es decir

$$\begin{array}{r} 8 \quad 9 \quad 1 \quad 1 \quad 2 \quad 1 \\ \times \quad 7 \quad 3 \quad 1 \quad 7 \quad 3 \quad 1 \\ \hline 56 \quad 27 \quad 1 \quad 7 \quad 6 \quad 1 \end{array}$$

Puesto que

$$56 + 27 + 1 + 7 + 6 + 1 = 98,$$

se tiene que $b = 8$.

- c) Dígito de control **c**: Se procede como en los casos anteriores para el campo 10, es decir

$$\begin{array}{r} 1 \quad 5 \quad 0 \quad 6 \quad 2 \quad 3 \\ \times \quad 7 \quad 3 \quad 1 \quad 7 \quad 3 \quad 1 \\ \hline 7 \quad 15 \quad 0 \quad 42 \quad 6 \quad 3 \end{array}$$

Puesto que

$$7 + 15 + 0 + 42 + 6 + 3 = 73,$$

se tiene que $c = 3$.

- d) Dígito de control **d**: Se procede como en los casos anteriores para la concatenación de los campos 3, 4, 5, 7, 8, 10 y 11 (sustituyendo las letras por números de acuerdo a la tabla anterior). Es decir,

$$\begin{array}{r} 33 \quad 34 \quad 35 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 3 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 35 \\ \times \quad 7 \quad 3 \quad 1 \quad 7 \quad 3 \quad 1 \quad 7 \quad 3 \quad 1 \quad 7 \quad 3 \quad 1 \quad 7 \quad 3 \quad 1 \quad 7 \quad 3 \quad 1 \quad 7 \\ \hline 231 \quad 102 \quad 35 \quad 7 \quad 6 \quad 3 \quad 28 \quad 15 \quad 6 \quad 21 \quad 3 \quad 2 \quad 21 \quad 12 \quad 5 \quad 42 \quad 21 \quad 8 \quad 245 \\ \\ 8 \quad 9 \quad 1 \quad 1 \quad 2 \quad 1 \quad 8 \quad 1 \quad 5 \quad 0 \quad 6 \quad 2 \quad 3 \quad 3 \\ \times \quad 3 \quad 1 \quad 7 \quad 3 \quad 1 \quad 7 \quad 3 \quad 1 \quad 7 \quad 3 \quad 1 \quad 7 \quad 3 \quad 1 \\ \hline 24 \quad 9 \quad 7 \quad 3 \quad 2 \quad 7 \quad 24 \quad 1 \quad 35 \quad 0 \quad 6 \quad 14 \quad 9 \quad 3 \end{array}$$

Puesto que la suma de los resultados obtenidos es

$$\begin{aligned} &231 + 102 + 35 + 7 + 6 + 3 + 28 + 15 + 6 + 21 + 3 + 2 + 21 + 12 + 5 + 42 + 21 \\ &+ 8 + 245 + 24 + 9 + 7 + 3 + 2 + 7 + 24 + 1 + 35 + 0 + 6 + 14 + 9 + 3 = 957, \end{aligned}$$

se tiene que $d = 7$.



- b) $ak \equiv bk \pmod{m}$ para todo $k \in \mathbb{Z}$ (el recíproco no es cierto: $8 \equiv 2 \pmod{6}$ pero $4 \not\equiv 1 \pmod{6}$).
- c) $a^r \equiv b^r \pmod{m}$ para todo $r \in \mathbb{N}$. \square

Ejercicio 1.6.5 Si $a = a_1d, b = b_1d$ y $m = m_1d$ demostrar que

$$a \equiv b \pmod{m} \text{ si y sólo si } a_1 \equiv b_1 \pmod{m_1}. \quad \square$$

Definición 1.6.6 Llamaremos *congruencia lineal* a una ecuación de la forma

$$ax \equiv b \pmod{m}$$

donde $a, b \in \mathbb{Z}$, con $a \neq 0$, y $m \in \mathbb{N}$ son dados. \square

El primer problema que se plantea es si las congruencias tienen solución, es decir, si existe x que satisfaga que ax y b son congruentes módulo m , y si esta solución es única o puede haber varias. Para dilucidar el problema del número de soluciones vamos primero a estudiar dónde se debe buscar.

Proposición 1.6.7 Sea $ax \equiv b \pmod{m}$ una congruencia lineal. Si α es una solución de la misma y $\beta \equiv \alpha \pmod{m}$ entonces β es también una solución de la congruencia.

DEMOSTRACIÓN. Basta tener en cuenta que

$$\beta \equiv \alpha \pmod{m} \Rightarrow a\beta \equiv a\alpha \pmod{m}.$$

Como $a\alpha \equiv b \pmod{m}$, la propiedad transitiva asegura que $a\beta \equiv b \pmod{m}$, es decir, que β es solución de la congruencia. \square

Observación 1.6.8 Este resultado nos indica que basta con que encontremos las soluciones de la congruencia que estén en \mathbb{Z}_m . \square

Ejemplo 1.6.9 La congruencia $2x \equiv 1 \pmod{6}$ no tiene solución en \mathbb{Z}_6 , mientras que la congruencia $2x \equiv 2 \pmod{8}$ tiene dos soluciones en \mathbb{Z}_8 ($x = 1$ y $x = 5$). \square

El siguiente resultado caracteriza cuándo una congruencia lineal tiene solución:

Teorema 1.6.10 Sean $a, b \in \mathbb{Z}$, $m \in \mathbb{N}$ y $d = \text{mcd}(a, m)$. La congruencia lineal $ax \equiv b \pmod{m}$ tiene solución si, y sólo si, d divide a b ; en ese caso, el número de soluciones en \mathbb{Z}_m es d .

DEMOSTRACIÓN. Primero se ve cuándo hay solución.

\Leftarrow Supongamos que d divide a b , digamos $b = dk$ para algún $k \in \mathbb{Z}$. Como $d = \text{mcd}(a, m)$, existen enteros u y v tales que

$$d = au + mv,$$

luego $b = auk + mvk$. Por lo tanto $b \equiv auk \pmod{m}$, así que uk es solución de la congruencia.

\Rightarrow Recíprocamente, si existe una solución α , existe algún $k \in \mathbb{Z}$ tal que

$$a\alpha - b = km,$$

luego $b = a\alpha - km$ y, como d divide a a y a m , también divide a b .

Veamos ahora el número de soluciones.

- a) En primer lugar supongamos que $d = 1$ y que hubiera dos soluciones, α y β en \mathbb{Z}_m . Entonces, tanto $a\alpha$ como $a\beta$ son congruentes con b módulo m , luego $a(\alpha - \beta)$ es múltiplo de m . Como a es primo con m , resulta que $\alpha - \beta$ es múltiplo de m y, al estar ambos en \mathbb{Z}_m , resulta que $\alpha - \beta = 0$, luego ambas soluciones son iguales.



b) Finalmente, si $d > 1$, sean

$$a = a_1 d, b = b_1 d \text{ y } m = m_1 d.$$

Como $\text{mcd}(a_1, m_1) = 1$, la congruencia $a_1 x \equiv b_1 \pmod{m_1}$ tiene solución única en \mathbb{Z}_{m_1} , digamos α . Como sabemos (véase el Ejercicio 1.6.5) que $ax \equiv b \pmod{m}$ si, y sólo si, $a_1 x \equiv b_1 \pmod{m_1}$, resulta que las soluciones de la congruencia original en \mathbb{Z}_m son

$$\alpha, \alpha + m_1, \alpha + 2m_1, \dots, \alpha + (d-1)m_1;$$

esto es, un total de d soluciones en \mathbb{Z}_m . \square

Una vez que hemos resuelto el problema de determinar cuándo hay solución, dónde hay que buscarla, y cuántas hay, veamos un ejemplo para ver cómo se obtienen.

Ejemplo 1.6.11 Consideremos la congruencia

$$56x \equiv 42 \pmod{105}. \quad (1.3)$$

Lo primero es ver si (1.3) tiene solución y, para ello, se calcula el máximo común divisor de 56 y 105. A partir de la tabla

i	r_i	q_i	x_i	y_i
0	105		1	0
1	56	1	0	1
2	49	1	1	-1
3	7	7	-1	2
	0			

se tiene que

$$\text{mcd}(56, 105) = 7$$

y la identidad de Bézout

$$7 = 105 \times (-1) + 56 \times 2.$$

Como 7 divide a 42, la congruencia (1.3) tiene solución, y hay un total de siete (el mcd) en \mathbb{Z}_{105} . Para calcularlas, seguimos el procedimiento aprendido al demostrar la existencia. Trabajamos con la congruencia auxiliar, de solución única, resultante de dividir todo por 7

$$8x \equiv 6 \pmod{15}, \quad (1.4)$$

que tiene asociada la identidad de Bézout

$$1 = 15 \times (-1) + 8 \times 2.$$

Multiplicando por $b_1 = 6$ la igualdad anterior, obtenemos que

$$6 = 6 \times 15 \times (-1) + 6 \times 8 \times 2 \equiv 8 \times (6 \times 2) = 8 \times 12 \pmod{15}$$

y, por tanto, $x = 12$ es solución de (1.4) (nos ha salido un número perteneciente a \mathbb{Z}_{15} ; si no fuera así tomaríamos su resto módulo 15). Y ahora que tenemos la solución única de la congruencia auxiliar (1.4), recordando de nuevo cómo demostramos el número de soluciones del problema original, ya sabemos que las siete soluciones de (1.3) son

$$12, 12+15 = 27, 12+2 \times 15 = 42, 12+3 \times 15 = 57, 12+4 \times 15 = 72, 12+5 \times 15 = 87, 12+6 \times 15 = 102.$$

Como nunca está de más comprobar que no nos hemos equivocado veamos, por ejemplo, que 102 es solución de (1.3). En efecto, puesto que

$$56 \times 102 = 5712 \text{ y } 5712 - 42 = 5670 = 54 \times 105,$$

se tiene que $56 \times 102 \equiv 42 \pmod{105}$. \square



Observación 1.6.12 En el caso particular de que $b = 1$, el Teorema 1.6.10 asegura que la congruencia lineal $ax \equiv 1 \pmod{m}$ tiene solución si, y sólo si, $\text{mcd}(a, m)$ divide a 1, es decir, si a y m son primos entre sí (en cuyo caso, la solución en \mathbb{Z}_m es única). Esto permite definir, cuando $\text{mcd}(a, m) = 1$, el *inverso* de a en \mathbb{Z}_m (que se suele denotar a^{-1}) como la solución de la congruencia $ax \equiv 1 \pmod{m}$ en \mathbb{Z}_m .

En la práctica, para encontrar a^{-1} basta con encontrar una identidad de Bézout de la forma

$$1 = ax + my$$

y tomar como a^{-1} el representante de x en \mathbb{Z}_m (es decir, el resto de dividir x entre m). \square

Ejemplo 1.6.13 Veamos que $7^{-1} = 16$ en \mathbb{Z}_{37} . En efecto, a partir de la tabla

i	r_i	q_i	x_i	y_i
0	37		1	0
1	7	5	0	1
2	2	3	1	-5
3	1	2	-3	16
	0			

se tiene que $1 = 37 \times (-3) + 7 \times 16$, de donde se deduce el resultado. \square

Una vez estudiadas las congruencias lineales, se generaliza la cuestión a sistemas de congruencias. Consideraremos para empezar dos congruencias, módulo respectivamente m y n . Estudiemos, en primer lugar, la relación entre las clases de congruencia módulo estos dos números.

Observación 1.6.14 Si $x \in \mathbb{Z}$ y m y n son dos números naturales, siempre se tiene la inclusión

$$[x]_{mn} \subset [x]_m \cap [x]_n,$$

puesto que si $y \in [x]_{mn}$, existe λ tal que $y = x + \lambda mn$, por lo que y es congruente con x , módulo m y también módulo n . Sin embargo, el otro contenido, en general, no se da. Basta tomar, por ejemplo, $x = 1, m = 2$ y $n = 6$; en este caso se tiene

$$7 \equiv 1 \pmod{2}, 7 \equiv 1 \pmod{6} \text{ pero } 7 \not\equiv 1 \pmod{12}. \quad \square$$

El siguiente resultado muestra que la igualdad se da cuando los módulos son primos entre sí.

Proposición 1.6.15 Si m y n son dos números naturales primos entre sí, se verifica que

$$[x]_m \cap [x]_n = [x]_{mn}.$$

DEMOSTRACIÓN. De acuerdo con la Observación 1.6.14, basta probar que

$$[x]_m \cap [x]_n \subset [x]_{mn}.$$

Si $y \in [x]_m \cap [x]_n$, existen λ y μ tales que

$$y = x + \lambda m = x + \mu n.$$

Por lo tanto, $\lambda m = \mu n$. Como $\text{mcd}(m, n) = 1$, se tiene que n divide a λ , digamos $\lambda = \lambda_1 n$. Con ello,

$$y = x + \lambda m = x + \lambda_1 nm,$$

y así $y \in [x]_{mn}$. \square

Estamos ya en condiciones de establecer una condición suficiente para que un sistema de dos congruencias tenga solución y estudiar su unicidad.



Teorema 1.6.16 Sean m_1 y m_2 dos naturales primos entre sí. Si cada una de las congruencias lineales

$$\begin{cases} a_1x \equiv b_1 \pmod{m_1} \\ a_2x \equiv b_2 \pmod{m_2} \end{cases}$$

tiene solución, entonces existe solución común a ambas congruencias en $\mathbb{Z}_{m_1m_2}$. Si, además,

$$\text{mcd}(a_1, m_1) = 1 \text{ y } \text{mcd}(a_2, m_2) = 1$$

entonces dicha solución es única en $\mathbb{Z}_{m_1m_2}$.

DEMOSTRACIÓN. Sean α_1 y α_2 soluciones respectivas de ambas congruencias. Como $\text{mcd}(m_1, m_2) = 1$, la identidad de Bézout para m_1 y m_2 nos dice que existen enteros u y v tales que

$$1 = m_1u + m_2v.$$

Pues bien, el número

$$\alpha = \alpha_2m_1u + \alpha_1m_2v$$

es solución común a ambas congruencias. En efecto, por un lado se tiene que

$$\alpha = \alpha_2m_1u + \alpha_1m_2v = \alpha_2m_1u + \alpha_1(1 - m_1u) = \alpha_1 + m_1(\alpha_2u - \alpha_1u)$$

y, por otro, que

$$\alpha = \alpha_2(1 - m_2v) + \alpha_1m_2v = \alpha_2 + m_2(\alpha_1v - \alpha_2v).$$

Hemos probado, entonces, que $\alpha \in [\alpha_1]_{m_1}$ y $\alpha \in [\alpha_2]_{m_2}$ por lo que, gracias a la Proposición 1.6.7, es solución de las dos congruencias. Si ahora consideramos β , el representante de α en $\mathbb{Z}_{m_1m_2}$, tenemos que

$$\beta \in [\alpha]_{m_1m_2} \subset [\alpha]_{m_1} \cap [\alpha]_{m_2} = [\alpha_1]_{m_1} \cap [\alpha_2]_{m_2},$$

esto es, que $\beta \in \mathbb{Z}_{m_1m_2}$ es solución del sistema original.

Probemos, a continuación, la unicidad. Supongamos que $\alpha, \beta \in \mathbb{Z}_{m_1m_2}$ son soluciones del sistema de congruencias. Por serlo de la primera, y tener ésta una única solución en \mathbb{Z}_{m_1} , α y β son congruentes módulo m_1 con dicha solución. Por la transitividad,

$$\alpha \equiv \beta \pmod{m_1}.$$

De igual forma se prueba que

$$\alpha \equiv \beta \pmod{m_2}.$$

En conclusión, se llega a

$$\alpha \in [\beta]_{m_1} \cap [\beta]_{m_2} = [\beta]_{m_1m_2},$$

de lo que se infiere que $\alpha = \beta$ por ser ambos números elementos de $\mathbb{Z}_{m_1m_2}$. \square

Ejemplo 1.6.17 Consideremos el sistema de congruencias lineales

$$\begin{cases} 3x \equiv 2 \pmod{8} \\ 2x \equiv 1 \pmod{5} \end{cases}$$

Los módulos 8 y 5 son primos entre sí, y cada congruencia tiene solución única, $\alpha_1 = 6$ y $\alpha_2 = 3$, respectivamente (compruébese). Escribiendo una Identidad de Bézout de 8 y 5, $1 = 2 \times 8 - 3 \times 5$, y multiplicando por las soluciones de forma cruzada como en la demostración anterior, obtenemos la única solución común en \mathbb{Z}_{40}

$$\alpha = 3 \times 2 \times 8 - 6 \times 3 \times 5 = 48 - 90 = -42 \equiv 38 \pmod{40}. \quad \square$$

El resultado anterior se generaliza a sistemas de una cantidad arbitraria de congruencias:



Teorema 1.6.18 (Chino del resto) Sean m_1, m_2, \dots, m_k números naturales primos entre sí dos a dos. Si cada una de las congruencias lineales

$$a_i x \equiv b_i \pmod{m_i}, \quad i = 1, 2, \dots, k$$

tiene solución, entonces existe solución común a todas ellas en $\mathbb{Z}_{m_1 m_2 \dots m_k}$. Si, además,

$$\text{mcd}(a_i, m_i) = 1, \quad i = 1, 2, \dots, k$$

entonces dicha solución es única en $\mathbb{Z}_{m_1 m_2 \dots m_k}$.

DEMOSTRACIÓN. Probamos la existencia de solución común por inducción sobre el número k de congruencias.

- a) Si $k = 2$, es el Teorema 1.6.16.
- b) Sea $k > 2$ y supongamos cierto el resultado para $k - 1$ congruencias. Entonces las $k - 1$ primeras congruencias del sistema tienen una solución común, digamos α . También cualquier solución de la congruencia

$$x \equiv \alpha \pmod{m_1 m_2 \dots m_{k-1}}$$

es solución de las $k - 1$ primeras congruencias. Basta considerar ahora el sistema de dos congruencias

$$\begin{cases} x \equiv \alpha \pmod{m_1 m_2 \dots m_{k-1}} \\ a_k x \equiv b_k \pmod{m_k} \end{cases} \quad (1.5)$$

Como $m_1 m_2 \dots m_{k-1}$ y m_k son primos entre sí, este sistema de dos congruencias tiene solución común, que lo es de todo el sistema.

Finalmente, si la solución común α de las $k - 1$ primeras congruencias del sistema es única, como la de la k -ésima también lo es, aplicando el mismo razonamiento que en el Teorema 1.6.16 al sistema (1.5) probamos la unicidad de solución común a todo el sistema. \square

Observación 1.6.19 Si los módulos de las congruencias no son primos entre sí dos a dos, puede o no haber solución. Así, por ejemplo, el sistema de congruencias

$$\begin{cases} x \equiv 2 \pmod{15} \\ x \equiv 1 \pmod{6} \end{cases}$$

no tiene solución, mientras que

$$\begin{cases} x \equiv 4 \pmod{15} \\ x \equiv 1 \pmod{6} \end{cases}$$

tiene la solución $x = 19$ en \mathbb{Z}_{30} . De hecho, se puede demostrar que una condición necesaria y suficiente para que el sistema de congruencias tenga solución es que las soluciones α_i de cada congruencia sean, a su vez, congruentes dos a dos módulo el máximo común divisor de los respectivos módulos, es decir,

$$\alpha_i \equiv \alpha_j \pmod{\text{mcd}(m_i, m_j)}. \quad \square$$

Ejemplo 1.6.20 El sistema de congruencias lineales

$$\begin{cases} 3x \equiv 2 \pmod{8} \\ 2x \equiv 1 \pmod{5} \\ 5x \equiv 2 \pmod{3} \end{cases}$$

tiene una única solución en \mathbb{Z}_{120} , puesto que los módulos 8, 5 y 3 son primos entre sí dos a dos, y cada congruencia tiene solución única, $\alpha_1 = 6$, $\alpha_2 = 3$ y $\alpha_3 = 1$, respectivamente (compruébese). En el



Ejemplo 1.6.17 hemos encontrado la solución común a las dos primeras en \mathbb{Z}_{40} , que es 38. Siguiendo lo aprendido en la demostración del teorema Chino del resto, consideramos el sistema de dos congruencias

$$\begin{cases} x \equiv 38 \pmod{40} \\ 5x \equiv 2 \pmod{3} \end{cases}$$

Una identidad de Bézout para 40 y 3 es $1 = 1 \times 40 - 13 \times 3$; multiplicando por las soluciones de forma cruzada obtenemos la única solución común en \mathbb{Z}_{120}

$$\alpha = 1 \times 1 \times 40 - 38 \times 13 \times 3 = 40 - 1482 = -2 \equiv 118 \pmod{120}. \quad \square$$

El teorema de Euler es la base en la que se asienta el procedimiento de cifrado exponencial y la clave RSA. Para llegar a él, comenzamos definiendo la función de Euler.

Definición 1.6.21 Sea $n \in \mathbb{N}$. Llamaremos $\varphi(n)$, la *función de Euler del número n* , al cardinal del conjunto

$$\{t \in \mathbb{N} : 1 \leq t \leq n \text{ y } \text{mcd}(t, n) = 1\}. \quad \square$$

Dado un número pequeño es muy fácil calcular su función de Euler. Vamos a obtener un procedimiento de carácter general mediante tres pasos sucesivos, según la estructura del número como producto de primos, usando en el tercero lo que hemos aprendido de congruencias.

Proposición 1.6.22 Si p es primo se verifica que

$$\varphi(p) = p - 1.$$

DEMOSTRACIÓN. Evidente. \square

Proposición 1.6.23 Si p es primo y $k > 1$ entonces

$$\varphi(p^k) = p^k - p^{k-1}.$$

DEMOSTRACIÓN. Los números hasta p^k que tienen factores comunes con p^k son múltiplos de p . Hasta p^k hay exactamente p^{k-1} múltiplos de p y, por lo tanto, se tiene el resultado. \square

Antes de dar el tercer paso, enunciaremos dos lemas auxiliares (se propone, como ejercicio, demostrarlos).

Lema 1.6.24 Sean $\alpha, \beta, a \in \mathbb{N}$. Si $\alpha \equiv \beta \pmod{a}$ y β es primo con a entonces α es primo con a .

Lema 1.6.25 Sean $a, b \in \mathbb{N}$. Un número $\gamma \in \mathbb{N}$ es primo con a y con b si, y sólo si, es primo con ab .

Teorema 1.6.26 Si a y b son dos números naturales primos entre sí, se verifica que

$$\varphi(ab) = \varphi(a)\varphi(b).$$

DEMOSTRACIÓN. Sean

$$A = \{\alpha_1, \alpha_2, \dots, \alpha_{\varphi(a)}\}, B = \{\beta_1, \beta_2, \dots, \beta_{\varphi(b)}\} \text{ y } C = \{\gamma_1, \gamma_2, \dots, \gamma_{\varphi(ab)}\}$$

respectivamente, los conjuntos de números primos con a , b y ab en \mathbb{Z}_a , \mathbb{Z}_b y \mathbb{Z}_{ab} . Vamos a construir una biyección de $A \times B$ en C . Para ello, dado $(\alpha_i, \beta_j) \in A \times B$, definimos el sistema de congruencias lineales

$$\begin{cases} x \equiv \alpha_i \pmod{a} \\ x \equiv \beta_j \pmod{b}. \end{cases}$$

Según sabemos, tiene solución única, digamos γ , en \mathbb{Z}_{ab} . Como $\gamma \equiv \alpha_i \pmod{a}$ y α_i es primo con a , el Lema 1.6.24 nos asegura que γ es primo con a . Del mismo modo, γ es primo con b y, por el Lema 1.6.25, lo es con ab . Es decir, $\gamma \in C$. Definimos $f : A \times B \rightarrow C$ tomando γ como la imagen de (α_i, β_j) .



Para ver la inyectividad, si

$$\gamma = f(\alpha_i, \beta_j) = f(\alpha_p, \beta_q),$$

se verifica que

$$\alpha_i, \alpha_p \equiv \gamma \pmod{a}$$

y, como ambos pertenecen a \mathbb{Z}_a , son iguales. Del mismo modo, $\beta_j = \beta_q$, por lo que f es inyectiva.

Finalmente, para ver que f es suprayectiva, sea $\gamma \in C$. Como γ es primo con ab , el Lema 1.6.25 indica que también lo es con a y con b . La congruencia $x \equiv \gamma \pmod{a}$ tiene solución, digamos α , en \mathbb{Z}_a . Como γ es primo con a , también lo es (gracias, de nuevo, al Lema 1.6.24) α que, por lo tanto, pertenece a A . Del mismo modo tenemos $\beta \in B$ con $\beta \equiv \gamma \pmod{b}$. Por la definición de f , se tiene que

$$f(\alpha, \beta) = \gamma$$

y, en consecuencia, f es suprayectiva. \square

Observación 1.6.27 Así pues, ya tenemos una forma de obtener la función de Euler de cualquier número. Basta expresarlo como producto de potencias de primos, y su función de Euler es el producto de las respectivas expresiones $(p^k - p^{k-1})$. \square

Ejemplo 1.6.28 Puesto que $420 = 2^2 \times 3 \times 5 \times 7$ se tiene que

$$\varphi(420) = (2^2 - 2^1)(3^1 - 3^0)(5^1 - 5^0)(7^1 - 7^0) = 2 \times 2 \times 4 \times 6 = 96. \quad \square$$

En estos términos, podemos ahora ver dos teoremas importantes en teoría de números, a saber:

Teorema 1.6.29 (Euler) Si a y n son dos números naturales primos entre sí, entonces

$$a^{\varphi(n)} \equiv 1 \pmod{n}.$$

DEMOSTRACIÓN. En primer lugar, vamos a demostrar la siguiente afirmación: considerando los conjuntos $A = \{v_1, v_2, \dots, v_{\varphi(n)}\}$ de números menores que n y primos con n , y $B = \{av_1, av_2, \dots, av_{\varphi(n)}\}$, entonces cada uno de los elementos de B es congruente módulo n con uno de los de A , y viceversa. En efecto, para cada $v_i \in A$, como a y n son primos entre sí, la congruencia

$$ax \equiv v_i \pmod{n}$$

tiene una única solución α en \mathbb{Z}_n que, razonando como en la demostración del Teorema 1.6.26, es primo con n . Así pues, $\alpha \in A$ y $a\alpha$ es un elemento de B congruente con v_i módulo n . Supongamos ahora que existe $\alpha \in A$ solución de las congruencias asociadas a v_i y v_j , es decir,

$$\begin{cases} a\alpha \equiv v_i \pmod{n} \\ a\alpha \equiv v_j \pmod{n}. \end{cases}$$

Obviamente, entonces, $v_i \equiv v_j \pmod{n}$, y como ambos son elementos de \mathbb{Z}_n , son iguales. Hemos probado, por tanto, que cada elemento de A es congruente con uno de B y que dos elementos distintos de A no pueden ser congruentes con el mismo elemento de B . Como ambos conjuntos tienen el mismo número de elementos, se llega a la afirmación previamente enunciada.

Ahora, llamando u y v , respectivamente, a los productos de los elementos de A y de B , se tiene que $v \equiv u \pmod{n}$. Pero

$$v = a^{\varphi(n)}u,$$

por lo que

$$a^{\varphi(n)}u \equiv u \pmod{n}$$

y, al ser u y n primos entre sí, existe el inverso de u en \mathbb{Z}_n y, por tanto,

$$a^{\varphi(n)} \equiv 1 \pmod{n}. \quad \square$$

Y ahora, como caso particular de este resultado, obtenemos:

Teorema 1.6.30 (Pequeño teorema de Fermat) Si p es primo y a no es múltiplo de p , entonces

$$a^{p-1} \equiv 1 \pmod{p}.$$

DEMOSTRACIÓN. Basta aplicar el Teorema 1.6.29 tomando $n = p$ y recordando que $\varphi(p) = p - 1$. \square



1.7 Aplicación: Criptografía

Un problema secular que ha adquirido nuevo aspecto modernamente es el del cifrado de mensajes. Consiste en transformar un mensaje claro en otro cifrado (*cifrar*) y en reconvertir el mensaje cifrado en mensaje claro (*descifrar*). Los procedimientos matemáticos más sencillos para ello han consistido, fundamentalmente, en añadir a los caracteres de los números (0, 1, 2, 3, 4, 5, 6, 7, 8, 9) la conversión de las letras del alfabeto a números mediante la tabla

A	B	C	D	E	F	G	H	I	J	K	L	M
10	11	12	13	14	15	16	17	18	19	20	21	22
N	O	P	Q	R	S	T	U	V	W	X	Y	Z
23	24	25	26	27	28	29	30	31	32	33	34	35

y asignando el número 36 al espacio en blanco; seguidamente, se aplica una regla que convierta estos números en otros de modo biunívoco, de forma que después se puedan descifrar y reconvertir a su valor original. Suelen considerarse varios tipos de procedimientos:

1.7.1 Procedimiento aditivo

La clave del cifrado es un número, digamos b , que se suma a cada uno de los del mensaje. Para obtener números que estén entre 0 y 36, se obtiene el resto módulo 37 de la suma del número original más b . El receptor del mensaje debe conocer b , y no tiene más que restar b módulo 37 o, lo que es lo mismo, sumar $37 - b$ para descifrar el mensaje.

Ejemplo 1.7.1 Veamos cómo cifrar el mensaje “EL 1 A LAS 3” con la clave aditiva $b = 17$. En primer lugar, “traducimos” cada carácter a su valor numérico, obteniendo

$$14 \ 21 \ 36 \ 1 \ 36 \ 10 \ 36 \ 21 \ 10 \ 28 \ 36 \ 3. \quad (1.6)$$

Su codificación es

$$31 \ 1 \ 16 \ 18 \ 16 \ 27 \ 16 \ 1 \ 27 \ 8 \ 16 \ 20, \quad (1.7)$$

lo que se traduce en “V1G1GRG1R8GK”. El receptor deberá sumar $37 - 17 = 20$ (módulo 37) en la cadena numérica (1.7) para obtener la cadena (1.6), que conduce al mensaje original. \square

1.7.2 Procedimiento multiplicativo

En lugar de sumar, se multiplica por un factor, digamos a ; es decir, el cifrado de la conversión numérica m de cada carácter será $ma \pmod{37}$. Para que el cifrado sea biyectivo es necesario que todas las congruencias lineales

$$ax \equiv y \pmod{37}$$

tengan una única solución para cualquier $y \in \mathbb{Z}_{37}$. Gracias al Teorema 1.6.10 podemos asegurar que esto ocurre si, y sólo si,

$$\text{mcd}(a, 37) = 1.$$

Para descifrar el mensaje se considera a^{-1} , inverso de a en \mathbb{Z}_{37} (véase la Observación 1.6.12). De esta forma, puesto que

$$aa^{-1} \equiv 1 \pmod{37},$$

obviamente se tiene que

$$(ma)a^{-1} \equiv m \pmod{37},$$

por lo que el número buscado m es el resto de dividir $(ma)a^{-1}$ entre 37.

Ejemplo 1.7.2 Veamos cómo cifrar el mensaje anterior “EL 1 A LAS 3” con la clave multiplicativa $a = 7$. Su traducción es (1.6) y su codificación viene dada por

$$24 \ 36 \ 30 \ 7 \ 30 \ 33 \ 30 \ 36 \ 33 \ 11 \ 30 \ 21, \quad (1.8)$$



lo que se traduce en “O U7UXU XBUL”. El receptor debe multiplicar en la cadena numérica (1.8) por el inverso de 7 en \mathbb{Z}_{37} que, como se vio en el Ejemplo 1.6.13, es 16, obteniendo

$$384\ 576\ 480\ 112\ 480\ 528\ 480\ 576\ 528\ 176\ 480\ 336$$

cuyos restos, módulo 37, constituyen la cadena (1.6) (compruébese). \square

1.7.3 Procedimiento exponencial

En este procedimiento, cuando se quiera cifrar un número m , se elevará éste a un exponente k y el resultado se reducirá módulo n , siendo k y n números naturales elegidos convenientemente. Para describir la manera de obtener de una forma eficaz el número cifrado, presentamos primero el algoritmo de exponenciación rápida, el cual minimiza el número de operaciones utilizando la idea de representar el exponente en base 2. Así, por ejemplo, puesto que

$$23 = 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0,$$

para todo $a \in \mathbb{N}$ se tiene que

$$a^{23} = \left(\left((a^2)^2 \right)^2 \right)^2 (a^2)^2 a^2 a,$$

lo cual conlleva sólomente 7 multiplicaciones.

Proposición 1.7.3 (Algoritmo de exponenciación rápida) Sean $a, c \in \mathbb{N}$. Si

$$c = 2^s + d_{s-1}2^{s-1} + \dots + d_22^2 + d_12 + d_0$$

con $d_i \in \{0, 1\}$ para $i = 0, 1, \dots, s - 1$, es el desarrollo en base 2 de c , y se consideran las sucesiones definidas mediante $z_0 = 1, b_0 = a$ y

$$\begin{cases} z_i = z_{i-1}(b_{i-1})^{d_{i-1}} \\ b_i = (b_{i-1})^2 \end{cases}$$

para $i = 1, 2, \dots, s$, entonces

$$a^c = z_s b_s.$$

DEMOSTRACIÓN. Basta tener en cuenta que

$$a^c = a^{2^s} (a^{2^{s-1}})^{d_{s-1}} \dots (a^{2^2})^{d_2} (a^2)^{d_1} a^{d_0}. \quad \square$$

Ejemplo 1.7.4 Para calcular 3^{23} construimos la tabla

i		d_i	z_i	b_i
0	23	1	1	3
1	11	1	3	9
2	5	1	27	81
3	2	0	2187	6561
4	1	1	2187	43046721
	0		94143178827	

y obtenemos que

$$3^{23} = 94143178827. \quad \square$$

Observación 1.7.5 Este algoritmo es particularmente sencillo de usar cuando, como ocurre en este procedimiento de encriptación, es necesario calcular el resto módulo n de la potencia de un número. En tal caso,



todas las operaciones se realizarán módulo n (lo denominaremos *algoritmo de exponenciación modular rápida*). Así, por ejemplo, para calcular el valor de 3^{23} módulo 19 construiremos la tabla

i		d_i	z_i	b_i
0	23	1	1	3
1	11	1	3	9
2	5	1	8	5
3	2	0	2	6
4	1	1	2	17
	0		15	

obteniendo

$$3^{23} \equiv 15 \pmod{19}. \quad \square$$

La siguiente proposición será de ayuda en el proceso de descifrado correspondiente.

Proposición 1.7.6 Sean $n, t \in \mathbb{N}$ tales que

$$t \equiv 1 \pmod{\varphi(n)}.$$

Dado $m \in \mathbb{N}$, si se verifica alguna de las siguientes hipótesis:

- m y n son primos entre sí.
- n es primo.
- n es producto de dos primos distintos.

entonces

$$m^t \equiv m \pmod{n}.$$

DEMOSTRACIÓN.

- a) Si $\text{mcd}(m, n) = 1$, usando el Teorema 1.6.29 (de Euler) se tiene que

$$m^t = m^{\varphi(n)\lambda+1} = m(m^{\varphi(n)})^\lambda \equiv m 1^\lambda \equiv m \pmod{n}.$$

- b) Sea n primo: si m es primo con n ya lo hemos demostrado en el apartado a); si no, es que m es múltiplo de n , luego

$$m \equiv 0 \pmod{n} \Rightarrow m^t \equiv 0 \equiv m \pmod{n}.$$

- c) Si $n = pq$ siendo p y q dos números primos distintos, por el Teorema 1.6.26, se tiene que

$$t \equiv 1 \pmod{\varphi(n)} \equiv 1 \pmod{\varphi(pq)} \equiv 1 \pmod{\varphi(p)\varphi(q)}.$$

Luego $t \equiv 1 \pmod{\varphi(p)}$ y $t \equiv 1 \pmod{\varphi(q)}$ por lo que, gracias al apartado b),

$$m^t \equiv m \pmod{p} \text{ y } m^t \equiv m \pmod{q}.$$

Como p y q son primos distintos, esto implica (véase la Proposición 1.6.15) que

$$m^t \equiv m \pmod{pq} \equiv m \pmod{n}. \quad \square$$

Observación 1.7.7 Este resultado se aplica para encriptar como sigue: se elige n primo o producto de dos primos distintos, se calcula $\varphi(n)$, se toma k primo con $\varphi(n)$ y se calcula j , el inverso de k en $\mathbb{Z}_{\varphi(n)}$, es decir, de forma que

$$kj \equiv 1 \pmod{\varphi(n)}.$$



Suponiendo que el emisor y el receptor de un mensaje (o, más precisamente, de su conversión numérica m) conocen los números k y j , el emisor mandará el mensaje codificado que es r , el resto de la división de m^k entre n . Para descodificar, teniendo en cuenta que

$$m^{kj} \equiv m \pmod{n},$$

el receptor calculará

$$r^j \equiv (m^k)^j \equiv m^{kj} \equiv m \pmod{n},$$

y así recuperará (la conversión numérica de) el mensaje original m .

Obviamente, este procedimiento sólo es válido cuando el número m es menor que n . Por ello, en la práctica, cuando el mensaje es largo, se agrupan sus caracteres (escribiendo las cifras como 00, 01, 02, . . . , 09 para que todos los caracteres estén representados por un número de dos cifras) en *paquetes* m_i de la misma longitud ℓ de forma que el número 36^ℓ . 36 sea menor que n (por ejemplo, para $n = 10^8$, podemos tomar $\ell = 4$, puesto que $36^4 < n$). En el caso de que el número de caracteres del mensaje no sea múltiplo de ℓ , el último paquete se completa con espacios en blanco. □

Ejemplo 1.7.8 Para ilustrar este método, consideramos un caso sencillo con n pequeño y un mensaje corto. Tomamos $n = 5 \times 11 = 55$, lo que nos obliga a tomar paquetes que contienen un único carácter. Veamos cómo cifrar el mensaje “ADN”. Puesto que

$$\varphi(55) = 4 \times 10 = 40,$$

tomamos, por ejemplo, $k = 3$ (que es primo con 40). A partir de la tabla

i	r_i	q_i	x_i	y_i
0	40		1	0
1	3	13	0	1
2	1	3	1	-13
	0			

deducimos que el inverso de $k = 3$ en \mathbb{Z}_{40} es

$$j = -13 \equiv 27 \pmod{40}.$$

Traduciendo cada carácter a su valor numérico, el mensaje se convierte en 101323. Como nuestros paquetes se corresponden con un único carácter, debemos codificar 3 números: $m_1 = 10$, $m_2 = 13$ y $m_3 = 23$. Para ello, empleando el *algoritmo de exponenciación modular rápida* calculamos m_i^3 a partir de la siguiente tabla

i	d_i	$z_i^{(1)}$	$b_i^{(1)}$	$z_i^{(2)}$	$b_i^{(2)}$	$z_i^{(3)}$	$b_i^{(3)}$
0	3	1	1	10	1	13	23
1	1	1	10	45	13	4	23
	0		10		52		12

Los tres paquetes de mensaje codificados que se envían son 10, 52 y 12.

Para descifrar los tres paquetes recibidos, el receptor debe calcular el resto, módulo $n = 55$, de los números recibidos (10, 52 y 12) elevados a $j = 27$. Empleando de nuevo el algoritmo de exponenciación modular rápida tenemos

i	d_i	$z_i^{(1)}$	$b_i^{(1)}$	$z_i^{(2)}$	$b_i^{(2)}$	$z_i^{(3)}$	$b_i^{(3)}$
0	27	1	1	10	1	52	12
1	13	1	10	45	52	9	12
2	6	0	10	45	28	26	1
3	3	1	10	45	28	16	1
4	1	1	10	45	8	36	1
	0		10		13		23

Por tanto, el descifrado de los tres paquetes es, respectivamente, 10, 13 y 23 (lo cual se corresponde con el mensaje inicial “ADN”). □



1.7.4 El procedimiento RSA

Todos los procedimientos anteriores se basan en sistemas de *clave privada*, en que los datos para el cifrado (y el descifrado) han de mantenerse secretos (conocidos únicamente por el emisor y el receptor) si no se quiere que cualquiera pueda interpretar el código. Además, los procedimientos aditivo y multiplicativo, así como el exponencial con n “pequeño”, son susceptibles de descifrado por mero empleo de prueba y error, aunque esto pueda requerir una cantidad de tiempo progresivamente mayor.

El salto cualitativo siguiente son los sistemas de *clave pública / clave privada*, y el primero ha sido el sistema *RSA*, así llamado por los nombres de sus autores, Rivest, Shamir y Adelman. Éste es un sistema exponencial en el cual n es producto de dos primos “grandes” distintos p y q . Aunque se hagan públicos el exponente k y el módulo n , si no se sabe factorizar n , no se puede calcular $\varphi(n) = (p-1)(q-1)$ y, por lo tanto, no se puede encontrar el número j necesario para descifrar el código. Por ello, la idea fundamental es utilizar un número n que no sea factorizable en un tiempo “razonable”, lo cual exige actualmente usar dos factores de unas 150 cifras y, por lo tanto, n de unas 300.

El sistema de envío de mensajes encriptados funciona de la siguiente forma: Si A quiere mandar un mensaje a B debe consultar la clave pública (n, k) de B . Con ella A cifra el mensaje (elevándolo a k) y se lo envía a B ; lo único que debe hacer B es descifrarlo mediante su clave privada j (elevando a j el número recibido).

Este sistema sirve también para firmar los mensajes y comprobar su integridad; en concreto, sirve para que el receptor B esté seguro de que quien envía el mensaje es, efectivamente, el supuesto emisor A , y de que el mensaje no ha sido manipulado por un tercero. Para conseguir esto, basta que A envíe un mensaje en el que, junto con el texto cifrado mediante la clave pública de B , vaya un *resumen* del texto que haya sido cifrado mediante la clave privada de A . Este resumen se obtiene mediante la aplicación de un tipo específico de funciones que se conocen con el nombre de funciones *hash* y cuyas características esenciales son: que no se puede recuperar el texto original a partir del resumen; y que pequeños cambios en el texto alteran fuertemente el resumen, de modo que la probabilidad de que dos textos distintos generen el mismo resumen es ínfima. Cuando B recibe el mensaje, descifra el primer texto mediante su clave privada, y el segundo con la clave pública de A ; el resultado de aplicar la función hash al primer texto debe coincidir con el segundo. El segundo texto enviado se conoce como *firma digital* y esta coincidencia asegura que el mensaje no ha sido manipulado por terceros y que quien envía el mensaje es, efectivamente, A .

Ejemplo 1.7.9 Para ejemplificar el uso del procedimiento RSA consideramos también un valor pequeño de n y el mismo mensaje “ADN” del Ejemplo 1.7.8. Supongamos que el emisor A tiene como clave pública $(n_A, k_A) = (91, 5)$ mientras que la del receptor B es $(n_B, k_B) = (55, 3)$. Según lo visto en el Ejemplo 1.7.8, el texto cifrado que se envía está formado por los tres paquetes 10, 52 y 12. Para obtener la firma digital, supongamos que el resumen del mensaje fuera “N”. Como la clave privada de A es $j_A = 29$ (puesto que $\varphi(91) = 72$ y $5 \times 29 = 145 \equiv 1 \pmod{72}$), el paquete del resumen $m = 23$ debe elevarse a 29 módulo $n_A = 91$, obteniéndose 04 (compruébese). De esta forma, el mensaje que recibe B es

$$10\ 52\ 12\ |04.$$

Para descifrar el primer trozo, B utiliza su clave privada que es, como se vio en el Ejemplo 1.7.8, $j_B = 27$. Elevando los tres paquetes 10, 52 y 12 a 27 módulo $n_B = 55$, obtiene 10, 13 y 23 (compruébese) lo que, obviamente, se corresponde con el mensaje original “ADN”. Al elevar el paquete de la firma digital 04 a la clave pública de A , $k_A = 5$, módulo $n_A = 91$ obtiene “N”, comprobando que es lo que le sale al hacer el resumen del mensaje recibido. Esta coincidencia le confirma que el mensaje ha sido enviado por A y no ha sido manipulado. \square

1.8 Problemas

1.8.1 Determinar los números enteros que, al dividirlos por un número $a \in \mathbb{N}$, dan un cociente igual a su resto.



1.8.2 Determinar el menor entero positivo que dé restos

$$1, 2, 3, 4, 5, 6, 7, 8, 9,$$

respectivamente, al dividirlo por

$$2, 3, 4, 5, 6, 7, 8, 9, 10.$$

1.8.3 Para cada entero positivo $n > 1$, encontrar n números compuestos consecutivos.

1.8.4 Probar que el número formado por 300 unos no es un cuadrado perfecto.

1.8.5 Probar que la suma de cuatro números naturales consecutivos no es un cuadrado perfecto.

1.8.6 Encontrar el máximo común divisor y una identidad de Bézout para los números 10672 y 4147.

1.8.7 Encontrar el mínimo común múltiplo de los números anteriores.

1.8.8 El reverso del D.N.I. de un funcionario tiene la forma:

IDESPABC654321a50414974* 701001bM121221cESP JIMENEZ<GOMEZ<<ANTONIO
--

a) ¿Cuál es su número de registro personal?

b) Determinar la letra del N.I.F. y los cuatro dígitos de control de su D.N.I.

1.8.9 Sea \mathfrak{R} la relación definida en \mathbb{N} por:

$$a \mathfrak{R} b \text{ si y sólo si } a^2 \equiv b^2 \pmod{7}.$$

Probar que \mathfrak{R} es una relación de equivalencia y obtener las clases de equivalencia.

1.8.10 Sea $f : \mathbb{Z}_{12} \rightarrow \mathbb{Z}_{12}$ definido por

$$f([x]_{12}) = [x^2]_{12}.$$

¿Es f una aplicación? ¿Es inyectiva? ¿Quién es $\text{Im}(f)$?

1.8.11 Determinar el valor, con los módulos correspondientes, de las siguientes expresiones:

$$12427 \pmod{10}, 12112 \times 347 \pmod{3}, 22^{1327} \pmod{21} \text{ y } 10^{123} \pmod{8}.$$

1.8.12 Hallar el resto de la división del número $n = 1! + 2! + \dots + 99! + 100!$ entre 12.

1.8.13 Demostrar que para todo $k \geq 1$ se verifica que

$$7 \mid (5^{2k} + 3 \times 2^{5k-2}).$$

1.8.14 Usando congruencias, probar las reglas de divisibilidad por 8, por 9 y por 11.

1.8.15 Demostrar que si $3 \mid (a^2 + b^2)$ entonces $3 \mid a$ y $3 \mid b$.

1.8.16 Demostrar que $5^x + 2 = 7^y$ no tiene soluciones en \mathbb{N} distintas de la trivial $x = y = 1$.

1.8.17 Calcular, si existen, inversos de 24 y de 31 en \mathbb{Z}_{70} .

1.8.18 Sea p un número primo y a un entero no nulo. Demostrar que a es su propio inverso módulo p si, y sólo si, $a \equiv 1 \pmod{p}$ o $a \equiv -1 \pmod{p}$.

1.8.19 Sea $a \in \mathbb{Z} \setminus \{0\}$ y p primo. Demostrar que si p no divide a a entonces a^{p-2} es el inverso de a en \mathbb{Z}_p .



1.8.20 Resolver la congruencia $35x \equiv 119 \pmod{139}$.

1.8.21 Encontrar todas las soluciones en \mathbb{Z}_{60} de la congruencia

$$18x \equiv 54 \pmod{60}.$$

1.8.22 Supongamos que ha caído una gota de tinta sobre un papel y no se ve bien una de las cifras del D.N.I. 2765*372 T. ¿Cuál es el valor de la cifra *?

1.8.23 Resolver el sistema de congruencias

$$\begin{cases} 19x \equiv 103 \pmod{990} \\ x \equiv 91 \pmod{251}. \end{cases}$$

1.8.24 Justificar si el sistema de congruencias

$$\begin{cases} 12x \equiv 19 \pmod{31} \\ 9x \equiv 17 \pmod{32} \end{cases}$$

tiene solución y resolverlo en caso afirmativo.

1.8.25

a) Encontrar las soluciones que tiene la congruencia $12x \equiv 15 \pmod{99}$ en \mathbb{Z}_{99} .

b) Hallar todas las soluciones enteras del sistema de congruencias

$$\begin{cases} 12x \equiv 15 \pmod{99} \\ 7x \equiv 24 \pmod{32}. \end{cases}$$

1.8.26

a) Hallar, justificando la respuesta, todas las soluciones de la congruencia

$$8x \equiv 6 \pmod{42}.$$

b) Determinar cuántas soluciones tiene en \mathbb{Z}_{462} el sistema de congruencias

$$\begin{cases} 8x \equiv 6 \pmod{42} \\ 3x \equiv 5 \pmod{11}, \end{cases}$$

justificando la respuesta. Hallar la menor solución entera de dicho sistema que es mayor que 1000.

1.8.27 ¿De cuántas maneras se pueden obtener 430 € en billetes de 20 y de 50 euros?

1.8.28 ¿De cuántas formas pueden distribuirse 355 euros en billetes de 5, 10 y 50 euros de modo que haya el mismo número de billetes de 5€ que de 10€?

1.8.29 Encontrar el menor número natural múltiplo de 20 que tiene resto 4 cuando se divide por 7 y cuando se divide por 9.

1.8.30 Los alumnos de Bachillerato de un Instituto deben abonar unos gastos. Los de primero pagan 16'90€ y, los de segundo, 16'75€. Si en total se recaudan 740'15€, ¿cuántos alumnos de cada curso han efectuado el pago?

1.8.31 Hallar, usando un sistema de dos congruencias, el menor número natural terminado en 3 cuyo doble tiene resto 5 al dividirlo por 7.



1.8.32 Resolver el sistema de congruencias

$$\begin{cases} x \equiv 2 \pmod{3} \\ 2x \equiv 3 \pmod{5} \\ 3x \equiv 7 \pmod{11}. \end{cases}$$

1.8.33 Encontrar el menor número natural múltiplo de 11 que tiene resto 1 cuando se divide por 3, por 5 y por 7.

1.8.34 Dos comerciantes se reparten una partida de teléfonos móviles, de modo que uno se queda con $\frac{2}{7}$ del total y el otro con $\frac{5}{7}$. El primero los revende en lotes de 51 y el segundo en lotes de 31. Cuando terminan, al primero le han sobrado 46 teléfonos y al segundo 19. ¿Cuál es el menor número inicial de teléfonos que hace esto posible?

1.8.35 Con los estudiantes de primer curso pueden formarse equipos de fútbol (11 jugadores) sobrando 4 estudiantes, equipos de balonmano (7 jugadores) sobrando 2 estudiantes y equipos de baloncesto (5 jugadores) sobrando 4 estudiantes.

- Determinar los posibles valores del número de estudiantes de primero, sabiendo que son menos de 500.
- Si se forman equipos de los mismos deportes sólo con los estudiantes matriculados en grupos de tarde se obtienen, exactamente, los mismos restos. ¿Cuántos estudiantes de primero están matriculados en grupos de mañana?

1.8.36 Determinar el número de primos con 21 que son menores que él y explicitarlos.

1.8.37 Determinar el resto de la división de $(102^{73} + 55)^{37}$ entre 111.

1.8.38 Calcular, justificando la respuesta, el resto de dividir:

$$8^{1000001} \text{ entre } 25, 9^{240002} \text{ entre } 35, 12^{1200001} \text{ entre } 21 \text{ y } 14^{240002} \text{ entre } 35.$$

1.8.39 Comprobar que si a y m son primos entre sí entonces $a^{\varphi(m)-1}$ es el inverso de a módulo m .

1.8.40 Aplicando el procedimiento aditivo con $b = 7$, cifrar el mensaje

“LOS ESTUDIANTES DE MATEMATICAS SON MUY ESCOGIDOS”.

Y descifrar, para el mismo procedimiento con $b = 10$, el mensaje

“WS9MVK4O9NOV9MKTO0Y9O19BEJC”.

1.8.41 Aplicando el procedimiento multiplicativo con $a = 11$, cifrar el primer mensaje del Problema 1.8.40. ¿Qué debemos hacer cuando lo recibamos para descifrarlo?

1.8.42 ¿Qué condición debe cumplir un número a para poder ser usado como clave de cifrado con el método multiplicativo? Si $a = 19$, descifrar el mensaje “Y7W5UC”.

1.8.43 Descifrar el mensaje “3FI” que ha sido cifrado usando el procedimiento exponencial con $n = 41$ y $k = 7$.

1.8.44 Queremos enviar el mensaje “PIDE 3” a alguien cuya clave pública en el sistema RSA es $(n, k) = (51, 5)$. ¿Cuál es el mensaje enviado? Si esta persona recibe el mensaje “17453415464534” ¿cuál es su significado?

1.8.45 En una clave pública RSA (n, k) con $n = 143$, ¿de cuántas maneras distintas se pueden elegir claves de encriptado k ?



1.8.46 Decidir si las siguientes parejas de claves de encriptado RSA son válidas. En caso afirmativo, hallar la clave de desencriptado: i) (57, 14) ii) (85, 9) iii) (36, 3).

1.8.47 Descifrar el mensaje “54 10” sabiendo que ha sido cifrado mediante el procedimiento RSA con $n = 85$ y $k = 3$.

1.8.48 Encriptar el mensaje “LOL” mediante el procedimiento RSA con $n = 51$ y $k = 23$.

1.8.49

- a) Encriptar el mensaje “HOY” mediante el procedimiento RSA con $n = 39$ y $k = 5$.
- b) Calcular la clave de desencriptado y describir el proceso que debe seguir el receptor del mensaje para desencriptarlo.

1.8.50 Desencriptar el mensaje “50 47 62” sabiendo que se ha encriptado usando el procedimiento RSA con $n = 65$ y $k = 43$. ¿De cuántas maneras válidas distintas se podrían elegir claves de encriptado k para este valor de n ?





2 Grupos de simetrías. Mosaicos

El objetivo de este capítulo es estudiar y clasificar todas las posibles formas de cubrir el plano con copias idénticas de una “ficha” o “pieza” (*tesela*) de manera periódica; es decir, los mosaicos o teselaciones periódicas (al final del capítulo veremos también algunas teselaciones “aperiódicas”). La herramienta para esta clasificación es el estudio de qué movimientos del plano (aplicaciones biyectivas del plano que respetan la distancia entre puntos, es decir, *isometrías*) preservan la teselación, es decir, el estudio de las *simetrías* de la teselación. A su vez, para este estudio, es fundamental el hecho de que las simetrías de una figura plana tienen una estructura algebraica de *grupo*. Por tanto deberemos estudiar algunas nociones de teoría de grupos, aplicarlas al grupo de los movimientos del plano y sus subgrupos, aplicar a su vez estos resultados al estudio de los retículos, y podremos entonces encontrar los 17 grupos cristalográficos planos, que son todos los posibles grupos de simetrías de un mosaico del plano y, por tanto, clasifican todas las formas posibles de teselar un plano de forma periódica. La consecuencia es una primera mirada sobre la realización de las estructuras algebraicas en la geometría a través del concepto central de simetría.

2.1 Las simetrías rotacionales de un tetraedro

¿Cuántas simetrías rotacionales tiene un tetraedro regular? Es decir, ¿cuántos movimientos rígidos del espacio llevan el tetraedro sobre sí mismo? Contemos los ejes y las rotaciones por cada eje:

- Hay un eje de rotación que pasa por el centro de cada cara y el vértice opuesto (en total, 4 ejes), y las rotaciones de ángulos $\frac{2\pi}{3}$ y $\frac{4\pi}{3}$ dejan invariante el tetraedro.
- Otro eje de rotación es el que une los puntos medios de aristas no concurrentes (otros 3 ejes). En ese caso, hay que girar un ángulo π .
- Si añadimos la identidad, que es trivialmente una simetría, el total es de $8 + 3 + 1 = 12$.

Un prisma hexagonal y una pirámide dodecagonal tienen también 12 simetrías rotacionales. En efecto,

- En el primer caso, las cinco rotaciones cuyo eje es el que une el centro de ambas bases y de ángulo $\frac{k\pi}{3}$, $k = 1, 2, \dots, 5$; las tres rotaciones de ángulo π cuyos ejes unen los puntos medios de las seis caras rectangulares opuestas; las tres rotaciones de ángulo π cuyos ejes unen los puntos medios de las aristas opuestas que no delimitan las bases; y la identidad.
- En el segundo, las 11 rotaciones cuyo eje une el vértice de la pirámide con el punto medio de la base, de ángulos $\frac{k\pi}{6}$, $k = 1, 2, \dots, 11$, y la identidad.

¿Tienen entonces estas tres figuras el mismo comportamiento en lo que a las simetrías rotacionales se refiere? ¿Tienen “la misma simetría”? La respuesta es no. Aunque el número de simetrías coincida, las relaciones entre ellas son diferentes. Por ejemplo, el orden sucesivo en el que se apliquen varias simetrías de la pirámide no influye en el resultado final (que es la rotación cuyo ángulo es la suma total de los ángulos de las rotaciones aplicadas). Sin embargo, esto no se aplica a los otros dos casos (véase la Figura 2.1). Además, para el tetraedro, hay tres rotaciones (no triviales) distintas que, aplicadas dos veces, dejan al tetraedro en la posición de partida (es decir, que al componerse con ellas mismas dan como resultado la identidad): son las que tienen ejes que unen puntos medios de aristas. Sin embargo, para el prisma hexagonal, hay siete de estas rotaciones (todas las de ángulo π) y sólo una en el caso de la pirámide.

2.2 Grupos. Isomorfismos y homomorfismos de grupos

2.2.1 Definición axiomática de grupo

Parece claro, pues, que las tres figuras consideradas en el apartado anterior exhiben un comportamiento diferente en lo que a sus simetrías se refiere. Dicho de otra forma, en el conjunto de las simetrías de una



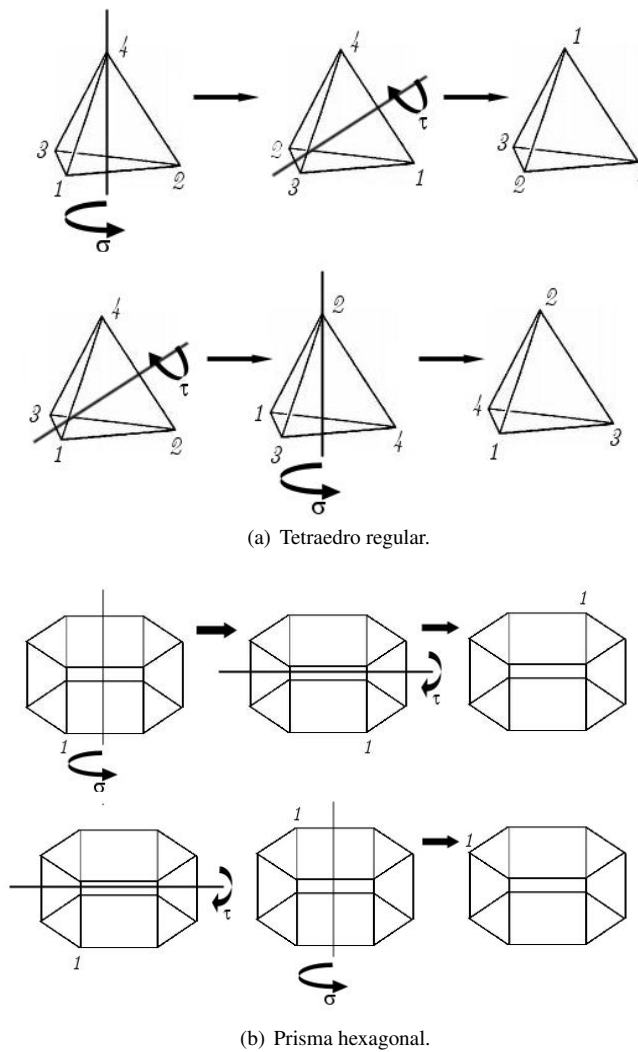


Figura 2.1: Simetrías en un tetraedro y en un prisma hexagonal.

figura hay una estructura más profunda y más rica que la puramente conjuntista, es decir, hay más información que la simplemente numérica. De hecho, en los tres casos las simetrías de cada figura tienen en sus relaciones rasgos comunes, aunque dichas relaciones acaben mostrando comportamientos distintos. Así, la identidad es siempre una simetría de una figura; la composición de dos simetrías es una simetría; la composición con la identidad no tiene ningún efecto; la composición es asociativa (es decir, dadas tres simetrías ρ_1, ρ_2 y ρ_3 tales que $\rho_1 \circ \rho_2 = \tau_1$ y $\rho_2 \circ \rho_3 = \tau_2$, entonces al aplicar sucesivamente las tres simetrías es $\rho_1 \circ \rho_2 \circ \rho_3 = \tau_1 \circ \rho_3 = \rho_1 \circ \tau_2$); y toda simetría tiene otra simetría que deshace sus efectos, una inversa. Aislando estas propiedades, podemos dar una definición axiomática de una estructura que se comporta, genéricamente, como las simetrías de una figura:

Definición 2.2.1 Un grupo es un par (G, \cdot) formado por un conjunto G y una aplicación

$$\begin{aligned} \cdot: G \times G &\rightarrow G \\ (a, b) &\mapsto \cdot(a, b) = a \cdot b \end{aligned}$$

que verifican las siguientes propiedades:

- 1) Propiedad asociativa: $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ para todo $a, b, c \in G$.
- 2) Existencia de elemento neutro: Existe $e \in G$ tal que $a \cdot e = e \cdot a = a$ para todo $a \in G$.



3) Existencia de inverso: Para todo $a \in G$ existe $b \in G$ tal que $a \cdot b = b \cdot a = e$.

Se dice que el grupo es *conmutativo* (o *abeliano*¹) si, además de las tres propiedades anteriores, se verifica

4) Propiedad conmutativa: $a \cdot b = b \cdot a$ para todo $a, b \in G$. \square

Ejemplo 2.2.2

- Las simetrías del tetraedro, del prisma hexagonal, de la pirámide dodecagonal, ... forman grupos.
- Los números enteros \mathbb{Z} , racionales \mathbb{Q} , o reales \mathbb{R} , con la suma, constituyen grupos abelianos. También lo son números racionales (o reales) no nulos con el producto (debemos quitar el cero pues éste carece de inverso), así como los números complejos \mathbb{C} con la suma, o los complejos no nulos con el producto.
- Las aplicaciones biyectivas del conjunto $\{1, 2, \dots, n\}$ en sí mismo, con la composición de aplicaciones, forman un grupo (no abeliano). A este grupo se le llama el *grupo simétrico*, S_n , o el grupo de *permutaciones* de n elementos.
- Las matrices 2×2 , con la suma de matrices, son un grupo abeliano. Las matrices 2×2 con determinante no nulo, con el producto de matrices, son un grupo no abeliano. A este grupo se le llama el *grupo general lineal* de orden 2, $GL(2, \mathbb{R})$ o, simplemente, $GL(2)$ (obviamente, en vez de 2, se puede usar cualquier otro número natural n). También son grupo con el producto:
 - las matrices de determinante 1 (el *grupo especial lineal*, $SL(n)$).
 - las matrices ortogonales, es decir, aquellas que verifican $A^T = A^{-1}$ (el *grupo ortogonal*, $O(n)$).
 - las matrices ortogonales con determinante 1 (el *grupo especial ortogonal*, $SO(n)$).
- Las clases de restos módulo $n \in \mathbb{N}$, \mathbb{Z}_n , son un grupo con la suma. Para p primo, $\mathbb{Z}_p \setminus \{0\}$ es también grupo con el producto.
- El *grupo de Klein* $\mathbb{Z}_2 \oplus \mathbb{Z}_2 = \{(0, 0), (1, 0), (0, 1), (1, 1)\}$ es un grupo con la suma, definida componente a componente (es un caso particular de suma directa de grupos abelianos²). \square

Todos estos ejemplos se acomodan a la definición que hemos obtenido “destilando” el comportamiento de las simetrías respecto a la composición. ¿Cuál es la ventaja de dar una definición axiomática, en vez de estudiar específicamente el caso que nos interese? Pues que, a partir de los axiomas de la definición, se puede demostrar *con una sola demostración* que ciertas propiedades se verifican *en todos los casos en que se cumpla la definición*, ya sean simetrías de figuras espaciales, matrices, números, permutaciones o clases de restos. Por ejemplo:

Proposición 2.2.3 Sea (G, \cdot) un grupo. Entonces:

- El elemento neutro e de G es único.
- Cada elemento de G tiene un único inverso. (Esto permite definir, para cada $a \in G$ su inverso, que denotaremos por a^{-1}).
- Sea $a \in G$. Si existe $b \in G$ tal que $a \cdot b = e$ entonces $b = a^{-1}$.

DEMOSTRACIÓN.

- Supongamos que e y e' son dos elementos neutros. Entonces, por ser e neutro, se verifica que

$$e \cdot e' = e'. \quad (2.1)$$

Pero, por ser e' también elemento neutro, se tiene que

$$e \cdot e' = e. \quad (2.2)$$

Por tanto, de (2.1) y (2.2), se concluye que $e = e'$.

¹Por el matemático noruego Niels Henrik Abel (1802–1829).

²Sean $(G, +)$ y $(H, +)$ dos grupos abelianos. La *suma directa* de G y H , que representaremos como $G \oplus H$, es el grupo constituido por el conjunto $G \times H$ con la operación $(g_1, h_1) + (g_2, h_2) = (g_1 + g_2, h_1 + h_2)$ para todo $g_1, g_2 \in G$ y $h_1, h_2 \in H$.



b) Sea $a \in G$ y supongamos que $b, c \in G$ son dos inversos de a . Entonces, por definición, se verifica que

$$b = b \cdot e = b \cdot (a \cdot c) = (b \cdot a) \cdot c = e \cdot c = c.$$

c) Como $a \in G$, existe $a^{-1} \in G$. Entonces,

$$b = e \cdot b = (a^{-1} \cdot a) \cdot b = a^{-1} \cdot (a \cdot b) = a^{-1} \cdot e = a^{-1}. \quad \square$$

2.2.2 Subgrupos. Grupos finitos. Orden

El grupo de rotaciones del espacio que dejan invariante un prisma n -gonal se denomina *grupo diédrico n -ésimo*, y se denota por D_n . Normalmente, se visualiza más bien como el grupo de transformaciones del plano que deja invariante un polígono regular de n lados, y así lo definiremos más adelante. La visualización se puede hacer intuitivamente, haciendo que la altura del prisma tienda hacia 0. Las bases coinciden entonces en un polígono plano regular de n lados. De este modo, las rotaciones alrededor del eje que une el centro de las bases se convierten en rotaciones planas alrededor del centro del polígono; y las rotaciones de ángulo π respecto a ejes paralelos a las bases tienen el efecto de “voltear” el polígono alrededor de dicho eje: es decir, el mismo efecto que una reflexión (que es una transformación del plano) respecto a ese eje.

Vimos, al principio de este capítulo, que D_6 tiene 12 elementos. Visualicemos D_6 como las rotaciones y reflexiones planas que dejan invariante un hexágono regular, e imaginemos ahora que incrustamos un triángulo equilátero en el hexágono, uniendo vértices alternos, como se muestra en la Figura 2.2.

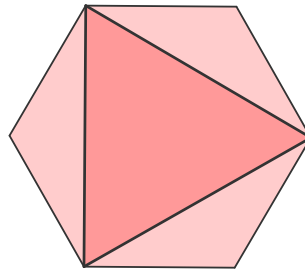


Figura 2.2: Triángulo incrustado en un hexágono.

Entonces, parte de las simetrías del hexágono también son simetrías del triángulo: las rotaciones alrededor del centro de ambos, y ángulos $\frac{2\pi}{3}$ y $\frac{4\pi}{3}$, y las reflexiones con respecto a las rectas que unen los vértices del triángulo con los puntos medios de los lados opuestos. En total, seis simetrías. De este modo, visualizamos D_3 como subconjunto de D_6 , pero de forma que D_3 , con la operación de D_6 restringida a sus elementos, es también grupo. Decimos que D_3 es *subgrupo* de D_6 . En general,

Definición 2.2.4 Sea (G, \cdot) un grupo. Un subconjunto $H \subseteq G$ es un *subgrupo* de G (y se escribe $H \leq G$) si (H, \cdot) es grupo. \square

Ejemplo 2.2.5

a) $(2\mathbb{Z}, +) \leq (\mathbb{Z}, +) \leq (\mathbb{Q}, +) \leq (\mathbb{R}, +) \leq (\mathbb{C}, +)$.

b) $SO(n) \leq SL(n) \leq GL(n)$. \square

Notación 2.2.6 Sea (G, \cdot) un grupo. Si $x \in G$ y $n \in \mathbb{Z}$, vamos a denotar por x^n a

$$x^n = \begin{cases} x \cdot \overset{(n)}{\cdot} \cdot x & \text{si } n > 0 \\ e & \text{si } n = 0 \\ x^{-1} \cdot \overset{(-n)}{\cdot} \cdot x^{-1} & \text{si } n < 0. \end{cases} \quad \square$$



Definición 2.2.7 Sea (G, \cdot) un grupo y $x \in G$. El conjunto

$$\langle x \rangle = \{x^n : n \in \mathbb{Z}\}$$

es, trivialmente, un subgrupo abeliano de G , y se llama el *subgrupo generado por el elemento x* . Si existe un elemento $y \in G$ tal que

$$G = \langle y \rangle$$

se dice que G es un *grupo cíclico*. \square

Definición 2.2.8 Si G es un grupo con una cantidad finita de elementos, se llama *orden de G* , y lo denotaremos por $|G|$, al cardinal de G (en el caso de que G no sea un conjunto finito, diremos que tiene *orden infinito*). Si $x \in G$, llamaremos *orden de x* a

$$\text{ord}(x) = |\langle x \rangle|. \quad \square$$

Observación 2.2.9 El orden de un elemento x , si es finito, es el menor entero positivo n tal que $x^n = e$. \square

Ejemplo 2.2.10 El orden de D_6 es 12. De sus elementos, la identidad tiene orden 1, la rotación de ángulo π y las reflexiones tienen orden 2, las rotaciones de ángulo $\frac{k\pi}{3}$ ($k = 1, 5$) tienen orden 6, y las de ángulo $\frac{2k\pi}{3}$ ($k = 1, 2$) tienen orden 3. \square

Ejemplo 2.2.11 El orden de todos los elementos del grupo de Klein es 2 (excepto el elemento neutro, $(0, 0)$, que tiene orden 1). \square

Acercas de los subgrupos, nos será útil saber que:

Proposición 2.2.12 Sea (G, \cdot) un grupo y H un subconjunto no vacío de G . Se verifica que:

$$H \leq G \Leftrightarrow x \cdot y^{-1} \in H \text{ para todo } x, y \in H.$$

DEMOSTRACIÓN.

\Rightarrow Evidente.

\Leftarrow Veamos que H es un subgrupo de G :

- 1) La propiedad asociativa se verifica automáticamente por ser G grupo.
- 2) El elemento neutro $e \in G$ está, de hecho, en H , pues dado cualquier $x \in H$ (y hay alguno, pues $H \neq \emptyset$) se tiene que $e = x \cdot x^{-1} \in H$.
- 3) El inverso de cualquier $x \in H$ también está en H , ya que $e, x \in H \Rightarrow x^{-1} = e \cdot x^{-1} \in H$.
- 4) Pero, además, para que (H, \cdot) sea un grupo, debe ser $\cdot : H \times H \rightarrow H$. En efecto, $\forall x, y \in H \Rightarrow x, y^{-1} \in H \Rightarrow x \cdot y = x \cdot (y^{-1})^{-1} \in H$. \square

Proposición 2.2.13 Sea (G, \cdot) un grupo y $H_1, H_2 \subseteq G$. Si $H_1 \leq G$ y $H_2 \leq G$ entonces $H_1 \cap H_2 \leq G$.

DEMOSTRACIÓN. Usando la Proposición 2.2.12, basta probar que

$$x, y \in H_1 \cap H_2 \Rightarrow x \cdot y^{-1} \in H_1 \cap H_2.$$

Si $x, y \in H_1 \cap H_2$, entonces $x, y \in H_1$ y, por ser H_1 un grupo, se verifica que $x \cdot y^{-1} \in H_1$; análogamente, $x \cdot y^{-1} \in H_2$. Así pues, $x \cdot y^{-1} \in H_1 \cap H_2$. \square



2.2.3 Isomorfismos: cuándo dos grupos son “iguales”

Ejemplo 2.2.14 Consideremos el conjunto formado por las rotaciones en el plano alrededor del origen, de ángulos $0, \frac{\pi}{2}, \pi$ y $\frac{3\pi}{2}$, que denotamos $\sigma_0, \sigma_1, \sigma_2$ y σ_3 , respectivamente. Este conjunto, con el producto de composición, constituye un grupo G_1 . La tabla de multiplicar de este grupo sería

	σ_0	σ_1	σ_2	σ_3
σ_0	σ_0	σ_1	σ_2	σ_3
σ_1	σ_1	σ_2	σ_3	σ_0
σ_2	σ_2	σ_3	σ_0	σ_1
σ_3	σ_3	σ_0	σ_1	σ_2

Por otra parte, llamamos G_2 al grupo formado por el conjunto de números complejos $\{1, i, -1, -i\}$ dotado de la multiplicación habitual; la tabla de multiplicar del grupo sería

	1	i	-1	$-i$
1	1	i	-1	$-i$
i	i	-1	$-i$	1
-1	-1	$-i$	1	i
$-i$	$-i$	1	i	-1

Finalmente, la tabla de “multiplicación” de $G_3 = \mathbb{Z}_4$ (i.e., el conjunto $\{0, 1, 2, 3\}$ con la suma módulo 4) es

	0	1	2	3
0	0	1	2	3
1	1	2	3	0
2	2	3	0	1
3	3	0	1	2

Existe una completa analogía entre las tres tablas. Si definimos una aplicación $f: G_1 \rightarrow G_2$ mediante

$$\sigma_0 \mapsto 1, \sigma_1 \mapsto i, \sigma_2 \mapsto -1, \sigma_3 \mapsto -i,$$

entonces esta aplicación es biyectiva y verifica que

$$f(x \circ y) = f(x) \cdot f(y),$$

ya que las tablas de multiplicar en ambos grupos son idénticas a través de f . Esto no es extraño, pues multiplicar por un número complejo de módulo 1 equivale a realizar un giro de ángulo igual al argumento del número. Igualmente, la aplicación $g: G_2 \rightarrow G_3$ definida mediante

$$1 \mapsto 0, i \mapsto 1, -1 \mapsto 2, -i \mapsto 3,$$

es biyectiva y verifica que

$$g(x \cdot y) = g(x) + g(y).$$

De nuevo, no hay sorpresas: basta pensar que $G_2 = \{i^0, i^1, i^2, i^3\}$ y que al multiplicar dos potencias de la misma base se suman los exponentes. Es decir que, para nosotros, los tres grupos son “el mismo” a todos los efectos. Más precisamente, diremos que f y g son “isomorfismos” y G_1, G_2 y G_3 son “isomorfos”. \square

Definición 2.2.15 Sean (G, \cdot) y $(G', *)$ grupos. Una aplicación $f: G \rightarrow G'$ es un *isomorfismo* (de grupos) si es biyectiva y verifica que

$$f(x \cdot y) = f(x) * f(y)$$

para todo $x, y \in G$. En ese caso, los grupos G y G' se dicen *isomorfos* y se escribe $G \cong G'$. \square

Ejemplo 2.2.16 Son isomorfismos:

a) La aplicación $(\mathbb{R}, +) \rightarrow (\mathbb{R}^+, \cdot)$ dada por $x \mapsto \exp(x)$.



- b) La aplicación $D_3 \rightarrow S_3$ que a cada simetría le hace corresponder el efecto de la simetría sobre los vértices del triángulo (comprobar como ejercicio). \square

Proposición 2.2.17 Sean G y G' dos grupos y sea $f: G \rightarrow G'$ un isomorfismo entre ellos. Entonces:

- a) G y G' tienen el mismo cardinal.
 b) Si e y e' son, respectivamente, los elementos neutros de G y G'

$$f(e) = e' \text{ y } f(x^{-1}) = (f(x))^{-1}.$$

- c) Si H es subgrupo de G , $f(H)$ es subgrupo de G' (Los isomorfismos llevan subgrupos a subgrupos).
 d) Si $x \in G$, $\text{ord}(x) = \text{ord}(f(x))$ (Los isomorfismos preservan el orden de los elementos).
 e) Si G es un grupo abeliano G' también lo es.

DEMOSTRACIÓN.

- a) Trivial (pues todo isomorfismo es una biyección).
 b) Para todo $x \in G$ se tiene

$$f(x) = f(xe) = f(x)f(e),$$

así que

$$e' = (f(x))^{-1}f(x) = (f(x))^{-1}f(x)f(e) = e'f(e) = f(e),$$

de modo que $f(e)$ es el elemento neutro de G' . En cuanto al inverso, se tiene que

$$f(x)f(x^{-1}) = f(xx^{-1}) = f(e) = e'$$

por lo que $f(x^{-1})$ es el inverso de $f(x)$.

- c) Sea $H \leq G$. Basta observar que para todo $f(x), f(y) \in f(H)$ se verifica que

$$f(x)f(y)^{-1} = f(xy^{-1}) \in f(H),$$

ya que $xy^{-1} \in H$.

- d) En efecto, puesto que $f(\langle x \rangle)$ es un subgrupo y se verifica que

$$f(x^n) = (f(x))^n,$$

se tiene que

$$f(\langle x \rangle) = \langle f(x) \rangle.$$

Consecuentemente,

$$\text{ord}(x) = |\langle x \rangle| = |f(\langle x \rangle)| = |\langle f(x) \rangle| = \text{ord}(f(x)).$$

- e) Basta tener en cuenta que para todo $x' = f(x)$ e $y' = f(y)$ se verifica que

$$x'y' = f(x)f(y) = f(xy) = f(yx) = y'x'. \quad \square$$

Observación 2.2.18 Con las dos últimas afirmaciones de la Proposición 2.2.17 podemos formalizar que los tres grupos de simetrías de la Sección 2.1 son “distintos” en el sentido de que no pueden ser isomorfos. \square



2.2.4 Homomorfismos de grupos. Núcleo e imagen

Definición 2.2.19 Una aplicación $f: G \rightarrow G'$ es un *homomorfismo de grupos* si se verifica que

$$f(xy) = f(x)f(y) \text{ para todo } x, y \in G. \quad \square$$

Ejemplo 2.2.20 El determinante de una matriz es un homomorfismo de grupos entre $GL(n)$ y $(\mathbb{R} \setminus \{0\}, \cdot)$. Nótese que no es un isomorfismo. \square

Observación 2.2.21 Las afirmaciones segunda y tercera de la Proposición 2.2.17 se verifican también para homomorfismos, pues en su demostración no se utiliza que la aplicación en cuestión sea biyectiva. \square

Definición 2.2.22 Sea $f: G \rightarrow G'$ un homomorfismo de grupos.

- a) $\ker(f) = \{x \in G : f(x) = e'\}$ es el *núcleo* de f .
 b) $\text{im}(f) = f(G) = \{f(x) : x \in G\}$ es la *imagen* de f . \square

Ejemplo 2.2.23 Si $f: GL(n) \rightarrow (\mathbb{R} \setminus \{0\}, \cdot)$ es el homomorfismo definido como $f(A) = \det(A)$ se tiene que $\ker(f) = SL(n)$ e $\text{im}(f) = \mathbb{R} \setminus \{0\}$. \square

Proposición 2.2.24 Si $f: G \rightarrow G'$ es un homomorfismo de grupos se tiene que $\ker(f) \leq G$ e $\text{im}(f) \leq G'$.

DEMOSTRACIÓN. Para demostrar que $\ker(f) \leq G$, basta observar que para todo $x, y \in \ker(f)$ se verifica

$$f(xy^{-1}) = f(x)(f(y))^{-1} = e',$$

por lo que $xy^{-1} \in \ker(f)$. Por otra parte, la imagen de G es subgrupo de G' , ya que los homomorfismos llevan subgrupos en subgrupos, como se dice en la Observación 2.2.21. \square

Ejercicio 2.2.25 Sea $f: G \rightarrow G'$ un homomorfismo de grupos. Demostrar:

- a) La aplicación f es inyectiva $\Leftrightarrow \ker(f) = \{e\}$.
 b) La aplicación f es sobreyectiva $\Leftrightarrow \text{im}(f) = G'$. \square

2.2.5 Subgrupos normales. El grupo cociente

Dado $H \leq G$, consideremos la relación de equivalencia dada por

$$x_H \sim y \Leftrightarrow xy^{-1} \in H.$$

La clase de equivalencia de x es

$$Hx = \{hx : h \in H\},$$

pues

$$y_H \sim x \Leftrightarrow yx^{-1} = h \in H \Leftrightarrow y = hx \text{ con } h \in H.$$

Ejemplo 2.2.26 Consideremos $G = \mathbb{Z}_{20}$ y $H = \{0, 4, 8, 12, 16\}$ con la suma módulo 20. Entonces $H \leq G$, dos elementos de G están relacionados cuando su diferencia está en H (es decir, si su diferencia es múltiplo de 4) y la clase de equivalencia de $x \in G$ será $\{x, x+4, x+8, x+12, x+16\}$ (entendiendo la suma módulo 20). \square

Cada una de estas clases de equivalencia tiene exactamente $|H|$ elementos. En efecto, la aplicación

$$\begin{aligned} H &\rightarrow Hx \\ h &\mapsto hx \end{aligned}$$

es trivialmente sobreyectiva; por otra parte, es inyectiva, ya que si $hx = h'x$ entonces $(hx)x^{-1} = (h'x)x^{-1}$ y, por tanto, $h = h'$.

Puesto que G es la unión disjunta de sus clases de equivalencia, obtenemos el siguiente resultado:



Teorema 2.2.27 (Lagrange) Si G es un grupo finito y $H \leq G$ se verifica que $|H|$ divide a $|G|$.

DEMOSTRACIÓN. Si llamamos G/H al conjunto cociente de G por la relación de equivalencia $H \sim$,

$$G/H = \{Hx : x \in G\},$$

como los Hx son una partición de G , cada Hx tiene $|H|$ elementos, y hay $|G/H|$ de ellas, tenemos que $|G| = |G/H||H|$. \square

Definición 2.2.28 $H \leq G$ se llama *subgrupo normal* (y se escribe $H \triangleleft G$) si se verifica que

$$ghg^{-1} \in H \text{ para todo } g \in G \text{ y } h \in H. \quad \square$$

Observación 2.2.29 Si G es abeliano todos sus subgrupos son normales. \square

Ejemplo 2.2.30 $SL(2)$ es un subgrupo normal de $GL(2)$ puesto que se tiene la propiedad

$$\det(ABA^{-1}) = \det(B) = 1 \text{ para todo } A \in GL(2), B \in SL(2). \quad \square$$

Observación 2.2.31 Lo interesante es que si $H \triangleleft G$, entonces la operación de G le da estructura de grupo a G/H . En efecto, podemos definir el producto de dos clases mediante

$$(Hx)(Hy) = Hxy$$

sin que la definición dependa del representante elegido, ya que

$$(h_1x_1h_2y)(xy)^{-1} = h_1x_1h_2yy^{-1}x^{-1} = h_1x_1h_2x^{-1} = h_1(x_1h_2x^{-1}) \in H,$$

por ser H normal. \square

Proposición 2.2.32 Si $f: G \rightarrow G'$ es un homomorfismo, entonces $\ker(f) \triangleleft G$.

DEMOSTRACIÓN. Para todo $x \in \ker(f)$ y $z \in G$ se verifica que

$$f(zxz^{-1}) = f(z)f(x)f(z)^{-1} = f(z)f(z)^{-1} = e',$$

luego $z x z^{-1} \in \ker(f)$. \square

2.3 Isometrías del plano y matrices ortogonales

Definición 2.3.1 Una *isometría del plano* es una aplicación biyectiva $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ tal que

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

o, equivalentemente,

$$\|f(x) - f(y)\| = \|x - y\|. \quad \square \tag{2.3}$$

Observación 2.3.2

a) Las isometrías preservan ángulos: dado un ángulo \widehat{abc} , como los triángulos abc y $f(a)f(b)f(c)$ son “iguales” (por coincidir la longitud de sus tres lados), se verifica que

$$f(a)\widehat{f(b)f(c)} = \widehat{abc}.$$

b) En la definición anterior no es necesario pedir que la aplicación f sea biyectiva, pues esta propiedad se puede deducir de (2.3). \square

Proposición 2.3.3 Las isometrías del plano, con la composición, forman un grupo que denotamos E_2 .



DEMOSTRACIÓN. Veamos que E_2 cumple las propiedades de grupo:

a) Por la definición es muy fácil comprobar que la composición de isometrías es isometría

$$\|(g \circ f)(x) - (g \circ f)(y)\| = \|g(f(x)) - g(f(y))\| \stackrel{g \text{ isom}}{=} \|f(x) - f(y)\| \stackrel{f \text{ isom}}{=} \|x - y\|,$$

y la propiedad asociativa se verifica automáticamente.

b) Nótese que la identidad es, trivialmente, una isometría.

c) Por otra parte, dada una isometría f , su inversa (que existe por ser f biyectiva) cumple que

$$\|f^{-1}(x) - f^{-1}(y)\| = \|f(f^{-1}(x)) - f(f^{-1}(y))\| = \|x - y\|. \quad \square$$

Ejemplo 2.3.4 Son isometrías:

a) La *traslación* de vector v , $t_v: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dada por $t_v(x) = v + x$.

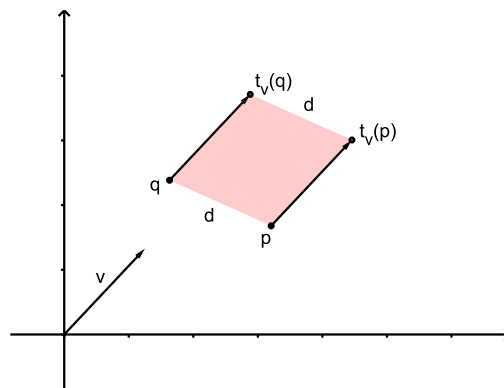


Figura 2.3: Traslación de vector v .

En efecto, en la Figura 2.3 se observa que, dados dos puntos p y q , los puntos p , q , $t_v(p)$ y $t_v(q)$ forman un paralelogramo, de donde $d(t_v(p), t_v(q)) = d = d(p, q)$. En realidad, esto puede verse directamente, pues

$$\|t_v(x) - t_v(y)\| = \|v + x - v - y\| = \|x - y\|.$$

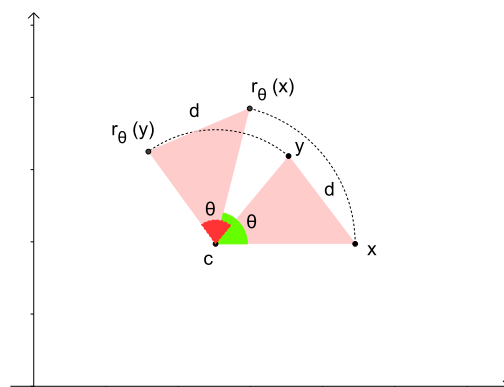
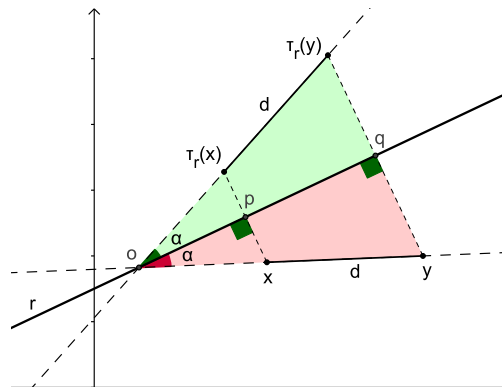


Figura 2.4: Rotación r_θ de ángulo θ alrededor de c .

b) Las *rotaciones*. Observamos en la Figura 2.4, por un lado, que $d(c, x) = d(c, r_\theta(x))$, y que $d(c, y) = d(c, r_\theta(y))$. Además, se tiene que

$$\widehat{xc r_\theta(x)} + r_\theta(\widehat{xc r_\theta(y)}) = \widehat{xc r_\theta(y)} = \widehat{xcy} + \widehat{yc r_\theta(y)}.$$

Figura 2.5: Reflexión τ_r respecto a la recta r .

Como $\widehat{ocr_\theta(x)} = \widehat{ycr_\theta(y)} = \theta$, se deduce que $\widehat{xcy} = r_\theta(x)\widehat{cr_\theta(y)}$. De este modo los triángulos $ctxy$ y $cr_\theta(x)r_\theta(y)$ son iguales y, por tanto, $d(x, y) = d(r_\theta(x), r_\theta(y))$.

- c) Las reflexiones respecto a una recta. En la Figura 2.5 se observa que los triángulos opx y $op\tau_r(x)$ son ambos rectángulos y de catetos iguales dos a dos, por lo que también coinciden sus hipotenusas. Se puede razonar análogamente con los triángulos ogy y $oq\tau_r(y)$. Se concluye que $d(\tau_r(x), \tau_r(y)) = d(o, \tau_r(y)) - d(o, \tau_r(x)) = d(o, y) - d(o, x) = d(x, y)$. \square

Definición 2.3.5 Una aplicación $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ es *lineal* si verifica:

- 1) $f(x + y) = f(x) + f(y)$ para todo $x, y \in \mathbb{R}^2$.
- 2) $f(\lambda x) = \lambda f(x)$ para todo $x \in \mathbb{R}^2$ y $\lambda \in \mathbb{R}$. \square

Observación 2.3.6 Toda aplicación lineal $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ viene determinada por las imágenes de una base. En efecto, si escribimos los elementos de \mathbb{R}^2 en forma de vectores columna y escogemos la base canónica

$$\mathcal{B} = \left\{ e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\},$$

denotando las imágenes por f de e_1 y e_2 por

$$f \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix} \text{ y } f \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} b \\ d \end{pmatrix},$$

se verifica que

$$\begin{aligned} f(x) &= f \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = f \left(x_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right) = x_1 f \begin{pmatrix} 1 \\ 0 \end{pmatrix} + x_2 f \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ &= x_1 \begin{pmatrix} a \\ c \end{pmatrix} + x_2 \begin{pmatrix} b \\ d \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = Mx. \end{aligned}$$

Por tanto, f queda determinada por la matriz

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

que se denomina *matriz asociada* a la aplicación lineal f . \square

Proposición 2.3.7 Las isometrías del plano que dejan fijo el origen son aplicaciones lineales.



DEMOSTRACIÓN. Sea $f: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ una isometría verificando $f(0) = 0$. Tenemos entonces que

$$\|f(y)\| = \|f(y) - f(0)\| = \|y - 0\| = \|y\| \text{ para todo } y \in \mathbb{R}^2.$$

a) Veamos que $f(\lambda x) = \lambda f(x)$ para todo $x \in \mathbb{R}^2$ y $\lambda \in \mathbb{R}$. Obviamente, la propiedad es cierta si $x = 0$ o $\lambda = 0$. En otro caso, puesto que f conserva la alineación de tres puntos, se verifica que $0, f(x)$ y $f(\lambda x)$ están alineados. Por tanto, $f(\lambda x) = \mu f(x)$ para algún $\mu \in \mathbb{R}$. Así pues,

$$|\mu| \|x\| = |\mu| \|f(x)\| = \|\mu f(x)\| = \|f(\lambda x)\| = \|\lambda x\| = |\lambda| \|x\|$$

y, con ello, $\mu = \pm\lambda$. Pero si $f(\lambda x) = -\lambda f(x)$ se llegaría a que

$$|\lambda + 1| \|x\| = \|(-\lambda - 1)f(x)\| = \|f(\lambda x) - f(x)\| = \|\lambda x - x\| = |\lambda - 1| \|x\|.$$

Como $x \neq 0$, $|\lambda + 1| = |\lambda - 1|$, lo que sólo se da si $\lambda = 0$ (llegando a una contradicción).

b) Veamos $f(x + y) = f(x) + f(y)$ para todo $x, y \in \mathbb{R}^2$. Si los puntos $x, 0$ e y están alineados entonces $y = \mu x$ y basta aplicar el apartado anterior. En otro caso, los puntos anteriores forman el triángulo $x0y$, el cual se transforma por f en el triángulo $f(x)0f(y)$; asimismo, el triángulo $x(x+y)y$ se transforma en el triángulo $f(x)p f(y)$ siendo $p = f(x+y)$. Por ser f una isometría y ser $0, x, x+y, y$ un paralelogramo, el cuadrilátero $0, f(x), p, f(y)$ es también un paralelogramo ("igual" al anterior). En consecuencia, se tiene que $p = f(x) + f(y)$ (véase la Figura 2.6). \square

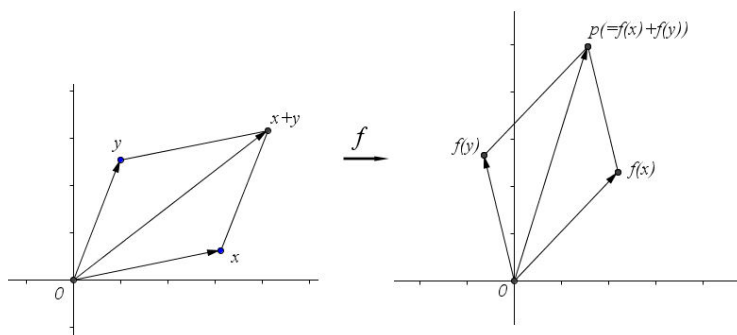


Figura 2.6: Linealidad de las isometrías.

Proposición 2.3.8 Las únicas isometrías del plano que dejan fijo el origen son las rotaciones en torno al origen y las reflexiones con eje por el origen. Las primeras vienen dadas por matrices de la forma

$$A_\theta = \begin{pmatrix} \cos \theta & -\operatorname{sen} \theta \\ \operatorname{sen} \theta & \cos \theta \end{pmatrix}$$

(cuyo determinante vale 1) siendo θ el ángulo de rotación y, las segundas, por matrices del tipo

$$B_\varphi = \begin{pmatrix} \cos \varphi & \operatorname{sen} \varphi \\ \operatorname{sen} \varphi & -\cos \varphi \end{pmatrix}$$

(de determinante -1) siendo $\frac{\varphi}{2}$ el ángulo que forma el eje con la horizontal.

DEMOSTRACIÓN. Por la Proposición 2.3.7, las isometrías que dejan fijo el origen son aplicaciones lineales y vienen, por tanto, dadas por una matriz M . Puesto que deben preservar la norma, M debe verificar que

$$M^T M = I_2$$

(esta propiedad se demuestra escribiendo $M^T M$ y usando la igualdad

$$x^T x = \|x\|^2 = \|Mx\|^2 = (Mx)^T Mx, x \in \mathbb{R}^2$$



aplicada a los vectores $x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $x = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ y $x = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

Por ser $GL(2)$ un grupo, se tiene que

$$M^T = M^{-1} \Rightarrow MM^T = I_2.$$

Consecuentemente, si

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

entonces

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} a^2 + b^2 & ac + bd \\ ac + bd & c^2 + d^2 \end{pmatrix}.$$

Por una parte, la relación

$$a^2 + b^2 = c^2 + d^2 = 1$$

implica que existen θ y φ tales que

$$a = \cos \varphi, b = \sin \varphi, d = \cos \theta \text{ y } c = \sin \theta.$$

pero, por otra parte, la relación

$$ac + bd = 0$$

implica que

$$0 = \cos \varphi \sin \theta + \sin \varphi \cos \theta = \sin(\theta + \varphi),$$

es decir,

$$\theta + \varphi = 0 \text{ ó } \theta + \varphi = \pi.$$

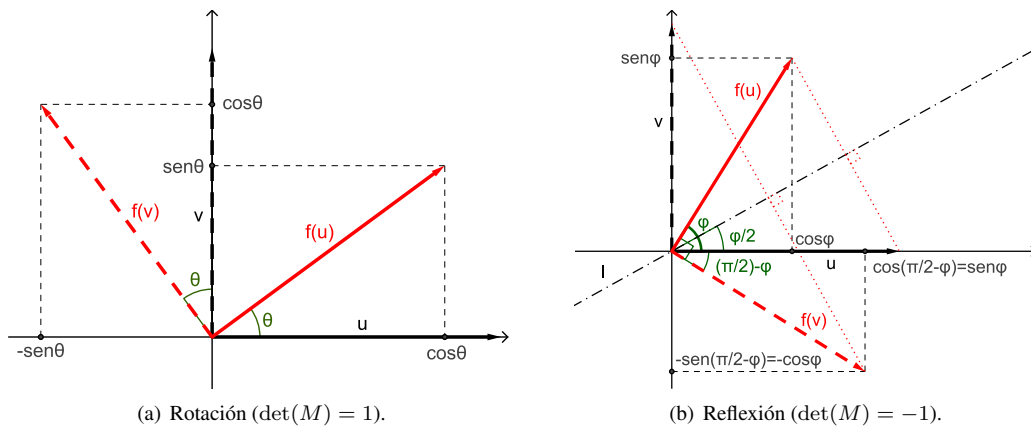


Figura 2.7: Rotación y reflexión.

a) En el primer caso, $\theta + \varphi = 0 \Rightarrow \cos \varphi = \cos \theta$ y $\sin \varphi = -\sin \theta$, por lo que

$$M = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.$$

Como una *rotación* de ángulo θ alrededor del origen tiene asociada esa matriz (véase la Figura 2.7(a)), la isometría es dicha rotación. Como $\det(M) = 1$ tenemos que $M \in SO(2)$.

b) En el segundo caso, $\theta + \varphi = \pi \Rightarrow \cos \theta = -\cos \varphi$ y $\sin \theta = \sin \varphi$, por lo que la matriz

$$M = \begin{pmatrix} \cos \varphi & \sin \varphi \\ \sin \varphi & -\cos \varphi \end{pmatrix}.$$

Como una *reflexión* respecto al eje por el origen de ángulo $\frac{\varphi}{2}$ con la horizontal tiene asociada esa matriz (véase la Figura 2.7(b)), la isometría es dicha reflexión. En este caso, $\det(M) = -1$. \square



Observación 2.3.9 De la Proposición 2.3.8 se deduce que toda matriz $M \in O(2)$ verifica

$$\det(M) = \pm 1$$

y, en el caso de que

$$\det(M) = -1 \Rightarrow M^2 = I_2.$$

Nótese que, en general el recíproco no es cierto (pues M podría ser la identidad o una rotación de ángulo $\theta = \pi$). \square

Proposición 2.3.10 Las únicas isometrías del plano son las rotaciones en torno al origen compuestas con las traslaciones y las reflexiones con eje por el origen compuestas con traslaciones.

DEMOSTRACIÓN. Supongamos que f es una isometría y que $f(0) = v$. Si denotamos por t_v la traslación de vector v , se tiene que la isometría

$$t_v^{-1} \circ f = g$$

fija el origen. El resultado se sigue aplicando la Proposición 2.3.8. \square

Observación 2.3.11 Así pues, lo que tenemos es una descripción de cualquier isometría del plano, f , en términos de una traslación t_v ,

$$t_v(x) = v + x,$$

y una transformación ortogonal f_M ,

$$f_M(x) = Mx,$$

mediante $f = (v, M)$ si

$$f = t_v \circ f_M,$$

es decir, si

$$f(x) = v + Mx.$$

Si $\det(M) = 1$ se habla de *isometrías directas* y, si no, de *isometrías inversas*. \square

Observación 2.3.12 Para calcular la expresión de la composición de (v, M) con (u, N) basta observar que

$$(t_v \circ f_M) \circ (t_u \circ f_N)(p) = v + M(u + Np) = (v + Mu) + MNp = t_{v+Mu} \circ f_{MN}(p),$$

luego

$$(v, M)(u, N) = (v + Mu, MN). \quad \square$$

Observación 2.3.13 Toda isometría f se corresponde de forma única con un par (v, M) . En efecto, si

$$v + Mx = f(x) = u + Nx \quad \text{con } v, u \in \mathbb{R}^2 \text{ y } M, N \in O(2)$$

entonces

$$v = f(0) = u \text{ y } Mx = Nx, x \in \mathbb{R}^2 \Rightarrow M = N. \quad \square$$

Clasificamos entonces las isometrías del plano en función de la descomposición propuesta:

Teorema 2.3.14 Los casos posibles de isometrías del plano se reducen a:

a) Traslaciones de vector v . Son de la forma (v, I_2) .

b) Rotaciones de centro c y ángulo θ . Son de la forma (v, A_θ) con

$$v = c - A_\theta c \text{ y } A_\theta = \begin{pmatrix} \cos \theta & -\text{sen } \theta \\ \text{sen } \theta & \cos \theta \end{pmatrix}$$



c) Reflexiones con eje una recta ℓ que forma un ángulo $\frac{\varphi}{2}$ con la horizontal. Son de la forma (v, B_φ) con

$$v = 2a \text{ y } B_\varphi = \begin{pmatrix} \cos \varphi & \text{sen } \varphi \\ \text{sen } \varphi & -\cos \varphi \end{pmatrix}$$

siendo a el punto de ℓ más cercano al origen.

d) Reflexiones con deslizamiento con eje una recta ℓ que forma un ángulo $\frac{\varphi}{2}$ con la horizontal y con vector de deslizamiento b (paralelo a ℓ). Son de la forma (v, B_φ) con

$$v = 2a + b \text{ y } B_\varphi = \begin{pmatrix} \cos \varphi & \text{sen } \varphi \\ \text{sen } \varphi & -\cos \varphi \end{pmatrix}$$

siendo a el punto de ℓ más cercano al origen.

DEMOSTRACIÓN.

a) Trivial.

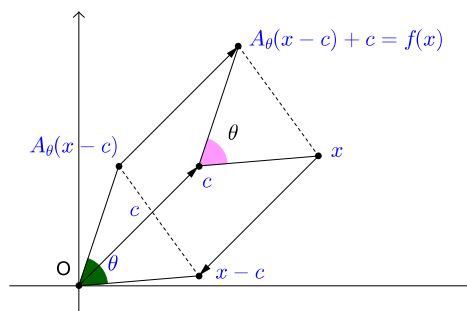


Figura 2.8: Descripción de una rotación.

b) Si f es una rotación de centro c y ángulo θ , en la Figura 2.8 se muestra cómo f se construye como una traslación de c al origen, el giro de ángulo θ respecto a éste y, finalmente, la traslación del origen a c , es decir,

$$f(x) = c + A_\theta(x - c),$$

por lo que

$$f(x) = (c - A_\theta c) + A_\theta x.$$

Recíprocamente, toda isometría $f = (v, A_\theta)$ con $\theta \neq 0$ se corresponde con una rotación. En efecto, de ser f una rotación tendría un punto fijo c (el centro de rotación) que debería verificar

$$f(c) = c \Rightarrow v + A_\theta c = c \Rightarrow (I_2 - A_\theta)c = v.$$

Así c es solución del sistema lineal

$$\begin{pmatrix} 1 - \cos \theta & \text{sen } \theta \\ -\text{sen } \theta & 1 - \cos \theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}.$$

Como el determinante de la matriz del sistema es

$$(1 - \cos \theta)^2 + \text{sen}^2 \theta = 1 - 2 \cos \theta + \cos^2 \theta + \text{sen}^2 \theta = 2 - 2 \cos \theta = 2(1 - \cos \theta) \neq 0$$

(ya que $\theta \neq 0$), se tendría que

$$c = (I_2 - A_\theta)^{-1}v.$$

Si llamamos g a la rotación de centro este valor de c y ángulo θ entonces $f \equiv g$ pues, como hemos visto antes, se tiene que

$$\begin{aligned} g(p) &= (c - A_\theta c) + A_\theta p = (I_2 - A_\theta)c + A_\theta p \\ &= (I_2 - A_\theta)(I_2 - A_\theta)^{-1}v + A_\theta p \\ &= v + A_\theta p = f(p), \quad p \in \mathbb{R}^2. \end{aligned}$$



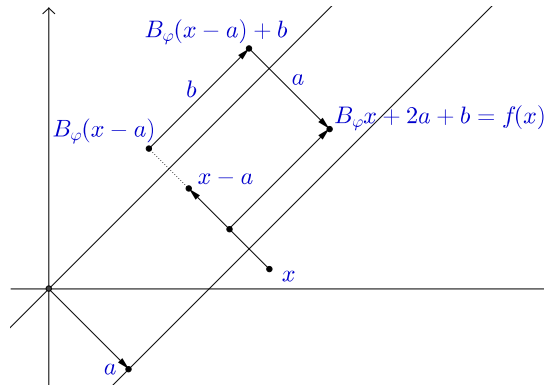


Figura 2.9: Descripción de una reflexión con deslizamiento.

d) (y c)) Sea m la recta paralela a ℓ por el origen y f_{B_φ} la reflexión respecto a m . Para construir la reflexión sobre ℓ con deslizamiento b , basta trasladar ℓ hasta m , reflejar en m mediante f_{B_φ} , sumar el vector b y, finalmente, volver a trasladar hasta ℓ , es decir (véase la Figura 2.9),

$$f(x) = a + B_\varphi(x - a) + b = a + B_\varphi x - B_\varphi a + b = a + B_\varphi x + a + b = 2a + b + B_\varphi x,$$

ya que $B_\varphi a = -a$. Recíprocamente, si f es una isometría del tipo $f = (v, B_\varphi)$, consideramos

$$b = \frac{v + B_\varphi v}{2} \text{ y } a = \frac{v - B_\varphi v}{4}.$$

La justificación de esta elección queda ilustrada en la Figura 2.10. Distinguimos dos casos:

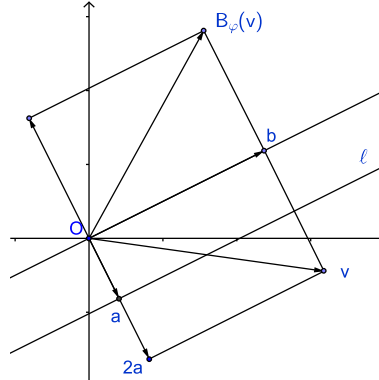


Figura 2.10: Vectores a , b y v de una reflexión con deslizamiento.

i) Supongamos, en primer lugar, que $b \neq 0$. Puesto que $B_\varphi^2 = I_2$,

$$B_\varphi b = \frac{B_\varphi v + B_\varphi^2 v}{2} = \frac{B_\varphi v + v}{2} = b.$$

Es decir, el vector b queda invariante por B_φ y, por tanto, es un vector director del eje de la reflexión B_φ (esto es, la recta que pasa por el origen y forma un ángulo $\frac{\varphi}{2}$ con la horizontal). Por otra parte, como $B_\varphi a = -a$, se verifica que a es perpendicular a dicho eje. (Esto puede verse directamente:

$$\begin{aligned} a^T b &= \frac{1}{8} (v - B_\varphi v)^T (v + B_\varphi v) = \frac{1}{8} (v^T v - v^T B_\varphi^T v + v^T B_\varphi v - v^T B_\varphi^T B_\varphi v) \\ &= \frac{1}{8} (v^T v - v^T B_\varphi v + v^T B_\varphi v - v^T I_2 v) = 0. \end{aligned}$$

Así pues, la recta ℓ de dirección b que pasa por el punto a tiene la propiedad de que a es el punto de ℓ más cercano al origen. Si llamamos g a la reflexión respecto a ℓ con deslizamiento b , por lo visto anteriormente, se tiene que

$$g(p) = 2a + b + B_\varphi p = \frac{v - B_\varphi v}{2} + \frac{v + B_\varphi v}{2} + B_\varphi p = v + B_\varphi p = f(p), \quad p \in \mathbb{R}^2,$$

por lo que $f \equiv g$, es decir, f es una reflexión con deslizamiento.

- ii) Si $b = 0$, entonces $B_\varphi v = -v$ y $a = \frac{v}{2}$. Así pues, el vector v (y, por tanto, a) es perpendicular al eje de reflexión de B_φ . Razonando como antes, la recta ℓ paralela al eje de B_φ que pasa por el punto a tiene la propiedad de que a es el punto de ℓ más cercano al origen. Ahora, si llamamos g a la reflexión respecto a ℓ , se verifica que

$$g(p) = 2a + B_\varphi p = v + B_\varphi p = f(p), \quad p \in \mathbb{R}^2,$$

por lo que $f \equiv g$, es decir, f es una reflexión. \square

Ejemplo 2.3.15

- a) Para calcular las ecuaciones de una rotación de centro $(1, 2)$ y ángulo $\frac{\pi}{2}$, en primer lugar calculamos la matriz de la rotación, que es

$$A_{\frac{\pi}{2}} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Para encontrar el vector v usamos que el centro c queda invariante por la rotación, es decir,

$$v + A_{\frac{\pi}{2}} c = c \Rightarrow v + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \Rightarrow v = \begin{pmatrix} 2 \\ -1 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

Recíprocamente, conocidas las ecuaciones de esta isometría

$$f(x, y) = (3 - y, 1 + x),$$

tenemos que

$$f \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

de donde

$$v = \begin{pmatrix} 3 \\ 1 \end{pmatrix} \quad \text{y} \quad M = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Como M tiene determinante 1, se trata de una rotación. Para calcular el ángulo θ usamos que

$$\begin{cases} \cos \theta = 0 \\ \text{sen } \theta = 1 \end{cases} \Rightarrow \theta = \frac{\pi}{2}.$$

Finalmente, el centro c se obtiene usando que $f(c) = c$, es decir, resolviendo el sistema

$$\begin{pmatrix} 3 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} \Leftrightarrow \begin{cases} c_1 + c_2 = 3 \\ -c_1 + c_2 = 1, \end{cases}$$

cuya solución es $c = (1, 2)$.

- b) Las ecuaciones de la reflexión con respecto a la recta $x - y - 1 = 0$ pueden calcularse de la siguiente forma: puesto que $(1, 1)$ es un vector director de la recta, el ángulo $\frac{\varphi}{2}$ que ésta forma con la horizontal verifica que

$$\tan \frac{\varphi}{2} = \frac{1}{1} = 1 \Rightarrow \frac{\varphi}{2} = \frac{\pi}{4} \Rightarrow \varphi = \frac{\pi}{2}.$$



De esta forma, la matriz de esta reflexión es

$$B_{\frac{\pi}{2}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Sabemos que $v = 2a$ siendo a el punto de la recta más cercano al origen; éste lo calculamos como la intersección del eje de reflexión con la recta perpendicular a él que pasa por el origen, es decir,

$$\begin{cases} x - y = 1 \\ x + y = 0. \end{cases}$$

Por tanto, $a = (\frac{1}{2}, -\frac{1}{2})$ y, en consecuencia, $v = (1, -1)$.

Recíprocamente, conocidas las ecuaciones de esta isometría

$$g(x, y) = (1 + y, -1 + x),$$

tenemos que

$$g \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

de donde

$$v = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \text{ y } M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Como M tiene determinante -1 , se trata de una reflexión o una reflexión con deslizamiento (según sea nulo o no b). Usando las fórmulas, puesto que $Mv = B_{\varphi}v = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$, tenemos

$$a = \frac{v - B_{\varphi}v}{4} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} \text{ y } b = \frac{v + B_{\varphi}v}{2} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

luego se trata de una reflexión; el eje de simetría es la recta perpendicular al vector a que pasa por el punto a , es decir, es una recta de la forma

$$x - y = c,$$

donde c se determina como

$$c = \frac{1}{2} + \frac{1}{2} = 1.$$

Así, el eje de reflexión es la recta $x - y - 1 = 0$.

- c) Para obtener las ecuaciones de la reflexión con respecto a la recta $x - y - 1 = 0$ y deslizamiento $b = (2, 2)$ se procede como en el caso anterior, obteniéndose el mismo vector a y la misma matriz B_{φ} ; el vector v es ahora

$$v = 2a + b = \begin{pmatrix} 3 \\ 1 \end{pmatrix}.$$

Recíprocamente, conocidas las ecuaciones de esta isometría

$$h(x, y) = (3 + y, 1 + x),$$

tenemos que

$$h \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

de donde

$$v = \begin{pmatrix} 3 \\ 1 \end{pmatrix} \text{ y } M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$



Como M tiene determinante -1 , se trata de una reflexión o una reflexión con deslizamiento (según sea nulo o no b). Usando las fórmulas, puesto que $Mv = B_\varphi v = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$, tenemos

$$a = \frac{v - B_\varphi v}{4} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} \quad \text{y} \quad b = \frac{v + B_\varphi v}{2} = \begin{pmatrix} 2 \\ 2 \end{pmatrix},$$

luego se trata de una reflexión con deslizamiento $b = (2, 2)$. Para calcular el eje de simetría podemos proceder como antes, calculando la recta perpendicular al vector a que pasa por el punto a . Otra forma de pensarlo es considerar la recta con vector director b que pasa por a , es decir,

$$2x - 2y = c,$$

donde c se determina como

$$c = 2 \cdot \frac{1}{2} + 2 \cdot \frac{1}{2} = 2.$$

Así, el eje de reflexión es la recta $x - y - 1 = 0$. \square

2.4 Grupo diédrico e isometrías del plano

En la Definición 2.2.7 se introdujo el concepto de grupo cíclico como el generado por un elemento. Esta noción puede extenderse a dos elementos de la siguiente forma:

Definición 2.4.1 Si un grupo G tiene ciertos elementos x, y tales que

$$\begin{aligned} G = \langle x, y \rangle &= \{x^{j_1} y^{k_1} x^{j_2} y^{k_2} \dots x^{j_n} y^{k_n} : n \geq 0, j_i, k_i \in \mathbb{Z}\} \\ &= \{e, x, y, x^2, x^3, x^{-1}, y^2, xy, yx^3y, x^2y^{-3}xy^3x^{-1}y^2, \dots\} \end{aligned}$$

se dice que G está *generado* por x e y (o que x e y *generan* G). \square

Vamos a buscar elementos que generen el grupo diédrico, que introdujimos en la Sección 2.2.2. Para ello, comencemos por dar su definición de manera formal, y estudiar algunas de sus propiedades.

Definición 2.4.2 Llamaremos *grupo diédrico* D_n al grupo formado por las simetrías de un n -gono regular (es decir, las isometrías que lo dejan fijo). \square

Observación 2.4.3 El orden de D_n es $2n$ puesto que un n -gono regular tiene $2n$ simetrías. En efecto: una simetría f lleva vértices en vértices por lo que, fijado uno de ellos, por ejemplo el p_1 , sólo tenemos n posibles imágenes para él. Por otra parte, como los transformados de dos vértices vecinos son también vecinos entre sí, si $f(p_1) = p_k$, hay dos posibilidades (entendiendo, si $k = n$, que $p_{n+1} = p_1$):

- $f(p_2) = p_{k+1}$ (y, por tanto, $f(p_n) = p_{k-1}$). f es entonces una rotación.
- $f(p_2) = p_{k-1}$ (y, por tanto, $f(p_n) = p_{k+1}$). Ahora f es una reflexión.

Así pues, tenemos un total de $2n$ simetrías. \square

Ejemplo 2.4.4 El grupo diédrico D_5 de simetrías de un pentágono regular está generado por la rotación σ de ángulo $\frac{2\pi}{5}$ alrededor del centro del pentágono, y por una reflexión τ respecto a cualquier eje que une un vértice con el punto medio del lado opuesto, es decir,

$$D_5 = \langle \sigma, \tau \rangle.$$

En efecto, las simetrías del pentágono son los giros alrededor de su centro, de ángulos $\frac{2k\pi}{5}$, $k = 1, 2, 3, 4$ (pero estos son σ^k), la identidad $I_2 = \sigma^0 = \sigma^5$, y las reflexiones respecto a los ejes que unen cada vértice con el punto medio del lado opuesto. Una de estas reflexiones es τ . Si numeramos los vértices del pentágono



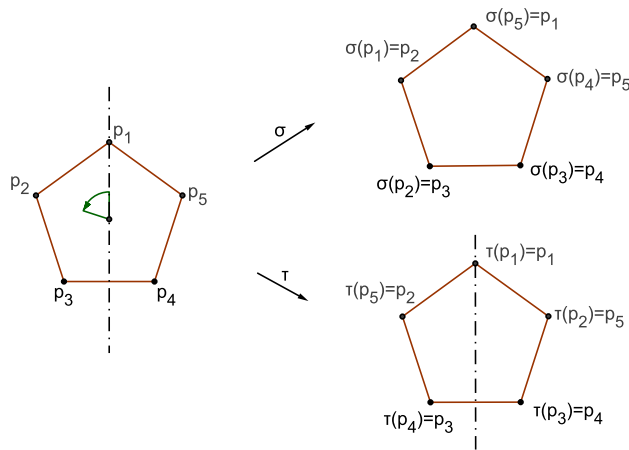


Figura 2.11: Los generadores de D_5 .

como p_1, p_2, p_3, p_4, p_5 , siendo p_1 el vértice que queda fijo por τ (ver Figura 2.11), la acción de τ sobre los vértices es

$$\begin{aligned} \tau : p_1 &\mapsto p_1 \\ p_2 &\leftrightarrow p_5 \\ p_3 &\leftrightarrow p_4. \end{aligned}$$

Si ahora giramos mediante σ , tenemos que $\sigma \circ \tau$ se comporta así:

$$\begin{aligned} \sigma\tau : p_1 &\xrightarrow{\tau} p_1 \xrightarrow{\sigma} p_2 \\ p_2 &\mapsto p_5 \mapsto p_1 \\ p_3 &\mapsto p_4 \mapsto p_5 \\ p_4 &\mapsto p_3 \mapsto p_4 \\ p_5 &\mapsto p_2 \mapsto p_3 \end{aligned}$$

es decir, $\sigma\tau = \sigma \circ \tau$ es la reflexión cuyo eje pasa por p_4 y el punto medio de $\overline{p_1p_2}$. De modo análogo, se ve que $\sigma^2\tau$ es la reflexión cuyo eje pasa por p_2 , $\sigma^3\tau$ aquella cuyo eje pasa por p_5 y $\sigma^4\tau$ la que tiene el eje por p_3 . Así pues, tenemos 10 elementos distintos de D_5 y, por tanto,

$$D_5 = \{I_2, \sigma, \sigma^2, \sigma^3, \sigma^4, \tau, \sigma\tau, \sigma^2\tau, \sigma^3\tau, \sigma^4\tau\}.$$

Veamos que $\sigma\tau = \tau\sigma^4$ probando que $\sigma\tau$ y $\tau\sigma^4$ tienen el mismo efecto sobre los vértices. En efecto,

$$\begin{aligned} \tau\sigma^4 : p_1 &\xrightarrow{\sigma^4} p_5 \xrightarrow{\tau} p_2 \\ p_2 &\mapsto p_1 \mapsto p_1 \\ p_3 &\mapsto p_2 \mapsto p_5 \\ p_4 &\mapsto p_3 \mapsto p_4 \\ p_5 &\mapsto p_4 \mapsto p_3 \end{aligned}$$

De manera más general, se puede probar que

$$\sigma^k\tau = \tau\sigma^{5-k}, \quad k = 1, 2, 3, 4.$$

Nótese que esta propiedad no es más que la formalización de la idea de que reflejar en un eje y luego rotar el polígono es como rotar el polígono *en sentido contrario* y luego reflejar. \square



Observación 2.4.5 En el caso general, puede demostrarse que el grupo diédrico D_n está generado por dos elementos, σ y τ (es decir, $D_n = \langle \sigma, \tau \rangle$) que verifican

$$\sigma^n = I_2, \tau^2 = I_2 \text{ y } \sigma^k \tau = \tau \sigma^{n-k}, k = 1, 2, \dots, n-1,$$

donde I_2 denota la identidad, σ es una rotación de ángulo $\frac{2\pi}{n}$ y τ es una reflexión (aunque la demostración en el caso n par es ligeramente distinta a la que hemos indicado para D_5). \square

Observación 2.4.6 Se puede probar que todo grupo generado por dos elementos a y b que verifiquen que

$$b^2 = e \text{ y } ab = ba^{n-1} (= ba^{-1}),$$

siendo n el orden de a (i.e. $a^n = e$) es isomorfo a D_n . Nótese que, en particular, las propiedades $ab = ba^{n-1}$ y $a^n = e$ implican que

$$a^k b = ba^{n-k}, k = 1, 2, \dots, n-1.$$

En efecto, el caso base de la inducción es nuestra hipótesis y, en general,

$$a^{k+1}b = aa^k b = aba^{n-k} = ba^{n-1}a^{n-k} = ba^{2n-(k+1)} = ba^n a^{n-(k+1)} = ba^{n-(k+1)}. \quad \square$$

El resultado fundamental de esta sección clasifica los subgrupos finitos de las isometrías del plano que fijan el origen:

Proposición 2.4.7 *Todo subgrupo finito $G \leq O(2)$ es cíclico o diédrico.*

DEMOSTRACIÓN. Distinguimos dos casos:

a) Supongamos primero que $G \leq SO(2)$ y veamos que G es cíclico. En esta situación todos los elementos de G son rotaciones y, como hay un número finito, existe θ tal que es el menor ángulo no nulo de las rotaciones de G . Veamos que existe $n \in \mathbb{N}$ tal que

$$\theta = \frac{2\pi}{n}.$$

En efecto, A_θ debe tener orden finito, digamos m , por tenerlo G . Como

$$A_\theta^m = I_2 \Rightarrow \text{existe } k \in \{1, 2, \dots, m-1\} \text{ tal que } m\theta = 2\pi k \Rightarrow \theta = \frac{2k\pi}{m}.$$

Haciendo la división entera de m entre k , existen $n \in \mathbb{N}$ y $0 \leq r < k$ tales que

$$m = kn + r.$$

Así pues,

$$(n+1)\theta = (n+1)\frac{2k\pi}{m} = 2(kn+k)\frac{\pi}{m} = 2(m+k-r)\frac{\pi}{m} = 2\pi + 2(k-r)\frac{\pi}{m},$$

por lo que

$$A_\theta^{n+1} = A_\varphi \text{ siendo } \varphi = 2(k-r)\frac{\pi}{m} \leq \theta.$$

Si $r \neq 0$ el ángulo φ de la rotación $A_\varphi \in G$ sería menor que θ (lo que contradice la hipótesis). En conclusión,

$$r = 0 \Rightarrow m = kn \Rightarrow \theta = \frac{2k\pi}{m} = \frac{2\pi}{n}.$$

Finalmente, es obvio que $G = \langle A_\theta \rangle$ pues, si no, existiría en G una rotación A_ψ con $\psi \in (k\theta, (k+1)\theta)$ para algún k y, por tanto, $A_{\psi-k\theta}$ estaría en G , lo que no es posible al ser $\psi - k\theta < \theta$.



b) Supongamos ahora que G no es subgrupo de $SO(2)$. Consideramos

$$H = G \cap SO(2)$$

que es un subgrupo de $SO(2)$ que, por a)), es cíclico: $H = \langle \sigma \rangle$ con $\sigma^n = I_2$. El conjunto cociente G/H tiene, exactamente, dos elementos: H y $H\tau$ para algún $\tau \in G \setminus H$ (en efecto, si $\tau' \in G \setminus H$, entonces $\tau\tau'^{-1} = \tau\tau' \in H$, por ser su determinante 1, y así $H\tau = H\tau'$). Por lo tanto, todo elemento de G está en una de las dos clases de equivalencia anteriores y, así, usando que $H = \langle \sigma \rangle$, podemos afirmar que

$$G = \{I_2, \sigma, \sigma^2, \dots, \sigma^{n-1}, \tau, \sigma\tau, \sigma^2\tau, \dots, \sigma^{n-1}\tau\}.$$

Además, al ser G un grupo, se tiene que $G = \langle \sigma, \tau \rangle$. Para ver que $G \cong D_n$, como sabemos que $\tau^2 = I_2$ y $\sigma^n = I_2$, gracias a la Observación 2.4.6, sólo nos resta comprobar que

$$\sigma\tau = \tau\sigma^{n-1}.$$

Pero el determinante de $\sigma\tau$ es -1 , es decir, $\sigma\tau$ es una reflexión, y por tanto

$$\sigma\tau = (\sigma\tau)^{-1} = \tau^{-1}\sigma^{-1} = \tau\sigma^{n-1}. \quad \square$$

2.5 Grupos cristalográficos planos

Pretendemos ahora *realizar* o identificar ciertos subgrupos del grupo de isometrías de todo el plano como los grupos de simetría de un patrón o configuración infinita (pero discreta) de puntos del plano (infinita, para que tenga sentido que las traslaciones sean simetrías de la configuración; si no, estaríamos hablando, directamente, de subgrupos de $O(2)$); discreta, para que haya un dominio fundamental compacto, es decir, que las teselas con las que vamos a recubrir el plano tengan un área finita y no nula). Como veremos, a esos patrones de puntos los llamaremos *retículos*.

Notación 2.5.1 Denotamos por π el homomorfismo $\pi: E_2 \rightarrow O(2)$ dado por

$$\pi(g) = M$$

para

$$g = (v, M) = t_v \circ f_M. \quad \square$$

Observación 2.5.2 π es, en efecto, un homomorfismo, pues

$$\pi((v_1, M_1)(v_2, M_2)) = \pi(v_1 + M_1v_2, M_1M_2) = M_1M_2 = \pi(v_1, M_1)\pi(v_2, M_2)$$

(véase la Observación 2.3.12). \square

Definición 2.5.3 Dado un subgrupo $G \leq E_2$, definimos su *subgrupo de traslaciones* como

$$T = G \cap \ker(\pi) = \ker(\pi|_G),$$

y el *grupo puntual* de G como

$$J = \pi(G) \leq O(2). \quad \square$$

Ejemplo 2.5.4 Consideremos el conjunto de puntos del plano formado por un panel de celdas hexagonales como las de la Figura 2.12, donde los vectores u y v unen centros de hexágonos contiguos y son linealmente independientes. Sea G el grupo de las simetrías de este conjunto, es decir, el formado por todas las traslaciones de vector $nu + mv$ con $n, m \in \mathbb{Z}$, las rotaciones alrededor del centro de algún hexágono de ángulo $\frac{k\pi}{3}$ con $k = 0, 1, \dots, 5$, las reflexiones respecto a ejes que pasan por vértices opuestos o puntos medios de lados opuestos de un hexágono, y sus composiciones. En este caso, pues, no es difícil ver que

$$T = \langle t_u, t_v \rangle \text{ y } J = D_6. \quad \square$$



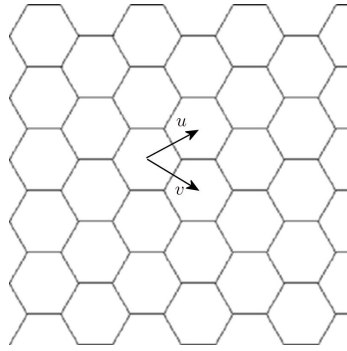


Figura 2.12: Panel de celdas hexagonales.

Definición 2.5.5 Los *grupos cristalográficos planos* son aquellos subgrupos de E_2 cuyos subgrupos de traslaciones tienen dos generadores independientes y cuyo grupo puntual es finito. \square

Observación 2.5.6

- La primera condición asegura que el grupo está formado por las simetrías de una configuración auténticamente bidimensional, y no unidimensional. Es decir, que tratamos con grupos realmente del plano, y no extensiones triviales al plano de grupos de simetrías de la recta (los cuales conducen a los denominados *frisos*).
- La segunda condición tiene que ver con la condición discreta (no continua) del retículo; implica que hay un subconjunto del plano (una celda del retículo) que se va a repetir indefinidamente hasta cubrir todo el plano. Sin esta condición, los puntos del retículo podrían estar infinitamente cerca, y la tesela con la que cubrimos el plano podría ser infinitamente pequeña. \square

Para obtener la clasificación de estos grupos (que nos van a dar todos los posibles grupos de simetrías de mosaicos del plano), necesitamos algunos resultados acerca de los subgrupos de traslaciones y los grupos puntuales.

2.6 Retículos y su clasificación

Nota: En lo que resta de este capítulo supondremos que G es un grupo cristalográfico plano. \square

Definición 2.6.1 Un *retículo* (plano) es un subconjunto R de \mathbb{R}^2 para el que existen dos vectores $u, v \in \mathbb{R}^2$ de modo que

$$R = \{mu + nv : m, n \in \mathbb{Z}\}. \quad \square$$

Observación 2.6.2 Nótese que R es el subgrupo de $(\mathbb{R}^2, +)$ generado por los vectores u y v . \square

Proposición 2.6.3 Sean $G \leq E_2$ y T el subgrupo de traslaciones de G . Entonces, el conjunto

$$L = \{t(0) : t \in T\}$$

es el retículo generado por a y b , siendo:

- a un elemento de norma minimal de L .
- b un elemento de L no alineado con a , de longitud mínima.

Diremos que L es el retículo asociado al grupo cristalográfico plano G .

DEMOSTRACIÓN. Por ser G cristalográfico plano, $T = \langle t_u, t_v \rangle$ para ciertos vectores u, v . Además, como L es isomorfo a T mediante el isomorfismo $t_w \mapsto w = t_w(0)$, se verifica que L es el subgrupo de $(\mathbb{R}^2, +)$ generado por los vectores u y v , es decir, L es el retículo asociado a esos mismos vectores.



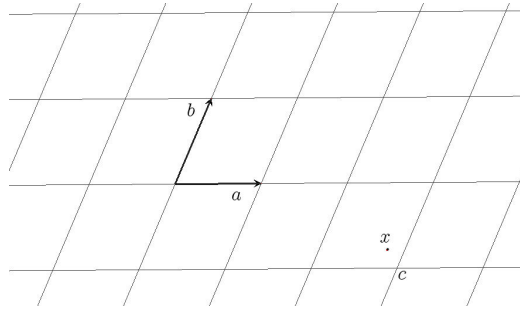


Figura 2.13: División del plano en paralelogramos.

Por lo tanto, si escogemos a y b en L como indica el enunciado, el retículo L' generado por a y b está contenido en L ya que las combinaciones enteras de a y b son, en particular, combinaciones enteras de u y v (por serlo a y b). El retículo L' genera una división del plano en paralelogramos como se ve en la Figura 2.13.

Veamos que $L' = L$ argumentando por reducción al absurdo: supongamos que existe $x \in L$ tal que $x \notin L'$; entonces, escogemos el paralelogramo que contiene a x y llamamos c a la esquina del paralelogramo más cercana a x (véase, nuevamente, la Figura 2.13). Veamos que

$$\|x - c\| < \|b\|. \tag{2.4}$$

En efecto, el punto del paralelogramo más distante de su esquina más cercana (y, por tanto, de todas las esquinas) es su centro, y la distancia del centro del paralelogramo a su esquina más cercana es

$$\min \left\{ \left\| \frac{a+b}{2} \right\|, \left\| \frac{a-b}{2} \right\| \right\} < \frac{\|a\| + \|b\|}{2} \leq \|b\|$$

donde se ha usado que, por definición, a y b no están alineados y $\|a\| \leq \|b\|$.

Por otra parte, por construcción, $x - c = \alpha a + \beta b$ con $\alpha, \beta \in [-1, 1]$. Veamos que β no puede ser nulo: Por una parte, si lo fuera y $\alpha \in (-1, 1)$ se llegaría a que $\|x - c\| < \|a\|$, lo que contradice que a sea de norma mínima en L . Pero tampoco puede ocurrir que $\beta = 0$ y $\alpha = \pm 1$ ya que, como $x - c \in L$ y $c \in L'$, tenemos que $x - c \notin L'$ y, en particular,

$$x - c \neq \pm a \in L'.$$

Así pues, $\beta \neq 0$ o, lo que es lo mismo, $x - c$ es un vector de L que no está alineado con a ; y gracias a (2.4) sabemos que tiene longitud estrictamente menor que b . Esto contradice la elección de b . Deducimos, en consecuencia, que no hay ningún elemento en $L \setminus L'$. \square

Observación 2.6.4 Los retículos se pueden clasificar en función del ángulo y magnitud relativa de a y b . La forma en que se han elegido a y b asegura que $\|a\| \leq \|b\| \leq \|a - b\|$. Sustituyendo, en caso de que sea necesario, b por $-b$ para asegurar que $\|a - b\| \leq \|a + b\|$, se tiene la siguiente cadena de desigualdades:

$$\|a\| \leq \|b\| \leq \|a - b\| \leq \|a + b\|$$

El teorema de Pitágoras descarta la posibilidad de que

$$\|b\| = \|a - b\| = \|a + b\|$$

por lo que los casos posibles son (véase la Figura 2.14):

- a) *Oblicuo*: $\|a\| < \|b\| < \|a - b\| < \|a + b\|$.
- b) *Rectangular*: $\|a\| < \|b\| < \|a - b\| = \|a + b\|$.



c) *Rectangular centrado*: $\|a\| < \|b\| = \|a - b\| < \|a + b\|$. En este caso, las direcciones a y $2b - a$ son perpendiculares.

d) *Cuadrado*: $\|a\| = \|b\| < \|a - b\| = \|a + b\|$.

e) *Hexagonal*: $\|a\| = \|b\| = \|a - b\| < \|a + b\|$.

Nótese que el caso restante (denominado *rómbico*)

$$\|a\| = \|b\| < \|a - b\| < \|a + b\|,$$

puede asimilarse al rectangular centrado con $a_0 = a - b$ y $b_0 = a$. En este caso, el generador a_0 del retículo ya no sería el elemento de norma mínima pero, aún así, lo asociamos al caso rectangular centrado porque el hecho característico de este caso es que las direcciones a_0 y $2b_0 - a_0$ son perpendiculares. \square

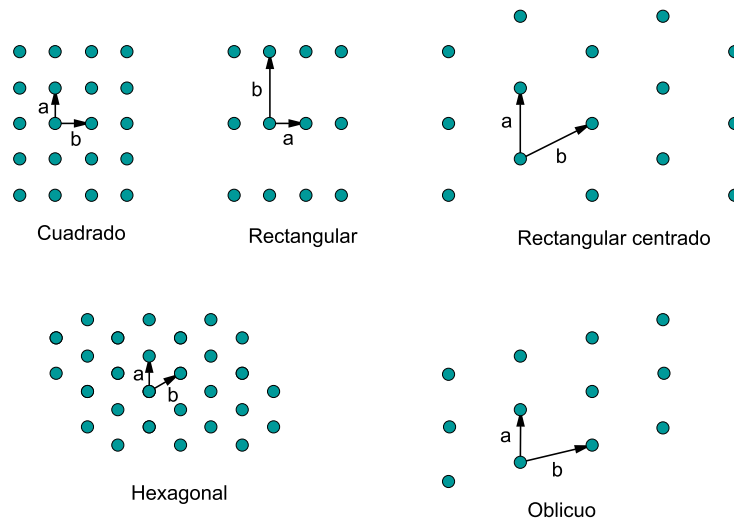


Figura 2.14: Clasificación de los retículos.

Observación 2.6.5 A pesar de que $J = \pi(G) \leq O(2)$, en general, J no es un subgrupo de G porque los elementos de J no tienen por qué estar en G . Por ejemplo, si tomamos como G el grupo generado por la traslación $t_v = (v, I_2)$ con $v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ y la reflexión con deslizamiento $\tau = (w, B_\pi)$ con $w = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, entonces (compruébese) el grupo de traslaciones y el grupo puntual de G son, respectivamente,

$$T = \langle t_v, t_{2w} \rangle \text{ y } J = \{I_2, B_\pi\}.$$

Sin embargo, $B_\pi \notin G$. En efecto, si $B_\pi \in G$ entonces la traslación $t_w = f_{B_\pi} \circ \tau$ debería estar en G (y, por tanto en T). Pero esto no ocurre, dado que w no puede ser una combinación entera de v y $2w$ (que son perpendiculares). De hecho, en G sólo hay traslaciones y reflexiones con deslizamiento, todas de orden infinito. \square

No obstante, el siguiente resultado nos asegura que los elementos de J son simetrías del retículo L :

Proposición 2.6.6 Para todo $A \in J$ y $p \in L$ se verifica que $f_A(p) \in L$. Dicho de otro modo,

$$f_A(L) = L,$$

por lo que f_A es una simetría de L .

DEMOSTRACIÓN. Sea $A \in J$. Entonces existe t_v tal que

$$t_v \circ f_A \in G.$$



Queremos ver que para todo $p = t_p(0) \in L$ se verifica que $f_A(p) \in L$. Como $T = \ker(\pi|_G)$, se verifica que $T \triangleleft G$ (véase la Proposición 2.2.32) y, por tanto,

$$(t_v \circ f_A) \circ t_p \circ (t_v \circ f_A)^{-1} \in T \text{ y } (t_v \circ f_A) \circ t_p \circ (t_v \circ f_A)^{-1}(0) \in L.$$

Pero

$$(t_v \circ f_A) \circ t_p \circ (t_v \circ f_A)^{-1}(0) = v + A(p + A^{-1}(0 - v)) = v + Ap - v = Ap = f_A(p). \quad \square$$

Lema 2.6.7 *Sea σ una rotación de centro c de orden n . Entonces existe $m \in \mathbb{N}$ tal que σ^m es una rotación con el mismo centro y de ángulo $\frac{2\pi}{n}$.*

DEMOSTRACIÓN. Sea θ el ángulo de σ ; como el orden de σ es n , existe $k \in \mathbb{N}$ tal que $n\theta = 2\pi k$, con $\text{mcd}(n, k) = 1$. Sea m solución de la congruencia $kx \equiv 1 \pmod{n}$, es decir, $km = nt + 1$ para algún $t \in \mathbb{Z}$. Ahora podemos escribir

$$m\theta = \frac{2\pi km}{n} = \frac{2\pi(nt + 1)}{n} = 2\pi t + \frac{2\pi}{n},$$

para concluir que $m\theta$, el ángulo de la rotación σ^m , es efectivamente $\frac{2\pi}{n}$. \square

Proposición 2.6.8 (Restricción cristalográfica) *Toda rotación de un grupo cristalográfico plano tiene orden 2, 3, 4 ó 6.*

DEMOSTRACIÓN. Las rotaciones de orden $n \geq 7$ tienen (véase el Lema 2.6.7) una potencia cuya es una rotación de ángulo $\frac{2\pi}{n} < \frac{\pi}{3}$. Sea A la matriz de esta rotación; supongamos que $A \in J$ y lleguemos a una contradicción. En efecto, por la Proposición 2.6.6, se tendría que $f_A(a) \in L$ y $f_A(a) - a \in L$, siendo a el generador de L de norma mínima; pero entonces se tendría la contradicción

$$\|f_A(a) - a\| < \|a\|,$$

por ser $\|f_A(a)\| = \|a\|$ y el ángulo de rotación menor que $\frac{\pi}{3}$ (véase la Figura 2.15).

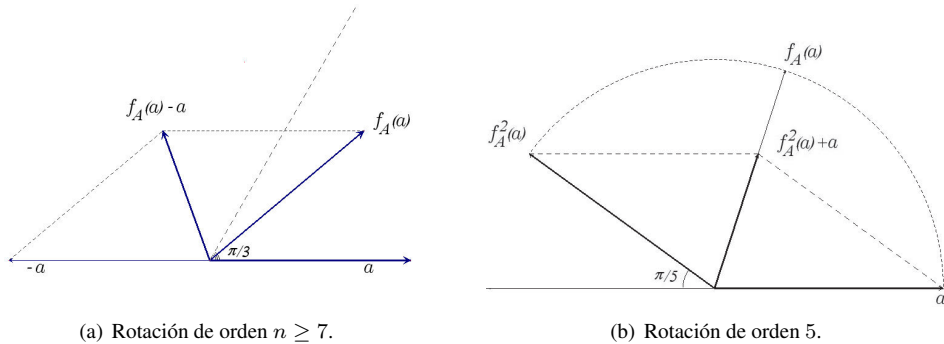


Figura 2.15: Rotaciones.

En cuanto a las rotaciones de orden 5, todas ellas poseen una potencia que es una rotación de ángulo $\frac{2\pi}{5}$ (nuevamente, gracias al Lema 2.6.7). En ese caso, si A es la matriz de esta rotación, se llegaría a que

$$\|f_A^2(a) + a\| < \|a\|$$

(véase, de nuevo, la Figura 2.15) lo cual constituye, de nuevo, una contradicción. \square

Para finalizar, enunciemos los siguientes resultados:

Proposición 2.6.9 *Un isomorfismo entre grupos cristalográficos planos respeta traslaciones, rotaciones, reflexiones y reflexiones con deslizamiento.*

DEMOSTRACIÓN. Véase [1, Teorema 25.5, pág. 152]. \square

Proposición 2.6.10 *Dos grupos cristalográficos planos isomorfos tienen grupos puntuales isomorfos.*

DEMOSTRACIÓN. Véase [1, Corolario 25.6, pág. 153]. \square



2.7 Aplicación: Clasificación de los mosaicos del plano

En esta sección se da una clasificación (salvo isomorfismo) de los grupos cristalográficos planos. Para ello, para cada uno de los 5 tipos de retículo, se estudia cuáles son las distintas posibilidades para el grupo puntual correspondiente, teniendo en cuenta la restricción cristalográfica y, también, que el grupo puntual debe dejar fijo el retículo. De alguna forma, pretendemos así “eliminar” las traslaciones, que corresponderían a la repetición periódica de la baldosa hasta cubrir el plano, para centrarnos en las simetrías de la baldosa misma. No obstante, en algunos casos, el mismo retículo puede tener un mismo grupo puntual que provenga de grupos de simetría distintos: en ese caso hay que analizar si las reflexiones que aparecen en el grupo puntual provienen de reflexiones de G o de reflexiones con deslizamiento, y considerar el ángulo que forman los ejes de las dichas reflexiones con los generadores del retículo.

Una vez clasificados los grupos cristalográficos planos, tenemos una clasificación, por sus grupos de simetría, de todas las configuraciones de puntos del plano que están creadas por la repetición periódica por traslaciones de una región finita del plano (que se corresponde con la *tesela* o *baldosa*), es decir, habremos dado un modelo de clasificación de los *mosaicos* o *pavimentos* periódicos del plano. Como veremos, hay exactamente 17 modelos (todos ellos pueden encontrarse en la decoración de la Alhambra de Granada).

Para denominar a cada uno de los tipos de grupos cristalográficos planos, introducimos una de las notaciones estándar, la cual está formada por cuatro caracteres (que, a veces, se abrevian a tres o a dos):

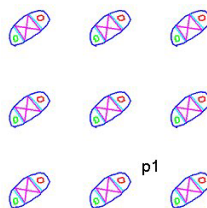
- a) En lo referente al retículo (Primer carácter: p o c):
 - i) p si se usa una celda primitiva del retículo (una copia del paralelogramo básico generado por a y b , sin puntos del retículo en su interior).
 - ii) c si se usa como unidad básica un conjunto de varias celdas primitivas, junto con su centro (que es un punto del retículo).
- b) En lo referente al grupo puntual:
 - i) Segundo carácter: 1, 2, 3, 4 ó 6 indica la existencia de una rotación del orden correspondiente (1 indicaría la ausencia de rotaciones, *i.e.*, que la única “rotación” presente es la identidad).
 - ii) Tercer y cuarto caracteres (m , g o 1):
 - m indica que hay una reflexión (*mirror*).
 - g indica que hay una reflexión con deslizamiento (*glide*).
 - 1 indica que no hay ninguna de ambas.

El tercer carácter se refiere a reflexiones o deslizamientos con eje perpendicular a alguna de las direcciones del retículo (eje principal). El cuarto carácter, con respecto a un eje (eje secundario) que o bien es perpendicular al anterior (en los casos rectangular o rectangular centrado), o bien forma con él un ángulo de $\pi/4$ (en el caso cuadrado) o es paralelo a un generador (en el caso hexagonal).

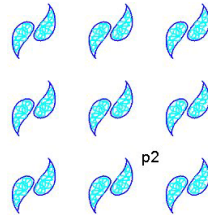
A menudo, cuando no aportan información relevante, se utiliza la “notación abreviada”, en la que se omite alguno de los caracteres anteriores. Así, por ejemplo, $p2 \equiv p211$ o $cm \equiv c2mm$.

Como ya se ha dicho, vamos a realizar una clasificación mediante los 5 tipos posibles de retículo estudiando, en cada caso, cuáles son las distintas posibilidades para el grupo puntual correspondiente, teniendo en cuenta que éste debe actuar sobre el retículo. Tenemos los siguientes casos:

- a) L es oblicuo. En este caso, sólo I_2 y $A_\pi = -I_2$ preservan el retículo. Por tanto, hay dos posibilidades:
 - $(p1) = (p111)$ Si $J = \{I_2\}$, en cuyo caso G está generado por las dos traslaciones.



- $(p2) = (p211)$ Si $J = \{I_2, -I_2\}$. Entonces, de hecho, en G sólo hay traslaciones y giros de ángulo π .

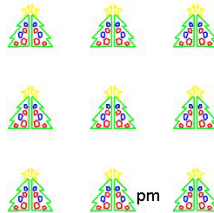


b) L es rectangular. Las transformaciones ortogonales que preservan L son

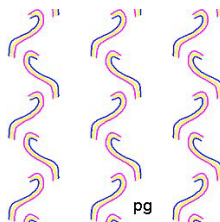
$$\{I_2, -I_2, B_0, B_\pi\}$$

y J es un subgrupo de este grupo. Si $J = \{I_2\}$ ó $J = \{I_2, -I_2\}$ volvemos a obtener los casos $p1$ y $p2$ (ya que el retículo rectangular es un caso particular del retículo oblicuo). Por tanto, los nuevos casos que aparecen son:

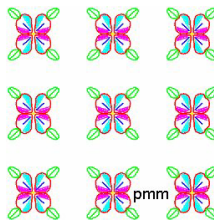
- $(pm) = (p1m1)$ Si $J = \{I_2, B_0\}$ y G contiene una reflexión de eje horizontal (en el caso $J = \{I_2, B_\pi\}$, el eje es vertical).



- $(pg) = (p1g1)$ Si con el mismo $J = \{I_2, B_0\}$, G no contiene una reflexión. En ese caso G debe contener una reflexión con deslizamiento (v, B_0) con v horizontal (de forma análoga, si cambiamos B_0 por B_π , la reflexión con deslizamiento es vertical).

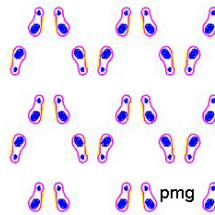


- $(pmm) = (p2mm)$ Si $J = \{I_2, -I_2, B_0, B_\pi\}$ y G contiene una reflexión de eje horizontal y otra de eje vertical. Dado que la composición de estas reflexiones es una rotación de ángulo π (pues $B_0 B_\pi = -I_2$), la existencia de mm implica que debe haber un 2. Es por ello que con la notación no abreviada, se le denomina $p2mm$.



- $(pmg) = (p2mg)$ Si con el mismo $J = \{I_2, -I_2, B_0, B_\pi\}$, G tiene una reflexión horizontal pero no una vertical (o viceversa). Entonces B_π debe provenir de una reflexión con deslizamiento de

eje vertical (respectivamente, horizontal).

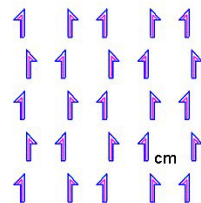


- $(pgg) = (p2gg)$ Si, con el mismo $J = \{I_2, -I_2, B_0, B_\pi\}$, G no contiene reflexiones (y, por tanto, G contiene reflexiones con deslizamiento con ejes horizontales y verticales).

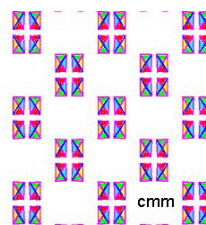


c) L es rectangular centrado. Las transformaciones que respetan el retículo coinciden con el caso anterior. En los casos $J = \{I_2\}$ o $J = \{I_2, -I_2\}$ volvemos a obtener, como en el caso oblicuo y rectangular, grupos del tipo $p1$ y $p2$. Sin embargo, las otras posibilidades para el grupo puntual nos proporcionan dos nuevos grupos (pues el hecho de que el retículo sea distinto hace que el mismo J produzca un G diferente al del caso rectangular):

- $(cm) = (c1m1)$ Si $J = \{I_2, B_0\}$ (análogamente, si $J = \{I_2, B_\pi\}$) entonces B_0 debe provenir de una reflexión o de una reflexión con deslizamiento. Pero de hecho se puede demostrar que proviene de ambas. Es decir que, para este grupo puntual, existen tanto reflexiones como reflexiones con deslizamiento con eje perpendicular a uno de los lados del rectángulo centrado. Por existir reflexiones sin deslizamiento, la notación apropiada es $c1m1$.



- $(cmm) = (c2mm)$ Si $J = \{I_2, -I_2, B_0, B_\pi\}$ y, como antes, B_0 y B_π provienen de reflexiones (sin y con deslizamiento) con ejes perpendiculares a los lados del rectángulo centrado.



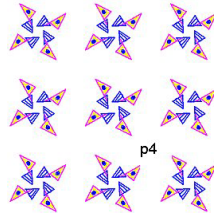
d) L es cuadrado. El grupo de transformaciones ortogonales que lo preservan es

$$D_4 = \langle A_{\frac{\pi}{2}}, B_0 \rangle,$$

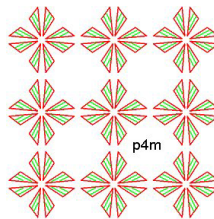
del cual J es un subgrupo. Los únicos casos nuevos son los que incluyen $A_{\frac{\pi}{2}}$ en J (puesto que un retículo cuadrado es un caso particular de un retículo rectangular, el resto de casos vuelve a proporcionarnos los ya conocidos $p1, p2, pm, pg, pmm, pmg, pgg, cm$ y cmm):



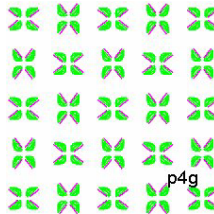
- $(p4) = (p411)$ Si $J = \langle A_{\frac{\pi}{2}} \rangle$. No aparecen reflexiones ni reflexiones con deslizamiento.



- $(p4m) = (p4mm)$ Si $J = \langle A_{\frac{\pi}{2}}, B_0 \rangle = D_4$ y B_0 proviene de una reflexión en G . Existen también, como composición de reflexiones y rotaciones, reflexiones en ejes que forman ángulos de $\pi/4$ con los generadores del retículo, por lo que su notación no abreviada es $p4mm$.

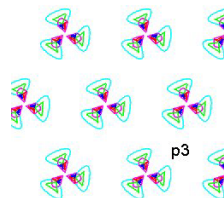


- $(p4g) = (p4gm)$ Si $J = \langle A_{\frac{\pi}{2}}, B_0 \rangle = D_4$ y B_0 no proviene de ninguna reflexión en G . Por lo tanto, G contiene una reflexión con deslizamiento de eje perpendicular a un lado del retículo. El análisis de este caso demuestra que en G existen reflexiones en ejes que forman ángulos de $\pi/4$ con las direcciones del retículo; por eso la notación completa es $p4gm$.



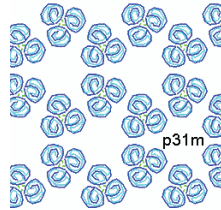
e) L es hexagonal. Entonces J es un subgrupo de D_6 y los únicos casos nuevos surgen cuando J incluye a $A_{\frac{2\pi}{3}}$ o $A_{\frac{\pi}{3}}$. Las otras posibles elecciones de J vuelven a arrojar como resultado grupos de simetría del tipo $p1$, $p2$, cm y cmn , ya que éste es un caso particular de retículo rectangular centrado (en el sentido de que $2b - a$ es perpendicular a a). Analizando los nuevos casos:

- $(p3) = (p311)$ Si $J = \langle A_{\frac{2\pi}{3}} \rangle$. No aparecen reflexiones ni reflexiones con deslizamiento.

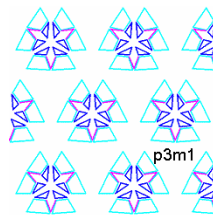


- $(p31m)$ Si $J = \langle A_{\frac{2\pi}{3}}, B_0 \rangle$. Se demuestra entonces que B_0 procede necesariamente de una reflexión, cuyo eje es paralelo a una dirección del retículo. Se puede comprobar que no hay reflexiones

ni reflexiones con deslizamiento que tengan ejes perpendiculares a a o b (de ahí el 1 en $p31m$).

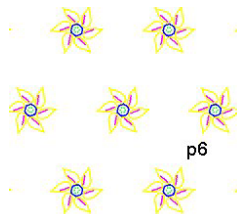


- $(p3m1)$ Si $J = \langle A_{\frac{2\pi}{3}}, B_{\frac{\pi}{3}} \rangle$. De nuevo, se comprueba que $B_{\frac{\pi}{3}}$ procede necesariamente de una reflexión cuyo eje es perpendicular a un generador. El análisis detallado posterior muestra que no hay reflexiones ni reflexiones con deslizamiento con un eje paralelo a a o b , lo que explica el nombre de este caso.

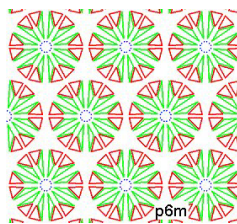


Una nota importante es que en parte de la literatura, y en particular, en [1], los nombres de los casos $p3m1$ y $p31m$ están permutados respecto a los que hemos presentado aquí.

- $(p6) = (p611)$ Si $J = \langle A_{\frac{\pi}{3}} \rangle$. No aparecen reflexiones ni reflexiones con deslizamiento.



- $(p6m) = (p6mm)$ Si $J = \langle A_{\frac{\pi}{3}}, B_0 \rangle = D_6$. Se demuestra que B_0 proviene de una reflexión.



Faltaría demostrar que los 17 grupos son genuinamente diferentes, i.e., que no son isomorfos entre sí. Es claro que si los grupos puntuales correspondientes no son isomorfos, tampoco pueden serlo los grupos cristalográficos. En cuanto a los grupos que tienen grupos puntuales isomorfos, hay que encontrar diferencias en G , como la existencia, número u orientación de los ejes de reflexiones o reflexiones con deslizamiento.

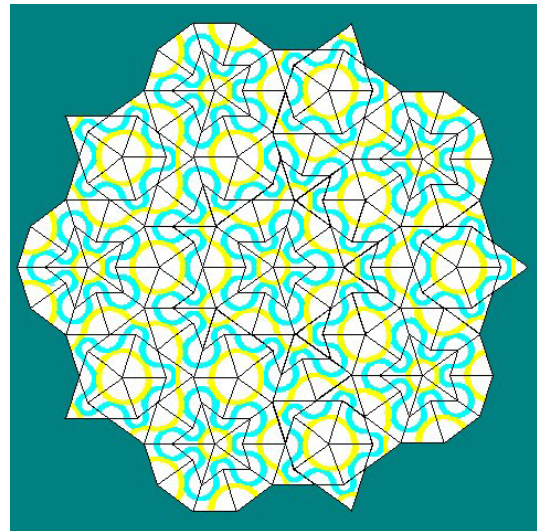
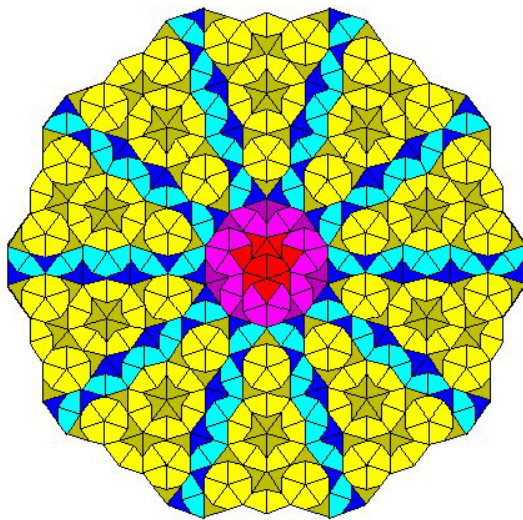
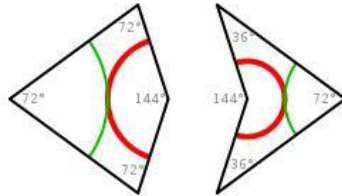
Cerramos el capítulo con una breve mención, mediante un ejemplo, de teselaciones o pavimentos aperiódicos. La idea es comparar con la situación desarrollada hasta ahora:

Definición 2.7.1 Un conjunto de polígonos se llama *aperiódico* si puede usarse para recubrir el plano, pero todos los mosaicos que se forman son no periódicos, es decir, ninguna traslación es una simetría del mosaico. \square



Observación 2.7.2 Las técnicas que deben utilizarse para demostrar que ningún pavimento del plano formado por un conjunto determinado de polígonos puede ser periódico son muy complicadas. Hay muchos problemas que permanecen abiertos (sin resolver) en este ámbito. \square

Ejemplo 2.7.3 Penrose propone varios ejemplos, el más famoso de los cuales tiene sólo dos polígonos, llamados “dardos” y “cometas” (*darts and kites*)



2.8 Problemas

2.8.1 Encontrar todas las simetrías rotacionales de un cubo.

2.8.2 ¿Cuáles de los siguientes conjuntos de matrices con coeficientes reales forman un grupo con el producto de matrices?

a) Las matrices $\begin{pmatrix} a & b \\ b & c \end{pmatrix}$ con $ac \neq b^2$.

b) Las matrices $\begin{pmatrix} a & b \\ c & a \end{pmatrix}$ con $a^2 \neq bc$.

c) Las matrices $\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$ con $ac \neq 0$.

d) Las matrices con coeficientes enteros y determinante no nulo.

2.8.3 Sea (G, \cdot) un grupo. Dados $x, y \in G$ demostrar que

$$(x \cdot y)^{-1} = y^{-1} \cdot x^{-1} \text{ y } (x^{-1})^{-1} = x.$$

2.8.4 Encontrar todos los subgrupos de \mathbb{Z}_4 y de \mathbb{Z}_7 .



2.8.5 Comprobar si son o no subgrupos los siguientes subconjuntos de $GL(2)$:

- a) $H_1 = \{M \in GL(2) : \text{tr}(M) = 2\}$, donde $\text{tr}(M) = a + d$ es la *traza* de la matriz $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.
- b) $H_2 = \{M \in GL(2) : \det(M) \in \mathbb{Q}\}$.
- c) $H_3 = \left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} : a, b \in \mathbb{R} \text{ con } a^2 - b^2 \neq 0 \right\}$.
- d) $H_4 = \left\{ \begin{pmatrix} 0 & b \\ -b & 0 \end{pmatrix} : b \in \mathbb{R} \setminus \{0\} \right\}$.

2.8.6 Razonar si los siguientes subconjuntos constituyen subgrupos:

- a) $\{0, 2, 4, 6, 8\} \subset (\mathbb{Z}_9, +)$
- b) $\{I_2, \sigma^3, \tau, \tau\sigma^3\} \subset D_6$.
- c) $\{2k : k \in \mathbb{Z} \setminus \{0\}\} \subset (\mathbb{R} \setminus \{0\}, \cdot)$.

2.8.7 Sea (G, \cdot) un grupo y $H = \{x \in G : x \cdot y = y \cdot x \ \forall y \in G\}$. ¿Es H subgrupo de G ? Razonar la respuesta.

2.8.8 Hallar el orden de todos los elementos de $\mathbb{Z}_5, \mathbb{Z}_9$ y \mathbb{Z}_{12} .

2.8.9 Calcular el orden de los siguientes elementos:

- a) de 3 en (G, \cdot) siendo $G = \{1, 3, 5, 9, 11, 13\}$ y \cdot la multiplicación módulo 14.
- b) de 4 en (G, \cdot) siendo $G = \{1, 2, 4, 5, 7, 8\}$ y \cdot la multiplicación módulo 9. ¿Es (G, \cdot) cíclico?
- c) de $\sigma\tau$ y σ^2 en $D_5 = \langle \sigma, \tau \rangle$.
- d) de A_π y $A_{\frac{\pi}{3}}$ en $O(2)$.

2.8.10 Comprobar que $\{1, 2, 4, 7, 8, 11, 13, 14\}$ forman un grupo con la multiplicación módulo 15. Hallar el orden de todos sus elementos.

2.8.11 Consideramos el grupo $G = \{1, 3, 5, 7, 9, 11, 13, 15\}$ con la operación producto módulo 16. Estudiar si los siguientes conjuntos son subgrupos de G , y si son cíclicos:

$$A = \{1, 3, 9, 11\}, B = \{1, 5, 7, 13\} \text{ y } C = \{1, 7, 9, 15\}.$$

2.8.12 Demostrar que

$$G = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right\}$$

forma un grupo con el producto de matrices y hallar el orden de sus elementos. ¿Es G isomorfo a \mathbb{Z}_4 ?

2.8.13 Probar que las matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ con $a, b, c, d \in \mathbb{Z}$ y $ad - bc = 1$ forman un grupo con el producto.

Si $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ y $B = \begin{pmatrix} 0 & 1 \\ -1 & -1 \end{pmatrix}$, hallar los órdenes de A, B, AB y BA .

2.8.14 Demostrar que $\{1, 2, 4, 5, 7, 8\}$ es un grupo con la multiplicación módulo 9 y probar que es isomorfo a \mathbb{Z}_6 .

2.8.15 Comprobar que $\{1, 3, 7, 9, 11, 13, 17, 19\}$ es un grupo con la multiplicación módulo 20. ¿Por qué este grupo no es isomorfo a \mathbb{Z}_8 ?



2.8.16 ¿Cuáles de las siguientes aplicaciones definen homomorfismos de $\mathbb{C} \setminus \{0\}$ en $\mathbb{C} \setminus \{0\}$?

- a) $z = a + bi \mapsto \bar{z} = a - bi$.
 b) $z = a + bi \mapsto z^2 = (a^2 - b^2) + 2abi$.
 c) $z = a + bi \mapsto iz = -b + ai$.
 d) $z = a + bi \mapsto |z| = \sqrt{a^2 + b^2}$.

2.8.17 ¿Alguna de las siguientes aplicaciones define un homomorfismo de $GL(2)$ en $GL(2)$?

- a) $A \mapsto A^T$.
 b) $A \mapsto (A^{-1})^T$.
 c) $A \mapsto A^2$.

2.8.18 Sea

$$G = \left\{ \begin{pmatrix} a & b \\ b & a \end{pmatrix} : a, b \in \mathbb{R}, a^2 \neq b^2 \right\},$$

y consideramos el grupo (G, \cdot) donde \cdot es el producto de matrices. Decidir si la aplicación $f: G \rightarrow G$ dada por

$$f \left(\begin{pmatrix} a & b \\ b & a \end{pmatrix} \right) = \begin{pmatrix} a^2 + b^2 & 2ab \\ 2ab & a^2 + b^2 \end{pmatrix}$$

es un homomorfismo de grupos. En caso afirmativo, hallar su núcleo e imagen.

2.8.19 Obtener el núcleo y la imagen de los homomorfismos:

- a) $f: O(2) \rightarrow (\mathbb{R} \setminus \{0\}, \cdot)$ dado por $f(M) = \det(M)$.
 b) $g: \mathbb{Z}_{12} \rightarrow \mathbb{Z}_{12}$ dado por $g(n) = 3n$.

2.8.20 ¿Es $O(2)$ subgrupo normal de $GL(2)$?

2.8.21 Demostrar que $SO(2)$ es un subgrupo de $O(2)$, y comprobar si es o no un subgrupo normal.

2.8.22 Se consideran los conjuntos de matrices

$$D = \left\{ M = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \text{ con } a, b \in \mathbb{R} \setminus \{0\} \right\}$$

y

$$T = \left\{ M = \begin{pmatrix} a & c \\ 0 & b \end{pmatrix} \text{ con } a, b \in \mathbb{R} \setminus \{0\} \text{ y } c \in \mathbb{R} \right\}.$$

- i) Probar que $T \leq GL(2)$ y $D \leq T$. ¿Es D un subgrupo normal de T ?
 ii) Sea $f: T \rightarrow D$ la aplicación definida mediante

$$f \left(\begin{pmatrix} a & c \\ 0 & b \end{pmatrix} \right) = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}.$$

Demostrar que f es un homomorfismo y determinar su núcleo. ¿Es f inyectivo?

2.8.23 Escribir los siguientes movimientos en la forma (v, M) :

- a) Rotación antihoraria de ángulo $\frac{\pi}{6}$ alrededor del punto $(-1, 1)$.
 b) Rotación horaria de ángulo $\frac{\pi}{3}$ alrededor del punto $(1, 2)$.
 c) Reflexión en la recta $x + y + 3 = 0$.



d) Reflexión en la recta $\sqrt{3}y - x = 4$.

2.8.24 Probar que la composición de dos giros de ángulo π es una traslación.

2.8.25 Demostrar que toda isometría inversa (es decir, cuya matriz asociada tiene determinante -1) se puede descomponer como una reflexión seguida de un giro de ángulo π .

2.8.26 Demostrar que una reflexión en m seguida de una reflexión en m' es una traslación si $m \parallel m'$ y una rotación en caso contrario.

2.8.27 Clasificar las siguientes isometrías determinando, en cada caso, los elementos que las caracterizan (centro, ángulo, eje, deslizamiento, ...):

a) $f(x, y) = \left(3 - \sqrt{3} - \frac{1}{2}x - \frac{\sqrt{3}}{2}y, -3 - \sqrt{3} + \frac{\sqrt{3}}{2}x - \frac{1}{2}y \right)$.

b) $g(x, y) = \left(\frac{3}{5}x + \frac{4}{5}y - 14, \frac{4}{5}x - \frac{3}{5}y + 3 \right)$.

c) $h(x, y) = \left(6 - \frac{5}{13}x - \frac{12}{13}y, 4 - \frac{12}{13}x + \frac{5}{13}y \right)$.

2.8.28 Clasificar las isometrías dadas por:

a) $f(x, y) = \left(-1 + \frac{4x}{5} + \frac{3y}{5}, 3 - \frac{3x}{5} + \frac{4y}{5} \right)$

b) $g(x, y) = (-9 - y, -3 - x)$

c) $h(x, y) = \left(\frac{12}{13}x + \frac{5}{13}y - 2, \frac{5}{13}x - \frac{12}{13}y + 10 \right)$,

determinando, en cada caso, los elementos que las caracterizan (centro, ángulo, eje, deslizamiento, ...).

2.8.29 Escribir las siguientes isometrías en notación (v, M) :

a) f , la rotación de centro $(1, 3)$ y ángulo $\frac{\pi}{2}$.

b) g , la reflexión con deslizamiento de eje $x - y - 2 = 0$ y deslizamiento $(1, 1)$.

Calcular y clasificar la isometría $g \circ f$.

2.8.30 Determinar la expresión en la forma (v, M) de las siguientes isometrías del plano:

a) La rotación de centro $(3, 0)$ y ángulo $\frac{\pi}{6}$.

b) La reflexión con deslizamiento de eje la recta $y + \frac{\sqrt{3}}{3}x = \sqrt{3}$ y vector de deslizamiento $b = (\sqrt{3}, -1)$.

2.8.31 a) Expresar en forma (v, M) la rotación f de ángulo $\frac{\pi}{3}$ alrededor del punto $(2, 2)$.

b) Clasificar la isometría

$$g(x, y) = \left(\sqrt{3} + \frac{x}{2} + \frac{\sqrt{3}}{2}y, 1 + \frac{\sqrt{3}}{2}x - \frac{y}{2} \right),$$

determinando los elementos que la caracterizan.

c) ¿Puede ser $g \circ f$ una rotación? ¿Y $f \circ g$?



2.8.32 Sea f el giro de centro $(1, 1)$ y ángulo $\frac{\pi}{2}$, y g la reflexión con deslizamiento con eje $y = x + 1$ y vector de deslizamiento $(1, 1)$. Escribir la isometría $g \circ f$ en la notación (v, M) , clasificarla y determinar los elementos geométricos que la caracterizan.

2.8.33 Una rotación f de centro $(-2, -2)$ y ángulo $\theta = -\frac{\pi}{2}$ se compone con una isometría desconocida g , de forma que se obtiene $h = g \circ f = (u, N)$ siendo

$$u = (4, 2) \text{ y } N = \begin{pmatrix} \frac{3}{5} & -\frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{pmatrix}.$$

- Escribir f en forma (v, M) .
- Clasificar h y obtener sus elementos principales.
- Hallar g .

2.8.34 Se consideran las isometrías $f(x, y) = (1 - y, x - 1)$ y $g(x, y) = (2 + y, x)$.

- Clasificarlas y describir sus elementos geométricos.
- Escribir $g \circ f$ en notación (v, M) .
- Clasificar el retículo generado por $u_1 = (1, 1)$, $u_2 = (1, -1)$. ¿Es $g \circ f$ una simetría del retículo?

2.8.35 a) Sea f la reflexión respecto a la recta $-2x + 3y = 4$ con deslizamiento $b = (3, 2)$. Encontrar sus ecuaciones.

- Clasificar y describir los elementos geométricos de la isometría

$$g(x, y) = (2 + x, -2 - y).$$

- Clasificar la isometría $f \circ g$ y determinar sus elementos geométricos.

2.8.36 Expresar en forma (v, M) la rotación f de ángulo $\frac{\pi}{4}$ alrededor del punto $(1, 1)$ y la reflexión g con respecto a la recta $x + y = 2$. Clasificar la isometría $f \circ g$ y encontrar sus elementos geométricos relevantes.

2.8.37 Escribir en forma (v, M) la reflexión f respecto a la recta $y = 2 - x$, y la reflexión con deslizamiento g con eje $x = 1$ y deslizamiento $w = (0, 2)$. Clasificar $g \circ f$ y dar sus elementos geométricos.

2.8.38 Sean f una rotación de centro c y ángulo θ y g una rotación de centro c' y ángulo θ' . ¿Puede ser $g \circ f$ una reflexión?

2.8.39 ¿Cuál es el resultado de un giro de ángulo π seguido de una traslación?

2.8.40 Mostrar que dos reflexiones conmutan si y sólo si sus ejes coinciden o son perpendiculares.

2.8.41 Probar que un giro de ángulo π conmuta con una reflexión si y sólo si su centro está en el eje de ésta.

2.8.42 Probar que una traslación conmuta con una reflexión si y sólo si su vector es paralelo al eje de ésta.

2.8.43 Demostrar que el conjunto de todas las rotaciones del plano alrededor de un punto dado P es un grupo con la composición de funciones. ¿Es esto cierto si consideramos todas las reflexiones respecto a rectas que pasan por P ? ¿Y si consideramos el conjunto de las rotaciones en torno a P y las reflexiones respecto a rectas por P ?

2.8.44 Encontrar los subgrupos de D_5 . Decidir cuáles de ellos son normales.



2.8.45 Consideremos un hexágono regular, y tracemos una diagonal. Hacer una lista de las simetrías del hexágono que dejan fijos todos los puntos de la diagonal, y otra de las que llevan la diagonal sobre sí misma. Demostrar que ambos conjuntos son subgrupos del grupo D_6 de las simetrías del hexágono.

2.8.46 Dibujar los retículos generados por los siguientes pares de vectores, indicando de qué tipo son

a) $\mathbf{a} = (-1, -\sqrt{3})$ y $\mathbf{b} = (1, -\sqrt{3})$.

b) $\mathbf{a} = (1, 0)$ y $\mathbf{b} = (2, -4)$.

c) $\mathbf{a} = (-2, 0)$ y $\mathbf{b} = (-1, 3)$.

2.8.47 Sea G un grupo cristalográfico con retículo cuadrado. ¿Cuáles son las distintas posibilidades para el grupo puntual de G ?

2.8.48 Si un mosaico no posee simetrías que sean rotaciones ni reflexiones, ¿cuál puede ser su grupo cristalográfico?

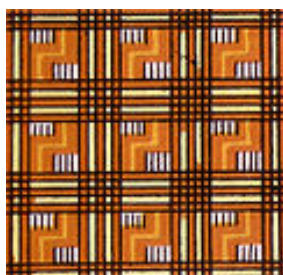
2.8.49 Sea G un grupo cristalográfico plano cuyo grupo de traslaciones está generado por los vectores $u = (2, 0)$ y $v = (3, 2)$.

a) Clasificar el retículo asociado al grupo G , determinando generadores de norma mínima.

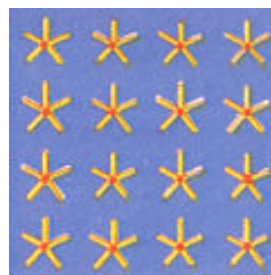
b) Decidir cuáles pueden ser los grupos puntuales de G .

2.8.50 ¿Qué matrices de rotación pueden ser elementos del grupo puntual J de un mosaico cuyo retículo es cuadrado?

2.8.51 Analizar los patrones siguientes para averiguar su grupo cristalográfico



Hawaii



Egipto 1



Egipto 2

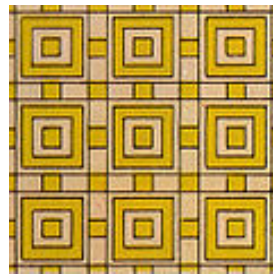


La Alhambra





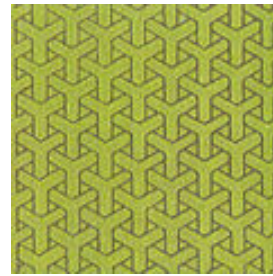
Tapiz persa



Asiria



Linóleo (EEUU)



China

2.8.52 “Jugar” con el programa Kali (en <http://www.geom.uiuc.edu/java/Kali/welcome.html>: versión java para usar desde el navegador; en <http://www.geometrygames.org>: versión ejecutable para Windows/Linux o Mac). ¿Podrías deducir la correspondencia entre las notaciones de los grupos cristalográficos que usa el programa (esa notación se llama de “orbifolds”) y nuestras notaciones, a partir de las simetrías de los mosaicos que se crean?

3 Trigonometría plana y esférica. Aplicaciones

3.1 Introducción y contenidos

Etimológicamente *trigonometría* significa “medida de los triángulos”. La trigonometría es una rama de la matemática que se encarga de estudiar las relaciones entre los lados y los ángulos de los triángulos, así como de las características y aplicaciones de las funciones trigonométricas de ángulos. En términos generales, la trigonometría se subdivide en:

- a) *Trigonometría plana*, que estudia las figuras contenidas en un plano.
- b) *Trigonometría esférica*, que estudia el establecimiento de las propiedades y relaciones que satisfacen los elementos de triángulos definidos en la superficie de una esfera mediante arcos de círculos máximos, así como de la resolución de los mismos.

A partir de una revisión de conceptos y relaciones de la trigonometría plana para su aplicación en problemas concretos de navegación y geodesia, se introducen los principales conceptos de la geometría de la esfera para pasar posteriormente a hacer una comparación entre las geometrías de la esfera y del plano e introducir el concepto de *triángulo esférico*, sus elementos y las principales expresiones que relacionan entre sí dichos elementos. Considerando la superficie de la Tierra como una esfera, se introducen conceptos geodésicos (como *meridiano*, *paralelo*, *polos*, *latitud* y *longitud*), de Astronomía (como los *sistemas de coordenadas astronómicas horizontal*, *ecuatorial horario* y *ecuatorial absoluto*) y de navegación (como *rumbo* y *derrota*) y se plantean diversos problemas de navegación y de Astronomía de posición para cuya resolución se utilizarán triángulos esféricos.

Finalmente, y considerando el elipsoide como aproximación de la figura de la Tierra, se presenta el *sistema de coordenadas geodésico* que permite fijar la posición de un observador sobre la superficie terrestre. Finalizamos el capítulo describiendo el problema de posicionamiento por satélites y en particular el posicionamiento con GPS.

En todo el capítulo se considerará que los ángulos se miden, como es tradicional en el contexto en que nos encontramos, en grados sexagesimales en lugar de en radianes.

3.2 Trigonometría plana

3.2.1 Revisión

Funciones trigonométricas: sen, cos, tan, csc, sec, cot

Sea $\{O; X, Y\}$ un sistema de coordenadas rectangular en el plano y P un punto del plano, cuya posición respecto de este sistema de coordenadas viene dada por (x, y) siendo x la *abscisa* e y la *ordenada* del punto P . Sea $r = \sqrt{x^2 + y^2}$ la distancia del punto P al origen O y α el ángulo entre el eje X y la dirección \vec{OP} (véase la Figura 3.1).

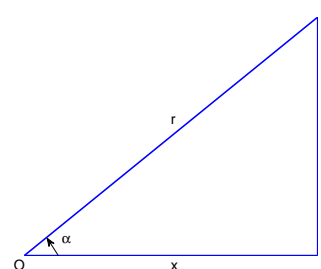


Figura 3.1: Triángulo rectángulo.

Se definen seis funciones trigonométricas de α en términos de la abscisa, ordenada y distancia de P :

Función	Definición	Función	Definición
<i>seno</i>	$\text{sen } \alpha = \frac{\text{ordenada}}{\text{distancia}} = \frac{y}{r}$	<i>cosecante</i>	$\text{csc } \alpha = \frac{\text{distancia}}{\text{ordenada}} = \frac{r}{y}$
<i>coseno</i>	$\text{cos } \alpha = \frac{\text{abscisa}}{\text{distancia}} = \frac{x}{r}$	<i>secante</i>	$\text{sec } \alpha = \frac{\text{distancia}}{\text{abscisa}} = \frac{r}{x}$
<i>tangente</i>	$\text{tan } \alpha = \frac{\text{ordenada}}{\text{abscisa}} = \frac{y}{x}$	<i>cotangente</i>	$\text{cot } \alpha = \frac{\text{abscisa}}{\text{ordenada}} = \frac{x}{y}$



Valores de las funciones trigonométricas para 0° , 90° , 180° y 270°

Ángulo α	$\text{sen } \alpha$	$\text{cos } \alpha$	$\text{tan } \alpha$	$\text{csc } \alpha$	$\text{sec } \alpha$	$\text{cot } \alpha$
0°	0	1	0	$\pm\infty$	1	$\pm\infty$
90°	1	0	$\pm\infty$	1	$\pm\infty$	0
180°	0	-1	0	$\pm\infty$	-1	$\pm\infty$
270°	-1	0	$\pm\infty$	-1	$\pm\infty$	0

Relaciones fundamentales

$\text{sen } \alpha = \frac{1}{\text{csc } \alpha}$	$\text{cos } \alpha = \frac{1}{\text{sec } \alpha}$	$\text{tan } \alpha = \frac{1}{\text{cot } \alpha}$
$\text{csc } \alpha = \frac{1}{\text{sen } \alpha}$	$\text{sec } \alpha = \frac{1}{\text{cos } \alpha}$	$\text{cot } \alpha = \frac{1}{\text{tan } \alpha}$
$\text{tan } \alpha = \frac{\text{sen } \alpha}{\text{cos } \alpha}$	$\text{cot } \alpha = \frac{\text{cos } \alpha}{\text{sen } \alpha}$	

Identidades

$\text{sen}^2 \alpha + \text{cos}^2 \alpha = 1$	$1 + \text{tan}^2 \alpha = \text{sec}^2 \alpha$	$1 + \text{cot}^2 \alpha = \text{csc}^2 \alpha$
---	---	---

Funciones trigonométricas de dos ángulos

SUMA DE ÁNGULOS:

$$\begin{aligned} \text{sen}(\alpha + \beta) &= \text{sen } \alpha \text{cos } \beta + \text{cos } \alpha \text{sen } \beta \\ \text{cos}(\alpha + \beta) &= \text{cos } \alpha \text{cos } \beta - \text{sen } \alpha \text{sen } \beta \\ \text{tan}(\alpha + \beta) &= \frac{\text{tan } \alpha + \text{tan } \beta}{1 - \text{tan } \alpha \text{tan } \beta} \end{aligned}$$

Observación 3.2.1 Veamos, a modo de ejemplo, cómo demostrar la primera de estas fórmulas para el caso particular en el que $0 \leq \alpha, \beta, \alpha + \beta \leq 90^\circ$ (las demostraciones para ángulos en otros cuadrantes son análogas, y el resto de identidades de la lista, así como las que vendrán a continuación, son consecuencias más o menos directas de la primera).

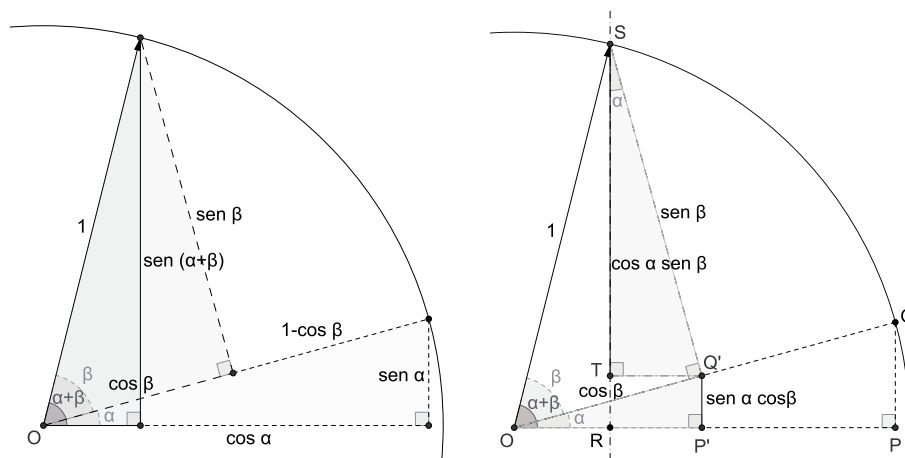


Figura 3.2: Demostración gráfica de la fórmula del seno de la suma.

En la primera Figura de 3.2 hemos construido tres triángulos rectángulos de hipotenusa 1, con un vértice común en un punto O , y cuyos ángulos en O son α , β , y $\alpha + \beta$. Para calcular $\text{sen}(\alpha + \beta)$ en función de las razones trigonométricas de α y β , descomponemos el lado \overline{RS} (cuya longitud es, justamente, $\text{sen}(\alpha + \beta)$) en dos segmentos, $\overline{RS} = \overline{RT} + \overline{TS}$ (véase la segunda Figura en 3.2). Ahora, $\overline{RT} = \overline{P'Q'}$ = $\text{sen } \alpha \cos \beta$, por ser el cateto opuesto al ángulo α en el triángulo $OP'Q'$, cuya hipotenusa $\overline{OQ'}$ mide $\cos \beta$. Por otra parte, el ángulo γ del triángulo TSQ' es justamente α , por ser $\overline{TS} \perp \overline{OP}$ y $\overline{Q'S} \perp \overline{OQ}$. En consecuencia, $\overline{TS} = \cos \alpha \text{sen } \beta$, por ser el cateto contiguo al ángulo α en el triángulo TSQ' , cuya hipotenusa $\overline{SQ'}$ mide $\text{sen } \beta$. En consecuencia, se obtiene la fórmula buscada.

DIFERENCIA DE ÁNGULOS:

$$\begin{aligned}\text{sen}(\alpha - \beta) &= \text{sen } \alpha \cos \beta - \cos \alpha \text{sen } \beta \\ \cos(\alpha - \beta) &= \cos \alpha \cos \beta + \text{sen } \alpha \text{sen } \beta \\ \tan(\alpha - \beta) &= \frac{\tan \alpha - \tan \beta}{1 + \tan \alpha \tan \beta}\end{aligned}$$

ÁNGULO DOBLE:

$$\begin{aligned}\text{sen } 2\alpha &= 2 \text{sen } \alpha \cos \alpha \\ \cos 2\alpha &= \cos^2 \alpha - \text{sen}^2 \alpha = 1 - 2 \text{sen}^2 \alpha = 2 \cos^2 \alpha - 1 \\ \tan 2\alpha &= \frac{2 \tan \alpha}{1 - \tan^2 \alpha}\end{aligned}$$

ÁNGULO MITAD:

$$\begin{aligned}\text{sen } \frac{\alpha}{2} &= \pm \sqrt{\frac{1 - \cos \alpha}{2}} \\ \cos \frac{\alpha}{2} &= \pm \sqrt{\frac{1 + \cos \alpha}{2}} \\ \tan \frac{\alpha}{2} &= \pm \sqrt{\frac{1 - \cos \alpha}{1 + \cos \alpha}} = \frac{\text{sen } \alpha}{1 + \cos \alpha} = \frac{1 - \cos \alpha}{\text{sen } \alpha}\end{aligned}$$

PRODUCTOS DE SENOS Y COSENOS:

$$\begin{aligned}\text{sen } \alpha \cos \beta &= \frac{1}{2} (\text{sen}(\alpha + \beta) + \text{sen}(\alpha - \beta)) \\ \cos \alpha \text{sen } \beta &= \frac{1}{2} (\text{sen}(\alpha + \beta) - \text{sen}(\alpha - \beta)) \\ \cos \alpha \cos \beta &= \frac{1}{2} (\cos(\alpha + \beta) + \cos(\alpha - \beta)) \\ \text{sen } \alpha \text{sen } \beta &= -\frac{1}{2} (\cos(\alpha + \beta) - \cos(\alpha - \beta))\end{aligned}$$

SUMA Y DIFERENCIA DE SENOS Y COSENOS:

$$\begin{aligned}\text{sen } A + \text{sen } B &= 2 \text{sen } \frac{A+B}{2} \cos \frac{A-B}{2} \\ \text{sen } A - \text{sen } B &= 2 \cos \frac{A+B}{2} \text{sen } \frac{A-B}{2} \\ \cos A + \cos B &= 2 \cos \frac{A+B}{2} \cos \frac{A-B}{2} \\ \cos A - \cos B &= -2 \text{sen } \frac{A+B}{2} \text{sen } \frac{A-B}{2}\end{aligned}$$



Valores de las funciones trigonométricas para 30° , 45° y 60°

Ángulo α	$\text{sen } \alpha$	$\text{cos } \alpha$	$\text{tan } \alpha$	$\text{csc } \alpha$	$\text{sec } \alpha$	$\text{cot } \alpha$
30°	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{3}$	2	$\frac{2\sqrt{3}}{3}$	$\sqrt{3}$
45°	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$	1	$\sqrt{2}$	$\sqrt{2}$	1
60°	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$	$\sqrt{3}$	$\frac{2\sqrt{3}}{3}$	2	$\frac{\sqrt{3}}{3}$

Valores de las funciones trigonométricas para $-\alpha$, $90^\circ \pm \alpha$, $180^\circ \pm \alpha$, $270^\circ \pm \alpha$ y $k \cdot 360^\circ \pm \alpha$ siendo α un ángulo del primer cuadrante

Ángulo	sen	cos	tan	csc	sec	cot
$-\alpha$	$-\text{sen } \alpha$	$\text{cos } \alpha$	$-\text{tan } \alpha$	$-\text{csc } \alpha$	$\text{sec } \alpha$	$-\text{cot } \alpha$
$90^\circ \pm \alpha$	$\text{cos } \alpha$	$\mp \text{sen } \alpha$	$\mp \text{cot } \alpha$	$\text{sec } \alpha$	$\mp \text{csc } \alpha$	$\mp \text{tan } \alpha$
$180^\circ \pm \alpha$	$\mp \text{sen } \alpha$	$-\text{cos } \alpha$	$\pm \text{tan } \alpha$	$\mp \text{csc } \alpha$	$-\text{sec } \alpha$	$\pm \text{cot } \alpha$
$270^\circ \pm \alpha$	$-\text{cos } \alpha$	$\pm \text{sen } \alpha$	$\mp \text{cot } \alpha$	$-\text{sec } \alpha$	$\pm \text{csc } \alpha$	$\mp \text{tan } \alpha$
$k \cdot 360^\circ \pm \alpha$ ($k \in \mathbb{Z}$)	$\pm \text{sen } \alpha$	$\text{cos } \alpha$	$\pm \text{tan } \alpha$	$\pm \text{csc } \alpha$	$\text{sec } \alpha$	$\pm \text{cot } \alpha$

Triángulos planos. Leyes y fórmulas generales

Consideremos un triángulo de lados a , b , c y sean A , B , C los ángulos opuestos a los lados a , b , c , respectivamente, como se muestra en la Figura 3.3:

TEOREMA DEL SENO:

$$\frac{a}{\text{sen } A} = \frac{b}{\text{sen } B} = \frac{c}{\text{sen } C}$$

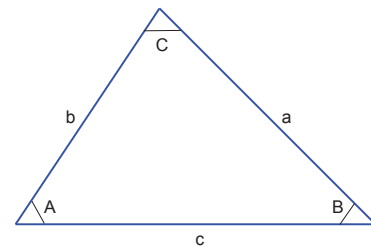


Figura 3.3: Triángulo de lados a , b y c .

TEOREMA DEL COSENO:

$$\begin{aligned} a^2 &= b^2 + c^2 - 2bc \cos A \\ b^2 &= c^2 + a^2 - 2ca \cos B \\ c^2 &= a^2 + b^2 - 2ab \cos C \end{aligned}$$

FÓRMULAS DE PROYECCIÓN:

$$\begin{aligned} a &= b \cos C + c \cos B \\ b &= c \cos A + a \cos C \\ c &= a \cos B + b \cos A \end{aligned}$$



FÓRMULAS DE MOLLWEIDE:

$\frac{a+b}{c} = \frac{\cos \frac{A-B}{2}}{\operatorname{sen} \frac{C}{2}}$	$\frac{a-b}{c} = \frac{\operatorname{sen} \frac{A-B}{2}}{\cos \frac{C}{2}}$
$\frac{b+c}{a} = \frac{\cos \frac{B-C}{2}}{\operatorname{sen} \frac{A}{2}}$	$\frac{b-c}{a} = \frac{\operatorname{sen} \frac{B-C}{2}}{\cos \frac{A}{2}}$
$\frac{c+a}{b} = \frac{\cos \frac{C-A}{2}}{\operatorname{sen} \frac{B}{2}}$	$\frac{c-a}{b} = \frac{\operatorname{sen} \frac{C-A}{2}}{\cos \frac{B}{2}}$

3.2.2 Aplicaciones de la trigonometría plana a la navegación y Geodesia

Los conceptos y relaciones de la trigonometría plana presentados en la sección anterior se pueden aplicar a problemas concretos de navegación y geodesia, entre los que podemos citar:

- La determinación de la distancia de un punto a un objeto inaccesible, conociendo una distancia base y midiendo dos ángulos.
- La determinación de la altura de un objeto sobre una base inaccesible y de altura desconocida (por ejemplo, la altura de una torre en una colina), midiendo desde dos puntos, cuya distancia es conocida, los ángulos que forman con la horizontal los puntos inferior y superior del objeto.
- El *problema de Snellius–Pothenot*, en el que se determina la distancia de un punto M a tres puntos a partir de la distancia entre estos tres puntos y de las medidas angulares efectuadas desde ellos a M .
- El *problema de Hansen*, en el que se determina, conocida la distancia entre dos puntos A_1 y A_2 , la distancia entre dos puntos M y M' y las distancias de estos a A_1 y A_2 ; para ello se utilizan las medidas angulares efectuadas desde estos dos puntos a los puntos M y M' .
- El *principio de triangulación* para la medida de un arco de meridiano, aplicando el teorema del seno a una cadena de triángulos, estando los vértices de dichos triángulos situados a ambos lados de un segmento de meridiano.

Estas aplicaciones de la trigonometría plana serán tratadas en la sección de problemas de este capítulo.

3.3 Trigonometría esférica

3.3.1 Geometría esférica

Definición 3.3.1

- Una *esfera* es un subconjunto de \mathbb{R}^3 formado por los puntos cuya distancia a un punto fijo (*centro*) es menor o igual que una cierta cantidad r (*radio*). Es decir, es el conjunto de \mathbb{R}^3 definido por

$$E = \{(x, y, z) \in \mathbb{R}^3 : (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 \leq r^2\}$$

($r > 0$ es el *radio* de la esfera y (x_0, y_0, z_0) el *centro* de la misma).

- La *superficie de una esfera* de centro en el punto (x_0, y_0, z_0) y radio $r > 0$ es el subconjunto de \mathbb{R}^3 definido por

$$S = \{(x, y, z) \in \mathbb{R}^3 : (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2\}. \quad \square$$



Observación 3.3.2 Cuatro puntos no coplanarios del espacio euclídeo \mathbb{R}^3 definen una superficie esférica y sólo una. \square

Proposición 3.3.3 La intersección de un plano con una esfera de radio r es un círculo que tiene por radio $r' = \sqrt{r^2 - d^2}$ siendo d la distancia del centro de la esfera al plano de corte.

DEMOSTRACIÓN. Ejercicio. \square

Definición 3.3.4

- a) Un círculo máximo es la intersección de una esfera con un plano que contiene el centro de dicha esfera; la palabra “máximo” se refiere a que se trata de un círculo de radio r (véase la Proposición 3.3.3), que es el mayor posible en la esfera. Con un abuso de lenguaje, es a la circunferencia que delimita un círculo máximo a la que se denomina *círculo máximo*. Si el plano no pasa por el centro de la esfera entonces éste la interseca en un círculo menor; análogamente, se usa la expresión *círculo menor* para denominar a la circunferencia que lo delimita.
- b) Cada círculo máximo define dos puntos sobre la esfera que se llaman sus *polos*: son los extremos del diámetro perpendicular al círculo máximo (dicho de otro modo: el círculo máximo está dado por la intersección de un plano –que pasa por el centro– con la esfera; los polos son los puntos de intersección de la recta perpendicular al plano por el centro con la esfera). \square

Ejemplo 3.3.5 Los polos del ecuador son, precisamente, los polos norte y sur. \square

Definición 3.3.6

- a) Diremos que dos puntos son antipodales si son extremos de un mismo diámetro.
- b) Dados dos puntos A y B de la esfera, existe un círculo máximo (único, salvo que los puntos sean antipodales, en cuyo caso hay infinitos) que pasa por ambos (basta considerar el plano que pasa por A, B y el centro de la esfera). Se define la *distancia esférica* entre dos puntos A y B como el menor de los arcos de extremos A y B del círculo máximo que pasa por ellos. \square

Observación 3.3.7

- a) La distancia esférica entre cada punto de cada círculo máximo y sus polos es 90° .
- b) La longitud de arco ℓ correspondiente a una distancia esférica α en una esfera de radio r es

$$\ell = \alpha r \frac{\pi}{180}.$$

En el caso particular de que $r = 1$, se tiene que $\ell = \alpha \frac{\pi}{180}$. \square

Proposición 3.3.8 Un arco de círculo máximo es la distancia más corta sobre la esfera entre dos puntos de la esfera.

DEMOSTRACIÓN. Vamos a demostrar únicamente que, dados dos puntos en la esfera, el arco de circunferencia que proporciona la menor distancia entre ellos es el que sigue el círculo máximo que los une. Basta probar que, para una longitud de cuerda fija $2c$, la longitud de arco más corta es la correspondiente al radio más grande o, lo que es lo mismo, al arco 2β más pequeño (véase la Figura 3.4); esto es, que la longitud de un arco de cuerda fija es una función creciente del arco.

Ahora bien, expresando β en radianes, puesto que la longitud de arco viene dada por

$$\ell(\beta) = 2\beta r = 2\beta \frac{c}{\sin \beta} = 2c \frac{\beta}{\sin \beta}$$

derivando esta función de β obtenemos

$$\ell'(\beta) = 2c \frac{\sin \beta - \beta \cos \beta}{\sin^2 \beta}.$$

Es fácil comprobar (hágase) que, si $0 < x \leq \frac{\pi}{2}$, entonces $\tan x > x$. Esto prueba que ℓ' es positiva en $(0, \frac{\pi}{2}]$, es decir, la función ℓ es creciente en ese intervalo. \square



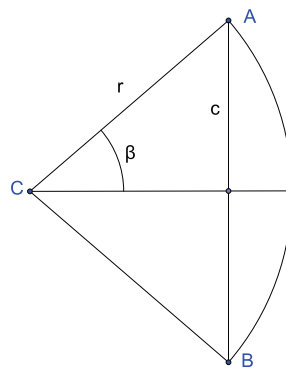


Figura 3.4: Longitud de arco.

Observación 3.3.9 (Comparación de la geometría en la esfera y la geometría en el plano)

La Proposición 3.3.8 indica que los círculos máximos juegan en la esfera el papel que juegan las rectas en el plano (son “el camino más corto entre dos puntos”). También se pueden definir los triángulos sobre la esfera (lo haremos más adelante; véase la Figura 3.5) como el resultado de unir tres puntos “no alineados” (es decir, que no estén en un círculo máximo) mediante “rectas” (o sea, mediante círculos máximos). Los ángulos de un triángulo esférico serán los ángulos que forman los planos que contienen a cada lado (por ser un arco de círculo máximo, cada lado está en un plano que pasa por los vértices y el centro de la esfera). Esto permite definir una geometría (conocida como *geometría esférica*) que tiene algunas características especiales.

Veamos, en primer lugar, algunas cuestiones relativas a la geometría en el plano. La geometría usual en el plano es la conocida como geometría euclídea y se basa en los cinco postulados (o axiomas) que Euclides estableció en sus *Elementos* en el siglo III a.c. Los postulados se suelen enunciar como:

- 1) Por dos puntos se puede trazar una y sólo una recta que los une.
- 2) Cualquier segmento puede prolongarse indefinidamente.
- 3) Se pueden trazar circunferencias de centro y radio arbitrarios.
- 4) Todos los ángulos rectos son iguales.
- 5) Por un punto exterior a una recta se puede trazar una y sólo una paralela a dicha recta.

El quinto postulado siempre fue polémico, y mucha gente (al parecer, el propio Euclides) intentó demostrarlo a partir de los cuatro anteriores. Estos intentos llevaron a la aparición, en el siglo XVIII, de las geometrías no euclídeas, tras probarse que el quinto postulado no es consecuencia de los cuatro anteriores, pero su negación tampoco los contradice. La geometría esférica es un ejemplo de geometría no euclídea: en ella se cumplen los cuatro primeros postulados, pero no se cumple el quinto, puesto que dos rectas (es decir, dos círculos máximos) siempre se cortan en dos puntos (los extremos del diámetro intersección de los dos planos que los contienen) y, por tanto, no existen rectas paralelas. A continuación se recogen algunas de las diferencias entre la geometría en la esfera y la geometría en el plano:

PLANO	ESFERA
Dos rectas en el plano son paralelas o se cortan en <i>un único</i> punto	Dos círculos máximos se cortan siempre en <i>dos</i> puntos de la superficie de la esfera
La suma de los ángulos de un triángulo plano es <i>igual</i> a dos rectos (180°)	La suma de los ángulos de un triángulo esférico es variable y siempre <i>mayor</i> que dos rectos
Un triángulo plano <i>solamente</i> puede tener <i>un</i> ángulo recto	Un triángulo esférico puede tener <i>dos</i> o <i>tres</i> ángulos rectos
Dos perpendiculares a una recta son <i>paralelas</i>	Dos círculos máximos perpendiculares a otro <i>se cortan en los polos</i> de éste
El lugar geométrico de los puntos que equidistan de una recta está formado por dos rectas paralelas a la dada	El lugar geométrico de los puntos que equidistan de un círculo máximo está formado por <i>dos círculos menores</i>



3.3.2 Triángulos esféricos

Definición 3.3.10 Sea E una esfera de centro O y radio $r > 0$ como la de la Figura 3.5.

- a) Sean A , B y C tres puntos de la esfera E que no están sobre el mismo círculo máximo. Uniéndolos por arcos de círculos máximos BC , CA y AB (de longitud $< 180^\circ$), la figura resultante es un *triángulo esférico*. Los tres puntos definen un plano que no pasa por el centro de la esfera.

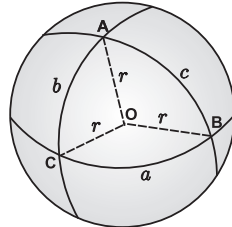


Figura 3.5: Triángulo esférico.

- b) En todo triángulo esférico de vértices A , B y C sobre la esfera E , pueden distinguirse 6 ángulos:
- A , B y C : *ángulos diedros* entre los planos que contienen a los círculos máximos que se cortan en dichos puntos.
 - a , b y c : distancias esféricas sobre los arcos BC , CA y AB , respectivamente. Es decir, a , b y c son los lados del triángulo esférico (opuestos a los ángulos A , B y C) que, de hecho, se miden en grados sexagesimales.

La *trigonometría esférica* trata de las relaciones trigonométricas que existen entre los seis elementos, tres lados y tres ángulos, de un triángulo esférico.

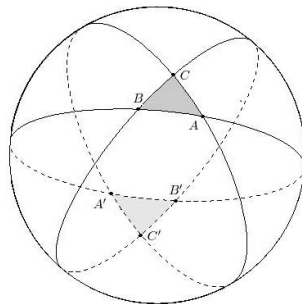


Figura 3.6: Triángulo esférico y triángulo antipodal.

- c) Si A' , B' y C' son los puntos antipodales de los vértices A , B y C del triángulo esférico, entonces los triángulos en que queda dividida la esfera son (véase la Figura 3.6):
- ABC : triángulo esférico dado.
 - ABC' , $AB'C$ y $A'BC$: *triángulos adyacentes* al ABC (por tener con él un lado común).
 - $A'B'C'$: *triángulo antipodal* (o *simétrico*) al dado respecto al centro.
 - $A'B'C$, $A'BC'$ y $AB'C'$: triángulos adyacentes al $A'B'C'$, o antipodales de ABC' , $AB'C$ y $A'BC$, respectivamente, respecto del centro.
- d) En el triángulo esférico ABC se definen las cantidades:
- Exceso esférico*: $\epsilon = A + B + C - 180^\circ$.
 - Perímetro* del triángulo esférico: $p = 2s = a + b + c$. La cantidad s se denomina *semiperímetro*.
 - Defecto esférico*: $d = 360^\circ - p$. \square



Definición 3.3.11 Se denomina *triángulo polar* relativo al triángulo esférico ABC de lados a , b y c , al triángulo esférico cuyos vértices son los polos \bar{A} , \bar{B} y \bar{C} de los lados del triángulo (para cada lado del triángulo se escoge el polo que está en el hemisferio exterior al triángulo original según se muestra en la Figura 3.7). \square

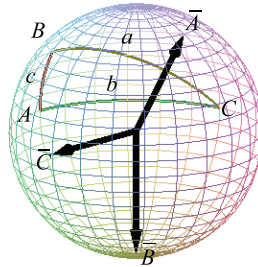


Figura 3.7: Triángulo polar.

Observación 3.3.12 A la vista de la Figura 3.8 es fácil ver que

$$\bar{a} = 180^\circ - A$$

y lo mismo ocurre con los otros dos lados del triángulo polar ($\bar{b} = 180^\circ - B$ y $\bar{c} = 180^\circ - C$). \square

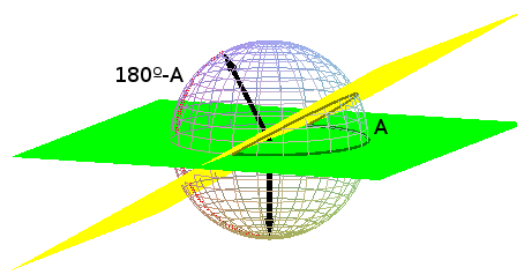


Figura 3.8: Triángulo polar.

Proposición 3.3.13 El triángulo polar del triángulo polar de un triángulo esférico ABC es el triángulo original ABC .

DEMOSTRACIÓN. Sin pérdida de generalidad, podemos suponer que el radio de la esfera es $r = 1$. El lado del triángulo polar correspondiente al vértice A , \bar{a} , es el arco del círculo máximo que une los polos de los lados b y c . Por tanto, sobre la esfera, \bar{a} es el arco de círculo máximo entre los vectores unitarios normales a los planos generados por b y c . Así, el polo (\bar{A}) de \bar{a} es un vector unitario normal a ambos vectores unitarios normales, es decir, (\bar{A}) está a la vez en los planos generados por b y c (y, obviamente, en la esfera), por lo que $(\bar{A}) = b \cap c = A$ (de hecho, podría ser $(\bar{A}) = A$ o $(\bar{A}) = A' = -A$ pero, por escoger en cada caso el vector normal en el hemisferio exterior al triángulo, puede comprobarse gráficamente que debe ser $(\bar{A}) = A$). \square

Corolario 3.3.14 En el triángulo polar se verifica que $\bar{A} = 180^\circ - a$, $\bar{B} = 180^\circ - b$ y $\bar{C} = 180^\circ - c$.

DEMOSTRACIÓN. Basta pensar en la propiedad $\bar{a} = 180^\circ - A$ aplicada al triángulo polar $\bar{A}\bar{B}\bar{C}$. \square

Definición 3.3.15 Un *huso* es la región de la esfera delimitada por dos arcos de círculos máximos (véase la Figura 3.9) que, como ya se ha dicho, se cortan en puntos antipodales. \square



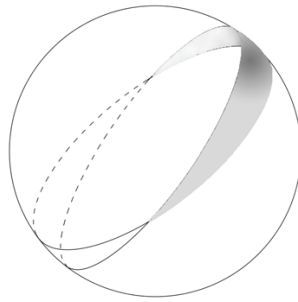


Figura 3.9: Huso.

Proposición 3.3.16 *El área de un huso en una esfera de radio r es $2\alpha \frac{\pi}{180} r^2$, siendo α el ángulo que forman los dos planos determinados por ambos círculos máximos.*

DEMOSTRACIÓN. Basta tener en cuenta que el área debe ser proporcional al ángulo y, dado que el área de la esfera es $4\pi r^2$, el área de un hemisferio (un huso de ángulo 180°) es $2\pi r^2$; de este modo, si el huso tiene ángulo α en vez de 180° , su área será $2\alpha \frac{\pi}{180} r^2$. \square

Proposición 3.3.17 (Propiedades de los triángulos esféricos)

- La suma de los tres lados es menor que 360° . Es decir, el defecto esférico es positivo.
- Cualquier lado de un triángulo esférico es menor que la suma de los otros dos y mayor que su diferencia.
- La suma de los ángulos de un triángulo esférico es mayor que dos rectos (180°) y menor que seis (540°). Por tanto, el exceso esférico es positivo y menor que 360° .
- El menor de los ángulos de un triángulo esférico difiere de la suma de los otros dos en menos de dos rectos (180°).
- La suma de los excesos/defectos esféricos de un triángulo esférico y sus adyacentes es 360° .
- El exceso/defecto esférico de un triángulo esférico es igual al defecto/exceso esférico de su triángulo polar.
- El área de un triángulo esférico es igual a

$$\frac{\pi}{180}(A + B + C - 180^\circ) r^2 = \frac{\pi}{180} \epsilon r^2.$$

DEMOSTRACIÓN. El resultado fundamental, en el que se basan los restantes, es g) y, por ello, lo demostramos primero:

- Dado un triángulo esférico ABC como en la Figura 3.6 consideramos los husos generados por los círculos máximos asociados a los lados del triángulo.

En la Figura 3.10 se muestran los dos husos asociados a los lados del triángulo esférico:

- Husos asociados a los lados b y c con área (sombreada) $4A \frac{\pi}{180} r^2$.
- Husos asociados a los lados a y c con área (sombreada) $4B \frac{\pi}{180} r^2$.
- Husos asociados a los lados a y b con área (sombreada) $4C \frac{\pi}{180} r^2$.

La unión de todas las zonas sombreadas cubre toda la esfera, aunque los triángulos ABC y su antipodal ($A'B'C'$ en la Figura 3.10) se cubren 3 veces (una por cada par de husos). Nótese que ambos triángulos tienen el mismo área. De este modo, se verifica que

$$4r^2(A + B + C) \frac{\pi}{180} = 4\pi r^2 + 2(\text{área}(ABC) + \text{área}(A'B'C')) = 4\pi r^2 + 4 \text{área}(ABC)$$

o, equivalentemente,

$$\text{área}(ABC) = \frac{\pi}{180}(A + B + C - 180^\circ) r^2.$$



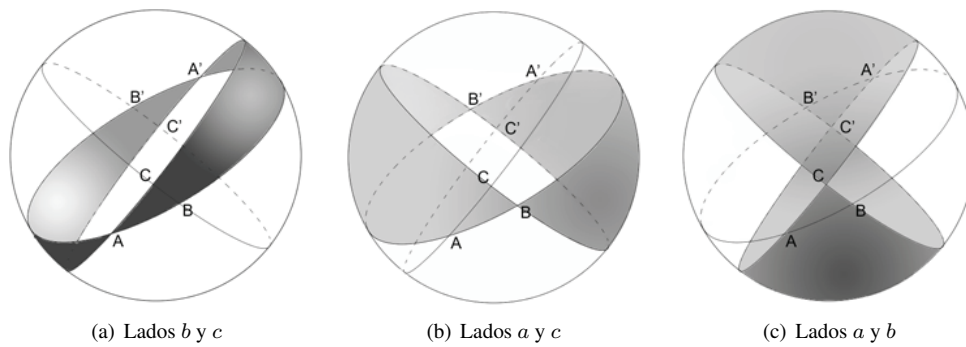


Figura 3.10: Husos asociados a los lados del triángulo esférico.

- c) Puesto que el área de un triángulo esférico puede oscilar entre 0 y $2\pi r^2$ (ya que todo triángulo da lugar a la vez a su triángulo antipodal, con el mismo área), el tamaño máximo de un triángulo es, por tanto, un hemisferio. De esta forma, por el apartado g), se tiene que

$$0 \leq \frac{\pi}{180} (A + B + C - 180^\circ) r^2 \leq 2\pi r^2 \Rightarrow 180^\circ \leq A + B + C \leq 540^\circ.$$

- f) Se verifica que

$$\bar{\varepsilon} = \bar{A} + \bar{B} + \bar{C} - 180^\circ = (180^\circ - a) + (180^\circ - b) + (180^\circ - c) - 180^\circ = 360^\circ - (a + b + c) = \delta.$$

Como el triángulo polar del triángulo polar es el triángulo original (véase la Proposición 3.3.13), se tiene también que $\bar{\delta} = \varepsilon$.

- a) Dado el triángulo ABC, se tiene que su polar verifica $\bar{\varepsilon} \geq 0$ luego, por el apartado f), es $\delta \geq 0$. Es decir,

$$360^\circ - (a + b + c) = \delta \geq 0 \Rightarrow a + b + c \leq 360^\circ.$$

- e) Un triángulo esférico ABC y sus adyacentes, unidos al triángulo esférico antipodal $A'B'C'$ y sus triángulos adyacentes (que son antipodales de los triángulos adyacentes a ABC) cubren toda la esfera. Por lo tanto, la suma de sus áreas es $4\pi r^2$ y, así, la suma de las áreas de ABC y sus triángulos adyacentes es $2\pi r^2$. Pero dicha suma es el producto de r^2 por la suma de sus excesos esféricos en radianes, de modo que dicha suma de excesos esféricos es $2\pi \text{ rad} = 360^\circ$. Si consideramos esta igualdad aplicada al triángulo polar del original, resulta que la suma de los excesos del triángulo polar y sus adyacentes (que es la suma de los defectos del triángulo original y sus adyacentes) es también de 360° .

- b) Puesto que $b + c$ es la longitud de un camino para ir de B a C sobre la esfera, se tiene que $b + c \geq a$ (pues a es el camino más corto de B a C; de hecho, la desigualdad es un menor estricto, pues A, B y C no están sobre un círculo máximo). Análogamente, se tiene que $b \leq a + c$ y, por tanto, que $b - c \leq a$.

- d) Puesto que $\bar{a} \leq \bar{b} + \bar{c}$, tendremos

$$180^\circ - A \leq (180^\circ - B) + (180^\circ - C),$$

de donde se concluye que

$$-A \leq 180^\circ - (B + C) \Rightarrow B + C - A \leq 180^\circ.$$

La desigualdad es cierta, en particular, si A es el menor de los tres ángulos. \square



A continuación definimos lo que se conoce con el nombre de coordenadas esféricas:

Definición 3.3.18 (Coordenadas esféricas) Consideremos un sistema de coordenadas cartesianas $\{O; X, Y, Z\}$. Todo punto $P \in \mathbb{R}^3$ puede representarse mediante sus *coordenadas esféricas* (r, θ, ψ) (véase la Figura 3.11):

$$\begin{cases} r : & \text{coordenada radial } (r > 0) \\ \theta : & \text{coordenada polar } (-90^\circ \leq \theta \leq 90^\circ) \\ \psi : & \text{coordenada acimutal } (0 \leq \psi \leq 360^\circ). \end{cases}$$

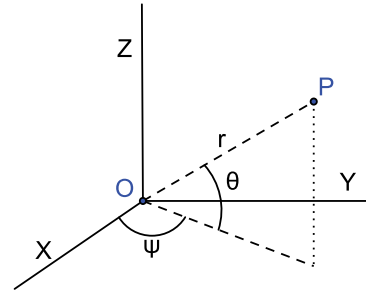


Figura 3.11: Coordenadas esféricas.

Las coordenadas cartesianas (x, y, z) de un punto P y sus coordenadas esféricas (r, θ, ψ) vienen relacionadas por las siguientes expresiones:

$$\begin{cases} x = r \cos \theta \cos \psi \\ y = r \cos \theta \sin \psi \\ z = r \sin \theta \end{cases} \quad \text{y} \quad \begin{cases} r = \sqrt{x^2 + y^2 + z^2} \\ \psi = \arctan \frac{y}{x} \\ \theta = \arctan \frac{z}{\sqrt{x^2 + y^2}}. \quad \square \end{cases}$$

Veamos ahora las principales relaciones entre los elementos de un triángulo esférico ABC de ángulos A, B, C y lados a, b, c como el de la Figura 3.37:

Teorema 3.3.19 (Fórmulas de Bessel)

a) *Primera fórmula de Bessel (Fórmula del coseno para lados):*

$$\begin{cases} \cos a = \cos b \cos c + \sin b \sin c \cos A \\ \cos b = \cos c \cos a + \sin c \sin a \cos B \\ \cos c = \cos a \cos b + \sin a \sin b \cos C. \end{cases}$$

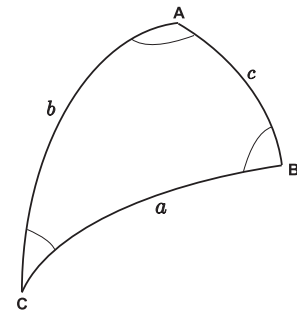


Figura 3.12: Triángulo esférico.

b) *Segunda fórmula de Bessel (Fórmula del seno):*

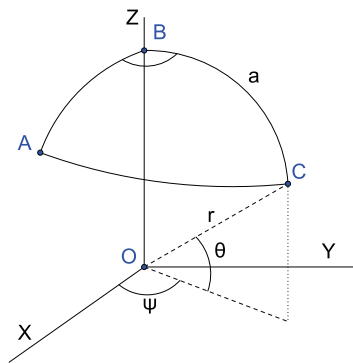
$$\frac{\sin A}{\sin a} = \frac{\sin B}{\sin b} = \frac{\sin C}{\sin c}.$$

c) *Tercera fórmula de Bessel (Analogía de Bessel):*

$$\begin{cases} \sin a \cos B = \cos b \sin c - \sin b \cos c \cos A \\ \sin a \cos C = \cos c \sin b - \sin c \cos b \cos A \\ \sin b \cos A = \cos a \sin c - \sin a \cos c \cos B \\ \sin b \cos C = \cos c \sin a - \sin c \cos a \cos B \\ \sin c \cos A = \cos a \sin b - \sin a \cos b \cos C \\ \sin c \cos B = \cos b \sin a - \sin b \cos a \cos C. \end{cases}$$

DEMOSTRACIÓN. Sea $\{O; X, Y, Z\}$ un sistema de referencia euclídeo tridimensional de forma que su origen O coincida con el centro de la esfera que contiene el triángulo esférico de vértices A, B y C y lados a, b y c . El eje Z viene determinado por la dirección del vértice B y el plano XZ contiene al lado c (véase la Figura 3.13).

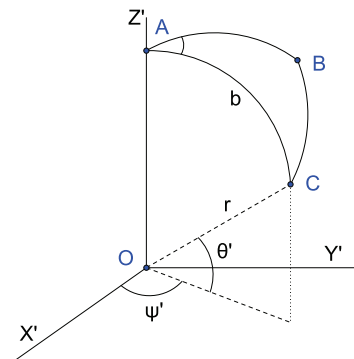


Figura 3.13: Coordenadas esféricas respecto a $\{O; X, Y, Z\}$.

Entonces, utilizando la relación entre las coordenadas cartesianas y las coordenadas esféricas, para el vértice $C = (x, y, z)$ se tiene que

$$\begin{cases} x = r \cos \theta \cos \psi = r \cos(90^\circ - a) \cos B = r \operatorname{sen} a \cos B \\ y = r \cos \theta \operatorname{sen} \psi = r \cos(90^\circ - a) \operatorname{sen} B = r \operatorname{sen} a \operatorname{sen} B \\ z = r \operatorname{sen} \theta = r \operatorname{sen}(90^\circ - a) = r \cos a. \end{cases}$$

Consideremos ahora el sistema de referencia euclídeo tridimensional $\{O; X', Y', Z'\}$ de tal forma que $Y' \equiv Y$ y el eje Z' está definido por la dirección del vértice A. Consecuentemente, el eje X' también pertenece al plano que contiene a los ejes X, Z y Z' (véase la Figura 3.14).

Figura 3.14: Coordenadas esféricas respecto a $\{O; X', Y', Z'\}$.

Respecto a este sistema de referencia, las coordenadas cartesianas (x', y', z') del vértice C son:

$$\begin{cases} x' = r \cos \theta' \cos \psi' = r \cos(90^\circ - b) \cos(180^\circ - A) = -r \operatorname{sen} b \cos A \\ y' = r \cos \theta' \operatorname{sen} \psi' = r \cos(90^\circ - b) \operatorname{sen}(180^\circ - A) = r \operatorname{sen} b \operatorname{sen} A \\ z' = r \operatorname{sen} \theta' = r \operatorname{sen}(90^\circ - b) = r \cos b. \end{cases}$$

Una rotación de ángulo c alrededor del eje $Y' \equiv Y$, permite hacer coincidir ambos sistemas de referencia y, consecuentemente, obtener las relaciones entre las coordenadas del vértice C a partir de la relación

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cos c & 0 & -\operatorname{sen} c \\ 0 & 1 & 0 \\ \operatorname{sen} c & 0 & \cos c \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$



De esta relación se obtienen las expresiones

$$\begin{cases} -r \operatorname{sen} b \cos A = r \cos c \operatorname{sen} a \cos B - r \operatorname{sen} c \cos a \\ r \operatorname{sen} b \operatorname{sen} A = r \operatorname{sen} a \operatorname{sen} B \\ r \cos b = r \operatorname{sen} c \operatorname{sen} a \cos B + r \cos c \cos a \end{cases}$$

de las cuales se deducen, respectivamente, una de las terceras, segundas y primeras fórmulas de Bessel. Permutando circularmente las letras, se obtienen las restantes fórmulas. \square

Observación 3.3.20 A partir de las fórmulas de Bessel obtenidas en el Teorema 3.3.19 pueden demostrarse las siguientes relaciones:

a) *Fórmula de las cuatro partes (o de las cotangentes):*

$$\begin{cases} \cos a \cos B = \operatorname{sen} a \cot c - \operatorname{sen} B \cot C \\ \cos a \cos C = \operatorname{sen} a \cot b - \operatorname{sen} C \cot B \\ \cos b \cos A = \operatorname{sen} b \cot c - \operatorname{sen} A \cot C \\ \cos b \cos C = \operatorname{sen} b \cot a - \operatorname{sen} C \cot A \\ \cos c \cos A = \operatorname{sen} c \cot b - \operatorname{sen} A \cot B \\ \cos c \cos B = \operatorname{sen} c \cot a - \operatorname{sen} B \cot A. \end{cases}$$

b) *Fórmula del coseno (para ángulos):*

$$\begin{cases} \cos A = -\cos B \cos C + \operatorname{sen} B \operatorname{sen} C \cos a \\ \cos B = -\cos C \cos A + \operatorname{sen} C \operatorname{sen} A \cos b \\ \cos C = -\cos A \cos B + \operatorname{sen} A \operatorname{sen} B \cos c. \end{cases}$$

c) *Fórmula análoga (para ángulos):*

$$\begin{cases} \operatorname{sen} A \cos b = \cos B \operatorname{sen} C + \operatorname{sen} B \cos C \cos a \\ \operatorname{sen} B \cos c = \cos C \operatorname{sen} A + \operatorname{sen} C \cos A \cos b \\ \operatorname{sen} C \cos a = \cos A \operatorname{sen} B + \operatorname{sen} A \cos B \cos c \end{cases}$$

(véanse los ejercicios correspondientes en la hoja de problemas). \square

A partir de las fórmulas del ángulo mitad de la trigonometría plana, y sustituyendo las fórmulas del coseno, se pueden obtener un grupo de fórmulas que explicitan la tangente del ángulo diedro mitad.

Teorema 3.3.21 (Fórmulas de Borda) Si $s = \frac{a+b+c}{2}$ es el semiperímetro del triángulo definido por los lados a , b y c , se verifica que:

$$\begin{cases} \tan \frac{A}{2} = \sqrt{\frac{\operatorname{sen}(s-b) \operatorname{sen}(s-c)}{\operatorname{sen} s \operatorname{sen}(s-a)}} \\ \tan \frac{B}{2} = \sqrt{\frac{\operatorname{sen}(s-a) \operatorname{sen}(s-c)}{\operatorname{sen} s \operatorname{sen}(s-b)}} \\ \tan \frac{C}{2} = \sqrt{\frac{\operatorname{sen}(s-a) \operatorname{sen}(s-b)}{\operatorname{sen} s \operatorname{sen}(s-c)}}. \end{cases}$$



DEMOSTRACIÓN. A partir de la primera fórmula de Bessel

$$\cos a = \cos b \cos c + \sin b \sin c \cos A$$

se deduce que

$$\cos A = \frac{\cos a - \cos b \cos c}{\sin b \sin c},$$

por lo que

$$\begin{aligned} 1 - \cos A &= \frac{\sin b \sin c - \cos a + \cos b \cos c}{\sin b \sin c} \\ &= \frac{\cos(b-c) - \cos a}{\sin b \sin c} = -2 \frac{\sin \frac{b-c+a}{2} \sin \frac{b-c-a}{2}}{\sin b \sin c}, \end{aligned} \quad (3.1)$$

donde se han aplicado las relaciones

$$\begin{cases} \cos(\alpha - \beta) = \cos \alpha \cos \beta + \sin \alpha \sin \beta \\ \cos \alpha - \cos \beta = -2 \sin \frac{\alpha + \beta}{2} \sin \frac{\alpha - \beta}{2}. \end{cases}$$

Análogamente,

$$1 + \cos A = -2 \frac{\sin \frac{a+b+c}{2} \sin \frac{a-b-c}{2}}{\sin b \sin c}. \quad (3.2)$$

A partir de las expresiones (3.1) y (3.2), se verifica que

$$\tan^2 \frac{A}{2} = \frac{1 - \cos A}{1 + \cos A} = \frac{\sin \frac{a+b-c}{2} \sin \frac{a+c-b}{2}}{\sin \frac{a+b+c}{2} \sin \frac{b+c-a}{2}}. \quad (3.3)$$

Puesto que para el semiperímetro s se verifican las relaciones

$$\frac{a+b-c}{2} = s-c, \quad \frac{c+a-b}{2} = s-b \quad \text{y} \quad \frac{b+c-a}{2} = s-a, \quad (3.4)$$

al sustituirlas en la expresión (3.3), se obtiene la relación buscada

$$\tan \frac{A}{2} = \sqrt{\frac{\sin(s-b) \sin(s-c)}{\sin s \sin(s-a)}}. \quad \square$$

A partir de las fórmulas de Borda, y teniendo en cuenta las fórmulas del ángulo suma de la trigonometría plana, se obtiene:

Teorema 3.3.22 (Analogías de Gauss–Delambre)

$$\begin{cases} \sin \frac{A+B}{2} \cos \frac{c}{2} = \cos \frac{a-b}{2} \cos \frac{C}{2} \\ \cos \frac{A+B}{2} \cos \frac{c}{2} = \cos \frac{a+b}{2} \sin \frac{C}{2} \\ \sin \frac{A-B}{2} \sin \frac{c}{2} = \sin \frac{a-b}{2} \cos \frac{C}{2} \\ \cos \frac{A-B}{2} \sin \frac{c}{2} = \sin \frac{a+b}{2} \sin \frac{C}{2}. \end{cases}$$

DEMOSTRACIÓN. Probemos, por ejemplo, la tercera de estas fórmulas. Utilizamos las expresiones (3.1), (3.2) y (3.4), encontradas en la deducción de las fórmulas de Borda, para calcular $\sin \frac{A}{2}$ y $\cos \frac{A}{2}$:

$$\begin{cases} \sin \frac{A}{2} = \sqrt{\frac{1 - \cos A}{2}} = \sqrt{\frac{\sin(s-b) \sin(s-c)}{\sin b \sin c}} \\ \cos \frac{A}{2} = \sqrt{\frac{1 + \cos A}{2}} = \sqrt{\frac{\sin s \sin(s-a)}{\sin b \sin c}}. \end{cases}$$



Análogamente,

$$\begin{cases} \operatorname{sen} \frac{B}{2} = \sqrt{\frac{\operatorname{sen}(s-c)\operatorname{sen}(s-a)}{\operatorname{sen} c \operatorname{sen} a}} \\ \operatorname{cos} \frac{B}{2} = \sqrt{\frac{\operatorname{sen} s \operatorname{sen}(s-b)}{\operatorname{sen} c \operatorname{sen} a}}. \end{cases}$$

Sustituyendo estas relaciones en la expresión

$$\operatorname{sen} \frac{A-B}{2} = \operatorname{sen} \frac{A}{2} \operatorname{cos} \frac{B}{2} - \operatorname{cos} \frac{A}{2} \operatorname{sen} \frac{B}{2}$$

y operando, se obtiene la relación

$$\operatorname{sen} \frac{A-B}{2} \operatorname{sen} \frac{c}{2} = \operatorname{sen} \frac{a-b}{2} \operatorname{cos} \frac{C}{2}. \quad \square$$

Observación 3.3.23 Permutando circularmente las letras, se obtienen las otras fórmulas que completan este grupo. \square

Dividiendo las analogías de Gauss–Delambre se obtienen:

Teorema 3.3.24 (Analogías de Neper)

$$\begin{cases} \tan \frac{a+b}{2} = \frac{\operatorname{cos} \frac{A-B}{2}}{\operatorname{cos} \frac{A+B}{2}} \tan \frac{c}{2} \\ \tan \frac{a-b}{2} = \frac{\operatorname{sen} \frac{A-B}{2}}{\operatorname{sen} \frac{A+B}{2}} \tan \frac{c}{2} \\ \tan \frac{A+B}{2} = \frac{\operatorname{cos} \frac{a-b}{2}}{\operatorname{cos} \frac{a+b}{2}} \cot \frac{C}{2} \\ \tan \frac{A-B}{2} = \frac{\operatorname{sen} \frac{a-b}{2}}{\operatorname{sen} \frac{a+b}{2}} \cot \frac{C}{2}. \end{cases}$$

DEMOSTRACIÓN. Por ejemplo, para obtener la expresión de $\tan \frac{A-B}{2}$, basta tener en cuenta que

$$\tan \frac{A+B}{2} = \frac{\operatorname{sen} \frac{A+B}{2}}{\operatorname{cos} \frac{A+B}{2}}$$

y utilizar las analogías de Gauss–Delambre. \square

Observación 3.3.25 Permutando circularmente las letras, se obtienen otras ocho fórmulas que completan este grupo. \square

3.4 Aplicaciones de la trigonometría esférica a la navegación

Como primera aproximación, se puede considerar la Tierra como una esfera girando 360° por día alrededor de un eje fijo (*eje de rotación*). Este eje interseca a la superficie de la Tierra en dos puntos llamados *polos geográficos norte y sur* (denotados por N y S).

- El círculo máximo cuyos polos son N y S es el *ecuador*.



- Los círculos menores paralelos al ecuador se denominan *paralelos*.
- Los círculos máximos perpendiculares al ecuador que pasan por los polos N y S se denominan *meridianos*. Llamaremos *meridiano superior* de un punto al arco de meridiano que tiene por extremos los polos y que contiene al punto.

El eje de la Tierra es una elección natural como eje de un sistema de coordenadas sobre la superficie de la Tierra. También se considera el ecuador y el *meridiano cero* (o de *referencia*) que es el *meridiano de Greenwich* (con propiedad, el meridiano superior de Greenwich). La posición de cualquier punto X sobre la superficie terrestre viene determinada por dos coordenadas angulares, la *latitud* Φ y la *longitud* Λ (véase la Figura 3.15):

- La *latitud* Φ es el ángulo medido a lo largo del meridiano superior del punto X desde el ecuador hasta X. Toma valores comprendidos entre -90° y 90° . El complementario de la latitud, $90^\circ - \Phi$, se denomina *colatitud*.
- La *longitud* Λ es el ángulo comprendido entre el meridiano cero y el meridiano superior del punto X, a lo largo del ecuador. Sus valores están, por tanto, entre -180° y 180° .

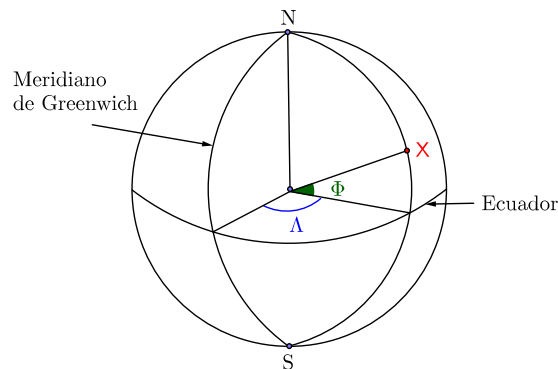


Figura 3.15: Latitud y longitud del punto X.

Observación 3.4.1

- Nótese que la latitud Φ y la longitud Λ juegan el papel de los ángulos θ y ψ , respectivamente, de las coordenadas esféricas. La única diferencia radica en que la longitud se suele considerar entre -180° y 180° (en vez de entre 0° y 360°).
- Habitualmente no se usan valores negativos para la longitud y la latitud. Cuando la longitud se mide hacia el este del meridiano cero (lo correspondiente a ángulos positivos) se habla de longitud este ($\Lambda = x \text{ E}$) y si se mide hacia el oeste (lo correspondiente a ángulos negativos) se habla de longitud oeste ($\Lambda = x \text{ W}$). Análogamente, para la latitud se habla de latitud norte ($\Phi = x \text{ N}$ si el ángulo de latitud x se mide hacia el polo norte) y de latitud sur ($\Phi = x \text{ S}$ si el ángulo de latitud x se mide hacia el polo sur). \square

Observación 3.4.2 Nótese que todos los puntos que tienen el mismo meridiano superior tienen la misma longitud y que todos los puntos de un mismo paralelo tienen la misma latitud. \square

Definición 3.4.3 Considerando la Tierra como una esfera de radio 6370 km, se llama *milla geográfica* (o *milla marina*) a la longitud del arco de un círculo máximo de ángulo un minuto. Una milla geográfica son 1853 m (compruébese). \square

Definición 3.4.4

- Se denomina *ruta* o *derrota* entre dos puntos a cualquier trayectoria que los una.
- Una ruta es *ortodrómica* si sigue un arco de círculo máximo.



- c) El *rumbo* geográfico de una ruta en un punto es el ángulo que forma el meridiano que pasa por dicho punto con la ruta (en el sentido de la misma), medido desde el norte y en sentido horario (norte–este–sur–oeste). □

Observación 3.4.5 Una forma alternativa de indicar el rumbo, en vez de usar el ángulo entre 0° y 360° dado por la definición anterior, es indicar el menor ángulo que forma la ruta con las direcciones de los puntos cardinales. La información proporcionada por el ángulo resultante, entre 0° y 45° , se completa con la indicación del punto cardinal con respecto al que se mide el ángulo, y en dirección a qué otro punto cardinal se está midiendo. Así, por ejemplo, 0° correspondería a N, 30° a N30°E, 300° a W30°N, 45° a N45°E o E45°N (o, simplemente, NE), y 160° sería S20°E.

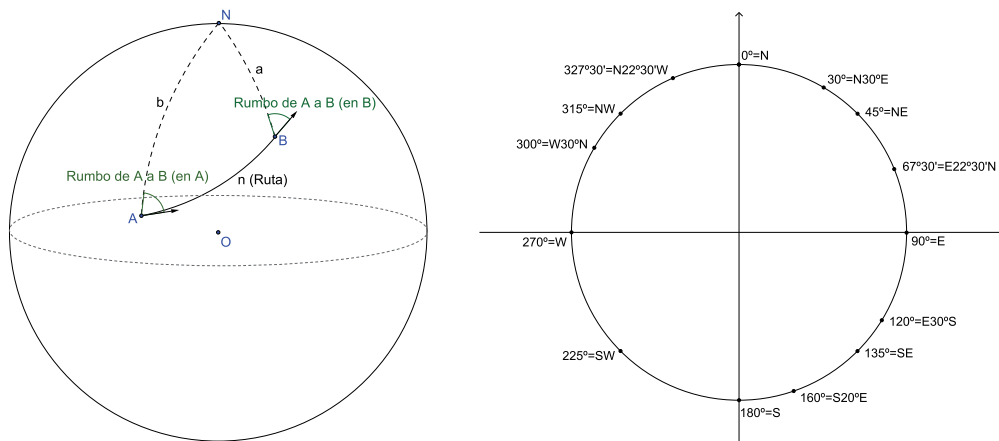


Figura 3.16: Ruta y rumbo.

En la sección de problemas se estudian algunas aplicaciones de las nociones anteriores a la navegación.

3.5 Sistemas de coordenadas en la esfera celeste

3.5.1 La esfera celeste

Un observador en la superficie terrestre puede pensarse como situado en el centro de una esfera imaginaria de radio arbitrario, denominada *esfera celeste*, sobre la cual los astros parecen moverse de este a oeste.

La *Astronomía de Posición* es la parte de la Astronomía que estudia la posición y el movimiento de los cuerpos materiales del universo en el espacio y en el tiempo, mediante medidas efectuadas en observaciones astronómicas. El término de “posición” se usa con el significado de “dirección aparente” y se expresa en función de dos coordenadas angulares. Para este objetivo, el tamaño del radio de la esfera celeste es irrelevante y se considera que toma un valor igual a la unidad.

Cuando se toma como centro de la esfera celeste la posición del observador sobre la superficie de la Tierra se habla de *esfera topocéntrica*. Cuando se considera el centro de la Tierra como centro de la esfera celeste, ésta se denomina *esfera geocéntrica*.¹

Definición 3.5.1 En una esfera celeste de centro O se consideran los siguientes elementos:

- Eje celeste*: Recta paralela al eje de rotación de la Tierra que pasa por O.
- Polos celestes*: Cada uno de los puntos en que el eje celeste interseca a la esfera celeste. Se denomina *polo celeste norte*, P_N (*polo celeste sur*, P_S) al polo celeste que está, respecto a O, en el mismo sentido que está el polo geográfico norte N (polo geográfico sur S) respecto al centro de la Tierra.

¹Si el centro considerado es el centro del Sol o el centro de otros cuerpos del sistema solar, se habla de *esfera heliocéntrica* o *planetocéntrica*.



- c) *Ecuador celeste*: Círculo máximo de la esfera celeste perpendicular al eje celeste.
- d) *Meridiano celeste*: Círculo máximo que pasa por los polos celestes.
- e) *Paralelo celeste*: Círculo menor paralelo al ecuador celeste. □

También se suelen considerar los siguientes conceptos, ligados a la posición del observador, que se supone situado sobre la tierra:

Definición 3.5.2 En una esfera celeste se denomina:

- a) *Vertical del lugar de observación*: Dirección de la fuerza de gravedad en el lugar de observación.
- b) *Horizonte celeste*: Círculo máximo de la esfera celeste perpendicular a la vertical. El horizonte celeste divide a la esfera en dos *hemisferios* que determinan la parte visible del cielo y la que no lo es.
- c) *Cénit y nadir*: Cada uno de los puntos en que la vertical interseca a la esfera celeste. Llamaremos *cénit* al que se encuentra en el hemisferio visible y *nadir* al otro.
- d) *Vertical celeste*: Círculo máximo que pasa por el cenit y el nadir. □

3.5.2 Sistemas de coordenadas celestes

Un *sistema de coordenadas* celestes viene dado por:

- a) Un *origen* de coordenadas: centro de la esfera celeste, O .
- b) Un *plano fundamental* que pasa por el centro de la esfera celeste y cuya intersección con la misma determina el *círculo máximo fundamental*. El *eje fundamental*, o perpendicular a este círculo que pasa por el origen, determina los *polos* de la misma.
- c) Un punto arbitrario del círculo máximo fundamental que se denomina *punto cero* (o *punto fundamental*) del sistema.
- d) Un *sentido* de medida de ángulos: entre los polos (indicando el polo en cuya dirección aumenta el tamaño de los ángulos) y en el círculo máximo fundamental (indicando, a partir del punto cero, el sentido en que crece el tamaño de los ángulos).

Cada elección de los elementos anteriores proporciona un sistema de coordenadas celestes. Se suelen considerar como origen de coordenadas el lugar de observación sobre la superficie de la Tierra (*coordenadas topocéntricas*) o el centro de masas de la Tierra (*coordenadas geocéntricas*). Como plano fundamental se suele tomar el horizonte celeste (como se hace en el sistema horizontal) o el ecuador (como en los sistemas ecuatoriales, tanto el horario como el absoluto). A continuación se describen estos sistemas de coordenadas:

Sistema horizontal (o altacimutal)

Se trata de un sistema topocéntrico en el que se considera la esfera celeste centrada en el observador. Como eje fundamental se toma la vertical del lugar de observación, por lo que el plano fundamental es el que contiene al horizonte celeste y los polos serán el cenit Z y el nadir Z' .

Denominaremos *meridiano del observador* al meridiano celeste definido por el plano que contiene al meridiano terrestre del observador (nótese que no es otro que el meridiano que pasa por el cenit Z y por el polo celeste norte P_N). El horizonte celeste corta al meridiano del observador en dos puntos: N (el más cercano a P_N) y S (el otro). Entre ellos, y a distancia de 90° , se encuentran los puntos E y W (E es el más cercano al norte, en sentido horario desde el cenit). Estos cuatro puntos sobre el horizonte se denominan *puntos cardinales norte, sur, este y oeste*, respectivamente (véase la Figura 3.17).

El *sistema de coordenadas horizontal* (o *altacimutal*) tiene los siguientes elementos de referencia:



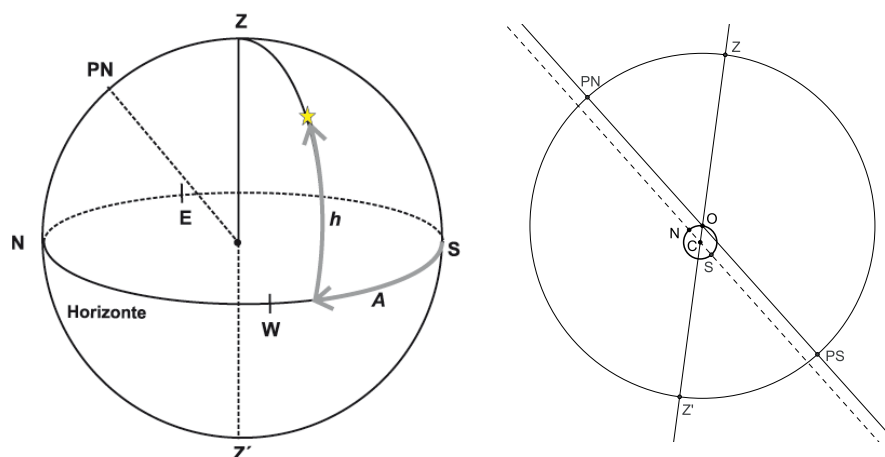


Figura 3.17: Sistema de coordenadas horizontal.

Plano fundamental:	<i>Horizonte</i>
Polo:	<i>Z (cénit)</i>
Punto cero:	<i>S (sur)</i>

Respecto al sistema de coordenadas horizontal, las coordenadas de un astro X son (A, h) o (A, z) , donde se denomina:

- Acimut* (A): ángulo medido a lo largo del horizonte, contado positivo desde el Sur y en sentido S-W-N-E, hasta el meridiano del astro.
- Altura* (h): ángulo medido en el vertical celeste del astro (ZXZ') desde el horizonte hasta el astro. El complementario de la altura es la *distancia cenital*, z .

$A = \angle SZX$	$(0^\circ \leq A \leq 360^\circ)$
$h = 90^\circ - z = 90^\circ - ZX$	$(-90^\circ \leq h \leq 90^\circ)$

Observación 3.5.3 Este sistema de coordenadas tiene el inconveniente de ser un sistema local, pues depende del lugar de observación: las coordenadas de un punto de la esfera celeste serán distintas para dos observadores situados en dos puntos diferentes de la Tierra (pues difieren sus respectivos horizontes celestes). Además, este sistema es instantáneo, en el sentido de que los valores de las coordenadas dependen del momento en que se realice la medición. \square

Sistema ecuatorial horario

Al igual que el sistema horizontal, se toma como centro de la esfera celeste la posición del observador, por lo que se trata de un sistema topocéntrico. En este sistema se toman los siguientes elementos de referencia:

Plano fundamental:	<i>Ecuador celeste</i>
Polo:	P_N (<i>polo celeste norte</i>)
Punto cero:	P_0

donde P_0 es el punto intersección del ecuador celeste con el meridiano del observador más cercano al sur S .

En el sistema ecuatorial horario (véase la Figura 3.18), las coordenadas de un astro X son (H, δ) , donde se denomina:

- Ángulo horario* (H): ángulo medido sobre el ecuador celeste y que forman los planos meridianos del observador y del astro (desde el meridiano del observador hacia el oeste). En esta coordenada, en lugar



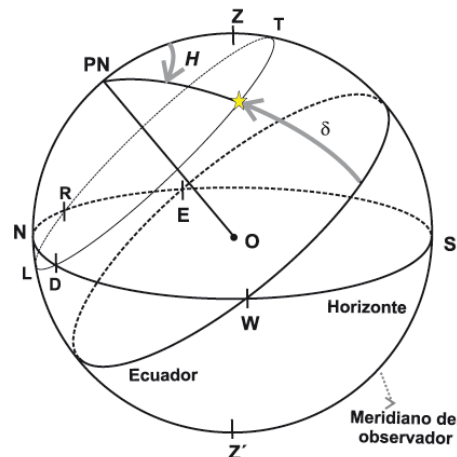


Figura 3.18: Sistema de coordenadas ecuatorial horario.

de utilizar grados sexagesimales, se utilizan horas (h), minutos terrestres (m) y segundos terrestres (s), teniendo en cuenta la siguiente equivalencia

$$1^h = 15^\circ, 1^m = 15' \text{ y } 1^s = 15''.$$

Nótese que una hora (respectivamente, un minuto o un segundo terrestre) es el ángulo que gira la Tierra en una hora de tiempo (respectivamente, en un minuto o en un segundo de tiempo).

- ii) *Declinación* (δ): ángulo medido en el meridiano de la estrella ($P_N X P_S$), desde el ecuador hasta la estrella. El complementario de la declinación es la *distancia polar*.

$$\begin{aligned} H &= \angle Z P_N X & (0^h \leq H \leq 24^h) \\ \delta &= 90^\circ - \angle P_N X & (-90^\circ \leq \delta \leq 90^\circ) \end{aligned}$$

Observación 3.5.4 Un observador situado en un punto de la superficie de la Tierra ve aparecer las estrellas por el este, describiendo un arco sobre el horizonte hasta que desaparecen por el oeste, alcanzando el punto más alto sobre el horizonte al cruzar el meridiano del observador. Este movimiento aparente del cielo tal como aparece a un observador situado sobre la superficie de la Tierra es debido a la rotación terrestre, que hace que los astros a lo largo del día describan un paralelo de declinación. Durante este movimiento (conocido como *movimiento diurno*), la declinación de una estrella no varía pero, en cambio, sí lo hace el ángulo horario. Por lo tanto, en este sistema, la coordenada ángulo horario es tanto local como instantánea. \square

Sistema de coordenadas ecuatorial absoluto

Es un sistema geocéntrico, por lo que se considera la esfera celeste con la Tierra situada en su centro. Antes de describir sus elementos de referencia, introduzcamos algunos conceptos.

El plano de la órbita de la Tierra se conoce como *plano de la eclíptica* y el círculo máximo que es la intersección de este plano con la esfera celeste se conoce con el nombre de *eclíptica*. A lo largo del movimiento anual, el eje de la Tierra mantiene una dirección que forma un ángulo $\epsilon = 23^\circ 27'$ con la normal al plano de la eclíptica (la *oblicuidad de la eclíptica*). Los puntos de intersección de la eclíptica y del ecuador celeste son:

- punto Aries* (γ): punto en el que el Sol, en su movimiento “aparente”, pasa del hemisferio sur celeste al norte. Esto ocurre al comienzo de la primavera (alrededor del 21 de marzo). También se conoce como *equinoccio de primavera*.
- punto Libra* (Ω): punto diametralmente opuesto a γ . En él, el Sol pasa del hemisferio norte celeste al sur. Esto ocurre al comienzo del otoño (alrededor del 22 de septiembre) y se conoce también como *equinoccio de otoño*.



La recta que une estos dos puntos se denomina *línea de los equinoccios*. El sistema de coordenadas ecuatorial absoluto (véase la Figura 3.19) tiene los siguientes elementos de referencia:

Plano fundamental:	<i>Ecuador celeste</i>
Polo:	P_N (polo celeste norte)
Punto cero:	γ (punto Aries)

Respecto al sistema de coordenadas ecuatorial absoluto, las coordenadas de un astro X son (α, δ) , siendo:

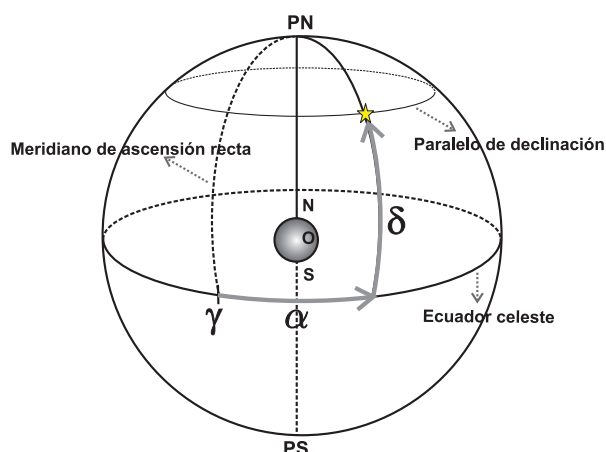


Figura 3.19: Sistema de coordenadas ecuatorial absoluto.

- i) *Ascensión recta* (α): ángulo medido en el plano del ecuador desde el punto Aries hasta el meridiano del astro y medido en sentido antihorario desde P_N .
- ii) *Declinación* (δ). Se define igual que en el sistema ecuatorial horario.

$$\alpha = \angle \gamma P_N X \quad (0^h \leq \alpha \leq 24^h)$$

$$\delta = 90^\circ - \angle P_N X \quad (-90^\circ \leq \delta \leq 90^\circ)$$

Observación 3.5.5 Nótese que los puntos situados en el ecuador celeste tienen declinación nula (en particular, los equinoccios). \square

Transformaciones entre sistemas de coordenadas celestes

Los sistemas de coordenadas celestes nos permiten desde localizar objetos cuya posición conocemos previamente, a deducir la posición aproximada de un objeto que estamos observando, para identificarlo. Conociendo las coordenadas de un astro podemos localizarlo en el cielo, ya sea directamente mediante los círculos graduados de un telescopio o indirectamente mediante cartas celestes, catálogos y planisferios.

Básicamente, los telescopios se pueden clasificar, atendiendo a su montura (que es la parte mecánica que une el trípode o base al instrumento óptico), en dos tipos, según los planos de referencia que utilicen (coordenadas): telescopios de *montura altacimutal* y telescopios de *montura ecuatorial*. La montura tiene como objetivo proveer de movimiento controlado al telescopio. La más simple es la montura altacimutal, que realiza movimientos horizontales y verticales (acimut y altura, respectivamente). Este tipo de montura también se presenta en los teodolitos. La montura ecuatorial, que utiliza como plano fundamental el ecuador celeste, usa las coordenadas ecuatoriales absolutas, ascensión recta y declinación. La montura ecuatorial tiene la ventaja de que para mantener el telescopio fijo en una dirección, es decir, para contrarrestar el movimiento aparente de un objeto celeste debido a la rotación de la Tierra, basta con hacer girar respecto a uno de los ejes (el de la ascensión recta) dejando fijo el otro (el de la declinación). Sin embargo, tiene el inconveniente de ser un mecanismo más complicado (y caro) que una montura altacimutal. Estas monturas,



en cambio, necesitan un movimiento simultáneo (y diferente según la posición) en los ejes de altura y acimut para mantener el telescopio apuntado a un objeto estelar fijo.

Puesto que, además, la ascensión recta y la declinación son las coordenadas utilizadas en la descripción y elaboración de las cartas celestes, catálogos y planisferios, se hace necesario conocer las transformaciones entre los distintos sistemas de coordenadas, tanto para orientar la montura de un telescopio o teodolito para observar un astro, como para identificar un objeto observado (deducir su ascensión recta y declinación) a partir de su posición altacimutal o ecuatorial horaria.

Observación 3.5.6 (Planisferio celeste) Un planisferio es una representación plana de un hemisferio de la esfera celeste. Se consigue mediante la *proyección estereográfica*, la cual utiliza como centro de proyección uno de los polos celestes (si se quiere representar el hemisferio norte, se considera el polo sur y el polo norte para la representación del hemisferio sur). El plano de proyección es tangente a la esfera en el antípoda del centro de proyección (en este caso será un plano paralelo al ecuador y que pasa por el polo norte o sur, dependiendo del hemisferio que se quiera representar). La proyección estereográfica ecuatorial hace corresponder a cada punto A del hemisferio considerado un punto A' en el plano de proyección que es la intersección de éste con la recta que une el centro de proyección con el punto A . La proyección estereográfica tiene las siguientes propiedades:

- Los meridianos en la esfera celeste se transforman en rectas que concurren en el transformado del polo (que es el centro del planisferio).
- Los paralelos en la esfera celeste se transforman en circunferencias concéntricas en el planisferio.
- Es conforme, lo que quiere decir que si dos curvas sobre la superficie de la esfera celeste se cortan en un determinado ángulo, sus proyecciones se cortan en el mismo ángulo. Se trata, por tanto, de una proyección que conserva los ángulos. Esto es importante pues permite el reconocimiento “rápido” de las constelaciones. \square

En esta sección vamos a obtener las relaciones existentes entre las coordenadas horizontales y ecuatoriales, para lo que serán de utilidad los conceptos y fórmulas de trigonometría esférica estudiados con anterioridad.

Proposición 3.5.7 (Coordenadas horizontales vs. horarias) Consideremos un sistema de coordenadas horizontal y un sistema de coordenadas ecuatorial horario, centrados ambos en el mismo lugar de observación O de latitud Φ . Entonces, las coordenadas ecuatoriales horarias se escriben en función de las horizontales mediante las fórmulas:

$$\begin{cases} \operatorname{sen} \delta = \operatorname{sen} \Phi \operatorname{sen} h - \cos \Phi \cos h \cos A \\ \tan H = \frac{\cos h \operatorname{sen} A}{\operatorname{sen} h \cos \Phi + \cos h \operatorname{sen} \Phi \cos A} \end{cases}$$

Recíprocamente, las coordenadas horizontales se escriben en función de las ecuatoriales horarias mediante

$$\begin{cases} \operatorname{sen} h = \operatorname{sen} \Phi \operatorname{sen} \delta + \cos \Phi \cos \delta \cos H \\ \tan A = \frac{\cos \delta \operatorname{sen} H}{-\operatorname{sen} \delta \cos \Phi + \cos \delta \operatorname{sen} \Phi \cos H} \end{cases}$$

DEMOSTRACIÓN. Las coordenadas horizontales y las coordenadas ecuatoriales horarias tienen una relación que puede visualizarse con ayuda de la Figura 3.20. Nótese que, puesto que el ecuador terrestre y el celeste están en planos paralelos, el ángulo que forma con el ecuador celeste la recta que une el origen y el cénit es, exactamente, la latitud del observador.

En primer lugar, supongamos conocidas la altura h y el acimut A . Si fijamos nuestra atención en el triángulo esférico representado en la Figura 3.20, de él conocemos dos lados ($90^\circ - h$ y $90^\circ - \Phi$) y el ángulo comprendido ($180^\circ - A$). Podemos, por tanto, gracias a la primera fórmula de Bessel, determinar el lado opuesto escribiendo

$$\cos(90^\circ - \delta) = \cos(90^\circ - \Phi) \cos(90^\circ - h) + \operatorname{sen}(90^\circ - \Phi) \operatorname{sen}(90^\circ - h) \cos(180^\circ - A),$$



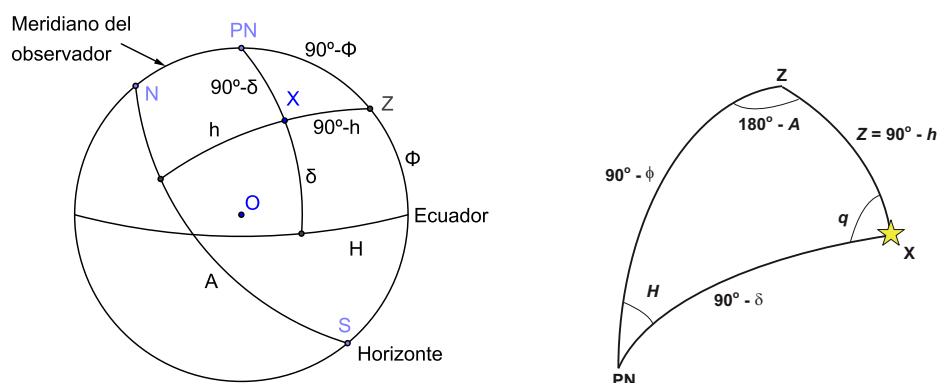


Figura 3.20: Relación entre coordenadas horizontales y ecuatoriales horarias.

es decir,

$$\sin \delta = \sin \Phi \sin h - \cos \Phi \cos h \cos A.$$

A continuación, podríamos despejar $\cos H$ mediante la primera fórmula de Bessel, pero nos quedaría en función de δ . Para escribir el ángulo horario H en función de h y A usamos la segunda y tercera fórmulas de Bessel:

$$\begin{cases} \frac{\sin(90^\circ - \delta)}{\sin(180^\circ - A)} = \frac{\sin(90^\circ - h)}{\sin H} \\ \sin(90^\circ - \delta) \cos H = \cos(90^\circ - h) \sin(90^\circ - \Phi) - \sin(90^\circ - h) \cos(90^\circ - \Phi) \cos(180^\circ - A) \end{cases}$$

o, equivalentemente,

$$\begin{cases} \cos \delta \sin H = \cos h \sin A \\ \cos \delta \cos H = \sin h \cos \Phi + \cos h \sin \Phi \cos A. \end{cases}$$

Haciendo el cociente entre estas dos últimas expresiones obtenemos la relación buscada

$$\tan H = \frac{\cos h \sin A}{\sin h \cos \Phi + \cos h \sin \Phi \cos A}.$$

La relación recíproca se obtiene de forma similar, teniendo en cuenta que también en este caso se conocen dos lados ($90^\circ - \delta$ y $90^\circ - \Phi$) y el ángulo comprendido (H) del triángulo esférico de la Figura 3.20. \square

Observación 3.5.8 Las relaciones anteriores proporcionan el ángulo H (respectivamente A) con ambigüedad, pues queda determinado a través de su tangente. Es fácil ver que H y A están, ambos, en el mismo intervalo $[0^\circ, 180^\circ]$ o $[180^\circ, 360^\circ]$. Las otras dos coordenadas (δ y h) quedan unívocamente determinadas en su intervalo de definición (esto es, $[-90^\circ, 90^\circ]$) al ser calculadas mediante su seno.

Definición 3.5.9 Se denomina *hora sidérea* (o *tiempo sidéreo*) de un punto P de la superficie terrestre al ángulo horario del punto Aries cuando se toma P como origen del sistema de coordenadas ecuatorial horario. \square

Teniendo en cuenta la relación de tamaño existente entre los radios de la esfera celeste y de la Tierra, se puede suponer que para ambos sistemas de coordenadas ecuatoriales (absoluto y horario) ambos polos celestes norte y ambos ecuadores celestes coinciden. Bajo este supuesto, ambas declinaciones son la misma. La relación entre el ángulo horario y la ascensión recta viene dada por:

Proposición 3.5.10 (Relación fundamental de la Astronomía de posición) Considerando un sistema de coordenadas ecuatorial horario centrado en un observador O , si H es el ángulo horario en dicho sistema, se verifica que

$$\theta = H + \alpha,$$

siendo θ la hora sidérea de O , es decir, en cada instante de tiempo, y para cualquier astro, la suma de su ángulo horario con su ascensión recta coincide con la hora sidérea de O .



DEMOSTRACIÓN. Basta observar que, en la Figura 3.21, el ángulo horario del astro es AHX y el tiempo sidéreo es $AH\gamma$. \square

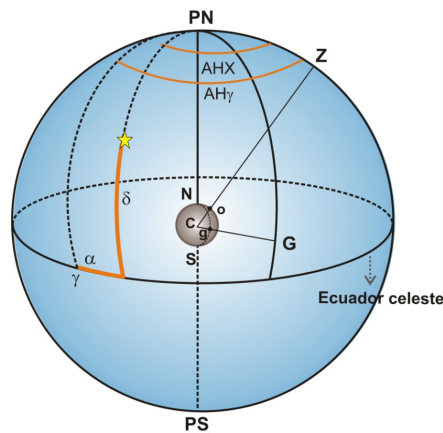


Figura 3.21: Relación fundamental de la Astronomía de posición.

Observación 3.5.11 (Coordenadas horarias vs. absolutas) Así pues, el cambio entre coordenadas ecuatoriales horarias y absolutas es muy sencillo: la declinación es la misma y el ángulo horario y la ascensión rectas están ligadas por la relación fundamental de la Astronomía de posición a través del tiempo sidéreo del lugar de observación. \square

Observación 3.5.12 (Coordenadas horizontales vs. absolutas) El cambio de coordenadas horizontales a coordenadas ecuatoriales absolutas (y viceversa) debe hacerse en dos pasos, a través de las coordenadas ecuatoriales horarias. \square

3.6 El elipsoide terrestre

3.6.1 Forma y dimensiones de la Tierra. Reseña histórica

Las grandes civilizaciones de la antigüedad (egipcia, babilonia y sumeria) se basaron en la mitología y las representaciones simbólicas para describir la forma y la situación de la Tierra en el cosmos. Por lo tanto, se puede decir que, para ellos, la astronomía era esencialmente mítica: se pensaba que los astros, aún siendo inaccesibles, podían influir en los seres humanos. Utilizaron la periodicidad de los movimientos celestes para establecer calendarios, cuya principal utilidad era la supuesta predicción de las voluntades divinas. Es a partir del siglo VI antes de nuestra era cuando los griegos instauran el racionalismo científico y aparecen las primeras “teorías del mundo”. En esa época, ya se entiende que un modelo sólo puede ser válido si las predicciones son ratificadas por las observaciones.

Así, Tales de Mileto (624–546 a.C.) imaginaba una Tierra plana flotando sobre el agua, rodeada de un cielo hemisférico. Anaximandro (610–545 a.C.), sin embargo, afirmaba que la Tierra es un cilindro, que no reposa en nada, y nosotros vivimos en la parte plana superior. Para Pitágoras de Samos (segunda mitad del siglo VI a.C.), el orden y la arquitectura del mundo no podían estar gobernados más que por el número y las configuraciones geométricas más perfectas y más simples. La Tierra no podía concebirse de otro modo que esférica, en armonía con la bóveda celeste. Aparte de esta concepción dogmática, las observaciones de fenómenos terrestres o celestes confirmaron bastante rápidamente a los filósofos griegos la validez de esta representación de la Tierra, que fue definitivamente admitida por los sucesores de Pitágoras (como Philolaos y Héraclides). Sólo la rotación de la Tierra sobre sí misma así como su lugar en el rango de planetas fueron objetos de especulación.

Admitida la forma de la Tierra como esférica, faltaba determinar las dimensiones de su circunferencia. La primera medida científica de la circunferencia terrestre fue efectuada por Eratóstenes (276–194 a.C.). Para ello empleó un método trigonométrico, además de las nociones de latitud y longitud. Por referencias



obtenidas de un papiro de su biblioteca, sabía que en Asuán el día del solsticio de verano los objetos no proyectaban sombra alguna y la luz alumbraba el fondo de los pozos; esto significaba que la ciudad estaba situada justamente sobre la línea del trópico, y su latitud era igual a la de la eclíptica que ya conocía. Eratóstenes, suponiendo que Asuán y Alejandría tenían la misma longitud (realmente distan 3°) y que el Sol se encontraba tan alejado de la Tierra que sus rayos podían suponerse paralelos, midió la sombra en Alejandría el mismo día del solsticio de verano al mediodía, demostrando que el cénit de la ciudad distaba $\frac{1}{50}$ parte de la circunferencia (es decir, $7^\circ 12'$) del de Alejandría. Posteriormente, tomó la distancia estimada por las caravanas que comerciaban entre ambas ciudades, de donde dedujo que la circunferencia de la Tierra era de 250000 estadios (unidad de longitud griega, que tomaba como patrón la longitud del estadio de Olimpia, que equivalía a $174'125$ m), resultado que posteriormente elevó hasta 252000 estadios, de modo que a cada grado correspondieran 700 estadios. Aparte de una supuesta medida de Posidonius, Ptolomeo y ciertos astrónomos árabes refinaron las medidas y corroboraron las de Eratóstenes.

Hay que esperar al siglo XVI para obtener una medida más precisa del meridiano. La tentativa fue realizada en 1550 por Fernel (1497–1558), que medirá la longitud de un arco de meridiano de 1° entre París y Amiens. Para ello, construyó un aparato para contar el número de vueltas de rueda de un carro, sobre un camino casi rectilíneo entre París y Amiens. Encuentra 57077 toesas (1 toesa = $1'949$ m). No es hasta el año 1615 cuando el matemático y geómetra holandés Snellius (1580–1626) mide un arco de meridiano de 1° por un nuevo método basado en la construcción, a lo largo del meridiano a medir, de una red de triángulos: un lado (base) y dos ángulos del primer triángulo medidos sobre el terreno, lo que permite calcular, por vía trigonométrica, todos los lados del triángulo y deducir la longitud buscada. Este procedimiento se denomina *triangulación* y fundó los métodos de la geodesia moderna. El anteojo astronómico, aparecido al principio del siglo XVII, permitió construir instrumentos geodésicos con una precisión mayor que la de los que utilizaron Fernel y Snellius.

Picard (1620–1682) retomó en 1665, con nuevos instrumentos, la medida del meridiano de París a Amiens y la determinó como 57060 toesas para 1° de meridiano. En 1671 Picard publica una obra titulada *Medida de la Tierra* en la que la forma esférica de la Tierra aparece ya como una aproximación de la figura real.

Polémica sobre la forma de la Tierra

Las medidas del meridiano comenzadas a principios del siglo XVII y proseguidas a lo largo de todo el siglo XVIII se van haciendo cada vez más precisas, no tardando en plantearse cuestiones relacionadas con la figura de la Tierra. Se encuentra que un *esferoide* (elipsoide de revolución) sería la mejor aproximación de la Tierra, pero la cuestión de saber cómo era el eje de revolución del esferoide fue objeto de una viva polémica: ¿la Tierra está achatada en los polos o en el ecuador?

Huygens (1629–1695) y Newton (1642–1727) llegaron a deducir, utilizando distintas consideraciones, que la figura de la Tierra era un esferoide ampliado en el ecuador y achatado en los polos, lo que conllevaría a una menor longitud del grado de meridiano medido cerca del ecuador que un grado del mismo meridiano medido en las vecindades de los polos. Además, Newton encontró que el aplanamiento tenía un valor de $\frac{1}{230}$.

Las medidas de Cassini y de La Hire parecían ir en contra de las afirmaciones de Newton y Huygens. Surgió una polémica que sólo pudo resolverse mediante numerosas medidas efectuadas en expediciones astrogeodésicas en diversos lugares de la Tierra próximos al polo norte y al ecuador. Así, la Academia de Ciencias de París decide hacer la medida de dos arcos de grado de meridiano, una en Perú por Godin, Bouguer y La Condamine, de 1735 a 1743, y en el que participaron los españoles Jorge Juan y Antonio de Ulloa, y la segunda en Laponia en 1736 por Maupertuis y Clairaut, entre otros. Los resultados de las medidas confirmaron lo enunciado por Huygens y Newton respecto a que la Tierra estaba achatada por los polos, contra la afirmación de Cassini.

Unidades de longitud y su relación con las dimensiones terrestres

Estos trabajos astrogeodésicos encontraron una aplicación en la elaboración de un sistema métrico encargado en 1790 a una comisión científica formada por Borda, Lagrange, Laplace, Monge y Condorcet. Se decidió, en 1791, adoptar para la unidad de longitud, el *metro*, como la diez millonésima parte de un



cuarto de meridiano del esferoide terrestre. En 1837 el sistema métrico decimal fue declarado obligatorio en Francia y paulatinamente fue adoptado por casi todos los países salvo los anglosajones (quienes sólo recientemente lo están introduciendo progresivamente).

Después, en 1875, la Convención del Metro instituyó una Oficina Internacional de Pesos y Medidas cuya sede se fijó en París donde, en el pabellón de Breteuil se guardan el metro internacional (de platino e iridio), como también el kilogramo internacional. Sin embargo, los avances incesantes de la técnica obligaron a una redefinición del metro ya para comienzos de los años sesenta. El primero de enero de 1961 se define el *metro* como “la longitud igual a 1650763’73 veces la longitud de onda en el vacío de la radiación correspondiente a la transición entre los niveles $2p_{10}$ y $5d_5$ del átomo de cripton 86”. Finalmente, el 20 de Octubre de 1983, se adoptó como definición de metro “la longitud de trayecto recorrido en el vacío por la luz durante un tiempo de $1/299792458$ de segundo”.

La Tierra representada por elipsoides

Las medidas geodésicas continuaron a lo largo de todo el siglo XIX con técnicas cada vez más precisas, con el fin de conseguir elipsoides que se aproximasen mejor a la Tierra. Es de destacar que la mayoría de los grandes matemáticos de esta época, tales como Gauss, Lagrange, Monge, Laplace, Bessel y Clairaut contribuyeron a potenciar la Geodesia. En particular, los rápidos progresos del análisis matemático –cálculo diferencial e integral– permitieron elaborar una teoría del potencial, de la mecánica de fluidos y de medios deformables que conduciría a modelizaciones de masas de fluidos homogéneos en rotación uniforme por elipsoides de revolución (*esferoide de Mac–Laurin*) o por elipsoides arbitrarios (*elipsoide de Jacobi*).

En la segunda mitad del siglo XX, el uso de satélites artificiales ha proporcionado nuevas técnicas, renovando los métodos tradicionales de la geodesia:

- i) Reemplazando redes de triángulos en la superficie de la Tierra por redes de tetraedros desde la Tierra a la órbita del satélite, obteniéndose medidas mucho más precisas, rápidas y sobre una gran distancia.
- ii) El análisis de variaciones de órbitas de satélites permite, por métodos de la mecánica celeste aplicados a las perturbaciones, deducir los parámetros geométricos con mayor precisión.
- iii) Al igual que en otras disciplinas, los avances informáticos han permitido realizar cálculos simbólicos y numéricos antaño imposibles.

3.6.2 Posición de un observador en la superficie de la Tierra

En la Sección 3.4 se han visto algunas aplicaciones de la trigonometría esférica a la navegación, como el cálculo de rumbos y la determinación de la posición de un observador (latitud y longitud) en la superficie de la Tierra considerada esférica. La resolución de estos problemas está ligada a la resolución de triángulos esféricos. En el caso de la navegación, el triángulo esférico viene dado por la *coordenadas geográficas* de dos puntos y uno de los polos geográficos; esto es, si los vértices vienen dados por ambos puntos y por el polo norte, los lados serán las colatitudes respectivas (complementarios de las latitudes) y la ruta; los ángulos respectivos al polo norte y a ambos puntos de la superficie terrestre, los determinan la diferencia de longitudes y los rumbos recíprocos, respectivamente.

Sin embargo, la forma de la Tierra no se corresponde con una esfera, ya que la diferencia entre el radio en el ecuador terrestre y el radio en los polos de alrededor de 21 km. Aunque esta diferencia es relativamente pequeña (0.3% radio ecuatorial = 6378 km), es imprescindible tenerla en cuenta en la conformación de mapas, en el cálculo de eclipses, la estimación de trayectorias de satélites, el posicionamiento GPS, etc.

La disciplina que se ocupa de estudiar la figura geométrica precisa de la Tierra y los métodos que se emplean para la determinación de dicha figura es la *Geodesia*. Como ya se ha comentado, antes del advenimiento de los satélites artificiales el trabajo geodésico se realizaba por métodos de triangulación y de gravimetría hechos sobre el terreno. Con dichos satélites se ha conseguido un mayor conocimiento sobre la forma verdadera de nuestro planeta.

Para representar la forma de la Tierra se suelen utilizar tres tipos de superficies (véase la Figura 3.22):



- a) *Superficie topográfica*: figura real de la Tierra formada por los mares, océanos y continentes.
- b) *Geoide*: figura geométrica que busca representar la verdadera forma de la Tierra haciendo que la figura coincida con el nivel medio de los océanos y continúe sobre las áreas continentales como la superficie imaginaria (a nivel promedio del mar), la cual tiene, por definición, la propiedad de que cualquier lugar de su superficie es perpendicular a la dirección de la fuerza de la gravedad.
- c) *Elipsoide de referencia*: superficie matemática apta para realizar cálculos geodésicos. Puesto que el geoides es imposible de identificar con una figura geométrica sencilla, (pues resulta completamente irregular), se suele adoptar como figura geométrica apropiada, en muy buena aproximación, un *elipsoide de revolución*, llamado también *esferoide*, cuya forma tridimensional resulta de rotar por completo una elipse sobre su eje menor. Un esferoide o elipsoide de revolución queda determinado por el *radio ecuatorial* a , que juega el papel del semieje mayor del elipsoide, y el *radio polar* b , relacionados a través de relaciones como el *aplanamiento* $\alpha = \frac{a-b}{a}$ o la *primera excentricidad* $e^2 = \frac{a^2-b^2}{a^2}$.

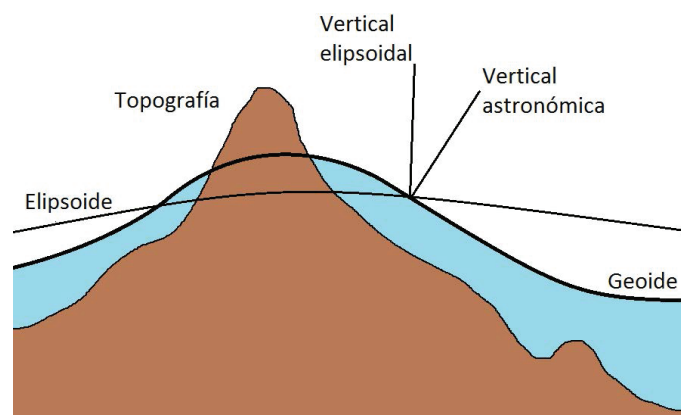


Figura 3.22: Tipos de superficie de la Tierra.

Sobre la superficie terrestre, podemos definir tres sistemas de coordenadas: el *sistema de coordenadas astronómicas*, que es completamente independiente de la forma y dimensiones de la Tierra y se determina a partir de la vertical astronómica, y los *sistemas de coordenadas geodésicas* y *geocéntricas*, los cuales se basan en la utilización de un elipsoide de revolución como aproximación de la Tierra.

Estos sistemas de coordenadas sobre la superficie de la Tierra se conocen con el nombre de *sistemas de coordenadas geográficas*. La posición de un punto sobre la Tierra expresada en términos de sus coordenadas en uno de estos sistemas se denomina *posición geográfica*. El plano fundamental para los tres sistemas es el *ecuador terrestre* y las coordenadas angulares son *latitud* y *longitud*, a las que se añadirá el adjetivo *geocéntrico*, *geodésico* y *astronómico* (o *geográfico*), respectivamente. Presentaremos a continuación las definiciones de las distintas coordenadas de un observador en la superficie de la Tierra.

Coordenadas geocéntricas

Este sistema de coordenadas tiene como origen el centro de masas de la Tierra. El plano fundamental es, como en los tres sistemas, el ecuador terrestre. Las coordenadas geocéntricas son (véase la Figura 3.23):

- a) φ' , *latitud geocéntrica* de un punto sobre la superficie terrestre: ángulo existente entre la recta que pasa por el punto y el centro del planeta, y el ecuador terrestre. El rango de valores que toma es

$$-90^\circ (90^\circ\text{S}) \leq \varphi' \leq 90^\circ (90^\circ\text{N}).$$

- b) λ' , *longitud geocéntrica* de un punto sobre la superficie terrestre: ángulo medido sobre el ecuador terrestre, desde el meridiano cero (o de referencia) hasta el meridiano del punto correspondiente.
- c) ρ , *distancia radial* de un punto sobre la superficie terrestre: distancia en línea recta existente entre dicho punto y el centro de masas de la Tierra.



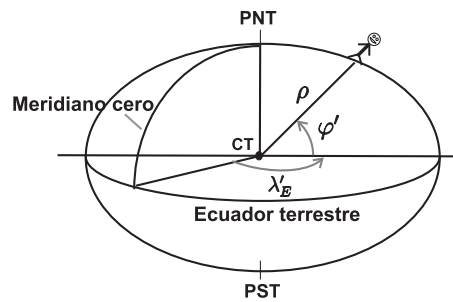


Figura 3.23: Coordenadas geocéntricas.

Coordenadas geodésicas

Este sistema de coordenadas descansa enteramente en un esferoide (elipsoide de revolución) de referencia que hay que especificar de entrada. Un esferoide queda determinado cuando se adoptan valores específicos del radio ecuatorial terrestre a y del aplastamiento α (o un parámetro equivalente). La importancia de este sistema de coordenadas radica en que la latitud geodésica es la que se encuentra en los mapas y atlas geográficos. Las coordenadas geodésicas son (véase la Figura 3.24):

- a) φ , *latitud geodésica* de un punto sobre la superficie terrestre: ángulo existente entre la normal al esferoide en dicho punto y el ecuador terrestre. El rango de valores que toma es

$$-90^\circ (90^\circ\text{S}) \leq \varphi \leq 90^\circ (90^\circ\text{N}).$$

La latitud geodésica puede llegar a diferir de la latitud geocéntrica hasta unos $11''.5$ a una latitud de 45° . La diferencia entre estas dos latitudes se conoce con el nombre de *ángulo de la vertical*.

- b) λ , *longitud geodésica* de un punto sobre la superficie terrestre: definida de la misma forma que la longitud geocéntrica, de forma que $\lambda = \lambda'$.
- c) h , *altura* de un observador sobre el elipsoide: distancia al esferoide medida a lo largo de la normal a dicho esferoide. En una primera aproximación se puede tomar h de un determinado lugar como su altura sobre el nivel del mar.

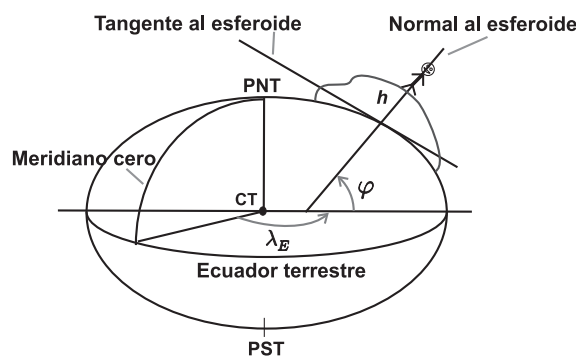


Figura 3.24: Coordenadas geodésicas.

Coordenadas astronómicas (o geográficas)

Cuando se determinan la latitud y la longitud mediante observaciones astronómicas, esto es, con respecto al polo celeste y al meridiano local a través de la vertical local, a los valores obtenidos de estos ángulos se les adiciona el adjetivo de geográficos (o también astronómicos). Las coordenadas astronómicas (o geográficas) son (véase la Figura 3.25):



- a) ϕ , *latitud geográfica* de un punto sobre la superficie terrestre: ángulo existente entre la dirección de la plomada (vertical local) en dicho punto y el ecuador terrestre. Puesto que la vertical local de un punto está afectada por las anomalías gravitacionales locales (montañas prominentes, depósitos subterráneos muy densos,...) y los campos gravitacionales cambiantes de la Luna, El Sol y los océanos –lo que implica que la vertical extendida hasta el centro de la Tierra no pasa por el centro del esferoide– existirá una pequeña diferencia en dirección entre la vertical de dicho punto y la normal al esferoide (la que define φ). La inclinación de la vertical local a la normal al esferoide de referencia se conoce con el nombre de *desviación de la vertical* o *deflexión vertical*. Por tanto, lo que diferencia la latitud geográfica de la latitud geodésica es la desviación de la vertical.

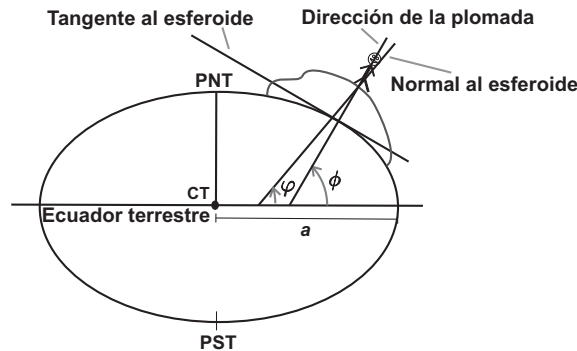


Figura 3.25: Coordenadas astronómicas (o geográficas).

- b) Λ , *longitud geográfica* de un punto sobre la superficie terrestre: ángulo entre el plano del meridiano astronómico de dicho punto y el plano del primer meridiano que pasa por Greenwich. El meridiano astronómico es el plano que pasa por el observador y contiene la vertical y una paralela a la dirección del eje de rotación. Como ya se apuntó, la vertical de un punto no necesariamente pasa por el centro del esferoide, por lo que el meridiano astronómico no coincide, por lo general, con el meridiano geodésico (que sí pasa por el centro del esferoide). De ahí que las longitudes geográfica y geodésica difieran entre sí una pequeña cantidad.

3.7 Posicionamiento utilizando satélites artificiales

La *geodesia espacial* se basa en la observación o recepción de señales, formadas por radiaciones del espectro electromagnético, procedentes de objetos que no estén físicamente ligados al suelo terrestre. Utiliza fundamentalmente los satélites artificiales aunque también se incluyen las técnicas de observación a globos en la atmósfera, a nuestro satélite natural y a cuásares extragalácticos. Los sistemas de satélites tienen diferentes aplicaciones como posicionamiento, vigilancia, comunicaciones, meteorología, geodesia, astronomía, geofísica, teledetección, etc. Sin entrar en los detalles de la geodesia espacial, indicaremos brevemente el principio de posicionamiento sobre la base de la utilización de satélites artificiales, en particular, el sistema de posicionamiento global (NAVSTAR–GPS) “NAVigation Satellite Timing And Ranging–Global Positioning System”, que se convirtió en uso corriente desde principios de los años 90 del siglo pasado.

El *problema de posicionamiento por satélites* consiste en obtener nuestra situación respecto al origen, o vector de posición (véase la Figura 3.26). Para ello debemos conocer la posición del satélite respecto al origen, o vector satélite, en cada instante de tiempo, y obtener, a partir de nuestras observaciones, el vector observación, entre el satélite y nuestra estación.

Componentes principales del sistema NAVSTAR–GPS

El sistema NAVSTAR–GPS está constituido por 3 componentes o segmentos diferenciados: *espacial*, de *control* y de *utilización*. A continuación, describimos cada una de ellas:

- a) *Componente espacial*: El sistema GPS está compuesto por una constelación de 24 satélites operativos (más cuatro en reserva) distribuida en 6 planos orbitales igualmente espaciados de 55° de inclinación con



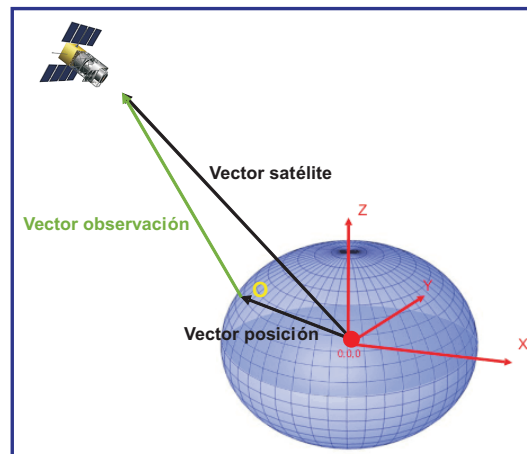


Figura 3.26: GPS.

respecto al plano ecuatorial. Cada plano contiene 4 satélites a una altitud media aproximada de 20180 km, en órbitas de baja excentricidad (de 0.03 a 0.3), con un periodo orbital de 11 horas y 58 minutos. Por ello, la misma configuración aparece cada día a la misma hora menos 4 minutos. Un receptor fijo en la Tierra observará a un satélite dado dos veces al día y habrá al menos 4 satélites observables desde la Tierra en cualquier momento. En cada satélite opera un reloj atómico de cesio y emite señales electromagnéticas que incluyen un código que señala al receptor la hora a la que se emite la señal y la posición del satélite con respecto a un sistema de referencia que gira con la Tierra y está centrado en ella (véase la Figura 3.27).

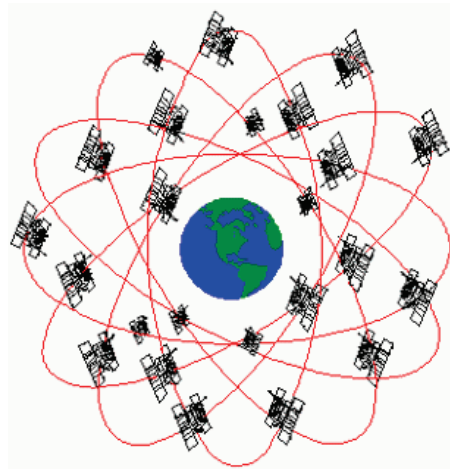


Figura 3.27: GPS: Componente espacial.

- b) *Componente de control*: Consta de 5 estaciones terrestres que controlan los satélites GPS comprobando su estado operativo y su posición exacta en el espacio. La estación principal está situada en Colorado Springs, la cual transmite correcciones sobre los datos de las órbitas de los satélites y sus relojes. Las otras 4 estaciones cuasiecuatoriales están en Hawaii (Pacífico oriental), Kwajalein (Pacífico occidental), Ascensión (Atlántico sur) y Diego García (Índico). Están regularmente espaciadas en longitud y tienen coordenadas muy precisas; recibiendo continuamente señales de los satélites que estén sobre el horizonte local con la que obtienen la información necesaria para establecer con gran precisión las órbitas de todos los satélites (véase la Figura 3.28).
- c) *Componente de utilización (segmento de usuario)*: Son los elementos e instrumentos utilizados para hallar coordenadas de un punto. El equipo esencial está formado por antena y receptor. A la antena siempre estarán referidas las coordenadas halladas, y va unida directamente o mediante un cable al

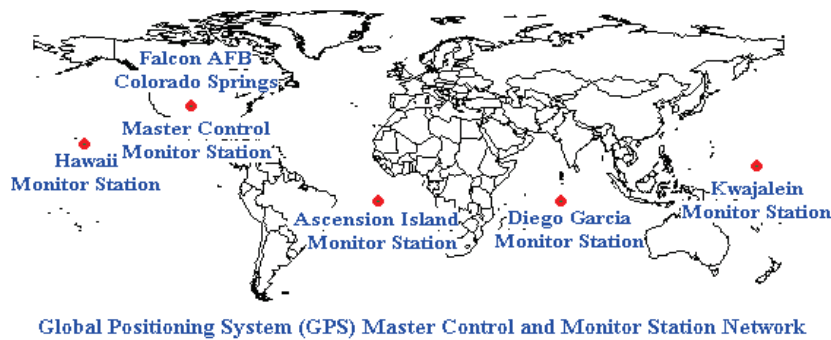


Figura 3.28: GPS: Componente de control.

receptor. Analizando señales de al menos 4 satélites, un receptor en la superficie de la Tierra puede conocer su posición (latitud, longitud y altitud) con una exactitud del orden del centímetro (véase la Figura 3.29).

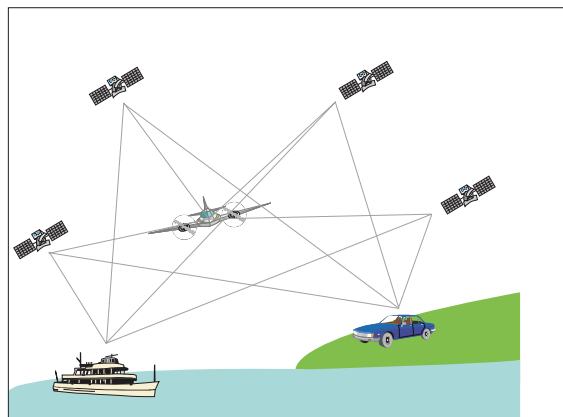


Figura 3.29: GPS: Componente de utilización.

Posicionamiento con GPS

El principio matemático básico del posicionamiento GPS es la *triangulación*. Para calcular la posición de un punto será necesario que el receptor GPS determine con exactitud la distancia que lo separa de los satélites. Para ello, el GPS mide el tiempo que tarda una señal (que contiene las coordenadas del satélite y la hora a la que ha sido emitida) en llegar a la Tierra y, utilizando el valor de la velocidad de la luz en la atmósfera, calcula la distancia entre el satélite y el receptor. Para poder medir este tiempo, los satélites necesitan disponer de relojes muy precisos.

Imaginemos un observador situado en la superficie de la Tierra, y un satélite S_1 que emite una señal en el instante t_1 . El observador recibe la señal en el instante t siendo $t - t_1$ el tiempo de propagación de la señal desde la posición ocupada por el satélite en el instante t_1 al observador supuesto fijo. Hagamos la hipótesis de que los relojes a bordo del satélite y el utilizado por el observador indican estrictamente la misma hora; en este caso, se puede concluir que el observador está situado sobre una esfera de radio $c(t - t_1)$, siendo c la velocidad de la luz (caso A de la Figura 3.30).

Consideramos un segundo satélite S_2 sincronizado con el primero, que emite una señal en un instante t_2 . Esta señal será captada por el observador en el mismo instante t que la señal anterior, siendo $t - t_2$ el tiempo que tarda en llegar la señal desde la posición ocupada por el satélite S_2 en el instante t_2 hasta el observador. Entonces, el observador estará en la intersección entre la primera esfera de radio $c(t - t_1)$ y otra centrada en el segundo satélite S_2 y de radio $c(t - t_2)$. La intersección de ambas esferas es un círculo, y en algún punto de ese círculo es donde se encuentra el observador (caso B de la Figura 3.30).



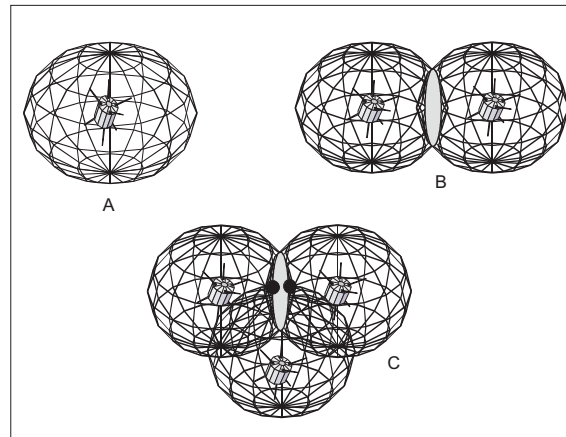


Figura 3.30: Posicionamiento con GPS.

Análogamente se considera un tercer satélite S_3 , sincronizado con los dos satélites anteriores. El observador estará situado en la intersección de 3 esferas teniendo cada una por centro las posiciones de los satélites S_1 , S_2 y S_3 en los instantes de emisión de la señal. La intersección de estas 3 esferas se reduce a dos puntos (caso C de la Figura 3.30). Para decidir cuál es la verdadera localización necesitamos de una cuarta medida. Generalmente uno de los dos puntos no es real debido a que está demasiado lejos de la Tierra o a una velocidad imposible. Como ya se ha indicado, cada punto de la Tierra recibe la señal de, al menos, cuatro satélites; el uso de estas mediciones adicionales será muy útil por las razones que se explican a continuación.

Supongamos que son 4 los satélites accesibles desde la posición del observador. Si (x, y, z) son las coordenadas cartesianas del observador y (x_i, y_i, z_i) para $i = 1, 2, 3, 4$ son las posiciones de los 4 satélites en los instantes de emisión de la señal, t_1, t_2, t_3 y t_4 , entonces la posición del observador resulta de la solución del sistema de 4 ecuaciones

$$\sqrt{(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2} = c(t - t_i), \quad i = 1, 2, 3, 4,$$

cuyas incógnitas son (x, y, z) y t . Para resolver este sistema, elevamos al cuadrado las cuatro ecuaciones, obteniendo

$$(x - x_i)^2 + (y - y_i)^2 + (z - z_i)^2 = c^2(t - t_i)^2, \quad i = 1, 2, 3, 4. \quad (3.5)$$

A continuación, las restamos dos a dos (primera menos segunda, segunda menos tercera y tercera menos cuarta), llegando al sistema lineal

$$\begin{cases} (x_2 - x_1)x + (y_2 - y_1)y + (z_2 - z_1)z - c^2(t_2 - t_1)t = b_1 \\ (x_3 - x_2)x + (y_3 - y_2)y + (z_3 - z_2)z - c^2(t_3 - t_2)t = b_2 \\ (x_4 - x_3)x + (y_4 - y_3)y + (z_4 - z_3)z - c^2(t_4 - t_3)t = b_3 \end{cases} \quad (3.6)$$

donde los segundos miembros

$$\begin{cases} b_1 = \frac{x_2^2 + y_2^2 + z_2^2 - (x_1^2 + y_1^2 + z_1^2) + c^2(t_1^2 - t_2^2)}{2} \\ b_2 = \frac{x_3^2 + y_3^2 + z_3^2 - (x_2^2 + y_2^2 + z_2^2) + c^2(t_2^2 - t_3^2)}{2} \\ b_3 = \frac{x_4^2 + y_4^2 + z_4^2 - (x_3^2 + y_3^2 + z_3^2) + c^2(t_3^2 - t_4^2)}{2} \end{cases}$$

son cantidades conocidas. Para encontrar la solución del sistema lineal (3.6) se considera t como parámetro y se calculan x, y, z en función de t . Para determinar t se sustituyen los valores obtenidos de x, y, z en cualquiera de las cuatro ecuaciones cuadráticas (3.5). Finalmente, con ese valor de t se determinan las coordenadas del observador (x, y, z) .



Ejemplo 3.7.1 Supongamos que a un receptor GPS le llegan las siguientes señales de cuatro satélites: la primera con la posición $(2,4,0)$ y la hora $5'75$; la segunda con la posición $(4,0,1)$ y la hora 11 ; la tercera con la posición $(2,2,3)$ y la hora 11 ; la cuarta con la posición $(4,2,0)$ y la hora $5'75$. Estamos suponiendo que la unidad de longitud es 10^4 km y que la unidad de tiempo es el milisegundo. En este caso, el valor de la velocidad de la luz (tomamos la del vacío) es $0'03$. El sistema de ecuaciones (3.6) puede escribirse como

$$\begin{cases} 2x - 4y + z - 0'03^2 5'25t = -1'5396 \\ -2x + 2y + 2z = 0 \\ 2x - 3z + 0'03^2 5'25t = 1'5396 \end{cases}$$

obteniéndose los valores

$$\begin{cases} x = -0'0024t + 0'7698 \\ y = -0'0024t + 0'7698 \\ z = 0. \end{cases} \quad (3.7)$$

Introduciendo estos valores de x, y y z en, por ejemplo, la primera de las ecuaciones cuadráticas

$$(x - 2)^2 + (y - 4)^2 + z^2 = 0'03^2(t - 5'75)^2$$

se hallan dos valores

$$t = \begin{cases} 134'8145 \\ -99'4587. \end{cases}$$

El valor negativo no tiene sentido (las señales llegan al receptor, obviamente, después de salir de los satélites). Sustituyendo el valor positivo de t en (3.7) obtenemos

$$\begin{cases} x = 0'4513 \\ y = 0'4513 \\ z = 0. \end{cases}$$

Se trata, al ser la coordenada z nula, de un punto que se encuentra en el ecuador (su latitud valdrá, por tanto, cero). La longitud se calcula mediante

$$\Lambda = \arctan \frac{y}{x} = \arctan 1.$$

Teniendo en cuenta que y es positiva y, por tanto, nos encontramos al este del meridiano cero, tendremos que el punto se encuentra a una longitud $\Lambda = 45^\circ$ E. Por último, calculando su distancia al origen obtenemos $0'6382$, es decir, 6382 km. Teniendo en cuenta que el radio del ecuador es de 6378 km, el punto está, aproximadamente, a 4000 metros sobre el nivel del mar. \square

Finalmente, debe señalarse que en la mayoría de ocasiones se cuenta con un número de satélites accesibles $n > 4$. Esto da lugar a n ecuaciones cuadráticas de la forma (3.5) y, restándolas dos a dos, a un sistema lineal del tipo 3.6 con $n - 1$ ecuaciones y 4 incógnitas. Se tiene, por tanto, un *sistema sobredeterminado* (más ecuaciones que incógnitas) que, en general, no tiene solución. Este tipo de problemas se resuelve mediante el método de *mínimos cuadrados* cuyo estudio excede los contenidos de este curso.

Próximamente, Europa estará dotada de un sistema llamado GALILEO, comparable al sistema americano GPS. El sistema GALILEO contará con 30 satélites en órbita media terrestre con una altura de 23222 km de la superficie terrestre, estarán distribuidos en 3 planos orbitales, inclinados 56° con respecto al ecuador (10 satélites en cada plano) y periodo de revolución de 14 horas.



3.8 Problemas

3.8.1 El valor exacto de $\tan 120^\circ$ es:

- a) $-\sqrt{3}$ b) $\frac{\sqrt{3}}{3}$ c) $\pm\infty$ d) Ninguna de las anteriores.

3.8.2 El valor exacto de $\sin(-135)^\circ$ es:

- a) 1 b) $-\frac{\sqrt{2}}{2}$ c) 0 d) Ninguna de las anteriores.

3.8.3 El valor exacto de $\cos(-330)^\circ$ es:

- a) $\frac{1}{2}$ b) $\frac{\sqrt{3}}{2}$ c) 0 d) Ninguna de las anteriores.

3.8.4 Cuando el Sol está a 30° por encima del horizonte, la longitud de la sombra proyectada por un edificio de 20 m de altura es:

- a) $\frac{20\sqrt{3}}{3}$ b) $20\sqrt{2}$ c) $20\sqrt{3}$ d) Ninguna de las anteriores.

3.8.5 El ángulo de elevación del Sol en el momento en que un edificio tiene la misma altura que la longitud de su sombra proyectada en un plano horizontal es:

- a) 45° b) 90° c) 30° d) Ninguna de las anteriores.

3.8.6 La expresión simplificada de $\tan^2 \alpha \cos^2 \alpha + \cot^2 \alpha \sin^2 \alpha$ es:

- a) $-\cos^2 \alpha$ b) $\sin^2 \alpha$ c) 1 d) Ninguna de las anteriores.

3.8.7 La expresión simplificada de $\tan \alpha + \frac{\cos \alpha}{1 + \sin \alpha}$ es:

- a) $\sin \alpha$ b) $\sec \alpha$ c) 0 d) Ninguna de las anteriores.

3.8.8 La expresión $\frac{\tan \alpha - \sin \alpha}{\sin^3 \alpha}$ es equivalente a:

- a) $\sec \alpha$ b) $\frac{\sec \alpha}{1 + \cos \alpha}$ c) $\frac{\sec \alpha}{1 - \cos \alpha}$ d) Ninguna de las anteriores.

3.8.9 Son soluciones de la ecuación trigonométrica $2 \sec \alpha = \tan \alpha + \cot \alpha$:

- a) 0° y 45° b) 90° y 180° c) 0° y 180° d) Ninguna de las anteriores.

3.8.10 Son soluciones de la ecuación trigonométrica $\tan 2\alpha + 2 \sin \alpha = 0$:

- a) 0° , 45° y 300° b) 90° , 180° y 300° c) 0° , 60° , 180° y 300° d) Ninguna de las anteriores.



3.8.11 Consideremos un triángulo de lados a , b , c y sean A , B , C los ángulos opuestos a los lados a , b , c , respectivamente, como se muestra en la Figura 3.31. Demostrar los siguientes resultados:

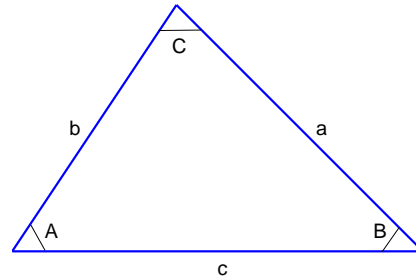


Figura 3.31: Triángulo de lados a , b y c .

a) *Teorema del seno:*

$$\frac{a}{\operatorname{sen} A} = \frac{b}{\operatorname{sen} B} = \frac{c}{\operatorname{sen} C}.$$

b) *Teorema del coseno:*

$$\begin{cases} a^2 = b^2 + c^2 - 2bc \cos A \\ b^2 = c^2 + a^2 - 2ca \cos B \\ c^2 = a^2 + b^2 - 2ab \cos C. \end{cases}$$

c) *Fórmulas de Mollweide:*

$$\begin{cases} \frac{a+b}{c} = \frac{\cos \frac{A-B}{2}}{\operatorname{sen} \frac{C}{2}}, & \frac{a-b}{c} = \frac{\operatorname{sen} \frac{A-B}{2}}{\cos \frac{C}{2}} \\ \frac{b+c}{a} = \frac{\cos \frac{B-C}{2}}{\operatorname{sen} \frac{A}{2}}, & \frac{b-c}{a} = \frac{\operatorname{sen} \frac{B-C}{2}}{\cos \frac{A}{2}} \\ \frac{c+a}{b} = \frac{\cos \frac{C-A}{2}}{\operatorname{sen} \frac{B}{2}}, & \frac{c-a}{b} = \frac{\operatorname{sen} \frac{C-A}{2}}{\cos \frac{B}{2}}. \end{cases}$$

d) *Fórmulas de proyección:*

$$\begin{cases} a = b \cos C + c \cos B \\ b = c \cos A + a \cos C \\ c = a \cos B + b \cos A. \end{cases}$$

3.8.12 Resolución de triángulos planos: Consiste en determinar, a partir de ciertos datos, los tres lados y los tres ángulos de un triángulo plano. Resolver los triángulos planos correspondientes a las siguientes situaciones:

a) *Conocidos tres lados:* se aplica el teorema del coseno.

Ejemplo: $a = 3$, $b = 4$ y $c = 5$.

b) *Conocidos dos lados y el ángulo comprendido entre ellos:* se aplica el teorema del coseno.

Ejemplo: $a = 3$, $b = 4$ y $C = 60^\circ$.

c) *Conocido un lado y los dos ángulos en sus extremos:* se aplica el teorema del seno.

Ejemplo: $a = 3$, $B = 45^\circ$ y $C = 105^\circ$.

d) Estudiar qué situaciones pueden darse cuando se conocen dos lados (por ejemplo, a y b) y un ángulo que no es el comprendido entre ellos (por ejemplo, A o B).

e) Estudiar qué situaciones pueden darse cuando se conocen un lado y los tres ángulos, pero no se sabe cuál de ellos es el opuesto al lado conocido.



3.8.13 Determinación de la distancia de un punto a un objeto inaccesible: Consideremos dos lugares de observación O_1 y O_2 y un punto A inaccesible. Supuesta conocida la distancia base $a = d(O_1, O_2)$ y medidos los ángulos con un *teodolito*²

$$\vartheta_1 = \angle(\overline{O_1O_2}, \overline{O_1A}) \text{ y } \vartheta_2 = \angle(\overline{O_2O_1}, \overline{O_2A}),$$

hallar las distancias de los puntos de observación al punto inaccesible (véase la Figura 3.32).

APLICACIÓN: $a = 20$ m, $\vartheta_1 = 60^\circ$ y $\vartheta_2 = 30^\circ$.

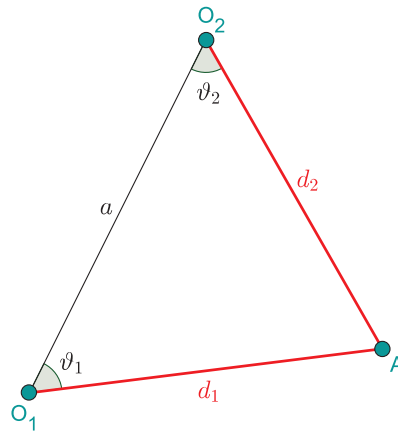


Figura 3.32: Distancia a un objeto inaccesible.

3.8.14 Determinación de la altura de un objeto inaccesible: Una torre está situada sobre una base a la que no se puede acceder (pues pasa un río, hay un obstáculo, ...). Sobre el nivel del suelo se sitúan dos teodolitos en los puntos C y D , separados una distancia $d = d(C, D)$. Estos puntos están localizados en el mismo plano que O y B (puntos inferior y superior de la torre). Desde C y D se miden, respectivamente, los ángulos α_1 y α_2 , tal y como se indica en la Figura 3.33. Determinar la altura $h = d(O, B)$ de la torre.

APLICACIÓN: $d = 50$ m, $\alpha_1 = 60^\circ$ y $\alpha_2 = 45^\circ$.

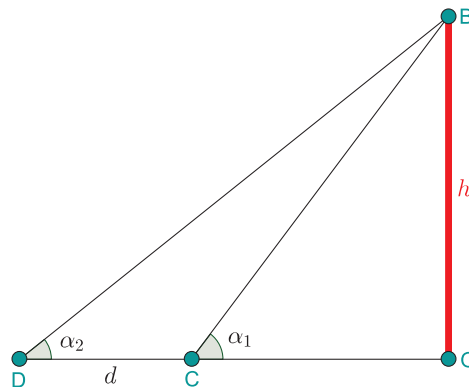


Figura 3.33: Altura de la torre.

3.8.15 Determinación de la altura de un objeto sobre una base inaccesible y de altura desconocida: Una torre está situada sobre una colina. Sobre el nivel del suelo se sitúan dos teodolitos en los puntos C y D , separados una distancia $d = d(C, D)$. Estos puntos están localizados en el mismo plano que A y B (puntos inferior y superior de la torre). Se miden desde C los ángulos α_1 y β_1 y desde D el ángulo α_2 , tal y como se indica en la Figura 3.34. Determinar la altura $h = d(A, B)$ de dicha torre.

APLICACIÓN: $d = 50$ m, $\alpha_1 = 60^\circ$, $\beta_1 = 30^\circ$ y $\alpha_2 = 45^\circ$.

²Instrumento de precisión que se compone de un círculo horizontal y un semicírculo vertical, ambos graduados y provistos de anteojos, para medir ángulos en sus planos respectivos.



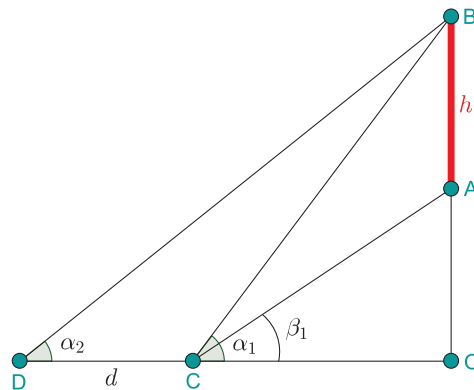


Figura 3.34: Altura de la torre de base inaccesible.

3.8.16 Problema de Snellius–Pothenot³: Sean O , A_1 y A_2 tres estaciones costeras no alineadas y M un navío observable desde estas tres estaciones que se encuentra situado en el plano (O, A_1, A_2) (véase la Figura 3.35). Conocidas las distancias $a_1 = d(O, A_1)$ y $a_2 = d(O, A_2)$ y los ángulos $\vartheta = \angle(\overline{OA_1}, \overline{OA_2})$, $\alpha_1 = \angle(\overline{MO}, \overline{MA_1})$ y $\alpha_2 = \angle(\overline{MO}, \overline{MA_2})$, hallar la distancia del barco a las tres estaciones costeras.

APLICACIÓN: $a_1 = 4$ km, $a_2 = 5$ km, $\vartheta = 55^\circ$, $\alpha_1 = 30^\circ$ y $\alpha_2 = 45^\circ$.

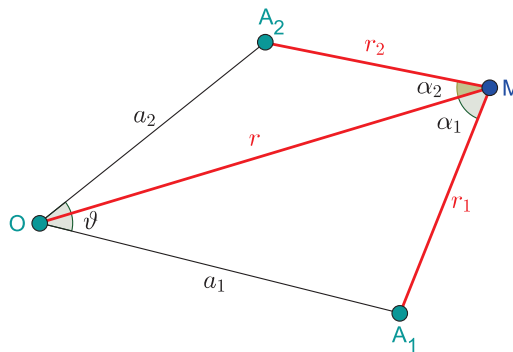


Figura 3.35: Problema de Snellius–Pothenot.

3.8.17 Problema de Hansen: Supongamos que A_1 y A_2 son dos estaciones costeras cuya distancia $a = d(A_1, A_2)$ es conocida y sean M y M' dos navíos observables desde estas dos estaciones. Si se miden los ángulos $\alpha = \angle(\overline{MA_1}, \overline{MA_2})$, $\beta = \angle(\overline{MA_2}, \overline{MM'})$, $\alpha' = \angle(\overline{M'A_1}, \overline{M'A_2})$ y $\beta' = \angle(\overline{M'A_1}, \overline{M'M})$ (véase la Figura 3.36), determinar las distancias:

- $r = d(M, M')$.
- $r_1 = d(A_1, M)$ y $r_2 = d(A_2, M)$.
- $r'_1 = d(A_1, M')$ y $r'_2 = d(A_2, M')$.

APLICACIÓN: $a = 1$ km, $\alpha = 28.75^\circ$, $\beta = 29^\circ$, $\alpha' = 28.25^\circ$ y $\beta' = 73^\circ$.

³Se trata de un problema de determinación de distancias que puede aplicarse a la navegación y fue resuelto en 1617 por Willebrord Snel van Royen (1580–1626), también conocido como Snellius, y por el francés Pothenot en 1692.



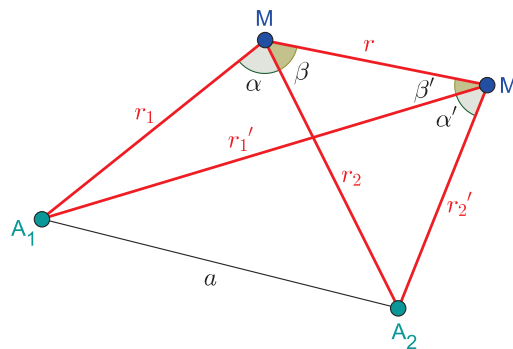


Figura 3.36: Problema de Hansen.

En los Problemas 3.8.18 a 3.8.23, ABC representa un triángulo esférico de vértices A, B, C y lados a, b, c como el de la Figura 3.37.

3.8.18 Demostrar los siguientes resultados:

a) *Fórmula del coseno para ángulos:*

$$\begin{cases} \cos A = -\cos B \cos C + \sin B \sin C \cos a \\ \cos B = -\cos C \cos A + \sin C \sin A \cos b \\ \cos C = -\cos A \cos B + \sin A \sin B \cos c. \end{cases}$$

b) *Analogía de Bessel para ángulos:*

$$\begin{cases} \sin A \cos b = \cos B \sin C + \sin B \cos C \cos a \\ \sin B \cos c = \cos C \sin A + \sin C \cos A \cos b \\ \sin C \cos a = \cos A \sin B + \sin A \cos B \cos c. \end{cases}$$

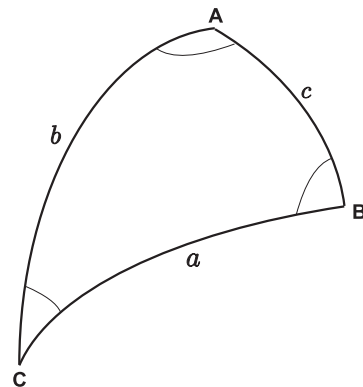


Figura 3.37: Triángulo esférico.

(**Indicación:** aplicar las fórmulas de Bessel al triángulo polar).

3.8.19 Demostrar la *fórmula de las cuatro partes* (o de las *cotangentes*):

$$\begin{cases} \cos a \cos B = \sin a \cot c - \sin B \cot C \\ \cos a \cos C = \sin a \cot b - \sin C \cot B \\ \cos b \cos A = \sin b \cot c - \sin A \cot C \\ \cos b \cos C = \sin b \cot a - \sin C \cot A \\ \cos c \cos A = \sin c \cot b - \sin A \cot B \\ \cos c \cos B = \sin c \cot a - \sin B \cot A. \end{cases}$$

(**Indicación:** aplicar la Analogía de Bessel para ángulos y la segunda fórmula de Bessel).

3.8.20 Establecer las fórmulas de Bessel y las de los Problemas 3.8.18 y 3.8.19 para un triángulo esférico recto en $C = 90^\circ$.

3.8.21 Establecer las fórmulas de Bessel y las de los Problemas 3.8.18 y 3.8.19 para un triángulo esférico recto en $c = 90^\circ$.

3.8.22 **Resolución de triángulos esféricos:** Consiste en determinar, a partir de ciertos datos, los tres lados y los tres ángulos de un triángulo esférico. Resolver los triángulos esféricos correspondientes a las siguientes situaciones:



- a) Conocidos tres lados: se aplica la primera fórmula de Bessel.
Ejemplo: $a = 145^{\circ}30'34.1''$, $b = 70^{\circ}32'35.8''$ y $c = 125^{\circ}30'17.6''$.
- b) Conocidos tres ángulos: se aplica la fórmula del coseno para ángulos.
Ejemplo: $A = 141^{\circ}28'21''$, $B = 43^{\circ}3'41''$ y $C = 79^{\circ}2'37''$.
- c) Conocidos dos lados y el ángulo comprendido entre ellos: se aplica la primera fórmula de Bessel.
Ejemplo: $a = 57^{\circ}22'11''$, $b = 72^{\circ}12'19''$ y $C = 94^{\circ}1'49''$.
- d) Conocido un lado y los dos ángulos en sus extremos: se aplica la fórmula del coseno para ángulos.
Ejemplo: $A = 170^{\circ}35'47.5''$, $B = 100^{\circ}50'25.3''$ y $c = 60^{\circ}22'30.5''$.

3.8.23 Razonar si son posibles las siguientes configuraciones para un triángulo esférico:

- a) $a = 70^{\circ}$, $b = 100^{\circ}$ y $c = 50^{\circ}$.
- b) $a = 65^{\circ}$, $b = 120^{\circ}$ y $c = 35^{\circ}$.
- c) $a = 100^{\circ}$, $b = 120^{\circ}$ y $c = 150^{\circ}$.
- d) $A = 60^{\circ}$, $B = 70^{\circ}$ y $C = 90^{\circ}$.
- e) $A = 60^{\circ}$, $B = 115^{\circ}$ y $C = 145^{\circ}$.
- f) $A = 60^{\circ}$, $B = 20^{\circ}$ y $C = 90^{\circ}$.

En caso afirmativo, hallar el exceso esférico, el defecto esférico y el área de cada triángulo.

3.8.24 Dado un triángulo esférico:

- a) Demostrar que si tiene dos lados iguales entonces tiene dos ángulos iguales.
- b) Si los tres lados son iguales, ¿es cierto que los tres ángulos también lo son? Razonar la respuesta.

3.8.25 Un barco navega 55 millas hacia el este a lo largo del paralelo de latitud $44^{\circ}30'$ N. Determinar el incremento en longitud.

3.8.26 Dos ciudades A y B están situadas sobre el mismo paralelo de latitud Φ . Sabiendo que la distancia existente entre ellas, medida sobre el círculo máximo que las une, es L kilómetros, expresar la distancia entre ellas medida sobre el paralelo, en función de Φ . Aplicarlo al caso en que $L = 1000$ km y $\Phi = 45^{\circ}$ N.

3.8.27 Conocidas las coordenadas geográficas de Berlín ($\Phi_B = 52^{\circ}31'7''$ N, $\Lambda_B = 13^{\circ}24'30''$ E) y San Petersburgo ($\Phi_{SP} = 59^{\circ}56'30''$ N, $\Lambda_{SP} = 30^{\circ}18'24''$ E), determinar la distancia en km entre las dos ciudades, considerando como valor del radio terrestre 6371 km.

3.8.28 Un barco navega 120 millas de un punto A ($\Phi = 45^{\circ}$ N, $\Lambda = 70^{\circ}$ W) a un punto B a lo largo de un círculo máximo con rumbo inicial 30° (o, lo que es lo mismo, N 30° E). Hallar la latitud y longitud de B.

3.8.29 Encontrar la distancia, rumbo inicial y rumbo final en el viaje desde Honolulu ($\Phi_H = 21^{\circ}18.3' N$, $\Lambda_H = 157^{\circ}52.3' W$) a San Francisco ($\Phi_{SF} = 37^{\circ}47.5' N$, $\Lambda_{SF} = 122^{\circ}25.7' W$).

3.8.30 Conocidas las coordenadas geográficas de Madrid ($\Phi_M = 40^{\circ}24' N$, $\Lambda_M = 3^{\circ}30' W$) y Buenos Aires ($\Phi_{BA} = 34^{\circ}35' S$, $\Lambda_{BA} = 58^{\circ}15' W$), determinar:

- a) La distancia entre las dos ciudades.
- b) El rumbo inicial y final en la ruta más corta entre Madrid y Buenos Aires.
- c) Las coordenadas geográficas del punto de la ruta anterior que equidista de ambas ciudades.



3.8.31 Un avión despegue de Londres (latitud 50° , longitud 0°) con su piloto en estado semiinconsciente, con rumbo desconocido (sobre un círculo máximo hacia el cuadrante sureste) y vuela una distancia indeterminada hasta que el piloto recupera el uso de sus facultades y vira 90° hacia la derecha. Al cabo de 1800 millas volando sobre un círculo máximo, vuelve a cruzar el meridiano de Greenwich y, en ese momento, lleva un rumbo de 240° (o, lo que es lo mismo, $W 30^\circ S$). ¿Cuál es su latitud en ese momento?

3.8.32 De un punto P situado en el Ecuador sale un buque A que después de recorrer 4800 millas alcanza la latitud de $60^\circ N$, encontrándose al este del meridiano de P . Otro buque B sale del mismo punto P con rumbo 30° y recorre 3600 millas. Determinar la distancia en millas entre ambos buques.

3.8.33 Sabiendo que las coordenadas geográficas de Madrid son: latitud $\Phi_M = 40^\circ 24' 30'' N$ y longitud $\Lambda_M = 3^\circ 41' 4'' W$, se pide:

- Hallar el punto A sobre el ecuador cuya distancia angular a Madrid es de 90° , y que se encuentra al este de Madrid.
- Un avión que se dirige desde Madrid hacia el punto A se ve obligado a aterrizar a los $\frac{2}{3}$ de su camino. Determinar la posición geográfica del punto de aterrizaje.

3.8.34 Un barco parte del punto A del meridiano de Greenwich que tiene latitud $\Phi = 30^\circ S$, con un rumbo inicial de 45° (navegando todo el tiempo por un círculo máximo).

- Encontrar las coordenadas del punto B en el que el barco cruzará el ecuador.
- Determinar el rumbo final en B .
- Si continúa navegando manteniendo el mismo rumbo, encontrar la latitud del punto C de la ruta que tiene longitud $\Lambda = 80^\circ E$.

3.8.35 Un objeto volador se mueve sobre un meridiano desde el Polo Norte hasta llegar al Ecuador. En ese momento gira 30° a la izquierda y recorre 3000 millas, también por un círculo máximo, antes de estrellarse. Se pide:

- Dibujar un triángulo esférico asociado al problema, describiendo los ángulos, puntos y círculos máximos que intervienen.
- Determinar la latitud del punto en que se estrella el objeto.
- Sabiendo que la longitud de dicho punto es $12^\circ E$, hallar la longitud del punto en que el objeto giró.

3.8.36 Se consideran tres estaciones de seguimiento espacial A , B y C en la superficie terrestre. La estación A se encuentra en el punto en el que el meridiano de Greenwich corta al ecuador, la estación B en el Polo Norte, mientras que la estación C está situada en el hemisferio sur, sobre el meridiano $45^\circ E$ y a una distancia de 3600 millas de la estación A . Con motivo del amerizaje de una cápsula espacial, el buque encargado de recogerla se sitúa en el corte del paralelo $30^\circ N$ con el meridiano $22^\circ E$. La cápsula cae al mar en un punto de coordenadas ($\Phi = 40^\circ 20' N$, $\Lambda = 27^\circ 40' E$).

- Determinar las coordenadas geográficas de C .
- Sabiendo que el buque navega a 25 millas/hora, calcular el tiempo que tardará en llegar al punto de amerizaje de la cápsula, así como el rumbo inicial que debe tomar el buque en su ruta hasta ella.

3.8.37 Un dirigible navega por un círculo máximo desde un punto A del hemisferio sur, con una velocidad de 200 millas/h y rumbo inicial $N 30^\circ E$, hasta un punto B en el ecuador, al cual llega con rumbo final $N 15^\circ E$. Determinar, razonando la respuesta:

- La diferencia de longitudes entre los puntos A y B .
- La latitud del punto A .



c) El tiempo que tarda en llegar de A hasta B.

3.8.38 Un buque zarpa de un puerto A situado en el hemisferio sur con un rumbo inicial de 35° y navega hasta cortar al ecuador en un punto B, momento en que gira 80° hacia la izquierda y sigue, siempre navegando por círculo máximo, hasta cortar al meridiano del puerto de origen en un punto C cuya longitud es $16^\circ 40' E$. Si el buque navega a 20 millas por hora y ha tardado 8 días en recorrer el tramo AB, determinar:

- El rumbo final en C.
- La distancia recorrida por el buque.
- Las coordenadas de los puntos A, B y C.

3.8.39 De un punto A situado en el ecuador y en el meridiano de longitud $60^\circ E$ sale un dirigible a las 8 de la mañana con una velocidad de 120 millas por hora recorriendo el ecuador hacia el oeste. De otro punto B cuya situación geográfica es $(45^\circ S, 80^\circ E)$ sale en su persecución un avión, con velocidad de 360 millas por hora y navegando por un círculo máximo, logrando alcanzarle en un punto C a las 6 de la tarde. Determinar:

- Las coordenadas del punto de encuentro C.
- La hora de salida del avión.

3.8.40 Se consideran dos puntos A y B, ambos de longitud $60^\circ E$, siendo B el punto medio del arco del meridiano que une A con el Polo Norte. De estos dos puntos salen, siguiendo un círculo máximo y en el mismo instante, dos aviones con destino al punto C del ecuador de longitud 0° . El avión que parte de A viaja a 400 millas por hora, llegando con un rumbo final de 225° .

- Determinar el rumbo inicial en A y la distancia entre A y C.
- Determinar la velocidad que debe llevar el avión que parte de B para que ambos aviones lleguen a C simultáneamente.

3.8.41 Se considera un sistema de coordenadas horizontales (A, h) y un sistema de coordenadas ecuatoriales horarias (H, δ) , ambos con el mismo origen O de latitud Φ .

- Demostrar que se tiene la siguiente relación entre las coordenadas horizontales y las ecuatoriales horarias:

$$\begin{pmatrix} \cos \delta \cos H \\ \cos \delta \sin H \\ \sin \delta \end{pmatrix} = A_{\Phi-90^\circ} \begin{pmatrix} \cos h \cos A \\ \cos h \sin A \\ \sin h \end{pmatrix}$$

donde, en general, A_θ es la matriz de rotación en \mathbb{R}^3 de ángulo θ en el plano perpendicular al segundo vector de la base

$$A_\theta = \begin{pmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{pmatrix}.$$

Dar una interpretación geométrica de este hecho.

- Utilizar el apartado anterior para determinar una relación análoga entre las coordenadas ecuatoriales horarias y las horizontales. (**Indicación:** ¿cuál es la inversa de la matriz A_θ ?).

3.8.42 Razonar si son posibles las siguientes configuraciones del triángulo esférico $P_N ZX$ (P_N posición del Polo Norte, Z posición del cénit del observador y X posición del astro) determinando, en caso afirmativo, el área, el defecto esférico y el exceso esférico del triángulo polar asociado:

- El acimut del astro es $A = 120^\circ$, su ángulo horario es $H = 4^h$ y el ángulo $\sphericalangle(P_N X Z)$ es de 45° .
- El astro tiene una altura $h = 60^\circ$ y una declinación $\delta = 45^\circ$ cuando es observado desde un lugar de latitud $\Phi = 10^\circ N$.



- c) El acimut del astro es $A = 120^\circ$, su ángulo horario es 6^h y el ángulo $\sphericalangle(P_N X Z) = 45^\circ$.
- d) Se observa el astro desde un lugar de latitud 50°N y tiene una altura de 30° y una declinación de 45° .
- e) El acimut del astro es $A = 80^\circ$, su ángulo horario es 8^h y el ángulo $\sphericalangle(P_N X Z) = 30^\circ$.

3.8.43 Desde el Observatorio de Oslo de latitud $\Phi = 60^\circ 12' 42.5''\text{N}$ se ha observado la estrella γ Draconis. Sabiendo que las coordenadas ecuatoriales horarias de esta estrella a las $5^h 7^m$ de tiempo sidéreo local son $H = 11^h 10^m 38.28^s$ y $\delta = 51^\circ 29' 23.64''$, calcular las coordenadas horizontales y la ascensión recta de la estrella.

3.8.44 A las $23^h 39^m 18.64^s$ de tiempo sidéreo local se observa en el horizonte de Sevilla, de latitud $\Phi = 37^\circ 23' 10''\text{N}$, la estrella Castor cuyo acimut es $A = 228^\circ 15' 23.52''$. Determinar las coordenadas ecuatoriales absolutas.

3.8.45 Con un teodolito situado en un lugar de latitud $\Phi = 80^\circ 23' 45.1''\text{N}$ se ha observado la estrella β Umi de declinación $\delta = 74^\circ 11' 54.37''$ obteniendo un acimut $A = 160^\circ 23' 9''$. Hallar el ángulo horario y la distancia cenital de dicha estrella.

3.8.46 Encontrar la latitud de un observador en el hemisferio norte si se conoce, en un instante determinado, el ángulo horario del Sol $H_\odot = 1^h 34^m 24^s$, la altura del Sol $h_\odot = 40^\circ 10'$ y la declinación del Sol $\delta_\odot = 15^\circ 38'$.

3.8.47 En un lugar de latitud $\Phi = 44^\circ 56' 45''\text{N}$ y tiempo sidéreo local $21^h 12^m 55^s$, un astro tiene como ascensión recta $\alpha = 18^h 42^m 55^s$ y declinación $\delta = 22^\circ 34' 2''$.

- a) Determinar sus coordenadas horizontales.
- b) Transcurridas 9 horas y media desde el instante anterior, ¿cuál será la declinación del astro? ¿Y su acimut?





4 Dinámica discreta

La dinámica se ocupa del estudio de los cambios a lo largo del tiempo. La dinámica discreta estudia dichos cambios en una sucesión de instantes de tiempo (años, meses, generaciones, . . .) de forma que el estado del sistema en un instante de tiempo se supone dependiente del estado en instantes anteriores. Para estudiar estos cambios se intenta encontrar una relación entre lo que ocurre en un determinado instante y lo que ocurrirá en el siguiente, es decir, encontrar una relación *causa-efecto*. Dicha relación es lo que se conoce como *modelo* del problema. Una vez establecido un modelo, se considera la teoría matemática necesaria para analizarlo y las herramientas gráficas y de cálculo que conduzcan a su resolución.

La modelización dinámica es el arte de modelizar fenómenos que cambian con el tiempo. El procedimiento general que suele seguirse para crear un modelo es el siguiente:

1. Identificamos una situación del mundo real que se quiere estudiar y hacemos hipótesis sobre esa situación.
2. Traducimos las hipótesis anteriores a relaciones matemáticas.
3. Utilizamos nuestros conocimientos matemáticos para analizar o resolver las relaciones anteriores.
4. Traducimos la solución obtenida al mundo real para aprender más acerca del modelo original.

Conviene tener en cuenta lo siguiente:

- a) La relación matemática no es, en general, la solución. Por ejemplo, para construir un cuadrado de 4 m^2 , la traducción matemática es $x^2 = 4$, donde x es la longitud de un lado; nótese que ésta no es la solución del problema, pues las posibles soluciones son $x = 2$ y $x = -2$.
- b) Debemos asegurarnos de que la solución obtenida tiene sentido en la situación considerada. Así, en el ejemplo anterior, $x = -2$ carece de sentido.
- c) Con bastante frecuencia ninguna de las soluciones matemáticas tiene sentido, por lo que debemos reconsiderar las hipótesis originales y rehacer todos los cálculos.

4.1 Sistemas dinámicos discretos. Terminología

En esta sección vamos a introducir los conceptos básicos de la teoría de sistemas dinámicos discretos. Comenzamos con algunos ejemplos:

Ejemplo 4.1.1 Supongamos que el 1 de Enero de 2010 abrimos una cuenta en el banco con 1000 € a interés compuesto del 10% anual. De esta forma, el 1 de Enero de 2011 tendremos 1100 € en nuestra cuenta (los 1000 € originales más el 10% de los 1000 € de interés). El 1 de Enero de 2012 no tendremos 1200 € sino que, para nuestra satisfacción, tendremos 1210 € en nuestra cuenta (los 1100 € que teníamos más el 10% de los 1100 € de interés). Destacamos lo siguiente:

- a) El conocimiento de lo que es cierto hoy (tenemos 1000 € en el banco).
- b) La repercusión que tienen los actos actuales en el futuro (podemos predecir cuánto dinero tendremos en nuestra cuenta en cualquier instante futuro¹).

Traslademos la situación anterior al lenguaje de las matemáticas: el 1 de Enero de 2010 es el tiempo 0, el 1 de Enero de 2011 es el tiempo 1, el 1 de Enero de 2012 es el tiempo 2, . . . De esta forma, el 1 de Enero de 2028 es el tiempo $2028 - 2010 = 18$ (los demás días son irrelevantes, pues el dinero de nuestra cuenta permanece inalterable hasta el 1 de Enero del siguiente año). Denotemos por $A(0)$ la cantidad (en euros)

¹Excepto factores desconocidos como tasas, cambios en los tipos de interés o quiebra del banco.



que tenemos en nuestra cuenta en el tiempo 0, es decir, $A(0) = 1000$. De la misma forma, $A(1) = 1100$ y $A(2) = 1210$. Nuestro objetivo es *predecir* cuánto dinero tendremos en nuestra cuenta en el tiempo n , siendo n un año futuro. Por ejemplo, si $n = 18$ (año 2028), ¿cuánto es $A(18)$? Puesto que la cantidad en el tiempo 1 es la cantidad en el tiempo 0 más el interés de la cuenta en el tiempo 0 (10 % de $A(0)$), podemos escribir

$$A(1) = A(0) + 0'1 A(0) = 1'1 A(0).$$

Análogamente,

$$A(2) = 1'1 A(1), A(3) = 1'1 A(2), \dots, A(18) = 1'1 A(17)$$

y, así, sucesivamente. Necesitamos una expresión más manejable para las expresiones anteriores. Textualmente, las ecuaciones anteriores dicen que “la cantidad en nuestra cuenta el próximo año es la cantidad que tenemos en nuestra cuenta este año más el interés de la cuenta en este año”. Si éste es el año n , el año próximo es el año $n + 1$; si denotamos, respectivamente, la cantidad en el banco en cada uno de estos años por $A(n)$ y $A(n + 1)$, la sentencia anterior se expresa matemáticamente como

$$A(n + 1) = A(n) + 0'1 A(n) = 1'1 A(n), \quad n = 0, 1, 2, \dots \quad (4.1)$$

Nótese que la ecuación (4.1) no resuelve nuestro problema, pero le redefine matemáticamente. Para resolver nuestro problema necesitamos algún método que permita determinar la cantidad de euros que tenemos en nuestra cuenta en cualquier instante de tiempo futuro.

- Un método para encontrar una solución es el cálculo directo: con vistas a calcular $A(18)$, puesto que $A(0) = 1000$, por sustitución directa obtenemos que

$$A(1) = 1'1 A(0) = 1100.$$

Repitiendo este proceso, se obtiene que

$$A(2) = 1'1 A(1) = 1210, \quad A(3) = 1'1 A(2) = 1331$$

y, así, sucesivamente. Por este procedimiento llegaríamos a calcular $A(18)$ de una forma larga y tediosa.

- En lugar de esto, podemos utilizar un ordenador: escribiendo un programa sencillo (en casi cualquier lenguaje de programación), podríamos calcular $A(100)$ o $A(1000)$ rápidamente.
- No obstante, podemos obtener la mejor solución de (4.1) (en el sentido de ser la más cómoda para trabajar) haciendo las sustituciones

$$A(2) = 1'1 A(1) = (1'1)^2 A(0), \quad A(3) = 1'1 A(2) = (1'1)^3 A(0)$$

y, en general,

$$A(n) = (1'1)^n A(0), \quad n = 0, 1, 2, \dots \quad (4.2)$$

La ecuación (4.2) es lo que nosotros entenderemos por “solución” de nuestro problema. Utilizaremos esta ecuación y un ordenador (o calculadora) para obtener, de una forma rápida y sencilla, los euros que tendremos en nuestra cuenta en cualquier instante de tiempo futuro. En particular, para el caso que nos ocupa, se tiene que

$$A(18) = (1'1)^{18} 1000 \approx 5560 \text{€}. \quad \square$$

El ejemplo anterior es significativo del resto de este capítulo, en el sentido de que se utiliza la información en un determinado instante para deducir lo que ocurrirá en instantes posteriores. Este tipo de problemas aparece en crecimiento de poblaciones, genética o juegos de azar, por citar algunas situaciones.

Una vez analizado un problema, el siguiente paso es su generalización. La razón de esto es que las generalizaciones son, frecuentemente, tan sencillas de estudiar como los casos particulares y tienen la ventaja de que se pueden aplicar en situaciones más generales. Con vistas a generalizar el Ejemplo 4.1.1 supongamos que el interés del banco no es el 10 % sino el $I = 100i$ % (en el caso anterior, $i = 0'1$). La cantidad de



dinero que tendremos el próximo año será la cantidad que tenemos este año más i veces de la cantidad que tenemos este año, es decir,

$$A(n+1) = A(n) + iA(n) = (1+i)A(n), \quad n = 0, 1, 2, \dots \quad (4.3)$$

Nótese que la ecuación (4.3) coincide con (4.1) cuando $i = 0$. Desde el punto de vista matemático es tan sencillo trabajar con la ecuación (4.3) como con la ecuación (4.1). Además, la ecuación (4.3) tiene la ventaja de que, si el banco cambia el tipo de interés, sabremos lo que ocurrirá con nuestro dinero sin tener que realizar trabajo adicional. De forma análoga a como se hizo en el Ejemplo 4.1.1, la solución de la ecuación (4.3) es

$$A(n) = (1+i)^n A(0), \quad n = 0, 1, 2, \dots \quad (4.4)$$

Ejemplo 4.1.2 Consideremos un cultivo celular en el que las células se reproducen por mitosis de forma que, por cada célula aparecen dos nuevas células transcurrida una unidad de tiempo y, supongamos también, que en el tiempo considerado las células no se mueren. La manera más sencilla de modelizar este problema es considerar el cambio en el tamaño del cultivo en una unidad de tiempo. Sea $A(0) = a_0$ el número de células que hay en el instante $t = 0$ (el tiempo en el que comenzamos nuestras observaciones). De esta forma, la diferencia $A(1) - A(0)$ representa el cambio en la cantidad de células tras la primera unidad de tiempo; pero este cambio es igual al número de nuevas células, es decir, el número de células que han aparecido en ese periodo de tiempo. Por tanto,

$$A(1) - A(0) = A(0).$$

Análogamente, el cambio desde el periodo 1 al periodo 2, $A(2) - A(1)$, es igual al número de células nacidas en la población de tamaño $A(1)$, es decir,

$$A(2) - A(1) = A(1).$$

Continuando esta argumentación, se obtiene que el cambio en el tamaño del cultivo en un determinado periodo de tiempo es igual al tamaño que tenía la población en el inicio del periodo, esto es,

$$A(n+1) - A(n) = A(n)$$

o, equivalentemente,

$$A(n+1) = 2A(n), \quad n = 0, 1, \dots$$

Si suponemos que inicialmente en el cultivo hay $A(0) = 100$ células, para los sucesivos periodos de tiempo se van obteniendo los tamaños

$$A(1) = 200, A(2) = 400, A(3) = 800, A(4) = 1600, A(5) = 3200, A(6) = 6400, \dots$$

Estos valores crecen exponencialmente a infinito (no es difícil conjeturar que $A(n) = 100 \times 2^n$); de hecho, por ejemplo, $A(100)$ es del orden de 10^{30} , por lo que este modelo no parece muy realista.

Para mejorar nuestro modelo, supongamos que el número de células que surgen en un periodo n es proporcional al tamaño de la población en ese periodo (es decir, el número de nacimientos en el periodo n es igual a $bA(n)$, siendo b la *tasa de natalidad*) y que el número de muertes en el periodo n es proporcional al tamaño de la población en ese periodo (es decir, el número de muertes en el periodo n es igual a $dA(n)$, siendo d la *tasa de mortalidad*). Puesto que el cambio en el tamaño del cultivo en un periodo es igual al balance entre nacimientos y muertes, se tiene que

$$A(n+1) - A(n) = bA(n) - dA(n)$$

o, equivalentemente,

$$A(n+1) = (1+r)A(n), \quad n = 0, 1, \dots$$

siendo

$$r = b - d$$



la *tasa neta* del crecimiento del cultivo. Si suponemos que inicialmente $A(0) = 100$ y consideramos un valor razonable de la tasa neta de crecimiento, por ejemplo, $r = 0'2$, se obtiene que $A(10) = 619'174$, $A(20) = 3833'760$, $A(50) = 910043'815$ y $A(100) = 8281797452'201$ (de hecho, $A(n)$ va a infinito exponencialmente). Esta es la teoría de Malthus, según la cual las poblaciones crecen exponencialmente; Malthus predijo una catástrofe mundial que todavía no ha ocurrido. En la Sección 4.4.2 estudiaremos un modelo más adecuado para el crecimiento de una población. \square

Ejemplo 4.1.3 Leonardo de Pisa (más conocido por Fibonacci) planteó y resolvió el siguiente problema sobre el crecimiento de una población de conejos: “Una pareja de conejos tarda un mes en alcanzar la edad fértil, a partir de ese momento cada mes engendra una pareja de conejos que, a su vez, tras ser fértiles engendrarán cada mes una pareja de conejos. ¿Cuántos conejos habrá al cabo de un determinado número de meses?”. En el instante inicial (el nacimiento de la pareja original) tenemos una pareja de conejos. Obviamente, al finalizar el primer mes seguirá habiendo en la población una única pareja de conejos. Trascorridos 2 meses habrá 2 parejas (la original y su descendencia). Al cabo de tres meses habrá 3 parejas: las dos que teníamos más la descendencia de la original. En general, en cada mes tendremos tantas parejas como había el mes anterior más las descendientes de las parejas que había hace dos meses. Se obtiene así la sucesión de Fibonacci, que se genera mediante

$$A(n+2) = A(n+1) + A(n), \quad n = 0, 1, 2, \dots,$$

con $A(0) = 1$ y $A(1) = 1$. Así pues, el número de parejas de conejos es, en los sucesivos meses,

$$1, 1, 2, 3, 5, 8, 13, 21, \dots$$

Más adelante regresaremos sobre este ejemplo. \square

Las ecuaciones que describen una relación entre un instante de tiempo e instantes de tiempo anteriores (como las ecuaciones (4.1) y (4.3)) se denominan *sistemas dinámicos discretos* o *ecuaciones en diferencias*. Las expresiones explícitas (o en forma cerrada) (como las ecuaciones (4.2) y (4.4)) se denominan *soluciones* del sistema dinámico correspondiente. De manera informal, un sistema dinámico es una sucesión de números definida de forma recursiva, es decir, hay una regla que relaciona cada término de la sucesión con términos anteriores de la misma. Por ejemplo, para la sucesión de potencias de dos $1, 2, 4, 8, 16, \dots$ la regla que los determina es $A(0) = 1$ y

$$A(n+1) = 2A(n), \quad n = 1, 2, \dots$$

es decir, cada término de la sucesión es el doble que el anterior. Como acabamos de ver, la sucesión se obtiene a partir de la regla de formación y el primer término. Veamos otro ejemplo: a partir de

$$\begin{cases} A(0) = 0'1 \\ A(n+1) = 2A(n)(1 - A(n)), \quad n \in \mathbb{N} \cup \{0\} \end{cases}$$

se obtiene la sucesión

$$A(1) = 2A(0)(1 - A(0)) = 0'18, \quad A(2) = 2A(1)(1 - A(1)) = 0'2952,$$

$$A(3) = 2A(2)(1 - A(2)) = 0'4161, \quad A(4) = 2A(3)(1 - A(3)) = 0'4859, \dots$$

Los sistemas dinámicos pueden venir dados en muchas formas distintas pero, como veremos, diferentes tipos de ecuaciones (aparentemente distintas) pueden ser tratadas de forma parecida. Por esto dividiremos estos sistemas en grandes clases y estudiaremos cada una de ellas por separado.

Definición 4.1.4 Un *sistema dinámico discreto de primer orden* es una sucesión de números $\{A(n)\}_{n=0}^{\infty}$ de forma que cada elemento de la misma está relacionado con el anterior mediante una relación del tipo

$$A(n+1) = f(A(n), n) \tag{4.5}$$

para alguna función f . \square



Observación 4.1.5

- a) El sistema dinámico es de primer orden porque cada elemento de la sucesión depende sólo del elemento anterior.
- b) En todo lo que sigue, con bastante frecuencia omitiremos el término *discreto* (hablaremos, simplemente, de *sistemas dinámicos*) e identificaremos el sistema dinámico con la regla que lo define. □

Ejercicio 4.1.6 Encontrar los primeros seis valores $\{A(n)\}_{n=1}^6$ del sistema dinámico

$$A(n+1) = 1 + 2A(n), \quad n = 0, 1, 2, \dots$$

si $A(0) = 5$.

SOLUCIÓN.

$$A(1) = 1 + 10 = 11, \quad A(2) = 1 + 22 = 23, \quad A(3) = 1 + 46 = 47, \\ A(4) = 1 + 94 = 95, \quad A(5) = 1 + 190 = 191 \quad \text{y} \quad A(6) = 1 + 382 = 383. \quad \square$$

Definición 4.1.7 Consideremos un sistema dinámico de primer orden de la forma (4.5).

- a) Llamaremos *solución general* del sistema dinámico (4.5) a la familia de todas las sucesiones $\{A(n)\}_{n=0}^{\infty}$ que satisfacen el sistema dinámico cuando se sustituyen los valores $A(n)$ y $A(n+1)$.
- b) Llamaremos *solución particular* del sistema dinámico (4.5) con dato inicial $A(0) = a_0$ a una sucesión $\{A(n)\}_{n=0}^{\infty}$ que satisfice el sistema dinámico cuando se sustituyen los valores de $A(n)$ y $A(n+1)$ y, además, satisface la condición inicial $A(0) = a_0$. □

Observación 4.1.8 En la solución general de un sistema dinámico de primer orden suele aparecer una constante arbitraria que queda determinada cuando se fija un valor inicial $A(0)$. □

Definición 4.1.9 Un sistema dinámico discreto de primer orden se denomina *lineal de coeficientes constantes* si es de la forma

$$A(n+1) = rA(n) + b, \quad n = 0, 1, 2, \dots$$

con $r, b \in \mathbb{R}$. Diremos que el anterior sistema es:

- a) *homogéneo* si $b = 0$.
- b) *afín* si $b \neq 0$. □

Ejemplo 4.1.10 De los sistemas dinámicos de primer orden

$$A(n+1) = 3A(n) \quad \text{y} \quad A(n+1) = 2A(n) + 5$$

(las funciones que los definen son, respectivamente, $f(x) = 3x$ y $f(x) = 2x + 5$), el primero es lineal homogéneo y el segundo es afín. □

Ejemplo 4.1.11 La solución general del sistema dinámico lineal

$$A(n+1) = 3A(n), \quad n = 0, 1, 2, \dots \quad (4.6)$$

es

$$A(n) = c3^n, \quad n = 0, 1, 2, \dots \quad (4.7)$$

con $c \in \mathbb{R}$. En efecto, basta observar que

$$A(n+1) = c3^{n+1} = 3c3^n = 3A(n), \quad n = 0, 1, 2, \dots$$

La solución particular del sistema dinámico (4.6) con dato inicial

$$A(0) = 2$$

es

$$A(n) = 2 \times 3^n, \quad n = 0, 1, 2, \dots$$

pues basta elegir la constante c de (4.7) de forma que $c = c3^0 = A(0) = 2$. □



Definición 4.1.12 Un sistema dinámico discreto de orden 2 es una sucesión de números $\{A(n)\}_{n=0}^{\infty}$ de forma que

$$A(n+2) = f(A(n+1), A(n), n) \quad (4.8)$$

para alguna función f . (Esta definición se extiende fácilmente a sistemas de orden $m \in \mathbb{N}$). \square

Ejemplo 4.1.13

- a) $A(n+2) = A(n+1) + A(n)$ es un sistema dinámico de segundo orden.
- b) $A(n+3) = A(n+1) - 5A(n)$ es un sistema dinámico de tercer orden. Aunque el término $A(n+2)$ no aparece, los términos importantes para determinar el orden son el primero y el último utilizados (en este caso, $A(n)$ y $A(n+3)$). \square

Consideremos el sistema dinámico de segundo orden

$$A(n+2) = 2A(n+1) - A(n), \quad n = 0, 1, 2, \dots \quad (4.9)$$

que es equivalente a las ecuaciones

$$\begin{cases} A(2) = 2A(1) - A(0), \\ A(3) = 2A(2) - A(1) = 2(2A(1) - A(0)) - A(1) = 3A(1) - 2A(0), \\ A(4) = 2A(3) - A(2) = 2(3A(1) - 2A(0)) - (2A(1) - A(0)) \\ \quad = 4A(1) - 3A(0), \dots \end{cases}$$

Como se observa, los términos $A(0)$ y $A(1)$ son necesarios para determinar cualquier término $A(n)$ de la sucesión anterior. Por ejemplo, si $A(0) = 2$ y $A(1) = 3$, los primeros términos del sistema dinámico (4.9) son

$$A(0) = 2, A(1) = 3, A(2) = 4, A(3) = 5, A(4) = 6, \dots$$

(compruébese; de hecho, $A(n) = n + 2$). Los valores $A(0) = 2$ y $A(1) = 3$ se denominan *valores iniciales* del correspondiente sistema dinámico. Para encontrar los valores de $A(n)$ para todo n se necesitan algunos valores iniciales para que el sistema pueda comenzar; esto es debido a que, en sistemas dinámicos de orden superior, la primera ecuación siempre tiene más de una incógnita, por lo que debemos determinar todas, salvo una, de esas incógnitas. En general, para un sistema dinámico de orden $m \in \mathbb{N}$ debemos dar m *valores iniciales*.

4.2 Sistemas dinámicos lineales de primer orden

Cuando se modelizan problemas del mundo real se obtienen modelos complejos (en general, no lineales) acordes con la realidad, pero difíciles de resolver o sin solución explícita. Una primera aproximación a la solución del problema pasa por considerar los aspectos más básicos de la situación, lo que conduce a modelos más simples que pueden dar lugar a sistemas dinámicos lineales (posiblemente no homogéneos). Los sistemas lineales son, en general, fáciles de resolver; con ellos se puede obtener la solución de un problema matemático, aunque ésta corresponderá a un modelo no totalmente realista. Muchas de las aplicaciones que veremos se modelizan razonablemente mediante sistemas dinámicos lineales; en otros casos, el uso de sistemas lineales debe entenderse como un primer paso en el camino de la solución, como una aproximación al modelo no lineal.

La primera parte de esta sección está dedicada a estudiar el comportamiento cualitativo de las soluciones de los sistemas lineales, la segunda a encontrar soluciones de los mismos y, en la última, presentamos distintas aplicaciones que se modelizan mediante este tipo de sistemas.



4.2.1 Puntos de equilibrio: estabilidad, análisis gráfico

En esta sección introducimos el concepto de punto de equilibrio de un sistema dinámico discreto, el cual sólo tiene sentido en los denominados sistemas autónomos:

Definición 4.2.1 Un sistema dinámico discreto de primer orden es *autónomo* si la función f de (4.5) no depende de n (es decir, cuando $A(n+1) = f(A(n))$). En caso contrario, el sistema dinámico se llama *no autónomo*. □

Ejemplo 4.2.2 Supongamos que pedimos prestado a un amigo $A(0)$ euros a un interés de 5% anual y que le pagamos 100 € al año. El sistema dinámico que describe lo que le debemos a nuestro amigo cada año es

$$A(n+1) = 1.05 A(n) - 100. \quad (4.10)$$

- Por ejemplo, si la deuda original es $A(0) = 1000$ € entonces

$$A(1) = 950, A(2) = 897.5, A(3) = 842.4, A(4) = 784.5, A(5) = 723.7, \dots$$

Este comportamiento hace pensar que la deuda va decreciendo; cuando alcancemos un año n en el que $A(n)$ sea negativo entonces habremos cancelado nuestra deuda.

- En cambio, si la deuda original es $A(0) = 3000$ € entonces

$$A(1) = 3050, A(2) = 3102.5, A(3) = 3157.6, A(4) = 3215.5, A(5) = 3276.3, \dots$$

Ahora parece que deberemos a nuestro amigo cada vez más y más dinero y nunca podremos cancelar nuestra deuda.

El problema es que si restringimos nuestros pagos a 100 € al año existe un límite superior de dinero que podemos pedir si queremos llegar, en algún momento, a devolver el dinero que nos han prestado. En el caso que nos ocupa este límite son 2000 €, de forma que nuestros pagos mensuales son iguales a la deuda contraída, pues si $A(0) = 2000$ € entonces

$$A(1) = 2000, A(2) = 2000, A(3) = 2000, \dots$$

y, en general,

$$A(n) = 2000, n \in \mathbb{N}. \quad (4.11)$$

La solución constante (4.11) se dice que es un “punto de equilibrio” (o “punto fijo”) del sistema dinámico (4.10). Los puntos de equilibrio de un sistema dinámico son de vital importancia pues, a menudo, nos dicen qué pasará eventualmente en nuestro sistema. □

Definición 4.2.3 Un número a es un *punto de equilibrio* (o *punto fijo*) del sistema autónomo

$$A(n+1) = f(A(n)), n = 0, 1, 2, \dots \quad (4.12)$$

si $A(n) = a$ es una solución constante de (4.12). Es decir, $A(n) = a$ para todos los valores de $n \in \mathbb{N}$ cuando el dato inicial es $A(0) = a$. □

Es relativamente sencillo encontrar puntos de equilibrio a de un sistema dinámico autónomo.

Proposición 4.2.4 La condición necesaria y suficiente para que un número a sea punto de equilibrio del sistema dinámico (4.12) es que $a = f(a)$.

DEMOSTRACIÓN. Basta observar que si $A(n) = a$ entonces $A(n+1) = a$ para todo $n \in \mathbb{N} \cup \{0\}$. □

Ejemplo 4.2.5 Para encontrar los puntos de equilibrio del sistema dinámico

$$A(n+1) = 2A(n) - 3 \quad (4.13)$$

basta resolver la ecuación $a = 2a - 3$. Por tanto, $A(n) = 3$ es el único punto de equilibrio del sistema anterior. Nótese que si $A(0) = 3$ entonces $A(1) = 2A(0) - 3 = 6 - 3 = 3$, $A(2) = 2A(1) - 3 = 6 - 3 = 3$ y, en general, $A(n) = 3$ para todo $n \in \mathbb{N}$. □



En general, se tiene el siguiente resultado:

Proposición 4.2.6 *El único punto de equilibrio del sistema dinámico afín*

$$A(n+1) = rA(n) + b \quad (4.14)$$

con $r \neq 1$ es

$$a = \frac{b}{1-r}. \quad (4.15)$$

DEMOSTRACIÓN. Basta observar que (4.15) es la única solución de la ecuación $a = ra + b$ con $r \neq 1$. \square

Observación 4.2.7 Nótese que el sistema dinámico (4.14) no tiene puntos de equilibrio cuando $r = 1$ (pues la ecuación $a = a + b$ con $b \neq 0$ carece de soluciones). \square

Observación 4.2.8 Un sistema dinámico puede tener muchos puntos de equilibrio. En general, cuanto más “no lineal” sea un sistema dinámico, más puntos de equilibrio puede tener. Por ejemplo, el sistema dinámico $A(n+1) = A^2(n)$ tiene dos puntos de equilibrio: $a = 0$ y $a = 1$ (soluciones de la ecuación $a - a^2 = 0$). En general, habrá que considerar el número de ceros que tenga la función $x - f(x)$. \square

Definición 4.2.9 Un punto de equilibrio a del sistema dinámico (4.12) es:

a) *estable (atractivo o atractor)* si existe $\varepsilon > 0$ de forma que si $|A(0) - a| < \varepsilon$ entonces

$$\lim_{n \rightarrow +\infty} A(n) = a.$$

b) *inestable (repulsivo o repulsor)* si existe $\varepsilon > 0$ de forma que si $0 < |A(0) - a| < \varepsilon$ entonces

$$|A(n) - a| > \varepsilon$$

para algún valor de n (aunque no necesariamente todos).

c) *neutro* si no es estable ni inestable. \square

Observación 4.2.10 Puesto que $|A(n) - a|$ mide la distancia que hay entre a y $A(n)$, la definición anterior establece que un punto de equilibrio es estable si la sucesión $\{A(n)\}_{n=0}^{\infty}$ tiende a a siempre que $A(0)$ esté suficientemente próximo a a ; análogamente, un punto de equilibrio es inestable cuando, sin importar cómo de cercano esté $A(0)$ de a , el valor $A(n)$ permanece alejado (al menos una cantidad ε) de a (normalmente, estos valores $A(n)$ o van a infinito o bien se acercan a otro punto de equilibrio). Una de las partes más importantes de este capítulo es la dedicada al estudio de la estabilidad de los puntos de equilibrio puesto que para un gran número de aplicaciones es importante determinar el comportamiento de las soluciones para “tiempos largos”. \square

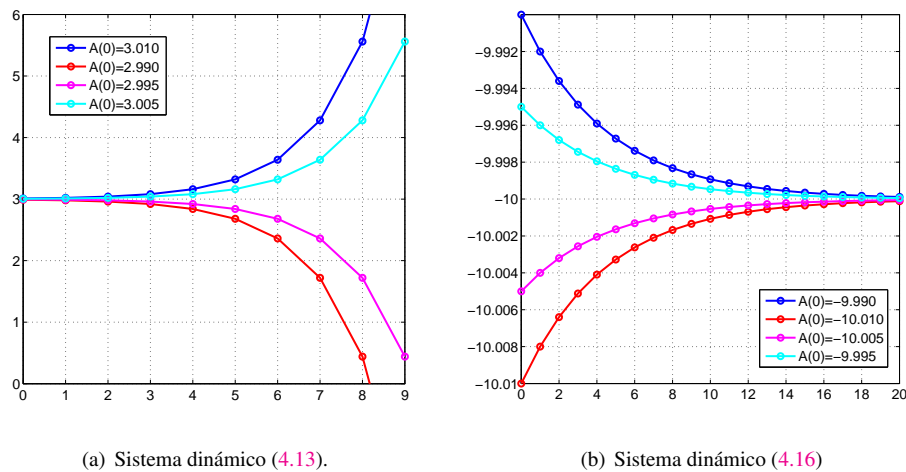
Veamos algunos ejemplos relativos a la estabilidad de puntos de equilibrio.

Ejemplo 4.2.11 Como hemos visto en el Ejemplo 4.2.5, $a = 3$ es el único punto de equilibrio del sistema dinámico (4.13).

En la Tabla 4.1 se muestran los primeros valores $A(n)$ del sistema dinámico (4.13) para cuatro datos iniciales distintos, cercanos al punto de equilibrio: $A(0) = 3'01$, $A(0) = 2'99$, $A(0) = 2'995$ y $A(0) = 3'005$. Asimismo, en la parte izquierda de la Figura 4.1 se muestran los puntos $(n, A(n))$ del sistema dinámico (4.13) para los datos iniciales anteriores (los puntos $(n, A(n))$ se han unido mediante segmentos que no tienen ningún significado, pero que hacen más fácil de entender el comportamiento de $A(n)$ a medida que n aumenta). De lo anterior se intuye (para la demostración rigurosa basta aplicar el Teorema 4.2.13 que enunciaremos a continuación) que el punto de equilibrio $a = 3$ del sistema dinámico (4.13) es inestable. \square



n	$A(n)$	$A(n)$	$A(n)$	$A(n)$
0	3'010	2'990	2'995	3'005
1	3'020	2'980	2'990	3'010
2	3'040	2'960	2'980	3'020
3	3'080	2'920	2'960	3'040
4	3'160	2'840	2'920	3'080
5	3'320	2'680	2'840	3'160
6	3'640	2'360	2.680	3'320
7	4'280	1'720	2'360	3'640
8	5'560	0'440	1'720	4'280
9	8'120	-2'120	0'440	5'560
10	13'240	-7'240	-2.120	8'120

TABLA 4.1: Valores $A(n)$ del sistema dinámico (4.13) con distintos datos iniciales.Figura 4.1: Puntos $(n, A(n))$ de sistemas dinámicos con distintos datos iniciales.

Ejemplo 4.2.12 Como se comprueba fácilmente, $a = -10$ es el único punto de equilibrio del sistema dinámico

$$A(n+1) = 0'8A(n) - 2. \quad (4.16)$$

En la Tabla 4.2 se muestran los primeros valores $A(n)$ del sistema dinámico (4.16) para cuatro datos iniciales distintos cercanos al punto de equilibrio: $A(0) = -9'99$, $A(0) = -10'01$, $A(0) = -10'005$ y $A(0) = -9'995$. Asimismo, en la parte derecha de la Figura 4.1 se muestran los puntos $(n, A(n))$ del sistema dinámico (4.16) para los datos iniciales anteriores. Podemos intuir que el punto de equilibrio $a = -10$ del sistema dinámico (4.16) es estable (de hecho, como veremos en el Teorema 4.2.13, para cualquier dato inicial $A(0)$ se verifica que $A(n) \rightarrow -10$ cuando $n \rightarrow +\infty$). \square

Mientras que para sistemas dinámicos no lineales puede ser complicado determinar cuándo un punto de equilibrio es estable o no, es muy sencillo para sistemas dinámicos afines como los anteriores:

Teorema 4.2.13 Consideremos el sistema dinámico afín (4.14) con $r \neq 1$. El punto de equilibrio $a = \frac{b}{1-r}$ dado en (4.15) es:

a) estable si $|r| < 1$. De hecho,

$$\lim_{n \rightarrow +\infty} A(n) = a$$

para cualquier valor de $A(0)$.



0	-9'990000	-10'01000	-10'00500	-9'995000
1	-9'992000	-10'00800	-10'00400	-9'996000
2	-9'993600	-10'00640	-10'00320	-9'996800
3	-9'994880	-10'00512	-10'00256	-9'997440
4	-9'995904	-10'00410	-10'00205	-9'997952
5	-9'996723	-10'00328	-10'00164	-9'998362
6	-9'997379	-10'00262	-10'00131	-9'998689
7	-9'997903	-10'00210	-10'00105	-9'998951
8	-9'998322	-10'00168	-10'00084	-9'999161
9	-9'998658	-10'00134	-10'00067	-9'999329
10	-9'998926	-10'00107	-10'00054	-9'999463

TABLA 4.2: Valores $A(n)$ del sistema dinámico (4.16) con distintos datos iniciales.

b) inestable si $|r| > 1$. De hecho,

$$\lim_{n \rightarrow +\infty} |A(n)| = +\infty$$

para cualquier valor de $A(0) \neq a$.

c) neutro cuando $r = -1$. En este caso, para todo $n \in \mathbb{N} \cup \{0\}$ se verifica que

$$\begin{cases} A(2n) = A(0) \\ A(2n+1) = A(1) \end{cases}$$

(por lo que diremos que este punto de equilibrio es un ciclo de orden 2).

DEMOSTRACIÓN. Puesto que $a = ra + b$, para todo $n \in \mathbb{N}$ se tiene que

$$|A(n) - a| = |rA(n-1) + b - ra - b| = |r| |A(n-1) - a|.$$

De esta forma, se verifica que

$$|A(1) - a| = |r| |A(0) - a|, |A(2) - a| = |r| |A(1) - a| = |r|^2 |A(0) - a|,$$

$$|A(3) - a| = |r| |A(2) - a| = |r|^3 |A(0) - a|, \dots$$

y, en general (pruébese por inducción),

$$|A(n) - a| = |r|^n |A(0) - a|, \quad n \in \mathbb{N} \cup \{0\}. \quad (4.17)$$

Teniendo en cuenta que

$$\lim_{n \rightarrow +\infty} |r|^n = \begin{cases} 0 & \text{si } |r| < 1 \\ +\infty & \text{si } |r| > 1, \end{cases} \quad (4.18)$$

a partir de (4.17) se tiene que:

a) si $|r| < 1$ entonces

$$\lim_{n \rightarrow +\infty} |A(n) - a| = \lim_{n \rightarrow +\infty} |r|^n |A(0) - a| = 0$$

para cualquier valor de $A(0)$. Por tanto, a es un punto de equilibrio estable.

b) si $|r| > 1$ entonces

$$\lim_{n \rightarrow +\infty} |A(n) - a| = \lim_{n \rightarrow +\infty} |r|^n |A(0) - a| = +\infty$$

para cualquier valor de $A(0) \neq a$. Por tanto, a es un punto de equilibrio inestable.



c) si $r = -1$ entonces

$$|A(n) - a| = |A(0) - a|, \quad n \in \mathbb{N} \cup \{0\}.$$

Por tanto, a es un punto de equilibrio neutro. Además, puesto que para todo n se verifica que

$$A(n) = -A(n-1) + b = -(-A(n-2) + b) + b = A(n-2),$$

se tiene que

$$\begin{cases} A(0) = A(2) = A(4) = \dots = A(2n) = \dots \\ A(1) = A(3) = A(5) = \dots = A(2n+1) = \dots \quad \square \end{cases}$$

Ejemplo 4.2.14 El Teorema 4.2.13 nos permite afirmar que el punto de equilibrio del sistema dinámico afín:

a) $A(n+1) = 0'5 A(n) + 3$ es $a = 6$ y es estable.

b) $A(n+1) = -3A(n) + 8$ es $a = 2$ y es inestable. \square

A continuación vamos emplear una técnica gráfica que nos conducirá a un mejor conocimiento de los sistemas dinámicos y los puntos fijos. Esta técnica nos ayudará a “ver” qué está ocurriendo en una ecuación particular.

Ejemplo 4.2.15 Consideremos el sistema dinámico

$$A(n+1) = 1'5 A(n) - 1, \quad n \in \mathbb{N} \cup \{0\} \quad (4.19)$$

con $A(0) = 3$. Si denotamos por x el valor $A(n)$ y por y el valor siguiente $A(n+1)$ podemos expresar el sistema dinámico anterior como

$$y = 1'5x - 1.$$

Así, haciendo $x = A(0) = 3$ se tiene que $y = A(1) = 1'5 A(0) - 1 = 4'5 - 1 = 3'5$; ahora denotamos al valor actual $A(1) = 3'5 = x$ y hacemos $y = 1'5x - 1 = 5'25 - 1 = 4'25 = A(2)$; denotamos al valor actual $A(2) = 4'25 = x$ y hacemos $y = 1'5x - 1 = 6'375 - 1 = 5'375 = A(3)$ y así, sucesivamente. En general, hacemos los siguientes pasos:

1. Hacemos $x = 3$.
2. Hacemos $y = 1'5x - 1$.
3. Reemplazamos el antiguo valor de x por el nuevo y y obtenemos un nuevo x .
4. Volvemos al paso 2 para obtener un nuevo valor y .

Veamos el proceso anterior de forma gráfica. Comenzamos dibujando las rectas $y = 1'5x - 1$ e $y = x$. El punto intersección de estas dos rectas ($x = 2$) es el punto en el cual ambos valores de x e y coinciden (obviamente, se trata del punto de equilibrio $a = 2$). Gráficamente, el paso 2 nos dice que vamos verticalmente de $x = A(0) = 3$ a la recta $y = 1'5x - 1$ para llegar al punto $(A(0), A(1)) = (3, 3'5)$ (véase la Figura 4.2). El paso 3 nos dice que debemos ir horizontalmente a la recta $y = x$ para llegar al punto $(A(1), A(1)) = (3'5, 3'5)$ (que cambia el antiguo valor de x por el nuevo x). A continuación, volviendo al paso 2, vamos verticalmente a $y = 1'5x - 1$ para llegar al punto $(A(1), A(2)) = (3'5, 4'25)$. Este proceso se va repitiendo sucesivamente. Nótese que las abscisas x de los puntos de la recta $y = 1'5x - 1$ son, precisamente, los valores $A(n)$ del sistema dinámico (4.19). Recuérdese que si $r > 1$ el punto fijo es repelente para un sistema dinámico afín; esto se visualiza gráficamente en la Figura 4.2 donde los puntos se mueven alejándose del punto intersección de las rectas (el carácter repulsivo cuando $|r| > 1$ es equivalente a que la pendiente de la recta $y = 1'5x - 1$ sea mayor (en valor absoluto) que la de la recta $y = x$). \square



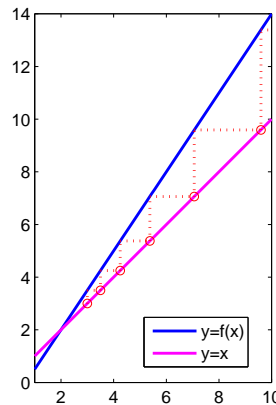


Figura 4.2: Gráfica del sistema dinámico (4.19) con $f(x) = 1.5x - 1$ y $A(0) = 3$.

Definición 4.2.16 Consideramos un sistema autónomo de primer orden

$$A(n+1) = f(A(n)), \quad n \in \mathbb{N} \cup \{0\} \quad (4.20)$$

con $A(0)$ dado. Dibujamos las gráficas de la curva

$$y = f(x) \quad (4.21)$$

y de la recta

$$y = x. \quad (4.22)$$

Tomamos $A(0)$ como primer valor de x y vamos verticalmente hasta el punto intersección con la curva (4.21); desde este punto, vamos horizontalmente hasta el punto de la corte con la recta (4.22); la abscisa x de este punto es $A(1)$. Repetimos estos pasos para obtener $A(2), A(3), \dots$. La gráfica resultante se denomina *representación gráfica* o, simplemente, *gráfica del sistema dinámico* (4.20). \square

Para las representaciones gráficas de sistemas dinámicos afines se tiene el siguiente resultado (que es una reafirmación del Teorema 4.2.13):

Teorema 4.2.17 Consideremos la representación gráfica del sistema dinámico definido por la recta $y = rx + b$ (con $r \neq 1$). El punto intersección de esta recta con la recta $y = x$ es (a, a) siendo $a = \frac{b}{1-r}$, y la sucesión de puntos determinadas por la gráfica:

- converge a (a, a) si $|r| < 1$.
- diverge a $\pm\infty$ si $|r| > 1$.
- oscila alrededor de un cuadrado centrado en (a, a) si $r = -1$. \square

El ejemplo siguiente ratifica el resultado anterior y también justifica porqué estas gráficas se denominan, a veces, “telarañas”:

Ejemplo 4.2.18 Consideremos el sistema dinámico

$$A(n+1) = -0.8A(n) + 3.6, \quad n \in \mathbb{N} \cup \{0\} \quad (4.23)$$

con $A(0) = -4$. En la Figura 4.3 se muestra la representación gráfica este sistema dinámico. Como la pendiente r de la recta $y = -0.8x + 3.6$ es negativa pero $|r| = 0.8 < 1$ se tiene que la sucesión de las abscisas de los puntos de la gráfica converge (de forma oscilante) al punto de equilibrio estable $a = 2$. \square

Veamos ahora un ejemplo en el que $r = -1$:



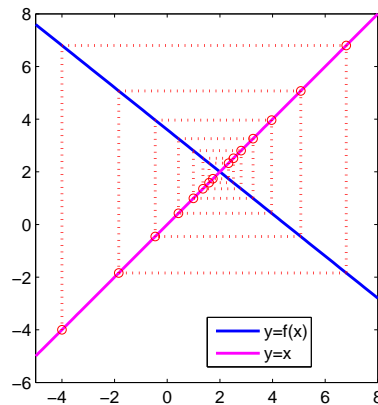


Figura 4.3: Gráfica del sistema dinámico (4.23) con $f(x) = -0.8x + 3.6$ y $A(0) = -4$.

Ejemplo 4.2.19 Consideremos el sistema afín

$$A(n+1) = -A(n) + 4, \quad n \in \mathbb{N} \cup \{0\} \quad (4.24)$$

con $A(0) = 6$. En la Figura 4.4 se muestra la representación gráfica de este sistema dinámico. Como la pendiente de la recta $y = -x + 4$ es $r = -1$ se tiene que la sucesión de puntos determinados por la gráfica oscila alrededor de un cuadrado² centrado en $(a, a) = (2, 2)$. Esto significa que la solución es un *ciclo de orden 2*, oscilando en los valores $x = -2$ y $x = 6$; por tanto, el punto fijo $a = 2$ es neutro. En general, siempre que obtengamos una curva cerrada en nuestra gráfica significa que tenemos un *ciclo*³. □

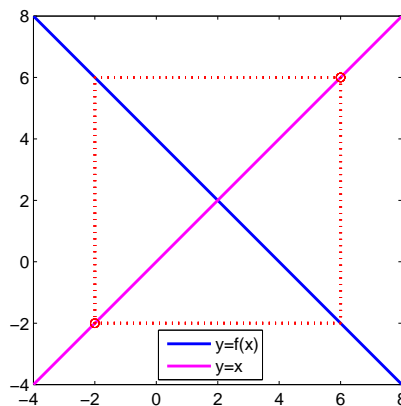


Figura 4.4: Gráfica del sistema dinámico (4.24) con $f(x) = -x + 4$ y $A(0) = 6$.

4.2.2 Solución general y particular del sistema homogéneo

Como ya se vió en la Definición 4.1.9, los sistemas dinámicos lineales homogéneos de primer orden con coeficientes constantes son de la forma

$$A(n+1) = rA(n), \quad n = 0, 1, 2, \dots \quad (4.25)$$

siendo r una constante. En el Ejemplo 4.1.11 obtuvimos la solución general y una particular de un sistema dinámico de este tipo. En general, se tiene el siguiente resultado:

²En algunas ocasiones, debido a que la escala en los ejes coordenados suele ser diferente, el dibujo puede ser un rectángulo.

³Este hecho es importante y regresaremos a él en el estudio de sistemas dinámicos no lineales.



Teorema 4.2.20

a) La solución general del sistema dinámico lineal homogéneo (4.25) es

$$A(n) = cr^n, \quad n = 0, 1, 2, \dots \quad (4.26)$$

con $c \in \mathbb{R}$.

b) La solución particular del sistema dinámico lineal homogéneo (4.25) con dato inicial

$$A(0) = a_0$$

es

$$A(n) = a_0 r^n, \quad n = 0, 1, 2, \dots$$

DEMOSTRACIÓN.

a) Por un lado, es obvio que para todo $c \in \mathbb{R}$ se verifica

$$cr^{n+1} = rcr^n, \quad n = 0, 1, 2, \dots,$$

por lo que la sucesión $A(n) = cr^n$ es solución del sistema dinámico. Por otra parte, cualquier solución del sistema dinámico tiene que ser de esa forma pues, razonando por inducción, $A(1) = A(0)r$ y si $A(n) = A(0)r^n$ entonces

$$A(n+1) = rA(n) = rA(0)r^n = A(0)r^{n+1}, \quad n = 0, 1, 2, \dots$$

b) Basta elegir la constante c en (4.26) de forma que

$$c = cr^0 = A(0) = a_0. \quad \square$$

Ejemplo 4.2.21 La solución general del sistema dinámico lineal homogéneo

$$A(n+1) = -3A(n), \quad n = 0, 1, 2, \dots \quad (4.27)$$

es

$$A(n) = c(-3)^n, \quad n = 0, 1, 2, \dots$$

con $c \in \mathbb{R}$ (ya que $r = -3$) y la solución particular del sistema dinámico (4.27) con dato inicial

$$A(0) = -3$$

es

$$A(n) = (-3) \times (-3)^n = (-3)^{n+1}, \quad n = 0, 1, 2, \dots \quad \square$$

Observación 4.2.22 ¿Cómo encontrar la solución particular del sistema dinámico lineal homogéneo (4.25) si en lugar de darnos $A(0)$ nos dan, por ejemplo, $A(1) = a_1$? Elegimos la constante c en (4.26) de forma que

$$cr = A(1) = a_1 \Rightarrow c = \frac{a_1}{r} = A(0).$$

Sustituyendo en (4.26) el valor de c obtenido, se tiene que la solución particular buscada es

$$A(n) = \frac{a_1}{r} r^n = a_1 r^{n-1}, \quad n = 0, 1, 2, \dots$$

En general, si $A(m) = a_m$, se tiene que $A(0) = \frac{a_m}{r^m}$ y

$$A(n) = \frac{a_m}{r^m} r^n = a_m r^{n-m}, \quad n = 0, 1, 2, \dots \quad \square$$

Por el Teorema 4.2.13 sabemos que el único punto de equilibrio $a = 0$ del sistema dinámico (4.25) es:



- a) estable si $|r| < 1$ y $\lim_{n \rightarrow +\infty} A(n) = 0$ para cualquier valor de $A(0)$.
- b) inestable si $|r| > 1$ y $\lim_{n \rightarrow +\infty} |A(n)| = +\infty$ para cualquier valor de $A(0) \neq 0$.
- c) neutro cuando $r = -1$ y, en este caso, $a = 0$ es un ciclo de orden 2.

Maticemos un poco más lo anterior (véase la Figura 4.5):

- a) Si $0 < r < 1$ la solución $A(n)$ va exponencialmente a 0. Más concretamente,

$$\begin{cases} \text{Si } a_0 > 0 \Rightarrow \text{la solución } A(n) \text{ decrece hacia } 0 \\ \text{Si } a_0 < 0 \Rightarrow \text{la solución } A(n) \text{ crece hacia } 0. \end{cases}$$

- b) Si $-1 < r < 0$ la solución $A(n)$ va a 0 de forma oscilante.
- c) Si $r > 1$ la solución $A(n)$ va exponencialmente a $+\infty$ (si $a_0 > 0$) o $-\infty$ (si $a_0 < 0$).
- d) Si $r < -1$ la solución $A(n)$ oscila hacia infinito con amplitud creciente.

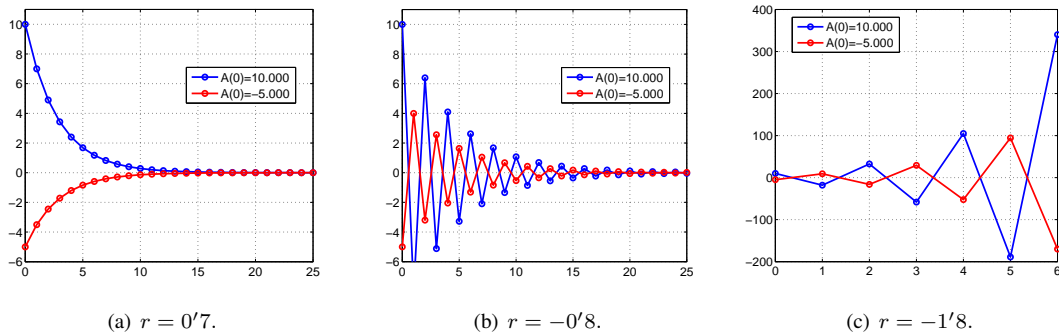


Figura 4.5: Puntos $(n, A(n))$ del sistema dinámico (4.25) con $f(x) = rx$.

Una propiedad que aparece con alguna frecuencia (por ejemplo, en el caso de los ciclos de orden 2) es la periodicidad de las soluciones:

Definición 4.2.23 Una solución $A(n)$ de un sistema dinámico es *periódica* si existe $m \in \mathbb{N}$ de forma que

$$A(n + m) = A(n), \quad n = 0, 1, 2, \dots \quad (4.28)$$

El número natural m más pequeño para el que se verifica (4.28) se denomina *periodo* de la solución. \square

Ejemplo 4.2.24 La solución particular del sistema dinámico lineal homogéneo

$$A(n + 1) = -A(n), \quad n = 0, 1, 2, \dots \quad (4.29)$$

con dato inicial

$$A(0) = -6$$

es

$$A(n) = (-6) \times (-1)^n = 6 \times (-1)^{n+1}, \quad n = 0, 1, 2, \dots \quad (4.30)$$

Puesto que

$$A(n + 2) = 6 \times (-1)^{n+3} = 6 \times (-1)^{n+1} = A(n), \quad n = 0, 1, 2, \dots$$

se tiene que la solución (4.30) es periódica de periodo 2 (véase la Figura 4.6)⁴. \square

Observación 4.2.25 Más adelante consideraremos otros ejemplos de soluciones periódicas con periodos más grandes. \square

⁴Nótese que también se verifica que $A(n + 4) = A(n)$, $A(n + 6) = A(n)$, ... pero $m = 2$ es el número más pequeño con la propiedad anterior.



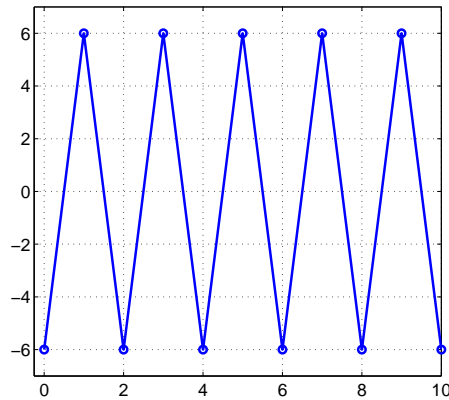


Figura 4.6: Puntos $(n, A(n))$ del sistema dinámico (4.2.24) con $A(0) = -6$.

4.2.3 Solución general y particular del sistema afín

Comenzamos con el estudio de un sistema dinámico afín de primer orden de la forma

$$A(n+1) = rA(n) + b, \quad n = 0, 1, 2, \dots \quad (4.31)$$

siendo r y b constantes. Se tiene el siguiente resultado:

Teorema 4.2.26

a) La solución general del sistema dinámico afín (4.31) es

$$A(n) = \begin{cases} cr^n + \frac{b}{1-r} & \text{si } r \neq 1 \\ c + nb & \text{si } r = 1 \end{cases} \quad (4.32)$$

para $n = 0, 1, 2, \dots$ con $c \in \mathbb{R}$.

b) La solución particular del sistema del sistema dinámico afín (4.31) con dato inicial

$$A(0) = a_0$$

es

$$A(n) = \begin{cases} \left(a_0 - \frac{b}{1-r}\right)r^n + \frac{b}{1-r} & \text{si } r \neq 1 \\ a_0 + nb & \text{si } r = 1 \end{cases}$$

para $n = 0, 1, 2, \dots$

DEMOSTRACIÓN.

a) Distinguiamos dos casos:

i) $r \neq 1$ Consideremos el punto de equilibrio

$$a = \frac{b}{1-r}$$

del sistema (4.31) y definamos la sucesión

$$E(n) = A(n) - a, \quad n = 0, 1, 2, \dots \quad (4.33)$$



Como $a = ra + b$,

$$E(n+1) = A(n+1) - a = rA(n) + b - ra - b = r(A(n) - a) = rE(n). \quad (4.34)$$

La solución general del sistema dinámico homogéneo (4.34) es, como se vió en el Teorema 4.2.20,

$$E(n) = cr^n, \quad n = 0, 1, 2, \dots$$

con $c \in \mathbb{R}$. Llevando este valor a (4.33), se tiene que

$$A(n) = cr^n + \frac{b}{1-r}, \quad n = 0, 1, 2, \dots$$

ii) $\boxed{r=1}$ En este caso, el sistema se escribe $A(n+1) = A(n) + b$ con lo que

$$A(1) = A(0) + b, A(2) = A(1) + b = A(0) + 2b, A(3) = A(2) + b = A(0) + 3b$$

y, en general,

$$A(n) = A(0) + nb = c + nb$$

siendo $c = A(0)$ un valor arbitrario.

b) Basta elegir, de forma adecuada, la constante c en (4.32) para que $A(0) = a_0$. \square

Ejemplo 4.2.27

a) Puesto que el punto de equilibrio del sistema dinámico afín

$$A(n+1) = 2A(n) + 5, \quad n = 0, 1, 2, \dots \quad (4.35)$$

es

$$a = \frac{5}{1-2} = -5$$

se tiene que la solución general del sistema (4.35) es

$$A(n) = c2^n - 5, \quad n = 0, 1, 2, \dots$$

con $c \in \mathbb{R}$ y la solución particular del sistema dinámico (4.35) con dato inicial

$$A(0) = 3$$

es

$$A(n) = (3+5)2^n - 5 = 2^{n+3} - 5, \quad n = 0, 1, 2, \dots$$

(véase la Figura 4.7).

b) La solución general del sistema dinámico afín

$$A(n+1) = A(n) - 2, \quad n = 0, 1, 2, \dots \quad (4.36)$$

es

$$A(n) = c - 2n, \quad n = 0, 1, 2, \dots$$

con $c \in \mathbb{R}$ y la solución particular del sistema dinámico (4.36) con dato inicial

$$A(0) = 9$$

es

$$A(n) = 9 - 2n, \quad n = 0, 1, 2, \dots$$

(véase la Figura 4.8). Hay que destacar la relación lineal en la figura y que, ahora, no hay puntos de equilibrio. \square



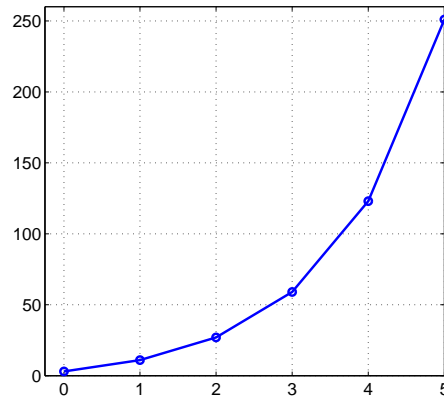


Figura 4.7: Puntos $(n, A(n))$ del sistema dinámico (4.35) con $A(0) = 3$.

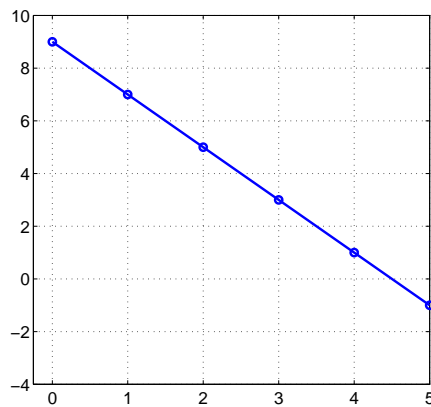


Figura 4.8: Puntos $(n, A(n))$ del sistema dinámico (4.36) con $A(0) = 9$.

Observación 4.2.28

- a) Cuando $r \neq 1$ y $a_0 = \frac{b}{1-r}$, la solución particular del sistema dinámico afín (4.31) es

$$A(n) = \frac{b}{1-r}, \quad n = 0, 1, 2, \dots$$

Es decir, la solución es constante (lo cual no debe sorprender debido a que el valor inicial coincide con el punto de equilibrio).

- b) El Teorema 4.2.26 se puede utilizar para demostrar el Teorema 4.2.13. En efecto, la propiedad (4.18) determina que

$$\begin{cases} \lim_{n \rightarrow +\infty} A(n) = \lim_{n \rightarrow +\infty} \left(cr^n + \frac{b}{1-r} \right) = \frac{b}{1-r} & \text{si } |r| < 1 \\ \lim_{n \rightarrow +\infty} |A(n)| = \lim_{n \rightarrow +\infty} \left| cr^n + \frac{b}{1-r} \right| = +\infty & \text{si } |r| > 1. \end{cases}$$

Por otra parte, si $r = -1$, se tiene que

$$A(n) = c(-1)^n + \frac{b}{2}$$



por lo que, para todo $n \in \mathbb{N} \cup \{0\}$, se verifica que

$$A(n+2) = c(-1)^{n+2} + \frac{b}{2} = c(-1)^n + \frac{b}{2} = A(n)$$

y, por tanto,

$$\begin{cases} A(0) = A(2) = A(4) = \dots = A(2n) = \dots \\ A(1) = A(3) = A(5) = \dots = A(2n+1) = \dots \quad \square \end{cases}$$

4.2.4 Aplicaciones

Vamos a estudiar algunas situaciones reales que se pueden modelizar mediante sistemas afines y que son de índoles muy diversas: tipos de interés, hipotecas, datación arqueológica, genética, etc.

Finanzas

Consideremos una situación general de un producto financiero que tiene un interés nominal del I por ciento, compuesto en m periodos anuales. Supongamos que en cada periodo se aporta o detrae una cantidad b (si $b > 0$ se tratará de una aportación, y de un reembolso en caso contrario). Si suponemos un capital inicial a_0 , el sistema dinámico

$$A(n+1) = \left(1 + \frac{i}{m}\right) A(n) + b,$$

donde $i = \frac{I}{100}$ y $A(0) = a_0$, describe la relación entre el capital invertido en dicho producto al finalizar el periodo $n+1$ y el capital invertido al final del periodo n . Puesto que $r = 1 + \frac{i}{m} \neq 1$, la solución de dicho sistema es

$$A(n) = \left(1 + \frac{i}{m}\right)^n \left(a_0 + \frac{mb}{i}\right) - \frac{mb}{i} \quad (4.37)$$

Vamos a usar esta situación general para describir distintas situaciones de las finanzas:

Tasa anual equivalente

Analicemos este párrafo, extraído de un anuncio:

3% T.A.E.* DURANTE LOS 4 PRIMEROS MESES.

*T.A.E. calculada para cualquier importe. Abono mensual de intereses. Tipo de interés nominal anual 2'96% durante 4 meses (3% T.A.E.) aplicable a partir de la fecha del primer ingreso.

Las letras T.A.E. se refieren a la Tasa Anual Equivalente. ¿Por qué se distingue entre interés nominal anual (2'96% en este caso) y T.A.E. (3% para el ejemplo)? La tasa o tipo de interés anual equivalente representa la rentabilidad que se obtendría si se mantuviera el dinero, en las mismas condiciones de capitalización, durante un año. En nuestro ejemplo, la rentabilidad que se obtendría si el capital total al cabo de esos 4 meses (capital inicial más intereses) se reinvirtiera por otro periodo igual en las mismas condiciones y, a su vez, el dinero obtenido se volviera a reinvertir por otros 4 meses. Finalmente (al cabo de un año) tendríamos 1'03 veces el capital inicial, es decir, habríamos obtenido una rentabilidad anual del 3%.

Usando la fórmula (4.37) podemos deducir cuál es el interés nominal de este producto. Puesto que

$$1'03a_0 = A(12) = \left(1 + \frac{i}{12}\right)^{12} a_0,$$

podemos despejar

$$i = 12\left((1'03)^{\frac{1}{12}} - 1\right) = 0'029595$$

para obtener que el interés nominal anual $I = 100i$ es, redondeando a dos decimales, 2'96.



En general, el interés nominal de un producto que se capitaliza en m periodos anuales con una T.A.E. $T = 100t$ viene dado por

$$I = 100m \left((1 + t)^{\frac{1}{m}} - 1 \right).$$

Cuadro de amortización de una hipoteca

Supongamos que queremos pedir un crédito hipotecario y el banco nos ofrece un tipo de interés nominal del 4%. Necesitamos pedir 200000€ y queremos saber cuánto tendríamos que pagar al mes si solicitamos el crédito a, por ejemplo, 15 años. Según la fórmula (4.37) tendríamos que la cantidad que debemos al banco en el n -ésimo mes viene dada por:

$$A(n) = \left(1 + \frac{0'04}{12} \right)^n \left(200000 + \frac{12b}{0'04} \right) - \frac{12b}{0'04}$$

siendo $-b$ la cantidad que debemos pagar al mes. Para determinar b debemos utilizar que en el mes 180 ya no deberemos nada, es decir, $A(180) = 0$. Así pues,

$$0 = \left(1 + \frac{0'04}{12} \right)^{180} \left(200000 + \frac{12b}{0'04} \right) - \frac{12b}{0'04},$$

de donde

$$b = -\frac{0'04}{12} \left(\frac{\left(1 + \frac{0'04}{12} \right)^{180} 200000}{\left(1 + \frac{0'04}{12} \right)^{180} - 1} \right) = -1479'38 \text{ €}.$$

Por tanto, tendríamos que pagar del orden de 1500€ al mes.

Nótese que el primer mes estaremos pagando de intereses

$$200000 \frac{0'04}{12} = 666'67 \text{ €}$$

por lo que habremos amortizado $812'71 = 1479'38 - 666'67 = 200000 - A(1)$ euros de capital.

En general, para un crédito de una cantidad C a N años y a un tipo de interés nominal $I = 100i$, la cantidad que se debe al banco en el mes n -ésimo se expresa

$$A(n) = \left(1 + \frac{i}{12} \right)^n \left(C - \frac{12c}{i} \right) + \frac{12c}{i}.$$

siendo c la cuota mensual, que vendrá dada por

$$c = -b = \frac{iC}{12} \left(\frac{\left(1 + \frac{i}{12} \right)^{12N}}{\left(1 + \frac{i}{12} \right)^{12N} - 1} \right).$$

Finalmente, los intereses que se pagan en la cuota del mes $n + 1$ son

$$\text{int}(n + 1) = \frac{i}{12} A(n) = \left(1 + \frac{i}{12} \right)^n \left(\frac{iC}{12} - c \right) + c = c - \left(1 + \frac{i}{12} \right)^n \left(c - \frac{iC}{12} \right),$$

mientras que el capital amortizado en ese mes es

$$\text{am}(n + 1) = c - \text{int}(n + 1) = \left(1 + \frac{i}{12} \right)^n \left(c - \frac{iC}{12} \right).$$



Datación mediante carbono 14

El material radiactivo con el paso del tiempo se va desintegrando, es decir, va perdiendo actividad. Si denotamos por $A(n)$ la cantidad de material que al cabo de n años se mantiene radiactiva, y por p la probabilidad de que un átomo se desintegre, la cantidad de material activo cada año será proporcional a la cantidad que permanecía activa en el año anterior con constante $r = 1 - p$. Tendremos, por tanto,

$$A(n+1) = rA(n).$$

Se sabe que la *semivida* de esta materia radiactiva (es decir, el tiempo que tarda en reducirse a la mitad la cantidad de materia que mantiene radiactividad) es, para el carbono 14, de 5730 años. Por otra parte, la proporción de carbono 14 es aproximadamente constante en la atmósfera y en los seres vivos (en las plantas, los isótopos que pierden actividad se sustituyen por otros a través de la fotosíntesis; en los animales, por la ingesta de plantas o de otros animales que se alimentan de plantas). Cuando un ser vivo muere, deja de suplir el carbono 14 que decae en su actividad, por lo que midiendo la proporción que de este material tiene un resto orgánico, respecto a la cantidad que tendría en caso de estar vivo, se puede datar el momento del fallecimiento del ser del que proviene el resto.

En concreto, usando que la semivida del carbono 14 es, como se ha dicho, 5730 años, se tiene

$$\frac{A(0)}{2} = A(5730) = r^{5730}A(0),$$

por lo que para este material tendremos

$$r = (0'5)^{\frac{1}{5730}} = 0'999879.$$

Ejemplo 4.2.29 Un arqueólogo encuentra un hueso de un animal que tiene un 70% del carbono 14 original y quiere saber de cuándo procede. Llamando año cero al año en que murió el animal y año n al año actual, podemos escribir

$$A(n) = 0'7A(0) = (0'5)^{\frac{n}{5730}}A(0),$$

es decir, $n = 2948'5$. El hueso tendría, por tanto, entre de 2900 y tres mil años. \square

Evolución genética

Muchos de los rasgos de los individuos están determinados por los genes que heredan de sus padres. La mayoría de los genes está formada por dos variantes o *alelos* que denotaremos por A y a . Cada individuo hereda un alelo de cada uno de sus padres y el rasgo viene determinado por un par (A, A) , (A, a) , (a, A) o (a, a) . Los individuos que tiene los pares de alelos (A, A) y (a, a) se denominan *homocigóticos* y los que tienen el par (A, a) (trabajaremos con genes en los que el orden de los alelos no importa) se denominan *heterocigóticos*. Estos tres serán los *genotipos* posibles. De cada uno de los dos alelos, en cada individuo se manifiesta sólo uno; su *fenotipo*. En los homocigóticos se manifiesta el único alelo que tienen, mientras que en los heterocigóticos se manifiesta siempre el A (que se denomina *dominante*, mientras que al alelo a se le denomina *recesivo*).

Si en la generación inicial (generación 0) hay T individuos, de los cuales F tienen sexo femenino y M masculino ($T = F + M$), en total habrá $2T$ alelos. De ellos, suponemos que hay una proporción $p(0)$ de tipo A y $q(0)$ de tipo a (obviamente $p(0) + q(0) = 1$). En total habrá $2p(0)T$ alelos A y $2q(0)T$ alelos a . Asimismo, suponemos que las proporciones se mantienen cuando se toman sólo machos o sólo hembras.

El número de alelos A que provienen de una hembra son $p(0)F$, mientras que el número de alelos a es $q(0)F$; análogamente, tenemos $p(0)M$ alelos A y $q(0)M$ alelos a provenientes de un macho. Así pues, bajo la hipótesis de que los apareamientos se producen de manera aleatoria, la proporción de individuos homocigóticos A en la generación 1 será

$$u = \frac{p(0)^2 FM}{FM} = p(0)^2,$$



mientras que la de homocigóticos a vendrá dada por

$$w = \frac{q(0)^2 FM}{FM} = q(0)^2.$$

Finalmente, la proporción de heterocigóticos será

$$v = \frac{2p(0)q(0)FM}{FM} = 2p(0)q(0).$$

Nótese que $u + v + w = (p(0) + q(0))^2 = 1$, como no podía ser de otra forma.

Cabe ahora preguntarse cuál es la proporción total de alelos A y de alelos a en la generación 1. Si llamamos T a la cantidad de individuos de la generación 1, el número de alelos A será $(2u + v)T$ y el número de alelos a será $(v + 2w)T$, mientras que el total de alelos será $2T$. Por tanto, la proporción de alelos A será

$$p(1) = \frac{(2p(0)^2 + 2p(0)q(0))T}{2T} = p(0)(p(0) + q(0)) = p(0)$$

y la de alelos a

$$q(1) = \frac{(2p(0)q(0) + 2q(0)^2)T}{2T} = q(0)(p(0) + q(0)) = q(0).$$

De esta forma, hemos llegado a la sorprendente conclusión de que, a pesar que uno de los alelos sea dominante frente al otro, la proporción de alelos de cada tipo se mantiene constante de generación en generación. También es constante la proporción de individuos heterocigóticos, homocigóticos A y homocigóticos a . Este hecho, que ocurre cuando no hay influencias externas (mutación, desventajas de uno de los alelos) se conoce como *Ley de Hardy-Weinberg*.

Algunas veces, los caracteres recesivos tienen algún tipo de desventaja. Supongamos, por ejemplo, que los individuos homocigóticos a mueren antes de alcanzar la edad adulta reproductiva (se suele decir que es un *rasgo letal*). Denotemos por $p(n)$ y $q(n)$ la proporción de alelos A y a , respectivamente, en la generación n ; la proporción de genotipos homocigóticos A , heterocigóticos y homocigóticos a en la generación $n + 1$ vendrá dada, como antes, por

$$u = p(n)^2, v = 2p(n)q(n) \text{ y } w = q(n)^2,$$

respectivamente.

Si T es el número total de nacidos en la generación $n + 1$, el número de individuos adultos será

$$(u + v)T = Tp(n)(p(n) + 2q(n)) = Tp(n)(1 + q(n)),$$

donde hemos usado que $p(n) = 1 - q(n)$.

La cantidad total de alelos en los adultos de la generación $n + 1$ será, por tanto, el doble:

$$2Tp(n)(1 + q(n)).$$

El número de total de alelos a en esta generación coincide con la cantidad de individuos heterocigóticos, esto es,

$$vT = 2Tp(n)q(n),$$

por lo que la proporción de alelos a en la generación $n + 1$ es

$$q(n + 1) = \frac{2Tp(n)q(n)}{2Tp(n)(1 + q(n))} = \frac{q(n)}{1 + q(n)}. \quad (4.38)$$

Éste es un sistema dinámico no lineal, pero con un pequeño truco vamos a poder transformarlo en uno lineal y resolverlo. Llamando

$$Q(n) = \frac{1}{q(n)}$$



la ecuación (4.38) se transforma en

$$\frac{1}{Q(n+1)} = \frac{1}{1 + \frac{1}{Q(n)}},$$

lo cual es equivalente (si $Q(n) \neq 0$, hipótesis más que razonable) a

$$\frac{1}{Q(n+1)} = \frac{1}{1 + Q(n)}$$

o, de otra forma,

$$Q(n+1) = Q(n) + 1.$$

La solución de este sistema es

$$Q(k) = k + Q_0,$$

por lo que tendremos

$$q(k) = \frac{q_0}{kq_0 + 1}.$$

Hemos probado el siguiente

Principio de selección: Supongamos que los individuos homocigóticos del rasgo recesivo a no se reproducen, pero sí lo hacen los homocigóticos dominantes y los heterocigóticos. Entonces, la proporción de alelos a en la n -ésima generación viene dada por

$$q(n) = \frac{q_0}{nq_0 + 1},$$

donde q_0 es la proporción inicial de alelos a . \square

Es claro que la cantidad $q(n)$ tiende a cero cuando n tiende a infinito. Pero en una situación verosímil, en la que la proporción inicial de alelos con el rasgo recesivo letal será pequeña, este decrecimiento es muy lento. Veamos un ejemplo:

Ejemplo 4.2.30 Supongamos que en una población hay un rasgo letal con una proporción de alelos a que viene dada por $q(0) = 0'02$. ¿Cuántas generaciones son necesarias para que esta proporción se reduzca a la mitad? Para ello, debe cumplirse

$$0'01 = q(n) = \frac{q(0)}{q(0)n + 1} = \frac{0'02}{0'02n + 1},$$

es decir, $n = 50$. Por lo tanto, serían necesarias 50 generaciones para alcanzar tal reducción. Nótese que, en el caso de los seres humanos, esto supone del orden de 1250 años. Esta lenta reducción explica por qué la eugenesia negativa que se intentó llevar a cabo a principios del Siglo XX en algunos países, no sólo es de dudosa categoría ética, sino que además es ineficaz. \square

4.3 Sistemas dinámicos lineales de segundo orden

En general, los sistemas dinámicos no lineales no pueden resolverse de forma elemental, salvo unos cuantos casos de sistemas dinámicos de primer orden. Por consiguiente, al tratar de resolver sistemas dinámicos de orden superior, limitaremos nuestra atención a los *sistemas dinámicos lineales de coeficientes constantes*. Como veremos, nuestra capacidad para resolver un sistema lineal de orden n con coeficientes constantes, dependerá de la aptitud para encontrar las raíces de un polinomio de grado n .

En toda esta sección, consideraremos únicamente sistemas dinámicos lineales de segundo orden con coeficientes constantes de la forma

$$A(n+2) = aA(n+1) + bA(n) + c, \quad n = 0, 1, \dots \quad (4.39)$$

donde $a, b, c \in \mathbb{R}$. Si $c = 0$ diremos que el sistema lineal (4.39) es *homogéneo* y, en caso contrario, diremos que es *afín*.



4.3.1 Sistemas homogéneos: Polinomio característico, solución general y particular

Ejemplo 4.3.1 Un estudio realizado sobre la contaminación de los Los Grandes Lagos (Erie y Ontario, el segundo situado aguas abajo del primero) entre Estados Unidos y Canadá permite modelizar la contaminación de sus aguas mediante dos sistemas dinámicos. Más concretamente, sean $A(n)$ y $B(n)$ las cantidades de polución en los lagos Erie y Ontario, respectivamente, en el año n ; suponemos que se verifican las siguientes relaciones

$$A(n+1) = 0'62A(n) \text{ y } B(n+1) = 0'87B(n) + 0'38A(n).$$

Una forma de resolver el sistema anterior consiste en despejar $A(n)$ en la segunda ecuación, es decir,

$$A(n) = \frac{1}{0'38} B(n+1) - \frac{0'87}{0'38} B(n),$$

y sustituir en la primera, obteniendo

$$\frac{1}{0'38} B(n+2) - \frac{0'87}{0'38} B(n+1) = A(n+1) = 0'62A(n) = 0'62 \left(\frac{1}{0'38} B(n+1) - \frac{0'87}{0'38} B(n) \right).$$

Tras unas simplificaciones, se obtiene el sistema dinámico de segundo orden

$$B(n+2) = (0'87 + 0'62)B(n+1) - (0'87)(0'62)B(n), \quad (4.40)$$

Como puede comprobarse directamente, la sucesión

$$B(n) = c_1(0'87)^n + c_2(0'62)^n, \quad n = 0, 1, \dots$$

con $c_1, c_2 \in \mathbb{R}$, es solución del sistema (4.40). Destaquemos que $0'87$ y $0'62$ son las raíces de la ecuación de segundo grado⁵

$$x^2 - (0'87 + 0'62)x + (0'87)(0'62) = 0. \quad \square$$

Lo anterior da pie a la siguiente:

Definición 4.3.2 La *ecuación característica* asociada a un sistema dinámico lineal homogéneo de segundo orden de la forma

$$A(n+2) = aA(n+1) + bA(n), \quad n = 0, 1, \dots \quad (4.41)$$

con $a, b \in \mathbb{R}$ es

$$x^2 - ax - b = 0.$$

El polinomio de segundo grado

$$P(x) = x^2 - ax - b \quad (4.42)$$

se denomina *polinomio característico* asociado al sistema dinámico (4.41). \square

Observación 4.3.3 Si introducimos la notación

$$v(n) = \begin{pmatrix} A(n+1) \\ A(n) \end{pmatrix},$$

es fácil ver que el sistema dinámico de segundo orden (4.41) puede escribirse como

$$v(n+1) = Mv(n),$$

donde

$$M = \begin{pmatrix} a & b \\ 1 & 0 \end{pmatrix}.$$

Esto es, puede escribirse como un sistema dinámico de orden uno para una sucesión de vectores de \mathbb{R}^2 , en lugar de un sistema de orden 2 para una sucesión de números reales. Nótese que la matriz M tiene como polinomio característico, precisamente, el polinomio que hemos denominado polinomio característico del sistema dinámico (4.41). \square

⁵En general, si x_1 y x_2 son las raíces de la ecuación $ax^2 + bx + c = 0$ con $a \neq 0$, se verifica que $x_1 + x_2 = -\frac{b}{a}$ y $x_1x_2 = \frac{c}{a}$.



Como hemos visto en el Ejemplo 4.3.1, las raíces del polinomio característico nos van a permitir obtener la solución general de cualquier sistema dinámico lineal de segundo orden. Concretamente,

Teorema 4.3.4 (Raíces distintas) Consideremos un sistema dinámico lineal homogéneo de segundo orden de la forma (4.41) y sean r y s con $r \neq s$ las dos raíces (reales o complejas) del polinomio característico asociado (4.42). Entonces,

a) la solución general del sistema dinámico (4.41) viene dada por

$$A(n) = c_1 r^n + c_2 s^n, \quad n = 0, 1, \dots \quad (4.43)$$

con $c_1, c_2 \in \mathbb{R}$.

b) la solución particular de (4.41) que prescribe los datos iniciales $A(0) = a_0$ y $A(1) = a_1$ es

$$A(n) = \frac{a_1 - sa_0}{r - s} r^n + \frac{a_1 - ra_0}{s - r} s^n, \quad n = 0, 1, \dots \quad (4.44)$$

DEMOSTRACIÓN.

a) Veamos que toda solución del sistema dinámico (4.41) se escribe en la forma (4.43), razonando por inducción. En primer lugar, cualesquiera que sean los valores de $A(0)$ y $A(1)$ siempre podemos escribirlos como

$$\begin{cases} A(0) = c_1 + c_2 \\ A(1) = c_1 r + c_2 s. \end{cases}$$

Esto es cierto porque, al ser $r \neq s$, el sistema con incógnitas c_1 y c_2 tiene solución única.

Supongamos ahora que $A(n+1)$ y $A(n)$ se expresan según (4.43). Puesto que r y s son raíces del polinomio característico (4.42), se verifica que

$$ar + b = r^2 \quad \text{y} \quad as + b = s^2. \quad (4.45)$$

Así pues, podemos escribir

$$\begin{aligned} A(n+2) &= aA(n+1) + bA(n) = a(c_1 r^{n+1} + c_2 s^{n+1}) + b(c_1 r^n + c_2 s^n) \\ &= c_1 r^n (ar + b) + c_2 s^n (as + b) \\ &= c_1 r^n r^2 + c_2 s^n s^2 \\ &= c_1 r^{n+2} + c_2 s^{n+2} \end{aligned}$$

b) Debemos elegir las constantes c_1 y c_2 para que se verifiquen las condiciones iniciales. Por tanto, estas constantes vienen dadas como la solución del sistema lineal

$$\begin{cases} c_1 + c_2 = A(0) = a_0 \\ rc_1 + sc_2 = A(1) = a_1 \end{cases} \Rightarrow c_1 = \frac{a_1 - sa_0}{r - s} \quad \text{y} \quad c_2 = \frac{a_1 - ra_0}{s - r}. \quad \square$$

Ejemplo 4.3.5 Consideremos el sistema dinámico lineal de segundo orden

$$\begin{cases} A(0) = 7, \quad A(1) = -6 \\ A(n+2) = -A(n+1) + 6A(n), \quad n = 0, 1, \dots \end{cases} \quad (4.46)$$

Puesto que el polinomio característico asociado al sistema anterior es

$$P(x) = x^2 + x - 6 = (x - 2)(x + 3),$$

se tiene que la solución general de (4.46) es

$$A(n) = c_1 2^n + c_2 (-3)^n, \quad n = 0, 1, \dots$$



con $c_1, c_2 \in \mathbb{R}$. Determinemos las constantes c_1 y c_2 para que se satisfagan las condiciones iniciales

$$\begin{cases} c_1 + c_2 = A(0) = 7 \\ 2c_1 - 3c_2 = A(1) = -6 \end{cases} \Rightarrow c_1 = 3 \text{ y } c_2 = 4.$$

Por tanto, la solución del sistema dinámico (4.46) es

$$A(n) = 3 \times 2^n + 4(-3)^n, \quad n = 0, 1, \dots \quad \square$$

Ejemplo 4.3.6 Consideremos el sistema dinámico lineal de segundo orden

$$A(n+2) = A(n+1) + A(n), \quad n = 0, 1, \dots \quad (4.47)$$

La secuencia de números $\{A_n\}_{n=0}^{\infty}$ se denomina *sucesión de Fibonacci* (aparece tanto en Matemática Aplicada como en Biología). Como el polinomio característico asociado al sistema (4.47)

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

tiene por raíces

$$r = \frac{1 + \sqrt{5}}{2} \text{ y } s = \frac{1 - \sqrt{5}}{2},$$

se tiene que el término general de la sucesión de Fibonacci es

$$A(n) = c_1 \left(\frac{1 + \sqrt{5}}{2} \right)^n + c_2 \left(\frac{1 - \sqrt{5}}{2} \right)^n, \quad n = 0, 1, \dots$$

con $c_1, c_2 \in \mathbb{R}$. El resultado es sorprendente, dado que si $A(0), A(1) \in \mathbb{N}$, cada número natural $A(n)$ se expresa en términos del número irracional $\sqrt{5}$. Es un mero ejercicio comprobar que la sucesión de Fibonacci tradicional

$$1, 1, 2, 3, 5, 8, 13, 21, \dots$$

se genera cuando se toman los valores

$$c_1 = \frac{5 + \sqrt{5}}{10} \text{ y } c_2 = \frac{5 - \sqrt{5}}{10}. \quad \square$$

Ejemplo 4.3.7 Consideremos el sistema dinámico lineal homogéneo de segundo orden

$$\begin{cases} A(0) = 2, \quad A(1) = 6 \\ A(n+2) = 2A(n+1) - 5A(n), \quad n = 0, 1, \dots \end{cases} \quad (4.48)$$

Puesto que el polinomio característico asociado al sistema anterior es

$$P(x) = x^2 - 2x + 5$$

y tiene por raíces

$$x = \frac{2 \pm \sqrt{4 - 20}}{2} = 1 \pm 2i,$$

siendo $i = \sqrt{-1}$ la *unidad imaginaria*, se tiene que la solución general de (4.48) es

$$A(n) = c_1(1 + 2i)^n + c_2(1 - 2i)^n, \quad n = 0, 1, \dots$$

con $c_1, c_2 \in \mathbb{R}$. Determinemos las constantes c_1 y c_2 para que se satisfagan las condiciones iniciales

$$\begin{cases} c_1 + c_2 = A(0) = 2 \\ (1 + 2i)c_1 + (1 - 2i)c_2 = A(1) = 6 \end{cases} \Rightarrow c_1 = 1 - i \text{ y } c_2 = 1 + i.$$

Por tanto, la solución del sistema dinámico (4.48) es

$$A(n) = (1 - i)(1 + 2i)^n + (1 + i)(1 - 2i)^n, \quad n = 0, 1, \dots$$

Como se aprecia, el resultado obtenido es todavía más sorprendente que el del Ejemplo 4.3.6, dado que la sucesión $A(0) = 2, A(1) = 6, A(2) = 2, A(3) = -26, A(4) = -62, A(5) = 6, \dots$ está formada por números enteros y cada entero $A(n)$ se expresa en términos de números complejos. \square



A continuación, vamos a ver cómo se expresa la solución de un sistema dinámico lineal de segundo orden cuando el polinomio característico asociado tiene una única raíz real doble. Previamente al teorema que nos dará dicha expresión, veamos un ejemplo.

Ejemplo 4.3.8 Supongamos que tenemos un coche que permite recorrer 20 km por litro de combustible y que comenzamos a conducir con 15 litros en el depósito. Denotemos por $A(n)$ la cantidad de combustible que nos queda tras recorrer n km (nótese que $A(0) = 15$). Puesto que consumimos 0'05 litros de combustible por kilómetro, se tiene que

$$A(n+1) = A(n) - 0'05, \quad n = 0, 1, \dots$$

es decir, por cada kilómetro que recorremos tenemos 0'05 litros menos de combustible. Por el Teorema 4.2.26 sabemos que la solución particular del sistema dinámico

$$\begin{cases} A(0) = 15 \\ A(n+1) = A(n) - 0'05, \quad n = 0, 1, \dots \end{cases}$$

es

$$A(n) = 15 - 0'05n, \quad n = 0, 1, \dots$$

En general, el sistema dinámico que describe el consumo de combustible en un vehículo tras recorrer n km es

$$A(n+1) = A(n) - c,$$

donde c es la cantidad de combustible que se necesita para recorrer 1 km. De nuevo, el Teorema 4.2.26 nos asegura que la solución particular del sistema dinámico

$$\begin{cases} A(0) = a_0 \\ A(n+1) = A(n) - c, \quad n = 0, 1, \dots \end{cases}$$

es

$$A(n) = c_1 + c_2n, \quad n = 0, 1, \dots$$

siendo $c_1 = a_0$ y $c_2 = -c$. Otra forma de abordar este problema consiste en escribir

$$A(n+2) - A(n+1) = A(n+1) - A(n)$$

(el consumo empleado para recorrer un kilómetro (de la $n+1$ a la $n+2$) es el mismo que en recorrer otro kilómetro (de la n a la $n+1$)) o, equivalentemente,

$$A(n+2) = 2A(n+1) - A(n), \quad n = 0, 1, \dots \quad (4.49)$$

Se prueba por inducción que la solución general del sistema dinámico anterior es

$$A(n) = c_1 + c_2n, \quad n = 0, 1, \dots \quad (4.50)$$

con $c_1, c_2 \in \mathbb{R}$. En efecto, a partir de $A(0)$ y $A(1)$ es claro que la elección

$$\begin{cases} c_1 = A(0) \\ c_2 = A(1) - A(0) \end{cases}$$

permite escribir

$$\begin{cases} A(0) = c_1 + c_2 \cdot 0 \\ A(1) = c_1 + c_2 \cdot 1. \end{cases}$$

Si ahora suponemos que $A(n+1)$ y $A(n)$ se escriben en la forma (4.50) tendremos que

$$\begin{aligned} A(n+2) &= 2A(n+1) - A(n) = 2(c_1 + c_2(n+1)) - (c_1 + c_2n) \\ &= c_1 + c_2n + 2c_2 = c_1 + c_2(n+2). \end{aligned}$$



Conociendo los valores iniciales $A(0) = a_0$ y $A(1) = a_1$, los valores c_1 y c_2 se toman de forma que se verifique

$$\begin{cases} c_1 = A(0) = a_0 \\ c_1 + c_2 = A(1) = a_1 \end{cases} \Rightarrow c_1 = a_0 \text{ y } c_2 = a_1 - a_0.$$

Por tanto, la solución del sistema dinámico

$$\begin{cases} A(0) = a_0, A(1) = a_1 \\ A(n+2) = 2A(n+1) - A(n), n = 0, 1, \dots \end{cases}$$

viene dada por

$$A(n) = a_0 + (a_1 - a_0)n, n = 0, 1, \dots \quad \square$$

El polinomio característico asociado al sistema dinámico lineal de segundo orden (4.49) es

$$P(x) = x^2 - 2x + 1 = (x - 1)^2$$

y tiene a $r = 1$ como una raíz doble. En general, ¿qué ocurre si r es una raíz doble del polinomio característico? La respuesta la encontramos en el siguiente resultado:

Teorema 4.3.9 (Raíz doble) Consideremos un sistema dinámico lineal homogéneo de segundo orden de la forma (4.41) y sea $r \neq 0$ la única raíz (doble) del polinomio característico asociado (4.42). Entonces,

a) la solución general del sistema dinámico (4.41) viene dada por

$$A(n) = (c_1 + c_2n)r^n, n = 0, 1, \dots \quad (4.51)$$

con $c_1, c_2 \in \mathbb{R}$.

b) la solución particular de (4.41) que prescribe los datos iniciales $A(0) = a_0$ y $A(1) = a_1$ es

$$A(n) = \left(a_0 + \frac{a_1 - ra_0}{r} n \right) r^n, n = 0, 1, \dots \quad (4.52)$$

DEMOSTRACIÓN. En primer lugar, nótese que si el polinomio característico (4.42) tiene a r como única raíz real doble, éste es de la forma

$$P(x) = x^2 - ax - b = (x - r)^2 = x^2 - 2rx + r^2 \Rightarrow a = 2r \text{ y } b = -r^2,$$

por lo que el sistema dinámico (4.41) se escribe como

$$A(n+2) = 2rA(n+1) - r^2A(n), n = 0, 1, \dots \quad (4.53)$$

a) Consideremos una nueva sucesión $B(n)$ definida mediante

$$A(n) = r^n B(n). \quad (4.54)$$

Sustituyendo en el sistema dinámico (4.53) obtenemos

$$r^{n+2}B(n+2) = A(n+2) = 2rA(n+1) - r^2A(n) = r^{n+2}(2B(n+1) - B(n))$$

para $n = 0, 1, \dots$. Puesto que $r \neq 0$, dividiendo por r^{n+2} en la expresión anterior, obtenemos

$$B(n+2) = 2B(n+1) - B(n), n = 0, 1, \dots,$$

es decir, hemos llegado al sistema dinámico (4.49) cuya solución general, según se ha visto en el Ejemplo 4.3.8, es

$$B(n) = c_1 + c_2n, n = 0, 1, \dots$$

con $c_1, c_2 \in \mathbb{R}$. Sustituyendo este valor en (4.54) se tiene que la solución general del sistema dinámico (4.53) es la dada en (4.51).



b) Debemos elegir las constantes c_1 y c_2 para que se verifiquen las condiciones iniciales. Por tanto, estas constantes vienen dadas como la solución del sistema lineal

$$\begin{cases} c_1 = A(0) = a_0 \\ rc_1 + rc_2 = A(1) = a_1 \end{cases} \Rightarrow c_1 = a_0 \text{ y } c_2 = \frac{a_1 - ra_0}{r}. \quad \square$$

Ejemplo 4.3.10 Consideremos el sistema dinámico lineal homogéneo de segundo orden

$$\begin{cases} A(0) = 3, A(1) = 4 \\ A(n+2) = 4A(n+1) - 4A(n), n = 0, 1, \dots \end{cases} \quad (4.55)$$

Puesto que el polinomio característico asociado al sistema anterior es

$$P(x) = x^2 - 4x + 4 = (x - 2)^2,$$

se tiene que la solución general de (4.55) es

$$A(n) = (c_1 + c_2n)2^n, n = 0, 1, \dots$$

con $c_1, c_2 \in \mathbb{R}$. Determinemos las constantes c_1 y c_2 para que se satisfagan las condiciones iniciales

$$\begin{cases} c_1 = A(0) = 3 \\ 2c_1 + 2c_2 = A(1) = 4 \end{cases} \Rightarrow c_1 = 3 \text{ y } c_2 = -1.$$

Por tanto, la solución del sistema dinámico (4.55) es

$$A(n) = (3 - n)2^n, n = 0, 1, \dots \quad \square$$

4.3.2 Sistemas afines: Solución general y particular

Para los sistemas afines de segundo orden la solución se expresa según se recoge en el siguiente resultado:

Teorema 4.3.11 Consideremos el sistema dinámico afín de segundo orden

$$A(n+2) = aA(n+1) + bA(n) + c, n = 0, 1, \dots \quad (4.56)$$

y sean r y s las raíces del polinomio característico asociado

$$P(x) = x^2 - ax - b. \quad (4.57)$$

La solución general del sistema dinámico (4.56) es

$$A(n) = \begin{cases} c_1r^n + c_2s^n + \frac{c}{1-a-b} & \text{si } r \neq s \text{ con } r \neq 1 \text{ y } s \neq 1 \\ (c_1 + c_2n)r^n + \frac{c}{1-a-b} & \text{si } r = s \neq 1 \\ c_1r^n + c_2 + \frac{c}{2-a}n & \text{si } r \neq 1 \text{ y } s = 1 \\ c_1 + c_2n + \frac{c}{2}n^2 & \text{si } r = s = 1 \end{cases}$$

para $n = 0, 1, 2, \dots$, donde las constantes $c_1, c_2 \in \mathbb{R}$ dependen de los valores iniciales.

DEMOSTRACIÓN. Es claro que si $A(n)$ y $B(n)$ son soluciones del sistema afín (4.56), entonces su diferencia es solución del sistema homogéneo asociado (4.41), correspondiente a tomar $c = 0$. Por tanto, la solución general del sistema afín (4.56) se obtiene como la suma de $A_h(n)$, solución general del sistema homogéneo asociado (4.41), y $A_p(n)$, una solución particular del sistema afín (4.56). Es decir, la solución general de (4.56) puede escribirse como

$$A(n) = A_h(n) + A_p(n), n = 0, 1, 2, \dots$$



Por los Teoremas 4.3.4 y 4.3.9 sabemos que la solución general del sistema lineal homogéneo asociado (4.41) viene dada por

$$A_h(n) = \begin{cases} c_1 r^n + c_2 s^n & \text{si } r \neq s \\ (c_1 + c_2 n) r^n & \text{si } r = s. \end{cases}$$

Queda por obtener, en cada caso, una solución particular del sistema no homogéneo (4.56). Veamos, en primer lugar, que se verifica

$$\begin{cases} r \neq 1 \text{ y } s \neq 1 & \text{si } b \neq 1 - a \\ r \neq 1 \text{ y } s = 1 & \text{si } b = 1 - a \text{ y } a \neq 2 \\ r = s = 1 & \text{si } b = -1 \text{ y } a = 2. \end{cases} \quad (4.58)$$

En efecto, basta observar que si una de las raíces es 1, entonces $P(1) = 0$ y, por tanto,

$$b = 1 - a,$$

en cuyo caso, las raíces de (4.57) son

$$x = \frac{a \pm \sqrt{a^2 + 4(1-a)}}{2} = \frac{a \pm \sqrt{(a-2)^2}}{2} = \frac{a \pm (a-2)}{2},$$

es decir,

$$s = 1 \text{ y } r = a - 1.$$

A la vista de (4.58), distinguimos los siguientes casos:

a) $r \neq 1$ y $s \neq 1$: Buscamos una solución particular de (4.56) de la forma

$$A_p(n) = K.$$

Sustituyendo en el sistema (4.56), obtenemos que

$$A_p(n+2) = aA_p(n+1) + bA_p(n) + c \Rightarrow K = aK + bK + c \Rightarrow (1-a-b)K = c \Rightarrow K = \frac{c}{1-a-b}$$

(recuérdese que, en este caso, $b \neq 1 - a$).

b) $r \neq 1$ y $s = 1$: Buscamos una solución particular de (4.56) de la forma

$$A_p(n) = Kn.$$

Puesto que, ahora $b = 1 - a$, al sustituir en el sistema (4.56), obtenemos que

$$\begin{aligned} A_p(n+2) &= aA_p(n+1) + (1-a)A_p(n) + c \Rightarrow K(n+2) = aK(n+1) + (1-a)Kn + c \\ &\Rightarrow (K - aK - (1-a)K)n + 2K - aK - c = 0 \Rightarrow (2-a)K = c \Rightarrow K = \frac{c}{2-a} \end{aligned}$$

(recuérdese que $a \neq 2$).

c) $r = s = 1$: Buscamos una solución particular de (4.56) de la forma

$$A_p(n) = Kn^2.$$

Puesto que, en este caso, $b = -1$ y $a = 2$, al sustituir en el sistema (4.56), obtenemos que

$$\begin{aligned} A_p(n+2) &= 2A_p(n+1) - A_p(n) + c \Rightarrow K(n+2)^2 = 2K(n+1)^2 - Kn^2 + c \\ &\Rightarrow K(n^2 + 4n + 4) = 2K(n^2 + 2n + 1) - Kn^2 + c \\ &\Rightarrow (K - 2K + K)n^2 + (4K - 4K)n + 4K - 2K - c = 0 \\ &\Rightarrow 2K = c \Rightarrow K = \frac{c}{2}. \quad \square \end{aligned}$$



Ejemplo 4.3.12 El polinomio característico asociado al sistema dinámico

$$A(n+2) = 5A(n+1) - 6A(n) - 3, \quad n = 0, 1, 2, \dots$$

es

$$P(x) = x^2 - 5x + 6 = (x-2)(x-3),$$

por lo que la solución general del mismo es

$$A(n) = c_1 2^n + c_2 3^n - \frac{3}{2}, \quad n = 0, 1, 2, \dots \quad \square$$

Ejemplo 4.3.13 Dado el sistema dinámico

$$A(n+2) = 3A(n+1) - 2A(n) - 5, \quad n = 0, 1, 2, \dots$$

su polinomio característico es

$$P(x) = x^2 - 3x + 2 = (x-1)(x-2),$$

por lo que su solución general viene dada por

$$A(n) = c_1 + c_2 2^n + 5n, \quad n = 0, 1, 2, \dots \quad \square$$

Ejemplo 4.3.14 El polinomio característico asociado al sistema dinámico

$$A(n+2) = 2A(n+1) - A(n) + 4, \quad n = 0, 1, 2, \dots$$

es

$$P(x) = x^2 - 2x + 1 = (x-1)^2.$$

Así pues, este sistema dinámico tiene como solución general

$$A(n) = c_1 + c_2 n + 2n^2, \quad n = 0, 1, 2, \dots \quad \square$$

4.3.3 Aplicaciones

Veamos un par de ejemplos de aplicación de los sistemas lineales de segundo orden con coeficientes constantes: el primero alecciona sobre el comportamiento que debe tenerse en los casinos, mientras que el segundo modeliza la endogamia genética.

Apuestas

Supongamos que vamos a realizar apuestas sucesivas en un juego que paga uno a uno, es decir, por cada euro jugado podemos ganar un euro o perderlo (esto ocurre, por ejemplo, al apostar en la ruleta a rojo/negro, impar/par, pasa/falta, ...; lo veremos en detalle más adelante). La situación que contemplamos es la siguiente: comenzamos con n euros y apostamos un euro cada vez hasta que, o bien nos arruinamos (nos quedamos sin dinero), o bien alcanzamos una cantidad de N euros (obviamente, $N > n$), en cuyo caso nos retiramos. Además, supondremos que en cada lance hay una probabilidad p de ganar y una probabilidad $q = 1 - p$ de perder.

Denotemos por $P(n)$ la probabilidad de arruinarse si en un momento dado se tienen n euros. Es claro que $P(0) = 1$ y $P(N) = 0$. Estas dos igualdades jugarán el papel de condiciones “iniciales” para el sistema dinámico al que llegaremos. Vamos a intentar calcular $P(n+1)$, la probabilidad de arruinarse si se tienen $n+1$ euros. Pueden ocurrir dos cosas: la primera (Caso 1) que en el siguiente lance ganemos y, después nos arruinemos y la segunda (Caso 2) que en el primer lance perdamos y, después, nos arruinemos. El valor buscado del $P(n+1)$ será la suma de la probabilidad de que ocurra lo primero más la probabilidad de que ocurra lo segundo.



La probabilidad de que ocurra el Caso 1 es el producto de la probabilidad de ganar en el lance por la probabilidad de arruinarse cuando se tienen $n+2$ euros (los $n+1$ de partida más el que acabamos de ganar), es decir,

$$pP(n+2),$$

mientras que la probabilidad de que ocurra el Caso 2 es el producto de la probabilidad de perder en el lance por la probabilidad de arruinarse cuando se tienen n euros (los $n+1$ de partida menos el que acabamos de perder),

$$qP(n).$$

Por consiguiente,

$$P(n+1) = pP(n+2) + qP(n), \quad n = 0, 1, \dots, N-2,$$

o, de forma equivalente,

$$P(n+2) = \frac{1}{p}P(n+1) - \frac{q}{p}P(n), \quad n = 0, 1, \dots, N-2.$$

La ecuación característica de este sistema es

$$x^2 - \frac{1}{p}x + \frac{q}{p} = 0,$$

cuyas raíces son $x = 1$ y $x = \frac{q}{p}$. Supongamos por el momento que $p \neq q$ (es decir, $p \neq 0,5$); entonces, ambas raíces son distintas y la solución del sistema viene dada, llamando $r = \frac{q}{p}$, por

$$P(n) = c_1(1)^n + c_2r^n = c_1 + c_2r^n.$$

Para determinar las constantes c_1 y c_2 utilizamos las condiciones “iniciales” y obtenemos

$$\begin{cases} 1 = P(0) = c_1 + c_2 \\ 0 = P(N) = c_1 + c_2r^N, \end{cases}$$

es decir,

$$c_2 = \frac{1}{1-r^N} \quad \text{y} \quad c_1 = \frac{r^N}{r^N-1}.$$

Consecuentemente,

$$P(n) = \frac{r^N - r^n}{r^N - 1}.$$

En el caso de que $p = q = 0,5$, la solución del sistema dinámico, al tener el polinomio característico $x = 1$ como raíz doble, viene dada por

$$P(n) = (c_1 + c_2n)1^n = c_1 + c_2n,$$

lo que, usando las dos condiciones “iniciales”, conduce a

$$P(n) = \frac{N-n}{N}.$$

La ruleta francesa

En la ruleta francesa hay 37 casillas numeradas del 0 al 36. Los números del 1 al 36 están coloreados, la mitad de rojo y la mitad de negro. El 0 es de un color distinto (verde). Son muchos los tipos de apuestas que pueden realizarse, pero las que son a uno a uno son las de color (se apuesta a rojo o a negro y se gana si el número que sale es del color elegido), paridad (se apuesta a par o impar y se gana si el número que sale



es de la paridad elegida; el cero se considera que no es ni par ni impar) o pasa/falta (si se elige “pasa”, se gana si sale un número entre 19 y 36; si se elige “falta”, se gana si sale un número entre 1 y 18).

Es claro que, en los tres casos, la probabilidad de ganar es

$$p = \frac{18}{37} = 0'486.$$

Estudiemos cuál es la probabilidad de arruinarnos si entramos con 30 euros y hacemos apuestas sucesivas de un euro en la ruleta hasta, o bien arruinarnos o bien conseguir doblar dicha cantidad (es decir, $n = 30$ y $N = 60$). Puesto que, en este caso,

$$r = \frac{q}{p} = \frac{\frac{19}{37}}{\frac{18}{37}} = \frac{19}{18} = 1'0\overline{5},$$

tendremos que

$$P(30) = \frac{r^{60} - r^{30}}{r^{60} - 1} = 0'835.$$

Es decir, la probabilidad de arruinarnos es muy alta. Y cuantas más veces juguemos es peor: la misma situación para $n = 100$ y $N = 200$ (es decir, también queremos doblar la cantidad inicial, pero necesitamos más lances para llegar a ella) da una probabilidad de arruinarse de 0'995.

Como se puede observar, la presencia del 0 es lo que desequilibra la balanza a favor de la banca. Si no existiera esta casilla, la probabilidad de ganar sería

$$p = \frac{18}{36} = 0'5,$$

y la probabilidad de arruinarnos en caso de que quisiéramos doblar el dinero inicial sería, como ya se ha dicho,

$$P(n) = \frac{N - n}{N} = \frac{2n - n}{2n} = \frac{n}{2n} = \frac{1}{2},$$

es decir, nos encontraríamos ante un juego justo.

Hay que destacar el muy distinto resultado para situaciones que están muy cercanas (la diferencia entre las dos probabilidades de ganar, $\frac{18}{37}$ y $\frac{1}{2}$, es sólo de 14 milésimas). La moraleja de todo este asunto es que la mejor estrategia para intentar doblar el dinero inicial es apostar todo una única vez y marcharse, tanto si se gana como si se pierde.

Endogamia

Nuestra intención es mostrar cómo, en poblaciones pequeñas y aisladas, los rasgos se perpetúan, incluso aunque originalmente fueran escasos. Vamos a considerar un ejemplo extremo, pero que da cuenta de cómo es la situación en general.

Spongamos que un determinado rasgo tiene cuatro alelos a, b, c y d y estamos interesados en los individuos homocigóticos, esto es, los que tienen los genes $(a, a), (b, b), (c, c)$ o (d, d) . Partimos de una pareja en la que la hembra es (a, b) y el macho (c, d) . Éstos se acoplan y tienen un hijo y una hija en la generación 1. Estos dos hermanos se acoplan, a su vez, y tienen un hijo y una hija (nietos de la pareja original) y así sucesivamente. Denominamos $P(n)$ la probabilidad de que un individuo de la generación n sea homocigótico (da igual el tipo).

Por otra parte, si llamamos $A(n)$ la probabilidad de que, dada una pareja (heterosexual) de hermanos de la generación n , al tomar un alelo de cada uno de ellos, ambos sean del mismo tipo, es claro que

$$P(n + 1) = A(n).$$

En la generación 1, cada individuo toma un alelo de su madre (a o b) y uno de su padre (c o d); por tanto, $P(1) = 0$. De forma más obvia aún, $P(0) = 0$.



¿Cuál es la probabilidad de que un individuo de la generación 2 sea homocigótico? Veamos la probabilidad de que sea (a, a) , por ejemplo, y multipliquemos por cuatro. Para llegar en la segunda generación a un individuo (a, a) hace falta que la madre pase el alelo a a su hija (lo que ocurre con probabilidad $\frac{1}{2}$), que se lo pase al hijo (análogamente, con probabilidad $\frac{1}{2}$), que la hija se lo pase al nieto (otra vez probabilidad $\frac{1}{2}$) y que el hijo se lo pase también al nieto (probabilidad $\frac{1}{2}$ de nuevo). Así la probabilidad de que el nieto sea (a, a) es

$$\left(\frac{1}{2}\right)^4 = \frac{1}{16}.$$

La probabilidad de que el individuo de la segunda generación sea homocigótico es, por tanto, $P(2) = 0'25$.

Consideremos ahora las distintas formas en que un individuo de la segunda generación puede conseguir su par de alelos a partir de los de sus abuelos. Hay tres posibilidades: Los dos alelos son copia de uno de los de alguno de sus abuelos, esto es, (a, a) , (b, b) , (c, c) o (d, d) ; el individuo tiene el mismo par de alelos que tenía uno de sus abuelos, (a, b) o (c, d) ; cada alelo proviene de uno de los dos abuelos, (a, c) , (a, d) , (b, c) o (b, d) .

En general, para la generación $n + 2$ consideramos los siguientes casos:

C1 Los dos alelos de un individuo de la generación $n + 2$ son copia de uno de los de alguno de sus abuelos de la generación n . Los dos abuelos tienen, entre ellos, cuatro alelos. La probabilidad de que el nieto repita uno en concreto de ellos es (como en el caso de la segunda generación) $\frac{1}{16}$. Así, la probabilidad de C1 es $P(C1) = 0'25$.

C2 El individuo tiene los dos mismos alelos que uno de sus abuelos. Razonando como antes, para que un abuelo (x, y) dé un nieto (x, y) una posibilidad es que el abuelo pase el alelo x a su hija (lo que ocurre con probabilidad $\frac{1}{2}$), que le pase el alelo y al hijo (análogamente, con probabilidad $\frac{1}{2}$), que la hija le pase al nieto el alelo x (otra vez probabilidad $\frac{1}{2}$) y que el hijo le pase el alelo y al nieto (probabilidad $\frac{1}{2}$ de nuevo); esto tiene probabilidad $\frac{1}{16}$. La otra posibilidad es la análoga, intercambiando los papeles de x e y . Así pues, la probabilidad de que el individuo tenga dos alelos distintos provenientes de uno de los abuelos es $\frac{1}{8}$. Como el individuo tiene dos abuelos, la probabilidad de que el individuo tenga dos alelos distintos provenientes del mismo abuelo es $P(C2) = 0'25$.

C3 Cada alelo proviene de uno de los dos abuelos. Como éste es el caso complementario de los otros dos, debe ocurrir que

$$P(C1) + P(C2) + P(C3) = 1,$$

por lo que $P(C3) = 0'5$.

Estamos ya en disposición de calcular $P(n + 2)$, la probabilidad de que un individuo de la generación $n + 2$ sea homocigótico. Vamos a distinguir tres casos: que ocurra C1 y el individuo sea homocigótico, que ocurra C2 y el individuo sea homocigótico y, finalmente, que ocurra C3 y el individuo sea homocigótico.

c1 Si los alelos de un individuo son ambos copia del mismo alelo de uno de sus abuelos, es que el individuo es homocigótico, por lo que $P(c1) = P(C1) = 0'25$.

c2 Si el individuo tiene los dos mismos alelos que uno de sus abuelos, sólo puede ser homocigótico si ese abuelo lo era. Pero la probabilidad de que un individuo de la generación n sea homocigótico es lo que hemos denominado $P(n)$. Por tanto,

$$P(c2) = P(C2)P(n) = 0'25P(n).$$

c3 Si cada alelo proviene de uno de los dos abuelos, ¿cuál es la probabilidad de que ambos sean iguales? La respuesta ya la hemos dado: es lo que hemos llamado $A(n)$, la probabilidad de que dos alelos, uno tomado de un hermano (el abuelo) y otro tomado de una hermana (la abuela) sean iguales. Como ya habíamos dicho, $A(n) = P(n + 1)$. Podemos concluir entonces que

$$P(c3) = P(C3)P(n + 1) = 0'5P(n + 1).$$



Recopilando los tres casos tenemos que

$$P(n+2) = 0'5P(n+1) + 0'25P(n) + 0'25.$$

El polinomio característico de este sistema dinámico tiene como raíces

$$r = \frac{1 + \sqrt{5}}{4} \simeq 0'809 \text{ y } s = \frac{1 - \sqrt{5}}{4} \simeq -0'309,$$

por lo que la solución general del sistema dinámico afín de segundo orden es

$$P(n) = c_1 \left(\frac{1 + \sqrt{5}}{4} \right)^n + c_2 \left(\frac{1 - \sqrt{5}}{4} \right)^n + 1$$

(véase el Teorema 4.3.11), donde c_1, c_2 dependen de los valores iniciales.

En nuestro caso, el valor que toman las constantes c_1 y c_2 es

$$c_1 = -\frac{3\sqrt{5} + 5}{10} \text{ y } c_2 = \frac{3\sqrt{5} - 5}{10} \quad (4.59)$$

(compruébese). No obstante, merece la pena destacar que, independientemente del valor que tomen las constantes c_1 y c_2 , la solución siempre va a tender al valor de equilibrio $a_0 = 1$ cuando n tiende a infinito. Esto debe interpretarse como que, esta evolución endogámica conduce de forma segura a individuos homocigóticos. Como hay cuatro alelos, la probabilidad de que un individuo sea, por ejemplo, (a, a) tras un número alto de generaciones es cercana a 0.25. Si el alelo a causara un rasgo no deseado, aunque fuera recesivo y poco probable en la generación inicial, una apareamiento endogámico como el descrito llevaría a una alta proporción de individuos con este rasgo. Es por esto por lo que la endogamia en humanos se desaconseja fuertemente (en prácticamente todas las civilizaciones el incesto está repudiado).

4.4 Sistemas dinámicos no lineales: Estabilidad y caos

4.4.1 Estabilidad

Ejemplo 4.4.1 Para encontrar los puntos de equilibrio del sistema dinámico no lineal

$$A(n+1) = (A(n) + 4)A(n) + 2 \quad (4.60)$$

resolvemos la ecuación $a = (a + 4)a + 2$, es decir,

$$a^2 + 3a + 2 = 0 \Rightarrow (a + 2)(a + 1) = 0.$$

Por tanto, $A(n) = -2$ y $A(n) = -1$ son los dos puntos de equilibrio del sistema anterior.

En la Tabla 4.3 se muestran los primeros valores $A(n)$ del sistema dinámico (4.60) para tres datos iniciales distintos: $A(0) = -1'01$, $A(0) = -0'99$ y $A(0) = -2'4$. Asimismo, en la Figura 4.9 se muestran los puntos $(n, A(n))$ del sistema dinámico (4.60) para los datos iniciales anteriores. A la vista de los resultados anteriores, intuitivamente se ve que

$$\begin{cases} A(0) = -1'01 \Rightarrow \lim_{n \rightarrow +\infty} A(n) = -2 \\ A(0) = -0'99 \Rightarrow \lim_{n \rightarrow +\infty} A(n) = +\infty \\ A(0) = -2'4 \Rightarrow \lim_{n \rightarrow +\infty} A(n) = -2. \end{cases}$$

Este tipo de comportamientos dió lugar a las nociones de punto de equilibrio estable, inestable y neutro recogidas en la Definición 4.2.9. \square

Veamos una primera caracterización de la estabilidad de los puntos de equilibrio de los sistemas no lineales de primer orden.



n	$A(n)$	$A(n)$	$A(n)$
0	-1'010000	$-9'900000 \times 10^{-1}$	-2'400000
1	-1'019900	$-9'799000 \times 10^{-1}$	-1'840000
2	-1'039404	$-9'593960 \times 10^{-1}$	-1'974400
3	-1'077255	$-9'171433 \times 10^{-1}$	-1'999345
4	-1'148542	$-8'274214 \times 10^{-1}$	-2'000000
5	-1'275020	$-6'250593 \times 10^{-1}$	-2'000000
6	-1'474404	$-1'095381 \times 10^{-1}$	-2
7	-1'723748	$1'573846 \times 10^0$	-2
8	-1'923685	$1'077238 \times 10^1$	-2
9	-1'994176	$1'611336 \times 10^2$	-2
10	-1'999966	$2'661057 \times 10^4$	-2
11	-2'000000	$7'082287 \times 10^8$	-2

TABLA 4.3: Valores $A(n)$ del sistema dinámico (4.60) con distintos datos iniciales.

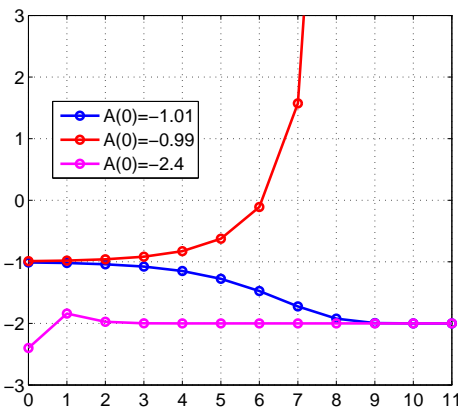


Figura 4.9: Puntos $(n, A(n))$ del sistema dinámico (4.60) con distintos datos iniciales.

Teorema 4.4.2 Consideremos el sistema autónomo

$$A(n + 1) = f(A(n)), \quad n = 0, 1, 2, \dots$$

Sea a un punto de equilibrio de dicho sistema y supongamos que f tiene derivada continua en un entorno de a . El punto de equilibrio a es:

a) estable si $|f'(a)| < 1$.

b) inestable si $|f'(a)| > 1$.

Si $|f'(a)| = 1$, el resultado no es concluyente.

DEMOSTRACIÓN.

a) En el caso de que $|f'(a)| < 1$ se verifica que existen $\varepsilon > 0$ (suficientemente pequeño) y $b \in (|f'(a)|, 1)$ tales que

$$|f'(x)| < b, \quad x \in I = (a - \varepsilon, a + \varepsilon).$$

Eligiendo $A(0) \in I$, veamos que $|A(1) - a|$ es más pequeño que $|A(0) - a|$. En efecto, puesto que $a = f(a)$ (por ser punto de equilibrio), se tiene que

$$|A(1) - a| = |f(A(0)) - a| = |f(A(0)) - f(a)|. \tag{4.61}$$



Por otro lado, el *teorema del Valor Medio* garantiza la existencia de un punto x_0 , localizado entre $A(0)$ y a , tal que

$$f(A(0)) - f(a) = f'(x_0)(A(0) - a) \quad (4.62)$$

por lo que, sustituyendo (4.62) en (4.61), se verifica que

$$|A(1) - a| = |f(A(0)) - a| = |f'(x_0)||A(0) - a|. \quad (4.63)$$

Ahora bien, como x_0 está entre $A(0)$ y a , se tiene que $x_0 \in I$ y, por tanto,

$$|f'(x_0)| < b.$$

De esta forma, a partir de (4.63), se verifica que

$$|A(1) - a| < b|A(0) - a|.$$

Dado que $A(1)$ está más cercano a a que $A(0)$, se tiene que $A(1) \in I$, por lo que podemos reiterar el proceso anterior, obteniendo que

$$|A(2) - a| < b|A(1) - a| < b^2|A(0) - a|$$

y, en general,

$$|A(n) - a| < b^n|A(0) - a|, \quad n \in \mathbb{N}.$$

Consecuentemente, al ser $b < 1$, se verifica que

$$\lim_{n \rightarrow +\infty} |A(n) - a| = 0$$

o, equivalentemente,

$$\lim_{n \rightarrow +\infty} A(n) = a,$$

lo que indica que a es un punto de equilibrio estable.

b) Si $|f'(a)| > 1$ se verifica que existen $\varepsilon > 0$ (suficientemente pequeño) y $b \in (1, |f'(a)|)$ tales que

$$|f'(x)| > b, \quad x \in I = (a - \varepsilon, a + \varepsilon). \quad (4.64)$$

Eligiendo $A(0) \in I$, $A(0) \neq a$, veamos que existe algún valor de $n \in \mathbb{N}$ para el cual $A(n) \notin I$. Comencemos viendo que $|A(1) - a|$ es mayor que $|A(0) - a|$. En efecto, argumentando como en el apartado a), se verifica que

$$|A(1) - a| = |f(A(0)) - f(a)| = |f'(x_0)||A(0) - a| > b|A(0) - a|,$$

por lo que pueden presentarse dos casos:

- i) Si $A(1) \notin I$, ya tenemos el resultado.
- ii) Si $A(1) \in I$, repetimos el proceso con $A(2)$, obteniendo que

$$|A(2) - a| > b|A(1) - a| > b^2|A(0) - a|.$$

Nuevamente,

- 1) Si $A(2) \notin I$, ya tenemos el resultado.
- 2) Si $A(2) \in I$, repetimos el proceso con $A(3)$ y, así, sucesivamente.

En general, se demuestra por inducción que: o bien $A(n) \notin I$ o bien

$$|A(n+1) - a| > b|A(n) - a| > b^{n+1}|A(0) - a|.$$

Así pues, como I es un intervalo finito y

$$\lim_{n \rightarrow +\infty} b^n = +\infty$$

(ya que $b > 1$), tomando n suficientemente grande se tiene que, o bien $A(n) \notin I$, o bien la distancia de $A(n+1)$ al valor a es mayor que la longitud de I y $A(n+1) \notin I$. \square



Observación 4.4.3 La tangente a una curva en un punto (cuya pendiente es su derivada) es, en algún sentido, la mejor aproximación lineal de la curva en ese punto. Una de las principales razones de la importancia de la derivada es que, en muchos casos, esta aproximación lineal nos da toda la información que necesitamos. En particular, si a es un punto de equilibrio y $|f'(a)| < 1$ entonces, como acabamos de ver, se verifica que

$$\lim_{n \rightarrow +\infty} A(n) = a$$

siempre que $A(0)$ esté suficientemente cerca de a . Cuando $f'(a) = \pm 1$ la derivada primera no da suficiente información; ésta puede obtenerse estudiando las derivadas sucesivas, aunque aquí no lo haremos. \square

Ejemplo 4.4.4

a) El sistema dinámico (4.60) viene definido por la función $f(x) = x^2 + 4x + 2$; puesto que $f'(-2) = 0 < 1$ y $f'(-1) = 2 > 1$, se tiene que los puntos de equilibrio $a = -2$ y $a = -1$ son, respectivamente, estable e inestable.

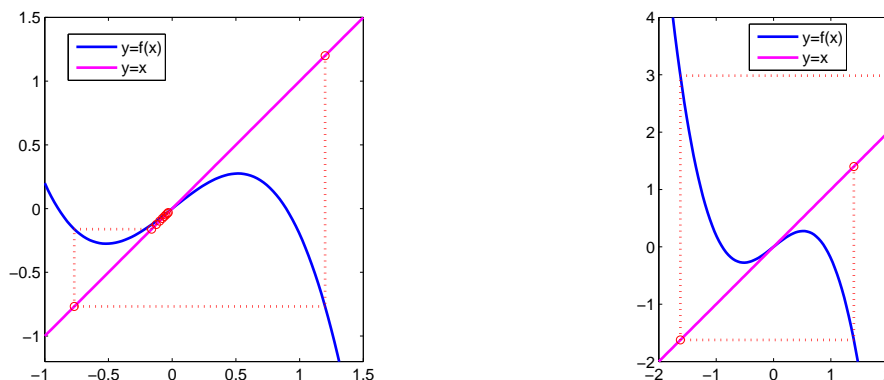
b) Los puntos de equilibrio del sistema dinámico no lineal

$$A(n + 1) = 0'8 A(n) - A^3(n), \quad n \in \mathbb{N} \cup \{0\}. \tag{4.65}$$

son las raíces de la ecuación

$$a = 0'8a - a^3 \Rightarrow a^3 = -0'2a \Rightarrow a = 0 \text{ (las otras dos raíces son complejas conjugadas).}$$

En este caso $f(x) = 0'8x - x^3$ y $f'(x) = 0'8 - 3x^2$. En consecuencia, puesto que $f'(0) = 0'8$, el punto de equilibrio $a = 0$ es estable. En la Figura 4.10 se ve cómo los puntos de la gráfica son atraídos hacia el origen cuando $A(0)$ está cercano a 0 ($A(0) = 1'2$ en este caso), mientras que van formando ciclos hacia afuera (con $|A(n)| \rightarrow +\infty$ si $n \rightarrow +\infty$) cuando $A(0)$ está suficientemente alejado de 0 (por ejemplo, tomando $A(0) = 1'4$).



(a) $A(0) = 1'2$.

(b) $A(0) = 1'4$.

Figura 4.10: Gráfica del sistema dinámico (4.65) con $f(x) = 0'8x - x^3$.

c) Los puntos de equilibrio del sistema dinámico no lineal

$$A(n + 1) = 3'2 A(n) - 0'8 A^2(n), \quad n \in \mathbb{N} \cup \{0\}. \tag{4.66}$$

son las raíces de la ecuación

$$a = 3'2a - 0'8a^2 \Rightarrow 0'8a^2 = 2'2a \Rightarrow a = 0 \text{ y } a = 2'75.$$

Puesto que $f(x) = 3'2x - 0'8x^2$, tenemos

$$f'(x) = 3'2 - 1'6x = 1'6(2 - x).$$



Como $f'(0) = 3/2$ y $f'(2/75) = -1/2$, ambos puntos de equilibrio son inestables. En la Figura 4.11 se ve cómo los puntos de la gráfica son repelidos de los puntos de equilibrio a pesar de comenzar con valores próximos a éstos. Nótese cómo, aunque ambos puntos de equilibrio son inestables, no sabemos cuál va a ser el comportamiento de $A(n)$ (en los dos casos parece que se llega a un ciclo de orden 2).

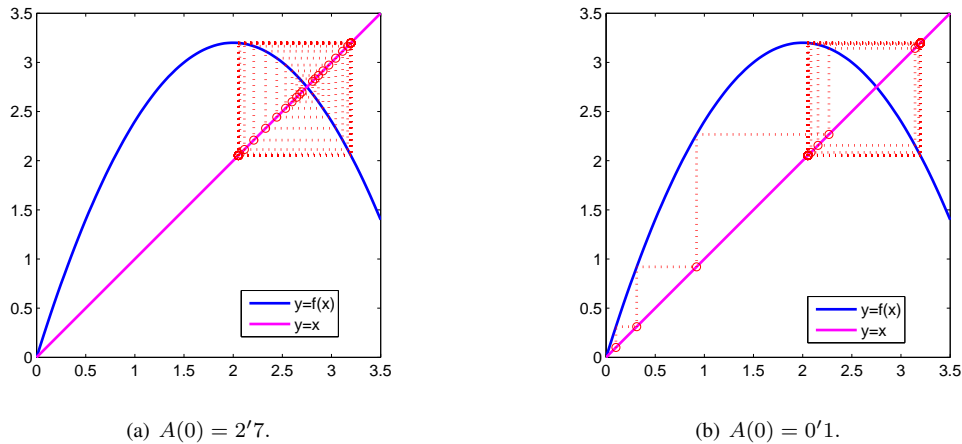


Figura 4.11: Gráfica del sistema dinámico (4.66) con $f(x) = 3/2x - 0/8x^2$.

d) Los puntos de equilibrio del sistema dinámico no lineal

$$A(n + 1) = A^3(n) - A^2(n) + 1, \quad n \in \mathbb{N} \cup \{0\}. \tag{4.67}$$

son las raíces de la ecuación

$$a = a^3 - a^2 + 1 \Rightarrow 0 = a^3 - a^2 - a + 1 = (a + 1)(a - 1)^2 \Rightarrow a = -1 \text{ y } a = 1.$$

Puesto que $f(x) = x^3 - x^2 + 1$ se tiene que

$$f'(x) = 3x^2 - 2x = x(3x - 2).$$

Por un lado, puesto que $f'(-1) = 5$, el punto de equilibrio $a = -1$ es inestable. Por otra parte, como

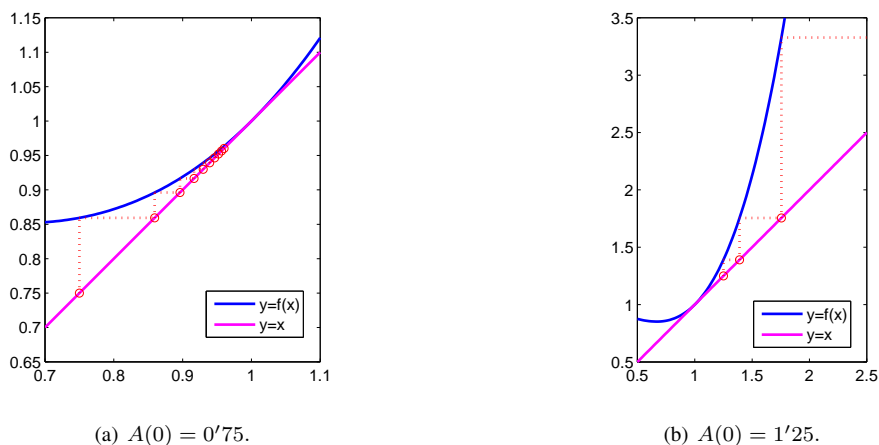


Figura 4.12: Gráfica del sistema dinámico (4.67) con $f(x) = x^3 - x^2 + 1$.

$f'(1) = 1$, el Teorema 4.4.2 no nos permite concluir la estabilidad o no del punto de equilibrio $a = 1$; no obstante, de la Figura 4.12 parece deducirse que el punto de equilibrio $a = 1$ es estable por la izquierda e inestable por la derecha. Este tipo de puntos de equilibrio suelen denominarse *semiestables*. □



Observación 4.4.5 Cuando un punto de equilibrio es estable, un problema interesante (y, a menudo, complicado) consiste en determinar el “intervalo maximal” (c, d) en el que se verifique que

$$\lim_{n \rightarrow +\infty} A(n) = a \text{ para todo } A(0) \in (c, d).$$

Mediante técnicas gráficas, estos intervalos maximales pueden determinarse de forma bastante precisa. Por ejemplo, consideremos el sistema dinámico no lineal

$$A(n + 1) = 1'5 A(n) - 0'5 A^2(n), \quad n \in \mathbb{N} \cup \{0\}. \tag{4.68}$$

Los puntos fijos del sistema dinámico (4.68) son las raíces de la ecuación

$$a = 1'5a - 0'5a^2 \Rightarrow 0'5a^2 = 0'5a \Rightarrow a = 0 \text{ y } a = 1$$

(siendo $a = 0$ inestable y $a = 1$ estable). En la Figura 4.13 se muestra la representación gráfica del sistema dinámico (4.68) cuando $A(0) = 0'2$; se aprecia que los puntos van hacia el punto $a = 1$. Si construimos las gráficas asociadas a otros valores de $A(0)$ podemos observar que, cuando $A(0)$ está “razonablemente” cercano a 1, los puntos son atraídos por $a = 1$ mientras que, si $A(0)$ está cercano a 0, son repelidos de 0 (aunque no necesariamente vayan a infinito). De hecho, puede comprobarse que

$$\begin{cases} \lim_{n \rightarrow +\infty} A(n) = 1 & \text{si } A(0) \in (0, 3) \\ \lim_{n \rightarrow +\infty} |A(n)| = +\infty & \text{si } A(0) \notin [0, 3], \end{cases}$$

por lo que $(0, 3)$ es el intervalo maximal del punto de equilibrio $a = 1$.

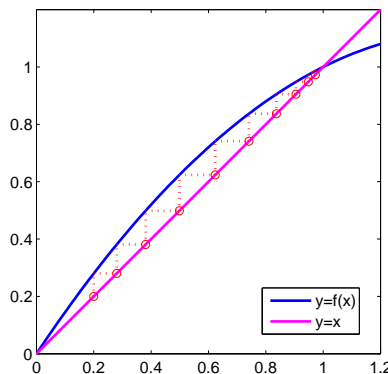


Figura 4.13: Gráfica del sistema dinámico (4.68) con $f(x) = 1'5x - 0'5x^2$ y $A(0) = 0'2$.

En otras ocasiones, los resultados obtenidos con el ordenador pueden ayudar a estimar dicho intervalo maximal. Así, por ejemplo, en la Figura 4.10 puede verse que dicho intervalo es mayor que $(-1'2, 1'2)$, pero más pequeño que $(-1'4, 1'4)$; dando a $A(0)$ valores comprendidos entre $1'2$ y $1'4$ podemos estimar el intervalo maximal con la precisión deseada. □

4.4.2 La ecuación logística

Estamos interesados en estudiar la evolución de una determinada especie. Para ello, denotemos por $A(n)$ el número de individuos de la población en el instante de tiempo n y supongamos que por cada individuo existente en el periodo n habrá, por término medio, α individuos en el periodo $n + 1$, es decir,

$$A(n + 1) = \alpha A(n), \quad n = 0, 1, 2, \dots \tag{4.69}$$

conocida como *ecuación de Malthus*, propuesta por un economista y pensador del siglo XIX para estimar la evolución de la población humana. En el caso de que $\alpha > 1$, es decir, si existe algún crecimiento vegetativo



en la población, los valores de $A(n)$ crecen en progresión geométrica y se disparan de forma exponencial, razón por la que esta ecuación desató una fuerte polémica entre los contemporáneos de Malthus, suponiendo el primer aldabonazo en la conciencia colectiva de la humanidad sobre el problema de la sobrepoblación del planeta.

En 1976, el biólogo Robert May formuló otra ecuación (que difería de la de Malthus) para estudiar el crecimiento de una población de insectos en un ecosistema cerrado. May tuvo en cuenta los efectos de saturación del ecosistema: cuando la población se acerca al máximo posible que el medio ambiente puede sustentar, el parámetro α debe disminuir, lo que equivale a considerar que este parámetro sea función del número de individuos. Con ello se llega a una ecuación del tipo

$$A(n+1) = \alpha(A(n))A(n), \quad n = 0, 1, 2, \dots$$

Podemos tomar como unidad de medida el tamaño máximo posible de población, de manera que $A(n)$ exprese la fracción de población existente en el periodo n con respecto al nivel máximo de población; en tal caso, se tendrá que $A(n) \in [0, 1]$. May formuló la hipótesis de que $\alpha(A(n))$ debería decrecer linealmente cuando $A(n)$ creciera, hasta hacerse nula cuando $A(n)$ tomara el valor 1; para ello consideró la función

$$\alpha(A(n)) = r(1 - A(n)),$$

que conduce a la *ecuación logística*

$$A(n+1) = rA(n)(1 - A(n)), \quad n = 0, 1, 2, \dots \quad (4.70)$$

con dato inicial $A(0) \in [0, 1]$. El hecho de que los valores de $A(n)$ deban permanecer en el intervalo $[0, 1]$ restringe los posibles valores del parámetro r , de forma que $r \in [0, 4]$; esto es así debido a que la parábola $f(x) = rx(1-x)$ vale cero en 0 y 1 y toma su valor máximo $\frac{r}{4}$ en $x = \frac{1}{2}$.

Nótese que para valores pequeños de $A(n)$ se verifica que $1 - A(n) \simeq 1$, por lo que la ecuación (4.70) equivale a la ecuación de Malthus (4.69) con parámetro $\alpha = r$

$$A(n+1) \simeq rA(n), \quad n = 0, 1, 2, \dots$$

El término $-rA^2(n)$ es un *término de amortiguamiento* puesto que amortigua el crecimiento de la población, evitando que se vaya a infinito.

Quedan varias preguntas que deben ser respondidas como, por ejemplo, ¿Se puede encontrar la solución general de la ecuación logística?, ¿cuáles son los puntos de equilibrio de la ecuación logística?, ¿para qué valores de r los puntos de equilibrio son estables?, cuando un punto de equilibrio es estable, ¿para qué datos iniciales a_0 la solución $A(n)$ tiende a algún punto de equilibrio? Para dar una respuesta negativa a la primera pregunta, en la mayoría de los casos va a ser imposible encontrar la solución general de un sistema dinámico no lineal: podemos determinar con el ordenador los valores $A(0), A(1), \dots, A(n)$ para valores grandes de n pero lo que no vamos a encontrar, en general, es una “simple” expresión algebraica para $A(n)$ en términos de n . Este hecho proporciona mucha inseguridad pues, independientemente del número de valores $A(n)$ que determinemos, siempre queda la duda de que algo extraño puede ocurrir en el siguiente valor que calculemos. Ante la imposibilidad de encontrar la solución de nuestra ecuación, desarrollaremos técnicas para determinar cuándo un punto de equilibrio es estable (nótese que el hecho de que un punto de equilibrio sea atractor proporciona mucha información acerca del comportamiento, para tiempos grandes, de las soluciones de un sistema dinámico).

Ejemplo 4.4.6 En el caso de que $r = 2$, la ecuación logística (4.70) toma la forma

$$A(n+1) = 2A(n)(1 - A(n)), \quad n = 0, 1, \dots \quad (4.71)$$

En la Figura 4.14 se muestran las gráficas de los primeros valores $(n, A(n))$ para los datos iniciales $a_0 \in \{0'05, 0'75, 0'9, -0'1\}$ (nótese que el dato inicial $a_0 = -0'1$ no es realista, pero es interesante desde el punto de vista matemático). Busquemos los puntos de equilibrio del sistema dinámico (4.71):

$$a = 2a(1 - a) = 2a - 2a^2 \Leftrightarrow a(2a - 1) = 0 \Leftrightarrow a = 0 \text{ o } a = \frac{1}{2} = 0'5.$$



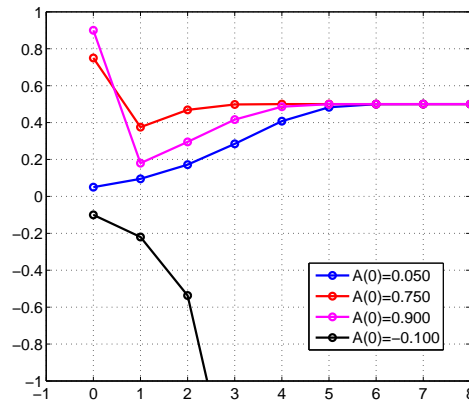


Figura 4.14: Puntos $(n, A(n))$ de la ecuación logística (4.71) para diversos datos iniciales.

Esto tiene sentido, pues una población de tamaño cero permanece siempre en cero y, si $a_0 = 0.5$, la razón de crecimiento es uno y, de nuevo, el tamaño de la población permanece constante. Puesto que

$$f(x) = 2x(1 - x) = 2x - 2x^2,$$

se tiene que

$$f'(x) = 2 - 4x = 2(1 - 2x)$$

y, por tanto,

$$f'(0) = 2 > 1 \text{ y } f'(0.5) = 0 < 1,$$

por lo que el Teorema 4.4.2 nos permite afirmar que $a = 0$ es un punto de equilibrio inestable del sistema dinámico (4.71) mientras que $a = 0.5$ es un punto de equilibrio estable de dicho sistema. Nótese en la Figura 4.14 que la línea horizontal $y = 0.5$ es el límite al que tiende el tamaño de la población cuando $a_0 \in \{0.05, 0.75, 0.9\}$; de hecho, puede demostrarse que

$$\lim_{n \rightarrow +\infty} A(n) = 0.5$$

para cualquier dato inicial $a_0 \in (0, 1)$ (los extremos de este intervalo maximal de convergencia son las raíces de la función f). □

Observación 4.4.7 En general, los puntos de equilibrio de la ecuación logística (4.70) con $r \neq 0$ ⁶ son

$$a = 0 \text{ y } a = \frac{r - 1}{r}.$$

En efecto, basta observar que

$$a = ra - ra^2 \Leftrightarrow a(ra - r + 1) = 0 \Leftrightarrow a = 0 \text{ o } a = \frac{r - 1}{r}.$$

Para estudiar la estabilidad de estos puntos de equilibrio, puesto que

$$f'(x) = r - 2rx = r(1 - 2x),$$

se tiene que

$$f'(0) = r \text{ y } f'\left(\frac{r - 1}{r}\right) = r\left(1 - \frac{2(r - 1)}{r}\right) = r\frac{2 - r}{r} = 2 - r$$

⁶Cuando $r = 0$ la ecuación logística (4.70) toma la forma $A(n + 1) = 0$, $n = 0, 1, 2, \dots$ y tiene un único punto de equilibrio $a = 0$ que, además, es estable.



y, por tanto,

$$\left\{ \begin{array}{l} a = 0 \text{ es un punto de equilibrio} \\ a = \frac{r-1}{r} \text{ es un punto de equilibrio} \end{array} \right\} \left\{ \begin{array}{l} \text{estable si } r \in (-1, 1) \\ \text{inestable si } r \in (-\infty, -1) \cup (1, +\infty) \\ \text{estable si } r \in (1, 3) \\ \text{inestable si } r \in (-\infty, 0) \cup (0, 1) \cup (3, +\infty). \end{array} \right.$$

Para el caso que nos ocupa de la modelización del tamaño de una población, en el que los valores admisibles del parámetro r están en el intervalo $[0, 4]$, tendremos

$$\left\{ \begin{array}{l} a = 0 \text{ es un punto de equilibrio} \\ a = \frac{r-1}{r} \text{ es un punto de equilibrio} \end{array} \right\} \left\{ \begin{array}{l} \text{estable si } r \in [0, 1) \\ \text{inestable si } r \in (1, 4] \\ \text{estable si } r \in (1, 3) \\ \text{inestable si } r \in (0, 1) \cup (3, 4]. \end{array} \right.$$

Debe observarse que, para el segundo valor del punto de equilibrio a , si $r \in (0, 1)$ entonces $a \notin [0, 1]$, por lo que es un caso en el que no estamos interesados. Finalmente, destacamos que queda por determinar la estabilidad de los puntos de equilibrio anteriores cuando el valor absoluto de la derivada es uno, esto es, la estabilidad de $a = 0$ para $r = 1$ y de $a = \frac{2}{3}$ para $r = 3$ (puede demostrarse que ambos son estables). \square

Ejemplo 4.4.8 Supongamos ahora que tomamos $r = 2'5$ en la ecuación logística (4.70), por lo que ésta toma la forma

$$A(n+1) = 2'5A(n)(1 - A(n)), \quad n = 0, 1, 2, \dots \quad (4.72)$$

Como se ha visto en la Observación 4.4.7, los puntos de equilibrio de la ecuación logística (4.72) son $a = 0$ y $a = \frac{2'5-1}{2'5} = \frac{3}{5} = 0'6$, siendo el primero inestable y el segundo estable.

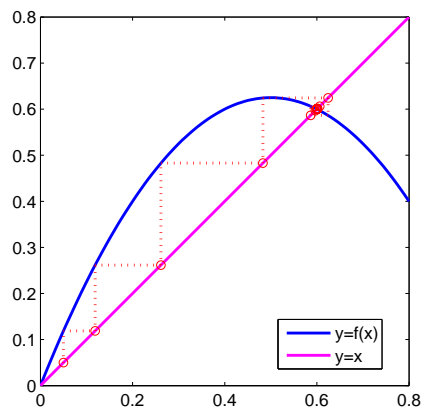


Figura 4.15: Gráfica del sistema dinámico (4.72) con $a_0 = 0'05$.

Por ser $a = 0'6$ estable, se tendrá que

$$\lim_{n \rightarrow +\infty} A(n) = 0'6$$

siempre que el primer valor $A(0)$ esté cerca de $0'6$. De hecho, se puede demostrar que cuando $r \in (1, 3)$ dicha propiedad se verifica para todo $A(0) \in (0, 1)$. En la Figura 4.15 se ilustra este hecho cuando se toma $A(0) = 0'05$.



También se puede probar un comportamiento análogo para el otro punto de equilibrio $a = 0$ y los valores $r \in [0, 1)$ en los que éste es estable; esto es, se verifica que

$$\lim_{n \rightarrow +\infty} A(n) = 0$$

para todo $A(0) \in [0, 1]$. \square

4.4.3 Ciclos y caos

En la Sección 4.4.2 se ha demostrado que cuando $r \in [0, 3]$ uno de los dos puntos de equilibrio de la ecuación logística es estable y que, para valores $r \in (3, 4]$, los dos puntos de equilibrio son inestables. En esta sección vamos a estudiar el comportamiento de la sucesión $\{A(n)\}_{n=0}^{\infty}$ para estos últimos valores del parámetro.

Ejemplo 4.4.9 Consideremos la ecuación logística con $r = 3'2$, es decir,

$$A(n + 1) = 3'2A(n)(1 - A(n)), \quad n = 0, 1, 2, \dots \tag{4.73}$$

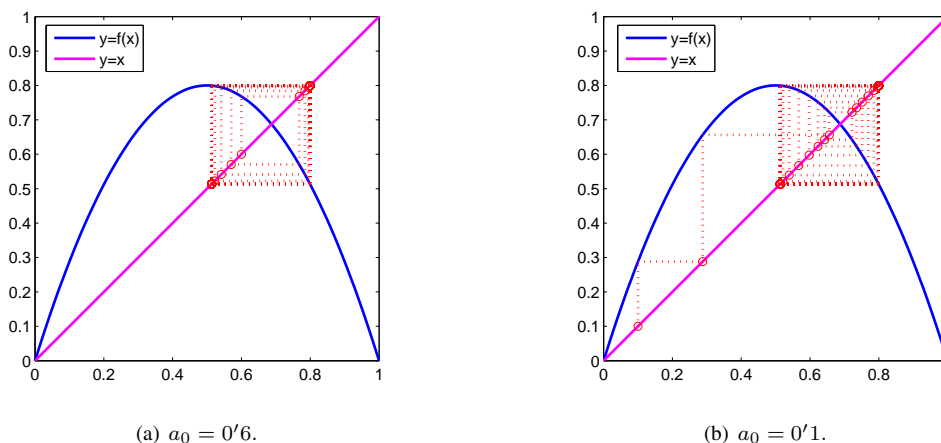


Figura 4.16: Gráfica del sistema dinámico (4.73).

Como ya sabemos, los dos puntos de equilibrio $a = \frac{11}{16} = 0'6875$ y $a = 0$ son inestables. En la Figura 4.16 se muestran las representaciones gráficas del sistema (4.73) donde se han tomado datos iniciales cercanos a los de equilibrio: $a_0 = 0'6$ en la gráfica izquierda y $a_0 = 0'1$ en la gráfica derecha. En ambos casos parece que la sucesión $\{A(n)\}_{n=0}^{\infty}$ tiende a un ciclo de orden 2. \square

Demos una primera definición de la estabilidad de un ciclo de segundo orden:

Definición 4.4.10 Un ciclo de orden 2 de un sistema dinámico de primer orden que tome los valores a_1 y a_2 es *estable* (o *atractor*) si existen entornos (c_1, d_1) y (c_2, d_2) de a_1 y a_2 , respectivamente, de forma que:

- a) Si $A(0) \in (c_1, d_1) \Rightarrow \lim_{n \rightarrow +\infty} A(2n) = a_1$ y $\lim_{n \rightarrow +\infty} A(2n + 1) = a_2$.
- b) Si $A(0) \in (c_2, d_2) \Rightarrow \lim_{n \rightarrow +\infty} A(2n) = a_2$ y $\lim_{n \rightarrow +\infty} A(2n + 1) = a_1$.

Es decir, si cuando $A(0)$ está próximo a a_1 se verifica que $A(n)$ tiende a a_1 cuando n es par y a a_2 cuando n es impar (y, de forma similar, cuando $A(0)$ está próximo a a_2). \square

Ejemplo 4.4.11 Si $r = 3'2$ se verifica que los números

$$a_1 = \frac{21 + \sqrt{21}}{32} \simeq 0'7995 \quad \text{y} \quad a_2 = \frac{21 - \sqrt{21}}{32} \simeq 0'5130$$



constituyen un ciclo de orden 2 para el sistema dinámico (4.73). En efecto, si $A(0) = \frac{21 + \sqrt{21}}{32}$ puede comprobarse que

$$A(1) = 3'2 \frac{21 + \sqrt{21}}{32} \left(1 - \frac{21 + \sqrt{21}}{32} \right) = \frac{21 - \sqrt{21}}{32} = a_2$$

y

$$A(2) = 3'2 \frac{21 - \sqrt{21}}{32} \left(1 - \frac{21 - \sqrt{21}}{32} \right) = \frac{21 + \sqrt{21}}{32} = a_1.$$

De esta forma, continuando el proceso, se obtiene que

$$A(2n) = \frac{21 + \sqrt{21}}{32} = a_1 \text{ y } A(2n + 1) = \frac{21 - \sqrt{21}}{32} = a_2$$

para $n = 0, 1, 2, \dots$, por lo que los puntos a_1 y a_2 constituyen un ciclo de orden 2, aparentemente estable (como veíamos en la Figura 4.16). Para estudiar la estabilidad de este ciclo expresaremos $A(n + 2)$ en términos de $A(n)$:

$$\begin{aligned} A(n + 2) &= f(f(A(n))) = 3'2 f(A(n))(1 - f(A(n))) \\ &= (3'2)^2 A(n)(1 - A(n))(1 - 3'2 A(n)(1 - A(n))) \\ &= \frac{256}{125} A(n)(1 - A(n))(16(A(n))^2 - 16A(n) + 5), \quad n = 0, 1, 2, \dots \end{aligned} \quad (4.74)$$

Nótese que, llamando

$$g(x) = f(f(x)) = \frac{256}{125} x(1 - x)(16x^2 - 16x + 5),$$

el sistema dinámico (4.74) puede escribirse en la forma

$$B(n + 1) = g(B(n)), \quad n = 0, 1, 2, \dots \quad (4.75)$$

donde $B(n) = A(2n)$.

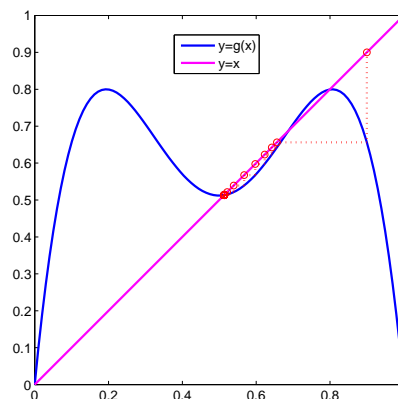


Figura 4.17: Gráfica del sistema dinámico (4.75) con $B(0) = 0'9$.

En la Figura 4.17 se muestra la representación gráfica del sistema dinámico (4.75) tomando como dato inicial $B(0) = A(0) = 0'9$. Los puntos de corte de la curva $y = g(x)$ con la recta $y = x$ son, precisamente, los valores

$$a_1 = \frac{21 + \sqrt{21}}{32} \simeq 0'7995 \text{ y } a_2 = \frac{21 - \sqrt{21}}{32} \simeq 0'5130$$

(que definen el ciclo de orden 2), junto con los otros dos puntos de equilibrio inestables del sistema (4.73) ($a = 0$ y $a = \frac{1}{16} = 0'6875$).



Para demostrar algebraicamente que a_1 y a_2 forman un ciclo estable de orden 2, vamos a estudiar separadamente la sucesión de términos de índice par y la sucesión de términos de índice impar, utilizando para ambas el resultado del Teorema 4.4.2. Para ello, calcularemos la derivada de $g(x) = f(f(x))$ en los valores a_1 y a_2 : si la derivada está comprendida entre -1 y 1 en ambos casos, se tendrá que la sucesión $\{A(n)\}_{n=0}^{\infty}$ tenderá a a_1 para n par y a a_2 para n impar (o viceversa). Por la *regla de la cadena* se tiene que

$$g'(x) = f'(f(x))f'(x)$$

siendo $f(x) = 3'2x - 3'2x^2$. Puesto que a_1 y a_2 forman un ciclo de orden 2 del sistema dinámico

$$A(n+1) = f(A(n)), \quad n = 0, 1, 2, \dots \quad (4.76)$$

se verifica que

$$a_2 = f(a_1) \text{ y } a_1 = f(a_2).$$

Por tanto, se tiene que

$$g'(a_1) = f'(f(a_1))f'(a_1) = f'(a_2)f'(a_1) = (3'2 - 6'4a_2)(3'2 - 6'4a_1) = \frac{4}{25} = 0'16 < 1,$$

por lo que a_1 es un punto de equilibrio estable para el sistema dinámico (4.75). Análogamente, puesto que

$$g'(a_2) = f'(f(a_2))f'(a_2) = f'(a_1)f'(a_2) = \frac{4}{25} = 0'16 < 1,$$

se tiene que a_2 también es un punto de equilibrio estable para el sistema dinámico (4.75). Por tanto, se tiene que a_1 y a_2 forman un ciclo estable de orden 2 del sistema dinámico (4.73). \square

La manera de proceder, en el caso general, es la siguiente:

- 1) Se hallan, si existen, dos números a_1 y a_2 que formen un ciclo de orden 2. Puesto que a_1 y a_2 satisfacen

$$a_2 = f(a_1) \text{ y } a_1 = f(a_2),$$

deben ser solución de la ecuación

$$a = g(a) = f(f(a)). \quad (4.77)$$

- 2) Se determina la estabilidad del ciclo de orden 2 obtenido, estudiando el tamaño de la derivada de la función g .

Obviamente, se tiene el siguiente resultado:

Teorema 4.4.12 Si un número es solución de la ecuación (4.77) entonces es un punto de equilibrio o uno de los valores de un ciclo de orden 2. \square

La noción general de estabilidad de un ciclo de orden n se recoge en la siguiente definición:

Definición 4.4.13 Un ciclo de orden k que toma valores a_1, a_2, \dots, a_k de un sistema dinámico de primer orden

$$A(n+1) = f(A(n))$$

es *estable*, *neutro* o *inestable* si alguno de los valores a_i es un punto de equilibrio estable, neutro o inestable, respectivamente, del sistema dinámico gobernado por la función $g = f \circ \dots \circ f$. \square

Observación 4.4.14 La definición anterior es consistente puesto que es fácil probar que todos los elementos de un ciclo tienen el mismo carácter en lo que a la estabilidad se refiere. \square

Para determinar la estabilidad del ciclo de orden 2, utilizaremos el siguiente resultado:

Teorema 4.4.15 Supongamos que el sistema dinámico (4.76) tiene un ciclo en orden 2 que toma los valores a_1 y a_2 . Si la función f tiene derivada continua en sendos entornos de a_1 y de a_2 , este ciclo de orden 2 es:



a) estable si $|f'(a_1)f'(a_2)| < 1$.

b) inestable si $|f'(a_1)f'(a_2)| > 1$.

DEMOSTRACIÓN. El ciclo de orden 2 del sistema dinámico (4.76) es estable si los puntos de equilibrio a_1 y a_2 son estables para el sistema dinámico asociado a $g = f \circ f$, es decir, si se verifica que

$$|g'(a_1)| < 1 \text{ y } |g'(a_2)| < 1.$$

Ahora bien, puesto que

$$g'(x) = f'(f(x))f'(x), \quad a_2 = f(a_1) \text{ y } a_1 = f(a_2),$$

se tiene que

$$\begin{cases} g'(a_1) = f'(f(a_1))f'(a_1) = f'(a_2)f'(a_1) \\ g'(a_2) = f'(f(a_2))f'(a_2) = f'(a_1)f'(a_2), \end{cases}$$

de donde se sigue el resultado. \square

Ejemplo 4.4.16 Utilicemos los Teoremas 4.4.12 y 4.4.15 para encontrar los valores de r en los que la ecuación logística tiene un ciclo de orden 2 estable. Puesto que

$$f(x) = rx(1-x),$$

se tiene que

$$\begin{aligned} a = f(f(a)) &\Leftrightarrow a = rf(a)(1-f(a)) = r^2a(1-a)(1-ra(1-a)) \\ &\Leftrightarrow a = (r^2a - r^2a^2)(1-ra+ra^2) \\ &\Leftrightarrow a = -r^3a^4 + 2r^3a^3 - r^2(r+1)a^2 + r^2a \\ &\Leftrightarrow -r^3a^4 + 2r^3a^3 - r^2(r+1)a^2 + (r^2-1)a = 0 \\ &\Leftrightarrow -a(ra-r+1)(r^2a^2 - r(r+1)a + r+1) = 0 \end{aligned}$$

(donde, en la última factorización se ha tenido en cuenta que $a = 0$ y $a = \frac{r-1}{r}$ son puntos de equilibrio y, por tanto, raíces de la ecuación $a = f(f(a))$). Así pues, los números que forman el ciclo de orden 2 son las raíces de la ecuación de segundo grado

$$r^2a^2 - r(r+1)a + r+1 = 0,$$

es decir,

$$\begin{aligned} a &= \frac{r(r+1) \pm \sqrt{r^2(r+1)^2 - 4r^2(r+1)}}{2r^2} \\ &= \frac{r(r+1) \pm \sqrt{r^2(r+1)(r+1-4)}}{2r^2} \\ &= \frac{r+1 \pm \sqrt{(r+1)(r-3)}}{2r} \end{aligned}$$

que, en el rango $r > 3$ que estamos estudiando, son dos raíces reales y distintas. Para analizar la estabilidad de este ciclo de orden dos, tengamos en cuenta que

$$f'(x) = r - 2rx$$

y apliquemos el Teorema 4.4.15. Puesto que

$$\begin{aligned} f'(a_1)f'(a_2) &= (r - 2ra_1)(r - 2ra_2) \\ &= \left(r - \left(r+1 + \sqrt{(r+1)(r-3)}\right)\right) \left(r - \left(r+1 - \sqrt{(r+1)(r-3)}\right)\right) \\ &= \left(-1 - \sqrt{(r+1)(r-3)}\right) \left(-1 + \sqrt{(r+1)(r-3)}\right) \\ &= 1 - (r+1)(r-3) = 1 - (r^2 - 2r - 3) = 4 + 2r - r^2, \end{aligned}$$



se tiene que

$$\begin{aligned} |f'(a_1)f'(a_2)| < 1 &\Leftrightarrow |4 + 2r - r^2| < 1 \Leftrightarrow -1 < 4 + 2r - r^2 < 1 \\ &\Leftrightarrow \left\{ \begin{array}{l} r^2 - 2r - 5 < 0 \Leftrightarrow r \in (1 - \sqrt{6}, 1 + \sqrt{6}) \\ r^2 - 2r - 3 > 0 \Leftrightarrow r \in (-\infty, -1) \cup (3, +\infty) \end{array} \right\} \\ &\Leftrightarrow r \in (1 - \sqrt{6}, -1) \cup (3, 1 + \sqrt{6}) \simeq (-1'4495, -1) \cup (3, 3'4495). \end{aligned}$$

Puesto que sólo nos interesan valores $r \in [0, 4]$, el Teorema 4.4.15 nos permite afirmar que el ciclo de orden 2 es estable si $3 < r < 1 + \sqrt{6}$ e inestable si $r \in (1 + \sqrt{6}, 4]$. \square

Acabamos de ver que para poblaciones con tasa de crecimiento grandes (mayores que 3), pero no demasiado grandes (menores que $1 + \sqrt{6} \simeq 3'4495$) el tamaño de la población formará una oscilación estable.

La pregunta que podemos formularnos ahora es ¿qué ocurre si $r > 1 + \sqrt{6}$? Sabemos que los dos puntos de equilibrio y el ciclo de orden 2 son todos inestables. En este caso, nos preguntamos si podría haber ciclos de orden 3, 4, ... Un ciclo de orden 3 consiste en tres números a_1, a_2 y a_3 tales que

$$a_2 = f(a_1), a_3 = f(a_2) \text{ y } a_1 = f(a_3).$$

Nótese que $a_i = f(f(f(a_i)))$ para $i = 1, 2, 3$. Para encontrar los ciclos de orden 3 habría que resolver la ecuación

$$a = f(f(f(a))).$$

Por ejemplo, en la ecuación logística, se tiene que $f(f(f(x)))$ es un polinomio de grado 8. Para encontrar un ciclo de orden 4 debemos factorizar un polinomio de grado 16 y, en general, para encontrar un ciclo de orden $k \geq 2$ debemos factorizar un polinomio de grado 2^k . Uno puede imaginarse que, en la práctica, resulte muy complicado (o, incluso, imposible) obtener ciclos de orden k de forma exacta.

Ejemplo 4.4.17 Consideremos la ecuación logística con $r = 3'5$, es decir,

$$A(n+1) = 3'5A(n)(1 - A(n)), \quad n = 0, 1, \dots \quad (4.78)$$

En la Figura 4.18 se muestra un ciclo de orden 4 formado por los números

$$\begin{cases} a_1 \simeq 0'8749972636024664, a_2 \simeq 0'3828196830173181, \\ a_3 \simeq 0'8269407065914337, a_4 \simeq 0'5008842103068593, \end{cases} \quad (4.79)$$

que parece ser estable (funciona bien para casi cualquier a_0 que se tome entre 0 y 1). \square

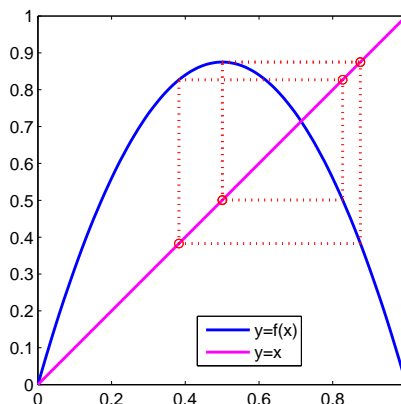


Figura 4.18: Gráfica del sistema dinámico (4.78) con $a_0 = 0'5$.



Ejemplo 4.4.18 Consideremos la ecuación logística con $r = 3'835$

$$A(n+1) = 3'835A(n)(1 - A(n)), \quad n = 0, 1, \dots \quad (4.80)$$

En la Figura 4.19 se muestra un ciclo de orden 3 formado por los números

$$a_1 \simeq 0'152074266627797, a_2 \simeq 0'494514368360185 \text{ y } a_3 = 0'958634596587540, \quad (4.81)$$

que parece ser estable (funciona bien para casi cualquier a_0 que se tome entre 0 y 1). \square

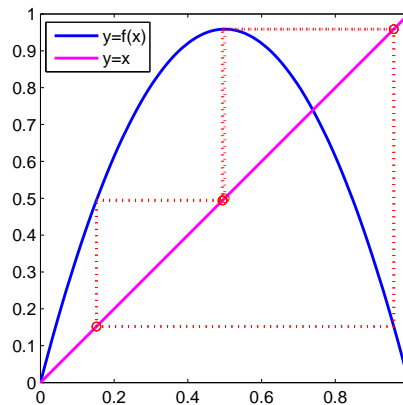


Figura 4.19: Gráfica del sistema dinámico (4.78) con $a_0 = 0'5$.

Para analizar la estabilidad de los ciclos de orden k , se puede utilizar la siguiente generalización del Teorema 4.4.19:

Teorema 4.4.19 Supongamos que el sistema dinámico (4.76) tiene un ciclo de orden $k \geq 2$ con valores a_1, a_2, \dots, a_k . Este ciclo de orden k es:

- a) estable si $|f'(a_1)f'(a_2) \cdots f'(a_k)| < 1$.
 b) inestable si $|f'(a_1)f'(a_2) \cdots f'(a_k)| > 1$. \square

Ejemplo 4.4.20 Estudiemos la estabilidad de los ciclos de los Ejemplos 4.4.17 y 4.4.18:

- a) La función f asociada al sistema dinámico (4.78) es

$$f(x) = 3'5x - 3'5x^2 \Rightarrow f'(x) = 3'5 - 7x.$$

Puesto que

$$|f'(a_1)f'(a_2)f'(a_3)f'(a_4)| \simeq |-0'0305| < 1$$

se verifica que el ciclo de cuarto orden (4.79) es estable.

- b) La función f asociada al sistema dinámico (4.80) es

$$f(x) = 3'835x - 3'835x^2 \Rightarrow f'(x) = 3'835 - 7'67x.$$

Puesto que

$$|f'(a_1)f'(a_2)f'(a_3)| \simeq |-0'3950| < 1$$

se verifica que el ciclo de tercer orden (4.81) es estable. \square



Como hemos visto, en la familia de sistemas dinámicos asociados a la ecuación logística con $0 \leq r \leq 4$, las soluciones que comienzan en casi cualquier punto del intervalo $(0, 1)$ (excepto los puntos de equilibrio y ciclos inestables) convergen a un punto de equilibrio si $r \in [0, 3]$ y a un ciclo de orden 2 si $r \in (3, 1 + \sqrt{6})$. Se puede demostrar que, sucesivamente, van apareciendo ciclos estables de órdenes 4, 8, 16, 32, ... según va creciendo r . Este fenómeno de duplicación del periodo diverge hacia infinito antes de llegar al valor $r = 4$, para pasar después a una sucesión de intervalos donde, en algunos de ellos, la solución “rellena” ciertos subintervalos de $[0, 1]$ o, en otros casos, converge a ciclos de órdenes variables como podían ser 3, 5, 6, ... y en los que se producen también ciertos fenómenos de duplicación del periodo. Este comportamiento en el que pequeñas variaciones de r dan como resultado cambios enormes en el comportamiento de la solución es el que hace que un sistema como éste reciba el nombre de *sistema dinámico caótico*. En el siguiente ejercicio se pueden visualizar algunos de estos comportamientos:

Ejercicio 4.4.21 Comprobar, utilizando el software `Sisdin` (ver [11]), que para $r = 0'5$ la sucesión $A(n)$ tiende a 0 cualquiera que sea el dato inicial $a_0 \in (0, 1)$. Cuando $r = 0'7$ ocurre lo mismo, pero para $r = 1'5$ los valores $A(n)$ se acercan a $\frac{1}{3}$. En cambio, cuando $r = 2$, se acercan a $0'5$.

A partir de $r > 3$, aparece una *bifurcación* (se pasa de un punto de equilibrio estable a un ciclo de orden dos estable) ya que, por ejemplo, con $r = 3'2$, se tiene un ciclo estable de orden 2 y los sucesivos términos $A(n)$ tienden, de manera oscilante, a los valores $0'5130456 \dots$ y $0'799456 \dots$. Al aumentar r , cada uno de estos límites se bifurca a su vez. Así, por ejemplo, para $r = 3'5$ aparece un ciclo estable de orden 4 dado por $0'38282 \dots, 0'82694 \dots, 0'50088 \dots$ y $0'87500 \dots$; para $r = 3'55$ se tiene un ciclo estable de orden 8.

Pero si seguimos aumentando r , la situación comienza a volverse caótica. A veces aparece un cierto orden: por ejemplo, el parámetro $r = 3'627$ proporciona un ciclo estable de orden 6, para $r = 3'702$ se tiene un ciclo estable de orden 7, con $r = 3'74$ tenemos un ciclo de orden 5 estable, mientras que el valor $r = 3'835$ da lugar a un ciclo estable de orden 3. Sin embargo, si tomamos como parámetro $r = 3'65$ la solución obtenida a partir de cualquier valor inicial (salvo el punto de equilibrio $a \simeq 0'72602739$) parece rellenar el conjunto $(0'2914, 0'6775) \cup (0'7537, 0'9125)$ \square

Los fenómenos anteriores se pueden visualizar muy bien mediante un gráfico en el que representaremos en el eje de abscisas los valores del parámetro $r \in [0, 4]$ y, para cada valor de r , los puntos $(r, A(n))$ correspondientes a puntos suficientemente avanzados de la solución que comienza en un punto a_0 , aleatoriamente elegido en el intervalo $(0, 1)$ para el sistema dinámico correspondiente a la ecuación logística (el punto a_0 puede ser el mismo para todos los valores del parámetro). Este gráfico se conoce con el nombre de *diagrama de Feigenbaum*, por estar íntimamente ligado a los trabajos del físico Mitchell Feigenbaum.

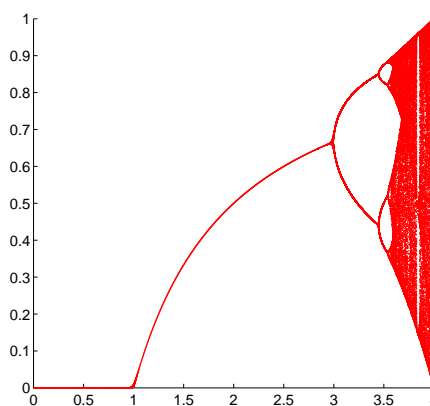


Figura 4.20: Diagrama de Feigenbaum para la ecuación logística.

En la Figura 4.20 se muestra el diagrama de Feigenbaum correspondiente a la ecuación logística para valores del parámetro $r \in [0, 4]$. Los puntos de corte de la gráfica obtenida con una recta vertical levantada sobre cada valor de r corresponden a los puntos límites de la solución que comienza en casi cualquier punto del intervalo $(0, 1)$. Más concretamente, si para un cierto valor $r = r_1$ sólo hay un punto de corte (r_1, a_1)

significa que a_1 es un punto de equilibrio estable de la ecuación logística con $r = r_1$; si para otro valor $r = r_2$ hay dos puntos de corte (r_2, a_1) y (r_2, a_2) , significa que los valores a_1 y a_2 constituyen un ciclo estable de orden 2 de la ecuación logística de parámetro $r = r_2$; y, así, sucesivamente. Para aquellos valores de r en los que la recta vertical corta a la gráfica en una nube de puntos de $(0, 1)$, lo que ocurre es que la solución, para casi cualquier valor inicial $a_0 \in (0, 1)$, de la ecuación logística con este parámetro r vaga incontroladamente, rellenando algún subconjunto de $(0, 1)$.

Recuérdese que para $r \in [0, 1]$ el punto $a = 0$ es el único punto de equilibrio estable; éste es el motivo de que el diagrama de Feigenbaum parezca no existir para estos valores del parámetro (en realidad, lo que ocurre es que todos los puntos ahí representados son de la forma $(r, 0)$ y se confunden con el eje de abscisas); por esta razón, el diagrama de Feigenbaum se suele representar en el intervalo $[1, 4]$.

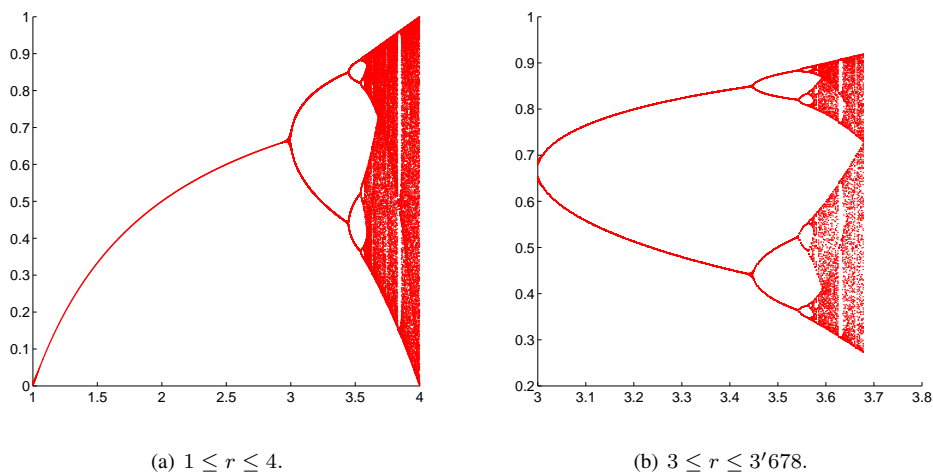


Figura 4.21: Diagrama de Feigenbaum para la ecuación logística.

El diagrama de Feigenbaum tiene un carácter autosemejante, puesto que dentro de él se pueden encontrar copias semejantes al conjunto total. Esto puede verse en la Figura 4.21, donde el segundo diagrama es una ampliación del primero. De hecho, puede demostrarse que este conjunto tiene estructura *fractal*.

4.5 Problemas

4.5.1 Encontrar los seis primeros términos de la sucesión

$$A(n+1) = 1 + 2A(n)$$

partiendo del valor inicial $A(0) = 5$.

4.5.2 Supongamos que una persona enferma toma, cada 8 horas, una pastilla que contiene 200 mg de un cierto principio activo (del cual suponemos que pasa directamente a la sangre y que el cuerpo elimina, cada 8 horas, un 20% de la cantidad que de él haya en el torrente sanguíneo). Determinar la cantidad que habrá de dicho principio en la sangre del enfermo tras 2 días de tratamiento.

4.5.3 Determinar cuáles de los siguientes sistemas dinámicos de primer orden son lineales de coeficientes constantes homogéneos y cuáles son afines:

a) $A(n+3) = \frac{2}{A(n+2)}$

b) $A(k) = A(k-1) + 1$

c) $A(n+1) = (A(n))^2 + 3$



d) $A(n+4) = -A(n+3)$

e) $A(n+1) = -3A(n) + n.$

4.5.4 Encontrar los eventuales puntos de equilibrio de los siguientes sistemas dinámicos y clasificarlos en estables, inestables ó neutros:

a) $A(n+1) = 2A(n) + 6$

b) $A(n+1) = -\frac{A(n)}{2}$

c) $A(n+1) = -4A(n) + 7$

d) $A(n+1) = A(n) + 2$

e) $A(n+1) = 0'4A(n) - 1.$

4.5.5 Determinar el valor de b para que el sistema dinámico

$$A(n+1) = -0'3A(n) + b$$

tenga como punto de equilibrio $a = 5$.

4.5.6 Determinar el valor de r para que el sistema dinámico

$$A(n+1) = rA(n) + 2r$$

tenga como punto de equilibrio $a = 2$.

4.5.7 Esbozar la representación gráfica de los siguientes sistemas dinámicos y deducir, a partir de ella, la estabilidad de los correspondientes puntos de equilibrio:

a) $A(n+1) = -1'5A(n) + 5$ con $A(0) = 3$

b) $A(n+1) = 0'5A(n) - 5$ con $A(0) = 1$

c) $A(n+1) = -A(n) + 3$ con $A(0) = 1$

d) $A(n+1) = 1'5A(n) - 5$ con $A(0) = 9$

e) $A(n+1) = A(n) + 3$ con $A(0) = 0.$

4.5.8 Encontrar la solución general y los eventuales puntos de equilibrio de los siguientes sistemas dinámicos:

a) $A(n+1) = 3A(n)$

b) $A(n+1) = -0'5A(n)$

c) $A(n+1) = A(n) + 3$

d) $A(n+1) = 1'5A(n) - 5$

e) $3A(n+1) + 1 = 2A(n) - 4$

f) $2A(n+1) - 3A(n) = 8.$

4.5.9 Encontrar la solución particular de los sistemas dinámicos del Problema 4.5.8 cuando el dato inicial es $A(0) = 4$.

4.5.10 Se considera el sistema dinámico

$$A(n+1) = 3A(n) - 4.$$



a) Si $A(0) = 3$, ¿para qué valor de k se verifica que $A(k) = 245$?

b) Si $A(6) = -1456$, ¿cuánto vale $A(0)$?

4.5.11 Se considera el sistema dinámico lineal

$$A(n+1) = 3A(n) + \lambda n + 5, \quad n = 0, 1, 2, \dots \quad \text{con } \lambda \in \mathbb{R}. \quad (4.82)$$

a) Determinar la solución general del sistema homogéneo asociado a (4.82).

b) Estudiar para qué valores de λ el sistema (4.82) tiene puntos de equilibrio y determinarlos. Estudiar la estabilidad de los mismos.

c) En el caso concreto en que $\lambda = 2$, es decir, para el sistema dinámico $A(n+1) = 3A(n) + 2n + 5$, $n = 0, 1, 2, \dots$

i) Encontrar una solución particular de la forma $\alpha n + \beta$ con $\alpha, \beta \in \mathbb{R}$.

ii) Escribir la solución general.

iii) Encontrar la solución que verifica $A(1) = 2$.

4.5.12 Supongamos que un banco paga un 8% de interés anual, compuesto trimestralmente, y hacemos un depósito de 1000 €.

a) Modelizar este problema mediante un sistema dinámico que describa la cantidad de dinero en la cuenta en un trimestre en función de la cantidad que había en el trimestre anterior.

b) ¿Cuánto dinero habrá al cabo de un año?

c) ¿Cuánto dinero habrá al cabo de n trimestres? ¿Y al cabo de k años?

d) Generalizar la situación anterior para un interés anual I , compuesto en m periodos durante un año, para un depósito inicial de a_0 euros.

4.5.13 Se abre con 10000 € una cuenta que proporciona un 4% de interés nominal anual que se capitaliza mensualmente. ¿Qué cantidad debe aportarse cada mes para que al cabo de 10 años se haya doblado la cantidad invertida? Con dicha aportación mensual ¿cuántos años deben transcurrir (desde la apertura de la cuenta) para que el capital inicial se triplique?

4.5.14 Un producto financiero que capitaliza trimestralmente tiene una tasa anual equivalente (TAE) del 4%. ¿Cuál es su tipo de interés nominal?

4.5.15 Calcular la T.A.E. de un depósito bancario a 6 meses que ofrece un T.I.N. del 3% con capitalización trimestral de los intereses.

4.5.16 Un banco ofrece una cuenta que capitaliza mensualmente y que tiene una T.A.E. del 6%. Si se ingresan inicialmente 1000 €, ¿cuánto dinero habrá en la cuenta al cabo de 5 meses?

4.5.17 Calcular la T.A.E. de un depósito que ofrece como interés nominal el 1'50% el primer año, 2'25% el segundo año y 3'00% el tercero, si se supone que los intereses se capitalizan anualmente. Ídem para el caso de que lo hagan semestralmente. Ídem para el caso de que lo hagan mensualmente.

4.5.18 Supongamos que cualquiera puede abrir una cuenta de un banco al 4% T.A.E. durante 5 años (es decir, el mercado asegura esa rentabilidad durante 5 años). Por otra parte, supongamos que hace algún tiempo compramos un producto financiero que asegura a su propietario, dentro de cinco años, 10000 €. ¿Cuál sería la cantidad mínima por la que deberíamos vender este producto si alguien estuviera interesado en comprárnoslo? (Esta cantidad en la que habría que valorar ese producto a día de hoy se suele denominar *valor actual*).

4.5.19 ¿Cuál es el valor actual de un producto financiero cuyo valor dentro de 5 años será de 100000 €, si se supone una inflación del 4% anual?



4.5.20 Calcular cuánto dinero podemos pedir prestado al 4% a 20 años, si sabemos que podemos pagar un máximo de 1000 € al mes.

4.5.21 ¿A cuántos años debe pedirse una hipoteca de 200000 €, si se quiere pagar una cuota mensual de 1000 €, cuando el T.I.N. es del 4%?

4.5.22 Un banco ofrece un depósito con vencimiento a 12 meses, pago de intereses trimestral y tipo de interés nominal 4%. El depósito admite ser cancelado de forma anticipada con una penalización que consiste en restar a los intereses producidos hasta ese momento el interés que produciría el capital inicial al 2% T.I.N. durante el periodo que va desde la fecha de cancelación hasta el vencimiento.

- Calcular la T.A.E. de este depósito ¿Cuál es el capital obtenido, transcurridos los 12 meses, si se han depositado 1000 €?
- ¿Qué capital se obtiene si se depositan 1000 € y el depósito se cancela transcurridos 9 meses? ¿Cuál es la T.A.E. de la rentabilidad obtenida?

4.5.23 Necesitamos pedir una hipoteca de 200000 € y podemos pagar 1000 € al mes. Si el banco nos ofrece un tipo de interés nominal anual del 3%, ¿a cuántos años debemos pedir el préstamo?

4.5.24 Si nos ofrecen un producto que proporciona un interés nominal del 2'5% capitalizado semestralmente, durante 15 años, ¿qué cantidad debemos invertir si queremos tener, al cabo de los 15 años, 20000 €?

4.5.25 Supongamos que se solicita un crédito hipotecario de 300000 € a un tipo de interés nominal del 4%.

- ¿Cuál sería la cuota mensual si el periodo de amortización son 20, 30 o 40 años? Concluir la influencia del aumento de la vida del crédito en el tamaño de la cuota.
- Calcular, en los tres casos, cuánto se paga de intereses y cuánto de capital en los primeros meses.

4.5.26 La ONCE ofrece el *Cupón fin de semana*, cuyo premio es un sueldo de 6000 € al mes durante 25 años. ¿Cuál es el valor actual de dicho premio si suponemos una inflación del 4% (es decir, qué capital habría que invertir al 4% nominal para obtener, durante 25 años, esa rentabilidad mensual)?

4.5.27 Si se invierten 10000 € en una cuenta que proporciona el 4'8% de interés nominal, con capitalizaciones mensuales, ¿cuántos años deben pasar para que el saldo de dicha cuenta sea de 25000 €? Ídem si cada mes se aportan 100 €.

4.5.28 Ante la reforma del sistema de pensiones, una persona de 40 años comienza a ingresar, en una cuenta que capitaliza mensualmente y proporciona un 6% T.I.N., una aportación fija mensual. Lo hace así durante 30 años, con la intención de poder sacar de esta cuenta, a partir de cumplir 70 años, 10000 € al mes durante 20 años. ¿Cuál debe ser el valor de la aportación mensual?

4.5.29 A una persona de 38 años le tocan 2 millones de euros en la lotería. Decide asegurarse su jubilación y, para ello, invierte en una cuenta bancaria a un interés nominal del 4% con capitalización mensual. Cuando tenga 65 años comenzará a retirar, cada mes, 30000 €. ¿Durante cuánto tiempo podrá hacerlo? ¿Cuál es el valor actual de los primeros 30000 € que cobra si se supone una inflación anual del 5%?

4.5.30 Un estudiante recibe una beca-préstamo de 50000 € al iniciar sus estudios, dinero que debe comenzar a devolver al cabo de 4 años.

- En el momento de recibir el dinero, lo ingresa en una cuenta que proporciona un 2% de interés nominal, capitalizado mensualmente. Para mantenerse mientras estudia, saca de dicha cuenta una cantidad de b euros al mes. Sabiendo que al concluir los 4 años de carrera el saldo de la cuenta es de 2000 €, determinar el valor de b .
- Calcular la tasa anual equivalente (TAE) de la cuenta anterior.



- c) En el momento de acabar, el estudiante devuelve los 2000 € sobrantes y comienza a amortizar el préstamo, con un tipo de interés nominal del 3% y una cuota mensual de 1000 €. ¿Cuánto tiempo tardará en cancelar el préstamo?

4.5.31 Un especulador quiere aprovechar los distintos tipos de interés en los países de la eurozona para enriquecerse, pidiendo prestado en un país 100000 € e invirtiéndolos en una cuenta en otro país.

- a) La cuenta en la que deposita los 100000 € proporciona un 4'5% de interés nominal anual y capitaliza mensualmente. ¿Qué cantidad debe retirar cada mes para que la cuenta quede a cero transcurridos 10 años?
- b) El préstamo de 100000 € tiene un interés nominal anual del 4%. El especulador pretende que la cuota del préstamo coincida con la cantidad retirada del depósito del apartado anterior. ¿A cuántos años debe pedir el crédito para que esto ocurra?
- c) Comprobar que con esta operación el especulador termina ganando más de 3000 €.
- d) Calcular la tasa anual equivalente (T.A.E.) de la cuenta del apartado a).

4.5.32 Supongamos que tenemos una muestra de 15 g de material radiactivo y que, al cabo de un año, sólo quedan 14'7 g activos. ¿Cuál es la semivida de este material?

4.5.33 ¿Qué porcentaje del carbono 14 original tiene un hueso cuya antigüedad se ha datado en, aproximadamente, 3000 años?

4.5.34 Si la semivida de un cierto material radiactivo es 100 años y en la actualidad tenemos 1 g de dicho material, ¿qué cantidad de dicho material permanecerá activa dentro de un año?

4.5.35 En una excavación arqueológica se descubre el mango de madera de un hacha que tiene un 30% de su cantidad original de carbono 14. Datar, de forma aproximada, dicha herramienta.

4.5.36 Encontrar la solución general de los siguientes sistemas dinámicos lineales homogéneos de orden dos:

- a) $A(n+2) = A(n+1) + 2A(n)$
- b) $A(n+2) = -A(n+1) - A(n)$
- c) $A(n+2) = -2A(n+1) - A(n)$
- d) $A(n+2) = 2A(n+1) + A(n)$
- e) $A(n+2) = 4A(n+1) - 3A(n)$
- f) $A(n+2) = 6A(n+1) - 9A(n)$.

4.5.37 Hallar la solución particular de los sistemas del Problema 4.5.36 cuando los datos iniciales son $A(0) = 0$ y $A(1) = 1$.

4.5.38 Encontrar la solución general de los siguientes sistemas dinámicos lineales afines de orden dos:

- a) $A(n+2) = A(n+1) + 2A(n) - 2$
- b) $A(n+2) = -A(n+1) - A(n) + 6$
- c) $A(n+2) = -2A(n+1) - A(n) + 12$
- d) $A(n+2) = 2A(n+1) + A(n) - 8$
- e) $A(n+2) = 4A(n+1) - 3A(n) - 10$
- f) $A(n+2) = 6A(n+1) - 9A(n) + 24$



g) $A(n+2) = 2A(n+1) - A(n) + 14$.

4.5.39 Hallar la solución particular de los sistemas del Problema 4.5.38 cuando los datos iniciales son $A(0) = 0$ y $A(1) = 1$.

4.5.40 Se considera el sistema dinámico

$$A(n+2) = aA(n+1) + A(n) + 1 \text{ siendo } a \in \mathbb{R}.$$

a) Determinar la solución general del sistema dinámico anterior en función de los valores del parámetro a .

b) Si $a = 0$, hallar la solución particular que verifica $A(1) = \frac{1}{2}$ y $A(2) = 2$.

4.5.41 Determinar, en función del parámetro real λ , la expresión de la solución general del sistema dinámico

$$\lambda A(n+2) = 2A(n+1) - A(n) + 1, \quad n = 0, 1, 2, \dots$$

4.5.42 Supongamos un juego que paga uno a uno en el que la probabilidad de ganar es $0'49$.

a) Si se comienza con 50€ y se tiene intención de retirarse cuando se doble esta cantidad, ¿cuál es la probabilidad de arruinarse si se hacen apuestas de 1€ ? ¿Y si las apuestas son de 5€ ?

b) Si se mantiene la idea de retirarse cuando se tengan 100€ , ¿con qué cantidad debería comenzarse para tener, al menos, un 50% de probabilidad de obtener los 100€ ? Ídem si la probabilidad de ganar en cada apuesta es $0'48$.

4.5.43 Hacer un estudio análogo al realizado con la ruleta francesa ($(n, N) = (30, 60)$ y $(n, N) = (100, 200)$) para la ruleta americana, en la que además del 0 , existe una casilla marcada con 00 , el cual no es ni rojo ni negro, ni par ni impar.

4.5.44 Encontrar los puntos de equilibrio de los siguientes sistemas dinámicos y clasificarlos:

a) $A(n+1) = 1'7A(n) - 0'14A^2(n)$

b) $A(n+1) = 0'8A(n) + 0'1A^2(n)$

c) $A(n+1) = 3A(n) - A^2(n) + 3$

d) $A(n+1) = 0'2A(n) + 0'2A^3(n)$

e) $A(n+1) = 1'4A(n) - 0'2A^2(n) + 3$

f) $A(n+1) = \frac{\pi}{2} \text{sen}(A(n))$.

4.5.45 Estudiar la estabilidad de los puntos de equilibrio del sistema dinámico $A(n+1) = \frac{1}{A(n)+2}$.

4.5.46 Determinar los puntos de equilibrio del sistema dinámico

$$A(n+1) = A(n) + \text{sen}(A(n)), \quad n = 0, 1, 2, \dots$$

y estudiar la estabilidad de los mismos.

4.5.47 Determinar el valor de c para que 2 sea punto de equilibrio del sistema $A(n+1) = cA^2(n) + c$. ¿Es estable dicho punto de equilibrio?

4.5.48 Determinar el valor de c para que $\frac{\pi}{2}$ sea punto de equilibrio del sistema $A(n+1) = c \text{sen}^2(A(n)) + c$. ¿Es estable dicho punto de equilibrio?



4.5.49 Determinar los valores de a y b para que el sistema dinámico $A(n+1) = aA^3(n) + b$ tenga a 0 y 1 como puntos de equilibrio. ¿Son estables?

4.5.50 Se consideran los sistemas dinámicos:

a) $A(n+1) = 5A(n) - A^2(n) - b$

b) $A(n+1) = -bA(n) + A^3(n)$.

Encontrar, en función del parámetro b , los puntos de equilibrio de dichos sistemas y estudiar su estabilidad. Utilizar la representación gráfica para intuir la estabilidad de los puntos de equilibrio en los casos dudosos.

4.5.51 Para cada $\lambda \geq 0$ se considera el sistema dinámico

$$A(n+1) = \lambda A(n)^2(2\lambda A(n) + 1).$$

Encontrar sus puntos de equilibrio, en función de λ , y determinar la estabilidad de los mismos.

4.5.52 Se considera el sistema dinámico no lineal

$$\begin{cases} A(0) = a_0 \in \mathbb{R} \\ A(n+1) = \frac{A(n)}{\sqrt{2 - A^2(n)}}, n = 0, 1, 2, \dots \end{cases} \quad (4.83)$$

a) Determinar los puntos de equilibrio del sistema dinámico (4.83) así como la estabilidad de los mismos.

b) Para datos iniciales $a_0 \in (0, 1)$ se considera el cambio de variable

$$B(n) = \frac{1}{A^2(n)}.$$

i) Demostrar que $\{B(n)\}_{n=0}^{\infty}$ es solución de un sistema dinámico afín de la forma

$$\begin{cases} B(0) = b_0 \\ B(n+1) = rB(n) + b, n = 0, 1, 2, \dots \end{cases} \quad (4.84)$$

para ciertos valores de b_0, r y b .

ii) Encontrar la solución del sistema (4.84) para los valores de b_0, r y b determinados en el apartado anterior.

iii) Explicitar la solución del sistema dinámico (4.83) y comprobar que está bien definida para los datos iniciales considerados, es decir, cuando $a_0 \in (0, 1)$.

4.5.53 Se considera la ecuación logística para los valores del parámetro $r = 3'1$ y $r = 3'6$. Encontrar, en ambos casos, un ciclo de orden 2 y estudiar su estabilidad. Ratificar los resultados obtenidos dibujando la gráfica de los sistemas.





5 Teoría de grafos y aplicaciones

Hay muchas situaciones que pueden ser modeladas por medio de grafos: mapas de transporte, torneos de deporte, red de internet, redes sociales, algoritmos informáticos, . . . A lo largo de este capítulo se analizarán varios ejemplos importantes, siendo el colofón el que da lugar a la base del algoritmo *PageRank* que utiliza el buscador *Google*.

5.1 Definiciones. Lema del apretón de manos

Definición 5.1.1 Un *grafo* $G = (V, A)$ consiste en un conjunto no vacío V (cuyos elementos se denominan *vértices*) y una familia A de parejas no ordenadas de vértices; estas parejas de vértices se denominan *aristas* y diremos que cada arista une los dos vértices que la forman y que incide en cada uno de ellos. Estas parejas pueden aparecer repetidas en la familia A ; además, pueden tener los dos elementos iguales. Si una arista se repite se denomina *múltiple*; las aristas que unen un mismo vértice se denominan *bucles* (o *lazos*).

- a) Si en A no hay aristas múltiples ni bucles, se dice que G es un *grafo simple*.
- b) Si en A hay aristas múltiples, pero no bucles, diremos que G es un *multigrafo*.
- c) Si en A hay bucles (y, eventualmente, aristas múltiples) se dice que G es un *pseudografo*. □

Observación 5.1.2 Salvo que se diga lo contrario, supondremos que V y A son conjuntos finitos. □

Observación 5.1.3 Es importante hacer notar que en la literatura no siempre se utilizan estos mismos nombres y que muchas veces es difícil saber si un resultado se prueba para grafos simples, multigrafos o para pseudografos en general. □

Ejemplo 5.1.4

- a) Una liguilla entre los cuatro equipos de fútbol madrileños (a : Atlético de Madrid, g : Getafe, m : Real Madrid y r : Rayo Vallecano) en la que jueguen todos contra todos una vez, constituye un grafo si se consideran como aristas los partidos celebrados. El conjunto de vértices es $V = \{a, g, m, r\}$, siendo la familia de aristas $A = \{\{a, g\}, \{a, m\}, \{a, r\}, \{g, m\}, \{g, r\}, \{m, r\}\}$. Se trata de un grafo simple.
- b) En el conjunto $V = \{0, 1, 2, 3\}$ se considera la relación de congruencia módulo 2, y se toman como aristas la familia $A = \{\{0, 0\}, \{0, 2\}, \{1, 1\}, \{1, 3\}, \{2, 2\}, \{3, 3\}\}$, es decir, los pares de números que están relacionados mediante dicha relación. Puesto que existen bucles, se trata de un pseudografo. □

Observación 5.1.5 Un grafo suele visualizarse como un dibujo en el que cada vértice se representa mediante un punto y cada arista mediante una línea que une los puntos que representan sus dos vértices (véase la Figura 5.1). Es importante destacar que un mismo grafo se puede representar por dibujos muy distintos, lo que puede dar lugar a confusiones.

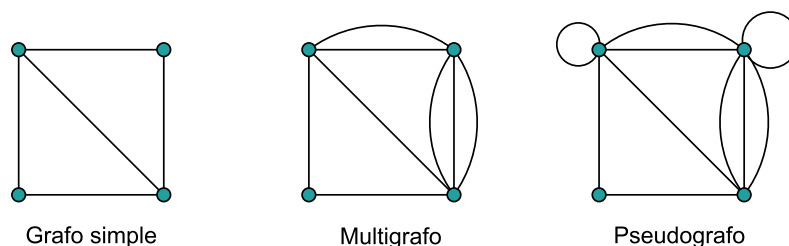


Figura 5.1: Tipos de grafos.



Ejemplo 5.1.6 En la Figura 5.2 encontramos representaciones de los siguientes grafos:

- a) Un torneo de cualquier deporte se puede modelizar mediante un grafo en el que los vértices son los equipos y hay una arista que une dos equipos si éstos se enfrentan. Por ejemplo, en un torneo de tipo liga, en el que cada equipo juega contra todos los demás, a una vuelta, se tendría un grafo simple.
- b) La red del Metro de Madrid constituye un grafo donde los vértices son las estaciones y hay una arista que une dos estaciones cuando son consecutivas en la misma línea. La existencia de estaciones como Argüelles y Moncloa (consecutivas en las líneas 3 y 6) hace que este grafo sea un multigrafo.
- c) Una relación de equivalencia da lugar a un grafo en el que los vértices son los elementos del conjunto y hay una arista entre ellos si están relacionados. Se trata de un pseudografo pues, a pesar de no tener aristas múltiples, tiene un bucle para cada vértice.

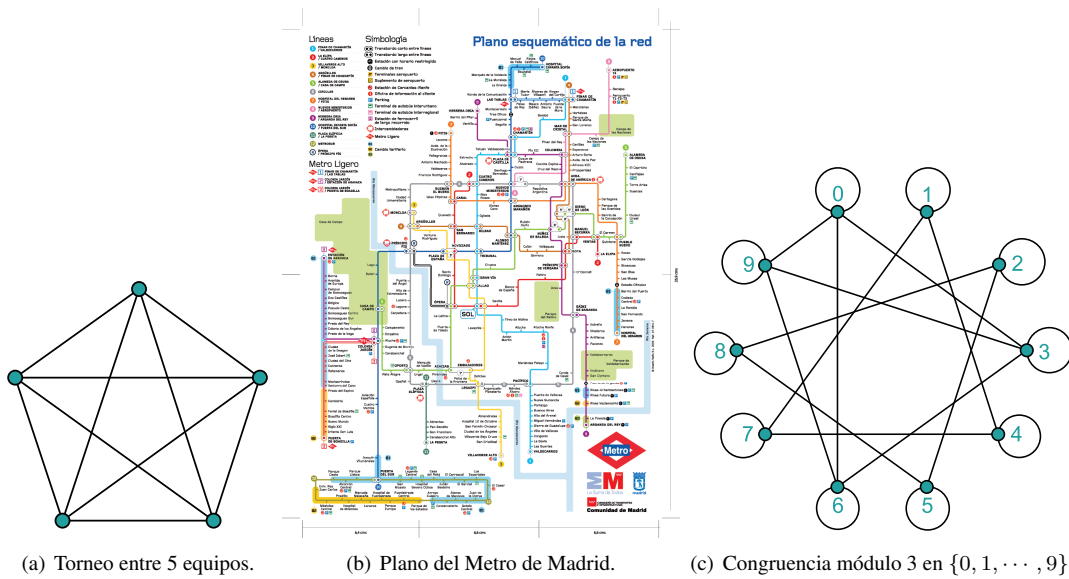


Figura 5.2: Grafos del Ejemplo 5.1.6.

La siguiente definición introduce el concepto de orientación en un grafo, que implica que cada arista se recorre en un sentido determinado:

Definición 5.1.7 Un grafo (V, A) se dice *orientado* (o *dirigido*) si las parejas de vértices de A que definen las aristas son pares ordenados. □

Observación 5.1.8 Intuitivamente, en los grafos orientados las aristas tienen un origen y un final. En este contexto, también se habla de grafos dirigidos simples, multigrafos dirigidos y pseudografos dirigidos. □

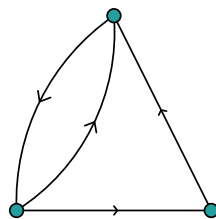


Figura 5.3: Grafo orientado.

Ejemplo 5.1.9 Los mensajes de correo electrónico enviados entre un grupo de colaboradores se puede modelizar mediante un multigrafo orientado: los vértices son cada uno de los colaboradores y hay una arista entre dos colaboradores cuando el primero ha enviado un mensaje al segundo. Si se admite que uno se escriba a sí mismo (por ejemplo, para tener una copia de seguridad de los mensajes), se trata de un pseudografo dirigido. □

A continuación damos la noción de camino; intuitivamente, una cadena de aristas consecutivas.

Definición 5.1.10 Sea $G = (V, A)$ un grafo no orientado y $v, v' \in V$. Llamaremos *camino de longitud k* entre v y v' (o que une v y v') a una secuencia de aristas a_1, a_2, \dots, a_k de forma que

$$a_1 = \{v, v_1\}, a_2 = \{v_1, v_2\}, \dots, a_{k-1} = \{v_{k-2}, v_{k-1}\}, a_k = \{v_{k-1}, v'\}.$$

Llamaremos *camino cerrado* (o *circuito*) a un camino en el que $v = v'$. Si todas las aristas son distintas, hablaremos de *camino simple*. \square

Observación 5.1.11 Nótese que, si el grafo es simple, todo camino queda determinado por la secuencia de los vértices (que pueden repetirse). Si no lo es, pueden existir caminos distintos que pasen por los mismos vértices, en el mismo orden. \square

Definición 5.1.12 En un grafo simple no orientado, se denomina *ciclo* (o *camino cíclico*) a todo camino simple cerrado en el que no hay vértices repetidos (salvo, obviamente, los extremos). Si un grafo no tiene ciclos, se denomina *árbol*. \square

Ejemplo 5.1.13 Consideremos el grafo de la Figura 5.4(a). Caminos de v_1 a v_3 :

- De longitud 1: $\{v_1, v_3\}$.
- De longitud 2: $\{v_1, v_2\}, \{v_2, v_3\}$ y $\{v_1, v_5\}, \{v_5, v_3\}$.
- De longitud 3: $\{v_1, v_5\}, \{v_5, v_4\}, \{v_4, v_3\}$, pero también $\{v_1, v_5\}, \{v_1, v_5\}, \{v_1, v_3\}, \dots$

Además, $\{v_5, v_3\}, \{v_3, v_2\}, \{v_2, v_1\}, \{v_1, v_3\}, \{v_3, v_5\}$ es un camino cerrado y $\{v_5, v_3\}, \{v_3, v_2\}, \{v_2, v_5\}$ un ciclo. En la Figura 5.4(b) se muestra un árbol. \square

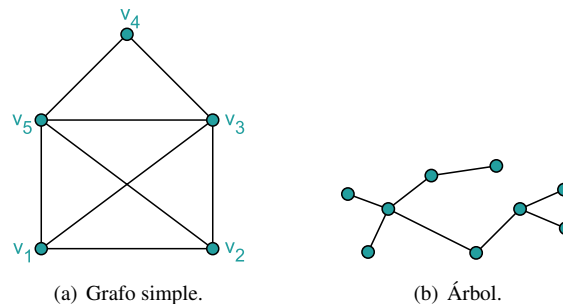


Figura 5.4: Grafos del Ejemplo 5.1.13.

Observación 5.1.14 Si el grafo está orientado, el camino que une v con v' debe seguir la orientación de las aristas. \square

A continuación se definen distintos tipos de grafos simples:

Definición 5.1.15 Diremos que un grafo simple no orientado es:

- Un *grafo completo* cuando cada vértice se une con el resto de los vértices.
- Un *grafo bipartito* si el conjunto de vértices se divide en dos conjuntos disjuntos de tal manera que las aristas únicamente unen vértices de un subconjunto con vértices del otro subconjunto. Diremos que es un *grafo bipartito completo* si cada vértice del primer subconjunto está unido mediante una arista con cada vértice del segundo.
- Un *ciclo* si consiste únicamente en un camino cíclico.
- Una *rueda* si se trata de un ciclo más un vértice adicional unido a cada uno de los demás mediante una arista. \square



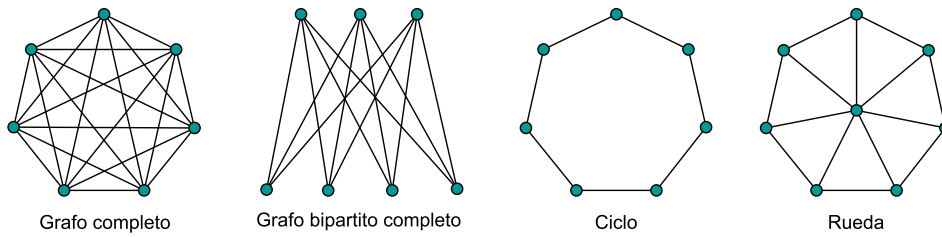


Figura 5.5: Algunos ejemplos de grafos simples.

Ejemplo 5.1.16

- a) Un torneo de tipo liga a una vuelta, en el que cada equipo juega contra todos los demás, es un grafo completo.
- b) Un grupo de amigos heterosexuales (que serán los vértices) en el que las aristas las forman parejas que han mantenido alguna relación entre ellos, constituye un grafo bipartito. Si se han dado todos los emparejamientos posibles, estamos ante un grafo bipartito completo.
- c) Una red local de ordenadores constituye un ciclo.
- d) Una red local de ordenadores junto con el servidor constituyen una rueda. □

A continuación, introducimos el concepto de conexión, que tiene que ver con la idea intuitiva de dividir un grafo en partes separadas entre sí:

Definición 5.1.17 Diremos que un grafo no orientado es *conexo* si cualquier par de vértices puede unirse mediante un camino (véase la Figura 5.6). □

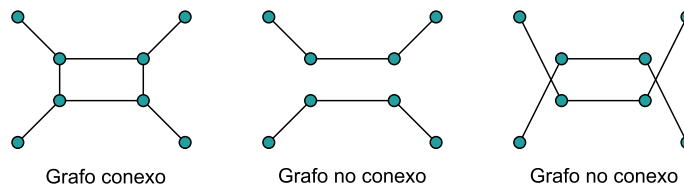


Figura 5.6: Conexión de grafos no orientados.

Para grafos orientados la noción de conexión es más sutil. Antes de establecer su definición, introducimos el siguiente concepto:

Definición 5.1.18 Llamaremos *grafo no orientado subyacente* a un grafo orientado dado, al grafo que resulta de suprimir en éste la dirección en las aristas, eliminando el orden entre los dos vértices que las componen (Figura 5.7). El grafo no orientado subyacente tiene el mismo número de aristas que el grafo orientado del que proviene. □

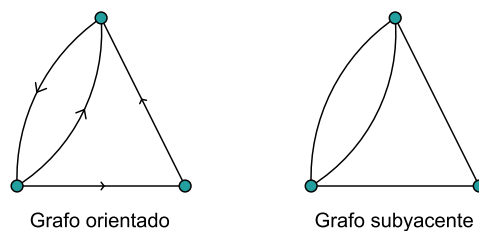


Figura 5.7: Grafo orientado y grafo no orientado subyacente.



Definición 5.1.19 Un grafo orientado $G = (V, A)$ se dice que es:

- Fuertemente conexo* si dados $v, v' \in V$ hay un camino que va de v a v' .
- Débilmente conexo* si el grafo no orientado subyacente es conexo. \square

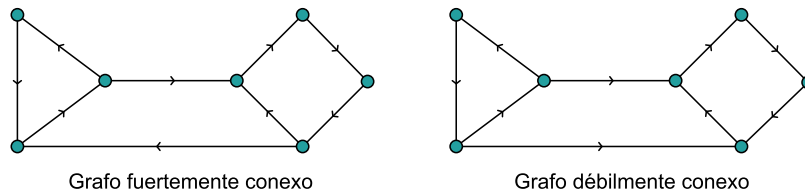


Figura 5.8: Conexión de grafos orientados.

Observación 5.1.20 Obviamente, todo grafo orientado fuertemente conexo es débilmente conexo. \square

Otras nociones importantes están relacionadas con la construcción de grafos nuevos a partir de grafos preexistentes.

Definición 5.1.21

- Un *subgrafo* de un grafo $G = (V, A)$ es un grafo $H = (W, F)$ que verifica $W \subset V$ y $F \subset A$.
- La *unión* de dos grafos $G_1 = (V_1, A_1)$ y $G_2 = (V_2, A_2)$ es un nuevo grafo $G = (V, A)$ con $V = V_1 \cup V_2$ y $A = A_1 \cup A_2$. \square

Un primer resultado relevante de la teoría de grafos es el lema que vamos a ver a continuación. Primero introducimos la siguiente noción:

Definición 5.1.22 El *grado de un vértice* $v \in V$ es el número de aristas $\text{grad}(v)$ que inciden en él.¹ En el caso de que de un vértice v no parta ninguna arista (es decir, $\text{grad}(v) = 0$), se dice que el vértice v está *aislado*. \square

Observación 5.1.23 Aunque por el momento no lo necesitamos, cuando el grafo está orientado se habla entonces de *grado saliente*, $S\text{grad}(v)$, cuando se cuentan el número de aristas que parten de $v \in V$, y de *grado entrante*, $E\text{grad}(v)$, para el número de aristas que terminan en v . \square

Teorema 5.1.24 (Lema del apretón de manos) Para todo grafo $G = (V, A)$ se verifica que

$$\sum_{v \in V} \text{grad}(v) = 2a,$$

siendo a el número de aristas.

DEMOSTRACIÓN. La demostración es sencilla. Como cada arista tiene dos extremos, al contar todos los grados estamos contando cada una de las aristas dos veces. \square

Corolario 5.1.25 La suma de los grados de los vértices de un grafo es par. En consecuencia, un grafo no puede tener un número impar de vértices de grado impar. \square

Ejemplo 5.1.26 Si se quieren conectar 9 ordenadores en red con 3 puertos cada uno aprovechando todos los puertos, obtendríamos un grafo imposible. Esto nos demuestra la imposibilidad de tal red. \square

¹Si de un vértice surge una arista que vuelve al mismo vértice, esta arista cuenta como 2 para el cómputo del grado de ese vértice.



5.2 Isomorfismo de grafos

Si se han construido grafos modelando situaciones, se puede ver que personas distintas habrán dado grafos aparentemente distintos pero que contienen la misma información: mismo número de vértices y mismo número de aristas que, a su vez, realizan las mismas conexiones entre los vértices:

Ejemplo 5.2.1 Consideremos un grupo de cinco personas: María López, que nació un primero de mayo; Raúl López, nacido un 12 de mayo; Raúl García, que cumple años el 13 de mayo; Teresa García, que nació un 13 de enero; y, finalmente, María García, nacida un primero de enero. Ordenamos las personas en el orden que se han enunciado (y las denominamos v_1, v_2, v_3, v_4 y v_5 , respectivamente) y consideramos el conjunto de vértices $V = \{v_1, v_2, v_3, v_4, v_5\}$. Vamos a definir dos grafos: en el primero, diremos que hay una arista entre dos vértices si las personas correspondientes tienen el mismo nombre o el mismo apellido; en el segundo, tendremos una arista si las personas correspondientes han nacido en el mismo mes o en el mismo día del mes. En el primer caso, la familia de aristas vendrá dada por

$$A_1 = \{\{v_1, v_2\}, \{v_1, v_5\}, \{v_2, v_3\}, \{v_3, v_4\}, \{v_3, v_5\}, \{v_4, v_5\}\}$$

mientras que, en el segundo caso, tendremos

$$A_2 = \{\{v_1, v_2\}, \{v_1, v_3\}, \{v_1, v_5\}, \{v_2, v_3\}, \{v_3, v_4\}, \{v_4, v_5\}\}.$$

Obviamente, estos dos grafos no son el mismo, aunque puedan representarse (Figura 5.9) mediante dibujos que son, salvo las etiquetas de los vértices, el mismo.

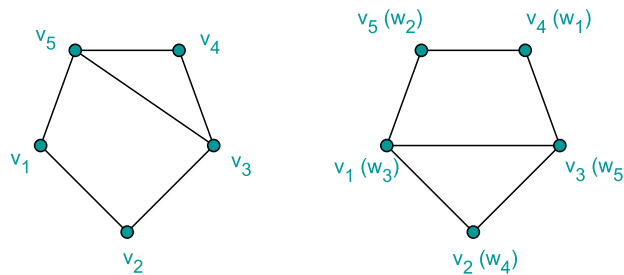


Figura 5.9: Ejemplo de grafos isomorfos.

Nótese que si reordenamos los vértices del segundo grafo de la siguiente forma

$$w_1 = v_4, w_2 = v_5, w_3 = v_1, w_4 = v_2, w_5 = v_3,$$

las aristas quedan

$$\begin{aligned} & \{\{w_3, w_4\}, \{w_3, w_5\}, \{w_2, w_3\}, \{w_4, w_5\}, \{w_1, w_5\}, \{w_1, w_2\}\} \\ & = \{\{w_1, w_2\}, \{w_1, w_5\}, \{w_2, w_3\}, \{w_3, w_4\}, \{w_3, w_5\}, \{w_4, w_5\}\}, \end{aligned} \tag{5.1}$$

esto es, las mismas que en el primer grafo. \square

Este ejemplo motiva que se introduzca la siguiente definición.

Definición 5.2.2 Un *isomorfismo* entre dos grafos $G_1 = (V_1, A_1)$ y $G_2 = (V_2, A_2)$ consiste en dos aplicaciones biyectivas

$$\phi_V : V_1 \rightarrow V_2 \text{ y } \phi_A : A_1 \rightarrow A_2$$

de forma que

$$a \in A_1 \text{ une } v, v' \in V_1 \text{ si, y sólo si, } \phi_A(a) \in A_2 \text{ une } \phi_V(v), \phi_V(v') \in V_2. \quad \square$$

Observación 5.2.3 La definición anterior es válida para grafos simples, multigrafos y pseudografos. Para el caso de un grafo simple, un isomorfismo queda determinado por una aplicación biyectiva $\phi : V_1 \rightarrow V_2$ de forma que $v, v' \in V_1$ están conectados por una arista si, y sólo si, $\phi(v), \phi(v') \in V_2$ están conectados por una arista. \square



Definición 5.2.4 Dos grafos se dice que son *isomorfos* si existe un isomorfismo entre ellos. El conjunto de todos los grafos isomorfos entre sí se denomina *clase de isomorfía*. □

Notación 5.2.5 A partir de ahora denotaremos por:

- a) K_n la clase de isomorfía del *grafo completo* de n vértices.
- b) $K_{m,n}$ la clase de isomorfía del *grafo bipartito completo* de m y n vértices, en donde todos los vértices del primer subconjunto de m vértices se conectan con todos los del segundo subconjunto de n vértices.
- c) C_n el ciclo de n vértices.
- d) R_n la rueda de $n + 1$ vértices. □

Observación 5.2.6 En la práctica, cuando se dice que se toma el grafo K_n , se quiere decir que se toma cualquiera de los elementos de su clase de isomorfía. Lo mismo con el resto de grafos incluido en la notación anterior. □

Ejemplo 5.2.7 Los grafos de la Figura 5.5 son, respectivamente, del tipo $K_7, K_{3,4}, C_7$ y R_7 . □

Observación 5.2.8 Dados dos grafos, puede suceder que existan varios isomorfismos entre ellos. El cómputo del número de isomorfismos entre dos grafos puede ser una tarea difícil aunque en algunos casos sea factible. Por ejemplo, es obvio que un grafo es isomorfo consigo mismo. Los isomorfismos de un grafo en sí mismo se denominan *automorfismos*. Es fácil ver que los automorfismos del grafo completo K_n son, precisamente, las permutaciones de n elementos (los n vértices), por lo que el número de automorfismos de K_n es $n!$ □

Observación 5.2.9 Hay varios criterios sencillos que pueden ayudar a determinar si dos grafos no son isomorfos:

- a) Si dos grafos no tienen el mismo número de vértices, no pueden ser isomorfos. Lo mismo ocurre con el número de aristas.
- b) Si dos grafos no tienen la misma cantidad de vértices de un mismo grado, no pueden ser isomorfos.
- c) La conexión se preserva bajo isomorfismo de grafos. Por tanto, si un grafo es conexo y el otro no, no pueden ser isomorfos.
- d) La existencia de ciclos de una longitud dada se preserva bajo isomorfismo de grafos. Así pues, si un grafo tiene un ciclo de una longitud y el otro no, no pueden ser isomorfos.

No obstante, que dos grafos tengan el mismo número de vértices y de aristas, la misma cantidad de vértices de un mismo grado, la misma cantidad de componentes conexas y la misma cantidad de ciclos de igual longitud, no demuestra que sean isomorfos. Véase, por ejemplo, la Figura 5.10, en la que ambos grafos representados tienen 8 vértices (4 de grado 1, 2 de grado 2 y 2 de grado 3), tienen 7 aristas, son conexos y no tienen ciclos. Sin embargo, no son isomorfos, puesto que en el de la izquierda los 4 vértices de grado 1 son adyacentes a vértices de grado 3, mientras que en el de la derecha uno de ellos es adyacente a uno de grado 2.

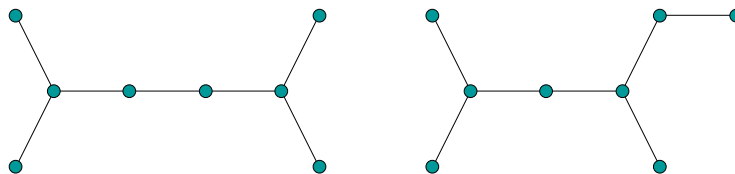


Figura 5.10: Grafos no isomorfos.

En cualquier caso, para probar que dos grafos son isomorfos hace falta encontrar un isomorfismo entre ellos. Así, por ejemplo, en los grafos de la Figura 5.11 el isomorfismo es el que a cada vértice v_i le hace corresponder el vértice w_i , para $i = 1, 2, 3, 4, 5, 6$. □



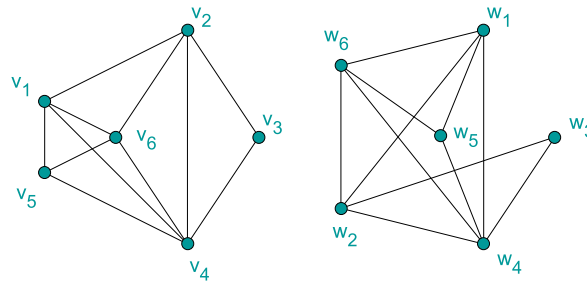


Figura 5.11: Isomorfismo entre dos grafos.

5.3 Matriz de adyacencia

Existen múltiples maneras útiles de representar los grafos. La representación del mismo puede ser esencial a la hora de poder determinar si dos grafos dados son isomorfos. Una de estas maneras es por medio de una matriz, como se explica a continuación.

Definición 5.3.1 Sea $G = (V, A)$ un grafo no orientado en el que se supone que los vértices están ordenados $\{v_1, v_2, \dots, v_n\}$. La *matriz de adyacencia* del grafo G respecto de ese orden es una matriz $M \in \mathcal{M}_{n \times n}(\mathbb{N})$ con $M = (m_{ij})_{i,j=1}^n$, siendo

$$m_{ij} = \text{número de aristas que conectan } v_i \text{ con } v_j.$$

En la notación anterior m_{ij} denota el elemento de la i -ésima fila y la j -ésima columna de la matriz M . \square

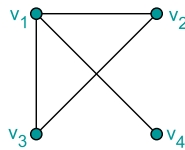


Figura 5.12: Grafo simple.

Ejemplo 5.3.2 El grafo de la Figura 5.12 tiene como matriz de adyacencia, cuando se consideran los vértices ordenados (v_1, v_2, v_3, v_4) , la matriz

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}.$$

Sin embargo, la matriz

$$\begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

es la matriz de adyacencia del mismo grafo cuando se considera como orden de los vértices (v_2, v_4, v_1, v_3) . \square

Ejercicio 5.3.3 Dibujar un grafo no orientado cuya matriz de adyacencia sea

$$\begin{pmatrix} 1 & 3 & 0 & 0 \\ 3 & 0 & 1 & 1 \\ 0 & 1 & 2 & 2 \\ 0 & 1 & 2 & 0 \end{pmatrix}. \quad \square$$



Observación 5.3.4 (Propiedades)

a) En un grafo simple se tiene

$$\begin{cases} m_{ij} \in \{0, 1\} & \text{para todo } i, j \\ m_{ii} = 0 & \text{para todo } i. \end{cases}$$

b) La matriz de adyacencia es simétrica, es decir, $m_{ij} = m_{ji}$.

c) Se puede definir la matriz de adyacencia de un grafo orientado como

$$m_{ij} = \text{número de aristas que van de } v_i \text{ a } v_j.$$

En ese caso, la matriz M ya no tiene por qué ser simétrica.

d) Como los n vértices se pueden ordenar de $n!$ maneras, puede haber en principio $n!$ matrices de adyacencia distintas para un mismo grafo. Sin embargo puede haber grafos en los que órdenes distintos produzcan la misma matriz. Por ejemplo, el grafo completo de n vértices produce siempre la matriz

$$\begin{pmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 0 \end{pmatrix}. \quad \square$$

Respecto al cambio de orden, existe la noción de matriz cambio de ordenación con la que se puede pasar de una matriz de adyacencia a otra. Pensemos por un momento en que en un grafo se intercambia el orden de dos vértices. ¿Cómo afectaría esto a la matriz de adyacencia? Obviamente, si hemos cambiado el vértice v_i con el vértice v_j en la matriz quedarán intercambiadas la columna i con la columna j y la fila i con la fila j . Esto equivale a multiplicar la matriz de adyacencia, por la derecha y por la izquierda, respectivamente, por una matriz P^{ij} , la cual es el resultado de intercambiar las filas (o las columnas) i y j en la matriz identidad. Esta matriz P^{ij} es simétrica.

Supongamos que tenemos una ordenación \mathcal{O}_1 de los vértices de un grafo, $\{v_1, v_2, \dots, v_n\}$, y consideremos otra ordenación de los vértices \mathcal{O}_2 , que se obtendrá a partir de \mathcal{O}_1 mediante una permutación σ ; es decir, la nueva ordenación será $\{v_{\sigma(1)}, v_{\sigma(2)}, \dots, v_{\sigma(n)}\}$. Llamemos $M_{\mathcal{O}_1} = (m_{ij})_{i,j=1}^n$ a la matriz de adyacencia correspondiente a la ordenación \mathcal{O}_1 y $M_{\mathcal{O}_2}$ a la correspondiente a la ordenación \mathcal{O}_2 . Por la propia definición de matriz de adyacencia, es claro que

$$(M_{\mathcal{O}_2})_{ij} = m_{\sigma(i)\sigma(j)}.$$

Esto es, las columnas y las filas de la nueva matriz de adyacencia serán las de la original, permutadas según la permutación σ . Por lo tanto, para construir $M_{\mathcal{O}_2}$ basta considerar la matriz resultado de permutar las columnas de la identidad según σ , es decir,

$$C_{\mathcal{O}_2\mathcal{O}_1} = (e_{\sigma(1)}, e_{\sigma(2)}, \dots, e_{\sigma(n)}),$$

donde e_i es el i -ésimo vector columna de la base canónica. A partir de ella tenemos:

$$M_{\mathcal{O}_2} = (C_{\mathcal{O}_2\mathcal{O}_1})^T M_{\mathcal{O}_1} C_{\mathcal{O}_2\mathcal{O}_1}.$$

La matriz $C_{\mathcal{O}_2\mathcal{O}_1}$ se denomina *matriz de cambio de ordenación* (de la ordenación \mathcal{O}_1 a la ordenación \mathcal{O}_2 en este caso). Lo que hemos probado es el siguiente resultado:

Proposición 5.3.5 Si \mathcal{O}_1 y \mathcal{O}_2 son dos ordenaciones de un grafo $G = (V, A)$, $M_{\mathcal{O}_1}$ y $M_{\mathcal{O}_2}$ son las matrices de adyacencia de G respecto de los órdenes \mathcal{O}_1 y \mathcal{O}_2 , respectivamente, y $C_{\mathcal{O}_2\mathcal{O}_1}$ es la matriz de cambio de ordenación de la primera a la segunda, entonces

$$M_{\mathcal{O}_2} = (C_{\mathcal{O}_2\mathcal{O}_1})^T M_{\mathcal{O}_1} C_{\mathcal{O}_2\mathcal{O}_1}. \quad \square \tag{5.2}$$



Ejemplo 5.3.6 Es claro que $\sigma = \{2, 4, 1, 3\}$ es la permutación que lleva la primera ordenación del Ejemplo 5.3.2 a la segunda. Por tanto, la matriz de cambio de ordenación es

$$C_{\mathcal{O}_2\mathcal{O}_1} = (e_2, e_4, e_1, e_3) = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}.$$

De esta forma, la matriz de adyacencia en la segunda ordenación se obtiene, teniendo en cuenta (5.2), como

$$\begin{aligned} (C_{\mathcal{O}_2\mathcal{O}_1})^T M_{\mathcal{O}_1} C_{\mathcal{O}_2\mathcal{O}_1} &= \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}^T \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} = M_{\mathcal{O}_2}. \quad \square \end{aligned}$$

Observación 5.3.7 Puesto que

$$(C_{\mathcal{O}_2\mathcal{O}_1})^T C_{\mathcal{O}_2\mathcal{O}_1} = \begin{pmatrix} e_{\sigma(1)}^T \\ e_{\sigma(2)}^T \\ \dots \\ e_{\sigma(n)}^T \end{pmatrix} (e_{\sigma(1)}, e_{\sigma(2)}, \dots, e_{\sigma(n)}) = I_n,$$

es obvio que

$$(C_{\mathcal{O}_2\mathcal{O}_1})^T = (C_{\mathcal{O}_2\mathcal{O}_1})^{-1}. \quad (5.3)$$

Así pues, la relación (5.2) puede interpretarse como un cambio de base, en concreto, el paso de una base dada a la que resulta al reordenar los vectores de dicha base de acuerdo con el nuevo orden.

Por otra parte, la permutación que pasa de la segunda ordenación de los vértices, $\{v_{\sigma(1)}, v_{\sigma(2)}, \dots, v_{\sigma(n)}\}$, a la primera, $\{v_1, v_2, \dots, v_n\}$, no es otra que σ^{-1} . Por tanto, la definición de matriz de cambio de ordenación que hemos dado nos indica que

$$C_{\mathcal{O}_1\mathcal{O}_2} = (e_{\sigma^{-1}(1)}, e_{\sigma^{-1}(2)}, \dots, e_{\sigma^{-1}(n)}).$$

Para que la definición sea coherente, lo razonable sería que cambiar de ordenación y, luego, deshacer el cambio, fuera como no hacer nada. En otras palabras, que

$$C_{\mathcal{O}_1\mathcal{O}_2} C_{\mathcal{O}_2\mathcal{O}_1} = I_n. \quad (5.4)$$

Esto es así, puesto que al multiplicar $C_{\mathcal{O}_1\mathcal{O}_2}$ por $C_{\mathcal{O}_2\mathcal{O}_1}$ la matriz resultante tiene las mismas columnas que la primera, pero ordenadas según el orden que proporciona σ . Es decir, la i -ésima columna de la matriz producto es la columna de $C_{\mathcal{O}_1\mathcal{O}_2}$ que ocupa el lugar $\sigma(i)$, esto es, $e_{\sigma^{-1}(\sigma(i))} = e_i$.

Como consecuencia de (5.3) y (5.4) se obtiene

$$C_{\mathcal{O}_1\mathcal{O}_2} = (C_{\mathcal{O}_2\mathcal{O}_1})^T,$$

esto es, la matriz del cambio de una ordenación a otra es la traspuesta de la matriz del cambio de la segunda ordenación a la primera. \square

Veamos algunas propiedades con la matriz de adyacencia:

Proposición 5.3.8 *Dos grafos no orientados son isomorfos si, y sólo si, se pueden escoger órdenes en los vértices de ambos grafos respecto de los cuales las matrices de adyacencia sean iguales.*



DEMOSTRACIÓN. En efecto, es obvio que si dos grafos tienen la misma matriz de adyacencia, la aplicación que lleva el i -ésimo vértice del primer grafo al i -ésimo vértice del segundo grafo es un isomorfismo. Recíprocamente, si dos grafos son isomorfos por medio de un isomorfismo $f = (\phi_V, \phi_A) : G_1 \rightarrow G_2$, entonces dado un orden \mathcal{O}_1 en el primer grafo, tomaremos como i -ésimo vértice del segundo grafo la imagen por ϕ_V del i -ésimo vértice de G_1 . Esto define un orden \mathcal{O}_2 en G_2 con el cual la matriz de adyacencia tiene que ser la misma que la de G_1 con \mathcal{O}_1 . \square

Observación 5.3.9 Se tiene el mismo resultado si ambos grafos son orientados. Sin embargo, un grafo no orientado y uno orientado pueden tener la misma matriz de adyacencia sin ser isomorfos. Por ejemplo, el grafo simple no orientado $V = \{a, b\}$, $A = \{\{a, b\}\}$ y el grafo simple orientado $V = \{a, b\}$, $A = \{(a, b), (b, a)\}$ (véase la Figura 5.13) tienen la misma matriz de adyacencia y, obviamente, no son isomorfos. \square



Figura 5.13: Grafos no isomorfos con la misma matriz de adyacencia.

Observación 5.3.10 Como se ha visto en la demostración de la Proposición 5.3.8, un isomorfismo entre dos grafos isomorfos G_1 y G_2 puede relacionarse con la matriz C de cambio de ordenación que hemos visto anteriormente, sin más que tomar $C = C_{\mathcal{O}_2\mathcal{O}_1}$, siendo \mathcal{O}_1 y \mathcal{O}_2 las ordenaciones allí citadas. \square

Proposición 5.3.11 Un grafo no orientado es no conexo si, y sólo si, se puede dar un orden de los vértices de tal manera que la matriz de adyacencia sea de la forma

$$M = \begin{pmatrix} M_1 & 0 \\ 0 & M_2 \end{pmatrix}. \quad (5.5)$$

Un grafo es bipartito si, y sólo si, se puede dar un orden de los vértices de tal manera que la matriz de adyacencia se escriba como

$$M = \begin{pmatrix} 0 & M_1 \\ M_2 & 0 \end{pmatrix}. \quad (5.6)$$

DEMOSTRACIÓN. Es evidente que si la matriz de adyacencia es como en (5.5), los vértices implicados en la primera submatriz M_1 nunca se conectarán con los vértices implicados en M_2 . Entonces G no es conexo.

Por otra parte, si G no es conexo, tomemos un vértice v que no se pueda conectar con un camino con algún otro w . Consideremos todos los vértices (supongamos que hay r , incluido el propio v) que sí se pueden conectar mediante un camino con v . Consideramos un orden \mathcal{O} en G de tal manera que estos vértices sean los primeros r . Es evidente que la matriz de adyacencia tiene que ser de la forma (5.5) en donde M_1 es una submatriz $r \times r$. La submatriz M_2 es de tamaño distinto de cero ya que w ocupa un lugar posterior al lugar r en el orden \mathcal{O} .

Para el caso bipartito, simplemente hay que observar que estos grafos tienen sus vértices clasificados en dos subconjuntos de forma natural. Cualquier orden que tenga primero a los de un subconjunto y luego a los otros tendrá una matriz como (5.6). Lo recíproco es obvio. \square

Por último, hay una propiedad muy útil asociada a la matriz de adyacencia para contar “caminos” de cierta longitud entre dos vértices.

Teorema 5.3.12 Sea $G = (V, A)$ un grafo (orientado o no orientado) y sea M la matriz de adyacencia de G respecto de cierto orden $\{v_1, v_2, \dots, v_n\}$ de V . Se verifica que el número de caminos diferentes de longitud k entre v_i y v_j es igual a la entrada (i, j) de la matriz M^k .

DEMOSTRACIÓN. Lo probamos por inducción sobre k :

a) Para $k = 1$ el resultado es inmediato.



- b) Suponiendo cierto el resultado para k lo mostramos para $k + 1$. Como $M^{k+1} = M^k M$, la entrada (i, j) de M^{k+1} es igual a

$$b_{i1}m_{1j} + b_{i2}m_{2j} + \cdots + b_{in}m_{nj},$$

siendo $M^k = (b_{ij})$. Por la hipótesis de inducción, b_{is} nos da el número de caminos de v_i a v_s . Por otra parte, un camino de longitud $k + 1$ se descompone en un camino arbitrario de longitud k de v_i a un vértice v_s más un camino de longitud 1 de v_s a v_j . Así la fórmula anterior nos da todos los caminos de longitud $k + 1$, completándose la demostración. \square

5.4 Caminos eulerianos y hamiltonianos

5.4.1 Caminos eulerianos

Definición 5.4.1 Un camino euleriano en un grafo $G = (V, A)$ es un camino simple que contiene todas las aristas. Si existe un camino euleriano en G se dice que G es un grafo euleriano. \square

Observación 5.4.2 En la ciudad prusiana de Königsberg (hoy Kaliningrado, Rusia) era tema de distracción el intentar descubrir un itinerario para sus paseos de forma que pudiesen cruzar los siete puentes que cruzaban el río Pregel pasando por cada uno sólo una vez. Leonhard Euler, en su libro *Solutio Problematis ad Geometriam Situs Pertinentis* (1736), obra que puede considerarse como nacimiento de la Topología, proporcionó una demostración completa de la inexistencia de dicho paseo dando un resultado general para todas las posibles configuraciones de puentes y ríos. A modo de ilustración, de los trabajos de Euler podemos extraer la página recogida en la Figura 5.14, en la que se puede ver un esquema de dicha ciudad y de sus puentes.



Figura 5.14: Texto de Leonhard Euler.

En la Facultad de Matemáticas de la UCM se cuenta con la escultura “Homenaje a Euler”, de J. Vaquero Turcios, que modela la ciudad y sus famosos 7 puentes. Para simplificar el problema, es importante advertir que la disposición de las zonas de la ciudad separadas por los distintos brazos del río Pregel es equivalente al grafo de la Figura 5.15.



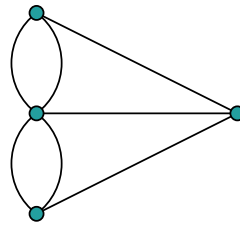


Figura 5.15: Grafo de los brazos del río Pregel.

Nuestro problema se reduce a continuación a determinar si el grafo anterior es recorrible sin levantar el lápiz del papel y sin trazar dos veces ninguna arista, es decir, si el grafo admite un recorrido euleriano.² □

Veamos un resultado que nos facilitará la caracterización de los grafos eulerianos:

Lema 5.4.3 *En un grafo no orientado conexo $G = (V, A)$ dos vértices cualesquiera pueden unirse mediante un camino simple.*

DEMOSTRACIÓN. Dados dos vértices v y v' de G existe un camino que los une, por ser G conexo. Consideremos, de todos los caminos que los unen, uno de longitud mínima. Supongamos que este camino no es simple y lleguemos a contradicción. En tal caso, el camino sería $C = a_1, a_2, \dots, a_k$ de forma que

$$a_1 = \{v, v_1\}, a_2 = \{v_1, v_2\}, \dots, a_{k-1} = \{v_{k-2}, v_{k-1}\}, a_k = \{v_{k-1}, v'\},$$

con, al menos, una arista repetida. Supongamos que $a_i = a_j$ con $1 \leq i < j$. En el caso de que el camino $a_i, a_{i+1}, \dots, a_{j-1}, a_j$ no sea cerrado (es decir, en ambas posiciones i y j la arista se recorre en el mismo sentido) entonces el camino $a_1, a_2, \dots, a_{i-1}, a_j, \dots, a_k$ llevaría de v a v' y tendría una longitud menor que el camino C , lo cual contradice la hipótesis. El otro caso es que el camino $a_i, a_{i+1}, \dots, a_{j-1}, a_j$ sea un circuito; entonces el camino $a_1, a_2, \dots, a_{i-1}, a_{j+1}, \dots, a_k$ llevaría de v a v' y tendría una longitud menor que el camino C , lo cual también contradice la hipótesis. □

Ahora podemos dar condiciones necesarias y suficientes para determinar si un grafo (multigrafo, pseudografo) es euleriano. En concreto, se tiene el siguiente resultado:

Teorema 5.4.4 (Euler) *Un grafo no orientado conexo es euleriano si, y sólo si, tiene 2 o ningún vértice de grado impar.*

- Si tiene 2 vértices impares, el camino euleriano comienza en cualquiera de ellos y termina en el otro.*
- Si no hay vértices de grado impar, el camino euleriano puede comenzar en cualquier vértice.*

DEMOSTRACIÓN.

⇒ Supongamos que el grafo es euleriano. Si realizamos un camino recorriendo de forma continua todo el grafo, cuando se pasa por un vértice cualquiera, se entra y se sale, es decir, se “consumen” dos aristas confluyentes en él. Como hay que recorrer todas las aristas, todos los extremos que concurren en un vértice cualquiera se van “consumiendo” de dos en dos. Por tanto, el grado de cualquier vértice debe ser par, salvo el punto de partida y de finalización si son distintos. En este caso, su grado será exactamente una unidad (la salida para el primero y la llegada para el segundo) más un número par por el mismo razonamiento anterior.

⇐ La argumentación de esta demostración se ejemplifica en la Figura 5.16. Si existen dos vértices de grado impar, podemos considerar, gracias al Lema 5.4.3, un camino simple que los une. Si hemos recorrido todo el grafo, ya hemos terminado. Si no es así, el subgrafo resultante al eliminar las aristas de este camino y, tras ello, los vértices que queden aislados, tendrá todos los vértices de grado par y, posiblemente, varias componentes conexas.

²Contrariamente a lo que se puede leer en muchos sitios, Euler no redujo en su trabajo el mapa de Königsberg al clásico grafo con 4 vértices de la figura anterior. La primera publicación impresa de dicha transformación es de W. Rouse Ball y data de 1892.



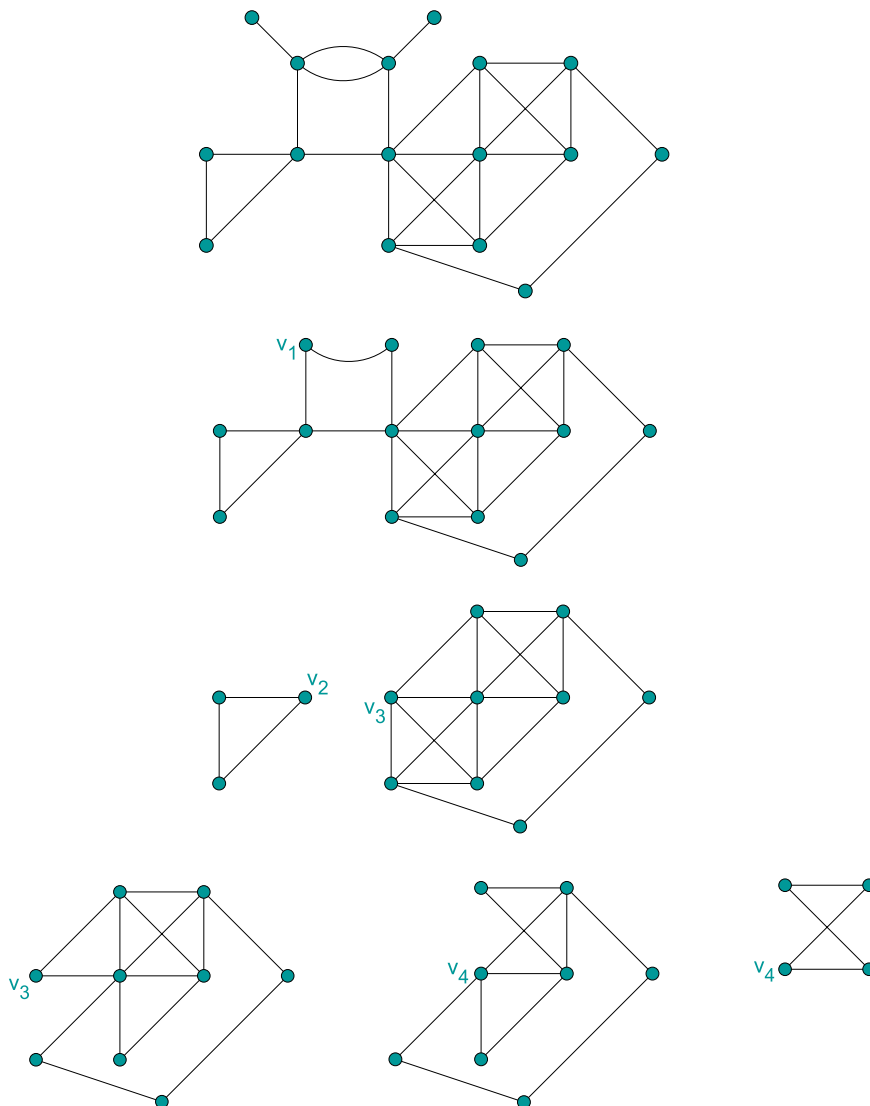


Figura 5.16: Ejemplo para la demostración del Teorema 5.4.4.

Fijemos nuestra atención en una de estas componentes. Es seguro que tiene, al menos, un vértice común con el camino simple eliminado (si no, el grafo original no sería conexo). Partiendo de dicho vértice v , construimos un circuito simple que comience y acabe en v . (Existe un camino así puesto que, partiendo de v , al ir pasando por los vertices vamos consumiendo las aristas de 2 en 2, de forma que, si podemos llegar a un vértice distinto de v siempre podemos salir de él; como hay un número finito de vértices, en algún momento llegaremos al vértice v). Si este camino pasa por todas las aristas de la componente conexa en la que estamos, lo adjuntamos al camino original, con la idea de que el camino simple original, cuando llegue a este circuito se desvíe por él y, luego, continúe. En otro caso, se considera el subgrafo resultante al eliminar las aristas de este circuito y, tras ello, los vértices aislados, el cual tendrá todos los vértices de grado par y, posiblemente, varias componentes conexas.

Basta iterar este procedimiento en las componentes conexas que vayan resultando sucesivamente para agotar todas las aristas (pues hay un número finito de ellas). En caso de que no hubiera dos vértices de grado impar la demostración sería la misma sin más que obviar el primer camino simple considerado. En la Figura 5.17 se muestra un camino euleriano para el grafo de la Figura 5.16. \square

De forma similar se puede demostrar el siguiente resultado:



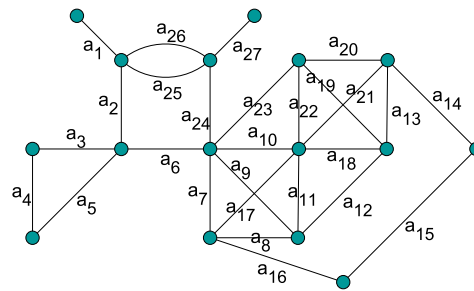


Figura 5.17: Camino euleriano en el grafo de la Figura 5.16.

Teorema 5.4.5 (Teorema de Euler para grafos orientados) *Un grafo orientado débilmente conexo es euleriano (respetando en el recorrido la orientación de las aristas) si, y sólo si, se satisface alguna de las condiciones siguientes:*

1) Todos los vértices $v \in V$ verifican

$$\text{Sgrad}(v) = \text{Egrad}(v)$$

y, en ese caso, el camino euleriano puede comenzarse en cualquier vértice.

2) Todos los vértices $v \in V$ satisfacen $\text{Sgrad}(v) = \text{Egrad}(v)$ salvo dos vértices v_1 (el de salida) y v_2 (el de llegada) que satisfacen

$$\text{Sgrad}(v_1) = \text{Egrad}(v_1) + 1 \text{ y } \text{Sgrad}(v_2) = \text{Egrad}(v_2) - 1. \quad \square$$

5.4.2 Caminos hamiltonianos

Una cuestión similar a la planteada por los puentes de Königsberg es la de determinar en un grafo un camino continuo que pase por todos los vértices una y sólo una vez. Históricamente, el problema germinal de esta situación se puede situar en el juego “La vuelta al mundo” de W. Hamilton (el mismo que descubrió los cuaterniones). El juego consiste en determinar un camino a lo largo de las aristas de un dodecaedro que pase por todos los vértices.

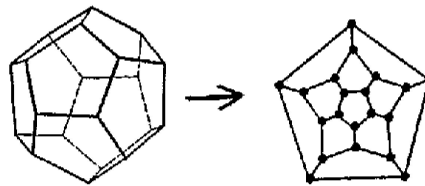


Figura 5.18: Dodecaedro de Hamilton.

Para estudiar el problema es interesante transformar el *dodecaedro de Hamilton* en un grafo (véase la Figura 5.18). Esta transformación está relacionada con la sección sobre grafos planos que se estudia más adelante.

Definición 5.4.6 Un *camino hamiltoniano* en un grafo $G = (V, A)$ es un camino simple $\{v_1, v_2, \dots, v_n\}$ que pasa por todos los vértices $v_i \in V$ una y sólo una vez. Si el camino, una vez llegado al último punto v_n , puede continuarse con una arista más hasta el punto de partida v_1 , se dice que es *cerrado*. Diremos que G es un *grafo hamiltoniano* si contiene un camino hamiltoniano. \square

Observación 5.4.7 Todo grafo hamiltoniano tiene un subgrafo que contiene todos sus vértices (es decir, al que se llega eliminando solamente aristas) que es euleriano. Este criterio puede resultar útil para probar que un grafo no es hamiltoniano. \square



Son muchos los problemas que se pueden modelizar mediante grafos y en los que la solución depende de la existencia o no de caminos hamiltonianos en el grafo correspondiente. Quizás el más famoso es el que se conoce como el problema del viajante: un viajante (representante o vendedor de un producto) visita varias ciudades para comercializar su producto; ¿puede visitar todas las ciudades (vértices) sin pasar dos veces por la misma ciudad? Es evidente que esto depende de cómo estén dispuestas las carreteras (aristas) que unen las ciudades. En general, los problemas de distribución o reparto (por ejemplo, la recogida de basuras en una ciudad) están ligados a los caminos hamiltonianos.

Aunque la definición de grafo euleriano y hamiltoniano son similares (cambiando aristas por vértices) sin embargo son conceptos que no tienen una gran relación. En la Figura 5.19 pueden verse ejemplos de un grafo hamiltoniano que no es euleriano, un grafo euleriano que no es hamiltoniano y un grafo que tiene ambas propiedades, pero con un camino distinto para cada una de ellas.

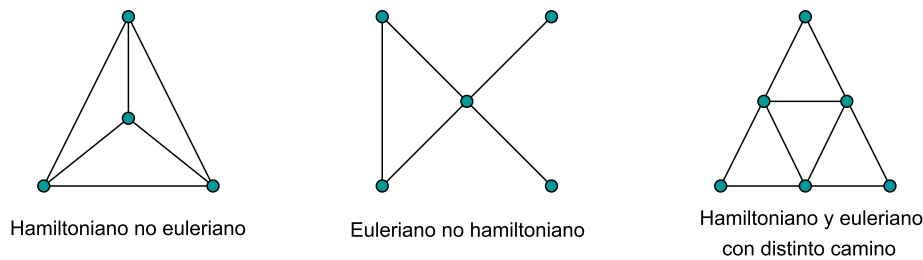


Figura 5.19: Grafos eulerianos y hamiltonianos.

De hecho, mientras que los teoremas de Euler caracterizan los grafos (orientados y no orientados) que son eulerianos, no hay condiciones necesarias y suficientes para determinar si un grafo admite un camino (cerrado o no) hamiltoniano. Hay muchos resultados parciales; de entre ellos, destaca el siguiente (del que no se incluye demostración):

Teorema 5.4.8 *Todo grafo conexo simple de $n \geq 3$ vértices de forma que el grado de cada vértice sea, al menos, $\frac{n}{2}$ admite un camino hamiltoniano cerrado.* \square

Un camino hamiltoniano cerrado que recorre el dodecaedro de Hamilton (el grafo que lo representa) está recogido en la Figura 5.20.

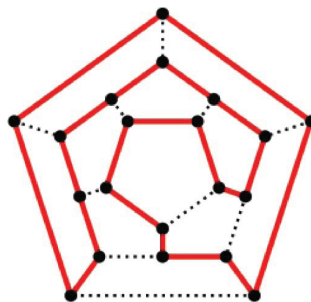


Figura 5.20: Camino hamiltoniano cerrado en el dodecaedro de Hamilton.

5.5 Grafos planos

Definición 5.5.1 Un grafo es *plano* si puede ser dibujado en el plano \mathbb{R}^2 sin que se corte ninguna arista. \square

Observación 5.5.2 Los grafos planos juegan un papel importante en muchas situaciones. Por ejemplo, el diseño de placas electrónicas en donde los vértices son los dispositivos electrónicos y las aristas los caminos impresos en la placa. El cruce de aristas representa un cortocircuito. \square



Ejemplo 5.5.3

- a) Un ejemplo sencillo de grafo no plano es el grafo completo K_5 de 5 vértices. En la Observación 5.5.9 veremos que, efectivamente, no admite una representación plana; sin embargo, es muy sencillo comprobar que K_2 , K_3 y K_4 sí son planos. Evidentemente K_n con $n > 5$ tampoco es plano, pues contiene a K_5 como subgrafo.
- b) Otro ejemplo importante es el problema de las tres casas y los tres pozos (el grafo bipartito $K_{3,3}$): tres vecinos, Antonio, Bernardo y Carla se llevan muy mal entre ellos. Desgraciadamente conviven en una parcela con tres pozos tal y como se muestra en la Figura 5.21. ¿Pueden construir cada uno de ellos 3 caminos uniendo su casa con cada uno de los pozos de tal manera que los caminos no se crucen? (así se evitan encuentros desagradables). Si tal configuración de caminos existiese, escogiendo el circuito

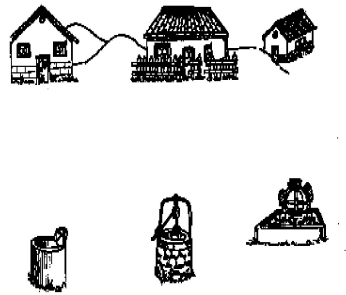


Figura 5.21: Problema de las tres casas y los tres pozos.

determinado por el siguiente orden de vértices

$$C_1P_1C_2P_2C_3P_3C_1$$

(donde C_1, C_2, C_3 representan las 3 casas y P_1, P_2, P_3 los tres pozos) quedaría por trazar las aristas C_1P_2 , C_2P_3 y C_3P_1 . Sin embargo, el circuito mencionado es una curva cerrada simple con una región “dentro” y otra “fuera” (teorema de la Curva de Jordan). Así, C_1P_2 se puede trazar por dentro y C_2P_3 por fuera (o al contrario), imposibilitando el trazado de C_3P_1 . \square

Los dos ejemplos anteriormente vistos son esencialmente los grafos que caracterizan la no planaridad. En efecto:

Definición 5.5.4 Una *subdivisión elemental* en un grafo consiste en añadir un vértice w en una arista $\{u, v\}$; es decir, añadir un vértice nuevo w al conjunto de vértices y reemplazar una arista $\{u, v\}$ por dos aristas $\{u, w\}$ y $\{w, v\}$. Dos grafos se dicen que son *homeomorfos* si ambos son resultado de la realización de subdivisiones elementales del mismo grafo (véase la Figura 5.22). \square

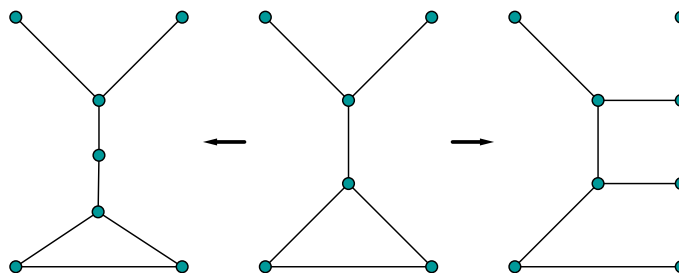


Figura 5.22: Grafos homeomorfos.



Teorema 5.5.5 *Un grafo no es plano si, y sólo si, contiene un subgrafo homeomorfo a K_5 o $K_{3,3}$.*

DEMOSTRACIÓN. No se incluye la prueba. \square

Ejemplo 5.5.6 De los grafos representados en la Figura 5.23, el grafo (a) no es plano puesto que contiene un subgrafo $K_{3,3}$ (basta eliminar la arista horizontal gruesa en azul; los tres vértices alineados en horizontal serían uno de los dos conjuntos de vértices y los tres vértices restantes el otro). El grafo (b) es trivialmente plano (basta desplazar el vértice rojo dentro del ciclo formado por los cuatro vértices azules).

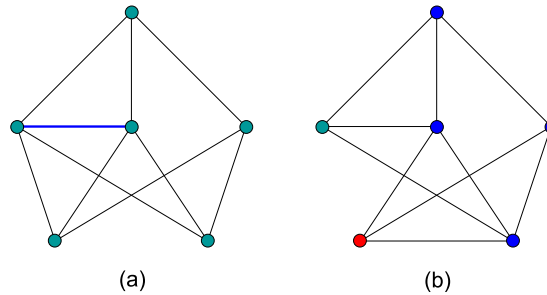


Figura 5.23: Grafo no plano y grafo plano.

5.5.1 Fórmula de Euler

Un grafo plano divide el plano en varias regiones disjuntas, incluida una región no acotada exterior. Si c denota el número de tales regiones, v el número de vértices y a el número de aristas, se tiene el siguiente resultado:

Teorema 5.5.7 (Euler) *Si G es un grafo plano conexo, entonces $c + v = a + 2$.*

DEMOSTRACIÓN. Se trata de una prueba similar a una demostración por inducción en el número de aristas. Partimos del subgrafo G_1 formado por una arista y sus dos vértices (o el mismo) de G y vamos a ir añadiéndole aristas hasta terminar con el grafo G .

- Para G_1 se cumple la igualdad de Euler.
- Ahora añadimos una arista a G_1 de forma que uno de los extremos sea un vértice de G_1 . Si el otro vértice no es de G_1 , lo añadimos. Así construimos G_2 y, de manera recurrente, los grafos G_3, G_4, \dots hasta llegar al grafo G . Sean v_n, a_n y c_n el número de vértices, aristas y regiones de G_n . Ya hemos visto que $c_1 - a_1 + v_1 = 2$. Si ahora suponemos que la fórmula de Euler es cierta para G_{n-1} , veamos que también es cierta para G_n . Si la arista nueva tiene los dos vértices en G_{n-1} , entonces es fácil ver que $c_n = c_{n-1} + 1, a_n = a_{n-1} + 1$ y $v_n = v_{n-1}$, por lo que se sigue cumpliendo la fórmula de Euler. Si, por el contrario, hay que añadir un vértice a la nueva arista, no se crea ninguna nueva región y, por tanto, $c_n = c_{n-1}, a_n = a_{n-1} + 1, v_n = v_{n-1} + 1$ y se sigue cumpliendo la igualdad. \square

Corolario 5.5.8 *En un grafo simple plano conexo con $v \geq 3$ se cumple que $a \leq 3v - 6$.*

DEMOSTRACIÓN. Para probar este resultado utilizaremos el concepto de grado de una región. En un grafo simple plano conexo, llamaremos grado de una región al número de aristas que se recorren al dibujar su frontera, de modo que una arista se cuenta dos veces si se pasa dos veces por ella (véase la Figura 5.24).

Pues bien: hacemos notar que la suma de los grados de todas las regiones es $2a$, ya que cada arista la estamos contando dos veces (bien porque es frontera de dos regiones distintas, bien porque la recorremos dos veces al dibujar la frontera de una región). Por otra parte, la hipótesis $v \geq 3$ obliga a que el grado de cada región sea, al menos, tres, por lo que la suma de los grados de todas las regiones (es decir, $2a$) es mayor o igual que $3c$. Así pues, hemos obtenido que

$$c \leq \frac{2}{3}a.$$

El resultado se concluye sustituyendo esta desigualdad en la fórmula de Euler. \square



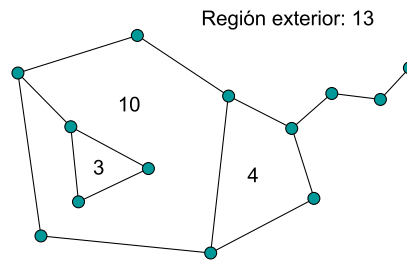


Figura 5.24: Grado de una región.

Observación 5.5.9 K_5 no es plano pues, en él, se tiene que $a = 10$, $v = 5$ y $10 > 15 - 6 = 9$. □

Observación 5.5.10 La fórmula de Euler también es válida para poliedros. De hecho, un poliedro se puede desarrollar en el plano sin roturas deformándolo de tal manera que se preserven aristas, caras y vértices. El resultado es un grafo plano. En ese caso a y v son las aristas y vértices del poliedro y el número de regiones c corresponde con el número de caras (véase la Figura 5.25). □

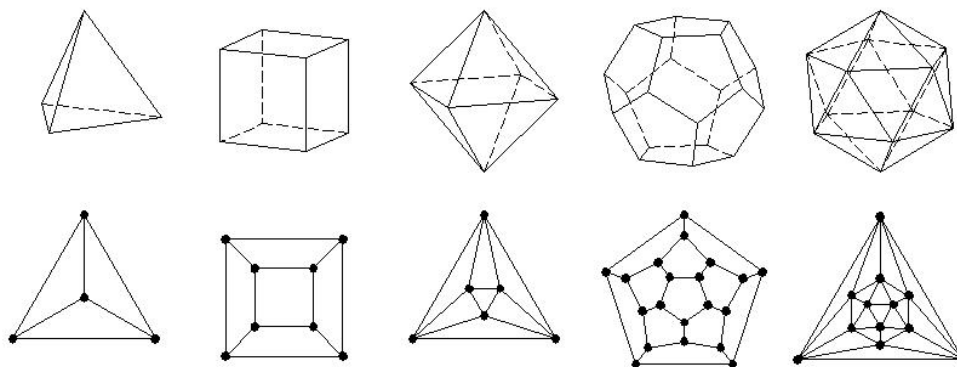


Figura 5.25: Poliedros.

5.6 Coloración de grafos

Definición 5.6.1 Una *coloración de un grafo simple* es una asignación de colores a cada vértice de forma que no haya dos vértices adyacentes con el mismo color. El número mínimo de colores $\chi(G)$ necesarios se denomina *número cromático de vértices*. □

Ejemplo 5.6.2 Supongamos que el anfitrión de una fiesta quiere regalar a cada uno de sus invitados un ramo de rosas. Dos invitados pueden ser “amigos” o “desconocidos” entre ellos. Para quedar bien, el anfitrión quiere que no haya dos conocidos con un ramo del mismo color. ¿Cuál es el mínimo número de colores requeridos para poder llevar a cabo su objetivo? □

Observación 5.6.3 También se puede introducir la noción de *coloración de aristas* de forma que dos aristas del mismo color nunca concurren en un mismo vértice. □

Desde luego, para todo grafo de n vértices se verifica que $\chi(G) \leq n$. En el caso del grafo completo $G = K_n$, la cota anterior se alcanza, es decir, $\chi(K_n) = n$. Pero hay algunos resultados en donde se puede bajar de dicho valor máximo.



Teorema 5.6.4 (Brooks, 1941) Sea $G = (V, A)$ un grafo conexo que no es completo ni es un ciclo de longitud impar. Entonces

$$\chi(G) \leq \max_{v \in V} \text{grad}(v).$$

DEMOSTRACIÓN. No se incluye la prueba. \square

Observación 5.6.5 Nótese que en el caso de un ciclo de longitud impar, el número mínimo de colores es 3, a pesar de que $\max_{v \in V} \text{grad}(v) = 2$. \square

Ejemplo 5.6.6 El grafo de la Figura 5.26 (a) admite una coloración mediante tres colores, como se muestra en la Figura 5.26 (b). Además, su número cromático es 3 puesto que contiene un ciclo de orden impar (el ciclo C_5 marcado en grueso y anaranjado en la Figura 5.26 (b)). \square

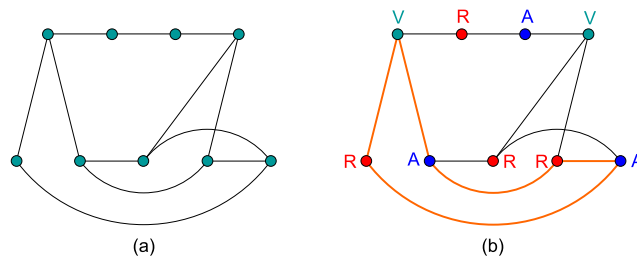


Figura 5.26: Coloración de un grafo.

En el caso de los grafos planos, se puede conseguir una cota más restrictiva a la del Teorema de Brooks. En general, si G es un grafo plano, siempre se verifica que $\chi(G) \leq 4$. Este resultado es conocido como el *teorema de los cuatro colores*, y se puede enunciar como sigue:

Teorema 5.6.7 (de los cuatro colores) *Cualquier mapa se puede colorear con, a lo sumo, 4 colores de tal manera que no haya dos países fronterizos del mismo color.* \square

Observación 5.6.8 Pasar de una mapa a un grafo es sencillo: los vértices representarán a los países y las aristas se establecerán únicamente si dos países tienen frontera común. \square

5.7 Teoría espectral de grafos. Aplicación a Google

Desde hace unos años, *Google* se ha convertido en el buscador estándar en la red. Uno de sus secretos es el algoritmo *PageRank* que utiliza para ordenar los resultados de las búsquedas. Este algoritmo conjuga una interesante aplicación de la Teoría de Grafos con el Álgebra Lineal.

Google fue creado en 1998 por Sergei Brin y Lawrence Page en la Universidad de Stanford. El nombre es una variación sobre el término *googol*, es decir, el número 10^{100} , haciendo referencia al gran volumen de datos con los que el programa tiene que trabajar. De hecho, según la página de Google, atiende a más de 250 millones de consultas diarias e indexa varios miles de millones de páginas web.

El problema central que vamos a abordar es el siguiente: si ya tenemos los resultados de una búsqueda dada, ¿cómo los mostramos?, ¿en qué orden?, ¿cómo podemos hacer una jerarquía? Necesitamos un criterio de ordenación, una asignación de importancias a cada página de la red encontrada

Páginas	P_1, P_2, \dots, P_N
Importancia	x_1, x_2, \dots, x_N

Google usa, además de la herramienta básica de ordenación que veremos aquí, otros criterios:

- a) No es lo mismo que un cierto término aparezca en una página en el título, en negrita, en un tipo de letra pequeña, etc. Es decir, el formato y el tipo de letra aportan una información que puede ser de utilidad.



- b) Para búsquedas combinadas de varios términos simultáneos, no es lo mismo que esos términos aparezcan “lejos” o “cerca” los unos de los otros en la página web en la que se encuentren.

Salvo por estos aspectos (y los vinculados a la estrategia de negocio de la empresa), el modelo de ordenación de Google se basa en lo siguiente:

- a) Primer paso: descripción de la información, con un grafo orientado o dirigido G . Cada sitio de la red es un vértice, y hay una arista (dirigida) entre P_i y P_j si desde la página P_i hay un enlace a la página P_j . Construimos la traspuesta de la matriz de adyacencia; es decir, la matriz $M = (m_{ij})$ donde m_{ij} es el número de enlaces que van de P_j a P_i (si no hay enlaces de P_j a P_i entonces $m_{ij} = 0$ y supondremos que una página siempre se enlaza a sí misma, por lo que los elementos diagonales valen 1).

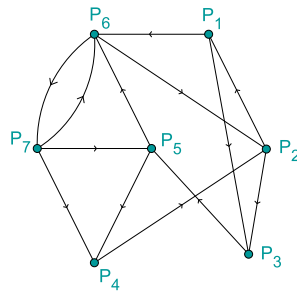


Figura 5.27: Grafo de las páginas de Internet.

Por ejemplo, suponiendo que sólo hubiera siete páginas en la red y que éstas estuvieran enlazadas como se muestra en la Figura 5.27, la matriz M correspondiente sería

$$M = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

- b) La importancia de una página P_i será mayor en función de dos factores:

- 1) que sea muy probable llegar a P_i ; por ejemplo, si hay muchos enlaces desde distintas páginas a P_i .
- 2) que las citas que llegan a P_i sean de sitios que, a su vez, tienen gran importancia.

El modelo asigna un valor de importancia x_i a una página (vértice) P_i en función de lo probable que sea llegar a ella. Se considera que una parte de esta probabilidad es uniforme (todas las páginas tendrían la misma probabilidad de ser visitadas aleatoriamente) y ésta sería $\frac{1}{N}$, siendo N el número de páginas en Internet. La otra parte de esta probabilidad se considera proporcional a la suma de las importancias de los vértices de los que salen aristas confluyentes en P_i o, lo que es lo mismo, de las páginas que enlazan con P_i . La proporcionalidad vendrá dada en función de los enlaces que salgan de cada página. Así, en el ejemplo, la importancia x_6 de la página P_6 vendrá dada a partir de:

$$\frac{1}{3}x_1 + \frac{1}{3}x_5 + \frac{1}{3}x_6 + \frac{1}{4}x_7$$

donde estamos ponderando lo que aportan las páginas P_1 , P_5 y P_6 con un tercio, porque de cada una de ellas salen 3 enlaces, y lo que aporta la página P_7 con un cuarto, porque de ella salen 4 enlaces. En general, la importancia de la página P_i se expresará mediante

$$\sum_{j=1}^N \frac{m_{ij}}{k_j} x_j,$$



donde k_j es el número de enlaces que salen de la página P_j . Esta expresión es la coordenada i -ésima del vector $\widetilde{M}x$, donde \widetilde{M} es la matriz que tiene por elementos

$$\widetilde{m}_{ij} = \frac{m_{ij}}{k_j}.$$

En nuestro ejemplo,

$$\widetilde{M} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{2} & 0 & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{3} & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{3} & 0 & \frac{1}{4} \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{4} \end{pmatrix}$$

Al dividir por k_j estamos ponderando el hecho de que no es lo mismo estar enlazado desde una página *selectiva* que sólo tiene un enlace, que desde una página *promiscua* que tiene muchos. Nótese que

$$\sum_{i=1}^N \frac{m_{ij}}{k_j} = 1. \quad (5.7)$$

Finalmente, la importancia de la página P_i se define como:

$$x_i = \frac{1-d}{N} + d \sum_{j=1}^N \frac{m_{ij}}{k_j} x_j, \quad (5.8)$$

para algún d entre 0 y 1 (Google toma $d = 0.85$). Sumando en i y usando (5.7) tenemos

$$\sum_{i=1}^N x_i = (1-d) + d \sum_{i=1}^N \sum_{j=1}^N \frac{m_{ij}}{k_j} x_j = (1-d) + d \sum_{j=1}^N \sum_{i=1}^N \frac{m_{ij}}{k_j} x_j = (1-d) + d \sum_{j=1}^N x_j$$

y, por lo tanto,

$$\sum_{i=1}^N x_i = 1.$$

Así pues, a partir de la expresión (5.8), podemos escribir la siguiente igualdad vectorial:

$$x = \left(\frac{1-d}{N} \mathbb{1}_N + d\widetilde{M} \right) x,$$

donde por $\mathbb{1}_N$ denotamos la matriz cuadrada de orden N cuyos elementos son todos 1.

En definitiva, el modelo *PageRank* busca un autovector $x = (x_i)$ de la matriz

$$\mathcal{A} = \frac{1-d}{N} \mathbb{1}_N + d\widetilde{M}$$

asociado al autovalor $\lambda = 1$ cuyas coordenadas sean todas ellas estrictamente positivas y que, además, verifique la propiedad

$$\sum_{i=1}^N x_i = 1.$$



Nuestro primer propósito es demostrar que el problema de encontrar un autovector asociado al autovalor $\lambda = 1$ con las características apuntadas, tiene solución. Posteriormente, una vez resuelto el problema teórico, describiremos el algoritmo que nos conducirá a la solución y nos interesaremos por su implementación en el ordenador.

En primer lugar, veamos que $\lambda = 1$ es en verdad autovalor de \mathcal{A} . Para ello, basta considerar la relación (5.7), la cual nos asegura que las columnas de \mathcal{A} suman, todas ellas, uno. Por tanto, las columnas de $\mathcal{A} - I_N$ suman, todas ellas, cero. Esto implica que las filas de $\mathcal{A} - I_N$ son linealmente dependientes y, por tanto, el determinante de la matriz $(\mathcal{A} - I_N)$ es nulo; esto es, $\lambda = 1$ es autovalor de \mathcal{A} .

Con vistas a la demostración del resto del aserto enunciamos (sin demostración) el siguiente resultado. Se trata de un teorema ya clásico (es de principios del siglo XX) muy anterior a la existencia de los ordenadores, pero que ha sido básico para el desarrollo del método *PageRank*.

Teorema 5.7.1 (Perron, 1907) *Sea M una matriz (cuadrada) con elementos positivos. La matriz M tiene un (único) autovalor λ con las propiedades:*

- λ es positivo y simple (multiplicidad algebraica).
- El autoespacio asociado a λ (que es, por tanto, de dimensión 1) está generado por un vector $x = (x_i)$ con $x_i > 0$.
- Para cualquier otro autovalor μ de la matriz M se verifica que $|\mu| < \lambda$. \square

El autovalor λ del teorema de Perron se suele denominar *autovalor dominante*.

El teorema se aplica a nuestra matriz, puesto que todos los elementos de \mathcal{A} son positivos. Si probásemos que el autovalor dominante de \mathcal{A} es $\lambda = 1$, habríamos demostrado que el problema planteado tiene (una única) solución. Para ello, nos bastará demostrar que todos los autovalores de \mathcal{A} tienen módulo menor o igual que 1:

Proposición 5.7.2 *Sea μ un autovalor de $\mathcal{A} = (a_{ij})_{i,j=1}^N$ y sea $v \neq 0$ un autovector asociado. Entonces $|\mu| \leq 1$*

DEMOSTRACIÓN. Denominando $m = \sum_{i=1}^N |v_i| > 0$, podemos escribir

$$\begin{aligned} |\mu|m &= |\mu| \sum_{i=1}^N |v_i| = \sum_{i=1}^N |\mu v_i| = \sum_{i=1}^N \left| \sum_{j=1}^N a_{ij} v_j \right| \leq \sum_{i=1}^N \sum_{j=1}^N |a_{ij}| |v_j| \\ &= \sum_{j=1}^N \sum_{i=1}^N |a_{ij}| |v_j| = \left(\sum_{i=1}^N |a_{ij}| \right) \left(\sum_{j=1}^N |v_j| \right) = \sum_{j=1}^N |v_j| = m, \end{aligned}$$

donde en el penúltimo paso se ha utilizado que las columnas de \mathcal{A} suman 1. Basta dividir por m para tener el resultado. \square

El problema está, desde el punto de vista teórico, resuelto: Nótese que si el autovector x proporcionado por el Teorema de Perron no verifica

$$\sum_{i=1}^N x_i = 1,$$

basta considerar el vector \tilde{x} definido por

$$\tilde{x} = \frac{1}{\sum_{i=1}^N x_i} x.$$

Nos queda por analizar cómo encontrar una solución en un tiempo que sea suficientemente pequeño como para que resulte útil. La respuesta la da el método de la potencia para el cálculo de un autovector asociado al autovalor dominante:



Observación 5.7.3 (Método de la potencia) Supongamos, por simplificar, que tenemos una matriz A diagonalizable con una base de autovectores $\{v_1, v_2, \dots, v_N\}$ correspondiente a los autovalores

$$\lambda_1 > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_N|.$$

Tomemos un vector arbitrario

$$w_0 = c_1 v_1 + c_2 v_2 + \dots + c_N v_N,$$

con $c_1 \neq 0$. Entonces,

$$Aw_0 = \sum_{i=1}^N c_i A v_i = \sum_{i=1}^N c_i \lambda_i v_i$$

e, iterando,

$$A^k w_0 = \sum_{i=1}^N c_i \lambda_i^k v_i = c_1 \lambda_1^k v_1 + \sum_{i=2}^N c_i \lambda_i^k v_i.$$

Dividiendo por λ_1^k en ambos lados,

$$\frac{1}{\lambda_1^k} A^k w_0 = c_1 v_1 + \sum_{i=2}^N c_i \left(\frac{\lambda_i}{\lambda_1} \right)^k v_i.$$

Como

$$\left| \frac{\lambda_i}{\lambda_1} \right| < 1, \quad i = 2, 3, \dots, N,$$

se tiene que

$$\frac{1}{\lambda_1^k} A^k w_0 \xrightarrow{k \rightarrow \infty} c_1 v_1. \quad \square$$

En el caso particular de nuestra matriz de Google \mathcal{A} , puesto que $\lambda_1 = 1$, tendremos que la sucesión de vectores w_k , donde

$$w_k = \mathcal{A}^k w_0,$$

converge a un autovector asociado al autovalor dominante. El proceso se detendrá cuando dos vectores seguidos de esta sucesión w_k y w_{k+1} estén suficientemente próximos. Entonces, se toma como solución el vector

$$\frac{1}{\sum_{i=1}^N (w_{k+1})_i} w_{k+1}.$$

Para acabar, nos quedarían por analizar los aspectos más puramente computacionales del cálculo del autovector mediante el método de la potencia. Para ello, teniendo en cuenta que

$$w_k = \mathcal{A} w_{k-1},$$

lo que tenemos que saber calcular de una forma eficaz es el producto de la matriz \mathcal{A} por un vector cualquiera b . Si volvemos ahora a la definición de \mathcal{A} , denotando por $\mathbf{1}_N$ el vector columna de tamaño N cuyas coordenadas valen todas 1, escribimos

$$\mathcal{A}b = \left(\frac{1-d}{N} \mathbf{1}_N + d\tilde{M} \right) b = \left(\sum_{i=1}^N b_i \right) \frac{1-d}{N} \mathbf{1}_N + d\tilde{M}b.$$

Así pues, si sabemos calcular, de forma eficaz, el producto de la matriz \tilde{M} por un vector, tendremos el trabajo hecho.

La matriz \tilde{M} tiene muy pocos elementos no nulos; es una matriz *dispersa* (*sparse* en inglés). Debemos encontrar una forma de almacenar la matriz en la memoria del ordenador (y, más importante, de operar con ella) que sólo involucre a los elementos no nulos.



Una de las maneras que se suelen utilizar para guardar en memoria una matriz como ésta es almacenar sólo los elementos no nulos como se indica a continuación. Se utilizan tres vectores L1, L2 y L3, actuando los dos primeros a modo de punteros. En L1 se guardan, acumulativamente, el número de elementos no nulos de cada fila, en L2 sus columnas (colocando, para cada fila, en último lugar la columna del elemento diagonal) y en L3 los coeficientes correspondientes, en ese orden. Así, para la matriz de nuestro ejemplo

$$\widetilde{M} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{2} & 0 & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{3} & 0 & \frac{1}{4} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{3} & 0 & \frac{1}{4} \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{4} \end{pmatrix}$$

se tendría

$$L1 = (0, 2, 5, 8, 11, 14, 18, 20)$$

$$L2 = (\underbrace{2, 1}, \underbrace{4, 6, 2, 1}, \underbrace{2, 3, 5, 7, 4}, \underbrace{3, 7, 5, 1}, \underbrace{5, 7, 6}, \underbrace{6, 7})$$

$$L3 = (\underbrace{\frac{1}{3}, \frac{1}{3}}, \underbrace{\frac{1}{2}, \frac{1}{3}, \frac{1}{3}, \frac{1}{3}}, \underbrace{\frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{2}}, \underbrace{\frac{1}{2}, \frac{1}{4}, \frac{1}{3}, \frac{1}{3}}, \underbrace{\frac{1}{4}, \frac{1}{3}, \frac{1}{4}}, \underbrace{\frac{1}{3}, \frac{1}{4}}).$$

Si se ha usado este tipo de almacenamiento para la matriz \widetilde{M} , la operación $w = \mathcal{A}b$ queda implementada, con la sintaxis de Matlab, mediante las líneas que vienen a continuación:

```
w=zeros(N,1);
for i=1:N
    for j=L1(i)+1:L1(i+1)
        w(i)=w(i)+L3(j)*b(L2(j));
    end
end
w=(1-d)*sum(b)/N+d*w;
```

5.8 Problemas

5.8.1 El grafo de intersección de una colección de conjuntos A_1, A_2, \dots, A_n es el grafo que tiene un vértice por cada conjunto y una arista conectando dos vértices si estos conjuntos tienen intersección no vacía. Determinar los grafos de intersección de las siguientes colecciones:

- a) $A_1 = \{\dots, -4, -3, -2, -1, 0\}$, $A_2 = \{\dots, -2, -1, 0, 1, 2, \dots\}$, $A_3 = \{2n : n \in \mathbb{Z}\}$, $A_4 = \{2n+1 : n \in \mathbb{Z}\}$ y $A_5 = \{3n : n \in \mathbb{Z}\}$.
- b) $A_1 = \{x \in \mathbb{R} : x < 0\}$, $A_2 = \{x \in \mathbb{R} : -1 < x < 0\}$, $A_3 = \{x \in \mathbb{R} : 0 < x < 1\}$, $A_4 = \{x \in \mathbb{R} : x < -1\}$ y $A_5 = \mathbb{R}$.

5.8.2 Determinar cuándo el grafo completo K_n es bipartito. Lo mismo con los grafos C_n (ciclo de n vértices), R_n (rueda de $n+1$ vértices) o Q_n (esqueleto de vértices y aristas del cubo de \mathbb{R}^n).

5.8.3 Probar que todo grafo simple siempre tiene dos vértices con el mismo grado.

5.8.4 Es conocido el juego del zorro, la cabra, el saco de coles y el hortelano. El hortelano quiere cruzar un río con su pequeño bote. Lo malo es que en el bote solo cabe (además de él) una sola cosa más. Además, obviamente, la cabra no puede estar con el zorro o con el saco de coles si no está el hortelano presente. Construir el grafo bipartito de las posibles configuraciones. Construir una estrategia para pasar todo al otro lado del río.



5.8.5 ¿Cuántas aristas tiene un grafo de 10 vértices si todos ellos son de grado 4? ¿Y un grafo con 9 vértices si todos ellos son de grado 3?

5.8.6 ¿Cuál es el mínimo número posible de vértices de un grafo de 7 aristas si cada vértice tiene grado menor o igual que 3? Razonar la respuesta.

5.8.7 Probar que si en un grafo simple de n vértices, cada uno tiene grado mayor o igual a $\frac{n-1}{2}$ entonces el grafo es conexo.

5.8.8 Sea G un grafo. Si v es el número de vértices, a el número de aristas,

$$M = \max_{v \in V} \text{grad}(v) \text{ y } m = \min_{v \in V} \text{grad}(v),$$

probar que

$$m \leq \frac{2a}{v} \leq M.$$

5.8.9 El grafo complementario \bar{G} de un grafo simple G es aquél que tiene los mismos vértices que G de forma que dos vértices están unidos en \bar{G} si, y sólo si, no lo están en G . Determinar \bar{K}_n , $\bar{K}_{m,n}$, \bar{C}_n , \bar{R}_n y \bar{Q}_n .

5.8.10 En una clase hay 32 estudiantes. ¿Es posible que nueve de ellos tengan 3 amigos (en clase) cada uno, once tengan 4 amigos cada uno y doce tengan 5 amigos cada uno? Inventar otra distribución, modificando ligeramente la anterior, que sí sea posible.

5.8.11 Se tiene una caja con 6 agujas de punto largas. Intentar colocarlas sobre una mesa de tal manera que cada una de ellas toque exactamente a otras dos. Volver a colocarlas de forma que cada una de ellas toque exactamente a otras tres. Intentar hacer esto último con 7 agujas (en vez de 6).

5.8.12 ¿Cuántos grafos simples no orientados distintos hay (salvo isomorfía) de 3 ó 4 vértices?

5.8.13 ¿Cuántos grafos simples no isomorfos hay que tengan 5 vértices y 3 aristas? Razonar la respuesta.

5.8.14 Dibujar el grafo definido por la matriz de adyacencia

$$\begin{pmatrix} 0 & 1 & 3 & 0 & 4 \\ 1 & 2 & 1 & 3 & 0 \\ 3 & 1 & 1 & 0 & 1 \\ 0 & 3 & 0 & 0 & 2 \\ 4 & 0 & 1 & 2 & 3 \end{pmatrix}.$$

5.8.15 Dibujar el grafo dado por la matriz de adyacencia

$$\begin{pmatrix} 0 & 2 & 3 & 0 \\ 1 & 2 & 2 & 1 \\ 2 & 1 & 1 & 0 \\ 1 & 0 & 0 & 2 \end{pmatrix}.$$

5.8.16 ¿Qué relación existe entre la matriz de adyacencia de un grafo simple G y la de su complementario \bar{G} ? Si G_1 y G_2 son dos grafos simples, demostrar que:

$$G_1 \text{ y } G_2 \text{ son isomorfos} \Leftrightarrow \bar{G}_1 \text{ y } \bar{G}_2 \text{ son isomorfos.}$$

5.8.17 Se considera un grafo no orientado con matriz de adyacencia

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Decidir, justificando la respuesta, si el grafo es conexo y si es bipartito.



5.8.22 Decidir si es euleriano el grafo no orientado cuya matriz de adyacencia es

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

5.8.23 Se cuenta con un alambre de 120 cm de longitud. ¿Se puede usar dicho alambre para formar las aristas de un cubo de 10 cm de lado sin realizar ningún corte? Si no es posible, ¿cuál es el mínimo número de cortes necesarios?

5.8.24 Determinar cuáles de los siguientes grafos son eulerianos:

- a) El grafo obtenido al representar las comunidades autónomas de la España peninsular como vértices y dibujar las aristas únicamente cuando dos comunidades son vecinas.
- b) El grafo obtenido con el problema de Euler aplicado a la disposición de los puentes de la actual Kaliningrado (véase la Figura 5.29).
- c) Los grafos de la Figura 5.30 (si el grafo es dirigido, el camino euleriano debe respetar la orientación de las aristas):

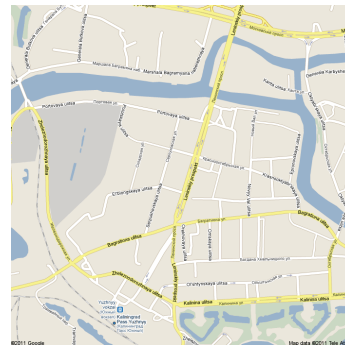


Figura 5.29: Ciudad de Kaliningrado.

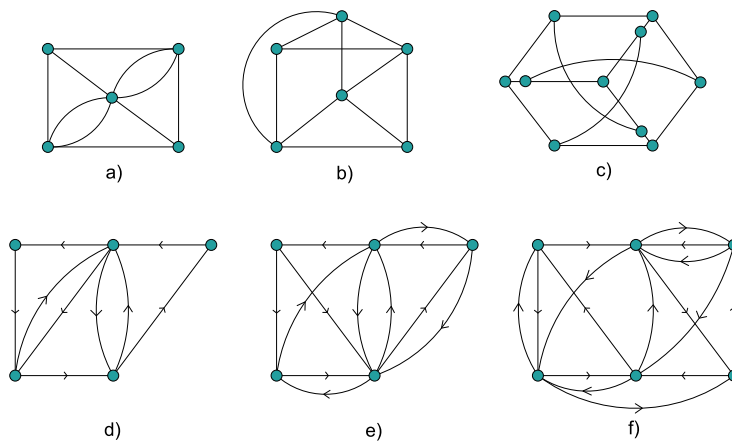


Figura 5.30: Grafos del apartado c) del Problema 5.8.24.

5.8.25 Sea G un grafo no orientado, simple, conexo, euleriano, con un número impar de vértices n y tal que su grafo complementario \overline{G} es conexo.

- a) Dado un vértice v de G de grado k , calcular el grado de v en \overline{G} .
- b) Demostrar que \overline{G} es euleriano.

5.8.26 Dar condiciones necesarias y suficientes para que un grafo contenga un camino que sea, simultáneamente, euleriano y hamiltoniano.

5.8.27 Se consideran las dos parejas de grafos de la Figura 5.31.

- a) Estudiar razonadamente si las parejas (a)-(b) y (c)-(d) son isomorfas.



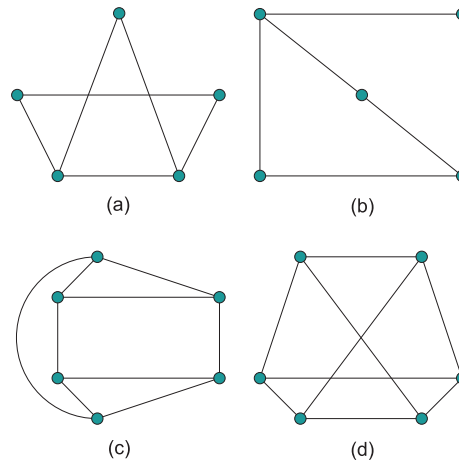


Figura 5.31: Grafos del Problema 5.8.27.

- b) Estudiar si el grafo (a) contiene un camino euleriano, y si el grafo (d) contiene un camino hamiltoniano. En cada caso, si la respuesta es afirmativa, encontrar ejemplos de dichos caminos.

5.8.28 Si un grafo es euleriano, ¿lo son también sus grafos homeomorfos? Responder a la misma pregunta si el grafo es hamiltoniano en vez de euleriano.

5.8.29 En un tablero 3×3 se coloca un caballo en una esquina. Se quiere que el caballo recorra todos los escaques una y sólo una vez saltando según las reglas del ajedrez para esta figura. Dibujar el grafo que define la situación (los escaques son los vértices y el salto de caballo las aristas). ¿Es posible ese recorrido? Repetir el mismo ejercicio con un tablero 4×3 .

5.8.30 Sea G un grafo simple conexo con 7 vértices y 16 aristas, ¿puede ser plano?

5.8.31 Un grafo G tiene 13 aristas, 2 vértices de grado 5 y el resto de sus vértices tienen grado 4. ¿Cuántos vértices tiene G ? En el caso de que G sea simple y conexo, ¿puede ser plano?

5.8.32 Determinar cuáles de los grafos de la Figura 5.32 admiten una representación plana.

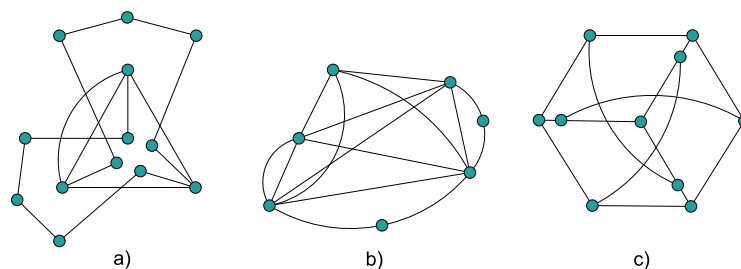


Figura 5.32: Grafos del Problema 5.8.32.

5.8.33

- a) Se tiene un grafo plano con 30 aristas de forma que divide el plano en 20 regiones. ¿Cuántos vértices tiene?
- b) Hay 20 puntos en el interior de un cuadrado. Estos puntos se unen entre ellos y con los vértices del cuadrado por segmentos de tal manera que el cuadrado queda dividido en triángulos y no se producen cortes de segmentos. Determinar el número de triángulos resultante.



5.8.34 Sea G un grafo con k componentes conexas ($k \geq 1$). Hallar una fórmula de Euler que relacione a , v , c y k .

5.8.35 Si G es un grafo simple bipartito plano, demostrar que $a \leq 2v - 4$ si $v \geq 3$.

5.8.36 ¿Cuál es el número cromático de los grafos $K_{m,n}$, C_n y R_n ?

5.8.37 Determinar el número cromático de los grafos de la Figura 5.33.

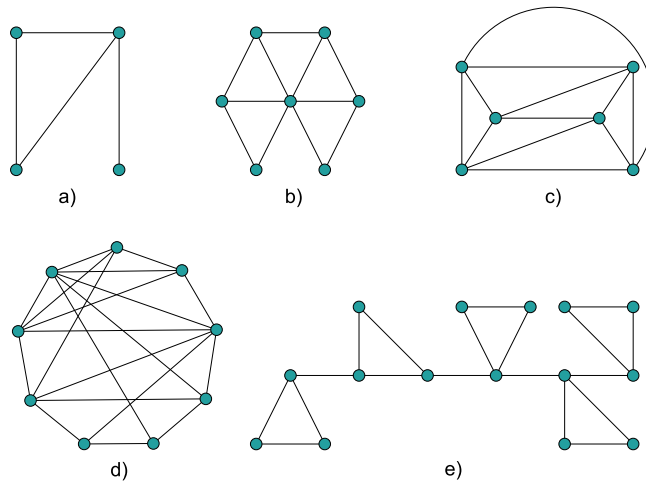


Figura 5.33: Grafos del Problema 5.8.37.

5.8.38 Dados los grafos de la Figura 5.34, decidir si son eulerianos y calcular su número cromático. ¿Son bipartitos?

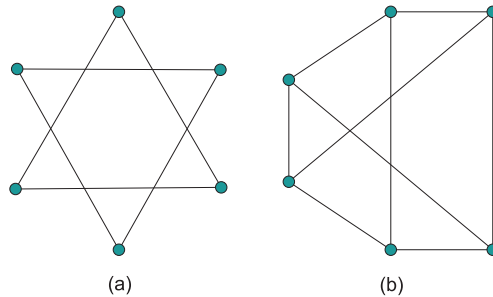


Figura 5.34: Grafos del Problema 5.8.38.

5.8.39 Se consideran los grafos G_1, G_2 y G_3 con matrices de adyacencia

$$M_1 = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}, M_2 = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix} \text{ y } M_3 = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 \end{pmatrix},$$

respectivamente.

- a) Representarlos gráficamente, y estudiar si son isomorfos dos a dos.
- b) Decidir razonadamente si G_1 es plano, si G_2 es euleriano, y calcular el número cromático de G_3 .



5.8.40 Se consideran los grafos de la Figura 5.35.

- Decidir cuáles de las parejas de grafos (a)-(b), (c)-(d) y (e)-(f) son isomorfas y cuáles no.
- Calcular los números cromáticos de los grafos (a) y (f).
- Decidir si son planos los grafos (b) y (e).

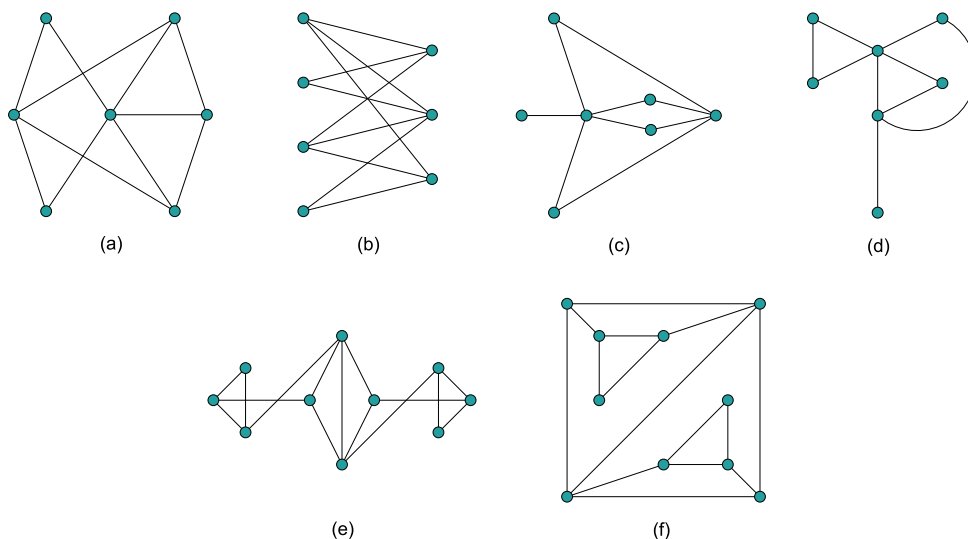


Figura 5.35: Grafos del Problema 5.8.40.

5.8.41 Se consideran los grafos de la Figura 5.36.

- Razonar cuáles de las parejas de grafos (a)-(b), (c)-(d) y (e)-(f) son isomorfas y cuáles no.
- ¿Son planos los grafos (e) y (b)?
- Escribir la matriz de adyacencia de (d). ¿Cuántos caminos hay de su vértice inferior izquierdo a su vértice superior izquierdo, que tengan longitud menor o igual que dos?

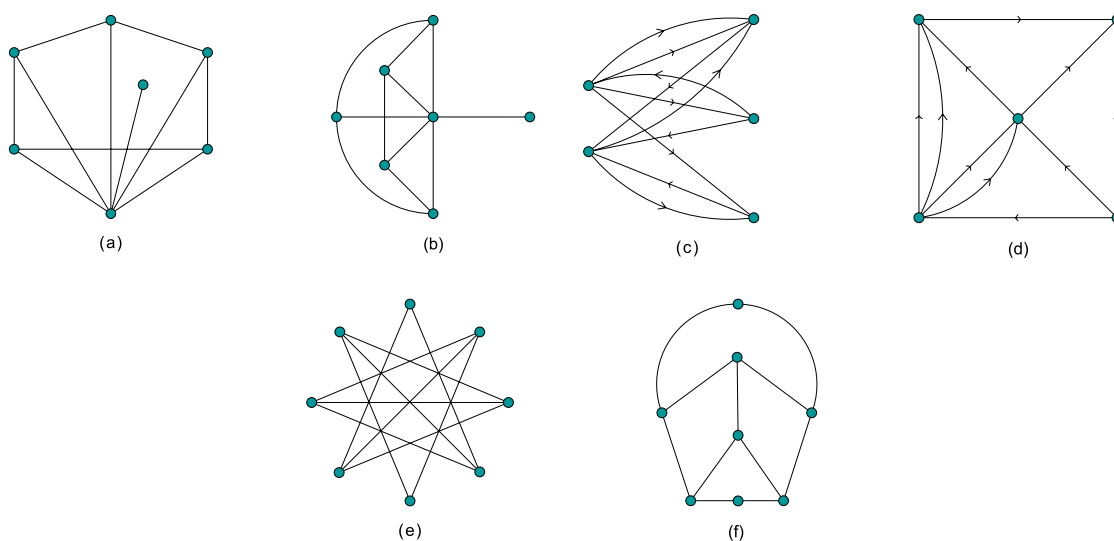


Figura 5.36: Grafos del Problema 5.8.41.



5.8.42 Se consideran las tres parejas de grafos de la Figura 5.37:

- a) Determinar cuáles de las parejas (a)-(b), (c)-(d) y (e)-(f) son isomorfas y cuáles no.
- b) Hallar el número cromático de los grafos (a) y (c).
- c) Determinar si los grafos (b) y (f) son planos.

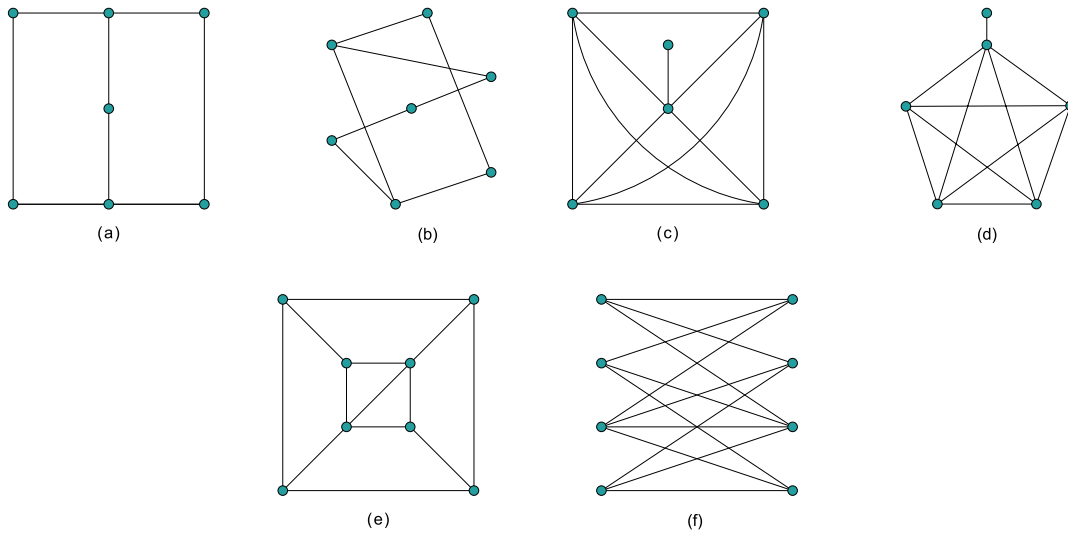


Figura 5.37: Grafos del Problema 5.8.42.

5.8.43 Se consideran los grafos de la Figura 5.38.

- a) ¿Cuáles de las parejas de grafos (a)-(b), (c)-(d) y (e)-(f) son isomorfas?
- b) Estudiar si (c) y (e) son eulerianos. En caso afirmativo, encontrar un camino euleriano.
- c) Decidir si los grafos (b) y (d) son planos.
- d) Calcular, razonadamente, el número cromático de (a) y (f).

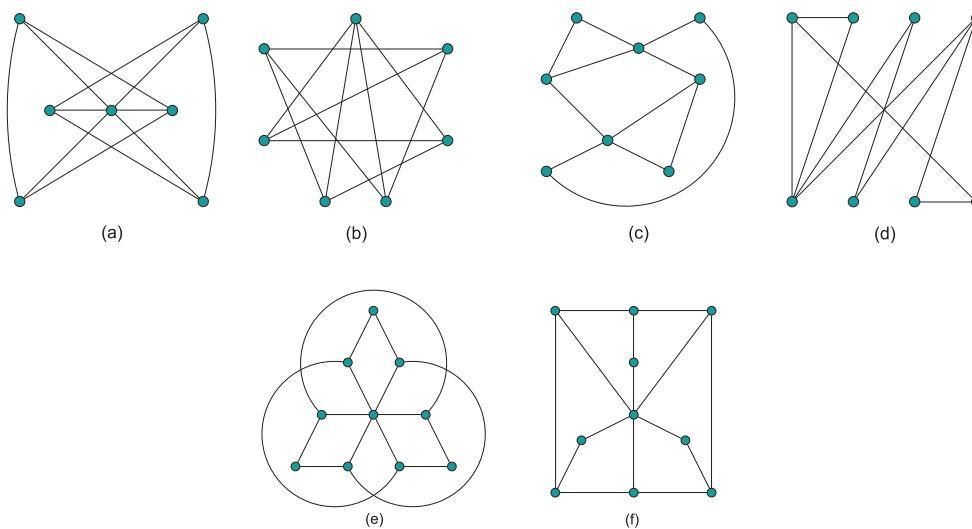


Figura 5.38: Grafos del Problema 5.8.43.

5.8.44 Dos estaciones de TV producirán interferencias si emiten a la misma frecuencia y están a distancia menor de 150 km. ¿Cuál es el mínimo número de frecuencias necesarias para la red de emisoras A, B, C, D, E y F si en la siguiente tabla se expresan todas las posibles distancias en km entre ellas?

	A	B	C	D	E	F
A	—	85	175	200	50	100
B	85	—	125	175	100	160
C	175	125	—	100	200	250
D	200	175	100	—	210	220
E	50	100	200	210	—	100
F	100	160	250	220	100	—

5.8.45 Una liga de fútbol está formada por 6 equipos. En la siguiente matriz se indican las victorias acaecidas entre los partidos jugados

$$A = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \end{pmatrix},$$

donde un 1 (resp. 0) en la fila i columna j con $i \neq j$ significa una victoria (resp. derrota) del equipo i en su casa frente al equipo j . Así, por ejemplo, el cuarto equipo ha vencido en su casa a los equipos 1, 2, 3 y 5 (los unos de la cuarta fila) y ha vencido fuera de su casa a los equipos 2, 3, 5 y 6 (los ceros de la cuarta columna, salvo el cuarto). ¿Cuántos partidos ha ganado, en total, cada equipo? Calcular, utilizando el ordenador, los autovalores y un autovector con entradas positivas asociado al autovalor dominante de A . ¿Cuál será el mejor equipo de acuerdo con el *ranking system* usado por *Google*?





Bibliografía

- [1] M. A. Armstrong: *Groups and symmetry*. Springer–Verlag, Nueva York, 1988.
- [2] F. Ayres: *Teoría y problemas de trigonometría plana y esférica*. McGraw–Hill, 1976.
- [3] M. Berrocoso, M. E. Ramírez, J. M. Enríquez–Salamanca y A. Pérez–Peña: *Notas y apuntes de trigonometría esférica y astronomía de posición*. Universidad de Cádiz, 2003.
- [4] T. S. Blyth y E. F. Robertson: *Sets and mappings*. Essential Student Algebra, vol. 1, Chapman and Hall, Londres y Nueva York, 1986.
- [5] J. Casey: *A treatise of spherical trigonometry and its application to Geodesy and Astronomy with numerous examples*. Merchant Books, 2007.
- [6] R. L. Devaney: *An Introduction to Chaotic Dynamical Systems*. Addison-Wesley, 1989.
- [7] F. J. Cirre: *Matemática discreta*. Base Universitaria, Editorial Anaya, Madrid, 2004.
- [8] P. Fernández: El secreto de Google y el Álgebra lineal. *Boletín de la Sociedad Española de Matemática Aplicada*, 30 (2004), 115–141. Un enlace a este trabajo puede encontrarse en la página personal del autor: http://www.uam.es/personal_pdi/ciencias/gallardo/index.htm
- [9] T. Gowers: *Matemáticas. Una breve introducción*. Alianza Editorial, Madrid, 2008.
- [10] B. Grünbaum y G. C. Shephard: *Tilings and patterns*. W. H. Freeman and Company, Nueva York, 1987.
- [11] J. A. Infante y J. M. Rey: Software *Sisdin* para el estudio y visualización de sistemas dinámicos discretos. http://www.mat.ucm.es/momat/software/Sisdin_setup_win32.exe.
- [12] M. A. Martín, M. Morán y M. Reyes: *Iniciación al caos*. Educación matemática en secundaria, Editorial Síntesis, Madrid, 1995.
- [13] G. Rodríguez y A. J. Gil: *Problemas de Astronomía*. Editorial Complutense, Madrid, 1993.
- [14] K. H. Rosen: *Matemática discreta y sus aplicaciones*. McGraw–Hill, Madrid, 2004.
- [15] J. T. Sandefur: *Discrete Dynamical Systems. Theory and Applications*. Clarendon Press, Oxford, 1990.
- [16] I. Stewart: *¿Juega Dios a los dados?* Crítica, 1997.

