POLITEXT 99

Matemática discreta

POLITEXT

Francesc Comellas - Josep Fàbrega
Anna Sànchez - Oriol Serra

Matemática discreta

EDICIONS UPC

La presente obra fue galardonada en el segundo concurso
"Ajuts a l'elaboració de material docent" convocado por la UPC.

Traducción al castellano de la obra original en catalán
Matemàtica discreta, realizada por Gabriel Valiente.

Primera edición: febrero de 2001

Diseño de la cubierta: Manuel Andreu

© Edicions UPC, 2001
Edicions de la Universitat Politècnica de Catalunya, SL
Jordi Girona Salgado 31, 08034 Barcelona
Tel.: 934 016 883 Fax: 934 015 885
Edicions Virtuals: www.edicionsupc.es
E-mail: edicions-upc@upc.es

Producción: Grup Artyplan-Artimpres S. A.
Agricultura 21, Nave 5, 08980 Sant Feliu de Ll. (Barcelona)

Depósito legal: B-7.069-2001
ISBN: 84-8301-456-4

Índice General

Prólogo **iv**

1 Algoritmos **1**
1.1 Introducción 1
1.2 Algoritmos y máquina de Turing 4
1.3 Lenguaje algorítmico 7
1.4 Análisis de algoritmos 9
1.5 Comparación de algoritmos 15
1.6 Clasificación de algoritmos 20

Enumeración **25**

2 Combinaciones y permutaciones **27**
2.1 Selecciones ordenadas y no ordenadas 27
2.2 Algunos ejemplos de aplicación 32
2.3 Propiedades de los coeficientes binomiales 38

3 Principios básicos de enumeración **49**
3.1 Cardinales de conjuntos 50
3.2 Principio de inclusión-exclusión 51
3.3 Biyecciones. Números de Catalan. Particiones 58
3.4 El principio del palomar y el teorema de Ramsey 64

4 Funciones generadoras **73**
4.1 Ecuaciones de recurrencia 74
4.2 Funciones generadoras 78
4.3 Ecuaciones de recurrencia lineales 84
4.4 Números combinatorios 89

Teoría de grafos **101**

5 Grafos y digrafos **103**
5.1 Definiciones básicas . . . 103
5.2 Caminos, conectividad y distancia . . . 106
5.3 Operaciones entre grafos . . . 108
5.4 Digrafos . . . 110
5.5 Representación matricial . . . 111
5.6 Grafos y redes de interconexión . . . 114
5.7 Planaridad: la fórmula de Euler . . . 115
5.8 Caracterización de los grafos planares . . . 117

6 Árboles **123**
6.1 Árboles . . . 124
6.2 Árboles generadores . . . 127
6.3 Número de árboles generadores . . . 128
6.4 Obtención de todos los árboles generadores . . . 132
6.5 Árboles generadores de coste mínimo . . . 133

7 Circuitos y ciclos **141**
7.1 Grafos eulerianos . . . 141
7.2 Ciclos hamiltonianos . . . 149
7.3 Ciclos fundamentales . . . 156
7.4 Análisis de redes eléctricas . . . 162

8 Flujos, conectividad y apareamientos **173**
8.1 Redes de transporte . . . 174
8.2 El teorema del flujo máximo–corte mínimo . . . 177
8.3 Conectividad . . . 180
8.4 Los teoremas de Menger . . . 181
8.5 Apareamientos en grafos bipartitos . . . 184
8.6 El teorema de Hall . . . 185

Estructuras algebraicas **191**

9 Introducción a las estructuras algebraicas **193**
9.1 Relaciones . . . 193
9.2 Aplicaciones . . . 199

9.3 Operaciones 203
9.4 Estructuras algebraicas 208

10 Grupos **215**
10.1 Definiciones y propiedades 216
10.2 Grupos abelianos finitos 224
10.3 Grupos de permutaciones 229
10.4 Digrafos de Cayley 240
10.5 Enumeración de Pólya 244

11 Anillos y cuerpos **257**
11.1 Definiciones y propiedades 257
11.2 El anillo de los polinomios 266
11.3 Cuerpos finitos 275

12 Estructuras combinatorias **289**
12.1 Diseños combinatorios 289
12.2 Geometrías finitas 305
12.3 Cuadrados latinos 314

Índice de materias **331**

Prólogo

La *matemática discreta* es una rama de las matemáticas que trata las estructuras finitas y numerables. Esta definición, forzosamente imprecisa, queda mejor delimitada cuando se da una descripción de sus contenidos. A grandes rasgos, las lineas básicas de las que se ocupa la matemática discreta son las técnicas de enumeración, las estructuras combinatorias, la teoría de grafos y las estructuras algebraicas. Asimismo, la algorítmica es una herramienta imprescindible para la construcción de soluciones a los problemas que se tratan.

Aunque históricamente éstas eran áreas que no formaban un cuerpo estructurado, el progreso de la informática y de las técnicas de computación les ha dado un impulso decisivo y las ha convertido en una de las ramas de la matemática aplicada con más vitalidad.

Este impulso ha influido también en el diseño de los *curricula* en las enseñanzas de ingeniería y matemáticas alrededor del mundo. En este sentido, en nuestro país, la implantación de nuevos planes de estudio y la reforma de los existentes hace que la matemática discreta haya sido introducida como un elemento importante de la formación básica.

El libro de texto que se propone ha sido pensado para servir de soporte a cursos básicos de matemática discreta. Así, los conocimientos de matemáticas que se presuponen en el lector son los que corresponden a unos primeros cursos universitarios de álgebra y cálculo. El texto contiene material más que suficiente para cubrir dos cuatrimestres lectivos, y facilita así una cierta flexibilidad en la elección de los temas a explicar. Desde el punto de vista pedagógico se ha hecho un esfuerzo especial para presentar los temas de una forma simple pero rigurosa. Como cualquier texto de matemáticas, los problemas al final de los capítulos y los ejercicios insertados en el texto constituyen un elemento importante del libro.

El contenido del libro se estructura en un capítulo inicial sobre algorítmica seguido de tres partes dedicadas a la enumeración, la teoría de grafos y las estructuras algebraicas discretas.

En el capítulo inicial se introducen las nociones básicas de recursividad, lenguajes algorítmicos y complejidad de algoritmos. En la primera parte, se hace un repaso de la combinatoria elemental, se discuten principios básicos de enumeración y se presentan técnicas de enumeración más elaboradas basadas en las funciones generadoras y las ecuaciones de recurrencia. Paralelamente, se van introduciendo también algunos temas clásicos de combinatoria como,

por ejemplo, las particiones de conjuntos y de enteros, desarreglos o la teoría de Ramsey, entre otros.

La segunda parte presenta los temas básicos de la teoría de grafos. Se introducen en primer lugar los elementos básicos de la teoría y la terminología. A continuación se estudian los árboles, en cierto sentido la clase más simple de grafos, a pesar de tener numerosas aplicaciones en áreas muy diversas. En particular, se trata también la obtención de árboles generadores de coste mínimo, que constituye un problema clásico en investigación operativa. Sigue el estudio de la estructura cíclica de un grafo y su aplicación al análisis de redes eléctricas. También se tratan los problemas clásicos de existencia de circuitos eulerianos y ciclos hamiltonianos, y su relación con ciertos problemas de optimización combinatoria como pueden ser los problemas del viajante o del cartero chino. El último capítulo de esta parte estudia tres temas aparentemente no relacionados, pero que resultan estar estrechamente ligados: flujos en redes de transporte, conectividad de grafos y apareamientos en grafos bipartitos. Todos ellos tienen numerosas aplicaciones a problemas de optimización y de asignación y en el diseño de redes de interconexión.

Finalmente, la última parte del libro está dedicada a estudiar las estructuras algebraicas discretas. Después de introducir las operaciones binarias y sus propiedades, se presentan los conceptos básicos de la teoría de grupos. Se describen las propiedades más significativas de los grupos cíclicos y de los grupos de permutaciones, y se dedica una atención especial a la representación de grupos per medio de los grafos de Cayley. A continuación se tratan estructuras algebraicas definidas a partir de dos operaciones: anillos y cuerpos. En particular, se estudia el anillo de polinomios y se aplica a la construcción de cuerpos finitos. Las estructuras combinatorias estudian de manera sistemática las relaciones de incidencia entre determinados objetos y ciertos subconjuntos de estos objetos. En el último capítulo se introducen los diseños combinatorios como modelos generales de estas estructuras. En particular, se introducen las llamadas *geometrías finitas* y se particulariza en el estudio de planes afines y proyectivos finitos. El capítulo se acaba con el estudio de cuadrados latinos y la construcción de conjuntos de cuadrados latinos mutuamente ortogonales.

Los autores quieren agradecer especialmente el interés con que el profesor José Luis Andrés Yebra ha revisado el manuscrito de este libro. Sus correcciones y sugerencias han sido una ayuda muy valiosa. En este sentido, queremos manifestar también nuestro agradecimiento al profesor Miquel Àngel Fiol Mora, así como a Javier Ozón Górriz.

Agradecemos también el soporte institucional de la Universitat Politècnica de Catalunya y la confianza depositada en el proyecto de este libro que se ha manifestado en el otorgamiento de una ayuda para su elaboración.

Los autores
Barcelona, 23 de marzo de 1994

Índice de Materias

1–factor, 190

adyacente
    desde, 110
    hacia, 110
Albertson, algoritmo de, 153
algoritmo, 2, 5
    de burbujas, 17
    de inserción, 16
    de inserción mínima, 154
    genético, 156
    recursivo, 9
anillo, 210, 257
    íntegro, 259
    abeliano, 258
    centro, 285
    cociente, 262
    de Boole, 284
    euclídeo, 270
    principal, 263
    producto cartesiano, 285
    unitario, 258
anulador por la izquierda, 285
apareamiento, 184
    completo, 185
    perfecto, 184
aplicación, 199
    biyectiva, 201
    exhaustiva, 201
    inyectiva, 201
árbol, 124
    binario, 124
    de decisión, 124
    generador, 127
    generador de coste mínimo, 133
arborescencia, 139
arco, 110
arista, 103
arista-conectividad, 180
aristas
    independientes, 103, 184
    paralelas, 104
automorfismo, 213
    de un grafo, 120
autovalores de un grafo, 121

Bézout, identidad de, 271
BIBD, 298
Binet–Cauchy, teorema de, 130
biyecciones, 58
bloque, 189, 289
    de un grafo, 189
Bose, teorema de, 318
bosque, 126
Bruch–Ryser–Chowla, teorema de, 304
Burnside, lema de, 246

camino, 106
    hamiltoniano, 150
caminos
    internamente disyuntos, 181

capacidad, 174
característica, 259
cardinales de conjuntos, 50
Cauchy, teorema de, 229
Cayley, fórmula de, 131
Cayley, teorema de, 235
centro, 107, 285
Church–Turing, hipótesis de, 6
ciclo, 106
    fundamental, 158, 161
    hamiltoniano, 150
circuito, 106
    euleriano, 142
    euleriano dirigido, 147
clase de equivalencia, 197
clases laterales, 219
cociclo, 160
    fundamental, 160
coeficiente, 266
coeficientes binomiales, 31, 38, 75, 80
    propiedad de la adición, 39
coeficientes multinomiales, 43
combinaciones, 27
    con repetición, 28
    sin repetición, 28
complejidad
    espacial, 10
    temporal, 10
complemento de un grafo, 109
componentes, 107
composición de aplicaciones, 202
conectividad, 180
conexo
    débilmente, 111
    fuertemente, 111
    unilateralmente, 111
congruencia módulo *n*, 197
conjunto
    cociente, 199
    separador, 181
contracción, 296
    de una arista, 131
correspondencia, 194
corrientes de ciclo, 164
corte
    s–t corte, 175
    mínimo, 177
    simple, 160
coste, 133
cuadrados latinos, 314
    ortogonales, 316
cuerda, 127
cuerpo, 210, 260

de Bruijn
    digrafo de, 148
    secuencia de, 148
defecto de un grafo bipartito, 185
desarreglos, 56, 75, 89
descomposición de un grafo en subgrafos, 146
diámetro, 107
diferencia simétrica, 284
digrafo, 110
    de Cayley, 242
    hamiltoniano, 152
    línea, 122
    simétrico, 111
    simétrico asociado a un grafo, 111
Dirac, teorema de, 152
Dirichlet, principio de, 64
diseño, 289
    complementario, 292
    derivado, 303
    dual, 291

incompleto, 292
isomorfo, 290
regular, 292
residual, 303
s-derivado, 296
s-residual, 297
simétrico, 301
simple, 290
uniforme, 292
distancia, 106
media, 107
divisores de cero, 258
dominio, 200

ecuaciones de recurrencia, 74
lineales, 84
elemento
inverso, 207
neutro, 207
primitivo, 281
endomorfismo, 213
epimorfismo, 212
de anillos, 264
equilibrado, 298
Eratóstenes, 53
Erdős–Szekeres, teorema de, 65
espacio, 9
estabilizador, 245
estructura
algebraica, 208
cociente, 213
Euclides
algoritmo de, 4, 12, 271
teorema de, 270
Euler
fórmula de, 116
función ϕ de, 55
excentricidad, 107

fórmula del binomio, 35, 80
fórmula multinomial, 35
factorial, 29
Ferrers, diagramas de, 62
Fleury, algoritmo de, 168
flujo, 174
máximo, 177
neto entrante, 174
neto saliente, 174
valor del, 175
Ford–Fulkerson
algoritmo de, 179
teorema de, 177
función polinómica, 274
funciones generadoras, 78
exponenciales, 83
ordinarias, 78

geometría lineal finita, 305
giro, 121
grado, 105, 266
de entrada, 110
de salida, 110
máximo, 105
mímino, 105
secuencia de grados, 120
grafo, 103
d–regular, 105
k–arista conexo, 180
k–conexo, 180
r–partito, 109
bipartito, 109
bipartito completo, 109
camino, 108
ciclo, 108
completo, 108
conexo, 106
de Cayley, 242

dirigido, 110
euleriano, 142
hamiltoniano, 150
hipohamiltoniano, 169
nulo, 109
planar, 115
subyacente, 111
trivial, 107
vértice-simétrico, 121
vértice-transitivo, 121
grupo, 209, 216
alternado, 238
cíclico, 224, 225
cociente, 220
de los cuaternones, 242
de permutaciones, 230
diédrico, 231
isomorfo, 221
presentación, 240
producto cartesiano, 222
relaciones, 241
simétrico, 230

Hall, teorema de, 186
hipercubo, 110, 253
homomorfismo, 263
canónico, 264
de grupos, 221

ideal, 261
bilateral, 261
maximal, 265
principal, 262
índice de ciclos, 248
intersección de grafos, 109
invariante de un grafo, 105
isomorfismo, 212
de anillos, 264
de grafos, 105

Königsberg, problema de los puentes de, 141
Kirchoff, leyes de, 163
Kuratowski, teorema de, 119

Lagrange, teorema de, 220
lazos, 104
lista de incidencia, 114

máquina de Turing, 4
determinista, 5
no determinista, 21
matriz
de adyacencia, 112
de ciclos fundamentales, 159
de grados, 139
de impedancias, 164
de impedancias de ciclo, 165
de incidencia, 113, 290
de incidencia reducida, 128
Menger, teorema de, 182, 183
mergesort, 17
Meyniel, teorema de, 153
MOLS, 317
monoide, 209
monomorfismo, 212
de anillos, 264
morfismo, 212
de anillos, 263
MTND, 21
multigrafo, 104
euleriano, 142
mutación, 156

número ciclomático, 158
números
combinatorios, 89

de Bell, 92, 95
de Catalan, 58, 59, 76, 90
de Fibonacci, 77
de Ramsey, 67
de Stirling, 92
de primer tipo, 96, 253
de segundo tipo, 92
piramidales, 42
triangulares, 41
NP-C, 21

operación binaria, 203
asociativa, 206
conmutativa, 206
distributiva, 206
órbita, 245
orden, 103, 225, 304, 309
Ore, teorema de, 151

palabras de alfabetos, 32
particiones, 58, 77, 91
conjugadas, 62
de conjuntos, 92
de un entero, 61
Pascal, triángulo de, 39
permutación
ciclo, 233
signatura, 238
transposición, 235
permutaciones, 27
con repetición, 28
sin repetición, 28
Petersen, grafo de, 120
plano
afín, 312
proyectivo, 306
población, 156
polinomio, 266
característico, 86
divisor, 268
irreductible, 269
mónico, 266
mcd, 269
mcm, 270
primo, 268, 269
producto, 266
suma, 266
polinomios coprimos, 270
Prüfer, secuencia de, 138
principio
de adición, 50
de dualidad, 307
de inclusión-exclusión, 51
criba, 69
del palomar, 64
problema
(Δ, D), 115
de los matrimonios, 186
del cartero chino, 168
del conector, 133
del viajante, 154
tipo NP, 21
tipo P, 20
procedimiento, 9
producto cartesiano, 193
de grafos, 109
producto directo, 223
proporción áurea, 88
puente, 108

quicksort, 19

raíz, 274
multiplicidad, 275
radio, 107
Ramsey, teorema de, 64

recorrido, 106, 200
    euleriano, 142
recursividad, 9
red
    de interconexión, 115
    de transporte, 174
región, 116
relació
    de equivalencia, 197
relación, 194
    binaria, 194
    de orden, 196
    inversa, 200

secuencia de aumento, 177
semigrupo, 209
series formales, 78
simulated annealing, 156
sistema de representantes diferentes, 187
Steiner, sistema de, 298
subanillo, 260
subdivisión
    de un grafo, 118
    de una arista, 118
subespacio
    de ciclos, 158
    de cociclos, 160
subestructura, 211
subgrafo, 104
    generador, 104
    inducido, 104
subgrupo, 218
    índice, 219
    generado, 225
    normal, 220
    propio, 219
    trivial, 219
suma binaria de grafos, 110
suma de grafos, 109

t-diseño, 294
tamaño, 103
teoría de grafos, 101
teorema
    de factorización, 269
    del flujo máximo–corte mínimo, 177
tiempo, 9
torneo, 153
torres de Hanoi, 13
transformación elemental, 132
transversal, 187
Tutte, teorema de, 152

unión de grafos, 109

vértice, 103
    de corte, 107
vértices
    adyacentes, 103
    apareados, 184
    independientes, 103
vértices y aristas incidentes, 103
variedades, 289
vectores ortogonales, en grafos, 158

Whitney, teorema de, 184

Capítulo 1

Algoritmos

1. Introducción
2. Algoritmos y máquina de Turing
3. Lenguaje algorítmico
4. Análisis de algoritmos
5. Comparación de algoritmos
6. Clasificación de algoritmos

El objetivo de este capítulo es proporcionar el marco formal y las herramientas básicas que permitan la descripción y el estudio de los algoritmos que se presentarán al largo de este libro. Después de una introducción general al concepto de algoritmo, este se precisa en la sección 2 dentro del modelo de cómputo de la máquina de Turing en el que se consideran los algoritmos. A continuación se presenta la notación o lenguaje algorítmico que se usará, y en la sección 4 se dan los fundamentos del análisis de algoritmos, que nos permitirá realizar la comparación entre diferentes algoritmos de resolución de un mismo problema (sección 5) e introducir la clasificación general de algoritmos de la sección 6, objetivo fundamental del capítulo.

1.1 Introducción

El concepto general de algoritmo es seguramente conocido por el lector. El Diccionario de la Lengua Española de la Real Academia Española define el término *algoritmo* como "conjunto ordenado y finito de operaciones que permite hallar la solución de un problema". Básicamente un algoritmo es, entonces, una descripción de cómo se realiza una tarea: la resolución de un problema concreto. Por procedimiento entendemos una secuencia de instrucciones tales que hechas en el orden adecuado llevan al resultado deseado. Otra característica que se desprende

de la definición es que un mismo algoritmo puede resolver todos los problemas de una misma clase. En un sentido general, la noción de algoritmo no es exclusiva de la matemática. También tienen una relación muy próxima al algoritmo una receta de cocina, una partitura musical, una guía turística de una ciudad o un museo, un manual de instrucciones de un electrodoméstico, etc.

El primer punto que se debe considerar es la identificación de los problemas que resolverá el algoritmo. Consideraremos problemas bien definidos y adecuados para ser solucionados mediante el uso de computadores digitales. También se debe tratar de problemas que pueden caracterizarse por su dimensión. Para cada tamaño puede haber diferentes *instancias* del problema. Por ejemplo, si el problema considerado es la ordenación de un conjunto de elementos, la dimensión del problema sería la cardinalidad del conjunto y para una dimensión fijada se pueden dar instancias diferentes que corresponderán al conjunto concreto de elementos que se considere.

El algoritmo debe tener en cuenta el *modelo* abstracto dentro del cual tratemos el problema. No usaremos el mismo modelo si tenemos que ordenar un conjunto de libros por temas que si queremos ordenar un conjunto de números naturales. El modelo, de hecho, es un inventario de las herramientas a nuestra disposición para tratar el problema y una descripción de la manera de usarlas.

Para la solución del problema se tratará de codificar la instancia considerada para que constituya la entrada del algoritmo. El algoritmo produce una salida que, una vez decodificada, es la solución del problema. Solucionar el problema quiere decir ser capaces de realizar este proceso para cualquier tamaño e instancia del problema.

Habitualmente, los algoritmos se escriben pensando en el ejecutor o procesador del algoritmo (máquina o persona). El procesador debe ser capaz de interpretarlo y de ejecutarlo. Si se trata de una máquina, será preciso que esté en su lenguaje específico. De esta manera hay una serie de pasos intermedios entre la descripción del algoritmo y el programa final que lo ejecuta; véase la figura 1.1.

Una característica importante deseable para un algoritmo es su transportabilidad, consecuencia de su generalidad, la cual debe permitir la adaptación del algoritmo para que pueda ser ejecutado por máquinas muy diferentes.

Podemos ahora precisar un poco más el concepto de algoritmo: Además de estar constituido por una secuencia de operaciones que llevan a la resolución de un tipo concreto de problemas dentro de un modelo prefijado, en un algoritmo es preciso que se verifique:

Existencia de un conjunto de entradas. Debe existir un conjunto específico de objetos cada uno de los cuales son los datos iniciales de un caso particular del problema que resuelve el algoritmo. Este conjunto se llama conjunto de entradas del problema.

Definibilidad. Cada paso debe ser definible de forma precisa y sin ninguna ambigüedad.

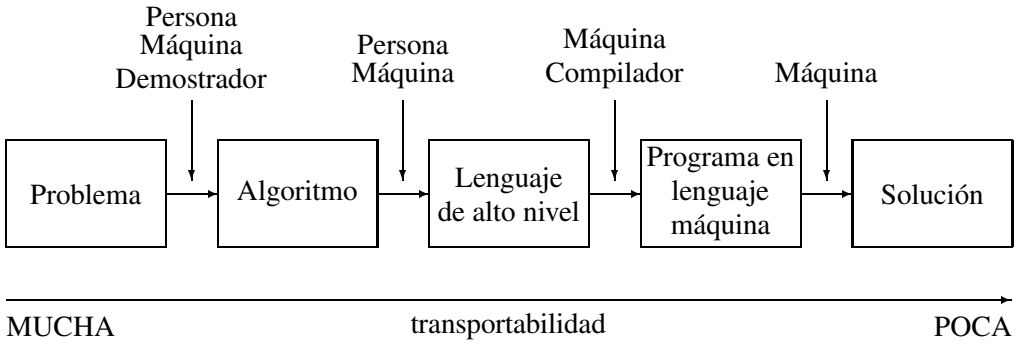


Figura 1.1: Proceso de solución de un problema mediante un algoritmo

Efectividad. Todas las operaciones a realizar en el algoritmo han de ser suficientemente básicas para que se puedan hacer en un tiempo finito en el procesador que ejecuta el algoritmo.

Finitud. El algoritmo debe acabar en un número finito de pasos. Un *método de cálculo* puede no tener esta restricción.

Corrección. El algoritmo debe ser capaz de encontrar la respuesta correcta al problema planteado.

Predecibilidad. Siempre consigue el mismo resultado para un mismo conjunto de entradas.

Existencia de un conjunto de salidas. El resultado del algoritmo asociado al conjunto de entradas.

También se consideran características importantes de un algoritmo la *claridad* y la *concisión*: Un algoritmo claro y breve resulta más sencillo de programar, a la vez que es más sencillo comprobar si es correcto.

Con esta definición comprobamos que una receta de cocina no encaja totalmente con la noción de algoritmo. La receta tiene un conjunto de entradas (ingredientes) y de salidas (el plato guisado); las instrucciones pueden, en principio, hacerse efectivas (se supone que se dispone de los utensilios convenientes—el *hardware*—y que el cocinero no es torpe); se verifica la finitud, pero falla la definibilidad, ya que son ambigüas frases típicas de una receta como 'añadir una *pizca* de sal', 'revolver *lentamente*', o bien, '*esperar* que se haya reducido el líquido', y también falla su predecibilidad.

Un mismo problema se puede resolver usando algoritmos diferentes. Los algoritmos podrán tener una complejidad diferente, la cual afectará al tiempo de ejecución y a la ocupación

de memoria, y posiblemente se distinguirán también por la exactitud de los resultados a que llevan.

Consideremos, por ejemplo, el cálculo del máximo común divisor de dos enteros. Recordemos que el máximo común divisor, mcd, de dos enteros positivos es el entero positivo más grande que los divide. Por ejemplo $\mathrm{mcd}(225,945) = 45$. Si alguno de los enteros es cero, el mcd se define como el otro entero. Así $\mathrm{mcd}(14,0) = 14$ y $\mathrm{mcd}(0,0) = 0$.

Un primer procedimiento para encontrarlo puede consistir en descomponer cada uno de los números en primos y considerar el producto de las potencias más bajas de cada primo común. En el caso de 180 y 380 escribiremos $180 = 2^2 \cdot 3^2 \cdot 5$ y $380 = 2^2 \cdot 5 \cdot 19$ y por tanto el $\mathrm{mcd}(180,380) = 2^2 \cdot 5 = 20$. Este método, si bien es el que se suele enseñar en las escuelas, es muy ineficiente cuando los números considerados son grandes (más de 5 dígitos), dado que la descomposición en primos es difícil. El mismo problema puede ser resuelto, como estudiaremos en la sección 4, usando el clásico algoritmo de Euclides, que consiste en dividir el número más grande por el más pequeño y retener el resto. De forma sucesiva se hace el cociente del divisor y el resto anteriores hasta que el resto sea cero. En este caso, el último resto calculado diferente de cero es el máximo común divisor de los dos números iniciales. Aplicándolo al mismo ejemplo de antes: 380 dividido por 180 tiene por resto 20. 180 dividido por 20 tiene resto cero. Así, el máximo común divisor de 380 y 180 es 20. Más adelante entraremos en detalle sobre el análisis de algoritmos, el objetivo del cual es precisamente el estudio y la comparación de algoritmos.

Hay otros puntos en relación a la algorítmica que son importantes pero en los cuales no entraremos. Mencionemos en primer lugar la cuestión del diseño de algoritmos: ¿Existen algoritmos para diseñar algoritmos? ¿Todo proceso tiene un algoritmo que lo describe? Este tipo de cuestiones lleva a un área importante: la teoría de la computabilidad. Otro aspecto es la comprobación de lo que realmente hace un algoritmo. Métodos como la especificación formal (por ejemplo, el *Vienna Development Method* o **Z**) son útiles para este tipo de estudios.

1.2 Algoritmos y máquina de Turing

La máquina de Turing no es un objeto físico, sino un artificio matemático que nos proporciona un modelo de computación en el cual encuadramos el análisis de nuestros algoritmos. Es preciso decir que hay otros modelos igualmente válidos, pero no tan comunes en la literatura. Destacamos las *máquines de acceso aleatorio*, mucho más realistas que la de Turing en el sentido que emulan en su estructura un ordenador real, véase por ejemplo [2] o [6]. Todos los modelos son equivalentes, pero la máquina de Turing tiene un carácter más elemental que hace más sencillo comprender su aplicación en el estudio de algoritmos.

Una *máquina de Turing* consiste en un cabezal de lectura/escritura por el cual pasa una

cinta infinita que puede moverse hacia adelante y hacia atrás. La cinta se encuentra dividida en casillas que pueden estar vacías o llevar un símbolo de un determinado alfabeto. De hecho, con un solo símbolo (además de la posibilidad de dejar la casilla vacía) es suficiente, ya que cualquier otro alfabeto lo podríamos expresar en secuencias de este símbolo y casillas vacías. En nuestro ejemplo, figura 1.2, se usa un solo símbolo (un cuadrado negro).

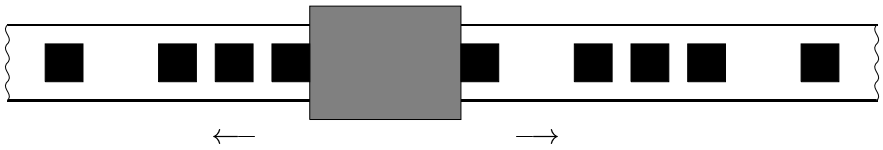


Figura 1.2: Máquina de Turing

El cabezal se encuentra en cada instante en un estado concreto de entre un número finito de estados posibles diferentes. La máquina funciona paso a paso y cada uno de los pasos consiste en situar el cabezal delante de una casilla y, después de leer el contenido de la casilla, realizar, de acuerdo con el resultado de la lectura y el estado interno de la máquina, las tres acciones siguientes: borrar la casilla y dejarla vacía, o bien, escribir un símbolo que puede ser el mismo que había antes, pasar a un nuevo estado (que puede ser el mismo) y finalmente mover la cinta una casilla en una de las dos direcciones posibles, o bien, acabar el cómputo. La conducta global de la máquina viene determinada por un *conjunto de instrucciones*, las cuales indican, para cada posible estado y símbolo leído, las tres acciones que es preciso hacer. Si se han de dar datos iniciales, éstos se escriben en la cinta de acuerdo con el sistema de codificación considerado. El cabezal se sitúa en la casilla inicial y, una vez se acaba el cómputo, la máquina entra en un estado especial de *parada* y deja de funcionar. La posible respuesta se encontrará escrita en la misma cinta a partir de la posición donde se encuentre el cabezal. Cada instrucción se puede representar con una *quíntupla* $(e_a, s_a; e_d, s_d, m)$, donde e_a indica el estado de la máquina cuando se lee el símbolo s_a, y e_d es el estado de la máquina después de dejar escrito el símbolo s_d en la casilla procesada y mover la cinta a la derecha o la izquierda según indica m.

Ya que tanto el conjunto de estados como el de símbolos es finito, cualquier cómputo puede ser especificado completamente por un conjunto finito de quíntuplas. Supongamos también que el cómputo es determinístico en el sentido que, dados el estado de la máquina y el símbolo de la cinta antes de cualquier acción, existe una quíntupla que determina el nuevo estado de la máquina, el símbolo a escribir en la cinta y el movimiento que debe hacer. Se suele decir entonces que la máquina de Turing es *determinista*.

Gracias a la máquina de Turing, un algoritmo puede ser definido de forma precisa como

una secuencia de instrucciones que determina totalmente la conducta de la máquina para la resolución del problema considerado. La llamada *hipótesis de Church–Turing* dice que todo algoritmo puede ser descrito (o implantado) en una máquina de Turing.

Ejemplo 1.1. En este ejemplo, el alfabeto que se usará tiene un solo símbolo, el 1. Los enteros positivos se representan por una secuencia de unos. El número total de unos es el entero considerado (representación unaria). La máquina tiene cinco estados posibles etiquetados 0,1,2,3 y A (el estado especial de parada). El programa tiene por objeto determinar si un cierto entero es par o impar. Si es par, la máquina escribirá un 1 y se parará. Si es impar dejará la casilla vacía y también se parará. La salida del programa, 1 o vacío, se encontrará en la cinta a continuación del entero con una casilla vacía como separador. Se supone que el cabezal se encuentra, en el momento de iniciar la computación, sobre el primer dígito del entero, y que el resto de dígitos están a la derecha de éste. La figura 1.3 especifica las quíntuplas que forman el conjunto de instrucciones y la acción del programa. Si la entrada es el entero 4 (1111 en la notación unaria empleada), D y E expresan el movimiento de la cinta hacia la derecha o la izquierda, el símbolo $*$ indica la irrelevancia de expresar el contenido de la posición correspondiente de la quíntupla y el símbolo b indica que la casilla está en blanco. La posición del cabezal viene dada por la flecha.

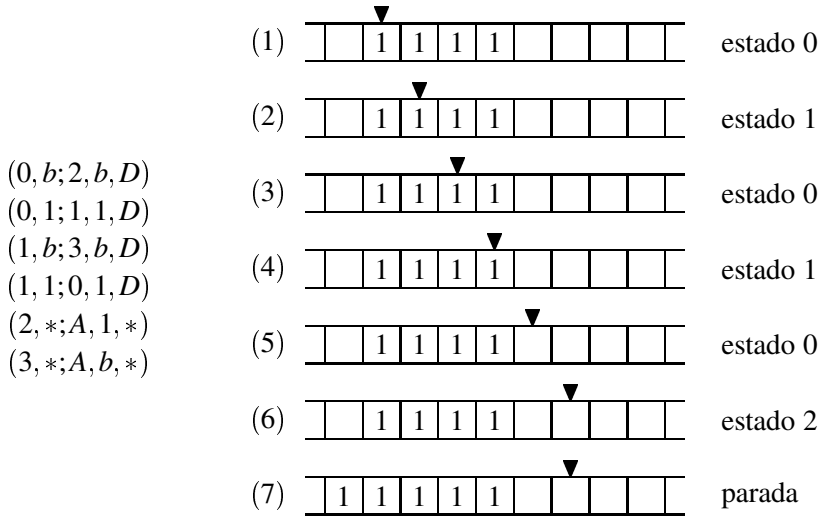


Figura 1.3: Conjunto de intrucciones y ejecución de un programa para determinar la paridad del entero 4 en una máquina de Turing

Ejemplo 1.2. El conjunto de instrucciones de la figura 1.4 describe una implantación del al-

goritmo de Euclides en una máquina de Turing. Se supone que los dos enteros están escritos en la cinta en representación unaria y separados por una casilla vacía. El cabezal se encuentra, en el momento de iniciar el cómputo, sobre esta casilla. Se deja al lector la comprobación de su funcionamiento.

```
(0,b;0,b,D)     (0,1;1,1,E)     (1,b;2,1,D)     (1,1;1,1,E)
(2,b;10,b,D)    (2,1;3,b,D)     (3,b;4,b,D)     (3,1;3,1,D)
(4,b;4,b,D)     (4,1;5,b,D)     (5,b;7,b,E)     (5,1;6,1,E)
(6,b;6,b,E)     (6,1;1,1,E)     (7,b;7,b,E)     (7,1;8,1,E)
(8,b;9,b,E)     (8,1;8,1,E)     (9,b;2,b,D)     (9,1;1,1,E)
(10,b;A,b,*)    (10,1;10,1,D)
```

Figura 1.4: Conjunto de intrucciones que describen la implantación del algoritmo de Euclides en una máquina de Turing

1.3 Lenguaje algorítmico

Un algoritmo puede ser expresado o formulado de maneras muy diferentes. Podemos usar únicamente el idioma habitual o emplear presentaciones gráficas más o menos complicadas e independientes de un lenguaje natural como, por ejemplo, un diagrama de flujos. Muy a menudo se usan metalenguajes que combinan una descripción en un lenguaje natural con unos símbolos o palabras clave que expresan algunas de las acciones bien definidas que es preciso efectuar.

De hecho, y pensando en el ejemplo de la receta de cocina, ésta equivaldría al algoritmo, pero podría ser escrita en catalán, francés o ruso (diferentes *lenguajes de 'programación'*), a pesar que el resultado obtenido, por ejemplo un pastel, sería equivalente fuera cual fuese el lenguaje empleado. Por otra parte, aunque en una receta de cocina no se acostumbra a usar un metalenguaje, sí que se suele presentar de forma estructurada ordenando los ingredientes uno bajo el otro, indicando claramente los diferentes pasos, etc.

En este libro usamos esencialmente el lenguaje natural con algunas palabras y símbolos que tienen el papel de metalenguaje. La indentación del texto también es importante y permite distinguir las diferentes partes y estructuras del algoritmo. La tabla 1.1 muestra algunos de los principales elementos de este metalenguaje.

Nos hemos decidido por esta aproximación, porque al mismo tiempo que permite dar una cierta estructura visual al algoritmo que facilita su lectura y su comprensión se evita la complejidad a que lleva muy a menudo un diagrama de flujos. Al mismo tiempo es suficientemente

Tabla 1.1: Principales elementos de la notación algorítmica que se usa en este libro

Asignación

variable $\leftarrow$ *expresión*

El valor de *expresión* pasa a ser el nuevo valor de *variable*.

Decisión

Si *condición* **entonces hacer** *instrucción*$_1$ **si no hacer** *instrucción*$_2$

Si es cierta la *condición* se ejecuta la *instrucción*$_1$, si no se ejecuta la *instrucción*$_2$.

Repetición

Hacer *instrucción*$_1 \ldots$ *instrucción*$_n$ **hasta que** *condición*

Hasta que sea cierta la *condición* se ejecutan todas las instrucciones desde *instrucción*$_1$ hasta *instrucción*$_n$.

También hay otras maneras de hacer la repetición:

Para *variable* $=$ *val*$_{inicial}$ **hasta** *val*$_{final}$ **hacer** *instrucción*$_1 \ldots$ *instrucción*$_n$.

Repetir *val*$_{final} - $ *val*$_{inicial}$ veces la ejecución de todas las instrucciones desde *instrucción*$_1$ hasta *instrucción*$_n$.

Repetir hasta *condición instrucción*$_1 \ldots$ *instrucción*$_n$

Repetir, hasta que sea cierta *condición*, todas las instrucciones desde *instrucción*$_1$ hasta *instrucción*$_n$.

Mientras *condición* **hacer** *instrucción*$_1 \ldots$ *instrucción*$_n$

Repetir, mientras sea cierta *condición*, todas las instrucciones desde *instrucción*$_1$ hasta *instrucción*$_n$.

concreto para poder ser ejecutado con facilidad. Todos los algoritmos que se describen en este libro, de hecho, tendrían que poderse programar sin demasiadas dificultades en cualquier lenguaje estructurado de alto nivel como *C* o *Pascal*.

Finalmente, un elemento importante en el lenguaje algorítmico y que usamos a menudo es la *recursividad*. Los algoritmos que consideramos usualmente son por su propia definición *iterativos* (se encaminan hacia la solución paso a paso). Un algoritmo se llama *recursivo* si se define a sí mismo. En un algoritmo iterativo, una parte de él puede ejecutarse diversas veces (por ejemplo, en una estructura repetitiva); sin embargo, en un algoritmo recursivo se realiza la reejecución del algoritmo completo desde el comienzo (normalmente con un conjunto diferente de entradas). También usaremos a veces *procedimientos* que se pueden considerar como algoritmos y que se llaman desde dentro del algoritmo principal. Lógicamente pueden existir procedimientos recursivos; por extensión también se llama *algoritmo recursivo* aquel que contiene algún procedimiento recursivo.

1.4 Análisis de algoritmos

Para resolver un problema determinado y una vez decidido el modelo para el tratamiento del problema, pueden existir diversos algoritmos. El análisis de algoritmos es entonces esencial para la realización práctica y eficiente del algoritmo, como por ejemplo su programación en un ordenador. Siempre es de interés perfeccionar el método de resolución de un problema allá donde signifique un esfuerzo menor y, si tenemos en cuenta la figura esquemática del proceso que nos lleva del problema a la solución (Fig. 1.1), lógicamente será más sencillo mejorar el algoritmo que la codificación en lenguaje máquina.

Una manera de comparar dos algoritmos que resuelven un mismo problema podría ser representarlos en un determinado lenguaje de programación, encontrar la solución al problema usando el mismo ordenador y medir el tiempo que tardan en encontrar la solución. Como el tiempo que tarda el ordenador en encontrar la solución es proporcional al número de operaciones elementales que este debe efectuar (sumas, productos, etc.), se suele llamar *tiempo de ejecución* a este número de operaciones.

De los diferentes parámetros que caracterizan un algoritmo, uno de los que nos puede interesar más es precisamente el *tiempo*, o número de operaciones básicas que necesita el algoritmo para solucionar el problema en función del tamaño de la entrada. Otro parámetro que a veces se considera es el *espacio* de memoria que necesita el algoritmo para su ejecución. Como normalmente estaremos más interesados en la rapidez que en el almacenamiento en memoria, cuando analizemos un algoritmo estudiaremos, si no se dice lo contrario, el tiempo de ejecución.

Estos dos parámetros, así como también la misma comparación de algoritmos, mantienen su sentido y son perfectamente tratables en el contexto de la máquina de Turing: La comple-

jidad temporal de una máquina de Turing determinista en función del número de casillas que ocupa la entrada vendrá dada, en este caso, por el número máximo de pasos que puede llegar a hacer la máquina considerando todas las posibles entradas del mismo tamaño. La complejidad espacial viene dada por la distancia máxima—en número de casillas—que puede desplazarse la cinta por delante del cabezal a partir de la casilla inicial.

Además, por ejemplo, si un cierto algoritmo tiene una complejidad temporal polinómica en términos de una máquina de Turing, mantiene la complejidad polinómica usando argumentos más informales basados en el número de operaciones. Por esta razón usaremos habitualmente esta manera de analizar la eficiencia de un algoritmo para evitar la carga que supone trabajar con máquinas de Turing.

Vamos a verlo en un ejemplo: Consideremos dos posibles algoritmos para el cálculo de x^n donde x es un real y n un natural. El primero simplemente va haciendo un bucle acumulando el producto de x por sí mismo n veces:

Entrada: x: real, n: entero.
Algoritmo x^n, (1)

1. $y \leftarrow 1.0$, $i \leftarrow 0$
2. $y \leftarrow yx$, $i \leftarrow i+1$
3. **Si** $i = n$ **entonces salir**
sino ir al paso 2

Salida: y.

Observemos que, si calculamos por ejemplo x^{10000}, este algoritmo efectuará 10000 veces el bucle principal.

El segundo algoritmo es conceptualmente más complicado. Para calcular x^n usa la representación binaria de n para determinar cuales de las potencias $x, x^2, x^4, \ldots$ son necesarias para construir x^n y así considerar sólo su producto. Usa dos operadores, `and` y `shr`, que muchos lenguajes de programación tienen incorporados. El primero actúa bit a bit sobre la representación binaria de los dos enteros considerados y retorna el entero que resulta de tomar 1 si los dos bits son 1 y 0 en cualquier otro caso. Por ejemplo: 923 `and` $123 = 1110011011_2$ `and` $0001111011_2 = 0000011011_2 = 27$. El segundo desplaza, en la reprentación binaria, todos los dígitos una posición hacia la derecha añadiendo un 0 a la izquierda y pierde el dígito de la derecha. Equivale a hacer la división entera por 2. Por ejemplo: `shr` $235 =$ `shr` $11101011_2 = 01110101_2 = 117$.

Entrada: x: real, n: entero.

Algoritmo x^n, (2)

1. $y \leftarrow 1.0$, $z \leftarrow x$.
2. **Si** $(n \texttt{ and } 1) \neq 0$ **entonces** $y \leftarrow yz$.
3. $z \leftarrow zz$.
4. $n \leftarrow \texttt{ shr } n$.
5. **Si** $n \neq 0$ **entonces** ir al paso 2.

Salida: y.

Aplicándolo también para calcular x^{10000} y como $10000 = 10011100010000_2$, es decir, tiene 14 dígitos binarios, el algoritmo sólo hace 14 veces su bucle principal. Así, en general, el primer algoritmo efectúa un número de pasos aproximadamente igual a n, mientras que, para el segundo, el número de pasos es proporcional al número de dígitos que tiene n representado en binario ($\log_2 n$). El segundo algoritmo necesita un tiempo de ejecución mucho menor que el primero para conseguir el mismo resultado.

En este punto conviene introducir una notación útil para estimar la eficiencia de los algoritmos. Cuando describimos que el número de operaciones, $f(n)$, que efectúa un algoritmo depende de n, se pueden ignorar contribuciones pequeñas. Lo que es preciso conocer es un orden de magnitud para $f(n)$ válido para todo n salvo, quizá, un número finito de casos especiales. Así:

Definición 1.3. Sea f una función de $\mathbb{N}$ en $\mathbb{N}$. Decimos que $f(n)$ es $O(g(n))$ si existe una constante k positiva tal que $f(n) \leqslant kg(n)$ para todo $n \in \mathbb{N}$ (salvo posiblemente un número finito de excepciones).

Con esta notación se dice que el primer algoritmo para el cálculo de x^n es $O(n)$ y el segundo $O(\log n)$.

Podemos ahora dar los pasos que es preciso efectuar en el análisis de algoritmos:

1. Describir el algoritmo de forma precisa.
2. Definir el tamaño n de una instancia característica del problema.
3. Calcular $f(n)$, número de operaciones que efectúa el algoritmo para resolver el problema.

De hecho, incluso para un mismo tamaño de datos n, el algoritmo puede tener, como veremos, comportamientos muy diferentes dependiendo de la instancia considerada. En este caso será preciso hacer tres tipos diferentes de análisis:

1. Análisis del caso más favorable: Qué rapidez posee el algoritmo bajo las condiciones más favorables.

2. Análisis del caso medio: Cuál es el rendimiento medio del algoritmo considerando todos los posibles conjuntos de datos iniciales (normalmente asumiendo que todos son igualmente probables).

3. Análisis del caso más desfavorable: Qué rapidez posee en el peor caso (es decir, para datos de entrada que hacen que el algoritmo tenga el peor rendimiento).

La primera opción no suele ser demasiado útil. La segunda es, en general, complicada de calcular y podría ser poco significativa si muchos de los conjuntos de datos afectan poco o son improbables en la práctica. El estudio del caso más desfavorable acostumbra a ser el más utilizado. De todas maneras es preciso tratarlo, teniendo presente siempre el problema que se está solucionando.

Vamos a utilizar nuevamente el algoritmo de Euclides para ilustrar ahora el análisis de algoritmos. Este algoritmo, como bien sabemos, encuentra el máximo común divisor de un par de números enteros positivos.

La recursión juega un papel importante en el algoritmo de Euclides. El punto esencial está en el hecho de que cualquier factor común de m y n lo es también de $m-n$, y también que cualquier factor común de n y $m-n$ lo es de m, ya que $m = n+(m-n)$. Así $\mathrm{mcd}(m,n) = \mathrm{mcd}(m-n,n)$.

Consideremos los enteros m y n y vamos a encontrar su mcd. Supongamos que dividiendo m por n encontramos un cociente q_1 y un resto r_1, es decir, $m = nq_1 + r_1$ con $q_1 \geqslant 0$ y $0 \leqslant r_1 < n$. Como $r_1 = m - nq_1$, aplicando q_1 veces este razonamiento, obtendremos que $\mathrm{mcd}(m,n) = \mathrm{mcd}(n,r_1)$. Repitiendo el proceso encontramos:

$$\begin{array}{rcll} m & = & nq_1 + r_1 & 0 \leqslant r_1 < n \\ n & = & r_1q_2 + r_2 & 0 \leqslant r_2 < r_1 \\ r_1 & = & r_2q_3 + r_3 & 0 \leqslant r_3 < r_2 \\ & \vdots & & \vdots \\ r_{k-1} & = & r_kq_{k+1} + r_{k+1} & 0 \leqslant r_{k+1} < r_k \\ r_k & = & r_{k+1}q_{k+2} & \end{array}$$

Y por tanto:

$$\mathrm{mcd}(m,n) = \mathtt{mcd}(n,r_1) = \mathrm{mcd}(r_1,r_2) = \cdots = \mathrm{mcd}(r_k,r_{k+1}) = r_{k+1}$$

Por ejemplo, los restos sucesivos que encontramos cuando se calcula el mcd de 315 y 91 son 42 y 7 y entonces $\mathrm{mcd}(315,91) = 7$.

Dar el algoritmo es ahora directo:

Entrada: m, n: enteros.
Algoritmo Euclides

1. Dividir m por n. Sea r el resto ($0 \leqslant r < n$).
2. **Si** $r = 0$ **entonces salir**.
3. $m \leftarrow n$, $n \leftarrow r$. Ir al paso 1.

Salida: n.

Analizaremos este algoritmo estudiando el peor caso. Este pasará cuando en cada paso decrece mínimamente la sucesión. Así comencemos por el final: El último valor será 0. El penúltimo 1. Ahora tiene que seguir el valor más pequeño tal que dividido por 1 dé 0 de resto: 1. A continuación el valor más pequeño que dividido por 1 dé 1 de resto: 2. La sucesión que construimos será:

$$0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots$$

Curiosamente, esta sucesión es la conocida *sucesión de Fibonacci*[1]. El peor caso para el algoritmo de Euclides corresponde, entonces, a considerar dos números de Fibonacci sucesivos. Procuremos estimar el orden del algoritmo observando la sucesión: Si los enteros iniciales considerados son los primeros términos de la sucesión, con 2 dígitos son precisos 6 pasos, y con 3 dígitos son precisos 12 pasos, para 100 dígitos son precisos unos 500 pasos. Así vemos que es de orden aproximadamente igual al número de dígitos de los enteros afectados. Para n grande, podemos ver fácilmente que el número medio de iteraciones para calcular $\mathrm{mcd}(m,n)$ es $O(\log n)$. En efecto, usando que el número de Fibonacci, número k es $n = F_k = (((1+\sqrt{5})/2)^k - ((1-\sqrt{5})/2)^k)/\sqrt{5}$—sección 4.3—, y considerando k grande, entonces n es proporcional a $((1+\sqrt{5})/2)^k$. Tomando logaritmos vemos que $k = O(\log n)$ y, como k es precisamente el índice del número, n es por tanto la cantidad de pasos a efectuar cuando se aplica el algoritmo.

El análisis resultará más sencillo en el caso del algoritmo de resolución de un problema que es en cierta manera clásico: el problema de las torres de Hanoi. Este juego, propuesto en 1883 por el matemático francés Édouard Lucas, consiste en tres palos, uno de los cuales contiene un número determinado de discos de diferentes tamaños puestos uno sobre el otro de mayor a menor. Se trata de pasar todos los discos a otro palo, dejándolos también de mayor a menor, y de acuerdo con las reglas:

1. En cada movimiento sólo se puede mover el disco de arriba de un palo a otro.

2. No se puede poner un disco sobre uno de diámetro más pequeño.

[1] Véase el capítulo 4.

El problema consiste en presentar un algoritmo óptimo para resolver el problema dando el número de movimientos que es preciso hacer.

Comenzaremos proponiendo un algoritmo recursivo.

Entrada: $N\_discos, Palo\_inicial, Palo\_final$
Algoritmo HANOI
Procedimiento $H(n,r,s)$. [Mueve n discos del palo r al s]
Si $n = 1$ **entonces** mueve un disco de r a s.
sino hacer $H(n-1,r,t)$ [t es el palo que no es ni r ni s]
mueve un disco de r a s [Sólo queda el de abajo de todo]
$H(n-1,t,s)$
1. *H(N_discos,Palo_inicial,Palo_final)* [Llamada principal]
Salida: Los movimientos que se van haciendo.

La figura 1.5 muestra gráficamente cómo se ejecuta el algoritmo.

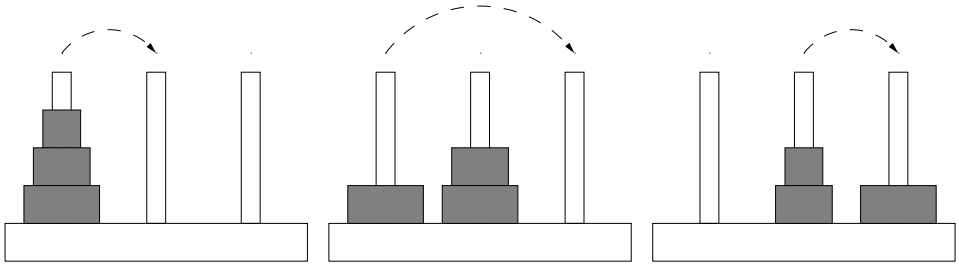


Figura 1.5: Resolución del problema de las torres de Hanoi en el caso de tres discos

Para estudiar su eficiencia vamos a determinar el total de movimientos $h(n)$ que efectúa el algoritmo cuando hay n discos en el palo p_1 y los queremos pasar al palo p_3. Está claro que $h(1) = 1$. El algoritmo dice que, si $n > 1$ y suponiendo que los discos $d_1, d_2, \ldots, d_n$ (d_n es el de abajo de todo) están en el palo p_1, lo que es preciso hacer es pasar los $n-1$ discos $d_1, d_2, \ldots, d_{n-1}$ al palo p_2 (esto son $h(n-1)$ movimientos), después pasar d_n a p_3 y finalmente pasar los $n-1$ discos de p_2 a p_3. El número total de movimientos será:

$$h(n) = 2h(n-1) + 1, \quad n \geqslant 2$$

Solucionar esta ecuación recursiva no es difícil y, de hecho, en el capítulo 4 se tratarán métodos para resolver este tipo de ecuaciones. Aquí veremos cómo podemos encontrar $h(n)$ por inducción. Observemos primero que $h(0) = 0$, $h(1) = 1$, $h(2) = 3$, $h(3) = 7$, $h(4) = 15$, $h(5) = 31$,

etc. Esto nos sugiere que quizá $h(n) = 2^n - 1$. Demostrémoslo. Supongámoslo cierto para $n-1$, es decir, que es cierto que $h(n-1) = 2^{n-1} - 1$. Demostraremos que es cierto para n. En efecto:

$$h(n) = 2h(n-1) + 1 = 2(2^{n-1} - 1) + 1 = 2^n - 1$$

En cuanto al comportamiento del algoritmo, resulta que, tal como se ha planteado el problema, el mejor caso coincide con el peor y con el medio. Así, el algoritmo es $O(2^n)$. Si lo hiciésemos a mano, y moviésemos 1 disco por segundo, en el caso de 8 discos tendríamos que hacer 255 movimientos y tardaríamos unos 4 minutos. Suponiendo un movimiento cada microsegundo, para mover 60 discos harían falta 36600 años. Ya vemos, entonces, que un análisis de este algoritmo nos lleva a detectar un comportamiento fundamentalmente diferente al del algoritmo de Euclides. Finalmente, y para este algoritmo, podemos tratar también la cuestión de la complejidad y demostrar que no es posible encontrar un algoritmo que resuelva el problema con menos movimientos. Supongamos que $\tilde{h}(n)$ sea el número mínimo de movimientos para el algoritmo mejor posible. Consideremos la relación entre $\tilde{h}(n+1)$ y $\tilde{h}(n)$. Para el caso con $n+1$ discos, el disco de abajo de todo se tendrá que mover en algún momento. Antes de hacerlo, los otros n discos habrán de estar en uno de los otros palos. Esto quiere decir que al menos se habrán hecho $\tilde{h}(n)$ movimientos y, similarmente, cuando se haya cambiado de sitio el disco de abajo, se tendrán que hacer al menos $\tilde{h}(n)$ movimientos más para poner el resto de discos encima de él. Por tanto $\tilde{h}(n+1)$ debe valer como mínimo $2\tilde{h}(n)+1$. Dado que, lógicamente, $\tilde{h}(1) = 1$, resulta de la ecuación última que $\tilde{h}(n) \geqslant h(n)$ para todo n. Así, $\tilde{h}(n)$ es el número de movimientos para el mejor algoritmo.

1.5 Comparación de algoritmos

En la sección anterior hemos visto como es posible estudiar el comportamiento de un algoritmo y encontrar su complejidad temporal (o espacial) en función del tamaño de los datos de entrada. Dada, por ejemplo, la relación directa que hay entre la complejidad temporal de un algoritmo y el tiempo de cálculo de su implantación, podemos ver fácilmente qué tipo de problemas son tratables en la práctica. Si usásemos una máquina que realizase una operación básica cada microsegundo, podríamos escribir la tabla 1.2.

Observar que, aunque mejoremos mucho la rapidez de las máquinas, la tabla no se modifica esencialmente. Suponiendo que lleguemos a tener en un futuro próximo máquinas 1000 veces más rápidas, esto simplemente significaría dividir por 1000 los valores de la tabla, cosa que, evidentemente, no afecta el comportamiento asintótico. Es más importante, entonces, conseguir buenos algoritmos para que ninguna instancia del problema conlleve una resolución con un tiempo superior al polinómico.

Tabla 1.2: Tiempo de cálculo según la complejidad (una instrucción por microsegundo)

	$n = 10$	$n = 20$	$n = 40$	$n = 60$	$n = 1000$
n	0.00001 s	0.00002 s	0.00004 s	0.00006 s	0.001 s
n^2	0.0001 s	0.0004 s	0.0016 s	0.0036 s	1 s
n^3	0.001 s	0.008 s	0.064 s	0.216 s	17 m
2^n	0.001 s	1 s	12.7 días	36 534 años Cromagnon	
$n!$	3.629 s	77 094 años Neanderthal	$2.6\ 10^{34}$ años Edad del Universo $1.5\ 10^{10}$ años	$2.6\ 10^{68}$ años	

Así pues, la elección del algoritmo es una parte esencial asociada también a su análisis. Para ilustrar este hecho dedicaremos esta sección a comparar diversos algoritmos de ordenación de listas. El objetivo de los algoritmos es ordenar una lista de elementos de un conjunto según la relación de orden que hay. Esto incluye la ordenación alfabética de nombres, la ordenación de reales, etc.

El primer algoritmo considerado es el llamado *algoritmo de inserción*. Consiste en ir considerando uno a uno los elementos de la lista, a partir del segundo, e insertarlos en la posición que les corresponda comparando con los anteriores hasta encontrar su sitio.

Entrada: Una lista, $L = \{a_1, a_2, \ldots, a_n\}$, con n elementos no ordenados.
Algoritmo INSERCIÓN
  Para $i = 2$ **hasta** n
    $tmp \leftarrow a_i$
    $j \leftarrow i - 1$
    repetir hasta que $j = 0$ **o bien** $tmp \geqslant n_j$
      $n_{j+1} \leftarrow n_j$
      $j \leftarrow j - 1$
    $n_{j+1} \leftarrow tmp$
Salida: La lista L ordenada.

El número de comparaciones que efectúa el algoritmo es el aspecto crítico en su análisis. En el caso peor, cuando la lista esté ordenada al revés, serán precisas $1 + 2 + 3 + \cdots + (n-1) = n(n-1)/2$ comparaciones, esto es, un polinomio $O(n^2)$ en el tamaño de entrada.

¿Cuál es el número de comparaciones en el caso medio? El algoritmo de inserción es precisamente uno de aquellos casos en que el análisis del caso medio no es mucho más difícil de efectuar que el análisis del peor caso. Es preciso en primer lugar encontrar el valor medio de la posición en la cual insertar el elemento a_i, ya que esto nos da el número medio de comparaciones para cada i. Consideremos que todas las $i!$ posibles ordenaciones son igualmente probables. Con esto, la posición donde insertar a_i puede ser cualquiera entre la primera (a_i es menor que cualquiera de los $i-1$ valores ya ordenados) y la i (a_i es más grande que a_{i-1}). Si el sitio correcto donde insertar a_i es la posición j, entonces la cantidad de comparaciones es $i-j+1$ donde $j = 2,3,\ldots,i$. Si a_i se tuviese que insertar en la posición $j = 1$, entonces el número de comparaciones es $i-1$. El número medio de comparaciones resulta ser:

$$\begin{aligned}
\frac{1}{i}\left(i-1+\sum_{j=2}^{i}(i-j+1)\right) &= \frac{1}{i}\left[i-1+\sum_{k=1}^{i-1}k\right] \\
&= \frac{1}{i}\left(i-1+\frac{i(i-1)}{2}\right) \\
&= \frac{(i-1)(i+2)}{2i} = \frac{i^2+i-2}{2i} \\
&= \frac{i}{2}+\frac{1}{2}-\frac{1}{i}
\end{aligned}$$

Para encontrar el número total de comparaciones sólo es preciso sumar esto entre 2 y n y encontramos finalmente:

$$\frac{1}{2}\left[\frac{n(n+1)}{2}-1\right]$$

Otro algoritmo $O(n^2)$ es el *algoritmo de burbujas*. Este algoritmo es simple de describir. La idea es pasar por la lista intercambiando dos elementos sucesivos si están en orden incorrecto. Al final de la primera pasada, el elemento más grande quedará situado en el sitio correcto. Se hace ahora otra pasada y será el segundo más grande el que se situará en su sitio. En total serán precisas $n-1$ pasadas para conseguir ordenar la lista. También vemos que en la segunda pasada no es preciso llegar al final, sólo es preciso considerar los $n-1$ primeros elementos. Análogamente, $n-2$ en la tercera, etc. El número total de comparaciones es el mismo que con el algoritmo de inserción. El nombre del algoritmo proviene de la asociación con el hecho de que en cada pasada el elemento mayor se va desplazando hacia un extremo como si fuese una burbuja dentro de un líquido que subiese a la superficie.

Otro algoritmo conocido de ordenación es el algoritmo *mergesort*. Es un ejemplo clásico de la técnica algorítmica conocida como *dividir y vencer*. La idea consiste en dividir la lista en dos por la mitad y ordenar cada mitad recursivamente. Al final se combinan las dos listas manteniendo el orden creciente.

Damos primero el algoritmo de combinación de listas:

Entrada: Dos listas $L_1 = \{a_1, a_2, \ldots, a_r\}$ y $L_2 = \{b_1, b_2, \ldots, b_s\}$
Algoritmo COMBINA(L_1,L_2) `[`L_1 y L_2 tienen r y s elementos ordenados`]`

1. $i \leftarrow 1$, $j \leftarrow 1$, $k \leftarrow 1$.
2. **2.1**. **Si** $a_i \leqslant b_j$ **hacer** $c_k \leftarrow a_i$ y **Si** $i < r$ **hacer** $i \leftarrow i+1$.
sino hacer $c_k \leftarrow b_j$ y **Si** $j < s$ **hacer** $j \leftarrow j+1$
2.2. **Si** $i = r$ **hacer** $a_r \leftarrow b_s$ y **Si** $j = s$ **fer** $b_s \leftarrow a_r$.
3. $k \leftarrow k+1$.
4. **Si** $k \leqslant r+s$ volver al paso 2.
Salida: La lista ordenada $L = \{c_1, c_2, \ldots, c_{r+s}\}$.

El algoritmo MERGESORT de ordenación de listas para el caso de una lista de naturales viene dado por:

Entrada: La lista $L = \{a_i, a_{i+1}, \ldots, a_j\}$ con n naturales.
Algoritmo MERGESORT

1. **Si** L tiene un sólo elemento la lista está ordenada. **Salir**.
2. $k \leftarrow \lfloor \frac{i+j}{2} \rfloor$.
$L_1 \leftarrow \{a_i, a_{i+1}, \ldots, a_k\}$.
$L_2 \leftarrow \{a_{k+1}, a_{k+2}, \ldots, a_j\}$.
3. Ordenar por separado L_1 y L_2 usando MERGESORT. `[llamada recursiva]`
4. $L \leftarrow$ COMBINA(L_1, L_2).
Salida: La lista ordenada L.

Podemos estudiar un poco la complejidad de este algoritmo. Para la etapa de combinación de listas se hacen $n-1$ comparaciones. Así, el total de pasos vendrá dado por

$$M(n) = M(\left\lfloor \frac{n}{2} \right\rfloor) + M(\left\lceil \frac{n}{2} \right\rceil) + n - 1, \quad M(1) = 0$$

Se trata, como en el caso de las torres de Hanoi, de solucionar esta ecuación recursiva. La estudiaremos en el caso en que n sea una potencia de 2. De entrada, cambiamos $n-1$ por n (encontramos una cota superior, de todas maneras $n-1$ correspondía al caso peor de la combinación de listas). Como $n = 2^k$, introduciendo $m(k) = M(2^k)$ la ecuación nos quedará: $m(k) = 2m(k-1) + 2^k$, $\quad m(0) = 0$. Esta ecuación es fácil de resolver.

Ejercicio 1.4. Demostrar por inducción que la solución de la ecuación

$$m(k) = 2m(k-1) + 2^k, \quad m(0) = 0$$

es $m(k) = k2^k$.

Así resulta que el número total de pasos es $M(n) \leqslant n\log n$. No es difícil de ver que este resultado es válido para cualquier n aunque no sea una potencia de 2. En resumen, estamos ante un algoritmo de ordenación $O(n\log n)$. Un inconveniente del algoritmo es la ocupación de memoria. Observad que en la etapa de combinación se crea una nueva lista del mismo tamaño que la inicial (de hecho puede ser modificado per evitar esto, a costa de aumentar su complejidad).

Finalmente presentamos el algoritmo *quicksort*. También es de la familia de algoritmos de *dividir y vencer*. Mientras en el algoritmo anterior se trata de partir la lista lo más exactamente posible, ahora lo que se hace es partirla creando nuevas listas de forma que una sublista tenga elementos inferiores o iguales a la otra.

Entrada: La lista $L = \{a_1, a_2, \ldots, a_n\}$ de n naturales.
Algoritmo QUICKSORT

1. Si L tiene un solo elemento la lista está ordenada. **Salir**.
2. Considerar a_1 el primer elemento de la lista.
3. Separar el resto de la lista en dos sublistas L_1 y L_2
de forma que L_1 contiene los elementos anteriores a a_1 y L_2 los posteriores.
4. Ordenar por separado L_1 y L_2 con QUICKSORT. `[llamada recursiva]`
5. Concatenar la primera lista, el elemento a_1 y la segunda lista.

Salida: La lista ordenada L.

Ejercicio 1.5. Escribir un algoritmo para separar una lista de acuerdo con el paso 3 del algoritmo QUICKSORT.

Se puede ver que QUICKSORT tiene una complejidad temporal—número de comparaciones—$O(n\log n)$ (como MERGESORT). ¿Cuál de los dos es preferible?

1. El peor caso de QUICKSORT se sabe que es de orden más grande que el peor de MERGESORT. En general, los casos peores son poco probables. Así, este aspecto lo hemos de tener en cuenta, pero no es decisivo.

2. La utilización de la memoria en QUICKSORT es de orden n, mientras que en MERGESORT es de $2n$. Ésta es una cuestión importante si queremos emplear ordenadores pequeños para tratar grandes cantidades de datos.

3. Si se hace un análisis de la complejidad temporal para el caso medio, el factor resultante se inclina ligeramente a favor de MERGESORT. Si tenemos en cuenta que el movimiento de elementos de la lista no ha sido contado (sólo las comparaciones), en la práctica resulta que QUICKSORT es más rápido.

Con este estudio se ha pretendido mostrar que el análisis de algoritmos puede ayudar de forma importante en el diseño de un algoritmo para la resolución de un problema concreto. Sin embargo, como veremos más adelante, hay problemas que son intrínsecamente difíciles de tratar y para los cuales no se conocen algoritmos eficientes. Ser capaz de identificarlos y conocer sus limitaciones también será de utilidad.

1.6 Clasificación de algoritmos

Para poder discutir de una forma abstracta y general la eficiencia de diferentes algoritmos en la solución de problemas y así poder hacer una clasificación, es preciso en primer lugar hacer una consideración sobre los tipos de problemas que se nos pueden plantear:

- Problemas de búsqueda: Encontrar una cierta X en los datos de entrada que satisfaga la propiedad P.
- Problemas de transformación: Transformar los datos de entrada para que satisfagan la propiedad P.
- Problemas de construcción: Construir un conjunto C que satisfaga la propiedad P.
- Problemas de optimización: Encontrar el mejor X que satisface la propiedad P.
- Problemas de decisión: Decidir si los datos de entrada satisfacen la propiedad P.

Para una discusión abstracta sobre la posibilidad de resolver problemas mediante algoritmos eficientes, resulta conveniente reformularlos en términos de problemas de decisión, en el sentido de buscar una respuesta del tipo si/no. Esto nos permitirá poder comparar problemas muy diferentes. Por ejemplo, el problema de multiplicar dos enteros (dados x e y, ¿cuál es el valor de su producto?) puede presentarse también como: Dados los enteros x, y y z, ¿es cierto que $xy = z$?

Un problema de decisión se dice que es de tipo **P** si para su resolución son precisos algoritmos que realicen un número de operaciones básicas $O(p)$, donde p es un polinomio en

el tamaño del problema. Hay un tipo de problemas especialmente difíciles de tratar. Son los llamados **NP**, que quiere decir *no determinístico polinómico*. Estos problemas pueden ser resueltos en tiempo polinómico con una máquina de Turing no determinista o MTND, que puede ser definida como un conjunto de MTD procesando la cinta en paralelo. En este caso, la complejidad temporal en función de la longitud n de la entrada viene dada por el número de pasos máximo que puede hacer, considerando todas las posibles entradas de longitud n y tomando como tiempo correspondiente a una entrada el de la máquina de Turing determinista que ha tardado menos que todas las que computaban en paralelo. Una solución de un problema NP puede ser comprobada en tiempo polinómico.

Para muchos problemas NP se conocen cotas lineales eficientes, sin embargo las cotas conocidas para el caso peor son exponenciales.

Los problemas tipo P se incluyen claramente en el conjunto de problemas tipo NP. Una controversia conocida en el mundo de la algorítmica es si P es igual a NP. Se piensa que $P \neq NP$. El hecho que esta cuestión sea difícil de responder reside en la dificultad de probar que un problema *no* se puede resolver en tiempo polinómico. Para esto sería preciso considerar *todos* los posibles algoritmos de resolución y demostrar que todos ellos son ineficientes.

Se dice que un problema de decisión X se puede *transformar polinómicamente* en otro problema de decisión Y si, dadas entradas A y B, que pueden construirse una de la otra en tiempo polinomial respecto del original, A hace que X tenga respuesta afirmativa si y solamente si B hace que Y tenga respuesta afirmativa. Con esta noción se puede introducir la definición de problema **NP-C** (*NP-completo*). Un cierto problema X es del tipo NP-C si, además de ser NP, todos los otros problemas NP se pueden transformar en X polinómicamente. Se demuestra [4] [7] que, si se sabe resolver de forma eficiente uno de los problemas NP-C, quedan resueltos todos. De la misma manera, si se demuestra que uno cualquiera de los problemas de esta categoría no es tratable, todos los otros tampoco lo son.

Finalmente conviene comentar que, a pesar de la dificultad de dar algoritmos eficientes para la solución general de problemas del tipo NP-C, se conocen buenos métodos heurísticos que llevan a soluciones cuasi-óptimas. Así, en el capítulo 7 se presentarán algunos de ellos.

Notas históricas y bibliográficas

Aunque las primeras referencias escritas de algoritmos son de los griegos (por ejemplo, el algoritmo de Euclides, *Elementos*, libro 7, prop. 1-2), el nombre tiene origen en el matemático persa Abu Kha'far Muhammad ibn Musá Abdallah *al-Hwarizmi* al-Madjusi al-Qutrubulli, nacido en Hwarizm (Urgenč, Uzbekistán) hacia el año 780 y que trabajó gran parte de su vida en Bagdad. Así, la palabra algoritmo en realidad viene del nombre del pueblo donde nació este matemático, más conocido por su obra *Kitab al-jabr wa'l-muqabala*, que precisamente

da nombre a otra parte importante de las matemáticas: el álgebra. A pesar de los orígenes históricos realmente antiguos de ejemplos concretos de algoritmos, la formulación precisa de la noción de algoritmo es de este siglo. La aportación más importante es sin duda la de Alan Turing, matemático inglés que en el año 1932 introdujo el concepto de *máquina de Turing*, que ha contribuido enormemente al desarrollo de la teoría de la computabilidad. Otro hito a indicar fueron los resultados de S. A. Cook (1971) y L. Levin (1973) sobre los problemas NP-C. La algorítmica es una de las áreas científicas más activas. Entre los últimos avances interesantes es preciso destacar los trabajos (1988 y 1990) sobre complejidad estructural de José L. Balcázar, Josep Díaz y Joaquim Gabarró, profesores de la UPC, y el resultado de 1990 de Carsten Lund, Lance Fortnow y Howard Karloff de la Universidad de Chicago, Noam Nisan del MIT y Adi Shamir del Weizmann Institute, que de forma resumida dice que prácticamente cualquier problema, incluyendo los NP, tiene una verificación probabilista sencilla.

Bibliografía

[1] M. Abellanas, D. Lodares. *Análisis de Algoritmos y Teoría de Grafos.* RA-MA Editorial, 1990.

[2] A. V. Aho, J. E. Hopcroft, J. D. Ullman. *The Design and Analysis of Computer Algorithms.* Addison-Wesley, 1974.

[3] J. L. Balcázar, J. Diaz, J. Gabarró. *Structural Complexity I (II)*, Springer–Verlag, 1988 (1990).

[4] S. A. Cook. "The complexity of theorem proving procedures", *Proceeding of the 3rd. Annual ACM Symposium on the Theory of Computation*, pp. 151–158, 1971.

[5] D. Harel. *Algorithmics. The Spirit of Computing*. Addison-Wesley, 1987.

[6] L. Kučera. *Combinatorial Algorithms.* Adam Hilger, 1990.

[7] L. Levin. "Universal search problems", *Problems of Information Transmission*, **9**, pp. 265–266, 1973.

[8] S. B. Maurer, A. Ralston. *Discrete Algorithmic Mathematics.* Addison-Wesley, 1991.

[9] G. J. E. Rawlins. *Compared to What? An Introduction to the Analysis of Algorithms.* Freeman, 1992.

Problemas

1. Se considera una máquina de Turing donde la cinta usa un alfabeto con los símbolos x, 1 y 2 (además de la posibilidad de tener la casilla vacía, b). Los estados son cinco, de los cuales dos son estados especiales de parada que llamamos A_{SI} y A_{NO}. Dad un programa que compruebe si un entero m, $m > 1$ divide exactamente a otro entero n. Para ello usar la representación unaria para m y n y entrar los datos en la cinta inicialmente como

   $$[\ldots, x, 1, 1, \ldots, 1, b, 1, 1, \ldots, 1, 1, x, \ldots]$$

   con m unos en la parte a la izquierda del cabezal, que se encontrará situado sobre la casilla vacía, y n unos a la derecha.

2. Sea P_n el peor caso, en cuanto a comparaciones, en que se puede encontrar el algoritmo QUICKSORT para ordenar una lista de n elementos.

   (a) Justificar las ecuaciones:

   $$P_n = n - 1 + \max_{0 \leqslant i < n} (P_i + P_{n-i-1}); \quad P_0 = 0$$

   (b) Demostrar que la solución (única) de las ecuaciones anteriores es $P_n = n(n-1)/2$

3. (a) ¿Qué tipo de lista inicial lleva al peor comportamiento para el algoritmo de burbujas, en cuanto a comparaciones?

   (b) Realizar el análisis del peor caso para este algoritmo.

Parte I Enumeración

En esta parte se presentan diversas técnicas para contar los elementos de un conjunto. Paralelamente a la descripción de técnicas usuales de enumeración, se presentan también problemas clásicos de combinatoria, a los cuales se aplican los resultados que se van obteniendo.

Las técnicas más sencillas basadas en la enumeración de permutaciones y combinaciones se tratan en el primero de los tres capítulos que componen esta primera parte, y se aplican estas técnicas a problemas de enumeración de funciones entre conjuntos, palabras de alfabetos, distribuciones, particiones de enteros y la fórmula del binomio. En el segundo capítulo se analizan algunos principios básicos de enumeración, especialmente el principio de inclusión-exclusión, y se consideran los problemas de desarreglos, función de Euler, números de Catalan y particiones de enteros. Aunque no son estrictamente principios de enumeración, se incluyen también en este capítulo el principio del palomar y el teorema de Ramsey. El último capítulo de esta parte está dedicado a las ecuaciones de recurrencia y las funciones generadoras. Aunque hemos intentado mantener un nivel asequible, estos son los temas que requieren más nivel matemático, y resultarán más fáciles al lector que tenga cierta familiaridad con las series de potencias. Esta primera parte se cierra con los números de Stirling y de Bell. Algunos temas de enumeración que requieren conocimientos adicionales se ven en otras partes del texto. Así, por ejemplo, ciertos problemas de enumeración de grafos se tratan en la segunda parte, y la teoría de enumeración de Pólya se presenta en la tercera parte.

Capítulo 2

Combinaciones y permutaciones

1. Selecciones ordenadas y no ordenadas
2. Algunos ejemplos de aplicación
3. Propiedades de los coeficientes binomiales

En este capítulo se exponen los problemas más simples de enumeración que forman parte de la combinatoria elemental. Los modelos básicos se basan en la enumeración de selecciones ordenadas y no ordenadas, con o sin repetición, de los elementos de un cierto conjunto. En la sección 1 se obtienen las fórmulas de enumeración de estas selecciones. A pesar de su simplicidad, estos problemas de enumeración permiten resolver una diversidad considerable de problemas, de los cuales hay algunos ejemplos interesantes en la sección 2: el número de palabras que pueden formarse a partir de un alfabeto, el número de soluciones de ciertas ecuaciones enteras, el número de aplicaciones entre dos conjuntos, la fórmula del binomio y problemas relacionados, y los problemas de distribuciones. Los llamados coeficientes binomiales tienen una importancia singular y permiten expresar muchos de los resultados de enumeración; la tercera sección está dedicada a analizar las propiedades más importantes de estos números.

2.1 Selecciones ordenadas y no ordenadas

Comenzaremos con un recorrido por la combinatoria elemental contando de cuántas maneras diferentes se pueden seleccionar un cierto número de elementos de un conjunto. Para contar este número es preciso fijar los criterios con que se diferencia una selección de otra. Aquí tendremos en cuenta dos tipos de criterios: el orden de los elementos y el número de veces que puede aparecer cada uno.

Si distinguimos dos selecciones: cuando tienen elementos diferentes, o bien, cuando los elementos aparecen en un orden diferente, hablaremos de *permutaciones*. En cambio, si no distinguimos dos selecciones que sólo difieren en la ordenación de sus elementos, entonces hablaremos de *combinaciones*. Por otra parte, si cada elemento puede aparecer como mucho una vez, hablaremos de selecciones *sin repetición*, mientras que, si no hay esta restricción, hablaremos de seleccciones *con repetición*. Por ejemplo, en el conjunto

$$X = \{1,2,3,4\}$$

podemos formar 16 permutaciones, con repetición, de dos elementos,

11	12	13	14
21	22	23	24
31	32	33	34
41	42	43	44

12 permutaciones, sin repetición, de dos elementos,

	12	13	14
21		23	24
31	32		34
41	42	43	

10 combinaciones, con repetición, de dos elementos,

11	12	13	14
	22	23	24
		33	34
			44

y 6 combinaciones, sin repetición, de dos elementos,

	12	13	14
		23	24
			34

En esta sección obtendremos una fórmula para la enumeración del número de selecciones diferentes de k elementos, tomados de un conjunto X de n elementos, que identificaremos con $\{1,2,\ldots,n\}$.

Lo que resulta más sencillo de contar es el número de selecciones ordenadas con repetición de k elementos. Llamamos PR_n^k a este número, que se lee "permutaciones con repetición de n

elementos tomados de k en k". Tenemos n elecciones para el primer elemento de la selección, y para cada elección podemos formar todas las permutaciones con repetición de n elementos, pero ahora tomados de $k-1$ en $k-1$. Es decir,

$$PR_n^k = nPR_n^{k-1}$$

Aplicando esta fórmula sucesivamente, y teniendo en cuenta que $PR_n^1 = n$ (hay n elecciones para el último elemento), obtenemos

$$PR_n^k = nPR_n^{k-1} = n^2PR_n^{k-2} = \cdots = n^k$$

En el ejemplo anterior obtenemos $PR_4^2 = 4^2 = 16$ permutaciones con repetición de cuatro elementos tomados de 2 en 2.

Consideremos ahora las permutaciones sin repetición de n elementos tomados de k en k, el número de las cuales denotaremos por P_n^k. Como cada elemento puede aparecer como mucho una sola vez en la selección, es preciso que $k \leqslant n$. Podemos calcular este número con un argumento similar al anterior. Tenemos n opciones para el primer elemento y para cada uno podemos formar las permutaciones de los $n-1$ elementos restantes (ahora un elemento puede salir en la selección como mucho una vez) tomados de $k-1$ en $k-1$, de manera que

$$P_n^k = nP_{n-1}^{k-1}$$

Como $k \leqslant n$, la aplicación sucesiva de esta relación lleva a $P_{n-(k-1)}^1 = n-(k-1)$ (el último elemento se puede escoger entre los $n-(k-1)$ que aún no han sido escogidos), de donde

$$P_n^k = n \cdot (n-1) \cdots (n-k+1)$$

En particular, cuando $k = n$, obtenemos las *permutaciones* de n elementos, el número de las cuales se denota simplemente como $P_n = n \cdot (n-1) \cdots 3 \cdot 2 \cdot 1$. Este número se representa con el símbolo $n!$ y se lee *factorial* de n. En particular, el símbolo factorial permite escribir P_n^k de manera más económica como

$$P_n^k = \frac{n!}{(n-k)!}$$

notación que se puede extender al caso $n = k$ si adoptamos el convenio que $0! = 1$. Veremos más adelante que este convenio tiene una justificación combinatoria y permite además dar cohesión a muchas notaciones, de manera que es universalmente aceptado.

Se puede considerar una situación intermedia entre las permutaciones con repetición y las permutaciones sin repetición y también es fácil de contar. Consiste en fijar de entrada el número de veces que cada elemento debe aparecer en la selección. Por ejemplo, se pueden formar 12 permutaciones con los elementos de $X = \{1,2,3,4\}$ de manera que 1 aparezca exactamente dos veces, 2 y 3 aparezcan una sola vez y 4 ninguna,

1123	1213	1231	2113	2131	2311
1132	1312	1321	3112	3121	3211

En general, supongamos que el elemento i aparece k_i veces en la selección. Formemos un nuevo conjunto

$$Y = \{1_1, \ldots, 1_{k_1}, 2_1, \ldots, 2_{k_2}, \ldots, n_1, \ldots, n_{k_n}\}$$

de $k = k_1 + k_2 + \cdots + k_n$ elementos, en el cual hemos distinguido provisionalmente las k_i apariciones de cada elemento. Formemos ahora las $k!$ permutaciones de los elementos de este conjunto y agrupemos aquellas que difieren sólo en una permutación de los elementos del mismo tipo. En el ejemplo anterior obtendríamos los grupos siguientes,

$1_1 1_2 23$	$1_1 21_2 3$	$1_1 231_2$	$21_1 1_2 3$	$21_1 31_2$	$231_1 1_2$
$1_2 1_1 23$	$1_2 21_1 3$	$1_2 231_1$	$21_2 1_1 3$	$21_2 31_1$	$231_2 1_1$
$1_1 1_2 32$	$1_1 31_2 2$	$1_1 321_2$	$31_1 1_2 2$	$31_1 21_2$	$321_1 1_2$
$1_2 1_1 32$	$1_2 31_1 2$	$1_2 321_1$	$31_2 1_1 2$	$31_2 21_1$	$321_2 1_1$

En general, cada grupo tiene $k_1! \cdot k_2! \cdots k_n!$ elementos, que representan de hecho la misma permutación cuando dejamos de distinguir los subíndices. Así, el número de permutaciones de n elementos en los cuales el elemento i aparece k_i veces, que denotaremos por $P_n^{k_1,\ldots,k_n}$, vale

$$P_n^{k_1,\ldots,k_n} = \frac{k!}{k_1! \cdot k_2! \cdots k_n!}, \quad k = k_1 + k_2 + \cdots + k_n$$

Vamos a considerar ahora selecciones no ordenadas o combinaciones. Dos selecciones serán diferentes si y sólo si tienen elementos diferentes. Primero contaremos las combinaciones sin repetición de n elementos tomados de k en k (donde k no puede ser más grande que n), y denotaremos este número como C_n^k. Para esto consideramos provisionalmente todas las P_n^k selecciones ordenadas, y agrupamos aquellas que sólo difieren en el orden de sus elementos. En el ejemplo al comienzo de la sección, en el que se formaban selecciones de dos elementos de un conjunto de cuatro, obtendríamos los grupos,

12 21	13 31	14 41
	23 32	24 42
		34 43

En general, cada uno de los grupos contiene $k!$ permutaciones que representan la misma combinación, de manera que

$$C_n^k = \frac{P_n^k}{k!} = \frac{n!}{(n-k)!\,k!}, \quad 0 \leqslant k \leqslant n$$

Este número aparece con tanta frecuencia en la combinatoria que recibe también una notación y denominación especiales: se llama *coeficiente binomial* y se denota por $\binom{n}{k}$,

$$\binom{n}{k} = \frac{n!}{(n-k)!\,k!}$$

(se dice 'n sobre k' o 'n escoge k'). En una sección posterior trataremos algunas de las propiedades y aplicaciones de estos números. De momento observemos que este número coincide también con el de permutaciones de dos elementos en que uno se repite $k_1 = k$ veces y el otro $k_2 = n-k$ veces,

$$P_2^{k,n-k} = \frac{n!}{k!\,(n-k)!} = \binom{n}{k} = C_n^k \tag{2.1}$$

En la sección siguiente discutiremos una interpretación combinatoria de este resultado, pero de momento nos será útil para contar el último de los números que nos interesan aquí: el de las *combinaciones con repetición* de n elementos tomados de k en k, que denotaremos por CR_n^k (aquí no es preciso que $k \leqslant n$). Observemos que si agrupamos las permutaciones con repetición que sólo difieren en el orden, no todos los grupos tienen el mismo tamaño. En el ejemplo del comienzo de la sección, el grupo que corresponde a la combinación $\{1,1\}$ tiene un único elemento, mientras que el que corresponde a $\{1,2\}$ contiene las dos permutaciones 12 y 21. Por lo tanto, la técnica que hemos usado antes para contar combinaciones sin repetición ya no nos es útil aquí.

En este caso nos serviremos de una estrategia ingeniosa. Pongamos $n-1$ barras que definen n espacios, uno para cada elemento del conjunto original. Identifiquemos cada una de las combinaciones con repetición poniendo en cada uno de estos espacios tantas estrellas como elementos correspondientes haya en la combinación. En el ejemplo tendríamos las correspondencias,

$$\begin{array}{llll}
11 \leftrightarrow \star\star|\,|\,| & 12 \leftrightarrow \star|\,\star|\, & \\
13 \leftrightarrow \star|\,|\,\star| & 14 \leftrightarrow \star|\,|\,|\star \\
22 \leftrightarrow |\star\star|\,| & 23 \leftrightarrow |\star|\,\star| \\
24 \leftrightarrow |\star|\,|\star & 33 \leftrightarrow |\,|\star\star| \\
34 \leftrightarrow |\,|\star|\star & 44 \leftrightarrow |\,|\,|\star\star
\end{array}$$

De acuerdo con esta correspondencia, hay tantas combinaciones con repetición de n elementos tomados de k en k como permutaciones de dos elementos (barras y estrellas) con exactamente $n-1$ barras y k estrellas. Ya hemos visto en la expresión 2.1 que este número coincide con el de las combinaciones de $k+n-1$ elementos tomados de k en k, de manera que,

$$CR_n^k = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!\,(n-1)!}$$

Con este número se completan las expresiones más comunes de la combinatoria elemental que resumimos en la tabla que sigue.

Tabla 2.1: Número de selecciones

	permutaciones	combinaciones
sin repetición	$P_n^k = \dfrac{n!}{(n-k)!}, \quad k \leqslant n$	$C_n^k = \binom{n}{k} = \dfrac{n!}{(n-k)!\,k!}, \quad k \leqslant n$
con repetición	$PR_n^k = n^k$	$CR_n^k = \binom{n+k-1}{k} = \dfrac{(n+k-1)!}{(n-1)!\,k!}$
con repeticiones $k_1, \ldots, k_n$ $k = k_1 + \cdots + k_n$	$P_n^{k_1,\ldots,k_n} = \dfrac{k!}{k_1! \cdot k_2! \cdots k_n!}$	1

2.2 Algunos ejemplos de aplicación

El problema de contar el número de selecciones de elementos de un conjunto que hemos usado como modelo en la sección anterior es sólo una referencia a la cual se pueden reducir muchos de los problemas de la combinatoria elemental. En esta sección expondremos unos cuantos de estos problemas y su relación con el problema original.

Palabras de alfabetos

Dado un alfabeto de n símbolos, $A = \{1, 2, \ldots, n\}$, queremos contar el número de posibles palabras diferentes que se pueden formar de acuerdo con diversos criterios.

El número de palabras de una longitud fijada k que se pueden formar con n símbolos coincide con el de selecciones ordenadas con repetición:

$$PR_n^k = n^k$$

Por ejemplo, con ocho bits (ceros o unos) se pueden formar $2^8 = 256$ palabras. En el caso de que todos los símbolos de una palabra sean diferentes (por tanto $k \leqslant n$), podemos formar

tantas palabras como selecciones ordenadas de k elementos de un conjunto de n, es decir, el número de permutaciones de n elementos tomados de k en k,

$$P_n^k = \frac{n!}{(n-k)!}$$

Si fijamos la cantidad de cada uno de los símbolos que aparece en cada palabra, es decir, exigimos que haya k_i símbolos i para $1 \leqslant i \leqslant n$ (y por tanto la palabra tiene longitud $k = k_1 + k_2 + \cdots + k_n$), el número de palabras que se pueden formar es el de permutaciones con repeticiones fijadas,

$$P_k^{k_1,\ldots,k_n} = \frac{k!}{k_1! \cdots k_n!}$$

En particular, si el alfabeto sólo tiene dos símbolos, el número de palabras en que uno de los símbolos aparece exactamente k veces y el otro $n-k$ es

$$P_2^{k,n-k} = \frac{k!}{k!\,(n-k)!} = \binom{n}{k}$$

es decir, el número de combinaciones de n elementos tomados de k en k.

Conjuntos y aplicaciones

El número de permutaciones con repetición de k elementos de un conjunto de tamaño n se puede interpretar también como una aplicación que le asigna uno de los n elementos a cada una de las k posiciones de la selección. En la figura siguiente se ilustra esta asignación para $k = 7$ y $n = 5$.

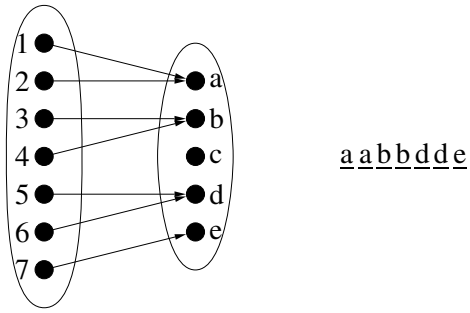


Figura 2.1: Relación entre aplicaciones y selecciones

Con esta analogía, si $B = \{1, 2, \ldots, k\}$ y $A = \{a_1, a_2, \ldots, a_n\}$ son dos conjuntos finitos, cada aplicación $f : B \to A$ se puede identificar con la selección ordenada $(f(1), \ldots, f(k))$ de elementos de A. El número de aplicaciones de B en A es entonces el de permutaciones con repetición $PR_n^k = n^k$ (por ello se suele representar el conjunto de aplicaciones de B en A por A^B).

Por otra parte, el número de aplicaciones inyectivas (es decir, dos posiciones diferentes no pueden tener el mismo elemento) que se pueden definir de B en A es el número de permutaciones de n elementos tomados de k en k, P_n^k (y entonces tiene que ser $k \leqslant n$). En una sección posterior se trata la enumeración de aplicaciones exhaustivas (problema 5 del capítulo siguiente).

Las combinaciones se asocian de manera natural con subconjuntos. Una selección no ordenada de k elementos de un conjunto de tamaño n es de hecho un subconjunto de tamaño k. El número de subconjuntos de k elementos de un conjunto de tamaño n es entonces el de las combinaciones C_n^k. Aquí admitimos que hay un subconjunto de cero elementos (el subconjunto vacío) y uno de n elementos (el conjunto de partida), de donde

$$C_n^0 = \frac{n!}{n!\,0!} = C_n^n = 1$$

cosa que justifica el convenio que hemos adoptado de escribir $0! = 1$.

Una manera de representar los subconjuntos de tamaño k de un conjunto de tamaño n consiste en enumerar los elementos del conjunto, $A = \{a_1, a_2, \ldots, a_n\}$, y expresar cada subconjunto $B \subset A$ como una palabra de longitud n de 0's y 1's, $x_1 x_2 \ldots x_n$, $\quad x_i \in \{0, 1\}$, de manera que x_i vale 1 si el elemento a_i pertenece a B y vale cero de otro modo. Por ejemplo, el subconjunto $B = \{a_2, a_5\}$ de $A = \{a_1, a_2, a_3, a_4, a_5, a_6\}$ se representaría por la palabra 010010, mientras que el conjunto A entero se representa por la palabra 111111. Volvemos a encontrar, entonces, que el número de palabras de longitud n con k 1's y $(n-k)$ 0's es el mismo que el número de subconjuntos de tamaño k de un conjunto de tamaño n, $\binom{n}{k}$.

Binomios y otras expresiones aritméticas

El origen de la denominación *coeficiente binomial* para los números $\binom{n}{k}$ se encuentra en el cálculo del desarrollo de la expresión binomial

$$(a+b)^n = \underbrace{(a+b)\cdot(a+b)\cdots(a+b)}_{n}$$

donde n es un número entero. Desarrollando el producto de los n paréntesis, se obtiene una expresión en la cual aparecen términos del estilo $a^i b^{n-i}$ con $0 \leqslant i \leqslant n$. Cada término $a^i b^{n-i}$ aparece al escoger a en i de los paréntesis y b en los $n-i$ restantes al hacer el producto. Como

esta elección se puede hacer de $\binom{n}{i}$ maneras diferentes, se obtiene la expresión

$$(a+b)^n = \binom{n}{n}a^n b^0 + \binom{n}{n-1}a^{n-1}b^1 + \cdots + \binom{n}{0}a^0 b^n = \sum_{i=0}^{n}\binom{n}{i}a^i b^{n-i}$$

llamada *fórmula del binomio*, en la cual los coeficientes de cada monomio son números combinatorios.

De manera similar se obtiene la llamada fórmula multinomial que permite expresar el desarrollo de

$$(a_1+a_2+\cdots+a_n)^k = \underbrace{(a_1+a_2+\cdots+a_n)\cdots(a_1+a_2+\cdots+a_n)}_{k}$$

Razonando de la misma manera que en el caso anterior, el coeficiente de $a_1^{k_1}\cdot a_2^{k_2}\cdots a_n^{k_n}$ en este desarrollo es el número de elecciones de a_i en k_i de los paréntesis, para cada i, $1 \leqslant i \leqslant n$, de donde

$$(a_1+a_2+\cdots+a_n)^k = \sum \frac{k!}{k_1!\cdot k_2!\cdots k_n!}a_1^{k_1}\cdot a_2^{k_2}\cdots a_n^{k_n}$$

donde el sumatorio se extiende a todas las combinaciones de enteros no negativos $k_1, k_2, \ldots, k_n$ tales que $k_1+k_2+\cdots+k_n = k$. De aquí proviene también la denominación de coeficiente multinomial para $P_n^{k_1 k_2 \ldots k_n}$. Por analogía con el coeficiente binomial, el multinomial se expresa también como

$$\binom{k_1+k_2+\cdots+k_n}{k_1,k_2,\cdots,k_n} = \frac{(k_1+k_2+\cdots+k_n)!}{k_1!\cdot k_2!\cdots k_n!}$$

Supongamos ahora que extendemos la expresión multinomial al caso de que haya una cantidad infinita numerable de sumandos. Para simplificar la situación supondremos que los sumandos son potencias de a,

$$(\sum_{i\geqslant 0} a^i)^n = (1+a+a^2+a^3+\cdots)^n$$

En este caso obtendremos una expresión del estilo

$$c_0 + c_1 a + c_2 a^2 + c_3 a^3 + \cdots$$

donde el coeficiente c_i corresponde al número de maneras de escoger un sumando en cada uno de los n paréntesis $(1+a+a^2+a^3+\cdots)$ de manera que la suma de las potencias sea i. Para calcular este coeficiente se puede seguir la estrategia siguiente. Identificamos cada uno de los paréntesis con una bola numerada (de 1 a n), y extraemos una selección no ordenada de i bolas

con repeticiones permitidas. Identificamos la selección con una elección de potencias en cada paréntesis de la manera siguiente. Si en la selección hay r_1 bolas 1, r_2 bolas 2, y en general r_i bolas i, tomamos a^{r_1} en el primer paréntesis, a^{r_2} en el segundo y, en general, a^{r_i} en el i-ésimo, $i \leqslant n$. Así, hay tantas combinaciones con repetición de n elementos tomados de i en i como maneras de escoger un sumando en cada paréntesis de manera que las potencias escogidas sumen i, o sea que $c_i = \binom{n+i-1}{i}$. De aquí que

$$(\sum_{i\geqslant 0} a^i)^n = (1+a+a^2+a^3+\cdots)^n = \sum_{i\geqslant 0} \binom{n+i-1}{i} a^i \tag{2.2}$$

Ecuaciones enteras

Directamente relacionado con los últimos ejemplos, se puede considerar el número de maneras diferentes en que se puede escribir un entero como suma ordenada de enteros no negativos, es decir, el número de soluciones de la ecuación

$$k = x_1 + x_2 + \cdots + x_n, \quad x_i \geqslant 0$$

con x_i entero no negativo. Por ejemplo, 4 se puede expresar como suma de tres enteros no negativos de las 15 maneras siguientes

$$\begin{array}{lll} 0+0+4 & 0+4+0 & 4+0+0 \\ 0+2+2 & 2+0+2 & 2+2+0 \\ 0+1+3 & 0+3+1 & 1+0+3 \\ 1+3+0 & 3+0+1 & 3+1+0 \\ 1+1+2 & 1+2+1 & 2+1+1 \end{array}$$

Este número se puede contar de la manera siguiente. Consideremos combinaciones con repetición de k elementos de $X = \{1,2,\ldots,n\}$. En cada combinación llamamos x_j al número de veces que aparece el elemento j. Entonces, $x_1 + x_2 + \cdots + x_n = k$, y cada una de las combinaciones corresponde a una solución de la ecuación. Además, dos combinaciones diferentes dan dos soluciones diferentes. En el ejemplo anterior, la combinación con repetición 2333 del conjunto $\{1,2,3\}$ corresponde a la solución $x_1 = 0$, $x_2 = 1$ y $x_3 = 3$. El número de soluciones es entonces

$$CR_n^k = \binom{n+k-1}{k}$$

Hay otra manera de obtener el mismo resultado. Ponemos k como la suma de k 1's. Cada expresión de k como suma de n enteros no negativos se corresponde con una agrupación de los

k unos en n grupos, cosa que se puede conseguir poniendo $n-1$ 'separaciones' en el grupo de k unos. Por ejemplo, en la lista anterior,

$$\begin{aligned} 4 = &\quad 1+1+2 \to \quad 1\bullet 1\bullet 11 \\ 4 = &\quad 1+0+3 \to \quad 1\bullet\bullet 111 \\ 4 = &\quad 0+2+2 \to \quad \bullet 11\bullet 11 \end{aligned}$$

Así, hay $n+k-1$ posiciones en las cuales es preciso poner $n-1$ separaciones y k unos. Esto se puede hacer de $\binom{n+k-1}{k}$ maneras.

Este problema admite diversas variaciones. Por ejemplo, el número de soluciones de la ecuación

$$r = y_1 + y_2 + \cdots + y_n, \quad y_i > 0$$

donde y_i son ahora enteros positivos y $r > n$ se puede obtener del problema anterior poniendo $x_i = y_i - 1$ (que será no negativo) y $k = r - n$, es decir, el número de soluciones vuelve a ser

$$\binom{n+k-1}{k} = \binom{r-1}{n-1}$$

Si hacemos ahora $z_i = x_i + a$, $1 \leqslant i \leqslant n$ con a un entero positivo, el número de soluciones de la ecuación

$$t = z_1 + z_2 + \cdots + z_n$$

con $z_i \geqslant a$ y $t > na$ es nuevamente

$$\binom{t-na+n-1}{n-1}$$

Ejercicio 2.1. ¿Cuál es el número de soluciones enteras de la ecuación

$$8 = x_1 + x_2 + x_3$$

de manera que x_i sean enteros entre 1 y 4?

Distribuciones

Para acabar esta pequeña lista de ejemplos consideremos el problema siguiente. Supongamos que tenemos n cajas numeradas y k bolas que se ponen en las cajas. El objetivo es contar cuántas disposiciones diferentes de las bolas en las cajas se pueden obtener atendiendo a diversos criterios.

Si cada caja puede contener como mucho una bola, y estas están numeradas (es decir, son distinguibles), el número de disposiciones diferentes es el de permutaciones de n elementos tomados de k en k, P_n^k (y tiene que ser $k \leqslant n$).

Si cada caja puede contener más de una bola, el número de disposiciones diferentes es el de permutaciones con repetición, PR_n^k.

Si las bolas no son distinguibles (sólo diferenciamos dos disposiciones por el número de bolas en cada caja), tenemos CR_n^k disposiciones diferentes, y en el caso de que cada caja pudiese contener sólo una bola, habría C_n^k bolas.

Este ejemplo tiene una aplicación interesante a la física de partículas. El estado macroscópico de un sistema de k partículas se identifica dando el nivel energético de cada una de ellas (y cada una puede estar en $n > k$ niveles). Así entonces, cada estado se corresponde con una manera de poner k bolas (las partículas) en n cajas (los niveles). En el modelo de Maxwell-Boltzmann, se asume que las partículas son distinguibles (no es lo mismo que la partícula 1 tenga nivel a y la 2 nivel b, que la 1 tenga nivel b y la 2 nivel a) de manera que hay PR_n^k estados diferentes. En el modelo de Bose-Einstein, se considera que las partículas no son distinguibles, de manera que el número de estados es CR_n^k. En el modelo de Fermi-Dirac se asume que las partículas son indistinguibles y que cumplen el principio de exclusión de Pauli, según el cual dos partículas diferentes no pueden estar en el mismo estado. El número de estados es entonces C_n^k. Cada una de estas hipótesis corresponde al comportamiento empírico de diferentes tipos de partículas subatómicas.

Ejercicio 2.2. ¿De cuántas maneras se pueden poner $k > n$ bolas en n cajas de manera que cada caja contenga al menos una bola si a) las bolas son distinguibles, b) las bolas no son distinguibles?

Ejercicio 2.3. ¿De cuántas maneras se pueden distribuir k tareas en n procesadores de manera que cada procesador tenga asignada como mucho una tarea?

2.3 Propiedades de los coeficientes binomiales

Como ya hemos comentado antes, los coeficientes binomiales aparecen con tanta frecuencia que resulta interesante conocer algunas de sus propiedades. En las demostraciones de estas propiedades usaremos siempre que sea posible argumentos combinatorios, aunque no sean, en muchos casos, la única vía de demostración. Hasta que no se indique lo contrario, supondremos siempre que, en la expresión $\binom{n}{k}$, n y k son enteros no negativos y que $k \leqslant n$.

Si tenemos en cuenta que $\binom{n}{k}$ cuenta el número de palabras de longitud n con exactamente k ceros y $n-k$ unos, está claro que la elección de la posición de los k ceros determina la de los

$n-k$ unos, de manera que

$$\binom{n}{k} = \binom{n}{n-k} \tag{2.3}$$

Esta es la *propiedad de simetría* de los coeficientes binomiales, y dice que la sucesión

$$\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n-1}, \binom{n}{n}$$

tiene una simetría central. Los primeros y últimos términos de la sucesión son $1, n, n(n-1)/2, \ldots$ y $\ldots, n(n-1)/2, n, 1$ respectivamente. Comparando dos términos consecutivos, tenemos

$$\frac{\binom{n}{k}}{\binom{n}{k+1}} = \frac{k+1}{n-k} \begin{cases} < 1 & \text{si } k < \lfloor \frac{n-1}{2} \rfloor \\ \geqslant 1 & \text{si } k \geqslant \lfloor \frac{n-1}{2} \rfloor \end{cases}$$

de manera que la sucesión anterior crece para $0 \leqslant k \leqslant \lfloor \frac{n-1}{2} \rfloor$ y decrece para $\lfloor \frac{n-1}{2} \rfloor + 1 \leqslant k \leqslant n$. Si n es par, el valor más grande es $\binom{n}{n/2}$, mientras que si n es impar, los términos más grandes son $\binom{n}{(n-1)/2} = \binom{n}{(n+1)/2}$.

Hay un algoritmo clásico para calcular $\binom{n}{k}$ para valores moderados de n que se llama *triángulo de Pascal* (o también *triángulo de Tartaglia*) y que se basa en la siguiente propiedad de los coeficientes binomiales:

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}, \quad 0 < k < n \tag{2.4}$$

Esta es la *propiedad de la adición* de los coeficientes binomiales. Desde el punto de vista combinatorio, la expresión se puede deducir de la manera siguiente. La familia de subconjuntos de tamaño k de un conjunto X de tamaño n se puede partir en dos subfamilias: la de los subconjuntos que *no* contienen un cierto elemento $a \in X$, de los cuales hay tantos como elecciones de k elementos entre los $n-1$ que quedan, $\binom{n-1}{k}$, y la de los que contienen a, tantos como elecciones de $k-1$ elementos entre los $n-1$ diferentes de a, $\binom{n-1}{k-1}$. De aquí la igualdad 2.4.

Usando la relación anterior, se pueden disponer los números combinatorios en un triángulo

$$\begin{array}{ccccccccc} & & & \binom{1}{0} & & \binom{1}{1} & & & \\ & & \binom{2}{0} & & \binom{2}{1} & & \binom{2}{2} & & \\ & \binom{3}{0} & & \binom{3}{1} & & \binom{3}{2} & & \binom{3}{3} & \\ \binom{4}{0} & & \binom{4}{1} & & \binom{4}{2} & & \binom{4}{3} & & \binom{4}{4} \end{array}$$

donde el primer y el último número de cada fila valen 1 y cada uno de los otros es la suma de los dos que tiene encima en la fila anterior. Numéricamente,

$$\begin{array}{ccccccc} & & 1 & & 1 & & \\ & 1 & & 2 & & 1 & \\ 1 & & 3 & & 3 & & 1 \\ 1 & & 4 & & 6 & & 4 & & 1 \end{array}$$

La identidad 2.4 permite obtener otras expresiones combinatorias. Usándola de forma iterada podemos escribir

$$\begin{aligned} \binom{n}{k} &= \binom{n-1}{k} + \binom{n-1}{k-1} = \\ &= \binom{n-1}{k} + \binom{n-2}{k-1} + \binom{n-2}{k-2} = \\ &= \binom{n-1}{k} + \binom{n-2}{k-1} + \binom{n-3}{k-2} + \binom{n-3}{k-3} \\ &= \binom{n-1}{k} + \binom{n-2}{k-1} + \cdots + \binom{n-k}{1} + \binom{n-k-1}{0} \end{aligned}$$

Esta igualdad se puede expresar de una manera más legible poniendo $n+k+1$ en el lugar de n:

$$\sum_{i=0}^{k} \binom{n+i}{i} = \binom{n}{0} + \binom{n+1}{1} + \cdots + \binom{n+k}{k} = \binom{n+k+1}{k} \tag{2.5}$$

que proporciona la suma de coeficientes binomiales contiguos en los cuales la parte superior y la inferior difieren siempre en una constante (n en la expresión anterior), motivo por el cual nos referiremos a ella como la *propiedad de la adición paralela*. Sobre el Triángulo de Pascal, esta suma es la de los términos de un segmento de diagonal como en la figura 2.2.

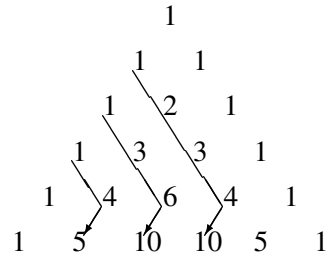


Figura 2.2: La propiedad de la adición paralela

Como el triángulo de Pascal tiene un eje de simetría central, se tendría que poder obtener una fórmula similar a la igualdad 2.5 aplicando esta simetría sobre las diagonales de la

figura 2.2. Efectivamente, de esta manera se obtiene la fórmula

$$\sum_{i=k}^{n}\binom{i}{k}=\binom{k}{k}+\binom{k+1}{k}+\cdots+\binom{n}{k}=\binom{n+1}{k+1} \tag{2.6}$$

que proporciona la suma de coeficientes binomiales contiguos en los que el índice superior sigue el índice de sumación y el inferior es constante, y que llamaremos *propiedad de la adición superior*. Sobre el Triángulo de Pascal, esta propiedad corresponde a sumas diagonales como las de la figura 2.3.

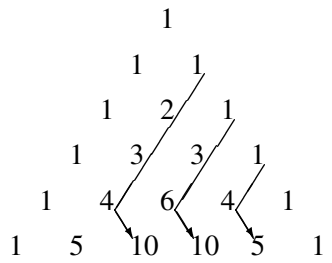


Figura 2.3: La propiedad de la adición superior

Ejercicio 2.4. Obtener la propiedad de la adición superior de las dos maneras siguientes.

1. Usando la propiedad de adición como para obtener la igualdad 2.4, pero descomponiendo el otro sumando.

2. Con el argumento combinatorio siguiente. Tenéis un conjunto de $n+1$ bolas numeradas de 0 a n y formáis la familia de subconjuntos de tamaño $k+1$. Esta familia se puede partir en las subfamilias formadas por los subconjuntos que tienen la bola más alta numerada i, $i=k,k+1,\ldots,n$.

Las dos propiedades anteriores son muy similares y tienen diversas aplicaciones particulares. Por ejemplo, para $n=1$ obtenemos,

$$\sum_{i=0}^{k-1}\binom{1+i}{i}=1+2+\cdots+(k-1)+k=\binom{k+1}{k-1}=\binom{k+1}{2}=\frac{k(k+1)}{2} \tag{2.7}$$

que porporciona una fórmula para la suma de los k primeros naturales. Los números que se obtienen, $\binom{k+1}{2}$, se llaman *números triangulares* porque cuentan el número de términos en las n primeras líneas del triángulo de Pascal (o en una disposición triangular de bolas). Los primeros son los de la figura 2.4.

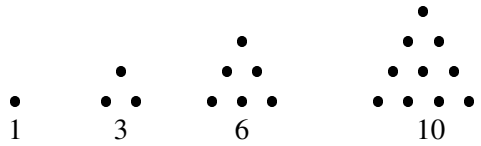


Figura 2.4: Representación geométrica de los primeros números triangulares

La fórmula 2.6 para $k = 2$ da la expresión

$$\sum_{i=2}^{n} \binom{i}{2} = \binom{2}{2} + \binom{3}{2} + \cdots + \binom{n}{2} = \binom{n+1}{3} = \frac{(n+1)n(n-1)}{6}$$

que corresponde a la suma de los n primeros números triangulares. Por similitud con éstos, los números $\binom{n+1}{3}$ se llaman *números piramidales*, ya que cuentan el número de elementos de una pirámide de base triangular y $n-1$ pisos, siendo el piso i un triángulo como los de la figura 2.4. La suma de números piramidales daría también el número de puntos de un objeto ... en cuatro dimensiones.

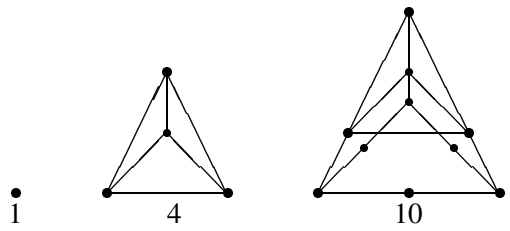


Figura 2.5: Representación geométrica de los primeros números piramidales

Expresiones más complejas involucran sumas de productos de coeficientes binomiales. La más conocida de estas expresiones es la *convolución de Vandermonde*,

$$\sum_{i=0}^{k} \binom{n}{i}\binom{m}{k-i} = \binom{n+m}{k} \tag{2.8}$$

que se deduce con el argumento combinatorio siguiente. Para obtener todos los subconjuntos de tamaño k de un conjunto de n bolas blancas y m bolas negras, podemos formar todos los que no tienen ninguna bola blanca (los hay $\binom{n}{0}\binom{m}{k}$), los que tienen una (los hay $\binom{n}{1}\binom{m}{k-1}$) y

así sucesivamente hasta los que tienen todas las bolas blancas (contando que, cuando $a < b$, entonces $\binom{a}{b} = 0$). Observemos que, para $m = 1$, reobtenemos la fórmula de adición.

Los coeficientes multinomiales se pueden obtener también como productos de coeficientes binomiales. Recordemos que

$$\binom{k}{k_1, \ldots, k_n} = \frac{k!}{k_1! \cdots k_n!}, \quad k = k_1 + \cdots + k_n$$

cuenta el número de permutaciones de n elementos en las cuales se repite k_i veces el elemento i. Para contar este número podíamos haber hecho lo siguiente. Escogemos primero las k_1 posiciones del elemento 1 (hay $\binom{k}{k_1}$ maneras de hacerlo). Después escogemos las k_2 posiciones del elemento 2 en las posiciones que quedan (de $\binom{k-k_1}{k_2}$ maneras) y así sucesivamente para obtener,

$$\binom{k}{k_1, \ldots, k_n} = \binom{k}{k_1}\binom{k-k_1}{k_2}\binom{k-k_1-k_2}{k_3}\cdots\binom{k_n}{k_n} \tag{2.9}$$

Por ejemplo,

$$\binom{k}{k_1, k_2} = \binom{k}{k_1}\binom{k-k_1}{k_2}$$

Ejercicio 2.5. Demostrar la identidad

$$\binom{k}{k_1}\binom{k_1}{k_2} = \binom{k}{k_2}\binom{k-k_2}{k_1-k_2}$$

La fórmula del binomio

$$(a+b)^n = \binom{n}{0}a^n b^0 + \binom{n}{1}a^{n-1}b^1 + \cdots + \binom{n}{n-1}a^1 b^{n-1} + \binom{n}{n}a^0 b^n$$

permite también obtener diversas expresiones con números combinatorios. Poniendo $a = b = 1$ obtenemos

$$\binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n-1} + \binom{n}{n} = 2^n \tag{2.10}$$

En términos de conjuntos, la identidad dice que un conjunto de n elementos tiene un total de 2^n subconjuntos, que es también el número de palabras de longitud n que se pueden formar con ceros y unos y también la suma de todos los coeficientes binomiales de una línea horizontal del Triángulo de Pascal.

Poniendo $a = 1$ y $b = -1$ en la fórmula binomial se obtiene

$$\binom{n}{0} - \binom{n}{1} + \cdots + (-1)^{n-1}\binom{n}{n-1} + (-1)^n\binom{n}{n} = 0 \tag{2.11}$$

es decir, que la suma con signos alternados de cada línea del Triángulo de Pascal da cero, otra consecuencia de su simetría central.

Acabamos esta sección con algunos comentarios sobre la fórmula del binomio. Tal como ha sido escrita y deducida, la fórmula sólo es válida para valores de n naturales. Lo que resulta más sorprendente es que, con una extensión adecuada (pero bien natural) de los coeficientes binomiales $\binom{x}{k}$ para valores reales de x, Newton obtuvo una fórmula del binomio válida para cualquier potencia (negativa, fraccionaria o irracional). En la definición del coeficiente binomial, podemos escribir, para n y k enteros no negativos,

$$\binom{n}{k} = \frac{n!}{(n-k)!\,k!} = \frac{n(n-1)\cdots(n-k+1)}{k!}$$

La parte derecha de esta igualdad permite definir $\binom{x}{k}$ para x real de la manera siguiente,

$$\binom{x}{k} = \frac{x(x-1)\cdots(x-k+1)}{k!} \qquad k \in \mathbb{N} \tag{2.12}$$

Así, por ejemplo,

$$\binom{-2}{3} = \frac{(-2)(-3)(-4)}{3\cdot 2\cdot 1} = -4$$

o

$$\binom{1/2}{3} = \frac{(1/2)(-1/2)(-3/2)}{3\cdot 2\cdot 1} = \frac{1}{2^4}$$

Observemos que si x es un entero menor que k, el numerador en la expresión 2.12 es cero. En esta definición, el miembro inferior del coeficiente binomial, k, es siempre un entero no negativo. Se puede generalizar la definición para todos los enteros poniendo

$$\binom{x}{k} = 0 \qquad k \text{ entero negativo}$$

Hay una relación precisa entre esta generalización de los coeficientes binomiales y la versión original cuando x es un entero negativo, $x = -n$. Esta relación, llamada *propiedad de inversión*, se obtiene de la manera siguiente,

$$\begin{aligned}\binom{-n}{k} &= \frac{(-n)(-n-1)\cdots(-n-k+1)}{k!} = \\ &= (-1)^k \frac{n(n+1)\cdots(n+k-1)}{k!} = \\ &= (-1)^k \binom{n+k-1}{k}\end{aligned}$$

La fórmula es, de hecho, simétrica, de manera que se puede escribir

$$\binom{r}{k} = (-1)^k \binom{k-r-1}{k} \qquad r \in \mathbb{Z} \tag{2.13}$$

Ejercicio 2.6. Hemos demostrado la fórmula anterior cuando r es un entero negativo. Demostrarla cuando r es un entero positivo.

Esta extensión de los coeficientes binomiales permite generalizar la fórmula del binomio a cualquier entero, positivo o negativo. Recordemos que en la fórmula 2.2 de la sección anterior habíamos obtenido,

$$(1+a+a^2+a^3+\cdots)^n = \sum_{i\geqslant 0} \binom{n+i-1}{i} a^i$$

Cada uno de los factores es la suma de una serie geométrica de razón a. Si $0 < a < 1$, esta suma vale $\frac{1}{1-a}$, de manera que, usando la ecuación anterior, se puede escribir

$$(1+a)^{-n} = \left(\frac{1}{(1-(-a))}\right)^n = \sum_{k\geqslant 0} \binom{n+k-1}{k} a^k (-1)^k$$

donde ahora el sumatorio de la derecha tiene una cantidad no finita de términos. Usando la fórmula de inversión, podemos escribir

$$(1+a)^r = \sum_{k\geqslant 0} \binom{r}{k} a^k \qquad r \in \mathbb{Z} \tag{2.14}$$

que proporciona una generalización de la fórmula del binomio para cualquier entero r. Si r es positivo, los términos del sumatorio con el índice inferior k más grande que r son nulos y el sumatorio tiene de hecho $r+1$ términos, mientras que si r es negativo, el sumatorio tiene infinitos términos.

Se puede demostrar (usando el desarrollo en serie de Taylor de la función $f(x) = (1+x)^r$ alrededor del origen) que la fórmula 2.14 es válida para cualquier valor real de r. En un capítulo posterior, al tratar funciones generadoras, usaremos esta fórmula y comentaremos qué papel juega la convergencia de la serie que se obtiene.

Notas bibliográficas

Desde un punto de vista histórico, el desarrollo inicial de la combinatoria elemental se produjo con el nacimiento del cálculo de probabilidades, a finales del siglo XVII y en relación sobre todo a problemas relacionados con los juegos de azar. Una de las primeras obras que sistematiza los

primeros resultados en el cálculo de probabilidades, el *Ars conjectandi* de J. Bernouilli (1713), contiene ya una expresión de la fórmula del binomio que después generalizará Newton. Este origen hace que en muchos textos de probabilidad haya una buena introducción a esta parte de la combinatoria. Un ejemplo excelente en este sentido es el del libro de Feller [3].

El material cubierto en este capítulo se puede encontrar en cualquier texto elemental de combinatoria o de matemática discreta. El libro de Anderson [1] contiene una exposición a nivel sencillo, mientras que en el de Berge [2] se hace una exposición más compacta. En lo que respecta a las propiedades de los números combinatorios, el texto de Graham, Knuth y Patashnik [4] es una referencia completa y de lectura agradecida.

Bibliografía

[1] I. Anderson. *Introducción a la Combinatoria*, Ed. Vicens Vives, 1992.

[2] C. Berge. *Principes de Combinatorie*, Dunod, Paris, 1968.

[3] W. Feller. *An Introduction to Probability Theory and its Applications* (Vol. I), Wiley Interscience, 1968.

[4] R. Graham, D. E. Knuth, O. Patashnik. *Concrete Mathematics*, Addison Wesley, 1989.

Problemas

1. Queremos codificar los símbolos alfanuméricos (28 letras y 10 cifras) en palabras de una cierta longitud k de un alfabeto binario $A = \{0, 1\}$. ¿Cuál es la mínima longitud necesaria para poderlo hacer?

2. ¿Es suficiente con palabras de hasta 4 de longitud para representar todas las letras del alfabeto ordinario en lenguaje Morse (el lenguaje Morse dispone sólo de dos símbolos: punto y raya)?

3. ¿De cuántas maneras se pueden escoger tres números del 1 al 9 de manera que no salgan dos consecutivos?

4. Las placas de matrícula tienen cuatro dígitos numéricos seguidos de dos alfabéticos. ¿Cuántos coches se pueden matricular? Una vez se han agotado, se propone que las matrículas puedan estar formadas por seis dígitos alfanuméricos (es decir, cifras del 0 al 9 o letras de la A a la Z). ¿Cuántas matrículas nuevas se pueden hacer? Una vez agotadas éstas, ¿qué estrategia proporcionaría más matrículas nuevas, hacer matrículas con 7 dígitos, o bien añadir un símbolo al alfabeto?

5. ¿Cuántos números hay entre 100 y 900 que tengan las cifras diferentes? ¿Cuántos números más grandes que 6600 con todas las cifras diferentes y sin ninguna de las cifras 7, 8 ni 9?

6. ¿Cuántas palabras de longitud 4 se pueden formar con las cinco vocales sin que se repita ninguna? ¿Y de longitud 5 (también sin que se repita ninguna)?

7. Un código de colores con barras usa 6 colores para pintar 4 barras, pero dos barras consecutivas no pueden tener el mismo color. ¿Cuántas palabras diferentes se pueden formar?

8. En un alfabeto de 10 consonantes y 5 vocales, ¿cuántas palabras de cinco letras sin dos vocales seguidas ni tres consonantes seguidas se pueden formar?

9. La música serial se basa en el principio de que en cualquier línea melódica han de aparecer los 12 tonos de la escala antes de repetirse alguno. ¿Cuántas líneas melódicas de 12 notas se pueden formar según este principio?

10. En problemas de diseño de redes de interconexión se suelen usar grafos que tienen por vértices palabras de un alfabeto. Por ejemplo, los llamados grafos de Kautz tienen por vértices las palabras de longitud k qur se pueden formar de un alfabeto de n símbolos con la condición que dos letras consecutivas no pueden ser iguales. ¿Cuántas de estas palabras hay?

11. Sea $A = \{1, 2, \ldots, n\}$ y $X = \{x_1, x_2, \ldots, x_k\}$ un conjunto de k símbolos. Una aplicación $f : X \to A$ es *ordenada* si $f(x_1) \leqslant f(x_2) \leqslant \cdots \leqslant f(x_k)$ y *estrictamente ordenada* si las desigualdades son estrictas. ¿Cuántas aplicaciones ordenadas y cuántas estrictamente ordenadas hay de X en A?

12. En una reunión de una empresa hay ocho representantes de los accionistas, seis representantes de acreedores, cuatro representantes de los trabajadores y tres técnicos. Para resolver más ágilmente la organización de la reunión deciden nombrar una comisión formada por tres representantes de los accionistas, dos representantes de acreedores, un representante de los trabajadores y un técnico. ¿Cuántas comisiones diferentes se podrían formar? Si uno de los accionistas se niega en rotundo a formar parte de la comisión con dos de los representantes de los trabajadores, a los cuales tiene manía, ¿cuántas comisiones se podrían formar?

13. Una empresa de sondeos escoge una muestra de 20 estudiantes al azar de entre una comunidad de 500 estudiantes para hacer una encuesta. ¿Cuántas muestras diferentes puede obtener? Uno de los estudiantes está encantado de que le pasen la encuesta. ¿Cuántas de estas muestras contienen a este estudiante?

14. ¿De cuántas maneras se pueden poner n bolas numeradas en k cajas numeradas de manera que en cada caja haya al menos una bola? ¿Y si las bolas no están numeradas?

15. Usando la propiedad de inversión, demostrar que la suma parcial de los términos de una línea del Triángulo de Pascal con los signos alternados vale

$$\sum_{i=0}^{k}(-1)^i\binom{n}{i}=(-1)^k\binom{n-1}{k}$$

16. Demostrar la identidad

$$\sum_{i=0}^{k}\binom{k+r}{i}x^iy^k-i=\sum_{i=0}^{k}\binom{-r}{i}(-x)^i(x+y)^{k-i}$$

Usando esta identidad para $x=-1$ e $y=1$, o bien, $x=y=1$ y $r=k+1$, obtener, respectivamente,

$$\begin{aligned}\sum_{i=0}^{k}\binom{k+r}{i}(-1)^k &= \binom{-r}{k}, \quad k \text{ entero positivo},\\ \sum_{i=0}^{k}\binom{2k+1}{k} &= \sum_{i=0}^{k}\binom{k+i}{i}2^{k-i}=2^k\end{aligned}$$

Capítulo 3

Principios básicos de enumeración

1. Cardinales de conjuntos
2. Principio de inclusión-exclusión
3. Biyecciones
4. Principio del palomar y teorema de Ramsey

En este capítulo se describen y desarrollan algunos principios básicos de enumeración que, a pesar de ser de una gran simplicidad, se convierten en herramientas sistemáticas útiles en problemas combinatorios cuyo uso puede llegar a ser importante. Algunos de estos principios se han usado de manera implícita en el capítulo anterior.

En la primera sección se describen algunas relaciones entre operaciones entre conjuntos y sus cardinales. Una de estas relaciones, la del cardinal de la unión de conjuntos, da lugar al llamado *principio de inclusión-exclusión*, que se describe en la sección 2. Como ejemplo de una aplicación relativamente sofisticada de este principio, se tratan la función ϕ *de Euler* y el llamado problema de los *desarreglos*. En la sección 3 se da una ilustración de cómo se pueden relacionar estructuras aparentemente inconexas para enumerarlas más fácilmente. El ejemplo nos lleva a introducir los llamados *números de Catalan*, que tienen un interés combinatorio considerable. En el mismo contexto se introduce también el problema de las *particiones* de un entero positivo. En la última sección se desarrolla otro principio elemental, el llamado *principio del palomar*. En este caso no se trata de resolver un problema de enumeración, sino la existencia de una determinada configuración o propiedad combinatoria a través de hipótesis muy generales. Como aplicación de este principio simple se dan dos resultados clásicos, el teorema de Erdős-Szekeres y el teorema de Ramsey.

3.1 Cardinales de conjuntos

Los problemas de enumeración son de hecho problemas de cálculo de cardinales de conjuntos finitos. Por esto resulta interesante conocer cómo se traducen las operaciones usuales entre conjuntos en operaciones aritméticas entre los respectivos cardinales. En esta sección veremos algunas situaciones sencillas en las cuales esta traducción es posible.

Si A es un conjunto finito, $|A|$ denota el número de elementos, o cardinal, de A.

Como consecuencia directa de la definición de unión de conjuntos, tenemos:

Principio de adición. Sean A y B dos conjuntos finitos disyuntos, $A \cap B = \emptyset$. Entonces

$$|A \cup B| = |A| + |B|$$

Este resultado se extiende por inducción al caso de r conjuntos si son disyuntos dos a dos: Si $A_1, A_2, \ldots, A_k$ son conjuntos finitos y $A_i \cap A_j = \emptyset$ para cualquier par de índices $i, j \in \{1, 2, \ldots, k\}$, entonces

$$|A_1 \cup A_2 \cup \cdots \cup A_k| = |A_1| + |A_2| + \cdots + |A_k|$$

Para la intersección no hay una relación general que ligue $|A \cap B|$ con $|A|$ y $|B|$. Sí que la hay en cambio para la complementación. Si Ω es un conjunto finito y $A \subset \Omega$, denotamos por $\Omega \setminus A$ el complemento de A en Ω (es decir, el conjunto de elementos de Ω que no están en A).

Proposición 3.1. Sea Ω un conjunto finito y $A \subset \Omega$. Entonces,

$$|\Omega \setminus A| = |\Omega| - |A|$$

Demostración. Podemos escribir Ω como la unión disyunta $A \cup (\Omega \setminus A)$, de manera que, por el principio de adición, $|\Omega| = |A| + |\Omega \setminus A|$. □

Esta última proposición se usa generalmente cuando resulta más sencillo calcular el cardinal del complemento de un subconjunto que el del propio subconjunto. Por ejemplo, el número de palabras de cinco letras que contienen al menos una vocal se calcula fácilmente como el total de palabras menos las que no tienen ninguna vocal, $5^{27} - 5^{22}$.

Una de las operaciones entre conjuntos que admite una traducción directa en términos de cardinales es la del producto cartesiano.

Proposición 3.2. Sean A y B dos conjuntos finitos y $A \times B$ su producto cartesiano. Entonces,

$$|A \times B| = |A| \cdot |B|$$

Demostración. El producto cartesiano se puede escribir como la unión disyunta $A \times B = \cup_{a \in A} \{a\} \times B$, donde cada uno de los términos de la unión tiene cardinal $|B|$ y los hay tantos como elementos de A. □

La proposición anterior se puede extender por inducción a cualquier número finito de factores. Si $A_1, \ldots, A_r$ son conjuntos finitos, entonces

$$|A_1 \times \cdots \times A_r| = |A_1| \cdots |A_r|$$

Observemos que, si $A = \{x_1, \ldots, x_n\}$, el producto cartesiano r veces $A \times \cdots \times A$ tiene por elementos las permutaciones con repetición de los elementos de A tomados de r en r, de manera que reencontramos el resultado

$$PR_n^r = |\underbrace{A \times \cdots \times A}_{r}| = |A|^r = n^r$$

El resultado de la proposición anterior se conoce a veces como la *regla del producto* y se expresa diciendo que, si para realizar un proceso en dos etapas hay n_1 maneras de hacer la primera y, para cada una de ellas, n_2 de realizar la segunda, entonces el número total de maneras de realizar el proceso es el producto $n_1 n_2$. Si llamamos A al conjunto de maneras de realizar la primera etapa y B al conjunto de maneras de realizar la segunda, el conjunto de maneras de realizar el proceso es $A \times B$. Este es el procedimiento que se ha usado en el capítulo anterior para calcular el número de permutaciones. Por ejemplo, las permutaciones de n elementos sin repetición se pueden formar en un proceso de n etapas. En la primera hay n elecciones posibles del primer elemento, en la segunda $n-1$ elecciones, y así sucesivamente hasta la última que admite una única opción, de manera que el número total de permutaciones es el producto $n!$.

3.2 Principio de inclusión-exclusión

En la sección anterior se ha relacionado el cardinal de la unión de dos conjuntos disyuntos con el cardinal de cada uno de ellos a través de la igualdad $|A \cup B| = |A| + |B|$. Cuando los conjuntos no son disyuntos se puede obtener una fórmula que hace intervenir el cardinal de la intersección.

Proposición 3.3. Sean A y B dos conjuntos finitos. Entonces,

$$|A \cup B| = |A| + |B| - |A \cap B|$$

Demostración. La unión $A \cup B$ se puede poner como la unión disyunta,

$$A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$$

donde $|A \setminus B| = |A| - |A \cap B|$ y $|B \setminus A| = |B| - |B \cup A|$. Usando ahora el principio de adición, se obtiene la expresión del enunciado. □

El llamado *principio de inclusión-exclusión* es una extensión de este resultado al caso de la unión de n conjuntos y tiene una expresión un poco más compleja. La deducimos primero para el caso de la unión de tres conjuntos, $A \cup B \cup C$. En la figura 3.1 está representada la situación con diagramas de Venn.

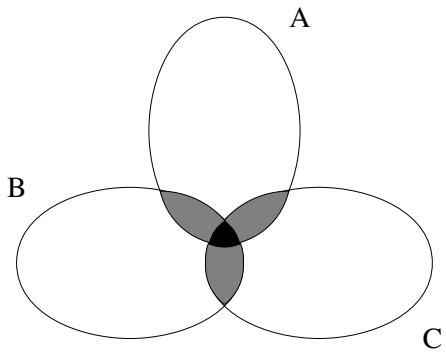


Figura 3.1: La unión $A \cup B \cup C$

Si los tres conjuntos fuesen disyuntos, tendríamos $|A \cup B \cup C| = |A| + |B| + |C|$. Si no lo son, en la expresión $|A| + |B| + |C|$ se cuentan dos veces los elementos que están en la intersección de sólo dos de los tres subconjuntos (los de la zona cuadriculada en la figura) y tres veces los elementos de $A \cap B \cap C$ (los de la zona rayada de la figura). Restando los cardinales de $A \cap B$, $A \cap C$ y $B \cap C$, habremos descontado una vez los elementos que están en la zona cuadriculada, pero también habremos descontado tres veces los elementos que están en $A \cap B \cap C$. Sumando una vez el cardinal de esta zona rayada obtenemos,

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$

La fórmula general de inclusión-exclusión tiene un aspecto similar y se puede razonar de la misma manera, aunque aquí haremos una demostración diferente. El nombre de inclusión-exclusión proviene del hecho de que en la expresión de la derecha hay elementos que se van incluyendo y excluyendo alternativamente.

Proposición 3.4 (Principio de inclusión-exclusión). Sean $A_1, A_2, \ldots, A_n$ conjuntos finitos.

Entonces,

$$\begin{aligned} |A_1 \cup A_2 \cup \cdots \cup A_n| &= \sum_{i=1}^{n} |A_i| - \sum_{i<j} |A_i \cap A_j| + \cdots + (-1)^{r-1} \sum_{i_1<\cdots<i_r} |A_{i_1} \cap \cdots \cap A_{i_r}| \\ &\quad + \cdots + (-1)^{n-1} |A_1 \cap A_2 \cap \cdots \cap A_n| \end{aligned} \tag{3.1}$$

Demostración. Sólo es preciso comprobar que cada elemento de la unión se cuenta una sola vez en la expresión de la derecha de la igualdad. Supongamos que $x \in A_1 \cup \cdots \cup A_n$ es un elemento que está exactamente en $p \leqslant n$ de los n conjuntos, $x \in A_{i_1} \cap \cdots \cap A_{i_p}$. Entonces x está contado $p = \binom{p}{1}$ veces en el primer sumando $\sum_{i=1}^{n} |A_i|$. En el segundo sumando, $\sum_{i<j} |A_i \cap A_j|$, está contado una vez para cada elección de dos de los subconjuntos $A_{i_1}, \ldots, A_{i_p}$, es decir, $\binom{p}{2}$ veces. En general, en el sumando r-ésimo el elemento x está contado $\binom{p}{r}$ veces mientras $r \leqslant p$ y ninguna vez si $r > p$. Por tanto, el elemento x está contado

$$\binom{p}{1} - \binom{p}{2} + \cdots + (-1)^{r-1} \binom{p}{r} + \cdots + (-1)^{p-1} \binom{p}{p}$$

De acuerdo con la fórmula 2.11 de las propiedades de los números binomiales que hemos deducido en el capítulo anterior, esta expresión vale $\binom{p}{0} = 1$. □

Ejercicio 3.5. Demostrar el principio de inclusión-exclusión por inducción sobre el número de conjuntos en la unión.

Desde el punto de vista de la notación, el principio de inclusión-exclusión se puede expresar de una manera más compacta, pero también más críptica. Sea $K = \{1, 2, \ldots, n\}$ el conjunto de los subíndices. Para cada $T \subset K$ llamamos $N(T) = |\cap_{i \in T} A_i|$. Entonces, la fórmula 3.1 se puede escribir

$$|A_1 \cup A_2 \cup \cdots \cup A_n| = \sum_{T \subset K} (-1)^{|T|+1} N(T) \tag{3.2}$$

donde el sumatorio se extiende a todos los subconjuntos de K.

A continuación veremos algunos ejemplos de aplicación de este principio.

La criba de Eratóstenes

El principio de inclusión-exclusión fue originalmente formulado en términos diferentes por Sylvester (1800–1850) inspirado en un método atribuído a Eratóstenes (siglo V aC) para encontrar los números primos. El método de Eratóstenes consiste en hacer una lista de números naturales de 1 a n. Se comienza marcando el 1 y todos los números pares (múltiplos de 2)

más grandes que 2. A continuación se busca el primer número sin marca (el 3) y se marcan todos los múltiplos de 3 más grandes que 3. Se busca el primer número sin marca (el 5) y se marcan todos los múltiplos de 5 más grandes que 5. Prosiguiendo de esta manera hasta agotar los números de la lista, todos los que no llevan marca son números primos. Para encontrar los números primos del 1 al 20, las fases del proceso serían,

$$\boxed{1}\ 2\ 3\ \boxed{4}\ 5\ \boxed{6}\ 7\ \boxed{8}\ 9\ \boxed{10}\ 11\ \boxed{12}\ 13\ \boxed{14}\ 15\ \boxed{16}\ 17\ \boxed{18}\ 19\ \boxed{20}$$

$$\boxed{1}\ 2\ 3\ \boxed{4}\ 5\ \boxed{6}\ 7\ \boxed{8}\ \boxed{9}\ \boxed{10}\ 11\ \boxed{12}\ 13\ \boxed{14}\ \boxed{15}\ \boxed{16}\ 17\ \boxed{18}\ 19\ \boxed{20}$$

El proceso continúa marcando los múltiplos de $5, 7, 11, 13, 17$ y 19, que no producen nuevas marcas en la lista, de manera que los números sin marcas en la última lista son números primos.

La versión de Sylvester es una generalización de este método. Supongamos que los N elementos de un conjunto Ω pueden tener hasta k propiedades diferentes $p_1, \ldots, p_k$ (en el método de Eratóstenes las propiedades son 'ser divisible por 2', 'ser divisible por 3', etc.). Supongamos que queremos contar cuántos de los elementos *no* tienen ninguna de las propiedades. Si $T = \{i_1, \ldots, i_r\} \subset K = \{1, 2, \ldots, k\}$, llamamos $N(T)$ al número de elementos que tienen las propiedades $p_{i_1}, \ldots, p_{i_r}$. El número de elementos que no tiene ninguna de las propiedades es $N(\emptyset)$. Entonces,

$$N(\emptyset) = N - \sum_{T \subset K} (-1)^{|T|} N(T) \tag{3.3}$$

Esta expresión se obtiene de la proposición 3.4 simplemente llamando A_i al conjunto de los elementos que tienen la propiedad p_i. Entonces, el número que se busca es el cardinal del conjunto de elementos que no están en ninguno de los A_i, es decir,

$$N(\emptyset) = |\Omega \setminus (A_1 \cup \cdots \cup A_k)| = |\Omega| - |A_1 \cup \cdots \cup A_k|$$

Esta es la forma con que se enuncia habitualmente el principio de inclusión-exclusión. Enunciado de esta manera se puede dar la generalización siguiente. Para cada $r = 1, \ldots, k$, llamamos $N_r = \sum_{T \subset K, |T| = r} N(T)$. Entonces:

Proposición 3.6. El número de elementos de un conjunto X que tienen exactamente r de las propiedades $p_1, \ldots, p_k$ es

$$N(r) = N_r - N_{r+1} + \cdots + (-1)^{k-r} N_k$$

Ejercicio 3.7. Demostrar esta última proposición.

Ejercicio 3.8. ¿Cuántos números hay entre 1 y 1000 que no sean divisibles ni por 3 ni por 7?

En particular, si para $r = 0$ definimos $N_0 = N$, la proposición 3.6 permite tener la fórmula de inclusión-exclusión expresada aún de otra manera,

$$N(0) = N - N_1 + \cdots + (-1)^r N_r + \cdots + (-1)^k N_k \tag{3.4}$$

Ejercicio 3.9. Demostrar que los r primeros sumandos de la ecuación 3.4 proporcionan una cota *superior* de $N(0)$ si r es impar y una cota *inferior* si r es par.

La función ϕ de Euler

Un problema relacionado con el anterior es el de contar cuántos números hay entre 1 y n que sean relativamente primos con n, es decir, que no tengan ningún divisor diferente del 1 en común con n. La función que da este número para cada n se llama *función* ϕ *de Euler* y juega un papel importante en la teoría de números. Por ejemplo, los primeros valores de ϕ son $\phi(1) = \phi(2) = 1$, $\phi(3) = \phi(4) = 2$ y $\phi(5) = 4$. Obtendremos ahora una expresión de $\phi(n)$ usando el principio de inclusión-exclusión. Un resultado básico de aritmética (el teorema de factorización) establece que cada número admite una descomposición única como producto de números primos,

$$n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_k^{\alpha_k}$$

Cualquier divisor de n debe ser dividido por uno o más de los números primos que aparecen en su descomposición. Llamamos A_i al conjunto de números entre 1 y n divisibles por p_i. Los números relativamente primos con n no tienen divisores comunes con n, de manera que son precisamente los que no están en ninguno de los conjuntos A_i. Así entonces,

$$\phi(n) = n - |A_1 \cup A_2 \cup \cdots \cup A_k|$$

donde el cardinal de esta unión se puede calcular usando el principio de inclusión-exclusión. Para ello, observemos que los números entre 1 y n divisibles por p son $p, 2p, 3p, \ldots$ hasta llegar a n y los hay por tanto n/p. Así entonces,

$$\begin{aligned} |A_i| &= \frac{n}{p_i}, \\ |A_i \cap A_j| &= \frac{n}{p_i p_j}, \quad i \neq j \\ &\vdots \quad\quad \vdots \\ |A_1 \cap A_2 \cap \cdots \cap A_k| &= \frac{n}{p_1 p_2 \cdots p_k} \end{aligned}$$

Entonces,

$$\phi(n) = n - \sum_i \frac{n}{p_i} + \sum_{i<j} \frac{n}{p_i p_j} - \cdots + (-1)^k \frac{n}{p_1 p_2 \cdots p_k} \tag{3.5}$$

Esta fórmula se expresa habitualmente de una manera más compacta como un producto en lugar de una suma,

$$\phi(n) = n\left(1-\frac{1}{p_1}\right)\left(1-\frac{1}{p_2}\right)\cdots\left(1-\frac{1}{p_k}\right) \tag{3.6}$$

Ejercicio 3.10. Demostrar que efectivamente las expresiones 3.5 y 3.6 son equivalentes.

Ejercicio 3.11. Obtener la expresión de $\phi(n)$ cuando *i)* n es un número primo, *ii)* n es un producto de dos primos.

Ejercicio 3.12. Demostrar que si n y n' son enteros relativamente primos, entonces

$$\phi(nn') = \phi(n)\phi(n')$$

Una de las propiedades interesantes de la función ϕ es la siguiente:

Proposición 3.13. Sea D el conjunto de divisores de un número natural n. Entonces,

$$n = \sum_{d\in D}\phi(d)$$

Demostración. Si $X = \{1,2,\ldots,n\}$, llamamos $A_i = \{r \in X : \mathsf{mcd}(r,n) = i\}$ (recordemos que mcd indica el máximo común divisor). Está claro que si i no divide a n entonces $A_i = \emptyset$, de manera que $X = \cup_{d\in D}A_d$ y que esta unión es disyunta. Así entonces,

$$n = |X| = \sum_{d\in D}|A_d|$$

Observemos que, para cada divisor $d \in D$, los elementos de A_d tienen la forma md, donde m es relativamente primo con n y $1 \leqslant m \leqslant n/d$. Por lo tanto, $|A_d| = \phi(n/d)$. Observemos finalmente que $\sum_{d\in D}\phi(n/d) = \sum_{d\in D}\phi(d)$. □

Desarreglos

Una de las aplicaciones clásicas del principio de inclusión-exclusión es el llamado problema de los desarreglos. Una vez, un oficinista tenía que poner diez cartas dirigidas a diez clientes en sus respectivos sobres; como tenía mucha prisa, puso cada carta en un sobre sin mirar si coincidía el domicilio y resultó que ninguno de los clientes recibió la carta que le iba dirigida. El hombre pensó que había sido verdadera mala suerte no adivinar ni una. Tener mala suerte querría decir que, de todas las posibilidades, había relativamente pocas de que sucediese este completo desastre. En esta sección procuraremos evaluar la mala suerte del oficinista.

Consideremos la selección ordenada $12\ldots n$ de n símbolos y una permutación $a_1a_2\ldots a_n$ de esta selección. Diremos que esta permutación es un *desarreglo* de la selección original si

ningún elemento está en su sitio. Por ejemplo, 4123 es un desarreglo de 1234, mientras que 1423 no lo es, ya que 1 está en su sitio. El problema que consideraremos es el de contar cuántos desarreglos se pueden formar de $12\ldots n$.

Llamaremos A_i al conjunto de todas las permutaciones de $12\ldots n$ que dejan el elemento i en su sitio. Entonces, el número D_n de desarreglos es el número de permutaciones que no están en ninguno de los conjuntos A_i, es decir,

$$D_n = n! - |A_1 \cup A_2 \cup \cdots \cup A_n|$$

Otra vez, el cardinal de este último conjunto se puede calcular por el principio de inclusión-exclusión. Observemos que $|A_i| = (n-1)!$, ya que, si el elemento i se tiene que quedar en su sitio, nos quedan las permutaciones de los otros $n-1$ elementos. De manera similar, para cada una de las $\binom{n}{2}$ elecciones de pares i, j, $|A_i \cap A_j| = (n-2)!$. En general, para cada una de las $\binom{n}{p}$ elecciones de p índices, $|A_{i_1} \cap \cdots \cap A_{i_p}| = (n-p)!$. Por tanto,

$$D_n = n! - \binom{n}{1}(n-1)! + \binom{n}{2}(n-2)! + \cdots + (-1)^p \binom{n}{p}(n-p)! + \cdots + (-1)^n \binom{n}{n}$$

Esta expresión se puede escribir de una manera más sencilla como

$$D_n = n!\left[1 - \frac{1}{1!} + \frac{1}{2!} + \cdots + (-1)^p \frac{1}{p!} + \cdots + (-1)^n \frac{1}{n!}\right]$$

Observemos que, escrita de esta manera, la expresión entre corchetes es la suma parcial n-ésima de

$$\frac{1}{e} = 1 - \frac{1}{1!} + \frac{1}{2!} + \cdots + (-1)^p \frac{1}{p!} + \cdots + (-1)^n \frac{1}{n!} + \cdots = \sum_{i \geqslant 0} \frac{(-1)^i}{i!}$$

Las primeras sumas parciales de esta serie numérica son

$$1, 0, 0.5, 0.33, 0.375, 0.366, 0.3680, 03678, \ldots$$

que convergen rápidamente hacia $1/e$. De esta manera, para n 'grande', se puede poner que D_n es aproximadamente igual a $n!/e$. Así, para valores de n grandes, la proporción de desarreglos sobre el total de permutaciones es aproximadamente $1/e$. Para el caso de los 10 sobres del oficinista, hay 1.334.961 desarreglos entre las $10! = 3.628.800$ permutaciones posibles, de manera que la proporción de desarreglos es 0.36787946 (mientras que $1/e$ es 0.36787944...). Esto puede ser una medida de su mala suerte: aproximadamente un 37% de las posibilidades son desarreglos.

Ejercicio 3.14. ¿Cuántas permutaciones de $12\ldots n$ hay que tengan el 1 y el 2 en su sitio? ¿Cuántas permutaciones de $12\ldots n$ hay en las cuales exactamente r elementos están en su sitio?

Ejercicio 3.15. Un tarotólogo afirma tener poderes telepáticos. Para probarlo pide que se reordenen seis cartas sobre la mesa y dice que adivinará cuáles son con los ojos cerrados. Si adivina 3 de las 6 cartas, ¿se puede decir que ha sido por azar o esto demuestra efectivamente una habilidad especial?

3.3 Biyecciones. Números de Catalan. Particiones

En esta sección veremos otro principio elemental de enumeración que, como los que hemos visto anteriormente, permite edificar técnicas relativamente sofisticadas desde una observación casi trivial.

Una manera de contar los elementos de un conjunto es relacionarlos con los de un conjunto que ya sabemos contar. En particular, si se puede describir una biyección $\phi : A \to B$ entre los conjuntos A y B, los dos conjuntos tienen el mismo número de elementos. En esta sección desarrollaremos aplicaciones de este principio.

La construcción de biyecciones para conseguir enumerar los elementos de un conjunto se ha usado anteriormente de manera implícita en diversas ocasiones. Recordemos, por ejemplo, como se obtenía en el capítulo anterior la fórmula de enumeración de las combinaciones con repetición de n elementos tomados de k en k, CR_n^k. Si A es el conjunto de estas combinaciones, establecíamos una biyección de A en el conjunto B de todas las secuencias de longitud $n+k-1$ con exactamente n ceros y $k-1$ unos. Esta biyección estaba definida con la regla

$$\begin{array}{c} \underbrace{11\ldots1}_{k_1}\underbrace{22\ldots2}_{k_2}\ldots\underbrace{nn\ldots n}_{k_n} \\ \updownarrow \\ \underbrace{00\ldots0}_{k_1}\,1\,\underbrace{00\ldots0}_{k_2}\,1\,\ldots 1\,\underbrace{00\ldots0}_{k_n} \end{array}$$

A continuación se establecía una biyección entre B y el conjunto C de todos los subconjuntos de tamaño $k-1$ de un conjunto $X = \{x_1, \ldots, x_{n+k-1}\}$ de $n+k-1$ elementos de acuerdo con la asociación

$$\begin{array}{ccc} \alpha = \alpha_1\alpha_2\ldots\alpha_{n+k-1} & \qquad & \alpha_i \in \{0,1\}, \quad \sum_i \alpha_i = k-1 \\ \updownarrow & & \\ X_\alpha = \{x_{i_1}, x_{i_2}, \ldots, x_{i_{k-1}}\} \subset X, & & x_i \in X_\alpha \Leftrightarrow \alpha_i = 1 \end{array}$$

Este último conjunto tiene $\binom{n+k-1}{k-1}$ elementos, de manera que obteníamos el número CR_n^k de combinaciones con repetición de n elementos tomados de k en k como

$$CR_n^k = |A| = |B| = |C| = \binom{n+k-1}{k-1}$$

En esta sección ilustraremos el uso de esta técnica a través de dos ejemplos. En el primero introduciremos otros números combinatorios importantes, los *números de Catalan*, que permiten resolver algunos problemas de enumeración interesantes. El segundo trata el problema de las *particiones* de un entero. Estos dos temas volverán a ser tratados en el capítulo siguiente.

Números de Catalan

En una expresión aritmética con paréntesis debe haber tantos paréntesis abiertos como cerrados. Además, no se puede cerrar un paréntesis que no se haya abierto antes. Prescindiendo de todo lo que no sean paréntesis, una expresión como

$$()(())$$

sería admisible, mientras que

$$)(())($$

no tiene corrección sintáctica.

Dado un cierto número $2n$ de paréntesis (n abiertos y n cerrados), contaremos cuántas expresiones correctas se pueden formar, entendiendo por correctas aquellas en que no se cierra un paréntesis que no se haya abierto. Llamamos A al conjunto de estas expresiones. Será más cómodo trabajar identificando un paréntesis abierto con 1 y un paréntesis cerrado con -1. Entonces, se puede establecer una biyección entre A y el conjunto B de todas las selecciones ordenadas de 1's y -1's de longitud $2n$. La corrección sintáctica de los paréntesis se traduce en las secuencias $x_1, x_2, \ldots, x_{2n}$, $x_i \in \{1,-1\}$ en que la suma de los primeros k dígitos no es nunca negativa y la suma de todos vale cero:

$$B = \{x_1, x_2, \ldots, x_{2n} : x_i \in \{-1,1\}, \quad \sum_{i=1}^{k} x_i \geqslant 0, \quad \sum_{i=1}^{2n} x_i = 0\}$$

Para visualizar el problema de enumeración que queremos resolver, representaremos cada uno de los elementos de B como un camino de $(0,0)$ a $(2n,0)$ en $\mathbb{Z} \times \mathbb{Z}$ formado uniendo (i,j) con $(i+1, j+1)$ si $x_i = 1$, y con $(i+1, j-1)$ si $x_i = -1$. En la figura 3.2 hay un ejemplo de esta representación.

De esta manera establecemos una biyección entre los elementos de B y los del conjunto C de todos los caminos de $(0,0)$ a $(2n,0)$ en $\mathbb{Z} \times \mathbb{Z}$ que unen puntos (i,j) con $(i+1, j \pm 1)$ y que no atraviesan el eje de abcisas. Este es aún un conjunto difícil de contar. Una solución ingeniosa para hacerlo es la siguiente.

Consideremos el conjunto C_t de todos los caminos de $(0,0)$ a $(2n,0)$ en $\mathbb{Z} \times \mathbb{Z}$ que unen puntos (i,j) con $(i+1, j \pm 1)$, prescindiendo de la condición que no atraviesen el eje de abcisas.

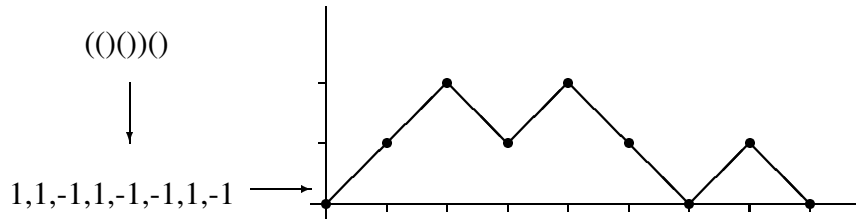


Figura 3.2: Representación geométrica de los elementos de B

Hay tantos caminos en C_t como secuencias ordenadas con n 1's y $n-1$'s (prescindiendo de que la suma de los primeros i dígitos sea no negativa), es decir,

$$|C_t| = \binom{2n}{n}$$

De estos caminos, los que no están en C son los que cortan la recta $y = -1$, es decir, contienen algún punto de la forma $(i,-1)$. Llamamos C_{-1} al conjunto de estos caminos. Entonces,

$$|C| = |C_t| - |C_{-1}|$$

Volveremos a usar una biyección para contar los caminos de C_{-1}. Consideremos la parte del camino entre $(0,0)$ y el primer punto que toca la recta $y = -1$. Si hacemos una reflexión de esta parte del camino con eje de simetría la recta $y = -1$, obtenemos un camino de $(0,-2)$ a $(2n,0)$. En la figura 3.3 hay un ejemplo de este tipo de reflexión.

Recíprocamente, dado un camino cualquiera de $(0,-2)$ a $(2n,0)$, la reflexión con eje de simetría la recta $y = -1$ de la parte del camino que va de $(0,-2)$ al primer punto que corta el eje $y = -1$ da un camino de C_{-1}. Esta reflexión parcial establece entonces una biyección entre los caminos de C_{-1} y el conjunto D de todos los caminos de $(0,-2)$ a $(2n,0)$. De éstos hay tantos como permutaciones de $n+1$ 1's y $n-1$ -1's, es decir,

$$|C_{-1}| = |D| = \binom{2n}{n-1}$$

Por tanto,

$$|C| = |C_t| - |C_{-1}| = \binom{2n}{n} - \binom{2n}{n-1} = \binom{2n}{n} - \frac{n}{n+1}\binom{2n}{n} = \frac{1}{n+1}\binom{2n}{n}$$

Después de diversas transformaciones, hemos llegado finalmente a contar nuestro conjunto original. El resultado que se obtiene es el llamado n-ésimo número de *Catalan*,

$$C_n = \frac{1}{n+1}\binom{2n}{n}$$

Los primeros de estos números son,

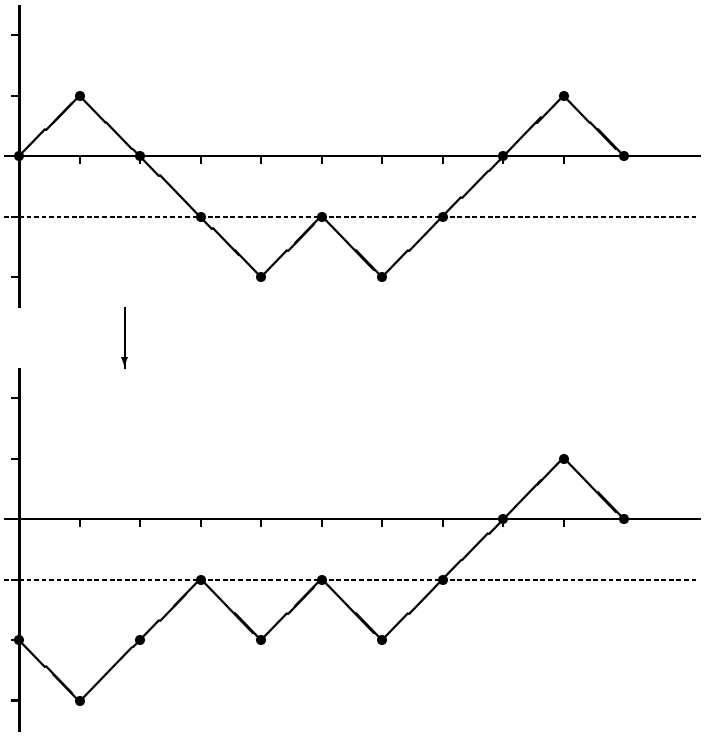


Figura 3.3: Reflexión parcial de los caminos de C_{-1}

n	1	2	3	4	5	6
C_n	1	2	5	14	42	132

Los números de Catalan aparecen en diversos problemas de enumeración aparentemente inconexos. Algunos de estos problemas, relacionados con la enumeración de árboles, se tratarán en un capítulo posterior.

Ejercicio 3.16. (Problema de los votantes) En una elección entre dos candidatos, A y B, hay $2n$ electores. Si el resultado final es de empate, ¿de cuántas maneras podría salir el escrutinio de manera que el candidato A no tuviese nunca menos votos que el candidato B?

Ejercicio 3.17. Sea $X = \{1, 2, \ldots, n\}$. ¿Cuántas selecciones ordenadas con repetición de $\{1, 2, \ldots, n\}$ y longitud n, $x_1 \ldots x_n$, $x_i \in X$ se pueden formar de manera que $x_j \leqslant x_{j+1}$ y $x_j \leqslant j$, $j = 1, \ldots, n$?

Particiones de un entero

Una partición de un entero positivo n es una representación de n como suma de enteros positivos. En la sección 2 del capítulo anterior se ha tratado el problema de enumerar las maneras de

escribir un número natural n como suma de enteros no negativos, entendiendo por diferentes dos expresiones si el orden con que aparecen los sumandos es diferente. El número de *particiones ordenadas* de n en k sumandos positivos hemos visto que era $\binom{n+k-1}{k}$. El problema es mucho más complejo si lo que pretendemos es contar el número de particiones no ordenadas, o simplemente *particiones*, de un entero positivo n en k partes. Llamamos $p_k(n)$ a este número. por ejemplo, $p_2(4) = 2$, y las dos particiones de 4 en dos partes son

$$4 = 3+1 \quad \text{y} \quad 4 = 2+2$$

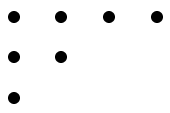
Escribiremos cada partición no ordenada de n en k partes dándolas en orden decreciente. Así, $p_k(n)$ es el número de soluciones de

$$n = x_1 + x_2 + \cdots + x_k, \quad x_1 \geqslant x_2 \geqslant \cdots \geqslant x_k \geqslant 1$$

e identificamos la partición con la k-tupla $x = (x_1, x_2, \ldots, x_k)$. Una manera de representar cada una de las particiones es por medio de los llamados *diagramas de Ferrers*. El diagrama de Ferrers de la partición $x = (x_1, x_2, \ldots, x_k)$ se obtiene poniendo k filas de puntos, con x_i puntos en la fila i. Así, por ejemplo, la partición de 7 en tres partes

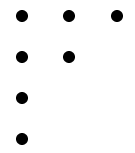
$$7 = 4+2+1$$

se representa con el diagrama



En lo que sigue veremos algunos ejemplos de resultados sobre particiones que se pueden obtener por medio de biyecciones.

La partición x' es *conjugada* de la partición x cuando su diagrama de Ferrers se obtiene del de x cambiando filas por columnas. La partición de 7 conjugada de $(4,2,1)$ es entonces la correspondiente al diagrama

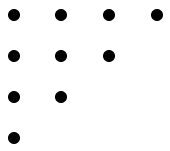


Proposición 3.18. (i) El número $p_k(n)$ de particiones de n en k partes coincide con el número de particiones de n en las cuales la parte más grande es k.

(ii) El número de particiones de n en les que la parte más grande es k o menor que k coincide con el número de particiones de n en como mucho k partes.

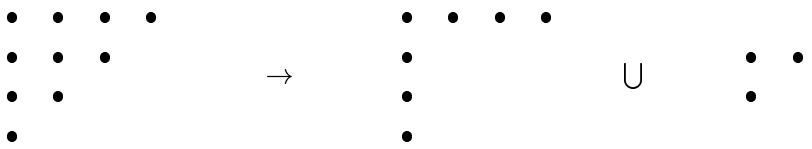
Demostración. La conjugación es una biyección entre el conjunto de particiones de n en k partes y el conjunto de particiones de n en las cuales la parte más grande es k. La segunda afirmación se demuestra de la misma manera. □

Decimos que una partición es autoconjugada si coincide con su partición conjugada. Por ejemplo, la partición $(4,3,2,1)$ de 10 es autoconjugada. El diagrama de Ferrers de una partición autoconjugada es simétrico respecto de la diagonal principal. El de la partición $(4,3,2,1)$, por ejemplo, es el siguiente:



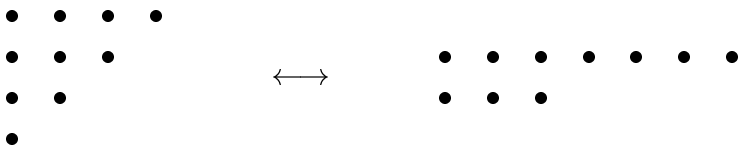
Proposición 3.19. El número de particiones de n en partes pares y diferentes coincide con el número de particiones autoconjugadas.

Demostración. Definiremos una biyección σ entre el conjunto $P_c(n)$ de particiones autoconjugadas y el conjunto $P_{sd}(n)$ de partes pares y diferentes. El diagrama de Ferrers de una partición autoconjugada se puede ver como una unión de 'L's invertidas:



La primera de estas 'L's invertidas está formada por la primera fila y la primera columna. Suprimiéndola, se obtiene la segunda 'L' con la primera fila y la primera columna del diagrama que queda, y así sucesivamente. Para cada partición autoconjugada x, llamamos $\sigma(x)$ a la partición que tiene por filas estas 'L's invertidas. Está claro que $\sigma(x) \in P_{sd}(n)$. Recíprocamente, para cada partición $x \in P_{sd}(n)$, plegando cada fila de longitud $2k+1$ por el punto $k+1$ construimos una sucesión de 'L's invertidas que corresponden a una partición de $P_c(n)$, de manera que σ es efectivamente una biyección entre los dos conjuntos. □

En la figura siguiente se ilustra la biyección σ definida en esta última demostración para una partición de 10.



Proposición 3.20. El número de particiones de n en partes pares coincide con el número de particiones de n en partes diferentes.

Demostración. Sea $P_s(n)$ el conjunto de particiones de n en partes pares y $P_d(n)$ el conjunto de particiones de n en partes diferentes. Definiremos una biyección σ entre los dos conjuntos de la manera siguiente. Para cada partición $x \in P_d(n)$, dejamos invariantes las partes impares y convertimos cada parte par de tamaño $m_i 2^j$, donde m_i es impar, en 2^j partes de tamaño m_i. Está claro, entonces, que $\sigma(x) \in P_s(n)$ y que la aplicación es inyectiva. Para ver que es exhaustiva construimos su inversa. Para cada partición $y \in P_s(n)$, sea m_i el número de partes de tamaño i. Si $m_i \leqslant 1$, dejamos invariante la parte correspondiente. Si $m_i > 1$, escribimos m_i en base 2, $m_i = \sum_j a_j 2^j$, y creamos una parte de tamaño $i2^j$ para cada j tal que $a_j = 1$. Es fácil comprobar que de esta manera hemos construido efectivamente una biyección entre $P_s(n)$ y $P_d(n)$. □

Ejercicio 3.21. Demostrar por medio de una biyección que el número de particiones de n en exactamente tres partes coincide con el número de particiones de $2n$ en tres partes tales que la suma de dos cualesquiera de las partes es más grande que la tercera.

3.4 El principio del palomar y el teorema de Ramsey

Acabamos este capítulo desarrollando otro principio simple con resultados bien poco triviales. A diferencia de los principios de las secciones anteriores, el que veremos aquí no proporciona una herramienta de enumeración, sino que garantiza sólo la existencia de configuraciones particulares en hipótesis muy generales. La filosofía general se podría resumir diciendo que determinadas configuraciones son inevitables cuando los conjuntos son suficientemente grandes.

El *principio del palomar* se conoce también como *principio de Dirichlet* y viene a decir que, si muchas palomas se ponen en un palomar que tiene pocos agujeros, en alguno de los agujeros tiene que haber más de una paloma. Más formalmente se puede enunciar de la manera siguiente.

Principio de Dirichlet En una selección de $k+1$ elementos, o más, de un conjunto de k elementos, algún elemento aparece dos o más veces.

De acuerdo con este principio, en una reunión de más de doce personas, al menos dos han de haber nacido en el mismo mes; o aún, en una ciudad de más de dos millones de habitantes, al menos dos personas deben tener el mismo número de cabellos (no hay casos conocidos de personas con más de 2 millones de cabellos).

Una aplicación menos evidente es la siguiente.

Proposición 3.22. En una sucesión estrictamente creciente de $2n$, $n > 1$, números enteros entre 1 y $3n$, $1 \leqslant a_1 < a_2 < \cdots < a_{2n} \leqslant 3n$, debe haber dos, $a_i < a_j$, la diferencia entre los cuales sea exactamente $a_j - a_i = n - 1$.

Demostración. Consideremos la sucesión formada por la sucesión original y los elementos obtenidos sumándoles $(n-1)$,

$$a_1, a_2, \ldots, a_{2n}, a_1 + (n-1), a_2 + (n-1), \ldots, a_{2n} + (n-1)$$

Esta sucesión tiene $4n$ números entre 1 y $4n-1$, de manera que, por el principio de Dirichlet, debe tener dos números iguales. Como en la colección original todos los números son diferentes, debe ser $a_j = a_i + (n-1)$ para algunos i, j. □

Ejercicio 3.23. Demostrar que en la proposición anterior no se puede asegurar que haya dos términos de la colección la diferencia entre los cuales sea exactamente n (comprobarlo primero para $n = 2$ y $n = 3$).

Ejercicio 3.24. Un jugador de ajedrez quiere jugar un máximo de 45 partidas en un mes (de 30 días) para no fatigarse. Para mantenerse en forma, pero, quiere jugar al menos una partida cada día. Demostrar que hay un período de como mucho 14 días consecutivos en los cuales juega exactamente 14 partidas.

Una generalización sencilla del principio de Dirichlet es la siguiente.

Proposición 3.25. En una selección de m elementos de un conjunto de tamaño $k < m$ hay algún elemento que aparece al menos $\lceil \frac{m}{k} \rceil$ veces.

Por ejemplo, en una reunión de 40 personas, al menos 4 han nacido en el mismo mes, ya que $\lceil 40/12 \rceil = 4$. Si se reparten 10 personas en dos grupos, uno de los dos tiene 5 personas o más, y si son 11, alguno de los dos grupos tiene al menos 6 personas. Si un ordenador tiene 8000 bits de memoria libre en ocho posiciones diferentes, una de las posiciones tiene al menos 1000 bits libres. O también, si la media de edad de un grupo de personas es de 20 años, alguna debe tener 20 años o más y alguna otra 20 años o menos.

Una aplicación más difícil de esta segunda versión del principio de Dirichlet es el siguiente teorema del gran matemático húngaro del siglo XX Paul Erdős.

Teorema 3.26 (Erdős–Szekeres, 1935). En una sucesión cualquiera de $n^2 + 1$ enteros diferentes, o bien hay una subsucesión estrictamente creciente de $n+1$ elementos, o bien hay una estrictamente decreciente de $n+1$ elementos.

Demostración. Llamamos $a_1, a_2, \ldots, a_{n^2+1}$ a la sucesión original. Para cada i, llamamos c_i a la longitud de la subsucesión creciente más larga que comienza en a_i. Si alguno de los

c_i es más grande que $n+1$, ya hemos acabado. En caso contrario, tenemos n^2+1 enteros $c_1,\ldots,c_{n^2},c_{n^2+1}=1$ entre 1 y n, de manera que, por el principio de Dirichlet, debe haber $\left\lceil \frac{n^2+1}{n} \right\rceil = n+1$ enteros iguales. Supongamos que $c_{i_1}=c_{i_2}=\cdots=c_{i_{n+1}}$, donde $i_1<i_2<\cdots<i_{n+1}$. Entonces los enteros correspondientes, $a_{i_1},a_{i_2},\ldots,a_{i_{n+1}}$ forman una subsucesión decreciente. En efecto, si fuese $a_{i_k}<a_{i_{k+1}}$, entonces, usando la subsecuencia creciente más larga que comienza en $a_{i_{k+1}}$, que tiene longitud $c_{i_{k+1}}$, obtendríamos una subsecuencia creciente desde a_{i_k} de longitud $c_{i_k}=c_{i_{k+1}}+1$, contradiciendo que estos dos números sean iguales. □

Analizemos por ejemplo la sucesión

$$10,3,2,1,6,5,4,9,8,7$$

Aquí $n=3$, y las subsecuencias crecientes más largas desde cada término tienen longitudes

i	1	2	3	4	5	6	7	8	9	10
c_i	1	3	3	3	2	2	2	1	1	1

de manera que no hay ninguna subsecuencia de longitud 4. Hay en cambio cuatro 1 entre los c_i para $i=1,8,9,10$. Efectivamente, los términos $1,8,9,10$ de la sucesión forman una subsucesión decreciente.

El resultado del teorema anterior es el mejor posible en el sentido que con menos de n^2+1 enteros se pueden formar sucesiones que no satisfacen el enunciado. Por ejemplo, en la sucesión

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

de $9=3^2$ enteros, las subsecuencias crecientes y decrecientes más largas tienen 3 elementos.

Ejercicio 3.27. Dar ejemplos de sucesiones con 4^2 y 5^2 términos que no contengan subsucesiones crecientes ni decrecientes de longitud más grande que 4 y 5 respectivamente.

Una de las generalizaciones más complejas del principio de Dirichlet es la que formuló Ramsey en el año 1930. La riqueza de las extensiones y las aplicaciones de su resultado han dado lugar a toda una teoría, que se llama *teoría de Ramsey*. Antes de introducir el resultado general veremos una aplicación sencilla.

Proposición 3.28. En un grupo de seis personas, o bien hay tres que se conocen mútuamente, o bien hay tres que no se conocen entre ellas.

Demostración. Llamamos a a una de las personas. Por el principio de Dirichlet, de las cinco que quedan debe haber o bien tres que conocen individualmente a a, o tres que no la conocen.

Supongamos que b, c y d conocen a a. Si entre ellas hay dos que se conocen, ya hemos encontrado el trío de conocidos comunes. Si no, b, c y d forman un trío de mutuamente desconocidos.

Si en cambio hay tres personas b, c y d que no conocen a a, o bien dos de ellas no se conocen (y forman con a un trío de desconocidos), o bien se conocen mutuamente y forman un trío de conocidos comunes. □

El enunciado general del teorema de Ramsey clasifica los subconjuntos de un cierto tamaño k de un conjunto X en dos clases disyuntas y asegura que, si $|X|$ es suficientemente grande, entonces se puede encontrar un subconjunto de un cierto tamaño de manera que todos sus k-subconjuntos sean sólo de una de las dos clases. En la proposición anterior, clasificamos todas las parejas (2-subconjuntos) en dos clases, las de parejas de conocidos y parejas de desconocidos. Entonces, o bien hay un subconjunto de 3 elementos donde todas las parejas son de conocidos, o bien uno de 3 elementos donde todas son de desconocidos.

Teorema 3.29 (Ramsey, 1930). Sea X un conjunto de N elementos y clasifiquemos todos sus subconjuntos de k elementos en dos clases disyuntas, A y B. Sean $p,q \geqslant k$ dos enteros. Entonces, si $N \geqslant R(p,q,k)$, un número que no depende de X sino sólo de p, q y k, o bien existe un subconjunto A de tamaño p que tiene todos los k-subconjuntos en la clase A, o bien existe otro B de tamaño q que tiene todos los k-subconjuntos en la clase B.

Observar que el teorema sólo asegura la existencia de un cierto número $R(p,q,k)$, que se llama *número de Ramsey*, y de los subconjuntos A o B. De hecho se conocen pocos valores de $R(p,q,k)$. La proposición 3.28 es equivalente a que $R(3,3,2) \leqslant 6$, ya que en un conjunto de seis elementos se pueden encontrar subconjuntos de 3 elementos de manera que todas las parejas que contiene sean de una clase. De hecho, ya no se puede asegurar lo mismo en un conjunto de cinco elementos. Por ejemplo, si clasificamos las parejas del conjunto $X = \{1,2,3,4,5\}$ en las clases

$$\begin{aligned} A &= \{\{1,2\},\{2,3\},\{2,4\},\{3,5\},\{4,5\}\} \\ B &= \{\{1,4\},\{1,5\},\{2,3\},\{2,5\},\{3,4\}\} \end{aligned}$$

todos los subconjuntos de tres elementos contienen dos parejas de una clase y una de la otra, de manera que $R(3,3,2) = 6$.

Demostración del teorema. Observemos en primer lugar que para $k = 1$ el teorema es cierto: Si clasificamos todos los elementos en dos clases A y B, por el principio del palomar en una de las dos hay al menos $\left\lceil \frac{|X|}{2} \right\rceil$ elementos. Por tanto, si $|X| = p+q-1$, o bien hay un subconjunto A de p elementos con todos los elementos en la clase A, o, en caso contrario, han de quedar al

menos q elementos de la clase B. Así, $R(p,q,1) \leqslant p+q-1$ y no es difícil ver que, de hecho, vale la igualdad.

Otro caso en que resulta fácil probar el teorema y dar el valor de $R(p,q,k)$ es para $p=k$ o para $q=k$. Si $p=k$, tomamos A como uno de los conjuntos de la clase A y, si no hay ninguno, $B=X$ tiene todos los subconjuntos de la clase B, de manera que $R(k,q,k)=q$. Similarmente, $R(p,k,k)=p$.

Demostraremos ahora el caso general de la existencia de $R(p,q,k)$, $p>k$, $q>k$, por inducción. Supongamos que el teorema es cierto para $R(p,q,k-1)$ y todos los valores de p y q. Supongamos también que es cierto para $R(p',q',k)$ si $p'<p$ o $q'<q$. En particular, llamamos $p_1=R(p-1,q,k)$ y $q_1=R(p,q-1,k)$.

Sea X un conjunto con al menos $R(p_1,q_1,k-1)$ elementos y sea $x_0 \in X$. Consideremos $X_0=X\setminus\{x_0\}$ y clasifiquemos todos los subconjuntos de tamaño $(k-1)$ de X_0 en las clases A' y B' de manera que $U\subset X_0$ está en la clase A' (respectivamente, B') si y sólo si $U\cup\{x_0\}\subset X$ está en la clase A (respectivamente, B). Entonces, o bien hay un conjunto A' de tamaño p_1 con todos los subconjuntos de tamaño $k-1$ en la clase A', o bien hay un conjunto B' de tamaño q_1 con todos los subconjuntos de tamaño $k-1$ en la clase B'.

En el primer caso, como $p_1=R(p-1,q,k)$, o bien existe un subconjunto $A\subset A'$ de tamaño $(p-1)$ que tiene todos los subconjuntos de tamaño k en la clase A, caso en el cual $A\cup\{x_0\}$ tiene tamaño p y la misma propiedad, o bien existe un conjunto $B\subset A'$ de tamaño q que tiene todos los subconjuntos de tamaño q de la clase B, tal y como queríamos demostrar. Un razonamiento similar se aplicaría si lo que existe es un conjunto B' de tamaño q_1. □

Aunque se conocen muy pocos valores de los números de Ramsey $R(p,q,k)$, la demostración anterior proporciona una cota superior de estos números.

Corolario 3.30. Los números de Ramsey satisfacen la desigualdad

$$R(p,q,k) \leqslant R(R(p-1,q,k),R(p,q-1,k),k-1)+1$$

En particular, para $k=2$, la desigualdad anterior se puede escribir de la manera siguiente:

Corolario 3.31. $R(p,q,2) \leqslant \binom{p+q-2}{p-1}$

Demostración. Tenemos $R(p,2,2)=p=\binom{p}{p-1}$, de manera que el resultado es cierto para $q=2$. Similarmente, $R(2,q,2)=\binom{q}{1}$. Supongamos que el resultado es cierto para $R(p',q',2)$ si $p'<p$ o bien $q'<q$. Usando la desigualdad del corolario anterior y el hecho de que $R(p,q,1)=$

$p+q-1$,

$$\begin{aligned} R(p,q,2) &\leqslant R(R(p-1,q,2),R(p,q-1,2),1)+1 \\ &= R(p-1,q,2)+R(p,q-1,2) \\ &\leqslant \binom{p+q-3}{p-2}+\binom{p+q-3}{p-1} \\ &= \binom{p+q-2}{p-1} \end{aligned}$$

□

Notas bibliográficas

El desarrollo y la sistematización de los principios que se han visto en este capítulo se han producido básicamente durante el siglo XX, a medida que los métodos de enumeración y la combinatoria en general han ido formando un cuerpo teórico con identidad propia. El principio de inclusión-exclusión, formulado en su versión moderna por Da Silva (1854) y Sylvester (1883), forma parte de los llamados *métodos de criba (sieve methods)* utilizados en teoría de números. Sylvester fue también quien introdujo los diagramas de Ferrers para el estudio de particiones. A pesar de su nombre, E. Catalan (1814–1894) fue un matemático belga que estudió la formación correcta de paréntesis. El principio de reflexión que se ha usado para obtener los números de Catalan fue introducido por el matemático francés D. André a comienzos del siglo XX. El Teorema de Ramsey ha sido la fuente de toda una rama de la combinatoria, llamada justamente Teoría de Ramsey, que gira alrededor de la idea de que un conjunto llega a contener subconjuntos con propiedades preestablecidas siempre que se tome un número suficiente de elementos.

Los temas que aparecen en este capítulo se pueden encontrar en la mayoría de textos de combinatoria o matemática discreta. Los libros de Biggs [1] y de Hall [2] son ejemplos excelentes. El texto de Stanton y White [3] pone énfasis en el aspecto constructivo y algorítmico de estos problemas.

Bibliografía

[1] N. L. Biggs. *Matemática Discreta*, Ed. Vicens Vives, 1994.

[2] M. Hall. *Combinatorial Theory*, Wiley Interscience, 1986.

[3] D. Stanton, D. White. *Constructive Combinatorics*, UTM, Springer Verlag, 1986.

Problemas

1. ¿Cuántos números hay entre 1000 y 10000 que no sean divisibles por 3 ni por 7?

2. ¿Cuál es el número entero más pequeño que no tiene más de cuatro divisores primos?

3. Demostrar que el número de enteros $n \leqslant N$ que no son divisibles por los primos del conjunto $P = \{2,3,5,7\}$ es

$$N - \sum_{i \in P} \lfloor N/i \rfloor + \sum_{i,j \in P} \lfloor N/ij \rfloor - \sum_{i,j,k \in P} \lfloor N/ijk \rfloor + \lfloor N/210 \rfloor$$

   Calcular este número si $N = 210k$.

4. Demostrar que el número de matrices cuadradas (a_{ij}) de tamaño 3×3 y coeficientes a_{ij} enteros no negativos tales que la suma de los coeficientes de cada fila y la suma de los coeficientes de cada columna vale $r \in \mathbb{N}$ es

$$\binom{r+2}{2}^2 - 3\binom{r+3}{4}$$

5. Demostrar que el número $e(n \to m)$ de aplicaciones exhaustivas de un conjunto X de n elementos en un conjunto Y de m elementos viene dado por la expresión

$$e(n \to m) = m^n - \binom{m}{1}(m-1)^n + \binom{m}{2}(m-2)^n - \cdots + (-1)^{m-1} m$$

6. Demostrar que el número de palabras de longitud $2n$ que se pueden formar de un alfabeto de n letras sin que dos letras seguidas sean iguales es

$$\frac{1}{2^n}\left((2n!) - \binom{n}{1}2(2n-1)! + \binom{n}{2}2^2(2n-2)! - \cdots (-1)^n 2^n n!\right)$$

7. Usando el principio de inclusión-exclusión, contar cuántas combinaciones con repetición de tamaño 11 se pueden formar con 3 elementos x_1, x_2, x_3 si x_1 no puede aparecer más de $r_1 = 3$ veces, x_2 no puede aparecer más de $r_2 = 4$ veces y x_3 no puede aparecer más de $r_3 = 6$ veces (llamamos A_i al conjunto de combinaciones con más de r_i elementos x_i, $i = 1,2,3$).

8. Un ordenador efectúa el producto de n números por parejas usando la propiedad asociativa (por ejemplo, $x_1x_2x_3x_4$ se puede multiplicar haciendo primero x_1x_2, el resultado por x_3 y el resultado por x_4, es decir, $(((x_1x_2)x_3)x_4)$. Si no se puede alterar el orden de los elementos, ¿de cuántas maneras diferentes se puede efectuar el producto de n números?

9. Demostrar, a partir de una biyección, la identidad

$$\binom{n}{k}\binom{k}{m} = \binom{n}{m}\binom{n-m}{k-m}$$

10. Si $X = \{1,2,\ldots,n\}$, una función $f : X \longrightarrow X$ es *monótona* si $x \leqslant y \Rightarrow f(x) \leqslant f(y)$. Demostrar que el número de aplicaciones monótonas que verifican $f(i) \leqslant i$, $i \in X$ es el número de Catalan C_n.

11. Usando una biyección con las secuencias de longitud $(m+n)$ de ceros y unos con m ceros, demostrar que el número de diagramas de Ferrers que caben en un rectángulo $m \times n$ es $\binom{m+n}{n}$.

12. Demostrar que hay tantas particiones autoconjugadas de n como particiones con una parte de tamaño k^2 y el resto de tamaño $\leqslant 2k$ para algún k.

13. Tenemos una caja con i bolas de color c_i para $i = 1,\ldots,n$. Demostrar que, dado $k < n$, el mínimo número de bolas que es preciso extraer para tener la seguridad que habrá k del mismo color es $(k-1)(n-(k/2)+1)+1$.

14. Demostrar que en cualquier subconjunto Y de tamaño n de $X = \{1,2,\ldots,2n\}$ hay un par de elementos $a,b \in Y$ tales que a divide a b.

15. Dada cualquier sucesión $a_1,\ldots,a_p$ de p números naturales, ¿hay alguna subsecuencia de términos consecutivos $a_i, a_{i+1},\ldots,a_{i+k}$ tal que p divide a la suma $a_i + a_{i+1} + \cdots + a_{i+k}$?

16. Demostrar que, para cualquier partición de los pares de elementos de un conjunto X de cardinal 17 en tres clases A,B,C, o bien hay tres pares de la clase A, o bien tres de la clase B o bien tres de la clase C.

Capítulo 4

Funciones generadoras

1. Ecuaciones de recurrencia
2. Funciones generadoras
3. Ecuaciones de recurrencia lineales
4. Números combinatorios

En este capítulo se tratan técnicas de resolución de problemas de enumeración en los cuales se obtienen expresiones del resultado en términos del tamaño del problema. Cuando estas expresiones relacionan la solución del problema en tamaños diferentes, se obtienen las llamadas *ecuaciones de recurrencia*, que se exponen en la primera sección y donde se obtienen también ecuaciones de recurrencia para muchos de los problemas considerados en los dos capítulos anteriores. En particular se introduce la famosa sucesión de Fibonacci, que es un ejemplo muy representativo. Las *funciones generadoras*, que se introducen en la sección 2, constituyen una de las herramientas más versátiles para el tratamiento de problemas de enumeración. En esta sección se ven algunas de las manipulaciones más comunes de las funciones generadoras ordinarias, se expone un ejemplo de aplicación en relación a los coeficientes binomiales y se introducen también las funciones generadoras exponenciales. Aunque no es estrictamente necesario, el conocimiento de los desarrollos en serie de potencias de funciones de variable real o compleja puede hacer más fácil y rica la lectura de esta parte. En la sección 3 se usan funciones generadoras para obtener un procedimiento sistemático de resolución de las llamadas ecuaciones de recurrencia *lineales*. La resolución hace intervenir la descomposición de fracciones racionales en fracciones simples. En esta cuestión, también el conocimiento de este tipo de descomposición (que se usa también para el cálculo de primitivas de funciones racionales) hará más ágil la lectura de algunos fragmentos. Finalmente, en la última sección se hace una revisión de los números combinatorios bajo la óptica de las funciones generadoras. Aparte de los

coeficientes binomiales, discutidos en la sección 2, se revisan los desarreglos, los números de Catalan y las particiones. Por otra parte, se completa la descripción de números combinatorios con los llamados *números de Stirling* y *números de Bell*.

4.1 Ecuaciones de recurrencia

El enunciado de un problema combinatorio de enumeración hace intervenir uno o más números naturales que dan las dimensiones del problema. Por ejemplo, en la enumeración de los subconjuntos de tamaño k de un conjunto de tamaño n, las dimensiones del problema son n y k. A menudo es fácil, si no trivial, resolver los problemas en dimensiones pequeñas. También es frecuente que la solución se pueda expresar en términos del mismo problema en dimensiones más pequeñas. Las expresiones que dan estas relaciones son lo que se llama *ecuaciones de recurrencia*.

Comencemos con un problema conocido, el de contar el número de subconjuntos de tamaño k del conjunto $X = \{x_1, x_2, \ldots, x_n\}$. Este número no depende de los elementos del conjunto X, sino sólo de su tamaño n y del tamaño k de los subconjuntos que queremos extraer. Llamamos a este número, provisionalmente, por $C(n,k)$. Es fácil contar cuántos subconjuntos hay de tamaño 1, $C(n,1) = n$. También es fácil ver que $C(n,n) = 1$. En cuanto a la solución general, de todos los subconjuntos de tamaño k que podemos formar en el conjunto X, los hay que contienen el elemento x_1 y los hay que no. Los que contienen el elemento x_1 son los que se pueden formar añadiendo $k-1$ elementos de los $n-1$ elementos de X diferentes de x_1. Este es, sin embargo, el mismo problema: ¿cuántos subconjuntos de tamaño $k-1$ se pueden formar de un conjunto de $n-1$ elementos? De acuerdo con nuestra notación, este número es $C(n-1,k-1)$. De forma similar, los subconjuntos que no contienen x_1 son los de tamaño k que se pueden formar con los $n-1$ elementos de X diferentes de x_1. Así obtenemos la relación

$$C(n,k) = C(n-1,k) + C(n-1,k-1) \tag{4.1}$$

Una *ecuación de recurrencia* para una sucesión de números $f(n_1, n_2, \ldots, n_k)$, donde $n_1, \ldots, n_k$ son las variables de la ecuación (usualmente enteros positivos) es una expresión que da el valor de $f(n_1, n_2, \ldots, n_k)$ en términos de $f(n'_1, n'_2, \ldots, n'_k)$ para valores $n'_i \leqslant n_i$ más pequeños de las variables. Una ecuación de recurrencia determina los valores de la sucesión si se establece (i) el dominio de validez de la ecuación en términos de las variables y (ii) un conjunto suficiente de valores particulares de la sucesión.

En el ejemplo anterior, la relación 4.1 es una ecuación de recurrencia de dos variables, válida para enteros positivos n,k con $k < n$, que determina los valores de la sucesión $C(n,k)$ si se especifican por ejemplo los valores $C(n,1)$ y $C(n,n)$, ya que la aplicación reiterada de la recurrencia 4.1 conduce a números de estos dos tipos.

Las ecuaciones de recurrencia son muy útiles cuando se quiere resolver un problema particular (para ciertos valores concretos de n y de k en el ejemplo anterior), ya que el cálculo se puede mecanizar con mucha facilidad. En otras palabras, una ecuación de recurrencia es ya un algoritmo recursivo para hacer cálculos. Lo único que se necesita es un punto de partida para poner en marcha el proceso. En el nuestro ejemplo, podríamos calcular $C(4,3)$ sabiendo que $C(n,1) = n$, $C(n,n) = 1$, para todo $n \in \mathbb{N}$, haciendo

$$C(4,3) = C(3,3) + C(3,2) = 1 + (C(2,2) + C(2,1)) = 1 + (2+1) = 4$$

En cambio, las ecuaciones de recurrencia tienen el inconveniente que no dan ninguna información sobre el valor de las soluciones si no se calculan explícitamente. Por esto resulta interesante intentar encontrar técnicas que permitan reducir una ecuación de recurrencia a lo que se llama una *solución cerrada*, que quiere decir una expresión que pueda ser evaluada en una cantidad fija de operaciones aritméticas (sumas y diferencias, productos y cocientes, exponenciación y radicación). En problemas combinatorios se admite también como expresión cerrada el factorial de un número (aunque de hecho corresponde a una cantidad variable de multiplicaciones) y a menudo también expresiones con sumatorios. Así, por ejemplo, la solución que hemos obtenido en el capítulo 2

$$C(n,k) = \binom{n}{k} = \frac{n!}{k!\,(n-k)!}$$

se admite como una expresión cerrada, mientras que la relación 4.1 ciertamente no lo es.

En esta sección veremos como muchos de los problemas de los dos capítulos anteriores admiten también un tratamiento en términos de ecuaciones de recurrencia. En secciones posteriores se analizarán técnicas específicas para resolver algunas de estas recurrencias.

Los *coeficientes binomiales* satisfacen un número considerable de ecuaciones de recurrencia de las cuales se han visto ejemplos en la última sección del capítulo 2. La más característica es la ecuación 4.1, que hemos discutido más arriba y que proporciona la base para construir el triángulo de Tartaglia.

Recordemos que un *desarreglo* de n símbolos es una permutación que no deja ningún símbolo en su sitio. El número D_n de desarreglos de n símbolos se puede obtener también por una ecuación de recurrencia. Consideremos el conjunto A_i de los desarreglos de $12\ldots n$ en los cuales 1 ocupa la posición $i \neq 1$. Sea A_i^1 el conjunto de estos desarreglos en que i ocupa la posición 1, y $A_i^0 = A_i \setminus A_i^1$ el resto. Para $n = 5$ y $i = 2$ tendremos, por ejemplo,

A_2^1	A_2^0		
21453	31254	41253	51234
21534	31524	41523	51423
	31452	41532	51432

Si se intercambian los símbolos 1 e i en la identidad, para completar un desarreglo es preciso mover de su sitio los $(n-2)$ símbolos que quedan y distribuirlos en las $(n-2)$ posiciones libres. Por tanto, hay tantos elementos en A_i^1 como desarreglos de los $(n-2)$ símbolos diferentes de 1 y de i, $|A_i^1| = D_{n-2}$. Por otra parte, si i no ocupa la posición 1, llamando '1' al símbolo i establecemos una biyección entre A_i^0 y los desarreglos de $(n-1)$ símbolos (tenemos los símbolos $1 \dots n$ sin el símbolo i y las posiciones $1 \dots n$ sin la posición i). Así, $|A_j^0| = D_{n-1}$. Entonces, $|A_i| = |A_i^1| + |A_i^0| = D_{n-2} + D_{n-1}$ para cualquier i. De

$$D_n = |A_2 \cup \dots A_n|$$

donde la unión es disyunta, obtenemos la ecuación de recurrencia

$$D_n = (n-1)(D_{n-2} + D_{n-1}) \tag{4.2}$$

válida para todos los enteros positivos $n > 2$, y que determina los valores de D_n a partir de $D_1 = 0$ y $D_2 = 1$. A partir de estos valores, se obtiene la sucesión

$$0, 1, 2, 9, 44, 265, \dots$$

Ejercicio 4.1. Encontrar una ecuación de recurrencia para el número D_n^k de permutaciones de $12 \dots n$ que dejan exactamente k símbolos en su sitio. Usarla para obtener los 5 primeros valores de D_5^2.

Los *números de Catalan* satisfacen también una ecuación de recurrencia característica. Recordemos que los números de Catalan C_n cuentan, entre otras cosas, el número de secuencias de $2n$ paréntesis sintácticamente correctos (es decir, sin que se cierre un paréntesis que no ha sido abierto antes). Dada una de estas secuencias, llamamos k al entero más pequeño tal que la subsucesión de los primeros $2k$ paréntesis también es sintácticamente correcta. Por ejemplo, en las 5 sucesiones correctas de 6 paréntesis,

((())), (()()), (())(), ()()(), ()(())

los valores de k son 3, 3, 2, 1 y 1 respectivamente. Todas las secuencias que tienen un valor fijado de k se obtienen concatenando una secuencia correcta de longitud $k-1$ cerrada entre paréntesis con una secuencia correcta de longitud $(n-k)$, de manera que el número de secuencias con este valor es $C_{k-1}C_{n-k}$. Así pues, tenemos la ecuación de recurrencia

$$C_n = \sum_{k=1}^{n} C_{k-1}C_{n-k} = \sum_{k=0}^{n-1} C_k C_{n-k-1} \tag{4.3}$$

válida para enteros positivos $n > 2$ si definimos $C_0 = 1$. A partir de los valores obvios $C_1 = 1$ y $C_2 = 2$, se obtiene la secuencia

$$1, 2, 5, 14, 42, 132, \dots$$

Ejercicio 4.2. Demostrar la ecuación 4.3 usando la expresión $C_n = \frac{1}{n+1}\binom{2n}{n}$ y las propiedades de los coeficientes binomiales.

Las *particiones* de un entero n en $k \leqslant n$ partes, $p_k(n)$, se pueden obtener también por medio de una recurrencia. Para cada partición

$$n = x_1 + x_2 + \cdots + x_k, \quad x_1 \geqslant x_2 \geqslant \cdots \geqslant x_k \geqslant 1$$

tenemos una expresión del tipo

$$n - k = y_1 + y_2 + \cdots + y_k, \quad y_1 \geqslant y_2 \geqslant \cdots \geqslant y_k \geqslant 0$$

restando 1 a cada x_i. Si algunos de los valores y_j valen cero, tenemos en realidad una partición de $(n-k)$ en menos de k partes, de manera que obtenemos la ecuación de recurrencia

$$p_k(n) = p_k(n-k) + p_{k-1}(n-k) + \cdots + p_2(n-k) + p_1(n-k) \tag{4.4}$$

válida para enteros positivos $n \geqslant k > 1$. Para $k = 2$, por ejemplo, los valores de $p_1(n) = 1$ y $p_2(2) = p_2(3) = 1$ determinan la sucesión $p_2(n)$, $n \geqslant 2$

$$1, 1, 2, 2, 3, 4, 4, 4, \ldots$$

Acabaremos esta sección con una de las ecuaciones de recurrencia más célebres: la que da lugar a la llamada *sucesión de Fibonacci*. Uno de los problemas que lleva a esta sucesión es el de determinar el número de secuencias de longitud n con los símbolos $0, 1$ de manera que no haya dos ceros seguidos. Llamamos S_n al conjunto de estas sucesiones y $s_n = |S_n|$. Por ejemplo, las sucesiones de longitud 3 de ceros y unos sin dos ceros seguidos son

$$111 \quad 110 \quad 101 \quad 011 \quad 010$$

de manera que $s_3 = 5$. Obtendremos ahora una ecuación de recurrencia para s_n. Hay tantas sucesiones de S_n que acaban en 1 como s_{n-1}. En cambio, si acaban en cero, el penúltimo dígito debe ser 1, de manera que el número de sucesiones de longitud n sin dos ceros consecutivos y acabadas en cero es s_{n-2}. Tenemos por tanto la ecuación de recurrencia

$$s_n = s_{n-1} + s_{n-2} \tag{4.5}$$

válida para cualquier entero positivo $n > 2$. A partir de los valores $s_1 = 2$ y $s_2 = 3$ obtenemos la sucesión

$$2, 3, 5, 8, 13, 21, \ldots$$

La ecuación de recurrencia 4.5 es la que define los *números de Fibonacci* salvo que los valores iniciales son diferentes que los de la sucesión s_n. Si llamamos F_n al número n-ésimo de Fibonacci, tenemos

$$F_n = F_{n-1} + F_{n-2}, \quad F_0 = 0, \quad F_1 = 1 \tag{4.6}$$

de manera que $s_n = F_{n-2}$ para $n \geqslant 2$ (también es común la asignación de valores iniciales $F_0 = F_1 = 1$, con lo cual $s_n = F_{n-1}$). La popularidad de la sucesión de Fibonacci reside tanto en la frecuencia con que aparece en problemas combinatorios y su amplia aplicabilidad como en la simplicidad de la ecuación de recurrencia que la define. En la sección 3 se tratará la resolución de una clase general de ecuaciones de recurrencia y, en particular, obtendremos una expresión cerrada de estos números.

Ejercicio 4.3. Supongamos que en una población de conejos en cada período hay tantos machos como hembras y se aparean para tener dos conejos, un macho y una hembra. Los recién nacidos se incorporan al ciclo reproductor dos períodos después de haber nacido. Encontrar una ecuación de recurrencia para la sucesión u_n que da el número de individuos en el período n. Dar los primeros cinco términos de la sucesión si inicialmente hay una única pareja de recién nacidos.

4.2 Funciones generadoras

De acuerdo con lo que se exponía en la sección anterior, muchas veces un problema combinatorio se puede interpretar como la determinación de una secuencia u_n de números, cada uno de los cuales es la solución del problema de tamaño n. El concepto de función generadora permite trabajar con la secuencia entera 'almacenándola' en una función. En esta sección veremos de qué manera se hace esto y qué ventajas supone para la resolución de los problemas de enumeración.

Dada una secuencia de números u_n, $n \geqslant 0$, se llama *función generadora ordinaria* de esta secuencia la expresión

$$U(x) = u_0 + u_1 x + u_2 x^2 + \cdots = \sum_{n \geqslant 0} u_n x^n$$

La expresión anterior es lo que se llama una *serie formal* y la sucesión $\{u_0, u_1, \ldots\}$ es su *sucesión de coeficientes*. Las series formales son una extensión de los polinomios. De hecho, los polinomios son las series formales con sólo un número finito de elementos no nulos[1].

[1] En el capítulo 12 hay una sección dedicada a tratar el anillo de polinomios, una lectura de la cual puede ayudar a familiarizarse en algunas de las cuestiones que trataremos aquí.

Como en éstas, se pueden definir las operacionas algebraicas de suma y producto así como también la operación de derivación. En primer lugar, diremos que dos series formales son iguales si tienen la misma sucesión de coeficientes. Dadas dos expresiones $U(x) = \sum_{n\geqslant 0} u_n x^n$ y $V(x) = \sum_{n\geqslant 0} v_n x^n$, definimos su suma como la expresión

$$U(x) + V(x) = \sum_{n\geqslant 0} (u_n + v_n) x^n \tag{4.7}$$

y su producto como

$$U(x)V(x) = \sum_{n\geqslant 0} c_n x^n \tag{4.8}$$

donde

$$c_n = u_0 v_n + u_1 v_{n-1} + \cdots + u_{n-1} v_1 + u_n v_0$$

(recordar la suma y el producto de polinomios).

Diremos que $U(x)$ es la inversa respecto del producto de $V(x)$ si el producto de las dos series es $1 = 1 + 0 \cdot x + 0 \cdot x^2 + \cdots$, y escribimos

$$U(x) = \frac{1}{V(x)}$$

Por ejemplo, la inversa respecto del producto de

$$V(x) = 1 + x + x^2 + \cdots$$

(la serie asociada a la sucesión que tiene todos los términos iguales a 1) es la serie $U(x) = 1 - x$, ya que, de acuerdo con la ley del producto que hemos definido,

$$(1 - x)(1 + x + x^2 + \cdots) = 1$$

y escribimos

$$1 + x + x^2 + x^3 + \cdots = \sum_{n\geqslant 0}^{\infty} x^n = \frac{1}{1 - x} \tag{4.9}$$

Esta última igualdad no se debe entender como una igualdad de funciones. Si substituimos x por $1/2$, el valor numérico de los dos lados de la igualdad es el mismo. En cambio, si substituimos x por 2, a la izquierda de la igualdad obtenemos una serie numérica divergente y a la derecha obtenemos -1. Este hecho, sin embargo, no nos impide considerar la igualdad 4.9 como una igualdad válida en el conjunto de las series formales. Desde este punto de vista, tenemos el resultado siguiente:

Proposición 4.4. Una serie formal $U(x) = \sum_{n\geqslant 0} u_n x^n$ es invertible si y sólo si $u_0 \neq 0$.

Demostración. La serie $U(x)$ es invertible si existe una serie $V(x) = \sum_{k\geqslant 0} v_k x^k$ de manera que $W(x) = U(x)V(x) = 1$. De acuerdo con la definición del producto de series, los coeficientes de $W(x)$ satisfacen las relaciones,

$$w_0 = u_0 v_0 = 1$$

de donde tiene que ser $v_0 = 1/u_0$ (que está bien definido si $u_0 \neq 0$)

$$w_1 = u_1 v_0 + u_0 v_1 = 0$$

de donde tiene que ser $v_1 = u_1 v_0/u_0$, y, en general,

$$w_h = u_h v_0 + u_{h-1} v_1 + \cdots + u_1 v_{h-1} + u_0 v_h$$

que proporciona recurrentemente el coeficiente v_h en términos de los coeficientes de $U(x)$ y los coeficientes $v_{h-1}, \ldots, v_0$ encontrados anteriormente. Así entonces, este procedimiento permite identificar los coeficientes de una serie $V(x)$ inversa de $U(x)$. □

Otra de las operaciones útiles en el conjunto de las series formales es la de *derivación*. La derivada de la serie $U(x) = \sum_{n\geqslant 0} u_n x^n$ se define como

$$U'(x) = \sum_{n\geqslant 1} n u_n x^{n-1}$$

Una de las ventajas de concentrar la secuencia $\{u_n\}_{n\in\mathbb{N}}$ en una expresión global $U(x) = \sum_{n\geqslant 0} u_n x^n$ es justamente la posibilidad de efectuar el tipo de manipulaciones algebraicas que hemos descrito hasta aquí. Muchas veces, el uso de estas manipulaciones proporciona expresiones explícitas de los términos u_n de la secuencia. En lo que sigue veremos una primera ilustración de la versatilidad de las funciones generadoras revisando los coeficientes binomiales y algunos problemas combinatorios asociados a estos números en la perspectiva de las funciones generadoras.

Recordemos cómo obteníamos la fórmula del binomio en la sección 3 del capítulo 2 por medio de un argumento combinatorio. Al desarrollar el producto

$$\underbrace{(1+x)\cdots(1+x)}_{n}$$

el coeficiente de x^k cuenta el número de maneras de escoger x en k de los paréntesis y 1 en los $n-k$ restantes. En otras palabras, el coeficiente de x^k es el número de combinaciones de n

elementos tomados de k en k, $\binom{n}{k}$. Para n fijo, tenemos entonces que la función generadora de la secuencia $u_k = \binom{n}{k}$ (recordemos que $\binom{n}{k} = 0$ si $k > n$) es

$$U_n(x) = (1+x)^n = \sum_{k \geqslant 0} \binom{n}{k} x^k \tag{4.10}$$

En este caso la función generadora tiene sólo un número finito de sumandos, de manera que se puede interpretar también como una función de x. De aquí se obtienen, por ejemplo, las relaciones

$$U_n(1) = \sum_{k \geqslant 0} \binom{n}{k} = 2^n$$

que da el número total de subconjuntos de un conjunto de n elementos, o

$$U_n(-1) = \sum_{k \geqslant 0} \binom{n}{k} (-1)^k = 0$$

Derivando la ecuación 4.10, obtenemos

$$U_n'(x) = \sum_{k \geqslant 1} \binom{n}{k} k x^{k-1}$$

que coincide con la derivada usual de $(1+x)^n$, $U'(x) = n(1+x)^{n-1}$, de donde se obtiene

$$U_n'(1) = \binom{n}{1} + 2\binom{n}{2} + \cdots + n\binom{n}{n} = n2^{n-1}$$

A partir de la igualdad $U_{2n} = (1+x)^{2n} = (1+x)^n (1+x)^n = (U_n(x))^2$, igualando los coeficientes del mismo grado en los dos lados de la igualdad y usando la identidad $\binom{n}{k} = \binom{n}{n-k}$, se obtiene la relación

$$\binom{2n}{n} = \binom{n}{0}^2 + \binom{n}{1}^2 + \cdots + \binom{n}{n}^2$$

Estos son algunos ejemplos de cómo el uso de funciones generadoras proporciona resultados de otro modo difíciles de obtener y de demostrar.

Consideremos ahora el producto

$$\underbrace{(1+x+x^2) \cdots (1+x+x^2)}_{n}$$

El coeficiente de x^k cuenta el número de maneras de escoger 1, x o x^2 en cada uno de los paréntesis de manera que la suma total de exponentes sea k. Si identificamos cada paréntesis

con un símbolo, la potencia cero se puede asociar al hecho que no se escoge el símbolo, la potencia 1 al hecho que se escoge una vez y la potencia 2 al hecho que se escoge dos veces. Así, el coeficiente de x^k cuenta el número de combinaciones de n símbolos tomados de k en k de manera que cada símbolo puede aparecer hasta dos veces. Por ejemplo, las combinaciones de 4 elementos tomados de 3 en 3 admitiendo hasta dos apariciones de cada elemento son

123	134	124	234
112	113	114	
221	223	224	
331	332	334	
441	442	443	

y hay tantas como el coeficiente de x^3 en

$$(1+x+x^2)^4 = 1+4x+10x^2+16x^3+\cdots$$

La función generadora de este tipo de combinaciones es entonces

$$U_n(x) = (1+x+x^2)^n \tag{4.11}$$

Ejercicio 4.5. Sea u_k el número de combinaciones de n elementos tomados de k en k en las cuales el elemento i puede aparecer hasta r_i veces. Encontrar la función generadora de la sucesión u_k. Usando esta función generadora, calcular el número de combinaciones de 4 elementos tomados de 3 en 3 de manera que el elemento x_1 puede aparecer una vez, el elemento 2 puede aparecer dos veces y los elementos 3 y 4 pueden aparecer tres veces.

Llevando al límite la generalización que se propone en el ejercicio anterior, se puede obtener el número de combinaciones con repetición de n elementos tomados de k en k: si no hay limitaciones en el número de veces que puede aparecer un elemento, la función generadora es

$$U(x) = (1+x+x^2+\cdots)^n$$

La expresión $V(x) = 1+x+x^2+\cdots = \sum_{k\geqslant 0} x^k$ es la función generadora de la secuencia $v_k = 1$, $k \geqslant 0$. Recordemos que en la ecuación 4.9, habíamos obtenido una expresión más compacta de esta función como $V(x) = 1/(1-x)$. Así entonces, la función generadora del número de combinaciones con repetición de n elementos tomados de k en k, $u_k = \binom{n+k-1}{k}$, es

$$U(x) = \frac{1}{(1-x)^n} = \sum_{k\geqslant 0} \binom{n+k-1}{k} x^k \tag{4.12}$$

El conocimiento de esta función generadora es útil para tratar y resolver problemas similares. Por ejemplo, el número de combinaciones con repetición de n elementos tomados de k

en k de manera que cada elemento aparece un número *par* de veces es, siguiendo los mismos argumentos,

$$W(x) = (1+x^2+x^4+\cdots)^n = \frac{1}{(1-x^2)^n} \tag{4.13}$$

Ejercicio 4.6. Determinar la función generadora de las combinaciones con repetición de n elementos tomados de k en k, $k > 0$, en los cuales cada elemento aparece un número impar de veces. Usar esta función generadora para determinar el número de combinaciones de 4 elementos tomados de 3 en 3 en las cuales cada elemento aparece un número impar de veces.

Acabamos esta sección viendo brevemente otra clase de funciones generadoras. Además de la función generadora *ordinaria* de una sucesión $\{u_n\}_{n\in\mathbb{N}}$, se consideran habitualmente otros tipos de funciones generadoras. Particularmente interesantes son las funciones generadoras exponenciales. La *función generadora exponencial* de la sucesión $\{u_n\}_{n\in\mathbb{N}}$ es

$$E(x) = \sum_{n\geqslant 0} \frac{u_n}{n!} x^n$$

La diferencia con las funciones generadoras ordinarias es entonces que los coeficientes se dividen por $n!$. Por ejemplo, del análisis elemental sabemos que la función generadora de la sucesión $(1,1,1,\dots)$ es la función exponencial,

$$\sum_{n\geqslant 0} \frac{1}{n!} x^n = e^x$$

La ventaja de esta modificación respecto de las funciones generadoras ordinarias consiste sobre todo en la propiedad siguiente:

Proposición 4.7. Si $E(x)$ y $G(x)$ son las funciones generadoras exponenciales de las secuencias $\{u_n\}_{n\in\mathbb{N}}$ y $\{v_n\}_{n\in\mathbb{N}}$ respectivamente, entonces $E(x)G(x)$ es la función generadora de la secuencia $\{s_n\}_{n\in\mathbb{N}}$ con

$$s_n = \sum_{k=0}^{n} \binom{n}{k} u_k v_{n-k}$$

Demostración. Se trata simplemente de efectuar el producto

$$E(x)G(x) = \left(\sum_{k\geqslant 0} \frac{u_k}{k!} x^k\right)\left(\sum_{r\geqslant 0} \frac{v_r}{r!} x^r\right) = \sum_{k,r\geqslant 0} \frac{u_k v_r}{k!\, r!} x^{k+r}$$

Al reunir todos los coeficientes de una misma potencia n obtenemos el coeficiente

$$\sum_{k+r=n} \frac{u_k v_r}{k!\, r!}$$

de manera que el coeficiente n-ésimo de la función generadora exponencial $E(x)G(x)$ es

$$n! \sum_{k+r=n} \frac{u_k v_r}{k!\, r!} = \sum_{k=0}^{n} \binom{n}{k} u_k v_{n-k}$$

□

Recordemos que, si $U(x)$ y $V(x)$ son las funciones generadoras ordinarias de las sucesiones $\{u_n\}_{n\in\mathbb{N}}$ y $\{v_n\}_{n\in\mathbb{N}}$ respectivamente, entonces $U(x)V(x)$ es la función generadora ordinaria de la sucesión $\{c_k = \sum_{n=0}^{k} u_n v_{k-n}\}_{k\in\mathbb{N}}$. El uso de funciones generadoras exponenciales proporciona una herramienta alternativa para describir combinaciones de sucesiones en términos de funciones generadoras. En la última sección de este capítulo se verán algunas situaciones en que el uso de las funciones generadoras exponenciales resulta más adecuado que el de funciones generadoras ordinarias.

4.3 Ecuaciones de recurrencia lineales

Una de las aplicaciones de las funciones generadoras es la de resolución de ecuaciones de recurrencia. La idea básica de esta aplicación está en el hecho que la traslación de índice en una función generadora se traduce simplemente en una expresión algebraica: el producto por una potencia de x. Por ejemplo, las series $U(x) = \sum_{n\geqslant 0} u_n x^n$ y $V_1(x) = \sum_{n\geqslant 1} u_{n-1} x^n$, que corresponden a las secuencias $(u_0, u_1, u_2, \dots)$ y $(0, u_0, u_1, \dots)$ respectivamente, están relacionadas por la igualdad,

$$V_1(x) = xU(x)$$

De manera similar, la serie que resulta de $U(x)$ desplazando los índices h posiciones hacia adelante es

$$V_h(x) = \sum_{n\geqslant h} u_{n-h} x^n = x^h U(x) \tag{4.14}$$

que corresponde a la succesión de coeficientes $(\underbrace{0, \dots, 0}_{h}, u_0, u_1, \dots)$.

Este hecho permite en general analizar las ecuaciones de recurrencia mediante funciones generadoras. En esta sección se usará para obtener un procedimiento sistemático de resolución de una clase amplia de ecuaciones de recurrencia: las ecuaciones de recurrencia lineales.

La secuencia u_n satisface una ecuación de recurrencia *lineal homogenea* de orden k si

$$u_n = a_1 u_{n-1} + a_2 u_{n-2} + \cdots + a_k u_{n-k}, \quad n \geqslant k \tag{4.15}$$

para ciertas constantes $a_1, \ldots, a_k$. Los términos de la secuencia quedan determinados por la ecuación y un conjunto de k valores (usualmente se dan los valores de $u_0, u_1, \ldots, u_{k-1}$). Por ejemplo, la ecuación 4.6 de la sección anterior que da los números de Fibonacci,

$$F_{n+2} = F_{n+1} + F_n, \quad n \geqslant 0, \quad F_0 = F_1 = 1$$

es una ecuación lineal de orden 2.

Sea $U(x) = \sum_{n\geqslant 0} u_n x^n$ la función generadora de una sucesión que satisface una ecuación de recurrencia lineal de orden k como 4.15.

Multiplicamos los dos lados de esta ecuación por x^n y sumamos para todos los valores de $n \geqslant k$ para obtener

$$\sum_{n\geqslant k} u_n x^n = a_1 \sum_{n\geqslant k} u_{n-1} x^n + \cdots + a_k \sum_{n\geqslant k} u_{n-k} x^n \tag{4.16}$$

Para expresar esta igualdad en términos de $U(x) = \sum_{n\geqslant 0} u_n x^n$, llamamos $U_h(x)$ al polinomio de los primeros h coeficientes,

$$U_h(x) = u_0 + u_1 x + \cdots + u_{h-1} x^{h-1}, \quad h \geqslant 1$$

Entonces,

$$\sum_{n\geqslant k} u_n x^n = U(x) - U_k(x), \quad \sum_{n\geqslant k} u_{n-1} x^n = x(U(x) - U_{k-1}(x)), \quad \cdots, \sum_{n\geqslant k} u_{n-k} x^n = x^k U(x)$$

Substituyendo estas expresiones en la ecuación 4.16, obtenemos

$$U(x)(1 - a_1 x - \cdots - a_k x^k) = U_k(x) - a_1 x U_{k-1}(x) - \cdots - a_{k-1} x^{k-1} U_1(x) \tag{4.17}$$

A la derecha de la igualdad hay una suma de polinomios de grado como mucho $k-1$ que denotaremos por $C(x)$. Observemos que los coeficientes de $C(x)$ se pueden obtener de los valores de $u_0, u_1, \ldots, u_{k-1}$. De esta manera tenemos la primera parte de la resolución de ecuaciones de recurrencia lineales.

Proposición 4.8. Si $\{u_n\}_{n\in\mathbb{Z}}$ satisface una ecuación de recurrencia lineal de orden k

$$u_n = a_1 u_{n-1} + a_2 u_{n-2} + \cdots + a_k u_{n-k} \qquad n \geqslant k$$

entonces la función generadora $U(x) = \sum_{n\geqslant 0} u_n x^n$ es

$$U(x) = \frac{C(x)}{1 - a_1 x - \cdots - a_k x^k}$$

para un cierto polinomio $C(x)$ de grado como mucho $k-1$, los coeficientes del cual quedan determinados por los valores de $u_0, u_1, \ldots, u_{k-1}$.

Esta proposición proporciona una expresión cerrada de la función generadora. Más adelante veremos com obtener los coeficientes u_n a partir de esta expresión, con lo que se completa la resolución de la ecuación de recurrencia inicial. Antes, sin embargo, veamos cómo se aplica para encontrar la función generadora de la sucesión $\{F_n\}_{n\in\mathbb{N}}$ de los números de Fibonacci. Recordemos que estos números satisfacen la ecuación

$$F_n = F_{n-1} + F_{n-2}, \quad n \geqslant 2, \quad F_0 = 0, \quad F_1 = 1$$

De acuerdo con la proposición anterior, la función generadora $F(x) = \sum_{n\geqslant 0} F_n x^n$ es

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

donde $C(x)$ es un polinomio de grado como mucho 1, es decir, $C(x) = c_0 + c_1 x$. Para determinar estos coeficientes, escribimos

$$C(x) = F(x)(1 - x - x^2) = (F_0 + F_1 x + \cdots)(1 - x - x^2) = F_0 + (F_1 - F_0)x + \cdots$$

de manera que $c_0 = F_0 = 0$ y $c_1 = (-1)F_0 + F_1 = 1$. Así entonces,

$$F(x) = \frac{x}{1 - x - x^2} \tag{4.18}$$

Una vez se ha obtenido una expresión 'cerrada' como la ecuación

$$U(x) = \frac{C(x)}{1 - a_1 x - \cdots - a_k x^k} = \frac{C(x)}{Q(x)} \tag{4.19}$$

de la función generadora $U(x)$, el paso siguiente consiste en obtener una expresión explícita de los términos de la sucesión u_n. Para ello se define lo que se llama *polinomio característico* de la ecuación de recurrencia 4.15 como el polinomio

$$P(t) = t^k - a_1 t^{k-1} - \cdots - a_k = t^k Q(\frac{1}{t})$$

Si $\lambda_1, \ldots, \lambda_s$ son las raíces de este polinomio con multiplicidades $m_1, \ldots, m_s$, la descomposición

$$P(t) = (t - \lambda_1)^{m_1} \cdots (t - \lambda_s)^{m_s}$$

proporciona una descomposición del denominador de la expresión 4.19 de la forma $Q(x) = x^k P(\frac{1}{x}) = (1 - x\lambda_1)^{m_1} \cdots (1 - x\lambda_s)^{m_s}$, de donde

$$U(x) = \frac{C(x)}{(1 - \lambda_1 x)^{m_1} \cdots (1 - \lambda_s x)^{m_s}}$$

Descomponiendo esta fracción racional en fracciones simples, obtenemos una expresión del tipo

$$U(x) = \sum_{i=1}^{s} \left(\frac{c_{i1}}{(1-\lambda_i x)} + \cdots + \frac{c_{im_i}}{(1-\lambda_i x)^{m_i}} \right) = \sum_{i=1}^{s} \sum_{j=1}^{m_i} \frac{c_{ij}}{(1-\lambda_i x)^j} \tag{4.20}$$

donde c_{ij} son constantes que se pueden determinar a partir de los coeficientes de $C(x)$ y de las raíces λ_i del polinomio característico[2].

Cada uno de los sumandos de esta expresión corresponde a una serie formal del tipo $c/(1-\lambda x)^m$. Como hemos visto en la sección anterior al tratar las combinaciones con repetición (ecuación 4.12), la expresión general de esta serie es

$$\frac{c}{(1-\lambda x)^m} = c \sum_{n \geqslant 0} \binom{n+m-1}{n} (\lambda x)^n$$

de manera que el coeficiente n-ésimo de la serie es $c\binom{n+m-1}{n}\lambda^n$. A partir de ésta se puede entonces obtener una expresión general del coeficiente u_n de $U(x)$ por medio de la ecuación 4.20,

$$u_n = \sum_{i=1}^{s} \sum_{j=1}^{m_i} c_{ij} \binom{n+j-1}{n} (\lambda_i)^n \tag{4.21}$$

Así se obtiene una expresión cerrada de la sucesión $\{u_n\}$. Observemos que las constantes que aparecen a la derecha de la igualdad son: (*i*) las raíces λ_i del polinomio característico (que tiene por coeficientes los de la recurrencia) y (*ii*) los coeficientes c_{ij} que se obtienen a partir de estas raíces y de los coeficientes del polinomio $C(x)$, determinados a partir de k valores de la sucesión. La ecuación 4.21 tiene un aspecto complicado que puede oscurecer el resultado. En realidad, lo que hemos demostrado se puede poner de forma más simple como en la proposición siguiente:

Proposición 4.9. Sea u_n una sucesión que satisface la ecuación de recurrencia lineal

$$u_n = a_1 u_{n-1} + \cdots + a_k u_{n-k} \qquad n \geqslant k$$

y sean $\lambda_1, \ldots, \lambda_s$ las raíces del polinomio característico de la ecuación,

$$P(x) = x^k - a_1 x^{k-1} - \cdots - a_k$$

con multiplicidades $m_1, \ldots, m_s$. Entonces, existen polinomios $P_1(x), \ldots, P_s(x)$, donde cada $P_i(x)$ tiene grado como mucho $m_i - 1$, de manera que

$$u_n = \sum_{i=1}^{s} P_i(n) \lambda_i^n \tag{4.22}$$

[2]El lector que no conozca estas cuestiones puede consultar, por ejemplo, [2, Cap. 5].

Además, los coeficientes de los polinomios $P_i(x)$ se pueden determinar a partir de los k primeros valores de la sucesión, $u_0, u_1, \ldots, u_{k-1}$.

En la resolución práctica de ecuaciones de recurrencia, no es preciso disponer de una expresión explícita como la de la ecuación 4.21. El enunciado de la proposición anterior resulta suficiente y la ecuación 4.22 proporciona una manera más simple y directa de obtener los coeficientes u_n. Para ello sólo es preciso determinar los coeficientes de los s polinomios $P_i(x)$ que aparecen en esta ecuación. Esto se puede hacer poniendo en la ecuación 4.22 los valores iniciales $u_0, u_1, \ldots, u_{k-1}$ (o cualquier otro conjunto de k valores conocidos). Veremos ahora un ejemplo para encontrar una expresión cerrada de la sucesión de Fibonacci,

$$F_n = F_{n-1} + F_{n-2}, \quad n \geqslant 2, \quad F_0 = 0, \quad F_1 = 1$$

El polinomio característico de la recurrencia es $x^2 - x - 1$, que tiene las raíces

$$\lambda_1 = \frac{1+\sqrt{5}}{2} \quad \text{y} \quad \lambda_2 = \frac{1-\sqrt{5}}{2}$$

ambas de multiplicidad 1. De acuerdo con la proposición 4.9, el término general de la sucesión es

$$F_n = c_1 (\frac{1+\sqrt{5}}{2})^n + c_2 (\frac{1-\sqrt{5}}{2})^n \tag{4.23}$$

Usando los valores iniciales $F_0 = 0$ y $F_1 = 1$, obtenemos

$$\left.\begin{aligned} c_1 + c_2 &= 0 \\ c_1\lambda_1 + c_2\lambda_2 &= 1 \end{aligned}\right\}$$

de donde $c_1 = \frac{1}{\sqrt{5}}$ y $c_2 = \frac{1}{\sqrt{5}}$, y

$$F_n = \frac{1}{\sqrt{5}} \left((\frac{1+\sqrt{5}}{2})^n - (\frac{1-\sqrt{5}}{2})^n \right) \tag{4.24}$$

Resulta bastante curioso que la solución venga dada en términos de $\frac{1+\sqrt{5}}{2}$. Este número es el llamado *número de oro*, que da la *proporción áurea*, considerada por los antiguos griegos como la relación perfecta entre medidas y que se usaba tanto en las construcciones arquitectónicas como en escultura. Esta proporción aparece también en ciertas manifestaciones orgánicas naturales (por ejemplo, las medidas de los huesos de los dedos humanos siguen esta proporción). Resulta curioso también que la combinación de números irracionales en la ecuación 4.24 dé siempre un número natural F_n, cosa que sería difícil de demostrar si no dispusiésemos de una definición previa de los términos de la sucesión. Estos hechos sorprendentes forman parte de la popularidad de la sucesión de Fibonacci.

4.4 Números combinatorios

En esta sección revisaremos algunos de los números combinatorios que hemos obtenido en los capítulos anteriores bajo la óptica de las funciones generadoras. Completaremos también el repertorio de números combinatorios introduciendo los números de Stirling y de Bell. Recordemos que la otra clase importante de números combinatorios, los coeficientes binomiales, ya ha sido tratada en la sección 2.

Desarreglos

En la sección 2 hemos obtenido la ecuación de recurrencia

$$D_n = (n-1)(D_{n-1} + D_{n-2}), \quad n \geqslant 2, \quad D_0 = 1, \quad D_1 = 0 \tag{4.25}$$

para el número de desarreglos de n símbolos, D_n. De manera similar a la manera como se ha obtenido la función generadora de los coeficientes de una ecuación de recurrencia lineal, esta recurrencia conduce a una ecuación en $D(x)$.

Ejercicio 4.10. Sea $D(x)$ la función generadora (ordinaria) de la sucesión $\{D_n\}_{n\in\mathbb{N}}$ del número de desarreglos de n símbolos. Multiplicando por x^n la igualdad 4.25 y sumando para $n \geqslant 2$, demostrar que $D(x)$ satisface la ecuación $D(x)(1-x^2) - D'(x)(x+x^3) = 1$.

Aquí, sin embargo, el uso de la función generadora exponencial

$$E(x) = \sum_{n\geqslant 0} \frac{D_n}{n!} x^n$$

resulta más efectivo. Por ello, observemos que el número total de permutaciones de n símbolos es $n!$, y que cualquier permutación deja algún número k de elementos fijados, $k = 0, 1, \ldots, n$. El número de permutaciones que dejan exactamente k símbolos fijados es $\binom{n}{k} D_{n-k}$, ya que para cada una de las $\binom{n}{k}$ elecciones de k elementos que quedan fijados podemos realizar D_{n-k} desarreglos con el resto. Así pues,

$$n! = \sum_{k=0}^{n} \binom{n}{k} D_{n-k} \tag{4.26}$$

El aspecto de esta ecuación recuerda la expresión de los coeficientes de un producto de funciones generadoras exponenciales: si $G(x) = e^x$ es la función generadora de la sucesión $(1, 1, \ldots)$ y $E(x)$ es la función generadora exponencial de $\{D_n\}_{n\in\mathbb{N}}$, la ecuación 4.26 indica que el producto $E(x)G(x) = E(x)e^x$ es la función generadora exponencial $H(x)$ de la sucesión $\{\frac{1}{n!}\}_{n\in\mathbb{N}}$. Escribiendo

$$\frac{1}{1-x} = 1 + x + x^2 + \cdots = \frac{0!}{0!} + \frac{1!}{1!}x + \frac{2!}{2!}x^2 + \cdots = H(x)$$

obtenemos con sorprendente facilidad

$$E(x) = \frac{1}{1-x} e^{-x} \tag{4.27}$$

Esta expresión permite reobtener los valores de D_n tomando el coeficiente de grado n en ambos lados de la igualdad. Para obtener el coeficiente de x^n a la derecha de la igualdad es preciso hacer el producto de $(1+x+x^2+\cdots)$ y $(1-x+x^2/2!-x^3/3!+\cdots)$, de donde

$$D_n = n! \left(1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + (-1)^n \frac{1}{n!} \right) \tag{4.28}$$

Números de Catalan

Recordemos de la sección anterior que los números de Catalan satisfacen la ecuación de recurrencia

$$C_n = C_0 C_{n-1} + C_1 C_{n-2} + \cdots + C_{n-2} C_1 + C_{n-1} C_0 \tag{4.29}$$

La forma de esta recurrencia recuerda la del producto de dos series formales. Llamamos

$$C(x) = C_0 + C_1 x + C_2 x^2 + \cdots$$

a la función generadora de los números de Catalan. Observemos que el coeficiente u_{n-1} del producto $C(x) \cdot C(x)$ coincide, de acuerdo con la relación 4.29, con C_n, de manera que la sucesión de coeficientes de $(C(x))^2$ es la sucesión de números de Catalan trasladada una unidad. En términos de la función generadora, esto indica que tenemos la relación

$$C(x) - 1 = x(C(x))^2 \tag{4.30}$$

donde hemos restado $1 = C_0$ a la derecha de la igualdad, ya que a la izquierda no hay 'término independiente'. Si miramos esta relación como una ecuación de segundo grado con incógnita $C(x)$,

$$x(C(x))^2 - C(x) + 1 = 0$$

y resolvemos la ecuación como resolveríamos una ecuación de segundo grado, obtenemos

$$C(x) = \frac{1 \pm \sqrt{1-4x}}{2x} \tag{4.31}$$

Considerada como una función de variable real, el signo '+' hace que el valor de la función tienda a ∞ cuando x tiende a cero. En cambio, con la elección del signo '$-$', este límite vale $1 = C_0 = C(0)$. Así entonces, obtenemos la expresión de la función generadora de los números de Catalan de la forma

$$C(x) = \frac{1 - \sqrt{1-4x}}{2x} \tag{4.32}$$

Particiones

En esta parte introduciremos el uso de las funciones generadoras para tener una herramienta más en el análisis de las particiones de enteros positivos y veremos también cómo permite obtener resultados de forma simple.

El tipo de funciones generadoras que dan las particiones de un entero son similares a las que dan las combinaciones con repetición. Observemos primero que el coeficiente de x^n en el desarrollo de la expresión

$$(1+x+x^2+\cdots)(1+x^2+(x^2)^2+\cdots)\cdots(1+x^k+(x^k)^2+\cdots)$$

da el número de maneras de expresar n en sumandos menores o iguales a k. En efecto, al hacer el producto se obtiene x^n cada vez que n se puede expresar como

$$n = n_1 + 2n_2 + 3n_3 + \cdots + kn_k \tag{4.33}$$

Identificamos esta descomposición de n con la partición que tiene n_1 '1's, n_2 '2's, n_3 '3's, $\ldots$, n_k 'k's. Recíprocamente, cada una de estas descomposiciones proporciona una única descomposición de n como 4.33. Así pues, la función generadora $P_{\leqslant k}(x)$ del número $p_{\leqslant k}(n)$ de particiones de n en partes más pequeñas o iguales a k es

$$P_{\leqslant k}(x) = \frac{1}{(1-x)(1-x^2)\cdots(1-x^k)} \tag{4.34}$$

Este argumento puede hacerse extensivo a otras particiones similares. Por ejemplo, las funciones generadoras del número de particiones de n en partes pares (respectivamente, impares) más pequeñas que $2k$ (respectivamente, $2k-1$) son

$$P_{p,\leqslant k}(x) = \frac{1}{(1-x^2)(1-(x^2)^2)\cdots(1-(x^2)^k)}$$

$$P_{s,\leqslant k}(x) = \frac{1}{(1-x)(1-x^3)\cdots(1-x^{2k-1})}$$

Haciendo extensivo el razonamiento sin limitar el tamaño de las partes, obtenemos la función generadora de las particiones de un entero

$$P(x) = \frac{1}{\prod_{i\geqslant 1}(1-x^i)}$$

la de las particiones de un entero en partes pares,

$$P_p(x) = \frac{1}{\prod_{i\geqslant 1}(1-(x^2)^i)}$$

y la de particiones en partes impares,

$$P_s(x) = \frac{1}{\prod_{i\geqslant 1}(1-x^{2i-1})}$$

Un argumento similar, aún, proporciona el número de particiones de n en partes *diferentes*, $p_{\neq}(n)$,

$$P_{\neq}(x) = (1+x)(1+x^2)(1+x^3)\cdots = \prod_{i\geqslant 1}(1+x^i) \tag{4.35}$$

Con el uso de estas funciones generadoras se pueden obtener la mayoría de los resultados que hemos visto en el capítulo anterior. Para ilustrar el tipo de argumentos que se pueden usar, demostraremos aquí el siguiente:

Proposición 4.11. El número $p_{\neq}(n)$ de particiones de n en partes diferentes coincide con el número $p_s(n)$ de particiones de n en partes impares.

Demostración. A partir de la igualdad $(1-x^{2i}) = (1-x^i)(1+x^i)$, tenemos

$$\prod_{i\geqslant 1}(1-x^{2i}) = \left(\prod_{i\geqslant 1}(1+x^i)\right)\left(\prod_{i\geqslant 1}(1-x^i)\right) = P_{\neq}(x)\left(\prod_{i\geqslant 1}(1-x^i)\right)$$

de donde

$$P_{\neq}(x) = \frac{1}{\prod_{i\geqslant 1}(1-x^{2i-1})} = P_s(x)$$

□

Números de Stirling y de Bell

Relacionado con el problema de las particiones de un entero positivo, otro problema clásico es el de calcular el número de particiones de un conjunto. Una *partición* de un conjunto de n elementos $X = \{1,2,\ldots,n\}$ es una descomposición de X en unión disyunta de subconjuntos. Por ejemplo, las particiones de $X = \{1,2,3\}$ son

$$\{1\}\,\{2,3\}; \quad \{1,2\}\{3\}; \quad \{1,3\}\,\{2\}; \quad \{1\}\,\{2\}\{3\}; \quad \{1,2,3\}$$

El número de particiones de un conjunto de n elementos en k subconjuntos no vacíos se llama *número de Stirling de segundo tipo* y se denota por $\left\{{n \atop k}\right\}$. Por ejemplo, para cualquier n, $\left\{{n \atop 1}\right\} = 1$ (la única partición en un solo conjunto es X mismo) y $\left\{{n \atop n}\right\} = 1$ (todos los conjuntos de la partición tienen un solo elemento). Por convenio, tomaremos el valor $\left\{{n \atop 0}\right\} = 0$ si $n > 0$ y $\left\{{n \atop k}\right\} = 0$ si $k > n$.

Ejercicio 4.12. Demostrar que, para $n > 0$, $\left\{ {n \atop 2} \right\} = 2^{n-1} - 1$

Los números de Stirling tienen un cierto parentesco con los coeficientes binomiales. En particular, satisfacen una ecuación de recurrencia similar que deduciremos ahora. El conjunto P_k de particiones del conjunto $X = \{1, 2, \ldots, n\}$ en k partes no vacías se puede poner como unión disyunta del conjunto P_k^1 de particiones en las cuales $\{1\}$ es una parte, y su complementario P_k^0. Podemos establecer una biyección entre P_k^1 y el conjunto de particiones de $\{2, \ldots, n\}$ en $(k-1)$ partes simplemente eliminando la parte $\{1\}$, de manera que $|P_k^1| = \left\{ {n-1 \atop k-1} \right\}$. Por otro lado, cada partición de P_k^0 se puede obtener añadiendo el elemento 1 en una de las partes de una partición de $\{2, \ldots, n\}$ en k partes (hay k posibilidades) de manera que $|P_k^0| = k \left\{ {n-1 \atop k} \right\}$. De aquí,

$$\left\{ {n \atop k} \right\} = \left\{ {n-1 \atop k-1} \right\} + k \left\{ {n-1 \atop k} \right\} \qquad n > k > 0 \tag{4.36}$$

Observemos la similitud con la recurrencia $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ de los coeficientes binomiales. Como en éstos, la ecuación 4.36 permite calcular los números de Stirling para valores pequeños en una estructura similar al triángulo de Tartaglia:

$$\begin{array}{ccccccc} & & & 0 & 1 & & \\ & & 0 & 1 & 1 & & \\ & 0 & 1 & 3 & 1 & & \\ 0 & 1 & 7 & 6 & 1 & & \end{array}$$

Obtendremos ahora una expresión explícita de los números de Stirling de segundo tipo usando funciones generadoras. Como los coeficientes binomiales, estos números dependen de dos variables, n y k. En el caso de los coeficientes binomiales, hemos encontrado una función generadora para n fijado. Aquí un procedimiento similar encaja mal con la forma de la recurrencia a causa del factor k que multiplica al segundo sumando. En lugar de ello, podemos considerar la función generadora de $\left\{ {n \atop k} \right\}$ para k fijo,

$$S_k(x) = \sum_{n \geqslant 0} \left\{ {n \atop k} \right\} x^n$$

Multiplicando los dos lados de la igualdad 4.36 por x^n y sumando para $n \geqslant 0$ (recordemos que $\left\{ {n \atop k} \right\} = 0$ si $n < k$), obtenemos

$$S_k(x) = xS_{k-1}(x) + kxS_k(x) \tag{4.37}$$

de manera que

$$S_k(x) = \frac{x}{(1-kx)} S_{k-1}(x)$$

Escribiendo ahora $S_{k-1}(x)$ en términos de $S_{k-2}(x)$, ésta en términos de $S_{k-3}(x)$ y reiterando el proceso hasta $S_0(x) = 1$, obtenemos

$$S_k(x) = \frac{x^k}{(1-kx)(1-(k-1)x)\cdots(1-x)} \tag{4.38}$$

Esta es una función generadora racional, de manera que podemos usar la descomposición en fracciones simples como en la sección anterior para las recurrencias lineales, y obtenemos

$$S_k(x) = x^k \sum_{j=1}^{k} \frac{c_j}{(1-jx)} \tag{4.39}$$

para ciertas constantes $c_1, \ldots, c_k$. Para obtener una expresión de estas constantes, el procedimiento habitual consiste en multiplicar los dos lados de la ecuación que se obtiene de 4.38 y 4.39

$$\frac{1}{(1-kx)(1-(k-1)x)\cdots(1-x)} = \sum_{j=1}^{k} \frac{c_j}{(1-jx)}$$

por $(1-jx)$ y tomar $x = 1/j$, con lo que la expresión de la derecha es directamente c_j. De esta manera se obtiene (ver los detalles en el ejercicio 4.13)

$$c_j = \frac{1}{k!}(-1)^{k-j}\binom{k}{j} \tag{4.40}$$

Cada uno de los sumandos de la derecha de la igualdad 4.39 corresponde a la serie formal

$$\frac{c_j}{(1-jx)} = \sum_{n \geqslant 0} c_j j^n x^n$$

de manera que el coeficiente (n)-ésimo de la serie $S_k(x)$ es

$$\left\{ {n \atop k} \right\} = c_1 + c_2 2^n + \cdots + c_k k^n$$

Substituyendo los valores de las constantes que hemos encontrado en 4.40, obtenemos finalmente una forma cerrada para los números de Stirling de segundo tipo:

$$\left\{ {n \atop k} \right\} = \frac{1}{k!} \sum_{j=1}^{k} (-1)^{k-j} \binom{k}{j} j^n \qquad k \geqslant 1 \tag{4.41}$$

Ejercicio 4.13. Demostrar la fórmula de los coeficientes c_j de la expresión

$$\frac{1}{(1-x)(1-2x)\cdots(1-kx)} = \sum_{j=1}^{k} \frac{c_j}{(1-jx)}$$

Para ello multiplicar los dos lados de la igualdad por $(1-jx)$ y tomar $x = 1/j$, con lo que se obtiene

$$c_j = \frac{1}{j^{k-1}}(j-1)(j-2)\cdots 2\cdot 1\cdot(-1)\cdot(-2)\cdots(j-k)$$

Volviendo ahora al problema combinatorio asociado a los números de Stirling de segundo tipo, supongamos ahora que queremos contar el número de particiones de un conjunto de n elementos sin especificar el número de partes. Este es el llamado *número de Bell* de orden n, B_n y vale

$$B_n = \sum_{k=1}^{n} \left\{ {n \atop k} \right\} \tag{4.42}$$

Está claro que la fórmula 4.41, que proporciona los números de Stirling $\left\{ {n \atop k} \right\}$, puede servir para calcular los números de Bell. Los primeros de estos números, con el convenio que $B_0 = 1$, son

$$1, 1, 2, 5, 15, 52, \ldots$$

Hay otra manera de obtener los números de Bell. En primer lugar obtendremos una ecuación de recurrencia para B_{n+1}. Dado el conjunto $X = \{1, 2, \ldots, n, n+1\}$, consideremos el elemento 1. En cada partición de X, el elemento 1 está en una parte de $(k+1)$ elementos para una cierta $k = 0, 1, \ldots, n$. Hay $\binom{n}{k}$ maneras de escoger los otros elementos de esta parte, y hay B_{n-k} particiones de los $(n-k)$ elementos que quedan. Así entonces,

$$B_{n+1} = \sum_{k \geqslant 0} \binom{n}{k} B_{n-k} \tag{4.43}$$

La forma de esta ecuación de recurrencia sugiere el producto de funciones generadoras exponenciales de la proposición 4.7. Intentemos entonces obtener esta función generadora,

$$E(x) = \sum_{n \geqslant 0} B_n \frac{x^n}{n!}$$

Para ello, multiplicamos los dos lados de la ecuación 4.43 por $x^n/n!$ y sumamos para $n \geqslant 0$:

$$\sum_{n \geqslant 0} B_{n+1} \frac{x^n}{n!} = \sum_{n \geqslant 0} \sum_{k \geqslant 0} \binom{n}{k} B_{n-k} \frac{x^n}{n!} \tag{4.44}$$

En la parte izquierda de la igualdad tenemos la función generadora exponencial de la secuencia $\{B_n\}_{n \in \mathbb{N}}$ pero trasladada una unidad. En el caso de las funciones generadoras ordinarias, esta traslación se traduce en multiplicar por x. En el caso de las funciones generadoras exponenciales corresponde en cambio a la operación de derivación.

Ejercicio 4.14. Si $E(x) = \sum_{n\geqslant 0} u_n \frac{x^n}{n!}$ es la función generadora exponencial de la sucesión $(u_0, u_1, u_2, \dots)$, demostrar que la función generadora exponencial de la sucesión $(0, u_0, u_1, \dots)$ es $E'(x)$.

En cuanto al lado de la derecha, el coeficiente de $\frac{x^n}{n!}$ es

$$\sum_{k\geqslant 0} \binom{n}{k} B_{n-k}$$

que corresponde al coeficiente n-ésimo de la función generadora exponencial del producto de $E(x)$ y $G(x) = e^x$, siendo esta última la función generadora exponencial de la sucesión $(1, 1, 1, \dots)$. Con todo esto,

$$E'(x) = E(x)e^x \tag{4.45}$$

de donde $E'(x)/E(x) = e^x - 1$ y

$$E(x) = e^{e^x - 1} \tag{4.46}$$

Esta última expresión es una ilustración más de la versatilidad y la potencia del uso de las funciones generadoras para la resolución de problemas de enumeración.

Hemos estado hablando de los números de Stirling de segundo tipo. Esto, está claro, es porque hay una clase de números que se llaman números de Stirling de *primer tipo*. Acabaremos este capítulo introduciéndolos. En este caso, la función generadora servirá de instrumento para definir esta clase de números.

Recordemos que en el capítulo 2 hemos considerado la extensión de los coeficientes binomiales $\binom{m}{n}$ al caso en que m es un número real x cualquiera, tomando

$$\binom{x}{n} = \frac{1}{n!} x(x-1)\cdots(x-n+1)$$

Al desarrollar la expresión $x(x-1)\cdots(x-n+1)$, se obtiene una serie formal con los coeficientes de grado más grande n nulos,

$$x(x-1)\cdots(x-n+1) = \sum_{k=1}^{n} a_k x^k \tag{4.47}$$

Los coeficientes de esta serie son los llamados *números de Stirling de primer tipo* y se denotan por $s(n,k)$. Por convenio, $s(0,0) = 0$. Además, $s(n,k) = 0$ para $k \geqslant n$. Para valores pequeños de n, se puede obtener fácilmente una lista de los valores de $s(n,k)$:

$$\begin{array}{lll} n=1 & x & s(1,1)=1 \\ n=2 & x(x-1)=x^2-x & s(2,1)=-1,\ s(2,2)=1 \\ n=3 & x(x-1)(x-2)=x^3-3x^2+2x & s(3,1)=2,\ s(3,2)=-3,\ s(3,3)=1 \end{array}$$

En particular, la función generadora $s_n(x)$ de los números de Stirling $s(n,k)$ para n fijo es

$$s_n(x) = x(x-1)\cdots(x-(n-1))$$

Algunas veces se hace referencia a los números de Stirling de primer tipo como los valores absolutos de los coeficientes que aparecen en el desarrollo 4.47. La notación

$$\begin{bmatrix} n \\ k \end{bmatrix} = (-1)^{(n-k)} s(n,k) \tag{4.48}$$

está bastante extendida y la adoptaremos aquí.

El motivo por el cual estos números tienen un denominación tan próxima a los números de Stirling de segundo tipo es que estos últimos satisfacen una relación en cierta manera dual a la de los primeros. Así como podemos escribir

$$x(x-1)\cdots(x-n+1) = \sum_{k\geqslant 0} \begin{bmatrix} n \\ k \end{bmatrix} (-1)^{(n-k)} x^k$$

también podemos escribir

$$x^n = \sum_{k\geqslant 0} \begin{Bmatrix} n \\ k \end{Bmatrix} x(x-1)\cdots(x-k+1)$$

Ejercicio 4.15. Demostrar la identidad anterior usando la relación de recurrencia de los números de Stirling de segundo tipo.

Con estas relaciones se obtienen una buena cantidad de identidades combinatorias que relacionan coeficientes binomiales y números de Stirling, algunas de las cuales se consideran en los ejercicios al final del capítulo. En los ejercicios del capítulo 10 se dará también una interpretación combinatoria de los números de Stirling de primer tipo. Aquí acabaremos la sección obteniendo una relación de recurrencia para estos números que vuelve a tener similitud con las de los coeficientes binomiales y de los números de Stirling de segundo tipo que hemos considerado en esta sección.

Proposición 4.16. Los números $\begin{bmatrix} n \\ k \end{bmatrix}$ satisfacen la ecuación de recurrencia

$$\begin{bmatrix} n \\ k \end{bmatrix} = (n-1)\begin{bmatrix} n-1 \\ k \end{bmatrix} + \begin{bmatrix} n-1 \\ k-1 \end{bmatrix} \tag{4.49}$$

Demostración. Escribimos la función generadora $s_n(x)$ de los números $s(n,k) = (-1)^{(n-k)} \begin{bmatrix} n \\ k \end{bmatrix}$ como

$$x(x-1)\cdots(x-n+2)(x-n+1) =$$
$$x(x-1)\cdots(x-n+2)x - x(x-1)\cdots(x-n+2)(n-1)$$

de manera que

$$s_n(x) = xs_{n-1}(x) - (n-1)s_n(x)$$

Igualando ahora los coeficientes de mismo grado, obtenemos

$$s(n,k) = s(n-1,k-1) - (n-1)s(n-1,k)$$

que escrito en la notación $\begin{bmatrix} n \\ k \end{bmatrix}$ proporciona la identidad de la proposición. □

La ecuación de recurrencia 4.49 permite construir el triángulo de los números de Stirling de primer tipo de forma similar a como se ha hecho para los coeficientes binomiales y para los números de Stirling de segundo tipo,

$$\begin{array}{ccccccc} & & & 0 & & 1 & \\ & & 0 & & 1 & & 1 \\ & 0 & & 2 & & 3 & 1 \\ 0 & & 6 & & 11 & 6 & 1 \end{array}$$

Notas bibliográficas

Quizá la primera referencia a una ecuación de recurrencia es el *Liber Abaci* (1203) de Leonardo da Pisa (conocido también como Fibonacci), donde se trata el problema de crecimiento de una población que da lugar a la sucesión de Fibonacci. Fue el matemático francés del siglo XIX, F. E. A. Lucas, quien dio el nombre a la sucesión.

Las funciones generadoras surgen inicialmente como una rama del análisis que se llamó análisis combinatorio, aunque la manipulación de series ya había sido un recurso muy utilizado por matemáticos como Newton, Euler, Gauss o Lagrange. Uno de los textos que intentan sistematizar el análisis combinatorio es el libro clásico de MacMahon (1917) [3]. Por otra parte, en el libro de H. S. Wilf [4] se puede encontrar un texto moderno y sugerente sobre el tema.

Los números combinatorios que se han tratado en este capítulo no agotan la familia de números combinatorios útiles para problemas de enumeración. En el libro de Graham, Knuth y Patashnik [1] hay un extenso capítulo dedicado a los números combinatorios.

Bibliografía

[1] R. L. Graham, D. E. Knuth, O. Patashnik. *Concrete Mathematics*, Addison Wesley, 1991.

[2] H. Childs. *A Concrete Introduction to Higher Algebra*, UTM, Springer Verlag, 1979.

[3] P. A. MacMahon. *Combinatory Analysis*, Chelsea Publishing Company, NY, 1960.

[4] H. S. Wilf. *Generatingfunctionology*, Academic Press, 1990.

Problemas

1. Sea A_n el número de maneras de subir n escalones en n pasos si en cada paso podemos subir uno o dos escalones indistintamente. Demostrar que la función generadora de la sucesión $\{A_n\}$ es $A(x) = 1/(1-x-x^2)$.

2. Una triangulación de un polígono regular es una partición del polígono en triángulos. Encontrar una ecuación de recurrencia para el número T_n de triangulaciones de un polígono regular de n lados. *Indicación: los números T_n satisfacen una ecuación de recurrencia similar a la de los números de Catalan.*

3. Llamar $f(n)$ al número de regiones en que el plano queda dividido por n rectas (por ejemplo, $f(1) = 2$ y $f(2) = 4$; suponer que no hay dos rectas paralelas ni que tres cualesquiera se puedan cortar en un único punto). Demostrar que $f(n) = 1 + n(n+1)/2$.

4. Encontrar una ecuación de recurrencia que dé el número $f(n)$ de palabras de longitud n de palabras de un alfabeto $\{A, B, C\}$, de manera que los símbolos A y B no aparezcan consecutivamente. Demostrar que la función generadora de la secuencia $f(n)$ es $(1+x)/(1-2x-x^2)$ y deducir que $f(n) = (1/2)((1+\sqrt{2})^n + (1-\sqrt{2})^n)$.

5. Encontrar la función generadora del número $f(k)$ de palabras de longitud k del alfabeto $\{0, 1, -1, 2, -2\}$ en las cuales hay un número par de ceros.

6. Sea $E(x)$ la función generadora exponencial de la sucesión $(u_0, u_1, \dots)$. Demostrar que la función generadora exponencial de la sucesión $(0, u_0, u_1, \dots)$ es $E'(x)$ y que, en general, la derivada k-ésima de $E(x)$ es la función generadora exponencial de la sucesión $(\underbrace{0, \dots, 0}_{k}, u_0, u_1, \dots)$.

   Usar el resultado anterior para demostrar que la función generadora exponencial de la sucesión de Fibonacci F_n es $E(x) = (1/\sqrt{5})(\lambda_1 e^{\lambda_1 x} - \lambda_2 e^{\lambda_2 x})$, donde λ_1, λ_2 son las raíces positiva y negativa respectivamente de $P(x) = x^2 - x - 1$.

7. Usando las expresiones $F_n = F_{n+2} - F_{n+1}$ y $F_n = F_{n-2} + F_{n-1}$, donde F_n es el n-ésimo número de Fibonacci, demostrar las identidades

   (a) $F_{n+k} = F_k F_{n+1} + F_{k-1} F_n$

(b) $F_{2n+1} = F_{n+1}^2 + F_n^2$

(c) $F_{2n} = F_n F_{n+1} + F_{n-1} F_n$

Deducir en particular que F_{kn} es un múltiplo de F_n.

8. Demostrar que cualquier número natural n se puede expresar de manera única como $n = F_{k_1} + F_{k_2} + \cdots + F_{k_r}$, donde $k_i \geqslant k_{i+1} + 2$, $i = 1, \ldots, r-1$ (escoger F_{k_1} como el número más grande de Fibonacci entre los menores que n y demostrar el enunciado por inducción).

9. En la transmisión de mensajes formados por palabras de ceros y unos, cada '0' se transmite en una unidad de tiempo y cada '1' en dos unidades de tiempo. Determinar el número $N(k)$ de mensajes que se pueden transmitir en k unidades de tiempo.

10. Determinar el número $h(n)$ de movimientos que es preciso hacer para resolver el problema de las torres de Hanoi del capítulo 1, sabiendo que $h(n) = 2h(n-1) + 1$, $n \geqslant 2$.

11. Encontrar la función generadora de la sucesión $\{2^n + 3^n\} = \{2, 5, 13, \ldots\}$.

12. Encontrar la expresión del término general de la recurrencia lineal homogénea

$$u_n = 5u_{n-1} + 6u_{n-2}, \quad u_0 = 0, \quad u_1 = 1$$

13. Usando la notación $[x]_n = x(x-1)\cdots(x-n+1)$, demostrar que

$$[x+y]_n = \sum_{k \geqslant 0} [x]_k [y]_{n-k}$$

14. Demostrar que $\left\{ {n \atop n-1} \right\} = \left[{n \atop n-1} \right] = \binom{n}{2}$

Parte II Teoría de grafos

Una de la partes de la matemática discreta que en estos últimos años ha experimentado un desarrollo más notable es la la *teoría de grafos*. Enmarcada dentro de la combinatoria, esta teoría permite modelar de forma simple cualquier sistema en el cual exista una relación binaria entre ciertos objetos; y es por esto que su ámbito de aplicación es muy general y cubre áreas que van desde la misma matemática—topología, probabilidad, análisis numérico, etc.—hasta las ingenierías eléctrica, de telecomunicación e informática, la investigación operativa, la sociología o, incluso, la lingüística.

En los capítulos siguientes se presentan los temas más importantes de la teoría de grafos: grafos y digrafos; planaridad; árboles y árboles generadores; grafos eulerianos y hamiltonianos; ciclos y cociclos fundamentales; flujos en redes de transporte; conectividad y apareamientos.

Capítulo 5

Grafos y digrafos

1. Definiciones básicas
2. Caminos, conectividad y distancia
3. Operaciones entre grafos
4. Digrafos
5. Representación matricial
6. Grafos y redes de interconexión
7. Planaridad: la fórmula de Euler
8. Caracterización de los grafos planares

En este capítulo se estudian los conceptos más básicos de la teoría de grafos y se introduce la relación de la teoría con una de sus aplicaciones importantes: el diseño de redes de interconexión. La última parte del capítulo estudia los grafos planares. Una de las aplicaciones interesantes del tema es el diseño de circuitos integrados e impresos.

5.1 Definiciones básicas

Un *grafo* $G = (V, E)$ es una estructura combinatoria constituida por un conjunto $V = V(G)$ de elementos llamados *vértices* y un conjunto $E = E(G)$ de pares no ordenados de vértices distintos llamados *aristas*. Si la arista $e = \{u, v\} = uv$ relaciona los vértices u y v, se dice que u y v son vértices *adyacentes* y también que el vértice u (o v) y la arista e son *incidentes*. De otro modo, los vértices se llaman *independientes*. Las aristas $e = uv$ y $f = wz$ son *aristas independientes* si no tienen vértices en común, es decir, $\{u, v\} \cap \{w, z\} = \emptyset$. El número de vértices de G, $|V(G)|$, es el *orden* del grafo y el número de aristas $|E(G)|$ es su *tamaño*. A menudo resulta útil representar un grafo mediante un dibujo donde los vértices son puntos y

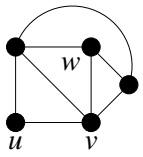


Figura 5.1: Grafo con orden 5 y tamaño 8

las aristas líneas que unen los vértices adyacentes. Así, por ejemplo, en el grafo representado en la figura 5.1, de orden 5 y tamaño 8, los vértices u y v son adyacentes, mientras que u y w son vértices independientes. A veces conviene ampliar esta definición de grafo para permitir la existencia de *lazos*, es decir, aristas que unen un vértice con él mismo, y *aristas paralelas* que unen un mismo par de vértices. En este texto, un grafo con lazos y/o con aristas paralelas se llamará *multigrafo*. En la figura 5.2 se muestra un multigrafo con un lazo l y aristas paralelas e y f.

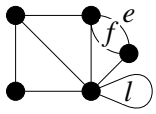


Figura 5.2: Multigrafo

Un grafo $G' = (V', E')$ es un *subgrafo* de $G = (V, E)$ si $V' \subset V$ y $E' \subset E$. Cuando $V' = V$, el subgrafo G' se llama *subgrafo generador* de G. Dado $V' \subset V$, si el subgrafo $G' = (V', E')$ contiene todas las aristas que unen en G dos vértices de V', entonces se dice que G' es el *subgrafo inducido* por V' y se denota con $G[V']$. Por ejemplo, dado $W \subset V(G)$, $G - W = G[V \setminus W]$ es el subgrafo que resulta de suprimir en G los vértices del conjunto W y todas las aristas incidentes con estos vértices. En particular, dado un vértice $u \in V$, el subgrafo $G - u = G - \{u\}$ es el obtenido eliminando el vértice u y todas las aristas incidentes con u. En cambio, la supresión de aristas no implica la eliminación de los vértices incidentes con estas aristas. Dado el grafo G y el subconjunto de aristas $F \subset E(G)$, el subgrafo $G - F$ es simplemente $(V, E \setminus F)$. Si F está formado por una única arista e, entonces el grafo $G - F$ se denotará por $G - e$. Ver la figura 5.3.

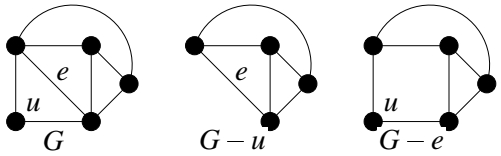


Figura 5.3: Supresión de un vértice y de una arista

Dado $u \in V(G)$, $\Gamma(u)$ denota el conjunto de vértices adyacentes con u y su número, $d(u)$, es el *grado* del vértice u. El *grado mímino* y el *grado máximo* son parámetros del grafo definidos por $\delta = \delta(G) = \min_{u\in V}\{d(u)\}$ y $\Delta = \Delta(G) = \max_{u\in V}\{d(u)\}$ respectivamente. Si $\delta = \Delta = d$, se dice que G es un grafo *d–regular*. Cuando G es un multigrafo, el grado de un vértice se define como el número de aristas incidentes con este vértice (contando cada lazo dos veces).

Al sumar todos los grados de los vértices de un grafo, cada arista $e = uv$ se cuenta dos veces (una vez desde cada uno de los dos vértices u y v incidentes con e). Así, se tiene el resultado siguiente:

Teorema 5.1. Dado un grafo (multigrafo) $G = (V, E)$, se cumple

$$\sum_{u\in V} d(u) = 2|E|$$

Este resultado se conoce como el lema de las manos estrechadas (*handshaking lemma*) porque se puede formular diciendo que en toda reunión de personas el número total de manos que se estrechan, cuando las personas se saludan entre ellas, es siempre par. Una consecuencia inmediata del teorema anterior es la siguiente:

Corolario 5.2. En todo grafo (multigrafo) G el número de vértices con grado impar es par.

Demostración. Separando en la suma $\sum_{u\in V} d(u)$ los términos correspondientes a vértices de grado par de aquellos correspondientes a vértices de grado impar se tiene

$$2|E| = \sum_{d(u)\ \text{par}} d(u) + \sum_{d(u)\ \text{impar}} d(u)$$

donde, siendo el primer sumando y la suma pares, el segundo sumando ha de ser también par. □

Dos grafos G y H se llaman *isomorfos* si existe una biyección $\phi : V(G) \longrightarrow V(H)$ entre los correspondientes conjuntos de vértices, llamada *isomorfismo*, que preserva las adyacencias, es decir, $uv \in E(G) \Longleftrightarrow \phi(u)\phi(v) \in E(H)$. Dos grafos isomorfos sólo se diferencian por la rotulación de los vértices (y, en general, por su representación gráfica). Por ejemplo, los grafos de la figura 5.4 son isomorfos con $\phi(i) = i'$.

Cuando dos grafos son isomorfos tienen, obviamente, el mismo orden, el mismo tamaño y el mismo número de vértices de un grado determinado. Todos estos parámetros son ejemplos de parámetros invariantes por isomorfismos. Un *invariante* de un grafo G es un número asociado a G que toma el mismo valor para cada grafo isomorfo a G. Dado un grafo, no se conoce ningún conjunto completo de invariantes, es decir, un conjunto de invariantes que determinen el grafo salvo isomorfismos.

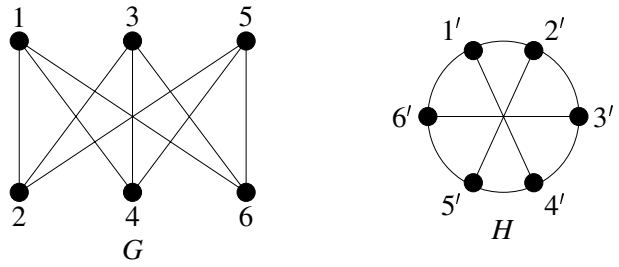


Figura 5.4: Grafos isomorfos

5.2 Caminos, conectividad y distancia

Dado un grafo $G = (V, E)$, una secuencia de vértices $u_0, u_1, \ldots, u_l$ con $u_{i-1}u_i \in E$, $1 \leqslant i \leqslant l$, y $u_{i-1}u_i \neq u_{j-1}u_j$ si $i \neq j$, se llama un *recorrido* R de longitud l entre u_0 y u_l. Es preciso notar que todas las aristas de un recorrido son distintas. Cuando interese considerar en G recorridos que también repiten aristas se indicará explícitamente. Un *circuito* es un recorrido cerrado, es decir, un recorrido en el cual $u_0 = u_l$. Cuando todos los vértices de R son distintos se tiene un *camino*, y un *ciclo* es un camino cerrado. Por ejemplo, en el grafo de la figura 5.5, u, v, w, z, v, t es un recorrido; u, v, z, w, t es un camino; u, v, w, z, v, t, u es un circuito y, finalmente, u, v, t, u es un ciclo. A veces interesa considerar un camino o un ciclo $C = u_0, u_1, \ldots, u_l$ como un subgrafo de G con $V(C) = \{u_0, u_1, \ldots, u_l\}$ y $E(C)$ el conjunto de aristas correspondiente.

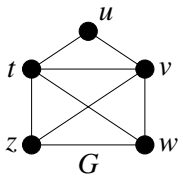


Figura 5.5: Recorridos en un grafo

Si entre todo par de vértices de G existe un camino, el grafo se dice *conexo* y, en este caso, la *distancia*, $d(u, v)$, entre dos vértices u y v es la longitud mínima de un camino entre estos vértices. En un grafo conexo G, la distancia es una métrica, ya que para todo $u, v, w \in V(G)$ se cumple:

- $d(u, v) \geqslant 0$ y $d(u, v) = 0$ si y sólo si $u = v$;
- $d(u, v) = d(v, u)$;
- $d(u, v) \leqslant d(u, w) + d(w, v)$.

Si G es conexo, su *diámetro*, $D = D(G)$, es la más grande de las distancias en el grafo:

$$D = \max_{u,v \in V(G)} \{d(u,v)\} \tag{5.1}$$

mientras que la *distancia media*, $\overline{D}$, se define como

$$\overline{D} = \frac{1}{|V(G)|^2} \sum_{u,v \in V(G)} d(u,v) \tag{5.2}$$

La *excentricidad* de un vértice $u \in V(G)$ es la máxima de las distancias entre u y los otros vértices de G. Desde este punto de vista, el diámetro del grafo también se puede definir como la máxima de las excentricidades. Por otra parte, la mínima de las excentricidades de los vértices del grafo, $r = r(G) = \min_{u \in V(G)} e(u)$, se llama *radio*. El *centro* del grafo, $Z(G)$, es el conjunto de vértices con excentricidad igual a r.

Proposición 5.3. Si G es un grafo conexo, entonces

$$r(G) \leqslant D(G) \leqslant 2r(G)$$

Demostración. La desigualdad $r(G) \leqslant D(G)$ es consecuencia inmediata de las definiciones de radio y diámetro. Por otra parte, si u y v son vértices de G tales que $d(u,v) = D(G)$ y $w \in Z(G)$, se tiene, por la desigualdad triangular, $D(G) \leqslant d(u,w) + d(w,v) \leqslant 2r(G)$. □

Ejercicio 5.4. Dar ejemplos de grafos tales que $r(G) = D(G)$ y $r(G) = 2D(G)$.

Un grafo G no conexo consta de dos o más *componentes* G_i donde cada $G_i = (V_i, E_i)$ es un subgrafo inducido que es conexo y maximal respecto de esta propiedad, en el sentido que, si $w \in V \setminus V_i$, entonces el subgrafo inducido $G[V_i \cup \{w\}]$ es no conexo.

Se dice que el vértice $w \in V(G)$ es un *vértice de corte* si el número de componentes de $G - w$ es más grande que el número de componentes de G. En particular, si G es conexo, la supresión de un vértice de corte desconecta G en dos o más componentes. Por ejemplo, en la figura 5.6 el vértice w es un vértice de corte y su supresión da lugar a un grafo con tres componentes.

Ejercicio 5.5. Demostrar que un vértice w es de corte si y sólo si existen $u, v \in V(G)$, $u, v \neq w$, tales que w pertenece a cada camino entre u y v.

Un grafo *trivial* es un grafo con un único vértice. Salvo este caso, se tiene el resultado siguiente:

Teorema 5.6. Todo grafo no trivial tiene al menos dos vértices que no son de corte.

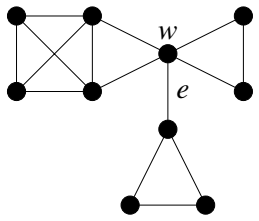


Figura 5.6: w es un vértice de corte y e es un puente

Demostración. Si el teorema fuese falso existiría un grafo conexo no trivial G con todos sus vértices de corte excepto, como mucho, uno. Sean $u, w \in V(G)$ tales que $d(u,w) = D(G)$ y supongamos que w es vértice de corte. Por otra parte, sea v un vértice tal que u y v pertenecen a componentes diferentes de $G - w$. Como todos los caminos entre u y v pasan por w, llegamos a la contradicción $d(u,v) > d(u,w) = D(G)$. □

En cuanto a las aristas, el concepto análogo al de vértice de corte es el de arista puente. Dado un grafo G, una arista $e \in E(G)$ es un *puente* si el número de componentes de $G - e$ es más grande que el número de componentes de G.

Ejercicio 5.7. Demostrar que si e es un puente, entonces $G - e$ tiene exactamente un componente más que G.

5.3 Operaciones entre grafos

Muchas veces conviene expresar la estructura de un grafo G en términos de grafos más simples. Antes de introducir diversas operaciones entre grafos que permiten estas descomposiciones, se consideran algunas clases de grafos particularmente interesantes.

En el *grafo completo* de orden n, K_n, cada vértice es adyacente a todos los otros. Así, el número de aristas de K_n es $\binom{n}{2}$. Un grafo formado por un único ciclo de longitud n se llama *grafo ciclo* de orden n y se denota C_n. Un *grafo camino* de orden n, P_n, es aquel formado por un único camino de longitud $n-1$. En la figura 5.7 se representan K_4, C_4 y P_4 respectivamente.

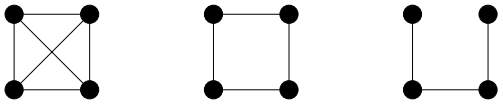


Figura 5.7: Grafos completo, ciclo y camino de orden 4

Ejercicio 5.8. Sea G un grafo con n vértices y k componentes. Demostrar la desigualdad siguiente:

$$|E(G)| \leqslant \frac{(n-k)(n-k+1)}{2}$$

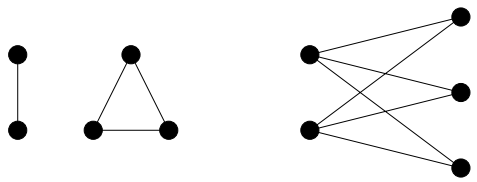
Un grafo G es *bipartito* con *clases de vértices* V_1 y V_2 si $V(G) = V_1 \cup V_2$, $V_1 \cap V_2 = \emptyset$, y cada arista une un vértice de V_1 con un vértice de V_2. En el grafo *bipartito completo*, $K_{n,m}$, $|V_1| = n$, $|V_2| = m$, y cada vértice de V_1 es adyacente con todos los vértices de V_2. Por ejemplo, los grafos representados en la figura 5.4 son isomorfos a $K_{3,3}$. De forma más general, un grafo G es *r–partito* si $V(G) = \bigcup_{i=1}^{r} V_i$, $V_i \cap V_j = \emptyset$, $i \neq j$, y cada arista une vértices de clases diferentes.

El *complemento* $\overline{G}$ de un grafo G es el grafo con conjunto de vértices $V(\overline{G}) = V(G)$ y $uv \in E(\overline{G})$ si y sólo si los vértices u y v son independientes en G. El complementario del grafo completo K_n no tiene ninguna arista y se llama grafo *nulo* N_n de orden n.

Dados dos grafos G y H, su *unión* $G \cup H$ es el grafo $(V(G) \cup V(H), E(G) \cup E(H))$. Por ejemplo, si G es un grafo no conexo con componentes $G_1, G_2, \ldots, G_n$, entonces $G = G_1 \cup G_2 \cup \cdots \cup G_n$. Ver la figura 5.8a. También, si $G_1, G_2, \ldots, G_n$ son subgrafos generadores de G disyuntos en aristas y $\bigcup_i E(G_i) = E(G)$, entonces podemos expresar G como $G = G_1 \cup G_2 \cup \cdots \cup G_n$.

La *intersección* $G \cap H$ es el grafo $(V(G) \cap V(H), E(G) \cap E(H))$.

La *suma* $G + H$ de los grafos G y H, con conjuntos de vértices disyuntos, tiene también conjunto de vértices $V(G) \cup V(H)$, pero ahora el conjunto de aristas es $E(G) \cup E(H) \cup \{uv \mid u \in V(G) \text{ y } v \in V(H)\}$. Por ejemplo, el grafo bipartito completo $K_{m,n}$ se puede expresar como $\overline{K_m} + \overline{K_n}$ (Fig. 5.8b).



(a) (b)

Figura 5.8: (a) $K_2 \cup K_3$ (b) $\overline{K_2} + \overline{K_3}$

El *producto cartesiano* $G \times H$, con $V(G) \cap V(H) = \emptyset$, es el grafo que tiene por conjunto de vértices $V(G) \times V(H)$ y dos vértices $(u_g, u_h), (v_g, v_h) \in V(G \times H)$ son adyacentes si y sólo si

$$u_g = v_g \text{ y } u_h v_h \in E(H)$$

o

$$u_h = v_h \text{ y } u_g v_g \in E(G)$$

Por ejemplo, el *hipercubo* de dimensión n, Q_n, que tiene por conjunto de vértices

$$V(Q_n) = \{u = (x_1, x_2, \ldots, x_n) \mid x_i \in \{0,1\}\}$$

siendo dos vértices u y v adyacentes si y sólo si las secuencias correspondientes difieren sólo en un dígito, puede expresarse como

$$Q_n = \overbrace{K_2 \times K_2 \times \cdots \times K_2}^{n}$$

El orden de Q_n es 2^n. Ver la figura 5.9.

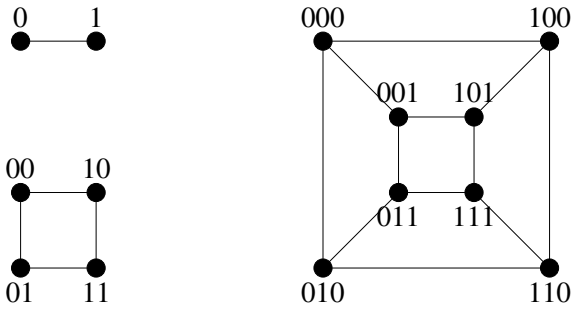


Figura 5.9: Los hipercubos de dimensión 1, 2 y 3

Finalmente, otra operación interesante es la *suma binaria* de G y H definida por $G \oplus H = (V(G) \cup V(H), E(G) \triangle E(H))$ donde $\triangle$ denota la diferencia simétrica de conjuntos. Es decir, e es una arista de $G \oplus H$ si e es arista de G o de H, pero no de los dos grafos a la vez. Cuando los grafos G y H tienen conjuntos de aristas disyuntos, $G \oplus H$ es isomorfo a $G \cup H$. Esta operación se utilizará en el capítulo 7 en relación a la estructura de ciclos de un grafo.

5.4 Digrafos

El concepto de *grafo dirigido* o *digrafo* deriva directamente del de grafo exigiendo que las aristas, ahora llamades *arcos*, sean pares ordenados de vértices distintos. Así, un digrafo $G = (V,A)$ es una estructura combinatoria formada por un par (V,A) de conjuntos disyuntos tales que $A \subset V \times V$. Si $a = (u,v) \in A$, decimos que el vértice u es *adyacente hacia el* vértice v y que v es *adyacente desde u*. Como en el caso no dirigido, cuando se permiten lazos y/o arcos paralelos se tiene un multidigrafo; ver la figura 5.10.

En un digrafo $G = (V,A)$ es preciso distinguir entre *grado de entrada* $d^-(u) = |\Gamma^-(u)|$ y *grado de salida* $d^+(u) = |\Gamma^+(u)|$, siendo ahora $\Gamma^-(u) = \{w \in V : (w,u) \in A\}$ y $\Gamma^+(u) =$

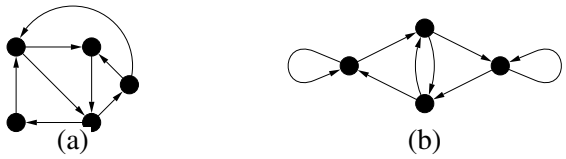


Figura 5.10: (a) Digrafo (b) Multidigrafo

$\{w \in V : (u,w) \in A\}$ los conjuntos de vértices adyacentes hacia y desde u respectivamente. El digrafo es d–regular si $d^+(u) = d^-(u) = d$, para todo $u \in V$. El resultado correspondiente al teorema 5.1 es el siguiente:

$$\sum_{u \in V} d^+(u) = \sum_{u \in V} d^-(u)$$

Los conceptos de recorrido, camino, circuito y ciclo tienen ahora un carácter dirigido. Por ejemplo, un camino desde el vértice u_0 hasta el vértice u_l es una sucesión de vértices distintos $u_0, u_1, \ldots, u_l$ tal que u_{i-1} es adyacente hacia u_i, $1 \leqslant i \leqslant l$. La distancia en un digrafo no tiene la propiedad simétrica, ya que, en general, $d(u,v) \neq d(v,u)$. El diámetro y la distancia media se definen también por las ecuaciones 5.1 y 5.2.

Un digrafo G es *simétrico* si dados $u, v \in V(G)$ se tiene $(u,v) \in A(G)$ si y sólo si $(v,u) \in A(G)$. Así, un grafo G se puede representar por su *digrafo simétrico asociado* G^* obtenido a partir de G substituyendo cada arista uv de G por el par de arcos (u,v), (v,u).

Dado un digrafo G, su grafo (multigrafo) *subyacente* es el grafo (multigrafo) que resulta de G cuando se suprime la orientación de los arcos.

En un digrafo se pueden distinguir diferentes tipos de conectividad. El digrafo G es *débilmente conexo* si el grafo (multigrafo) que resulta al suprimir las direcciones de los arcos es conexo. G es *unilateralmente conexo* si para todo $u, v \in V(G)$ existe un camino de u hacia v o un camino de v hacia u. Finalmente, el digrafo G es *fuertemente conexo* si para todo $u, v \in V(G)$ existe un camino de u hacia v. La conectividad fuerte implica la unilateral y esta implica la débil.

Como en el caso de grafos, un isomorfismo entre dos digrafos es una biyección entre sus conjuntos de vértices que preserva las adyacencias dirigidas.

5.5 Representación matricial

Como ya se ha indicado, un grafo o un digrafo puede visualizarse mediante un dibujo en que cada vértice se representa por un punto y cada arista o arco por una línea o línea dirigida respectivamente. No obstante, cuando se requiere el procesamiento por ordenador, resulta más

conveniente disponer de representaciones matriciales del grafo o digrafo. En esta sección se discuten las dos más importantes.

Matriz de adyacencia

La *matriz de adyacencia* de un grafo G con conjunto de vértices $V = \{v_1, v_2, \ldots, v_n\}$ es la matriz cuadrada $A = A(G)$, $n \times n$, definida por:

$$(A)_{ij} = \begin{cases} 1, & \text{si } v_i v_j \in E(G) \\ 0, & \text{de otro modo} \end{cases} \tag{5.3}$$

La matriz A es simétrica con elementos nulos en la diagonal. Por otra parte, el número de elementos iguales a 1 en la fila (o columna) i de $A(G)$ es $d(v_i)$, el grado del vértice v_i o, lo que es equivalente, el número de caminos de longitud 1 que comienzan en el vértice v_i. De forma más general, las potencias de A dan información sobre los caminos en G.

Teorema 5.9. El elemento $(A^k)_{ij}$ es igual al número de recorridos (pudiendo repetir vértices y/o aristas) de longitud k entre v_i y v_j.

Demostración. Por inducción sobre k. La proposición se cumple si $k = 1$. Para $k > 1$:

$$(A^k)_{ij} = (A^{k-1}A)_{ij} = \sum_{l=1}^{n} (A^{k-1})_{il} (A)_{lj}$$

El término general $(A^{k-1})_{il}(A)_{lj}$ de la suma anterior sólo es no nulo si $(A^{k-1}) \geqslant 1$ y $(A)_{lj} = 1$. Pero en este caso, si $(A^{k-1})_{il} = m$, existen m recorridos de longitud k entre v_i y v_j que acaban con la arista $v_l v_j$. Como v_l es un vértice cualquiera, sumando para todo l se tiene el resultado. □

Corolario 5.10. En un grafo conexo G, la distancia entre dos vértices v_i y v_j es k si y sólo si k es el menor entero no negativo tal que $(A^k(G))_{ij} \neq 0$.

Ejemplo: La matriz de adyacencia del grafo G de la figura 5.5, ordenando los vértices de la forma $v_1 = u$, $v_2 = v$, $v_3 = w$, $v_4 = z$ y $v_5 = t$, es:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Calculando A^2:

$$A^2 = \begin{pmatrix} 2 & 1 & 2 & 2 & 1 \\ 1 & 4 & 2 & 2 & 3 \\ 2 & 2 & 3 & 2 & 2 \\ 2 & 2 & 2 & 3 & 2 \\ 1 & 3 & 2 & 2 & 4 \end{pmatrix}$$

Así, por ejemplo, el elemento $(A^2)_{25} = 3$ indica la existencia de tres recorridos de longitud 2 entre los vértices $v_2 = v$ y $v_5 = t$: v,u,t; v,w,t y v,z,t. Por otra parte, la existencia de elementos nulos en la matriz A y el hecho que todos los elementos de A^2 sean diferentes de cero demuestra que el diámetro del grafo es 2 (¿por qué?).

Las permutaciones de las filas y de las columnas correspondientes de la matriz $A(G)$ equivalen a reordenar los vértices del grafo. Así se tiene el resultado siguiente:

Teorema 5.11. Dos grafos G y H son isomorfos si y sólo si

$$A(H) = P^{-1}A(G)P$$

donde P es una matriz de permutaciones.

La ecuación 5.3 define también la matriz de adyacencia de un multigrafo sin aristas paralelas. En este caso, la presencia de un lazo en el vértice v_i corresponde a tener $(A)_{ii} = 1$.

La matriz de adyacencia de un digrafo se define de forma análoga. Ahora bien, cualquier matriz binaria $n \times n$ (y no sólo las simétricas) puede ser la matriz de adyacencia de un digrafo de orden n. El número de unos en la fila i de la matriz es $d^+(v_i)$, mientras que $d^-(v_i)$ corresponde al número de unos en la columna i.

Matriz de incidencia

Sea G un grafo con $V(G) = \{v_1, v_2, \ldots, v_n\}$ y $E(G) = \{e_1, e_2, \ldots, e_m\}$. La *matriz de incidencia* $B = B(G)$ es la matriz $n \times m$ definida por:

$$(B)_{ij} = \begin{cases} 1 & \text{si } v_i \text{ es incidente con } e_j \\ 0 & \text{de otro modo} \end{cases} \tag{5.4}$$

Como en el caso de la matriz de adyacencia, el número de unos en la fila i de B corresponde al grado del vértice v_i. En cambio, cada columna contiene exactamente dos unos.

Las matrices de incidencia permiten representar multigrafos sin lazos. En este caso, las columnas repetidas de $B(G)$ manifiestan la existencia de aristas paralelas.

Si G es un digrafo, $B(G)$ se define como

$$(B)_{ij} = \begin{cases} 1, & \text{si } v_i \text{ es incidente hacia el arco } a_j \\ -1, & \text{si } v_i \text{ es incidente desde el arco } a_j \\ 0, & \text{de otro modo} \end{cases}$$

Las matrices de adyacencia y de incidencia de un grafo G representan su estructura y por ello han de estar relacionadas. Para expresar esta relación, sea D la matriz diagonal tal que $(D)_{ii}$ es el grado $d(v_i)$ del vértice i-ésimo de $V(G)$.

Teorema 5.12. $BB^t = A + D$

Demostración. Sea $V(G) = \{v_1, v_2, \ldots, v_n\}$ y $E(G) = \{e_1, e_2, \ldots, e_m\}$. Si $i \neq j$,

$$(BB^t)_{ij} = \sum_{k=1}^{m} b_{ik} b_{jk} = a_{ij}$$

ya que $b_{ik}b_{jk}$ sólo es diferente de 0 (y vale 1) si $e_k = v_i v_j$.

De otro modo, si $i = j$,

$$(BB^t)_{ii} = \sum_{k=1}^{m} b_{ik} b_{ik} = d(v_i)$$

ya que, ahora, $\sum_{k=1}^{m} b_{ik}b_{ik}$ cuenta el número de aristas incidentes con v_i. □

Ejercicio 5.13. Comprobar el teorema anterior en el grafo de la figura 5.5.

Las matrices $A(G)$ y $B(G)$ constituyen la base de las estructuras de datos más comúnmente utilizadas para representar un grafo o un digrafo en la memoria de un ordenador. Sin embargo, si $|E(G)|$ es mucho menor que $|V(G)|^2$, la forma más económica de representación será una *lista de incidencia*, que es una lista encadenada que da para cada vértice v las aristas o los arcos de los cuales v es incidente.

5.6 Grafos y redes de interconexión

El diseño de redes de interconexión se ha convertido en un problema fundamental en las ingenierías de telecomunicación y telemática, así como también en las ciencias de los computadores. Así, por ejemplo, la tecnología *VLSI* de integración de circuitos a gran escala permite, actualmente, la construcción de sistemas de cálculo constituidos por miles de procesadores conectados entre ellos, mediante módulos de memoria compartidos, con el objetivo de aprovechar el paralelismo inherente a las tareas que les son asignadas. Estos sistemas multiprocesadores requieren redes de interconexión de gran complejidad y es preciso señalar que la eficacia y el

rendimiento del sistema dependen, en gran parte, de la elección adecuada del mecanismo de interconexión entre sus diferentes elementos. Otros ejemplos de sistemas que requieren redes de interconexión complejas los encontramos en las diferentes redes telefónicas y telemáticas nacionales e internacionales.

Un grafo puede modelar una red de interconexión. Los vértices del grafo corresponden a los nodos de la red y las aristas a los enlaces. Si estos enlaces tienen carácter unidireccional, es decir, si sólo es posible la comunicación en un sentido, entonces la topología de la red correspondrá a un digrafo.

Los parámetros y las propiedades más importantes a tener en cuenta en el diseño de la red corresponden a parámetros análogos en el grafo que la modela. Ver, por ejemplo, la referencia [2]. Así, el retardo máximo de las comunicaciones entre nodos de la red será, en general, proporcional al diámetro D del grafo que la modela. Por otra parte, por razones de carácter técnico, el número máximo de enlaces incidentes con un nodo determinado suele ser limitado. Esto hace que el grado máximo Δ del grafo correspondiente esté también acotado. En este sentido, un problema de optimización interesante es el llamado *problema* (Δ, D) que consiste en encontrar el número máximo de vértices que puede tener un grafo con grado máximo Δ y diámetro D, así como también obtener grafos óptimos con este orden máximo. Ver los problemas 19 y 20 al final del capítulo.

Otro parámetro fundamental a tener en cuenta a la hora de diseñar una red de interconexión es su fiabilidad, es decir, la capacidad de la red para continuar funcionando, quizá de forma degradada, aunque falle alguno de sus elementos—nodos y/o enlaces. La fiabilidad de la red corresponde a la conectividad del grafo que la modela. La conectividad, que se estudiará en el capítulo 8, mide el número mínimo de vértices y/o aristas que se han de suprimir en el grafo para desconectarlo. Por ello interesa que el grafo que modela la red tenga un valor de conectividad tan grande com sea posible.

Finalmente, otra característica que es preciso señalar es que la red permita algoritmos de encaminamiento de los mensajes que sean a la vez simples y eficientes. Esta característica se debe tener en cuenta en el grafo que modela la red a la hora de definir las reglas de adyacencia entre sus vértices.

5.7 Planaridad: la fórmula de Euler

De forma intuitiva, un grafo es planar cuando se puede dibujar en el plano sin que sus aristas se crucen. De manera más precisa, diremos que un grafo G con orden n y tamaño m es *planar* si es posible distinguir en el plano un conjunto V de n puntos distintos—que corresponden a los vértices de G—y un conjunto E de m curvas, disyuntas dos a dos, excepto posiblemente en sus extremos—que corresponden a las aristas de G—de tal forma que, si la curva C corresponde

a la arista uv, entonces sólo los puntos extremos de C corresponden a vértices de G, precisamente u y v. Diremos que el par (V,E) es una *realización* plana de G. Por ejemplo, el grafo K_4 representado en la figura 5.11(a) es planar y en la figura 5.11(b) se da una realización plana del mismo. Una realización plana de un grafo planar G divide el plano en un cierto número

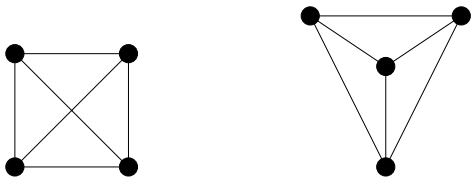


Figura 5.11: (a) K_4 (b) Realización plana de K_4

de regiones, siendo una *región* de G una porción maximal del plano tal que dos puntos cualesquiera de la misma pueden unirse mediante una curva C, de tal forma que ningún punto de C corresponde a un vértice de G ni pertenece a ninguna de las curvas correspondientes a las aristas de G. Toda realización plana de un grafo planar G da lugar a una región no acotada llamada la *región exterior* de G.

Ejercicio 5.14. Demostrar que todo grafo G que no contiene ciclos es planar y que toda realización plana de G da lugar a una única región.

El número de regiones, $r = r(G)$, no depende de la realización plana de G que se considere. En efecto, se cumple el resultado siguiente:

Teorema 5.15 (fórmula de Euler). Sea r el número de regiones en una realización plana de un grafo conexo y planar con orden n y tamaño m. Entonces,

$$n - m + r = 2 \tag{5.5}$$

Demostración. El resultado se puede demostrar por inducción sobre el número de aristas m. La fórmula 5.5 se cumple trivialmente si $m = 0$, ya que, en este caso, se tiene $n = 1$ y $r = 1$. Supongamos que el teorema es cierto para todos los grafos conexos y planares con número de aristas $k < m$, $m \geqslant 1$, y demostremos que entonces también se cumple si el tamaño del grafo es m. Sea G un grafo conexo y planar con m aristas. Si G no contiene ciclos, entonces $r = 1$ y, tal como se demostrará en el capítulo siguiente, su orden n tiene que ser $m + 1$. Por tanto, la fórmula 5.5 se cumple en este caso. De otro modo, si G contiene algún ciclo, consideremos una realización plana de G con r regiones y sea e una arista que pertenezca a un ciclo de G. Suprimiendo e, obtenemos una realización plana de $G - e$ con $r - 1$ regiones. Por otra parte,

el orden y el tamaño de $G-e$ son n y $m-1$ respectivamente. Así pues, por la hipótesis de inducción, se tiene $n-(m-1)+(r-1)=2$, es decir $n-m+r=2$. □

Ejercicio 5.16. Demostrar que si G es planar con orden n, tamaño m y k componentes, entonces se cumple $n-m+r=1+k$, siendo r el número de regiones de toda realización plana de G.

Los dos resultados siguientes, que dan condiciones necesarias para que un grafo sea planar, son consecuencia de la fórmula de Euler.

Teorema 5.17. Sea G un grafo planar y conexo con n vértices, $n \geqslant 3$, y m aristas. Entonces,

$$m \leqslant 3n-6$$

Demostración. Podemos suponer que G contiene algún ciclo, ya que, de otro modo, $m=n-1$ (como se ha indicado antes) y el teorema se cumple. En toda realización plana de G, la frontera de cada región contiene un número de aristas f_i más grande o igual a 3. Así, $\sum_{i=1}^{r} f_i \geqslant 3r$. Pero la suma anterior vale $2m$, ya que cada arista pertenece a la frontera de dos regiones (si existen vértices de grado 1, la suma anterior es menor que $2m$). Por tanto, $2m \geqslant 3r$. Aplicando ahora la fórmula de Euler, se tiene $2m \geqslant 3r = 3(2-n+m) = 6-3n+3m$, es decir, $m \leqslant 3n-6$. □

Ejercicio 5.18. Demostrar la validez del teorema 5.17 aunque G sea un grafo no conexo.

Corolario 5.19. Todo grafo planar tiene como mínimo un vértice de grado menor que 6.

Demostración. Recordemos que la suma de los grados de los vértices de un grafo vale $2m$, donde m es el número de aristas. Por tanto, si todos los vértices tuviesen grado más grande o igual a 6, se tendría $2m \geq 6n$, siendo n el orden del grafo. Pero si G es planar, por el teorema anterior, $2m \leqslant 6n-12 < 6n$. Así, algún vértice debe tener grado menor que 6. □

5.8 Caracterización de los grafos planares

Los grafos K_5 y $K_{3,3}$ juegan un papel esencial en la caracterización de los grafos planares.

Teorema 5.20. Los grafos K_5 y $K_{3,3}$ no son planares.

Demostración. El orden y el tamaño de K_5 son $n=5$ y $m=10$ respectivamente. Por tanto, $10=m>3n-6=9$, cosa que contradice el teorema 5.17. Así, K_5 no es planar.

El grafo $K_{3,3}$ es bipartito. No contiene, por tanto, ciclos de longitud 3. Así, si $K_{3,3}$ fuese planar, la frontera de cada región de una realización plana del grafo sería un ciclo de longitud al menos 4. Por tanto, sumando para todas las regiones, se tendría $4r \leqslant 2m$, donde $m=9$ es

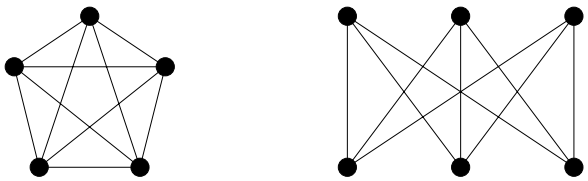


Figura 5.12: Los grafos K_5 y $K_{3,3}$

el número de aristas. Así, sería $4r \leq 18$. Por otra parte, dado que el orden de $K_{3,3}$ es $n = 6$, la fórmula de Euler implicaría $4r = 4(2 - n + m) = 4(2 - 6 + 9) = 20$, y se llegaría a una contradicción. Por tanto, tampoco $K_{3,3}$ es planar. □

Si G es un grafo planar, la "inserción" de un vértice de grado dos no modifica su planaridad. Por ejemplo, el grafo H de la figura 5.13 se ha obtenido del grafo G de la misma figura

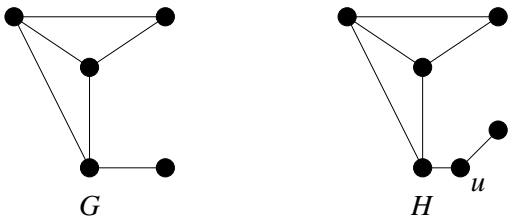


Figura 5.13: Subdivisión de una arista

insertando el vértice u.

De manera más formal, se dice que una arista $e = uv$ de un grafo G se *subdivide* si e se sustituye por un camino u, t, v de longitud dos, donde t es un nuevo vértice añadido a $V(G)$. El grafo H es una *subdivisión* del grafo G si H es isomorfo a G o se obtiene de G mediante una sucesión de subdivisiones de aristas.

La demostración de las proposiciones siguientes se deja como ejercicio.

Proposición 5.21. Si G es un grafo no planar, entonces todo grafo H, subdivisión de G, es no planar.

Proposición 5.22. Si G es un grafo planar, entonces cada subgrafo de G es planar.

Por tanto, si un grafo G contiene algún subgrafo que sea subdivisión de K_5 o $K_{3,3}$, entonces G es no planar. Es decir, si G es un grafo planar, entonces no puede contener ningún subgrafo que sea subdivisión de K_5 o $K_{3,3}$. Curiosamente, esta condición necesaria es también suficiente.

Esta proposición constituye el teorema de Kuratowski, que da la caracterización de los grafos planares.

Teorema 5.23. Un grafo es planar si y sólo si no contiene ningún subgrafo que sea subdivisión de K_5 o $K_{3,3}$.

Notas bibliográficas

Los libros de Bollobás [3], Bondy y Murty [4], Chartrand y Lesniak [5] y Harary [6] son textos clásicos que cubren los diversos temas de la teoría de grafos con más profundidad de lo que se hace en este libro. El libro de Wilson [7] constituye una buena introducción. También es preciso mencionar el texto de Basart [1]. El artículo de Bermond, Delorme y Quisquater [2] da una visión excelente de la aplicación de la teoría de grafos al diseño de buenas topologías para redes de interconexión.

Finalmente, el tema de planaridad y la demostración del teorema de Kuratowski se puede estudiar, por ejemplo, en el libro de Chartrand y Lesniak [5] mencionado antes.

Bibliografía

[1] J. M. Basart. *Grafs: Fonaments i Algorismes*, Publicacions de la Universitat Autònoma de Barcelona, 1993.

[2] J.–C. Bermond, C. Delorme, J.–J. Quisquater. "Strategies for interconection networks: some methods from graph theory", *J. Parallel and Distributed Computing*, **3**, pp. 433–449, 1986.

[3] B. Bollobás. *Graph Theory*, Springer, 1979.

[4] J. A. Bondy, U. S. R. Murty. *Graph Theory with Applications*, North Holland, 1976.

[5] G. Chartrand, L. Lesniak. *Graphs and Digraphs*, Wadsworth & Brooks, 1986.

[6] F. Harary. *Graph Theory*, Addison Wesley, 1972.

[7] R. J. Wilson. *Introducción a la Teoría de Grafos*, Alianza Universidad, vol. 367, Alianza Editorial, 1983.

Problemas

1. (a) Determinar todos los grafos de orden 4 y tamaño 3 con conjunto de vértices $V = \{1,2,3,4\}$;

   (b) determinar todos los grafos no isomorfos de orden 4 y tamaño 3;

   (c) determinar todos los grafos no isomorfos de orden 4;

   (d) determinar todos los grafos no isomorfos de orden 5.

2. Una lista $(d_1, d_2, \ldots, d_n)$ de enteros no negativos, $d_1 \leqslant d_2 \leqslant \cdots \leqslant d_n$, se dice que es la *secuencia de grados* de un grafo G de orden n si los vértices de G se pueden etiquetar de la forma $v_1, v_2, \ldots, v_n$ con $d(v_i) = d_i$, $1 \leqslant i \leqslant n$. Dibujar un grafo de orden 8 y secuencia de grados $(2,3,3,3,3,3,3,4)$.

3. Demostrar que la relación "ser isomorfo a" es, en la colección de todos los grafos, una relación de equivalencia.

4. Sea G un grafo con exactamente dos vértices de grado impar. Demostrar que existe un camino entre estos dos vértices.

5. Demostrar que para todo grafo G de orden 6, o bien G, o bien $\overline{G}$ contiene un triángulo (es decir, contiene K_3 como subgrafo).

6. Demostrar que, si G es un grafo no conexo, entonces su complemento $\overline{G}$ es conexo.

7. Demostrar que todo grafo G de orden n, $n > 1$, y tamaño más grande que $(n-1)(n-2)/2$ es conexo.

8. Sea G un grafo con n vértices tal que $d(u) \geqslant (n-1)/2$ para todo vértice u. Demostrar que G es conexo.

9. Sea e una arista de un grafo conexo G. Demostrar que $G - e$ es conexo si y sólo si e no pertenece a ningún ciclo de G.

10. Demostrar que un grafo es bipartito si y sólo si todos sus ciclos tienen longitud par.

11. Sean P_1 y P_2 dos caminos distintos entre dos vértices u, v de un grafo G. Demostrar que $P_1 \oplus P_2$ es un ciclo o un conjunto de ciclos en G.

12. Un *automorfismo* de un grafo G es un isomorfismo de G en él mismo. Dar ejemplos de automorfismos no triviales en los grafos siguientes: K_n, $K_{n,m}$ y el *grafo de Petersen P* representado en la figura 5.14.

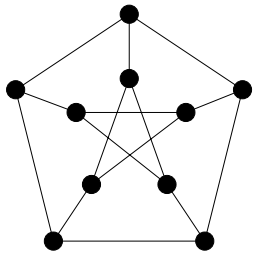


Figura 5.14: El grafo de Petersen P

13. Un grafo G es *vértice-transitivo* o *vértice-simétrico* si para todo $u, v \in V(G)$ existe un automorfismo ϕ de G tal que $\phi(u) = v$. Demostrar que el grafo de Petersen es vértice-transitivo.

14. El *giro* de un grafo G, $g = g(G)$, es la menor longitud de un ciclo contenido en G. Determinar los grafos 3-regulares con orden mínimo y

    (a) $g = 4$;

    (b) $g = 5$.

15. Sea Q_n el hipercubo de dimensión n.

    (a) Calcular su tamaño;

    (b) calcular su diámetro;

    (c) demostrar que es un grafo bipartito.

16. Sea G un digrafo con grado mínimo de salida $\delta^+(G) = \min_{u \in V(G)} \{d^+(u)\} > 0$. Demostrar que G contiene un ciclo dirigido.

17. Los *autovalores* de un grafo G son los autovalores de su matriz de adyacencia.

    (a) Calcular los autovalores de K_4, $K_{1,3}$ y del grafo de Petersen;

    (b) demostrar que la suma de los autovalores de G vale 0;

    (c) demostrar que la suma de sus cuadrados es $2|E(G)|$;

    (d) determinar los autovalores de K_n.

18. Sea G un grafo conexo. Demostrar que G es d–regular si y sólo si d es autovalor de G.

19. Demostrar que el número máximo de vértices que puede tener un grafo con grado máximo Δ y diámetro D es

$$n(\Delta, D) = \begin{cases} 2D+1, & \text{si } \Delta = 2 \\ \frac{\Delta(\Delta-1)^D-2}{\Delta-2}, & \text{si } \Delta > 2 \end{cases}$$

El número $n(\Delta, D)$ se conoce como *cota de Moore*.

20. Demostrar que el número máximo de vértices que puede tener un digrafo d–regular con diámetro D es

$$n(d, D) = \begin{cases} D+1, & \text{si } d = 1 \\ \frac{d^{D+1}-1}{d-1}, & \text{si } d > 1 \end{cases}$$

21. El *digrafo línea* $LG = (V_L, A_L)$ de un digrafo $G = (V, A)$ es el digrafo que se obtiene tomando como vértices los arcos de G, es decir, $V_L = A$, siendo $u = (u, u') \in V_L$ adyacente hacia $v = (v, v') \in V_L$ si y sólo si $u' = v$. Suponer que G es fuertemente conexo. Demostrar que:

    (a) si G es d–regular, entonces LG también lo es;

    (b) si G no es un ciclo dirigido, entonces $D(LG) = D(G) + 1$;

    (c) determinar LG y $L(LG)$ si G es el digrafo completo de orden 3 en que cada vértice es adyacente hacia los otros dos.

22. Demostrar que el grafo de Petersen contiene un subgrafo que es subdivisión de $K_{3,3}$ y que, por tanto, es no planar.

Capítulo 6

Árboles

1. Árboles
2. Árboles generadores
3. Número de árboles generadores
4. Obtención de todos los árboles generadores
5. Árboles generadores de coste mímimo

Supongamos que se quiere comunicar n nodos utilizando una red de interconexión que tenga el menor número posible de enlaces. Para permitir la comunicación entre dos nodos cualesquiera, el grafo G correspondiente a esta red tendrá que ser conexo y, por otra parte, no podrá contener ciclos, porque si tuviese y uv fuese una arista perteneciente a un ciclo, entonces el grafo $G - uv$ sería aún conexo y tendría menor número de aristas que G. Por ejemplo, el grafo "estrella" de la figura 6.1 muestra la solución al problema planteado que tiene diámetro mínimo $D = 2$.

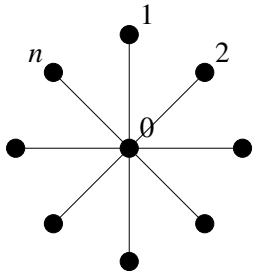


Figura 6.1: Estrella

La solución encontrada corresponde a un tipo de grafo conexo particularmente simple y con muchas aplicaciones: un árbol. En este capítulo estudiaremos las propiedades más importantes

de los árboles y también consideraremos los árboles como subgrafos generadores de otro grafo G. Este estudio lo aplicaremos a uno de los problemas clásicos de investigación operativa: la determinación de los árboles generadores de coste mínimo, y lo usaremos en el capítulo siguiente para el análisis de la estructura de ciclos fundamentales de un grafo.

6.1 Árboles

Un *árbol* es un grafo conexo y sin ciclos. La figura 6.2 muestra, por ejemplo, un árbol con orden 7. Este tipo de estructura combinatoria aparece en muchas aplicaciones de naturaleza

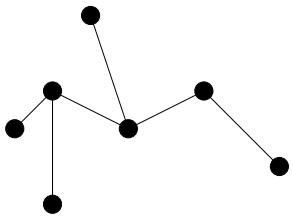


Figura 6.2: Árbol

distinta. Por ejemplo, en el diseño de algoritmos y estructuras de datos se utilizan los llamados *árboles de decisión* y, en particular, los *árboles binarios*, en los cuales existe un único vértice r, llamado raíz, que tiene grado 2 y los vértices restantes tienen grado 1 o 3. Así, en el árbol binario de la figura 6.3, los vértices v y w representan posibles alternativas a partir de u.

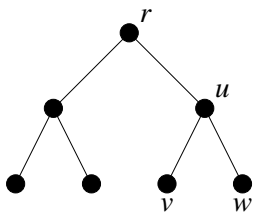


Figura 6.3: Árbol binario

El teorema siguiente da diversas caracterizaciones de los árboles. Si u, v son vértices independientes de un grafo $G = (V, E)$, representaremos por $G + uv$ el grafo resultante de añadir a G la arista uv, es decir, $G + uv = (V, E \cup \{uv\})$.

Teorema 6.1. Dado un grafo G, las proposiciones siguientes son equivalentes:

(a) G es un árbol.

(b) G es conexo y no tiene ciclos.

(c) Entre cada par de vértices de G existe un único camino.

(d) G es conexo con orden n y tamaño $n-1$.

(e) G es conexo, pero $G-e$ es no conexo para toda arista $e \in E(G)$.

(f) G es acíclico, pero $G+uv$ contiene un ciclo para todo par u, v de vértices independientes.

Demostración. (a) $\Longleftrightarrow$ (b). Esta equivalencia corresponde a la definición de árbol.

(b) $\Longleftrightarrow$ (c). Sea G un árbol. Si entre dos vértices dados de G hubiese dos caminos diferentes, C_1 y C_2, entonces su unión $C_1 \cup C_2$ sería un subgrafo de G que contendría al menos un ciclo. Recíprocamente, si entre cada par de vértices distintos de un grafo G hay un único camino, G es conexo y también tiene que ser acíclico; de otro modo, si Γ fuese un ciclo de G, existirían al menos dos caminos distintos entre cada par de vértices de Γ.

(a) $\Longleftrightarrow$ (d). Esta equivalencia se puede demostrar por inducción sobre el número de vértices. Si G es un árbol con $n = 1$, $n = 2$ o $n = 3$ vértices, se cumple trivialmente que su número de aristas es $n-1$; ver la figura 6.4. Supongamos ahora que todo árbol de orden k menor que n

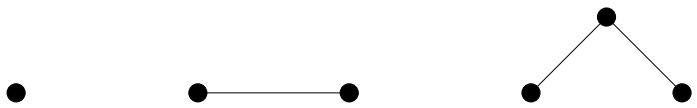


Figura 6.4: Árboles de orden 1, 2 y 3

tiene tamaño $k-1$. Sea G un árbol de orden n y sea $e = uv \in E(G)$. El camino u, v es el único en el árbol G entre los vértices u y v. Por tanto, el grafo $G-e$ está constituído por dos componentes conexos T_u y T_v que, por ser G acíclico, tampoco contienen ciclos y que, por tanto, son árboles. Si T_u y T_v tienen orden n_u y n_v, respectivamente, entonces, por la hipótesis de inducción, sus tamaños son, respectivamente, $n_u - 1$ y $n_v - 1$. Por tanto, el número de aristas de G es $(n_u-1)+(n_v-1)+1 = (n_u+n_v-1) = n-1$. Así pues, (a) $\Longrightarrow$ (d). Para demostrar la implicación contraria, (d) $\Longrightarrow$ (a), es preciso notar, en primer lugar, que la hipótesis de que G sea conexo es necesaria, ya que la figura 6.5 muestra un ejemplo de grafo con $|E(G)| = |V(G)| - 1$ que no es árbol. En segundo lugar, el teorema 1.1 del capítulo 7 implica que todo grafo con orden n y tamaño $n-1$, $n > 1$, tiene al menos dos vértices con grado 1 (ver el problema 1 al final del capítulo). Planteando ahora la hipótesis de inducción de que todo grafo conexo con menos de n vértices y tamaño una unidad menor que su orden es árbol (la hipótesis se cumple

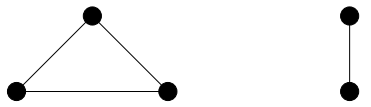


Figura 6.5: Grafo con $|E(G)| = |V(G)| - 1$ que no es árbol

otra vez trivialmente para los primeros casos), sea G un grafo conexo con n vértices y $n-1$ aristas. Si $u \in V(G)$ es uno de los vértices de G con grado 1, consideremos el grafo $G-u$ que tiene orden $n-1$ y tamaño $n-2$. Por la hipótesis de inducción, $G-u$ es árbol y, por tanto, G tiene que ser también árbol.

(a) $\Longleftrightarrow$ (e). Sea G conexo. Si G tuviese ciclos y e fuese una arista perteneciente a algún ciclo de G, entonces $G-e$ sería aún conexo. Por tanto, G no contiene ciclos y es un árbol. Recíprocamente, si G es árbol y $e = uv \in E(G)$, la arista e es el único camino entre sus vértices terminales u y v. Por tanto, $G-e$ es no conexo. Notemos, entonces, que en un árbol cada arista es una arista puente.

(a) $\Longleftrightarrow$ (f). Sea G un árbol. Dados dos vértices independientes u, v de G, sea C el único camino en G entre estos dos vértices. En el grafo $G+uv$, la arista uv forma con C un ciclo. Así, (a) $\Longrightarrow$ (f). Recíprocamente, si G no contiene ciclos, pero para todo par de vértices independientes u, v, el grafo $G+uv$ ya no es acíclico, entonces G debe ser conexo, de otro modo $G+uv$ no tendría ningún ciclo si u y v son vértices que pertenecen a componentes distintos de G. □

Un grafo conexo de orden n debe tener al menos $n-1$ aristas. En este sentido, los árboles son los grafos conexos de tamaño más pequeño. Recordemos también que el hecho de que un grafo conexo sin ciclos de orden n tenga tamaño $n-1$ se usó en el capítulo anterior para demostrar la fórmula de Euler y sus consecuencias.

Cada arista de un árbol es un puente. En cuanto a los vértices, se tiene el resultado siguiente, la demostración del cual se deja como ejercicio.

Teorema 6.2. Un vértice v de un árbol T es vértice de corte si y sólo si $d(v) > 1$.

El concepto de árbol se puede generalizar de la forma siguiente. Un *bosque* es un grafo acíclico. Así, cada componente de un bosque es un árbol.

Ejercicio 6.3. Demostrar que si G es un bosque con orden n y k componentes, entonces G tiene $n-k$ aristas.

6.2 Árboles generadores

Un *árbol generador* de un grafo G es un subgrafo generador de G que es árbol. Notemos que, si T es un árbol generador de G, entonces T es un árbol maximal contenido en G en el sentido siguiente: si $e \in E(G)$ es una arista que no pertenece a $E(T)$, entonces $T+e$ ya es un subgrafo de G que contiene un ciclo. La figura 6.6 muestra un grafo G y uno de sus árboles generadores.

Figura 6.6: T es un árbol generador de G

Teorema 6.4. Todo grafo conexo tiene un árbol generador.

Demostración. Dado $G=(V,E)$ conexo, sea $u \in V$ un vértice cualquiera. Para cada $v \in V$, $v \neq u$, escogemos un vértice w_v adyacente con v y tal que $d(u,w_v) = d(u,v)-1$ (w_v puede coincidir con u). Sea $E' \subset E$ el conjunto de aristas de la forma $w_v v$ obtenido de esta manera. Entonces, el subgrafo $T=(V,E')$ es árbol generador de G. En efecto, por construcción existe en T un camino entre u y cualquier otro vértice v. Así, T es un subgrafo conexo. Por otra parte, si T tuviese un ciclo Γ y w fuese un vértice de Γ tal que la distancia $d(u,w)$ fuese máxima, entonces el vértice w sólo podría ser adyacente hacia otro vértice del ciclo, cosa que contradice el hecho de que en un ciclo cada vértice es adyacente hacia otros dos. □

A partir de esta demostración resulta sencillo formular un algoritmo para obtener un árbol generador de un grafo conexo. Además, el árbol generador T obtenido tiene la propiedad de preservar las distancias desde el vértice u. Es decir, $d_G(u,v) = d_T(u,v)$, para todo $v \in V(G)$. En la figura 6.7 se muestra un árbol generador del grafo G de la figura 6.6, obtenido con el procedimiento descrito. Cada vértice de G se ha etiquetado con su distancia al vértice u indicado en la figura.

En la sección 6.5 se presentan, en un contexto más general que el considerado aquí, dos algoritmos clásicos de obtención de árboles generadores.

Si T es un árbol generador, se llaman *cuerdas* las aristas de G que no son aristas de T. Así, si c es una cuerda, el subgrafo $T+c = (V(G), E(T) \cup \{c\})$ contiene exactamente un ciclo. Por

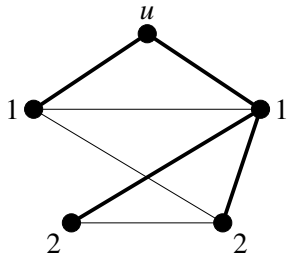


Figura 6.7: Obtención de un árbol generador

ejemplo, si G y T son los indicados en la figura 6.6, la figura 6.8 muestra el ciclo Γ_c creado por la cuerda c. Esta propiedad será ampliamente usada en el próximo capítulo, cuando se estudie la estructura de ciclos fundamentales de un grafo.

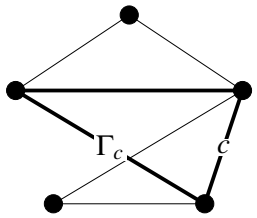


Figura 6.8: Ciclo creado por la cuerda c

6.3 Número de árboles generadores

Como se ha visto en el capítulo anterior, una posible representación matricial de un grafo conexo G no trivial se obtiene considerando su matriz de incidencia B. Las filas de B corresponden a los vértices del grafo y cada columna, que contiene exactamente dos unos, corresponde a una arista de G. Así, la información de la estructura de G queda también recogida en la llamada *matriz de incidencia reducida*, B_r, obtenida de B suprimiendo la fila correspondiente a un vértice dado v_r, que se puede tomar como vértice de referencia.

Ejemplo: El grafo G de la figura 6.9, con $V(G) = \{1,2,3,4\}$ y conjunto de aristas $E(G) = \{e_1, e_2, e_3, e_4\}$, donde $e_1 = 12$, $e_2 = 13$, $e_3 = 14$, $e_4 = 34$, tiene como matriz de incidencia B y

matriz de incidencia reducida B_r,

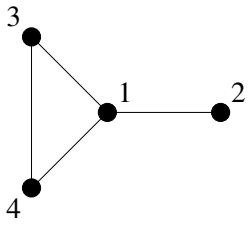


Figura 6.9: Obtención de B_r

$$B = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}, \quad B_r = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

donde se ha tomado como vértice v_r el vértice 1.

Ejercicio 6.5. Reconstruir el grafo del ejemplo anterior a partir de la matriz de incidencia reducida B_r.

Ahora, sea M una matriz obtenida de B_r cambiando, en cada columna de B_r, una de las entradas igual a 1 por -1. Además, si G tiene orden n y H es un subgrafo con $n-1$ aristas, designaremos por M_H la submatriz de M, cuadrada de orden $n-1$, las columnas de la cual corresponden a las aristas de H.

Ejercicio 6.6. Comprobar que la matriz M es la matriz de incidencia reducida de la matriz de incidencia de un digrafo obtenido de G orientando las aristas del grafo.

El resultado siguiente relaciona la estructura de H con el determinante de M_H.

Teorema 6.7. El subgrafo T es árbol generador de G si y sólo si $|det(M_T)| = 1$.

Demostración. Sea T un árbol generador de G, sea $u_1 \neq v_r$ un vértice de G que tenga grado uno en T y denotemos por b_1 la arista de T con la cual este vértice incide. En general, sea u_i, $i = 2, \ldots, n-1$, un vértice diferente de v_r que tenga grado uno en el árbol $T - \{u_1, \ldots, u_{i-1}\}$ y b_i la arista de este árbol con la cual u_i es incidente. Ahora, permutando adecuadamente las filas y las columnas de M_T, es posible obtener una matriz M'_T tal que $|(M'_T)_{ij}| = 1$ si u_i es incidente con b_j y $(M'_T)_{ij} = 0$ de otro modo. Pero, por construcción, M'_T es triangular inferior con $|(M'_T)_{ii}| = 1$, $i = 1, 2, \ldots, n-1$. Por tanto, $|det(M'_T)| = |det(M_T)| = 1$.

En cambio, si H es un subgrafo generador de G no conexo y con $n-1$ aristas, la suma de las filas de M_H correspondientes a los vértices de un componente de H que no contenga el vértice v_r da el vector 0, ya que en cada columna sumamos 1 y -1. Por tanto, $det(M_T) = 0$.

Finalmente, si H no es subgrafo generador de G y no contiene v_r, entonces la suma de las filas de M_H da nuevamente el vector 0. Si H no es subgrafo generador y contiene v_r, entonces alguna de las filas de M_H es el vector 0. En cualquier caso, $det(M_T) = 0$. □

Ejercicio 6.8. Construir las matrices M_H de los subgrafos generadores del grafo de la figura 6.9.

El resultado anterior permite, ahora, contar el número de árboles generadores de un grafo conexo no trivial.

Teorema 6.9. El número $\tau(G)$ de árboles generadores de un grafo conexo G no trivial es

$$\tau(G) = det(MM^T)$$

Demostración. Se usará un resultado clásico de álgebra de determinantes, conocido como *teorema de Binet–Cauchy*: Si P y Q son matrices $r \times s$ y $s \times r$, respectivamente, con $r \leqslant s$, y P_i y Q_i, $i = 1, 2, \ldots, \binom{s}{r}$, son, respectivamente, las submatrices cuadradas $r \times r$ de P y Q de tal forma que, si P_i contiene las columnas $n_{i_1}, n_{i_2}, \ldots, n_{i_r}$ de P, entonces Q_i contiene las filas correspondientes $n_{i_1}, n_{i_2}, \ldots, n_{i_r}$ de Q, se cumple la fórmula $det(PQ) = \sum_i det(P_i)det(Q_i)$.

Aplicando este resultado,

$$det(MM^T) = \sum_i det(M_i)det(M_i^T) = \sum_i (det(M_i))^2$$

donde cada M_i corresponde a un subgrafo con $n-1$ aristas. Sin embargo, $(det(M_i))^2 = 1$ si y sólo si las columnas de M_i definen un árbol generador de G y $(det(M_i))^2 = 0$, de otro modo. Por tanto, $det(MM^T)$ cuenta el número total de árboles generadores. □

Ejemplo: Consideremos el grafo de la figura 6.6 con los vértices numerados como se indica en la figura 6.10 y las aristas ordenadas de la forma siguiente: $e_1 = 12$, $e_2 = 13$, $e_3 = 23$, $e_4 = 25$, $e_5 = 34$, $e_6 = 35$, $e_7 = 45$. Tomando como vértice de referencia $v_r = 5$, una posible matriz M es la siguiente:

$$M = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & -1 \end{pmatrix}$$

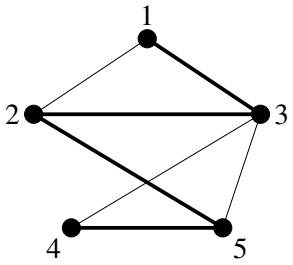


Figura 6.10: Árboles generadores y matriz de incidencia

Por ejemplo, al árbol generador definido por el conjunto de aristas $\{e_2, e_3, e_4, e_7\}$ corresponde la submatriz cuadrada de orden 4,

$$M_T = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

que tiene determinante 1. Sin embargo, al subgrafo H definido por el conjunto de aristas $\{e_1, e_2, e_3, e_7\}$, corresponde la submatriz

$$M_H = \begin{pmatrix} 1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

con determinante nulo.

El número de árboles generadores de G es $det(MM^T) = 21$.

Tomando $G = K_n$, se obtiene el siguiente corolario del teorema anterior, conocido como *fórmula de Cayley*, la demostración del cual se deja como ejercicio.

Corolario 6.10. El número de árboles de orden n, $n \geqslant 2$, es n^{n-2}.

En este resultado, los vértices de los árboles están numerados y dos árboles se consideran distintos cuando no son idénticos, aunque sean isomorfos.

Dado un grafo conexo no trivial $G = (V, E)$ y $e = uv \in E$, denotamos por G_e el multigrafo resultante de "contraer" la arista e, es decir, G_e se obtiene suprimiendo de G la arista e e identificando sus vértices terminales u, v; ver la figura 6.11. Un árbol generador de un multigrafo es, como en el caso de un grafo, un subgrafo que es árbol y contiene todos los vértices. El número $\tau(G)$ de árboles generadores de G verifica la siguiente fórmula recursiva:

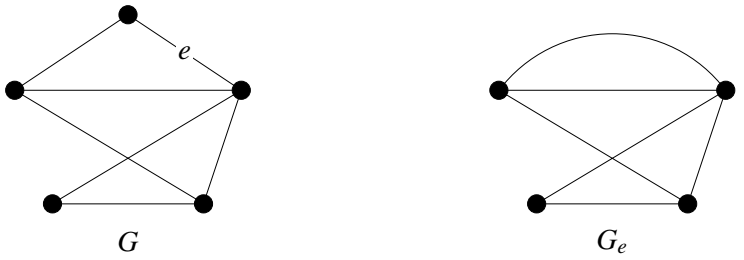


Figura 6.11: Contracción de una arista

Teorema 6.11. Dada $e \in E(G)$, se cumple $\tau(G) = \tau(G-e) + \tau(G_e)$.

Demostración. Sólo es preciso notar que los árboles generadores de G que no contienen la arista e se corresponden con los árboles generadores de $G-e$, mientras que los que sí contienen la arista e se corresponden con los árboles generadores de G_e. □

Ejercicio 6.12. Calcular el número de árboles generadores del multigrafo G_e de la figura 6.11.

6.4 Obtención de todos los árboles generadores

Sean T y T' dos árboles generadores de un grafo conexo no trivial G. Sea F el conjunto de aristas de T que son cuerdas respecto a T', y F' el conjunto de aristas de T' que son cuerdas respecto a T. Es decir, $F = E(T) \setminus E(T')$ y $F' = E(T') \setminus E(T)$. Dada $f \in F$, alguna de las aristas que constituyen el único ciclo Γ_f de $T' + f$ debe pertenecer a F', de otro modo T no sería acíclico. Sea F_γ el conjunto de estas aristas de Γ_f que también pertenecen a F'. Alguna $f' \in F_\gamma$ debe ser tal que f pertenezca también al único ciclo $\Gamma_{f'}$ de $T + f'$; de otro modo, la unión de los ciclos Γ_f y $\Gamma_{f'}$, $f' \in F_\gamma$, tendría un ciclo constituído únicamente por aristas de T.

Si ahora se considera el árbol $T_1 = (T + f') - f$, el número de aristas comunes a T_1 y T' es una unidad más grande que el de aristas comunes a T y T'. Si decimos que T_1 se ha obtenido de T mediante una *transformación elemental*, iterando el proceso descrito se obtiene el resultado siguiente:

Teorema 6.13. A partir de un árbol generador T se obtiene cualquier otro árbol generador T' mediante una sucesión de k transformaciones elementales, siendo k el número de aristas de T' que son cuerdas en T.

Ejercicio 6.14. Sean T y T' los árboles generadores indicados en las figuras 6.6 y 6.7 respectivamente. Obtener uno del otro mediante transformaciones elementales.

6.5 Árboles generadores de coste mínimo

Una *función de coste* definida sobre el conjunto de aristas de un grafo conexo no trivial $G = (V, E)$ es una aplicación c que asigna a cada arista $e \in E$ un número real no negativo $c(e)$. Si T es un árbol generador de G se define su coste, $c(T)$, mediante la fórmula:

$$c(T) = \sum_{e \in E(T)} c(e)$$

En muchas aplicaciones surge la cuestión de determinar un árbol generador T_m que tenga coste mínimo. Por ejemplo, si $c(e)$, $e = uv$, representa el coste de conectar directamente los nodos u y v de una red, y se trata de determinar la red más económica, entonces la solución correspondrá a encontrar el árbol generador de coste mínimo del grafo completo determinado por los nodos (¿por qué?). Este es un problema clásico conocido como el *problema del conector*.

Dos algoritmos clásicos para determinar un árbol generador de coste mínimo T_m de un grafo $G = (V, E)$ son los siguientes:

Entrada: $G = (V, E)$: un grafo conexo.
Algoritmo KRUSKAL

1. $F_0 \leftarrow \emptyset$.
2. **Para** $k = 1$ **hasta** $|V| - 1$ **hacer**
   $F_k \leftarrow F_{k-1} \cup \{b_k\}$, donde b_k es una arista de coste mínimo entre las aristas e de $E \setminus F_{k-1}$ tales que el subgrafo $H = (V, F_{k-1} \cup \{e\})$ es acíclico.
3. $T_m = (V, F_{|V|-1})$.

Salida: T_m: árbol generador de coste mínimo.

Entrada: $G = (V, E)$: un grafo conexo.
Algoritmo PRIM

1. Escoger un vértice $v_0 \in V$.
   Sea $V_0 = \{v_0\}$ y T_0 el árbol constituído por este único vértice.
2. **Para** $k = 1$ **hasta** $|V| - 1$ **hacer**
   Sea $v_k \in V \setminus V_{k-1}$ adyacente con algún vértice $w \in V_{k-1}$ y tal que la arista $e = v_k w$ tenga coste mínimo entre todas las que no pertenecen a $E(T_{k-1})$ y son incidentes con algún vértice de T_{k-1}.
   $V_k \leftarrow V_{k-1} \cup \{v_k\}$

$T_k \leftarrow T_{k-1} + e$

3. T_m=$T_{|V|-1}$.

Salida: T_m: árbol generador de coste mínimo.

Teorema 6.15. Los algoritmos de Kruskal y de Prim determinan un árbol generador de coste mínimo.

Antes de demostrar el teorema anterior, consideremos un resultado útil a la hora de decidir si un árbol generador T tiene coste mínimo. Designamos por $C(T)$ el conjunto de árboles generadores que se obtienen de T mediante una transformación elemental. Es decir, $T' = (T+c)-e$, donde c es una cuerda respecto a T y e es una arista de T que pertenece al único ciclo Γ_c de $T+c$.

Lema 6.16. T es un árbol generador de coste mínimo si y sólo si $c(T') \geqslant c(T)$ para todo $T' \in C(T)$.

Demostración. La condición es evidentemente necesaria. Para demostrar la suficiencia sea T un árbol generador que verifique la condición enunciada y sea T_m un árbol generador de coste mínimo. Sea F el conjunto de aristas de T que son cuerdas respecto a T_m y F_m el conjunto de aristas de T_m que son cuerdas respecto a T. Dada $f \in F$, sea $h \in F_m$ una arista del único ciclo Γ_f de $T_m + f$ tal que f pertenece también al único ciclo Γ_h de $T+h$. Consideremos el árbol generador $T' = (T+h)-f$. Dado que $T' \in C(T)$, se tiene $c(T') \geqslant c(T)$ y, por tanto, $c(h) \geqslant c(f)$. Por otra parte, si consideramos el árbol generador $T'_m = (T_m + f) - h \in C(T_m)$, obtenemos también $c(f) \geqslant c(h)$. Así, $c(f) = c(h)$ y, por tanto, $c(T') = c(T)$. Por otra parte, T' tiene con T_m una arista común más que T.

Vemos ahora que T' satisface también la condición $c(T'') \geqslant c(T')$ para todo $T'' \in C(T')$. Supongamos que exista un árbol generador $T'' = (T'+h')-f' \in C(T')$ tal que $c(T'') < c(T')$, donde h' es una cuerda respecto a T' y f' es una arista de T'. Si el único ciclo $\Gamma_{h'}$ de $T'+h'$ no contiene a h, entonces $\Gamma_{h'}$ es también el único ciclo de $T+h'$, y $(T+h')-f' \in C(T)$ tendría un coste menor que T. De otro modo, si el ciclo $\Gamma_{h'}$ contiene también la arista h, entonces la unión de $\Gamma_{h'}$ y Γ_h (el único ciclo de $T+h$) contiene el único ciclo $\Gamma'_{h'}$ que resulta añadiendo la cuerda h' al árbol T. Además, f' debe pertenecer también a $\Gamma'_{h'}$ y, otra vez, $(T+h')-f'$ tendría coste menor que T.

Dado que T' satisface también la condición del enunciado, podemos iterar el proceso y obtener una sucesión de árboles generadores $T, T', \ldots, T_m$, todos ellos con el mismo coste, que acaba en T_m. Así, $c(T) = c(T_m)$ y T tiene coste mínimo. □

Demostración del teorema 6.15. (a) Consideremos en primer lugar el algoritmo de Kruskal. El subgrafo T_m que se obtiene es un árbol generador. En efecto, T_m tiene $n-1$ aristas y, por construcción, es acíclico. Veamos que también es conexo. Sean $T_1, T_2, \ldots, T_r$ los componentes conexos de T_m, que son por tanto árboles. Así,

$$n-1 = |E(T)| = \sum_{i=1}^{r}(|V(T_i)|-1) = \sum_{i=1}^{r}|V(T_i)| - r = n-r$$

de donde se concluye $r=1$. Si existiese $T' = (T_m + h) - f \in C(T_m)$ con $c(T') < c(T_m)$, esto implicaría $c(h) < c(f)$ y, por tanto, cuando en el paso (2) del algoritmo se ha seleccionado la arista f, sería preciso haber seleccionado antes la arista h. Por tanto, T_m verifica la condición suficiente expresada en el lema 6.16 y T_m tiene coste mínimo.

(b) Por construcción, la sucesión $T_0, T_1, \ldots, T_{n-1} = T_m$ obtenida aplicando el algoritmo de Prim es una sucesión de subgrafos que son árboles, con órdenes $1, 2, \ldots, n$, respectivamente, que acaba, por tanto, en un árbol generador. Supongamos otra vez que exista $T' = (T_m + h) - f \in C(T_m)$ tal que $c(T') < c(T_m)$ y, por tanto, $c(h) < c(f)$. Designemos por Γ_h el único ciclo de T_m que contiene a f y h y sea T_k el árbol de la sucesión anterior tal que $T_k = T_{k-1} + f$. Sea $f_1 \neq f$ la otra arista de Γ_h, que no pertenece a $E(T_{k-1})$, incidente con un vértice de T_{k-1}. Por el proceso de obtención de T_k a partir de T_{k-1} se debe cumplir $c(f_1) \geqslant c(f)$. Sea ahora T_{k_1}, $k_1 > k$, el árbol de la sucesión construida por el algoritmo de la forma $T_{k_1} = T_{k_1-1} + f_1$, creado al añadir la arista f_1. Si $f_2 \neq f_1$ es la otra arista de Γ_h incidente con vértices de T_{k_1-1}, se tiene $c(f_2) \geqslant c(f_1)$. Repitiendo el razonamiento, obtendremos finalmente una arista f_s de Γ_h, con $c(f_s) \geqslant \cdots \geqslant c_(f_2) \geqslant c(f_1)$ y tal que f_s y h son incidentes con vértices de T_{k_s-1}. Por tanto, $c(h) \geq c(f_s)$, con lo que se concluye también que $c(h) \geqslant c(f)$, en contradicción con $c(h) < c(f)$. Así, también en este caso, T_m verifica la condición suficiente expresada en el lema 6.16 y tiene, por tanto, coste mínimo. □

Las figuras 6.12 y 6.13 muestran secuencias de subgrafos obtenidos a partir del grafo G de la figura 6.6, con los costes que se indican, cuando se aplican los algoritmos de Kruskal y Prim respectivamente. Notemos que, tal como se ha dicho en la demostración del teorema 6.15, los subgrafos generados por el algoritmo de Prim son árboles, mientras que, en general, esto no es así en caso de aplicar el algoritmo de Kruskal.

Notas bibliográficas

Los libros de Bondy y Murty [1] y de Chartrand y Lesniak [3] cubren de forma excelente la temática del capítulo. El libro de Harary [4] contiene un apéndice con los diagramas de todos los árboles de orden menor o igual a 10. La aplicación de los árboles a las ciencias de la computación puede estudiarse en el libro de Knuth [5]. El resultado sobre el número de

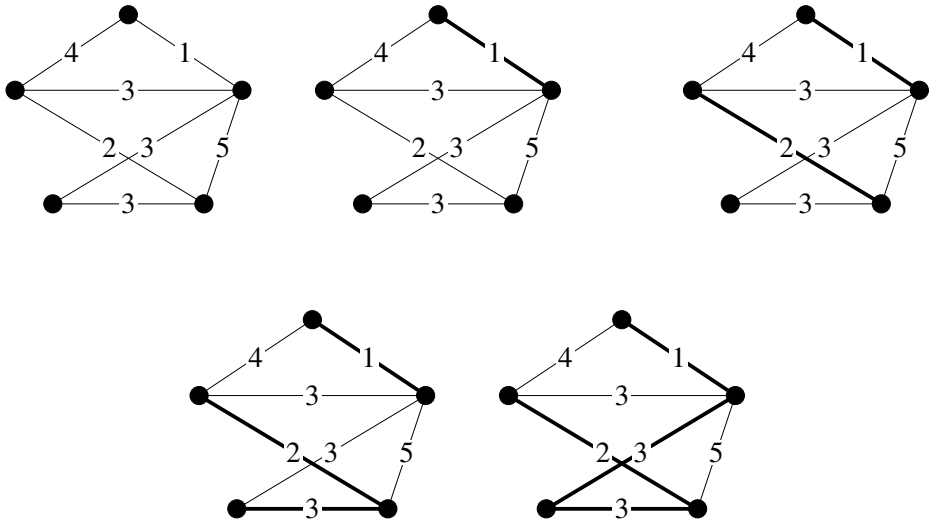


Figura 6.12: Desarrollo del algoritmo de Kruskal

árboles generadores no idénticos se debe a Cayley [2], y en [7] hay una recopilación de diversas demostraciones del teorema. En cuanto a los árboles generadores, el algoritmo de Kruskal fue descrito en [6] y la referencia para el algoritmo de Prim es [8].

Bibliografía

[1] J. A. Bondy, U. S. R. Murty. *Graph Theory with Applications*, North Holland, 1976.

[2] A. Cayley. "A theorem on trees", *Quart. J. Math.*, **23**, pp. 376–378, 1889.

[3] G. Chartrand, L. Lesniak. *Graphs and Digraphs*, Wadsworth & Brooks, 1986.

[4] F. Harary. *Graph Theory*, Addison Wesley, 1972.

[5] D. E. Knuth. *The Art of Computing Programming*, vol. 1, Addison Wesley, 1968.

[6] J. B. Kruskal. "On the shortest spanning tree of a graph and the traveling salesman problem", *Proc. Amer. Math. Soc.*, **7**, pp. 48–50, 1956 .

[7] J. W. Moon. "Various proofs of Cayley's formula for counting trees", *A Seminar on Graph Theory*, Holt, Rinehart and Winston, pp. 70–78, 1967.

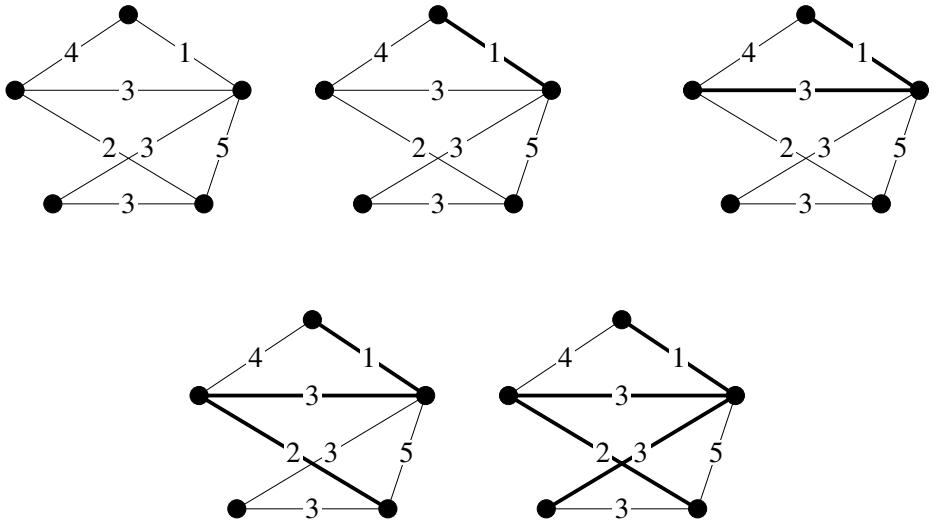


Figura 6.13: Desarrollo del algoritmo de Prim

[8] R. C. Prim. "Shortest connection network and some generalizations", *Bell System Tech. J.*, **36**, pp. 1389–1401, 1957.

Problemas

1. Demostrar que en todo grafo con n vértices y $n-1$ aristas existen al menos dos vértices de grado 1.

2. Demostrar que un árbol es un grafo bipartito.

3. Determinar todos los árboles de orden 4. Cuántos hay que sean no isomorfos?

4. Demostrar el teorema 6.2: Un vértice v de un árbol T es vértice de corte si y sólo si $d(v) > 1$.

5. Sea T un árbol con n vértices y sea G un grafo con grado mínimo $\delta(G) \geqslant n-1$. Demostrar que T es subgrafo de G (es decir, existe un subgrafo de G isomorfo a T).

6. Recordemos que el centro de un grafo G es el conjunto de vértices de G que tienen excentricidad mínima. Demostrar que el centro de un árbol está formado por un único vértice o por dos vértices adyacentes.

7. Sean r y D el radio y el diámetro de un árbol T. Demostrar que D es igual a $2r$ o a $2r-1$. Relacionar este resultado con el del problema 6.

8. Sea T un árbol binario con orden n.

   (a) Demostrar que n es impar;

   (b) demostrar que el número de vértices de grado 1 es $(n+1)/2$.

9. Sea r el vértice raíz de un árbol binario T con orden n. Se dice que $v \in V(T)$ está en el nivel i si $d(r,v) = i$. Sea l_m el nivel máximo. Demostrar que

$$\lceil \log_2 (n+1) - 1 \rceil \leqslant l_m \leqslant \frac{n-1}{2}$$

10. Demostrar que un subgrafo H de un grafo conexo G es subgrafo de un árbol generador de G si y sólo si H es acíclico.

11. Demostrar la fórmula de Cayley (corolario 6.10) aplicando el teorema 6.9 al grafo completo K_n.

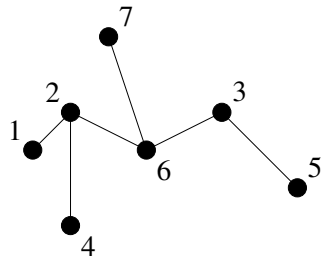


Figura 6.14: Secuencia asociada a T

12. Sea $V = \{1, 2, \ldots, n\}$, $n > 2$, y T un árbol con conjunto de vértices V. Asociemos a T una secuencia $(t_1, t_2, \ldots, t_{n-2})$ de longitud $n-2$ (llamada *secuencia de Prüfer*) construida con elementos de V mediante el procedimiento siguiente. Suponiendo V ordenado, sea v_1 el primer vértice de grado uno en T y sea t_1 el vértice adyacente a v_1. Sea ahora v_2 el primer vértice de grado uno en $T - v_1$ y sea t_2 el vértice adyacente en $T - v_1$ a v_2. Esta operación se repite hasta que t_{n-2} ha quedado definido y sólo quedan dos vértices.

   (a) Dar la secuencia correspondiente al árbol de la figura 6.14.

   (b) Dar el árbol que tiene asociada la secuencia $(4, 4, 3, 1, 1)$.

   (c) Sea T_n el conjunto de árboles de orden n, $n > 2$, con conjunto de vértices V. Demostrar que el procedimiento descrito define una correspondencia biyectiva entre T_n y V^{n-2}.

(d) Utilizando este hecho, demostrar el resultado dado en el corolario 6.10.

13. Una *arborescencia* A con raíz r es un digrafo tal que el grafo resultante de quitar las direcciones de los arcos es un árbol y con la propiedad de que existe un único camino dirigido desde r hacia cualquier otro vértice de A; ver la figura 6.15. Demostrar que el número de arborescencias con n vértices etiquetados es n^{n-1}.

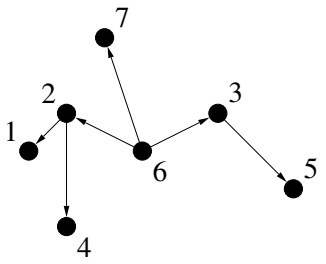


Figura 6.15: Arborescencia con raíz $r = 6$

14. Usando la fórmula recursiva dada en el teorema 6.11, calcular el número de árboles generadores del grafo de la figura 6.10.

15. Sea M la matriz obtenida a partir de la matriz de incidencia reducida B_r, tal como se explica en la sección 6.3. Demostrar que si G es un grafo conexo con n vértices, $n \geqslant 2$, entonces el rango de M es $n-1$.

16. La matriz de grados de un grafo G con conjunto de vértices $\{v_1, v_2, \ldots, v_n\}$ es la matriz diagonal $\mathrm{diag}(d(v_1), d(v_2), \ldots, d(v_n))$. Demostrar la siguiente versión del teorema 6.9: Si G es un grafo conexo no trivial con matriz de adyacencia A y matriz de grados C, entonces el número de árboles generadores de G es el valor de un cofactor cualquiera de la matriz $C - A$.

17. Dado un un grafo conexo G, considerar el algoritmo siguiente:

    (a) Si G no tiene ciclos, entonces $T = G$. Fin.

    (b) De otro modo, sea Γ un ciclo de G y $e \in E(\Gamma)$ tal que $G' = G - e$ sea conexo. Tomar $G = G'$ y volver al paso anterior.

    Demostrar que este procedimiento está bien definido y da un árbol generador de G. ¿Cuántas iteraciones se requieren para obtener el árbol generador?

18. Sea F el conjunto de cuerdas de un grafo conexo G respecto de un cierto árbol generador. Demostrar que si Γ es un ciclo de G, entonces alguna arista de Γ pertenece a F.

19. Sea G un grafo conexo no trivial con una función de coste c definida sobre su conjunto de aristas tal que $c(e) \neq c(f)$ si e y f son aristas distintas. Demostrar que, en este caso, el árbol generador de coste mínimo es único.

20. Considerar un grafo conexo no trivial con una función de coste definida sobre su conjunto de aristas. Suponer que existe una arista e cuyo coste es menor que el de cualquier otra arista. Demostrar que todo árbol generador de coste mínimo contiene la arista e.

21. Sea G un grafo conexo con orden n y tamaño m. Consideremos el grafo T_G que tiene por vértices los árboles generadores de G, y dos árboles son adyacentes en T_G si se obtienen uno del otro por una transformación elemental.

    (a) Demostrar que T_G es conexo;

    (b) demostrar que $D(T_G) \leqslant \min(n-1, m-n+1)$;

    (c) determinar T_G si $G = K_4$.

22. La tabla siguiente da los costes de las conexiones entre diferentes nodos.

	A	B	C	D	E	F
A	–	5	6	4	3	7
B	5	–	2	4	8	5
C	6	2	–	4	8	8
D	4	4	4	–	2	5
E	3	8	8	2	–	4
F	7	5	8	5	4	–

    Determinar una red más económica que permita la comunicación entre dos nodos cualesquiera usando: (a) el algoritmo de Kruskal; (b) el algoritmo de Prim.

Capítulo 7

Circuitos y ciclos

1. Grafos eulerianos
2. Grafos hamiltonianos
3. Ciclos fundamentales
4. Análisis de redes eléctricas

Este capítulo se dedica a analizar la estructura cíclica de los grafos desde dos puntos de vista. Primero, dado un grafo G, se estudia la existencia en G de circuitos y ciclos particularmente interesantes: los circuitos eulerianos y los ciclos hamiltonianos. Veremos que, aunque los problemas de encontrar circuitos y ciclos de este tipo se formulan de manera muy similar, tienen, en cambio, soluciones bien diferentes. Como aplicaciones, se consideran el conocido problema del viajante y el de las secuencias de de Bruijn. En segundo lugar, se ve como todos los ciclos de G se obtienen a partir de los llamados ciclos fundamentales respecto de un árbol generador del grafo. Como aplicación, se presentan las ideas básicas del análisis por ciclos de una red eléctrica *RLC*.

7.1 Grafos eulerianos

El origen de la teoría de grafos se asocia a menudo con la resolución que dio Euler del llamado problema de los puentes de Königsberg (1736). Esta antigua ciudad prusiana, dividida por el río Pregel, que bordea la isla de Kneiphof, tenía siete puentes dispuestos como indica la figura 7.1. Los habitantes de esta ciudad se planteaban la cuestión siguiente: ¿es posible, paseando, hacer un recorrido que pase una única vez por cada uno de los siete puentes? La resolución que dio Euler de este problema no solamente respondía a esta cuestión, sino que introducía la noción de grafo y resolvía al mismo tiempo un problema de carácter más general.

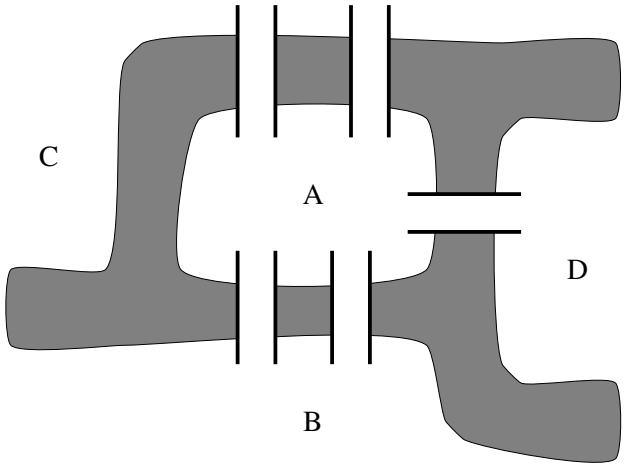


Figura 7.1: Los siete puentes de la ciudad de Königsberg

Dado un grafo (multigrafo) G, se dice que un *circuito* en G es *euleriano* si usa una única vez cada una de sus aristas. En el caso que un circuito así exista, se dice que el *multigrafo* G es *euleriano*. De forma similar, un recorrido que pasa una única vez por cada una de las aristas de G es un *recorrido euleriano*. En la figura 7.2, el grafo (a) no contiene ningún recorrido ni circuito euleriano, el grafo (b) contiene recorridos pero no circuitos eulerianos y el grafo (c) contiene circuitos eulerianos. La solución del problema de los puentes de Königsberg es

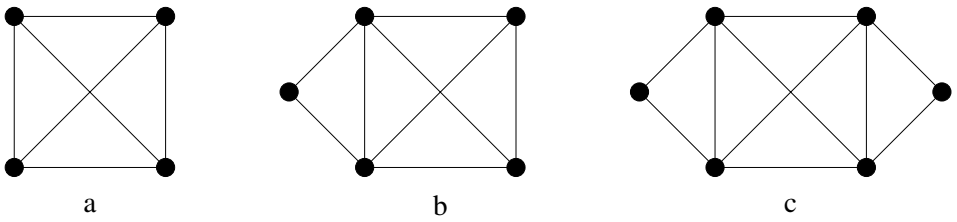


Figura 7.2: Grafos con y sin recorridos o circuitos eulerianos

equivalente a determinar si el multigrafo de la figura 7.3, obtenido asociando a cada región un vértice y a cada puente una arista, contiene o no un recorrido euleriano. Está claro que si un multigrafo es euleriano tiene que ser conexo, salvo vértices aislados. Por otra parte, no es difícil observar que, en un grafo euleriano, todos los vértices tienen que tener grado par. De hecho, estas dos condiciones son también suficientes para la existencia de circuitos eulerianos.

Teorema 7.1. Un multigrafo G es euleriano si y sólo si es conexo (salvo vértices aislados) y

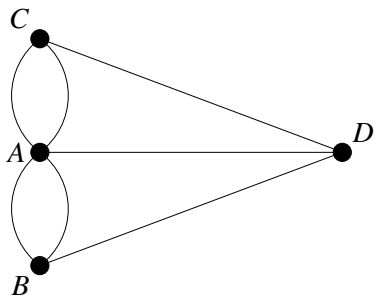


Figura 7.3: El grafo del problema de los puentes de Königsberg

todos sus vértices tienen grado par.

Demostración. Está claro que la existencia de vértices aislados no afecta la propiedad de ser euleriano. Por tanto, podemos suponer que G no tiene vértices aislados.

Necesidad. Si G tiene un circuito euleriano, este circuito conecta todos los vértices y G es conexo. Al recorrer el circuito, cada vez que entremos en un vértice por una determinada arista tenemos que salir por una arista diferente, de manera que cada vértice tiene que tener un número par de aristas incidentes.

Suficiencia. Supongamos ahora que G es conexo y que todos sus vértices tienen grado par. Construimos un recorrido a partir de un vértice arbitrario v_0 sin usar la misma arista dos veces. Si llegamos a un vértice $x \neq v_0$ habremos usado un número impar de aristas incidentes a x, de manera que podemos salir de x por una nueva arista. Cuando esto ya no sea posible, estaremos forzosamente en el vértice de salida v_0 y habremos descrito un circuito C_0.

Si C_0 contiene todas las aristas, ya hemos acabado. En caso contrario, eliminamos de G todas las aristas de C_0, de manera que obtenemos un grafo G' que vuelve a tener todos los vértices de grado par (aunque no tiene por qué ser conexo). Sea H un componente conexo de G' que tenga alguna arista. Como el multigrafo de partida G es conexo, H tiene que tener algún vértice de C_0, digámosle v_1. Construimos un recorrido dentro de H a partir de v_1 de forma similar al que hemos hecho desde v_0 en G y obtenemos un circuito C_1 que no contiene ninguna arista común con C_0. Entonces, $C_0 \cup C_1$ es un circuito con un número de aristas estrictamente más grande que C_0.

Si $C_0 \cup C_1$ contiene todas las aristas, ya hemos acabado. En caso contrario se puede repetir el procedimiento anterior para obtener una sucesión de circuitos con número de aristas estrictamente creciente. Como el número de aristas de G es finito, con este procedimiento acabamos construyendo un circuito euleriano. □

Esta demostración proporciona, entonces, un método constructivo para obtener un circuito euleriano, una vez sabemos que las condiciones del teorema se satisfacen. De hecho se puede usar el razonamiento para obtener un algoritmo que construye un circuito euleriano, cuando es posible.

Este algoritmo, claramente, necesita un mecanismo generador de recorridos. Así, comenzaremos dando un procedimiento *RecorrerCamino* que tiene como objetivo ampliar un recorrido dado. En este procedimiento V y E son el conjunto de vértices y de aristas del grafo, v es un vértice cualquiera de V y $\Gamma(w)$ es el conjunto de aristas incidentes con el vértice w. El procedimiento simplemente añade aristas nuevas, mientras sea posible, a un recorrido dado.

Entrada: $G = (V, E)$: un grafo; $v \in V$.
Procedimiento RECORRERCAMINO

1. $P \leftarrow \emptyset$ [P es la secuencia de aristas del recorrido]
2. $U \leftarrow E$ [U es el conjunto de aristas que aún no están en el recorrido]
3. $w \leftarrow v$ [Inicializamos w, final actual del recorrido]
4. **Mientras** $\Gamma(w) \cap U \neq \emptyset$ **hacer**
Tomar $e \in \Gamma(w) \cap U$
$x \leftarrow$ otro extremo de e
$U \leftarrow U - e$
Añadir e a P
$w \leftarrow x$

Salida: P: camino.

En cada iteración, *RecorrerCamino* toma una arista arbitraria, por la cual aún no se ha pasado, de todas las que inciden con el vértice actual y la añade al recorrido P. El procedimiento continúa hasta que no hay aristas libres a partir del vértice actual.

El algoritmo *Euler* usa este procedimiento para construir un circuito euleriano en un grafo conexo con todos los vértices de grado par. Observemos la recursividad del algoritmo, que proviene de la demostración en la cual se basa.

Entrada: $G = (V, E)$: un grafo.
Algoritmo EULER
Procedimiento $Euler(V', E', v')$
RecorrerCamino(V', E', v') [Encuentra un circuito P en $G(V', E')$]
Si P no es un circuito euleriano de $G(V', E')$ **entonces hacer**

Borrar de E las aristas que hay en P
Denotar los componentes no triviales del subgrafo
que resulta $G_1(V_1,E_1), G_2(V_2,E_2), \ldots, G_j(V_j,E_j)$
sea v_i un vértice de P en V_i
Para $i = 1$ **hasta** j **hacer**
$Euler(V_i, E_i, v_i)$ [Encuentra un circuito euleriano C en G_i]
Pone C en P en la posición v_i
$C \leftarrow P$
Retorna C

1. $v \leftarrow$ cualquier vértice de V.
2. $Euler(V, E, v)$.

Salida: C [Circuito Euleriano de $G = (V,E)$].

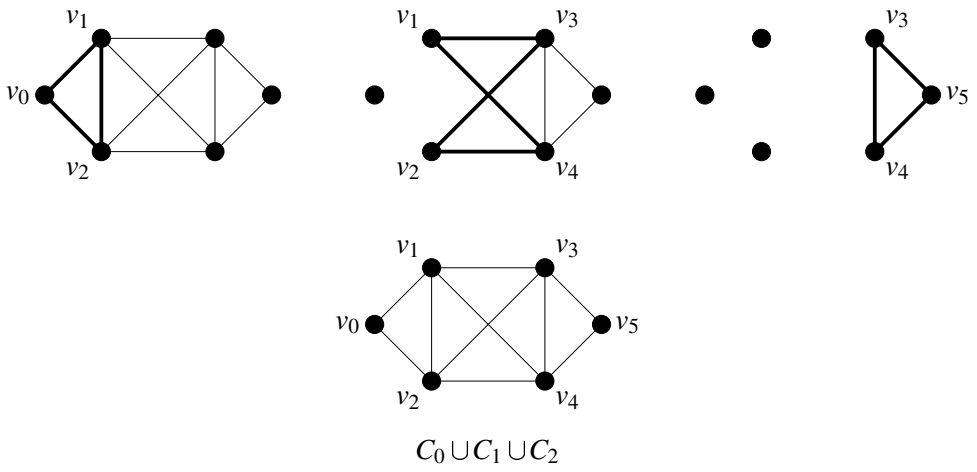


Figura 7.4: Construcción de un circuito euleriano

Usamos el grafo de la figura 7.2 para mostrar el funcionamiento del algoritmo tal como se muestra en la figura 7.4.

Cuando se llama al procedimiento Euler, este llama a *RecorrerCamino*, el cual inicia un recorrido en $v = v_0$ y produce un circuito. En el grafo del ejemplo suponemos que se obtiene el circuito $C_0 = v_0, v_1, v_2, v_0$. Resulta $P = C_0$. Como aún no tenemos el circuito euleriano construido, borramos las aristas correspondientes al circuito encontrado C_0 y repetimos el proceso en el resto del grafo comenzando, por ejemplo, en v_1. Supongamos que encontra-

mos ahora el circuito $C_1 = v_1, v_4, v_2, v_3, v_1$. Entonces, P será $v_0, v_1, v_4, v_2, v_3, v_1, v_2, v_0$. Finalmente, comenzando en v_4, encontramos $C_2 = v_4, v_3, v_5, v_4$. El circuito euleriano resulta ser $v_0, v_1, v_4, v_3, v_5, v_4, v_2, v_3, v_1, v_2, v_0$.

El teorema 7.1 permite también resolver la cuestión relativa a la existencia de recorridos eulerianos. Esto se debe al hecho que un multigrafo G admite un recorrido euleriano con vértices terminales u y v si y sólo si $G+uv$ admite un circuito euleriano. Con esta observación es fácil demostrar el resultado siguiente.

Teorema 7.2. Un multigrafo G contiene un recorrido euleriano si y sólo si es conexo (salvo vértices aislados) y el número de vértices de grado impar es 0 o 2.

Cabe observar que, en particular, los extremos de cualquier recorrido euleriano han de ser los vértices de grado impar.

Este resultado da, entonces, una respuesta negativa al problema de los puentes de Königsberg, ya que el grafo de la figura 7.3 no satisface las condiciones necesarias.

Una propiedad característica de los grafos eulerianos es la de admitir siempre alguna descomposición (no necesariamente única) en ciclos. Para hacer precisa esta idea, recordemos que, dados dos grafos $G_1 = (V_1, E_1)$ y $G_2 = (V_2, E_2)$, su unión es $G_1 \cup G_2 = (V_1 \cup V_2, E_1 \cup E_2)$ y su intersección es $G_1 \cap G_2 = (V_1 \cap V_2, E_1 \cap E_2)$. Recordemos también que $G_1 \oplus G_2 = (V_1 \cup V_2, E_1 \triangle E_2)$, donde $E_1 \triangle E_2$ son todas las aristas que pertenecen a E_1 o a E_2. pero no a los dos. Finalmente, se dice que $G = (V, E)$ se *descompone* en los subgrafos G_1 y G_2 cuando $G_1 \cup G_2 = G$ y $G_1 \cap G_2 = N$, donde N es un grafo nulo. En otras palabras, se dice que G se descompone en los subgrafos G_1 y G_2 si $G = G_1 \oplus G_2 = G_1 \cup G_2$. La descomposición en un número cualquiera de subgrafos se define de forma similar.

Teorema 7.3. Un multigrafo conexo G es euleriano si y sólo si admite una descomposición en ciclos.

Demostración. Suficiencia: Sean $C_1, C_2, \ldots, C_n$ ciclos disyuntos en aristas tales que $G = C_1 \oplus C_2 \oplus \cdots \oplus C_n$. Como el grado de cada vértice en cada ciclo es dos, deducimos que cada vértice de G tiene grado par.

Necesidad: Supongamos que G es euleriano, y sea C un circuito euleriano de G. Si en C no hay vértices repetidos, G es ya un ciclo. Si no, sea v el primer vértice que se repite al recorrer C desde un vértice inicial v_0 y llamamos C_1 al ciclo que se describe entre las dos primeras apariciones de v. Entonces, el grafo $G' = G \setminus E(C_1)$ continúa teniendo todos los vértices de grado par y es conexo salvo vértices aislados, y por tanto es euleriano. Iterando este procedimiento obtenemos una secuencia de ciclos disyuntos en aristas cuya unión es G. □

Esta descomposición no necesariamente es única, ya que puede haber más de un circuito euleriano y cada uno de ellos se puede recorrer desde cada uno de sus vértices.

En la figura 7.5 se ilustra la descomposición de un grafo euleriano siguiendo el procedimiento que proporciona la demostración del teorema anterior.

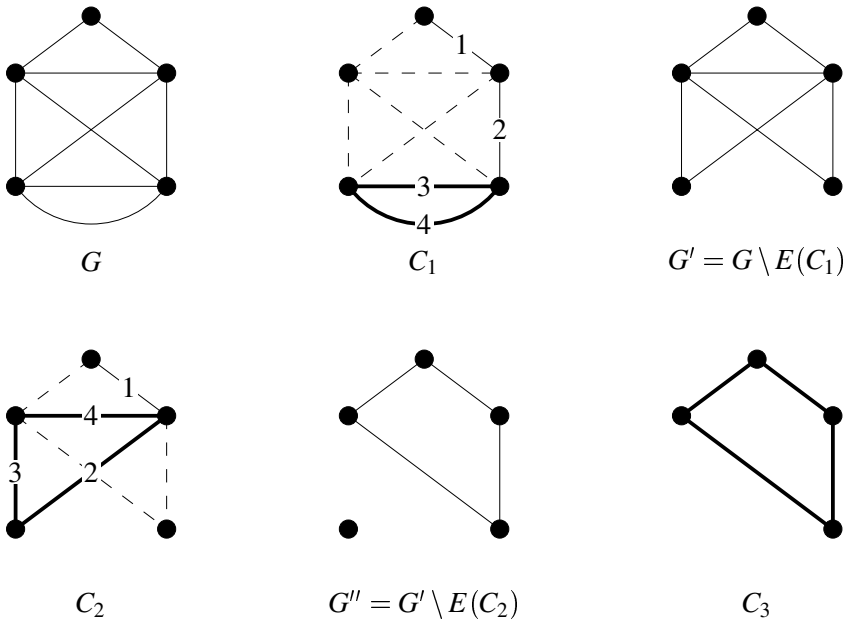


Figura 7.5: Descomposición del grafo G en los ciclos C_1, C_2 y C_3

La propiedad de ser euleriano se puede extender también al caso de digrafos. Un *recorrido* en un multidigrafo G es *euleriano* si pasa por todos sus arcos una única vez. Si el recorrido es cerrado diremos que se trata de un *circuito euleriano dirigido* y diremos también que G es *euleriano*.

La caracterización de digrafos eulerianos se recoge en los resultados siguientes, similares a los que se incluyen en los teoremas 7.1 y 7.2.

Como en el caso no dirigido, si un multidigrafo es euleriano tiene que ser fuertemente conexo (excepto si tiene vértices aislados). Además, al recorrer el circuito, cada vez que incidimos en un vértice x por una arista, es preciso salir por otra arista incidente desde este mismo vértice y, por tanto, $d^+(x) = d^-(x)$ para todo $x \in V(G)$. Estas dos condiciones necesarias son también suficientes. La demostración de la suficiencia es análoga a la del teorema 7.1 y se propone como ejercicio.

Teorema 7.4. Un multidigrafo G es euleriano si y sólo si es fuertemente conexo (salvo vértices aislados) y para todo $x \in V(G)$, $d^+(x) = d^-(x)$.

La existencia de recorridos eulerianos en multidigrafos viene caracterizada en el teorema siguiente, cuya demostración se propone también como ejercicio.

Teorema 7.5. Un multidigrafo G contiene un recorrido euleriano desde el vértice u hasta el vértice v, no adyacente hacia u, si y sólo si el multigrafo subyacente es conexo (salvo vértices aislados), y para todo $x \in V(G) \setminus \{u,v\}$, $d^+(x) = d^-(x)$, mientras que $d^+(u) = d^-(u)+1$ y $d^-(v) = d^+(v)+1$.

Secuencias de de Bruijn

Una aplicación interesante de los circuitos eulerianos es la obtención de las llamadas secuencias de de Bruijn. Estas secuencias aparecen en el estudio de los registros cíclicos de desplazamiento, tema este que tiene una amplia aplicación técnica en las telecomunicaciones, la teoría de códigos, la criptografía y las ciencias de la computación.

Dado un alfabeto S con d símbolos podemos formar d^k palabras $x = x_0x_1\ldots x_{k-1}$, $x_i \in S$, de longitud k. Una *secuencia de de Bruijn* es una secuencia circular $x_0, x_1, \ldots, x_{n-1}$, $x_i \in S$, de longitud n, con la propiedad siguiente: para cada palabra x de longitud k, existe un único i tal que $x = x_ix_{i+1}\ldots x_{i+k-1}$, donde los subíndices se toman módulo n. Claramente, $n \geqslant d^k$. Veremos a continuación que para todo d y k existen soluciones con $n = d^k$.

Para demostrar este resultado se introduce la familia de los digrafos de de Bruijn. Esta familia de grafos dirigidos es también interesante en el diseño de grandes redes de interconexión, es decir, redes con muchos nodos de tal forma que cada nodo está conectado como mucho a un número dado de nodos vecinos y la distancia máxima entre nodos también está acotada. El lector interesado en este tema puede consultar por ejemplo [3].

El *digrafo de de Bruijn* $B(d,k-1) = (V,A)$, $d > 1$, $k > 1$ tiene por conjunto de vértices V el de las d^{k-1} palabras de longitud $k-1$, siendo el vértice correspondiente a la palabra $x = x_0, x_1, \ldots, x_{k-2}$ adyacente hacia los vértices de la forma $y = x_1, x_2, \ldots, x_{k-1}$. Cabe notar que cada arco $a = (x,y) \in A$ queda identificado con una palabra $x_0, x_1, \ldots, x_{k-1}$ de longitud k y que $B(d,k-1)$ es un multidigrafo, ya que contiene lazos. Los digrafos $B(2,2)$ y $B(2,3)$ se representan en la figura 7.6.

Proposición 7.6. $B(d,k-1)$ es un digrafo regular con grado d y tiene d^k arcos.

Demostración. De acuerdo con la regla de adyacencia que define los arcos, cada vértice $x_0, x_1, \ldots, x_{k-2}$ es adyacente hacia los d vértices diferentes $x_1, \ldots, x_{k-2}, s$, $s \in S$. Por otra parte, $x_0, x_1, \ldots, x_{k-2}$ es adyacente desde los d vértices $s, x_0, x_1, \ldots, x_{k-3}$, $s \in S$. Esto demuestra que el grado de entrada y el grado de salida de cada vértice es d y, por tanto, $B(d,k-1)$ es d–regular. Así, $|A| = d|V| = d^k$. □

De acuerdo con la condición necesaria y suficiente dada en la sección anterior, para que un digrafo sea euleriano, la proposición anterior permite formular el resultado siguiente:

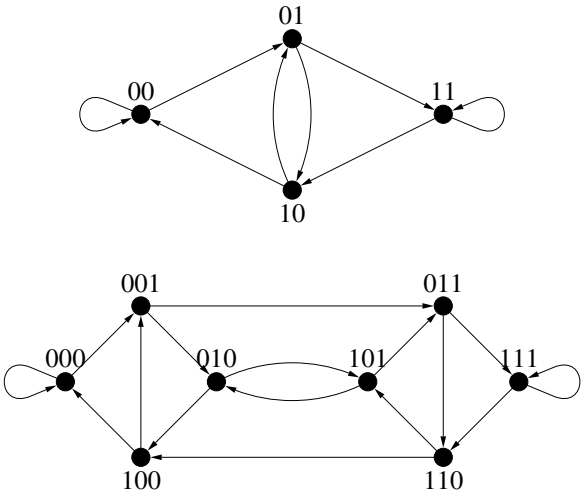


Figura 7.6: Digrafos $B(2,2)$ y $B(2,3)$

Teorema 7.7. Para todo $d > 1$ y $k > 1$, el digrafo de de Bruijn $B(d,k-1)$ es euleriano.

Consideremos ahora un circuito euleriano C en $B(d,k-1)$. Como cada arco de $B(d,k-1)$ corresponde a una palabra de longitud k y dos arcos consecutivos del circuito considerado son de la forma $a_0 = x_0x_1\ldots x_{k-1}$ y $a_1 = x_1x_2\ldots x_k$, es evidente que C da una secuencia de de Bruijn de longitud mínima d^k.

Corolario 7.8. Para todo $d > 1$ y $k > 1$ existen secuencias de de Bruijn de longitud d^k.

Ejemplo: El circuito euleriano

$$\begin{aligned} 000 &\to 001 \to 011 \to 110 \to 101 \to 011 \to 111 \to 111 \to 110 \\ &\to 100 \to 001 \to 010 \to 101 \to 010 \to 100 \to 000 \to 000 \end{aligned}$$

en $B(2,3)$ proporciona la secuencia de de Bruijn 0001101111001010 de longitud 16 para palabras binarias de longitud 4.

7.2 Ciclos hamiltonianos

El problema de buscar recorridos cerrados que pasen por todas las aristas de un grafo una única vez se puede modificar ligeramente para considerar caminos cerrados que pasen por todos los vértices una única vez. El célebre matemático irlandés Sir William Rowan Hamilton propuso

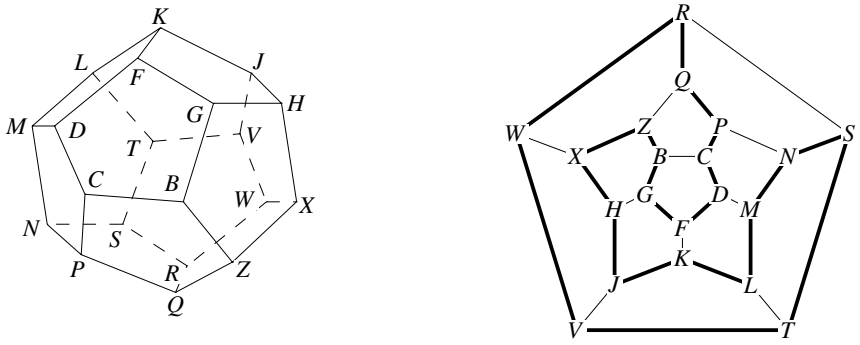


Figura 7.7: La vuelta al mundo de Hamilton y el ciclo en el grafo

en 1886 el problema de encontrar un itinerario para recorrer veinte ciudades alrededor del mundo puestas en los vértices de un dodecaedro de manera que se pase una única vez por cada ciudad y se vuelva a la de salida. Esto corresponde a encontrar un ciclo cerrado que pase una única vez por todos los vértices del grafo del dodecaedro (ver la figura 7.7).

Por ello se llama *ciclo hamiltoniano* en un grafo a un camino cerrado que pasa una única vez por cada vértice (estrictamente hablando, dos veces por el vértice inicial que es también el final). Un grafo es *hamiltoniano* si contiene un ciclo hamiltoniano. De forma similar, un camino *hamiltoniano* es un camino que pasa una única vez por cada vértice.

A pesar de ser nociones próximas, la propiedad de ser hamiltoniano y de ser euleriano son independientes. En la figura 7.8 se puede ver el grafo más pequeño que es (a) euleriano y hamiltoniano a la vez, (b) euleriano pero no hamiltoniano, (c) no euleriano pero hamiltoniano y (d) ni hamiltoniano ni euleriano.

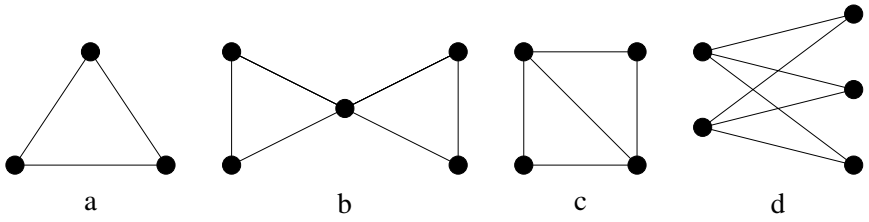


Figura 7.8: Grafo (a) euleriano y hamiltoniano; (b) euleriano y no hamiltoniano; (c) hamiltoniano y no euleriano; (d) ni hamiltoniano ni euleriano

Otra diferencia importante entre estas dos nociones es que, así como saber si un grafo es o no euleriano se responde con un teorema sencillo y definitivo en términos de la estructura del grafo, el problema de saber si es hamiltoniano resulta mucho más difícil: de hecho este es uno de los grandes problemas abiertos en la teoría de grafos y hasta ahora no se conoce ningún resultado que dé condiciones necesarias y suficientes para responder a esta cuestión. Los tipos de resultados que se conocen son, o bien resultados generales que dan condiciones suficientes, o bien resultados específicos relativos a familias particulares de grafos. A continuación exponemos ejemplos de los dos tipos. En toda esta sección supondremos que G es un grafo conexo.

Intuitivamente está claro que si el grafo tiene un número suficiente de aristas es más fácil poder recorrer un ciclo hamiltoniano. Hay una gran cantidad de resultados que hacen precisa esta intuición. Muchos de ellos se basan en la observación siguiente:

Lema 7.9. Sea G un grafo de n vértices y u, v dos vértices no adyacentes tales que $d(u) + d(v) \geqslant n$. Entonces, G es hamiltoniano si y sólo si $G + uv$ es hamiltoniano.

Demostración. Está claro que, si G es hamiltoniano, $G + uv$ también lo es.

Supongamos que $G + uv$ es hamiltoniano y sea H un ciclo hamiltoniano de este grafo. Si no contiene la arista uv, H también es un ciclo hamiltoniano en G. En caso contrario, $H' = H - uv$ es un camino hamiltoniano en G, y sea $u = w_1 w_2 \dots w_{n-1} w_n = v$. Llamamos U a los vértices tales que su sucesor en el camino es adyacente a u, $U = \{w_k : u w_{k+1} \in E(G)\}$ y V a los vértices adyacentes a v, $V = \{w_k : v w_k \in E(G)\}$. Tenemos $|U \cup V| < n$, ya que v no pertenece a ninguno de los dos conjuntos, de manera que

$$n > |U \cup V| = |U| + |V| - |U \cap V| = d(u) + d(v) - |U \cap V| \geqslant n - |U \cap V|$$

o sea, que $U \cap V$ no es vacío. Si $w_k \in U \cap V$, podemos formar el ciclo hamiltoniano

$$u = w_1, w_2, \dots, w_k, v, w_{n-1}, \dots, w_{k+1}, u$$

tal como se indica en la figura 7.9. □

Con este lema no es difícil demostrar uno de los resultados clásicos sobre esta cuestión.

Teorema 7.10 (Ore, 1961). Si para cada par de vértices no adyacentes u, v de un grafo G de n vértices, $n \geqslant 3$, se satisface $d(u) + d(v) \geqslant n$, entonces G es hamiltoniano.

Demostración. De acuerdo con el lema anterior y las hipótesis del teorema, si u y v son dos vértices no adyacentes, G es hamiltoniano si y sólo si lo es también $G + uv$. En realidad podemos añadir a G todas las aristas que unen vértices no adyacentes sin alterar la propiedad que el grafo sea hamiltoniano. Así, entonces, G es hamiltoniano si y sólo si el grafo completo

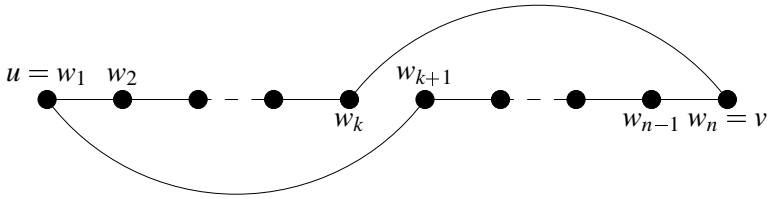


Figura 7.9:

K_n que se obtiene de esta manera lo es también. Está claro que en K_n podemos describir un ciclo hamiltoniano simplemente visitando cualquier vértice no recorrido hasta que esto no sea posible y volviendo entonces al vértice inicial. □

En particular, se obtiene también este otro resultado clásico:

Teorema 7.11 (Dirac, 1952). Un grafo G de n vértices, $n \geqslant 3$, tal que su grado mínimo satisface $\delta(x) \geqslant n/2$, es hamiltoniano.

En cuanto a resultados relativos a familias de grafos, resulta obvio por ejemplo que los ciclos C_n son hamiltonianos. Esto es también cierto para los grafos completos K_n, tal como hemos mencionado en la demostración del teorema anterior. En cambio no lo es para los bipartitos completos $K_{n,m}$ con $n \neq m$. En el ejercicio 15 se dan indicaciones para estudiar el problema en los grafos bipartitos. Quizá el tipo de resultados más fuertes relativos a familias especiales de grafos sea el que hace referencia a los grafos planares. Tal como se ha visto en el capítulo 5, un grafo es planar cuando se puede dibujar en el plano de manera que sus aristas no se corten. Por otra parte, un grafo es k–conexo si la supresión de menos de k vértices no lo desconecta (capítulo 8). La caracterización de grafos planares hamiltonianos está históricamente relacionada con el célebre teorema de los cuatro colores, que dice que cualquier mapa dibujado sobre un papel admite una coloración de países, de manera que dos países adyacentes tengan colores diferentes, usando sólo cuatro colores. No se habría tenido que esperar 150 años para ver este teorema demostrado si se hubiese podido demostrar que cualquier grafo planar 3-regular y 3-conexo admite un ciclo hamiltoniano (cosa que no es cierta). Lo que sí se puede asegurar es el resultado siguiente, cuya demostración se puede encontrar por ejemplo en [2].

Teorema 7.12 (Tutte, 1956). Un grafo planar 4-conexo es hamiltoniano.

Como es habitual, las nociones anteriores se generalizan fácilmente al caso de grafos dirigidos. Intuitivamente parece que será aún más difícil encontrar caminos dirigidos hamiltonianos, ya que la orientación es una restricción añadida. Sin embargo, el problema de saber si un digrafo D es hamiltoniano se puede reducir al caso no dirigido con la construcción siguiente.

Construimos un grafo G que tiene, por cada vértice u de D, un camino de longitud 3, $x_u y_u z_u$. Si u es adyacente hacia v en el digrafo D, entonces x_u es adyacente a z_v en el grafo G. Es fácil comprobar que D es hamiltoniano si y sólo si G lo es. En la figura 7.10 hay dibujados un digrafo D y el grafo G construido a partir de D de esta manera.

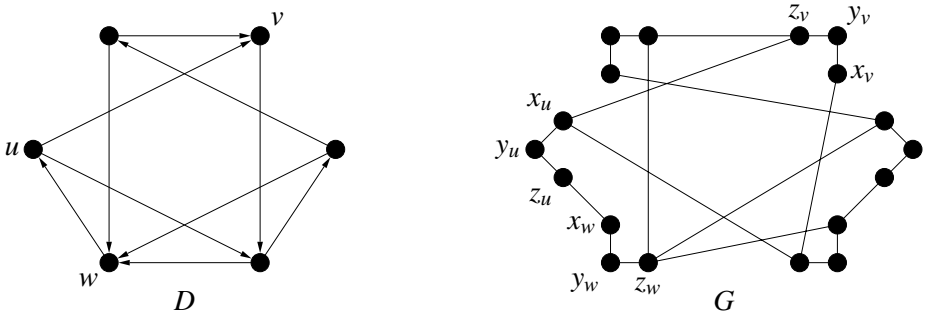


Figura 7.10: El digrafo D es hamiltoniano si y sólo si G lo es

Además, se pueden encontrar en el caso dirigido resultados similares a los del caso no dirigido, aunque las técnicas de demostración no siempre son similares. Por ejemplo, el teorema equivalente al de Ore en el caso dirigido es el teorema de Meyniel, cuya demostración se puede ver por ejemplo en [2].

Teorema 7.13. Sea D un digrafo fuertemente conexo de orden n tal que para cada par de vértices u y v no adyacentes (ni uv ni vu son arcos del digrafo) se satisface $d(u)+d(v) \geqslant 2n-1$, donde $d(x) = d^+(x) + d^-(x)$. Entonces D es hamiltoniano.

Una aplicación inmediata de este resultado es que los digrafos fuertemente conexos que se obtienen orientando las aristas de un grafo completo son hamiltonianos (estos grafos se llaman *torneos*, y han sido extensamente estudiados).

Algoritmos para encontrar ciclos hamiltonianos

Determinar si un grafo $G(V,A)$ es hamiltoniano es un problema NP-completo. Los algoritmos exactos que dan un circuito hamiltoniano solamente son aplicables cuando $|V|$ es pequeña, o bien sobre determinados tipos de grafos. En general se aplican algoritmos aproximados que no pueden garantizar, caso de no encontrar un ciclo hamiltoniano, que este no exista.

Para grafos de Ore, es decir, que satisfacen las condiciones del teorema 7.10, existen algoritmos que encuentran un ciclo hamiltoniano. El más conocido es de orden $O(|V|^3)$ y fue propuesto por Bondy y Chvátal (1974). Un algoritmo de orden $O(|V|^2)$ ha sido propuesto recientemente [9].

Entrada: $G = (V, E)$: un grafo de Ore.
Algoritmo ALBERTSON

1. Crear un camino maximal $P : u, \ldots, x_k, \ldots, v$.
2. **Mientras** $|V(C)| \neq |V(G)|$ **hacer**
  Si u es adyacente a v **entonces hacer** $C : u, \ldots, v, u$
    sino encontrar k tal que u sea adyacente a $x_k + 1$ y v a x_k
3. $C \leftarrow u, x_{k+1}, x_{k+2}, ..., v, x_k, x_{k-1}, ..., x, u$
4. Encontrar $x \in V(G - C)$ y crear P^*, un camino que contenga x y todo C.
5. Hacer P un camino maximal que contenga P^*

Salida: C [Ciclo Hamiltoniano de $G = (V, E)$].

El problema del viajante

Una variante al problema de encontrar un ciclo hamiltoniano en un grafo es el conocido problema del viajante. En este problema, un viajante tiene que visitar un conjunto de ciudades de forma que pase solamente una vez per cada una y que el trayecto total realizado sea mínimo. El modelo será entonces un grafo con pesos en las aristas que representen las distancias. Entonces se trata de encontrar un ciclo hamiltoniano tal que la suma de los pesos de las aristas sea mínima. El problema es también NP-completo. Por ello, los esfuerzos se han concentrado en dar algoritmos aproximados y métodos heurísticos. Los métodos heurísticos podrían ser clasificados entre aquellos que construyen una posible solución y los que intentan mejorar sistemáticamente una solución inicial. Correspondiente al primer tipo hay un método heurístico sencillo conocido como del *vecino más próximo*. Se comienza en un vértice y se añade la arista de distancia mínima. A continuación se van poniendo aristas de distancia mínima en cada extremo del camino. Este método no es muy eficaz, ya que puede dejar de considerar aristas cortas, y cerrar el camino para hacer el ciclo suele ser muy costoso.

Un manera de reducir los problemas que aparecen con este algoritmo consiste en comenzar con un ciclo corto y expandirlo insertando vértices de manera que, en cada paso, el peso total de las aristas del ciclo aumente el mínimo posible. Esta técnica es conocida como *algoritmo de inserción mínima*. También podemos encontrarnos con dificultades provinientes, esencialmente, del hecho que un grafo arbitrario con pesos no tiene por qué verificar la desigualdad triangular respecto de los pesos. Si ésta se verifica podemos dar el algoritmo siguiente:

Entrada: $G = (V, E)$: un grafo con pesos que satisface la desigualdad triangular.
Algoritmo INSERCIÓN MÍNIMA

1. Seleccionar un vértice cualquiera y considerarlo un ciclo C_1 de G.
2. $i \leftarrow 1$
3. **Si** $i = |V|$ **entonces** $C = C_{|V|}$. Ir a la salida.
sino si C_i ha sido escogido, $1 \leqslant i \leqslant |V|$, entonces dar un vértice
v_i que no esté en C_i próximo a dos vértices consecutivos w_i y w_{i+1} de C_i.
4. C_{i+1} se forma insertando v_i entre w_i y w_{i+1}

Salida: C [Ciclo Hamiltoniano que aproxima el ciclo del viajante].

Entre el segundo tipo de métodos hay los conocidos métodos de intercambio de aristas de Lin (1965) y Lin y Kernighan (1973). Estos métodos se basan en modificar un ciclo inicial intercambiando aristas (dos o más). Ver la figura 7.11.

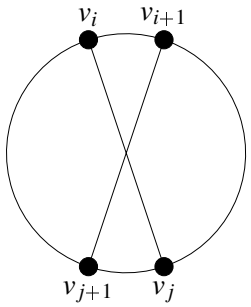


Figura 7.11:

Damos a continuación un algoritmo basado en el intercambio de dos aristas.

Entrada: $G = (V, E)$: un grafo con pesos.
Algoritmo LIN Y KERNIGHAN
1. Considerar un ciclo hamiltoniano inicial $C : v_1, v_2, \ldots, v_n, v_1$ de G.
2. **Para** i, j tales que $1 < i+1 < j < |V|$
Obtener un nuevo ciclo
$C_{ij} = v_1, v_2, \ldots, v_i, v_j, v_{j-1}, \ldots, v_{i+1}, v_{j+1}, v_{j+2}, \ldots, v_{|V|}, v_1$
Si $w(v_i, v_j) + w(v_{i+1}, v_{j+1}) < w(v_i, v_{i+1}) + w(v_j, v_{j+1})$ **entonces** $C \leftarrow C_{ij}$

Salida: C [Ciclo Hamiltoniano que aproxima el ciclo del viajante].

Más recientemente, y también basados en el intercambio de aristas, se han dado algoritmos que consiguen buenos resultados para problemas con un número elevado de ciudades. Una familia de algoritmos se basa en el recocido simulado (*simulated annealing*). El método proviene de la analogía entre los estados de un sistema físico, por ejemplo un líquido, y los estados que puede tomar el problema de optimización. Esencialmente, el algoritmo es el mismo que antes, sólo que ahora los cambios se aceptan con una cierta probabilidad $e^{-\Delta E/(KT)}$. De esta manera, el algoritmo admite cambios que empeoren el ciclo encontrado y así es posible evitar mínimos locales. El parámetro T se asocia con la temperatura del estado físico que modela (a más temperatura, más desorden) y E con la energía (en el modelo físico se trata de bajar la temperatura para conseguir un estado de energía mínima) [13], [1]. Otra familia de algoritmos que se ha aplicado con éxito al problema del viajante, la familia de los *algoritmos genéticos*, se basa en la mecánica de la selección natural y la genética [10]. Posibles soluciones al problema, generadas aleatoriamente, forman una *población*. En cada iteración se genera una nueva población reproduciendo y cruzando entre sí las soluciones de la generación anterior seleccionadas probabilísticamente de acuerdo con su coste. Un mecanismo de *mutación* permite reducir las posibilidades de encontrar mínimos locales.

7.3 Ciclos fundamentales

Dado un grafo G simple y conexo con n vértices y m aristas, sea $\mathcal{C}$ el conjunto de todos los subgrafos de G que son descomponibles en ciclos o, de manera equivalente, que son ciclos o suma $\oplus$ de ciclos disyuntos en aristas. Estudiaremos en esta sección cómo se puede distinguir en G una colección de $m-n+1$ ciclos, llamados *fundamentales*, a partir de los cuales se puede expresar cualquier subgrafo $C \in \mathcal{C}$. En la sección siguiente se verá una aplicación importante de este resultado al análisis de redes eléctricas.

Subespacio de ciclos

En primer lugar demostremos que el conjunto $\mathcal{C}$ es estable por la operación $\oplus$.

Teorema 7.14. Si $C_1, C_2 \in \mathcal{C}$, entonces $C_1 \oplus C_2 \in \mathcal{C}$.

Demostración. Si C_1 y C_2 no tienen vértices en común, el resultado es trivial. De otro modo, sea v un vértice común a C_1 y C_2 y supongamos que ninguna de las aristas incidentes con v es común a C_1 y C_2. Como v tiene en C_1 y C_2 grado par, también tendrá grado par en $C = C_1 \oplus C_2$ (ver la figura 7.12a). Por otra parte, si la arista e incidente con v es común a C_1 y C_2, e no aparece en $C_1 \oplus C_2$ y, por tanto, el grado resultante de v en C continuará siendo par (figura 7.12b). En cualquiera de los casos, cada vértice $v \in V(C)$ tiene en C grado par y, por

tanto, el grafo C es euleriano. De acuerdo con el teorema 7.3, C se puede descomponer en ciclos disyuntos en aristas, es decir, $C \in \mathcal{C}$. □

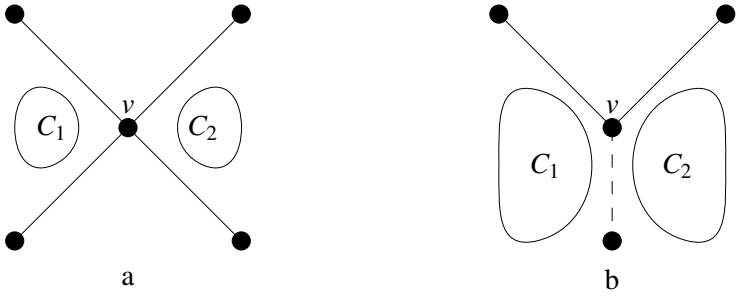


Figura 7.12:

Sea $E(G) = \{e_1, e_2, \ldots, e_m\}$. Podemos asociar a G un espacio vectorial finito $E = E(G)$ de dimensión m de la forma siguiente: cada vector $\mathbf{h} = (h_1, h_2, \ldots, h_m) \in E$, con $h_i = 0, 1$, $1 \leqslant i \leqslant m$, representa el subgrafo H de G inducido por las aristas $e_i \in E(G)$ tales que $h_i = 1$. Dados $\mathbf{h}, \mathbf{g} \in E$, $\mathbf{h} \oplus \mathbf{g}$ es el vector $(h_1 \oplus g_1, h_2 \oplus g_2, \ldots, h_m \oplus g_m)$, donde $\oplus$ es la suma módulo 2. Análogamente, si $\alpha = 0, 1$ y $\mathbf{h} \in E$, $\alpha h = (\alpha h_1, \alpha h_2, \ldots, \alpha h_m)$, donde αh_i se calcula multiplicando módulo 2. Notemos que la operación interna $\oplus$ definida en E es compatible con la suma $\oplus$ de subgrafos de G, es decir, si el vector $\mathbf{h}$ corresponde al subgrafo H y el vector $\mathbf{k}$ al subgrafo K, entonces $\mathbf{h} \oplus \mathbf{k}$ corresponde a $H \oplus K$ (no es preciso considerar los posibles vértices aislados). Teniendo presente esta identificación, usaremos indistintamente las notaciones H o $\mathbf{h}$ para referirnos a un determinado subgrafo o al vector que lo representa.

Ejemplo: En el grafo de la figura 7.13, los subgrafos H y K vienen representados por los vectores $h = (0, 1, 1, 1, 1, 1, 0, 0)$ y $k = (1, 0, 0, 1, 0, 1, 1, 1)$ respectivamente. Así, el vector que corresponde a $H \oplus K$ es $h \oplus k = (1, 1, 1, 0, 1, 0, 1, 1)$.

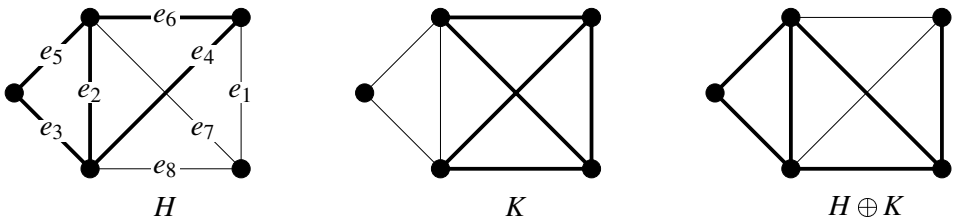


Figura 7.13: Suma $\oplus$ de subgrafos

En particular, el teorema 7.14 nos dice que C corresponde a un subespacio vectorial de E, que podemos llamar *subespacio de ciclos*.

Se puede formular una caracterización interesante del subespacio C de los ciclos a partir de la matriz de incidencia $\mathbf{B}$ del grafo G. Esta caracterización se utilizará en la sección siguiente, donde se estudiará la aplicación de la teoría de los ciclos fundamentales al análisis de redes eléctricas. Notemos que cada fila de $\mathbf{B}$ es un vector $\mathbf{b} \in E$.

Dados $\mathbf{h}, \mathbf{k} \in E$ definimos $\mathbf{h} \cdot \mathbf{k} \equiv \sum_{\oplus} h_i k_i = h_1 k_1 \oplus h_2 k_2 \oplus \cdots \oplus h_m k_m$ y diremos que $\mathbf{h}$ y $\mathbf{k}$ son *ortogonales* si $\mathbf{h} \cdot \mathbf{k} = \mathbf{0}$. De acuerdo con esta definición, dos vectores son ortogonales cuando corresponden a subgrafos disyuntos en aristas o con un número par de aristas comunes.

Teorema 7.15. El vector $\mathbf{c}$ pertenece a C si y sólo si $\mathbf{c}$ es ortogonal a cada fila de $\mathbf{B}$.

Demostración. (a) Supongamos que $\mathbf{c} \in C$. Sea $v \in V(C)$, donde C es el subgrafo representado por el vector $\mathbf{c}$, y sea $\mathbf{b}_v$ la fila de $\mathbf{B}$ que corresponde al vértice v. Como el número de aristas de C incidentes en v es par, claramente $\mathbf{b}_v \cdot \mathbf{c} = \mathbf{0}$. Si $v \notin V(C)$ también, trivialmente, $\mathbf{b}_v$ es ortogonal a $\mathbf{c}$.

(b) Sea $\mathbf{c} \in E$ ortogonal a cada vector fila de $\mathbf{B}$. Esto quiere decir que cada vértice v del grafo tiene en C grado par (o cero). (¿Por qué?) Así, C es euleriano y se descompone en ciclos disyuntos en aristas. □

Consideremos ahora un árbol generador T del grafo G y sea c una cuerda respecto a T. El subgrafo $T + c$ contiene exactamente un ciclo C_c de G. Este ciclo C_c creado añadiendo la cuerda c, lo llamamos *ciclo fundamental respecto a* T. Como todo árbol generador tiene $n-1$ aristas, existen $\mu = m - n + 1$ ciclos fundamentales. El parámetro μ se conoce como el *número ciclomático* de G.

Ejemplo: Para el grafo G de la figura 7.14, con $E(G) = \{c_1, c_2, c_3, c_4, r_5, r_6, r_7, r_8\}$, los ciclos fundamentales asociados a las cuerdas c_1, c_2, c_3, c_4 respecto del árbol generador considerado son:

- $C_1 = (1,0,0,0,0,1,1,0)$
- $C_2 = (0,1,0,0,0,0,1,1)$
- $C_3 = (0,0,1,0,1,0,1,1)$
- $C_4 = (0,0,0,1,0,1,1,1)$

Teorema 7.16. Los μ ciclos fundamentales respecto a un árbol generador T constituyen una base del subespacio de ciclos C.

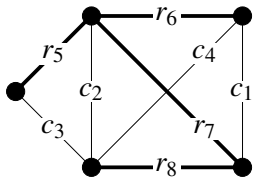


Figura 7.14: Ciclos fundamentales

Demostración. Sean $C_1, C_2, \ldots, C_\mu$ los ciclos fundamentales asociados a las cuerdas $c_1, c_2, \ldots, c_\mu$ (respecto al árbol generador T). Se tiene que demostrar que $C_1, C_2, \ldots, C_\mu$ son linealmente independientes y que generan C.

(a) Cada C_i contiene exactamente una cuerda, es decir, la cuerda c_i. Por tanto, el componente i–ésimo de $\alpha_1 C_1 \oplus \alpha_2 C_2 \oplus \cdots \oplus \alpha_\mu C_\mu$ es α_i. Así, entonces, $\alpha_1 C_1 \oplus \alpha_2 C_2 \oplus \cdots \oplus \alpha_\mu C_\mu = \mathbf{0}$ implica necesariamente $\alpha_1 = \alpha_2 = \cdots = \alpha_m = 0$.

(b) Sea $C \in C$ y $c_{i_1}, c_{i_2}, \ldots, c_{i_r}$, $1 \leqslant r \leqslant \mu$, las aristas de C que son cuerdas respecto al árbol generador T considerado. Por otra parte, sea $\Gamma = C_{i_1} \oplus C_{i_2} \oplus \cdots \oplus C_{i_r}$ (en C) y supongamos que $\Gamma \neq C$. Es evidente que las cuerdas de Γ son también $c_{i_1}, c_{i_2}, \ldots, c_{i_r}$. Pero entonces el subgrafo $\Gamma \oplus C$ no contiene ninguna cuerda y, por tanto, es imposible que pertenezca a C, cosa que contradice el teorema 7.14. Así, $C = \Gamma = C_{i_1} \oplus C_{i_2} \oplus \cdots \oplus C_{i_r}$, y C es generado por los ciclos fundamentales. □

Ejemplo: Para el grafo de la figura 7.14, el ciclo $C = (0,0,1,1,1,1,0,0)$ es una combinación de los ciclos fundamentales $C_3 = (0,0,1,0,1,0,1,1)$ y $C_4 = (0,0,0,1,0,1,1,1)$, ya que las aristas de C que son cuerdas respecto del árbol generador considerado son c_3 y c_4.

Matriz de ciclos fundamentales

Si $C_1, C_2, \ldots, C_\mu$, son los ciclos fundamentales de G, la *matriz de ciclos fundamentales* $\mathbf{C}_f = \mathbf{C}_f(G)$ es la matriz binaria $\mu \times m$ que tiene por elementos (c_{ij}):

$$(c_{ij}) = \begin{cases} 1 & \text{si la arista } e_j \text{ pertenece al ciclo } C_i, \\ 0 & \text{de otro modo;} \end{cases}$$

es decir, sus vectores fila son los vectores asociados a los ciclos fundamentales.

El teorema 7.15 nos dice que

$$\mathbf{B}\,\mathbf{C}_f^T = \mathbf{C}_f\,\mathbf{B}^T = \mathbf{0}$$

siendo $\mathbf{B}$ la matriz de incidencia.

Cociclos

Un concepto dual del de ciclo, y también interesante por sus aplicaciones, es el de corte simple o cociclo. Un conjunto de aristas de corte en un grafo $G = (V, E)$ (que continuamos suponiendo simple y conexo) es un conjunto $S \subset E$ tal que $G - S = (V, E \setminus S)$ es no conexo. Un *corte simple* o *cociclo* es un conjunto de corte S minimal, es decir, tal que ningún subconjunto propio desconecta el grafo.

Sea T un árbol generador de G. Si r es una arista de T, sean V_1 y V_2 los conjuntos de vértices correspondientes a los dos componentes de $T - r$. Notemos que V_1 y V_2 constituyen una partición de $V(G)$. El corte simple, S_r, de G asociado a esta partición se llama el corte fundamental o *cociclo fundamental* asociado a r. Por tanto, el número de cociclos fundamentales de G es el número de aristas de un árbol generador, es decir, $n-1$. Así como cada ciclo fundamental contiene exactamente una cuerda y el resto de las aristas del ciclo son aristas del árbol generador, cada cociclo fundamental contiene exactamente una arista del árbol generador y el resto de elementos son cuerdas.

Ejemplo: Para el grafo de la figura 7.14, los 4 cociclos fundamentales asociados a las aristas del árbol generador considerado son:

- $S_5 = \{r_5, c_3\}$
- $S_6 = \{r_6, c_1, c_4\}$
- $S_7 = \{r_7, c_1, c_2, c_3, c_4\}$
- $S_8 = \{r_8, c_2, c_3, c_4\}$

Sea S el conjunto de los subgrafos de G que son descomponibles en cociclos. El conjunto S es también estable por la operación $\oplus$ (problema 19). Además, cada corte simple S, lo podemos representar mediante un vector $s \in E$. De hecho, S corresponde a un subespacio vectorial de E llamado *subespacio de cociclos*. Una base del subespacio de cociclos la constituye un conjunto de $n-1$ cociclos fundamentales respecto de un árbol generador del grafo.

Para finalizar esta discusión, presentamos los teoremas siguientes que relacionan los ciclos y cociclos fundamentales respecto de un árbol generador T.

Teorema 7.17. Cada arista r de T, que determina el cociclo fundamental S_r, aparece en cada ciclo fundamental asociado con las aristas de S_r que son cuerdas, y r no aparece en ningún otro ciclo fundamental.

Demostración. Sea $S_r = \{r, c_1, \ldots, c_k\}$, donde $c_1, \ldots, c_k$ son cuerdas. Como cada ciclo tiene con cada cociclo un número par de aristas en común (¿por qué?), si C_{c_i} es el ciclo fundamental asociado a $c_i \in S_r$, entonces $S_r \cap E(C_{c_i}) = \{r, c_i\}$. Por tanto, r aparece en cada ciclo fundamental

asociado con una cuerda de S_r. Con un razonamiento similar se demuestra que r no puede aparecer en ningún otro ciclo fundamental. □

Teorema 7.18. Cada cuerda c, que determina el ciclo fundamental C_c, aparece en cada cociclo fundamental asociado a las aristas (respecto de T) de C_c, y c no aparece en ningún otro cociclo fundamental.

La demostración de este resultado se deja como ejercicio (ver el problema 20). El lector tendría que comprobar los dos teoremas anteriores en el grafo de la figura 7.14.

Ciclos fundamentales en digrafos

La teoría de los ciclos fundamentales se puede extender sin dificultad al caso dirigido. Cuando G es un digrafo (sin lazos) con conjunto de arcos $A(G) = \{a_1, a_2, \ldots, a_m\}$, los vectores ciclo se definen de la manera siguiente: sea G^* el grafo subyacente que resulta de G al suprimir la orientación de los arcos. Supongamos que G^* es conexo, aunque puede tener aristas paralelas. Si C es un ciclo de G^*, orientamos C de manera arbitraria, por ejemplo en sentido horario. Entonces, el vector $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_m)$, que representa a C en el digrafo G, tiene el componente $\gamma_i = 1$ si el arco a_i pertenece a C y las orientaciones coinciden, $\gamma_i = -1$ si el arco a_i pertenece a C y las orientaciones no coinciden. Si el arco a_i no pertenece a C, entonces $\gamma_i = 0$.

Si G tiene n vértices, un árbol generador T de G es un subgrafo de G^* que es árbol y tal que $V(T) = V(G^*)$. Los ciclos fundamentales se definen ahora como en el caso no dirigido.

Ejemplo: Para el digrafo G de la figura 7.15, los ciclos fundamentales asociados a las cuerdas c_1, c_2, c_3, c_4 respecto del árbol generador considerado son:

- $C_1 = (1,0,0,0,0,1,-1,0)$
- $C_2 = (0,-1,0,0,0,0,1,1)$
- $C_3 = (0,0,1,0,1,0,1,1)$
- $C_4 = (0,0,0,-1,0,1,-1,-1)$

donde $A(G) = \{c_1, c_2, c_3, c_4, r_5, r_6, r_7, r_8\}$ y los ciclos se consideran orientados en sentido horario.

Ahora asociamos a G el espacio vectorial real $\mathbf{R}^m$. El subespacio de ciclos C es el subespacio generado por los vectores ciclo. Con estas definiciones, los teoremas 7.15 y 7.16 también se verifican para digrafos. En cuanto al teorema 7.15, la ortogonalidad entre vectores se entiende en el sentido usual en $\mathbf{R}^m$, y la matriz $\mathbf{B}$ es la matriz de incidencia del digrafo, tal como se ha definido en el capítulo 7. Por ejemplo, demostremos para el caso dirigido la proposición siguiente (que forma parte del teorema 7.15).

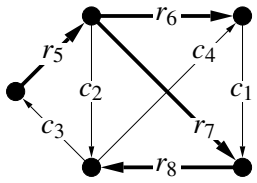


Figura 7.15: Ciclos fundamentales en el digrafo G

Proposición 7.19. Si el vector **c** es ortogonal a cada fila de la matriz **B**, entonces **c** es combinación lineal de ciclos fundamentales.

Demostración. Sean γ_{i_j}, $1 \leqslant j \leqslant \mu$ los componentes de **c** que corresponden a las cuerdas respecto del árbol generador T considerado. Demostremos que $\mathbf{c} = \sum_{j=1}^{\mu} \gamma_{i_j} C_{i_j}$, donde los C_{i_j} son los ciclos fundamentales. En efecto, si consideramos el vector $\mathbf{d} = \mathbf{c} - \sum_{j=1}^{\mu} \gamma_{i_j} C_{i_j}$, **d** sólo puede tener diferente de cero los componentes que corresponden a aristas de T. Además, **d** también es ortogonal a cada fila de **B** (¿por qué?). Consideremos ahora un vértice v que tenga grado 1 en T y sea a_k la arista de T incidente con v. Entonces, por la ortogonalidad supuesta, el componente d_k de **d** tiene que ser 0. Repitiendo este razonamiento con el árbol T_1 que resulta de T al suprimir los vértices de grado 1, demostramos que también aquellos componentes del vector **d** que corresponden a vértices de grado 2 en T son 0. Iterando el proceso se obtiene $\mathbf{d} = \mathbf{0}$, es decir, $\mathbf{c} = \sum_{j=1}^{\mu} \gamma_{i_j} C_{i_j}$, tal como se quería demostrar. □

Ejemplo: Para el digrafo de la figura 7.15, el ciclo $C = (0,0,1,-1,1,1,0,0)$ es $C_3 + C_4$ donde C_3 y C_4 son los ciclos fundamentales $(0,0,1,0,1,0,1,1)$ y $(0,0,0,-1,0,1,-1,-1)$.

La última sección del capítulo se dedica a presentar la aplicación de los ciclos fundamentales en el análisis de redes eléctricas.

7.4 Análisis de redes eléctricas

El comportamiento de una red eléctrica depende de las características de los elementos que la componen y de su topología, es decir, de la manera como estos elementos están interconectados. Es en este punto donde la teoría de grafos proporciona una herramienta matemática que permite el análisis sistemático. Esta aplicación de la teoría de grafos fue introducida por G. Kirchhoff en el año 1847 y en la actualidad tiene una gran importancia, dado que constituye la base de los programas para ordenador que permiten hacer el análisis automático de grandes redes eléctricas.

Sólo consideraremos redes *RLC* formades por resistencias R, bobinas L y condensadores C además de fuentes de tensión y de corriente independientes. Esta formulación es suficientemente general, ya que cualquier elemento eléctrico con dos terminales, lineal, pasivo e invariante en el tiempo, puede ser modelado por una combinación de elementos R, L y C.

Cada elemento R, L o C de la red, lo representaremos por una arista e orientada de manera arbitraria. Asociadas con esta arista e tenemos las variables $v(t)$ e $i(t)$ que corresponden al valor de la tensión y la intensidad de la corriente en el elemento considerado y en el instante de tiempo t. El signo de estas variables se toma de acuerdo con la orientación de e tal como se indica en la figura 7.16. En cada elemento, la corriente y la tensión se relacionan de acuerdo con la ley física que gobierna su comportamiento. Así, en una resistencia se cumple la ley de Ohm $v(t) = Ri(t)$, en una bobina la ley de Faraday $v(t) = L\frac{di(t)}{dt}$ y en un condensador la tensión es proporcional a la carga eléctrica acumulada, es decir, $v(t) = \frac{1}{C}\int i(\tau)\, d\tau$, siendo R, L y C los valores de la resistencia, la autoinducción y la capacidad respectivamente. La red resultante de

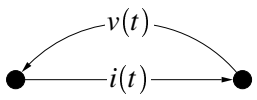


Figura 7.16: Variables $v(t)$ e $i(t)$

interconectar los diferentes elementos que la componen queda modelada por un digrafo G. Por el hecho de interconectar los elementos, se cumplen unas restricciones adicionales conocidas como *leyes de Kirchoff*.

Ley de Kirchoff de las corrientes Para cada vértice u de G, la suma algebraica de las corrientes que salen de u vale 0.

Ley de Kirchoff de las tensiones Para cada circuito C de G, la suma algebraica de las tensiones en las aristas de C vale 0.

Análisis por ciclos

Sea $\mathbf{B} = (b_{ij})$ la matriz de incidencia de G. La matriz $\mathbf{B}$ es $n \times m$ donde n y m son el orden y el tamaño de G respectivamente. Sea también $\mathbf{i}(t) = (i_1(t), i_2(t), \ldots, i_m(t))^T$ el vector que tiene por componentes las corrientes en cada arista. De acuerdo con la ley de Kirchoff de las corrientes, en el vértice r–ésimo de G se cumple $\sum_{k=1}^{m} b_{rk} i_k(t) = 0$, o bien matricialmente

$$\mathbf{B}\mathbf{i}(t) = \mathbf{0} \tag{7.1}$$

Por la ecuación 7.1, el vector $\mathbf{i}(t)$ es ortogonal a cada vector fila de la matriz de incidencia $\mathbf{B}$. Tal como se ha explicado en la sección anterior (proposición 7.19), esto significa que $\mathbf{i}(t)$ pertenece al subespacio de ciclos C del espacio vectorial $E = \mathbf{R}^m$ asociado a G. Sea $\mu = m - n + 1$ el número de ciclos fundamentales de G respecto de un árbol generador T. Si $\mathbf{C}_f$ es la matriz $\mu \times m$ de ciclos fundamentales respecto de T, sus filas $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_\mu$ constituyen una base de C y, por tanto, $\mathbf{i}(t) = \sum_{k=0}^{\mu} i_{ck}(t)\mathbf{c}_k$, siendo $i_{c1}(t), i_{c2}(t), \ldots, i_{c\mu}(t)$ los componentes del vector corriente $\mathbf{i}(t)$. Matricialmente:

$$\mathbf{i}(t) = \mathbf{C}_f^T \mathbf{i}_c(t) \tag{7.2}$$

donde $\mathbf{i}_c(t)$ es el vector $(i_{c1}(t), i_{c2}(t), \ldots, i_{c\mu}(t))^T$. Las corrientes $i_{c1}(t), i_{c2}(t), \ldots, i_{c\mu}(t)$ se llaman *corrientes de ciclo* y, por la ecuación 7.2, la corriente $i_k(t)$ que atraviesa el elemento k–ésimo de la red es combinación lineal de las μ corrientes de ciclo.

Substituyendo 7.2 en 7.1 tenemos la ley de Kirchoff de las corrientes expresada por las corrientes de ciclo:

$$\mathbf{B}\,\mathbf{C}_f^T \mathbf{i}_c(t) = \mathbf{0} \tag{7.3}$$

Consideremos ahora una red *RLC* en la cual todas las fuentes de energía sean generadores de tensión. Si $\mathbf{e}_c(t) = (e_1(t), e_2(t), \ldots, e_\mu(t))^T$ es el vector que tiene por componente k–ésimo la tensión suministrada por los generadores del k–ésimo ciclo fundamental $\mathbf{c}_k$, la ley de Kirchoff de las tensiones se puede formular como:

$$\mathbf{C}_f\,\mathbf{v}(t) = \mathbf{e}_c(t) \tag{7.4}$$

donde $\mathbf{v}(t) = (v_1(t), v_2(t), \ldots, v_m(t))^T$ es el vector tensión que tiene por componentes las tensiones en las aristas correspondientes a los elementos pasivos R, L y C.

Tal como se ha dicho antes, la tensión $v_k(t)$ y la corriente $i_k(t)$ se relacionan en cada elemento según la ley física que describe su comportamiento. Si $V_k(s)$ e $I_k(s)$ son las transformadas de Laplace de las variables $v_k(t)$ e $i_k(t)$ respectivamente, y suponemos que las condiciones iniciales de la red son nulas (es decir, para cada bobina $i(0) = 0$ y para cada condensador $v(0) = 0$), se puede escribir la siguiente ecuación matricial:

$$\mathbf{V}(s) = \mathbb{Z}(s)\,\mathbf{I}(s) \tag{7.5}$$

donde $\mathbb{Z}(s)$ es la matriz diagonal $m \times m$ llamada *matriz de impedancias*, tal que $(\mathbb{Z}(s))_{kk} = Z_k(s)$ es la impedancia del elemento k–ésimo de la red. Así, $V_k(s) = Z_k(s)I_k(s)$, donde $Z_k(s)$ es igual a R, Ls o $\frac{1}{Cs}$, según si el elemento considerado es una resistencia, una bobina o un condensador.

De la ecuación obtenida al transformar por Laplace la ecuación 7.4 y de 7.5:

$$\mathbf{C}_f\,\mathbb{Z}(s)\,\mathbf{I}(s) = \mathbf{E}_c(s) \tag{7.6}$$

Por otra parte, transformando la ecuación 7.2, $\mathbf{i}(t) = \mathbf{C}_f^T \mathbf{i}_c(t)$, y substituyendo en 7.6:

$$(\mathbf{C}_f \mathbb{Z}(s) \mathbf{C}_f^T) \mathbf{I}_c(s) = \mathbf{E}_c(s) \tag{7.7}$$

donde la matriz $\mathbf{C}_f \mathbb{Z}(s) \mathbf{C}_f^T \equiv \mathbb{Z}_c(s)$ es $\mu \times \mu$ y se llama *matriz de impedancias de ciclo*.

La solución del sistema 7.7 da las corrientes de ciclo y la ecuación 7.2 proporciona las corrientes en cada elemento de la red.

Finalmente, indicamos que, utilizando los cociclos fundamentales, se puede hacer el análisis de la red usando como variables independientes las tensiones de $n-1$ de los nodos respecto de un nodo de referencia. Naturalmente, este análisis es dual del presentado en esta sección.

Notas históricas y bibliográficas

El trabajo original de Euler sobre el problema de los puentes de Königsberg se publicó en [7]. Curiosamente, Euler enuncia una de las implicaciones (si el grado de todos los vértices es par, el grafo tiene un circuito euleriano) sin demostrarla. Esta parte de la demostración no apareció publicada hasta 1873 en un artículo póstumo del joven matemático alemán Carl Hierholzer [11]. Tanto las circunstancias históricas de los dos trabajos como una transcripción al inglés de los textos originales se pueden encontrar en el libro de Biggs, Lloyd y Wilson sobre la historia de la teoría de grafos [4]. El lector que quiera profundizar más en el tema de los ciclos eulerianos en grafos y digrafos puede consultar por ejemplo [2].

El problema de las secuencias de de Bruijn fue introducido y tratado por este matemático holandés en 1946 en [5], donde presenta también los digrafos que llevan su nombre. La relación entre las secuencias de de Bruijn y los registros cíclicos de desplazamiento, así como también algunas de sus aplicaciones se pueden encontrar por ejemplo en [14]. Los digrafos de los cuales se derivan las secuencias tienen muchas propiedades interesantes que se discuten, por ejemplo, en [3], donde también hay un análisis de su aplicación al diseño de redes.

A pesar de que les dio el nombre, Hamilton no fue el primero en estudiar ciclos que pasan una única vez por cada vértice. El matemático inglés Kirkman había dirigido a la Royal Society el mismo problema para los grafos de los poliedros [12]. No obstante, Hamilton lo hizo popular, sobre todo porque inventó un juego de mesa basado en la existencia de ciclos hamiltonianos en el grafo del icosaedro, *The icosian game*, del cual estaba muy orgulloso, pero que fue un fracaso comercial. Nuevamente, el lector encontrará una excelente exposición histórica y transcripciones de los artículos originales en [4]. Sobre el estudio de ciclos hamiltonianos se puede encontrar un resumen extenso y bibliografía adicional en [2].

El problema del viajante es uno de los problemas de la algorítmica que ha hecho correr más tinta, no solamente por las aplicaciones que se derivan de su solución, sino también porque constituye un test de la eficacia de nuevos métodos en este área. Aunque se puede encontrar

una exposición relativamente completa en [8], hay libros enteramente dedicados al problema, como por ejemplo [13]. En la referencia anterior [8] se puede encontrar también el tratamiento algorítmico de los problemas de circuitos eulerianos y ciclos hamiltonianos.

Kirchoff enunció sus dos famosas leyes para el análisis de redes eléctricas a los 20 años cuando aún era estudiante de física. En [4] hay también una transcripción del trabajo original de Kirchoff y referencias históricas. En [6] hay una exposición completa de la aplicación de los ciclos y los cociclos fundamentales al análisis de redes eléctricas.

Como en el resto de capítulos de esta parte, el libro de Wilson [15] es una buena introducción de los temas que se tratan en este capítulo.

Bibliografía

[1] E. Aarts, J. Korst. *Simulated Annealing and Boltzmann Machines*, John Wiley & Sons, Chichester, 1989.

[2] L. Beinecke, R. Wilson (Eds.). *Selected Topics in Graph Theory*, Academic Press, 1978.

[3] J. C. Bermond, C. Peyrat. "De Bruijn and Kautz networks, a competition for the hypercube", *Hypercube and Distributed Computers*, Elsevier Sc. Publ, North Holland, pp. 279–293, 1989.

[4] N. L. Biggs, E. K. Lloyd, R. J. Wilson. *Graph Theory 1736–1936*, Oxford University Press, 1976.

[5] N. G. de Bruijn. "A combinatorial problem", *Koninklijke Nederlandse Academie van Watenschappen Proc.*, **A49**, pp. 758–764, 1946.

[6] N. Deo. *Graph Theory with Applications to Engineering and Computer Science*, Prentice Hall, 1974.

[7] L. Euler. "Solutio problematis ad geometriam situs pertinentis", *Comentarii Academiae Scientiarum Imperialis Petropoitanae*, **8**, pp. 128–140, 1736.

[8] A. Gibbons. *Algorithmic Graph Theory*, Cambridge University Press, 1985.

[9] R. Gould. *Graph Theory*, The Benjamin/Cummings Publishing Company, Inc., 1988.

[10] D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, 1989.

[11] C. Hierholzer. "Über die Möglichkeit, einen Linienzug ohne Wiederholung und ohne Unterbrechnung zu Umfahren", *Mathematische Annalen*, **6**, pp. 30–32, 1873.

[12] T.P. Kirkman., "On the representation of polyhedra", *Philosophical Transactions of the Royal Society*, **146**, pp. 413–418, 1856.

[13] L. Lawler, J. K. Lenstra (Eds.). *The Travelling Salesman Problem*, John Wiley & Sons, 1987.

[14] H. Stone. *Discrete Mathematical Structures and their Applications*, SRA Computer Science Series, 1974.

[15] R. Wilson. *Introducción a la Teoría de Grafos*, Alianza Universidad, vol. 367, 1983.

Problemas

1. Estudiar si los grafos siguientes son eulerianos:
   (a) El grafo completo de cuatro y el de cinco vértices
   (b) El grafo del cubo
   (c) El grafo de Petersen
2. ¿Para qué valores de n el grafo completo de n vértices es euleriano?
3. Estudiar para qué valores de n y m los grafos bipartitos completos $K_{n,m}$ son eulerianos.
4. ¿Cuántas veces (como mínimo) se tiene que alzar el lápiz del papel para dibujar la figura 7.17 sin repetir ninguna línea?

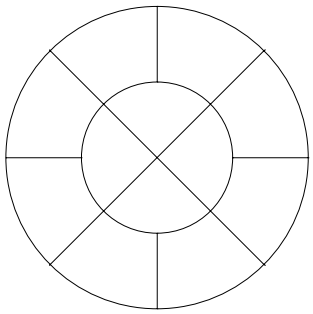


Figura 7.17:

5. Las autoridades actuales de la antigua ciudad de Königsberg han decidido finalmente construir los puentes que sea preciso para satisfacer el antiguo capricho de sus habitantes. ¿Cuántos han de construir como mínimo y dónde es preciso ponerlos?

6. Demostrar que un multidigrafo G es euleriano si y sólo si es fuertemente conexo (salvo vértices aislados) y para todo $x \in V(G)$, $d^+(x) = d^-(x)$ (teorema 7.4).

7. Demostrar que un multidigrafo G contiene un camino euleriano desde el vértice u hasta el vértice v, $u \neq v$ si y sólo si el multigrafo subyacente es conexo (salvo vértices aislados), y para todo $x \in V(G) \setminus \{u, v\}$, $d^+(x) = d^-(x)$, mientras que $d^+(u) = d^-(u) + 1$ y $d^-(v) = d^+(v) + 1$ (teorema 7.5).

8. *Problema del cartero chino*. Un cartero tiene que recorrer todas las calles de su pueblo para repartir las cartas. Si el mapa del pueblo es el de la figura 7.18, proponer un recorrido para el cartero saliendo de la oficina de correos (señalada con un cuadrado negro) y volviendo a la oficina de correos de manera que la distancia recorrida sea mínima.

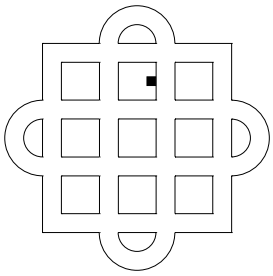


Figura 7.18:

9. Una modificación del procedimiento *RecorrerCamino* del algoritmo *Euler* es la siguiente: Si $G = (V, E)$ es el grafo original, en cada paso sea G' el grafo con conjunto de vértices V (los del grafo original) y aristas, todas las aristas aún no consideradas del grafo original. Ahora, en lugar de escoger una arista no usada cualquiera a partir del punto actual, se escoge una arista tal que su exclusión del grafo no incrementa el número de componentes no triviales del grafo G'. Si no hay ninguna arista de este tipo se considera una arista cualquiera, el algoritmo así modificado es conocido como *algoritmo de Fleury*. Aplicar el algoritmo de Fleury al grafo de la figura 7.4.

10. Existe un teorema que dice: Si G tiene un circuito euleriano, entonces el algoritmo de Fleury lo encuentra. Es decir, con este algoritmo, no es preciso hacer llamadas recursivas. El algoritmo no acaba hasta que se ha pasado por todas las aristas. Demostrar el teorema.

    Indicación: Mostrar que, si G tiene un circuito euleriano, entonces la única vez en que no hay una arista desde el vértice actual v, cuya exclusión no incrementa el número de

componentes del grafo, es cuando el grado de v en G' es 1. Esto significa que cuando no hay aristas libres incidentes en el vértice considerado, no las hay en ninguna parte. ¿Por qué? ¿Por qué esto es importante?

11. *Problema del hotel de Baltimore*. En el hotel Hilton de Baltimore no hay llaves en las puertas de las habitaciones, sino cerraduras electrónicas que se abren tecleando un código secreto de cinco dígitos (0 o 1). Un malhechor quiere entrar en la habitación de Madame Castafiore para robarle las joyas, y teclea todos los números de cinco cifras desde 00000 hasta 11111, o sea, que teclea 160 cifras. Demostrar que habría podido abrir la puerta con una quinta parte de cifras y dar la secuencia que es preciso teclear.

12. *Problema de la localización de la posición de un disco*. Para localizar la posición de un disco magnético, se inscriben bits (ceros o unos) de manera que una lectora lee cuatro dígitos consecutivos. ¿Cuántos dígitos se pueden registrar y en qué orden para que cada posición de la lectora corresponda a una posición diferente del disco?

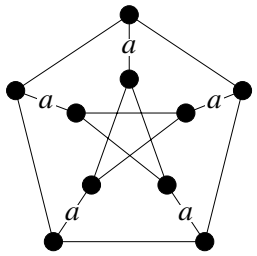


Figura 7.19: El grafo de Petersen

13. Demostrar que el grafo de Petersen (en la figura 7.19) no es hamiltoniano, pero que el grafo que se obtiene suprimiendo uno cualquiera de los vértices lo es. (Un grafo con esta propiedad se dice que es *hipohamiltoniano*. El de Petersen es el grafo hipohamiltoniano más pequeño en número de vértices.)

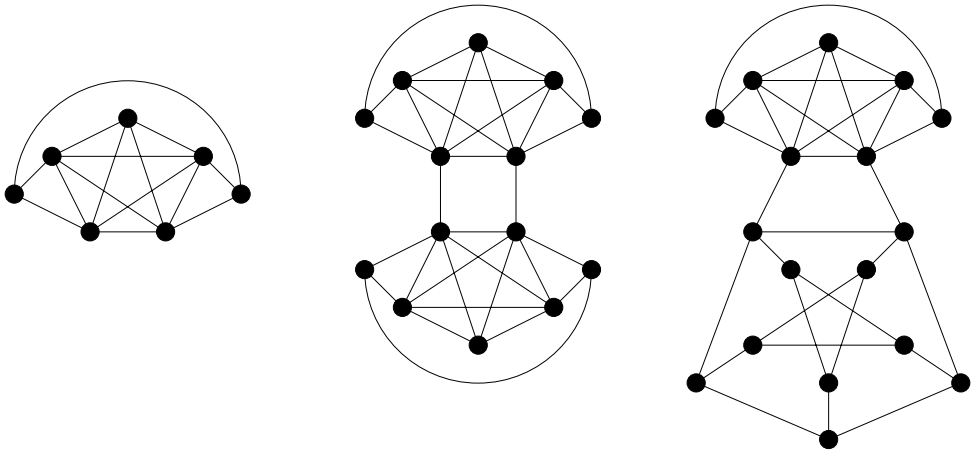
    Indicación: ¿Cuántas aristas de tipo a tendría que tener un ciclo hamiltoniano en un grafo de Petersen?

14. En los grafos dibujados en la figura 7.20 mostrar que:

    (a) El grafo (a) es hamiltoniano

    (b) El grafo (b) es hamiltoniano

    (c) El grafo (c) no es hamiltoniano



a b c

Figura 7.20:

15. (a) Demostrar que un grafo bipartito de orden impar no es hamiltoniano.

    (b) Demostrar que un grafo bipartito con conjuntos estables U y V no puede ser hamiltoniano si $|U| \neq |V|$.

    (c) Determinar qué grafos bipartitos completos son hamiltonianos.

16. El origen real del estudio de los grafos hamiltonianos proviene del viejo problema de determinar si con el movimiento del caballo se pueden recorrer todos los cuadros de un tablero de ajedrez pasando por cada uno una única vez. La respuesta es que sí para un tablero de 8×8 cuadros (Vandermonde, 1771; Kirkman, 1856). Considerar un grafo que represente el problema y demostrar que, si el tablero es de 5×5 cuadros, la respuesta es que no.

17. (a) Demostrar que el grafo completo K_n con n impar se puede descomponer en $n-2$ caminos hamiltonianos disyuntos en aristas.

    (b) Demostrar que el grafo completo K_n con n par se puede descomponer en n ciclos hamiltonianos disyuntos en aristas.

18. Demostrar que cada ciclo C tiene con cada cociclo S un número par de aristas en común.

19. Demostrar que la suma $\oplus$ de cociclos es también un cociclo o es la unión de cociclos disyuntos en aristas.

20. Demostrar el teorema 7.18: Cada cuerda c, que determina el ciclo fundamental C_c, aparece en cada cociclo fundamental asociado con las aristas (respecto de T) de C_c, y c no aparece en ningún otro cociclo fundamental.

21. Demostrar que, en un grafo conexo G, un vértice v es de corte si y sólo si existen dos aristas e y e' incidentes con v y tales que ningún ciclo de G las contiene a las dos.

Capítulo 8

Flujos, conectividad y apareamientos

1. Redes de transporte
2. El teorema del flujo máximo–corte mínimo
3. Conectividad
4. Los teoremas de Menger
5. Apareamientos en grafos bipartitos
6. El teorema de Hall

En este capítulo se estudian tres temas aparentemente no relacionados, pero que, como se verá, poseen un vínculo estrecho. Se comienza, en la sección 1, estudiando los flujos en redes de transporte, cuestión de gran aplicación a ciertos problemas de investigación operativa. En primer lugar, se definen los conceptos de red de transporte, de valor del flujo y de capacidad de un corte. Después de establecer que el valor del flujo es menor o igual que la capacidad de cualquier corte, se demuestra en la sección 2 el clásico teorema del flujo máximo–corte mínimo y se presenta el algoritmo de Ford y Fulkerson, que permite encontrar el flujo máximo en la red.

En las secciones 3 y 4 del capítulo se estudian los importantes conceptos de conectividad y arista-conectividad. Cuando el grafo que se considera modela una red de interconexión, las conectividades constituyen medidas de la vulnerabilidad de la red ante el fallo de nodos y/o enlaces. Los teoremas de Menger constituyen los resultados clásicos sobre conectividad y relacionan las conectividades locales con el número máximo de caminos internamente disyuntos en el grafo. Se presenta la relación estrecha que hay entre estos teoremas y el teorema del flujo máximo–corte mínimo.

En las últimas secciones del capítulo se presenta otro de los resultados básicos de la teoría de grafos y la combinatoria: el teorema de Hall. Este resultado proporciona la condición

necesaria y suficiente para que se puedan aparear los vértices de un grafo bipartito. El teorema de Hall constituye, también, el resultado principal de la teoría transversal de conjuntos, y se aplica a problemas de asignación de tareas. La demostración que se presenta del teorema de Hall usa, de nuevo, el teorema del flujo máximo–corte mínimo.

8.1 Redes de transporte

En ciertas aplicaciones interesa determinar el flujo máximo (de un fluido, datos, etc) que fluye a través de una red desde un cierto nodo s hasta otro nodo t, cuando los enlaces de la red tienen una capacidad limitada de transmisión del flujo. El modelo que propociona la teoría de grafos para resolver este problema lo constituyen las llamadas redes de transporte.

Una *red de transporte* $X = (G, s, t, c)$ es un digrafo $G = (V, A)$ con dos vértices distinguidos, s y t, y una función c llamada *capacidad* que asigna a cada arco $a = (u, v) \in A(G)$ un valor entero no negativo $c(a) = c(u, v)$ llamado la capacidad del arco a.

Dada una red de transporte X, una función $\phi : A \longrightarrow \mathbb{Z}$ que cumple

$$0 \leqslant \phi(a) = \phi(u, v) \leqslant c(a) \qquad \text{para cada } a = (u, v) \in A(G) \tag{8.1}$$

y

$$\sum_{v \in \Gamma^+(u)} \phi(u, v) = \sum_{v \in \Gamma^-(u)} \phi(v, u) \qquad \text{para cada } u \in V(G) - \{s, t\} \tag{8.2}$$

se llama un *flujo* en X. Diremos también que $\phi(a)$ es el flujo que atraviesa el arco a.

La condición 8.1 acota el flujo que atraviesa un arco determinado por su capacidad. Por otra parte, dado $u \in V(G)$ diremos que

$$\sum_{v \in \Gamma^+(u)} \phi(u, v) - \sum_{v \in \Gamma^-(u)} \phi(v, u)$$

es el *flujo neto saliente* de u y, análogamente,

$$\sum_{v \in \Gamma^-(u)} \phi(v, u) - \sum_{v \in \Gamma^+(u)} \phi(u, v)$$

es el *flujo neto entrante* en u, entonces la condición 8.2 es una condición de equilibrio que dice que el flujo neto saliente de (o entrante en) cada vértice u diferente de s y t vale cero. En la figura 8.1 se representa una red de transporte con un flujo asociado. El primero de los dos valores indicados en cada arco a es $c(a)$ y el segundo es $\phi(a)$.

Ejercicio 8.1. Demostrar que toda red de transporte admite al menos un flujo.

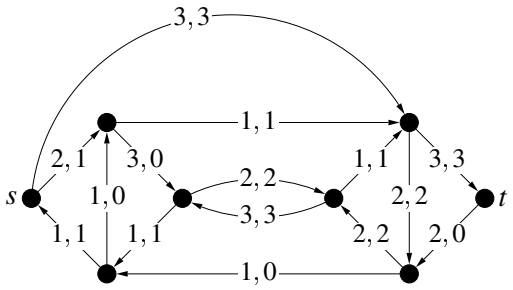


Figura 8.1: Red de transporte

El *valor del flujo*, $\mathrm{val}(\phi)$, se define por

$$\mathrm{val}(\phi) = \sum_{v\in\Gamma^+(s)} \phi(s,v) - \sum_{v\in\Gamma^-(s)} \phi(v,s) \tag{8.3}$$

y corresponde al flujo neto saliente del vértice s. De hecho, el valor del flujo es también el flujo neto entrante en t.

Proposición 8.2. Dada una red de transporte X y un flujo ϕ en la red, se cumple

$$\mathrm{val}(\phi) = \sum_{v\in\Gamma^+(s)} \phi(s,v) - \sum_{v\in\Gamma^-(s)} \phi(v,s) = \sum_{v\in\Gamma^-(t)} \phi(v,t) - \sum_{v\in\Gamma^+(t)} \phi(t,v)$$

Demostración. Sea $G=(V,A)$ el digrafo correspondiente a X. Se cumple

$$\sum_{u\in V} \sum_{v\in\Gamma^+(u)} \phi(u,v) = \sum_{u\in V} \sum_{v\in\Gamma^-(u)} \phi(v,u) \tag{8.4}$$

ya que cada lado de esta igualdad es $\sum_{a\in A}\phi(a)$. Aplicando la condición 8.2, la igualdad anterior es, simplemente:

$$\sum_{v\in\Gamma^+(s)} \phi(s,v) + \sum_{v\in\Gamma^+(t)} \phi(t,v) = \sum_{v\in\Gamma^-(s)} \phi(v,s) + \sum_{v\in\Gamma^-(t)} \phi(v,t)$$

de donde se obtiene el resultado enunciado. □

En la red de la figura 8.1, el valor del flujo es 3.

Dado $S\subset V$, denotamos por $(S,\overline{S})$ el conjunto de todos los arcos (u,v) con $u\in S$ y $v\in\overline{S}$, donde $\overline{S}=V\setminus S$. Un *s–t corte* es un conjunto $F=(S,\overline{S})$ tal que $s\in S$ y $t\in\overline{S}$. La suma $\sum_{a\in F} c(a)$ de las capacidades de los arcos que forman F es la capacidad $c(F)$ del *s–t* corte. Por ejemplo, la figura 8.2 muestra un *s–t* corte de capacidad 6 para la red de transporte de la figura 8.1, donde $S=\{s,1,2,3\}$ y $\overline{S}=\{4,5,6,t\}$.

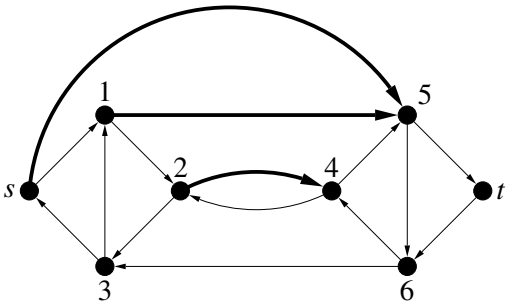


Figura 8.2: *s–t* corte

Dado un *s–t* corte $F = (S, \overline{S})$, sea $\overline{F}$ el conjunto de arcos $(\overline{S}, S)$.

Notemos que un *s–t* corte F desconecta G en el sentido que $G - F$ no contiene ningún camino dirigido de s hacia t. Por ello, se cumple el resultado siguiente:

Proposición 8.3. Sea X una red de transporte, ϕ un flujo en X y F un *s–t* corte. Entonces, $\text{val}(\phi) \leqslant c(F)$.

Demostración. Sumando para todos los vértices de S el flujo neto saliente de cada uno de estos vértices, se obtiene, teniendo en cuenta que $s \in S$, $t \in \overline{S}$ y las ecuaciones 8.2 y 8.3:

$$\text{val}(\phi) = \sum_{u \in S} \left(\sum_{v \in \Gamma^+(u)} \phi(u,v) - \sum_{v \in \Gamma^-(u)} \phi(v,u) \right) \tag{8.5}$$

Sea G_S el subgrafo dirigido inducido por S, es decir, $V(G_S) = S$ y $(u,v) \in A(G_S)$ si y sólo si $(u,v) \in A(G)$. Tenemos:

$$\sum_{u \in S} \sum_{v \in \Gamma^+(u)} \phi(u,v) = \sum_{a \in A(G_S)} \phi(a) + \sum_{a \in F} \phi(a)$$

Análogamente,

$$\sum_{u \in S} \sum_{v \in \Gamma^-(u)} \phi(v,u) = \sum_{a \in A(G_S)} \phi(a) + \sum_{a \in \overline{F}} \phi(a)$$

Por tanto, a partir de la ecuación 8.5 y aplicando la condición 8.1:

$$\text{val}(\phi) = \sum_{a \in F} \phi(a) - \sum_{a \in \overline{F}} \phi(a) \leqslant \sum_{a \in F} \phi(a) \leqslant \sum_{a \in F} c(a) = c(F) \tag{8.6}$$

□

8.2 El teorema del flujo máximo–corte mínimo

Un *flujo máximo* en una red de transporte X es un flujo ϕ con la propiedad que $\mathrm{val}(\phi) \geqslant \mathrm{val}(\phi')$, siendo ϕ' cualquier otro flujo en la red. Dado que el número de flujos que podemos definir en X es finito, siempre existe un flujo máximo.

Un s–t corte F se dice que es un *corte mínimo* si $c(F) \leqslant c(F')$ para cualquier otro s–t corte F'. Notemos que, si ϕ y F son un flujo y un corte en X tales que $\mathrm{val}(\phi) = c(F)$, entonces, por la proposición 8.3, ϕ es un flujo máximo y F es un corte mínimo. Un caso particular en que esto sucede lo da el resultado siguiente:

Lema 8.4. Sea ϕ un flujo en X y F un s–t corte. Si $\phi(a) = c(a)$ para todo $a \in F$ y $\phi(a) = 0$ para todo $a \in \overline{F}$, entonces ϕ es un flujo máximo y F es un corte mínimo.

Demostración. Por 8.6,

$$\mathrm{val}(\phi) = \sum_{a \in F} \phi(a) - \sum_{a \in \overline{F}} \phi(a) = \sum_{a \in F} c(a) = c(F)$$

□

El resultado más importante sobre redes de transporte lo da el teorema siguiente, debido a Ford y Fulkerson y conocido también como el teorema del flujo máximo–corte mínimo.

Teorema 8.5. En cualquier red de transporte, el valor de un flujo máximo es igual a la capacidad de un corte mínimo.

Demostración. Sea ϕ un flujo máximo en la red de transporte $X = (G, s, t, c)$ que se considera. Demostremos la existencia de un s–t corte F que cumple las condiciones enunciadas en el lema anterior, es decir, $\phi(a) = c(a)$ para todo $a \in F$ y $\phi(a) = 0$ para todo $a \in \overline{F}$.

Sea $S \subset V(G)$ definido recursivamente por las condiciones siguientes:

(a) $s \in S$;

(b) si $u \in S$ y $\phi(u,v) < c(u,v)$, entonces $v \in \Gamma^+(u)$ pertenece a S;

(c) si $u \in S$ y $\phi(v,u) > 0$, entonces $v \in \Gamma^-(u)$ pertenece a S.

Demostremos que $t \notin S$. En efecto, si fuese $t \in S$, tendría que existir una secuencia P de vértices distintos

$$s = u_0, u_1, u_2, \ldots, u_k, u_{k+1}, \ldots, u_{n-1}, u_n = t \tag{8.7}$$

donde, o bien $(u_{k-1}, u_k) \in A(G)$ y $\phi(u_{k-1}, u_k) < c(u_{k-1}, u_k)$, o bien $(u_k, u_{k-1}) \in A(G)$ y $\phi(u_k, u_{k-1}) > 0$, $k = 1, 2, \ldots, n$. Denominamos P una *secuencia de aumento* y denotamos por

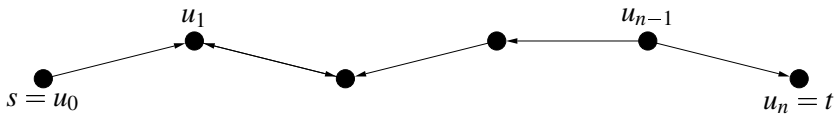


Figura 8.3: Secuencia de aumento

$A(P)$ el conjunto de arcos correspondientes; ver la figura 8.3. En el primer caso, cuando $a = (u_{k-1}, u_k) \in A(G)$ y $\phi(a) < c(a)$, sea $\varepsilon_k = c(a) - \phi(a)$. De otro modo, cuando $a = (u_k, u_{k-1}) \in A(G)$ y $\phi(a) > 0$, sea $\varepsilon_k = \phi(a)$. Sea $\varepsilon = \min_k \varepsilon_k$, $k = 1, 2, \ldots, n$.

Ahora, definimos una función ϕ' en $A(G)$ de la forma siguiente:

$$\phi'(a) = \begin{cases} \phi(a) + \varepsilon & \text{si } a = (u_{k-1}, u_k) \in A(P) \quad \text{para algún } k,\ 1 \leqslant k \leqslant n; \\ \phi(a) - \varepsilon & \text{si } a = (u_k, u_{k-1}) \in A(P) \quad \text{para algún } k,\ 1 \leqslant k \leqslant n; \\ \phi(a) & \text{si } a \notin A(P). \end{cases} \tag{8.8}$$

Claramente se cumple $0 \leqslant \phi'(a) \leqslant c(a)$ para todo $a \in A(G)$ y, por tanto, ϕ' verifica la condición 8.1. Es fácil comprobar que ϕ' también verifica la condición 8.2 y, por tanto, ϕ' es un flujo en X. Supongamos, por ejemplo, que para algún k, $1 \leqslant k \leqslant n-1$, se tuviese $(u_{k-1}, u_k) \in A(P)$ y $(u_k, u_{k+1}) \in A(P)$. Entonces, si $u = u_k$,

$$\sum_{v \in \Gamma^+(u)} \phi'(u, v) = \phi(u, u_{k+1}) + \varepsilon + \sum \phi(u, v)$$

donde la última suma se extiende a todos los vértices v adyacentes desde u que no pertenecen a P. Análogamente,

$$\sum_{v \in \Gamma^-(u)} \phi'(v, u) = \phi(u_{k-1}, u) + \varepsilon + \sum \phi(v, u)$$

donde, ahora, la última suma se extiende a todos los vértices v adyacentes hacia u que no pertenecen a P. Por tanto, $\sum_{v \in \Gamma^+(u)} \phi'(u, v) = \sum_{v \in \Gamma^-(u)} \phi'(v, u)$. De la misma forma se comprueba que la condición 8.2 se cumple en las otras situaciones posibles: cuando, para algún k, $1 \leqslant k \leqslant n-1$, se tiene $(u_k, u_{k-1}) \in A(P)$ y $(u_{k+1}, u_k) \in A(P)$, o cuando, para algún k, $1 \leqslant k \leqslant n-1$, u_k es adyacente hacia (desde) u_{k-1} y u_{k+1}.

Calculemos $\mathrm{val}(\phi')$. Si $s = u_0$ es adyacente hacia u_1, entonces

$$\begin{aligned} \mathrm{val}(\phi') &= \sum_{v \in \Gamma^+(s)} \phi'(s, v) - \sum_{v \in \Gamma^-(s)} \phi'(v, s) \\ &= \sum_{v \in \Gamma^+(s) - \{u_1\}} \phi(s, v) + (\phi(s, u_1) + \varepsilon) - \sum_{v \in \Gamma^-(s)} \phi(v, s) \\ &= \sum_{v \in \Gamma^+(s)} \phi(s, v) - \sum_{v \in \Gamma^-(s)} \phi(v, s) + \varepsilon \\ &= \mathrm{val}(\phi) + \varepsilon \end{aligned}$$

De la misma forma, $\text{val}(\phi') = \text{val}(\phi) + \varepsilon$ si $s = u_0$ es adyacente desde u_1. En cualquier caso, se llega a la contradicción que ϕ no es un flujo máximo. Por tanto, en contra de lo que se había supuesto, el vértice t no pertenece al conjunto S.

Así, el conjunto de arcos $F = (S, \overline{S})$ es un s–t corte. Sin embargo, por la definición de S, $\phi(a) = c(a)$ para todo $a \in F$ y $\phi(a) = 0$ para todo $a \in \overline{F}$. Por el lema 8.4, F tiene capacidad mínima. □

Ejercicio 8.6. Completar la comprobación que el flujo ϕ' cumple la condición 8.2.

La demostración anterior proporciona un método de obtención de un flujo máximo en una red de transporte. El algoritmo, también debido a Ford y Fulkerson, construye recursivamente y comenzando con un flujo dado (por ejemplo, el flujo nulo) una secuencia de flujos que acaba en un flujo máximo.

Entrada: $X = (G, s, t, c)$: una red de transporte.
Algoritmo FORD Y FULKERSON

1. Etiquetar s con $(-, \infty)$.
2. Repetir los pasos **3** y **4** mientras se pueda o hasta que t quede etiquetado.
3. **Si** v es un vértice no etiquetado adyacente desde un vértice etiquetado u
y $\phi(u,v) < c(u,v)$, **entonces** etiquetar v con $(u^+, \varepsilon(v))$
donde $\varepsilon(v) = \min\{\varepsilon(u), c(u,v) - \phi(u,v)\}$.
4. **Si** v es un vértice no etiquetado adyacente hacia un vértice etiquetado u
y $\phi(v,u) > 0$, **entonces** etiquetar v con $(u^-, \varepsilon(v))$
donde $\varepsilon(v) = \min\{\varepsilon(u), \phi(v,u)\}$.
5. **Si** t ha quedado etiquetado, **entonces**, volviendo hacia atrás a partir de t, se
encuentra una secuencia de aumento P: $s = u_0, u_1, \ldots, u_{n-1}, u_n = t$
donde, para $1 \leqslant k \leqslant n$, u_k está etiquetado $(u_{k-1}^+, \varepsilon(u_k))$ si $(u_{k-1}, u_k) \in A(P)$
y u_k está etiquetado $(u_{k-1}^-, \varepsilon(u_k))$ si $(u_k, u_{k-1}) \in A(P)$.
5.1. En el primer caso, cambiar $\phi(u_{k-1}, u_k)$ por $\phi(u_{k-1}, u_k) + \varepsilon(t)$.
5.2. En el segundo caso, cambiar $\phi(u_k, u_{k-1})$ por $\phi(u_k, u_{k-1}) - \varepsilon(t)$.
5.3. Borrar las etiquetas y volver al paso **1**.
6. Si t no ha quedado etiquetado, entonces ϕ es un flujo máximo.

Salida: ϕ: un flujo máximo.

Ejercicio 8.7. Aplicar el algoritmo de Ford y Fulkerson a la red de transporte de la figura 8.1.

Las definiciones de capacidad y flujo pueden ser generalizadas para permitir valores reales no negativos. En este caso, el teorema del flujo máximo–corte mínimo es aún válido, pero puede ocurrir que el algoritmo descrito no converja hacia un flujo máximo.

En algunas aplicaciones, en lugar de considerar un único vértice s y un único vértice t, conviene considerar conjuntos disyuntos S y T y definir ahora el valor del flujo como

$$\mathrm{val}(\phi) = \sum_{a \in (S,\overline{S})} \phi(a) - \sum_{a \in (\overline{S},S)} \phi(a) \tag{8.9}$$

Naturalmente, la condición de equilibrio 8.2 sólo se tiene que cumplir ahora para los vértices del conjunto $V(G) - (S \cup T)$. Esta situación se reduce a la aquí considerada si añadimos dos nuevos vértices s, t de tal forma que s sea adyacente hacia cada vértice x de S, t sea adyacente desde cada vértice y de T y, además, $c(s,x) = M$, para cada $x \in S$ y $c(y,t) = M$ para cada $y \in T$, siendo M un entero suficientemente grande. Se deja como ejercicio que el lector complete los detalles.

Ejercicio 8.8. Demostrar que una forma equivalente de la expresión 8.9 es:

$$\mathrm{val}(\phi) = \sum_{a \in (\overline{T},T)} \phi(a) - \sum_{a \in (T,\overline{T})} \phi(a)$$

8.3 Conectividad

Sea $G \neq K_n$ un grafo no dirigido. La *conectividad* $\kappa(G)$ de G es el mínimo número de elementos de un conjunto $S \subset V(G)$ tal que $G - S$ es no conexo. Por ejemplo, si G es conexo y tiene un vértice de corte, entonces $\kappa(G) = 1$. Se excluyen los grafos completos de esta definición, dado que K_n es el único grafo de orden n que no puede ser desconectado eliminando vértices. Como la supresión de $n-1$ vértices cualesquiera reduce K_n al grafo trivial K_1 constituido por un único vértice, se define, en este caso, $\kappa(K_n) = n - 1$.

Diremos también que G es k–conexo si $\kappa(G) \geqslant k$. Así, si G es k–conexo, entonces $G = K_{k+1}$ o G tiene al menos $k+2$ vértices y ningún subconjunto S de menos de k vértices lo deconecta (es decir, $G - S$ es aún conexo).

Análogamente, la *arista-conectividad* $\lambda(G)$ de un grafo $G \neq K_1$ es la mínima cardinalidad de un conjunto $F \subset E(G)$ tal que $G - F$ es no conexo. Por definición, $\lambda(K_1) = 0$. Como en el caso de vértices, G es k–arista-conexo si $\lambda(G) \geqslant k$.

Por ejemplo, el grafo de la figura 8.4 tiene conectividad 2 y arista-conectividad 3.

Tal como se ha dicho en la introducción del capítulo, cuando el grafo que se considera modela una red de interconexión, las conectividades constituyen medidas de la vulnerabilidad de la red ante el fallo de nodos y/o enlaces.

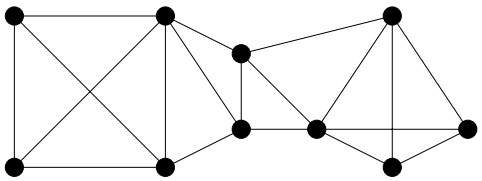


Figura 8.4: $\kappa(G) = 2, \lambda(G) = 3$

La eliminación de una arista uv no supone la eliminación de sus vértices terminales u y v. En cambio, si $u \in V(G)$, entonces el subgrafo $G - u$ no contiene ni el vértice suprimido u ni las aristas incidentes con este vértice. Estas consideraciones justifican la desigualdad $\kappa(G) \leqslant \lambda(G)$. Por otra parte, si u es un vértice con grado $\delta = \delta(G)$ (el grado mínimo del grafo), la eliminación de las δ aristas incidentes con u dejan u desconectado del resto del grafo. Por tanto, $\lambda(G) \leqslant \delta(G)$. Así, se tiene el siguiente teorema, la demostración del cual dejamos como ejercicio.

Teorema 8.9. Para todo grafo G,

$$\kappa(G) \leqslant \lambda(G) \leqslant \delta(G)$$

Cuando G es 1–conexo, existe un camino entre cada par de vértices del grafo. La extensión de este resultado a grafos n–conexos, $n \geqslant 1$, constituye el teorema de Whitney. Consideremos antes, sin embargo, dos resultados clásicos sobre conectividad: los teoremas de Menger.

8.4 Los teoremas de Menger

Sean u y v vértices no adyacentes de un grafo G. Se dice que $S \subset V(G)$ es un conjunto u–v *separador* si $G - S$ es no conexo y u y v pertenecen a componentes distintos de $G - S$. Es decir, S es un conjunto u–v separador, $u, v \in V(G) \setminus S$, si y sólo si cada camino entre u y v contiene algún elemento de S.

Análogamente, dados dos vértices cualesquiera de G, $F \subset E(G)$ es un conjunto u–v arista-separador si $G - F$ es no conexo y u y v pertenecen a componentes distintos de $G - F$.

Dados dos caminos entre u y v, diremos que son *internamente disyuntos* si los únicos vértices que tienen en común son precisamente los vértices terminales u y v. De forma similar, dos caminos entre u y v son arista-disyuntos si no tienen aristas en común.

Los teoremas de Menger relacionan el número mínimo de elementos que tiene que tener un conjunto que separe dos vértices dados con el número máximo de caminos disyuntos entre

estos vértices. La demostración que presentaremos usará el teorema del flujo máximo–corte mínimo. Comenzamos enunciando el teorema relativo a aristas:

Teorema 8.10. Sean u y v vértices de G. El número mínimo de aristas de un conjunto u–v arista-separador es igual al número máximo de caminos arista-disyuntos entre u y v.

Demostración. Podemos suponer que G es conexo. Sea m el número máximo de caminos arista-disyuntos entre u y v y sea n el número mínimo de aristas de un conjunto u–v arista-separador. Claramente,

$$n \geqslant m \tag{8.10}$$

Asociamos a G una cierta red de transporte $X = (G^*, u, v, c)$. El digrafo G^* correspondiente a X es el digrafo simétrico asociado a G y se obtiene a partir de G de la forma siguiente: $V(G^*) = V(G)$, y si $xy \in E(G)$ entonces $(x,y), (y,x) \in A(G^*)$. Ver la figura 8.5. La función

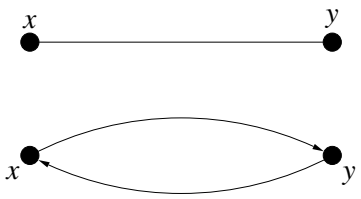


Figura 8.5: Obtención del digrafo G^* asociado a G

capacidad c toma el valor $c(a) = 1$ para todo $a \in A(G^*)$.

Si F es un u–v corte en G^*, entonces las aristas correspondientes constituyen en G un conjunto u–v arista-separador. Por tanto,

$$n \leqslant |F| = c(F) \tag{8.11}$$

En G^* existen m caminos dirigidos (sin arcos comunes) de u hacia v. Si ψ es una función que asigna valor 1 a cada arco de estos caminos dirigidos y asigna valor 0 a los arcos restantes de G^*, entonces ψ es un flujo en X de valor m. Así, si ϕ es un flujo máximo en X, entonces $\mathrm{val}(\phi) \geqslant m$. Por otra parte, sea ϕ un flujo máximo. Para cada $a \in A(G^*)$, $\phi(a)$ vale 1 o 0. Sea D el subgrafo dirigido de G^* definido por los arcos a tales que $\phi(a) = 1$. En D debe existir al menos un camino desde u hasta v (¿por qué?). Como ϕ es un flujo, la condición 8.2 implica $d^+(w) = d^-(w)$ para todo $w \in V(D) - \{u, v\}$, y, por la proposición 8.2, se cumple $d^+(u) - d^-(u) = d^-(v) - d^+(v) = \mathrm{val}(\phi)$, donde los grados indicados se consideran en D.

Estas condiciones implican la existencia en D, y por tanto en G^*, de $\mathrm{val}(\phi)$ caminos desde u hasta v disyuntos en arcos (demostrarlo como ejercicio; ver el teorema 7.5). Así, $m \geqslant \mathrm{val}(\phi)$ y, por tanto,

$$m = \mathrm{val}(\phi) \tag{8.12}$$

Ahora, considerando en la red un corte mínimo F, teniendo en cuenta 8.11 y 8.12, y aplicando el teorema de flujo máximo–corte mínimo, se tiene $n \leqslant c(F) = \mathrm{val}(\phi) = m$. Así, por 8.10, $n = m$. □

Aplicando este teorema y teniendo en cuenta la definición de k–arista-conectividad se obtiene el corolario siguiente:

Corolario 8.11. Un grafo G ($\neq K_1$) es k–arista-conexo si y sólo si entre cada par de vértices de G existen al menos k caminos arista-disyuntos.

El teorema de Menger para vértices es el siguiente:

Teorema 8.12. Sean u y v vértices no adyacentes de G. El número mínimo de vértices de un conjunto u–v separador es igual al número máximo de caminos internamente disyuntos entre u y v.

Demostración. Podemos suponer que G es conexo. Sea m el número máximo de caminos disyuntos entre u y v y sea n el número mínimo de vértices de un conjunto u–v separador. Claramente,

$$n \geqslant m \tag{8.13}$$

Sea G^* el digrafo simétrico asociado a G. Sea ahora G' el digrafo obtenido a partir de G^* mediante el procedimiento siguiente:

1. Remplazamos cada vértice $x \in V(G^*) - \{u, v\}$ por el par de vértices x', x'' junto con el arco (x', x'').

2. Si $(x, y) \in A(G^*)$, con x, y diferentes de u, v, entonces reemplazamos el arco (x, y) por el arco (x'', y').

3. Si $(u, x) \in A(G^*)$, con $x \neq v$, entonces reemplazamos (u, x) por (u, x'). Análogamente, si $(x, u) \in A(G^*)$, con $x \neq v$, entonces reemplazamos (x, u) por (x'', u).

4. Si $(x, v) \in A(G^*)$, con $x \neq u$, entonces reemplazamos (x, v) por (x'', v). Finalmente, si $(v, x) \in A(G^*)$, con $x \neq u$, entonces reemplazamos (v, x) por (v, x').

Cada camino P entre u y v en el grafo G corresponde a un camino dirigido P^* desde u hasta v en el digrafo G^* y corresponde, también, a un camino dirigido P' desde u hasta v en G', donde P' se obtiene de P^* reemplazando cada vértice interno w de P^* por el arco (w', w''). Recíprocamente, cada camino dirigido desde u hasta v en G' corresponde a un camino dirigido en G^* obtenido contrayendo los arcos de la forma (w', w''), y corresponde también, por tanto, a un camino entre u y v en G. Además, dos caminos P_1 y P_2 en G entre u y v son internamente disyuntos si y sólo si los correspondientes P_1' y P_2' en G' no tienen arcos en común. Por tanto, el número máximo de caminos disyuntos en G entre u y v es también el número máximo de caminos dirigidos de u a v en G' que no tienen arcos en común.

Si ϕ es un flujo máximo en la red de transporte $X = (G', u, v, c)$, donde la función capacidad c toma el valor $c(a) = 1$ para todo $a \in A(G')$, entonces, tal como se ha establecido en la demostración del teorema de Menger para aristas, $\mathrm{val}(\phi) = m$.

Por otra parte, si F es un corte mínimo en X, los arcos $a \in A(G')$ que forman F son de la forma $a = (u, x')$, $a = (x', x'')$, $a = (y'', x')$ o $a = (x'', v)$, donde x, y son vértices distintos de u, v (en G). En cualquier caso, asociamos al arco a el vértice $v_a = x \in V(G)$. Sea $V_a = \{v_a \mid a \in F\} \subset V(G)$. Claramente, $|V_a| \leqslant |F|$. Además, fácilmente se comprueba que V_a constituye, en G, un conjunto u–v separador. Por tanto, $n \leqslant |V_a| \leqslant |F| = c(F)$.

Ahora, aplicando otra vez el teorema del flujo máximo–corte mínimo, se tiene $n \leqslant c(F) = \mathrm{val}(\phi) = m$. Así, por 8.13, $n = m$. □

Ejercicio 8.13. Determinar los digrafos G^* y G' de la demostración anterior que corresponden al grafo G de la figura 8.4.

Usando este teorema cuando u y v son vértices no adyacentes de G y considerando también el caso $uv \in E(G)$, se puede demostrar el resultado siguiente, llamado teorema de Whitney, que constituye la formulación para vértices del resultado dado por el corolario 8.11.

Teorema 8.14. Un grafo G con $n \geqslant k+1$ vértices es k–conexo si y sólo si entre cada dos vértices distintos de G existen al menos k caminos internamente disyuntos.

8.5 Apareamientos en grafos bipartitos

Un *apareamiento* M en un grafo $G = (V, E)$ es un subconjunto de *aristas independientes*, es decir, si $uv, st \in M$, entonces $\{u, v\} \cap \{s, t\} = \emptyset$. Si $uv \in M$, diremos que u y v son *vértices apareados*. Por ejemplo, en la figura 8.6, $M = \{12, 34\}$ es un apareamiento.

Un *apareamiento perfecto* es un apareamiento en el cual todo vértice del grafo es incidente con alguna arista del apareamiento. No todos el grafos contienen un apareamiento perfecto (por ejemplo, los que tienen un número impar de vértices como sucede en la figura 8.6).

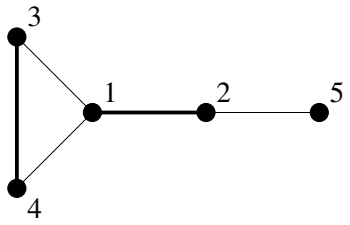


Figura 8.6: Apareamiento

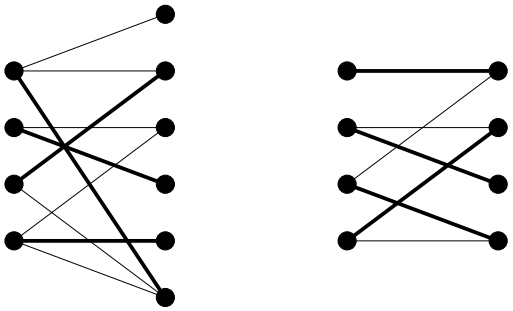


Figura 8.7: Apareamientos completo y perfecto

Dado que, en general, las cuestiones de apareamientos se aplican a resolver problemas de asignaciones, lo que quiere decir que el grafo a considerar es bipartito, nos centraremos en este tipo de grafos.

Sea M un apareamiento en un grafo bipartito $G = (U \cup W, E)$. Si M aparea todos los vértices de U con vértices de W, se dice que M es un *apareamiento completo* de U en W. Por tanto, si M es un apareamiento completo, $|M| = |U|$. Cuando $|U| = |W|$, un apareamiento completo también es perfecto. Los ejemplos de la figura 8.7 muestran un apareamiento completo y un apareamiento perfecto en grafos bipartitos. Sea $S \subset U$, entonces indicamos por $\Gamma(S)$ el conjunto de vértices de W que son adyacentes con algún vértice de S. Se llama defecto de S, y lo denotaremos por $\eta(S)$, a $|S| - |\Gamma(S)|$. El *defecto* del grafo bipartito G es $\eta(G) = \max\{\eta(S) \mid S \subset U\}$. Como $\eta(\emptyset) = |\emptyset| - |\Gamma(\emptyset)| = 0$, se tiene $\eta(G) \geqslant 0$.

8.6 El teorema de Hall

Volviendo a la cuestión de la existencia de un apareamiento completo en un grafo bipartito $G = (U \cup W, E)$, una condición necesaria para su existencia es claramente que G tenga defecto nulo, es decir, que para todo subconjunto S de U se cumpla $|S| \leqslant |\Gamma(S)|$. Curiosamente,

esta condición es también suficiente. Esto es lo que afirma el llamado teorema de Hall. Para demostrar este resultado aplicaremos otra vez el teorema del flujo máximo–corte mínimo.

Teorema 8.15. Un grafo bipartito $G = (U \cup W, E)$ admite un apareamiento completo de U en W si y sólo si para todo $S \subset U$ se verifica $|S| \leqslant |\Gamma(S)|$.

Demostración. Demostramos sólo la suficiencia. Supongamos, por tanto, que para todo $S \subset U$ se verifica $|S| \leqslant |\Gamma(S)|$.

Sea G' el digrafo con conjunto de vértices $V(G') = V(G) \cup \{s,t\}$, donde s y t son dos nuevos vértices añadidos a $V(G)$, y conjunto de arcos $A(G') = A_s \cup A_{UV} \cup A_t$, donde $A_s = \{(s,u) \mid u \in U\}$, $A_{UV} = \{(u,w) \mid uw \in E\}$ y $A_t = \{(w,t) \mid w \in W\}$. Sea X la red de transporte (G', s, t, c) donde $c(a) = 1$ para todo $a \in A_s \cup A_t$ y $c(a) = M$ para todo $a \in A_{UV}$, siendo $M > |U|$.

Sea F un s–t corte en X. Si F contiene algún arco de A_{UV}, entonces $c(F) \geqslant M > |U|$. De otro modo, el corte F debe tener la estructura $F = (P, \overline{P})$ con $P = \{s\} \cup S \cup T$, $S \subset U$, $T \subset W$ y $\Gamma(S) \subset T$. Naturalmente, $\overline{P} = (U \setminus S) \cup (W \setminus T) \cup \{t\}$. Pero, en este caso:

$$c(F) = |U \setminus S| + |T| = |U| - |S| + |T| \geqslant |U| - (|S| - |\Gamma(S)|) \geqslant |U|$$

Así, cualquier s–t corte tiene capacidad más grande o igual a $|U|$. Por otra parte, A_s constituye un corte mínimo con capacidad $|U|$. Por tanto, si ϕ es un flujo máximo en X, por el teorema del flujo máximo–corte mínimo, su valor es $\mathrm{val}(\phi) = |U|$. Pero un flujo de valor $|U|$ en X corresponde a un apareamiento completo de U en W. □

Observemos que la necesidad de la condición del teorema de Hall también se puede demostrar usando la red de transporte X considerada. En efecto, si $|S| > |\Gamma(S)|$ para cierto $S \subset U$, entonces considerando el s–t corte $F = (P, \overline{P})$ con $P = \{s\} \cup S \cup \Gamma(S)$ se tendría

$$c(F) = |U \setminus S| + |\Gamma(S)| = |U| - |S| + |\Gamma(S)| < |U|$$

Así, un flujo máximo en X tiene un valor inferior a $|U|$ y no es posible un apareamiento completo.

Ejercicio 8.16. (Problema de los matrimonios.) Sean un grupo de muchachos y un grupo de muchachas tales que a cada muchacha le agrada alguno de los muchachos. ¿Cuáles son las condiciones para que todas las muchachas se puedan casar con un muchacho que les agrade?

La condición $|S| \leqslant |\Gamma(S)|$ puede ser difícil de comprobar. El resultado que se presenta a continuación proporciona una condición suficiente para la existencia de un apareamiento completo.

Corolario 8.17. Sea $G = (U \cup W, E)$ un grafo bipartito. Si existe un entero $k > 0$ tal que $d(u) \geqslant k \geqslant d(w)$ para todo vértice $u \in U$ y todo vértice $w \in W$, entonces existe un apareamiento completo de U en W.

Demostración. Consideremos un subconjunto $S \subset U$. Como el conjunto de aristas $E_{\Gamma(S)}$ incidentes con vértices de $\Gamma(S)$ contiene el conjunto de aristas E_S incidentes con vértices de S, tenemos la relación

$$k|\Gamma(S)| \geqslant |E_{\Gamma(S)}| \geqslant |E_S| \geqslant k|S|$$

de donde se obtiene la condición suficiente del teorema de Hall. □

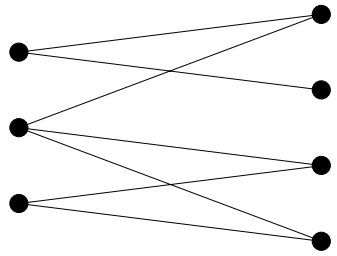


Figura 8.8: Existencia de un apareamiento completo

El resultado anterior se aplica, en particular, cuando el grafo es k–regular.

Ejercicio 8.18. Aplicar el resultado anterior al grafo de la figura 8.8 y encontrar un apareamiento completo.

Hay muchos problemas de optimización combinatoria que pueden ser tratados gracias al teorema de Hall. Uno de ellos, el que dio lugar al estudio de Hall, se formula de la forma siguiente: dada una colección de conjuntos $S_1, S_2, \ldots, S_n$ no vacíos encontrar lo que se llama un *sistema de representantes diferentes*, o *transversal*, es decir, un conjunto $\{s_1, s_2, \ldots, s_n\}$ tal que $s_i \in S_i$, $1 \leqslant i \leqslant n$, y $s_i \neq s_j$ si $i \neq j$.

Muchas veces, a pesar de no existir un apareamiento completo, interesa encontrar un apareamiento de cardinalidad máxima. A continuación se demuestra un resultado que permite saber cuál es la cardinalidad de este apareamiento a partir del defecto del grafo.

Teorema 8.19. Dado un grafo bipartito $G = (U \cup W, E)$, el máximo número de vértices de U que se pueden aparear con vértices de W es $|U| - \eta(G)$ y este apareamiento existe.

Demostración. La demostración es similar a la del teorema de Hall. Se contruye una red de transporte X de la misma manera y se considera un s–t corte F. Si F contiene algún arco de A_{UV}, entonces $c(F) \geqslant M > |U| \geqslant |U| - \eta(G)$. De otro modo, el corte F debe tener la estructura $F = (P, \overline{P})$ con $P = \{s\} \cup S \cup T$, $S \subset U$, $T \subset W$ y $\Gamma(S) \subset T$. Pero, también en este caso:

$$c(F) = |U| - |S| + |T| \geqslant |U| - (|S| - |\Gamma(S)|) = |U| - \eta(S) \geqslant |U| - \eta(G)$$

Así, cualquier s–t corte tiene capacidad más grande o igual a $|U| - \eta(G)$.

Ahora, sea $S \subset U$ tal que $\eta(S) = \eta(G)$. Si $P = \{s\} \cup S \cup \Gamma(S)$, entonces $F = (P, \overline{P})$ es un corte mínimo, ya que $c(F) = |U| - |S| + |\Gamma(S)| = |U| - \eta(S) = |U| - \eta(G)$.

Así, un flujo máximo en X, que corresponde a un apareamiento de cardinalidad máxima en G, tiene valor $|U| - \eta(G)$. □

Notas bibliográficas

La referencia para el teorema de flujo máximo–corte mínimo, debido a Ford y Fulkerson, es [5]. También es preciso mencionar el libro [6] de estos mismos autores. Una mejora del algoritmo de Ford y Fulkerson, debida a Edmonds y Karp, se encuentra en [4]. Resultados recientes sobre flujos en redes se dan en [7].

La relación entre conectividad, arista-conectividad y grado mínimo, dada por el teorema 8.9, es debida a Whitney [10]. El libro de Chartrand y Lesniak [2] contiene una excelente exposición del tema de conectividad. Para ver las aplicaciones al estudio de la vulnerabilidad de una red de interconexión se puede consultar [1].

Finalmente, la referencia para el teorema de Hall es [8]. Un buen algoritmo para encontrar apareamientos en grafos bipartitos se da en [3]. En [9] se describe una aplicación interesante de este teorema al problema de optimización del número de conexiones de redes multibus para sistemas multiprocesadores.

Bibliografía

[1] J.–C. Bermond, N. Homobono, C. Peyrat. "Large fault-tolerant interconnection networks", *Graphs and Combinatorics*, **5**, pp. 107–123, 1989.

[2] G. Chartrand, L. Lesniak. *Graphs and Digraphs*, Wadsworth & Brooks, 1986.

[3] J. Edmonds. "Paths, trees and flowers", *Canad. J. Math.*, **17**, pp. 449–467, 1965.

[4] J. Edmonds, R. M. Karp. "Theoretic improvements in algorithmic efficiency for network flow problems", *J. Assoc. Comput. Mach.*, **19**, pp. 248–264, 1972.

[5] L. R. Ford, D. R. Fulkerson. "Maximal flow through a network", *Canad. J. Math.*, **8**, pp. 399–404, 1956.

[6] L. R. Ford, D. R. Fulkerson. *Flows in networks*, Princeton University Press, 1962.

[7] A. V. Goldberg, E. Tardos, R. E. Tarjan. "Network flow algorithms", *Tech. Rep.*, CS–TR–216–89, Princeton University.

[8] P. Hall. "On representatives of subsets", *J. London Math. Soc.*, **10**, pp. 26–30, 1935.

[9] T. Lang, M. Valero, M. A. Fiol. "Reduction of connections for multibus organization", *IEEE Trans. Computers*, **C-32**, n. 8, pp. 707–716, 1983.

[10] H. Whitney. "Congruent graphs and the connectivity of graphs", *Amer. J. Math.*, **54**, pp. 150–168, 1932.

Problemas

1. Considerar la red de transporte $X = (G, s, t, c)$ con conjunto de vértices $V(G) = \{s, 1, 2, 3, 4, 5, t\}$ y con arcos y capacidades de los arcos dados por la tabla siguiente:

Arco	$(u,1)$	$(2,u)$	$(u,3)$	$(1,2)$	$(2,3)$	$(1,4)$
Capacidad	5	4	4	3	2	4

Arco	$(1,5)$	$(3,5)$	$(5,4)$	$(4,t)$	$(5,t)$
Capacidad	2	3	2	3	4

   (a) Determinar un corte de capacidad mínima.

   (b) Aplicando el algoritmo de Ford y Fulkerson, determinar un flujo máximo.

2. Sea ϕ un flujo máximo y F un corte mínimo en una red de transporte. Demostrar que $\phi(a) = c(a)$ para todo $a \in F$ y $\phi(a) = 0$ para todo $a \in \overline{F}$.

3. Demostrar el teorema 8.9: para todo grafo G, $\kappa(G) \leqslant \lambda(G) \leqslant \delta(G)$.

4. Determinar un grafo G con $\kappa(G) = 3$, $\lambda(G) = 4$ y $\delta(G) = 5$.

5. Un grafo no trivial conexo y sin vértices de corte se llama *bloque*. Un *bloque de un grafo* G es un subgrafo de G que es bloque y que es maximal respecto de esta propiedad. Demostrar que:

   (a) un bloque de G es un subgrafo inducido de G;

   (b) dos bloques de G distintos tienen como mucho un vértice en común que es un vértice de corte de G;

   (c) G se puede expresar como unión de sus bloques.

6. Demostrar el resultado siguiente usado en la demostración del teorema 8.10: Sea D un digrafo conexo con vértices u y v tales que $d^+(u) = d^-(u) + k$, $d^-(v) = d^+(v) + k$, donde k es un entero positivo, y tal que $d^+(w) = d^-(w)$ para todo $w \in V(D) - \{u, v\}$. Entonces existen k caminos desde u hasta v que son disyuntos en arcos.

7. Demostrar que en todo grafo 2-conexo dos vértices cualesquiera pertenecen a un ciclo común.

8. Comprobar los teoremas de Menger en el grafo de Petersen P. ¿Cuál es el valor de $\lambda(P)$ y $\kappa(P)$?

9. Sea G un grafo y sean U y W subconjuntos disyuntos de $V(G)$. Un camino de la forma $u, z_1, \ldots, z_n, v$ con $u \in U$ y $v \in V$, diremos que es un camino U–V. También diremos que $S \subset V(G)$ es un conjunto U–V separador si $G - S$ no contiene ningún camino entre vértices de U y vértices de W. Demostrar la siguiente generalización del teorema de Menger: El número mínimo de vértices de un conjunto U–V separador es igual al número máximo de caminos U–V que son disyuntos en vértices.

10. Enunciar y demostrar los teoremas de Menger para digrafos.

11. Sea G un grafo con n vértices, conectividad $\kappa \geqslant 1$ y diámetro D. Demostrar que $n \geqslant \kappa(D-1)+2$.

12. Determinar un apareamiento completo en el hipercubo Q_n.

13. Sea S la colección de conjuntos $\{3,5\}$, $\{3,4\}$, $\{1,3\}$, $\{1,5\}$ y $\{2,4\}$. Encontrar un transversal para S asociando a S un determinado grafo bipartito y encontrando un apareamiento completo en este grafo.

14. Formular una condición necesaria y suficiente para la existencia de un transversal para la colección de conjuntos no vacíos $S_1, S_2, \ldots, S_n$.

15. Un 1–*factor* de un grafo G es un subgrafo generador de G que es 1-regular.

    (a) Demostrar que G contiene un 1-factor si y sólo si G contiene un apareamiento perfecto;

    (b) demostrar que todo grafo bipartito regular contiene un 1-factor.

16. Un grafo G se dice *k–arista-colorable* si existe una aplicación c entre $E(G)$ y un conjunto C de k elementos, llamados *colores*, tal que las aristas incidentes con un mismo vértice tienen colores distintos asignados. Demostrar que si G es bipartito, entonces el mínimo valor de k tal que G es k–colorable es $\Delta(G)$.

    Indicación: demostrar primero que G es subgrafo de un subgrafo bipartito Δ–regular.

Parte III Estructuras algebraicas

Las estructuras algebraicas constituyen una de las herramientas básicas para tratar la mayor parte de los problemas asociados a la matemática discreta. En este libro, hemos pretendido dar las primeras nociones algebraicas para poder tratar, a título de ejemplo, algunos de los problemas más conocidos dentro de este ámbito.

En esta última parte, como se ha hecho en el resto del libro, todos los conjuntos a los cuales nos referiremos serán conjuntos finitos o numerables. El primer capítulo de esta parte introduce los conceptos de relaciones, aplicaciones, operaciones, y se acaba presentando las estructuras algebraicas que se tratarán en los capítulos siguientes. Algunas de las nociones que se consideran corresponden a una formación básica y, por tanto, han sido ya utilizadas a lo largo del libro, pero por razones de coherencia formal se ha considerado conveniente reconsiderarlas y agruparlas ordenadamente en su contexto original. El capítulo siguiente está dedicado al estudio de los grupos como modelo más completo de estructura definida a partir de una operación. El tercer capítulo trata las estructuras algebraicas con dos operaciones: anillos y cuerpos. Además de su interés intrínseco como estructuras discretas, los grupos, anillos y cuerpos ofrecen una variedad considerable de aplicaciones. Por ejemplo, la teoría de enumeración de Pólya se estudia al final del capítulo de grupos y en el último capítulo se presentan algunas aplicaciones que constituyen tradicionalmente temas propios de la matemática discreta: diseños combinatorios, geometrías finitas y cuadrados latinos.

Capítulo 9

Introducción a las estructuras algebraicas

1. Relaciones
2. Aplicaciones
3. Operaciones
4. Estructuras algebraicas

La base que fundamenta esta última parte descansa sobre la noción elemental de correspondencia o relación. Como caso particular, aparece el concepto de aplicación a partir del cual se obtiene la noción de operación que abrirá las puertas que conducen al mundo de las estructuras algebraicas. Éste es, por tanto, un capítulo introductorio que nos permitirá definir las bases que se desarrollarán en los capítulos siguientes, dedicados al estudio de las estructuras algebraicas más relevantes.

9.1 Relaciones

Comenzaremos considerando el conjunto formado por todos los posibles pares ordenados formados a partir de los elementos de dos conjuntos. Así pues, dados dos conjuntos A y B, su *producto cartesiano* se define como

$$A \times B = \{(a,b) \mid a \in A,\, b \in B\}$$

El hecho de que se trate de pares ordenados hace que $A \times B \neq B \times A$ si $A \neq B$.

Estos pares ordenados se pueden interpretar como relaciones entre los elementos de un conjunto con los del otro. Esta interpretación conduce al concepto siguiente, básico en toda esta parte.

Dados dos conjuntos A y B, se llama *correspondencia* o *relación* de A a B a cualquier subconjunto R del producto cartesiano $A \times B$.

Como ejemplo ilustrativo de esta definición podemos considerar la siguiente relación entre los conjuntos $A = \{a, b\}$ y $B = \{1, 2, 3\}$:

$$R = \{(a, 1), (a, 2), (b, 2)\}$$

Para visualizar estas relaciones es útil usar el grafo de la relación. Éste es un digrafo bipartito $(A \cup B, R)$, que tiene A y B como partes estables y hay un arco de $a \in A$ hacia $b \in B$ si y sólo si $(a, b) \in R$.

La relación del ejemplo anterior se representaría con el grafo siguiente:

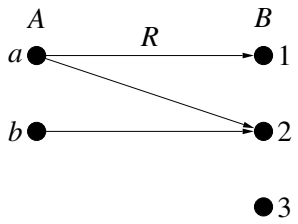


Figura 9.1: Grafo de la relación R

Si $A = B$ diremos que la relación es *binaria* sobre A. Si R es una relación binaria sobre A y $(a, a') \in R$ entonces diremos que a está relacionado con a' y lo denotaremos como aRa'.

En el caso que la relación sea binaria, el grafo que la representa tiene por vértices los elementos del conjunto A y los arcos quedan determinados por las relaciones, es decir, hay un arco de a hacia a' si y sólo si $(a, a') \in R$. Este grafo lo notaremos como,

$$G = (A, R)$$

Como ejemplos ilustrativos podemos considerar los grafos de las relaciones siguientes:

1. La relación $\{(1, 1), (1, 3), (2, 4)\}$ en el conjunto $A = \{1, 2, 3, 4\}$ se puede representar por el grafo de la figura 9.2.

2. El grafo de la relación "ser menor o igual que" en el conjunto $A = \{1, 2, 3, 4\}$ se puede representar como en la figura 9.3.

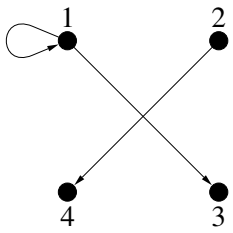


Figura 9.2:

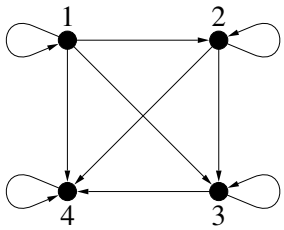


Figura 9.3:

Hay muchas relaciones que tienen en común propiedades con significación especial. Dada una relación R definida sobre un conjunto A, diremos que R es

- *reflexiva* si y sólo si para todo $a \in A$, aRa;
- *simétrica* si y sólo si para todo $a, b \in A$, $aRb \Rightarrow bRa$;
- *antisimétrica* si y sólo si para todo $a, b \in A$, aRb y $bRa \Rightarrow a = b$;
- *transitiva* si y sólo si para todo $a, b, c \in A$, aRb y $bRc \Rightarrow aRc$.

Es fácil construir ejemplos de relaciones que verifiquen algunas de estas propiedades:

1. Si sobre el conjunto de los seres humanos escogemos como relación "ser hermano de", podemos comprobar fácilmente que se verifican las propiedades simétrica y transitiva.
2. Si en el conjunto anterior consideramos la relación "ser estudiante de", ninguna de estas propiedades se cumplen en general.
3. Si la relación que escogemos en el mismo conjunto es "ser más alto que", se verifican sólo las dos últimas propiedades.

4. Si la relación considerada es "ser más alto o igual que", entonces se cumplen todas excepto la segunda de estas propiedades.

La agrupación de algunas de estas propiedades conduce a determinadas clases de relaciones que, por su interés, tienen un nombre propio que las representa.

Diremos que una relación binaria R definida sobre un conjunto A es de *orden* si es reflexiva, antisimétrica y transitiva.

Ejemplos inmediatos de relaciones de orden son los siguientes:

1. "Ser más pequeño" sobre el conjunto de los números naturales.

2. La inclusión no estricta es también una relación de orden sobre el conjunto de las partes de cualquier conjunto. En particular, si $A = \{a, b\}$ podemos representar esta relación con el grafo siguiente,

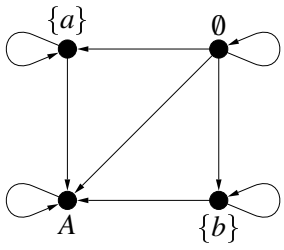


Figura 9.4: Grafo de una relación de orden

Para simplificar el grafismo podemos suprimir los autoenlaces que representan la reflexividad, así como también las aristas que se deducen de la transitividad. Siguiendo este criterio, la representación de la relación anterior es la que figura en 9.5.

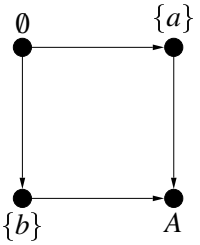


Figura 9.5: Grafo simplificado de una relación de orden

Es útil observar que el grafo de una relación de orden no puede contener ciclos (excepto autoenlaces). Esto quiere decir que hay pares de elementos no relacionados. En el ejemplo anterior, $\{a\}$ no está relacionado con $\{b\}$. Estas situaciones no se presentan si el orden es total.

Una relación de orden R sobre el conjunto A, se dice que es *total* si para todo $a, b \in A$, aRb o bRa. En caso contrario, se dice que el orden es *parcial*.

Así, el grafo que reprenta un orden total tiene que ser una "cadena".

Como ejemplos de relaciones de orden podemos considerar la relación de inclusión sobre las partes de un conjunto como relación de orden parcial. Por ejemplo, si $A = \{a, b\}$, $\{a\}$ y $\{b\}$ no están relacionados, mientras que la relación binaria "ser menor o igual" sobre los números naturales es una relación de orden total.

Otro tipo importante de relación, que conduce a la identificación de objetos equivalentes respecto a alguna propiedad común, es la siguiente.

Se dice que una relación R sobre un conjunto A es de *equivalencia* si es reflexiva, simétrica y transitiva.

Un ejemplo importante de este tipo de relación en el conjunto de los enteros es la llamada *relación congruencia módulo n*. Se dice que dos enteros x, y son congruentes módulo n si y sólo si $x - y$ es un múltiplo de n, es decir, $x = y + kn$ para algún entero k, y se denota normalmente por

$$x \equiv y \pmod{n}$$

Es fácil verificar que la relación de congruencia satisface las propiedades reflexiva, simétrica y transitiva.

Si R es una relación de equivalencia definida sobre A, llamamos *clase de equivalencia* de un elemento $a \in A$ al conjunto de elementos que están relacionados con a según R, y lo denotamos por $[a] = \{x \in A \mid xRa\}$.

Toda relación de equivalencia proporciona una clasificación o partición del conjunto original en subconjuntos que representan las clases de equivalencia originadas por medio de la relación.

Recordemos en primer lugar la definición de partición.

Una colección de subconjuntos propios de un conjunto A, $\{A_i\}_{i \in I}$, es una *partición* de A si y sólo si satisface las dos condiciones siguientes:

1. $\cup_{i \in I} A_i = A$.
2. $A_i \cap A_j = \phi, \quad \forall i, j \in I, \quad i \neq j$.

Proposición 9.1. Si R es una relación de equivalencia sobre un conjunto A, entonces la colección de clases de equivalencia $\{[a], a \in A\}$ es una partición de A.

Demostración.

1. La unión de clases de equivalencia es A. En primer lugar $[a] \subseteq A$, para todo $a \in A$ y, por tanto, $\cup_{a\in A}[a] \subseteq A$. También es cierta la inclusión contraria ya que para todo $a \in A$, $a \in [a]$ (como mínimo $[a]$ contiene a debido a la reflexividad de la relación). De aquí $A \subseteq \cup_{a\in A}[a]$.

2. Todas las clases son disyuntas. Es decir, si $[a] \cap [b] \neq \phi$, entonces es preciso ver que estas clases coinciden. Si suponemos que $x \in [a] \cap [b]$, esto significa que aRx y xRb y, por tanto, aRb. Entonces para todo $y \in [a]$, yRa y aRb implican yRb, de manera que $[a] \subset [b]$. De forma análoga se ve que $[b] \subset [a]$.

□

Observemos que la reflexividad de la relación nos permite demostrar que las clases de equivalencia cubren todo el conjunto A, mientras que la simetría y la transitividad nos garantizan que las clases son disyuntas.

De hecho, también es cierto que toda partición permite definir (de manera formal) una relación de equivalencia sobre el conjunto unión de estas partes. Para ver esto, si $\{A_i\}_{i\in I}$ es una partición del conjunto A, definimos la relación de equivalencia R de la forma siguiente: aRb si y sólo si a y b pertenecen a un mismo conjunto A_i de la partición. Es inmediato comprobar que esta relación verifica las tres propiedades que la hacen de equivalencia.

Hay dos ejemplos extremos (poco interesantes) de relaciones de equivalencia que siempre se pueden definir sobre un conjunto.

1. Uno es la *relación trivial* en la cual cada elemento sólo está relacionado con sí mismo. Con esta relación se obtienen tantas clases como elementos tiene el conjunto de partida y cada clase contiene sólo un elemento.

2. En el extremo opuesto podemos definir la *relación universal* en la cual cada elemento está relacionado con cualquier otro. Esta relación únicamente proporciona una clase de equivalencia que coincide con el propio conjunto de partida.

Un ejemplo no trivial de relación de equivalencia es el de congruencia módulo n en $\mathbb{Z}$. En particular,

1. Si $n = 2$ esta relación permite clasificar los enteros en dos subconjuntos, el de los números pares y el de los números impares, $\mathbb{Z} = [0] \cup [1]$.

2. Si tomamos $n = 3$, $\mathbb{Z}$ queda dividido en tres clases, $\mathbb{Z} = [0] \cup [1] \cup [2]$, donde

$$\begin{aligned} [0] &= \{0, \pm 3, \pm 6, \pm 9, \dots\}, \\ [1] &= \{1, 1 \pm 3, 1 \pm 6, 1 \pm 9, \dots\}, \\ [2] &= \{2, 2 \pm 3, 2 \pm 6, 2 \pm 9, \dots\} \end{aligned}$$

La partición de los elementos de un conjunto en clases de equivalencia permite considerar un nuevo conjunto (desde la perspectiva de la relación de equivalencia) con menos elementos, el constituido por sus clases de equivalencia, que formalmente definimos de la forma siguiente.

Dada una relación de equivalencia R sobre un conjunto A, el *conjunto cociente* de A módulo R es el conjunto que tiene por elementos las clases de equivalencia y lo notaremos como $A/R = \{[a] \mid a \in A\}$.

Sobre el conjunto de los enteros, las relaciones de congruencia de los ejemplos anteriores dan lugar a los siguientes conjuntos cociente:

1. Si R es la relación de congruencia módulo 2, entonces $\mathbb{Z}/R = \{[0], [1]\}$.

2. Si R es la relación de congruencia módulo 3, entonces $\mathbb{Z}/R = \{[0], [1], [2]\}$.

Las relaciones de equivalencia, aunque directamente no dan lugar a ningún tipo especial de construcción algebraica, son imprescindibles para trabajar a un cierto nivel con cualquiera de ellas.

9.2 Aplicaciones

Las aplicaciones o funciones discretas son un caso particular de relación o correspondencia entre dos conjuntos finitos o numerables, en la cual a cada elemento del primer conjunto le hacemos corresponder un único elemento del segundo conjunto. Este tipo de relación es una de las más utilizadas en todo lo referente a la matemática discreta.

De forma precisa, se dice que una relación f sobre el conjunto $X \times Y$ es una *aplicación* o función discreta si y sólo si para todo $x \in X$ existe un único $y \in Y$ tal que xfy. Si xfy o $(x, y) \in f$, se dice que f envía x a y y lo denotamos escribiendo $f(x) = y$. También se dice que y es la imagen de x por f, o bien, que x es una antiimagen de y.

El conjunto imagen de X a través de f es el subconjunto de Y sobre el que se envía algún elemento de X y habitualmente se denota como $f(X)$ o también como $Im\, f$. Es decir,

$$f(X) = Im\, f = \{y \in Y \mid \exists x \in X : f(x) = y\}$$

Se dice que X es el *dominio* de la aplicación f y se denota como $Dom f = X$. Se dice también que el *recorrido* de f es Y. Habitualmente, para expresar el dominio y el recorrido de una aplicación f, se utiliza la notación $f : X \longrightarrow Y$.

Si consideramos los conjuntos de números naturales o enteros, podemos definir las aplicaciones siguientes:

1. $f : \mathbb{N} \longrightarrow \mathbb{N},\ f(n) = n+1$
2. $f : \mathbb{N} \longrightarrow \mathbb{Z},\ f(n) = n-1$
3. $f : \mathbb{N} \longrightarrow \{1\},\ f(n) = 1$
4. $f : \mathbb{Z} \longrightarrow \mathbb{Z},\ f(z) = 3z-17.$

En el grafo de una aplicación $f : X \longrightarrow Y$, de cada vértice de X tiene que salir una única arista hacia algun vértice de Y. Los grafos siguientes representan algunas aplicaciones o funciones discretas.

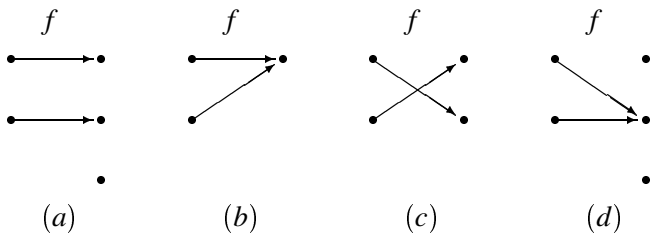


Figura 9.6: Grafos de aplicaciones

A menudo, para definir aplicaciones sobre conjuntos finitos, se especifica el valor que toma la aplicación sobre cada elemento de su dominio.

Tiene interés considerar, como se verá más adelante, la restricción de una aplicación a un subconjunto del dominio y también su relación inversa.

Dada una aplicación $f : X \longrightarrow Y$ y un subconjunto $X' \subset X$, una *restricción* de f sobre X' es una aplicación $f' : X' \longrightarrow Y$ que coincide con f si consideramos f restringida a X'. Lo denotamos como $f_{|X'} = f'$. También se dice que f es una *extensión* de f'.

La *relación inversa* de una aplicación $f : X \longrightarrow Y$ es el conjunto de pares ordenados $\{(y,x) \mid (x,y) \in f\}$ y se denota como f^{-1}.

Como sugiere la definición, el grafo de la inversa de una aplicación se obtiene invertiendo el sentido de los arcos en el grafo de la función original. La inversa de una función $f : X \longrightarrow Y$ es siempre una relación sobre $Y \times X$, pero no necesariamente es una aplicación, como se puede observar en la figura anterior.

Hay ciertos tipos de funciones que por su comportamiento reciben un nombre especial. Entre las más comunes se encuentran las siguientes.

Una aplicación $f : X \longrightarrow Y$ se llama

- *inyectiva* si y sólo si para todo $x, x' \in X$, si $x \neq x'$, entonces $f(x) \neq f(x')$.
- *exhaustiva* si y sólo si para todo $y \in Y$, existe $x \in X$ tal que $f(x) = y$.
- *biyectiva* si y sólo si es inyectiva y exhaustiva.

En la figura 9.6 hay representada en primer lugar una aplicación inyectiva, seguida de una exhaustiva y una biyectiva. La última de las aplicaciones queda fuera de esta clasificación.

Observar también que, si $f : X \longrightarrow Y$ es una biyección, entonces $f^{-1} : Y \longrightarrow X$ es una aplicación y es también biyectiva.

Una interpretación a veces útil de la clasificación anterior es la siguiente. Si $f : X \longrightarrow Y$ es una aplicación y $b \in Y$ es un valor arbitrario, entonces decir que

a) la solución de la ecuación $f(x) = b$, en caso de existir, es única es equivalente a decir que f es inyectiva;

b) la ecuación $f(x) = b$ admite solución en x es equivalente a decir que f es exhaustiva;

c) existe una única solución de la ecuación $f(x) = b$ es equivalente a decir que f es biyectiva.

La proposición siguiente dice que, en algunos casos, estas tres condiciones son equivalentes.

Proposición 9.2. Si X e Y son dos conjuntos finitos con el mismo número de elementos, entonces $f : X \longrightarrow Y$ es inyectiva si y sólo si f es exhaustiva.

Demostración. En general, si X e Y son finitos, es fácil ver que f es inyectiva si y sólo si $|X| = |f(X)|$ y f es exhaustiva si y sólo si $|f(X)| = |Y|$. Cuando $|X| = |Y|$, estas dos condiciones son equivalentes. □

Cabe observar que este resultado sólo es cierto si X e Y son finitos. Como ejemplo, si consideramos la aplicación $f : \mathbb{N} \longrightarrow \mathbb{N}$, $f(n) = 2n$, que envía los números naturales a los números pares, ésta es claramente una aplicación inyectiva, pero en cambio no es exhaustiva.

De la demostración de este resultado se deduce el *principio de Dirichlet* o también llamado *principio del palomar*, que ha sido introducido en el capítulo 5 y que, expresado en términos de esta proposición, dice que si $|X| > |Y|$, entonces algún elemento de Y tiene que tener más de una antiimagen.

Uno de los recursos más potentes para la obtención de nuevas funciones a partir de otras ya conocidas se obtiene a partir de la composición de funciones.

Dadas dos funciones $f : X \longrightarrow Y$ y $g : Y \longrightarrow Z$ se define la *composición* de f con g, que se denota como $g \circ f$, como aquella aplicación $g \circ f : X \longrightarrow Z$ tal que $(g \circ f)(x) = g(f(x))$.

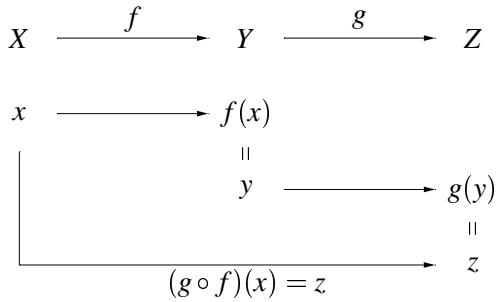


Figura 9.7: Composición de aplicaciones

A título de ejemplo ilustrativo se pueden considerar los grafos que figuran en 9.7, que

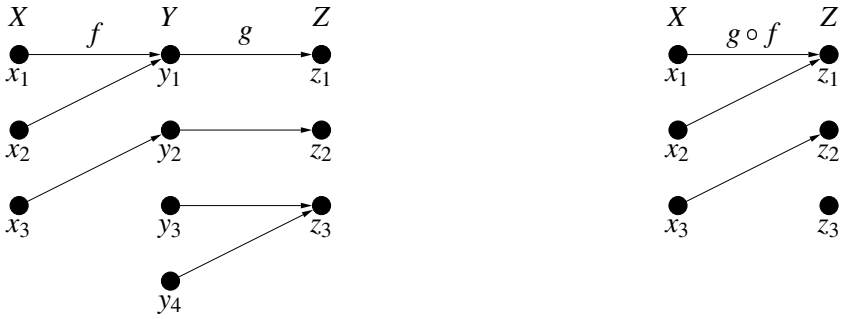


Figura 9.8: Grafo de una composición de aplicaciones

representan el grafo de la composición $g \circ f$ a partir de los grafos de las aplicaciones f y g.

Una cuestión interesante que se plantea constantemente en matemáticas es la de saber cuándo, al combinar dos entidades con propiedades comunes, se obtiene un resultado con la misma propiedad. Para la composición de aplicaciones tenemos el resultado siguiente:

Proposición 9.3. Dadas dos aplicaciones, $f : X \longrightarrow Y$ y $g : Y \longrightarrow Z$, se puede afirmar que:

a) si f, g son inyectivas, entonces $g \circ f$ es también inyectiva;

b) si f, g son exhaustivas, entonces $g \circ f$ es también exhaustiva.

Demostración. a) Para demostrar que $g \circ f$ es inyectiva, sean $x, x' \in X$ tales que $(g \circ f)(x) = (g \circ f)(x')$. Entonces, de $g(f(x)) = g(f(x'))$ se deduce que $f(x) = f(x')$ ya que g es inyectiva. Como f es también inyectiva deducimos que $x = x'$, y por tanto obtenemos la inyectividad de $g \circ f$.

b) Dado $z \in Z$, como g es exhaustiva, existe $y \in Y$ tal que $g(y) = z$. Por otra parte, como f es también exhaustiva, existe $x \in X$ tal que $f(x) = y$. Así, $z = g(y) = g(f(x)) = (g \circ f)(x)$ y, por tanto, $Im(g \circ f) = Z$. □

Cabe observar que, de esta proposición, se deduce directamente la biyección de la composición, siempre que las funciones originales sean biyectivas.

9.3 Operaciones

Las operaciones binarias más familiares son las operaciones aritméticas de la suma y el producto. Cada una de estas operaciones es una regla que asocia a cada par de números otro número bien definido. El concepto genérico de operación binaria es una generalización de esta idea.

Una *operación binaria*, a veces llamada también *ley de composición interna*, sobre un conjunto A es una aplicación de $A \times A$ sobre A.

$$\begin{array}{rccc} f: & A \times A & \longrightarrow & A \\ & (a,b) & \longrightarrow & f(a,b) \end{array}$$

Habitualmente, las aplicaciones que representan operaciones binarias se denotan mediante algún símbolo que une los elementos operados, por ejemplo,

$$f(a,b) = a \star b, \quad f(a,b) = a \perp b, \quad f(a,b) = a + b$$

Seguidamente damos los ejemplos de operaciones binarias más utilizadas.

1. Además de las operaciones aritméticas elementales, en el conjunto de los números naturales se pueden definir muchas otras operaciones, como por ejemplo

$$\begin{array}{ccc} \mathbb{N} \times \mathbb{N} & \longrightarrow & \mathbb{N} \\ (n,m) & \longrightarrow & n(m+1) \end{array}$$

2. La composición de aplicaciones definidas de un conjunto X en él mismo, que denotamos como $F(X)$, constituye un ejemplo importante de operación no aritmética.

$$\begin{array}{ccc} F(X) \times F(X) & \longrightarrow & F(X) \\ (f,g) & \longrightarrow & f \circ g \end{array}$$

3. Si consideramos $\wp(X)$, el conjunto de las partes del conjunto X, la unión y la intersección son dos ejemplos importantes de operaciones.

$$\begin{array}{rcl} \wp(\mathbf{X}) \times \wp(\mathbf{X}) & \longrightarrow & \wp(\mathbf{X}) \\ (A,B) & \longrightarrow & A \cup B \end{array} \qquad \begin{array}{rcl} \wp(\mathbf{X}) \times \wp(\mathbf{X}) & \longrightarrow & \wp(\mathbf{X}) \\ (A,B) & \longrightarrow & A \cap B \end{array}$$

Una estructura algebraica importante basada en estas operaciones es la llamada *de Boole*, que se define en el problema 7 del penúltimo capítulo.

Otros ejemplos importantes de operaciones aritméticas son la suma y el producto sobre enteros módulo n y constituyen lo que se llama *aritmética modular*. Estas operaciones se definen de manera natural, es decir, asignando a la suma de clases la clase de la suma y como producto de clases la clase del producto.

$$\begin{array}{rcl} \mathbb{Z}_n \times \mathbb{Z}_n & \longrightarrow & \mathbb{Z}_n \\ ([a],[b]) & \longrightarrow & [a]+[b]=[a+b] \end{array} \qquad \begin{array}{rcl} \mathbb{Z}_n \times \mathbb{Z}_n & \longrightarrow & \mathbb{Z}_n \\ ([a],[b]) & \longrightarrow & [a]\cdot[b])=[a\cdot b] \end{array}$$

Es preciso comprobar que estas operaciones están bien definidas, es decir, que no dependen de los representantes escogidos en cada clase. Ciertamente, si a, a' son dos representantes cualesquiera de la clase $[a]$ y b, b' dos de la clase $[b]$, entonces podemos escribir $a' = a + hn$ y $b' = b + kn$, y por tanto $a' + b' = a + b + pn$ y $a' \cdot b' = a \cdot b + qn$, $h, k, p, q \in \mathbb{Z}$, de donde

$$[a' + b'] = [a + b] \quad y \quad [a' \cdot b'] = [a \cdot b]$$

La aritmética computacional ofrece muchos ejemplos de operaciones binarias sobre conjuntos finitos. Cada computador tiene un repertorio de operaciones aritméticas sobre los números enteros, que normalmente incluyen sumas, diferencias, multiplicaciones y divisiones. A causa de la propia estructura del computador, sólo un subconjunto finito de enteros pueden ser manipulados. Por tanto, en la práctica, las operaciones aritméticas son modulares.

Se pueden describir otros tipos de operaciones binarias diferentes de las operaciones aritméticas brevemente comentadas. El hecho de que una señal pueda tomar valores sobre el conjunto $\mathbb{Z}_2 = \{0, 1\}$ hace que cualquier dispositivo con dos entradas y una salida represente una operación binaria sobre $\mathbb{Z}_2$.

Una operación binaria se puede describir tabulando los valores de los pares asociados a su dominio. Esta tabulación normalmente se llama *tabla de composición* de la operación. Como ejemplo consideremos la operación definida sobre el conjunto $A = \{a, b, c, d\}$ descrita en la tabla 9.1.

En particular son útiles las tablas de operaciones aritméticas modulares. Como ejemplos, podemos considerar las que figuran en las tablas 9.2 y 9.3.

Estas operaciones se dicen binarias por indicar que cada par ordenado de elementos de A es enviado por la operación a un nuevo elemento de A. Si son ternas ordenadas de elementos

Tabla 9.1: Tabla de una operación binaria

$\star$	a	b	c	d
a	b	c	d	d
b	a	b	c	d
c	c	a	c	d
d	a	b	a	b

Tabla 9.2: Tabla de la suma en $\mathbb{Z}_4$

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

de A las que son enviadas a A por la operación, diremos que la operación es *ternaria*. De forma similar, si la operación nos proporciona las imágenes en A de n-tuplas de A, entonces hablaremos de una operación *n-ária*.

Una manera sencilla de construir operaciones n-árias es la de componer recursivamente operaciones binarias. Por ejemplo, a partir de una operación binaria $f : A \times A \longrightarrow A$, podemos construir la operación ternaria

$$\begin{array}{rccc} f: & A\times A\times A & \longrightarrow & A \\ & (a,b,c) & \longrightarrow & f(f(a,b),c) \end{array}$$

Esta es la manera en que los ordenadores realizan operaciones sobre un conjunto de n

Tabla 9.3: Tabla del producto en $\mathbb{Z}_4$

$\cdot$	0	1	2	3
0	0	0	0	0
1	0	1	2	3
2	0	2	0	2
3	0	3	2	1

entradas, ya que internamente el ordenador sólo admite operaciones binarias. Por ejemplo, la operación ternaria $f(a,b,c) = a+b+c$ se efectúa haciendo

$$f(a,b,c) = (a+b)+c$$

Las operaciones binarias pueden tener propiedades que resultan imprescindibles para la construcción de estructuras algebraicas. Las que presentan un interés más relevante en este sentido son las siguientes.

Dadas dos operaciones binarias, $\star$ y $\perp$, definidas sobre un conjunto A, diremos que la operación $\star$ es

- *asociativa* si y sólo si, para todo $a,b,c \in A$, $a\star(b\star c) = (a\star b)\star c$;
- *conmutativa* si y sólo si, para todo $a,b \in A$, $a\star b = b\star a$;
- *distributiva* respecto de $\perp$ si y sólo si, para todo $a,b,c \in A$, se satisfacen las igualdades siguientes:

$$\begin{aligned} a\star(b\perp c) &= (a\star b)\perp(a\star c) \\ (b\perp c)\star a &= (b\star a)\perp(c\star a) \end{aligned}$$

Observar que las dos igualdades que se tienen que cumplir (para que $\star$ sea distributiva respecto de $\perp$) coinciden cuando $\star$ es conmutativa.

En los conjuntos de los números naturales, enteros y racionales, tanto la suma como el producto son operaciones asociativas y conmutativas, y el producto es distributivo respecto de la suma, pero la suma no es distributiva respecto del producto.

Cabe observar que, en $\mathbb{Q}$, la diferencia y el cociente no son ni asociativas ni conmutativas, y también que, en $\mathbb{Z}$, el cociente no está definido, ya que hay pares de números (por ejemplo, $(3,2)$) que no tienen una imagen definida. Similarmente, en $\mathbb{N}$ no se puede definir el cociente y, en este caso, tampoco la diferencia.

Ejercicio 9.4. Comprobar que la unión y la intersección sobre el conjunto de las partes de cualquier conjunto son operaciones asociativas y conmutativas y, en este caso, que la unión es distributiva respecto de la intersección y también en sentido contrario.

La composición de aplicaciones constituye un ejemplo importante de operación asociativa y no conmutativa. La asociatividad se comprueba fácilmente y, en cuanto a la conmutatividad, podemos considerar, por ejemplo

$$f,g : \mathbb{N} \longrightarrow \mathbb{N}$$

tales que, $f(n) = 2n$ y $g(m) = m+1$. Entonces, $(g \circ f)(n) = 2n+1$, mientras que $(f \circ g)(m) = 2(m+1)$.

Un conjunto con una operación binaria, independientemente de las propiedades de esta operación, puede admitir ciertos elementos que por su comportamiento respecto de la operación se llaman *singulares*. Así, dado un conjunto A y una operación binaria $\star$, se dice que $e \in A$ es un elemento *neutro* respecto de $\star$ si y sólo si

$$a \star e = e \star a = a \qquad \forall a \in A$$

Lema 9.5. En cada operación binaria existe como máximo un elemento neutro.

Demostración. Supongamos que e y e' fuesen dos elementos neutros respecto de $\star$. Entonces se tendría que verificar que $e' = e \star e' = e' \star e = e$; por tanto, si existe neutro, éste es único. □

Si A tiene elemento neutro e respecto de $\star$, entonces se dice que $a' \in A$ es un elemento *inverso* de $a \in A$ respecto de $\star$ si y sólo si $a \star a' = a' \star a = e$.

Lema 9.6. Si $\star$ es una operación asociativa sobre el conjunto A, y admite un elemento neutro $e \in A$, entonces cada elemento $a \in A$ admite como máximo un elemento inverso $a' \in A$.

Demostración. Si $a', a'' \in A$ fuesen dos inversos de $a \in A$, entonces

$$a' = a' \star e = a' \star (a \star a'') = (a' \star a) \star a'' = a''$$

□

La existencia de neutros e inversos respecto de las operaciones aritméticas elementales puede comprobarse fácilmente en los conjuntos de números con que se trabaja habitualmente. Así, tenemos:

1. En $\mathbb{N}$ no hay elemento neutro respecto de la suma y, por tanto, no tiene sentido hablar de inversos respecto de esta operación. Si consideramos el producto, el 1 es el elemento neutro. Como no existe ningún natural diferente del 1, que multiplicado por otro dé 1, ningún elemento (diferente de 1) tiene inverso.

2. Si consideramos la suma en $\mathbb{Z}$, tenemos el cero como neutro y $-a$ como inverso de $a \in Z$, mientras que con el producto, aunque también exista neutro, el 1, ningún elemento tiene inverso.

3. En el conjunto $\mathbb{Q}$, tanto la suma como el producto admiten elementos neutros e inversos respecto de las dos operaciones.

4. Las operaciones aritméticas elementales sobre $\mathbb{Z}_n$ admiten como neutros las clases del $[0]$ y del $[1]$ respectivamente. El inverso respecto de la suma de una clase $[a]$ es claramente la clase $[-a]$. La existencia de inversos respecto del producto no es tan evidente como en el caso de la suma; en el próximo capítulo veremos que $[a] \in \mathbb{Z}_n$ admite inverso respecto del producto si y sólo si $mcd(a,n) = 1$.

   Como ejemplo, si consideramos en $\mathbb{Z}_9$ los elementos 2, 4, 5, 7 y 8, estos tienen inverso, como es fácil comprobar, $(2 \times 5 \equiv 1 \pmod 9)$, $4 \times 7 \equiv 1 \pmod 9$, $(8 \times 8 \equiv 1 \pmod 9)$, mientras que no hay ningún entero s tal que $3 \times s \equiv 1 \pmod 9$ ni $6 \times s \equiv 1 \pmod 9$.

Ejemplos no numéricos de elementos singulares los encontramos al considerar:

1. El conjunto de las partes de un conjunto, $\wp(X)$, con la unión y la intersección como operaciones admite como neutros respectivos el $\emptyset$ y el propio conjunto X. Es fácil comprobar que ningún subconjunto no trivial de X admite inverso respecto de la unión ni respecto de la intersección.

2. El conjunto de las aplicaciones, $F(X)$, definidas desde un conjunto cualquiera X en él mismo respecto de la composición admite siempre la aplicación identidad como neutro, ya que para todo $x \in X$, y para toda $f \in F(X)$,

$$(f \circ Id)(x) = f(Id(x)) = f(x) = Id(f(x)) = (Id \circ f)(x)$$

   Recordemos que la aplicación inversa de $f \in F(X)$ sólo existe si f es biyectiva, entonces la existencia de aplicaciones inversas queda restringida al subconjunto de aplicaciones biyectivas, que denotamos como $F^*(X)$.

9.4 Estructuras algebraicas

En esta última sección se introduce la noción de estructura algebraica, así como también ciertos aspectos generales, teniendo en cuenta que en los capítulos sucesivos se desarrollarán con más precisión los modelos más importantes de estas estructuras.

La idea básica subyacente en la definición de estructura algebraica es la de un conjunto con una o varias operaciones, aunque puede intervenir más de un conjunto, así como también otros tipos de relaciones. En general, las operaciones definidas pueden ser n-arias, pero si no se especifica lo contrario, aquí consideraremos únicamente operaciones binarias y nos referiremos a ellas directamente como operaciones.

Una *estructura algebraica* es una n-tupla cuyos elementos son conjuntos y relaciones entre estos conjuntos, de las cuales se destacan en particular las operaciones y también, si se

quiere, los elementos singulares que puedan tener estos conjuntos respecto de las operaciones asociadas. Para denotarla podemos escribir

$$(X, Y, \dots, R_1, R_2, \dots, \star, \perp, \dots, e, e', \dots)$$

donde $X, Y, \dots$ son conjuntos, $R_1, R_2, \dots$ relaciones definidas en estos conjuntos, $\star, \perp, \dots$ son operaciones n-arias (en general) sobre estos conjuntos, y $e, e', \dots$ elementos singulares de estas operaciones.

De hecho, al largo de todo este capítulo se han estado utilizando ya ejemplos de estructuras algebraicas, entre las cuales tenemos:

1. $(\mathbb{N}, \leqslant)$. Relación de orden total sobre los naturales.
2. $(\mathbb{N}, =)$. Relación de equivalencia, también sobre los naturales.
3. $(\wp(X), \cup, \emptyset)$. La operación unión que tiene por neutro el conjunto vacío.
4. $(F(X), \circ, Id)$. La composición de aplicaciones con la aplicación identidad como neutro.
5. $(\mathbb{Q}, +, \cdot, 0, 1)$. La suma y el producto sobre los racionales con el 0 como neutro de la suma y el 1 como neutro del producto.

Es preciso mencionar que las estructuras más importantes están definidas sobre un único conjunto en el cual hay definidas una o dos operaciones. Los elementos singulares normalmente no se especifican si son fácilmente deducibles. A continuación daremos una clasificación ordenada de estas estructuras. Primero introduciremos aquellas que están definidas a partir de una única operación.

Dado un conjunto X y una operación $\star$ sobre X, diremos que la estructura algebraica $(X, \star)$ es un:

- *semigrupo* si y sólo si $\star$ es asociativa;
- *monoide* si y sólo si $\star$ es asociativa y X tiene elemento neutro;
- *grupo* si y sólo si $\star$ es asociativa, X tiene elemento neutro y cada elemento de X tiene inverso.

Si $\star$ es conmutativa diremos que la estructura correspondiente es abeliana o conmutativa.

Entre los ejemplos que se han tratado al largo del capítulo es rutinario comprobar la estructura algebraica que corresponde a algunos de ellos.

Ejercicio 9.7. Comprobar las siguientes afirmaciones.

1. $(\mathbb{Z}, \times)$ es un monoide abeliano.

2. $(\mathbb{Z}, +)$ es un grupo abeliano.
3. $(\wp(X), \cup)$ es también un monoide abeliano.
4. $(F(X), \circ)$ es un monoide no abeliano.
5. $(F^*(X), \circ)$ es un grupo no abeliano.

Dado un conjunto X y dos operaciones, $\star$ y $\perp$ sobre X, diremos que $(X, \star, \perp)$ es un:

- *anillo* si y sólo si $(X, \star)$ es un grupo abeliano, $\perp$ es asociativa y distributiva respecto de $\star$,
- *anillo unitario* si y sólo si es un anillo y $(X, \perp)$ tiene elemento neutro,
- *anillo unitario abeliano* si y sólo si es un anillo y $(X, \perp)$ tiene elemento neutro y $(X, \perp)$ es conmutativa,
- *cuerpo* si y sólo si es un anillo unitario abeliano y $(X, \perp)$ tiene elementos inversos.

De forma similar a como hemos hecho con los ejemplos sobre estructuras algebraicas con una única operación, podemos aprovechar aquí también ejemplos ya tratados con dos operaciones.

Ejercicio 9.8. Comprobar las afirmaciones siguientes.

1. $(\mathbb{Z}, +, \times)$ es un anillo unitario abeliano.
2. $(\mathbb{Z}_n, +, \times)$ es también un anillo unitario abeliano.
3. $(\mathbb{Q}, +, \times)$ es un cuerpo abeliano.
4. $(\mathbb{Z}_p, +, \times)$ es un cuerpo abeliano si y sólo si $p \in Z$ es un número primo.

Ejercicio 9.9. Comprobar que el conjunto de las aplicaciones entre números racionales, respecto de la suma y el producto, definidas a continuación, $(F(\mathbb{Q}), +, \times)$, tiene estructura de anillo unitario abeliano.

Para toda $f, g \in F(\mathbb{Q})$, y para todo $q \in \mathbb{Q}$, definimos:

$$\begin{array}{rcl} F(\mathbb{Q}) \times F(\mathbb{Q}) & \longrightarrow & F(\mathbb{Q}) \\ (f,g)(q) & \longrightarrow & (f+g)(q) = f(q) + g(q) \end{array}$$

$$\begin{array}{rcl} F(\mathbb{Q}) \times F(\mathbb{Q}) & \longrightarrow & F(\mathbb{Q}) \\ (f,g)(q) & \longrightarrow & (f \times g)(q) = f(q) \times g(q) \end{array}$$

Observar que la unicidad en la suma y el producto de números racionales se transmite a la suma y el producto de aplicaciones racionales. Es decir, estas operaciones están bien definidas.

Una vez sabemos lo que significa que un conjunto X tenga una determinada estructura algebraica, es natural plantearse la cuestión siguiente: al considerar un subconjunto $Y \subset X$, ¿es posible que esta estructura se mantenga al restringirla a Y? Esta cuestión da lugar a un concepto muy utilizado en este ámbito, el de subestructura.

Dada una estructura algebraica $(X, \star)$ y un subconjunto $X' \subset X$, se dice que $(X', \star)$ es una *subestructura* de la anterior si y sólo si la operación $\star$ es cerrada en X', es decir,

$$x' \star y' = z' \in X' \qquad \forall x', y' \in X'$$

y además mantiene las propiedades y los elementos singulares que definen la estructura original.

Observar que la asociatividad y la conmutatividad de una operación se mantienen en cualquier subconjunto del conjunto de partida.

De forma general, dada una estructura algebraica,

$$(X, Y, \dots, R_1, R_2, \dots, \star, \perp, \dots, e, e', \dots)$$

y una familia de subconjuntos $X' \subset X$, $Y' \subset Y$, ..., se dice que la estructura algebraica

$$(X', Y', \dots, R_1, R_2, \dots, \star, \perp, \dots, e, e', \dots)$$

es una subestructura de la estructura original si y sólo si las relaciones, las operaciones y los elementos singulares originales se mantienen con las mismas propiedades al considerar la estructura original restringida a la familia de subconjuntos.

Entre los ejemplos anteriores podemos observar que algunas de las estructuras algebraicas son subestructuras de otras.

1. $(\mathbb{N}, \times)$ es un submonoide abeliano del monoide abeliano $(\mathbb{Z}, \times)$. Pero, si consideramos como operación la suma, la estructura de grupo que hay en $\mathbb{Z}$ se pierde en $\mathbb{N}$, ya que este último conjunto no contiene los elementos inversos.

2. $(\mathbb{Z}, +)$ es un subgrupo abeliano del grupo abeliano $(\mathbb{Q}, +)$. Pero no es cierto que el anillo unitario abeliano $(\mathbb{Z}, +, \times)$ sea una subestructura del cuerpo abeliano $(\mathbb{Q}, +, \times)$, ya que los elementos inversos respecto del producto en $\mathbb{Q}$ no están en $\mathbb{Z}$.

Dos estructuras algebraicas diferentes pueden compartir características similares. Si una estructura está definida sobre un conjunto X y una estructura similar lo está sobre un conjunto

Y, la similitud de estas estructuras se pone de manifiesto por medio de una aplicación entre los conjuntos X e Y que conserva las características de la estructura.

Dadas dos estructuras algebraicas $(X,\star)$ e $(Y,\perp)$, la aplicación $f : X \longrightarrow Y$ es un *morfismo* de $(X,\star)$ en $(Y,\perp)$ si y sólo si

$$f(x)\perp f(x') = f(x\star x') \qquad \forall x,x' \in X$$

La definición dice que la imagen de la composición de dos elementos coincide con la composición de las imágenes de cada uno de ellos.

Si $f : X \longrightarrow Y$ es un morfismo de $(X,\star)$ en $(Y,\perp)$, se dice que $(f(X),\perp)$ es la *imagen homomórfica* de X por f.

Observar que la operación $\perp$ será siempre cerrada en $f(X)$.

Por ejemplo, $(\mathbb{Z},\times)$ y $(\mathbb{N}\cup\{0\},\times)$ son homomórficos, ya que la aplicación

$$|\,| : \mathbb{Z} \longrightarrow \mathbb{N}\cup\{0\}$$

que envía cada entero z a su módulo $|z|$, satisface la condición de morfismo, es decir,

$$|z\times z'| = |z|\times|z'| \qquad \forall z,z' \in \mathbb{Z}$$

Como los morfismos son aplicaciones, éstas pueden ser inyectivas, exhaustivas y biyectivas. En cada caso reciben también nombres especiales.

Si f es un morfismo de $(X,\star)$ en $(Y,\perp)$, entonces diremos que f es un:

- *monomorfismo* si y sólo si f es inyectiva,
- *epimorfismo* si y sólo si f es exhaustiva,
- *isomorfismo* si y sólo si f es biyectiva.

Como ejemplos de esta clasificación podemos considerar los siguientes:

1. La aplicación identidad $Id : (\mathbb{N},\times) \longrightarrow (\mathbb{Z},\times)$, que envía cada número natural a él mismo, es un monomorfismo.
2. La aplicación $|\,| : (\mathbb{Z},\times) \longrightarrow (\mathbb{N}\cup\{0\},\times)$, que envía cada número entero z a su módulo $|z|$, es un epimorfismo.
3. La congruencia módulo n, $[\,] : (\mathbb{Z},+) \longrightarrow (\mathbb{Z}_n,+)$, que envía cada entero z a su clase $[z]$, es un ejemplo importante de epimorfismo.

Como una congruencia módulo n es una relación de equivalencia, ésta induce una partición del conjunto en clases y da lugar al conjunto cociente, que permite introducir la noción de *estructura cociente*. Más concretamente y en general:

Una relación de equivalencia R sobre una estructura $(X,\star)$ se dice que es *compatible* respecto de una operación $\star$ si y sólo si

$$xRx',\ yRy' \quad \Rightarrow \quad (x\star y)R(x'\star y')$$

Es decir, si la operación no depende de los representantes elegidos.

Si R es compatible con $\star$, entonces se dice que R induce una *estructura cociente*, $(X/R,\star_R)$ donde X/R es el conjunto de las clases de equivalencia de X módulo R y $\star_R$ es la operación inducida por R, es decir,

$$[x\star y] = [x]\star_R[y] \qquad \forall x,y\in X$$

Observar que no hay ambigüedad en la definición de $\star_R$, ya que no depende de los representantes escogidos en cada clase. Es por ello que se ha introducido la noción de compatibilidad. Ya hemos visto que la relación de congruencia módulo n en los enteros es compatible con la suma y con el producto. Justamente, $(\mathbb{Z}_n,+_n,\times_n)$ es la estructura cociente del anillo unitario $(\mathbb{Z},+,\times)$ por la relación de congruencia módulo n.

La compatibilidad de una relación respecto de una operación es una propiedad muy restrictiva. Es fácil encontrar ejemplos de particiones no compatibles con ciertas operaciones. Así, en $\mathbb{Z}$, la partición $A_1=\{1,2\}$ y $A_2=\mathbb{Z}\setminus A_1$ no es compatible con la suma, ya que al operar dos elementos de la clase A_1 podemos obtener un nuevo elemento de la misma clase $[1+1]=[2]=A_1$, o bien un elemento de la otra clase $[1+2]=[3]=A_2$.

Es interesante observar que la estructura algebraica $(X/R,\star_R)$ es una imagen homomórfica de la estructura $(X,\star)$ considerando el epimorfismo

$$f: X \longrightarrow X/R$$

que envía cada $x\in X$ a su clase $[x]$. Habitualmente, éste se llama *epimorfismo natural* de X sobre X/R.

A todo morfismo de una estructura algebraica sobre ella misma se le llama *endomorfismo*. Si el morfismo es biyectivo entonces se llama *automorfismo*.

La noción de morfismo se extiende a todas las estructuras algebraicas. Por ejemplo, un morfismo de la estructura $(X,\star,\perp)$ en la estructura $(Y,\star',\perp')$ es una aplicación $f: X\longrightarrow Y$ que satisface para todo $x,y\in X$,

$$f(x\star y)=f(x)\star' f(y) \quad y \quad f(x\perp y)=f(x)\perp' f(y)$$

es decir, respeta todas las operaciones y consiguientemente todas las propiedades que involucren a estas operaciones.

En general, la imagen homomórfica de una estructura algebraica es otra estructura con conjuntos, relaciones, operaciones y elementos singulares que se corresponden uno a uno con cada elemento de la estructura inicial. Además, las propiedades especiales de la estructura original se mantienen en la estructura imagen. Así, si $\star$ es asociativa en $(X, \star, \perp, e)$, $\star'$ lo es también en la imagen homomórfica $(f(X) \subset X', \star', \perp', e')$, como se puede comprobar fácilmente. En particular, el elemento neutro de la estructura original va a parar al elemento neutro de la estructura imagen, como se demuestra a continuación.

Lema 9.10. Dadas dos estructuras algebraicas $(X, \star, e)$ e $(Y, \perp, e')$ con elementos neutros respectivos e y e' y un morfismo $f : X \longrightarrow Y$, se cumple siempre que $e' = f(e)$.

Demostración. Para todo $x \in X$, la imagen homomórfica de $x \star e = x$ es

$$f(x) \perp f(e) = f(x)$$

y, por tanto, $f(e) = e'$. □

Capítulo 10

Grupos

1. Definiciones y propiedades
2. Grupos abelianos finitos
3. Grupos de permutaciones
4. Digrafos de Cayley
5. Teoría de enumeración de Pólya

La estructura de grupo es la más simple de las que se considerarán y también una de las que tiene una incidencia más extensa en sus aplicaciones.

En la sección 1 de este capítulo se revisa la definición de grupo que ya se ha enunciado en el capítulo anterior, se introduce la terminología básica y se ven las primeras propiedades. La sección 2 está dedicada al estudio de los grupos abelianos finitos, de los cuales se describe la estructura. Los grupos de permutaciones, y en particular los grupos simétrico y alternado, merecen una atención especial y en la sección 3 se consideran algunos aspectos algebraicos y combinatorios de estos grupos. Los grafos de Cayley proporcionan una manera de visualizar la estructura de un grupo. La interrelación entre la teoría de grupos y la teoría de grafos por medio de los grafos de Cayley es muy enriquecedora para ambas teorías y está relacionada con la descripción de un grupo mediante lo que se llaman *presentaciones*. Estas cuestiones se tratan en la sección 4. El capítulo se acaba con una aplicación de la teoría de grupos a un problema de enumeración que se conoce como teoría de Pólya. El objetivo es enumerar configuraciones diferentes sobre un conjunto que goza de ciertas simetrías.

10.1 Definiciones y propiedades

Tal como se ha introducido en la última sección del capítulo anterior, la estructura de grupo viene dada por la definición siguiente.

Un *grupo* es un par $(G,\cdot)$ formado por un conjunto y una operación binaria que cumple:

G0 La operación es cerrada, es decir, $a\cdot b\in G$, para todo $a,b\in G$.

G1 La operación es asociativa, es decir, $(a\cdot b)\cdot c=a\cdot(b\cdot c)$, para todo $a,b,c\in G$.

G2 El conjunto G tiene elemento neutro, que se denotará por e, respecto de la operación.

G3 Cada elemento de G tiene inverso respecto de la operación. El inverso del elemento $a\in G$ se denotará por a^{-1}.

Si además la operación es conmutativa, se dice que el grupo es *abeliano*.

Las propiedades descritas en los axiomas $G1$, $G2$ y $G3$ son una abstracción de las propiedades que satisfacen las operaciones elementales en los conjuntos de números. Así, los conjuntos de números enteros o racionales con la suma, $(\mathbb{Z},+)$, $(\mathbb{Q},+)$, son ejemplos de grupos abelianos. De la misma manera, el conjunto de números racionales con el producto, $(\mathbb{Q}^*,\cdot)$, tiene también estructura de grupo abeliano.

Por abuso de notación nos referiremos a menudo a un grupo indicando sólo su conjunto base, dejando de lado la referencia a la operación. Así pues, hablaremos del grupo G en lugar de hablar del grupo $(G,\cdot)$, de manera que debe quedar sobreentendido a qué operación se hace referencia. Siguiendo los modelos aritméticos de los conjuntos de números, la notación genérica de la operación es '$\cdot$' (notación multiplicativa) y entonces el elemento neutro se denota por 'e' o por '1'. A menudo se escribe 'ab' en lugar de '$a\cdot b$'. Cuando el grupo es abeliano, la operación se denota por '$+$' (notación aditiva), el elemento neutro se denota por '0' y el inverso de x por '$-x$'.

Algunas de las propiedades elementales que se derivan de los axiomas de grupo son las siguientes:

Proposición 10.1. En un grupo $(G,\cdot)$ se cumple:

1. El elemento neutro es único.
2. El elemento inverso de cada elemento es único.
3. El inverso de a^{-1} es a, es decir, $(a^{-1})^{-1}=a$.
4. El inverso de $a\cdot b$ es $b^{-1}\cdot a^{-1}$, es decir, $(a\cdot b)^{-1}=b^{-1}\cdot a^{-1}$.
5. La ecuación $a\cdot x=b$ tiene una solución única $x=a^{-1}\cdot b\in G$.

Ejercicio 10.2. Demostrar las propiedades enunciadas en la proposición anterior e indicar cuáles de los axiomas de grupo se usan.

La última de las propiedades, que asegura que en un grupo una ecuación del tipo $ax = b$ tiene siempre una solución única en x, es característica de la estructura de grupo. Este punto de vista es importante ya que, históricamente, los objetivos iniciales del álgebra estaban ligados a la resolución de ecuaciones.

Es preciso observar también el cambio de orden en la escritura de los elementos en la propiedad 3 de la proposición. Si G es un grupo abeliano, se puede escribir $(a \cdot b)^{-1} = a^{-1} \cdot b^{-1}$. Esta relación aparentemente más natural puede dejar de cumplirse si el grupo no es abeliano.

En este capítulo nos centraremos sobre todo en grupos finitos, es decir, en los que el conjunto de base del grupo es finito.

Una de las maneras de representar la estructura de un grupo finito es dando la tabla de la operación. Por ejemplo, la tabla siguiente corresponde a la de un grupo de cuatro elementos, $G = \{e, a, b, c\}$.

Tabla 10.1: Tabla de un grupo de cuatro elementos

·	e	a	b	c
e	e	a	b	c
a	a	b	c	e
b	b	c	e	a
c	c	e	a	b

En la tabla de un grupo se pueden visualizar fácilmente los dos últimos axiomas de la estructura: la fila y la columna que corresponden al elemento neutro son idénticas a la fila y la columna cero respectivamente. Además, en cada fila y en cada columna aparece una única vez este elemento neutro al operar cada elemento con su único inverso. Otra propiedad característica más de la tabla de un grupo es que en cada una de las filas y de las columnas aparece una única vez cada uno de los elementos del grupo. Esto es porque si $x \cdot y = x \cdot z$, multiplicando los dos lados de la igualdad por x^{-1} obtenemos $y = z$. La tabla muestra también si el grupo es abeliano, caso en el que hay una simetría respecto de la diagonal principal como en la tabla anterior. La propiedad asociativa es la que no queda reflejada en la tabla y se tiene que verificar de manera exhaustiva (y a menudo tediosa).

Ejercicio 10.3. Usando las propiedades que se han mencionado de la tabla de un grupo, demostrar que hay un único grupo de dos elementos y un único grupo de tres elementos.

Subgrupos

De acuerdo con la noción genérica de subestructura, se dice que un subconjunto $H \subset G$ es un *subgrupo* de G si con la operación '$\cdot$' restringida a los elementos de H se satisfacen los axiomas de grupo. Por ejemplo, el subconjunto formado por los elementos $H = \{e, b\}$ en el grupo de la tabla anterior (10.1) es un subgrupo, ya que la operación restringida a este subconjunto es cerrada, tiene elemento neutro e y el elemento b tiene como inverso el mismo b. La tabla de este subgrupo está representada en 10.2.

Tabla 10.2: Subgrupo del grupo de la tabla 10.1

$\cdot$	e	b
e	e	b
b	b	e

De hecho, 10.2 corresponde a la tabla del único grupo de dos elementos. En realidad no es preciso comprobar los cuatro axiomas de grupo para determinar si un subconjunto es o no un subgrupo.

Proposición 10.4. Sea $(G, \cdot)$ un grupo y $H \subset G$. Entonces, $(H, \cdot)$ es un subgrupo de $(G, \cdot)$ si y sólo si se satisface la relación

$$a \cdot b^{-1} \in H \qquad \forall a, b \in H$$

Si G es finito, $(H, \cdot)$ es un subgrupo si y sólo si la operación es cerrada en H.

Demostración. Supongamos que se satisface la relación. Si $a \in H$, tomando $b = a$ obtenemos $a \cdot a^{-1} = e \in H$ de manera que H contiene el elemento neutro. Entonces, tomando $a = e$, para cualquier elemento $b \in H$, $e \cdot b^{-1} = b^{-1} \in H$, de manera que cualquier elemento de H tiene inverso en H. Dados dos elementos $a, b \in H$, tenemos que $a \cdot (b^{-1})^{-1} = ab \in H$, de manera que la operación es cerrada. Finalmente, la propiedad asociativa se hereda directamente de la misma propiedad en G. Recíprocamente, si $(H, \cdot)$ es un subgrupo de $(G, \cdot)$, está claro que se satisface la relación, es decir, si $a, b \in H$ entonces $a, b^{-1} \in H$ y por tanto $a \cdot b^{-1} \in H$. Finalmente, en caso que G sea finito, basta que la operación sea cerrada en H. La demostración se deja como ejercicio. □

Ejercicio 10.5. Demostrar que en el enunciado de la proposición anterior se puede substituir la relación $a \cdot b^{-1} \in H$ para todo $a, b \in H$ por

$$a^{-1} \cdot b \in H \qquad \forall a, b \in H$$

De manera simplificada se escribe $H < G$ para denotar que $(H, \cdot)$ es un subgrupo de $(G, \cdot)$. Está claro que el subconjunto formado sólo por el elemento neutro es un subgrupo de G. Todo el grupo G es también un subgrupo de él mismo. Estos dos se llaman subgrupos *triviales* de G, mientras que los subgrupos no triviales se llaman también subgrupos *propios*.

Ejercicio 10.6. Demostrar que $(\mathbb{Z}, +) < (\mathbb{Q}, +)$.

Ejercicio 10.7. Demostrar que $n\mathbb{Z} = \{nk, k \in \mathbb{Z}\}$, el conjunto de múltiplos de un número entero n, es un subgrupo de $(\mathbb{Z}, +)$.

Un subgrupo propio H de G induce en G una relación de equivalencia R_H definida por

$$aR_Hb \Leftrightarrow a^{-1} \cdot b \in H, \qquad \forall a, b \in G$$

Ejercicio 10.8. Demostrar que R_H es efectivamente una relación de equivalencia.

De acuerdo con la última proposición (y el ejercicio que sigue), si $a, b \in H$, entonces $a^{-1} \cdot b \in H$, de manera que aR_Hb. Recíprocamente, si aR_Hb y $a \in H$, entonces $b = a \cdot (a^{-1}b) \in H$. Esto quiere decir que los elementos de H forman una de las clases de equivalencia. De manera similar se puede ver que la clase de un elemento $a \in G$ es justamente el conjunto de los elementos $aH = \{ax, x \in H\}$ y que todas ellas se pueden describir de esta manera.

Ejercicio 10.9. Demostrar esta última afirmación: Si H es un subgrupo de G, las clases de equivalencia de la relación R_H son los conjuntos de la forma $aH = \{ax, x \in H\}$ para cada $a \in G$.

Por ejemplo, si G es el grupo de cuatro elementos de la tabla 10.1 y H el subgrupo de dos elementos de la tabla 10.2, $H = b + H = \{e, b\}$ y $a + H = \{a, c\}$ son las clases de equivalencia de la relación R_H (usamos la notación aditiva).

A causa de su forma, las clases de equivalencia por la relación R_H se llaman *clases laterales por la izquierda* de G módulo H. Está claro que todas las clases tienen el mismo cardinal (si aH, bH son dos clases, la aplicación $f : aH \longrightarrow bH$ dada por $f(ax) = bx$ es una biyección). Si G es un grupo finito, el número de clases se llama *índice* de H en G y se denota por $|G : H|$. Así pues,

$$|G : H| = \frac{|G|}{|H|}$$

Esto lleva a uno de los primeros resultados que se obtuvieron en la teoría de grupos.

Teorema 10.10 (Teorema de Lagrange). Sea G un grupo finito y H un subgrupo propio de G. Entonces $|H|$ es un divisor de $|G|$.

El teorema de Lagrange limita el número de subgrupos que puede tener un grupo. Por ejemplo, un grupo de orden primo no puede tener ningún subgrupo propio. El recíproco del teorema de Lagrange no es necesariamente cierto: el hecho que k sea un divisor de $|G|$ no quiere decir que G tenga que tener un subgrupo de orden k (o que no pueda tener más de uno).

Si $H < G$, habríamos podido definir también la relación de equivalencia ${}_HR$ dada por

$$a\,{}_HR\,b \Leftrightarrow a \cdot b^{-1} \in H$$

En este caso, las clases de equivalencia son Ha, $a \in G$ y se llaman *clases laterales por la derecha* de G módulo H. Si G es un grupo abeliano, entonces $aH = Ha$, para todo $a \in G$; en este caso, las clases laterales por la derecha coinciden con las clases por la izquierda. Si G no es abeliano, las clases por la derecha no coinciden necesariamente con las clases por la izquierda y las dos relaciones dan lugar a particiones diferentes.

Si $xH = Hx$ para todo $x \in G$, se dice que H es un subgrupo *normal* de G y se indica escribiendo $H \triangleleft G$. En este caso, la operación '$\cdot$' del grupo G induce una operación en el conjunto de G/H de clases de equivalencia definida como

$$(xH) \cdot (yH) = (x \cdot y)H$$

Ejercicio 10.11. Comprobar que, si H es un subgrupo normal de G, la operación está bien definida, es decir, el resultado no depende del representante que se escoge en cada clase. Más concretamente, si $xH = x'H$ e $yH = y'H$, entonces $(x \cdot y)H = (x' \cdot y')H$. Esto no es necesariamente cierto si H no es un subgrupo normal.

No es difícil comprobar que el conjunto G/H con esta operación vuelve a tener estructura de grupo. De hecho, la clase eH es su elemento neutro y el elemento inverso de xH es $x^{-1}H$. Este grupo se llama *grupo cociente* de G módulo H.

Ejercicio 10.12. Demostrar que, efectivamente, si $H \triangleleft G$, entonces G/H con la operación $(aH) \cdot (bH) = abH$ tiene estructura de grupo.

En el capítulo anterior hemos visto un ejemplo importante de grupo cociente. Recordemos que la relación de congruencia módulo n en el conjunto $\mathbb{Z}$ de los números enteros está definida como

$$x \equiv y \pmod{n} \Leftrightarrow n|(x-y)$$

Si llamamos $n\mathbb{Z}$ al subgrupo de los múltiplos de n introducido en el ejercicio 10.7, vemos que la relación de congruencia es una relación de equivalencia módulo este subgrupo. Como $(\mathbb{Z},+)$ es un grupo abeliano, $n\mathbb{Z}$ es un subgrupo normal, de manera que se puede definir el grupo cociente $\mathbb{Z}/n\mathbb{Z}$, que habitualmente se denota por $\mathbb{Z}_n$, con la operación

$$(a+n\mathbb{Z}) + (b+n\mathbb{Z}) = (a+b) + n\mathbb{Z}$$

Tanto $\mathbb{Z}$ como $n\mathbb{Z}$ son grupos infinitos, pero $\mathbb{Z}_n$ es un grupo finito de orden n.

Morfismos de grupos

Particularizaremos ahora al caso de la estructura de grupo otra de las definiciones generales que se han dado en el capítulo anterior. Una aplicación

$$f : G \longrightarrow H$$

entre dos grupos $(G,\cdot)$, $(H,\circ)$ es un *morfismo* de grupos si

$$f(a\cdot b) = f(a)\circ f(b) \qquad \forall a,b \in G$$

es decir, es lo mismo operar dos elementos en G y aplicar la función f que aplicar la función f a los dos elementos y operar las imágenes en H.

Ejercicio 10.13. Sea $f : G \longrightarrow H$ un morfismo de grupos.

1. Demostrar que $f(e_G) = e_H$.
2. Demostrar que $f(a^{-1}) = (f(a))^{-1}$.

Si la función f es biyectiva, se dice que es un *isomorfismo* de grupos y también que los dos grupos $(G,\cdot)$ y $(H,\circ)$ son *isomorfos*. Dos grupos isomorfos tienen las mismas propiedades algebraicas y, desde el punto de vista de la estructura, difieren sólo en la denominación de sus elementos. Por ejemplo, el conjunto $H = \{0,1\}$ con la operación 'o exclusiva' que tiene por tabla:

$\oplus$	0	1
0	0	1
1	1	0

es isomorfo al grupo representado en la tabla 10.2, donde el isomorfismo viene dado por $f(e) = 0$ y $f(b) = 1$. Desde el punto de vista de la estructura, los dos grupos son entonces idénticos.

Ejercicio 10.14. Sea $f : G \longrightarrow H$ un morfismo del grupo $(G,\cdot)$ en el grupo $(H,\circ)$.

1. Demostrar que el subconjunto $G_0 = \{x \in G \mid f(x) = e_H\}$ es un subgrupo de G. Demostrar además que se trata de un subgrupo normal.
2. Demostrar que el subconjunto $Im\, f$ es un subgrupo de H.
3. Demostrar que f es un morfismo inyectivo si y sólo si G_0 es el subgrupo trivial $G_0 = \{e_G\}$.
4. Demostrar que la aplicación $\hat{f} : G/G_0 \longrightarrow Im\, f$ dada por $\hat{f}(x\cdot G_0) = f(x)$ está bien definida y es un isomorfismo de grupos.

Producto cartesiano de grupos

El producto cartesiano de grupos proporciona una manera de generar nuevos grupos a partir de otros conocidos.

Dados dos grupos $(G_1, \star_1, e_1)$ y $(G_2, \star_2, e_2)$, podemos definir una operación $\star$ sobre $G_1 \times G_2$ de forma natural:

$$(g_1, g_2) \star (h_1, h_2) = (g_1 \star_1 h_1, g_1 \star_2 h_2)$$

Se puede comprobar fácilmente que este producto es asociativo, como consecuencia de la asociatividad en los grupos originales. Está claro que (e_1, e_2) es el elemento neutro de esta nueva operación y que, para cada $(g_1, g_2) \in G_1 \times G_2$, $(g_1^{-1}, g_2^{-1}) \in G_1 \times G_2$ es su inverso. Así,

$$(G_1 \times G_2, \star)$$

es un grupo que se llama *producto cartesiano* de los grupos G_1 por G_2.

Ejercicio 10.15. Demostrar que $G_1 \times G_2$ es abeliano si y sólo si lo son G_1 y G_2.

Ejercicio 10.16. Comprobar que $\mathbb{Z}_2 \times \mathbb{Z}_2$ es un grupo no isomorfo a $\mathbb{Z}_4$.

De forma similar se puede definir el producto cartesiano de n grupos, $G_1, G_2, \ldots, G_n$,

$$G_1 \times G_2 \times \cdots \times G_n$$

Ejercicio 10.17. Comprobar que $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_3$ es un grupo isomorfo a $(\mathbb{Z}_2 \times \mathbb{Z}_2) \times \mathbb{Z}_3$.

De los subgrupos propios de G_1 y G_2 podemos obtener también de manera natural subgrupos propios de $G_1 \times G_2$.

Ejercicio 10.18. Si H_1 y H_2 son subgrupos propios de G_1 y G_2 respectivamente, entonces $H_1 \times H_2$ es un subgrupo propio de $G_1 \times G_2$.

En particular, G_1 y G_2 se pueden considerar subgrupos de $G_1 \times G_2$ por medio de las identificaciones siguientes:

$$\begin{aligned} G_1 &\leftrightarrow G_1' = G_1 \times \{e_2\} \\ G_2 &\leftrightarrow G_2' = \{e_1\} \times G_2 \end{aligned}$$

Es inmediato ver que estas identificaciones representan isomorfismos que identifican cada $g_1 \in G_1$ con $(g_1, e_2) \in G_1'$ y, similarmente, cada $g_2 \in G_2$ con $(e_1, g_2) \in G_2'$. Así, cada elemento $(g_1, g_2) \in G_1 \times G_2$ se identifica de manera única con $((g_1, e_2), (e_1, g_2)) \in G_1' \times G_2'$. Es preciso observar que, $G_1' \cap G_2' = (e_1, e_2)$ y además se cumple que $(g_1, e_2) \star (e_1, g_2) = (e_1, g_2) \star (g_1, e_2)$ para todo $g_1 \in G_1$ y $g_2 \in G_2$.

Recíprocamente, nos podemos plantear la cuestión siguiente: ¿es posible expresar un grupo G como producto cartesiano de otros grupos de órdenes (evidentemente) más pequeños (y por tanto más manejables)?

Esta cuestión sugiere la definición siguiente. Se dice que un grupo G es *producto directo* de sus subgrupos H y H' si

1. $G = \{hh' \mid h \in H, h' \in H'\}$;

2. $H \cap H' = \{e\}$, e es el elemento neutro de G;

3. H y H' son subgrupos normales en G.

En particular, el producto cartesiano $G = G_1 \times G_2$ es también el producto directo de los subgrupos G'_1 y G'_2 de G,

$$G = G_1 \times G_2 \simeq G'_1 \times G'_2$$

de manera que las nociones de producto cartesiano y producto directo de grupos son equivalentes.

Observar que, si G es abeliano, la tercera condición la cumple cualquiera de sus subgrupos. Así, para obtener una descomposición de G como producto directo de otros grupos, será preciso buscar entre los subgrupos que tengan intersección trivial. En este caso, se usa a veces $G = H \oplus H'$ para denotar que el producto directo de H y H' es abeliano, siguiendo la costumbre de utilizar la notación aditiva en el caso de grupos abelianos.

Así, por ejemplo, $(\mathbb{Z}_6, +)$ tiene dos subgrupos propios,

$$\begin{aligned} H &= \{0,3\} \simeq \mathbb{Z}_2 \\ H' &= \{0,2,4\} \simeq \mathbb{Z}_3 \end{aligned}$$

cuya intersección es el elemento neutro de $(\mathbb{Z}_6, +)$. Podemos obtener los elementos de $\mathbb{Z}_6$ a partir de los elementos de $\mathbb{Z}_2 \times \mathbb{Z}_3$, mediante la identificación que figura a continuación:

$\mathbb{Z}_6$	$\leftrightarrow$	$\mathbb{Z}_2 \times \mathbb{Z}_3$
0	$\leftrightarrow$	(0,0)
1	$\leftrightarrow$	(1,1)
2	$\leftrightarrow$	(0,2)
3	$\leftrightarrow$	(1,0)
4	$\leftrightarrow$	(0,1)
5	$\leftrightarrow$	(1,2)

Tabla 10.3: Tabla de $(\mathbb{Z}_2 \times \mathbb{Z}_3, +)$

+	(0,0)	(1,1)	(0,2)	(1,0)	(0,1)	(1,2)
(0,0)	(0,0)	(1,1)	(0,2)	(1,0)	(0,1)	(1,2)
(1,1)	(1,1)	(0,2)	(1,0)	(0,1)	(1,2)	(0,0)
(0,2)	(0,2)	(1,0)	(0,1)	(1,2)	(0,0)	(1,1)
(1,0)	(1,0)	(0,1)	(1,2)	(0,0)	(1,1)	(0,2)
(0,1)	(0,1)	(1,2)	(0,0)	(1,1)	(0,2)	(1,0)
(1,2)	(1,2)	(0,0)	(1,1)	(0,2)	(1,0)	(0,1)

Con esta identificación es fácil comprobar que la tabla 10.3 correspondiente a

$$(\mathbb{Z}_2 \times \mathbb{Z}_3, +)$$

coincide con la tabla de $(\mathbb{Z}_6, +)$.

Por tanto, podemos afirmar que $\mathbb{Z}_6 = H_1 \oplus H_2 \simeq \mathbb{Z}_2 \times \mathbb{Z}_3$.

Ejercicio 10.19. Comprobar que $\mathbb{Z}_{12} \simeq \mathbb{Z}_4 \oplus \mathbb{Z}_3$. Comprobar también que $\mathbb{Z}_{12}$ no es producto directo de $\mathbb{Z}_2$ y $\mathbb{Z}_6$. ¿Por qué?

En la sección siguiente se usarán descomposiciones en productos directos para obtener la clasificación de los grupos abelianos.

10.2 Grupos abelianos finitos

En esta sección se describe una clase importante de grupos finitos: los grupos abelianos. Los grupos abelianos más simples son los grupos *cíclicos*, que se verán en primer lugar. A continuación se verá que todos los grupos abelianos se pueden expresar como productos directos de grupos cíclicos.

Grupos cíclicos

Los grupos cíclicos proporcionan el ejemplo más sencillo de grupo finito. Para introducirlos se considera primero el concepto de orden de un elemento en un grupo. Sea G un grupo finito y g un elemento de G. Indicaremos por

$$g^k = \underbrace{gg\ldots g}_{k}$$

Consideremos la sucesión de elementos $g, g^2, g^3, \ldots, g^k, \ldots$. Como el grupo es finito, en esta sucesión no todos los elementos pueden ser diferentes, de manera que para algunos índices m, n tendremos $g^m = g^n$. Supongamos que $m < n$. Multiplicando ambos lados de la igualdad por $(g^m)^{-1} = g^{-m}$, obtenemos $g^{n-m} = e$. Así pues, tenemos:

Proposición 10.20. Si G es un conjunto finito y $g \in G$, existe un entero k tal que $g^k = e$.

La proposición anterior justifica la definición siguiente.

Sea G un grupo finito y $g \in G$. Se llama *orden* de g, y se indica por $|g|$, al menor número natural k para el cual $g^k = e$.

Si g tiene orden k, entonces los elementos $g, g^2, \ldots, g^{k-1}, g^k = e$ son todos diferentes. Efectivamente, si hubiese dos iguales, $g^p = g^q, p < q < k$, entonces $g^{q-p} = e$, contrariamente al hecho que k es el menor natural con esta propiedad. Además, $g^m = g^{m+k}$ para todo $m \in \mathbb{N}$, de manera que la secuencia infinita de potencias de g tiene período k, es decir, los elementos de la secuencia se repiten cada k posiciones. Está claro que el producto de dos potencias de g es otra potencia de g, de manera que la operación del grupo es cerrada en el subconjunto $H = \{g, g^2, \ldots, g^{k-1}, g^k = e\}$. Según la proposición 10.4, tenemos el resultado siguiente.

Proposición 10.21. Sea G un grupo finito y $g \in G$ un elemento de orden k. Entonces

$$H = \{g, g^2, \ldots, g^{k-1}, g^k = e\}$$

es un subgrupo de G de orden $k = |g|$.

En particular, según el teorema de Lagrange, tenemos:

Corolario 10.22. Sea G un grupo finito y $g \in G$. Entonces $|g|$ es un divisor de $|G|$.

Al subgrupo H de las potencias de un elemento $g \in G$ se lo llama subgrupo *generado* por g, y también se dice que g *genera* H. La estructura cíclica de este grupo sugiere la definición siguiente.

Un grupo finito G es *cíclico* si contiene un elemento g que genera todo el grupo, es decir, todos los elementos de G se expresan como potencia de g.

Ejercicio 10.23. Demostrar que un grupo cíclico es abeliano.

Ejercicio 10.24. Sean G y H dos grupos cíclicos del mismo orden k, generados por los elementos $g \in G$ y $h \in H$ respectivamente. Demostrar que la aplicación $f : G \longrightarrow H$ definida por $f(g^n) = h^n, n = 1, 2, \ldots, k$ es un isomorfismo de grupos.

Según el enunciado de este último ejercicio, vemos que, salvo isomorfismo, hay como mucho un único grupo cíclico para cada orden k. Vamos a ver que hay efectivamente uno para cada orden.

Proposición 10.25. El grupo $\mathbb{Z}_n$ es un grupo cíclico de orden n.

Demostración. Cualquier elemento $k+n\mathbb{Z} \in \mathbb{Z}_n$ se puede escribir como $k+n\mathbb{Z} = k(1+n\mathbb{Z})$, de manera que la clase del 1 genera todo el grupo. □

Los grupos cíclicos aparecen en muchas y diversas aplicaciones, de modo que resulta útil adquirir una cierta habilidad para operar en estos grupos (lo que se llama *aritmética modular*). La manera habitual de trabajar en estos grupos consiste en tomar como representantes de las n clases los enteros de 0 a $n-1$ y efectuar la suma ordinaria entre estos elementos, buscando después el representante correspondiente. Así, por ejemplo, los grupos cíclicos de órdenes 4 y 5 tienen las tablas siguientes:

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

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

Así pues, hay un y sólo un grupo cíclico de orden n para cada entero positivo, el grupo $\mathbb{Z}_n$. Este es el único grupo de orden n cuando n es un número primo.

Proposición 10.26. Si p es un número primo, hay un único grupo de orden p y este es cíclico.

Demostración. Sea G un grupo de orden p y g un elemento cualquiera de G diferente de e. Según el corolario 10.22, el orden de g es un divisor de $|G| = p$. Como p es primo, tiene que ser $|g| = p$, de manera que el subgrupo generado por g es todo G. □

Los grupos cíclicos tienen otra particularidad en relación al teorema de Lagrange: para cada divisor k de n, existe un único subgrupo de $\mathbb{Z}_n$ de orden k. Esto se puede ver a partir de las proposiciones siguientes.

Proposición 10.27. Cualquier subgrupo de un grupo cíclico es cíclico.

Proposición 10.28. Sea k un divisor de n y $h = n/k$. El conjunto de múltiplos de h en $\mathbb{Z}_n$ forma un subgrupo de orden k.

Ejercicio 10.29. Demostrar las proposiciones anteriores y deducir que para cada divisor k de n hay un único subgrupo de orden k de $\mathbb{Z}_n$.

Proposición 10.30. Si n y m son enteros primos entre sí, entonces

$$\mathbb{Z}_n \times \mathbb{Z}_m \simeq \mathbb{Z}_{nm}$$

Este resultado es consecuencia de la proposición 10.33 que veremos más adelante. En general, $\mathbb{Z}_n$ es isomorfo al producto directo de los subgrupos cíclicos que tienen órdenes divisores de n y estos órdenes son primos entre sí.

Teorema 10.31. Sea $\mathbb{Z}_n$ un grupo cíclico de orden $n = p_1^{n_1} p_2^{n_2} \cdots p_s^{n_s}$, donde p_i son números primos diferentes. Entonces,

$$\mathbb{Z}_n \simeq \mathbb{Z}_{p_1^{n_1}} \times \mathbb{Z}_{p_2^{n_2}} \times \cdots \times \mathbb{Z}_{p_s^{n_s}}$$

Así, por ejemplo, sabemos ya que $\mathbb{Z}_6 \simeq \mathbb{Z}_2 \times \mathbb{Z}_3$. Si observamos la tabla 10.3, podemos comprobar que $(1,1) \in \mathbb{Z}_2 \times \mathbb{Z}_3$ genera todo $\mathbb{Z}_2 \times \mathbb{Z}_3$ y por tanto deducimos que $\mathbb{Z}_2 \times \mathbb{Z}_3$ es cíclico. Esta es otra manera de comprobar que $\mathbb{Z}_2 \times \mathbb{Z}_3$ es isomorfo a $\mathbb{Z}_6$.

El ejercicio siguiente muestra, en cambio, que no siempre se puede descomponer $\mathbb{Z}_n$ en producto de grupos cíclicos de órdenes divisores de n.

Ejercicio 10.32. Comprobar que el grupo cíclico de cuatro elementos, $\mathbb{Z}_4$, no es isomorfo a $\mathbb{Z}_2 \times \mathbb{Z}_2$. ¿Por qué?

Grupos abelianos

Conocer la estructura interna de un grupo es en general un problema difícil. Si el grupo es abeliano, este problema tiene solución, como veremos a continuación.

En el apartado anterior hemos estudiado el modelo más sencillo de grupo abeliano, aquel que está generado por un único elemento. Si consideramos ahora un grupo G generado por un conjunto de elementos $S = \{g_1, g_2, \ldots, g_r\} \subset G$ (es decir, cualquier elemento de G se puede obtener como 'producto' de elementos de S) tales que conmuten entre ellos, $g_i g_j = g_j g_i$, para todo i, j, entonces cualquier elemento g de G se puede expresar como producto de potencias de estos generadores:

$$g = g_1^{x_1} g_2^{x_2} \cdots g_r^{x_r}$$

Como consecuencia, G es un *grupo abeliano*.

La relación entre los órdenes de dos elementos y el orden del producto de estos elementos es importante para conocer cuál es la estructura interna del grupo que generan. El resultado siguiente es útil en este sentido.

Proposición 10.33. Sean g y h dos elementos de un grupo abeliano G con órdenes respectivos n y m tales que $\text{mcd}(n,m) = 1$. Entonces el orden de gh es nm.

Demostración. Si $k = nm$, podemos decir que $(gh)^k = g^k h^k = 1$ y, por tanto, el orden de gh tiene que ser un divisor de k. Supongamos ahora que este orden fuese $k' < k$. En este caso tendríamos

que $(gh)^{k'} = g^{k'}h^{k'} = 1$, y de aquí que $g^{k'} = (h^{-1})^{k'}$. Ahora bien, el orden de este elemento $|g^{k'}|$ tiene que dividir al orden de g, $|g| = n$ y el orden de h, $|h| = m$. Como $\mathrm{mcd}(n,m) = 1$, el orden de $g^{k'}$ y de $(h^{-1})^{k'}$ tiene que ser 1. Por otra parte, $1 = (hh^{-1})^{k'} = h^{k'}(h^{-1})^{k'} = h^{k'}$, de donde se deduce que k' tiene que ser un múltiplo de n, de m y del mínimo común múltiplo. Como $\mathrm{mcd}(n,m) = 1$, tiene que ser $k' = \mathrm{mcm}(n,m) = nm$. □

Ejercicio 10.34. Demostrar la proposición 10.30 usando la proposición anterior.

Para caracterizar los órdenes de los elementos de un grupo abeliano finito, es útil considerar el orden máximo de sus elementos, llamado *exponente del grupo*.

Proposición 10.35. El orden de cualquier elemento de un grupo finito abeliano divide al exponente del grupo.

Demostración. Sea n el exponente de un grupo abeliano G y sea $g \in G$ tal que $|g| = n$. Supongamos que existiese un elemento $g' \in G$ de orden $|g'| = n'$ tal que n' no dividiese a n. Si $d = \mathrm{mcd}(n,n')$, entonces $|g^d| = n/d$ es relativamente primo con n', $\mathrm{mcd}(n/d,n') = 1$, y, por tanto, la proposición anterior nos dice que $|g'g^d| = n'n/d$. Pero como hemos supuesto que n' no divide a n, $n' > d$ y el orden de $|g'g^d| > n$ en contradicción con el hecho que n es el exponente del grupo. □

En particular, si el orden de un grupo abeliano es su exponente, entonces el grupo es cíclico. Si el grupo está generado por más de un elemento, $G = \langle g_1, g_2, \ldots, g_r \rangle$, y los órdenes de sus generadores son primos entre sí, entonces el orden de $g_1 g_2 \cdots g_r$ es producto de los órdenes de todos los generadores, que es justamente el orden de G y por tanto en este caso el grupo es también cíclico. Como consecuencia, si un grupo finito abeliano no es cíclico, debe tener como mínimo dos generadores con órdenes no primos entre sí.

A continuación describiremos la estructura general de los grupos abelianos. Las demostraciones de los resultados que siguen no se incluirán en este texto a causa de su nivel de dificultad. El lector interesado las puede encontrar, por ejemplo, en [3].

El primer paso para determinar la estructura de un grupo abeliano es un resultado similar al de la proposición 10.30. Para cada primo p, llamamos $G(p)$ al conjunto de todos elementos de G que tienen orden una potencia de p.

Ejercicio 10.36. Demostrar que si $G(p) \neq \emptyset$, entonces $G(p)$ es un subgrupo de G. Demostrar que si p divide a n, entonces $G(p)$ tiene orden una potencia de p.

Recordemos que el orden de cualquier elemento divide al orden del grupo. El siguiente teorema asegura que la afirmación recíproca también es cierta para divisores primos de $|G|$ y constituye la clave para la clasificación de los grupos abelianos.

Teorema 10.37 (Cauchy). Si p es un primo que divide a $|G|$, entonces hay algún elemento $g \in G$ que tiene orden p.

Este teorema asegura que $G(p) \neq \emptyset$ si y sólo si p divide a n. Como, para dos primos diferentes p, q, $G(p) \cap G(q) = \emptyset$, se obtiene en particular:

Proposición 10.38. Si G es un grupo abeliano de orden $n = p_1^{n_1} p_2^{n_2} \cdots p_s^{n_s}$, entonces

$$G \simeq G(p_1) \times \cdots \times G(p_s)$$

Según la proposición anterior, sólo es preciso determinar la estructura de cada uno de los grupos abelianos $G(p)$ de orden una potencia de p. Si cada uno de ellos es cíclico, de acuerdo con los comentarios anteriores, G también es cíclico. Si $G(p)$ no es cíclico, se puede expresar también como producto de grupos cíclicos.

Proposición 10.39. Si G es un grupo abeliano de orden p^k, p primo, entonces

$$G \simeq \mathbb{Z}_{p^{r_1}} \times \cdots \times \mathbb{Z}_{p^{r_t}}$$

para algunos $r_1, \ldots, r_t$ tales que $k = r_1 + \cdots + r_t$.

Según la proposición anterior, cada descomposición de k en suma de enteros $r_1 + \cdots + r_t$ proporciona un grupo abeliano de orden p^k y todos se pueden obtener así. Por ejemplo, los únicos grupos abelianos de orden $n = 3^3$ son $\mathbb{Z}_{27}$, $\mathbb{Z}_9 \times \mathbb{Z}_3$ y $\mathbb{Z}_3 \times \mathbb{Z}_3 \times \mathbb{Z}_3$, que son diferentes entre sí.

Las proposiciones 10.38 y 10.39 proporcionan una caracterización completa de la estructura general de cualquier grupo abeliano finito.

Teorema 10.40. Si G es un grupo abeliano finito de orden $n = p_1^{n_1} p_2^{n_2} \cdots p_s^{n_s}$, entonces

$$G \simeq \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_t}$$

donde cada n_i es una potencia de alguno de los primos de la descomposición de n y $n_1 \cdots n_t = n$.

Ejercicio 10.41. Encontrar todos los grupos abelianos de orden $n = 24$.

10.3 Grupos de permutaciones

Como ya se ha mencionado en el capítulo anterior, el conjunto de aplicaciones biyectivas de un conjunto en él mismo forma un grupo con la composición de aplicaciones. El elemento neutro

del grupo es la aplicación identidad, y la aplicación inversa de una aplicación f está definida como $f^{-1}(y) = x \Leftrightarrow f(x) = y$.

Cuando X es un conjunto finito, estas aplicaciones se llaman *permutaciones*. Cualquier conjunto de permutaciones que forme grupo, se llama *grupo de permutaciones* y el conjunto de todas las permutaciones de n elementos es lo que se llama *grupo simétrico* de n símbolos que se denota por S_n y tiene $n!$ elementos.

Los grupos de permutaciones constituyeron uno de los estímulos principales para el estudio de los grupos finitos. De hecho, todo grupo finito se puede interpretar como un grupo de permutaciones. Este resultado, conocido como el teorema de Cayley, así como un estudio detallado de esta clase de grupos se verá más adelante en esta sección. Antes, sin embargo, discutiremos un ejemplo importante de grupos de permutaciones: los grupos de simetrías.

Grupos de simetrías

La estructura de grupo aparece de forma natural en el estudio de simetrías. En términos generales, una simetría sobre un conjunto es una biyección entre sus elementos que respeta su estructura. Las simetrías, por tanto, se pueden componer y el conjunto de todas ellas tiene estructura de grupo con la composición. En esta sección ilustramos este hecho a partir de un grupo de simetrías particular, el de los movimientos rígidos de un polígono regular que dejan su forma invariante.

Consideremos un triángulo equilátero de vértices ABC. Una rotación de $\pi/3$ radianes con centro el baricentro del triángulo lleva el vértice A al B, el B al C, el C al A y deja el triángulo invariante. Este es entonces un movimento rígido del triángulo que deja su forma invariante (véase la figura 10.1).

Llamamos g a este movimento. Si lo aplicamos dos veces (es decir, g^2), tenemos otro movimiento que también deja el triángulo invariante. Si lo aplicamos tres veces tenemos los vértices del triángulo en la posición inicial.

Los dos movimientos, g y g^2, no dejan ningún vértice fijo y son los únicos con esta propiedad. El triángulo admite, sin embargo, otros movimientos que lo dejan también invariante. El movimiento de rotación de π radianes sobre el eje dado por cada una de las alturas deja también el triángulo invariante. Los hay tres de estos movimientos, uno para cada una de las alturas. Llamamos a a la rotación que deja fijo el vértice A y, de manera similar, b y c a las que dejan fijos los vértices B y C respectivamente (véase la figura 10.1). Cada uno de estos movimientos queda identificado por su acción sobre los vértices A, B y C del triángulo, de manera que se puede ver el grupo de simetrías como un grupo de permutaciones de tres elementos. Los cinco movimientos que hemos visto y la identidad constituyen, por tanto, todas las simetrías del triángulo, ya que sólo hay seis permutaciones de tres elementos. En 10.4 está la tabla del grupo que forman. De la tabla se desprende que el grupo de simetrías de un triángulo es no abeliano.

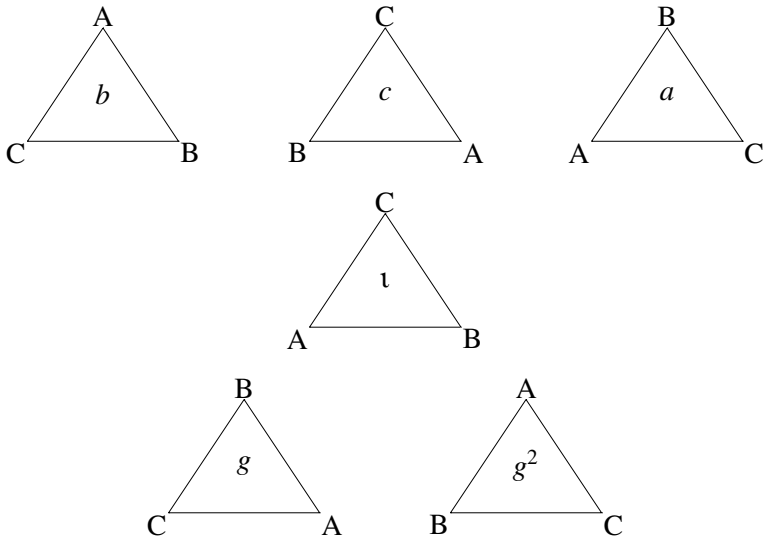


Figura 10.1: Simetrías del triángulo

Este es el grupo no abeliano más pequeño y forma parte de la familia de los llamados *grupos diédricos*, que se denotan por D_n y de los que hay uno para cada entero positivo n. La característica de todos estos grupos no abelianos es que tienen tamaño $2n$ y contienen un subgrupo cíclico de tamaño n que, además, es un subgrupo normal. En la tabla 10.4 se puede comprobar esta afirmación para $n = 3$. Para cada n, el grupo de simetrías de un polígono regular de n vértices es precisamente el grupo diédrico D_{2n}. Todas las simetrías se obtienen por una rotación de $2\pi/n$ radianes que, aplicada reiteradamente, proporciona las n simetrías que son giros y que dan lugar al subgrupo cíclico de D_{2n}. Los otros movimientos son rotaciones de π radianes en torno a un eje que pasa por un vértice y por el centro del polígono. En la figura 10.2 se ilustra

Tabla 10.4: Tabla de composición de las simetrías del triángulo

$\cdot$	e	g	g^2	a	b	c
e	e	g	g^2	a	b	c
g	g	g^2	e	c	a	b
g^2	g^2	e	g	b	c	a
a	a	b	c	e	g	g^2
b	b	c	a	g^2	e	g
c	c	a	b	g	g^2	e

la situación para el caso del pentágono.

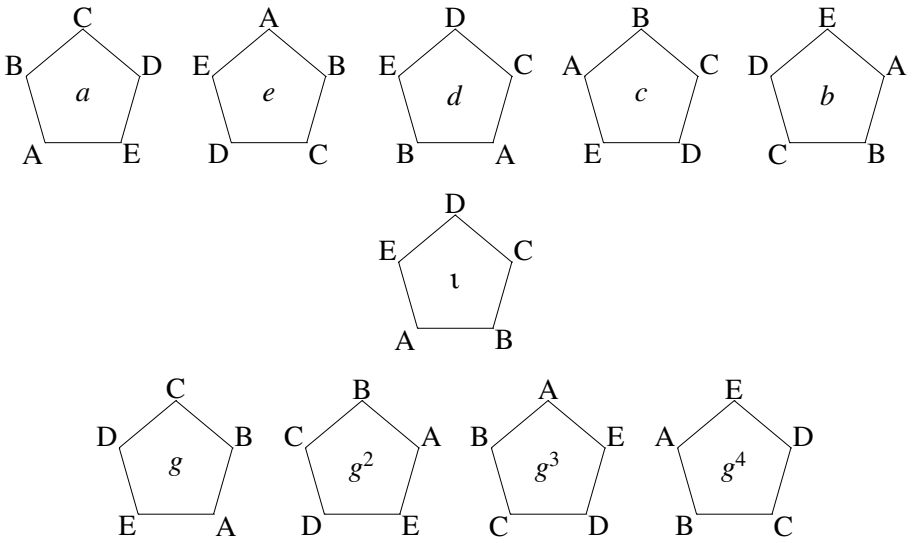


Figura 10.2: Simetrías del pentágono regular

El conjunto de automorfismos de un grafo (véase el problema 5.12) proporciona otro ejemplo importante de grupo de simetrías. En este caso, una simetría de un grafo es una biyección entre sus vértices que conserva las adyacencias. Por ejemplo, el grupo de automorfismos de un grafo completo de n vértices es el grupo simétrico de n símbolos, ya que cada biyección entre los vértices es un automorfismo del grafo.

Ejercicio 10.42. Demostrar que el grupo de automorfismos de un ciclo de orden n es el grupo diédrico D_n.

Notación cíclica de las permutaciones

Volvamos ahora al estudio general de los grupos de permutaciones. Para estudiarlos, el nombre que se da a los elementos del conjunto X donde se aplican resulta irrelevante, de manera que consideraremos $X = \{1, 2, \ldots, n\}$.

Para hacer más manejable el estudio de las permutaciones, conviene desarrollar una cierta notación. En primer lugar, si σ, τ son dos permutaciones de n elementos, su producto $\sigma\tau$ representa la composición leída de derecha a izquierda, es decir, aplicando τ en primer lugar y σ después (los algebristas suelen preferir la lectura inversa). Denotaremos la permutación identidad con la letra ι.

Para representar una permutación de n elementos, Lagrange usaba una notación matricial en la que colocaba los elementos de 1 a n en la primera fila y la lista de sus imágenes en la

segunda:

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma(1) & \sigma(2) & \cdots & \sigma(n) \end{pmatrix}$$

Esta es la notación *tabular*. Observar que la segunda fila es siempre una permutación de la primera, de donde viene la denominación de *permutaciones*.

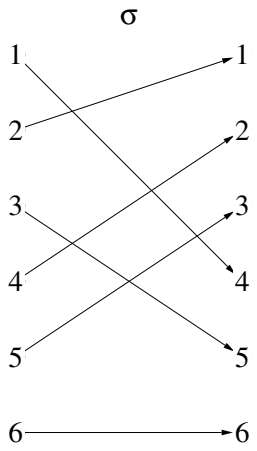
Hay otra notación que a menudo es útil, la notación *cíclica* introducida por Cauchy. Partiendo de un elemento $x_0 \in X$, consideremos la imagen de x_0, $\sigma(x_0)$, la imagen de éste, $\sigma(\sigma(x_0)) = \sigma^2(x_0)$, y así sucesivamente hasta que vuelve a aparecer $x_0 = \sigma^{j_0}(x_0)$. Si $j_0 = n$, escribimos

$$\sigma = (x_0, \sigma(x_0), \sigma^2(x_0), \ldots, \sigma^{n-1}(x_0))$$

mientras que, si $j_0 < n$, tomamos cualquier elemento que aún no haya aparecido, x_1, y consideramos las imágenes $\sigma(x_1), \sigma^2(x_1), \ldots$ hasta que vuelve a aparecer x_1. Iterando este procedimiento hasta que han aparecido todos los elementos, obtenemos una representación de la permutación como

$$\sigma = (x_0, \sigma(x_0), \ldots, \sigma^{j_0-1}(x_0))(x_1, \sigma(x_1), \ldots, \sigma^{j_1-1}(x_1)) \cdots (x_k, \sigma(x_k), \ldots, \sigma^{j_k-1}(x_k))$$

Por ejemplo, la permutación de 6 elementos



se escribe en notación tabular y en la notación cíclica como

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 1 & 5 & 2 & 3 & 6 \end{pmatrix} = (142)(35)(6)$$

Cada uno de los paréntesis en la última notación se llama *ciclo* de la permutación y la longitud de cada ciclo es el número de elementos que tiene. En el ejemplo, la permutación σ se escribe

como un ciclo de longitud 3, uno de longitud 2 y uno de longitud 1. Para simplificar la notación, a veces se dejan de lado los ciclos de longitud 1. Dos ciclos son *disyuntos* si no tienen ningún elemento en común. Una manera de expresar lo que hemos obtenido es la siguiente:

Proposición 10.43. Cada permutación $\sigma \in S_n$ se puede expresar de manera única como producto de ciclos disyuntos.

Ejercicio 10.44. Escribir en notación cíclica la permutación $\sigma\tau$, donde

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 1 & 5 & 2 & 3 & 6 \end{pmatrix} \quad \text{y} \quad \tau = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 2 & 6 & 1 & 4 & 5 \end{pmatrix}$$

Ejercicio 10.45. Escribir en notación tabular la permutación $\sigma\tau$, donde $\sigma = (154)(236)$ y $\tau = (136)(25)(3)$.

La estructura cíclica de una permutación define en cierta manera su estructura algebraica.

Lema 10.46. El orden de una permutación $\sigma \in S_n$ es el mínimo común múltiplo de las longitudes de los ciclos de su descomposición cíclica.

Demostración. Consideremos primero el caso en que σ se escribe como un único ciclo de longitud k, $\sigma = (x_1 x_2 \cdots x_k)$ (sin contar los ciclos de longitud 1). Entonces, σ^i envía x_1 a x_{1+i}, x_2 a x_{2+i} y, en general, x_m al símbolo con subíndice $m+i \pmod k$. En particular, σ^i es la permutación identidad si y sólo si i es un múltiplo de n. Si σ se escribe como el producto de ciclos disyuntos $c_1 c_2 \cdots c_r$, de longitudes $l_1, l_2, \ldots, l_r$, entonces $\sigma^i = c_1^i c_2^i \cdots c_r^i$, y este producto es la identidad si y sólo si cada c_j^i es la identidad, de manera que i tiene que ser múltiplo de cada una de las longitudes. Recíprocamente, si i es múltiplo de todas las longitudes l_i, entonces $\sigma^i = \iota$. □

Así, por ejemplo, el orden de la permutación $(14)(25)(36)$ es 2, y el de la permutación $(135)(24)$ es 6.

Teorema de Cayley

Hay muchos subconjuntos de permutaciones que forman grupo respecto de la composición, es decir, que son subgrupos del grupo simétrico. Por ejemplo, el subgrupo generado por una permutación que consta de un único ciclo,

$$c_n = (12 \cdots n)$$

es un grupo cíclico de orden n, es decir, $\langle c_n \rangle \simeq \mathbb{Z}_n$.

El teorema de Cayley proporciona un resultado que justifica el interés de los grupos de permutaciones para el estudio de los grupos finitos.

Teorema 10.47 (Cayley). Cualquier grupo G de n elementos es isomorfo a un grupo de permutaciones de n símbolos.

Demostración. Como hemos hecho en el ejemplo anterior, identificamos cada elemento $g \in G$ con la permutación σ_g de n símbolos definida por $\sigma_g(x) = gx$. Esta identificación proporciona una biyección f entre los elementos de G y n permutaciones de S_n. Para ver que es un morfismo, es preciso comprobar que $f(gh) = f(g)f(h)$. Pero $f(gh)$ es la permutación dada por $f(gh)(x) = ghx$ y $f(g)f(h)(x) = f(g)(hx) = ghx$, de manera que f es efectivamente un isomorfismo. □

Tabla 10.5: La tabla de D_3

$\cdot$	e	g	g^2	a	b	c
e	e	g	g^2	a	b	c
g	g	g^2	e	c	a	b
g^2	g^2	e	g	b	c	a
a	a	b	c	e	g	g^2
b	b	c	a	g^2	e	g
c	c	a	b	g	g^2	e

Para ilustrar este resultado utilizaremos el grupo diédrico de 6 elementos. Podemos identificar cada elemento del grupo con una permutación de seis elementos, de la manera siguiente,

$$\begin{array}{ccc} e & \longrightarrow & (1)(2)(3)(4)(5)(6) \\ g & \longrightarrow & (123)(465) \\ g^2 & \longrightarrow & (132)(456) \\ a & \longrightarrow & (14)(25)(36) \\ b & \longrightarrow & (15)(26)(34) \\ c & \longrightarrow & (16)(24)(35) \end{array}$$

donde hemos cambiado e, g, g^2, a, b, c por $1, 2, 3, 4, 5, 6$. Es fácil comprobar que esta identificación es en realidad un isomorfismo, es decir, que resulta lo mismo operar con los elementos del grupo diédrico que componer las correspondientes permutaciones. Esto es lo que dice el teorema de Cayley, que proporciona un contexto general para todos los grupos finitos.

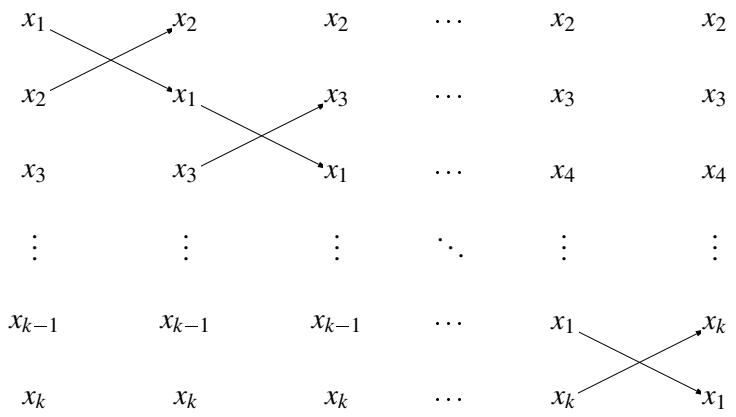
Transposiciones. El grupo alternado.

Una permutación que se escribe como un único ciclo de longitud 2 se llama *transposición*. Escribiremos $\tau_{ij} = (ij)$. El conjunto de transposiciones es interesante por el hecho que cual-

quier permutación se puede expresar como producto de transposiciones (no necesariamente disyuntas). Podemos expresar este resultado de la forma siguiente:

Proposición 10.48. El conjunto de todas las transposiciones de n símbolos genera todo S_n.

Demostración. Un ciclo $(x_1x_2\dots x_k)$ se puede expresar en términos de las transposiciones según el esquema siguiente:



o sea, que

$$(x_1x_2\dots x_k) = (x_1x_k)(x_1x_{k-1})\cdots(x_1x_3)(x_1x_2)$$

Ahora, cualquier permutación se puede expresar como producto de ciclos disyuntos, y cada uno de ellos se puede expresar como producto de transposiciones. □

Ejercicio 10.49. Escribir la permutación $(125)(346)$ como producto de transposiciones.

Ejercicio 10.50. Dar una cota superior del número de transposiciones que aparecen en la expresión de una permutación de S_n como producto de transposiciones.

En realidad no son precisas todas las transposiciones para generar todo el grupo simétrico.

Proposición 10.51. Las n transposiciones de la forma $(1i)$, $i = 2,3,\dots,n$ generan el grupo simétrico.

Demostración. Observemos simplemente que cualquier transposición (ij) se puede escribir como $(1i)(1j)(1i)$. □

Ejercicio 10.52. Demostrar que se puede generar todo el grupo S_n a partir de

1. las transposiciones $(12),(23),\dots,((n-1)n)$;

2. la transposición (12) y el ciclo $(23\dots n)$.

Según la proposición y el ejercicio anteriores, está claro que una permutación admite en general diversas expresiones diferentes como producto de transposiciones. Sin embargo, todas estas expresiones tienen una cosa en común.

Proposición 10.53. Si $\sigma = \tau_1\tau_2\cdots\tau_k = \tau'_1\tau'_2\cdots\tau'_{k'}$ son dos expresiones de la permutación σ como producto de transposiciones, entonces k y k' tienen la misma paridad.

Demostración. Sea π una permutación cualquiera y c el número de ciclos en su expresión cíclica (que es única). Sea $\tau = (ij)$ una transposición. Si i, j pertenecen al mismo ciclo $(ix_2\cdots x_{k-1}jx_{k+1}\cdots x_m)$ de la expresión cíclica de π, entonces

$$(ij)(ix_2\cdots x_{k-1}jx_{k+1}\cdots x_m) = (ix_2\cdots x_{k-1})(jx_{k+1}\cdots x_m)$$

(véase la figura 10.3), mientras que el resto de ciclos quedan inalterados por la acción de τ. Si, en cambio, i, j pertenecen a ciclos diferentes, $(ix_2\cdots x_m)$ y $(jy_2\cdots y_{m'})$, entonces $(ij)(ix_2\cdots x_m)$ $(jy_2\cdots y_{m'}) = (ix_2\cdots x_m jy_2\cdots y_{m'})$ y el resto de ciclos quedan inalterados por la acción de τ. En el primer caso, el número de ciclos de $\tau\pi$ es $c+1$ y en el segundo es $c-1$.

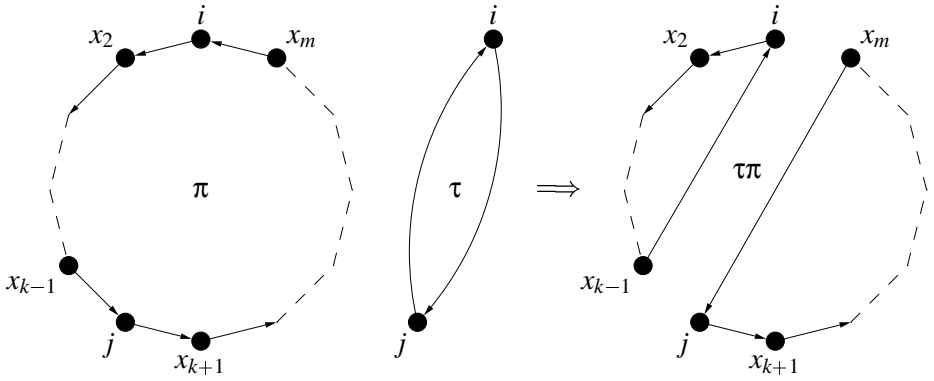


Figura 10.3: Composición de un ciclo y una transposición

Supongamos entonces que el número de ciclos en la expresión cíclica de $\sigma = \tau_1\tau_2\cdots\tau_k$ es c. El número de ciclos de τ_k es $(n-1)$ (contamos también los ciclos de longitud 1). Aplicando iteradamente el resultado anterior, cada vez que se aplica una transposición el número de ciclos del producto aumenta o disminuye en una unidad, y el resultado final tiene que ser c, de manera que $c = (n-1) - a + b$, siendo a el número de veces que disminuye y b el número de veces que aumenta, con $a + b = k$. Así pues, $c = (n-1) + k - 2a$. Haciendo lo mismo con la segunda descomposición obtendremos $c = (n-1) + k' - 2a'$ para un cierto a'. Restando las

dos igualdades, se ve que $k-k' = 2(a-a')$. Por tanto, o bien k y k' son ambos pares, o bien son ambos impares. □

Las permutaciones que se escriben como producto par de transposiciones se llaman permutaciones *pares* y las que no, se llaman *impares*. La *signatura* de una permutación es $sgn(\sigma) = 1$ si σ es par y $sgn(\sigma) = -1$ si es impar.

Ejercicio 10.54. Estudiar la paridad de un ciclo de orden k.

Ejercicio 10.55. Considerar el polinomio de n variables

$$P(x_1, x_2, \ldots, x_n) = \prod_{i<j} (x_i - x_j)$$

Demostrar que

$$P(x_1, x_2, \ldots, x_n) = (-1)^{sgn(\sigma)} P(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)})$$

Si τ es una transposición cualquiera, la aplicación $f_\tau : S_n \longrightarrow S_n$ definida como $f_\tau(\sigma) = \tau\sigma$ es una biyección que envía las permutaciones pares a las impares y viceversa, de manera que S_n contiene $n!/2$ permutaciones pares y un número igual de impares. Otra particularidad de esta clasificación de las permutaciones de S_n es la siguiente.

Proposición 10.56. El conjunto de permutaciones pares es un subgrupo de S_n.

Demostración. Sólo es preciso ver que el conjunto de permutaciones pares es cerrado por la composición. Esto es evidente, ya que si $\sigma = \tau_1\tau_2\cdots\tau_k$ y $\sigma' = \tau'_1\tau'_2\cdots\tau'_{k'}$, entonces su producto se puede escribir como producto de transposiciones $\sigma\sigma' = \tau_1\tau_2\cdots\tau_k\tau'_1\tau'_2\cdots\tau'_{k'}$ de longitud $(k+k')$, que es par si k y k' son pares. □

El subgrupo de permutaciones pares se llama subgrupo *alternado*, se denota por $Alt(n)$ y tiene $n!/2$ elementos. Como la relación de equivalencia inducida en S_n por este subgrupo sólo tiene dos clases, A_n es obviamente un subgrupo normal de S_n. Este subgrupo tiene una importancia singular por la conexión que establecieron Galois y Abel entre la posibilidad de resolver una ecuación de grado n $a_0x^n + a_{n-1}x^{n-1} + \cdots + a_{n-1}x + a_n = 0$ con una cantidad finita de operaciones elementales, con la existencia de subgrupos normales de A_n. Se puede demostrar que A_n no tiene subgrupos normales para $n \geqslant 5$, cosa que proporciona el argumento para asegurar que las ecuaciones de grado mayor que cuatro no se pueden resolver, en general, por radicales.

Grupos de matrices

Una manera de representar una permutación σ de n elementos consiste en considerar una matriz cuadrada $P_\sigma = (p_{ij})$ de orden n en la que

$$p_{ij} = \begin{cases} 1 & \text{si } \sigma(i) = j \\ 0 & \text{de otro modo} \end{cases}$$

Así, por ejemplo, a la permutación $\sigma = (142)(35)(6)$ le corresponde la matriz

$$P_\sigma = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Observemos que en cada fila y en cada columna hay exactamente un 1 y el resto de elementos son ceros. Este tipo de matrices se llaman justamente *matrices de permutaciones*. Observar que el determinante de cualquiera de estas matrices es ± 1. Diremos P_n al conjunto de las matrices de permutaciones de orden n.

El interés de esta representación proviene del hecho que la composición de permutaciones se traduce justamente en el producto de matrices.

Lema 10.57. Sean σ, τ dos permutaciones de S_n y P_σ, P_τ sus representaciones matriciales. Entonces:

1. $P_{\tau\sigma} = P_\sigma P_\tau$, es decir, la composición de permutaciones se traduce en producto de matrices.

2. $P_{\sigma^{-1}}$ es la matriz transpuesta ${P_\sigma}^T$.

Demostración. Sean $P_\sigma = (p_{ij})$ y $Q_\tau = (q_{ij})$. Entonces, su producto es $R = P_\sigma Q_\tau = (r_{ij})$ donde $r_{ij} = \sum_{k=1}^{n} p_{ik} q_{kj}$. El término r_{ij} vale 1 si y sólo si existe algún valor de k tal que $p_{ik} = q_{kj} = 1$ y en cualquier otro caso $r_{ij} = 0$. Como para cada fila (i) y para cada columna (j) existe un y sólo un valor de k tal que $p_{ik} = 1$ y $q_{kj} = 1$, $R = (r_{ij})$ es una matriz de permutaciones. Además, si π es la permutación asociada a R, $\pi(i) = j \Leftrightarrow r_{ij} = 1 \Leftrightarrow p_{ik} = q_{kj} = 1$ para un único k y $\tau\sigma(i) = \tau(k) = j$, de manera que $\pi = \tau\sigma$.

Por otra parte, $(p_{ji}) = P_{\sigma^{-1}}$, ya que $\sigma(i) = j$ si y sólo si $i = \sigma^{-1}(j)$. Por tanto, $P_{\sigma^{-1}}$ es la matriz transpuesta de P_σ. □

Como consecuencia directa del lema anterior, el conjunto P_n de todas las matrices de permutaciones de orden n tiene estructura de grupo con el producto de matrices, y este grupo es isomorfo a S_n.

Se pueden definir grupos de matrices más generales que los grupos de matrices de permutaciones. Por ejemplo, el conjunto de matrices cuadradas invertibles con coeficientes en $\mathbb{Q}$, $\mathbb{R}$ o $\mathbb{C}$ tienen estructura de grupo con el producto. Estos son ejemplos de grupos infinitos que, en general, no son abelianos.

El producto de matrices se puede definir también cuando los términos son los elementos de $\mathbb{Z}_n$ y las operaciones de suma y producto se hacen módulo n. Cualquier subconjunto de matrices invertibles donde la operación producto sea cerrada proporciona un nuevo ejemplo de grupo de matrices que, en este caso, es finito (y, en general, no abeliano). Los grupos de matrices proporcionan entonces una fuente importante de ejemplos de grupos finitos.

Por ejemplo, el conjunto de todas las matrices cuadradas 2×2 invertibles con términos de $\mathbb{Z}_2$ forma un grupo de 6 elementos:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Ejercicio 10.58. Demostrar que el grupo anterior es isomorfo al grupo diédrico de 6 elementos.

Ejercicio 10.59. Considerar el grupo de las matrices invertibles de la forma

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}$$

que tienen sus términos en $\mathbb{Z}_3$. ¿Es isomorfo al grupo del ejercicio anterior?

10.4 Digrafos de Cayley

Una buena manera de estudiar los grupos finitos consiste en describirlos a través de lo que se llaman *presentaciones*. Recordemos que, dado un grupo G, se dice que $S \subset G$ es un conjunto de generadores de G si cada elemento de G se puede expresar como producto de elementos de S, y se escribe

$$G = \langle S \rangle$$

Por ejemplo, si G es un grupo cíclico, hay un elemento $g \in G$ tal que cualquier elemento $x \in G$ se expresa como $x = g^k$ para una cierta potencia k de g, es decir, $G = \langle \{g\} \rangle$. El grupo diédrico de seis elementos introducido en la sección anterior no es cíclico. Esto quiere decir que se

precisa más de un elemento para conseguir un conjunto de generadores. Se puede comprobar fácilmente que el subconjunto $S = \{g, a\}$ es un conjunto de generadores de D_6.

Ejercicio 10.60. Expresar todos los elementos de D_6 como productos de elementos de $S = \{g, a\}$. ¿Son generadores los subconjuntos $S' = \{g, b\}$ y $S'' = \{a, b\}$? ¿Cuántos subconjuntos de dos generadores tiene el grupo?

Una lista de generadores de un grupo no es suficiente para determinar (salvo isomorfismos) de qué grupo se trata. Por ello es preciso indicar cuando dos expresiones diferentes corresponden al mismo elemento del grupo. Por ejemplo, en el caso del grupo cíclico de n elementos tenemos $g = g^{n+1}$ o $g^2 = g^{2n+2}$. En el caso del grupo diédrico generado por $\{g, a\}$, tenemos $g = ag^2a$. Todas estas igualdades se pueden expresar poniendo la identidad a un lado de la igualdad. Por ejemplo, las dos igualdades anteriores para el caso del grupo cíclico se expresarían como $g^n = e$ y $g^{2n} = e$, mientras que la identidad anterior del grupo diédrico se podría expresar como $gaga = e$. En este contexto, cada una de estas expresiones igualadas al elemento neutro se llama una *relación*.

El objetivo de presentar un grupo a través de generadores y relaciones consiste en encontrar un conjunto mínimo de relaciones a partir de las cuales se puedan obtener todos los elementos del grupo. Un conjunto de relaciones con esta propiedad se llama conjunto de relaciones *definidor* del grupo. Por ejemplo, en el caso del grupo cíclico de orden n, basta con la relación $g^n = e$ para deducir todas las demás. El par formado por un conjunto S de generadores y un conjunto R de relaciones definidoras se llama una *presentación* del grupo, y se denota por

$$G = \langle S | R \rangle$$

En el caso del grupo cíclico, entonces, tenemos la presentación

$$G = \langle g | g^n = e \rangle$$

y esta expresión determina el grupo. Como norma general, cuando se escribe $g^n = e$ en una presentación se sobreentiende que n es la potencia más pequeña de g que da e. De la misma manera, cuando hay varios generadores, se sobreentiende que ninguna subexpresión de una relación da el elemento neutro. Habitualmente, las relaciones se escriben simplemente como g^n en lugar de $g^n = e$.

Encontrar una presentación no es fácil en general, especialmente en lo que respecta a encontrar un conjunto reducido de relaciones definidoras. En general, una de las relaciones que se incluye es la que da el orden de los elementos. Una presentación de D_6 con los generadores g, a incluiría entonces las relaciones $g^3 = e$ y $a^2 = e$. Pero éstas no son definidoras del grupo. Por ejemplo, no se puede deducir que $gaga = e$. Con ésta se obtiene ya un conjunto de relaciones definidoras del grupo:

$$D_6 = \langle g, a | g^3, a^2, gaga \rangle$$

Determinar cuál es la tabla del grupo a partir de una presentación, o si un conjunto de relaciones es definidor de un grupo del cual tenemos la tabla, puede ser una tarea penosa, pero una presentación suele ser una manera económica y precisa de identificar un grupo. Los digrafos de Cayley proporcionan una visualización de un grupo expresado en términos de generadores que resulta muy útil.

Dado un grupo G y un conjunto de generadores S, el *digrafo de Cayley* de G respecto de S es el digrafo $Cay(G,S)$ que tiene por conjunto de vértices los elementos del grupo y por conjunto de arcos $\{(x,xs), x \in G, s \in S\}$. Si $S = S^{-1}$ e identificamos cada arco del digrafo con una arista, obtenemos el *grafo de Cayley* de G respecto de S. En las figuras 10.4 y 10.5 hay dibujados los digrafos de Cayley $Cay(\mathbb{Z}_6, \{2,3\})$ y $Cay(D_6, \{g,a\})$ y los grafos de Cayley $Cay(\mathbb{Z}_5, \{1,-1\})$ y $Cay(S_3\{(12),(13)\}$.

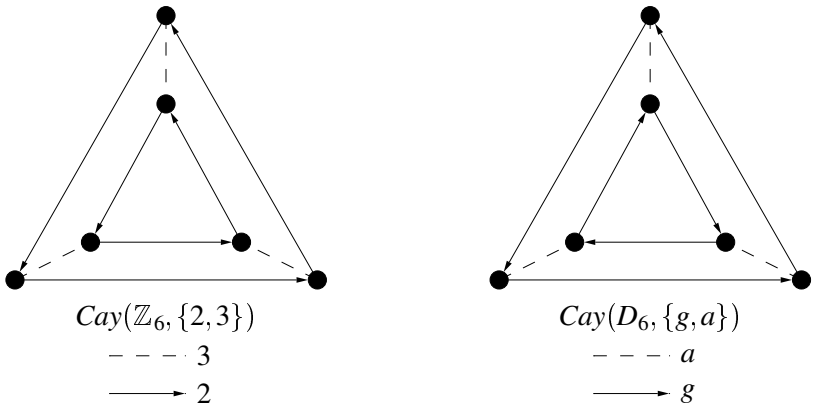


Figura 10.4: Ejemplos de digrafos de Cayley

El digrafo de Cayley del grupo cíclico de n elementos (con un único generador) es un ciclo de n vértices. Cada ciclo de un digrafo de Cayley corresponde a una relación, y un conjunto de relaciones es definidora del grupo si y sólo si todos los ciclos del digrafo se pueden descomponer en ciclos correspondientes a las relaciones definidoras.

Ejercicio 10.61. El *grupo de los cuaternones* es un grupo de ocho elementos que se define a través de la presentación siguiente:

$$Q_8 = \langle a,b | a^4, b^4, abab^{-1}, a^2b^2 \rangle$$

Dibujar el digrafo de Cayley $Cay(Q_8, \{a,b\})$.

El interés de los digrafos de Cayley está en la interrelación entre propiedades combinatorias del digrafo y propiedades algebraicas del grupo. No es difícil ver las siguientes interrelaciones.

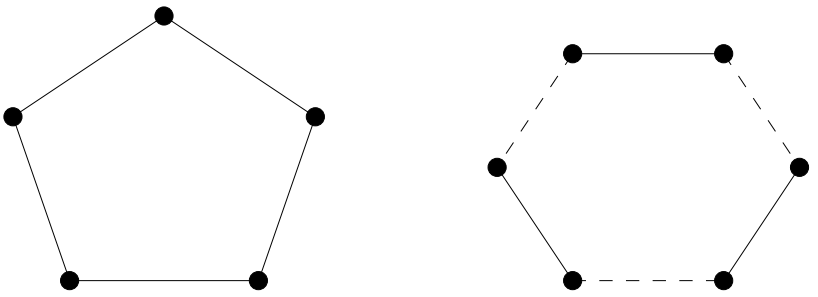


Figura 10.5: Ejemplos de grafos de Cayley

Proposición 10.62. Sea G un grupo y S un conjunto de generadores de G. Entonces $Cay(G,S)$ es un digrafo regular de grado $|S|$ fuertemente conexo.

Una característica especial de los digrafos de Cayley es la siguiente (véase el problema 5.13).

Proposición 10.63. Los digrafos de Cayley son vértice-simétricos.

Demostración. Dado un digrafo de Cayley, $Cay(G,S)$, y dos vértices $g,h \in G$ del digrafo, la aplicación $f_{gh} : G \longrightarrow G$ definida como $f_{gh}(x) = (hg^{-1})x$ satisface:

1. es una biyección, ya que $(hg^{-1})x = (hg^{-1})y$ implica $x = y$;
2. $f_{gh}(g) = h$;
3. es un automorfismo del digrafo, ya que (x, xs) es un arco si y sólo si $((hg^{-1})x, (hg^{-1})xs)$ lo es.

Por tanto, para cada par de vértices hay un automorfismo del digrafo que envía uno al otro. □

Desde el punto de vista de la teoría de grafos, los digrafos de Cayley suministran ejemplos numerosos y variados de digrafos vértice-simétricos. No todos los digrafos vértice-simétricos son, sin embargo, digrafos de Cayley. El ejemplo más pequeño en número de vértices de digrafo vértice-simétrico que no es de Cayley es el digrafo de Petersen, obtenido a partir del grafo del mismo nombre reemplazando cada arista por un arco. En cambio, el ciclo de n vértices es el digrafo de Cayley de un grupo cíclico de orden n respecto de un generador, y el digrafo completo de n vértices es el digrafo de Cayley de cualquier grupo G de orden n

respecto del conjunto de generadores $S = G \setminus \{e\}$. Los llamados digrafos de doble enlace, donde cada nodo $i \in \{0, 1, \ldots, n-1\}$ es adyacente a $i \pm a \pmod n$ y a $i \pm b \pmod n$ con $a, b \in \{0, 1, \ldots, n-1\}$, y que son utilizados en el diseño de redes de interconexión, son los grafos de Cayley $Cay(\mathbb{Z}_n, \{a, b, -a, -b\})$.

10.5 Enumeración de Pólya

La teoría de enumeración de Pólya tiene origen en el intento de enumerar diferentes compuestos químicos orgánicos. La diferencia entre compuestos se establece por la naturaleza y posición de los átomos en la estructura de una molécula, pero no entre aquellos que sólo difieren por ciertas simetrías. Por ejemplo, los dos primeros compuestos de la figura 10.6 son el mismo, mientras que el primero y el último no son químicamente iguales.

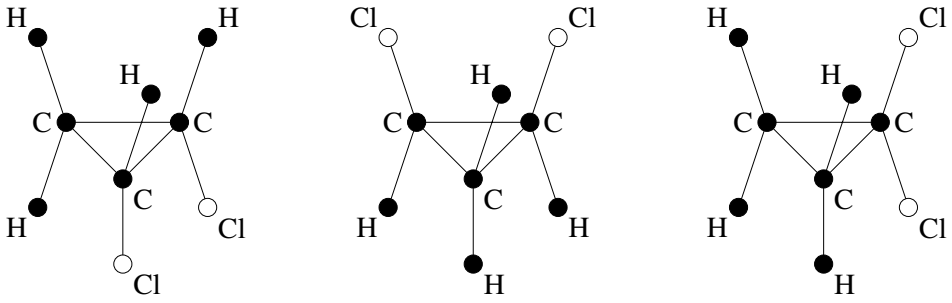


Figura 10.6: Simetría de compuestos químicos

El resultado que permite resolver este tipo de problemas de enumeración se conoce como el teorema de Pólya. Su objetivo es enumerar clases diferentes de configuraciones construidas sobre un objeto que tiene una cierta simetría. Para la exposición de la teoría de Pólya, substituiremos átomos por etiquetas o *colores*, que se asignan a diferentes elementos de un conjunto. Este conjunto gozará de ciertas simetrías que harán que diferentes maneras de colorear los vértices resulten equivalentes. El objetivo será contar cuántas clases de maneras equivalentes de hacer la coloración hay.

Supongamos, por ejemplo, que queremos asignar uno de los dos colores $\{\bullet, \circ\}$ a los vértices del grafo de la figura 10.7. De las $PR_4^2 = 16$ posibles maneras de hacer esta asignación, no consideraremos, sin embargo, diferentes aquellas que no se pueden distinguir sin enumerar explícitamente los vértices, es decir, las configuraciones encuadradas en la figura 10.8.

Con este criterio de diferenciación se obtienen, pues, 9 maneras diferentes en lugar de las 16 originales.

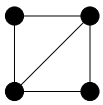


Figura 10.7:

La situación es, en general, la siguiente. Tenemos un conjunto $X = \{1, 2, \ldots, n\}$ y un grupo de permutaciones G de los elementos de X. En el ejemplo, X es el conjunto de vértices del grafo, y G es el grupo de automorfismos del grafo (si numeramos los vértices de 1 a 4 en sentido horario, $G = \{\iota, (13), (24), (13)(24)\}$). Finalmente tenemos un conjunto $C = \{c_1, \ldots, c_k\}$ de etiquetas o colores (en el ejemplo $C = \{\bullet, \circ\}$). Cada aplicación

$$f : X \longrightarrow C$$

proporciona una manera de asignar colores a los elementos de X. Diremos que es una *coloración* de los elementos de X. Llamamos C^X al conjunto de estas coloraciones, que tiene k^n elementos. Lo que decide el criterio de diferenciación de dos coloraciones es la acción del grupo G. Diremos que dos coloraciones $f, f' \in C^X$ son *G-equivalentes* si existe alguna permutación $\sigma \in G$ de manera que $f' = f\sigma$. En el ejemplo anterior, las coloraciones de la figura 10.9 son equivalentes, ya que, numerando los vértices en sentido horario y tomando la permutación $\sigma = (24) \in G$, tenemos que $f' = f\sigma$.

De forma similar, diremos que dos coloraciones son G-diferentes si no son G-equivalentes. El teorema de Pólya proporciona una manera de obtener este número de configuraciones diferentes en términos del tamaño n de X, del número k de colores y de la estructura del grupo G de permutaciones. El resultado se basa en la enumeración de lo que se llaman *órbitas* de un grupo de permutaciones. La *órbita* de un elemento $x \in X$, que denotaremos por O_x, es el conjunto de elementos $y \in X$ para los cuales hay alguna permutación en G que envía x a y, es decir:

$$O_x = G(x) = \{g(x), g \in G\} \subset X$$

En el grupo de automorfismos $G = \{\iota, (13), (24), (13)(24)\}$ del grafo del ejemplo anterior, la órbita del 1 es $\{1, 3\}$ y la del 2 es $\{2, 4\}$. En general, el número de órbitas de un grupo de permutaciones está relacionado con el número de puntos que deja fijos cada una de las permutaciones: como más puntos fijos dejen los elementos de G, más órbitas tiene el grupo. Para hacer precisa esta afirmación necesitamos otra definición. Dado un elemento $x \in X$, el *estabilizador* de x en G es el conjunto G_x de permutaciones que dejan x fijo,

$$G_x = \{g \in G \mid g(x) = x\} \subset G$$

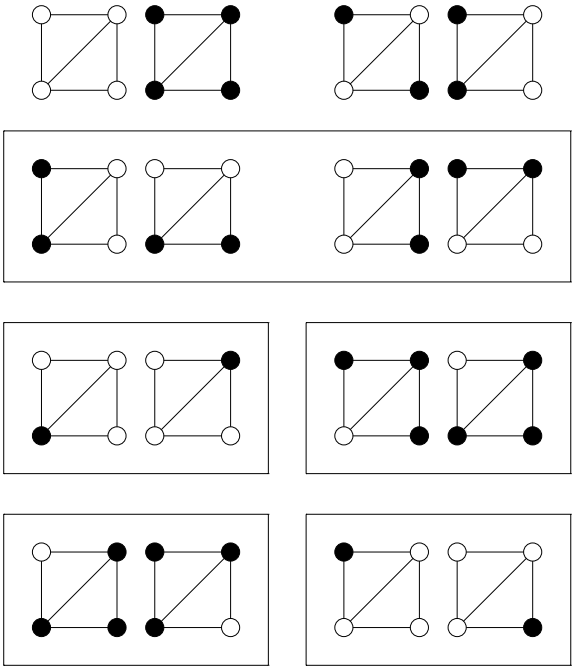


Figura 10.8: Coloraciones equivalentes

Ejercicio 10.64. Demostrar que G_x es un subgrupo de G. Demostrar que todas las permutaciones de una clase lateral hG_x envían x al mismo elemento $y = h(x)$ (si $h_1, h_2 \in hG_x$, entonces $h_1(x) = h_2(x) = y$).

Finalmente, para cada $g \in G$, diremos $fix(g) = \{x \in X \mid g(x) = x\} \subset X$ al conjunto de puntos de X que quedan fijos por g.

Lema 10.65 (Lema de Burnside). El número de órbitas de un grupo de permutaciones G que actúa sobre el conjunto X es

$$\frac{1}{|G|} \sum_{g \in G} fix(g)$$

Demostración. De acuerdo con lo que dice el ejercicio 10.64, si y está en la órbita de x, hay tantas permutaciones que envían x a y como el número de elementos de G_x. Por tanto, $|G| = |G_x| \cdot |O_x|$. Por otra parte, está claro que $1 = \sum_{y \in O_x} (1/|O_x|)$, de manera que el número de órbitas

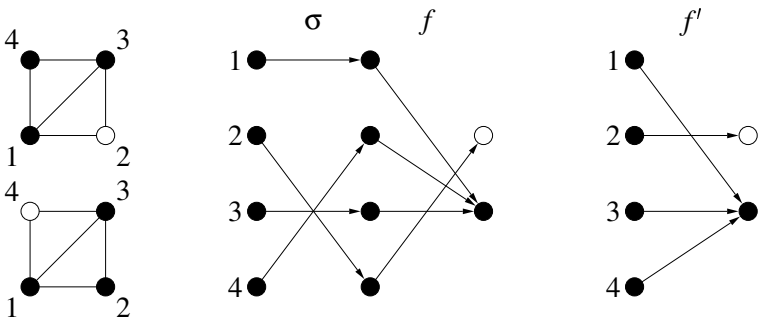


Figura 10.9: Dos coloraciones G-equivalentes

es

$$\sum_{x\in G}\frac{1}{|O_x|}=\frac{1}{|G|}\sum_{x\in G}|G_x|$$

Consideremos ahora los pares $\{(g,x)\in G\times X \mid g(x)=x\}$. Hay dos maneras de contar estos pares: para cada $x\in X$ el número de pares es G_x, mientras que para cada $g\in G$, el número de pares es $fix(g)$. Por tanto, $\sum_{x\in G}|G_x|=\sum_{g\in G}fix(g)$, de donde se deduce el enunciado del lema. □

El lema anterior se puede interpretar diciendo que el número de órbitas coincide con el número medio de puntos fijos de cada permutación de G.

El problema de enumeración que nos ocupa se puede formular en términos del número de órbitas de un cierto grupo de permutaciones. Recordemos que dos coloraciones $f,f'\in C^X$ son G-equivalentes si hay alguna permutación $\sigma\in G$ tal que $f=f'\sigma$, y nuestro objetivo es contar cuántas clases de equivalencia hay.

Podemos interpretar G como un grupo de permutaciones sobre C^X si definimos, para cada $f\in C^X$, $\overline{\sigma}(f)=f\sigma$. Si dos de estas coloraciones f,f' son diferentes, entonces $f\sigma, f'\sigma$ también lo son, de manera que la aplicación $\overline{\sigma}$ es inyectiva. Como C^X es finito, también es biyectiva, es decir, $\overline{G}=\{\overline{\sigma},\sigma\in G\}$ es un grupo de permutaciones de C^X. Observemos finalmente que dos coloraciones son G-equivalentes si y sólo si pertenecen a la misma órbita de $\overline{G}$.

Ejercicio 10.66. Demostrar que, si C tiene más de un color, $\overline{\sigma}=\overline{\sigma}'$ si y sólo si $\sigma=\sigma'$.

La versión más simple del teorema de Pólya es el resultado de aplicar el lema de Burnside al grupo $\overline{G}$. Llamamos $c(\sigma)$ al número de ciclos en la descomposición de σ. Por ejemplo, si $\sigma=(12)(3)(456)$, entonces $c(\sigma)=3$.

Teorema 10.67. Sea G un grupo de permutaciones de $X = \{1, 2, \ldots, n\}$. El número de coloraciones G–diferentes de X con los colores de $C = \{c_1, c_2, \ldots, c_k\}$ es

$$|G(C^X)| = \frac{1}{|G|} \sum_{\sigma \in G} k^{c(g)}$$

Demostración. Como ya hemos mencionado, el número de coloraciones G-diferentes coincide con el número de órbitas de $\overline{G}$. Según el lema de Burnside, este número de órbitas es

$$\frac{1}{|\overline{G}|} \sum_{\overline{\sigma} \in \overline{G}} fix(\sigma)$$

Si $\sigma = (x_{11}x_{12}\ldots x_{1j_1})\cdots(x_{r1}x_{r2}\ldots x_{rj_r})$ es la descomposición de σ en ciclos disyuntos, $\overline{\sigma}(f) = f\sigma = f$ si y sólo si f es constante en cada uno de los ciclos de σ. Si $c(\sigma) = r$ es el número de ciclos de σ, hay $k^{c(\sigma)}$ posibles coloraciones con esta propiedad. Así pues, $|fix(\sigma)| = k^{c(\sigma)}$. El teorema se obtiene entonces observando que $|G| = |\overline{G}|$ (véase el ejercicio 10.66). □

En el ejemplo que hemos tratado antes, $G = \{\iota, (12), (34), (12)(34)\}$ tiene una permutación de cuatro ciclos (la identidad), dos de tres ciclos y una de dos ciclos. Por tanto, el número de coloraciones diferentes con dos colores del grafo del ejemplo es

$$|G(X)^C| = \frac{1}{4}(2^4 + 2^3 + 2^3 + 2^2) = 9$$

Observar que en la aplicación del teorema es preciso considerar también los ciclos de longitud 1. En particular, en el sumatorio aparece siempre el término $|C|^{|X|}$ que corresponde a la contribución de la permutación identidad.

Es preciso notar también que la aplicación de este resultado exige conocer el número y longitud de los ciclos en la descomposición cíclica de cada permutación de G. Esta información se conoce para algunos de los grupos más comunes, pero puede ser difícil de obtener en general. Una manera de condensar esta información sobre la estructura de los ciclos de las permutaciones de G es lo que se llama el *índice de ciclos* del grupo, que se define de la manera siguiente. Si una permutación $\sigma \in G$ tiene λ_1 ciclos de longitud 1, λ_2 ciclos de longitud 2, ..., λ_n ciclos de longitud n, se dice que σ es del tipo $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$. Por ejemplo, la permutación $(12)(3)(456)$ de un grupo de permutaciones de seis elementos es del tipo $(1,1,1,0,0,0)$. Si llamamos $h(\lambda)$ al número de permutaciones de G de tipo λ, el índice de ciclos de G se define como

$$P_G(x_1, x_2, \ldots, x_n) = \frac{1}{|G|} \sum h(\lambda) x_1^{\lambda_1} x_2^{\lambda_2} \cdots x_n^{\lambda_n}$$

donde el sumatorio se extiende a todos los λ posibles, es decir, a todos aquellos que satisfacen $1 \cdot \lambda_1 + 2 \cdot \lambda_2 + \cdots + n \cdot \lambda_n = n$. El índice de ciclos del grupo de automorfismos del grafo de nuestro ejemplo es

$$P_G(x_1, x_2, x_3, x_4) = x_1^4 + 2x_1^2 x_2 + x_2^2 \tag{10.1}$$

El teorema 10.67 se puede expresar en términos del índice de ciclos de G como:

Teorema 10.68. El número de coloraciones G-diferentes de X con los colores de $C = \{c_1, \ldots, c_k\}$ es

$$|G(C^X)| = P_G(k, \ldots, k)$$

Así pues, el conocimiento del polinomio de ciclos de un grupo de permutaciones permite resolver el problema de enumeración que nos hemos planteado.

Ejercicio 10.69. Encontrar el polinomio enumerador de ciclos de un grupo cíclico de orden p donde p es un número primo. Calcular cuántas secuencias de ceros y unos de longitud 7 se pueden formar si consideramos iguales dos secuencias que sólo difieren en una traslación cíclica de los dígitos (por ejemplo, las secuencias 1000000 y 0100000 se consideran iguales).

El teorema de enumeración de Pólya va un poco más allá de los enunciados de los teoremas 10.70 y 10.68, y permite calcular el número de configuraciones diferentes en las cuales aparecen un determinado número de colores. Para ello asociamos a cada coloración f un *peso* en términos de los colores de C de la manera siguiente. Si la coloración f asigna el color c_i a α_i de los elementos de X, $1 \leqslant i \leqslant k$, el peso de f es

$$p(f) = c_1^{\alpha_1} \cdot c_2^{\alpha_2} \cdots c_k^{\alpha_k}$$

Por ejemplo, la coloración de la figura 10.10 tiene peso $p(f) = \bullet^3 \circ^1$.

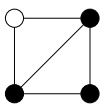


Figura 10.10: Una coloración de peso $p(f) = \bullet^3 \circ^1$

Teorema 10.70 (Pólya, 1935). El número de coloraciones G-diferentes de X con los colores de C que usan α_i veces el color i, $1 \leqslant i \leqslant n$, es el coeficiente de $c_1^{\alpha_1} \cdot c_2^{\alpha_2} \cdots c_k^{\alpha_k}$ en la expresión

$$P_G((c_1 + \cdots + c_k), (c_1^2 + \cdots + c_k^2), \ldots, (c_1^n + \cdots + c_k^n))$$

donde $P_G(x_1, \ldots, x_n)$ es el índice de ciclos de G.

Demostración. El objetivo es ahora contar el número de coloraciones G-diferentes con el mismo peso. Daremos una idea de la demostración sin entrar en los detalles.

Denotamos por $\alpha = (\alpha_1, \ldots, \alpha_k)$ el peso de de la coloración f y denotamos por $(C^X)_\alpha$ las coloraciones de peso α. La aplicación $\tilde{\sigma}(f) = f\sigma$ es una permutación en $(C^X)_\alpha$, de manera que el conjunto $\tilde{G}$ de estas permutaciones es un grupo de permutaciones de $(C^X)_\alpha$. El lema de Burnside nos dice ahora que el número de coloraciones $\tilde{G}$-diferentes de $(C^X)_\alpha$ es

$$\frac{1}{\tilde{G}} \sum_{\tilde{\sigma} \in \tilde{G}} fix(\tilde{\sigma})$$

Las coloraciones fijadas por $\tilde{\sigma}$ son las que toman un valor constante sobre cada ciclo de la permutación σ. Por tanto, el número de coloraciones de peso α que quedan fijadas por una permutación σ de tipo $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)$ es el coeficiente de $c_1^{\alpha_1} \cdot c_2^{\alpha_2} \cdots c_k^{\alpha_k}$ en el desarrollo de

$$(c_1 + \cdots + c_k)^{\lambda_1} (c_2^2 + \cdots + c_k^2)^{\lambda_2} \cdots (c_1. + \cdots + c_k^n)^{\lambda_n}$$

de donde se obtiene el resultado. □

Veamos cómo se aplicaría este teorema en nuestro ejemplo. Recordando el índice de ciclos de G en la ecuación 10.1, tenemos

$$\begin{aligned} & P_G((\circ + \bullet), (\circ^2 + \bullet^2), (\circ^3 + \bullet^3), (\circ^4 + \bullet^4)) \\ &= \frac{1}{4}(\circ + \bullet)^4 + 2(\circ + \bullet)^2(\circ^2 + \bullet^2) + (\circ^2 + \bullet^2)^2 \\ &= \frac{1}{4}(4\circ^4 + 8\circ^3\bullet + 12\circ^2\bullet^2 + 8\circ\bullet^3 + 4\bullet^4) \end{aligned}$$

En la expresión se puede leer el número de coloraciones G-diferentes para cada distribución de colores. Por ejemplo, el coeficiente de $\circ\bullet^3$ nos dice que hay dos coloraciones G-diferentes que usan tres veces el color '$\bullet$' y una vez el color '$\circ$'.

Notas bibliográficas

La teoría de grupos es una disciplina que ha alcanzado una extensión enorme, especialmente en este siglo, y hay por tanto una bibliografía muy extensa. Como ejemplo de monografías

especializadas en el tema (para las cuales este capítulo podría ser una introducción), se puede mencionar el libro de Robinson [3], mientras que el de Wielandt [5] es una referencia clásica de los grupos de permutaciones. A un nivel más accesible y orientado a las aplicaciones, el libro de Stone [4] es una buena referencia. El texto original en el que Pólya introduce su teoría de enumeración se puede encontrar en una traducción inglesa [2]. Finalmente, el libro de Budden [1], de lectura amena e instructiva, hace una revisión bastante exhaustiva de todos los conceptos de la teoría de grupos y presenta algunas aplicaciones insólitas.

Bibliografía

[1] F. J. Budden. *The Fascination of Groups*. Cambridge University Press, 1972.

[2] G. Pólya, R. C. Read. *Combinatorial Enumeration of Groups, Graphs and Chemical Compounds*. Springer-Verlag, 1987.

[3] D. J. S. Robinson. *A Course in The Theory of Groups*. Springer-Verlag, 1982.

[4] H. S. Stone. *Discrete Mathematical Structures and their Applications*. Science Research Asociates, 1973.

[5] H. Wielandt. *Finite Permutation Groups*. Academic Press, 1964.

Problemas

1. Demostrar que un subgrupo de índice dos es siempre normal.

2. Demostrar que en un grupo de orden par siempre existe un elemento diferente del neutro de orden dos.

3. Demostrar que en cualquier grupo G se cumple para todo $a, b \in G$ que

   (a) $|ab| = |ba|$

   (b) $|aba^{-1}| = |b|$

   donde $|g|$ denota el orden de un elemento.

4. Demostrar que si H_1 y H_2 son subgrupos propios de un grupo finito G, entonces H_1H_2 es subgrupo de G si y sólo si $H_1H_2 = H_2H_1$. Demostrar también que

$$|H_1H_2| = |H_1| \cdot |H_2| / |H_1 \cap H_2|$$

5. Demostrar que $\mathbb{Z} \times \mathbb{Z}$ no tiene subgrupos de la forma $\mathbb{Z}_n \times \mathbb{Z}_m$.

6. Demostrar que si H_1 y H_2 son subgrupos normales de los grupos G_1 y G_2 respectivamente, entonces $H_1 \times H_2$ es subgrupo normal de $G_1 \times G_2$ y

$$(G_1 \times G_2)/(H_1 \times H_2) \simeq (G_1/H_1) \times (G_2/H_2)$$

7. Demostrar que si en un grupo finito G se cumple que para cualquier par de subgrupos F y H, $F \subset H$ o bien $H \subset F$, entonces G es cíclico y tiene orden potencia de un primo.

8. Demostrar que todo grupo cociente de un grupo cíclico es cíclico.

9. Demostrar que un grupo de orden $2p$, donde p es un número primo, o bien es cíclico, o bien es isomorfo al grupo diédrico D_p.

10. Demostrar que un grupo G es abeliano si y sólo si la aplicación $\phi : G \longrightarrow G$ dada por $\phi(g) = g^2$ es un endomorfismo de G.

11. Determinar el número de homomorfismos diferentes entre $\mathbb{Z}_2$ y $\mathbb{Z}_3$. Determinar este número, en general para $\mathbb{Z}_n$ y $\mathbb{Z}_m$ en función de n y de m.

12. Demostrar que la signatura de una permutación coincide con la de su inversa.

13. ¿Cuántos ciclos diferentes de longitud n hay en S_n?

14. Demostrar que, en un grupo de permutaciones, o bien la signatura de todas las permutaciones es par, o bien la mitad tiene signatura par y la otra mitad impar. ¿Cuál es la signatura de un grupo de permutaciones de orden impar?

15. Observar que la permutación

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 5 & 2 & 6 & 1 & 3 & 4 \end{pmatrix}$$

se puede expresar como producto de ciclos de longitud 3 como $(642)(531)(432)$. Demostrar que cualquier permutación de signatura par se puede expresar como producto de ciclos de longitud 3 (en general con intersecciones no vacías).

16. Demostrar que el número de permutaciones de S_n que se expresan en notación cíclica como un producto de k_1 1-ciclos, k_2 2-ciclos, en general k_j j-ciclos, $j = 1, \ldots, n$, es

$$\frac{1}{1^{k_1} 2^{k_2} \cdots n^{k_n}} \begin{pmatrix} n \\ k_1, \ldots, k_n \end{pmatrix}$$

siempre que $k_1 + 2k_2 + \cdots + nk_n = n$. ¿Cuántos n-ciclos hay en S_n?

17. Recordar que los números de Stirling de segundo tipo $\left\{ {n \atop k} \right\}$ cuentan el número de subconjuntos de tamaño k de un conjunto de n elementos. Sea ahora $s(n,k)$ el número de maneras de poner n elementos en k-ciclos. Por ejemplo, los elementos de $\{1,2,3,4\}$ se pueden poner en 2-ciclos como

$$\begin{array}{cccc} (123)(4) & (124)(3) & (134)(2) & (234)(1) \\ (132)(4) & (142)(3) & (143)(2) & (243)(1) \\ (12)(34) & (13)(24) & (14)(23) & \end{array}$$

de manera que $s(4,2) = 11$. Demostrar que $s(n,k)$ satisface la ecuación de recurrencia

$$s(n,k) = (n-1)s(n-1,k) + s(n-1,k-1)$$

Usando los resultados de la última sección del capítulo de funciones generadoras, deducir que $s(n,k) = \left[{n \atop k} \right]$, el número de Stirling de primer tipo.

18. Dibujar el grafo de Cayley del grupo alternado de 4 símbolos, A_4 respecto de los generadores $\sigma = (123)$ y $\tau = (12)(34)$.

19. Demostrar que todas las relaciones de un grupo finito G se pueden expresar a partir de las relaciones correspondientes a un conjunto de circuitos fundamentales del digrafo de Cayley $Cay(G,S)$.

20. Considerar el grafo que tiene por vértices los subconjuntos de $X = \{1,2,\ldots,n\}$, y en el que dos vértices son adyacentes cuando los correspondientes subconjuntos difieren en exactamente un elemento. Demostrar que el grafo que se obtiene es isomorfo al grafo de Cayley $Cay(\mathbb{Z}_2 \times \cdots \times \mathbb{Z}_2, \{e_1,\ldots,e_n\})$, donde $e_i = (0,\ldots,0,\overset{i)}{1},0,\ldots,0)$ (este grafo se llama *hipercubo de dimensión n*; ¿pueden imaginar por qué?).

21. Dado el digrafo de Cayley $Cay(G,S)$ dibujado en la figura 10.11, dar una presentación del grupo G.

22. Demostrar que el índice de ciclos de $\mathbb{Z}_6$ es

$$P_{\mathbb{Z}_6}(x_1,x_2,x_3,x_4,x_5,x_6) = \frac{1}{6}(x_1^6 + x_2^3 + 2x_3^2 + 2x_6)$$

Demostrar que, en general, el índice de ciclos de un grupo cíclico $\mathbb{Z}_n$ es

$$P_{\mathbb{Z}_n}(x_1,\ldots,x_n) = \frac{1}{n}\sum_{d|n} \phi(d) x_d^{n/d}$$

donde ϕ es la función de Euler.

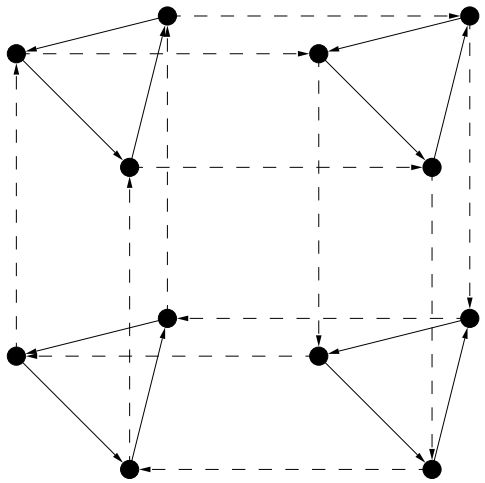


Figura 10.11: $Cay(G,S)$

23. Dos secuencias $(x_1x_2\ldots x_n)$, $(y_1y_2\ldots y_n)$ se dice que son cíclicamente iguales si sólo difieren en una rotación módulo n, es decir, $x_i = y_{i+k \bmod n}$ para $1 \leqslant i \leqslant n$. Por ejemplo, las secuencias 10011 y 00111 son cíclicamente iguales. Esta simetría circular aparece muy frecuentemente. ¿Cuántas secuencias de ceros y unos de longitud n circularmente diferentes hay? ¿Cuántas de estas secuencias de longitud 6 tienen exactamente tres unos?

24. Demostrar que el índice de ciclos del grupo D_3 de simetrías de un triángulo es

$$P_{D_3}(x_1,x_2,x_3) = \frac{1}{6}(x_1^3 + 3x_1x_2^2 + 2x_3^6)$$

En general, demostrar que el índice de ciclos del grupo D_n de simetrías de un polígono regular de n lados es

$$P_{D_n}(x_1,\ldots,x_n) = \frac{1}{2}P_{\mathbb{Z}_n}(x_1,\ldots,x_n) + \frac{1}{4}x_2^{n/2} + \frac{1}{4}x_1^2x_2^{(n/2)-1}$$

si n es par, y

$$P_{D_n}(x_1,\ldots,x_n) = \frac{1}{2}P_{\mathbb{Z}_n}(x_1,\ldots,x_n) + \frac{1}{2}x_1x_2^{(n-1)/2}$$

si n es impar.

25. ¿De cuántas maneras se pueden etiquetar con tres colores los vértices de un polígono regular de n lados si no distinguimos dos maneras que sólo difieren por una simetría del polígono?

26. Demostrar que el índice de ciclos del grupo simétrico S_n es

$$P(S_n;x_1,\ldots,x_n)=\sum_{\lambda}\frac{1}{\lambda_1!2^{\lambda_2}!\cdots n^{\lambda_n}\lambda_n!}x_1^{\lambda_1}x_2^{\lambda_2}\cdots x_n^{\lambda_n}$$

donde el sumatorio se extiende a $\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_n)$ tales que $\lambda_1+2\lambda_2+\cdots+n\lambda_n=n$. (Recordar la fórmula de Cauchy.)

27. Demostrar que el índice de ciclos del grupo alternado A_n es

$$P(A_n;x_1,\ldots,x_n)=\sum_{\lambda}\frac{1+(-1)^{\lambda_2+\lambda_4+\cdots}}{\lambda_1!2^{\lambda_2}!\cdots n^{\lambda_n}\lambda_n!}x_1^{\lambda_1}x_2^{\lambda_2}\cdots x_n^{\lambda_n}$$

donde el sumatorio se extiende a $\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_n)$ tales que $\lambda_1+2\lambda_2+\cdots+n\lambda_n=n$.

Capítulo 11

Anillos y cuerpos

1. Definiciones y propiedades
2. El anillo de los polinomios
3. Cuerpos finitos

En el capítulo anterior se ha estudiado la estructura algebraica más completa definida a partir de una operación, la estructura de grupo. En este capítulo iniciaremos el estudio de estructuras algebraicas definidas a partir de dos operaciones, los anillos y los cuerpos, introducidas en el primer capítulo de esta última parte.

Los sistemas de numeración algebraicamente más completos están construidos a partir de dos operaciones: la suma y el producto. Esto hace que sea importante el estudio de conjuntos que se comporten de forma similar desde el punto de vista algebraico.

La primera sección de este capítulo está dedicada al estudio de las propiedades básicas de los anillos y de los cuerpos. Se introducen las nociones de ideal y anillo cociente, que serán útiles más adelante. La segunda sección se dedica al estudio de un ejemplo importante de anillo, el anillo de los polinomios, que se utilizará para la construcción de cuerpos finitos en la última sección de este capítulo. Las aplicaciones basadas en estas estructuras se estudiarán en el último capítulo.

11.1 Definiciones y propiedades

Recordemos que un *anillo* $A = (A, \star, \circ)$ es una estructura algebraica en la cual A es un conjunto y '$\star$', '$\circ$' son operaciones binarias definidas sobre A que satisfacen las condiciones siguientes:

A1 $(A, \star)$ es un grupo abeliano.

A2 $(A, \circ)$ es un semigrupo.

A3 '$\circ$' es distributiva respecto de '$\star$'. Esto es, para todo $a, b \in A$,

$$\begin{cases} a \circ (b \star c) = (a \circ b) \star (a \circ c) \\ (a \star b) \circ c = (a \circ c) \star (b \circ c) \end{cases}$$

El ejemplo más sencillo y representativo de estructura de anillo lo encontramos en $(\mathbb{Z}, +, \cdot)$, el anillo de los enteros. De hecho, éste es un ejemplo de *anillo unitario*, es decir, admite elemento neutro respecto de la segunda operación. Hay, sin embargo, autores que incluyen dentro de los axiomas de anillo la existencia de este elemento neutro. Esto se debe al hecho de que la mayoría de los anillos más utilizados cumplen este requisito.

Por similitud con $(\mathbb{Z}, +, \cdot)$, cuando tratemos con un anillo unitario cualquiera, en general nos referiremos a la suma y al producto como primera y segunda operación respectivamente y utilizaremos el 0 y el 1 como neutros respectivos. Para abreviar la notación, escribiremos ab en lugar de $a \cdot b$.

Está claro que los axiomas de anillo son una abstracción del comportamiento de los números enteros respecto de las operaciones aritméticas elementales: la suma y el producto. Sin embargo, $(\mathbb{Z}, +, \cdot)$ tiene, además, otras propiedades referidas a la segunda operación que permiten refinar esta estructura. Así, la conmutatividad de esta segunda operación conlleva la estructura de *anillo abeliano*. Esta propiedad no la comparten, sin embargo, todos los anillos, como es el caso del anillo de las matrices cuadradas de orden 2 sobre $\mathbb{Z}$, $(M_2(\mathbb{Z}), +, \cdot)$.

Ejercicio 11.1. Demostrar que $(M_2(\mathbb{Z}), +, \cdot)$ es un anillo unitario no abeliano.

Una clase importante de anillos abelianos unitarios finitos es $(\mathbb{Z}_n, +, \cdot)$, el anillo de los enteros módulo n.

Ejercicio 11.2. Demostrar que $(\mathbb{Z}_n, +, \cdot)$ es un anillo unitario abeliano.

La *ley de simplificación* es otra propiedad importante que cumplen los números enteros, es decir, para todo $a, b, c \in \mathbb{Z}^* = \mathbb{Z} \setminus \{0\}$ se verifica

$$ab = ac \Longrightarrow b = c$$

Esta propiedad está relacionada con la definición siguiente.

Diremos que el anillo $(A, +, \cdot)$ admite *divisores de cero* si existen $a, b \in A \setminus \{0\}$ tales que $ab = 0$.

Ejercicio 11.3. Demostrar que en un anillo A se verifica la ley de simplificación si y sólo si A no tiene divisores de cero.

El anillo $\mathbb{Z}$ de los enteros no tiene divisores de cero, pero es fácil encontrar ejemplos de anillos que sí tienen. En $\mathbb{Z}_6$, por ejemplo, se cumple que $[2][3] = [4][3] = [0]$ y por tanto $[2]$, $[3]$ y $[4]$ son divisores de cero. Cabe observar, sin embargo, que $[2] \neq [4]$. Por ello, en $\mathbb{Z}_6$ no es válida la ley de simplificación.

Se puede comprobar fácilmente que $\mathbb{Z}_3$ o $\mathbb{Z}_5$ no tienen divisores de cero. En el ejercicio siguiente hay clasificados los valores de n para los cuales $\mathbb{Z}_n$ admite la ley de simplificación.

Ejercicio 11.4. Demostrar que $(\mathbb{Z}_n, +, \cdot)$ admite divisores de cero si y sólo si n no es primo.

Un anillo abeliano sin divisores de cero se llama *anillo íntegro* o *anillo de integridad*. Si, además, el anillo es unitario diremos que se trata de un *dominio de integridad*. Así diremos que $(\mathbb{Z}, +, \cdot)$ es un dominio de integridad, mientras que $(\mathbb{Z}_n, +, \cdot)$ en general sólo tiene estructura de anillo unitario abeliano.

El hecho de que, en $\mathbb{Z}_n$, la clase de n sea la clase del cero

$$[n] = [\underbrace{1+1+\cdots+1}_{n}] = [0]$$

sugiere la siguiente abstracción relativa a los anillos unitarios abelianos en general.

Se define la *característica* de un anillo unitario abeliano A como el mínimo número natural n tal que

$$n \cdot 1 = \underbrace{1+1+\cdots+1}_{n} = 0$$

en A. Si este número no existe, se dice que el anillo tiene característica cero. De hecho, podríamos interpretar la característica de un anillo unitario abeliano como el orden del subgrupo aditivo generado por el 1.

Está claro que la característica de $\mathbb{Z}_n$ es n. Un ejemplo también claro de anillo de característica cero lo encontamos en $\mathbb{Z}$.

Proposición 11.5. La característica de un dominio de integridad es cero o es un número primo.

Demostración. Sea A un dominio de integridad de característica $n_0 \neq 0$. Si existiesen $a, b \in \mathbb{N}$ tales que $n_0 = ab$, entonces querría decir que

$$(\underbrace{1+\cdots+1}_{a})(\underbrace{1+\cdots+1}_{b}) = \underbrace{1+\cdots+1}_{n_0} = 0$$

y A tendría divisores de cero, en contra de lo que hemos supuesto. □

Ejercicio 11.6. Demostrar que si la característica de A es p, entonces

1. $n \cdot 1 = \underbrace{1+1+\cdots+1}_{n} = 0$ implica que p divide a n;
2. para todo $a \in A$, $p \cdot a = \underbrace{a+a+\cdots+a}_{p} = 0$.

Finalmente, si los elementos no nulos de un anillo tienen estructura de grupo abeliano respecto del producto, diremos que este anillo es un *cuerpo* y lo notaremos habitualmente con la letra K. Dicho de una otra manera, $(K,+,\cdot)$ es un cuerpo si $(K,+)$ y $(K^*,\cdot)$ son grupos abelianos y el producto es distributivo respecto de la suma.

Es fácil observar que en un cuerpo K no pueden existir elementos a y b de K^* tales que $ab = 0$, ya que, multiplicando por la izquierda por a^{-1}, deduciríamos que b tiene que ser cero, en contra de lo que hemos supuesto. Por tanto, podemos afirmar lo siguiente:

Proposición 11.7. Todo cuerpo es un dominio de integridad.

El ejemplo más pequeño de cuerpo lo encontramos en $\mathbb{Z}_2$. De hecho, es fácil obtener toda una familia de cuerpos finitos no triviales, como muestra el resultado siguiente:

Proposición 11.8. $(\mathbb{Z}_p,+,\cdot)$ es un cuerpo si y sólo si p es primo.

Demostración. Si $(\mathbb{Z}_p,+,\cdot)$ es un cuerpo, tiene que ser un dominio de integridad. Por el ejercicio 11.4, p tiene que ser primo. En este caso, basta con ver que cada elemento tiene inverso. Para cada $b \in \mathbb{Z}_p$, el conjunto $b\mathbb{Z}_p = \{bx, x \in \mathbb{Z}_p\}$ tiene p elementos diferentes, ya que se satisface la ley de simplificación. Como $b0 = 0$, existe $x \in \mathbb{Z}_p \setminus \{0\}$ tal que $bx = 1$ y por tanto x es inverso de b. □

Ejercicio 11.9. Adaptar esta última demostración para ver que todo dominio de integridad finito es un cuerpo.

Otros ejemplos bien conocidos de cuerpos no finitos de característica cero los encontramos en $(\mathbb{Q},+,\cdot)$, $(\mathbb{R},+,\cdot)$ o $(\mathbb{C},+,\cdot)$. En la última sección de este capítulo construiremos nuevas familias de cuerpos finitos de característica p a partir del ya conocido $\mathbb{Z}_p$. Para poder hacer estas construcciones es necesaria la noción de subanillo.

Se dice que un subconjunto B de un anillo $(A,+,\cdot)$ es un *subanillo* de A si con las operaciones '$+$' y '$\cdot$' restringidas a los elementos de B se satisfacen los axiomas de anillo. Así pues, B tiene que ser un subgrupo del grupo aditivo de A y una parte estable de A por la multiplicación.

Proposición 11.10. Sea $(A,+,\cdot)$ un anillo y $B \subset A$. Entonces para que $(B,+,\cdot)$ sea subanillo de $(A,+,\cdot)$ es necesario y suficiente que $(B,+)$ sea subgrupo de $(A,+)$ y que el producto sea cerrado en B.

Ejercicio 11.11. Demostrar que $\mathbb{Z}$ es un subanillo de $\mathbb{Q}$ considerado como anillo.

Ejercicio 11.12. Determinar los subanillos de $\mathbb{Z}_5$ y $\mathbb{Z}_6$.

Ejercicio 11.13. Caracterizar en general los subanillos de $\mathbb{Z}_n$.

Es fácil observar que, si A es un anillo abeliano o íntegro, también lo es el subanillo B; pero A puede ser unitario sin que lo sea B, como muestra el resultado siguiente.

Proposición 11.14. Los subanillos de $(\mathbb{Z},+,\cdot)$ son los conjuntos $n\mathbb{Z}$.

Demostración. Tal como se vio en el capítulo anterior, los únicos subgrupos de $(\mathbb{Z},+)$ son los conjuntos $n\mathbb{Z} = \{nk, k \in \mathbb{Z}\}$, conjunto de múltiplos de un entero n. Estos conjuntos son claramente estables por la multiplicación y por tanto son los únicos subanillos de $\mathbb{Z}$. □

Observar que para $n \geqslant 2$, $n\mathbb{Z}$ no es unitario.

Ideales y anillo cociente

Siguiendo un proceso paralelo al descrito en el capítulo anterior para la obtención de la estructura de grupo cociente, definiremos aquí la noción de anillo cociente. Por ello introducimos en primer lugar la noción de relación de equivalencia compatible con la estructura de anillo.

Diremos que una relación de equivalencia R es compatible por la derecha con la estructura de anillo si y sólo si R es compatible con la suma y el producto de este anillo. Es decir, para todo $a, b \in A$ y para cualquier elemento $x \in A$ se cumple que

$$aRb \iff \begin{cases} (a+x)R(b+x) \\ axRbx \end{cases}$$

De forma similar, R es compatible por la izquierda si la segunda condición es $xaRxb$. Sabemos que una relación R que cumpla la primera de estas condiciones tiene la forma $aRb \Leftrightarrow a-b \in B$, donde B es un subgrupo aditivo de $(A,+)$. Para la segunda de estas condiciones necesitamos también que $x(a-b)$ o $(a-b)x$ sean elementos de B para cualquier $x \in A$. Esto obliga a restringir las relaciones de equivalencia compatibles con la estructura de anillo a ciertos subgrupos del grupo aditivo, llamados *ideales por la izquierda* o *por la derecha* según sea la relación.

Un subconjunto $I \subset A$ es un *ideal* por la *derecha* del anillo A si $(I,+)$ es un subgrupo de $(A,+)$ y para todo $a \in I$ y para todo $x \in A$ se cumple $ax \in I$. El ideal es por la *izquierda* si esta segunda condición es $xa \in A$. Por ejemplo, $\{0\}$ es un ideal de cualquier anillo, como también lo es el anillo entero A. Los ideales diferentes de $\{0\}$ y A se llaman *ideales propios*.

De forma similar al caso de grupos, si un ideal por la izquierda coincide con el correspondiente ideal por la derecha, se dice que el ideal es *bilateral*. Los ideales bilaterales juegan en

los anillos un papel similar al de los subgrupos normales en los grupos. Observemos que si A es un anillo abeliano, sus ideales son bilaterales.

Es preciso tener en cuenta que es posible que un ideal tenga estructura de anillo no unitario, aunque provenga de un anillo con unidad. Este es el caso de los ideales de $\mathbb{Z}$, $n\mathbb{Z}$.

Ejercicio 11.15. Demostrar que $n\mathbb{Z}$ son los únicos ideales de $\mathbb{Z}$.

Hemos visto que las únicas relaciones de equivalencia compatibles con la estructura de anillo son de la forma $aRb \Leftrightarrow a-b \in I$, donde I es un ideal del anillo. Así, como en el caso de los grupos, las clases de equivalencia inducidas a partir de una de estas relaciones serán de la forma $a+I$ con $a \in A$ y las notaremos como $[a]_I = \{a+I, a \in A\}$ (o simplemente $[a]$ si la referencia al ideal se sobreentiende). El conjunto formado por estas clases se llama *conjunto cociente módulo I* y se representa por $A/I = \{[a], a \in A\}$. Si I es un ideal bilateral, el conjunto A/I tiene estructura de anillo con las operaciones inducidas de A y se llama *anillo cociente módulo I*. También se conoce como *anillo factor* de A respecto I.

Ejercicio 11.16. Si I es un ideal bilateral de un anillo A, comprobar que las operaciones siguientes están bien definidas, para todo $[a],[b] \in A/I$:

$$\begin{cases} [a]+_I[b] &= [a+b] \\ [a]\cdot_I[b] &= [ab] \end{cases}$$

La comprobación del resultado siguiente es rutinaria y la dejamos como ejercicio para el lector.

Proposición 11.17. Si I es un ideal bilateral de un anillo A, entonces $(A/I, +_I, \cdot_I)$ es un anillo.

Observar que los anillos cocientes de $\mathbb{Z}$ son justamente los $\mathbb{Z}_n$.

En lo que sigue consideraremos anillos abelianos, de manera que los ideales serán bilaterales y nos referiremos a ellos simplemente como ideales.

Como en el caso de subgrupos, la intersección de ideales es también un ideal. En particular tiene un interés especial considerar la intersección de todos los ideales que contienen un determinado subconjunto $X \subset A$. Entonces se dice que este ideal está *generado* por X y se denota por $I = (X)$. En otras palabras, el ideal generado por X es el ideal más pequeño que contiene a X. Los ideales generados por un solo elemento tienen un interés especial y se llaman *ideales principales*.

Observemos que el ideal generado por un solo elemento $a \in A$ tiene que contener los elementos de la forma na para cualquier $n \in \mathbb{Z}$, ya que tiene que ser subgrupo del grupo aditivo de A. También tiene que contener los elementos de la forma xa para todo $x \in A$. Por tanto,

tiene que contener todo elemento de la forma $na+xa$, con $n \in \mathbb{Z}$ y $x \in A$. Estos elementos son suficientes para formar un ideal, ya que

$$\begin{aligned}(na+xa)-(n'a+x'a) &= (n-n')a+(x-x')a\\ y(na+xa) &= (yn)a+(yx)a\end{aligned}$$

Observar que $n-n' \in \mathbb{Z}$ y que $(x-x'),\ yn,\ yx \in A$.

Si el anillo A es unitario podemos identificar $n \in \mathbb{Z}$ con $n \cdot 1 \in A$, que es el elemento que consiste en sumar n veces el neutro del producto de A. De donde, $na = (n \cdot 1)a \in Aa$. De aquí obtenemos el resultado siguiente:

Proposición 11.18. En un anillo unitario abeliano A, los ideales generados por $a \in A$ son de la forma aA.

En general, el ideal generado por un conjunto de elementos de un anillo unitario abeliano, $\{a_1, a_2, \ldots, a_n\} \subset A$ está descrito por los elementos de la forma

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n \qquad x_i \in A, \forall i$$

Así, por ejemplo, el ideal de $\mathbb{Z}$ generado por el 2 y por el 3 es de la forma

$$\{2x+3y,\ x,y \in \mathbb{Z}\}$$

Hay anillos en que todos sus ideales son principales. En este caso se dice que el *anillo* es *principal*. Este es el caso de $\mathbb{Z}$ que tomamos como ejemplo sencillo, reiterado e ilustrativo de la mayor parte de las cuestiones consideradas en esta sección. Cabe observar que la noción de anillo principal va ligada a los anillos unitarios abelianos, sobre los cuales, como hemos mencionado anteriormente, trabajaremos a partir de ahora.

Ejercicio 11.19. Demostrar que $\mathbb{Z}$ es un anillo principal.

Ejercicio 11.20. Demostrar que en $2\mathbb{Z}$, el ideal I generado por 4 no es $\{4x, x \in 2\mathbb{Z}\}$. ¿Por qué?

Está claro que si I es un ideal de un anillo A, también es ideal de todos los subanillos B de A que lo contengan. Sin embargo, es preciso observar que en sentido contrario no es cierto. Por ejemplo, $n\mathbb{Z}$ es un ideal de $\mathbb{Z}$, pero no es un ideal de $\mathbb{Q}$.

Morfismos de anillos

Diremos que una aplicación f de un anillo A sobre un anillo A' es un *morfismo entre anillos*, o bien un *homomorfismo de anillos*, si y sólo si f respeta la suma y el producto. Es decir, para todo $a, b \in A$,

$$\begin{aligned}f(a +_A b) &= f(a) +_{A'} f(b)\\ f(a \cdot_A b) &= f(a) \cdot_{A'} f(b)\end{aligned}$$

Es preciso observar que, de hecho, no es necesario suponer que A' sea un anillo; es suficiente considerar A' como un conjunto con suma y producto, como pone de manifesto el ejercicio siguiente.

Ejercicio 11.21. Comprobar que si A es un anillo y $f : A \longrightarrow A'$ es un morfismo, donde A' es un conjunto con suma y producto, entonces

1. $f(A)$ es una parte estable de A';
2. $(f(A), +_{A'}, \cdot_{A'})$ es un anillo.

Es fácil comprobar que ciertas propiedades algebraicas de A se transmiten a través de f tal como se indica en los ejercicios siguientes.

Ejercicio 11.22. Demostrar que, si f es un morfismo entre los anillos A y $f(A)$, entonces

1. $f(0) = 0$ y $f(-a) = -f(a)$;
2. $f^{-1}(0)$ es un ideal de A, llamado *núcleo de* f.

Ejercicio 11.23. Demostrar que si A es un anillo unitario abeliano, entonces

1. $f(A)$ es también abeliano;
2. $f(1) = 1$;
3. $f(a^{-1}) = (f(a))^{-1}$, siempre que $a^{-1} \in A$.

Recordemos que los subgrupos normales están íntimamente relacionados con los morfismos entre grupos. En el caso de anillos, serán los ideales los que jugarán un papel similar.

Así, la aplicación $f : A \longrightarrow A/I$, donde I es un ideal de A tal que $f(x) = x + I$ para todo $x \in A$, es un homomorfismo exhaustivo o *epimorfismo*, llamado *homomorfismo canónico*.

Ejercicio 11.24. Demostrar que si $f : A \longrightarrow A/I$ es homomorfismo canónico, entonces el núcleo de f es el ideal I.

Un morfismo entre anillos es inyectivo si y sólo si $f^{-1}(0) = \{0\}$. En este caso se dice que f es un *monomorfismo*.

Además, como en el caso de grupos, si el morfismo es biyectivo se dice que los anillos son *isomorfos*.

Los homomorfismos conservan también los ideales en un cierto sentido, como se puede comprobar mediante el siguiente resultado que usaremos más adelante.

Proposición 11.25. Si $f : A \longrightarrow A'$ es un homomorfismo de anillos e I' es un ideal de A', entonces $I = f^{-1}(I')$ es un ideal de A que contiene al núcleo de f.

Demostración. Si $a, b \in I$, entonces $f(a-b) = f(a) - f(b) \in I'$ de manera que $a - b \in I$. Similarmente, para todo $x \in A$, $f(ax) = f(a)f(x) \in I'$ de manera que $ax \in I$, e I es un ideal de A que contiene a $f^{-1}(0)$. □

Ideales maximales y cuerpos

El hecho que $\mathbb{Z}/n\mathbb{Z} = \mathbb{Z}_n$ sea un anillo unitario abeliano, y que para determinados valores de n tenga estructura de cuerpo, no es un hecho particular de $\mathbb{Z}$, sino que lo comparte con cualquier anillo con las mismas características. Encontrar condiciones por las cuales un anillo cociente es un cuerpo es el objetivo que nos proponemos en esta parte.

Un ideal I en A se dice *maximal* si $I \neq A$ y no existe ningún otro ideal entre I y A.

Ejercicio 11.26. Demostrar que el ideal $(p) = p\mathbb{Z}$ es maximal en $\mathbb{Z}$ si y sólo si p es primo.

Una manera útil de obtener cuerpos a partir de un anillo unitario abeliano consiste en localizar sus ideales maximales y construir el correspondiente anillo cociente. Éste es justamente el argumento utilizado en la última sección de este capítulo para la construcción de cuerpos finitos. Por ello vemos primero cuáles son los ideales de un cuerpo.

Proposición 11.27. Un anillo unitario abeliano K es un cuerpo si y sólo si sus únicos ideales son $\{0\}$ y K.

Demostración. Supongamos primero que K es un cuerpo e $I \neq \{0\}$ es un ideal de K. Entonces, si $a \in I$, $a^{-1}a = 1 \in I$ y por tanto $I = K$.

Recíprocamente, sea $I = (a)$ el ideal generado por $a \in K^* = K \setminus \{0\}$. Si $I = aK = K$, entonces existe $x \in K^*$ tal que $ax = 1$, de manera que x es el inverso de a. □

Esta última propiedad permite ver el resultado siguiente:

Proposición 11.28. Si M es un ideal maximal de un anillo unitario abeliano A, entonces A/M es un cuerpo.

Demostración. Consideremos el epimorfismo canónico $f : A \longrightarrow A/M$. Ya hemos visto en el ejercicio 11.23 que A/M es abeliano y unitario. Según la proposición 11.25, si $[J]$ es un ideal de A/I, entonces $I = f^{-1}([I])$ es un ideal de A que contiene a M. Por tanto, o bien $I = A$ y $[I] = A/M$, o bien $I = M$ y $[I] = [0]$. Así, A/M no tiene ideales propios y por tanto es un cuerpo. □

11.2 El anillo de los polinomios

En esta sección se estudia un ejemplo importante de anillo, el anillo de los polinomios, que se utilizará en la sección siguiente para la construcción de cuerpos de orden finito.

Normalmente se interpretan los polinomios como expresiones formales del tipo "$a_0+a_1x+\cdots+a_nx^n$" con "indeterminada" x. El origen de esta expresión se pondrá de manifesto al final de esta sección. De momento nos preguntamos, ¿qué representa esta x y cómo se hace para sumarla o multiplicarla? Podemos resolver esta cuestión introduciendo los polinomios de una manera aparentemente más formal, pero más precisa y más útil. La respuesta parte de una observación sencilla: un polinomio "$a_0+a_1x+\cdots+a_nx^n$" queda determinado por la secuencia $(a_0,a_1,\dots,a_n)$.

Definimos el conjunto de los *polinomios* sobre un anillo unitario abeliano A como el conjunto de todas las sucesiones de elementos de A que tienen un número finito de elementos no nulos

$$a=(a_0,a_1,\dots,a_n,0,0,\dots)$$

Diremos que $a_0,a_1,\dots,a_n$ son los *coeficientes* del polinomio a. Si n es el entero más grande para el cual $a_n\neq 0$, diremos que el polinomio a tiene *grado* n y lo notaremos escribiendo $gr(a)=n$. Si $a_n=1$ diremos que el polinomio a es *mónico*. A los polinomios de grado cero se los llama *constantes*. Es preciso observar que el polinomio nulo, el que tiene todos sus coeficientes cero, que denotaremos directamente como 0, no tiene grado según esta regla, pero se interpreta también como un polinomio constante y se dice, formalmente, que tiene grado $-\infty$.

Definiremos dos operaciones, la suma y el producto, que permitirán estructurar este conjunto como el propio anillo A sobre el cual se ha contruido.

Dados dos polinomios $a=(a_0,a_1,\dots)$ y $b=(b_0,b_1,\dots)$ con coeficientes en un anillo unitario abeliano A, definimos el *polinomio suma* y el *polinomio producto*

$$\begin{aligned} a+b &= (a_0+b_0,a_1+b_1,\dots)\\ ab &= (c_0,c_1,\dots), \qquad c_k=\sum_{i+j=k}a_ib_j=\sum_{i=0}^{k}a_ib_{k-i}\end{aligned}$$

Observar que, en general,

$$\begin{aligned} gr(a+b) &\leqslant \max\{gr(a),gr(b)\}\\ gr(ab) &\leqslant gr(a)+gr(b)\end{aligned}$$

Ejercicio 11.29. Si $a=(3,3,6,0,\dots)$ y $b=(3,-3,3,-1,0,\dots)$ son polinomios en $\mathbb{Z}_7[x]$, calcular los polinomios $a+b$ y ab. Repetir el cálculo si los polinomios a y b se consideran en $\mathbb{Z}_9[x]$.

La suma y el producto de polinomios involucra sólo sumas y productos de elementos del anillo de base A. Teniendo en cuenta esta observación, es fácil deducir que los polinomios respecto de la suma se comportan como grupo abeliano con el polinomio 0 como elemento neutro y que respecto del producto se comportan como un semigrupo abeliano con elemento neutro el polinomio constante $1 = (1, 0, \dots)$. La distributividad del producto respecto de la suma es también consecuencia directa de las observaciones anteriores. Así, el conjunto de polinomios con coeficientes en un anillo unitario abeliano tiene también estructura de anillo unitario abeliano.

La justificación de la notación clásica a que aludíamos al comienzo de esta sección descansa en el hecho de identificar el polinomio $(0, 1, 0, \dots)$ con x. De esta manera tenemos que, $x \cdot x = x^2 = (0, 0, 1, 0, \dots)$, $x^3 = (0, 0, 0, 1, 0, \dots)$ y así sucesivamente y, por convenio, $x^0 = (1, 0, \dots) = 1$. Por tanto, podemos escribir

$$a = (a_0, a_1, \dots, a_n, 0, \dots) = a_0 + a_1 x + \cdots + a_n x^n = \sum_{i=0}^{n} a_i x^i$$

Cabe observar que ahora x no es una "indeterminada", sino un polinomio especial. Es habitual, como consecuencia de esta expresión, denotar el conjunto de los polinomios con coeficientes en A por $A[x]$ y los polinomios de $A[x]$ por $a(x)$, para diferenciarlos, si es necesario, de los propios elementos del anillo, $a \in A$. Observemos también que A se puede interpretar como el conjunto de los polinomios constantes, es decir, cada elemento $a \in A$ se asocia con el polinomio $a(x) = a + 0x + 0x^2 + \cdots \in A[x]$. Identificaremos por tanto los polinomios constantes con los elementos del anillo.

Los resultados siguientes nos garantizan en qué condiciones está permitida la ley de simplificación para los polinomios.

Lema 11.30. Si A es un dominio de integridad y $a(x), b(x) \in A[x]$, entonces $gr(a(x)b(x)) = gr(a(x)) + gr(b(x))$.

Demostración. Si $a(x)$ y $b(x)$ tienen grados $n, m \geqslant 0$ respectivamente, entonces el coeficiente de grado $m+n$ de $c(x) = a(x)b(x)$ es $c_{m+n} = a_n b_m \neq 0$, ya que a_n y b_m son no nulos y A no tiene divisores de cero. Si alguno de los grados es $-\infty$, entonces $a(x)b(x) = 0$ y también vale la igualdad. □

Cabe observar que en esta demostración ha sido útil la asignación de $-\infty$ como grado del polinomio nulo en lugar de asignarle grado cero, como haríamos con los otros polinomios constantes.

Proposición 11.31. Si A es un dominio de integridad, $A[x]$ también lo es.

Demostración. Sabemos que $A[x]$ es un anillo unitario abeliano. Tenemos que ver entonces que, si A es íntegro, también lo es $A[x]$. Si consideramos $a(x), b(x) \in A[x]$ tales que $a(x)b(x) = 0$, como $gr(ab) = gr(a) + gr(b)$, deducimos que $a(x) = 0$ o $b(x) = 0$. □

El resultado siguiente demuestra la imposibilidad de ampliar la estructura algebraica de los polinomios para obtener la estructura de cuerpo, demostrando la imposibidad de obtener inversos para todos los elementos de $A[x]$, independientemente de las propiedades de A.

Proposición 11.32. Si A es un anillo íntegro unitario, los únicos elementos invertibles de $A[x]$ son los elementos invertibles de A.

Demostración. $a(x), b(x) \in A[x]^* = A[x] \setminus \{0\}$ son inversos si y sólo si $a(x)b(x) = 1$. Como $gr(ab) = gr(a) + gr(b) = 0$, deducimos que $gr(a) = gr(b) = 0$ y por tanto $a(x) = a \in A$ y $b(x) = a^{-1}$. □

Divisibilidad en K[x]

Una vez definido y estructurado el conjunto de polinomios $A[x]$ con coeficientes sobre un anillo unitario abeliano A, nos interesa factorizar estos polinomios de forma similar a como factorizamos los números enteros, es decir, descomponiéndolos como productos de elementos tan "simples" como sea posible. En el caso de los enteros, sabemos que estos elementos son los números primos. Como veremos a continuación, en el caso de los polinomios, estos elementos "simples" se llaman también polinomios *primos*. Para ello necesitamos introducir las definiciones, las notaciones y las propiedades correspondientes al concepto de divisibilidad en el ámbito de $A[x]$, y observaremos que son similares a las propias en el caso de $\mathbb{Z}$.

Dados dos polinomios $a(x)$ y $b(x)$ de $A[x]$, diremos que $b(x)$ es un *divisor* de $a(x)$, o que $b(x)$ *divide* a $a(x)$, y lo denotaremos escribiendo

$$b(x)|a(x)$$

si y sólo si existe un polinomio $c(x) \in A[x]$ tal que $a(x) = b(x)c(x)$.

Observemos que los polinomios constantes correspondientes a los elementos invertibles de A, $U = \{u \in A,\ \exists u' \in A,\ uu' = 1\}$ son divisores de cualquier polinomio $a(x) \in A[X]$, ya que $a(x) = uu'a(x)$, donde u' es el inverso de u. Por el mismo motivo, $ua(x)$ es divisor de $a(x)$ para todo $u \in U$. Éstos se llaman *divisores triviales*. Es preciso observar que, sea cual sea A, exceptuando $\mathbb{Z}_2$, como mínimo se tienen asegurados ± 1 y $\pm a(x)$ como divisores triviales de $a(x)$. En $\mathbb{Z}[x]$ éstos son los únicos, mientras que si A es un cuerpo, cualquiera de sus elementos, salvo el cero, es un divisor trivial de $a(x)$. Por otra parte, si $gr(a) > gr(b) > 0$, diremos que $b(x)$ es un *divisor propio* de $a(x)$.

Todo polinomio que no tiene divisores propios, es decir, que no tiene otros divisores que los triviales, se dice que es *primo* o *irreductible*. Está claro que todo polinomio de grado 1 es irreductible. En el cuerpo de los números complejos $\mathbb{C}$, el *teorema fundamental del álgebra* nos garantiza que éstos son los únicos polinomios irreductibles de $\mathbb{C}[x]$. Desgraciadamente, no hay resultados teóricos tan sencillos para conocer los polinomios irreductibles en otros anillos de polinomios. Por ello, nos tendremos que conformar con resultados parciales que nos ayudarán a estudiar la situación en cada caso.

En nuestro contexto, nos interesan particularmente los polinomios definidos sobre un cuerpo y en especial sobre cuerpos finitos. En lo que sigue centraremos nuestra atención en el caso particular de estos polinomios.

A continuación enunciaremos un resultado teórico, el *teorema de factorización*, esencial para todo lo que sigue, cuya demostración evitaremos, ya que no tiene ningún interés práctico y es excesivamente laboriosa[1].

Teorema 11.33 (Teorema de factorización). Dado un cuerpo K, cada polinomio $k(x) \in K[x]$ admite una representación única de la forma

$$k(x) = kp_1(x)p_2(x)\cdots p_m(x)$$

con $k \in K$, $p_1(x), p_2(x), \ldots, p_m(x) \in K[x]$ polinomios mónicos irreductibles.

Esta descomposición en factores primos tiene en la práctica una dificultad fundamental: la inexistencia de métodos sencillos para encontrar en general estos polinomios irreductibles. Los conceptos y los resultados siguientes están dedicados al estudio de este problema.

Sea K un cuerpo. Para cualquier par de polinomios $a(x)$ y $b(x)$ de $K[x]$ se define su *máximo común divisor* como el polinomio de grado más grande que los divide a ambos. Es decir, $D(x) \in K[x]^* = K[x] \setminus \{0\}$ es un máximo común divisor de los polinomios $a(x)$ y $b(x)$, y escribimos $D(x) = \mathrm{mcd}(a(x), b(x))$ si se verifican las condiciones siguientes:

1. $D(x)|a(x)$ y $D(x)|b(x)$;
2. si $D'(x)|a(x)$ y $D'(x)|b(x)$, entonces $D'(x)|D(x)$.

Es preciso observar que, tal como está definido el máximo común divisor de dos polinomios, éste no es único, ya que si $D(x)$ es un máximo común divisor, entonces $kD(x)$ también lo es, para todo $k \in K$. Tiene sentido, por tanto, escoger el polinomio más sencillo que represente toda esta familia. Éste es el polinomio mónico correspondiente, que denotaremos como

$$d(x) = \mathrm{mcd}(a(x), b(x))$$

[1] El lector que esté interesado puede encontrar esta demostración en [3], por ejemplo.

Los factores en común que tienen dos polinomios en sus factorizaciones son los factores de su máximo común divisor. Diremos que dos polinomios $a(x), b(x) \in K^*[x]$ son *primos entre sí* o *coprimos* si $\mathrm{mcd}(a(x), b(x)) = 1$.

De forma similar, se define el *mínimo común múltiplo* entre $a(x)$ y $b(x)$, $M(x) = \mathrm{mcm}(a(x), b(x))$, si se verifican las condiciones siguientes:

1. $a(x)|M(x)$ y $b(x)|M(x)$;
2. si $a(x)|M'(x)$ y $b(x)|M'(x)$, entonces $M(x)|M'(x)$.

Como antes, se define $m(x) = \mathrm{mcm}(a(x), b(x))$ como el polinomio mónico de entre los que satisfacen las dos propiedades anteriores.

El teorema de factorización proporciona una manera teórica de encontrar el máximo común divisor de dos polinomios: sólo es preciso seleccionar los factores comunes de sus factorizaciones. En la práctica, sin embargo, resulta en general difícil encontrar estas factorizaciones. Es importante, por tanto, disponer de algoritmos eficientes que permitan la obtención directa de estos máximo común divisores. El teorema de Euclides, que enunciaremos a continuación, y una consecuencia inmediata de éste, conducirán a un algoritmo de estas características: el algoritmo de Euclides.

Teorema 11.34 (Euclides). Dados $a(x), b(x) \in K^*[x]$, existen dos únicos polinomios $q(x), r(x) \in K[x]$ tales que

$$a(x) = b(x)q(x) + r(x), \qquad gr(r) < gr(b)$$

La expresión anterior es la llamada *división euclídea*, de $a(x)$ por $b(x)$. En general, los anillos para los cuales es válida la división euclídea se llaman anillos *euclídeos*. La demostración del teorema anterior prueba la existencia de los polinomios $q(x), r(x) \in K[x]$, llamados respectivamente *cociente* y *resto*. Así, $K[x]$ es euclídeo. El resultado siguiente es una consecuencia importante del teorema anterior.

Corolario 11.35. $\mathrm{mcd}(a(x), b(x)) = \mathrm{mcd}(b(x), r(x))$.

Demostración. Si $d(x) = \mathrm{mcd}(a(x), b(x))$, significa que $a(x) = d(x)a_1(x)$ y $b(x) = d(x)b_1(x)$ para ciertos polinomios $a_1(x), b_1(x)$. De la igualdad del teorema de Euclides, $a(x) = b(x)q(x) + r(x)$, deducimos que $r(x) = d(x)(a_1(x) - b_1(x)q(x))$, de donde $d(x)|r(x)$. Por otra parte, de la misma igualdad se deduce que cualquier divisor común $d'(x)$ de $b(x)$ y $r(x)$ divide también a $a(x)$. □

Algoritmo de Euclides

En las condiciones del teorema de Euclides podemos usar el corolario anterior tantas veces como sea posible, es decir, hasta que obtengamos como resto el polinomio nulo. Esto es,

$$\begin{array}{lcll} a(x) & = & b(x)q_1(x)+r_1(x), & gr(r_1)<gr(b) \\ b(x) & = & r_1(x)q_2(x)+r_2(x), & gr(r_2)<gr(r_1) \\ r_1(x) & = & r_2(x)q_3(x)+r_3(x), & gr(r_3)<gr(r_2) \\ & \vdots & & \\ r_{n-2}(x) & = & r_{n-1}(x)q_n(x)+r_n(x), & gr(r_n)<gr(r_{n-1}) \\ r_{n-1}(x) & = & r_n(x)q_{n+1}(x)+0 & \end{array}$$

Aplicando reiteradamente el corolario anterior a las sucesivas expresiones, obtenemos

$$\mathrm{mcd}(a(x),b(x)) = \mathrm{mcd}(b(x),r_1(x)) = \cdots = \mathrm{mcd}(r_n(x),0) = r_n(x)$$

Es decir, el máximo común divisor es justamente el último resto diferente de cero.

Como ejemplo, podemos comprobar que $a(x) = x^4+3x^2-4x+1$ y $b(x) = x^2-3x$ son polinomios primos entre sí en $\mathbb{Z}_5$:

$$\begin{array}{rcl} x^4+3x^2+x+1 & = & (x^2+2x)(x^2+3x+2)+(2x+1) \\ x^2+2x & = & (2x+1)(3x+2)+3 \\ 2x+1 & = & 3(4x+2)+0 \end{array}$$

La simplicidad y la utilidad de este algoritmo son bien evidentes y hacen innecesario cualquier elogio. Una consecuencia directa también muy importante de este resultado es la llamada *identidad de Bézout*, que esencialmente consiste en "recorrer" el algoritmo de Euclides en sentido contrario, mediante substituciones sucesivas de los restos. Más concretamente, en el algoritmo anterior podemos escribir

$$r_n(x) = r_{n-2}(x) - r_{n-1}(x)q_n(x)$$

De la misma manera, cada resto $r_i(x)$ se puede expresar en términos de restos anteriores hasta obtener una expresión de $r_n(x)$ en términos de los polinomios iniciales $a(x)$ y $b(x)$.

Es preciso observar que el polinomio $r_n(x)$ que se obtiene en el algoritmo no es necesariamente mónico, de manera que tomaremos el correspondiente polinomio monico $d(x) = ur_n(x)$, donde u es el inverso del coeficiente de grado más grande de $r(x)$, como máximo común divisor. Con estas observaciones tenemos:

Teorema 11.36 (Identidad de Bézout). Dados $a(x),b(x) \in K[x]^*$ con $\mathrm{mcd}(a(x),b(x)) = d(x)$, existen dos polinomios $s(x),t(x) \in K[x]$ tales que

$$a(x)s(x)+b(x)t(x) = d(x)$$

Si tomamos como ejemplo los polinomios anteriores en $\mathbb{Z}_5[x]$, $a(x) = x^4 + 3x^2 - 4x + 1$ y $b(x) = x^2 - 3x$, obtenemos la identidad de Bézout a partir del máximo común divisor, que hemos calculado previamente a partir del algoritmo de Euclides, de la forma siguiente:

$$\begin{aligned} 3 &= (x^2+2x) - (2x+1)(3x+2) \\ &= (x^2+2x) - [(x^4+3x^2+x+1) - (x^2+2x)(x^2+3x+2)](3x+2) \\ &= (x^2+2x)[1+(x^2+3x+2)(3x+2)] - (x^4+3x^2+x+1)(3x+2) \\ &= (x^2+2x)(3x^3+x^2+2x) - (x^4+3x^2+x+1)(3x+2) \end{aligned}$$

Por tanto, multiplicando por 2 los dos lados de la igualdad, obtenemos:

$$1 = (x^2+2x)\underbrace{(x^3+2x^2+4x)}_{t(x)} + (x^4+3x^2+x+1)\underbrace{(4x+1)}_{s(x)}$$

Ejercicio 11.37. Usar el algoritmo de Euclides para encontrar el máximo común divisor $d(x)$ de los polinomios

$$\begin{aligned} a(x) &= 1+2x^2 \\ b(x) &= 1+2x+x^2 \end{aligned}$$

en $\mathbb{Z}_3[x]$. Obtener después los polinomios $s(x)$ y $t(x)$ que permiten escribir la identidad de Bézout $a(x)s(x) + b(x)t(x) = d(x)$.

Los ideales de K[x]

La noción de divisibilidad introducida aquí para los polinomios es válida también para otros anillos unitarios abelianos, en particular para $\mathbb{Z}$. La simplicidad de toda la teoría de la divisibilidad en $\mathbb{Z}$ proviene del hecho que $\mathbb{Z}$ es un anillo principal. Veremos en esta sección que $K[x]$ es también un anillo principal, siendo K un cuerpo.

En primer lugar es preciso encontrar una traducción del concepto de divisor en términos de ideales.

Proposición 11.38. En un anillo euclídeo,

$$b|a \Longleftrightarrow (a) \subset (b)$$

donde (a) y (b) son los ideales generados por a y b respectivamente.

Ejercicio 11.39. Demostrar la proposición anterior.

Es preciso observar que hemos traducido las nociones clásicas de divisibilidad en términos de inclusión de subconjuntos. Esta nueva manera de interpretar la divisibilidad tiene ventajas conceptuales que aprovecharemos en la sección siguiente.

La medida de proximidad, en cuanto a la divisibilidad, entre dos elementos de un anillo íntegro unitario A la da su máximo común divisor. En términos de ideales, la expresión de este máximo común divisor es muy simple.

Proposición 11.40. En un anillo euclídeo,

$$d = \mathrm{mcd}(a,b) \Longleftrightarrow (d) = (a,b)$$

Demostración. Observemos que el ideal generado por a y b está formado por los elementos de la forma $ax+by$, $x,y \in A$, y, por tanto, $(d) = (a,b) = (a) + (b)$. □

Observemos que de la demostración anterior se deduce de forma inmediata la identidad de Bézout.

Ejercicio 11.41. Sean a y b dos elementos primos entre sí en un anillo íntegro unitario. Encontrar la expresión correspondiente en términos de ideales.

La proposición anterior permite interpretar el anillo de polinomios sobre un cuerpo como un anillo principal.

Proposición 11.42. Si K es un cuerpo, $K[x]$ es un anillo principal.

Demostración. Tenemos que demostrar que los ideales de $K[x]$ son principales. Sea $I \neq \{0\}$ un ideal de $K[x]$ y $b(x)$ un polinomio de grado mínimo entre los polinomios de $I \setminus \{0\}$. Para cualquier $a(x) \in I$, la división euclídea permite escribir $a(x) = b(x)q(x) + r(x)$, para ciertos polinomios $b(x), r(x)$ y $gr(r(x)) < gr(b(x))$. Pero $r(x) = a(x) - b(x)q(x) \in I$, de manera que, siendo $b(x)$ el polinomio de grado mínimo de $I \setminus \{0\}$ tiene que ser $r(x) = 0$ y $a(x) \in (b(x))$. □

Podemos afirmar por tanto que los ideales I de $K[x]$ son de la forma $I = (a(x))$ con $a(x) \in K[x]$ mónico.

Para la construcción de los cuerpos finitos en la última sección, se usará el resultado siguiente, que ya hemos demostrado en general para cualquier anillo unitario abeliano.

Teorema 11.43. Si K es un cuerpo y $M(x)$ es un ideal maximal de $K[x]$, entonces $K[x]/M(x)$ es un cuerpo.

Es importante, por tanto, conocer cuáles son los ideales maximales de $K[x]$. Recordemos que, siguiendo también en este aspecto la similitud con el anillo de los enteros, los ideales maximales de $K[x]$ son justamente los generados por sus polinomios primos o irreductibles.

Proposición 11.44. Un ideal $(a(x))$ es maximal de $K(x)$ si y sólo si $a(x)$ es un polinomio primo.

Demostración. Ya hemos visto que todos los ideales de $K(x)$ son principales. Si $a(x)$ no es primo, se puede poner $a(x) = p(x)q(x)$, donde $p(x)$ y $q(x)$ son polinomios de grado mayor o igual que 1. Entonces, el ideal $(a(x))$ está estrictamente incluido en $(p(x))$, y no es maximal. Recíprocamente, si $a(x)$ no es maximal, $((a(x))$ está estrictamente incluido en $(p(x))$ para algún $p(x)$, de manera que $a(x) = p(x)q(x)$ y $a(x)$ no es primo. □

Raíces de un polinomio

Es frecuente pensar en funciones cuando se habla de polinomios. Esta es la razón que lleva a la notación clásica de los polinomios, a la que aludíamos al comienzo de esta sección. Pero, de hecho, no toda función polinómica está representada por un único polinomio, aunque sí sea cierto en sentido contrario. A continuación definiremos las funciones polinómicas y estudiaremos la relación entre sus ceros y la factorización de los polinomios correspondientes.

Sea K un cuerpo y $a(x) \in K[x]$, $a(x) = \sum_{i=0}^{n} a_i x^i$. Se define la *función polinómica* asociada al polinomio $a(x)$ como la aplicación

$$\begin{array}{rcl} \overline{a} : K & \longrightarrow & K \\ k & \longrightarrow & \overline{a}(k) = \displaystyle\sum_{i=0}^{n} a_i k^i \end{array}$$

Tomamos como ejemplo los polinomios $a(x) = x - 2$ y $b(x) = x^3 - 2$ en $\mathbb{Z}_3$. Las funciones polinómicas asociadas a estos dos polinomios son la misma, como se puede comprobar en la tabla siguiente:

$$\begin{array}{l} \overline{a}(0) = \overline{b}(0) = 1 \\ \overline{a}(1) = \overline{b}(1) = 2 \\ \overline{a}(2) = \overline{b}(2) = 0 \end{array}$$

En cambio, los dos polinomios no tienen los mismos coeficientes (no son el mismo). Comportamientos como el de este ejemplo son frecuentes en funciones polinómicas definidas sobre $\mathbb{Z}_p$, donde p es un número primo. De hecho, serán éstos los polinomios con los cuales trabajaremos. Se debe decir, sin embargo, que si el cuerpo es de característica cero, como por ejemplo $\mathbb{Q}$, $\mathbb{R}$ o $\mathbb{C}$, cada función polinómica tiene asociado un único polinomio y por tanto en estos casos no hay inconveniente en identificar los dos conceptos.

Se dice que $\alpha \in K$ es una *raíz* (o *cero*) del polinomio $a(x) \in K[x]$ si $\overline{a}(\alpha) = 0$, es decir, si es un cero de la función polinómica correspondiente.

El resultado siguiente da, mediante una condición muy sencilla, la herramienta clave que permite relacionar los ceros de una función polinómica con los factores de un polinomio.

Teorema 11.45. $\alpha \in K$ es una raíz de $a(x) \in K[x]$ si y sólo si $(x - \alpha) | a(x)$.

Demostración. Observemos en primer lugar que $gr(a) \geqslant 1$, ya que si $gr(a) = 0$ querría decir que $a(x)$ es un polinomio constante y por tanto no tiene raíces. Al hacer la división euclídea de $a(x)$ por $x-\alpha$ tenemos $a(x) = (x-\alpha)q(x)+r(x)$ con $gr(r) < 1$; como α es raíz de $a(x)$, $r(x) = 0$ y de aquí $(x-\alpha)|a(x)$. En sentido contrario sólo es preciso observar que si $(x-\alpha)|a(x)$, entonces existe un $q(x) \in K[x]$ tal que $a(x) = (x-\alpha)q(x)$ y, por tanto, α es raíz de $a(x)$. □

Se dice que $\alpha \in K$ es una raíz de *multiplicidad* $m \geqslant 1$ del polinomio $a(x) \in K[x]$ si y sólo si $(x-\alpha)^m$ divide a $a(x)$, pero no lo hace $(x-\alpha)^{m+1}$.

El teorema siguiente establece la relación entre el número de raíces de un polinomio y su grado. El resultado aparentemente inocente es esencial para tratar el problema central que nos ocupa, la factorización polinómica.

Teorema 11.46. Si el grado de $a(x) \in K[x]$ es n, entonces la suma de las multiplicidades de las raíces es como máximo n.

Demostración. Si $a(x)$ tiene raíces $\alpha_1, \ldots, \alpha_k$ con multiplicidades $m_1, \ldots, m_k$, entonces el polinomio $b(x) = (x-\alpha_1)^{m_1} \cdots (x-\alpha_k)^{m_k}$ divide a $a(x)$ y, por tanto, $gr(b(x)) = m_1 + \cdots + m_k \leqslant gr(a(x))$. □

A veces se utiliza una versión diferente de este teorema que dice que el número de raíces diferentes de un polinomio es como máximo igual a su grado.

11.3 Cuerpos finitos

Sabemos que $(\mathbb{Z}_p, +, \cdot)$ es un cuerpo de p elementos si p es un número primo. ¿Existen cuerpos finitos de cualquier orden? Responder a esta cuestión es el primer objetivo que nos proponemos en esta sección. Estudiaremos, para comenzar, la estructura interna que deben tener estos cuerpos mediante el estudio de su grupo aditivo y multiplicativo. Describiremos de manera breve las razones de su existencia, que dependerá, como veremos, de la existencia de ciertos polinomios. Finalmente describiremos la manera general de obtenerlos mediante ejemplos ilustrativos y comentaremos, para acabar, también brevemente, que los cuerpos que hemos aprendido a construir son de hecho los únicos posibles.

Para comenzar recordemos que, si K es un cuerpo finito, su característica tiene que ser un número primo.

Teorema 11.47. Si K es un cuerpo de característica p, $|K| = p^n$, $n \in \mathbb{N}$.

Demostración. Demostraremos que el grupo aditivo de K es isomorfo al producto directo de n grupos cíclicos de orden p.

En primer lugar, observemos que el subgrupo cíclico de $(K,+)$ generado por un elemento $k \in K$ tiene orden p, como consecuencia directa de su característica. Es decir,

$$\langle k \rangle = \{k, k+k, \ldots, \underbrace{k+\cdots+k}_{p}\} = \{mk, m \in \mathbb{Z}_p\} \simeq \mathbb{Z}_p$$

Supongamos ahora que $\{g_1, g_2, \ldots, g_n\}$ sea un conjunto de generadores mínimo de K, es decir, que ningún subconjunto de éste genere todo K. Esto quiere decir que para cualquier elemento $k \in K$ existen n enteros $\{m_1, m_2, \ldots, m_n\}$, tales que

$$k = \sum_{i=1}^{n} m_i g_i$$

Demostraremos que las p^n posibles combinaciones de expresiones de esta forma son diferentes y dan lugar por tanto a p^n elementos diferentes de K. Supongamos que dos sumatorios diferentes diesen lugar al mismo elemento de K, es decir,

$$k = \sum_{i=1}^{n} m_i g_i = \sum_{i=1}^{n} m'_i g_i$$

Si j fuese la primera posición del sumatorio tal que $m_j \neq m'_j$, tendríamos que

$$(m_j - m'_j) g_j = \sum_{i=j+1}^{n} (m_i - m'_i) g_i$$

Ahora bien, como $(m_j - m'_j) = (m_j - m'_j) \cdot 1 \neq 0$ tiene inverso en K, obtendríamos

$$g_j = (m_j - m'_j)^{-1} \sum_{i=j+1}^{n} (m_i - m'_i) g_i$$

Esto contradice la hipótesis de que $\{g_1, g_2, \ldots, g_n\}$ es un conjunto de generadores mínimo.

Podemos asociar por tanto, de manera única, cada elemento $k = \sum_{i=1}^{n} m_i g_i$ de K con la n-tupla $\{m_1, m_2, \ldots, m_n\}$ de elementos de $\mathbb{Z}_p$. Esta asociación es por tanto una biyección de K en $\underbrace{\mathbb{Z}_p \times \mathbb{Z}_p \times \cdots \times \mathbb{Z}_p}_{n} = (\mathbb{Z}_p)^n$, que es de hecho un isomorfismo entre $(K,+)$ y $((\mathbb{Z}_p)^n,+)$. Por tanto,

$$(K,+) \simeq ((\mathbb{Z}_p)^n,+)$$

□

Acabamos de demostrar que el orden de cualquier cuerpo finito tiene que ser una potencia de un número primo, p^n. Esto lo hemos hecho estudiando la estructura de su grupo aditivo y

hemos visto que éste debe ser isomorfo a $((\mathbb{Z}_p)^n, +)$. Es, por tanto, razonable interpretar los elementos del cuerpo K como polinomios con coeficientes en $\mathbb{Z}_p$ de grado inferior a n. Sin embargo, si bien está claro que este conjunto es adecuado como modelo para el grupo aditivo del cuerpo, está claro también que no lo es para su grupo multiplicativo. Sólo es preciso observar que el producto no es cerrado en este conjunto, ya que el producto de dos polinomios sobre un cuerpo íntegro tiene por grado la suma de los grados de los polinomios correspondientes. Convendrá rectificar entonces el modelo, de manera que sea también compatible con el grupo multiplicativo. Antes de presentar un modelo apropiado con la estructura de cuerpo, estudiemos cómo tendría que ser su grupo multiplicativo.

Hemos visto que dos cuerpos finitos del mismo orden tienen sus grupos aditivos isomorfos. Veremos ahora que también son isomorfos sus grupos multiplicativos.

Teorema 11.48. El grupo multiplicativo de un cuerpo finito es cíclico.

Demostración. Supongamos que K es un cuerpo de orden p^n, con p primo y n un número natural. Como el orden de cualquier elemento diferente de cero de un grupo finito multiplicativo divide al orden del grupo, tenemos que, si $k \in K^*$, entonces $k^{p^n-1} = 1$. Esto es equivalente a decir que la ecuación

$$x^{p^n-1} - 1 = 0$$

tiene $p^n - 1$ ceros en K.

Por otra parte, está claro que el grupo multiplicativo de un cuerpo es abeliano y por tanto podemos utilizar el resultado siguiente, que es consecuencia (no directa) del teorema de Lagrange para grupos abelianos: "el exponente de un grupo es múltiplo de todos los órdenes de los elementos del grupo". De aquí que, si m es el exponente de $(K^*, \cdot)$, cada elemento de K^* satisface la ecuación $x^m - 1 = 0$, y por tanto, existen $p^n - 1$ raíces diferentes en esta ecuación. Pero, como cada polinomio de grado m tiene como máximo m raíces, deducimos que $m = p^n - 1$. Por tanto, el elemento que tiene orden m genera todo el grupo multiplicativo, es decir, el grupo $(K^*, \cdot)$ es cíclico. □

Hasta ahora hemos demostrado que, si existe un cuerpo finito K, debe cumplir las condiciones siguientes:

1. $|K| = p^n$;
2. $(K, +) \simeq ((\mathbb{Z}_p)^n, +)$;
3. $(K^*, \cdot)$ es cíclico.

La pregunta que nos formulamos a continuación tiene una apariencia sencilla, teniendo en cuenta la información de que disponemos, pero realmente no es así. Dado cualquier número primo p y cualquier número natural n, ¿existe un cuerpo de orden p^n?

Sabemos ya que, si en el anillo de los números enteros $(\mathbb{Z}, +, \cdot)$ definimos la relación de equivalencia módulo un número primo p, obtenemos el cuerpo $(\mathbb{Z}_p, +, \cdot)$ de orden p. De forma similar, si en el anillo de los polinomios $(\mathbb{Z}_p[x], +, \cdot)$ definimos la relación de equivalencia módulo un polinomio primo de grado n, obtendremos un cuerpo de orden p^n, llamado *cuerpo de Galois* de orden p^n en honor de Evariste Galois (1811–1832) y denotado por $GF(p^n)$ o simplemente $\mathbf{F}_q$, con $q = p^n$. Éste es entonces el modelo que aventurábamos anteriormente.

Teorema 11.49. Si $p(x) \in \mathbb{Z}_p[x]$ es un polinomio primo de grado n, entonces $\mathbb{Z}_p[x]/p(x)$ es un cuerpo de orden p^n, llamado cuerpo de Galois y denotado por $\mathbf{F}_{p^n}$.

Demostración. Por el hecho de ser $\mathbb{Z}_p[x]$ un anillo principal, sus ideales maximales están justamente generados por polinomios primos y por tanto $\mathbb{Z}_p[x]/p(x)$ es un cuerpo. □

El problema, entonces, consiste en asegurar la existencia de algún polinomio primo con coeficientes sobre cualquier cuerpo $\mathbb{Z}_p$ y de cualquier grado n. Demostrar la existencia de estos polinomios es laborioso y requiere la introducción de conceptos algebraicos que van más allá de los propósitos de este libro, pero alentamos al lector interesado y lo dirigimos a [3] [5].

Teorema 11.50. Para cualquier número natural n y cualquier número primo p, existe un polinomio primo $p(x) \in \mathbb{Z}_p[x]$ de grado n.

Muy esquemáticamente, la demostración consiste en caracterizar los elementos del cuerpo a partir de las raíces del polinomio $(x^{p^n-1} - 1)x = x^{p^n} - x$. Trivialmente, localizamos los elementos neutros del cuerpo como raíces de este polinomio. El resto de los elementos serán también raíces del polinomio y, por tanto, de los factores de su descomposición. Más concretamente: se caracterizan los elementos del cuerpo a partir de las raíces de los polinomios irreductibles de grado d, d divisor de n. Se demuestra que las raíces de un polinomio irreductible de grado d sobre $\mathbf{F}_p$ son las p-ésimas potencias de una de sus raíces α: $\alpha, \alpha^p, \alpha^{p^2}, \ldots, \alpha^{p^{d-1}}$, para algún elemento α del cuerpo. A partir de este resultado se deduce que los factores de $x^{p^n} - x$ son todos los polinomios irreductibles sobre $\mathbf{F}_p$ de grado d. Finalmente se demuestra que, para $d = n$, el número de estos polinomios es como mínimo 1.

Sabiendo de su existencia, la obtención práctica de estos polinomios es también en general muy laboriosa. Por este motivo se han construido unas tablas con todos los polinomios irreductibles de grado n sobre $\mathbf{F}_p$ para valores razonablemente moderados de p y de n. En la tabla 11.1 presentamos una muestra que incluye los polinomios irreductibles sobre $\mathbf{F}_2$ y sobre $\mathbf{F}_3$ de grados 1, 2 y 3.

Tabla 11.1: Polinomios irreductibles de grado 1, 2 y 3 sobre $\mathbf{F}_p$

grado	$\mathbf{F}_2$	$\mathbf{F}_3$
1	x	x
	$x+1$	$x+1$
		$x+2$
2	x^2+x+1	x^2+1
		x^2+2x+2
		x^2+x+2
3	x^3+x+1	x^3+2x+1
	x^3+x^2+1	x^3+2x^2+1
		x^3+x^2+2
		x^3+2x^2+2
		x^3+x^2+x+2
		x^3+x^2+2x+1
		x^3+2x^2+x+1
		x^3+2x^2+2x+2

Observemos en primer lugar que se consideran sólo polinomios mónicos, ya que sabemos que un polinomio $p(x) \in \mathbf{F}_p[x]$ es irreductible si y sólo si lo es el polinomio $kp(x)$, para todo $k \in \mathbf{F}_p^*$.

Para los valores de p y n que aparecen en la tabla 11.1 es razonablemente sencillo deducir los posibles polinomios primos. Para ello, procedemos de forma similar a como lo haríamos para determinar si un número es primo. Claramente, los únicos polinomios irreductibles de grado 1 son los que figuran en la tabla. Los polinomios reducibles de grado 2 serán producto de dos polinomios irreductibles de grado 1; así tenemos que $x \cdot x = x^2$, $x(x+1) = x^2+x$ y $(x+1)(x+1) = x^2+1$ son los únicos polinomios reducibles de grado 2 sobre $\mathbb{Z}_2$ y por tanto x^2+x+1 representa el único polinomio irreductible de grado 2 sobre $\mathbb{Z}_2$. De hecho, hay cuatro polinomios de grado 3 que contienen únicamente factores de grado 1 y dos que contienen factores de grado 1 y de grado 2. Por tanto, sólo dos de los ocho posibles polinomios de grado 3 sobre $\mathbb{Z}_2$ son irreductibles. En general, es preciso observar que hay p^n polinomios mónicos de grado $n \geqslant 1$. Algunos de estos polinomios se obtienen como producto de factores de grado más pequeño, pero, de hecho, no existen p^n maneras diferentes de combinar polinomios irreductibles de grado menor que n para obtener uno de grado n.

Conociendo ya algunos polinomios irreductibles, pasamos a construir los cuerpos correspondientes.

Comenzamos con $\mathbf{F}_4$. Para ello, consideramos en $\mathbb{Z}_2[x]$ la relación de equivalencia módulo el único polinomio irreductible de grado dos que tenemos, $p(x) = x^2 + x + 1$. Las clases que se obtienen quedan representadas por los polinomios $\{0, 1, x, x+1\}$. Así,

$$\mathbb{Z}_2[x]/(x^2 + x + 1) = \{[0], [1], [x], [x+1]\}$$

Recordemos que en general la suma y el producto de clases están definidos a partir de sus representantes. Por tanto,

$$\{[0], [1], [x], [x+1]\} = \{0, 1, [x], [x]+1\}$$

Si representamos la clase $[x] = \{x + (x^2 + x + 1)c(x), c(x) \in \mathbb{Z}_2[x]\}$ por α, las tablas 11.2 y 11.3 muestran el comportamiento de los elementos del cuerpo $\mathbb{Z}_2[x]/(x^2 + x + 1)$ respecto de la suma y el producto.

Tabla 11.2: Tabla de $(\mathbf{F}_4, +)$

$+$	0	1	α	$\alpha+1$
0	0	1	α	$\alpha+1$
1	1	0	$\alpha+1$	α
α	α	$\alpha+1$	0	1
$\alpha+1$	$\alpha+1$	α	1	0

Tabla 11.3: Tabla de $(\mathbf{F}_4(\alpha), \cdot)$

$\cdot$	1	α	$\alpha+1$
1	1	α	$\alpha+1$
α	α	$\alpha+1$	1
$\alpha+1$	$\alpha+1$	1	α

Es preciso observar que mediante la relación de equivalencia definida se ha conseguido tener el producto cerrado en el conjunto de polinomios sobre $\mathbb{Z}_2$ de grado inferior a 2.

Es interesante observar también que, en este caso, α es un generador del grupo multiplicativo $(\mathbb{Z}_2^\star[x]/(x^2+x+1),\cdot) = (\mathbf{F}_4(\alpha),\cdot)$, es decir, podemos obtener todos los elementos del cuerpo salvo el cero, como las 2^2-1 potencias sucesivas de α:

$$\{\mathbb{Z}_2^\star[x]/(x^2+x+1)\} = \mathbf{F}_4(\alpha) = \{1,\alpha,\alpha+1\} = \{\alpha^3,\alpha,\alpha^2\}$$

Es interesante, para facilitar los cálculos en las tablas de multiplicar, obtener un generador simple. En general se dice que un generador del grupo multiplicativo del cuerpo es un *elemento primitivo*. Cabe observar que este elemento siempre existe (el grupo multiplicativo del cuerpo es siempre cíclico).

Para construir $\mathbf{F}_8$ tenemos dos posibles elecciones para el polinomio irreductible, x^3+x+1 o x^3+x^2+1.

En primer lugar, calculamos los 8 polinomios de grado inferior a 3 con coeficientes en $\mathbb{Z}_2$:

$$\{0,1,x,x+1,x^2,x^2+1,x^2+x,x^2+x+1\}$$

Observemos que la tabla de sumar se obtiene directamente. Nos centraremos por tanto en la tabla de multiplicar.

Está claro que, multiplicando los representantes de cada clase y calculando los restos módulo x^3+x+1 o x^3+x^2+1, obtendremos unos representantes mónicos de los elementos de $\mathbb{Z}_2[x]/(x^3+x+1)$ y $\mathbb{Z}_2[x]/(x^3+x^2+1)$ respectivamente. Pero estos cálculos en general son tediosos y es por ello interesante encontrar un elemento primitivo del cuerpo.

Si consideramos $\mathbb{Z}_2[x]/(x^3+x+1)$, por ejemplo, podemos comprobar que las 2^3-1 potencias sucesivas de x dan lugar a toda una familia de representantes del grupo multiplicativo del cuerpo y, por tanto, del propio cuerpo.

$$\{x,x^2,x^3,x^4,x^5,x^6,x^7\} = \{x,x^2,x+1,x^2+x,x^2+x+1,x^2+1,1\}$$

Consecuentemente, podemos representar los elementos del cuerpo por las sucesivas potencias de $[x]$:

$$\{[x],[x]^2,[x]^3=[x]+1,[x]^4=[x]^2+[x],[x]^5=[x]^2+[x]+1,[x]^6=[x]^2+1,[x]^7=1\}$$

Igual que en el ejemplo anterior, la clase de $[x]$, que denotaremos también por α, es un elemento primitivo del cuerpo que ahora denotamos por $\mathbf{F}_8(\alpha)$. Así, para construir la tabla 11.4 usaremos las potencias de α. Para encontrar los polinomios correspondientes a las potencias de α sólo es preciso deshacer los cambios que figuran a la derecha en la tabla 11.4.

Si utilizamos x^3+x^2+1 para construir $\mathbf{F}_8$, podemos comprobar, también en este caso, que a partir de las 2^3-1 potencias sucesivas de x obtenemos toda una familia de representantes de los elementos del cuerpo diferentes de cero. Pero en este caso las relaciones son las siguientes:

$$\{[x],[x^2],[x^3]=[x^2+1],[x^4]=[x^2+x+1],[x^5]=[x+1],[x^6]=[x^2+x],[x^7]=1\}$$

Tabla 11.4: Tabla de $(\mathbf{F}_8(\alpha),\cdot)$

$\cdot$	α	α^2	α^3	α^4	α^5	α^6	1
α	α^2	α^3	α^4	α^5	α^6	1	α
α^2	α^3	α^4	α^5	α^6	1	α	α^2
α^3	α^4	α^5	α^6	1	α	α^2	α^3
α^4	α^5	α^6	1	α	α^2	α^3	α^4
α^5	α^6	1	α	α^2	α^3	α^4	α^5
α^6	1	α	α^2	α^3	α^4	α^5	α^6
1	α	α^2	α^3	α^4	α^5	α^6	1

$\alpha^1 = [x]$
$\alpha^2 = [x]^2$
$\alpha^3 = [x]+1$
$\alpha^4 = [x]^2+[x]$
$\alpha^5 = [x]^2+[x]+1$
$\alpha^6 = [x]^2+1$
$\alpha^7 = 1$

Igual que en el ejemplo anterior, la clase de $[x]$, que denotaremos ahora por β, es un elemento primitivo del cuerpo, denotado por $\mathbf{F}_8(\beta)$. Y, también igual que antes, construimos la tabla del producto a partir de las potencias de β (Tabla 11.5).

Tabla 11.5: Tabla de $(\mathbf{F}_8(\beta),\cdot)$

$\cdot$	β	β^2	β^3	β^4	β^5	β^6	1
β	β^2	β^3	β^4	β^5	β^6	1	β
β^2	β^3	β^4	β^5	β^6	1	β	β^2
β^3	β^4	β^5	β^6	1	β	β^2	β^3
β^4	β^5	β^6	1	β	β^2	β^3	β^4
β^5	β^6	1	β	β^2	β^3	β^4	β^5
β^6	1	β	β^2	β^3	β^4	β^5	β^6
1	β	β^2	β^3	β^4	β^5	β^6	1

$\beta^1 = [x]$
$\beta^2 = [x]^2$
$\beta^3 = [x]^2+1$
$\beta^4 = [x]^2+[x]+1$
$\beta^5 = [x]+1$
$\beta^6 = [x]^2+[x]$
$\beta^7 = 1$

Si comparamos las tablas 11.4 y 11.5, vemos que son evidentemente idénticas. Sin embargo, después de substituir en estas tablas las sucesivas potencias por los polinomios correspondientes, las tablas no coinciden. Es fácil comprobar que $\beta+1$ es, de hecho, otro elemento primitivo de x^3+x+1. Esto quiere decir que existe un isomorfismo ϕ de $(\mathbb{Z}_2[x]/(x^3+x+1)$ en $(\mathbb{Z}_2[x]/(x^3+x^2+1)$

$$\phi : \mathbf{F}_8(\alpha) \longrightarrow \mathbf{F}_8(\beta)$$

tal que $\phi(\alpha) = \beta+1$.

Este hecho no es casual. En general, se demuestra que todos los posibles cuerpos de un mismo orden son isomorfos. De hecho, hemos visto ya que todos los cuerpos de orden p^n tienen sus grupos aditivos isomorfos a $((\mathbb{Z}_p)^n, +)$ y sus grupos multiplicativos son cíclicos de orden $p^n - 1$. Queda, por tanto, por ver que la estructura del cuerpo no depende del generador que escogemos ni en particular, entonces, del polinomio irreductible escogido. Por razones similares a las aludidas cuando planteábamos la existencia de un polinomio irreductible de cualquier grado sobre cualquier $\mathbb{Z}_p$ con p primo, prescindiremos de presentar aquí la demostración formal sobre la existencia en general de un isomorfismo entre dos cuerpos cualesquiera del mismo orden y recomendamos, al lector interesado en esta cuestión, la misma bibliografía.

El teorema siguiente recoge de manera concisa el resultado que acabamos de mencionar.

Teorema 11.51. Para cada número primo p y para cada número natural n, hay un único cuerpo, salvo isomorfismos, de orden p^n llamado cuerpo de Galois, $\mathbf{GF}(p^n)$.

Notas bibliográficas

Aunque la parte introductoria de este capítulo, es decir, la que se refiere a las definiciones y primeras propiedades de los anillos, se puede encontrar en cualquier libro introductorio de álgebra, nosotros preferimos recomendar, incluso para esta parte, libros de cariz aplicado, ya que éste será nuestro último interés y es conveniente no dispersar los objetivos en cuestiones extremadamente teóricas. En este sentido el libro de Birkhoff y Bartee [2] puede ser de gran ayuda. Es interesante tener más de una referencia, y tanto el libro del Stone [5] como el de Childs [4] pueden también ser útiles en esta primera parte. Para la última parte, la que corresponde a los cuerpos finitos, recomendamos el libro de Lindl [3], teniendo en cuenta que, si bien su calidad es indudable, su nivel es superior al de este libro. Para compensar este desnivel, el libro de Biggs [1] es ideal. También se encuentra explicado este tema a un nivel intermedio en los otros libros recomendados para la primera parte.

Bibliografía

[1] N. L. Biggs. *Matemática Discreta*, Vicens Vives, 1993.

[2] G. Birkhoff, T. C. Bartee. *Modern Applied Algebra*, McGraw-Hill, 1970.

[3] R. Lidl, G. Pilz. *Applied Abstract Algebra*, Springer-Verlag, 1984.

[4] L. Childs. *A Concrete Introduction to Higher Algebra*, Springer-Verlag, 1988.

[5] H. S. Stone. *Discrete Mathematical Structures and their Applications*, Science Research Asociates, SRA, 1973.

Problemas

1. Demostrar que el producto en un anillo A es conmutativo si y sólo si para todo $a,b \in A$,
$$(a+b)^2 = a^2 + 2ab + b^2$$

2. Comprobar que el conjunto de aplicaciones de un anillo A sobre él mismo, $F(A)$, tiene estructura de anillo con la suma y el producto de aplicaciones, es decir, que para toda $f,g \in F(A)$ y para todo $x \in A$ se define:
$$\begin{aligned}(f+g)(x) &= f(x)+g(x)\\ (fg)(x) &= f(x)g(x)\end{aligned}$$

3. Demostrar, de forma general, que el conjunto de aplicaciones de un conjunto cualquiera X sobre un anillo A, $F(X,A)$, tiene con las mismas operaciones del ejercicio anterior, estructura de anillo.

4. Sea $(G,+)$ un grupo abeliano. Demostrar que el conjunto de aplicaciones de G en G con la suma y la composición de aplicaciones, $(End(G),+,\circ)$, es un anillo unitario. ¿Tiene divisores de cero? Estudiar el caso de $G = \mathbb{Z}_2 \times \mathbb{Z}_2$.

5. Demostrar que si un anillo $(A \neq \{0\},+,\cdot)$ es unitario, entonces los elementos neutros de la suma y el producto son diferentes.

6. Un anillo A en el cual todo elemento $a \in A$ es independiente de la segunda operación, es decir, $a^2 = a$, se llama *anillo de Boole*. Demostrar que
   (a) A es de característica 2 y es abeliano;
   (b) para todo $a,b \in A$ se cumple que $ab(a+b) = 0$;
   (c) si A es íntegro, entonces $A = \{0\}$ o bien $A \simeq \mathbb{Z}_2$;
   (d) sólo hay un anillo de Boole de cuatro elementos.

7. Si X es un conjunto cualquiera, demostrar que $(P(X),\triangle,\cap)$ es un anillo de Boole, con la llamada *diferencia simétrica*, $\triangle$, como primera operación. Es decir, para todo $C,D \subseteq X$
$$C \triangle D = (C \cup D) - (C \cap D)$$

8. Consideremos en el producto cartesiano de dos anillos A_1 y A_2 las operaciones suma y producto siguientes:
$$\begin{aligned}(a_1,a_2)+(b_1,b_2) &= (a_1+b_1,a_2+b_2)\\ (a_1,a_2)(b_1,b_2) &= (a_1b_1,a_2b_2)\end{aligned}$$

(a) Demostrar que $(A_1 \times A_2, +, \cdot)$ es un anillo, llamado *producto cartesiano*.

(b) Estudiar cómo se trasladan la conmutatividad, la integridad y la existencia de unidad de A_1 y A_2 a $A_1 \times A_2$.

9. El conjunto de elementos de un anillo A que conmutan con todos los elementos de A

$$Z(A) = \{x \in A \mid xa = ax, \forall a \in A\}$$

se llama *centro* de A. Demostrar que es un subanillo de A.

10. Dado un anillo unitario A y un elemento $a \in A$, demostrar que la aplicación $\phi : A \longrightarrow A$ definida por $\phi(x) = axa^{-1}$ es un automorfismo de A.

11. Demostrar que el conjunto de elementos invertibles de un anillo unitario A

$$U(A) = \{u \in A \mid \exists a \in A, au = ua = 1\}$$

es un subgrupo del grupo multiplicativo A^*. Calcular $U(\mathbb{Z})$, $U(\mathbb{Z}_3)$, $U(\mathbb{Z}_6)$ y en general $U(\mathbb{Z}_n)$.

12. Demostrar que la imagen homomórfica de un cuerpo es un cuerpo.

13. Dado un cuerpo K, demostrar que la aplicación $f : K \longrightarrow K$ tal que $f(k) = k^p$ es un automorfismo si K tiene característica p.

14. Dado un subconjunto propio Y de un anillo A, consideremos el llamado *anulador por la izquierda* de Y. Éste es el conjunto X de elementos de A tal que

$$X(Y) = \{x \in A \mid xy = 0, \forall y \in Y\}$$

Demostrar que es un ideal por la izquierda de A.

15. Si A es un anillo unitario, consideremos el producto cartesiano $\mathbb{Z} \times A$ con las operaciones siguientes:

$$\begin{aligned} (n,a)+(m,b) &= (n+m, a+b) \\ (n,a)(m,b) &= (nm, nb+ma+ab) \end{aligned}$$

(a) Demostrar que $\mathbb{Z} \times A$ es un anillo unitario.

(b) Demostrar que $\{0\} \times A$ es un ideal bilateral de $\mathbb{Z} \times A$.

16. Dar ejemplos de polinomios $a(x)$ y $b(x)$ de $\mathbb{Z}_6[x]$ tales que $gr(ab) < gr(a) + gr(b)$. ¿Hay polinomios con estas condiciones en $\mathbb{Z}_5[x]$? ¿Por qué?

17. Comprobar que en $\mathbb{Z}_{12}[x]$ la igualdad siguiente es cierta:

$$(x+1)(x+11) = (x+7)(x+5)$$

¿Es cierta también en $\mathbb{Z}_{13}[x]$? ¿Por qué?

18. Encontrar un polinomio $a(x) \in \mathbb{Z}_n[x]$ no nulo tal que su función polinómica $\bar{a}(x)$ sea nula.

19. Demostrar que en $\mathbb{Z}_p[x]$ (p es primo) los polinomios

$$\begin{cases} a(x) &= x^p \\ b(x) &= x \end{cases}$$

definen la misma función polinómica.

20. ¿Se puede efectuar la divisió euclídea en $\mathbb{Z}[x]$ de

$$\begin{cases} a(x) &= 5x^3 + 2x - 1 \\ b(x) &= x^2 - 3x + 11 \ ? \end{cases}$$

¿Por qué?

21. Sean $a(x), b(x) \in K[x]$ dos polinomios primos entre sí. Demostrar que

$$a(x)|b(x)c(x) \Rightarrow a(x)|c(x)$$

22. Comprobar que $(x+1)^3 = x^3 + 1$ en $\mathbb{Z}_3[x]$. Encontrar para qué valores de n es cierto en $\mathbb{Z}_n[x]$ que

$$(x+1)^n = x^n + 1$$

23. Sea $a(x) \in \mathbb{Z}_p[x]$. Demostrar que, en general, si p es primo, entonces se cumple

(a) $(a(x))^p = a(x^p)$

(b) $(a(x))^{p^n} = a(x^{p^n})$

24. Demostrar que los polinomios $a(x) = x^2 + 1$ y $b(x) = 2x$ son polinomios primos entre sí en $\mathbb{Z}[x]$. Deducir que $\mathbb{Z}[x]$ no es principal.

25. Demostrar que en $\mathbb{Z}[x]$ el ideal generado por $x^2 + 1$ es primo pero no es maximal. ¿Por qué?

26. Factorizar $3x^2 + 2x - 1$ en $\mathbb{Q}[x]$, $\mathbb{Z}_3[x]$ y $\mathbb{Z}[x]$.

27. Encontrar un polinomio irreductible de grado tres en $\mathbb{Z}_5[x]$.

28. Demostrar que ax^2+bx+c es irreductible en $\mathbb{Z}_p[x]$ (p es primo), si y sólo si b^2-4ac no es un cuadrado en $\mathbb{Z}_p$.

29. Demostrar que el grupo aditivo de $\mathbf{F}_9$ no es cíclico.

30. Construir dos representaciones de $\mathbf{F}_9$ y comprobar que son isomórficas.

Capítulo 12

Estructuras combinatorias

1. Diseños combinatorios
2. Geometrías finitas
3. Cuadrados latinos

Las estructuras combinatorias estudian de manera sistemática la selección de objetos según unas reglas específicas, o bien, de forma equivalente, las relaciones de incidencia entre determinados objetos y ciertos subconjuntos de estos objetos. La construcción de estas estructuras depende en gran medida de las estructuras algebraicas que se han descrito en los capítulos anteriores, aunque están relacionadas también con otras ramas de la matemática, como por ejemplo la teoría de números o las geometrías finitas entre otros.

Este capítulo pretende dar una visión general sobre estas estructuras, y dedica una atención especial a algunas de ellas como ejemplos ilustrativos importantes. La primera sección está dedicada a introducir los diseños combinatorios como modelos generales de estructuras combinatorias. En la segunda sección se introducen las geometrías finitas y se particulariza el estudio en los planos proyectivos y los planos afines como modelos geométricos de ciertos diseños combinatorios. El capítulo finaliza con el estudio de los cuadrados latinos como modelo combinatorio útil para contrastar diferentes aspectos de un mismo fenómeno.

12.1 Diseños combinatorios

Un *diseño combinatorio* es un par $D = (V, B)$ formado por un conjunto de elementos $V = \{x_1, x_2, \ldots, x_v\}$, llamado conjunto de *variedades*, y una familia de subconjuntos de estas variedades $B = \{B_i, B_i \subset V\}$, llamados *bloques* del diseño.

El número de variedades de un diseño D se denota por $|V| = v$ y el número de sus bloques por $b = |B|$.

Un diseño es, entonces, un sistema general de incidencia que nos dice cuándo un elemento (variedad) $x_i \in V$ está en un determinado subconjunto (bloque) $B_j \in B$. Una forma útil y sencilla de representar el sistema es a partir de lo que se llama su *matriz de incidencia*.

$$A = (a_{ij})_{v \times b}, \quad a_{ij} = \begin{cases} 1, & \text{si } x_i \in B_j \\ 0, & \text{si } x_i \notin B_j \end{cases}$$

Por ejemplo,

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

corresponde a la matriz de incidencia de un diseño $D = (V, B)$, con conjunto de variedades $V = \{1, 2, 3, 4, 5\}$ y conjunto de bloques:

$$B = \{\{4\}, \{4\}, \{2, 4\}, \{2, 3, 4\}\}$$

Observar que se admite la posibilidad que B tenga bloques repetidos. Cuando éste no es el caso, se dice que el diseño es *simple*.

Se dice que dos diseños $D = (V, B)$ y $DE = (V', B')$ son *isomorfos* si se puede obtener uno del otro reordenando variedades o bloques. Es decir, sus matrices de incidencia, A y A', se obtienen una de la otra intercambiando filas (variedades) o columnas (bloques). Más concretamente, D es isomorfo a D' si existen matrices de permutaciones P y Q (de dimensiones $v \times v$ y $b \times b$ respectivamente) tales que,

$$A' = PAQ$$

Por ejemplo, el diseño D' que tiene por matriz de incidencia

$$A' = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

es isomorfo al diseño D descrito anteriormente, ya que existen dos matrices de permutaciones P y Q tales que:

$$A' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} = PAQ$$

Esto corresponde a intercambiar las variedades y los bloques según las biyecciones siguientes:

$$\begin{array}{lcl} x_1 & \to & x_1 = x'_1 \\ x_2 & \to & x_4 = x'_2 \\ x_3 & \to & x_3 = x'_3 \\ x_4 & \to & x_5 = x'_4 \\ x_5 & \to & x_2 = x'_5 \end{array} \qquad \begin{array}{lcl} B_1 & \to & B_4 = B'_1 \\ B_2 & \to & B_3 = B'_2 \\ B_3 & \to & B_2 = B'_3 \\ B_4 & \to & B_1 = B'_4 \end{array}$$

Hay diversas maneras de utilizar un diseño $D = (V, B)$ para construir otros diseños relacionados con él. Quizá la más sencilla de todas es la que corresponde al llamado *diseño dual*, en el cual las funciones de las variedades y de los bloques se intercambian. De forma más precisa, si denotamos el diseño dual de $D = (V, B)$ por

$$D^T = (V^T, B^T) = (B, V)$$

entonces cada variedad $x \in V$ corresponde a un bloque de B^T y cada bloque $B \in B$ corresponde a una variedad de V^T. Las incidencias en D^T vienen dadas por la regla siguiente: una variedad $B \in V^T$ está en el bloque $x \in B^T$ si y sólo si $x \in V$ es de $B \in B$.

Por ejemplo, dado el diseño $D = (V, B)$, con $V = \{1, 2, 3, 4, 5, 6\}$ y el conjunto B formado por los bloques:

$$\begin{array}{ll} B_1 = \{1, 2, 3\} & B_5 = \{2, 4, 5\} \\ B_2 = \{1, 4, 5\} & B_6 = \{2, 4, 6\} \\ B_3 = \{1, 2, 6\} & B_7 = \{3, 4, 5\} \\ B_4 = \{1, 3, 6\} & B_8 = \{3, 5, 6\} \end{array}$$

su diseño dual $D^T = (V^T, B^T)$ tiene como conjunto de variedades $V^T = \{1, 2, 3, 4, 5, 6, 7, 8\}$ y como conjunto de bloques B^T el formado por

$$\begin{array}{ll} B_1 = \{1, 2, 3, 4\} & B_4 = \{2, 5, 6, 7\} \\ B_2 = \{1, 3, 5, 6\} & B_5 = \{2, 5, 7, 8\} \\ B_3 = \{1, 4, 7, 8\} & B_6 = \{3, 4, 6, 8\} \end{array}$$

A partir de la definición está claro que el diseño dual del diseño dual es el propio diseño, $(D^T)^T = D$. Se puede demostrar como ejercicio la relación entre las matrices de incidencia de un diseño y su dual:

Proposición 12.1. Si A es la matriz de incidencia de un diseño D, entonces la matriz transpuesta A^T es la matriz de incidencia del diseño dual D^T.

A partir de un diseño $D = (V, B)$ también se puede definir la estructura que se obtiene al reemplazar cada bloque $B_i \in B$ por su complemento $\bar{B}_i = V \setminus B_i$. De esta manera se obtiene otro diseño sobre el mismo conjunto de variedades con el mismo número de bloques, llamado *diseño complementario* de D y que denotaremos por

$$\bar{D} = (\bar{V}, \bar{B}) = (V, B)$$

Por ejemplo, la matriz

$$\bar{A} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

es la matriz de incidencia del diseño complementario del primer diseño considerado. Los respectivos conjuntos de bloques figuran a continuación:

$$\begin{aligned} B &= \{\{4\},\{4\},\{2,4\},\{2,3,4\}\} \\ \bar{B} &= \{\{1,2,3,5\},\{1,2,3,5\},\{1,3,5\},\{1,5\}\} \end{aligned}$$

Diseños regulares

El estudio sistemático de estas estructuras hace necesaria la consideración de ciertas restricciones sobre las relaciones de incidencia. Las más básicas son las que figuran a continuación y caracterizan los diseños que las cumplen.

Se dice que un diseño $D = (V, B)$ es

a) *incompleto* si existe algún bloque $B_i \in B$ tal que, $|B_i| < v$;

b) *uniforme* si cada bloque tiene el mismo número (k) de variedades;

c) *regular* si cada variedad pertenece al mismo número (r) de bloques.

Es fácil comprobar que los subconjuntos de $V = \{1,2,3,4,5,6\}$

$$\begin{array}{ll} B_1 = \{1,2,3\} & B_5 = \{2,4,5\} \\ B_2 = \{1,4,5\} & B_6 = \{2,4,6\} \\ B_3 = \{1,2,6\} & B_7 = \{3,4,5\} \\ B_4 = \{1,3,6\} & B_8 = \{3,5,6\} \end{array}$$

constituyen los bloques de un diseño que cumple estas condiciones.

Estas restricciones en las relaciones de un diseño se traducen en la correspondiente matriz de incidencia en las condiciones siguientes:

a) existe alguna columna con algún 0;

b) cada columna tiene el mismo número (k) de 1;

c) cada fila tiene el mismo número (r) de 1.

La matriz de incidencia del diseño anterior es:

$$A = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$$

Ejercicio 12.2. Demostrar que si $A \in M_{v\times b}(\mathbb{Z}_2)$ es la matriz de incidencia de un diseño uniforme y regular, entonces, si J_n denota la matriz $n \times n$ con todos los términos iguales a 1, se cumple:

b) $J_v A = kJ_{v\times b}$

c) $AJ_b = rJ_{v\times b}$

Éstas fueron las condiciones que propuso R. A. Fisher en 1940 para obtener un buen modelo estadístico para comparar diferentes marcas de un mismo producto. En este caso, un determinado número de personas, b, tiene que probar v marcas de un determinado producto de manera que cada persona tiene que probar el mismo número, k, de marcas y cada marca tiene que ser probada por el mismo número, r, de personas. Está claro que si cada persona prueba todas las marcas, entonces el problema tiene solución, considerando, por ejemplo, la relación de parámetros, $b = v = k = r$. Pero esta solución es demasiado costosa y es conveniente buscar otras con $k < v$.

Esta sección la dedicaremos al estudio de este tipo de estructuras, a las que nos referiremos como diseños regulares, o directamente como diseños, e indicaremos sus parámetros con (v,k,r). El número de bloques de un (v,k,r) diseño no es arbitrario. Es inmediato comprobar que si existe un diseño regular, sus parámetros tienen que cumplir la relación siguiente:

Proposición 12.3. En todo diseño regular de parámetros (v,k,r) se cumple:

$$bk = rv$$

Demostración. Es suficiente observar que el número (n) de incidencias se puede contar de dos maneras diferentes:

1. $n = bk$, ya que hay b bloques con k variedades;
2. $n = rv$, ya que cada variedad pertenece a r bloques.

□

Ejercicio 12.4. Demostrar que si una matriz $A \in M_{v\times b}(\mathbb{Z}_2)$ cumple las condiciones del ejercicio 12.2, entonces A es la matriz de incidencia de un diseño regular con parámetros (v,k,r).

t-diseños

Se puede aumentar la regularidad de un diseño exigiendo que cada t-subconjunto de variedades, donde $1 \leqslant t \leqslant k \leqslant v$, esté contenido en el mismo número (λ) de bloques. En este caso, el sistema correspondiente se llama *t-diseño* y se denota por

$$t\text{-}(v,k,\lambda)$$

Los diseños regulares son por tanto 1-diseños, tomando $t = 1$ y $\lambda = r$. La proposición 12.3 se puede obtener también como caso particular del teorema siguiente, que nos dice cuál es la relación que tienen que mantener los parámetros de cualquier t-diseño. Su demostración es también una generalización del razonamiento utilizado en el caso de diseños regulares.

Teorema 12.5. El número de bloques de un t-(v,k,λ) diseño es

$$b = \lambda \frac{\binom{v}{t}}{\binom{k}{t}}$$

Demostración. Sea (V,B) un t-(v,k,λ) diseño. Para obtener el resultado contaremos de dos maneras diferentes el número de pares (T,B), donde T es un t-subconjunto de V y $B \in B$ es un bloque que contiene el subconjunto T.

El número de t-subconjuntos contenidos en un bloque es $\binom{k}{t}$ y por tanto hay $b\binom{k}{t}$ de estos pares. Por otra parte, $\binom{v}{t}$ es el número de t-subconjuntos de V y λ es el número de bloques que contienen T y por tanto el número de pares (T,B) es también $\lambda\binom{v}{t}$. □

Esta condición, contrariamente al caso de los 1-diseños, no es suficiente para la existencia del diseño correspondiente. Por ejemplo, se ha demostrado la no existencia de ningún diseño con parámetros 2-$(43,7,1)$ o 10-$(16,72,1)$. Es necesario, por tanto, encontrar recursos que faciliten la obtención en general de t-diseños, o al menos, que nos aseguren su no existencia.

Siguiendo el mismo tipo de razonamiento, se puede generalizar el resultado anterior:

Teorema 12.6. El número de bloques, λ_s, de un t-(v,k,λ_t) diseño que contienen un determinado s-subconjunto $S \subset V$, con $s \leqslant t$, es

$$\lambda_s = \lambda_t \frac{\binom{v-s}{t-s}}{\binom{k-s}{t-s}}$$

Demostración. Sea (V,B) un t-(v,k,λ) diseño. Contemos ahora de dos maneras diferentes el número de pares (T,B), donde T es un t-subconjunto de V que contiene un determinado s-subconjunto S, y $B \in B$ es un bloque que contiene T.

El número de t-subconjuntos que contienen S y están contenidos en un determinado bloque es $\binom{k-s}{t-s}$, y el número de bloques que contienen S es λ_s. Por otra parte, el número de t-subconjuntos que contienen S es $\binom{v-s}{t-s}$, y el número de veces (bloques) que aparece cada t-subconjunto es λ_t. De donde:

$$\lambda_s \binom{k-s}{t-s} = \lambda_t \binom{v-s}{t-s}$$

□

Observemos que el valor λ_s, obtenido en el teorema anterior, no depende del conjunto S que se ha considerado: todos los subconjuntos de tamaño s están contenidos en λ_s bloques. Como consecuencia inmediata deducimos que un t-diseño tiene que ser también un s-diseño para $1 \leqslant s \leqslant t$.

Corolario 12.7. Todo t-diseño es también un s-diseño, para todo $1 \leqslant s \leqslant t$.

Por ejemplo, los subconjuntos de $\{1,2,3,4,5,6,7,8\}$

$$\begin{array}{ll} \{1,2,3,5\} & \{2,3,4,7\} \\ \{1,2,4,8\} & \{2,3,6,8\} \\ \{1,2,6,7\} & \{2,4,5,6\} \\ \{1,3,4,6\} & \{2,5,7,8\} \\ \{1,3,7,8\} & \{3,4,5,8\} \\ \{1,4,5,7\} & \{3,5,6,7\} \\ \{1,5,6,8\} & \{4,6,7,8\} \end{array}$$

constituyen los bloques de un 3-$(8,4,1)$ diseño , ya que, como es fácil comprobar, cualquier 3-subconjunto aparece una única vez. Es fácil comprobar también que cualquier 2-subconjunto aparece en tres bloques diferentes, por ejemplo el par $\{1,3\}$ aparece en los bloques $\{1,2,3,5\}$, $\{1,3,4,6\}$ y $\{1,3,7,8\}$. Así, estos bloques son también bloques de un 2-$(8,4,3)$ diseño y también de un 1-$(8,4,7)$ diseño. Es inmediato comprobar, sin embargo, que no constituyen un 4-diseño, ya que no aparece, por ejemplo, el 4-subconjunto $\{1,2,3,4\}$.

A causa de la dificultad que supone la construcción de un t-diseño en general, es interesante como mínimo poder obtener algunos diseños a partir de otros conocidos. En este sentido, el teorema anterior (12.6) nos proporciona también un recurso útil.

Dado un t-diseño $D = (V,B)$ y un s-subconjunto S de V, con $s \leqslant t$, se define lo que llamaremos diseño *s-derivado* de D respecto de S y que denotaremos por

$$D_S = (V_S, B_S)$$

donde $V_S = V \setminus S$ y

$$B_S = \{B_i \setminus S, B_i \in B : S \subset B_i\}$$

Es decir, un diseño s-derivado es el que se obtiene de un t-diseño suprimiendo $s \leqslant t$ variedades de V y de cada uno de los bloques de D que contienen cada una de estas s variedades. Si S sólo tiene un elemento, entonces se dice que D_S es una *contracción* de D respecto del 1-subconjunto S.

Por ejemplo, los diseños s-derivados del diseño 3-$(8,4,1)$ descrito anteriormente son los diseños 2-$(7,3,1)$ y el 1-$(6,2,1)$, que figuran a continuación y que se obtienen suprimiendo de 3-$(8,4,1)$ los subconjuntos $S = \{8\}$ y $S = \{7,8\}$ respectivamente.

$$\{1,2,4\},\ \{2,3,6\},\ \{1,3,7\},\ \{2,5,7\},\ \{1,5,6\},\ \{3,4,5\},\ \{4,6,7\}$$
$$\{1,3\},\ \{2,5\},\ \{4,6\}$$

Cabe observar que 2-$(7,3,1)$ es una contracción de 3-$(8,4,1)$.

Es inmediato comprobar, como consecuencia del teorema 12.6, que D_S es efectivamente un t-diseño con los parámetros que figuran a continuación.

Corolario 12.8. Si existe un t-(v,k,λ) diseño, existe también el diseño s-derivado con parámetros $(t-s)$-$(v-s,k-s,\lambda)$, para todo $1 \leqslant s \leqslant t$.

En el mismo sentido se puede definir lo que llamaremos diseño *s-residual* de un t-diseño, $D=(V,B)$ respecto de un s-subcunjunto S de V, con $s \leqslant t$,

$$D^S = (V^S, B^S)$$

donde $V^S = V \setminus S$ y

$$B^S = \{B_i \in B : B_i \cap S = \emptyset\}$$

Es decir, un diseño s-residual es el que se obtiene de un t-diseño suprimiendo un subconjunto de cardinal inferior a t del conjunto de variedades V y como bloques se consideran los del diseño original que no contienen ningún elemento de este subconjunto.

Por ejemplo, los diseños s-residuales que se obtienen del diseño 3-$(8,4,1)$ descrito anteriormente suprimiendo del conjunto de variedades los subconjuntos $S=\{8\}$ y $S=\{7,8\}$ son respectivamente los que figuran a continuación:

$$\{1,2,3,5\},\ \{1,2,6,7\},\ \{1,3,4,6\},\ \{1,4,5,7\},\ \{2,3,4,7\},\ \{2,4,5,6\},\ \{3,5,6,7\}$$
$$\{1,2,3,5\},\ \{1,3,4,6\},\ \{2,4,5,6\}$$

Cabe observar que a diferencia de los diseños s-derivados, los diseños s-residuales tienen todos la misma uniformidad, es decir, los bloques que se obtienen son todos del mismo tamaño que los originales.

La proposición siguiente permite deducir que esta construcción proporciona efectivamente un nuevo t-diseño.

Proposición 12.9. El número de bloques de un t-(v,k,λ_t) diseño que no contienen ninguna variedad de un determinado s-subconjunto $S \subset V$, con $s \leqslant t$, es

$$\lambda^s = \lambda_t \frac{\binom{v-s}{k}}{\binom{v-t}{k-t}}$$

Teorema 12.10. Si existe un t-(v,k,λ) diseño, existe también el diseño s-residual con parámetros $(t-s)$-$(v-s,k,\mu)$, para todo $1 \leqslant s < t$.

Ejercicio 12.11. Demostrar el teorema 12.10 usando la proposición 12.9.

Se puede demostrar sin excesiva dificultad que el diseño complementario de un t-diseño es también un t-diseño. Al final de esta sección se deducirán los parámetros del complementario

de un 2-diseño. Usando el mismo tipo de razonamiento, se pueden deducir los parámetros del complementario de un t-diseño en general.

Desgraciadamente no existen resultados generales que faciliten la obtención de t-diseños. Un breve repaso histórico dará una idea de la dificultad del problema y de su estado actual.

Es preciso mencionar que, para $t > 4$, se conocen muy pocos t-diseños. Por ejemplo, no fue hasta 1976 que se obtuvieron los primeros 5-diseños de parámetros

$$5\text{-}(12,6,1), \quad 5\text{-}(24,6,1), \quad 5\text{-}(28,6,1), \quad 5\text{-}(48,6,1), \quad 5\text{-}(84,6,1)$$

En 1978, W. Mills construyó un $5\text{-}(72,6,1)$ diseño y desde entonces hasta 1986, en que D. Kreher y S. Radziszowski encontraron el primer 6-diseño, $6\text{-}(14,7,4)$, no se obtuvo nada nuevo. De hecho, los especialistas en el tema conjeturaban la no existencia de tales diseños. Fue L. Teirlinck, en el año 1987, quien contradijo esta sospecha demostrando la existencia de un t-diseño para cualquier valor de t. Pero los diseños que se obtienen a partir de su demostración tienen unos parámetros extraordinariamente grandes y, por tanto, la obtención de ejemplos pequeños es aún un problema abierto.

Dentro de la clasificación general de los t-diseños hay, sin embargo, dos casos especialmente importantes de los cuales es más fácil obtener información. Éstos son, de hecho, los que dieron origen a esta teoría de diseños y que están relacionados, cronológicamente hablando, con problemas geométricos surgidos en el siglo pasado (J. Steiner, 1844) y con problemas estadísticos tratados en este siglo (R. A. Fisher, 1940):

1. Los llamados *sistemas de Steiner* son t-diseños en los cuales $\lambda = 1$ y se denotan habitualmente por

$$S(t,k,v)$$

2. Los llamados *BIBD* (abreviación de su denominación inglesa *Balanced Incomplete Block Design*) son 2-diseños, es decir, $t = 2$. En este caso se dice que λ es el parámetro que mide el equilibrio del diseño y el 2-diseño correspondiente se llama *equilibrado* y se denota normalmente por

$$(v,b,r,k,\lambda)\text{-BIBD}$$

¿Existen diseños para cualquier valor de estos parámetros? La respuesta a esta cuestión no es hoy en día aún del todo satisfactoria, como veremos a continuación.

Sistemas de Steiner

Cualquier sistema de Steiner, además de las condiciones generales como t-diseño, cumple la condición adicional siguiente:

Teorema 12.12. En cualquier sistema de Steiner, $S(t,k,v)$, se cumple:

$$v \geqslant (t+1)(k-t+1)$$

Demostración. Sea (V,B) un $S(t,v,k)$ diseño. En primer lugar, observemos que en cualquier sistema de Steiner dos bloques diferentes tienen como máximo $(t-1)$ variedades en común.

Observemos también que debe existir algún $(t+1)$-subconjunto que no esté en ningún bloque. Sea X alguno de estos $(t+1)$-subconjuntos.

Para cada uno de los $(t+1)$ t-subconjuntos $T \subset X$ de V, existe un único bloque, $B_T \in B$, que contiene T. Cada uno de estos B_T bloques contiene $k-t$ variedades que no son de X y cada variedad de $V \setminus X$ está contenida como máximo en uno de estos B_T bloques, ya que estos B_T bloques tienen siempre $t-1$ variedades de X en común.

Deducimos, por tanto, que la unión de todos estos B_T bloques contiene

$$v \geqslant |\cup B_T| = |X| + \sum_{T \subset X} |B_T \setminus X| = (t+1) + (t+1)(k-t)$$

variedades, como queríamos demostrar. □

Como consecuencia de este teorema podemos deducir, por ejemplo, la no existencia del 10-diseño que hemos mencionado anteriormente, $S(10,16,72)$, ya que $72 < 11 \cdot 7$.

Los sistemas de Steiner más conocidos y más sencillos son los llamados *sistemas triples de Steiner* constituidos por bloques de tamaño tres, es decir, $S(2,3,v)$ y que habitualmente se denotan por

$$\mathrm{STS}(v)$$

Como ejemplo consideremos $\mathrm{STS}(9)$, que tiene por bloques los que figuran a continuación:

$$\begin{array}{ll} \{1,2,3\} & \{2,4,8\} \\ \{1,4,7\} & \{2,5,9\} \\ \{1,5,8\} & \{2,6,7\} \\ \{1,6,9\} & \{3,4,9\} \\ \{4,5,6\} & \{3,5,7\} \\ \{7,8,9\} & \{3,6,8\} \end{array}$$

Como consecuencia del teorema 12.5, el número de *triplos* (bloques) de un sistema triple de Steiner es $b = v(v-1)/6$ y ésta es por tanto una condición sencilla para deducir la no existencia de un $\mathrm{S}(2,3,v)$. Por ejemplo, no existe ningún $\mathrm{S}(2,3,8)$. Esta expresión nos permite deducir también que el sistema triple de Steiner más pequeño es el $\mathrm{STS}(7)$ descrito a continuación:

$$\begin{array}{ll} \{1,2,4\} & \{2,3,5\} \\ \{1,3,7\} & \{2,6,7\} \\ \{1,5,6\} & \{3,4,6\} \\ & \{4,5,7\} \end{array}$$

Proposición 12.13. El número de bloques en un sistema triple de Steiner, STS(v) es

$$b = v(v-1)/6$$

Ejercicio 12.14. Demostrar que no existe ningún STS(v) para valores de $v = 4, 5, 6$.

Los 2-sistemas de Steiner son un caso particular de BIBD y los trataremos en el apartado siguiente. Un ejemplo de 3-sistema de Steiner está descrito en 12.1. Los sistemes de Steiner más interesantes son los que corresponden a valores de $t > 3$ y son también los de más difícil obtención. Por ejemplo, para $t \geqslant 4$, únicamente se conocen las contracciones de los 5-diseños mencionados anteriormente (12.1):

$$\mathrm{S}(4,5,11), \quad \mathrm{S}(4,5,23), \quad \mathrm{S}(4,5,27), \quad \mathrm{S}(4,5,47), \quad \mathrm{S}(4,5,83)$$

BIBD

Si centramos ahora la atención en los 2-diseños (BIBD), el problema en general de su obtención es también un problema abierto, pero en este caso hay más resultados parciales y de más fácil tratamiento, como veremos a continuación.

En primer lugar sabemos que en cualquier 2-(v,k,λ) diseño se cumple:

1. $bk = vr$
2. $r(k-1) = \lambda(v-1)$

La primera de estas condiciones es consecuencia de la regularidad del diseño y la segunda se obtiene del teorema 12.6 tomando $s = 1$.

Estas condiciones necesarias, sin embargo, sabemos también que no son suficientes. En particular, no existe ningún $(43,7,1)$-BIBD.

El siguiente teorema, conocido como *desigualdad de Fisher*, proporciona una condición muy sencilla para la no existencia de BIBD. Este resultado se obtiene teniendo en cuenta el comportamiento de las matrices de incidencia de estos diseños.

Proposición 12.15. Si A es la matriz de incidencia de un (v,b,r,k,λ)-BIBD, entonces

$$AA^T = (r-\lambda)I + \lambda J$$

Demostración.

$$AA^T = (c_{ij})_{v\times v}, \qquad \begin{cases} c_{ii} &= \sum_{h=1}^{b} a_{ih}^2 = \sum_{h=1}^{b} a_{ih} = r \\ c_{ij} &= \sum_{h=1}^{b} a_{ih}a_{jh} = \lambda \end{cases}$$

Observar que los elementos de la diagonal de AA^T son todos iguales a r como consecuencia de la regularidad del diseño, y el resto son todos iguales a λ como consecuencia de su equilibrio. Así,

$$AA^T = \begin{pmatrix} r & \lambda & \cdots & \lambda \\ \lambda & r & & \lambda \\ \vdots & & \ddots & \vdots \\ \lambda & & \cdots & r \end{pmatrix}$$

□

Teorema 12.16. El número b de bloques de un (v,b,r,k,λ)-BIBD es como mínimo igual al número v de variedades:

$$b \geqslant v$$

Demostración. Calculamos de forma inmediata el determinante de AA^T, restando la primera fila de las otras filas y sumando a la primera columna la suma del resto de columnas, para obtener

$$|AA^T| = \begin{vmatrix} (r+(v-1)\lambda) & \lambda & \cdots & \lambda \\ 0 & (r-\lambda) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & (r-\lambda) \end{vmatrix}, \quad |AA^T| = (r+(v-1)\lambda)(r-\lambda)^{v-1}$$

Usando la condición que proporciona el equilibrio del diseño y teniendo en cuenta que $r > \lambda$, deducimos que

$$|AA^T| = (r+(k-1)r)(r-\lambda)^{v-1} = rk(r-\lambda)^{v-1} \neq 0$$

Finalmente, se obtiene el resultado teniendo en cuenta que

$$v = rang(AA^T) \leqslant rangA \leqslant \min(v,b)$$

□

Si se considera el caso límite de la desigualdad de Fisher, es decir, igual número de bloques que de variedades, se obtienen los llamados *diseños simétricos*, que se denotan habitualmente por

$$(v,k,\lambda)\text{-SD}$$

De la igualdad $bk = vr$, deducimos que en un diseño simétrico, $k = r$.

El ejemplo más pequeño y también más conocido de diseño simétrico es el $(7,3,1)$-SD, que ya hemos considerado anteriormente como sistema triple de Steiner y que trataremos de nuevo en la sección siguiente como ejemplo importante de geometría finita.

La simetría de los diseños es una característica que se pierde con facilidad cuando se manipulan estos diseños para obtener otros. Por ejemplo, el dual de un diseño simétrico no siempre es simétrico.

Ejercicio 12.17. Buscar un ejemplo de un diseño simétrico tal que su dual no lo sea.

Es fácil comprobar, sin embargo, que el diseño complementario de un diseño simétrico es también simétrico.

Proposición 12.18. El diseño complementario de un diseño simétrico (v,k,λ)-SD es también un diseño simétrico de parámetros $(v, v-k, b-2r+\lambda)$-SD, si $b-2r+\lambda > 0$.

Demostración. Sea $D = (V,B)$ un diseño simétrico de parámetros (v,k,λ)-SD. Es inmediato comprobar, a partir de la definición, que el diseño complementario $\bar{D} = (\bar{V},\bar{B})$ cumple $|\bar{V}| = v$, $|\bar{B}| = b = v$ y que todo $\bar{B} \in \bar{B}$ tiene cardinal $|\bar{B}| = v-k$.

El parámetro de equilibrio de $\bar{D}$, $\bar{\lambda}$, se obtiene descontando del número de bloques b de B aquellos que contienen una determinada pareja $\{x,y\}$ de V, λ, y también aquellos bloques que sólo contienen alguno de los elementos x o y, r. Tenemos, por tanto:

$$\bar{\lambda} = b - 2(r-\lambda) - \lambda = b - 2r + \lambda$$

□

Como ejemplo, el diseño $(7,4,2)$-SD que figura a continuación es complementario del $(7,3,1)$-SD descrito anteriormente como sistema triple de Steiner.

$$\begin{array}{ll} \{3,5,6,7\} & \{1,4,6,7\} \\ \{2,4,5,6\} & \{1,3,4,5\} \\ \{2,3,4,7\} & \{1,2,5,7\} \\ & \{1,2,3,6\} \end{array}$$

Se pueden obtener 2-diseños de dos maneras especiales a partir de diseños simétricos, considerando en los dos casos como punto de referencia un bloque del diseño original. Estas construcciones aparecen como generalizaciones de ciertas construcciones geométricas que veremos en la próxima sección. La denominación de estas construcciones puede llevar a confusión, ya que son similares a otras introducidas por t-diseños.

Dado un diseño simétrico $D = (V, B)$, se define el diseño *derivado* de D respecto de un bloque $B \in B$,

$$D_B = (V_B, B_B)$$

como el diseño que tiene por conjunto de variedades el propio B, $V_B = B$ y como conjunto de bloques el que se obtiene de intersectar todos los otros bloques de B con el propio B:

$$B_B = \{B_i \cap B,\ B_i \in B \setminus B\}$$

Si consideramos el diseño $(15, 7, 3)$-SD, con conjunto de bloques

$$\begin{array}{lcllcl}
B_1 & = & \{1,2,3,4,5,6,7\} & B_8 & = & \{2,4,6,8,10,12,14\} \\
B_2 & = & \{1,2,3,8,9,10,11\} & B_9 & = & \{2,4,6,9,11,13,15\} \\
B_3 & = & \{1,2,3,12,13,14,15\} & B_{10} & = & \{2,5,7,8,10,13,15\} \\
B_4 & = & \{1,4,5,8,9,12,13\} & B_{11} & = & \{2,5,7,9,11,12,14\} \\
B_5 & = & \{1,4,5,10,11,14,15\} & B_{12} & = & \{3,4,7,8,11,12,15\} \\
B_6 & = & \{1,6,7,8,9,14,15\} & B_{13} & = & \{3,4,7,9,10,13,14\} \\
B_7 & = & \{1,6,7,10,11,12,13\} & B_{14} & = & \{3,5,6,8,11,13,14\} \\
 & & & B_{15} & = & \{3,5,6,9,10,12,15\}
\end{array}$$

su diseño derivado respecto el bloque B_1 es el diseño que tiene como conjunto de variedades $V_B = B_1 = \{1,2,3,4,5,6,7\}$ y tiene por bloques

$$\begin{array}{lcllcl}
B_1 & = & \{1,2,3\} & B_8 & = & \{2,4,6\} \\
B_2 & = & \{1,2,3\} & B_9 & = & \{2,5,7\} \\
B_3 & = & \{1,4,5\} & B_{10} & = & \{2,5,7\} \\
B_4 & = & \{1,4,5\} & B_{11} & = & \{3,4,7\} \\
B_5 & = & \{1,6,7\} & B_{12} & = & \{3,4,7\} \\
B_6 & = & \{1,6,7\} & B_{13} & = & \{3,5,6\} \\
B_7 & = & \{2,4,6\} & B_{14} & = & \{3,5,6\}
\end{array}$$

Comprobar que identificando los bloques iguales el diseño que se obtiene es isomorfo al diseño simétrico $(7, 3, 1)$-SD.

Se define también el *residual* de D respecto de un bloque $B \in B$,

$$D^B = (V^B, B^B)$$

como el diseño que tiene por conjunto de variedades el que se obtiene suprimiendo B de V, $V^B = V \setminus B$ y como conjunto de bloques el que se obtiene suprimiendo los elementos de B de todos los otros bloques:

$$B^B = \{B_i \setminus B,\ B_i \in B \setminus B\}$$

Como ejemplo, consideremos el diseño residual del diseño $(7,3,1)$-SD respecto del bloque $B = \{3,4,6\}$. Así tenemos $V^B = \{1,2,5,7\}$ y como conjunto de bloques se obtiene

$$B^B = \{\{1,2\},\{2,5\},\{5,7\},\{5,1\},\{7,2\},\{7,1\}\}$$

Se puede comprobar sin dificultad que en los dos casos se obtienen BIBD con los parámetros que figuran a continuación:

Proposición 12.19. Dado un diseño simétrico de parámetros (v,k,λ)-SD, su diseño derivado tiene parámetros $(k, v-1, k-1, \lambda, \lambda-1)$-BIBD.

Proposición 12.20. Dado un diseño simétrico de parámetros (v,k,λ)-SD, su diseño residual tiene parámetros $(v-k, v-1, k, k-\lambda, \lambda)$-BIBD.

Los diseños simétricos constituyen la clase más estudiada de diseños. En particular cumplen una condición muy simple respecto de lo que se llama *orden* del diseño, que se define como $k-\lambda$. Esta condición fue obtenida por Bruch, Ryser y Chowla en 1950 y se conoce directamente como la condición BRC.

Teorema 12.21. Si el número v de variedades de un (v,k,λ)-SD es par, entonces el orden del diseño, $k-\lambda$, es un cuadrado.

Demostración. La matriz de incidencia del diseño (v,k,λ)-SD es cuadrada y entonces

$$|AA^T| = |A|^2 = (r+(v-1)\lambda)(r-\lambda)^{v-1} = k^2(k-\lambda)^{v-1}$$

por tanto, $(k-\lambda)^{v-1}$ debe ser un cuadrado. Si v es par, entonces $k-\lambda$ es un cuadrado. □

Los mismos autores obtuvieron también una condición necesaria para la existencia de un diseño simétrico con un número impar de variedades. La demostración en este caso es más complicada y usa resultados de teoría de números que no tienen cabida en este libro.

Teorema 12.22. Si el número v de variedades de un (v,k,λ)-SD es impar, entonces existen tres números enteros x,y,z tales que

$$z^2 = (k-\lambda)x^2 + (-1)^{(v-1)/2}\lambda y^2$$

No se conoce ningún conjunto de parámetros que cumpla las condiciones de los teoremas 12.21 o 12.22 para el cual no exista el diseño simétrico correspondiente. Pero la suficiencia de este resultado es aún un problema abierto. Por ejemplo, no se ha podido determinar la existencia de un diseño de parámetros $(111,11,1)$-SD.

12.2 Geometrías finitas

Una *geometría finita* es un sistema particular de incidencia en el cual, a partir de una determinada axiomática, se define una cierta familia de subconjuntos de un conjunto finito de elementos llamados puntos. En particular, una geometría finita es un diseño combinatorio en el que se consideran las variedades como puntos.

En función de la axiomática definida aparecen diferentes estructuras geométricas. Unas de las más sencillas son las llamadas *geometrías lineales finitas*, en las cuales la axiomática se refiere a propiedades que deben cumplir ciertos subconjuntos de puntos llamados líneas o rectas. Así, si $P = \{p_1, p_2, \ldots, p_v\}$ representa un conjunto de puntos, el conjunto de líneas será un determinado subconjunto de las partes de P, $L = \{l_1, l_2, \ldots, l_b\} \subset \mathcal{P}(P)$ y la correspondiente geometría se representa por

$$G = (P, L)$$

El número de puntos de una línea $l \in L$ lo notaremos por $|l| = |\{p \in P, p \in l\}|$. Como veremos a continuación, es útil considerar el conjunto de líneas que contienen un determinado punto p. Representamos este conjunto por $L_p = \{l \in L \mid p \in l\}$.

De hecho son las *geometrías casi-lineales* las que tienen la axiomática más simple:

QL0 Para todo $l \in L$, $|l| \geqslant 2$

QL1 Para todo $p, q \in P$, $|L_p \cap L_q| \leqslant 1$

El primero de estos axiomas no es propio de estas geometrías, sino que lo comparten todas las geometrías finitas no triviales. El segundo es por tanto el que las caracteriza y asegura que dos puntos cualesquiera están como máximo en una línea. Si imponemos que haya exactamente una línea que los contenga, obtenemos la axiomática propia de una *geometría lineal finita*:

L0 Para todo $l \in L$, $|l| \geqslant 2$

L1 Para todo $p, q \in P$, $|L_p \cap L_q| = 1$

Planos proyectivos

La geometría proyectiva tiene sus orígenes en el siglo IV (Pappus de Alejandría) pero no fue hasta el siglo XVI, mediante los pintores flamencos, que se le dio importancia, y aún se demoró tres siglos más para hacer sistemático y riguroso su estudio (Boole, Cayley, Sylvester, siglo XIX). La versión finita de las geometrías proyectivas tiene múltiples aplicaciones en combinatoria relacionadas con la construcción de ciertos diseños simétricos y también con la obtención de lo que se llaman *cuadrados latinos*, que estudiaremos en la próxima sección.

Añadiendo condiciones a las descritas anteriormente para geometrías lineales, se obtienen tipos especiales de geometrías lineales finitas. En particular si se considera que dos líneas diferentes siempre tienen un único punto en común, y que existen como mínimo cuatro puntos no colineales tres a tres (esto, como veremos, evita casos triviales), se obtiene lo que se llama *plano proyectivo finito*. Así, un plano proyectivo es una geometría lineal finita que cumple los axiomas siguientes:

P0 Para todo $l \in L$, $|l| \geqslant 2$

P1 Para todo $p, q \in P$, $|L_p \cap L_q| = 1$

P2 Para todo $l, l' \in L$, $|l \cap l'| = 1$

P3 Existen $p, q, s, t \in P$, no colineales tres a tres

Observar que en un plano proyectivo todas las rectas se cortan, de manera que no se satisface el axioma de Euclides, que afirma que por un punto exterior a una recta pasa una única paralela.

El plano proyectivo más pequeño es el plano de Fano que aparece en la figura 12.1.

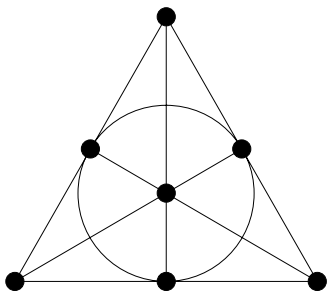


Figura 12.1: Plano de Fano

¿Existen planos proyectivos con cualquier número de puntos? Comprobaremos que la respuesta a esta pregunta es negativa si descartamos los casos llamados *degenerados* que se muestran en la figura 12.2.

Ejercicio 12.23. Comprobar que no existe ningún plano proyectivo no degenerado de cuatro puntos.

Es preciso observar en primer lugar que el axioma **P3** se impone para eliminar los casos degenerados. Es preciso observar también que **P1** y **P2** son condiciones duales, es decir, que se obtienen una de la otra intercambiando puntos por líneas. En particular, también se verifica el dual de **P3**.

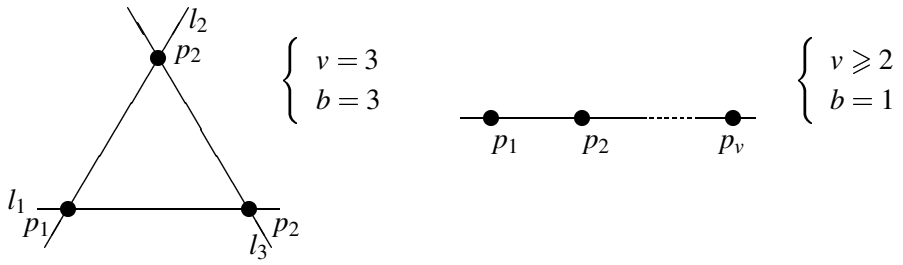


Figura 12.2: Planos proyectivos degenerados

Proposición 12.24. En un plano proyectivo hay al menos cuatro líneas tales que tres cualesquiera de ellas no contienen un mismo punto.

Demostración. Sean $p, q, r, s \in P$ cuatro puntos no colineales tres a tres. Entonces tres de las líneas $l_{pq}, l_{qr}, l_{ps}, l_{rs}$ no tienen ningún punto en común. Por ejemplo, si l_{pq}, l_{ps}, l_{rs} tuviesen un punto t en común, las líneas l_{pqt}, l_{pst} y l_{rst} coincidirían, ya que las dos primeras y las dos últimas tendrían dos puntos en común y psr serían colineales. □

Este resultado junto con los tres axiomas **P1**, **P2** y **P3** hacen que cualquier resultado sobre planos proyectivos tenga su dual (intercambiando puntos por líneas). Se dice por tanto que los planos proyectivos verifican lo que se llama *principio de dualidad*.

El comportamiento regular de los planos proyectivos se evidencia mediante los siguientes resultados que ponen de manifiesto al mismo tiempo este principio de dualidad.

Teorema 12.25. En un plano proyectivo, todas las líneas contienen el mismo número de puntos y cada punto pertenece al mismo número de líneas.

Demostración. Para demostrar que cada línea contiene el mismo número de puntos, establecemos una biyección entre los puntos de dos líneas diferentes l y l'. Para ello, consideremos $x \in P$ tal que no sea un punto de $l \cup l'$.

La *proyección* sobre l' de cada punto $p \in l$ respecto al punto x es el punto $p' \in l'$ que se obtiene como intersección de la línea l_{xp} con l':

$$p' = l' \cap l_{xp}$$

tal como se puede ver en la figura 12.3.

Observar que si $p, q \in l$, $p \neq q$, entonces $p' \neq q'$ (axioma **P2**). Por tanto, la proyección es una biyección.

Por el principio de dualidad, también es cierto que cada punto pertenece al mismo número de líneas. □

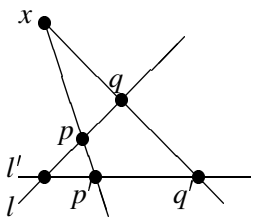


Figura 12.3: Proyección respecto a x

Teorema 12.26. En un plano proyectivo, el número de puntos que contiene cada línea es igual al número de líneas que pasan por cada punto.

Demostración. Sea (P,L) un plano proyectivo. Consideremos $l \in L$ y $x \in P \setminus l$. Entonces la aplicación que a cada punto $p \in l$ le asigna la línea l_{px} es una biyección entre el conjunto de puntos de l y el conjunto de líneas que pasa por x. □

Teorema 12.27. Un plano proyectivo con $m+1$ puntos en cada línea tiene m^2+m+1 puntos.

Demostración. Si v es el número de puntos de un plano proyectivo, entonces

$$v = (m+1)m+1$$

donde $(m+1)$ es $|L_p|$, es decir, el número de líneas que contienen un determinado punto $p \in P$, y m es $|l|-1$, $l \in L_p$.

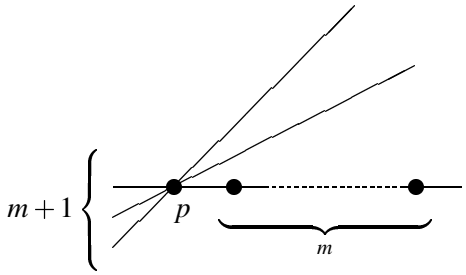


Figura 12.4: Número de puntos de un plano proyectivo

Observemos que éstos son efectivamente todos los puntos del plano, ya que, si existiese algún punto q que no fuese de L_p, también tendría que existir (axioma **P1**) la línea $l_{pq} \in L_p$ que lo une a p. □

El principio de dualidad nos garantiza que el número de líneas de un plano proyectivo es el mismo que el número de puntos, es decir, m^2+m+1.

En particular, no existen planos proyectivos con 5 o 6 puntos. Observar también que para $m = 1$, el plano que se obtiene es degenerado. Por tanto, tal como se ha comentado anteriormente, el plano proyectivo más pequeño es el plano de Fano con $m = 2$. Para $m = 3$ se obtiene un plano proyectivo con 13 puntos, que está representado en la figura 12.5.

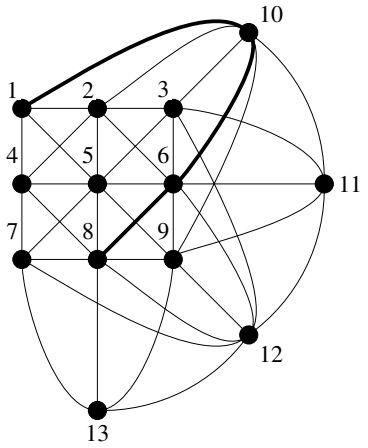


Figura 12.5: Plano proyectivo de 13 puntos

Al número m se le dice *orden* del plano proyectivo. Un plano proyectivo de orden m se denota habitualmente por

$$PG(2,m)$$

De momento sabemos que cualquier plano proyectivo debe tener m^2+m+1 puntos; sin embargo, ¿existe un plano proyectivo para cualquier valor de m? Podemos obtener la respuesta identificando los planos proyectivos con unos ciertos diseños simétricos. Si se consideran los puntos de $PG(2,m)$ como variedades y las líneas como bloques, se obtiene un 2-diseño simétrico con los parámetros que figuran a continuación:

$$\begin{array}{rcl} (V,B) & \leftrightarrow & (P,L) \\ (m^2+m+1,m+1,1)\text{-SD} & \leftrightarrow & PG(2,m) \\ v=b & \leftrightarrow & m^2+m+1 \\ k=r & \leftrightarrow & m+1 \\ \lambda & \leftrightarrow & 1 \end{array}$$

Se puede observar como consecuencia del axioma **P1** que $\lambda = 1$, y se puede comprobar también que se cumplen las condiciones de regularidad de todo diseño simétrico. Por ejemplo,

el plano proyectivo de la figura 12.5 corresponde al diseño $(13,4,1)$-SD, cuyos bloques figuran a continuación:

$$\begin{array}{lcllcl} B_1 & = & \{1,2,3,11\} & B_8 & = & \{3,5,7,10\} \\ B_2 & = & \{1,4,7,13\} & B_9 & = & \{3,6,9,13\} \\ B_3 & = & \{1,5,9,12\} & B_{10} & = & \{3,12,8,4\} \\ B_4 & = & \{1,10,6,8\} & B_{11} & = & \{4,5,6,11\} \\ B_5 & = & \{2,6,12,7\} & B_{12} & = & \{7,8,9,11\} \\ B_6 & = & \{2,4,10,9\} & B_{13} & = & \{10,11,12,13\} \\ B_7 & = & \{2,5,8,13\} & & & \end{array}$$

Está claro, por tanto, que todo plano proyectivo es una representación geométrica de un determinado diseño simétrico. En sentido contrario también es cierto, es decir, cualquier $(m^2+m+1,m+1,1)$-SD cumple los axiomas de plano proyectivo, como veremos a continuación:

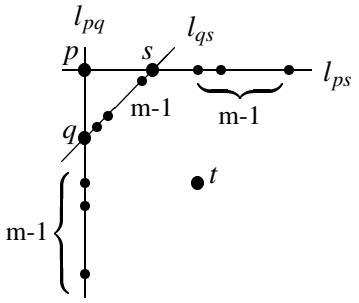
Teorema 12.28. $PG(2,m)$ existe si y sólo si existe el diseño simétrico

$$(m^2+m+1,m+1,1)\text{-SD}$$

Demostración. Es suficiente demostrar que los axiomas de los planos proyectivos se cumplen en el diseño $(m^2+m+1,m+1,1)$-SD para todo $m \geqslant 2$.

Los dos primeros axiomas se deducen directamente a partir de la propia definición de los parámetros k y λ, y **P2** se cumple como consecuencia de la simetría del diseño, que en términos geométricos equivale a la dualidad de **P1**.

Para demostrar que también se cumple el axioma **P3**, será preciso encontrar cuatro puntos no colineales tres a tres. Para ello, consideremos la única línea ($\lambda = 1$) que contiene dos puntos cualesquiera $p,q \in P$, $l_{pq} \in L$. Como el diseño es incompleto, existe $s \in P \setminus l_{pq}$ y por tanto



podemos considerar las líneas que unen el punto s con los puntos p y q, l_{sp} y l_{sq} (observar que

son únicas, ver la figura 12.2). Entonces,

$$|l_{pq} \cup l_{sp} \cup l_{sq}| = 3(m-1)+3 = 3m < m^2+m+1$$

De donde deducimos que

$$\exists t \in P \setminus l_{pq} \cup l_{sp} \cup l_{sq}$$

□

El resultado siguiente nos proporciona condiciones sencillas para determinar la no existencia de planos proyectivos para determinados órdenes.

Teorema 12.29. Si existe un diseño $(m^2+m+1, m+1, 1)$-SD con $m \equiv 1,2 \pmod 4$, entonces existe $a, b \in \mathbb{Z}$ tal que $m = a^2+b^2$.

En particular, no existen planos proyectivos de orden 6, 14, 21, 22, ... La demostración de este resultado es consecuencia del teorema 12.22 para diseños simétricos en general. Esquemáticamente,

$$v-1 = m(m+1) \equiv 0 \pmod 2$$

de donde, claramente, v es impar para todo valor de m, y el teorema 12.22 asegura en este caso la existencia de una terna de números enteros $(x,y,z) \in (\mathbb{Z}^3)^*$ tal que

$$z^2 = \underbrace{(k-\lambda)}_{m} x^2 + (-1)^{(v-1)/2} \underbrace{\lambda}_{1} y^2$$

Si $m \equiv 1,2 \pmod 4$, se obtiene que $z^2+y^2 = mx^2$ y un resultado de teoría de números asegura que, en este caso, existe $a, b \in \mathbb{Z}$ tal que $m = a^2+b^2$.

Es preciso mencionar, sin embargo, que este teorema no proporciona condiciones suficientes para la existencia de planos proyectivos de estos órdenes. Por ejemplo, se ha mencionado ya la falta de información sobre la existencia de un diseño de parámetros $(111, 11, 1)$-SD, que evidentemente cumple las condiciones del teorema. Por tanto, la existencia del plano proyectivo de orden 10, $PG(2,10)$, es un problema por resolver.

Si $m \equiv 0,3 \pmod 4$, del mismo teorema 12.22 se obtiene que $z^2-y^2 = mx^2$ y por ejemplo los puntos $(1,(m-1)/2,(m+1)/2)$ y $(1,(m-4)/4,(m+4)/4)$ son soluciones de esta ecuación, con lo que no se puede concluir nada sobre la existencia o no de un plano proyectivo con estos órdenes.

En cuanto a resultados concretos de existencia de planos proyectivos, al final de la sección dedicada a cuadrados latinos se verá un procedimiento constructivo que asegura la existencia de $P(2,m)$ cuando m es una potencia de un número primo. Esta construcción está basada en los cuerpos de Galois.

Planos afines

La geometría afín está intrínsecamente relacionada con la geometría proyectiva, aunque, de hecho, la geometría afín sigue los postulados de la geometría euclídea, mientras que en la geometría proyectiva, como ya hemos visto, no es así. La relación que hay entre las dos se pondrá de manifiesto, como veremos, a través del diseño residual del diseño simétrico asociado a un plano proyectivo.

Se dice que una geometría finita $G = (P,L)$ es un *plano afín* si cumple las condiciones siguientes:

A0 Para todo $l \in L$, $|l| \geqslant 2$

A1 Para todo $p,q \in P$, $|L_p \cap L_q| = 1$

A2 Para todo $l \in L$, y para todo $p \in P \setminus l$, existe una única línea $l' \in L_p$ tal que $|l \cap l'| = 0$

A3 Existen $p,q,s \in P$, no colineales

Los axiomas de planos afines y planos proyectivos difieren sólo en los dos últimos y, de hecho, la diferencia substancial es entre los axiomas **P2** y **A2**. Cabe observar que son los axiomas **P2** y **A2** los que contraponen estas geometrías lineales respecto de la geometría euclídea.

Denotamos por $G^{l^*} = G^* = (P^*, L^*)$ el diseño residual que se obtiene suprimiendo una línea $l^* \in L$ y sus puntos del diseño $G = (P,L) = (m^2+m+1, m+1, 1)$-SD. Demostraremos que este diseño residual es un plano afín. En primer lugar, los parámetros del diseño que se obtiene son los siguientes:

$$\begin{array}{rcl} G = (P,L) & \leftrightarrow & G^* = (P^*, L^*) \\ (m^2+m+1, m+1, 1)\text{-SD} & \leftrightarrow & (m^2, m^2+m, m, m+1, 1)\text{-BIBD} \end{array}$$

$$\begin{array}{rcl} |P| = m^2+m+1 & \leftrightarrow & |P^*| = m^2 \\ |L| = m^2+m+1 & \leftrightarrow & |L^*| = m^2+m \\ |l| = m+1 & \leftrightarrow & |l| = m \\ |L_p| = m+1 & \leftrightarrow & |{L^*}_p| = m+1 \\ |L_p \cap L_q| = 1 & \leftrightarrow & |{L^*}_p \cap {L^*}_q| = 1 \end{array}$$

Ejercicio 12.30. Comprobar que los parámetros del diseño residual de un plano proyectivo $PG(2,m)$ son efectivamente los que se han descrito.

Teorema 12.31. El diseño residual de un plano proyectivo es un plano afín.

Demostración. Comprobemos que $G^* = (P^*, L^*)$ cumple los axiomas de plano afín. Los dos primeros se deducen directamente a partir de la definición de diseño residual.

A0 Para todo $l \in L^*$, $|l| = m \geqslant 2$

A1 Para todo $p, q \in P^*$, $|L_p^* \cap L_q^*| = 1$

A2 Para que se cumpla este axioma es preciso comprobar que para todo $l \in L^*$, y para todo $p \in P^* \setminus l$, existe una única línea $l_p \in L_p^*$ tal que $|l \cap l_p| = 0$.

Para ello, consideremos la única línea $l_{pp^*} \in L$, donde p^* es el punto $l^* \cap l$ (véase la figura 12.6). Como $l \cap l_{pp^*} = p^*$ en G, entonces, $l \cap l_{pp^*} = \emptyset$ en G^*.

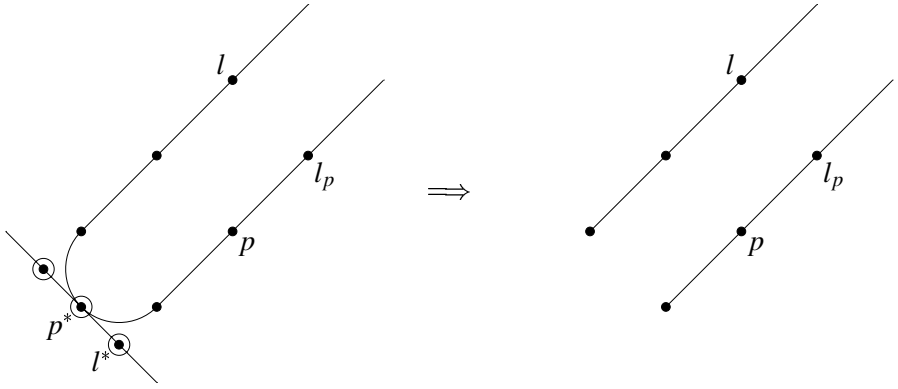


Figura 12.6: Obtención de un plano afín a partir del plano proyectivo

A3 este axioma es consecuencia directa de **P3**, teniendo en cuenta que suprimiendo una línea en un plano proyectivo siempre quedan como mínimo tres puntos no colineales.

□

Denotamos por $AG(2,m)$ el plano afín de orden m, es decir, el plano afín que tiene m puntos en cada línea.

Podemos invertir el proceso que hemos seguido para obtener un plano afín a partir de un plano proyectivo, añadiendo al plano afín la línea l^*, formada por los puntos correspondientes a las intersecciones entre líneas que no tienen intersección en $AG(2,m)$. Así,

$$G = G^* \cup l^*$$

En la figura 12.7 hay representado el plano afín de orden dos y el correspondiente plano proyectivo.

Ejercicio 12.32. Comprobar que efectivamente $PG(2,m) = AG(2,m) \cup l^*$.

Teorema 12.33. $AG(2,m)$ existe si y sólo si existe $PG(2,m)$.

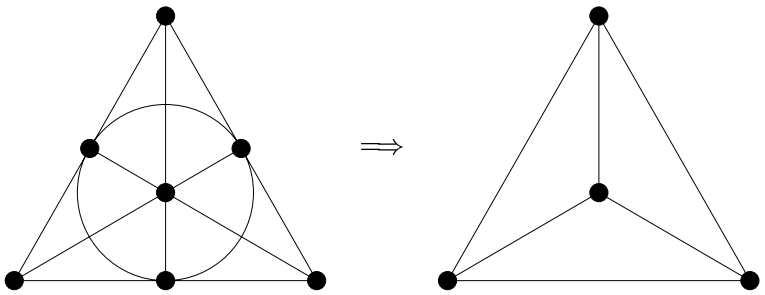


Figura 12.7: $AG(2,2)$ y $PG(2,2)$

12.3 Cuadrados latinos

Un *cuadrado latino* de orden n es una matriz de orden $n \times n$ cuyos términos son elementos de un conjunto cualquiera S de tamaño n, de manera que cada fila y cada columna contenga todos los elementos de S.

Por ejemplo, la matriz

$$\begin{array}{ccc} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{array}$$

es un cuadrado latino de orden 3.

Está claro que cada fila y cada columna de un cuadrado latino es una permutación de los elementos de S. También está claro que un cuadrado latino es un diseño completo simétrico con n bloques repetidos, cada uno de ellos igual a S.

Es fácil ver que existen cuadrados latinos de cualquier orden. Sólo es preciso identificar S con un grupo G del mismo orden y considerar como cuadrado latino Q, la tabla de su operación, de manera que a $q_{ij} \in Q$ le corresponda $g_k \in G$ si y sólo si $g_k = g_i g_j$. Cabe observar que, de esta manera, ningún elemento se repite en ninguna fila ni en ninguna columna. Por ejemplo, los dos cuadrados latinos siguientes se corresponden con las tablas de los grupos $\mathbb{Z}_4$ y $\mathbb{Z}_2 \times \mathbb{Z}_2$ (en este último se identifica $(0,0)$ con 0, $(0,1)$ con 1, $(1,0)$ con 2 y $(1,1)$ con 3):

$$\begin{array}{cccc} 0 & 1 & 2 & 3 \\ 1 & 2 & 3 & 0 \\ 2 & 3 & 0 & 1 \\ 3 & 0 & 1 & 2 \end{array} \qquad \begin{array}{cccc} 0 & 1 & 2 & 3 \\ 1 & 0 & 3 & 2 \\ 2 & 3 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{array}$$

Dos cuadrados latinos de tamaño n son *equivalentes* si es posible deducir uno del otro mediante una permutación de los símbolos. Por ejemplo, los cuadrados que figuran a continuación

son equivalentes:

$$\begin{array}{cc} 1 & 2 \\ 2 & 1 \end{array} \qquad \begin{array}{cc} 2 & 1 \\ 1 & 2 \end{array}$$

$$\begin{array}{ccc} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{array} \qquad \begin{array}{ccc} 3 & 2 & 1 \\ 2 & 1 & 3 \\ 1 & 3 & 2 \end{array}$$

En general, las tablas de grupos no isomorfos proporcionan ejemplos de cuadrados latinos no equivalentes.

Como el nombre de los elementos de S es irrelevante, supondremos que $S = \{1, 2, \ldots, n\}$. Un cuadrado latino $Q = (q_{ij})$ de orden n se dice *normalizado* si los términos de la primera columna aparecen en el orden natural, es decir, para todo i, $q_{i1} = i$. Observemos que siempre podemos obtener cuadrados latinos normalizados permutando los nombres de los símbolos.

Un cuadrado latino $Q = (q_{ij})$ de orden n se llama *idempotente* si los términos de la diagonal aparecen en el orden natural, es decir, para todo i, $q_{ii} = i$.

Cuadrados latinos mutuamente ortogonales

El llamado *problema de los 36 oficiales* (L. Euler, 1782) dio origen a lo que se conoce hoy en día por *cuadrados latinos mutuamente ortogonales*.

El problema consistía en encontrar, dadas 6 graduaciones y 6 regimientos, una formación de 6×6 oficiales tal que en cada fila y en cada columna hubiese un oficial de cada regimiento y de cada graduación.

Una posible ordenación de los oficiales de manera que en cada fila y en cada columna haya sólo un oficial de cada graduación es la que figura a continuación:

$$\begin{array}{cccccc} G_1 & G_6 & G_2 & G_5 & G_3 & G_4 \\ G_4 & G_2 & G_6 & G_3 & G_1 & G_5 \\ G_2 & G_5 & G_3 & G_6 & G_4 & G_1 \\ G_5 & G_3 & G_1 & G_4 & G_6 & G_2 \\ G_6 & G_1 & G_4 & G_2 & G_5 & G_3 \\ G_3 & G_4 & G_5 & G_1 & G_2 & G_6 \end{array}$$

Una posible ordenación de los oficiales de manera que en cada fila y en cada columna haya sólo un oficial de cada regimento es la que figura a continuación:

$$\begin{array}{cccccc} R_1 & R_2 & R_3 & R_4 & R_5 & R_6 \\ R_3 & R_1 & R_2 & R_6 & R_4 & R_5 \\ R_2 & R_3 & R_1 & R_5 & R_6 & R_4 \\ R_4 & R_5 & R_6 & R_1 & R_2 & R_3 \\ R_6 & R_4 & R_5 & R_3 & R_1 & R_2 \\ R_5 & R_6 & R_4 & R_2 & R_3 & R_1 \end{array}$$

Es preciso observar que cada fila y cada columna de estas ordenaciones contiene todas las graduaciones (regimientos) o, equivalentemente, ninguna fila o columna contiene ninguna graduación (regimiento) más de una vez.

El problema tendrá solución si, superponiendo las dos ordenaciones, cada par (G_i, R_j) aparece en la formación una única vez. En este caso, esto no es así, como se puede observar en el cuadro siguiente:

$$\begin{array}{cccccc} 11 & 62 & 23 & 54 & 35 & 46 \\ 43 & 21 & 62 & 36 & 14 & 55 \\ 22 & 53 & 31 & 65 & 46 & 14 \\ 54 & 35 & 16 & 41 & 62 & 23 \\ 66 & 14 & 45 & 23 & 51 & 32 \\ 35 & 46 & 54 & 12 & 23 & 61 \end{array}$$

Por ejemplo, la pareja 62 aparece tres veces, mientras que las parejas 33 o 44, entre otras, no aparecen.

Euler usaba el alfabeto griego para denotar las graduaciones y el alfabeto romano para denotar los regimientos, y por ello denominaba greco-romanos a estos cuadrados. Esta es también la razón por la cual los cuadrados latinos se denominan de esta manera.

¿Es posible obtener alguna ordenación de manera que este problema tenga solución?

Euler conjeturó que si $n \equiv 2 \pmod 4$, entonces no existe ningún cuadrado greco-romano de orden n. En esta sección trabajaremos sobre esta conjetura.

El problema de los 36 oficiales es un problema que exige la existencia de dos cuadrados latinos tales que superponiéndolos aparezcan todas las posibles parejas o, de forma equivalente, no se repita ninguna. Para comenzar definiremos este concepto.

Dos cuadrados latinos $A = (a_{ij})$ y $B = (b_{ij})$ de tamaño n son *ortogonales* si los n^2 pares ordenados $(a_{ij}, b_{ij}) \in A \times B$ son todos diferentes. Lo denotaremos escribiendo

$$A \perp B \qquad (\text{o } B \perp A)$$

En primer lugar es preciso observar que no existen cuadrados latinos ortogonales de tamaño 2. Si tomamos $n = 3$, podemos considerar un ejemplo clásico introducido por Fisher (1926). De hecho, fue Fisher quien recuperó y utilizó de forma sistemática los cuadrados latinos para tratar esencialmente experimentos sobre agricultura. En su ejemplo se trataba de estudiar la incidencia conjunta de tres fertilizantes $\{f_1, f_2, f_3\}$ y tres insecticidas $\{i_1, i_2, i_3\}$ sobre un campo dividido en tres parcelas $\{P_1, P_2, P_3\}$, durante tres años consecutivos, $\{A_1, A_2, A_3\}$. Por ello, es preciso combinar en cada año y cada parcela una pareja formada por un fertilizante y un insecticida de manera que todas las parejas hayan sido probadas.

Las figuras 12.8 y 12.9 muestran que las dos condiciones son compatibles en el sentido de que es posible obtener todas las combinaciones, es decir, existen dos cuadrados latinos ortogonales de tamaño tres.

	A_1	A_2	A_3
P_1	f_1	f_2	f_3
P_2	f_2	f_3	f_1
P_3	f_3	f_1	f_2

	A_1	A_2	A_3
P_1	i_1	i_2	i_3
P_2	i_3	i_1	i_2
P_3	i_2	i_3	i_1

Figura 12.8: Cuadrados latinos de tamaño tres

	A_1	A_2	A_3
P_1	11	22	33
P_2	23	31	12
P_3	32	13	21

Figura 12.9: Cuadrados latinos ortogonales de tamaño tres

En la figura 12.10 hay un ejemplo de tres cuadrados latinos ortogonales dos a dos de orden 4. Desde el punto de vista del diseño de experimentos, se pueden interpretar como tres aspectos diferentes de un mismo fenómeno que se quieren contrastar dos a dos. Es inmediato comprobar que estos tres cuadrados son mutuamente ortogonales. Esto nos lleva a la definición siguiente.

Se dice que una familia $A_1, A_2, \ldots, A_k$ de cuadrados latinos del mismo orden constituye un conjunto de *MOLS* (*Mutually Orthogonal Latin Squares*) si $A_i \perp A_j$ para todo i, j, $i \neq j$.

Es natural plantearse la cuestión siguiente. ¿En qué condiciones existe una familia de MOLS? En esta sección trataremos este problema y daremos un método constructivo para encontrar familias de MOLS para ciertos valores de orden n.

1	2	3	4
2	1	4	3
3	4	1	2
4	3	2	1

1	3	4	2
2	4	3	1
3	1	2	4
4	2	1	3

1	4	2	3
2	3	1	4
3	2	4	1
4	1	3	2

Figura 12.10: Cuadrados latinos mutuamente ortogonales

Proposición 12.34. Si A y B son cuadrados latinos ortogonales, los correspondientes cuadrados normalizados también lo son.

Demostración. Sean $A = (a_{ij})$ y $B = (b_{ij})$ dos cuadrados latinos ortogonales. Si efectuamos las permutaciones σ_1 y σ_2 de $\{1,2,\ldots,n\}$ en A y B respectivamente, cada par (a_{ij}, b_{ij}) se transforma en $(\sigma_1(a_{ij}), \sigma_2(b_{ij}))$, de manera que continúa habiendo todos los pares y la condición de ortogonalidad se mantiene. Sólo es preciso entonces efectuar las permutaciones necesarias para normalizar cada uno de los cuadrados. □

Proposición 12.35. Si existe una familia de k MOLS de orden n, entonces $k \leqslant n-1$.

Demostración. Sea $A_1, A_2, \ldots, A_k$ una familia de MOLS normalizada de orden n. Si $A_p = (a_{ij}^p)$ y $A_q = (a_{ij}^q)$ son dos cuadrados cualesquiera de esta familia, sólo es preciso observar que

$$(a_{12}^p, a_{12}^q) \neq (i,i), \qquad 1 \leqslant i \leqslant n$$

ya que, si no, $(a_{12}^p, a_{12}^q) = (a_{1i}^p, a_{1i}^q)$, cosa que contradice la condición de ortogonalidad entre A_p y A_q. Por tanto,

$$a_{12}^1, a_{12}^2, \ldots, a_{12}^k$$

son todos diferentes y, consecuentemente, $k \leqslant n-1$. □

Un *conjunto completo de MOLS* de orden n es un conjunto de $n-1$ MOLS de orden n.

Si $N(n)$ representa el número máximo de MOLS de orden n, sabemos que $N(2) = 1$, $N(3) = 2$ y $N(4) = 3$.

¿Para qué valores de n, $N(n) = n-1$? Es decir, ¿para qué valores de n existe un conjunto completo de MOLS de orden n?

El teorema siguiente, debido a Bose (1938), garantiza la existencia de un conjunto completo de MOLS para cualquier potencia de un número primo. La demostración es constructiva y consiste esencialmente en identificar el conjunto de variedades con los elementos de un cuerpo del mismo orden.

Teorema 12.36. Si p es un número primo, $N(p^k) = p^k - 1$, para todo $k \in \mathbb{N}$.

Demostración. Identifiquemos el conjunto de $p^k = n$ variedades con el cuerpo de Galois del mismo orden.

$$V \leftrightarrow GF(n) = \{f_0 = 0, f_1 = 1, f_2, \ldots, f_{n-1}\}$$

donde $f_i = \alpha^{i-1}$, $2 \leqslant i \leqslant n-1$, siendo α un elemento primitivo del cuerpo.

A partir de los elementos del cuerpo definimos la familia siguiente de matrices:

$$\begin{array}{rcll} A_l & = & (a^l_{ij}) & 1 \leqslant l < n-1 \\ a^l_{ij} & = & f_l f_j + f_i & 0 \leqslant i, j \leqslant n-1 \end{array}$$

(aquí los índices van de 0 a $n-1$ en lugar de ir de 1 a n como es habitual). En primer lugar, demostraremos que estas matrices son cuadrados latinos. Para ello, comprobemos que los elementos de cada fila y de cada columna son todos diferentes. Si $a^l_{ij} = a^l_{ik}$, entonces $f_l f_j + f_i = f_l f_k + f_i$ y, como $f_l \neq 0$, deducimos que $f_j = f_k$ y por tanto $j = k$. Razonando de forma similar, se demuestra que los elementos de cualquier columna son todos diferentes.

Demostraremos ahora que la familia de matrices definida constituye un conjunto completo de MOLS. Para ello es preciso demostrar que cualquier par de estas matrices son ortogonales. Supongamos que no lo son. Entonces, existen dos pares iguales ocupando posiciones diferentes, es decir,

$$(a^l_{ij}, a^m_{ij}) = (a^l_{hk}, a^m_{hk})$$

de donde

$$\left\{ \begin{array}{rcl} f_l f_j + f_i & = & f_l f_k + f_h \\ f_m f_j + f_i & = & f_m f_k + f_h \end{array} \right.$$

Restando estas dos igualdades deducimos que $f_j = f_k$ y, por tanto, $j = k$ e $i = h$. □

Como ejemplo de aplicación del teorema anterior, veamos cómo se construyen conjuntos completos de MOLS de órdenes tres y cuatro. Para los de orden tres, consideramos $GF(3) = (\mathbb{Z}_3, +, \cdot)$ y obtenemos los dos cuadrados ortogonales a partir de las igualdades:

$$\begin{array}{rclrcl} a^1_{0j} & = & j & a^2_{0j} & = & 2j \\ a^1_{1j} & = & j+1 & a^2_{1j} & = & 2j+1 \\ a^1_{2j} & = & j+2 & a^2_{2j} & = & 2j+2 \end{array}$$

de donde

$$A_1 = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 0 \end{pmatrix} \qquad A_2 = \begin{pmatrix} 0 & 2 & 1 \\ 1 & 0 & 2 \\ 2 & 1 & 0 \end{pmatrix}$$

Para los de orden cuatro, consideremos $GF(4) = (\mathbb{Z}_2[x]/(x^2+x+1), +, \cdot)$, donde

$$\mathbb{Z}_2[x]/(x^2+x+1) = \{f_0 = 0,\ f_1 = 1,\ f_2 = x,\ f_3 = x+1\}$$

En este caso, los cuadrados ortogonales se obtienen a partir de las igualdades siguientes:

$$\begin{array}{lll} a^1_{0j} = f_j & a^2_{0j} = f_2 f_j & a^3_{0j} = f_3 f_j \\ a^1_{1j} = f_j + 1 & a^2_{1j} = f_2 f_j + 1 & a^3_{1j} = f_3 f_j + 1 \\ a^1_{2j} = f_j + f_2 & a^2_{2j} = f_2 f_j + f_2 & a^3_{2j} = f_3 f_j + f_2 \\ a^1_{3j} = f_j + f_3 & a^2_{3j} = f_2 f_j + f_3 & a^3_{3j} = f_3 f_j + f_3 \end{array}$$

$$A_1 = \begin{pmatrix} 0 & 1 & x & x+1 \\ 1 & 0 & x+1 & x \\ x & x+1 & 0 & 1 \\ x+1 & x & 1 & 0 \end{pmatrix} \qquad A_2 = \begin{pmatrix} 0 & x & x+1 & 1 \\ 1 & x+1 & x & 0 \\ x & 0 & 1 & x+1 \\ x+1 & 1 & 0 & x \end{pmatrix}$$

$$A_3 = \begin{pmatrix} 0 & x+1 & 1 & x \\ 1 & x & 0 & x+1 \\ x & 1 & x+1 & 0 \\ x+1 & 0 & x & 1 \end{pmatrix}$$

Para simplificar la notación, podemos expresar las anteriores matrices identificando $f_2 = x$ con 2 y $f_3 = x+1$ con 3, y obtener

$$A_1 = \begin{pmatrix} 0 & 1 & 2 & 3 \\ 1 & 0 & 3 & 2 \\ 2 & 3 & 0 & 1 \\ 3 & 2 & 1 & 0 \end{pmatrix} \qquad A_2 = \begin{pmatrix} 0 & 2 & 3 & 1 \\ 1 & 3 & 2 & 0 \\ 2 & 0 & 1 & 3 \\ 3 & 1 & 0 & 2 \end{pmatrix} \qquad A_3 = \begin{pmatrix} 0 & 3 & 1 & 2 \\ 1 & 2 & 0 & 3 \\ 2 & 1 & 3 & 0 \\ 3 & 0 & 2 & 1 \end{pmatrix}$$

Realmente no es necesario hacer todos estos cálculos. Los conjuntos completos de MOLS que proporciona el teorema 12.36 siguen un comportamiento general sencillo que explicitamos a continuación:

Método constructivo de un conjunto completo de $n-1$ MOLS para $n = p^k$, p primo.

1. A_1 es la tabla del grupo aditivo $(GF(n), +)$, ya que $f_1 = 1$ y, por tanto, $a^1_{ij} = f_j + f_i$
2. Todos los cuadrados tienen la primera columna igual, ya que $f_0 = 0$ y, por tanto, $a^l_{i0} = f_i$

3. Las columnas restantes de cada A_i se obtienen haciendo la permutación cíclica $(23\ldots n)$ de las columnas del cuadrado anterior A_{i-1}. Este comportamiento se debe al carácter cíclico de $(GF(n)^*,\cdot)$, como demostramos a continuación. Si α es un elemento primitivo de $GF(n)$,

$$f_i = \alpha^{i-1} \qquad 1 \leqslant i \leqslant n-1$$

Entonces, para $2 \leqslant j \leqslant n-1$ y $l \geqslant 2$,

$$a_{ij}^{l+1} = f_{l+1}f_j + f_i = \alpha^l\alpha^{j-1} + f_i = \alpha^{l-1}\alpha^j + f_i = f_l f_{j+1} + f_i = a_{i(j+1)}^l$$

Es preciso mencionar que Wernicke (1910) enunció el recíproco del teorema anterior, es decir, si existe un conjunto completo de MOLS de orden n, entonces n es una potencia de un primo. Se detectaron errores en la demostración de este resultado y hasta ahora continúa siendo un problema abierto.

Una manera de obtener nuevas familias de MOLS a partir de otras la proporciona el resultado siguiente, obtenido por MacNeish (1922).

Teorema 12.37. Si existen dos familias de k MOLS de órdenes respectivos n y m, entonces existe una nueva familia de k MOLS de orden nm.

Demostración. Sean F_1 y F_2 dos familias de k MOLS de órdenes respectivos n y m:

$$\begin{aligned} F_1 &= \{A_1, A_2, \ldots, A_k\} \\ F_2 &= \{B_1, B_2, \ldots, B_k\} \end{aligned}$$

Definimos una nueva familia $F_3 = \{C_1, C_2, \ldots, C_k\}$ de orden nm a través del producto cartesiano de las matrices de las familias anteriores:

$$C_l = A_l \times B_l = \begin{pmatrix} (a_{11}^l, B_l) & \cdots & (a_{1n}^l, B_l) \\ \vdots & \ddots & \vdots \\ (a_{n1}^l, B_l) & \cdots & (a_{nn}^l, B_l) \end{pmatrix}$$

donde

$$(a_{ij}^l, B_l) = \begin{pmatrix} (a_{ij}^l, b_{11}^l) & \cdots & (a_{ij}^l, b_{1m}^l) \\ \vdots & \ddots & \vdots \\ (a_{ij}^l, b_{m1}^l) & \cdots & (a_{ij}^l, b_{mm}^l) \end{pmatrix}$$

Es preciso demostrar que F_3 es una familia de MOLS.

En primer lugar, comprobemos que los elementos de F_3 son cuadrados latinos. Para ello sólo es preciso observar que dos elementos cualesquiera de una fila (columna) de C_l, $1 \leqslant l \leqslant k$ son diferentes ya que A_l y B_l son cuadrados latinos:

$$\begin{cases} (a^l_{ij}, b^l_{uv}) & \neq & (a^l_{ij'}, b^l_{uv'}) \\ (a^l_{ij}, b^l_{uv}) & \neq & (a^l_{i'j}, b^l_{u'v}) \end{cases}$$

Comprobemos ahora que los elementos de F_3 son mutuamente ortogonales. Para ello, supongamos que existen dos parejas de elementos iguales a $(C_l, C_h) \in F_3 \times F_3$,

$$((a^l_{ij}, b^l_{uv}), (a^h_{ij}, b^h_{uv})) = ((a^l_{i'j'}, b^l_{u'v'}), (a^h_{i'j'}, b^h_{u'v'}))$$

Entonces, igualando componentes y teniendo en cuenta que $A_l \perp A_h$ y $B_l \perp B_h$, deducimos que las posiciones también tienen que coincidir. □

Como consecuencia directa de este teorema tenemos el resultado siguiente:

Teorema 12.38. Si $n = p_1^{n_1} p_2^{n_2} \cdots p_s^{n_s}$ es la descomposición en factores primos de n, entonces

$$N(n) \geqslant \min\{p_i^{n_i} - 1, 1 \leqslant i \leqslant s\}$$

El único caso en que el más pequeño de los $p_i^{n_i}$ en la descomposición en factores primos de n es 2 se da cuando n es par pero no divisible por cuatro. Así pues,

Corolario 12.39. Si $n \not\equiv 2 \pmod 4$, entonces $N(n) \geqslant 2$.

Teniendo en cuenta este resultado, deducimos por ejemplo que hay como mínimo dos cuadrados latinos de orden 12 mutuamente ortogonales, $N(12) \geqslant 2$. Si $n = 2m$, donde m es un número impar, sólo deducimos que $N(n) \geqslant 1$. Recordemos que la conjetura de Euler decía que, en este caso, $N(n) = 1$. Para $m = 1$, está claro que $N(2) = 1$. Para $m = 3$, Terry obtuvo en 1901 todas las posibles parejas de cuadrados latinos de orden 6 (9.408, considerando sólo cuadrados latinos normalizados) y no encontró ninguno que fuese ortogonal. Por tanto, el problema de los 36 oficiales no tiene solución y parte de la conjetura de Euler es cierta. Pero no fue hasta 1960 que Bose, Parker y Shrikhande demostraron, mediante diseños experimentales, que la conjetura es falsa, excepto justamente en los casos conocidos $n = 2, 6$.

Teorema 12.40. $N(n) \geqslant 2, \quad n \equiv 2 \pmod 4, \quad n \neq 2, 6.$

La demostración de este resultado es muy larga y, por este motivo, se ha intentado encontrar demostraciones más sencillas, pero de momento no se ha conseguido.

MOLS y planos proyectivos

La existencia de un conjunto completo de MOLS, tal como veremos a continuación, equivale a la existencia de un plano proyectivo. La demostración de este resultado es constructiva y pensamos que muy instructiva. En particular, proporciona un método para construir los planos proyectivos de orden $n = p^k$, p primo, a partir de la familia completa de MOLS que se ha descrito en el teorema 12.36 del apartado anterior.

Teorema 12.41. Existe un plano proyectivo de orden m, si y sólo si existe un conjunto completo de MOLS de orden m.

Demostración. Sea $PG(2,m)$ un plano proyectivo de orden m. Consideramos una línea cualquiera l de $PG(2,m)$ y escogemos dos puntos arbitrarios p y q de l. Consideramos ahora las intersecciones entre los conjuntos de líneas siguientes:

$$\begin{aligned} L_p \setminus l &= \{l_p^1, l_p^2, \ldots, l_p^m\} \\ L_q \setminus l &= \{l_q^1, l_q^2, \ldots, l_q^m\} \end{aligned}$$

que denotamos por

$$p_{ij} = l_p^i \cap l_q^j$$

y $p_1, \ldots, p_{m-1}$ son los puntos de l diferentes de p y q.

Definimos la familia de matrices de orden $m \times m$,

$$F = \{A_k = (a_{ij}^k),\ 1 \leqslant k \leqslant m-1\}$$

de manera que a_{ij}^k representa la línea de $PG(2,m)$ que pasa por los puntos p_{ij} y $p_k \in l \setminus \{p,q\}$.

Demostraremos que F es un conjunto completo de MOLS. En primer lugar, es preciso comprobar que los elementos de F son cuadrados latinos. Cualquier matriz A_k sólo tiene m elementos diferentes, ya que por p_k sólo pasan $(m+1)$ rectas, y de éstas la recta l no intersecta con p_{ij}. Por otra parte, los elementos de una fila (columna) son todos diferentes. Efectivamente, si $a_{ij}^k = a_{ij'}^k$ ($a_{ij}^k = a_{i'j}^k$), entonces $p_{ij} = p_{ij'}$ ($p_{ij} = p_{i'j}$), ya que $a_{ij}^k, a_{ij'}^k \in L_{p_k}$ ($a_{ij}^k, a_{i'j}^k \in L_{p_k}$), y, por tanto, $j = j'$ ($i = i'$).

Es preciso comprobar también que los elementos de F son mutuamente ortogonales. Supongamos lo contrario, es decir, que $A_k, A_k' \in F$ son tales que alguna pareja, $(a_{ij}^k, a_{ij}^{k'})$, aparece más de una vez, o sea

$$(a_{ij}^k, a_{ij}^{k'}) = (a_{i'j'}^k, a_{i'j'}^{k'})$$

Entonces, igualando componentes, obtenemos que los puntos p_k, p_{ij} y $p_{i'j'}$ están en una misma línea l', y, de forma similar, los puntos $p_{k'}$, p_{ij} y $p_{i'j'}$ están sobre otra línea l'', de manera que $l' \cap l'' = \{p_{ij}, p_{i'j'}\}$, lo que contradice el axioma **P2**.

En sentido contrario, sea ahora $F = \{A_1, A_2, \ldots, A_{m-1}\}$ un conjunto completo de MOLS. Podemos construir un plano proyectivo de orden n de la forma siguiente.

Consideremos una malla cuadrada de tamaño $m \times m$ y definamos el conjunto de líneas formado por las líneas horizontales y las líneas verticales, es decir, $2m$ líneas. Definamos también el conjunto de puntos formado por las intersecciones de las líneas de la malla, más los puntos x e y que se obtienen de intersectar las líneas horizontales entre ellas y las verticales entre ellas. De esta manera tenemos $m^2 + 2$ puntos.

Consideremos ahora, para cada cuadrado latino $A_i \in F$, el conjunto L_i formado por todas las líneas que se obtienen al unir las diferentes posiciones que ocupa un mismo elemento de A_i y denotemos por x_i el punto, exterior a la malla, donde hacemos intersectar estas líneas. Observemos que, por el hecho de ser A_i un cuadrado latino, $|L_i| = m$, y que, por el hecho de ser F una familia de MOLS, $x_i \neq x_j$ si $i \neq j$. Finalmente añadimos una nueva línea que contiene los $(m-1)$ puntos x_i y además el x y el y.

De esta manera tenemos un conjunto de puntos P y un conjunto de líneas L tales que:

$$\begin{cases} |P| & = & m^2 + |\{x,y\}| + |\{x_i \mid 1 \leqslant i \leqslant m\}| = m^2 + m + 1 \\ |L| & = & 2m + (m-1)m + 1 = m^2 + m + 1 \\ |l| & = & m+1, \forall l \in L \\ |L_p| & = & m+1, \forall p \in P \\ |L_p \cap L_q| & = & 1, \forall p, q \in P \end{cases}$$

Éste es un 2-diseño simétrico con parámetros $(m^2 + m + 1, m + 1, 1)$-SD y, por tanto, un plano proyectivo de orden m, $PG(2,m)$. □

En la figura 12.11 se muestra por ejemplo la construcción del plano proyectivo de orden tres, $PG(2,3)$, a partir del siguiente conjunto completo de MOLS:

$$A_1 = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 1 \end{pmatrix} \qquad A_2 = \begin{pmatrix} 0 & 2 & 1 \\ 1 & 0 & 2 \\ 2 & 1 & 0 \end{pmatrix}$$

Como consecuencia del teorema anterior, deducimos que existe un plano proyectivo de orden cualquier potencia de un primo. De la misma manera, se podría haber relacionado la existencia de un conjunto completo de MOLS de orden m con la de un plano afín del mismo orden (recordemos que hay un plano proyectivo de orden m si y sólo si hay un plano afín del mismo orden).

Corolario 12.42. Si $m = p^k$, con p primo, entonces existen $PG(2,m)$ y $AG(2,m)$.

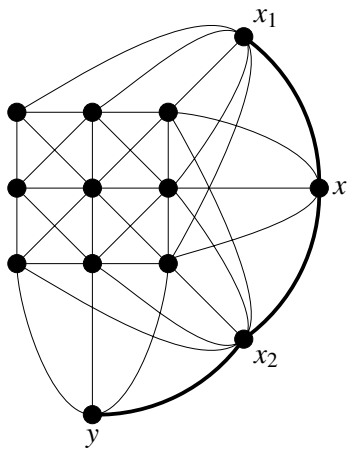


Figura 12.11: $PG(2,3)$

Notas bibliográficas

Son muchos los libros de combinatoria que dedican una parte importante al estudio de los diseños, y hay otros más específicos que se dedican exclusivamente al estudio de los diseños. Aquí hemos escogido una muestra que creemos suficiente para cubrir un margen amplio de niveles de exigencia.

El texto de Anderson [1] presenta una visión global de las posibilidades del tema que puede ser útil en un nivel básico. En el libro de Wallis [6] se tratan todos los aspectos básicos relacionados con el tema de forma clara y más extensa. El texto de Street y Street [4] puede ser un libro complementario y presenta diferentes métodos para la construcción explícita de diseños que no han tenido un espacio en este libro. En este sentido recomendamos también el libro de Hall [2], que además contiene de forma muy comprensible toda la información básica sobre esta teoría. Los lectores más interesados se pueden dirigir a [3] o a [5]. En el primero se tratan todas estas cuestiones desde un punto de vista más formal, mientras que, en el segundo, el estilo es más conciso pero hay más información. Ambos textos son de un nivel más exigente que los anteriores.

Bibliografía

[1] I. Anderson. *A First Course in Combinatorial Mathematics*, Oxford University Press, 1979.

[2] M. Hall. *Combinatorial Theory*, John Wiley & Sons, 1986.

[3] D. R. Hughes, F. C. Piper. *Design Theory*, Cambridge University Press, 1988.

[4] A. P. Street, D. J. Street. *Combinatorics of Experimental Designs*, Oxford Science, 1986.

[5] J. H. van Lint, R. M. Wilson. *A Course in Combinatorics*, Cambridge University Press, 1993.

[6] W. D. Wallis. *Combinatorial Designs*, Marcel Dekker, 1988.

Problemas

1. Estudiar los valores de $r = \lambda_1, \lambda_2$ y λ_3 en la siguiente estructura combinatoria E definida sobre el conjunto $V = \{1,2,3,4,5,6,7\}$ y que tiene por conjunto de bloques los que figuran a continuación:

$$\begin{array}{lcllcl} B_1 & = & \{1,2,4\} & B_8 & = & \{1,2,4\} \\ B_2 & = & \{1,3,7\} & B_9 & = & \{1,3,7\} \\ B_3 & = & \{1,5,6\} & B_{10} & = & \{1,5,6\} \\ B_4 & = & \{2,3,5\} & B_{11} & = & \{2,3,5\} \\ B_5 & = & \{2,6,7\} & B_{12} & = & \{2,6,7\} \\ B_6 & = & \{3,4,6\} & B_{13} & = & \{3,4,6\} \\ B_7 & = & \{4,5,7\} & B_{14} & = & \{4,5,7\} \end{array}$$

   Comprobar que si se define la relación de equivalencia R que identifica bloques iguales, entonces, E/R es isomorfo a STS(7).

2. Comprobar que el conjunto de bloques que figuran a continuación, obtenidos a partir del conjunto $V = \{1,2,3,4,5,6,7\}$, no es un 2-diseño:

$$\begin{array}{lcllcl} B_1 & = & \{1,2,4\} & B_8 & = & \{1,2,4\} \\ B_2 & = & \{1,3,7\} & B_9 & = & \{1,3,7\} \\ B_3 & = & \{1,5,6\} & B_{10} & = & \{1,5,6\} \\ B_4 & = & \{2,3,5\} & B_{11} & = & \{2,3,6\} \\ B_5 & = & \{2,6,7\} & B_{12} & = & \{2,5,7\} \\ B_6 & = & \{3,4,6\} & B_{13} & = & \{3,4,5\} \\ B_7 & = & \{4,5,7\} & B_{14} & = & \{4,6,7\} \end{array}$$

   Comprobar también que la estructura definida en este ejercicio no es isomorfa a la definida en el ejercicio anterior.

3. Estudiar los parámetros de la estructura siguiente e interpretarla como una 2-estructura y también como una 1-estructura:

$$V = \{1,2,3,4,5,6,7\} \qquad \begin{aligned} B_1 &= \{1,2,3,6\} \\ B_2 &= \{1,2,5,7\} \\ B_3 &= \{1,3,4,5\} \\ B_4 &= \{1,4,6,7\} \\ B_5 &= \{2,3,4,7\} \\ B_6 &= \{2,4,5,6\} \\ B_7 &= \{3,5,6,7\} \end{aligned}$$

   Comparar la matriz de incidencia de esta estructura con la matriz de incidencia de $\mathrm{STS}(7)$.

4. Demostrar, de forma constructiva, que la condición de la proposición 12.3, $bk = rv$ es también suficiente para la existencia de un diseño regular con parámetros (v,k,r).

5. Un diseño regular $D = (V,B)$ con parámetros (v,k,r) se dice *trivial* si cada k-subconjunto de V está contenido como mínimo en un bloque de B. Demostrar que un diseño D es trivial si y sólo si D es un t-diseño para todo t tal que $0 \leqslant t \leqslant k$.

6. Demostrar que un 2-diseño con $v = 8$ y $k = 3$ es trivial.

7. Demostrar que no puede existir un sistema de Steiner con parámetros $S(5,7,13)$.

8. Demostrar que si existe un sistema de Steiner $S(3,4,v)$, entonces $v = 6n+2$, o bien, $v = 6n+4$, para algún natural n. (Se demuestra que éstas son también condiciones suficientes.)

9. Demostrar que no puede existir ningún 4-diseño con parámetros $(11,7,2)$.

10. Demostrar que para el diseño 5-$(24,8,1)$, todos los λ_4, λ_3, λ_2 y λ_1 son enteros.

11. Demostrar que si D es un 2-diseño con parámetros (v,k,λ), entonces son equivalentes las afirmaciones siguientes:

    (a) $b = v$;

    (b) $k = r$;

    (c) D^T es un 2-diseño;

    (d) D y D^T son diseños simétricos con parámetros (v,k,λ)-SD.

12. Demostrar que no existen 2-diseños simétricos con parámetros:

(a) $(4,7,1)$-SD

(b) $(22,7,2)$-SD

(c) $(29,8,2)$-SD

13. Demostrar que hay un único 3-$(8,4,1)$ diseño (salvo isomorfismos).

14. Examinar todos los posibles conjuntos de parámetros para un diseño simétrico con $\lambda = 1$ y $k \leqslant 24$. Decidir cuándo:

    (a) existe algún diseño simétrico con estos parámetros;

    (b) no existe diseño simétrico con estos parámetros;

    (c) no se puede decidir.

15. Examinar todos los posibles conjuntos de parámetros con $\lambda = 2$ y $k \leqslant 16$. Decidir cuándo:

    (a) no puede existir ningún diseño simétrico $(v,k,2)$-SD;

    (b) el teorema BRC no dá información para la existencia de diseños simétricos con estos parámetros.

16. Demostrar que, si un cuadrado latino de tamaño n tiene un subcuadrado latino de tamaño $m < n$, entonces $2m \leqslant n$.

17. Demostrar que el siguiente cuadrado latino no se corresponde con la tabla multiplicativa de ningún grupo finito.

1	2	3	4	5
2	1	5	3	4
3	4	1	5	2
4	5	2	1	3
5	3	4	2	1

18. Demostrar que no existe ningún cuadrado latino ortogonal con

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

19. Demostrar que el único cuadrado latino normalizado de orden 4 que admite cuadrados latinos ortogonales es

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

20. Usando el ejercicio anterior, demostrar que, salvo isomorfismos, existe sólo un plano afín y un plano proyectivo de orden 4.

21. De forma similar al ejercicio anterior, demostrar que los planos afines y proyectivos de orden 3 son únicos.

22. Un cuadrado latino se dice *auto-ortogonal* si es ortogonal a su propio transpuesto.

    (a) Demostrar que, en un cuadrado latino auto-ortogonal, los elementos de la diagonal tienen que estar ordenados consecutivamente, es decir: 1, 2, 3, ...

    (b) Demostrar que no existen cuadrados latinos auto-ortogonales de tamaño 3.

    (c) Encontrar cuadrados latinos auto-ortogonales de tamaño 4 y tamaño 5.